2018

Influence of Commercial DNA Extraction Kit Choice on Prokaryotic Community Metrics in Marine Sediment

Gustavo Ramirez
University of Rhode Island, garamirez@uri.edu

Dennis Graham
University of Rhode Island, dgraham@gso.uri.edu

See next page for additional authors

Authors

Gustavo Ramirez, Dennis Graham, and Steven D'Hondt

Influence of commercial DNA extraction kit choice on prokaryotic community metrics in marine sediment

Gustavo A. Ramírez [ID],* Dennis Graham, Steven D'Hondt
Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island

Abstract

Commercial DNA extraction kits are widely used for cultivation-free surveys of marine sediment. However, the consequences of popular extraction-kit choices for sequence-based biological inferences about marine sedimentary communities have not previously been exhaustively assessed. To address this issue, we extracted DNA from multiple paired subsamples of marine sediment using two popular commercial extraction kits (MO BIO Laboratories PowerSoil® DNA isolation kit and MP Biomedicals FastDNA™ Spin Kit for Soil). We report comparisons of (1) total DNA yield, (2) extract purity, (3) gene-targeted quantification, and (4) post-sequencing ecological inferences in near-seafloor (< 1 meter below seafloor [mbsf]) and subsurface (> 1 mbsf) marine sediment. In near-seafloor sediment, the MP Biomedicals FastDNA™ Spin Kit for Soil exhibits higher extraction yields, higher 16S rRNA gene loads, higher taxonomic diversity, and lower contaminant loads. In subseafloor sediment, both kits yield similar values for all of these parameters. The MO BIO Laboratories PowerSoil® DNA isolation kit generally co-extracts less protein with the DNA in both near-seafloor and subseafloor sediment. For samples from all depths, both kits exhibit similar depth-dependent community richness patterns, taxonomic composition, and ordination-based similarity trends. We conclude that, despite kit-specific differences in extract yields, purity and reagent contaminant loads, ecological inferences based on next-generation sequencing of DNA extracted using these popular commercial kits are robustly comparable, particularly for subseafloor sediment samples.

Microbes in marine sediment comprise a sizable fraction of Earth's biosphere (D'Hondt et al. 2004; Kallmeyer et al. 2012). Community-wide adaptions to starvation (Jørgensen and D'Hondt 2006), extreme energy limitation (Hoehler and Jørgensen 2013), and slow biomass turnover rates (Lomstein et al. 2012) make cultivation-based ecological interrogation of this habitat prohibitively difficult (Cragg et al. 1990; Parkes et al. 2014; Jørgensen and Marshall 2016). In contrast, the cost of DNA sequencing, an alternative to cultivation-based surveys, is at an all-time low (Muir et al. 2016) and fuels the "next-generation sequencing (NGS)" revolution (Park and Kim 2016). NGS has contributed significantly to understanding of subseafloor sedimentary life (Biddle et al.

2008, 2018; Jørgensen et al. 2012; Orsi et al. 2013, 2018; Spang et al. 2015; Starnawski et al. 2017; Karst et al. 2018) and is now a standard tool for ecological studies of this and other marine habitats (Orcutt et al. 2011).

The first step in NGS is nucleic-acid extraction from the sample matrix (Lombard et al. 2011). Given the remarkable complexity of marine sediment, a universally optimal DNA extraction method is unrealistic. Therefore, numerous protocols exist that aim to optimize specific aspects of the extraction process (Lipp et al. 2008; Kallmeyer and Smith 2009; Lloyd et al. 2010; Morono et al. 2013, 2014; Lever et al. 2015). Different methods of DNA extraction are known to differently affect DNA yield and quality (Mahmoudi et al. 2011; Knauth et al. 2013) and these consequences must be considered in cross-study comparisons (Felczykowska et al. 2015). Standardization of nucleic-acid extraction technique is a known challenge for surveying microbial biogeography across subseafloor habitats (Orcutt et al. 2011). Although formal efforts have been made to bring this issue to light (Orcutt et al. 2013), no consensus has been reached.

Commercial DNA-extraction kits provide standardized, organic-solvent-free alternatives to the laborious tasks of in-house reagent preparation and protocol optimization (Tan and
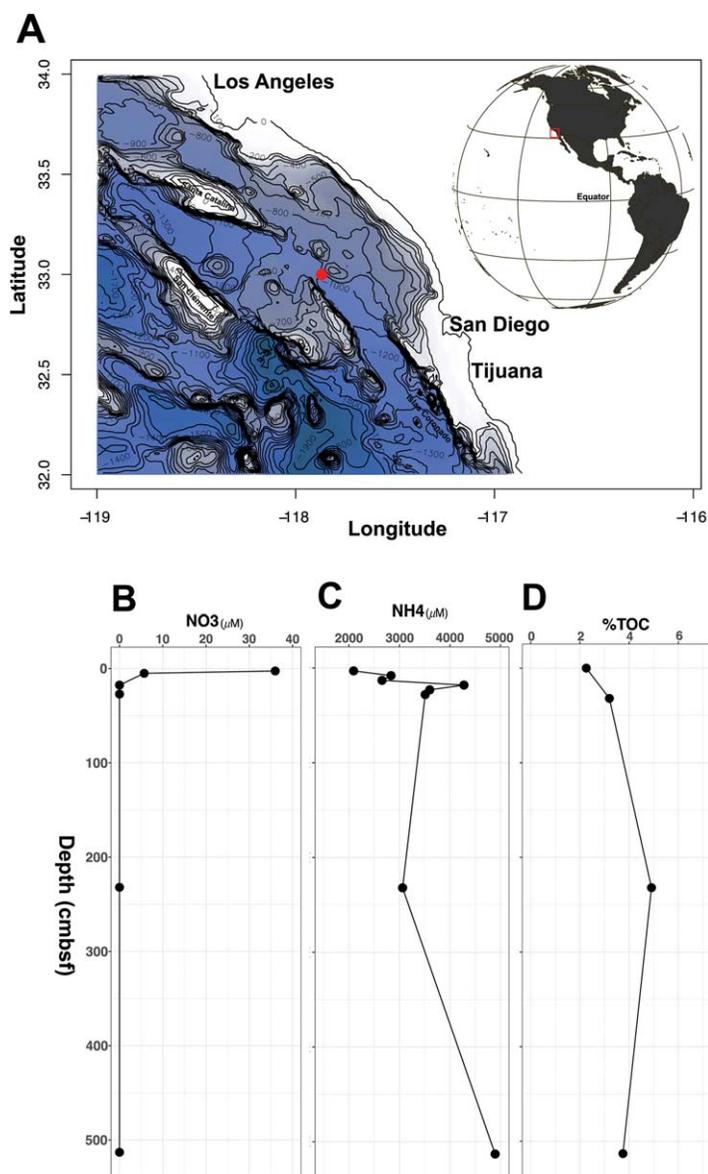

Fig. 1. (A) Bathymetric map of the Border Lands region, highlighting the California Channel Islands, with our coring site depicted as a red circle. Data for the map was compiled from the NOAA bathymetry library and visualized using the R-package *marmap*. (**B–D**) Depth variation of interstitial nitrate (**B**), ammonium(**C**), and percent (wt/wt) organic carbon at our coring site (**D**). Oxygen was below detection at all sampled sediment depths.

Yiap 2009). They are increasingly popular for studies of subseafloor habitats. Drawbacks of commercial kits include the unknown composition of proprietary reagents, protocol inflexibility to address sample-specific needs (Lever et al. 2015), and the pervasiveness of kit-specific contaminants (Salter et al. 2014). Surprisingly, given their popularity, a complete assessment of marine sediment extraction kit choice on NGS-based results, from physical characterization of extracts to postsequencing ecological inferences, is lacking. To begin to

address this issue, we extracted DNA from marine sediment using two commercial extraction kits (MO BIO Laboratories PowerSoil® DNA isolation kit and MP Biomedicals FastDNA^TM Spin Kit for Soil) that are widely used for studies of this habitat (Francis et al. 2005; Schippers et al. 2005; Edgcomb et al. 2011; Jørgensen et al. 2012; Spang et al. 2015; Walsh et al. 2016). To assess the consequences of kit choice, we compare (1) total DNA yields, (2) extract purities, (3) gene-targeted quantification, and (4) post-sequencing ecological inferences.

