





Characterization and Demonstration of Mock Communities as Control Reagents for Accurate Human Microbiome Community Measurements

 Dieter M. Tourlousse,^a Koji Narita,^{b,c} Takamasa Miura,^d Akiko Ohashi,^a Masami Matsuda,^a Yoshifumi Ohyama,^d Mamiko Shimamura,^d Masataka Furukawa,^d Ken Kasahara,^{b,c} Keishi Kameyama,^{b,e} Sakae Saito,^f Maki Goto,^f Ritsuko Shimizu,^f Riko Mishima,^g Jiro Nakayama,^{d,g} Koji Hosomi,^h Jun Kunisawa,^h Jun Terauchi,^{b,i}  Yuji Sekiguchi,^a Hiroko Kawasaki^d

^aBiomedical Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan

^bJapan Microbiome Consortium (JMBC), Osaka, Japan

^cChitose Laboratory Corp., Kawasaki, Kanagawa, Japan

^dBiological Resource Center, National Institute of Technology and Evaluation (NITE), Kisarazu, Chiba, Japan

^eInstitute of Food Sciences and Technologies, Ajinomoto Co., Inc., Kawasaki, Kanagawa, Japan

^fTohoku Medical Megabank Organization, Tohoku University, Aoba-ku, Sendai, Miyagi, Japan

^gDepartment of Bioscience and Biotechnology, Faculty of Agriculture, Graduate School, Kyushu University, Nishi-ku, Fukuoka, Japan

^hLaboratory of Vaccine Materials, Center for Vaccine and Adjuvant Research, and Laboratory of Gut Environmental System, National Institutes of Biomedical Innovation, Health and Nutrition (NIBIOHN), Ibaraki, Osaka, Japan

ⁱOno Pharmaceutical Co., Ltd., Osaka, Japan

ABSTRACT Standardization and quality assurance of microbiome community analysis by high-throughput DNA sequencing require widely accessible and well-characterized reference materials. Here, we report on newly developed DNA and whole-cell mock communities to serve as control reagents for human gut microbiota measurements by shotgun metagenomics and 16S rRNA gene amplicon sequencing. The mock communities were formulated as near-even blends of up to 20 bacterial species prevalent in the human gut, span a wide range of genomic guanine-cytosine (GC) contents, and include multiple strains with Gram-positive type cell walls. Through a collaborative study, we carefully characterized the mock communities by shotgun metagenomics, using previously developed standardized protocols for DNA extraction and sequencing library construction. Further, we validated fitness of the mock communities for revealing technically meaningful differences among protocols for DNA extraction and metagenome/16S rRNA gene amplicon library construction. Finally, we used the mock communities to reveal varying performance of metagenome-based taxonomic profilers and the impact of trimming and filtering of sequencing reads on observed species profiles. The latter showed that aggressive preprocessing of reads may result in substantial GC-dependent bias and should thus be carefully evaluated to minimize unintended effects on species abundances. Taken together, the mock communities are expected to support a myriad of applications that rely on well-characterized control reagents, ranging from evaluation and optimization of methods to assessment of reproducibility in interlaboratory studies and routine quality control.

IMPORTANCE Application of high-throughput DNA sequencing has greatly accelerated human microbiome research and its translation into new therapeutic and diagnostic capabilities. Microbiome community analyses results can, however, vary considerably across studies or laboratories, and establishment of measurement standards to improve accuracy and reproducibility has become a priority. The here-developed mock communities, which are available from the NITE Biological Resource Center (NBRC) at the National Institute of Technology and Evaluation (NITE, Japan), provide well-characterized control reagents that allow users to judge the accuracy of their measurement results. Widespread and consistent adoption of the mock communities will improve reproducibility and comparability

Editor Paul A. Jensen, University of Illinois at Urbana Champaign

Copyright © 2022 Tourlousse et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Yuji Sekiguchi, y.sekiguchi@aist.go.jp, or Hiroko Kawasaki, kawasaki-hiroko@nite.go.jp.

The authors declare no conflict of interest.

Received 19 October 2021

Accepted 1 February 2022

Published 2 March 2022

of microbiome community analyses, thereby supporting and accelerating human microbiome research and development.

KEYWORDS control reagents, human microbiome, metagenomics, standards

In the last decade, research into the human microbiome has greatly benefited from widespread access to high-throughput DNA sequencing as a powerful tool for interrogating complex microbial ecosystems, such as those found in the human gastrointestinal tract. This has enabled large-scale cohort studies that revealed numerous correlations between the taxonomic and functional composition of human-associated microbiota and host health (1, 2). As the mechanistic underpinnings of disease-causing alterations in the human microbiome are now starting to be better understood (3, 4), microbiome science is anticipated to lead to a myriad of new therapeutic and diagnostic applications in coming years (5–7). With this has also come an increased emphasis on quality assurance and standardization of methods for microbiome community analysis (8–10).

Reference materials form the basis of quality assurance and standardization across analytical disciplines. In the microbiome field, reference materials have generally taken the form of defined mixtures of genomic DNA or whole cells from several to a few tens of distinct microorganisms. Such mock communities have been used to characterize protocol-dependent biases and optimize or validate methods (11–15), to assess interlaboratory reproducibility in collaborative studies (16–18), and to evaluate run-to-run variability (19). In addition to mock communities, human whole stool samples have also been employed in several of the above studies and their development as reference materials is ongoing (20, 21).

Mock communities with known composition, in contrast to whole stool samples, are intended to provide a ‘ground truth’ to which measurement results can be compared in order to assess accuracy (11). To ensure fit for purpose, this requires careful characterization of the mock communities with respect to their purity and quantitative composition. Analysis of new mock communities should thus ideally be performed using multiple well-established orthogonal methods, preferably across multiple laboratories. Further, complete genome sequences should be available for all strains in order to facilitate interpretation of mock community measurement results.

Here, we report on the development and evaluation of mock communities intended to serve as control reagents for human microbiome community measurements by high-throughput DNA sequencing. The mock communities consist mainly of bacteria that are prevalent in the human gut and were formulated as near-even mixtures of genomic DNA or whole cells of 20 (DNA mock) or 18 (cell mock) bacterial strains. As part of the development of the mock communities, complete genome sequences were newly generated for 12 strains. We carefully characterized the mock communities by shotgun metagenomics, using previously validated standard operating procedures (SOPs) for DNA extraction and metagenomic library construction (11), across multiple laboratories in a collaborative study. Further, we established their fitness to reveal variability in measurement results generated by different methods and/or by changing laboratories, covering protocols for DNA extraction and metagenome and 16S rRNA gene amplicon sequencing. Finally, we explored use of the mock communities to compare metagenome-based taxonomic profilers and to assess the impact of read trimming and filtering on observed species profiles. The here-described mock communities, which are available from the NITE Biological Resource Center (NBRC) at the National Institute of Technology and Evaluation (NITE, Japan), and demonstration of their utility in typical usage cases are expected to contribute to ongoing efforts to improve consistency of human microbiota measurements through better standardization and quality assurance.

RESULTS AND DISCUSSION

The mock communities consist mostly of bacteria that are prevalent in the human gastrointestinal tract, covering the phyla *Bacteroidetes*, *Actinobacteriota*, *Verrucomicrobiota*, *Firmicutes* and *Proteobacteria* (Table 1 and Fig. 1a). In addition, two species associated with

TABLE 1 Bacterial species included in the mock communities

Species	Culture collection	Nucleotide accession	Genome size (bp)	GC content (%)	16S rRNA genes	Cell wall (Gram-type) ^a	Relative abundance in DNA mock ^b (%)	Relative abundance in cell mock ^c (%)
<i>Bacteroides uniformis</i>	NBRC 113350	AP019724 – AP019728	4,989,532	46.2	4	–	4.7	5.6
<i>Blautia</i> sp.	NBRC 113351	CP084061	6,247,046	46.7	5	+	4.5	5.6
<i>Enterocloster clostridioformis</i>	NBRC 113352	BJLB01000001 – BJLB01000002	5,687,315	48.9	5	+	5.3	5.6
<i>Parabacteroides distasonis</i>	NBRC 113806	AP019729	5,179,960	45.0	7	–	4.8	5.6
<i>Bacillus subtilis</i> subsp. <i>subtilis</i>	NBRC 13719	AP019714 – AP019715	4,295,305	43.3	10	+	5.2	5.6
<i>Streptococcus mutans</i>	NBRC 13955	AP019720	2,018,796	36.9	5	+	6.9	5.6
<i>Pseudomonas putida</i>	NBRC 14164	AP013070	6,156,701	62.3	7	–	3.9	5.6
<i>Lactobacillus delbrueckii</i> subsp. <i>delbrueckii</i>	NBRC 3202	AP019750	1,910,306	50.1	8	+	3.6	5.6
<i>Escherichia coli</i>	NBRC 3301	CP048439 – CP048440	4,755,096	50.8	7	–	5.6	5.6
<i>Flavonifractor plautii</i>	NBRC 113805	CP084007	4,277,038	60.4	3	+	3.7	5.6
<i>Staphylococcus epidermidis</i>	NBRC 113846	CP084008 – CP084011	2,520,735	32.2	6	+	4.8	5.6
<i>Cutibacterium acnes</i> subsp. <i>acnes</i>	NBRC 113869	CP084017	2,560,907	60.0	3	+	5.0	5.6
<i>Bifidobacterium longum</i>	NBRC 114370	CP084012 – CP084013	2,594,022	60.1	5	+	5.7	5.6
<i>Anaerostipes caccae</i>	NBRC 114412	CP084016	3,284,789	44.5	4	±	5.3	5.6
<i>Ruminococcus gnavus</i>	NBRC 114413	CP084014 – CP084015	3,757,469	42.5	5	+	5.6	5.6
<i>Megasphaera massiliensis</i>	NBRC 114414	CP084019	2,610,024	50.6	7	–	4.8	0 ^b
<i>Megamonas funiformis</i>	NBRC 114415	CP084018	2,464,533	31.5	6	–	3.7	0 ^b
<i>Collinsella aerofaciens</i>	NBRC 114504	CP084004 – CP084006	2,278,612	60.3	5	+	6.2	5.6
<i>Bifidobacterium longum</i> subsp. <i>longum</i>	NBRC 114494	CP084020 – CP084022	2,534,372	60.1	4	+	4.7	5.6
<i>Akkermansia muciniphila</i>	NBRC 114322	CP084201 – CP084202	2,788,458	55.7	3	–	6.0	5.6

^aThe symbols +, – and ± indicate strains with Gram-positive, Gram-negative and Gram-variable type cells walls, respectively.

