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

Short title: Gastrointestinal microbiota of *Peromyscus leucopus*

Ana Milovic [1], Khalil Bassam [1,2], Hanjuan Shao [1], Ioulia Chatzistamou [3], Danielle M. Tufts [4], Maria Diuk-Wasser [4], and 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

* Corresponding author

Email: abarbour@uci.edu

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 deermouse 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.

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

58 *Peromyscus* species can live twice as long as the laboratory mouse or rat (12). *P. leucopus* differs
59 in its social behavior and reproductive physiology from rodents that are traditional
60 experimental models (13, 14).

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

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

76 Accordingly, we carried out a combined microbiologic and metagenomic study of the GI
77 microbiome of *P. leucopus*. Our study focused on animals of a stock colony that has for many

78 years been the major source of animals for different laboratories and spin-off breeding
79 programs, including our own, in North America. The study extended to samples of *P. leucopus*
80 deermice in their natural environments and, for a comparative animal, vivarium-reared *Mus*
81 *musculus* under similar husbandry. While our investigations revealed similarities between the
82 microbiota of the white-footed deermouse and the house mouse, there were also notable
83 differences. These included a greater abundance and diversity of lactobacilli in *P. leucopus*. The
84 investigation of four *Lactobacillus* species, particularly in their niches in the stomach of *P.*
85 *leucopus*, was a special emphasis. A comparison of the GI microbiota of a natural population of
86 *P. leucopus* and the stock colony animals revealed several species in common, albeit with larger
87 variance among the wild animals .

88

89 Results and discussion

90 High coverage sequencing of fecal metagenome

91 Since there was only limited information in the literature on the GI microbiota of *P.*
92 *leucopus* (22, 23), we began with untargeted assessment of microbiome constituents and their
93 diversity in a sample of fecal pellets collected from two adult males and two adult females of
94 the same birth cohort and shipment.

95 DNA was extracted and used for library construction; 332,279,332 paired-end reads of
96 average length 247 nt were obtained 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://mgrast.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 200,000 reads and a species count of 9000 (Fig S1 of Supplementary information). The alpha diversity was 250 species. By phylum 94% of the matched reads were either Bacteroidetes (60%) or Firmicutes (34%) (Fig S2 of Supplementary information). 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%) (Fig S3 of Supplementary information).

A portion of the DNA was also submitted for commercial 16S rRNA metagenomics analysis of microbiota. As illustrated in Fig 1 and detailed in Tables S1 and S2 of Supplementary information, 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 Bacterioidales (*Bacterioidaceae*, *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*, *Ruminococcoceae*, Thermoanaerobacterales Family III, *Erysipelotrichaceae*, and

Veillonellaceae). Two exceptions were organisms of the families *Spirochaetaceae* of the phylum Spirochaetes and *Helicobacteraceae* of the phylum Proteobacteria.

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 criteria (*x*-axis) and by genome-wide gene (*y*-axis). The values of the two methods were normalized by *Z*-score. The different taxa are indicated in the graph capital letters, which defined in the box to 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.

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 ≥ 30X 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).

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 (Table S2 of Supplementary information). *Enterobacteriaceae* at 0.3% accounted for a relatively small proportion of matched reads.

As in humans (27), *Bacteroidaceae*, *Lachnospiraceae*, Prevotellaceae, and *Ruminococcaceae* 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 the metagenome in a European population (27) and by 16S sequencing ≤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).

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 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

155 designated LL2. The whole genome sequence of isolate LL2's chromosome and plasmids was
156 sequenced and assembled using a hybrid of long reads and short reads (Table 1) for an overall
157 coverage of 90X. The chromosome in two contigs of 3.4 Mb and 1.6 Mb totaled 5.0 Mb with a GC
158 content of 50%.

159

Table 1. Resources from this study

Organism and strain	Description	BioProject	BioSample	SRA or MG-RAST [a]	Accession No.
Lactobacillus johnsonii LL8	2,045,501 bp WGS chromosome; 40 contigs	PRJNA589091	SAMN13266521	SRX7128459	WKKC00000000
Lactobacillus johnsonii LL8	75,746 bp plasmid	PRJNA589091	SAMN13266521	SRX7128459	CM019125
Escherichia coli LL2	Chromosome; 3,345,873 and 1,610,537 bp contigs	PRJNA533838	SAMN11468944	SRR9087223 SRR9087224	VBVB01000001 VBVB01000002
Escherichia coli LL2	121,192 bp plasmid	PRJNA533838	SAMN11468944	SRR9087223 SRR9087224	CM017030
Escherichia coli LL2	90,617 bp plasmid	PRJNA533838	SAMN11468944	SRR9087223 SRR9087224	CM017032
Escherichia coli 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 *Helicobacter* sp. LL4	290,716 bp 172,100 bp 203,294 bp chromosome fragments	PRJNA540317	SAMN11533720	mgm4799371.3 mgm4799372.3	MN577567 MN577568 MN577569
Uncultured *Helicobacter* 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 *Spirochaetaceae* bacterium LL50	277,828 bp chromosome fragment	PRJNA540317	SAMN11533720	mgm4799371.3	MN577574

				mgm4799372.3	
Uncultured *Prevotella* 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 *Mycoplasmataceae* bacterium LL85	126,601 bp 100,243 bp chromosome fragments	PRJNA540317	SAMN11533720	mgm4799371.3 mgm4799372.3	MN991199 MN991200
Uncultured Muribaculaceae bacterium LL71	125,326 bp 103,384 bp chromosome fragments	PRJNA540317	SAMN11533720	mgm4799371.3 mgm4799372.3	MT002444 MT002445
Tritrichomonas sp. LL5	1,501 bp of small subunit ribosomal RNA	PRJNA540317	SAMN13920683	mgm4864879.3	MN120899
Tritrichomonas sp. LL5	989 bp partial iron hydrogenase gene of	PRJNA540317	SAMN13920683	mgm4864879.3	MN985504

	hydrogenosome				
Tritrichomonas sp. LL5	5298 bp fragment with DNA polymerase type B, organellar and viral family protein	PRJNA540317	SAMN13920683	mgm4864879.3	MT002461
Uncultured *Lactobacillus* sp. ("peromysci") BI7442	*rpsA, ftsK, ftsZ, dnaA, dnaN, recD, ileS, recA, topA*	PRJNA593618	SAMN13482862	SRX7285441	MN792760 - MN792768
Uncultured *Lactobacillus animalis* 7442BI	51 large and small ribosomal proteins	PRJNA593618	SAMN13482862	SRX7285441	MN817867 - MN817918

[a] SRA, Sequence Read Archive accession number; MG-RAST, mg-rast.org metagenomics analysis server sequence file number

[b] n.a., not applicable

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.

While the origin of this *E. coli* strain in the colony animals is obscure, it appears to be stably maintained among the gut microbiota of this population of *P. leucopus*. This adaptation 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

187 different species were identified. The genomes of three of organisms, namely *L. animalis* strain
188 LL1, *L. reuteri* strain LL7, and a new species, designated as *Lactobacillus* sp. LL6 and
189 provisionally named as "L. peromysci", have been reported (24). The fourth genome, of the LL8
190 strain of *L. johnsonii*, is described first here (Table 1). *L. johnsonii*'s chromosome from cumulative
191 contigs was 2,045,501 bp, about the same size as that of "L. peromysci" at 2,067,236 bp, but
192 shorter than the 2,280,577 bp length for *L. animalis* and 2,205,740 bp for *L. reuteri*. The % GC
193 content of "L. peromysci" at 33.5 was closer to *L. johnsonii* (34.4) than to either *L. animalis* (41.0)
194 or *L. reuteri* (38.9).

195 Fig 2 is a distance phylogram of 1385 aligned sites of 16S ribosomal RNA genes for the
196 four different lactobacilli. These were distributed across four major groups of the genus
197 *Lactobacillus*. The phylogenetic relationships were examined in more depth by multilocus
198 sequence typing of the 53 genes for ribosomal proteins. These were identified in the genomes,
199 compared with other deposited sequences in the ribosomal MLST database
200 (https://pubmlst.org) (31), concatenated, and then aligned with analogously concatenated DNA
201 sequences from related species (Table S3 of Supplementary information). Bacteria with identical
202 sequences for the 53 ribosomal protein genes were not found in the rMLST database of 133,460
203 profiles. The % GC contents of the concatenated coding sequences were 39.5, 40.8, 42.2, and 42.3
204 for *L. johnsonii*, "L. peromysci", *L reuteri*, and *L. animalis*, respectively. Fig 3 shows the distance
205 phylograms for ~20 kb of aligned positions for the four species, each grouped with other strains
206 or species within their respective phylogenetic clusters.

Fig 2. Neighbor-joining distance phylogram of 1420 aligned positions of 16S ribosomal RNA genes of the culture isolates of four *Lactobacillus* species from *Peromyscus leucopus* and selected other *Lactobacillus* spp. The sources for the accession numbers for the strains are given in Methods (*L. animalis*, *L. reuteri*, and "L. peromysci") or in Table 1. The other organisms represented are from Reference RNA sequences database of the National Center for Biotechnology Information; the accession numbers are given after the species name. The scale for distance by criterion of observed differences is indicated. Percent bootstrap (100 iterations) support values of ≥ 90% at a node are shown.

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 ≥ 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.

"Lactobacillus peromysci" was distant from other sequenced lactobacilli by rMLST (panel A), as well as by its 16S ribosomal RNA gene (Fig 2). The nearest taxon in the sequence alignment was *L. intestinalis*, which was first isolated from the intestines of *Mus musculus* and other murids (32). The unique ST for the rMLST for strain LL6 of this organism is 115326.

Draft and complete genomes of numerous *L. reuteri* strains have been sequenced, for example, strain Byun-re-01, which was isolated from *M. musculus* small intestine (33). Many of these are utilized in the fermented foods industry, such as production of kimchi, or as dietary supplements, but others were isolated as constituents of the GI microbiota of several varieties of animals. *L. reuteri* strain LL7 was in a cluster that mainly comprised isolates from *M. musculus* (panel B).

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

246 observed in other *Lactobacillus* spp. (37). In all genomes there was evidence of lysogenic
247 bacteriophages or their remnants. All species except *L. reuteri* discernibly had coding sequences
248 for Class I or Class III bacteriocins or their specific transport and immunity proteins (Table S4 of
249 Supplementary information).

250 Table 2 summarizes differentiating genetic profiles among the four species for 11 selected
251 genes or pathways. Two species, *L. reuteri* and "L. peromysci", had coding sequences for a
252 urease, which could provide for tolerance of acidic conditions, such as in the stomach. A urease
253 had previously been identified in a *L. reuteri* strain that was considered a gut symbiont in
254 rodents (38). The four species had *secY1-secA1* transport and secretion systems. Accessory Sec
255 systems (*secY2-secA2*) were identified in genomes of *L. reuteri*, *L. johnsonii*, and *L. animalis* but
256 not in "L. peromysci". The LL7 strain of *L. reuteri* on its megaplasmid also had coding sequences
257 for a third SecY-SecA system. An accessory Sec system was involved with adhesion and biofilm
258 formation in the Lactobacillales bacterium *Streptococcus pneumoniae* (39). A coding sequence for
259 an IgA protease was identified in *L. johnsonii* but not in the other three species. An IgA protease
260 in another strain of *L. johnsonii* was associated with long-term persistence in the gut of mice (40).
261 The presence or absence of other genes or pathways that differentiated between the four species
262 were an L-rhamnose biosynthesis pathway in one species, a *luxS* gene associated with a quorum
263 sensing system in *L. reuteri* and *L. johnsonii* (41), a type 1 CRISPR-Cas3 array in "L. peromysci"
264 (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).

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
reuteri	+	+	+	-	-	+	-	-	-	+	-
johnsonii	-	+	-	+	-	+	-	+	+	-	-
animalis	-	+	-	-	-	-	-	+	+	-	+
"peromysci"	+	-	-	-	+	-	+	+	+	-	-

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* represenstatives (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.

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, as well as analysis of the 16S ribosomal RNA gene sequence from a fecal sample from another animal (Fig S5 of Supplementary information), 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. 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).

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

295 metagenome of *P. leucopus* and from other sources. The respective phylogenetic analyses used
296 concatenated sequences of the following: ribosomal protein genes (panel A); the DNA gyrase A
297 (GyrA), phenylalanyl t-RNA synthase, alpha subunit (PheS), and chromosomal replication
298 iniator protein (DnaA) (panel B); DNA-directed RNA polymerase, beta-subunit (RpoB) (panel
299 C); and DNA gyrase B (GyrB), tyrosyl t-RNA synthase (TyrS), and DnaA (panel D). The distance
300 criteria were Jukes-Cantor for the codon-aligned nucleotide sequences and Poisson for amino
301 acid sequences. The scales for distance are shown in each panel. Percent bootstrap (100
302 iterations) support values of ≥ 80% at a node are shown.

303 ***Spirochaetaceae***

304 Panels B, C, and D of Fig 4 are phylograms of three other types of bacteria that were
305 identified among the high-coverage metagenomic contigs (Tables 1 and S5). The uncultured
306 spirochete LL50 was placed in the genus *Treponema* by the MG-RAST analysis program. Yet
307 species in this genus are highly divergent and include free-living organisms in a variety of
308 environments, symbionts of termites, the agent of syphilis, and gut residents, such as *T.*
309 *porcinum*, which was isolated from the feces of pigs (52). More distant still was the agent of
310 Lyme disease, *Borreliella burgdorferi*, of the family *Borreliaceae* (53). In our view naming the
311 organism as a "treponeme" would provide little insight about its role in the microbiome and
312 may even be misleading.

313 **Seven other bacterial taxa**

The *Mycoplasmataceae* bacterium LL85 (panel C of Fig 4 and Table S5) was unlike any other mollicute represented in the database but was in cluster with vertebrate-associated species, like *M. pneumoniae*. But there is also deep branching in this tree, as the tree including as outgroup two Firmicutes shows. Panel D is a phylogram of selected alphaproteobacteria and includes the organism LL75 (Table S5). 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.

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 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 ~10^6 per milliliter of unconcentrated cecal fluid (Fig 5; S1 File). 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 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 μm.

Fig 6 shows phylograms of nucleotide sequence of the small subunit (SSU) ribosomal RNA gene (panel A; Table S5) and of the partial amino acid sequence of the iron hydrogenase of the hydrogenosome of anaerobic protozoa (panel B; Table S5) (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 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 (Fig S4 of Supplementary information). *T. foetus* is a sexually-transmitted pathogen of cattle (61) and a cause of chronic diarrhea in domestic cats (62).

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

371 hydrogenase protein (panel B) of *Tritrichomonas* sp. LL5 of *P. leucopus* and selected other
372 parabasilids and other microbes. The distance criteria were observed differences for nucleotide
373 alignment and Poisson for amino acid alignment. The scales for distance are shown in each
374 panel. Percent bootstrap (100 iterations) support values of ≥ 80% at a node are shown.

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

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

383 The preceding study revealed several microbes that were either undescribed species or
384 genera, e.g. "L. peromysci" or the *Candidatus* Melainabacteria bacterium, or new strains of
385 known microbial species, e.g. *L. animalis* LL1 and *T. muris* LL4. These novelties notwithstanding,
386 to what extent did the gut microbiota of this deermouse resemble that of the typical laboratory
387 animal, a house mouse that was maintained under similar husbandry conditions, including
388 diet? That question motivated the following experiment.

