Received: 16 March 2023 | Accepted: 19 October 2023

DOI: 10.1111/1462-2920.16530

RESEARCH ARTICLE

ENVIRONMENTAL MICROBIOLOGY Applied Microbiology International

Simultaneous analysis of seven 16S rRNA hypervariable gene regions increases efficiency in marine bacterial diversity detection

Kleopatra Leontidou[1] | **Ion L. Abad-Recio**[2] | **Verena Rubel**[1] | **Sabine Filker**[3] | **Martin Däumer**[4] | **Alexander Thielen**[4] | **Anders Lanzén**[2,5] | **Thorsten Stoeck**[1]

[1]Ecology Group, Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau, Kaiserslautern, Germany

[2]Marine Ecosystems Functioning, AZTI, Marine Research, Basque Research and Technology Alliance, Pasia, Gipuzkoa, Spain

[3]Molecular Ecology Group, Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau, Kaiserslautern, Germany

[4]SeqIT, Laboratory for Molecular Diagnostics and Services, Kaiserslautern, Germany

[5]IKERBASQUE, Basque Foundation for Science, Bilbao, Bizkaia, Spain

Correspondence
Thorsten Stoeck, Ecology Group, Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau, Kaiserslautern, Germany.
Email: stoeck@rhrk.uni-kl.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Numbers: STO 414/19-1, Fi 2089/3-1

Abstract

Environmental DNA sequencing is the gold standard to reveal microbial community structures. In most applications, a one-fragment PCR approach is applied to amplify a taxonomic marker gene, usually a hypervariable region of the 16S rRNA gene. We used a new reverse complement (RC)-PCR-based assay that amplifies seven out of the nine hypervariable regions of the 16S rRNA gene, to interrogate bacterial communities in sediment samples collected from different coastal marine sites with an impact gradient. In parallel, we employed a traditional one-fragment analysis of the hypervariable V3–V4 region to investigate whether the RC-PCR reveals more of the 'unseen' diversity obtained by the one-fragment approach. As a benchmark for the full deck of diversity, we subjected the samples to PCR-free metagenomic sequencing. None of the two PCR-based approaches recorded the full taxonomic repertoire obtained from the metagenomics datasets. However, the RC-PCR approach detected 2.8 times more bacterial genera compared to the near-saturation sequenced V3–V4 samples. RC-PCR is an ideal compromise between the standard one-fragment approach and metagenomics sequencing and may guide future environmental sequencing studies, in which bacterial diversity is a central subject.

INTRODUCTION

High-throughput sequencing (HTS) technologies and computational abilities to process and analyse massive sequence datasets have remarkably transformed our understanding of microbial diversity. As a result of these technological achievements, a unified scaling law of biodiversity predicts that Earth is home to upward of one trillion (10^{12}) microbial species (Locey & Lennon, 2016). A large fraction of this diversity resides in marine water and sediment (Walsh et al., 2016). As part of complex ecological networks, these bacteria are essential elements in the oceanic carbon pump (Sogin et al., 2006) and play pivotal roles in primary production and nutrient recycling (Falkowski et al., 2008). Changes in the composition of bacterial communities may therefore have profound effects on the functioning of marine ecosystems, and, therefore, monitoring changes in the ocean microbiome on local, regional, and global scales is vital to underpin strategies that address environmental and societal challenges of the 21st century (Tara Ocean Foundation et al., 2022). Furthermore, due to their sensitive and fast reaction to environmental changes, bacteria have recently attracted attention as

Lanzén Anders and Stoeck Thorsten shared senior authorship.

bioindicators in environmental biomonitoring (Cordier et al., 2021; Frühe et al., 2021; Pawlowski et al., 2018). This include, for example, the effects of industrial activities such as oil extraction (Coelho et al., 2016; Cordier et al., 2018; Laroche et al., 2018; McFarlin et al., 2017; Nguyen et al., 2018) and marine finfish aquaculture (Aylagas et al., 2021; Frühe et al., 2021; Keeley et al., 2018), as well as urban and industrial discharge (e.g., Lanzén et al., 2021; Obi et al., 2016; Clark et al., 2020; Moreira et al., 2022) on marine ecosystems.

