Supplementary Figure 1. The sampled single species populations of threespine stickleback (a & b) and Eurasian perch (c & d) exhibit unimodal variation in both morphology and diet. Unimodal phenotype distributions within each fish population (a & c), demonstrating that these are single populations\(^5-8\). See Supplementary Table 5 for loadings of morphological traits on the principal component axes. Diet variation for stickleback (b) and perch (d) are presented as two plots per species, showing raw $\delta^{13}$C and $\delta^{15}$N values per fish and converted into $\alpha$ and $\text{tpos}$. Among-individual variation in carbon and nitrogen stable isotope ratios was an order of magnitude greater than the variance observed when individuals are fed a shared diet\(^9\). The range of $\delta^{13}$C variation spans nearly the full range of $\delta^{13}$C differences between basal pelagic and littoral consumers (filter-feeding mussels and epibenthic grazing snails). Following Post et al\(^{10}\) we used mussel and snail $\delta^{13}$C (means indicated by arrows on the horizontal axis) and $\delta^{15}$N to convert fish
isotope ratios into measures of proportion benthic carbon and trophic position. A 95% density ellipse indicates the major axis of covariation between $\alpha$ and $tpos$, captured by the first principal component axis of diet scores (70.1% of variation in stickleback, 79.8% in perch). This major axis is highly parallel to isotopic measures of dietary differentiation between benthic and limnetic species pairs\(^{11}\) (benthic / limnetic is often used instead of littoral / pelagic) and between parapatric littoral versus pelagic stickleback populations in adjoining lake basins\(^{12}\). We plot the centroid values of $\alpha$ and $tpos$ for three pairs of benthic (green) and limnetic populations (blue), including Priest Lake species pairs (Pri-B and Pri-L), Paxton Lake species pairs (Pax-B and Pax-L) and a parapatric benthic and limnetic population pair (Dug = Dugout Lake, Orm = Ormond Lake). There is appreciable diet divergence between male and female stickleback (ANOVA, $\alpha$: $F_{1,163}=60.0$, $p<0.0001$; $tpos$ $F_{1,163}=56.7$, $p<0.0001$), In perch, sex has no effect on $\alpha$ or $tpos$ ($p = 0.759$ and 0.878 respectively).
Supplementary Figure 2. Relative abundance (averaged across hosts) of microbial Classes and Orders across male and female (a) stickleback and (b) perch exhibit very weak between-sex differences in gut microbiota. In stickleback, PerMANOVAs of Jaccard distance
matrices calculated from relative abundance data found significant between-sex differences in microbiota composition at the level of Class (p=0.016), Order (p=0.038), Family (p=0.038) and Genus (p=0.048) but not Phylum (p=0.088). For each taxonomic rank, sex explained only ~1% of the variation in microbial taxon relative abundance. We therefore conclude that there are differences in microbial community composition between male and female stickleback, but these are negligible compared to the heterogeneity among individuals. Note that this conclusion does not preclude the existence of stronger sex differences in the form of sex*diet interactions. In perch, main effects of sex are not supported by perMANOVAs (p>0.25 for tests at all taxonomic ranks).
Supplementary Figure 3. Canonical correlations between microbial community structure and stickleback traits are sex specific. CCA correlations are shown for stickleback males (top two panels) and females (bottom panels). The scatterplots in the left column indicate the correlation between host traits and microbiota structure, with the horizontal axis representing a multivariate weighted combination of $\alpha$, $tpos$, and morphological PC1-4 (see Supplementary Table 6), and the vertical axis a combination of microbial PCoA traits 1-27 (top 50% of variation). The loadings of host traits and microbial PCoA axes on the top two CCA axes are plotted in the right column for males (top) and females (bottom). For example, in males the
major axis (dimension 1) of trait variation is largely distinguished by variation between individuals with high values of PC4 and \( tpos \) (negative axis 1 scores) to individuals with high \( \alpha \) (positive axis scores). Similarly, the first canonical Y axis of the microbiota distinguishes individuals with large values of uwPCoA7&10, from individuals with large values of unweighted principal coordinate axes (uwPCoA4, 15 and 20). The primary female trait axis is different from the male trait axis, separating individuals with high \( tpos \) (but not PC4) versus high \( \alpha \). In females, the microbial canonical axes are also substantially different than in males: the first axis separates females with high scores on uwPCoA4 & 14 & 27, from females with high uwPCoA 8 & 9 & 22 & 26. See Supplementary Information Table 2 for details of significance, correlations, and loadings.
Supplementary Figure 4. The canonical correlations between host traits and microbiota composition entail different loadings of host traits (top panel) and microbial axes (bottom panel) in males versus females. Loadings indicate the relative weight of each host trait (two diet...
measures and morphological principal component axes 1 to 4) or microbial unweighted PCoA axis (axes 1-21, top 50% of variation) in driving trait-microbe correlations. If males and females exhibit similar trait-microbe associations, their canonical axis loadings should be similar (positively correlated). As indicated by correlation coefficients and p values at the top of each panel, these correlations are not significantly different from zero, both for male versus female traits that drive microbial variation, or for male versus female microbial axes that respond to diet. Unlike stickleback, in perch loadings of host traits on the predictor (X) canonical axes are similar in males versus females (r=0.4559 p=0.022), and males and females exhibit modestly similar microbial loadings on the leading CCA Y axis (r=0.1329 p=0.037). This significant correlation between male and female CCA loadings makes the absence of such correlations in stickleback all the more striking.
Supplementary Figure 5. There are significant effects of diet on inferred gut microbiome functional composition in stickleback. (a) Individual variation in gene ontology category frequencies. As in Fig. 1, individuals are represented as stacked columns whose color distribution represents the relative abundance of gene ontology (GO) categories inferred from 16S sequences and published microbial genome databases, using PICRUSt. Individuals are sorted along the horizontal axis based on the first principal component axis of gut microbiome gene composition. GO categories are ordered from most abundant (bottom, membrane transport) to least abundant (top), combining the rarest GO categories as “other” for clarity. Although the taxonomic composition of the gut microbiota varies among individuals, GO composition is comparatively similar among individuals, at least to the resolution available with PICRUSt. This suggests substantial functional redundancy among many taxonomically disparate microbiota within each population. The equivalent figure for perch looks very similar, so is not included. (b) Inferred gut microbiome functional composition in lab stickleback fed littoral (left of solid line) versus pelagic (right) diets. Individuals here are sorted along the horizontal first by diet treatment, then within diet sorted by the first principal component axis of GO category variation.
