Evaluation of 96-well high-throughput DNA extraction methods for 16S rRNA gene metabarcoding

Marie-Pierre Chapuis, Laure Benoit, Maxime Galan

Received: 22 September 2022 | Revised: 8 March 2023 | Accepted: 2 May 2023

DOI: 10.1111/1755-0998.13812

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

Evaluation of 96-well high-throughput DNA extraction methods for 16S rRNA gene metabarcoding

Marie-Pierre Chapuis[1,2] | Laure Benoit[1,2] | Maxime Galan[3]

[1]CBGP, CIRAD, Montpellier SupAgro, INRAE, IRD, University of Montpellier, Montpellier, France

[2]CIRAD, CBGP, Montpellier, France

[3]CBGP, INRAE, Montpellier SupAgro, INRAE, IRD, University of Montpellier, Montpellier, France

Correspondence
Marie-Pierre Chapuis, CIRAD, UMR CBGP, F-34398 Montpellier, France.
Email: marie-pierre.chapuis@cirad.fr

Funding information
Centre de Coopération Internationale en Recherche Agronomique pour le Développement

Handling Editor: Sebastien Calvignac-Spencer

Abstract

Gaining meaningful insights into bacterial communities associated with animal hosts requires the provision of high-quality nucleic acids. Although many studies have compared DNA extraction methods for samples with low bacterial biomass (e.g. water) or specific PCR inhibitors (e.g. plants), DNA extraction bias in samples without inherent technical constraint (e.g. animal samples) has received little attention. Furthermore, there is an urgent need to identify a DNA extraction methods in a high-throughput format that decreases the cost and time for processing large numbers of samples. We here evaluated five DNA extraction protocols, using silica membrane-based spin columns and a 96-well microplate format and based on either mechanical or enzymatic lysis or a combination of both, using three bacterial mock communities and Illumina sequencing of the V4 region of the 16SrRNA gene. Our results showed that none of the DNA extraction methods fully eliminated bias associated with unequal lysis efficiencies. However, we identified a DNA extraction method with a lower bias for each mock community standard. Of these methods, those including an enzymatic lysis showed biases specific to some bacteria. Altogether, these results again demonstrate the importance of DNA extraction standardization to be able to compare the microbiome results of different samples. In this attempt, we advise for the use of the 96-well DNeasy Blood and Tissue kit (Qiagen) with a zirconia bead-beating procedure, which optimizes altogether the cost, handling time and bacteria-specific effects associated with enzymatic lysis.

KEYWORDS
16 rRNA gene, animal, bacteria, DNA extraction, high-throughput, metabarcoding, next-generation sequencing

1 | INTRODUCTION

Analysis of animal microbiota is a fundamental issue in ecology and evolution since it can help understand how animal species cope with challenges associated with environmental and evolutionary changes (e.g. Groussin et al., 2017; Lindsay et al., 2020). Since microbial symbionts have a mutually beneficial relationship with their host and play important roles in the immune and physiological systems, they likely affect their ecology and evolution (i.e. nutritional ecology and host range, life history, behaviour). Furthermore, the complex microbial communities associated with animals play a significant role in human and animal health and in agriculture,

and their analysis could be useful to discover unexplored pathogens (Galan et al., 2016) and develop future research on biological control innovation (Zhang et al., 2018). Microbial studies can also help shape conservation initiatives in the face of global change (Littleford-Colquhoun et al., 2022). Yet, beyond model systems, current knowledge of microbial communities associated with animals remains limited (Frago et al., 2020; Weinroth et al., 2022). This limited knowledge is partly because recognizing cryptic, diverse and numerous microorganisms hosted by animals is a difficult task. Nevertheless, the design of standard taxonomic barcodes for polymerase chain reaction (PCR) amplification (e.g. 16S ribosomal RNA (rRNA) gene for bacteria identification) and the latest advances in high-throughput sequencing have recently increased the sensitivity and accuracy of microbial communities profiling. Consequently, gene sequencing studies of animal host-associated microbial populations have increased in the past 10 years (Frago et al., 2020; Weinroth et al., 2022).

The potential of microbial metabarcoding is limited by biases introduced by essential processing steps that can modify compositional and abundance data sets (Pollock et al., 2018). The efficiency of the cell lysis step in the DNA extraction protocol is known to have among the largest impact on the apparent abundance of taxa in bacterial communities (Brooks et al., 2015; Sinha et al., 2017). Indeed, bacteria have different structures of cell walls, which enclose their cytoplasm and genomic contents, more or less hard to lyse. Notably, some classes of bacteria carry thick layers of peptidoglycan, like gram-positive bacteria and spore-forming bacteria. Harsh sample treatment could affect DNA quality, particularly for the gram-negative (thin-walled) bacteria, while mild process may cause partial lysis, particularly for gram-positive (thick-walled) bacteria (Bag et al., 2016). Both cocktail of lytic enzymes (Yuan et al., 2012) and bead beating using either silica, ceramic or zirconia beads (de Boer et al., 2010; Javal et al., 2022; Sohrabi et al., 2016) have been shown to improve DNA isolation from gram-positive and spore-forming bacteria (reviewed in Pollock et al., 2018). Based on this observation, many distinct DNA extraction protocols that incorporate mechanical or chemical lysis or a combination of both, including commercial kits dedicated to microbiome applications, are now available. Beyond the type of lysis, other sources of variability between DNA extraction methods are reagent microbial contamination that is ubiquitous and falsely inflates specific richness within samples (Salter et al., 2014) and either low- or high-throughput format that dictates different constraints on manual processing of individual samples, elution volume and cross-contamination. As a result, biases potentially specific to DNA extraction protocols can produce variations in bacterial diversity for the same sample (reviewed in Pollock et al., 2018).

Biases in 16S rRNA studies due to DNA extraction have received much attention over the past decade, but the literature has predominantly focused on protocols for samples that combine the issue of lysing a diverse array of microbes to additional challenges specific to the sample type. Notably, samples from soil (Dopheide et al., 2019; Tiago et al., 2015), plants (Giangacomo et al., 2021) and faeces (Gryp et al., 2020; Hart et al., 2015; Kennedy et al., 2014; Knudsen et al., 2016; Salonen et al., 2010; Wu et al., 2010) are complex environments that contain chemical inhibitors that reduce DNA purity and PCR efficiency (e.g. debris and calcium ions in environmental samples and organic matter such as humic acid, bile salts and polysaccharides in biological samples). Additionally, water samples (Deiner et al., 2015; Djurhuus et al., 2017; Liu et al., 2019; Shi et al., 2020; Wang et al., 2020) and noninvasive tissue samples used in health and disease study or clinical diagnosis (e.g. cuticular, nasal or saliva swabs; Abusleme et al., 2014; Birer et al., 2017; Bjerre et al., 2019; Mattei et al., 2019; Teng et al., 2018; Vesty et al., 2017; Zhou et al., 2019) may contain small quantities of bacterial biomass, that could result in a high risk of contamination from exogenous sources and a low bacterial DNA yield, which can complicate downstream processes.

By contrast, the selection of a suitable procedure for bacterial DNA extraction from samples without inherent technical constraint or potential bias, that is, containing a mid or high bacterial biomass and without specific PCR inhibitors, has not received the same attention. Yet, studies of animal microbiome often focus on the gut where resides the most diverse and abundant microbial assemblage (with the exception of mammals and birds, for which the faeces are often studied instead; but see Čížková et al., 2021) in order to study symbiotic associations (e.g. Javal et al., 2022) or interactions between microbiome and parasites (e.g. Guiver et al., 2022). A major advantage of samples without inherent constraint is that bias in compositional and abundance data of the DNA extraction methods can be assessed based on the analysis of mock community standards of known composition. Indeed, while previous studies proposed some standard DNA extraction methods for a given sample type (e.g. the DNeasy PowerSoil Kit is utilized for buccal and faecal samples in the standardized DNA extraction protocol established by the Human Microbiome Project; https://www.hmpdacc.org), their validation of a bacterial DNA extraction method was limited by the need to rely on real samples of unknown bacterial composition in order to take into account the complexity of the environmental or biological matrix (Greathouse et al., 2019). Although these existing studies compared DNA extraction methods based on DNA yield, total number of passing-filter reads or specific richness within samples, none of these criteria are expected to inform on the accuracy of the microbial community diversity.

Furthermore, previous bacterial DNA bias studies have predominantly focused on DNA extraction based on microfuge tubes rather than on 96-well microplates. While efforts to increase the throughput of bacterial DNA extraction have recently been reported, they mainly concern human microbiomes in the context of clinical studies where samples are time-sensitive (Marotz et al., 2018; Shaffer et al., 2021, 2022). Yet, there is also a need for high-throughput (HTP) sample processing for analyses of large sample sizes in the context of animal studies, in order to decrease both the time and cost of processing, which can be prohibitive for most laboratories studying ecology and evolution (e.g. Abbate et al., 2020). In this attempt, methods based on

silica adsorption membrane are promising since they are known to produce DNA of high quality (e.g. Schiebelhut et al., 2017) and do not require the substantial equipment startup costs of other modern solid-phase methods using magnetic fields beads (such as the ThermoScientific® KingFisher or Maxwell® automated machines). Although some studies have explored the microbial communities' biases due to DNA extraction based on silica membrane spin-columns using microcentrifuge tubes (Gryp et al., 2020; Hart et al., 2015; Kennedy et al., 2014; Tiago et al., 2015; Wu et al., 2010), to our knowledge, none have looked at these biases using a HTP 96-well microplate format.

In this study, we aimed to identify a high-throughput format DNA extraction method that yields, at a reasonable price, fast results with no loss of bacterial taxonomic representation and the least bias in abundance data. We evaluated the performance of five different extraction protocols based on a 96-well microplate format and three different commercial kits using silica adsorption membranes and commonly used to extract bacterial DNA from animal tissues. We chose two relatively expensive commercial kits designed for microbiome or metagenome applications and a cheaper kit dedicated to DNA extraction for animal tissues. We included two homemade procedures based on the former commercial kit, adding a preliminary additional cell lysis procedure, either a bead beating using zirconia beads or an enzymatic lysis. These five HTP 96 DNA extraction protocols were evaluated based on bacterial taxonomic composition and abundance inferred from Illumina amplicon sequencing of the V4 region of bacterial 16SrRNA gene. To this aim, we used three available commercial mock community standards for bacteria whole cells that were contrasted by including either phylogenetically distant or close bacterial strains, which are distributed in either even or log proportions.

2 | MATERIALS AND METHODS

Figure 1 provides a schematic view of our methodological workflow, which is detailed below.

