University of South Bohemia in České Budějovice
Faculty of Science

Evolution and genomics of symbionts in Hippoboscidae

Master thesis

Bc. Eva Šochová

Supervisor: RNDr. Filip Husník

České Budějovice 2016

Šochová, E., 2014: Evolution and genomics of symbionts in Hippoboscidae. Mgr. Thesis, in English. – 44 p ., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Anottation:

Obligately blood-sucking parasites harbour symbiotic bacteria providing them B-vitamins and cofactors missing from their blood diet. Within Hippoboscoidea group (parasites of birds and mammals), tsetse flies as medically important vectors have been studied extensively while bat flies and louse flies tend to be neglected. This thesis is composed of two complementary manuscripts focused on phylogeny and origin of bacterial symbionts in Hippoboscidae family (manuscript 1) and their genome evolution (manuscript 2). First, phylogenetic approach was employed to determine lineages of obligate and facultative symbionts present in this group. Second, genomic and phylogenomic analyses were carried out to better understand evolution of obligate endosymbionts from the *Arsenophonus* genus in this group. Results of the two studies indicate that relationships between Hippoboscoidea and their symbionts are extremely dynamic with frequent replacements of obligate symbionts. This hypothesis is supported by both phylogenetic and genomic evidence, in particular, *Arsenophonus* endosymbionts of Hippoboscidae represent several distinct lineages (of likely different ages) with noticeable differences in genome features and metabolic capabilities. The data presented in this thesis thus greatly extend our knowledge about evolution and genomics of symbiotic bacteria in Hippoboscidae and bloodsucking hosts in general.

Prohlašuji, že svoji diplomovou práci jsem vypracovala samostatně, pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své diplomové práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích, 22. dubna 2016

Eva Šochová

Acknowledgements:

Firstly, I would like to thank my supervisor Filip Husník for his guidance and great amount of valuable advices. Secondly, I have to thank everybody who collected samples used in this study. Finally, my special thank belongs to my boyfriend to stand by me in these such hard days.

Content

Introduction

The term symbiosis was firstly implemented by a German botanist and mycologist Albert Bernhard Frank in the nineteen century as a co-existence of two different organisms [1]. According to this broad definition, symbiosis includes three different interactions - mutualism, commensalism, and parasitism. Biology is complex and these terms are, of course, arbitrary, so it is not always possible to distinguish between them. For example, a typical reproductive parasite *Wolbachia* can in some situations behave like a mutualist, e.g by protecting the host from viruses [2, 3]. In a similar way, each of these associations can switch to a different one or form gradients. For instance, a mutualist can become a parasite when it is over-abundant [4].

Probably the most essential step in the eukaryotic evolution was the origin of mitochondria (and later on plastids) via endosymbiosis of an archaeal cell with bacteria of α-proteobacterial and cyanobacterial origin [5, 6]. Obligate intracellular symbionts of insects seem to resemble eukaryotic organelles in many mechanistic ways (e.g. by their small genome size and host dependence [7]), but there are also some clear differences [8, 9]. In most cases, they are needed to supplement nutritionally unbalanced diets of their hosts. They reside in a specialized host-derived symbiotic organ called bacteriome and strictly co-evolve with their hosts for millions of years due to vertical transmission. On the contrary, facultative endosymbionts do not co-evolve with their hosts and can also use horizontal transmission or reproductive manipulation(s) to spread through the host population. Their presence in the host is not restricted to special cells and they can invade variety of organs [8, 10, 11]. Endosymbioses are often quite dynamic, with symbiont loss, replacement or complementation usually taking place once the ancient endosymbiont reaches genome size of less than ~500 genes [12–16]. This phenomenon was extensively studied in sap-feeding insect, but very little attention has been paid to blood-feeding systems.

„*Nothing in biology makes sense except in the light of evolution*" (Dobzhansky 1973). This master thesis uses bloodsucking flies from the Hippoboscoidea superfamily as a model group to unravel general mechanisms of endosymbiosis evolution and genomics in bloodsucking insects. The superfamily consists of four obligately blood-sucking families: Glossinidae, Nycteribidae, Streblidae, and Hippoboscidae. Glossinidae is a basal and species-poor clade which harbours an obligate endosymbiont *Wigglessworthia glossinidia* [17] and a facultative endosymbiont *Sodalis glossinidius* [18], Nycteribidae, Streblidae, and Hippoboscidae (together called Pupipara) are species-rich lineages associated with

Arsenophonus bacteria [19–21] and *Sodalis* bacteria in the Hippoboscidae family [19, 22, 23]. However, the evolution of symbiosis in this group is believed to be much more complex and likely influenced by symbiont replacements and horizontal transmission of symbionts because *Arsenophonus* and *Sodalis* bacteria belong to widespread endosymbiotic clades infecting a great number of hosts [20, 24, 25].

The primary aim of my master thesis was to complement work from my bachelor thesis [26] with genome data and comprehensive phylogenetic and phylogenomic analyses. In particular, to infer evolutionary history of Hippoboscidae and their symbionts. However, I have generated enough data to prepare two separate manuscripts which complement each other and to co-author one more article [27]. The first manuscript focuses on phylogeny of the Hippoboscidae family and its three endosymbionts and will be submitted to an evolutionary journal (such as BMC Evolutionary Biology or similar). The second manuscript focuses on comparative genomics of two *Arsenophonus* endosymbionts from avian Hippoboscidae (*Ornithomya biloba* and *Crataerina pallida*) and will be submitted to more genomics-oriented journal (such as Genome Biology and Evolution or similar). Here I present drafts of both these manuscripts as my master thesis.

(1) Complex Evolution of Symbiosis in Louse Flies

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

Authors' contributions:
ES: ~70%
FH: ~15%
EN: ~5%
AH: ~5%
VH: ~5%

FH and VH designed the study. ES obtained most of the sequence data, prepared alignments, inferred phylogenies, and prepared draft manuscript. ES and FH participated in evolutionary interpretation of results. FH participated in manuscript preparation. EN provided some sequence data from her previous work. AH collected samples of African louse flies. ES, FH, EN, and VH read and approved the final manuscript.

(2) Insight into genomes of obligate *Arsenophonus* endosymbionts of two avian louse flies, *Ornithomya biloba* and *Crataerina pallida*

Eva Šochová[1,*], Filip Husník[2, 3], Petr Heneberg[4], Václav Hypša[1, 3]

Authors' contributions:

ES: ~60%

FH: ~30%

PH: ~5%

VH: ~5%

FH and ES designed the study. ES prepared gDNA of *O. biloba* for Illumina MiSeq 300-200 bp paired-end sequencing, carried out assembly of genomes, inferred phylogenies, reconstructed B-vitamin pathways, and prepared draft manuscript. FH prepared gDNA of *O. biloba* and *C. pallida* for Illumina HiSeq 100 bp paired-end sequencing and performed microscopy examination of their endosymbionts and also participated in manuscript preparation. ES and FH participated in COG comparisons and interpretation of results. PH collected samples of *O. biloba*. ES, FH, and VH read and approved the final manuscript.

References

1. Frank AB, Trappe JM: **On the nutritional dependence of certain trees on root symbiosis with belowground fungi (an English translation of A.B. Frank's classic paper of 1885).** *Mycorrhiza* 2005, **15**:267–75.

2. Teixeira L, Ferreira A, Ashburner M: **The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*.** *PLoS Biol* 2008, **6**:e2.

3. Bian G, Xu Y, Lu P, Xie Y, Xi Z: **The endosymbiotic bacterium *Wolbachia* induces resistance to dengue virus in *Aedes aegypti*.** *PLoS Pathog* 2010, **6**:e1000833.

4. Weeks AR, Turelli M, Harcombe WR, Reynolds KT, Hoffmann AA: **From parasite to mutualist: rapid evolution of *Wolbachia* in natural populations of *Drosophila*.** *PLoS Biol* 2007, **5**:e114.

5. Moreira D, López-García P: **Symbiosis between methanogenic Archaea and δ-Proteobacteria as the origin of Eukaryotes: The Syntrophic Hypothesis**. *J Mol Evol* 1998, **47**:517–530.

6. Williams TA, Foster PG, Cox CJ, Embley TM: **An archaeal origin of eukaryotes**

supports only two primary domains of life. *Nature* 2013, **504**:231–6.

7. Moran NA, Bennett GM: **The Tiniest Tiny Genomes**. 2014(May):195–215.

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

9. Bennett GM, Moran NA: **Heritable symbiosis : The advantages and perils of an evolutionary rabbit hole**. 2015, **2015**.

10. Baumann P: **Biology bacteriocyte-associated endosymbionts of plant sap-sucking insects.** *Annu Rev Microbiol* 2005, **59**:155–89.

11. Moran NA, McCutcheon JP, Nakabachi A: **Genomics and evolution of heritable bacterial symbionts.** *Annu Rev Genet* 2008, **42**(July):165–90.

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

13. Conord C, Despres L, Vallier A, Balmand S, Miquel C, Zundel S, Lemperiere G, Heddi A: **Long-term evolutionary stability of bacterial endosymbiosis in curculionoidea: additional evidence of symbiont replacement in the dryophthoridae family.** *Mol Biol Evol* 2008, **25**:859–68.

14. McCutcheon JP, McDonald BR, Moran NA: **Convergent evolution of metabolic roles in bacterial co-symbionts of insects.** *Proc Natl Acad Sci U S A* 2009, **106**:15394–9.

15. Bennett GM, Moran NA: **Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a phloem-feeding insect.** *Genome Biol Evol* 2013, **5**:1675–88.

16. Koga R, Moran NA: **Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont.** *ISME J* 2014, **8**:1237–46.

17. Aksoy S: ***Wigglesworthia*** **gen. nov. and** ***Wigglesworthia glossinidia*** **sp. nov., taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies.** *Int J Syst Bacteriol* 1995, **45**:848–51.

18. Dale C, Maudlin I: ***Sodalis*** **gen. nov. and** ***Sodalis glossinidius*** **sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly** ***Glossina morsitans morsitans*****.** *Int J Syst Bacteriol* 1999, **49 Pt 1**:267–75.

19. 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*****.** *Appl Environ Microbiol* 2006, **72**:2997–3004.

20. Nováková E, Hypša V, Moran NA: ***Arsenophonus*****, an emerging clade of intracellular**

symbionts with a broad host distribution. *BMC Microbiol* 2009, **9**:143.

21. Hosokawa T, Nikoh N, Koga R, Satô M, Tanahashi M, Meng X-Y, Fukatsu T: **Reductive genome evolution, host-symbiont co-speciation and uterine transmission of endosymbiotic bacteria in bat flies.** *ISME J* 2012, **6**:577–87.

22. Nováková E, Hypša V: **A new *Sodalis* lineage from bloodsucking fly *Craterina melbae* (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont *Sodalis glossinidius*.** *FEMS Microbiol Lett* 2007, **269**:131–5.

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

24. Morse SF, Bush SE, Patterson BD, Dick CW, Gruwell ME, Dittmar K: **Evolution, multiple acquisition, and localization of endosymbionts in bat flies (Diptera: Hippoboscoidea: Streblidae and Nycteribiidae).** *Appl Environ Microbiol* 2013, **79**:2952–61.

25. Duron O, Schneppat UE, Berthomieu A, Goodman SM, Droz B, Paupy C, Nkoghe JO, Rahola N, Tortosa P: **Origin, acquisition and diversification of heritable bacterial endosymbionts in louse flies and bat flies.** *Mol Ecol* 2014, **23**:2105–17.

26. ŠOCHOVÁ E: **Intracelulární symbionti krevsajících dvoukřídlých skupiny Hippobosccoidea**. *ŠOCHOVÁ, Eva. Intracelulární symbionti krevsajících dvoukřídlých skupiny Hippobosccoidea. Č. Bud., 2014. bakalářská práce (Bc.). JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH. Přírodovědecká fakulta* 2014.

27. 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.

Complex Evolution of Symbiosis in Louse Flies

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

*Correspondence: sochova.e@seznam.cz
[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, ASCR, v.v.i., České Budějovice, Czech Republic
[4] Department of Biodiversity, School of Molecular and Life Sciences, Faculty of Science and Agriculture, University of Limpopo, South Africa

Abstract

Background: Symbiotic interactions between insects and bacteria are pervasive and represent a continuum of associations from greatly intimate (obligate symbiosis) to less stable (facultative symbiosis). Blood-sucking insects are no exception to this pattern. Obligate endosymbionts are hypothesized to supplement B-vitamins and cofactors missing from the insect blood diet while the role of facultative endosymbionts is less understood in these systems. Here we focus on the stability and dynamics of obligate symbioses in one bloodsucking group (Hippoboscidae) and analyse it using phylogenetic approach.

Results: We have inferred phylogenies of the host lineage and three genera of symbionts. Phylogeny of Hippoboscoidea was difficult to resolve as different genes/analyses frequently inferred contradictory topologies. We confirmed monophyly of Glossinidae, but monophyly of Nycteribiidae, Streblidae, and Hippoboscidae was not strongly supported.

In total, we obtained 65 endosymbiont 16S rRNA gene sequences: 27 for *Arsenophonus*, 12 for *Sodalis*, and 26 for *Wolbachia*. We detected a new obligate lineage of *Sodalis* co-evolving with Olfersini group. In addition to this obligate lineage, there are also several facultative lineages of *Sodalis* in Hippoboscidae. In a similar way, *Arsenophonus* endosymbionts represent obligate endosymbiotic lineages co-evolving with their hosts, as well as facultative infections incongruent with the host phylogeny. Finally, *Wolbachia* strains in Hippoboscidae fall into three supergroups: A, B, and the most common F.

Conclusions: We have untangled surprisingly dynamic, yet selective, evolution of symbiosis within louse flies. The dynamicity is strongly shaped by endosymbiont replacements, but 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.

Keywords: Symbiont replacement, *Arsenophonus*, *Sodalis*, Louse flies, Phylogeny

Background

Symbiosis is a ubiquitous interaction appearing in all domains of life. Such symbiotic associations are very common between insects and their symbiotic bacteria and are often considered to be evolving together as a holobiont [1]. Insect endosymbiotic bacteria are traditionally classified as obligate or facultative. The definition is based on arbitrary characteristics such as tissue localization, mode of transmission, contribution to the host fitness, genome size, or AT content (extensively reviewed by [2]). In reality, there is a gradient of interactions and these categories are only indicative. Any establishment of a symbiotic association brings not only advantages, but also several disadvantages to both partners. Perhaps the most crucial is that after entering the host, the endosymbiont genome tends to decay due to genetic drift [3] and the host is becoming dependent on such a degenerating symbiont [4, 5]. As symbionts are essential for the host, the host can try to escape from this evolutionary 'rabbit hole' by acquisition of novel symbionts or via endosymbiont replacement and supplementation [6]. This phenomenon is known in almost all symbiotic groups of insects and it was especially studied in the sap-feeding group Auchenorrhyncha [7–9], while only few studies were performed in blood-sucking groups.

Blood-sucking hosts from diverse groups such as sucking lice [10–13], chewing lice [14], bed bugs [15, 16], kissing bugs [17–21], ticks [22, 23], tsetse flies [24, 25], bat flies [26, 27], louse flies [26, 28, 29], and leeches [30] have established symbiotic associations with bacteria from different lineages, mostly α-proteobacteria [15] and γ-proteobacteria [10, 11, 14, 17, 24, 25, 27–31]. 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 [16, 32–39], but experimental evidence supporting this hypothesis is scarce [15, 16, 40, 41]. 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 [42–47].

Hippoboscoidea superfamily is formed by four families (Glossinidae, Nycteribiidae, Streblidae, and Hippoboscidae) which are all obligately blood-sucking and tightly associated with endosymbionts. Its monophyly was confirmed by numerous studies [48–51], but inner phylogeny of this group has not been fully resolved yet. Glossinidae is monophyletic and sister to remaining three groups forming a monophyletic group called Pupipara [50]. Both groups associated with bats form one branch, where Nycteribiidae seems to be monophyletic while monophyly of Streblidae was not conclusively confirmed [49, 50]. According to these studies, Hippoboscidae is also a monophyletic group, but its exact position is not well-resolved.