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

397 The reproducibility between replicate library constructions from the same sample was
398 assessed with quantitations of reads assigned by taxonomic family for specimens from seven *P.*
399 *leucopus* among the 20 total. Pairwise coefficients of determination (R^2) for the 91 possible
400 combinations were calculated (Table S8 of Supplementary information). The mean (95% CI) of
401 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
402 non-replicate pairs. We concluded that most of the variation between samples was attributable
403 to inter-specimen differences in the microbiota and not to technical issues in library preparation
404 or sequencing.

405 The prevalences of different taxonomic families in the *P. leucopus* and *M. musculus* gut
406 metagenomes were similar (left panel of Fig 7 and Table S9 of Supplementary information). But
407 a few families stood out as either more or less common in the deermice. Notable among these
408 were *Lactobacillaceae*, *Helicobacteriaceae*, and *Spirochaetaceae*, which were approximately 4x, 8x,

409 and 2x, respectively, more prevalent on average among microbiota of *P. leucopus* than in *M.*
410 *musculus*. There was no evidence of *Tritrichomonas* sp. in the BALB/c mice by this analysis, but
411 direct examination of intestinal contents was not carried out.

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

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

424 **Fig 8.** Heat map-formatted shading matrix of KEGG Orthology gene level annotations of gut
425 metagenomes of *P. leucopus* and *M. musculus*. The annotations were generated by
426 MicrobiomeAnalyst (https://www.microbiomeanalyst.ca). Columns are grouped by species and
427 by sex within each species. Individual animal identifications are given on the *x*-axis below the
428 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).

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 Table S11 of the Supplementary information. 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 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.

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 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.

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's Index 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. The greater scattering of the samples from Block Island animals corresponded to the alpha diversity of this group.

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 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 μm. Panels C and D, Gram stain (C) and Wright-Giemsa stain (D) of sections of squamous epithelium. Bar, 100 μm. Red arrowheads indicate gram-positive bacteria in a biofilm.

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 Wesney 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.

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 μm. Panel F, phase microscopy of wet mount of unconcentrated broth culture of "L. peromysci". Bar, 10 μm.

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) x 10^9 than in males at 0.76 (0.40-1.4) x 10^9 (*t*-test $p = 0.02$ for log-transformed values) (Table 3). There was no discernible association with place of birht, and there was no difference between females and males in the proportions of the lactobacilli were identified as "L. peromysci", *L. animalis*, and *L.*

526 *reuteri*/*L. johnsonii*. For five animals, whose lactobacilli were subjected to 16S ribosomal RNA
527 gene PCR and sequencing for confirmation, the *L. reuteri*/*L. johnsonii*-type colonies are
528 predominantly *L. reuteri*. But *L. johnsonii* was confirmed to be present as well and outnumbered
529 *L. reuteri* in one animal. The results for 3 animals that had been on a 9% fat content diet, which
530 was part of the breeding program, instead of 6% fat content were not distinguishable from those
531 for the other 17.

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

Animal ID	Sex	Log_{10} total cfu/gm	% colony type		
			animalis	"peromysci"	*reuteri/johnsonii*
22403	F	10.9	3	32	65
22404	F	11.3	61	4	35
25053	F	10.7	37	60	3
25054	F	11.0	96	4	<1 [a]
25055	F	10.4	94	6	<1
25065	F	8.9	50	19	31
25062	F	8.5	60	14	26 (26/0) [b]
25063	F	8.8	52	12	36 (33/3)
25058	F	8.3	82	14	5 (5/0)
22401	M	8.6	33	19	48
22375	M	8.8	58	3	40
22420	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] <1, below limit of detection by serial dilution on plates

[b] (/), % *reuteri*/% *johnsonii* by PCR and 16S ribosomal rDNA sequencing

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 (Table S7 of Supplementary information). 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 were similar by these measures for all these groups. 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

548 in the stomach extract, while a typically abundnant genus in the intestine, *Prevotella*, was
549 present in small numbers in some samples. The lactobacilli in the stomach were about as
550 numerous as the stomach tissue cells constituting the sample.

551 **Fig 13.** Box-whisker plots of log-transformed normalized reads of total metagenomes of the
552 stomachs of 9 *P. leucopus* (left panel) and *M. musculus* (right panel) that mapped to
553 chromosomes of *Lactobacillus* spp. or other bacteria by host species and by sex. The references to
554 which reads were mapped were complete chromosomes or partial chromosomes of organisms
555 listed in Table 1. "*Lactobacillus*" in the first position of each panel were the cumulative reads for
556 the four individual *Lactobacillus* species in this analysis.

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

563 One limitation of this experiment is that we may have overlooked species that were not
564 identified because they were not cultivable by our method and conditions, which were
565 microaerophilic, not strictly anaerobic. That said, if cells of such non-cultivable lactobacilli had
566 been present in the feces or stomach, their numbers did not reach a threshold for assembly into
567 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 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 (Table S14 of Supplementary information). 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

587 Bray-Curtis dissimilarity of the inbred BALB/c *M. musculus*, outbred LL stock *P. leucopus*, and
588 the natural population of *P. leucopus* of Block Island.

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

598 Using the chromosome sequences of the four *Lactobacillus* species and partial
599 chromosome sequences of *Spirochaetaceae* bacterium LL50, *Prevotella* sp. LL70, and *Helicobacter*
600 sp. LL4 as references, we mapped and counted reads, as described for the LL stock and *M.*
601 *musculus* study above. Fig 14 summarizes results for the 18 animals grouped by sex. Lactobacilli
602 were common but, as seen with family level matching, there was greater variation between
603 samples of the different animals than was observed for colony animals. There was also
604 substantial variation in prevalences of the *Spirochaetaceae* bacterium and the *Prevotella* species. In
605 most of the samples there was scant evidence of the *Helicobacte*r species but in two animals,
606 there were higher numbers of this organism, reaching 7% of the total reads in one fecal sample.

607 **Fig 14.** Box-whisker plots of log-transformed normalized reads of fecal metagenomes of 5
608 female (red) and 4 male (blue) *P. leucopus* of a natural population of Block Island, Rhode Island.
609 The reference genomes and other sequences were those described for Fig 13 and in addition the
610 partial chromosome sequence of *Spirochaetaceae* sp. LL50. As an estimate of the number of
611 mammalian cells in the extract, "nuclei" corresponded with normalized reads mapped to *P.*
612 *leucopus* genome.

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

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

627 aligned with the corresponding sequences of the LL6 isolate of "L. peromysci". Isolate 7442BI
628 was comparatively more distant from the LL1 isolate of *L. animalis* in the stock colony but still
629 clustered with it rather than with other examples of *L. animalis*.

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

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

642 This survey also documented that a strain or strains of "L. peromysci" and *L. animalis* are
643 present in the native deermice. The high degree of sequence identity between two "L.
644 peromysci" examples, whose origins were North Carolina and a New England island, long
645 separated from the mainland, suggest that this newly-discovered species is autochthonous and
646 plausibly a narrowly host-restricted symbiont of *P. leucopus*. Host-range restrictions of

647 lactobacilli for the stomachs of mice were demonstrated by Wesney and Tannock (71).
648 Supporting an assignment of a symbiont lifestyle was its smaller genome size and lower % GC
649 content of this species in comparison with *L. reuteri* and *L. johnsonii* with their broader host
650 ranges (79).

651 Conclusions

652 Six decades ago René Dubos (of the epigraph), Russell Schaedler, and their colleagues at
653 what is now Rockefeller University reported in a series of ground-breaking papers on the "fecal
654 flora" of mice and variations in that microflora between mouse strains (80-82). They associated
655 differences in gastrointestinal flora with growth rates of the mice and the mouse's susceptibility
656 to infection and endotoxin. A featured group of bacteria in their studies were lactobacilli. They
657 showed that this group of bacteria were present in large numbers in the feces and that they
658 predominated (up to 10^9 cfu per g of homogenate) in the stomachs of the mice (70), similarly to
659 what we observed in *P. leucopus*. As their studies first intimated, the rodent may plausibly owe
660 as much to the genomes of their microbiota as to the nuclei and mitochondria of their somatic
661 cells for either ameliorating or exasperating disease (83).

662 There are also implications of our findings for development of oral vaccines that target *P.*
663 *leucopus* to block transmission of pathogens either from tick to the reservoir or from the
664 reservoir to the tick. Two of the candidate vehicles for the bait delivery of recombinant vaccine
665 antigens to rodents have been *E. coli* and a *Lactobacillus* species (30, 84). In neither case were the

666 strains known to be adapted for life in *P. leucopus*. Success rate for achieving a protective
667 response may be enhanced by use as the bacterial vehicle microbes that are adapted to *P.*
668 *leucopus*. Such organisms presumably would more likely than an allochthonous bacterium to
669 stably colonize and then proliferate to numbers large enough for the recombinant protein to
670 elicit the sought-after immune responses.

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

683 Materials and methods

684 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/cAnNCrl (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. 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 *Borreliella 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 x 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 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 Nikkor 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 and cecum, and from bacteria harvested from broth cultures were extracted with ZymoBIOMICS™ DNA Miniprep or Microprep kits (Zymo Research, Irvine, CA). Freshly-dissected, unwashed tissues were cut into small pieces before trituration and then homogenization in the lysis buffer. DNA concentration was determined by 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 for the genus *Lactobacillus*: forward 5'-CCTAATACATGCAAGTCG and reverse 5'-GGTTCTCCTACGGCTA. The Platinum Taq polymerase and master mix

762 (ThermoFisher Scientific) contained uracil-DNA glycosylase. On a T100 thermal cycler (BioRad)
763 PCR conditions (°C for temperature) The conditions were 37° for 10 min, 94° for min, 40 cycles
764 of 94° for 10 s, 55° for 30 s, and 72° for 45 s. The 1.5 kb PCR product was isolated and purified
765 from agarose gel using the NucleoSpin Gel and PCR Clean-up kit (Takara). The product was
766 subjected to Sanger dideoxy sequencing at GENEWIZ (San Diego, CA).

767 Whole genome sequencing, assembly, and annotation

768 Long reads were obtained using an Oxford Nanopore Technology MinION Mk1B
769 instrument with Ligation Sequencing Kit, R9.4.1 flow cell, MinKnow v. 19.6.8 for primary data
770 acquisition, and Guppy v. 3.2.4 for base calling with default settings. Paired-end short reads
771 were obtained on a MiSeq sequencer with paired-end v2 Micro chemistry and 150 cycles
772 (Illumina, San Diego, CA). The library was constructed using the NEXTflex Rapid DNA-Seq kit
773 (Bioo Scientific, Austin, TX), the quality of sequencing reads was analyzed using FastQC
774 (Babraham Bioinformatics), and reads were trimmed of Phred scores <15 and corrected for poor-
775 quality bases using Trimmomatic (88). A hybrid assembly was carried out with Unicycler v.0.4.7
776 (89) with default settings and 16 threads on the High Performance Computing cluster of the
777 University of California Irvine. Assembly of short reads alone were performed with the
778 Assembly Cell program of CLC Genomics Workbench v. 11 (Qiagen). Annotation was provided
779 by the NCBI Prokaryotic Genome Annotation Pipeline (90). Putative bacteriocins and their
780 associated transport and immunity functions were identified by BAGEL4 (91, 92). For other
781 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 analysis

The same DNA extract used for the metagenome sequencing at University of California Irvine was submitted to ID Genomics, Inc. (Seattle, WA) and subjected 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). The output of the MG-RAST protocol was analyzed in *R* using the *vegan* package (https://cran.r-project.org, https://github.com/vegandevs/vegan). Alpha diversity was expressed by the Shannon's Diversity Index, which accounts for evenness as well as richness (96). Beta diversity 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 non-metric 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

820 chromosomes and large contigs (100). For *Lactobacillus* spp. and the *Helicobacter* sp. the lactic
821 acid bacteria database and *Helicobacter* database, respectively, options were chosen. Alignments
822 and phylogenetic analysis were carried out with the SeaView v. 4 suite (101).

823 Data availability

824 The resources for the several new sequences for genomes, large contigs, and individual
825 genes that described here are listed in Table 1. The accession numbers for the annotated
826 genomes and plasmids of *L. animalis* LL1, *L. reuteri* LL7, and "L. peromysci" LL6 are given in
827 Bassam et al. (24). Fig 2 and its legend provides accession numbers for 16S ribosomal RNA gene
828 sequences of other *Lactobacillus* species and strains. The nucleotide sequences of ribosomal
829 proteins for different species and strains of *Lactobacillus* and *Helicobacter* were obtained from the
830 Ribosomal MLST database of PubMLST (https://pubmlst.org/rmlst/) and given in Table S3 of
831 Supplementary information. Additional accession numbers for individual genes of other
832 organisms as references are given in Table S5 of Supplementary information. Hyperlinks to the
833 long (Nanopore) and short (Illumina) sequence reads at the Sequence Read Archive or MG-
834 RAST database are provided in Table 1 and Tables S7 and S13 of Supplementary information.

835 Statistical analysis

836 Normalized reads and other values whose distributions spanned more than one order of
837 magnitude were log-transformed before parametric analysis by 2-tailed *t*-test. Inverse
838 transformation was carried out to provide nonparametric means and corresponding asymmetric

839 95% confidence intervals. The *Z*-score was the number of standard deviations below or above
840 the population mean a give raw value was. The False Discovery Rate (FDR) with corrected *p*
841 value was estimated by the method of Benjamini and Hochberg (102). The box-whisker plot
842 graphs were made with SYSTAT v. 13.1 software (Systat Software, Inc.).

843 Acknowledgements

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

846 The research was supported by National Institutes of Health grant R21 AI136523 to A.G.B.,
847 National Science Foundation grants 1736150 and 1755670 to H. Kiaris (University of South
848 Carolina), and National Science Foundation (Division of Integrative Organismal Systems) grant
849 IOS1755286 to D.M.T. and M.A.D.

850

851 References

852 1. Dubos R. Autobiographical Manuscript. New York: Rockefeller University; 1981. p.
853 Folder 14.

854 2. Sangodeyi F. Rene Dubos and the Emerging Science of Human Microbial Ecology New
855 York: Rockefeller University; 2012 [https://www.issuelab.org/resources/27965/27965.pdf]

856 3. Hall ER. The Mammals of North America. Second ed. New York: John Wiley and Sons;
857 1981.

858 4. Lackey JA, Huckaby DG, Ormiston BG. Peromyscus leucopus. Mammalian Species.
859 1985(247):1-10.

860 5. Munshi-South J, Kharchenko K. Rapid, pervasive genetic differentiation of urban white-
861 footed mouse (Peromyscus leucopus) populations in New York City. Mol Ecol. 2010.

862 6. Crossland JP, Dewey MJ, Barlow SC, Vrana PB, Felder MR, Szalai GJ. Caring for
863 *Peromyscus* spp. in research environments. Lab Animal. 2014;43(5):162-6.

864 7. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees,
865 and divergence times. Mol Biol Evol. 2017;34:1812-9.

866 8. Steppan SJ, Schenk JJ. Muroid rodent phylogenetics: 900-species tree reveals increasing
867 diversification rates. PLoS One. 2017;12(8):e0183070.

868 9. Schrago CG, Voloch CM. The precision of the hominid timescale estimated by relaxed
869 clock methods. J Evol Biol. 2013;26(4):746-55.

870 10. Schug MD, Vessey SH, Korytko AI. Longevity and survival in a population of white-
871 footed mice (Peromyscus leucopus). J Mammal. 1991;72:360-6.