2.1 | Bacterial mock community samples

We used three commercial bacterial mock community standards prepared as a mixture of whole cells of multiple bacteria species fully sequenced, characterized and authenticated. The 20 Strain Even Mix Whole Cell Material standard (ATCC MSA-2002; American Type Culture Collection) contained equal numbers of cells of twenty bacteria species and the Microbial Community Standard I (ZymoBIOMICS D6300) and Standard II (ZymoBIOMICS D6310) contained eight other bacteria species, respectively, in equal or logarithmic numbers of cells (hereafter named ATCC, ZBI and ZBII, respectively; see Table 1 for detailed taxonomic composition). The logarithmic community standard increased dominance of one species (95.9%) and secondary species are in a proportion below 2.8%. Since the 16S rRNA gene varies in the number of copies within a genome (from 2 to 14), the number of 16S rRNA gene copies in each mock community standard does not reflect the number of cells (or genome copies) of bacteria species. We obtained this expected value by correcting the cell proportions provided by the supplier by the number of 16S rRNA gene copies of each bacterial genome (see Table 1 for detailed genomic composition and expectations in gene copies for each cell community standard, hereafter referred as to *EXP*). In order to evaluate PCR and sequencing bias independently of the DNA extraction step, we included genomic-DNA mass-balanced standards (hereafter referred as to *DNA Standard*) of same bacteria for each mock community standard: 20 Strain Even Mix Genomic Material (ATCC MSA-1002), Microbial Community DNA Standard I (ZymoBIOMICS D6305) and DNA Standard II (ZymoBIOMICS D6311).

2.2 | HTP 96 DNA extraction methods

High-throughput DNA extractions were performed according to five different protocols, based on three different commercial kits: the DNeasy PowerSoil HTP 96 DNA Kit (Qiagen; MoBio PowerSoil Kit Cat#: 12955-4) and the ZymoBIOMICS 96 DNA Kit (ZymoResearch Cat#: D4303), both designed for purifying bacterial DNA from a variety of sample inputs, and, the DNeasy 96 Blood and Tissue kit (Qiagen Cat#: 69582) dedicated to DNA extraction for animal tissues and sometimes also used for microbial community analyses from these hosts (e.g. Abbate et al., 2020; Rombaut et al., 2017; Rynkiewicz et al., 2015). Three methods, called PS, ZB and BT, are based on the use of the respective commercial kit following the manufacturer's recommendations with a preliminary step of overnight proteinase K lysis. We also included two homemade modified versions of the BT method. In the method called BTB (for 'Blood & Tissue kit + Beads'), a mechanic cell lysis procedure was added after the proteinase K lysis: the lysate was bead-beaten for 5 min with 500 mg of 0.45/0.55 mm zirconia beads Zirmil® Y (Saint-Gobain), in a TissueLyser II (Qiagen) at the maximum speed setting (30 Hz) during 5 min. The beads were removed after centrifugation at 6000 *g* for 30 s and by pipeting the supernatant in new collection microtubes. In the method called BTE (for 'Blood & Tissue kit + Enzymes'), an enzymatic cell lysis procedure was added as a preliminary step to the proteinase K lysis: cells were suspended in 180 µL ATL buffer of the DNeasy Blood and Tissue kit (Qiagen) with 20 µL lysozyme (10 mg/mL, Sigma-Aldrich), 5 µL mutanolysin (10 KU/mL, A&A Biotechnology) and 0.2 µL lysostaphin (10 mg/mL, Sigma-Aldrich) and incubated for 30 min at 37°C, then 20 min at 50°C. After this additional step, 20 µL of proteinase K was added for overnight lysis at 56°C. The cost of the five methods was calculated per sample using the list price for necessary supplies, reagents and kits as of November 2018 (standard laboratory equipment was excluded). Total time for handling 2×96 samples for each DNA extraction method was calculated for each method by counting the minutes spent

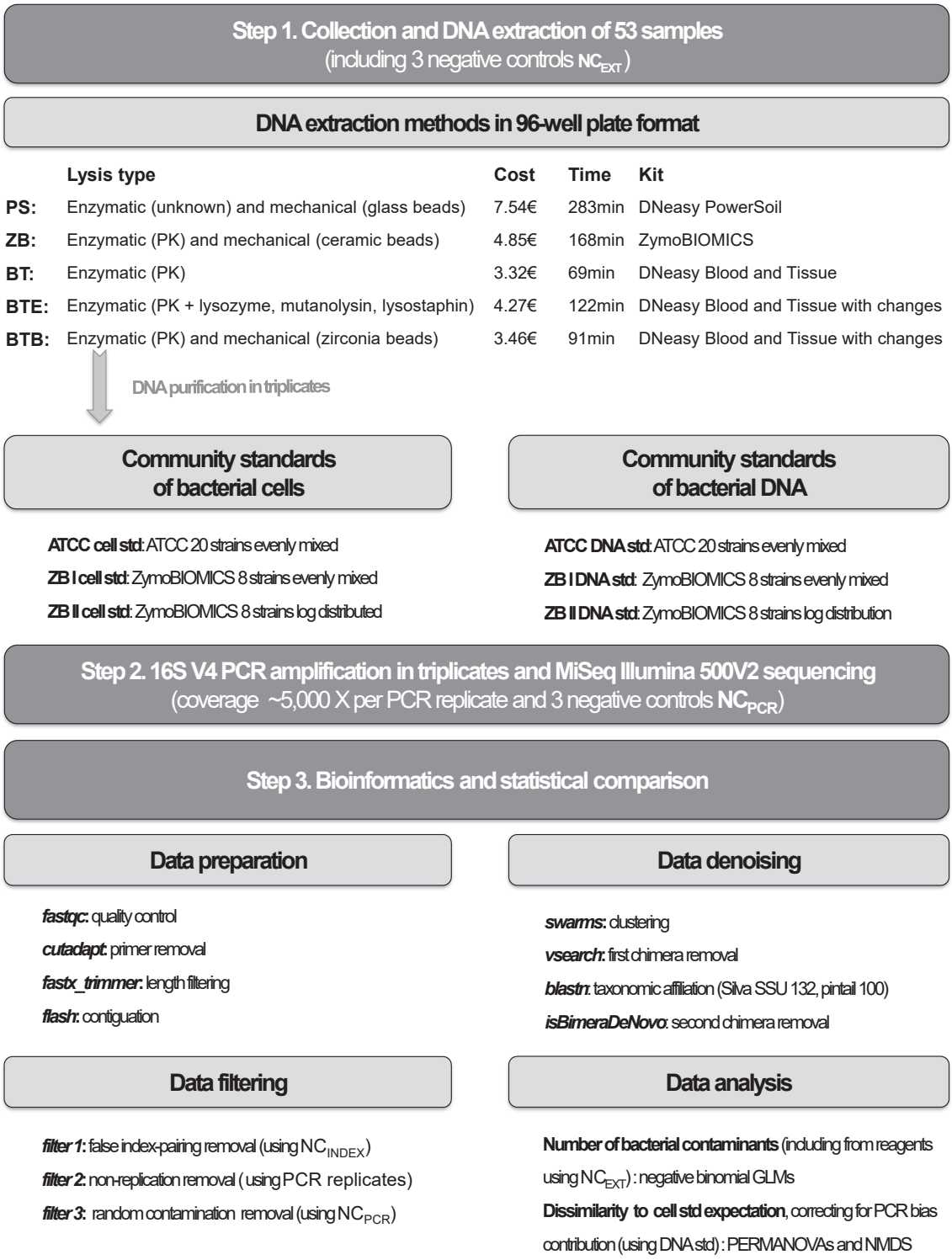

FIGURE 1 Schematic representation of our workflow for the evaluation of five HTP 96 DNA extraction methods, using three mock community standards of bacterial cells. NC_{EXT}, negative controls for DNA extraction; NC_{PCR}, negative controls for PCR; PK, proteinase K.

performing each step of the protocol during laboratory manipulation. DNA extractions were performed in three technical replicates for each method and bacterial cell community standard. The final elution was 200 μL for all methods. The final DNA yield (ng) was assessed using the Qubit® dsDNA HS Assay kit with a Qubit® 2.0 fluorometer.

2.3 | 16S rRNA gene sequencing

We used universal primers to amplify a 251-bp portion of the hypervariable V4 region of the bacterial 16S rRNA gene (16S-V4F:587 GTGCCAGCMGCCGCGGTAA; 16S-V4R: GGACTACH VGGGTWTCTAATCC) and a dual-index method to multiplex the

TABLE 1 Detailed composition of the three commercial community standards used in this study.

Standard ID	ATCC ID	Genbank ID	Species	Gram status	Genome size (Mb)	Gene copy number	Supplier proportion (%)	EXP (%)
ATCC	ATCC 17978	NZ_CP009257.1	*Acinetobacter baumannii*	Negative	4.34	6	5	5.3
ATCC	ATCC 17982	NZ_DS264586.1	*Actinomyces odontolyticus*	Positive	2.39	2	5	1.8
ATCC	ATCC 10987	NC_003909.8	*Bacillus cereus*	Positive	5.42	12	5	10.5
ATCC	ATCC 8482	NC_009614.1	*Bacteroides vulgatus*	Negative	5.16	7	5	6.1
ATCC	ATCC 15703	NC_008618.1	*Bifidobacterium adolescentis*	Positive	2.09	5	5	4.4
ATCC	ATCC 35702	NC_009617.1	*Clostridium beijerinckii*	Positive	6.49	14	5	12.3
ATCC	ATCC 11828	NC_006085.1	*Cutibacterium acnes*	Positive	2.56	4	5	3.5
ATCC	ATCC BAA-816	NC_001263.1	*Deinococcus radiodurans*	Negative	3.29	7	5	6.1
ATCC	ATCC 47077	NC_017316.1	*Enterococcus faecalis*	Positive	3.36	4	5	3.5
ATCC	ATCC 700926	NC_000913.3	*Escherichia coli*	Negative	4.64	7	5	6.1
ATCC	ATCC 700392	NC_000915.1	*Helicobacter pylori*	Negative	1.67	2	5	1.8
ATCC	ATCC 33323	NC_008530.1	*Lactobacillus gasseri*	Positive	1.89	6	5	5.3
ATCC	ATCC BAA-335	NC_003112.2	*Neisseria meningitidis*	Negative	2.27	4	5	3.5
ATCC	ATCC 33277	NC_010729.1	*Porphyromonas gingivalis*	Negative	2.35	4	5	3.5
ATCC	ATCC 9027	NC_009656.1	*Pseudomonas aeruginosa*	Negative	6.26	4	5	3.5
ATCC	ATCC 17029	NZ_AKVW01000001.1	*Rhodobacter sphaeroides*	Negative	4.6	3	5	2.6
ATCC	ATCC BAA-1556	NC_007795.1	*Staphylococcus aureus*	Positive	2.82	6	5	5.3
ATCC	ATCC 12228	NC_004461.1	*Staphylococcus epidermidis*	Positive	2.56	5	5	4.4
ATCC	ATCC BAA-611	NC_004116.1	*Streptococcus agalactiae*	Positive	2.16	7	5	6.1
ATCC	ATCC 700610	NC_004350.2	*Streptococcus mutans*	Positive	2.03	5	5	4.4
ZBI	ATCC 6633	NA	*Bacillus subtilis*	Positive	4.09	9	12.5	17.4
ZBI	ATCC 7080	NA	*Enterococcus faecalis*	Positive	3.08	4	12.5	9.9
ZBI	NA	NA	*Escherichia coli*	Negative	5.10	6–7	12.5	10.1
ZBI	ATCC 14931	NA	*Lactobacillus fermentum*	Positive	2.10	5	12.5	18.4
ZBI	ATCC 19117	NA	*Listeria monocytogenes*	Positive	2.97	5–6	12.5	14.1
ZBI	ATCC 15442	NA	*Pseudomonas aeruginosa*	Negative	6.51	4	12.5	4.2
ZBI	NA	NA	*Salmonella enterica*	Negative	4.86	5–7	12.5	10.4
ZBI	NA	NA	*Staphylococcus aureus*	Positive	2.74	5	12.5	15.5
ZBII	ATCC 6633	NA	*Bacillus subtilis*	Positive	4.09	9	0.89	1.2
ZBII	ATCC 7080	NA	*Enterococcus faecalis*	Positive	3.08	4	0.00089	0.00067
ZBII	NA	NA	*Escherichia coli*	Negative	5.10	6–7	0.089	0.07
ZBII	ATCC 14931	NA	*Lactobacillus fermentum*	Positive	2.10	5	0.0089	0.012

