Assessing the Differences in 16S rRNA Gene Sequencing Due to High Fidelity DNA Polymerase

Influence of DNA Polymerase on PCR bias, chimera formation, and errors when sequencing 16S rRNA genes from complex communities

Marc A Sze[1] and Patrick D Schloss[1 †]

† To whom correspondence should be addressed: pschloss@umich.edu

1 Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI

Co-author e-mails:

- marcsze@med.umich.edu

Abstract

Introduction

16S rRNA gene sequencing is a powerful and widely used tool for surveying the structure of microbial communities (1–3). This approach has exploded in popularity with the advent of high throughput sequencing where it is possible to characterize numerous samples with thousands of sequences per sample. Numerous factors can impact how a natural community is represented by the sequencing data including the method of acquiring samples (4–8), storage conditions (4–6, 9–12), extraction methods (13), amplification conditions (8, 14, 15), sequencing method (15–17), and analytical pipeline (15, 18–20). The increased sampling depth that is now available relative to previous Sanger-based methods is expected to compound the impacts of an investigator's choices and the interpretation of their results.

One step in the generation of 16S rRNA gene sequence data that has been long known to have a significant impact on the description of microbial communities is the choice of conditions for PCR amplification (8, 14, 15). Factors such as the choice of primers have an obvious impact on which populations will be amplified (18, 21). However, a variety of PCR artifacts can also impact the perception of a community including the formation of chimeras (14, 22–24), misincorporation of nucleotides (23, 25, 26), preferential amplification of populations leading to bias (24, 27–33), and accumulation of random amplification events leading to PCR drift (24, 27, 32, 34). Many bioinformatic tools have been developed to identify chimeras, there are significant sensitivity and specificity tradeoffs (14, 35). Others have attempted to account for PCR bias using modeling approaches (29, 36); however, these have been developed for idealized situations. Laboratory-based solutions to minimize chimera formation have also been proposed such as minimizing the amount of template DNA in the PCR, minimizing the number of rounds of PCR, minimizing the amount of shearing in the template DNA, and using DNA polymerases that have a proof-reading ability (14, 23). To minimize PCR drift, some investigators will pool technical replicate PCRs hoping to average out the drift (34). Other factors that have been shown to impact the formation of PCR artifacts are outside the control of an investigator including the fraction of DNA bases that are guanines or cytosines, the variation in the length of the targeted region across the community, the sequence of the DNA that flanks the template, and the genetic diversity of the community (28, 30–33). Early investigations of the factors that lead to the formation of PCR artifacts focused on analyzing binary mixtures of genomic DNA and 16S rRNA gene fragments to explore PCR biases and chimera formation. Although these studies were instrumental in forcing researchers to be cautious about the interpretation of their results, we have a poor understanding of how these factors affect the formation of PCR artifacts in more complex communities.

The influence that the choice of DNA polymerase has on the formation of PCR artifacts has not been well studied. There has been recent interest in how the choice of the hypervariable region and data analysis pipelines impact the sequencing error rate (15, 18–20); however, these studies use the same DNA polymerase in the PCR step and implicitly assume that the rate of nucleotide misincorporation from PCR are significantly smaller than those from the sequencing phase. There has been more limited interest in the impact that DNA polymerase choice has on the formation of chimeras (23, 37). A recent study found differences in the number of OTUs and chimeras between normal and high fidelity DNA polymerases (37). The authors of this study could reduce the difference between two polymerases by optimizing the annealing and extension steps within the PCR protocol (37). Yet this optimization was specific for the community they were analyzing (i.e. captive and semi-captive red-shanked doucs) and assumed that if the two polymerases generate the same community structure that the community structure was correct. In fact, the community structure generated by both methods was not free of artifacts, but has the same artifacts. A challenge in these types of experiments is having a priori knowledge of the true community representation. A mock community with known composition allows researchers to quantify the sequencing error rate, fraction of chimeras, and bias (19); however, mock communities have a limited phylogenetic diversity relative to natural communities. Natural communities, in contrast, have an unknown community composition making absolute measurements impossible. They can be used to validate results from mock communities and to understand the relative impacts of artifacts on the ability to differentiate biological and methodological sources of variation. Given the large number of DNA polymerases available to researchers it is unlikely that a specific recommendation is possible. Rather, the development of general best practices and understanding the impact of PCR artifacts on an analysis are desired.

This study investigated the impact of choice of high fidelity DNA polymerase and the number of rounds of amplification on the formation of PCR artifacts using a mock community and human stool samples. These PCR artifacts included (i) the effect of the polymerase on the error rate of the bases represented in the final sequences, (ii) the fraction of sequences that appeared to be chimeras and the ability to detect those chimeras, (iii) the bias of preferentially amplifying one fragment over another in a mixed pool of templates, and (iv) inter-sample variation in community structure of samples amplified with the same polymerase across the amplification process. To characterize these factors we sequenced a mock community of 8 organisms with known sequences and community structure and human fecal samples with unknown sequences and community structures. We sequenced the V4 region of the 16S rRNA genes from a mock community by generating paired 250 nt reads on the Illumina MiSeq platform. This region and sequencing approach was used because it has been shown to result in a relatively low sequencing error rate and is a widely used protocol. To better understand the impact of DNA polymerase choice on PCR artifacts, we selected five high fidelity DNA polymerases and amplified the communities using 15, 20, 25, 30, and 35 rounds of amplification (Table 1). Collectively, our results suggest that the number of rounds and to a lesser degree the choice of DNA polymerase used in PCR impact the sequence data, the effects are consistent and are smaller than the biological differences between individuals.

