Supplemental figures

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], and Michael W. Nachman[1]

[1] Department of Integrative Biology and Museum of Vertebrate Zoology, University of California Berkeley, Berkeley, CA, 94720, USA.

[2] Department of Biology, Monmouth University, West Long Branch, NJ, 07764, USA.

[3] Department of Neurobiology and Behavior, Cornell University, Ithaca, NY, 14853, USA.

[4] California Institute for Quantitative Biosciences, University of California Berkeley, Berkeley, CA, 94720, USA.

Table of contents

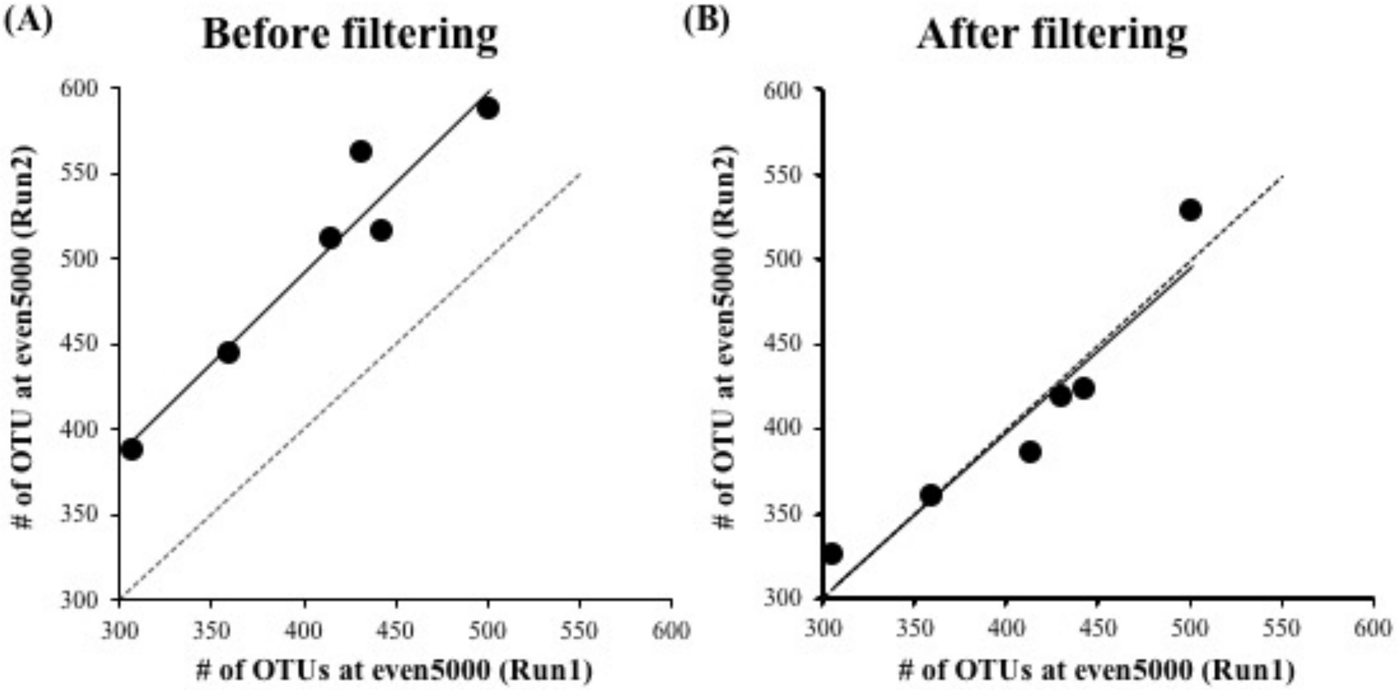


Figure S1. Correlations between OTU counts on the same six samples between two lanes of Illumina MiSeq before and after filtering. Dotted lines show the expected regression between the two samples ran on different lanes of Illumina sequencing. (A) Run2 shows greater number of OTU counts compared to run1 despite rarefying to an even depth of 5000 reads. (B) Rare OTUs with relative abundance less than 8.0×10^{-6} were removed from the run2 OTU table to account for the lane bias.