(Continues)

TABLE 1 (Continued)

Standard ID	ATCC ID	Genbank ID	Species	Gram status	Genome size (Mb)	Gene copy number	Supplier proportion (%)	EXP (%)
ZBII	ATCC 19117	NA	*Listeria monocytogenes*	Positive	2.97	5–6	89.1	95.9
ZBII	ATCC 15442	NA	*Pseudomonas aeruginosa*	Negative	6.51	4	8.9	2.8
ZBII	NA	NA	*Salmonella enterica*	Negative	4.86	5–7	0.089	0.07
ZBII	NA	NA	*Staphylococcus aureus*	Positive	2.74	5	0.000089	0.0001

Note: The eight bacterial genomes from ZymoBIOMICS standards can be downloaded at https://s3.amazonaws.com/zymo-files/BioPool/ZymoBIOMICS.STD.refseq.v2.zip. Genome size and 16S rRNA gene copy number per genome for each bacterial species are provided by the supplier. The column 'Supplier proportion' refers to relative proportions in genome copies for ATCC standards whereas it refers to relative proportions in total DNA amount for ZymoBIOMICS. *EXP* refers to relative proportions in 16S rRNA gene copies and was computed considering variation among bacteria in genome sizes and/or in numbers of 16S rRNA gene copies if applicable. When the genome size is variable within the bacterial species, we considered a median value (i.e. for *Escherichia coli*, *Listeria monocytogenes* and *Salmonella enterica*).

Abbreviations: ATCC, standard from the American Type Culture Collection; ZBI, standard from ZymoBIOMICS with bacteria in equal numbers; ZBII, standard from ZymoBIOMICS with bacteria in logarithmic numbers.

PCR products in an Illumina MiSeq sequencing run (see Galan et al., 2016; protocol derived from Kozich et al., 2013). All pre-PCR laboratory manipulations were conducted with filter tips under a sterile hood in a DNA-free room, that is, a room dedicated to the preparation of PCR mix and equipped with hoods that are kept free of DNA by UV irradiation and bleach treatment. DNA amplification was performed in 5 μL of Multiplex PCR kit (Qiagen) master mix, with 4 μL of combined i5 and i7 primers (2.5 μM each) and 2 μL of genomic DNA. The PCR began with an initial denaturation at 95°C for 15 min; followed by 35 cycles of denaturation at 95°C for 20 s, annealing at 55°C for 15 s and extension at 72°C for 5 min; followed by a final extension step at 72°C for 10 min. PCR products (3 μL) were verified by electrophoresis in a 1.5% agarose gel. The putative presence of contamination was checked along the whole laboratory procedure using different negative controls (five negative controls for DNA extraction, that is, one for each protocol, and three negative controls for PCR). A total of 45 DNA samples and eight negative controls were amplified and indexed in three independent PCR reactions using different dual-index combinations and then sequenced on the MiSeq Illumina platform. We performed a run of 2 × 251 bp paired-end sequencing shared with samples of another project.

2.4 | Data denoising and filtering

Bioinformatics steps were performed with the FROGS pipeline (Escudié et al., 2018) available at https://github.com/geraldinepascal/FROGS.git. An exception was the preparation of Illumina reads since the preprocessing step of FROGS fails for data produced following the dual-index method of Kozich et al. (2013). Instead, we used a homemade Shell script (see the section 'Data availability') to trim primers with cutadapt v. 1.9.1 (Martin, 2011) and merge paired-end reads into contigs with FLASH v. 1.2.11 (Magoc & Salzberg, 2011). We then used the FROGS pipeline to sort reads by length (expected value of 251b ± 10b) cluster reads in fine-scale molecular operational taxonomic units (fine-scale OTUs) with SWARM v2 and a local clustering threshold using a d-value of 3 (Mahé et al., 2014, 2015), dereplicate reads, remove chimeras using VSEARCH (Rognes et al., 2016) and affiliate a taxonomy for each fine-scale OTU using NCBI Blast+ on the Silva SSU version 132 database including only the reference reads with a pintail quality score of 100. Finally, since the remaining chimera were visually detected in the abundance table (sequences formed by two more abundant sequences of the same sample), we used the *isBimeraDeNovo* function from the *dada2* package (Callahan et al., 2019) to detect and remove the residual chimeras. Our two-step approach to remove chimeras, which combines first a conservative algorithm (e.g. a cross-sample validation and the absence of parameters for tuning sensitivity) and second a more sensitive algorithm, is likely to be effective while limiting the risk of removing sequences falsely detected as chimeric.

We filtered for false positives generated by other processes than the DNA extraction step, that is, arising from PCR library preparation

or index switching during sequencing. Note that we did not filter for false positives generated by the DNA extraction step in order to evaluate bacterial contaminants that arose from reagents in DNA extraction kits (see below). Filtering was inspired by the strategy proposed by Galan et al. (2016) and automated with homemade R scripts (see the section 'Data Availability'). First, we discarded positive results associated with putative incorrect assignments of reads due to the generation of false index-pairing generated during the Illumina sequencing (Kircher et al., 2012), using a maximum false index-pairing rate for each PCR product of 0.02%, based on estimates from Galan et al. (2016). Then, a given bacterial OTU was considered positive only if the three PCR library replicates were positive (Alberdi et al., 2018; Galan et al., 2016; Robasky et al., 2014). This step allowed us to remove inconsistent OTUs and control for random contamination that may occur during the preparation of PCR 96-well microplates. At this stage, the reads of the replicated PCRs of each bacterial OTU were summed for each of the 53 samples. Finally, we discarded positive results associated with read counts below a bacterial OTU-specific threshold that control for PCR reagent contamination using PCR negative controls.

2.5 | Evaluation of DNA extraction methods

In order to evaluate the HTP 96 DNA extraction methods, we focused on the accuracy of representation of bacterial species, in both composition (i.e. true and false positives) and evenness (i.e. relative abundances). All statistical analyses were performed with the R software v. 4.2.1 and appropriate packages (R Development Core Team, 2022), and data visualization was performed with *ggplot2* (Wickham, 2016).

For each of the five HTP 96 DNA extraction methods (PS, ZB, BT, BTE, BTB), we first counted the bacterial OTUs of the mock community standards that were recovered from Illumina amplicon sequencing of the V4 region of bacterial 16SrRNA gene (true positives). We then identified the bacterial OTUs that were not expected in the samples (false positives) and tested for the effect of the HTP 96 DNA extraction protocol on their number. To this aim, we performed negative binomial generalized linear models (GLMs) for each of the three bacterial mock community standard (ATCC, ZBI, ZBII), using the function *glm.nb* from the *MASS package* (Venables et al., 2002), and tested the factor significance with an analysis of variance (type III), using the *Anova* functions from the *car* package (Fox & Weisberg, 2019). When significant, we used the *emmeans* package (Lenth, 2023) for pairwise comparisons between extraction methods' marginal mean estimates using a Tukey's HSD procedure for multiple testing correction. Furthermore, in an attempt to disentangle bacteria that arose from reagents in DNA extraction kits (i.e. kitome) or from random contamination (i.e. well-to-well, aerosol or investigators contamination) during the DNA extraction step, we used negative samples that control for DNA extraction reagent contamination.

For each mock community standard (ATCC, ZBI, ZBII), we measured the dissimilarity between the observed and the expected relative abundances in 16S rRNA gene copies of bacteria, using the Bray–Curtis dissimilarity, using the function *vegdist* from the *vegan* package (Dixon, 2003). This metric is commonly used in microbiome analyses and is easy to interpret, with 0% indicating that two samples share the exact same relative abundance of each OTU and 100% indicating that two samples share none of the OTUs. Since the Bray–Curtis dissimilarity assumes that the total numbers of reads are equal across samples (Bray & Curtis, 1957), the metric values were computed on rarefied data using the minimum sample size, using the function *rrarefy* from the *vegan* package (Dixon, 2003). While rarefaction guarantees an absence of effect of differences in sample sizes, this procedure is expected to decrease the statistical power for sample comparisons, as a result of omitted read counts and added noise from the random sampling (McMurdie & Holmes, 2014). Thus, preliminary we verified that our sample sizes were large enough to withstand the loss of data and power (see the Results section).

Based on these dissimilarity values, we first represented DNA purification method differences in a reduced dimensional space with a nonmetric multidimensional scaling (NMDS) plot, using the function *metaMDS* from the *vegan* package (Dixon, 2003). We then used nonparametric permutational MANOVA (PERMANOVA) with 1000 permutations to assess for a potential effect of the HTP 96 DNA extraction method, using the function *adonis* from the *vegan* package (Dixon, 2003). When significant, we used post-hoc pairwise comparisons between purification methods using the *pairwiseAdonis* package (Arbizu, 2017). Given a sample size of three DNA extraction replicates only, the statistical power of these pairwise tests is expected to be very low, even for a large size effect (e.g. estimates of 20% and 34% at a threshold of 0.05 and 0.1, respectively, for a size effect of 0.6, using the *pwr.anova.test* function of the *pwr* package; Champely, 2020). Thus, *p*-values near the standard threshold of statistical significance were also considered to interpret the trends (i.e. *p-values* ≤.1). In order to evaluate the contribution of the PCR bias in these dissimilarity values to expectation, we also computed the dissimilarity value to expectation of the abundance table observed from the corresponding manufacturer's DNA standard (*DNA*). Note that our experimental design actually did not allow us to distinguish the PCR bias from the bias related to the sequencing process. However, it is likely that Illumina sequencing plays a limited role in amplicon read abundance bias relative to PCR, as shown for several artificial communities sequenced at an amplicon of the Cytochrome c oxidase I (COI) (see tables 1 and S1 in Galan et al. (2018)).

3 | RESULTS

The total cost and handling time of the five different DNA extraction methods ranged in a similar order (Figure 1), with higher values for the methods based on commercial kits dedicated to bacteriome applications (see details on min counts for each step of the methods' protocol in the Table S1 and details on costs for supplies, reagents and kits of each method in the Table S2). The total handling

time is closely related to the number of steps needed to carry out each method (see Table S1, e.g. BT: 13 steps, 69 min vs. PS: 36 steps, 283 min).

