Published online 15 December 2016

Nucleic Acids Research, 2017, Vol. 45, No. 4 ***e23***
doi: 10.1093/nar/gkw984

Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing

Dieter M. Tourlousse, Satowa Yoshiike, Akiko Ohashi, Satoko Matsukura, Naohiro Noda and Yuji Sekiguchi*

Biomedical Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8566, Japan

Received August 4, 2016; Revised October 11, 2016; Editorial Decision October 12, 2016; Accepted October 24, 2016

ABSTRACT

High-throughput sequencing of 16S rRNA gene amplicons (16S-seq) has become a widely deployed method for profiling complex microbial communities but technical pitfalls related to data reliability and quantification remain to be fully addressed. In this work, we have developed and implemented a set of synthetic 16S rRNA genes to serve as universal spike-in standards for 16S-seq experiments. The spike-ins represent full-length 16S rRNA genes containing artificial variable regions with negligible identity to known nucleotide sequences, permitting unambiguous identification of spike-in sequences in 16S-seq read data from any microbiome sample. Using defined mock communities and environmental microbiota, we characterized the performance of the spike-in standards and demonstrated their utility for evaluating data quality on a per-sample basis. Further, we showed that staggered spike-in mixtures added at the point of DNA extraction enable concurrent estimation of absolute microbial abundances suitable for comparative analysis. Results also underscored that template-specific Illumina sequencing artifacts may lead to biases in the perceived abundance of certain taxa. Taken together, the spike-in standards represent a novel bioanalytical tool that can substantially improve 16S-seq-based microbiome studies by enabling comprehensive quality control along with absolute quantification.

INTRODUCTION

High-throughput sequencing of 16S rRNA gene amplicons (16S-seq) permits efficient characterization of tens to hundreds of microbiota in a single sequencing run (1,2). Combined with fallen costs and increased accessibility of bioinformatics tools (3–5), this has created numerous opportunities for adopting 16S-seq in a range of fields that rely upon detailed microbiota measurements. However, the reproducibility of 16S-seq within and across independent investigations have been documented to be relatively poor (6–8), thereby reducing confidence in 16S-seq data reliability and complicating meta-analysis of independently generated datasets. This is mainly due the wide range of experimental variables that may introduce bias at all steps of a typical 16S-seq workflow, including sample storage and nucleic acid extraction (9–11), primer choice, polymerase chain reaction (PCR) amplification and sequencing platform (12–15), and read data processing and analysis (16,17). As the scale of 16S-seq-based microbiome studies continues to expand and the technology is being increasingly adopted in critical diagnostics settings, a pressing need exists for novel tools that allow routine and comprehensive quality control of the entire measurement procedure, from sample processing to data analysis (18,19).

Further, and measurement biases aside, an important limitation of current 16S-seq procedures is that only relative abundance data are generated, by expressing taxon abundances as proportions of total reads. Interpretation of microbial community dynamics based on solely relative abundances can however be misleading because fluctuations in the absolute abundance of one species may cause an apparent change in the measured (relative) abundance of other species (20,21). Undoubtedly, the availability of straightforward methodologies for quantifying absolute microbial abundances through 16S-seq would be highly beneficial to enable more informative comparative analyses of taxon abundances across samples.

The utilization of synthetic spike-in standards is a promising strategy for addressing some of the technical challenges associated with 16S-seq. Synthetic spike-in standards are relatively well established in the field of RNA-seq (22,23) but have, to the best of our knowledge, not yet been fully explored for 16S-seq. Akin to widely used mock communities, spike-in sequences can serve as ground truths to verify measurement accuracy and reproducibility as well as to evaluate and/or fine-tune bioinformatics pipelines (24–

*To whom correspondence should be addressed. Tel: +81 29 861 7866; Fax: +81 29 861 6400; Email: y.sekiguchi@aist.go.jp

26). A key benefit of spike-in controls, as compared to mock communities, is that the former are added directly to the sample(s) under investigation and hence better assess measurement performance and data quality on a per-sample basis. Concurrently, enumeration of spike-in reads can be used for absolute quantification or read count normalization, based on the known amount of spike-ins added to the samples. Such a strategy, using genomic DNA or cells from selected microorganisms, was recently demonstrated for quantifying total 16S rRNA gene abundances in soil (27) and adjustment of read counts to total microbial loads (28). A drawback of these studies was however that the spike-ins needed to be carefully selected to ensure their absence in the studied microbiomes, such that depending upon the analyzed samples different standards may be required.

In this study, we have developed and tested a set of synthetic spike-in standards for use in 16S-seq experiments. The spike-ins represent artificial 16S rRNA genes with *in silico* designed variable regions lacking identity to nucleotide sequences in public databases, permitting robust tracing of spike-in reads in 16S-seq data from any microbiome sample. Using defined mixtures and environmental microbiota, we performed a series of experiments to characterize the performance of the spike-in standards. Further, we demonstrated the utility of staggered spike-in mixtures for performing quality control on a per-sample basis and simultaneously quantify absolute microbial abundances. Overall, the spike-in standards provide a novel and powerful resource for advancing 16S-seq-based microbiome studies, with their universal nature promoting routine and widespread usage.

MATERIALS AND METHODS

Design and synthesis of the spike-in sequences

Spike-in sequences consisted of conserved regions identical to those of selected natural 16S rRNA genes and artificial variable regions (Figure 1A). The latter were bioinformatically designed starting from randomly generated 12-mers that were progressively concatenated into longer sequences upon evaluation of homopolymer tracts, G+C content and sequence identity (both within and between sequences). This resulted in a set of random sequences (∼2000 bp in length) that satisfied the following criteria: uniform G+C content, no homopolymers of >3 bp, no repeats exceeding 16 bp (as determined by BLAST search of all 1-bp moving window 20-mers) and no self-complementary regions exceeding 10 bp (as determined using Mfold, 29). In addition, the optimized set of artificial sequences contained no between-sequence BLAST hits of >18 bp and shared negligible identity with sequences in NCBI's nt, est and est_human nucleotide sequence databases (web-Blast performed in October 2010). Artificial 16S rRNA genes were then constructed by replacing the variable regions of selected natural 16S rRNA gene sequences (Table 1) with the artificial sequences generated above. Reassessment of the spike-in sequences described in this work by Blast search against more recent releases of a range of NCBI databases (web-Blast performed in September 2016) verified that they shared only negligible identity with known sequences. Complete spike-in sequences are provided in Supplementary Table S1 and are also available in the GenBank/EMBL/DDBJ database under accession numbers LC140931–LC140942. Full-length spike-in sequences (∼1500 bp) were chemically synthesized and inserted into pUC19 plasmid cloning vector by Takara Bio's Dragon Genomics Center (Otsu, Japan).

Preparation of spike-in standards

Plasmid cloning vectors with spike-in sequence inserts were transformed into ECOS Competent *Escherichia coli* JM109 (Nippon Gene, Toiya, Japan) following the manufacturer's instructions. Plasmid DNA was extracted from overnight liquid cultures using the QIAGEN Plasmid Midi Kit. Plasmid DNA was then linearized using the following single-cutting restriction enzymes, according to the manufacturer's instructions: BpmI (New England Biolabs) for spike-ins Ec5001, Ec5002, Ec5005, Ec5502 and Ga5501; BsaI-HF (New England Biolabs) for Ec5003, Ec5004, Ec6001, Bv5501, Ca5501 and Tb5501; and ScaI (TaKaRa Bio) for Ec5501. Linearized plasmid DNA was purified using the Agencourt AMPure XP system (Beckman Coulter) and size and integrity were verified by electrophoresis using the Bioanalyzer 2100 with a DNA 12000 Kit (Agilent). DNA concentrations were determined with a high-sensitivity Quant-iT dsDNA Assay Kit (Invitrogen) using a Qubit Fluorometer 3.0 (Life Technologies). Plasmid DNA was diluted to 10 ng/μl in Tris-EDTA (TE) buffer (pH 8.0) and distributed in single-use aliquots stored at −80°C. Spike-in sequences were verified by Sanger sequencing (see Supplementary Methods for details) and experimentally determined sequences were in all cases in agreement with designed sequences. Spike-in standard mixes were prepared based on estimated copy numbers and stored in TE buffer at −20°C until use.

Mock community preparation

Mock communities were prepared from linearized plasmid DNA containing cloned near-full-length 16S rRNA genes of 15 different bacteria, namely *Nitrobacter winogradskyi* (ATCC 14123), *Nitrosomonas europaea* (ATCC 19178), *Pseudomonas putida* (strain KT2440), *Gemmatimonas aurantiaca* (strain T-27), *Bacillus subtilis* (ATCC 6051), *Desulfovibrio vulgaris* (strain Hildenborough), *Clostridium acetobutylicum* (ATCC 824), *Microlunatus phosphovorus* (strain NM-1), *Anaerolinea thermophila* (strain UNI-1), *Treponema bryantii* (ATCC 33254), *Deinobacter grandis* (DSM 3963), *Bacteroides vulgatus* (ATCC 8482), *E. coli* (strain DH5a), *Chloroflexus aurantiacus* (strain J-10-fl), *Desulfitobacterium hafniense* (strain DCB-2). Details on the preparation of the plasmids are provided in the Supplementary Methods. Inserts were verified by Sanger sequencing and experimentally determined sequences used as references for data analyses. Two types of mock communities were prepared by mixing plasmid DNAs at defined concentrations, namely: (i) an even mock with equimolar amounts of all templates (1.1×10^4 copies/PCR reaction) and (ii) a staggered mock with template concentrations ranging from 7.6×10^1 to 3.8×10^4 copies/PCR reaction (Supplementary Table S4).