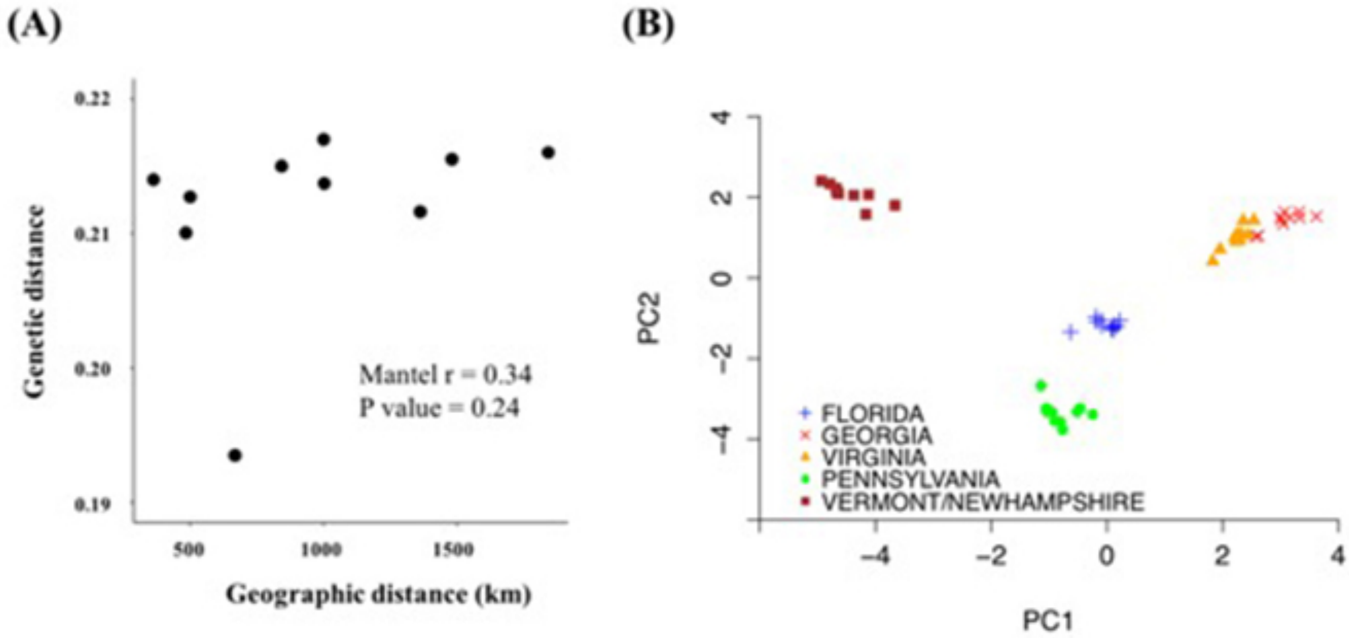


Figure S2. There is no evidence of isolation by distance. (A) Genetic distance (calculated by ngsDist) and geographic distance between wild mice populations are not significantly correlated (Mantel test, r = 0.34, p = 0.24). (B) Genetic PCA plot adopted from Phifer-Rixey et al. (2018).

