

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

Jana Říhová¹, Giampiero Batani¹, Sonia Rodríguez-Ruano¹, Jana Martinu², Eva Novakova¹, and Václav Hypša¹

¹University of South Bohemia Faculty of Science

²Jihoceska Univerzita v Českých Budejovicích

May 5, 2020

Abstract

Phylogenetic diversity of symbiotic bacteria in sucking lice suggests that lice have experienced a complex history of symbiont acquisition, loss, and replacement during their evolution. By combining metagenomics and amplicon screening across several populations of two louse genera (*Polyplax* and *Hoplopleura*) we describe a novel louse symbiont lineage related to *Neisseria* and *Snodgrassella*, and show its independent origin within dynamic lice microbiomes. While the genomes of these symbionts are highly similar in both lice genera, their respective distributions and status within lice microbiomes indicate that they have different functions and history. In *Hoplopleura acanthopus*, the *Neisseria*-related bacterium is a dominant obligate symbiont universally present across several host's populations, and seems to be replacing a presumably older and more degenerated obligate symbiont. In contrast, the *Polyplax* microbiomes are dominated by the obligate symbiont *Legionella polyplacis*, with the *Neisseria*-related bacterium co-occurring only in some samples and with much lower abundance.

Introduction

An increasing number of studies demonstrate ubiquity and high diversity of insect-associated microbiomes (Douglas, 2015; Engel, & Moran, 2013). 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, 1989). 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 (since 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 co-phylogeny with the host (Chen, Li, & Aksoy, 1999; Clark, Moran, Baumann, & Wernegreen, 2000; Sauer, Stackebrandt, Gadau, Holldobler, & Gross, 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, Clark et al., 2000). 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 are less modified, retain more free-living-like characteristics, and some are supposed to be the intermediate stages of evolution towards obligate symbionts. *Wigglesworthia* 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 theoretical work postulates that after a certain amount of coevolutionary time, the symbiotic bacterium becomes too degenerated and functionally inadequate, and it has to be replaced (or accompanied) by another symbiont. While it would be interesting to see how the 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 relies on metagenomic information and phylogenetic reconstructions, likely 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, Outreman, Mieuzet, & Simon, 2015; Manzano-Marin, Szabo, Simon, Horn, & Latorre, 2017; Meseguer et al., 2017).

Amongst hematophagous (blood-feeding) insects which live exclusively on vertebrate blood, sucking lice of the order Anoplura, with more than 500 spp. (Light, Smith, Allen, Durden, & Reed, 2010), are the most ancient and diversified group. Accordingly, they possess a high diversity of symbiotic bacteria (Allen, Burleigh, Light, & Reed, 2016; Boyd, Allen, de Crécy-Lagard, & Reed, 2014; Boyd et al., 2016; Fukatsu et al., 2009; Hypsa & Krizek, 2007). Depending on interpretation, the 16S rRNA gene-based phylogenies for the available taxa suggest 5-6 independent symbiotic lineages. However, none of them is a universal louse symbiont distributed across the whole order (e.g. like *Buchnera* in aphids). The distribution of louse symbionts suggests a relatively recent origin of each 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 has been investigated. The actual number of symbiotic lineages is therefore likely to be much higher. Of the currently known lineages, genomic data are only available for four; three of them showing clear signatures of P-symbionts: *Riesia* spp., *Puchtella pedicinophila*, and *Legionellapolyplacis* (Table 1). Correspondingly, each of these lineages has been found in two to four related host species as a result of co-phylogenetic 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 diversity and distribution of the known symbionts in sucking lice thus indicate that this insect group has been undergoing particularly dynamic acquisition, loss, and replacement of symbionts. In this study, we analyze the background of these processes by combining genomic and amplicon approaches across several populations of the louse genera *Polyplax* and *Hoplopleura*. We reveal a new symbiotic lineage related to the genera *Neisseria* and *Snodgrassella* (the latter being a symbiont of bees). We show that these bacteria established their symbiotic relationships independently with the two louse genera, and we prove their intracellular localization in host's bacteriocytes. Based on the phylogeny-dependent diversity of the microbiome profiles, we suggest rapid microbiome changes at the host population level, possibly underlying the dynamic processes of symbiont acquisition, loss, and replacement in these blood sucking insects.

