


Highly Reproducible 16S Sequencing Facilitates Measurement of Host Genetic Influences on the Stickleback Gut Microbiome

Clayton M. Small,^a Mark Currey,^a Emily A. Beck,^a Susan Bassham,^a  William A. Cresko^a

^aInstitute of Ecology and Evolution, University of Oregon, Eugene, Oregon, USA

ABSTRACT Multicellular organisms interact with resident microbes in important ways, and a better understanding of host-microbe interactions is aided by tools such as high-throughput 16S sequencing. However, rigorous evaluation of the veracity of these tools in a different context from which they were developed has often lagged behind. Our goal was to perform one such critical test by examining how variation in tissue preparation and DNA isolation could affect inferences about gut microbiome variation between two genetically divergent lines of threespine stickleback fish maintained in the same laboratory environment. Using careful experimental design and intensive sampling of individuals, we addressed technical and biological sources of variation in 16S-based estimates of microbial diversity. After employing a two-tiered bead beating approach that comprised tissue homogenization followed by microbial lysis in subsamples, we found an extremely minor effect of DNA isolation protocol relative to among-host microbial diversity differences. Abundance estimates for rare operational taxonomic units (OTUs), however, showed much lower reproducibility. Gut microbiome composition was highly variable across fish—even among cohoused siblings—relative to technical replicates, but a subtle effect of host genotype (stickleback line) was nevertheless detected for some microbial taxa.

IMPORTANCE Our findings demonstrate the importance of appropriately quantifying biological and technical variance components when attempting to understand major influences on high-throughput microbiome data. Our focus was on understanding among-host (biological) variance in community metrics and its magnitude in relation to within-host (technical) variance, because meaningful comparisons among individuals are necessary in addressing major questions in host-microbe ecology and evolution, such as heritability of the microbiome. Our study design and insights should provide a useful example for others desiring to quantify microbiome variation at biological levels in the face of various technical factors in a variety of systems.

KEYWORDS DNA isolation, fish model, host-microbe systems, microbial ecology, repeatability, reproducibility


From early development through senescence, animal and plant hosts interact with their resident microbiota through complex host-microbe relationships, resulting in a diversity of both positive and negative outcomes for host health and fitness. For example, outstanding questions regarding the mappings of host genetic variation to microbiome variation and their role in diseases such as diabetes, obesity, and inflammatory bowel disease (IBD), are a major focus of host-microbe systems biology (1). Indeed, the recognized importance of host-microbe interactions has led to a recent spike in interdisciplinary research efforts, complete with accelerated tool development both molecular and computational in nature. This rapid progress, however, has in some cases meant a lag in the thorough evaluation of the veracity and efficacy of these tools. Understanding host-microbe relationships from ecological, evolutionary, and dis-

Citation Small CM, Currey M, Beck EA, Bassham S, Cresko WA. 2019. Highly reproducible 16S sequencing facilitates measurement of host genetic influences on the stickleback gut microbiome. *mSystems* 4:e00331-19. <https://doi.org/10.1128/mSystems.00331-19>.

Editor Haiyan Chu, Institute of Soil Science, Chinese Academy of Sciences

Copyright © 2019 Small et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to William A. Cresko, wcresko@uoregon.edu.

 Using careful experimental design and intensive sampling of individual stickleback fish, we addressed technical and biological sources of variation in 16S-based estimates of microbial diversity. We found an extremely minor effect of DNA isolation protocol.

Received 22 May 2019

Accepted 19 July 2019

Published 13 August 2019

ease perspectives hinges on estimation of microbial diversity in samples from various host body sites. Although quantification of the microbiome may now be achieved using shotgun metagenomic approaches (2), for instance to measure disease-microbiome associations (3), marker-based techniques such as high-throughput 16S rRNA amplicon sequencing are still the most cost-effective, straightforward, and commonly applied methods for microbial community profiling. However, as researchers extend their work beyond routinely characterized environments such as soil and human fecal samples and into new, diverse study systems (4–6), the adoption and extension of previously optimized techniques should occur cautiously and intentionally. Methodologies for 16S amplicon sequencing should ideally be evaluated at multiple stages (i.e., sample collection and handling through analysis), compared with multiple alternative options, and evaluated with respect to the discriminatory power and precision of diversity analyses based on them. The Microbiome Quality Control Project (7), for example, has addressed some of these issues for human stool and artificial microbial communities, including an effective quantification of laboratory-to-laboratory variation. Other diverse endeavors have evaluated effects of DNA isolation attributes on sequencing-based community inference in corals (8), fleas (9), human saliva (10), and marine biofilms (11), for example, but the foci of these studies did not include quantifying reproducibility and its uncertainty using large samples of among-individual variation.

Research aims may require the direct sampling of whole host organs in animal models, such as the gastrointestinal tract, to obtain an unfiltered view of internal host-microbe relationships. However, with these shifts in sampling strategy come a slew of considerations and obstacles, in part because commercially available kits are designed and optimized for a narrow range of sample types such as soil or human stool. Potential problems with sample processing and library preparation based on these unrefined protocols may include poor DNA integrity, inadequate DNA quantity, host and reagent contamination, and low repeatability, all of which may vary depending on the sample type.

We compared the gut microbiomes of two divergent populations of threespine stickleback (*Gasterosteus aculeatus*) that have been maintained in the same lab environment to evaluate whether our biological conclusions could be affected by the use of three popular DNA isolation protocols. Threespine stickleback fish have repeatedly colonized a diversity of freshwater habitats from ancestral marine populations, resulting in exceptional degrees of within- and among-population genetic and phenotypic variation for countless traits (12–17). As in humans, stickleback populations segregate genetic variation that has arisen and been shaped by natural processes in the wild, unlike the variation generated via laboratory-induced mutations in many model organisms (18). This feature makes stickleback an excellent model for understanding the role of standing host genetic variation in determining phenotypes germane to host-microbe interactions, including microbial community structure itself (19–21). In our initial attempts to isolate DNA from adult stickleback guts for 16S sequencing, we found commonly used DNA isolation protocols, including one specifically designed for microbial samples, untenable owing to low quality and high variance of DNA yield, and fragmentation both within and among protocols. This outcome prompted us to optimize these DNA isolation protocols for adult stickleback guts.

We employed careful experimental design and thorough sampling of laboratory-raised hosts to address both technical and biological sources of variation in 16S-based estimates of microbial diversity (Fig. 1). The relative contributions of individual host and DNA isolation protocol to variation in 16S-based diversity estimates have not been satisfactorily measured in previous studies, due to insufficient biological (individual-level) replication, inadequate parameter estimation, or both. To address this, the technical objectives of our study included a careful comparison of operational taxonomic unit (OTU) relative abundance and diversity (both alpha and beta) across libraries generated from three DNA isolation protocols, followed by formal quantification of reproducibility and its uncertainty. The biological objective of our study was to test for differences in relative OTU abundance and diversity arising from genetic differences

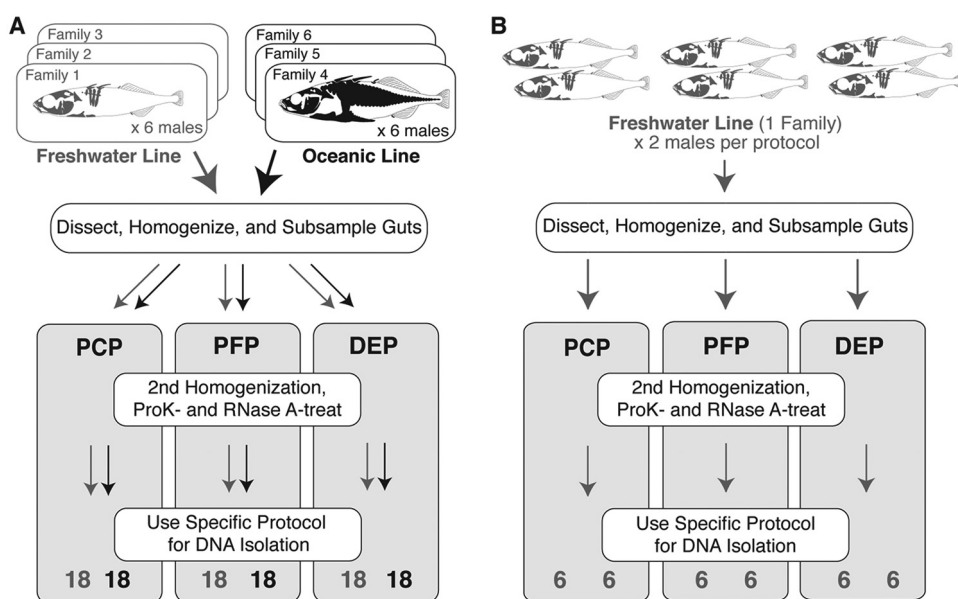


FIG 1 Experimental design to evaluate technical (DNA isolation protocol) and biological (individual and population) variation in 16S sequencing-based diversity metrics (A) and within-individual precision for these metrics (B). The stickleback lines (populations), families, and sample processing steps, and sample sizes used in the current study are shown. In the first experiment (A), we assigned one of three homogenate subsamples from each fish gut to one of the three DNA isolation protocols, phenol-chloroform protocol (PCP), PowerFecal protocol (PFP), or DNeasy protocol (DEP), for a total of 108 extractions across 36 fish. In the second experiment (B), we assigned all six homogenate subsamples from a given fish gut to one of the three DNA isolation protocols. Each protocol was represented by two fish, for a total of 36 extractions.

between two stickleback laboratory lines, one descended from a freshwater lake population and the other from an oceanic population, but both raised and housed in a common environment. The factorial nature of our study design also permitted assessment of statistical interactions between DNA isolation protocol and host genotype, that is, whether any dependency of biological inferences on DNA isolation protocol choice might exist. We also performed a separate experiment in which we measured the precision of each DNA isolation protocol using replicate samples from the same individual.

In general, we found the stickleback gut microbiome to be highly variable even among full siblings reared together and that variation due to the two host genetic backgrounds (population of origin) was detected but smaller than individual-level variation. Unoptimized tissue processing had a major effect on the yield and integrity of DNA isolated using different protocols. However, after employing a two-step bead beating approach consisting of initial tissue homogenization followed by microbial lysis in homogenate subsamples, we found an extremely minor effect of DNA isolation protocol on the ability to understand microbial diversity using 16S data. This is an important finding for those researchers faced with the decision of having to choose among available protocols. Our results indicate that so long as bias during initial tissue processing steps is minimized, the actual choice of kit may be relatively unimportant. Our protocol optimization, study design, and insights both technical and biological should be useful to others who seek to quantify microbial community structure in fish guts and other tissue types using high-throughput 16S sequencing.

