

RESEARCH ARTICLE

Lactobacilli and other gastrointestinal microbiota of *Peromyscus leucopus*, reservoir host for agents of Lyme disease and other zoonoses in North America

Ana Milovic¹, Khalil Bassam^{1,2}, Hanjuan Shao¹, Ioulia Chatzistamou³, Danielle M. Tufts⁴, Maria Diuk-Wasser⁴, Alan G. Barbour^{1,5*}

1 Department of Microbiology & Molecular Genetics, University of California Irvine, Irvine, California, United States of America, **2** Faculty of Medicine, American University of Beirut, Beirut, Lebanon, **3** Department of Pathology, Microbiology and Immunology, School of Medicine, University of South Carolina, Columbia, South Carolina, United States of America, **4** Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, New York, United States of America, **5** Department of Medicine, University of California Irvine, Irvine, California, United States of America

* abarbour@uci.edu



OPEN ACCESS

Citation: Milovic A, Bassam K, Shao H, Chatzistamou I, Tufts DM, Diuk-Wasser M, et al. (2020) Lactobacilli and other gastrointestinal microbiota of *Peromyscus leucopus*, reservoir host for agents of Lyme disease and other zoonoses in North America. PLoS ONE 15(8): e0231801. <https://doi.org/10.1371/journal.pone.0231801>

Editor: Ulrike Gertrud Munderloh, University of Minnesota, UNITED STATES

Received: March 27, 2020

Accepted: July 15, 2020

Published: August 20, 2020

Copyright: © 2020 Milovic et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: The research was supported by National Institutes of Health (nih.gov) grant R21 AI136523 to A.G.B., National Science Foundation (nsf.gov) grants 1736150 and 1755670 to I.C. and H. Kiaris (University of South Carolina), and National Science Foundation (Division of Integrative Organismal Systems) grant IOS1755286 to D.M.T.

Abstract

The cricetine rodent *Peromyscus leucopus* is an important reservoir for several human zoonoses, including Lyme disease, in North America. Akin to hamsters, the white-footed deer mouse has been unevenly characterized in comparison to the murid *Mus musculus*. To further understanding of *P. leucopus*' total genomic content, we investigated gut microbiomes of an outbred colony of *P. leucopus*, inbred *M. musculus*, and a natural population of *P. leucopus*. Metagenome and whole genome sequencing were combined with microbiology and microscopy approaches. A focus was the genus *Lactobacillus*, four diverse species of which were isolated from forestomach and feces of colony *P. leucopus*. Three of the species—*L. animalis*, *L. reuteri*, and provisionally-named species “*L. peromysci*”—were identified in fecal metagenomes of wild *P. leucopus* but not discernibly in samples from *M. musculus*. *L. johnsonii*, the fourth species, was common in *M. musculus* but absent or sparse in wild *P. leucopus*. Also identified in both colony and natural populations were a *Helicobacter* sp. in feces but not stomach, and a *Tritrichomonas* sp. protozoan in cecum or feces. The gut metagenomes of colony *P. leucopus* were similar to those of colony *M. musculus* at the family or higher level and for major subsystems. But there were multiple differences between species and sexes within each species in their gut metagenomes at orthologous gene level. These findings provide a foundation for hypothesis-testing of functions of individual microbial species and for interventions, such as bait vaccines based on an autochthonous bacterium and targeting *P. leucopus* for transmission-blocking.

and M.A.D. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Epigraph: “I have always looked at problems from an ecological point of view, by placing most emphasis not on the living things themselves, but rather on their inter-relationships and on their interplay with surroundings and events.” René Dubos, 1981 [1, 2]

Peromyscus leucopus, the white-footed deermouse, is one of the most abundant wild mammals in central and eastern United States and adjacent regions of Canada and Mexico [3, 4]. The rodent is an omnivore, consuming a variety of seeds, such as oak acorns, as well as insects and other invertebrates. Its wide geographic range extends from rural areas to suburbs and even cities, and it is especially common in areas where humans and wildland areas interface [5]. Conditions permitting, *P. leucopus* is procreatively proliferant, with 20 or more litters during a female’s period of fecundity [6].

Although commonly called a “mouse”, this species and other members of the genus *Peromyscus* belong to the family Cricetidae, which includes hamsters and voles, and not the family Muridae, which includes *Mus* and *Rattus*. The pairwise divergence time for the genera *Peromyscus* and *Mus* is estimated to be ~27 million years ago [7, 8], approximately the time since divergence of the family Hominidae, the great apes and hominids, from Cercopithecidae, the Old World monkeys [7, 9]. While only a minority of a birth cohort of *P. leucopus* typically survive the predation and winter conditions of their first year in nature [10, 11], in captivity *Peromyscus* species can live twice as long as the laboratory mouse or rat [12]. *P. leucopus* differs in its social behavior and reproductive physiology from rodents that are traditional experimental models [13, 14].

P. leucopus also merits attention as a natural host and keystone reservoir for several tick-borne zoonoses of humans (reviewed in [15]). These include Lyme disease, babesiosis, anaplasmosis, a form of relapsing fever, ehrlichiosis, and a viral encephalitis. For humans these infections are commonly disabling and sometimes fatal, but *P. leucopus* is remarkably resilient in the face of persistent infections with these pathogens, singly or in combination. How this species tolerates infections to otherwise thrive as well as it does is poorly understood.

P. leucopus’ importance as a pathogen reservoir, its resilience in the face of infection, and its appealing features as an animal model [6, 16], prompted our genetic characterization of this species, beginning with sequencing and annotating its nuclear and mitochondrial genomes [17, 18]. The present study represents the third leg of this project, namely the microbial portion of the total animal “genome” for this species. Given the development of bait-delivered oral vaccines targeting *P. leucopus* [19] and plans to genetically modify and release this species [20, 21], pushing ahead on these interventional fronts without better understanding *Peromyscus* microbiota, the gastrointestinal (GI) tract’s in particular, seemed shortsighted.

Accordingly, we carried out a combined microbiologic and metagenomic study of the GI microbiome of *P. leucopus*. In our exploration of bacteria, protozoa, and DNA viruses in this species’ gut, we included whole genome sequences from cultured organisms as well as large contigs assembled from metagenome reads of uncultured organisms in our in-depth exploration of this. Attention initially focused on animals of a stock colony that has for many years been the major source of animals for different laboratories and spin-off breeding programs, including our own, in North America. The study extended to samples of *P. leucopus* deermice in their natural environments and, for a comparative animal, vivarium-reared *Mus musculus* under similar husbandry. While our investigations revealed similarities between the microbiota of the white-footed deermouse and the house mouse, there were also notable differences. These included a greater abundance and diversity of lactobacilli in *P. leucopus*. The investigation of four *Lactobacillus* species, particularly in their niches in the stomach of *P. leucopus*, was a special emphasis. A comparison of the GI microbiota of a natural population of *P. leucopus*

and the stock colony animals revealed several species in common, albeit with larger variance among the wild animals.

Results and discussion

High coverage sequencing of fecal metagenome

At the start of our investigation there was only limited information in the literature on the GI microbiota of *P. leucopus*, and the studies were restricted to data on bacterial 16S ribosomal RNA genes sequences, as in the examples of references [22] and [23]. For a fuller description of the microbiota, including parasites and DNA viruses, as well as bacteria, of representatives of this species, we took an untargeted, metagenomic approach. This was followed by bioinformatic characterization of the breadth of organisms represented in the metagenome and then assembly and annotation of large chromosomal contigs of selected microorganisms in the metagenome.

We began with deep sequencing of a library of DNA extracted from a pooled sample of fecal pellets collected from two adult males and two adult females of the LL stock colony. This yielded 332,279,332 paired-end reads of average length 247 nt, or $\sim 8 \times 10^{10}$ bp, after quality control and trimming of adapters. The mean % GC content was 47; 90% of the trimmed reads had PHRED scores of ≥ 30 . The reads were characterized as to families of bacteria, parasites, and DNA viruses at the metagenomic server MG-RAST (<http://mgrass.org>). Annotated proteins accounted for 65% of the reads, followed by unknown proteins at 34%, and then ribosomal RNA (rRNA) genes at 0.8%. The rarefaction curve became asymptotic at $\sim 200,000$ reads and a species count of 9000 (S1 Fig). The alpha diversity for this pooled sample was 250 species. By phylum 94% of the matched reads were either Bacteroidetes (60%) or Firmicutes (34%) (S2 Fig). Higher level functional categories included carbohydrates (16.4% of reads), clustering-based subsystems (14.8%), protein metabolism (8.9%), amino acids and derivatives (7.8%), RNA metabolism (6.6%), and DNA metabolism (5.7%) (S3 Fig).

A portion of the DNA was also submitted to a reference laboratory for 16S rRNA determinations from metagenomic analysis of microbiota. As illustrated in Fig 1 and detailed in S1 and S2 Tables, for the 20 most abundant taxa at the family or higher, there was concordance between the methods in the rankings. The most common families by the metagenomic accounting were members of the gram-negative bacterial order Bacteroidales (*Bacteroidaceae*, *Porphyromonadaceae*, *Prevotellaceae*, and *Rikenellaceae*), the gram-positive phylum Actinobacteria (*Bifidobacteriaceae* and *Coriobacteriaceae*), or the gram-positive phylum Firmicutes (*Bacillaceae*, *Enterococcaceae*, *Lactobacillaceae*, unclassified Clostridiales, *Eubacteriaceae*, *Lachnospiraceae*, *Peptococcaceae*, *Ruminococcaceae*, Thermoanaerobacterales Family III, *Erysipelotrichaceae*, and *Veillonellaceae*). Two exceptions were organisms of the families *Spirochaetaceae* of the phylum Spirochaetes and *Helicobacteraceae* of the phylum Proteobacteria.

A de novo assembly yielded 16,945 ungapped contigs of ≥ 10 kb from 197,369,943 reads and totaling 385 Mb of sequence with an average coverage of 104X. Of the total, 219 contigs were ≥ 100 kb in length and with ≥ 30 X coverage. These were used in searches of non-redundant nucleotide and protein databases for provisional classifications. The identified taxa included Bacteroidales, Clostridia, *Clostridiaceae*, Erysipelotrichales, *Lactobacillaceae*, *Muribaculaceae*, Firmicutes, and *Spirochaetaceae*. Three organisms represented among the high coverage contigs could be unambiguously classified as to species: *Lactobacillus animalis*, which is considered in detail below, and two *Parabacteroides* species: *distasonis* and *johnsonii*. Another *Lactobacillus* species represented among the highly ranked contigs could not be identified with a known species represented in the database [24].

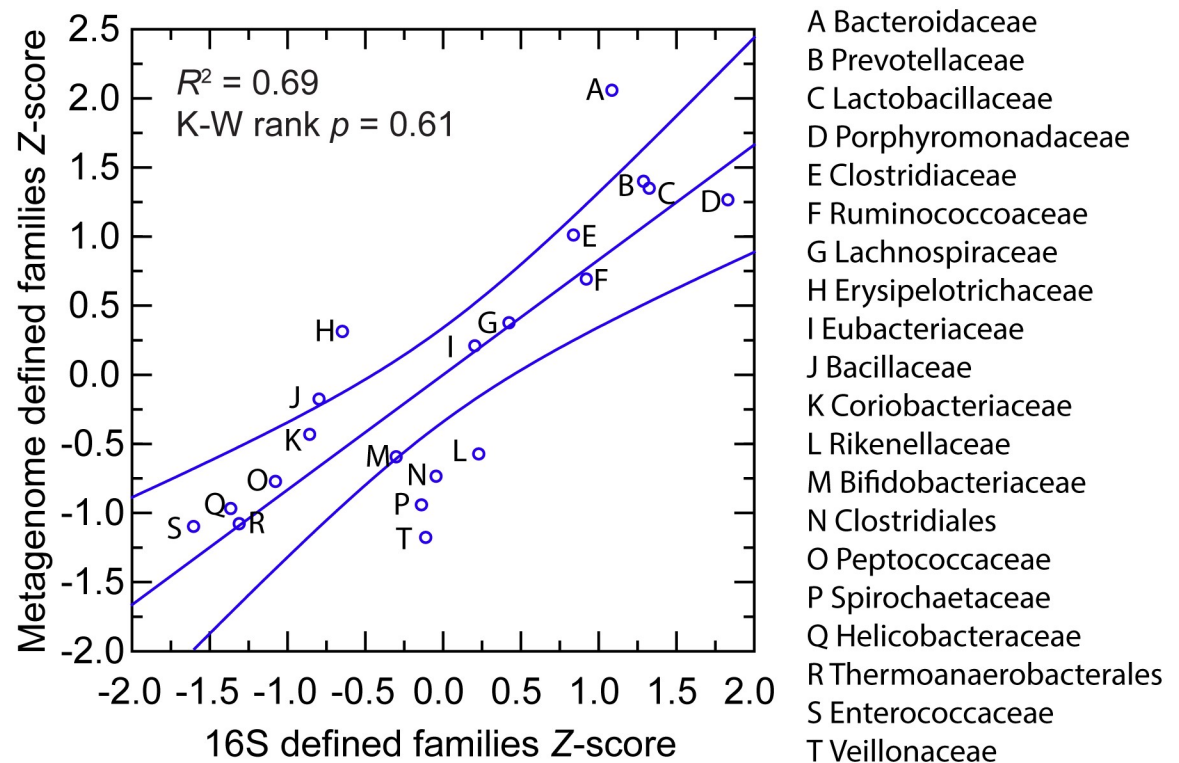


Fig 1. Scatter plot of relative abundances of commonly occurring bacterial families or orders in fecal metagenomes of *Peromyscus leucopus* LL stock by 16S ribosomal RNA gene criterion (x -axis) and by genome-wide metagenome criterion (y -axis). The values of the two methods were normalized by Z -score. The different taxa are indicated in the graph by capital letters, which are defined in the box to the right. The linear regression curve and its 95% confidence interval is shown. The coefficient of determination (R^2) value and the Kruskal-Wallis (K-W) test by ranks p value are given.

<https://doi.org/10.1371/journal.pone.0231801.g001>

Among the high coverage contigs were also representatives of Rhodospirillales of the class Alphaproteobacteria, *Mycoplasmataceae* of Mollicutes, and the little-characterized group of bacteria called Elusimicrobia [25]. Nearly as prevalent were organisms closely related to the phylum-level designation *Candidatus* Melainabacteria [26]. On the list of organisms identified by searches with metagenomic contigs of databases and cumulatively accounting for 95% of the matched reads were the unexpected finding of the protozoan taxon *Trichomonadidae* with 41,614 or 0.12% of the reads (S2 Table). *Enterobacteriaceae* at 0.3% accounted for a relatively small proportion of matched reads.

As in humans [27], *Bacteroidaceae*, *Lachnospiraceae*, *Prevotellaceae*, and *Ruminococcoaceae* were abundant in the gut metagenome and cumulatively accounted for approximately half of the identified families in the *P. leucopus* sample. One difference between humans and this *P. leucopus* sample was the much higher prevalence in *P. leucopus* of the family *Lactobacillaceae*, which on average represented only ~0.2% of metagenomes in a European population [27] and by 16S sequencing $\leq 0.4\%$ on the fecal microbiota in other studies [28]. A higher proportion of lactobacilli in the fecal microbiota was previously noted in other rodents [29] and specifically in *Peromyscus* species [22].

Selected taxa

Escherichia coli. Although *Enterobacteriaceae* were infrequently represented among the metagenomic sequences, their cultivability under routine laboratory conditions and the

availability of a vast database for some species prompted our isolation of *Enterobacteriaceae* from LL stock *P. leucopus* fecal pellets on selective media. The predominant isolate on the plates was an *Escherichia coli*, which we designated LL2. The whole genome sequence of isolate LL2's chromosome and plasmids was sequenced and assembled using a hybrid of long reads and short reads (Table 1) for an overall coverage of 90X. The chromosome in two contigs of 3.4 Mb and 1.6 Mb totaled 5.0 Mb with a GC content of 50%.

E. coli LL2 had the following MLST schema types (<http://pubmlst.org> or <http://enterobase.org>): Achtman ST-278, Pasteur ST-357, and ribosomal protein ST-122394. The ribosomal protein profile was unique among thousands of isolates in the database. The 121 kb, 56.5 kb, and 91 kb plasmids of strain LL2 were similar to the following *E. coli* plasmids, respectively: a 185 kb plasmid (NC_007675) found in an avian strain, a 58 kb plasmid (CP024858) of a multiply antibiotic-resistant human isolate, and an 89 kb plasmid (CM007643) in an organism isolated from sewage. *E. coli* LL2 was susceptible to ampicillin, ciprofloxacin, gentamicin, and sulfamethoxazole-trimethoprim by in vitro testing.

The chromosome was notable for the following: CRISPR-Cas1 and -Cas3 arrays; ISas1, ISNCY, IS3, IS110 and IS200 family transposases; restriction-modification systems; fimbria and curli biosynthesis and transport systems; type II toxin-antitoxin systems; and type II, type III and type VI secretion systems. The plasmids encoded fimbrial and pilin proteins, type I, type II, and type IV secretion systems, colicins, CdiA-type contact-dependent inhibition toxin, and three conjugative transfer systems, but no discernible coding sequences for antibiotic resistance.

Serial dilutions of feces of LL stock 20 animals (11 females and 9 males) in phosphate-buffered saline and plated on agar selective for gram-negative enteric bacteria yielded a mean (asymmetric 95% confidence interval) of 3,491 (677–18,010) colony-forming units (cfu) of *E. coli* per g of feces. This low density was consistent with the findings from metagenomic sequencing.

The origin of this *E. coli* strain in these animals is obscure. Its source may have been the vivarium for the breeding stock of the colony, which includes other *Peromyscus* species besides *P. leucopus*. But the strain appears to be stably maintained among the gut microbiota of this population of *P. leucopus*. This adaptation and its susceptibilities to commonly-used antibiotics may make it a candidate as a vector for delivering oral vaccines to this species [30].

