School of Doctoral Studies in Biological Sciences
University of South Bohemia in České Budějovice
Faculty of Science

Genome evolution of endosymbiotic bacteria in blood-sucking insects

Ph.D. Thesis

RNDr. Jana Říhová

Supervisor: prof. RNDr. Václav Hypša, CSc.
Specialist: doc. RNDr. Eva Nováková, Ph.D.

Faculty of Science, University of South Bohemia in České Budějovice

České Budějovice 2023

This thesis should be cited as:

Říhová, J., 2023: Genome evolution of endosymbiotic bacteria in blood-sucking insects. Ph.D. Thesis. University of South Bohemia, Faculty of Science, School of Doctoral Studies in Biological Sciences, České Budějovice, Czech Republic, České Budějovice, Czech Republic, 111 pp.

Annotation

The research on insect-bacteria symbiosis is experiencing a rapid burst leading to the identification of much higher complexity and dynamics of the insect-bacteria associations than ever before envisaged. The generally accepted concept of obligate bacterial symbionts as providers of essential compounds (vitamins, aminoacids) has been further developed into the framework of microbiomes, i.e. communities of microorganisms associated with the host. The majority of the knowledge has been derived from plant sap-feeders, while data on blood-feeding insects remain fragmented and even inconsistent. In this thesis, we employ amplicon, genomic and *in-vivo* approaches to investigate several symbiotic systems of obligate blood-feeders and address the following questions: What is the diversity of symbiotic bacteria in our studied systems? What are the metabolic capacities of these symbionts? What is the role of symbionts in the biology of their hosts? What are the possible sources of bacteria prone to establish symbiosis with obligate blood-feeders? Where these symbionts reside in the host´s body? What is the coevolutionary pattern of these bacteria with their hosts? As an answer to these questions, I present the results detailed in the course of this thesis, which are preceded by a comprehensive introduction to the studied topics and wraps up with a summary of the main results.

Declaration

I hereby declare that I am the author of this dissertation and that I have used only those sources and literature detailed in the list of references.

České Budějovice, 22. 2. 2023

…….

Jana Říhová

This thesis originated from the **Faculty of Science, University of South Bohemia** supporting doctoral studies in the Parasitology study program.

Financial support

The research presented in this thesis was financially supported by the grants of the Czech Science Foundation (projects #18-07711S, #20-07674S and #21-02532S of Václav Hypša) and Grant Agency of the University of South Bohemia (SGA 2016 of Jana Říhová).

Acknowledgements

Firstly, I would like to thank my great supervisor Václav Hypša for his patience, support, wise and very valuable advices and guidance throughout my whole studies. I believe that he is one of the top lead scientists in this field of research, and it has been my honor to be part of his group. Specifically, I have always appreciated his way of teaching me how to communicate my research with the science world. Sometimes it has been challenging to stay motivated and "sell the story", but we have never given up. Secondly, Eva Nováková as my supervisor-specialist deserves many thanks for boosting our papers with supplementary analyses and new fresh ideas and also for introducing me to many new methods. She is a very experienced and professional leader known for her research achievements across the whole world. I would also like to thank my US supervisor Jessica E. Light, who I very appreciate for the warm welcome to her team (even in the bad COVID times) and for letting me use her sample collections and all of her laboratory gear. She also introduced me to other scientists in lice-oriented research (Kayce C. Bell, Lance A. Durden and others), for which I am very grateful.

Many thanks go to all of my current and former colleagues (Jana Martinů, Jitka Kocková, Anetka Trefancová, Lucka Albrechtová, Vita Kuročenko, Jirka Ťápal, Honza Zima, Damián Dedecius, Jana Kvičerová, Sonia Rodríguez-Ruano, Anna M. Floriano, Míša Fickerová, Katka Žižková, Klára Charvátová, Hassan Tarabai, Anna Mácová and many others) from our laboratory or abroad for any help, always very friendly atmosphere and countless inspiring ideas flourishing not only this research, but my life as well. It has been my greatest honor to work with you all and also spend some fun time with you!

Special thanks I would like to dedicate to my beloved fiancé (Benjamin S. Martin), first family (mum, dad, my brothers and their wives and others in heaven) and my second family (friends) who have unconditionally supported me and have never ever given up on me. They have provided me with their endless support, kindness and their way of understanding. Certainly, I never ever would have been able to get this far without them.

Let me finish with this cliché:

Thank you all for everything, you are simply the best and will be forever remembered!

List of papers and author's contribution

The thesis is based on the following papers (listed chronologically):

I. **Říhová, J.**, Batani, G., Rodríguez-Ruano, S. M., Martinů, J., Vácha, F., Nováková, E., & Hypša, V. (2021). A new symbiotic lineage related to *Neisseria* and *Snodgrassella* arises from the dynamic and diverse microbiomes in sucking lice. *Molecular Ecology*, *30*(9), 2178-2196. doi: 10.1111/mec.15866; (IF = 6.622)

JŘ, VH and EN designed the study. JŘ and JM carried out sample collection. JŘ was responsible for the sequencing part, assemblies of genomes, genomic comparisons, phylogenetic analyses of the hosts and the symbionts, and partially a reconstruction of metabolic analyses. EN and SMRR were responsible for the analyses of amplicon sequencing data. JŘ and GB developed the protocol for visualization of symbionts using the FISH method. FV and VH designed the detailed overview of metabolic pathways. All authors participated in data analyses and manuscript writing. JŘ contribution was 50%.

II. **Říhová, J.**, Bell, K. C., Nováková, E., & Hypša, V. (2022). *Lightella neohaematopini*: A new lineage of highly reduced endosymbionts coevolving with chipmunk lice of the genus *Neohaematopinus*. *Frontiers in Microbiology*, *13*, 900312. doi: 10.3389/fmicb.2022.900312; (IF = 6.064)

JŘ, VH and EN designed the study. JŘ participated in the genomic, phylogenetic and metabolic analyses of the symbionts with VH. The metagenomic binning of reads was done by EN. KCB carried out the phylogenetic analyses of the louse hosts. JŘ, VH and EN wrote the manuscript. JŘ contribution was 70%.

Co-author agreement

Prof. RNDr. Václav Hypša, CSc., the supervisor of this Ph.D. thesis and the co-author of the papers I and II, fully acknowledges the stated contribution of RNDr. Jana Říhová to these manuscripts.

……

prof. RNDr. Václav Hypša, CSc.

Doc. RNDr. Eva Nováková, Ph.D. the supervisor-specialist of this Ph.D. thesis and the co-author of the papers I and II, fully acknowledges the stated contribution of RNDr. Jana Říhová to these manuscripts.

……

doc. RNDr. Eva Nováková, Ph.D.

Contents

INTRODUCTION

Symbiosis as a driving force of evolution

Symbiosis between organisms is an important driving force in nature and it plays a key role in many ecological and evolutionary processes. The research of symbiotic relationships brought a new perspective on the evolution of eukaryotic organisms. Examples of symbiosis can be found in all kingdoms of life. One of the most essential symbioses originated during the establishment of the eukaryotic cell via the symbiosis of the ancestor of the eukaryotic cell with α-proteobacteria (later mitochondria) and cyanobacteria (later plastids; Gray and Doolitle, 1982; Margulius, 1970).

The symbiotic relationship of bacteria with insects is a common feature which has been extensively studied. In 1965 zoologist and cytologist Paul Buchner was studying insect morphology and he noticed a considerable diversity of microorganisms living inside the insect bodies. He described many types of bacteria with different shapes, which enter mutualistic relationships with their hosts (Buchner, 1965). With the development of molecular methods, we could observe that symbiotic relationships underpin many crucial biological processes of hosts, both within individuals and also at the level of their evolutionary and ecological diversifications. These symbiotic relationships are typically observed in insect hosts which feed on nutrient-poor sources, such as vertebrate blood or plant phloem/xylem. The bacterial symbionts are known to utilize their metabolic pathways to compensate for the deficiency of essential vitamins or amino acids in the insect´s diet (Moran et al., 2008; Moya et al., 2008; Baumann, 2005; Akman et al., 2002; Nogge, 1976).

Basic characterization of symbiotic forms

According to the level of integration with the host, we differentiate between obligate (primary, P-) and facultative (secondary, S-) symbionts. P-symbionts are vertically (most often transovarially) transferred to progeny, and they usually share a long coevolutionary history with their hosts. These symbionts are typically harbored in a specialized organ, formerly mycetome (Douglas, 1989), or more appropriately bacteriome (Baumann et al., 1995). Bacteriome-associated symbionts, for example *Buchnera aphidicola* in aphids, are "domesticated" by the host and they have lost their ability to invade naive hosts, so they are dependent on host-mediated mechanisms of transmission (McCutcheon and Moran, 2012). Genomes of P-symbiont usually undergo an extensive reduction by gene loss and have low GC content compared to the genomes of their free-living relatives. However, these symbionts retain genes important for the host and their own survival, such as genes involved in the biosynthesis of essential metabolites deficient in the insect diet. Functional dependencies can lead to many interesting molecular features such as the complementarity of metabolic pathways of the host and symbiont or the gain of missing genes through horizontal gene transfer (HGT). For example, *Wolbachia*, a frequent insect parasite, turned into a mutualistic

bacterium in bedbug *Cimex lectularius* via horizontal acquisition of an operon coding for biotin biosynthesis (Gerth and Bleidorn, 2016; Nikoh et al., 2014). I found the same trait in *Legionella* symbiont of the louse *Polyplax serrata*, which is also presumed to have established nutritional symbiosis due to horizontally acquired operon, completing the metabolic pathway of biotin (Říhová et al., 2017).

On the other hand, facultative S-symbionts retain the ability of horizontal transfer, and they can repeatedly invade new unrelated hosts (Moran et al., 2008; Dale and Moran, 2006; Darby et al., 2005; Dale and Maudlin, 1999; Hypša and Dale, 1997). These symbionts can provide benefits to their host, e. g., symbiotic *Hamiltonela defensa* mediates protection of the aphid host against parasitoid wasps through the production of toxins, which harms wasp larvae (Moran et al., 2008; Oliver et al., 2003). Several endosymbionts, such as *Wolbachia* and *Cardinium*, are described as reproductive parasites that manipulate their hosts in various ways through cytoplasmic incompatibility, feminization, male-killing or thelytokous parthenogenesis (White et al., 2013). These strategies lead to an increase of the number of infected hosts within the population, even if the fitness of the host is reduced (Brownlie and Johnson, 2009; Gotoh et al., 2007; Hunter et al., 2003). During the long-term interaction of S-symbiont with the host, evolutionary forces can drive both to establish an obligate relationship. Also, a metabolically more efficient S-symbiont can contribute to the new functions and eventually it can replace the decaying P-symbiont (Manzano-Marín et al., 2020, Koga and Moran, 2014; Toenshoff et al., 2012), or they can establish a cooperative system together. The symbionts then cooperate in the provision of essential metabolites to the host by the complementarity of their metabolic pathways (Husník et al., 2013; Koga et al., 2013; McCutcheon and von Dohlen, 2011).

Methodological advances in host-symbiont research

Significant progress in the study of symbionts was achieved by introduction and advances of several new methods allowing better and more specific visualization, screening through host populations, and detailed genomic analyses. The method enabling visualization of endosymbionts *in-vivo* is based on *in-situ* hybridization of fluorescent-labelled probes on symbiotic DNA or RNA (FISH). The fluorescent signals then allow for the detection of culturable but more importantly also unculturable organisms (typically P-symbionts), and therefore it can elucidate the composition and the distribution of complex microbial communities inside the host. Moreover, using the *in-vivo* methods we can also study the morphology of symbionts, the number of bacterial cells, or their spatial distribution in a host body (Moter and Göbel, 2000). For example, implementation of this method in *Polyplax-Legionella* model enabled us to visualize the pathway of symbiont´s transmission to the next generation through female ovaries (Říhová, 2017). Formerly, many screening studies on the endosymbiotic communities were based on Sanger sequencing of bacterial DNA amplified with specific primers. These techniques rely on strong prior assumptions concerning the bacterial strains generally present in the studied organisms. In the past decade, symbiosis research experienced a remarkable shift from these single-gene analyses (with

limited throughput) towards high-throughput sequencing methods. Such methods consisting of many different strategies that are based on a combination of techniques specific for template preparation, nucleic acid sequencing and imaging, and data analyses including genome alignment and assembly (Lemmon and Lemmon, 2013; Metzker, 2009).

One of the major high-throughput sequencing methods used in microbiome-oriented research is the amplicon sequencing of PCR amplified fragments of 16S rRNA bacterial gene. This method allows for the detection of all bacteria species in the sample (Degnan and Ochman, 2012). This method was implemented for the study of environmental sample communities and later it was also used for the investigation of bacterial communities of the aphid microbiome (Gauthier et al., 2015; Bansal et al., 2014; Jing et al., 2014; Jones et al., 2011). Nowadays amplicon sequencing has become an essential method for studying composition and changes of microbiomes and for comparing them on a broader scale, for example, in the context of ontogenic stages, environment, and species identity (Brown et al. 2020).

The advantage of the high-throughput sequencing is the investigation of the diversity of bacteria in sample without its prior knowledge. Moreover, the abundances of sequencing reads assigned to each bacterial strain can serve as a proxy for the relative abundance of each type of bacteria in a sample (Bansal et al., 2014; Otani et al., 2014). This procedure is designed to cluster sequences based on their similarity into higher units called OTUs (operational taxonomic units) or ASVs (amplicon sequence variants). Results of amplicon sequencing in our model *Polyplax serrata* lice revealed a surprising diversity of bacterial OTUs (**MS1**). Amplicon analyses of bacterial OTUs show multiple dominant OTUs mostly assigned to P-symbiont *Legionella polyplacis* and other possibly coevolving P-symbionts, *Buchnera* and *Arsenophonus*, presented in a non-random pattern and possessing low GC content (<50%). In several samples, we also found the dominant OTU assigned to Neisseriaceae with a higher GC content (>50%) indicating either an environmental contamination or a bacterium in an early stage of symbiosis (**MS1**).

Results of amplicon sequencing could also serve as the background for the targeting samples for the consequent whole-genome sequencing. The genomic data for symbionts could be generated either from the pure material (single-cell templates, bacterial cultures, etc.) or symbiont-containing tissues (bacteriomes, gut, whole abdomen, etc.). The latter approach is more commonly used in symbiont research as a majority of symbiotic bacteria is not culturable on growth media (Singh et al., 2019; Kukuchi and Fukatsu, 2013; Tringe and Rubin, 2005). When using the host tissues, the resulting dataset will contain metagenomic data for microbiota and host, which could be used for multiple analyses (study of evolutionary patterns, metabolic complementation, genetic exchange and/or modification between symbionts, host-symbiont coevolution, etc.). On the other hand, the assembly of individual genomes from metagenomic data could pose several methodical challenges. In particular, the occurrence of similar strains/species – sometimes present at greatly uneven ratios – could drastically reduce the quality of the reconstructed genomes (Meiser et al., 2017). A prior knowledge of the general genomic characteristics of symbionts (e.g., based on amplicon screening), could be important for correct data processing.

Methodical advances have reached the stage where microbial genomes can be sequenced routinely by short-read Illumina sequencing, which produce highly accurate reads (99,9%) with length 75-300 bp (Goodwin et al, 2016). Being the most economical, Illumina currently dominates among sequencing techniques used in symbionts-oriented research. Using the short reads, however, challenges the assembly process towards the complete genomes (especially larger genomes of S-symbiont; Cahill et al., 2010; Whiteford et al., 2005). These difficulties are mainly caused by sequence repetitiveness in free-living bacteria and young symbionts (Kingsford et al., 2010; Whiteford et al., 2005). This could be particularly overcome by using long-read (single-molecule) sequencing technologies such as the Oxford Nanopore and PacBio (Goldstein et al., 2019; Rhoads and Au, 2015). Previously, these techniques had been generating long reads with higher error rates and thus the assembly of long reads required polishing with Illumina short reads (Magi et al., 2018; George et al., 2017; Sović et al., 2017). However, with the current progress of certain platforms and used chemistry, error rates of long reads are highly similar to error rates of Illumina sequencing technologies (98% for Oxford Nanopore PromethION and 99% for PacBio HiFi; Lin et al., 2021).

The recently established Hi-C method also provides a crucial improvement to the efficiency of genome assembly (Burton et al., 2014). Although this technique was primarily developed for chromosomal assemblies of eukaryotes, in combination with long reads it could also be highly efficient for assemblies of large bacterial genomes from metagenomic templates (Stewart et al., 2018; Press et al., 2017; Marbouty et al., 2014). The Hi-C method introduces cross-links in the spatial structure of all DNA molecules within a single cell, producing a library of ligated fragments found in close physical proximity. Combined with regular sequencing techniques (short reads/long reads), sequence and proximity information from Hi-C libraries allows for metagenomic deconvolution (including plasmids), yielding reference quality assemblies (as demonstrated even for the highly complex cattle rumen microbial community; Stewart et al., 2018).

Genomic background of endosymbiosis

The research based on the analyses of high-throughput data resulted in general concept of typical characteristics describing genomes of symbionts and their evolution. A common feature documented across a spectrum of symbiotic bacteria is the genome reduction in comparison to the genome sizes of free-living relatives. Their genomes can be extensively economized (in extreme cases resembling the genomes of organelles, i.e., mitochondria or chloroplast). These extreme genome reductions have been described in several symbionts of phytophagous insects (reaching 112 kbp; Bennett et al., 2016). In contrast, the smallest genomes of symbionts reported from hematophagous insects oscillate around 500 kbp (**MS2**; Boyd et al., 2017; Říhová et al., 2017; Boyd et al. 2014; Kirkness et al. 2010). One of the plausible explanations for the minimal size differences could be the level of integration within the host cells, where the tiny symbionts of phytophagous insects are restricted in host-derived symbiosomal membrane, while symbionts of hematophagous insects are usually localized in the cytoplasm of the host cells and retain more sets of genes, e.g., coding for the

synthesis of their bacterial cell wall (Husník, 2018; Moran and Bennett, 2014; Rio et al., 2012; McCutcheon and Moran, 2012). From an evolutionary perspective, all obligate symbionts have been generally presumed to begin as facultative ones (Nikoh et al., 2014). The whole evolutionary process of symbiotic transformation, from free-living bacterium to P-symbiont via S-symbiont, shows many typical characteristics, which are similar in the broad spectrum of phylogenetic lineages.

The early stages of genome reduction are represented by some more recently evolved symbionts and are characterized by the proliferation of mobile elements, the formation of pseudogenes, multiple genomic rearrangements and the deletion of chromosome fragments. However, in the genomes of ancient symbionts, which have coevolved with their host for millions of years, most of the mobile elements and pseudogenes have been eliminated (McCutcheon and Moran, 2012). Genome reduction is also usually connected with other genome changes, such as an extremely fast evolution of sequences, change in matching of codons, bias in nucleotide composition towards AT bases and lower thermal stability in RNAs/proteins. The main cause of these irreversible changes is a small effective population size and the asexuality of the symbionts in a host-restricted environment (Moran, 1996). The population structure of host-restricted symbionts is significantly influenced by the length of the relationship with the host and the decreasing chance for recombination between strains in different hosts. The lack of recombination leads to an accumulation of deleterious mutations and an increased drift, which eliminates non-essential genes from the genome (McCutcheon and Moran, 2012). For example, it is assumed that the loss of genes coding for SOS reparation system in *Wigglesworthia glossinidia* induced an accumulation of point mutations in nucleotide composition. These mutations possibly lead to the replacement of G/C by A/T (Akman et al., 2002). Despite the massive gene losses, particular genes have a key role in biology of the bacteria or the host, and they are usually retained in these genomes, e. g., heat shock protein (chaperon) complexes buffering a lower stability of proteins, DnaK and DnaJ complex, or genes involved in the biosynthesis of vitamins and amino acids (McCutcheon and Moran, 2012).

Over some evolutionary period, the need for supplementation of these compounds can lead to a state where the host is existentially dependent on a rapidly degenerating symbiont. Since the reduction of genome can be considerably fast, many metabolic pathways of symbionts can be disrupted and in extreme cases such a symbiont can balance on the edge of extinction. This phenomenon of symbionts' genome decays is known as a rabbit hole (Bennett & Moran 2015) and it can end up in death of both partners. The trajectory towards extinction can be slowed down by selection affecting either host or symbiont (Wernegreen, 2002), however the selection impact can be insufficient. Is there any way to escape from this detrimental circle? One of the possible mechanisms of coping with the degenerating symbiont is complementing its lost functions with metabolically more efficient bacteria. It can result in a complex of several symbionts cooperating in one host, e.g., the endosymbiotic complex of Auchenorrhyncha (Koga et al., 2013).

An alternative process, and possible compensation of important gene loss is gene acquisition via horizontal gene transfer (HGT), which is assumed to be one of the main driving forces of bacterial evolution (Heuer and Smalla, 2007). HGT can also occur between host and symbiotic genome. For example, in the genome of citrus mealybug *Planococcus citri* the researchers found 22 HGTs from a diverse spectrum of bacteria (Husník et al., 2013). In blood-sucking insects, HGT events are common between microbial inhabitants where diverse ecological types of bacteria in different stages of transition (free-living bacteria, host-associated symbionts, bacterial pathogens, and random environmental contamination) get into close contact, which subsequently generates suitable conditions for HGTs events. The particularly interesting example of such event could be the horizontal transfer of operon coding for biotin synthesis, which contributed to the emergence of a novel nutritional symbioses of bedbugs with *Wolbachia* (Nikoh et al., 2014), rodent lice with *Legionella* (Říhová et al., 2017), *H. marginatum* with *Midichloria* (Buysse et al., 2021), white flies with *Cardinium* (Santos-Garcia et al., 2014) and planthoppers with *Wolbachia* (Ju et al., 2020).

Phylogenetic analyses of streamlined symbiotic genomes

The degenerative processes in the genomes of symbionts often affect the accuracy of phylogenetic analyses. P-symbionts seem to be introducing many artifacts into phylogenetic reconstructions due to the changes in nucleotide frequencies favoring adenine-thymine bases and rapid sequence evolution (Husník et al., 2011). As a result of these changes, symbiotic lineages may experience parallel shifts of nucleotide composition and these convergences lead to the introduction of homoplasies. In the phylogenetic trees, such fast-evolving sequences of P-symbionts appear on long branches and tend to cluster together, which has been described in many studies as the long-branch attraction phenomenon (LBA; Ruano-Rubio and Fares, 2007; Bergsten, 2005).

The traditionally used 16S rDNA marker has been employed as a predominant marker for inferring the phylogenetic position of newly described symbionts. Therefore, many symbiotic lineages are represented only by this gene, which has been shown to be inadequate for inferring a reliable phylogenic position particularly within Enterobacteriaceae, where the majority of symbionts originate (Husník et al., 2011; Naum et al., 2008). The progress of whole-genome sequencing enabled the use of different phylogenetic markers for reconstruction of multigene phylogenetic matrices. However, the phylogenetic position could also be strongly influenced by the method of tree inference used. As was shown by Husník et al. (2011), the LBA could be reduced by introducing non-parametric mixture models like CAT, which count for the site-specific features of protein evolution. Using this model to infer phylogeny of a multigene matrix of Enterobacteriaceae, polyphyly of the symbiotic clusters was displayed and an independent origin of certain symbiotic lineages was inferred by multiple studies (**MS2**; Nishide et al. 2022; Klein et al. 2016; Husník et al. 2011).

Origin of symbiotic bacteria in insects

The diversity of symbiotic bacteria across insects documents many independent origins of these associations. The establishment of symbiotic relationship is only possible with a sufficient flux of free-living bacteria from the environment into insect bodies, and their repeated transformation into symbiotic forms. The empirical studies confirm the presence of various bacterial taxa in guts and other tissues of insects. Among the taxa most frequently found in the symbiotic associations are the genera *Arsenophonus* and *Sodalis* (Santos-Garcia et al., 2017; Nováková et al., 2015). The overrepresentation of these clades in symbiotic relationships could hardly be interpreted as an effect of sampling bias. It is believed that certain bacterial groups are more prone to evolve intimate interactions with eukaryotic host cells. However, there is no solid evidence why these groups are particularly prone to (or capable of) undergoing transformation into a symbiotic lifestyle.

Symbiotic bacteria can have various origins, including environmental bacteria or bacterial pathogens (Sachs et al. 2014; Sachs et al., 2011). In theory, an evolutionary continuum exists from free-living or pathogenic bacteria leading towards facultative associations with host and eventually evolving into heritable obligate symbiosis. Despite diverse evolutionary origins and metabolic capacities, the transition to a mutualistic form often includes convergent processes, such as a reduction of superfluous functions, the loss of virulence factors, and either novel acquisition or preservation of traits advantageous to the host (McCutcheon & Moran, 2012). However, the forces which drive bacteria into such transitions have not yet been fully understood, possibly because of a lack of the symbiotic genera which would manifest different evolutionary stages. Alternatively, a comparative genomic analyses of symbionts with their closely related free-living relatives provide a feasible tool to examine the evolutionary route and adaptive mechanisms accompanying the transition to symbiosis (Zheng et al., 2017; Boscaro et al., 2013). For instance, an opportunistic human infective *Sodalis praecaptivus* serves as a suitable comparison for understanding the evolution of symbiosis of insect-associated lineages of *Sodalis* (Chari et al., 2015).

From a host perspective, microbiome composition is to some extent host-specific and keeps some degree of phylogenetic conservativeness. In general, blood-feeding insects have diverse life strategies that affect their microbiome composition. One of the possible routes of microbiome colonization in blood-feeding animals is through their blood meal (Husník, 2018). The blood-feeding strategy can result in an incidental intake of pathogens which then could colonize the host's tissues such as salivary glands (Strand, 2018), or an acquisition of bacteria from the skin microbiome of the host (Husník, 2018). In Anoplura, many independent symbiotic lineages emerged from diverse bacterial groups which suggests that lice experienced a complex history of symbiont acquisitions, lossess, and replacements throughout their evolution. However, what is the possible source of such bacteria prone to enter symbiosis with lice? We address this question in ***Ongoing research I.***

Microbiomes of blood-feeding animals

Origin of blood-feeding in evolution of organisms

The blood-feeding strategy has multiple independent origins in many groups of organisms reaching from invertebrates (nematodes, annelids, platyhelminths, arthropods) to vertebrates (some parasitic fishes and birds, vampire bats). The most diverse groups of blood-feeding parasites are recruited from arthropods within mites, ticks and insects. The vast majority of blood-feeders (app. 14000 species; Adams, 1999) could be found in insects, including mosquitoes, sand flies, black flies, tabanids, biting midges, tsetse flies, louse flies, bat flies, bed bugs, triatomines, fleas and lice (see below Table 1; Duron & Gottlieb, 2020; Husník, 2018; Budachetri et al., 2014; Weiss & Aksoy, 2011; Adams, 1999). However, all the mentioned groups are greatly variable in dependence on a blood meal during their ontogeny.

Blood is a generous source of proteins and has a high content of salt, however it is believed to be poor in B-vitamins and cofactors (Ribeiro & Arcà, 2009). In general, insects cannot synthesize eight B-vitamins that function as coenzymes in various essential enzymatic reactions, and are therefore needed for the successful development (Douglas, 2017). Several groups of blood-feeding insects do not feed on blood during their whole lifecycle, but they also seek alternative food sources, typically in a larval stage (e.g., mosquitoes). Therefore, these groups do not depend on typical obligate mutualists but they possess a complex microbiomes in their guts. However, these groups do not fall into the scope of our research. In contrast, the insect living on vertebrate blood as an exclusive source of nutrients, (e.g., bed bugs, kissing bugs, lice, hippoboscoid flies; obligate blood-feeders thereafter) are dependent on symbiotic bacteria which compensate the host with B-vitamins and cofactors deficient in blood diet (Nikoh et al., 2014; Kirkness et al., 2010). An overview of the main lineages of symbiotic bacteria found in particular groups of blood-sucking insects is presented in Table 1.

Table 1. Selected blood-feeding insects and the major bacterial members of their microbiome. The groups consisting of obligate blood-feeders are indicated in bold.

Insect host lineage	**P-symbionts**	**S-symbionts or pathogens**
Mosquitoes (Diptera)	Diverse gut symbionts (*Serratia, Asaia*, etc.)	*Wolbachia* spp., *Rickettsia* spp., *Cardinium* spp.
Sand flies (Diptera)	Diverse gut symbionts	*Wolbachia* spp., *Cardinium* spp., *Rickettsia* spp., *Spiroplasma* spp., *Bartonella* spp.
Black flies (Diptera)	Diverse gut symbionts	*Wolbachia* spp., *Francisella tularensis*
Tabanids (Diptera)	Diverse gut symbionts	*Wolbachia* spp., *Spiroplasma* spp. *Rickettsia* spp.
Biting midges (Diptera)	Diverse gut symbionts	*Wolbachia* spp., *Cardinium* sp., *Rickettsia* spp.
Fleas (Siphonaptera)	Diverse gut symbionts	*Yersinia* spp., *Rickettsia* spp., *Cardinium* sp., *Bartonella* spp., *Wolbachia* spp., *Spiroplasma* spp.
Tsetse flies (Diptera)	*Wigglesworthia glossinidia*	*Sodalis glossinidius*, *Wolbachia* sp., *Burkholderia* sp., *Pantoea* sp.
Louse flies (Diptera)	*Arsenophonus* *Sodalis* (and allied)	*Arsenophonus*, *Sodalis* (and allied) *Wolbachia* spp., *Bartonella* spp. *Pectobacterium*-allied
Bat flies (Diptera)	*Arsenophonus* (and allied), *Aschnera chinzeii*	*Wolbachia* spp., *Bartonella* spp.
Kissing bugs (Heteroptera)	Diverse gut symbionts	*Arsenophonus triatominarum* *Wolbachia* spp., *Bartonella* spp.
Bed bugs (Heteroptera)	*Wolbachia* spp.	unidentified Enterobacteriales
Chewing lice (Phthiraptera)	Enterobacteriaceae *Wolbachi*a spp.	-
Sucking lice (Phthiraptera)	*Puchtella* spp. *Arsenophonus*-allied (*Riesia* spp.) *Sodalis*-allied Neisseriaceae-allied *Legionella polyplacis* *Lightella neohaematopini* *Haematopinicola* spp. Enterobacteriaceae	Neisseriaceae-allied unidentified Enterobacteriales *Rickettsia prowazekii* *Rickettsia* sp. *Bartonella quintana* *Borrelia recurrentis* *Wolbachia* spp. *Salmonella typhi*
Rhynchophthirina lice (Phthiraptera)	Enterobacteriaceae	-

*I note that this table is not fully exhaustive, and it shows only the major bacterial symbionts reported to date.

The diversity of symbiotic lineages in obligate blood-feeders

The obligate blood-feeders form an important ecological group of insects which is known to utilize symbiotic bacteria for their metabolic functions. The host-symbiotic systems are either monospecific (where the host maintains relationships with a single or a few microbial species), or more symbionts can work in complex microbial communities. In this section, I would like to present an overview of certain groups of obligate blood-feeders used as model organisms in our laboratory and provide a brief insight into their symbioses.

- Hippoboscoidea (Diptera)

The only obligate blood-feeders in holometabolous insects that maintain obligate intracellular bacteria are Hippoboscoidea (bat flies, louse flies and tsetse flies). This group developed a unique reproduction strategy called adenotrophic viviparity, where larvae are developing inside the female´s body (in the uterus) and receive nutrients through specialized structures called 'milk glands'. For symbionts 'milk glands' also serve as a transport route to the uterus during larvae development (Aksoy et al., 1997). While facultative symbionts could invade many tissues (reproductive organs, bacteriome, midgut, fat body; Nováková et al., 2015; Balmand et al., 2013), the obligate symbionts usually reside in bacteriome and 'milk glands'.

Glossinidae (tsetse flies), a basal and species-poor clade of Hippoboscoidea, harbors an obligate endosymbiont *Wigglesworthia glossinidia* (Aksoy, 1995) and a facultative endosymbiont *Sodalis glossinidius* (Dale and Maudlin, 1999). In contrary, bat flies (Nycteribidae and Streblidae) and louse flies (Hippoboscidae) are species-rich lineages associated mainly with *Arsenophonus*-allied (Duron et al., 2014; Morse et al., 2013; Hosokawa et al., 2012) and *Sodalis*-allied bacteria (Šochová et al., 2017; Chrudimský et al., 2012; Nováková and Hypša, 2007). Apart from these two genera, Hippoboscidae harbor a substantial diversity of facultative symbionts and pathogens. The presence of an *Arsenophonus* and *Sodalis* species representing S-symbionts, numerous *Wolbachia* strains (Speer et al., 2022; Šochová et al., 2017), and several *Bartonella* species (Halos et al., 2004) suggest repeated symbiont acquisition from the environment and their replacements, pointing to a complex and dynamic symbiotic system with many interesting aspects for following research.

- Triatominae (Heteroptera: Reduviidae)

Most of the obligate blood-feeders belong to hemimetabolous insects (Anoplura and some Heteroptera, e.g., Cimicidae and Triatominae). Triatomines are considered obligate blood-feeders usually feeding on reptiles, birds and mammals. Additionally, it has been proved that triatomines can use also alternative feeding strategies as conspecific coprophagy (feeding on feces of triatomine), haemolymphagy (feeding on arthropod haemolymph), and kleptohematophagy (piercing through an abdomen of other triatomine and feeding on its gut content; Schmidt et al., 2019; Durán et al. 2016; Alves et al., 2011; Sandoval et al., 2000). Unlike other groups of obligate blood-feeders, Triatominae lack specialized bacteriome-associated bacteria but maintain a diverse extracellular microbiota in the gut

lumen (Rodríguez-Ruano et al., 2018). The study of triatomine gut microbiota has gained relevance in the past decades due to its possible effect on the host's vectorial competence of *Trypanosoma cruzi* (Guarneri and Schaub, 2021; Müller et al., 2019; Díaz et al., 2016; Azambuja et al., 2004). This research could have a prospective use in the control strategies of triatomine bugs.

In recent studies (Brown et al., 2020; Rodríguez-Ruano et al., 2018) it was shown that the composition of gut microbiome is significantly influenced by triatomine ontogeny, species identity, and the environment. For example, during the ontogenetic development of some species of wild triatomines, microbiomes are characterized by decrease in bacterial diversity, where *Dietzia* is becoming a dominant taxon in older instars of majority of studied triatomine species. However, in other triatomine species, the *Dietzia* is not presented, which is pointing to a species-specific microbiome composition also modulated by specific habitat factors (Brown et al. 2020). Recently, it has been demonstrated, that several species of *Rhodnius* harbor the vertically transmitted *Wolbachia* symbiont which can act as a nutritional symbiont possibly due to the provisioning of biotin (Filée et al., 2022) as shown in other *Wolbachia* nutritional mutualists (Nikoh et al., 2014). It was hypothesized that these *Wolbachia* symbionts may also help to complement the metabolic capacity of the *Rhodococcus rhodnii* gut symbionts (Filée et al., 2022). However, to address the exact symbiotic role of both bacteria, *in-vivo* functional tests with *Rhodnius* cured with *R. rhodnii* and/or *Wolbachia* are required.