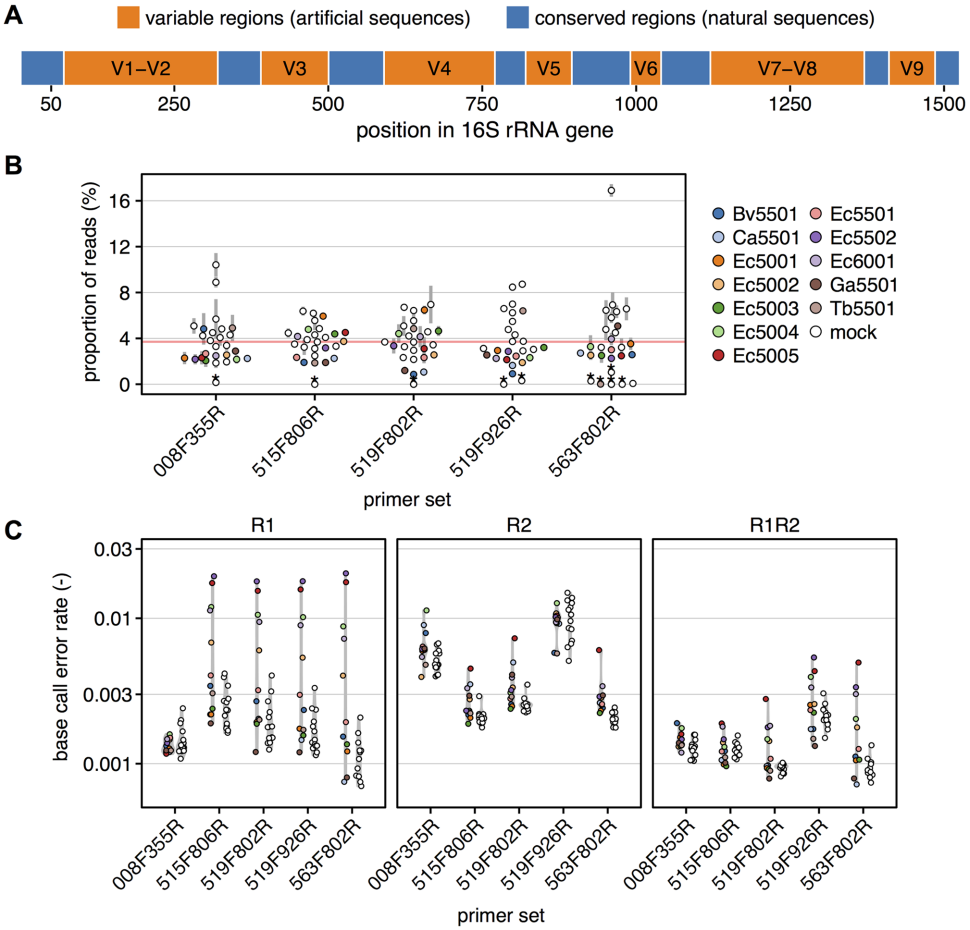

Figure 1. Design and basic characterization of synthetic 16S rRNA gene spike-in standards. (**A**) Layout of the spike-in standards consisting of natural conserved regions and bioinformatically designed (artificial) variable regions. (**B**) Bee swarm plot of the proportion of reads assigned to the different sequence templates in sample E, processed with varying primer sets as shown on the *x*-axis. Data represent the mean and standard deviation of three technical replicates. Sequences with mismatches to the primers are marked with an asterisk. The horizontal red line represents the expected equal proportion (3.7%). Note that all mock community members are represented by empty symbols (marked as 'mock' in the legend). (**C**) Base call error rates for read 1 (R1), read 2 (R2) and merged reads (R1R2), following default read quality trimming (condition 'Qtrim5:15', see Materials and Methods). Three technical replicates were combined and error rates calculated as the ratio of the total number of errors divided by the total number of sequenced bases for each of the templates and primer sets. Note that the same color scheme is used for panels B and C.

Table 1. Summary of the synthetic 16S-seq spike-in standards developed in this study

Identifier	GenBank accession number	Reference 16S rRNA gene sequence, species and strain (GenBank accession number)	Length spike-in (bp)	G+C content spike-in (%)
Ec5001	LC140931	*E. coli* strain ATCC 11775 (X80725, AF233451)	1525	51.3
Ec5002	LC140932	*E. coli* strain ATCC 11775 (X80725, AF233451)	1525	52.1
Ec5003	LC140933	*E. coli* strain ATCC 11775 (X80725, AF233451)	1525	51.7
Ec5004	LC140934	*E. coli* strain ATCC 11775 (X80725, AF233451)	1525	51.7
Ec5005	LC140935	*E. coli* strain ATCC 11775 (X80725, AF233451)	1525	51.5
Ec5501	LC140936	*E. coli* strain ATCC 11775 (X80725, AF233451)	1525	55.3
Ec5502	LC140937	*E. coli* strain ATCC 11775 (X80725, AF233451)	1525	56.2
Ec6001	LC140938	*E. coli* strain ATCC 11775 (X80725, AF233451)	1525	57.2
Bv5501	LC140939	*B. vulgatus* strain JCM 5826 (NR_112946)	1520	55.5
Ca5501	LC140940	*C. acetobutylicum* strain ATCC 824 (X78070)	1495	55.8
Ga5501	LC140941	*G. aurantiaca* strain T-27 (AB072735)	1508	57.9
Tb5501	LC140942	*T. bryantii* strain DSM 1788 (NR_118718, plus in-house generated sequence)	1554	56.2

Environmental samples and DNA extraction

Soil sample was collected from a small forest at the premises of AIST (Ibaraki, Japan) and sludge samples (from an activated sludge system) were obtained from a local municipal wastewater treatment plant (Ibaraki, Japan). Extraction of DNA was performed with the FastDNA SPIN Kit for Soil (MP Biomedicals) according to the manufacturer's instructions, starting from 100 mg of soil or 60 mg of centrifuged (10 min at 16 000 *g*) sludge biomass. If applicable, a defined amount of spike-in standards was added to the samples following cell lysis by bead beating. Size of the extracted DNA was evaluated by gel electrophoresis and yields quantified with a high-sensitivity Quant-iT dsDNA Assay Kit. Extracted DNA was stored in nuclease-free H_2O at −20°C until use.

Quantitative real-time PCR

Total 16S rRNA gene copy numbers in environmental DNA extracts were measured by real-time quantitative PCR (qPCR), using the same primers as used for preparation of the sequencing libraries (Supplementary Table S3), except that Illumina adapters were not present in the qPCR primers. Reactions (20 μl) contained 1× reaction buffer (Power SYBR Green PCR Master Mix; Thermo Fisher Scientific), 500 nM of forward and reverse primer (synthesized by Tsukuba Oligo Service, Tsukuba, Japan) and 5 μl of DNA template. Real-time PCR was performed on a ViiA 7 Real-Time PCR System (Applied BioSystems) with the following thermal cycling conditions: 95°C for 10 min and 40 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 30 s. A standard curve was prepared from a 10-fold serial dilution of linearized pUC19 plasmid DNA containing the near-full-length 16S rRNA gene of *E. coli* strain DH5a. All reactions were set up in triplicate and analysis of the qPCR data was performed using the ViiA 7 Software.

Amplicon library preparation and high-throughput sequencing

Libraries for 16S rRNA gene amplicon sequencing were prepared according to Illumina's dual indexing strategy using a two-step PCR protocol. Various PCR primers previously described in the literature were used, targeting the V1–V2, V4 and V4–V5 variable regions of the 16S rRNA gene (see Supplementary Table S3 for primer sequences and references). The majority of experiments described in the body text were performed with primer sets 515F806R and 008F355R and AmpliTaq Gold LD DNA polymerase; the corresponding experimental procedures are described below and details for additional primer sets and DNA polymerases are provided in the Supplementary Methods. In brief, first-round PCR reactions (15 μl) contained 1× PCR buffer (Applied Biosystems), 250 nM each of forward and reverse primer (synthesized by Tsukuba Oligo Service, Tsukuba, Japan), 200 μM of each deoxynucleotide triphosphate (Applied Biosystems), 0.375 units of AmpliTaq Gold LD (Thermo Fisher Scientific) and 1.5 μl of template DNA. Template DNA consisted of 1.5 ng of environmental DNA or 3×10^5 total copies of plasmid DNA (Supplementary Table S4). Thermal cycling conditions were as follows: enzyme activation for 9 min at 95°C, amplification for 25–30 cycles at 95°C for 45 s, 50°C for 45 s and 72°C for 1 min, followed by final extension for 5 min at 72°C. Amplicons were purified using the Agencourt AMPure XP system (Beckman Coulter) following the manufacturer's protocol. Second-round PCR reactions to attach Nextera (XT) indices and sequencing adapters (Illumina) followed the manufacturer's protocol, with the minor modification that reactions were scaled down to 30 μl. Amplicons were purified as above using the AMPure XP system and quantified by a D1000 or HS D1000 ScreenTape Assay kit using the 2200 TapeStation System (Agilent). Amplicons were mixed in equimolar concentrations and sequenced on an Illumina MiSeq instrument using V2 chemistry, producing 2 × 250 bp paired reads. PhiX DNA was added to the libraries at a concentration of 10–30%. Details about the experimental procedures for 454 sequencing are provided in the Supplementary Methods.

