

Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies

Daryl M Gohl¹, Pajau Vangay², John Garbe³, Allison MacLean¹, Adam Hauge^{1,9}, Aaron Becker¹, Trevor J Gould⁴, Jonathan B Clayton⁵, Timothy J Johnson⁵, Ryan Hunter⁶, Dan Knights^{7,8} & Kenneth B Beckman¹

Amplicon-based marker gene surveys form the basis of most microbiome and other microbial community studies. Such PCR-based methods have multiple steps, each of which is susceptible to error and bias. Variance in results has also arisen through the use of multiple methods of next-generation sequencing (NGS) amplicon library preparation. Here we formally characterized errors and biases by comparing different methods of amplicon-based NGS library preparation. Using mock community standards, we analyzed the amplification process to reveal insights into sources of experimental error and bias in amplicon-based microbial community and microbiome experiments. We present a method that improves on the current best practices and enables the detection of taxonomic groups that often go undetected with existing methods.

The ability to profile microbial communities using NGS has sparked enormous interest in studying the microbiome^{1,2}. Pioneering work has established experimental methods and analytical frameworks for studying the microbiome, including marker gene surveys, in which a portion of a conserved sequence such as the 16S ribosomal RNA (rRNA) gene is amplified, sequenced, and used to quantify the organisms or operational taxonomic units (OTUs) that make up a microbial community^{2–8}.

It is well known that PCR-based marker gene surveys are prone to biases and errors that have resulted in concerns about accuracy, reproducibility, and contamination in microbiome studies^{9–12}. Documented sources of error and bias include undersampling¹³, differential extraction efficiency^{14–16}, amplification of contaminating DNA from reagents⁹, sample storage conditions^{17,18}, amplification parameters (e.g., enzyme, annealing temperature, time, ramp rates,

and cycle number)^{19–23}, starting template concentration^{21,24}, template properties (e.g., GC content and secondary structure)^{11,25,26}, primer mismatches or degeneracies^{27–30}, polymerase error^{23,31–34}, chimeric reads^{23,35}, random errors³⁶, and systematic PCR errors^{28,37}. Finally, sequencing also has an error rate, and the imaging-based nature of Illumina sequencing in particular means that the sequencing of amplicons is particularly challenging^{23,38}.

The recognition of these errors, biases, and limitations has led to a proliferation of methods for amplicon library preparation. These methods differ in whether the sequencing adaptors are added by ligation^{39–42}, single-step PCR^{31,43–46}, or multi-step PCR^{47–49}; whether indices for sample multiplexing are read in-line^{39–42,44} or using separate index reads^{31,43,45–49}; whether single^{39,43–46,49} or dual^{31,40,41,47,48} indices are used; and whether standard adaptor-targeting^{39–42,44,45,47–49} or custom amplicon-targeting^{31,43,46} sequencing primers are used. In addition, some methods use frameshifting bases to boost sequence quality^{41,44,48,49} or unique molecular identifiers to correct sequencing errors within consensus families^{48,49}. Although attempts have been made to standardize these methods^{2,43} and to examine the effects of amplification conditions on bias^{10,11,23,24,33,50}, the parameter space of these different protocols has not been systematically explored⁵.

Here we compare several methods for amplicon-based library preparation and explore their main parameters. We examine the accuracy of different methods and dissect the amplification process to provide insight into the sources of experimental error and bias. We show that the signatures of sources of error and bias were evident in nonhuman primate (NHP) and human-associated microbial samples that we sequenced using different methods, and that errors of the size we detected are likely to substantially affect experimental conclusions. Finally, we suggest best practices for conducting amplicon-based marker gene surveys.

RESULTS

We set out to establish a robust, high-throughput amplicon-based microbiome profiling protocol within a university core facility (the University of Minnesota Genomics Center). Initially, we implemented two protocols: the Earth Microbiome Project (EMP) protocol, a widely adopted standardized method⁴³, and a dual-indexing (DI) protocol similar to that described in an Illumina technical note⁴⁷.

The EMP protocol targets variable region 4 of the 16S rRNA gene, amplifying for 35 PCR cycles with Taq polymerase. Functionalities required for sequencing (flow cell adaptors and indices) are added directly to amplification primers; after PCR, samples are normalized,

¹University of Minnesota Genomics Center, Minneapolis, Minnesota, USA.

²Biomedical Informatics and Computational Biology, University of Minnesota, Minneapolis, Minnesota, USA. ³Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota, USA. ⁴University of Minnesota Informatics Institute, Minneapolis, Minnesota, USA. ⁵Department of Veterinary and Biomedical Sciences, University of Minnesota, St. Paul, Minnesota, USA. ⁶Department of Microbiology and Immunology, University of Minnesota, Minneapolis, Minnesota, USA. ⁷Biotechnology Institute, University of Minnesota, St. Paul, Minnesota, USA. ⁸Department of Computer Science and Engineering, University of Minnesota, Minneapolis, Minnesota, USA. ⁹Present address: Illumina, San Diego, California, USA. Correspondence should be addressed to D.M.G. (dmgohl@umn.edu).

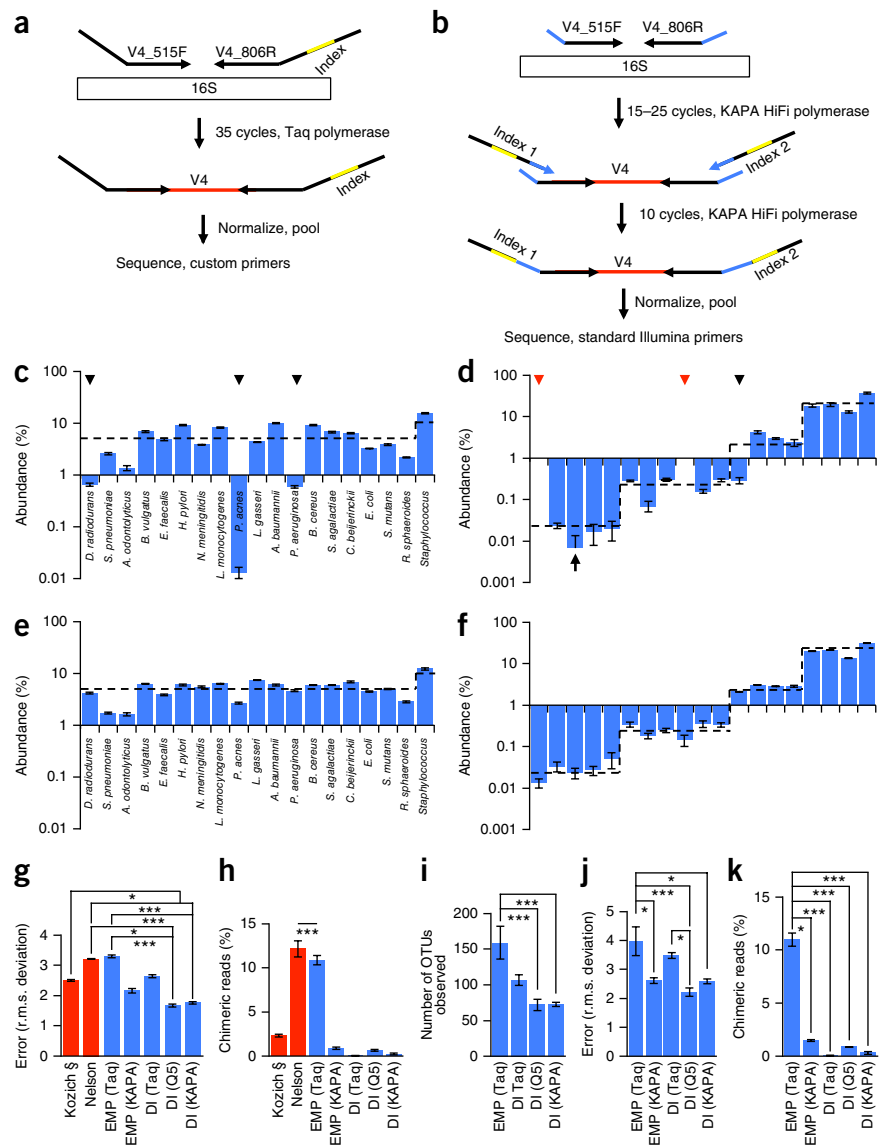
Received 9 November 2015; accepted 11 May 2016; published online 25 July 2016; doi:10.1038/nbt.3601

Figure 1 Protocols for 16S rRNA gene microbiome profiling and the effect of method and enzyme choice on the accuracy of microbiome profiling. **(a)** Outline of the EMP protocol, which uses a single round of amplification with primers targeting the V4 variable region of the 16S rRNA gene and containing Illumina flow cell adaptors and a single index sequence. **(b)** The DI protocol used in this study, in which the V4 variable region of the 16S rRNA gene is amplified with primers containing common adaptor sequences, and then the Illumina flow cell adaptors and dual indices are added in a secondary amplification.

(c) HM-276D even mock community mean abundance data measured using the EMP (Taq) protocol. Arrowheads indicate that the observed abundance deviated by more than fivefold from the expected value. Note that the V4 region did not distinguish the two species of *Staphylococcus* in the mock community, so the expected abundance of that genus is 10% (in **c–f**, expected abundance is indicated by the dashed line). **(d)** HM-277D staggered mock community mean abundance data measured using the EMP (Taq) protocol.