Glossinidae (tsetse flies) harbour three different symbiotic bacteria: obligate symbiont *Wigglesworthia glossinidia* which is essential for the host survival [4], facultative symbiont *Sodalis glossinidius* which was suggested to cooperate with *Wiggleswothia* on thiamine biosynthesis [46], and reproductive manipulator *Wolbachia* [52]. Nycteribiidae, Streblidae (bat flies), and Hippoboscidae (louse flies) are associated with *Arsenophonus* bacteria [26,

31, 53–55]. On one hand, *Arsenophonus* bacteria form clades of obligatory lineages coevolving with their hosts, but on the other hand, there were detected several loosely associated *Arsenophonus* lineages likely representing facultative symbionts spreading horizontally across populations [53–55]. *Wolbachia* infection was found in all Hippoboscoidea groups [27, 38, 52, 56]. Finally, Hippoboscidae are also infected by several distinct lineages of *Sodalis*-like bacteria [28, 29, 31] likely representing similar spectrum of symbioses as observed for *Arsenophonus*.

As outlined above, Hippoboscoidea represents a group of blood-sucking insects with strikingly dynamic symbioses. Obligate symbionts from *Arsenophonus* and *Sodalis* clades tend to come and go, hampering flawless host-symbiont co-phylogenies often seen in other insect-bacteria systems. However, why are the endosymbiont replacements so common and what keeps the symbiont consortia limited to only the specific bacterial clades remains unknown. Tsetse flies as medically important vectors of pathogens are undoubtedly the most studied Hippoboscoidea lineage, but 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 phylogeny of Hippoboscidae and their symbiont lineages and try to untangle their relationship to the host, in particular if they are obligate co-evolving lineages, facultative infections, or if they likely represent recent symbiont replacements just re-starting the obligate relationship.

Results

Hippoboscidae phylogeny

We reconstructed host phylogeny using three markers: 16S rRNA, EF and COI (including three genomic COI sequences). However, Hippoboscoidea phylogeny was difficult to clearly resolve with our three-gene dataset. Therefore, we assembled and annotated mitochondrial genomes of four main louse fly lineages (supplementary figure Fig. S1) and used them for phylogenetic reconstruction as well. 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 Numts can thus also contribute to intricacy of louse fly phylogenies. According to our analyses, Hippoboscoidea represented a well-supported monophyletic clade (supplementary figures Fig. S2-18). Glossinidae formed a well-defined monophyletic group, but monophyly of the remaining three families (Hippoboscidae, Nycteribidae, 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, supplementary figures Fig. S2-18).

Arsenophonus and Sodalis phylogenies

In total, 65 endosymbiont 16S rRNA gene sequences were obtained in this study and four sequences of the 16S rRNA gene were mined from our *Arsenophonus* genome data. The genus *Arsenophonus* was identified in 27 cases, 12 sequences were similar to *Sodalis*-allied species, and 26 sequences were from *Wolbachia*. Putative obligatory and facultative symbiont assignment was based on GC content, branch length, and phylogenetic analyses (Table S3).

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, particularly within the Nycteribiidae, Streblidae, and several Hippoboscidae lineages (Fig. 2, supplementary figures Fig. S19, and Fig. S20). 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 the Hippoboscidae, *Arsenophonus* sequences from the Ornithomyinae subfamily form a monophyletic clade congruent with Ornithomyinae topology while two other obligatory *Arsenophonus* clades were detected in the genera *Lipoptena* and *Melophagus* (Fig. 2, supplementary figure Fig. S19). All other *Arsenophonus* sequences from the Hippoboscidae either represent facultative symbionts or young obligate symbioses which are impossible to reliably detect by phylogenetic methods (but see the discussion for *Hippobosca* and *Crataerina*).

Most of the likely facultative endosymbionts of the Hippoboscidae cluster within a clade of short-branched species comprising also the well-known species *Arsenophonus arthropodicus* and *Arsenophonus nasoniae*. Interestingly, both obligate and facultative lineages were detected from several species, e.g. *Ornithomya biloba* and *Ornithomya avicularia*. 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 they clustered within the outgroup in the BI analyses (supplementary figure Fig. S20) and with long-branched species in the ML analyses suggesting long branch attraction (supplementary figure Fig. S23).

Sodalis phylogeny reconstruction using 16S rDNA sequences revealed an obligatory endosymbiont from the tribe Olfersini including the genera *Pseudolynchia* and *Icosta* and several facultative lineages (Fig. 3, supplementary figures Fig. S24-26).

Wolbachia MLST analysis

Sequences of 16S rDNA were used only for *Wolbachia* detection and approximate supergroup determination (Fig. 4, supplementary figures Fig. S27). The MLST analysis was performed with ten selected species (one of them obtained from genomic data of *O. biloba*; see Table S3). Supergroup A was detected from 6 species, supergroup B was detected from 7 species, and supergroup F was detected from 17 species (including *M. ovinus* which is somewhat distant; Fig. 4, supplementary figure Fig. S28).

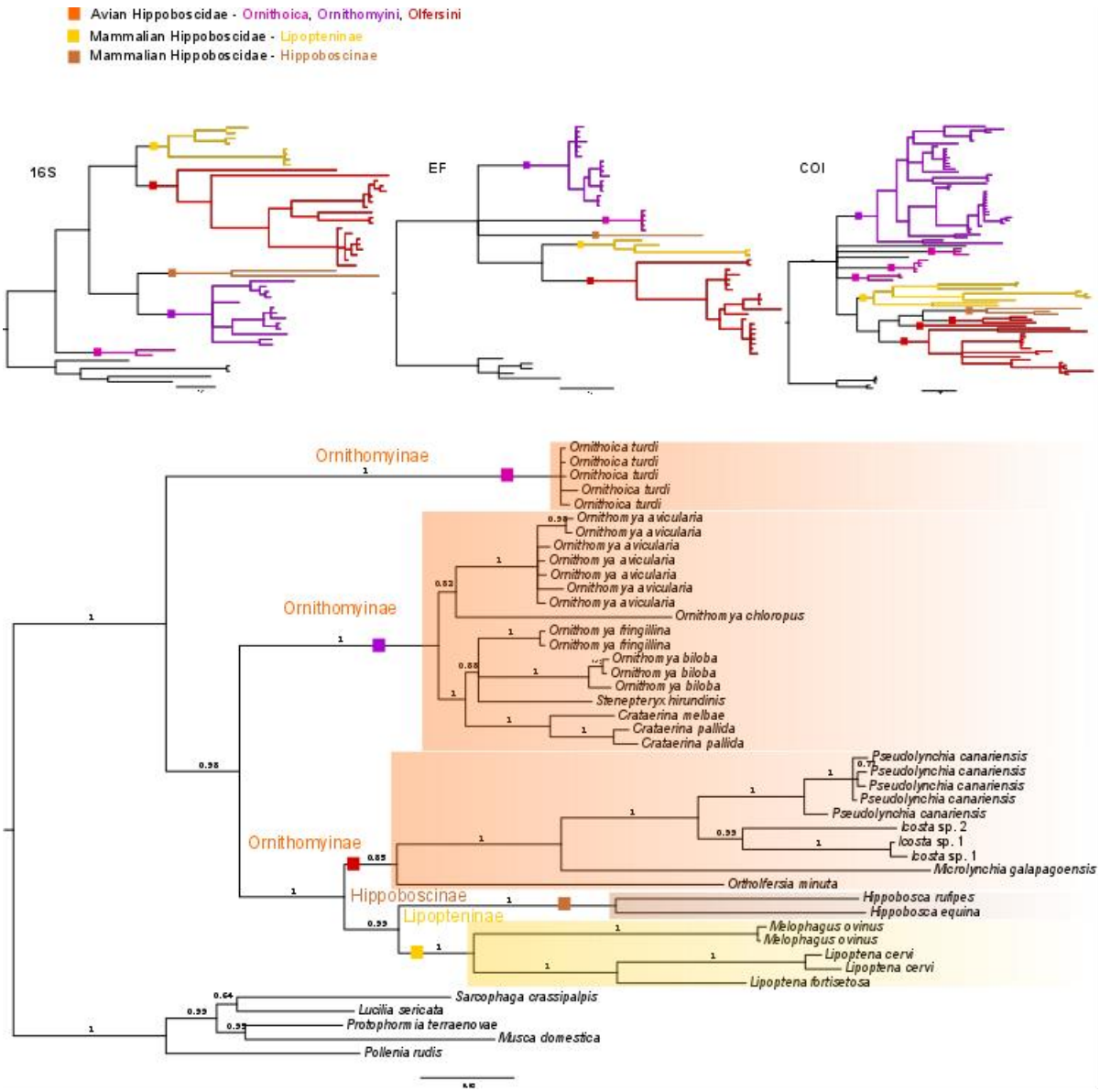

Figure 1 Host phylogeny derived from concatenation of three genes: 16S rRNA, EF, and COI. The phylogeny was reconstructed by BI analysis. Three smaller trees on the top of the figure represent outlines of three separate phylogenetic trees based on BI analyses of 16S rRNA, EF, and COI genes**.** Full versions of these phylogenies are included in supplementary figures (FIG_S2-4). 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 supplementary figures and genomic COI sequences are labelled with gDNA (FIG_S1-15).

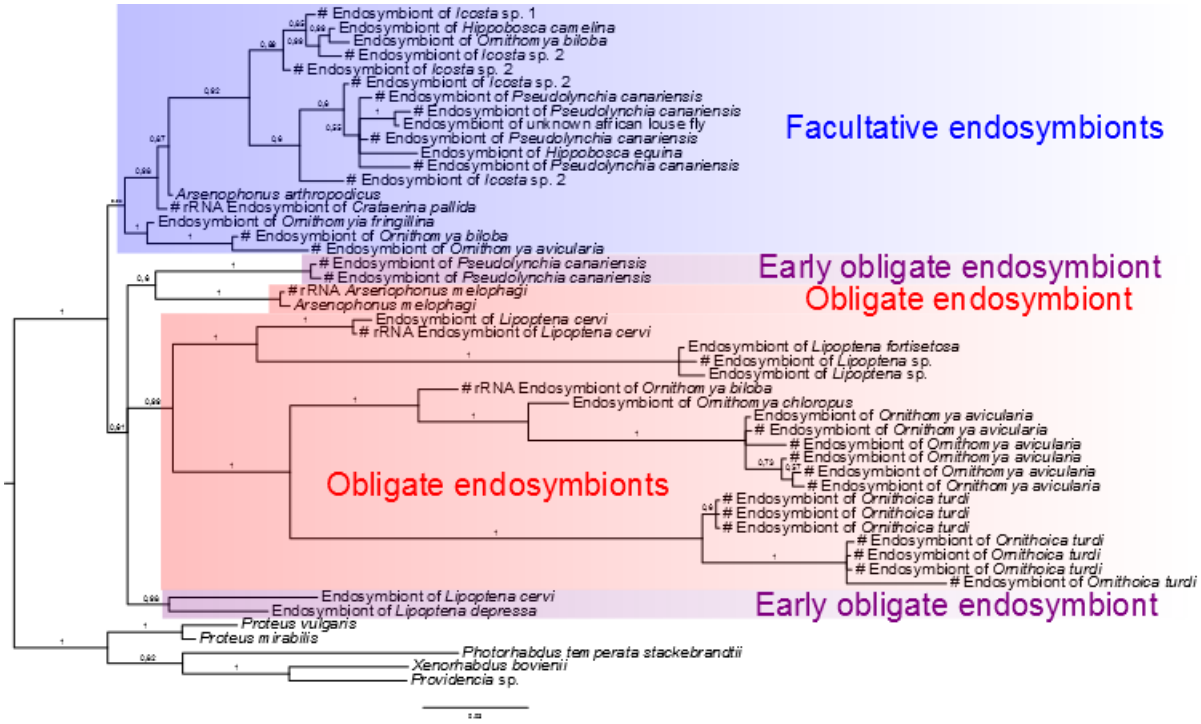

Figure 2 16S rRNA phylogeny of *Arsenophonus* in Hippoboscidae inferred by BI analysis. Taxa labelled with # are newly sequenced in this study. Genomic sequences are labelled with rRNA. Facultative symbionts, which have no co-evolutionary pattern with their hosts, are in blue, obligate symbionts with topologies congruent with their hosts, are in red, and symbionts, which are supposed to be undergoing recent genome reduction, are in purple. Phylogenetic reconstructions of *Arsenophonus* in entire Hippoboscoidea and Hippoboscoidea including problematic sequences (JX853024, JX853062, JX853027, KC597723, and KC597745) are included in supplementary figures (FIG_S16-20).

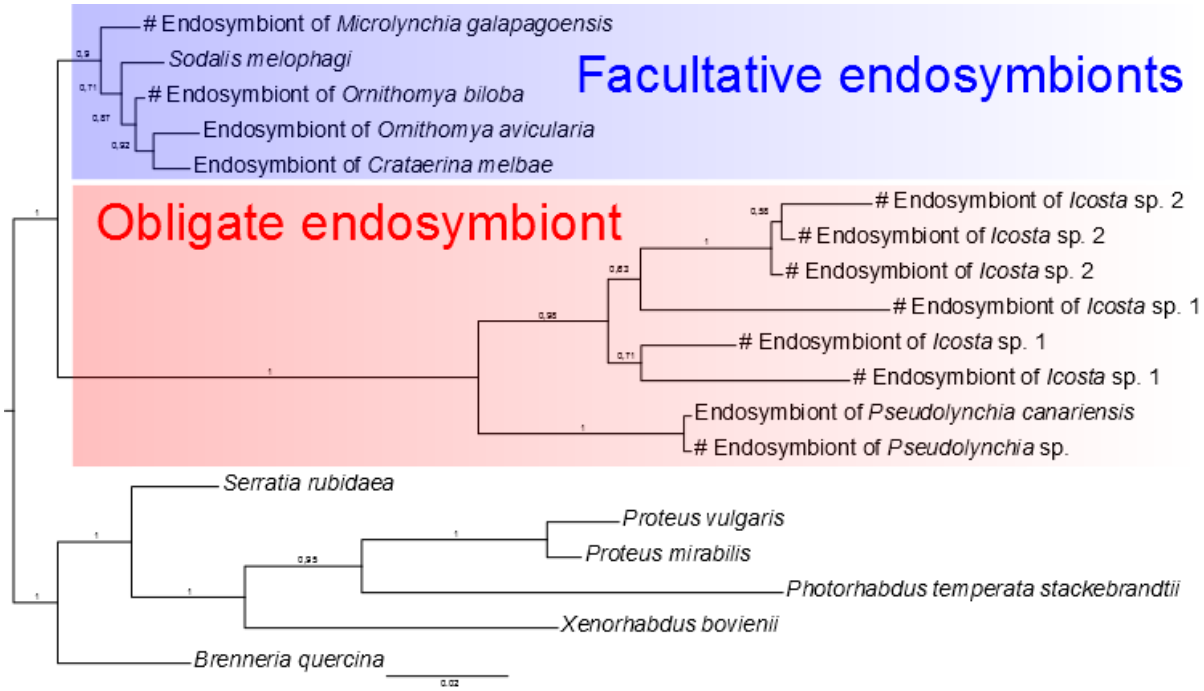

Figure 3 16S rRNA phylogeny of *Sodalis* in Hippoboscidae inferred by BI analysis. Taxa labelled with # are newly sequenced in this study. Facultative symbionts, which have no co-evolutionary pattern with host, are in blue. Obligate symbiont representing a new obligate lineage of *Sodalis*-like bacteria, with phylogeny congruent with its host is in red. Phylogenetic reconstructions of all *Sodalis*-like bacteria are included in supplementary figures (FIG_S21-23).