Illumina read data processing and analysis

The following main bioinformatics tools were used: QIIME v1.8.0 (4), Trimmomatic v0.32 (30), Cutadapt v1.8.3 (31) and USEARCH v8.1.1861 (32). Dual barcodes were extracted from the index fastq files and concatenated using the QIIME script extract_barcodes.py. Demultiplexing was performed using the QIIME script split_libraries_fastq.py; no errors were allowed in the barcodes and quality trimming and filtering were suppressed by passing the options -r 251 -p 0 -n 251 -q 0. Minimal filtering of the read data was performed in two steps. Firstly, raw read pairs with a length of <100 bases were eliminated using Trimmomatic (option PE MINLEN:100). Secondly, primer trimming was performed using Cutadapt, allowing up to two mismatches and requiring that primer sequences were anchored at the beginning of the read; reads lacking an identifiable primer sequence were discarded. We then generated three datasets with varying degrees of quality trimming, namely: 'noQtrim': no quality trimming; 'Qtrim5:15': read trimming by Trimmomatic with options SLIDINGWINDOW:5:15 MINLEN75; and 'Qtrim5:20': read trimming by Trimmomatic with options SLIDINGWINDOW:5:20 MINLEN75. The 'Qtrim5:15' condition was set as the default in our pipeline and used unless stated otherwise. Following quality trimming, reads were merged using USEARCH's -fastq_mergepairs command with the following options: -fastq_minovlen 20 -fastq_maxdiffpct 25 -fastq_nostagger -fastq_maxdiffs 9999. A summary of the number of reads retained following each of the processing steps is provided in Supplementary Table S6. Details for processing of 454 read data are provided in the Supplementary Methods.

Direct comparison of processed reads against the reference sequences was performed by global alignment using USEARCH's usearch_global command with the following options: -fulldp -id 0.90 -maxaccepts 9999 -maxrejects 9999 -top_hit_only. Outputs were generated in the Sequence Alignment/Map format for subsequent calculation of base call error rates, as detailed in the Supplementary Methods. Prior to read mapping, putative chimeric sequences were identified using UCHIME (33) in database mode with default settings. For generation of *de novo* operational tax-

onomic units (OTUs), we used the UPARSE (5) pipeline with default settings, unless stated otherwise. Briefly, this included dereplication of the reads; abundance-based sorting of the dereplicated reads, with removal of singletons if applicable; generation of OTUs with a radius of 3%; chimera filtering of OTU centroids using UCHIME in reference mode against the Broad Microbiome Utilities' 16S Gold reference database (version microbiomeutil-r20110519, supplemented with the spike-in sequences); and finally, mapping of all reads against the OTU centroids by usearch_global with default parameters, at 97% sequence identity. Main command lines for read processing and analyses are provided in the Supplementary Methods.

Downstream data analysis

Generated data files were imported into the R statistical computing environment (v3.2.2 available at https://www.R-project.org/; 34) and analyzed using a suite of packages and functions. Random subsampling of OTU tables was performed using the function 'rrarefy', beta diversity was calculated as Bray-Curtis dissimilarities using the function 'vegdist' based on rarefied and square-root transformed OTU counts, unconstrained sample ordination was performed by principle coordinates analysis (PCoA) using the function 'capscale', rarefaction curves of the number of observed OTUs were generated using the function 'rarecurve' and OTU richness was estimated using estimateR, all as implemented in the R package vegan (v2.3-5 available at http://CRAN.R-project.org/package=vegan; 35). Dose-response curves for individual spike-in standards were fitted with a Poisson generalized linear model (GLM), using the function 'glm' of the R package MASS (v7.3-45 available at http://CRAN.R-project.org/package=MASS; 36). Standard curves based on aggregated dose-response curves were generated as negative binomial GLMs by regressing read counts to spike-in amounts using the MASS function 'glm.nb'. Approximate confidence and prediction intervals were generated using the function 'interval' from the R package HH (v3.1-31 available at http://CRAN.R-project.org/package=HH; 37). For absolute quantification, the slope of the negative binomial GLM regression fit was fixed at 1. The intercept of the fitted model was then antilog-transformed and used as scaling factor to convert read counts to absolute copy numbers. Graphics were prepared using R's ggplot2 package (v2.1.0 available at http://CRAN.R-project.org/package=ggplot2; 38). Main R commands are provided in the Supplementary Methods.

Data availability

All raw read data have been deposited in the NCBI sequence read archive (SRA) under accession number SRA434741 (BioProject SRP076838). An overview of all sequencing libraries analyzed in this study is provided in Supplementary Table S5.

RESULTS AND DISCUSSION

Developing a set of synthetic 16S rRNA gene spike-in standards

We generated synthetic 16S rRNA genes by replacing the variable regions of selected natural 16S rRNA genes with *in silico* generated random sequences with negligible identity to nucleotide sequences in public databases (Figure 1A and Supplementary Table S1). In this layout, the conserved regions act as primer binding sites for PCR amplification and the variable regions allow clear identification of spike-in reads during bioinformatics analysis. The set of spike-in standards developed in this work consisted of twelve sequences that were designed based on the 16S rRNA genes of five bacterial species representing phyla that are ubiquitous in diverse environments, namely *E. coli, B. vulgatus*, *C. acetobutylicum*, *G. aurantiaca* and *T. bryantii* (Table 1). The variable regions V1–V2 and V7–V8 were concatenated by omitting the interspacing conserved regions, as illustrated in Figure 1A. The G+C content of the spike-in standards based on *E. coli* varied between 51.3 and 57.2% while the other spike-ins had a G+C content of 55.5 to 57.9% (Table 1 and Supplementary Table S2). Full-length spike-in sequences were chemically synthesized and inserted into a plasmid cloning vector; spike-ins were used in the form of linearized plasmid DNA.

Basic characterization of the spike-in standards

We first evaluated the ability of commonly used PCR primers targeting the V1–V2, V4 and V4–V5 regions of the 16S rRNA gene to amplify the spike-in standards (see Supplementary Table S3 for primer sequences). Agarose gel electrophoresis showed a single band of the expected size for all spike-ins and primer combinations (data not shown), verifying that the random sequences did not compromise amplification specificity. Using this set of primers, we next generated 16S-seq libraries starting from an equimolar pool of the spike-in standards and 15-species plasmid-based mock community; this mixture was designated as sample E in Supplementary Table S4. Unless stated otherwise, sequencing was performed using an Illumina MiSeq, generating 2 × 250 bp paired reads.

Quantitativeness of the 16S-seq data was assessed based on minimally processed reads (see Materials and Methods for details) in order to mitigate sequence-dependent biases that may be introduced upon read processing. Spike-in reads accounted for 35.8±4.2% (range: 31.6–41.0%) of total reads in a given library, which was comparable to the expected proportion of 44.4%. The distribution of reads assigned to each of the reference sequences is shown in Figure 1B for R1 reads; analyses based on R1 and R2 reads were highly consistent (Pearson's r of 0.99; Supplementary Figure S1). Variability in read counts (quantified as the coefficient of variation) ranged from 39.8% (primer set 008F355R) to 55.2% (primer set 519F802R) for the spike-in standards and from 40.4% (primer set 515F806R) to 99.6% (primer set 563F802R) for the mock community members. The higher variability for the mock community was mostly attributed to low read counts for sequence templates with mismatches to the primers (Figure 1B). Similarly, spike-

in standard Tb5501 displayed poor representation in the 563F802R library, due to a single-base mismatch with the 563F primer. Comparison of sample E libraries with those from a mock-community-only sample (sample Ec in Supplementary Table S4) further verified that the spike-in standards did not affect detection efficiencies of the mock community members (Supplementary Figure S2). Additionally, normalized read counts for the spike-ins and mock community members in sample E libraries produced with two types of Taq DNA polymerases were highly correlated (Supplementary Figure S3). In comparison, correlations of spike-in reads in the Taq DNA polymerase libraries with those in a library generated with KOD polymerase were weaker, which was mirrored by weaker correlations for the mock community members in these libraries (Supplementary Figure S3). Taken together, these data verified that the spike-in standards were compatible with 16S-seq; no major quantitative biases were evident and detection rates of the spike-ins were on par with those of the natural 16S rRNA gene sequences in the mock community.

We next evaluated sequence-level quality of the spike-in standards by determining base call error rates for minimally processed reads as well as reads subjected to window-based quality trimming (see Materials and Methods for details). As illustrated in Figure 1C and Supplementary Figure S4, error rates for the spike-in standards were largely comparable with those of the mock community members. Read quality trimming, as expected, resulted in lower error rates and substantial variability in error rates was observed depending on read direction, template and sequenced region (Supplementary Figure S4). Surprisingly, we found that error rates for R1 reads of a number of the *E. coli*-based spike-ins were elevated for primer sets amplifying the V4 region of the 16S rRNA gene (Figure 1C). This was not due to differences with the reference sequences since base call error rates at the corresponding positions in overlapping R2 reads were not higher (Supplementary Figure S5). Additionally, error rate profiles for reads generated by 454 pyrosequencing did not display increased base call inaccuracies (Supplementary Figure S6); this also underscored that sequencing behavior may vary considerably between platforms. Inspection of Phred quality score (Q-score) profiles indicated that Illumina reads with higher base call error rates displayed a conspicuous deterioration in sequence quality (Supplementary Figure S7). Comparison of Q-score profiles from the 515F806R and 563F802R libraries further showed that the decline in base quality occurred at consistent positions (Supplementary Figure S8), suggesting that specific sequence features triggered the sharp deterioration in quality. We note that similar Q-score profiles were also observed for a number of natural 16S rRNA gene sequences in the mock community (e.g. for *G. aurantiaca* amplified with primer set 519F802R; Supplementary Figure S7). In accordance with our findings, sequence-specific biases in read quality have previously also been reported for 16S-seq reads generated using the IonTorrent (12), Roche 454 and Illumina MiSeq platforms (39). While more detailed investigation was beyond the scope of this study, excessive phasing/prephasing triggered by specific sequence characteristics was presumed to be responsible for this effect (40,41). As expected, read merging was effective in partially removing errors associated with low-quality reads, resulting in error rates on the order of several errors per 1000 sequenced bases (Figure 1C). Taken as a whole, these data validated that the sequencing characteristics of the spike-in standards were broadly comparable with those of natural 16S rRNA genes and that the spike-ins may thus be used as references for estimating base call error rates. We further cautiously contend that spike-in sequences with inherently lower raw base qualities may serve as valuable guides to fine-tune read processing parameters; this will be explored in more detail in a subsequent section.