Lactobacillus. We isolated lactobacilli from fecal pellets of stock colony *P. leucopus* on plates of selective medium that were incubated under microaerophilic and hypercapnic conditions at 37 °C. Four different species were identified. The genomes of three of organisms, namely *L. animalis* strain LL1, *L. reuteri* strain LL7, and a new species, designated as *Lactobacillus* sp. LL6 and provisionally named as “*L. peromysci*”, have been reported [24]. The fourth genome, of the LL8 strain of *L. johnsonii*, is described first here (Table 1). *L. johnsonii*'s chromosome from cumulative contigs was 2,045,501 bp, about the same size as that of “*L. peromysci*” at 2,067,236 bp, but shorter than the 2,280,577 bp length for *L. animalis* and 2,205,740 bp for *L. reuteri*. The % GC content of “*L. peromysci*” at 33.5 was closer to *L. johnsonii* (34.4) than to either *L. animalis* (41.0) or *L. reuteri* (38.9).

Fig 2 is a distance phylogram of 1385 aligned sites of 16S ribosomal RNA genes for the four different lactobacilli. These were distributed across four major groups of the genus *Lactobacillus*. The phylogenetic relationships were examined in more depth by multilocus sequence typing of the 53 genes for ribosomal proteins. These were identified in the genomes, compared with other deposited sequences in the ribosomal MLST database (<https://pubmlst.org>) [31], concatenated, and then aligned with analogously concatenated DNA sequences from related species (S3 Table). Bacteria with identical sequences for the 53 ribosomal protein genes were not found in the rMLST database of 133,460 profiles. The % GC contents of the concatenated

Table 1. Resources from this study.

Organism and strain	Description	BioProject	BioSample	SRA or MG-RAST ^a	Accession No.
Cultured <i>Lactobacillus johnsonii</i> LL8	2,045,501 bp WGS chromosome; 40 contigs	PRJNA589091	SAMN13266521	SRX7128459	WKKC00000000
Cultured <i>Lactobacillus johnsonii</i> LL8	75,746 bp plasmid	PRJNA589091	SAMN13266521	SRX7128459	CM019125
Cultured <i>Escherichia coli</i> LL2	Chromosome; 3,345,873 and 1,610,537 bp contigs	PRJNA533838	SAMN11468944	SRR9087223 SRR9087224	VBVB01000001 VBVB01000002
Cultured <i>Escherichia coli</i> LL2	121,192 bp plasmid	PRJNA533838	SAMN11468944	SRR9087223 SRR9087224	CM017030
Cultured <i>Escherichia coli</i> LL2	90,617 bp plasmid	PRJNA533838	SAMN11468944	SRR9087223 SRR9087224	CM017032
Cultured <i>Escherichia coli</i> LL2	56,474 bp plasmid	PRJNA533838	SAMN11468944	SRR9087223 SRR9087224	CM017031
Gut metagenome	50–450 kb contigs of fecal metagenome	PRJNA540317	SAMN11533720	mgm4799371.3 mgm4799372.3	JAAGKN000000000
Uncultured <i>Helicobacter</i> sp. LL4	290,716 bp 172,100 bp 203,294 bp chromosome fragments	PRJNA540317	SAMN11533720	mgm4799371.3 mgm4799372.3	MN577567 MN577568 MN577569
Uncultured <i>Helicobacter</i> sp. LL4	16S ribosomal RNA gene, partial	n.a. ^b	n.a.	n.a.	MT114577
Uncultured Candidatus Melainabacteria bacterium isolate LL20	270,170 bp chromosome fragment	PRJNA540317	SAMN11533720	mgm4799371.3 mgm4799372.3	MN577570
Uncultured Elusimicrobia bacterium LL30	232,820 bp 215,518 bp chromosome fragments	PRJNA540317	SAMN11533720	mgm4799371.3 mgm4799372.3	MN577571 MN577572
Uncultured Clostridiales bacterium LL40	438,773 bp chromosome fragment	PRJNA540317	SAMN11533720	mgm4799371.3 mgm4799372.3	MN577573
Uncultured <i>Spirochaetaceae</i> bacterium LL50	277,828 bp chromosome fragment	PRJNA540317	SAMN11533720	mgm4799371.3 mgm4799372.3	MN577574
Uncultured <i>Prevotella</i> sp. LL70	456,702 bp 379,405 bp chromosome fragments	PRJNA540317	SAMN11533720	mgm4799371.3 mgm4799372.3	MN990733 MN990734
Uncultured Rhodospirillales bacterium LL75	145,048 bp 150,471 bp 130,339 bp 104,625 bp 177,737 bp chromosome fragments	PRJNA540317	SAMN11533720	mgm4799371.3 mgm4799372.3	MN990728 MN990729 MN990730 MN990731 MN990732
Uncultured <i>Mycoplasmataceae</i> bacterium LL85	126,601 bp 100,243 bp chromosome fragments	PRJNA540317	SAMN11533720	mgm4799371.3 mgm4799372.3	MN991199 MN991200
Uncultured <i>Muribaculaceae</i> bacterium LL71	125,326 bp 103,384 bp chromosome fragments	PRJNA540317	SAMN11533720	mgm4799371.3 mgm4799372.3	MT002444 MT002445
Uncultured <i>Tritrichomonas</i> sp. LL5	1,501 bp of small subunit ribosomal RNA	PRJNA540317	SAMN13920683	mgm4864879.3	MN120899
Uncultured <i>Tritrichomonas</i> sp. LL5	989 bp partial iron hydrogenase gene of hydrogenosome	PRJNA540317	SAMN13920683	mgm4864879.3	MN985504
Uncultured <i>Tritrichomonas</i> sp. LL5	5298 bp fragment with DNA polymerase type B, organellar and viral family protein	PRJNA540317	SAMN13920683	mgm4864879.3	MT002461
Uncultured <i>Lactobacillus</i> sp. (“peromysci”) B17442	<i>rpsA</i> , <i>ftsK</i> , <i>ftsZ</i> , <i>dnaA</i> , <i>dnaN</i> , <i>recD</i> , <i>ileS</i> , <i>recA</i> , <i>topA</i>	PRJNA593618	SAMN13482862	SRX7285441	MN792760 - MN792768
Uncultured <i>Lactobacillus animalis</i> 7442BI	51 large and small ribosomal proteins	PRJNA593618	SAMN13482862	SRX7285441	MN817867 - MN817918

<https://doi.org/10.1371/journal.pone.0231801.t001>

coding sequences were 39.5, 40.8, 42.2, and 42.3 for *L. johnsonii*, “*L. peromysci*”, *L. reuteri*, and *L. animalis*, respectively. Fig 3 shows the distance phylograms for ~20 kb of aligned positions for the four species, each grouped with other strains or species within their respective phylogenetic clusters.

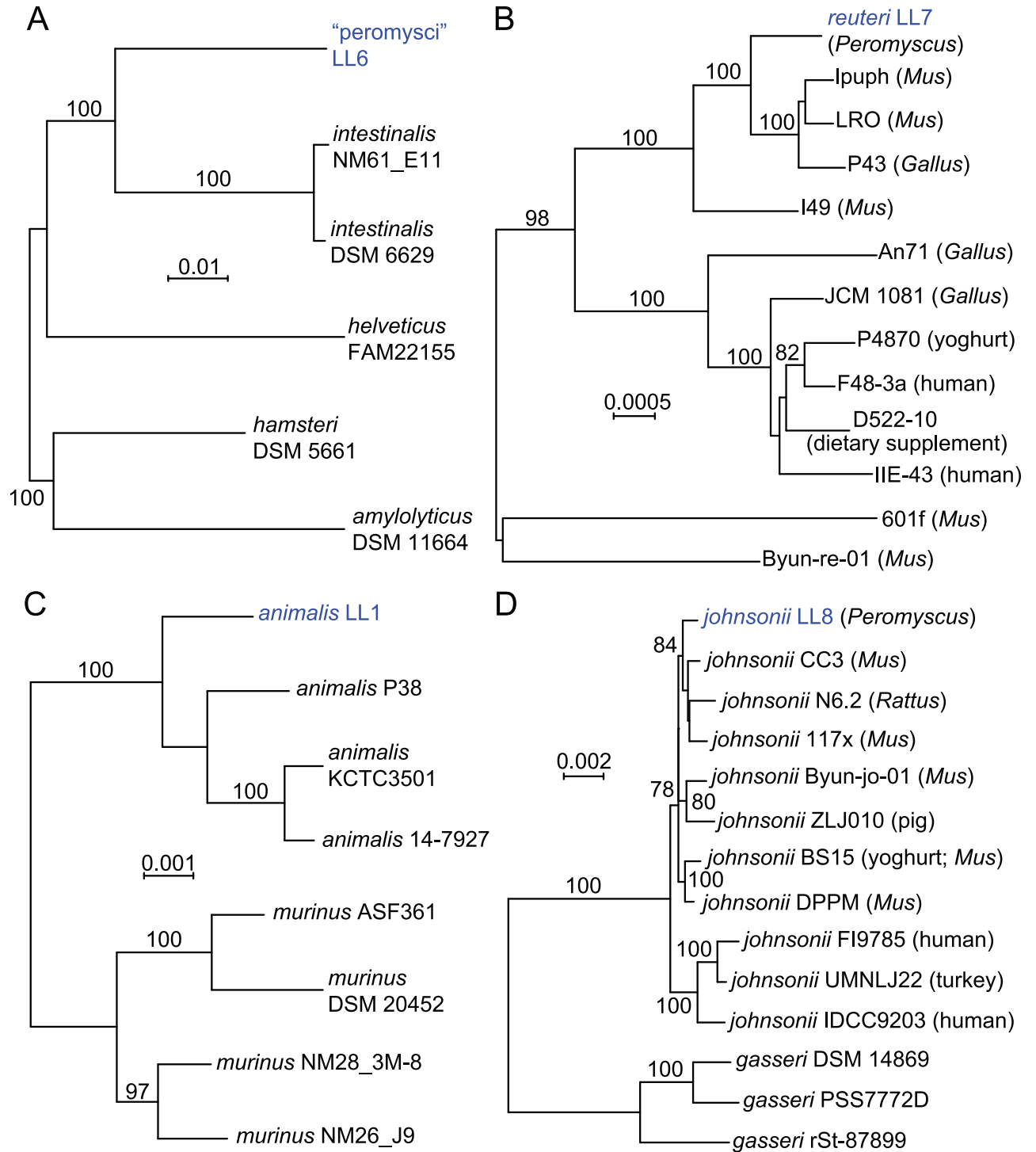


Fig 3. Neighbor-joining distance phylograms of codon-aligned, concatenated nucleotide sequences for complete sets of ribosomal proteins of “*L. peromysci*” (panel A), *L. reuteri* (panel B), *L. animalis* (panel C), and *L. johnsonii* (panel D) of *P. leucopus* compared with *Lactobacillus* spp. (strain identifier) of other sources. The scales for distance by Jukes-Cantor criterion are indicated in each panel. Percent bootstrap (100 iterations) support values of $\geq 75\%$ at a node are shown. In panels B and D the host animal or other origin for a given isolate are given in parentheses.

<https://doi.org/10.1371/journal.pone.0231801.g003>

Table 2. Selected genes and pathways in four species of *Lactobacillus* of the gastrointestinal microbiota of *Peromyscus leucopus*.

Species	Urease	secY2-secA2	secY3-secA3	IgA protease (pfam07580)	L-rhamnose pathway	luxS	Type 1 CRISPR Cas	Thiamine biosynthesis	Nitroreductase nfnB-nifU	Arginine deiminase/repressor	Type VII secretion system
<i>reuteri</i>	+	+	+	-	-	+	-	-	-	+	-
<i>johnsonii</i>	-	+	-	+	-	+	-	+	+	-	-
<i>animalis</i>	-	+	-	-	-	-	-	+	+	-	+
"peromysci"	+	-	-	-	+	-	+	+	+	-	-

<https://doi.org/10.1371/journal.pone.0231801.t002>

L. animalis and *L. murinus* are closely related species that primarily have been associated with GI microbiota of rodents and some other mammals. Isolate LL1 grouped with representatives of *L. animalis* in the analysis (panel C) and not *L. murinus* [34]. LL1's 16 ribosomal RNA sequence was identical to that of the type strain ATCC 35046 of *L. animalis* [35] at 1488 of 1489 positions (GCA_000183825) [36]. Another pair of closely-related species are *L. johnsonii* and *L. gasseri*, for which there are several sequenced genomes. The LL8 isolate from fecal pellets of *P. leucopus* clustered with *L. johnsonii* strains from mice and rats (panel D). More distant were strains of *L. johnsonii* isolated from humans and a bird; more distant still were representatives of *L. gasseri*.

Plasmids were identified in each of the four species on the basis of a circularly permuted sequence for a contig and presence of coding sequences that were homologous to known plasmid replication or partition proteins (Table 1). Large plasmids of 179 kb and 76 kb were present in *L. reuteri* and *L. johnsonii*, respectively. *L. animalis* and "L. peromysci" had small plasmids of 4 kb and 7 kb, respectively. Megaplasmids of greater than 100 kb have been observed in other *Lactobacillus* spp. [37]. In all genomes there was evidence of lysogenic bacteriophages or their remnants. All species except *L. reuteri* discernibly had coding sequences for Class I or Class III bacteriocins or their specific transport and immunity proteins (S4 Table).

Table 2 summarizes differentiating genetic profiles among the four species for 11 selected genes or pathways. Two species, *L. reuteri* and "L. peromysci", had coding sequences for a urease, which could provide for tolerance of acidic conditions, such as in the stomach. A urease had previously been identified in a *L. reuteri* strain that was considered a gut symbiont in rodents [38]. The four species had *secY1-secA1* transport and secretion systems. Accessory Sec systems (*secY2-secA2*) were identified in genomes of *L. reuteri*, *L. johnsonii*, and *L. animalis* but not in "L. peromysci". The LL7 strain of *L. reuteri* on its megaplasmid also had coding sequences for a third SecY-SecA system. An accessory Sec system was involved with adhesion and biofilm formation in the Lactobacillales bacterium *Streptococcus pneumoniae* [39]. A coding sequence for an IgA protease was identified in *L. johnsonii* but not in the other three species. An IgA protease in another strain of *L. johnsonii* was associated with long-term persistence in the gut of mice [40]. The presence or absence of other genes or pathways that differentiated between the four species were an L-rhamnose biosynthesis pathway in one species, a *luxS* gene associated with a quorum sensing system in *L. reuteri* and *L. johnsonii* [41], a type 1 CRISPR-Cas3 array in "L. peromysci" [42], pathways for thiamine biosynthesis [43] and for reduction of nitrate [44] in three species, an arginine deiminase and its repressor in *L. reuteri* [45], and a type VII secretion system in *L. animalis* [46].

Of the four species found in *P. leucopus* feces, only *L. johnsonii* and *L. reuteri* have been commonly isolated from human feces [28]. While various strains of *L. reuteri*, *L. johnsonii*, and either *L. animalis* or the closely related *L. murinus* have been observed among the GI microbiota of *M. musculus* representatives [47], an organism similar to "L. peromysci" has not. Whether this is an indication of a restricted host range or a specific adaptation for this

bacterium is considered below. *Lactobacillaceae* organisms were previously noted in the gut microbiota of *Peromyscus* species but were not further classified as to species or otherwise characterized [22, 23].

Helicobacter. Among the assembled metagenomic contigs were three totaling 666,100 bp of a *Helicobacter* genome (Table 1). The contigs had non-overlapping in genetic content, and blast searches with translated genes from each of the 3 contigs yielded the identical rankings of taxa for homologous proteins. On these bases, we concluded that the contigs represented a single type of *Helicobacter* bacterium, and designated it strain LL4. Using the DNA sequences for 53 ribosomal proteins of this organism, we compared it with similar sets from other *Helicobacter* spp. (panel A of Fig 4). This analysis showed that the organism was near-identical to orthologous sequences of *Helicobacter* sp. MIT 05–5293 (accession JROZ02000000), which had been cultivated from a wild *P. leucopus* captured in Massachusetts ([48]; J.G. Fox, personal communication). This finding indicated that the organism was autochthonous for *Peromyscus* and had not been acquired from another rodent housed in the same vivarium. Baxter et al. reported “abundant representatives” of the genus *Helicobacter* in wild *P. maniculatus* and *P. leucopus* captured in Michigan [22].

Strains LL4 and MIT 05–5293 are in a cluster of species known as “enterohepatic” *Helicobacter* for their primary residence in the intestine rather than the stomach and for their frequent presence in liver tissue [49]. These species may not be benign. *H. hepaticus* is associated with hepatitis, bowel inflammation, and carcinoma [50], and *H. typhlonius* is associated with reduced fecundity in mice [51].

Spirochaetaceae. The phylogenetic relationship of the uncultured *Spirochaetaceae* bacterium LL50 to selected other spirochetes (S5 Table) is shown in panel B of Fig 4. The organism was classified in the genus *Treponema* by the MG-RAST analysis program. Bacteria that have assigned to this genus are highly divergent and include free-living organisms in a variety of environments, symbionts of termites, the agent of syphilis, and gut residents, such as *T. porcinum*, which was isolated from the feces of pigs [52]. More distant still was the agent of Lyme disease, *Borrelia burgdorferi*, of the family *Borreliaceae* [53]. In our view naming the organism as a “treponeme” on the evidence to date would provide little insight about its role in the microbiome and may even be misleading.

Seven other bacterial taxa. The *Mycoplasmataceae* bacterium LL85 (panel C of Fig 4) was unlike any other mollicute represented in the database but was in a cluster with vertebrate-associated species, like *M. pneumoniae*. But there is also deep branching in this tree, as the tree including two Firmicutes as outgroups shows (S5 Table). Panel D is a phylogram of selected alphaproteobacteria and includes the organism LL75 (S5 Table). The algorithmic analysis identified this at the genus level as *Azospirillum*, which is a largely uncharacterized taxon with highly divergent members. While assignments as to genus or family are uncertain at this time, LL75 clustered within the order Rhodospirillales and not with rhizobacteria (S5 Table).

