TITLE: Species identity dominates over environment in shaping the microbiota of small mammals

AUTHORS: S. C. L. Knowles, R. M. Eccles, L. Baltrūnaitė

JOURNAL: Ecology Letters

PUBLISHER: Wiley

PUBLICATION DATE: 13 March 2019 (online)

DOI: https://doi.org/10.1111/ele.13240

Species identity dominates over environment in shaping the microbiota of small mammals

Knowles, S. C. L* [1,2], Eccles, R. M. [3] and Baltrunaite, L.[4]

[1] Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot, Berkshire, UK

[2] Department of Pathobiology and Population Sciences, The Royal Veterinary College, Hawkshead Lane, Hatfield, Herfordshire, AL9 7TA, UK.

[3] Centre for Genomic Research, Biosciences Building, Crown Street Liverpool, United Kingdom, L69 7ZB, UK.

[4] Nature Research Centre, Akademijos Str. 2, LT–08412 Vilnius, Lithuania

Statement of authorship: SCLK and LM collected the data, SCLK and RE performed laboratory work, SCLK performed analyses and wrote the manuscript, and all authors contributed to editing it.

Data accessibility statement: Raw sequence data from this work is available from the European Nucleotide Archive (doi: xxx-xxx-xx). Associated metadata and further information about the study can be found at the Open Science Framework project page, "Comparative gut microbiota of small mammals".

Running title: The mammalian gut microbiota: genetics vs environment

Keywords:
Microbiota, phylogenetic, mammal, rodent, microbiome, symbiont, community, cospeciation, 16S, Bacteroidales, codiversification

Article type: Letter

Abstract word count: 149

Main text word count: 5000

References: 76

5 figures, 1 table

*Author for correspondence *:*

Dr Sarah Knowles
The Royal Veterinary College
Hawkshead Lane
Hatfield
Hertfordshire, UK
AL9 7TA
Tel: +44 (0)1707 667061
Email: sknowles@rvc.ac.uk

Abstract

The mammalian gut microbiota is considered pivotal to host fitness, yet the determinants of community composition remain poorly understood. Laboratory studies show that environmental factors, particularly diet, are important, while comparative work emphasises host genetics. Here we compare the influence of host genetics and the environment on the microbiota of sympatric small mammals (mice, voles, shrews) across multiple habitats. While sharing a habitat caused some microbiota convergence, the influence of species identity dominated. In all three host genera, an individual's microbiota was more similar to conspecifics living elsewhere than to heterospecifics at the same site. Our results suggest this species-specificity arises in part through host-microbe codiversification. Stomach contents analysis suggested diet also shapes the microbiota, but where diet is itself influenced by species identity. In this way, we can reconcile the importance of both diet and genetics, while showing that species identity is the major factor shaping the microbiota.

Introduction

All animals have evolved in a bacterial world, and harbour a diverse community of microbial symbionts colonising internal and external surfaces (McFall-Ngai *et al.* 2013). The mammalian gut houses a particularly dense and diverse community of microbes that performs many important functions for the host. These include the provision of otherwise inaccessible nutrients from food (Rosenbaum *et al.* 2016), protection from pathogenic infections (Buffie & Pamer 2013), and detoxification of poisonous compounds (Kohl *et al.* 2014). Despite this, we are only just beginning to understand the processes shaping the composition of host-associated microbial communities over evolutionary and ecological timescales (Foster *et al.* 2017) .

77 Both controlled experiments in laboratory animals and human studies have shown that
78 environmental factors can strongly affect gut microbiota composition. In particular, diet is a
79 major influence, with both short-term diet shifts and long-term dietary habits affecting these
80 communities (David *et al.* 2014; Carmody *et al.* 2015; Sonnenburg *et al.* 2016; Griffin *et al.*
81 2017). A host's social and physical environment is also important. When mice are cohoused,
82 their microbiota composition converges (Hildebrand *et al.* 2013; Seedorf *et al.* 2014; Griffin *et*
83 *al.* 2017), and cohabiting, unrelated humans are more similar in their gut microbiota than those
84 living apart (Song *et al.* 2013). Strong environmental effects have also been reported in studies
85 of wild animals, including seasonal and habitat differences (Maurice *et al.* 2015; Amato *et al.*
86 2016; Ren *et al.* 2017). While genetic effects on the gut microbiota have been detected in
87 laboratory and human studies (Wang *et al.* 2018), these are often rather weak, and within-
88 species studies typically emphasize the strong influence of environmental factors, such as diet
89 (Carmody *et al.* 2015; Rothschild *et al.* 2018; Weissbrod *et al.* 2018).
90
91 In parallel, a growing number of phylogenetic studies have shown the importance of host
92 genetics in shaping the microbiota. These have found either that microbiota composition
93 recapitulates the host phylogeny (known as 'phylosymbiosis'), or shows species-specificity,
94 with that of conspecifics being more similar than that of heterospecifics. Although not
95 universally detected (Dietrich *et al.* 2014; Sanders *et al.* 2014; Baxter *et al.* 2015; Martinson
96 *et al.* 2017), such host phylogenetic effects have been found in a diverse range of taxa,
97 including mammals, insects and birds (Ochman *et al.* 2010; Phillips *et al.* 2012; Brooks *et al.*
98 2016; Amato *et al.* 2018; Nishida & Ochman 2018). Furthermore, recent work has provided
99 evidence for cospeciation among mammals and their gut microbes (Moeller *et al.* 2016). While
100 these findings suggest an important role for host genetics, a challenge is that in wild settings
101 such patterns can have a range of explanations, including environmental ones. In particular,
102 a major confound is that different species often occur in different habitats, such that
103 phylogenetic patterns may be driven by environmental ones (Brooks *et al.* 2016; Groussin *et*
104 *al.* 2017).

A major open question, therefore, is whether host genetics or a shared environment dominates in shaping the microbiota. Answering this requires the effects of habitat and host genetics to be disentangled in a natural setting. To do this, we performed a cross-factorial comparison, characterising the microbiota of individuals from multiple species within each of three widespread small mammal genera (*Apodemus* mice, *Microtus* voles, and *Sorex* shrews) across the same set of five contrasting habitats. In this way, we are able to test whether a shared evolutionary history (belonging to the same species) or instead a shared environment (being in the same habitat) dominates in determining gut microbiota composition.

Materials and Methods

Trapping, sample collection and diet analysis

Trapping took place between 14th and 27th August 2014 at five sites (BG, CC, LM, PL, LM, WF) within 3 to 23km of each other, near Vilnius in Lithuania (Fig. S1). Sites represented contrasting habitats where we expected to trap multiple species from three common genera – *Apodemus* (mice), *Microtus* (vole) and *Sorex* (shrew). They were far enough apart that animals should not regularly move between sites, but not so distant as to introduce major within-species genetic differentiation, which could confound habitat-related microbiota differences. The species studied have small home ranges, with the widest ranging (*Apodemus* spp.) rarely moving more than 0.25km (Andreassen *et al.* 1998; Wang & Grimm 2007; Yletyinen & Norrdahl 2008; Stradiotto *et al.* 2009) such that animals should not have moved between sites. Genetic differentiation is also not expected to be strong at this spatial scale (Gauffre *et al.* 2008). Snap traps baited with bread soaked in oil were set at dusk for 2-3 nights per site, and retrieved the next morning. Animals were placed in sterile bags and kept on ice for immediate transport to the lab for dissection. Animals were keyed to species using morphological characteristics, and age (juvenile, sub-adult, adult), sex, body mass and reproductive status were recorded (Supplementary Information). To explore the role of dietary

differences in driving microbiota differences, we examined each individual's stomach contents. Stomach contents were inspected under a dissecting microscope to determine the relative abundance of broad dietary categories (e.g. seed, vegetative parts, insect, fungi; Fig. S2). An approximately 10mm section of the distal colon (in rodents) or simple gut (in shrews) was removed for microbiota characterisation. The contents were placed in RNALater™ and refrigerated at the end of each day. Because shrews degrade more quickly post mortem, shrews were dissected before rodents. Utensils were cleaned thoroughly with 70% ethanol and flamed between dissections. At the end of fieldwork, samples were spun down, RNALater™ removed and samples were stored at -80^{o}C. Five months later, samples were transported frozen to the UK and stored at -80^{o}C before DNA extraction. To test how lethal trapping might have affected microbiota composition, we performed a limited amount of live-trapping on three nights at two sites (PL and WF), using small Sherman traps (2 x 2.5 x 6.5") baited with grain, carrot and bedding. Animals were transported to the Nature Research Center, where they were humanely killed by cervical dislocation, and immediately dissected to take gut content samples, which were stored and processed exactly as described above.

16S rRNA gene sequencing

Genomic DNA was extracted from gut content samples using the MoBio™ PowerSoil kit, according to manufacturer's instructions. The V4 region of the bacterial 16S rRNA gene was amplified using primers 515F/806R (Caporaso *et al.* 2011), with library preparations following a two-step (tailed-tag) approach with dual-indexing (D'Amore *et al.* 2016). Primer sequences are given in Table S1. Amplicon libraries were sequenced on an Illumina® MiSeq with 250bp paired-end reads. Full details of sequencing methodology are in Supplementary Information.

