AEM
Journals.ASM.org

Evaluating Bias of Illumina-Based Bacterial 16S rRNA Gene Profiles

Katherine Kennedy,[a] Michael W. Hall,[a] Michael D. J. Lynch,[a] Gabriel Moreno-Hagelsieb,[b] Josh D. Neufeld[a]

University of Waterloo, Department of Biology, Waterloo, Ontario, Canada[a]; Wilfrid Laurier University, Department of Biology, Waterloo, Ontario, Canada[b]

Massively parallel sequencing of 16S rRNA genes enables the comparison of terrestrial, aquatic, and host-associated microbial communities with sufficient sequencing depth for robust assessments of both alpha and beta diversity. Establishing standardized protocols for the analysis of microbial communities is dependent on increasing the reproducibility of PCR-based molecular surveys by minimizing sources of methodological bias. In this study, we tested the effects of template concentration, pooling of PCR amplicons, and sample preparation/interlane sequencing on the reproducibility associated with paired-end Illumina sequencing of bacterial 16S rRNA genes. Using DNA extracts from soil and fecal samples as templates, we sequenced pooled amplicons and individual reactions for both high (5- to 10-ng) and low (0.1-ng) template concentrations. In addition, all experimental manipulations were repeated on two separate days and sequenced on two different Illumina MiSeq lanes. Although within-sample sequence profiles were highly consistent, template concentration had a significant impact on sample profile variability for most samples. Pooling of multiple PCR amplicons, sample preparation, and interlane variability did not influence sample sequence data significantly. This systematic analysis underlines the importance of optimizing template concentration in order to minimize variability in microbial-community surveys and indicates that the practice of pooling multiple PCR amplicons prior to sequencing contributes proportionally less to reducing bias in 16S rRNA gene surveys with next-generation sequencing.

PCR represents an essential molecular tool for effective surveys of 16S rRNA genes from a wide range of environments. As with all molecular methods for profiling microbial communities, understanding and mitigating biases associated with PCR amplification are important for improving experimental design and robustness, in addition to establishing standardized methodological protocols. PCR biases can be classified into two general categories: selection, which is the result of inherent differences in amplification efficiencies, and drift, which is the result of stochastic fluctuations and, therefore, nonreproducible. One aspect of selection is the tendency toward a 1:1 ratio of all products due to more abundant templates being less available for amplification because of reannealing (1). Although selection bias caused by differences in primer binding efficiency is not easily managed, reducing the number of amplification cycles was proposed in order to limit the tendency toward this homogeneous product ratio (2–4). Because the effects of drift differ across replicate amplifications, it was theorized that such effects might cancel each other if the reactions were pooled, thereby increasing reproducibility (2, 3).

In addition to selection and drift, massively parallel "next-generation" sequencing can be susceptible to unique sources of representational bias. Specifically, multiplex sequencing allows parallel processing of multiple samples by using unique, sample-specific "barcodes." These barcodes can be added by ligation, but modified barcoded primers are commonly employed because they involve fewer protocol steps and may contribute less bias (5). However, others have tested barcoded primers in triplicate on the same sample from the mouse gut lumen and demonstrated that these primers resulted in less reproducible data sets and that the bias they introduced could not be predicted from the secondary structure of the primer variants (6). Another study demonstrated multiple sources of bias, including template GC content, associated with pyrosequencing-based characterization of defined mixtures of bacterial and archaeal sequence templates (7). Other biases inherent in molecular approaches to community profiling include the DNA extraction protocol (8) and sample storage conditions (9). For example, frozen samples better maintain alpha diversity and differ least in beta diversity (10). Importantly, low template concentrations may be particularly susceptible to bias due to the increased impact of stochastic processes during PCR (11).

Despite knowledge about sources of bias associated with PCR-based methodologies, the effects of amplicon pooling and other potential sources of variability (e.g., primer barcodes and template concentration) have not been assessed rigorously and simultaneously for Illumina-based microbial-community profiling. Here, we used a barcode-based 16S rRNA gene-profiling method (12) to test the influence of template concentration, PCR product pooling, repeating experimental procedures, and processing within multiple Illumina flow cells on the resulting soil and fecal bacterial-community profiles. The results showed a significant reduction in sample profile variability with increased template concentration, whereas amplicon pooling, barcode selection, experimental manipulations, and Illumina flow cells did not contribute significantly to amplification bias, as measured by profile variability.