Materials and Methods

DNA template preparation

Polyplax serrata lice (n=25) were collected from yellow-necked wood mice (*Apodemus flavicollis*) captured in Czech Republic (Struzna) and Germany (Baiersbronn) during 2011. *Hoplopleura acanthopus* lice (n=40) were obtained from common voles (*Microtus arvalis*) trapped in Czech Republic (Hlinsko) during 2014. All samples were stored in 96% ethanol at -20°C. Total DNA was extracted from whole louse abdomens using QIAamp DNA Micro Kit (Qiagen, Hilden, Germany) according to manufacturer protocol. DNA quality was assessed by gel electrophoresis and its' concentration measured with Qubit High sensitivity kit.

Genome sequencing and assembly

We sequenced the *Polyplax* lice pooled sample on one lane of Illumina HiSeq2000 (GeneCore, Heidelberg, Germany) 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 dataset 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 the 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 and their origin was further confirmed by blast analyses of individual genes as specified below.

To sequence a complete *Hoplopleura acanthopus* lice metagenome, we employed Illumina MiSeq (GeneCore, Heidelberg, Germany) 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 5 *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, D’Hert, Schultz, Cruets, & Van Broeckhoven, 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 (for details see SupplementaryInformation/SupplementaryText).

We annotated genomes of both *Neisseria* -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 WNLJ000000000 (draft genome of *Polyplax serrata* symbiont in 39 contigs). To assess phylogenetic origin of individual genes, we blasted each gene against the nr database and categorised its putative origin based on the returned hits (for details see SupplementaryInformation/SupplementaryText).

Genome comparison

Genome synteny was analyzed using ProgressiveMauve (Darling, Mau, & Perna, 2010). The complete closed genome of the *Neisseria* -related symbiont from *H. acanthopus* was aligned to the 39 contigs of the *Neisseria* -related symbiont from *P. serrata*, yielding 27 Locally Collinear Blocks (Figure 1). The resulting synteny regions are explored in more detail in SupplementaryData1. To provide a comparison of selected metabolic pathways, we extended the analysis from our previous work on *Legionella polyplacis* (Rihova, Novakova, Husnik, & Hypsa, 2017) which followed the selection of genes published by Moran, McCutcheon and Nakabachi (2008). The comparison was based on the RAST annotation for the two *Neisseria* -related symbionts, and the KEGG database (Kanehisa, Sato, Kawashima, Furumichi, & Tanabe, 2016) for our other bacteria.

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 from 2014-2018. Details on sample origin, including the sampling locations, GPS coordinates, and rodent host species, are provided in SupplementaryData2. DNA templates were extracted from each individual using QIAamp DNA Micro Kit (Qiagen, Hilden, Germany). Two independent multiplexed 16S rRNA gene amplicon libraries were built for each of the louse genera. Detailed description of the library designs and MiSeq sequencing is provided in SupplementaryInformation/SupplementaryText. Amplicon data for the lice microbiomes (demultiplexed, quality filtered and merged) are available at <https://www.ebi.ac.uk/ena/data/view/PRJEB35541>.

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 and clustered into OTUs (operational taxonomic units) according to the workflow implementing USEARCH v9.0.1001 scripts. All putative contaminants found in negative controls were discarded and highly abundant, potentially symbiotic, OTUs were identified. Details

on the read processing, decontamination and OTU filtering, including the accession links to the raw data, are provided in the SupplementaryInformation/SupplementaryText.

Phylogeny

Phylogenetic analysis of the *Neisseria* -related symbionts was performed on two different matrices, the concatenated “multigene matrix” and the “16S matrix”. To avoid a possible artefact due to HGT, the “multigene matrix” was composed of 10 genes with reliably supported origin within Neisseriales (the genes selected arbitrarily from the blast category “Neisseriales”; details on the assignment to the categories are provided in SupplementaryInformation/SupplementaryText) which were present in both genomes. 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 outgroups (SupplementaryData3). For each gene, the sequences were aligned in MAFFT v7.450 using the E-INS-i setting (Kato, Misawa, Kuma, & Miyata, 2002). Ambiguously aligned positions and divergent blocks were discarded using Gblocks v. 091b (Castresana, 2000). The LG +G+I was determined as best fitting model for all matrices by Akaike information criterion (AIC) using smart model selection of PhyML (Lefort, Longueville, & Gascuel, 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 “multigene matrix”. Bayesian inference of the “multigene matrix” was conducted in MrBayes v3.2.5 using LG +G+I evolutionary model (Ronquist et al., 2012). Four chains were run for 20 000 000 generations with sampling frequency set to 1 000 generations. Convergence was checked in Tracer v1.6.0. (Rambaut, Drummond, Xie, Baele, & Suchard, 2018).