Results

Sequencing errors vary by the number of cycles and the DNA polymerase used in PCR. The presence of sequence errors can confound the ability to accurately classify 16S rRNA gene sequences and group sequences into operational taxonomic units (OTUs). More importantly, sequencing errors themselves can alter the representation of the community. Therefore, it is important to minimize the number of sequencing errors. Using a widely-used approach that generates the lowest reported error rate, we quantified the error rate by sequencing the V4 region of the 16S rRNA genes from an 8 member mock community. We also removed any contigs that were at least three bases more similar to a chimera of two references than to a single reference sequence [ref]. Regardless of the polymerase, the error rate increased with the number of rounds of amplification. Using 30 rounds of PCR is a common approach across diverse types of samples. Among the data generated using 30 rounds of PCR the Accuprime polymerase had the highest error rate (i.e. 0.XXX%) followed by the Platinum (i.e. 0.XXX%), Phusion (i.e. 0.XXX%), Kappa (i.e. 0.XXX%), and Q5 (i.e. 0.XXX%) polymerases (Figure 1). When we applied the pre.clustering denoising step, which merges the counts of reads within 2 nt of a more abundant sequence, the error rates dropped considerably such that the Platinum polymerase had the highest error rate (i.e. 0.XXX%) followed by the Accuprime (i.e. 0.XXX%), Phusion (i.e. 0.XXX%), Kappa (i.e. 0.XXX%), and Q5 (i.e. 0.XXX%) polymerases (Figure 1). Although specific recommendations are difficult to make because the impact of the actual community structure and concentration of the initial DNA template are likely to have an impact on the results, it is clear that using as few PCR cycles as necessary and a polymerase with the lowest possible error rate is a good guide to minimizing the impact of polymerase on the error rate.

The fraction of sequences identified as being chimeric varies by the number of cycles and the DNA polymerase used in PCR. Chimeric PCR products can significantly confound downstream analyses. Although numerous bioinformatic tools exist to identify and remove chimeric sequences with high specificity, their sensitivity is relatively low and can be reduced by the presence of sequencing errors. We identified those contigs that were at least three bases more similar to a chimera of two references than to a single reference sequence. As expected from previous studies, the number of chimeras increased with rounds of amplification. Interestingly, the fraction of chimeric sequences from the mock community varied by the type of polymerase used. After 30 rounds of PCR, the Platinum polymerase had the highest chimera rate (i.e. XX%) followed by the Q5 (i.e. XX%), Phusion (i.e. XX%), Kappa (i.e. XX%), and Accuprime (i.e. XX%) polymerases. We used the UChime algorithm to detect chimeras in our pre-clustered mock community data and calculate the algorithm's sensitivity and specificity. For all polymerases, the specificity was above XX% and showed a weak association with the number of cycles used. There was considerable inter-polymerase and inter-round of amplification variation in the sensitivity of UChime to detect the chimeras from the mock community. This suggested that the residual error rate after pre-clustering the sequence data did not compromise the sensitivity. In general, the sensitivity of UChime varied between XX and XX%. The generalizability of these results is limited because we used a single mock community with limited genetic diversity. Although we did not know the true chimera rate for our four human stool samples, we were able to calculate the fraction of sequences that UChime identified as being chimeric. These results followed those from the mock communities; additional rounds of amplification significantly increased the rate of chimeras and there was variation between the polymerases that we used. Although it was not possible to identify the features of a polymerase that resulted in higher rates of chimeras, it is clear that using the smallest number of PCR cycles possible will minimize the impacts of chimeras.

At the sequence level, PCR amplification bias is subtle. Since researchers began using PCR to sequence 16S rRNA genes there has been concern that amplifying fragments from a mixed template pool could lead to a biased representation in the pool of products and would confound downstream analyses. The mock community was generated by mixing equal amounts of genomic DNA from each of the 8 bacteria resulting in uneven representation of the *rrn* operons across the bacteria. The vendor of the mock community subjects each lot of genomic DNA to shotgun sequencing to more accurately quantify the actual abundance of each organism in the community. We compared the expected relative abundance of the 16S rRNA genes from each bacterium in the mock community to the data we generated across rounds of amplification and polymerase. Interestingly, for some bacteria, their representation became less biased with additional rounds of PCR (e.g. *L. fermentum*), while others became more biased (e.g. *E. faecalis*), and others had little change (e.g. *B. subtilis*). The percentage of bases in the V4 region that were guanines or cytosines was not predictive of the amount of bias. Across the strains there was no variation in the length of their V4 regions and they each had the same sequence in the region that the primers annealed. One of the bacteria represented in the mock community, *S. enterica*, had 6 identical copies of the V4 region and 1 copy that differed from those by one nucleotide. The dominant copy had a thymidine and the rare copy had a guanine. We used the sequence data to calculate the ratio of the dominant to rare variants from *S. enterica* (Figure. XXX). The Accuprime, Phusion, Platinum, and Q5 polymerases converged to a ratio of 5.XXX; however, the ratio for the Kappa polymerase varied between 6.XXX and 7.XXX for 25 to 35 rounds of PCR. Given the subtle nature of the variation in the relative abundances of each 16S rRNA gene fragment, it was not possible to create generalizable rules that would explain the bias and the previous observation regarding the guanine and cytosine relative abundance was not observed with our data.

