Received: 16 January 2019 | Revised: 26 April 2019 | Accepted: 20 May 2019

DOI: 10.1111/mec.15139

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Host genetic determinants of the gut microbiota of wild mice

Taichi A. Suzuki[1] | Megan Phifer-Rixey[2] | Katya L. Mack[1] | Michael J. Sheehan[3] | Dana Lin[1] | Ke Bi[4] | Michael W. Nachman[1]

[1]Department of Integrative Biology and Museum of Vertebrate Zoology, University of California Berkeley, Berkeley, California, USA

[2]Department of Biology, Monmouth University, West Long Branch, New Jersey, USA

[3]Department of Neurobiology and Behavior, Cornell University, Ithaca, New York, USA

[4]California Institute for Quantitative Biosciences, University of California Berkeley, Berkeley, California, USA

Correspondence
Taichi A. Suzuki, Department of Microbiome Science, Max Planck Institute for Developmental Biology, Max-Planck-Ring 5, 72076 Tübingen, Germany.
Email: taichi.suzuki@tuebingen.mpg.de

Funding information
NIH, Grant/Award Number: RO1 GM074245 and RO1 GM127468; Directorate for Biological Sciences and NSF, Grant/Award Number: 1501646

Abstract

Identifying a common set of genes that mediate host–microbial interactions across populations and species of mammals has broad relevance for human health and animal biology. However, the genetic basis of the gut microbial composition in natural populations remains largely unknown outside of humans. Here, we used wild house mouse populations as a model system to ask three major questions: (a) Does host genetic relatedness explain interindividual variation in gut microbial composition? (b) Do population differences in the microbiota persist in a common environment? (c) What are the host genes associated with microbial richness and the relative abundance of bacterial genera? We found that host genetic distance is a strong predictor of the gut microbial composition as characterized by 16S amplicon sequencing. Using a common garden approach, we then identified differences in microbial composition between populations that persisted in a shared laboratory environment. Finally, we used exome sequencing to associate host genetic variants with microbial diversity and relative abundance of microbial taxa in wild mice. We identified 20 genes that were associated with microbial diversity or abundance including a macrophage-derived cytokine (*IL12a*) that contained three nonsynonymous mutations. Surprisingly, we found a significant overrepresentation of candidate genes that were previously associated with microbial measurements in humans. The homologous genes that overlapped between wild mice and humans included genes that have been associated with traits related to host immunity and obesity in humans. Gene–bacteria associations identified in both humans and wild mice suggest some commonality to the host genetic determinants of gut microbial composition across mammals.

KEYWORDS
coevolution, GWAS, mammals, metagenomics, microbiome, *Mus musculus*

1 | INTRODUCTION

Host-associated microbial communities play an important role in health and fitness (McFall-Ngai et al., 2013). Compositional and functional variation in the gut microbiota has been linked to a variety of diseases in humans and laboratory mouse models including obesity, inflammatory bowel disease and autism (e.g., Hsiao et al., 2013; Marchesi et al., 2007; Turnbaugh et al., 2006). Links between the gut microbiota and fitness-related traits have also been reported in wild mammals including traits related to digestion, immunity and behaviour (Suzuki, 2017). Therefore, understanding the mechanisms governing the maintenance and function of gut microbial communities is important in medicine and animal biology more broadly.

Host genetics may play an important role in structuring gut microbial communities. For example, genome-wide markers have been associated with overall differences in the microbiome (i.e.,

beta-diversity) in humans (Blekhman et al., 2015; Ma et al., 2014). Twin studies have shown that monozygotic twins tend to have more similar microbial composition compared to dizygotic twins (Goodrich, Davenport, Beaumont, et al., 2016; Goodrich et al., 2014). Mouse knockout experiments have identified genes involved in immunity, metabolism and behaviour that affect the gut microbiota (Spor, Koren, & Ley, 2011). Mouse quantitative trait locus (QTL) mapping studies have also identified multiple genomic regions associated with the relative abundance of different microbial taxa (Benson et al., 2010; Leamy et al., 2014; McKnite et al., 2012; Org et al., 2015; Wang et al., 2015).

In human populations, microbiome genome-wide association studies (mGWAS) have identified specific candidate genes associated with natural variation of the gut microbiota (Blekhman et al., 2015; Bonder et al., 2016; Davenport et al., 2015; Goodrich, Davenport, Beaumont, et al., 2016; Knights et al., 2014; Turpin et al., 2016; Wang et al., 2016), and a few gene–bacteria associations have been replicated in multiple human populations (Goodrich, Davenport, Waters, Clark, & Ley, 2016; Hall, Tolonen, & Xavier, 2017). However, this approach has not been used to look for gene–bacteria associations in wild mammals. Genes identified from human mGWAS are often compared with those identified in laboratory mice, but gene–bacteria associations identified in a controlled laboratory environment may differ from those in a complex natural environment. In fact, the function and composition of the gut microbiota in laboratory mice are known to differ from those of their wild relatives (Rosshart et al., 2017). Samples from a wild population would provide an opportunity for mGWAS that is more directly comparable to human mGWAS.

Wild house mice (*Mus musculus domesticus*) are globally distributed and live in a wide range of environments in association with humans (Phifer-Rixey & Nachman, 2015). The house mouse is a powerful model because it is possible to disentangle variables using experimental manipulation (Wang et al., 2015, 2014) and assess the functions of the microbiome using germ-free mice (Rosshart et al., 2017). Previous work has shown that geographic and genetic distances (Linnenbrink et al., 2013), diet as measured by stable isotopes (Wang et al., 2014), reproductive status, body size, age, viral and parasite infection status (Weldon et al., 2015), gut regions (Suzuki & Nachman, 2016) and altitude (Suzuki, Martins, & Nachman, 2018) are associated with compositional differences in the gut microbiota of wild house mice. However, there have been no previous efforts to identify specific genes underlying compositional variation in the gut microbiota of wild mice.

Here, we characterize natural variation in the gut microbiota of wild house mice sampled from five populations along an environmental gradient in eastern North America and identify specific host genes associated with gut microbial composition. First, we show that genetic distance correlates with microbial composition (i.e., beta-diversity), both within and between populations. Second, using a common garden experiment, we show that differences in the gut microbiota among wild populations persist in the laboratory, suggesting that they are not driven by the environment such as diet. Third, we identify genome-wide gene–bacteria associations in wild mice using the complete exome sequences of all mice. Finally, we document significant overlap between the genes associated with microbiota variation in mice and genes associated with microbiota variation in humans.

2 | MATERIALS AND METHODS

2.1 | Sample collection

We collected a total of 50 adult house mice (*Mus musculus domesticus*) from five populations in eastern North America during the summer of 2012 (Phifer-Rixey et al., 2018) and generated 80 laboratory-reared individuals in the laboratory for a common garden experiment (see below). For wild-caught individuals, we collected ten mice each from each of five populations: Florida (FL), Georgia (GA), Virginia (VA), Pennsylvania (PA) and New Hampshire—Vermont (NH-VT) (summarized in Table S1). Sherman live traps were used with peanut butter and oats as bait. Each mouse was caught a minimum of 500 m from all other mice to avoid sampling close relatives. Animals were kept in Sherman traps, euthanized by cervical dislocation, and all tissues and external measurements were collected within 24 hr after capture. Caecum and liver were stored in liquid nitrogen in the field and then stored in a deep freezer (−80°C) until sequencing. Diet was inferred using carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$) stable isotopes from mouse hair following the protocol of Suzuki and Nachman (2016). Climate data were inferred using the first two principal components (Table S2) calculated from 19 climatic variables downloaded from WorldClim database (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) and the R package "dismo".

