Evaluation of Compatibility of 16S rRNA V3V4 and V4 Amplicon Libraries for Clinical Microbiome Profiling

Po-Yu Liu (OrcID: 0000-0003-1290-0850),[a] Wei-Kai Wu,[b,c] Chieh-Chang Chen,[a,d,e] Suraphan Panyod,[c] Lee-Yan Sheen,[c] Ming-Shiang Wu[a,d]#

[a]Department of Internal Medicine, National Taiwan University College of Medicine, Taipei, Taiwan

[b]Department of Internal Medicine, National Taiwan University Hospital Bei-Hu Branch, Taipei, Taiwan

[c]Institute of Food Science and Technology, National Taiwan University, Taipei, Taiwan

[d]Division of Gastroenterology and Hepatology, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan

[e]Graduate Institute of Clinical Medicine, National Taiwan University College of Medicine, Taipei, Taiwan

#Address correspondence to Ming-Shiang Wu, mingshiang@ntu.edu.tw

Running Title: 16S rRNA V3V4 and V4 Amplicon Compatibility

Word count:

Abstract: 247 words

Main text: 4729 words

ABSTRACT

Sequencing of the 16S rRNA gene by Illumina next-generation sequencing is broadly used in microbiome studies. Different hypervariable regions of the 16S rRNA gene, V3V4 (amplified with primers 341F–805R) or V4 (V4O; primers 515F–806R), are selected, depending on the targeted resolution. However, in population-based clinical studies, combining V3V4 and V4 data from different studies for a meta-analysis is challenging. Reads generated by short-read (150-bp) high-throughput sequencing platforms do not fully recover the V4 region read-length. Here, we evaluated the compatibility of 16S rRNA V3V4 and V4 amplicons for microbiome profiling. We compared taxonomic compositions obtained by the analysis of V3V4 and V4 amplicons, and V4 fragments trimmed from V3V4 amplicons. We also evaluated an alternative V4 region (V4N; primers 519F–798R) designed for efficient stitching with 150-bp paired-end sequencing. First, we simulated a global investigation of environmental prokaryotes *in silico*. This revealed that V4O primers recovered the highest proportion of fragments (81.7%) and most phyla, including archaea. Empirical sequencing of standard (mock) and human fecal samples revealed biased patterns of each primer that were similar to the ones determined by *in silico* simulation. Further, for human fecal microbiome profiling, the between-sample variance was greater than the systematic bias of each primer. The use of trimmed V4 fragments and single-end amplicons resulted in the same systematic bias. In conclusion, paired-end V4O sequencing yielded the most accurate data for both, simulation and mock community sequencing; the V4O amplicons were compatible with trimmed V4 sequences for microbiome profiling.

IMPORTANCE

Next-generation sequencing of the 16S rRNA gene is a commonly used approach for clinical microbiome studies. Different amplicons of the 16S rRNA hypervariable regions are used in different studies, which creates incompatible sequence features when comparing and integrating data among studies by using 16S denoising pipelines. Here we compared the type of data and coverage obtained when different 16S rRNA amplicons were analyzed. *In silico* and empirical analyses of the human fecal microbiome revealed that the V3V4 amplicons are compatible with V4 amplicons after trimming up to the same region. These observations demonstrate that reconciling the compatibility of clinical microbiome data from different studies improve not only the sample size but also the confidence of the hypothesis tested.

INTRODUCTION

The human microbiome affects human health in numerous ways (1-7). Notably, the gut microbiome mediates host metabolic and physiological status, such as digestion, immune response, neuron transmission, and circulation (8). Clinical research of the microbiome is rapidly advancing, propelled by sequencing of the 16S rRNA gene (abbreviated "16S") using the next-generation sequencing (NGS) technology. Although long-read sequencing technologies have matured in recent years, the well-developed analysis pipelines for massive microbiome taxonomic profiling (e.g., QIIME2) are mainly based on the Illumina sequencer systems (9).

Accurate evaluation of the microbiota heavily depends on the primers used (10, 11). Further, lower-level taxonomic resolution bias can arise when non-representative regions are amplified (12). The 16S V3V4 (primers 341F–805R) and V4 (primers 515F–806R) hypervariable regions are most frequently used for human microbiome profiling (6, 13). When an amplicon library is prepared using Nextera XT two-step polymerase chain reaction (PCR), the expected insert sizes from these regions are 465 bp and 291 bp, respectively. (14, 15). The V3V4 and V4 amplicons are fully recovered on the Illumina MiSeq platform, which can be used to sequence up to 600 nucleotides from both ends of an amplicon [(300 bp)×2].

When the National Institute of Health Human Microbiome Project (HMP) started in 2008, investigations of the human microbiota using Roche/454 pyrosequencing focused on the 16S V3V5 region (primers 357F–926R) (16). However, the V3V4 region became the mainstream amplicon target in microbiota studies since Illumina released a recommended library preparation protocol for sequencing on the MiSeq platform (15). Although the outputs of both sequencing approaches were comparable (17), the Illumina platform generated much more reads than pyrosequencing. After HMP, in 2010, the Earth Microbiome Project (EMP) was initiated, as a

global investigation of environmental and host-associated microbiomes (18, 19). The 16S V4 region (primers 515F–806R) was amplified for sequencing in the EMP projects (18), including human-associated microbiome studies (20) and the American Gut Project (21). This allowed for a more representative prokaryotic profiling since the V4 universal primer pair effectively captures both bacterial and archaeal 16S sequences. However, the EMP V4 libraries were constructed using custom sequencing primers according to Caporaso et al. (22), and generated by 150-bp paired-end (PE) sequencing with additional 14 cycles for barcode tagging on MiSeq or HiSeq. The amplicon insert size of the custom V4 library is expected to be 252 bp, while the two-step PCR method generates 291 bp from the same targeted region (14). This means that the Nextera XT two-step PCR method is only suitable for platforms with an over 150-bp PE sequencing capacity (250-bp PE and 300-bp PE). Nonetheless, the V4 region is analyzed regardless of which protocol is followed. Data from different protocols are supposed to be ideally integrated amplicon sequence variants (ASVs) for meta-analysis microbiota profiling by using denoising algorithms (23) [e.g., DADA2 (24), Deblur (25), and UNOISE3 (26, 27)] in the QIIME2 pipeline (28).

The sequencing cost per megabase has exceeded Moore's Law, which describes a trend in doubling of computing power that conceives the improvement of DNA sequencing capacity yearly and even over the exponential trend (29). The Illumina sequencing platforms generate between several tens to hundreds of millions of reads, enabling deep profiling of a large number of samples during a single PE run at a fraction of the cost of a study. The output of more than 100,000 reads per sample is suggested and sufficient for microbiota investigations (15). Reduced sequencing runs are preferred in large sample-size studies, especially clinical cohort studies, to avoid batch effects. Higher throughput sequencers, such as NextSeq and HiSeq, which are

common in academic core laboratories, generate reads with a low batch effect and at a low cost per sample with a single run (30). However, the confidence for stitching both ends to recover the full V4 region after quality trimming of thus generated PE sequences [(150 bp)×2, the maximum read length of NextSeq] is low. Hence, it is difficult for the higher throughput platforms to meet the demands for cost–benefit and data yields.

In response to the changes of amplicon sequencing methods in the clinical gut microbiota research and considering the cost–benefit of sequencing, we here compared several pairs of sequencing and analysis approaches. We first conducted an *in silico* PCR simulation of a global investigation of environmental prokaryotes by capturing the 16S V3V4 and V4 regions, as well as the V4 fragments trimmed from V3V4 from the SILVA 132 ribosomal RNA NR 99 database (DB). We tested primer and analysis bias by analyzing a mock microbial community. We then sequenced human fecal samples and analyzed the data using the QIIME2 pipeline with the DADA2 plugin, which is currently the most accurate sample inference method (denoising). We then evaluated the compatibility (including the taxonomic abundance consistency and coverage) of the amplicons of different 16S hypervariable regions. The analysis revealed primer- and analysis method-associated bias. Based on the findings, we propose optimized analytical options for clinical population-based and meta-analysis studies.

RESULTS

Cost evaluation of 16S amplicon sequencing using the Illumina MiSeq and NextSeq platforms. NGS sequencers are versatile platforms for sequencing-based studies. The cost–benefit ratio (output quantity and quality) of sequencing has to be considered when choosing a suitable platform. Focusing on the 16S amplicon sequencing, we compared the cost effectiveness

of the Illumina mid to mid–high throughput platforms (MiSeq and NextSeq platforms), which the majority of academic core laboratories are commonly equipped with (Table 1).

The maximal flow cell configurations of MiSeq and NextSeq are 600 (300-bp PE) and 300 (150-bp PE) cycles, respectively. Based on the Illumina Nextera XT library preparation method for 16S amplicons, the V3V4 and V4 regions are fully recovered using the MiSeq 600-cycle reagent kits; on the other hand, only the V4 region can be in "theory" recovered using the NextSeq 300-cycle reagent kits. Fixing the output at 100K PE reads per sample, the MiSeq can process 175 samples per run (96 samples per run as the regular configuration), using the default library preparation method (Nextera XT). The mid-output and high-output flow cells of the NextSeq can process up to 384 samples per run (all index pairs), accounting for 39.4% and 12.8% reads of a single run, respectively. The 16S amplicon sequencing costs in US dollars are $15.2 per 100K reads per sample for the MiSeq, decreasing to $3.4 and $2.8 for the NextSeq mid-output and high-output platforms, accordingly.

Since the Illumina NGS base quality decreases toward the 3'-end over the read (Fig. 1A), the overlapping area of the PE sequences should be sufficiently long to allow the assembly by high-quality base pairs (Fig. 1B). However, when the V4 amplicon sequencing is performed via short read-lengths (e.g., 150-bp PE), gaps exist between the two amplicon ends after quality trimming (Fig. 1C). Caporaso et al. (29) modified the V4 library preparation method to make it suitable for use with 150-bp PE sequencing. The modified method generates 252-bp amplicons instead of 291-bp amplicons obtained with the Nextera two-step PCR approach. However, although the modified method increases both, the V4 amplicon assembly efficacy and sequencing capacity (up to 975 and 2168 samples per run on the NextSeq mid-output and high-output platforms), it requires manual alteration of the sequencing software configuration. The

above comparisons revealed that it is necessary to optimize sequencing configurations (i.e., library construction by amplifying suitable hypervariable region and sequencing with long enough configuration) for improving the cost-benefit of 16S amplicon sequencing. We, therefore, further proceeded to test if the shorter (V4) amplicons harbor equivalent or better taxonomic profiling capacities compared to the longer (V3V4) amplicons.

Comparison of taxonomic profiling capacities of the 16S V3V4 and V4 regions by an *in silico* PCR simulation. To evaluate the primer efficacies of 16S V3V4 (primers 341F–805R) and V4 primer pairs (including V4 original, V4O, primers 515F–806R; alternative V4, V4N, primers 519F–798R; tV4O, V4O region trimmed from V3V4 fragment; and tV4N, V4N region trimmed from V3V4 fragment), we simulated PCR capture of the targeted fragments from the DB [SILVA 16S gene database (NR 132 99%)](31, 32). The simulation encompassed an investigation of the primer-dependency of detected bacterial and archaeal profiles because the *in silico* PCR captured all targeted sequences from the global environment.

The DB contains 369,953 representative 16S sequences (Table 2). Our targeted approaches extracted 59.2% to 81.7% of sequences from the DB. Although the V3V4 primers captured the longest fragments, they extracted 77.3% of all DB sequences, while the V4O primers extracted 81.7% of all sequences. The V4N-captured sequences covered most of the V4O region but were shorter by approximately 11 bp. This reduced the capture rate of V4N to 66.2%. The tV4 primers recovered 72.5% (tV4O) and 59.2% (tV4N) of sequences from the DB, and 93.8% (tV4O) and 76.5% (tV4N) of the V3V4 sequences.

Based on the assigned taxonomy analysis, all primers captured similar proportions of the major phyla in *in silico* PCR (Fig. 2A). However, the overall differences from the DB composition ranged from 6.44% to 36.69% of the phyla proportions. The V4O resulted in the

171 least differences (6.44%) from the DB, followed by V3V4 (15.23%) and V4N (30.0%). The
172 trimmed V4 approaches led to datasets that differed by 16.9% (tV4O) and 36.69% (tV4N) from
173 the DB. Both V4O and V4N approaches captured the maximum archaea (5.26% and 3.40%,
174 respectively; Table S1), while the V3V4 approach captured the fewest archaea (0.02%). The two
175 trimmed V4 approaches, designed based on V3V4, did not efficiently capture the archaea (0.02%
176 for tV4O and 0.005% for tV4N).