Materials and procedures

Sample collection

All samples are from Site SR1703-10-17PC/TC, cored by R/V *Sally Ride* expedition SR1703 to the USA-México Border Lands in the Eastern Pacific Ocean in February 2017 (mobilization: San Diego, Scripts Institute of Oceanography, demobilization: San Diego, Scripts Institute of Oceanography). Site SR1703-10-17PC/TC is located at latitude: 33°00.959′N, longitude: 117°52.114′W, in water depth of 915 m (Fig. 1A). The samples span the sediment depths from 0 cm to 513 cm below seafloor. The samples were taken immediately after core recovery, stored at −80°C, and transported to shore for DNA extraction, 16S rRNA gene quantification and sequencing. Subsamples of each sample were separately processed with the *Mo Bio PowerSoil® DNA Isolation Kit* and the *MP Biomedicals FastDNA^TM Spin Kit for Soil* as detailed below.

Total organic carbon and geochemical description

Briefly, we measured total organic carbon (TOC) in sediment and pore water by pyrolysis, following the procedures of Verardo et al. (1990). Ammonium concentrations were measured fluorometrically, following the procedures of Holmes et al. (1999). We measured oxygen and nitrate using optodes and ion chromatography with UV-absorbance detection, respectively, as detailed elsewhere (D'Hondt et al. 2015).

Sediment characterization

A full physical and geochemical description of the sediment from a nearby site, down to hundreds of meters below seafloor (mbsf), (Tanner Basin, ODP Site 1014) is found elsewhere (Lyle et al. 1997). Briefly, sediment from this area is siliciclastic clay containing calcareous nannofossil and foraminifers. At our coring site, oxygen was not detected, nitrate decreased to below detection within a few centimeter below seafloor (cmbsf), ammonium concentrations increased with sediment depth and TOC (%wt) remained within a 2–5% range (Fig. 1B–D), matching the observed range of TOC values at ODP Site 1014 at equivalent depths (4–5% for 30–490 cmbsf). Thus, we examined a high biomass anoxic habitat where organic-matter oxidation is likely coupled to sulfate reduction. A constant sedimentation rate of ~ 11.5 cm kyr^{-1} throughout the late Quaternary (Lyle et al. 1997) dates our deepest sample (513 cmbsf) to ~ 44.6 kyr.

DNA extraction: Mo Bio PowerSoil® DNA Isolation Kit

Chromosomal DNA extractions were performed in triplicate on all samples using the MO BIO PowerSoil® DNA Isolation Kit (MO BIO, Carlsbad, California, U.S.A.). Briefly, as recommended by the manufacturer, 0.25 g of sediment sample were vortexed, in a PowerBead Tube (2 mL volume), horizontally using a flat-bed vortex adapter pad (MO BIO Catalog# 1300-V1–24) at maximum speed for 10 min. In sum, all steps, involving proprietary reagents and spin filters, were performed as outlined in the manufacture's protocol. Final extracts were eluted in 100 μL of sterile PCR grade DNA-free water and stored at -80°C. Extraction blanks were extracted in parallel.

DNA extraction: MP Biomedicals FastDNA™ Spin Kit for Soil

Chromosomal DNA extractions were performed in triplicate on all samples using the MP Biomedicals FastDNA™ Spin Kit for Soil (MP Biomedicals, Santa Ana, California, U.S.A.). Briefly, as recommended for marine sediment by the manufacturer, 0.5 g of sediment sample was automatically homogenized by bead beating in Lysing Matrix E tubes (2 mL volume) using a FastPrep-96™ homogenizer at speed 5.5 twice for 40 s (with a 2 min rest on ice). In sum, all steps involving proprietary reagents and spin filters were performed as outlined in the manufacture's protocol. Final extracts were eluted in 100 μL of sterile PCR grade DNA-free water and stored at -80°C. Extraction blanks were extracted in parallel.

Extraction yield and purity

DNA extraction yields were calculated using a Qubit® 2.0 fluorometer with BR (broad range) dsDNA assay reagents (Invitrogen, Life Technologies). Both A_{260}/A_{280} and A_{260}/A_{230} absorbance ratios were measured using a Nanodrop 1000 spectrophotometer (Thermo Scientific).

Qunatitative PCR

Bacterial 16S rRNA gene quantification was performed with Bacteria-specific primers Bac341f (5′-CCT ACG GGW GGC WGC A-3′) and Uni518r (5′-ATT ACC GCG GCT GG-3′) using the following quantitative polymerase chain reaction (q-PCR) thermocycling program: 95°C for 15 min, ×40 (95°C for 15 s, 58°C for 30 s, 72°C for 30 s) with 1 : 10 diluted DNA extracts. Archaeal 16S rRNA gene quantification was performed with Uni515F (5′-CAG CMG CCG CGG TAA-3′) and Arch908r (5′-CCC GCC AAT TCC TTT AAG TT-3′) using the following thermocycling program: 95°C for 15 min, ×40 (95°C for 30 s, 58°C for 30 s, 72°C for 45 s) with 1 : 10 diluted DNA extracts. Standards for both primer sets consisted of dilutions of size-verified, gel-extracted, purified PCR amplicon (*Vibrio fischeri* for Bacteria and *Halobacterium* sp. NRC-1 for Archaea; Carolina Bio. Supp. catalogue #: 115722 and 154801, respectively). A range of 10^2–10^8 16S rRNA gene copy numbers per microliter for standards was used in each

run. The R^2 value of all standard curves was > 0.98 with amplification efficiencies of 110%. All samples, each representing experimental triplicates, were run in technical triplicates; thus $n = 9$ for each data point shown with bars representing standard error of the mean.

16S rRNA gene amplification and amplicon sequencing

The V4 hypervariable region of the 16S rRNA gene was targeted using the following universal prokaryotic primers: 518F (5′-GTG YCA GCM GCC GCG GTA A-3′) and 806R (5′-GGA CTA CNV GGG TWT CTA AT-3′), with partial Nextera adapters (Caporaso et al. 2012). Polymerase chain reactions (PCRs) were performed in triplicate for each extract (1 : 10 diluted) using the following thermocycling program: 94°C for 3 min; ×32 (94°C for 45 s, 50°C for 60 s, 72°C for 90 s); 72°C for 10 m and a 4°C hold. Last, pooled triplicate amplicons were cleaned with AMPure® XP beads (Agencourt Bioscience Corporation, Beckman Coulter). Sequencing was performed at the University of Rhode Island Next-Generation Sequencing (NGS) Facility using the Illumina MiSeq platform with V2 chemistry kit (2 × 250 bp, 500 cycles) reagents. All sequencing data has been submitted to the NCBI SRA repository under accession numbers: SRX3512769:SRX3512778, SRX3512787, and SRX3512788.