Bioinformatic processing

Sequence data was processed through the DADA2 pipeline (v1.4) in R to infer amplicon sequence variants (ASVs) (Callahan *et al.* 2016a, 2017) (Supplementary Information). Briefly, reads were trimmed and filtered for quality, ASVs inferred, putative chimeras removed and

taxonomy assigned using the 13.8 Greengenes database clustered at 97% identity. A *phyloseq* object (McMurdie & Holmes 2013) was created for further processing and analysis. ASVs taxonomically assigned as chloroplast or mitochondria were removed, as well as those (1.3% ASVs) where a phylum was not assigned, after which the dataset contained 18,402 ASVs. The R package iNEXT (Chao *et al.* 2014; Hsieh *et al.* 2016) was used to create sample completeness and rarefaction curves. Sample completeness plateaued by approximately 10,000 reads (Fig. S3), such that all samples except one (with 26 reads) were retained, spanning a read count (before further filtering for beta diversity analyses, see below) of 11,794 to 931,354.

Statistical analyses

All analyses were carried out in R version 3.4.3 (R Core Team 2017). Since the 14 samples from live-caught animals did not cluster strongly within host genera (Fig. S4), these were pooled with the 211 other samples during analysis. We confirmed that capture method did not strongly influence community composition in multivariate PERMANOVAs (see below).

Alpha diversity analyses

For alpha diversity analyses, filtering was limited to the removal of ASVs assigned as chloroplast, mitochondria or with phylum unassigned. We used additive diversity partitioning (Crist *et al.* 2003) to ask at what level bacterial diversity arose – was the greatest turnover in ASV richness seen when sampling a new host species or family, or was the majority already present within species, with only relative abundances changing at higher taxonomic ranks? This method partitions total diversity (γ diversity) into that occurring at the within-individual (α diversity) and subsequent hierarchical levels – between individuals, species, genera and families (β diversities). We used the *adipart* function (Crist *et al.* 2003) in package *vegan* (Oksanen *et al.* 2017) to do this, using asymptotic estimates of ASV richness per sample calculated in package iNEXT as the response.

187

188 *Beta diversity analyses*

189 For beta diversity analyses, further (abundance) filtering was performed by only retaining
190 ASVs with more than 1 copy in at least 5% of samples, to remove potential contaminants and
191 sequencing artefacts. This resulted in a dataset containing 2474 ASVs, with sample read
192 count ranging from 9,291 to 72,1783. We also tested a more permissive abundance filter,
193 retaining ASVs with more than 1 copy in at least 3 samples, leading to a total of 8005 ASVs.
194 Since results were very similar and conclusions unchanged, results using this alternative filter
195 are not reported further. A phylogenetic tree was constructed from ASVs using the method
196 described by Callahan *et al.* (Callahan *et al.* 2016b), and read counts were normalised using
197 cumulative-sum scaling (CSS) in the *MetagenomeSeq* package (Paulson *et al.* 2013).
198 Pairwise dissimilarities were calculated using four beta-diversity metrics (Jaccard distance,
199 Bray-Curtis dissimilarity, weighted and unweighted UniFrac distances) in packages *vegan* and
200 *phyloseq*, and used in principle coordinates analysis (PCoA). To examine the relative extent
201 to which species and capture site predicted microbiota composition in each host genus, four
202 analytical approaches were used: (1) Hierarchical clustering to visualise whether microbiota
203 samples predominantly clustered by species or site (2) permutational analysis of variance
204 (PERMANOVA) (3) comparisons of mean dissimilarity values between pairs of samples
205 according to whether they belonged to the same species and/or were captured at the same
206 site and (4) Random Forest Classifier (RFC) models, assessing how accurately samples could
207 be assigned to species and capture site respectively.

208

209 *Hierarchical clustering*

210 Hierarchical clustering was performed with the UPGMA algorithm using *hclust* in R. Trees
211 were visualised using packages *ape* (Paradis *et al.* 2004) and *dendextend* (Galili 2015).

212

213 *PERMANOVAs*

PERMANOVAs were performed using the *adonis* function in package *vegan,* with 10,000 permutations. Since *adonis* tests terms sequentially (a term's explanatory power depends on what is fitted before it), univariate models including either species or site were constructed to compare the variance explained by each, with extraction batch (15 levels) as a blocking factor. Subsequently, to explore the influence of other variables, models were constructed including species, site, age, sex and reproductive status (4-levels: reproductive male, non-reproductive male, pregnant female, non-pregnant female), a linear term for body mass and several methodological variables: sequencing run, raw read count (linear term), capture method, and the maximum time interval between trap collection and dissection (linear term). Dispersion tests using function *betadisper* were performed to assess whether significant species or site effects could be influenced by differences in group dispersion (Anderson 2001).

Permutation tests on pairwise dissimilarity metrics

We tested whether mean community dissimilarity values differed according to whether or not individuals belonged to the same species or came from the same site, using Monte Carlo permutations of category labels to generate null distributions of dissimilarity values appropriate to each comparison (Sanders *et al.* 2014). We used 1-tailed p-values, as there is an *a priori* expectation that animals belonging to the same taxon or present at the same site, should be more similar than those from different taxa or sites.

Random Forest Classifier models

A Random Forest Classifier (RFC) supervised learning algorithm was implemented in package *randomForest,* to classify microbiota samples according to either host species or capture site (Breiman 2001; Knights *et al.* 2011). Models were run on CSS-normalised ASV counts with 100,000 trees, and the out-of-bag error rate used as a measure of classification accuracy. We also used cross-validation to assess the performance of models created using 70% of the data as applied to the remaining 30%, though results were extremely similar to out-of-bag error estimates and are not reported further. To establish which ASVs were most important in driving

species distinguishability, we examined their importance scores (Mean Decrease Gini) in RFC models, and the taxonomic distribution of the most important ASVs relative to all ASVs identified.

Variability in strength of the species signal

Evidence from other mammalian groups suggests some gut microbes coexist stably and cospeciate with their hosts (Moeller *et al.* 2016). Because host speciation events are recent on the scale of bacterial phylogenies, this should result in sister host species containing sister symbiont lineages differing largely at a fine taxonomic scale. A corollary is that differences between symbiotic communities arising through cospeciation should decay at broader bacterial phylogenetic scales (Sanders *et al.* 2014). To test for this pattern, we assessed how sensitive the host species signal was to the level of bacterial phylogenetic or taxonomic resolution used. We used the *tip_glom* function in *phyloseq* to group bacterial sequences into OTUs with progressively lower phylogenetic resolution, and the *tax_glom* function to group bacterial ASVs at the Family, Order, Class or Phylum level (using the subset of ASVs assigned to at least Family level). We then examined how this affected species-distinguishability within each host genus, as represented by either R^2 for the species term in a univariate PERMANOVA based on Bray-Curtis dissimilarity, or the out-of-bag error rate for species classification in RFC models. We also examined how species distinguishability varied among the four dissimilarity metrics used (in PERMANOVA analyses and PCoA plots), which differ in the extent to which they account for phylogenetic relatedness among ASVs.

Analysis of diet and microbiota composition

Diet could vary as a result of phylogenetic effects (host species have evolved different dietary preferences) or environmental effects (hosts eat different things in different habitats), such that dietary variation could contribute to microbiota differences across species, capture sites, or both. Therefore, for each host genus we examined variation in diet according to species and capture site, and whether diet similarity predicted microbiota similarity. Bray-Curtis

dissimilarity was calculated from proportional stomach contents data, for individuals with both microbiota and diet data (n=215). We used permutation tests identical in format to those described above for analysing the microbiota, to assess pairwise differences among individuals in diet according to species and site. For each host genus, we used Mantel tests in *vegan* to assess whether diet composition predicted microbiota composition (Bray-Curtis dissimilarity).

Results

We characterised the gut microbiota from ten species of mouse, vole and shrew captured at 5 sites in Lithuania (225 individuals, Table S2, Fig. S1). The majority of species were captured in all five habitats providing a large number of sympatric and allopatric comparisons, both within and across species, to evaluate drivers of gut microbiota composition.