At the community level, the effects of PCR amplification bias grow with additional rounds of PCR. Because the distance between samples and across rounds of PCR could be artificially inflated due to sequencing errors and chimeras, we used analyzed the alpha and beta diversity of the mock community data at different phases of the sequence curation pipeline. First, we removed the chimeras from the mock community data as described above and mapped the individual reads to the OTUs that the 16S rRNA gene fragments would cluster into if there were no sequencing errors. This gave us a community distribution that reflected the distribution following PCR without any artifacts. Alpha diversity metrics such as richness and Shannon and Inverse Simpson diversity metrics increased with the number of rounds of PCR for all polymerases except for the Kappa polymerase, where the richness remained constant but the diversity decreased. These data suggest that PCR had the effect of making the community distribution more even than it was originally. Next, we used the true sequence diversity, but removed chimeras based on mapping reads to all possible chimeras between reference sequences, and clustered the reads to OTUs. The same trends were observed as with perfect sequence data except that the richness and diversity metrics trended higher. Finally, we used the observed sequence data and used the UCHIME algorithm to identify chimeras. Again, we observed the same trends as with the other analyses except the richness and diversity metrics trended higher. These comparisons demonstrated that although the bias at the sequence level was subtle, PCR does have a bias that is exacerbated by errors and chimeras. When we measured the Bray-Curtis distance between the communities observed after 25 rounds of amplification and those at 30 and 35 rounds of amplification the distances between 25 and 35 rounds were higher than between 25 and 30 rounds for each of the polymerases by an average of 0.XX units. The Platinum polymerase varied the most across rounds of amplification (25 vs 30 rounds: 0.XX; 25 vs 35 rounds: 0.XX). Although the distances between samples were small, the ordination of these distances showed a clear change in community structure with increasing rounds of PCR. This observation was supported by our statistical analysis, which revealed that the number of rounds of PCR (R^2=0.XX, P=0.XXXX) was a larger factor than the choice of polymerase (R^2=0.XX, P=0.XXXX). These results demonstrate that subtle differences in relative abundances can have an impact on overall community structure. This variation underscores the importance of only comparing sequence data that have been generated using the same PCR conditions.

The choice of polymerase or the number of rounds of amplification have little impact on the relative interpretation of community-wide metrics of diversity. The biases that we observed at the population and community levels using mock community data appeared to be small relative to the expected differences between biological samples. To study this further, we calculated alpha and beta-diversity metrics using the human stool samples for each of the polymerases and rounds of amplification. We calculated the number of observed OTUs, Shannon diversity, inverse Simpson diversity index for each condition and donor. Although there were clear differences between conditions, the relative ordering of the stool samples did not meaningfully vary across conditions. When we characterized the variation between rounds of amplification using human stool samples, the distance between the 25 and 30 rounds and 25 and 35 rounds varied considerably between samples and polymerases. In general the inter-round variation was lowest for the data generated using the Kappa and Accuprime polymerases. The data generated using the Platinum polymerase was consistent across rounds, but it was more biased than these polymerases. Considering the average distance across the four samples varied between 0.XX and 0.XX, regardless of the polymerases and number of rounds of amplification, any bias due to amplification is unlikely to obscure community-wide differences between samples. In support of this was our principle coordinates analysis of the Bray-Curtis distances, which revealed distinct clusters by donor. Within each cluster there were no obvious patterns related to the polymerase or number of rounds of PCR. However, our statistical analysis using adonis revealed statistically significant differences in the community structures with the subject explaining the most variation (R^2=0.XXX, P=0.XXXX), followed by the number of rounds of PCR (R^2=0.XXX, P=0.XXXX) and the choice of polymerase (R^2=0.XXX, P=0.XXXX). Together, these results indicate that for a coarse analysis of communities, the choice of number of rounds of amplification or polymerase are not important, but that they must be consistent across samples. It is difficult to develop a specific recommendation based on the level of bias across rounds of PCR or polymerases; however, the general suggestion is to use as few rounds of amplification as possible.

There is little evidence of a relationship between polymerase or number of rounds of amplification on PCR drift. There have been concerns that the same template DNA subjected to the same PCR conditions could result in different representations of communities because of random drift. To test this, we determined the average Bray-Curtis distance between replicate reactions using the same polymerase and number of rounds of amplification. Using the mock community data there were no obvious trends. The average Bray-Curtis distance within a set of conditions varied by 0.XX to 0.XX units. Although we did not obtain replicates of each of the stool samples, the intra-sample variation for each set of conditions was consistent and varied between 0.XX and 0.XX units. These data suggest that amplicon sequencing is robust to random variation in amplification and that any differences are likely to be smaller than what is considered biologically relevant.