For the common garden experiment, we captured live animals close to the northern and southern populations using Sherman live traps during summer, 2013 from Saratoga Springs, NY, and Gainsville, FL. Within each location, animals were collected from at least 10 sites that were a minimum of 500 m apart. Animals were housed in the laboratory at 23°C with a light cycle of 10 hr dark and 14 hr light. Teklad Global food (18% Protein Rodent Diet) was fed ad libitum. Wild-caught mice from the same population were paired to produce F1 offspring. Ten independent crosses were conducted, and offspring were housed with their own littermates. Body weight and fresh faecal samples were collected from 40 individuals per population representing four adults (two females and two males) each from the 10 independent crosses of wild-caught founders. Faecal samples were stored in −80°C until sequencing. Detailed information on laboratory-reared animals is in Table S3. All procedures involving animals were reviewed and approved by the IACUC at the University of Arizona (07-004) and at the University of California Berkeley (R361-0514).

2.2 | 16S rRNA gene sequencing

We extracted DNA from caecal and faecal samples from wild-caught and laboratory-reared mice, respectively. Although these sample types are different, faecal samples are known to closely match individual differences in caecal samples in wild house mice (Suzuki & Nachman, 2016). We followed the DNA extraction protocol described in Suzuki and Nachman (2016). Briefly, we added a

bead-beating step before step 4 in the protocol from QIAamp DNA Stool Minikit (Qiagen). The V4 region of the 16S rRNA gene was amplified and multiplexed using the standard primers and barcodes described in Caporaso et al. (2012), and the samples were sequenced on two lanes of 150-bp pair-end Illumina MiSeq at the Next Generation Sequencing Core Facility at Argonne National Laboratory. Negative controls were included in every set of amplifications. To avoid potential lane bias, the same DNA aliquots of six samples from the first lane (MPR108, MPR114, MPR120, MPR135, MPR138 and MPR144) were run on the second lane. Lane 1 included all the wild-caught populations, and Lane 2 included all the laboratory-reared populations plus these six controls.

2.3 | Mouse exome data

We used exome data from Phifer-Rixey et al. (2018). That study is based on the same individuals used in this study. Briefly, DNA was extracted from frozen liver, kidney or spleen. Genomic libraries were enriched for mouse exons using a NimbleGen in-solution capture array (SeqCap EZ) and sequenced using 100-bp pair-end Illumina HiSeq2000. After quality filtering and SNP discovery, we further filtered the SNPs to only include those with a minor allele frequency of 5% or greater. This resulted in 279,278 SNPs. Each SNP was annotated to a single gene or multiple genes using Variant Effect Predictor in Ensembl. Additional details of quality filtering and SNP discovery are given in Phifer-Rixey et al. (2018).

2.4 | 16S data processing

We processed all of the 16S data in QIIME version 1.9.0 (Caporaso et al., 2010). The forward reads were demultiplexed and quality-filtered using default parameters using *split_libraries_fastq.py*. Chimeric sequences were removed using USEARCH 6.1 (Edgar, Haas, Clemente, Quince, & Knight, 2011). A subsampled open-reference OTU picking approach (*pick_open_reference_otus.py*) was employed with default parameters. OTUs at 97% similarity were generated using UCLUST (Edgar, 2010), and taxa were assigned based on the GREENGENE database 13.8 (DeSantis et al., 2006). To remove sequence errors and very rare OTUs, OTUs with <10 reads across all samples were removed. A phylogenetic tree was created using FASTTREE (Price, Dehal, & Arkin, 2009). The OTU table was rarefied to an even depth of 5,000 reads. Two samples (FL08M1 and FL08M2) were removed from all analyses due to low sequence reads (<200 reads).

Despite rarefying the reads to equal depth for all samples, the OTU counts were consistently higher in Lane 1 compared to Lane 2 for the six control samples (Figure S1). This lane bias is likely due to the greater average sequence depth of Lane 2 (68,196 reads per sample) compared to Lane 1 (13,918 reads per sample) resulting in an excess of rare OTUs in Lane 2. To account for this, we removed rare OTUs from Lane 2 to normalize the OTU counts between lanes before rarefaction (i.e., OTUs with a relative abundance $<8.0 \times 10^{-6}$ were removed) (Figure S1). Since Lane 1 included all wild-caught mice and Lane 2 included all laboratory-reared mice, all conclusions derived from comparisons within these groups (essentially all major conclusions; see Results2.2) were not affected by lane bias. Moreover, conclusions drawn from comparisons between wild-caught and laboratory-reared mice remained the same with or without correcting for lane bias (see Results2.2). The OTU table corrected for lane bias was used for all analyses presented below.

2.5 | Statistical analysis

We calculated beta-diversity measurements (i.e., Bray–Curtis dissimilarity, Binary-Sorensen-Dice, unweighted and weighted UniFrac distances) among all individuals using *beta_diversity.py* in QIIME. Pairwise distances for geography (km) were calculated based on GPS coordinates of the sampling locations. Pairwise distances for host genomes were calculated based on the exome data (~280,000 SNPs) using ngsDist (Vieira, Lassalle, Korneliussen, & Fumagalli, 2016), which takes into account uncertainty of the genotype calls. We used Mantel tests to test for correlations between beta-diversity and eight predictor variables (genetic distance, geographic distance, body weight, BMI, diet (δ^{13}C and δ^{15}N), climate PC1 and climate PC2) among all 50 wild-caught mice. We used Partial Mantel tests to ask whether host genetic distance is independently associated with Bray–Curtis dissimilarity while controlling for the effects of other variables. We also made comparisons among individuals within populations using Spearman's rho correlation. Specifically, we compared Bray–Curtis dissimilarity and geographic distance and we also compared Bray–Curtis dissimilarity and genetic distance. We also assessed the effects of geography while controlling for genetics (and the effects of genetics while controlling for geography) by comparing the residuals in a covariate regression with Bray–Curtis dissimilarity. Similarly, to test for correlations between beta-diversity and body size measurements, Spearman's rho correlation with residuals between body size and latitude was used to control for the known effect of latitude on body size (Table S4).

We calculated Bray–Curtis dissimilarity separately among mice within the northern-wild population (NH-VT), the southern-wild population (FL), the northern-laboratory population (NY), and the southern-laboratory population (FL). Similarly, Bray–Curtis dissimilarity was calculated separately between the two wild populations and between the two laboratory populations. We tested whether Bray–Curtis dissimilarity between population comparisons were significantly greater than within-population comparisons in both wild-caught and laboratory-reared individuals using Wilcoxon permutation tests based on 9,999 Monte Carlo resampling with the "Wilcox_test" function in the R package "coin". We calculated alpha-diversity using *alpha_diversity.py* and relative abundances of bacterial taxa using *summarize_taxa.py* using the rarefied OTU table. We used phylogenetic diversity (Faith, 1992) as an alpha-diversity measurement, and we focused on the relative abundances of 17 bacterial genera that were present in at least 50% of the individuals and had an average relative abundance of >1% across all individuals. The motivation for selecting a small set of common bacterial genera is both statistical (e.g., to minimize 0 values and multiple testing) and biological (e.g., narrower taxonomic groups [e.g.,

genera] have been shown to be associated with host genomic regions better than broader taxonomic groups [e.g., classes or phyla] (Benson et al., 2010)). Wilcoxon tests and Kruskal–Wallis tests were used for all pairwise and group comparisons unless otherwise stated.