- Heteroptera: Cimicidae

In bedbugs, *Wolbachia* was the first symbiont discovered to have established an obligate relationship with *Cimex lectularius* (Hosokawa et al., 2010). The most remarkable feature of this *Wolbachia* strain is the presence of a horizontally acquired operon coding for B7 vitamin (biotin) synthesis (Nikoh et al., 2014). The *Wolbachia* strains were later shown to co-diverge with the bedbug lineages, maintaining symbiotic relationships across several species (Balvín et al., 2018). More recently, the amount of genomic data from diverse bed bugs shed light on the intriguing complexity of their symbiotic systems (unpublished data of our laboratory). Similar to lice, the bed bugs possess a diversity of symbiotic bacteria suggesting a complex process of acquisitions, losses, and replacements underlined by a notable host-symbiont preference (unpublished data of our laboratory).

- Phthiraptera

Within the ectoparasitic order Phthiraptera, endosymbiotic bacteria have been well documented in Anoplura, Rhynchophthirina and also in some Amblycera (Allen et al. 2016; Boyd and Reed, 2012; Hypša and Křížek, 2007). The Rhynchophthirina feed from the pooled blood of their hosts and their symbionts belong to an independent lineage within the family Enterobacteriaceae (***Ongoing research II***, Hypša & Křížek, 2007). Within Amblycera (part of chewing lice), *Menacanthus eurysternus* feeds on blood and maintains symbiosis with the two different strains of *Wolbachia*. One of them possesses an extremely small genome (733,850 bp) and a horizontally acquired capacity for pantothenate synthesis (Mahmood et al., 2023).

Anoplura (sucking lice) are the oldest and the most diverse group of strictly obligate blood-feeders, comprising more than 500 species (Light et al., 2010). Since Ries (1931), who clarified the structure of bacteriomes on tissue sections, bacterial symbiosis in lice has become of interest to many scientists (Boyd et al., 2017; Allen et al., 2016; Fukatsu et al., 2009; Hypša and Křížek, 2007; Sasaki-Fukatsu et al., 2006; Buchner, 1965; Bewig and Schwartz, 1956; Puchta, 1955; Puchta, 1954). Bacteriomes of sucking lice have variable structures and shapes and they are situated in different parts of the louse body. For example, in *Haematopinus apri*, the symbionts are localized in the bacteriocyte-like structures recognized as extracellular cavities of a rounded shape, surrounded by the midgut epithelial cells. Also, in females, these symbionts reside in the "depot bacteriomes" from which the symbiotic bacteria migrate to the ovarial ampullae and infect the developing oocytes (Nishide et al. 2022; Buchner, 1965; Ries, 1931). However, symbionts could also be localized in an unpaired oval bacteriome, so called stomach disc, situated on the ventral side of midgut, as in human louse *Pediculus humanus* (Buchner, 1965).

Only a few percent of louse species have been inspected for a presence of symbiotic bacteria, however even in this fraction, a high diversity of symbiotic lineages has been found (***Ongoing research II***; **MS2**; **MS1**; Nishide et al., 2022; Říhová et al., 2017; Allen et al., 2016; Boyd et al., 2017; Boyd et al., 2016; Boyd et al., 2014; Kirkness et al., 2010; Fukatsu et al., 2009; Allen et al., 2007; Hypša and Křížek, 2007). It suggests first that the real number of lineages could be potentially much higher, and second that these symbionts are possibly of recent origins. Most louse symbionts originate in the order Enterobacteriales (***Ongoing research II***; **MS2**; Nishide et al., 2022; Boyd et al., 2017; Allen et al., 2016; Boyd et al., 2014; Allen et al., 2007; Hypša and Křížek, 2007), but none of them are fixed across multiple lice genera (as we can see in *Buchnera* symbiont of aphids), which suggests high rates of replacements or acquisition/losses in lice. This diversity seems to be quite unique among other obligate blood-feeders for which the data are available. For example, in Hippoboscidae, their symbionts originate in *Arsenophonus* or *Sodalis* genera or bacteria with uncertain position (***Ongoing research II***, Šochová et al., 2017). To address whether and why lice are prone to acquire symbionts from diverse bacterial groups, we extended the knowledge of their symbiont´s diversity and set up comprehensive genome-based research to infer their precise phylogenetic positions and elucidate their metabolic capacities.

Our previous research: Background for this thesis

First, when I came to our laboratory, I started processing louse samples from the local populations of mice collected during the long-term research of my colleagues. My bachelor topic followed up on the detection of a symbiont of the lice *Polyplax serrata* and *Polyplax spinulosa* with a possibly unique phylogenetic position among the well-known pathogens from the order Legionellales (Gammaproteobacteria; Hypša and Křížek, 2007). During my bachelor studies, we sequenced the genome of the symbiont from the *Polyplax serrata* louse and newly implemented many bioinformatic approaches and appliances for the genomic survey. Also, we clarified the phylogenetic position of this symbiont, which clustered on a

long branch among *Legionella* pathogens (Říhová, 2015). When screening the *Polyplax serrata* data for contigs of symbionts, we discovered additional potentially symbiotic bacterium, clustering according to our preliminary analyses within the order Neisseriales (Betaproteobacteria). These results pointed to a more complex symbiotic system. A screening of the population of lice *Polyplax serrata* provided data for the phylogenetic analysis based on 16S rRNA gene, where the sequence of Legionellales symbionts from different localities clustered within a monophyletic group. The results showed that we do not deal with a contaminations, but with the symbiont that is coevolving with the hosts.

I then focused on the basic genome characterization of the two symbionts and their phylogenetic origins. While the Legionellales symbiont clustered directly within the genus *Legionella*, the Neisseriales symbiont was rather distantly related to the genus *Neisseria*. The Neisseriales symbiont was shown to cluster on a basal offshoot of Neisseriaceae close to the bee (and bumble bee) gut symbiont *Snodgrassella alvi*. Using bioinformatic analyses, we assembled the genomes of these two symbionts into several contigs presenting draft genomes. Basic genomic characteristics also confirmed the endosymbiotic nature of these bacteria. When we screened the genome of Legionellales symbiont for horizontally transferred genes, we found biotin operon genes, acquired from Alphaproteobacteria, closely related to the biotin operon found in the *Wolbachia* nutritional symbiont of *Cimex lectularius* (Nikoh et al., 2014). Metabolic pathways for B vitamins were reconstructed for *Legionella* and Neisseriales symbionts and compared for their possible complementarity. Finally, trial experiments for FISH visualization of symbionts in *Polyplax serrata* samples were introduced.

In my diploma thesis I proceeded with a thorough study of these two bacterial symbionts. The genome of Legionellales symbiont was assembled to the circular chromosome using advanced methodical approaches. A phylogenetic analyses based on multigene protein matrix confirmed its phylogenetic position among *Legionella* pathogens and we proposed its name *Candidatus* Legionella polyplacis. The complex study of genome, e. g., detailed genome characteristics and FISH analyses, proved an obligate character of this symbiont and a common coevolutionary history with the host by vertical transfer to the progeny. We also clarified the phylogenetic origin of horizontally transferred biotin operon genes relative to operons of other insect symbionts. These results were published in *Genome Biology and Evolution* journal and were presented as a part of my diploma thesis (Říhová, 2017; Říhová et al., 2017).

The second part of my diploma thesis was focused on a screening of the Neisseriales symbiont bacteria (later named as Neisseriaceae-related symbiont in **MS1**). We confirmed the distribution of the Neisseriaceae-related symbiont in the population of lice in Stružná region using appropriate PCR markers. Moreover, the genome assembly was qualitatively improved by filtering out the chimeric regions. Using a complete genomic set, we reconstructed metabolic pathways for B-vitamins, and we checked for the complementarity of these pathways with *Ca*. Legionella polyplacis (for simplicity later referred as *Legionella polyplacis*). Based on the described research we designed aims for the dissertation thesis.

AIMS AND SCOPE OF THE THESIS

The main aim of this thesis builds on newly obtained experience with the high-throughput techniques (such as amplicon and genome sequencing) and several unique observations that we accumulated during our previous research. In the original plan, the scope of the thesis was focused solely on the genomics of symbiotic bacteria in lice. However, we pointed out that the original aims should serve as a framework, and they would be adapted to the ongoing research and new partial results. Therefore, we also present partial results for several new studies extending the thesis with genomic surveys of the symbiotic bacteria of other obligate blood-feeders and the identification of a possible environmental source of symbiotic bacteria for the studied insects.

The main goal:

To investigate diversities of microbiomes in several louse species, and to sequence genomes of their symbionts with the aim to obtain a wider background for inferring evolutionary features and significance of these symbionts for the louse hosts. I approached this goal in the two following projects:

1. Collecting new samples of lice from local populations (*Polyplax* spp., *Hoplopleura* spp.) for studying the diversity and composition of their microbiomes

Starting hypothesis: Based on Allen et al. (2016) and our basic PCR screening using either gammaproteobacterial or general bacterial primers, no symbiotic bacteria have been detected in lice of genus *Hoplopleura* collected from Arvicolinae mice. However, this is in sharp contrast with generally accepted paradigm that highly specialized obligate blood-feeders as lice usually need P-symbionts to compensate for the lack of essential metabolites. This general presumption led us to consider two different hypotheses. Either we potentially found the system where no symbiont is needed for host survival or the used methods were not sufficiently sensitive to detect the P-symbiont.

Achievements: Using high-throughput amplicon sequencing, we screened several populations of local *Hoplopleura* species (*H. acanthopus* and *H. edentula*) and generated their microbiome profiles. Based on the screening, we found several OTUs representing possible P-symbionts of *Hoplopleura* lice. Particularly surprising in context of my previous study (Říhová, 2017) was presence of a dominant OTU assigned to Neisseriaceae. We sequenced a pooled sample of *H. acanthopus* and assembled the genome of its Neisseriaceae-related symbiont into circular chromosome. In phylogenetic analyses it clusters as a sister taxon to the previously described Neisseriaceae-related symbiont from unrelated louse *Polyplax serrata* (Říhová, 2017). These two symbionts form a new monophyletic lineage of louse symbionts clustering as a most basal offshoot of Neisseriaceae. These interesting findings led us to assemble additional samples from local populations of *Polyplax serrata* lice and to screen their microbiomes using amplicon approaches. The population-wide screening shows

that while Neisseriaceae-related symbiont of *H. acanthopus* was with high consistency present as the most dominant bacterium, the Neisseriaceae-related symbiont of *P. serrata* symbiont was found only in some samples and with much lower abundance. Using whole-genome approaches we sequenced the draft genome of Neisseriaceae-related symbiont of *P. serrata*. Considering the genome characteristics and distribution pattern of both Neisseriaceae-related symbionts we suggest that these symbiont are in an early/intermediate stage of symbiogenesis. Particularly, in *H. acanthopus*, Neisseriaceae-related symbiont seems to be evolving towards P-symbiont. Consistently with these findings, we show the localization of Neisseriaceae-related symbiont of *H. acanthopus* in the putative bacteriocytes localized below louse ovarial ampulla and in the cells of ovariole sheaths. The results of this study were published in Manuscript #1 (**MS1**).

2. Collecting of new samples of other lice species (collected in other countries) for studying the diversity and composition of their microbiomes

Starting hypothesis: Our previous results and other studies pointed out a high diversity of louse symbionts from diverse phylogenetic groups, and absence of a common coevolving symbiont (such as *Buchnera* in aphids). To better understand evolutionary origins and the functional role of symbionts in lice, we needed more data from a wider taxonomic diversity of lice and their symbiotic bacteria. To extend the current dataset, I settled a collaboration with Dr. Kayce C. Bell (Department of Mammalogy, Natural History Museum of Los Angeles County, Los Angeles, CA, USA), who provided metagenomic data for *Neohaematopinus pacificus* chipmunk lice. As it was shown in coevolutionary analyses of chipmunks and their *N. pacificus* lice, the lice form a cluster of separated lineages, where each lineage is specific to one or several related chipmunk species (Bell et al., 2021). The genomic data of the *N. pacificus* lice was used for consequent genomic analyses.

Achievements: Using this data, we assembled multiple genomes of symbionts of *N. pacificus* lice. We showed that these symbionts are closely related to *Puchtella,* a symbiont of louse *Pedicinus baddii* (Boyd et al., 2017). Considering the phylogenetic distances between the two louse species (i.e., *N. pacificus* and *P. baddii*; Light et al., 2010) we suggest that both symbionts established their relationship independently, each with different species. The whole new lineage of these symbionts of *N. pacificus* coevolves with the host and displays features typical for highly reduced P-symbionts. We proposed name *Ca.* Lightella neohaematopini for all included strains from *N. pacificus*. Comparison of metabolic pathways suggests that *L. neohaematopini* most likely supplement host with several B-vitamins. Interestingly, some of those louse samples were shown to harbor bacterium closely related to Neisseriaceae-related louse symbionts previously described in **MS1**. The main characteristics of the draft genome of this Neisseriaceae-related symbiont imply that it represents evolutionary younger and less degenerated S-symbiont. Further characterization of these bacteria extended the known diversity of the louse-associated symbiotic lineages and point at the exceptional dynamic processes of symbionts' acquisitions and replacements in lice. The results of this study were published in Manuscript #2 (**MS2**).

Ongoing research and future prospects:

The two topics presented in this section cover the currently assembled preliminary results as a follow up to the two published studies (**MS1** and **MS2**). Both topics were designed for running grant projects. In respect of this thesis, it extends the two main studies with broader evolutionary context, studying the putative sources and origins of symbiotic bacteria (***Ongoing research I***). In addition, the progress and affordability of advanced sequencing technologies enabled the study of other obligate blood-feeders and their symbiotic systems. The results will be used for a comprehensive comparative study of the metabolic properties among symbionts of obligate blood-feeders (***Ongoing research II***).

- Screening of environmental samples to assess putative source(s) of symbiotic bacteria

Starting hypothesis: Based on previous research, we assumed that the phylogenetic diversity of symbiotic bacteria in sucking lice shows many independent origins of such associations. It could only be possible with multiple acquisitions of free-living bacteria from the environment into lice bodies, and their repeated transformation into symbiotic forms. Anoplura are obligate blood-feeders living permanently in the host fur. Therefore, we hypothesized that potential sources of bacteria prone to enter the symbiotic relationships could be fur microbiome, microbiome of mouse hair follicles/epidermis, pathogenic bacteria of mice or bacterial symbionts/pathogens of other ectoparasites.

Achievements: To test this hypothesis, we designed an experiment where we screened different samples (fur swabs, pieces of mice skin, spleens and ectoparasites) collected from local populations of 3 mouse species (*Apodemus flavicollis*, *Microtus arvalis* and *Clethrionomys glareolus*) using amplicon sequencing. Surprisingly, we found 6 different OTUs assigned to Neisseriaceae (OTU 5 was the most abundant across the samples; see below). Based on the amplicon screening results, we chose one sample containing Neisseriaceae OTU 5 from each studied species of mice. Using whole-genome sequencing we assembled three genome drafts of environmental Neisseriaceae bacteria (eNeisseriaceae bacteria; "e" stands for environmental). In the phylogenetic analyses, all three eNeisseriaceae bacteria cluster as closely related sister taxa on a long monophyletic branch together with the previously described louse symbionts. An early evolutionary split between eNeisseriaceae bacteria and Neisseriaceae-related symbionts is also supported by highly similar characteristics and synteny between each form (environmental vs. symbiotic). The results of this study are presented in ***Ongoing research I***.

- Sequencing, genome assemblies and phylogenetic reconstructions of the new symbionts from obligate blood-feeders – background for the comprehensive comparative studies

Starting hypothesis: While some symbiotic systems in obligate blood-feeders have been well-described, many others remain unexplored. During my PhD studies, we established multiple projects uncovering the diversity of symbiotic bacteria in such insects and raised several questions concerned with their evolutionary origin and functional roles for their host.

The evolutionary events and functional co-dependencies with the host led to a deterioration of genomes of symbiotic bacteria and their convergence towards the preservation of particular metabolic pathways serving to the host for the synthesis of compounds missing in the blood meal (i.e., B-vitamins and cofactors). In the light of accumulated research, we hypothesized that symbiotic bacteria of obligate blood-feeders share some common metabolic patterns (e.g., synthesis of riboflavin). To further investigate these functional and taxonomic convergences, we first needed to extend a dataset of symbiotic bacteria across the other groups of these insects.

Achievements: During my PhD studies, I collected a diverse sample-set of obligate blood -feeders and generated their microbiome profiles. Using this data, we identified the samples with consistent microbial patterns. The selected samples of suitable DNA quality were sent for whole-genome sequencing. The metagenomic data obtained from sequencing were assembled according to a well-established pipeline. As a result, we retrieved twenty new genomes of symbiotic bacteria of obligate blood-feeders. To assess phylogenetic positions of these symbionts, we used genomic data to reconstruct the multigene phylogenetic matrices and then inferred phylogenetic trees. Also, we show the comparisons of the main genome features of these new symbionts with their close relatives. The results of this study are presented in ***Ongoing research II***.

REFERENCES

Adams, T. S. (1999). Hematophagy and hormone release. *Annals of the Entomological Society of America*, *92*(1), 1-13. doi: 10.1093/aesa/92.1.1

Akman, L., Yamashita, A., Watanabe, H., Oshima, K., Shiba, T., Hattori, M., Aksoy, S. (2002). Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nature Genetics*, 32(3), 402-407. doi: 10.1038/ng986

Aksoy, S., Chen, X. A., & Hypša, V. (1997). Phylogeny and potential transmission routes of midgut-associated endosymbionts of tsetse (Diptera: Glossinidae). *Insect Molecular Biology*, *6*(2), 183-190. doi: 10.1111/j.1365-2583.1997.tb00086.x

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

Allen, J. M., Burleigh, J. G., Light, J. E., & Reed, D. L. (2016). Effects of 16S rDNA sampling on estimates of the number of endosymbiont lineages in sucking lice. *PeerJ*, *4*, e2187. doi: 10.7717/peerj.2187

Allen, J. M., Reed, D. L., Perotti, M. A., & Braig, H. R. (2007). Evolutionary relationships of "*Candidatus* Riesia spp.," endosymbiotic Enterobacteriaceae living within hematophagous primate lice. *Applied and Environmental Microbiology*, *73*(5), 1659-1664. doi: 10.1128/AEM.01877-06

Alves, C. L., Araujo, R. N., Gontijo, N. F., & Pereira, M. H. (2011). Importance and physiological effects of hemolymphagy in triatomines (Hemiptera: Reduviidae). *Journal of Medical Entomology*, *48*(2), 372-381. doi: 10.1603/ME10151

Azambuja, P., Feder, D., & Garcia, E. S. (2004). Isolation of *Serratia marcescens* in the midgut of *Rhodnius prolixus*: impact on the establishment of the parasite *Trypanosoma cruzi* in the vector. *Experimental Parasitology*, *107*(1-2), 89-96. doi: 10.1016/j.exppara.2004.04.007

Balvín, O., Roth, S., Talbot, B., & Reinhardt, K. (2018). Co-speciation in bedbug *Wolbachia* parallel the pattern in nematode hosts. *Scientific Reports*, *8*(1), 8797. doi: 10.1038/s41598-018-25545-y

Bansal, R., Mian, M. R., & Michel, A. P. (2014). Microbiome diversity of *Aphis glycines* with extensive superinfection in native and invasive populations. *Environmental Microbiology Reports*, *6*(1), 57-69. doi: 10.1111/1758-2229.12108

Balmand, S., Lohs, C., Aksoy, S., & Heddi, A. (2013). Tissue distribution and transmission routes for the tsetse fly endosymbionts. *Journal of Invertebrate Pathology*, *112*, 116-122. doi: 10.1016/j.jip.2012.04.002

Baumann, P., Baumann, L., Lai, C. Y., Rouhbakhsh, D., Moran, N. A., & Clark, M. A. (1995). Genetics, physiology, and evolutionary relationships of the genus *Buchnera*: intracellular symbionts of aphids. *Annual Review of Microbiology*, *49*(1), 55-94. doi: 10.1146/annurev.mi.49.100195.000415

References

Baumann, P. (2005). Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annual Review of Microbiology, 59,* 155-189. doi: 10.1146/annurev.micro.59.030804.121041

Bell, K. C., Allen, J. M., Johnson, K. P., Demboski, J. R., & Cook, J. A. (2021). Disentangling lousy relationships: Comparative phylogenomics of two sucking louse lineages parasitizing chipmunks. *Molecular Phylogenetics and Evolution, 155,* 106998. doi: 10.1016/j.ympev.2020.106998.

Bennett, G. M., Abbà, S., Kube, M., & Marzachì, C. (2016). Complete genome sequences of the obligate symbionts "*Candidatus* Sulcia muelleri" and "*Ca*. Nasuia deltocephalinicola" from the pestiferous leafhopper *Macrosteles quadripunctulatus* (Hemiptera: Cicadellidae). *Genome Announcements, 4*(1), e01604-15. doi: 10.1128/genomeA.01604-15

Bennett, G. M., & Moran, N. A. (2015). Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. *Proceedings of the National Academy of Sciences of the United States of America, 112*(33), 10169-10176. doi: 10.1073/pnas.1421388112

Bergsten, J. (2005). A review of long-branch attraction. *Cladistics, 21*(2), 163-193. doi: 10.1111/j.1096-0031.2005.00059.x

Bewig, F., & Schwartz, W. (1956). Untersuchungen über die symbiose von tieren mit pilzen und bakterien. *Archives of Microbiology, 24* (2), 174-208. doi: 10.1007/BF00408632

Boyd, B. M., Allen, J. M., Nguyen, N. P., Vachaspati, P., Quicksall, Z. S., Warnow, T., … & Reed, D. L. (2017). Primates, lice and bacteria: speciation and genome evolution in the symbionts of hominid lice. *Molecular Biology and Evolution, 34*(7), 1743-1757. doi: doi.org/10.1093/molbev/msx117

Boyd, B. M., & Reed, D. L. (2012). Taxonomy of lice and their endosymbiotic bacteria in the post-genomic era. *Clinical Microbiology and Infection, 18*(4), 324-331. doi: 10.1111/j.1469-0691.2012.03782.x

Buchner, P. (1965). Endosymbiosis of animals with plant microorganims. *Ecology and Evolution*. London: *Croom Helm* Ltd., 1-28.

Budachetri, K., Browning, R. E., Adamson, S. W., Dowd, S. E., Chao, C. C., Ching, W. M., & Karim, S. (2014). An insight into the microbiome of the *Amblyomma maculatum* (Acari: Ixodidae). *Journal of Medical Entomology, 51*(1), 119-129. doi: 10.1603/ME12223

Buysse, M., Floriano, A. M., Gottlieb, Y., Nardi, T., Comandatore, F., Olivieri, E., ... & Duron, O. (2021). A dual endosymbiosis supports nutritional adaptation to hematophagy in the invasive tick *Hyalomma marginatum*. *eLife, 10,* e72747. doi: 10.7554/eLife.72747

Boscaro, V., Fokin, S. I., Schrallhammer, M., Schweikert, M., & Petroni, G. (2013). Revised systematics of *Holospora*-like bacteria and characterization of "*Candidatus* Gortzia infectiva", a novel macronuclear symbiont of *Paramecium jenningsi*. *Microbial Ecology, 65*, 255-267. doi: 10.1007/s00248-012-0110-2

Boyd, B. M., Allen, J. M., Nguyen, N. P., Vachaspati, P., Quicksall, Z. S., Warnow, T., … & Reed, D. L. (2017). Primates, lice and bacteria: speciation and genome evolution in the

symbionts of hominid lice. *Molecular Biology and Evolution, 34*(7), 1743-1757. doi: doi.org/10.1093/molbev/msx117

Boyd, B. M., Allen, J. M., Koga, R., Fukatsu, T., Sweet, A. D., Johnson, K. P., & Reed, D. L. (2016). Two bacterial genera, *Sodalis* and *Rickettsia*, associated with the seal louse *Proechinophthirus fluctus* (Phthiraptera: Anoplura). *Applied and Environmental Microbiology, 82*(11), 3185-3197. doi: 10.1128/AEM.00282-16

Boyd, B. M., Allen, J. M., de Crécy-Lagard, V., & Reed, D. L. (2014). Genome sequence of *Candidatus* Riesia pediculischaeffi, endosymbiont of chimpanzee lice, and genomic comparison of recently acquired endosymbionts from human and chimpanzee lice. *G3: Genes, Genomes, Genetics, 4*(11), 2189-2195. doi: 10.1534/g3.114.012567

Brown, J. J., Rodríguez-Ruano, S. M., Poosakkannu, A., Batani, G., Schmidt, J. O., Roachell, W., ... & Nováková, E. (2020). Ontogeny, species identity, and environment dominate microbiome dynamics in wild populations of kissing bugs (Triatominae). *Microbiome*, *8*(1), 1-16. doi: 10.1186/s40168-020-00921-x

Brownlie, J. C., & Johnson, K. N. (2009). Symbiont-mediated protection in insect hosts. *Trends in Microbiology*, *17*(8), 348-354. doi: 10.1016/j.tim.2009.05.005

Burton, J. N., Liachko, I., Dunham, M. J., & Shendure, J. (2014). Species-level deconvolution of metagenome assemblies with Hi-C–based contact probability maps. *G3: Genes, Genomes, Genetics, 4*(7), 1339-1346. doi: 10.1534/g3.114.011825

Cahill, M. J., Köser, C. U., Ross, N. E., & Archer, J. A. (2010). Read length and repeat resolution: Exploring prokaryote genomes using next-generation sequencing technologies. *PloS One, 5*(7), e11518. doi: 10.1371/journal.pone.0011518

Chari, A., Oakeson, K. F., Enomoto, S., Jackson, D. G., Fisher, M. A., & Dale, C. (2015). Phenotypic characterization of *Sodalis praecaptivus* sp. nov., a close non-insect-associated member of the *Sodalis*-allied lineage of insect endosymbionts. *International Journal of Systematic and Evolutionary Microbiology*, *65*(5), 1400-1405. doi: 10.1099/ijs.0.000091

Chrudimský, T., Husník, F., Nováková, E., & Hypša, V. (2012). *Candidatus* Sodalis melophagi sp. nov.: Phylogenetically independent comparative model to the tsetse fly symbiont *Sodalis glossinidius*. *PLoS One*, *7*(7), e40354. doi: 10.1371/journal.pone.0040354

Dale, C., & Moran, N. A. (2006). Molecular interactions between bacterial symbionts and their hosts. *Cell*, *126*(3), 453-465. doi: 10.1016/j.cell.2006.07.014

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

Degnan, P. H., & Ochman, H. (2012). Illumina-based analysis of microbial community diversity. *The ISME Journal*, *6*(1), 183-194. doi: 10.1038/ismej.2011.74

Darby, A. C., Chandler, S. M., Welburn, S. C., & Douglas, A. E. (2005). Aphid-symbiotic bacteria cultured in insect cell lines. *Applied and Environmental Microbiology*, *71*, 4833-4839. doi: 10.1128/AEM.71.8.4833-4839.2005

References

Díaz, S., Villavicencio, B., Correia, N., Costa, J., & Haag, K. L. (2016). Triatomine bugs, their microbiota and *Trypanosoma cruzi*: Asymmetric responses of bacteria to an infected blood meal. *Parasites & Vectors*, *9*(1), 1-11. doi: 10.1186/s13071-016-1926-2

Douglas, A. E. (2017). The B vitamin nutrition of insects: The contributions of diet, microbiome and horizontally acquired genes. *Current Opinion in Insect Science*, *23*, 65-69. doi: 10.1016/j.cois.2017.07.012

Douglas, A. E., (1989). Mycetocyte symbiosis in insects. *Biological Reviews*, 64(4), 409-434. doi: 10.1111/j.1469-185X.1989.tb00682.x

Durán, P., Siñani, E., & Depickère, S. (2016). On triatomines, cockroaches and haemolymphagy under laboratory conditions: New discoveries. *Memórias do Instituto Oswaldo Cruz*, *111*, 605-613. doi: 10.1590/0074-02760160027

Duron, O., & Gottlieb, Y. (2020). Convergence of nutritional symbioses in obligate blood feeders. *Trends in Parasitology*, *36*(10), 816-825. doi: 10.1016/j.pt.2020.07.007

Duron, O., Schneppat, U. E., Berthomieu, A., Goodman, S. M., Droz, B., Paupy, C., ... & Tortosa, P. (2014). Origin, acquisition and diversification of heritable bacterial endosymbionts in louse flies and bat flies. *Molecular Ecology*, *23*(8), 2105-2117. doi: 10.1111/mec.12704

Filee, J., Agesilas-Lequeux, K., Lacquehay, L., Berenger, J. M., Dupont, L., Mendonca, V., ... & Harry, M. (2022). *Wolbachia* genomics reveals a potential for a nutrition-based symbiosis in blood-sucking Triatomine bugs. *bioRxiv*, 2022-09. doi: 10.1101/2022.09.06.506778

Fukatsu, T., Hosokawa, T., Koga, R., Nikoh, N., Kato, T., Hayama, … & Tanaka, I. (2009). Intestinal endocellular symbiotic bacterium of the macaque louse *Pedicinus obtusus*: Distinct endosymbiont origins in anthropoid primate lice and the old world monkey louse. *Applied Environmental Microbiology*, *75*(11), 3796-3799. doi: 10.1128/aem.00226-09

Gaio, A. D. O., Gusmão, D. S., Santos, A. V., Berbert-Molina, M. A., Pimenta, P. F., & Lemos, F. J. (2011). Contribution of midgut bacteria to blood digestion and egg production in *Aedes aegypti* (Diptera: Culicidae)(L.). *Parasites & Vectors*, *4*(1), 1-10. doi: 10.1186/1756-3305-4-105

Gauthier, J. P., Outreman, Y., Mieuzet, L., & Simon, J. C. (2015). Bacterial communities associated with host-adapted populations of pea aphids revealed by deep sequencing of 16S ribosomal DNA. *PloS One*, *10*(3), e0120664. doi: 10.1371/journal.pone.0120664

George, S., Pankhurst, L., Hubbard, A., Votintseva, A., Stoesser, N., Sheppard, A. E., ... & Phan, H. T. (2017). Resolving plasmid structures in Enterobacteriaceae using the MinION nanopore sequencer: Assessment of MinION and MinION/Illumina hybrid data assembly approaches. *Microbial Genomics*, *3*(8), e000118. doi: 10.1099/mgen.0.000118

Gerth, M., & Bleidorn, C. (2016). Comparative genomics provides a timeframe for *Wolbachi*a evolution and exposes a recent biotin synthesis operon transfer. *Nature Microbiology*, *2*(3), 1-7. doi: 10.1038/nmicrobiol.2016.241

References

Goldstein, S., Beka, L., Graf, J., & Klassen, J. L. (2019). Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing. *BMC Genomics*, *20*(1), 1-17. doi: 10.1186/s12864-018-5381-7

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, *17*(6), 333-351. doi: 10.1038/nrg.2016.49

Gotoh, T., Noda, H., & Ito, S. (2007). *Cardinium* symbionts cause cytoplasmic incompatibility in spider mites. *Heredity*, *98*(1), 13-20. doi: 10.1038/sj.hdy.6800881

Gray, M. W., & Doolittle, W. F. (1982). Has the endosymbiont hypothesis been proven? *Microbiological Reviews*, *46*(1), 1. doi: 10.1128/mmbr.46.1.1-42.1982

Guarneri, A. A., & Schaub, G. A. (2021). Interaction of triatomines with their bacterial microbiota and trypanosomes. Triatominae - The biology of Chagas disease vectors, *Entomology in Focus*, Cham: Springer International Publishing, 345-386. doi: 10.1007/978-3-030-64548-9_15

Halos, L., Jamal, T., Maillard, R., Girard, B., Guillot, J., Chomel, B., ... & Boulouis, H. J. (2004). Role of Hippoboscidae flies as potential vectors of *Bartonella* spp. infecting wild and domestic ruminants. *Applied and Environmental Microbiology*, *70*(10), 6302-6305. doi: 10.1128/AEM.70.10.6302-6305.2004

Heuer, H., & Smalla, K. (2007). Horizontal gene transfer between bacteria. *Environmental Biosafety Research*, *6*(1-2), 3-13. doi: 10.1051/ebr:2007034

Hosokawa, T., Nikoh, N., Koga, R., Satô, M., Tanahashi, M., Meng, X. Y., & Fukatsu, T. (2012). Reductive genome evolution, host-symbiont co-speciation and uterine transmission of endosymbiotic bacteria in bat flies. *The ISME Journal*, *6*(3), 577-587. 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. *Proceedings of the National Academy of Sciences*, *107*(2), 769-774. doi: 10.1073/pnas.0911476107

Hunter, M. S., Perlman, S. J., & Kelly, S. E. (2003). A bacterial symbiont in the Bacteroidetes induces cytoplasmic incompatibility in the parasitoid wasp *Encarsia pergandiella*. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1529), 2185-2190. doi: 10.1098/rspb.2003.2475

Husník, F. (2018). Host-symbiont-pathogen interactions in blood-feeding parasites: Nutrition, immune cross-talk and gene exchange. *Parasitology*, *145*(10), 1294-1303. doi: 10.1017/S0031182018000574

Husník F., Nikoh, N., Koga, R., Ross, L., Duncan, R. P., Fujie, M., … & McCutcheon, J. P. (2013). Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell*, *153*(7), 1567-1578. doi: 10.1016/j.cell.2013.05.040

Husník, F., Chrudimský, T., & Hypša, V. (2011). Multiple origins of endosymbiosis within the Enterobacteriaceae (γ-Proteobacteria): convergence of complex phylogenetic approaches. *BMC Biology*, *9*(1), 1-18. doi: 10.1186/1741-7007-9-87

References

Hypša, V., & Křížek, J. (2007). Molecular evidence for polyphyletic origin of the primary symbionts of sucking lice (Phthiraptera, Anoplura). *Microbial Ecology, 54*(2), 242-251. doi: 10.1007/s00248-006-9194-x

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

Jing, X., Wong, A. C. N., Chaston, J. M., Colvin, J., McKenzie, C. L., & Douglas, A. E. (2014). The bacterial communities in plant phloem-sap-feeding insects. *Molecular Ecology*, *23*(6), 1433-1444. doi: 10.1111/mec.12637

Jones, R. T., Bressan, A., Greenwell, A. M., & Fierer, N. (2011). Bacterial communities of two parthenogenetic aphid species cocolonizing two host plants across the Hawaiian Islands. *Applied and Environmental Microbiology*, *77*(23), 8345-8349. doi: 10.1128/AEM.05974-11

Ju, J. F., Bing, X. L., Zhao, D. S., Guo, Y., Xi, Z., Hoffmann, A. A., ... & Hong, X. Y. (2020). *Wolbachia* supplement biotin and riboflavin to enhance reproduction in planthoppers. *The ISME Journal*, *14*(3), 676-687. doi: 10.1038/s41396-019-0559-9

Kikuchi, Y., & Fukatsu, T. (2014). Live imaging of symbiosis: Spatiotemporal infection dynamics of a GFP-labelled *Burkholderia* symbiont in the bean bug *Riptortus pedestris*. *Molecular Ecology*, *23*(6), 1445-1456. doi: 10.1111/mec.12479

Kingsford, C., Schatz, M. C., & Pop, M. (2010). Assembly complexity of prokaryotic genomes using short reads. *BMC Bioinformatics*, *11*, 1-11. doi: 10.1186/1471-2105-11-21

Kirkness, E. F., Haas, B. J., Sun, W., Braig, H. R., Perotti, M. A., Clark, J. M., ... & Gerlach, D. (2010). Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proceedings of the National Academy of Sciences*, *107*(27), 12168-12173. doi: 10.1073/pnas.1003379107

Klein, A., Schrader, L., Gil, R., Manzano-Marín, A., Flórez, L., Wheeler, D., ... & Oettler, J. (2016). A novel intracellular mutualistic bacterium in the invasive ant *Cardiocondyla obscurio*r. *The ISME Journal*, *10*(2), 376-388. doi: 10.1038/ismej.2015.119

Koga, R., & Moran, N. A. (2014). Swapping symbionts in spittlebugs: Evolutionary replacement of a reduced genome symbiont. *The ISME Journal*, *8*(6), 1237-1246. doi: 10.1038/ismej.2013.235

Koga, R., Bennett, G. M., Cryan, J. R., & Moran, N. A. (2013). Evolutionary replacement of obligate symbionts in an ancient and diverse insect lineage. *Environmental Microbiology*, *15*(7), 2073-2081. doi: 10.1111/1462-2920.12121

Lemmon, E. M., & Lemmon, A. R. (2013). High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, *44*, 99-121. doi: 10.1146/annurev-ecolsys-110512-135822

Light, J. E., Smith, V. S., Allen, J. M., Durden, L. A., & Reed, D. L. (2010). Evolutionary history of mammalian sucking lice (Phthiraptera: Anoplura). *BMC Evolutionary Biology*, *10*(1), 1-15. doi: 10.1186/1471-2148-10-292

References

Lin, M. Y., Koppers, N., Denton, A., Schlüter, U., & Weber, A. P. (2021). Whole genome sequencing and assembly data of *Moricandia moricandioides* and *M. arvensis*. *Data in Brief*, *35*, 106922. doi: 10.1016/j.dib.2021.106922

Magi, A., Semeraro, R., Mingrino, A., Giusti, B., & D'aurizio, R. (2018). Nanopore sequencing data analysis: State of the art, applications and challenges. *Briefings in Bioinformatics*, *19*(6), 1256-1272. doi: 10.1093/bib/bbx062

Mahmood, S., Nováková, E., Martinů, J., Sychra, O., & Hypša, V. (2023). Supergroup F *Wolbachia* with extremely reduced genome: Transition to obligate insect symbionts. *Microbiome*, *11*(1), 1-16. doi: 10.1186/s40168-023-01462-9

Marbouty, M., Cournac, A., Flot, J. F., Marie-Nelly, H., Mozziconacci, J., & Koszul, R. (2014). Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. *eLife*, *3*, e03318. doi: 10.7554/eLife.03318

Margulis, L. (1970). Origin of eukaryotic cells: Evidence and research implications for a theory of the origin and evolution of microbial, plant and animal cells on the precambrian Earth. *Yale University Press*.