The 53 PCR products, including mock community standard samples and negative controls, in three PCR replicates, generated a total of 646,166 reads and 312 bacterial fine-scale OTUs. After the first filtering step controlling for false index-pairing during sequencing, we kept 645,762 reads representing 258 bacterial OTUs. After the second filtering step considering only OTUs positive in all PCR triplicates, we kept 641,854 reads representing 95 bacterial OTUs. After the third filtering step using PCR negative controls, we kept 638,209 reads, still representing 95 OTUs. Because no OTU was removed, this step indicated the absence of OTUs specific to PCR reagent bacterial contaminants. All samples reached the rarefaction plateau, indicating saturation (see rarefaction curves in the Figure S1) and adequate sample size (i.e. a mean of 12,932 passing-filter 16S reads excluding DNA extraction negative controls). Yet, we detected a significant effect of the HTP 96 DNA extraction protocol on sample sizes by using a negative binomial GLM (*p-values* < .001). Pairwise comparisons between extraction methods' estimated marginal means revealed only a few significant comparisons out of 30 (i.e. ZB vs. BTE and BT vs. BTE for ATCC standard and ZB vs. PS, BT and BTB for ZBI standard and ZB vs. PS and BTB vs. PS for ZBII), which nevertheless suggested that the ZB commercial kit produced slightly higher read numbers (Figure S2), justifying data rarefaction prior to dissimilarity computation (using the minimum sample size, i.e. 8314 passing-filter 16S reads).

3.1 | Bacterial community composition

Taxonomic composition of the DNA samples in Figure 2 showed that all HTP 96 DNA extraction methods recovered the expected bacterial OTUs from all mock community standards. However, we observed a few oddities largely independent of the HTP 96 DNA extraction method (see also Figure S3). As for the ATCC standard of 20 bacteria, the two *Staphylococcus* species (*S. aureus* and *S. epidermis*) share the same sequence of the V4 region of bacterial 16SrRNA gene, leading to 19 expected bacterial OTUs. Furthermore, although used primers perfectly match its sequence, the bacteria *Cutibacterium acnes*, expected to be found in a few hundred of reads with our sequencing coverage, was unexpectedly not found by any of the tested DNA extraction method, nor in the corresponding commercial DNA standard. Satisfactorily, log-distributed taxa of the ZBII standard were detected provided that our deep coverage (i.e. about 13,000 passing-filter reads per sample) was sufficient to detect at least a few reads, with an order of prevalence preserved in most replicates. For example, although *Escherichia coli* and *Salmonella enterica*, were expected at a low proportion of 0.07%, corresponding to an expectation of only 13 reads for our mean sample coverage (about 13,000 passing-filter 16S reads), both bacteria were systematically found. The rare gram-positive bacteria *Lactobacillus fermentum* (expected at 0.012%) at the limit of the detection threshold (a single-read expectation) was detected systematically by the two DNA extraction commercial kits dedicated to bacteriome applications (PS and ZB), sporadically by the BTB method and in none of the replicates of the BT and BTE methods. Unexpectedly, the extremely rare gram-positive bacteria *S. aureus* (expected at 0.0001%) and *Enterococcus faecalis* (expected at 0.00067%) were detected at least in a triplicate by the PS and ZB methods, despite of a null expectation with our coverage. The BT commercial kit and its derivatives did not recover the presence of these bacteria below our detection threshold, with the exception of the BTB method that detected *E. faecalis* once.

Figures 2 and S3 also showed that DNA samples recovered a total of 50 OTUs that were not expected in any of the mock community standards (false positives), with an average of four bacterial contaminants per DNA sample. As for the ATCC mock community standard, we detected a higher false-positive rate for the PS commercial kit (negative binomial GLM; *p-value* = .003 and significant *p-values* for PS vs. BTE and BTB pairwise comparisons). For the ZBI mock community standard with bacteria evenly mixed, no significant difference in the number of bacterial contaminants was found among HTP 96 DNA extraction methods (negative binomial GLM; *p-value* = .09). As for the ZBII mock community standard with log-distributed bacteria, the ZB commercial kit dedicated to microbiome applications showed a higher number of false positives than all other methods (negative binomial GLM; *p-value* = .008), with only the pairwise comparison with PS being significant (*p-value* = .015). Despite their high number, false-positive OTUs generated by the DNA extraction step represented an overall low read proportion, with an average of 1.7%, 0.1% and 1.0% for the ATCC, ZBI and ZBII mock community standards, respectively. Note that such relative abundances will not affect much the Bray–Curtis dissimilarity values, which are not sensitive to differences in the relative abundance for the rarest OTUs.

In addition, 30 foreign bacterial OTUs were detected in the DNA extraction negative controls, including 25 in the single ZB method (see their taxonomy and abundance in Table 2). Other HTP 96 DNA extraction methods led to none, a single or two bacterial OTUs in their DNA extraction negative controls. Out of these 30 reagent contaminants, ten were also found in our DNA samples, with a sequence that either was similar to that of a bacteria expected in commercial community standards (i.e. *Acinetobacter baumannii*, *Enterococcus faecalis*, *Salmonella enterica* of the ZBII mock community standard) or that corresponds to one of the 50 false positives found in our DNA samples (Table 2). Out of these 10 bacterial OTUs observed both in our DNA samples and in the DNA extraction negative controls, six (all in ZB negative controls) would have been removed from the abundance table using the filtering-specific threshold of Galan et al. (2016).

3.2 | Bacterial community evenness

Bray–Curtis dissimilarity values to cell standard expectations (*EXP*) computed on rarefied data ranged from 17% to 52% and were close

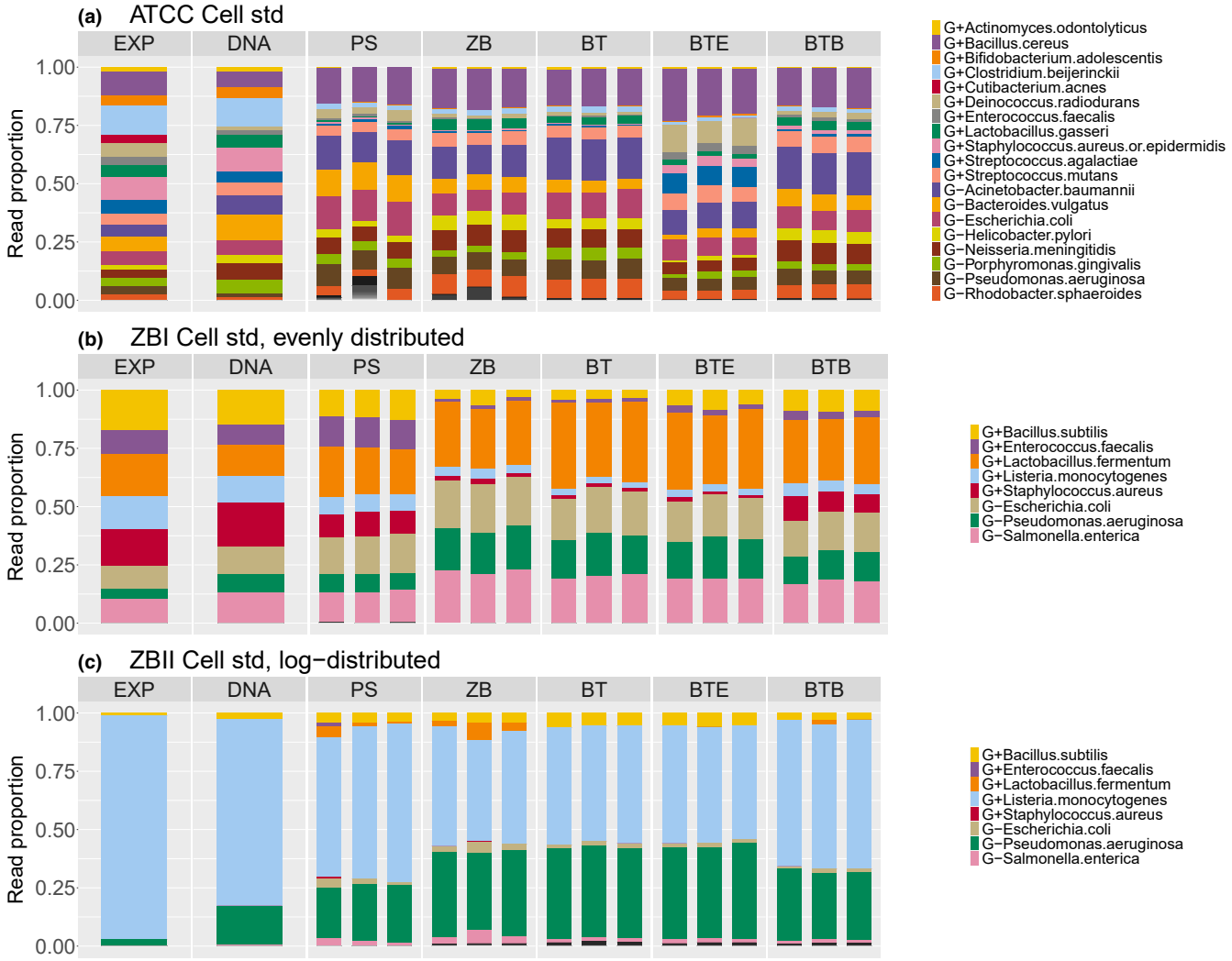

FIGURE 2 Taxonomic composition of the DNA samples obtained from the five HTP 96 DNA extraction methods compared in this study. The relative abundances of expected bacterial OTUs (true positives) are represented by different colours and those of false positives are represented in a grey scale. *EXP* refers to the expected relative proportions in 16S rRNA gene copies of each bacteria (see Table 1 for further details) and DNA to the manufacturer's DNA standard of the corresponding mock community standards. ATCC, standard from the American Type Culture Collection; BTB, DNeasy Blood and Tissue (Qiagen)+mechanic cell lysis procedure; BT, DNeasy Blood and Tissue (Qiagen); BTE, DNeasy Blood and Tissue (Qiagen)+enzymatic cell lysis procedure; PS, DNeasy PowerSoil HTP 96 DNA (Qiagen); ZBII, standard from ZymoBIOMICS with bacteria in logarithmic numbers; ZBI, standard from ZymoBIOMICS with bacteria in equal numbers; ZB, ZymoBIOMICS HTP 96 DNA (ZymoResearch).

to each other for the different commercial community standards (i.e. means after PCR bias correction of 40%, 37% and 41%, for the ATCC, ZBI and ZBII standards, respectively; Figure 3). Since we were primarily interested in the bias due to DNA extraction, we estimated PCR bias contribution to Bray–Curtis dissimilarity values to cell standard expectations, using a DNA standard from the corresponding manufacturer (*DNA*). Dissimilarity values of the DNA standard were 18%, 11% and 15% for ATCC, ZBI and ZBII bacterial mock communities, respectively. Thus, the PCR bias expectedly increased with the number of bacterial taxa present (e.g. higher in the ATCC standard with 20 taxa) and the unevenness of proportions (e.g. higher in the ZBII standard with log-distributed taxa). Overall, the PCR bias accounted for a large third of the dissimilarity to cell standard expectation, which after correction averaged 25% (Figure 3).