MATERIALS AND METHODS

Stool samples (S1 and S3) were collected from two healthy adults and frozen at −20°C immediately after collection. These stool samples were

Received 6 May 2014 **Accepted** 1 July 2014
Published ahead of print 7 July 2014
Editor: K. E. Wommack
Address correspondence to Josh D. Neufeld, jneufeld@uwaterloo.ca.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.01451-14.
Copyright © 2014, American Society for Microbiology. All Rights Reserved.
doi:10.1128/AEM.01451-14
The authors have paid a fee to allow immediate free access to this article.

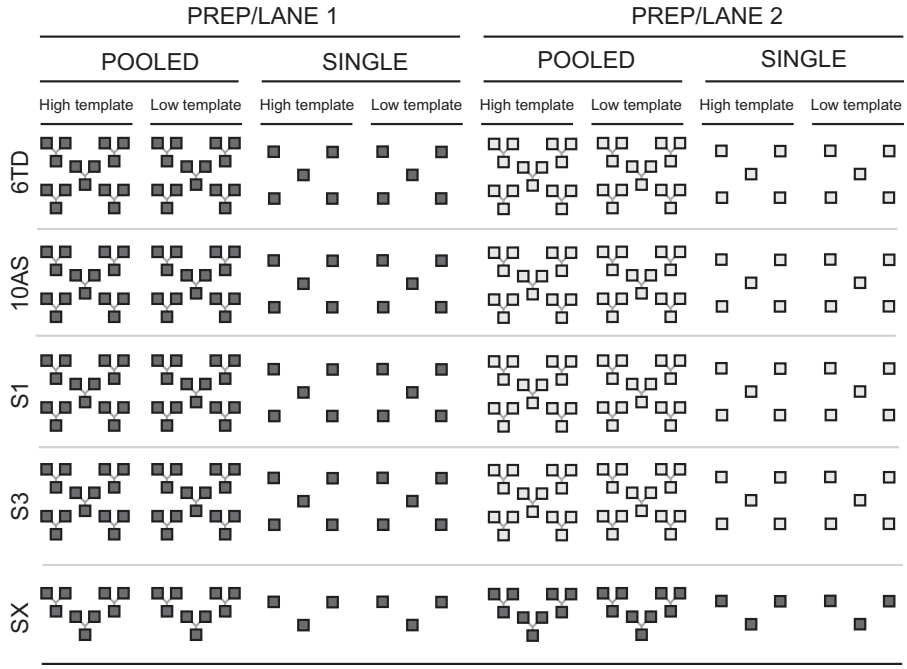

FIG 1 Graphical representation of the experimental design, including soil samples (6TD and 10AS), stool samples (S1, S3, and SX), and treatments. The dark squares represent amplicons prepared on the first day of experimental manipulation (i.e., intended for the first MiSeq lane, except for sample SX), whereas the light squares represent amplicons prepared on a second experimental day (i.e., intended for the second MiSeq lane).

used in a previous study on the bacterial biogeography of the human gastrointestinal tract (13). In addition, two soil samples were selected from the Canadian MetaMicrobiome Library (CM^2BL) for inclusion in this study, including soil from a temperate deciduous forest (6TD) and an agricultural soybean field (10AS). These soil samples represented various land usage histories and chemistry profiles, most notably pH (i.e., pH 6.4 for 6TD and pH 7.6 for 10AS). Additional metadata for the stool and soil samples can be accessed in the corresponding publication (13) and website (http://www.cm2bl.org/), respectively.

DNA extraction. For each sample, DNA was extracted from five 0.25-g aliquots using the PowerSoil DNA Isolation Kit (MoBio) according to the manufacturer's instructions, with minor modifications. Namely, we included an additional 40-s bead-beating step and heating to 70°C for 10 min prior to the contaminant binding step. A final ethanol precipitation was used to concentrate and purify the DNA. Nucleic acids were run on a 1% agarose gel for gel-based quantification, followed by spectrophotometric quantification using a Nanodrop 1000 (Thermo Scientific, USA) and Qubit fluorometric quantification (Life Technologies, USA). Final DNA concentrations were estimated by averaging consistent measurements from each of these parallel quantification efforts.