Black arrowheads indicate that the observed abundance deviated by more than fivefold from the expected value; red arrowheads indicate taxa that had no mapped reads (dropouts) and thus could not be plotted on a log scale. The arrow indicates an error bar with a lower bound of zero that could not be plotted on a log scale. **(e)** HM-276D even mock community mean abundance data measured using the DI (KAPA) protocol. **(f)** HM-277D staggered mock community mean abundance data measured using the DI (KAPA) protocol.

(g) Mean r.m.s. deviation for the HM-276D even mock community data measured using the indicated methods. § HM-278D expected abundance values were used to calculate the r.m.s. deviation for this data set. **(h)** Mean percentage of chimeric reads observed with the indicated methods. **(i)** Average number of OTUs observed with the indicated methods. **(j)** Mean r.m.s. deviation for the HM-277D staggered mock community data measured using the indicated methods. **(k)** Mean percentage of chimeric reads observed with the indicated methods. Error bars are \pm s.e.m.; $n = 3$ (**c,d**) or 4 (**e,f**). (**g,h**), $n = 12, 2, 3, 3, 4, 4, 4$; (**i**), $n = 3, 4, 4, 4, 4$; (**j,k**), $n = 3, 3, 4, 4, 4, 4$. * $P < 0.05$, *** $P < 0.01$, determined by analysis of variance with Tukey's honest significant difference *post hoc* test.



pooled, and sequenced using custom sequencing primers that overlap the amplification primers (Fig. 1a). The DI protocol uses a single pair of PCR primers with 5' adaptor tails to amplify samples in a 'primary' amplification. A 'secondary' PCR adds flow cell adaptors and indices (Fig. 1b). A two-step amplification is advantageous because it avoids potential bias due to the use of different indices in each primary amplification, and it offers increased versatility by minimizing the number of primers that must be synthesized.

Initially, we tested two optimizations that should improve accuracy. First, we used a proofreading polymerase optimized for fidelity, processivity, and low GC bias (KAPA HiFi polymerase)³². Second, we optimized the template concentration to minimize the number of PCR cycles. To assess accuracy, we sequenced two mock microbial communities from the Human Microbiome Project Consortium: an even mock community (HM-276D, with 20 bacterial species at equimolar rRNA operon counts) and a staggered mock community (HM-277D,

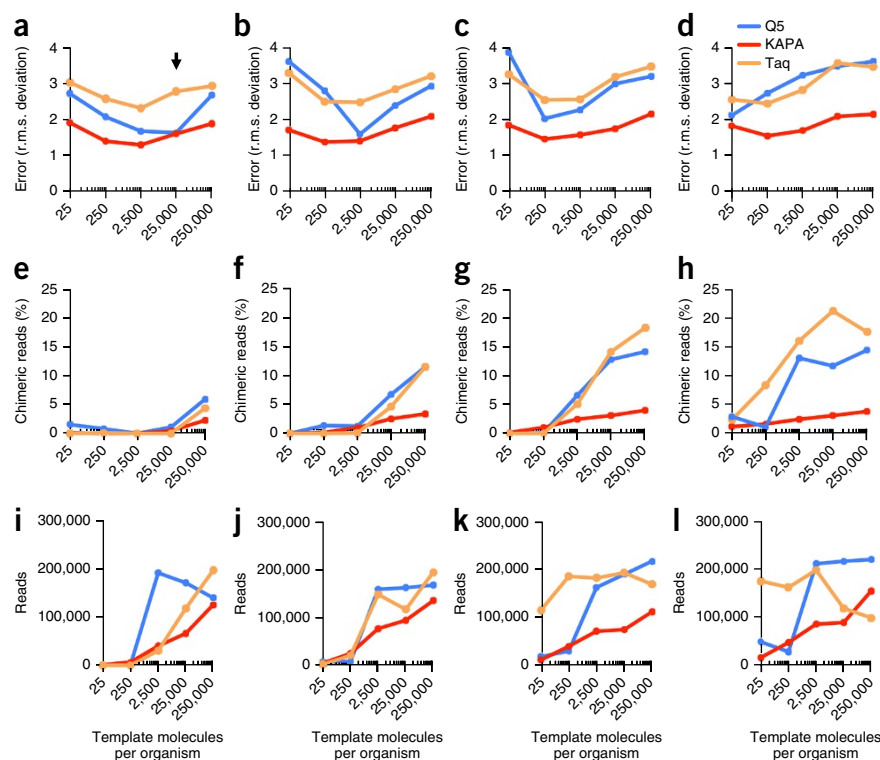
with the same 20 species present at abundances spanning four orders of magnitude)^{3,4}.

The EMP protocol misestimated the abundances of three organisms in the even mock community by more than sevenfold (Fig. 1c). For the staggered mock community, two organisms (*Deinococcus radiodurans* and *Propionibacterium acnes*) were completely undetected (Fig. 1d). The DI (KAPA) protocol was more accurate, with maximal overestimation of approximately 2.3-fold and maximal underestimation of approximately 3.1-fold across both mock communities (Fig. 1e,f, Supplementary Fig. 1).

Polymerase choice affects accuracy

The accuracy of mock community quantification can be expressed as the root mean square deviation (r.m.s. deviation) between observations and expectations (Fig. 1g). In addition to making direct comparisons of the DI and EMP methods, we studied the effect of polymerase

Figure 2 The effect of enzyme choice, PCR cycle number, and template concentration on accuracy, chimera formation, and sample balance. Plots are for the HM-276D even mock community at five different starting template concentrations amplified for different numbers of cycles using KAPA HiFi, Q5, and Taq polymerase. (a–d) Error observed with (a) 20, (b) 25, (c) 30, and (d) 35 amplification cycles. (e–h) The percentage of chimeric reads with (e) 20, (f) 25, (g) 30, and (h) 35 amplification cycles. (i–l) Total number of reads with (i) 20, (j) 25, (k) 30, and (l) 35 amplification cycles. The arrow in **a** indicates the conditions that were used in the experiments with the DI protocol in **Figure 1**.



choice, by using KAPA HiFi polymerase in the EMP method and by comparing three different polymerases in the DI method (Fig. 1g). All methodological variants were highly precise, indicating, as has been previously reported⁵, that the presence of the index sequences in EMP amplification primers is unlikely to significantly affect accuracy (Supplementary Fig. 1). There was also a strong correlation between our EMP data and previously published EMP mock community results⁵¹ (Fig. 1g,h, Supplementary Fig. 1).

However, the methods differed dramatically in their overall accuracy (Fig. 1g, Supplementary Fig. 1). The DI protocol with a proofreading polymerase was considerably more accurate than either the standard EMP protocol or published results from an alternative to the EMP protocol^{31,51}. The DI protocol and substitution of KAPA HiFi for Taq in the EMP method also resulted in a significant reduction of the proportion of chimeric reads (Fig. 1h). In addition, the number of OTUs and genus-level taxa observed was significantly higher in samples amplified using the EMP protocol (Fig. 1i, Supplementary Fig. 1), probably because of the higher number of PCR cycles.

We observed similar patterns of accuracy and chimerism with the staggered mock community (Fig. 1j,k, Supplementary Fig. 1). The species dropouts seen in the EMP (Taq) staggered mock community data were also observed in the EMP (KAPA) data set, and *P. acnes* was absent from the DI (Taq) data set. In contrast, the DI (Q5) and DI (KAPA) data sets, in addition to having considerably higher overall accuracy, showed no such species dropout (Fig. 1f, Supplementary Fig. 1). For the staggered mock community, we noted a significant improvement in accuracy when we used KAPA HiFi in the EMP protocol (Fig. 1j). These results demonstrate that both amplification protocol and polymerase choice have a notable effect on accuracy.