872 11. Bunikis J, Tsao J, Luke CJ, Luna MG, Fish D, Barbour AG. Borrelia burgdorferi infection
873 in a natural population of Peromyscus leucopus mice: a longitudinal study in an area where
874 Lyme Borreliosis is highly endemic. J Infect Dis. 2004;189(8):1515-23.

875 12. Sacher G, Hart R. Longevity, aging, and comparative cellular and molecular biology of
876 the house mouse, Mus musculus, and the white-footed mouse, Peromyscus leucopus. Birth
877 Defects, Orig Artic Ser;(United States). 1978;14(1).

878 13. Shorter KR, Owen A, Anderson V, Hall-South AC, Hayford S, Cakora P, et al. Natural
879 genetic variation underlying differences in Peromyscus repetitive and social/aggressive
880 behaviors. Behavior Genet. 2014;44(2):126-35.

881 14. Veres M, Duselis AR, Graft A, Pryor W, Crossland J, Vrana PB, et al. The biology and
882 methodology of assisted reproduction in deer mice (Peromyscus maniculatus). Theriogenology.
883 2012;77(2):311-9.

884 15. Barbour AG. Infection resistance and tolerance in Peromyscus spp., natural reservoirs of
885 microbes that are virulent for humans. Semin Cell Devel Biol. 2017;61(1):115-22.

886 16. Bedford NL, Hoekstra HE. Peromyscus mice as a model for studying natural variation.
887 eLife. 2015;4:e06813.

888 17. Barbour AG, Shao H, Cook VJ, Baldwin-Brown J, Tsao JI, Long AD. Genomes, expression
889 profiles, and diversity of mitochondria of the White-footed Deermouse Peromyscus leucopus,
890 reservoir of Lyme disease and other zoonoses. Sci Rep. 2019;9(1):17618.

891 18. Long A, Baldwin-Brown J, Tao Y, Cook V, Balderrama-Gutierrez G, Corbett-Detig R, et
892 al. The genome of Peromyscus leucopus, natural host for Lyme disease and other emerging
893 infections. Sci Advances. 2019;5(7):eaaw6441.

894 19. Gomes-Solecki MJ, Brisson DR, Dattwyler RJ. Oral vaccine that breaks the transmission
895 cycle of the Lyme disease spirochete can be delivered via bait. Vaccine. 2006;24(20):4440-9.

896 20. Najjar DA, Normandin AM, Strait EA, Esvelt KM. Driving towards ecotechnologies.
897 Pathogens Global Health. 2017;111(8):448-58.

898 21. Specter M. Annals of Science: Rewriting the Code of Life. The New Yorker. 2017;93(1);
899 January 2.

900 22. Baxter NT, Wan JJ, Schubert AM, Jenior ML, Myers P, Schloss PD. Intra- and
901 interindividual variations mask interspecies variation in the microbiota of sympatric
902 peromyscus populations. Appl Environ Microbiol. 2015;81(1):396-404.

903 23. Kohl KD, Dearing MD, Bordenstein SR. Microbial communities exhibit host species
904 distinguishability and phylosymbiosis along the length of the gastrointestinal tract. Mol Ecol.
905 2018;27(8):1874-83.

906 24. Bassam K, Milovic A, Barbour AG. Genome sequences of three Lactobacillus species
907 strains of the stomach of the White-footed Deermouse (Peromyscus leucopus). Microbiol Res
908 Announc. 2019;8(40).

909 25. Herlemann DP, Geissinger O, Ikeda-Ohtsubo W, Kunin V, Sun H, Lapidus A, et al.
910 Genomic analysis of "Elusimicrobium minutum," the first cultivated representative of the
911 phylum "Elusimicrobia" (formerly termite group 1). Appl Environ Microbiol. 2009;75(9):2841-9.

912 26. Soo RM, Skennerton CT, Sekiguchi Y, Imelfort M, Paech SJ, Dennis PG, et al. An
913 expanded genomic representation of the phylum Cyanobacteria. Genome Biol Evol.
914 2014;6(5):1031-45.

915 27. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, et al. Population-level
916 analysis of gut microbiome variation. Science. 2016;352(6285):560-4.

917 28. Walter J. Ecological role of lactobacilli in the gastrointestinal tract: implications for
918 fundamental and biomedical research. Appl Environ Microbiol. 2008;74(16):4985-96.

919 29. Tannock G. The microecology of lactobacilli inhabiting the gastrointestinal tract. In:
920 Marshall KC, editor. Advances in Microbial Ecology. 11. Boston, MA: Springer; 1990. p. 147-71.

921 30. Gomes-Solecki M, Richer L. Recombinant E. coli dualistic role as an antigen-adjuvant
922 delivery vehicle for oral immunization. In: Pal U, Buyuktanir O, editors. Borrelia burgdorferi
923 Methods in Molecular Biology. New York: Humana Press; 2018. p. 347-57.

924 31. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, et al. Ribosomal
925 multilocus sequence typing: universal characterization of bacteria from domain to strain.
926 Microbiology. 2012;158(Pt 4):1005-15.

927 32. Fujisawa T, Itoh K, Benno Y, Mitsuoka T. Lactobacillus intestinalis (ex Hemme 1974) sp.
928 nov., nom. rev., isolated from the intestines of mice and rats. Int J Syst Bacteriol. 1990;40(3):302-
929 4.

930 33. Kim D, Cho MJ, Cho S, Lee Y, Byun SJ, Lee S. Complete Genome Sequence of
931 Lactobacillus reuteri Byun-re-01, isolated from mouse small intestine. Microbiol Res Announc.
932 2018;7(17).

933 34. Sun Z, Harris HM, McCann A, Guo C, Argimon S, Zhang W, et al. Expanding the
934 biotechnology potential of lactobacilli through comparative genomics of 213 strains and
935 associated genera. Nat Commun. 2015;6:8322.

936 35. Dent V, Williams R. Lactobacillus animalis sp. nov., a new species of Lactobacillus from
937 the alimentary canal of animals. Zentralblatt für Bakteriologie Mikrobiologie und Hygiene: I
938 Abt Originale C: Allgemeine, angewandte und ökologische Mikrobiologie. 1982;3(3):377-86.

939 36. Nam SH, Choi SH, Kang A, Kim DW, Kim RN, Kim A, et al. Genome sequence of
940 Lactobacillus animalis KCTC 3501. J Bacteriol. 2011;193(5):1280-1.

941 37. Claesson MJ, Li Y, Leahy S, Canchaya C, van Pijkeren JP, Cerdeno-Tarraga AM, et al.
942 Multireplicon genome architecture of Lactobacillus salivarius. Proc Natl Acad Sci U S A.
943 2006;103(17):6718-23.

944 38. Frese SA, Benson AK, Tannock GW, Loach DM, Kim J, Zhang M, et al. The evolution of
945 host specialization in the vertebrate gut symbiont Lactobacillus reuteri. PLoS Genet.
946 2011;7(2):e1001314.

947 39. Bandara M, Skehel JM, Kadioglu A, Collinson I, Nobbs AH, Blocker AJ, et al. The
948 accessory Sec system (SecY2A2) in Streptococcus pneumoniae is involved in export of
949 pneumolysin toxin, adhesion and biofilm formation. Microbes Infect. 2017;19(7-8):402-12.

950 40. Denou E, Pridmore RD, Berger B, Panoff J-M, Arigoni F, Brüssow H. Identification of
951 genes associated with the long-gut-persistence phenotype of the probiotic Lactobacillus
952 johnsonii strain NCC533 using a combination of genomics and transcriptome analysis. J
953 Bacteriol. 2008;190(9):3161-8.

954 41. Wilson CM, Aggio RB, O'Toole PW, Villas-Boas S, Tannock GW. Transcriptional and
955 metabolomic consequences of LuxS inactivation reveal a metabolic rather than quorum-sensing
956 role for LuxS in Lactobacillus reuteri 100-23. J Bacteriol. 2012;194(7):1743-6.

957 42. Hidalgo-Cantabrana C, Goh YJ, Pan M, Sanozky-Dawes R, Barrangou R. Genome editing
958 using the endogenous type I CRISPR-Cas system in Lactobacillus crispatus. Proc Natl Acad Sci
959 U S A. 2019;116(32):15774-83.

960 43. Saulnier DM, Santos F, Roos S, Mistretta TA, Spinler JK, Molenaar D, et al. Exploring
961 metabolic pathway reconstruction and genome-wide expression profiling in Lactobacillus
962 reuteri to define functional probiotic features. PLoS One. 2011;6(4):e18783.

963 44. Moreno-Vivian C, Cabello P, Martinez-Luque M, Blasco R, Castillo F. Prokaryotic nitrate
964 reduction: molecular properties and functional distinction among bacterial nitrate reductases. J
965 Bacteriol. 1999;181(21):6573-84.

966 45. Hwang H, Lee JH. Characterization of arginine catabolism by lactic acid bacteria isolated
967 from kimchi. Molecules (Basel, Switzerland). 2018;23(11).

968 46. Singer JR, Blosser EG, Zindl CL, Silberger DJ, Conlan S, Laufer VA, et al. Preventing
969 dysbiosis of the neonatal mouse intestinal microbiome protects against late-onset sepsis. Nat
970 Med. 2019;25(11):1772-82.

971 47. Peña JA, Li SY, Wilson PH, Thibodeau SA, Szary AJ, Versalovic J. Genotypic and
972 phenotypic studies of murine intestinal lactobacilli: species differences in mice with and without
973 colitis. Appl Environ Microbiol. 2004;70(1):558-68.

974 48. Sheh A, Shen Z, Fox JG. Draft genome sequences of eight enterohepatic Helicobacter
975 species isolated from both laboratory and wild rodents. Genome Announc. 2014;2(6).

976 49. Suerbaum S, Josenhans C, Sterzenbach T, Drescher B, Brandt P, Bell M, et al. The
977 complete genome sequence of the carcinogenic bacterium Helicobacter hepaticus. Proc Natl
978 Acad Sci U S A. 2003;100(13):7901-6.

979 50. Fox JG, Ge Z, Whary MT, Erdman SE, Horwitz BH. Helicobacter hepaticus infection in
980 mice: models for understanding lower bowel inflammation and cancer. Mucosal Immunol.
981 2011;4(1):22-30.

982 51. Bracken TC, Cooper CA, Ali Z, Truong H, Moore JM. Helicobacter infection significantly
983 alters pregnancy success in laboratory mice. J Am Assoc Lab Animal Sci. 2017;56(3):322-9.

984 52. Nordhoff M, Taras D, Macha M, Tedin K, Busse HJ, Wieler LH. Treponema berlinense sp.
985 nov. and Treponema porcinum sp. nov., novel spirochaetes isolated from porcine faeces. Int J
986 Syst Evol Microbiol. 2005;55(Pt 4):1675-80.

987 53. Barbour AG. *Borreliaceae*. In: Whitman WB, Rainey R, P. Kämpfe P, Trujillo M, Chun J,
988 DeVos P, et al., editors. Bergey's Manual of Systematics of Archaea and Bacteria. 2018. p. 1-9.

989 54. Geissinger O, Herlemann DP, Morschel E, Maier UG, Brune A. The ultramicrobacterium
990 "Elusimicrobium minutum" gen. nov., sp. nov., the first cultivated representative of the termite
991 group 1 phylum. Appl Environ Microbiol. 2009;75(9):2831-40.

992 55. Lagkouvardos I, Lesker TR, Hitch TCA, Galvez EJC, Smit N, Neuhaus K, et al. Sequence
993 and cultivation study of Muribaculaceae reveals novel species, host preference, and functional
994 potential of this yet undescribed family. Microbiome. 2019;7(1):28.

995 56. Doran DJ. A catalogue of the protozoa and helminths of North American rodents. I.
996 Protozoa and Acanthocephala. Am Midland Naturalist. 1954;52(1):118-28.

997 57. Ericsson AC, Gagliardi J, Bouhan D, Spollen WG, Givan SA, Franklin CL. The influence
998 of caging, bedding, and diet on the composition of the microbiota in different regions of the
999 mouse gut. Sci Rep. 2018;8(1):4065.

1000 58. Steiner JM, Schwamberger S, Pantchev N, Balzer HJ, Vrhovec MG, Lesina M, et al. Use of
1001 ronidazole and limited culling to eliminate Tritrichomonas muris from laboratory mice. J Am
1002 Assoc Lab Animal Sci. 2016;55(4):480-3.

1003 59. Koyama T, Endo T, Asahi H, Kuroki T. Life cycle of Tritrichomonas muris. Zentralblatt
1004 für Bakteriologie, Mikrobiologie und Hygiene Series A: Medical Microbiology, Infectious
1005 Diseases, Virology, Parasitology. 1987;264(3-4):478-86.

1006 60. Hackstein JH, Akhmanova A, Boxma B, Harhangi HR, Voncken FG. Hydrogenosomes:
1007 eukaryotic adaptations to anaerobic environments. Trends Microbiol. 1999;7(11):441-7.

1008 61. Rae DO, Crews JE. Tritrichomonas foetus. Vet Clin North Am Food Animal Practice.
1009 2006;22(3):595-611.

1010 62. Yao C, Koster LS. Tritrichomonas foetus infection, a cause of chronic diarrhea in the
1011 domestic cat. Vet Res. 2015;46:35.

1012 63. Escalante NK, Lemire P, Cruz Tleugabulova M, Prescott D, Mortha A, Streutker CJ, et al.
1013 The common mouse protozoa Tritrichomonas muris alters mucosal T cell homeostasis and
1014 colitis susceptibility. J Exp Med. 2016;213(13):2841-50.

1015 64. Howitt MR, Lavoie S, Michaud M, Blum AM, Tran SV, Weinstock JV, et al. Tuft cells,
1016 taste-chemosensory cells, orchestrate parasite type 2 immunity in the gut. Science.
1017 2016;351(6279):1329-33.

1018 65. Chudnovskiy A, Mortha A, Kana V, Kennard A, Ramirez JD, Rahman A, et al. Host-
1019 rotozoan interactions protect from mucosal infections through activation of the inflammasome.
1020 Cell. 2016;167(2):444-56.e14.

1021 66. Lange BM, Croteau R. Isopentenyl diphosphate biosynthesis via a mevalonate-
1022 independent pathway: Isopentenyl monophosphate kinase catalyzes the terminal enzymatic
1023 step. Proc Natl Acad Sci USA. 1999;96(24):13714-9.

1024 67. Hammes WP, Vogel RF. The genus Lactobacillus. In: Wood B, Holzapfel W, editors. The
1025 Genera of Lactic Acid Bacteria: Springer; 1995. p. 19-54.

1026 68. Salzman NH, de Jong H, Paterson Y, Harmsen HJM, Welling GW, Bos NA. Analysis of
1027 16S libraries of mouse gastrointestinal microflora reveals a large new group of mouse intestinal
1028 bacteria. Microbiol. 2002;148(Pt 11):3651-60.

1029 69. Carleton MD. A survey of gross stomach morphology in New World Cricetinae
1030 (Rodentia, Muroidea), with comments on functional interpretations. Miscellaneous Publications,
1031 University of Michigan No 1461973.

1032 70. Savage DC, Dubos R, Schaedler RW. The gastrointestinal epithelium and its
1033 autochthonous bacterial flora. J Exp Med. 1968;127(1):67-76.

1034 71. Wesney E, Tannock GW. Association of rat, pig, and fowl biotypes of lactobacilli with the
1035 stomach of gnotobiotic mice. Microbial Ecol. 1979;5(1):35-42.

1036 72. Dubos R, Schaedler RW, Costello R, Hoet P. Indigenous, normal, and autochthonous
1037 flora of the gastrointestinal tract J Exp Med. 1965;122:67-76.