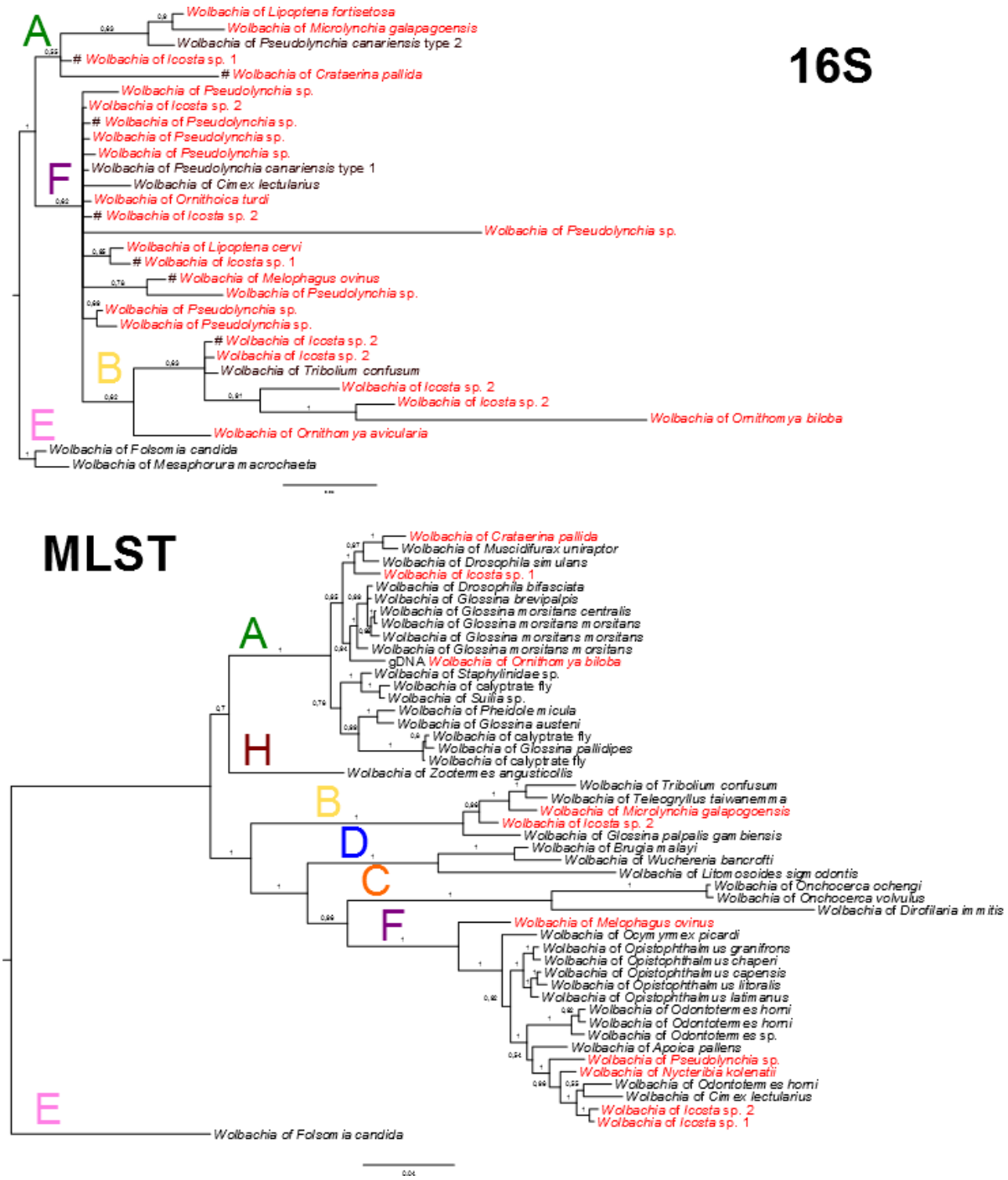

Figure 4 *Wolbachia* phylogeny inferred from 16S rRNA and MLST genes by BI analysis. Colour letters upon branches correspond to *Wolbachia* supergroups. Taxa in red represent *Wolbachia* bacteria from Hippoboscidae and Nycteribidae which are newly sequenced in this study. Taxa labelled with # in the 16S tree represent taxa which were used for the MLST analysis. *Wolbachia* from *O. biloba*, which was obtained from genomic data, is labelled with gDNA. Additional phylogenies of *Wolbachia* are included in supplementary figures (FIG_S24-25). Supergroup E was used for rooting both trees.

Discussion

Hippoboscidae phylogeny: an unfinished portrait

With respect to medical and veterinary importance, numerous studies were carried out to reconstruct phylogeny of the Hippoboscoidea group [48–51]. They confirmed that the Hippoboscoidea superfamily is monophyletic, but its inner topology was never fully resolved. This study was performed to bring better insight into Hippoboscidae phylogeny, its relationship to three remaining Hippoboscoidea groups and three genera of insect endosymbionts. Our three-gene dataset did not produce a clear solution for the Hippoboscoidea phylogeny. We thus assembled mitochondrial genomes of four main lineages of louse flies and, although there is very limited sampling of mitochondrial genomes for Hippoboscoidea, we used them for phylogenetic reconstruction as well to show congruency of topologies suggesting that mitochondrial genomes will be valuable for further phylogenetic analyses of this group. As a consequence, we were not able to infer its inner phylogeny using mitochondrial genomes because of missing data for bat flies (supplementary figures Fig. S2, Fig. S3). According to analyses based on our three gene dataset, tsetse flies form a monophyletic group (supplementary figures Fig. S8-11 and Fig. S15-18) as previously described [50, 51]. Nevertheless, monophyly of the remaining three families (Nycteribidae, Streblidae, and Hippoboscidae) was not supported even though preceding studies confirmed monophyly of Nycteribiidae [49–51] and Hippoboscidae [48–50]. Finally, Streblidae lineage remains polyphyletic [49, 51]. Within Hippoboscidae, groups Lipopteninae, Hippoboscinae, Ornithomyini and Olfersini are well-defined and monophyletic, but their exact relationship is still not clear. The most problematic taxa are Hippoboscinae group and also the genus Ornithoica with positions depending on used genes/analyses (Fig. 1; supplementary figures Fig. S4-18). A possible explanation for these incongruences in topologies can be that there was a rapid radiation from the ancestor of Hippoboscoidea group into main subfamilies of Hippoboscidae, and consequently all three selected genes carry very weak phylogenetic signal for this period of Hippoboscidae evolution. The most problematic marker for reconstruction of Hippoboscoidea phylogeny is COI because of missing data (only short sequences are available especially for Nycteribiidae and Streblidae in GenBank; supplementary figures Fig. S4-18). COI phylogenies which were also detected in this study are known to suffer from numerous pseudogenes called Numts [57]. On the other hand, EF seems to be a very good marker (supplementary figures Fig. S4-18), but the biggest disadvantage of this gene is no taxon sampling for Hippoboscoidea superfamily in GenBank.

Hidden endosymbiont diversity within Hippoboscidae family

Within Hippoboscidae, bacteria from three different endosymbiotic genera were described: *Arsenophonus* [26, 31, 53–55], *Sodalis* [28, 29, 31], and *Wolbachia* [31, 38]. The most attention has been paid to *Arsenophonus* as supposedly the most common endosymbiont of Hippoboscidae. As it was suggested by several studies, its evolution has been influenced by not only vertical transmission but also horizontal transfers with possible symbiont replacements [53–55]. Different lineages of *Arsenophonus* bacteria have probably

established obligatory lifestyle in Hippoboscoidea at least five times: three times within the Hippoboscidae (*Melophagus ovinus* and related species, *Lipoptena* spp. and related species, and Ornithomyinae group), once within the Nycteribiidae, and once within the Streblidae (Fig. 2, supplementary figure Fig. S19).

Similar results are apparent from previous studies suggesting that obligate *Arsenophonus* endosymbiont of the Nycteribiidae, described as *Aschnera chinzeii* [27], forms a monophyletic clade congruent with the host phylogeny (designed as clade A by [54]; [55, 58]), as well as obligate *Arsenophonus* endosymbiont of Streblidae forms a monophyletic group, but the congruence with the host phylogeny was not confirmed (clade B by [54], or ALO_2 by [55]). According to these studies, there is also a clade formed by diverse species of all Pupipara with no phylogenetic relationship (clade C by [54]; [55]). Our results did not confirm this clade, but the taxa were rather scattered on short branches in contrast to obligate endosymbionts (supplementary figures Fig. S19-23). We propose them to be facultative endosymbionts which is also supported by their relationship to *Arsenophonus atrthropodicus* [31] (Fig. 2). Obligate endosymbionts are known to often evolve from facultative symbionts which are no longer capable of horizontal transmission between hosts [2]. Endosymbionts with ongoing recent genome reduction, which we call early obligate endosymbionts, can be also found on branches between endosymbionts labelled as facultative. Thanks to their recent change of lifestyle, they in many ways resemble facultative endosymbionts, e.g. their positions in phylogenetic trees are not stable and differ by used analysis and taxon sampling (Fig. 2, supplementary figures Fig. S19-23). Such nascent stage of endosymbiosis was shown for obligate *Arsenophonus* endosymbiont of *C. pallida* (Šochová in prep. 2016) and similar results can be expected for *Arsenophonus* endosymbionts of *Hippobosca* species. Finally, the unstable position of *Riesia pediculicola* was clearly recognized as a consequence of long branch attraction (see supplementary figures Fig. S19-23).

One γ-proteobacterial symbiont included into the *Arsenophonus* clade was also described from Nycterophyliinae and Trichobiinae (Streblidae) ([56]; clade D by [54]; ALO-1 by [55]). However, our results do not support its placement within the clade as these sequences were placed into outgroup in our BI analysis or attracted by long branches in the ML analysis (supplementary figures Fig. S20, Fig. S23). They also clearly resemble a sequence from *Trichobius yunkeri* (DQ314776 by [26]) which was suggested to be an artificial chimerical product [53]. Therefore, these sequences were excluded from our further analyses since they likely represent either a lineage outside of the *Arsenophonus* clade or PCR artefacts.

In contrast to *Arsenophonus*, only a few studies were performed on *Sodalis*-like endosymbiotic bacteria within Hippoboscidae [28, 29, 38]. Dale et al. [31] detected a putative obligate endosymbiont from *Pseudolynchia canariensis* which was suggested to be *Sodalis*. No additional data were published about this symbiont since then. We detected this symbiont in all studied members of the Olfersini group and according to our results, it is obligate *Sodalis*-like endosymbiont forming a monophyletic clade congruent with the Olfersini phylogeny (Fig. 3, supplementary figures Fig. S24-26). Similarly to *Arsenophonus*,

Sodalis bacteria also establish facultative associations such as with *Melophagus ovinus* and *Ornithomya avicularia* [29] or *Ornithomya biloba* (this study). *Sodalis* endosymbiont from *Crataerina melbae* was suggested to be obligate [28], but our study did not confirm this hypothesis (supplementary figures Fig. S24, Fig. S26). Interestingly, facultative *Sodalis* endosymbiont of *Microlynchia galapagoensis* was inferred to be closely related to likely free-living *Biostraticola tofi* clustering within the *Sodalis* clade (supplementary figures Fig. S24, Fig. S26). These results suggest that there are several lineages of facultative *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 environment as implied by their relationship to e.g. *Sodalis praecaptivus* [59] (Fig. 3, supplementary figures Fig. S24-26).

Coinfections of obligate and facultative *Arsenophonus* strains in Hippoboscidae (or potentially *Sodalis* in Olfersini) are extremely difficult to recognize using only the 16S rRNA gene. Facultative endosymbionts retain up to seven copies of this gene often causing false variability in phylogenetic analyses [18]. Consequently, 16S rDNA of facultative endosymbionts tends to be amplified more likely in PCR than from obligate endosymbionts due to its higher copy number and lower frequency of mutations in primer binding sites. To resolve this problem, multi locus analyses should be implemented to infer overall evolutionary relationships of all endosymbionts within Hippoboscoidea. Since our data are likely also influenced by this setback, we do not dare to speculate which of the detected facultative *Arsenophonus* lineages represent 'ancestors' of the several distinct obligate lineages or which of them were involved in the recent replacement scenario. However, that the replacement or independent origin scenario happens is nicely illustrated by endosymbionts from the Olfersini group (Fig. 2, Fig. 3).

To complement the picture of Hippoboscidae endosymbiosis, we also reconstructed *Wolbachia* evolution. Louse flies were found to be infected by 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 (Fig. 4) as it was previously described [60, 61]. Since *Wolbachia* seems to be one of the most common donor of genes horizontally transferred to insect genomes, including tsetse flies [62–64], we cannot rule out that some of *Wolbachia* sequences detected in this study represent HGT insertions into the respective host genomes. *Wolbachia* from the supergroup A in *Glossina mossitans morsitans* (Glossinidae) was proposed to cause cytoplasmic incompatibility [52], but the biological role of *Wolbachia* in Hippoboscidae was never examined. 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 [56], Nycteribiidae [27], Amblycera [65], and Cimicidae where it plays a role of an obligate nutritional endosymbiont [15, 16]. Given the broad distribution of this lineage in Hippoboscidae, evaluation of its interactions with the host are an interesting goal for future studies.

Why are Hippoboscidae-symbiont associations so dynamic?

According to our results, symbiosis in the Hippoboscidae group is very dynamic and influenced by frequent symbiont replacements. Facultative *Arsenophonus* and *Sodalis* infections seem to be the best resources for endosymbiotic counterparts, but it remains unclear why just these two genera.

Sodalis glossinidius possesses modified outer membrane protein (OmpA) which is playing an important role in the interaction with the host immune system [66, 67]. Both *Sodalis* and *Arsenophonus* bacteria retain genes for the type III secretion system [29, 68–70] allowing pathogenic bacteria to invade eukaryotic cells. Moreover, several strains of these bacteria are cultivable in laboratory [17, 25, 29, 31, 71, 72] suggesting that they should be able to survive horizontal transmission, e.g. *Arsenophonus nasoniae* is able to spread by horizontal transfer between species [73], while *Sodalis*-allied bacteria have several times successfully replaced ancient symbionts [8, 74]. Whereas facultative endosymbionts of Hippoboscoidea are widespread in numerous types of tissues such as milk glands, bacteriome, haemolymph, gut, fat body, and reproductive organs [25, 31, 38, 75], obligate endosymbionts are restricted to the bacteriome and milk glands [24, 38, 54, 75, 76]. Entrance into milk glands ensures vertical transmission of facultative endosymbiont to progeny and better establishment of the infection. This enables the endosymbiont to hitch-hike with the obligate endosymbiont and because the obligate endosymbiont is inevitably degenerating [3, 77], the new co-symbiont can gradually replace it if needed. For instance, *Sodalis melophagi* was shown to appear in both milk glands and bacteriome and to code the same full set of B-vitamin pathways (including in addition a thiamine pathway) as the obligate endosymbiont *Arsenophonus melophagi [38]*. This situation suggests that it could be potentially capable to shift from facultative to obligatory lifestyle and replace the *Arsenophonus melophagi* endosymbiont.

We suggest that the evolution of endosymbiosis in Hippoboscoidea can be explained by the following scenario highly similar to a scenario already suggested for the evolution of symbiosis in *Columbicola* lice [78] and mealybugs [79]. An ancestral Pupipara endosymbiont was likely either from the *Arsenophonus* or *Sodalis* lineage (given our finding of the obligate *Sodalis* lineage in Olfersini). Since then, the symbiont was in different lineages repeatedly replaced by new *Arsenophonus* (or *Sodalis* in Olfersini if not ancestral) bacteria as supported by different levels of genome reduction in separate *Arsenophonus* lineages ([38]; Nováková in prep. 2016; Šochová in prep. 2016**)** and incongruent host-symbiont phylogenies (this study). The observed incongruences of *Arsenophonus* phylogeny with the host and no genome synteny across *Arsenophonus* from distinct Hippoboscidae, therefore; simply reflect endosymbiont lineages of different ages and thus at different stages of genome reduction.

Conclusions

Hippoboscoidea superfamily forms a monophyletic group with poorly resolved inner topology. We reconstructed its phylogeny using concatenated matrix of 15 mitochondrial

genes and three other markers: EF, COI, and 16S rRNA. Our results confirmed monophyly of tsetse flies whereas monophyly of bat flies and louse flies was not strongly supported. These results were affected especially by missing data and short sequences of COI available especially for Nycteribiidae and Streblidae and no taxon sampling of Hippoboscoidea for EF in GenBank.