Note that the rank order of GO categories is slightly different than in panel (a). There was no correlation between diet effects on GO composition in the lab males versus diet effect on GO composition in wild males (using $\alpha$ effects, $r=0.083$ $p=0.618$; using $tpos$, $r=-0.012$, $p=0.939$). (c) Diet effects on inferred GO counts in wild-caught fish, separated by host sex and species. The top heatmap indicates effects of trophic position, bottom heatmap indicates effects of the proportion littoral carbon. Colors indicate the effect size (slope / standard error) from quasibinomial GLMs in which the relative abundance of each GO category was regressed on $\alpha$ and $tpos$ within each sex and species. Green (red) squares indicate a GO category that is more abundant with higher (lower) trophic position or more littoral (pelagic) carbon. White stars indicate significant GLM effects at ($p<0.05$). White inequality symbols indicate which sex exhibits a more positive diet effect (significant sex*diet interaction in a species-wide GLM). GO categories are ordered on the horizontal axis in order of overall effect size (Fisher’s combined probability) from least significant on the left to most significant on the right. We list GO subcategories, beginning with an abbreviation indicating higher categories: CP = Cellular Process, EI = Environmental Information Processing, GI = Genetic Information Processing, HD = Human Diseases, M = Metabolism, OS = Organismal Systems. Wild stickleback with more littoral diets (higher $\alpha$) had significantly more genes involved in sensory systems ($p=0.0026$), signaling interactions ($p=0.032$), and significantly fewer genes involved in RNA processing ($p=0.028$). Stickleback with higher trophic position had more genes involved in enzyme families ($p=0.0393$), immune system disease ($p=0.0475$), membrane transport ($p=0.0078$), cell motility ($p=0.0421$), and signal transduction ($p=0.0266$), and fewer genes involved in sensory systems ($p=0.0016$), metabolic diseases ($p=0.0398$), cofactor/vitamin metabolism ($p=0.0029$), nervous system ($p=0.0010$), energy metabolism ($p=0.0052$), and RNA processing ($p=0.0185$). Perch with
more littoral diets exhibited more genes involved in RNA processing \((p=0.0004)\), membrane transport \((p=0.0090)\), transcription \((p=0.0119)\), and fewer genes involved in terpenoid/polyketide metabolism \((p=0.0136)\), endocrine system \((p=0.0332)\), transport and catabolism \((p=0.0366)\). Perch with higher trophic position had more genes involved in RNA processing \((p<0.0001)\), but \(tpos\) had no significant effect on any other GO category \((all \ p>0.1)\). Note that although some of these are not strongly significant and might be discarded with FDR corrections \((e.g., \ q\text{-values} > 0.05)\), too many are significant to be plausibly explained as type I error. Significant sex*\(\alpha\) interactions were observed in stickleback for sensory system genes \((p=0.0088)\). Sex*\(tpos\) interactions were observed for genes involved in nervous systems \((p=0.024)\), energy metabolism \((p=0.024)\), membrane transport \((p=0.034)\), and cofactor/vitamin metabolism \((p=0.042)\). For littoral carbon, there was no correlation between male and female diet effects on GO category relative abundances \((r=-0.098, p=0.556)\), but we did observe a significant positive correlation between male and female GO category responses to trophic position \((r=0.4022, p=0.0123)\). For both diet metrics, females exhibited systematically stronger diet effects on GO category composition than was seen in males \((\alpha \text{-}t=-6.22 \ p<0.0001; \ tpos \text{-}t=-6.09 \ p<0.0001)\). In perch, significant sex*\(\alpha\) interactions were observed for carbohydrate metabolism \((p=0.0199)\), metabolic diseases \((p=0.0287)\), enzyme families \((p=0.0297)\), signal transduction \((p=0.0372)\), cell growth/death \((p=0.0455)\), infections disease \((p=0.0485)\). One sex*\(tpos\) interaction (nervous system; \(p=0.0243\)) did not survive FDR correction and no others were detected. Perch male and female microbiome responses to diet were uncorrelated \((r=-0.11 \text{ for } \alpha; \ r=0.011 \text{ for } tpos, \text{ both with } p>0.4)\).
Supplementary Figure 6. The presence or absence of some microbial taxa depends on diet treatment in lab-reared male stickleback. A discriminant function analysis of unweighted PCoA axes identified a single major axis of microbial variation separating *Daphnia*- and chironomid-fed male stickleback. The presence/absence of microbial taxa was regressed against this major axis, using logistic regression (best fit line plotted for each taxon) to identify clades underlying the diet effect. Dots in each panel represent individual fish, and the significance of the logistic regression is indicated above each panel. We use logistic regression of presence/absence here to demonstrate that diet effects entail variation in microbial presence/absence, as well as variation in relative abundance (the focus of most other OTU-based analyses in this paper). We also focus here on microbial Orders, simply to illustrate the point that diet effects are observed not just at the OTU level (the focus of most other analyses in this paper), but also at higher taxonomic ranks.
Supplementary Figure 7. Comparison of lab versus wild stickleback gut microbiota reveals weak similarity in microbial composition and no similarity in diet effects on microbiota. (a) Correlations between the log_{10} relative abundance of microbial taxa in the wild versus lab-reared
stickleback, focusing on taxa present in both samples. Correlation coefficients and corresponding p-values are provided for Phyla, Classes, Orders, Families, Genera, and OTUs. Overall, few natural OTUs were found in the lab samples: of the 3867 OTUs identified in the lab-reared males, 1899 OTUs were also observed in our sample of wild stickleback (9343 OTUs), and only 571 were observed in at least two individuals in both the lab and wild samples. Thus, only 14.7% of microbial OTUs from the lab are also repeatedly found in the wild. Conversely only 6.1% of wild OTUs are recovered in the lab. (b) There is no correlation between the diet effects on OTU abundance in wild versus lab-reared stickleback. Of the shared microbial OTUs, we found significant microbe-diet associations in both lab and field samples. There were 115 shared OTUs that exhibited significant positive (74 OTUs) or negative (41 OTUs) correlations with littoral diet (\(\alpha\)) in wild male stickleback, and 12 OTUs that exhibited significant effects of diet in lab fish (10 more common in littoral diets). However, only two OTUs exhibited significant associations with diet in both lab and wild samples, one of which exhibited opposite responses in lab versus field samples. Diet effects are measured as t statistics in which the slope of relative abundance with diet is divided by the standard error of the slope, from GLMs of each microbial OTU on either % littoral carbon (wild) or diet treatment (1 = chironomid larvae [littoral prey], 0 = daphnia [pelagic prey]). Filled symbols indicate statistically significant effects, open symbols are non-significant. The color of filled symbols indicates in which sample the diet effect is significant (red for wild, green for lab, and blue for both). Similar results (no correlation between lab and diet effects) were observed when analyzing higher taxonomic ranks instead (e.g., Family \(\rho=0.105, p=0.234\))
Supplementary Figure 8. An arbitrarily selected example of a microbial OTU (unknown genus and species of Pirellulaceae) exhibits a significant sex*diet interaction in lab-reared male and female stickleback fed two different diets (chironomids or Daphnia). This OTU is more common in chironomid-fed than in Daphnia-fed females (p = 0.000053) but unaffected by male diet (p=0.294; the trend is opposite to that in females) resulting in a significant sex*diet interaction (p=0.000042; with FDR correction q=0.0102). Females are indicated by red dots and trendline, males by blue dots and line. Lines indicate estimated effects from quasibinomial.