McCutcheon, J. P., & Moran, N. A. (2012). Extreme genome reduction in symbiotic bacteria. *Nature Reviews Microbiology*, 10(1), 13-26. doi: 10.1038/nrmicro2670

McCutcheon, J. P., & Von Dohlen, C. D. (2011). An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Current Biology*, *21*(16), 1366-1372. doi: 10.1016/j.cub.2011.06.051

Meiser, A., Otte, J., Schmitt, I., & Grande, F. D. (2017). Sequencing genomes from mixed DNA samples-evaluating the metagenome skimming approach in lichenized fungi. *Scientific Reports*, *7*(1), 1-13. doi: 10.1038/s41598-017-14576-6

Metzker, M. L. (2009). Sequencing in real time. *Nature Biotechnology*, *27*(2), 150-151. doi: 10.1038/nbt0209-150

Moran, N. A., & Bennett, G. M. (2014). The tiniest tiny genomes. *Annual Review of Microbiology*, *68*, 195-215. doi: 10.1146/annurev-micro-091213-112901

Moran, N. A., McCutcheon, J. P., & Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics*, *42*, 165-190. doi: 10.1146/annurev.genet.41.110306.130119

Moran, N. A. (1996). Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences*, *93*(7), 2873-2878. doi: 10.1073/pnas.93.7.2873

Morse, S. F., Bush, S. E., Patterson, B. D., Dick, C. W., Gruwell, M. E., & Dittmar, K. (2013). Evolution, multiple acquisition, and localization of endosymbionts in bat flies (Diptera: Hippoboscoidea: Streblidae and Nycteribiidae). *Applied and Environmental Microbiology*, *79*(9), 2952-2961. doi: 10.1128/AEM.03814-12

Moter, A., & Göbel, U. B. (2000). Fluorescence in situ hybridization (FISH) for direct visualization of microorganisms. *Journal of Microbiological Methods*, *41*(2), 85-112. doi: 10.1016/S0167-7012(00)00152-4

References

Moya, A., Peretó, J., Gil, R., & Latorre, A. (2008). Learning how to live together: Genomic insights into prokaryote-animal symbioses. *Nature Reviews Genetics*, *9*(3), 218-229. doi: 10.1038/nrg2319

Müller, J. N., Gonçalves, T. C. M., Ricardo-Silva, A. H., Souza, A. C., Santos, F. M., Santos, R., ... & Carbajal-de-la-Fuente, A. L. (2019). Does antennal sensilla pattern of different populations of *Triatoma maculata* (Hemiptera: Reduviidae) reveal phenotypic variability? *Parasites & Vectors*, *12*, 1-6. doi: 10.1186/s13071-019-3856-2

Naum, M., Brown, E. W., & Mason-Gamer, R. J. (2008). Is 16S rDNA a reliable phylogenetic marker to characterize relationships below the family level in the Enterobacteriaceae? *Journal of Molecular Evolution*, *66*, 630-642. doi: 10.1007/s00239-008-9115-3

Nikoh, N., Hosokawa, T., Moriyama, M., Oshima, K., Hattori, M., & Fukatsu, T. (2014). Evolutionary origin of insect-*Wolbachia* nutritional mutualism. *Proceedings of the National Academy of Sciences*, *111*(28), 10257-10262. doi: 10.1073/pnas.1409284111

Nishide, Y., Oguchi, K., Murakami, M., Moriyama, M., Koga, R., & Fukatsu, T. (2022). Endosymbiotic bacteria of the boar louse *Haematopinus apri* (Insecta: Phthiraptera: Anoplura). *Frontiers in Microbiology*, *13*, 2852. doi: 10.3389/fmicb.2022.962252

Nogge, G. (1976). Sterility in tsetse flies (*Glossina morsitans* Westwood) caused by loss of symbionts. *Experientia*, *32*(8), 995-996. doi: 10.1007/BF01933932

Nováková, E., Husník, F., Šochová, E., & Hypša, V. (2015). *Arsenophonus* and *Sodalis* symbionts in louse flies: an analogy to the *Wigglesworthia* and *Sodalis* system in tsetse flies. *Applied and Environmental Microbiology*, *81*(18), 6189-6199. doi: 10.1128/AEM.01487-15

Nováková, E., & Hypša, V. (2007). A new *Sodalis* lineage from bloodsucking fly *Craterina melbae* (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont *Sodalis glossinidius*. *FEMS Microbiology Letters*, *269*(1), 131-135. doi: 10.1111/j.1574-6968.2006.00620.x

Oliver, K. M., Russell, J. A., Moran, N. A., & Hunter, M. S. (2003). Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proceedings of the National Academy of Sciences*, *100*(4), 1803-1807. doi: 10.1073/pnas.0335320100

Otani, S., Mikaelyan, A., Nobre, T., Hansen, L. H., Koné, N. G. A., Sørensen, S. J., ... & Poulsen, M. (2014). Identifying the core microbial community in the gut of fungus-growing termites. *Molecular Ecology*, *23*(18), 4631-4644. doi: 10.1111/mec.12874

Press, M. O., Wiser, A. H., Kronenberg, Z. N., Langford, K. W., Shakya, M., Lo, C. C., ... & Liachko, I. (2017). Hi-C deconvolution of a human gut microbiome yields high-quality draft genomes and reveals plasmid-genome interactions. *biorxiv*, 198713. doi: 10.1101/198713

Puchta, O. (1955). Experimentelle untersuchungen über die bedeutung der symbiose der kleiderlaus *Pediculus vestimenti* burm. *Zeitschrift für Parasitenkunde*, *17*(1). doi: 10.1007/BF00260226

Puchta, O. (1954). Experimentelle untersuchungen über die symbiose der kleiderlaus *Pediculus vestimenti* burm. *Naturwissenschaften*, *41*, 71-72. doi: 10.1007/BF00634192

References

Rhoads, A., & Au, K. F. (2015). PacBio sequencing and its applications. *Genomics, Proteomics & Bioinformatics*, *13*(5), 278-289. doi: 10.1016/j.gpb.2015.08.002

Ries, E. (1931). Die Symbiose der läuse und federlinge. *Zeitschrift für Morphologie und Ökologie der Tiere*, *20*(2-3), 233-367.

Ribeiro, J. M., & Arcà, B. (2009). From sialomes to the sialoverse: An insight into salivary potion of blood-feeding insects. *Advances in Insect Physiology*, *37*, 59-118. doi: 10.1016/S0065-2806(09)37002-2

Říhová, J., Nováková, E., Husník, F., & Hypša, V. (2017). *Legionella* becoming a mutualist: Adaptive processes shaping the genome of symbiont in the louse *Polyplax serrata*. *Genome Biology and Evolution*, *9*(11), 2946-2957. doi: 10.1093/gbe/evx217

Říhová, J. (2017). Evoluce genomu bakteriálních symbiontů vši *Polyplax serrata* (Phthiraptera, Anoplura). Master Thesis.

Říhová, J. (2015). Symbiotické bakterie vší rodu *Polyplax*: Základní fylogenetická a genomická charakterizace. Bachelor thesis.

Rio, R. V., Symula, R. E., Wang, J., Lohs, C., Wu, Y. N., Snyder, A. K., ... & Aksoy, S. (2012). Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: Glossinidae) obligate symbiont *Wigglesworthia*. *mBio*, *3*(1), e00240-11. doi: 10.1128/mBio.00240-11

Ruano-Rubio, V., & Fares, M. A. (2007). Artifactual phylogenies caused by correlated distribution of substitution rates among sites and lineages: The good, the bad, and the ugly. *Systematic Biology*, *56*(1), 68-82. doi: 10.1080/10635150601175578

Rodríguez-Ruano, S. M., Škochová, V., Rego, R. O., Schmidt, J. O., Roachell, W., Hypša, V., & Nováková, E. (2018). Microbiomes of North American Triatominae: the grounds for Chagas disease epidemiology. *Frontiers in Microbiology*, *9*, 1167. doi: 10.3389/fmicb.2018.01167

Sachs, J. L., Skophammer, R. G., Bansal, N., & Stajich, J. E. (2014). Evolutionary origins and diversification of proteobacterial mutualists. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1775), 20132146. doi: 10.1098/rspb.2013.2146

Sachs, J. L., Skophammer, R. G., & Regus, J. U. (2011). Evolutionary transitions in bacterial symbiosis. *Proceedings of the National Academy of Sciences*, *108*(supplement_2), 10800-10807. doi: 10.1073/pnas.1100304108

Sandoval, C. M., Joya, M. I., Gutierez, R., & Angulo, V. M. (2000). Cleptohaematophagy of the Triatomine bug *Belminus herreri*. *Medical and Veterinary Entomology*, *14*(1), 100-101. doi: 10.1046/j.1365-2915.2000.00210.x

Santos-Garcia, D., Rollat-Farnier, P. A., Beitia, F., Zchori-Fein, E., Vavre, F., Mouton, L., ... & Silva, F. J. (2014). The genome of *Cardinium* cBtQ1 provides insights into genome reduction, symbiont motility, and its settlement in *Bemisia tabaci*. *Genome Biology and Evolution*, *6*(4), 1013-1030. doi: 10.1093/gbe/evu077

Sasaki-Fukatsu, K., Koga, R., Nikoh, N., Yoshizawa, K., Kasai, S., Mihara, M., ... & Fukatsu, T. (2006). Symbiotic bacteria associated with stomach discs of human lice. *Applied and Environmental Microbiology*, *72*(11), 7349-7352. doi: 10.1128/AEM.01429-06

References

Schmidt, J. O., Dorn, P. L., & Klotz, S. A. (2019). Second-best is better than nothing: Cockroaches as a viable food source for the kissing bug *Triatoma recurva* (Hemiptera: Reduviidae). *Journal of Medical Entomology*, *56*(3), 651-655. doi: 10.1093/jme/tjy233

Singh, B., Mal, G., Gautam, S. K., Mukesh, M., Singh, B., Mal, G., ... & Mukesh, M. (2019). Insect gut - A treasure of microbes and microbial enzymes. *Advances in Animal Biotechnology*, 51-58. doi: 10.1007/978-3-030-21309-1_5

Šochová, E., Husník, F., Nováková, E., Halajian, A., & Hypša, V. (2017). *Arsenophonus* and *Sodalis* replacements shape evolution of symbiosis in louse flies. *PeerJ*, 5, e4099. doi:-10.7717/peerj.4099

Sović, I., Križanović, K., Skala, K., & Šikić, M. (2016). Evaluation of hybrid and non-hybrid methods for de novo assembly of nanopore reads. *Bioinformatics*, *32*(17), 2582-2589. doi: 10.1093/bioinformatics/btw237

Speer, K. A., Teixeira, T. S. M., Brown, A. M., Perkins, S. L., Dittmar, K., Ingala, M. R., ... & Clare, E. L. (2022). Cascading effects of habitat loss on ectoparasite-associated bacterial microbiomes. *ISME Communications*, *2*(1), 67. doi: 10.1038/s43705-022-00153-0

Stewart, R. D., Auffret, M. D., Warr, A., Wiser, A. H., Press, M. O., Langford, K. W., ... & Watson, M. (2018). Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nature Communications*, *9*(1), 870. doi: 10.1038/s41467-018-03317-6

Strand, M. R. (2018). Composition and functional roles of the gut microbiota in mosquitoes. *Current Opinion in Insect Science*, *28*, 59-65. doi: 10.1016/j.cois.2018.05.008

Toenshoff, E. R., Gruber, D., & Horn, M. (2012). Co-evolution and symbiont replacement shaped the symbiosis between adelgids (Hemiptera: Adelgidae) and their bacterial symbionts. *Environmental Microbiology*, *14*(5), 1284-1295. doi: 10.1111/j.1462-2920.2012.02712.x

Tringe, S. G., & Rubin, E. M. (2005). Metagenomics: DNA sequencing of environmental samples. *Nature Reviews Genetics*, *6*(11), 805-814. doi: 10.1038/nrg1709

White, J. A., Giorgini, M., Strand, M. R., & Pennacchio, F. (2013). Arthropod endosymbiosis and evolution. *Arthropod Biology and Evolution: Molecules, Development, Morphology*, 441-477. doi: 10.1007/978-3-642-36160-9_17

Whiteford, N., Haslam, N., Weber, G., Prügel-Bennett, A., Essex, J. W., Roach, P. L., ... & Neylon, C. (2005). An analysis of the feasibility of short read sequencing. *Nucleic Acids Research*, *33*(19), e171-e171. doi: 10.1093/nar/gni170

Weiss, B., & Aksoy, S. (2011). Microbiome influences on insect host vector competence. *Trends in Parasitology*, *27*(11), 514-522. doi: 10.1016/j.pt.2011.05.001

Wernegreen, J. J. (2002). Genome evolution in bacterial endosymbionts of insects. *Nature Reviews Genetics*, *3*(11), 850-861. doi: 10.1038/nrg931

Zheng, P., Wang, M., Li, C., Sun, X., Wang, X., Sun, Y., & Sun, S. (2017). Insights into deep-sea adaptations and host-symbiont interactions: A comparative transcriptome study on *Bathymodiolus* mussels and their coastal relatives. *Molecular Ecology*, *26*(19), 5133-5148. doi: 10.1111/mec.14160

MANUSCRIPT #1

A new symbiotic lineage related to *Neisseria* and *Snodgrassella* arises from the dynamic and diverse microbiomes in sucking lice

Molecular Ecology (2021), *30*: 2178-2196
https://doi.org/10.1111/mec.15866

Received: 17 November 2020 | Revised: 16 February 2021 | Accepted: 23 February 2021

DOI: 10.1111/mec.15866

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

A new symbiotic lineage related to *Neisseria* and *Snodgrassella* arises from the dynamic and diverse microbiomes in sucking lice

Jana Říhová[1] | Giampiero Batani[1] | Sonia Maria Rodríguez-Ruano[1] | Jana Martinů[1,2] | František Vácha[3] | Eva Nováková[1,2] | Václav Hypša[1,2]

[1]Department of Parasitology, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

[2]Institute of Parasitology, Biology Centre, ASCR, v.v.i, České Budějovice, Czech Republic

[3]Department of Chemistry, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

Correspondence
Václav Hypša, Department of Parasitology, Faculty of Science, University of South Bohemia, Branišovská 1760, 37005 České Budějovice, Czech Republic.
Email: vacatko@prf.jcu.cz

Funding information
Grantová Agentura České Republiky, Grant/Award Number: GA18-07711S

Abstract

The phylogenetic diversity of symbiotic bacteria in sucking lice suggests that lice have a complex history of symbiont acquisition, loss, and replacement throughout their evolution. These processes have resulted in the establishment of different, phylogenetically distant bacteria as obligate mutualists in different louse groups. By combining metagenomics and amplicon screening across several populations of three louse species (members of the genera *Polyplax* and *Hoplopleura*) we describe a novel louse symbiont lineage related to *Neisseria* and *Snodgrassella*, and show its independent origin in the two louse genera. While the genomes of these symbionts are highly similar, their respective distributions and status within lice microbiomes indicate that they have different functions and history. In *Hoplopleura acanthopus*, the Neisseriaceae-related bacterium is a dominant obligate symbiont present across several host populations. In contrast, the *Polyplax* microbiomes are dominated by the obligate symbiont *Legionella polyplacis*, with the Neisseriaceae-related bacterium co-occurring only in some samples and with much lower abundance. The results thus support the view that compared to other exclusively blood feeding insects, Anoplura possess a unique capacity to acquire symbionts from diverse groups of bacteria.

KEYWORDS
amplicon sequencing, genome evolution, lice, microbiome, symbiosis

1 | INTRODUCTION

An increasing number of studies demonstrate ubiquity and high diversity of insect-associated microbiomes (Douglas, 2015; Engel & Moran, 2013; Yun et al., 2014). These microbial communities, composed of various pathogens, commensals and random contaminants, can serve as natural sources of beneficial symbiotic bacteria. In some insects they give rise to highly specialized, maternally-transmitted mutualists called primary symbionts (P-symbionts), which contribute to the host's metabolism (Douglas, 2015). However, depending on richness and dynamics, the microbiomes usually contain several symbiotic bacteria in various evolutionary stages. In their typical form, P-symbionts are readily recognized by several features (because they are indispensable mutualists): they are universally present in all individuals, as a rule inhabiting specialized host's organs called bacteriomes (Baumann, 2005), and their genomes are significantly reduced with a strong AT bias (Moran, 1996). One specific feature of P-symbionts is their cophylogeny with the host (Chen et al., 1999; Dhami et al., 2013; Moran et al., 1993; Sauer et al., 2000). For example, two of the most studied P-symbionts, *Buchnera* in aphids and *Wigglesworthia* in tsetse flies, were acquired at the beginning of their hosts' diversification and strictly mirror their entire phylogeny (Chen et al., 1999; Moran et al., 1993). Other P-symbionts are restricted to some of the host's lineages, indicating that they are either recently

acquired symbionts or remnants of an ancient symbiont lost in some of the host lineages (Bennett & Moran, 2013).

In some insects, several different P-symbionts may coexist and/or can be accompanied by various secondary symbionts (S-symbionts). The latter usually possess less modified genomes than typical P-symbionts, retain more free-living-like characteristics, and some are supposed to be the intermediate stages of evolution towards obligate symbionts. The P-symbiont *Wigglesworthia glossinidia* represents a typical example of this as it is often accompanied by the S-symbionts *Sodalis glossinidius* and *Wolbachia* (Aksoy, 2000). The complexity of symbiotic associations is obviously due to an ongoing process of symbiont acquisition/loss/replacement, which is well known from several bacteria-insect models and has a well-developed theoretical background (Bennett & Moran, 2015). The theory postulates that after a certain amount of co-evolutionary time, the symbiotic bacterium becomes too degenerated and functionally inadequate, and it has to be replaced (or accompanied) by another symbiont (McCutcheon et al., 2019). While it would be interesting to see how microbiome diversity and dynamics relate to the complexity of symbiosis in different insect groups, there is very little information available today. The majority of studies on insects and their P- and S-symbionts rely on metagenomic information and phylogenetic reconstructions, probably missing a substantial part of microbiome diversity. The introduction of amplicon approaches recently demonstrated that this method can significantly improve our insight into microbiome composition, even in extensively studied model systems (Doudoumis et al., 2017; Gauthier et al., 2015; Manzano-Marín et al., 2017; Meseguer et al., 2017).

With more than 500 spp., the suborder Anoplura is the most ancient and diversified group of insects feeding exclusively on vertebrate blood in all life stages (Light et al., 2010). Accordingly, all investigated sucking lice possess bacteriomes inhabited by obligatory P-symbionts, which are generally supposed to provide the hosts with compounds missing in their blood diets, most typically B vitamins (Allen et al., 2016; Boyd et al., 2014, 2016; Fukatsu et al., 2009; Hypša & Křížek, 2007). This distinguishes Anoplura from the two larger suborders, the chewing lice Amblycera and Ischnocera, which feed on hair or feathers of their hosts. Information on chewing lice symbiosis is scarce and fragmented. Typical bacteriomes seem to be absent in all Amblycera and developed only in some groups of Ischnocera (Perotti et al., 2008). Even less is known on the small suborder Rhynchophthirina. Symbionts of these hematophagous lice belong to an independent lineage within the family Enterobacteriaceae (Hypša & Křížek, 2007). Molecular analyses indicate that Anoplura host high diversity of symbiotic bacteria. Depending on interpretation, the 16S rRNA gene-based phylogenies for the available taxa suggest 5–6 independent symbiotic lineages of Gammaproteobacteria. However, none of them is a dominant louse symbiont broadly distributed across the order (e.g., like *Buchnera* in aphids). The distribution of these symbionts among louse groups suggests a relatively recent origin of each symbiotic lineage and hence a high rate of acquisition/loss/replacement processes. Moreover, compared to the extensively screened phytophagous groups, only a small fraction of sucking lice diversity (approximately 5%) has been investigated. The actual number of symbiotic lineages is therefore likely to be much higher. This diversity of P-symbionts makes Anoplura potentially unique among the obligate blood-feeding insects. For example, all Glossinidae possess *Wigglesworthia* as a universal obligate symbiont (Chen et al., 1999; Moran et al., 1993). While this may be attributed to their low taxonomic diversity (single genus *Glossina*), the family Hippoboscidae with approx. 21 genera (Dick, 2006) provides a more appropriate comparison. This group has experienced several symbiont replacements but all of their obligate symbionts seem to originate from two bacterial lineages, *Arsenophonus* and *Sodalis* (Šochová et al., 2017). In another related dipteran family Nycteribiidae, all obligate symbionts form a single monophyletic clade related to *Arsenophonus* (Hosokawa et al., 2012). Unfortunately, no such comparison is possible with other exclusively blood feeding groups (i.e., Triatominae, Polyctenidae, Cimicidae) due to the lack of relevant information. To seriously address whether and why Anoplura are able to acquire their obligate symbionts from an unusually broad range of bacteria, we first need more complete knowledge on the overall diversity and genomic/metabolic capacity of their symbionts.

Of the currently known lineages of louse symbionts, genomic data are only available for four; three of them show clear signatures of P-symbionts: *Riesia* spp. (louse genera *Pediculus* and *Phthirus* from hominid hosts), *Puchtella pedicinophila* (*Pedicinus* from red colobus monkey and macaques), and *Legionella polyplacis* (*Polyplax serrata* from field mice *Apodemus* spp.; Table 1). Correspondingly, each of these lineages has been found in two to four related host species, as a result of cophylogenetic processes. The fourth lineage, the *Sodalis*-like symbiont from *Proechinophthirus fluctus*, possesses a significantly larger genome exceeding 2 Mbp, and GC content 50%, which the authors interpret as possible evidence of recent replacement of a more ancient and now extinct endosymbiont (Boyd et al., 2016). The distribution of known symbionts in sucking lice thus indicates that this insect group has been undergoing particularly dynamic acquisition, loss, and replacement processes, resulting in high diversity of their symbiotic bacteria.

In this study, we explore the unique diversity of symbiosis in Anoplura by characterizing symbionts from two louse genera, *Hoplopleura* and *Polyplax*. We combine metagenomic and amplicon analyses to address two specific questions. First, we screen diverse populations of *P. serrata* to determine, if the previously described *Legionella polyplacis* (Říhová et al., 2017) is a fixed obligate symbiont. We also use amplicon screening of *Polyplax serrata* and *Hoplopleura* spp. to assess microbiome diversity as a potential source of mutualistic symbionts. Based on the genealogy-dependent diversity of the microbiome profiles, we suggest rapid microbiome changes at the louse population level, reflecting the generally dynamic processes of symbiont acquisition, loss, and replacement in these blood sucking insects. Second, during the screening, we identified a new symbiotic lineage related to the betaproteobacterial genera *Neisseria* and *Snodgrassella* (the latter being a symbiont of bees) in both louse genera. We use a comparative genomic approach to show that these bacteria established their symbiotic relationships independently

TABLE 1 Comparison of the main genomic characteristics

Bacterium - louse host(s)	Genome size (Mb)	GC%	No. of protein coding genes	Source
"Neisseriaceae" – Hoplopleura acanthopus	1.6	33.4	1421	This study
"Neisseriaceae" – Polyplax serrata	1.8	34	1789	This study
Legionella polyplacis – Polyplax serrata + P. spinulosa	0.5	23	473	CP021497.1
Riesia spp. – *Pediculus* + *Phthirus* (4 spp.)	0.5–0.6	25–28.5	476–557	Boyd et al., (2017)
Puchtella pedicinophila – Pedicinus obtusus	0.6	24.2	564	Boyd et al., (2017)_
Sodalis – Proechinophthirus fluctus	2.2	50	1287[a]	Boyd et al., (2016)
Snodgrassella alvi	2.5	41.3	2125	CP007446.1
Neisseria meningitidis	2.2	51.7	2118	NC_003112.2
Neisseriaceae (722 genomes)	1.7–4.9 (avg. 2.2)	41.4–68.4 (avg. 51.7)	1795–4433 (avg. 2245)	JGI database
Snodgrassella (65 genomes)	1.3–3 (avg. 2.4)	37.8–47.8 (avg. 41.8)	1462–5539 (avg. 2444)	JGI database

Louse symbionts are highlighted by a grey background.

[a]Number of genes annotated in the shotgun library avaiable in GenBank under the accession numbers LECR01000001-92.

with the two louse genera, and for *Hoplopleura* we show their intracellular localization in the host's bacteriocytes.

2 | MATERIALS AND METHODS

2.1 | Amplicon library preparation and sequencing

Samples of 190 *Polyplax serrata*, 14 *Hoplopleura acanthopus*, and 2 *Hoplopleura edentula* were collected from different rodent species and populations between 2014–2018. DNA templates, extracted from each individual using the QIAamp DNA Micro Kit (Qiagen), were amplified in two independent, multiplexed 16S rRNA gene libraries. The first library for *Polyplax serrata* samples was constructed according to EMP protocols (http://www.earthmicrobiome.org/protocols-and-standards/16s/) with 515F/806R primers (Apprill et al., 2015; Parada et al.,2016) producing 300–350 bp amplicons of the V4 hypervariable region. Altogether 16 negative controls were used to examine the extraction and amplification procedures, i.e., no template extractions and PCR water templates. The library was sequenced in a single MiSeq run using v2 chemistry with 2 × 150 bp output. The second library for *Hoplopleura* species was constructed using a double barcoding strategy with 515F and 926R primers (Parada et al., 2016; Walters et al., 2016) yielding longer amplicons of 450 bp. A total of 16 *Hoplopleura* samples were part of a pooled library of 432 samples containing eight negative extraction and amplification controls as well as six positive controls. The positive controls, comprising three samples of commercially available mock communities with equal composition of 10 bacterial species and three samples with a staggered profile (ATCC Microbiome Standards), were used to confirm the barcoding output and evaluate any amplification bias and the depth of the sequencing. The data from the second library were generated in a single MiSeq run using v3 chemistry with 2 × 300 bp output. Complete metadata including rodent host species, sampling locality, GPS coordinates and barcode assignments for analysed samples are available in Data S1.

2.2 | Processing and analyses of 16S rRNA gene amplicons

The amplicons, originating from two different library designs, were processed as two independent data sets. The raw data were quality checked in FastQC (Andrews, 2010) and trimmed using USEARCH v9.0.1001 (Edgar, 2013). The reads were further processed into OTUs (operational taxonomic units) with an in-house workflow implementing USEARCH v9.0.1001 scripts as described previously (Brown et al., 2020; Rodríguez-Ruano et al., 2020). Demultiplexed, merged, trimmed and quality filtered reads were clustered at 100% identity providing a representative set for de novo OTU picking. OTUs were defined using the USEARCH global alignment option at 97% identity. Since the RDP taxonomy classifier (Wang et al., 2007) failed to classify *Arsenophonus* OTUs to the genus level, taxonomy was assigned to the representative sequences using the best BLAST hits (Camacho et al., 2009) against the SILVA 132 database (Quast et al., 2013). Results of both approaches are available in Data S1.

Chloroplast and mitochondrial OTUs, OTUs of extremely low abundance (as recommended by Bokulich et al., 2013), and singletons were filtered out from the final OTU table (Data S1) using QIIME 1.9 (Caporaso et al., 2010). Amplicon data (demultiplexed, quality filtered and merged) are available at https://www.ebi.ac.uk/ena/data/view/PRJEB35541.

Since the negative controls for the *Polyplax* data set contained a considerable number of bacterial reads, 9299 on average compared to 161 retrieved for negative controls in Hoplopleura library, the OTU tables were filtered for potential contaminants. These were defined as OTUs comprising more than 1% of reads in any negative control found in more than one fourth of the controls for both

Polyplax and *Hoplopleura* libraries. Details on negative control read counts and identification of potential contaminants are provided in Data S1. Since the primary characterization of bacterial diversity in this study centres on symbiotic taxa, only dominant taxa were considered, i.e., OTUs comprising 10 or more percent of the reads within a particular sample. Under the assumption of high symbiont prevalence within the host populations, our final selection of taxa includes OTUs that occur in more than five individuals across the analysed *Polyplax* samples, or in at least two *Hoplopleura* samples. Considering the apparent nonrandom distribution of the dominant taxa among the *Polyplax* populations, we specifically tested the effects of *Polyplax* genealogy and geographic origin (see Host phylogenetic background) on the structure of complete microbiomes. We calculated community quantitative measures (Richness and Shannon index) and ordination analyses (nonmetric multidimensional scaling, NMDS; based on Bray-Curtis dissimilarities) for the complete microbiomes using the normalised decontaminated data set. Rarefaction, performed at 3000 sequences per sample, and all statistical analyses were executed in R with the "vegan" package and adonis function (https://cran.r-project.org; https://github.com/vegandevs/vegan).

2.3 | DNA template preparation for metagenomics

Polyplax serrata lice (n = 25) were collected from six specimens of yellow-necked wood mice (*Apodemus flavicollis*) captured in the Czech Republic (Struzna) and Germany (Baiersbronn) in 2011. *Hoplopleura acanthopus* lice (n = 40) were obtained from a single heavily infested specimen of a common vole (*Microtus arvalis*) trapped in the Czech Republic (Hlinsko) in 2014. All samples were stored in 96% ethanol at −20°C. Total DNA was extracted from whole louse abdomens using a QIAamp DNA Micro Kit (Qiagen) its quality assessed by gel electrophoresis, and concentration measured with a Qubit High sensitivity kit.

2.4 | Genome sequencing and assembly

We sequenced the *Polyplax* lice pooled sample on one lane of Illumina HiSeq2000 (GeneCore) using 2 × 100 paired-end reads. Read quality was checked using FastQC (Andrews, 2010) and quality trimming was performed using the BBtools package (https://jgi.doe.gov/data-and-tools/bbtools). The resulting data set contained 309,892,186 reads. We used SPAdes assembler 3.10 (Bankevich et al., 2012) to build the assembly, implementing careful options and enabling mismatch corrections. To check for bacterial plasmid(s) we submitted the complete assembly (124,985 contigs) to PlasmidFinder (Carattoli et al., 2014) with sensitivity set to three different thresholds (95%, 85%, and 60%). We identified bacterial contigs by blasting *Snodgrassella alvi* wkB2 genome against the assembly using custom BLAST in the program Geneious (Kearse et al., 2012). This procedure retrieved 39 contigs which were putatively assigned to Neisseriales. For 32 contigs their Neisseriales origin was further confirmed by BLAST analyses of individual genes as specified below. Since the contig carrying 5S and 23S rRNA genes was not reliably assembled, it is not included in the final assembly. The remaining six contigs were of non-Neisseriales origin and were removed from the assembly.

To sequence the *Hoplopleura acanthopus* metagenome, we employed Illumina MiSeq (GeneCore) and Oxford-Nanopore (University of Urbana, Illinois, USA) technology. We constructed the Illumina library from the total DNA of 35 individuals and sequenced it in four runs of Illumina MiSeq using v2 500 cycle chemistry. We used the same procedure for quality checking and filtering as described for the *Polyplax* data set. The resulting number of reads was 34,406,078. We used high molecular weight DNA from five *H. acanthopus* as a template for Oxford-Nanopore sequencing on GridIONx5. The total number of reads was 1,653,194. The quality of Nanopore reads was checked using NanoPack tools (De Coster et al., 2018) and quality filtering was performed using Filtlong (https://github.com/rrwick/Filtlong).

To assemble the *H. acanthopus* metagenome we employed a hybrid approach combining the Illumina and Nanopore data. We used two assemblers, Flye (Kolmogorov et al., 2019) and Canu (Koren et al., 2017), to generate contigs from Nanopore reads. While the Flye assembly resulted in 724 contigs, the Canu assembler generated 2762 contigs. The Nanopore filtered reads were mapped back onto both assemblies using Minimap2 (Li, 2018). To polish the contigs subsets we used consensus calling in Racon (Vaser et al., 2017) followed by two iterations of Medaka polish (https://github.com/nanoporetech/medaka). To obtain optimum sequence correctness, the resulting contigs of the two assemblers were polished with Illumina trimmed reads using Minimap2 alignment and Racon contigs consensus polish. The corrected Flye and Canu assemblies consisted of 702 and 2721 contigs, respectively. We identified three bacterial contigs using the same BLAST procedure as described above for *P. serrata* and assembled them into a single linear sequence (with more than 500 bp congruent overlaps) using the De novo Assembly tool in Geneious. The genome was completed and closed with an additional 1643 bp contig retrieved from the corrected Canu assembly into the 1,607,498 bp long circular genome. For both symbionts, completeness of their genomes was assessed in BUSCO v4.0.6 (Simão et al., 2015; Data S2), and distributions of their contigs within the complete metagenomic assemblies was visualized in Blobtools (Laetsch & Blaxter, 2017). The table of taxonomic annotation was prepared in two steps. First, a BLAST search (with discontinuous megablast algorithm) was done on the complete assembly. Since this step only recognized 20 of the 32 candidate symbiont contigs in the metagenomic assembly of *P. serrata* (probably due to the aberrant nature of the symbiont sequences and high degree of HGT), we completed the taxonomic assignment with a more precise method based on BLASTX and multihit analysis performed on individual genes rather than complete contigs (Data S2).