We revealed unexpected complexity of symbiosis evolution within the Hippoboscidae family. Louse flies have established symbiotic association with bacteria from three different genera: *Arsenophonus*, *Sodalis*, and *Wolbachia*. *Arsenophonus* and *Sodalis* represent both obligate and facultative endosymbionts while the role of *Wolbachia* remains unclear. These results suggest very dynamic evolutionary scenario shaped by frequent symbiont replacements and turnovers. However, the mechanism driving these dynamic, yet selective, origins of obligate endosymbioses remains elusive.

Methods

Sample collection and DNA isolation

Samples of louse flies were collected in six countries (South Africa, Papua New Guinea, Ecuador – Galapagos, 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, *fts*Z, *gat*B, *hcp*A; see Table S2). All primers used in this study are summarized in supplementary table 2. PCR reaction was performed under standard conditions using High Fidelity PCR Enzyme Mix (Thermo Scientific) and Hot Start Tag DNA Polymerase (Qiagen) 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) according to the manufacturer´s protocols. Inserts from selected colonies were amplified using T7 and SP6 primers or isolated from plasmids using the Plasmid Miniprep Spin Kit (Jetquick). 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* [38], *Lipoptena cervi* (Nováková in prep. 2016), *Ornithomya biloba*, and *Crataerina pallida* (Šochová in prep. 2016) as well as their endosymbionts (see Table S1).

Alignments and phylogenetic analyses

Assembly of raw sequences was performed in Geneious v8.1.7 [80]. Datasets were composed of the assembled sequences, mined genomic sequences, available sequences in GenBank (see supplementary Table 4), or the *Wolbachia* MLST database. Sequences were aligned with Mafft v7.017 [81, 82] implemented in Geneious using an E-INS-i algorithm with default parameters. The alignment was not trimmed as trimming resulted in loss of most data. Phylogenetic analyses were carried out using maximum likelihood (ML) in PhyML v3.0 [83, 84] and Bayesian inference (BI) in MrBayes v3.1.2 [85]. The $GTR+I+\Gamma$ evolutionary model was selected in jModelTest [86] 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 [87] was used for convergence and burn-in examination. Phylogeny trees were visualised, rooted, and preliminary adjusted in FigTree v1.4.2 [88]. Final graphical adjustments of all phylogeny trees were performed in Inkscape v0.91 [89].

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 [90]. 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 redirect us to assemble mitochondrial genomes of four louse fly lineages and reconstruct phylogeny using these genomes. Contigs of mitochondrial genomes were identified in genomic data of *M. ovinus*, *L. cervi*, *O. biloba*, and *C. pallida* using BLASTn and tBLASTn searches [91]. 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 [92]. 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 reconstruction of phylogeny. Phylogeny reconstruction of concatenated matrix was performed as described above.

Abbreviations

EF: Nuclear gene for elongation factor
COI: Mitochondrial gene for cytochrome oxidase subunit I
16S rRNA: Mitochondrial/bacterial gene 16S rRNA
MLST: Multi Locus Sequence Typing

Competeting interests

The authors declare that they have no competing interests.

Author´s contributions
FH and VH designed the study. ES obtained most of the sequence data, prepared alignments, inferred phylogenies, and prepared draft manuscript. ES and FH participated in evolutionary interpretation of results. FH participated in manuscript preparation. EN provided some sequence data from her previous work. AH collected samples of African louse flies. ES, FH, EN, and VH read and approved the final manuscript.

Acknowledgement
This work was supported by the Grant Agency of the Czech Republic (grant 13-01878S to VH).

Additional files

Additional file 1: Additional methodology and result tables. Table S1 includes detailed sample information, Table S2 summarises primers used in this study, Table S3 summarises sequencing results of this study, Table S4 contains summary of mitochondrial genes used for phylogenetic reconstruction, and Table S5 contains all accession numbers of GenBank sequences used in this study.

Additional file 2: All phylogenetic trees reconstructed in this study. There are included both BI and ML figures of host phylogeny based on EF, COI, 16S rRNA, and 15 mitochondrial genes, as well as figures of symbiont phylogeny based on 16S rDNA and *Wolbachia* MLST.

References

1. Zilber-Rosenberg I, Rosenberg E: **Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution.** *FEMS Microbiol Rev* 2008, **32**:723–35.
2. Moran NA, McCutcheon JP, Nakabachi A: **Genomics and evolution of heritable bacterial symbionts.** *Annu Rev Genet* 2008, **42**(July):165–90.
3. Moran NA: **Accelerated evolution and Muller's rachet in endosymbiotic bacteria.** *Proc Natl Acad Sci* 1996, **93**:2873–2878.
4. Pais R, Lohs C, Wu Y, Wang J, Aksoy S: **The obligate mutualist *Wigglesworthia glossinidia* influences reproduction, digestion, and immunity processes of its host, the tsetse fly.** *Appl Environ Microbiol* 2008, **74**:5965–74.
5. Koga R, Tsuchida T, Sakurai M, Fukatsu T: **Selective elimination of aphid endosymbionts: effects of antibiotic dose and host genotype, and fitness consequences.** *FEMS Microbiol Ecol* 2007, **60**:229–39.
6. Bennett GM, Moran NA: **Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole**. *Proc Natl Acad Sci* 2015, **112**:201421388.
7. Bennett GM, Moran NA: **Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a phloem-feeding insect.** *Genome Biol Evol* 2013, **5**:1675–88.
8. Koga R, Bennett GM, Cryan JR, Moran NA: **Evolutionary replacement of obligate symbionts in an ancient and diverse insect lineage.** *Environ Microbiol* 2013, **15**:2073–81.
9. Koga R, Moran NA: **Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont.** *ISME J* 2014, **8**:1237–46.
10. Allen JM, Reed DL, Perotti MA, Braig HR: **Evolutionary relationships of "*Candidatus* Riesia spp.," endosymbiotic enterobacteriaceae living within hematophagous primate lice.** *Appl Environ Microbiol* 2007, **73**:1659–64.
11. Hypša V, Křížek J: **Molecular evidence for polyphyletic origin of the primary symbionts of sucking lice (Phthiraptera, Anoplura)**. *Microb Ecol* 2007, **54**:242–251.
12. Fukatsu T, Hosokawa T, Koga R, Nikoh N, Kato T, Hayama S, Takefushi H, Tanaka I: **Intestinal endocellular symbiotic bacterium of the macaque louse *Pedicinus obtusus*: Distinct endosymbiont origins in anthropoid primate lice and the old world monkey louse.** *Appl Environ Microbiol* 2009, **75**:3796–9.
13. Allen JM, Burleigh JG, Light JE, Reed DL: **Effects of 16S rDNA sampling on estimates of endosymbiont lineages in sucking lice**. 2015.
14. 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**:6660–8.
15. Hosokawa T, Koga R, Kikuchi Y, Meng X-Y, Fukatsu T: ***Wolbachia* as a bacteriocyte-associated nutritional mutualist.** *Proc Natl Acad Sci U S A* 2010, **107**:769–74.
16. Nikoh N, Hosokawa T, Moriyama M, Oshima K, Hattori M, Fukatsu T: **Evolutionary origin of insect-*Wolbachia* nutritional mutualism**. *Proc Natl Acad Sci* 2014, **111**:10257–10262.
17. 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**:1140–4.
18. 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 Appl Microbiol* 2008, **31**:88–100.
19. Ben-Yakir D: **Growth retardation of *Rhodnius prolixus* symbionts by immunizing host against Nocardia (Rhodococcus) rhodnii**. *J Insect Physiol* 1987, **33**:379–383.
20. Beard CB, Mason PW, Aksoy S, Tesh RB, Richards FF: **Transformation of an insect symbiont and expression of a foreign gene in the Chagas' disease vector *Rhodnius prolixus*.** *Am J Trop Med Hyg* 1992, **46**:195–200.
21. Pachebat JA, van Keulen G, Whitten MMA, Girdwood S, Del Sol R, Dyson PJ, Facey PD: **Draft genome sequence of *Rhodococcus rhodnii* strain LMG5362, a symbiont of *Rhodnius prolixus* (Hemiptera, Reduviidae, Triatominae), the principle vector of *Trypanosoma cruzi*.** *Genome Announc* 2013, **1**:e00329–13.
22. Zhong J: ***Coxiella*-like endosymbionts.** *Adv Exp Med Biol* 2012, **984**:365–79.
23. Liu L, Li L, Liu J, Hu Y, Liu Z, Guo L, Liu J: **Coinfection of *Dermacentor silvarum* Olenev (Acari: Ixodidae) by *Coxiella*-Like, *Arsenophonus*-like, and *Rickettsia*-like symbionts.** *Appl Environ Microbiol* 2013, **79**:2450–4.
24. Aksoy S: ***Wigglesworthia* gen. nov. and *Wigglesworthia glossinidia* sp. nov., taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies.** *Int J Syst Bacteriol* 1995, **45**:848–51.
25. Dale C, Maudlin I: ***Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*.** *Int J Syst Bacteriol* 1999, **49 Pt 1**:267–75.
26. 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**:64–8.
27. Hosokawa T, Nikoh N, Koga R, Satô M, Tanahashi M, Meng X-Y, Fukatsu T: **Reductive genome evolution, host-symbiont co-speciation and uterine transmission of endosymbiotic bacteria in bat flies.** *ISME J* 2012, **6**:577–87.
28. Nováková E, Hypša V: **A new *Sodalis* lineage from bloodsucking fly *Craterina melbae* (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont *Sodalis glossinidius*.** *FEMS Microbiol Lett* 2007, **269**:131–5.
29. Chrudimský T, Husník F, Nováková E, Hypša V: **Candidatus Sodalis melophagi sp. nov.: phylogenetically independent comparative model to the tsetse fly symbiont Sodalis glossinidius.** *PLoS One* 2012, **7**:e40354.
30. Kikuchi Y, Fukatsu T: **Endosymbiotic bacteria in the esophageal organ of glossiphoniid leeches**. *Appl Environ Microbiol* 2002, **68**:4637–4641.
31. 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*.**

Appl Environ Microbiol 2006, **72**:2997–3004.
32. 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*.** *Genome Biol Evol* 2015, **7**:2871–84.
33. Boyd BM, Allen JM, de Crécy-Lagard V, Reed DL: **Genome sequence of *Candidatus* Riesia pediculischaeffi, endosymbiont of chimpanzee lice, and genomic comparison of recently acquired endosymbionts from human and chimpanzee lice.** *G3 (Bethesda)* 2014, **4**:2189–95.
34. Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, Lee SH, Robertson HM, Kennedy RC, Elhaik E, Gerlach D, Kriventseva E V, 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, Krause E, Mittapalli O, 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.
35. Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S: **Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*.** *Nat Genet* 2002, **32**:402–407.
36. Rio RVM, Symula RE, Wang J, Lohs C, Wu Y, Snyder AK, Bjornson RD, Oshima K, Biehl BS, Perna NT, Hattori M, Aksoy S: **Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: glossinidae) obligate symbiont *Wigglesworthia*.** *MBio* 2012, **3**:e00240–11.
37. Smith TA, Driscoll T, Gillespie JJ, Raghavan R: **A *Coxiella*-like endosymbiont is a potential vitamin source for the lone star tick.** *Genome Biol Evol* 2015, **7**:831–8.
38. 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.
39. Boyd BM, Allen JM, Koga R, Fukatsu T, Sweet AD, Johnson KP, Reed DL: **Two bacteria, *Sodalis* and *Rickettsia*, associated with the seal louse *Proechinophthirus fluctus* (Phthiraptera: Anoplura).** *Appl Environ Microbiol* 2016:AEM.00282–16.
40. Michalkova V, Benoit JB, Weiss BL, Attardo GM, Aksoy S: **Vitamin B6 generated by obligate symbionts is critical for maintaining proline homeostasis and fecundity in tsetse flies.** *Appl Environ Microbiol* 2014, **80**:5844–53.
41. Snyder AK, Rio RVM: **"*Wigglesworthia morsitans*" folate (vitamin B9) biosynthesis contributes to tsetse host fitness.** *Appl Environ Microbiol* 2015, **81**:5375–86.
42. Geiger A, Ravel S, Frutos R, Cuny G: ***Sodalis glossinidius* (Enterobacteriaceae) and vectorial competence of *Glossina palpalis gambiensis* and *Glossina morsitans morsitans* for *Trypanosoma congolense* savannah type.** *Curr Microbiol* 2005, **51**:35–40.
43. Geiger A, Ravel S, Mateille T, Janelle J, Patrel D, Cuny G, Frutos R: **Vector competence of *Glossina palpalis gambiensis* for *Trypanosoma brucei* s.l. and genetic diversity of the symbiont *Sodalis glossinidius*.** *Mol Biol Evol* 2007, **24**:102–9.
44. Toh H, Weiss BL, Perkin SAH, Yamashita A, Oshima K, Hattori M, Aksoy S: **Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of Sodalis glossinidius in the tsetse host.** *Genome Res* 2006, **16**:149–156.
45. Snyder AK, Deberry JW, Runyen-Janecky L, Rio RVM: **Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts.** *Proc Biol Sci* 2010, **277**:2389–97.
46. Belda E, Moya A, Bentley S, Silva FJ: **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* 2010, **11**:449.
47. Wang J, Brelsfoard C, Wu Y, Aksoy S: **Intercommunity effects on microbiome and GpSGHV density regulation in tsetse flies.** *J Invertebr Pathol* 2013, **112 Suppl**:S32–9.
48. Nirmala X, Hypša V, Žurovec M: **Molecular phylogeny of Calyptratae (Diptera: Brachycera): the evolution of 18S and 16S ribosomal rDNAs in higher dipterans and their use in phylogenetic inference.** *Insect Mol Biol* 2001, **10**:475–85.
49. 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 Phylogenet Evol* 2006, **38**:155–70.
50. Petersen FT, Meier R, Kutty SN, Wiegmann BM: **The phylogeny and evolution of host choice in the Hippoboscoidea (Diptera) as reconstructed using four molecular markers.** *Mol Phylogenet Evol* 2007, **45**:111–22.
51. Kutty SN, Pape T, Wiegmann BM, Meier R: **Molecular phylogeny of the Calyptratae (Diptera: Cyclorrhapha) with an emphasis on the superfamily Oestroidea and the position of Mystacinobiidae and McAlpine's fly**. *Syst Entomol* 2010, **35**:614–635.
52. Pais R, Balmand S, Takač P, Alam U, Carnogursky J, Brelsfoard C, Galvani A, Aksoy S, Medlock J, Heddi A, Lohs C: ***Wolbachia* symbiont infections induce strong cytoplasmic incompatibility in the tsetse fly *Glossina morsitans***. *PLoS Pathog* 2011, **7**:e1002415.
53. Nováková E, Hypša V, Moran NA: ***Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution.** *BMC Microbiol* 2009, **9**:143.
54. Morse SF, Bush SE, Patterson BD, Dick CW, Gruwell ME, Dittmar K: **Evolution, multiple acquisition, and localization of endosymbionts in bat flies (Diptera: Hippoboscoidea: Streblidae and Nycteribiidae).** *Appl Environ Microbiol* 2013, **79**:2952–61.
55. Duron O, Schneppat UE, Berthomieu A, Goodman SM, Droz B, Paupy C, Nkoghe JO, Rahola N, Tortosa P: **Origin, acquisition and diversification of heritable bacterial endosymbionts in louse flies and bat flies.** *Mol Ecol* 2014, **23**:2105–17.
56. Morse SF, Dick CW, Patterson BD, Dittmar K: **Some like it hot: evolution and ecology of novel endosymbionts in bat flies of cave-roosting bats (Hippoboscoidea, Nycterophiliinae).** *Appl Environ Microbiol* 2012, **78**:8639–49.
57. Black Iv WC, Bernhardt SA: **Abundant nuclear copies of mitochondrial origin (NUMTs) in the *Aedes aegypti* genome.** *Insect Mol Biol* 2009, **18**:705–13.
58. Wilkinson DA, Duron O, Cordonin C, Gomard Y, Ramasindrazana B, Mavingui P, Goodman SM, Tortosa P: **The bacteriome of bat flies (Nycteribiidae) from the Malagasy region: a community shaped by host ecology, bacterial transmission mode, and host-vector specificity.** *Appl Environ Microbiol* 2016(January):AEM.03505–15.
59. Clayton AL, Oakeson KF, Gutin M, Pontes A, Dunn DM, von Niederhausern AC, Weiss RB, Fisher M, Dale C: **A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect-bacterial symbioses.** *PLoS Genet* 2012, **8**:e1002990.
60. Schilthuizen M, Stouthamer R: **Horizontal transmission of parthenogenesis-inducing microbes in *Trichogramma* wasps.** *Proc Biol Sci* 1997, **264**:361–6.
61. Gerth M, Gansauge M-T, Weigert A, Bleidorn C: **Phylogenomic analyses uncover origin and spread of the *Wolbachia* pandemic.** *Nat Commun* 2014, **5**:5117.
62. Husník F, Nikoh N, Koga R, Ross L, Duncan RP, Fujie M, Tanaka M, Satoh N, Bachtrog D, Wilson ACC, Dohlen CD Von, Fukatsu T, Mccutcheon JP: **Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis**. *Cell* 2013, **153**:1567–1578.
63. Sloan DB, Nakabachi A, Richards S, Qu J, Murali SC, Gibbs RA, Moran NA: **Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont**