GLMs, with 95% confidence intervals indicated by shading. A small value is added to create non-zero relative abundances in fish where the OTU was never observed (cluster of small points at the bottom of the graph), to allow us to plot log-transformed abundances. Slight horizontal jitter is added to distinguish overlapping points.
Supplementary Figure 9. Experimental diet treatments (chironomid versus *Daphnia*) effects on OTU abundance in females (horizontal axis) are only weakly correlated with diet effects on OTU abundances in males (vertical axis). Each point indicates the diet effect size in females and males for a given OTU. Effect sizes are measured by quasibinomial GLM slope estimates divided by their standard errors to yield t-statistics, which are plotted on the x axis for females and y axis for males. Positive values indicate OTUs that are more common in *Daphnia*-fed fish, negative values indicate OTUs more common in chironomid-fed fish. Each point is a
separate OTU. Open circles are OTUs with no significant diet effect in either sex. Red or blue points indicate significant diet effects in females or males respectively, green points indicate significant diet effects in both sexes. Black points indicate OTUs with significant diet effect in an overall model with both sexes combined, but not significant in either sex alone. Triangles indicate OTUs with significant sex*diet interactions. The diagonal dotted line indicates the anticipated line of equality if diet effects are identical in males versus females. Unlike in wild fish, where diet effects were uncorrelated, in lab stickleback we found a weak positive correlation between diet effects in male versus female stickleback in the lab (r=0.171, p=0.0077). Thirty-two OTUs (out of 241 common OTUs) were significantly affected by female diet. All but one of these OTUs were more common in chironomid-fed females. Eight OTUs were significantly affected by male diet (all more common in chironomid-fed males). Of these, only three OTUs exhibited significant effects in both sexes. The absolute magnitude of diet effects in females (measured by the t-statistic effect size estimate from the quasibinomial GLM) is consistently stronger than the magnitude of diet effects in males (paired t-test, t = -2.74, df = 241, p=0.0065). This result corroborates the observation from wild-caught fish that diet effects were stronger in females than in males.
Supplementary Figure 10. Selected examples exhibit OTUs with sex-dependent diet effects on the human microbiota. Such interactions occur when diet (here, individuals’ saturated fatty acid intake) has a different effect on OTU relative abundance in females (black lines and points) than in males (red lines and points). For each panel we plot the trend estimated with a sex-specific quasibinomial GLM, and provide measures of statistical support for diet PC1 effects on the OTU in each sex, and for a sex*diet interaction. We focus here on saturated fatty acid intake to make the point that sex*diet interaction effects can be observed for individual dietary measures as well as the dietary principal component axes used here for other analyses of human diet-microbe associations.
Supplementary Figure 11. Diet effects on the gut microbiota of male and female mice show positively correlated diet effects between sexes, albeit with some sex*diet interactions. (a) Binomial mixed-model GLMs revealed significant effects of sex, diet treatment (regular chow versus high-fat diet), and sex*diet interactions on the relative abundance of microbial genera in the guts of various strains of laboratory mice. To present selected examples, each panel presents the relative abundance (number of sequence reads out of total read depth per individual) of a given microbial genus, as a function of host diet and sex. Quasibinomial GLM fits with confidence intervals are plotted for each sex (red points + curves for females, blue points and
curves for males). Individuals were either fed a regular chow diet or a high fat diet (left and right of the solid vertical line, respectively. Four microbial genera showed significant sex by diet interactions that survived tests for multiple comparisons: A) Lactobacillus, B) Allistipes, C) Lachnospiraceae, and D) an unknown Clostridiales genus. Individuals where the microbe was not observed are shown at a log relative abundance of -7. Some horizontal and vertical scatter are added to each point to separate overlapping points. (b) Diet effects are significantly correlated (p<0.0001) between male and female mice. Each point represents the female and male diet effect on one of the common genera (found in at least 5 hosts). Diet effect is measured as a t-statistic, the quasibinomial GLM slope of genus relative abundance as a function of diet, standardized by the standard error of that slope. Open points are not statistically significant (p>0.05). Red points are significant in females, blue significant in males, and green points are significant in both. Horizontal and vertical lines visually distinguish OTUs with positive effect estimates in both sexes (top right quadrant), or negative effects in both sexes (bottom left), or opposite effects in males versus females (top left or bottom right).
Supplementary Note 1

QIIME bioinformatic steps to identify microbial sequences

Analysis protocol:
Analysis performed with QIIME 1.5.0-dev. `split_libraries_fastq.py` was performed with default parameters. OTUs were picked using a closed-reference OTU picking protocol against the Greengenes 12_10 release. Weighted and unweighted UniFrac, PD, and observed species were computed for all samples.