^bRelative abundances represent values assigned during formulation of the mock communities, based on quantification of the total DNA content of individual strains prior to mixing.

^c*M. massiliensis* and *M. funiformis* were excluded from the cell mock community.

human skin microbiota, namely, *Staphylococcus epidermidis* and *Cutibacterium acnes* subsp. *acnes*, were included, along with the GC-rich bacterium *Pseudomonas putida*.

While based on their composition the mock communities are mainly intended for assessment of fecal microbiota measurements, they are expected to be more broadly applicable. Specifically, species capture a wide range of genomic GC contents (from 31.5% for *Megamonas funiformis* NBRC 114415 to 62.3% for *P. putida* NBRC 14164) and include a suite of strains with reported Gram-positive type cell walls (hereafter referred to as Gram-positives). Both genomic GC content and cell wall structure represent two important sources of bias in metagenomic analyses, and the mock communities thus provide appropriate reagents for challenging protocols.

As part of the development of the mock communities, we generated complete genome sequences for 12 strains lacking publicly available genomes (Table S1), by short- and long-read sequencing (see Materials and Methods for details). Next, to formulate the DNA mock community, we determined the total DNA concentration of genomic DNA stocks of each strain by fluorometry. The stocks were then combined to generate a mixture containing nearly equal genome copy numbers of each strain, resulting in a mock community with relative abundances ranging from 3.6% for *Lactobacillus delbrueckii* subsp. *delbrueckii* NBRC 3202 to 6.9% for *Streptococcus mutans* NBRC 13955 (Table 1). A similar approach was followed for the cell mock community, by measuring the total DNA content of whole cell stocks through acid-catalyzed release of adenine from whole cells. This technique was originally devised for determining the efficiency of microbial cell lysis and DNA recovery (22) and we previously demonstrated that it could provide a reliable basis for quantifying whole cell mock communities (11). Cell stocks of individual strains were then combined into a mixture containing equal amounts of total DNA, or cell equivalents assuming a constant genome copy number per cell, for each of the strains. Note that two strains, namely, *Megasphaera massiliensis* NBRC 114414 and *M. funiformis* NBRC 114415, were omitted from the cell mock community, such that each strain had a relative abundance of 5.6% (Table 1).

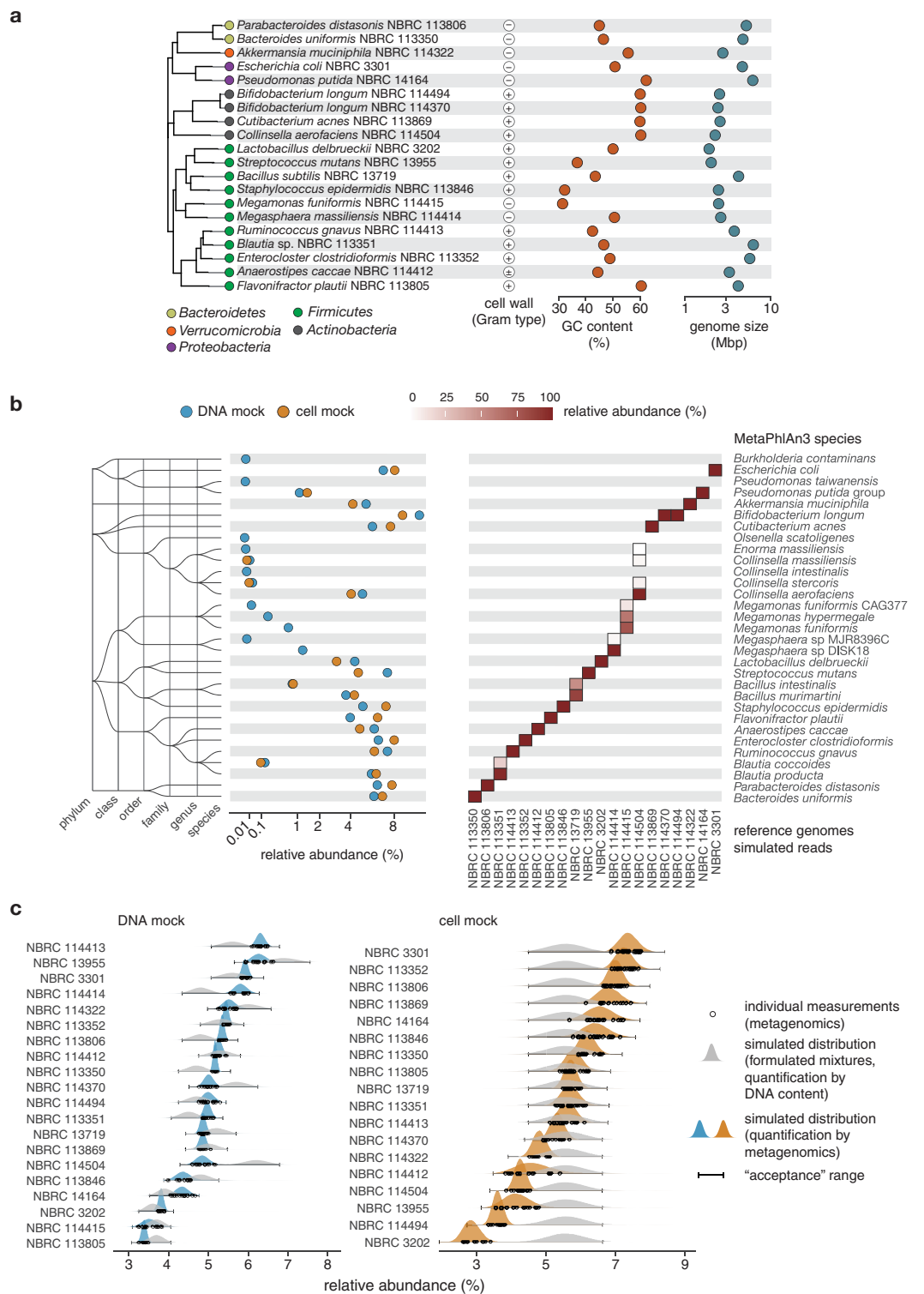


FIG 1 (a) Bacterial species included in the mock communities. The phylogenetic tree was inferred based on single-copy marker genes, using the GTDB-Tk. Phylum-level taxonomic assignments are indicated by colored circles shown at the leaves. Species features, that is Gram type, genomic GC content and genome size, are shown. (b) Characterization of the mock communities by shotgun metagenomics and taxonomic profiling with MetaPhlAn3. Species profiles were generated based on the combined sequencing data of all replicated measurements, performed following SOPs, for both mock communities ($n = 16$ and $n = 20$ for the DNA and cell mock community, respectively, covering multiple aliquots, SOPs and laboratories; Table S3 in the supplemental material). The tree represents a taxonomic dendrogram. The heatmap depicts species profiles of 10 million *in silico* generated reads (5 million 151-bp paired-end reads) for each of the genomes shown on the x axis, with fill colors showing estimated relative abundances as indicated in the legend. (c) Relative abundance of each strain in the mock communities as (Continued on next page)

To characterize the mock communities and establish their fitness for purpose in typical use cases, we set up a collaborative study involving five laboratories (Fig. S1 and Table S2 in the supplemental material). Firstly, to assess purity and quantify strain-wise abundances using a technique orthogonal to total DNA quantification used during formulation (see above), three of the laboratories analyzed the mock communities by shotgun metagenomics, following our previously developed SOPs (11). Secondly, to establish fitness for purpose of the mock communities in revealing differences among protocols, two of the laboratories also evaluated non-SOPs for DNA extraction and metagenomics library construction. Further, all five laboratories performed 16S rRNA gene (V4 hypervariable region) amplicon sequencing, using both a standardized protocol (Illumina's two-step tailed PCR protocol) or alternative custom methods (Table S2).

Characterization of the mock communities by shotgun metagenomics. Based on the collaborative study outlined above, a total of 16 and 20 metagenomics measurements were performed for the DNA and cell mock communities, respectively, following our previously developed SOPs (Table S3 in the supplemental material). In short, the SOP for DNA extraction (denoted as protocol N) uses the ISOSPIN Fecal DNA kit, employing three rounds of bead beating to ensure effective lysis of Gram-positives. For library construction, two SOPs were employed, namely, protocol B (QIAseq FX DNA Library kit) involving enzymatic DNA fragmentation, and protocol K (SMARTer ThruPLEX DNA-Seq kit) using physical DNA fragmentation by focused ultrasonication. Sequencing was performed on an Illumina NextSeq (Lab A, note that libraries prepared by Lab B were sequenced by Lab A as shown in Fig. S1) or HiSeq instrument (Lab C), generating 2×151 and 2×101 bp reads, respectively. Replicates covered three aliquots for each mock community (Lab A), protocol N for DNA extraction (Lab A and B), and protocols B and K for library construction (Lab A, B and C).

To assess purity of the mock communities, we analyzed the sequencing data by MetaPhlAn3 (23), a well-established tool for metagenome profiling based on marker genes. Inspection of individual sequencing libraries showed that species profiles were highly consistent across replicates, without indication of library-specific impurities (Fig. S2 in the supplemental material). The lack of contaminants was further verified through analysis of sequencing data combined across all available replicates for both mock communities, in order to allow detection of potentially lower abundance contaminants (Fig. 1b). More specifically, detected bacteria belonging to genera not part of the mock communities, including *Burkholderia* and *Olsenella*, were estimated to be of very low abundance (less than 0.001% in the combined sequencing reads). For other genera, namely, *Bacillus*, *Megamonas*, *Megasphaera* and *Collinsella*, unexpected species were concluded not to represent contaminants because they were also identified in the corresponding simulated reads, and thus presumed to reflect incorrect identifications by MetaPhlAn3. Taken together, these data verified that the formulated mock communities can, for most practical purposes, be considered as free of contaminants.