Sequence analyses

All sequences were processed with *mothur* v.1.34.4 (Schloss et al. 2009) following the *mothur* Illumina MiSeq Standard Operation Procedure (Kozich et al. 2013). Briefly, forward and reverse reads were merged into 1.7 million contigs. Only contigs meeting the following criteria were retained: maximum homopolymers of 6 bp, minimal length of 288 bp, maximum length of 294 bp and zero ambiguities. Only 1.03 million contigs ascribed to these criteria. All sequences from extraction blanks were removed from all other groups, thus eliminating blank extractions completely from further analyses. The remaining 688,196 sequences were aligned to the *mothur*-recreated Silva SEDD v119 database (Yarza et al. 2010), trimmed to the V4-hypervarieble alignment region and subsequently preclustered at 1% dissimilarity using the *pre.cluster* (diffs = 3, for ∼ 300 bp amplicons, as suggested in the *mothur* SOP) command. Spurious sequence generation was mitigated by abundance ranking sequences and merging with rare sequences if sequences differed by three base pairs, as outlined elsewhere (Kozich et al. 2013). Chimera screening and removal was performed by implementation of de novo mode of UCHIME (Edgar et al. 2011). Using the average neighbor method, a distance matrix was generated clustering 642,732 sequences into operational taxonomic units (OTUs) at 3% or higher dissimilarity cut off. Taxonomic classification of OTUs was done with *mothur* using the SILVA v119 database. A summary of group-specific sequence numbers following quality control commands is provided in the Supporting Information Table S1.
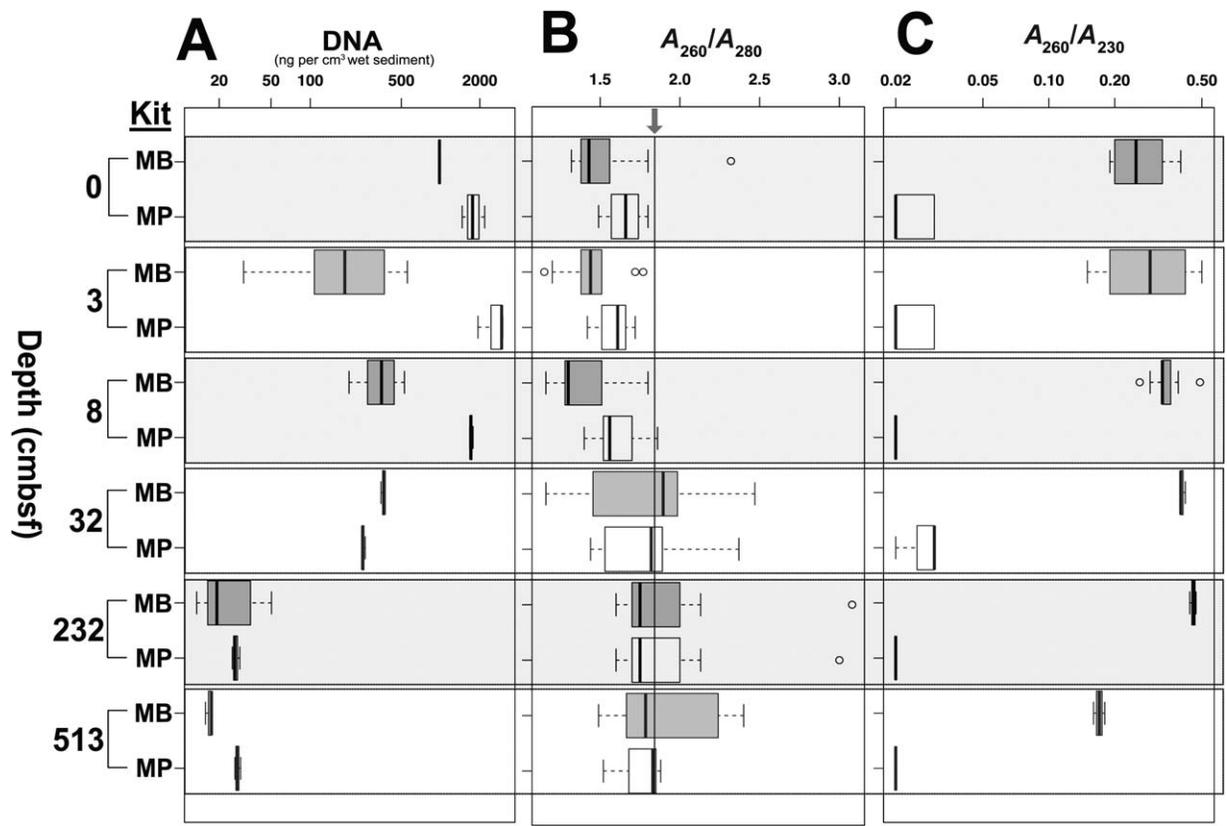
Fig. 2. (**A**) DNA extraction yields in ng of DNA per cm^3 of sediment for both extraction kits at all sampled sediment depths. (**B**) A_{260}/A_{280} ratios for both extraction kits at all sampled sediment depths. The arrow points to a vertical line that depicts the expected A_{260}/A_{280} ratio of 1.8 for pure extracts. (**C**) A_{260}/A_{230} ratios for both extraction kits at all sampled sediment depths. Note all observations fall below the expected A_{260}/A_{230} ratio (1.8 or higher).

Statistical analyses and visualizations

All community analyses were performed on a subsampled dataset ($n = 32,031$ sequences per group). Estimations of sample richness (Observed OTU counts, Chao1, and Shannon), principal coordinate analysis (PCoA), using Bray-Curtis distances, and OTU abundance heatmaps were generated in *RStudio* version 0.98.1091 (Racine 2012) using the packages *vegan* version 2.3-0 (Oksanen et al. 2015) and *phyloseq* (McMurdie and Holmes 2013).

Assessment

Pore-water geochemistry and TOC

The sediment was free of dissolved oxygen at all sampled depths. Nitrate was 36 μM at the sediment/water interface and was not detected at 17.5 cmbsf (Fig. 1B). Ammonium generally increased with sediment depth from $\sim$ 2 mM near the sediment/water interface to $\sim$ 4.9 mM at 513 cmbsf (Fig. 1C). It increases to 40 mM at 450 mbsf at the nearby Tanner Basin ODP drill site 1014 (Lyle et al. 1997). TOC generally increased with sediment depth from 2% at 0 cmbsf to $\sim$ 4–5% at 232 cmbsf and 523 cmbsf (Fig. 1D).