Table 1 lists five other types of novel bacteria that were identified in the *P. leucopus* gut metagenome and partially sequenced and annotated. These were a *Candidatus* Melainabacteria bacterium (isolate LL20), an Elusimicrobia bacteria (isolate LL30), a Clostridiales bacterium (isolate LL40), a *Prevotella* species (isolate LL70), and a *Muribaculaceae* bacterium (isolate LL71). *Candidatus* Melainabacteria is either a non-photosynthetic sister phylum of cyanobacteria or a class within the phylum Cyanobacteria [26]. Besides a variety of environmental sources, including hot springs and microbial mats, these poorly-characterized organisms have also been identified in the feces of humans and other animals. The phylum Elusimicrobia, formerly “Termite Group 1” [54], is a strictly-anaerobic, deeply-branched lineage of gram-negative bacteria, representatives of which were first observed in the hindgut of termites [25]. The family *Muribaculaceae* (formerly “family S24-7”) of the order Bacteroidales

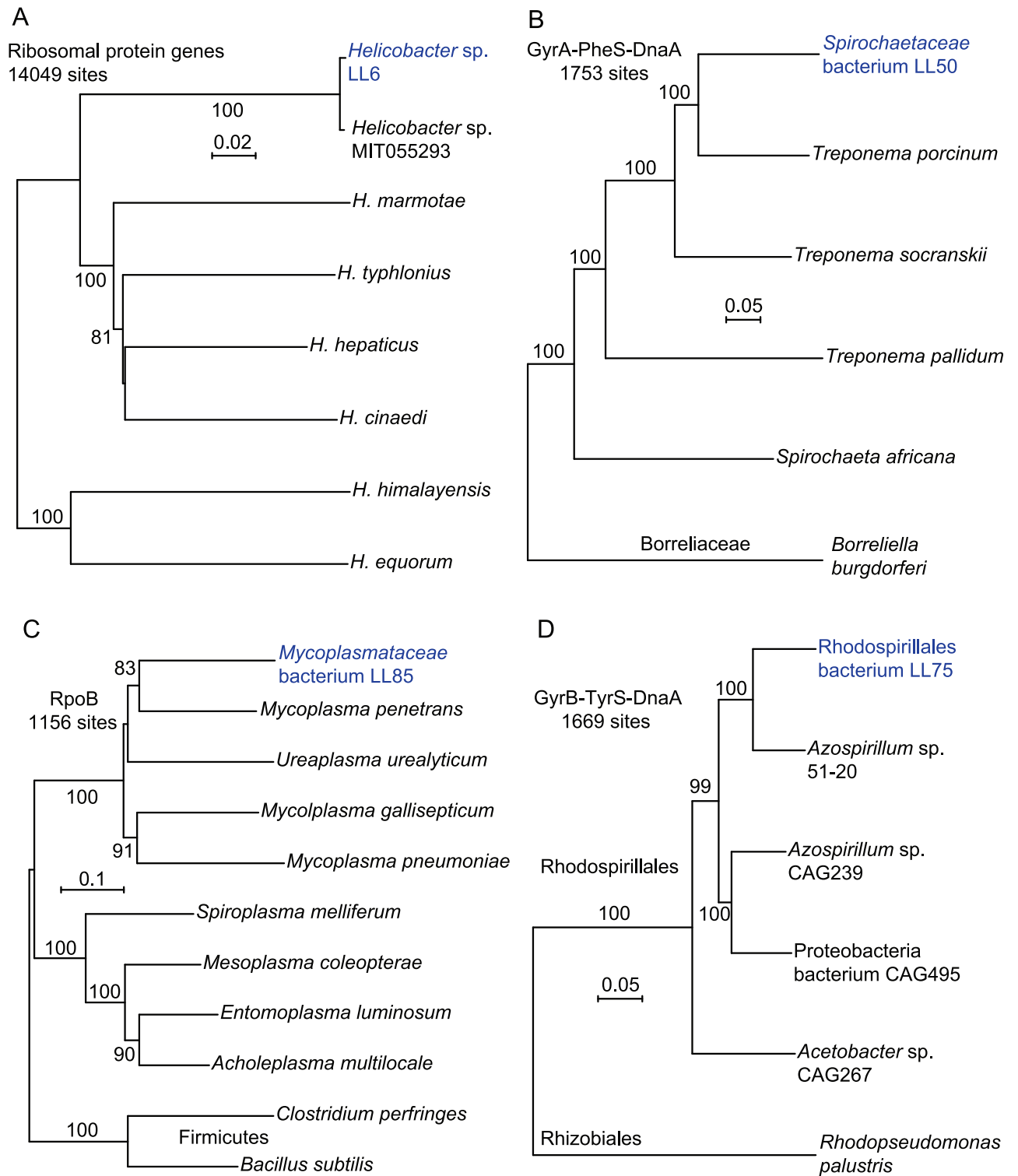


Fig 4. Neighbor-joining distance phylograms of concatenated nucleotide (panel A) or amino sequences (panels B-C) of *Helicobacter* spp. (panel A), *Spirochaetaceae* bacteria (panel B), Mollicutes and Firmicutes bacteria (panel C) and Rhodospirillales bacteria (panel D) of gut metagenome of *P. leucopus* and from other sources. The accession numbers for the other bacteria represented in the phylograms are given in S5 Table. The respective phylogenetic analyses used concatenated sequences of the following: ribosomal protein genes (panel A); the DNA gyrase A (GyrA), phenylalanyl t-RNA synthase, alpha subunit (PheS), and chromosomal replication initiator protein (DnaA) (panel B); DNA-directed RNA polymerase, beta-subunit (RpoB) (panel C); and DNA gyrase B (GyrB), tyrosyl t-RNA synthase (TyrS), and DnaA (panel D). The distance criteria were Jukes-Cantor for the codon-aligned nucleotide sequences and Poisson for amino acid sequences. The scales for distance are shown in each panel. Percent bootstrap (100 iterations) support values of $\geq 80\%$ at a node are shown.

<https://doi.org/10.1371/journal.pone.0231801.g004>

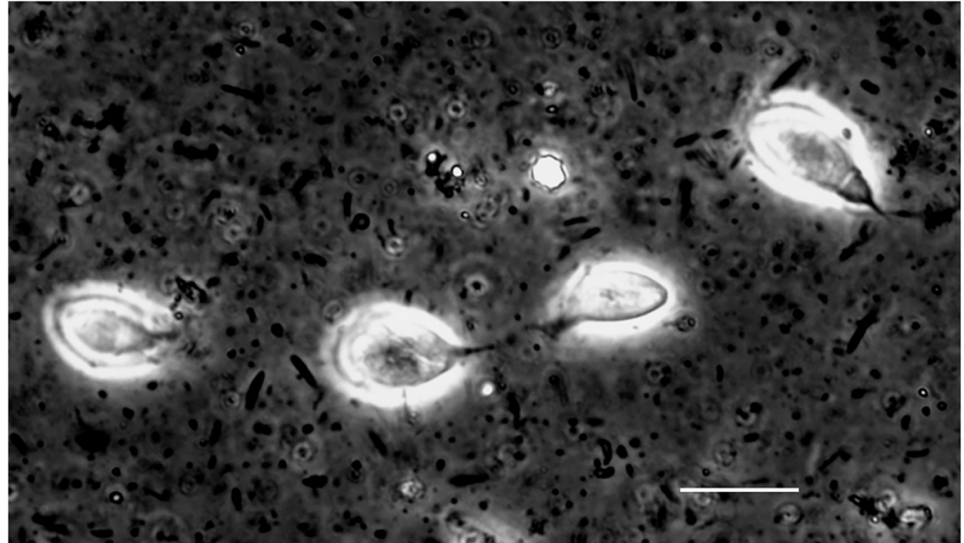


Fig 5. Photomicrograph of live *Tritrichomonas* flagellated protozoan in the cecal fluid of *P. leucopus* LL stock. Four organisms against the background of intestinal bacteria were visualized in the wet mount by differential interference microscopy. Bar, 10 μ m.

<https://doi.org/10.1371/journal.pone.0231801.g005>

were first identified among gut microbiota of mice and subsequently in the intestines of other animals, including humans and ruminants [55].

DNA viruses. Of 112,677,080 reads of the metagenome high-coverage sequencing of the LL stock animals, 97,147 (0.09%) were assigned to one of 28 DNA virus families. Three classifications accounted 92% of the reads: *Siphoviridae* (50%), which are bacteriophages with long contractile tails; *Myoviridae* (21%), which are bacteriophages with contractile tails; and “unclassified viruses” (21%). At the species level, 31,812 (68%) of the 46,904 *Siphoviridae*-matching reads mapped specifically to bacteriophages of *Lactobacillus* spp.

***Tritrichomonas* protozoan.** Intestinal flagellated protozoa named “*Trichomonas muris*” or “*Tritrichomonas muris*” had previously been identified in wild *P. leucopus* and *P. maniculatus* [56]. While laboratory mice are typically free of intestinal protozoa [57], the anaerobic *Tritrichomonas muris* has been reported in some populations of colony *M. musculus* [58]. To further investigate the protozoa that were provisionally identified as “*Trichomondidae*” at the family level in the metagenome analysis, we euthanized 14 healthy adult animals (6 females and 8 males) and examined fresh cecal contents by phase microscopy. Six of the LL stock animals had been born at the PGSC facility, and 8 had been born at U.C. Irvine.

In each of the 14 animals examined there were numerous motile flagellates consistent in morphology with *T. muris* in their ceca [59]. These were each at a cell density of $\sim 10^6$ per milliliter of unconcentrated cecal fluid (Fig 5; S1 Movie). Entire ceca and their contents from two adult females and two adult males were subjected to DNA extraction, library preparation from the DNA, sequencing, and de novo assembly of contigs.

Fig 6 shows phylograms of nucleotide sequence of the small subunit (SSU) ribosomal RNA gene (panel A; S5 Table) and of the partial amino acid sequence of the iron hydrogenase of the hydrogenosome of anaerobic protozoa (panel B; S5 Table) [60]. The SSU of isolate LL5 indicates that it is probably synonymous with *Tritrichomonas muris*, for which only a SSU sequence was available. The sequence of the iron hydrogenase further supported placement in the genus *Tritrichomonas*. *Histomonas melagridis*, a sister taxon by this analysis, is recognized

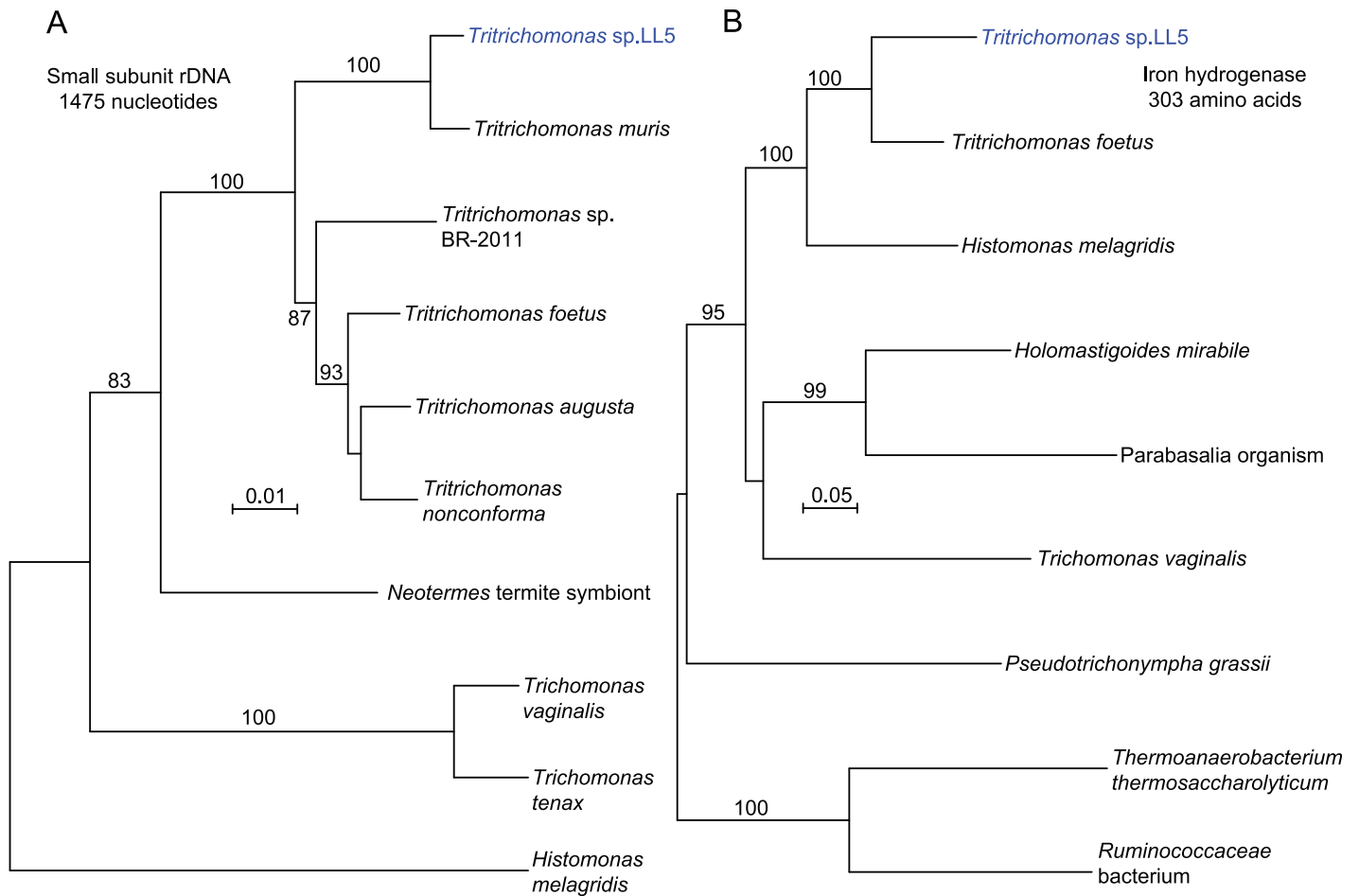


Fig 6. Neighbor-joining distance phylograms of nucleotide (panel A) or amino acid (panel B) partial sequences of small subunit ribosomal RNA gene (rDNA) (panel A) and iron hydrogenase protein (panel B) of *Tritrichomonas* sp. LL5 of *P. leucopus* and selected other parabasilids and other microbes. The distance criteria were observed differences for nucleotide alignment and Poisson for amino acid alignment. The scales for distance are shown in each panel. Percent bootstrap (100 iterations) support values of $\geq 80\%$ at a node are shown.

<https://doi.org/10.1371/journal.pone.0231801.g006>

as a pathogen of poultry. Another sequence of the LL5 organism encodes a type B DNA polymerase (Table 1), which likewise matched closely with an ortholog in the *Tritrichomonas foetus* genome sequence (S4 Fig). *T. foetus* is a sexually-transmitted pathogen of cattle [61] and a cause of chronic diarrhea in domestic cats [62].

Whether the *T. muris* is a commensal shared across natural populations of *Peromyscus* or a parasite acquired from another rodent during the colony's history in a vivarium remains to be determined. As related below, there is sequence evidence of the same or related organism in several wild animals. Whatever the case, these organisms may have an effect on immune responses of *P. leucopus*, as has been reported for *T. muris* in *M. musculus* [63–65], and their presence needs to be taken into account in interpreting experimental results in the laboratory and in applications for field interventions.

Comparative study of GI microbiota of *P. leucopus* and *M. musculus*

The preceding study revealed several microbes that were either undescribed species or genera, e.g. “*L. peromysci*” or the *Candidatus* Melainabacteria bacterium, or new strains of known

microbial species, e.g. *L. animalis* LL1 and *T. muris* LL4. These novelties notwithstanding, to what extent did the gut microbiota of this deermouse resemble that of the typical laboratory animal, a house mouse that was maintained under similar husbandry conditions, including diet? That question motivated the following experiment.

Fecal pellet samples from 20 adult *P. leucopus* (10 females and 10 males) and 20 adult BALB/c *M. musculus* (10 females and 10 males) were obtained and stored frozen at -80°C until processing. All animals were approximately 10 weeks old. The animals were housed in the same vivarium facility, though in different rooms. The pellets were subjected to total DNA extractions, and paired-end Illumina sequencing with 250 cycles of indexed libraries were carried out. There were means (95% CI) of $3.4 (3.1\text{--}3.7) \times 10^6$ post-quality control reads for *P. leucopus* samples and $3.4 (3.2\text{--}3.6) \times 10^6$ for *M. musculus* samples (S6 and S7 Tables).

The reproducibility between replicate library constructions from the same sample was assessed with quantitations of reads assigned by taxonomic family for specimens from seven *P. leucopus* among the 20 total. Pairwise coefficients of determination (R^2) for the 91 possible combinations were calculated (S8 Table). The mean (95% CI) of R^2 values were 0.999 (0.999–1.0) for the 7 pairs of replicates and 0.930 (0.915–0.944) for the 84 non-replicate pairs. We concluded that most of the variation between samples was attributable to inter-specimen differences in the microbiota and not to technical issues in library preparation or sequencing.

The prevalences of different taxonomic families in the *P. leucopus* and *M. musculus* gut metagenomes were similar (left panel of Fig 7 and S9 Table). But a few families stood out as either more or less common in the deermice. Notable among these were *Lactobacillaceae*, *Helicobacteriaceae*, and *Spirochaetaceae*, which were approximately 4x, 8x, and 2x, respectively, more prevalent on average among microbiota of *P. leucopus* than in *M. musculus*. There was no evidence of *Tritrichomonas* sp. in the BALB/c mice by this analysis, but direct examination of intestinal contents was not carried out.