Systematic examination of factors affecting accuracy

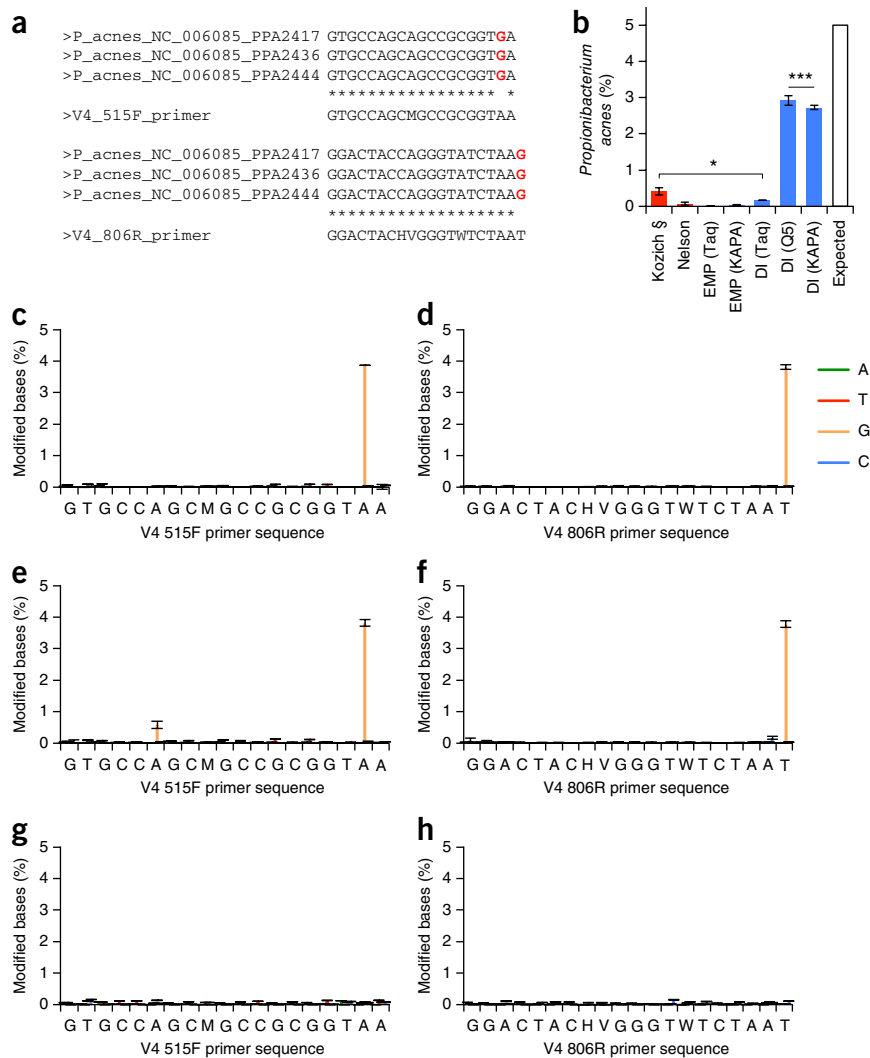
Because these protocols differed in multiple parameters, we next systematically compared the effects of template concentration, enzyme choice, PCR cycle number, and annealing temperature on accuracy. We amplified a dilution series of the even mock community with three different polymerases (Taq, Q5, and KAPA HiFi), for 20, 25, 30, or 35 primary PCR cycles. Across this parameter space, KAPA HiFi was much more accurate than either Q5 or Taq (Fig. 2a–d). An exception was the run with 20 cycles and 25,000 template molecules per organism, in which Q5 and KAPA HiFi led to essentially identical r.m.s. deviation values. By coincidence, these were the conditions

used to collect the data in **Figure 1**. With Q5 polymerase, accuracy decreased with increasing cycle number and template concentration, correlating with the appearance of chimeric reads (Fig. 2e–h). Notably, chimerism with KAPA HiFi was substantially lower than with other enzymes across many conditions, including the high template concentrations and cycle numbers that favor chimera formation (Fig. 2g,h). In contrast, we found that small differences in annealing temperature did not have much of a role in determining differential accuracy (Supplementary Note 1, Supplementary Fig. 2). Finally, even after PicoGreen-based normalization, a correlation remained between template concentration and sample balance, particularly when a proofreading polymerase was used (Fig. 2i–l). Reducing the polymerase concentration was found to improve sample balance (Supplementary Note 2, Supplementary Fig. 3).

Primer editing by proofreading polymerases

One factor contributing to the differences in accuracy among methods was the dropout of *P. acnes* in all but the DI method with proofreading polymerases. The V4 515F and V4 806R primers are reported to perform poorly for detection of *P. acnes*⁶, because of mismatches with the *P. acnes* 16S rRNA gene (Fig. 3a). We detected low levels of *P. acnes* using the EMP protocol and the DI (Taq) protocol; the species was also effectively absent in a published mock community EMP data set⁵¹ (Fig. 3b). Surprisingly, we observed relatively high levels of *P. acnes* with the DI (Q5) and DI (KAPA) protocols (Fig. 3b). When we examined the portion of reads corresponding to the amplification primers for the DI (Q5) and DI (KAPA) data sets, we found that approximately 4% of the V4 515F primer sequences had been edited from A>G at position 18 and approximately 4% of the V4 806R primer sequences had been edited from T>G at position 20, modifications matching the *P. acnes* template sequence (Fig. 3c–f; also see Supplementary Note 3 and Supplementary Fig. 4). No such modifications were observed in the DI (Taq) data set (Fig. 3g,h). These results demonstrate that proofreading polymerases can edit

Figure 3 Primer editing by proofreading polymerases allows recovery of organisms with mismatches to the amplification primers. (a) Alignment of the V4 primer regions of the *P. acnes* 16S rRNA gene to the V4 515F and V4 806R primer sequences. Positions with mismatches to the V4 515F and V4 806R primers are in red. (b) Mean percentage of reads mapped to *P. acnes* with the indicated methods. *** $P < 0.01$, * $P < 0.05$, determined by analysis of variance with Tukey's honest significant difference *post hoc* test. (c–h) Mean percentage of edited bases in different primer regions in HM-276D even mock community data measured with variations of the DI protocol. (c) Data for the V4 515F primer region measured with the DI protocol with Q5 polymerase. (d) Data for the V4 806R primer region measured with the DI protocol with Q5 polymerase. (e) Data for the V4 515F primer region measured with the DI protocol with KAPA HiFi polymerase. (f) Data for the V4 806R primer region measured with the DI protocol with KAPA HiFi polymerase. (g) Data for the V4 515F primer region measured with the DI protocol with Taq polymerase. (h) Data for the V4 806R primer region measured with the DI protocol with Taq polymerase. Error bars are \pm s.e.m.; $n = 4$ (c–h). For (b), $n = 12, 2, 3, 3, 4, 4, 4$.



amplification primers in a PCR reaction, permitting the amplification of sequences from organisms whose templates contain primer mismatches.

Custom primers can cause organism dropout

We were initially puzzled by the very low levels of *P. acnes* seen with the EMP (KAPA) protocol. However, one difference between the EMP and DI protocols is that the EMP sequencing protocol uses custom sequencing primers that fully overlap the V4 515F and V4 806R amplification primers. The discovery of primer editing reveals a major disadvantage to using custom sequencing primers. In amplicons, accurately edited primer sites will be mismatched to the sequencing primer, causing the amplicons to drop out at the sequencing stage (**Supplementary Fig. 5**).

Amplification biases affect measures of species abundance

Despite the improved accuracy observed with the DI (KAPA) method, a residual error of roughly 35% remained (**Supplementary Fig. 1**). For specific organisms, the patterns of deviation between observed and expected frequencies are highly complex. Whereas the abundances of some templates, such as *Escherichia coli* (**Fig. 4a–d**), are relatively consistent across conditions, those of others, such as *Pseudomonas aeruginosa* and *Bacteroides vulgatus*, vary as a function of the polymerase used, template concentration, or PCR cycle number (**Fig. 4e–l**). These nonlinearities in amplification probably result from subtle differences in amplification efficiency due to template and amplicon properties, as well as complex interactions among the many template and amplicon molecules in the reaction^{10,28,37}.

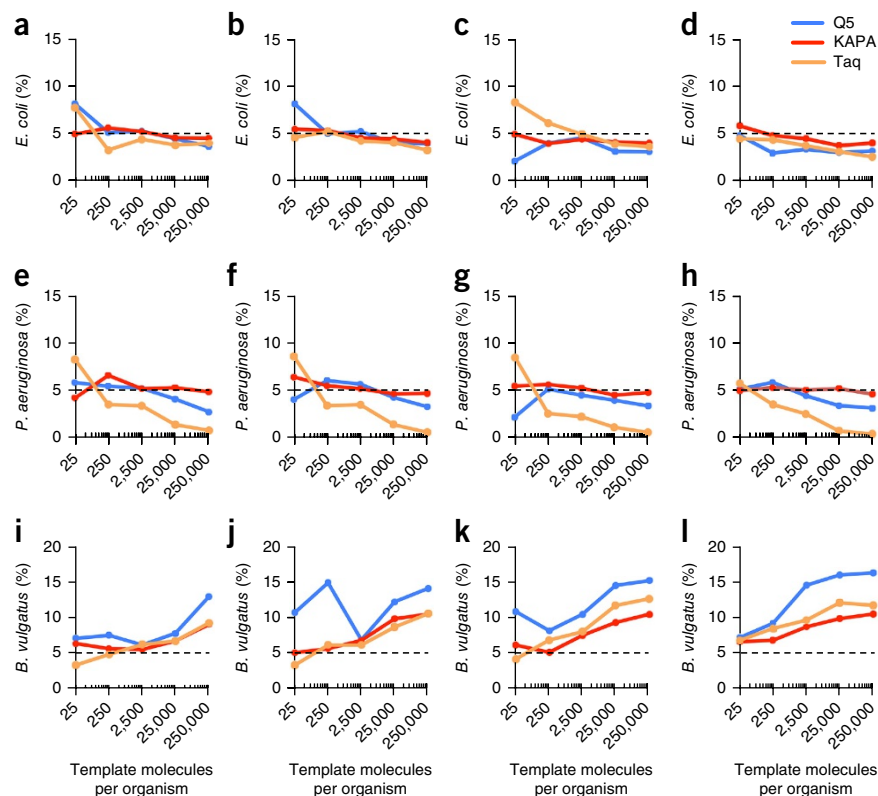
Error and bias in complex samples

To determine whether the sources of error and bias seen in mock community sequencing experiments are evident in more complex

samples, we amplified a collection of fecal samples from wild and semi-captive red-shanked doucs (*Pygathrix nemaeus*), as well as a set of sputum and nasal mucus samples from subjects with cystic fibrosis and paired healthy controls, using either the EMP (Taq) or the DI (KAPA) method. Principal component analysis indicated a strong effect of method, as the NHP samples clearly separated by methodology along the PC2 axis (**Fig. 5a**). However, at a high level, both EMP (Taq) and DI (KAPA) methods led to a similar separation of wild and semi-captive samples along the PC1 axis (**Fig. 5a**). As in mock community data sets, there was an increase in the number of OTUs with the EMP method, particularly for OTUs with low read numbers, which could reflect PCR errors from the non-proofreading polymerase and the increased number of amplification cycles (**Fig. 5b**). A large proportion of chimeric reads were seen with the EMP (Taq) method, whereas the DI (KAPA) method produced very few chimeric reads (**Fig. 5c**).