Discussion

Our results suggest that the number of rounds of PCR and to a lesser degree the choice of DNA polymerase impact the analysis of 16S rRNA gene sequence data from bacterial communities. Although it was not possible to make direct connections between PCR conditions and specific sources of bias, we were able to identify general recommendations that reduce the amount of error, chimera formation, and bias. Researchers should strive to minimize the number of rounds of PCR and should use a high fidelity polymerase. Although specific PCR conditions impact the precise interpretation of the data, the effects were consistent and were smaller than the biological differences between individuals. Based on these observations, within a study, amplicons must be generated by consistent protocols to yield meaningful comparisons. When comparing across studies, values like richness, diversity, and relative abundances must be made in relative and not absolute terms.

The observed sequencing error rates and alpha diversity metrics followed the manufacturers' measurements of their polymerases' fidelity. Accuprime and Platinum have fidelity that are approximately 10-times higher than that of Taq whereas the fidelity of Phusion, Q5, and Kappa are more than 100 times higher. Among these polymerases, the Kappa polymerase resulted in the the lowest error rate, lowest chimera rate, and least bias across rounds of PCR. Considering polymerase development is an active area of commercial development and it likely that new polymerases will come to market, it is important for researchers to understand how changing the polymerase impacts downstream analyses for their type of samples.

Over the past 20 years, a large literature has attempted to document various PCR biases and underscores the fact that data based on amplification of DNA from a mixed community are not a true representation of the actual community. In addition to obvious biases imposed by primer selection, other factors inherent in PCR can influence the representation of communities. Factors that can lead to preferential amplification of one fragment over another have included guanine and cytosine composition, length, flanking DNA composition, amount of DNA shearing, and number of rounds of PCR (24, 27–33). In addition, environmental and reagent contaminants can also have a significant impact on the analysis of low biomass samples (38). Less well understood is the effect of phylogenetic diversity on bias and chimera formation. Communities with low phylogenetic diversity may be more prone to chimera formation since chimeras are more likely to form among closely related sequences (14, 35). The interaction of these various influences on PCR artifacts are complex and difficult to tease apart. Minimizing the level of DNA shearing and using the fewest number of rounds of PCR with a polymerase that has the highest possible fidelity are strategies that can be employed to minimize the formation of chimeras. Although care should always be taken when choosing a polymerase for 16S rRNA gene sequencing, our observations show that the differences between a variety of polymerases are smaller than the actual biological variation in fecal communities between individuals. Therefore, if the biological signal of interest is similar to differences in fecal bacterial communities found between individuals, then the type of high fidelity DNA polymerase used will only minimally change the results.

Even with these strategies it is impossible to remove all PCR artifacts. Beyond the imperfections of the best polymerases, sometimes difficult to lyse organisms require stringent lysis steps and low biomass samples require additional rounds of PCR. A host of bioinformatics tools are available for removing residual sequencing errors (18, 39–41). These tools struggle to correctly differentiate true biological diversity (i.e. Amplicon Sequencing Variants or ASVs) and PCR or sequencing errors. Although these methods remove residual errors, they also risk splitting intragenomic variants into separate ASVs, merging 16S rRNA gene sequences from different taxa into the same ASV. Other tools are available for removing chimeras (14, 35) where there is a trade off between the sensitivity of detecting chimeras and the specificity of correctly calling a sequence a chimera. In recent years, parameters for these algorithms have been changed to increase their sensitivity with little evaluation of the effects on the specificity of the algorithms (39, 41). Others recommend removing any read that has an abundance below a specified threshold (e.g. removing all sequences that only appear once) (20, 39–41). This method must be approached with caution as such approaches are likely to introduce a different bias of the community representation and ignore the fact that artifacts may be quite abundant. Ultimately, researchers must test their hypotheses with multiple methods to validate the claims they reach with any one method (42). All methods have biases and limitations and we must use complementary methods to obtain robust results.

Materials & Methods

Mock community. The ZymoBIOMICS™ Microbial Community DNA Standard (Zymo, CA, USA) was used for mock communities and the bacterial component was made up of *Pseudomonas aeruginosa*, *Escherichia coli*, *Salmonella enterica*, *Lactobacillus fermentum*, *Enterococcus faecalis*, *Staphylococcus aureus*, *Listeria monocytogenes*, and *Bacillus subtilis* at equal genomic DNA abundance (https://web.archive.org/web/20171217151108/http://www.zymoresearch.com:80/microbiomics/microbial-standards/zymobiomics-microbial-community-standards). The actual relative abundance for each bacterium was obtained from Zymo's certificate of analysis for the lot (Lot: ZRC187325), which they determined using shotgun metagenomic sequencing (https://github.com/SchlossLab/Sze_PCRSeqEffects_mSphere_2019/data/references/ZRC187325.pdf).