To determine the relative abundance of each strain, we used kallisto's quantification workflow involving pseudo-mapping of paired reads to the reference genomes and estimation of read counts for each strain by expectation maximization (24). Firstly, nearly all reads could be assigned to the reference genomes, with mapping rates of, on average, 99% for both mock communities. This verified that the reference genomes were suitable for quantitative analyses. Secondly, variability in strain-wise abundances among aliquots, protocols (two SOPs were used for library construction, see above) and laboratories were small (Fig. S3 in the supplemental material), with a variation around the grand mean of all replicates of 2.7% (qmCV, see Materials and Methods for definition) and 5.7% for the DNA and cell mock community,

FIG 1 Legend (Continued)

assigned based on total DNA content, during formulation, and as measured by shotgun metagenomics. Empty circles shown individual shotgun metagenomics measurement results as mentioned for panel b. Gray and blue/orange density plots show the distribution of simulated strain-wise abundances as indicated in the legend. 'Acceptance' ranges for strain-wise abundances are shown as error bars. Note that strains NBRC 114414 (*M. massiliensis*) and NBRC 114415 (*M. funiformis*) were not included in the cell mock community.

respectively. Thirdly, estimated relative abundances for the different strains were in reasonably good agreement with values assigned during formulation based on total DNA quantification (Fig. 1c). For the DNA mock community, abundances of individual strains deviated 1.1-fold (gmAFD as calculated based on the mean of strain-wise abundances across replicates, see Materials and Methods for definition) from the assigned values. For the cell mock community, differences between assigned and measured values were slightly higher, with a gmAFD of 1.2-fold. Further, Gram-positives in the cell mock community had a total abundance of $63.5 \pm 0.4\%$ (mean and sd across replicates), in good agreement with the expected value of 67.2%.

Following the strategy that we described previously (11, also see Materials and Methods), we next calculated guidance values for allowable errors (that is, the acceptable level of deviation from the expected composition) by considering differences between strain-wise abundances assigned during formulation by total DNA quantification and the above metagenomics measurement results. These values, which are provided in Table S4 in the supplemental material, can be used to judge the accuracy of measurement results. Although a more rigorous and formal appraisal will be warranted toward achieving consensus on the acceptable level of errors (9), setting acceptance values for key performance parameters is a prerequisite for enabling meaningful use of reference materials.

As can be observed in Fig. 1c, agreement between relative abundances assigned during formulation and values measured by metagenomics was lower for a few strains. Here, expected values were calculated based on total DNA concentrations of DNA or cell stocks of each strain prior to mixing. Compared to digital PCR, total DNA measurements were used as they were presumed to be less sensitive to varying genome coverage, e.g., near the origin of replication in growing cells, which may lead to biased quantification depending on the genes targeted by digital PCR. With respect to the cell mock community, we previously reported that mixtures formulated based on total DNA quantification, through measurement of adenine content in whole cells, showed better concordance with metagenome measurements, as compared to mixtures fabricated based on cell counts determined by flow cytometry (11). In addition, DNA content can also account for potential variations in genome ploidy.

After having characterized the mock communities, we next leveraged the collaborative study results to establish their fitness for purpose in two typical use scenarios. Firstly, we evaluated the utility of the mock communities to reveal differences in performance of methods and/or laboratories, covering DNA extraction, shotgun metagenomics and 16S rRNA gene amplicon sequencing. Secondly, we used the shotgun metagenomics data to evaluate different taxonomic profiling tools and to assess the impact of trimming and filtering (that is, pre-processing) of reads on downstream analysis results.

Comparison of protocols and laboratories. As summarized in Table S2 and Fig. S1 in the supplemental material, each of the participating laboratories analyzed the DNA and cell mock communities, performing DNA extraction, library construction and sequencing. Extraction of DNA from the cell mock community was performed either following our previously established SOP (protocol N) or using custom protocols (see Supplementary Methods for details). Libraries for shotgun metagenomics, from both the DNA mock community and extracted DNA for the cell mock community, were generated by three laboratories, following either our SOPs (protocols B and K) or using alternative methods based on commercial kits. Further, all laboratories also generated amplicon libraries of the V4 hypervariable region of the 16S rRNA gene, following either Illumina's standard two-step tailed PCR protocol with KAPA HiFi DNA polymerase, which we here considered as SOP, or using locally preferred protocols using different polymerases (see Supplementary Methods for details, Table S2).

For the shotgun metagenome data, abundances of the strains were determined by kallisto as above. For the 16S rRNA gene amplicon data, reads were trimmed based on quality and filtered, paired reads merged, and the abundance of each strain estimated using USEARCH's annot command, requiring a sequence identity of $\geq 97\%$ to the reference sequences. For comparison with the metagenomics-based species profiles, sequence counts for the amplicon data were corrected for 16S rRNA gene copy number and converted to relative abundances.

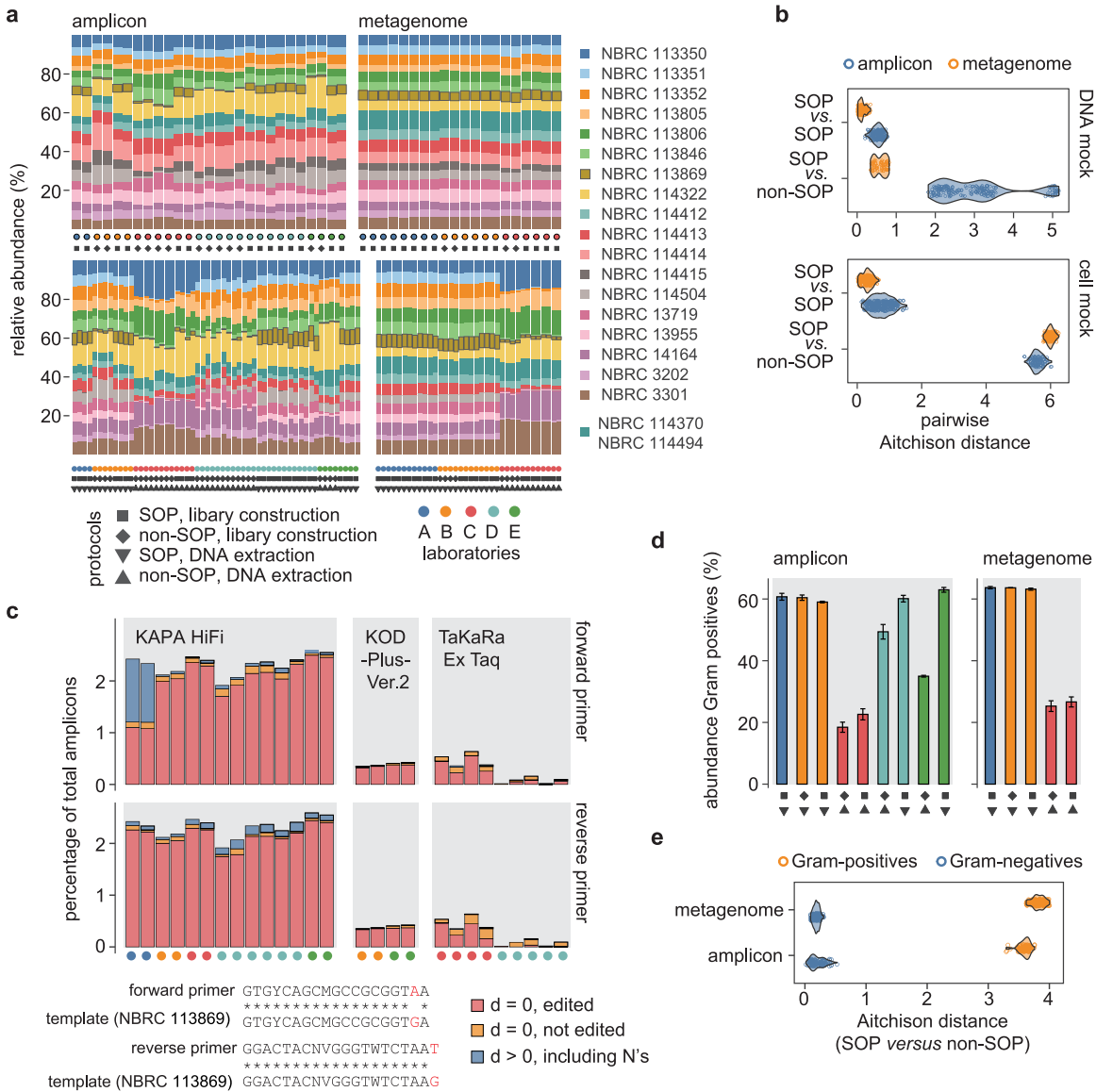


FIG 2 (a) Stacked bar charts of individual measurement results generated in the collaborative study for the DNA (top) and cell (bottom) mock community by amplicon and metagenome sequencing. Symbols below each bar indicate the type of protocol (SOP versus non-SOP for DNA extraction and library construction) and laboratory, as indicated in the legend. Note that the two strains of *Bifidobacterium longum* (NBRC 114370 and NBRC 114494) were analyzed as a single species. (b) Violin plots of pairwise Aitchison distances for measurements performed using SOPs and non-SOPs for library construction (top) and DNA extraction (bottom). Individual data points were overlaid with jitter. For the cell mock community, only protocols employing SOPs for library construction were included to highlight the effect of DNA extraction. (c) Bar charts of the relative abundance (without 16S rRNA gene copy number correction) of *C. acnes* strain NBRC 113869 for amplicon libraries prepared following the SOP using KAPA HiFi DNA polymerase and non-SOPs using alternative polymerases as indicated. Red and orange bars represent amplicons perfectly matching the expected primer sequences (Hamming distance, *d*, of zero), with and without editing, respectively. Blue bars indicate amplicons with at least a single mismatch, including undetermined (N) bases, to the expected edited or unedited primer sequences. The location of the primer mismatches to the template sequences is shown below the plot. (d) Cumulative relative abundance of Gram-positives measured for the cell mock community by amplicon and metagenome sequencing. Bars are colored by laboratory and symbols below the bars show the type of protocol, as in panel a. Bar heights represent the mean and error bars show standard deviations. (e) Violin plots of the distribution of pairwise Aitchison distances for the subcomposition of Gram-positives and negatives, considering SOPs and non-SOPs for DNA extraction and SOPs only for library construction.

The DNA mock community revealed high reproducibility of species profiles generated by different laboratories when using SOPs for both metagenome and amplicon library construction (Fig. 2a). To quantify differences between replicates, we examined variability in measured compositions based on Aitchison distances. While similar analysis using Bray-Curtis dissimilarities yielded qualitatively consistent results, the Aitchison distance is considered a superior metric because it better accounts for compositionality of the data.

This analysis showed that between-laboratory dispersion of species profiles measured by amplicon sequencing was, as expected, higher than for metagenome sequencing (P -value < 0.01 ; Mann-Whitney U test of pairwise Aitchison distances; Fig. 2b). Use of non-SOPs for library construction had only a small effect on species profiles measured by metagenomics but led to much larger differences for amplicon sequencing. As assessed using the cell mock community, the SOP for DNA extraction similarly yielded highly reproducible results across laboratories, both as measured by metagenome and amplicon sequencing (Fig. 2b). On average, use of non-SOPs for DNA extraction led to the largest variability in species profiles, underscoring that DNA extraction represents a dominant source of variability in metagenomic analysis methods.