We detected primer editing in both NHP and human-associated samples amplified with the DI (KAPA) method, resulting in differential detection of a number of taxa compared with EMP (Taq). For instance, a set of TM7 phylum OTUs were present at up to 1.8% in the NHP samples analyzed with the DI (KAPA) protocol, but they were absent from the data sets generated with EMP (Taq) (**Fig. 5d**). Reads from this taxon had an A>C edit at position 18 of the V4 515F primer (**Fig. 5e**). Detecting the TM7 phylum is especially noteworthy,

Figure 4 Nonlinearities in amplification lead to a complex pattern of amplification biases that differentially affect different templates. Plots are for the HM-276D even mock community at five different starting template concentrations amplified for different numbers of cycles using KAPA HiFi, Q5, and Taq polymerase. (a–d) Percent abundance of *E. coli* with (a) 20 cycles, (b) 25 cycles, (c) 30 cycles, and (d) 35 cycles. (e–h) Percent abundance of *P. aereginosa* with (e) 20 cycles, (f) 25 cycles, (g) 30 cycles, and (h) 35 cycles. (i–l) Percent abundance of *B. vulgatus* with (i) 20 cycles, (j) 25 cycles, (k) 30 cycles, and (l) 35 cycles. Expected abundance is indicated by the dashed line.



as shotgun experiments have shown that a large fraction of TM7 bacteria have mismatches predicted to disrupt amplification with the V4 515F and V4 806R primers^{52,53}. Additional examples of taxa with primer mismatches and edits were also detected at appreciable levels in the DI (KAPA) data sets, but not in EMP (Taq) data sets (Supplementary Note 4, Supplementary Figs. 6 and 7).

Modeling observed error distributions in published data sets

To determine whether the mock community error distributions could be expected to affect biological conclusions, we re-noised a collection of published data sets using the observed error distributions for different methods (Supplementary Note 5, Supplementary Table 1, Supplementary Fig. 8). Recovery of differentiated taxa in re-noised data was compared against the original publication, with results reported as recall (the fraction of original significantly differentiated taxa correctly identified) and precision (the fraction of original significantly differentiated taxa among all significantly differentiated taxa identified). We found that the DI (KAPA) method was able to recall, on average, a higher fraction of the original significantly differentiated taxa than the EMP (Taq) method (Fig. 6a). Similar trends were seen with precision (Fig. 6b). In pairwise comparison of all five methods, the DI method had higher mean precision and recall than the EMP method for any given polymerase, and KAPA and Q5 polymerases had higher mean precision and recall than Taq for any given method. Note that we have modeled only quantitative errors. Because the extent to which qualitative dropout of taxa due to primer mismatches manifests in published data is unknown, such dropout was not included in this modeling. Thus, our estimates of recall and precision are probably conservative. Although it is unclear whether the magnitude of errors in the mock community data sets is representative of more complex communities, these results show that such errors have at least the potential to affect experimental results.

DISCUSSION

We compared various methods and explored the parameter space for amplicon-based NGS library preparation. We examined the effects of adaptor addition, template concentration, enzyme choice, annealing temperature, and PCR cycle number. Our results shed light on the sources of error and bias in amplicon-based marker gene experiments. We propose that best practice should include the use of a proofreading polymerase and a highly processive polymerase, that sequencing

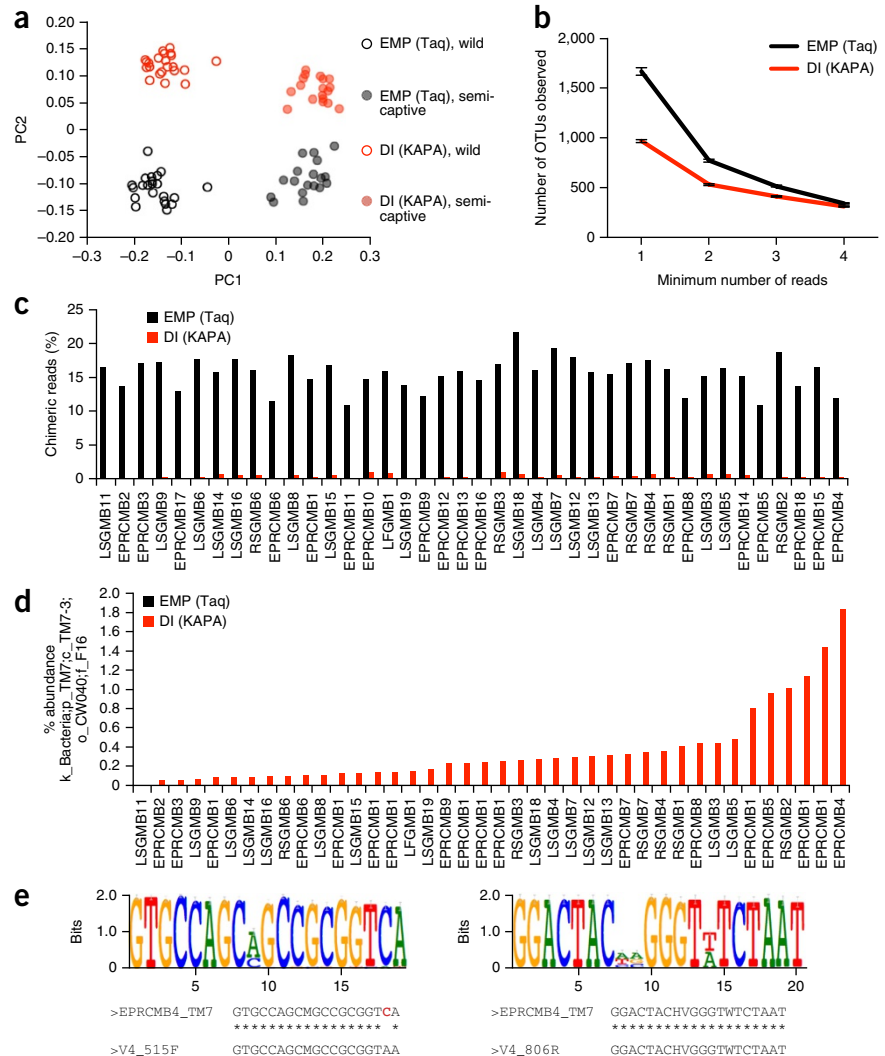
primers that overlap with amplification primers should not be used, that template concentration should be optimized, and that the number of PCR cycles should be minimized.

The substantial differences in quantification detected in this study with different methods, or even with variants of a method, suggest that caution is needed in comparisons of data produced with different protocols. Furthermore, results of amplicon-based microbiome experiments should be verified using orthogonal approaches. The formation of PCR chimeras is not random; rather, it is a function of template abundance³⁵ and sequence homology^{54,55}. Therefore, removal of chimeric reads by computational approaches may skew the resulting microbial profile. One better solution is to prevent chimeric reads from occurring in the first place. We found that using a highly processive polymerase dramatically reduced chimera formation across a broad range of PCR conditions, including high template concentrations and PCR cycle numbers (Figs. 1h,k, 2e–h and 5c). Minimizing PCR cycle number by optimizing the starting template concentration is another, complementary strategy that can be used to reduce the number of chimeric reads^{23,56}. The PCR conditions that minimized chimera formation were also associated with the greatest accuracy, with the exception of very low template concentrations (Fig. 2a–d).

We had hypothesized that using a proofreading polymerase would improve accuracy by minimizing substitution errors⁵⁷, thereby reducing the number of spurious OTUs (Figs. 1g,i,j and 5b, Supplementary Fig. 1). However, an unexpected outcome of using a proofreading polymerase was primer editing. In fact, we were able to detect *P. acnes* only when we used a proofreading polymerase (Fig. 3b), and this detection was enabled by primer editing. The efficiency of primer editing is remarkable given that editing must occur in essentially every PCR cycle in order to be observed in the final sequencing reads.

Primer mismatches in the 3–4 bp at the 3' end of the primer^{27,58,59} are most deleterious to amplification. The frequency of primer

Figure 5 Comparison of EMP (Taq) and DI (KAPA) methods applied to NHP fecal samples. (a) Principal component analysis of NHP fecal samples showing the first two principal components (PC1 and PC2, respectively). Data points representing individual samples are color-coded by method and subject captivity status. (b) Average number of OTUs observed with the indicated methods. Error bars are \pm s.e.m. $n = 40$. (c) Percentage of chimeric reads produced by the EMP (Taq) and DI (KAPA) method. (d) Percent abundance of the *k__Bacteria*; *p__TM7*; *c__TM7-3*; *o__CW040*; *f__F16* taxon as measured by either the EMP (Taq) or DI (KAPA) method. (e) Logo plots and alignments of V4 515F (left) and V4 806R (right) primer sequences to the corresponding region in reads assigned to the *k__Bacteria*; *p__TM7*; *c__TM7-3*; *o__CW040*; *f__F16* taxon in the EMPMB4 sample. A position with mismatch to the V4 515F primer is in red.