Evaluation of staggered spike-in mixtures

The spike-in standards are intended to be employed as mixtures of multiple plasmid DNAs at a range of concentrations. Similar to strategies adopted for RNA-seq (22,23), this permits dose-response curves to be efficiently generated for each sample based on the read counts of multiple spike-ins with varying input amounts. The resultant dose-response curves can then be analyzed to gauge linearity of the assay and provide an estimate of the limit-of-detection (LOD) of the measurement. Concurrently, scaling factors for absolute quantification can be obtained, with the benefit that the response of multiple spike-ins is taken into account.

We prepared and evaluated four spike-in mixtures; all mixtures had a dynamic range of $\sim 2^{10}$ and their compositions were designed such that each spike-in was evaluated across most of the dynamic range in the different mixtures (Supplementary Table S4). Samples were prepared by combining the spike-in mixtures with the even mock community, mirroring the composition of sample E, and were designated Sb1 to Sb4 in Supplementary Table S4. In addition, a selected spike-in mixture, namely mix 3, was also added to the staggered mock community; the latter had a dynamic range of $\sim 2^{9}$, with template concentration ranging from 7.6×10^{1} to 3.8×10^{4} copies/PCR reaction (sample Q in Supplementary Table S4). Amplicon libraries for the Sb samples were prepared with primer set 515F806R while sample Q was analyzed with both primer sets 515F806R and 008F355R. Read data were processed with default quality trimming settings ('Qtrim5:15') and count data were generated by tallying reads mapped against the reference sequences at a global identity threshold of 97%, after removal of putative chimera by UCHIME.

We inspected dose-response characteristics of individual spike-ins by plotting read counts, after adjustment for uneven sequencing depth by single rarefaction (4600 spike-in reads per library), as a function of spike-in amount across the four Sb samples. For all spike-ins, read counts increased monotonically with spike-in amount (Figure 2A) and Pearson's correlation coefficients were ~0.99 for both level-level and log–log correlations (Supplementary Table S7). Small deviations from linearity could be discerned for a number of spike-in standards (e.g. for spike-in Ec5001 in Supplementary Figure S9). This may be attributed to subtle variations in amplification efficiency across the Sb samples due to their varying initial template proportions and complex interplay among the different templates/amplicons (9,25). As expected, spike-in detection efficiencies quantified as the back-transformed intercept of the Poisson GLM fit with

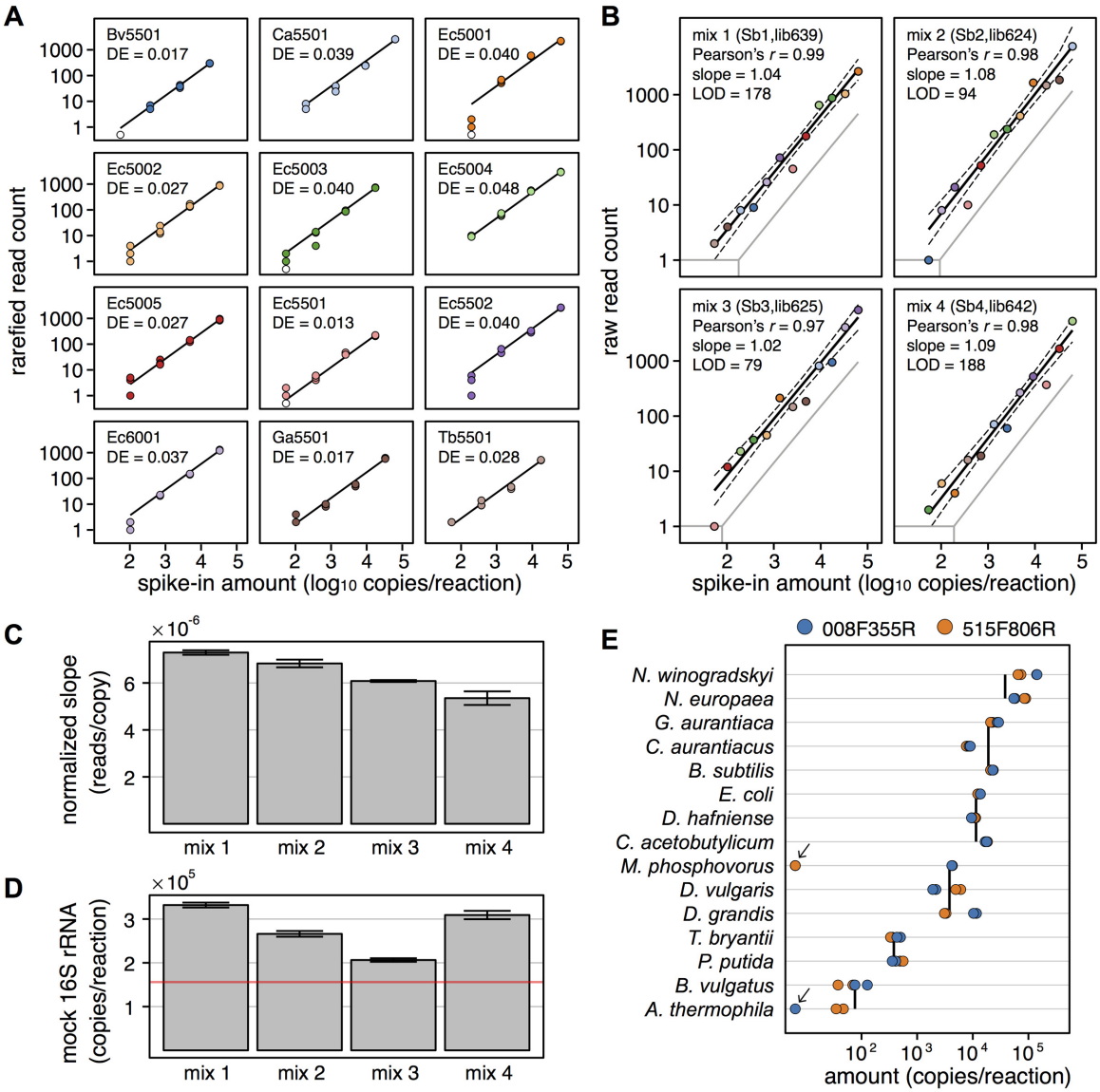

Figure 2. Assessment of the quantitative performance of the spike-in standards. (**A**) Dose-response curves for individual spike-in standards based on their varying amounts in mixes 1 to 4 (samples Sb1–Sb4 in Supplementary Table S4). Symbols show data from individual replicates ($n = 2$ or 3) and empty symbols indicate a read count of zero. Solid black lines display the fitted Poisson GLMs with constrained slope of 1. Indicated detection efficiencies (DEs) were obtained as the antilogarithm of the intercept of the Poisson GLMs and are expressed as number of reads per copy number. (**B**) Aggregated dose-response curves for the different spike-in mixes. Data from a representative library for each of the mixes are shown. Solid black lines represent the negative binomial GLM regression fits and dashed black lines indicate approximate 95% confidence intervals. Solid gray lines show the lower bound of the 90% prediction intervals; the intersection with the horizontal line at a raw read count of one represents the estimated LOD. Symbols are colored based on spike-in identity as in panel A. (**C**) Normalized scaling factors for the different spike-in mixes. Scaling factors were calculated as the antilogarithm of the intercept of negative binomial GLM regression fits with fixed slope of 1, followed by normalization to the total number of spike-in reads for comparison. (**D**) Bar chart of total 16S rRNA gene copy numbers in the even mock community, as estimated based on the standard curve-derived scaling factors. The red line shows the expected value. For panels C and D, data represent the mean and standard deviation of two or three technical replicates. (**E**) Scatter plot of expected (black line segments) and measured (circles) absolute amounts of the different mock community members in sample Q, expressed as copy numbers per PCR reaction. Arrows point to templates with mismatches to the primers. Symbols show data from two technical replicates.

slope constraint were variable among the spike-in standards, consistent with the variable detection rates observed for sample E (Figure 1B). Taken together, these data verified that quantitative detection of the spike-in standards was reliable, validating their further use as quantification standards in 16S-seq.

Strong linear relationships between spike-in read count and input amount were also observed for aggregated dose-response curves based on multiple spike-in standards with varying input concentrations in a single mixture (Figure 2B). The constant relationship between spike-in amount and read count was further evident from the slopes of the negative binomial GLM fits relating read count to spike-in amount (1.08±0.04, Supplementary Table S8). As such, these data indicated that, based on aggregated dose-response curves, enumeration of the spike-in standards was

stable across a dynamic range of at least 2^{10} for concentrations ranging from 5.4×10^{1} and 6.3×10^{4} copies per reaction.

Standard curves for absolute quantification were obtained by negative binomial GLM regression analysis of the aggregated dose-response curves, with the slope of the model fixed at 1. The antilogarithm of the fitted intercept was then used as library-specific scaling factor to convert read counts to absolute copy numbers. As shown in Figure 2C, scaling factors normalized to the total number of spike-in reads deviated, on average, 10.5% from the overall mean across mixtures. We note that scaling factors based on individual spike-ins varied considerably as a results of their differential detection efficiencies (Supplementary Figure S10). This highlighted the benefit of utilizing multiple spike-in standards to 'average out' differential detection rates, which may be expected to reduce systematic biases in measured absolute copy numbers.