Each sample was diluted to generate high and low template concentrations. Samples S3, 6TD, and 10AS were diluted to 10 ng/μl and 0.1 ng/μl, and sample S1 was diluted to 5 ng/μl and 0.1 ng/μl, because less DNA was recovered from this stool sample overall. For each of the 8 samples (i.e., S1 high, S1 low, S3 high, S3 low, 6TD high, 6TD low, 10AS high, and 10AS low), 20 PCR amplifications were performed (Fig. 1). Of these, five were not pooled and the remaining samples were pooled in triplicate, resulting in five pooled products and five individual products for each sample. SX replicates are from sample S3 but were separated to assess interlane variability during sequencing.

PCR. Reactions were performed on the same thermocycler in randomized 96-well plates to limit bias due to possible thermal profile variations across wells. The entire procedure was performed on two separate days to assess possible variability introduced by sample manipulation. SX samples were processed only once in order to run identical products in both sequencing lanes.

The combined V3-V4 regions were amplified using modified 341F and 816R primers containing a 6-base barcode, the Illumina adapter sequence, and regions for binding of the sequencing primers (12). The number of cycles was 30: 30 s at 95°C, 30 times 15 s at 95°C followed by 30 s at 50°C, then 30 s at 68°C, and finally 5 min at 68°C. Each reaction was done in a 25-μl volume consisting of the following components: 2.5 μl Thermal Polymerase buffer (10×; NEB), 0.05 μl deoxynucleoside triphosphates (dNTPs; 100 μM), 0.05 μl primer 341F (100 μM), 0.5 μl primer 816R (10 μM), 0.125 μl *Taq* polymerase, 1.5 μl bovine serum albumin (BSA), 1 μl template (at the concentrations described above), and 19.3 μl PCR water.

Illumina library construction. Products for each sample were gel purified to remove primers and primer dimers by separating them on a 1% agarose gel and using a QIAquick gel extraction kit (Qiagen, Mississauga, Ontario, Canada). The products for each sample were mixed in equal nanogram amountsand quantified using a NanoDrop ND2000 spectrophotometer (Thermo Scientific, Wilmington, DE) before being sequenced in two lanes of an Illumina MiSeq at Argonne National Laboratory (Lemont, IL) using paired-end sequencing as previously described (12), followed by paired-end sequence assembly with PANDAseq (18). Analyses of rarefied data sets (4,158 sequences per sample) were performed using QIIME (14) through AXIOME (15), with nonmetric multidimensional scaling (NMDS) being performed for all samples together and both NMDS and PERMDISP (19) (with analysis of variance [ANOVA]) applied to each sample group individually (S1, S3, 6TD, 10AS, and SX).

Nucleotide sequence accession number. All sequence data are available through the European Bioinformatics Institute (EBI) with project accession number PRJEB6592.

RESULTS

Testing DNA extracts from four soil and fecal samples as templates, we performed duplicate experimental manipulations on separate days, each involving the sequencing of pooled and individual reactions for both high (5- to 10-ng) and low (0.1-ng) template concentrations (Fig. 1). Two Illumina MiSeq runs generated 250-base paired-end sequences for each of the duplicate experi-

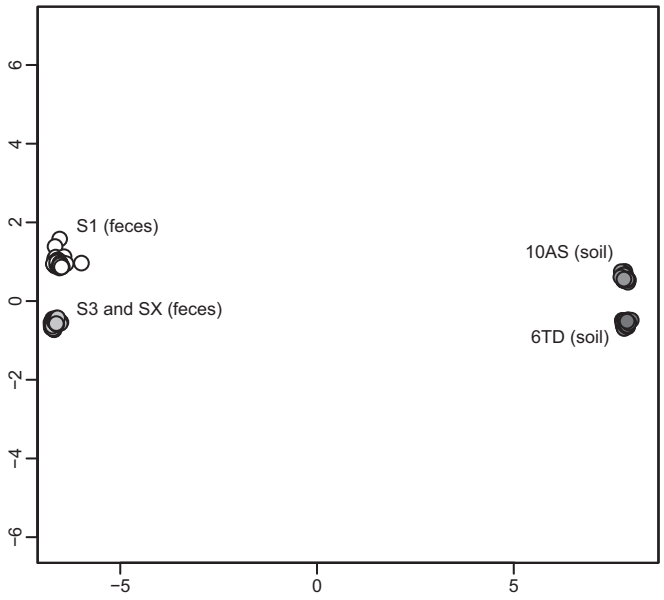

FIG 2 NMDS based on a Bray-Curtis dissimilarity matrix of all sample replicates from Illumina data for S1, S3, 6TD, and 10AS.

