Title: Patterns of microbiome variation among infrapopulations of permanent bloodsucking parasites

Authors: Jorge Doña[1, 2]*, Stephany Virrueta Herrera[1], Tommi Nyman[3], Mervi Kunnasranta[4,5], and Kevin P. Johnson[1]*

Author details:

[1]Illinois Natural History Survey, Prairie Research Institute, University of Illinois at Urbana-Champaign, 1816 S. Oak St., Champaign, Illinois 61820, USA.

[2]Departamento de Biología Animal, Universidad de Granada, Granada, 18001, Spain.

[3]Department of Ecosystems in the Barents Region, Norwegian Institute of Bioeconomy Research, Svanhovd 35, 9925 Svanvik, Norway.

[4]Department of Environmental and Biological Sciences, University of Eastern Finland, Yliopistokatu 7, 80101 Joensuu, Finland.

[5]Natural Resources Institute Finland, Joensuu, Finland.

Email addresses: Jorge Doña: jorged@illinois.edu; Stephany Virrueta Herrera: svirru2@illinois.edu; Tommi Nyman: tommi.nyman@nibio.no; Mervi Kunnasranta: mervi.kunnasranta@uef.fi; Kevin P. Johnson: kdpjohnso@illinois.edu

*Correspondence to: jorged@illinois.edu, kpjohnso@illinois.edu.

Abstract:

While interspecific variation in microbiome composition can often be readily explained by factors such as host species identity, there is still limited knowledge of how microbiomes vary at scales lower than the species level (e.g., between individuals or populations). Here, we evaluated variation in microbiome composition of individual parasites among infrapopulations (i.e., populations of parasites of the same species living on a single host individual). To address this question, we used genome-resolved and shotgun metagenomic data of 17 infrapopulations (balanced design) of the permanent, bloodsucking seal louse *Echinophthirius horridus* sampled from individual Saimaa ringed seals *Pusa hispida saimensis*. Both genome-resolved and read-based metagenomic classification approaches consistently show that parasite infrapopulation identity is a significant factor that explains both qualitative and quantitative patterns of microbiome variation at the intraspecific level. This study contributes to the general understanding of the factors driving patterns of intraspecific variation in microbiome composition, especially of bloodsucking parasites, and has implications for understanding how well-known processes occurring at higher taxonomic levels, such as phylosymbiosis, might arise in these systems.

Keywords: *genome-resolved metagenomics, host-symbiont, intraspecific variation, lice, microbiota, shotgun metagenomics, symbiont*.

Introduction

Patterns of inter- and intraspecific variation in microbiome composition of animals have received much attention because the microbiome may influence many biological processes that have considerable effects on the host (Clemente et al. 2012; Le Chatelier et al. 2013; Rothschild et al. 2018; Rudman et al. 2019; Velazquez et al. 2019). For instance, particular microbiome compositions have been found to drive genomic adaptation (Rudman et al. 2019) or to confer protection against pathogens (Velazquez et al. 2019).

In general, both stochastic (e.g., dispersal, or ecological drift) and deterministic (e.g., host immunological regulation, or microbe–microbe interactions) processes operate across multiple spatial scales to shape the composition of animal microbiomes (Adair and Douglas 2017; Kohl 2020). In particular, among the many determinants shaping microbiome composition, host species identity has been repeatedly identified as a key factor determining the composition of animal microbiomes (Brooks et al. 2016; Mazel et al. 2018; Nishida and Ochman 2018; Lutz et al. 2019; Knowles et al. 2019; Lim and Bordenstein 2020; Song et al. 2020). In other words, microbiomes of individuals of the same species tend to be more similar than to those of another species. This pattern is generally the result of filtering microbial taxa by the host (e.g., through host diet, habitat, or immune system, Adair and Douglas 2017) or result from host-microbe coevolution (Lim and Bordenstein 2020). When this process exhibits phylogenetic signal, the pattern is known as phylosymbiosis (i.e., microbial community relationships that recapitulate the phylogeny of their host, Brucker and Bordenstein 2013; Brooks et al. 2016; Lim and Bordenstein 2020). Nonetheless, several aspects of the variation of animal microbiomes are yet to be better understood (Lim and Bordenstein 2020). In particular, for non-human animals, there is still much to learn about how microbiomes vary at scales below the species level, such as between populations (Blekhman et al.

2015; Kohl et al. 2018; Rothschild et al. 2018; Campbell et al. 2020; Fountain-Jones et al. 2020) or ecotypes (Agany et al. 2020).

An area of focus on understanding intraspecific variation in microbiome composition has been bloodsucking parasites. In these parasites, previous studies have consistently found a major role of the host species in shaping microbiome composition in the parasites (Osei-Poku et al. 2012; Zhang et al. 2014; Swei and Kwan 2017; Zolnik et al. 2018; Landesman et al. 2019; Lee et al. 2019; Muturi et al. 2019). However, in ticks (*Ixodes scapularis*), host individual identity of the blood meal was even more important than host species identity in explaining microbiome composition (Landesman et al. 2019). These results suggested that individual host identity of the blood meal might be an important factor that shapes parasite microbiomes at the intraspecific level (Landesman et al. 2019). In theory, microbiomes of individual bloodsucking parasites could vary due to: 1) the individual parasite immune system that may impose selection on different bacterial taxa (Blekhman et al. 2015; Suzuki et al. 2019), 2) differences in the source of the blood meal that may transfer or disperse particular bacterial taxa, or modulate bacteria by creating specific conditions during digestion (Rothschild et al. 2018), 3) microbe–microbe interactions (Hassani et al. 2018), and 4) stochastic processes (e.g., ecological drift) (Lankau et al. 2012). However, for most species, and for bloodsucking parasites in particular, the nature of intraspecific variation in microbiomes and the relative importance of factors shaping this variation remain understudied.

Sucking lice (Phthiraptera: Anoplura) are permanent blood-feeding ectoparasites that live in the fur or hairs of mammals. The sucking lice of pinnipeds (seals, sea lions, and walrus) are of particular interest because of their need to adapt to the aquatic lifestyle of their hosts (Durden and Musser 1994; Leonardi et al. 2013). There is evidence that the sucking lice of seals and sea lions

have codiversified with their hosts (Kim 1971, 1975, 1985; Leonardi et al. 2019). In addition, the sucking lice of pinnipeds represent an interesting system in which to study the variation in microbiome composition and the drivers of this variation at an intraspecific level because: 1) these lice have well defined, isolated populations (infrapopulations) on individual seal hosts, due to an expected low rate of horizontal dispersal among host individuals, which is only possible during the seals' haul-out periods on land or ice (Kim 1985; Leonardi et al. 2013, 2019); and 2) these lice feed only upon the blood of their host (Snodgrass 1944; Kim 1985), so that it can be assumed that individuals from the same infrapopulation feed upon "exactly" the same resource (i.e., the blood of the individual seal on which they occur).

Here, we used genome-resolved approaches (the construction of draft microbial genomes from short-read shotgun sequencing data; Bowers et al. 2017; Uritskiy et al. 2018) and metagenomic classification tools (taxonomic classification of individual sequencing reads; Menzel et al. 2016) to infer patterns of microbiome variation among individuals of the sucking seal louse *Echinophthirius horridus* (von Olfers, 1816) inhabiting individual Saimaa ringed seals *Pusa hispida saimensis* (Nordquist, 1899). Our sampling design, involving analysis of two individual lice from each of 17 seals, allowed us to evaluate the degree to which variation in microbiome composition among individual lice is explained by the infrapopulation (the identity of the seal host).

Materials and Methods

Sampling, DNA extraction, and sequencing

Thirty-four individual lice were sampled from 17 individual Saimaa ringed seals (*Pusa hispida saimensis*), which is an endemic endangered landlocked subspecies of the ringed seal living in

106 freshwater Lake Saimaa in Finland (e.g., Nyman et al. 2014). Individual lice were collected from
107 seals found dead or from seals that were live-captured for telemetry studies (e.g., Niemi et al.
108 2019), and placed in 2-ml screw-cap tubes with 99.5% ethanol. Lice from a single seal individual
109 were put in the same tube. Prior to DNA extraction, each louse individual was rinsed with 95%
110 ethanol and placed alone in a new sterile vial; then, the remaining ethanol was evaporated at room
111 temperature.

112 Whole lice were ground up individually, and genomic DNA was extracted using the Qiagen
113 QIAamp DNA Micro Kit (Qiagen, Valencia, CA, U.S.A.). The standard protocol was modified
114 so that specimens were incubated in ATL buffer and proteinase K at 55 (insert degree) C for 48 h
115 instead of the recommended 1 – 3 h, as well as by substituting buffer AE with buffer EB (elution
116 buffer). This was done to ensure maximal yield (greater than 5 ng) of DNA from each louse. Each
117 DNA extract was quantified with a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, U.S.A.)
118 following the manufacturer's recommended protocols.

119 Shotgun genomic libraries were prepared from the extracts with Hyper Library construction kits
120 (Kapa Biosystems, Wilmington, MA, U.S.A.), and the libraries were quantitated by qPCR and 150
121 bp pair-end sequenced on one lane of an Illumina NovaSeq 6000 sequencer (Albany, New York).
122 FASTQ files from sequence data were generated and demultiplexed with bcl2fastq v.2.20. All
123 library preparations, sequencing, and FASTQ file generation were carried out at the Roy J. Carver
124 Biotechnology Center (University of Illinois, Urbana, IL, U.S.A.). Raw reads were subsequently
125 deposited to the NCBI GenBank SRA database (Table S1).