2.5 | Genome annotation

We annotated genomes of both Neisseriaceae-related symbionts using RAST (Aziz et al., 2008) and deposited them in GenBank under

the accession numbers CP046107 (closed genome of the symbiont from *H. acanthopus*) and WNLJ00000000 (draft genome of *P. serrata* symbiont in 32 contigs).

To assess the phylogenetic origins of individual genes, we first blasted the complete set of protein-coding genes against the nonredundant (nr) protein database using the BLASTP algorithm (Altschul et al., 1990) set to retrieve one hit. The genes which returned members of Neisseriales were assigned to a category "Neisseriales" (Data S2) and considered to be inherited from the Neisseriales ancestor. The rest of the genes were blasted again with BLAST parameters set to 10 hits. Based on the results, the genes were assigned to the following categories: "Mixed" if the hits contained any Neisseriales together with other bacterial groups, "Putative HGT" if the hits did not contain any of the Neisseriales (or even any of Betaproteobacteria; designated by bold italics ***Putative HGT*** in the Data S2), "E" if the hits were eukaryotic, and "No hit" if no hit was obtained. Two bacteria were removed from the nr database prior to this analysis, *Francisella* sp. (accession number GCA_003248485.1) and *Haemophilus parainfluenzae* (GCA_003240835.1). During our preliminary analysis, genes of these two bacteria formed a substantial part of the best hits. Upon closer inspection, we found these two bacteria misclassified (most probably being Neisseriaceae or perhaps chimeras obtained from environmental metagenomic samples). The results of BLASTP analysis also served for identification of possible pseudogenes: following the criterion used by Waterworth et al., (2020), we screened for genes (except for those in the category of hypothetical proteins) which were truncated more than 20% when compared to their closest BLAST hits. Since determination of potential pseudogenes is an ambiguous procedure dependent on the method and criteria used, we complemented this simple straightforward approach with additional analysis performed by a new bioinformatic tool Pseudofinder, which utilizes more complex criteria (Syberg-Olsen et al., 2020).

2.6 | Genome comparison

Three different comparisons were made for the genomes of the two Neisseriaceae-related symbionts. First, the sets of shared and unique genes were obtained by comparing annotation lists of the two symbionts. We did not have a complete closed genome for the *P. serrata* symbiont, so we specifically tested the absence of the genes identified as unique for *H. acanthopus* symbiont. This was done by mapping *P. serrata* metagenomic reads on the set of *H. acanthopus* unique genes using the BBMap tool (Bushnell, 2014). Second, genome synteny was analysed using ProgressiveMauve (Darling et al., 2010). The complete closed genome of the Neisseriaceae-related symbiont from *H. acanthopus* was aligned to the 32 contigs of the Neisseriaceae-related symbiont from *P. serrata*, yielding 34 locally collinear blocks. The resulting synteny regions are explored in more detail in Data S2. Third, using the KEGG database (Kanehisa et al., 2016), we reconstructed the main features of the metabolic capacity of the Neisseriaceae-related symbiont from *Hoplopleura* and made comparisons with several additional bacteria. The set of compared bacteria included the Neisseriaceae-related symbiont of *Polyplax*, four other louse symbionts with complete genomes available (see Table1), *Wigglesworthia glossinidia* (accession number NC_004344.2), *Snodgrassella alvi* (accession number SAMN03081540), and *Neisseria meningitidis* (accession number NC_003112.2).

2.7 | Assessing metabolic capacity

Based solely on the genome content, it is not possible to determine the actual capacity of a symbiont to produce a particular metabolite (e.g., vitamin) and provide it to the host. For example, enzymes for complete biosynthetic pathways are not required if the symbiont obtains some of the intermediate metabolites from the host. In some systems this leads to host-symbiont cooperation in synthesizing the product (Husník et al., 2013). We therefore focused on comparing the potential metabolic capacities (i.e., the completeness of metabolic pathways) in different bacteria, rather than assessing their actual metabolic roles. To compare the encoded metabolic capacities, we established four categories of completeness for the main metabolic pathways. The first "functional" category includes pathways where all genes of a particular biosynthetic cascade are present, and the substrate is either glucose or another common metabolic intermediate, e.g. essential amino acid. The second category comprises of pathways with a single missing gene, considered as "likely functional". In this category, the same enzyme is usually absent in most of the studied symbionts (for example phosphatases [3.1.3.1] or [3.1.3.104] in folate or riboflavin biosynthetic pathways, respectively, or dehydrogenase [1.1.1.95] in the serine pathway), indicating that its function is probably substituted by an alternative molecule. The third category contains pathways with two or more missing genes, considered as "degraded" and nonfunctional. For two pairs of amino acids (Glutamine/Glutamate, Serine/Glycine) we established a fourth category of "cyclic pathways". Most of the symbionts have the capacity for interconverting between these amino acids but cannot synthesize them from glucose.

2.8 | Screening of louse SRA databases

We screened 23 samples (representing 17 louse species of 11 genera; Data S3) of louse metagenomic data available in the SRA database (Leinonen et al., 2011) for the presence of Neisseriaceae related symbionts. For each sample we performed BLASTN with the V4 hypervariable region of the *H. acanthopus* symbiont (i.e., the same region as utilized in the amplicon analysis) used as a query sequence. The mapped reads retrieved from the SRA databases were used as queries in a subsequent BLAST analysis with the aim to obtain their taxonomic assignment. Numbers of the mapped reads for each SRA library and their taxonomic assignments are provided in the Data S3.

2.9 | Host phylogenetic background

We used 379 bp long sequences of the COI gene (amplified with L6625 and H7005 (Hafner et al., 1994) primers) to determine the phylogenetic background of 190 *Polyplax serrata* samples. For phylogenetic reconstruction of *Hoplopleura* host species we amplified a 968 bp long region of the COI gene using LCO1490 (Folmer et al., 1994) and H7005 (Hafner et al., 1994) primers. PCR products were enzymatically purified and sent for Sanger sequencing. All sequences are available in GenBank under the accession numbers provided in Data S4. The matrices were aligned using E-INS-i algorithm of MAFFT v7.450 (Katoh et al., 2002) in Geneious software. Ambiguously aligned positions and divergent blocks were discarded using GBLOCKS v. 091b (Castresana, 2000). Phylogenetic relationships were reconstructed by Bayesian inference with the GTR + I + G best-fit model selected according to a corrected Akaike information criterion using jModelTest2 v2.1.10 (Darriba et al., 2012; Guindon & Gascuel, 2003). *Polyplax spinulosa* was used as the outgroup for the *Polyplax serrata* data set and the *Polyplax serrata* sequence was used as the outgroup for the *Hoplopleura* spp. data set. Bayesian analyses conducted in MrBayes v3.2.4 (Ronquist et al., 2012) consisted of two parallel Markov chain Monte Carlo simulations with four chains run for 10,000,000 generations and sampling frequency of 1000 generations. The convergence of parameter estimates and their ESS values was checked in software TRACER v1.7 (Rambaut et al., 2018).

2.10 | Phylogenetic origin of the symbionts

Phylogenetic analysis of the Neisseriaceae-related symbionts was performed on two different matrices, the concatenated amino-acid "61-gene matrix" and the nucleotide "16S matrix". To construct the "61-gene matrix", we identified a set of 61 orthologues using Orthofinder (Emms & Kelly, 2019). For the analysis, we only retained the orthologues with reliably supported origins within Neisseriales (i.e., the orthologues from the BLAST category "Neisseriales"; details on the assignment to the categories are provided above in the section Genome annotation) and were present in both genomes, because the Neisseriaceae-related symbionts possessed a high proportion of genes potentially acquired by HGT. Two Betaproteobacteria of the order Burkholderiales, *Burkholderia cepacia* and *Acidovorax* sp. KKS102, one gammaproteobacterium, *Legionella pneumophila* subsp. *pneumophila* str. Philadelphia 1, and one alphaproteobacterium, *Rhizobium leguminosarum*, were used as the outgroups (Data S4). For each gene, the sequences were aligned in MAFFT v7.450 using the E-INS-i setting. Ambiguously aligned positions and divergent blocks were discarded using GBLOCKS v. 091b. The best fitting models, TN93 + G + I for the "16S matrix" and LG + G + I + F for the "61-gene matrix", were determined by Akaike information criterion (AIC) using smart model selection of PhyML (Lefort et al., 2017). Maximum-likelihood phylogenetic reconstructions were performed using online PhyML server v3.0 (Guindon et al., 2010) with 100 bootstrap replicates for each single-gene alignment and also for the concatenated "61-gene matrix". Bayesian inference of the "61-gene matrix" was conducted in MrBayes v3.2.5 using the LG + G + I + F evolutionary model. Four chains were run for 20,000,000 generations with sampling frequency set to 1000 generations. Convergence was checked in TRACER v1.6.0.

The "16S matrix" was designed with the aim to obtain a wider phylogenetic context by including the bacteria for which the 16S rRNA gene sequence is the only available marker. The 16S rRNA gene sequences were retrieved by BLASTN from GenBank (Data S4). Two Betaproteobacteria, *Taylorella equigenitalis* str. 09-09 and *Advenella kashmirensis* str. cv4, and one alphaproteobacterium, *Rhizobium capsici* str. IMCC34666, were used as outgroups. The matrix was prepared with the same procedure as the "61-gene matrix" and analyzed by maximum likelihood (ML) and Bayesian inference (BI). The evolutionary models best fitting to the data set were selected according to AIC using jModelTest2 v2.1.10. ML analysis and 100 bootstrap replicates were performed in PhyML using selected TN93 + G + I evolutionary model. BI analysis was performed in MrBayes v3.2.5, using the GTR + G + I substitution model running four chains for 10,000,000 generations and checked for convergence as was previously described.

2.11 | Localization analysis of symbionts

We fixed *H. acanthopus* tissues by incubation in 4% paraformaldehyde solution at 4°C for 39 h. Subsequently, insects were transferred and kept at 4°C in Carnoy's solution for 27 h, 2% hydrogen peroxide ethanol solution for 3 days and 6% hydrogen peroxide ethanol solution for 10 days to quench tissue autofluorescence. We then washed specimens with 400 µl of hybridization buffer (900 mM NaCl, 20 mM Tris-HCl pH 7.4, 0.01% sodium dodecyl sulphate, SDS, and 30% formamide) at 46°C for 10 min, prehybridized with 200 µl of hybridization buffer at 46°C for 1 h, and hybridized at 46°C for 3 h with 400 µl of hybridization buffer containing probes. All fluorescent probes were obtained from Sigma-Aldrich and used at a concentration of 2.5 ng μl^{-1} for Cyanine 5 (Cy5)-labelled EUB338 (5′-GCTGCCTCCCGTAGGAGT-3′), targeting all bacteria (Amann et al., 1990), and 6-carboxyfluorescein (6-Fam)-labelled beta-572 (5′-TTAACCGTCTGCGCTCGCTT-3′), targeting the family Neisseriaceae (Martinson et al., 2012). We pre-evaluated the required stringency of the hybridization conditions in silico using mathFISH (Yilmaz et al., 2011). Following hybridization, we washed specimens twice with 400 µl of prewarmed wash buffer (20 mM Tris-HCl, 5 mM EDTA, 0.01% SDS and 112 mM NaCl) at 48°C for 10 min. All incubations were carried out with ongoing shaking at 300 rpm. We then placed lice on microscope slides, incubated them in ~50 µl of 4′,6-diamidino-2-phenylindole (DAPI) solution (1 ng/µl) in the dark for 10 min, and mounted slides in Mowiol anti-fading medium (Kuraray Europe GmbH). We captured the fluorescent signals with a laser scanning confocal microscope Olympus FV3000 (Olympus). We acquired at least three confocal stacks (up to 30 scans per optical slice) at 100×, 400× and 630× magnifications, a colour depth of 24 bit and a resolution from 1 to 2 µm per pixel (depending on the fluorochrome) by investigating multiple regions from each of the

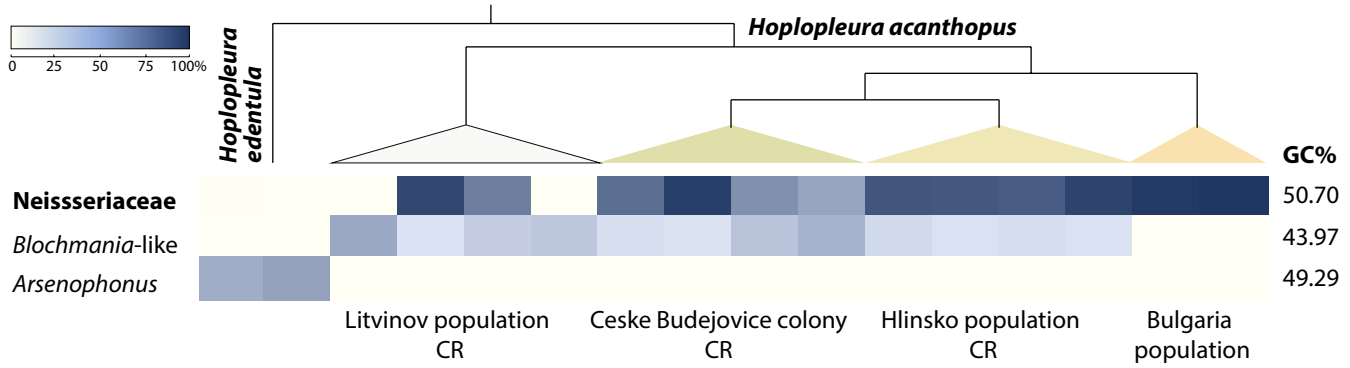

FIGURE 1 Dominant taxa found in *Hoplopleura edentula* and four different populations of *Hoplopleura acanthopus*. The heat map is based on relative abundance of the bacterial OTUs comprising 10 or more percent of the reads within a particular sample. The phylogenetic scheme simplifies the detailed phylogenetic relationships of the hosts as shown in Figure S1 [Colour figure can be viewed at wileyonlinelibrary.com]

three replicate specimens. Resultant images were processed with the ImageJ distribution Fiji (Schindelin et al., 2012; Schneider et al., 2012).

3 | RESULTS

3.1 | Amplicon based screening

Sequencing of multiplexed 16S rRNA gene libraries produced high quality amplicon data. On average, we retrieved 21,205 and 20,201 merged 16S rRNA bacterial sequences across individual *Polyplax* and *Hoplopleura* samples, respectively. The average read number for the even and staggered standards, included in the *Hoplopleura* library as positive controls (see Methods), equalled 32,513. For the even mock communities, the sequencing recovered all 10 bacterial taxa in comparable abundances. The data from the staggered communities, designed for testing sequencing sensitivity and PCR bias, confirmed that our approach can reveal complete microbiome profiles, including low abundant taxa. In particular, sequences of *Bifidobacterium adolescentis* (ATCC15703) and *Deinococcus radiodurans* (ATCC BAA-816), both present in 0.04% of the original mock DNA template, comprised on average 0.02% and 0.01% of the reads among the three sequenced mock communities. We observed preferential amplification of a dominant community component, *Staphylococcus epidermidis* (ATCC 12228), in the staggered standards. Compared to the mock template where this taxon comprises 44.78% of the total DNA, the average read abundance equalled to 70.12%. Potential preferential amplification of some taxa in the analysed microbiomes would cause a systematic error, inherent to the entire data set, and thus not affecting abundance-based dissimilarity measures and clustering analyses.

3.2 | Distribution of dominant bacterial taxa

Microbiomes of the tested species/lineages of the lice are dominated by several bacterial taxa with a complex distribution pattern. For two of these bacteria, *Legionella polyplacis* from *Polyplax* and Neisseriaceae from *Hoplopleura* lice, the data on their distribution could be complemented by complete genome characteristics. For other bacteria, partial 16S rRNA gene amplicons provide two (not entirely independent) kinds of information: taxonomical assignment and GC content. Among the dominant taxa (see methods for definition), the highest abundance was detected for the Neisseriaceae OTU, which was present in 12 of the 14 *H. acanthopus* samples, in two of them being the only dominant symbiont. Two additional dominant OTUs were taxonomically assigned to the genera *Blochmannia* (within 12 specimens total, in two of them as the only symbiont) and *Arsenophonus* (the only symbiont detected in *H. edentula*; Figure 1; details on host phylogeny in Figure S1). Both of these genera are endosymbioic with insects. *Blochmannia* is a P-symbiont of carpenter ants (Sauer et al., 2000) and *Arsenophonus* is a widely distributed bacterium with a broad range of symbiotic associations (from parasitism to mutualism) in different insect groups (Nováková et al., 2009). The two OTUs related to these genera possess sequence characteristics (GC content of 43.9% and 49.3%) and distributions among the two *Hoplopleura* species and *H. acanthopus* populations which imply their symbiotic nature.

For *Polyplax serrata* a comprehensive population-wide amplicon screening revealed (besides *L. polyplacis*) nine dominant taxa assigned to the genus or family level (Figure 2; details on host phylogeny in Figure S2). The distribution of OTUs with a low GC content, i.e., *Buchnera* (45.1% GC) and *Arsenophonus* 2 (49.4% GC), appears to be nonrandom with respect to the host genealogy, possibly indicating putative obligate coevolving symbionts. For the other taxa, the taxonomic assignment and GC content >50% (with the exception of *Cloacibacterium*) indicate that the bacteria may represent environmental contamination or very early symbiotic associations, e.g., Neisseriaceae taxon and *Arsenophonus* 1. However, it is important to note that OTU taxonomic assignments are based on a short sequence and should be interpreted as approximate affiliations rather than precise phylogenetic position, particularly in the case of symbionts with highly derived genomes like *Buchnera* and *Blochmannia*.

Community quantitative measures and the statistical evaluation of clustering patterns performed for the complete microbiome data

(Figure S3) further reveal that the communities are shaped by *P. serrata* genealogy and geographical origin. *P. serrata* microbiomes significantly differ among the major phylogenetic clades (A, W, E and N in Figure 2; R^2 = 0.037, $p \leq 0.001$,) as well as individual phylogenetic lineages within the W clade (W1 to W7 in Figure 2; R^2 = 0.16, $p \leq 0.001$). The most variability among the microbiomes in *P. serrata* W clade is then explained by their geographical origin (Figure S3; R^2 = 0.34, $p \leq 0.001$). Similar, statistically significant dissimilarities among microbiomes from different genetic clusters and localities were identified in the E clade (Figure S4).

3.3 | Genomes of the Neisseriaceae-related symbionts

The complete closed genome of the obligate *Hoplopleura acanthopus* symbiont is 1,607,498 bp long with 33.4% GC content and coding density 83.48%. It contains 1,421 predicted protein coding genes, nine genes coding rRNAs, and 39 genes coding tRNAs (Data S2). The 9 rRNA genes represent three complete 16S-23S-5S rRNA gene operons, but they are arranged in an unlinked manner known for some other bacteria (Brewer et al., 2020), including other P-symbionts (Munson et al., 1993). The 23S rRNA and 5S rRNA genes are placed in close proximity, while the 16S rRNA gene is separated by long stretches of DNA.

The genome draft of the Neisseriaceae-related symbiont in *Polyplax serrata* consisted of 32 contigs (Figure 3) which sum to 1,814,374 bp, with the GC content 33.7% and coding density 89.36%. It contains 1748 predicted protein coding genes and 35 tRNA genes (Data S2). Due to the fragmentation of the genome in contigs, the rRNA genes were not reliably assembled and their number remains unclear. Of the 36 genes absent in this draft genome annotation but present in the *H. acanthopus* symbiont assembly, the *P. serrata* metagenomic reads mapped only on two genes, uracil-DNA glycosylase family 1 and Fe-4S ferredoxin iron-sulphur binding. However, both genes were found in the complete metagenomic assembly of *P. serrata*. The sequence of Fe-4S ferredoxin iron-sulphur binding is present in the draft genome but not recognized and annotated as a functional gene by the RAST server. The gene for uracil-DNA glycosylase family 1 was assembled together with a gene for a sodium-dependent transporter into a short two-gene contig. Since this contig has an order of magnitude lower coverage than the rest of the assembly (2.9 in contrast to approx. 30) and the sodium-dependent transporter yielded *Cutibacterium* (Actinobacteria) when blasted against GenBank nr database, we removed the contig from the assembly.

In both symbiont genomes, the lengths of some annotated loci (excluding hypothetical proteins) were less than 80% of the closest sequences retrieved by BLAST, possibly indicating a pseudogenization process. In the *H. acanthopus* symbiont, only 20 such loci were identified, often determined as mobile elements and transposases, either during the annotation process or by the BLAST analyses (Data S5). In the *P. serrata* symbiont, the proportion of the truncated loci was considerably higher, containing 105 loci from various categories (Data S5). Pseudofinder analysis yielded similar numbers of potential

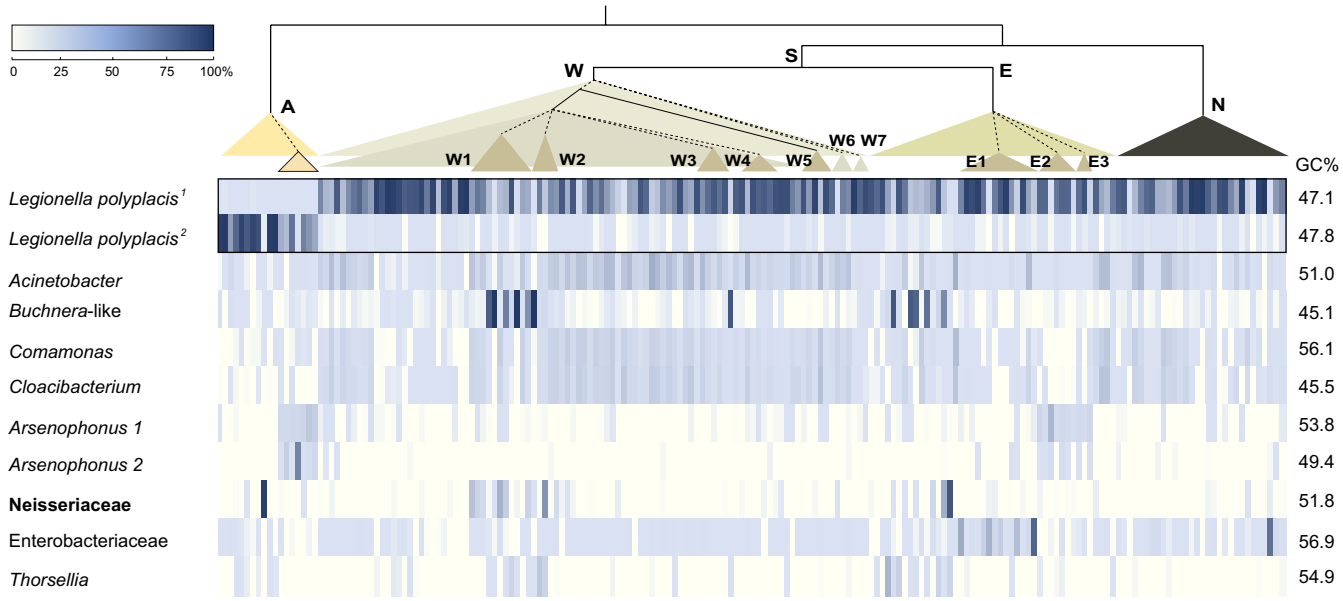

FIGURE 2 Dominant taxa found in *Polyplax serrata* samples from distinct populations. The heat map is based on relative abundance of the bacterial OTUs comprising 10 or more percent of the reads within a particular sample. The phylogenetic scheme, which simplifies the COI based phylogeny of individual samples, is provided in Figure S3. Designation of the branches is based on mtDNA structure described in the study by Martinů et al., (2018): A, lineage specific to *Apodemus agrarius* and *Apodemus uralensis*, S, lineage specific to *Apodemus flavicollis* (W, west sublineage, E, east sublineage), N, nonspecific lineage from *Apodemus flavicollis* and *Apodemus sylvaticus*. W1–W7 and E1–E3 stand for individual genetic clusters within west and east sublineages as specified in Figure S2. The numbers for *Legionella polyplacis* designate two different OTUs, reflecting evolutionary changes accumulated after the split of *Polyplax serrata* lineage [Colour figure can be viewed at wileyonlinelibrary.com]

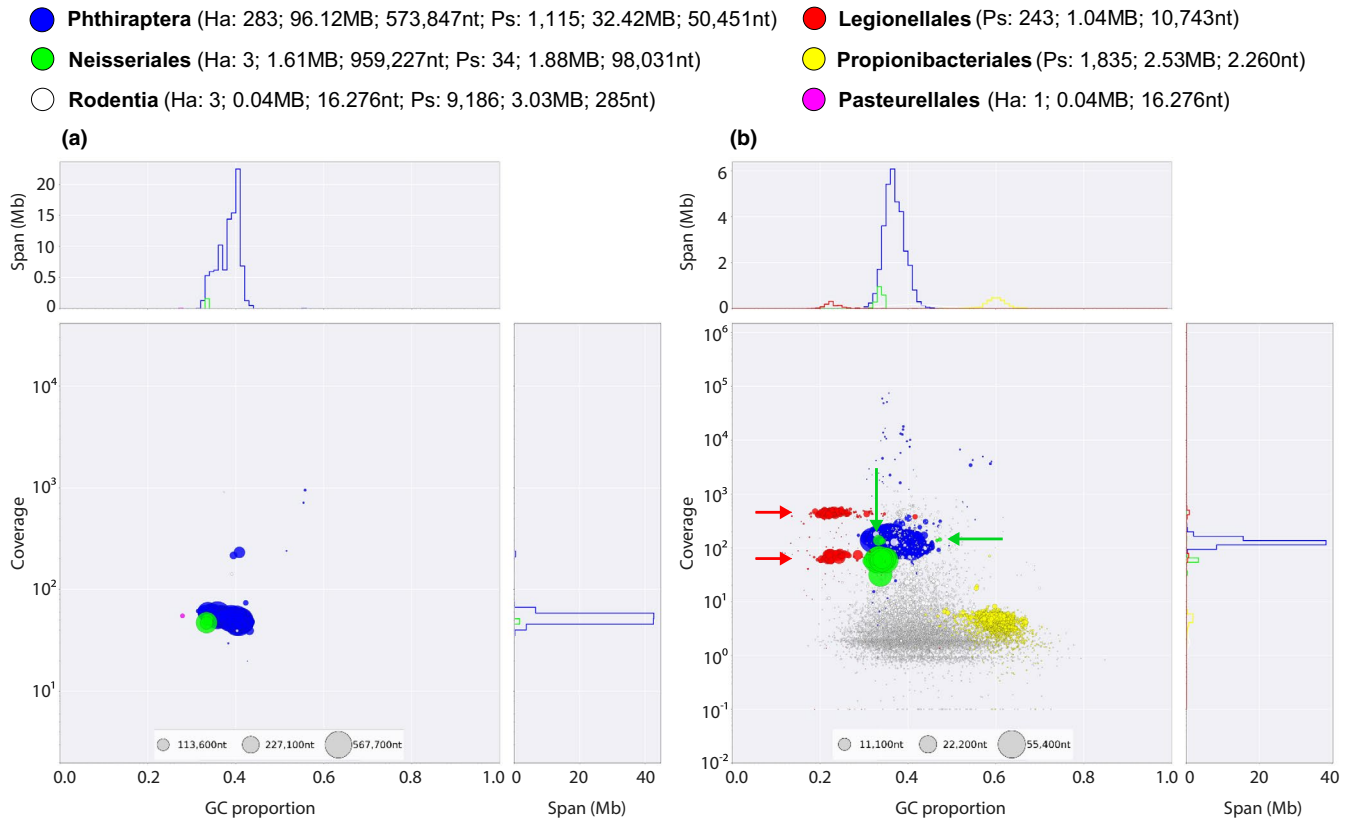

FIGURE 3 Blobplots of *Hoplopleura acanthopus* (a) and *Polyplax serrata* (b) metagenomic assemblies. Contigs are plotted based on their GC content and coverage, scaled by span, and coloured by their taxonomic assignments. For each taxon, the number of assigned contigs, their total length (span), and the average contig length are provided in the bracket. Red horizontal arrows: two strains of *L. polyplacis* separated by different coverage; green vertical arrow: two contigs containing phage-associated and secretion system associated genes, causing seemingly higher coverage; green horizontal arrow: two short contigs containing ribosomal RNAs. The rRNA containing contigs were not reliably assembled (probably chimeric assembly of two or more rRNA copies) and therefore they are not included in the final Neisseriaceae-*Ps* genome [Colour figure can be viewed at wileyonlinelibrary.com]

pseudogenes (11 and 106 in the *H. acanthopus* and *P. serrata* symbiont, respectively), only partially overlapping with the former analysis (Data S5).

BLAST analysis against the NCBI nr database showed that a large proportion of the genes in both symbionts did not yield Neisseriales within the best hits (see Methods and Data S2). This is in remarkable contrast to *Legionella polyplacis*, for which 96.3% of the genes were assigned to the genus *Legionella* by the BLAST search. The difficulty with taxonomic assignment was also reflected in the Blobtools analyses, since some of the contigs could only be assigned to Neisseriales by a more thorough multi-hit BLAST analysis of individual genes (Data S2; Figure 3). This assignment was further confirmed by the high proportion of genes shared in the two Neisseriaceae-related symbionts and considerable degree of their synteny (Data S2). When aligned with Mauve software the syntenies were placed into 34 locally collinear blocks, the longest blocks exceeded 100 kb with a span of 100 genes (Figure 4).

To make the following text intelligible, we designate Neisseriaceae-related symbionts as Neisseriaceae-*Ha* (symbiont of *Hoplopleura acanthopus*) and Neisseriaceae-*Ps* (symbiont of *Polyplax serrata*). Genome size and content of these two symbionts, in comparison to other lice symbionts (*Riesia* spp., *Puchtella pedicinophila*, *Legionella polyplacis*, and the symbiont from *Proechinophthirus*), *Snodgrassella* and *Neisseria*, are summarized in Table 1.

3.4 | Metabolic capacity

With their genome sizes and gene contents, the two Neisseriaceae-related symbionts appear to be amongthe less reduced and therefore most metabolically diverse symbionts sequenced from lice (Table 1). This is well demonstrated by a comparison of important cellular and metabolic functions. For example, within the categories associated with recombination and repair, or cell wall biosynthesis, the two symbionts retained higher numbers of genes than the other louse symbionts, with the only exception of *Sodalis* from *Proechinophthirus fluctus* (Figure 5b). Similarly, reconstruction of the main metabolic features inferred for the genome of Neisseriaceae-*Ha* symbiont (Figure 5a) shows a relatively high degree of completeness when compared to the other

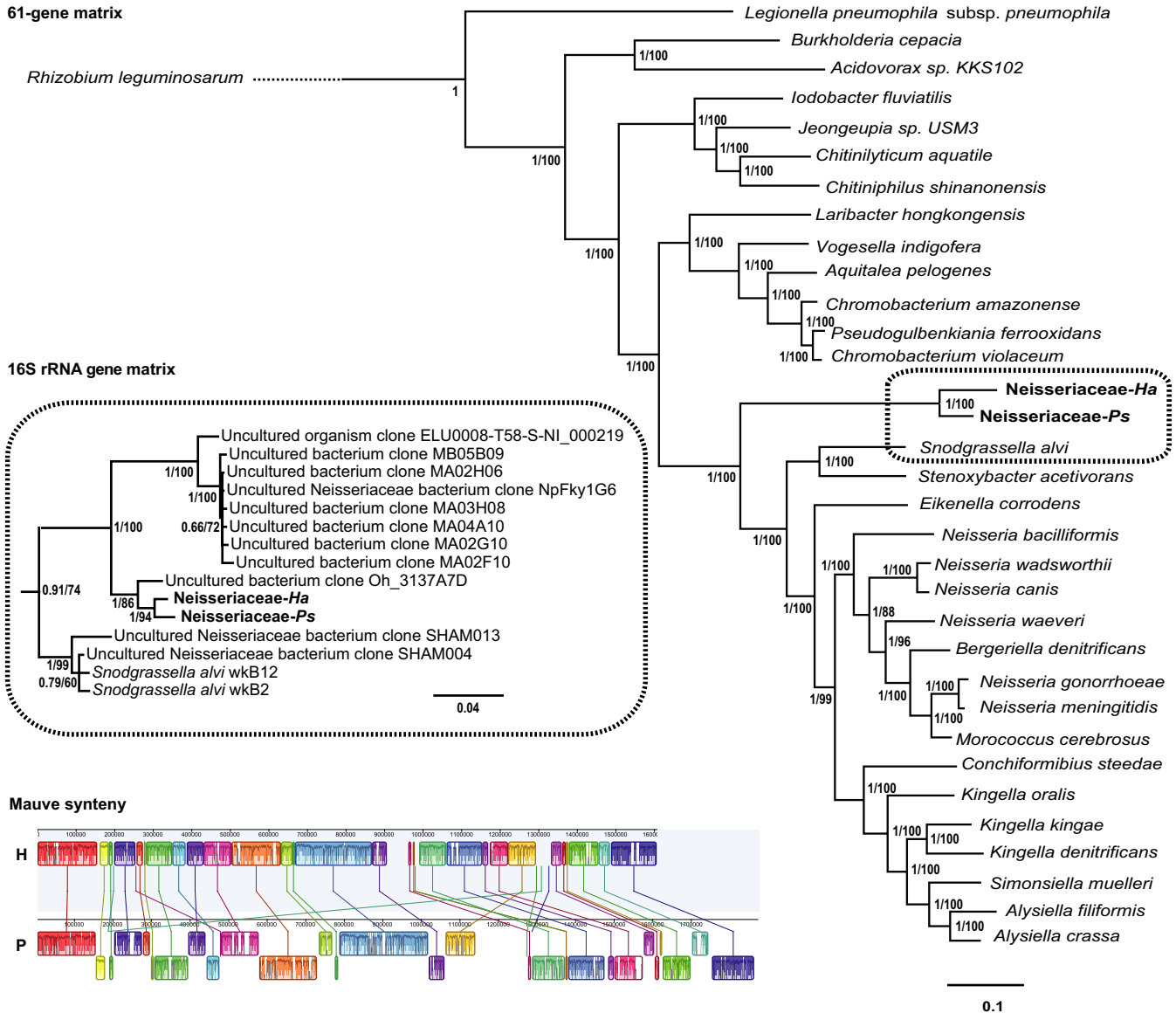

FIGURE 4 Phylogenetic relationships of the two Neisseriaceae-related symbionts. 61-gene matrix: Bayesian analysis of the amino-acid multigene matrix; the numbers at the nodes show posterior probabilities/bootstrap supports obtained by the maximum likelihood analysis in PhyML. 16S matrix: part of the tree obtained by the Bayesian analysis of the nucleotide 16S matrix showing relationships of the two Neisseriaceae-related symbionts to several Uncultured bacteria (see Figure S5 for the complete tree); the numbers at the nodes show posterior probabilities/bootstrap supports obtained by the maximum likelihood analysis in PhyML. The dashed boxes indicate corresponding parts of the 61-gene and 16S rRNA trees. Mauve synteny: an overview of strong synteny between the two symbionts; H, Neisseriaceae-*Ha*; P, Neisseriaceae-*Ps* (see Data S2 for a complete list of the genes) [Colour figure can be viewed at wileyonlinelibrary.com]

bacteria we included (Figure 5c). The only bacteria with higher or comparable metabolic capacities were the nonsymbiotic *Neisseria meningitidis*, the bee symbiont *Snodgrassella*, and the closely related Neisseriaceae-*Ps* symbiont. Regarding the metabolites usually associated with the insect-bacteria symbiosis, the two Neisseriaceae-related symbionts possess the highest capacity for amino acid synthesis among all the louse symbionts compared here. They both share the same pattern of the present/absent pathways, except for histidine biosynthesis, which is complete in Neisseriaceae-*Ps* but missing in Neisseriaceae-*Ha*. Another metabolic difference between these two symbionts consists in their capacities to synthesize B vitamins. While both possess potentially functional pathways for thiamine, riboflavin, biotin and folate, Neisseriaceae-*Ps* genome also encodes for synthesis of niacin and possibly pyridoxine.