In addition to serving as control reagents for comparing results across methods or laboratories, the mock communities may also assist in identifying systematic bias due to specific protocols. For instance, we found that the abundance of strain *C. acnes* NBRC 113869 was significantly higher in amplicon libraries generated using the SOP compared to the evaluated non-SOPs (P -value < 0.01 , Welch's t test), and consistent with abundances estimated by metagenomics (Fig. 2a). Based on previous studies (12, 25), it was hypothesized that the superior recovery of *C. acnes* by the SOP was associated with the use of KAPA HiFi DNA polymerase because this enzyme can efficiently amplify sequence templates with mismatches to the primers by so-called primer editing. Indeed, inspection of the primer regions in the sequenced amplicons showed that, on average, 96% of sequences were edited to match the template sequence (Fig. 2c). While KOD-Plus- DNA polymerase also showed high primer-editing activity, overall recovery of *C. acnes* remained however much lower (Fig. 2c). Based on this finding, and in line with the best practices of Gohl et al. (12), the use of polymerases with strong proof-reading/primer-editing activity is thus recommended for amplicon-based microbiome studies, at least provided that non-target amplification, due to for example excess of host DNA in tissue-derived samples, is not of concern or can be suppressed. This is expected to improve accuracy by enabling efficient amplification of sequence templates with mismatches to the used PCR primers, especially mismatches near the 3'-end of commonly used 16S rRNA gene amplification primers (25).

In similar fashion, the cell mock community revealed that differences in species profiles between the SOP and non-SOPs for DNA extraction were mainly due to variable total recovery of Gram-positives, as consistently observed by metagenome and amplicon sequencing (Fig. 2d). Further, differences in species profiles within the subcomposition of Gram-positives were much larger than within the subcomposition of Gram-negatives (Fig. 2e). This indicated that the efficiency of DNA recovery between Gram-positive species/strains is more variable than for Gram negatives, as has been reported previously by us (11) and others (15).

Taken together, these data showed that the mock communities can reveal technically meaningful differences in measurement results due to varying protocols and laboratories. They thus provide valuable control reagents to improve consistency, especially when used in conjunction with defined acceptance criteria for validating protocols and routine quality assurance. For instance, application of the above-defined guidance values showed that all here evaluated non-SOPs for DNA extraction resulted in higher than allowable errors whereas non-SOPs for metagenome library construction showed only slightly different performance than the SOPs (Fig. S4 in the supplemental material). Further, the mock communities also revealed important sources of bias and thus provide a useful resource for development and optimization of protocols.

Comparison of taxonomic profilers and impact of read processing. Mock communities can further serve as control reagents for assessing performance of bioinformatics procedures. Here, we demonstrate this by comparing different taxonomic profiling tools and evaluating the effect of quality trimming and filtering of reads on species abundances.

Comparison of taxonomic profilers. To compare metagenome-based taxonomic profilers, we used the sequencing data generated as part of the characterization of the mock communities, providing a total of 16 data sets for the DNA mock community as detailed above (Fig. S1 and Table S3 in the supplemental material). To account for differences in sequencing depth, data were randomly subsampled to 4 million read pairs

per sample, unless stated otherwise. Subsampled reads were then subjected to profiling by MetaPhlan3 (23), mOTUs2 (26), kraken2 (27), and kraken2 followed by bracken (28). We note that this selection of tools was not intended to serve as a comprehensive benchmark but rather capture algorithms that are widely used in the microbiome field and showed overall favorable performance in recent comparisons (29, 30). For mOTUs2 and MetaPhlan3, we used the databases provided with the tools. For kraken2 and bracken, we obtained publicly available indexes and k-mer distributions build based on the Genome Taxonomy Database (GTDB, release 95; referred to as kraken2_gtdb and bracken_gtdb hereafter) and RefSeq (referred to as kraken2_refseq, and bracken_refseq). Further, for kraken2, we applied a relatively lenient threshold (0.05) for the confidence score, following a recent benchmarking study (29). For bracken, we set the abundance threshold to 1000, such that species with less than 1,000 reads (prior to read reassignment) were excluded from the final species profiles.

To gauge precision of the profilers, we evaluated the number of species classified at varying minimum relative abundance thresholds (29). As displayed in Fig. 3a, all profilers classified a comparable number of high-abundance species, close to the expected value. Further, for the marker-gene based methods mOTUs2 and MetaPhlan3, the number of identified species remained relatively constant for decreasing minimum abundance thresholds. In contrast, kraken2 led to a rapid proliferation of the number of identified species at lower abundances. This was especially pronounced for kraken2_gtdb, which classified a considerably number of species with relative abundances on the order of 0.1%. As the DNA mock community contains only 20 species, virtually all these species thus represent false positives (that is, species incorrectly classified as present in the mock community). Processing of the kraken2 profiles by bracken, as expected, substantially reduced the number of low-abundance species. Inspection of the species profiles further showed that false positives for kraken2, and by extension bracken, were often due to species belonging to the same genus as species known to be present in the mock community (Fig. S5 in the supplemental material).

We further compared profilers based on detection of expected species (that is, true positives as identified based on the assigned species names for each of the profilers) and agreement with the expected composition. Note that for kraken2 and bracken, we focused on profiles based on the RefSeq database (that is, kraken2_refseq and bracken_refseq) because of the challenge of handling the vast number of false positives for kraken2_gtdb. We stress however that the inflation of false positives for kraken2_gtdb is not indicative of poor performance of the GTDB *per se* but rather highlights the challenge of using comprehensive databases with read assignments by kraken2. For all profilers, expected species accounted for the majority of tallied species abundances, ranging from $92.4 \pm 0.6\%$ (mean and sd) for mOTUs2 to $97.0 \pm 0.2\%$ for bracken_refseq. As shown in Fig. 3b, estimated species abundances varied however considerably among tools and not all expected species were identified by all tools. For instance, MetaPhlan3 failed to detect *B. subtilis* but rather identified two other, false positive *Bacillus* species, namely, *B. intestinalis* and *B. murimartini* (see also Fig. S6 in the supplemental material). Similarly, the species *Megasphaera massiliensis* was not identified by kraken2_refseq and bracken_refseq but rather classified as predominantly *M. elsdenii* (Fig. S6). As depicted in Fig. 3b, abundant 'unexpected' taxa typically represented species belonging to the same genus as genera present in the mock communities or taxa not fully annotated to the species level (e.g., *Pseudomonas* sp. for mOTUs2, *Megasphaera* sp. DISK 18 for MetaPhlan3 and *Lachnospirillum* sp. YL32 for kraken2_refseq). Agreement with the expected even species profiles, expressed as Bray-Curtis similarities, ranged from $78.6 \pm 0.2\%$ (bracken_refseq) to $85.0 \pm 0.4\%$ (MetaPhlan3), which was substantially lower than for kallisto ($95.6 \pm 0.2\%$; Fig. S7 in the supplemental material).

Further, Fig. 3b also revealed higher variation in estimated species abundances across replicates for mOTUs2, with a variability of 29.2% (qmCV calculated around the grand mean of all measurements), compared to 5.2% (MetaPhlan3) or less for the other tools. This was also reflected in terms of pairwise Aitchison distances, with mOTUs2 showing significantly larger dispersion between replicates than the other profilers (P -value < 0.01 , Mann-Whitney U test on pairwise distances), both within a single laboratory and across

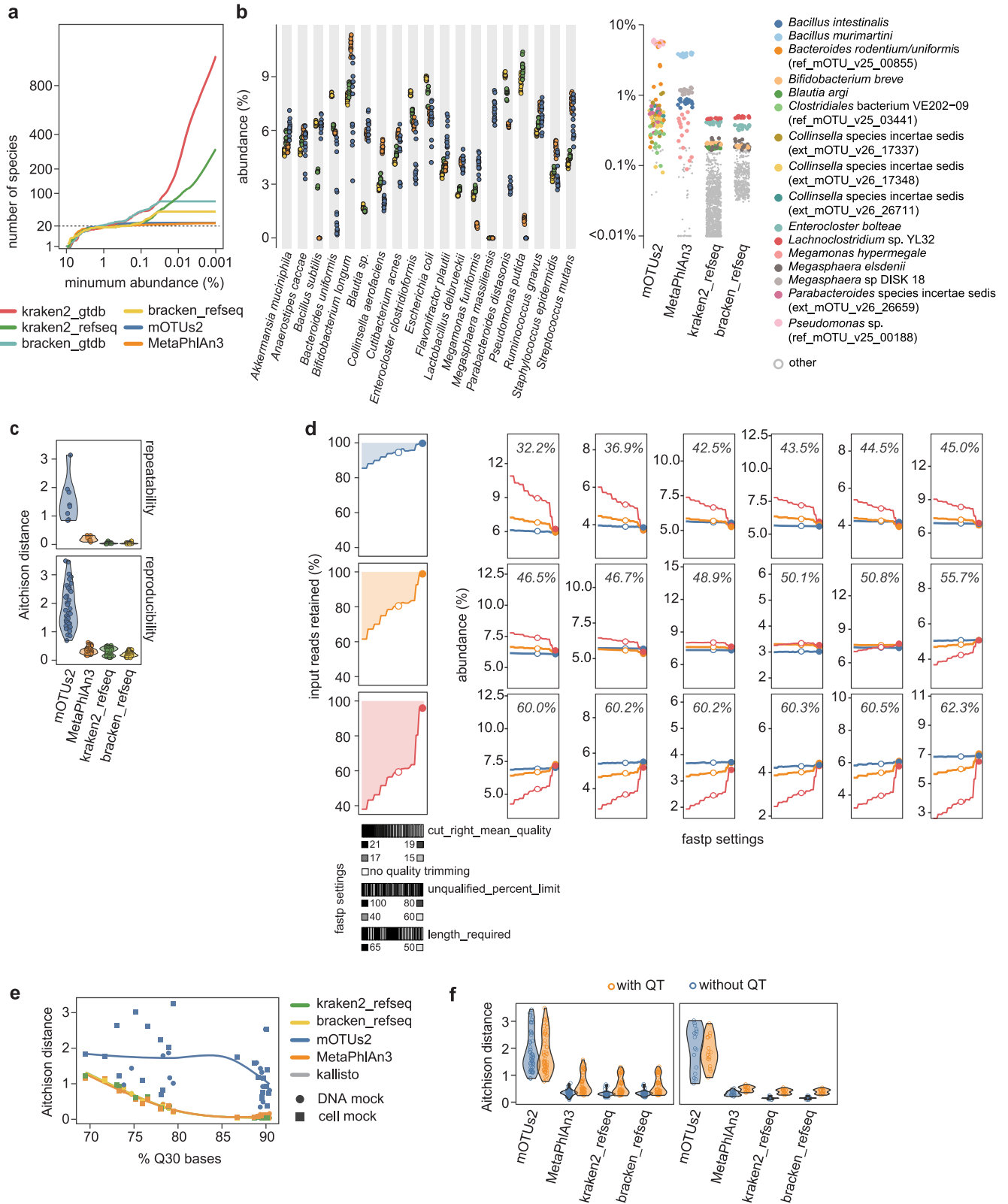


FIG 3 (a) Number of species classified by each of the profilers plotted as a function of minimum abundance threshold. Data are shown as the mean (colored solid lines) and standard deviation (ribbons, if visible) of 16 replicated measurement of the DNA mock community, following SOPs for library construction. The dashed horizontal line indicates the expected number of species. (b) Estimated abundances of expected species in the DNA mock community. Symbols represent results of individual measurements and are colored according to the taxonomic profiler as in panel a. The graph on the right shows the abundance of false positives (that is, species that are not included in mock the community but classified as present), with the abundant