genomes in sap-feeding insects. *Mol Biol Evol* 2014, **31**:857–71.
64. 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: **Presence of extensive *Wolbachia* symbiont insertions discovered in the genome of its host *Glossina morsitans morsitans*.** *PLoS Negl Trop Dis* 2014, **8**:e2728.
65. Covacin C, Barker SC: **Supergroup F *Wolbachia* bacteria parasitise lice (Insecta: Phthiraptera).** *Parasitol Res* 2007, **100**:479–85.
66. Weiss BL, Wu Y, Schwank JJ, Tolwinski NS, Aksoy S: **An insect symbiosis is influenced by bacterium-specific polymorphisms in outer-membrane protein A.** *Proc Natl Acad Sci U S A* 2008, **105**:15088–93.
67. Weiss BL, Maltz M, Aksoy S: **Obligate symbionts activate immune system development in the tsetse fly.** *J Immunol* 2012, **188**:3395–403.
68. Dale C, Young SA, Haydon DT, Welburn SC: **The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion**. *Proc Natl Acad Sci* 2001, **98**:1883–1888.
69. Wilkes TE, Darby AC, Choi J-H, Colbourne JK, Werren JH, Hurst GDD: **The draft genome sequence of *Arsenophonus nasoniae*, son-killer bacterium of *Nasonia vitripennis*, reveals genes associated with virulence and symbiosis.** *Insect Mol Biol* 2010, **19 Suppl 1**:59–73.
70. 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: **Genome degeneration and adaptation in a nascent stage of symbiosis.** *Genome Biol Evol* 2014, **6**:76–93.
71. Darby AC, Choi J-H, Wilkes T, Hughes MA, Werren JH, Hurst GDD, Colbourne JK: **Characteristics of the genome of *Arsenophonus nasoniae*, son-killer bacterium of the wasp *Nasonia*.** *Insect Mol Biol* 2010, **19 Suppl 1**:75–89.
72. Chari A, Oakeson KF, Enomoto S, Jackson DG, Fisher MA, Dale C: **Phenotypic characterization of *Sodalis praecaptivus* sp. nov., a close non-insect-associated member of the *Sodalis*-allied lineage of insect endosymbionts.** *Int J Syst Evol Microbiol* 2015, **65**(Pt 5):1400–5.
73. Duron O, Wilkes TE, Hurst GDD: **Interspecific transmission of a male-killing bacterium on an ecological timescale.** *Ecol Lett* 2010, **13**:1139–48.
74. Conord C, Despres L, Vallier A, Balmand S, Miquel C, Zundel S, Lemperiere G, Heddi A: **Long-term evolutionary stability of bacterial endosymbiosis in curculionoidea: additional evidence of symbiont replacement in the dryophthoridae family.** *Mol Biol Evol* 2008, **25**:859–68.
75. Heddi A, Lohs C, Aksoy S, Balmand S: **Tissue distribution and transmission routes for the tsetse fly endosymbionts**. *J Invertebr Pathol* 2012, **112 Suppl**:S116–22.
76. Attardo GM, Lohs C, Heddi A, Alam UH, Yildirim S, Aksoy S: **Analysis of milk gland structure and function in *Glossina morsitans*: milk protein production, symbiont populations and fecundity.** *J Insect Physiol* 2008, **54**:1236–42.
77. Wernegreen JJ: **Genome evolution in bacterial endosymbionts of insects.** *Nat Rev Genet* 2002, **3**:850–61.
78. Smith WA, Oakeson KF, Johnson KP, Reed DL, Carter T, Smith KL, Koga R, Fukatsu T, Clayton DH, Dale C: **Phylogenetic analysis of symbionts in feather-feeding lice of the genus *Columbicola* : evidence for repeated symbiont replacements**. *BMC Evol Biol* 2013, **13**:1.
79. Husník F, McCutcheon JP: **Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis.** Cold Spring Harbor Labs Journals; 2016.
80. 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: **Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.** *Bioinformatics* 2012, **28**:1647–9.
81. Katoh K: **MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform**. *Nucleic Acids Res* 2002, **30**:3059–3066.
82. Katoh K, Asimenos G, Toh H: **Multiple alignment of DNA sequences with MAFFT.** *Methods Mol Biol* 2009, **537**:39–64.
83. Guindon S, Delsuc F, Dufayard J-F, Gascuel O: *Bioinformatics for DNA Sequence Analysis*. *Volume 537*. Totowa, NJ: Humana Press; 2009. [*Methods in Molecular Biology*]
84. Guindon S, Gascuel O: **A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood**. *Syst Biol* 2003, **52**:696–704.
85. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogenetic trees.** *Bioinformatics* 2001, **17**:754–755.
86. Posada D: **Selection of models of DNA evolution with jModelTest.** *Methods Mol Biol* 2009, **537**:93–112.
87. **Tracer** [http://tree.bio.ed.ac.uk/software/tracer/]
88. **FigTree** [http://tree.bio.ed.ac.uk/software/figtree/]
89. **Inkscape** [https://inkscape.org/en/]
90. Smith SA, Dunn CW: **Phyutility: a phyloinformatics tool for trees, alignments and molecular data.** *Bioinformatics* 2008, **24**:715–6.
91. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403–10.
92. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2.** *Nat Methods* 2012, **9**:357–9.

Insight into genomes of obligate *Arsenophonus* endosymbionts of two avian louse flies, *Ornithomya biloba* and *Crataerina pallida*

Eva Šochová[1,*], Filip Husník[2, 3], Petr Heneberg[4], 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, ASCR, v.v.i., České Budějovice, Czech Republic
[4] Third Faculty of Medicine, Charles University in Prague, Czech Republic
*Correspondence: sochova.e@seznam.cz

Data deposition:

Abstract

Insects feeding on nutrient poor diet commonly establish co-operations with symbiotic bacteria to compensate this deficiency. Endosymbionts of hematophagous host are supposed to complement them B-vitamins and cofactors missing from their blood meal. Experimental evidence supporting this type of nutrient provisioning is scarce and limited to model species, and therefore our knowledge predominantly comes from genome data. Up to date, genome data are available only for endosymbiotic bacteria from mammalian blood-feeders. There should be expected a difference between endosymbionts of mammalian and avian parasites as composition of their blood can be diverse. Here we present the first genomes of obligate endosymbionts of avian parasites from exclusively blood-sucking Hippoboscidae group, *Candidatus* Arsenophonus ornithomyarum and *Candidatus* Arsenophonus crataerinae. These bacteria not only represent a comparative model to mammal-parasite-endosymbiont models but also help us to better understand the evolution and origin of symbiosis within louse flies which was previously shown to be very dynamic. In terms of B-vitamin supplementation, we did not observe a remarkable difference between endosymbionts of avian and mammalian parasites. However, our data support previous hypothesis that *Arsenophonus* bacteria established endosymbiosis several times independently within Hippoboscidae. *A*. *ornithomyarum* represents ancient lineage of obligate endosymbiont while *A*. *cratarinae* is very likely recently acquired. Interestingly, *A*. *ornithomyarum* possesses a plasmid with horizontally transferred pantothenate (B5) biosynthesis genes acquired from *Sodalis* bacteria, which commonly infect Hippoboscidae. This finding nicely exemplifies 'intracellular arena' hypothesis suggesting that bacteria co-infecting a single host can exchange genes.

Keywords: *Arsenophonus*, Obligate endosymbionts, Louse flies, Bird parasites, B-vitamins, Plasmid

Introduction

Hematophagous organisms feeding solely on nutritionally unbalanced blood diet frequently establish associations with endosymbiotic bacteria. Although symbioses of blood-sucking hosts have originated multiple times independently (Ben-Yakir 1987; Aksoy 1995; Hypša & Dale 1997; Dale & Maudlin 1999; Trowbridge et al. 2006; Allen et al. 2007; Křížek & Hypša 2007; Nováková & Hypša 2007; Hosokawa et al. 2010; Chrudimský et al. 2012; Allen et al. 2015), they seem to converge on similar genome content in not only insects (Akman et al. 2002; Kirkness et al. 2010; Rio et al. 2012; Boyd et al. 2014; Nikoh et al. 2014; Nováková et al. 2015), but also across other systems such as leeches (Manzano-Marín et al. 2015) or ticks (Smith et al. 2015). Since blood is notoriously poor in B-vitamins and cofactors, symbionts are hypothesized to supplement this deficiency. As a consequence, most attention in genomic studies has been paid to their capacity to synthesize B-vitamins (Akman et al. 2002; Kirkness et al. 2010; Rio et al. 2012; Boyd et al. 2014; Nikoh et al. 2014; Manzano-Marín et al. 2015; Nováková et al. 2015; Smith et al. 2015; Boyd et al. 2016). However, it is important to note that as laboratory rearing of hematophagous hosts is not always possible, experimental evidence supporting B-vitamin provision is scarce and limited to model species such as bedbugs and tsetse flies (Hosokawa et al. 2010; Snyder et al. 2010; Michalkova et al. 2014; Nikoh et al. 2014; Snyder & Rio 2015), so comparative genome data from non-model systems can still provide us with valuable information.

There are fourteen whole genomes of endosymbionts from blood-sucking hosts available to date (see supplementary table S1) which represent a small fraction when compared to numerous endosymbiont genomes from sap-feeding hosts (reviewed in Moran & Bennett 2014). On one hand, blood-feeding and sap-feeding symbioses seem to be very similar in regards to the supplementation of nutritionally poor diet. But on the other hand, there is likely a significant functional difference between these systems. Sap-feeding insects complement enzymatic machinery of their symbionts with usually highly reduced genomes (Hansen & Moran 2011; Husník et al. 2013; Sloan et al. 2014; Luan et al. 2015). These endosymbionts are in most cases separated from the host cytoplasm by host-derived symbiosomal membrane (McLean & Houk 1973). This outermost membrane is hypothesized to represent barrier from nutrients in the host cytoplasm in order to enable host control of nutrient supply to the endosymbiont. Consequently, essential amino acid biosynthesis is regulated by hosts in bacteriocytes of hemipteran insects (Price et al. 2014; Duncan et al. 2014). In stark contrast, endosymbionts of blood-sucking insects have genomes that do not show extreme reduction (roughly less than 500 genes), are in most cases not engulfed by the symbiosomal membrane and can thus freely use nutrients from the host cytoplasm, and were never shown to complement host pathways in a highly interdependent manner.

Among bloodsucking insects, tsetse flies (Hippoboscoidea: Glossinidae), medically important vectors of *Trypanosoma* spp., are the most studied symbiotic system as

their endosymbionts are potential targets for vector control (Medlock et al. 2013). Complete genome data are available for not only all tsetse fly endosymbionts, *Wigglesworthia glossinidia*, *Sodalis glossinidius*, and *Wolbachia* wGm (Akman et al. 2002; Toh et al. 2006; Rio et al. 2012; Brelsfoard et al. 2014), but also for the insect host (International Glossina Genome Initiative 2014). Paired host-symbiont data are also currently available for two more species infecting human: body lice (Hampton 2010; Kirkness et al. 2010) and bedbugs (Nikoh et al. 2014; Benoit et al. 2016; Rosenfeld et al. 2016), but all other systems were inspected without the host genome (Kumar et al. 2013; Pachebat et al. 2013; Boyd et al. 2014; Nováková et al. 2015) (Nováková et al. 2016).

Comparative genomic analyses suggest that pathways for biosynthesis of B vitamins biotin (B7), riboflavin (B2), pyridoxine (B6) are conserved among all species sequenced to date. However, there are several notable exceptions among the different systems. For example, all obligate endosymbionts originating from the *Arsenophonus* clade (i.e. from louse flies *Melophagus ovinus* and *Lipoptena fortisetosa*, and lice *Pediculus humanus* and *Pediculus pediculischaeffi)* do not possess genes for thiamine (B1) biosynthesis, but only code genes for its ABC transporter (Kirkness et al. 2010; Boyd et al. 2014; Nováková et al. 2015)(Nováková et al. 206). It is therefore uncertain if the host and symbiont only compete for thiamine from the blood meal or if it is somehow involved in the establishment and maintenance of these symbioses.

Two B-vitamins that are well-supported as essential and likely provided from symbionts to their bloodsucking hosts are B7 and B2, biotin and riboflavin (Akman et al. 2002; Kirkness et al. 2010; Rio et al. 2012; Boyd et al. 2014; Nikoh et al. 2014; Manzano-Marín et al. 2015; Nováková et al. 2015; Smith et al. 2015). For instance, origin of the nutritional symbiosis between *Wolbachia* (wCle) and *Cimex lectularius* bedbug is a consequence of acquisition of biotin operon by horizontal gene transfer (Nikoh et al. 2014). Moriyama et al. (2015) also proposed that riboflavin provided by *Wolbachia* underlines its fitness contribution to its host over reproductive manipulation to spread in the host population.

Strikingly, all bloodsucking organisms described above as having endosymbiont genomes analysed (with the exception of one genome from a leech) feed primarily on mammals. However, there are numerous bloodsucking parasites that specialize on other vertebrates such as birds, amphibians or reptiles. Blood composition in these animals can be, however, quite different from mammals (e.g. due to presence of nucleated red blood cells or different levels of some B-vitamins). It is unknown if these differences in the parasite blood meal drive gene loss and retention of its symbionts. Louse flies (Hippoboscidae) feeding on birds are in our opinion a particularly suitable model group to study this question. They are closely related to (and likely evolved within) a clade of parasites feeding on mammals (tsetse flies, bat flies, and mammalian louse flies) and thus reducing differences caused by the host phylogenetic origin. Moreover, they have obligate endosymbionts from the *Arsenophonus* clade making gene content of the putative ancestral symbiont(s) of the mammalian and avian parasites highly similar and thus also reducing

differences caused by the symbiont phylogenetic origin. If the blood composition of birds and mammals does not make a difference, we would expect the symbionts to converge on similarly reduced gene/pathway content as in already described symbionts of mammalian louse flies (Trowbridge et al. 2006; Chrudimský et al. 2012; Hosokawa et al. 2012; Morse et al. 2012, 2013; Nováková et al. 2015).
Here we present complete genome sequence of an obligate endosymbiont of *Ornithomya biloba* (Hippoboscoidea: Hippoboscidae), *Candidatus* Arsenophonus ornithomyarum, and a draft genome sequence of an obligate endosymbiont of *Craterina pallida* (Hippoboscoidea: Hippoboscidae)*, Candidatus* Arsenophonus crataerinae*.* For simplicity, we refer to these endosymbionts without the *Candidatus* denomination hereafter. We also reconstructed *Arsenophonus* phylogeny based on phylogenomic data available for the *Arsenophonus* clade (using 23 genes from 8 complete genomes of *Arsenophonus* bacteria) and analysed the genomes with particular attention paid to bloodsucking members of the clade.