The total number of mock community 16S rRNA gene copies in the Sb samples, as estimated based on the standard curve-derived scaling factors, are depicted in Figure 2D. Values varied, on average, 15.1% from the overall mean across mixtures and agreed reasonably well with expected concentrations. Still, systematic overestimation of total 16S rRNA gene copy numbers was observed, ranging from as low as 1.3-fold for mix 3 to up to 2-fold for mix 1. Although deviations from expected abundances may in part be due to inadvertent DNA quantification and/or mixing inaccuracies, it appeared that PCR bias skewed the proportion of spike-in reads (Supplementary Figure S11), resulting in a proportional bias in predicted copy numbers.

Estimated absolute copy numbers of the mock community templates in sample Q are depicted in Figure 2E. Correlation coefficients between measured and expected values were 0.92 and 0.83 for primer sets 515F806R and 008F355R, respectively, and deviated, on average, 1.5-fold and 1.7-fold from the expected concentrations. Except for species that were not detected due to primer mismatches, differences ranged from 1.01-fold (for *D. hafniense* with primer set 515F806R) to 3.74-fold (for *N. winogradskyi* with primer set 008F355R). Cumulatively, the mock community was predicted to contain a total 16S rRNA gene copy number of 2.6×10^{5} copies/reaction (primer set 515F806R) and 3.2×10^{5} copies/reaction (primer set 008F355R). Deviation from the expected value (1.6×10^{5} copies/reaction) appeared again to be attributed to PCR bias, aside from DNA quantification and mixing errors.

In addition to absolute quantification, the relationship between read count and spike-in amount may also be exploited to estimate the LOD of the measurements. Here, the LOD was defined as the copy number for which at least a single read can be expected with 95% probability, as determined based on approximate prediction intervals of the unconstrained negative binomial GLM fits (Figure 2A). For the Sb samples, estimated LODs ranged from 87 to 246 copies/reaction (Supplementary Table S8) and, as expected, a strong linear relationship existed between sequencing depth (total number of spike-in reads) and LOD (Supplementary Figure S12). We point out that LODs corresponded to a fitted read count of 8±1 and may hence be considered as relatively conservative.

Application of the spike-in standards for quality control

Having characterized the spike-in standards using samples with defined composition, we next evaluated the spike-ins with environmental microbiota and demonstrated their utility for quality control (described in this section) and absolute quantification (described in a subsequent section). In short, we obtained three environmental samples (designated as sludge1, sludge2 and soil) and generated 16S-seq libraries from unamended and spiked DNA extracts as well as samples spiked at the point of DNA extraction (Supplementary Table S4). Sequencing libraries were prepared with primer sets 515F806R and/or 008F355R. Unless stated otherwise, reads were processed with default read trimming ('Qtrim5:15') and count data generated based on *de novo* OTU clustering using the UPARSE pipeline, at a nominal identity threshold of 97%. Spike-in standards accounted for 16.5±2.9% of total reads in a given library (Supplementary Table S9).

As shown in Figure 3A for primer set 515F806R, normalized read counts for the 100% spike-in mix were highly consistent with normalized read counts in the spiked environmental DNA extracts. Similarly, relative abundances of individual OTUs in unamended and spiked samples were highly comparable (Figure 3B). The lack of effect of the spike-in standards on measured community structure was also evident from the strongly overlapping PCoA ordinations of spiked and unamended samples (Figure 3C; see Supplementary Figure S13 for primer set 008F355R data). Further, rarefaction curves of the number of observed OTUs were virtually identical for samples with spike-in standards added as compared to unamended samples (Figure 3D), as were Chao1 richness estimates (Figure 3E; see Supplementary Figure S14 for primer set 008F355R data). Taken together, these data validated that: (i) quantitative performance of the spike-in standards was not impacted by complex 16S rRNA gene pools present in the environmental microbiota, and (ii) the spike-in standards did not disturb observed microbial community structure and composition.

Quantitative QC. Addition of a common spike-in mixture to all samples permits assessment of measurement quantitativeness and consistency across libraries, including evaluation of the effect of read data processing. To demonstrate this approach, we subjected a set of libraries from environmental samples amended with spike-in mix 3 and evaluated quantitativeness in terms of the correlation between observed and modeled read counts. The latter were obtained as described above based on negative binomial GLM analysis with slope constraint. As shown in Figure 4A for primer set 515F806R, mild read processing ('noQtrim' and 'Qtrim5:15') yielded strong correlations between observed and predicted read counts, with a narrow distribution of Pearson's r values among libraries. In comparison, more stringent read processing ('Qtrim5:20') consistently reduced correlation strength for all libraries and also led to increased variation among libraries (Figure 4A).

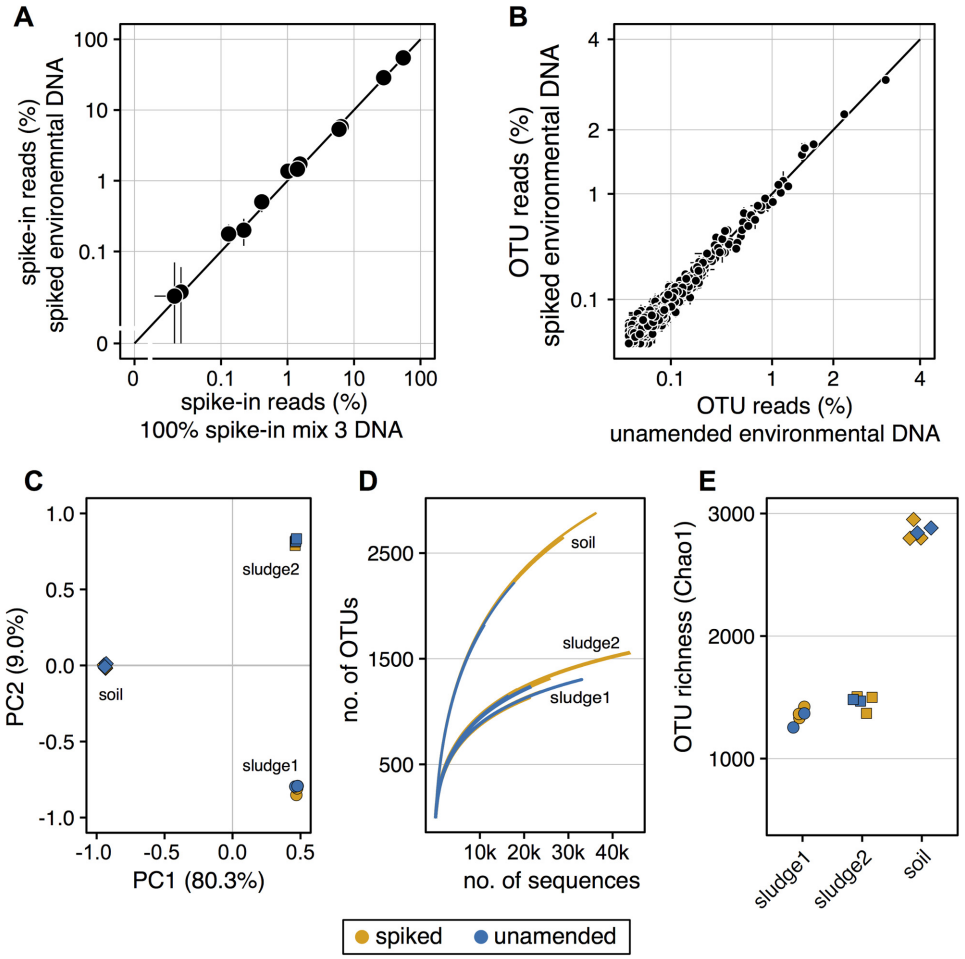

Figure 3. Testing of the spike-in standards with environmental microbiota. (**A**) Scatter plot of normalized spike-in read counts (relative abundance) in 100% spike-in mix 3 (*x*-axis) and spike-in mix added to DNA extracted from soil (*y*-axis). (**B**) Scatter plot of normalized OTU read counts for unamended (*x*-axis) and spiked (*y*-axis) soil DNA. For panels A and B, data represent the mean and standard deviation of three technical replicates and black lines represent the 1-to-1 diagonal. Note that axes in panels A and B were logarithm and square-root transformed, respectively, for visualization purposes. (**C**) Bray-Curtis dissimilarity PCoA ordination plot of unamended and spiked environmental DNA extracts. (**D**) Rarefaction curves showing observed OTU richness as a function of sequencing depth for unamended and spiked DNA extracts. (**E**) Estimated OTU richness (Chao1) of singly rarefied OTU tables for unamended and spiked samples. For panels C, D and E, data for individual libraries are shown ($n = 2$ and $n = 3$ for unamended and spiked samples, respectively). Note that symbol shape indicates sample type (soil, sludge1 and sludge2) and symbol colors distinguish unamended and spiked samples across panels C, D and E. All data are for primer set 515F806R; data for primer set 008F355R are provided in Supplementary Figures S13 and S14.