The most common approach applied by far to uncover microbial biodiversity in natural ecosystems is metabarcoding (or amplicon sequencing). This employs a targeted amplification of a taxonomic marker gene via polymerase chain reaction, the products of which are then sequenced, usually using an Illumina chemistry that produces millions of sequence reads (Mardis, 2013). Since its characterisation in the 1970s (Woese & Fox, 1975), the gold-standard taxonomic marker gene used for the classification of most bacteria is the 16S ribosomal RNA (rRNA) gene. In 1985, this gene was used for the first time to identify and classify uncultured bacteria in an environmental sample (Lane et al., 1985). Alternations of highly conserved and hypervariable regions are structuring this gene with an estimated substitution rate of the latter being ca. 7000 times higher than of the former (Van de Peer, 1996). While the highly conserved regions reflect the effects determined by function-related constraints that are conserved across higher-level taxonomic hierarchies, divergence in different 16S rRNA gene sequences is concentrated primarily in the hypervariable regions (Stackebrandt & Goebel, 1994). Therefore, these hypervariable regions are considered as ideal proxies to achieve a trustworthy level of taxonomic resolution on lower taxonomic levels (Kim et al., 2011).

Several hypervariable regions of the 16S rRNA gene have been used in environmental sequencing studies targeting bacterial communities, including the V1–V3 region (Harrison et al., 2018), the V4–V5 region (Milke et al., 2022), V4–V6 region (Walsh et al., 2016), the V5–V6 region (Jones et al., 2016), the V6–V8 region (Verhoeven et al., 2018) and the V6–V9 region (Brandt & House, 2016) or a single 16S hypervariable region, for example, V4 and V6 (Kerrigan et al., 2019). Among them, the presumably most popular fragment for bacterial diversity analysis of marine sediment is the V3–V4 region (Aylagas et al., 2021; Dully et al., 2021; Frühe et al., 2021), while the V4 region (primers 515F-806R) alone is also commonly used (Caporaso et al., 2011). The commonality of the above mentioned (and also other) studies is that they usually analyse only one or two full hypervariable regions to infer bacterial diversity because mainstream sequencing platforms (Illumina and previously Pyrosequencing and IonTorrent platforms) are limited by their capacity to sequence relatively short regions of DNA (Kerrigan et al., 2019). Although this has changed with the emergence of long-read sequencing technologies commercialized by Oxford Nanopore Technologies and Pacific Biosciences, short-read platforms still offer better cost-efficiency and do not require specific library preparation techniques such as amplicon circularisation in order to achieve relatively low error rates (Santos et al., 2020). Even though efforts were made to design universal PCR primers for the bacterial 16S rRNA gene that cover an as wide as possible range of bacteria (Klindworth et al., 2013), the 'universality' of each fragment-specific PCR primer set is not absolute (Kerrigan et al., 2019; Kim et al., 2008; Milke et al., 2022; Rosselli et al., 2016; Silverman et al., 2021; Stoeck et al., 2006). Furthermore, the design of PCR primers is based on a priori knowledge (i.e., available sequences in a 16S rRNA gene reference database). Sequences not available in a database may be characterised by a different primary structure at the PCR primer annealing sites, and thus, will not be amplified in the PCR reaction. Considering that from the one trillion estimated bacterial species (Locey & Lennon, 2016), we currently have reference sequences for approximately 9×10^6 species (included in the latest release of SILVA rRNA sequence database, https://www.arb-silva.de/documentation/release-1381/), it is reasonable to assume that the vast majority of species in a complex natural bacterial community remains undiscovered. Indeed, a modelling approach suggested that even with unlimited sampling and sequencing effort, the totality of microbial diversity in a marine tidal flat would be inaccessible with any single, and probably even all, PCR primer sets available (Hong et al., 2009).