177 We then compared the accuracy of the classification at each taxonomic level by the
178 different primer approaches (Fig. 2B). Over 90% of V4O-generated sequences were assigned
179 correct taxonomy, followed by the V3V4 and tV4O sequences (approximately 80% accuracy at
180 each level). However, at most 70% and 63% of the V4N and tV4N sequences, respectively, were
181 assigned the correct taxonomy (summarized in Table 3). This indicates that the V4O primer
182 would yield the best taxonomic profiles in a global microbial investigation.

183 **Mock community profiling by using seven different 16S amplicon analytic**
184 **approaches.** To empirically evaluate the compatibility of V3V4 and V4 primers, we sequenced
185 and analyzed a mock microbial community. The mock community represented eight species of
186 bacteria and two species of yeasts (see Materials and Methods) and is an artificial synthesized
187 microbial community that serves as a quantitative standard. We used seven analytical approaches,
188 all in conjunction with the QIIME2 DADA2 denoising pipeline (9, 24)(Fig. 3), i.e., we analyzed
189 PE V3V4 amplicons (V3V4), PE V4O amplicons (V4O), PE V4N amplicons (V4N), single-end
190 V4O amplicons (V4OSE), single-end V4N amplicons (V4NSE), and V4 amplicons trimmed
191 from V3V4 (trimmed V4O, tV4O; and trimmed V4N, tV4N). We also sequenced human fecal
192 samples from 10 volunteers using the same protocols utilized to evaluate the primer
193 compatibilities in real targeted samples.

First, we compared the data for mock community samples (simple sample composition) with human fecal samples (complex sample composition) using the seven analytic approaches (Fig. 3A). The beta-diversity (measured by Bray–Curtis dissimilarity) ordination plot revealed high homogeneity in mock community quantification of the seven approaches. The Bray–Curtis dissimilarity is a quantitative measurement. Accordingly, a zoomed-in view of the beta-diversity ordination for the mock community revealed that the PCR conditions (primers used) reflected the community quantitative composition (Fig. 3B). Specifically, the trimmed V4 data points were close to the V3V4 data point; the PE and SE V4O data points were clustered together; and the PE and SE V4N data points were proximal on the first plot axis.

The relative taxonomic abundances in the theoretical sample, as determined by the seven analytical approaches, were not significantly different (G-test; mean of G=1.31, P=0.99; Fig. 3C). The rates of misclassification (unassigned or assigned to yeasts) were less than 1% in all analyses. The differences in the determined relative abundance and the theoretical composition ranged from 8.14% (V4O) to 15.22% (V4N). Sample composition determined by the V3V4 approach (and its derived V4 approaches) differed by approximately 10% from the theoretical composition. This indicates that the different 16S regions and analytic approaches for profiling a simple composition microbiota, such as a synthesized mock community, are able to present similar quantitative and qualitative results.

Determination of sample heterogeneity and variation of taxa composition after 16S amplification with different primers. We next analyzed human fecal samples to evaluate the complementarity of different 16S amplification analytical approaches. The seven analytical approaches used were the same as those for the mock community analysis. We examined the beta diversity, as determined by the seven approaches (Fig. 4). The beta-diversity profiles for

fecal microbiota data points for the seven approaches reflected the 10 individual sample sources (Fig. 4A). We tested the sample and analytical approach heterogeneity by the analysis of similarities (ANOSIM). The between-sample variance was greater than within-sample variance (R=0.987, P=0.001; Fig. 4A and Fig. 4B); on the other hand, the between-analytical approach variance was not significantly greater than the within-analytical approach variance (R=0.016, P=0.22; Fig. 4C and Fig. 4D). In other words, the determined sample variation was relatively constant, regardless of the primers or analytical approaches used.

We next analyzed three alpha-diversity indices (namely, richness, Shannon entropy, and Simpson index of diversity), as determined by using the seven analytical approaches (Fig. S1). The three indices were not statistically significantly different when the different approaches were used (P>0.05). However, the differences in the richness index were marginally significant (P=0.08) for the seven approaches; this trend was attributed to reduced taxon numbers in samples analyzed using the trimmed V4O and trimmed V4N approaches (Fig. S1A).

Consistency of taxonomic abundances determined by different 16S amplicon analytical approaches. PCR artifacts (over-amplified amplicons and chimeras) interfere with taxonomic quantification and taxonomic structure profiling of microbiome samples. The artifacts are sequences that are amplified in a biased manner during PCR. No universal primer exists for a fully unbiased amplification. In the current study, we used the DADA2 denoising pipeline to reduce the confounding effect of PCR artifacts. We profiled the higher-level taxonomy compositions (the relative phylum abundance) by using stacked barplots (Fig. 5A). The proportion of each phylum was different for different analytical approach used, but the difference was not statistically significant (G-test; mean of G=0.81, P=1). However, the V4N-PE analytical

239 approach under-detected the phylum *Verrucomicrobia* (relative abundance 0.02% vs. 1% of
240 other approaches).

241 Because no theoretical reference exists for the human fecal microbiota composition, we
242 used the V3V4 approach (the longest sequence region; Fig. 5B) and the V4O approach (the most
243 accurate method based on *in silico* simulation and mock community experiments; Fig. 5C) as
244 benchmarks to evaluate the classification accuracy. The taxonomy data obtained by V3V4-
245 derived analyses (trimmed V4O and trimmed V4N) were the closest to those of the V3V4
246 approach above the genus level (Fig. 5B). Compared with the V3V4 data, the accuracy of V4O
247 and V4N (PE and SE) methods above the genus level was 78.9% to 91.3%. On the other hand,
248 compared with the V4O data, the taxonomy determined by the V4OSE method was the closest to
249 that of the V4O (PE) method (Fig. 5C). The other methods reached an 83.1% to 93.0% average
250 relative accuracy above the genus level. The rank abundance analysis at the family (Fig. S2A)
251 and genus levels (Fig. S2B) revealed that the taxon abundance was consistent with the V4O-
252 based classification accuracy.

253 We then performed pairwise analysis of taxonomic abundance correlations between the
254 seven analytical approaches (Fig. 6). We plotted the abundance of each taxonomy-assigned ASV
255 using pairwise-correlation scatter plots (Fig. 6, lower left panels) and noted the abundance
256 correlation coefficients (Fig. 6, upper right panels). The correlation coefficients between the
257 V3V4 data and other methods ranged from 0.54 to 0.69. The correlation coefficients between the
258 V4 method data, regardless of whether these were PE, SE, or trimmed methods, and the V4O and
259 V4N data were high (0.92 to 0.94). In addition, for the V4 PE data, we observed the best
260 correlation between the corresponding trimmed V4 methods (0.99 for V4O and tV4O; and 0.98
261 for V4N and tV4N) (summarized in Table 3). A high consistency among V4 approaches

262 presented by integrating both taxonomic assignment and quantification, and the V3V4 amplicons
263 would be compatible with V4 by trimming up to the same hypervariable region.
264
265 **DISCUSSION**
266 The increasing popularity and advances in sequencing technology prompt human and
267 clinical microbiome studies. However, the types of sequencing technology and reagents used
268 limit meaningful comparisons of the obtained data. Here, we compared the most popular
269 microbiome sequencing approaches that rely on the analysis of 16S amplicons (i.e., amplifying
270 different hypervariable regions and coupling with DADA2 sequence variants' denoising process).
271 We evaluated the accuracy and consistency of 16S amplicons, which are captured by V3V4 and
272 V4 regions. Our findings showed that V4O (515F-806R) primer yielded the most consistent,
273 complementary, and accurate taxonomic profile. Additionally, we implemented an integrated
274 approach for the V3V4 and V4 amplicons from different datasets by trimming V3V4 sequences
275 up to the V4 region.
276 The NGS technology and HMP (HMP 1 and iHMP) generate massive sequencing data,
277 promoting clinical microbiome association studies (33). These studies reveal numerous
278 previously unknown host–microbe interactions that underlie various health issues, especially
279 non-communicable diseases, with novel etiology linkages (3-7). For a population-based cohort
280 study in the clinical microbiome research, data throughput and quality, the cost–benefit aspect,
281 and validity of the analytical pipeline should all be considered. The 16S V3V4 and V4
282 hypervariable regions are widely selected for human microbiota profiling, but the fragments
283 amplified from different regions should be coupled with suitable sequencing conditions.
284 Accordingly, we here evaluated the compatibility of different amplicon libraries by *in silico* and

empirical approaches, coupled with a denoising algorithm to detect ASVs. The analysis demonstrated that the current widely used amplicon primers capture over 80% of the taxonomic information. The amplicon sequence length was a critical factor for taxonomic profiling as longer sequences contain more genetic information than shorter sequences (11). The primers were another critical factor, as the amplicons are more representative of the sample complexity when generated by more universal (conserved) primers. For example, while the V3V4 and V4O approaches share nearly identical reverse primers, the V3V4 library was characterized by a higher ASV richness index (ASV numbers; Fig. S1A) but contained less taxonomic information than the V4O library (the archaeal phyla were almost absent in the V3V4 library; Fig. 2A and Table 3).

The taxonomic assignment accuracy is crucial for the subsequent study design, such as strain isolation, identification, and clinical or commercial applications (12). The *in silico* PCR analysis performed herein demonstrated that the V4O taxonomy assignment is more accurate than the V3V4-based assignments. However, even though the V4N-captured sequences overlapped with most regions captured by the V4O method, the taxonomic accuracy (at most 70% accuracy at phylum level) was much lower than that of either V3V4 or V4O method. In practice, the tV4O approach (V4O trimmed from V3V4) was compatible for integrating V3V4-generated data with V4O-generated data (Fig. 6) when testing the human gut microbiome samples, which contain fewer archaea than environmental samples. Since short-read amplicon sequencing only extracts partial information for the 1.5-kb 16S rRNA gene, the taxonomic identification is at most limited at approximately 80% accuracy at the genus level by V3V4 primer amplification and at the species level by V4O primer amplification.

The V4 amplicon sequencing meets economic benefits for microbiome studies. Amplicon sequencing in many studies, not only environmental ecological studies but also human-associated studies, focuses on the V4 region and relies on the EMP protocol. However, the library construction methods are restricted by the maximum read length of a sequencer (30). In other words, the amplified V4 amplicon fragment size (PCR product without the adapter-linking sequences) is expected to be approximately 291 bp if the library is constructed by following the Nextera XT two-step dual index PCR protocol. The two-step dual index PCR method was officially designed for 16S V3V4 library construction, coupled with MiSeq 600-cycle (300 PE) sequencing by Illumina (15). Although the two-step dual index PCR is not exclusive to MiSeq and V3V4 amplicons (34), its sequencing outputs (i.e., the read length and PE overlap) should be precisely calculated based on the quality score distributions from the 5′-end to the 3′-end of the read (35). For example, the PE reads overlap by less than 10 bp when a V4 amplicon library is constructed by using Nextera XT kit and sequenced as 150 PE. Most V4 150 PE reads do not pass the quality filtering and cannot be used to assemble both ends. In addition, the DADA2 algorithm only accepts PE reads with a more than 20-bp overlap as default (24).

Caporaso et al. (36) developed customized sequencing primers based on HiSeq 150 PE. The ensuing customized sequencing procedures entailed modification of the library construction methods (forward primers were directly linked to barcode sequences and an adapter), the sequencing program (using the TruSeq workflow and ignoring an error message from the sequencer system), and increasing the run by additional 14 sequencing cycles. Consequently, the method not only solved the unassembled read problem (yielding 253-bp amplicon fragments) but also lowered the sequencing cost per sample (Table 1). However, even though in some studies, the early stages of sequencing do not follow the EMP or Caporaso protocols, the unassembled

reads can be analyzed by using a single-end pipeline. We here demonstrated that the single-end pipeline was comparable with the PE pipeline, with more than 90% confidence in the taxonomy assignment but a slightly inferior quantification (Pearson's correlation coefficient 0.75) (Table 3).