126 Metagenomic analyses

127

For the genome-resolved metagenomic analyses, we used the metaWRAP v1.1.5 pipeline (Uritskiy et al. 2018) along with all the recommended databases (i.e., Checkm_DB, NCBI_nt, and NCBI_tax). We used the metaWRAP Read_qc module with default parameters to quality trim the reads and to de-contaminate each sample from host reads. For decontamination, we ran a de-novo genome assembly of an individual louse of the same species, not included in this study, and with a high amount of sequencing data ("Echor52") in Abyss v2.0.1 (Jackman et al. 2017). Finally, we filtered out all non-bacterial reads from the contig file using Blobtools v1.0.1 (Laetsch and Blaxter 2017) and used this file to decontaminate all the other samples with the metaWRAP Read_qc module. Next, we co-assembled reads from all the samples with the metaWRAP Assembly module (--usemetaspades option) (Nurk et al. 2017). For this assembly, and because of memory limitations, we used BBNorm (sourceforge.net/projects/bbmap/) before assembly to reduce the coverage of the concatenated FASTQ file to a maximum of 100X and to discard reads with coverage under 3X. We binned reads with the metaWRAP Binning module (--maxbin2 --concoct --metabat2 options) (Alneberg et al. 2014; Wu et al. 2016; Kang et al. 2019) and then consolidated the resulting bins into a final bin set with both metaWRAP's Bin_refinement module (-c 50 -x 10 options) and the Reassemble_bins module. We quantified the bins resulting from the Bin_refinement module with Salmon (Patro et al. 2017) using the Quant_bins module with default parameters. Finally, we classified bins using the Classify_bins module. This module uses Taxator-tk, which gives highly accurate but conservative classifications (Dröge et al. 2015). Accordingly, we also uploaded our final metagenome-assembled genomes (MAGs) to MiGA for a complementary analysis to determine the most likely taxonomic classification and novelty rank of the bin (Rodriguez-R et al. 2018). We used the NCBI Genome database (Prokaryotes; February 26, 2020 version) for this analysis.

For the metagenomic classification of reads, we used the metagenomic classifier Kaiju (Menzel et al. 2016) with Reference database: nr (Bacteria and Archaea; Database date: 2017-05-16). We used the default parameters for these analyses – SEG low complexity filter: yes; Run mode: greedy; Minimum match length: 11; Minimum match score: 75; Allowed mismatches: 5. We then converted Kaiju's output files into a summary table at the genus and species level and filtered out taxa with low abundances (<0.1 % of the total reads). We also removed poorly identified taxa because they would artificially increase the similarity between our samples. Specifically, the following taxa were excluded: "NA", "Viruses", "archaeon", "uncultured bacterium", "uncultured Gammaproteobacteria bacterium" (Table S2 and S3).

Lastly, we used Decontam v1.2.1 to filter out bacterial taxa exhibiting known statistical properties of contaminants (Davis et al. 2018). We used the frequency method (*isContaminant* function) which is based on the inverse relationship between the relative abundance of contaminants and sample DNA concentration, and also has been found suitable for samples dominated by host DNA (Willner et al. 2012; Lusk 2014; Salter et al. 2014; Jervis-Bardy et al. 2015; Hooper et al. 2019; McArdle and Kaforou 2020). As input for Decontam analyses, we used the aforementioned total DNA concentration values. Then, as recommended, we explored the distribution of scores assigned by Decontam to assign the threshold according to bimodality between very low and high scores (Davis et al. 2018). For the MAGs matrix, no bimodality was found, and thus we used the 0.1 default value (Fig S1a). None of the MAGs were classified as contaminants, according to Decontam. For Kaiju matrices, a 0.3 threshold value was selected for the species matrix (Fig S1b) and 0.31 for the genus matrix (Fig S1c). Decontam filtered out a single species (*Clostridia* bacterium k32) from the species matrix and two genera (*Cupriavidus* and *Massilia*) from the genus matrix.

Statistical analyses

To visualize similarities of microbiome composition among louse individuals from the same or different individual seal hosts, we constructed non-metric multidimensional scaling (NMDS) ordinations based on Bray–Curtis and Jaccard (binary = T) dissimilarities using the phyloseq v1.26-1 R package (McMurdie and Holmes 2013). For the genome-resolved metagenomic analyses, we used the normalized MAGs compositional matrices resulting from Salmon. Specifically, these MAG counts are standardized to the individual sample size (MAG copies per million reads) and thus allow between-sample comparisons. For the Kaiju analyses, we used the rarefy_even_depth function of phyloseq (without replacement as in the hypergeometric model) to rarefy samples to the smallest number of classified sequences per individual observed (85,513, and 71,267 reads in genus and species matrices, respectively) (Weiss et al. 2017). To assess the influence of individual host identity on the microbiome composition of louse individuals, we conducted a permutational multivariate analysis of variance (PERMANOVA) (Anderson and Walsh 2013; Anderson 2014). PERMANOVA analyses were done using the *adonis2* function in vegan v2.5–4 (Oksanen et al. 2019), based on Bray–Curtis and Jaccard distance matrices with 100 iterations. In PERMANOVA analyses, for the individual host identity factor, our within-group sample size (n=2) was smaller than both the total number of groups (n=17) and the total sample size (n=34). Thus, to account for a potential deviation in F-statistics and R^2 values (Kelly et al. 2015), we wrote an R simulation that randomly subsampled the infrapopulations from which the louse came (5 infrapopulations per iteration). We ran 10 iterations and ran a PERMANOVA analysis for each iteration. Note that, for a few iterations, subsampled samples were too similar and PERMANOVA could not be done. In addition, we ran PERMANOVA analyses to explore additional factors (louse sex: male, female; sequencing lane: 1, 2; and host status: dead, alive) that

may explain variance in microbiome composition. Furthermore, we included significant factors as the first factors of the host identity PERMANOVA models (i.e., to obtain the variance explained by host identity after accounting for the variance explained by that factor). We also restricted the groups in which permutations could be done to only those with the same value of that vector using the strata argument (e.g., for a sample collected from a dead host, and for the host-status factor, permutations could only be done among other dead hosts). Lastly, we ran a Mantel test using the *mantel* function in vegan (method=spearman, permutations=9999) to explore if host locality (i.e., the coordinates in which each host was sampled) correlated with the microbiome composition of louse individuals. For this analysis, we ran 10 iterations of an R simulation in which we randomly subsampled one louse sample for each individual host and ran a Mantel test for each iteration. The following packages were used to produce the plots: ggplot2 v3.1.0.9 (Wickham 2016), grid v3.5.3 (R Core Team 2019), gridExtra v.2.3 (Auguie 2016), ggrepel v0.8.0 (Slowikowski et al. 2019), ggpubr v.0.2.5 (Kassambara 2018), and ggsci v2.9 (Xiao 2018).

Results

From the genome-resolved metagenomics pipeline, 13 high-quality bacterial metagenome-assembled genomes (MAGs) were obtained (Table 1; Fig 1). According to MiGA analyses, 10 of them (77%) likely belong to a species not represented in the NCBI Genome database.

Kaiju analyses recovered a higher diversity of microorganisms than did the genome-resolved approach. These differences are likely because of the quality-filtering parameters used in the genome-resolved metagenomics pipeline (i.e., these taxa may have been discarded because the completeness values of their bins were lower than 50% and/or their contamination values were higher than 10%). Nevertheless, bacterial taxa found in the genome-resolved metagenomic

219 approach were generally found also in Kaiju and with similar relative abundances (Fig 2), and a
220 similar pattern was found also at the species level (Fig S2).

221 Ordination and PERMANOVA analyses show a major role of infrapopulation identity in
222 explaining microbiome composition for both presence–absence and quantitative data. In the
223 genome-resolved metagenomic dissimilarity matrices, most (>84% in all cases) of the variance
224 was explained by infrapopulation identity (PERMANOVA: Bray-Curtis, R^2= 0.857, F=6.419,
225 P=0.001, Fig 3a; Jaccard, R^2= 0.842, F= 5.671, P=0.001; Fig S3a). Results from the simulations
226 were in line with the results of the regular model, and thus support that our results were not biased
227 by the sampling design [PERMANOVA: Bray-Curtis, R^2 (min= 0.65, max= 0.98, mean= 0.78); P
228 (min= 0.001, max= 0.019, n<0.05= 10/10); Jaccard, R^2 (min= 0.66, max= 1, mean= 0.86), P (min=
229 0.001, max= 0.106, n<0.05= 5/7)]. From all the additional factors examined, only host status (i.e.,
230 dead, alive) explained a significant amount of variance [PERMANOVA: Bray-Curtis, Host status:
231 R^2= 0.28, F= 12.72, P= 0.001, Louse sex: R^2= 0.08, F= 0.9, P= 0.554, Sequencing lane: R^2= 0.01,
232 F= 0.38, P= 0.878; Jaccard, Host status: R^2= 0.13, F= 4.93, P= 0.002, Louse sex: R^2= 0.03, F=
233 0.28, P= 0.867, Sequencing lane: R^2= 0, F= -0.01, P= 1; Mantel test, locality, Bray-Curtis: ρ (min=
234 -0.09, max= -0.09, mean= -0.09), P (min= 0.8749, max= 0.8867, n<0.05= 0/10); Jaccard: ρ (min=
235 -0.29, max= -0.29, mean= -0.29), P (min= 0.97742, max= 0.9777, n<0.05= 0/10)]. Including host
236 status in PERMANOVA analyses did not alter the results on the major influence of host identity
237 in explaining microbiome composition (PERMANOVA: Host identity, Bray-Curtis, R^2= 0.57, F=
238 4.58, P= 0.001; Jaccard, R^2= 0.71, F= 5.09, P= 0.002).