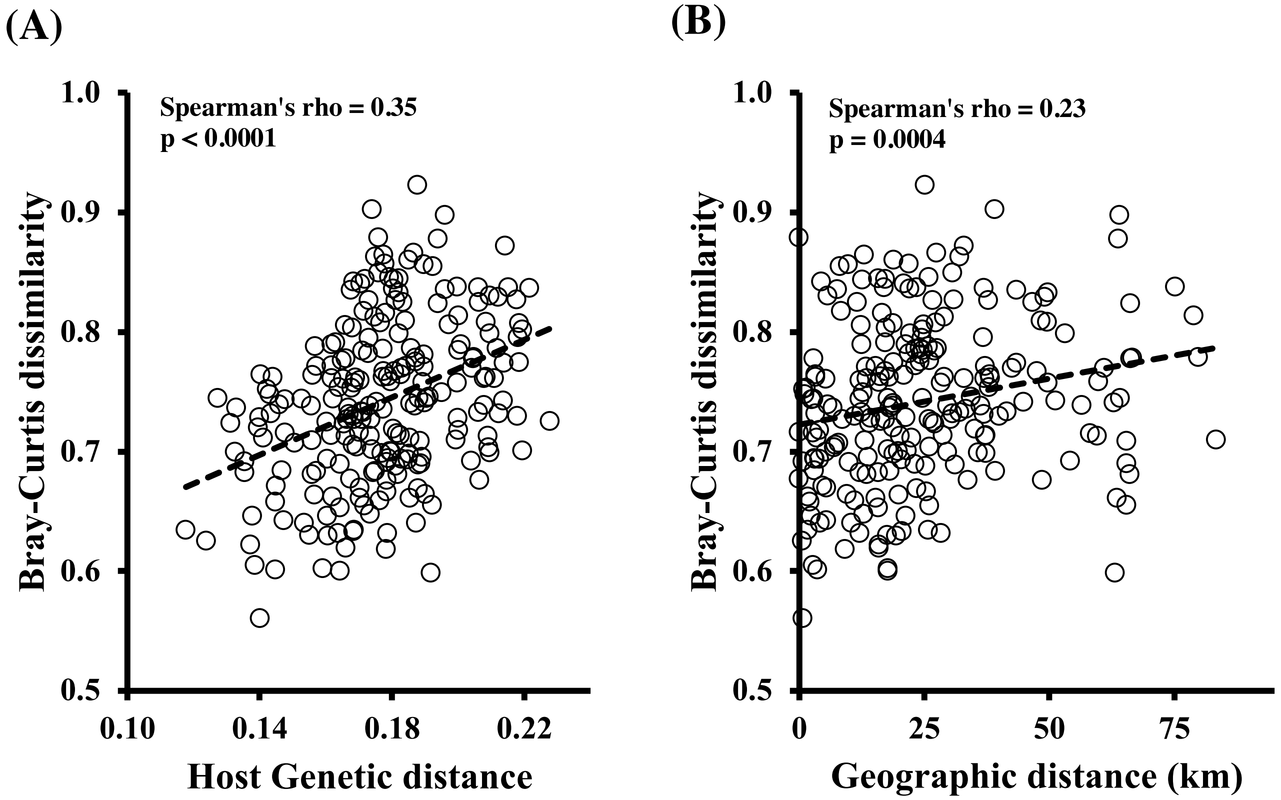


Figure S3. The effects of host genetic and geographic distances on Bray-Curtis dissimilarity within populations. (A) A positive correlation between Bray-Curtis dissimilarity and host genetic distance (calculated by ngsDist) within populations. (B) A positive correlation between Bray-Curtis dissimilarity and host geographic distance within populations. 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, the correlation between geographic distance and Bray-Curtis dissimilarity did not remain significant after controlling for genetic distance (Spearman's rho = 0.06, $p = 0.38$).