Human samples. Fecal samples were obtained from 4 individuals who were part of an earlier study (43). These samples were collected using a protocol approved by the University of Michigan Institutional Review Board. For this study, the samples were de-identified. DNA was extracted from the fecal samples using the MOBIO™ PowerMag Microbiome RNA/DNA extraction kit (now Qiagen, MD, USA).

PCR protocol. Five high fidelity DNA polymerases were tested including AccuPrime™ (ThermoFisher, MA, USA), KAPA HIFI (Roche, IN, USA), Phusion (New England Biolabs, MA, USA), Platinum (ThermoFisher, MA, USA), and Q5 (New England Biolabs, MA, USA). Manufacturer recommendations were followed except for the annealing and extension times, which were selected based on previously published protocols (18, 37). Primers targeting the V4 region of the 16S rRNA gene were used with modifications to generate MiSeq amplicon libraries [Reference 18; https://github.com/SchlossLab/MiSeq_WetLab_SOP/]. The number of rounds of PCR used for each sample and polymerase 15 and increased by 5 rounds up to 35 cycles. Insufficient PCR product was generated using 15 rounds and has not been included in our analysis.

Library generation and sequencing. Each PCR condition (i.e. combination of polymerase and number of rounds of PCR) were replicated four times for the mock community and one time for each fecal sample. Libraries were generated as described previously [Reference 18; https://github.com/SchlossLab/MiSeq_WetLab_SOP/]. The libraries were sequenced using the Illumina MiSeq sequencing platform to generate paired 250-nt reads.

Sequence processing. The mothur software program (v 1.41) was used for all sequence processing steps (44). The protocol has been previously published (18) (https://www.mothur.org/wiki/MiSeq_SOP). Briefly, paired reads were assembled using mothur's make.contigs command to correct errors introduced by sequencing (18). Any assembled contigs that contained an ambiguous base call, mapped to the incorrect region of the 16S rRNA gene, or appeared to be a contaminant were removed from subsequent analyses. Sequences were further denoised using mothur's pre.cluster command to merge the counts of sequences that were within 2 nt of a more abundant sequence. The VSEARCH implementation of UCHIME was used to screen for chimeras (35, 45). At various stages in the sequence processing pipeline for the mock community data, the mothur seq.error command was used to quantify the sequencing error rate as well as the true chimera rate. This command uses the true sequences from the mock community to generate all possible chimeras and removes any contigs that were at least three bases more similar to a chimera than to a reference sequence. The command then counts the number of substitutions, insertions, and deletions in the contig relative to the reference sequence and reports the error rate without the inclusion of chimeric sequences (19). The reference sequences and operon copy number for each bacterium were obtained from the ZymoBIOMICS™ Microbial Community DNA Standard protocol (https://web.archive.org/web/20181221151905/https://www.zymoresearch.com/media/amasty/amfile/attach/_D6305_D6306_ZymoBIOMICS_Microbial_Community_DNA_Standard_v1.1.3.pdf). Sequences were assigned to operational taxonomic units (OTUs) at a threshold of 3% dissimilarity using the OptiClust algorithm (46). To adjust for unequal sequencing when measuring alpha and beta diversity, all samples were rarefied to 1,000 sequences 1,000 times for downstream analysis. The Good's coverage for the samples at this level of sampling was routinely greater than 95%.

Statistical analysis. All analysis was done with the R (v 3.5.0) software package (47). Data transformation and graphing were completed using the tidyverse package (v 1.2.1) (48). The distance matrix data was analyzed using the adonis function within the vegan package (v 2.5.3) (49).

Reproducible methods. The data analysis code for this study can be found at https://github.com/SchlossLab/Sze_PCRSeqEffects_mSphere_2019. The raw sequences are available at the SRA (Accession SRP132931).

Acknowledgements

The authors thank the study participants in ERIN whose samples were utilized. We also would like to thank Judy Opp and April Cockburn for their effort in sequencing the samples as part of the Microbiome Core Facility at the University of Michigan. Additional thanks to members of the Schloss lab and Dr. Marcy Balunas for reading earlier drafts of the manuscript and providing helpful critiques. Salary support for Marc A. Sze came from the Canadian Institute of Health Research and NIH grant UL1TR002240. Salary support for Patrick D. Schloss came from NIH grants P30DK034933 and 1R01CA215574.

References

Figure 1. The error rate of assembled sequence reads increases with the number of rounds of PCR used and follows the relative error rates provided by the manufacturers; however, much of this error is mediated by denoising.

Figure 2. The fraction of all denoised sequences that were identified as being chimeric increases with the number of rounds of PCR used and varied between polymerases. (A) Sequencing of a mock community allowed us to identify the total fraction of sequences that were chimeric as well as the sensitivity and specificity of UCHIME to detect those chimeras. (B) Sequencing of four human stool samples after using one of five different polymerases again demonstrated increased rate of chimera formation with increasing number of rounds of PCR and variation across polymerases.

Figure 3. The relative abundances of reads mapped to reference sequences differed subtly from the expected relative abundances as determined by shotgun metagenomic sequencing. Bias did not increase with number of rounds of PCR or vary by polymerase or the guanine and cytosine content of the fragment. The expected relative abundance of each organism is indicated by the horizontal gray line. The percentage of bases that were guanines or cytosines within the V4 region of the 16S rRNA genes in each organism is indicated by the number in the lower left corner of each panel.