1038 73. Salas-Jara MJ, Ilabaca A, Vega M, Garcia A. Biofilm forming Lactobacillus: new
1039 challenges for the development of probiotics. Microorganisms. 2016;4(3).

1040 74. Almiron M, Traglia G, Rubio A, Sanjuan N. Colonization of the mouse upper
1041 gastrointestinal tract by Lactobacillus murinus: a histological, immunocytochemical, and
1042 ultrastructural study. Curr Microbiol. 2013;67(4):395-8.

1043 75. Dertli E, Mayer MJ, Narbad A. Impact of the exopolysaccharide layer on biofilms,
1044 adhesion and resistance to stress in Lactobacillus johnsonii FI9785. BMC Microbiol. 2015;15:8.

1045 76. Lin XB, Wang T, Stothard P, Corander J, Wang J, Baines JF, et al. The evolution of
1046 ecological facilitation within mixed-species biofilms in the mouse gastrointestinal tract. ISME J.
1047 2018;12(11):2770-84.

1048 77. Huang CI, Kay SC, Davis S, Tufts DM, Gaffett K, Tefft B, et al. High burdens of Ixodes
1049 scapularis larval ticks on white-tailed deer may limit Lyme disease risk in a low biodiversity
1050 setting. Ticks Tick Borne Dis. 2019;10(2):258-68.

1051 78. Finch C, Al-Damluji MS, Krause PJ, Niccolai L, Steeves T, O'Keefe CF, et al. Integrated
1052 assessment of behavioral and environmental risk factors for Lyme disease infection on Block
1053 Island, Rhode Island. PLoS One. 2014;9(1):e84758.

1054 79. Moran NA, McCutcheon JP, Nakabachi A. Genomics and evolution of heritable bacterial
1055 symbionts. Annu Rev Genet. 2008;42:165-90.

1056 80. Dubos R, Schaedler RW. Some biological effects of the digestive flora. Am J Med Sci.
1057 1962;244:265-71.

1058 81. Dubos RJ, Schaedler RW. The effect of diet on the fecal bacterial flora of mice and on their
1059 resistance to infection. J Exp Med. 1962;115:1161-72.

1060 82. Schaedler RW, Dubos RJ. The fecal flora of various strains of mice. Its bearing on their
1061 susceptibility to endotoxin. J Exp Med. 1962;115:1149-60.

1062 83. Rosshart SP, Vassallo BG, Angeletti D, Hutchinson DS, Morgan AP, Takeda K, et al. Wild
1063 mouse gut microbiota promotes host fitness and improves disease resistance. Cell.
1064 2017;171(5):1015-28.e13.

1065 84. Del Rio B, Seegers JF, Gomes-Solecki M. Immune response to Lactobacillus plantarum
1066 expressing Borrelia burgdorferi OspA is modulated by the lipid modification of the antigen.
1067 PLoS One. 2010;5(6):e11199.

1068 85. Carroll IM, Threadgill DW, Threadgill DS. The gastrointestinal microbiome: a malleable,
1069 third genome of mammals. Mamm Genome. 2009;20(7):395-403.

1070 86. Peromyscus Genetic Stock Center. Peromyscus Genetic Stock Center, Columbia, SC:
1071 University of South Carolina; 2017 [Available from: http://stkctr.biol.sc.edu].

1072 87. Tufts DM, Diuk-Wasser MA. Transplacental transmission of tick-borne Babesia microti in
1073 its natural host Peromyscus leucopus. Parasit Vectors. 2018;11(1):286.

1074 88. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence
1075 data. Bioinformatics. 2014:btu170.

1076 89. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome
1077 assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13(6):e1005595. doi:
1078 10.1371/journal.pcbi.1005595.

1079 90. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, et al. RefSeq: an
1080 update on prokaryotic genome annotation and curation. Nucl Acids Res. 2018;46(D1):D851-d60.

1081 91. de Jong A, van Heel AJ. BAGEL4 Groningen, the Netherlands: University of Groningen;
1082 2019 [Available from: http://bagel4.molgenrug.nl].

1083 92. van Heel AJ, de Jong A, Song C, Viel JH, Kok J, Kuipers OP. BAGEL4: a user-friendly
1084 web server to thoroughly mine RiPPs and bacteriocins. Nucl Acids Res. 2018;46(W1):W278-W81.

1085 93. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, et al. The metagenomics
1086 RAST server – a public resource for the automatic phylogenetic and functional analysis of
1087 metagenomes. BMC Bioinformatics. 2008;9(1):386.

1088 94. Gomez-Alvarez V, Teal TK, Schmidt TM. Systematic artifacts in metagenomes from
1089 complex microbial communities. ISME J. 2009;3(11):1314-7.

1090 95. Cox MP, Peterson DA, Biggs PJ. SolexaQA: At-a-glance quality assessment of Illumina
1091 second-generation sequencing data. BMC Bioinformatics. 2010;11(1):485.

1092 96. Hill MO. Diversity and evenness: a unifying notation and its consequences. Ecology.
1093 1973;54(2):427-32.

1094 97. Bray JR, Curtis JT. An ordination of the upland forest communities of southern
1095 Wisconsin. Ecological Monographs. 1957;27(4):325-49.

1096 98. Ward JH. Hierarchical grouping to optimize an objective function. J Am Stat Assoc.
1097 1963;58(301):236-44.

1098 99. Chong J, Liu P, Zhou G, Xia J. Using MicrobiomeAnalyst for comprehensive statistical,
1099 functional, and meta-analysis of microbiome data. Nature Protocols. 2020;15(3):799-821.

1100 100. Tanizawa Y, Fujisawa T, Nakamura Y. DFAST: a flexible prokaryotic genome annotation
1101 pipeline for faster genome publication. Bioinformatics. 2017;34(6):1037-9.

1102 101. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user
1103 interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010;27(2):221-4.

1104 102. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful
1105 approach to multiple testing. J Royal Stat Soc Series B (Methodological). 1995:289-300.

1106

Supporting information

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

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

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

Fig S4. 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 *Thraustothea 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 setstings of the LG model, 4 rate classes, and 100 replicates.

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

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

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

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

Table S5. Accession numbers of sequences of other species

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

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

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

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

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

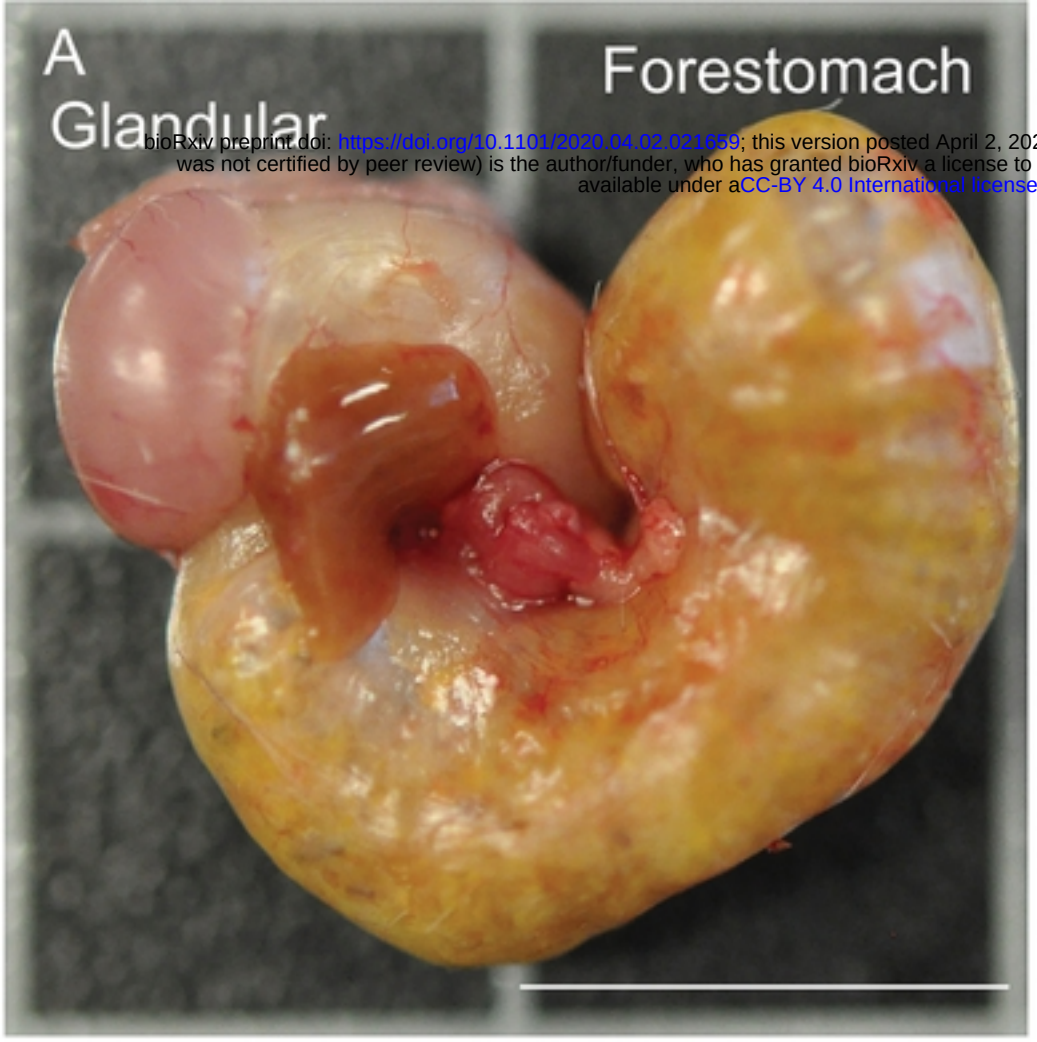
Table S11. Map-to-reference normalized PE250 reads (log10) for feces and stomach sample against *Lactobacillus* spp. and selected other bacteria

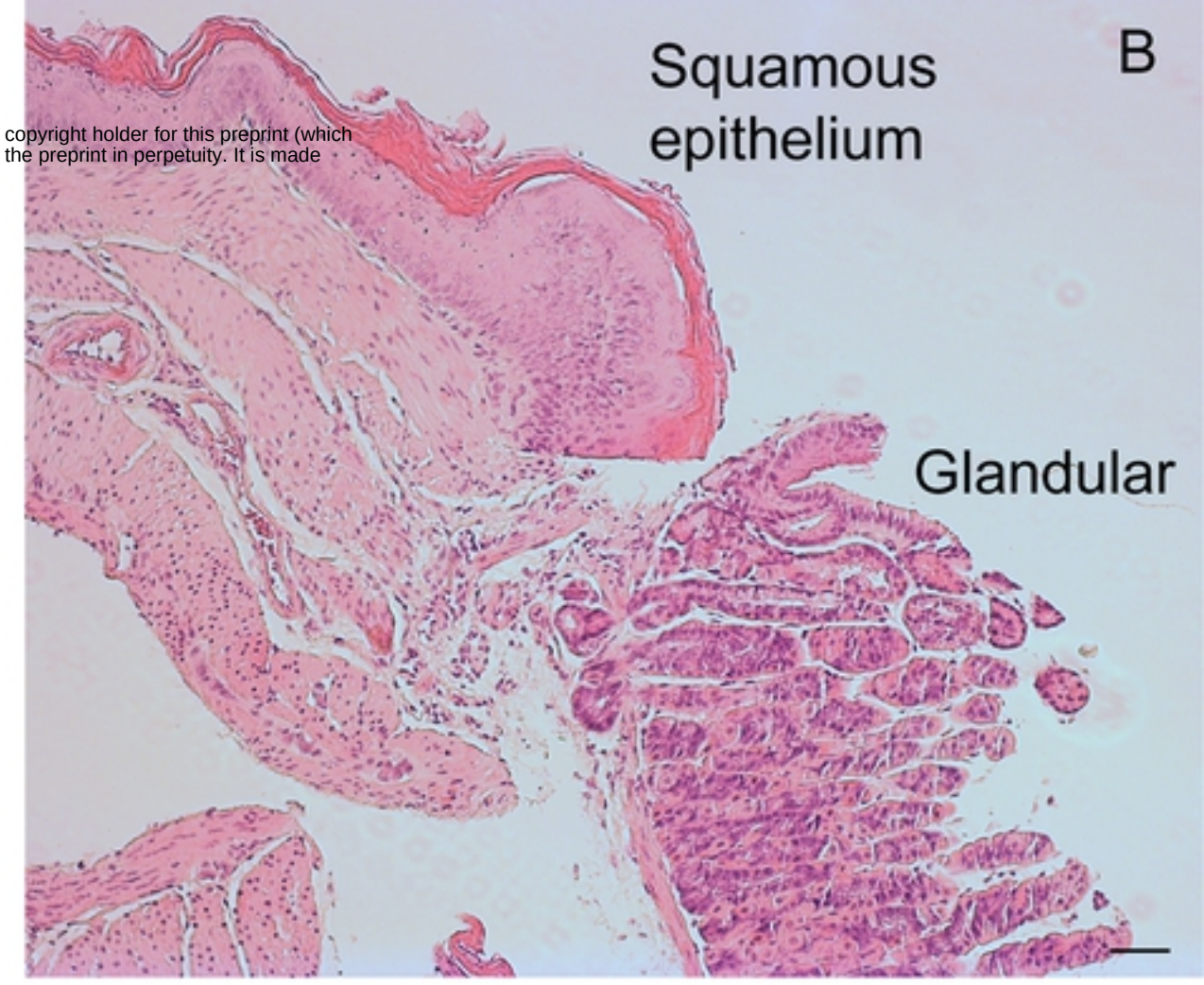
Table S12. Comparison of gut metagenomes of *Peromyscus leucopus* and *Mus musculus* by KEGG orthology gene