At the level of 86 operational KEGG pathways, the metagenomes of *P. leucopus* and *M. musculus* were nearly indistinguishable (right panel of Fig 7 and S10 Table). But, as shown in the heat map of Fig 8, at the homologous gene level there were many differences between these two species and also between females and males within each species (S12 and S13). Hierarchical clusters 2 and 4 of the analysis discriminated between mice and deermice regardless of sex, while clusters 1 and 3 signified marked differences by sex and less so by species.

As one example of differences between species, there was higher representation of genes of the mevalonate pathway in the gut metagenomes of *P. leucopus*. Beginning with acetyl-CoA and ending with isopentenyl pyrophosphate, the central intermediate in the biosynthesis of isoprenoids in all organisms [66], the coding sequences for the following ordered enzymes (with Enzyme Commission [EC] number) in the pathway were comparatively higher in frequency: acetyl-CoA C-acetyltransferase (EC:2.3.1.9), hydroxymethylglutaryl-CoA synthase (EC:2.3.3.10), hydroxymethylglutaryl-CoA reductase (EC:1.1.1.88), mevalonate kinase (EC:2.7.1.36), phosphomevalonate kinase (EC:2.7.4.2), and diphosphomevalonate decarboxylase (EC:4.1.1.33).

We further investigated specific differences between *P. leucopus* and *M. musculus* and between individual animals of each species in *Lactobacillus* spp. [67]. This was achieved by mapping reads to references of the chromosome sequences of the four species that had been isolated from the feces of LL stock *P. leucopus*. The caveat is that the lactobacilli in the mice would not be expected to be identical to the deermouse strains used as references. Fig 9 shows box plots for *Peromyscus* on the left and for *Mus* on the right for data given in S11 Table. Included in the analysis of *P. leucopus* gut metagenome reads were selected other bacteria that had been frequently identified among the metagenomic contigs and then further characterized by partial genome sequencing (see above).

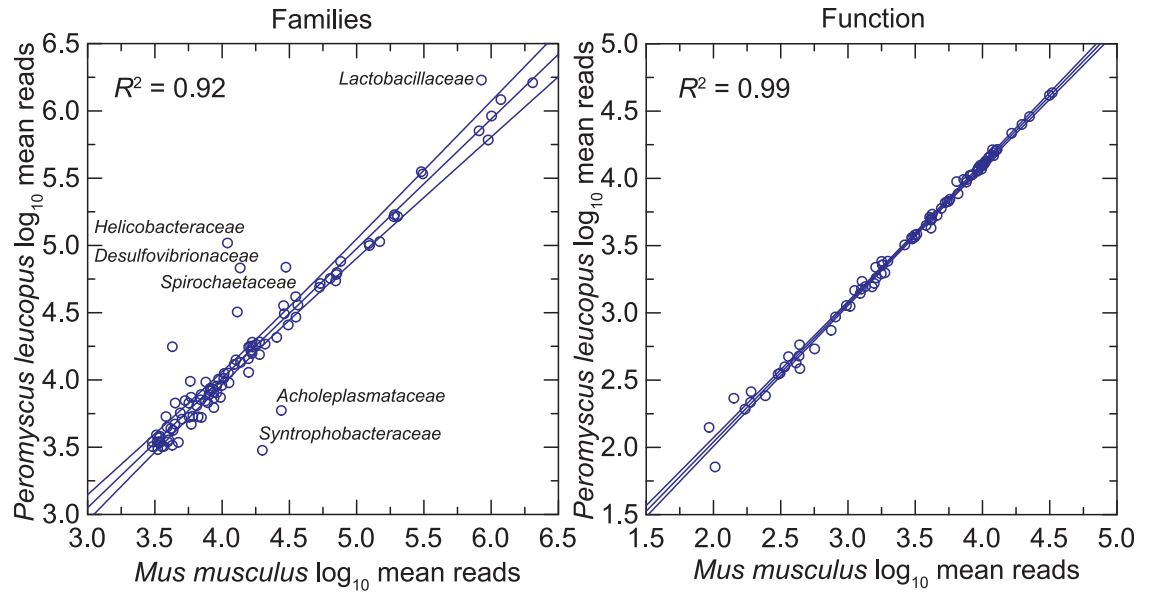


Fig 7. Scatter plots of log-transformed normalized reads of the gut metagenomes of 20 *P. leucopus* on the gut metagenomes of 20 *M. musculus* by bacterial families (left panel) or by function at the pathway level (right panel). The linear regression lines, their 95% confidence intervals, and coefficients of determination (R^2) are shown. Selected families that are comparatively more or less prevalent in *P. leucopus* are indicated.

<https://doi.org/10.1371/journal.pone.0231801.g007>

All four species of the lactobacilli were represented in each of the 20 *P. leucopus* metagenomes. “*L. peromysci*” and *L. reuteri* tended to be the most common and consistently represented, while *L. johnsonii* and *L. animalis* varied more in prevalences between animals. Other bacteria were also identified in the samples of all or most of the individual animals. The *Spirochaetaceae* bacterium was ~10-fold less abundant than the cumulative *Lactobacillus* spp. in the *P. leucopus* samples.

The mean number of lactobacilli in aggregate were ~2-fold more prevalent in *P. leucopus* females than males of the species (t -test $p = 0.03$). In *M. musculus* this sex difference for *Lactobacillus* was more pronounced; on average ~100-fold more reads from female mice mapped to *Lactobacillus* genomes than was found for male mice (t -test $p < 0.001$). The differences in amounts of fecal lactobacilli in the sample plausibly account for cluster 3 of the heatmap of Fig 8. *L. johnsonii* largely accounted for these differences between sexes in *M. musculus*; nearly all of the reads mapping to the *Lactobacillus* genus as a whole were mapping to the *L. johnsonii* genome. The three other species identified in *P. leucopus* were either not present or in much lower numbers in this sampling of *M. musculus*. Strains of *L. johnsonii* have been commonly detected in feces of laboratory mice [68].

A limitation to the study was that the LL stock animals were outbred, and the BALB/c mice were inbred. An inbred lineage derived from the LL stock population was not available. On the other hand, this distinction provided a comparison of microbiome diversities between an outbred and inbred population. As expected, there was greater alpha diversity among the outbred samples than the inbred (Fig 10). Another limitation was the dependence on fecal pellets collected at one time point. The samples were from similar age *P. leucopus* and *M. musculus* and were obtained from the animals and then processed on the same day, but for this study we did not assess variation within individuals over time. The gut microbiomes of the mothers of these animals were uncharacterized, so we could not evaluate maternal effects in this study.

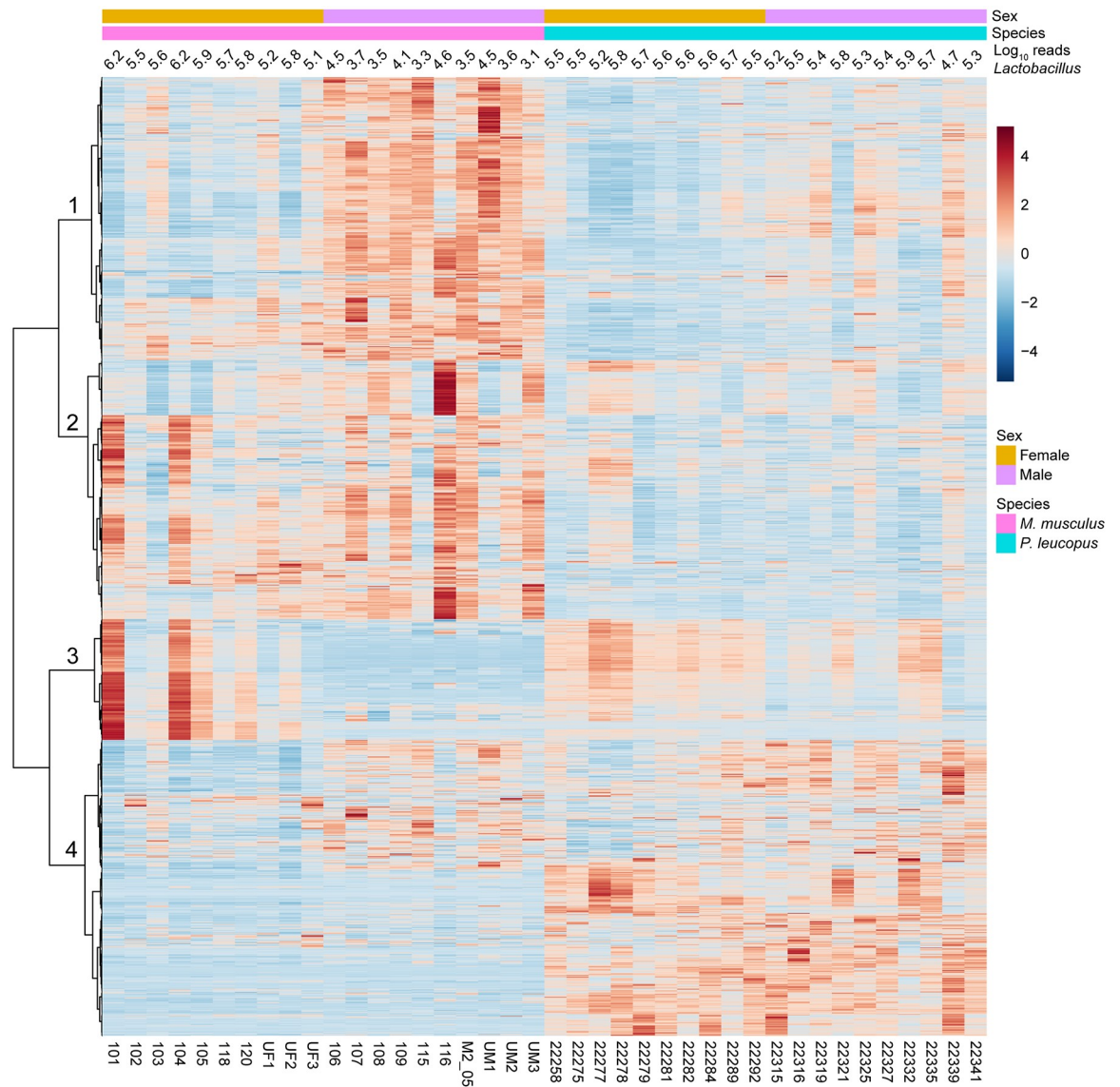


Fig 8. Heat map-formatted shading matrix of KEGG orthology gene level annotations of gut metagenomes of *P. leucopus* and *M. musculus*. The annotations were generated by MicrobiomeAnalyst (<https://www.microbiomeanalyst.ca>). Columns are grouped by species and by sex within each species. Individual animal identifications are given on the x-axis below the heat map. Above the heat map are the log-transformed reads mapping to the genus *Lactobacillus* for each animal's fecal sample. Clustering of rows of genes were by Pearson correlation coefficient. Four major clusters are labeled 1–4 on the y-axis. Scaling is by relative abundances from low (blue) to high (red).

<https://doi.org/10.1371/journal.pone.0231801.g008>

Lactobacilli of the stomach of *P. leucopus*

The differences between *P. leucopus* and *M. musculus* in the amount and species richness of the lactobacilli in their GI microbiota prompted further investigation of *P. leucopus* using histologic, microbiologic and genomic approaches. Fig 11 shows the gross morphology and histology of the stomach of representative LL stock *P. leucopus* animals [69]. The difference between forestomach with its stratified squamous epithelium and the discrete region lined with glandular mucosa are indicated in the dissecting scope and higher magnification light microscope views.

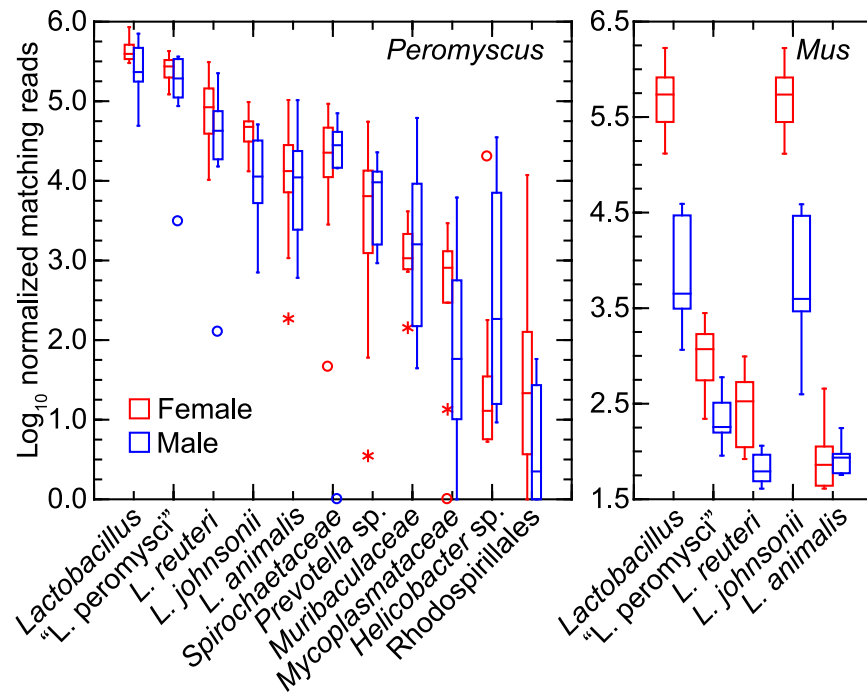


Fig 9. Box-whisker plots of log-transformed normalized reads of gut metagenomes of *P. leucopus* (left panel) and *M. musculus* (right panel) that mapped to chromosomes of *Lactobacillus* spp. or other bacteria by host species and grouped by sex. The references to which reads were mapped were complete chromosomes or partial chromosomes of organisms listed in Table 1. “*Lactobacillus*” in the first position of each panel were the cumulative reads for the four individual *Lactobacillus* species in this analysis.

<https://doi.org/10.1371/journal.pone.0231801.g009>

Staining of the sections of the fixed gastric tissue with Gram stain or Giemsa stain show a thick layer of gram-positive bacteria on the non-secretory epithelium portion of the stomach. This is similar to Savage et al. noted in the forestomachs of *M. musculus* [70]. The appearance is also consistent with the *Lactobacillus* biofilm that was described by Wesley et al. [71].

Two of the species, “*L. peromysci*” and *L. animalis*, could reliably be distinguished by their distinctive colony morphologies from the isolated strains of *L. reuteri* and *L. johnsonii*, which had colonies of similar appearance (Fig 12). The rough-surfaced, ropy colonies of “*L. peromysci*” and the compact smooth colonies of *L. animalis* were similar to what Dubos and colleagues described in their study of lactobacilli of the mouse stomach and gut [72].

We next used a different set of 20 animals of the LL stock, 6 (2 females and 4 males) of which were born at the PGSC facility and 14 (7 females and 7 males) of which were born at U. C. Irvine. All animals were housed at U.C. Irvine for at least 26 weeks before euthanasia, dissection, and cultivation of the stomach tissue and contents.

Mean (95% CI) colony forming units of lactobacilli per gram of stomach tissue on selective medium plates were ten-fold higher in females at $7.4 (1.1-47.4) \times 10^9$ than in males at $0.76 (0.40-1.4) \times 10^9$ (t -test $p = 0.02$ for log-transformed values) (Table 3). There was no discernible association with place of birth, and there was no difference between females and males in the proportions of the lactobacilli were identified as “*L. peromysci*”, *L. animalis*, and *L. reuteri*/*L. johnsonii*. For five animals, whose lactobacilli were subjected to 16S ribosomal RNA gene PCR and sequencing for confirmation, the *L. reuteri*/*L. johnsonii*-type colonies were predominantly *L. reuteri*. But *L. johnsonii* was confirmed to be present as well and outnumbered *L. reuteri* in one animal. The results for 3 animals that had been on a 9% fat content diet, which was part of

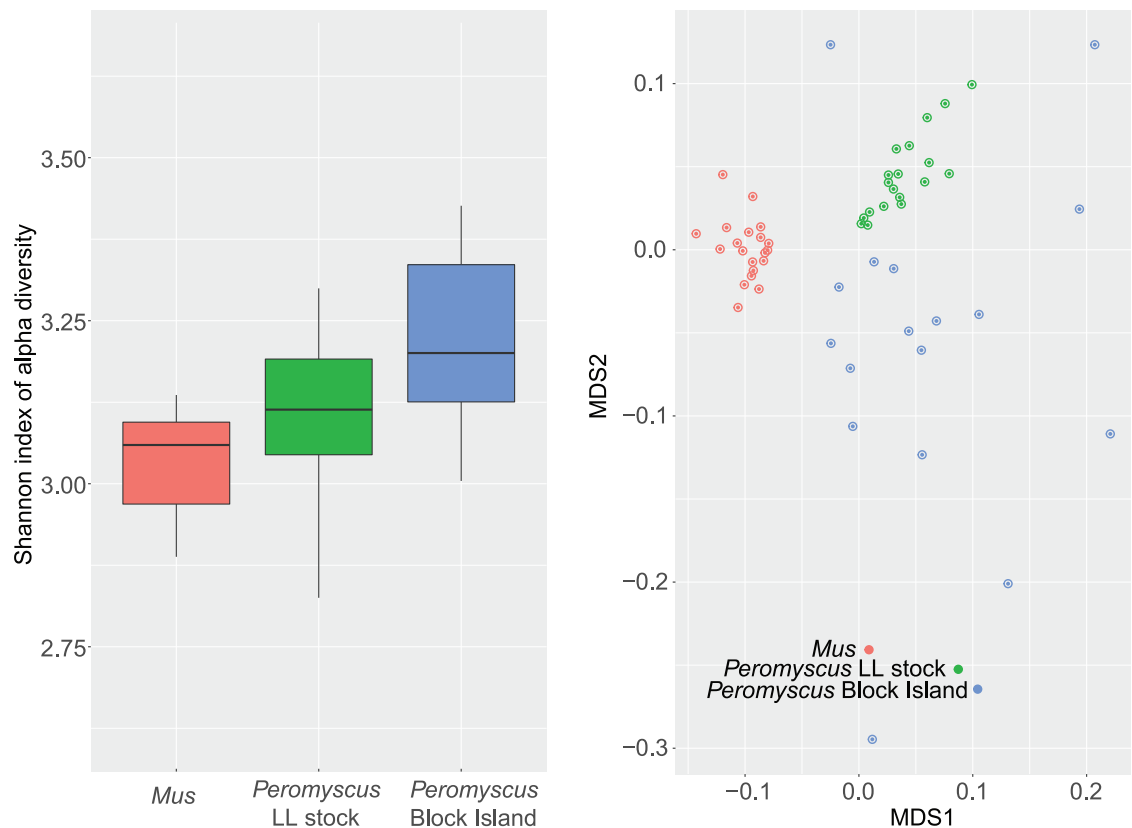


Fig 10. Alpha diversity (left) and beta diversity (right) of gut metagenomes of outbred *Peromyscus leucopus* (green), a natural population of *P. leucopus* (blue), and inbred *Mus musculus* (red). Left panel, box-whisker plots of Shannon index of alpha diversity for 20 BALB/c *M. musculus*, 20 LL stock colony *P. leucopus*, and 18 *P. leucopus* trapped on Block Island, RI. The 3 pairwise, 2-tailed *t*-test *p* values between the groups were ≤ 0.02 . Right panel, beta diversity by Bray-Curtis measure visualized by multi-dimensional scaling (MDS). A fuller description of the samples from the natural population from Block Island is provided below. The greater dispersion of values among these animals in panel B corresponded to the greater alpha diversity of this group (panel A).