mental manipulations. Following PANDAseq assembly at a high-quality assembly threshold, we obtained 3,296,834 assembled amplicon reads for lane 1 and 2,374,250 assembled reads for lane 2 (see Table S1 in the supplemental material). Samples that yielded fewer than 4,000 sequences (i.e., 11 out of 184 experimental samples) were removed from further analysis (see Table S1 in the supplemental material). Because the samples removed were roughly evenly distributed between high and low template concentrations, the sample source (i.e., S1, S3, 6TD, and 10AS), and pooled and individual samples, their removal was not expected to impact downstream analyses. The average number of sequences generated for each sample included in the analysis was 32,931 (range, 4,158 to 143,162).

Based on a Bray-Curtis dissimilarity matrix, an NMDS ordination of rarefied samples demonstrated that all replicates for each sample were highly consistent, separating distinctly by sample (Fig. 2), regardless of whether the samples were sequenced from pooled or from individual amplicons. Within the ordination space, the two stool samples were separated distinctly from the two soil samples, as expected. When sample separation was measured exclusively with pooled or individual samples using the multiresponse permutation procedure (MRPP), group separation (T) was greater for pooled samples ($T = -45.1$) than for individual samples ($T = -41.1$), and within-sample replicate heterogeneity (scatter [A]) was greater for individual sample amplicon data sets ($A = 0.58$, where an A value of 1 would indicate that all sample replicates were identical) than for pooled sample amplicon data sets ($A = 0.64$). This indicates that pooling samples improved the overall within-group homogeneity and slightly increased separation between groups.

Each sample was analyzed independently to better evaluate biases that might not be apparent within an all-sample analysis. Bray-Curtis-based NMDS ordinations were generated for each sample (Fig. 3) to evaluate the effects of the template concentration (i.e., high versus low), pooling (i.e., pooled versus individual), and lane (i.e., lane 1 versus lane 2). The results confirmed that high template concentrations (i.e., 5 to 10 ng per reaction) increased the accuracy of sample operational taxonomic unit (OTU) profiles compared to low template concentrations (i.e., 0.1 ng per reaction), especially for the two soil samples (i.e., 6TD and 10AS). Profile heterogeneity differences were less visually apparent for pooled and individual samples generated on separate days and run on a different MiSeq lane. A sample PCR for each lane was performed on a different day for all samples except the SX samples, which were derived from S3. Only one PCR preparation was performed for SX samples, and these same products were sequenced twice to assess variability associated with separate lanes (Fig. 1 and 3).

Quantitative data were obtained by using a PERMDISP analysis to test whether sequence profile variances were significantly different between groups. Within each group, distances from the centroid of that group to each sample were calculated (Fig. 4). The largest difference in variance was observed for high- and low-template-concentration groups. For all samples, high-template-concentration groups and pooled groups clustered more tightly than low-template-concentration groups and individual groups. The results showed that the effect of the template concentration was both large and significant for both soil samples and significant for samples S1 and SX (Fig. 3; see Table S2 in the supplemental material). Neither the effect of pooling nor that of the lane was significant for any samples. Minimization of sequence profile variance for pooled amplicon samples approached significance for both soil samples, although its effect was much less than for the template concentration.

To test whether pooling was of greater specific benefit for samples with low template concentrations, we performed an MRPP analysis on all samples divided into four groups: all individual high-template-concentration groups ($A = 0.70$; $T = -21.1$), all pooled high-template-concentration groups ($A = 0.75$; $T = -23.1$), all individual low-template-concentration groups ($A = 0.55$; $T = -20.8$), and all pooled low-template-concentration groups ($A = 0.60$; $T = -21.2$). The difference in within-group homogeneity due to pooling did not change between low-template-concentration samples ($\Delta A = 0.05$) and high-template-concentration samples ($\Delta A = 0.05$). However, these differences were less than the difference in within-group homogeneity between high-template-concentration samples and low-concentration samples for either pooled ($\Delta A = 0.15$) or individual ($\Delta A = 0.16$) samples.