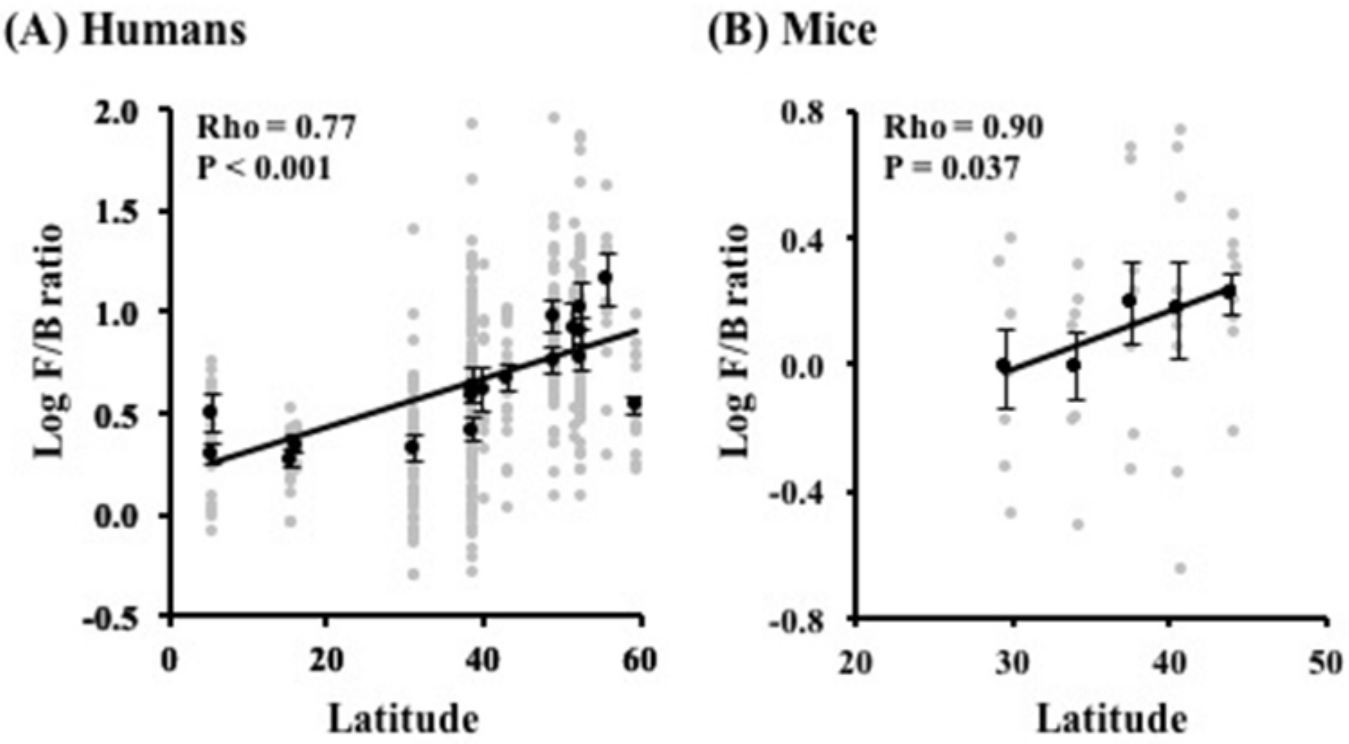