239 Similarly, in Kaiju matrices collapsed at the species level, most (>80% in all cases) of the variance
240 was also explained by infrapopulation identity (PERMANOVA: Bray–Curtis, R^2=0.804, F=4.346,
241 P=0.001, Fig 3b; Jaccard, R^2=0.803, F=4.319, P=0.001; Fig S3b). Again, results from simulations

were similar [PERMANOVA: Bray-Curtis, R^2 (min= 0.62, max= 0.88, mean= 0.75); P (min= 0.003, max= 0.058, n<0.05= 9/10); Jaccard, R^2 (min= 0.63, max= 0.95, mean= 0.76), P (min= 0.002, max= 0.09, n<0.05= 9/10)]. Of all the others factors examined, only host status explained a significant amount of variance [PERMANOVA: Bray-Curtis, Host-status: R^2= 0.22, F= 9.03, P= 0.001, Louse sex: R^2= 0.08, F= 0.81, P= 0.564, Sequencing lane: R^2= 0.01, F= 0.35, P= 0.859; Jaccard, Host-status: R^2= 0.21, F= 8.73, P= 0.001, Louse sex: R^2= 0.08, F= 0.88, P= 0.497, Sequencing lane: R^2= 0.01, F= 0.4, P= 0.825; Mantel test, locality, Bray-Curtis: ρ (min= 0.04, max= 0.04, mean= 0.04), P (min= 0.5637, max= 0.5785, n<0.05= 0/10); Jaccard: ρ (min= -0.03, max= -0.03, mean= -0.03), P (min= 0.5337, max= 0.5488, n<0.05= 0/10)]. PERMANOVA analysis accounting for host status did not alter the significance of host identity (PERMANOVA: Bray-Curtis, R^2= 0.52, F= 1.78, P= 0.007; Jaccard, R^2= 0.59, F= 3.37, P= 0.001).

Furthermore, results were consistent when using matrices collapsed at the genus level (>77% of variance explained in all cases) (PERMANOVA: Bray–Curtis, R^2= 0.865, F= 6.804, P= 0.001, Fig S4a; Jaccard, R^2= 0.774, F= 3.634, P= 0.001; Fig S4b). Once again, results from simulations were similar [PERMANOVA: Bray-Curtis, R^2 (min= 0.68, max= 0.96, mean= 0.8); P (min= 0.002, max= 0.073, n<0.05= 9/10); Jaccard, R^2 (min= 0.54, max= 0.86, mean= 0.73), P (min= 0.003, max= 0.061, n<0.05= 9/10)]. Additionally, of all the others factors examined, only host status explained a significant amount of variance [PERMANOVA: Bray-Curtis, Host-status: R^2= 0.3, F= 14, P= 0.001, Louse sex: R^2= 0.05, F= 0.51, P= 0.851, Sequencing lane: R^2= 0.01, F= 0.39, P= 0.753; Jaccard, Host-status: R^2= 0.18, F= 7.19, P= 0.002, Louse sex: R^2= 0.07, F= 0.75, P= 0.53, Sequencing lane: R^2= 0.01, F= 0.40, P= 0.75; Mantel test, locality, Bray-Curtis: ρ (min= 0.09, max= 0.09, mean= 0.09), P (min= 0.7198, max= 0.7344, n<0.05= 0/10); Jaccard: ρ (min= 0.02, max= 0.02, mean= 0.02), P (min= 0.4043, max= 0.4246, n<0.05= 0/10)]. Likewise,

PERMANOVA analysis accounting for host status did not alter the significance of host identity (PERMANOVA: Bray-Curtis, R^2= 0.56, F= 4.73, P= 0.001; Jaccard, R^2= 0.59, F= 2.96, P= 0.001).

Discussion

Two different metagenomic approaches support a major role of infrapopulation identity (ringed seal host individual) in explaining microbiome variation among individuals of the seal louse. In addition, highly similar results were found for approaches using either presence-absence or quantitative matrices, suggesting that not only is bacterial composition, but also bacterial abundance explained by infrapopulation identity. Our analyses were done on whole louse individuals and, thus, we cannot confidently differentiate between bacterial taxa inhabiting the lice (e.g., *Wolbachia* or *Hodgkinia*) from transient taxa present in the host blood meal (e.g., *Chlamydia*). Nevertheless, in line with current evidence on the determinants of microbiome composition of bloodsucking parasites, the louse blood meal from individual seals is the most likely candidate in explaining the patterns of microbiome variation across the louse infrapopulations found here. Indeed, many of the taxa found in our analyses have already been found in other bloodsucking parasites, thus supporting the influence of blood in shaping the composition of parasite microbiomes studied here (Jiménez-Cortés et al. 2018).

However, other factors in addition to blood may have contributed to the similarity of microbiomes between individual lice from the same seal host individual. Some similarity may have arisen from shared environmental bacteria, those on the surface of the louse from a shared environment (skin and fur of the host), or contamination between louse individuals in screw-cap tubes, and not filtered by our decontamination procedures. There may also be insect-specific bacterial taxa, independent from the host blood, that are shared horizontally between individual lice from the same infrapopulation. Finally, louse infrapopulations are known to typically be highly inbred, with a

high level of relatedness between individuals (Koop et al. 2014, DiBlasi et al 2018, Virrueta Herrera et al., in prep.). It may be that there are louse genetic factors that interact with the microbiome to produce a specific composition (Blekhman et al. 2015; Dobson et al. 2015; Suzuki et al. 2019).

Our results are congruent with previous findings on the influence of host blood on microbiomes of bloodsucking parasites. Specifically, several studies have found a major role of the specific host species from which a blood meal is taken in shaping microbiomes of other bloodsucking organisms, such as ticks (*Ixodes scapularis, Ixodes pacificus*) and mosquitoes (*Aedes aegypti*) (Swei and Kwan 2017; Landesman et al. 2019; Muturi et al. 2019). Furthermore, Landesman et al. (2019) showed that microbiomes of deer tick (*Ixodes scapularis*) nymphs were largely explained by the individual hosts of the tick, a result similar to the one obtained here. Interestingly, in that study, the percentage of variation explained was considerably lower (45%) than that found here (>77%). It may be that differences in parasite ecology, such as the whether the parasite is a permanent or a recurrent feeder (as are both the case in sucking lice) may modulate the extent to which host blood shapes parasite microbiomes. The differences in the proportion of variance explained by infrapopulation identity between the two studies could also be due to differences in experimental design, such as the number of sampled infrapopulations (3 in ticks, and 17 in the seal lice here) and whether the sample design is balanced (i.e., the same number of individual parasites sampled per infrapopulation).

The knowledge that blood from the same individual seal host may influence the similarity of the microbiome of blood-feeding lice from that host can potentially provide new insights into the influence of host blood on such parasites. There are least two not necessarily mutually exclusive processes may explain the influence of a host individual's blood on louse microbiomes. First, the

311 blood from a particular host individual may contain a specific composition of bacterial loads that
312 enter the louse on consumption of blood. Indeed, anopluran lice might have a higher likelihood
313 of being colonized by bacteria from host blood because they do not possess a peritrophic
314 membrane, an extracellular layer in the midgut that is composed of chitin, proteoglycans, and
315 proteins, which in most other insects surrounds the ingested food bolus and separates the gut
316 content, including bacteria, from the epithelium (Terra 2001; Waniek 2009). Indeed, the idea that
317 a lack of a peritrophic membrane may facilitate colonization of blood-feeding parasites by bacteria
318 present in the host blood has potentially also been supported by work on mouse fleas
319 (*Rhadinopsylla dahurica*), which also lack this membrane (Li et al. 2018). In this case, there was
320 evidence of homogenization (i.e., similar bacterial communities) between the host blood and the
321 parasite (whole flea individuals). The lack of a peritrophic membrane is often associated with
322 permanent parasites, such as blood-feeding lice, for which the continual availability of food means
323 that there is less selection for efficiency of digestion. Therefore, the presence versus absence of a
324 peritrophic membrane may explain the differences between lice and ticks (of which the latter
325 possess a peritrophic membrane) with regards to the influence of host blood on the composition of
326 the parasite microbiome.

327 A second possibility that could explain why host blood may influence louse microbiome
328 composition is that the conditions during blood digestion may alter bacterial taxa that were already
329 present in the louse. The specifics of blood digestion may have an individual host-specific
330 signature. Specifically, catabolism of blood meal leads to the generation of reactive oxygen
331 species that are known to alter the midgut bacterial composition and diversity of bloodsucking
332 parasites (Souza et al. 1997; Wang et al. 2011; Muturi et al. 2019). Also, the blood meals of
333 different host species are also known to differ in composition (e.g., total protein, hemoglobin, and

334 hematocrit content), and these differences may lead to a differential proliferation of microbial taxa
335 during digestion by the parasite (Souza et al. 1997; Wang et al. 2011; Muturi et al. 2019). It may
336 be the case that differences in blood composition among individuals even within the same host
337 species may be shaping the bacterial composition of lice in a manner that is specific to host
338 individuals.