The gut microbiota differs strongly among small mammal clades

Analysis of the full dataset showed that gut microbiota communities were clearly differentiated among the three host families – mice, voles and shrews (Murinae, Cricetidae and Soricidae). Principle coordinates analysis on both Bray-Curtis dissimilarities (Fig. 1A) and Unweighted UniFrac distances (Fig. S5) revealed clear clustering of samples by host family (PERMANOVA on Bray-Curtis dissimilarity, host family $F_{2,224}$=59.8, p=0.001, R^2=0.35). Indeed, broad differences in microbiota composition were evident in the relative abundance of bacterial phyla across host families (Fig. 1B). In rodents, the Bacteroidetes and Firmicutes phyla dominated, but voles tended to have higher relative abundance of Tenericutes and Spirochaetes than mice (Fig. 1B). The microbiota of common shrews (*Sorex araneus*) was often dominated by Proteobacteria, whereas in pygmy shrews (*Sorex minutus*) the Firmicutes were more dominant (Fig. 1B). In the rodent families where we sampled multiple genera, community composition was also structured by host genus (Fig. 1A, PERMANOVA on Bray-Curtis dissimilarity for host genus: mice $F_{1,67}$=19.7, p=0.001, R^2=0.23; voles $F_{1,111}$=25.8, p=0.001, R^2=0.19). RFC models also classified samples to host family or genus with 100% accuracy on

298 the basis of ASVs. Moreover, this signal remained strong even when higher bacterial
299 taxonomic units were used for classification; samples could be classified to host family 99.1%
300 of the time using bacterial families and 98.2% using bacterial phyla, with similar results for
301 classification to host genus (92.4% for family-level and 96.7% for phylum-level models
302 respectively). Thus, the gut microbiota composition of mice, voles and shrews found across
303 the same set of habitats is distinct even at the level of bacterial phyla.
304
305 As well as being compositionally different, microbiota diversity also varied across host
306 families, with voles (especially *Microtus*) having approximately double the richness and
307 Shannon diversity of mice and shrews (Fig. 1C), consistent with their more herbivorous diet
308 (Ley *et al.* 2008; Nishida & Ochman 2018). Across the total dataset, most diversity (42%
309 bacterial sequences) arose at the between-individual level. More than half the bacterial
310 diversity (59%) was observed within species, with the remainder at higher taxonomic levels
311 (9% between species, 12% between genera, 20% across host families, Table 1). Within the
312 three genera where multiple species were sampled (*Apodemus*, *Microtus* and *Sorex*), the
313 greatest proportion of richness again occurred at the between-individual level (51-57%), with
314 less (8-21%) arising across species. These results suggest that while some bacteria are
315 specific to a particular host species, genus or family, the majority of turnover in bacterial
316 diversity is seen across individuals, indicating these communities are highly individualized as
317 reported for the human microbiota (Ley *et al.* 2006; Faith *et al.* 2013).
318
319 ***Within host genera, the microbiota is shaped more strongly by species than capture***
320 ***site***
321 Both species identity and environment (capture site) shaped gut microbial communities within
322 each genus. However, across multiple analyses, species identity dominated. First, hierarchical
323 clustering according to Bray-Curtis dissimilarity showed that samples grouped primarily by
324 host species, with less prominent clustering by capture site that occurred largely within species
325 (Fig. 2A-C). Similar patterns were seen for the Jaccard and Unweighted UniFrac distances,

though clustering by species was less apparent using Weighted UniFrac (Fig. S6). Second, PCoA plots based on Bray-Curtis dissimilarity showed clear sample clustering by host species, but less so by capture site (Fig. 2D-F). Third, mean pairwise Bray-Curtis dissimilarity and Jaccard distance among samples was greater when comparing samples from different species than samples from different sites (Fig. 3A-C, Table S3). Most definitively, in all host genera an animal's microbiota composition was on average more similar to a conspecific caught elsewhere, than a heterospecific caught at the same site (Fig. 3A-C, Table S3). Fourth, Random Forest Classifier (RFC) models classified gut microbial communities from congeneric animals to host species with almost perfect accuracy (classification accuracy: *Apodemus* 100%, *Microtus* 98.8%, *Sorex* 97.7%) while classification accuracy according to capture site was poor (*Apodemus* 47.4%, *Microtus* 51.2%, *Sorex* 22.7%). Finally, univariate PERMANOVAs showed stronger effects of species than capture site (Table S4).

Taken together, these results indicate that species identity dominated over capture site in shaping gut microbiota beta diversity among congeneric small mammals. However, shared environment does play some role. Capture site explained a significant proportion of variance (9-13%) in all PERMANOVA models, alongside weaker effects of host age and methodological variables (Table S4). In the two rodent genera, mean pairwise Bray-Curtis dissimilarity among both con- and heterospecific individuals was also significantly lower when they were caught at the same site compared to different sites, indicating microbiota convergence when living in sympatry. Site effects were in the same direction but non-significant for shrews (Fig. 3A-C, Table S3). We even detected minor environmental convergence in the microbiota of animals from more distantly related groups; the microbiota of mice from the genera *Apodemus* and *Micromys* converged in sympatry (p=0.003), though we did not find such evidence for voles (*Microtus* vs *Myodes*, p=0.176).

Species-indicative microbial taxa

353 Our results suggest that each host species has a characteristic microbiota signature that
354 transcends the habitat they are in, and the other closely related species they mix with. But
355 which symbiont taxa are responsible for this? RFC models indicated that many of the top 20
356 most important sequence variants driving species distinguishability in rodents belonged to the
357 order Bacteroidales (90% for *Apodemus*, 100% for *Microtus*), and the majority (17/20 in both
358 cases) to one particular family within this order, S24-7. The family S24-7 was strongly over-
359 represented among species-indicative ASVs compared to all ASVs in the dataset, whereas
360 other common families including Lachnospiraceae and Rumminococcaceae were under-
361 represented (Fig. 4). Both S24-7 and its parent order Bacteroidales were also suggested to
362 be important for species distinguishability using other metrics, including the proportion of ASVs
363 in each taxon that were host species-specific (Fig. S7). Removal of S24-7 from the dataset
364 notably decreased the accuracy of RFC models in classifying *Microtus* samples to host
365 species (98.8% including vs 80.23% excluding this family), but classification accuracy
366 remained 100% for *Apodemus*, suggesting other species-indicative bacteria are also
367 important. Members of S24-7 were diverse and abundant in rodents, making up 19% and 30%
368 of ASVs in *Apodemus* and *Microtus* respectively, and ranging in mean relative abundance
369 across species from 39% to 53%. Further analysis showed that the S24-7 ASVs most
370 informative for distinguishing congeneric species in RFCs were scattered throughout this
371 family's phylogeny, as were species-specific ASVs (Fig. S8).
372 In contrast to rodents, species-indicative ASVs in *Sorex* shrews came from a much
373 broader range of taxa, with 50% Proteobacteria, 15% Tenericutes and the remainder from
374 other phyla (Fig. 4, Fig. S7). In all three genera, species-indicative ASVs in RFC models
375 generally had a higher than average relative abundance (Fig. S9). Overall, these findings
376 indicate that in rodents, the bacterial taxa most indicative of host species were not a random
377 subset of those present, but biased towards particular members of the Bacteroidales, whereas
378 species-indicative taxa in shrews belong to a much broader range of bacterial groups.
379
380 ***Species distinguishability is sensitive to bacterial phylogenetic resolution***

For the rodent genera (*Apodemus* and *Microtus*), the species signal was strongest when considering fine-scale bacterial phylogenetic resolution rather than deeper branching bacterial groups. Specifically, the host species signal decayed at broader phylogenetic scales, yet this pattern was not seen for the effect of capture site (Fig. 5A). Moreover, at a standardized phylogenetic resolution (ASVs), distance metrics that downweight the influence of recent bacterial evolution (UniFrac metrics) showed weaker species signals than those that do not (Jaccard distance and Bray-Curtis dissimilarity; Fig. 5B, Table S5). Finally, the finer the bacterial taxonomic resolution used, the greater the accuracy of RFC models at classifying congeneric rodent samples to host species. For *Apodemus*, species assignment accuracy dropped from 100% to 70%, and for *Microtus* from 99% to 57% when using phyla rather than ASVs as features (Table S6). It is important to note, however, that while species distinguishability declined at coarse bacterial taxonomic resolution, it was still detectable. Even at the level of bacterial classes, the microbiota of congeneric rodent species remained statistically distinct (PERMANOVA Species term: *Apodemus* R^2=0.075, p=0.034, *Microtus* R^2= 0.061, p=0.0313).

By contrast, in *Sorex* shrews the species signal was insensitive to bacterial phylogenetic resolution (Fig. 5A) and the dissimilarity metric used (Fig. 5B). RFC classification to species also remained relatively accurate whether ASVs or whole phyla were used as features (Table S6).

Association between host diet and the gut microbiota

The resolving power of stomach contents data differed among host genera. Stomach contents varied little among *Apodemus* mice, which have a diet heavily dominated by seeds that could not be visually distinguished. However, voles and shrews showed more variation in stomach contents (Fig. S2). Consistent with an effect of diet on the microbiota, diet similarity correlated positively with microbiota similarity among individuals in all three genera, with this correlation strongest for voles and marginally significant for mice and shrews (Mantel test on Bray-Curtis dissimilarities, *Microtus*: r=0.22, p=0.002; *Apodemus*: r: 0.07, p=0.074; *Sorex*: r=0.10,

p=0.070). Predictors of diet composition differed for the three genera. For *Apodemus*, where power to resolve dietary differences was weakest, we only detected a weak effect of capture site. However, *Microtus* diet was strongly predicted by species and less so by site, while *Sorex* diet only showed a species effect (Fig. 3D-F, Table S7). In the two groups where we find marked diet variation, therefore, species identity was the dominant predictor. Moreover, looking broadly across all groups and comparisons (Fig. 3) the patterns of similarity in diet resembled those in the microbiota, consistent with a role for diet in shaping site and species effects on the microbiota.