The ASV denoising methods (23) (e.g., DADA2(24), Deblur(25), and UNOISE3(27)) are replacing the traditional operational taxonomic unit clustering methods for choosing representative amplicon sequences. These denoising methods detect real biological sequence features missed by clustering, and denoised features are specific and reproducible (23). Therefore, we suggest trimming the V3V4 and V4 sequences to the same region to acquire the same representative ASVs when combining different libraries. Although the V3V4 approach under-detected the archaea, the analysis of trimmed V4 amplicons recovered the qualitative and quantitative aspects of the human gut microbiome with over 90% confidence (Table 3 and Fig. 6).

In conclusion, in the current study, we evaluated the compatibility of 16S V3V4 and V4 amplicons typically analyzed in clinical microbiome profiling studies by using a sequence variant-denoising pipeline. Our findings suggest that: (1) the analysis of the PE V4O amplicon (amplified using primers 515F–806R) results in the most accurate taxonomic assignment; (2) the V3V4 amplicon analysis is compatible with the V4 amplicon analysis after trimming to the same region; and (3) while mid-high throughput sequencers reduce the cost of sequencing per sample, only a customized V4 library is suitable for stitching PE reads for subsequent analyses (36). The findings are empirical and analytical suggestions for cost-effective population-based or meta-analysis clinical microbiome studies.

MATERIALS AND METHODS

Ethics statement and sample collection. The studies involving human fecal sample collection and informed consent from human participants were approved by the Institutional Review Board of National Taiwan University Hospital, Taipei, Taiwan (201606045RINB). Fecal samples from 10 healthy volunteers were collected during February 2017 at the National Taiwan University, as described by Wu et al. (37).

Primer selection and alternative V4 primer design. Two sets of 16S amplicon primers, which targeted the V3V4 (primers 341F–805R) and V4 (primers 515F–806R; V4O) regions, were selected from the Illumina-recommend (38) and EMP protocols (18), respectively. An alternative pair of V4 region-specific primers (V4N) consisted of modified primers of EMP V4 and Ghyselinck et al. (39): 519F, 5′-CAGCMGCCGCGGTAAT-3′, and 798R, 5′-GGGTWTCTAATCCKGTT-3′. The expected PCR product length was 279 bp. The V4N primers were synthesized with overhang adapters for index attachment and Illumina sequencing adapters(15). The coverage rate of each primer pair was evaluated by using *in silico* PCR simulation (see below).

***In silico* PCR simulation.** The read and taxonomy coverage rates of the SILVA 16S gene database (NR 132 99%) (31, 32) were evaluated by *in silico* PCR. The simulation was conducted by extracting the expected PCR fragments generated by amplification using the V3V4, V4O, and V4N primers. The *in silico* PCR pipeline was set-up using a UNIX shell script, as follows: (1) create a list of degenerate primer pairs, and link the forward primer to the reverse primer with ".*" from the 5′-end to the 3′-end (e.g., CCTACGGGAGGCAGCAG.*GGATTAGATACCCCAGTAGTC); (2) count the fragments extracted from the database fasta file by using UNIX command "*grep"* with the parameter *-c*; (3) obtain the targeted sequence fragments using the parameter *-o*; (4) extract the sequence ID using

the parameter *-B 1*; (5) count the length of *in silico* PCR products using the *awk '{ print length }'* UNIX command.

Sequencing library preparation for mock community and human fecal microbiomes. *DNA extraction.* Genomic DNA from a mock community standard (ZymoBIOMICS Microbial Community Standard, catalog no. D6300, ZYMO RESEARCH, CA, USA) and from stool samples from 10 volunteers was extracted using QIAamp® PowerFecal® DNA Kit (QIAGEN, catalog no. 12830–50; Hilden, Germany). The genomic DNA was stored at m –20°C until PCR amplification and amplicon sequencing.

Amplification and NGS sequencing. Two-step PCR was performed, following the Illumina protocol for 16S metagenomic sequencing library preparation. PCR was first performed to capture the 16S V3V4 (primers 341F–805R) and V4 hypervariable regions (primers 515F–806R and 519F–798R). Three libraries were then constructed by index PCR using the Nextera XT duel-Index PCR primers (15). The pooled libraries were PE-sequenced in the same run using the Illumina MiSeq reagent kit version 3 (San Diego, CA, USA) for 600 cycles at the Medical Microbiota Center of the First Core Laboratory, National Taiwan University College of Medicine.

Bioinformatic analysis for microbial taxonomic profiling. *Sequence denoising using DADA2 and the QIIME 2 pipeline.* The sequences were processed by using QIIME 2 pipeline (version 2019.10) (9). The primer sequences were trimmed from the raw reads in the three libraries by using the *cutadapt* plugin. The trimmed single-end or PE sequences were subsequently denoised using the *DADA2* plugin in QIIME2. To obtain qualified ASVs, the reads were truncated from the 3′-end based on the quality score distribution to the following read length: (1) V3V4-forward, 270 bp, and V3V4-reverse, 210 bp; (2) V4O-forward, 131 bp, and

V4O-reverse, 130 bp; and (3) V4N-forward, 134 bp, and V4N-reverse, 133 bp. In addition, theV3V4 reads were trimmed to the V4O and V4N read-length for further comparisons. High-confidence ASVs were then obtained by denoising, with quality filtering and chimera removal. The taxonomy was assigned using a naïve Bayes classifier trained on the SILVA 132 99% full-length 16S rRNA gene sequence database (31, 32).

Microbial biodiversity and statistical analyses. All statistical analyses were conducted with R version 3.6.1(40). Microbial community analyses were performed using the *vegan* R package (41). Alpha diversity indices, including the Shannon index and Simpson index, were determined by using the "diversity" function; the Richness index was calculated by using the "specnumber" function. The beta diversity was determined based on the Bray–Curtis dissimilarity and visualized by principal coordinates analysis. ANOSIM was used to test the heterogeneity among individuals, controlling for the different primers. All univariate analyses were conducted by the Kruskal–Wallis test with α=0.05 cut-off for significance and Dunn's test for post-hoc comparisons. Multiple-testing P-values were adjusted based on the false discovery rate by using the "p.adjust" function in R. The likelihood-ratio test (G-test with α=0.05 cut-off for significance) for the abundance profiles was performed by using the *RVAideMemoire* R package.

Data availability. Sequences generated in the course of the current study have been deposited in the Sequence Read Archive (SRA) database under the accession number PRJNA643648.

ACKNOWLEDGMENTS

We thank the research participants and research assistants from the Institute of Food Science and Technology, National Taiwan University, Taipei, Taiwan (Guan-Ling Ou) and National Taiwan University College of Medicine, Taipei, Taiwan (Yu-Tang Yang and Fang-Wei Kuo). We would also like to acknowledge the sequencing service provided by the Medical Microbiota Center of the First Core Laboratory, National Taiwan University College of Medicine, the computational resource support by Prof. Alex Hon-Tsen Yu at the Department of Life Science, National Taiwan University, and technical consulting by An-Chi Cheng at the University of Florida, Gainesville, FL, USA.

We declare that we have no competing interests.

P.Y.L. and W.K.W. conceived and planned the project. W.K.W., C.C.C., and S.P. were involved in sample collection, processing, and storage, and supervised the experiments. P.Y.L. conducted all bioinformatic and statistical analyses. P.Y.L. and W.K.W. were involved in data interpretation and manuscript planning. P.Y.L. drafted the manuscript. L.Y.S. and M.S.W. supervised the study. All authors approved submission of the final version.

REFERENCES

1. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. 2006. Microbial ecology: human gut microbes associated with obesity. Nature 444:1022-3.
2. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 444:1027-31.
3. Zhu W, Gregory JC, Org E, Buffa JA, Gupta N, Wang Z, Li L, Fu X, Wu Y, Mehrabian M, Sartor RB, McIntyre TM, Silverstein RL, Tang WHW, DiDonato JA, Brown JM, Lusis AJ, Hazen SL. 2016. Gut Microbial Metabolite TMAO Enhances Platelet Hyperreactivity and Thrombosis Risk. Cell 165:111-124.
4. Sampson TR, Debelius JW, Thron T, Janssen S, Shastri GG, Ilhan ZE, Challis C, Schretter CE, Rocha S, Gradinaru V, Chesselet MF, Keshavarzian A, Shannon KM, Krajmalnik-Brown R, Wittung-Stafshede P, Knight R, Mazmanian SK. 2016. Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson's Disease. Cell 167:1469-1480 e12.

5. Brennan CA, Garrett WS. 2016. Gut Microbiota, Inflammation, and Colorectal Cancer. Annu Rev Microbiol 70:395-411.
6. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, Andrews E, Ajami NJ, Bonham KS, Brislawn CJ, Casero D, Courtney H, Gonzalez A, Graeber TG, Hall AB, Lake K, Landers CJ, Mallick H, Plichta DR, Prasad M, Rahnavard G, Sauk J, Shungin D, Vazquez-Baeza Y, White RA, 3rd, Investigators I, Braun J, Denson LA, Jansson JK, Knight R, Kugathasan S, McGovern DPB, Petrosino JF, Stappenbeck TS, Winter HS, Clish CB, Franzosa EA, Vlamakis H, Xavier RJ, Huttenhower C. 2019. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature 569:655-662.
7. Hoyles L, Fernandez-Real JM, Federici M, Serino M, Abbott J, Charpentier J, Heymes C, Luque JL, Anthony E, Barton RH, Chilloux J, Myridakis A, Martinez-Gili L, Moreno-Navarrete JM, Benhamed F, Azalbert V, Blasco-Baque V, Puig J, Xifra G, Ricart W, Tomlinson C, Woodbridge M, Cardellini M, Davato F, Cardolini I, Porzio O, Gentileschi P, Lopez F, Foufelle F, Butcher SA, Holmes E, Nicholson JK, Postic C, Burcelin R, Dumas ME. 2018. Molecular phenomics and metagenomics of hepatic steatosis in non-diabetic obese women. Nat Med 24:1070-1080.
8. McFall-Ngai M, Hadfield MG, Bosch TC, Carey HV, Domazet-Loso T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF, Hentschel U, King N, Kjelleberg S, Knoll AH, Kremer N, Mazmanian SK, Metcalf JL, Nealson K, Pierce NE, Rawls JF, Reid A, Ruby EG, Rumpho M, Sanders JG, Tautz D, Wernegreen JJ. 2013. Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci U S A 110:3229-36.
9. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodriguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37:852-857.
10. Comeau AM, Douglas GM, Langille MG. 2017. Microbiome Helper: a Custom and Streamlined Workflow for Microbiome Research. mSystems 2.
11. Johnson JS, Spakowicz DJ, Hong BY, Petersen LM, Demkowicz P, Chen L, Leopold SR, Hanson BM, Agresta HO, Gerstein M, Sodergren E, Weinstock GM. 2019. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat Commun 10:5029.
12. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA, Gregory Caporaso J. 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6:90.
13. Wu WK, Chen CC, Liu PY, Panyod S, Liao BY, Chen PC, Kao HL, Kuo HC, Kuo CH, Chiu THT, Chen RA, Chuang HL, Huang YT, Zou HB, Hsu CC, Chang TY, Lin CL, Ho CT, Yu HT, Sheen LY, Wu MS. 2019. Identification of TMAO-producer phenotype and host-diet-gut dysbiosis by carnitine challenge test in human and germ-free mice. Gut 68:1439-1449.