Chromosomal DNA yields and extract purity

Extracted DNA quantities per cm^3 of wet sediment decrease as a function of sediment depth (Fig. 2A). Significantly higher DNA quantities were extracted using the MP kit in shallow (0–8 cmbsf) horizons ($p < 0.05$). In deeper horizons (> 8 cmbsf), extraction yields are similar for both kits and drop rapidly in magnitude with increasing depth, as expected with lower biomass loads (Fig. 2A). For both kits, extract concentrations from horizons deeper than 32 cmbsf are nearly two orders of magnitude lower than concentrations recovered from the sediment-water interface (0 cmbsf). The MB kit yielded higher concentrations than the MP kit at 32 cmbsf and 513 cmbsf. The 260/280 absorbance ratios are lowest in depths < 10 cmbsf independently of extraction kit used (Fig. 2B). In deeper horizons (32–513 cmbsf), mean 260/280 absorbance ratios are $\sim$ 1.8 but exhibit greater variability than at shallower depths. The 260/230 absorbance ratios, a secondary DNA purity measurement, are significantly different for
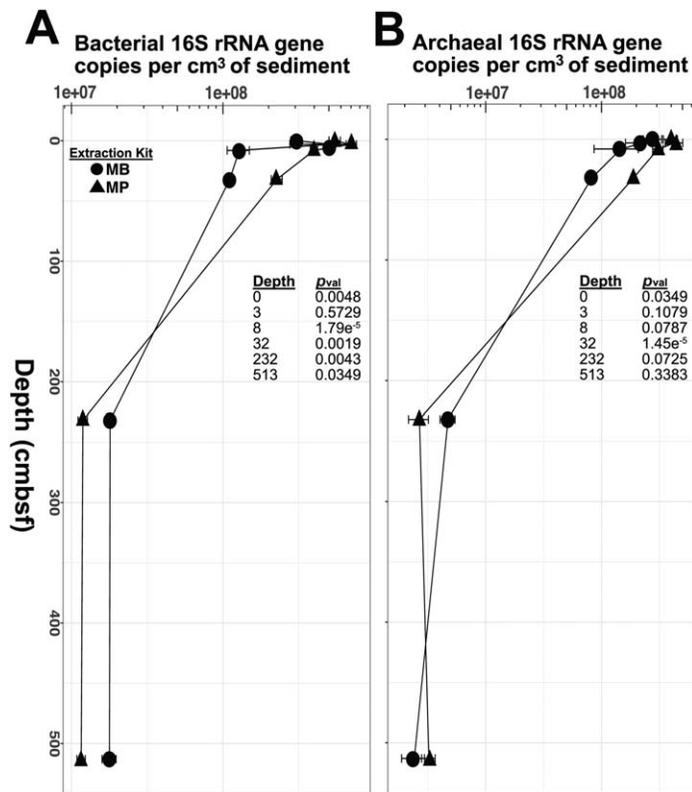
Fig. 3. (**A**) Bacteria-specific *q*-PCR gene loads per extraction kit at all sampled sediment depths. *p* values depicted in the figure show the results of student's *t*-test for all depth intervals. (**B**) Archaea-specific *q*-PCR gene loads per extraction kit at all sampled sediment depths. *p* values depicted in the figure show the results of student's *t*-test for all depth intervals.
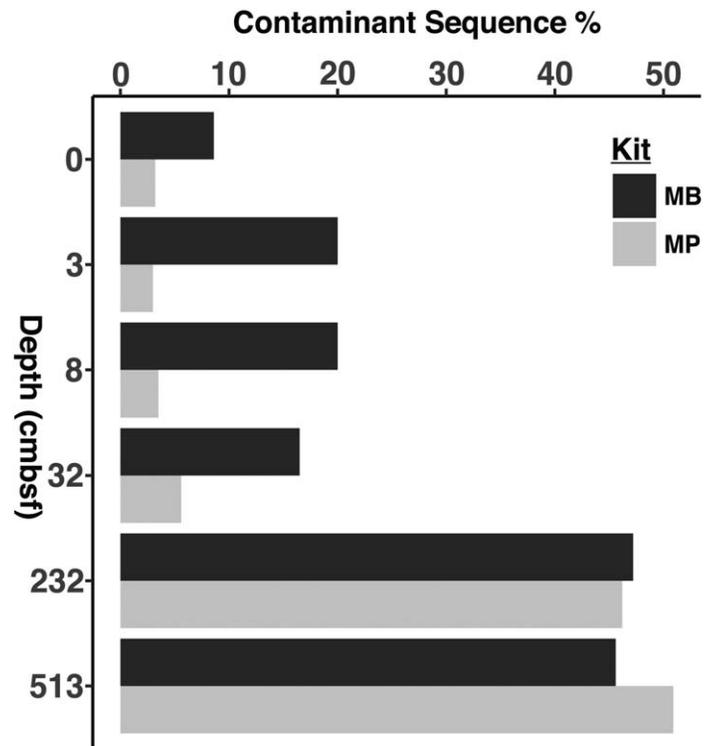

Fig. 4. Percent of sequences lost after subtraction of sequences from extraction blanks for each extraction kit across all sampled sediment depths.

MB and MP extracts ($p \leq 0.05$) at all sample depths, with MB 260/230 absorbance ratios consistently above 0.20 and MP ratios consistently below 0.05 (Fig. 2C).

Domain-specific 16S rRNA gene quantification

16S rRNA gene counts for Bacteria and Archaea are highest near the sediment/water interface and decrease as a function of sediment depth (Fig. 3). Count magnitudes for each domain are similar at the sediment-water interface. In deep samples, Archaea comprise an average of $\sim 18\%$ of the total prokaryotic 16S rRNA genes independently of extraction kit used and thus the Archaea to Bacteria ratio is lowest. Generally, significantly higher Bacterial and Archaeal gene counts were recovered with the MP extraction kit in the 0–32 cmbsf depth range ($p \leq 0.05$); however, at 3 cmbsf, domain-specific counts from both kits are statistically identical. In deeper horizons, significantly higher bacterial gene counts were recovered with the MB kit (Fig. 3A) but Archaeal gene counts were not consistently kit-dependent (Fig. 3B).

Contaminant sequences

Near-seafloor (0–32 cmbsf) samples processed with the MP kit contain a small fraction of sequences observed in MP

extraction blanks (3.0–5.6% of total sequences) (Fig. 4). A larger fraction of the sequences in near-seafloor samples processed with the MB kit are also present in the MB extraction blanks (8.6–30.0% of total sequences). For samples from greater depths (232 cmbsf and 513 cmbsf), sequence overlap between blanks and sediment samples is very high (> 45% of the sediment sequences) and does not depend on the kit (Fig. 4).

Prokaryotic community composition

Community composition from both MB and MP extracts is nearly identical for each sample (Fig. 5). Shallow samples (0–32 cmbsf) have sequences assigned to both Archaeal and Bacterial classes. Archaea are represented by unclassified Archaea, Thaumarchaeota, and Euryarchaeota. The Proteobacteria, Planctomycetes, Firmicutes, Chloroflexi, and candidate divisions OP8 (Aminicenantes) and JS1 (Atribacteria) represent the Bacteria. Deeper samples are exclusively dominated by Bacterial sequences; specifically, by Proteobacteria, Firmicutes, and Actinobacteria (> 2% relative abundance). Proteobacteria are abundant in all sampled depths. In shallow samples (0–32 cmbsf), Delta- and Gamma-proteobacteria are present (Fig. 5). At sediment depths greater than 32 cmbsf, the Gamma-proteobacteria constitute nearly 50% of the sequences and exclusively represent the Proteobacteria. In shallow samples, the Firmicute class Bacilli is only present
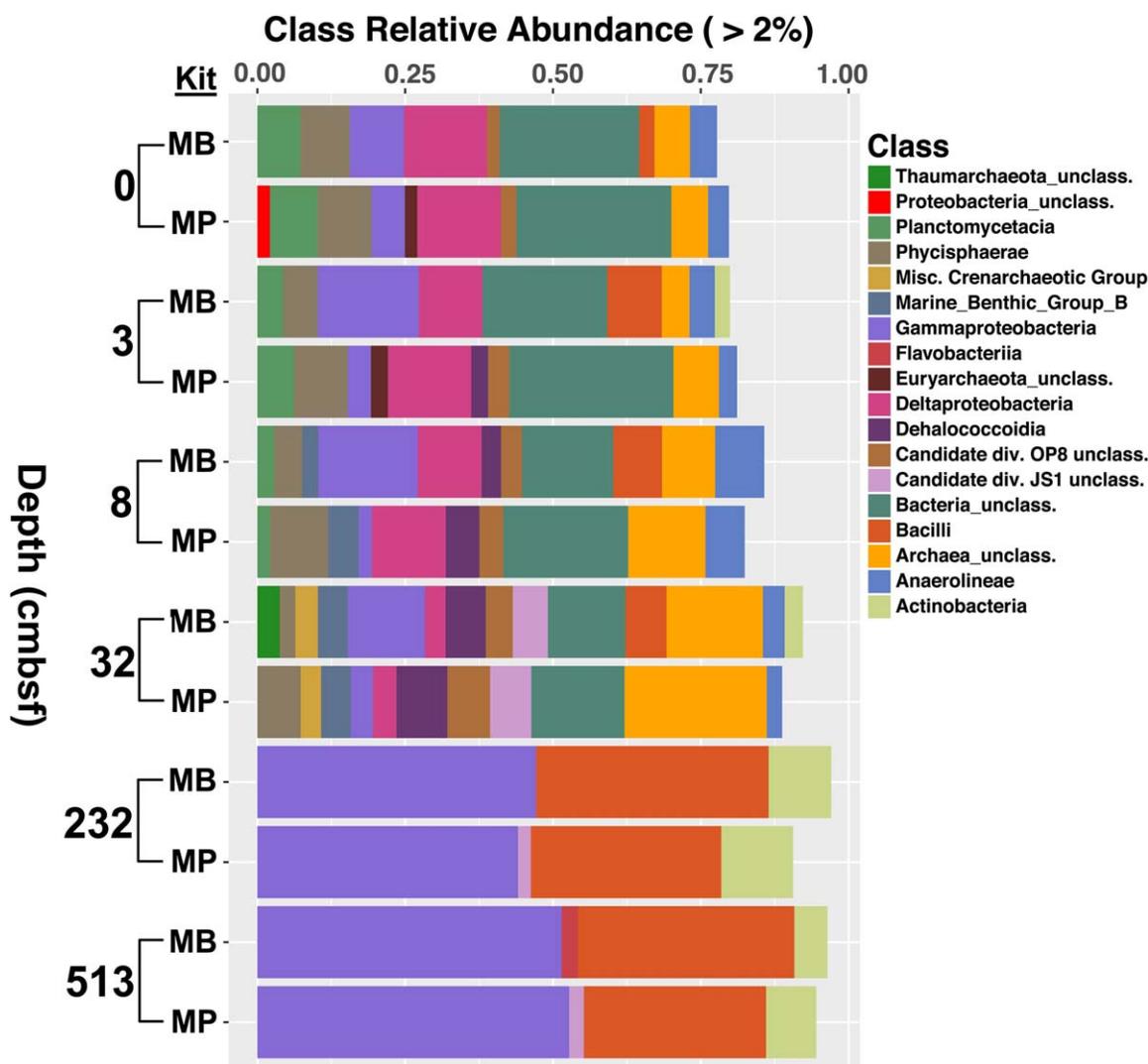
Fig. 5. Class-level community prokaryotic composition recovered from each extraction kit across all sampled sediment depths.