Discussion

The relative importance of host genetics and the environment in shaping the gut microbiota continues to be a topic of major debate (Spor *et al.* 2011). Important for this debate are differences in host phylogenetic scale. Within-species studies often report relatively weak genetic compared to environmental effects (Carmody *et al.* 2015; Rothschild *et al.* 2018), whereas across-species comparisons have tended to emphasize genetic effects, including a pattern of 'phylosymbiosis', wherein gut microbiota similarity among species mirrors the host phylogeny (Brucker & Bordenstein 2012; Brooks *et al.* 2016). Here we test the relative importance of host genetics and the environment where these two scales meet. We find that in three small mammal genera, host genetics (species identity) dominates over a shared environment in predicting gut microbiota composition. Specifically, an individuals' microbiota was on average more similar to conspecifics living elsewhere than to members of a closely related species living in the same location. Moreover, while environment (capture site) did shape the microbiota, this effect was largely within species; heterospecific rodents converged somewhat in gut microbiota composition when living in sympatry, but this was insufficient to override the strong influence of species identity.

Our finding of strong and consistent species differences in the mammalian microbiota implies that host phylogenetic effects previously documented at broad scales (Groussin *et al.*

436 2017; Moeller *et al.* 2017; Amato *et al.* 2018; Nishida & Ochman 2018) persist even among
437 closely related species living in sympatry. Consistent with our findings, work on primates has
438 shown that host phylogeny dominates over geography and dietary niche in shaping the gut
439 microbiota (Amato *et al.* 2018), and that although members of species living in closer
440 geographic proximity (Moeller *et al.* 2013) or even at a similar level in the forest canopy
441 (Perofsky *et al.* 2018) converge in their microbiota, community composition remains most
442 strongly predicted by species identity. By contrast, studies on the gut microbiota of distantly
443 related artiodactyl species (Moeller *et al.* 2017) and the skin microbiota of congeneric
444 salamanders (Muletz Wolz *et al.* 2017; Bird *et al.* 2018) suggest that a shared environment
445 can drive community similarity more strongly than host phylogenetic proximity. The
446 dominance of species identity over environment we find may therefore not be universal, and
447 further studies are needed to assess the generality of this pattern across different host taxa
448 and microbial community types.

449 What drives the species signature we find in the microbiota of congeneric small
450 mammals? Vertical inheritance and host-symbiont codiversification is one possibility, and
451 recent studies have provided evidence this process occurs for some mammalian gut bacteria
452 (Moeller *et al.* 2016; Groussin *et al.* 2017). Consistent with a role for codiversification, in mice
453 and voles we found that the microbiota of closely related species was most easily
454 distinguished when considering recently diverged bacterial groups (Fig. 5), a pattern also
455 recently found across a broader range of mammals (Groussin *et al.* 2017), but not in primates
456 (Sanders *et al.* 2014; Amato *et al.* 2018). The microbiota of different *Apodemus* species
457 (estimated to have diverged ~7 million years ago; Michaux *et al.* 2002; Suzuki *et al.* 2008)
458 was also more distinct than that of *Microtus* species (which diverged ~ 2 million years ago;
459 Bannikova *et al.* 2010). This is consistent with a positive correlation between microbiota
460 distinctness and host divergence time, as expected under codiversification and previously
461 shown for other mammals in the lab (Brooks *et al.* 2016) and the wild (Moeller *et al.* 2017). A
462 broader phylogenetic analysis using markers with greater resolution than 16S rRNA would be
463 needed to definitively test for codiversification between small mammals and their gut

464 microbes. However, it is also clear that codiversification cannot be the only process at play
465 here, as species distinguishability remained (albeit weaker in rodents) even at the level of
466 bacterial classes, which diverged long before their hosts. A range of other processes could
467 contribute to species distinctness in the gut microbiota (Davenport *et al.* 2017). For example,
468 closely related hosts (members of the same species) are more likely to share genetic or
469 behavioural mechanisms that drive the horizontal acquisition and retention of similar bacteria
470 from the environment. These include dietary preferences, innate and adaptive immune
471 components, gut morphology and mucus characteristics, all of which can differentially select
472 members of the microbiota (Kato *et al.* 2014; Carmody *et al.* 2015; Pabst *et al.* 2016; Sicard
473 *et al.* 2017; Amato *et al.* 2018). Of these mechanisms, those involving microbes binding to
474 diverse host epitopes, such as immunoglobulins or mucus glycans, also have the potential to
475 produce highly specific host-microbe interactions (Schroeder & Cavacini 2010; Naughton *et*
476 *al.* 2013), and generate species differences in the microbiota at a fine bacterial phylogenetic
477 scale, as observed here. Our data also suggest diet may play a role generating species
478 differences in the microbiota. In voles and shrews, we found species differences in diet that
479 were maintained in sympatry, and diet predicted microbiota variation. It is also noteworthy that
480 species-specificity in the shrew microbiota was insensitive to bacterial phylogenetic resolution,
481 and that the two shrew species studied differed strongly in diet, with *S. araneus* often having
482 eaten earthworms while *S. minutus* ate only arthropods. Host selection of different (deeply
483 diverged) gut microbes through contrasting diet may therefore play a more prominent role
484 shaping species distinguishability of the gut microbiota in this genus. Another possibility is that
485 the shrew gut microbiota includes more symbionts from their animal diet than the rodent
486 microbiota. This seems plausible given that Proteobacteria, the dominant phylum in the
487 earthworm microbiota (Liu *et al.* 2018), were much more abundant in earthworm-eating *S.*
488 *araneus* than *S. minutus*. Overall, our data suggest that dietary variation is more likely to drive
489 species differences in the microbiota than act as an environmental factor blurring them. In this
490 way, we can marry the statements that diet has important effects on the mammalian gut

491 microbiota, but that host genetics is still the ultimate force shaping these communities at this
492 host phylogenetic scale.
493 We also found that not all members of the microbiota were equally important for
494 distinguishing host species. Members of the order Bacteroidales were key drivers of host
495 species distinguishability in rodents. In particular, the family S24-7 were important, a group
496 found almost exclusively in the gut of homeothermic animals (Ormerod *et al.* 2016) and
497 abundant and diverse in the wild rodents we sampled. Why some bacterial groups are most
498 host specific than others is an interesting open question. One possibility is that some bacteria
499 are more amenable to host selection via immunity (Benson *et al.* 2010; Kurilshikov *et al.* 2017),
500 adhesion (McLoughlin *et al.* 2016) or consumption of host mucus (Sicard *et al.* 2017).
501 Interestingly, members of the S24-7 family vary in their trophic guild, with some degrading
502 plant glycans while others degrade host glycans (Ormerod *et al.* 2016), as well as their degree
503 of IgA coating (Bunker *et al.* 2015). Such differences in biology warrant further investigation
504 as potential mediators of host specificity.
505
506 In summary, we find across three small mammal genera that the gut microbiota is highly
507 species-specific, and that while sharing a habitat drives some convergence in community
508 composition among members of closely related species, this is insufficient to override the
509 dominant signature of species identity. Moreover, in rodents, host species distinguishability in
510 the microbiota was greatest at the tips of the bacterial phylogeny, and driven in large part by
511 members of the Bacteroidales. An important future goal will be to understand the processes
512 driving host specificity in the mammalian microbiota, and why different gut bacteria vary in the
513 strength of their association with a particular host species.
514
515 **Acknowledgements**
516 We would like to thank Kevin Foster for thoughtful comments on the manuscript and editing,
517 and Tim Barraclough and Tom Bell for supporting laboratory aspects of the work. This work
518 was funded by a NERC fellowship (NE/L011867/1) to SCLK.

519

520 **References**

521 Amato, K.R., Martinez-Mota, R., Righini, N., Raguet-Schofield, M., Corcione, F.P., Marini, E., *et al.*
522 (2016). Phylogenetic and ecological factors impact the gut microbiota of two Neotropical primate
523 species. *Oecologia*, 180, 717–733.

524 Amato, K.R., Sanders, J., Song, S.J., Nute, M., Metcalf, J.L., Thompson, L.R., *et al.* (2018).
525 Evolutionary trends in host physiology outweigh dietary niche in structuring primate gut
526 microbiomes. *ISME J.*, 11, 1.

527 Anderson, M.J. (2001). A new method for non parametric multivariate analysis of variance. *Austral*
528 *Ecol.*, 26, 32–46.

529 Andreassen, H.P., Hertzberg, K. & Ims, R.A. (1998). Space-use responses to habitat fragmentation
530 and connectivity in the root vole *Microtus oeconomus*. *Ecology*, 79, 1223–1235.