<https://doi.org/10.1371/journal.pone.0231801.g010>

the breeding program, instead of 6% fat content were not distinguishable from those for the other 17.

A separate group of 9 adult LL stock *P. leucopus* (5 females and 4 males) were euthanized after withholding food overnight, and the freshly-excised stomachs were subjected to DNA extraction without prior washing of the stomach. A mean (95% CI) of 477,688 (408,988–546,388) PE250 Illumina reads were obtained for the 9 samples (S7 Table). These were mapped to the four *Lactobacillus* genomes as references, as described above, as well as to partial chromosomes for *Prevotella* sp. LL70 and *Helicobacter* sp. LL4 (Table 1). For an estimate of the number of mammalian nuclei represented in the stomach extract the *P. leucopus* genome (accession NMRJ00000000.1) served as the reference. Fig 13 shows the distributions of normalized reads mapping to the references as well as to the *P. leucopus* genome and cumulatively to all *Lactobacillus* spp. Females and males in this sample were indistinguishable by group by these measures. For this group of animals and this analysis, we confirmed the high prevalence of “*L. peromysci*” in the stomach as well as the comparatively greater representation of *L. reuteri* over *L. johnsonii*. In this sample *L. animalis* was more variable in numbers between animals. As further evidence that the *Helicobacter* sp. was of the enterohepatic type, it was near undetectable in the stomach extract, while a typically abundant genus in the intestine,

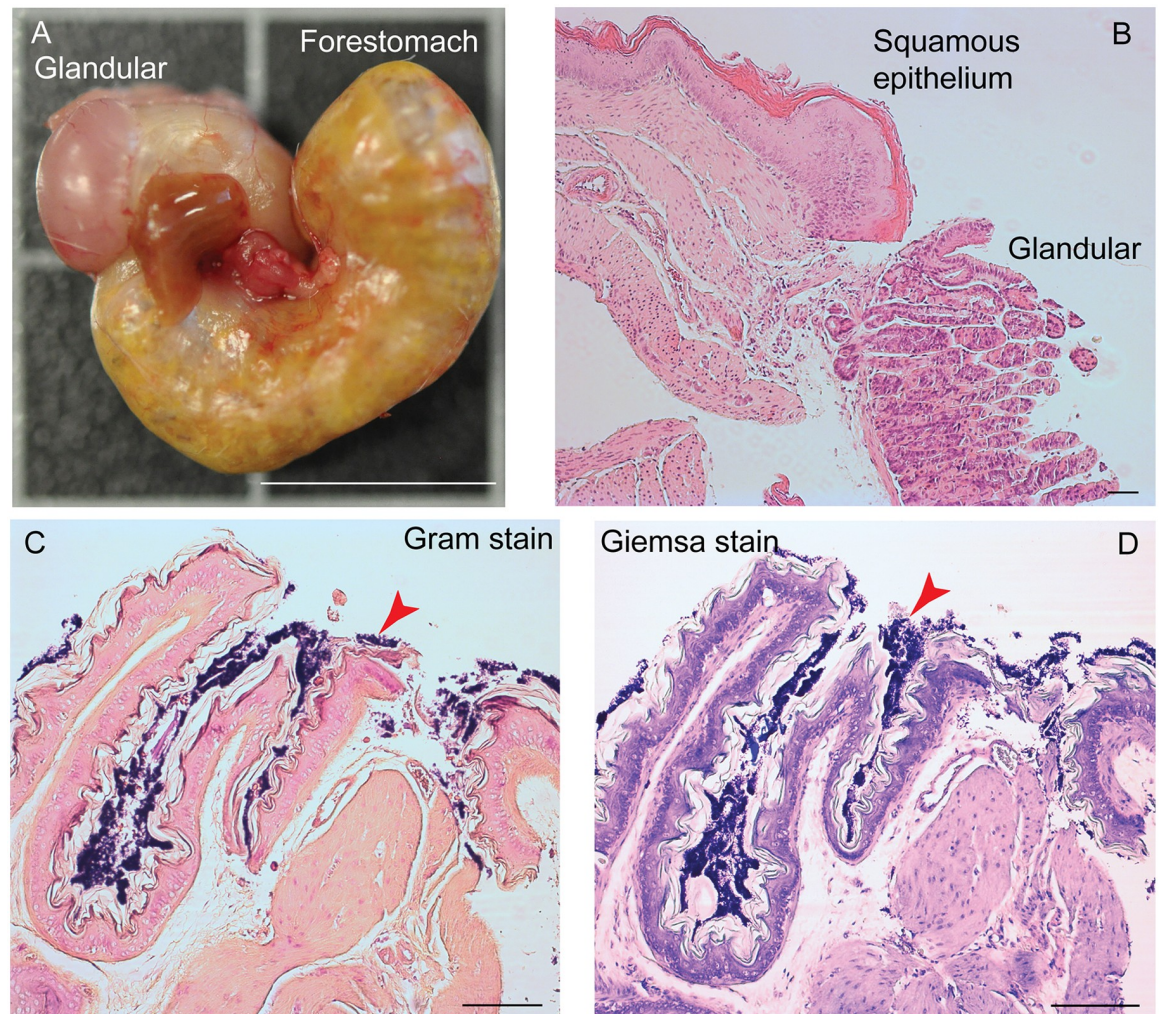


Fig 11. Gross morphology and histology of the stomach of *Peromyscus leucopus* LL stock. The glandular mucosa portions of the stomach and the forestomach with stratified squamous epithelium are indicated. Panel A, whole stomach after dissection. Portions of the esophagus and small intestine are juxtaposed in the center in this view. Bar, 1 cm. Panel B, histology of hematoxylin and eosin-stained section of junction of glandular and squamous epithelium parts. Bar, 100 μ m. Panels C and D, Gram stain (C) and Wright-Giemsa stain (D) of sections of squamous epithelium. Bar, 100 μ m. Red arrowheads indicate gram-positive bacteria in a biofilm.

<https://doi.org/10.1371/journal.pone.0231801.g011>

Prevotella, was present in small numbers in some samples. The lactobacilli in the stomach were about as numerous as the stomach tissue cells constituting the sample.

A strain of *L. reuteri* was shown to be the source of biofilm in the GI tract of mice in one study [73], and *L. murinus*, the sister taxa of *L. animalis* (Fig 3), accounted for the biofilm in another study of the upper GI tract of *M. musculus* [74]. *L. johnsonii* has also been demonstrated to produce an exopolysaccharide biofilm [75]. In a study of germ-free mice in which bacteria were experimentally introduced, *L. taiwanensis*, which is in the same cluster as *L. johnsonii* and *L. gasseri* [67], formed a mixed-species biofilm with *L. reuteri* [76].

One limitation of this experiment is that we may have overlooked species that were not identified because they were not cultivable by our method and conditions, which were micro-aerophilic, not strictly anaerobic. That said, if cells of such non-cultivable lactobacilli had been

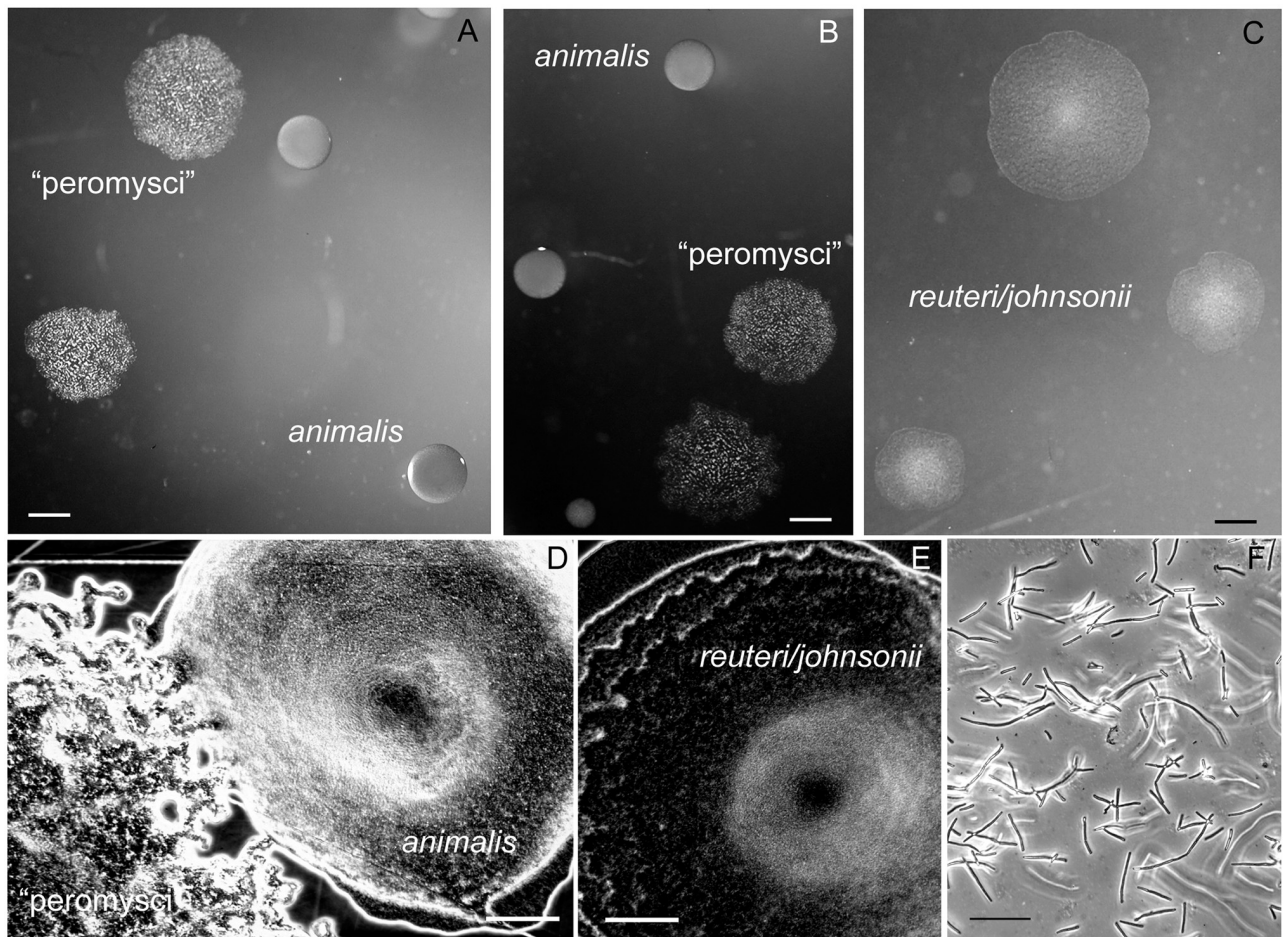


Fig 12. Colonies and cells of lactobacilli of the *P. leucopus* stomach and gut. Panels A-C show representative sizes and morphologies of colonies of "L. peromysci", *L. animalis*, and the less distinguishable *L. reuteri* and *L. johnsonii*. Bars, 1 mm. Panels D and E show magnified view of colonies of "L. peromysci" and *L. animalis* (D) and that of *L. reuteri* and *L. johnsonii* (E). Bar, 100 μ m. Panel F, phase microscopy of wet mount of unconcentrated broth culture of "L. peromysci". Bar, 10 μ m.

<https://doi.org/10.1371/journal.pone.0231801.g012>

present in the feces or stomach, their numbers did not reach a threshold for assembly into contigs of the de novo assembly of the high coverage sequencing and then detection.

Gut metagenomes of a natural population

The foregoing studies were of animals born and reared under controlled conditions, including the same diet and environmental parameters for all individuals in the group. Infectious diseases and predators were not a variable. The LL stock *P. leucopus* were outbred but the effective population size was small compared to a wild population [18]. Which of our findings would hold for animals sampled in their native habitats?

This particular study of a natural population had two specific purposes. The first was to assess the species richness or alpha diversity of microbiota within a given animal and differences in species composition or beta diversity between animals. The second objective's question of whether any of the *Lactobacillus* species we identified in the stock colony were present in natural populations. For this survey we used fecal pellets from *P. leucopus* that were

Table 3. Colony forming units (cfu) of *Lactobacillus* spp. in *Peromyscus leucopus* stomach.

Animal ID	Sex	Log ₁₀ total cfu/gm	% colony type		
			<i>animalis</i>	"peromysci"	<i>reuteri/johnsonii</i>
22403	F	10.9	3	32	65
22404 ^a	F	11.3	61	4	35
25053	F	10.7	37	60	3
25054	F	11.0	96	4	<1 ^b
25055	F	10.4	94	6	<1
25065	F	8.9	50	19	31
25062	F	8.5	60	14	26 (26/0) ^c
25063	F	8.8	52	12	36 (33/3)
25058	F	8.3	82	14	5 (5/0)
22401 ^a	M	8.6	33	19	48
22375	M	8.8	58	3	40
22420 ^a	M	9.7	48	19	33
22377	M	9.0	81	10	9
26050	M	8.4	87	6	7
25056	M	9.6	60	15	26
25010	M	9.0	84	16	<1
25060	M	8.8	55	23	23
25061	M	8.3	63	38	<1
25059	M	8.5	50	9	41 (36/5)
25011	M	9.0	61	24	14 (4/10)
Mean (95% CI)	n.a.	9.3 (8.9–9.8)	61 (51–70)	17 (11–23)	≤ 22

^a 9% fat content diet instead of 6% fat content diet.

^b <1, below limit of detection by serial dilution on plates

^c (*I*), *L. reuteri* to *L. johnsonii* ratio by PCR and 16S ribosomal rDNA sequencing

<https://doi.org/10.1371/journal.pone.0231801.t003>

individually captured and then released on Block Island, several miles off-shore from the North American mainland. At time of capture the animals were identified as to species, sex, and stage of maturity.

We analyzed the data from fecal pellets of 18 different animals (10 females and 8 males), the majority of which were adults, collected from *P. leucopus* captured at different locations on Block Island (S14 Table). As expected, there was greater variation between individual animals than was observed with the stock colony animals maintained under same conditions. Fig 10 compares the alpha diversity by Shannon index and beta diversity by Bray-Curtis dissimilarity of the inbred BALB/c *M. musculus*, outbred LL stock *P. leucopus*, and the natural population of *P. leucopus* of Block Island.

By algorithmic assignment of reads to taxonomic family, *Lactobacillaceae* was one of the most prevalent bacteria with a mean of ~5% of reads, but this was over a range of 0.3% to 20%. As was the case for the stock colony *P. leucopus*, the frequency of *Helicobacteraceae* varied more widely between sampled animals than for comparably-prevalent taxa: a mean of ~1% but over a range of 0.03% to 12%. The frequency of a parabasalid protist in the metagenomes, by the criterion of reads matching to Trichomonadidae, was similar to what we observed in the metagenome of the stock colony *P. leucopus*: the mean was 0.11% with a range of 0.02 to 0.62%. This was an indication that the *T. muris* was autochthonous in *P. leucopus*, but we did not have direct observation of the protozoa to confirm that.

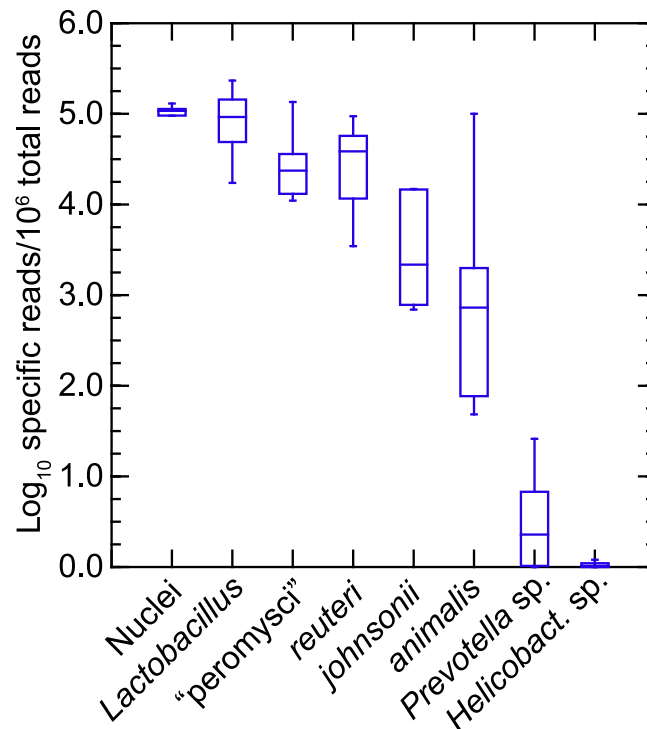


Fig 13. Box-whisker plots of log-transformed normalized reads of total metagenomes of the stomachs of 9 *P. leucopus* that mapped to chromosomes of *Lactobacillus* spp. or other bacteria by host species and by sex. The references to which reads were mapped were complete chromosomes or partial chromosomes of organisms listed in Table 1. "*Lactobacillus*" in the first position of each panel were the cumulative reads for the four individual *Lactobacillus* species in this analysis. The number of cell "nuclei" present in the stomach tissue and DNA extract were estimated by mapping reads to the whole genome of *P. leucopus* as described in the text.