in the MB communities, despite being extracted with both kits from sample depths > 32 cmbsf. Last, less abundant taxonomic guilds (< 2% of community) and the relative fraction of unclassified prokaryotic sequences decrease as a function of sediment depth for both extraction kits.

OTU-specific prevalence

Both kits yield generally similar abundances for the most abundant 100 OTUs in the dataset (Fig. 6). Despite this trend, a few exceptions exist, defined by substantially different (one order of magnitude or more) kit-specific OTU abundances within the same depth horizon (Fig. 6, red boxes). For the 10 most abundant OTUs, these differences are most accentuated in the 3 cmbsf and 8 cmbsf horizons where the MP kit yields much lower normalized abundances for OTUs 1–5 and 8. For the next 90 most abundant OTUs (10–100), across all sampled depth horizons, there is no consistency

regarding abundance and extraction kit since, for a single sediment depth, different OTUs where large abundance disparities are present differ in the extraction kit that reports the high vs. low values (e.g., at 232 cmbsf; OTUs 12, 73, 82, 86, 89, 90, and 92).

Alpha- and Beta-diversity metrics

For near-seafloor sediment (0–32 cmbsf), OTU counts, Chao1, and Shannon diversity estimators are consistently higher for DNA extracted with the MP kit than for DNA extracted with the MB kit. For deeper horizons (232 cmbsf and 513 cmbsf), both extraction kits yield nearly identical diversity estimates (Fig. 7A). PCoA analysis resolves shallow (0–32 cmbsf) samples as a function of extraction kit along axis 1 (47% of variance). Axis 2 (18.9% of variance) separates each shallow group as a function of sediment depth (Fig. 7B). The two kits yield slightly different results (different
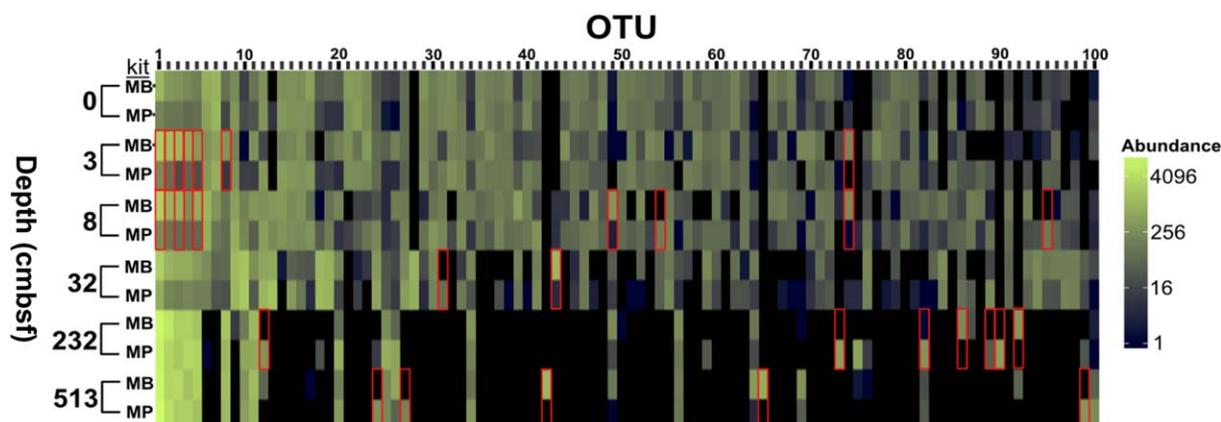
Fig. 6. Heat map depicting normalized abundances for the 100 most prevalent OTUs in our dataset as a function of extraction kit used and sediment depth interval. Red boxes highlight selected instances where, for specified OTUs, values reported by the two extraction kits differed by more than an order of magnitude.

Axis 1 loadings, but similar Axis 2 loadings) for near-seafloor samples (0–32 cmbsf). The results from the deeper horizons (232 cmbsf and 513 cmbsf) cluster tightly regardless of extraction kit used.

Discussion

MP extracts more DNA under near-seafloor conditions, but the same amount as MB under subseafloor conditions

The MP kit extracted more DNA from near-seafloor samples than the MB kit, specifically in the 0–8 cmbsf depth range (Fig. 2A). Similar observations have been made by others (Lever et al. 2015). Physical force alone may explain this observation, since the automated "bead beating" homogenization step of the MP protocol is more vigorous than the homologous horizontal vortexing step used in the MB protocol (see "Materials and procedures" section). Interestingly, for deeper horizons, extraction efficiency appears to be independent of homogenization vigor, since the MB kit reports higher yields at 32 cmbsf and 513 cmbsf. However, previous studies have shown that DNA extraction efficiency depends on both extraction protocol and taxonomic group (Morono et al. 2014). A detailed mechanism to explain our observations is unattainable because the reagent identities and chemical mechanisms of commercial extraction kits are proprietary; however, we speculate that (1) a decrease in biomass loads coupled with (2) a decrease in diversity affect (1) the amount of physical force needed for efficient community-wide lysis and (2) lower inter-taxonomic variation in membrane strength, respectively, eliminating the advantages observed for the MP protocol with near-seafloor sediment.

MP kit co-extracts more protein at all depths

In addition to DNA yield, extract purity is an important consideration for downstream applications (e.g., *q*-PCR, tag-sequencing, shot-gun metagenomics). The presence of co-extracted proteins and/or phenolic compounds (humic and fulvic acids) in near-seafloor (0–8 cmbsf) extracts from both kits is suggested by 260/280 nm absorbance ratios lower than 1.8 (Fig. 2B). This is a commonly reported issue for organic-rich marine sediment (Lloyd et al. 2010). For both kits, DNA extracts from samples of deep sediment (32–513 cmbsf) were notably cleaner (260/280 nm averages of $\sim$ 1.8) than those from near-seafloor samples. A secondary nucleic acid to protein ratio measure (260/230 nm), generally expected to be equal to or higher than 1.8 for "pure samples," shows that both kits co-extract protein but that the MP kit consistently co-extracts higher protein amounts relative to the MB kit across all sampled depths (Fig. 2C). Thus, despite the higher DNA yields obtained using the MP kit in shallow sediment, the vigorous homogenization in the MP protocol may also result in higher levels of inadvertent proteinaceous co-extraction.