Regarding general cell metabolism, glycolysis, as the core of energetic metabolism and a source of intermediates for all other metabolic pathways, is complete and probably functional in both Neisseriaceae-related symbionts. Among the other compared bacteria, this central metabolic pathway seems to be compromised in two symbionts. In *L. polyplacis* the pathway is highly incomplete, missing several genes. In *Sodalis*, the payoff phase is potentially inactive due

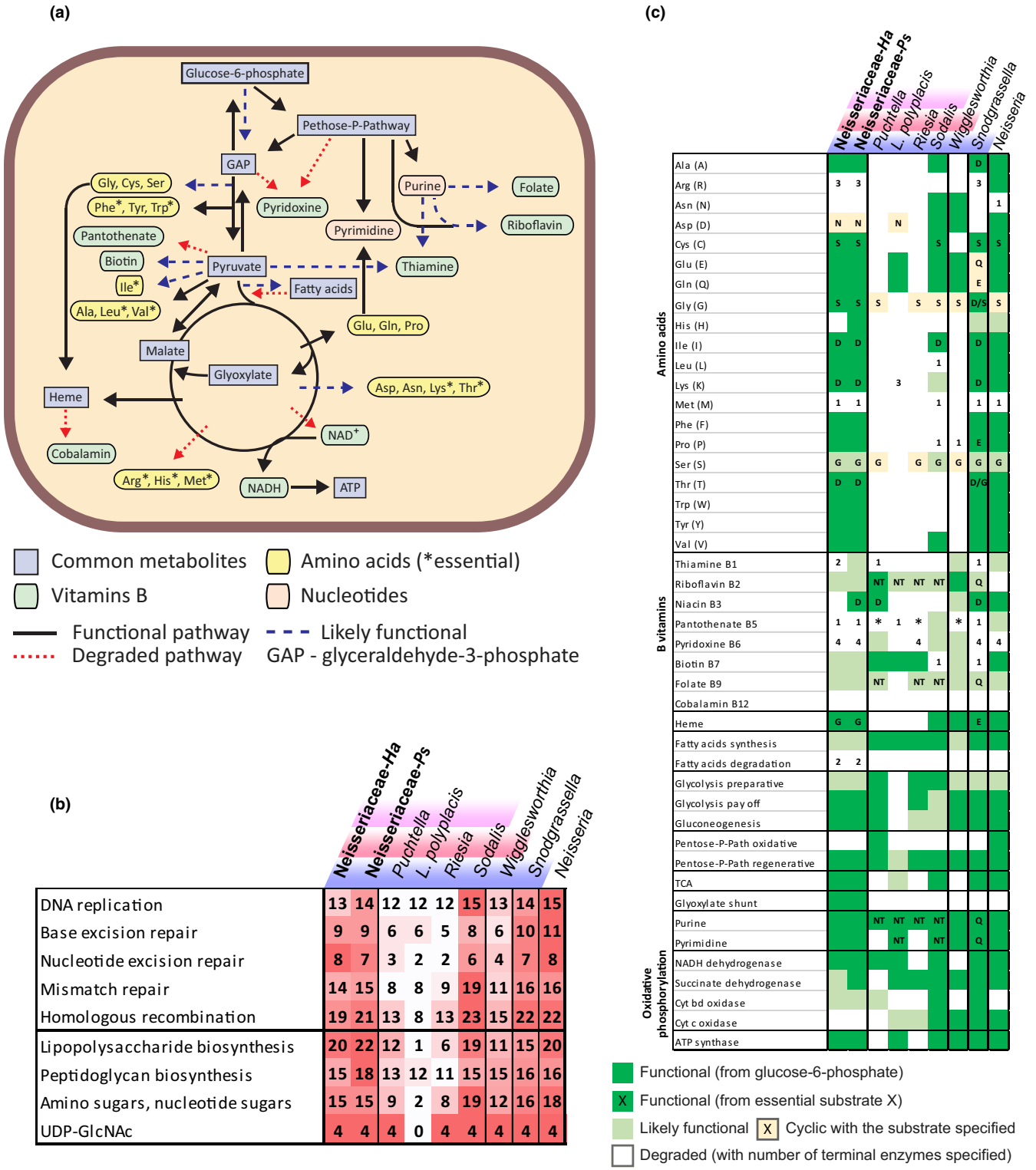

FIGURE 5 (a) Reconstruction of the main metabolic pathways of the Neisseriaecea-*Ha* symbiont. (b) Comparison of gene content (shown as number of genes) for important cellular functions in genomes of the two Neisseriaceae-related symbionts and other bacteria. C. Comparison of important metabolic pathways encoded by genomes of the two Neisseriaceae-related symbionts and other bacteria. The inbox letters are amino acids abbreviated by their common one letter code; NT stands for arbitrary nucleotide. Asterisk indicates a pathway of pantothenate synthesis starting from an intermediate 3-Methyl-2-oxobutanoate. Ecological characteristics of the compared taxa are indicated by background colouring of their names: blue, symbiotic bacteria in insects; red, symbionts of blood feeding insects; violet, louse symbionts [Colour figure can be viewed at wileyonlinelibrary.com]

to the missing phosphoglycerate kinase (2.7.3.2). Completeness of the citric acid cycle (tricarboxylic acid cycle, TCA) correlates with the degree of genome streamlining. In the three most reduced genomes (*Puchtella*, *Riesia* and *Wigglesworthia*) the cycle is strongly degraded. In *L. polyplacis* the production of citrate from oxaloacetate and Acetyl-CoA is hampered due to the missing citrate synthase [2.3.3.1] or its alternatives. Both Neisseriaceae-related symbionts, as well as the other compared bacteria, possess functional oxidative phosphorylation consisting of NADH dehydrogenase and ATP synthase, and at least one of two cytochrome oxidases, the c or bd type. De novo synthesis of purine and pyrimidine bases and/or nucleotides is fully functional in both Neisseriaceae-related symbionts and also in *Neisseria meningitidis* and *Wigglesworthia*. In the other compared bacteria this capacity is compromised, purine nucleotides can only be created by interconversion. In pyrimidine synthesis, the two strongly reduced symbionts *Puchtella* and *Riesia* not only lack the capacity for de novo biosynthesis of the nucleotides but also for their interconversion (details in Figure 5). Both Neisseriaceae-related symbionts (as well as *Sodalis* and *Wigglesworthia*) lack the initial enzyme for biosynthesis of fatty acids (S-acetyltransferase), although full biosynthetic capacity is preserved in the other louse symbionts. Degradation of fatty acids is dysfunctional in all the compared bacteria. In addition to the generally higher metabolic capacity, both Neisseriaceae-related symbionts also retain various genes connected to DNA exchange and/or transport, such as mobile elements, type IV pili, and secretion systems (Data S2).

3.5 | Phylogeny

In both phylogenies (the 16S rRNA gene and the 61-gene) the two Neisseriaceae-related symbionts cluster as sister taxa on a long common branch, placed firmly within Neisseriales with high nodal support. In the more robust multigene analysis the pair branches as an isolated offshoot at the base of Neisseriaceae (Figure 4). The analysis of 16S rRNA gene, for which a broader taxonomic spectrum is available, revealed additional close relatives which could not be included into the "61-gene matrix" due to lack of data (Figure S5; detail in Figure 4). All of them have been described as "*Uncultured bacterium*" from human and insect samples (oksanOteo et al., 2014). In contrast to the "61-gene matrix", the "16S matrix" placed the pair of Neisseriaceae-related symbionts into a monophyletic cluster together with *Snodgrassella* and several lineages of the "*Uncultured bacterium*".

3.6 | Localization of the lice symbionts

Both the specific Neisseriaceae and the universal bacterial probes hybridized to bacteria located within DAPI-stained cells (~20–30 μm). In females, they were detected in putative bacteriocytes which formed weakly adherent clusters found below ovarial ampulla (Figure 6), and in the cells of ovariole sheaths (Figure S6). In bacteriocytes, the two probes did not provide entirely overlapping patterns.

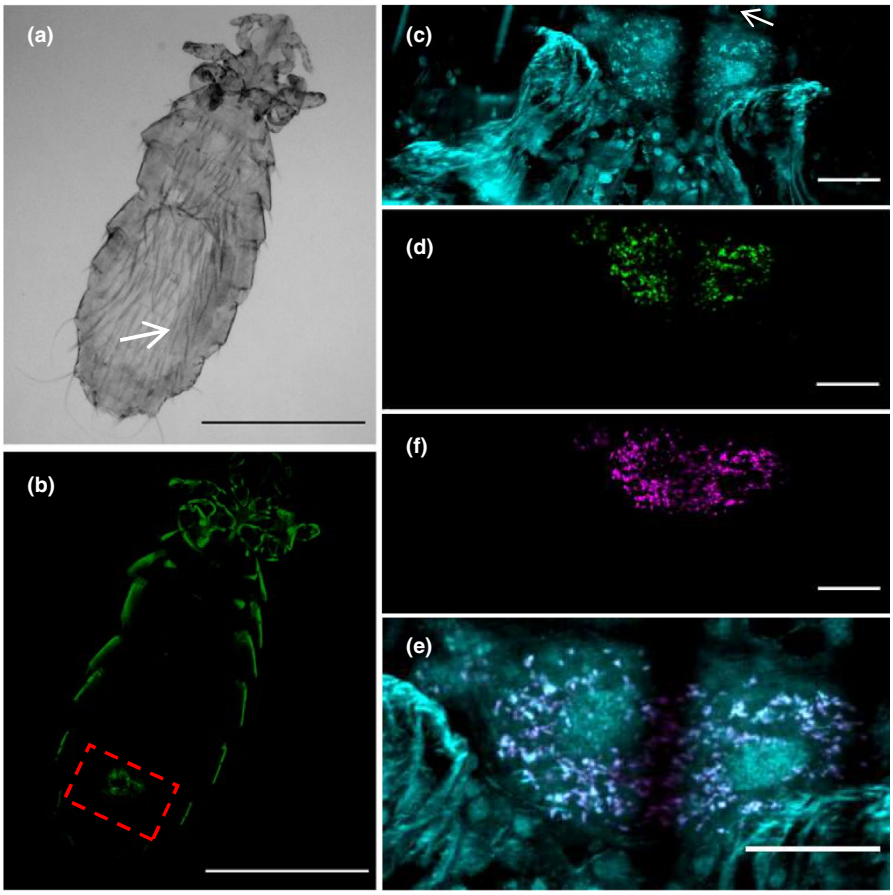

FIGURE 6 Light and confocal microscopy of a FISH-stained female *Hoplopleura acanthopus*. (a) Light microscopy image showing the louse body containing a developing egg (the dark area on the right side of the louse body marked by white arrow). (b) Hybridization signal for the Neisseriaceae specific probe beta-572 (green) shows the localization of the symbionts. The dashed red square defines the region shown in panels c, d and e. (c, d and e) Hybridization signals for DAPI (cyan), the Neisseriaceae specific probe beta-572 (green) and the generic bacterial probe EUB338 (magenta) are shown in panels c, d and e, respectively, and were combined in the merged colour image (f), with the Neisseriaceae-related symbionts appearing white. DAPI staining in panel c defines also the localization of the bacteriocytes below ovarial ampulla (o. a. indicated by the arrow). Scale bars: (a, b) 500 μm, (c, d, e and f) 20 μm [Colour figure can be viewed at wileyonlinelibrary.com]

However, it is difficult to decide whether this difference is due to the presence of two different bacteria within the bacteriocytes or due to the difference in signal intensity produced by chemically distinct chromophores. Many bacteriocytes containing intracellular bacteria were also found in the posterior part of the abdomen in male lice (Figure S7).

4 | DISCUSSION

The results of the combined genomic and amplicon analyses illustrate dynamic changes of microbiome composition among louse populations and species, indicating a role as a potential source of diverse mutualistic symbionts observed in Anoplura (Allen et al., 2016; Boyd et al., 2014, 2016; Fukatsu et al., 2009; Hypša & Křížek, 2007). The two Neisseriaceae-related symbionts represent a novel lineage, extending the known phylogenetic span of louse symbionts with a member of Betaproteobacteria. This finding supports the a priori expectation of this study that Anoplura possess a unique ability to acquire symbionts from diverse bacterial groups. Considering their genomic characteristics together with their distribution across the lice populations, we hypothesize that the two Neisseriaceae-related bacteria are symbionts in an early/intermediate stage of evolution (i.e., genome degeneration), which established their symbiosis independently within two different genera of lice. In *Hoplopleura*, they seem to be evolving towards typical P-symbionts. Consistent with this view, FISH analysis of the Neisseriaceae-*Ha* symbionts highlighted their localization within putative bacteriocytes (Figure 6 and Figure S7; negative control in Figure S8) and in cells of ovariole sheaths (Figure S6). Amplicon screening also showed that both Neisseriaceae-related symbionts are part of a rich and diverse microbiome (Figures 1 and 2), in which some of the other dominant taxa are closely related to other known obligate symbionts in insects. This complex picture suggests dynamic turnover of the symbiotic bacteria, with frequent acquisitions, losses and replacements, even within young phylogenetic lineages.

4.1 | Genomic and metabolic characteristics of the Neisseriaceae-related symbionts suggest an intermediate stage of evolution

We propose that the Neisseriaceae-related symbionts are in an early or intermediate stage of symbiosis evolution. We base this assertion on a generally accepted view that evolution from free-living bacteria to highly modified symbionts includes degenerative processes, such as genome shrinking, bias towards AT bases, and loss of metabolic capacities (for review see McCutcheon et al., 2019). For a comparison with other lice-associated symbionts, full genomes are currently available for four lineages. Three are highly reduced and display a strong compositional shift towards AT, typical for many P-symbionts (Table 1). The fourth possesses a large genome and is supposedly a recently acquired associate which replaced an older obligate symbiont (Boyd et al., 2016). The genomes of the two Neisseriaceae-related symbionts can thus be compared to various lice P-symbionts and to "free living" members of Neisseriales. Both Neisseriaceae-*Ha* and Neisseriaceae-*Ps* display significantly weaker genome degeneration than the highly reduced *Riesia*, *Puchtella*, and *Legionella polyplacis* (Boyd et al., 2017; Říhová et al., 2017). They have higher GC content, considerably larger genomes (approximately three times higher number of genes), and consequently more complete metabolic pathways (Figure 5). On the other hand, their genomes are recognizably reduced, and the GC content is decreased when compared to their relatives (e.g. the genus *Neisseria*, but also the bee symbiont *Snodgrassella*) or to the presumably young *Sodalis*-like symbiont from the louse *Proechinophthirus fluctus* (Boyd et al., 2016). When the two symbionts are compared to each other, the Neisseriaceae-*Ha* shows stronger reduction, retaining a lower number of genes, slightly lower GC content, and also significantly lower number of potential pseudogenes. This is consistent with its presumed role as an established obligate symbiont.

At the metabolic level the two Neisseriaceae-related symbionts share similar patterns of absent/present pathways, and their metabolic capacity is considerably higher than in typical obligate louse symbionts (*L. polyplacis*, *Riesia*, and *Puchtella*) or the tsetse fly symbiont *Wigglesworthia* (Figure 5). On the other hand, *Sodalis* from *Proechinophthirus fluctus*, also hypothesized to be a young symbiont that recently replaced an older symbiotic bacterium in the host (Boyd et al., 2016), retained its metabolic capacity to a comparable extent. The higher metabolic capacity of the two Neisseriaceae-related symbionts is well demonstrated for amino acid and B vitamin synthesis, two categories often considered crucial for insect-bacteria symbiosis. Compared to the 16 amino acid biosynthetic pathways in Neisseriaceae-*Ha* symbiont, the obligate symbionts *Riesia*, *Puchtela* and *L. polyplacis* possess only 2–4 potentially functional pathways, indicating that most of the amino acids produced by the Neisseriaceae-*Ha* symbiont may not be required by the host. In a similar way, the two Neisseriaceae-related symbionts can produce more B vitamins than the other louse symbionts. However, only the riboflavin (B2) and folate (B9) pathways seem to be preserved in all the compared louse symbionts. Biotin (B7) is the next most frequently present pathway, absent only in *Sodalis*. It should be noted that the degrees of completeness observed for different pathways do not necessarily reflect the actual role of the symbionts in provisioning metabolites to the host. This can be demonstrated by the highly degraded genomes of the obligate louse symbionts *Riesia*, *Puchtella* and *Legionella*. Based on the gene content, most of the B vitamins pathways are strongly degraded in these symbionts (Figure 5c). However, experimental evidence suggests that the body louse *Pediculus humanus* (and so possibly other species) may depend on provisioning of most B vitamins by the symbiont (Puchta, 1955). For pantothenate, this view is further supported by the presence of three important enzymes (panB, panC and panE) on the *Riesia* plasmid (Kirkness et al., 2010). The considerable incompleteness of the whole pantothenate pathway in *Riesia* (i.e., from glycolysis to pantothenate) shown in Figure 5c may indicate that this symbiont

is probably utilizing some of the host metabolites as a substrate for pantothenate biosynthesis. In our comparison we observed a similar situation, i.e., presence of these three enzymes but absence of the complete pathway, in two additional louse symbionts (asterisks in the Figure 5c). The genome-based comparison of metabolic capacities thus can serve as a base for assessing the degree of degradation rather than the actual metabolic role of the symbionts. It also indicates that the higher complexity of the two Neisseriaceae-related symbionts, compared to the strongly reduced genomes of the obligate symbionts, may be due to their early stage of symbiogenesis and weak degeneration rather than an adaptive preservation of the metabolic pathways. This conclusion is further supported by few differences between the two Neisseriaceae-related symbionts. The lower capacity of Neisseriaceae-*Ha* in comparison to Neisseriaceae-*Ps* (e.g. histidine, niacin, pyridoxine) indicates an ongoing process of reduction upon establishment of obligate symbiosis in the former symbiont. Based on the comparison with the more reduced obligate symbionts, we assume that the metabolic capacities of the two Neisseriaceae-related symbionts considerably exceed the metabolic demands of their hosts, and metabolic function of these symbionts cannot be assessed solely from their genome content.

If, as suggested by many of the described genomic features, Neisseriaceae-*Ha* is a young symbiont in an early state of evolution, we could assume that it only recently replaced a more ancient P-symbiont which was fulfilling the nutritional role prior to the acquisition of the Neisseriaceae-related symbiont. A similar scenario has been recently proposed for the *Sodalis* symbiont of *Proechinophthirus fluctus* by Boyd et al., (2016). As discussed below, a putative P-symbiont was indeed detected by amplicon screening in the majority of the *H. acanthopus* microbiomes. Also, the FISH survey indicated that the bacteriocytes might be inhabited by two different bacteria. However, this interpretation should be taken with caution since the non-overlapping signal may reflect different properties of the used chromophores (Figure 6; Demchenko, 2020; Eggeling et al., 1998; Mahmoudian et al., 2011).

4.2 | Distribution and origin of the symbionts in the lice microbiomes

Since the process of genome degradation - typical for symbiotic bacteria - accelerates once the bacterium becomes an obligate vertically-transmitted symbiont, we should expect, at least during the initial phase of the symbiogenesis, a correlation between the degree of genome degradation and the duration of host-symbiont coevolution (Moran, 1996). For example, among the symbionts of sucking lice, the high degree of genome degradation of *R. pediculicola* and *L. polyplacis* indicates a relatively long and intimate association with the host. In correspondence with this presumption, the *Riesia* lineage has been found in several louse species of two different genera, *Pediculus* and *Phthirus*, and *L. polyplacis* is hosted by at least two louse species, *P. serrata* and *P. spinulosa* (Hypša & Křížek, 2007). Moreover, our extensive amplicon screening shows that *L. polyplacis* is consistently present in a broad geographic and phylogenetic sample of *P. serrata* as a dominant bacterium (Figure 2). When compared to these two well documented examples of established P-symbionts, the Neisseriaceae-related symbionts, with an intermediate degree of the genome degeneration, show more restricted and patchy distributions. Only in *H. acanthopus* were they consistently present as the most dominant bacterium (the Neisseriaceae OTU in Figure 1), with the exception of two specimens from one population. This absence cannot be unequivocally explained from the data. It can for example reflect changes of the symbiont's titre during the physiological cycles. It is also interesting to note that absence in a specific host lineage of a similarly widespread, presumably obligate symbiont has been demonstrated in the ant *Cardiocondyla obscurior* and its symbiont *Westeberhardia* (Klein et al., 2016). The overall variability of the microbiomes was apparently correlated with lice genetic background: in Bulgarian samples the Neisseriaceae OTU was the only dominant bacterium, in other populations of *H. acanthopus* it was accompanied by an unknown bacterium which the BLAST search affiliated with *Blochmannia*, and in the two samples of *H. edentula* the only present OTU corresponded to the genus *Arsenophonus*. Since *P. serrata* is known to harbour the typical obligate P-symbiont *Legionella polyplacis* (Říhová et al., 2017), we screened this louse more extensively across several populations and genetic lineages. The results confirmed the ubiquitous presence of *L. polyplacis*, which in most cases was the most abundant OTU, with only occasional co-occurrence of the Neisseriaceae OTU (Figure 2). The split of *L. polyplacis* into two different OTUs correlated with host phylogeny, reflecting evolutionary changes during the evolution of the symbiont in distant host lineages, but certainly does not suggest the presence of two independent symbiotic lineages (in fact phylogenetic and genomic analyses confirm that *Polyplax-Legionella* co-evolution crosses the host species' boundaries, and the same symbiont is also present in the related louse species *P. spinulosa*; Hypša & Křížek, 2007). The split of *L. polyplacis* is also reflected in the bloptools plot with two distinct clusters assigned to Legionelalles (Figure 3). In contrast, the compact clustering of the Neisseriaceae-*Ps* contigs shows that no such differentiation due to long-term coevolution took place in this symbiont (Figure 3).

With regard to the general concept of symbiont acquisition, loss, and replacement within insects, and the high dynamism of louse microbiomes, two OTUs are of particular interest. Both OTUs assigned by BLAST to highly derived obligate symbionts, *Buchnera* in *P. serrata* and *Blochmannia* in *H. acanthopus*, seem to represent strongly derived symbiotic genomes of Enterobacteriaceae (Figures 1 and 2), for which BLAST assigned taxonomy reflects low GC content rather than real phylogenetic relationships. Since our metagenomic data did not yield any reliable information on either of these bacteria, it is difficult to hypothesize about their phylogenetic origin and function in the host. A genome reduction, deduced from the GC content of the 16S rRNA gene amplicon, suggests that they may represent some form of symbiotic bacterium, e.g., facultative symbionts or even scattered remains of ancient and novel symbionts, now retreating from the host's population and replaced with

more recent acquisitions. Coexistence of ancient and novel symbiont, with possible replacement of metabolic functions, has been recently documented for example in the system of *Buchnera* and *Erwinia*-related symbionts in *Cinara* aphids (Manzano-Marín et al., 2020). Interestingly, the FISH analysis shows that apart from the Neisseriaceae-*Ha* symbiont, the bacteriocytes of *H. acanthopus* may harbor another bacterium (Figure 6). Since the metagenomic assembly did not contain any other bacterial contigs, we were not able to identify the origin of this second symbiotic bacterium.

The diversity of microbiomes and the rapid process of symbiont acquisition/replacement make sucking lice an interesting model for studying the conditions and processes in the early stage of symbiogenesis. Considering the distribution patterns and the low degree of genome modifications in the two Neisseriaceae-related symbionts, it is unlikely that their occurrence in two different lice lineages is due to a common symbiotic origin in the *Hoplopleura-Polyplax* ancestor. The "common-origin" scenario would suppose that these symbionts were secondarily lost in many lineages of louse hosts, particularly in *Polyplax*. Also, the weak genomic modification suggests a recent origin rather than a long coevolutionary history dating back to the common ancestor of the two louse genera. While these features do not provide unequivocal evidence, an independent origin of the Neisseriaceae-related symbionts in each louse genus is a more parsimonious explanation. This poses an interesting question on the source of these symbionts and the mechanisms underlying their acquisition and symbiogenesis. Co-occurrences of closely related symbiotic bacteria in related insect hosts are usually consequences of either cospeciation or a tendency of specific bacterial lineages to frequently establish symbiosis with specific insect hosts (e.g., *Arsenophonus*, *Wolbachia*). However, neither of these explanations can be applied to the lice-Neisseriaceae association. Members of the family Neisseriaceae are only rarely found in symbiotic association with insects. The only well documented case of obligate symbiosis is the genus *Snodgrassella* found in several species of bees and bumblebees (Kwong & Moran, 2013). Based on the 16S rRNA gene phylogeny, the closest relative of the louse-associated Neisseriaceae is an uncultured bacterium described from a flea *Oropsylla hirsuta* (Jones et al., 2008), for which no other information is currently available. It is interesting to note that, similar to *Legionella polyplacis*, the Neisseriaceae-related symbionts originate from a bacterial lineage which is rarely found in insects and is a well-known vertebrate pathogen. Phylogenetic correlation between the vertebrate pathogens and symbionts of blood-feeding arthropods was previously reported in ticks (Ahantarig et al., 2013; Felsheim et al., 2009; Guizzo et al., 2017; Niebylski et al., 1997; Noda et al., 1997). For one of the tick symbionts, a *Francisella*-like bacterium, an origin from a mammalian pathogen was recently suggested by Gerhart et al., (2016). Based on the data available for the few louse genera examined so far, their microbiomes contain both vertebrate pathogens and typical insect-associated bacteria (e.g., *Arsenophonus*, *Sodalis*).

From a general point of view, the identification of the Neisseriaceae-related symbiotic lineage, phylogenetically distant from the previously known symbionts in Anoplura, supports the view that sucking lice are unique in their ability to repeatedly acquire obligate symbionts from diverse bacterial taxa. This result also demonstrates the need for more detailed high throughput screening of different sucking lice, to properly assess the significance of the two discussed ecological groups of bacteria (i.e., typical insect associates and vertebrate pathogens) as a source of nutritional symbionts.

ACKNOWLEDGEMENTS

This work was supported by the Grant Agency of the Czech Republic (grant GA18-07711S to V.H.). We would like to acknowledge the sequencing services of Genomics Core Facility, EMBL Heidelberg, Heidelberg, Germany and the DNA Services of W.M. Keck Center, University of Illinois at Urbana-Champaign, Illinois, USA. Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures" (CESNET LM2015042), is greatly appreciated. In addition, we thank Joel J. Brown for his language corrections on this manuscript.

AUTHOR CONTRIBUTIONS

Václav Hypša, Eva Nováková and Jana Říhová designed the study. Jana Říhová and Jana Martinů collected the samples. All authors contributed to data analyses. Václav Hypša, Eva Nováková, Jana Říhová, František Vácha and Giampiero Batani wrote the original draft. All authors participated in discussions and revised the manuscript. Václav Hypša supervised the whole project.

DATA AVAILABILITY STATEMENT

The genome assemblies are available from GenBank under the accession numbers CP046107 (closed genome of the symbiont from *H. acanthopus*) and WNLJ00000000 (draft genome of *P. serrata* symbiont in 32 contigs). Amplicon data for the lice microbiomes (demultiplexed, quality filtered and merged) are available at https://www.ebi.ac.uk/ena/data/view/PRJEB35541. OTU abundance tables, and accession numbers for individual samples of *Polyplax serrata* and *Hoplopleura* spp. (sequences of COI genes) are available in Data S1. The alignments for 61-gene matrix and 16S matrix are available from Dryad under doi link https://doi.org/10.5061/dryad.76hdr7ssn. The tree files in newick format are available in Data S6. Genome annotations of Neisseriaceae symbionts (gbk format with protein coding and RNA sequences from RAST and PROKKA servers) are available from Dryad under doi link https://doi.org/10.5061/dryad.j9kd51c9v.

ORCID

Jana Říhová https://orcid.org/0000-0001-7937-1960
Jana Martinů https://orcid.org/0000-0003-0296-7867
Václav Hypša https://orcid.org/0000-0001-5572-782X

REFERENCES

Ahantarig, A., Trinachartvanit, W., Baimai, V., & Grubhoffer, L. (2013). Hard ticks and their bacterial endosymbionts (or would be

pathogens). *Folia Microbiologica*, *58*(5), 419–428. https://doi.org/10.1007/s12223-013-0222-1

Aksoy, S. (2000). Tsetse – A haven for microorganisms. *Parasitology Today*, *16*(3), 114–118. https://doi.org/10.1016/S0169-4758(99)01606-3

Allen, J. M., Burleigh, J. G., Light, J. E., & Reed, D. L. (2016). Effects of 16S rDNA sampling on estimates of the number of endosymbiont lineages in sucking lice. *PeerJ*, *4*, e2187. https://doi.org/10.7717/peerj.2187

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Amann, R., Krumholz, L., & Stahl, D. (1990). Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *Journal of Bacteriology*, *172*(2), 762–770. https://doi.org/10.1128/jb.172.2.762-770.1990

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

Apprill, A., McNally, S., Parsons, R., & Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology*, *75*(2), 129–137. https://doi.org/10.3354/ame01753

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., … Zagnitko, O. (2008). The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics*, *9*(1), 75. https://doi.org/10.1186/1471-2164-9-75

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Baumann, P. (2005). Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annual Review of Microbiology*, *59*, 155–189. https://doi.org/10.1146/annurev.micro.59.030804.121041

Bennett, G. M., & Moran, N. A. (2013). Small, smaller, smallest: The origins and evolution of ancient dual symbioses in a phloem-feeding insect. *Genome Biology and Evolution*, *5*(9), 1675–1688. https://doi.org/10.1093/gbe/evt118

Bennett, G. M., & Moran, N. A. (2015). Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(33), 10169–10176. https://doi.org/10.1073/pnas.1421388112

Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., Mills, D. A., & Caporaso, J. G. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods*, *10*(1), 57–59. https://doi.org/10.1038/nmeth.2276

Boyd, B. M., Allen, J. M., de Crécy-Lagard, V., & Reed, D. L. (2014). Genome sequence of *Candidatus* Riesia pediculischaeffi, endosymbiont of chimpanzee lice, and genomic comparison of recently acquired endosymbionts from human and chimpanzee lice. *G3: Genes, Genomes, Genetics*, *4*(11), 2189–2195. https://doi.org/10.1534/g3.114.012567

Boyd, B. M., Allen, J. M., Koga, R., Fukatsu, T., Sweet, A. D., Johnson, K. P., & Reed, D. L. (2016). Two bacterial genera, *Sodalis* and *Rickettsia*, associated with the seal louse *Proechinophthirus fluctus* (Phthiraptera: Anoplura). *Applied and Environmental Microbiology*, *82*(11), 3185–3197. https://doi.org/10.1128/AEM.00282-16

Boyd, B. M., Allen, J. M., Nguyen, N.-P., Vachaspati, P., Quicksall, Z. S., Warnow, T., Mugisha, L., Johnson, K. P., & Reed, D. L. (2017). Primates, lice and bacteria: Speciation and genome evolution in the symbionts of hominid lice. *Molecular Biology and Evolution*, *34*(7), 1743–1757. https://doi.org/10.1093/molbev/msx117

Brewer, T. E., Albertsen, M., Edwards, A., Kirkegaard, R. H., Rocha, E. P., & Fierer, N. (2020). Unlinked rRNA genes are widespread among bacteria and archaea. *The ISME Journal*, *14*(2), 597–608. https://doi.org/10.1038/s41396-019-0552-3

Brown, J. J., Rodríguez-Ruano, S. M., Poosakkannu, A., Batani, G., Schmidt, J. O., Roachell, W., Zima, J., Hypša, V., & Nováková, E. (2020). Ontogeny, species identity and environment dominate microbiome dynamics in wild populations of kissing bugs (Triatominae). *Microbiome*, *8*, 146. https://doi.org/10.1186/s40168-020-00921-x

Bushnell, B. (2014). *BBMap: A fast, accurate, splice-aware aligner* (No. LBNL-7065E). Lawrence Berkeley National Laboratory, Berkeley.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., … Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*(5), 335. https://doi.org/10.1038/nmeth.f.303

Carattoli, A., Zankari, E., García-Fernández, A., Larsen, M. V., Lund, O., Villa, L., & Hasman, H. (2014). In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrobial Agents and Chemotherapy*, *58*(7), 3895–3903. https://doi.org/10.1128/AAC.02412-14

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, *17*(4), 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334

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*. *Journal of Molecular Evolution*, *48*(1), 49–58. https://doi.org/10.1007/PL00006444

Darling, A. E., Mau, B., & Perna, N. T. (2010). progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*, *5*(6), e11147. https://doi.org/10.1371/journal.pone.0011147

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772. https://doi.org/10.1038/nmeth.2109

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics*, *34*(15), 2666–2669. https://doi.org/10.1093/bioinformatics/bty149

Demchenko, A. P. (2020). Photobleaching of organic fluorophores: Quantitative characterization, mechanisms, protection. *Methods and Applications in Fluorescence*, *8*(2), 022001. https://doi.org/10.1088/2050-6120/ab7365

Dhami, M. K., Buckley, T. R., Beggs, J. R., & Taylor, M. W. (2013). Primary symbiont of the ancient scale insect family Coelostomidiidae exhibits strict cophylogenetic patterns. *Symbiosis*, *61*(2), 77–91. https://doi.org/10.1007/s13199-013-0257-8

Dick, C. W. (2006). *Checklist of world Hippoboscidae (Diptera: Hippoboscoidea)* (pp. 1–7). Department of Zoology, Field Museum Natural History.