Commands:
```
echo "split_libraries_fastq.py -i /Users/caporaso/analysis/bolnick_fish/11oct2012/Undetermined_S0_L001_R1_001.fastq.gz -b /Users/caporaso/analysis/bolnick_fish/11oct2012/Undetermined_S0_L001_I1_001.fastq.gz -o /Users/caporaso/analysis/bolnick_fish/11oct2012/slout_R1/ -m /Users/caporaso/analysis/bolnick_fish/11oct2012/meta_analysis_tmpudM30Tdfx7K1vmY0pRO1_map.txt --rev_comp_mapping_barcode" | qsub -keo -N bolnick_s11
```
```
echo "split_libraries_fastq.py -i /Users/shared/Illumina_hiseq_UCB/Illumina113011/fastq/s_5_1_sequences.fastq -b /Users/shared/Illumina_hiseq_UCB/Illumina113011/fastq/s_5_2_sequences.fastq -o /Users/caporaso/analysis/bolnick_fish/24oct2012/slout_perch5/ -m /Users/caporaso/analysis/bolnick_fish/perch_map.txt --rev_comp_mapping_barcode" | qsub -keo -N slper5
```
```
```
```
echo "split_libraries_fastq.py -i /Users/shared/Illumina_hiseq_UCB/Illumina_washu_051411/fastq_files/s_5_1_withindex_sequence.fastq -b /Users/shared/Illumina_hiseq_UCB/Illumina_washu_051411/fastq_files/s_5_1_withindex_sequence_barcode.fastq -o /Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ -m
```
```
Closed-reference OTU picking against Greengenes 12_10

```
echo "pick_reference_otus_through_otu_table.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_perch5/seqs.fna
-o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_perch5/ucrC_fast
/-p /Users/caporaso/analysis/bolnick_fish/ucrC_fast_params.txt -r
/Users/caporaso/data/gg_12_10_otus/rep_set/97_otus.fasta -t
/Users/caporaso/data/gg_12_10_otus/taxonomy/97_otu_taxonomy.txt -a0
50" | qsub -keo -N per5ucrC -l pvmem=8gb -q memroute
```

[lane 5 perch data was same samples as lane six, but much lower coverage so working only with lane 6 data]

```
echo "pick_reference_otus_through_otu_table.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_perch6/seqs.fna
-o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_perch6/ucrC_fast
/-p /Users/caporaso/analysis/bolnick_fish/ucrC_fast_params.txt -r
/Users/caporaso/data/gg_12_10_otus/rep_set/97_otus.fasta -t
/Users/caporaso/data/gg_12_10_otus/taxonomy/97_otu_taxonomy.txt -a0
50" | qsub -keo -N per5ucrC -l pvmem=8gb -q memroute
```

```
echo "pick_reference_otus_through_otu_table.py -i
/Users/caporaso/analysis/bolnick_fish/11oct2012/slout_R1/seqs.fna -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_lab/ucrC_fast/
-p /Users/caporaso/analysis/bolnick_fish/ucrC_fast_params.txt -r
/Users/caporaso/data/gg_12_10_otus/rep_set/97_otus.fasta -t
/Users/caporaso/data/gg_12_10_otus/taxonomy/97_otu_taxonomy.txt -a0
50" | qsub -keo -N labucrC -l pvmem=8gb -q memroute
```

```
echo "pick_reference_otus_through_otu_table.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/seqs.fna
-o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC_fast/
-p /Users/caporaso/analysis/bolnick_fish/ucrC_fast_params.txt -r
/Users/caporaso/data/gg_12_10_otus/rep_set/97_otus.fasta -t
/Users/caporaso/data/gg_12_10_otus/taxonomy/97_otu_taxonomy.txt -a0
50" | qsub -keo -N stiucrC -l pvmem=8gb -q memroute
```

Beta diversity (at 3000 and 10000 seqs/sample)

```
echo "beta_diversity_through_plots.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_perch6/ucrC_fast
/uclust_ref_picked_otus/otu_table.biom -o
```

23
/Users/caporaso/data/gg_12_10_otus/trees/97_otus.tree -e 10000 -m
/Users/caporaso/analysis/bolnick_fish/stickleback_map.txt -aO 50" | qsub -keo -N st10000 -l pvmem=4gb -q memroute

Convert biom OTU tables to txt
convert_biom.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_perch6/ucrC_fast/uclust_ref_picked_otus/otu_table.biom -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_perch6/ucrC_fast/uclust_ref_picked_otus/otu_table.txt --header_key taxonomy -b

convert_biom.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_lab/ucrC_fast/uclust_ref_picked_otus/otu_table.biom -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_lab/ucrC_fast/uclust_ref_picked_otus/otu_table.txt --header_key taxonomy -b

convert_biom.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC_fast/uclust_ref_picked_otus/otu_table.biom -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC_fast/uclust_ref_picked_otus/otu_table.txt --header_key taxonomy -b

Create master OTU table
echo "merge_otu_tables.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/combined_otu_table.biom" | qsub -keo -N mergefish -l pvmem=16gb -q memroute

per_library_stats.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/combined_otu_table.biom >
/Users/caporaso/analysis/bolnick_fish/24oct2012/combined_otu_table_per_lib_stats.txt

System configuration details (note that although qiime_config points to an old version of Greengenes, we're not using that in the analysis as 12_10 has been manually specified)
caporaso@compy ~> print_qiime_config.py

System information
Platform: linux2
Python version: 2.7.3 (default, Oct 3 2012, 11:10:11) [GCC 4.4.6 20120305 (Red Hat 4.4.6-4)]
Python executable: /Users/caporaso/.local/bin/python

Dependency versions

PyCogent version: 1.5.3
NumPy version: 1.5.1
matplotlib version: 1.1.0
biom-format version: 1.0.0c
QIIME library version: 1.5.0-dev, svn revision 3296
QIIME script version: 1.5.0-dev

PyNAST version (if installed): 1.1
RDP Classifier version (if installed): rdp_classifier-2.2.jar

QIIME config values

blastmat_dir: /Users/caporaso/blast-2.2.16/data
sc_queue: all.q
topiaryexplorer_project_dir: None
pynast_template_alignment_fp:
/Users/caporaso/data/gg_core/core_set_aligned.fasta.imputed
cluster_jobs_fp:
/Users/caporaso/bin/cluster_jobs_4.py
pynast_template_alignment_blastdb: None
assign_taxonomy_reference_seqs_fp:
/scratch/caporaso/gg_otus_4feb2011/rep_set/gg_97_otus_4feb2011.fasta

torque_queue: friendlyq
qiime_test_data_dir:
/scratch/caporaso/qiime_test_data
template_alignment_lanemask_fp:
/Users/caporaso/data/gg_core/lanemask_in_1s_and_0s.txt
jobs_to_start: 3
cloud_environment: False
qiime_scripts_dir:
/Users/caporaso/code/Qiime/scripts/
denoiser_min_per_core: 50
working_dir: None
python_exe_fp: python
temp_dir: /scratch/caporaso/temp/
Several non-gut microbiome samples should be excluded from the Stickleback analyses. Filtering those in the following steps:

Modified stickleback mapping file to exclude 6 problematic samples:
Stickleback1.143.filt
Stickleback2.273.liv
Stickleback2.301.liv
Stickleback2.327.liv
Stickleback2.330.filt
Stickleback1.20.liv
d7c7a80133f95ed83f65dcf443a0e3c7 stickleback_map.txt

Renamed original stickleback mapping file to stickleback_map_old.txt

Filter those six samples from the OTU table and distance matrix and re-generate principal coordinate matrices
pwd
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC_fast/uclust_ref_picked_otus

mv otu_table.biom otu_table.old.biom
mv otu_table.txt otu_table.old.txt
mv otu_table_per_lib_stats.txt otu_table_per_lib_stats.old.txt

filter_samples_from_otu_table.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC_fast/uclust_ref_picked_otus/otu_table.old.biom -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC_fast/uclust_ref_picked_otus/otu_table.biom --sample_id_fp
/Users/caporaso/analysis/bolnick_fish/stickleback_map.txt

# bdiv_even3000
mv unweighted_unifrac_dm.txt unweighted_unifrac_dm.old.txt
mv weighted_unifrac_dm.txt weighted_unifrac_dm.old.txt
mv unweighted_unifrac_pc.txt unweighted_unifrac_pc.old.txt
mv weighted_unifrac_pc.txt weighted_unifrac_pc.old.txt

filter_distance_matrix.py -i unweighted_unifrac_dm.old.txt -o unweighted_unifrac_dm.txt -t ../uclust_ref_picked_otus/otu_table.biom
filter_distance_matrix.py -i weighted_unifrac_dm.old.txt -o weighted_unifrac_dm.txt -t ../uclust_ref_picked_otus/otu_table.biom

principal_coordinates.py -i weighted_unifrac_dm.txt -o weighted_unifrac_pc.txt
principal_coordinates.py -i unweighted_unifrac_dm.txt -o unweighted_unifrac_pc.txt

# bdiv_even10000
mv unweighted_unifrac_dm.txt unweighted_unifrac_dm.old.txt
mv weighted_unifrac_dm.txt weighted_unifrac_dm.old.txt
mv unweighted_unifrac_pc.txt unweighted_unifrac_pc.old.txt
mv weighted_unifrac_pc.txt weighted_unifrac_pc.old.txt

filter_distance_matrix.py -i unweighted_unifrac_dm.old.txt -o unweighted_unifrac_dm.txt -t ../uclust_ref_picked_otus/otu_table.biom
filter_distance_matrix.py -i weighted_unifrac_dm.old.txt -o weighted_unifrac_dm.txt -t ../uclust_ref_picked_otus/otu_table.biom

principal_coordinates.py -i weighted_unifrac_dm.txt -o weighted_unifrac_pc.txt
principal_coordinates.py -i unweighted_unifrac_dm.txt -o unweighted_unifrac_pc.txt

Alpha rarefaction (at max 10000 seqs/sample - re-running to exclude samples with fewer than 10k sequences)

echo "filter_samples_from_otu_table.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_perch6/ucrC_fast/uclust_ref_picked_otus/otu_table.biom -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_perch6/ucrC_fast/uclust_ref_picked_otus/otu_table_mc10000.biom -n 10000 ;
a_alpha_rarefaction.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_perch6/ucrC_fast/uclust_ref_picked_otus/otu_table_mc10000.biom -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_perch6/ucrC_fast/arare_max10000/ -t
/Users/caporaso/data/gg_12_10_otus/trees/97_otus.tree -e 10000 -m
echo "filter_samples_from_otu_table.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC_fast/uclust_ref_picked_otus/otu_table_mc10000.biom -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC_fast/arare_max10000/ -t /Users/caporaso/data/gg_12_10_otus/trees/97_otus.tree -e 10000 -m /Users/caporaso/analysis/bolnick_fish/lab_map.txt -aO 50" | qsub -keo -N stastickleback0000 -l pvmem=4gb -q memroute

PICRUSt

Stickleback
echo "pick_reference_otus_through_otu_table.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/seqs.fna -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC_fast/picrust/ -p
/Users/caporaso/analysis/bolnick_fish/ucrC_fast_params.txt -r
/Users/caporaso/data/img_gg.otus_18may2012/rep_set/97_otus.img_gg_18may2012.corrected.fasta -aO 50" | qsub -keo -N stipicrust -l pvmem=8gb -q memroute

normalize_by_copy_number.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC

/Users/caporaso/analysis/bolnick_fish/perch_map.txt -aO 50" | qsub -keo -N perch10000 -l pvmem=4gb -q memroute
```
_fast_picrust/uclust_ref_picked_otus/otu_table.biom -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC
_fast_picrust/uclust_ref_picked_otus/otu_table.normed.biom

echo "predict_metagenomes.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC
_fast_picrust/uclust_ref_picked_otus/otu_table.normed.biom -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC
_fast_picrust/uclust_ref_picked_otus/picrust_metagenome_table.biom" |
qsub -keo -N stipi -l pvmem=8gb -q memroute

categorize_by_function.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC
_fast_picrust/uclust_ref_picked_otus/picrust_metagenome_table.biom -c
'KEGG Pathways' -l 2 -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC
_fast_picrust/uclust_ref_picked_otus/picrust_metagenome_table.keggl2.biom

categorize_by_function.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC
_fast_picrust/uclust_ref_picked_otus/picrust_metagenome_table.biom -c
'KEGG Pathways' -l 3 -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC
_fast_picrust/uclust_ref_picked_otus/picrust_metagenome_table.keggl3.biom

summarize_taxa_through_plots.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC
_fast_picrust/uclust_ref_picked_otus/picrust_metagenome_table.keggl2.biom -p
/Users/caporaso/analysis/bolnick_fish/picrust_summarize_params.txt -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC
_fast_picrust/uclust_ref_picked_otus/picrust_metagenome_table.kegg_summary_l2/

echo "core_diversity_analyses.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC
_fast_picrust/uclust_ref_picked_otus/picrust_metagenome_table.keggl2.biom -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_stickleback/ucrC
_fast_picrust/uclust_ref_picked_otus/core_div_keggl2_even1000000/ -m
/Users/caporaso/analysis/bolnick_fish/stickleback_map.txt -e 1000000 -
-nonphylogenetic_diversity --suppress_taxa_summary -aO 20" | qsub -keo
-N sticd2 -l pvmem=16gb -q memroute
```
```
Perch


```
Lab

echo "pick_reference_otus_through_otu_table.py -i
/Users/caporaso/analysis/bolnick_fish/11oct2012/slout_R1/seqs.fna -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_lab/ucrC_fast_pi
crust/ -p /Users/caporaso/analysis/bolnick_fish/ucrC_fast_params.txt -r
/Users/caporaso/data/img_gg_otus_18may2012/rep_set/97_otus_img_gg_18ma