339 Bloodsucking organisms, and anopluran lice in particular, are well known to rely on mutualistic
340 endosymbionts to complement deficiencies in their diet (Perotti et al. 2008; Boyd and Reed 2012;
341 Boyd et al. 2017; Jiménez-Cortés et al. 2018). Notwithstanding that several of the bacterial taxa
342 we found may not be stable inhabitants of lice, we did find evidence for the presence of several
343 louse-specific bacterial taxa. These include the obligate intracellular arthropod bacteria *Wolbachia*
344 (Werren 1997) and *Hodgkinia* (for which only endosymbionts of *Cicadas* are known; McCutcheon
345 et al. 2009). Accordingly, we explored our MAGs for genome characteristics typical of
346 endosymbionts. In particular, because endosymbiont genomes typically are small and have an AT
347 bias, we explored the relative position of the observed MAGs in a "Genome size ~ GC content"
348 correlation plot (Wernegreen 2015; Figure 4). Bin 1 appears to be the best candidate to be a
349 mutualistic endosymbiont, according to its relative position in the correlation plot. This MAG was
350 100% complete (according to CheckM; Parks et al. 2015), detected in most samples (prevalence =
351 71%), and classified with confidence as Flavobacteriaceae. MiGA analyses suggest it may even
352 belong to *Chryseobacterium* (p-value 0.585). Endosymbionts belonging to *Chryseobacterium* are
353 known in other arthropods (e.g., termites, mosquitoes, cockroaches, and ticks; Eutick et al. 1978;
354 Dugas et al. 2001; Campbell et al. 2004; Montasser 2005; Burešová et al. 2006). Additionally, we
355 conducted a preliminary investigation of the metabolic capabilities of this bacterium by
356 investigating the completeness of metabolic pathways using GhostKOALA (Kanehisa et al. 2016)

and KEGG-Decoder (Graham et al. 2018). This MAG has complete routes for synthesis of vitamin B (riboflavin), an essential amino acid (lysine), and several non-essential amino acids (e.g., serine; see Table S4), as well as many fully or partially missing routes that may be redundant or potentially shared (or synthesized along) with the louse (Table S4).

Overall, these results are congruent with what has been found for endosymbionts of bloodsucking parasites (Moriyama et al. 2015; Boyd et al. 2016; Santos-Garcia et al. 2017; Duron et al. 2018). Another anopluran pinniped louse (*Proechinophthirus fluctus*) has been found to have a *Sodalis* endosymbiont (Boyd et al. 2016), but we found no evidence of *Sodalis* in *Echinophthirius horridus*. Other species of Anoplura have yet other endosymbionts (Boyd et al. 2017, Říhová et al. 2017), suggesting that endosymbiont replacement is an ongoing and relatively common process within the order. Further research, including phylogenomic studies improving the phylogenetic placement of this potentially mutualistic bacterium and studies using fluorescence in situ hybridization (FISH) to ascertain the location of this bacterium in louse individuals, is needed to get deeper insight into the interaction of this bacterium with *E. horridus*.

Finally, at a broader scale, our results are congruent with previous studies that have found a major role of different levels of subdivision in shaping microbiomes in a wide range of systems. For instance, population identity has been found to largely explain microbiome composition of great apes (Campbell et al. 2020), American pikas (*Ochotona princeps*) (Kohl et al. 2018), and humans (Rothschild et al. 2018). Similar results have also been found across subdivision levels other than populations, such as ecotypes (human lice *Pediculus humanus*; Agany et al. 2020), or by spatial proximity (North American moose *Alces alces*; Fountain-Jones et al. 2020). On the other hand, while the influence of subdivision on microbiome composition is widely supported, much less is known about whether signs of phylosymbiosis can be found at these levels. Kohl et al. (2018)

found that populations largely explained microbiome variation in American pikas along with a phylosymbiotic pattern (i.e., closely related hosts had similar microbial communities). Interestingly, while we did not investigate the presence of phylosymbiosis here, our results suggest that a phylosymbiotic pattern is not always found at an intraspecific level. For instance, this may be the case of subdivided systems in which the food source (which allows the dispersal of microbes from one organism to another) constitutes the main determinant of microbiome composition, and populations are not necessarily composed of closely related organisms. However, phylosymbiotic patterns across species have been found in relatively similar systems. Therefore, bacterial dispersal, ecological drift, diversification, or microbe-microbe interactions may be the main factors explaining phylosymbiosis in these systems. More studies on the origin and prevalence of phylosymbiotic patterns within and across species are clearly needed.

Acknowledgments: We thank all researchers and students for collecting the lice analyzed in this study. Especially, we would like to thank researchers Miina Auttila, Vincent Biard, Meeri Koivuniemi, Lauri Liukkonen, Marja Niemi, Sari Oksanen, Mia Valtonen, and Eeva Ylinen for keeping eye on the lice during their own studies.

Author contributions: J.D., S.V.H, and K.P.J. conceived the study. T.N. and M.K. obtained samples. S.V.H. and K.P.J. collected the data. J.D. analyzed the data. T.N., M.V., and K.P.J. obtained financial support for the project. J.D. wrote the manuscript and all authors contributed to editing the manuscript.

Funding: This study was supported by the US National Science Foundation (DEB-1239788, DEB-1342604, and DEB-1926919 to K.P.J). Sequencing costs were supported by grants from

402 the Oskar Öflund Foundation, the Betty Väänänen Foundation, Societas Pro Fauna et Flora
403 Fennica, and the Nestori Foundation.

404 **Availability of data and material:** Raw sequence reads for all samples are available under
405 (NCBI) project *ACCESSION CODE*. Metagenomic assemblies are available at Figshare (doi:
406 10.6084/m9.figshare.12366575 —private link for review:
407 https://figshare.com/s/171207f7fb73f7d290a6). Metagenomic assemblies are also available from
408 NCBI Genome (*ACCESION CODES*).

409 **Ethics approval and consent to participate:** Telemetry studies have been approved by the local
410 environmental authority Centre for Economic Development, Transport and the Environment
411 (permit numbers: ESAELY/433/ 07.01/2012 and ESA-2008-L-519-254) and the Animal
412 Experiment Board in Finland (permit numbers: ESAVI/ 8269/04.10.07/2013 and ESAVI-2010-
413 08380/Ym-23).

414 **Competing interests:** The authors declare that they have no competing interests.
415

References:

Adair, K. L., and A. E. Douglas. (2017). Making a microbiome: the many determinants of host-associated microbial community composition. *Current Opinion in Microbiology*, 35, 23–29.

Agany, D. D. M., R. Potts, J. L. Gonzalez Hernandez, E. Z. Gnimpieba, and J. E. Pietri. (2020). Microbiome Differences between Human Head and Body Lice Ecotypes Revealed by 16S rRNA Gene Amplicon Sequencing. *Journal of Parasitology*, 106, 14.

Alneberg, J., B. S. Bjarnason, I. de Bruijn, M. Schirmer, J. Quick, U. Z. Ijaz, L. Lahti, N. J. Loman, A. F. Andersson, and C. Quince. (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods*, 11, 1144–1146.

Anderson, M. J. (2014). Permutational multivariate analysis of variance (PERMANOVA). *Wiley statsref: statistics reference online*, 1–15.

Anderson, M. J., and D. C. Walsh. (2013). PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: what null hypothesis are you testing? *Ecological Monographs*, 83, 557–574.

Auguie, B. (2016). gridExtra: Miscellaneous functions for "Grid" graphics. R package version 2.3. URL https://CRAN. R-project. org/package= gridExtra.

Blekhman, R., J. K. Goodrich, K. Huang, Q. Sun, R. Bukowski, J. T. Bell, T. D. Spector, A. Keinan, R. E. Ley, D. Gevers, and A. G. Clark. (2015). Host genetic variation impacts microbiome composition across human body sites. *Genome Biology*, 16, 191.

Bowers, R. M., N. C. Kyrpides, R. Stepanauskas, M. Harmon-Smith, D. Doud, T. B. K. Reddy, F. Schulz, J. Jarett, A. R. Rivers, E. A. Eloe-Fadrosh, S. G. Tringe, N. N. Ivanova, A. Copeland, A. Clum, E. D. Becraft, R. R. Malmstrom, B. Birren, M. Podar, P. Bork, G. M. Weinstock, G. M. Garrity, J. A. Dodsworth, S. Yooseph, G. Sutton, F. O. Glöckner, J. A. Gilbert, W. C. Nelson, S. J. Hallam, S. P. Jungbluth, T. J. G. Ettema, S. Tighe, K. T. Konstantinidis, W.-T. Liu, B. J. Baker, T. Rattei, J. A. Eisen, B. Hedlund, K. D. McMahon, N. Fierer, R. Knight, R. Finn, G. Cochrane, I. Karsch-Mizrachi, G. W. Tyson, C. Rinke, A. Lapidus, F. Meyer, P. Yilmaz, D. H. Parks, A. M. Eren, L. Schriml, J. F. Banfield, P. Hugenholtz, and T. Woyke. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology*, 35, 725–731.