Graphical representations of relative abundance data (Figure 2) for the two mock community standards with bacteria in equal numbers showed that this mean level of dissimilarity was partly related to a systematic under-representation of gram-positive bacteria relative to gram-negative bacteria. For the ATCC cell standard, the relative abundances of the gram-positive bacteria *Actinomyces odontolyticus*, *Bifidobacterium adolescentis*, *Clostridium beijerinckii* and *Staphylococcus aureus* and *epidermis*, were systematically under-estimated while relative abundances of the gram-negative bacteria *Escherichia coli*, *Pseudomonas aeruginosa* and *Rhodobacter sphaeroides* were systematically over-estimated (Figure 2a). For the ZBI cell standard, the relative abundance of the gram-positive bacteria *Bacillus subtilis* was systematically under-estimated while relative abundances of the gram-negative bacteria *E. coli*, *P. aeruginosa* and

TABLE 2 Read numbers for the 30 bacterial OTUs detected in the DNA extraction negative controls of the five HTP 96 methods used in this study.

Frogs Cluster ID	Silva taxonomy	NC.EXT.PS	NC.EXT.ZB	NC.EXT.BT	NC.EXT.BTE	NC.EXT.BTB
OTU_5[a]	Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; NA; NA	499	0	0	0	0
OTU_7[a]	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Moraxellaceae; Acinetobacter; NA	0	96	0	0	0
OTU_13[a]	Firmicutes; Bacilli; Lactobacillales; Enterococcaceae; Enterococcus; NA	0	0	0	0	193
OTU_20[b]	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Moraxellaceae; Acinetobacter; NA	0	2634	0	0	0
OTU_22[b]	Proteobacteria; Gammaproteobacteria; Betaproteobacteriales; Burkholderiaceae; NA; NA	0	2432	0	0	0
OTU_36[b]	Proteobacteria; Gammaproteobacteria; Betaproteobacteriales; Burkholderiaceae; Ralstonia; NA	0	878	0	0	0
OTU_37[b]	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Moraxellaceae; Acinetobacter; NA	0	12	0	0	0
OTU_40	Firmicutes; Bacilli; Bacillales; Paenibacillaceae; Paenibacillus; NA	0	0	0	0	704
OTU_41[b]	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Moraxellaceae; Acinetobacter; NA	0	605	0	0	0
OTU_43	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Moraxellaceae; Enhydrobacter; NA	0	0	448	0	0
OTU_64	Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Allorhizobium-Neorhizobium-Pararhizobium-Rhizobium; NA	0	253	0	0	0
OTU_77	Actinobacteria; Actinobacteria; NA; NA; NA; NA	0	0	117	0	0
OTU_79	Proteobacteria; Alphaproteobacteria; Rhizobiales; Xanthobacteraceae; NA; NA	0	175	0	0	0
OTU_86[b]	Proteobacteria; Gammaproteobacteria; Betaproteobacteriales; Burkholderiaceae; NA; NA	0	150	0	0	0
OTU_127	Proteobacteria; Gammaproteobacteria; Betaproteobacteriales; Burkholderiaceae; Aquabacterium; NA	0	89	0	0	0
OTU_133	Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Allorhizobium-Neorhizobium-Pararhizobium-Rhizobium; NA	0	72	0	0	0
OTU_143[b]	Proteobacteria; Gammaproteobacteria; Betaproteobacteriales; Burkholderiaceae; Aquabacterium; NA	0	142	0	0	0
OTU_147	Proteobacteria; Gammaproteobacteria; Betaproteobacteriales; Burkholderiaceae; NA; NA	0	64	0	0	0

TABLE 2 (Continued)

Frogs Cluster ID	Silva taxonomy	NC.EXT.PS	NC.EXT.ZB	NC.EXT.BT	NC.EXT.BTE	NC.EXT.BTB
OTU_173	Proteobacteria; Alphaproteobacteria; Azospirillales; Azospirillaceae; NA; NA	0	42	0	0	0
OTU_192	Proteobacteria; Gammaproteobacteria; Betaproteobacteriales; Burkholderiaceae; Aquabacterium; NA	0	37	0	0	0
OTU_202	Proteobacteria; Alphaproteobacteria; Caulobacterales; Caulobacteraceae; NA; NA	0	33	0	0	0
OTU_209	Proteobacteria; Alphaproteobacteria; Rhizobiales; Pleomorphomonadaceae; Chthonobacter; NA	0	29	0	0	0
OTU_227	Actinobacteria; Actinobacteria; Propionibacteriales; Nocardioidaceae; Nocardioides; NA	0	32	0	0	0
OTU_231	Proteobacteria; Gammaproteobacteria; NA; NA; NA; NA	0	20	0	0	0
OTU_244	Proteobacteria; Alphaproteobacteria; Rhizobiales; Devosiaceae; Devosia; NA	0	21	0	0	0
OTU_247	Proteobacteria; Gammaproteobacteria; Betaproteobacteriales; Burkholderiaceae; Noviherbaspirillum; NA	0	19	0	0	0
OTU_264	Proteobacteria; Alphaproteobacteria; Azospirillales; Azospirillaceae; Azospirillum; NA	0	17	0	0	0
OTU_291	Proteobacteria; Alphaproteobacteria; Caulobacterales; Caulobacteraceae; Brevundimonas; NA	0	14	0	0	0
OTU_344	Proteobacteria; Gammaproteobacteria; Betaproteobacteriales; Burkholderiaceae; Hydrogenophaga; NA	0	19	0	0	0
OTU_436	Proteobacteria; Alphaproteobacteria; Rhizobiales; Stappiaceae; Pannonibacter; NA	0	8	0	0	0

Abbreviations: NC.EXT.BT, DNA extraction negative control for the DNeasy Blood and Tissue (Qiagen); NC.EXT.BTB, DNA extraction negative control for the DNeasy Blood and Tissue (Qiagen) + mechanic cell lysis procedure; NC.EXT.BTE, DNA extraction negative control for the DNeasy Blood and Tissue (Qiagen) + enzymatic cell lysis procedure; NC.EXT.PS, DNA extraction negative control for the DNeasy PowerSoil HTP 96 DNA (Qiagen); NC.EXT.ZB, DNA extraction negative control for the ZymoBIOMICS HTP 96 DNA (ZymoResearch).

[a]OTU present in commercial community standards.

[b]Foreign OTU found in our DNA samples.

Salmonella enterica were systematically over-estimated (Figure 2b). Yet, there were also counter-examples with over-estimation of the relative abundance of some gram-positive bacteria with all HTP 96 DNA purification methods, such as *B. cereus* and *Acinetobacter baumannii* (in the ATCC standard) and *Lactobacillus fermentum* (in the ZBI standard).

Furthermore, the relative bacterial abundances varied significantly depending on the HTP 96 DNA extraction method used (p-values ≤ .003; PERMANOVAs). Overall, across commercial community standards, Bray–Curtis dissimilarity values to cell standard expectation averaged 31% for PS, 35% for BTB, 40% for BTE and 45% for BT and ZB (Figure 3). However, p-values for post-hoc pairwise comparisons (Figure 3) and projections along the NMDS biplot (Figure 4) showed that the performance of DNA extraction methods differed across mock community standards. For the ATCC standard, BTE outperformed other methods with a 10% decrease in dissimilarity values or more, followed by BTB that showed a difference of a single or a few percents (Figure 3; barely significant *p-values* of .1). Accordingly, there was a clear grouping of the BTE triplicates on one side of the biplot and triplicates of all other methods on the other side (Figure 4). Nevertheless, though closer in distance to the theoretical composition of the cell standard and to the observed composition of the DNA standard along the MDS1 axis, BTE was the furthest from the standards on the

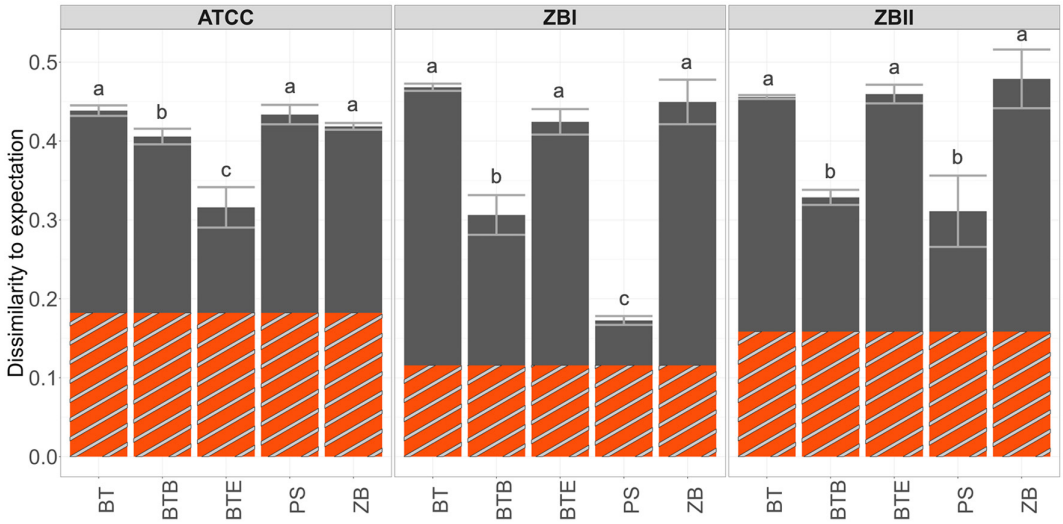

FIGURE 3 Bray–Curtis dissimilarity to expectation from three community standards of bacterial cells, for the five HTP 96 DNA extraction methods compared in this study. The hatched light red colour shows the PCR bias contribution to dissimilarity values to cell standard. It was computed as the Bray–Curtis dissimilarity value to the expectation of the commercial DNA standard of the corresponding community standard. See the legend of Figure 2 for details on acronyms of the five DNA extraction methods studied and the three mock community standards of bacterial cells. Letters show methods' difference based on *p-value* ≤ .1 in the post-hoc pairwise comparisons (see Materials and Methods for justification of this threshold).

MDS2 axis. The BTE profile (Figure 2) differed from other DNA extraction methods, with the negative (positive) bias in the relative abundance of several gram-positive (gram-negative) bacteria partly or mostly corrected (e.g. *Staphylococcus aureus* and *epidermis*, *Enterococcus faecalis*, *Streptococcus agalactiae*, *Acinetobacter baumannii*, *Helicobacter pylori* and *Neisseria meningitidis*). Yet, this was at the expense of a few other bacteria for which the overall bias was increased (e.g. *Bacillus cereus* and *Deinococcus radiodurans*).