y2012.corrected.fasta -ao 50" | qsub -keo -N labpicrust -l pvmem=8gb -q memroute

normalize_by_copy_number.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_lab/ucrC_fast_pi
crust/uclust_ref_picked_otus/otu_table.biom -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_lab/ucrC_fast_pi
crust/uclust_ref_picked_otus/otu_table.normed.biom

echo "predict_metagenomes.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_lab/ucrC_fast_pi
crust/uclust_ref_picked_otus/otu_table.normed.biom -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_lab/ucrC_fast_pi
crust/uclust_ref_picked_otus/picrust_metagenome_table.biom" | qsub -keo -N labpi -l pvmem=8gb -q memroute

categorize_by_function.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_lab/ucrC_fast_pi
crust/uclust_ref_picked_otus/picrust_metagenome_table.biom -c 'KEGG Pathways' -l 2 -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_lab/ucrC_fast_pi
crust/uclust_ref_picked_otus/picrust_metagenome_table.kegg12.biom

categorize_by_function.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_lab/ucrC_fast_pi
crust/uclust_ref_picked_otus/picrust_metagenome_table.biom -c 'KEGG Pathways' -l 3 -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_lab/ucrC_fast_pi
crust/uclust_ref_picked_otus/picrust_metagenome_table.kegg13.biom

summarize_taxa_through_plots.py -i
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_lab/ucrC_fast_pi
crust/uclust_ref_picked_otus/picrust_metagenome_table.kegg12.biom -p
/Users/caporaso/analysis/bolnick_fish/picrust_summarize_params.txt -o
/Users/caporaso/analysis/bolnick_fish/24oct2012/slout_lab/ucrC_fast_pi
crust/uclust_ref_picked_otus/kegg_summary_l2/

Converting Dan’s null samples to BIOM
def f(lines):
    otu_ids = [e.strip('"')[:-1] for e in lines[0].split()]
    sample_ids = []
    data = []
    for line in lines[1:]:
        fields = line.strip().split()
        sample_ids.append(fields[0].strip('"'))
        data.append([int(round(float(e))) for e in fields[1:]])
    return table_factory(array(data).T, sample_ids, otu_ids)

t = f(list(open('./nullOTUs.txt','U')))
open('nullOTUs.biom','w').write(t.getBiomFormatJsonString("Greg Caporaso"))
t = f(list(open('./nullOTUs10000.txt','U')))
open('nullOTUs10000.biom','w').write(t.getBiomFormatJsonString("Greg Caporaso"))
t = f(list(open('./nullOTUs3000.txt','U')))
open('nullOTUs3000.biom','w').write(t.getBiomFormatJsonString("Greg Caporaso"))

def g(lines,sid):
    otu_ids = []
    data = []
    for line in lines[1:]:
        line = line.strip().split()
        otu_ids.append(line[0].strip('"X' ).strip('"'))
        data.append(int(round(float(line[1]))))
    return table_factory(array([data]).T, [sid], otu_ids)

t = g(list(open('populationNullOTUs3000.txt','U')),'popNull')
open('populationNullOTUs3000.biom','w').write(t.getBiomFormatJsonString("Greg Caporaso"))

t = g(list(open('populationNullOTUs10000.txt','U')),'popNull')
open('populationNullOTUs10000.biom','w').write(t.getBiomFormatJsonString("Greg Caporaso"))

t = f(list(open('montecarlo_populationOTUs.txt','U')))  
open('montecarlo_populationOTUs.biom','w').write(t.getBiomFormatJsonString("Greg Caporaso"))

Alpha diversity on null OTU tables

```
```

echo "alpha_diversity.py -i /Users/caporaso/analysis/bolnick_fish/14mar2013/null_biom/nullOTUs10000.biom -m observed_species -o"
/Users/caporaso/analysis/bolnick_fish/14mar2013/null_biom/nullOTUs1000
0_observed_species.txt" | qsub -keo -N nullos10k -l pvmem=8gb -q
memroute

echo "alpha_diversity.py -i
/Users/caporaso/analysis/bolnick_fish/14mar2013/null_biom/nullOTUs3000
.biom -t /Users/caporaso/data/gg_12_10_otus/trees/97_otus.tree -m
PD_whole_tree -o
/Users/caporaso/analysis/bolnick_fish/14mar2013/null_biom/nullOTUs3000
_PD.txt" | qsub -keo -N nullPD3k -l pvmem=8gb -q memroute

echo "alpha_diversity.py -i
/Users/caporaso/analysis/bolnick_fish/14mar2013/null_biom/nullOTUs3000
.biom -m observed_species -o
/Users/caporaso/analysis/bolnick_fish/14mar2013/null_biom/nullOTUs3000
_observed_species.txt" | qsub -keo -N nullos3k -l pvmem=8gb -q
memroute

echo "alpha_diversity.py -i
/Users/caporaso/analysis/bolnick_fish/14mar2013/null_biom/populationNu
lOTUs3000.biom -t
/Users/caporaso/data/gg_12_10_otus/trees/97_otus.tree -m PD_whole_tree
-o
/Users/caporaso/analysis/bolnick_fish/14mar2013/null_biom/populationNu
lOTUs3000_PD.txt" | qsub -keo -N popPD3k -l pvmem=8gb -q memroute

echo "alpha_diversity.py -i
/Users/caporaso/analysis/bolnick_fish/14mar2013/null_biom/populationNu
lOTUs3000.biom -m observed_species -o
/Users/caporaso/analysis/bolnick_fish/14mar2013/null_biom/populationNu
lOTUs3000_observed_species.txt" | qsub -keo -N popos3k -l pvmem=8gb -q
memroute

echo "alpha_diversity.py -i
/Users/caporaso/analysis/bolnick_fish/14mar2013/null_biom/populationNu
lOTUs10000.biom -t
/Users/caporaso/data/gg_12_10_otus/trees/97_otus.tree -m PD_whole_tree
-o
/Users/caporaso/analysis/bolnick_fish/14mar2013/null_biom/populationNu
lOTUs10000_PD.txt" | qsub -keo -N popPD10k -l pvmem=8gb -q memroute

echo "alpha_diversity.py -i
/Users/caporaso/analysis/bolnick_fish/14mar2013/null_biom/populationNu
lOTUs10000.biom -m observed_species -o
/Users/caporaso/analysis/bolnick_fish/14mar2013/null_biom/populationNu
11OTUs10000_observed_species.txt" | qsub -keo -N popos10k -l pvmem=8gb -q memroute

echo "alpha_diversity.py -i
/Users/caporaso/analysis/bolnick_fish/14mar2013/null_biom/montecarlo_populationOTUs.biom -t
/Users/caporaso/data/gg_12_10_otus/trees/97_otus.tree -m PD_whole_tree -o
/Users/caporaso/analysis/bolnick_fish/14mar2013/null_biom/montecarlo_populationOTUs_PD.txt" | qsub -keo -N mcPD10k -l pvmem=8gb -q memroute

echo "alpha_diversity.py -i
/Users/caporaso/analysis/bolnick_fish/14mar2013/null_biom/montecarlo_populationOTUs.biom -m observed_species -o
/Users/caporaso/analysis/bolnick_fish/14mar2013/null_biom/montecarlo_populationOTUs_observed_species.txt" | qsub -keo -N mcos10k -l pvmem=8gb -q memroute

13 June 2013
Analysis of SA13009, SA13062, SA13064, and SA13069 These MiSeq sequence runs contain 16S amplicons from gut microbiota of experimentally-fed lab-reared male and female stickleback, as well as microbiota 16S amplicons from aquarium water and food sources.