Domain-specific gene quantification: Same depth-dependent trends with both kits

With both extraction kits, counts of Bacterial and Archaeal 16S rRNA genes decrease with increasing sediment depth (Fig. 3). Kit-specific differences are generally larger in magnitude for Bacteria, where kit choice significantly influences the gene quantities reported at all depths except at 3 cmbsf ($p < 0.05$) (Fig. 3A). In surface sediment, 16S gene count magnitudes for each domain are comparable ($\sim$ 1–5 $\times 10^8$ genes cm^{-3}); at depth, however, counts for Archaea (Fig. 3B) decrease relative to those of Bacteria ($\sim 10^6$ genes cm^{-3} vs. $\sim 10^7$ genes cm^{-3}, respectively). Previous work has shown that complex DNA-substrate interactions may favor the extraction of Bacterial over Archaeal DNA from marine basalt (Wang and Edwards 2009). Given that both extraction kits in this study yield gene counts for both Domains within the same range in shallow (0–32 cmbsf) samples, the dominance of Bacterial over Archaeal genes in deeper sediment (> 32 cmbsf) is likely not due to extraction
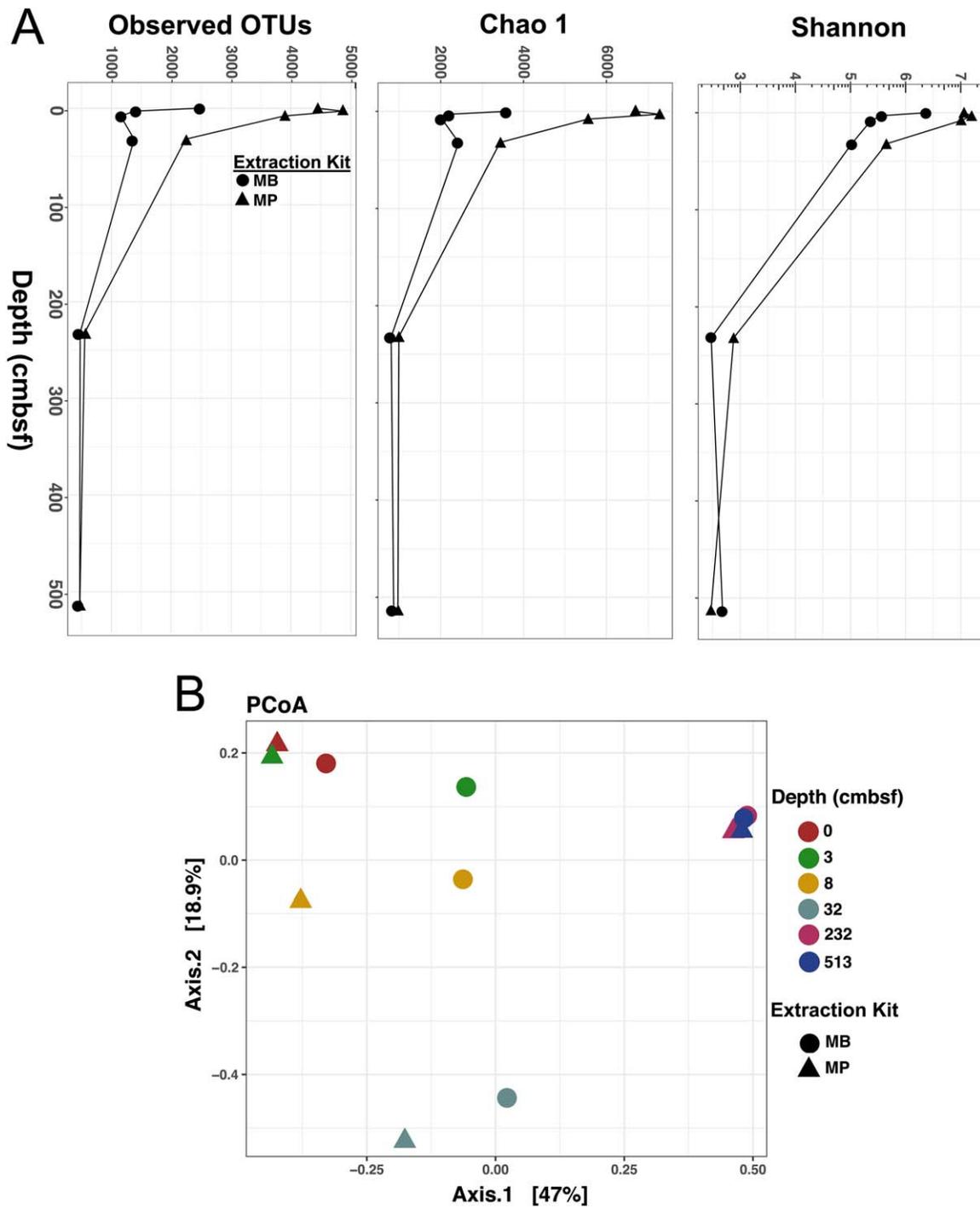
Fig. 7. (**A**) Alpha-diversity metrics (left: OTU counts normalized to 32,031 reads/sample, middle: Chao1, right: Shannon) depicted as a function of extraction kits at all sampled sediment depths. (**B**) Two-dimensional PCoA. Color depicts sediment depth interval and shape corresponds to extraction kit used (circles for MB and triangles for MP). Percentages in brackets represent the percent of total variance of the data set explain by each axis.

bias with either kit. It should be noted that lower Archaea counts at depth may still be a technical artifact because all available primers are inadequate at capturing the true breadth of diversity (e.g., Asgard Archaea) of this Domain (Karst et al. 2018). Overall, despite potential primer bias

and kit-specific significant differences in absolute quantities reported for Bacteria, both kits provide nearly identical narratives regarding the relative prevalence and abundance of Domain-specific 16S rRNA genes at all sampled sediment depths.

Contaminant sequence loads

Co-extraction and sequencing of kit-specific blanks allowed the in silico subtraction of contaminant sequences. Pre- and post-contaminant subtraction sequence loss shows that both kits yield extracts containing amplifiable kit-specific contaminant DNA (Fig. 4). The MB kit has higher contaminant-sequence percentages in surface sediment. Regardless of kit used, contaminant amplification increases substantially with sediment depth, to comprise nearly 50% of all sequences in the two deepest samples. Since (1) each extraction is assumed to contain the same initial amount of kit-contaminant DNA, (2) all sequencing occurred in a single sequencing run, and (3) the two subseafloor sediment horizons show the highest contaminant fractions, contaminant amplicon surges likely result from competitive PCR amplification (Salter et al. 2014). This is a longstanding problem for marine sediment that is exacerbated in low-biomass settings (Webster et al. 2003). We suggest addressing this issue in vitro, via larger extraction volumes to avoid template competition, and/or with arduous controls for efficient in silico subtraction.

Community composition is generally the same with both kits

Both extraction kits yield nearly identical community composition at the class levels (Fig. 5). However, from near-seafloor samples (0–32 cmbsf), Firmicutes (Bacilli class) are only represented in DNA extracted by the MB kit. Firmicutes are recovered by both kits in deeper samples. It is unknown if the Firmicute sequences extracted by the MB kit represent spores or vegetative cells. Given the following points: (1) Firmicutes dominate our two deeper samples (Fig. 5), (2) marine sediment may be supplied with 10^8 Firmicutes spores per m^2 per year (Hubert et al. 2009); we speculate that Firmicute spores are present in the near-seafloor sediment and more efficiently detected with the MB extraction protocol. Given that the MP kit more efficiently extracts bacterial spore DNA from soils compared to the MB kit (Dineen et al. 2010), an unsurprising result given the more vigorous homogenization step of the former, it is rather unexpected that the MB kit may be more efficient at this task in marine sediment.

Generally, extraction kits agree on OTU counts for discrete depths

Although normalized OTU counts are imperfect assessments of taxonomic distribution in nature (Weiss et al. 2017), comparing this metric for each sampled horizon as a function of extraction kit reveals insightful patterns. Kit-specific abundances for the top 100 OTUs clustered from the combined dataset show that relative abundances recovered with either kit generally agree to within one order of magnitude (Fig. 6). Exceptions to this trend, however, are observed (Fig. 6, red boxes). In near-seafloor samples (3 cmbsf and 8 cmbsf), six of the 10 most abundant OTUs are consistently under-recovered with the MP kit relative to the MB kit, despite both kits reporting similar values for this OTU subset at all other depths. Among the next most abundant 90 OTUs, across all sampled depths, kit-specific differences are sporadically seen; however, high vs. low values may be attributed to either extraction kit.