DISCUSSION

This study contributes to the understanding of bias in Illumina-based single-gene analysis by testing the overall contributions of several potential sources of experimental variability within a single experimental framework, with the same laboratory handling, samples, primers, and sequencer. Here, we compared sequence profiles from samples that were pooled in triplicate to profiles of individual reactions for high (i.e., 5 to 10 ng/μl) and low (i.e., 0.1 ng/μl) template concentrations, as well as for relatively high (soil) and low (stool) sample alpha diversities. Pooling of triplicate amplifications was first proposed to decrease PCR bias introduced by stochastic fluctuations in amplification efficiencies (drift) (3). However, the impact of pooling has not been demonstrated previously. Ideally, amplifications originating from the same DNA extract should overlap within an ordination. However, both PCR

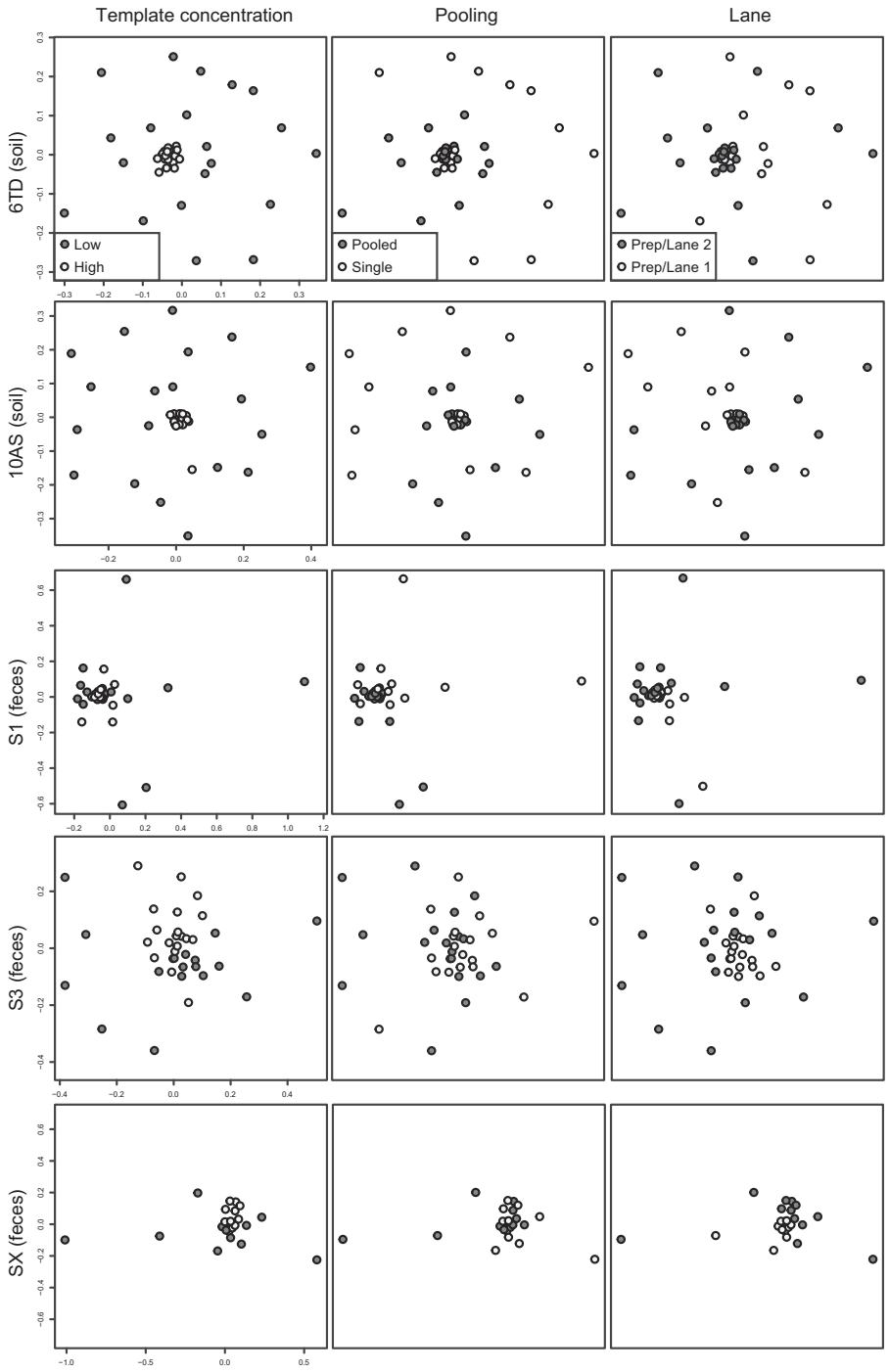

FIG 3 NMDS ordination of replicates for each sample using Bray-Curtis dissimilarity matrices. Open circles represent high-template-concentration, not-pooled, and lane 1 replicates; shaded circles represent low-template-concentration, pooled, and lane 2 replicates.

and sequencing are susceptible to biases that contribute variation among replicate amplifications of the same sample. Techniques that minimize bias result in replicate amplifications clustering more tightly, whereas techniques that increase bias also increase the variance between replicate amplifications.

Although we hypothesized that pooling of replicate PCR amplicons prior to sequencing would help increase sample profile separation and reduce within-sample heterogeneity for an all-sample distance matrix (Fig. 2), template concentration contributed far more to sample profile heterogeneity for the within-sample analysis. The impact of the template concentration was significant for all samples except S3 and was greatest for the high-diversity soil samples. Overall, the impact of the template concentration on profile heterogeneity corroborates previous results suggesting that low template concentrations result in greater stochastic fluctuations in PCR amplifications (3, 11). For the with-