Boyd, B. M., J. M. Allen, R. Koga, T. Fukatsu, A. D. Sweet, K. P. Johnson, and D. L. Reed. (2016). Two Bacterial Genera, Sodalis and Rickettsia, Associated with the Seal Louse Proechinophthirus fluctus (Phthiraptera: Anoplura). *Applied and Environmental Microbiology*, 82, 3185–3197.

Boyd, B. M., J. M. Allen, N.-P. Nguyen, P. Vachaspati, Z. S. Quicksall, T. Warnow, L. Mugisha, K. P. Johnson, and D. L. Reed. (2017). Primates, Lice and Bacteria: Speciation and Genome Evolution in the Symbionts of Hominid Lice. *Molecular Biology and Evolution,* 34, 1743–1757.

Boyd, B. M., and D. L. Reed. (2012). Taxonomy of lice and their endosymbiotic bacteria in the post-genomic era. *Clinical Microbiology and Infection*, 18, 324–331.

Brooks, A. W., K. D. Kohl, R. M. Brucker, E. J. van Opstal, and S. R. Bordenstein. (2016). Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History. *PLoS Biology*, 14, e2000225.

Brucker, R. M., and S. R. Bordenstein. (2013). The Hologenomic Basis of Speciation: Gut

461 Bacteria Cause Hybrid Lethality in the Genus Nasonia. *Science*, 341, 667–669.
462 Burešová, V., Z. Franta, and P. Kopáček. (2006). A comparison of Chryseobacterium
463 indologenes pathogenicity to the soft tick Ornithodoros moubata and hard tick Ixodes
464 ricinus. *Journal of Invertebrate Pathology*, 93, 96–104.
465 Campbell, C. L., D. L. Mummey, E. T. Schmidtmann, and W. C. Wilson. (2004). Culture-
466 independent analysis of midgut microbiota in the arbovirus vector Culicoides sonorensis
467 (Diptera: Ceratopogonidae). *Journal of Medical Entomology*, 41, 340–348.
468 Campbell, T. P., X. Sun, V. H. Patel, C. Sanz, D. Morgan, and G. Dantas. (2020). The
469 microbiome and resistome of chimpanzees, gorillas, and humans across host lifestyle and
470 geography. *ISME Journal*, 1–16.
471 Clemente, J. C., L. K. Ursell, L. W. Parfrey, and R. Knight. (2012). The impact of the gut
472 microbiota on human health: an integrative view. *Cell*, 148, 1258–1270.
473 Davis, N. M., D. M. Proctor, S. P. Holmes, D. A. Relman, and B. J. Callahan. (2018). Simple
474 statistical identification and removal of contaminant sequences in marker-gene and
475 metagenomics data. *Microbiome*, 6, 226.
476 Díaz-Sánchez, S., A. Estrada-Peña, A. Cabezas-Cruz, and J. de la Fuente. (2019). Evolutionary
477 Insights into the Tick Hologenome. *Trends in Parasitology*, 35, 725–737.
478 Dobson, A. J., J. M. Chaston, P. D. Newell, L. Donahue, S. L. Hermann, D. R. Sannino, S.
479 Westmiller, A. C.-N. Wong, A. G. Clark, and B. P. Lazzaro. (2015). Host genetic
480 determinants of microbiota-dependent nutrition revealed by genome-wide analysis of
481 Drosophila melanogaster. *Nature Communications*, 6, 1–9.
482 Dröge, J., I. Gregor, and A. C. McHardy. (2015). Taxator-tk: precise taxonomic assignment of
483 metagenomes by fast approximation of evolutionary neighborhoods. *Bioinformatics,* 31,
484 817–824.
485 Dugas, J. E., L. Zurek, B. J. Paster, B. A. Keddie, and E. R. Leadbetter. (2001). Isolation and
486 characterization of a Chryseobacterium strain from the gut of the American cockroach,
487 Periplaneta americana. *Archives of Microbiology*, 175, 259–262.
488 Durden, L. A., and G. G. Musser. (1994). The mammalian hosts of the sucking lice (Anoplura)
489 of the world: a host-parasite list. *Journal of Vector Ecology*, 19, 130–168.
490 Duron, O., O. Morel, V. Noël, M. Buysse, F. Binetruy, R. Lancelot, E. Loire, C. Ménard, O.
491 Bouchez, F. Vavre, and L. Vial. (2018). Tick-Bacteria Mutualism Depends on B Vitamin
492 Synthesis Pathways. *Current Biology*, 28, 1896-1902.e5.
493 Eutick, M. L., R. W. O'brien, and M. Slaytor. (1978). Bacteria from the gut of Australian
494 termites. *Applied and Environmental Microbiology*, 35, 823–828.
495 Fountain-Jones, N. M., N. J. Clark, A. C. Kinsley, M. Carstensen, J. Forester, T. J. Johnson, E.
496 A. Miller, S. Moore, T. M. Wolf, and M. E. Craft. (2020). Microbial associations and
497 spatial proximity predict North American moose (Alces alces) gastrointestinal
498 community composition. *Journal of Animal Ecology*, 89, 817–828.
499 Graham, E. D., J. F. Heidelberg, and B. J. Tully. (2018). Potential for primary productivity in a
500 globally-distributed bacterial phototroph. ISME Journal, 12, 1861–1866.
501 Hassani, M. A., P. Durán, and S. Hacquard. (2018). Microbial interactions within the plant
502 holobiont. *Microbiome*, 6, 58.
503 Hooper, R., J. C. Brealey, T. van der Valk, A. Alberdi, J. W. Durban, H. Fearnbach, K. M.
504 Robertson, R. W. Baird, M. B. Hanson, P. Wade, M. T. P. Gilbert, P. A. Morin, J. B. W.
505 Wolf, A. D. Foote, and K. Guschanski. (2019). Host-derived population genomics data
506 provides insights into bacterial and diatom composition of the killer whale skin.

Molecular Ecology, 28, 484–502.

Jackman, S. D., B. P. Vandervalk, H. Mohamadi, J. Chu, S. Yeo, S. A. Hammond, G. Jahesh, H. Khan, L. Coombe, R. L. Warren, and I. Birol. (2017). ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Research*, 27, 768–777.

Jervis-Bardy, J., L. E. Leong, S. Marri, R. J. Smith, J. M. Choo, H. C. Smith-Vaughan, E. Nosworthy, P. S. Morris, S. O'Leary, and G. B. Rogers. (2015). Deriving accurate microbiota profiles from human samples with low bacterial content through post-sequencing processing of Illumina MiSeq data. *Microbiome*, 3, 19. Springer.

Jiménez-Cortés, J. G., R. García-Contreras, M. I. Bucio-Torres, M. Cabrera-Bravo, A. Córdoba-Aguilar, G. Benelli, and P. M. Salazar-Schettino. (2018). Bacterial symbionts in human blood-feeding arthropods: Patterns, general mechanisms and effects of global ecological changes. *Acta Tropica*, 186, 69–101.

Kanehisa, M., Y. Sato, and K. Morishima. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of Molecular Biology*, 428, 726–731.

Kang, D. D., F. Li, E. Kirton, A. Thomas, R. Egan, H. An, and Z. Wang. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, 7:e7359.

Kassambara, A. (2018). Package 'ggpubr':'ggplot2'based publication ready plots. Version 0.2. See https://rpkgs. datanovia. com/ggpubr/index. html.

Kelly, B. J., Gross, R., Bittinger, K., Sherrill-Mix, S., Lewis, J. D., Collman, R. G., ... & Li, H. (2015). Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA. *Bioinformatics*, 31, 2461-2468.

Kim, K. C. (1985). Coevolution of Parasitic Arthropods and Mammals. Wiley.

Kim, K. C. (1975). Specific antiquity of the sucking lice and evolution of otariid seals. *Rapports et procès-verbaux des réunions*, 169, 544–549.

Kim, K. C. (1971). The Sucking Lice (Anoplura: Echinophthiriidae) of the Northern Fur Seal; Descriptions and Morphological Adaptation. *Annals of the Entomological Society of America*, 64, 280–292

Knowles, S. C. L., Eccles, R. M., & Baltrūnaitė, L. (2019). Species identity dominates over environment in shaping the microbiota of small mammals. *Ecology Letters*, 22, 826-837.

Kohl, K. D. (2020). Ecological and evolutionary mechanisms underlying patterns of phylosymbiosis in host-associated microbial communities. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, 375, 20190251.

Kohl, K. D., J. Varner, J. L. Wilkening, and M. D. Dearing. (2018). Gut microbial communities of American pikas (*Ochotona princeps*): Evidence for phylosymbiosis and adaptations to novel diets. *Journal of Animal Ecology*, 87, 323–330.

Laetsch, D. R., and M. L. Blaxter. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research*, 6, 1287.

Landesman, W. J., K. Mulder, B. F. Allan, L. A. Bashor, F. Keesing, K. LoGiudice, and R. S. Ostfeld. (2019). Potential effects of blood meal host on bacterial community composition in Ixodes scapularis nymphs. *Ticks and Tick-Borne Diseases*, 10, 523–527.