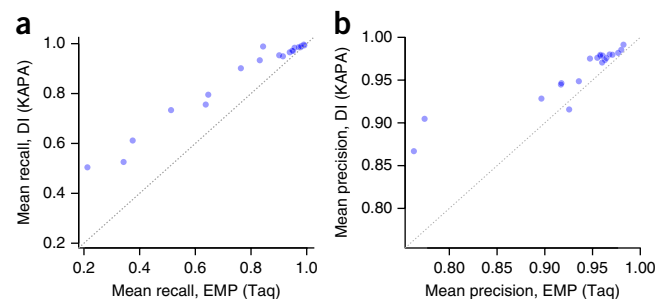
mismatches in 16S rRNA gene ‘universal’ primers has been estimated on the basis of comparison of shotgun data to the sequences of commonly used primers. These comparisons indicated that domain level non-coverage rates of more than 10% were present in 40 out of 56 primer–data set combinations²⁷. Such crucial primer mismatches with the V4 515F and V4 806R primers have been reported in shotgun reads from the TM7 phylum^{52,53}.

We observed primer editing in NHP and human-associated microbial communities, allowing recovery of multiple taxa (including from the TM7 phylum) with the DI (KAPA) method that were essentially absent from the EMP (Taq) data sets (Fig. 5d,e, Supplementary Figs. 6 and 7). Primer editing has another important implication for the design of marker gene NGS methods. Many protocols use custom sequencing primers that overlap with amplification primers, but one of the main benefits of using a proofreading polymerase—preventing dropout of taxa with primer mismatches—is negated when a custom sequencing primer is used (Supplementary Fig. 5). These results show that thorough optimization of protocols can increase the set of organisms identified with amplicon-based approaches.

With ever-decreasing sequencing costs, shotgun sequencing is becoming an alternative to marker gene studies. Although shotgun

sequencing obviates concerns about primer mismatches, methods for preparing shotgun libraries are not without bias; comparisons of even mock community data suggest that the quantitative accuracy of current shotgun approaches might not be greater than that of optimized amplicon-based approaches⁶⁰ (Supplementary Fig. 9). In addition, other factors such as the cost of library preparation, the need for more input material for shotgun sequencing, the ability of amplicon sequencing to detect less abundant organisms, the lack of a requirement for a reference genome database in marker gene surveys (a big issue for poorly characterized taxa, such as fungi), and the potential

Figure 6 Modeling the effect of errors of the magnitude measured for each method on the accuracy of published data sets. (a) Mean recall of the EMP (Taq) method versus that of the DI (KAPA) method. The P value for this comparison is 2.12×10^{-6} , calculated via paired Mann–Whitney U -test (95% confidence interval, $(-0.1319, -0.0292)$; effect size, -0.0823). The dashed diagonal line denotes no difference in mean recall; therefore points above the diagonal indicate treatment comparisons or data sets where DI (KAPA) recovered more of the original differentiated taxa. (b) Mean precision of the EMP (Taq) method versus mean precision of the DI (KAPA) method. The P value for this comparison is 3.18×10^{-5} , calculated via paired Mann–Whitney U -test (95% confidence interval, $(-0.0252, -0.0110)$; effect size, -0.0177). The dashed diagonal line denotes no difference in mean precision; therefore points above the diagonal indicate treatment comparisons or data sets where DI (KAPA) identified more of the original differentiated taxa and fewer false positives.



for host contamination (particularly in samples such as biopsies, in which the ratio of human to microbial DNA is high) will continue to make amplicon-based approaches attractive for many experiments.

To summarize, we report optimizations for amplicon-based microbiome surveys and show that these improvements can have a substantial effect on accuracy. Detailed protocols are available at Protocol Exchange (<http://www.nature.com/protocolexchange/>). It is likely that these principles could be applied more generally to any amplicon-based NGS experiment that aims to accurately quantify relative template abundance.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Accession codes. Sequencing data for this project are available through the NCBI Sequence Read Archive (study accession [SRP069981](#); BioProject [PRJNA305443](#)).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the staff of the University of Minnesota Genomics Center for helpful discussions and technical support. This work was supported by the Minnesota Partnership for Biotechnology and Medical Genomics (grant MNP IF #14.09). This work was carried out in part using computing resources at the University of Minnesota Supercomputing Institute. This work was also supported by the Margot Marsh Biodiversity Foundation and the US National Institutes of Health (PharmacoNeuroImmunology Fellowship NIH/NIDA T32 DA007097-32 to J.B.C.).

AUTHOR CONTRIBUTIONS

D.M.G. and K.B.B. conceived and designed the experiments, analyzed data, and wrote the manuscript; J.G. and T.J.G. contributed to the analysis; P.V. and D.K. carried out the modeling and helped write the manuscript; D.M.G., A.M., A.H., and A.B. conducted the experiments; J.B.C., T.J.J., and R.H. contributed experimental samples.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the [online version of the paper](#).

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Cho, I. & Blaser, M.J. The human microbiome: at the interface of health and disease. *Nat. Rev. Genet.* **13**, 260–270 (2012).
- Gilbert, J.A., Jansson, J.K. & Knight, R. The Earth Microbiome project: successes and aspirations. *BMC Biol.* **12**, 69 (2014).
- The Human Microbiome Project Consortium A framework for human microbiome research. *Nature* **486**, 215–221 (2012).
- Jumpstart Consortium Human Microbiome Project Data Generation Working Group Evaluation of 16S rDNA-based community profiling for human microbiome research. *PLoS One* **7**, e39315 (2012).
- Goodrich, J.K. *et al.* Conducting a microbiome study. *Cell* **158**, 250–262 (2014).
- Kuczynski, J. *et al.* Experimental and analytical tools for studying the human microbiome. *Nat. Rev. Genet.* **13**, 47–58 (2012).
- Caporaso, J.G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
- Schloss, P.D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
- Salter, S.J. *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12**, 87 (2014).
- Brooks, J.P. *et al.* The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiol.* **15**, 66 (2015).
- Pinto, A.J. & Raskin, L. PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PLoS One* **7**, e43093 (2012).
- Sinha, R., Abnet, C.C., White, O., Knight, R. & Huttenhower, C. The microbiome quality control project: baseline study design and future directions. *Genome Biol.* **16**, 276 (2015).
- Zhou, J. *et al.* Random sampling process leads to overestimation of β -diversity of microbial communities. *MBio* **4**, e00324–13 (2013).
- Yuan, S., Cohen, D.B., Ravel, J., Abdo, Z. & Forney, L.J. Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One* **7**, e33865 (2012).
- Kennedy, N.A. *et al.* The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. *PLoS One* **9**, e88982 (2014).
- Feinstein, L.M., Sul, W.J. & Blackwood, C.B. Assessment of bias associated with incomplete extraction of microbial DNA from soil. *Appl. Environ. Microbiol.* **75**, 5428–5433 (2009).
- Zhao, J. *et al.* Effect of sample storage conditions on culture-independent bacterial community measures in cystic fibrosis sputum specimens. *J. Clin. Microbiol.* **49**, 3717–3718 (2011).
- Cardona, S. *et al.* Storage conditions of intestinal microbiota matter in metagenomic analysis. *BMC Microbiol.* **12**, 158 (2012).
- Aird, D. *et al.* Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* **12**, R18 (2011).
- Ahn, J.-H., Kim, B.-Y., Song, J. & Weon, H.-Y. Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities. *J. Microbiol.* **50**, 1071–1074 (2012).
- Wu, J.-Y. *et al.* Effects of polymerase, template dilution and cycle number on PCR based 16 S rRNA diversity analysis using the deep sequencing method. *BMC Microbiol.* **10**, 255 (2010).
- Ishii, K. & Fukui, M. Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. *Appl. Environ. Microbiol.* **67**, 3753–3755 (2001).
- D'Amore, R. *et al.* A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. *BMC Genomics* **17**, 55 (2016).
- Kennedy, K., Hall, M.W., Lynch, M.D.J., Moreno-Hagelsieb, G. & Neufeld, J.D. Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles. *Appl. Environ. Microbiol.* **80**, 5717–5722 (2014).
- Hansen, M.C., Tolker-Nielsen, T., Givskov, M. & Molin, S. Biased 16S rDNA PCR amplification caused by interference from DNA flanking the template region. *FEMS Microbiol. Ecol.* **26**, 141–149 (1998).
- Reysenbach, A.L., Giver, L.J., Wickham, G.S. & Pace, N.R. Differential amplification of rRNA genes by polymerase chain reaction. *Appl. Environ. Microbiol.* **58**, 3417–3418 (1992).
- Mao, D.-P., Zhou, Q., Chen, C.-Y. & Quan, Z.-X. Coverage evaluation of universal bacterial primers using the metagenomic datasets. *BMC Microbiol.* **12**, 66 (2012).
- Polz, M.F. & Cavanaugh, C.M. Bias in template-to-product ratios in multitemplate PCR. *Appl. Environ. Microbiol.* **64**, 3724–3730 (1998).
- Hong, S., Bunge, J., Leslin, C., Jeon, S. & Epstein, S.S. Polymerase chain reaction primers miss half of rRNA microbial diversity. *ISME J.* **3**, 1365–1373 (2009).
- Klindworth, A. *et al.* Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **41**, e1 (2013).
- Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K. & Schloss, P.D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013).
- Quail, M.A. *et al.* Optimal enzymes for amplifying sequencing libraries. *Nat. Methods* **9**, 10–11 (2012).
- Schloss, P.D., Gevers, D. & Westcott, S.L. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One* **6**, e27310 (2011).
- Patin, N.V., Kunin, V., Lidström, U. & Ashby, M.N. Effects of OTU clustering and PCR artifacts on microbial diversity estimates. *Microb. Ecol.* **65**, 709–719 (2013).
- Haas, B.J. *et al.* Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* **21**, 494–504 (2011).
- Wagner, A. *et al.* Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift. *Syst. Biol.* **43**, 250–261 (1994).
- Suzuki, M.T. & Giovannoni, S.J. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl. Environ. Microbiol.* **62**, 625–630 (1996).
- Schirmer, M. *et al.* Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res.* **43**, e37 (2015).
- Zhou, H.-W. *et al.* BIPES, a cost-effective high-throughput method for assessing microbial diversity. *ISME J.* **5**, 741–749 (2011).
- Degnan, P.H. & Ochman, H. Illumina-based analysis of microbial community diversity. *ISME J.* **6**, 183–194 (2012).
- Gloor, G.B. *et al.* Microbiome profiling by illumina sequencing of combinatorial sequence-tagged PCR products. *PLoS One* **5**, e15406 (2010).
- Claesson, M.J. *et al.* Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res.* **38**, e200 (2010).
- Caporaso, J.G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621–1624 (2012).
- Fadrosh, D.W. *et al.* An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* **2**, 6 (2014).

