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

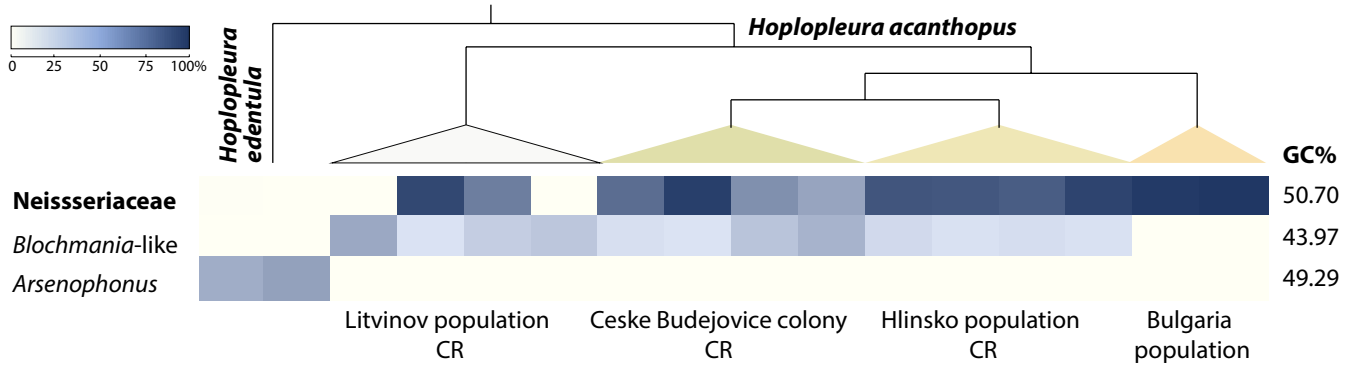


FIGURE 1 Dominant taxa found in *Hoplopleura edentula* and four different populations of *Hoplopleura acanthopus*. The heat map is based on relative abundance of the bacterial OTUs comprising 10 or more percent of the reads within a particular sample. The phylogenetic scheme simplifies the detailed phylogenetic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

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

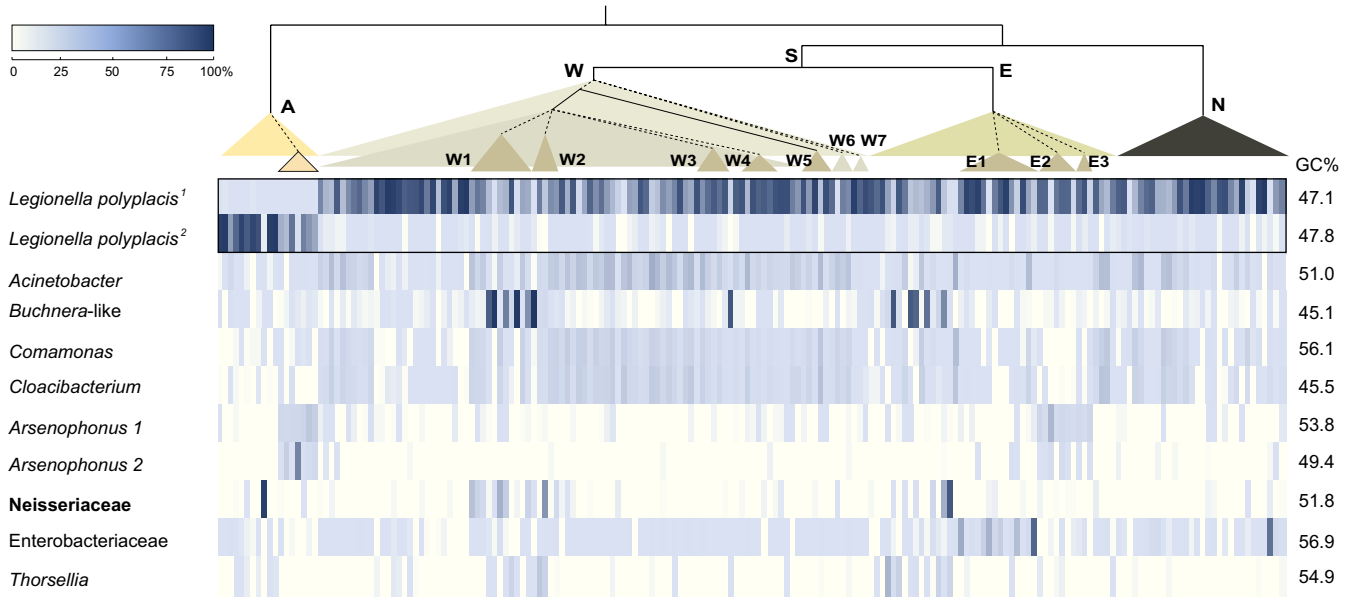


FIGURE 2 Dominant taxa found in *Polyplax serrata* samples from distinct populations. The heat map is based on relative abundance of the bacterial OTUs comprising 10 or more percent of the reads within a particular sample. The phylogenetic scheme, which simplifies the COI based phylogeny of individual 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