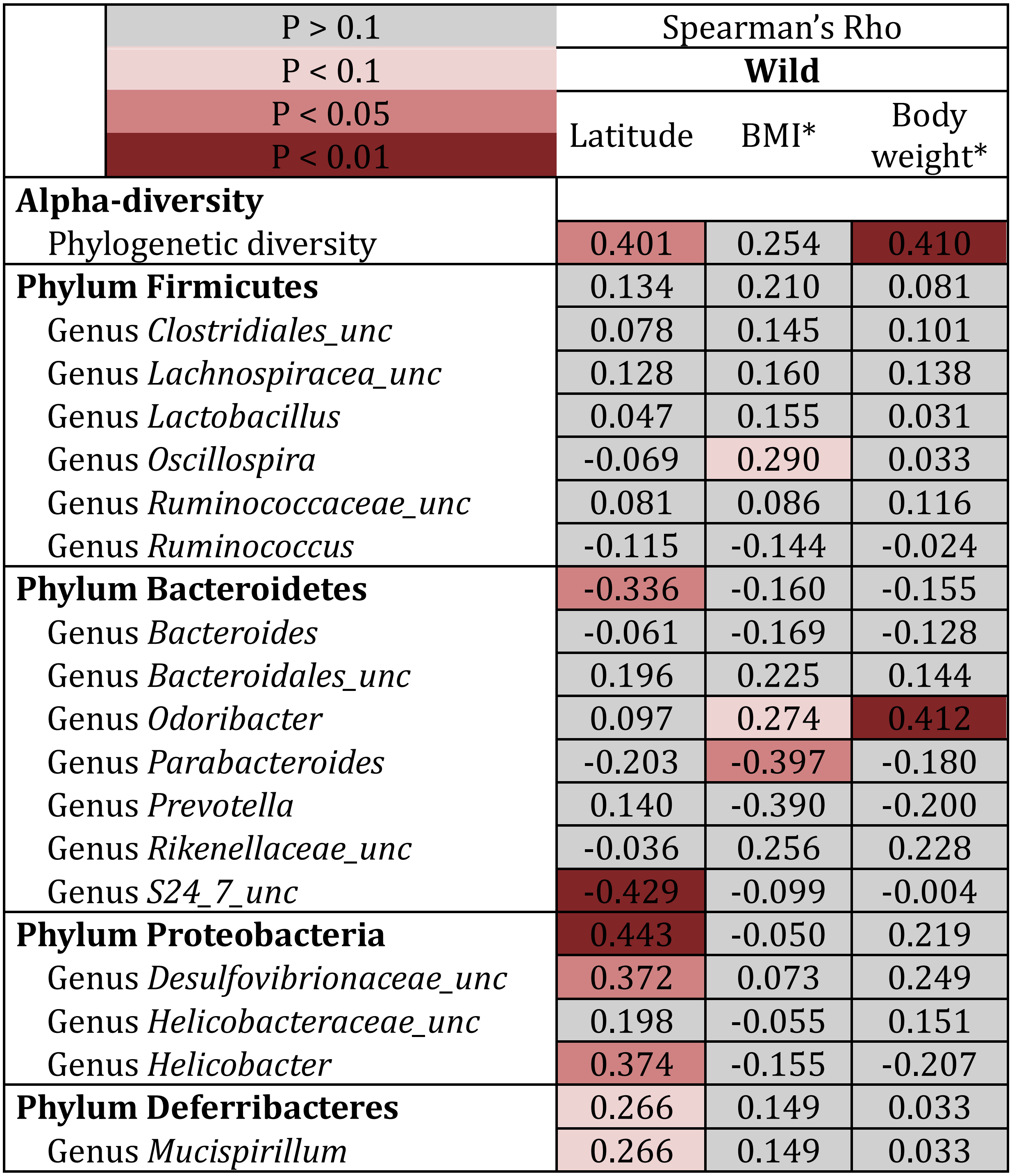