45. Bartram, A.K., Lynch, M.D.J., Stearns, J.C., Moreno-Hagelsieb, G. & Neufeld, J.D. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. *Appl. Environ. Microbiol.* **77**, 3846–3852 (2011).
46. Salipante, S.J. *et al.* Performance comparison of Illumina and ion torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. *Appl. Environ. Microbiol.* **80**, 7583–7591 (2014).
47. Illumina 16S metagenomic sequencing library preparation (Illumina Technical Note 15044223 Rev. A). *Illumina* http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf (2013).
48. Faith, J.J. *et al.* The long-term stability of the human gut microbiota. *Science* **341**, 1237439 (2013).
49. Lundberg, D.S., Yourstone, S., Mieczkowski, P., Jones, C.D. & Dangl, J.L. Practical innovations for high-throughput amplicon sequencing. *Nat. Methods* **10**, 999–1002 (2013).
50. Lee, C.K. *et al.* Groundtruthing next-gen sequencing for microbial ecology-biases and errors in community structure estimates from PCR amplicon pyrosequencing. *PLoS One* **7**, e44224 (2012).
51. Nelson, M.C., Morrison, H.G., Benjamino, J., Grim, S.L. & Graf, J. Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. *PLoS One* **9**, e94249 (2014).
52. Brown, C.T. *et al.* Unusual biology across a group comprising more than 15% of domain bacteria. *Nature* **523**, 208–211 (2015).
53. Eloe-Fadrosh, E.A., Ivanova, N.N., Woyke, T. & Kyrpides, N.C. Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. *Nat. Microbiol.* **1**, 15032 (2016).
54. Wang, G.C. & Wang, Y. Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. *Appl. Environ. Microbiol.* **63**, 4645–4650 (1997).
55. Wang, G.C. & Wang, Y. The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species. *Microbiology* **142**, 1107–1114 (1996).
56. Lahr, D.J.G. & Katz, L.A. Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. *Biotechniques* **47**, 857–866 (2009).
57. Kunkel, T.A. & Bebenek, K. DNA replication fidelity. *Annu. Rev. Biochem.* **69**, 497–529 (2000).
58. Ayyadevara, S., Thaden, J.J. & Shmookler Reis, R.J. Discrimination of primer 3'-nucleotide mismatch by taq DNA polymerase during polymerase chain reaction. *Anal. Biochem.* **284**, 11–18 (2000).
59. Bru, D., Martin-Laurent, F. & Philippot, L. Quantification of the detrimental effect of a single primer-template mismatch by real-time PCR using the 16S rRNA gene as an example. *Appl. Environ. Microbiol.* **74**, 1660–1663 (2008).
60. Jones, M.B. *et al.* Library preparation methodology can influence genomic and functional predictions in human microbiome research. *Proc. Natl. Acad. Sci. USA* **112**, 14024–14029 (2015).

ONLINE METHODS

Samples. The mock community DNA was obtained through BEI Resources, National Institute of Allergy and Infectious Diseases, US National Institutes of Health, as part of the Human Microbiome Project: Genomic Mock Community B (HM-276D, Even, High Concentration, v5.1H, and HM-277D, Staggered, High Concentration, v5.2H).

Fecal samples were collected opportunistically from semi-captive and wild red-shanked doucs (*P. nemaus*) immediately after defecation between 2012 and 2013. The red-shanked douc is a species of NHP that inhabits Lao People's Democratic Republic, Vietnam, and Cambodia. Doucs housed at the Endangered Primate Rescue Center in Ninh Binh, Vietnam, served as the semi-captive population. Doucs inhabiting Son Tra Nature Reserve, Da Nang, Vietnam, served as the wild population. Samples were placed on ice for transport and kept frozen at a maximum of -20°C until processing. DNA was extracted using the established method of Yu and Morrison⁶¹ with some modifications. Specifically, two rounds of bead-beating were carried out in the presence of NaCl and sodium dodecyl sulfate and were followed by ammonium acetate and isopropanol precipitations. Precipitated nucleic acids were treated with DNase-free RNase (Roche). DNA was purified using the QIAamp DNA Stool Mini Kit (Qiagen, Valencia, CA). DNA quantity was measured using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific).

For the human cystic fibrosis (CF) samples, 27 participants with CF were recruited during scheduled visits to the University of Minnesota Adult CF Center. Four non-CF subjects undergoing functional endoscopic sinus surgery were also recruited to serve as healthy controls. From each subject, a single sputum or sinus mucus sample was collected in a 50-ml conical tube that was then placed on ice and stored at -80°C before analysis. Samples were thawed to room temperature, and 0.5 ml of each sample was used for genomic DNA extraction using the PowerSoil DNA isolation kit (MoBio, Carlsbad, CA), which included bead-beating to lyse bacterial cells. Samples were obtained with prior informed consent from each subject enrolled in the study and with approval by the Institutional Review Board at the University of Minnesota (UMN IRB nos. 1401M47262 and 1404M49426).

PCR amplification. EMP method. Amplifications in the EMP method were carried out as previously described⁴³. Briefly, we diluted mock community samples 1:50 in sterile, nuclease-free water, such that the same template concentration was used in both EMP and DI amplification experiments (other samples were diluted 6.25-fold). We set up PCR reactions using the following recipe: 5 μl of diluted template DNA, 0.5 μl of forward primer (10 μM), 5 μl of reverse primer (0.8 μM), 1.5 μl of nuclease-free water, and 8 μl of 2.5 \times 5 PRIME HotMasterMix (5 PRIME, Gaithersburg, MD). We amplified samples using the following cycling conditions: 94 $^{\circ}\text{C}$ for 3 min; 35 cycles of 94 $^{\circ}\text{C}$ for 45 s, 50 $^{\circ}\text{C}$ for 60 s, and 72 $^{\circ}\text{C}$ for 90 s; and then a final extension at 72 $^{\circ}\text{C}$ for 10 min.

EMP method (substituting KAPA HiFi polymerase). We carried out amplifications using the EMP primers with KAPA HiFi polymerase as described above with the following modifications. The PCR recipe was as follows: 5 μl of diluted template DNA, 0.5 μl of forward primer (10 μM), 5 μl of reverse primer (0.8 μM), 3.5 μl of nuclease-free water, 4 μl of 5 \times KAPA HiFi buffer (KAPA Biosystems, Woburn, MA), 0.6 μl of 10 mM dNTPs (KAPA Biosystems, Woburn, MA), 1 μl of DMSO (Fisher Scientific, Waltham, MA), and 0.4 μl of KAPA HiFi Polymerase (KAPA Biosystems, Woburn, MA). We amplified samples using the following cycling conditions: 95 $^{\circ}\text{C}$ for 5 min; 35 cycles of 98 $^{\circ}\text{C}$ for 20 s, 55 $^{\circ}\text{C}$ for 15 s, and 72 $^{\circ}\text{C}$ for 1 min; and then a final extension at 72 $^{\circ}\text{C}$ for 10 min.

DI method. We amplified the V4 region of the 16S rRNA gene using a two-step PCR protocol. The primary amplification was done in a qPCR reaction, using ABI7900 so that the dynamics of the PCR reactions could be monitored. The following recipe was used: 3 μl of template DNA, 0.48 μl of nuclease-free water, 1.2 μl of 5 \times KAPA HiFi buffer (KAPA Biosystems, Woburn, MA), 0.18 μl of 10 mM dNTPs (KAPA Biosystems, Woburn, MA), 0.3 μl of DMSO (Fisher Scientific, Waltham, MA), 0.12 μl of ROX (25 μM) (Life Technologies, Carlsbad, CA), 0.003 μl of 1,000 \times SYBR Green, 0.12 μl of KAPA HiFi polymerase (KAPA Biosystems, Woburn, MA), 0.3 μl of forward primer (10 μM), and 0.3 μl of reverse primer (10 μM). Cycling conditions were as follows: 95 $^{\circ}\text{C}$ for 5 min, followed by 20 cycles of 98 $^{\circ}\text{C}$ for 20 s, 55 $^{\circ}\text{C}$ for 15 s, and 72 $^{\circ}\text{C}$ for 1 min.