2.6 | Wild mouse mGWAS and overlap with human mGWAS

We used a multivariate linear mixed model for association tests in GEMMA (version 0.94) using the exome (279,278 SNPs). The phylogenetic diversity measure and the relative abundances of 17 bacterial genera described above were Box-Cox-transformed following Goodrich, Davenport, Beaumont, et al. (2016) using the "PowerTransform" function in the R package "car". A multiple linear regression was used on the transformed microbial measurements to regress out the covariates including population structure, latitude and hidden factors. We calculated population structure using SNPRELATE (version 1.10.2) and used the first four genetic principal components (which together explain 18.5% of the genetic variation). Latitude was also used as a covariate to control for bacteria that vary latitudinally (Thompson et al., 2017). Hidden factors were calculated to account for experimental cofounders and batch effects

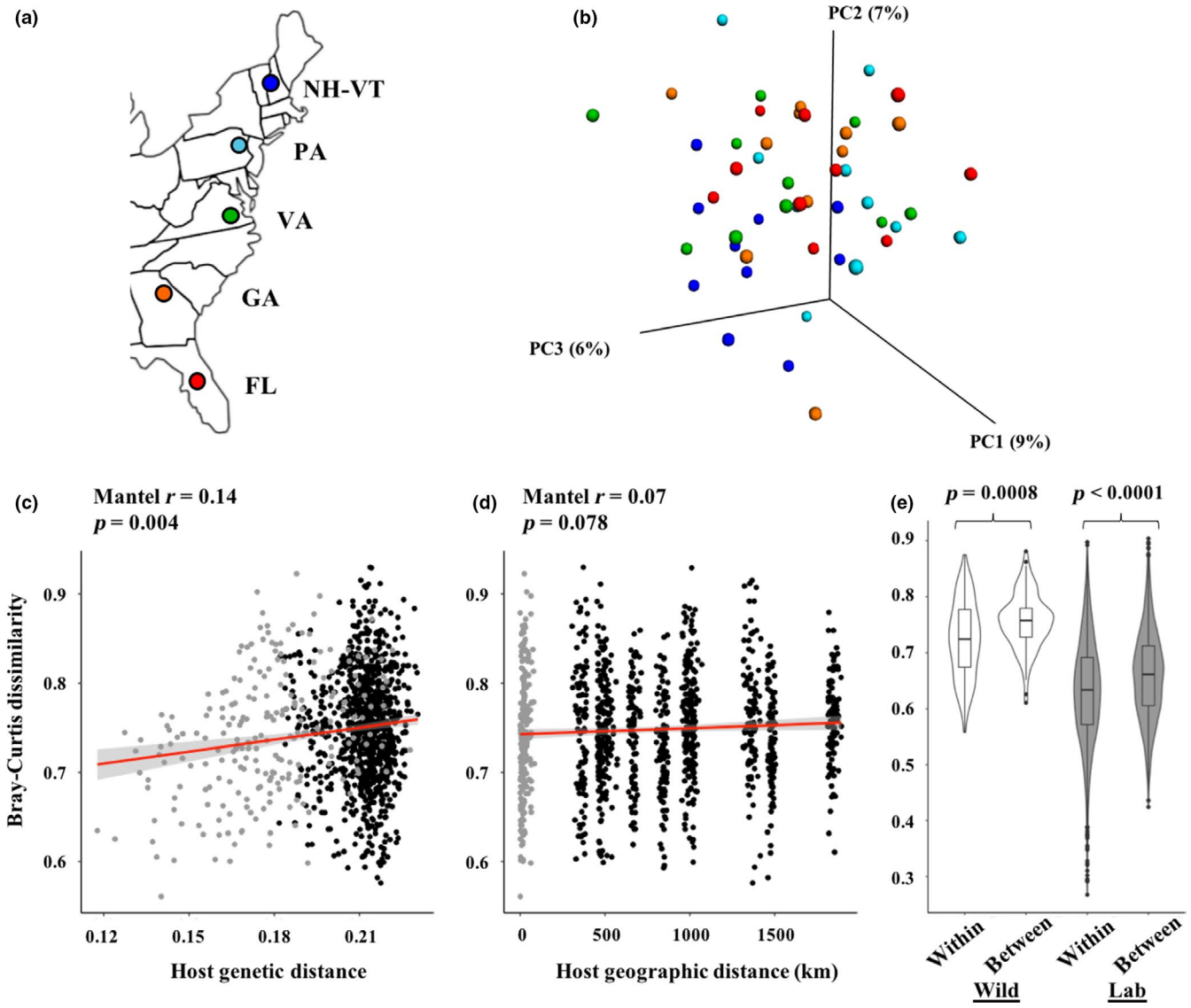

FIGURE 1 Population differences and host genetic distance are associated with compositional variation in the gut microbiota. (a) Sampling locations of five house mouse populations. (b) PCoA plot of Bray–Curtis dissimilarity. The colour corresponds to populations in (a). Populations show weak, but significant clustering (ADONIS R^2 = 0.095, p = 0.04). (c) A significant positive correlation between microbial distance and host genetic distance (calculated by ngsDist) and (d) a nonsignificant correlation between microbial distance and geographic distance using all individuals. Correlations for within-population comparisons (grey points) and between-population comparisons (black dots) are shown. (e) Violin box plots of Bray–Curtis dissimilarity within population comparisons (within) and between population comparisons (between) of the most northern (NH-VT) and southern (FL) populations in the wild (white bars) and laboratory (grey bars). p-values are Wilcoxon permutation tests

by inferring 10 cofounders in PEER (Stegle, Parts, Durbin, & Winn, 2010). We also accounted for relatedness by using a relatedness matrix estimated in GEMMA. Manhattan plots and QQ plots were generated using the R package "qqman." To control for false discovery, *q*-values were calculated using the R package "*q*-value" based on likelihood ratio *p*-values. Significant SNPs were called at a *q*-value <0.1. The functions of nonsynonymous SNPs were predicted using mouse genome version GRCm38 in SIFT (Sim et al., 2012).

To test for overlap of bacteria-associated genes identified in this study and previously published human mapping studies, we identified significant SNPs with a *q*-value <0.2 to increase power. We compiled SNPs and genes that were associated with microbial measurements (e.g., relative abundance of taxa, alpha- or beta-diversity) in seven human mGWAS studies (Table S5). To test whether the overlap of candidate genes in humans and mice was significant, we used a hypergeometric test with the "phyper" function in R. We used the total number of possible human-mouse orthologous genes in our exome (19,100) based on the Ensembl database (Sep 2017). To test whether the proportion of significant gene–bacteria associations was greater in the candidate gene set compared to all genes using the entire exome, we used a chi-square test with Yate's correction.

To calculate effect sizes and to ask whether different genotypes have significantly different relative abundances of taxa or alpha-diversity measurements, we used ANOVA. Residuals after covariate regression (i.e., population structure, latitude and hidden factors) on the Box-Cox-transformed microbial measurements were used for the analyses. We used a sign test to ask whether the directionality of genotype–bacteria associations within populations deviates significantly from the expected 50:50 ratio.

3 | RESULTS

3.1 | Host genetic distance and body size are associated with compositional variation in the gut microbiota

We found significant differences in gut microbial communities among five populations of house mice in eastern North America (Figure 1a,b) based on Bray–Curtis dissimilarity (ADONIS, $R^2 = 0.095$, $p = 0.04$), Binary-Sorensen-Dice ($R^2 = 0.099$, $p = 0.002$), and unweighted UniFrac distance ($R^2 = 0.097$, $p = 0.003$), but not based on weighted UniFrac distance ($R^2 = 0.087$, $p = 0.34$). While these results are consistent with findings in European populations of house mice (Linnenbrink et al., 2013), the population differences observed in the present study were relatively small. To understand how host genetic and environmental factors contribute to variation in gut microbial communities across all samples, we measured correlations between Bray–Curtis dissimilarity and eight putative predictor variables using Mantel tests (Table 1). We found that Bray–Curtis dissimilarity was significantly correlated with both host genetic distance and body mass index (BMI) after correcting for multiple tests (Mantel $r = 0.14$, $p = 0.004$ and Mantel $r = 0.25$, $p < 0.0001$, respectively) (Table 1 and Figure 1c). Diet ($\delta^{13}C$), climate PC1 and geographic distance (Figure 1d) also showed weak correlations with Bray–Curtis dissimilarity, but they were not significant (Table 1). Diet measurements ($\delta^{13}C$ and $\delta^{14}N$) did not vary among populations (Figure S1) or by latitude (Table S4). The overall results were similar using other beta-diversity measurements (Table S6).