As observed for sample E above, several spike-ins appeared to be sensitive to varying read trimming settings (Figure 4B), which contributed to the reduced correlations between predicted and modeled read counts when more aggressive read trimming was applied (Figure 4A). Therefore, we sought to evaluate the presence of OTUs with low read quality in our environmental 16S-seq datasets. In short, we generated OTU centroids based on strictly processed reads ('Qtrim5:20') and mapped all reads to this set of high-quality centroids. We found that for some OTUs the number of reads decreased dramatically in response to more aggressive quality trimming (Figure 4C and Supplementary Figure S15), in a fashion similar to that observed for the spike-in standards (Figure 4B). Further, base quality profiles for such OTUs displayed a characteristic decline in read quality (see inset of Figure 4C and Supplementary Figure S16), similar to the Q-score profiles observed for the spike-in standards (Supplementary Figure S7). We note that most OTU centroids (representative sequences) shared >99% sequence identity with sequences in the NCBI nt database and were found multiple times in the read data (Supplementary Table S10), suggesting that they represented genuine sequences rather than artifacts. Based on their unique sequencing characteristics, we contend that the spike-in standards may serve as valuable references for comprehensive assessment of data quality and bioinformatics procedures, although this will be dependent upon sequencing technology and sequenced region.

Qualitative QC. In addition to evaluating quantitative performance, the spike-in standards may also be used to evaluate qualitative accuracy in terms of base call error rates. As illustrated in Figure 4D, we found that more stringent read processing resulted in reduced and more consistent error rates among libraries. Combined with the effect of data processing on read counts, as described above, this

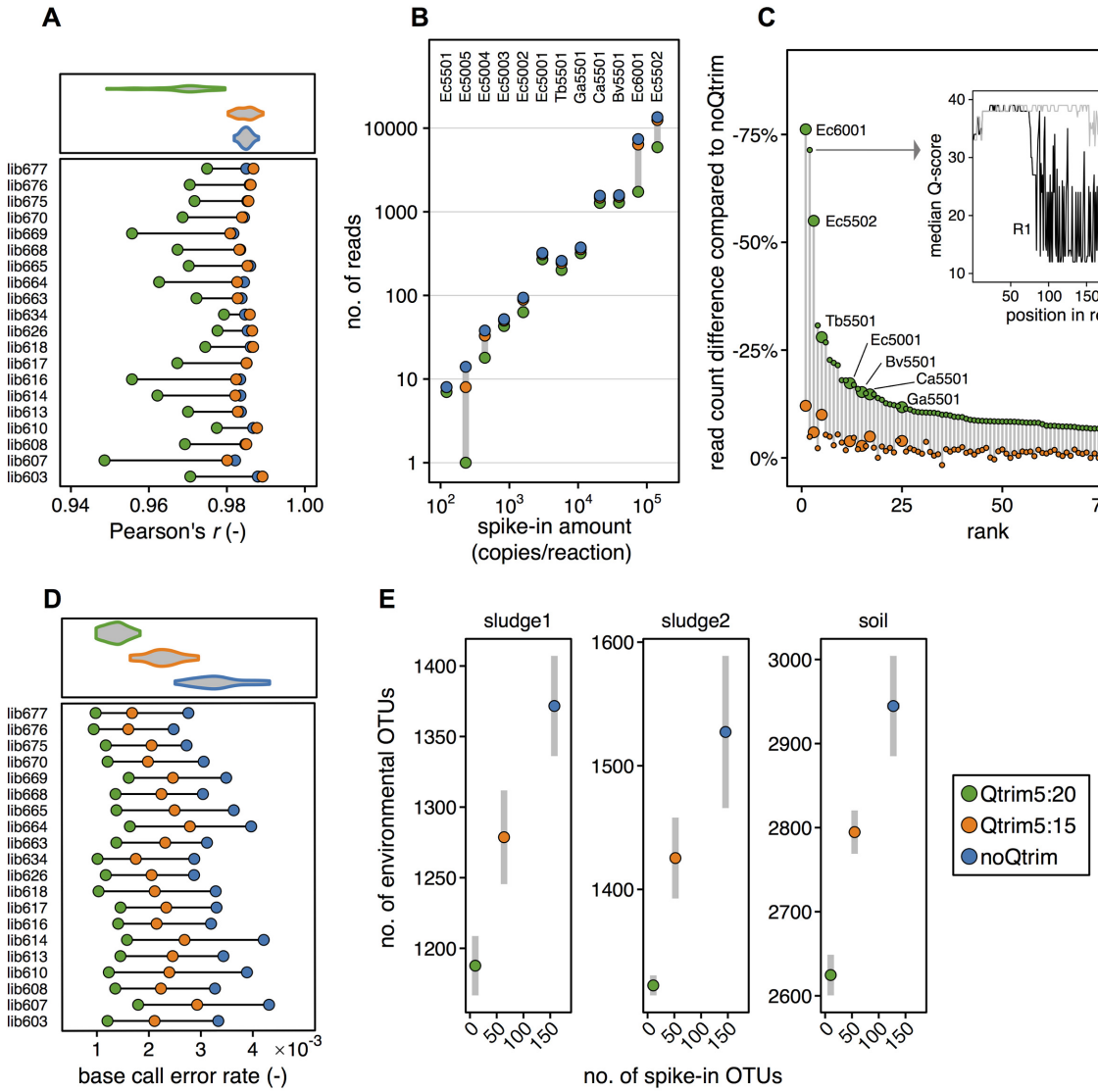

Figure 4. Utility of the spike-in standards for quality control. (**A**) Scatter plot of Pearson's correlation coefficients of the linear relationship between modeled and observed spike-in read counts across libraries, subjected to read processing with varying quality thresholds. The upper portion shows the distribution of Pearson's *r* values as violin plots. Symbols are colored based on the quality threshold applied during read processing and the same color scheme is used for all panels. (**B**) Scatter plot of the dose-response curves for the 100% spike-in mix 3 sample, showing the effect of read processing. Data from a representative library (namely, lib603) are shown. (**C**) Plot illustrating the effect of read processing on read counts for a spiked environmental microbiota sample. Data represent the percentage of reads removed as a result of read processing, as compared to read counts in the 'noQtrim' dataset. Larger symbols represent spike-in-derived OTUs and smaller symbols represent environmental OTUs in sample 'sludge1' (spike-ins added after DNA extraction); data for samples 'sludge2' and 'soil' are provided in the Supplementary Figure S15. Only OTUs with ≥50 reads in the 'noQtrim' data set were analyzed and data represent the mean of three replicated libraries. OTUs are ranked based on the data for set 'Qtrim5:20'. The inset shows the Q-score profiles for OTU250, as indicated by the arrow. (**D**) Scatter plot of base call error rates for the spike-in standards across libraries and subjected to read processing with varying quality thresholds. The upper portion shows the distribution of base call error rates as violin plots. (**E**) Plot of the number of observed environmental OTUs as a function the number of OTUs derived from the spike-in standards, including singletons, following single rarefaction to even depth. Data represent the mean and standard deviation of three technical replicates.

permits identification of appropriate read processing settings based on the compromise between quantitative accuracy and error rates.

In addition to serving as references for estimating error rates, the spike-ins can also be inspected to assess sequence-level accuracy of the generated OTUs as well as correctness of observed OTU richness. Firstly, using our default read processing settings followed by OTU demarcation by UPARSE at 97% sequence identity, all spike-in standards were assigned to 100% accurate OTUs of which the representative sequences were identical to the reference sequences (data not shown). Secondly, when singletons were retained, we found that the number of spike-in OTUs increased with decreasing read processing stringency and this effect was mirrored by a proportional increase in the number of environmental OTUs (Figure 4E). However, adjustment of the

number environmental OTUs based on the number of observed spike-in OTUs was challenging due to the large difference in alpha diversity in both sets.

As a whole, these analyses illustrated that data from spiked microbiota samples are valuable for generating a range of metrics and visualizations to evaluate the performance and consistency of the entire measurement procedure, including bioinformatics procedures.

Utilization of the spike-in standards for absolute quantification

We further assessed the utility of the spike-in standards for absolute quantification of total microbial abundances (that is, 16S rRNA gene copy numbers) in the sludge and soil samples. Briefly, in a first experiment, a common spike-in mix was added to a standard amount of DNA extracted from the different samples and microbial abundances estimated as total 16S rRNA gene copies per ng of DNA, using both primer sets 515F806R and 008F355R. In a second experiment, spike-in mix was added directly to the raw samples in order to quantify 16S rRNA gene copies per unit amount of sample; libraries for this experiment were prepared with primer set 515F806R. Read counts were generated based on *de novo* OTU clustering at 97% sequence identity, as in the previous section. As shown by two illustrative examples in Figure 5A, the spike-in standards captured the full range of read counts for the environmental OTUs. Standard curves for all samples are provided in Supplementary Figure S17 and two examples are shown in Figure 5B. As explained previously, scaling factors for absolute quantification were obtained by negative binomial GLM regression analysis and used to convert read counts to absolute 16S rRNA gene copy numbers.

Total 16S rRNA gene copies per ng of DNA generated based on the spike-in standard curves were generally in good agreement with those obtained by quantitative PCR using the same primer sets (Figure 5C). Across primer sets and samples, 16S-seq-based estimates differed roughly 30% from those obtained by qPCR. For primer set 008F355R, qPCR yielded lower estimates than 16S-seq while the opposite trend was observed for primer set 515F806R. For both primer sets, highest agreement between the 16S-seq and qPCR data was observed for the soil DNA sample, which had higher diversity than the sludge samples. As a whole, these data demonstrated that the spike-in standards provide a valuable tool for determining total microbial abundances in terms of 16S rRNA gene copy numbers, with estimates that were broadly consistent with those generated by qPCR. While we evaluated only a limited number of samples, similar performance can reasonably be expected for other microbiota.

Total microbial abundances estimated based on spike-in standards added prior to DNA extraction are shown in Figure 5D. For comparison, 16S rRNA gene abundances were also calculated based on DNA yield (ng DNA per mg of sample, wet-weight) and the number of 16S rRNA gene copies per ng of DNA (value derived from 16S-seq for comparability). Estimates based on the spike-in standards were roughly 40% higher than yield-based quantities, which was expected given that spike-in standards were able to account for losses during the DNA extraction/purification steps.