Materials and Methods

Sample preparation and sequencing

Samples of *Ornithomya biloba* and *Crataerina pallida* were collected in the Czech Republic from sand martin (*Riparia riparia*; 48.9344264N, 16.6248681E) and common swift (*Apus apus*; 50.0907972N, 15.0315103E), respectively, and stored in RNA later (Qiagen) at -20°C. Digestive systems were dissected in RNA later under Olympus SZ61 dissecting microscope and total genomic DNA (gDNA) was isolated from a single gut of each species. DNA extraction was performed using QIAamp DNA Micro Kit (Qiagen) according to the manufacturer's protocol. DNA concentrations were determined with Qubit Fluorometric Quantification (Invitrogen) and its quality was verified in 1 % agarose gel electrophoresis. In total, approximately one microgram of gDNA from *O*. *biloba* and three micrograms of gDNA from *C*. *pallida* were used for library preparations and the libraries were sequenced using Illumina MiSeq 300-200 bp paired-end sequencing (only O. biloba) and Illumina HiSeq 100 bp paired-end sequencing (both species) at Genecore sequencing facility in Heidelberg, Germany.

Microscopy

All procedures were performed following a protocol of Chrudimský et al. (2012). Midgut regions with bacteriomes from *O*. *biloba* and *C*. *pallida* were dissected directly into a 2.5% glutaraldehyde in 0.1 M phosphate buffer and prefixed at 4°C over night. Post-fixation of tissue was carried out in 2% osmium tetroxide in phosphate buffer for 60 min at 4°C. The samples were then dehydrated through ethanol series and embedded in Spurr resin. Uranyl acetate and lead citrate were

used for staining of ultrathin sections from samples. The ultratin sections were examined under transmission electron microscope JEOL JEM-1010.

Assembly and annotation endosymbiont genomes

Quality of raw reads was assessed by FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and overlapping paired-end reads from *O*. *biloba* were merged by PEAR tool v0.9.4 (Zhang et al. 2014). Raw data from *C*. *pallida* were trimmed according to sequence quality (< 31) and length (< 91), i. e. all reads with quality of sequence lower than 31 and shorter in length than 91 bp were discarded, using Sickle v1.33 (https://github.com/najoshi/sickle). De novo assembly of data from *O*. *biloba* was performed in SPAdes v3.6.1 (Bankevich et al. 2012) using merged k-mers (*-k 21,33,55,77,91,127,251*) and its *--careful* flag. MEGAHIT v0.2.0 (Li et al. 2015) assembler was used for de novo assembly of data from *C*. *pallida* using --no-mercy assembly option. Blobtools pipeline v0.9.12 (Kumar et al. 2013) and Bandage v0.7.0 (Wick et al. 2015) were used for visualisation of draft genome assemblies and exploration of endosymbiont composition. Pilon v 1.12 (Walker et al. 2014) was used for improve potential missassemblies in genome of obligate endosymbiont of *O*. *biloba*. Annotation of all endosymbiont genomes was carried out in Prokka annotating tool v1.10 (Seemann 2014) and visualized in Artemis genome browser v16.0.0 (Rutherford et al. 2000). Manual assessment of pseudogenes was performed for genes shorter than sixty percent of normal length of its ten top BlastP hits against NR and BlastX of intergenic regions. As the endosymbiont of *O*. *biloba* contains relatively long AT rich regions, several functional genes were annotated as pseudogenes due to homo-polymeric regions, but we cannot rule out that these regions are polymorphic in the symbiont population or that they are sometimes made functional by transcriptional slippage (Wernegreen et al. 2010).

Phylogenomics

Orthologous gene groups of 15 Enterobacteriaceae species were generated by the OrthoFinder program (Emms & Kelly 2015). Only single-copy genes present in all species (23 genes) were used for phylogenetic analyses. Matrices of individual genes were aligned by the MAFFT v7.017 E-INS-I algorithm (Katoh 2002; Katoh et al. 2009) implemented in Geneious and then concatenated in Geneious (Kearse et al. 2012). Ambiguously aligned positions were excluded by Gblocks v0.91 (Talavera & Castresana 2007). Concatenated amino-acid alignments were used for phylogeny reconstruction using Maximum likelihood (ML) and Bayesian inference (BI) phylogenetic methods. ML analysis was performed in Phyml v3.0 (Guindon & Gascuel 2003) under the LG+G model with subtree pruning and re-grafting tree search algorithm (SPR) and 100 bootstrap pseudo-replicates. BI was carried out in MrBayes 3.2.6 (Huelsenbeck & Ronquist 2001) under the LG+I+G model with one million generations. Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/) was used

for convergence and burn-in examination of BI runs. Concatenated amino-acid and Dayhoff6 recoded matrices were also analyzed under the CAT+GTR+G model in PhyloBayes MPI 1.6j (Lartillot et al. 2013) and PhyloBayes 4.1c (Lartillot et al. 2009), respectively. Posterior distributions obtained under two independent PhyloBayes runs were compared using tracecomp and bpcomp programs and runs were considered converged at maximum discrepancy values < 0.1 and minimum effective sizes > 100. Visualisation and rooting or tree was performed in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/) and graphical adjustments of the final trees was carried out in Inkscape v0.91 (https://inkscape.org/en/).

Reconstruction of metabolic pathways and comparative genomics

Three different tools were used for reconstruction of B-vitamin and co-factor pathways: Pathway Tools Software v19.5 (Dale et al. 2010), KAAS-KEGG Automatic Annotation Server (Moriya et al. 2007), and Blast2go Basic v3.2 (Conesa et al. 2005). Firstly, annotated draft genome assemblies of *A*. *ornithomyarum* and *A*. *crataerinae* were uploaded into the Pathway Tools Software and its Pathologic module where they were compared with species comparison and manually examined. Secondly, proteomes of *A*. *ornithomyarum* and *A*. *crataerinae* were uploaded into KAAS-KEGG Automatic Annotation Server using BBH (bi-directional best hit) assignment method and Blast2go Basic using default parameters. B-vitamin and co-factor pathway reconstructions were performed with EcoCyc (Keseler et al. 2013) and KEGG (Kanehisa et al. 2012) databases as guidelines. B-vitamin and co-factor pathways of facultative endosymbionts of *O*. *biloba* (*Sodalis* and *Wolbachia*), were also reconstructed using this approach.
COG (clusters of orthologous genes) categories of *A*. *ornithomyarum* and *A*. *crataerinae* were assigned using BlastP with an e-value cut-off of $1e^{-08}$ against the COG database (Tatusov et al. 2003). Complete genomes of four *Arsenophonus*-like bacteria from blood-sucking hosts were aligned using tBlastX and were visualized as linear with links connecting positions of blast hits in Processing3 (https://processing.org/).

Results

Endosymbiont diversity and tissue distribution

Three different bacteria were detected in our data from *O*. *biloba*. In MiSeq data, there were one obligate *Arsenophonus* symbiont, one facultative *Sodalis*-allied symbiont, and one facultative *Wolbachia* symbiont (supplementary Fig. S1a). In HiSeq data, we detected only one obligate *Arsenophonus* symbiont and one facultative *Wolbachia* symbiont (supplementary Fig. S1b). The de novo assembly of our data revealed that the genome of *Arsenophonus* was assembled into a single molecule and one plasmid while high number of repetitive sequences in *Sodalis* and

Wolbachia genomes made it impossible to assemble these genomes into circular molecules from our short-read data (supplementary Fig S2).
Two bacteria were detected in the data from *C. pallida*: one likely obligate *Arsenophonus* symbiont and one facultative *Wolbachia* symbiont (supplementary Fig. S4). From the de novo assembly, it was evident that the genome of *Wolbachia* was broken into small contigs with extremely low coverage coverage. Based on its presence as the only symbiont of *C*. *pallida* and tissue localization, we suggest the *Arsenophonus* symbiont to be relatively recent obligate endosymbiont in a nascent stage of genome reduction.
Arsenophonus symbionts of *O*. *biloba* and *C*. *pallida* were found to have similar cell shape as other obligate endosymbionts in insects and their localization in midgut bacteriome (supplementary Fig. S3, S5) and lumen of milk glands resembles localization of other endosymbionts within Hippoboscoidea group and confirms their likely obligate status.

Complete genome of *A*. *ornithomyarum* and draft genome of *A*. *crataerinae*

We assembled a complete genome of an obligate endosymbiont of *O*. *biloba* for which we propose the name *Candidatus* Arsenophonus ornithomyarum. The genome was assembled into one circular molecule with an average coverage of 118.9. As typical obligate endosymbiont, it has a reduced genome, but it does not belong to the most extremely reduced symbionts as it is apparent from intergenic regions and pseudogenes that its genome reduction is still ongoing. Its size is 874,730 bp with an average G+C content of 22.42% (Fig 1), coding density of 66.8%, and 15 pseudogenes (see also Table 1). It possesses one plasmid (11,716 bp) which encodes eight genes, among them a complete panthotenate (B5) pathway (*panBCE*), and was probably gained via horizontal gene transfer from *Sodalis*-like bacteria.
We also assembled draft genome of endosymbiont of *C*. *pallida* for which we propose the name *Candidatus* Arsenophonus crataerinae. The genome was assembled to 589 non-overlapping contigs spanning 2,985,179 bp with an average coverage of 39.1, an average G+C content of 38.35%, and a coding density of 79.8%. As the assembly was broken into numerous contigs, we were not able to determine the exact number of pseudogenes (see also Table 1).

Phylogenomic analysis of *Arsenophonus* bacteria

We reconstructed phylogeny of *Arsenophonus* endosymbionts including eight species for which whole genomes are available and using 23 single-copy orthologous genes of these taxa. We also included to our analyses obligate endosymbiont of leech *Haementeria officinalis*, *Providencia siddallii*, as a part of outgroup. According to our analyses, *Arsenophonus* bacteria form a monophyletic clade where endosymbionts of louse flies represent two independent lineages (Fig

2). The quite distant position of *Riesia* endosymbionts is probably a result of very rapid evolution of their sequences.

B-vitamin metabolism

Since endosymbionts of blood-sucking hosts are supposed to supplement them B-vitamins and cofactors missing in their diet, we examined and reconstructed pathways of these metabolites in *A*. *ornithomyarum*, *A*. *crataerinae*. We compared these two *Arsenophonus* endosymbionts from avian louse flies (*O*. *biloba*, *C*. *pallida*) with two *Arsenophonus* endosymbionts from mammalian louse flies (*M*. *ovinus*, *L*. *cervi*) (supplementary Fig. S6) and also in broad context with other endosymbionts of blood-sucking hosts for which genomic data are available (Table 2). *A*. *ornithomyarum* as other *Arsenophonus*-like bacteria from blood-sucking hosts is not able to synthesize thiamine (B1) and possesses only thiamine transporter. Interestingly, *A*. *crataerinae* is an exception of this pattern and possesses a complete thiamine pathway in addition to a thiamine transporter. Pantothenate (B5) pathway is preserved only in *A*. *ornithomyarum* and is encoded on a plasmid while CoA pathway is encoded on the genome and seems to be pseudogenized in *coaE* gene in this endosymbiont. Nicotinamide (B3) biosynthetic pathway is complete in both bacteria, but it differs from other *Arsenophonus* of mammalian louse flies in gene content and resembles the pathway of *Sodalis* bacteria of louse flies (supplementary Fig. S6 and Fig. S7). Additionally, *A*. *crataerinae* possesses *nadABC* genes for nicotinamide synthesis from aspartate which are missing in other *Arsenophonus* of Hippoboscidae. In regards to remaining B-vitamin pathways, riboflavin (B2), pyridoxine (B6), biotin (B7), and folate (B9) are intact in *A*. *ornithomyarum* and *A*. *crataerinae*. Similarly to obligate *Arsenophonus* endosymbionts, we compared B-vitamin metabolism of facultative *Sodalis* endosymbionts from one avian louse fly (*O*. *biloba*) to one mammalian louse fly (*M*. *ovinus*) (supplementary Fig S7). Both are able to synthesize all B-vitamins. Finally, *Wolbachia* from *O*. *biloba* has preserved only the riboflavin (B2) pathway.

Comparison of *Arsenophonus* bacteria genomes

We did not observe genomic stability in *Arsenophonus*-like endosymbionts of blood-sucking hosts. Genomes vary in size, coding density, and gene order (Table1, Fig 3**)**. The sizes of genomes range from 574 to 1,184 kb. In spite of this discrepancy, these endosymbionts do not differ dramatically in composition of gene functional categories (supplementary Table S2 and Fig S8). Most of genes (~30%) contribute to translation, replication and transcription as basic cellular processes. In addition to this, numerous genes (~10%) are involved in metabolic processes, especially coenzyme metabolism and transport (COG category H). Strikingly, a great number of genes (~11%) is involved in cell wall biogenesis (COG category M) in obligate *Arsenophonus* endosymbionts of Hippoboscidae in contrast to *R*. *pediculicola*.

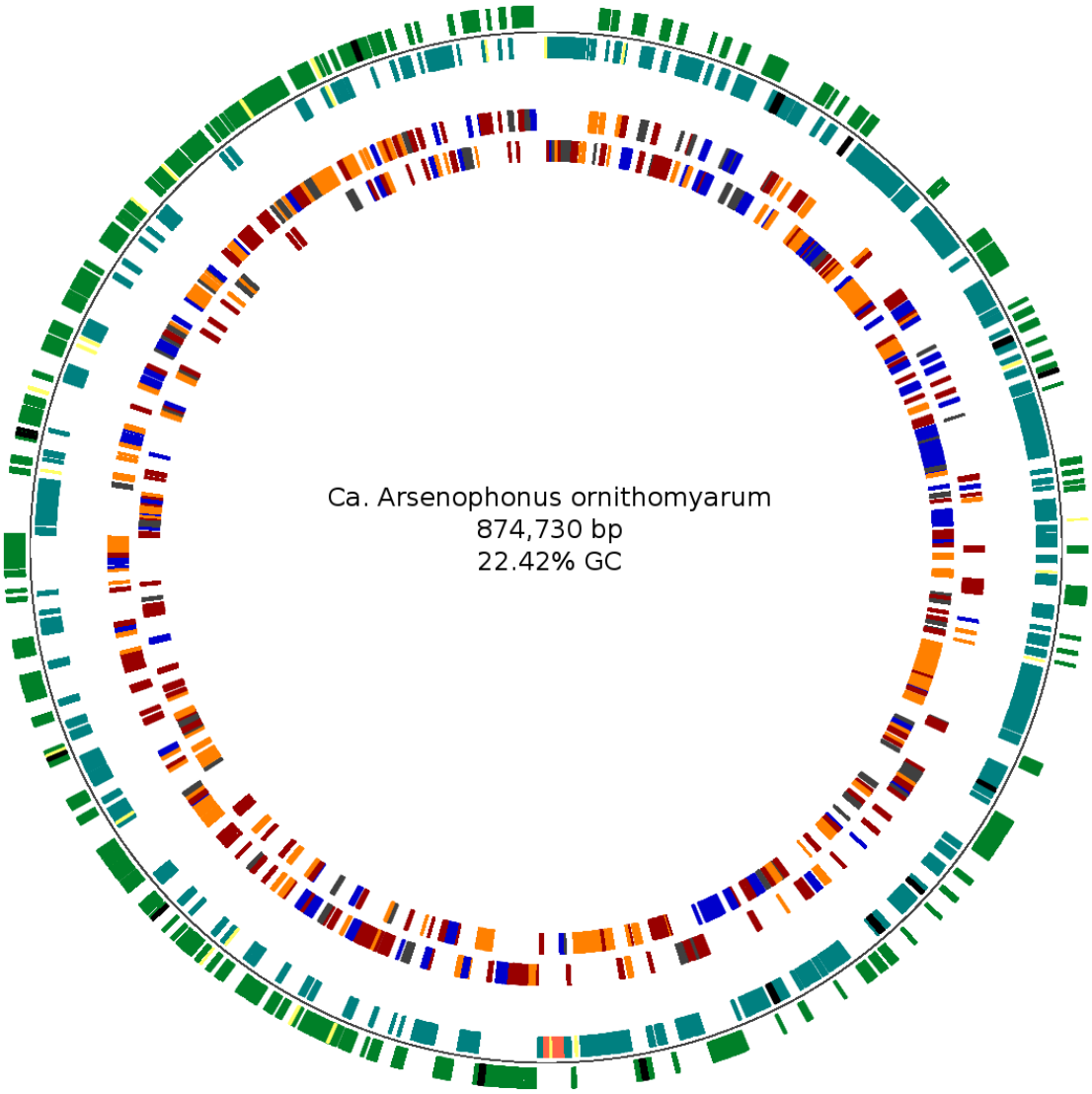

FIG. 1. – *Candidatus* Arsenophonus ornithomyarum genome. The outer circle represents gene composition of genome: protein-coding genes are green, pseudogenes are black, tRNAs are yellow, and rRNAs are orange. The inner circle shows three the most common cog categories: J (translation, ribosomal structure and biogenesis) is orange, M (cell wall/membrane/envelope biogenesis) is blue, and H (coenzyme transport and metabolism) is dark red.