RESULTS

Tissue subsampling and two-tiered bead beating improve gut DNA yield and integrity. We compared DNA yield and fragment size distribution between guts first homogenized with steel beads, subsampled, and then treated with a second bead beating step aimed at microbial lysis against guts handled without these modifications. By reducing tissue mass through measured, consistent subsampling and by including

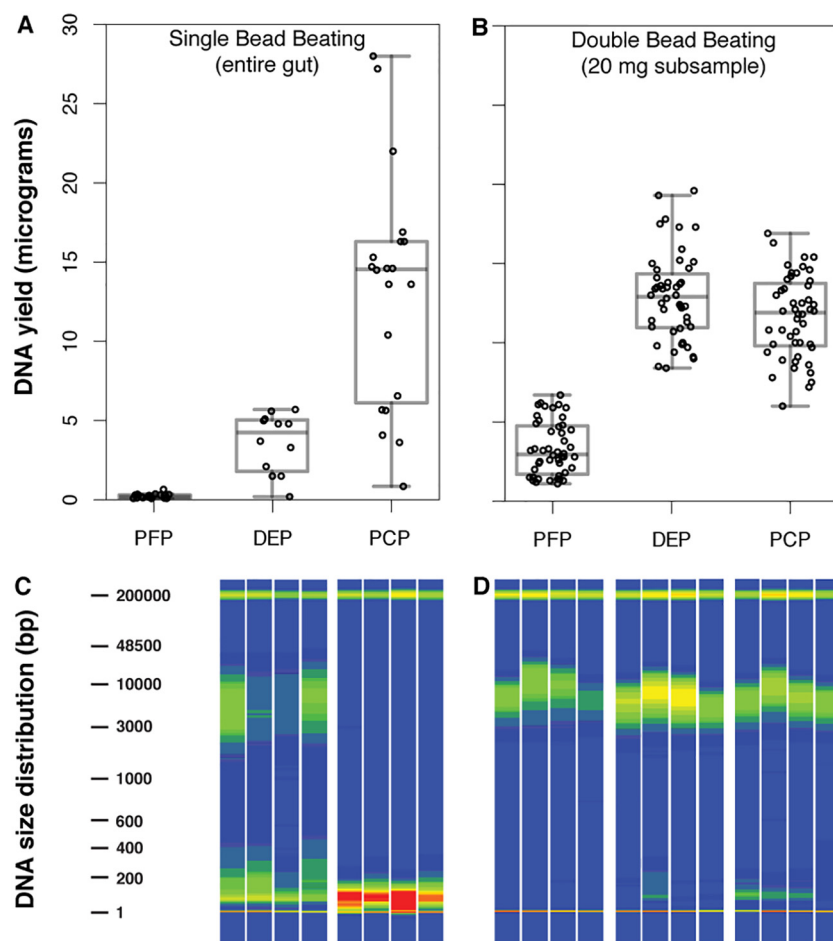


FIG 2 Gut homogenate subsampling and double bead beating improves DNA yield and integrity. (A) DNA yield boxplot for three protocols with single beating and no gut subsampling. (B) DNA yield boxplot for three protocols with double beating and subsampling. (C) Fragment analysis traces for two protocols with single beating and no subsampling. (D) Fragment analysis traces for three protocols with double beating and subsampling. Bands at 1 and 200,000 bp in panels C and D are lower and upper size standards. Dark green corresponds to approximately 400 relative fluorescence units (RFU), whereas dark red corresponds to approximately 2,300 RFU. Fragment analysis data were unavailable for singly beat PFP samples owing to insufficient DNA quantity. PCP, phenol-chloroform protocol; PFP, PowerFecal protocol; DEP, DNeasy protocol.

the second bead beating step, we achieved higher DNA yield and integrity and lower variance among individuals (Fig. 2). Subsampled, double-beat DNA isolates contained more micrograms of DNA on average (Fig. 2A and B). For column-based DNA isolation protocols (PowerFecal protocol [PFP] and DNeasy protocol [DEP]), median yield increased at least twofold. DNA integrity also improved with the modifications (Fig. 2C and D), especially in the case of phenol-chloroform protocol (PCP) isolations. Because whole guts were used for the unmodified protocols, within-fish comparisons of these unmodified protocols and testing of protocol-by-fish interactions were not possible.

Importantly, the among-sample variation, as estimated by the coefficient of variation for all three DNA isolation protocols, was also lower after subsampling and double bead beating (values of 0.557 versus 0.223 for PCP, 0.610 versus 0.509 for PFP, and 0.517 versus 0.214 for DEP [values for single versus double bead beating, respectively]). The coefficient of variation for DNA yield across all singly beat, whole-gut samples was 1.244, compared to 0.527 for doubly beat, subsampled guts. This difference was significant based on the asymptotic test described by Feltz and Miller (22), which assumes a χ^2 -distributed test statistic ($D'AD = 37.376$; $df = 1$; $P = 9.742e-10$).

Bacterial phyla of the stickleback gut microbiome are similar across studies and rearing environments. Rarefaction curves based on samples from 10 to 150,000

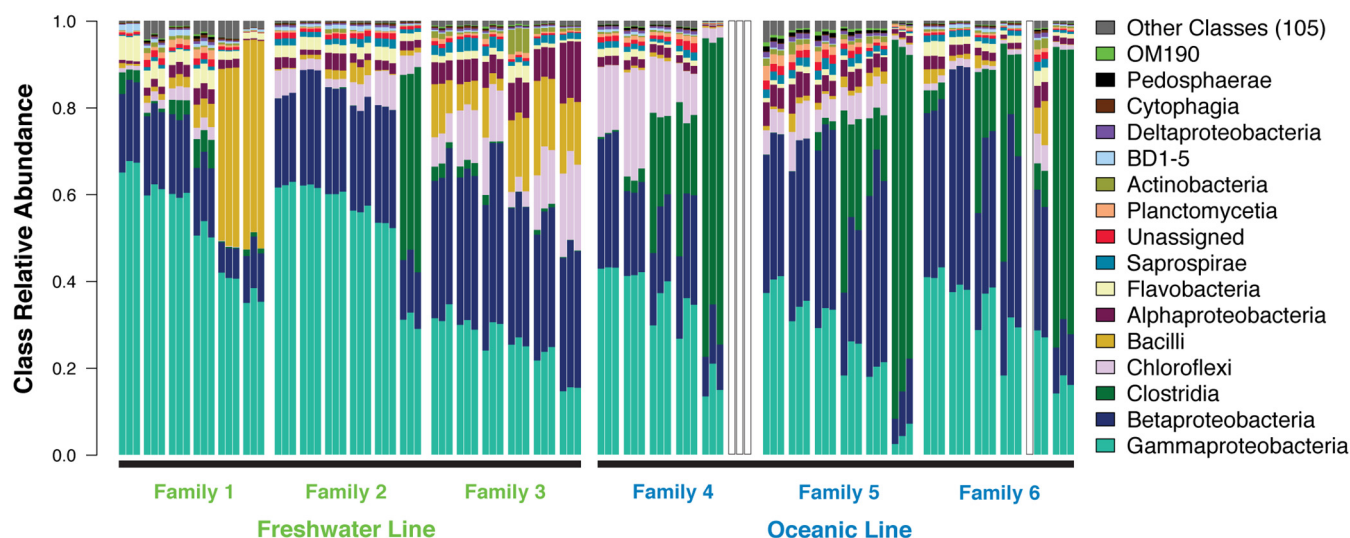


FIG 3 16S-based, class-level profiles of the stickleback gut microbiome vary substantially more by individual host than by DNA isolation method. Class relative abundances demonstrate substantial variation across individuals, families, and populations, but little variation among DNA isolation protocols within individuals. Each bar triplet denotes an individual fish gut, with individual bars representing PCP, PFP, and DEP DNA isolation methods, in that order. Individuals are sorted by mean *Gammaproteobacteria* abundance within each family. One individual from family 4 and the PCP library from an individual in family 6 were not analyzed owing to insufficient coverage.

sequences per library indicated that our final downsampling threshold of 105,000 sequences captured reasonable alpha diversity, given the rate of increase with sampling effort (see Fig. S1A and B in the supplemental material). Considering all 41 experimental fish (35 from the reproducibility experiment and 6 from the repeatability experiment) for which our sequence number threshold of 105,000 was reached, we recovered a mean per-individual OTU richness of 4,378.122 (standard error of the mean [SEM] = 291.360), which reflects OTUs summed across all libraries (each library downsampled to 105,000 sequences) per individual. At the phylum level, we observed a mean richness of 28.146 (SEM = 0.786). The major constituent phyla among the fish in our experiment included *Proteobacteria*, *Firmicutes*, *Chloroflexi*, *Bacteroidetes*, and *Cyanobacteria*, but we observed extensive among-individual variation (Fig. S1F). Phylum-level membership was comparable between the lab-reared fish in our study and both lab-reared and wild-caught fish from Bolnick et al. (23), with the exception of the greater relative abundance of the phylum *Chloroflexi* in our study (Fig. S1E).

Effects of individual hosts on microbial diversity are much larger than those of DNA isolation protocols. We evaluated the relative contributions of individual fish and DNA isolation protocol to overall variance in diversity metrics by treating the libraries from the three different protocols as repeated measurements of each fish. The contribution of DNA isolation protocol to variation in community composition was quite small relative to that of individual fish at the class (Fig. 3; see Fig. S2A in the supplemental material) and species (Fig. S2B) levels, as quantified by high reproducibility estimates for all five alpha diversity metrics (Table 1). Similarly, the effect of DNA isolation protocol on beta diversity was weak relative to the effect of individual (Fig. 4; Fig. S3), as reproducibility with respect to class and species Bray-Curtis dissimilarity and weighted UniFrac was high. Interestingly, reproducibility was substantially lower for unweighted UniFrac (Table 1; Fig. 4G to L). We also observed low reproducibility (although not as extreme as in the QIIME 1 analysis) for unweighted relative to weighted UniFrac after QIIME 2/Deblur denoising (Table 1), a method that should minimize the effects of erroneous sequences. In general, the QIIME 1 and QIIME 2 workflows yielded very similar patterns with respect to relative taxon abundance (Fig. S2A) and beta diversity (Table 1; Fig. 4; Fig. S3A to F). Last, confidence intervals (95% bootstrap) for reproducibility calculated separately for the two stickleback populations overlapped for all nine diversity metrics (see Data Set S1D in the supplemental

TABLE 1 Among-fish variation greatly exceeds within-fish (among-protocol) variation for several diversity metrics, as indicated by high reproducibility estimates^a

Diversity variable	Among-fish variation	Within-fish variation	Reproducibility (95% CI)
Class richness	77.409	11.080	0.875 (0.792, 0.917)
Class evenness	0.011	0.001	0.920 (0.885, 0.965)
Species richness	6606.324	123.393	0.982 (0.946, 0.991)
Species evenness	0.009	0.001	0.923 (0.886, 0.945)
Phylogenetic diversity	287.929	18.949	0.938 (0.870, 0.969)
Class Bray-Curtis	0.082	0.004	0.958 (0.912, 0.979)
Class Bray-Curtis (QIIME 2 with Deblur)	0.081	0.004	0.957 (0.921, 0.979)
Species Bray-Curtis	0.112	0.004	0.966 (0.938, 0.982)
Weighted UniFrac	0.070	0.003	0.955 (0.907, 0.978)
Weighted UniFrac (QIIME 2 with Deblur)	0.023	0.001	0.959 (0.926, 0.978)
Unweighted UniFrac	0.056	0.158	0.263 (0.207, 0.292)
Unweighted UniFrac (QIIME 2 with Deblur)	0.055	0.046	0.542 (0.466, 0.583)

^aVariance component and reproducibility estimates considering all individuals in experiment 1, along with bootstrap 95% confidence intervals (CIs) are shown. The reproducibility values are shown in boldface type.

material), suggesting that reproducibility was consistent for the two different host genetic backgrounds.