The observed correlation between microbial distance and genetic distance was independent of other variables including geographic distance (Table S7). First, there was no pattern of genetic isolation by distance among these populations (Figure S2). Second, the correlation between host genetic distance and Bray–Curtis dissimilarity remained significant after controlling for seven predictor variables, including geographic distance, using Partial Mantel tests (Table S7). When comparisons were made between individuals within populations, both host genetic distance and geographic distance showed significant correlations with Bray–Curtis dissimilarity (Figure S3). Consistent with the results among populations, we found that the correlation between host genetic distance and Bray–Curtis dissimilarity within populations remained significant after controlling for geographic distance using residuals of covariate regression (Spearman's $\rho = 0.26$, $p < 0.0001$). In contrast, we found that the correlation between geographic distance and Bray–Curtis dissimilarity did not remain significant after controlling for genetic distance (Spearman's $\rho = 0.06$, $p = 0.38$). These results suggest that host genetics and/or vertical transmission have stronger effects on the gut microbiota than the geographic distance between individuals.

The correlation between BMI and Bray–Curtis dissimilarity also remained significant after controlling for geographic distance (partial mantel $r = 0.25$, $p < 0.001$). This association is interesting because BMI and body weight vary clinally with latitude (Table S4), a pattern consistent with Bergmann's rule (Bergmann, 1847) and presumably reflecting thermoregulatory adaptation (Lynch, 1992; Phifer-Rixey et al., 2018). Moreover, at higher latitudes, individuals tended to have a greater ratio of Firmicutes to Bacteroidetes, which is associated with obesity in humans (Ley, Turnbaugh, Klein, & Gordon, 2006) and in laboratory mice (Ley et al., 2005; Turnbaugh et al., 2006) (Figure

TABLE 1 Correlations between microbial beta-diversity and predictor variables using Mantel test

		Bray–Curtis dissimilarity	
Predictor variables	***n***	**Mantel *r***	***p*-value[a]**
Genetic distance (ngsDist)	50	**0.14**	**0.004**
Geographic distance (km)	50	0.07	0.078
Body weight (g)	50	0.08	0.232
BMI (g/mm^2)	50	**0.25**	**<0.001**
Diet ($\delta^{13}C$)	50	0.14	0.044
Diet ($\delta^{15}N$)	50	0.04	0.67
ClimatePC1	50	0.10	0.066
ClimatePC2	50	0.01	0.913

[a]Bolded values show significance below Bonferroni-corrected *p*-value = 0.05/8 = 0.0063.

S4). We also identified various microbial taxa that correlated with latitude and body size after accounting for latitude (Figure S5). For example, the phylum Proteobacteria was the taxon whose relative abundance was most positively correlated with latitude (ρ = 0.443, p < 0.01). The phylogenetic diversity of the microbial community and the relative abundance of genus *Odoribacter* were most positively correlated with body weight after correcting for latitude (ρ = 0.410, p < 0.01 and ρ = 0.412, p < 0.01, respectively).

3.2 | Population differences in the microbiota persist in a common laboratory environment

We found a significant difference in Bray–Curtis dissimilarity among wild mice in comparisons within vs. between populations from the ends of the transect (NH-VT and FL) (Figure 1e), indicating that mice from these populations harbour compositionally distinct microbial communities. To test whether population differences in the microbiota were driven by environmental differences, we collected live animals close to the most northern and southern populations (NY and FL) and conducted a common garden experiment. Twenty unrelated wild mice were collected from each population and returned to the laboratory (10 males, 10 females). For each population, we created 10 crosses between wild-caught parents to produce 40 offspring which were reared under identical conditions. The laboratory-born mice showed major shifts in alpha-diversity and in the relative abundances of bacterial phyla and genera compared to the wild-caught animals (Table S8). For example, alpha-diversity measurements and the relative abundances of Firmicutes and Proteobacteria significantly decreased, and the relative abundance of Bacteroidetes significantly increased in laboratory-reared animals compared to wild-caught animals (Phylogenetic diversity and Shannon index, p < 0.001; all phyla, p < 0.001) (Table S8). Although there is a potential for batch effects since the microbiota of wild-caught and laboratory-reared individuals were sequenced on separate lanes, the same results were obtained with or without correcting for lane bias (Shannon index, p < 0.001; all phyla, p < 0.001). Interestingly, diet as assessed by stable isotopes showed some overlap between wild and laboratory mice (Figure S6). The laboratory diet was significantly different from that of the wild as assessed by δ^{15}N (laboratory mean: 6.5, wild mean: 7.1, p < 0.05) but not as assessed by δ^{13}C (laboratory mean: −19.0, wild mean: 19.1, p = 0.12) (Figure S6).

Despite the dramatic shifts in the microbiota from the wild to the laboratory environment (Figure S7), population differences in the microbiota persisted among laboratory-reared offspring (Figure 1e). Compositional differences in the microbiota both within and between populations were reduced in the laboratory setting compared to the wild (Figure 1e). However, the microbial community composition of laboratory populations was more similar, on average, to the wild populations from which they came than to the wild populations at the other end of the transect (Wilcoxon permutation test, p = 0.029, Figure S8). Overall, these results indicate that environmental differences alone (e.g., diet, temperature, etc.) cannot fully explain the population differences in the microbiota. The observed population differences are consistent either with a role for host genetics or simply with vertical transmission shaping the variation of the gut microbiota in wild mice.

3.3 | Identification of genetic loci underlying gut microbiota variation in wild mice

To identify host genes contributing to differences in the gut microbiota, we conducted a mGWAS using ~280,000 SNPs identified from sequencing the complete exomes of the 50 wild-caught mice (Phifer-Rixey et al., 2018). We searched for associations between host genetic variation and the relative abundances of 17 bacterial genera that were common (an average relative abundance of >1% and present in >50% of all individuals). We also searched for associations between host genetic variation and alpha-diversity represented by phylogenetic diversity. Analyses were done using multivariate linear mixed models in GEMMA while controlling for host population structure, relatedness, latitude and hidden factors. Among the 18 bacterial measurements, two bacteria genera (*Odoribacter* and *Bacteroides*) and phylogenetic diversity showed significant associations with host genetic loci (Table 2). Across all tests, we identified a total of 24 SNPs in 20 genes that passed a genome-wide significance threshold (q-value <0.1). Although none of the GO terms were significantly overrepresented after false discovery correction, the top three GO terms include mRNA transcription (*Mier1*, p = 0.006), protein lipidation (*Zdhhc7*, p = 0.04) and nucleobase-containing compound transport (*Slc35d1*, p = 0.07).

Although we attempted to account for population structure in identifying these genes using GEMMA, observed gene–bacteria associations might still be driven by differences among populations that are not fully accounted for by the model. To further account for population structure, we first looked at associations within individual populations and then examined whether the direction of the association was consistent among populations. Overall, most of the within-population genotype–bacteria comparisons showed the same direction as the all-population comparisons (37 out of 43 comparisons, sign test p < 0.0001, Table S9). Moreover, 20 of these 37 comparisons were individually significant (ANOVA, p < 0.05) despite the fact that these tests are underpowered with only 10 individuals per population (Table S9). Together, the results suggest that the observed genotype–bacteria associations are unlikely to be explained by population structure.

Among the 20 genes that were associated with bacterial measurements, the interleukin 12a gene (*IL12a*) included a SNP with the lowest p-value across all tests in this study (Table 2). *IL12* is a cytokine that plays a key role in innate and adaptive immunity by activating natural killer cells and regulating differentiation of T cells (Trinchieri, 1998). We identified six SNPs in *IL12a* that were significantly associated with the relative abundance of *Odoribacter* after accounting for population structure, latitude, and hidden factors as covariates (Figure 2a–d, Table 2). Three of these six SNPs were non-synonymous changes, and one of these (3_68695333) was predicted in silico to be deleterious (SIFT score = 0.002) (Table 2).