The primers for the primary amplification contained both 16S-specific primers (V4_515F and V4_806R) and adaptor tails for adding indices and Illumina flow cell adaptors in a secondary amplification. The following primers were used (16S-specific sequences in bold):

V4_515F_Nextera:
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGCCAGC
MGCCGCGGTAA
V4_806R_Nextera:
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGACTACHVG
GGTWTCTAAT

The amplicons from the primary PCR were diluted 1:100 in sterile, nuclease-free water, and a second PCR reaction was set up to add the Illumina flow cell adaptors and indices. The secondary amplification was done using the following recipe: 5 μl of template DNA, 1 μl of nuclease-free water, 2 μl of 5 \times KAPA HiFi buffer (KAPA Biosystems, Woburn, MA), 0.3 μl of 10 mM dNTPs (KAPA Biosystems, Woburn, MA), 0.5 μl of DMSO (Fisher Scientific, Waltham, MA), 0.2 μl of KAPA HiFi Polymerase (KAPA Biosystems, Woburn, MA), 0.5 μl of forward primer (10 μM), and 0.5 μl of reverse primer (10 μM). Cycling conditions were as follows: 95 $^{\circ}\text{C}$ for 5 min; ten cycles of 98 $^{\circ}\text{C}$ for 20 s, 55 $^{\circ}\text{C}$ for 15 s, and 72 $^{\circ}\text{C}$ for 1 min; and a final extension at 72 $^{\circ}\text{C}$ for 10 min. The following indexing primers were used (X indicates the positions of the 8-bp indices):

Forward indexing primer:
AATGATACGGCGACCACCGAGATCTACACXXXXXXXXXXTCGTCGG
CAGCGTC
Reverse indexing primer:
CAAGCAGAAGACGGCATACGAGATXXXXXXXXXXGTCTCGTGGGC
TCGG

Dilution series experiments. For the dilution series experiments, we used the DI-method primers (V4_515F_Nextera and V4_806R_Nextera, above) for all of the comparisons. We amplified a tenfold dilution series of the HM-276D mock community DNA for 20, 25, 30, or 35 cycles, using one of three different polymerases: KAPA HiFi HotStart (KAPA Biosystems, Woburn, MA), Q5 Hot Start High-Fidelity Master Mix (New England BioLabs, Ipswich, MA), or 5 PRIME HotMasterMix (5 PRIME, Gaithersburg, MD). PCR recipes and cycling conditions for the primary amplifications were as follows.

KAPA HiFi primary PCR recipe: 2.5 μl of DNA template, 3.5 μl of nuclease-free water, 2 μl of 5 \times KAPA HiFi buffer (KAPA Biosystems, Woburn, MA), 0.3 μl of 10 mM dNTPs (KAPA Biosystems, Woburn, MA), 0.5 μl of DMSO (Fisher Scientific, Waltham, MA), 0.2 μl of KAPA HiFi Polymerase (KAPA Biosystems, Woburn, MA), 0.5 μl of forward primer (10 μM), and 0.5 μl of reverse primer (10 μM).

KAPA HiFi cycling conditions: 95 $^{\circ}\text{C}$ for 5 min; 20, 25, 30, or 35 cycles of 98 $^{\circ}\text{C}$ for 20 s, 55 $^{\circ}\text{C}$ for 15 s, and 72 $^{\circ}\text{C}$ for 1 min; and 72 $^{\circ}\text{C}$ for 5 min.

Q5 PCR primary recipe: 2.5 μl of DNA template, 5 μl of 2 \times Q5 Hot Start High-Fidelity Master Mix, 0.5 μl of forward primer (10 μM), 0.5 μl of reverse primer (10 μM), and 1.5 μl of sterile, nuclease-free water.

Q5 cycling conditions: 98 $^{\circ}\text{C}$ for 30 s; 20, 25, 30, or 35 cycles of 98 $^{\circ}\text{C}$ for 20 s, 55 $^{\circ}\text{C}$ for 15 s, and 72 $^{\circ}\text{C}$ for 1 min; and 72 $^{\circ}\text{C}$ for 5 min.

5 PRIME Taq primary PCR recipe: 2.5 μl of DNA template, 4 μl of 2.5 \times 5 PRIME HotMasterMix, 0.5 μl of forward primer (10 μM), 0.5 μl of reverse primer (10 μM), and 2.5 μl of sterile, nuclease-free water.

5 PRIME Taq cycling conditions: 94 $^{\circ}\text{C}$ for 3 min; 20, 25, 30, or 35 cycles of 94 $^{\circ}\text{C}$ for 20 s, 55 $^{\circ}\text{C}$ for 15 s, and 72 $^{\circ}\text{C}$ for 1 min; and 72 $^{\circ}\text{C}$ for 5 min.

Primary PCRs were then diluted 1:100 in sterile, nuclease-free water, and a second PCR reaction was set up to add the Illumina flow cell adaptors and indices. For those reactions we used the following recipes (polymerase-specific cycling conditions were the same as above, but using ten cycles in the indexing step).

KAPA HiFi indexing PCR recipe: 5 μl of 1:100 DNA template, 1 μl of nuclease-free water, 2 μl of 5 \times KAPA HiFi buffer (KAPA Biosystems, Woburn, MA), 0.3 μl of 10 mM dNTPs (KAPA Biosystems, Woburn, MA), 0.5 μl of DMSO (Fisher Scientific, Waltham, MA), 0.2 μl of KAPA HiFi polymerase (KAPA Biosystems, Woburn, MA), 0.5 μl of forward primer (10 μM), and 0.5 μl of reverse primer (10 μM).

Q5 indexing PCR recipe: 5 μ l of 1:100 DNA template, 5 μ l of 2 \times Q5 Hot Start High-Fidelity Master Mix, and dried-down indexing primers (final concentration of 0.5 μ M for each primer).

5 PRIME Taq indexing PCR recipe: 5 μ l of 1:100 DNA template, 4 μ l of 2.5 \times 5 PRIME HotMasterMix, 1 μ l of sterile, nuclease-free water, and dried-down indexing primers (final concentration of 0.5 μ M for each primer).

KAPA HiFi concentration tests. For the KAPA HiFi concentration tests, we carried out amplifications using the KAPA HiFi primary PCR recipe and cycling conditions described in the “Dilution series experiments” section above, but we cut the amount of KAPA HiFi polymerase added to the 0.5 \times reactions in half (0.1 μ l per 10- μ l reaction) and reduced the amount added to the 0.25 \times reactions to one-fourth the 1 \times concentration (0.05 μ l per 10- μ l reaction); we added nuclease-free water to compensate for the missing volume. The indexing reactions for each of these conditions were carried out with the 0.5 \times concentration of KAPA HiFi polymerase, so the differences observed between these conditions are a result of the differing KAPA HiFi polymerase concentrations in the primary PCR reaction.

KAPA HiFi Readymix amplifications. We carried out KAPA HiFi ReadyMix PCRs as described above, using the DI primers (V4_515F_Nextera and V4_806R_Nextera, above) with the following recipes:

KAPA HiFi Readymix PCR recipe: 2.5 μ l of DNA template, 5 μ l of 2 \times KAPA HiFi HotStart Readymix, 0.5 μ l of forward primer (10 μ M), 0.5 μ l of reverse primer (10 μ M), and 1.5 μ l of sterile, nuclease-free water.

KAPA HiFi ReadyMix indexing PCR recipe: 5 μ l of 1:100 DNA template, 5 μ l of 2 \times KAPA HiFi HotStart Readymix, and dried-down indexing primers (final concentration of 0.5 μ M for each primer).

Annealing temperature comparisons. The 35-PCR-cycle HM-276D mock community dilution series experiments described above were also run with an annealing temperature of 50 $^{\circ}$ C instead of 55 $^{\circ}$ C using the commercial master mixes for each of the three polymerases (KAPA HiFi ReadyMix, Q5 Hot Start High-Fidelity Master Mix, and Taq 5 PRIME HotMasterMix). The same recipes and protocols were used to process these samples, with the exception of the altered annealing temperature.

Amplifying *C. jejuni* V4 and V3–V5 variable regions. We amplified DNA from a pure isolate of *Campylobacter jejuni* (81–176) using the V4 515F and V4 806R primers and the KAPA ReadyMix protocol described above, or using the KAPA HiFi (1 \times) protocol with primers for the V3–V5 variable region (**Supplementary Fig. 4**). The primer sequences for the primary amplification for the V3–V5 variable region were as follows (16S-specific sequences in bold):

V3F_Nextera:

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**CTACGGGAGG CAGCAG**

V5R_Nextera:

GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**CCGTCATTC MTTTRAGT**

Normalization and pooling of sequencing libraries. We quantified PCR products using a PicoGreen dsDNA assay kit (Life Technologies, Carlsbad, CA), normalized and pooled the samples, and concentrated approximately 1 μ g of material to 10 μ l using 1.8 \times AMPureXP beads (Beckman Coulter, Brea, CA). The pooled sample was then size-selected at 427 bp \pm 20% for the DI pools, or at 368 bp \pm 20% for the EMP pools, on a Caliper XT DNA 750 chip (Caliper Life Science, Hopkinton, MA). The size-selected material was cleaned up using AMPureXP beads and eluted in 20 μ l of EB buffer (10 mM Tris-HCl, pH 8.5). The final pooled sample was quantified using the PicoGreen dsDNA assay.