Lankau, E. W., P.-Y. Hong, and R. I. Mackie. (2012). Ecological drift and local exposures drive enteric bacterial community differences within species of Galapagos iguanas. *Molecular Ecology*, 21, 1779–1788.

Le Chatelier, E., T. Nielsen, J. Qin, E. Prifti, F. Hildebrand, G. Falony, M. Almeida, M.

Arumugam, J.-M. Batto, S. Kennedy, P. Leonard, J. Li, K. Burgdorf, N. Grarup, T. Jørgensen, I. Brandslund, H. B. Nielsen, A. S. Juncker, M. Bertalan, F. Levenez, N. Pons, S. Rasmussen, S. Sunagawa, J. Tap, S. Tims, E. G. Zoetendal, S. Brunak, K. Clément, J. Doré, M. Kleerebezem, K. Kristiansen, P. Renault, T. Sicheritz-Ponten, W. M. de Vos, J.-D. Zucker, J. Raes, T. Hansen, P. Bork, J. Wang, S. D. Ehrlich, and O. Pedersen. (2013). Richness of human gut microbiome correlates with metabolic markers. *Nature*, 500, 541–546.

Lee, S., J. Y. Kim, M. Yi, I.-Y. Lee, R. Fyumagwa, and T.-S. Yong. (2019). Comparative microbiomes of ticks collected from a black rhino and its surrounding environment. *International Journal of Parasitology Parasites and Wildlife*, 9, 239–243.

Leonardi, M. S., E. A. Crespo, J. A. Raga, and F. J. Aznar. (2013). Lousy mums: patterns of vertical transmission of an amphibious louse. *Parasitology Research*, 112, 3315–3323. Springer.

Leonardi, M. S., S. V. Herrera, A. Sweet, J. Negrete, and K. P. Johnson. (2019). Phylogenomic analysis of seal lice reveals codivergence with their hosts. *Systematic Entomology*, 44, 699–708.

Li, H., T. Li, and J. Qu. (2018). Stochastic processes govern bacterial communities from the blood of pikas and from their arthropod vectors. *FEMS Microbiology Ecology,* 94.

Lim, S. J., and S. R. Bordenstein. (2020). An introduction to phylosymbiosis. *Proceedings of the Royal Society B: Biological Sciences*, 287, 20192900.

Lusk, R. W. (2014). Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data. *PloS One,* 9.

Lutz, H. L., E. W. Jackson, P. W. Webala, W. S. Babyesiza, J. C. Kerbis Peterhans, T. C. Demos, B. D. Patterson, and J. A. Gilbert. (2019). Ecology and Host Identity Outweigh Evolutionary History in Shaping the Bat Microbiome. *mSystems,* 4.

Mazel, F., K. M. Davis, A. Loudon, W. K. Kwong, M. Groussin, and L. W. Parfrey. (2018). Is Host Filtering the Main Driver of Phylosymbiosis across the Tree of Life? *mSystems,* 3.

McArdle, A. J., and M. Kaforou. (2020). Sensitivity of shotgun metagenomics to host DNA: abundance estimates depend on bioinformatic tools and contamination is the main issue. *Access Microbiology*, doi: 10.1099/acmi.0.000104.

McCutcheon, J. P., B. R. McDonald, and N. A. Moran. *2009*. Origin of an Alternative Genetic Code in the Extremely Small and GC–Rich Genome of a Bacterial Symbiont. *PLoS Genetics*, 5, e1000565.

McMurdie, P. J., and S. Holmes. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS One,* 8.

Menzel, P., K. L. Ng, and A. Krogh. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications*, 7, 1–9.

Montasser, A. A. (2005). Gram-negative bacteria from the camel tick Hyalomma dromedarii (Ixodidae) and the chicken tick Argas persicus (Argasidae) and their antibiotic sensitivities. *Journal of the Egyptian Society of Parasitology*, 35, 95–106.

Moriyama, M., N. Nikoh, T. Hosokawa, and T. Fukatsu. (2015). Riboflavin Provisioning Underlies Wolbachia's Fitness Contribution to Its Insect Host. *mBio,* 6.

Muturi, E. J., C. Dunlap, J. L. Ramirez, A. P. Rooney, and C.-H. Kim. (2019). Host blood-meal source has a strong impact on gut microbiota of Aedes aegypti. *FEMS Microbiology Ecology*, 95.

598 Niemi, M., Liukkonen, L., Koivuniemi, M., Auttila, M., Rautio, A., & Kunnasranta, M. (2019).
599 Winter behavior of Saimaa ringed seals: Non-overlapping core areas as indicators of
600 avoidance in breeding females. *PloS one*, 14, 1.
601 Nishida, A. H., and H. Ochman. (2018). Rates of gut microbiome divergence in mammals.
602 *Molecular Ecology*, 27, 1884–1897.
603 Nurk, S., D. Meleshko, A. Korobeynikov, and P. A. Pevzner. (2017). metaSPAdes: a new
604 versatile metagenomic assembler. *Genome Research*, 27, 824–834.
605 Nyman, T., Valtonen, M., Aspi, J., Ruokonen, M., Kunnasranta, M. & Palo, J. U. (2014).
606 Demographic histories and genetic diversities of Fennoscandian marine and landlocked
607 ringed seal subspecies. *Ecology and Evolution*, 4, 3420-3434.
608 Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, R. B. O'hara, G. L. Simpson, P. Solymos, M.
609 H. H. Stevens, and H. Wagner. (2019). Vegan: community ecology package. R package
610 version 2.5-4.
611 Osei-Poku, J., C. M. Mbogo, W. J. Palmer, and F. M. Jiggins. (2012). Deep sequencing reveals
612 extensive variation in the gut microbiota of wild mosquitoes from Kenya. *Molecular*
613 *Ecology*, 21, 5138–5150.
614 Parks, D. H., M. Imelfort, C. T. Skennerton, P. Hugenholtz, and G. W. Tyson. (2015). CheckM:
615 assessing the quality of microbial genomes recovered from isolates, single cells, and
616 metagenomes. *Genome Research*, 25, 1043–1055.
617 Patro, R., G. Duggal, M. I. Love, R. A. Irizarry, and C. Kingsford. (2017). Salmon provides fast
618 and bias-aware quantification of transcript expression. *Nature Methods*, 14, 417.
619 Perotti, M. A., E. F. Kirkness, D. L. Reed, and H. R. Braig. (2008). Endosymbionts of lice. Pp.
620 223–238 *in* Insect Symbiosis, Volume 3. CRC Press.
621 R Core Team. (2019). R: A language and environment for statistical computing.
622 Rodriguez-R, L. M., S. Gunturu, W. T. Harvey, R. Rosselló-Mora, J. M. Tiedje, J. R. Cole, and
623 K. T. Konstantinidis. (2018). The Microbial Genomes Atlas (MiGA) webserver:
624 taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome
625 level. *Nucleic Acids Research*, 46, W282–W288.
626 Rothschild, D., O. Weissbrod, E. Barkan, A. Kurilshikov, T. Korem, D. Zeevi, P. I. Costea, A.
627 Godneva, I. N. Kalka, N. Bar, S. Shilo, D. Lador, A. V. Vila, N. Zmora, M. Pevsner-
628 Fischer, D. Israeli, N. Kosower, G. Malka, B. C. Wolf, T. Avnit-Sagi, M. Lotan-Pompan,
629 A. Weinberger, Z. Halpern, S. Carmi, J. Fu, C. Wijmenga, A. Zhernakova, E. Elinav, and
630 E. Segal. (2018) Environment dominates over host genetics in shaping human gut
631 microbiota. *Nature*, 555, 210–215.
632 Rudman, S. M., S. Greenblum, R. C. Hughes, S. Rajpurohit, O. Kiratli, D. B. Lowder, S. G.
633 Lemmon, D. A. Petrov, J. M. Chaston, and P. Schmidt. (2019). Microbiome composition
634 shapes rapid genomic adaptation of *Drosophila melanogaster*. *Proceedings of the*
635 *National Academy of Sciences USA*, 116, 20025–20032.
636 Salter, S. J., M. J. Cox, E. M. Turek, S. T. Calus, W. O. Cookson, M. F. Moffatt, P. Turner, J.
637 Parkhill, N. J. Loman, and A. W. Walker. (2014). Reagent and laboratory contamination
638 can critically impact sequence-based microbiome analyses. *BMC Biology*, 12, 87.
639 Springer.
640 Santos-Garcia, D., F. J. Silva, S. Morin, K. Dettner, and S. M. Kuechler. (2017). The All-
641 Rounder Sodalis: A New Bacteriome-Associated Endosymbiont of the Lygaeoid Bug
642 Henestaris halophilus (Heteroptera: Henestarinae) and a Critical Examination of Its
643 Evolution. *Genome Biology and Evolution*, 9, 2893–2910.