TABLE 2 Loci significantly associated with the relative abundances of bacterial taxa or phylogenetic diversity (q-value < 0.1) in wild mouse mGWAS

Chr	Bp	Annotated gene(s)	Associated microbial measurements	p-values[a]	Q-values	Effect size (%)[b]	Missense variant[c]
3	68695209	*Il12a*	*Odoribacter*	8.41E-07	0.042	23.1	Thr - Ser
3	68695333			1.10E-06	0.045	19.9	Ser - Met *
3	68695379			5.37E-07	0.033	18.0	–
3	68695382			5.37E-07	0.033	18.0	–
3	68695502			4.13E-07	0.033	22.2	–
3	68695548			7.79E-09	0.002	24.4	Gly - Ser
3	86138475	*Snord73a, Rnu73b, Rps3a1*	*Bacteroides*	1.77E-06	0.054	18.1	–
3	86138574			6.90E-07	0.028	18.4	–
3	86138625			2.00E-06	0.054	29.8	–
4	103170679	*Mier1, Slc35d1*	Phylogenetic diversity	2.19E-06	0.092	5.4	–
4	89692441	*Dmrta1*	*Bacteroides*	4.73E-07	0.023	27.9	–
5	90490831	*Afp*	*Bacteroides*	1.27E-08	0.002	27.0	–
5	90490846			1.27E-08	0.002	27.0	–
5	90491657			8.67E-08	0.007	17.5	–
6	121221243	*Tuba8, Gm15856*	*Bacteroides*	4.19E-06	0.086	23.4	–
6	121222841			3.03E-06	0.074	23.3	–
6	128374454	*Foxm1, Tex52*	Phylogenetic diversity	4.09E-07	0.051	33.3	–
6	128374521			4.09E-07	0.051	33.3	–
6	128374742			9.63E-07	0.055	34.0	–
8	120092803	*Zdhhc7, Gm20388, Gm15898*	*Bacteroides*	3.38E-07	0.021	38.2	–
8	13142468	*Cul4a*	*Bacteroides*	3.56E-06	0.079	22.5	–
8	16358320	*Csmd1*	Phylogenetic diversity	1.09E-06	0.055	17.0	–
11	3132802	*Sfi1, Pisd-ps1*	Phylogenetic diversity	8.39E-07	0.055	12.0	–
13	33671503	*Serpinb6d*	*Bacteroides*	1.38E-06	0.048	31.7	–

[a]Likelihood ratio p-values.
[b]ANOVA R^2 values. Residuals after covariate regression (i.e., Genetic PC1-4 and Latitude) on the Box-Cox-transformed relative abundance of bacterial taxa was used.
[c]Amino acid changes are shown; aCt/aGt (threonine/serine), aCg/aTg (serine/methionine) and Ggc/Agc (glycine/serine). * indicate significant deleterious changes predicted by SIFT (SIFT score <0.05).

3.4 | Homologous genes underlie gut microbiota variation in humans and mice

A common set of genes may underlie host–bacterial interactions across diverse mammals. To test this idea, we asked whether there was significant overlap between the genes underlying variation in the microbiota of mice and humans using two different approaches. First, we compiled genes that were associated with microbial measurements (e.g., relative abundance of taxa, alpha- or beta-diversity) in seven different human mGWAS. This comprised a set of 469 genes with one-to-one mouse-human orthologs (Table S5). We then conducted association analyses in GEMMA using this set of 469 genes in mice and found that 10 were significantly associated with one or more bacterial measurements (q-value < 0.1) (Table S10). This fraction of genes showing associations (10 out of 469 = 2.13%) is significantly greater than the fraction discovered in the initial analysis using all genes (20 out of 21,954 = 0.09%), suggesting that mouse mGWAS hits are overrepresented among genes previously identified in human mGWAS (chi-square test with Yate's correction p < 0.0001).

Second, we asked how many genes overlapped between the 469 genes identified in human mGWAS and the 20 mouse-human orthologous genes that were identified in the mouse mGWAS. Using the genome-wide cut-off of q-value < 0.1, there was only one gene, *Csmd1* that overlapped between these sets, and this degree of overlap was marginally not significant (hypergeometric test p = 0.06). However, when we made the genome-wide cut-off less stringent (q-value < 0.2), we identified 96 mouse-human orthologous genes

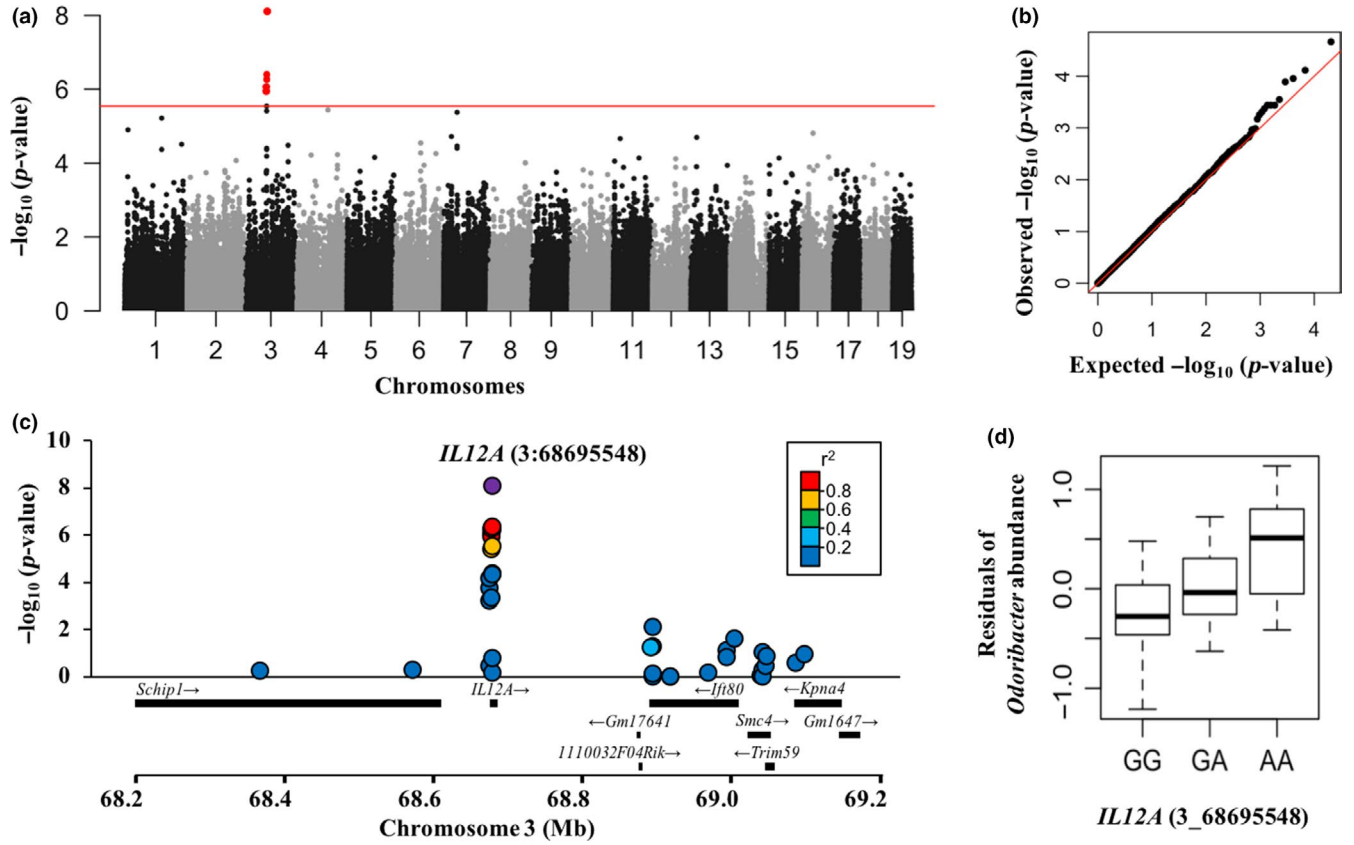

FIGURE 2 Results of mGWAS in wild mice. (a) Manhattan plot of *Odoribacter*. Six SNPs in *IL12A* on chromosome 3 are highlighted. Red line shows the genome-wide cut-off of the likelihood ratio test *p*-value (*q*-value < 0.1). (b) Quantile–quantile plot of *Odoribacter* *p*-values. The red diagonal line represents the expected distributions of *p*-values. (c) Zoom in plot around the SNP (3:68695548) that has the lowest *p*-value (represented by purple). Each dot is a SNP, and colours represent pairwise linkage disequilibrium measures (r^2) between the SNP (3:68695548) and SNPs within the surrounding 1 Mb window. r^2 was calculated based on genotype allele counts using PLINK. (d) Box plot of *Odoribacter* abundance and *IL12A* genotypes. A missense SNP in *IL12A* gene has significantly different abundances of *Odoribacter* (ANOVA $R^2 = 0.224$, $p = 0.0014$). Residual values were used for the y-axis controlling for population structure, latitude, and hidden factors using covariate regression. Whiskers indicate highest and lowest values