Sequencing. The sample pools were diluted to 2 nM on the basis of the PicoGreen measurements, and 10 μ l of the 2 nM pool was denatured with 10 μ l of 0.2 N NaOH, diluted to 8 pM in Illumina’s HT1 buffer, spiked with 15% PhiX, heat-denatured at 96 $^{\circ}$ C for 2 min, and sequenced with a MiSeq 600 cycle v3 kit (Illumina, San Diego, CA). For the EMP method, the custom sequencing primers described by Caporaso *et al.*⁴³ were added to the appropriate wells of the MiSeq reagent cartridge.

Analysis. We subsampled the mock community samples to a depth of 10,000 reads per sample. We then trimmed sequencing adaptor sequences using

Trimmomatic⁶² and used PANDAseq⁶³ to remove primer sequences (where applicable) and join paired-end reads. We converted FASTQ files to QIIME⁷ FASTQ format using a custom script. Next, we concatenated individual sample FASTA files into one FASTA file and ran chimera detection and removal using ChimeraSlayer’s usearch61 method³⁵. We mapped the resulting reads to a Human Microbiome Project mock community reference file⁴⁶ for the calculation of the percent abundance, r.m.s. deviation, and mean absolute percent error. For OTU picking, we used QIIME’s pick_open_reference_otus.py script with usearch61. We assigned taxonomy using QIIME’s assign_taxonomy.py script, which uses the RDP classifier and the Greengenes reference database clustered at 97% identity. We then collapsed reference-based OTUs according to taxonomy at the genus level using QIIME’s summarize_taxa.py script. We subsampled the NHP and human samples to an initial depth of 40,000 reads per sample and pre-processed them as described above. We subsampled the reads remaining after chimera detection down to a depth of 20,000 reads per sample, and we carried out OTU picking using QIIME’s pick_open_reference_otus.py script with usearch61. We assigned taxonomy as described previously, and we generated beta diversity principal coordinate analysis plots using QIIME’s beta_diversity_through_plots.py script. We analyzed the distribution of primer corrections by cataloging mismatches to the V4 primer sequences using custom Python scripts and BioPython⁶⁴. We trimmed Illumina adaptors using cutadapt⁶⁵, and we merged paired reads using PANDAseq⁶³. To filter out noise from indels in the primer regions, we used a threshold of a maximum of three mismatches per primer sequence for this analysis. We analyzed the primer sequences associated with the differentially abundant OTUs in the NHP and human data sets by searching for exact matches to the rep_set sequences from these OTUs in the untrimmed subsampled FASTQ files. We used WebLogo to generate the sequence logo plots⁶⁶.

Modeling. For each published data set, we summarized the OTU table at the genus level using QIIME⁷. We then applied error distributions for each of the different methods obtained from the mock community experiments described above to these taxon tables (we multiplied the relative abundances of each taxon by an error rate sampled with replacement). We filtered the taxon tables to remove rare taxa that were present in less than 0.01% of all samples and then arc-sin square root-transformed them for normality before testing for differences using a *t*-test. We adjusted the resulting *P* values for multiple comparisons using the false discovery rate method, and we considered any taxon with a resulting *q*-value \leq 0.05 as differentiated between the treatment groups. We used these same methods for genus-level taxon tables that did not have polymerase error distribution applied. We calculated recall as the number of correctly identified differentiated taxa found with the re-noised taxa table divided by the number of differentiated taxa found in the original taxa table. We calculated precision as the number of correctly identified differentiated taxa found with the modified taxa table divided by the number of all differentiated taxa found with the modified taxa table. We repeated the process of adding noise to the original taxa table and tests for differentiation 1,000 times to generate precision and recall values. We calculated *P* values for **Figure 6** using a paired Mann–Whitney *U*-test.

Optimized high-throughput amplicon-based microbiome profiling. Below is a step-by-step summary of the optimized DI (KAPA) method presented here. A more detailed protocol is available at Protocol Exchange (<http://www.nature.com/protocolexchange/>).

1. Optimize template concentration. Because template concentration can have large effects on amplification bias, it should be optimized before or concomitant with amplification. If samples have uniform DNA concentration and microbial load, it may be sufficient to use a consistent concentration of template DNA. For more complex or variable samples, a qPCR-based protocol for optimizing template concentration is described in the detailed protocol at Protocol Exchange.
2. Primary PCR amplification. Amplify samples using the following recipe (note: to avoid the need to pipette small volumes, this should be made up as a master mix with all components other than the template DNA, and 3 μ l of the master mix should then be mixed with 3 μ l of template DNA):
 - 3 μ l of template DNA
 - 0.6 μ l of nuclease-free water

- 1.2 µl of 5× KAPA HiFi buffer (Kapa Biosystems, Woburn, MA)
- 0.18 µl of 10 mM dNTPs (Kapa Biosystems, Woburn, MA)
- 0.3 µl of DMSO (Fisher Scientific, Waltham, MA)
- 0.12 µl of KAPA HiFi polymerase (Kapa Biosystems, Woburn, MA)
- 0.3 µl of V4_515F_Nextera primer (10 µM)
- 0.3 µl of V4_806R_Nextera primer (10 µM)

Use the following cycling conditions: 95 °C for 5 min, followed by 15–30 cycles of 98 °C for 20 s, 55 °C for 15 s, and 72 °C for 1 min. Hold at 4 °C. Note: for additional primer sets targeting other variable regions, see the detailed protocol at Protocol Exchange.

3. Secondary PCR amplification (indexing PCR). Dilute the amplicons from the primary PCR 1:100 in sterile, nuclease-free water, and set up a second PCR reaction to add the Illumina flow cell adaptors and indices using the following recipe (note: to avoid the need to pipette small volumes, this should be made up as a master mix with all components other than the template DNA and indexing primers, and 3 µl of the master mix should then be mixed with 5 µl of template DNA and 2 µl of indexing primer mix):

- 5 µl of template DNA
- 2 µl of 5× KAPA HiFi buffer (KAPA Biosystems, Woburn, MA)
- 0.3 µl of 10 mM dNTPs (KAPA Biosystems, Woburn, MA)
- 0.5 µl of DMSO (Fisher Scientific, Waltham, MA)
- 0.2 µl of KAPA HiFi Polymerase (KAPA Biosystems, Woburn, MA)
- 1 µl of forward indexing primer (5 µM)
- 1 µl of reverse indexing primer (5 µM)

Use the following cycling conditions: 95 °C for 5 min; ten cycles of 98 °C for 20 s, 55 °C for 15 s, and 72 °C for 1 min; and a final extension at 72 °C for 10 min. Hold at 4 °C. Note: for indexing primer sequences, see the detailed protocol at Protocol Exchange.

4. Sample normalization and pooling. Samples can be normalized and pooled in one of two ways:

- a. Concentration-based normalization. Quantify PCR products using a PicoGreen dsDNA assay kit (Life Technologies, Carlsbad, CA). Normalize samples to a consistent concentration by adding an appropriate volume of nuclease-free water and pool an equal volume of each of the normalized samples. Use a SpeedVac to concentrate the sample pool down to 20–100 µl and purify the sample pool using

1× AmpPureXP beads. Elute in 25 µl of nuclease-free water or elution buffer (10 mM Tris-Cl, 1 mM EDTA, pH 8.0). The library can be optionally size-selected using a Caliper XT DNA 750 chip (Caliper Life Science, Hopkinton, MA) or similar.

- b. Plate-based normalization. Purify the 10-µl indexing PCR reactions using a SequelPrep normalization plate according to the manufacturer's instructions. Dispense 10 µl of each sample to be pooled into a trough. Mix well and transfer from the trough to a 1.5-ml microfuge tube. For projects with large numbers of samples, the volume of pooled material can be adjusted downward. Use a SpeedVac to concentrate the sample pool down to 20–100 µl and purify the sample pool using 1× AmpPureXP beads. Elute in 25 µl of nuclease-free water or elution buffer (10 mM Tris-Cl, 1 mM EDTA, pH 8.0).
5. Library QC. Perform library QC by running the pooled library on an Agilent DNA HS chip, following the manufacturer's protocol and verifying that the size distribution is as expected. Determine the concentration of the pooled library by PicoGreen assay (according to the manufacturer's protocol).
6. Sequencing. Dilute the pooled library to 2 nM in elution buffer using the concentration determined by the PicoGreen assay and the expected size for the amplicon. Denature 10 µl of the 2 nM pool by mixing with 10 µl of 0.2 N NaOH for 5 min. Dilute denatured sample to 8 pM in Illumina's HT1 buffer, spike with 15% PhiX, and sequence using an Illumina MiSeq or HiSeq according to the manufacturer's instructions (Illumina, San Diego, CA).

61. Yu, Z. & Morrison, M. Improved extraction of PCR-quality community DNA from digests and fecal samples. *Biotechniques* **36**, 808–812 (2004).
62. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
63. Masella, A.P., Bartram, A.K., Truszkowski, J.M., Brown, D.G. & Neufeld, J.D. PANDAseq: paired-end assembler for Illumina sequences. *BMC Bioinformatics* **13**, 31 (2012).
64. Cock, P.J.A. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
65. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* **17**, 10–12 (2011).
66. Crooks, G.E., Hon, G., Chandonia, J.-M. & Brenner, S.E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).