Near-seafloor diversity is better captured with MP, subseafloor diversity is similar for both MP and MB

The number of OTUs per 32,031 reads is higher for the MP kit extractions in near-seafloor (0–32 cmbsf) sediment (Fig. 7A). Despite higher DNA yields with the MB kit at 32 cmbsf (Fig. 7A), OTU counts are higher for the MP extracts at this depth (Fig. 7A). This observation suggests that in high-biomass horizons and independently of bulk DNA yield, the MP extraction protocol favors lysis of a more diverse cellular cohort, resulting in higher OTU counts and better capturing alpha diversity. However, in the subseafloor sediment (at 232 cmbsf and 513 cmbsf), where 16S gene counts are lowest, both extraction kits yield nearly identical OTU counts (Fig. 7A). This convergence at depth may be explained by (1) subseafloor microbial populations being comprised of taxa equally susceptible to cell membrane disruption by both extraction protocols, (2) longer entombment times in deeper sediment weakening a fraction of the community that during early burial was more resistant to lysis by the MB kit (513 cmbsf cells are ∼ 44.6 kyr old), and/or (3) a lower silica matrix saturation point for MB relative to MP, an effect that exaggerates richness differences in near-seafloor horizons, where biomass is highest, but is negligible in deeper sediment with fewer than ∼ 10^8 16S rRNA genes cm^{-3}.

Ordination patterns not affected by extraction kit choice

Beta-diversity patterns, as inferred from multivariate ordination, follow similar depth-dependent trends as OTU counts; near-seafloor (0–32 cmbsf) horizons show kit-specific differences while subseafloor (232 cmbsf and 513 cmbsf) horizons cluster tightly (Fig. 7B). The MP extracts more clearly resolve ordination patterns (the MP-derived data are spread over a larger area in two-dimensional space, relative to the MB-derived data). Overall, the major conclusion drawn from ordination patterns—that near-seafloor communities are more dissimilar relative to subseafloor communities—may be drawn from the DNA extracted with either kit.

Comments and recommendations

Despite differences in DNA yield, protein co-extraction loads and absolute 16S rRNA gene counts, two popular DNA extraction kits (MO BIO Laboratories PowerSoil® DNA isolation kit and MP Biomedicals FastDNA™ Spin Kit for Soil) generate similar depth-dependent prokaryotic community composition, OTU-specific abundances and diversity trends for subseafloor sediment. Higher average extraction yields

are observed in near-seafloor sediment for the MP Biomedicals FastDNA™ Spin Kit for Soil. The MO BIO Laboratories PowerSoil® DNA isolation kit better avoided inadvertent co-extraction of protein. Extraction efficiencies decrease and relative contaminant sequence loads increase with both kits in deeper, lower biomass, sediment. Ecological inferences, i.e., nonparametric diversity estimates and multivariate statistical measures of community similarity with depth, specifically, drawn post-NGS analysis, are similar for DNA extracted with either kit. Given these results, we recommend (1) the MO BIO Laboratories PowerSoil® DNA isolation kit for minimal near-seafloor and subseafloor sediment proteinaceous co-extraction and (2) particularly for near-seafloor sediment, the MP Biomedicals FastDNA™ Spin Kit for Soil for larger extraction yields and resolution of higher taxonomic diversity. Results from DNA extracted with either kit are nearly identical for the subseafloor samples.

Dedication

This manuscript is dedicated to Augustus Huitzilin Ramírez-Tatone; welcome to Earth.

References

Biddle, J. F., S. Fitz-Gibbon, S. C. Schuster, J. E. Brenchley, and C. H. House. 2008. Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. Proc. Natl. Acad. Sci. USA **105**: 10583–10588. doi:10.1073/pnas.0709942105

Biddle, J. F., J. R. White, A. P. Teske, and C. H. House. 2011. Metagenomics of the subsurface Brazos-Trinity Basin (IODP site 1320): Comparison with other sediment and pyrosequenced metagenomes. ISME J. **5**: 1038–1047. doi:10.1038/ismej.2010.199

Caporaso, J. G., and others. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. **6**: 1621–1624. doi:10.1038/ismej.2012.8

Cragg, B. A., R. J. Parkes, J. C. Fry, R. A. Herbert, J. W. T. Wimpenny, and J. M. Getliff. 1990. Bacterial biomass and activity profiles within deep sediment layers. Proc. Integr. Ocean Drill. Program **112**: 607–619. doi:10.2973/odp.proc.sr.112.161.1990

D'Hondt, S., and others. 2004. Distributions of microbial activities in deep subseafloor sediments. Science **306**: 2216–2221. doi:10.1126/science.1101155

D'Hondt, S., and others. 2015. Presence of oxygen and aerobic communities from sea floor to basement in deep-sea sediments. Nat. Geosci. **8**: 299–304. doi:10.1038/ngeo2387

Dineen, S. M., R. T. Aranda, D. L. Anders, and J. M. Robertson. 2010. An evaluation of commercial DNA extraction kits for the isolation of bacterial spore DNA from soil. J. Appl. Microbiol. **109**: 1886–1896. doi:10.1111/j.1365-2672.2010.04816.x

Edgar, R. C., B. J. Haas, J. C. Clemente, C. Quince, and R. Knight. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics **27**: 2194–2200. doi:10.1093/bioinformatics/btr381

Edgcomb, V. P., D. Beaudoin, R. Gast, J. F. Biddle, and A. Teske. 2011. Marine subsurface eukaryotes: The fungal majority. Environ. Microbiol. **13**: 172–183. doi:10.1111/j.1462-2920.2010.02318.x

Felczykowska, A., A. Krajewska, S. Zielinska, and J. M. Los. 2015. Sampling, metadata and DNA extraction- important steps in metagenomi studies. Acta Biochim. Pol. **62**: 151–160. doi:10.18_916388/abp.2014

Francis, C. A., K. Roberts, J. Beman, A. Santoro, and B. B. Oakley. 2005. Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. Proc. Natl. Acad. Sci. USA **102**: 14683–14688. doi:10.1073pnas.0506625102

Hoehler, T. M., and B. B. Jørgensen. 2013. Microbial life under extreme energy limitation. Nat. Rev. Microbiol. **11**: 83–94. doi:10.1038/nrmicro2939

Holmes, R., A. Minot, R. Kérouel, B. Hooker, and B. Peterson. 1999. A simple and precise method for measuring ammonium in marine and freshwater ecosystems. Can. J. Fish. Aquat. Sci. **56**: 1801–1808. doi:10.1139/f99-128

Hubert, C., and others. 2009. A constant flux of diverse thermophilic bacteria into the cold Arctic Seabed. Science **325**: 1541–1544. doi:10.1126/science.1174012

Jørgensen, B. B., and S. D'Hondt. 2006. A starving majority deep beneath the seafloor. Science **314**: 932–934. doi:10.1126/science.1133796

Jørgensen, B. B., and I. P. Marshall. 2016. Slow microbial life in the seabed. Ann. Rev. Mar. Sci. **8**: 311–332. doi:10.1146/annurev-marine-010814-015535

Jørgensen, S., and others. 2012. Correlating microbial community profiles with geochemical data in highly stratified sediments from the Arctic Mid-Ocean Ridge. Proc. Natl. Acad. Sci. USA **109**: E2846–E2855. doi:10.1073/pnas.1207574109

Kallmeyer, J., and D. C. Smith. 2009. An improved electroelution method for separation of DNA from humic substances in marine sediment DNA extracts. FEMS Microbiol. Ecol. **69**: 125–131. doi:10.1111/j.1574-6941.2009.00684.x

Kallmeyer, J., R. Pockalny, R. R. Adhikari, D. C. Smith, and S. D'Hondt. 2012. Global distribution of microbial abundance and biomass in subseafloor sediment. Proc. Natl. Acad. Sci. USA **109**: 16213–16216. doi:10.1073/pnas.1203849109