14. 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. Nat Biotechnol 34:942-9.
15. Illumina. 2013. Illumina 16S metagenomic sequencing library preparation (Illumina Technical Note 15044223 Rev. A). http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf. Accessed 2018/10.
16. Human Microbiome Project C. 2012. Structure, function and diversity of the healthy human microbiome. Nature 486:207-14.
17. Sinclair L, Osman OA, Bertilsson S, Eiler A. 2015. Microbial community composition and diversity via 16S rRNA gene amplicons: evaluating the illumina platform. PLoS One 10:e0116955.
18. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, Navas-Molina JA, Janssen S, Kopylova E, Vazquez-Baeza Y, Gonzalez A, Morton JT, Mirarab S, Zech Xu Z, Jiang L, Haroon MF, Kanbar J, Zhu Q, Jin Song S, Kosciolek T, Bokulich NA, Lefler J, Brislawn CJ, Humphrey G, Owens SM, Hampton-Marcell J, Berg-Lyons D, McKenzie V, Fierer N, Fuhrman JA, Clauset A, Stevens RL, Shade A, Pollard KS, Goodwin KD, Jansson JK, Gilbert JA, Knight R, Earth Microbiome Project C. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551:457-463.
19. Gilbert JA, Jansson JK, Knight R. 2018. Earth Microbiome Project and Global Systems Biology. mSystems 3.
20. Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, Knights D, Gajer P, Ravel J, Fierer N, Gordon JI, Knight R. 2011. Moving pictures of the human microbiome. Genome Biol 12:R50.
21. McDonald D, Hyde E, Debelius JW, Morton JT, Gonzalez A, Ackermann G, Aksenov AA, Behsaz B, Brennan C, Chen Y, DeRight Goldasich L, Dorrestein PC, Dunn RR, Fahimipour AK, Gaffney J, Gilbert JA, Gogul G, Green JL, Hugenholtz P, Humphrey G, Huttenhower C, Jackson MA, Janssen S, Jeste DV, Jiang L, Kelley ST, Knights D, Kosciolek T, Ladau J, Leach J, Marotz C, Meleshko D, Melnik AV, Metcalf JL, Mohimani H, Montassier E, Navas-Molina J, Nguyen TT, Peddada S, Pevzner P, Pollard KS, Rahnavard G, Robbins-Pianka A, Sangwan N, Shorenstein J, Smarr L, Song SJ, Spector T, Swafford AD, Thackray VG, et al. 2018. American Gut: an Open Platform for Citizen Science Microbiome Research. mSystems 3.
22. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci U S A 108 Suppl 1:4516-22.
23. Nearing JT, Douglas GM, Comeau AM, Langille MGI. 2018. Denoising the Denoisers: an independent evaluation of microbiome sequence error-correction approaches. PeerJ 6:e5364.
24. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 13:581-3.
25. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, Knight R. 2017. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. mSystems 2.

26. Edgar RC, Flyvbjerg H. 2015. Error filtering, pair assembly and error correction for next-generation sequencing reads. Bioinformatics 31:3476-82.
27. Edgar RC. 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. bioRxiv doi:10.1101/081257:081257.
28. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope E, Da Silva R, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley G, Janssen S, Jarmusch AK, Jiang L, Kaehler B, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, et al. 2018. QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. PeerJ Preprints 6.
29. Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) www.genome.gov/sequencingcostsdata. Accessed 2020/07/06.
30. Holm JB, Humphrys MS, Robinson CK, Settles ML, Ott S, Fu L, Yang H, Gajer P, He X, McComb E, Gravitt PE, Ghanem KG, Brotman RM, Ravel J. 2019. Ultrahigh-Throughput Multiplexing and Sequencing of >500-Base-Pair Amplicon Regions on the Illumina HiSeq 2500 Platform. mSystems 4.
31. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590-6.
32. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glockner FO. 2014. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Research 42:D643-D648.
33. Integrative HMPRNC. 2014. The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. Cell Host Microbe 16:276-89.
34. Glenn TC, Pierson TW, Bayona-Vasquez NJ, Kieran TJ, Hoffberg SL, Thomas Iv JC, Lefever DE, Finger JW, Gao B, Bian X, Louha S, Kolli RT, Bentley KE, Rushmore J, Wong K, Shaw TI, Rothrock MJ, Jr., McKee AM, Guo TL, Mauricio R, Molina M, Cummings BS, Lash LH, Lu K, Gilbert GS, Hubbell SP, Faircloth BC. 2019. Adapterama II: universal amplicon sequencing on Illumina platforms (TaggiMatrix). PeerJ 7:e7786.
35. Lawrence CB, Solovyev VV. 1994. Assignment of position-specific error probability to primary DNA sequence data. Nucleic Acids Res 22:1272-80.
36. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6:1621-4.
37. Wu WK, Chen CC, Panyod S, Chen RA, Wu MS, Sheen LY, Chang SC. 2019. Optimization of fecal sample processing for microbiome study - The journey from bathroom to bench. J Formos Med Assoc 118:545-555.
38. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glockner FO. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res 41:e1.

586 39. Ghyselinck J, Pfeiffer S, Heylen K, Sessitsch A, De Vos P. 2013. The effect of primer
587 choice and short read sequences on the outcome of 16S rRNA gene based diversity
588 studies. PLoS One 8:e71360.
589 40. R Core Team. 2015. R: A language and environment for statistical computing, R
590 Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/.
591 41. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL,
592 Solymos P, Stevens MHH, Wagner H. 2015. vegan: community ecology package. R
593 package version 2.3-1, Oulu, Finland. https://CRAN.R-project.org/package=vegan.
594

595

596 **TABLE 1** Summary of the estimated 16S amplicon sequencing throughputs and costs for MiSeq and NextSeq platforms

	MiSeq	**NextSeq, mid-output**	**NextSeq, high-output**
Flow cell configuration	600 cycles (300-bp PE)	300 cycles (150-bp PE)	300 cycles (150-bp PE)
Throughput	13.2–15 Gb (44–50M reads)	32.5–39 Gb (216–260M reads)	100–120 Gb (660–800M reads)
Data quality[a]	>70%, >Q30	>75%, >Q30	>75%, >Q30
Targeted 16S hypervariable regions	V3V4, V4	V4	V4
No. samples per run per 100K PE reads (Nextera XT 2-Step PCR)[b]	175 (out of the possible 384 samples)	384 (39.4% read output of a single run)	384 (12.8% read output of a single run)
Maximum no. samples per run per 100K PE reads [Caporaso et al. (36), protocol, 1-Step PCR][c]	175 (out of possible 2168 samples)	975 (out of possible 2168 samples)	2168 (all possible index combinations, 72% read output of a single run)
Cost of sequencing per 100K-read output per sample[d]	US$15.2	US$3.4	US$2.8

[a]Data quality scores (average % of bases over the entire run), as reported by Illumina documents (15).

[b]Quality read outputs were determined as maximum Q30 proportion, 291-bp V4 insert size, and less than 10-bp overlap of 150-bp PE reads. PE: paired-end.

[c]V4 insert size of 252 bp.

[d]The costs (in US dollars) were determined based on the prices cited by the Medical Microbiota Center of the First Core Laboratory, National Taiwan University College of Medicine (Taipei, Taiwan).

604 **TABLE 2** Coverage rates of *in silico* PCR simulation with the V3V4 and V4 primers

	Sequences obtained after *in silico* PCR	**Proportion of all DB sequences[b]**	**Fraction of V3V4 sequences**
SILVA NR132 99 represented sequences	369,953	1.000	1.293
V3V4	286,083	0.773	1.000
V4O	302,383	0.817	1.057
V4N	244,801	0.662	0.856
tV4O[a]	268,308	0.725	0.938
tV4N[a]	218,908	0.592	0.765

605 [a]t in tV4O and tV4N, V4O fragments trimmed from V3V4 fragments.

606 [b]SILVA 132 database (16S rRNA gene clustering at 99%).

607

608 **TABLE 3** Summary of sequencing efficacies (*in silico* and empirical data), considering the coverage rate, taxonomic identification

609 rate, and taxonomic abundance, and depending on the primers and analysis approaches[a]

		V3V4	V4O	V4O SE	V4N	V4N SE	tV4O	tV4N
	Insert size	465 bp	291 bp	Up to 150 bp	279 bp	Up to 150 bp	291 bp	291 bp
	Platform	300 PE 250 PE	300 PE 250 PE	>150 SE	300 PE 250 PE 150 PE	>150 SE	300 PE 250 PE	300 PE 250 PE
***In silico* PCR**	Coverage rate of DB	0.773	0.817	–	0.662	–	0.725	0.592
	Accuracy of phylum classification from DB	84.77%	93.56%	–	70.00%	–	83.10%	63.31%
	Archaea	0.02%	Detectable (>4%)		Detectable (>3%)		0.02%	<0.01%
Empirical data	Phylum detection bias (vs. V3V4)	–	7.23%	6.99%	10.49%	10.36%	0.35%	0.34%
	Accuracy of phylum classification (vs. V3V4)	–	92.80%	93.02%	89.53%	89.65%	99.63%	99.67%
	Accuracy of phylum classification (vs. V4O)	92.80%	–	99.67%	90.32%	90.09%	92.76%	92.53%
	Taxonomic abundance correlation with V3V4	–	0.67	0.54	0.68	0.56	0.68	0.69
	Taxonomic abundance correlation with V4O	0.67	–	0.75	0.92	0.77	0.99	0.93

610 [a]SE for V4O SE and V4N SE, single-end analysis pipeline for V4O primer; t for tV4O and tV4N, V4O fragments trimmed from V3V4 fragments; DB, SILVA

611 132 database (16S rRNA gene clustering at 99%).

Figures

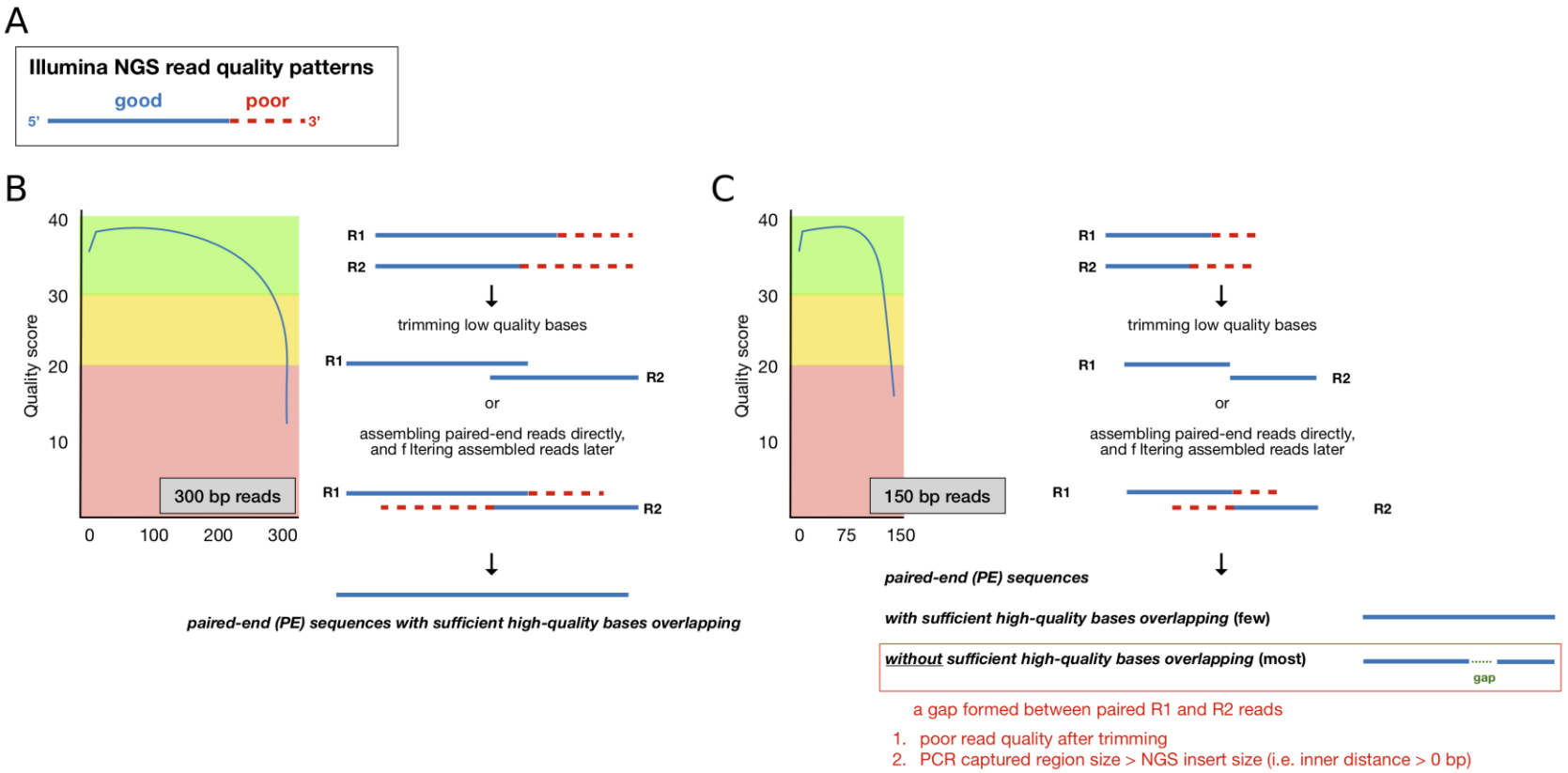


FIG 1 NGS read-quality pattern diagram. (A) Illumina read-quality patterns: from the 5′-end to 3′-end, the quality per base decreases and is usually trimmed (red dashed line) before assembly. (B) The 300-bp paired-end (PE) read-quality distribution and amplicon stitching procedures. The 300-bp PE reads with sufficiently long overlapping regions are stitched together. (C) The 150-bp PE read-

quality distribution and amplicon stitching procedures. Only a few read-pairs pass the quality filtering and assembly at both ends. Most 150-bp PE reads either fail to pass the quality control or fail to recover amplicon inserts that are too long.