The “16S matrix” was designed with the aim to obtain 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 the GenBank (SupplementaryData3). 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 our “multigene matrix” and analyzed by maximum likelihood (ML) and Bayesian inference (BI). The evolutionary models best fitting to dataset were selected according to the Akaike information criterion (AIC) using jModelTest2 v. 2.1.10 (Darriba, Taboada, Doallo, & Posada, 2012). ML analysis and 100 bootstrap replicates were performed in PhyML (Guindon & Gascuel, 2003) using selected TN93 +G+I evolutionary model. BI analysis was performed in MrBayes v. 3.2.5, using GTR +G+I substitution model running four chains for 10 000 000 generations and checked for convergence as was previously described.

Phylogenetic trees of the hosts, used as a background for the microbiome diversity, were reconstructed as described in SupplementaryInformation/SupplementaryText.

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 sulfate, SDS, and 30% formamide) at 46°C for 10 min, pre-hybridized 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 (St. Louis, Missouri, USA) and used at a concentration of 2.5 ng µl⁻¹ for Cyanine 5 (Cy5)-labelled EUB338 (5’- GCTGCCTCCCGTAGGAGT -3’), targeting all bacteria (Amann, Krumholz, & Stahl, 1990) and 6-carboxyfluorescein (6-Fam)-labelled beta-572 (5’- TTAACCGTCTGCGCTCGCTT -3’), targeting the family Neisseriaceae (Martinson, Moy, & Moran, 2012). We pre-evaluated the required stringency of the hybridization conditions *in silico* using mathFISH (Yilmaz, Parnerkar, & Noguera, 2011). Following hybridization, we washed specimens twice with 400 µl of pre-warmed 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 μ l of 4',6-diamidino-2-phenylindole (DAPI) solution (1 ng μ l⁻¹) in the dark for 10 min, and mounted slides in Mowiol anti-fading medium (Kuraray Europe GmbH, Tokyo, Japan). We captured the fluorescent signals with a laser scanning confocal microscope Olympus FV3000 (Olympus, Tokyo, Japan). We acquired at least three confocal stacks (up to 30 scans per optical slice) at 100x, 400x and 630x magnifications, a color depth of 24 bit and a resolution from 1 to 2 μ m per pixel (depending on the fluorochrome) by investigating multiple regions from each of the three replicate specimens. Resultant images were processed with the ImageJ distribution Fiji (Schindelin et al., 2012; Schneider, Rasband, & Eliceiri, 2012)

Results

Genomes of the Neisseria-related symbionts

The complete closed genome of the obligate *Hoplopleura acanthopussy* symbiont, obtained with a combination of Nanopore and Illumina data, is 1 607 498 bp long with 33.4% GC content. It contains 1 421 protein encoding genes, 9 genes coding rRNAs, and 39 genes coding tRNAs (SupplementaryData1). The 9 rRNA genes represent 3 complete 16S-23S-5S rRNA gene operons, but they are arranged in an unlinked manner known from many other bacteria (Brewer et al., 2019), including other P-symbionts (Munson, Baumann, & Baumann, 1993). The 23S rRNA and 5 rRNA genes are placed in close proximity, while the 16S rRNA gene is separated by long stretches of DNA. The genome draft of *Polyplax serrata* symbiont consisted of 39 contigs which sum to 1 839 042 bp, with the GC content 34.0%. In this assembly, we identified 1 789 protein encoding genes and 37 tRNA genes (SupplementaryData1). Due to the fragmentation of the genome in contigs, the rRNA genes were not reliably assembled and their number remains unclear. To make the following text intelligible, we recategorise *Neisseria*-related symbionts with the terms “*Neisseria*”-*H* (symbiont of *Hoplopleura*) and “*Neisseria*”-*P* (symbiont of *Polyplax*). Genome size and content of these two symbionts, in comparison to other lice symbionts (*Riesia* spp., *Puchtella pedicynophila*, *Legionella polyplacis*, and the symbiont from *Proechinophthirus*), *Snodgrassella* and *Neisseria*, are summarized in Table 1.