Figure 4. Despite evidence of subtle PCR bias at the genome level, there was significant evidence of bias using community-wide metrics that grew with the number of rounds of PCR when using a mock community. (A) With the exception of data collected using the Kappa polymerase, the OTU richness and Shannon diversity values increased with number of rounds of PCR and the inclusion of residual sequencing errors and chimeras. The horizontal black line indicates the expected richness and diversity if there were no bias, errors, or chimeras. (B) Relative to the mock community sampled after 25 rounds of PCR, the Bray-Curtis distance to the communities sampled after 30 and 35 rounds of PCR increased for all polymerases. (C) The variation between samples collected after 20, 25, 30, and 35 rounds of PCR for the five polymerases demonstrated a significant change in the community driven by the number of rounds of PCR and the polymerase used.

Figure 5. Sequencing of human stool samples indicated clear increase in bias with number of rounds of PCR, however, the bias appeared to be consistent within each sample. (A) With the exception of data collected using the Kappa polymerase, the OTU richness and Shannon diversity values increased with number of rounds of PCR. (B) Relative to the stool communities sampled after 25 rounds of PCR, the Bray-Curtis distance to the stool communities sampled after 30 and 35 rounds of PCR was inconsistent and there was little difference in variation for data collected using the Kappa polymerase. (C)

The variation between stool samples was larger than the amount of variation introduced by varying the number of rounds of PCR or polymerase.

Figure 6. There is little evidence for PCR drift adversely impacting the results of amplicon studies. The average distance between replicates of sequencing the same mock community or between the the human stool samples did not vary by number of rounds of PCR or by polymerase.

Figure S1: With the exception of the seuqence data generated using the Kappa polymerase, the ratio of the two *Salmonella enterica* V4 sequences to a value was lower than the expected ratio of 6:1. The two *S. enterica* V4 sequences differed by a single base.

1. **Gilbert JA**, **Jansson JK**, **Knight R**. 2018. Earth microbiome project and global systems biology. mSystems **3**. doi:10.1128/msystems.00217-17.

2. **Consortium HM**. 2012. Structure, function and diversity of the healthy human microbiome. Nature **486**:207–214. doi:10.1038/nature11234.

3. **Schloss PD**, **Girard RA**, **Martin T**, **Edwards J**, **Thrash JC**. 2016. Status of the archaeal and bacterial census: an update. mBio **7**. doi:10.1128/mbio.00201-16.

4. **Luo T**, **Srinivasan U**, **Ramadugu K**, **Shedden KA**, **Neiswanger K**, **Trumble E**, **Li JJ**, **McNeil DW**, **Crout RJ**, **Weyant RJ**, **Marazita ML**, **Foxman B**. 2016. Effects of specimen collection methodologies and storage conditions on the short-term stability of oral microbiome taxonomy. Applied and Environmental Microbiology **82**:5519–5529. doi:10.1128/aem.01132-16.

5. **Bassis CM**, **Nicholas M. Moore**, **Lolans K**, **Seekatz AM**, **Weinstein RA**, **Young VB**, **Hayden MK**. 2017. Comparison of stool versus rectal swab samples and storage conditions on bacterial community profiles. BMC Microbiology **17**. doi:10.1186/s12866-017-0983-9.

6. **Gorzelak MA**, **Gill SK**, **Tasnim N**, **Ahmadi-Vand Z**, **Jay M**, **Gibson DL**. 2015. Methods for improving human gut microbiome data by reducing variability through sample processing and storage of stool. PLOS ONE **10**:e0134802. doi:10.1371/journal.pone.0134802.

7. **Dominianni C**, **Wu J**, **Hayes RB**, **Ahn J**. 2014. Comparison of methods for fecal microbiome biospecimen collection. BMC Microbiology **14**:103. doi:10.1186/1471-2180-14-103.

8. **Santos QMB-d los**, **Schroeder JL**, **Blakemore O**, **Moses J**, **Haffey M**, **Sloan W**, **Pinto AJ**. 2016. The impact of sampling, PCR, and sequencing replication on discerning changes in drinking water bacterial community over diurnal time-scales. Water Research **90**:216–224. doi:10.1016/j.watres.2015.12.010.

9. **Sinha R**, **Chen J**, **Amir A**, **Vogtmann E**, **Shi J**, **Inman KS**, **Flores R**, **Sampson J**, **Knight R**, **Chia N**. 2015. Collecting fecal samples for microbiome analyses in epidemiology studies. Cancer Epidemiology Biomarkers & Prevention **25**:407–416. doi:10.1158/1055-9965.epi-15-0951.

10. **Amir A**, **McDonald D**, **Navas-Molina JA**, **Debelius J**, **Morton JT**, **Hyde E**, **Robbins-Pianka A**, **Knight R**. 2017. Correcting for microbial blooms in fecal samples during room-temperature shipping. mSystems **2**:e00199–16. doi:10.1128/msystems.00199-16.