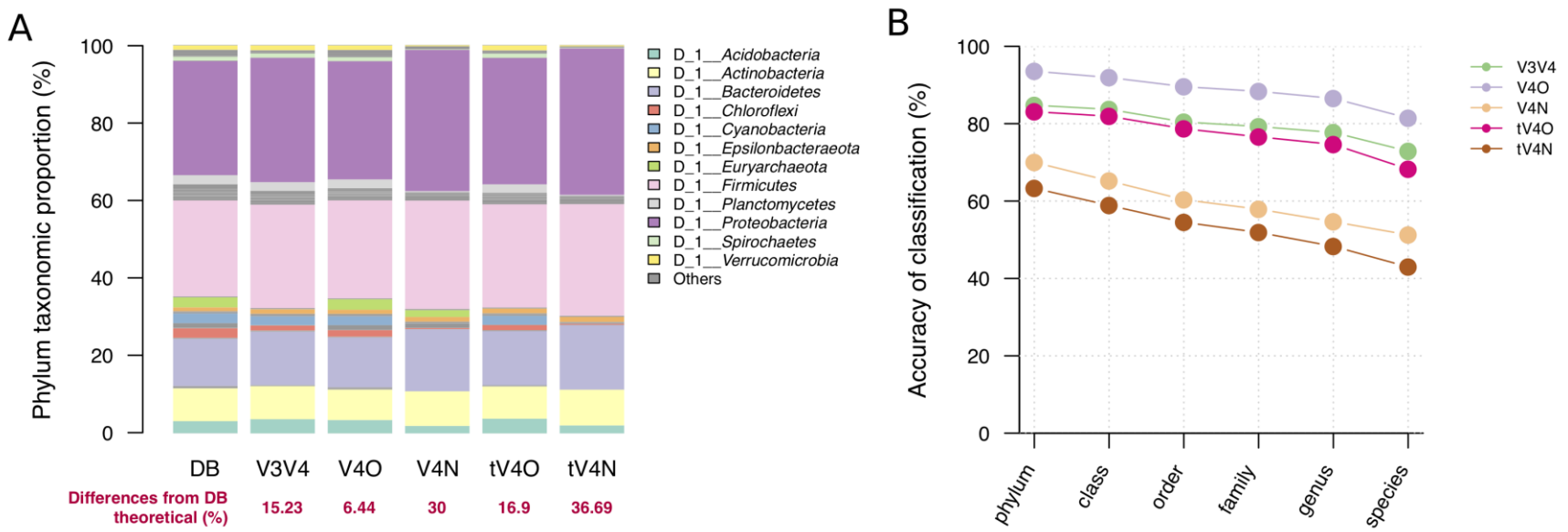


FIG 2 Taxonomic profiling via *in silico* PCR simulation. (A) Phylum composition of the SILVA 16S database and extracted phylum composition for the five tested conditions of *in silico* PCR. (B) Phylum to species taxonomic accuracies (against taxonomic assignment in the SILVA 16S database) for the five conditions of *in silico* PCR. The five conditions for *in silico* PCR were as follows: V3V4 (341F–805R), V4O (515F–806R), V4N (519F–798R), tV4O (trimmed V4O from V3V4), and tV4N (trimmed V4N fromV3V4).

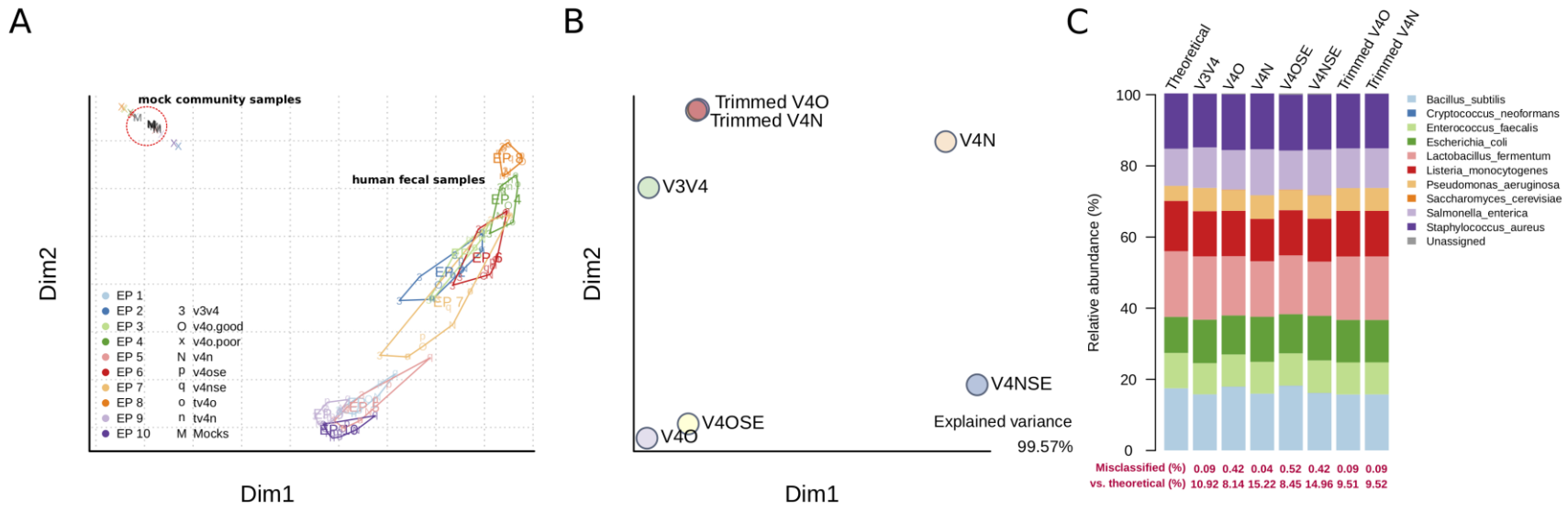
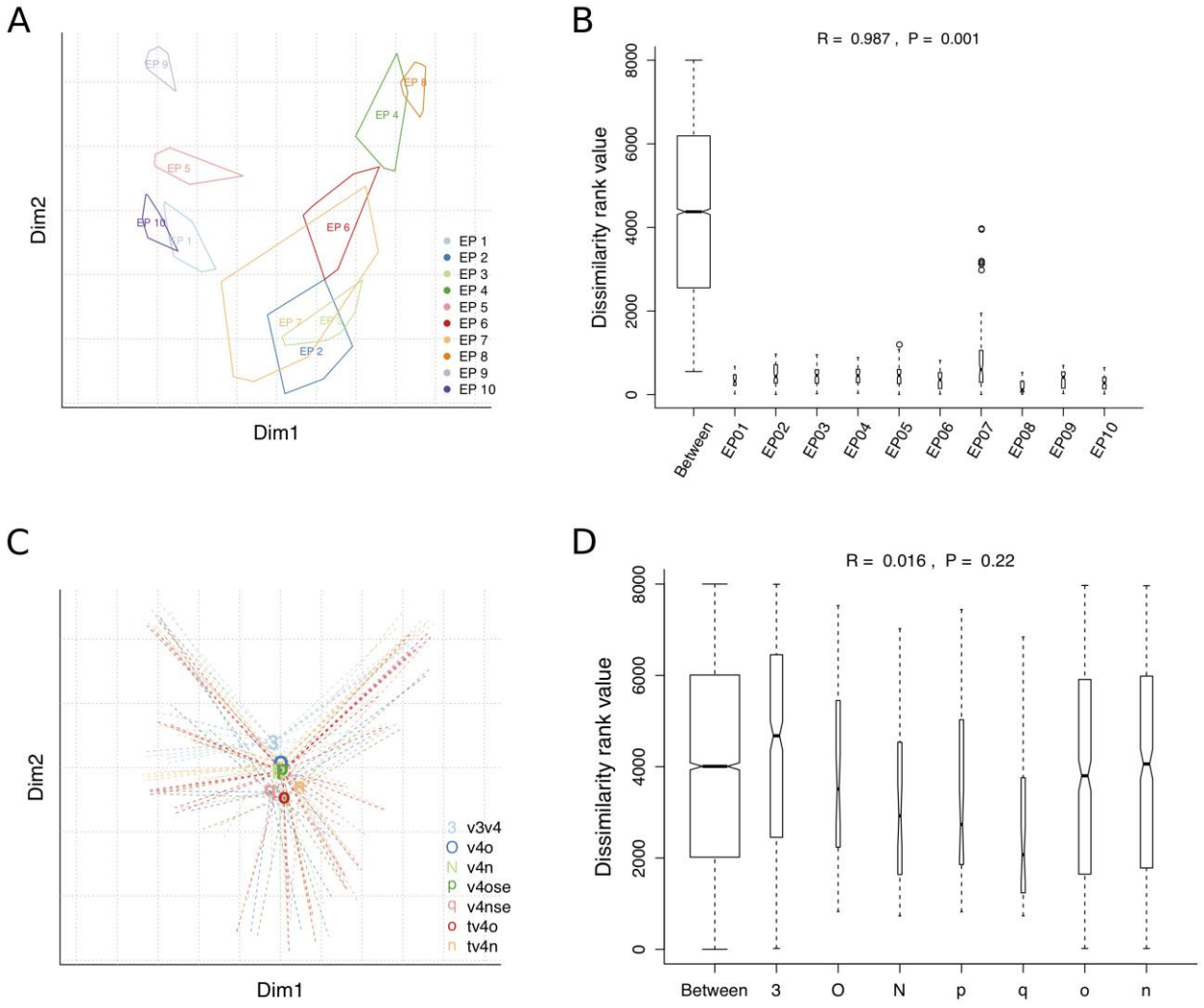


FIG 3 Compatibility evaluation for three primer-amplification and seven analytical approaches, using a mock microbial community sample. (A) Beta-diversity ordination (Bray–Curtis dissimilarity) of the human fecal microbiome (EP1–EP10; colored symbols) and mock community (M; black symbols) sequencing samples. Notation: 3, V3V4 amplicon library; O, successful assembly of V4O amplicon library; x, failed assembly of V4O amplicon library; N, V4N amplicon library; p, V4OSE, single-end analysis using V4O amplicon library; q, V4NSE, single-end analysis using V4N amplicon library; o, tV4O, trimmed V4O amplicon from V3V4 library; n, tV4N, trimmed V4N amplicon from V3V4 library. The mock community sample analyzed by seven analytical approaches clustered together and away from data for 10 individual human fecal microbiomes. (B) Beta-diversity ordination of the seven analytical approaches used for the analysis of mock community data. (C) The relative abundance of bacteria in the mock community, comparing the theoretical composition with that obtained by seven analytical approaches.