echo "split_libraries_fastq.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13009/Undetermined_S0_L001_R1_001.fastq.gz -b
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13009/Undetermined_S0_L001_I1_001.fastq -m
/Users/caporaso/analysis/bolnick_fish/13june2013/map_for_SA13009_CS.txt --rev_comp_mapping_barcodes -o
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13009/slout_R1/" | qsub -keo -N slSA13009

echo "split_libraries_fastq.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13062/Undetermined_S0_L001_R1_001.fastq.gz -b
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13062/Undetermined_S0_L001_I1_001.fastq -m
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13062_trc.txt -o
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13062/slout_R1_trc/
--barcode_type 6" | qsub -keo -N slSA13062

echo "split_libraries_fastq.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13063/Undetermined_
S0_L001_R1_001.fastq.gz -b
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13063/Undetermined_S0_L001_I1_001.fastq -m
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13063_trc.txt -o
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13063/slout_R1_trc/
--barcode_type 6" | qsub -keo -N slSA13063

echo "split_libraries_fastq.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13064/Undetermined_S0_L001_R1_001.fastq.gz -b
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13064/Undetermined_S0_L001_I1_001.fastq -m
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13064_trc.txt -o
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13064/slout_R1_trc/
--barcode_type 6" | qsub -keo -N slSA13064

echo "split_libraries_fastq.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13069/Undetermined_S0_L001_R1_001.fastq.gz -b
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13069/Undetermined_S0_L001_I1_001.fastq -m
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13069_trc.txt -o
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13069/slout_R1_trc/
--barcode_type 6" | qsub -keo -N slSA13069

SA13009

echo "pick_closed_reference_otus.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13009/slout_R1/seqs.fna -o
/Users/caporaso/data/gg_12_10_otus/taxonomy/97_otu_taxonomy.txt -aO 100" | qsub -keo -N SA13009ucrC -l pvmem=8gb -q memroute

SA13062

echo "pick_closed_reference_otus.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13062/slout_R1_trc/seqs.fna -o
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13062/slout_R1_trc/ucrC_fast/ -p
/Users/caporaso/analysis/bolnick_fish/ucrC_fast_params.txt -r
/Users/caporaso/data/gg_12_10_otus/rep_set/97_otus.fasta -t
/Users/caporaso/data/gg_12_10_otus/taxonomy/97_otu_taxonomy.txt -aO 50" | qsub -keo -N SA13062ucrC -l pvmem=8gb -q memroute

echo "print_biom_table_summary.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13062/slout_R1/ucrC_fast/otu_table.biom -t
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13062/slout_R1/ucrC_fast/otu_table_summary.txt" | qsub -keo -N SA13062summ -l pvmem=8gb -q memroute

echo "core_diversity_analyses.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13062/slout_R1_trc/ucrC_fast/otu_table.biom -t
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13062/slout_R1_trc/ucrC_fast/cd_3000/ -e 3000 -m
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13062_trc.txt -aO 50 --suppress_otu_category_significance" | qsub -keo -N SA13062cd3k -l pvmem=8gb -q memroute

SA13063

echo "pick_closed_reference_otus.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13063/slout_R1_trc/seqs.fna -o
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13063/slout_R1_trc/ucrC_fast/ -p
/Users/caporaso/analysis/bolnick_fish/ucrC_fast_params.txt -r
/Users/caporaso/data/gg_12_10_otus/rep_set/97_otus.fasta -t
/Users/caporaso/data/gg_12_10_otus/taxonomy/97_otu_taxonomy.txt -aO 50" | qsub -keo -N SA13063ucrC -l pvmem=8gb -q memroute

echo "print_biom_table_summary.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13063/slout_R1/ucrC
_fast/otu_table.biom -o /Users/caporaso/analysis/bolnick_fish/13june2013/SA13063/slout_R1/ucrC_fast/otu_table_summary.txt" | qsub -keo -N SA13063summ -l pvmem=8gb -q memroute


SA13064

/Users/caporaso/data/gg_12_10_otus/taxonomy/97_otu_taxonomy.txt -aO 50 | qsub -keo -N SA13064ucrC -l pvmem=8gb -q memroute

echo "print_biom_table_summary.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13064/slout_R1/ucrC_fast/otu_table.biom -o
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13064/slout_R1/ucrC_fast/otu_table_summary.txt" | qsub -keo -N SA13064summ -l pvmem=8gb -q memroute

echo "core_diversity_analyses.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13064/slout_R1_trc/ucrC_fast/otu_table.biom -t
/Users/caporaso/data/gg_12_10_otus/trees/97_otus.tree -o
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13064/slout_R1_trc/ucrC_fast/cd_3000/ -e 3000 -m
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13064_trc.txt -aO 50 --suppress_otu_category_significance" | qsub -keo -N SA13064cd3k -l pvmem=8gb -q memroute

echo "core_diversity_analyses.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13064/slout_R1_trc/ucrC_fast/otu_table.biom -t
/Users/caporaso/data/gg_12_10_otus/trees/97_otus.tree -o
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13064/slout_R1_trc/ucrC_fast/cd_10000/ -e 10000 -m
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13064_trc.txt -aO 50 --suppress_otu_category_significance" | qsub -keo -N SA13064cd10k -l pvmem=8gb -q memroute

SA13069

echo "pick_closed_reference_otus.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13069/slout_R1_trc/seqs.fna -o
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13069/slout_R1_trc/ucrC_fast/ -p
/Users/caporaso/analysis/bolnick_fish/ucrC_fast_params.txt -r
/Users/caporaso/data/gg_12_10_otus/rep_set/97_otus.fasta -t
/Users/caporaso/data/gg_12_10_otus/taxonomy/97_otu_taxonomy.txt -aO 50" | qsub -keo -N SA13069ucrC -l pvmem=8gb -q memroute