531 Bannikova, A.A., Lebedev, V.S., Lissovsky, A.A., Matrosova, V., Abramson, N.I., Obolenskaya, E. V,
532 *et al.* (2010). Molecular phylogeny and evolution of the Asian lineage of vole genus *Microtus*
533 (Rodentia : Arvicolinae) inferred from mitochondrial cytochrome b sequence. *Biol. J. Linn. Soc.*,
534 99, 595–613.

535 Baxter, N.T., Wan, J.J., Schubert, A.M., Jenior, M.L., Myers, P. & Schloss, P.D. (2015). Intra- and
536 interindividual variations mask interspecies variation in the microbiota of sympatric *Peromyscus*
537 populations. *Appl. Environ. Microbiol.*, 81, 396–404.

538 Benson, A.K., Kelly, S. a, Legge, R., Ma, F., Low, S.J., Kim, J., *et al.* (2010). Individuality in gut
539 microbiota composition is a complex polygenic trait shaped by multiple environmental and host
540 genetic factors. *Proc. Natl. Acad. Sci. U. S. A.*, 107, 18933–18938.

541 Bird, A.K., Prado-Irwin, S.R., Vredenburg, V.T. & Zink, A.G. (2018). Skin microbiomes of California
542 terrestrial salamanders are influenced by habitat more than host phylogeny. *Front. Microbiol.*, 9,
543 442.

544 Breiman, L. (2001). *Random Forests*. *Mach. Learn. 45(1), 5-32.*

545 Brooks, A.W., Kohl, K.D., Brucker, R.M., van Opstal, E.J. & Bordenstein, S.R. (2016).
546 Phylosymbiosis: relationships and functional effects of microbial communities across host
547 evolutionary history. *PLOS Biol.*, 14, e2000225.

548 Brucker, R.M. & Bordenstein, S.R. (2012). Speciation by symbiosis. *Trends Ecol. Evol.*, 27, 443–451.

549 Buffie, C.G. & Pamer, E.G. (2013). Microbiota-mediated colonization resistance against intestinal
550 pathogens. *Nat Rev Immunol*, 13, 790–801.
551 Bunker, J.J., Flynn, T.M., Koval, J.C., Jabri, B., Antonopoulos, D.A. & Benddelac, A. (2015). Innate
552 and adaptive humoral responses coat distinct commensal bacteria with immunoglobulin A.
553 *Immunity*, 43, 541–553.
554 Callahan, B.J., McMurdie, P.J. & Holmes, S.P. (2017). Exact sequence variants should replace
555 operational taxonomic units in marker-gene data analysis. *ISME J.*, 11, 2639–2643.
556 Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J. & Holmes, S.P. (2016a).
557 DADA2 : High resolution sample inference from amplicon data. *Nat. Methods*, 13, 581.
558 Callahan, B.J., Sankaran, K., Fukuyama, J.A., McMurdie, P.J. & Holmes, S.P. (2016b). Bioconductor
559 workflow for microbiome data analysis: from raw reads to community analyses. *F1000Research*,
560 5, 1492.
561 Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., *et al.*
562 (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample.
563 *Proc. Natl. Acad. Sci. U. S. A.*, 108 Suppl, 4516–22.
564 Carmody, R.N., Gerber, G.K., Luevano, J.M., Gatti, D.M., Somes, L., Svenson, K.L., *et al.* (2015). Diet
565 dominates host genotype in shaping the murine gut microbiota. *Cell Host Microbe*, 17, 72–84.
566 Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K., *et al.* (2014). Rarefaction
567 and extrapolation with Hill numbers: A framework for sampling and estimation in species
568 diversity studies. *Ecol. Monogr.*, 84, 45–67.
569 Crist, T.O., Veech, J.A., Gering, J.C. & Summerville, K.S. (2003). Partitioning species diversity across
570 landscapes and regions: a hierarchical analysis of α, β, and γ diversity. *Am. Nat.*, 162, 734–743.
571 D'Amore, R., Ijaz, U.Z., Schirmer, M., Kenny, J.G., Gregory, R., Darby, A.C., *et al.* (2016). A
572 comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA
573 community profiling. *BMC Genomics*, 17, 55.
574 Davenport, E.R., Sanders, J.G., Song, S.J., Amato, K.R., Clark, A.G. & Knight, R. (2017). The human
575 microbiome in evolution. *BMC Biol.*, 15, 127.
576 David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., *et al.* (2014).
577 Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, 505, 559–563.
578 Dietrich, C., Köhler, T. & Brune, A. (2014). The cockroach origin of the termite gut microbiota:

579 Patterns in bacterial community structure reflect major evolutionary events. *Appl. Environ.*
580 *Microbiol.*, 80, 2261–2269.
581 Faith, J.J., Guruge, J.L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A.L., *et al.*
582 (2013). The long-term stability of the human gut microbiota. *Science*, 341, 1237439.
583 Foster, K.R., Schluter, J., Coyte, K.Z. & Rakoff-Nahoum, S. (2017). The evolution of the host
584 microbiome as an ecosystem on a leash. *Nature*, 548, 43–51.
585 Galili, T. (2015). dendextend: An R package for visualizing, adjusting and comparing trees of
586 hierarchical clustering. *Bioinformatics*, 31, 3718–3720.
587 Gauffre, B., Estoup, A., Bretagnolle, V. & Cosson, J.F. (2008). Spatial genetic structure of a small
588 rodent in a heterogeneous landscape. *Mol. Ecol.*, 17, 4619–4629.
589 Griffin, N.W., Ahern, P.P., Cheng, J., Heath, A.C., Ilkayeva, O., Newgard, C.B., *et al.* (2017). Prior
590 dietary practices and connections to a human gut microbial metacommunity alter responses to
591 diet interventions. *Cell Host Microbe*, 21, 84–96.
592 Groussin, M., Mazel, F., Sanders, J.G., Smillie, C.S., Lavergne, S., Thuiller, W., *et al.* (2017).
593 Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nat.*
594 *Commun.*, 8, 14319.
595 Hildebrand, F., Nguyen, T.L.A., Brinkman, B., Yunta, R.G., Cauwe, B., Vandenabeele, P., *et al.*
596 (2013). Inflammation-associated enterotypes, host genotype, cage and inter-individual effects
597 drive gut microbiota variation in common laboratory mice. *Genome Biol.*, 14, R4.
598 Hsieh, T.C., Ma, K.H. & Chao, A. (2016). iNEXT: interpolation and extrapolation for species diversity.
599 R package version 2.0.8. *R-project*, 1–18.
600 Kato, L.M., Kawamoto, S., Maruya, M. & Fagarasan, S. (2014). The role of the adaptive immune
601 system in regulation of gut microbiota. *Immunol. Rev.*, 260, 67–75.
602 Knights, D., Costello, E.K. & Knight, R. (2011). Supervised classification of human microbiota. *FEMS*
603 *Microbiol. Rev.*, 35, 343–359.
604 Kohl, K.D., Weiss, R.B., Cox, J., Dale, C. & Dearing, M.D. (2014). Gut microbes of mammalian
605 herbivores facilitate intake of plant toxins. *Ecol. Lett.*, 17, 1238–1246.
606 Kurilshikov, A., Wijmenga, C., Fu, J. & Zhernakova, A. (2017). Host genetics and gut microbiome:
607 challenges and perspectives. *Trends Immunol.*, 38, 633–647.
608 Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., *et al.* (2008).

609 Evolution of mammals and their gut microbes. *Science*, 320, 1647–1651.
610 Ley, R.E., Peterson, D.A. & Gordon, J.I. (2006). Ecological and evolutionary forces shaping microbial
611 diversity in the human intestine. *Cell*, 124, 837–848.
612 Liu, D., Lian, B., Wu, C. & Guo, P. (2018). A comparative study of gut microbiota profiles of
613 earthworms fed in three different substrates. *Symbiosis*, 74, 21–29.
614 Martinson, V.G., Douglas, A.E. & Jaenike, J. (2017). Community structure of the gut microbiota in
615 sympatric species of wild *Drosophila*. *Ecol. Lett.*, 20, 629–639.
616 Maurice, C.F., Knowles, S.C., Ladau, J., Pollard, K.S., Fenton, A., Pedersen, A.B., *et al.* (2015).
617 Marked seasonal variation in the wild mouse gut microbiota. *ISME J.*, 9, 1–12.
618 McFall-Ngai, M., Hadfield, M.G., Bosch, T.C.G., Carey, H. V, Domazet-Lošo, T., Douglas, A.E., *et al.*
619 (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl. Acad.*
620 *Sci.*, 110, 3229–3236.
621 McLoughlin, K., Schluter, J., Rakoff-Nahoum, S., Smith, A.L. & Foster, K.R. (2016). Host selection of
622 microbiota via differential adhesion. *Cell Host Microbe*, 19, 550–559.
623 McMurdie, P.J. & Holmes, S. (2013). Phyloseq: An R package for reproducible interactive analysis
624 and graphics of microbiome census data. *PLoS One*, 8, e61217.
625 Michaux, J.R., Chevret, P., Filippucci, M. & Macholan, M. (2002). Phylogeny of the genus *Apodemus*
626 with a special emphasis on the subgenus *Sylvaemus* using the nuclear IRBP gene and two
627 mitochondrial markers: cytochrome b and 12S rRNA. *Mol. Phylogenet. Evol.*, 23, 123–136.
628 Moeller, A.H., Caro-Quintero, A., Mjungu, D., Georgiev, A. V., Lonsdorf, E. V., Muller, M.N., *et al.*
629 (2016). Cospeciation of gut microbiota with hominids. *Science* 353, 380–382.
630 Moeller, A.H., Peeters, M., Ndjango, J.B., Li, Y., Hahn, B.H. & Ochman, H. (2013). Sympatric
631 chimpanzees and gorillas harbor convergent gut microbial communities. *Genome Res.*, 23,
632 1715–1720.
633 Moeller, A.H., Suzuki, T.A., Lin, D., Lacey, E.A. & Wasser, S.K. (2017). Dispersal limitation promotes
634 the diversification of the mammalian gut microbiota. *Proc. Natl. Acad. Sci.*, 114, 13768–13773.
635 Muletz Wolz, C.R., Yarwood, S.A., Campbell Grant, E.H., Fleischer, R.C. & Lips, K.R. (2017). Effects
636 of host species and environment on the skin microbiome of Plethodontid salamanders. *J. Anim.*
637 *Ecol.*, 341–353.
638 Naughton, J.A., Mariño, K., Dolan, B., Reid, C., Gough, R., Gallagher, M.E., *et al.* (2013). Divergent