For the two ZB mock community standards with 8 bacteria either in equal (ZBI) or logarithm (ZBII) numbers, PS and BTB outperformed other methods (i.e. *p-values* of .1 approaching the standard significant threshold). More precisely, for the ZBI standard, PS showed dissimilarity to expectation 15% lower than BTB that in turn showed dissimilarity to expectation 15% lower than the three other methods (Figure 3). Accordingly, PS triplicates, closely followed by BTB triplicates, were closer to the standards than other methods along the first axis of the NMDS biplot (Figure 4). For the ZBII standard, BTB and PS both showed a mean drop of 15% in dissimilarity values relative to other methods (Figure 3; marginally significant *p-values* of .1). Their triplicates were close to each other and closer to the standards than other methods on the MDS1 axis (Figure 4). Figure 2 also showed that these two DNA extraction methods systematically represented more accurately the gram-positive bacteria *Listeria monocytogenes* and the gram-negative bacteria *Pseudomonas aeruginosa*, the two most abundant bacteria of the ZBII standard with logarithm distribution (i.e. 95.9% and 2.8%). For the ZBI standard with even distribution, the three other under-represented gram-positive bacteria (i.e. *Bacillus subtilis*, *Enterococcus faecalis*, *Staphylococcus aureus*) were also better recovered than with other DNA extraction methods.

4 | DISCUSSION

In this study, we compared five different DNA extraction methods commonly used to purify bacterial total DNA from samples without technical constraints beyond those associated with microbiota. All five methods are based on 96 microplate extraction kits using silica adsorption membranes that allow for decreasing both the cost and time of processing while increasing the throughput (up to 192 samples at the same time), which is of high relevance when large sample sizes are required. Using cell community standards with known species composition and abundances, we were able to evaluate each DNA extraction method on the accuracy of the bacterial community complexity (using compositional and abundance data).

Using Illumina amplicon sequencing of the V4 region of bacterial 16SrRNA gene, we showed a satisfactory recovery of bacteria for all HTP 96 DNA extraction methods. For the mock community standard with a log distribution, we observed that only the two commercial kits dedicated to microbiome applications (PS and ZB) recovered systematically the bacteria *Lactobacillus fermentum* at the very limit of the detection threshold (i.e. about 0.01%) and sporadically the bacteria *Staphylococcus aureus* and *Enterococcus faecalis* far below the detection threshold (i.e. ≤ 0.001%). The commercial kit dedicated to animal tissues (BT) and its version with an additional enzymatic lytic step (BTE) never recovered these rarest bacteria at counts within the statistical background, while the version with an additional mechanical lytic step (BTB) recovered two of them in a single DNA extraction replicate. This result can partly be explained by bias in estimation, since, in the same mock community of bacteria with even distribution, *L. fermentum* was consistently represented in higher relative abundance than expected for all methods and *E. faecalis* was found in higher relative abundance for the PS method. Nevertheless,

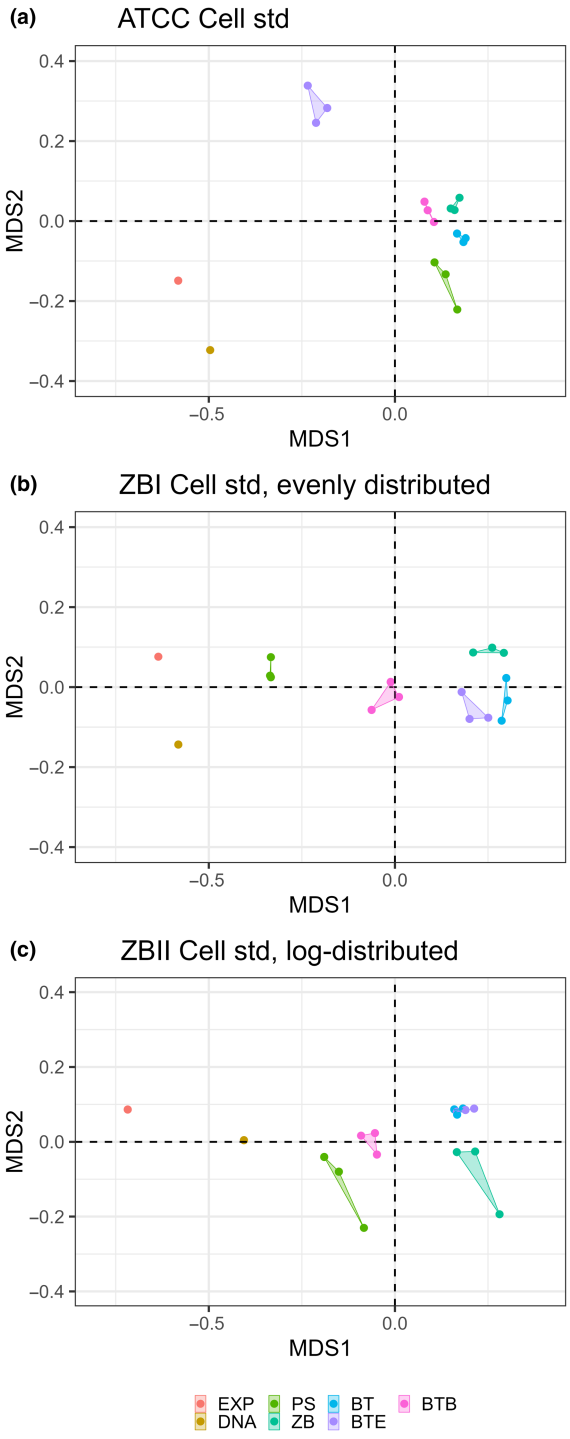

FIGURE 4 Nonmetric multidimensional scaling (NMDS) ordination based on Bray–Curtis dissimilarity of rarefied data for each commercial community standard, with convex hulls of HTP 96 DNA purification methods. (a) Standard from the American Type Culture Collection (stress value = 0.0001). (b) Standard from ZymoBIOMICS with bacteria in equal numbers (stress value = 0.0148). (c) Standard from ZymoBIOMICS with bacteria in logarithmic numbers (stress value = 0.0076). Each symbol represents a sample and colours represented the projections of the expected relative proportions in 16S rRNA gene copies of each bacteria (*EXP*; see Table 1 for further details), the manufacturer's DNA standard of the corresponding community standard (*DNA*) and the five DNA purification methods (see the legend of Figure 2 for details on their acronyms).

it is significant that all HTP 96 DNA extraction methods recovered all bacterial OTUs whose relative abundance was above the detection limit given our coverage (i.e. about 13,000 passing-filter reads per sample), even if this relative abundance was very low (i.e. 0.07%). This is even truer that we applied a conservative approach (i.e. positivity of three PCR replicates) that can precisely remove rare OTUs with a very low abundance (Alberdi et al., 2018).

Fifty false positives were recovered despite our data filtering process for false positives generated by other processes than the DNA extraction step, that is, index switching during sequencing or arising from PCR library preparation. Assuming the purity of commercial mock community standards and the efficiency of our filtering strategy for false positives generated during PCR and sequencing steps, this large number of foreign OTUs could only have been introduced during the DNA extraction step. Such a high level of bacterial contaminants would strongly affect richness estimation and stress the importance of data filtering. Random contamination during the DNA extraction step was suggested by the weak and inconsistent effect of HTP 96 DNA extraction method in numbers of bacterial contaminants, as well as the low repeatability across DNA extraction replicates. Yet, the majority of these exogenous OTUs (i.e. 30) were detected in the negative samples that control for DNA extraction reagent contaminants and were specific for each DNA extraction kit (see Table 2), including 25 in the single ZB method. Interestingly, out of these 30 reagent contaminants, only ten were also found in our DNA samples. This can easily be explained by the fact that contamination disproportionately affects samples with a low biomass of endogenous bacterial DNA (Salter et al., 2014). Nevertheless, the large number of bacteria observed in the ZB kitome is important to consider since such contaminant bacteria introduced prior to PCR may affect the sample bacterial taxonomic composition even in high-biomass samples (Callahan et al., 2017; Minich et al., 2019; Salter et al., 2014). Such bacterial OTUs could be removed from the abundance table for statistical analyses, using for example negative controls for DNA extraction and the corresponding filtering-specific threshold of Galan et al. (2016). In our case, such a filtering step led to the removal of 60% of bacterial contaminants found in DNA samples. Yet, the removal of background contaminants in DNA samples can be unsatisfactory in cases where they genuinely exist in DNA samples. In addition, some of the detected bacteria in DNA extraction negative controls can still originate from random contamination (i.e. well-to-well, aerosol or investigators' contamination) during the DNA extraction step (Davis et al., 2018; Karstens et al., 2019). Consequently, there is a risk that this approach removes a substantial proportion of bacteria that were biologically relevant (e.g. > 20% in Karstens et al., 2019). It is therefore recommended to use DNA extraction kits containing as little bacterial contaminant as possible, for example, BT or PS in this study.

Unsurprisingly, none of the five HTP 96 DNA extraction methods did recover the expected relative numbers of 16S rRNA gene copies. After correction for PCR bias contribution, we observed an

overall moderate level of dissimilarity to expectation similar across mock community standards (about 25%). The cell wall architecture was the greatest contributor to dissimilarity values, with a systematic under-representation of some gram-positive bacteria associated with the over-representation of gram-negative bacteria. This result suggests that none of these optimized HTP 96 DNA extraction method, even those dedicated to bacteriome applications, lysed gram-positive bacteria with a similar efficiency to gram-negative bacteria. Furthermore, a major finding was that this overall level of dissimilarity in bacterial abundances differed across HTP 96 DNA extraction methods in a dependent manner with the bacterial taxonomic composition. BTE outperformed other methods in correcting for the lysis bias between gram-positive and gram-negative bacteria of the ATCC community standard (at least for 6 out of the 20 bacteria in presence) while PS and BTB outperformed other methods in correcting the lysis bias for the eight bacteria of the ZBI and ZBII community standards, which differed by the relative abundance distribution (either even or logarithmic). Finally, we observed variations unique to a HTP 96 DNA extraction method, such as over-representation of some gram-positive bacteria, which tend to be associated with DNA extraction methods including enzymatic lysis additional to the standard proteinase K (e.g. *Bacillus cereus* and *Deinococcus radiodurans* with BTE for the ATCC standard and *Enterococcus faecalis* with PS for the ZBI and ZBII standards).

Altogether, these results again demonstrate the importance of DNA extraction standardization in microbiome analyses. While the best practice remains to use an identical DNA extraction protocol for all samples in comparisons, it is difficult to advise on which DNA extraction protocol to choose, since our study did not identify a HTP 96 DNA extraction protocol with systematically lower measures of bias than others for the different tested cell mock community standards. However, of the methods that stand out in either bacterial community (BTB, BTE and PS), the BTB method seems to be the best compromise, since it avoids the enzymatic procedure that increases the risk of specific effects on some gram-positive bacteria but also had the lowest cost (i.e. 3.46€ per sample vs. 4.27€ and 7.54€, respectively) and handling time (91 min for 2×96 samples vs. 122 min and 283 min, respectively). Overall, our study confirmed that the vigorous beating with beads is helpful for sample homogenization and mechanical cell lysis of bacteria (reviewed in Pollock et al., 2018). This result is of broad applicability in animal ecological studies, at least when tissue samples contain a mid- or high bacterial biomass and no PCR inhibitors but is all the more important for complex high diversity samples (e.g. gut samples of mammals) where there is a higher risk of missing gram-positive taxa due to bacterial lysis bias.