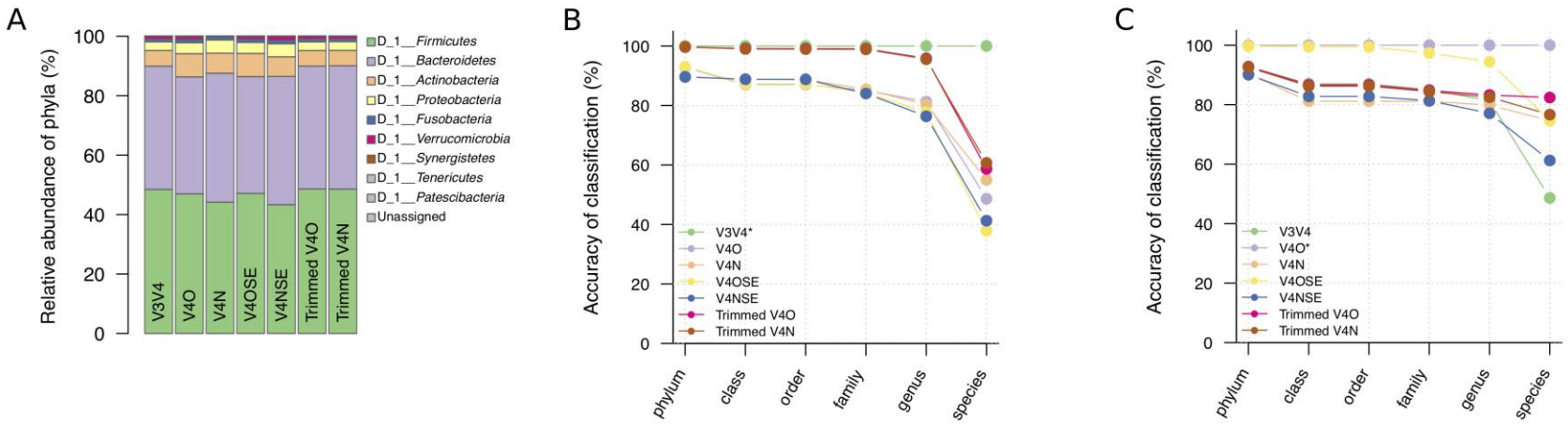
636

637



638

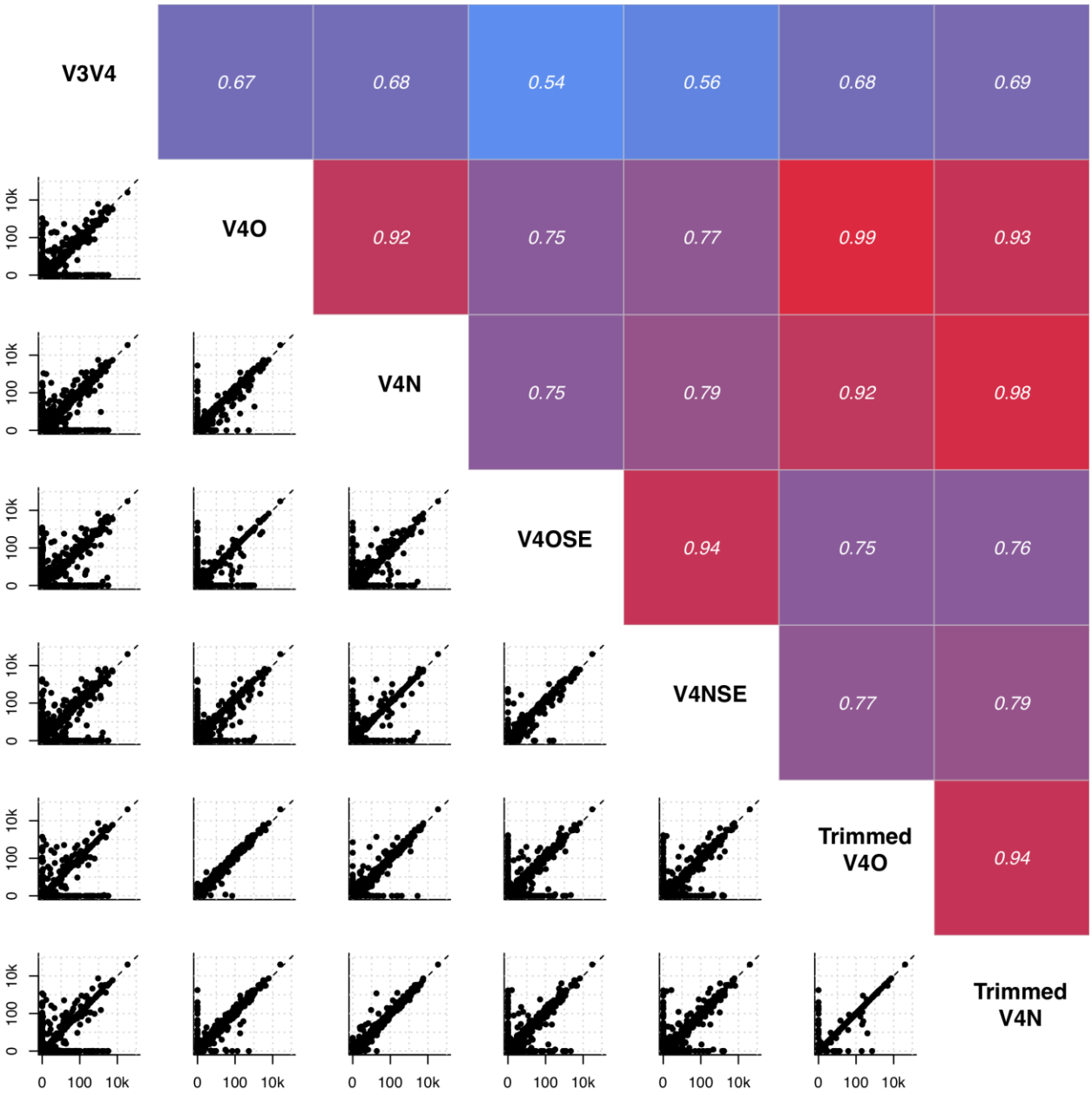
639 **FIG 4** Compatibility evaluation of three primer-amplification and seven analytical approaches
640 by testing the heterogeneity of 10 human fecal microbiomes. (A) Beta-diversity ordination
641 (Bray–Curtis dissimilarity) of 10 human fecal microbiome samples (colored-border polygons,
642 EP1–EP10). (B) Nested analysis of similarity (ANOSIM) of individual testing approaches,
643 nested by the analytical approach; R=0.987, P=0.001. (C) Beta-diversity ordination of 10 human
644 fecal microbiome samples (radial lines; the analytical approaches are labeled in the center). (D)
645 Nested ANOSIM of the analytical approaches nested by individuals; R=0.016, P=0.22.
646 Abbreviations are as in Fig. 3.



647

648 **FIG 5** Phylum compositions and taxonomic accuracy of the human fecal microbiomes analyzed by seven analytical approaches. (A)
649 Phylum compositions of the human fecal microbiome (10 individuals average per stacked bar) analyzed by seven analytical
650 approaches. (B) Phylum to species taxonomic accuracies based on the V3V4 taxonomy assignment. (C) Phylum to species taxonomic
651 accuracies based on the V4O taxonomy assignment. Abbreviations are as in Fig. 3.

652



653

654 **FIG 6** Taxonomic abundance correlations of seven analytical approaches. Lower left panels,

655 pairwise taxonomic abundance correlations. Each point is an amplicon sequence variant shared

656 by any two-analysis sets. Upper right panels, Pearson's correlation coefficients for pairwise

657 taxonomic abundance correlations; all tests were significant. Abbreviations are as in Fig. 3.

658

659

SUPPLEMENTARY FIGURE LEGENDS

FIG S1 Alpha-diversity indices obtained with seven analytical approaches by testing the heterogeneity of 10 human fecal microbiomes. (A) Richness (amplicon sequence variant number). (B) Shannon entropy. (C) Simpson index of diversity. All indices were tested by Kruskal–Wallis rank sum test with alpha = 0.05. The seven analytical approaches were as follows: paired-end V3V4 amplicons (V3V4), paired-end V4O amplicons (V4O), paired-end V4N amplicons (V4N), single-end V4O amplicons (V4OSE), single-end V4N amplicons (V4NSE), and V4 amplicons trimmed from V3V4 (trimmed V4O, tV4O; and trimmed V4N, tV4N).

FIG S2 Taxonomic rank abundances obtained with seven analytical approaches by testing the heterogeneity of 10 human fecal microbiomes. (A) Family rank abundance. Bar fill color corresponds to the phylum. (B) Genus rank abundance. Abbreviations are as in Fig. S1.