(Table S11) and eight genes overlapped with 469 human candidate genes (Table S12). The number of overlapped genes identified at *q*-value < 0.2 was greater than expected by chance (hypergeometric test $p = 0.0006$). Not surprisingly, all eight genes were also identified by the candidate gene approach mentioned above (Tables S10 and S12). Among the eight homologous overlapping genes between human and mouse mGWAS, all show expression in the brain of mice and humans (Table S12) and some have been associated with phenotypes related to obesity and immunity in other human GWAS.

4 | DISCUSSION

Understanding how mammalian hosts maintain the composition and function of the gut microbiota remains a major challenge in microbial ecology and biomedical research. We showed that differences in host genetics and body mass were significantly associated with compositional differences in the microbiota of wild mice. We did not find significant associations between the microbiota and diet as inferred by carbon and nitrogen stable isotopes. However, fine-scale dietary differences that are not reflected in isotope measurements might also contribute to variation in the wild mouse gut microbiota. In other studies, host genetic distance (Blekhman et al., 2015; Ma et al., 2014), body size measurements (e.g., Ley et al., 2006; Turnbaugh et al., 2009) and diet (David et al., 2014) have all been associated with differences in the gut microbial composition in humans. We found that mouse populations from higher latitudes had a greater ratio of Firmicutes/Bacteroidetes, a pattern also seen among human populations (Suzuki & Worobey, 2014). This is interesting since a higher ratio of Firmicutes/Bacteroidetes is known to be associated with obesity in humans (Ley et al., 2006) and in laboratory mice (Ley et al., 2005; Turnbaugh et al., 2006). The mice sampled here conform to Bergmann's rule with mice from higher latitudes exhibiting larger body sizes than mice from lower latitudes (Lynch, 1992; Phifer-Rixey et al., 2018). Further experiments, including transplants into gnotobiotic mice, would be useful for testing the role of the gut microbiome, if any, in adaptive host body size variation.

To test whether population differences in the gut microbiota were due to the environment or to host genetics, we identified population differences in the gut microbiota that persisted in a common

laboratory environment. Consistent with previous studies of wild mice raised in captivity (Wang et al., 2015, 2014), we observed a decrease in alpha-diversity and in the relative abundances of Firmicutes and Proteobacteria, and an increase in the relative abundance of Bacteroidetes, in laboratory-reared animals compared to wild-caught animals. Despite the changes in gut microbial composition that occurred when progeny of wild mice were raised in the laboratory, population differences persisted among the laboratory-reared mice. Moreover, the gut microbiota of laboratory-reared mice resembled the gut microbiota of the population of origin more than of the population at the other end of the transect. Similar patterns have also been found in other wild-derived mouse strains where population differences in the field were maintained in captivity for over 10 generations (Moeller, Suzuki, Phifer-Rixey, & Nachman, 2018). These observations suggest that environmental differences alone (e.g., diet, temperature) cannot explain the population differences in the microbiota. Instead, host genetics and/or vertical transmission must partly account for the observed population differences in the gut microbiota of wild mice. Experiments allowing the exchange of microbes between individuals from different populations could be used to directly test the effect of host genotype on the gut microbiota.

To explore the genetic basis of the gut microbiota, we identified both novel and previously known gene–bacteria associations in wild mice using a genome-wide mapping approach. The top association identified in this study was between the relative abundance of *Odoribacter* and a nonsynonymous SNP in *IL12a*, a cytokine that is involved in innate and adaptive immunity (Trinchieri, 1998). The upregulation of *IL12a* production has been linked to Crohn's disease in humans (Parronchi et al., 1997) and mucosal inflammation in mice (Liu et al., 2001). Furthermore, a recent study in humans demonstrated that inflammatory cytokine responses are associated with microbial taxa composition, metagenomic functional profiles and microbial metabolites (Schirmer et al., 2016). Interestingly, the relative abundance of *Odoribacter* was significantly correlated with tumour necrosis factor alpha (TNF-α) (Schirmer et al., 2016), which is another macrophage-derived cytokine that interacts with *IL12a* in mediating inflammatory responses in mammals (Ma, 2001). These observations lend further support to the role of *IL12a* in mediating host–microbial interactions in wild mice.

Finally, a significantly greater number of genes overlapped between human mGWAS and mouse mGWAS than expected by chance, including genes related to the nervous system, immunity and obesity. For example, a SNP in *Csmd1* is associated with alpha-diversity (i.e., phylogenetic diversity) in mice and showed the lowest *p*-value among the eight human-mouse overlapping genes. *Csmd1* is highly expressed in the central nervous system and in epithelial tissue and is involved in regulating the development of the central nervous system (Kraus et al., 2006). In humans, *Csmd1* is associated with beta-diversity of the gut microbiota (Wang et al., 2016), obesity-related traits (Comuzzie et al., 2012; Irvin et al., 2011; Liu et al., 2013), parasite infection status (Deng et al., 2013) and antibody response to smallpox vaccine (Ovsyannikova et al., 2012). Similarly, *Gpr158* is also highly expressed in mouse and human brains and is associated with bacterial taxa in the order Clostridiales in both wild mice and humans (Goodrich, Davenport, Beaumont, et al., 2016). *Gpr158* has been associated with variation in energy expenditure in a native American population that has a high prevalence of obesity (Piaggi et al., 2017). These results suggest host genes related to the nervous system, immunity and obesity may underlie gut microbial variation across diverse mammalian species.

In conclusion, we presented evidence that the host genome affects gut microbial composition within and between populations of wild mice using field observations and laboratory experiments. Gene–bacteria associations identified in wild mice and humans using similar mapping methods are strong candidates for genes influencing the mammalian gut microbial composition in a natural environment. Replicating these results in independent populations of wild mice and validating the functions of candidate SNPs in wild-derived inbred mice would further strengthen the observed gene–bacteria associations.

ACKNOWLEDGEMENTS

We thank Andrew Moeller, Mohamed Noor, the members of the Nachman laboratory, and three anonymous reviewers for discussions and comments on the manuscript. This work was supported by NIH grants (RO1 GM074245, RO1 GM127468) to MWN and an NSF grant (1501646) to TAS.

AUTHOR CONTRIBUTIONS

T.A.S. and M.W.N. designed the project. M.P.R. collected wild mouse tissues. M.J.S. and D.L. collected live mice and maintained wild-derived inbred lines. K.B. and M.P.R. generated exome data. T.A.S. generated 16S data. T.A.S. conducted common garden experiments and all analyses. K.L.M. assisted mGWAS data analyses. T.A.S. wrote the initial draft of the manuscript, and all authors contributed to revisions. M.W.N. supervised the project.

DATA ACCESSIBILITY

All museum specimens (skins and skulls) were prepared and have been deposited in the mammal collections of the Museum of Vertebrate Zoology at the University of California, Berkeley, with ancillary data uploaded to ARCTOS (see Table S1 for accession numbers). All 16S rRNA sequence data were uploaded to European Molecular Biology Laboratory, European Nucleotide Archive (ENA) database (PRJEB32701). The mouse exome sequence data are available under NCBI SRA (PRJNA397406) as described in Phifer-Rixey et al. (2018).