Table 1

General genome features of *Arsenophonus*-like bacteria.

Endosymbiont	Size (Mb)	GC (%)	rRNA	tRNA	CDC	Pseudogenes	Coding density	References
Arsenophonus nasoniae	3.57	37.37	8-10	52	3332	NA	NA	Darby et al. 2010
Arsenophonus nilaparvatae	2.96	37.59	NA	NA	2762	NA	NA	Fan et al. 2016
Arsenophonus crataerinae	2.99	38.35	6	40	2953	NA	79.8	This study
Arsenophonus melophagi	1.19	32.48	15	41	693	27	58.6	Nováková et al., 2015
Arsenophonus lipopteni	0.84	24.87	3	35	625	16	74.7	Nováková et al., 2016
Arsenophonus ornithomyarum	0.87 11,716 bp plasmid	22.42	3	35	596	15	66.8	This study
Riesia pediculicola	0.57 7,628 bp plasmid	28.5	NA	NA	556	NA	NA	Kirkness et al., 2010
Riesia pediculischaeffi	0.58 5,159 bp plasmid	31.8	NA	NA	585	NA	NA	Boyd et al., 2014

NA – not available

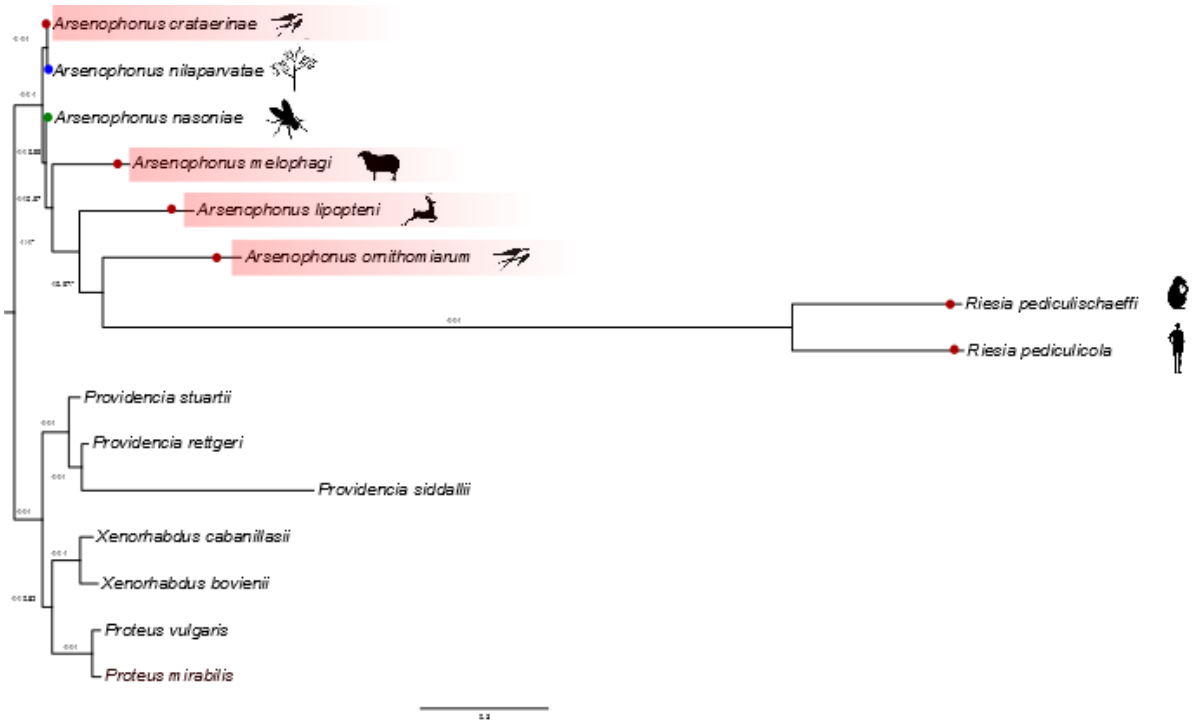

FIG. 2. – Phylogeny of *Arsenophonus* bacteria based on 23 single-copy orthologous genes. *Arsenophonus* endosymbionts of Hippoboscidae are red highlighted. Endosymbionts of blood-sucking parasites are labelled with red dot (they are also obligate ones), endosymbiont (facultative one) of phloem-sucking parasite is labelled with blue dot, and endosymbiont (facultative one) of parasitoid wasp is labelled with green dot. Silhouettes next to endosymbiont name belong to hosts of parasites which the endosymbionts colonize. Numbers at branch nodes indicate posterior probability values gained from MrBayes 3.2.6, PhyloBayes MPI 1.6j, and PhyloBayes 4.1c, respectively. Interestingly, taxa are aligned in the tree according to G+C content of their genomes (see Table 1).

Table 2

B-vitamin metabolism in obligate endosymbionts of blood-sucking hosts.

Endosymbiont	B1 (thiamine)	B2 (riboflavin)	B3 (nicotinamide)	B5 (pantothenate and acetyl CoA)	B6 (pyridoxine)	B7 (biotin)	B9 (folate)
Arsenophonus crataerinae	+ transporter	+	+	CoA only	+	+	+
Arsenophonus lipopteni	- transporter	+	+	CoA only	+	+	+
Arsenophonus melophagi	- transporter	+	+	CoA only	+	+	+
Arsenophonus ornithomyarum	- transporter	+	+	+ plasmid	+	+	+
Coxiella-like endosymbiont	+	+	+	+	+	+	+
Providencia siddallii	+	+	+	+	+	+	+
Riesia pediculicola	- transporter	+	+	+ plasmid	+	+	+
Riesia pediculischaeffi	- transporter	+	+	+ plasmid	+	+	+
Wigglesworthia glossinidia brevipalpis	+	+	+	+	+	+	+
Wigglesworthia glossinidia morsitans	+	+	+	+	+	+	+
***Wolbachia pipiens* str. wCle**	+ incomplete pseudogenes	+	-	-	+ incomplete	+	+ incomplete pseudogenes

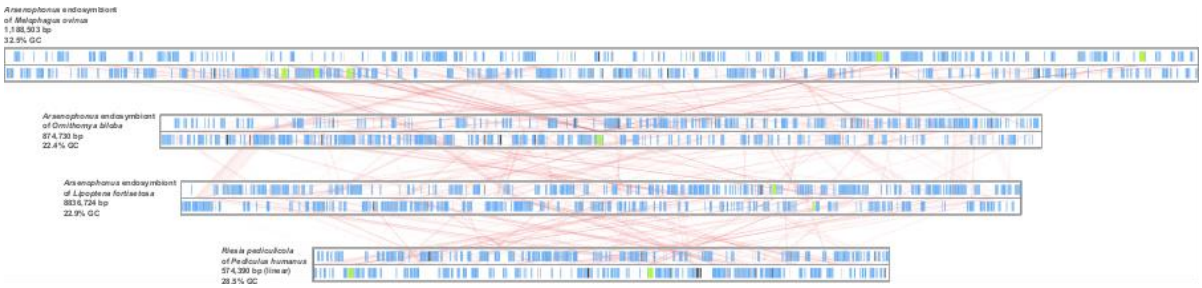

Fig. 3.–Genome alignments of *Arsenophonus* bacteria. Red lines connecting genomes show that there is no synteny between *Arsenophonus* genomes.

Discussion

Nowadays in the era of genomics, also plenty of endosymbiont genomes are being sequenced. Here we present the first genomes of endosymbionts of blood-sucking avian parasites. Both *Arsenophonus* endosymbionts are obligate endosymbionts, but they differ in age of their symbiotic association.

On one hand, *Arsenophonus ornithomyarum* represents an older obligate symbiont based on its genome features (874,730 bp; G+C content of 22.42%; coding density of 66.8%; Fig 1; Table 1). Interestingly, it also possesses a plasmid with *pan*BCE (11,716 bp) which was horizontally acquired from *Sodalis*-like bacteria. This finding is supported also by data from *O*. *biloba* without a recognizable *Sodalis* coinfection. According to 'intracellular arena' hypothesis bacteria co-infecting the same cellular environment inside their host can horizontally exchange DNA (Bordenstein & Wernegreen 2004). Bloodsucking hosts seem to support this hypothesis with one such HGT example from bedbugs (Nikoh et al. 2014) and one from louse flies (this study), both involved in B-vitamin biosynthesis.

On the other hand, *Arsenophonus crataerinae* symbiosis is presumably of much recent origin (total draft assembly size of 2,985,179 bp; G+C content of 38.35%; coding density of 79.8%; Table 1). Endosymbiont age is known to be correlated with its genome size, GC content, and coding density (McCutcheon & Moran 2012). Most of the oldest and most reduced endosymbionts tend to have GC content lower than 25% and coding density higher than 95% (Nakabachi et al. 2006; McCutcheon & Moran 2010; Sloan & Moran 2012). Interestingly, obligate *Arsenophonus* endosymbionts of louse flies possess relatively large genomes (the smallest one is *A*. *lipopteni* with 834 kb) with low coding densities in comparison to obligate endosymbionts of tsetse flies and primate lice, *W*. *glossinidia* and *Riesia* spp. (Table S1). It is thus likely that in Hippoboscidae group, the genome reduction of symbionts is still ongoing since the symbiosis does not last uninterrupted for many millions of years perhaps as a result of recurrent replacements (Nováková et al. 2009; Morse et al. 2013; Duron et al. 2014; Šochová in prep. 2016)

We have inferred *Arsenophonus* phylogeny (including four species four species from the Hippoboscidae family) using 23 single-copy orthologous genes (Fig 2). Interestingly, the multi-gene dataset produced phylogenetic trees very similar to single gene analyses based on 16S rDNA and *gro*EL genes (Nováková et al. 2009; Morse et al. 2013; Duron et al. 2014; Šochová in prep. 2016). The multi-gene analyses confirmed that *Arsenophonus* bacteria have established obligate associations at least three times within louse flies (Šochová in prep. 2016). We also consider the association of *A*. *crataerinae* with *C*. *pallida* as obligate. Although its position in the tree implies a close relationship to facultative *A*. *nilaparvatae* and its genome size does not indicate massive reduction, we provide several pieces of evidence supporting its obligate relationship with its host (Table, Fig S5, Fig S6). First, no other symbiotic microorganisms were detected in all inspected individuals. Second, its metabolic capabilities could provide B-vitamins to the host similarly as in

other louse flies. Third, it is present in the lumen of milk glands and bacteriome tissue.

In addition to *Arsenophonus*, we also detected two facultative endosymbionts within *O*. *biloba* (*Sodalis* and *Wolbachia* supergroup A), and one *Wolbachia* (supergroup A) infection within *C*. *pallida*. Both bacteria are known to be common endosymbionts of Hippoboscidae family (Dale et al. 2006; Nováková & Hypša 2007; Chrudimský et al. 2012; Šochová in prep. 2016; see also supplementary Fig S9). Total assembly size of the *Sodalis* draft genome is 4.976 Mb which makes it one of the biggest genomes from *Sodalis*-like bacteria recorded up to date. Accordingly, it shows much higher similarity to a free-living *Sodalis praecaptivus* species than to a facultative *Sodalis melophagi* endosymbiont from a related louse fly species. The *Sodalis* species detected from *O*. *biloba* thus represents an independent lineage of recently acquired facultative endosymbiont supporting a previous hypothesis that *Sodalis*-like bacteria are repeatedly acquired from environment in the Hippoboscidae group (Šochová in prep. 2016). That *Sodalis* bacteria are successful in establishment of endosymbiotic lifestyle via independent origins or repeated replacements was already well-established in numerous other insect groups (Conord et al. 2008; Koga et al. 2013; Smith et al. 2013; Michalik et al. 2014; Hosokawa 2016; Husník & McCutcheon 2016).

Since louse flies are exclusively feeding on blood meal which is deficient in B-vitamins and cofactors, their endosymbionts are suspected to supplement this scarcity. As they are parasites of birds as well as mammals, there could be a difference in vitamin supplementation to the hippoboscid host. *A*. *ornithomyarum* and *A*. *crataerinae* represent the first genomes of obligate endosymbionts of avian blood-sucking parasites which can be compared to obligate endosymbionts of mammalian bloodsucking parasites, including closely related species from the Hippoboscidae family, *A*. *melophagi* and *A*. *lipopteni*. Our results imply that obligate endosymbionts of avian and mammalian parasites differ solely in niacin (B3) biosynthesis. Both avian louse fly endosymbionts possess higher number of genes coding this pathway and *A*. *crataerinae* retains even the *nadABC* genes which were lost from all other *Arsenophonus* endosymbionts sequenced to date from bloodsucking insects (Kirkness et al. 2010; Boyd et al. 2014; Nováková et al. 2015)(Fig S6). We consider retention of these genes to be a consequence of its recent origin and unfinished genome reduction. Similarly, *A*. *crataerinae* is the first *Arsenophonus* endosymbiont which is able to synthesize thiamine (B1) and also import it from its environment. Other *Arsenophonus* bacteria of bloodsucking insects use a transporter to import thiamine from its environment (Table 2). Other interesting difference in B-vitamin metabolism is connected with pantothenate (B5) and acetyl CoA biosynthesis. Among louse flies endosymbionts, only *A*. *ornithomyarum* has a capability of B5 biosynthesis, and *panBCE* genes of this pathway are encoded by a plasmid which was acquired by horizontal gene transfer (HGT) from *Sodalis*-like bacteria. Similarly, *R*. *pediculicola* and *R*. *pediculischaeffi* retain pantothenate biosynthesis encoded on a plasmid (Kirkness et al. 2010; Boyd et al. 2014). However, it was never examined if these plasmids were gained by

HGT. In respect to other B-vitamins, riboflavin (B2), pyridoxine (B6), biotin (B7), and folate (B9) pathways are all preserved in *Arsenophonus* endosymbionts of louse flies.

Our data thus show no significant variation in B-vitamin supplementation in systems where blood-sucking parasites reside permanently on a bird or mammal host. However, there is a noticeable variability in thiamine synthesis among *Arsenophonus* symbionts. All obligate *Arsenophonus* endosymbionts with highly reduced genomes from blood-feeding hosts lost ability to synthesize thiamine. Interestingly, other obligate highly reduced endosymbiotic bacteria of blood-feeders such as *Wigglesworthia glossinidia*, *Providencia siddallii* or *Coxiella*-like endosymbiont retain this pathway as complete (Table 2). We hypothesize that differences in their host lifestyle could explain this gene loss. Most of louse flies and all lice are permanent blood-feeders while tsetse flies, ticks, and leeches suck blood intermittently. It therefore seems likely that permanent parasites have sufficient thiamine supply from their host blood not to require it from their endosymbionts. The only exception to this rule is *Wolbachia* wCle from bedbugs. In spite of bed bugs being intermittent parasites, this mutualist does not synthesize B1. It was shown that *Wolbachia* possesses solely biotin and riboflavin pathways in this symbiosis (Hosokawa et al. 2010; Nikoh et al. 2014; Moriyama et al. 2015).