639 mechanisms of interaction of *Helicobacter pylori* and *Campylobacter jejuni* with mucus and
640 mucins. *Infect. Immun.*, 81, 2838–2850.
641 Nishida, A.H. & Ochman, H. (2018). Rates of gut microbiome divergence in mammals. *Mol. Ecol.*, 27,
642 1884–97.
643 Ochman, H., Worobey, M., Kuo, C.H., Ndjango, J.B.N., Peeters, M., Hahn, B.H., *et al.* (2010).
644 Evolutionary relationships of wild hominids recapitulated by gut microbial communities. *PLoS*
645 *Biol.*, 8, 3–10.
646 Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., *et al.* (2017). vegan:
647 2.4-5, Community Ecology Package. R package version https://CRAN.R-
648 project.org/package=vegan.
649 Ormerod, K.L., Wood, D.L.A., Lachner, N., Gellatly, S.L., Daly, J.N., Parsons, J.D., *et al.* (2016).
650 Genomic characterization of the uncultured Bacteroidales family S24-7 inhabiting the guts of
651 homeothermic animals. *Microbiome*, 4, 36.
652 Pabst, O., Cerovic, V. & Hornef, M. (2016). Secretory IgA in the coordination of establishment and
653 maintenance of the microbiota. *Trends Immunol.*, 37, 287–296.
654 Paradis, E., Claude, J. & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R
655 language. *Bioinformatics*, 20, 289–290.
656 Paulson, J.N., Colin Stine, O., Bravo, H.C. & Pop, M. (2013). Differential abundance analysis for
657 microbial marker-gene surveys. *Nat. Methods*, 10, 1200–1202.
658 Perofsky, A.C., Lewis, R.J. & Meyers, L.A. (2018). Terrestriality and bacterial transfer: a comparative
659 study of gut microbiomes in sympatric Malagasy mammals. *ISME J.*, 15–18.
660 Phillips, C.D., Phelan, G., Dowd, S.E., McDonough, M.M., Ferguson, A.W., Delton Hanson, J., *et al.*
661 (2012). Microbiome analysis among bats describes influences of host phylogeny, life history,
662 physiology and geography. *Mol. Ecol.*, 21, 2617–2627.
663 R Core Team (2017). R: A language and environment for statistical computing. URL https://www.R-
664 project.org/.
665 Ren, T., Boutin, S., Humphries, M.M., Dantzer, B., Gorrell, J.C., Coltman, D.W., *et al.* (2017).
666 Seasonal, spatial, and maternal effects on gut microbiome in wild red squirrels. *Microbiome*, 5,
667 163.
668 Rosenbaum, M., Knight, R., Leibel, R.L., Science, C. & Jolla, L. (2016). The gut microbiota in human

669 energy homeostasis and obesity. *Trends Endocrinol. Metab.*, 26, 493–501.
670 Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., *et al.* (2018).
671 Environment dominates over host genetics in shaping human gut microbiota. *Nature*, 555, 210–
672 215.
673 Sanders, J., Kronauer, D.J.C., Vasconcelos, L., Frederickson, M. & Pierce, N.E. (2014). Stability and
674 phylogenetic correlation in gut microbiota: lessons from ants and apes. *Microb. Ecol.*, 23, 1268–
675 1283.
676 Schroeder, H. & Cavacini, L. (2010). Structure and function of immunoglobulins. *J Allergy Clin*
677 *Immunol*, 125, S41–S52.
678 Seedorf, H., Griffin, N.W., Ridaura, V.K., Reyes, A., Cheng, J., Rey, F.E., *et al.* (2014). Bacteria from
679 diverse habitats colonize and compete in the mouse gut. *Cell*, 159, 253–266.
680 Sicard, J.-F., Le Bihan, G., Vogeleer, P., Jacques, M. & Harel, J. (2017). Interactions of intestinal
681 bacteria with components of the intestinal mucus. *Front. Cell. Infect. Microbiol.*, 7, 387.
682 Song, S.J., Lauber, C., Costello, E.K., Lozupone, C.A., Humphrey, G., Berg-Lyons, D., *et al.* (2013).
683 Cohabiting family members share microbiota with one another and with their dogs. *Elife*, 2013,
684 1–22.
685 Sonnenburg, E.D., Smits, S.A., Tikhonov, M., Higginbottom, S.K., Wingreen, N.S. & Sonnenburg, J.L.
686 (2016). Diet-induced extinctions in the gut microbiota compound over generations. *Nature*, 529,
687 212–215.
688 Spor, A., Koren, O. & Ley, R. (2011). Unravelling the effects of the environment and host genotype on
689 the gut microbiome. *Nat. Rev. Microbiol.*, 9, 279–290.
690 Stradiotto, A., Cagnacci, F., Delahay, R., Tioli, S., Nieder, L. & Rizzoli, A. (2009). Spatial organization
691 of the yellow-necked mouse: effects of density and resource availability. *J. Mammal.*, 90, 704–
692 714.
693 Suzuki, H., Fillippucci, M.G., Chelomina, G.N., Sato, J.J., Serizawa, K. & Nevo, E. (2008). A
694 biogeographic view of *Apodemus* in Asia and Europe inferred from nuclear and mitochondrial
695 gene sequences. *Biochem. Genet.*, 46, 329–346.
696 Wang, J., Chen, L., Zhao, N., Xu, X., Xu, Y. & Zhu, B. (2018). Of genes and microbes: solving the
697 intricacies in host genomes. *Protein Cell*, 9, 446–461.
698 Wang, M. & Grimm, V. (2007). Home range dynamics and population regulation: An individual-based

699 model of the common shrew *Sorex araneus*. *Ecol. Modell.*, 205, 397–409.
700 Weissbrod, O., Rothschild, D., Barkan, E. & Segal, E. (2018). Host genetics and microbiome
701 associations from the lens of genome wide association studies. *Curr. Opin. Microbiol.*, 44, 9–19.
702 Yletyinen, S. & Norrdahl, K. (2008). Habitat use of field voles (*Microtus agrestis*) in wide and narrow
703 buffer zones. *Agric. Ecosyst. Environ.*, 123, 194–200.
704

Figure legends

Figure 1: Variation in gut microbiota composition across small mammal clades. (A) Principle coordinates (PCoA) plot based on Bray-Curtis dissimilarities indicating clustering of samples by host family and genus (B) Phylum-level gut microbiota composition by host species, with taxa unassigned to the phylum level removed (C) Asymptotic estimates of amplicon sequence variant (ASV) richness and Shannon diversity for each host species sampled, coloured by host genus, as estimated in R package iNEXT.