<https://doi.org/10.1371/journal.pone.0231801.g013>

In descending order the two most prevalent DNA viruses in the samples, with the exception of one animal, were bacteriophages of the families *Siphoviridae* and *Myoviridae* (data available under MG-RAST accession numbers of S14 Table). These were the two virus families that were also most prevalent in the feces of the stock colony animals. In the sample from the exceptional animal (number 7464 of S14 Table), the two most prevalent virus types were the genus *Dependovirus*, a group of DNA viruses dependent on another virus in the cell for replication, and the genus *Mastadenovirus*, an adenovirus that can provide that function. Other than the adenovirus in the one sample, there were no discernible representatives of herpesviruses, poxviruses, or parvoviruses, groups of DNA viruses which include pathogens for mammals. The methods used would not have identified RNA viruses other than retroviruses.

Using the chromosome sequences of the four *Lactobacillus* species and partial chromosome sequences of *Spirochaetaceae* bacterium LL50, *Prevotella* sp. LL70, and *Helicobacter* sp. LL4 as references, we mapped and counted reads, as described for the LL stock and *M. musculus* study above. Fig 14 summarizes results for the 18 animals grouped by sex. Lactobacilli were common but, as seen with family level matching, there was greater variation between samples of the different animals than was observed for colony animals. There was also substantial variation in prevalences of the *Spirochaetaceae* bacterium and the *Prevotella* species. In most of the

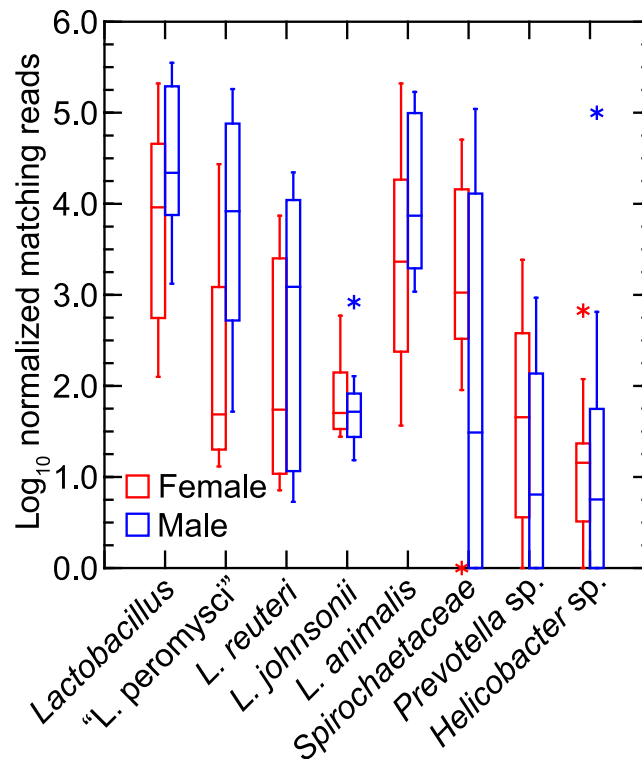


Fig 14. Box-whisker plots of log-transformed normalized reads of fecal metagenomes of 5 female (red) and 4 male (blue) *P. leucopus* of a natural population of Block Island, Rhode Island. The reference genomes and other sequences were those described for Fig 13 and in addition the partial chromosome sequence of *Spirochaetaceae* sp. LL50. As an estimate of the number of mammalian cells in the extract, "nuclei" corresponded with normalized reads mapped to *P. leucopus* genome. Points at a greater distance from the median than 1.5 times the interquartile range are plotted as asterisks.

<https://doi.org/10.1371/journal.pone.0231801.g014>

samples there was scant evidence of the *Helicobacter* species but in two animals, there were higher numbers of this organism, reaching 7% of the total reads in one fecal sample.

Among the lactobacilli used as references for this analysis, the two most prevalent species were "L. peromysci" and *L. animalis*. *L. reuteri* overall was about 10-fold lower in frequency, and *L. johnsonii* was about a hundred-fold lower in frequency. It is likely that reads called as *L. johnsonii* were the result of complete or partial matching to chromosomal loci that were highly conserved across the genus. Unlike the stock colony *P. leucopus* and the *M. musculus*, in this sampling of the Block Island population the samples from female animals had marginally lower representation of lactobacilli in the fecal samples than males.

To better characterize the two predominant *Lactobacillus* species in this set, we assembled contigs of reads mapping to "L. peromysci" or *L. animalis* from a higher coverage sequencing of the DNA of one of the Block Island samples. This yielded 51 of the 53 genes for ribosomal protein genes for a strain of *L. animalis*, which was designated 7442BI, and several core or housekeeping genes for the "L. peromysci"-like organism, which was designated BI7442 (Table 1). Fig 15 shows phylograms of DNA sequences for these and related *Lactobacillus* species or strains. The concatenated sequence of the BI7442 isolate was 99.2% identical over the 11,252 nt aligned with the corresponding sequences of the LL6 isolate of "L. peromysci".

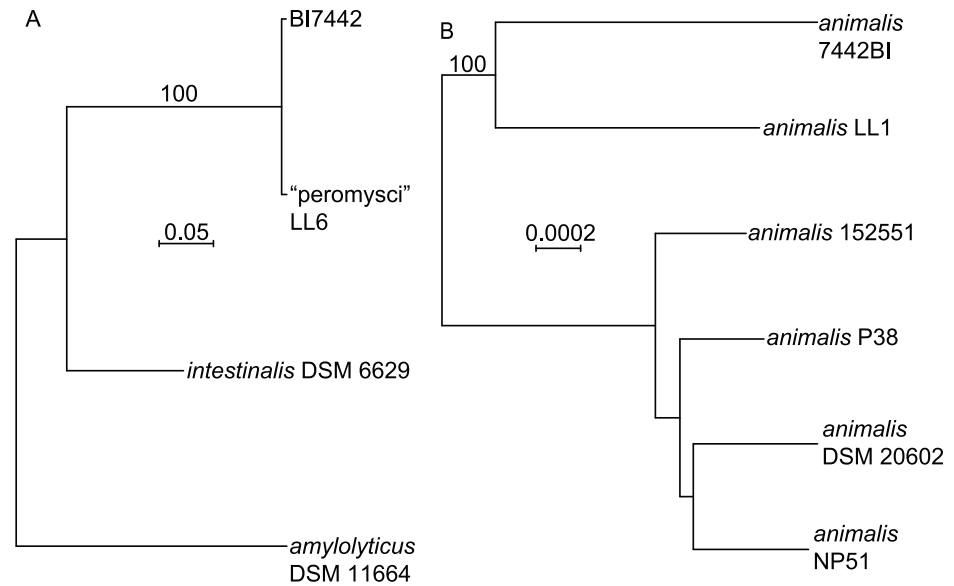


Fig 15. Distance phylograms of concatenated codon-aligned nucleotide sequences of two *Lactobacillus* spp. of *P. leucopus* of Block Island, Rhode Island. Panel A, 10,152 positions of *ftsK*, *ftsZ*, *dnaA*, *dnaN*, *ileS*, and *topA* of “*L. peromysci*” LL6 and BI7442 and two other *Lactobacillus* species. Panel B, 18,552 positions of 51 of 53 ribosomal protein genes of six *L. animalis* strains, including LL1 and 7442BI. The distance criteria were Jukes-Cantor. The scales for distance are shown in each panel. Percent bootstrap (100 iterations) support values of $\geq 80\%$ at a node are shown.

<https://doi.org/10.1371/journal.pone.0231801.g015>

Isolate 7442BI was comparatively more distant from the LL1 isolate of *L. animalis* in the stock colony but still clustered with it rather than with other examples of *L. animalis*.

The sample size was limited, and we did not attempt culture isolations from the pellets. But the source of samples from *P. leucopus* was notable for its location on an island where *B. burgdorferi* is enzootic [77] and the risk of infection for residents and visitors is high [78]. If there were to be future interventions targeting *P. leucopus* to interrupt disease transmission, Block Island would likely be a candidate site for this application.

This survey also documented that a strain or strains of “*L. peromysci*” and *L. animalis* are present in the native deermice. The high degree of sequence identity between two “*L. peromysci*” examples, whose origins were North Carolina and a New England island, long separated from the mainland, suggest that this newly-discovered species is autochthonous and plausibly a narrowly host-restricted symbiont of *P. leucopus*. Host-range restrictions of lactobacilli for the stomachs of mice were demonstrated by Wesley and Tannock [71]. Supporting an assignment of a symbiont lifestyle was its smaller genome size and lower % GC content of this species in comparison with *L. reuteri* and *L. johnsonii* with their broader host ranges [79].

Conclusions

Six decades ago René Dubos (of the epigraph), Russell Schaedler, and their colleagues at what is now Rockefeller University reported in a series of ground-breaking papers on the “fecal flora” of mice and variations in that microflora between mouse strains [80–82]. They associated differences in gastrointestinal flora with growth rates of the mice and the mouse’s susceptibility to infection and endotoxin. A featured group of bacteria in their studies were lactobacilli. They showed that this group of bacteria were present in large numbers in the feces

and that they predominated (up to 10^9 cfu per g of homogenate) in the stomachs of the mice [70], similarly to what we observed in *P. leucopus*. As their studies first intimated, the rodent may plausibly owe as much to the genomes of their microbiota as to the nuclei and mitochondria of their somatic cells for either ameliorating or exasperating disease [83].

There are also implications of our findings for development of oral vaccines that target *P. leucopus* to block transmission of pathogens either from tick to the reservoir or from the reservoir to the tick. Two of the candidate vehicles for the bait delivery of recombinant vaccine antigens to rodents have been *E. coli* and a *Lactobacillus* species [30, 84]. In neither case were the strains known to be adapted for life in *P. leucopus*. The success rate for achieving a protective response may be enhanced by use as the bacterial vehicle microbes that are adapted to *P. leucopus*. Such organisms presumably would more likely than an allochthonous bacterium to stably colonize and then proliferate to numbers large enough for the recombinant protein to elicit the sought-after immune responses.

Finally, this exploration and curation of microbes in the gut of the white-footed deermouse concludes the third leg of our project on the total genome of representative animals of the species: the nuclear genome [18], the mitochondrial genome [17], and now the GI microbiome [85]. This provides a foundation for testing of hypotheses by selective manipulation of the microbiota, for instance, by specifically targeting a certain species with a lytic phage or bacteriocin, to which it is not immune, and then evaluating the phenotype of the animal after this “knock-out”. Now that there is an annotated *P. leucopus* genome with millions of SNPs identified (UC Santa Cruz genome browser; <http://googl/LwHDr5>) it also feasible to investigate through forward genetics the host determinants of particular bacterial associations and for which there is evidence of variation within a population. An example would be the *Helicobacter* species that was highly variable in prevalence in both the wild animals and the stock colony animals.

Materials and methods

Colony animals

This study was carried out in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. At the University of California Irvine protocol AUP-18-020 was approved by the Institutional Animal Care and Use Committee (IACUC)-approved protocol. Adult outbred *P. leucopus* of the LL stock were purchased from the Peromyscus Genetic Stock Center (PGSC) of the University of South Carolina [86]. The closed colony of the LL stock was founded with 38 animals captured near Linville, NC in the mid-1980's. Some of the LL stock animals in the study were bred at the University of California Irvine's animal care facility from pairs originating at the PGSC. Adult BALB/cAnNCrI (BALB/c) *M. musculus* were purchased from Charles River. For the species comparison experiment both the PGSC-bred *P. leucopus* and *M. musculus* animals were housed in Techniplast individual ventilated cages in vivarium rooms with a 12 hour-12 hours light-dark cycle, an ambient temperature of 22 ± 1 °C, stable humidity, and on an ad libitum diet of 2020X Teklad global soy protein-free extruded rodent chow with 6% fat content (Envigo, Placentia, CA). Other animals of PGSC origin, including for the high-coverage gut metagenome study, were also housed under the same conditions and on the same diet. Twenty U.C. Irvine-bred animals were under the housing conditions and on the diet except for three (1 female and 2 males) that were on 2019 Teklad global protein extruded rodent chow with 9% fat content. Fecal samples were 2–3 pellets (100–150 mg) collected from an individual animal in temporary (< 1 h) isolation in a clean cage. The pellets were collected with sterile forceps and pooled in a single sterile 1.5 ml microfuge tube on ice. The pellets for DNA extractions

were stored frozen at -80°C . Pellets intended for cultivation, as described below, were processed as fresh specimens. Before euthanasia with carbon dioxide asphyxiation and cervical dislocation and then dissection of the stomach, food but not water was withheld for 12 h for selected animals. *P. leucopus* studied at the PGSC were under IACUC-approved protocol 2349-101211-041917 of the University of South Carolina and were euthanized by isoflurane inhalation.

Field site and animal trapping

The study was performed under IACUC-approved protocol AC-AAAS6470 of Columbia University [77]. Block Island, located 23 km from mainland Rhode Island, is part of the Outer Lands archipelagic region, which extends from Cape Cod, MA through to Staten Island, NY. Block Island is 25.2 sq. km, about 40% of which is maintained under natural conditions. The agent of Lyme disease *Borrelia burgdorferi* is enzootic on the island [87]. Animals were trapped at three locations: 1, a nature conservation area (41.15694, -71.58972); 2, private land with woodlots and fields (41.16333, -71.56611); and 3, Block Island National Wildlife Refuge (41.21055, -71.57222). Trapping was carried out during the May–August period with Sherman live traps (H.B. Sherman Traps, Inc. Tallahassee, FL) that were baited with peanut butter, oats, and sunflower seeds. Traps were arranged in nine 200 m transects with one trap placed every 10 m for a total of 180 traps at each location. Animals were removed from traps, weighed, sexed, and assessed as to age (adult, subadult, or juvenile) by pelage. Fecal pellets were collected and kept at -20°C on site, during shipment and until DNA extraction. The species identification of the source of the fecal pellets as *P. leucopus* was confirmed by sequencing of the D-loop of the mitochondrion as described [17].

Cultivation and enumeration of bacteria

Lactobacilli were initially isolated and then propagated on Rogosa SL agar plates (Sigma-Aldrich) in candle jars at 37°C . Gram-negative bacteria and specifically *Escherichia coli* were isolated and propagated on MacConkey Agar plates (Remel) incubated in ambient air at 37°C . For determinations of colony forming units (cfu) homogenates of stomach, cecum, or fecal pellets were suspended and the serially diluted in phosphate-buffer saline, pH 7.4, before plating in 100 μl volumes on solid media in 150 mm x 15 mm polystyrene Petri dishes. Colonies were counted manually. Liquid cultures of *Lactobacillus* spp. isolates or *E. coli* were in Difco Lactobacilli MRS Broth (Becton-Dickinson) or LB broth, respectively, and incubated at 37°C on a shaker. Bacteria were harvested by centrifugation at $8000 \times g$ for 10 min. Antibiotic susceptibilities were determined by standard disk testing on Mueller-Hinton Agar (Sigma-Aldrich) plates and ciprofloxacin 5 μg , gentamicin 10 μg , ampicillin 10 μg , and sulfamethoxazole 23.75 – trimethoprim 1.25 μg BBL Sensi-Disc antibiotic disks (Becton-Dickinson) according to the manufacturer instructions.

Histology

After the stomachs were removed from two euthanized *P. leucopus* LL stock adult females, they were opened longitudinally, gently flushed with PBS, and fixed in 10% buffered formalin (Thermo Fisher Scientific). Histological and histochemical analysis was performed on paraffin block sections of the stomach with Hematoxylin and Eosin, Wright-Giemsa and Gram stains (Abcam, Cambridge, UK).

Microscopy, photography, and video

Photographs of colonies on plates were taken with a Nikon Df DSLR camera and 60 mm Nikon AF-S Micro lens with illumination by incident light above and reflected light below the plates on an Olympus SZ40 dissecting scope. An Olympus BX60 microscope with attached QIClick CCD camera and Q-Capture Pro7 software (Teledyne Photometrics, Tucson, AZ) was used for low-magnification images of colonies under bright light microscopy and 400X images under phase and differential interference microscopy. Histology slides were examined on a Leica DM 2500 microscope equipped with a MC120 HD digital camera (Leica Microsystems, Buffalo Grove, IL).

DNA extractions

DNA from fresh and frozen fecal pellets, from tissue of stomach or cecum, and from bacteria harvested from broth cultures were extracted with ZymoBIOMICS DNA Miniprep or Microprep kits (Zymo Research). Freshly-dissected, unwashed gastric or cecal tissues were first cut into small pieces with sterile instruments before homogenization in the lysis buffer in an Omni Bead Ruptor 4 bead beater (Omni International). DNA concentrations were determined by both NanoDrop spectrophotometer and Qubit fluorometer (Thermo Fisher Scientific).

PCR

The near-complete 16S ribosomal RNA gene for *Lactobacillus* spp. was amplified using PCR using custom primers specific for the genus *Lactobacillus*: forward 5' -CCTAATACATG-CAAGTCG and reverse 5' -GGTTCCTCCTACGGCTA. The Platinum Taq polymerase and master mix (ThermoFisher Scientific) contained uracil-DNA glycosylase. On a T100 thermal cycler (BioRad) PCR conditions (°C for temperature) The conditions were 37° for 10 min, 94° for min, 40 cycles of 94° for 10 s, 55° for 30 s, and 72° for 45 s. The 1.5 kb PCR product was isolated and purified from agarose gel using the NucleoSpin Gel and PCR Clean-up kit (Takara). The product was subjected to Sanger dideoxy sequencing at GENEWIZ (San Diego, CA).