Although the overall variance in diversity metrics explained by differences in DNA isolation protocol was small relative to that explained by among-individual differences, we detected a significant effect of DNA isolation protocol for some measures via likelihood ratio tests comparing full and reduced linear mixed models (Table 2; Fig. S4 and Fig. S5). For example, total variation in class richness was explained significantly better by a model including DNA isolation protocol than by a model excluding the term. This effect size was small, however (Fig. S4A). Libraries from DNA isolated using DNeasy (DEP) yielded a modest increase in mean class richness from 52.663 to 54.892, with respect to PowerFecal (PFP), and 53.528 to 54.892 with respect to phenol-chloroform (PCP). We observed a similar trend for species richness and Faith's phylogenetic diversity (Table 2; Fig. S5A and C).

Beta diversity (as measured by class- and species-level Bray-Curtis dissimilarity, unweighted UniFrac, and weighted UniFrac) was significantly influenced by DNA isolation protocol, on average, in all cases ($P < 0.001$; see Data Set S1A for factorial permutational analysis of variance [PERMANOVA] hypothesis test statistics). The effect sizes were once again quite small in the context of among-individual variation, as reflected in nonmetric multidimensional scaling (nMDS) ordinations and pairwise library dissimilarity distributions (Fig. 4; Fig. S3A, C, D, and F).

Relative abundances of individual taxon groups were in some cases affected by DNA isolation protocol, based on comparison of lognormal Poisson generalized linear models. For 20 class-level and 89 species-level OTUs, the model including protocol was a better fit than the model excluding it based on Akaike information criterion (AIC), Bayesian information criterion (BIC), and the false discovery rate (FDR)-controlled likelihood ratio test (Data Set S1B and C). For instance, the class-level groups *Actinobacteria*, BD1-5 ("Gracilibacteria"), and *Thermomicrobia* tended to vary in abundance among DNA isolation protocols within fish consistently (Fig. S6A to C), albeit with quite small effect sizes. The mean downsampled read count for *Actinobacteria*, for example, was 1.317 times higher for PCP than for PFP methods. At the species level, taxonomy groups including *Agromyces* spp., an unassigned species from family *Rodobacteraceae*, and *Tsukamurella* spp. were among the most likely taxa affected by DNA isolation protocol (Fig. S7A to C), also to a minor degree.

We wanted to evaluate the potential for DNA isolation protocol differences to influence the ability to consistently measure among-host differences in relative OTU abundances and to ascertain whether this ability varied as a consequence of the scarcity of a given OTU. We found that reproducibility was indeed positively associated



(Continued on next page)

TABLE 2 Likelihood ratio test statistics^a

Diversity variable	Protocol		Population		Interaction	
	$\chi^2_{df=2}$	P value	$\chi^2_{df=1}$	P value	$\chi^2_{df=2}$	P value
Class richness	8.469	0.015	0.306	0.580	0.878	0.645
Class evenness	0.051	0.975	0.043	0.835	4.324	0.115
Species richness	5.354	0.069	0.230	0.631	5.658	0.059
Species evenness	4.095	0.129	2.243	0.134	11.175	0.004
Phylogenetic diversity	8.642	0.013	0.066	0.798	7.195	0.027

^aLikelihood ratio test (LRT) statistics with degrees of freedom (df) and *P* values for tests of effects of DNA isolation protocol, stickleback population, and interaction between the two. LRTs were conducted by comparing linear mixed models either including or excluding these fixed effects, plus random effects of fish and fish nested within family (see Materials and Methods). Tests with *P* values of <0.05 are shown in boldface type.

with average \log_{10} OTU abundance, based on a fitted logistic model (Fig. 5; Fig. S8A). The slope at inflection (γ) was significantly greater than zero ($\gamma = 0.352$; standard error [SE] = 0.015; $t = 23.977$; $P < 0.0001$). A similar relationship was observed for upper and lower 95% confidence interval (CI) bounds on reproducibility (Fig. 5). We also observed that the relationship between amplicon sequence variant (ASV) reproducibility and mean ASV abundance based on the QIIME 2 analysis was very similar to that for OTUs (Fig. 5; Fig. S8B), suggesting that erroneous sequences are not the fundamental cause of this pattern.

We noted a similar pattern when measuring OTU rarity in a different way: the number of individual hosts in which the OTU was detected. We once again observed that reproducibility of relative OTU abundance estimates was higher for common OTUs (Fig. S8C), and that the lower bound for reproducibility in our sample of hosts increased substantially when the OTU was present in at least 32 of 35 fish (Fig. S8D).

Effects of stickleback population on microbial diversity estimates are subtle and in limited cases contingent on DNA isolation protocol. We did not detect a statistically significant effect of host population (stickleback line) on any of the five alpha diversity metrics using likelihood ratio tests comparing nested full and reduced linear mixed models (Table 2), although class richness, class evenness, and species richness trended toward higher values in the oceanic population relative to the freshwater population (Fig. S4 and S5). We did detect a statistically significant interaction between stickleback population and DNA isolation protocol for species-level and phylogenetic alpha diversity metrics (Table 2). This implies that the effect of isolation protocol may differ depending on biological context, but these effect sizes were also quite small (Fig. S5B and C). For example, the maximum species evenness difference between protocol-population combinations was 0.070, and the maximum phylogenetic diversity difference was 5.370.

Beta diversity, as measured by class- and species-level Bray-Curtis dissimilarity, unweighted UniFrac, and weighted UniFrac, was not significantly influenced by host population after accounting for the nested nature of the data introduced by family structure (Fig. 4; see Data Set S1A for protocol-specific nested PERMANOVA hypothesis test statistics). Family itself was a significant determinant of community dissimilarity for all four metrics (Fig. 4; see Data Set S1A for factorial PERMANOVA hypothesis test statistics), although it should be noted that family was confounded by tank in our design. We did not detect a statistically significant interaction between host family and DNA isolation protocol for any of the four dissimilarity metrics assessed (Data Set S1A).

FIG 4 Legend (Continued)

(A, D, G, and J) Nonmetric multidimensional scaling (nMDS) ordinations from weighted and unweighted UniFrac, showing the three DNA isolation protocols from each individual connected as filled triangles. (B, E, H, and K) The same ordinations, but with individuals plotted as the centroid of each triplet from panels A, D, G, and J, and with 95% confidence ellipses drawn separately for each family. (C, F, I, and L) Pairwise dissimilarity matrix heatmaps representing all libraries. Panels A to C and G to I show QIIME 1-based analyses, and panels D to F and J to L show QIIME 2-based (denoised ASV) analyses. The library order is the same as in Fig. 3. Green ordination symbols represent the freshwater stickleback line, and blue symbols represent the oceanic stickleback line. Individual fish labeled by lowercase letters and corresponding arrows point to outliers in community space.

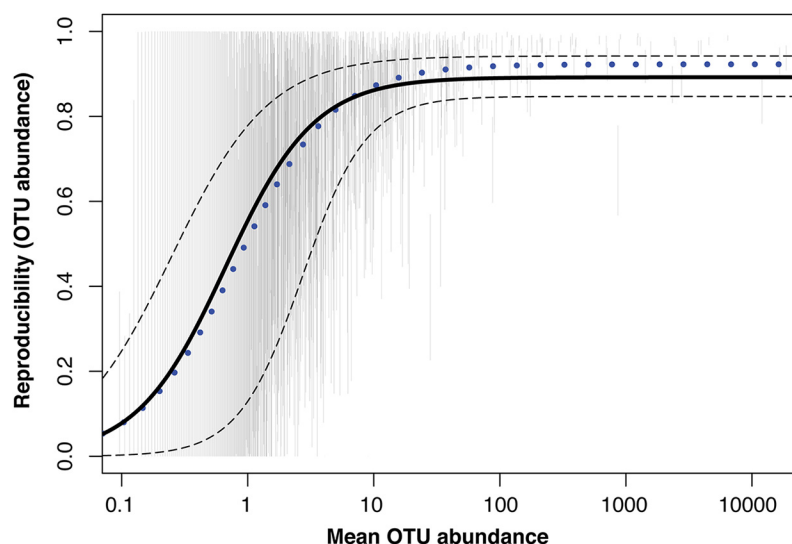


FIG 5 Reproducibility of OTU (and ASV) quantification across DNA isolation protocols increases nonlinearly with \log_{10} -transformed mean relative OTU abundance. Vertical gray lines represent 95% confidence intervals (CIs) for reproducibility estimates of 2,278 OTUs observed in at least 10 of 35 experimental fish. The solid black line represents predicted reproducibility values from a logistic model fit to the OTU data. Dashed lines represent predicted upper and lower bound CI values for reproducibility, also from logistic models fit to the OTU data. The blue dotted line represents predicted reproducibility values from a logistic model fit to 783 ASVs resulting from the QIIME 2 analysis and observed in at least 10 of 35 experimental fish.

Finally, among-fish community dissimilarity was highly correlated between DNA isolation protocols based on Mantel tests (Data Set S1D).

We also evaluated whether relative abundances for taxonomic groups (class- and species-level) might be affected by host population, based on comparison of lognormal Poisson generalized linear models accounting for family nestedness. Although no likelihood ratio tests were statistically significant after controlling the FDR at 0.1, 3 class-level groups and 25 species-level groups showed evidence for a population effect by virtue of a delta AIC of >2 , a delta BIC of >0 , and an uncorrected likelihood ratio test (LRT) P value of <0.05 (Data Set S1B and C). Class-level groups BD-7, an unassigned class from *Bacteroidetes*, and *Clostridia* were all three enriched in abundance in the oceanic population relative to the freshwater population (Fig. S6D to F). The effect size of population on the *Clostridia* group abundance, for example, was quite large. The mean oceanic *Clostridia* count was 23259.559 (SEM = 3880.291), whereas the mean freshwater *Clostridia* count was 3867.704 (SEM = 1417.775). Three species-level groups with especially strong tendencies toward host population differences were the oceanic enriched *Sphingobacterium multivorum* group, the freshwater-enriched *Plesiomonas shigelloides* group, and an oceanic enriched, unassigned group from the family *Clostridiaceae* (Fig. S7D and F).

We detected a statistically significant interaction between host population (accounting for family) and DNA isolation protocol for six class-level and 25 species-level groups (Data Set S1B and C), based on a delta AIC of >2 , a delta BIC of >0 , and an LRT FDR controlled at 0.10. However, effect sizes for this interaction type were again relatively small, as demonstrated by the abundance of a *Sphingobacterium multivorum* group across population-protocol combinations (Fig. S7D). In this case, the mean population difference in *S. multivorum* count was highest for DEP (4.497), followed by 1.688 and 2.912 for PCP and PFP, respectively.