AUTHOR CONTRIBUTIONS

LB, MG and MP designed the study; LB performed DNA extractions, quality control and sequencing; MPC wrote the Shell and R scripts, performed data analyses and wrote the manuscript with contributions from LB and MG.

ACKNOWLEDGEMENTS

MPC and LB were supported by the Insect Microbiome Project, funded by the French Agricultural Research Centre for International Development (CIRAD). Some of the data used in this work were produced through the GenSeq technical facilities of the Institut des Sciences de l'Evolution de Montpellier with the support of LabEx CeMEB, an ANR "Investissements d'avenir" program (AN-R-1-0-LAB-X-0-4-01).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Bacterial abundance data from MiSeq Illumina, Shell script for preprocessing of raw sequences and R scripts for data filtering and analyses were deposited in Cirad data repository (https://dataverse.cirad.fr/) and are available from this link: https://doi.org/10.18167/DVN1/D31UAV.

ORCID

Marie-Pierre Chapuis https://orcid.org/0000-0003-4405-7722
Laure Benoit https://orcid.org/0000-0003-3740-5346
Maxime Galan https://orcid.org/0000-0001-6981-5732

REFERENCES

Abbate, J. L., Galan, M., Razzauti, M., Sironen, T., Voutilainen, L., Henttonen, H., Gasqui, P., Cosson, J.-F., & Charbonnel, N. (2020). Pathogen community composition and co-infection patterns in a wild community of rodents [preprint]. *Ecology*. https://doi.org/10.1101/2020.02.09.940494

Abusleme, L., Hong, B.-Y., Dupuy, A. K., Strausbaugh, L. D., & Diaz, P. I. (2014). Influence of DNA extraction on oral microbial profiles obtained via 16S rRNA gene sequencing. *Journal of Oral Microbiology*, *6*, 1. https://doi.org/10.3402/jom.v6.23990

Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, *9*(1), 134–147. https://doi.org/10.1111/2041-210X.12849

Arbizu, P. M. (2017). pairwiseAdonis: Pairwise Multilevel Comparison using Adonis. R package version 0.4.1.

Bag, S., Saha, B., Mehta, O., Anbumani, D., Kumar, N., Dayal, M., Pant, A., Kumar, P., Saxena, S., Allin, K. H., Hansen, T., Arumugam, M., Vestergaard, H., Pedersen, O., Pereira, V., Abraham, P., Tripathi, R., Wadhwa, N., Bhatnagar, S., … Das, B. (2016). An Improved Method for High Quality Metagenomics DNA Extraction from Human and Environmental Samples. *Scientific Reports*, *6*(1), 26775. https://doi.org/10.1038/srep26775

Birer, C., Tysklind, N., Zinger, L., & Duplais, C. (2017). Comparative analysis of DNA extraction methods to study the body surface microbiota of insects: A case study with ant cuticular bacteria. *Molecular Ecology Resources*, *17*(6), e34–e45. https://doi.org/10.1111/1755-0998.12688

Bjerre, R. D., Hugerth, L. W., Boulund, F., Seifert, M., Johansen, J. D., & Engstrand, L. (2019). Effects of sampling strategy and DNA extraction on human skin microbiome investigations. *Scientific Reports*, *9*(1), 17287. https://doi.org/10.1038/s41598-019-53599-z

Bray, J. R., & Curtis, J. T. (1957). An ordination of the upland Forest communities of southern Wisconsin. *Ecological Monographs*, *27*(4), 325–349. https://doi.org/10.2307/1942268

Brooks, J. P., Edwards, D. J., Harwich, M. D., Rivera, M. C., Fettweis, J. M., Serrano, M. G., Reris, R. A., Sheth, N. U., Huang, B., Girerd, P., Strauss, J. F., Jefferson, K. K., & Buck, G. A. (2015). The truth about metagenomics: Quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiology*, *15*(1), 66. https://doi.org/10.1186/s12866-015-0351-6

Callahan, B. J., DiGiulio, D. B., Goltsman, D. S. A., Sun, C. L., Costello, E. K., Jeganathan, P., Biggio, J. R., Wong, R. J., Druzin, M. L., Shaw, G. M., Stevenson, D. K., Holmes, S. P., & Relman, D. A. (2017). Replication and refinement of a vaginal microbial signature of preterm birth in two racially distinct cohorts of US women. *Proceedings of the National Academy of Sciences*, *114*(37), 9966–9971. https://doi.org/10.1073/pnas.1705899114

Callahan, B. J., Wong, J., Heiner, C., Oh, S., Theriot, C. M., Gulati, A. S., McGill, S. K., & Dougherty, M. K. (2019). High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Research*, *47*(18), e103. https://doi.org/10.1093/nar/gkz569

Champely, S. (2020). pwr: Basic Functions for Power Analysis. R package version 1.3-0. https://CRAN.R-project.org/package=pwr

Čížková, D., Ďureje, Ľ., Piálek, J., & Kreisinger, J. (2021). Experimental validation of small mammal gut microbiota sampling from faeces and from the caecum after death. *Heredity*, *127*(2), 141–150. https://doi.org/10.1038/s41437-021-00445-6

Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., & Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*, *6*(1), 226. https://doi.org/10.1186/s40168-018-0605-2

de Boer, R., Peters, R., Gierveld, S., Schuurman, T., Kooistra-Smid, M., & Savelkoul, P. (2010). Improved detection of microbial DNA after bead-beating before DNA isolation. *Journal of Microbiological Methods*, *80*(2), 209–211. https://doi.org/10.1016/j.mimet.2009.11.009

Deiner, K., Walser, J.-C., Mächler, E., & Altermatt, F. (2015). Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation*, *183*, 53–63. https://doi.org/10.1016/j.biocon.2014.11.018

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, *14*(6), 927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x

Djurhuus, A., Port, J., Closek, C. J., Yamahara, K. M., Romero-Maraccini, O., Walz, K. R., Goldsmith, D. B., Michisaki, R., Breitbart, M., Boehm, A. B., & Chavez, F. P. (2017). Evaluation of filtration and DNA extraction methods for environmental DNA biodiversity assessments across multiple trophic levels. *Frontiers in Marine Science*, *4*, 314. https://doi.org/10.3389/fmars.2017.00314

Dopheide, A., Xie, D., Buckley, T. R., Drummond, A. J., & Newcomb, R. D. (2019). Impacts of DNA extraction and PCR on DNA metabarcoding estimates of soil biodiversity. *Methods in Ecology and Evolution*, *10*(1), 120–133. https://doi.org/10.1111/2041-210X.13086

Escudié, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K., Maman, S., Hernandez-Raquet, G., Combes, S., & Pascal, G. (2018). FROGS: Find, rapidly, OTUs with galaxy solution. *Bioinformatics*, *34*(8), 1287–1294. https://doi.org/10.1093/bioinformatics/btx791

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). SAGE.

Frago, E., Zytynska, S. E., & Fatouros, N. E. (2020). Microbial symbionts of herbivorous species across the insect tree. In *Advances in insect physiology* (Vol. *58*, pp. 111–159). Elsevier. https://doi.org/10.1016/bs.aiip.2020.04.002

Galan, M., Pons, J.-B., Tournayre, O., Pierre, É., Leuchtmann, M., Pontier, D., & Charbonnel, N. (2018). Metabarcoding for the parallel identification of several hundred predators and their prey: Application to bat species diet analysis. *Molecular Ecology Resources*, *18*(3), 474–489. https://doi.org/10.1111/1755-0998.12749

Galan, M., Razzauti, M., Bard, E., Bernard, M., Brouat, C., Charbonnel, N., Dehne-Garcia, A., Loiseau, A., Tatard, C., Tamisier, L., Vayssier-Taussat, M., Vignes, H., & Cosson, J.-F. (2016). 16S rRNA amplicon sequencing for epidemiological surveys of bacteria in wildlife. *MSystems*, *1*(4), e00032-16. https://doi.org/10.1128/mSystems.00032-16

Giangacomo, C., Mohseni, M., Kovar, L., & Wallace, J. G. (2021). Comparing DNA extraction and 16S rRNA gene amplification methods for plant-associated bacterial communities. *Phytobiomes Journal*, *5*(2), 190–201. https://doi.org/10.1094/PBIOMES-07-20-0055-R

Greathouse, K. L., Sinha, R., & Vogtmann, E. (2019). DNA extraction for human microbiome studies: The issue of standardization. *Genome Biology*, *20*(1), 212. https://doi.org/10.1186/s13059-019-1843-8

Groussin, M., Mazel, F., Sanders, J. G., Smillie, C. S., Lavergne, S., Thuiller, W., & Alm, E. J. (2017). Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nature Communications*, *8*(1), 14319. https://doi.org/10.1038/ncomms14319

Gryp, T., Glorieux, G., Joossens, M., & Vaneechoutte, M. (2020). Comparison of five assays for DNA extraction from bacterial cells in human faecal samples. *Journal of Applied Microbiology*, *129*(2), 378–388. https://doi.org/10.1111/jam.14608

Guiver, E., Galan, M., Lippens, C., Bellenger, J., Faivre, B., & Sorci, G. (2022). Increasing helminth infection burden depauperates the diversity of the gut microbiota and alters its composition in mice. *Current Research in Parasitology & Vector-Borne Diseases*, *2*, 100082. https://doi.org/10.1016/j.crpvbd.2022.100082

Hart, M. L., Meyer, A., Johnson, P. J., & Ericsson, A. C. (2015). Comparative evaluation of DNA extraction methods from feces of multiple host species for downstream next-generation sequencing. *PLoS One*, *10*(11), e0143334. https://doi.org/10.1371/journal.pone.0143334

Javal, M., Terblanche, J. S., Benoit, L., Conlong, D. E., Lloyd, J. R., Smit, C., & Chapuis, M.-P. (2022). Does host plant drive variation in microbial gut communities in a recently shifted Pest? *Microbial Ecology*. https://doi.org/10.1007/s00248-022-02100-x

Karstens, L., Asquith, M., Davin, S., Fair, D., Gregory, W. T., Wolfe, A. J., Braun, J., & McWeeney, S. (2019). Controlling for contaminants in low-biomass 16S rRNA gene sequencing experiments. *MSystems*, *4*(4), e00290–e00219. https://doi.org/10.1128/mSystems.00290-19

Kennedy, N. A., Walker, A. W., Berry, S. H., Duncan, S. H., Farquarson, F. M., Louis, P., Thomson, J. M., UK IBD Genetics Consortium, Satsangi, J., Flint, H. J., Parkhill, J., Lees, C. W., & Hold, G. L. (2014). The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. *PLoS One*, *9*(2), e88982. https://doi.org/10.1371/journal.pone.0088982

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, *40*(1), e3. https://doi.org/10.1093/nar/gkr771

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & 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. *Applied and Environmental Microbiology*, *79*(17), 5112–5120. https://doi.org/10.1128/AEM.01043-13