(Continued on next page)

laboratories (Fig. 3c). To clarify the source of this increased dispersion, we randomly subsampled the four deeply sequenced HiSeq libraries for the DNA mock community (Table S3 in the supplemental material) to between 0.5 and 20 million read pairs per sample and calculated differences between species profiles generated using the subsampled reads, for each library and subsampling depth separately. As shown in Fig. S8, mOTUs2 resulted in consistently higher differences between pairs of species profiles of subsampled reads, and this was correlated with subsampling depth. This observation suggested that the apparent poor replication for mOTUs2 seems to be due to stochastic noise arising from the smaller fraction of input reads mapped to a limited number (up to 10) of marker genes. In line with this, MetaPhlan3, which uses about 80 clade-specific marker genes per species, also resulted in slightly higher dissimilarities than kraken2_refseq and bracken_refseq (Fig. S8) but this effect was less pronounced than for mOTUs2. These observations suggest that deeper sequencing is beneficial for profiling tools that use a limited number of marker genes to improve reproducibility of inferred species abundances.

Species-dependent bias due to read trimming and filtering. To assess the impact of read trimming and filtering, we exploited the variable raw base quality across NextSeq 500 sequencing runs and libraries (Table S3 in the supplemental material). More specifically, unless stated otherwise, we processed the reads with fastp using a range of settings for the parameters `cut_right_mean_quality`, `length_required` and `unqualified_percent_limit` (Table S5). As depicted in Fig. 3d, the applied settings resulted in highly variable retention of raw reads, with lower-quality reads being discarded more aggressively when applying more stringent read trimming and filtering.

To inspect the impact of read trimming and filtering, we plotted the relative abundance of each strain in the processed data as a function of fastp settings, from more to less stringent (Fig. 3d). This revealed substantial differences in strain-wise abundances for the poorest quality read data upon more aggressive filtering/trimming of raw reads, and this effect was strongly associated with the genomic GC content of the strains. More specifically, reads from GC-rich genomes tended to be removed more aggressively (Fig. S9 in the supplemental material), leading to lower relative abundance of high-GC strains in the surviving reads after stringent trimming and filtering (Fig. 3d). This GC-dependent effect was also observed after processing of the reads using Cutadapt (Fig. S10), which employs a different algorithm for trimming of reads based on their base quality scores than fastp.

Increased bias due to read trimming and filtering was further consistently observed in species profiles generated using the above-described taxonomic profilers. More specifically, Aitchison distances between profiles generated based on reads that underwent only minimal preprocessing (settings 0_100_65; that is, no quality filtering and minimum length of 65 bp after adapter and trailing polyX trimming) and reads subjected to quality trimming (settings 15_100_65; that is, including 4 bp window-based trimming with quality threshold of 15) increased substantially for read data with lower bases qualities (Fig. 3e). As expected, this also led to larger differences between results generated by different laboratories due to

FIG 3 Legend (Continued)

species colored as indicated in the legend. (c) Violin plots showing the distribution of Aitchison distances, calculated based on the abundances of the expected species as shown in panel b, for each of the profilers. The top and bottom plots show all pairwise dissimilarities between technical replicates within a single laboratory (denoted as repeatability) and replicates from different laboratories (denoted as reproducibility), respectively. Larger pairwise distances indicate higher variability across replicated measurements. (d) Impact of read trimming/filtering on read retention and species abundances. The left three panels show the percentage of raw reads retained following read processing by fastp for three representative data sets with varying base quality (blue: Q30 bases of 90.1%; orange: 78.2%; red: 71.7%). Applied settings for read processing by fastp, are indicated below (also see Table S5 in the supplemental material) and sorted according to the percentage of retained reads for the data set with lowest raw base quality. The adjacent plots show the relative abundance of each of the strains, as determined by kallisto, for the same three data sets. Facets are sorted according to increasing genomic GC content, indicated in each of the plots. Filled and empty circles indicate the two settings evaluated for panels d and e. (e) Aitchison distances between species profiles of read data with quality trimming (settings 15_100_65 in Table S5 in the supplemental material) and without quality trimming (settings 0_100_65). Data points, each representing a different library, are plotted as a function of the libraries' raw base quality (x axis); colors and shapes reflect the mock community and profiling tool, respectively. Solid lines represent loess fits, for visualization purposes only. (f) Violin plots of pairwise Aitchison distances demonstrating the impact of quality trimming (settings 15_100_65, with QT and 0_100_65, without QT) on perceived reproducibility. Only pairwise comparisons for measurements performed by different laboratories following the SOPs were included.

variability in base quality (Fig. 3f). Finally, although omitting read trimming led to a slight increase in the number of unclassified reads (kraken2) or “unassigned” species (mOTUs2), no strong effect on the number of classified species was found (Fig. S11 in the supplemental material).

Finally, we also observed species-dependent variation in the quality of 16S rRNA gene amplicon sequencing reads. For instance, the reverse reads of *Parabacteroides distasonis* NBRC 113806 consistently displayed the lowest quality (Fig. S12 in the supplemental material), measured in terms of the expected errors as predicted based on the quality scores (31). This appeared to be due to a steep decline in raw base quality around position 65 (Fig. S13), presumably due excessive dephasing during sequencing. Of note is that, within our data set, such a drop in quality was not observed for reads generated with V3 chemistry by lab E; all other labs used V2 chemistry. These data highlighted that variation in raw base quality profiles among different species can lead to appreciable bias in taxonomic profiles depending on the applied settings for processing of raw reads, as we and others documented previously (32, 33).

Taken together, the above results validated the utility of the mock communities for assessing performance of bioinformatics pipelines. The substantial deviation from the expected composition observed for all profilers also underscored that for optimal use of the mock communities, quantification should employ dedicated reference sequences, preferably complete genomes, as this can reveal more subtle and meaningful differences between protocols. Here, we employed kallisto for quantification because of its speed and accuracy but other algorithms may also be used.

Further, our observations also point to a need to sufficiently scrutinize the effect of read quality trimming and filtering on metagenomics inferences, mirroring previous recommendations for RNA-seq (34). Here, we found that, for NextSeq 500 sequencing data, species-dependent bias due to read trimming and filtering was strongly associated with genomic GC content. In-depth investigation of whether similar effects occur on other sequencing platforms was beyond the scope of this study but variable GC-related bias depending on platform has previously been reported in the context of microbial (meta)genome sequencing (35). In any case, these observations suggest that read trimming may lead to substantial bias depending on the quality of the data, which may vary among sequencing runs. As such, it is advisable to perform only gentle trimming and/or filtering of reads, provided that the performance of downstream analyses is not negatively affected. In this context, inclusion of mock communities in each sequencing run will allow consistent identification of potentially aberrant sequencing runs and adequate evaluation of settings for preprocessing of reads to prevent bias.

To conclude, we have developed two mock communities intended to serve as control reagents for human microbiota analyses by high-throughput sequencing methods. Through a collaborative study, we carefully characterized the composition of the mock communities and demonstrated their utility in typical use cases. The mock communities provide appropriate control reagents that can be analyzed in each sequencing run to track consistency of the data and reveal potential systematic bias or drift due to e.g., varying base quality and read trimming or filtering. The mock communities, which are available from the NITE Biological Resource Center (NBRC) at the National Institute of Technology and Evaluation (NITE, Japan), are expected to improve standardization and quality assurance as microbiome research is rapidly being translated into new therapeutic and diagnostic applications.

MATERIALS AND METHODS

Bacterial strains and cultivation. Cells of all strains in the mock communities were obtained from NBRC at the National Institute of Technology and Evaluation (NITE, Japan). Liquid cultures were prepared using the media and cultivation conditions shown in Table S6. Cells were harvested during the late-log to stationary growth phase by centrifugation (15 min at $4,000 \times g$), washed with phosphate-buffered saline (PBS, pH 7.4), and stored at -80°C in PBS with 15% glycerol as cryoprotectant.

Formulation of the DNA mock community. High-molecular-weight DNA was extracted from the cultured cells using the MagAttract HMW DNA Kit (Qiagen), following the manufacturer’s instructions. To improve lysis efficiency, cells were pretreated with 8.3 mg mL^{-1} of lysozyme (Sigma) and 2800 units

mL^{-1} of achromopeptidase (FUJIFILM Wako Pure Chemical Corporation) for 15–60 min at 37°C , for strains NBRC 114412, NBRC 114370, NBRC 114494, NBRC 113351, NBRC 114504, NBRC 113805, NBRC 3202, NBRC 114413, NBRC 13955, NBRC 113869, NBRC 113846, NBRC 13719, NBRC 113806, and NBRC 114322. Purified DNA was treated with 0.5 mg mL^{-1} of RNase A (Nippon Gene Co., Ltd.) in Buffer AE (Qiagen; 10 mM Tris-HCl, 0.5 mM EDTA, pH 9.0) for 30 min at 37°C and cleaned up with phenol:chloroform:isoamyl alcohol (25:24:1, vol/vol) and chloroform. Cleaned-up DNA was recovered by ethanol precipitation, dissolved in Buffer EB (Qiagen; 10 mM Tris-HCl, pH 8.5), and stored at 4°C until use.

Total DNA concentrations were measured with the Quant-iT PicoGreen dsDNA Reagent (Thermo Fisher) using a Varioskan LUX Multimode Microplate Reader (Thermo Fisher). Measurements were performed in triplicate and averaged. DNA mass concentrations were converted to genome copy numbers based on the known size (or molecular weight) of the genomes. Genomic DNA from each strain were combined to obtain a near-even DNA mock community. Aliquots at a concentration of approximately $50 \text{ ng } \mu\text{L}^{-1}$ were prepared in Buffer EB and stored at -80°C until use.