echo "print_biom_table_summary.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/SA13069/slout_R1/ucrC_fast/otu_table.biom -o
Combined analyses

# Merge otu tables and mapping files from runs SA13009, SA13062, SA13064, and SA13069
merge_otu_tables.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/combined_17june2013/otu_table.biom

merge_mapping_files.py -m
map_for_SA13009_CS.txt,SA13062_trc.txt,SA13064_trc.txt,SA13069_trc.txt -o combined_17june2013/master_map.txt

# Split into lab and wild OTU tables
filter_samples_from_otu_table.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/combined_17june2013/otu_table.biom
tu_table.biom -o
/Users/caporaso/analysis/bolnick_fish/13june2013/combined_17june2013/otu_table_lab.biom --valid_states "Habitat:lab" -m
/Users/caporaso/analysis/bolnick_fish/13june2013/combined_17june2013/master_map.txt

filter_samples_from_otu_table.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/combined_17june2013/otu_table.biom -o
/Users/caporaso/analysis/bolnick_fish/13june2013/combined_17june2013/otu_table_wild.biom --valid_states "Habitat:lake,stream,marine" -m
/Users/caporaso/analysis/bolnick_fish/13june2013/combined_17june2013/master_map.txt

# run core diversity analyses on each

echo "core_diversity_analyses.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/combined_17june2013/otu_table_lab.biom -m
/Users/caporaso/analysis/bolnick_fish/13june2013/combined_17june2013/master_map.txt -o
/Users/caporaso/analysis/bolnick_fish/13june2013/combined_17june2013/lab_cd_3000/ -e 3000 -t
/Users/caporaso/data/gg_12_10_otus/trees/97_otus.tree --suppress_otu_category_significance -aO 50" | qsub -keo -N lab.cd.3k -l pvmem=8gb -q memroute

echo "core_diversity_analyses.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/combined_17june2013/otu_table_lab.biom -m
/Users/caporaso/analysis/bolnick_fish/13june2013/combined_17june2013/master_map.txt -o
/Users/caporaso/analysis/bolnick_fish/13june2013/combined_17june2013/lab_cd_10000/ -e 10000 -t
/Users/caporaso/data/gg_12_10_otus/trees/97_otus.tree --suppress_otu_category_significance -aO 50" | qsub -keo -N lab.cd.10k -l pvmem=8gb -q memroute

echo "core_diversity_analyses.py -i
/Users/caporaso/analysis/bolnick_fish/13june2013/combined_17june2013/otu_table_wild.biom -m
/Users/caporaso/analysis/bolnick_fish/13june2013/combined_17june2013/master_map.txt -o
/Users/caporaso/analysis/bolnick_fish/13june2013/combined_17june2013/wild_cd_3000/ -e 3000 -t
/Users/caporaso/data/gg_12_10_otus/trees/97_otus.tree -- suppress_otu_category_significance -aO 50" | qsub -keo -N wild.cd.3k -l pvmem=8gb -q memroute

Supplementary Methods

Use of quasibinomial general linear models in OTU analyses

A common practice in analyses of microbial communities is to rarify samples to a lowest common denominator of sequence read depth, or exclude samples with insufficient sequence reads. Doing so allows one to compare microbial community metrics such as species presence/absence or diversity (which are sensitive to sequencing depth) across samples with unequal sequence data. However, because rarification entails discarding data, it is preferable when possible to use analytical methods that allow for retention of all data while accounting for unequal amounts of information. Binomial generalized linear models (GLMs) provide exactly this service, as they model how independent variables predict the frequency of a particular outcome (e.g., counts of a particular OTU) out of some total number of observations (e.g., sequencing depth). The model aims to estimate how an unknown parameter (e.g., OTU frequency) varies as a function of independent variables (e.g., host diet), accounting for the fact that the observed data is a random variable which inevitably deviates from the OTU frequency. The extent of this deviation depends on sample size, consequently the retention of all data for each individual sample allows the model to account for the fact that different sample depths lead to different standard errors in estimating OTU frequency. Thus, binomial GLMs allow one to use all the available data and thereby obtain more correct representation of the error associated with estimating OTU frequencies.

The analysis uses a logit transformation of the original data, and this transformed data is then fit to a linear model. For plotting purposes, we convert model estimates (logit relative abundance, which readers find difficult to interpret) back into simple relative abundance which are more intuitive. The plotted curve fits and confidence intervals are also calculated based on
logit values, and back-calculated into relative abundances. Because of the back-calculation, curve fits and confidence intervals can appear, to the naked eye, to differ from what viewers might expect from a simple linear model. In viewing such plots in this paper, readers should keep in mind that curve fits are back-transformed and thus look different from typical linear models.

One drawback regarding binomial GLMs is that their estimates and p-values can be biased when the data is overdispersed (too many outliers, and the among-sample variance in OTU frequency is much greater than expected given mean OTU frequency. Most OTUs in this study exhibit over-dispersion. Consequently, we instead use quasibinomial GLMs that were developed to provide greater robustness to outliers and overdispersion\textsuperscript{1,2}. To confirm that the quasibinomial GLMs yielded suitable statistical inferences, we permuted our data (shuffling individual diet metrics) and recalcualted GLMs to confirm that the P-values obtained from the quasibinomial GLMs are a suitable measure of the type I error rate. That is, when there was a significant diet effect (P < 0.05), randomized data rarely generated effect size estimates equal to or greater than our observed effect sizes. Quasibinomial GLMs have been used in a few microbial community studies (e.g.,\textsuperscript{3,4}) but in our view represent an underutilized tool.
Supplementary References