Supplementary Figures

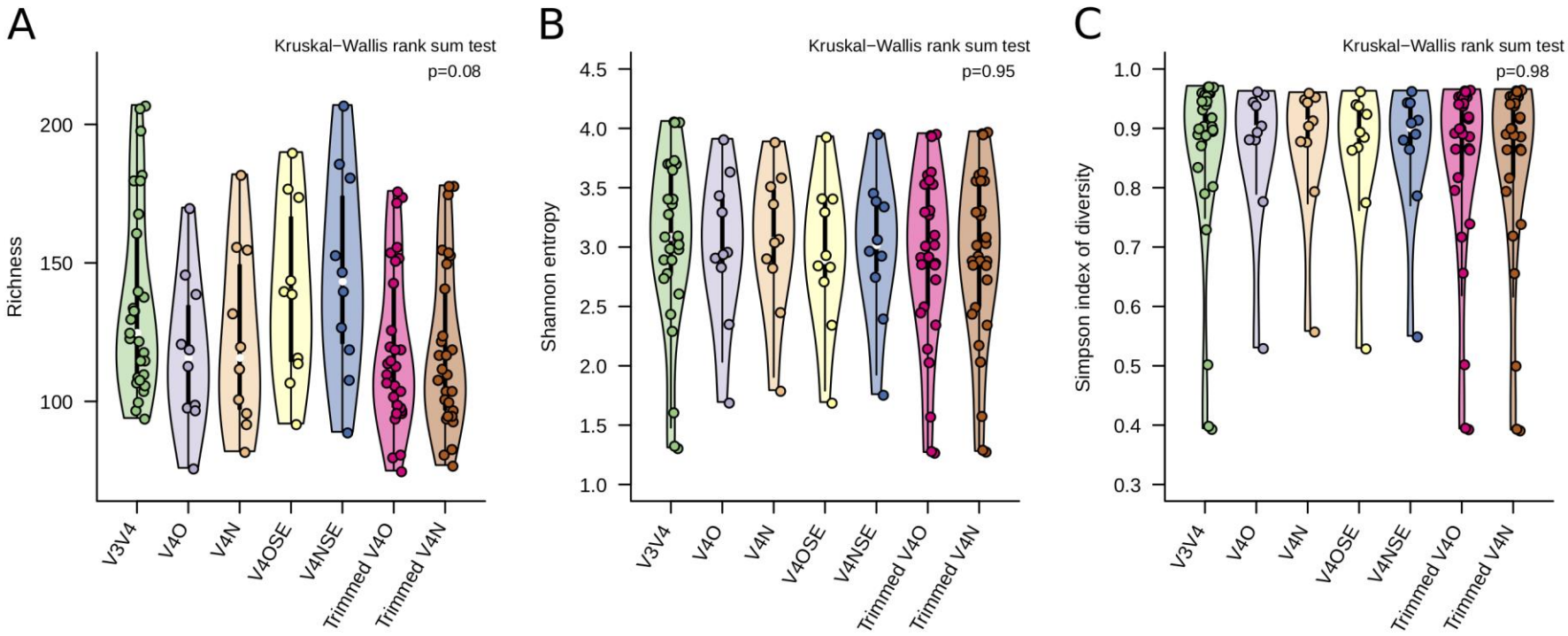


FIG S1

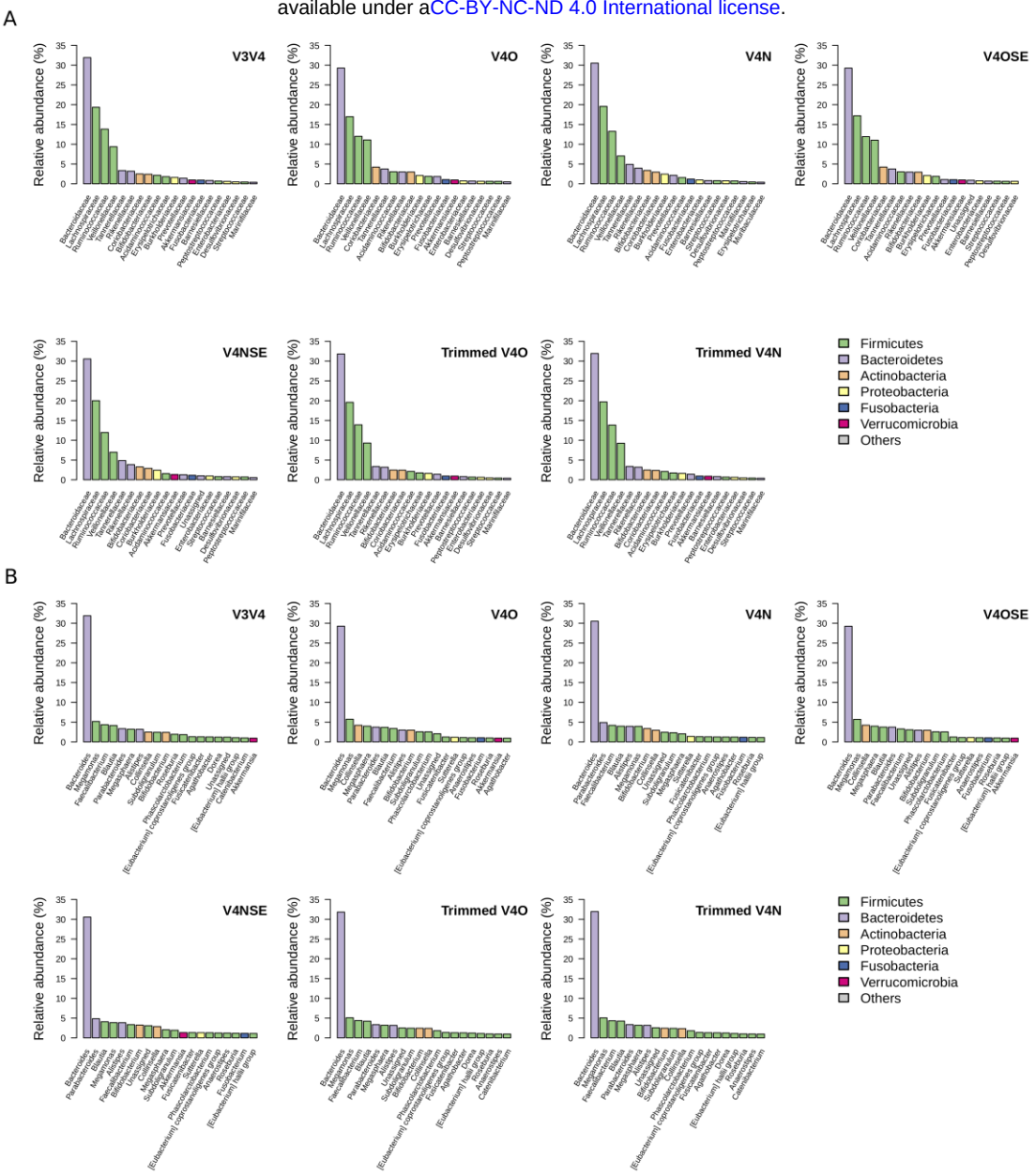


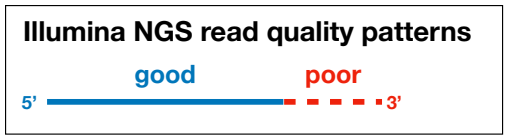
FIG S2

Supplementary Table

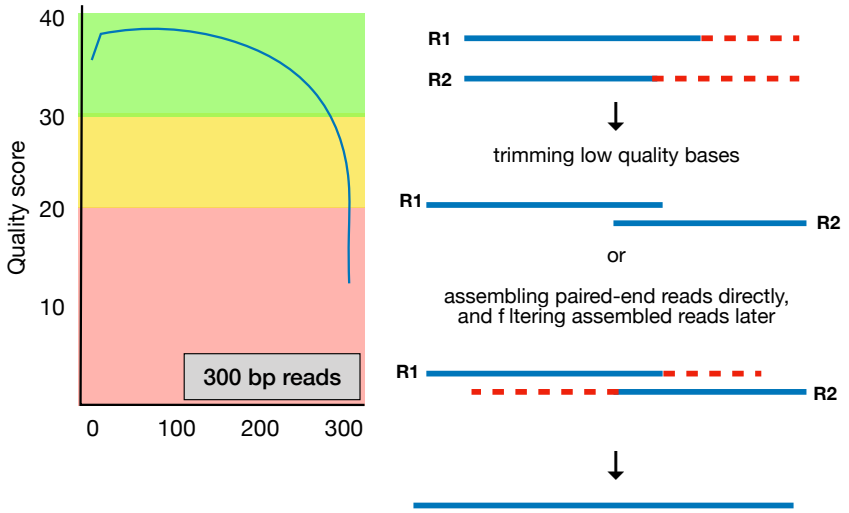
TABLE S1 *In-silico* archaeal sequence capture rates of the V3V4 and V4 primers

	Number of archaeal sequences	**Percentage of captured archaea within dataset**
SILVA NR132 99 represented sequences	19,501	5.27%
V3V4	69	0.024%
V4O	15,916	5.26%
V4N	8320	3.40%
tV4O	65	0.024%
tV4N	11	0.005%

A

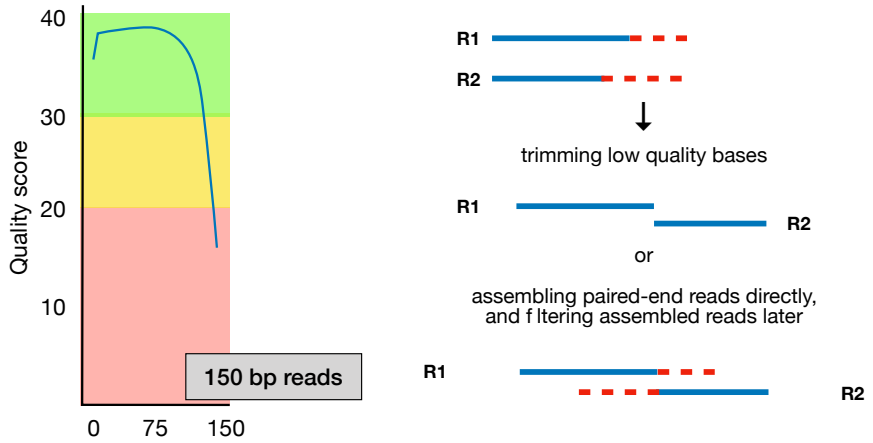


B



paired-end (PE) sequences with sufficient high-quality bases overlapping

C



paired-end (PE) sequences

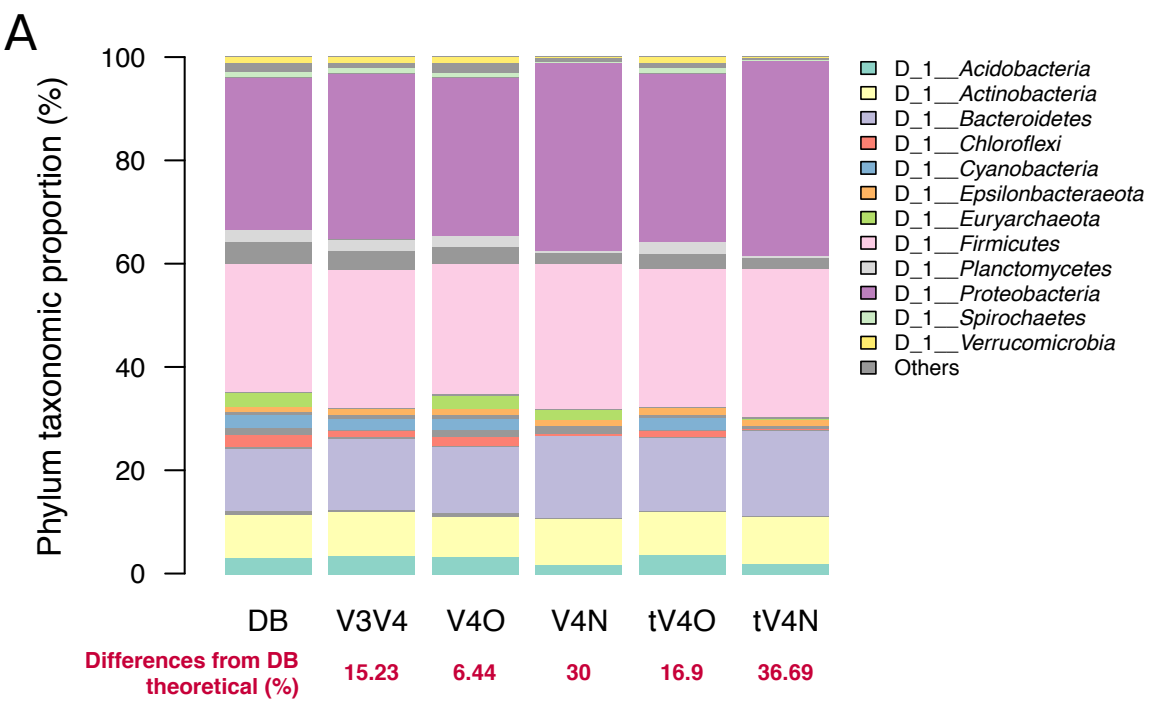
with sufficient high-quality bases overlapping **(few)**

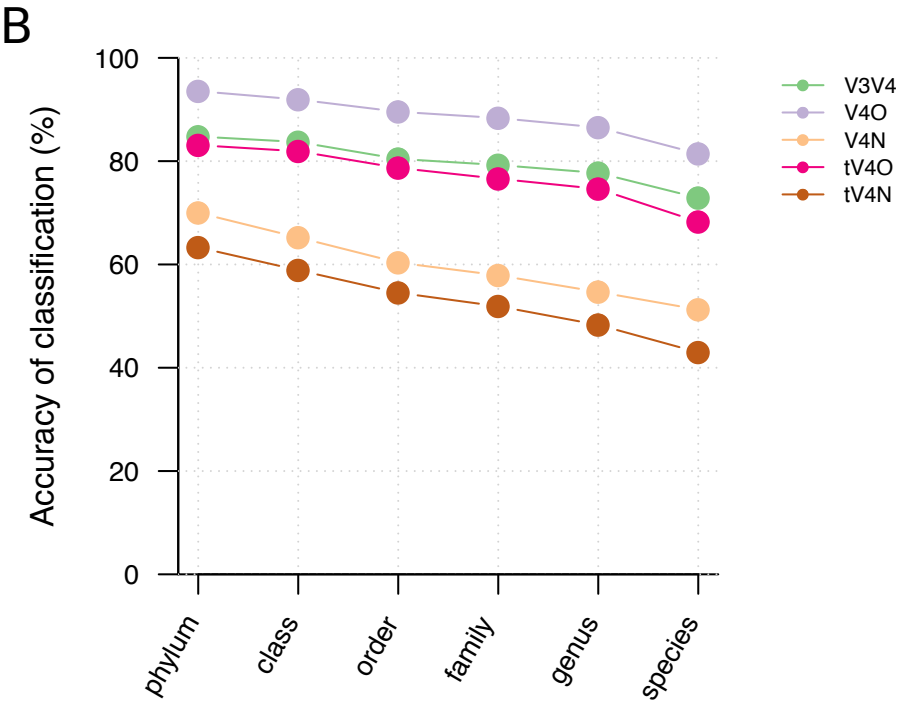
<u>without</u> sufficient high-quality bases overlapping **(most)**

gap

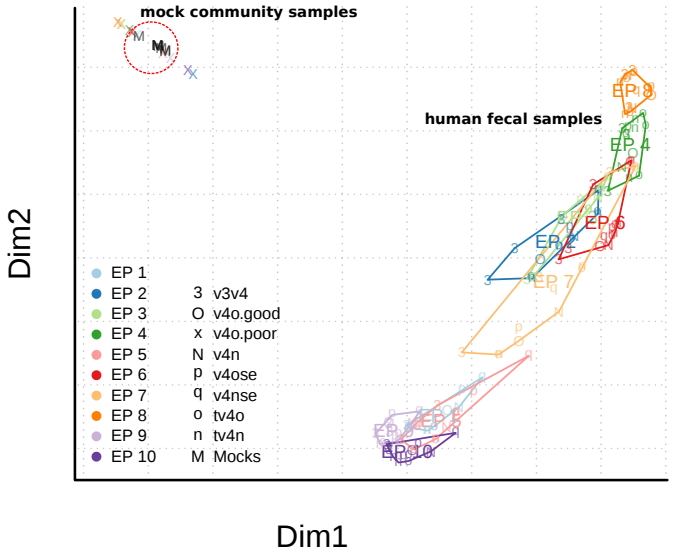
a gap formed between paired R1 and R2 reads