A solution to this challenge to the discovery of microbial diversity may be PCR-free technologies such as metagenomics (Liu, 2011; Logares et al., 2014; Pinto & Raskin, 2012; Thomas et al., 2012) or direct 16S rRNA-seq (Rosselli et al., 2016; Urich et al., 2008) approaches. However, because of the higher costs of these approaches coupled with the massive data that need to be analysed, they are not ideal for routine analyses such as biomonitoring of bacterial diversity, which may require the processing of hundreds of samples at a time (Cordier et al., 2021). Another alternative, which may at least reduce PCR-primer bias is a multiple primer approach, in which the same DNA template is amplified in parallel PCR reactions by several different primer-pairs, which are then all sequenced separately as individual sequencing samples (Pinto & Raskin, 2012). This approach can easily be up-scaled and automated but would increase the costs per sample with a factor equal to the number of PCR-primer pairs used.

The EasySeq™ 16S rRNA bacterial ID kit from NimaGen provides a cost-efficient solution for a

multiple primer approach, by amplifying seven of the nine hypervariable regions of the 16S rRNA gene (V1–V6 and V9) in only two multiplex PCR reactions. The kit employs the reverse complement (RC)-PCR technology, a one-step PCR assay that allows amplification, indexing, and tailing of target DNA regions in a single PCR tube. In contrast to conventional PCR, RC-PCR reactions do not contain target-specific primers for PCR amplification in the starting reaction mixture. Instead, target-specific primers are synthesized in the very first cycle of RC-PCR using index primers and RC-PCR probes as templates. The RC-PCR probes consist of (from 5′ to 3′) (i) the RC of the target-specific primer sequence, (ii) a universal tail region, and (iii) an extension blocker. The index primers consist of (from 5′ to 3′) (i) Illumina sequence adapter, (ii) index sequence, and (iii) the RC of the universal tail region. In the first reaction cycle of RC-PCR, the RC-PCR probes hybridize to the index primers via their universal tail regions. The DNA polymerase then extends the index primer by the target-specific primer sequence. From the second RC-PCR cycle onwards, these newly formed target specific index primers will amplify the target region, yielding amplicons fully decorated with universal tails, indexes, and Illumina sequence adapters. Notably, due to the 3′ extension blocker, the RC-PCR probe itself will not be extended, which ensures a constant concentration of the RC-PCR probes in the reaction for iterative synthesis of target-specific index primers. Since the concentrations of the target-specific primers are more aligned with the concentration of the target molecule, off-target priming and primer dimerization are highly reduced, a significant advantage over conventional PCR, which starts with a high excess of target-specific primers. RC-PCR reactions from one multiplex can be pooled and cleaned up together using magnetic beads. In the EasySeq™ 16S rRNA bacterial ID kit, two multiplex pools are generated (pool A: V1–V2, V4, V6; pool B: V3, V5, V9), which after separate clean-ups are combined into a single sample, which is then ready for sequencing. The entire RC-PCR process significantly decreases the hands-on time for library preparation with a minimal risk for sample swapping and sample cross-contamination (Kieser et al., 2020).

RC-PCR has proven its strong value, for example, in SARS-CoV-2 whole genome sequencing (Coolen et al., 2021; Schwarzer et al., 2021) and forensics (Bus et al., 2021). Most recently, the EasySeq™ 16S rRNA bacterial ID kit was applied in a clinical setting for the detection of bacterial pathogens in patient samples. The study revealed that the kit increased species discrimination and the sensitivity of clinically relevant bacteria detection compared to traditional 16S rRNA sequencing (Moorlag et al., 2023).