Precision of gut microbiome diversity measurements is high and similar across three DNA isolation protocols. We performed repeated measurements of individual stickleback gut microbiomes obtained from replicate aliquots from whole-gut homogenates and using a single DNA isolation protocol per gut (see Fig. 1B). 16S data generated from the same fish were extremely similar in taxonomic composition, relative

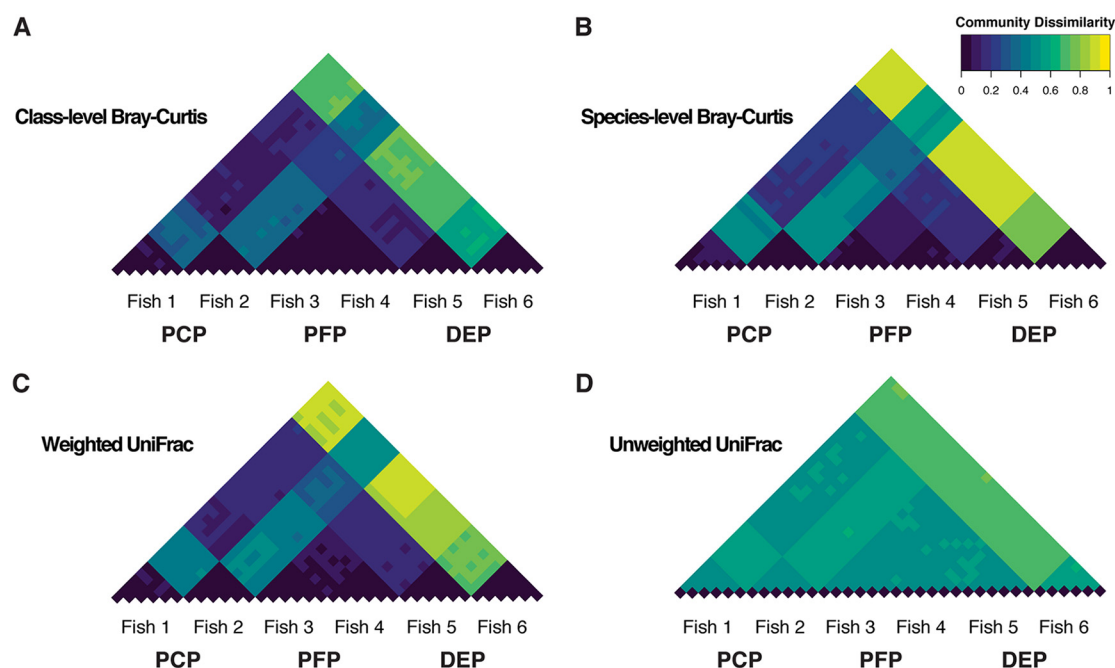


FIG 6 Precision of beta diversity measurements is consistently high among DNA isolation methods, but the relative magnitude of within- and among-fish community dissimilarity depends on the dissimilarity metric. Pairwise dissimilarity matrix heatmaps for the precision experiment, including class- and species-level Bray-Curtis (A and B), and weighted and unweighted UniFrac (C and D), illustrate low within-fish dissimilarity and higher among-fish dissimilarity. This pattern, however, is less evident for unweighted UniFrac.

to among-fish comparisons (Fig. S9A and D). We analyzed within-individual variation based on the six gut subsamples per fish and found no significant effect of DNA isolation protocol on precision for five alpha diversity metrics (Data Set S1E), including class-level richness and evenness (Fig. S9), species richness and evenness (Fig. S9E and F), and phylogenetic diversity (Fig. S9G). Similarly, we found no evidence for a significant effect of protocol on precision with respect to beta diversity (Data Set S1D), including class- and species-level Bray-Curtis dissimilarity and weighted and unweighted UniFrac (Fig. 6; Fig. S9H to K). Furthermore, and consistent with our across-protocol reproducibility analysis, the average degree of within-fish dispersion, relative to among-fish dispersion, was especially low for unweighted UniFrac (Fig. 6).

DISCUSSION

One surprising insight from the recent characterization of microbiomes using high-throughput sequencing has been the extensive diversity among individual hosts of the same species (24–27). The fundamental sources of this interindividual variation remain an active area of research. Rather than assuming that individual variation is vastly larger than technical variation, and therefore insignificant, we and others (7, 28, 29) argue that the relative magnitude of biological and technical variance components should be measured using strong experimental design. This is particularly relevant in the context of new and rapidly changing technologies for quantifying microbial diversity. Indeed, variation in microbial diversity metrics may be heavily influenced by technical factors in some cases, especially when molecular protocols are vastly different or suboptimal. For instance, extremely low yields of microbial DNA exacerbate the influence of contaminant species (30, 31), which could negatively impact biological inference. Furthermore, inferences based on some diversity metrics might be more susceptible to technical variation, particularly those metrics that are more heavily influenced by sequences from rare species whose abundance estimates may be more subject to sampling error.

In our experience, suspicions and intuition about the severe importance or unimportance of technical variation for 16S-based microbial ecology inference have been

extensively discussed. However, these effects have not been precisely quantified and evaluated outside the purview of “mock communities” (30, 32) or differences among research groups working on large-scale, collaborative efforts to understand the human microbiome (7). While these studies and several others (for example, references 9, 10, and 33 to 35), have been useful in identifying potential sources of technical variation that may or may not restrict or bias biological inferences based on among-individual variation, insufficient biological replication and/or inability to isolate specific technical factors have limited their scope of inference. For example, authors working on the Microbiome Quality Control Project point out that their carefully designed study was “unable to assign significance to any specific fixed effects (i.e., individual protocol variables), since in the small MBQC-base these were in large part confounded with individual handling and bioinformatics laboratories” (7).

To our knowledge, no prior study has sampled dozens of individual hosts, in combination with the controlled assignment of technical factor levels, to effectively quantify reproducibility (and its uncertainty) in a biologically relevant context, although smaller-scale studies have addressed similar themes (see above). We wanted to fill this important void with a well-replicated comparison involving one potential source of technical variation—DNA isolation protocol—and individual-level variation in the ecological and evolutionary context of stickleback host genetic differences.

One of our significant findings is that the earliest steps in sample handling are critical for improving downstream results. We found the process of dual bead beating with tissue homogenate subsampling essential to the quantity and quality of DNA isolated from adult threespine stickleback guts. Without this process, both lower yields with higher variance and more-fragmented DNA were certain. The PowerFecal column-based isolation protocol suffered most severely from a lack of double bead beating and subsampling, perhaps owing to an overloading of the column and subsequent failure to elute large DNA fragments. These results are significant, as low DNA yields are known to amplify any effects of contamination (30, 31). Decreasing among-sample variance in DNA attributes such as quantity, therefore, should reduce nonbiological variation among 16S-based microbial profiles. In principle, this will increase the power of statistical analyses, thereby reducing cost in the number of biological samples needed. In the specific case of our study, these modifications were absolutely essential to establishing a reasonable comparison of DNA isolation protocols. Before embarking on 16S sequencing for a large study, we recommend similar subsampling and optimization for large sample types or sample types that have not yet been tested with commercial kits. We also recommend that DNA yield and quality distributions for at least a random subset of samples be reported in published 16S and metagenomic studies.

Our results confirm that among-individual differences in stickleback gut communities are extensive (23), consistent with work on the gut microbiome of humans (25) and other hosts (27, 36). Although mean relative abundances of phyla in our experimental fish were qualitatively similar to those of other stickleback populations and environments (23), we found substantial variation in community composition even among male full siblings housed in the same tank. This is significant, as most ecological, evolutionary, and biomedical studies of host-associated microbes rely on an understanding of among-host differences in the microbiome. However, previous studies have not satisfactorily quantified the extent to which observed individual differences might be due to technical variation introduced by factors such as the DNA isolation protocol. We measured reproducibility (across three DNA isolation protocols) for a number of commonly used microbial diversity metrics. We found that alpha and beta diversity measurements of the stickleback gut microbiome were very reproducible, despite having applied three fundamentally different DNA isolation protocols. As stated previously, this high reproducibility is predicated upon the proper initial treatment of tissue through double beating and subsampling.

Interestingly, we observed an exception to high reproducibility in the case of unweighted UniFrac, although 95% CI lower bounds were still above zero. Because unweighted UniFrac does not account for differences in 16S sequence abundance, it

magnifies the effect of rare sequences (37). To evaluate whether this property was due entirely to rare, artificial OTUs originating from sequencing error (38, 39), we reanalyzed beta diversity reproducibility for unweighted and weighted UniFrac using a recent denoising approach to OTU/ASV picking (Deblur via QIIME 2). We found a substantial (>2-fold) increase in across-protocol reproducibility for unweighted UniFrac after denoising. However, despite this increased reproducibility for unweighted UniFrac in QIIME 2 compared to QIIME 1, a clear difference in reproducibility between weighted and unweighted metrics persisted in the QIIME 2 analyses (Table 1).

We emphasize that our data do not suggest that unweighted UniFrac-based measures of beta diversity are not reproducible in the general sense, but rather in the case of our study, they were lower than metrics that take abundance into account. The interpretation of reproducibility is contingent on the level of variation researchers wish to understand, and our objective was to study reproducibility across DNA isolation protocols (and repeatability across tissue subsamples) with respect to among-individual variation. Unweighted UniFrac is known to be especially sensitive to sampling bias (40). One recent study showed that unweighted UniFrac applied to resampled sequences from the same HMP tongue dorsum libraries projected large within-library variation relative to among-library variation, when this should be very low (41). This insight, along with our current study, suggests that sampling bias associated with rare sequences disproportionately affects the potential to explain among-individual variance with unweighted UniFrac compared to other metrics, an important consideration that researchers should make when interpreting microbiome data, especially in light of differences between similar, individual hosts. For example, the repeatability of measurement for a given trait has historically been understood as an upper bound on the heritability estimate for that trait (42–44).

Although reproducibility among DNA isolation methods was extremely high, we observed statistically significant effects of DNA isolation protocol on some measurements of the microbiota, including class richness, Faith's phylogenetic diversity, and relative abundance for at least 20 classes. Of these 20 classes, information regarding Gram stain was available for 17, and four of these (23.53%) were Gram positive. Of all 58 classes tested for an effect of DNA isolation protocol, Gram stain data were available for 41, and 5 of these (12.20%) were Gram positive. The four Gram-positive classes we identified as likely subject to an influence of DNA isolation protocol (*Actinobacteria*, *Acidimicrobiia*, *Bacilli*, and *Clostridia*) were all more abundant in phenol-chloroform (PCP) samples than in DNA samples isolated using column-based (PFP and DEP) protocols, so it is possible that the chemical properties of organic compounds like phenol subtly enrich representation from Gram-positive lineages. It should be noted, however, that the effect sizes for DNA isolation protocol in the above analyses were rather small (see Results), and our experimental design was well powered to detect even minor effects of DNA isolation protocol owing to many within-individual comparisons. Nevertheless, if researchers are interested in specific microbial lineages for a particular study, they should be aware that DNA isolation protocols may indeed influence abundance estimates for individual taxa.

We also examined the relationship between average OTU abundance and among-protocol reproducibility. On the basis of our stickleback data, a very clear, sigmoidal relationship suggested that reproducibility was indeed lowest, on average, for rare OTUs, and that it improved substantially up to a mean OTU count of 10. The nature of this function will almost certainly vary among systems and among sequencing depths (recall that these data were downsampled to 105,000 reads per library), but it provides a general reference for those interested in the reliable measurement of rare taxa with 16S sequencing. Even with high biological replication and relatively deep sampling, low repeatability for some organisms may be unavoidable. This pattern is fundamentally related to the reduced reproducibility we observed for unweighted UniFrac, in that increased sampling error for rare sequences makes among-individual comparisons less tractable.