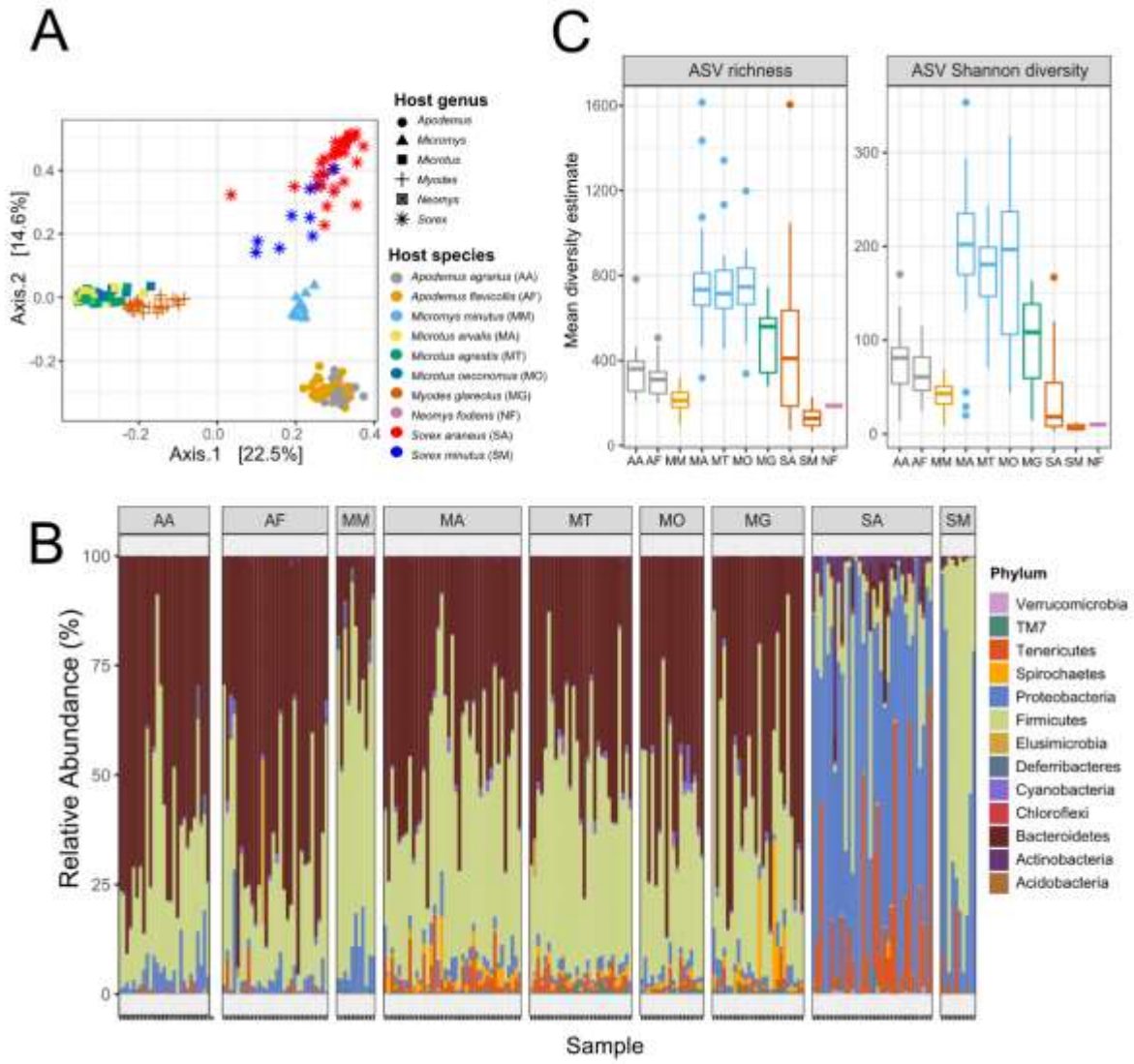

Figure 2: Clustering of gut microbial communities in three genera of small mammals according to species identity and capture site. Hierarchical clustering of samples according to Bray-Curtis dissimilarity. Dendrograms were constructed using UPGMA, with branches coloured according to host species, and bars indicating which host species and capture site each sample came from (A-C). Principle coordinate (PCoA) plots based on Bray-Curtis dissimilarity, with samples coloured by species, and capture sites indicated by symbols (D-F).

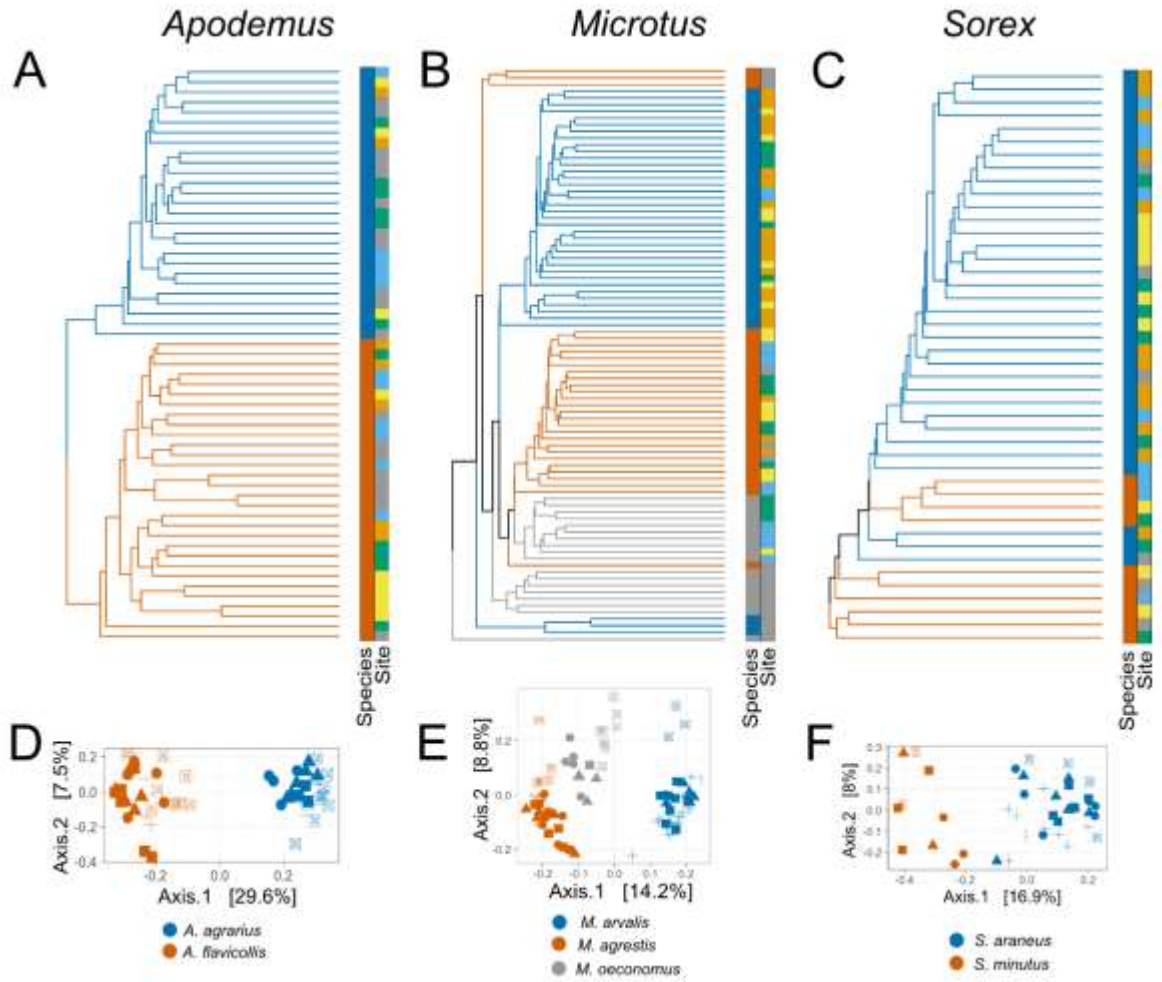

Figure 3: Pairwise differences in gut microbial community and diet composition according to species identity and capture site. Mean pairwise Bray-Curtis dissimilarities in microbiota composition (A-C) and stomach contents composition (D-F) according to whether samples came from the same species and/or the same capture site. Statistical significance is from Monte Carlo permutations: * p<0.05, ** p<0.001, ***p<0.0001, n.s. p>0.05. Black and dark grey bars indicate tests for species and site main effects respectively, while pale grey bars indicate tests involving species- or site-specific subsets of the data. Plots are based only on samples (n=215 in total) for which paired microbiota and diet data were available.