Karst, S. M., M. S. Dueholm, S. J. Mcilroy, R. H. Kirkegaard, P. H. Nielsen, and M. Albertsen. 2018. Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. Nat. Biotechnol. **36**: 190–195. doi:10.1038/nbt.4045

Knauth, S., H. Schmidt, and R. Tippkotter. 2013. Comparison of commercial kits for the extraction of DNA from paddy soils. Lett. Appl. Microbiol. **56**: 222–228. doi:10.1111/lam.12038

Kozich, J. J., S. L. Westcott, N. T. Baxter, S. K. Highlander, and P. D. Schloss. 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. doi:10.1128/AEM.01043-13

Lever, M. A., A. Torti, P. Eickenbusch, A. B. Michaud, T. Santl-Temkiv, and B. B. Jorgensen. 2015. A modular method for the extraction of DNA and RNA, and the separation of DNA pools from diverse environmental sample types. Front. Microbiol. **6**: 476. doi:10.3389/fmicb.2015.00476

Lipp, J. S., Y. Morono, F. Inagaki, and K. U. Hinrichs. 2008. Significant contribution of Archaea to extant biomass in marine subsurface sediments. Nature **454**: 991–994. doi:10.1038/nature07174

Lloyd, K. G., B. J. Macgregor, and A. Teske. 2010. Quantitative PCR methods for RNA and DNA in marine sediments: Maximizing yield while overcoming inhibition. FEMS Microbiol. Ecol. **72**: 143–151. doi:10.1111/j.1574-6941.2009.00827.x

Lombard, N., E. Prestat, J. D. van Elsas, and P. Simonet. 2011. Soil-specific limitations for access and analysis of soil microbial communities by metagenomics. FEMS Microbiol. Ecol. **78**: 31–49. doi:10.1111/j.1574-6941.2011.01140.x

Lomstein, B. A., A. T. Langerhuus, S. D'Hondt, B. B. Jorgensen, and A. J. Spivack. 2012. Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment. Nature **484**: 101–104. doi:10.1038/nature10905

Lyle, M., I. Koizumi, and C. Richter. 1997. ODP leg 167, site 104. Proc. ODP, Init. Repts. **167**: 175–221. doi:10.2973/odp.proc.ir.167.1997

Mahmoudi, N., G. F. Slater, and R. R. Fulthorpe. 2011. Comparison of commercial DNA extraction kits for isolation and purification of bacterial and eukaryotic DNA from PAH-contaminated soils. Can. J. Microbiol. **57**: 623–628. doi:10.1139/w11-049

Mcmurdie, P. J., and S. Holmes. 2013. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One **8**: e61217. doi:10.1371/journal.pone.0061217

Morono, Y., T. Terada, J. Kallmeyer, and F. Inagaki. 2013. An improved cell separation technique for marine subsurface sediments: Applications for high-throughput analysis using flow cytometry and cell sorting. Environ. Microbiol. **15**: 2841–2849. doi:10.1111/1462-2920.12153

Morono, Y., T. Terada, T. Hoshino, and F. Inagaki. 2014. Hot-alkaline DNA extraction method for deep-subseafloor archaeal communities. Appl. Environ. Microbiol. **80**: 1985–1994. doi:10.1128/AEM.04150-13

Muir, P., and others. 2016. The real cost of sequencing: Scaling computation to keep pace with data generation. Genome Biol. **17**: 53. doi:10.1186/s13059-016-0917-0

Oksanen, J., and others. 2015. vegan: Community ecology package. R package version 2.2-1; [accessed 2018 June 25]. Available from http://cran.r-project.org/package=vegan

Orcutt, B. N., J. B. Sylvan, N. J. Knab, and K. J. Edwards. 2011. Microbial ecology of the dark ocean above, at, and below the seafloor. Microbiol. Mol. Biol. Rev. **75**: 361–422. doi:10.1128/MMBR.00039-10

Orcutt, B. N., and others. 2013. Microbial activity in the marine deep biosphere: Progress and prospects. Front. Microbiol. **4**: 189. doi:10.3389/fmicb.2013.00189

Orsi, W. D., V. P. Edgcomb, G. D. Christman, and J. F. Biddle. 2013. Gene expression in the deep biosphere. Nature **499**: 205–208. doi:10.1038/nature12230

Orsi, W. D., T. A. Richards, and W. R. Francis. 2018. Predicted microbial secretomes and their target substrates in marine sediment. Nat. Microbiol. **3**: 32–37. doi:10.1038/s41564-017-0047-9

Park, S. T., and J. Kim. 2016. Trends in next-generation sequencing and a new era for whole genome sequencing. Int. Neurourol. J. **20**: S76–S83. doi:10.5213/inj.1632742.371

Parkes, R. J., B. Cragg, E. Roussel, G. Webster, A. Weightman, and H. Sass. 2014. A review of prokaryotic populations and processes in sub-seafloor sediments, including biosphere:geosphere interactions. Mar. Geol. **352**: 409–425. doi:10.1016/j.margeo.2014.02.009

Racine, J. S. 2012. RStudio: A platform-independent IDE for R and Sweave. J. Appl. Econ. **27**: 167–172. doi:10.1002/jae.1278

Salter, S. J., and others. 2014. Reagent and laboratory contaminantion can critically impact sequence-based microbiome analyses. BMC Biol. **12**: 87. doi:10.1186/s12915-014-0087-z

Schippers, A., L. N. Neretin, J. Kallmeyer, T. G. Ferdelman, B. A. Cragg, R. John Parkes, and B. B. Jørgensen. 2005. Prokaryotic cells of the deep sub-seafloor biosphere identified as living bacteria. Nature **433**: 861–864. doi:10.1038/nature03302

Spang, A., and others. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature **521**: 173–179. doi:10.1038/nature14447

Starnawski, P., and others. 2017. Microbial community assembly and evolution in subseafloor sediment. Proc. Natl. Acad. Sci. USA **114**: 2940–2945. doi:10.1073/pnas.1614190114

Tan, S. C., and B. C. Yiap. 2009. DNA, RNA, and protein extraction: The past and the present. J. Biomed. Biotechnol. **2009**: 1. doi:10.1155/2009/574398

Verardo, D., P. Froelich, and A. Mcintryre. 1990. Determinantion of organic carbon and nitrogen in marine sediments using the Carlo Erba N-1500 analyzer. Deep-Sea Res. Part A Oceanogr. Res. Pap. **37**: 157. doi:10.1016/0198-0149(90)90034-S

Walsh, E. A., J. B. Kirkpatrick, S. D. Rutherford, D. C. Smith, M. Sogin, and S. D'Hondt. 2016. Bacterial diversity and community composition from seasurface to subseafloor. ISME J. **10**: 979–989. doi:10.1038/ismej.2015.175

Wang, H., and K. J. Edwards. 2009. Bacterial and archaeal DNA extracted from inoculated experiments: Implication for the optimization of DNA extraction from deep-sea

basalts. Geomicrobiol. J. **26**: 463–469. doi:10.1080/01490450903060814

Webster, G., C. J. Newberry, J. C. Fry, and A. J. Weightman. 2003. Assessment of bacterial community structure in the deep sub-seafloor biosphere by 16S rDNA-based techniques: A cautionary tale. J. Microbiol. Methods **55**: 155–164. doi:10.1016/s0167-7012(03)00140-4

Weiss, S., and others. 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome **5**: 27. doi:10.1186/s40168-017-0237-y

Yarza, P., W. Ludwig, J. Euzéby, R. Amann, K.-H. Schleifer, F. O. Glöckner, and R. Rosselló-Móra. 2010. Update of the all-species living tree project based on 16S and 23S rRNA sequence analyses. Syst. Appl. Microbiol. **33**: 291–299. doi:10.1016/j.syapm.2010.08.001

Acknowledgments

Conflict of Interest

None declared.