644 Slowikowski, K., A. Schep, S. Hughes, S. Lukauskas, J. O. Irisson, and Z. N. Kamvar. (2019).
645 ggrepel: automatically position non-overlapping text labels with 'ggplot2'2018. URL
646 https://CRAN. R-project. org/package= ggrepel. R package version 0.8. 0.
647 Snodgrass, R. E. (1944). The feeding apparatus of biting and sucking insects affecting man and
648 animals. Smithsonian Miscellaneous Collections.
649 Song, S. J., J. G. Sanders, F. Delsuc, J. Metcalf, K. Amato, M. W. Taylor, F. Mazel, H. L. Lutz,
650 K. Winker, G. R. Graves, G. Humphrey, J. A. Gilbert, S. J. Hackett, K. P. White, H. R.
651 Skeen, S. M. Kurtis, J. Withrow, T. Braile, M. Miller, K. G. McCracken, J. M. Maley, V.
652 O. Ezenwa, A. Williams, J. M. Blanton, V. J. McKenzie, and R. Knight. (2020).
653 Comparative Analyses of Vertebrate Gut Microbiomes Reveal Convergence between
654 Birds and Bats. *mBio*, 11.
655 Souza, A. V. G., J. H. Petretski, M. Demasi, E. J. H. Bechara, and P. L. Oliveira. (1997). Urate
656 protects a blood-sucking insect against hemin-induced oxidative stress. *Free Radical*
657 *Biology & Medicine*, 22, 209–214.
658 Suzuki, T. A., M. Phifer-Rixey, K. L. Mack, M. J. Sheehan, D. Lin, K. Bi, and M. W. Nachman.
659 (2019). Host genetic determinants of the gut microbiota of wild mice. *Molecular*
660 *Ecology*, 28, 3197–3207.
661 Swei, A., and J. Y. Kwan. (2017). Tick microbiome and pathogen acquisition altered by host
662 blood meal. *ISME Journal*, 11, 813–816.
663 Terra, W. R. (2001). The origin and functions of the insect peritrophic membrane and peritrophic
664 gel. *Archives of Insect Biochemistry and Physiology*, 47, 47–61.
665 Uritskiy, G. V., J. DiRuggiero, and J. Taylor. (2018). MetaWRAP—a flexible pipeline for
666 genome-resolved metagenomic data analysis. *Microbiome,* 6, 158.
667 Velazquez, E. M., H. Nguyen, K. T. Heasley, C. H. Saechao, L. M. Gil, A. W. L. Rogers, B. M.
668 Miller, M. R. Rolston, C. A. Lopez, Y. Litvak, M. J. Liou, F. Faber, D. N. Bronner, C. R.
669 Tiffany, M. X. Byndloss, A. J. Byndloss, and A. J. Bäumler. (2019). Endogenous
670 Enterobacteriaceae underlie variation in susceptibility to Salmonella infection. *Nature*
671 *Microbiology*, 4, 1057–1064.
672 Wang, Y., T. M. G. Iii, P. Kukutla, G. Yan, and J. Xu. (2011). Dynamic Gut Microbiome across
673 Life History of the Malaria Mosquito Anopheles gambiae in Kenya. *PloS One*, 6, e24767.
674 Waniek, P. J. (2009). The digestive system of human lice: current advances and potential
675 applications. *Physiological Entomology*, 34, 203–210.
676 Weiss, S., Z. Z. Xu, S. Peddada, A. Amir, K. Bittinger, A. Gonzalez, C. Lozupone, J. R.
677 Zaneveld, Y. Vázquez-Baeza, A. Birmingham, E. R. Hyde, and R. Knight. (2017).
678 Normalization and microbial differential abundance strategies depend upon data
679 characteristics. *Microbiome*, 5, 27.
680 Wernegreen, J. J. (2015). Endosymbiont evolution: Predictions from theory and surprises from
681 genomes. *Annals of the New York Academy of Sciences*, 1360, 16–35.
682 Werren, J. H. (1997). Biology of wolbachia. *Annual Review of Entomology*, 42, 587–609.
683 Wickham, H. (2016). ggplot2: elegant graphics for data analysis. Springer.
684 Willner, D., J. Daly, D. Whiley, K. Grimwood, C. E. Wainwright, and P. Hugenholtz. (2012).
685 Comparison of DNA extraction methods for microbial community profiling with an
686 application to pediatric bronchoalveolar lavage samples. *PloS One*, 7.
687 Wu, Y.-W., B. A. Simmons, and S. W. Singer. (2016). MaxBin 2.0: an automated binning
688 algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics,* 32,
689 605–607.

690 Xiao, N. (2018). ggsci: scientific journal and sci-fi themed color palettes for 'ggplot2'. R
691 package version 2.9.
692 Zhang, X.-C., Z.-N. Yang, B. Lu, X.-F. Ma, C.-X. Zhang, and H.-J. Xu. (2014). The composition
693 and transmission of microbiome in hard tick, Ixodes persulcatus, during blood meal.
694 *Ticks and Tick-Borne Diseases*, 5, 864–870.
695 Zolnik, C. P., R. C. Falco, T. J. Daniels, and S.-O. Kolokotronis. (2018). Transient influence of
696 blood meal and natural environment on blacklegged tick bacterial communities. *Ticks*
697 *and Tick-Borne Diseases*, 9, 563–572.
698

699 **Figure and table legend:**

700 **Table 1.** Statistics of the MAGs assembled. MAG name indicates the name given to that bin for
701 this study (e.g., in Figure 1). The highest taxonomic rank with p-value ≤ 0.5 is shown in MiGA
702 ID. RPD ID is the result of the identification analysis using rRNA genes (16S) implemented in
703 MiGA; % indicates confidence in identification. Taxonomic novelty is a MiGA analysis that
704 indicates the taxonomic rank at which the MAG represents a novel taxon with respect to the NCBI
705 Genome database; highest taxonomic rank with p-value ≤ 0.01 are shown.

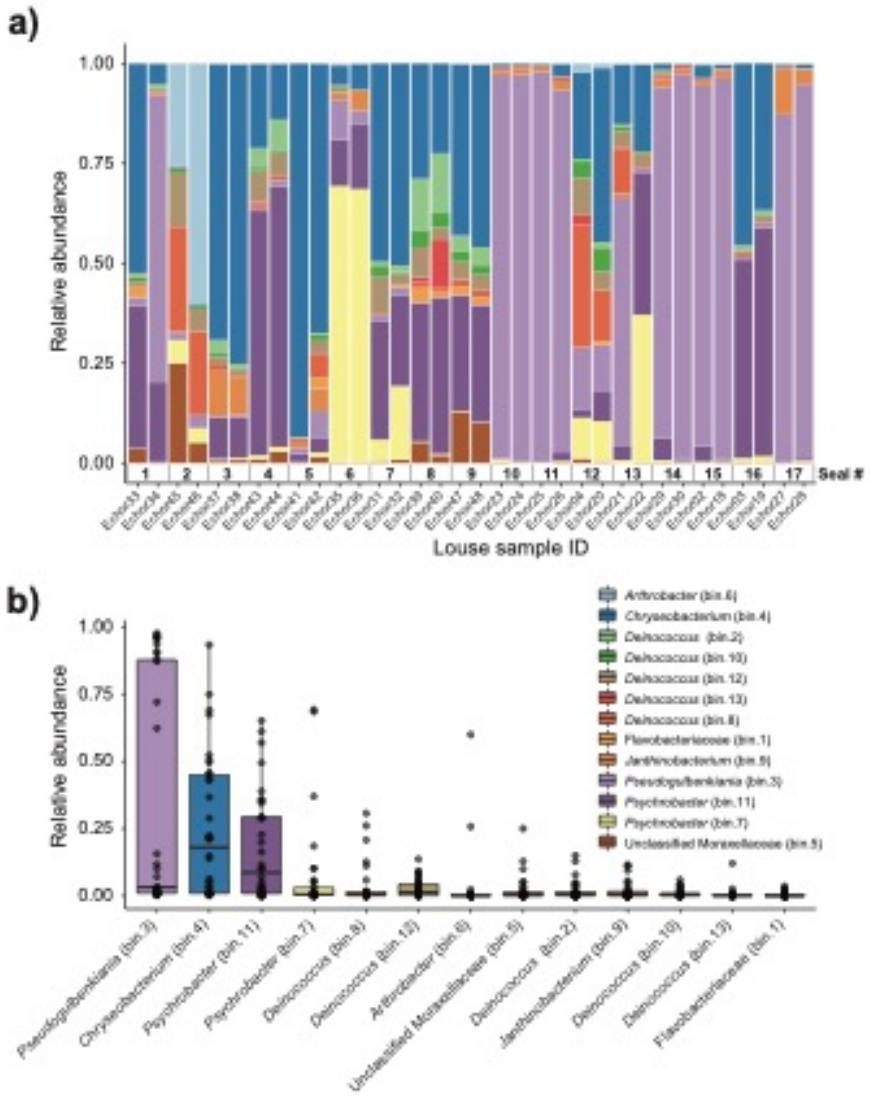
MAG name	Completeness (%)	Contamination (%)	N50 (bp)	Size (bp)	Taxator tk ID	MiGA ID	RDP ID	Taxonomic novelty
bin.1	100	1.07	57370	1869975	Flavobacteriaceae	Flavobacteriaceae*	NA	Species****
bin.4	99.26	0.24	81315	2500734	Flavobacteriaceae	Chryseobacterium*	Chryseobacterium (100.0%)	Species****
bin.2	98.51	0.42	36844	3101576	Deinococcus	Deinococcus grandis*	Deinococcus (100.0%)	Subspecies****
bin.7	97.75	0	16123	2650064	Moraxellaceae	Psychrobacter sp. PRwf-1*	NA	Subspecies****
bin.3	97.41	1.33	32961	4014303	Neisseriales	Pseudogulbenkiania*	NA	Species****
bin.11	95.65	0.92	69243	2786419	Moraxellaceae	Psychrobacter*	NA	Species****
bin.10	95.12	0	13409	2459723	Deinococcaceae	Deinococcus*	NA	Species****
bin.12	93.14	0.85	24793	2851493	Deinococcaceae	Deinococcus*	NA	Species****
bin.6	88.74	1.45	7283	1988194	Micrococcales	Arthrobacter*	NA	Species****
bin.13	77.11	0.64	3045	2627969	Deinococcaceae	Deinococcus*	NA	Species****
bin.5	74.13	0.61	24837	1635952	Moraxellaceae	unclassified Moraxellaceae*	Alkanindiges (99 %)	Species****
bin.8	67.76	0	10934	2837743	Deinococcaceae	Deinococcus*	NA	Species****

bin.9	61.13	0.30	2210	2110411	Janthinobacterium	Janthinobacterium sp. SNU WT3***	NA	Subspecies****