ORCID

Taichi A. Suzuki https://orcid.org/0000-0001-7800-8596
Michael J. Sheehan https://orcid.org/0000-0002-3949-7873
Dana Lin https://orcid.org/0000-0002-7952-2359

REFERENCES

Benson, A. K., Kelly, S. A., Legge, R., Ma, F., Low, S. J., Kim, J., … Pomp, D. (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(44), 18933–18938. https://doi.org/10.1073/pnas.1007028107

Bergmann, C. (1847). Uber die verhaltnisse der warmeokonomie der thiere zu ihrer grosse. *Göttinger Studien*, *3*(1), 595–708.

Blekhman, R., Goodrich, J. K., Huang, K., Sun, Q. I., Bukowski, R., Bell, J. T., … Clark, A. G. (2015). Host genetic variation impacts microbiome composition across human body sites. *Genome Biology*, *16*(1), 191. https://doi.org/10.1186/s13059-015-0759-1

Bonder, M. J., Kurilshikov, A., Tigchelaar, E. F., Mujagic, Z., Imhann, F., Vila, A. V., … Zhernakova, A. (2016). The effect of host genetics on the gut microbiome. *Nature Genetics*, *48*(11), 1407–1412. https://doi.org/10.1038/ng.3663

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., … Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*(5), 335–336. https://doi.org/10.1038/nmeth.f.303

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., … Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, *6*(8), 1621–1624. https://doi.org/10.1038/ismej.2012.8

Comuzzie, A. G., Cole, S. A., Laston, S. L., Voruganti, V. S., Haack, K., Gibbs, R. A., & Butte, N. F. (2012). Novel genetic loci identified for the pathophysiology of childhood obesity in the hispanic population. *PLoS ONE*, *7*(12), e51954.https://doi.org/10.1371/journal.pone.0051954

Davenport, E. R., Cusanovich, D. A., Michelini, K., Barreiro, L. B., Ober, C., & Gilad, Y. (2015). Genome-wide association studies of the human gut microbiota. *PLoS ONE*, *10*(11), e0140301. https://doi.org/10.1371/journal.pone.0140301

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., … Turnbaugh, P. J. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, *505*(7484), 559–563. https://doi.org/10.1038/nature12820

Deng, X., Sabino, E. C., Cunha-Neto, E., Ribeiro, A. L., Ianni, B., Mady, C., … Component, I. (2013). Genome wide association study (GWAS) of chagas cardiomyopathy in trypanosoma cruzi seropositive subjects. *PLoS ONE*, *8*(11), 4–10. https://doi.org/10.1371/journal.pone.0079629

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., … Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, *72*(7), 5069–5072. https://doi.org/10.1128/AEM.03006-05

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, *27*(16), 2194–2200. https://doi.org/10.1093/bioinformatics/btr381

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*, *61*(1), 1–10. https://doi.org/10.1016/0006-3207(92)91201-3

Goodrich, J. K., Davenport, E. R., Beaumont, M., Jackson, M. A., Knight, R., Ober, C., … Ley, R. E. (2016). Genetic determinants of the gut microbiome in UK twins. *Cell Host and Microbe*, *19*(5), 731–743. https://doi.org/10.1016/j.chom.2016.04.017

Goodrich, J. K., Davenport, E. R., Waters, J. L., Clark, A. G., & Ley, R. E. (2016). Cross-species comparisons of host genetic associations with the microbiome. *Science*, *352*(6285), 29–32. https://doi.org/10.1126/science.aad9379

Goodrich, J. K., Waters, J. L., Poole, A. C., Sutter, J. L., Koren, O., Blekhman, R., … Ley, R. E. (2014). Human genetics shape the gut microbiome. *Cell*, *159*(4), 789–799. https://doi.org/10.1016/j.cell.2014.09.053

Hall, A. B., Tolonen, A. C., & Xavier, R. J. (2017). Human genetic variation and the gut microbiome in disease. *Nature Reviews Genetics*, *18*(11), 690–699. https://doi.org/10.1038/nrg.2017.63

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978. https://doi.org/10.1002/joc.1276

Hsiao, E. Y., McBride, S. W., Hsien, S., Sharon, G., Hyde, E. R., McCue, T., … Mazmanian, S. K. (2013). Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell*, *155*(7), 1451–1463. https://doi.org/10.1016/j.cell.2013.11.024

Irvin, M. R., Wineinger, N. E., Rice, T. K., Pajewski, N. M., Kabagambe, E. K., Gu, C. C., … Arnett, D. K. (2011). Genome-wide detection of allele specific copy number variation associated with insulin resistance in African Americans from the hyperGEN study. *PLoS ONE*, *6*(8), 9–11. https://doi.org/10.1371/journal.pone.0024052

Knights, D., Silverberg, M. S., Weersma, R. K., Gevers, D., Dijkstra, G., Huang, H., … Xavier, R. J. (2014). Complex host genetics influence the microbiome in inflammatory bowel disease. *Genome Medicine*, *6*, 107. https://doi.org/10.1186/s13073-014-0107-1

Kraus, D. M., Elliott, G. S., Chute, H., Horan, T., Pfenninger, K. H., Sanford, S. D., … Holers, V. M. (2006). CSMD1 is a novel multiple domain complement-regulatory protein highly expressed in the central nervous system and epithelial tissues. *The Journal of Immunology*, *176*(7), 4419–4430. https://doi.org/10.4049/jimmunol.176.7.4419

Leamy, L. J., Kelly, S. A., Nietfeldt, J., Legge, R. M., Ma, F., Hua, K., … Pomp, D. (2014). Host genetics and diet, but not immunoglobulin A expression, converge to shape compositional features of the gut microbiome in an advanced intercross population of mice. *Genome Biology*, *15*(12), 552. https://doi.org/10.1186/s13059-014-0552-6

Ley, R. E., Bäckhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., & Gordon, J. I. (2005). Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(31), 11070–11075. https://doi.org/10.1073/pnas.0504978102

Ley, R. E., Turnbaugh, P. J., Klein, S., & Gordon, J. I. (2006). Microbial ecology: Human gut microbes associated with obesity. *Nature*, *444*(7122), 1022–1023. https://doi.org/10.1038/4441022a

Linnenbrink, M., Wang, J., Hardouin, E. A., Künzel, S., Metzler, D., & Baines, J. F. (2013). The role of biogeography in shaping diversity of the intestinal microbiota in house mice. *Molecular Ecology*, *22*, 1904–1916. https://doi.org/10.1111/mec.12206

Liu, C.-T., Monda, K. L., Taylor, K. C., Lange, L., Demerath, E. W., Palmas, W., … Fox, C. S. (2013). Genome-wide association of body fat distribution in African ancestry populations suggests new loci. *PLoS Genetics*, *9*(8), e1003681. https://doi.org/10.1371/journal.pgen.1003681

Liu, Z., Geboes, K., Heremans, H., Overbergh, L., Mathieu, C., Rutgeerts, P., & Ceuppens, J. L. (2001). Role of interleukin-12 in the induction of mucosal inflammation and abrogation of regulatory T cell function in chronic experimental colitis. *European Journal of Immunology*, *31*(5), 1550–1560. https://doi.org/10.1002/1521-4141(200105)31:5<1550:AID-IMMU1550>3.0.CO;2-3

Lynch, C. (1992). Clinal variation in cold adaptation in *Mus domesticus*: Verification of predictions from laboratory populations. *American Naturalist*, *139*(6), 1219–1236. https://doi.org/10.1086/285383

Ma, J., Coarfa, C., Qin, X., Bonnen, P. E., Milosavljevic, A., Versalovic, J., & Aagaard, K. (2014). mtDNA haplogroup and single nucleotide polymorphisms structure human microbiome communities. *BMC Genomics*, *15*(1), 257. https://doi.org/10.1186/1471-2164-15-257