Formulation of the cell mock community. For the cell mock community, total DNA concentrations in the cell stocks of each strain were determined by measuring adenine content following the method of de Bruin and Birnboim (22), with minor modifications as detailed previously (11). Two replicate measurements were performed and averaged. Based on the known base composition and size of the genomes, total DNA concentrations (in terms of genome copy numbers) were calculated. Cells from each strain were combined to obtain an even cell mock community, containing an equal genome copy number, or cell equivalents, of each strain. Aliquots at a concentration of approximately 4×10^{10} cells mL^{-1} , in PBS with 15% glycerol, were prepared and stored at -80°C until use.

Whole-genome sequencing and assembly. Long-read libraries for Nanopore sequencing were prepared with the Ligation Sequencing Kit (SQK-LSK109, Oxford Nanopore Technologies, ONT) and Native Barcoding Expansion pack (EXP-NBD104), following manufacturer's provided instructions. For strains NBRC 113846, NBRC 113869, NBRC 114322, NBRC 114370, NBRC 114412, NBRC 114413, NBRC 114414 and NBRC 114415, sequencing was performed on a MinION device (ONT) using an R9.4.1 flow cell. Bases were called with Guppy v4.5.3 in high accuracy mode (dna_r9.4.1_450bps_hac.cfg). For demultiplexing and trimming of barcodes, Guppy's guppy_barcode script was used, specifying command line flags `-barcode_kits "EXP-NBD104" -trim_barcodes -num_extra_bases_trim 2`. For strains NBRC 113805, NBRC 114494 and NBRC 114504, sequencing was performed on a GridION device (ONT), also using an R9.4.1 flow cell. Base calling and adapter trimming were performed using Guppy v1.8.5 and Porechop v0.2.3 (36, <https://github.com/rwrick/Porechop>), respectively. In all cases, ONT reads underwent additional trimming using Porechop v0.2.4, specifying option `-discard_middle`. Subsequently, a subset of high-quality reads with an estimated depth of $200\times$ was selected using FilTlong v0.2.0 (<https://github.com/rwrick/FilTlong>), with options `-min_length 1000 -min_mean_q 9 -mean_q_weight 30`.

For all strains, short-read libraries were prepared using the TruSeq DNA PCR-Free kit (Illumina), following recommended procedures, and sequenced on an Illumina MiSeq instrument using V3 chemistry, generating 2×301 bp reads. Sequencing reads were quality controlled using fastp v0.20.0 (37), specifying options `-trim_tail1 1 -trim_tail2 1 -cut_right -cut_right_window_size 4 -cut_right_mean_quality 18 -n_base_limit 0 -length_required 50`.

For all strains, long-read assemblies were generated using Flye v2.8.3 (38), specifying the expected genome size (`-genome-size`) and flags `-asm-coverage 50 -plasmids`, except for strain NBRC 114322 for which option `-meta` was specified. For strains NBRC 114494 and NBRC 114504, final assemblies were generated using Unicycler v0.4.7 (39), using the quality-controlled Illumina short reads and Flye assemblies as inputs. For the other strains (i.e., NBRC 113846, NBRC 113869, NBRC 114322, NBRC 114370, NBRC 114412, NBRC 114413, NBRC 114414, NBRC 114415, and NBRC 113805), the long-read assemblies were polished with the long reads using Racon v1.4.20 (40), specifying options `-m 8 -x -6 -g -8 -w 500`. For all strains except NBRC 113805, this was followed by polishing using ONT's Medaka v1.2.2 (<https://github.com/nanoporetech/medaka>). Finally, for all strains, including NBRC 113805, additional polishing was performed with the short reads using Pilon v1.23 (41); circular sequences were rotated between polishing rounds to ensure effective polishing of the ends of the sequences, including filling in a small number of bases typically missing from the ends of assemblies generated by Flye.

Genome assemblies were inspected with CheckM v1.1.3 (42), using lineage-specific marker gene sets (`lineage_wf` command). A phylogenetic tree was generated with the `infer` command of GTDB-Tk v1.6.0 (43), using the aligned single copy marker genes generated by the `classify_wf` command as inputs. Ribosomal RNA (rRNA) gene sequences were extracted from the genomes using Barnap v0.9 (<https://github.com/tseemann/barnap>). Recovered small subunit (16S) rRNA gene sequences were trimmed to the expected amplicon sequences (V4 hypervariable region, primer set 515F/806R, see below) using Cutadapt v3.4 (44) and unique sequence variants retained as reference sequences for analysis of the 16S rRNA gene amplicon sequencing data as detailed below.

Simulated reads were generated using BBMap's (v38.82; 45) `randomreads.sh` script, specifying command line flags `adderrors=f paired=t minlength = 151 maxlength = 151`. For consistency with the 'real' read data, simulated reads were processed by fastp, with command line flags as specified above.

Protocol (SOP) for DNA extraction. Extraction of DNA was performed using the ISOSPIN Fecal DNA kit (Nippon Gene Co., Ltd.), following our previously developed SOP based on the manufacturer's provided instructions (denoted as protocol N by Turlousse et al.) (11). More specifically, sample (here, $150 \mu\text{l}$ of cell mock community), along with $700 \mu\text{l}$ of FE1 Buffer and $10 \mu\text{l}$ of RNase, were added to the provided tubes with beads. Cells were disrupted by bead-beating for 1 min at a speed of 6 m s^{-1} using a FastPrep-24 instrument (MP Biomedicals). Bead beating was repeated three times, keeping the sample at room temperature for 5 min between rounds. Subsequently, $90 \mu\text{l}$ of FE2 Buffer was added and the sample centrifuged for 15 min at $12,000 \times g$. Supernatant (up to $500 \mu\text{l}$) was then recovered and mixed

with 0.4 volumes of FB Buffer and isopropanol. The sample was finally loaded onto the spin columns and cleaned up as per the provided instructions; purified DNA was eluted in 100 μ l of Tris-EDTA (pH 8.0) buffer. Detailed descriptions for the additional DNA extraction protocols (that is, non-SOPs) are provided in the Supplementary Methods.

Protocols (SOPs) for shotgun library construction and sequencing. Unless stated otherwise, construction of metagenome sequencing libraries was performed using the QIAseq FX DNA Library kit Qiagen; protocol B in Tourlousse et al. (11) and SMARTer ThruPLEX DNA-Seq kit (TaKaRa Bio; protocol K), according to manufacturer's provided instructions as follows.

For protocol B, enzymatic fragmentation reactions (50 μ l) contained 1 \times FX Buffer, 10 μ l of FX Enzyme Mix, and 50 ng of DNA template; reactions were incubated for 12 min at 32°C. For adapter ligation, 5 μ l of adapters were added, along with 20 μ l of DNA Ligase Buffer, 10 μ l of DNA Ligase, and 15 μ l of RNase-free H₂O; reactions were incubated for 15 min at 20°C. Adapter-ligated fragments were then purified and size-selected with the Agencourt AMPure XP PCR Purification system, using sequentially 1 and 0.8 volumes of bead solution, and eluted in 10 mM Tris-HCl buffer. Subsequently, adapter-ligated and size-selected fragments were amplified in PCRs (50 μ l) containing 1 \times HiFi PCR Master Mix, 1.5 μ l of Primer Mix and 23.5 μ l of DNA template. Thermal cycling conditions were as follows: 98°C for 2 min; 8 cycles of 98°C for 20 s, 60°C for 30 s and 72°C for 30 s; 72°C for 1 min. Finally, amplified libraries were purified as above, using 1 volume of bead solution, and eluted in 10 mM Tris-HCl (pH 8.0).

For protocol K, 1 μ g of DNA was subjected to focused ultrasonication using the Covaris M220 instrument, setting a target fragment size was 350 bp. Fragmented DNA was purified using the Agencourt AMPure XP PCR Purification system, bead-to-sample ratio of 1.8:1, and eluted in low-EDTA Tris-HCl buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0). Subsequently, DNA was subjected to 'template preparation' in reactions (13 μ l) containing 50 ng of fragmented DNA (10 μ l), 2 μ l of Template Preparation D Buffer and 1 μ l of Template Preparation D Enzyme; reactions were incubated for 25 min at 22°C and then for 20 min at 55°C. For 'library synthesis', 1 μ l each of Library Synthesis D Buffer and Library Synthesis D Enzyme were added and reactions incubated for 40 min at 22°C. For 'library amplification', 25 μ l of Library Amplification D Buffer, 1 μ l Library Amplification Enzyme, 5 μ l of Indexing reagents and 4 μ l of nuclease-free water were added and reaction subjected to thermal cycling as follows: 3 min at 72°C, 2 min at 85°C, 2 min at 98°C, 4 cycles of 98°C for 20 s, 67°C for 20 s, 72°C for 40, and 6 cycles of 98°C for 20 s and 72°C for 50 s. Finally, amplified libraries were purified with the Agencourt AMPure XP PCR Purification system, using 1 volume of beads, and eluted in 10 mM Tris-HCl (pH 8.0).

Libraries were quantified using the D5000 ScreenTape Assay system and 2200 TapeStation instrument (Agilent) and pooled at equimolar concentrations. Sequencing was performed using two different platforms, namely, a NextSeq 500 instrument using NextSeq 500/550 Mid Output Kit v2.5 (300 cycles, 2 \times 151 bp reads) at Laboratory A and a HiSeq 2500 instrument using the Rapid SBS Kit V2 (200 cycles, 2 \times 101 bp reads) at Laboratory C. All laboratories shared raw fastq files, after demultiplexing, with the central laboratory (that is, Lab A) for uniform processing and analysis, as described below.

Protocol (SOP) for amplicon library construction and sequencing. As SOP, we followed Illumina's two-step tailed PCR protocol to generate amplicon libraries targeting the V4 hypervariable region of the 16S rRNA gene. More specifically, first round PCRs contained 1 \times KAPA HiFi HotStart ReadyMix, 500 nM (each) forward (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGTCAGCMGCCGCGGTAA-3', synthesized by Hokkaido System Science Co., Ltd., the locus-specific 515F primer region is underlined) and reverse primer (GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGACTACNVGGGTWTCTAAT; the 806R primer region is underlined) and DNA template. Thermal cycling conditions were as follows: 95°C for 3 min; 25 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 30 s; and 72°C for 5 min. Amplicons were purified using the Agencourt AMPure XP PCR Purification system (bead-to-sample ratio of 1:1) and eluted in 10 mM Tris-HCl (pH 8.5). The Nextera XT Index Kit was then used to attach dual indexes and sequencing adapters, in PCRs (50 μ l) containing 1 \times KAPA HiFi HotStart ReadyMix, 5 μ l each of Index 1 and 2 primers, and 5 μ l of purified first-round PCR products. Thermal cycling conditions were as follows: 95°C for 3 min; 8 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 30 s; 72°C for 5 min. Amplicons were purified as above and quantified using the D1000 ScreenTape Assay system and 2200 TapeStation instrument (Agilent). Amplicon libraries were pooled at equimolar concentrations and sequenced on a MiSeq instrument using V2 (2 \times 251 bp) or V3 (2 \times 301 bp) chemistry, in the presence of up to 30% of phiX control DNA. Detailed descriptions for the additional PCR protocols (that is, non-SOPs) are provided in the Supplementary Methods.