This fully automatable process is nearly as cost-efficient as the amplification and sequencing of a single 16S rRNA gene fragment. Our expectation is that this approach will also unfold its power in microbial ecology studies by uncovering a notably larger fraction of the bacterial diversity in a variety of different coastal marine sediments compared to the traditional one-fragment approach. In specific, in this study, we address the following questions:

(i) To what extent (if any) does the RC-PCR protocol increase the fraction of bacterial diversity compared to traditional one-fragment studies? (ii) Which additional taxonomic entities (if any) are detected with individual hypervariable regions compared to traditional one-fragment studies? (iii) How does the employment of this workflow affect ecological information (alpha- and beta-diversity patterns) compared to traditional one-fragment studies? (iv) What taxonomic repertoire obtained from HTS is common across the individual hypervariable regions? To address these questions, we have analysed the same 10 coastal sediment samples using first, this new approach, second a traditional one-fragment analysis protocol (the hypervariable V3–V4 region), and, third, PCR-free metagenomics data, from which we infer microbial community structures that we consider as the 'full deck of diversity'.

EXPERIMENTAL PROCEDURES

Study sites, sampling, and DNA extraction

In total, we investigated bacterial community structures at 10 different coastal marine sites. Six sediment samples were collected from two aquaculture sites (Atlantic salmon, *Salmo salar*) in Scotland. One of these sites (DUN) was located near Oban and the other (LIS) in Loch Linnhe. Four further sampling locations were on the Basque coast (Spain, Bay of Biscay). Both aquaculture installations were sampled during the mid-production (DUN) and peak-production (LIS) period, respectively, in December 2020. Sample collection was described in detail previously (Dully et al., 2021). In brief, sediment was collected along a gradient of organic enrichment resulting from deposits of feed and fish faeces on the seabed, including the outer cage edge (CE, latitude Y 56.271, longitude X −5.279 at DUN; 56.501, −5.500 at LIS), an intermediate impact zone (Allowable Zone of Effect, AZE, latitude Y 56.271, longitude X −5.279 at DUN; 56.501, −5.500 at LIS) and a reference site unimpacted by aquaculture activities (REF, latitude Y 56.273, longitude X −5.272 at DUN; 56.499, −5.502 at LIS). At each site, two biological replicates were taken with a van Veen grab (0.1 m^2 area, DUN; 0.045 m^2 area, LIS). From each replicate, we sampled approximately 10 g of surface sediment (upper few millimetres) into a sterile plastic tube, which was then frozen at −20°C within a few hours of collection until further processing.

The four samples from the Bay of Biscay were all from tidal flats of three different estuaries of the rivers Oka, Urola, and Bidasoa. The four sites were chosen for having contrasting environmental statuses to obtain a spectrum of different bacterial communities. Two sites were sampled directly downstream of wastewater treatment plants: EOK05 (latitude Y 43.329, longitude X −2.672) in Oka, just outside the town Gernika, and EU08 (latitude Y 43.292, longitude X −2.245) in Urola, outside Zumaia. Both sites are considered impacted (Lanzén et al., 2021). EOK05 differed from EU08 by its hypoxic conditions and higher loads of organic matter (Lanzén et al., 2021). The other site from Oka was located near the mouth of the river in Sukarrieta, from a sandy beach, with very good environmental quality status (EOK20, latitude Y 43.394, longitude X −2.694) (Lanzén et al., 2021). Finally, EBI20 (latitude Y 43.370, longitude X −1.790) was from a small sandy beach in the town Hondarribia near the mouth of Bidasoa with relatively little impact (good environmental quality status, Lanzén et al., 2021). The impact categories were determined and classified according to the AZTI Marine Biological Index (for details we refer to Lanzén et al., 2021 and Dully et al., 2021). Samples were collected manually and followed the exact same procedure as described above for the aquaculture sites (Herlemann et al., 2011).

DNA was extracted from all samples following a previously described protocol (Frühe et al., 2021), employing the DNeasy PowerSoil kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The DNA concentration of the samples was measured with Nanodrop 2000 spectrophotometer (Peqlab, Erlangen, Germany) and it was in the range of 20–100 ng/μL.