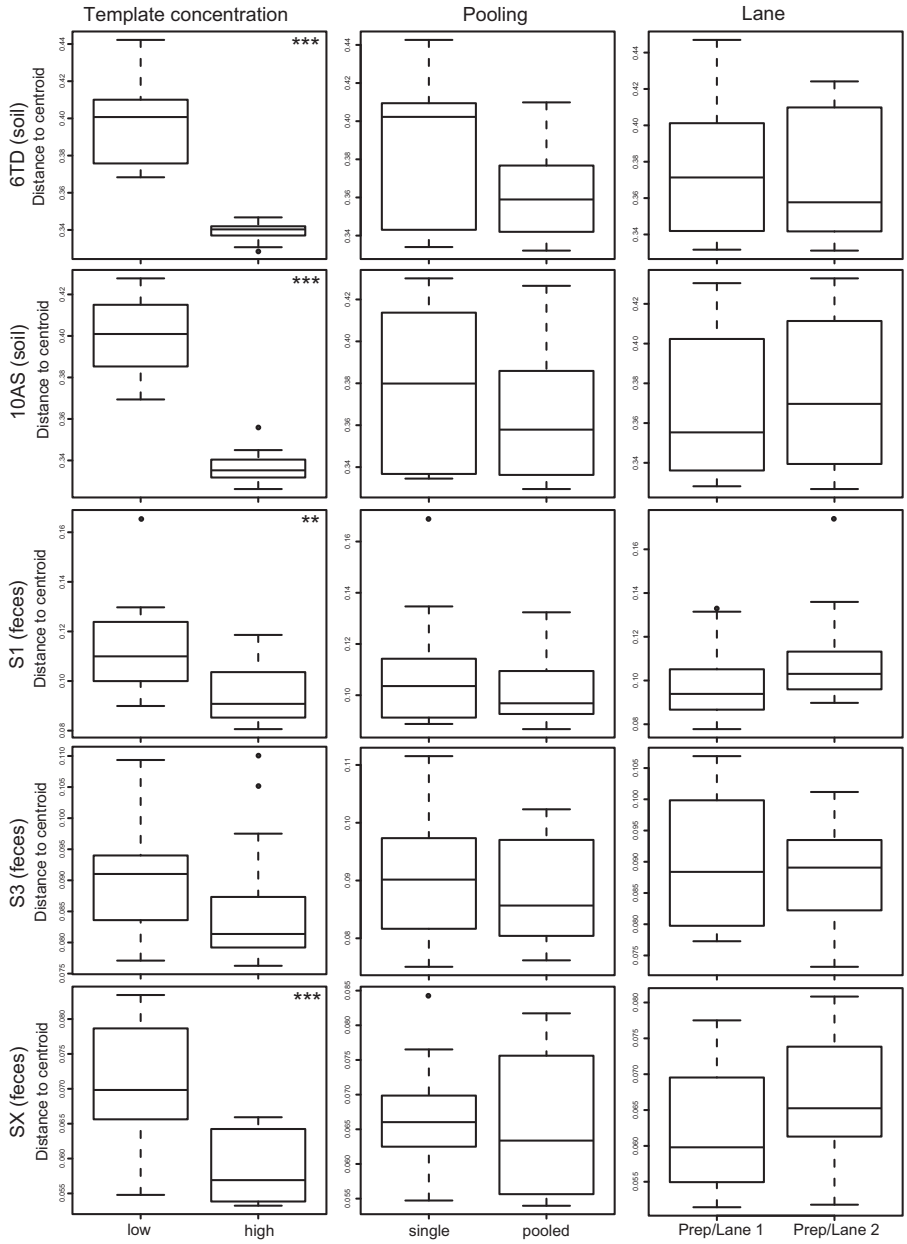

FIG 4 Distance to the centroid of a given sample group when using ordination distances for replicates of each sample. Within PERMDISP, ANOVA was implemented to test the statistical significance of sample distances to the corresponding centroids. The significance of differences is denoted by asterisks (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).

in-sample analysis, amplicon pooling reduced distances to centroids for replicates of all soil and fecal samples, although these differences did not reach significance. The implications are that pooling of triplicate reactions may be particularly helpful when an experimental design requires the separation of highly similar samples or treatment groups, such as stool samples from the same individual over time or for soil samples obtained from multiple plots from the same location. Although one would predict that pooling PCR amplifications might contribute distinct representation from low-abundance "rare biosphere" organisms, we did not detect any significant differences in key alpha diversity metrics (e.g., Shannon, Chao1, and numbers of OTUs) between pooled and individual amplicon data for each sample (data not shown).

The number of PCR amplifications is not strictly limited, so it is possible to both include more samples and pool triplicate amplifications. Indeed, protocols for both the Earth Microbiome Project and the Human Microbiome Project include pooling of triplicate reactions (16, 17). However, the increase in time and effort required to triple the number of amplifications being performed is likely not warranted in cases where samples are very