Whole genome sequencing, assembly, and annotation

Long reads were obtained using an Oxford Nanopore Technology MinION Mk1B instrument with Ligation Sequencing Kit, R9.4.1 flow cell, MinKnow v. 19.6.8 for primary data acquisition, and Guppy v. 3.2.4 for base calling with default settings. Paired-end short reads were obtained on a MiSeq sequencer with paired-end v2 Micro chemistry and 150 cycles (Illumina, San Diego, CA). The library was constructed using the NEXTflex Rapid DNA-Seq kit (Bioo Scientific, Austin, TX), the quality of sequencing reads was analyzed using FastQC (Babraham Bioinformatics), and reads were trimmed of Phred scores <15 and corrected for poor-quality bases using Trimmomatic [88]. A hybrid assembly was carried out with Unicycler v.0.4.7 [89] with default settings and 16 threads on the High Performance Computing cluster of the University of California Irvine. Assembly of short reads alone were performed with the Assembly Cell program of CLC Genomics Workbench v. 11 (Qiagen). Annotation was provided by the NCBI Prokaryotic Genome Annotation Pipeline [90]. Putative bacteriocins and their associated transport and immunity functions were identified by BAGEL4 [91, 92]. For other analyses paired-end reads were mapped with a length fraction of 0.7 and similarity fraction of 0.9 to whole genomes sequences or concatenated large contigs representing partial genomes (Table 1). Mapped reads were normalized for length of reference sequence and for total reads after quality control and removal of adapters.

Metagenome sequencing

The library was constructed using NEXTflex Rapid DNA-Seq kit v2 (Bioo Scientific) and the NEXTflex Illumina DNA barcodes after shearing the DNA with a Covaris S220 instrument, end repair and adenylation, and clean-up of the reaction mixture with NEXTFLEX Clean Up magnetic beads (Beckman Coulter, Brea, CA). The library was quantified by qPCR with the Kapa Sybr Fast Universal kit (Kapa Biosystems, Woburn, MA), and the library size was determined by analysis using the Bioanalyzer 2100 DNA High Sensitivity Chip (Agilent Technologies). Multiplexed libraries were loaded on either an Illumina HiSeq 2500 sequencer (Illumina, San Diego, CA) with paired-end chemistry for 250 cycles or a MiSeq Sequencer (Illumina, San Diego, CA) with paired-end v2 Micro chemistry and 150 cycles. The Illumina real time analysis software RTA 1.18.54 converted the images into intensities and base calls. De novo assemblies were performed with De Novo Assembly v. 1.4 of CLC Genomics Workbench v. 11 with the following settings: mismatch, insertion, and deletion costs of 3 each; length fraction of 0.3, and similarity fraction of 0.93.

16S ribosomal RNA gene analysis

The same DNA extract used for the metagenome sequencing at University of California Irvine was submitted to the reference laboratory ID Genomics, Inc. (Seattle, WA) and subjected to the company's 16S rRNA Metagenomics service (<http://idgenomics.com/our-services>), which used the 16S Metagenomics v. 1.01 program (Illumina). Of the 333,358 reads 82% were classified as to taxonomic family.

Metagenome analysis

Fastq files were uploaded to the metagenomic analysis server MG-RAST (<https://www.mg-rast.org>) [93]. Reads were joined using join paired reads function on the browser and filtered for *Mus musculus* v37 genome. Artificial replicate sequences produced by sequencing artifacts were removed by the method of Gomez-Alvarez et al. [94]. Low quality reads (Phred score <15 for no more than 5 bases) were removed using SolexQA, a modified DynamicTrim protocol [95]. When the term "alpha diversity" was reported in the results and tables for MG-RAST analyses, this stood for species richness or count of the species as calculated by MG-RAST. The output of the MG-RAST protocol was also analyzed in R using the *vegan* package (<https://cran.r-project.org>, <https://github.com/vegandevs/vegan>). By this means we also represented alpha diversity by the Shannon index, which accounts for evenness as well as richness [96]. Beta diversity as expressed as the Bray-Curtis Dissimilarity statistic [97] was calculated using the *avgdist* function with 1000 sample depth, the median as the function, and 100 iterations (<https://github.com/vegandevs/vegan/blob/master/man/avgdist.Rd>). Data was visualized using Nonmetric multidimensional scaling in two dimensions [98]. MicrobiomeAnalyst (<https://www.microbiomeanalyst.ca>) [99] was used for hierarchical clustering by distance criterion and by means of Pearson correlations. The DFAST prokaryotic genome annotation pipeline (<https://dfast.nig.ac.jp>) was used for annotation of incomplete chromosomes and large contigs [100]. For *Lactobacillus* spp. and the *Helicobacter* sp. the lactic acid bacteria database and *Helicobacter* database, respectively, options were chosen. Alignments and phylogenetic analysis were carried out with the SeaView v. 4 suite [101].

Statistical analysis

Normalized reads and other values whose distributions spanned more than one order of magnitude were log-transformed before parametric analysis by 2-tailed *t*-test. Inverse

transformation was carried out to provide nonparametric means and corresponding asymmetric 95% confidence intervals. Nonparametric analysis by rank order was by Kruskal-Wallis test. The Z-score was the number of standard deviations below or above the population mean a given raw value was. The False Discovery Rate (FDR) with corrected p value was estimated by the method of Benjamini and Hochberg [102]. The box-whisker plot graphs were made with SYSTAT v. 13.1 software (Systat Software, Inc.).

Supporting information

S1 Fig. Rarefaction curve for high-coverage sequencing of *Peromyscus leucopus* LL stock gut metagenome.

(EPS)

S2 Fig. MG-RAST analysis high-coverage sequencing of *Peromyscus leucopus* LL stock gut metagenome by phylum.

(EPS)

S3 Fig. MG-RAST analysis high-coverage sequencing of *Peromyscus leucopus* LL stock gut metagenome by subsystems.

(EPS)

S4 Fig. Maximum likelihood phylogram of 470 aligned amino acids of DNA polymerase type B, organellar and viral family protein of *Tritrichomonas* sp. LL5 (Table 1) and homologous proteins of other protozoa (*T. foetus*, *Trichomonas vaginalis*, *Entamoeba invadens*, and *Giardia* sp.), oocytes (*Aphanomyces astaci* and *Thraustotheca clavata*), and bacteria (Division WS6 bacterium and *Haliea* sp.). The GenBank accession numbers for the sequences are given next to organism name. The bootstrapped tree was generated with PhyML with the settings of the LG model, 4 rate classes, and 100 replicates.

(EPS)

S1 Table. Comparison of 16S sequence-based and metagenome-based identification of bacterial families in fecal pellets of LL stock *P. leucopus*.

(XLSX)

S2 Table. Metagenome by taxonomic family of fecal pellets of LL stock *P. leucopus*.

(XLSX)

S3 Table. Sources of coding sequences for ribosomal proteins at rMLST database of <http://mlst.org>.

(XLSX)

S4 Table. Putative bacteriocins and associated transport proteins and immunity proteins of 3 *Lactobacillus* species of *Peromyscus leucopus*.

(DOCX)

S5 Table. Accession numbers of sequences of other species.

(XLSX)

S6 Table. Statistics for gut metagenomes of *Peromyscus leucopus* and *Mus musculus*.

(XLSX)

S7 Table. *Peromyscus leucopus* and *Mus musculus* gut metagenome accession numbers.

(XLSX)

S8 Table. Replicates of libraries of *Peromyscus leucopus* DNA extracts of fecal pellets.
(XLSX)

S9 Table. Log10 of mean number of normalized reads for gut metagenomes of *Peromyscus leucopus* and *Mus musculus* by families with >3000 reads.
(XLSX)

S10 Table. Log10 of reads matched to function for metagenomes of fecal pellets of *Peromyscus leucopus* and *Mus musculus*.
(XLSX)

S11 Table. Map-to-reference normalized PE250 reads (log10) for feces and stomach sample against *Lactobacillus* spp. and selected other bacteria.
(XLSX)

S12 Table. Comparison of gut metagenomes of *Peromyscus leucopus* and *Mus musculus* by KEGG orthology gene.
(XLSX)

S13 Table. Reads of gut metagenomes of *Peromyscus leucopus* and *Mus musculus* by KEGG orthology gene for individual animals and data for analysis of [S11 Table](#).
(XLSX)

S14 Table. Block Island *Peromyscus leucopus* fecal metagenome.
(XLSX)

S1 Movie. Phase microscopy (400x magnification) of wet mount of *Tritrichomonas* sp. flagellates in the cecal contents of *Peromyscus leucopus*.
(MOV)

Acknowledgments

We thank Emma Keay and Vanessa Cook at the University of California Irvine and Asieh Naderi and Vimala Kaza at the University of South Carolina for their contributions to the study.

Author Contributions

Conceptualization: Ana Milovic, Alan G. Barbour.

Data curation: Khalil Bassam, Hanjuan Shao, Alan G. Barbour.

Formal analysis: Ana Milovic, Khalil Bassam, Hanjuan Shao, Alan G. Barbour.

Funding acquisition: Alan G. Barbour.

Investigation: Ana Milovic, Khalil Bassam, Hanjuan Shao, Ioulia Chatzistamou, Danielle M. Tufts, Maria Diuk-Wasser, Alan G. Barbour.

Methodology: Ana Milovic, Khalil Bassam, Hanjuan Shao, Ioulia Chatzistamou, Danielle M. Tufts, Maria Diuk-Wasser.

Resources: Danielle M. Tufts, Maria Diuk-Wasser.

Software: Khalil Bassam.

Supervision: Maria Diuk-Wasser, Alan G. Barbour.

Visualization: Ioulia Chatzistamou.

Writing – original draft: Alan G. Barbour.

Writing – review & editing: Ana Milovic.

References

1. Dubos R. Autobiographical Manuscript. New York: Rockefeller University; 1981. p. Folder 14.
2. Sangodeyi F. Rene Dubos and the Emerging Science of Human Microbial Ecology New York: Rockefeller University; 2012 [<https://www.issuelab.org/resources/27965/27965.pdf>]
3. Hall ER. The Mammals of North America. Second ed. New York: John Wiley and Sons; 1981.
4. Lackey JA, Huckaby DG, Ormiston BG. *Peromyscus leucopus*. Mammalian Species. 1985; 247:1–10.
5. Munshi-South J, Kharchenko K. Rapid, pervasive genetic differentiation of urban white-footed mouse (*Peromyscus leucopus*) populations in New York City. Mol Ecol. 2010; 19(19):4242–54. <https://doi.org/10.1111/j.1365-294X.2010.04816.x> PMID: 20819163
6. Crossland JP, Dewey MJ, Barlow SC, Vrana PB, Felder MR, Szalai GJ. Caring for *Peromyscus* spp. in research environments. Lab Animal. 2014; 43(5):162–6. <https://doi.org/10.1038/labam.504> PMID: 24751850
7. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol. 2017; 34:1812–9. <https://doi.org/10.1093/molbev/msx116> PMID: 28387841
8. Stepan SJ, Schenk JJ. Muroid rodent phylogenetics: 900-species tree reveals increasing diversification rates. PLoS One. 2017; 12(8):e0183070. <https://doi.org/10.1371/journal.pone.0183070> PMID: 28813483
9. Schrago CG, Voloch CM. The precision of the hominid timescale estimated by relaxed clock methods. J Evol Biol. 2013; 26(4):746–55. <https://doi.org/10.1111/jeb.12076> PMID: 23437917
10. Schug MD, Vessey SH, Korytko AI. Longevity and survival in a population of white-footed mice (*Peromyscus leucopus*). J Mammal. 1991; 72:360–6.
11. Bunikis J, Tsao J, Luke CJ, Luna MG, Fish D, Barbour AG. *Borrelia burgdorferi* infection in a natural population of *Peromyscus leucopus* mice: a longitudinal study in an area where Lyme Borreliosis is highly endemic. J Infect Dis. 2004; 189(8):1515–23. <https://doi.org/10.1086/382594> PMID: 15073690
12. Sacher G, Hart R. Longevity, aging, and comparative cellular and molecular biology of the house mouse, *Mus musculus*, and the white-footed mouse, *Peromyscus leucopus*. Birth Defects, Orig Artic Ser;(United States). 1978; 14(1).
13. Shorter KR, Owen A, Anderson V, Hall-South AC, Hayford S, Cakora P, et al. Natural genetic variation underlying differences in *Peromyscus* repetitive and social/aggressive behaviors. Behavior Genet. 2014; 44(2):126–35.
14. Veres M, Duselis AR, Graft A, Pryor W, Crossland J, Vrana PB, et al. The biology and methodology of assisted reproduction in deer mice (*Peromyscus maniculatus*). Theriogenology. 2012; 77(2):311–9. <https://doi.org/10.1016/j.theriogenology.2011.07.044> PMID: 21924468
15. Barbour AG. Infection resistance and tolerance in *Peromyscus* spp., natural reservoirs of microbes that are virulent for humans. Semin Cell Devel Biol. 2017; 61(1):115–22.
16. Bedford NL, Hoekstra HE. *Peromyscus* mice as a model for studying natural variation. eLife. 2015; 4:e06813.
17. Barbour AG, Shao H, Cook VJ, Baldwin-Brown J, Tsao JI, Long AD. Genomes, expression profiles, and diversity of mitochondria of the White-footed Deermouse *Peromyscus leucopus*, reservoir of Lyme disease and other zoonoses. Sci Rep. 2019; 9(1):17618. <https://doi.org/10.1038/s41598-019-54389-3> PMID: 31772306
18. Long A, Baldwin-Brown J, Tao Y, Cook V, Balderrama-Gutierrez G, Corbett-Detig R, et al. The genome of *Peromyscus leucopus*, natural host for Lyme disease and other emerging infections. Sci Advances. 2019; 5(7):eaaw6441.
19. Gomes-Solecki MJ, Brisson DR, Dattwyler RJ. Oral vaccine that breaks the transmission cycle of the Lyme disease spirochete can be delivered via bait. Vaccine. 2006; 24(20):4440–9. <https://doi.org/10.1016/j.vaccine.2005.08.089> PMID: 16198456
20. Najjar DA, Normandin AM, Strait EA, Esvelt KM. Driving towards ecotechnologies. Pathogens Global Health. 2017; 111(8):448–58. <https://doi.org/10.1080/20477724.2018.1452844> PMID: 29629849
21. Specter M. Annals of Science: Rewriting the Code of Life. The New Yorker. 2017; 93(1); January 2.

22. Baxter NT, Wan JJ, Schubert AM, Jenior ML, Myers P, Schloss PD. Intra- and interindividual variations mask interspecies variation in the microbiota of sympatric *Peromyscus* populations. *Appl Environ Microbiol*. 2015; 81(1):396–404. <https://doi.org/10.1128/AEM.02303-14> PMID: 25362056
23. Kohl KD, Dearing MD, Bordenstein SR. Microbial communities exhibit host species distinguishability and phyllosymbiosis along the length of the gastrointestinal tract. *Mol Ecol*. 2018; 27(8):1874–83. <https://doi.org/10.1111/mec.14460> PMID: 29230893
24. Bassam K, Milovic A, Barbour AG. Genome sequences of three *Lactobacillus* species strains of the stomach of the White-footed Deermouse (*Peromyscus leucopus*). *Microbiol Res Announc*. 2019; 8(40).
25. Herlemann DP, Geissinger O, Ikeda-Ohtsubo W, Kunin V, Sun H, Lapidus A, et al. Genomic analysis of "Elusimicrobium minutum," the first cultivated representative of the phylum "Elusimicrobia" (formerly termite group 1). *Appl Environ Microbiol*. 2009; 75(9):2841–9. <https://doi.org/10.1128/AEM.02698-08> PMID: 19270133
26. Soo RM, Skennerton CT, Sekiguchi Y, Imelfort M, Paech SJ, Dennis PG, et al. An expanded genomic representation of the phylum Cyanobacteria. *Genome Biol Evol*. 2014; 6(5):1031–45. <https://doi.org/10.1093/gbe/evu073> PMID: 24709563
27. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, et al. Population-level analysis of gut microbiome variation. *Science*. 2016; 352(6285):560–4. <https://doi.org/10.1126/science.aad3503> PMID: 27126039
28. Walter J. Ecological role of lactobacilli in the gastrointestinal tract: implications for fundamental and biomedical research. *Appl Environ Microbiol*. 2008; 74(16):4985–96. <https://doi.org/10.1128/AEM.00753-08> PMID: 18539818
29. Tannock G. The microecology of lactobacilli inhabiting the gastrointestinal tract. In: Marshall KC, editor. *Advances in Microbial Ecology*. 11. Boston, MA: Springer; 1990. p. 147–71.
30. Gomes-Solecki M, Richer L. Recombinant *E. coli* dualistic role as an antigen-adjuvant delivery vehicle for oral immunization. In: Pal U, Buyuktanir O, editors. *Borrelia burgdorferi Methods in Molecular Biology*. New York: Humana Press; 2018. p. 347–57. https://doi.org/10.1007/978-1-4939-7383-5_27 PMID: 29032558
31. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, et al. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology*. 2012; 158(Pt 4):1005–15. <https://doi.org/10.1099/mic.0.055459-0> PMID: 22282518
32. Fujisawa T, Itoh K, Benno Y, Mitsuoka T. *Lactobacillus intestinalis* (ex Hemme 1974) sp. nov., nom. rev., isolated from the intestines of mice and rats. *Int J Syst Bacteriol*. 1990; 40(3):302–4. <https://doi.org/10.1099/00207713-40-3-302> PMID: 2397198
33. Kim D, Cho MJ, Cho S, Lee Y, Byun SJ, Lee S. Complete Genome Sequence of *Lactobacillus reuteri* Byun-re-01, isolated from mouse small intestine. *Microbiol Res Announc*. 2018; 7(17).
34. Sun Z, Harris HM, McCann A, Guo C, Argimon S, Zhang W, et al. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. *Nat Commun*. 2015; 6:8322. <https://doi.org/10.1038/ncomms9322> PMID: 26415554
35. Dent V, Williams R. *Lactobacillus animalis* sp. nov., a new species of *Lactobacillus* from the alimentary canal of animals. *Zentralblatt für Bakteriologie Mikrobiologie und Hygiene: I Abt Originale C: Allgemeine, angewandte und ökologische Mikrobiologie*. 1982; 3(3):377–86.
36. Nam SH, Choi SH, Kang A, Kim DW, Kim RN, Kim A, et al. Genome sequence of *Lactobacillus animalis* KCTC 3501. *J Bacteriol*. 2011; 193(5):1280–1. <https://doi.org/10.1128/JB.01505-10> PMID: 21183665
37. Claesson MJ, Li Y, Leahy S, Canchaya C, van Pijkeren JP, Cerdeno-Tarraga AM, et al. Multireplicon genome architecture of *Lactobacillus salivarius*. *Proc Natl Acad Sci U S A*. 2006; 103(17):6718–23. <https://doi.org/10.1073/pnas.0511060103> PMID: 16617113
38. Frese SA, Benson AK, Tannock GW, Loach DM, Kim J, Zhang M, et al. The evolution of host specialization in the vertebrate gut symbiont *Lactobacillus reuteri*. *PLoS Genet*. 2011; 7(2):e1001314. <https://doi.org/10.1371/journal.pgen.1001314> PMID: 21379339
39. Bandara M, Skehel JM, Kadioglu A, Collinson I, Nobbs AH, Blocker AJ, et al. The accessory Sec system (SecY2A2) in *Streptococcus pneumoniae* is involved in export of pneumolysin toxin, adhesion and biofilm formation. *Microbes Infect*. 2017; 19(7–8):402–12. <https://doi.org/10.1016/j.micinf.2017.04.003> PMID: 28456649
40. Denou E, Pridmore RD, Berger B, Panoff J-M, Arigoni F, Brüssow H. Identification of genes associated with the long-gut-persistence phenotype of the probiotic *Lactobacillus johnsonii* strain NCC533 using a combination of genomics and transcriptome analysis. *J Bacteriol*. 2008; 190(9):3161–8. <https://doi.org/10.1128/JB.01637-07> PMID: 18223069