Figure S4. Correlations between obesity-associated composition (Log Firmicutes/Bacteroidetes ratio) and latitudes. Black dots are population average and gray dots are individual data points. (A) Figure modified from Suzuki and Worobey (2014). Significant positive correlation for population average (Rho = 0.77, P < 0.0001) and individual data points (Rho = 0.51, P < 0.0001) in healthy adult human populations. (B) Significant positive correlation for population average (Rho = 0.90, P = 0.037), but not for individual data points (Rho = 0.26, P = 0.11) in adult mouse populations. There are also weak trends of positive correlations between Log F/B ratio and Body weight (Rho = 0.29, p=0.0693) and BMI (Rho = 0.282, p = 0.0779) in wild mice. Error bars are SE.



$P > 0.1$	Spearman's Rho		
$P < 0.1$	**Wild**		
$P < 0.05$	Latitude	BMI*	Body weight*
$P < 0.01$			
Alpha-diversity			
Phylogenetic diversity	0.401	0.254	0.410
Phylum Firmicutes	0.134	0.210	0.081
Genus *Clostridiales_unc*	0.078	0.145	0.101
Genus *Lachnospiracea_unc*	0.128	0.160	0.138
Genus *Lactobacillus*	0.047	0.155	0.031
Genus *Oscillospira*	-0.069	0.290	0.033
Genus *Ruminococcaceae_unc*	0.081	0.086	0.116
Genus *Ruminococcus*	-0.115	-0.144	-0.024
Phylum Bacteroidetes	-0.336	-0.160	-0.155
Genus *Bacteroides*	-0.061	-0.169	-0.128
Genus *Bacteroidales_unc*	0.196	0.225	0.144
Genus *Odoribacter*	0.097	0.274	0.412
Genus *Parabacteroides*	-0.203	-0.397	-0.180
Genus *Prevotella*	0.140	-0.390	-0.200
Genus *Rikenellaceae_unc*	-0.036	0.256	0.228
Genus *S24_7_unc*	-0.429	-0.099	-0.004
Phylum Proteobacteria	0.443	-0.050	0.219
Genus *Desulfovibrionaceae_unc*	0.372	0.073	0.249
Genus *Helicobacteraceae_unc*	0.198	-0.055	0.151
Genus *Helicobacter*	0.374	-0.155	-0.207
Phylum Deferribacteres	0.266	0.149	0.033
Genus *Mucispirillum*	0.266	0.149	0.033

Figure S5. Correlations between microbial measurements (alpha-diversity and relative abundance of phyla and genera) and metadata (i.e. latitude, BMI, and body weight). * BMI and body weight are controlled for latitude by using the residuals of covariate regression. Spearman's rho values are shown and the p-values are indicated by color. The genera are chosen based on average relative abundance of >1% in the wild samples. Pregnant females and juveniles are excluded and only adults (n=40) are included in the analyses.