We designed a second, small experiment to compare precision of 16S-based com-

munity measurements (six gut subsamples per fish) among the three DNA isolation protocols. We observed no significant difference in precision for richness, evenness, phylogenetic diversity, or beta dispersion, among DNA isolation protocols. It should be noted, however, that our sample of individual fish per protocol was limited (just two), and that among-fish variation in microbial community structure was extensive (see Fig. S9D in the supplemental material). As a result, our power to detect among-protocol differences in precision was limited.

Although between-experiment comparisons were not a focus of our study, we did observe that one taxonomic group (family *Phormidiaceae*) was relatively high in abundance among most samples from the precision experiment (Fig. S9D) and low among most fish from the DNA protocol reproducibility experiment. This difference could reflect a temporal shift in *Phormidiaceae* (a cyanobacterial lineage) in our stickleback housing system, as these two experiments were conducted months apart.

With added confidence that optimized DNA isolation protocols contribute minimally to among-library variation, we then explored whether several factors might explain individual host differences in the stickleback gut microbiome, with a special interest in host genetic background. Recent studies of animal hosts, mostly featuring mammals, have reported a stronger influence of environmental variables relative to host genetic variation on gut microbiome variation (45–47). In our study, host family significantly explained beta diversity among individuals, but family effects were confounded by tank effects. Although all tanks in our study shared a common water system, the immediate tank environment is likely to influence host-associated microbes. Future studies should address host genetic effects like those at the family level by raising individuals related to different degrees in replicated common garden experiments informed by traditional quantitative genetics principles.

The population of origin was not a statistically significant factor for most of the gut microbiome traits we measured; however, individual species-level groups such as those associated with *Sphingobacterium multivorum* and *Plesiomonas shigelloides* showed strong evidence for an association with stickleback line. Taxonomic groups from the phylum *Firmicutes* (namely, the family *Clostridiaceae* and the genus *Turicibacter*) also differed in abundance between the two stickleback lines. Host genotype influences on *Firmicutes* appear to be common in mammals (48), and *Turicibacter* has been shown to be heritable in both humans and mice (49, 50). While a subtle effect of host genotype on the stickleback gut microbiome is consistent with the aforementioned insights from animal hosts, it should be interpreted with some caution, as the nested nature of our design and sampling only three families per population limited our statistical power to test population-level hypotheses. Future studies that experimentally control environmental effects carefully and sample more genetic variation at the population level (e.g., genome-wide association studies and large-scale common garden experiments) should provide the power to confirm these still largely untested contributions to among-individual microbiome variation. Notably, we detected minimal evidence for statistical interaction between host population and DNA isolation protocol. Again, although some of these tests were statistically significant, the associated effect sizes were rather small (Fig. S5B and C and Fig. S7D). The mechanistic causes of these subtle interactions are unknown, but it is possible that inorganic or organic compounds in the guts differing in concentration between freshwater and oceanic stickleback could copurify with DNA and affect downstream steps in library construction such as PCR, in a manner specific to DNA from some microbial lineages.

Our current study revealed high reproducibility across the three protocols we tested, and minimal concern that choice of DNA isolation protocol interacts with biological factors of interest. In our experience, the DNeasy protocol (DEP) required the shortest handling time, so we have adopted it in current studies of the stickleback gut microbiome. Negligible influence from these technical factors may not be the case for other biological systems or sample types, however, so we strongly encourage other researchers to design their studies in ways similar to those presented here in order to properly measure and minimize sources of technical variance. The payoff in limiting

technical variation is potentially large in terms of cost of reagents, time, and animal resources, especially when true biological signal is subtle. This concept is, of course, easily extended beyond 16S data sets to high-throughput RNA sequencing (RNA-Seq) and other high-throughput sequencing data. In summary, the complexity of communities and the sampling process can affect reproducibility and repeatability, but as we show, not always to a great extent. The magnitude of these effects depends on the biology of the system at hand and the diversity metric in question. Without properly quantifying relevant technical and biological variance components of sequencing-based microbial diversity metrics, however, it is impossible for our research community to move forward with confidence in addressing core questions about host-microbe interactions and microbial ecology in general.

MATERIALS AND METHODS

Rearing of adult stickleback, evaluation of DNA quality, and experimental design. (i) Threespine stickleback husbandry and collection of gut samples. We collected guts from male adult threespine stickleback (*Gasterosteus aculeatus*) derived from wild-caught Alaskan populations, which have been maintained in the laboratory for at least 10 generations. All individuals were raised to an age of 12 to 16 months, using standard protocols described in a previous publication (14). Briefly, fish were raised from embryos fertilized *in vitro*, and larvae were fed twice daily with brine shrimp nauplii and Zeigler larval AP100 diet. An equal parts mixture of Golden Pearl 800–1000 micron juvenile diet, Otohime C1, Zeigler zebrafish diet, and Hikari tropical micro pellets was fed twice daily to fish as juveniles and adults. Fish were housed in a large, single-source recirculating system (5 ppt salinity) with a 10% daily water change, in 20-liter tanks at a density of 20 to 30 fish per tank. Tanks were randomly positioned on a single shelving rack roughly equidistant from the incoming water source. We maintained fish in an approximately 1:1 sex ratio, with a photoperiod of 8-h light and 16-h dark (including 30-min dawn and dusk).

To reduce among-individual variation owing to sex (23), we sampled males only, as confirmed by DNA isolation from caudal fin clips and a PCR-based sex genotyping procedure (see reference 20). Fish were also not fed for 24 h prior to sampling to reduce the amount of food in the gut. Upon euthanasia by a lethal dose of MS222, the entire gastrointestinal tract of each fish, including the esophagus to just anterior of the urogenital opening, was carefully removed, weighed, and quickly flash frozen in liquid nitrogen in a screw-top tube containing nuclease-free homogenization beads (see below).

(ii) Initial assessment of DNA quality from three unmodified DNA isolation protocols. We evaluated yield and quality of DNA isolated using a standard phenol-chloroform-isoamyl alcohol protocol (PCP) and two commercial kit protocols commonly used in microbiome studies: MO BIO's PowerFecal kit (PFP) and the Qiagen's DNeasy Blood and Tissue kit (DEP). We dissected whole guts from 52 adult stickleback in our fish facility, randomly assigning 20 each to PCP and PFP, and 12 to DEP. In this initial assessment of unmodified DNA isolation protocols, we used the entire gut to be consistent with previous studies of stickleback gut microbiota (21, 23). In the case of PCP and DEP, each whole gut was dissected and flash frozen (see above) in a tube containing five nuclease-free 3.2-mm stainless steel beads (catalog no. SSB32; Next Advance). In the case of PFP, each gut was frozen in a tube containing ~1-mm garnet beads, which are standard issue for the kit. Next, we removed each tube containing a gut and beads from -80°C and followed the manufacturer's recommendations, with a few exceptions. We added 400 μl of Qiagen buffer ATL in the case of PCP and DEP, and 750 μl of Bead Solution in the case of PFP. We then homogenized guts in a Thermo Savant FastPrep FP120 with three 40-s bouts of beating at intensity level 6.5. Using the entire homogenate for each sample, we followed the instructions in the manuals of the DEP and PFP kits. In the case of PCP, we followed the same posthomogenization lysis instructions as for DEP, then combined a 650- μl aliquot of the lysate with 650 μl of 25:24:1 equilibrated phenol-chloroform-isoamyl alcohol in a phase-lock gel tube, mixed by inversion, and centrifuged at $18,000 \times g$ in a bench-top microcentrifuge for 5 min. We transferred the aqueous layer to a new phase-lock gel tube, added 500 μl of 24:1 chloroform-isoamyl alcohol, mixed by inversion, centrifuged again, and transferred the aqueous layer to a 1.5-ml tube. We precipitated the DNA using 450 μl of isopropanol and 5-min centrifugation at $5,800 \times g$, washed the pellet once with 70% ethanol and twice with 95% ethanol, air dried the pellets for 10 to 15 min, and resuspended the pellet in 100 μl of Qiagen buffer EB. We quantified DNA resulting from all three protocols using a Qubit 2.0 fluorometer (Invitrogen) and evaluated fragment length distributions using a fragment analyzer (Advanced Analytical).

(iii) Design of experiments. After optimizing these protocols (see below), we designed two separate experiments to evaluate the effects of several variables on inferred microbial community structure. For these experiments, we used two entirely different sets of fish sampled from our fish facility at different times and obtained multiple subsamples per fish gut. For the first experiment, each one of the 36 individual intestinal tracts was homogenized and then divided into three separate subsamples to be analyzed using different DNA isolation protocols. This design effectively allowed us to measure among-protocol technical variation based on within-host comparisons (reproducibility). These 36 fish were from two different lab lines, specifically a freshwater (FW) line derived from the natural population "Boot Lake," and an oceanic (OC) line derived from the natural population "Rabbit Slough." We sampled three different full-sib families from each line and six fish per family. Each family was housed in a different tank but in the same recirculating system. Our study design therefore enabled an assessment of the influence

of host genetic background on the microbiota, a topic of great interest for the field of host-microbe interactions (45, 46, 51). Figure 1A illustrates these components of the experimental design. Finally, using a sample of six FW males from a seventh full-sib family, we sampled two guts for each protocol type (six guts total), but we repeated six measurements (six subsamples) per homogenized gut to evaluate the within-fish repeatability (precision) of each DNA isolation protocol. Figure 1B reflects the precision assessment inherent in our study design.

Modified DNA isolation methods and Illumina 16S amplicon sequencing. (i) Standardized preprocessing with 20-mg subsampling. In order to effectively compare the three DNA methods using our experimental design, we standardized tissue preprocessing and introduced uniform mass tissue subsampling for all gut samples. We removed each gut (in a screw-cap tube with five nuclease-free 3.2-mm stainless steel beads) from -80°C , added $800\ \mu\text{l}$ of prewarmed Qiagen buffer ATL (with $0.5\ \mu\text{M}$ EDTA), and homogenized using the FastPrep FP120. We used three bouts of 40-s beating at intensity level 6.5 to achieve a homogeneous mixture and then pipetted volumes from each sample to achieve 20-mg subsamples, as calculated from the original mass of each gut. For the reproducibility experiment (Fig. 1A), three subsamples from each gut were taken, one for each of the three DNA isolation protocols described below. For the repeatability experiment (Fig. 1B), six subsamples from each gut were taken, all for a single DNA isolation protocol. Subsamples were transferred to screw-top tubes containing $100\ \mu\text{l}$ of 0.15-mm zirconium oxide beads (catalog no. ZrOB015; Next Advance) for future mechanical lysis of microbes, flash frozen in liquid nitrogen, and stored at -80°C . These aliquots then received one of the three DNA isolation treatments below. We also performed two “negative-control” DNA isolations for each isolation protocol, in which the protocol was conducted starting with no gut and with or without the addition of proteinase K (see below). Because our objectives did not include comparisons of accuracy among DNA isolation protocols, as others (30, 32) have evaluated this, we did not incorporate controlled assemblages of microbes (“mock communities”) in our experimental design.