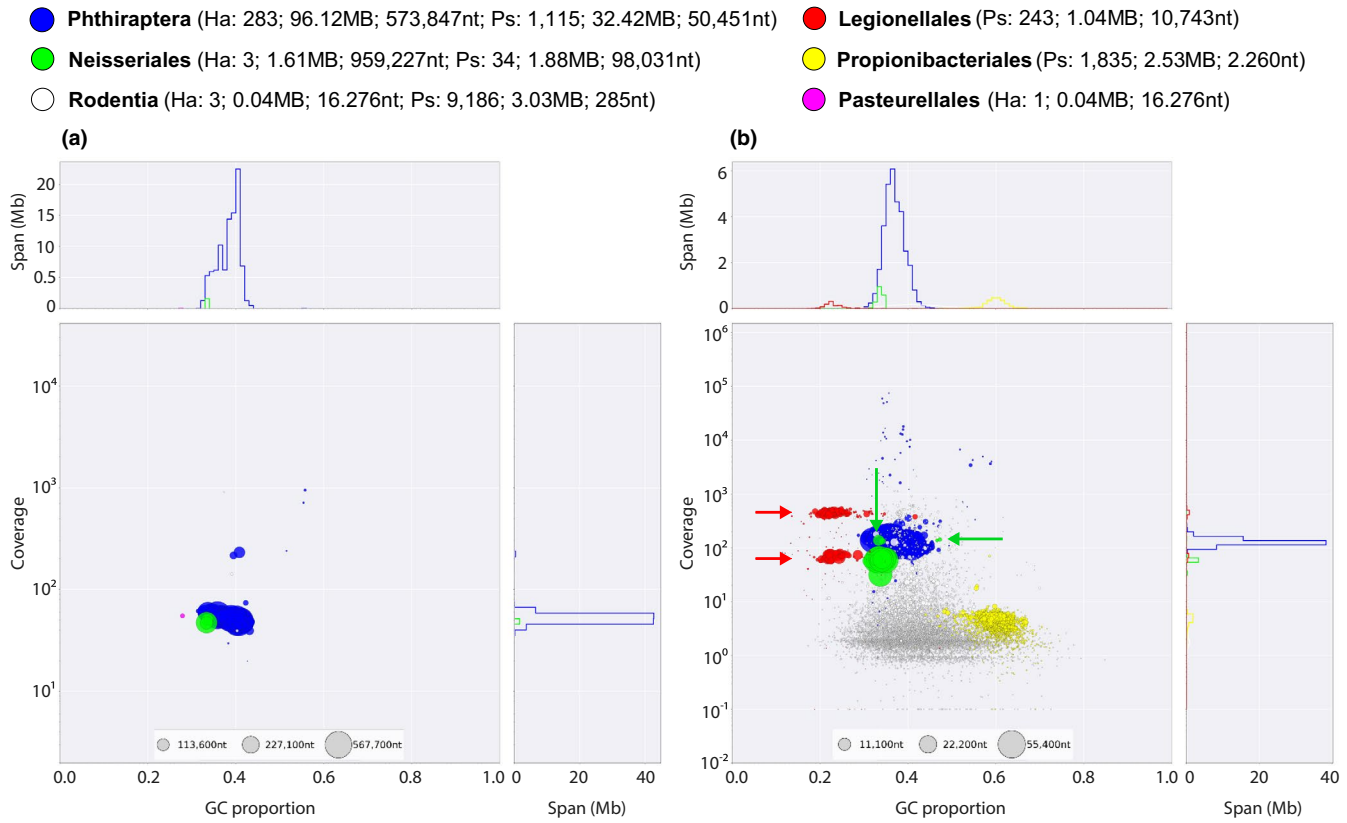


FIGURE 3 Blobplots of *Hoplopleura acanthopus* (a) and *Polyplax serrata* (b) metagenomic assemblies. Contigs are plotted based on their GC content and coverage, scaled by span, and coloured by their taxonomic assignments. For each taxon, the number of assigned contigs, their total length (span), and 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