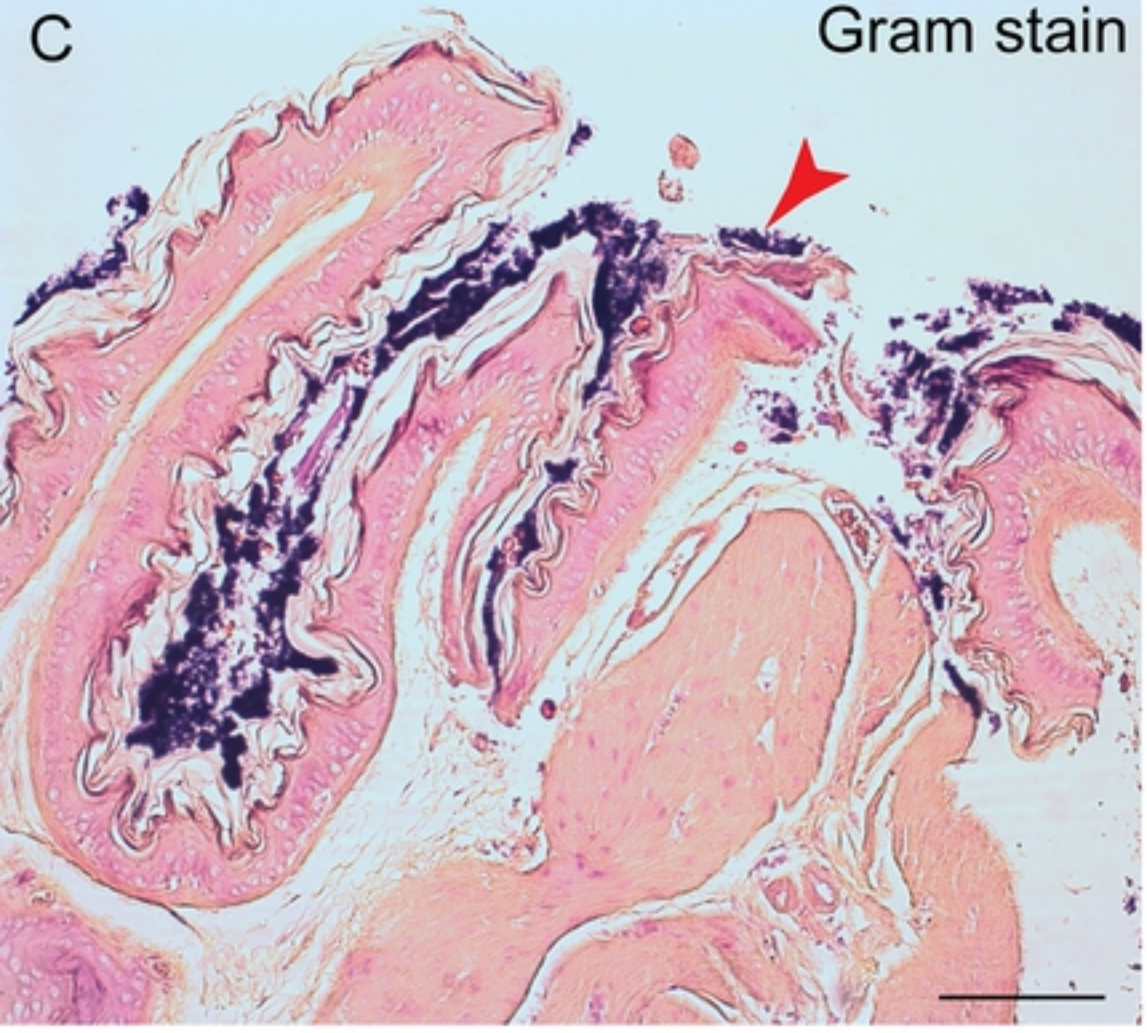
Table S13. Reads of gut metagenomes of *Peromyscus leucopus* and *Mus musculus* by KEGG orthology gene for individual animals and data for analysis of Table S11

1143 **Table S14. Block Island *Peromyscus leucopus* fecal metagenome**

1144 **Movie S1. Cecal content with *Tritrichomonas* flagellates**







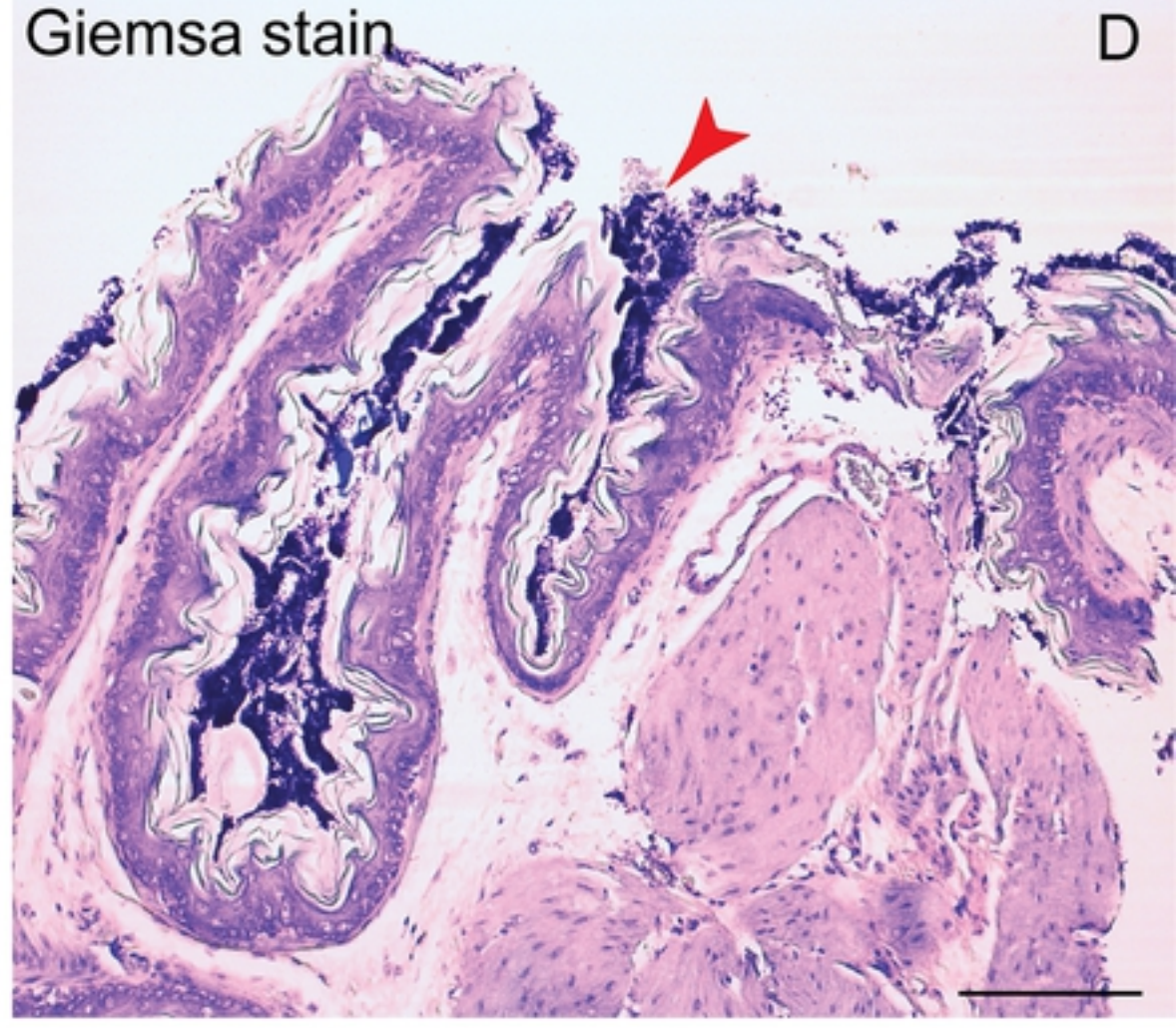


Figure 11

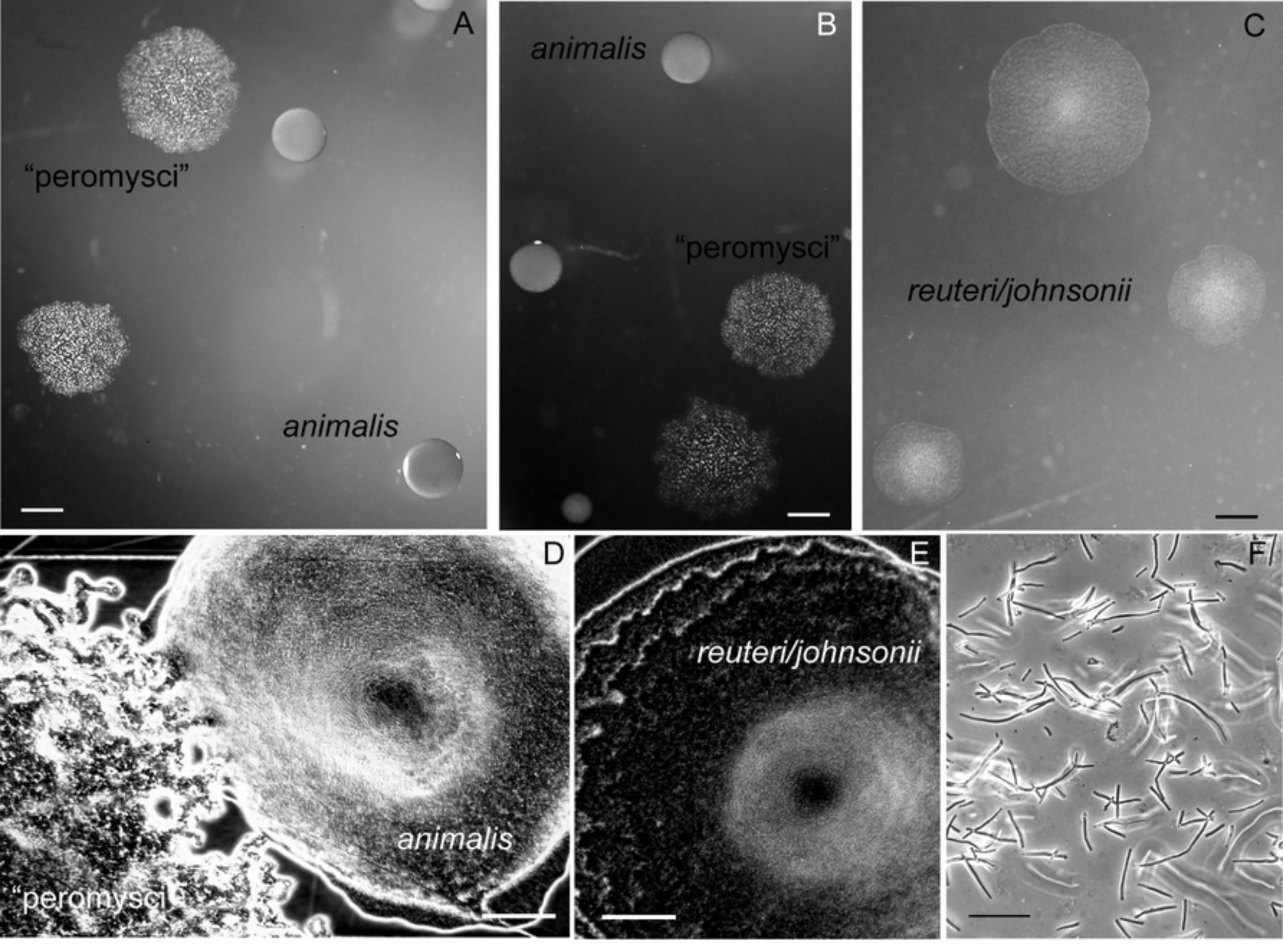

A
"peromysci"
animalis
B
animalis
"peromysci"
C
reuteri/johnsonii
D
animalis
"peromysci"
E
reuteri/johnsonii
F

Figure 12

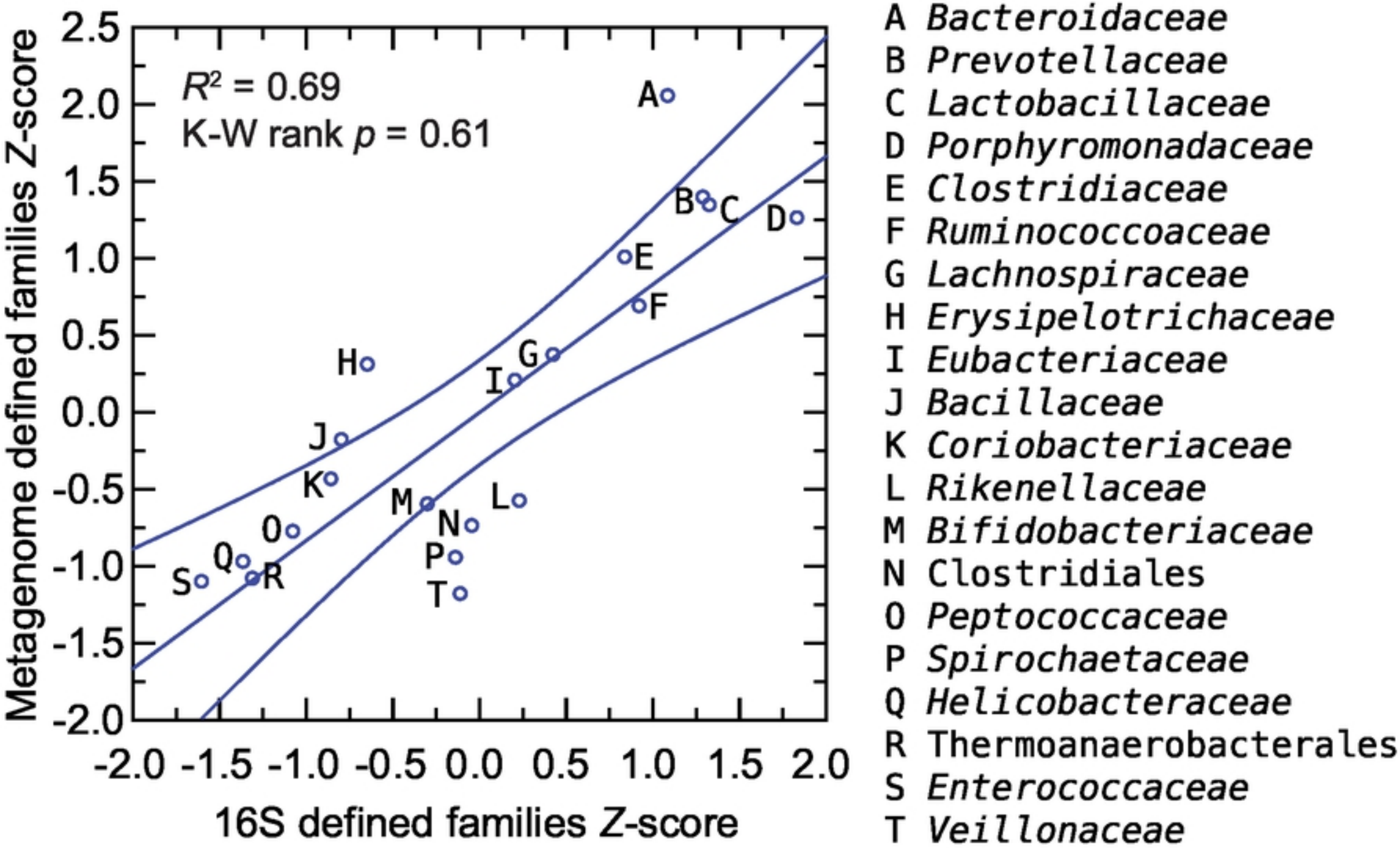

R^2 = 0.69
K-W rank p = 0.61
Metagenome defined families Z-score
16S defined families Z-score
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
A *Bacteroidaceae*
B *Prevotellaceae*
C *Lactobacillaceae*
D *Porphyromonadaceae*
E *Clostridiaceae*
F *Ruminococcoaceae*
G *Lachnospiraceae*
H *Erysipelotrichaceae*
I *Eubacteriaceae*
J *Bacillaceae*
K *Coriobacteriaceae*
L *Rikenellaceae*
M *Bifidobacteriaceae*
N Clostridiales
O *Peptococcaceae*
P *Spirochaetaceae*
Q *Helicobacteraceae*
R Thermoanaerobacterales
S *Enterococcaceae*
T *Veillonaceae*