As mentioned above, genomes of *Arsenophonus* endosymbionts have diverse sizes and when aligned, show no stability and differ markedly in gene order and coding densities (Fig 2). The same pattern could be seen among obligate intrabacterial-endosymbionts of mealybugs which were acquired several times independently (Husník & McCutcheon 2016). These data provide additional support for the hypothesis about independent origin of *Arsenophonus* endosymbiosis within Hippoboscidae. In terms of functional categories, we did not observe striking variation in distribution of genes into COG categories between *Arsenophonus*-like bacteria (Table S2, Fig S8). There was also no apparent contrast between bacteria from avian and mammalian parasites. The more reduced genomes possess proportionally lower number of genes in individual COG categories. The highest number of genes is involved in information storage and processing (J, L, K, A categories), i.e. fundamental cell processes. Coenzyme transport and metabolism (H category) is also one of the most abundant which nicely underlines nutritional basis of the symbioses. The only striking difference among *Arsenophonus* endosymbionts concerns cell wall biogenesis (M category). *R. pediculicola* possesses only about a half of genes involved in this process in contrast to *Arsenophonus* symbionts from louse flies. In this COG category and category I (lipid transport and metabolism), there is also a clear difference between endosymbionts of blood-feeding and sap-feeding insects which retain significantly lower number of genes contributing to cell envelope production (McCutcheon & Moran 2012). We note that this could be a consequence of their residence in host derived symbiosomal membrane which controls symbiont nutrient supply (Price et al. 2014; Duncan et al. 2014). Such an intimate integration and control likely contributes to their massive genome reduction and retention of only the most essential genes

(McCutcheon & Moran 2010; McCutcheon & von Dohlen 2011; Sloan & Moran 2012). On the other hand, endosymbionts of blood-sucking endosymbionts are not engulfed by such host-derived membrane and can freely uptake nutrients from their host cytoplasm. As a result, their genomes do not undergo such an extreme genome reduction. Since the host-symbiont integration is not so interdependent, they can be easily replaced by other bacteria. This phenomenon seems to be very common in the Hippoboscidae family (Šochová in prep, 2016). In lice, symbiont replacements have not yet been shown to be common, but it was hypothesized that their symbioses originated several times independently (Allen et al. 2015). Only tsetse fly symbiosis appears to be stable over many millions of years, but since the group contains only 22 species, replacements could be strongly influenced by chance (Chen et al. 1999).

A. *ornithomyarum* and *A*. *crataerinae* represent the very first genomes of endosymbionts from blood-sucking parasites of birds. Even though we did not observe a striking difference between these endosymbionts of blood-sucking parasites of birds and mammals, these genomes still provide valuable information about evolution of endosymbiosis in the Hippoboscidae group which seems to be much more tangled than anticipated. Obligate *Arsenophonus* symbioses were repeatedly and independently established in this blood-feeding group and led to independent genome reduction converging on similar gene content. Interestingly, our data also show that *Arsenophonus*-*Sodalis* coinfections in this blood-sucking parasitic group resulted in exchange of genetic material via a plasmid transfer essential for the nutritional role of the symbiosis.

Supplementary Material

Supplementary tables S1 and S2 and supplementary figures S1-S9 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

Acknowledgement

This work was supported by the Grant Agency of the Czech Republic (grant 13-01878S to VH).

Literature cited

Akman L et al. 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. Nat. Genet. 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. Int. J. Syst. Bacteriol. 45:848–51. doi: 10.1099/00207713-45-4-848.

Allen JM, Burleigh JG, Light JE, Reed DL. 2015. Effects of 16S rDNA sampling on estimates of endosymbiont lineages in sucking lice. doi: 10.7287/peerj.preprints.1457v1.

Allen JM, Reed DL, Perotti MA, Braig HR. 2007. Evolutionary relationships of '*Candidatus* Riesia spp.,' endosymbiotic enterobacteriaceae living within hematophagous primate lice. Appl. Environ. Microbiol. 73:1659–64. doi: 10.1128/AEM.01877-06.

Bankevich A et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19:455–77. doi: 10.1089/cmb.2012.0021.

Benoit JB et al. 2016. Unique features of a global human ectoparasite identified through sequencing of the bed bug genome. Nat. Commun. 7:10165. doi: 10.1038/ncomms10165.

Ben-Yakir D. 1987. Growth retardation of *Rhodnius prolixus* symbionts by immunizing host against

Nocardia (*Rhodococcus*) *rhodnii*. J. Insect Physiol. 33:379–383. doi: 10.1016/0022-1910(87)90015-1.
Bordenstein SR, Wernegreen JJ. 2004. Bacteriophage flux in endosymbionts (Wolbachia): infection frequency, lateral transfer, and recombination rates. Mol. Biol. Evol. 21:1981–91. doi: 10.1093/molbev/msh211.
Boyd BM et al. 2016. Two bacteria, Sodalis and Rickettsia, associated with the seal louse *Proechinophthirus fluctus* (Phthiraptera: Anoplura). Appl. Environ. Microbiol. AEM.00282–16–. doi: 10.1128/AEM.00282-16.
Boyd BM, Allen JM, de Crécy-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 (Bethesda). 4:2189–95. doi: 10.1534/g3.114.012567.
Brelsfoard C et al. 2014. Presence of extensive *Wolbachia* symbiont insertions discovered in the genome of its host *Glossina morsitans morsitans*. PLoS Negl. Trop. Dis. 8:e2728. doi: 10.1371/journal.pntd.0002728.
Conesa A et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 21:3674–6. doi: 10.1093/bioinformatics/bti610.
Conord 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. 25:859–68. doi: 10.1093/molbev/msn027.
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. 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*. Int. J. Syst. Bacteriol. 49 Pt 1:267–75. doi: 10.1099/00207713-49-1-267.
Dale JM, Popescu L, Karp PD. 2010. Machine learning methods for metabolic pathway prediction. BMC Bioinformatics. 11:15. doi: 10.1186/1471-2105-11-15.
Duncan RP et al. 2014. Dynamic recruitment of amino acid transporters to the insect/symbiont interface. Mol. Ecol. 23:1608–23. doi: 10.1111/mec.12627.
Duron O et al. 2014. Origin, acquisition and diversification of heritable bacterial endosymbionts in louse flies and bat flies. Mol. Ecol. 23:2105–17. doi: 10.1111/mec.12704.
Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16:157. doi: 10.1186/s13059-015-0721-2.
Guindon S, Gascuel O. 2003. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Syst. Biol. 52:696–704. doi: 10.1080/10635150390235520.
Hampton T. 2010. Body Louse Genome. JAMA. 304:733. doi: 10.1001/jama.2010.1081.
Hansen AK, Moran NA. 2011. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. Proc. Natl. Acad. Sci. U. S. A. 108:2849–54. doi: 10.1073/pnas.1013465108.
Hosokawa T et al. 2016. Obligate bacterial mutualists evolving from environmental bacteria in natural insect populations. Nat. Microbiol. 1:1–7. doi: 10.1038/nmicrobiol.2015.11.
Hosokawa T et al. 2012. Reductive genome evolution, host-symbiont co-speciation and uterine transmission of endosymbiotic bacteria in bat flies. ISME J. 6:577–87. doi: 10.1038/ismej.2011.125.
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. 107:769–74. doi: 10.1073/pnas.0911476107.
Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 17:754–755. doi: 10.1093/bioinformatics/17.8.754.
Husník F et al. 2013. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. Cell. 153:1567–78. doi: 10.1016/j.cell.2013.05.040.
Husník F, McCutcheon JP. 2016. Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. Cold Spring Harbor Labs Journals doi: 10.1101/042267.
Hypša V, 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. 47:1140–4. http://www.ncbi.nlm.nih.gov/pubmed/9336921.
Chen X, Li S, Aksoy S. 1999. Concordant evolution of a symbiont with its host insect species:

molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. J. Mol. Evol. 48:49–58. http://www.ncbi.nlm.nih.gov/pubmed/9873076.
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.
Initiative IGG. 2014. Genome sequence of the tsetse fly (*Glossina morsitans*): vector of African trypanosomiasis. Science. 344:380–6. doi: 10.1126/science.1249656.
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:D109–14. doi: 10.1093/nar/gkr988.
Katoh K. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066. doi: 10.1093/nar/gkf436.
Katoh K, Asimenos G, Toh H. 2009. Multiple alignment of DNA sequences with MAFFT. Methods Mol. Biol. 537:39–64. doi: 10.1007/978-1-59745-251-9_3.
Kearse M et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28:1647–9. doi: 10.1093/bioinformatics/bts199.
Keseler IM et al. 2013. EcoCyc: fusing model organism databases with systems biology. Nucleic Acids Res. 41:D605–12. doi: 10.1093/nar/gks1027.
Kirkness EF 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. 107:12168–73. 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. Environ. Microbiol. 15:2073–81. doi: 10.1111/1462-2920.12121.
Kŕířek J, Hypša V. 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.
Kumar S, Bala M, Raghava GPS, Mayilraj S. 2013. Draft Genome Sequence of *Rhodococcus triatomae* Strain BKS 15-14. Genome Announc. 1:e0012913. doi: 10.1128/genomeA.00129-13.
Kumar S, Jones M, Koutsovoulos G, Clarke M, Blaxter M. 2013. Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. Front. Genet. 4:237. doi: 10.3389/fgene.2013.00237.
Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics. 25:2286–8. doi: 10.1093/bioinformatics/btp368.
Lartillot N, Rodrigue N, Stubbs D, Richer J. 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Syst. Biol. 62:611–5. doi: 10.1093/sysbio/syt022.
Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 31:1674–6. doi: 10.1093/bioinformatics/btv033.
Luan J-B et al. 2015. Metabolic coevolution in the bacterial symbiosis of whiteflies and related plant sap-feeding insects. Genome Biol. Evol. 7:2635–47. doi: 10.1093/gbe/evv170.
Manzano-Marín A, Oceguera-Figueroa A, Latorre A, Jiménez-García LF, Moya A. 2015. Solving a Bloody Mess: B-vitamin independent metabolic convergence among gammaproteobacterial obligate endosymbionts from blood-feeding arthropods and the leech *Haementeria officinalis*. Genome Biol. Evol. 7:2871–84. doi: 10.1093/gbe/evv188.
McCutcheon JP, von Dohlen CD. 2011. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Curr. Biol. 21:1366–72. doi: 10.1016/j.cub.2011.06.051.
McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. Nat. Rev. Microbiol. 10:13–26. doi: 10.1038/nrmicro2670.
McCutcheon JP, Moran N a. 2010. Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. Genome Biol. Evol. 2:708–18. doi: 10.1093/gbe/evq055.
McLean DL, Houk EJ. 1973. Phase contrast and electron microscopy of the mycetocytes and symbiotes of the pea aphid, *Acyrthosiphon pisum*. J. Insect Physiol. 19:625–633. doi: 10.1016/0022-1910(73)90071-1.
Michalik A, Jankowska W, Kot M, Gołas A, Szklarzewicz T. 2014. Symbiosis in the green leafhopper,

Cicadella viridis (Hemiptera, Cicadellidae). Association in statu nascendi? Arthropod Struct. Dev. 43:579–87. doi: 10.1016/j.asd.2014.07.005.
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. 80:5844–53. doi: 10.1128/AEM.01150-14.
Moran NA, Bennett GM. 2014. The Tiniest Tiny Genomes. 195–215. doi: 10.1146/annurev-micro-091213-112901.
Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35:W182–5. doi: 10.1093/nar/gkm321.
Moriyama M, Nikoh N, Hosokawa T, Fukatsu T. 2015. Riboflavin provisioning underlies *Wolbachia*'s fitness contribution to Its insect host. MBio. 6:e01732–15–. doi: 10.1128/mBio.01732-15.
Morse SF et al. 2013. Evolution, multiple acquisition, and localization of endosymbionts in bat flies (Diptera: Hippoboscoidea: Streblidae and Nycteribiidae). Appl. Environ. Microbiol. 79:2952–61. doi: 10.1128/AEM.03814-12.
Morse SF, Dick CW, Patterson BD, Dittmar K. 2012. Some like it hot: evolution and ecology of novel endosymbionts in bat flies of cave-roosting bats (Hippoboscoidea, Nycterophiliinae). Appl. Environ. Microbiol. 78:8639–49. doi: 10.1128/AEM.02455-12.
Nakabachi A et al. 2006. The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. Science. 314:267. doi: 10.1126/science.1134196.
Nikoh N et al. 2014. Evolutionary origin of insect-*Wolbachia* nutritional mutualism. Proc. Natl. Acad. Sci. 111:10257–10262. doi: 10.1073/pnas.1409284111.
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. 81:6189–99. 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 Microbiol. Lett. 269:131–5. 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 Microbiol. 9:143. doi: 10.1186/1471-2180-9-143.
Pachebat JA et al. 2013. Draft genome sequence of *Rhodococcus rhodnii* Strain LMG5362, a Symbiont of *Rhodnius prolixus* (Hemiptera, Reduviidae, Triatominae), the Principle Vector of Trypanosoma cruzi. Genome Announc. 1:e00329–13. doi: 10.1128/genomeA.00329-13.
Price DRG et al. 2014. Aphid amino acid transporter regulates glutamine supply to intracellular bacterial symbionts. Proc. Natl. Acad. Sci. U. S. A. 111:320–5. doi: 10.1073/pnas.1306068111.
Rio RVM et al. 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.
Rosenfeld JA et al. 2016. Genome assembly and geospatial phylogenomics of the bed bug *Cimex lectularius*. Nat. Commun. 7:10164. doi: 10.1038/ncomms10164.
Rutherford K et al. 2000. Artemis: sequence visualization and annotation. Bioinformatics. 16:944–945. doi: 10.1093/bioinformatics/16.10.944.
Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 30:2068–9. doi: 10.1093/bioinformatics/btu153.
Sloan DB et al. 2014. Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. Mol. Biol. Evol. 31:857–71. doi: 10.1093/molbev/msu004.
Sloan DB, Moran NA. 2012. Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. Mol. Biol. Evol. 29:3781–92. doi: 10.1093/molbev/mss180.
Smith CL, Weiss BL, Aksoy S, Runyen-Janecky LJ. 2013. Characterization of the achromobactin iron acquisition operon in *Sodalis glossinidius*. Appl. Environ. Microbiol. 79:2872–81. doi: 10.1128/AEM.03959-12.
Smith TA, Driscoll T, Gillespie JJ, Raghavan R. 2015. A *Coxiella*-like endosymbiont is a potential vitamin source for the Lone Star tick. Genome Biol. Evol. 7:831–8. doi: 10.1093/gbe/evv016.
Snyder AK, Deberry JW, Runyen-Janecky L, Rio RVM. 2010. Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. Proc. Biol. Sci. 277:2389–97. doi: 10.1098/rspb.2010.0364.
Snyder AK, Rio RVM. 2015. '*Wigglesworthia morsitans*' folate (vitamin B9) biosynthesis contributes to tsetse host fitness. Appl. Environ. Microbiol. 81:5375–86. doi: 10.1128/AEM.00553-15.
Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and

ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56:564–77. doi: 10.1080/10635150701472164.
Tatusov RL et al. 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics. 4:41. doi: 10.1186/1471-2105-4-41.
Toh H et al. 2006. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. Genome Res. 16:149–156. 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). J. Invertebr. Pathol. 91:64–8. doi: 10.1016/j.jip.2005.08.009.
Walker BJ et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 9:e112963. doi: 10.1371/journal.pone.0112963.
Wernegreen JJ, Kauppinen SN, Degnan PH. 2010. Slip into something more functional: selection maintains ancient frameshifts in homopolymeric sequences. Mol. Biol. Evol. 27:833–9. doi: 10.1093/molbev/msp290.
Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 31:3350–2. doi: 10.1093/bioinformatics/btv383.
Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics. 30:614–20. doi: 10.1093/bioinformatics/btt593.