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

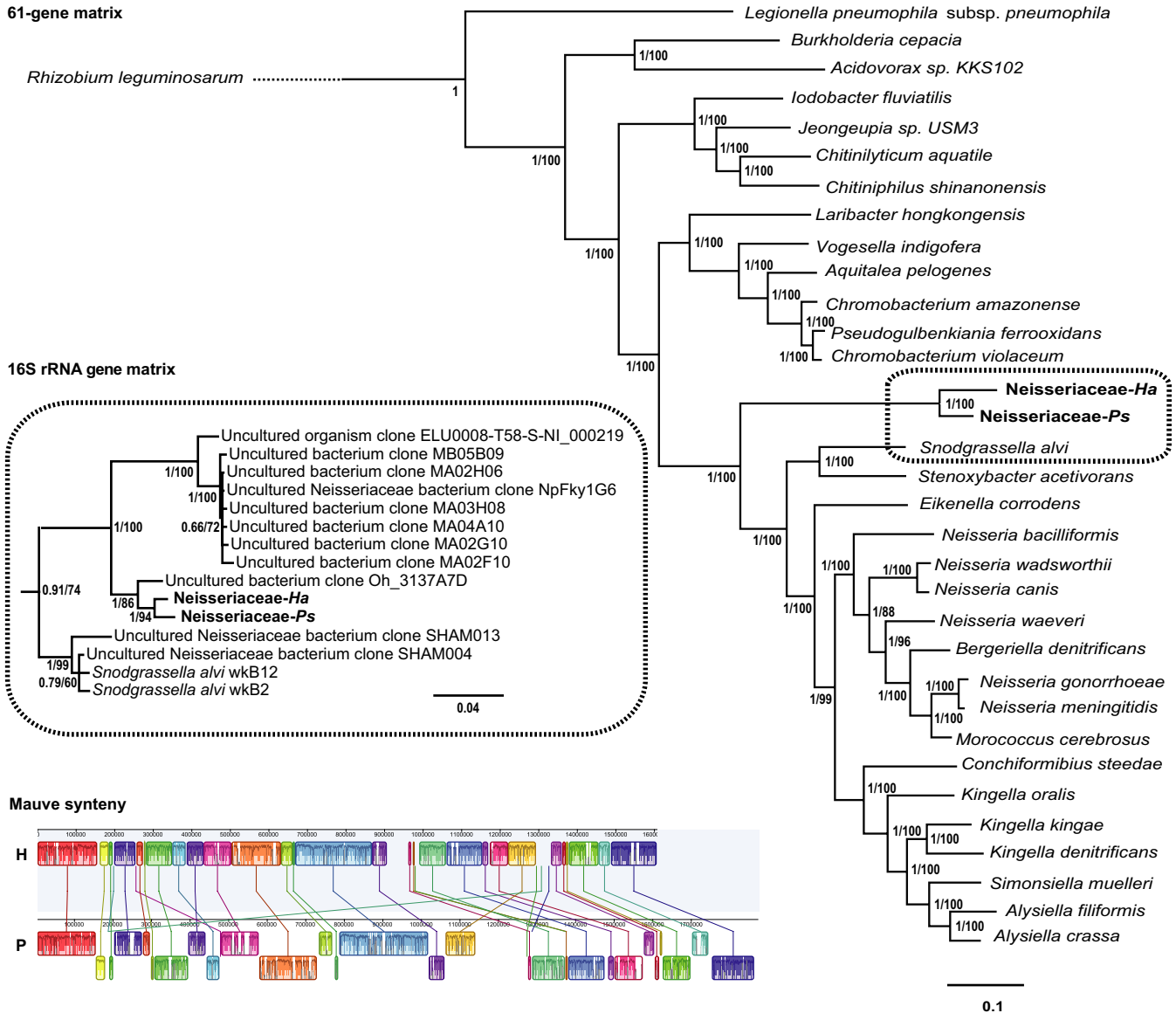
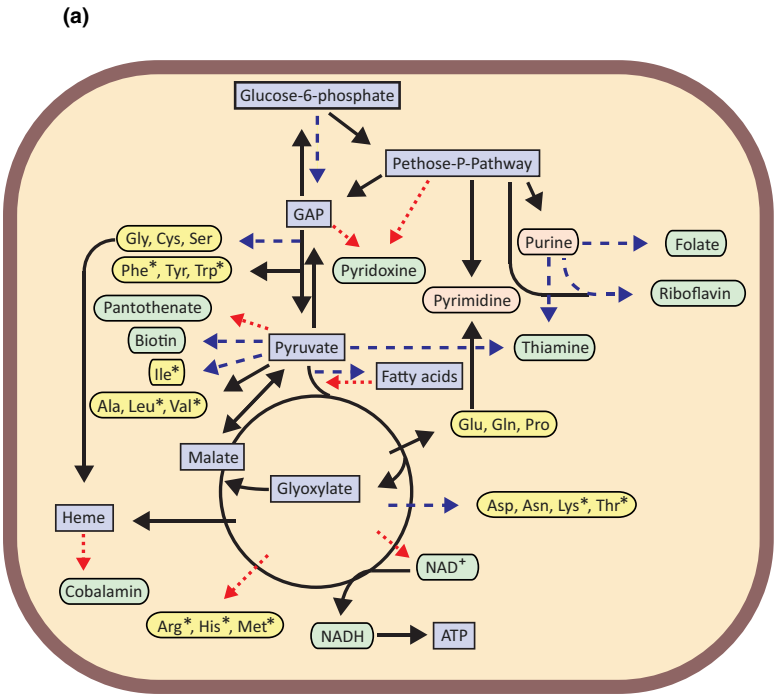