Figure 1

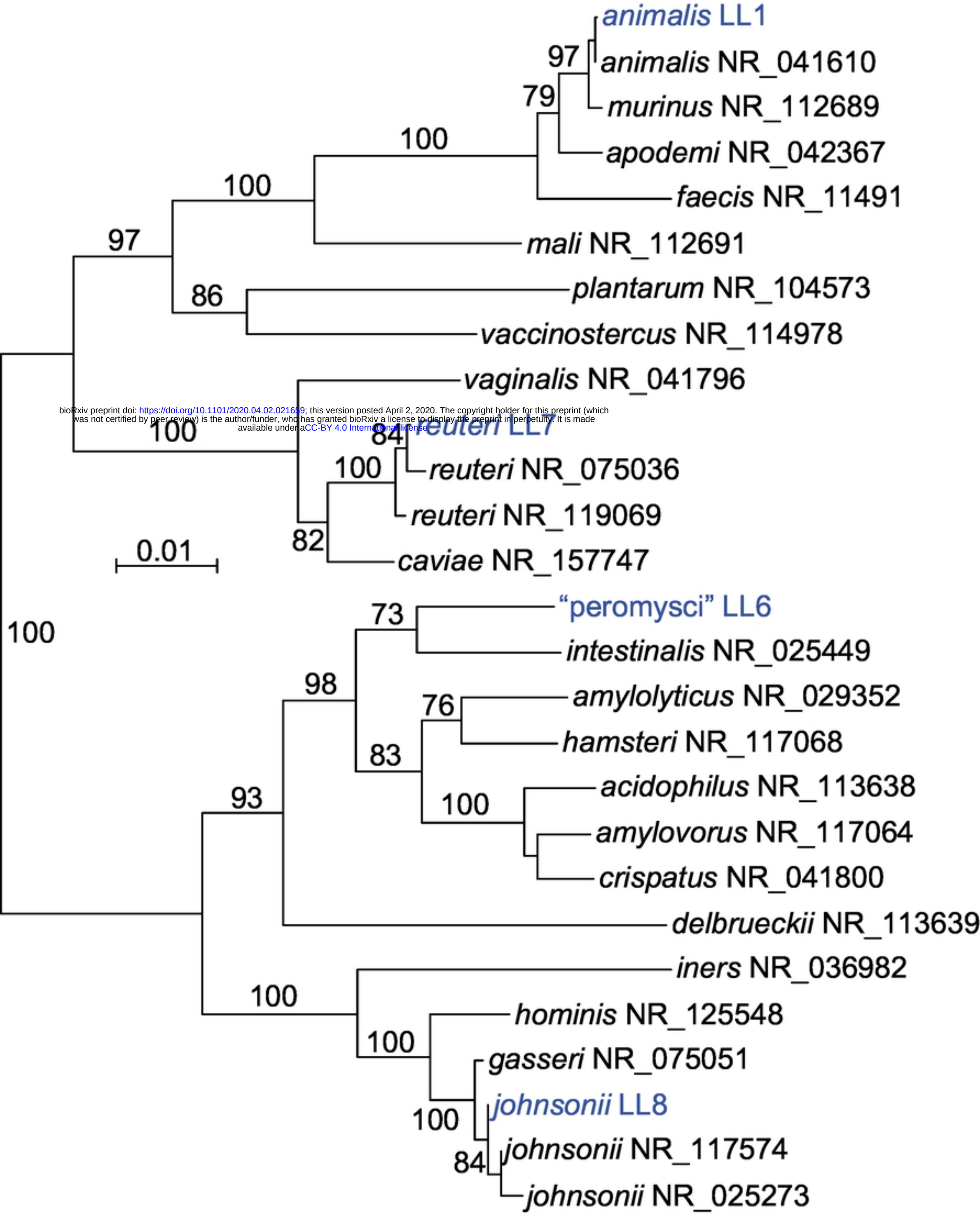


Figure 2

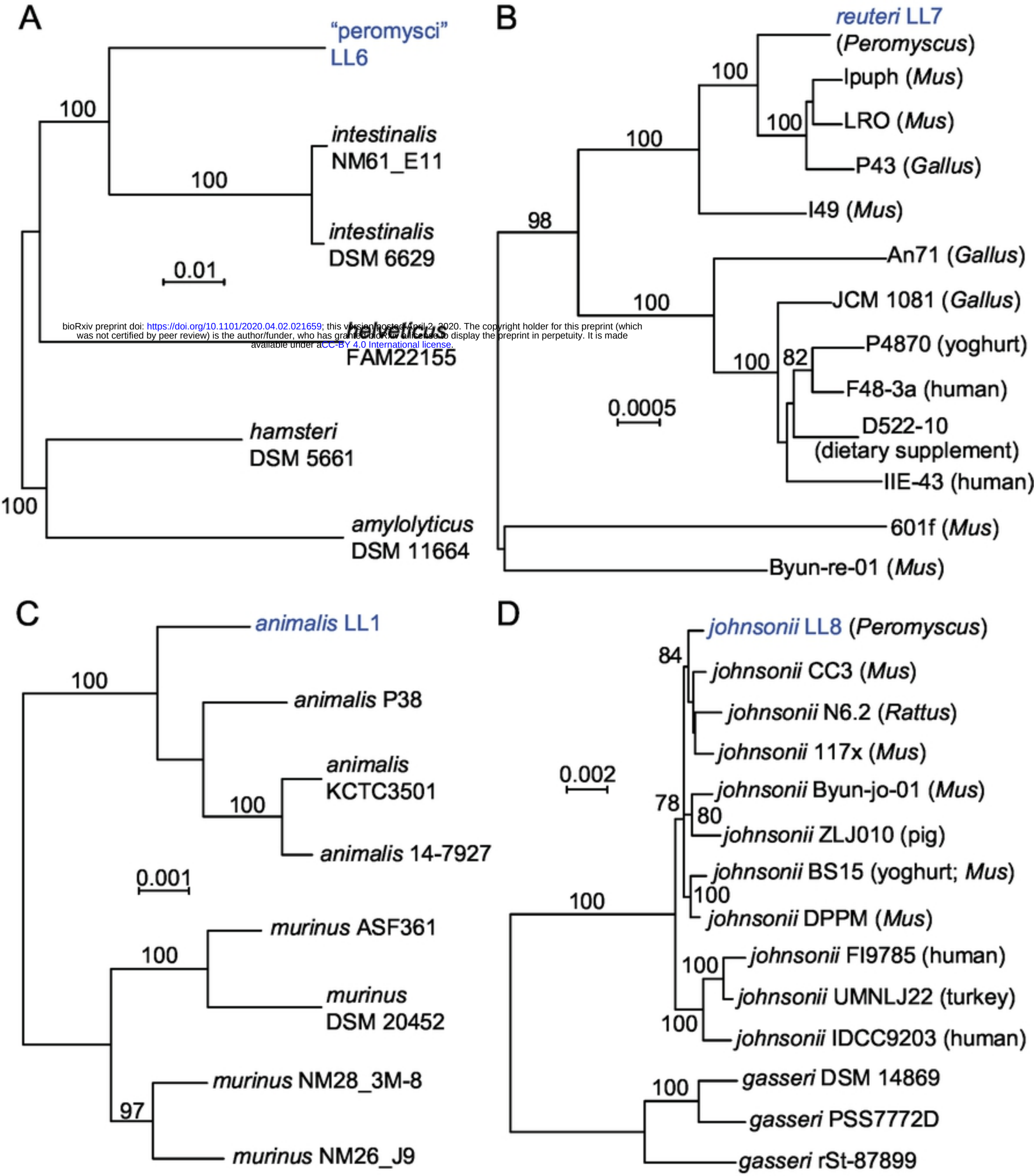


Figure 3

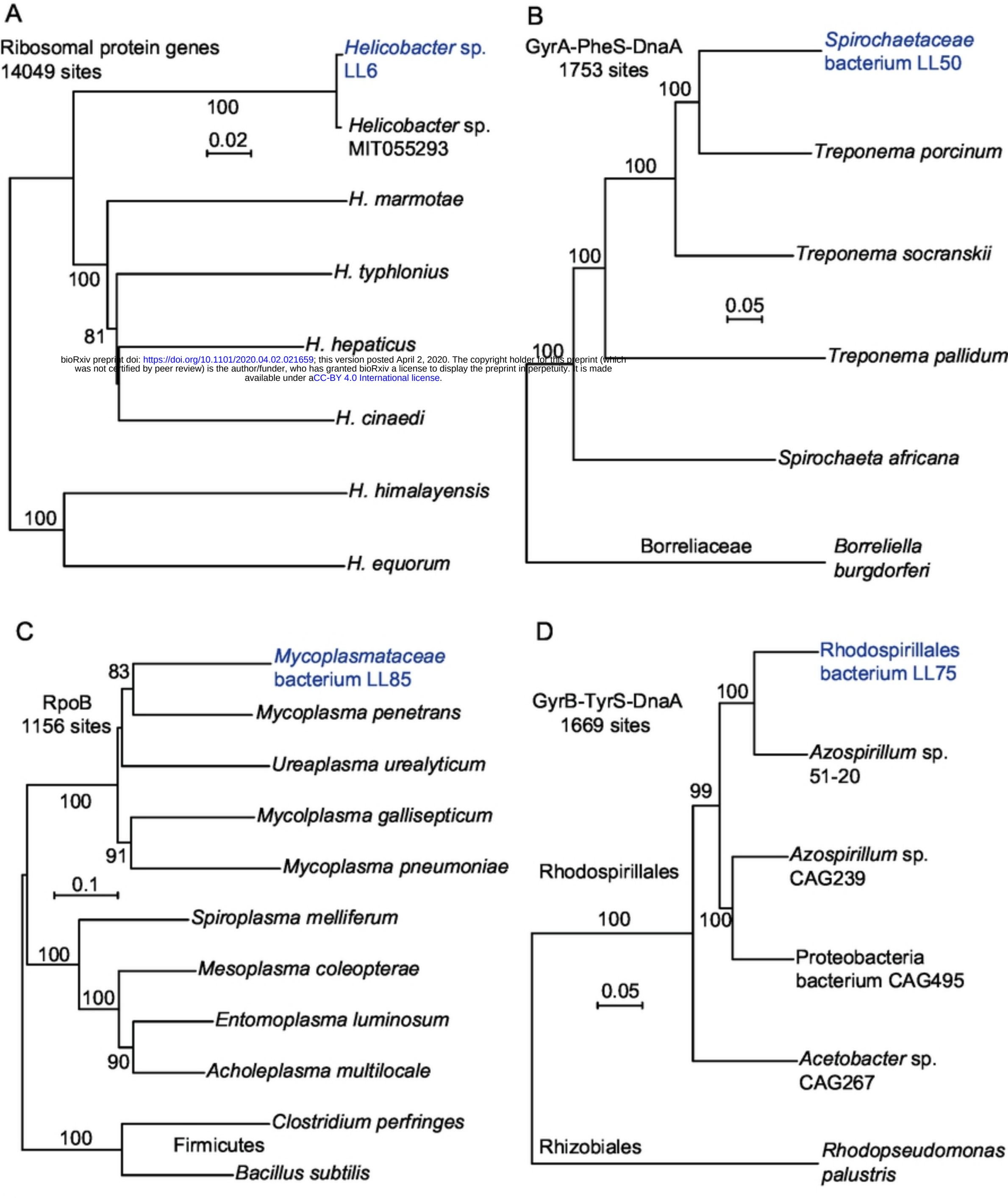


Figure 4

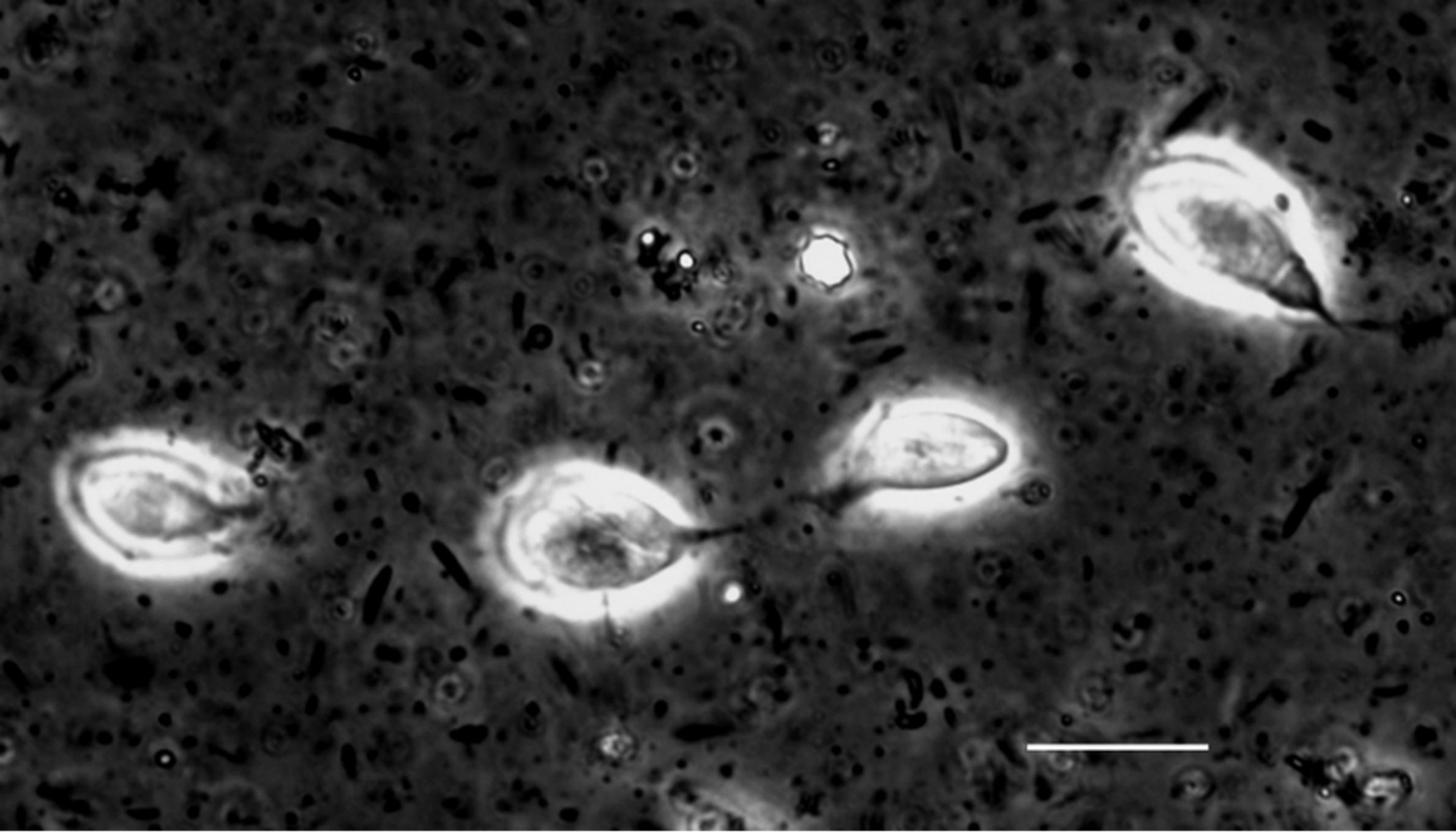


Figure 5

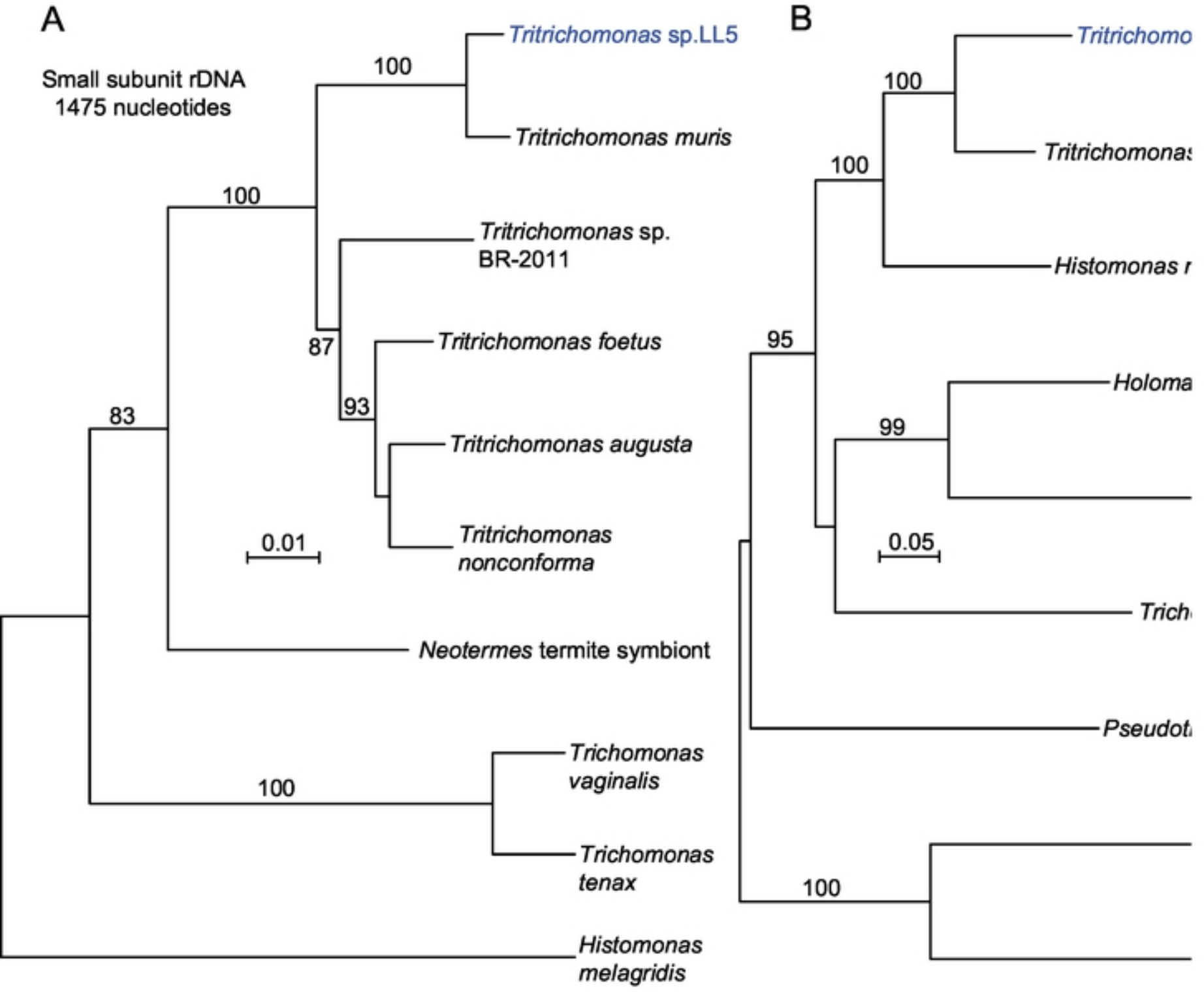

A
Small subunit rDNA
1475 nucleotides
Tritrichomonas sp.LL5
Tritrichomonas muris
Tritrichomonas sp.
BR-2011
Tritrichomonas foetus
Tritrichomonas augusta
Tritrichomonas
nonconforma
Neotermes termite symbiont
Trichomonas
vaginalis
Trichomonas
tenax
Histomonas