Knudsen, B. E., Bergmark, L., Munk, P., Lukjancenko, O., Priemé, A., Aarestrup, F. M., & Pamp, S. J. (2016). Impact of sample type and DNA isolation procedure on genomic inference of microbiome composition. *MSystems*, *1*(5), e00095-16. https://doi.org/10.1128/mSystems.00095-16

Lenth, R. (2023). emmeans: Estimated Marginal Means, aka Least-Squares Means. https://CRAN.R-project.org/package=emmeans

Lindsay, E. C., Metcalfe, N. B., & Llewellyn, M. S. (2020). The potential role of the gut microbiota in shaping host energetics and metabolic rate. *Journal of Animal Ecology*, *89*(11), 2415–2426. https://doi.org/10.1111/1365-2656.13327

Littleford-Colquhoun, B. L., Weyrich, L. S., Hohwieler, K., Cristescu, R., & Frère, C. H. (2022). How microbiomes can help inform conservation: Landscape characterisation of gut microbiota helps shed light on additional population structure in a specialist folivore. *Animal Microbiome*, *4*(1), 12. https://doi.org/10.1186/s42523-021-00122-3

Liu, M., Xue, Y., & Yang, J. (2019). Rare plankton subcommunities are far more affected by DNA extraction kits than abundant plankton. *Frontiers in Microbiology*, *10*, 454. https://doi.org/10.3389/fmicb.2019.00454

Magoc, T., & Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*(21), 2957–2963. https://doi.org/10.1093/bioinformatics/btr507

Mahé, F., Rognes, T., Quince, C., de Vargas, C., & Dunthorn, M. (2014). Swarm: Robust and fast clustering method for amplicon-based studies. *PeerJ*, *2*, e593. https://doi.org/10.7717/peerj.593

Mahé, F., Rognes, T., Quince, C., de Vargas, C., & Dunthorn, M. (2015). Swarm v2: Highly-scalable and high-resolution amplicon clustering. *PeerJ*, *3*, e1420. https://doi.org/10.7717/peerj.1420

Marotz, C., Amir, A., Humphrey, G., Gaffney, J., Gogul, G., & Knight, R. (2018). DNA extraction for streamlined metagenomics of diverse environmental samples. *BioTechniques*, *62*(6), 290–293. https://doi.org/10.2144/000114559

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

Mattei, V., Murugesan, S., Al Hashmi, M., Mathew, R., James, N., Singh, P., Kumar, M., Lakshmanan, A. P., Terranegra, A., Al Khodor, S., & Tomei, S. (2019). Evaluation of methods for the extraction of microbial DNA from vaginal swabs used for microbiome studies. *Frontiers in Cellular and Infection Microbiology*, *9*, 197. https://doi.org/10.3389/fcimb.2019.00197

McMurdie, P. J., & Holmes, S. (2014). Waste not, want not: Why rarefying microbiome data is inadmissible. *PLoS Computational Biology*, *10*(4), e1003531. https://doi.org/10.1371/journal.pcbi.1003531

Minich, J. J., Sanders, J. G., Amir, A., Humphrey, G., Gilbert, J., & Knight, R. (2019). Quantifying and understanding well-to-well contamination in microbiome research [Preprint]. *Genomics*. https://doi.org/10.1101/577718

Pollock, J., Glendinning, L., Wisedchanwet, T., & Watson, M. (2018). The madness of microbiome: Attempting to find consensus "best practice" for 16S microbiome studies. *Applied and Environmental Microbiology*, *84*(7), e02627–e02617. https://doi.org/10.1128/AEM.02627-17

R Development Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org

Robasky, K., Lewis, N. E., & Church, G. M. (2014). The role of replicates for error mitigation in next-generation sequencing. *Nature Reviews Genetics*, *15*(1), 56–62. https://doi.org/10.1038/nrg3655

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *4*, e2584. https://doi.org/10.7717/peerj.2584

Rombaut, A., Guilhot, R., Xuéreb, A., Benoit, L., Chapuis, M.-P., Gibert, P., & Fellous, S. (2017). Invasive Drosophila suzukii facilitates Drosophila melanogaster infestation and sour rot outbreaks in the vineyards. *Royal Society Open Science*, *4*(3), 170117.

Rynkiewicz, E. C., Hemmerich, C., Rusch, D. B., Fuqua, C., & Clay, K. (2015). Concordance of bacterial communities of two tick species and blood of their shared rodent host. *Molecular Ecology*, *24*(10), 2566–2579. https://doi.org/10.1111/mec.13187

Salonen, A., Nikkilä, J., Jalanka-Tuovinen, J., Immonen, O., Rajilić-Stojanović, M., Kekkonen, R. A., Palva, A., & de Vos, W. M. (2010). Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: Effective recovery of bacterial and archaeal DNA using mechanical cell lysis. *Journal of Microbiological Methods*, *81*(2), 127–134. https://doi.org/10.1016/j.mimet.2010.02.007

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., Parkhill, J., Loman, N. J., & Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, *12*(1), 87. https://doi.org/10.1186/s12915-014-0087-z

Schiebelhut, L. M., Abboud, S. S., Gómez Daglio, L. E., Swift, H. F., & Dawson, M. N. (2017). A comparison of DNA extraction methods for high-throughput DNA analyses. *Molecular Ecology Resources*, *17*(4), 721–729. https://doi.org/10.1111/1755-0998.12620

Shaffer, J. P., Carpenter, C. S., Martino, C., Salido, R. A., Minich, J. J., Bryant, M., Sanders, K., Schwartz, T., Humphrey, G., Swafford, A. D., & Knight, R. (2022). A comparison of six DNA extraction protocols for 16S, ITS and shotgun metagenomic sequencing of microbial communities. *BioTechniques*, *73*(1), 34–46. https://doi.org/10.2144/btn-2022-0032

Shaffer, J. P., Marotz, C., Belda-Ferre, P., Martino, C., Wandro, S., Estaki, M., Salido, R. A., Carpenter, C. S., Zaramela, L. S., Minich, J. J., Bryant, M., Sanders, K., Fraraccio, S., Ackermann, G., Humphrey, G., Swafford, A. D., Miller-Montgomery, S., & Knight, R. (2021). A comparison of DNA/RNA extraction protocols for high-throughput sequencing of microbial communities. *BioTechniques*, *70*(3), 149–159. https://doi.org/10.2144/btn-2020-0153

Shi, X. J., Liu, G., Zhang, M., Zhao, J., Li, H., Yang, Z., Bai, H., Liang, P., & Lu, Y. (2020). Membrane-sensitive bacterial DNA extractions and absolute quantitation of recovery efficiencies. *Science of the Total Environment*, *708*, 135125. https://doi.org/10.1016/j.scitotenv.2019.135125

Sinha, R., Abu-Ali, G., Vogtmann, E., Fodor, A. A., Ren, B., Amir, A., Schwager, E., Crabtree, J., Ma, S., Abnet, C. C., Knight, R., White, O., & Huttenhower, C. (2017). Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. *Nature Biotechnology*, *35*(11), 1077–1086. https://doi.org/10.1038/nbt.3981

Sohrabi, M., Nair, R. G., Samaranayake, L. P., Zhang, L., Zulfiker, A. H. M., Ahmetagic, A., Good, D., & Wei, M. Q. (2016). The yield and quality of cellular and bacterial DNA extracts from human oral rinse samples are variably affected by the cell lysis methodology. *Journal of Microbiological Methods*, *122*, 64–72. https://doi.org/10.1016/j.mimet.2016.01.013

Teng, F., Darveekaran Nair, S. S., Zhu, P., Li, S., Huang, S., Li, X., Xu, J., & Yang, F. (2018). Impact of DNA extraction method and targeted 16S-rRNA hypervariable region on oral microbiota profiling. *Scientific Reports*, *8*(1), 16321. https://doi.org/10.1038/s41598-018-34294-x

Tiago, M. T., de Naiara, M. B. R., de Shana, M. d. O. C., de Miliane, M. S. S., Everaldo, Z., & da Irene, S. C. (2015). Choice of DNA extraction protocols from gram negative and positive bacteria and directly from the soil. *African Journal of Microbiology Research*, *9*(12), 863–871. https://doi.org/10.5897/AJMR2014.7259

Venables, W. N., Ripley, B. D., & Venables, W. N. (2002). *Modern applied statistics with S* (4th ed.). Springer.

Vesty, A., Biswas, K., Taylor, M. W., Gear, K., & Douglas, R. G. (2017). Evaluating the impact of DNA extraction method on the representation of human Oral bacterial and fungal communities. *PLoS One*, *12*(1), e0169877. https://doi.org/10.1371/journal.pone.0169877

Wang, S., Zhu, Y., Yang, Y., Li, J., & Hoffmann, M. R. (2020). Electrochemical cell lysis of gram-positive and gram-negative bacteria: DNA extraction from environmental water samples. *Electrochimica Acta*, *338*, 135864. https://doi.org/10.1016/j.electacta.2020.135864

Weinroth, M. D., Belk, A. D., Dean, C., Noyes, N., Dittoe, D. K., Rothrock, M. J., Ricke, S. C., Myer, P. R., Henniger, M. T., Ramírez, G. A., Oakley, B. B., Summers, K. L., Miles, A. M., Ault-Seay, T. B., Yu, Z., Metcalf, J. L., & Wells, J. E. (2022). Considerations and best practices in

animal science 16S ribosomal RNA gene sequencing microbiome studies. *Journal of Animal Science*, *100*(2), skab346. https://doi.org/10.1093/jas/skab346

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. (2nd ed.) Springer International Publishing. Imprint: Springer. https://doi.org/10.1007/978-3-319-24277-4

Wu, G. D., Lewis, J. D., Hoffmann, C., Chen, Y.-Y., Knight, R., Bittinger, K., Hwang, J., Chen, J., Berkowsky, R., Nessel, L., Li, H., & Bushman, F. D. (2010). Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. *BMC Microbiology*, *10*(1), 206. https://doi.org/10.1186/1471-2180-10-206

Yuan, S., Cohen, D. B., Ravel, J., Abdo, Z., & Forney, L. J. (2012). Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One*, *7*(3), e33865. https://doi.org/10.1371/journal.pone.0033865

Zhang, F., Sun, X. X., Zhang, X. C., Zhang, S., Lu, J., Xia, Y. M., Huang, Y. H., & Wang, X. J. (2018). The interactions between gut microbiota and entomopathogenic fungi: A potential approach for biological control of *Blattella germanica* (L.): Biological control of *Blattella germanica*. *Pest Management Science*, *74*(2), 438–447. https://doi.org/10.1002/ps.4726

Zhou, X., Nanayakkara, S., Gao, J.-L., Nguyen, K.-A., & Adler, C. J. (2019). Storage media and not extraction method has the biggest impact on recovery of bacteria from the oral microbiome. *Scientific Reports*, *9*(1), 14968. https://doi.org/10.1038/s41598-019-51448-7

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Chapuis, M.-P., Benoit, L., & Galan, M. (2023). Evaluation of 96-well high-throughput DNA extraction methods for 16S rRNA gene metabarcoding. *Molecular Ecology Resources*, *00*, 1–17. https://doi.org/10.1111/1755-0998.13812