different. Importantly, we demonstrate here that PCR template concentration is an important consideration for reducing sample profile heterogeneity. We did not test a broad range of template concentrations with a large number of samples, partly because our data suggest that optimal template concentrations vary by sample type and would need to be optimized, depending on the require-

ments of individual studies. In addition, this study did not assess the effects of variable cycle numbers on PCR selection and drift, which are considered additional sources of bias (2–4). Nonetheless, our results did show that in all cases, higher template concentrations (i.e., 5 to 10 ng per reaction) resulted in increased amplification reproducibility compared to low template concentrations (i.e., 0.1 ng per reaction), which is likely comparable to the practice of minimizing PCR amplification cycles (2–4). Although many surveys of environmental samples generate low overall nucleic acid recoveries due to limited microbial biomass (e.g., surfaces, skin, and tissues) and would thus prevent using optimal template concentrations, we recommend that consistent and higher overall template concentrations (e.g., 5 to 10 ng) and pooling of low-template-concentration amplification replicates be used whenever possible for amplicon-based rRNA gene surveys of microbial communities.

ACKNOWLEDGMENTS

We acknowledge a social media conversation (http://www.webcitation.org/6QlllAxbS) with the following scientists that led to the conception and design of this study: Kristen DeAngelis, Rachel Poretsky Noah Fierer, Tracy Teal, Mike Cox, Aaron Saunders, Ameet Pinto, and Julia Gustavsen.