1. poor read quality after trimming
2. PCR captured region size > NGS insert size (i.e. inner distance > 0 bp)

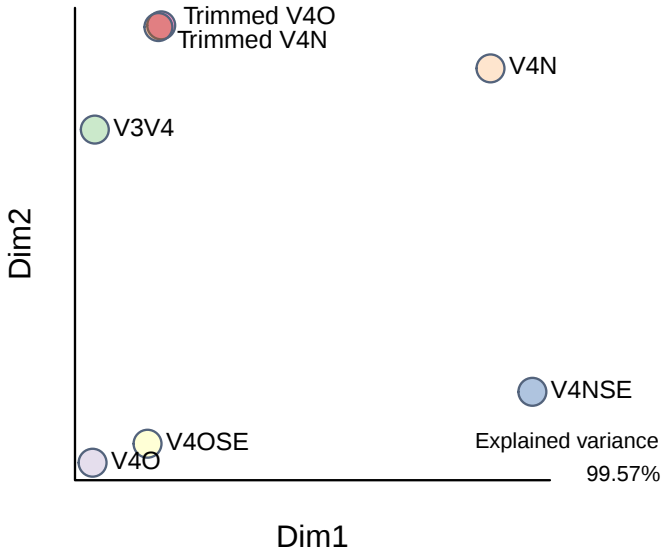

A
Phylum taxonomic proportion (%)
100
80
60
40
20
0
DB
V3V4
V4O
V4N
tV4O
tV4N
Differences from DB theoretical (%)
15.23
6.44
30
16.9
36.69
D_1__Acidobacteria
D_1__Actinobacteria
D_1__Bacteroidetes
D_1__Chloroflexi
D_1__Cyanobacteria
D_1__Epsilonbacteraeota
D_1__Euryarchaeota
D_1__Firmicutes
D_1__Planctomycetes
D_1__Proteobacteria
D_1__Spirochaetes
D_1__Verrucomicrobia
Others


B
Accuracy of classification (%)
100
80
60
40
20
0
phylum
class
order
family
genus
species
V3V4
V4O
V4N
tV4O
tV4N

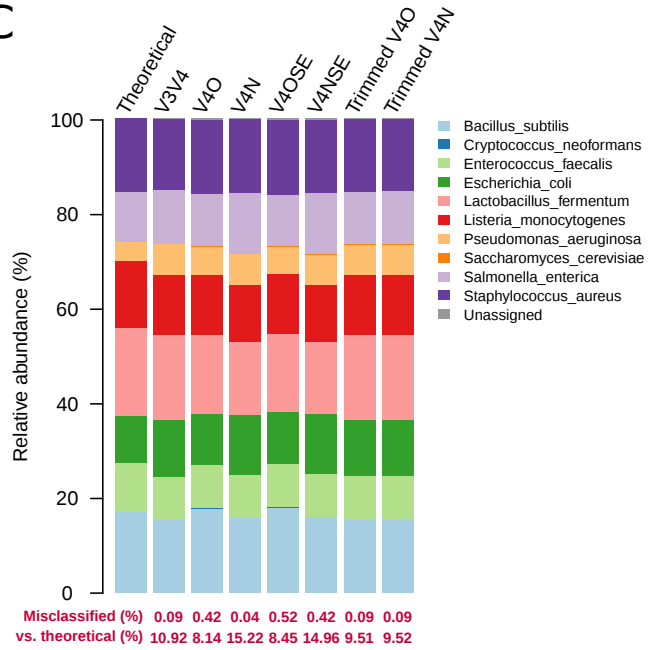
A

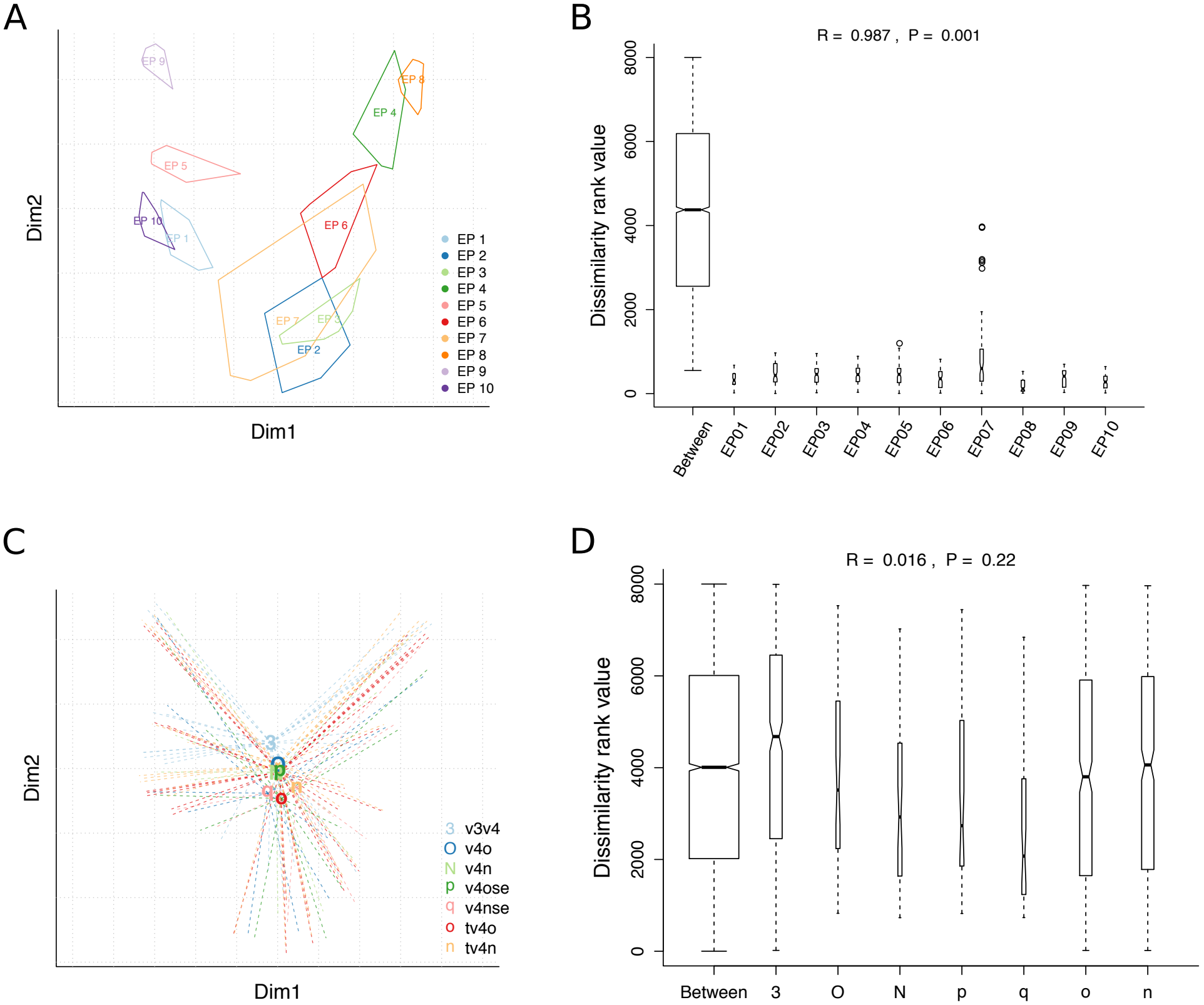

mock community samples
human fecal samples
EP 1
EP 2
EP 3
EP 4
EP 5
EP 6
EP 7
EP 8
EP 9
EP 10
3 v3v4
O v4o.good
x v4o.poor
N v4n
p v4ose
q v4nse
o tv4o
n tv4n
M Mocks
Dim2
Dim1

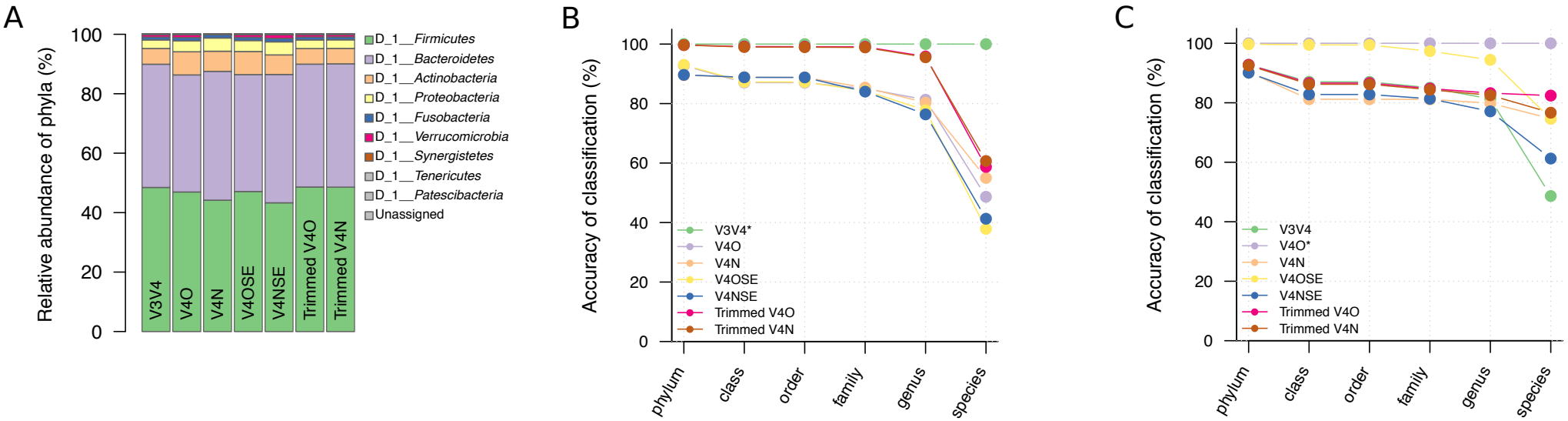
B


Trimmed V4O
Trimmed V4N
V4N
V3V4
V4NSE
V4OSE
V4O
Explained variance
99.57%
Dim2
Dim1

C


Theoretical
V3V4
V4O
V4N
V4OSE
V4NSE
Trimmed V4O
Trimmed V4N
Relative abundance (%)
100
80
60
40
20
0
Bacillus_subtilis
Cryptococcus_neoformans
Enterococcus_faecalis
Escherichia_coli
Lactobacillus_fermentum
Listeria_monocytogenes
Pseudomonas_aeruginosa
Saccharomyces_cerevisiae
Salmonella_enterica
Staphylococcus_aureus
Unassigned
Misclassified (%) 0.09 0.42 0.04 0.52 0.42 0.09 0.09
vs. theoretical (%) 10.92 8.14 15.22 8.45 14.96 9.51 9.52


A
Dim2
Dim1
EP 9
EP 8
EP 4
EP 5
EP 10
EP 1
EP 6
EP 7
EP 3
EP 2
EP 1
EP 2
EP 3
EP 4
EP 5
EP 6
EP 7
EP 8
EP 9
EP 10
B
R = 0.987 , P = 0.001
Dissimilarity rank value
0
2000
4000
6000
8000
Between
EP01
EP02
EP03
EP04
EP05
EP06
EP07
EP08
EP09
EP10
C
Dim2
Dim1
3
O
N
p
q
o
n
3 v3v4
O v4o
N v4n
p v4ose
q v4nse
o tv4o
n tv4n
D
R = 0.016 , P = 0.22
Dissimilarity rank value
0
2000
4000
6000
8000
Between
3
O
N
p
q
o
n


A
Relative abundance of phyla (%)
100
80
60
40
20
0
V3V4
V4O
V4N
V4OSE
V4NSE
Trimmed V4O
Trimmed V4N
D_1__Firmicutes
D_1__Bacteroidetes
D_1__Actinobacteria
D_1__Proteobacteria
D_1__Fusobacteria
D_1__Verrucomicrobia
D_1__Synergistetes
D_1__Tenericutes
D_1__Patescibacteria
Unassigned
B
Accuracy of classification (%)
100
80
60
40
20
0
V3V4*
V4O
V4N
V4OSE
V4NSE
Trimmed V4O
Trimmed V4N
phylum
class
order
family
genus
species
C
Accuracy of classification (%)
100
80
60
40
20
0
V3V4
V4O*
V4N
V4OSE
V4NSE
Trimmed V4O
Trimmed V4N
phylum
class
order
family
genus
species