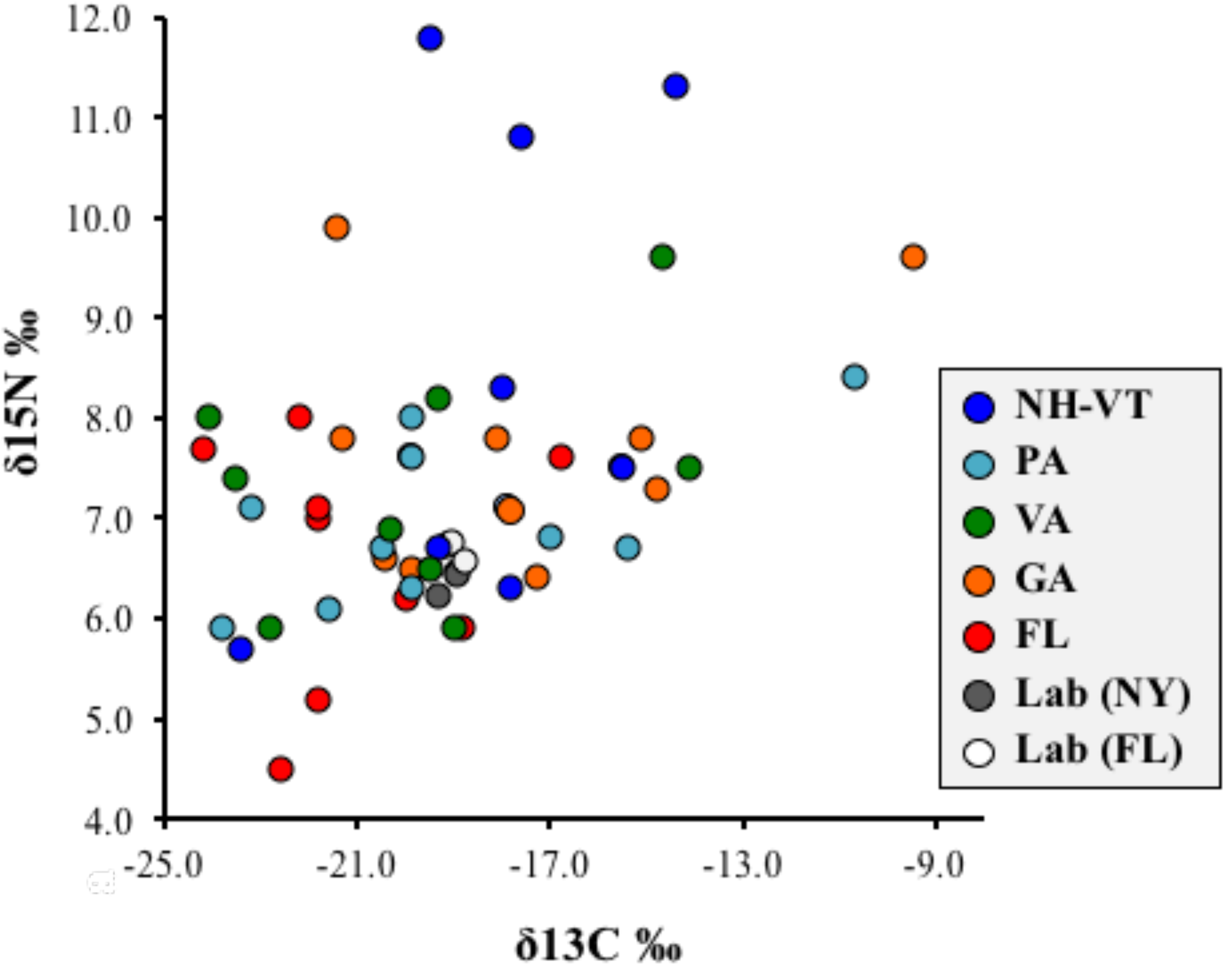


Figure S6. Dietary difference among wild and lab populations. No significant differences were observed among the populations (Kruskal-Wallis tests: δ13C ‰, p-value = 0.27 and δ15N ‰, p-value = 0.24). The lab diet was significantly different from that of the wild in terms of $\delta^{15}N$ (lab mean: 6.5, wild mean: 7.1, $p < 0.05$), but not in terms of $\delta^{13}C$ (lab mean: -19.0, wild mean: 19.1, p = 0.12)