This research was supported by the Canadian Institutes for Health Research (CIHR) (G.M.-H. and J.D.N.) and a Discovery Grant to J.D.N. from the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

1. **Suzuki MT, Giovannoni SJ.** 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. **62:**625–630.
2. **Inglis GD, Thomas MC, Thomas DK, Kalmokoff ML, Brooks SP, Selinger LB.** 2012. Molecular methods to measure intestinal bacteria: a review. J. AOAC Int. **95:**5–23. http://dx.doi.org/10.5740/jaoacint.SGE_Inglis.
3. **Polz MF, Cavanaugh CM.** 1998. Bias in template-to-product ratios in multitemplate PCR. Appl. Environ. Microbiol. **64:**3724–3730.
4. **Wagner A, Blackstone N, Cartwright P, Dick M, Misof B, Snow P, Wagner GP, Bartels J, Murtha M, Pendleton J.** 1994. Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift. Syst. Biol. **43:**250–261.
5. **Alon S, Vigneault F, Eminaga S, Christodoulou DC, Seidman JG, Church GM, Eisenberg E.** 2011. Barcoding bias in high-throughput multiplex sequencing of miRNA. Genome Res. **21:**1506–1511. http://dx.doi.org/10.1101/gr.121715.111.
6. **Berry D, Mahfoudh KB, Wagner M, Loy A.** 2011. Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. Appl. Environ. Microbiol. **77:**7846–7849. http://dx.doi.org/10.1128/AEM.05220-11.
7. **Pinto AJ, Raskin L.** 2012. PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. PLoS One **7:**e43093. http://dx.doi.org/10.1371/journal.pone.0043093.
8. **Martin-Laurent F, Philippot L, Hallet S, Chaussod R, Germon JC, Soulas G, Catroux G.** 2001. DNA extraction from soils: old bias for new microbial diversity analysis methods. Appl. Environ. Microbiol. **67:**2354–2359. http://dx.doi.org/10.1128/AEM.67.5.2354-2359.2001.
9. **Cardona S, Eck A, Cassellas M, Gallart M, Alastrue C, Dore J, Azpiroz F, Roca J, Guarner F, Manichanh C.** 2012. Storage conditions of intestinal microbiota matter in metagenomic analysis. BMC Microbiol. **12:**158. http://dx.doi.org/10.1186/1471-2180-12-158.
10. **Rubin BE, Gibbons SM, Kennedy S, Hampton-Marcell J, Owens S, Gilbert JA.** 2013. Investigating the impact of storage conditions on microbial community composition in soil samples. PLoS One **8:**e70460. http://dx.doi.org/10.1371/journal.pone.0070460.
11. **Chandler D, Fredrickson J, Brockman F.** 1997. Effect of PCR template concentration on the composition and distribution of total community 16S rDNA clone libraries. Mol. Ecol. **6:**475–482. http://dx.doi.org/10.1046/j.1365-294X.1997.00205.x.
12. **Bartram AK, Lynch MDJ, Stearns JC, Moreno-Hagelsieb G, Neufeld JD.** 2011. Generation of multi-million 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. Appl. Environ. Microbiol. **77:**3846–3852. http://dx.doi.org/10.1128/AEM.02772-10.
13. **Stearns JC, Lynch MDJ, Senadheera DB, Tenenbaum HC, Goldberg MB, Cvitkovitch DG, Croitoru K, Moreno-Hagelsieb G, Neufeld JD.** 2011. Bacterial biogeography of the human digestive tract. Sci. Rep. **1:**170. http://dx.doi.org/10.1038/srep00170.
14. **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, 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. http://dx.doi.org/10.1038/nmeth.f.303.
15. **Lynch MDJ, Masella AP, Hall MW, Bartram AK, Neufeld JD.** 2013. AXIOME: automated exploration of microbial diversity. GigaScience **2:**3. http://dx.doi.org/10.1186/2047-217X-2-3.
16. **Gilbert JA, Meyer F, Schriml L, Joint I, Mühling M, Field D.** 2010. Metagenomes and metatranscriptomes from the L4 long-term coastal monitoring station in the Western English Channel. Stand. Genomic Sci. **3:**183–193. http://dx.doi.org/10.4056/sigs.1202536.
17. **Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C.** 2009. The NIH human microbiome project. Genome Res. **19:**2317–2323. http://dx.doi.org/10.1101/gr.096651.109.
18. **Masella AP, Bartram AK, Brown DG, Truszkowski JM, Neufeld JD.** 2012. PANDAseq: Paired-eND Assembler for Illumina sequences. BMC Bioinformatics **13:**31. http://dx.doi.org/10.1186/1471-2105-13-31.
19. **Anderson MJ, Ellingsen KE, McArdle BH.** 2006. Multivariate dispersion as a measure of beta diversity. Ecol. Lett. **9:**683–693. http://dx.doi.org/10.1111/j.1461-0248.2006.00926.x.