All laboratories shared demultiplexed fastq files with the central laboratory (i.e., Lab A) for processing and analysis, as described below.

Shotgun sequencing data processing and analysis. Whole metagenome sequencing reads were pre-processed using fastp v0.20.0, specifying the following command line arguments, unless stated otherwise: `-trim_front1 5 -trim_front2 5 -trim_tail1 1 -trim_tail2 1 -unqualified_percent_limit 100 -trim_poly_x -poly_x_min_len 10 -n_base_limit 0 -low_complexity_filter -length_required 65`. Following pre-processing, if applicable, fastq files were randomly subsampled to 4 million reads using seqtk v1.3 (<https://github.com/lh3/seqtk>), using a constant seed for the forward and reverse sequencing read data.

Quantification of mock community members in the sequencing data was performed using kallisto v0.46.1 (24), using its quantification algorithm (kallisto quant) run with default command line arguments. Estimated counts assigned to each strain were summed across replicons (that is, chromosomes and plasmids, if applicable), normalized based on their effective length to obtain strain-wise coverages, and converted to relative abundances.

Taxonomic profiling was performed using the following tools and databases: (i) kraken2 v2.1.1 (27) using the Genome Taxonomy Database, release 95 (46) or RefSeq (47); (ii) bracken v2.6.0 (28) for reassigning reads at the species level; (iii) MetaPhlan3 v3.0.9 (23) using the mpa_v30_CHOCOPHlan_201901

database; and (iv) mOTUs2 v2.6.1 (26) using the default database associated with this version. Sources of the databases are provided in Table S7 in the supplemental material.

For kraken2, a confidence score threshold (command line flag `-confidence`) of 0.05 was used, following a recent benchmarking study (29). For bracken, a threshold (command line flag `-t`) of 1000 was used to accommodate the relative high coverage of the simple mock community. For kraken2 and bracken, generated reports were converted to mpa format using the script `kreport2mpa.py` (<https://github.com/jenniferlu717/KrakenTools>), specifying the command line arguments `-read_count -no-intermediate-ranks`. For MetaPhlAn3, bowtie2 v2.4.1 (48) was used to map the reads against the reference database, specifying options `-very-sensitive -no-unal`; fastq files of the forward and reverse reads were concatenated prior to mapping. Subsequently, reads with a mapping quality below 5 were removed using `samtools v1.10` (49) and the filtered read mappings were used for species-level profiling using MetaPhlAn3; note that no length filtering was employed as part of the MetaPhlAn3 pipeline but a minimal read length of 65 bp enforced during read preprocessing using `fastp` (see above). For mOTUs2, taxonomic profiles (mOTU level) were generated using the `profile` command, involving mapping of the reads (`map_tax`) to the marker gene database and generating profiles using `calc_mgc` and `calc_motu`. Command lines are provided in Table S7 in the supplemental material.

To identify expected species/strains in the outputs generated by each of the profilers, we parsed the species/mOTU-level assignments by `grep` using the species names, excluding subspecies names, shown in Table 1 as search terms, except for *Blautia* sp. NBRC 113351 (search term “*Blautia producta*”), *Bacteroides uniformis* NBRC 113350 (search term “*Bacteroides uniformis|Bacteroides rodentium|Vuniformis*”) and *Enterocloster clostridioformis* NBRC 113352 (search term “*Enterocloster clostridioformis|Clostridium clostridioforme*”).

Amplicon sequencing data processing and analysis. Trimming of primer sequences was performed using `Cutadapt v3.4`, specifying command line flags `-g YCAGCMGCCGCGTAA -G TACNVGGTWTCTAAT -no-indels -discard-untrimmed -max-n 0 -error-rate 0.2 -minimum-length 225 -length 225`. Subsequently, reads were truncated and filtered based on expected errors using the `DADA2's (v1.16; 50) filterAndTrim` function, specifying the following parameters: `truncLen = c(200,180)`, `trimLeft = c(0, 0)`, `maxN = c(0,0)`, `maxEE = c(4, 6)`, `truncQ = c(2)`. Trimmed reads were then merged using `USEARCH's (v11.0.667) (51) fastq_mergepairs` command, with options `-fastq_nostagger -fastq_pctid 80 -fastq_maxdiffs 999 -fastq_minovlen 100 -fastq_minmergelen 250 -fastq_maxmergelen 255`. Finally, merged reads were annotated with `USEARCH's annot` functionality, using the expected amplicon sequences extracted from the genomes (see above) as references. If applicable, counts assigned to different sequence variants for a given species were combined. Finally, read counts were corrected for 16S rRNA gene copy number, as determined based on the complete genome sequences, and species profiles expressed in terms of relative abundances, as percentages, based on the corrected read counts.

To investigate primer editing, we retained the leading 19 and 20 bp of the forward and reverse raw reads, respectively. This information was merged with the output of `USEARCH annot`, based on read names, and compared to the expected primer sequences. Similarly, to evaluate inspect strain-wise distributions of expected errors, `USEARCH's fastq_filter` command was used to append the expected errors to the read names (`option-fastq_eeout`), which was then merged with the read annotations for downstream analyses.

Downstream data analysis and visualization. All data were imported into R (v4.0.2) (52) for processing and visualization. Parsing of data was performed mainly using the packages `dplyr (v1.0.3) (53)` and `tidyr (v1.1.2) (54)`. All graphics were generated using `ggplot2 (v3.3.3) (55)`.

Differences between pairs of species-level taxonomic profiles were calculated as Aitchison distances (i.e., Euclidean distances after centered log ratio, `clr`, transformation) or as Bray-Curtis dissimilarities, using the function `vegdist` in the R package `vegan v2.5` (56). Variability between taxonomic profiles was calculated as the quadratic mean of species-wise coefficients of variation of abundances (denoted `qmCV`), as defined previously (11). Similarly, differences between two compositions (e.g., between measured and expected species profiles) were computed as the geometric mean of species-wise absolute fold differences (denoted `gmAFD`).

Guideline values for ‘allowable level of errors’ for the mock communities, here defined as the difference between values assigned by total DNA quantification during formulations and values measured by metagenomics, were calculated as described by Turlouze et al. (11). In short, we simulated formulated mock communities (that is, quantified based on DNA content) using R stats’ `rnorm` function, using the values assigned by total DNA measurements as mean and a standard deviation of 5% and 10% for the DNA and cell mock community, respectively. For the metagenomic measurements, generated using SOPs (see main text), simulated compositions were generated using the function `rnorm.acomp` in the R package `compositions (v2.0-1) (57)`. Here, the compositional center (that is, the closed geometric mean of strain-wise abundances) and (co)variance matrix (calculated using `compositions' var.acomp` function) were used, setting the off-diagonal of the (co)variance matrix to zero. Subsequently, differences between both simulated data sets were calculated for 10^5 random pairs and the 95% percentile set as the allowable level of errors. Based on the simulated data sets, we further defined ‘acceptance’ ranges for strain-wise abundances based on the absolute difference between both simulated data sets. More specifically, we calculated the 95% percentile of absolute differences and then calculated boundaries as the formulated abundances plus/minus the calculated 95% percentile.

Data availability. All sequencing data generated during the current study are available in NCBI’s Sequence Read Archive (SRA) under BioProject [PRJNA747117](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA747117); SRA accession numbers for individual data sets are provided in Table S3 in the supplemental material. Genome sequences have been deposited in `DBJ/EMBL/GenBank` under the accession numbers shown in Table 1.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 2.2 MB.

ACKNOWLEDGMENTS

This study was funded by the New Energy and Industrial Technology Development Organization (NEDO), Japan.

We declare no competing interests.