melagridis
100
100
87
93
83
100
0.01
B
Tritrichomo
Tritrichomonas
Histomonas
Holoma
Trich
Pseudot
100
100
95
99
0.05
100

Figure 6

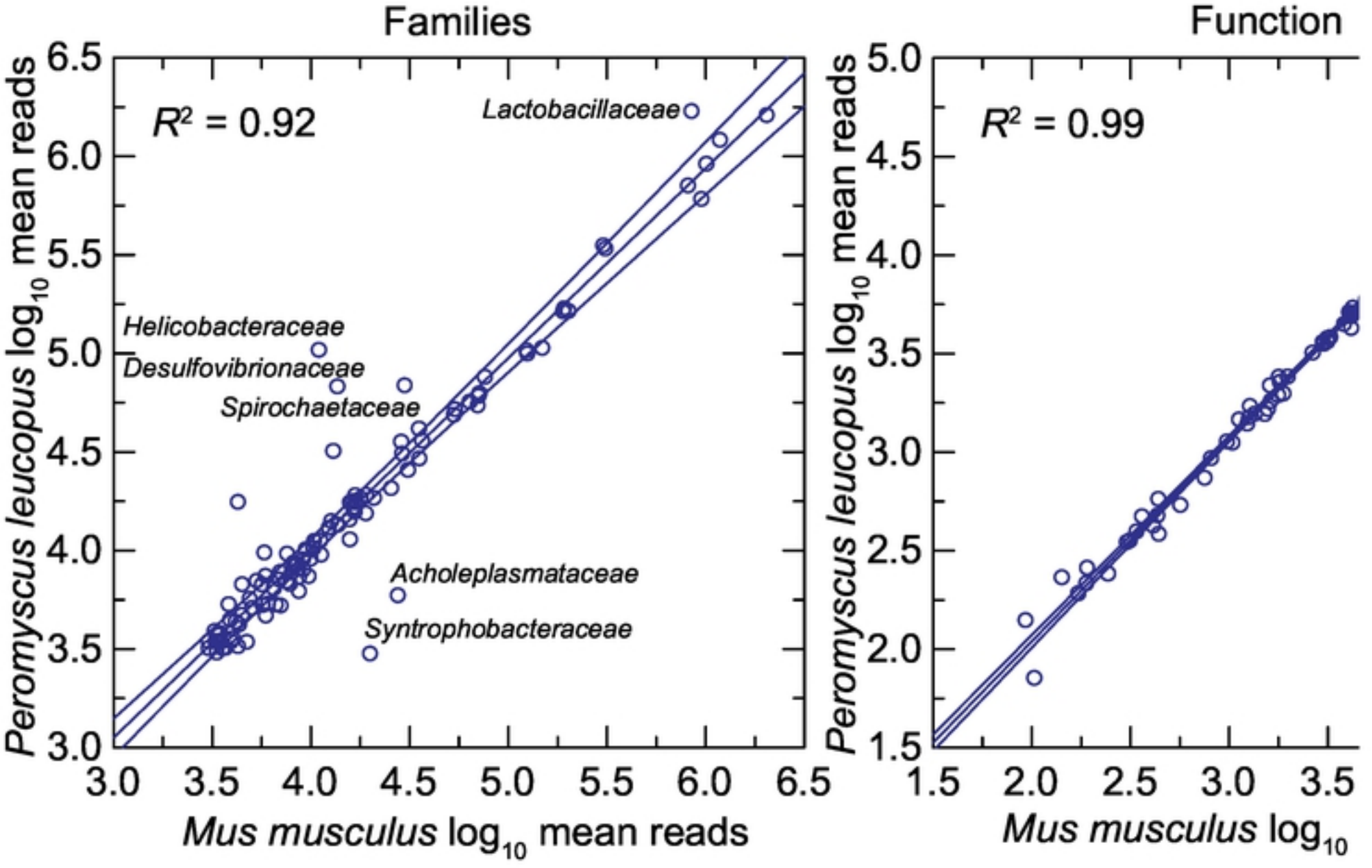

Families
Function
R^2 = 0.92
R^2 = 0.99
Lactobacillaceae
Helicobacteraceae
Desulfovibrionaceae
Spirochaetaceae
Acholeplasmataceae
Syntrophobacteraceae
Peromyscus leucopus log10 mean reads
Mus musculus log10 mean reads
Peromyscus leucopus log10 mean reads
Mus musculus log10
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5
1.5 2.0 2.5 3.0 3.5