11. **Lauber CL**, **Zhou N**, **Gordon JI**, **Knight R**, **Fierer N**. 2010. Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. FEMS Microbiology Letters **307**:80–86. doi:10.1111/j.1574-6968.2010.01965.x.

12. **Song SJ**, **Amir A**, **Metcalf JL**, **Amato KR**, **Xu ZZ**, **Humphrey G**, **Knight R**. 2016. Preservation methods differ in fecal microbiome stability, affecting suitability for field studies. mSystems **1**:e00021–16. doi:10.1128/msystems.00021-16.

13. **Costea PI**, **Zeller G**, **Sunagawa S**, **Pelletier E**, **Alberti A**, **Levenez F**, **Tramontano M**, **Driessen M**, **Hercog R**, **Jung F-E**, **Kultima JR**, **Hayward MR**, **Coelho LP**, **Allen-Vercoe E**, **Bertrand L**, **Blaut M**, **Brown JRM**, **Carton T**, **Cools-Portier S**, **Daigneault M**, **Derrien M**, **Druesne A**, **Vos WM de**, **Finlay BB**, **Flint HJ**, **Guarner F**, **Hattori M**, **Heilig H**, **Luna RA**, **Hylckama Vlieg J van**, **Junick J**, **Klymiuk I**, **Langella P**, **Chatelier EL**, **Mai V**, **Manichanh C**, **Martin JC**, **Mery C**, **Morita H**, **O'Toole PW**, **Orvain C**, **Patil KR**, **Penders J**, **Persson S**, **Pons N**, **Popova M**, **Salonen A**, **Saulnier D**, **Scott KP**, **Singh B**, **Slezak K**, **Veiga P**, **Versalovic J**, **Zhao L**, **Zoetendal EG**, **Ehrlich SD**, **Dore J**, **Bork P**. 2017. Towards standards for human fecal sample processing in metagenomic studies. Nature Biotechnology. doi:10.1038/nbt.3960.

14. **Haas BJ**, **Gevers D**, **Earl AM**, **Feldgarden M**, **Ward DV**, **Giannoukos G**, **Ciulla D**, **Tabbaa D**, **Highlander SK**, **Sodergren E**, **Methe B**, **DeSantis TZ**, **Petrosino JF**, **Knight R**, **and BWB**. 2011. Chimeric 16S rRNA sequence formation and detection in sanger and 454-pyrosequenced PCR amplicons. Genome Research **21**:494–504. doi:10.1101/gr.112730.110.

15. **Sinha R**, **Abu-Ali G**, **Vogtmann E**, **Fodor AA**, **Ren B**, **Amir A**, **Schwager E**, **Crabtree J**, **Ma S**, **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. Nature Biotechnology. doi:10.1038/nbt.3981.

16. **Meisel JS**, **Hannigan GD**, **Tyldsley AS**, **SanMiguel AJ**, **Hodkinson BP**, **Zheng Q**, **Grice EA**. 2016. Skin microbiome surveys are strongly influenced by experimental design. Journal of Investigative Dermatology **136**:947–956. doi:10.1016/j.jid.2016.01.016.

17. **Caporaso JG**, **Lauber CL**, **Walters WA**, **Berg-Lyons D**, **Lozupone CA**, **Turnbaugh PJ**, **Fierer N**, **Knight R**. 2010. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences **108**:4516–4522. doi:10.1073/pnas.1000080107.

18. **Kozich JJ**, **Westcott SL**, **Baxter NT**, **Highlander SK**, **Schloss PD**. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq illumina sequencing platform. Applied and Environmental Microbiology **79**:5112–5120. doi:10.1128/aem.01043-13.

19. **Schloss PD**, **Gevers D**, **Westcott SL**. 2011. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS ONE **6**:e27310. doi:10.1371/journal.pone.0027310.

20. **Bokulich NA**, **Subramanian S**, **Faith JJ**, **Gevers D**, **Gordon JI**, **Knight R**, **Mills DA**, **Caporaso JG**. 2012. Quality-filtering vastly improves diversity estimates from illumina amplicon sequencing. Nature Methods **10**:57–59. doi:10.1038/nmeth.2276.

21. **Parada AE**, **Needham DM**, **Fuhrman JA**. 2015. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environmental Microbiology **18**:1403–1414. doi:10.1111/1462-2920.13023.

22. **Wang GCY**, **Wang Y**. 1996. The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species. Microbiology **142**:1107–1114. doi:10.1099/13500872-142-5-1107.

23. **Potapov V**, **Ong JL**. 2017. Examining sources of error in PCR by single-molecule sequencing. PLOS ONE **12**:e0169774. doi:10.1371/journal.pone.0169774.

24. **Kebschull JM**, **Zador AM**. 2015. Sources of PCR-induced distortions in high-throughput sequencing data sets. Nucleic Acids Research gkv717. doi:10.1093/nar/gkv717.

25. **McInerney P**, **Adams P**, **Hadi MZ**. 2014. Error rate comparison during polymerase chain reaction by DNA polymerase. Molecular Biology International **2014**:1–8. doi:10.1155/2014/287430.