FIGURE 4 Phylogenetic relationships of the two Neisseriaceae-related symbionts. 61-gene matrix: Bayesian analysis of the amino-acid multigene matrix; the numbers at the nodes show posterior probabilities/bootstrap supports obtained by the maximum likelihood analysis in PhyML. 16S matrix: part of the 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)

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



(b)

	Neisseriaceae-Ha	Neisseriaceae-Ps	Puchtella	L. polyplacis	Riesia	Sodalis	Wigglesworthia	Snodgrassella	Neisseria
DNA replication	13	14	12	12	12	15	13	14	15
Base excision repair	9	9	6	6	5	8	6	10	11
Nucleotide excision repair	8	7	3	2	2	6	4	7	8
Mismatch repair	14	15	8	8	9	19	11	16	16
Homologous recombination	19	21	13	8	13	23	15	22	22
Lipopolysaccharide biosynthesis	20	22	12	1	6	19	11	15	20
Peptidoglycan biosynthesis	15	18	13	12	11	15	15	16	16
Amino sugars, nucleotide sugars	15	15	9	2	8	19	12	16	18
UDP-GlcNAc	4	4	4	0	4	4	4	4	4

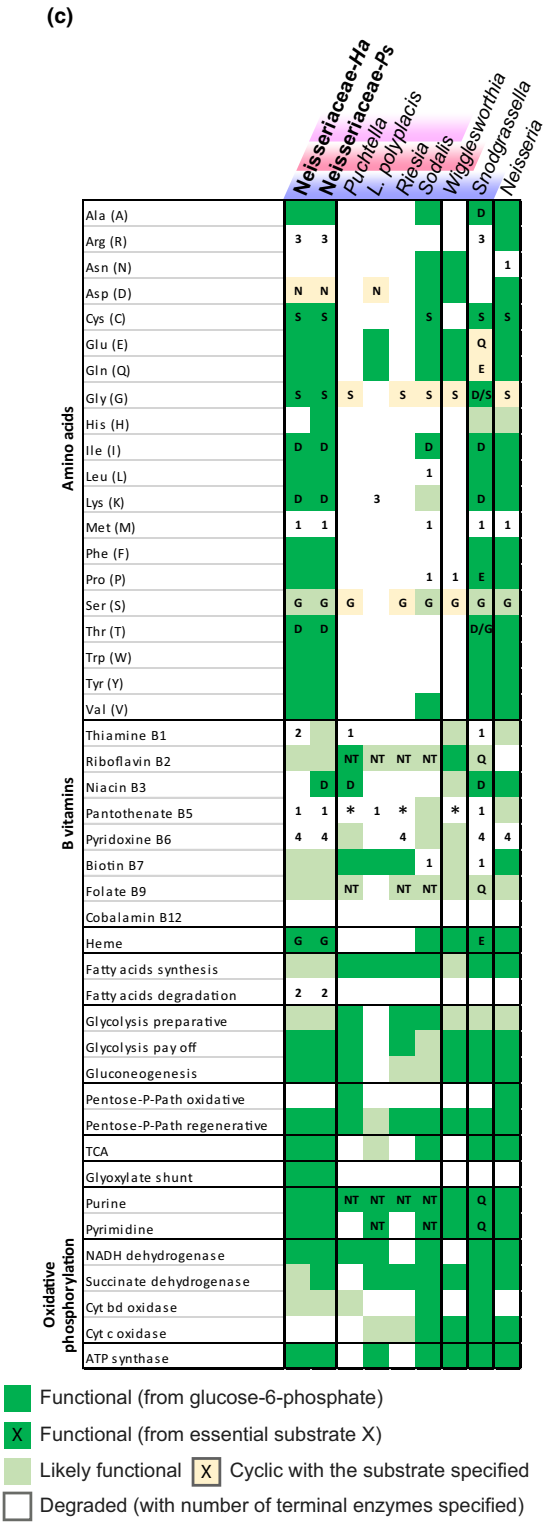


FIGURE 5 (a) Reconstruction of the main metabolic pathways of the Neisseriaecea-*Ha* symbiont. (b) Comparison of gene content (shown as number of genes) for important cellular functions in genomes of the two Neisseriaceae-related symbionts and other bacteria. C. Comparison of important metabolic pathways 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

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.

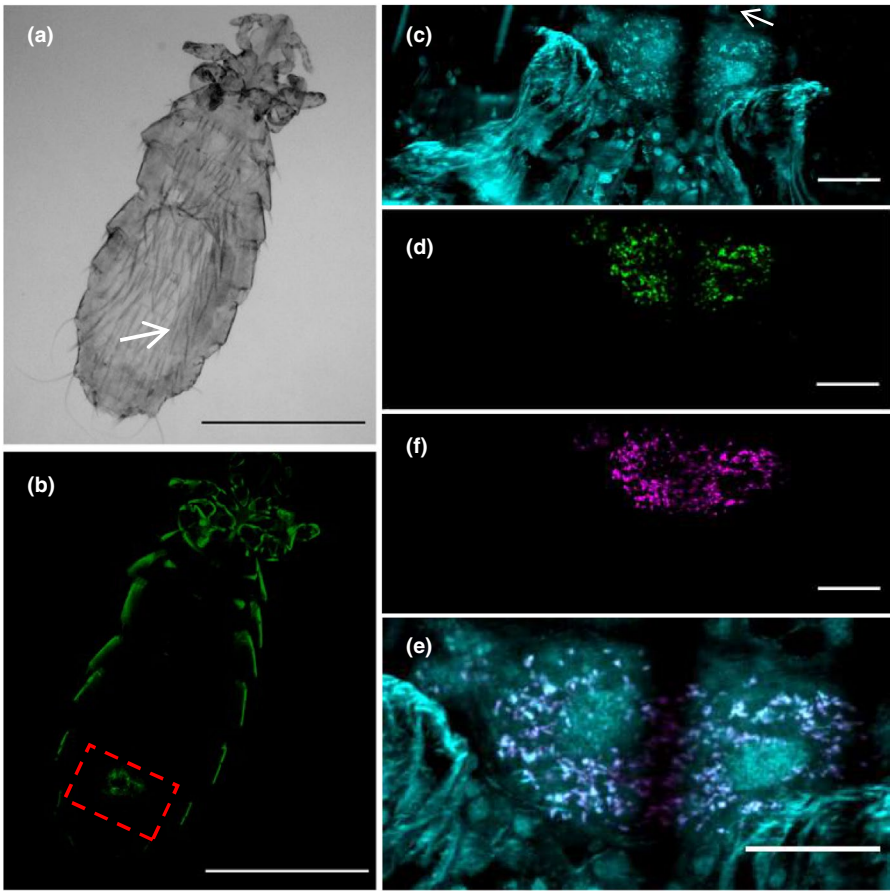


FIGURE 6 Light and confocal microscopy of a FISH-stained female *Hoplopleura acanthopus*. (a) Light microscopy image showing the louse body containing a developing egg (the dark area on the right side of the louse body marked by white arrow). (b) Hybridization signal for the Neisseriaceae specific probe 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

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;00:1–19. https://doi.org/10.1111/mec.15866