Doudoumis, V., Blow, F., Saridaki, A., Augustinos, A., Dyer, N. A., Goodhead, I., Solano, P., Rayaisse, J.-B., Takac, P., Mekonnen, S.,

Parker, A. G., Abd-Alla, A. M. M., Darby, A., Bourtzis, K., & Tsiamis, G. (2017). Challenging the *Wigglesworthia*, *Sodalis*, *Wolbachia* symbiosis dogma in tsetse flies: *Spiroplasma* is present in both laboratory and natural populations. *Scientific Reports*, *7*(1), 1–13. https://doi.org/10.1038/s41598-017-04740-3

Douglas, A. E. (2015). Multiorganismal insects: Diversity and function of resident microorganisms. *Annual Review of Entomology*, *60*, 17–34. https://doi.org/10.1146/annurev-ento-010814-020822

Edgar, R. C. (2013). UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, *10*(10), 996. https://doi.org/10.1038/nmeth.2604

Eggeling, C., Widengren, J., Rigler, R., & Seidel, C. A. M. (1998). Photobleaching of fluorescent dyes under conditions used for single-molecule detection: Evidence of two-step photolysis. *Analytical Chemistry*, *70*(13), 2651–2659. https://doi.org/10.1021/ac980027p

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 1–14. https://doi.org/10.1186/s13059-019-1832-y

Engel, P., & Moran, N. A. (2013). The gut microbiota of insects–diversity in structure and function. *FEMS Microbiology Reviews*, *37*(5), 699–735. https://doi.org/10.1111/1574-6976.12025

Felsheim, R. F., Kurtti, T. J., & Munderloh, U. G. (2009). Genome sequence of the endosymbiont *Rickettsia peacockii* and comparison with virulent *Rickettsia rickettsii*: Identification of virulence factors. *PLoS One*, *4*(12), e8361. https://doi.org/10.1371/journal.pone.0008361

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, *3*(5), 294–299. https://doi.org/10.1128/AEM.00226-09

Fukatsu, T., Hosokawa, T., Koga, R., Nikoh, N., Kato, T., Hayama, S.-I., Takefushi, H., & Tanaka, I. (2009). Intestinal endocellular symbiotic bacterium of the macaque louse *Pedicinus obtusus*: Distinct endosymbiont origins in anthropoid primate lice and the old world monkey louse. *Applied Environmental Microbiology*, *75*(11), 3796–3799. https://doi.org/10.1128/AEM.00226-09

Gauthier, J. P., Outreman, Y., Mieuzet, L., & Simon, J. C. (2015). Bacterial communities associated with host-adapted populations of pea aphids revealed by deep sequencing of 16S ribosomal DNA. *PLoS One*, *10*(3), e0120664. https://doi.org/10.1371/journal.pone.0120664

Gerhart, J. G., Moses, A. S., & Raghavan, R. (2016). A *Francisella*-like endosymbiont in the Gulf Coast tick evolved from a mammalian pathogen. *Scientific Reports*, *6*, 33670. https://doi.org/10.1038/srep33670

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, *52*(5), 696–704. https://doi.org/10.1080/10635150390235520

Guizzo, M. G., Parizi, L. F., Nunes, R. D., Schama, R., Albano, R. M., Tirloni, L., Oldiges, D. P., Vieira, R. P., Oliveira, W. H. C., Leite, M. D. S., Gonzales, S. A., Farber, M., Martins, O., Vaz, I. D. S., & Oliveira, P. L. (2017). A *Coxiella* mutualist symbiont is essential to the development of *Rhipicephalus microplus*. *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-17309-x

Hafner, M. S., Sudman, P. D., Villablanca, F. X., Spradling, T. A., Demastes, J. W., & Nadler, S. A. (1994). Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science*, *265*(5175), 1087–1090. https://doi.org/10.1126/science.8066445

Hosokawa, T., Nikoh, N., Koga, R., Satô, M., Tanahashi, M., Meng, X. Y., & Fukatsu, T. (2012). Reductive genome evolution, host–symbiont cospeciation and uterine transmission of endosymbiotic bacteria in bat flies. *The ISME Journal*, *6*(3), 577–587. https://doi.org/10.1038/ismej.2011.125

Husnik, F., Nikoh, N., Koga, R., Ross, L., Duncan, R. P., Fujie, M., Tanaka, M., Satoh, N., Bachtrog, D., Wilson, A. C. C., von Dohlen, C. D., Fukatsu, T., & McCutcheon, J. P. (2013). Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell*, *153*(7), 1567–1578. https://doi.org/10.1016/j.cell.2013.05.040

Hypša, V., & Křížek, J. (2007). Molecular evidence for polyphyletic origin of the primary symbionts of sucking lice (Phthiraptera, Anoplura). *Microbial Ecology*, *54*(2), 242–251. https://doi.org/10.1007/s00248-006-9194-x

Jones, R. T., McCormick, K. F., & Martin, A. P. (2008). Bacterial communities of *Bartonella*-positive fleas: Diversity and community assembly patterns. *Applied and Environmental Microbiology*, *74*(5), 1667–1670. https://doi.org/10.1128/AEM.02090-07

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, *44*(D1), D457–D462. https://doi.org/10.1093/nar/gkv1070

Katoh, K., Misawa, K., Kuma, K. I., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kirkness, E. F., Haas, B. J., Sun, W., Braig, H. R., Perotti, M. A., Clark, J. M., Lee, S. H., Robertson, H. M., Kennedy, R. C., Elhaik, E., Gerlach, D., Kriventseva, E. V., Elsik, C. G., Graur, D., Hill, C. A., Veenstra, J. A., Walenz, B., Tubio, J. M. C., Ribeiro, J. M. C., … Pittendrigh, B. R. (2010). Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proceedings of the National Academy of Sciences*, *107*(27), 12168–12173. https://doi.org/10.1073/pnas.1003379107

Klein, A., Schrader, L., Gil, R., Manzano-Marín, A., Flórez, L., Wheeler, D., Werren, J. H., Latorre, A., Heinze, J., Kaltenpoth, M., Moya, A., & Oettler, J. (2016). A novel intracellular mutualistic bacterium in the invasive ant *Cardiocondyla obscurior*. *The ISME Journal*, *10*(2), 376–388. https://doi.org/10.1038/ismej.2015.119

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, *37*(5), 540–546. https://doi.org/10.1038/s41587-019-0072-8

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, *27*(5), 722–736. https://doi.org/10.1101/gr.215087.116.

Kwong, W. K., & Moran, N. A. (2013). Cultivation and characterization of the gut symbionts of honey bees and bumble bees: Description of *Snodgrassella alvi* gen. nov., sp. nov., a member of the family Neisseriaceae of the Betaproteobacteria, and *Gilliamella apicola* gen. nov., sp. nov., a member of Orbaceae fam. nov., Orbales ord. nov., a sister taxon to the order 'Enterobacteriales' of the Gammaproteobacteria. *International Journal of Systematic and Evolutionary Microbiology*, *63*(6), 2008–2018. https://doi.org/10.1099/ijs.0.044875-0

Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research*, *6*, 1287. https://doi.org/10.12688/f1000research.12232.1

Lefort, V., Longueville, J. E., & Gascuel, O. (2017). SMS: Smart model selection in PhyML. *Molecular Biology and Evolution*, *34*(9), 2422–2424. https://doi.org/10.1093/molbev/msx149

Leinonen, R., Sugawara, H., & Shumway, M. (2011). The sequence read archive. *Nucleic Acids Research*, *39*, D19–D21. https://doi.org/10.1093/nar/gkq1019

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Light, J. E., Smith, V. S., Allen, J. M., Durden, L. A., & Reed, D. L. (2010). Evolutionary history of mammalian sucking lice (Phthiraptera: Anoplura). *BMC Evolutionary Biology*, *10*(1), 292. https://doi.org/10.1186/1471-2148-10-292

Mahmoudian, J., Hadavi, R., Jeddi-Tehrani, M., Mahmoudi, A. R., Bayat, A. A., Shaban, E., & Ghods, R. (2011). Comparison of the photobleaching and photostability traits of Alexa Fluor 568- and fluorescein isothiocyanate- conjugated antibody. *Cell Journal*, *13*(3), 169–172.

Manzano-Marín, A., Coeur d'acier, A., Clamens, A.-L., Orvain, C., Cruaud, C., Barbe, V., & Jousselin, E. (2020). Serial horizontal transfer of vitamin-biosynthetic genes enables the establishment of new nutritional symbionts in aphids' di-symbiotic systems. *The ISME Journal*, *14*(1), 259–273. https://doi.org/10.1038/s41396-019-0533-6

Manzano-Marín, A., Szabó, G., Simon, J. C., Horn, M., & Latorre, A. (2017). Happens in the best of subfamilies: Establishment and repeated replacements of co-obligate secondary endosymbionts within Lachninae aphids. *Environmental Microbiology*, *19*(1), 393–408. https://doi.org/10.1111/1462-2920.13633

Martinson, V. G., Moy, J., & Moran, N. A. (2012). Establishment of characteristic gut bacteria during development of the honeybee worker. *Applied and Environmental Microbiology*, *78*(8), 2830–2840. https://doi.org/10.1128/AEM.07810-11

Martinů, J., Hypša, V., & Štefka, J. (2018). Host specificity driving genetic structure and diversity in ectoparasite populations: Coevolutionary patterns in *Apodemus* mice and their lice. *Ecology and Evolution*, *8*(20), 10008–10022. https://doi.org/10.1002/ece3.4424

McCutcheon, J. P., Boyd, B. M., & Dale, C. (2019). The life of an insect endosymbiont from the cradle to the grave. *Current Biology*, *29*(11), 485–495. https://doi.org/10.1016/j.cub.2019.03.032

Meseguer, A. S., Manzano-Marín, A., Coeur D'Acier, A., Clamens, A. L., Godefroid, M., & Jousselin, E. (2017). *Buchnera* has changed flatmate but the repeated replacement of co-obligate symbionts is not associated with the ecological expansions of their aphid hosts. *Molecular Ecology*, *26*(8), 2363–2378. https://doi.org/10.1111/mec.13910

Moran, N. A. (1996). Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences*, *93*(7), 2873–2878. https://doi.org/10.1073/pnas.93.7.2873

Moran, N. A., Munson, M. A., Baumann, P., & Ishikawa, H. (1993). A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *253*(1337), 167–171. https://doi.org/10.1098/rspb.1993.0098

Munson, M. A., Baumann, L., & Baumann, P. (1993). *Buchnera aphidicola* (a prokaryotic endosymbiont of aphids) contains a putative 16S rRNA operon unlinked to the 23S rRNA-encoding gene: Sequence determination, and promoter and terminator analysis. *Gene*, *137*(2), 171–178. https://doi.org/10.1016/0378-1119(93)90003-L

Niebylski, M. L., Peacock, M. G., Fischer, E. R., Porcella, S. F., & Schwan, T. G. (1997). Characterization of an endosymbiont infecting wood ticks, *Dermacentor andersoni*, as a member of the genus *Francisella*. *Applied and Environmental Microbiology*, *63*(10), 3933–3940. https://doi.org/10.1128/AEM.63.10.3933-3940.1997

Noda, H., Munderloh, U. G., & Kurtti, T. J. (1997). Endosymbionts of ticks and their relationship to *Wolbachia* spp. and tick-borne pathogens of humans and animals. *Applied and Environmental Microbiology*, *63*(10), 3926–3932. https://doi.org/10.1128/AEM.63.10.3926-3932.1997

Nováková, E., Hypša, V., & Moran, N. A. (2009). *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. *BMC Microbiology*, *9*(1), 143. https://doi.org/10.1186/1471-2180-9-143

oksanOteo, J. A., Portillo, A., Portero, F., Zavala-Castro, J., Venzal, J. M., & Labruna, M. B. (2014). '*Candidatus* Rickettsia asemboensis' and *Wolbachia* spp. in *Ctenocephalides felis* and *Pulex irritans* fleas removed from dogs in Ecuador. *Parasites & Vectors*, *7*(1), 455. https://doi.org/10.1186/s13071-014-0455-0

Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, *18*(5), 1403–1414. https://doi.org/10.1111/1462-2920.13023

Perotti, M. A., Kirkness, E. F., Reed, D. L., Braig, H. R., Bourtzis, K., & Miller, T. A. (2008). Endosymbionts of lice. In T. A. Miller (Ed.), *Insect symbiosis* (Vol. *3*, pp. 205–220). CRC Press.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590–D596. https://doi.org/10.1093/nar/gks1219

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Říhová, J., Nováková, E., Husník, F., & Hypša, V. (2017). *Legionella* becoming a mutualist: Adaptive processes shaping the genome of symbiont in the louse *Polyplax serrata*. *Genome Biology and Evolution*, *9*(11), 2946–2957. https://doi.org/10.1093/gbe/evx217

Rodríguez-Ruano, S. M., Juhaňáková, E., Vávra, J., & Nováková, E. (2020). Methodological insight into mosquito microbiome studies. *Frontiers in Cellular and Infection Microbiology*, *10*, 86. https://doi.org/10.3389/fcimb.2020.00086

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*(3), 539–542. https://doi.org/10.1093/sysbio/sys029

Sauer, C., Stackebrandt, E., Gadau, J., Hölldobler, B., & Gross, R. (2000). Systematic relationships and cospeciation of bacterial endosymbionts and their carpenter ant host species: Proposal of the new taxon *Candidatus* Blochmannia gen. nov. *International Journal of Systematic and Evolutionary Microbiology*, *50*(5), 1877–1886. https://doi.org/10.1099/00207713-50-5-1877

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682. https://doi.org/10.1038/nmeth.2019

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. https://doi.org/10.1038/nmeth.2089

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Šochová, E., Husník, F., Nováková, E., Halajian, A., & Hypša, V. (2017). *Arsenophonus* and *Sodalis* replacements shape evolution of symbiosis in louse flies. *PeerJ*, *5*, e4099. https://doi.org/10.7717/peerj.4099

Syberg-Olsen, M., Garber, A., Keeling, P., McCutcheon, J., & Husník, F. (2020). *Pseudofinder*, *GitHub repository*. Retrieved from https://github.com/filip-husnik/pseudofinder/

Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, *27*(5), 737–746. https://doi.org/10.1101/gr.214270.116

Walters, W., Hyde, E. R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., Gilbert, J. A., Jansson, J. K., Caporaso, J. G., Fuhrman, J. A., Apprill, A., & Knight, R. (2016). Improved bacterial 16S rRNA gene (V4 and V4–5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *mSystems*, *1*(1), e00009–15. https://doi.org/10.1128/mSystems.00009-15

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, *73*(16), 5261–5267. https://doi.org/10.1128/AEM.00062-07

Waterworth, S. C., Flórez, L. V., Rees, E. R., Hertweck, C., Kaltenpoth, M., & Kwan, J. C. (2020). Horizontal gene transfer to a defensive symbiont with a reduced genome in a multipartite beetle microbiome. *MBio*, *11*(1), e02430-19. https://doi.org/10.1128/mBio.02430-19

Yilmaz, L. S., Parnerkar, S., & Noguera, D. R. (2011). mathFISH, a web tool that uses thermodynamics-based mathematical models for in silico evaluation of oligonucleotide probes for fluorescence in situ hybridization. *Applied and Environmental Microbiology*, *77*(3), 1118–1122. https://doi.org/10.1128/AEM.01733-10

Yun, J.-H., Roh, S. W., Whon, T. W., Jung, M.-J., Kim, M.-S., Park, D.-S., Yoon, C., Nam, Y.-D., Kim, Y.-J., Choi, J.-H., Kim, J.-Y., Shin, N.-R., Kim, S.-H., Lee, W.-J., & Bae, J.-W. (2014). Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host. *Applied and Environmental Microbiology*, *80*(17), 5254–5264. https://doi.org/10.1128/AEM.01226-14

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Říhová J, Batani G, Rodríguez-Ruano SM, et al. A new symbiotic lineage related to *Neisseria* and *Snodgrassella* arises from the dynamic and diverse microbiomes in sucking lice. *Mol Ecol*. 2021;30:2178–2196. https://doi.org/10.1111/mec.15866

MANUSCRIPT #2

Lightella neohaematopini: A new lineage of highly reduced endosymbionts coevolving with chipmunk lice of the genus *Neohaematopinus*

Frontiers in Microbiology (2022), *13*: 900312
https://doi.org/10.3389/fmicb.2022.900312

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 01 August 2022
DOI 10.3389/fmicb.2022.900312

OPEN ACCESS

EDITED BY
Chih-Horng Kuo,
Institute of Plant and Microbial Biology, Academia Sinica, Taiwan

REVIEWED BY
Daniel Tamarit,
Wageningen University and Research, Netherlands
Jennah Elizabeth Dharamshi,
Stockholm University, Sweden

*CORRESPONDENCE
Václav Hypša
vacatko@prf.jcu.cz

SPECIALTY SECTION
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

RECEIVED 20 March 2022
ACCEPTED 07 July 2022
PUBLISHED 01 August 2022

CITATION
Říhová J, Bell KC, Nováková E and Hypša V (2022) *Lightella neohaematopini*: A new lineage of highly reduced endosymbionts coevolving with chipmunk lice of the genus *Neohaematopinus*. *Front. Microbiol.* 13:900312. doi: 10.3389/fmicb.2022.900312

Lightella neohaematopini: A new lineage of highly reduced endosymbionts coevolving with chipmunk lice of the genus *Neohaematopinus*

Jana Říhová[1], Kayce C. Bell[2,3,4], Eva Nováková[1,5] and Václav Hypša[1,5]*

[1]Department of Parasitology, Faculty of Science, University of South Bohemia, České Budějovice, Czechia, [2]Department of Mammalogy, Natural History Museum of Los Angeles County, Los Angeles, CA, United States, [3]Department of Biology, Museum of Southwestern Biology, University of New Mexico, Albuquerque, NM, United States, [4]Department of Zoology, Denver Museum of Nature and Science, Denver, CO, United States, [5]Institute of Parasitology, Biology Centre, ASCR, v.v.i., České Budějovice, Czechia

Sucking lice (Anoplura) are known to have established symbiotic associations multiple times with different groups of bacteria as diverse as Enterobacteriales, Legionellales, and Neisseriales. This diversity, together with absence of a common coevolving symbiont (such as *Buchnera*, in aphids), indicates that sucking lice underwent a series of symbiont acquisitions, losses, and replacements. To better understand evolution and significance of louse symbionts, genomic and phylogenetic data are needed from a broader taxonomic diversity of lice and their symbiotic bacteria. In this study, we extend the known spectrum of the louse symbionts with a new lineage associated with *Neohaematopinus pacificus*, a louse species that commonly parasitizes North American chipmunks. The recent coevolutionary analysis showed that rather than a single species, these lice form a cluster of unique phylogenetic lineages specific to separate chipmunk species (or group of closely related species). Using metagenomic assemblies, we show that the lice harbor a bacterium which mirrors their phylogeny and displays traits typical for obligate mutualists. Phylogenetic analyses place this bacterium within Enterobacteriaceae on a long branch related to another louse symbiont, "*Candidatus* Puchtella pedicinophila." We propose for this symbiotic lineage the name "*Candidatus* Lightella neohaematopini." Based on the reconstruction of metabolic pathways, we suggest that like other louse symbionts, *L. neohaematopini* provides its host with at least some B vitamins. In addition, several samples harbored another symbiotic bacterium phylogenetically affiliated with the Neisseriales-related symbionts described

previously from the lice *Polyplax serrata* and *Hoplopleura acanthopus*. Characterizing these bacteria further extend the known diversity of the symbiotic associations in lice and show unique complexity and dynamics of the system.

KEYWORDS

genome evolution, insect symbionts, lice, symbiosis, coevolution, *Neohaematopinus pacificus*

Introduction

Establishment of an obligate mutualistic symbiosis with nutrient-providing bacteria is an important evolutionary event that allows some insect groups to exploit new ecological niches (Sudakaran et al., 2017). Due to mutual dependence, such host-symbiont systems often undergo long-term coevolution manifested by mutually mirrored phylogenies between the host and the symbiont (Moran et al., 1993; Chen et al., 1999; Sauer et al., 2000; Dhami et al., 2013). This type of essential nutritional mutualist is usually called a primary symbiont (P-symbiont) characterized by several typical features, such as genome reduction, low GC content, and genomic deterioration of metabolic capacities (Toft and Andersson, 2010). In several insect groups, particularly those feeding on plant xylem or phloem sap, coevolution of the symbiont with the host can reach hundreds of millions of years into the past (Moran et al., 1993; Spaulding and von Dohlen, 1998; Sauer et al., 2000). In contrast, secondary symbionts (S-symbionts) are symbiotic bacteria with diverse phenotypes, not essential for their hosts (Stouthamer et al., 1999; Zchori-Fein and Perlman, 2004; Moran et al., 2005; Toh et al., 2006; Nováková et al., 2009), which are usually acquired more recently and may accompany P-symbionts in a particular taxon, or even a population of the host (McCutcheon et al., 2019).

In sucking lice of the order Anoplura, several obligate symbiotic bacteria have been described, generally thought to provide the hosts with compounds missing in their diets (Hypsa and Krizek, 2007; Fukatsu et al., 2009; Boyd et al., 2014, 2016; Allen et al., 2016; Rihova et al., 2021). Although all extant Anoplura share similar lifestyles and diets, no common ancient P-symbiont has been maintained through all lineages. Instead, several groups of sucking lice have acquired their symbionts independently from bacterial taxa as diverse as Enterobacteriales (Fukatsu et al., 2009; Boyd et al., 2014, 2016), Legionellales (Rihova et al., 2017), and Neisseriales (Rihova et al., 2021). While from the deep phylogenetic perspective the symbionts of sucking lice seem to form a diverse assemblage of different bacteria (Allen et al., 2016), at the more recent phylogenetic level some symbiotic lineages show tendencies to coevolve with their host for a limited period, and display features common for P-symbionts. These are, for example, the *Puchtella* lineages described from closely related species of *Pedicinus* (Boyd et al., 2017) and the *Legionella polyplacis* known at least from two *Polyplax* host species, *P. serrata* and *P. spinulosa* (Hypsa and Krizek, 2007). Some other louse symbionts seem to have been acquired more recently, and their genomes appear to have undergone a transition from S-symbionts toward P-symbionts (Boyd et al., 2016; Rihova et al., 2021). This ambiguous mode of symbiosis, i.e., short periods of coevolution intermixed with frequent losses and acquisitions, raises a question on the significance of the symbionts for their louse hosts. While considered obligate mutualists, essential for the host survival and reproduction, the exact role of the louse symbionts is not clear. Seventy-year-old experimental evidence obtained for *Pediculus humanus* suggests that this louse may depend on provisioning of most B vitamins by its symbiont (Puchta, 1955). However, genomic comparison of the louse symbionts for which complete genomes are available shows considerable inconsistency in preservation of their metabolic pathways, including those for B vitamins (Rihova et al., 2021). To better understand the evolution and significance of louse-symbiont systems, much broader taxonomic diversity has to be examined using genomic and phylogenomic tools.

In this study, we extend the current list of the investigated louse-symbiont systems by taking advantage of the recently described codiversification between several chipmunk species (Rodentia: Sciuridae: *Tamias*) and their lice of the genus *Neohaematopinus* (Bell et al., 2021). These lice are primarily parasites of sciurid rodents in Asia and North and Central America, where one species, *N. pacificus*, was described from western North American chipmunks in the subgenus *Neotamias* (Bell et al., 2015, 2021). Coevolutionary analysis between the chipmunks and lice indicates that rather than a single polyxenic species, the lice classified within *N. pacificus* form a cluster of separated phylogenetic lineages each specific to a single chipmunk species or group of closely related species (Bell et al., 2021). Utilizing metagenomic data and phylogenetic background from this coevolutionary study, we identify a new putatively obligate and mutualistic symbiont of lice, show its coevolution with the louse hosts, and reconstruct its metabolic capacity related to its likely symbiotic role.

Materials and methods

Screening and assembly of chipmunk lice metagenomic data

Metagenomic data from 21 specimens of *Neohaematopinus* lice was downloaded from the NCBI SRA database (Sequence Read Archive; Leinonen et al., 2011; **Table 1**). Initially, we employed a comprehensive search for bacteria by classifying metagenomic reads from each sample using the GOTTCHA2-v.2.1.7 signature-based metagenomic taxonomic profiling tool (Freitas et al., 2015) implemented in KBase (Arkin et al., 2018) with database built from NCBI RefSeq Release 90. The full taxonomic output reports were summarized at the strain and genus level in R with package phyloseq (McMurdie and Holmes, 2013). All the bacterial strains with relative abundances above 0.1% in any sample were considered.

Assemblies of the metagenomic data were generated using SPAdes v.3.13.0 (Bankevich et al., 2012) with the option, meta. The CONCOCT v.1.1 software (Alneberg et al., 2014) implemented in Kbase (kb_concoct 1.3.4) was ran with default parameters, using Bowtie2 (Langmead and Salzberg, 2012) as a default read mapper, to organize the contigs into putative metagenome-assembled genomes (MAGs) based on their nucleotide composition and depth of coverage. Metagenome-assembled genomes were classified into bacterial taxa using GTDB-Tk (Chaumeil et al., 2020) against the Genome Taxonomy Database (GTDB) v.R06-RS202 (258,406 bacterial and archaeal genomes) based on average nucleotide identity with a reference genome, the genome's position in the reference tree, or relative evolutionary divergence and placement of the genome in the reference. The GOTTCHA read classification and GTDB-Tk classification of MAGs (**Supplementary Table 1**) both indicated the presence of a prospective obligate endosymbiont in majority of the samples. In the GOTTCHA classification (**Supplementary Table 1**), a number of reads were assigned to the taxa *Buchnera*, *Baumannia*, and *Evansia*. GTDB-Tk classification of MAGs revealed several Enterobacteriaceae, including *Puchtella* sp. str. PRUG, endosymbiont of the louse *Pedicinus baddi* (**Table 1**). Among the other assignments, a potentially interesting taxon in respect to endosymbiotic interactions, was the Neisseriaceae-related bacterium (see "Results and discussion" section).

Refining metagenome-assembled genomes and completing genome drafts

To refine Enterobacteriaceae and Neisseriaceae MAGs, i.e., to detect potentially missed contigs and to remove the incorrectly binned ones, we used the following procedure (scheme provided in **Supplementary Figure 1**). The Enterobacteriaceae MAGs were present in twelve louse metagenomic assemblies, in five of them by only few long contigs (**Table 1**). For these twelve meta-assemblies, we filtered all putative Enterobacteriaceae contigs based on their GC content and coverage. For each assembly, these values were inferred from the contigs present in its Enterobacteriaceae MAG. From these extended bins, we then selected subsets of all prokaryotic contigs based on the ORFs density as predicted in Geneious Prime v.2020.2.5 (Kearse et al., 2012). To confirm that this approach detected all contigs belonging to the Enterobacteriaceae symbiont, we performed an alternative BLAST-based screening. We prepared databases from all 21 MAGs and screened them by BLASTn search (set to default E-value 10.0 and one best hit) using all genes from *Puchtella* sp. str. PRUG (NZ_LKAS01000001.1 + NZ_LKAS01000002.1) and two close relatives *Wiglessworthia glossinidia* (BA000021.3; NC_016893.1) and *Blochmannia pensylvaticus* str. BPEN (NC_007292.1) as queries. In addition, we used annotated genes from the best draft (the sample SRR12483207; for annotation procedure see below) as queries to screen the remaining 20 meta-assemblies by BLASTx searches (Camacho et al., 2009) set to default E-value (10.0) and one best hit. For all the contigs retrieved by these approaches we checked their taxonomical origin by BLASTn search (set to default E-value 10.0 and three best hits) against the nucleotide (nt) database. The Neisseriaceae-related MAG was present in a single meta-assembly (SRR12483206) and was largely fragmented. To refine this MAG, we used the same approach as described above for the Enterobacteriaceae. To screen all other meta-assemblies for possible presence of this bacterium, we applied the BLAST-based approach with the genes of the two Neisseriaceae-related louse symbionts (CP046107 and WNLJ00000000) as queries.

For the putative taxonomic assignment of each gene in the least fragmented genome of the Enterobacteriaceae symbiont (L207) we employed BLASTx search against the nr database set to three best hits (**Supplementary Table 2**). The completeness of the L207 draft was assessed by BUSCO v.4 (Simao et al., 2015) and for the Enterobacteriaceae symbiont also by the size convergence (see "Results and discussion" section). All genome drafts were annotated by PROKKA (Seemann, 2014) and RAST (Aziz et al., 2008) and deposited in GenBank under the BioProject accession number PRJNA813538 for the twelve most complete assemblies of Enterobacteriaceae symbiont (L222, L469, L220, L207, L201, L221, L218, L208, L2017, L209, L214, and L204) and the two most complete assemblies of Neisseriaceae-related symbiont (N206 and N204). To detect candidate pseudogenes in the annotated L207 and N206 assemblies, we used bioinformatic tool Pseudofinder (Syberg-Olsen et al., 2021).

TABLE 1 Overview of metagenomic binning and taxonomy assignment.

SRR data code	Number of paired reads	Read length (nt)	Bin count	GTDB classified MAG bins	Assigned taxonomy: family; genus	Bin total length (nt) and contig count
SRR5088469	154952566	100	18	bin.004.fa	Enterobacteriaceae	458733 (5)
				bin.006.fa	Burkholderiaceae; *Variovorax*	1723338 (455)
				bin.009.fa	Burkholderiaceae; *Burkholderia*	7009440 (797)
				bin.015.fa	Hyphomicrobiaceae; *Hyphomicrobium*	422320 (157)
				bin.016.fa	Rhizobiaceae; *Mesorhizobium*	9399685 (16661)
				bin.017.fa	Sphingomonadaceae; *Sphingomonas*	4635438 (84)
SRR12483222	69397816	160	116	bin.022.fa	Enterobacteriaceae	464809 (14)
SRR12483221	34725622	160	40	bin.015.fa	Enterobacteriaceae	459627 (5)
				bin.024.fa	Xanthobacteraceae; *Bradyrhizobium*	7557328 (558)
SRR12483220	32486852	160	121	bin.044.fa	Enterobacteriaceae	460455 (6)
SRR12483219	55847912	160	156	none	NA	NA
SRR12483218	80798234	160	105	bin.012.fa	Enterobacteriaceae; *Puchtella*	440570 (34)
SRR12483217	27859822	160	72	bin.059.fa	Lactobacillaceae; *Lactobacillus*	254633 (74)
				bin.061.fa	Enterobacteriaceae; *Puchtella*	378565 (47)
SRR12483215	23212912	160	104	none	NA	NA
SRR12483214	27929696	160	166	bin.058.fa	Enterobacteriaceae; *Wigglessworthia*	202321 (55)
SRR12483213	41335766	160	47	bin.014.fa	Lactobacillaceae; *Lactobacillus*	500811 (86)
				bin.041.fa	Xanthobacteraceae; *Bradyrhizobium*	6871847 (970)
SRR12483212	27333640	160	96	none	NA	NA
SRR12483211	21275976	160	105	none	NA	NA
SRR12483210	46206712	160	130	none	NA	NA
SRR12483209	29071184	160	136	bin.012.fa	Enterobacteriaceae; *Puchtella*	317573 (72)
SRR12483208	38144898	160	107	bin.072.fa	Enterobacteriaceae; *Puchtella*	462982 (6)
SRR12483207	27742464	160	10	bin.019.fa	Enterobacteriaceae	449562 (3)
SRR12483206	38930820	160	60	bin.014.fa	Xanthobacteraceae; *Bradyrhizobium*	2938854 (759)
				bin.043.fa	Neisseriaceae	1148191 (218)
SRR12483204	71767598	160	78	bin.015.fa	Enterobacteriaceae	262 958 (53)
SRR12483203	20092028	160	43	none	NA	NA
SRR12483202	60251848	160	118	none	NA	NA
SRR12483201	34764372	160	92	bin.087.fa	Enterobacteriaceae	465231 (6)

Phylogenetic analyses

Phylogenetic position of the new Enterobacteriaceae symbiont

Because preliminary analyses suggested phylogenetic position of the dominant symbiont within Enterobacteriales, we downloaded a representative set of available proteomes for each major lineage of Enterobacteriales from NCBI and JGI databases (proteome accession numbers provided in **Supplementary Table 3A**) including the putatively closest relatives indicated by the binning process and BLASTp against NCBI database (*Puchtella* and *Wigglessworthia*). Several gammaproteobacteria representing other orders were used as outgroups: *Vibrio cholerae* O1 biovar eltor str. N16961 (Vibrionales), *Pseudomonas aeruginosa* PAO1 (Pseudomonadales), *Candidatus* Evansia muelleri CEM1.1 (Oceanospirillales), and *Xanthomonas citri* pv. *vignicola* strain CFBP7111 (Xanthomonadales). BLASTp searches with default parameters (E-value 10.0) and *Salmonella enterica* (accession NZ_CP065718) protein queries for 14 orthologs from Husnik et al. (2011) were used to retrieve sufficiently long and reliably aligned sequences (list of the orthologs provided in **Supplementary Table 3B**) from all the proteomes. For

the new Enterobacteriaceae symbiont, we included two of the least fragmented assemblies (L207 and L221), for which we identified the corresponding orthologs by BLASTp (with default E-value 10.0) and verified these by visual inspection and by the gene's annotations. The single-gene amino-acid matrices were aligned by MAFFT v.7.450 (Katoh et al., 2002) using the E-INS-i setting implemented in Geneious Prime v.2020.2.5 (Kearse et al., 2012), visually inspected, and the genes concatenated. Ambiguously aligned regions were removed by Gblocks (Castresana, 2000) using the less stringent option. The concatenated matrix ("phylogenetic matrix") was analyzed using maximum likelihood (ML) and Bayesian inference (BI) for the phylogenetic reconstruction. The resulting matrix consisted of 7,438 amino acids. The ML tree was inferred using PhyML v.3.3 (Guindon et al., 2009) implemented in Geneious Prime with 100 bootstrap replicates and run under the best fitting CpREV + G + I evolutionary model selected by Akaike criterion (AIC) by smart model selection (SMS) algorithm (Lefort et al., 2017) implemented in the PhyML v.3.0 web server (Guindon et al., 2010). The BI analyses were performed using two different approaches. First, MrBayes v.3.2.6 (Ronquist et al., 2012) was run with four chains for 2,000,000 generations under the CpREV + G + I model. The chain convergence was evaluated in Tracer v.1.6 (Rambaut et al., 2018) and by the standard deviation of split (<0.01) and PSRF+ (reached the value 1.0). Since the *Neohaematopinus* symbionts formed a very long branch, placed within a cluster of other long-branched symbiotic bacteria, we also used PhyloBayes MPI v.1.8 (Lartillot et al., 2013) with CAT-GTR model to minimize possible artifacts due to the aberrant character of the sequences. The analysis was run for 50,000 generations and the quality of convergence was evaluated by the bpcom and tracecomp assessments. The list of bipartition differences indicated that while for most bipartitions the parameter dropped below 0.1, for two bipartitions it remained above 0.3. These problematic bipartitions included taxa distant from our focus, particularly the species of *Erwinia*. To improve the convergence and verify correctness of the topology, we prepared a "reduced phylogenetic matrix" containing 68 taxa (designated in **Supplementary Table 3A**) and retained the matrix of 7,438 amino acids. This matrix was analyzed in PhyloBayes with the same parameters (CAT-GTR, 50,000 generations), resulting in a drop of the maxdiff parameter to 0.057.