PCR amplification and Illumina sequencing of the hypervariable V3–V4 16S rRNA gene fragment

The hypervariable V3–V4 region was already obtained previously in a different context (Dully et al., 2021). The PCR protocol for the amplification of this ca. 450 base pair (bp) fragment used the Bakt_341F (CCTACG GGNGGCWGCAG) and Bakt_805R (GACTACHVG GGTATCTAATCC) primer pair and consisted of an initial activation step of NEB's Phusion High-Fidelity DNA polymerase (NEB, USA) at 98°C for 30 s, followed by 27 identical three-step cycles consisting of 98°C for 10 s, 62°C for 30 s, and 72°C for 30 s; then a final 5-min extension at 72°C (Herlemann et al., 2011). From the resulting products, sequencing libraries were constructed using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina. The quality of the libraries was assessed with an Agilent Bioanalyzer 2100 system. V3–V4 libraries were sequenced by SeqIT (Kaiserslautern, Germany) on an Illumina MiSeq platform, generating 2 × 250 bp paired-end reads. All sequences can be accessed from NCBI's SRA database under BioProject ID PRJNA768445.

Reverse complement PCR (RC-PCR) workflow and Illumina sequencing

Using the EasySeq™ RC-PCR 16S rRNA variable region sequencing kit (Nimagen, Netherlands), we amplified the following hypervariable regions in two RC-PCR reactions for each of the 10 sediment samples: V1–V2, V3, V4, V5, V6, and V9. Thereby, RC-PCR reaction 1 (panel A) included a specific RC-PCR-primer mix to amplify the fragments V1–V2, V4, and V6, while reaction 2 (panel B) included a primer mix specific for the hypervariable regions V3, V5, and V9. The primary structures of the individual specific primers (also called RC-PCR probes) in both reaction mixes as well as the lengths of the individual hypervariable regions are provided in Table 1. According to the manufacturer's instructions, the RC-PCR reactions mixtures included 0.2 μL of the RC-PCR probe pools (panel A mix for RC-PCR multiplex and panel B mix for RC-PCR multiplex), 1.8 μL of the dilution buffer, 10 μL of the RC-PCR 2× Hotstart HiFi master mix and 5 μL of template DNA. RC-PCR reactions were performed in index-tubes of the 96-well plates provided with the kit, which contain pre-spotted primers harbouring the Illumina adapter sequences, the universal sequencing tail as well as unique dual indexes to tag each individual sample for downstream bioinformatic assignment of the amplified hypervariable regions to a specific DNA-sample. The touch-down RC-PCR program for both RC-PCR multiplexes 1 and 2 consisted of one activation cycle for 2 min at 98°C, followed by one cycle of 98°C for 10 s, 80°C for 1 s, 58°C for 10 min, and 72°C for 1 min, 2 cycles of 98°C for 10 s, 80°C for 1 s, 58°C for 90 min, 72°C for 30 s, 34 identical four-step cycles consisting of 95°C for 10 s, 80°C for 1 s, 58°C for 2 min, 72°C for 30 s, as per Nimagen's kit manual. The RC-PCR multiplex reactions 1 of all samples were then pooled into one single sample (Pool A consisting of the amplified hypervariable regions V1–V2, V4, and V6 of all 10 sediment samples) and all RC-PCR multiplex reactions 2 into another single sample (Pool B consisting of the amplified hypervariable regions V3, V5, and V9 of all 10 sediment samples). To remove unbound primers, primer dimers, and salts, these two sample pools A and B were then subjected to a magnetic bead cleanup (Nimagen's AmpliClean) according to the instructions in the EasySeq™ RC-PCR kit manual prior to Illumina sequencing. After quality check and quantification of the two libraries (Pool A and Pool B) on a Fluostar Optima system, the two sample pools were combined and sequenced by SeqIT (Kaiserslautern,

TABLE 1 PCR-primers used in this study.