When blasted against the NCBI nr database, a large proportion of the genes in both *Neisseria*-related symbionts did not yield Neisseriales within the best hits (see Methods and SupplementaryData1). This is in remarkable contrast to *Legionella polyplacis*, for which 96.3% of the genes were assigned to *Legionella* by the blast search. However, when compared to each other, the two *Neisseria*-related symbionts share a high proportion of their genes and even display a considerable degree of synteny (SupplementaryData1). When aligned with Mauve software the syntenies were placed into 27 Locally Collinear Blocks, the longest blocks extending 100 kb and hundred genes (Figure 1).

Phylogeny

In both phylogenies (the 16S rRNA gene and the multigene) the two *Neisseria*-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 1). 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 “multigene matrix” due to lack of data (SupplementaryInformation/SupplementaryFigure1; detail in Figure 1). All of them have been described as “*Uncultured bacterium*” from human and insect samples (Oteo et al., 2014). In contrast to the “multigene matrix”, the “16S matrix” placed the pair of *Neisseria*-related symbionts into a monophyletic cluster together with *Snodgrassella*.

Amplicon based screening

Sequencing of multiplexed 16S rRNA gene libraries produced high quality amplicon data. Similar numbers of merged 16S rRNA bacterial sequences were retrieved across individual *Polyplax* and *Hoplopleura* samples (21, 205 and 20, 201 on average). For the even mock communities included in the *Hoplopleura* library as positive controls (see Methods), 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. However, the staggered composition of the three mock samples did result in a preferential amplification of one of two dominating taxa, i.e. *Staphylococcus epidermidis* (ATCC 12228). Compared to the mock template where this taxon comprises 44.78% of the total DNA, the average read abundance equaled 70.12%. Therefore, no quantitative analyses of 16S rRNA gene amplicons were used in this study.

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, their symbiotic nature could be clearly demonstrated based on the complete genome characteristics and distribution within the host. For other bacteria, partial 16S rRNA gene amplicons provide two (not entirely independent) kinds of information: taxonomical assignment and GC content. Besides Neisseriaceae, two more OTUs, identified as dominant taxa (see methods for definition) and taxonomically assigned to *Blochmania* and *Arsenophonus* genera, were associated with *Hoplopleura* samples (Figure 2). In addition to the sequence characteristic (GC content of 43.9% and 49.3%), the distribution of these OTUs among the two *Hoplopleura* species and *H. acanthopus* populations points to their symbiotic nature.

For *Polyplax serrata* a comprehensive population-wide amplicon screening revealed (besides *L. polyplacis*) 9 dominant taxa assigned to the genus or family level (Figure 3). The distribution of OTUs with a low GC content, i.e. *Buchnera* (45.1% GC) and *Arsenophonus* 2 (49.4% GC), reflects the genealogy of the host and thus indicates putative obligate coevolving symbionts. For the other taxa, the taxonomic assignment and GC content >50% (with the exception of *Cloacibacterium*) indicates 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 the taxonomical assignment of the OTUs are based on a short sequence and should be interpreted as approximate affiliations rather than precise phylogenetic position, particularly compared to highly derived genomes like *Buchnera* and *Blochmannia*.

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, these putative bacteriocytes formed weakly adherent clusters found above ovarian ampullae. The two probes did not provide entirely overlapping patterns (Figure 4). However, it is difficult to decide whether this difference is due to 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 abdomen in male lice (Supplementary Information/Supplementary Figure 4).

Discussion

The results of the combined genomic and amplicon analyses which we present here, illustrate a positive relationship between the dynamics of louse microbiomes and the observed high diversity of symbiotic bacteria in lice (Allen et al., 2016; Boyd et al., 2016; Boyd et al., 2014; Fukatsu et al., 2009; Hypsa & Krizek, 2007). The two *Neisseria*-related symbionts, which are the primary focus of this study, represent a novel lineage, extending the known phylogenetic span of louse symbionts. Considering their genomic characteristics together with their distribution across the lice populations, we hypothesize that they are bacteria in an early/intermediate stage of evolution towards obligate symbiosis, which established their symbiosis independently within two different genera of lice. Consistent with this view, FISH analysis highlighted their localization within putative bacteriocytes (Figure 4 & Supplementary Information/Supplementary Figure 4), and amplicon screening showed that they are part of a rich microbiomes diversity (Figures 2 & 3), in which some of the other dominant taxa are closely related to the other known obligate symbionts in insects (the most degenerated ones may even be disappearing ancient P-symbionts). This complex picture suggests dy-

namic turnover of the symbiotic bacteria, their frequent acquisitions, losses and replacements even within local louse populations and young phylogenetic lineages.