706 Significance at p-value below: *0.5, **0.1, ***0.05, ****0.01

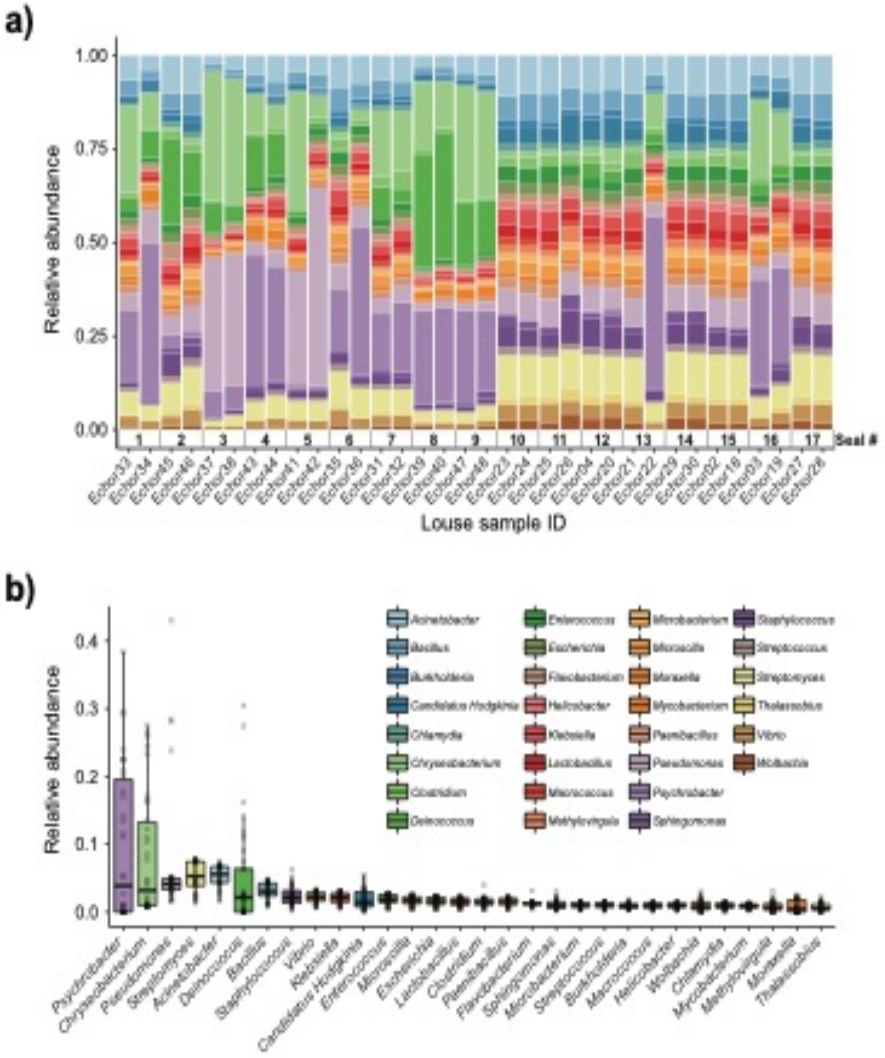
707

708 **Figure 1.** Genome-resolved metagenomic data. a) Stacked bar plot showing the relative
709 abundances of MAGs in each louse sample. Note that samples are ordered according to host
710 (i.e., samples from the same host are next to each other). b) Boxplot summarizing the relative
711 abundance of each MAG across the louse samples.
712 /



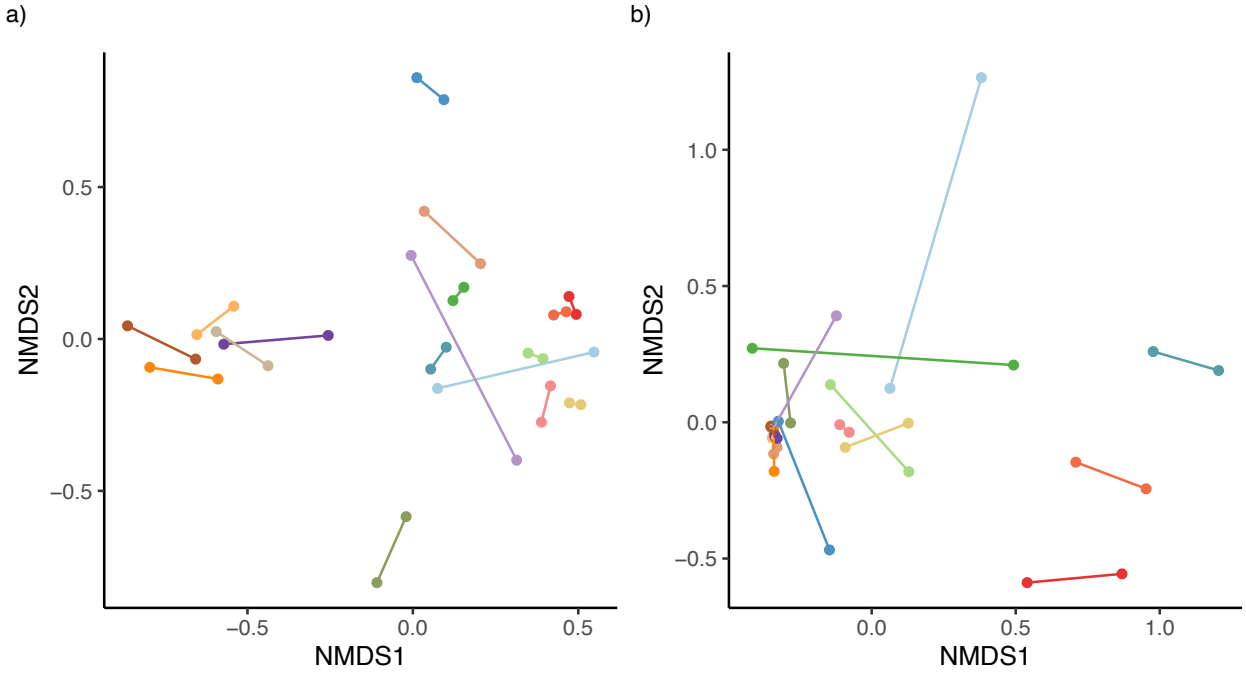
713

714 **Figure 2.** Kaiju data. a) Stacked bar plot showing bacterial relative abundances in each seal
715 louse sample. Note that samples are sorted according to host individual (i.e., samples from the
716 same host are next to each other). b) Boxplot summarizing the relative abundance of each taxon
717 across all louse samples.



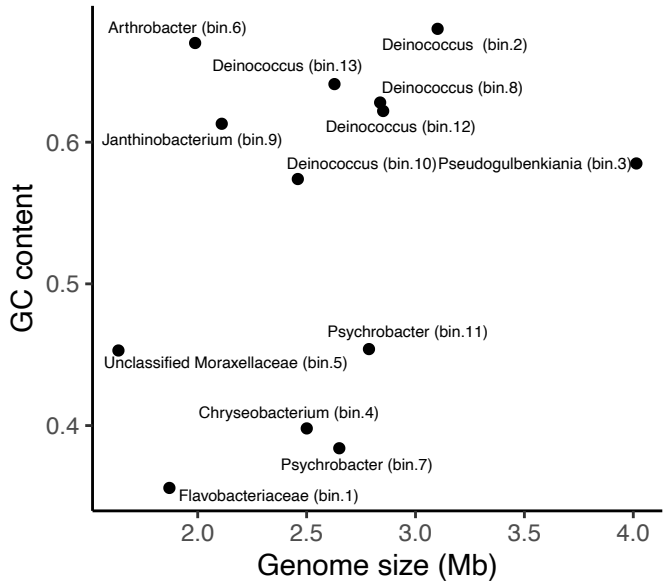
718 /
719

720 **Figure 3.** NMDS ordinations of seal louse microbiomes base on Bray–Curtis dissimilarity
721 matrices. a) MAG matrix, and b) Kaiju matrix (species level). Lice originating from the same
722 seal individual are colored similarly and connected by a line.
723 /



724

725 **Figure 4.** Scatter plot showing the relationship between genome size (Mb) and GC content (i.e.,
726 proportion of G and C sites) for sequenced MAGs.



727 /
728

729