Ma, X. (2001). TNF-α and IL-12:A balancing act in macrophage functioning. *Microbes and Infection*, *3*(2), 121–129. https://doi.org/10.1016/S1286-4579(00)01359-9

Marchesi, J. R., Holmes, E., Khan, F., Kochhar, S., Scanlan, P., Shanahan, F., … Wang, Y. (2007). Rapid and noninvasive metabonomic characterization of inflammatory bowel disease. *Journal of Proteome Research*, *6*(2), 546–551. https://doi.org/10.1021/pr060470d

McFall-Ngai, M., Hadfield, M. G., Bosch, T. C. G., Carey, H. V., Domazet-Lošo, T., Douglas, A. E., … Wernegreen, J. J. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(9), 3229–3236. https://doi.org/10.1073/pnas.1218525110

McKnite, A. M., Perez-Munoz, M. E., Lu, L. U., Williams, E. G., Brewer, S., Andreux, P. A., … Ciobanu, D. C. (2012). Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. *PLoS ONE*, *7*(6), e39191. https://doi.org/10.1371/journal.pone.0039191

Moeller, A. H., Suzuki, T. A., Phifer-Rixey, M., & Nachman, M. W. (2018). Transmission modes of the mammalian gut microbiota. *Science*, *362*(6413), 453–457. https://doi.org/10.1126/science.aat7164

Org, E., Parks, B. W., Joo, J. W. J., Emert, B., Schwartzman, W., Kang, E. Y., … Lusis, A. J. (2015). Genetic and environmental control of host-gut microbiota interactions. *Genome Research*, *25*(10), 1558–1569. https://doi.org/10.1101/gr.194118.115

Ovsyannikova, I. G., Kennedy, R. B., O'Byrne, M., Jacobson, R. M., Pankratz, V. S., & Poland, G. A. (2012). Genome-wide association study of antibody response to smallpox vaccine. *Vaccine*, *30*(28), 4182–4189. https://doi.org/10.1016/j.vaccine.2012.04.055

Parronchi, P., Romagnani, P., Annunziato, F., Sampognaro, S., Becchio, A., Giannarini, L., … Romagnani, S. (1997). Type 1 T-helper cell predominance and interleukin-12 expression in the gut of patients with Crohn's disease. *The American Journal of Pathology*, *150*(3), 823–832.

Phifer-Rixey, M., Bi, K. E., Ferris, K. G., Sheehan, M. J., Lin, D., Mack, K. L., … Nachman, M. W. (2018). The genomic basis of environmental adaptation in house mice. *PLoS Genetics*, *14*(9), e1007672. https://doi.org/10.1371/journal.pgen.1007672

Phifer-Rixey, M., & Nachman, M. W. (2015). Insights into mammalian biology from the wild house mouse *Mus musculus*. *Elife*, *2015*(4), 1–13. https://doi.org/10.7554/eLife.05959

Piaggi, P., Masindova, I., Muller, Y. L., Mercader, J., Wiessner, G. B., Chen, P., … Baier, L. J. (2017). A genome-wide association study using a custom genotyping array identifies variants in GPR158 associated with reduced energy expenditure in American Indians. *Diabetes*, *66*(8), 2284–2295. https://doi.org/10.2337/db16-1565

Price, M. N., Dehal, P. S., & Arkin, A. P. (2009). Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution*, *26*(7), 1641–1650. https://doi.org/10.1093/molbev/msp077

Rosshart, S. P., Vassallo, B. G., Angeletti, D., Hutchinson, D. S., Morgan, A. P., Takeda, K., … Rehermann, B. (2017). Wild mouse gut microbiota promotes host fitness and improves disease resistance. *Cell*, *171*(5), 1015–1028.e13. https://doi.org/10.1016/j.cell.2017.09.016

Schirmer, M., Smeekens, S. P., Vlamakis, H., Jaeger, M., Oosting, M., Franzosa, E. A., … Xavier, R. J. (2016). Linking the human gut microbiome to inflammatory cytokine production capacity. *Cell*, *167*(4), 1125–1136. https://doi.org/10.1016/j.cell.2016.10.020

Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*, *40*(W1), W452–W457. https://doi.org/10.1093/nar/gks539

Spor, A., Koren, O., & Ley, R. (2011). Unravelling the effects of the environment and host genotype on the gut microbiome. *Nature Reviews Microbiology*, *9*(4), 279–290. https://doi.org/10.1038/nrmicro2540

Stegle, O., Parts, L., Durbin, R., & Winn, J. (2010). A bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Computational Biology*, *6*(5), 1–11. https://doi.org/10.1371/journal.pcbi.1000770

Suzuki, T. A. (2017). Links between natural variation in the microbiome and host fitness in wild mammals. *Integrative and Comparative Biology*, *57*(4), 756–769. https://doi.org/10.1093/icb/icx104

Suzuki, T. A., Martins, F. M., & Nachman, M. W. (2018). Altitudinal variation of the gut microbiota in wild house mice. *Molecular Ecology*, 1–13. https://doi.org/10.1111/mec.14905 [Epub ahead of print]

Suzuki, T. A., & Nachman, M. W. (2016). Spatial heterogeneity of gut microbial composition along the gastrointestinal tract in natural populations of house mice. *PLoS ONE*, *11*(9), 1–15. https://doi.org/10.1371/journal.pone.0163720

Suzuki, T. A., & Worobey, M. (2014). Geographical variation of human gut microbial composition. *Biology Letters*, *10*(2), 20131037–20131037. https://doi.org/10.1098/rsbl.2013.1037

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., … Knight, R. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*, *551*(7681), 457–463. https://doi.org/10.1038/nature24621

Trinchieri, G. (1998). Interleukin-12: A cytokine at the interface of inflammation and immunity. *Advances in Immunology*, *70*(2), 83–243.

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., … Gordon, J. I. (2009). A core gut microbiome in obese and lean twins. *Nature*, *457*(7228), 480–484. https://doi.org/10.1038/nature07540

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., & Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, *444*(7122), 1027–1031. https://doi.org/10.1038/nature05414

Turpin, W., Espin-Garcia, O., Xu, W., Silverberg, M. S., Kevans, D., Smith, M. I., … Croitoru, K. (2016). Association of host genome with intestinal microbial composition in a large healthy cohort. *Nature Genetics*, *48*(11), 1413–1417. https://doi.org/10.1038/ng.3693

Vieira, F. G., Lassalle, F., Korneliussen, T. S., & Fumagalli, M. (2016). Improving the estimation of genetic distances from next-generation sequencing data. *Biological Journal of the Linnean Society*, *117*(1), 139–149. https://doi.org/10.1111/bij.12511

Wang, J., Kalyan, S., Steck, N., Turner, L. M., Harr, B., Künzel, S., … Baines, J. F. (2015). Analysis of intestinal microbiota in hybrid house mice reveals evolutionary divergence in a vertebrate hologenome. *Nature Communications*, *6*, 6440. https://doi.org/10.1038/ncomms7440

Wang, J., Linnenbrink, M., Künzel, S., Fernandes, R., Nadeau, M.-J., Rosenstiel, P., & Baines, J. F. (2014). Dietary history contributes to enterotype-like clustering and functional metagenomic content in the intestinal microbiome of wild mice. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(26), E2703–E2710. https://doi.org/10.1073/pnas.1402342111

Wang, J., Thingholm, L. B., Skiecevičienė, J., Rausch, P., Kummen, M., Hov, J. R., … Franke, A. (2016). Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nature Genetics*, *48*(11), 1396–1406. https://doi.org/10.1038/ng.3695

Weldon, L., Abolins, S., Lenzi, L., Bourne, C., Riley, E. M., & Viney, M. (2015). The gut microbiota of wild mice. *PLoS ONE*, *10*(8), 1–15. https://doi.org/10.1371/journal.pone.0134643

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Suzuki TA, Phifer-Rixey M, Mack KL, et al. Host genetic determinants of the gut microbiota of wild mice. *Mol Ecol*. 2019;28:3197–3207. https://doi.org/10.1111/mec.15139