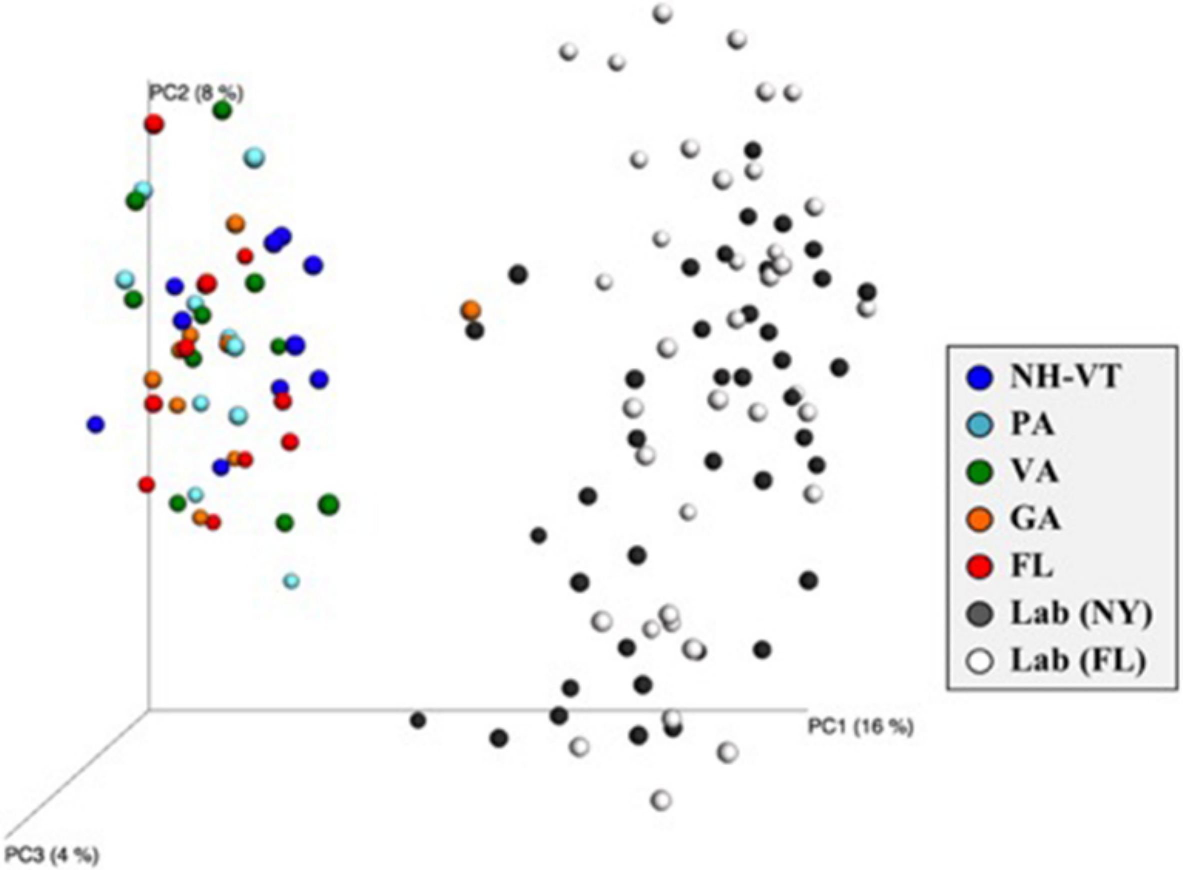


Figure S7. PCoA plot between wild and lab individuals using Bray-Curtis dissimilarity.

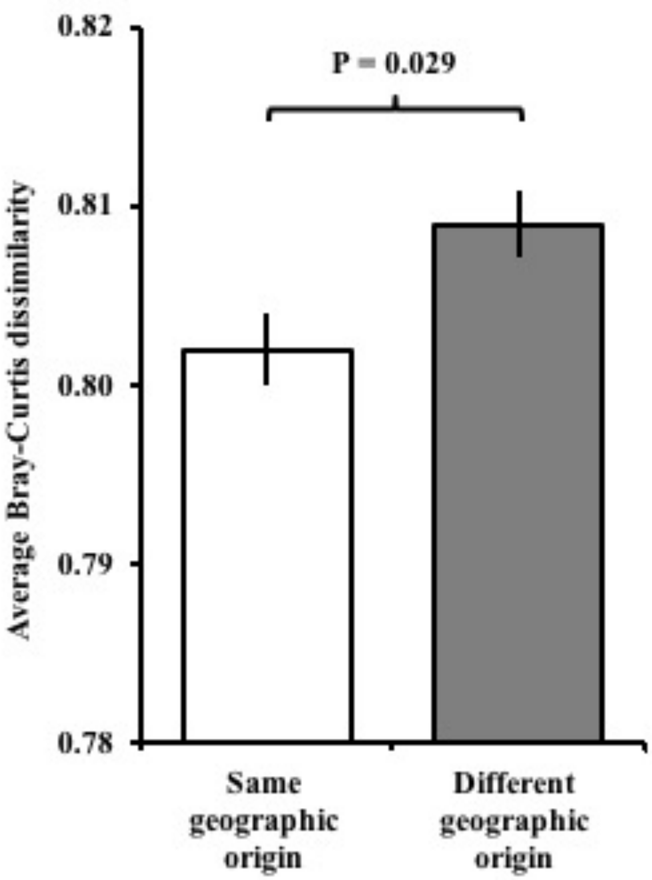


Figure S8. The microbial community composition is more similar between wild and lab reared animals that share the same geographic origin compared to those that do not share the geographic origin ($p = 0.029$). P-values are Wilcoxon permutation test based on 9999 Monte-Carlo resampling. Error bars are SE.