Host and the Enterobacteriaceae symbiont coevolution

To assess congruence between the host and the symbiont, we prepared host and Enterobacteriaceae symbiont matrices restricted to the datasets which provided sufficient data for the symbiont (12 host/symbiont samples sharing five orthologs; **Supplementary Table 3C**). Due to different characters and sizes of the matrices, the host and symbiont phylogenies were computed by different methods. The host phylogeny was reconstructed by two different methods. The first method employed ML analysis of 1,107 concatenated nuclear loci from Bell et al. (2021). The reconstruction was performed in IQ-TREE v2.0 (Minh et al., 2020) with the best model (GTR + F + R10) determined by ModelFinder (Kalyaanamoorthy et al., 2017) and support assessed by 1,000 ultrafast bootstrap replicates (Hoang et al., 2018), as implemented in IQ-TREE. In the second approach, we used IQ-TREE v2.0 to generate individual gene trees for the 1,107 nuclear loci and based on these trees, we then estimated a species tree in ASTRALIII v5.7.3 (Zhang et al., 2018), with local posterior probabilities (Sayyari and Mirarab, 2016). For the assemblies of the 12 symbionts, we concatenated five shared protein single-copy orthologs (COGs) determined by Orthofinder (Emms and Kelly, 2019) to build a "coevolutionary matrix" (list of used COGs provided in **Supplementary Table 3D**). The sequences were aligned in MAFFT v.7.450 with E-INS-i settings and processed by Gblocks using the less stringent options. The resulting matrix consisted of 1,150 amino acids. For the phylogenetic reconstructions we used ML and BI methods. The ML tree was reconstructed with 100 bootstrap replicates using the web-based PhyML with best fitting model determined by AIC using SMS function of the online PhyML server v.3.0 (HIVb + G + F). The BI trees were inferred using MrBayes v.3.2.7 under JTT + G + F evolutionary model selected by jModelTest2 v.2.1.10 (Darriba et al., 2012). Four chains were run for 10,000,000 generations and chain convergences were checked in Tracer v.1.6 and evaluated by standard deviation split (<0.01) and PSRF+ (reached the value 1.0). Both the host and the symbiont trees were rooted to fit the *N. pacificus* topologies published in Bell et al. (2021). To evaluate the possible effect of gene number on the symbiont phylogeny (number of single-gene orthologs shared by all samples shown in **Supplementary Table 3C**), we also analyzed by ML, a taxonomically reduced matrix (11 symbiont samples), containing fifty shared protein single-copy orthologs (**Supplementary Table 3D**) determined by OrthoFinder. The sequences were aligned in MAFFT v.7.450 with E-INS-i settings and processed by Gblocks using the less stringent options. The resulting alignment included 6,379 amino acids. The ML tree was reconstructed with 100 bootstrap replicates using the web-based PhyML with the best fitting model determined by AIC using SMS function in the online PhyML server v.3.0 (Q.bird + G + F).

Phylogenetic position of the Neisseriaceae-related symbiont

To determine the phylogenetic position of the bacterium binned using CONCOCT and classified as a member of Neisseriaceae, we prepared a dataset of Neisseriales

proteomes available from the NCBI database, including the putatively closest relatives indicated by BLASTp against NCBI database (Neisseriaceae-related symbionts of lice). We used representatives from other bacterial orders as outgroups, two betaproteobacteria from the order Burkholderiales, *Acidovorax* sp. KKS102 and *Burkholderia cepacia*, one alphaproteobacterium *Rhizobium leguminosarum*, and one gammaproteobacterium, *Legionella pneumophila* subsp. *pneumophila* str. Philadelphia 1. Accession numbers of the proteomes are provided in **Supplementary Table 3E**. For the Neisseriaceae-related symbionts, we included sequences from the two most complete assemblies (N206 and N204). A concatenated amino acid matrix ("Neisseriales matrix") was constructed for 30 single-gene orthologs (COGs) identified by OrthoFinder v.2.4.0 and processed in the same manner as described above for Enterobacteriaceae symbiont matrices (list of COGs provided in **Supplementary Table 3F**). The resulting matrix consisted of 4,465 amino acids. Phylogenetic reconstructions were performed by ML and BI methods. We used PhyML for ML reconstruction using 100 bootstrap replicates and the LG + G + I + F evolutionary model as determined by AIC using SMS algorithm in the online PhyML server v.3.0. The selected LG + G + I + F evolutionary model was also used for BI tree reconstruction in MrBayes v.3.2.7. The BI analysis was run in four chains for 10,000,000 generations and the convergences of the chains were checked in Tracer v.1.6 and evaluated by standard deviation split (<0.01) and PSRF+ (reached the value 1.0).

Genomes and metabolic pathways comparison

To compare the genome structure and reveal the possible rearrangements, we performed two synteny analyses. The first comparison included concatenated contigs representing five genome drafts of the new Enterobacteriaceae symbiont. The samples were selected to maximize two parameters, coverage of the symbiont's phylogenetic diversity and high completeness of the genome draft. Since several contigs contained only a single rRNA gene, and their position could not be reliably assessed, they were not included in the set. The second analysis compared these five genomes with the closest relative, i.e., *Puchtella* sp. str. PRUG (NZ_LKAS01000001.1 + NZ_LKAS01000002.1). The analyses were carried out using two programs, Mauve (Darling et al., 2010) and Clinker (Gilchrist and Chooi, 2020). Average nucleotide identity (ANI) of the genomes included in synteny analyses was calculated using a web-based ANI calculator (Yoon et al., 2017). Accession numbers for the corresponding genomes are provided in **Supplementary Table 3G**.

Assessment of metabolic capacities was based on the most complete and least fragmented genome drafts of the Enterobacteriaceae (L207) and Neisseriaceae (N206) symbionts. The Enzyme Commission (EC) numbers from the PROKKA output and the KEGG orthologs (KO) assigned by BlastKOALA (Kanehisa et al., 2016) were used to map the metabolic pathways in KEGG Mapper (last accessed: November 2021).[1] For the pathways assumed to potentially play a role in the symbiosis, absence of the genes identified as missing was verified using BLAST search. The metabolic reconstructions of B vitamins for the symbionts were compared to the metabolic pathways of several other louse symbionts with available genomes in NCBI, i.e., *Puchtella* sp. str. PRUG (NZ_LKAS01000001.1 + NZ_LKAS01000002.1), *Riesia pediculicola* USDA (GCF_000093065.1), *Legionella polyplacis* (GCA_002776555.1), *Sodalis*-like endosymbiont of *Proechinophthirus fluctus* (GCA_001602625.1), Neisseriaceae bacterium PsAf (GCA_017114885.1), and Neisseriaceae bacterium HaMa (GCA_016864895.1). The KO numbers for compared genomes are deposited in Mendeley Data under the doi link https://doi.org/10.17632/cks86467mv.6.

Proposal for the genus and species name of Enterobacteriaceae symbiont

As the results of our phylogenetic analyses show that Enterobacteriaceae symbionts form a new monophyletic lineage within the Enterobacteriaceae family, we propose the name for all included strains as "*Candidatus* Lightella neohaematopini," gen. nov., sp. nov. (hereafter, *Lightella neohaematopini* for a simple reference). The genus name *Lightella* refers to evolutionary biologist Jessica E. Light, who devoted part of her scientific career to phthirapteran research, and the species name *neohaematopini* refers to the genus of its insect host, *Neohaematopinus pacificus*.

Results and discussion

Taxonomic profiling of metagenomic reads and metagenomic bins

Initial rapid taxonomic assignment of metagenomic reads provided the first insight into the bacterial content of 21 SRA datasets. On average, only 0.4% reads from each dataset were assigned to Bacteria detecting altogether 359 unique bacterial strains (**Supplementary Table 1**). At the genus level, the highest proportion of bacterial reads were consistently classified to *Cutibacterium*, *Lactobacillus*, *Staphylococcus, and Bradyrhizobium*. In some of the samples, i.e., SRR12483201, SRR12483218, and SRR12483221, a number of bacterial reads were assigned to known symbiotic taxa, i.e., *Buchnera*,

1 http://www.genome.jp/kegg/mapper.html

Baumannia, and *Evansia* (**Supplementary Figure 2**). While the omnipresent taxa (*Cutibacterium*, *Lactobacillus*, *Staphylococcus*, *and Bradyrhizobium*) likely represent common human, environment or extraction kit related contamination typical for low biomass samples (Salter et al., 2014), identification of reads classified as obligate insect symbionts point out the association with similarly AT-rich bacteria.

The metagenomic assemblies reflected different qualities, sequencing depth and length of 21 SRA datasets (**Table 1**). The binning process resulted in a wide range in the number of bins (18–166) generated from each of the assemblies. Only 21 bins from 14 datasets were taxonomically assigned to Bacteria. The low number of bacterial bins (among 1,920 retrieved), further referred as MAGs, can be explained by the default settings used for the binning process (including min. contig length of 1,000 nt) and/or by the metagenomic origin of the SRA datasets containing high proportions of eukaryotic contigs. In contrast to the metagenomic taxonomic profiling of raw reads suggesting the omnipresence of *Cutibacterium*, *Lactobacillus*, *Staphylococcus*, *and Bradyrhizobium* genera, we have only retrieved MAGs for *Bradyrhizobium* and *Lactobacillus* from four of the meta-assemblies. It is difficult to determine the causes of this discrepancy, but it could be due to contamination with fragmented DNA, biased amplification during library preparation, or simply assignment inaccuracies stemming from the classifiers and databases we used. The majority of MAGs (12) assembled from twelve samples were identified as Enterobacteriaceae. While for seven of those the taxonomy was only resolved to the family level, five bins were assigned to symbiotic genera, i.e., *Puchtella* and *Wigglessworthia*. In addition, a single MAG representing another potentially symbiotic associate was identified as the family Neisseriaceae in a single dataset. With respect to the focus of this study, the two bacterial groups (Enterobacteriaceae and Neisseriaceae) were of particular interest and were included in further analyses.

Lightella neohaematopini: A new lineage of Enterobacteriaceae symbiont

Using BLASTx and BLASTn searches, the sequences of the new enterobacterial lineage were detected in all except one sample (SRR12483202), but the quality of these genome drafts (i.e., number of contigs and their overall lengths) varied among the SRA datasets (**Table 2**). The size comparisons indicate that the best drafts represent almost complete genomes, as with the growing quality of the assemblies, the drafts converge to the same size (**Table 2**). For eight assemblies, the size of the drafts ranged between 463,996 and 461,635 bp, with the longest contigs in the five best assemblies exceeding 217 kb. While this convergence suggested that the drafts reached a high level of completeness, BUSCO produced very low values. Depending on the database, completeness for sample L207 was estimated as 45.7, 41.3, and 45.2% for Proteobacteria, Gammaproteobacteria, and Enterobacteriales, respectively (**Supplementary Table 4**). Such low values, however, are common even for complete genomes of some symbiotic bacteria (*Candidatus* Riesia pediculischaeffi PTSU: 59.5%; *Sulcia muelleri* strain GWSS: 22.8%; values taken from the UniProt database).[2] The least fragmented assembly L207 (SRA sample SRR12483207) contained five contigs. This remaining fragmentation seems to be caused partly by the presence of two copies of the 23S rRNA gene, which were chimerically assembled in a single contig and could not be separated using this data (**Supplementary Figure 3**). This fact is also reflected by the presence of a short contig containing only the 23S rRNA gene, with coverage almost twice as high as that in the other contigs. Summarizing these parameters, the five least fragmented and most complete drafts (**Table 2**) were considered almost complete genomes and were used in the subsequent comparative analyses. The overview provided in the **Table 3** shows that this bacterium (represented by the L207 genome assembly) possesses features similar to the closest assigned taxon, *Puchtella* sp. str. PRUG (also see below for the phylogenetic analyses and **Supplementary Figure 4D** for 16S rRNA gene similarities), e.g., genome size, GC content, coding density, number of protein coding sequences, absence of transposases, mobile elements or phage-relates sequences. In contrast to this similarity and phylogenetic relatedness, virtually no synteny was detected between the *L. neohaematopini* (L207) and *Puchtella* sp. str. PRUG genomes. The dotplot produced by MUMMER v.3 (Kurtz et al., 2004), shown in **Supplementary Figure 4**, indicates that during their diversification from a common ancestor, one or both lineages underwent substantial genome rearrangements.

Annotation of the genome drafts resulted in similar sets of genes for the five best assemblies (annotations for corresponding genomes are deposited in Mendeley Data under the "doi" link https://doi.org/10.17632/cks86467mv.6). The number of protein coding genes varied between 435 and 443 (415–423 for genes with functional annotations and 19–23 for hypothetical genes). In three strains, the annotation contained two copies of 16S rRNA gene. In two strains the second copy was missing, most likely because of its position at the end of a contig (**Supplementary Figure 4**). Only one copy of 23S rRNA gene was found in the metagenomic assemblies and was placed on a short separated contig. Identical fragments of its sequences were, however, detected at the ends of two different contigs, suggesting that the 23S rRNA gene is also present in two copies and is one of the causes of the genome fragmentation (**Supplementary Figure 3**). The alignments obtained by Clinker and Mauve implemented in Geneious revealed the completely syntenic

2 www.uniprot.org/proteomes

10.3389/fmicb.2022.900312

TABLE 2 Lengths of symbiotic contigs extracted from *Neohaematopinus* lice meta-assemblies.

Neohaematopinus pacificus sample name	SSR data code	*Lightella neohaematopini draft genome code*	Number of *Lightella neohaematopini* contigs	Total length of *Lightella neohaematopini* contigs (bp)	Length of the longest *Lightella neohaematopini* contig	Neisseriaceae* symbiont draft genome Code	Number of Neisseriaceae* symbiont Contigs	Total length of Neisseriaceae* symbiont contigs (bp)	Length of the longest Neisseriaceae* Symbiont contig
DZTM1119Np	SRR12483222	L222	14	463,996	105,366	N222	3	6,579	3,880
DZTM377Np	SRR5088469	L469	7	463,572	217,438	N469	14	8,463	2,334
DZTM1701Np	SRR12483220	L220	7	463,522	188,166	N220	7	1,722	283
MVZ225305Np	SRR12483207	L207	5	462,938	217,084	N207	1	276	276
NK217036N	SRR12483201	L201	6	462,716	216,235	N201	24	13,211	5,588
DZTM1620Np	SRR12483221	L221	6	462,700	216,229	N221	7	1,765	290
DZTM2189Np	SRR12483218	L218	45	462,561	37,759	–	–	–	–
ZM.13956Np	SRR12483208	L208	6	461,635	216,026	N208	5	1,247	297
DZTM230Np	SRR12483217	L217	102	451,834	27,981	N217	5	1,314	282
ZM.13998Np	SRR12483209	L209	142	419,545	10,090	N209	6	1,484	278
DZTM2717Np	SRR12483214	L214	205	395,449	7,732	N214	6	2,034	689
MSB84515Np	SRR12483204	L204	105	347,616	11,111	N204	653	620,299	11,296
DZTM268Np	SRR12483215	L215	223	102,538	2,639	N215	30	7,499	295
DZTM708Np	SRR12483211	L211	231	92,062	1,150	N211	1	249	249
NK181766Np	SRR12483203	L203	144	51,326	1,557	–	–	–	–
DZTM203Np	SRR12483219	L219	134	39,424	660	N219	7	1,835	291
DZTM946Np	SRR12483210	L210	75	21,336	734	–	–	–	–
DZTM2776Np	SRR12483213	L213	3	12,058	1,097	N213	31	8,034	346
MVZ225310Np	SRR12483206	L206	46	11,224	3,732	N206	459	1,476,030	17,945
DZTM584Np	SRR12483212	L212	7	1,740	284	N212	4	1,058	307
NK215220Np	SRR12483202	–	–	–	–	–	–	–	–

The most complete and least fragmented genome drafts of *L. neohaematopini* are highlighted in green. The *L. neohaematopini* samples are ordered according to the total length of the concatenated genome drafts. *Neisseriaceae-related.

nature of the five strains (**Supplementary Figure 4**). Although the fragmentation into the contigs (up to seven) does not allow for an alignment in a strict sense, the complete syntenies along all contigs suggests that there has not been any rearrangement of the gene order. The few mismatches in the alignments were caused by the following: (1) Aligning inaccuracy at the ends of the contigs, (2) missing gene predictions or different annotations, and (3) real differences in absence/presence of the gene. In some cases, these causes are not easily recognized (e.g., missing annotation *vs.* absent sequence), but they account for only a small fraction of the genomic content (406 genes are present and identically annotated in all five genomes).

The BLAST-based verification of the common origin of all selected contigs produced hits of a broad taxonomic range. Although altogether the analysis supported *L. neohaematopini* as a member of Enterobacteriaceae, due to the aberrant nature

10.3389/fmicb.2022.900312

TABLE 3 Comparison of the main characteristics of the two new symbionts (the best drafts) with their closest relatives.

Genome	Louse host	Mammal host	Genome size (bp)	GC content (%)	Coding density (%)	CDS	Predicted proteins	Hypothetical proteins	Pseudogens	Transposases	Phage-related sequences	Mobile elements	BUSCO evaluation (%)
Lightella neohaematopini (L207)	*Neohaematopinus pacificus*	*Tamias alpinus*	462,938	22.2	90.5	449	443	23	29	0	0	0	45.2
Puchtella sp. str. PRUG	*Pedicinus badii*	*Procolobus rufomitratus*	558,122	24.2	92.4	602	547	34	47	0	0	0	72.7
Neisseriaceae-related symbiont (N206)	*Neohaematopinus pacificus*	*Tamias obscurus*	1,476,030	33.7	75.7	1,501	1,472	515	133	2	2	0	62.2
Neisseriaceae-related symbiont (PsAf)	*Polyplax serrata*	*Apodemus flavicollis*	1,814,374	33.7	89.4	1,739	1,660	336	106	5	0	0	85.8
Neisseriaceae-related symbiont (HaMa)	*Hoplopleura acanthopus*	*Microtus arvalis*	1,607,498	33.4	83.5	1,369	1,303	207	11	11	0	0	80.9

of the sequences (manifested by very low GC content of 22.2% and long branches in phylogenetic trees), it failed to reveal a specific closely related bacterium. Instead, the best hits included various taxa of Enterobacteriaceae, often the highly modified symbiotic bacteria such as *Buchnera*, *Blochmannia*, *Baumannia*, *Wigglessworthia*, and *Sodalis* (BLASTx results for L207 presented in **Supplementary Table 2**).

Phylogenetic position of *Lightella neohaematopini*

Considering different qualities of the genome drafts, we performed phylogenetic analyses with two different data sets. The first data set of 14 genes (7,438 amino acids), which included two lineages of *L. neohaematopini* (L207, L221) and 100 (the "phylogenetic matrix") or 63 (the "reduced phylogenetic matrix") additional members of Enterobacteriales, produced a tree in which the two strains of *L. neohaematopini* formed a monophyletic cluster on a long branch, placed as a sister group to *Puchtella* (**Figure 1** and **Supplementary Figures 5–7**). This relationship was obtained by both methods, maximum likelihood and Bayesian inference (MrBayes and PhyloBayes). However, the position of these two sister taxa (*L. neohaematopini* and *Puchtella*) as well as the arrangement of other symbiont taxa differed between the analyses. While maximum likelihood and the MrBayes analyses placed the *L. neohaematopini* + *Puchtella* lineage within a large cluster of symbionts on long branches (**Supplementary Figures 6**, **7**), the PhyloBayes analysis split this cluster into several distinct lineages (**Figure 1** and **Supplementary Figure 5**). In this tree, *L. neohaematopini* + *Puchtella* clustered in vicinity of *Wigglessworthia* and *Blochmannia*, two obligate insect symbionts with strongly modified genomes (i.e., highly reduced and AT-rich). Since highly aberrant sequences of symbiotic bacteria are known to be affected by long-branch attraction (Charles et al., 2001; Husnik et al., 2011), the cluster of long-branched symbionts produced by the maximum likelihood and MrBayes analyses is likely to be artificial and does not reflect real phylogenetic relationships. We therefore consider the PhyloBayes-derived topology, which is also compatible with the results obtained by Husnik et al. (2011), as a more reliable reconstruction of evolutionary history. However, with the current data on louse symbionts, the phylogenetic position of long-branched *L. neohaematopini* does not allow for interpretation of its symbiotic origin. Considering the phylogenetic distance between their hosts (Light et al., 2010), it is unlikely that *L. neohaematopini* and *Puchtella* share a common lice-associated symbiotic ancestor. We rather suggest that these bacteria established their relationship independently with different louse lineages and evolved into mutualists. This view is compatible with the broader phylogenetic picture showing that sucking lice

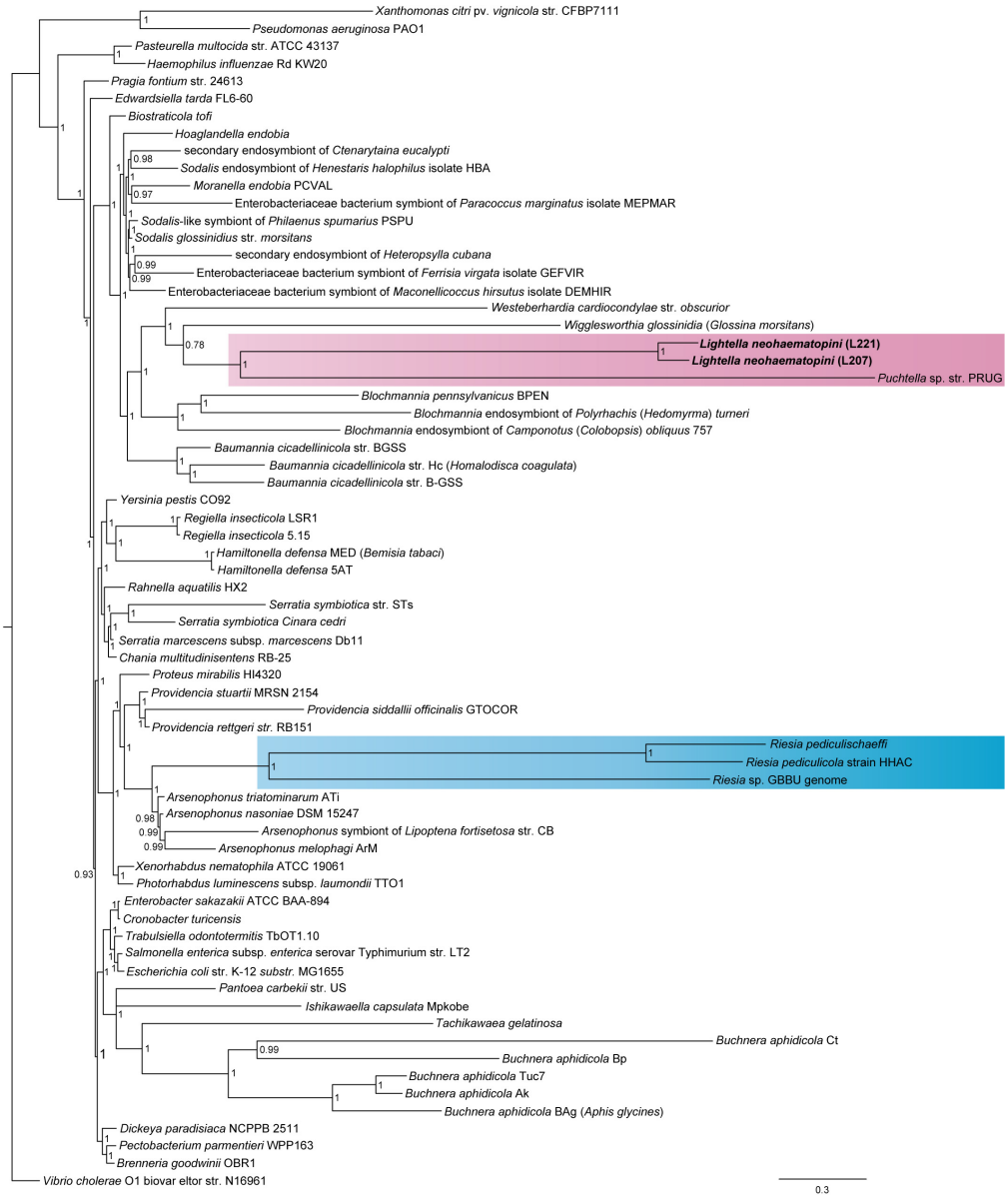

FIGURE 1
Position of *Lightella neohaematopini* within Enterobacteriales revealed by PhyloBayes analysis (BI under CAT-GTR model) of the concatenated 14-protein "reduced phylogenetic matrix" (7,438 aa). Clustering of *L. neohaemotopini* together with the *Puchtella* sp. str. PRUG is indicated by pink background. Clustering of another louse symbiont, the genus *Riesia*, within *Arsenophonus* cluster is highlighted by blue background. Values at the nodes show posterior probabilities.

acquired their symbionts several times from different groups of bacteria (Rihova et al., 2021). It should also be noted that no data are currently available on other *Neohaematopinus* species or related genera, such as *Haemodipsus* or *Sathrax*. If these lice possess symbionts related to *L. neohaematopini*, their inclusion in the analysis could in principle "break" the long branch of *L. neohaematopini* and allow for a more reliable phylogenetic placement, as well as evolutionary interpretation of its origin.

The second data set composed of 1,150 amino acids (five genes) from 12 strains of *L. neohaematopini* produced a ML tree (**Supplementary Figure 8**) and BI tree (**Figure 2**) with topologies closely corresponding to that of the host lice. Almost perfect congruence was found between the *L. neohaematopini* tree and the host tree produced by the ASTRAL-III method (**Figure 2**). The only discrepancy was the monophyly (louse tree) *vs.* paraphyly (symbionts) of the four-taxa cluster highlighted in **Figure 2**. This difference is most likely caused by an incorrect

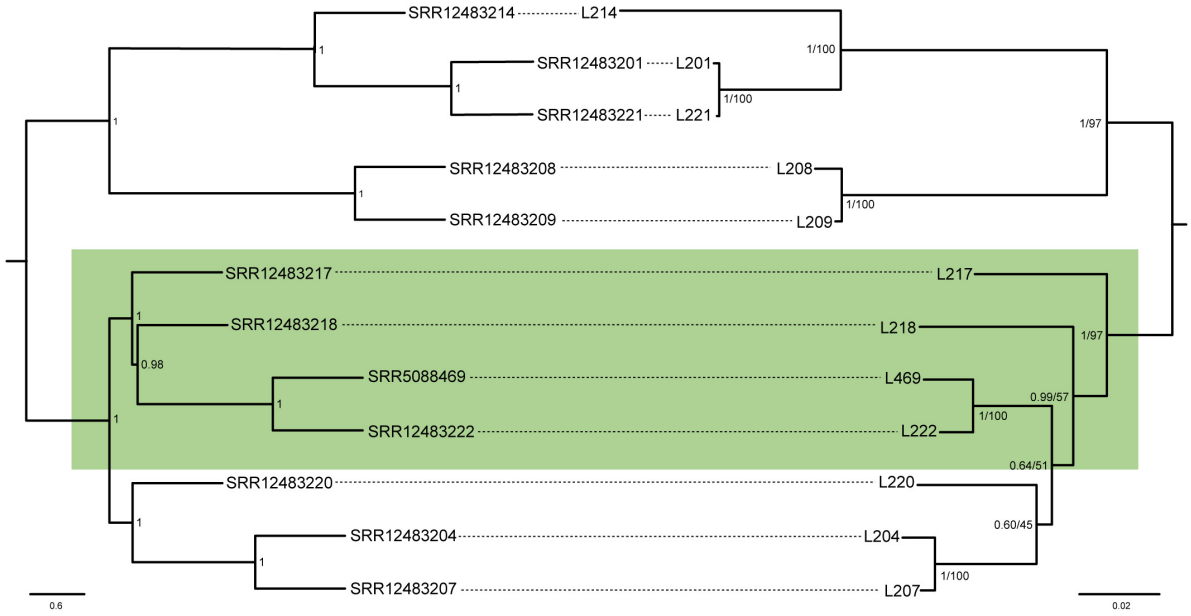

FIGURE 2
Coevolutionary reconstruction comparing phylogenetic trees of the lice and *Lightella neohaematopini*. The host tree was inferred by ASTRAL-III method from 1,107 individual gene trees, each tree derived from one of the 1,107 nucleotide matrices by IQ-TREE using GTR + F + R10 model. The symbiont tree was inferred by MrBayes (BI under JTT + G + F model) and PhyML (ML under HIVb + G + F model) from the amino-acid five-gene "coevolutionary matrix" (1,150 aa). Lengths of the branches in the symbiont tree correspond to the BI results. Values at the nodes of the symbiont tree show posterior probabilities/bootstrap. Values at the nodes of the host tree show local posterior probabilities. The green box highlights the conflicting parts of the phylogenies.

topology of one or both trees. As indicated in the symbiont tree in **Figure 2**, the conflicting phylogeny included two branches (indicated by green background in **Figure 2**) which are among the shortest in the tree and supported with low bootstrap values and posterior probabilities. Similarly, the host trees produced by two different methods (ASTRAL-III and IQ-TREE) differ in arrangement of these groups (**Figure 2** and **Supplementary Figure 9A**). The extension of the ortholog number by reducing the set of symbiont samples to 11 (L204 removed) resulted in slightly different topology but did not produce complete louse-symbiont congruence (**Supplementary Figure 9B**). Despite these minor incongruences, most likely caused by phylogenetic artifacts, the overall close correspondence between the host and the symbiont tree provides strong evidence for the common origin of these symbionts and their coevolution with the host.

Metabolic capacity and symbiotic role of *Lightella neohaematopini*

The genome of *L. neohaematopini* shows several features typical for obligate mutualists in insects (e.g., low GC content, reduced size), indicating a possible metabolic role in its louse host. In correspondence with the strong genome reduction, this new symbiont has considerably limited metabolic capacity (**Figure 3** and **Supplementary Figure 10**). Two categories of metabolites are most often considered compounds provided by an obligate symbiont to the blood-sucking host, amino acids and B vitamins (Nogge, 1981; Hosokawa et al., 2010; Snyder et al., 2010; Rihova et al., 2017). The *Neohaematopinus*-symbiont seems to have lost capacity for synthesis of most of the amino acids, only retaining pathways for lysine, glycine, and serine (for the serine/glycine pair, a category of "cyclic pathway" was established in Rihova et al. (2021), as many symbionts have the capacity for interconverting between them but cannot synthesize these amino acids from glucose; **Supplementary Figure 10**). On the other hand, *L. neohaematopini* retained considerable numbers of genes involved in synthesis of several B vitamins, namely riboflavin, folate, biotin, pantothenate, and pyridoxine (**Figure 3**). Of these vitamins, *L. neohaematopini* contains complete pathways for biotin. This pathway is also retained by three other louse-associated symbionts included in the comparison, *Puchtella* sp., *Riesia pediculicola*, and *L. polyplacis*, suggesting that biotin might be the most important compound provided to lice by their symbionts. In the riboflavin pathway, one gene is missing (5-[amino-6-(5-phospho-D-ribitylamino)uracil phosphatase; EC: 3.1.3.104] but as shown previously (and also here in **Figure 3**), this gene is missing in otherwise complete pathways of many other symbiotic bacteria, and it is thus likely that its function can be replaced by some other not yet identified gene (Rihova et al., 2021). The pathways for two additional B vitamins, folate and pyridoxine,

10.3389/fmicb.2022.900312

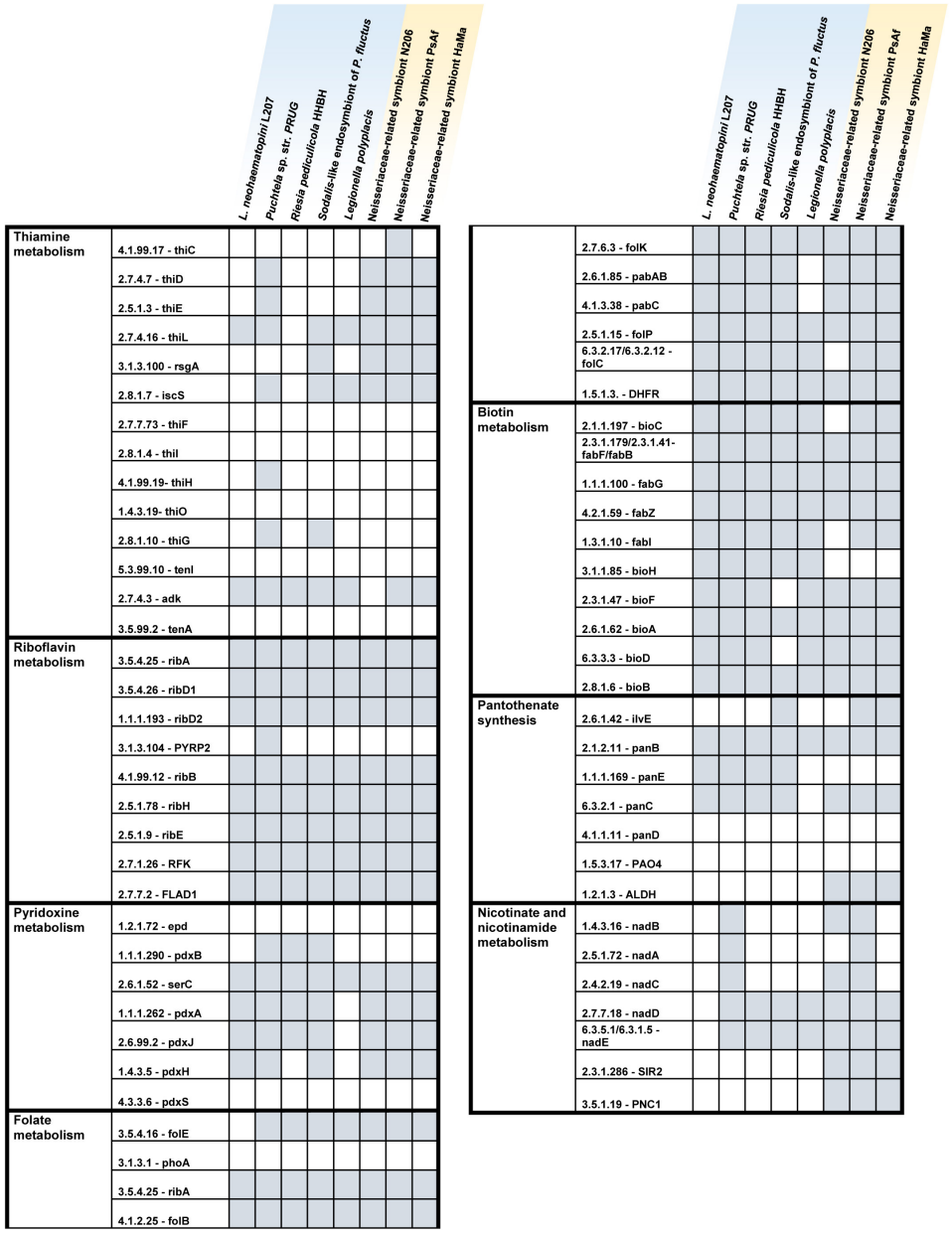

FIGURE 3
Comparison of B-vitamins pathways for *Lightella neohaematopini* L207, Neisseriaeceae-related symbiont N206 and other louse symbionts (blue background = γ-proteobacteria, yellow background = β-proteobacteria). Presence of the particular genes is indicated by gray background.

are missing two and three genes, respectively, some of them missing in all genomes included in the analysis. It is difficult to deduce from this pattern if *L. neohaematopini* can produce (or contribute to production) of these vitamins. It has been previously shown that some seemingly missing genes in these pathways playing role in mutualistic relationships can be present on a plasmid (Kirkness et al., 2010). However, this does not seem to be the case here: when using orthologs of these genes as BLAST query, we did not find their homologs in any of the 21 analyzed assemblies. For pantothenate synthesis, the symbiont's genome contains three genes (panB, panC, and panG) while the absence of the panD is likely to be compensated by the host (Price and Wilson, 2014). These pathways also do not contain any of the genes identified as potentially being pseudogenized. Altogether, this overview indicates a role for the symbiont in provisioning several B vitamins, or at least biotin, like some other louse symbionts (Boyd et al., 2017; Rihova et al., 2017, 2021). A comparison with its close relative, *Puchtella* sp. str. PRUG (symbiont of *P. baddi* from red colobus monkey *Procolobus rufomitratus*),

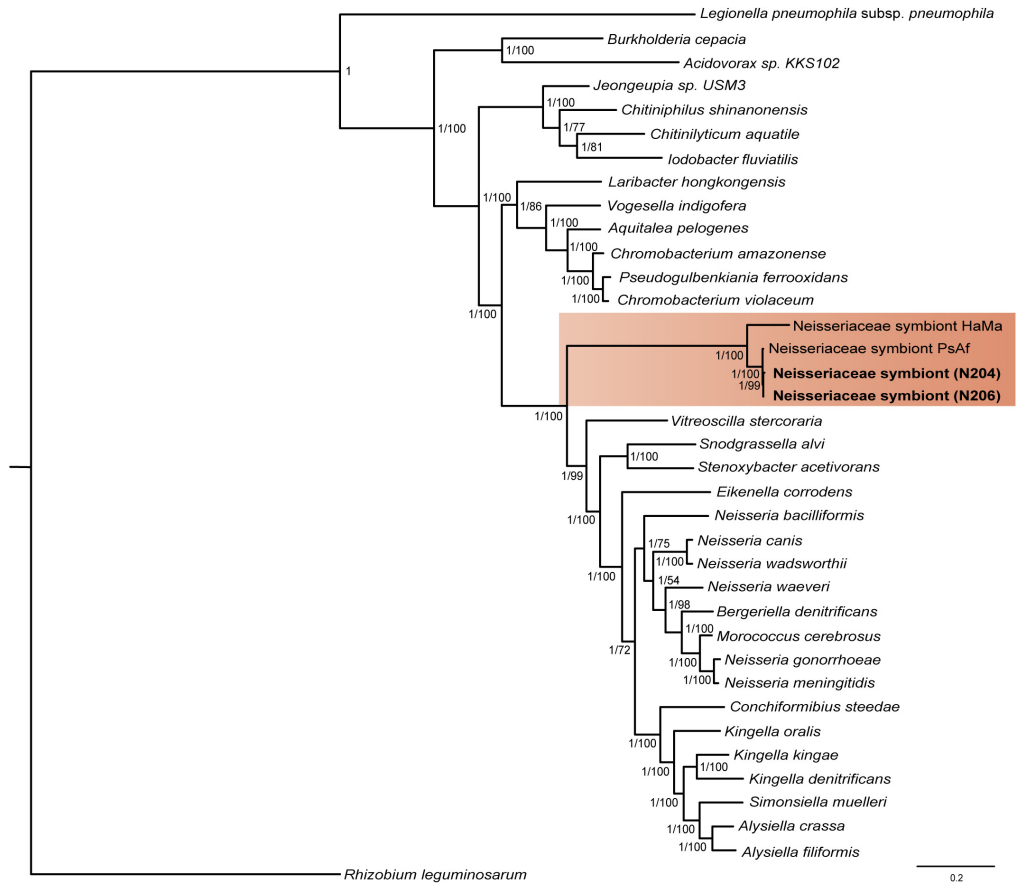

FIGURE 4
Clustering of the N206 and N204 symbionts together with the Neisseriaceae-related symbionts from Rihova et al. (2021), indicated by orange background. The tree was inferred by MrBayes (LG + G + I + F model) and PhyML (LG + G + I + F model) from the amino-acid 30-gene "Neisseriales matrix" (4,465 aa), lengths of the branches in the tree of the symbiont correspond to the BI tree results. Values at the nodes of the symbiont tree show posterior probabilities/bootstrap supports.

shows highly similar metabolic capacities of these two louse symbionts (**Figure 3**).