26. **Cline J**. 1996. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. Nucleic Acids Research **24**:3546–3551. doi:10.1093/nar/24.18.3546.

27. **Acinas SG**, **Sarma-Rupavtarm R**, **Klepac-Ceraj V**, **Polz MF**. 2005. PCR-induced sequence artifacts and bias: Insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Applied and Environmental Microbiology **71**:8966–8969. doi:10.1128/aem.71.12.8966-8969.2005.

28. **Polz MF**, **Cavanaugh CM**. 1998. Bias in template-to-product ratios in multitemplate PCR. Applied and Environmental Microbiology **64**:3724–3730.

29. **Brooks JP**, **David J Edwards**, **Harwich MD**, **Rivera MC**, **Fettweis JM**, **Serrano MG**, **Reris RA**, **Sheth NU**, **Huang B**, **Girerd P**, **Strauss JF**, **Jefferson KK**, **Buck GA**. 2015. The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. BMC Microbiology **15**. doi:10.1186/s12866-015-0351-6.

30. **Suzuki MT**, **Giovannoni SJ**. 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by pCR. Applied and environmental microbiology **62**:625–630.

31. **Chandler D**, **Fredrickson J**, **Brockman F**. 1997. Effect of pCR template concentration on the composition and distribution of total community 16S rDNA clone libraries. Molecular Ecology **6**:475–482.

32. **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. Systematic Biology **43**:250–261.

33. **Hansen MC**, **Tolker-Nielsen T**, **Givskov M**, **Molin S**. 1998. Biased 16S rDNA pCR amplification caused by interference from dNA flanking the template region. FEMS Microbiology Ecology **26**:141–149.

34. **Kennedy K**, **Hall MW**, **Lynch MDJ**, **Moreno-Hagelsieb G**, **Neufeld JD**. 2014. Evaluating bias of illumina-based bacterial 16S rRNA gene profiles. Applied and Environmental Microbiology **80**:5717–5722. doi:10.1128/aem.01451-14.

35. **Edgar RC**, **Haas BJ**, **Clemente JC**, **Quince C**, **Knight R**. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics **27**:2194–2200. doi:10.1093/bioinformatics/btr381.

36. **Edgar RC**. 2017. UNBIAS: An attempt to correct abundance bias in 16S sequencing, with limited success. doi:10.1101/124149.

37. **Gohl DM**, **Vangay P**, **Garbe J**, **MacLean A**, **Hauge A**, **Becker A**, **Gould TJ**, **Clayton JB**, **Johnson TJ**, **Hunter R**, **Knights D**, **Beckman KB**. 2016. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. Nature Biotechnology **34**:942–949. doi:10.1038/nbt.3601.

38. **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 Biology **12**. doi:10.1186/s12915-014-0087-z.

39. **Callahan BJ**, **McMurdie PJ**, **Rosen MJ**, **Han AW**, **Johnson AJA**, **Holmes SP**. 2016. DADA2: High-resolution sample inference from illumina amplicon data. Nature Methods **13**:581–583.

doi:10.1038/nmeth.3869.

40. **Edgar RC**. 2016. UNOISE2: improved error-correction for illumina 16S and ITS amplicon sequencing. doi:10.1101/081257.

41. **Amir A**, **McDonald D**, **Navas-Molina JA**, **Kopylova E**, **Morton JT**, **Xu ZZ**, **Kightley EP**, **Thompson LR**, **Hyde ER**, **Gonzalez A**, **Knight R**. 2017. Deblur rapidly resolves single-nucleotide community sequence patterns. mSystems **2**. doi:10.1128/msystems.00191-16.

42. **Schloss PD**. 2018. Identifying and overcoming threats to reproducibility, replicability, robustness, and generalizability in microbiome research. mBio **9**. doi:10.1128/mbio.00525-18.

43. **Seekatz AM**, **Rao K**, **Santhosh K**, **Young VB**. 2016. Dynamics of the fecal microbiome in patients with recurrent and nonrecurrent clostridium difficile infection. Genome Medicine **8**. doi:10.1186/s13073-016-0298-8.

44. **Schloss PD**, **Westcott SL**, **Ryabin T**, **Hall JR**, **Hartmann M**, **Hollister EB**, **Lesniewski RA**, **Oakley BB**, **Parks DH**, **Robinson CJ**, **Sahl JW**, **Stres B**, **Thallinger GG**, **Horn DJV**, **Weber CF**. 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology **75**:7537–7541. doi:10.1128/aem.01541-09.

45. **Rognes T**, **Flouri T**, **Nichols B**, **Quince C**, **Mahé F**. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ **4**:e2584. doi:10.7717/peerj.2584.

46. **Westcott SL**, **Schloss PD**. 2017. OptiClust, an improved method for assigning amplicon-based sequence data to operational taxonomic units. mSphere **2**:e00073–17. doi:10.1128/mspheredirect.00073-17.

47. **R Core Team**. 2018. R: A language and environment for statistical computing. R Foundation for Statistical ComputingVienna, Austria.

48. **Wickham H**. 2017. tidyverse: Easily install and load the 'tidyverse'.

49. **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**. 2018. vegan: Community ecology package.