(ii) Phenol-chloroform-isoamyl alcohol protocol (PCP). We removed homogenate subsamples from -80°C and added prewarmed Qiagen buffer ATL to bring the total ATL volume in the tube to $676\ \mu\text{l}$. We then homogenized the samples by two 40-s bouts in the FastPrep FP120 at level 6.5, briefly spun tubes, added $20\ \mu\text{l}$ of proteinase K (20 mg/ml), mixed by aspiration, and incubated at 56°C for 30 min. We added $4\ \mu\text{l}$ of RNase A (100 mg/ml), mixed by aspiration, incubated at 37°C for 30 min, and then transferred the entire volume of lysate to a new 1.5-ml tube, to which $500\ \mu\text{l}$ of 25:24:1 equilibrated phenol-chloroform-isoamyl alcohol was added. At this point, we carried out the remainder of the phenol-chloroform protocol exactly as described above.

(iii) MoBio PowerFecal protocol (PFP). We removed homogenate subsamples from -80°C and added prewarmed PowerFecal bead solution to bring the total volume of solution (ATL plus bead solution) in the tube to $750\ \mu\text{l}$. We then added $60\ \mu\text{l}$ of PowerFecal C1 solution, incubated at 65°C for 10 min, and then homogenized for two 40-s bouts in the FastPrep FP120 at level 6.5. We briefly spun tubes, added $20\ \mu\text{l}$ of proteinase K (20 mg/ml), mixed by aspiration, and incubated at 56°C for 30 min. Then we added $4\ \mu\text{l}$ of RNase A (100 mg/ml), mixed by aspiration, incubated at 37°C for 30 min, and followed the instructions in the PowerFecal manual, starting with step 7, which is a centrifugation for 1 min at $13,000 \times g$ to pellet and remove solids from the lysate. Finally, we eluted with $50\ \mu\text{l}$ Qiagen buffer EB and quantified DNA concentration as described above.

(iv) Qiagen DNeasy protocol (DEP). We removed homogenate subsamples from -80°C and added prewarmed Qiagen buffer ATL to bring the total ATL volume in the tube to $776\ \mu\text{l}$. We then homogenized samples by two 40-s bouts in the FastPrep FP120 at level 6.5, briefly spun tubes, added $20\ \mu\text{l}$ of proteinase K (20 mg/ml), mixed by aspiration, and incubated at 56°C for 30 min. We added $4\ \mu\text{l}$ of RNase A (100 mg/ml), mixed by aspiration, and incubated at 37°C for 30 min. We combined the entire volume of lysate with $800\ \mu\text{l}$ of Qiagen buffer AL and $800\ \mu\text{l}$ of 100% ethanol to a new 15-ml screw-cap tube to ensure adequate volume for the additional reagents. After briefly mixing the tube contents by aspiration, we transferred $600\ \mu\text{l}$ of the mixture to a DNeasy spin column, spun at $6,000 \times g$ in a bench-top microcentrifuge for 1 min, discarded flowthrough, and repeated four times, for the remainder of the mixture. Next, we added $500\ \mu\text{l}$ of Qiagen solution AW1, spun at $6,000 \times g$ for 1 min, added $500\ \mu\text{l}$ of Qiagen solution AW2, spun at $18,000 \times g$ for 3 min, added $500\ \mu\text{l}$ of 80% ethanol, and spun at $18,000 \times g$ for 3 min. Finally, we eluted DNA with $100\ \mu\text{l}$ buffer EB and quantified DNA concentration as described above.

(v) Construction of 16S rRNA gene amplicon libraries and Illumina sequencing. We submitted a $25\ \text{ng}/\mu\text{l}$ dilution from each gut DNA sample to the University of Oregon Genomics and Cell Characterization Core Facility (GC3F) for library amplification, cleanup, and sequencing. All six negative-control samples were not diluted, as the level of DNA in these samples was lower than the detection limit of our fluorometer. The GC3F generated 16S libraries from 200 ng of DNA template per sample using custom primers 515F (5'-AATGATACGGCGACCACCGAGATCTACACxxxxxxTATGGTAATTGTGTGCCAGCMGCCGCGGTAA3') and 806R (5'-CAAGCAGAAGACGGCATACGAGATxxxxxxAGTCAGTCAGCGGACTACHVGGGTWTCTAAT3'), which are based on the primers described by Caporaso et al. (68) and which amplify the “V4” 16S region but enable dual indexing (indexes represented by x's in the above sequences). A cocktail including $12.5\ \mu\text{l}$ NEBNext Q5 Hot Start HiFi PCR master mix, $4.5\ \mu\text{l}$ of $2.79\ \mu\text{M}$ primer mix, and $8\ \mu\text{l}$ of DNA template, was used for each library PCR. The thermal profile was as follows: initial denaturation at 98°C for 30 s, followed by 22 cycles, with 1 cycle consisting of 98°C for 10 s, 61°C for 20 s, and 72°C for 20 s, followed by a final extension step at 72°C for 2 min. Each library was cleaned twice using $20\ \mu\text{l}$ of Omega Mag-Bind RxnPure Plus beads and quantified by a Qubit fluorometer, at which point $9.235\ \text{ng}$ of DNA from each library (less for negative controls) were pooled. The GC3F quantified the library pool using quantitative PCR (qPCR), combined it with a complex RNA-Seq

library from an unrelated project, and sequenced 161-nucleotide (nt) paired-end reads in two Illumina HiSeq 2500 lanes.

Processing of Illumina 16S data and statistical inference. (i) Sequence filtering and OTU picking. Processing of sequences and OTU picking were primarily achieved using accessory scripts from QIIME version 1.9.1 (52) and to a lesser extent our own custom scripts. We overlapped ends of read pairs using QIIME's *join_paired_ends.py*, and we demultiplexed the merged reads using QIIME's *extract_barcodes.py* and *split_libraries_fastq.py*. We used default arguments, except that we allowed a maximum of two barcode errors when demultiplexing and invoked read truncation at 30 or more consecutive low-quality base calls. This filtering process yielded 119.94 million total reads from the 144 gut libraries, and 4,957 total reads from the six negative-control libraries. We performed open reference OTU picking using QIIME's *pick_open_reference_otus.py* with default settings (53, 54), which uses the Greengenes version 13.8 database as its reference (55). We then parsed OTUs by taxonomic assignment and removed all OTUs of mitochondrial or chloroplast origin to exclude the influence of host- and food-derived DNA. The total number of filtered OTU-assigned reads from gut libraries was 87.109 million (mean = 604,920.326; SEM = 26,291.498). Negative-control libraries produced exceedingly small numbers of OTU-assigned reads (418 for PCP, 348 for PCP_proK, 1,201 for PFP, 1,115 for PFP_proK, 644 for DEP, and 367 for DEP_proK). Given such a small likely contribution of contaminating template to gut libraries, we did not exclude gut OTUs based on information from the negative-control libraries.

To normalize coverage, we downsampled all libraries in the OTU table (including those from both reproducibility and repeatability experiments) to 105,000 sequences each, which we deemed an optimal trade-off between sequencing depth and retention of samples for analysis. This led to the exclusion of four libraries from the reproducibility study: all three samples from one FW fish (bringing the total number of fish analyzed in this experiment to 35) and the PCP library from a second FW fish. We then generated count summary tables for all taxonomy levels using QIIME's *summarize_taxa.py*, but downstream analyses described in this report feature phylum, class, species, or individual OTU counts (see Results). We used QIIME's *core_diversity_analyses.py* to generate phylogenetic diversity metrics separately for the reproducibility and precision studies, including Faith's phylogenetic diversity (56), unweighted UniFrac (57), and weighted UniFrac (37).

To evaluate the robustness of our primary results to different OTU picking strategies, we also performed ASV (amplicon sequence variant) definition and enumeration using the Deblur denoising approach (38), implemented in QIIME 2 (version 2018.8.0). Briefly, using QIIME 2, we merged the demultiplexed read pairs with *vsearch* (*join-pairs*), quality filtered using default settings of *quality-filter* (*q-score-joined*), and denoised using *deblur* (*denoise-16S*) with a trimming length of 251 nt. To assign taxonomy to the ASVs, we first trained a sequence classifier based on the GreenGenes 13_8 99%-clustered OTU database using *feature-classifier* (*fit-classifier-naive-bayes*), then applied it to our ASVs using *feature-classifier* (*classify-sklearn*). ASVs were filtered and enumerated in samples as described above, leading to 5852 ASVs and 58.598 million total ASV-assigned reads across gut libraries. Subsequent analyses relied on downsampling to 74,300 reads per sample, which was the deepest downsampling level that would allow direct comparison with the samples used in the primary (QIIME 1) analyses. Phylogenetic diversity metrics were calculated using *diversity* (*core-metrics-phylogenetic*) and based on a phylogenetic tree inferred for ASVs using *phylogeny* (*align-to-tree-mafft-fasttree*). Downstream analyses using the QIIME 1-generated OTUs (and a subset of these analyses using the QIIME 2-generated ASVs) were based on the respective downsampled count tables and phylogenetic diversity metrics described above and were conducted using version 3.3.2 of the R statistical language (58).

(ii) Reproducibility of alpha and beta diversity metrics. We evaluated relative contributions of biological (among-fish) and technical (within-fish) variation using a repeated-measures linear model framework. In particular, and following Lessels and Boag (59), we calculated "repeatability" (reproducibility in this case) as:

$$r_{\text{protocol}} = \frac{s_A^2}{(s_W^2 + s_A^2)}$$

where s_A^2 is the among-fish variance component and s_W^2 is the within-fish (among-protocol) variance component, as calculated from mean squares in an analysis of variance (ANOVA) (59). r_{protocol} ranges from 0 to 1, with high values indicating increasingly small contributions of DNA isolation protocol relative to individual host contributions, which are of interest to ecologists and evolutionary biologists. A perfect reproducibility of 1.0, for example, would indicate zero within-fish variance, that is, no effect of protocol in this situation. Alternatively, a repeatability of 0.5 would be interpreted as equal variance contributions from individual and protocol. We calculated reproducibility for class richness, class evenness, species richness, species evenness, and Faith's phylogenetic diversity using variance components estimated by the *lmer* function from the R package *lme4* (60). For each metric, we also resampled (with replacement) 35 individual fish 500 times and used the distribution of reproducibility values from the 500 bootstrap replicates to calculate 95% confidence intervals (CIs). We also applied this approach to multivariate measures of community dissimilarity (class- and species-level Bray-Curtis dissimilarity and weighted and unweighted UniFrac) by extracting the above variance components using the R function *nested.npmanova* from the *BiodiversityR* package (61). Furthermore, we calculated population-specific reproducibilities (and confidence intervals) for all of the above variables to evaluate whether reproducibility differed depending on host population. We also estimated reproducibility for weighted and unweighted UniFrac based on denoised ASVs generated from the QIIME 2 workflow to evaluate the potential influence of error introduced by OTU picking.