Genomic characteristics of the Neisseria-related symbionts suggest an intermediate stage of the evolution

We propose that the *Neisseria* -related symbionts are in an early or intermediate state of symbiosis evolution. We base this view on comparisons with other known symbionts. Full genomes are currently available for four lineages of lice-associated symbionts. Three are highly reduced and display strong compositional shift towards AT, typical for many P-symbionts (Table 1). The fourth possesses a large genome and is supposedly a recent acquisition (Boyd et al., 2016). The genomes of the two *Neisseria* -related symbionts can thus be compared to various lice P-symbionts and to “free living” members of Neisseriales. Both “*Neisseria* ”-H and “*Neisseria* ”-P display significantly weaker genome degeneration than the highly reduced *Riesia* , *Puchtella*, and *Legionella* (Boyd et al., 2017; Rihova et al., 2017). They have higher GC content, considerably larger genomes (approximately three times higher number of genes), and consequently more complete metabolic pathways (SupplementaryData1). On the other hand, their genomes are recognizably reduced and the GC content 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). The comparison of selected metabolic pathways shows that the *Neisseria* -related symbionts retain considerably greater numbers of genes in various categories than the reduced genomes of *L. polyplacis* and *R. pediculicola* (SupplementaryData1). This difference is particularly strong in the *Recombination and repair* category (as one example) but also in several amino acid biosynthesis pathways. An interesting example of differences between the two *Neisseria* -related symbionts is the complete histidine pathway in the “*Neisseria*- P” symbiont, entirely lost in the more reduced “*Neisseria*- H” (as well as in *L. polyplacis* and *R. pediculicola*). Another conspicuous difference between the *Neisseria* -related symbionts and the other two louse symbionts regards their capacity to build the cell walls. Similar to the *L. polyplacis* and *R. pediculicola* , the *Neisseria* -related symbionts retain the path for peptidoglycan synthesis, but unlike them, they also possess the genes for penicillin binding protein class A and a complete pathway for lipid A, required for synthesis of lipopolysaccharide. In contrast, both *Neisseria* -related symbionts seem to lack the rod-shape coding genes. Finally, the genomes of both *Neisseria* -related symbionts retain various genes connected to DNA exchange and/or transport, such as mobile elements, type IV pili, and secretion systems (SupplementaryData1). If, as suggested by many of the described genomic features, “*Neisseria* ”-H 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 *Neisseria* -related symbiont. As discussed below, such a putative P-symbiont was indeed detected 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 4).

Distribution and origin of the symbionts in the lice microbiomes

Since the process of genome degradation starts 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 at least by two louse species, *P. serrata* and *P. spinulosa* (Hypsa & Krizek, 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 3). When compared to these two well documented examples of established P-symbionts, the *Neisseria* -related symbionts, with an intermediate degree of the genome degeneration, show more restricted and patchy distribution (Figure 3). Only in *H. acanthopus* were they consistently present (the Neisseriaceae OTU) as the most dominant bacterium (with the exception of two specimens from one population; Figure 2), but were not found in the

two examined specimens of the related species (*H. edentula*). The overall variability of the microbiomes was apparently correlated with the lice genetic background: in Bulgarian samples the Neisseriaceae OTU was the only present bacterium, in other populations of *H. acanthopus* it was accompanied by an unknown bacterium which the blast search affiliated with *Blochmania*, and in the two samples of *H. edentula* the only present OTU corresponded to the genus *Arsenophonus*. Since *P. serrata* is known to harbor the typical obligate P-symbiont *Legionella polyplacis* (Rihova et al., 2017), we screened this louse more extensively across several populations and genetic lineages. The results confirmed a ubiquitous presence of *L. polyplacis*, which in most cases was the most abundant OTU, and only occasional co-occurrence of the Neisseriaceae OTU (Figure 3). The split of *L. polyplacis* into two different OTUs correlated to the host's phylogeny, reflects 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*; (Hypsa & Krizek, 2007). Three additional OTUs in *P. serrata* microbiomes show affinity to known insect symbionts, the *Buchnera* OTU and two *Arsenophonus* OTUs. Based on their genetic divergence and the differences in their GC content, the two *Arsenophonus* OTUs seem to represent two different lineages.