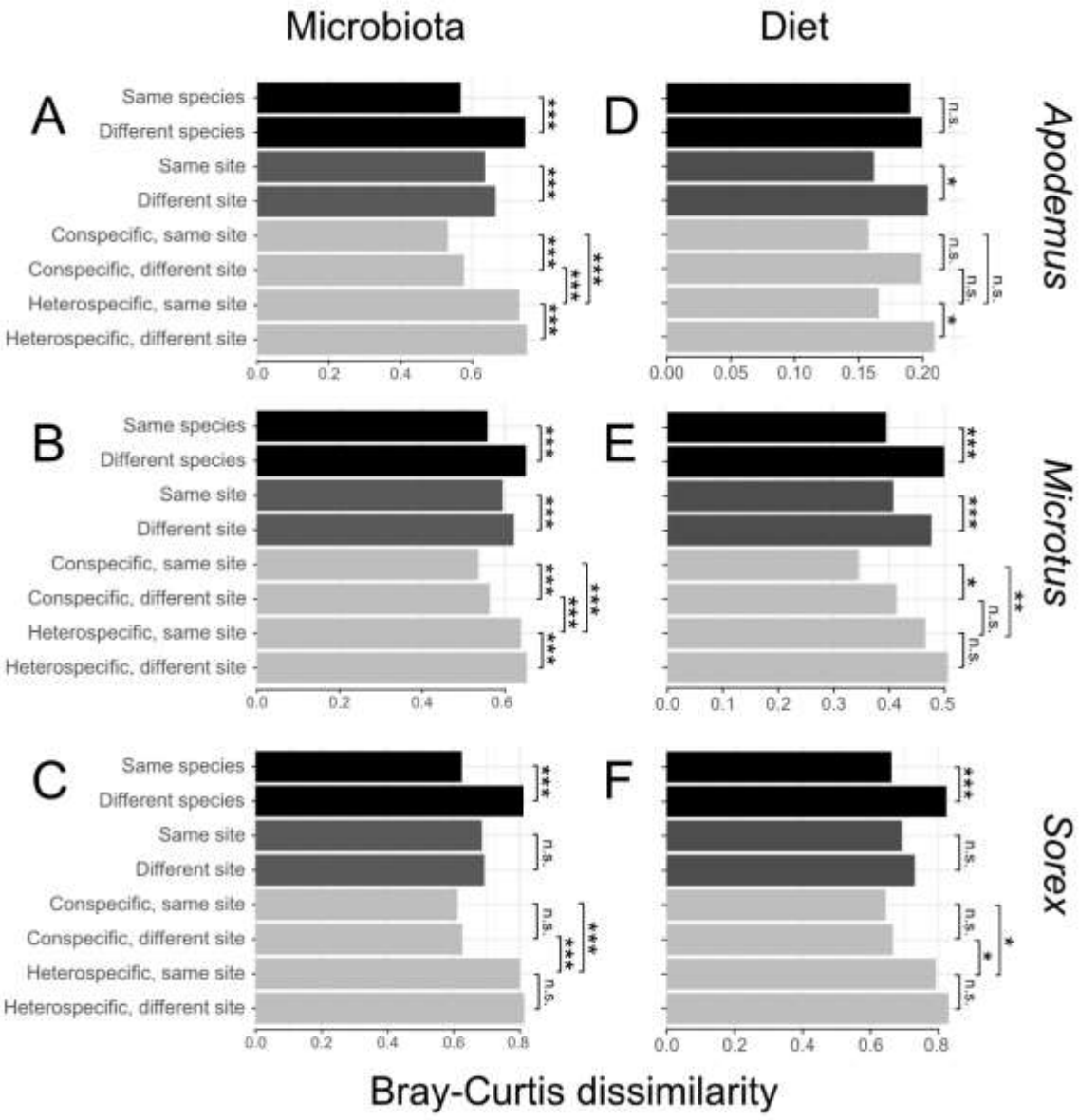

Figure 4: Representation of bacterial families among sequence variants most informative in species-classification RFC models compared to the full dataset. Bars indicate the proportion of sequence variants from each family that make up either the top 20 most important sequence variants for accurately assigning samples to host species in RFC models, compared to representation of the same families in the full dataset.

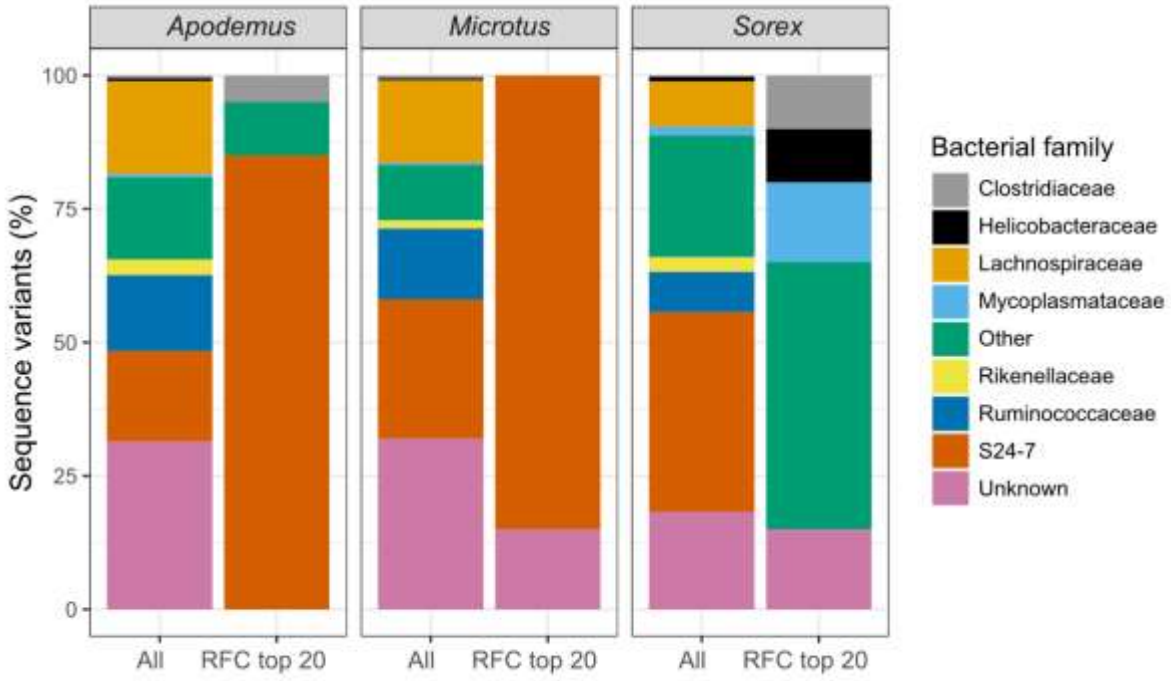

Figure 5: Factors affecting the strength of the host species signal in small mammal gut microbiota. (A) Strength of species and capture site effects within each host genus as estimated using R^2 from a PERMANOVA based on Bray-Curtis dissimilarity, with increasing agglomeration of branches (sequence variants) in the bacterial phylogeny. The x-axis indicates the parameter value used to define bacterial groups using the tip_glom function in *phyloseq* (B) PCoA plots showing how clustering of samples by host species within each genus varies across four dissimilarity metrics that differ in their sensitivity to the phylogenetic relatedness and abundance of bacterial sequence variants. R^2 values from PERMANOVAs testing the species effect are shown on each plot.

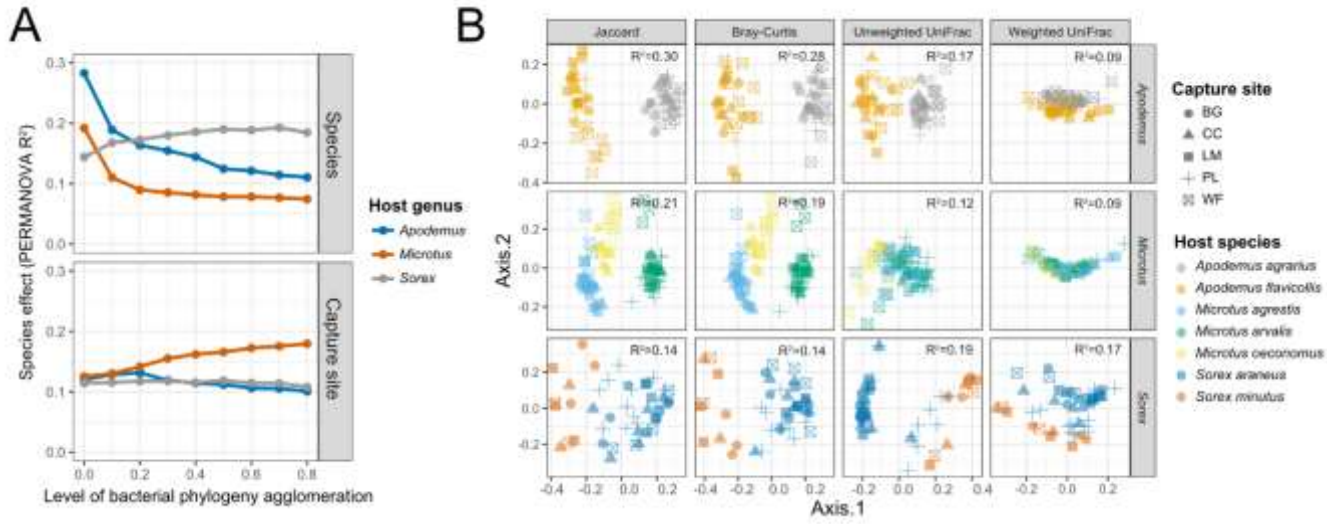

Table 1: **Hierarchical partitioning of total ASV richness**. Additive diversity partitioning was performed using the *adipart* function in *vegan*.

Host group	Mean # ASVs	Level	%
All (except *Neomys*)	2668	Within individual	16.92
	6555	Between individuals	41.58
	1393	Between species	8.84
	1918	Between genera	12.17
	3230	Between families	20.49
	15764	Total	100.00
Apodemus	1058	Within individual	40.15
	1347	Between individuals	51.09
	231	Between species	8.77
	2636	Total	100.00
Microtus	2093	Within individual	30.46
	3637	Between individuals	52.95
	1139	Between species	16.58
	6869	Total	100.00
Sorex	1190	Within individual	18.38
	3863	Between individuals	59.67
	1421	Between species	21.95
	6474	Total	100.00