(iii) Testing effects of DNA isolation protocol, host population, and their interaction on alpha diversity, beta diversity, and relative taxon abundances. For the same five alpha diversity variables mentioned above, we evaluated significance of the fixed effects of protocol and stickleback population using mixed linear models that included the random effect of individual nested within family. Note that family is indistinguishable from tank in our design, so family and tank effects cannot be separated. We fit full and reduced models using lmer from the R package lme4 (60) and tested null hypotheses of no population, protocol, and population-by-protocol interaction effects on each diversity variable using likelihood ratio tests.

We also tested the influence of these factors on four measures of community dissimilarity (beta diversity) using two permutational analysis of variance (PERMANOVA) tests (62), as necessitated by the complexity of our experimental design. First, we evaluated the effects of DNA isolation protocol, family (tank), and their interaction using the adonis2 function from the R package vegan (63), allowing within-fish comparisons by stratified permutation. Second, we evaluated the effect of population, accounting for nonindependence of individuals within the same family (tank), separately for the three DNA isolation protocols using the function nested.npmanova from the BiodiversityR package (61). Finally, to test whether among-fish community dissimilarity was correlated between DNA isolation protocol pairs, we performed Mantel tests using the mantel function from the R package vegan (63).

To evaluate effects of DNA isolation and stickleback population on relative abundances of class-level and species-level OTU groups (“L3” and “L7,” respectively, from *summarize_taxa.py*), we fit generalized linear mixed models that included the random effect of individual nested within family. We considered only those taxonomy groups represented by at least five counts in at least nine libraries. Given the overdispersed nature of these count data, we fit Poisson lognormal models using the glmer function from the R package lme4 (60) by including an observation-level effect in each model and by specifying the “Poisson” family of generalized linear model. Because 16S data provide information about relative, as opposed to absolute, abundances of the organisms in each sample, it should be acknowledged that differences in OTU and taxonomic group counts among samples could reflect compositional differences in the community as opposed to organism-specific ones. We evaluated the potential importance of each effect for each taxonomy group using false-discovery rate-controlled (64) likelihood ratio tests, Akaike information criterion (AIC) and Bayesian information criterion (BIC), in combination with interpretation of effect sizes.

(iv) OTU rarity and reproducibility of relative OTU abundance estimates. We measured the reproducibility of relative abundance estimates for 2,278 individual OTUs that were present in one or more libraries from at least 10 of the 35 fish from our reproducibility experiment. We estimated reproducibility using the same general repeated-measures framework above, except that we used the R package rptR (65) to calculate reproducibility and its 95% confidence interval for each OTU. We used the rpt function of rptR because it allows the flexibility of fitting an overdispersed Poisson generalized linear model and implements computationally efficient CI construction by parametric bootstrapping. We characterized the relationship between OTU abundance reproducibility and average OTU abundance by fitting three logistic models: one for the point estimate, one for its 95% CI upper bound, and one for its 95% CI lower bound. The logistic model parameterization was as follows:

$$y_i = \frac{\alpha}{1 + e^{(\beta - x_i)\gamma}}$$

where y is the reproducibility of abundance, its CI upper bound, or its CI lower bound for OTU i , x is the logarithm to base 10 of the among-library mean abundance of OTU i , α is the asymptote, β is the inflection point in units of x , and γ is the steepness of the relationship at inflection. We fit these logistic models using the R package nls2 (66). To evaluate whether denoised, QIIME2 sequence variants (ASVs) showed a similar relationship between reproducibility and mean abundance, we performed the same type of analysis as described above, but with reproducibility point estimates among 783 individual ASVs that were present in one or more libraries from at least 10 of the 35 fish.

(v) Diversity metric precision. We measured the precision with which each of several alpha and beta diversity metrics was estimated, comparing across the three DNA isolation protocols. This was made possible by sampling two guts for each protocol and six technical replicates per gut (Fig. 1B). In this case, within-fish variance was attributable only to tissue subsampling and technical differences among the six libraries prepared identically, and not protocol differences. Because we sampled only two fish for each DNA isolation protocol, we could not effectively apply the repeatability framework used in the reproducibility study (see above), which relies on adequately sampling across-individual variation. Instead, we conducted Levene’s tests according to Sokal and Rohlf (67) to test whether the average absolute deviation from group (individual gut) medians was significantly different among the three DNA isolation protocols. We applied this test to class- and species-level richness and evenness and to Faith’s phylogenetic diversity. We conducted the same type of test for class- and species-level Bray-Curtis dissimilarity and for weighted and unweighted UniFrac, but we considered distance between observation and group (individual gut) centroid as the response variable.

Data accessibility. Merged, cleaned 16S sequences for all individuals and sample metadata are available via the following figshare doi links: <https://doi.org/10.6084/m9.figshare.7616264.v1> and <https://doi.org/10.6084/m9.figshare.7616276.v1>, respectively.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mSystems.00331-19>.

FIG S1, TIF file, 2.7 MB.

FIG S2, TIF file, 0.8 MB.

FIG S3, TIF file, 1.7 MB.

FIG S4, TIF file, 0.8 MB.

FIG S5, TIF file, 0.6 MB.

FIG S6, TIF file, 1.3 MB.

FIG S7, TIF file, 1.3 MB.

FIG S8, TIF file, 0.9 MB.

FIG S9, TIF file, 1.6 MB.

DATA SET S1, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

We thank K. Milligan-Myhre and Z. Stephens for important discussions that helped motivate this work. H. Tavalire provided valuable recommendations regarding the OTU rarity analysis. K. Guillemin, R. Parthasarathy, B. Bohannan, J. Eisen, and many members of the University of Oregon (UO) META Center for Systems Biology provided helpful advice and feedback throughout the project. We are also grateful to M. Weitzman and D. Turnbull from the UO GC3F for critical assistance with library preparation and sequencing.

This work was funded by National Institutes of Health grants P50GM098911 (to W.A.C., K. Guillemin, et al.) and RR032670 (to W.A.C.) and Oregon Research Excellence Funds (to W.A.C.). E.A.B. was supported by a National Institutes of Health NRSA fellowship (F32GM122419).

We declare that we have no conflicts of interest.

C.M.S., M.C., S.B., and W.A.C. conceived and designed the study. C.M.S. and M.C. collected the data, and C.M.S. performed the analyses. C.M.S., M.C., and E.A.B. wrote an initial draft of the manuscript, and all authors contributed to subsequent versions.

REFERENCES

- Wu H, Tremaroli V, Backhed F. 2015. Linking microbiota to human diseases: a systems biology perspective. *Trends Endocrinol Metab* 26: 758–770. <https://doi.org/10.1016/j.tem.2015.09.011>.
- Sharpton TJ. 2014. An introduction to the analysis of shotgun metagenomic data. *Front Plant Sci* 5:209. <https://doi.org/10.3389/fpls.2014.00209>.
- Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng Y, Zhang D, Jie Z, Wu W, Qin Y, Xue W, Li J, Han L, Lu D, Wu P, Dai Y, Sun X, Li Z, Tang A, Zhong S, Li X, Chen W, Xu R, Wang M, Feng Q, Gong M, Yu J, Zhang Y, Zhang M, Hansen T, Sanchez G, Raes J, Falony G, Okuda S, Almeida M, LeChatelier E, Renault P, Pons N, Batto J-M, Zhang Z, Chen H, Yang R, Zheng W, Li S, Yang H, Wang J, Ehrlich SD, Nielsen R, Pedersen O, Kristiansen K, Wang J. 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490: 55–60. <https://doi.org/10.1038/nature11450>.
- Hyde ER, Navas-Molina JA, Song SJ, Kueneman JG, Ackermann G, Cardona C, Humphrey G, Boyer D, Weaver T, Mendelson JR, III, McKenzie VJ, Gilbert JA, Knight R. 2016. The oral and skin microbiomes of captive Komodo dragons are significantly shared with their habitat. *mSystems* 11:e00046-16. <https://doi.org/10.1128/mSystems.00046-16>.
- Nuccio EE, Anderson-Furgeson J, Estera KY, Pett-Ridge J, de Valpine P, Brodie EL, Firestone MK. 2016. Climate and edaphic controllers influence rhizosphere community assembly for a wild annual grass. *Ecology* 97: 1307–1318. <https://doi.org/10.1890/15-0882.1>.
- Torres JP, Tianero MD, Robes JMD, Kwan JC, Biggs JS, Concepcion GP, Olivera BM, Haygood MG, Schmidt EW. 2017. Stenotrophomonas-like bacteria are widespread symbionts in cone snail venom ducts. *Appl Environ Microbiol* 83:e01418-17. <https://doi.org/10.1128/AEM.01418-17>.
- Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, Schwager E, Crabtree J, Ma S, Microbiome Quality Control Project Consortium, Abnet CC, Knight R, White O, Huttenhower C. 2017. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. *Nat Biotechnol* 35:1077–1086. <https://doi.org/10.1038/nbt.3981>.
- Weber L, DeForce E, Apprill A. 2017. Optimization of DNA extraction for advancing coral microbiota investigations. *Microbiome* 5:18. <https://doi.org/10.1186/s40168-017-0229-y>.
- Lawrence AL, Hii SF, Chong R, Webb CE, Traub R, Brown G, Slapeta J. 2015. Evaluation of the bacterial microbiome of two flea species using different DNA-isolation techniques provides insights into flea host ecology. *FEMS Microbiol Ecol* 91:fiv134. <https://doi.org/10.1093/femsec/fiv134>.
- Raju SC, Lagstrom S, Ellonen P, de Vos WM, Eriksson JG, Weiderpass E, Rounge TB. 2018. Reproducibility and repeatability of six high-throughput 16S rDNA sequencing protocols for microbiota profiling. *J Microbiol Methods* 147:76–86. <https://doi.org/10.1016/j.mimet.2018.03.003>.
- Corcoll N, Osterlund T, Sinclair L, Eiler A, Kristiansson E, Backhaus T, Eriksson KM. 2017. Comparison of four DNA extraction methods for comprehensive assessment of 16S rRNA bacterial diversity in marine biofilms using high-throughput sequencing. *FEMS Microbiol Lett* 364: fnx139. <https://doi.org/10.1093/femsle/fnx139>.
- Bell MA, Foster SA (ed). 1994. The evolutionary biology of the threespine stickleback. Oxford University Press, Oxford, United Kingdom.
- Colosimo PF, Peichel CL, Nereng K, Blackman BK, Shapiro MD, Schluter D, Kingsley DM. 2004. The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biol* 2:E109. <https://doi.org/10.1371/journal.pbio.0020109>.
- Cresko WA, Amores A, Wilson C, Murphy J, Currey M, Phillips P, Bell MA, Kimmel CB, Postlethwait JH. 2004. Parallel genetic basis for repeated