Taken together, these results demonstrated the utility of the spike-in standards for quantifying total microbial loads as well as abundances of individual taxa (Figure 5E), expressed in terms of 16S rRNA gene copy numbers per unit of sample. We note however that estimated quantities ought to be interpreted with caution since the spike-in standards do not account for cell lysis efficiency. In addition, PCR bias resulting in skewed proportions of spike-in reads in the 16S-seq data will impact absolute quantities of all OTUs. Although we found differences between spike-in pools to be comparatively small (Figure 2C), it is therefore advisable to employ a common spike-in mixture for all samples to ensure optimal comparability, as was recommended previously (27,28).

Although the agreement between the qPCR and 16S-seq estimates underscored the promise of utilizing spike-in standards for concurrent determination of total 16S rRNA gene abundances and community structure, it is important to recognize that both techniques do not necessarily provide accurate estimates of the actual number of cells in the studied samples. This is due to variations in the number of 16S rRNA gene copy numbers per cell as well as biases introduced during sample preparation for molecular analyses. Single-cell techniques such as fluorescence *in situ* hybridization, although not without their own limitations (42), remain therefore better suited for accurate determination of absolute cell numbers for specific microbial taxa. In comparison, the approach described here provides absolute quantities that are most appropriately used for comparative analysis of taxon abundances across samples processed using the same methodology.

CONCLUSIONS

We have developed and implemented synthetic spike-in standards as a novel tool for advancing 16S-seq. Owing to their unique variable regions and naturally occurring conserved regions, the spike-ins can be applied to any microbiome sample and are compatible with common PCR primers targeting different regions of the 16S rRNA gene. Because synthetic spike-in standards can be added to all samples at different stages of the sample processing and library preparation workflow, they significantly extend the degree of quality control offered by co-sequenced mock communities. Furthermore, by enabling absolute quantification, or equivalently normalization of read counts to total 16S rRNA gene copy numbers (28), the methodology described here will facilitate demarcation of differentially abundant taxa across samples and broadly offer more precise insights into microbial community dynamics (21).

Toward further development, it will be advantageous to design additional spike-in sequences based on the conserved regions of a broader spectrum of microorganisms relevant to different ecosystems. Similarly, expanding the range of G+C contents, both in the conserved and artificial regions, will be useful since G+C content is known to have a considerable effect on PCR amplification efficiency (43). Finally, incorporating spike-in sequences with high sequence identity will allow more informative assessment

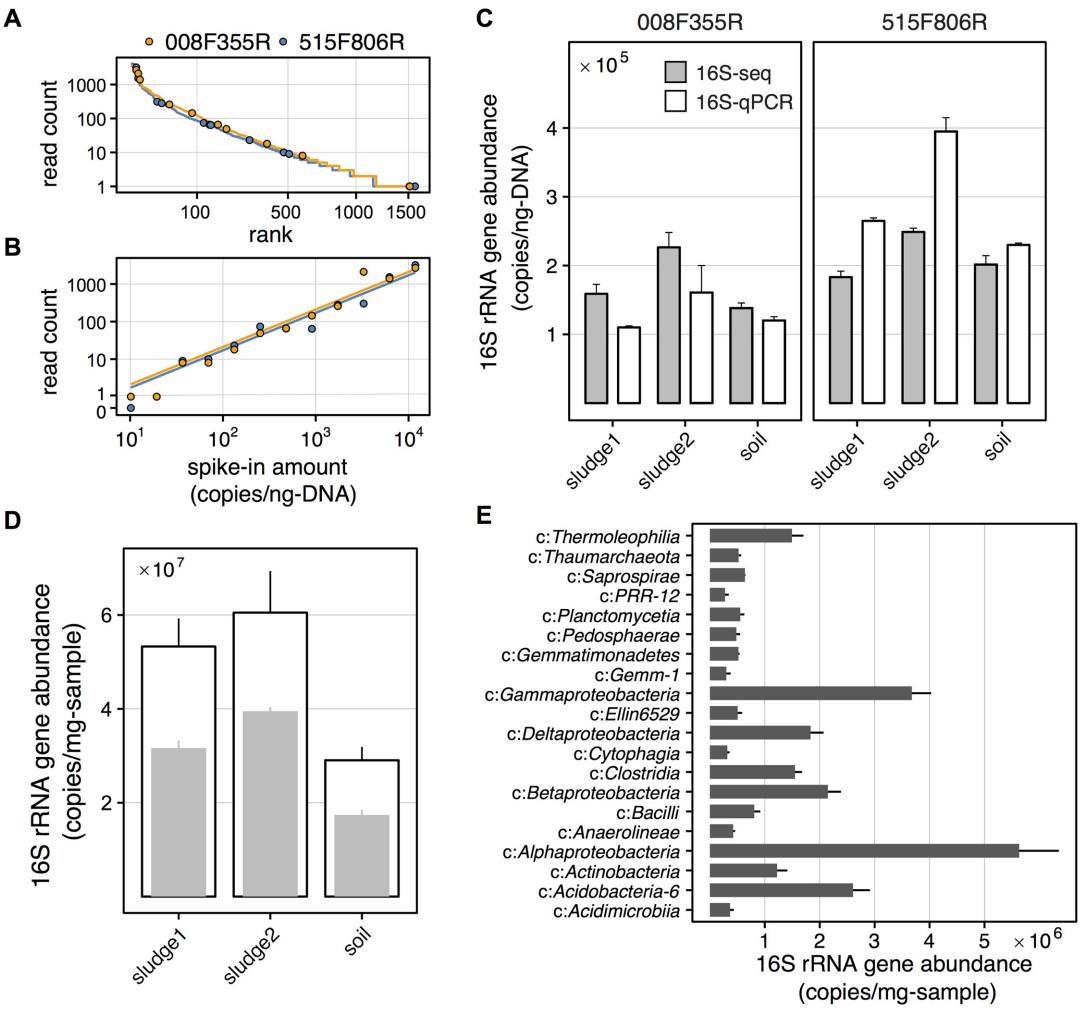

Figure 5. Utility of the spike-in standards for absolute quantification. (**A**) Rank abundance curves illustrating the range of read counts for the spike-in standards and environmental OTUs. Circles represent spike-in standards and are color-coded based on the primer set. (**B**) Representative standard curves for absolute quantification. Lines represent the negative binomial GLM regression fits with fixed slope of 1. For panels A and B, individual data from representative libraries are shown (lib608 and lib671 for primer sets 515F806R and 008F355R, respectively). (**C**) Comparison of total microbial abundances, expressed as 16S rRNA gene copy numbers per ng of DNA, as measured by 16S-seq and qPCR. Data for microbiota in three different environmental samples analyzed with two primer sets are shown. (**D**) Total microbial abundances (16S rRNA gene copy numbers) per unit mass of sample (mg wet-weight). Black empty bars represent estimates based on 16S-seq with spike-in standards added at the point of DNA extraction. Gray bars show the estimates based on DNA yield and 16S rRNA gene copy numbers per ng of DNA. The latter was estimated based on spike-in standards added to a standardized amount of DNA, as shown in panel C. (**E**) Illustrative example of absolute class-level taxon abundances per unit amount of sample (mg wet-weight) as estimated using the spike-in standards, added at the point of DNA extraction. Data for the 'soil' sample are shown. For panels C, D and E, data represent the mean and standard deviation of three technical replicates.

of the performance of OTU clustering and also accommodate emerging algorithms aimed at achieving sub-OTU or single-nucleotide level resolution in 16S-seq read data analysis (44–46).

The observation that certain spike-in displayed comparatively low sequencing quality also highlighted the need to better understand systematic biases and errors in Illumina read data (41). Notably, such reads were also found among environmental sequences, suggesting that taxa-specific biases may be introduced in the resultant data, to an extent determined by the stringency of read processing. The spike-in standards captured a range of read qualities and may hence serve as a useful reference for fine-tuning read processing parameters, with the recognition that this will dependent upon the 16S rRNA gene region sequenced as well as the sequencing technology used.

The spike-in standards can be readily adopted for a range of uses, in addition to routine quality control and quantification. For example, as was recently demonstrated for RNA-seq (23), control ratio mixtures may be used to assess technical performance of differential abundance measurements (Supplementary Figure S18). The spike-ins may further be explored as tools to assure sample identity and lack of contamination, in an approach previously described as SASI-Seq (47); this will be especially valuable in the context of actionable diagnostics based on 16S-seq.

To conclude, we believe that synthetic spike-ins provide a powerful tool to improve the reliability and accuracy of

16S-seq and should be routinely deployed in 16S-seq-based microbiome studies. This is expected to enhance confidence in the resultant data and facilitate their comparison and integration within and across studies, in addition to augmenting information content of the datasets by providing comparative absolute microbial abundances.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors wish to thank Kazunori Nakamura and Norihisa Matsuura for useful discussions.

FUNDING

This work was supported by the Ministry of Economy, Trade and Industry (METI), Japan. Funding for open access charge: National Institute of Advanced Industrial Science and Technology (AIST), Japan.

Conflict of interest statement. Some of the authors (S.M., N.N. and Y.S.) are named as inventors on a patent (Japanese patent application number 2014–089029) related to the spike-in standards presented on this study.