In respect 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 the *Buchnera* OTU from *P. serrata* and the *Blochmania* OTU from *H. acanthopus* seem to represent strongly derived symbiotic genomes (Figure 2 & 3), which blast-assigned taxonomy reflects the 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. However, strong genome reduction, deduced from the GC content of the 16S rRNA gene amplicon, suggests that they may represent the scattered remains of ancient symbionts, now retreating from the host's population and replaced with more recent acquisitions. Interestingly, the FISH analysis shows that apart from the "*Neisseria* -H" symbiont, the bacteriocytes of *H. acanthopus* harbor another bacterium (Figure 4). 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 symbionts acquisitions/replacements make sucking lice an interesting model for studying the conditions and processes in early stage of symbiogenesis. Considering the distribution patterns and the low degree of genome modifications in the two *Neisseria*-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 most parsimonious explanation is thus an independent origin of the symbiosis in each louse genus. 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 co-speciation 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-*Neisseria* 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 *Neisseria* is an uncultured bacterium described from a flea *Oropsylla hirsuta* (Jones, McCormick, & Martin, 2008), for which no other information is currently available. It is interesting to note that, similar to *Legionella polyplacis*, the *Neisseria*-like 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, Trinachartvanit, Baimai, & Grubhoffer, 2013; Felsheim, Kurtti, & Munderloh, 2009; Guizzo et al., 2017; Niebylski, Peacock, Fischer, Porcella, & Schwan, 1997; Noda, Munderloh, & Kurtti, 1997). For one of the tick symbionts, *Francisella*-like bacterium, the origin from mammalian pathogen was recently suggested by Gerhart, Moses and Raghavan (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*). A more detailed high throughput screening of different groups of sucking lice is now needed to assess significance of these two ecological groups of bacteria as a source of nutritional symbionts, and to study the processes during the early symbiogenesis.

Acknowledgments

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.

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. doi: 10.1007/s12223-013-0222-1
- Aksoy, S. (2000). Tsetse - A haven for microorganisms. *Parasitology Today*, 16 (3), 114-118. doi: 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. doi: 10.7717/peerj.2187
- 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. doi: 10.1128/jb.172.2.762-770.1990
- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.
- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., ... Meyer, F. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, 9 (1), 75. doi: 10.1186/1471-2164-9-75
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., ... Pyshkin, A. V. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19 (5), 455-477. doi: 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. doi: 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. doi: 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. doi: 10.1073/pnas.1421388112
- 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. doi: 10.1038/s41396-019-0552-3
- 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* (Phthi-

raptera: Anoplura). *Applied and Environmental Microbiology*, 82 (11), 3185-3197. doi: 10.1128/AEM.00282-16

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

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

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. doi: 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. doi: 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. doi: 10.1007/PL00006444

Clark, M. A., Moran, N. A., Baumann, P., & Wernegreen, J. J. (2000). Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. *Evolution*, 54 (2), 517-525. doi: 10.1111/j.0014-3820.2000.tb00054.x

Darling, A. E., Mau, B., & Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS One*, 5 (6), e11147. doi: 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), 772-772. doi: 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. doi: 10.1093/bioinformatics/bty149

Doudoumis, V., Blow, F., Saridaki, A., Augustinos, A., Dyer, N. A., Goodhead, I., ... & Parker, A. 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. doi: 10.1038/s41598-017-04740-3

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

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

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

Engel, P., & Moran, N. A. (2013). The gut microbiota of insects—diversity in structure and function. *FEMS Microbiology Reviews*, 37 (5), 699-735. doi: 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. doi: 10.1371/journal.pone.0008361

Fukatsu, T., Hosokawa, T., Koga, R., Nikoh, N., Kato, T., Hayama, S. I., ... & Tanaka, I. (2009). Intestinal endocellular symbiotic bacterium of the macaque louse *Pedicinus obtusus*: distinct endosymbiont origins

in anthropoid primate lice and the old world monkey louse. *Applied Environmental Microbiology* , 75 (11), 3796-3799. doi: 10.1128/AEM.00226-09

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

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. doi: 10.1038/srep33670

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

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. doi: 10.1093/sysbio/syq010

Guizzo, M. G., Parizi, L. F., Nunes, R. D., Schama, R., Albano, R. M., Tirloni, L., ... & Gonzales, S. A. (2017). A *Coxiella* mutualist symbiont is essential to the development of *Rhipicephalus microplus* . *Scientific Reports* , 7 (1), 1-10. doi: 10.1038/s41598-017-17309-x