Neisseriaceae-related symbionts

Apart from *L. neohaematopini*, the binning process detected a putative Neisseriales-related symbiont (**Table 1**). Its sequences were present in 17 assemblies, most of them in highly fragmented and incomplete form (**Table 2**). The most complete genome draft (N206) is 1,476,030 bp long (the longest contig reaching 17,945 bp) with a low GC content (33.7%). It contains 1,472 predicted protein coding genes, 515 of them annotated as hypothetical proteins (**Table 3**). The two assemblies (N206 and N204) included in the phylogenetic analysis cluster together with the Neisseriaceae-related symbionts PsAf and HaMa previously described from *Polyplax serrata* and *Hoplopleura acanthopus* lice, respectively (Rihova et al., 2021; **Figure 4**). The HaMa was also retrieved as the first BLAST hit when the contigs assigned to Neisseriaceae in the 17 assemblies were used as a query against the nt NCBI database. The clustering on a long common branch and high branch support (both ML and BI analyses) suggest that all these Neisseriaceae-related bacteria may have originated from a single symbiotic ancestor (**Supplementary Figure 11** and **Figure 4**). This phylogenetic position and the features typical for bacteria undergoing adaptation to symbiosis (e.g., low GC content and high number of pseudogenes) indicate the symbiotic nature of these bacteria. The main characteristics of the N206 genome, and its comparison to the other Neisseriaceae-related symbionts are summarized in **Table 3**. Since *N. pacificus* harbor *L. neohaematopini* as a typical obligate mutualistic P-symbiont, the significance of the Neisseriaceae-related symbiont for the host is unclear. Furthermore, the highly fragmented and possibly incomplete genome assembly (BUSCO completeness value 62.2%) does not allow for a reliable reconstruction of its metabolic capacities (in **Figure 3** and **Supplementary Figure 10** we show the capacities which can be inferred from this data).

It is, however, interesting to see that these bacteria are present in several louse genera, possibly playing the role of either S-symbionts (*N. pacificus, P. serrata*) or even P-symbionts (*H. acanthopus*).

The results presented in this study provide two pieces of information relevant for broader insight into the evolution of symbiosis between lice and bacteria. First, the new putative obligate mutualist, *L. neohaematopini*, extends the list of bacterial lineages which have established mutualistic symbiosis with different groups of sucking lice. This further supports the view that sucking lice underwent an exceptionally dynamic process of symbiont acquisitions and replacements. Phylogenetic match of *L. neohaematopini* with *Neohaematopinus pacificus* lice from different chipmunk species provides strong evidence that these symbionts were acquired in a single evolutionary event and have been maintained during the host's radiation. This codiversification process likely reflects the nutritional role of these bacteria. A comparison with other known louse symbionts suggests that the compounds most likely responsible for this dependence are several B vitamins. In contrast, the second potentially symbiotic bacterium is related to Neisseriaceae-related symbionts already known from two different groups of lice (Rihova et al., 2021). Interestingly, in *Neohaematopinus pacificus* lice the co-occurrence of these two bacteria resembles the situation in *Polyplax serrata*, where the mutualistic *Legionella polyplacis* is occasionally accompanied by the Neisseriaceae-related symbiont. We expect that addition of so far unexplored louse lineages will further extend overall diversity of their symbionts, but it can also help to resolve some of the current phylogenetic and evolutionary uncertainties.

Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding author.

Author contributions

VH, EN, and JŘ conceived the study. All authors contributed to the data analyses, interpretation of the results and writing the manuscript.

Funding

This work was supported by the Grant Agency of the Czech Republic (grant GA20-07674S to VH).

Acknowledgments

We would like to acknowledge Jessica E. Light and Julie Allen for their reviews of the article. We also highly appreciate access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures" (CESNET LM2015042).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.900312/full#supplementary-material

References

Allen, J. M., Burleigh, J. G., Light, J. E., and Reed, D. L. (2016). Effects of 16S rDNA sampling on estimates of the number of endosymbiont lineages in sucking lice. *PeerJ* 4:e2187. doi: 10.7717/peerj.2187

Alneberg, J., Bjarnason, B., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U., et al. (2014). Binning metagenomic contigs by coverage and composition. *Nat. Methods* 11, 1144–1146. doi: 10.1038/nmeth.3103

Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., et al. (2018). KBase: The United States department of energy systems biology knowledgebase. *Nat. Biotechnol.* 36, 566–569. doi: 10.1038/nbt.4163

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genom.* 9:75. doi: 10.1186/1471-2164-9-75

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Bell, K., Allen, J., Johnson, K., Demboski, J., and Cook, J. (2021). Disentangling lousy relationships: comparative phylogenomics of two sucking louse lineages parasitizing chipmunks. *Mol. Phylogenet. Evol.* 155:106998. doi: 10.1016/j.ympev.2020.106998

Bell, K. C., Matek, D., Demboski, J. R., and Cook, J. A. (2015). Expanded host range of sucking lice and pinworms of western North American chipmunks. *Comp. Parasitol.* 82, 312–321. doi: 10.1654/4756.1

Boyd, B., Allen, J., Koga, R., Fukatsu, T., Sweet, A., Johnson, K., et al. (2016). Two bacterial genera, *Sodalis* and *Rickettsia*, associated with the seal louse *Proechinophthirus fluctus* (Phthiraptera: Anoplura). *Appl. Environ. Microbiol.* 82, 3185–3197. doi: 10.1128/AEM.00282-16

Boyd, B. M., Allen, J. M., de Crécy-Lagard, V., and Reed, D. L. (2014). Genome sequence of *Candidatus* Riesia pediculischaeffi, endosymbiont of chimpanzee lice, and genomic comparison of recently acquired endosymbionts from human and chimpanzee lice. *G3* 4, 2189–2195. doi: 10.1534/g3.114.012567

Boyd, B. M., Allen, J. M., Nguyen, N. P., Vachaspati, P., Quicksall, Z., Warnow, T., et al. (2017). Primates, lice and bacteria: speciation and genome evolution in the symbionts of hominid lice. *Mol. Biol. Evol.* 34, 1743–1757. doi: 10.1093/molbev/msx117

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST plus: architecture and applications. *BMC Bioinform.* 10:421. doi: 10.1186/1471-2105-10-421

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334

Charles, H., Heddi, A., and Rahbe, Y. (2001). A putative insect intracellular endosymbiont stem clade, within the Enterobacteriaceae, infered from phylogenetic analysis based on a heterogeneous model of DNA evolution. *C. R. Acad. Sci. III* 324, 489–494. doi: 10.1016/S0764-4469(01)01328-2

Chaumeil, P., Mussig, A., Hugenholtz, P., and Parks, D. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36, 1925–1927. doi: 10.1093/bioinformatics/btz848

Chen, X. A., Li, S., and Aksoy, S. (1999). Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. *J. Mol. Evol.* 48, 49–58. doi: 10.1007/pl00006444

Darling, A. E., Mau, B., and Perna, N. T. (2010). progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. doi: 10.1371/journal.pone.0011147

Darriba, D., Taboada, G., Doallo, R., and Posada, D. (2012). jModelTest2: more models, new heuristics and parallel computing. *Nat. Methods* 9:772. doi: 10.1038/nmeth.2109

Dhami, M. K., Buckley, T. R., Beggs, J. R., and Taylor, M. W. (2013). Primary symbiont of the ancient scale insect family *Coelostomidiidae* exhibits strict cophylogenetic patterns. *Symbiosis* 61, 77–91. doi: 10.1007/s13199-013-0257-8

Emms, D., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238. doi: 10.1186/s13059-019-1832-y

Freitas, T. A. K., Li, P. E., Scholz, M. B., and Chain, P. S. (2015). Accurate read-based metagenome characterization using a hierarchical suite of unique signatures. *Nucleic Acids Res.* 43:e69. doi: 10.1093/nar/gkv180

Fukatsu, T., Hosokawa, T., Koga, R., Nikoh, N., Kato, T., Hayama, S., et al. (2009). 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.* 75, 3796–3799. doi: 10.1128/aem.00226-09

Gilchrist, C. L. M., and Chooi, Y. H. (2020). Clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* 37, 2473–2475. doi: 10.1093/bioinformatics/btab007

Guindon, S., Dufayard, J., Hordijk, W., Lefort, V., and Gascuel, O. (2009). PhyML: fast and accurate phylogeny reconstruction by maximum likelihood. *Infect. Genet. Evol.* 9, 384–385.

Guindon, S., Dufayard, J., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281

Hosokawa, T., Koga, R., Kikuchi, Y., Meng, X. Y., and Fukatsu, T. (2010). *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Proc. Natl. Acad. Sci. U. S. A.* 107, 769–774. doi: 10.1073/pnas.0911476107

Husnik, F., Chrudimsky, T., and Hypsa, V. (2011). Multiple origins of endosymbiosis within the Enterobacteriaceae (γ-Proteobacteria): convergence of complex phylogenetic approaches. *BMC Biol.* 9:87. doi: 10.1186/1741-7007-9-87

Hypsa, V., and Krizek, J. (2007). Molecular evidence for polyphyletic origin of the primary symbionts of sucking lice (*Phthiraptera*, *Anoplura*). *Microb. Ecol.* 54, 242–251. doi: 10.1007/s00248-006-9194-x

Kalyaanamoorthy, S., Minh, B., Wong, T., von Haeseler, A., and Jermiin, L. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587-589. doi: 10.1038/NMETH.4285

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, 457–462. doi: 10.1093/nar/gkv1070

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (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

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., and Sturrock, S. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kirkness, E. F., Haas, B. J., Sun, W., Braig, H. R., Perotti, M. A., and Clark, J. (2010). Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc. Natl. Acad. Sci.* 107, 12168–12173. doi: 10.1073/pnas.1003379107

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12. doi: 10.1186/gb-2004-5-2-r12

Langmead, B., and Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lartillot, N., Rodrigue, N., Stubbs, D., and Richer, J. (2013). PhyloBayes MPI: Phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* 62, 611–615. doi: 10.1093/sysbio/syt022

Lefort, V., Longueville, J., and Gascuel, O. (2017). SMS: smart model selection in PhyML. *Mol. Biol. Evol.* 34, 2422–2424. doi: 10.1093/molbev/msx149

Leinonen, R., Sugawara, H., and Shumway, M. (2011). The sequence read archive. *Nucleic Acids Res* 39, D19–D21. doi: 10.1093/nar/gkq1019

Light, J. E., Smith, V. S., Allen, J. M., Durden, L. A., and Reed, D. L. (2010). Evolutionary history of mammalian sucking lice (Phthiraptera: Anoplura). *BMC Evol. Biol.* 10:292. doi: 10.1186/1471-2148-10-292

McCutcheon, J. P., Boyd, B. M., and Dale, C. (2019). The life of an insect endosymbiont from the cradle to the grave. *Curr. Biol.* 29, 485–495. doi: 10.1016/j.cub.2019.03.032

McMurdie, P. J., and Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. doi: 10.1371/journal.pone.0061217

Minh, B., Schmidt, H., Chernomor, O., Schrempf, D., Woodhams, M., von Haeseler, A., et al. (2020). IQ-TREE2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37:2461. doi: 10.1093/molbev/msaa131

Moran, N., Munson, M., Baumann, P., and Ishikawa, H. (1993). A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc. R. Soc. B Biol. Sci.* 253, 167–171. doi: 10.1098/rspb.1993.0098

Moran, N. A., Russell, J. A., Koga, R., and Fukatsu, T. (2005). Evolutionary relationships of three new species of *Enterobacteriaceae* living as symbionts of aphids and other insects. *Appl. Environ. Microbiol.* 71, 3302–3310. doi: 10.1128/AEM.71.6.3302-3310.2005

Nogge, G. (1981). Significance of symbionts for the maintenance of an optimal nutritional state for successful reproduction in hematophagous *Arthropods*. *Parasitology* 82, 101–104.

Nováková, E., Hypša, V., and Moran, N. A. (2009). *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. *BMC Microbiol.* 9:143. doi: 10.1186/1471-2180-9-143

Price, D., and Wilson, A. (2014). A substrate ambiguous enzyme facilitates genome reduction in an intracellular symbiont. *BMC Biol.* 12:110. doi: 10.1186/s12915-014-0110-4

Puchta, O. (1955). Experimentelle untersuchungen über die bedeutung der symbiose der kleiderlaus *Pediculus vestimenti* burm. *Z. Parasitenkd.* 17, 1–40.

Rambaut, A., Drummond, A., Xie, D., Baele, G., and Suchard, M. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032

Rihova, J., Batani, G., Rodriguez-Ruano, S., Martinu, J., Vacha, F., Novakova, E., et al. (2021). A new symbiotic lineage related to *Neisseria* and *Snodgrassella* arises from the dynamic and diverse microbiomes in sucking lice. *Mol. Ecol.* 30, 2178–2196. doi: 10.1111/mec.15866

Rihova, J., Novakova, E., Husnik, F., and Hypsa, V. (2017). *Legionella* becoming a mutualist: adaptive processes shaping the genome of symbiont in the louse *Polyplax serrata*. *Genome Biol. Evol.* 9, 2946–2957. doi: 10.1093/gbe/evx217

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D., Darling, A., Hohna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539-542. doi: 10.1093/sysbio/sys029

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., et al. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 12:87. doi: 10.1186/s12915-014-0087-z

Sauer, C., Stackebrandt, E., Gadau, J., Holldobler, B., and Gross, R. (2000). Systematic relationships and cospeciation of bacterial endosymbionts and their carpenter ant host species: proposal of the new taxon *Candidatus* Blochmannia gen. nov. *Int. J. Syst. Evol. Microbiol.* 50, 1877–1886. doi: 10.1099/00207713-50-5-1877

Sayyari, E., and Mirarab, S. (2016). Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.* 33, 1654–1668. doi: 10.1093/molbev/msw079

Seemann, T. (2014). PROKKA: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Simao, F., Waterhouse, R., Ioannidis, P., Kriventseva, E., and Zdobnov, E. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Snyder, A. K., Deberry, J. W., Runyen-Janecky, L., and Rio, R. V. M. (2010). Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. *Proc. R. Soc. B Biol. Sci.* 277, 2389–2397. doi: 10.1098/rspb.2010.0364

Spaulding, A. W., and von Dohlen, C. D. (1998). Phylogenetic characterization and molecular evolution of bacterial endosymbionts in psyllids (Hemiptera: Sternorrhyncha). *Mol. Biol. Evol.* 15, 1506–1513. doi: 10.1093/oxfordjournals.molbev.a025878

Stouthamer, R., Breeuwer, J. A., and Hurst, G. D. (1999). *Wolbachia pipientis*: Microbial manipulator of arthropod reproduction. *Annu. Rev. Microbiol.* 53, 71–102.

Sudakaran, S., Kost, C., and Kaltenpoth, M. (2017). Symbiont acquisition and replacement as a source of ecological innovation. *Trends Microbiol.* 25, 375–390. doi: 10.1016/j.tim.2017.02.014

Syberg-Olsen, M., Graber, A. I., Keeling, P., McCutcheon, J., and Husnik, F. (2021). Pseudofinder: detection of pseudogenes in prokaryotic genomes. *bioRxiv* [Preprint]. doi: 10.1101/2021.10.07.463580

Toft, C., and Andersson, S. G. (2010). Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat. Rev. Genet.* 11, 465–475. doi: 10.1038/nrg2798

Toh, H., Weiss, B. L., Perkin, S. A., Yamashita, A., Oshima, K., Hattori, M., et al. (2006). Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Res.* 16, 149–156. doi: 10.1101/gr.4106106

Yoon, S., Ha, S., Lim, J., Kwon, S., and Chun, J. (2017). A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* 110, 1281–1286. doi: 10.1007/s10482-017-0844-4

Zchori-Fein, E., and Perlman, S. J. (2004). Distribution of the bacterial symbiont *Cardinium* in arthropods. *Mol. Ecol.* 13, 2009–2016. doi: 10.1111/j.1365-294X.2004.02203.x

Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform.* 19:153. doi: 10.1186/s12859-018-2129-y

ONGOING RESEARCH AND FUTURE PROSPECTS

In this section, I will present the two ongoing projects (***Ongoing research I*** and ***Ongoing research II***) extending my PhD thesis with additional results which are being prepared for several new manuscripts.

Ongoing research I

Screening of environmental samples to assess putative source(s) of symbiotic bacteria

Background

Current knowledge of symbiotic bacteria of insects covers the evolutionary trajectory from young symbionts to highly specialized and reduced symbionts. However, very little is known about the acquisition of bacteria from insects' environment and their transition to symbionts. This could be caused by several difficulties. For example, sometimes it is complicated to identify the closest free-living relatives of highly reduced symbionts, since symbionts often tend to form long branches with unstable position in phylogenetic trees (Husník et al., 2011). Another obstacle could be a big evolutionary "gap" between free-living relatives and the symbiont, i.e., missing intermediate stages. Therefore, systems with closely related symbiotic and free-living or pathogenic bacteria are valuable models for the investigation of changes related to the origin of symbiosis (Chari et al., 2015).

Anoplura, a group of obligate blood-feeding insects living exclusively on mammal hosts, are well-studied for their symbiotic relationship with bacteria of multiple phylogenetic origins (**MS2**; **MS1**; Nishide et al., 2022; Říhová et al., 2017; Allen et al., 2016; Boyd et al., 2017; Boyd et al., 2016; Boyd et al., 2014; Kirkness et al., 2010; Fukatsu et al., 2009; Allen et al., 2007; Hypša and Křížek, 2007). Currently accumulated data indicates that some of these bacterial lineages are frequently entering symbiotic relationships with unrelated louse genera. For example, recently described Neisseriaceae-related symbionts established symbiotic relationships independently with four louse genera, i.e., *Polyplax*, *Hoplopleura*, *Neohaematopinus* and *Haematopinus* (**MS2**; **MS1**; Nishide et al., 2022). Based on our comparisons of Neisseriaceae-related symbionts on the genome scale (**MS2**, **MS1**) we hypothesized that these bacteria are possibly in an early/intermediate stage of evolution and represent either S-symbionts (*Polyplax serrata*, *Neohaematopinus pacificus*) or even P-symbionts (*Hoplopleura acanthopus*).

In contrast to the wide distribution of the Neisseriaceae-related bacteria among louse lineages, we have no evidence of possible sources of such bacteria in a louse environment. In this study we designed an experiment for screening of several possible sources of symbiotic bacteria in

lice. We hypothesized that in our lice-mice model, the louse symbionts could originate from mouse fur microbiome, microbiome of mouse hair follicles/epidermis, pathogenic bacteria of mice or bacterial symbiont/pathogens of other ectoparasites.

Using a combination of amplicon and metagenomic approaches, we characterized environmental bacteria extracted from mouse fur, which are closely related to the Neisseriaceae-related symbionts of lice. Based on our results, we confirmed that various strains of these free-living Neisseriaceae-related bacteria are commonly present in the fur of studied species of mice, some are even closely related to the symbionts described in **MS1**. Therefore, we suggest that a fur microbiome could possibly be one of the sources of the bacteria, which can potentially establish symbiosis with lice. This study should serve as a methodical set-up for further screening analyses on a bigger dataset (currently in progress).

This part is comprised of unpublished data, which is present in the original thesis deposited at the Faculty of Science, University of South Bohemia.

Ongoing research II

Sequencing, genome assemblies and phylogenetic reconstructions of the new symbionts from obligate blood-feeders – background for the comparative study

Background

Diverse groups of blood-feeding invertebrates such as leeches, ticks, mosquitoes, kissing bugs, bed bugs, tsetse flies, louse flies, bat flies, sucking lice, chewing lice and Rhynchophthirina have established symbiotic associations with bacteria from different lineages, mostly belonging to gammaproteobacteria (**MS2**; Nishide et al., 2022; Boyd et al., 2017; Říhová et al., 2017; Šochová et al., 2017; Allen et al., 2016; Boyd et al., 2016; Boyd et al., 2014; Chrudimský et al., 2012; Hosokawa et al., 2012; Kirkness et al., 2010; Fukatsu et al., 2009; Allen et al., 2007; Fukatsu et al., 2007; Hypša and Křížek, 2007; Nováková and Hypša, 2007; Kikuchi and Fukatsu, 2002; Dale and Maudlin, 1999; Hyša and Dale, 1997; Aksoy, 1995) and alphaproteobacteria (Filée et al., 2022; Mahmood et al., 2013; Hosokawa, 2010). It is hypothesized, that the main role of the P-symbionts is to supplement their hosts with B-vitamins and cofactors, which are missing in the blood meal (Boyd et al., 2016; Manzano-Marín et al., 2015; Nováková et al., 2015; Smith et al., 2015; Nikoh et al., 2014; Rio et al., 2012; Akman et al., 2002). However, the experimental studies endorsing this evidence are very scarce (Snyder and Rio, 2015; Michalkova et al., 2014; Nikoh et al., 2014; Hosokawa et al., 2010).

The potential function of the S-symbionts for these hosts is even less understood since they can manifest many different phenotypes with either mutualistic (metabolic functions, protective functions; Hansen and Moran, 2014; Łukasik et al., 2013b; Oliver et al., 2009; Degnan and Moran, 2008; Oliver et al., 2003; Montllor et al., 2002) or harmful (manipulations of the host´s reproductive system; White et al., 2013; Rio et al., 2012; Duron et al., 2008) effects to their host. The extensive study of the multiple systems has resulted in many significant discoveries but to elucidate how widely relevant and applicable these findings are, a comprehensive comparative approach using a broad range of model systems is highly required. Here, using a combination of amplicon and genomic approaches, we present several new complete genomes and genome drafts of symbionts from multiple obligate blood-feeders (sucking lice, elephant lice and louse flies). The partial results presented in this section should draw the genomic and phylogenetic background for multiple comparative studies.

This part is comprised of unpublished data, which is present in the original thesis deposited at the Faculty of Science, University of South Bohemia.

SUMMARY OF THE MAIN RESULTS AND CONCLUSION

Distribution and diversity of the symbionts in the microbiomes of obligate blood-feeders

The results of the combined genomic and screening analyses revealed diversity and evolutionary features of symbiotic bacteria among several species (and populations) of obligate blood-feeders. The comparisons of the microbiome profiles allowed for the assessing of a distribution of the dominant microbial taxa in the studied samples. It revealed that several bacterial taxa display a complex distribution pattern, likely representing symbionts in different evolutionary stages of symbiogenesis. In particular, we show that sucking lice are exceptional among obligate blood-feeders, having a remarkably high diversity of their P-symbionts originated in taxonomically diverse bacterial groups. While some symbiotic bacteria of lice originate in bacterial groups commonly found in other insects (e.g., *Arsenophonus*/*Riesia* symbionts; **MS1**), others represent newly described bacterial lineages with different evolutionary origins (**MS1**, **MS2**, ***Ongoing research II***). This is in a sharp contrast to symbionts of hippoboscids, which are mainly represented by the genus *Arsenophonus* (***Ongoing research II***) and *Sodalis.*

Phylogenetic origins of the symbiotic bacteria in obligate blood-feeders

These interesting findings about diverse symbioses and their evolution raise questions on the source of these symbionts and a mechanism of symbiogenesis. In the phylogenetic analyses, most of the described bacterial symbionts cluster among symbionts of other insects within Enterobacteriaceae (**MS1**; ***Ongoing research II*)**, however some symbionts of lice (i.e., Neisseriaceae-related symbionts, *Legionella* and *Francisella* symbionts) originate from bacterial lineages which have not (or only rarely) been found in association with insects, but they are related to vertebrate pathogens (**MS1**, **MS2**, ***Ongoing research II*)**. It may suggest that these symbiotic lineages possibly originated from the vertebrate pathogens as proposed in *Francisella* symbionts of ticks. However, in ***Ongoing research I*** we show that such bacteria could also originate from bacteria which commonly inhabit the insect's environment. We expect that future study of these free-living/symbiotic relatives may help to elucidate the evolutionary changes accompanying a symbiogenesis process.

Genomic and metabolic characteristics of the new symbionts

The evolution from free-living bacteria to highly modified symbionts includes degenerative processes of their genomes. The effect of such processes was demonstrated in our comparisons of genome characteristics of the novel symbionts with their relatives. The genomes of these novel symbionts show typical features of genome reduction and represent a diverse spectrum of evolutionary stages with different metabolic capacities. While some of these novel symbionts are rather in an early/intermediate stage of symbiogenesis (Neisseriaceae-related symbionts of lice and short-branched *Arsenophonus* symbionts of hippoboscids; **MS1**, ***Ongoing research II***), others possess characteristics of strongly reduced P-symbionts (i.e., *Legionella* and *Lightella* symbionts of lice; P-symbionts of *H. suis*, *H. sciuricola*, *F. microcephala*, *F. texana*, *N. sciurinus*, *O. turdi* and *H. elephantis*;

MS1, **MS2**, ***Ongoing research II***). Interestingly, the genome characteristics of these highly reduced symbiotic bacteria from unrelated obligate blood-feeders indicate that they tend to converge to similar genome sizes and GC contents (***Ongoing research II***). To assess whether these symbionts also converge in metabolic capacities, we will design a robust background for metabolic comparisons (in preparation).

However, reliable reconstruction of metabolic pathways could be a relatively challenging task. As we show in **MS1** and **MS2**, a function of some genes missing in otherwise complete pathways (e.g., genes coding for phosphatases) could be replaced by other genes with the same function. Also, some seemingly missing genes in these pathways can be present on a plasmid as shown in *Riesia* symbionts of hominid lice (Kirkness et al., 2010). Moreover, in Říhová (2015) we show that while some genes could be clearly presented in the genome, they do not have to be correctly recognized and mapped by annotation and mapping tools. To reduce such bias, the comprehensive research and additional analyses need to be implemented (e.g., blasting of homologs of missing genes against genome assemblies) for reliable metabolic reconstructions. Also, since an absence of some genes could be compensated by the host or co-residing symbionts, it is difficult to infer whether some metabolic pathways are potentially functional and if they play a role in the host-symbiont metabolism. However, in the context of the presented metabolic reconstructions of louse symbionts, we show their possible role in the production (or contribution to production) of several B-vitamins (**MS1** and **MS2**).

Conclusion

In sum, the results presented in this thesis provide a broader insight into evolution of symbiosis between obligate blood-feeders and bacteria. Using multiple methodological approaches, we extent the known diversity of symbionts of obligate blood-feeders with several new lineages and show their evolutionary features. In particular, a high phylogenetic diversity of these lineages in the sucking lice is supporting the view that lice underwent dynamic turnovers of their symbionts. Also, we suggest that the insects' environment could serve as one of the possible sources of these symbiotic bacteria. Altogether, we propose that the investigation of these unique symbioses is important for untangling host-symbiont dependencies and for understanding their role in host ecology. It is fascinating to conclude that our findings contribute to pushing the limits of what we know about minimal genomes for sustaining a cellular life.

AUTHOR´S CURRICULUM VITAE

Name: RNDr. Jana Říhová

Date of birth: 07/20/1993

Nationality: Czech

Address: Svatojánská 103, 384 02 Lhenice, Czech Republic

Tel.: +420 725 151 505

E-mail: rihovj02@prf.jcu.cz, janca.rihova@gmail.com

ORCID: 0000-0001-7937-1960

Education

2018 – now Doctoral candidate in Parasitology - University of South Bohemia, Faculty of Science, České Budějovice

2018 RNDr. state rigorosum examination (RNDr. Thesis: *Legionella* becoming a mutualist: Adaptive processes shaping the genome of symbiont in the louse *Polyplax serrata* – published in 2017, *Genome Biology and Evolution*)

2015 – 2018 Master study of Parasitology - University of South Bohemia, Faculty of Science, České Budějovice

2012 – 2015 Bachelor study of Biology - University of South Bohemia, Faculty of Science, České Budějovice

2008 – 2012 Gymnasium K. V. Raise (High School), Hlinsko

Work experience

2019 – now Editor of mass media – Czech Society for Parasitology

2017 – now Research assistant (Grant projects of VH), Laboratory of Phylogeny and Evolution of Parasites, Faculty of Science, University of South Bohemia, České Budějovice

2017 Trial officer, i2L Research, South Bohemian Science and Technology Park, České Budějovice
Laboratory research focused on the determination of anthelmintic efficacy of the classified product against pinworms; Pesticides and insecticides testing; Breeding of mosquito's cultures

2017 Lector of science courses for kids, Věda nás baví, Faculty of Science, University of South Bohemia, České Budějovice

2015 – 2016 Laboratory technician, Laboratory of Phylogeny and Evolution of Parasites, Faculty of Science, University of South Bohemia, České Budějovice

2015 – 2016 Part-time employee (Grant projects of VH), Laboratory of Phylogeny and Evolution of Parasites, Faculty of Science, University of South Bohemia, České Budějovice

Teaching and supervision

2018 – 2020 Teaching assistant, Molecular phylogenetics, Faculty of Science, University of South Bohemia

2018 – 2019 Teaching assistant, Medical parasitology and diagnostic methods, Faculty of Science, University of South Bohemia

2018 – 2019 Teaching assistant, Medical parasitology for clinical biologists, Faculty of Science, University of South Bohemia

2018 – 2019 Co-supervision of Eliška Juhaňáková, Bc. Thesis, Evoluce symbiotických bakterií krevsajících bakterií dvoukřídlého hmyzu, Faculty of Science, University of South Bohemia

Internships

April – May 2021 Jessica E. Light laboratory, Department of Ecology and Conservation Biology, Texas A&M University, College Station, Texas, USA
Supervisor: Assoc. Prof. Jessica E. Light, Ph. D.

International conferences and talks

September 2022 **Říhová, J.**, Nováková, E., & Hypša, V. (2022). Fur microbiome as a putative source of symbiotic bacteria in lice [poster in English]. ICPOW 2022, 11th – 15th of September, Kruger National Park, South Africa

April 2021 Presentation at ECCB Seminar, Department of Ecology and Conservation Biology, Texas A&M University, Texas, USA; Symbionts and microbiome diversity of lice and keds; 45 mins + discussion

May 2016 **Říhová, J.**, Nováková, E., Husník, F., & Hypša, V. (2016). Phylogenetic and genomic characterization of bacterial endosymbionts of louse *Polyplax serrata* [poster in English]. XII. Czech and Slovak Parasitological Days, 16th – 20th of May 2016, Czech Society for Parasitology, Ledeč nad Sázavou, Czech Republic

Awards and funding

2018 Dean's award for excellent research results presented in Master Thesis

2016 Student Grant Agency of Faculty of Science, University of South Bohemia
Genomická a funkční charakterizace endosymbiontů vší rodu *Polyplax*
(PI: Jana Říhová)
Successful investigator – 3rd place

Trainings

2021 Operator of MiSeq Illumina Sequencer (GeneTICA)

List of publications (non-included in Ph.D. thesis)

Říhová, J., Nováková, E., Husník, F., & Hypša, V. (2017). *Legionella* becoming a mutualist: Adaptive processes shaping the genome of symbiont in the louse *Polyplax serrata*. *Genome Biology and Evolution*, *9*(11), 2946-2957. doi: 10.1093/gbe/evx217; (IF = 3.940)