Gene region	Forward primer sequence (5′-3′)	Reverse primer sequence (5′-3′)	Fragment size (bp)	Primers (forward-reverse)
V3–V4	CCTACGGGNGGCWGCAG	GACTACHVGGGTATCTAATCC	430	341F-805R
V1–V2	GAAGAGTTTGATCATGGCTCAG	CTGCTGCCTCCCGTAG	315	27F-342R
V3	CCAGACTCCTACGGGAGGCAGC	TACCGCGGCTGCTGCTGGCAC	181	334F-515R
V4	GTGCCAGCMGCCGCGGTAA	GGACTACHVGGGTWTCTAAT	291	515F-806R
V5	GATTAGATACCCTGGTAG	CTTGTGCGGGCCCCCGTCAATTC	153	786F-939R
V6	TCGATGCAACGCGAAGAA	ACATTTCACAACACGAGCTGACGA	99	986F-1085R
V9	TGYACACACCGCCCGTC	AAGGAGGTGATCCANCCYCA	135	1406F-1541R

Germany) on an Illumina MiSeq platform (2 x 250 base pairs), with a targeted ca. 300,000 reads per sample pool.

Metagenome sequencing

Shotgun metagenomic sequencing for the samples EBI20, EOK05, EOK20, and EU08 was conducted by SeqIT Kaiserslautern (Germany), and for the samples DUN_AZE (Dunstaffnage salmon farm Allowable Zone of Effect), DUN_CE (Dunstaffnage salmon farm Cage Edge), DUN_REF (Dunstaffnage salmon farm Reference), LIS_AZE (Lismore salmon farm Allowable Zone of Effect), LIS_CE (Lismore salmon farm Cage Edge), LIS_REF (Lismore salmon farm Reference) by the NGS Competence Center Tübingen (NCCT, Germany). Therefore, in the first step, individual sequencing libraries were constructed from the genomic DNA of all samples using Illumina's DNA Prep, (M) Tagmentation kit (formerly Nextera DNA Flex library Prep Kit) with an Integrated DNA Technology (IDT) for Illumina DNA/RNA Unique Dual (UD) Indexes Set. Quantification and quality controls were conducted with a Qubit (ds DNA Assay Kit, Thermofisher) and with a Bioanalyzer (Agilent High Sensitivity DNA Kit, Agilent) in case of NCCT and with a Fluostar Optima system in case of SeqIT. Sequencing was conducted on an Illumina NextSeq2000 (P3, 300 cycles, 2 × 150 bp paired-end) in case of SeqIT and on an Illumina NovaSeq 6000 with an S4 Reagent kit v1.5 (300 cycles, 2 × 150 bp paired-end) in case of NCCT.

Sequence data processing

Hypervariable V3-V4 16S rRNA gene amplicons: Original FASTQ files from Dully et al. (2021) used in this study are deposited in NCBI's SRA database under BioProject PRJNA768445. FastQC quality control tool (Andrews, 2010) was used for the first evaluation of the quality of the raw sequences based on an assigned quality score across all bases. Primers were removed from raw reads using cutadapt version 2.21.1 (Martin, 2011) setting an error rate, in terms of mismatches, insertions, and deletions, equal to 0.15. Subsequently, sequences were processed in R (R core Team 2022) using the Divisive Amplicon Denoising Algorithm (DADA2) (Callahan et al., 2016). In specific, sequences were first filtered and trimmed using the 'filterAndTrim' function of DADA2. Reads with a maximum number of expected errors (maxEE) equal to 2 for the forward reads and 5 for the reverse reads were discarded. The truncation length, which was set to 230 bp, was decided by choosing the sequence position with a Phred score of ≥30 (Q3) for at least 51% of all reads in a dataset (= base call accuracy 99.9%) (Ewing et al., 1998). The paired-end sequences were merged ('*mergePairs*' function) using a minimum overlap of 20 bp and an allowed mismatch of two bases. The *uchime_denovo* function of *vsearch* (Rognes et al., 2016) was employed to identify and exclude potential chimeras. The resulting high-quality sequence reads were used for taxonomic classification against the SILVA v138.1 database (Pruesse et al., 2007) using the *–sintax* algorithm of VSEARCH with 80% bootstrap cutoff. Sequences that could not be assigned to the targeted domain (Bacteria) as well as Amplicon sequence variants (ASVs) with less than 10 reads were eliminated from the final ASV-to-sample matrix.