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

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. doi: 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. doi: 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. doi: 10.1093/nar/gkf436

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... & Thierer, T. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* , 28 (12), 1647-1649. doi: 10.1093/bioinformatics/bts199

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 apicolagen* 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. doi: 10.1099/ijs.0.044875-0

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

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. doi: 10.1186/1471-2148-10-292

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. doi: 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. doi: 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. doi: 10.1002/ece3.4424
- Meseguer, A. S., Manzano-Marín, A., Coeur D'Acier, A., Clamens, A. L., Godefroid, M., & Jousset, 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. doi: 10.1111/mec.13910
- Moran, N. A. (1996). Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences* , 93 (7), 2873-2878. doi: 10.1073/pnas.93.7.2873
- Moran, N. A., McCutcheon, J. P., & Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics* , 42 , 165-190. doi: 10.1146/annurev.genet.41.110306.130119
- 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. doi: 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. doi: 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. doi: 10.1128/AEM.63.10.3926-3932.1997
- Oteo, 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. doi: 10.1186/s13071-014-0455-0
- 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. doi: 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. doi: 10.1093/gbe/evx217
- Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... & 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. doi: 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. doi: 10.1099/00207713-50-5-1877
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., ... & Tinevez, J. Y. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods* , 9 (7), 676-682. doi: 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. doi: 10.1038/nmeth.2089
- 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. doi: 10.1128/AEM.01733-10

Data accessibility

The genome assemblies are available from GenBank under the accession numbers CP046107 (closed genome of the symbiont from *H. acanthopus*) and WNLJ000000000 (draft genome of *Polyplax serrata* symbiont in 39 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* sp. (sequences of COI genes) are available in supplementary material.

Author Contributions

V.H., E.N. and J.R. designed the study. J.R. and J.M. collected the samples. All authors contributed to data analyses. V.H., E.N., J.R. and G.B. wrote the original draft. All authors participated in discussions and revised the manuscript. V.H. supervised the whole project.

Tables and Figures

Table 1. Comparison of main genomic characteristics. Louse symbionts highlighted by grey background.

Figure 1. Phylogenetic relationships of the two *Neisseria*-related symbionts. Multigene matrix: Bayesian analysis of the 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 16S matrix showing relationships of the two *Neisseria*-related symbionts to several Uncultured bacteria (see SupplementaryInformation/SupplementaryFigure1 for the complete tree); the numbers at the nodes show posterior probabilities/bootstrap supports obtained by the Maximum Likelihood analysis in PhyML. Mauve synteny: an overview of strong synteny between the two symbionts; H - “*Neisseria*”-H, P - “*Neisseria*”-P (see SupplementaryData1 for a complete list of the genes).

Figure 2. Highly abundant taxa found in *H. edentula* and four different populations of *H. acanthopus*. Detailed phylogenetic relationships of the hosts are shown in the SupplementaryInformation/SupplementaryFigure2.

Figure 3. Highly abundant taxa found in *Polyplax serrata* samples from distinct populations. The phylogenetic scheme simplifies the COI based phylogeny of individual samples provided in SupplementaryInformation/SupplementaryFigure3. Designation of the branches is based on mtDNA structure described in the study by Martinu et al. [50]: A = lineage specific to *Apodemus agrarius*, S = lineage specific to *Apodemus flavicollis* (W = west sublineage, E = east sublineage), N = nonspecific lineage from *A. flavicollis* and *A. sylvaticus*. The numbers for *Legionella polyplacis* designate two different OTUs, reflecting evolutionary changes accumulated after the split of *Polyplax serrata* lineages.

Figure 4. Light and confocal microscopy of a FISH-stained female *H. acanthopus*. (a) Light microscopy image showing the louse body containing a developing egg (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 panel 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 panel c, d and e, respectively, and were combined in the merged color image (f), with the *Neisseria*-related symbionts appearing white. DAPI staining in panel c defines also the localization of the bacteriocytes above ovarian ampullae.

Scale bars: (a, b) 500 μ m, (c, d, e and f) 20 μ m.

Hosted file

Table 1.docx available at <https://authorea.com/users/297294/articles/426418-a-new-symbiotic-lineage-related-to-neisseria-and-snodgrassella-arises-from-the-dynamic-and-diverse-microbiomes-in-sucking-lice>