REFERENCES

1. Hamady,M., Walker,J.J., Harris,J.K., Gold,N.J. and Knight,R. (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat. Methods*, **5**, 235–237.
2. Caporaso,J.G., Lauber,C.L., Walters,W.A., Berg-Lyons,D., Huntley,J., Fierer,N., Owens,S.M., Betley,J., Fraser,L., Bauer,M. *et al.* (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.*, **6**, 1621–1624.
3. Schloss,P.D., Westcott,S.L., Ryabin,T., Hall,J.R., Hartmann,M., Hollister,E.B., Lesniewski,R.A., Oakley,B.B., Parks,D.H., Robinson,C.J. *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.*, **75**, 7537–7541.
4. Caporaso,J.G., Kuczynski,J., Stombaugh,J., Bittinger,K., Bushman,F.D., Costello,E.K., Fierer,N., Peña,A.G., Goodrich,J.K., Gordon,J.I. *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods*, **7**, 335–336.
5. Edgar,R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods*, **10**, 996–998.
6. Clooney,A.G., Fouhy,F., Sleator,R.D., O' Driscoll,A., Stanton,C., Cotter,P.D. and Claesson,M.J. (2016) Comparing apples and oranges? Next generation sequencing and its impact on microbiome analysis. *PLoS One*, **11**, e0148028.
7. Hiergeist,A., Reischl,U., Gessner,A. and Priority Program 1656 Intestinal Microbiota Consortium & quality assessment participants (2016) Multicenter quality assessment of 16S ribosomal DNA-sequencing for microbiome analyses reveals high inter-center variability. Int. J. Med. Microbiol., **306**, 334–342.
8. Zhou,J., Wu,L., Deng,Y., Zhi,X., Jiang,Y.H., Tu,Q., Xie,J., Van Nostrand,J.D., He,Z. and Yang,Y. (2011) Reproducibility and quantitation of amplicon sequencing-based detection. *ISME J.*, **5**, 1303–1313.
9. Brooks,J.P., Edwards,D.J., Harwich,M.D. Jr, Rivera,M.C., Fettweis,J.M., Serrano,M.G., Reris,R.A., Sheth,N.U., Huang,B., Girerd,P. *et al.* (2015) The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiol.*, **15**, 66.
10. Choo,J.M., Leong,L.E. and Rogers,G.B. (2015) Sample storage conditions significantly influence faecal microbiome profiles. *Sci. Rep.*, **5**, 16350.
11. McCarthy,A., Chiang,E., Schmidt,M.L. and Denef,V.J. (2015) RNA preservation agents and nucleic acid extraction method bias perceived bacterial community composition. *PLoS One*, **10**, e0121659.
12. Salipante,S.J., Kawashima,T., Rosenthal,C., Hoogestraat,D.R., Cummings,L.A., Sengupta,D.J., Harkins,T.T., Cookson,B.T. and Hoffman,N.G. (2014) Performance comparison of Illumina and Ion Torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. *Appl. Environ. Microbiol.*, **80**, 7583–7591.
13. Pinto,A.J. and Raskin,L. (2012) PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PLoS One*, **7**, e43093.
14. Fouhy,F., Clooney,A.G., Stanton,C. and Claesson,M.J. (2016) 16S rRNA gene sequencing of mock microbial populations- impact of DNA extraction method, primer choice and sequencing platform. *BMC Microbiol.*, **16**, 123.
15. Wu,J.Y., Jiang,X.T., Jiang,Y.X, Lu,S.Y., Zou,F. and Zhou,H.W. (2010) Effects of polymerase, template dilution and cycle number on PCR based 16 S rRNA diversity analysis using the deep sequencing method. *BMC Microbiol.*, **10**, 255.
16. Schmidt,T.S., Matias Rodrigues,J.F. and von Mering,C. (2015) Limits to robustness and reproducibility in the demarcation of operational taxonomic units. *Environ. Microbiol.*, **17**, 1689–1706.
17. Bokulich,N.A., Subramanian,S., Faith,J.J., Gevers,D., Gordon,J.I., Knight,R., Mills,D.A. and Caporaso,J.G. (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods*, **10**, 57–59.
18. Sinha,R., Abnet,C.C., White,O., Knight,R. and Huttenhower,C. (2015) The microbiome quality control project: baseline study design and future directions. *Genome Biol.*, **16**, 276.
19. Stulberg,E., Fravel,D., Proctor,L.M., Murray,D.M., LoTempio,J., Chrisey,L., Garland,J., Goodwin,K., Graber,J., Camille Harris,C. *et al.* (2016) An assessment of US microbiome research. *Nat. Microbiol.*, **1**, 15015.
20. Tsilimigras,M.C.B. and Fodor,A.A. (2016) Compositional data analysis of the microbiome: fundamentals, tools, and challenges. *Ann. Epidemiol.*, **26**, 330–335.
21. Props,R., Kerckhof,F.M., Rubbens,P., De Vrieze,J., Hernandez Sanabria,E., Waegeman,W., Monsieurs,P., Hammes,F. and Boon,N. (2016) Absolute quantification of microbial taxon abundances. *ISME J.*, doi:10.1038/ismej.2016.117.
22. Jiang,L., Schlesinger,F., Davis,C.A., Zhang,Y., Li,R., Salit,M., Gingeras,T.R. and Oliver,B. (2011) Synthetic spike-in standards for RNA-seq experiments. *Genome Res.*, **21**, 1543–1551.
23. Munro,S.A., Lund,S.P., Pine,P.S., Binder,H., Clevert,D.A., Conesa,A., Dopazo,J., Fasold,M., Hochreiter,S., Hong,H. *et al.* (2014) Assessing technical performance in differential gene expression experiments with external spike-in RNA control ratio mixtures. *Nat. Commun.*, **5**, 5125.
24. Pelikan,C., Herbold,C.W., Hausmann,B., Müller,A.L., Pester,M. and Loy,A. (2015) Diversity analysis of sulfite- and sulfate-reducing microorganisms by multiplex dsrA and dsrB amplicon sequencing using new primers and mock community-optimized bioinformatics. *Environ. Microbiol.*, **18**, 2994–3009.
25. Gohl,D.M., Vangay,P., Garbe,J., MacLean,A., Hauge,A., Becker,A., Gould,T.J., Clayton,J.B., Johnson,T.J., Hunter,R. *et al.* (2016) Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat. Biotechnol.*, **34**, 942–949.
26. Kozich,J.J., Westcott,S.L., Baxter,N.T., Highlander,S.K. and Schloss,P.D. (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.*, **79**, 5112–5120.
27. Smets,W., Leff,J.W., Bradford,M.A., McCulley,R.L., Lebeer,S. and Fierer,N. (2016) A method for simultaneous measurement of soil bacterial abundances and community composition via 16S rRNA gene sequencing. *Soil Biol. Biochem.*, **96**, 145–151.
28. Stämmler,F., Gläsner,J., Hiergeist,A., Holler,E., Weber,D., Oefner,P.J., Gessner,A. and Spang,R. (2016) Adjusting microbiome profiles for differences in microbial load by spike-in bacteria. *Microbiome*, **4**, 28.
29. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.

30. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
31. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.*, **17**, 10–12.
32. Edgar,R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.
33. Edgar,R.C., Haas,B.J., Clemente,J.C., Quince,C. and Knight,R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194–2200.
34. R Core Team (2015) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.
35. Oksanen,J., Blanchet,F.G., Friendly,M., Kindt,R., Legendre,P., McGlinn,D., Minchin,P.R., O'Hara,R.B., Simpson,G.L., Solymos,P. *et al.* (2015) Vegan: Community Ecology Package. R Package Version 2.3–5.
36. Venables,W.N. and Ripley,B.D. (2002) *Modern Applied Statistics with S*. 4th Edn., Springer, NY.
37. Heiberger,R.M. (2016) HH: statistical analysis and data display: Heiberger and Holland. R Package Version 3.1–31.
38. Wickham,H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, NY.
39. Puente-Sánchez,F., Aguirre,J. and Parro,V. (2016) A novel conceptual approach to read-filtering in high-throughput amplicon sequencing studies. *Nucleic Acids Res.*, **44**, e40.
40. Nakamura,K., Oshima,T., Morimoto,T., Ikeda,S., Yoshikawa,H., Shiwa,Y., Ishikawa,S., Linak,M.C., Hirai,A., Takahashi,H. *et al.* (2011) Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res.*, **39**, e90.
41. Schirmer,M., Ijaz,U.Z., D'Amore,R., Hall,N., Sloan,W.T. and Quince,C. (2015) Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res.*, **43**, e37.
42. Amann,R. and Fuchs,B.M. (2008) Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. *Nat. Rev. Microbiol.*, **6**, 339–348.
43. Veal,C.D., Freeman,P.J., Jacobs,K., Lancaster,O., Jamain,S., Leboyer,M., Albanes,D., Vaghela,R.R., Gut,I., Chanock,S.J. *et al.* (2012) A mechanistic basis for amplification differences between samples and between genome regions. *BMC Genomics*, **13**, 455.
44. Callahan,B.J., McMurdie,P.J., Rosen,M.J., Han,A.W., Johnson,A.J. and Holmes,S.P. (2016) DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods*, **13**, 581–583.
45. Tikhonov,M., Leach,R.W. and Wingreen,N.S. (2015) Interpreting 16S metagenomic data without clustering to achieve sub-OTU resolution. *ISME J.*, **9**, 68–80.
46. Eren,A.M., Morrison,H.G., Lescault,P.J., Reveillaud,J., Vineis,J.H. and Sogin,M.L. (2015) Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J.*, **9**, 968–979.
47. Quail,M.A., Smith,M., Jackson,D., Leonard,S., Skelly,T., Swerdlow,H.P., Gu,Y. and Ellis,P. (2015) SASI-Seq: sample assurance Spike-Ins, and highly differentiating 384 barcoding for Illumina sequencing. *BMC Genomics*, **5**, 110.