REFERENCES

- Manor O, Dai CL, Kornilov SA, Smith B, Price ND, Lovejoy JC, Gibbons SM, Magis AT. 2020. Health and disease markers correlate with gut microbiome composition across thousands of people. *Nat Commun* 11:5206. <https://doi.org/10.1038/s41467-020-18871-1>.
- Salosensaari A, Laitinen V, Havulinna AS, Meric G, Cheng S, Perola M, Valsta L, Alfthan G, Inouye M, Watrous JD, Long T, Salido RA, Sanders K, Brennan C, Humphrey GC, Sanders JG, Jain M, Jousilahti P, Salomaa V, Knight R, Lahti L, Niranen T. 2021. Taxonomic signatures of cause-specific mortality risk in human gut microbiome. *Nat Commun* 12:2671. <https://doi.org/10.1038/s41467-021-22962-y>.
- Sipos A, Ujlaki G, Mikó E, Maka E, Szabó J, Uray K, Krasznai Z, Bai P. 2021. The role of the microbiome in ovarian cancer: mechanistic insights into oncogenesis and to bacterial metabolite signaling. *Mol Med* 27:33. <https://doi.org/10.1186/s10020-021-00295-2>.
- Fischbach MA. 2018. Microbiome: Focus on causation and mechanism. *Cell* 174:785–790. <https://doi.org/10.1016/j.cell.2018.07.038>.
- Malla MA, Dubey A, Kumar A, Yadav S, Hashem A, Abd Allah EF. 2018. Exploring the human microbiome: The potential future role of next-generation sequencing in disease diagnosis and treatment. *Front Immunol* 9:2868. <https://doi.org/10.3389/fimmu.2018.02868>.
- Schlager R. 2020. Microbiome diagnostics. *Clin Chem* 66:68–76. <https://doi.org/10.1373/clinchem.2019.303248>.
- Mimee M, Citorik RJ, Lu TK. 2016. Microbiome therapeutics: Advances and challenges. *Adv Drug Deliv Rev* 105:44–54. <https://doi.org/10.1016/j.addr.2016.04.032>.
- Greathouse KL, Sinha R, Vogtmann E. 2019. DNA extraction for human microbiome studies: the issue of standardization. *Genome Biol* 20:212. <https://doi.org/10.1186/s13059-019-1843-8>.
- Amos GCA, Logan A, Anwar S, Fritzsche M, Mate R, Bleazard T, Rijpkema S. 2020. Developing standards for the microbiome field. *Microbiome* 8:98. <https://doi.org/10.1186/s40168-020-00856-3>.
- Hornung BVH, Zwiittink RD, Kuijper EJ. 2019. Issues and current standards of controls in microbiome research. *FEMS Microbiol Ecol* 95:fiz045.
- Tourlousse DM, Narita K, Miura T, Sakamoto M, Ohashi A, Shiina K, Matsuda M, Miura D, Shimamura M, Ohyama Y, Yamazoe A, Uchino Y, Kameyama K, Arioka S, Kataoka J, Hisada T, Fujii K, Takahashi S, Kuroiwa M, Rokushima M, Nishiyama M, Tanaka Y, Fuchikami T, Aoki H, Kira S, Koyanagi R, Naito T, Nishiwaki M, Kumagai H, Konda M, Kasahara K, Ohkuma M, Kawasaki H, Sekiguchi Y, Terauchi J. 2021. Validation and standardization of DNA extraction and library construction methods for metagenomics-based human fecal microbiome measurements. *Microbiome* 9:95. <https://doi.org/10.1186/s40168-021-01048-3>.
- Gohl DM, Vangay P, Garbe J, MacLean A, Hauge A, Becker A, Gould TJ, Clayton JB, Johnson RJ, Hunter R, Knights D, Beckman KB. 2016. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat Biotechnol* 34:942–949. <https://doi.org/10.1038/nbt.3601>.
- Brooks JP, Edwards DJ, Harwich MD, Jr, Rivera MC, Fettweis JM, Serrano MG, Reris RA, Sheth NU, Huang B, Girerd P, Strauss JF, 3rd, Jefferson KK, Buck GA, Vaginal Microbiome Consortium. 2015. The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiol* 15:66. <https://doi.org/10.1186/s12866-015-0351-6>.
- Sui HY, Weil AA, Nuwagira E, Qadri F, Ryan ET, Mezzari MP, Phipatanakul W, Lai PS. 2020. Impact of DNA extraction method on variation in human and built environment microbial community and functional profiles assessed by shotgun metagenomics sequencing. *Front Microbiol* 11:953. <https://doi.org/10.3389/fmicb.2020.00953>.
- Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, Tramontano M, Driessen M, Hercog R, Jung FE, Kultima JR, Hayward MR, Coelho LP, Allen-Vercoe E, Bertrand L, Blaut M, Brown JRM, Carton T, Cools-Portier S, Daigheault M, Derrien M, Druésne A, de Vos WM, Finlay BB, Flint HJ, Guamer F, Hattori M, Heilig H, Luna RA, van Hylckama Vlieg J, Junick J, Klymiuk I, Langella P, Le Chatelier E, Mai V, Manichanh C, Martin JC, Mery C, Morita H, O'Toole PW, Orvain C, Patil KR, Penders J, Persson S, Pons N, Popova M, Salonen A, Saulnier D, Scott KP, Singh B, et al. 2017. Towards standards for human fecal sample processing in metagenomic studies. *Nat Biotechnol* 35:1069–1076. <https://doi.org/10.1038/nbt.3960>.
- Han D, Gao P, Li R, Tan P, Xie J, Zhang R, Li J. 2020. Multicenter assessment of microbial community profiling using 16S rRNA gene sequencing and shotgun metagenomic sequencing. *J Adv Res* 26:111–121. <https://doi.org/10.1016/j.jare.2020.07.010>.
- Sinha R, Abu-Alli G, Vogtmann E, Fodor AA, Ren B, Amir A, Schwager E, Crabtree J, Ma S, Abnet CC, Knight R, White O, Huttenhower C, Microbiome Quality Control Project Consortium. 2017. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. *Nat Biotechnol* 35:1077–1086. <https://doi.org/10.1038/nbt.3981>.
- O'Sullivan DM, Doyle RM, Temisak S, Redshaw N, Whale AS, Logan G, Huang J, Fischer N, Amos GCA, Preston MD, Marchesi JR, Wagner J, Parkhill J, Motro Y, Denise H, Finn RD, Harris KA, Kay GL, O'Grady J, Ransom-Jones E, Wu H, Laing E, Studholme DJ, Benavente ED, Phelan J, Clark TG, Moran-Gilad J, Huggett JF. 2021. An inter-laboratory study to investigate the impact of the bioinformatics component on microbiome analysis using mock communities. *Sci Rep* 11:10590. <https://doi.org/10.1038/s41598-021-89881-2>.
- Yeh YC, Needham DM, Sieradzki ET, Fuhrman JA. 2018. Taxon disappearance from microbiome analysis reinforces the value of mock communities as a standard in every sequencing run. *mSystems* 3:e00023–18. <https://doi.org/10.1128/mSystems.00023-18>.
- Allen-Vercoe E, Carmical JR, Forry SP, Gail MH, Sinha R. 2019. Perspectives for consideration in the development of microbial cell reference materials. *Cancer Epidemiol Biomarkers Prev* 28:1949–1954. <https://doi.org/10.1158/1055-9965.EPI-19-0557>.
- Mandal R, Cano R, Davis CD, Hayashi D, Jackson SA, Jones CM, Lampe JW, Latulippe ME, Lin NJ, Lippa KA, Piotrowski P, Da Silva SM, Swanson KS, Wishart DS. 2020. Workshop report: Toward the development of a human whole stool reference material for metabolomic and metagenomic gut microbiome measurements. *Metabolomics* 16:119. <https://doi.org/10.1007/s11306-020-01744-5>.
- de Bruin OM, Birnboim HC. 2016. A method for assessing efficiency of bacterial cell disruption and DNA release. *BMC Microbiol* 16:197. <https://doi.org/10.1186/s12866-016-0815-3>.
- Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, Mailyan A, Manghi P, Scholz M, Thomas AM, Valles-Colomer M, Weingart G, Zhang Y, Zolfo M, Huttenhower C, Franzosa EA, Segata N. 2021. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife* 10:e65088. <https://doi.org/10.7554/eLife.65088>.
- Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34:525–527. <https://doi.org/10.1038/nbt.3519>.
- Gohl DM, Auch B, Certano A, LeFrançois B, Bouevitch A, Doukhanine E, Fragel C, Macklaim J, Hollister E, Garbe J, Beckman KB. 2021. Dissecting and tuning primer editing by proofreading polymerases. *Nucleic Acids Res* 49:e87. <https://doi.org/10.1093/nar/gkab471>.
- Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh HJ, Cuenca M, Hingamp P, Alves R, Costea PI, Coelho LP, Schmidt TSB, Almeida A, Mitchell AL, Finn RD, Huerta-Cepas J, Bork P, Zeller G, Sunagawa S. 2019. Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat Commun* 10:1014. <https://doi.org/10.1038/s41467-019-08844-4>.
- Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol* 20:257. <https://doi.org/10.1186/s13059-019-1891-0>.
- Lu J, Breitwieser FP, Thielen P, Salzberg SL. 2017. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput Sci* 3:e104. <https://doi.org/10.7717/peerj-cs.104>.
- Ye SH, Siddle KJ, Park DJ, Sabeti PC. 2019. Benchmarking metagenomics tools for taxonomic classification. *Cell* 178:779–794. <https://doi.org/10.1016/j.cell.2019.07.010>.
- Parks DH, Rigato F, Vera-Wolf P, Krause L, Hugenholtz P, Tyson GW, Wood DLA. 2021. Evaluation of the Microba Community Profiler for taxonomic profiling of metagenomic datasets from the human gut microbiome. *Front Microbiol* 12:643682. <https://doi.org/10.3389/fmicb.2021.643682>.

31. Edgar RC, Flyvbjerg H. 2015. Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics* 31:3476–3482. <https://doi.org/10.1093/bioinformatics/btv401>.
32. Tourlousse DM, Yoshiike S, Ohashi A, Matsukura S, Noda N, Sekiguchi Y. 2017. Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing. *Nucleic Acids Res* 45:e23. <https://doi.org/10.1093/nar/gkw984>.
33. Prodan A, Tremaroli V, Brolin H, Zwinderman AH, Nieuwoudorp M, Levin E. 2020. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. *PLoS One* 15:e0227434. <https://doi.org/10.1371/journal.pone.0227434>.
34. Williams CR, Baccarella A, Parrish JZ, Kim CC. 2016. Trimming of sequence reads alters RNA-Seq gene expression estimates. *BMC Bioinformatics* 17: 103. <https://doi.org/10.1186/s12859-016-0956-2>.
35. Browne PD, Nielsen TK, Kot W, Aggerholm A, Gilbert MTP, Puetz L, Rasmussen M, Zervas A, Hansen LH. 2020. GC bias affects genomic and metagenomic reconstructions, underrepresenting GC-poor organisms. *Gigascience* 9:gjaa008. <https://doi.org/10.1093/gigascience/gjaa008>.
36. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genom* 3:e000132. <https://doi.org/10.1099/mgen.0.000132>.
37. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
38. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>.
39. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>.
40. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* 27:737–746. <https://doi.org/10.1101/gr.214270.116>.
41. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
42. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
43. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>.
44. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet J* 17:10–12. <https://doi.org/10.14806/ej.17.1.200>.
45. Bushnell B. 2014. BBMap: A fast, accurate, splice-aware aligner. <https://www.osti.gov/biblio/1241166>.
46. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil PA, Hugenholtz P. 2021. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res*.
47. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Thanki N, Yamashita RA, Zheng C, Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res* 46:D851–D860. <https://doi.org/10.1093/nar/gkx1068>.
48. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
49. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
50. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581–583. <https://doi.org/10.1038/nmeth.3869>.
51. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
52. R Core Team. 2020. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
53. Wickham H, François R, Henry L, Müller K. 2021. dplyr: A Grammar of Data Manipulation. R package version 1.0.3. <https://CRAN.R-project.org/package=dplyr>.
54. Wickham H. 2020. tidyr: Tidy Messy Data. R package version 1.1.2. <https://CRAN.R-project.org/package=tidyr>.
55. Wickham H. 2016. ggplot2: Elegant graphics for data analysis. New York: Springer-Verlag.
56. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlenn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2020. vegan: Community Ecology Package. R package version 2.5–7. <https://CRAN.R-project.org/package=vegan>.
57. van den Boogaart KG, Tolosana-Delgado R, Bren M. 2021. compositions: Compositional Data Analysis. R package version 2.0–1. <https://CRAN.R-project.org/package=compositions>.