41. Wilson CM, Aggio RB, O'Toole PW, Villas-Boas S, Tannock GW. Transcriptional and metabolomic consequences of LuxS inactivation reveal a metabolic rather than quorum-sensing role for LuxS in *Lactobacillus reuteri* 100–23. *J Bacteriol.* 2012; 194(7):1743–6. <https://doi.org/10.1128/JB.06318-11> PMID: 22287522
42. Hidalgo-Cantabrana C, Goh YJ, Pan M, Sanozky-Dawes R, Barrangou R. Genome editing using the endogenous type I CRISPR-Cas system in *Lactobacillus crispatus*. *Proc Natl Acad Sci U S A.* 2019; 116(32):15774–83. <https://doi.org/10.1073/pnas.1905421116> PMID: 31341082
43. Saulnier DM, Santos F, Roos S, Mistretta TA, Spinler JK, Molenaar D, et al. Exploring metabolic pathway reconstruction and genome-wide expression profiling in *Lactobacillus reuteri* to define functional probiotic features. *PLoS One.* 2011; 6(4):e18783. <https://doi.org/10.1371/journal.pone.0018783> PMID: 21559529
44. Moreno-Vivian C, Cabello P, Martinez-Luque M, Blasco R, Castillo F. Prokaryotic nitrate reduction: molecular properties and functional distinction among bacterial nitrate reductases. *J Bacteriol.* 1999; 181(21):6573–84. <https://doi.org/10.1128/JB.181.21.6573-6584.1999> PMID: 10542156
45. Hwang H, Lee JH. Characterization of arginine catabolism by lactic acid bacteria isolated from kimchi. *Molecules (Basel, Switzerland).* 2018; 23(11).
46. Singer JR, Blosser EG, Zindl CL, Silberger DJ, Conlan S, Laufer VA, et al. Preventing dysbiosis of the neonatal mouse intestinal microbiome protects against late-onset sepsis. *Nat Med.* 2019; 25(11):1772–82. <https://doi.org/10.1038/s41591-019-0640-y> PMID: 31700190
47. Peña JA, Li SY, Wilson PH, Thibodeau SA, Szary AJ, Versalovic J. Genotypic and phenotypic studies of murine intestinal lactobacilli: species differences in mice with and without colitis. *Appl Environ Microbiol.* 2004; 70(1):558–68. <https://doi.org/10.1128/aem.70.1.558-568.2004> PMID: 14711688
48. Sheh A, Shen Z, Fox JG. Draft genome sequences of eight enterohepatic *Helicobacter* species isolated from both laboratory and wild rodents. *Genome Announc.* 2014; 2(6).
49. Suerbaum S, Josenhans C, Sterzenbach T, Drescher B, Brandt P, Bell M, et al. The complete genome sequence of the carcinogenic bacterium *Helicobacter hepaticus*. *Proc Natl Acad Sci U S A.* 2003; 100(13):7901–6. <https://doi.org/10.1073/pnas.1332093100> PMID: 12810954
50. Fox JG, Ge Z, Whary MT, Erdman SE, Horwitz BH. *Helicobacter hepaticus* infection in mice: models for understanding lower bowel inflammation and cancer. *Mucosal Immunol.* 2011; 4(1):22–30. <https://doi.org/10.1038/mi.2010.61> PMID: 20944559
51. Bracken TC, Cooper CA, Ali Z, Truong H, Moore JM. *Helicobacter* infection significantly alters pregnancy success in laboratory mice. *J Am Assoc Lab Animal Sci.* 2017; 56(3):322–9.
52. Nordhoff M, Taras D, Macha M, Tedin K, Busse HJ, Wieler LH. *Treponema berlinense* sp. nov. and *Treponema porcinum* sp. nov., novel spirochaetes isolated from porcine faeces. *Int J Syst Evol Microbiol.* 2005; 55(Pt 4):1675–80. <https://doi.org/10.1099/ijs.0.63388-0> PMID: 16014500
53. Barbour AG. Borreliales. In: Whitman WB, Rainey R, Kämpfe P, Trujillo M, Chun J, DeVos P, et al., editors. *Bergey's Manual of Systematics of Archaea and Bacteria.* 2018. p. 1–9.
54. Geissinger O, Herlemann DP, Morschel E, Maier UG, Brune A. The ultramicrobacterium "*Elusimicrobium minutum*" gen. nov., sp. nov., the first cultivated representative of the termite group 1 phylum. *Appl Environ Microbiol.* 2009; 75(9):2831–40. <https://doi.org/10.1128/AEM.02697-08> PMID: 19270135
55. Lagkouvardos I, Lesker TR, Hitch TCA, Galvez EJC, Smit N, Neuhaus K, et al. Sequence and cultivation study of Muribaculaceae reveals novel species, host preference, and functional potential of this yet undescribed family. *Microbiome.* 2019; 7(1):28. <https://doi.org/10.1186/s40168-019-0637-2> PMID: 30782206
56. Doran DJ. A catalogue of the protozoa and helminths of North American rodents. I. Protozoa and Acanthocephala. *Am Midland Naturalist.* 1954; 52(1):118–28.
57. Ericsson AC, Gagliardi J, Bouhan D, Spollen WG, Givan SA, Franklin CL. The influence of caging, bedding, and diet on the composition of the microbiota in different regions of the mouse gut. *Sci Rep.* 2018; 8(1):4065. <https://doi.org/10.1038/s41598-018-21986-7> PMID: 29511208
58. Steiner JM, Schwamberger S, Pantchev N, Balzer HJ, Vrhovec MG, Lesina M, et al. Use of ronidazole and limited culling to eliminate *Tritrichomonas muris* from laboratory mice. *J Am Assoc Lab Animal Sci.* 2016; 55(4):480–3.
59. Koyama T, Endo T, Asahi H, Kuroki T. Life cycle of *Tritrichomonas muris*. *Zentralblatt für Bakteriologie, Mikrobiologie und Hygiene Series A: Med Microbiol, Infect Dis, Virol, Parasitol.* 1987; 264(3–4):478–86.
60. Hackstein JH, Akhmanova A, Boxma B, Harhangi HR, Voncken FG. Hydrogenosomes: eukaryotic adaptations to anaerobic environments. *Trends Microbiol.* 1999; 7(11):441–7. [https://doi.org/10.1016/S0966-842x\(99\)01613-3](https://doi.org/10.1016/S0966-842x(99)01613-3) PMID: 10542423

61. Rae DO, Crews JE. *Trichomonas foetus*. *Vet Clin North Am Food Animal Practice*. 2006; 22(3):595–611. <https://doi.org/10.1016/j.cvfa.2006.07.001> PMID: 17071355
62. Yao C, Koster LS. *Trichomonas foetus* infection, a cause of chronic diarrhea in the domestic cat. *Vet Res*. 2015; 46:35. <https://doi.org/10.1186/s13567-015-0169-0> PMID: 25880025
63. Escalante NK, Lemire P, Cruz Tleugabulova M, Prescott D, Mortha A, Streutker CJ, et al. The common mouse protozoa *Trichomonas muris* alters mucosal T cell homeostasis and colitis susceptibility. *J Exp Med*. 2016; 213(13):2841–50. <https://doi.org/10.1084/jem.20161776> PMID: 27836928
64. Howitt MR, Lavoie S, Michaud M, Blum AM, Tran SV, Weinstock JV, et al. Tuft cells, taste-chemosensory cells, orchestrate parasite type 2 immunity in the gut. *Science*. 2016; 351(6279):1329–33. <https://doi.org/10.1126/science.aaf1648> PMID: 26847546
65. Chudnovskiy A, Mortha A, Kana V, Kennard A, Ramirez JD, Rahman A, et al. Host-rotzoan interactions protect from mucosal infections through activation of the inflammasome. *Cell*. 2016; 167(2):444–56.e14. <https://doi.org/10.1016/j.cell.2016.08.076> PMID: 27716507
66. Lange BM, Croteau R. Isopentenyl diphosphate biosynthesis via a mevalonate-independent pathway: Isopentenyl monophosphate kinase catalyzes the terminal enzymatic step. *Proc Natl Acad Sci USA*. 1999; 96(24):13714–9. <https://doi.org/10.1073/pnas.96.24.13714> PMID: 10570138
67. Hammes WP, Vogel RF. The genus *Lactobacillus*. In: Wood B, Holzappel W, editors. *The Genera of Lactic Acid Bacteria*: Springer; 1995. p. 19–54.
68. Salzman NH, de Jong H, Paterson Y, Harmsen HJM, Welling GW, Bos NA. Analysis of 16S libraries of mouse gastrointestinal microflora reveals a large new group of mouse intestinal bacteria. *Microbiol*. 2002; 148(Pt 11):3651–60.
69. Carleton MD. A survey of gross stomach morphology in New World Cricetinae (Rodentia, Muroidea), with comments on functional interpretations. *Miscellaneous Publications, Museum of Zoology, University of Michigan No 146*. June 14, 1973. pp. 1–43.
70. Savage DC, Dubos R, Schaedler RW. The gastrointestinal epithelium and its autochthonous bacterial flora. *J Exp Med*. 1968; 127(1):67–76. <https://doi.org/10.1084/jem.127.1.67> PMID: 4169441
71. Wesney E, Tannock GW. Association of rat, pig, and fowl biotypes of lactobacilli with the stomach of gnotobiotic mice. *Microbial Ecol*. 1979; 5(1):35–42.
72. Dubos R, Schaedler RW, Costello R, Hoet P. Indigenous, normal, and autochthonous flora of the gastrointestinal tract *J Exp Med*. 1965; 122:67–76. <https://doi.org/10.1084/jem.122.1.67> PMID: 14325474
73. Salas-Jara MJ, Ilabaca A, Vega M, Garcia A. Biofilm forming *Lactobacillus*: new challenges for the development of probiotics. *Microorganisms*. 2016; 4(3).
74. Almiron M, Traglia G, Rubio A, Sanjuan N. Colonization of the mouse upper gastrointestinal tract by *Lactobacillus murinus*: a histological, immunocytochemical, and ultrastructural study. *Curr Microbiol*. 2013; 67(4):395–8. <https://doi.org/10.1007/s00284-013-0367-9> PMID: 23689939
75. Dertli E, Mayer MJ, Narbad A. Impact of the exopolysaccharide layer on biofilms, adhesion and resistance to stress in *Lactobacillus johnsonii* F19785. *BMC Microbiol*. 2015; 15:8. <https://doi.org/10.1186/s12866-015-0347-2> PMID: 25648083
76. Lin XB, Wang T, Stothard P, Corander J, Wang J, Baines JF, et al. The evolution of ecological facilitation within mixed-species biofilms in the mouse gastrointestinal tract. *ISME J*. 2018; 12(11):2770–84. <https://doi.org/10.1038/s41396-018-0211-0> PMID: 30013162
77. Huang CI, Kay SC, Davis S, Tufts DM, Gaffett K, Tefft B, et al. High burdens of *Ixodes scapularis* larval ticks on white-tailed deer may limit Lyme disease risk in a low biodiversity setting. *Ticks Tick-Borne Dis*. 2019; 10(2):258–68. <https://doi.org/10.1016/j.ttbdis.2018.10.013> PMID: 30446377
78. Finch C, Al-Damulji MS, Krause PJ, Niccolai L, Steeves T, O'Keefe CF, et al. Integrated assessment of behavioral and environmental risk factors for Lyme disease infection on Block Island, Rhode Island. *PLoS One*. 2014; 9(1):e84758. <https://doi.org/10.1371/journal.pone.0084758> PMID: 24416278
79. Moran NA, McCutcheon JP, Nakabachi A. Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet*. 2008; 42:165–90. <https://doi.org/10.1146/annurev.genet.41.110306.130119> PMID: 18983256
80. Dubos R, Schaedler RW. Some biological effects of the digestive flora. *Am J Med Sci*. 1962; 244:265–71. <https://doi.org/10.1097/00000441-196209000-00001> PMID: 13888073
81. Dubos RJ, Schaedler RW. The effect of diet on the fecal bacterial flora of mice and on their resistance to infection. *J Exp Med*. 1962; 115:1161–72. <https://doi.org/10.1084/jem.115.6.1161> PMID: 13888074
82. Schaedler RW, Dubos RJ. The fecal flora of various strains of mice. Its bearing on their susceptibility to endotoxin. *J Exp Med*. 1962; 115:1149–60. <https://doi.org/10.1084/jem.115.6.1149> PMID: 14497916

83. Rosshart SP, Vassallo BG, Angeletti D, Hutchinson DS, Morgan AP, Takeda K, et al. Wild mouse gut microbiota promotes host fitness and improves disease resistance. *Cell*. 2017; 171(5):1015–28.e13. <https://doi.org/10.1016/j.cell.2017.09.016> PMID: 29056339
84. Del Rio B, Seegers JF, Gomes-Solecki M. Immune response to *Lactobacillus plantarum* expressing *Borrelia burgdorferi* OspA is modulated by the lipid modification of the antigen. *PLoS One*. 2010; 5(6): e11199. <https://doi.org/10.1371/journal.pone.0011199> PMID: 20585451
85. Carroll IM, Threadgill DW, Threadgill DS. The gastrointestinal microbiome: a malleable, third genome of mammals. *Mamm Genome*. 2009; 20(7):395–403. <https://doi.org/10.1007/s00335-009-9204-7> PMID: 19629594
86. Peromyscus Genetic Stock Center. Peromyscus Genetic Stock Center, Columbia, SC: University of South Carolina; 2017 [<http://stkctr.biol.sc.edu>].
87. Tufts DM, Diuk-Wasser MA. Transplacental transmission of tick-borne *Babesia microti* in its natural host *Peromyscus leucopus*. *Parasit Vectors*. 2018; 11(1):286. <https://doi.org/10.1186/s13071-018-2875-8> PMID: 29728129
88. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
89. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol*. 2017; 13(6):e1005595. <https://doi.org/10.1371/journal.pcbi.1005595> PMID: 28594827
90. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, et al. RefSeq: an update on prokaryotic genome annotation and curation. *Nucl Acids Res*. 2018; 46(D1):D851–d60. <https://doi.org/10.1093/nar/gkx1068> PMID: 29112715
91. de Jong A, van Heel AJ. BAGEL4 Groningen, the Netherlands: University of Groningen; 2019 [<http://bagel4.molgenrug.nl>].
92. van Heel AJ, de Jong A, Song C, Viel JH, Kok J, Kuipers OP. BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucl Acids Res*. 2018; 46(W1):W278–W81. <https://doi.org/10.1093/nar/gky383> PMID: 29788290
93. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, et al. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*. 2008; 9(1):386.
94. Gomez-Alvarez V, Teal TK, Schmidt TM. Systematic artifacts in metagenomes from complex microbial communities. *ISME J*. 2009; 3(11):1314–7. <https://doi.org/10.1038/ismej.2009.72> PMID: 19587772
95. Cox MP, Peterson DA, Biggs PJ. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics*. 2010; 11(1):485.
96. Hill MO. Diversity and evenness: a unifying notation and its consequences. *Ecology*. 1973; 54(2):427–32.
97. Bray JR, Curtis JT. An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*. 1957; 27(4):325–49.
98. Ward JH. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc*. 1963; 58(301):236–44.
99. Chong J, Liu P, Zhou G, Xia J. Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. *Nature Protocols*. 2020; 15(3):799–821. <https://doi.org/10.1038/s41596-019-0264-1> PMID: 31942082
100. Tanizawa Y, Fujisawa T, Nakamura Y. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics*. 2017; 34(6):1037–9.
101. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*. 2010; 27(2):221–4. <https://doi.org/10.1093/molbev/msp259> PMID: 19854763
102. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Series B (Methodological)*. 1995:289–300.