Figure 7

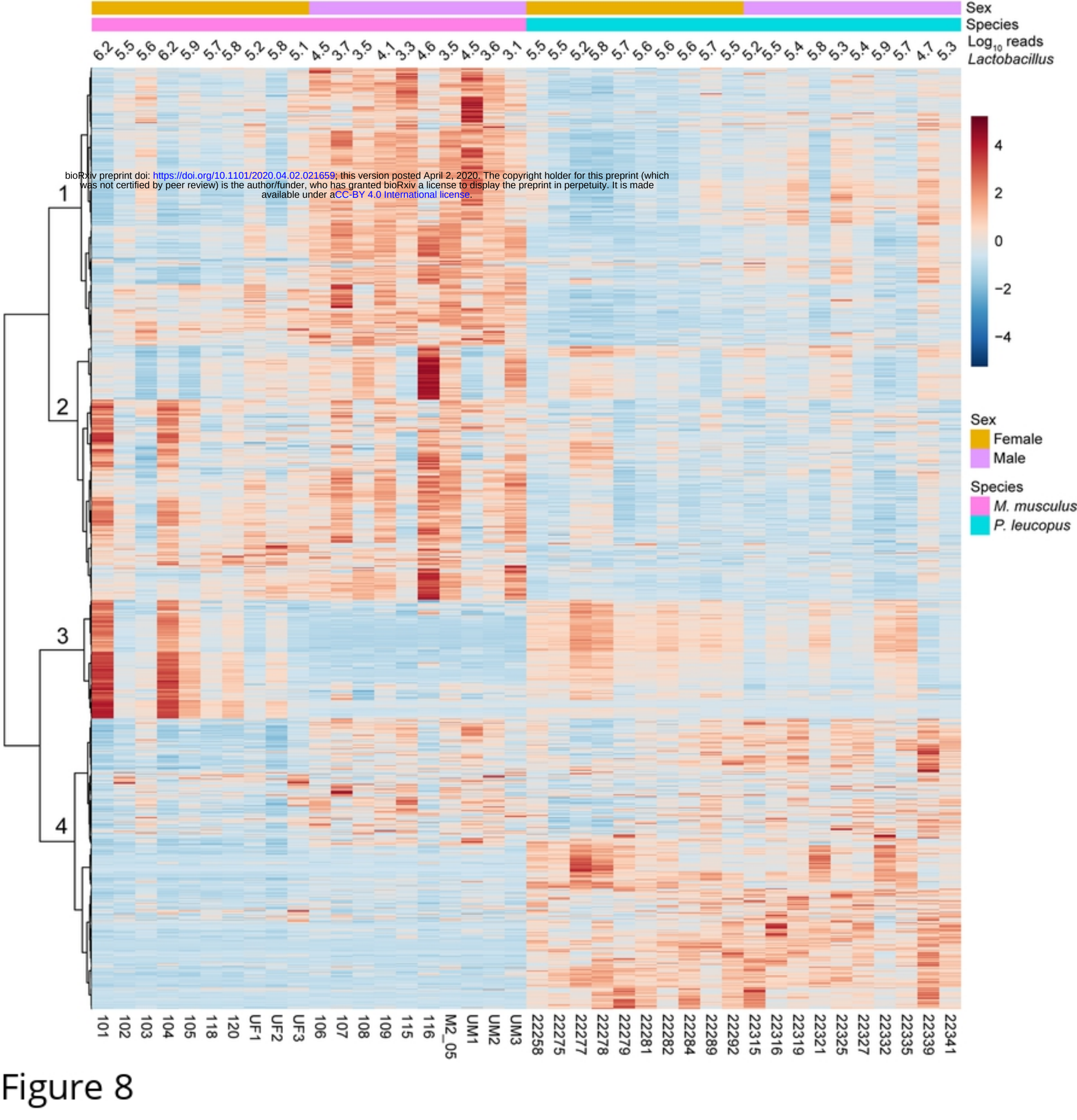


Figure 8

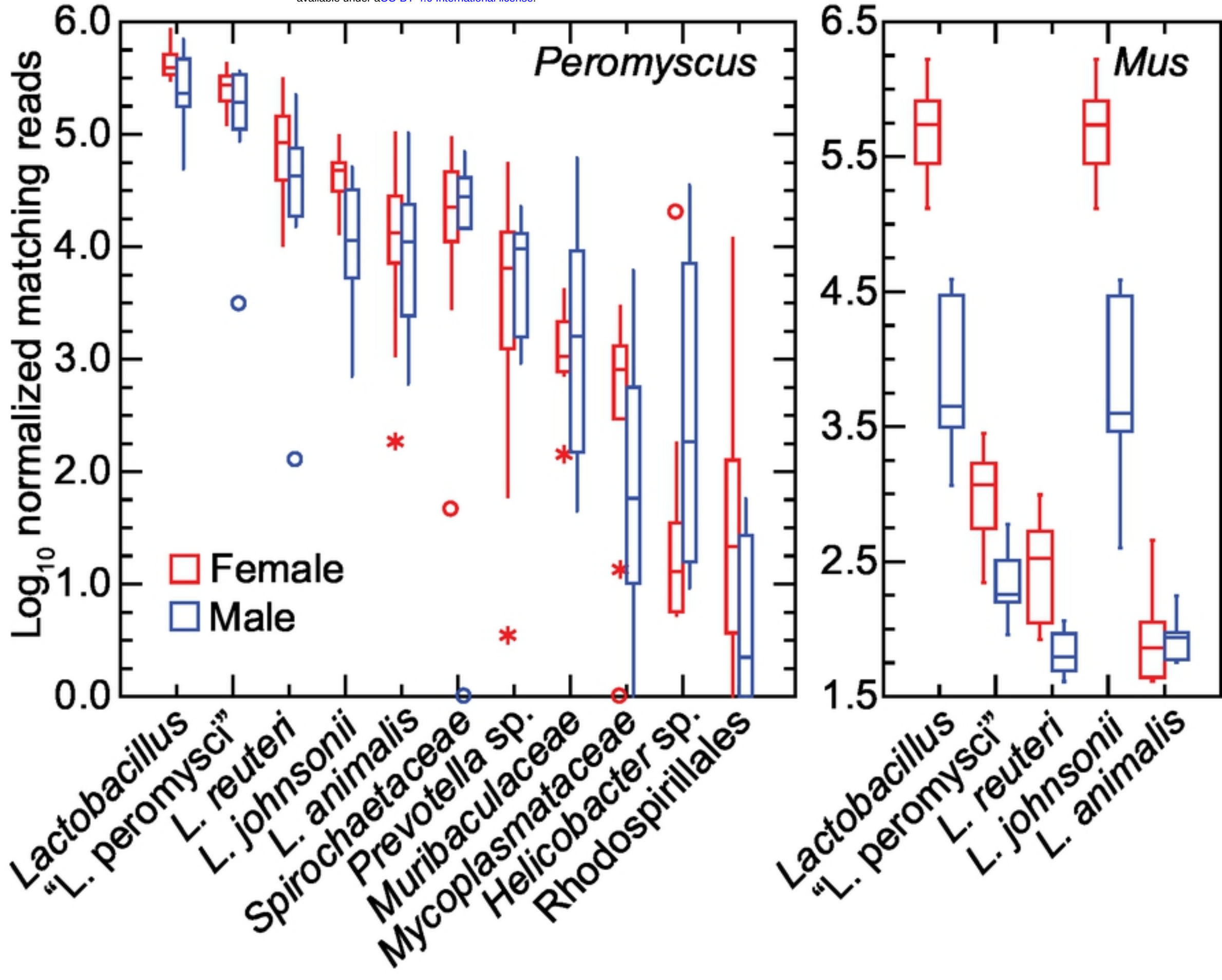


Figure 9

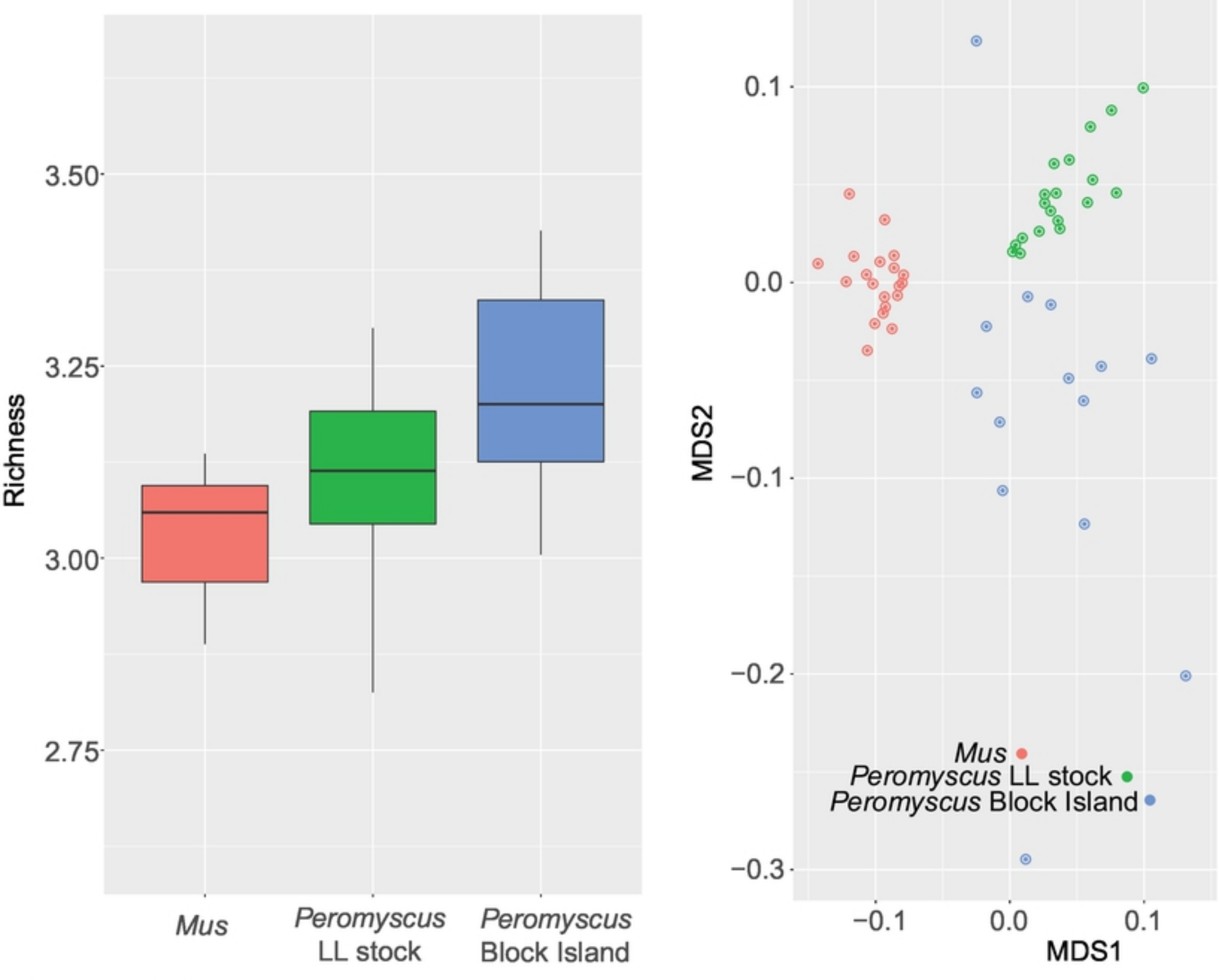

Richness
3.50
3.25
3.00
2.75
Mus
Peromyscus LL stock
Peromyscus Block Island
MDS2
0.1
0.0
−0.1
−0.2
−0.3
−0.1
0.0
0.1
MDS1
Mus
Peromyscus LL stock
Peromyscus Block Island

Figure 10

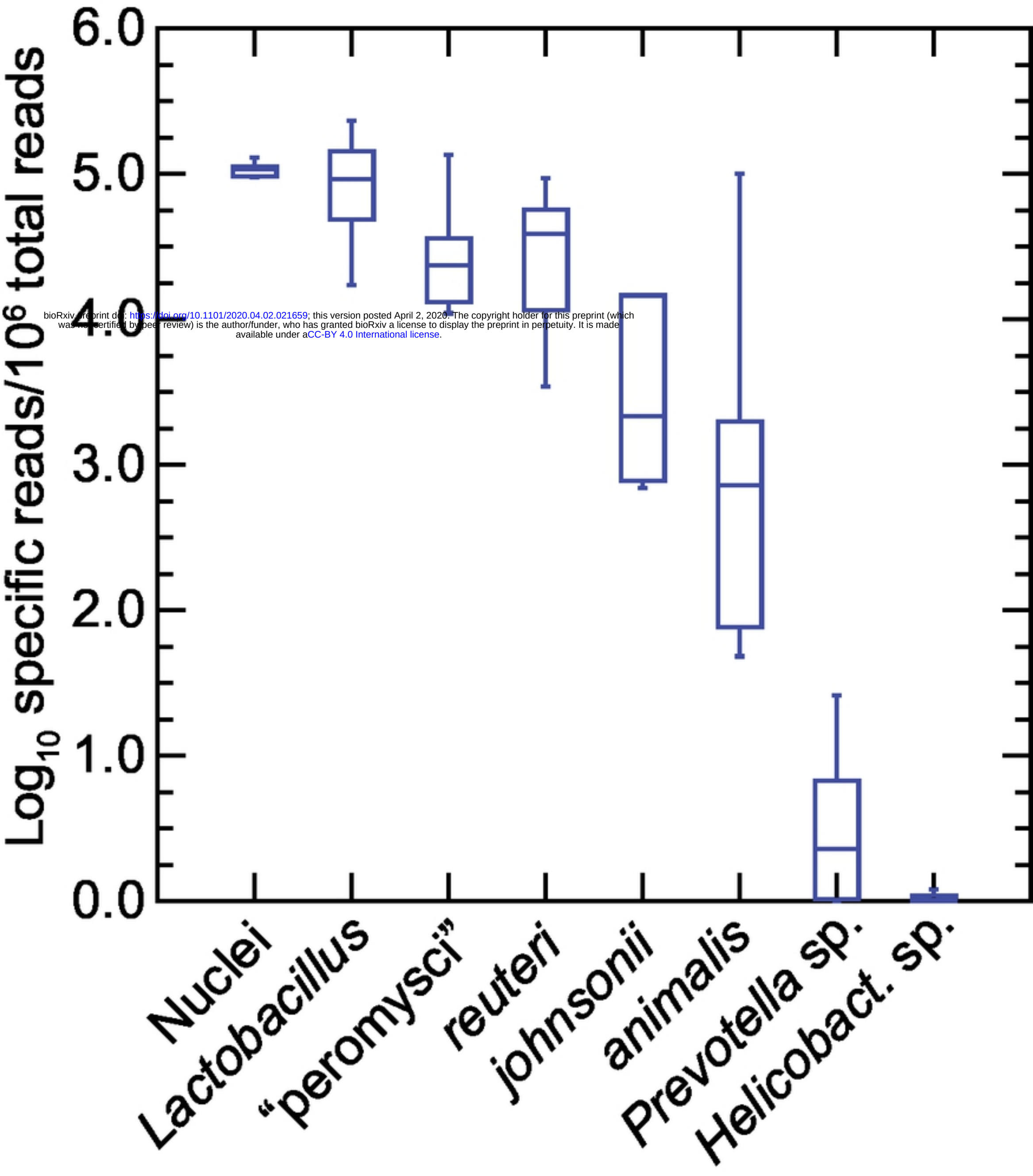


Figure 13

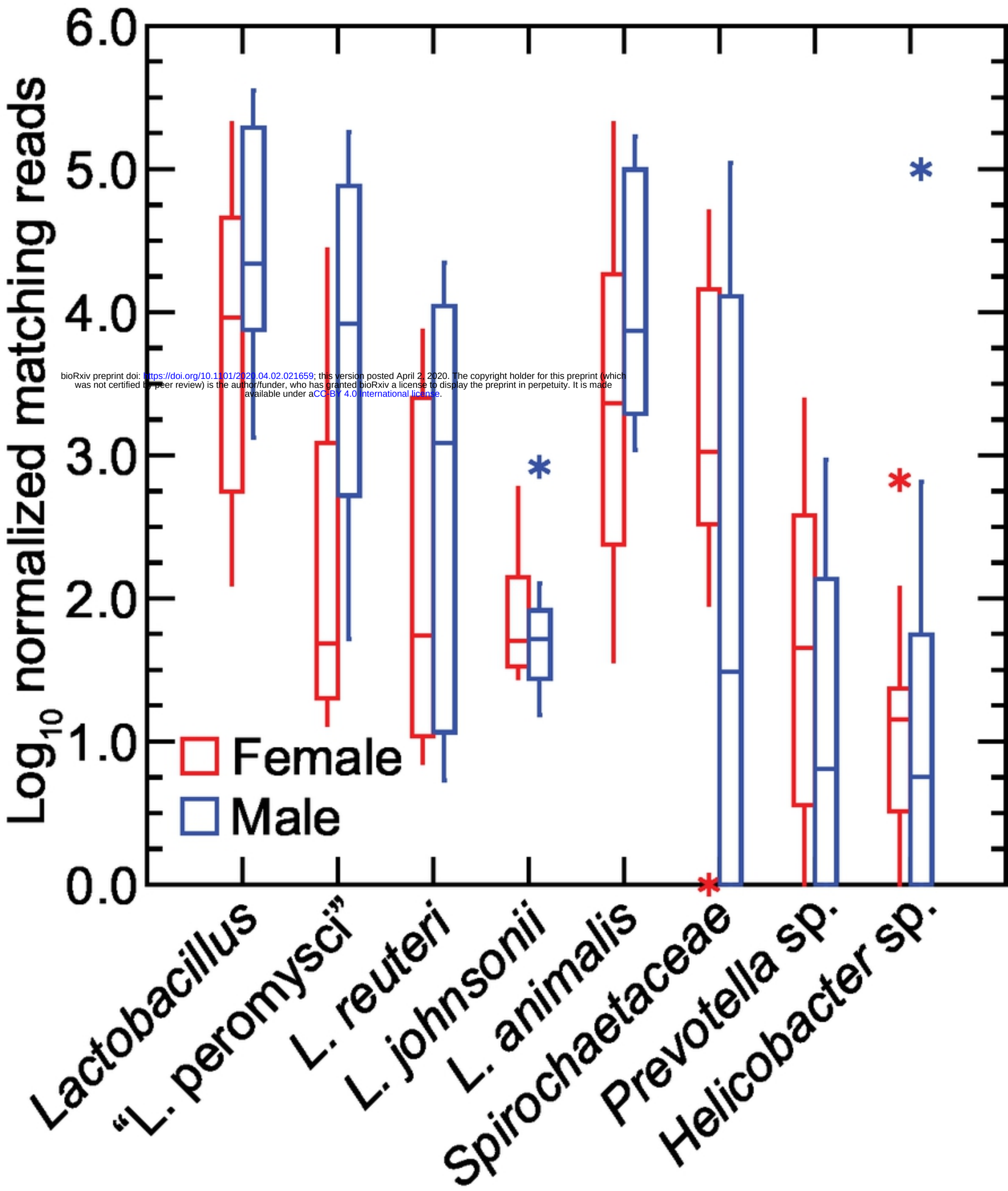


Figure 14

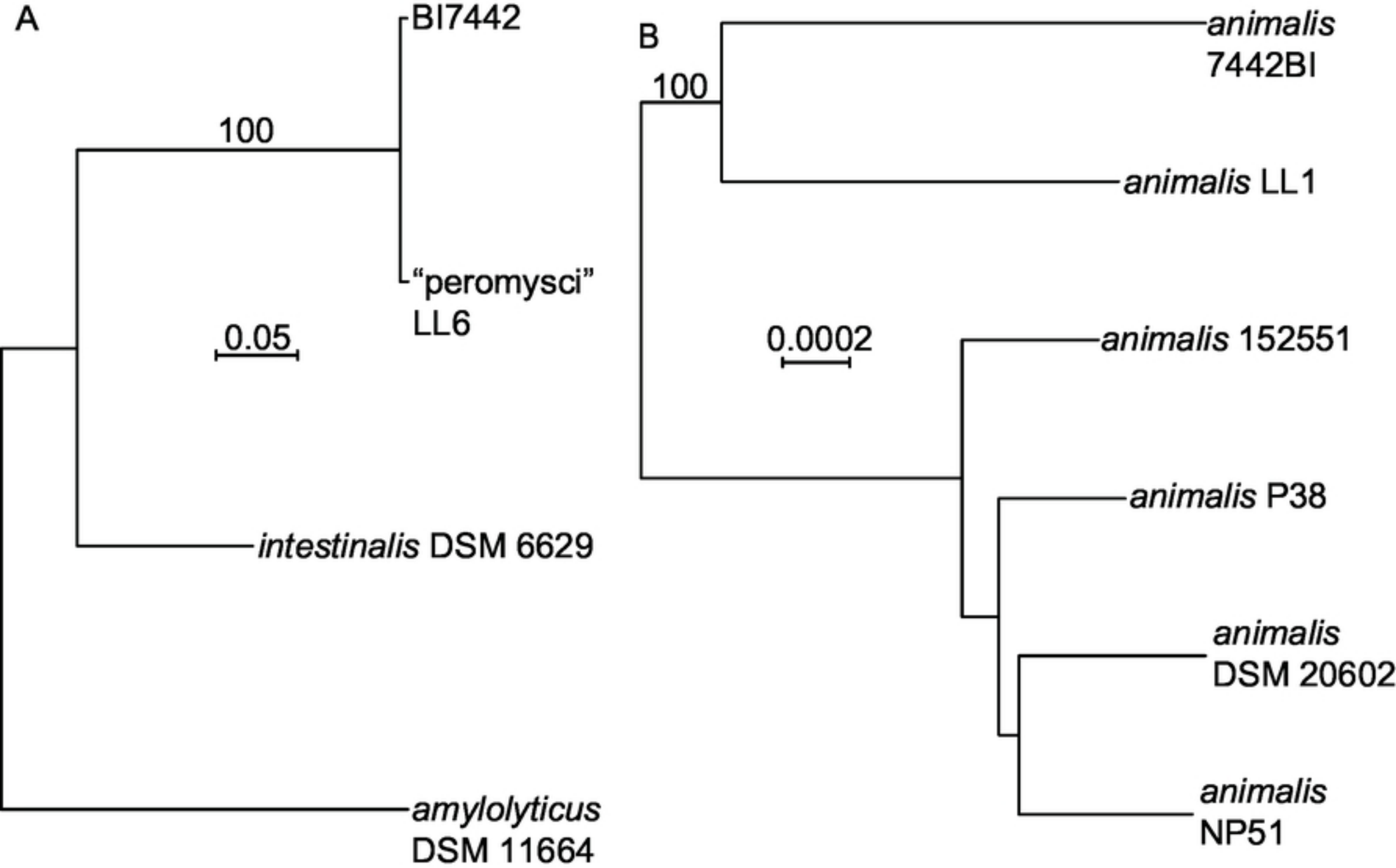

A
BI7442
100
"peromysci"
LL6
0.05
intestinalis DSM 6629
amylolyticus
DSM 11664
B
100
animalis
7442BI
animalis LL1
0.0002
animalis 152551
animalis P38
animalis
DSM 20602
animalis
NP51

Figure 15