- evolution of armor loss in Alaskan threespine stickleback populations. *Proc Natl Acad Sci U S A* 101:6050–6055. <https://doi.org/10.1073/pnas.0308479101>.
15. Cresko WA, McGuigan KL, Phillips PC, Postlethwait JH. 2007. Studies of threespine stickleback developmental evolution: progress and promise. *Genetica* 129:105–126. <https://doi.org/10.1007/s10709-006-0036-z>.
 16. Glazer AM, Killingbeck EE, Mitros T, Rokhsar DS, Miller CT. 2015. Genome assembly improvement and mapping convergently evolved skeletal traits in sticklebacks with genotyping-by-sequencing. *G3 (Bethesda)* 5:1463–1472. <https://doi.org/10.1534/g3.115.017905>.
 17. Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet* 6:e1000862. <https://doi.org/10.1371/journal.pgen.1000862>.
 18. Albertson RC, Cresko WA, Detrich HW, Postlethwait JH. 2009. Evolutionary mutant models for human disease. *Trends Genet* 25:74–81. <https://doi.org/10.1016/j.tig.2008.11.006>.
 19. Milligan-Myhre K, Small CM, Mittge EK, Agarwal M, Currey M, Cresko WA, Guillemin K. 2016. Innate immune responses to gut microbiota differ between oceanic and freshwater threespine stickleback populations. *Dis Model Mech* 9:187–198. <https://doi.org/10.1242/dmm.021881>.
 20. Small CM, Milligan-Myhre K, Bassham S, Guillemin K, Cresko WA. 2017. Host genotype and microbiota contribute asymmetrically to transcriptional variation in the threespine stickleback gut. *Genome Biol Evol* 9:504–520. <https://doi.org/10.1093/gbe/evx014>.
 21. Smith CC, Snowberg LK, Caporaso JG, Knight R, Bolnick DI. 2015. Dietary input of microbes and host genetic variation shape among-population differences in stickleback gut microbiota. *ISME J* 9:2515–2526. <https://doi.org/10.1038/ismej.2015.64>.
 22. Feltz CJ, Miller GE. 1996. An asymptotic test for the equality of coefficients of variation from k populations. *Stat Med* 15:647–658. [https://doi.org/10.1002/\(SICI\)1097-0258\(19960330\)15:6<647::AID-SIM184>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1097-0258(19960330)15:6<647::AID-SIM184>3.0.CO;2-P).
 23. Bolnick DI, Snowberg LK, Hirsch PE, Lauber CL, Org E, Parks B, Lusi AJ, Knight R, Caporaso JG, Svanbäck R. 2014. Individual diet has sex-dependent effects on vertebrate gut microbiota. *Nat Commun* 5:4500. <https://doi.org/10.1038/ncomms5500>.
 24. Adair KL, Wilson M, Bost A, Douglas AE. 2018. Microbial community assembly in wild populations of the fruit fly *Drosophila melanogaster*. *ISME J* 12:959–972. <https://doi.org/10.1038/s41396-017-0020-x>.
 25. Human Microbiome Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486:207–214. <https://doi.org/10.1038/nature11234>.
 26. Kueneman JG, Parfrey LW, Woodhams DC, Archer HM, Knight R, McKenzie VJ. 2014. The amphibian skin-associated microbiome across species, space and life history stages. *Mol Ecol* 23:1238–1250. <https://doi.org/10.1111/mec.12510>.
 27. Sullam KE, Rubin BE, Dalton CM, Kilham SS, Flecker AS, Russell JA. 2015. Divergence across diet, time and populations rules out parallel evolution in the gut microbiomes of Trinidadian guppies. *ISME J* 9:1508–1522. <https://doi.org/10.1038/ismej.2014.231>.
 28. Knight R, Vrbancac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, Gonzalez A, Kosciorek T, McCall LI, McDonald D, Melnik AV, Morton JT, Navas J, Quinn RA, Sanders JG, Swofford AD, Thompson LR, Tripathi A, Xu ZZ, Zaneveld JR, Zhu Q, Caporaso JG, Dorrestein PC. 2018. Best practices for analysing microbiomes. *Nat Rev Microbiol* 16:410–422. <https://doi.org/10.1038/s41579-018-0029-9>.
 29. Poussin C, Sierro N, Boue S, Battey J, Scotti E, Belcastro V, Peitsch MC, Ivanov NV, Hoeng J. 2018. Interrogating the microbiome: experimental and computational considerations in support of study reproducibility. *Drug Discov Today* 23:1644–1657. <https://doi.org/10.1016/j.drudis.2018.06.005>.
 30. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 12:87. <https://doi.org/10.1186/s12915-014-0087-z>.
 31. Bender JM, Li F, Adisetiyo H, Lee D, Zabih S, Hung L, Wilkinson TA, Pannaraj PS, She RC, Bard JD, Tobin NH, Aldrovandi GM. 2018. Quantification of variation and the impact of biomass in targeted 16S rRNA gene sequencing studies. *Microbiome* 6:155. <https://doi.org/10.1186/s40168-018-0543-z>.
 32. Yuan S, Cohen DB, Ravel J, Abdo Z, Forney LJ. 2012. Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One* 7:e33865. <https://doi.org/10.1371/journal.pone.0033865>.
 33. Kashinskaya EN, Andree KB, Simonov EP, Solovyev MM. 2017. DNA extraction protocols may influence biodiversity detected in the intestinal microbiome: a case study from wild Prussian carp, *Carassius gibelio*. *FEMS Microbiol Ecol* 93:fiw240. <https://doi.org/10.1093/femsec/fiw240>.
 34. Videvall E, Strandh M, Engelbrecht A, Cloete S, Cornwallis CK. 2017. Direct PCR offers a fast and reliable alternative to conventional DNA isolation methods for gut microbiomes. *mSystems* 2:e00132-17. <https://doi.org/10.1128/mSystems.00132-17>.
 35. Wagner Mackenzie B, Waite DW, Taylor MW. 2015. Evaluating variation in human gut microbiota profiles due to DNA extraction method and inter-subject differences. *Front Microbiol* 6:130. <https://doi.org/10.3389/fmicb.2015.00130>.
 36. Kreisinger J, Cizkova D, Vohanka J, Pialek J. 2014. Gastrointestinal microbiota of wild and inbred individuals of two house mouse subspecies assessed using high-throughput parallel pyrosequencing. *Mol Ecol* 23:5048–5060. <https://doi.org/10.1111/mec.12909>.
 37. Lozupone CA, Hamady M, Kelley ST, Knight R. 2007. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* 73:1576–1585. <https://doi.org/10.1128/AEM.01996-06>.
 38. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, Knight R. 2017. Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems* 2:e00191-16. <https://doi.org/10.1128/mSystems.00191-16>.
 39. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581–583. <https://doi.org/10.1038/nmeth.3869>.
 40. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. 2011. UniFrac: an effective distance metric for microbial community comparison. *ISME J* 5:169–172. <https://doi.org/10.1038/ismej.2010.133>.
 41. Wong RG, Wu JR, Gloor GB. 2016. Expanding the UniFrac toolbox. *PLoS One* 11:e0161196. <https://doi.org/10.1371/journal.pone.0161196>.
 42. Boake C. 1989. Repeatability: its role in evolutionary studies of mating behavior. *Evol Ecol* 3:173–182. <https://doi.org/10.1007/BF02270919>.
 43. Dohm MR. 2002. Repeatability estimates do not always set an upper limit to heritability. *Funct Ecol* 16:273–280.
 44. Falconer DS, Mackay TFC. 1996. Introduction to quantitative genetics, 4th ed. Longmans Green, Harlow, United Kingdom.
 45. Hildebrand F, Nguyen TL, Brinkman B, Yunta RG, Cauwe B, Vandenabeele P, Liston A, Raes J. 2013. Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. *Genome Biol* 14:R4. <https://doi.org/10.1186/gb-2013-14-1-r4>.
 46. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, Costea PI, Godneva A, Kalka IN, Bar N, Shilo S, Lador D, Vila AV, Zmora N, Pevsner-Fischer M, Israeli D, Kosower N, Malka G, Wolf BC, Avnit-Sagi T, Lotan-Pompan M, Weinberger A, Halpern Z, Carmi S, Fu J, Wijmenga C, Zernakova A, Elinav E, Segal E. 2018. Environment dominates over host genetics in shaping human gut microbiota. *Nature* 555:210–215. <https://doi.org/10.1038/nature25973>.
 47. Bledsoe JW, Waldbieser GC, Swanson KS, Peterson BC, Small BC. 2018. Comparison of channel catfish and blue catfish gut microbiota assemblages shows minimal effects of host genetics on microbial structure and inferred function. *Front Microbiol* 9:1073. <https://doi.org/10.3389/fmicb.2018.01073>.
 48. Goodrich JK, Davenport ER, Waters JL, Clark AG, Ley RE. 2016. Cross-species comparisons of host genetic associations with the microbiome. *Science* 352:532–535. <https://doi.org/10.1126/science.aad9379>.
 49. Benson AK, Kelly SA, Legge R, Ma F, Low SJ, Kim J, Zhang M, Oh PL, Nehrenberg D, Hua K, Kachman SD, Moriyama EN, Walter J, Peterson DA, Pomp D. 2010. Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc Natl Acad Sci U S A* 107:18933–18938. <https://doi.org/10.1073/pnas.1007028107>.
 50. Org E, Parks BW, Joo JW, Emert B, Schwartzman W, Kang EY, Mehrabian M, Pan C, Knight R, Gunsalus R, Drake TA, Eskin E, Lusi AJ. 2015. Genetic and environmental control of host-gut microbiota interactions. *Genome Res* 25:1558–1569. <https://doi.org/10.1101/gr.194118.115>.
 51. Kurilshikov A, Wijmenga C, Fu J, Zernakova A. 2017. Host genetics and gut microbiome: challenges and perspectives. *Trends Immunol* 38:633–647. <https://doi.org/10.1016/j.it.2017.06.003>.
 52. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD,

- Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JL, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336. <https://doi.org/10.1038/nmeth.f.303>.
53. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. 2010. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26:266–267. <https://doi.org/10.1093/bioinformatics/btp636>.
54. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
55. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72:5069–5072. <https://doi.org/10.1128/AEM.03006-05>.
56. Faith DP. 1992. Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61:1–10. [https://doi.org/10.1016/0006-3207\(92\)91201-3](https://doi.org/10.1016/0006-3207(92)91201-3).
57. Lozupone C, Knight R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71: 8228–8235. <https://doi.org/10.1128/AEM.71.12.8228-8235.2005>.
58. R Core Team. 2016. R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria.
59. Lessells CM, Boag PT. 1987. Unrepeatable repeatabilities: a common mistake. *The Auk* 104:116–121. <https://doi.org/10.2307/4087240>.
60. Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. *J Stat Soft* 67:1–48. <https://doi.org/10.18637/jss.v067.i01>.
61. Kindt R, Coe R. 2005. Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies. World Agroforestry Centre (ICRAF), Nairobi, Kenya.
62. Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 26:32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>.
63. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2017. vegan: community ecology package, v2.4-4. <https://CRAN.R-project.org/package=vegan>.
64. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B (Methodological)* 57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
65. Stoffel MA, Nakagawa S, Schielzeth H, Goslee S. 2017. rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods Ecol Evol* 8:1639–1644. <https://doi.org/10.1111/2041-210X.12797>.
66. Grothendieck G. 2013. nls2: non-linear regression with brute force, v0.2. <https://CRAN.R-project.org/package=nls2>.
67. Sokal RR, Rohlf FJ. 2011. Biometry + statistical tables. Macmillan Higher Education, London, United Kingdom.
68. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A* 108:4516–4522. <https://doi.org/10.1073/pnas.1000080107>.