Multiple V-regions datasets: Raw sequence reads obtained from the Pool A and Pool B libraries were first demultiplexed using cutadapt version 2.21.1. Thereby, specific tags served as identifiers for each of the 10 samples (six samples from Scottish aquaculture installation sites, four samples from the Bay of Biscay) and specific PCR primer sequences for the seven individual hypervariable region fragments (V1–V2, V3, V4, V5, V6, V9). Each individual dataset ($n = 60$, 10 samples × 6 hypervariable region fragments) was then processed separately as described above for the hypervariable V3–V4 16S rRNA gene region. Filter and trim parameters were adjusted for each individual dataset as necessary, and details are provided in Supplementary Table S1:. Original data files are deposited in NCBI's SRA database under BioProject PRJNA942496.

Metagenomic datasets: For an initial quality control, raw sequence reads were analysed with fastp (Chen

et al., 2018), which revealed duplication rate, GC content, and overrepresentation for each dataset. Reads with an average score < 30 and a read length < 145 bp were discarded. We then extracted the 16S rRNA gene fragment reads from the filtered metagenomic datasets using Bowtie 2 (version 0.23.2; Langmead & Salzberg, 2012) and the SILVA v138.1 reference database (default minimum score threshold of $20 + 8 * \ln(L)$ was used, where L is the read length). The reference sequences to which our query sequences were aligned were then used for taxonomic classification of the read mapping with the same SILVA v138.1 database using the *–sintax* algorithm of VSEARCH with 80% bootstrap cutoff, as used for amplicon data. To improve our classification accuracy, the reads that passed minimum thresholds with bowtie mapping were used for sequence assembly to full-length genes with EMIRGE (Miller et al., 2011; Miller et al., 2013). The interleaved 16S filtered reads were separated into forward and reverse reads, and these were assembled into full-length ribosomal genes using EMIRGE (nr iterations = 40, join_threshold = 0.97). Length-normalized abundance estimates were retrieved for abundance assessment which was used along with the obtained contig names to build an abundance table per sample and contig. Taxonomy was assigned to each contig using *–sintax* algorithm of VSEARCH with 80% bootstrap cutoff. Sequences that could not be assigned to the targeted domain (Bacteria) were not used for downstream analyses. Metagenome data files are available in NCBI's SRA database under BioProject PRJNA943618.

Statistical analysis

In accordance with the aim of this study, statistical analyses were conducted to compare the three different datasets (traditional hypervariable V3–V4 16S rRNA gene datasets, multiple hypervariable regions datasets obtained with Nimagen's EasySeq™ RC-PCR 16S rRNA variable region kit, metagenome datasets) in terms of the information that they provide about microbial community structures of the coastal sediment samples analysed. Rarefaction curves for all amplicon datasets (including the traditional V3–V4 amplicons as well as the RC-PCR obtained amplicons), which served to assess sampling saturation depth, were made with the 'rarecurve' function of the R package 'vegan' (Oksanen et al., 2020). Before executing the scripts to calculate rarefied ASV richness (vegan's 'specnumber' function) and Shannon index (H') (vegan's 'diversity' function) the original amplicon datasets (including the traditional V3–V4 amplicons as well as the RC-PCR obtained amplicons) were normalized to the same size using the 'rarefy_even_depth' function of the R package 'phyloseq' (McMurdie & Holmes, 2013). This normalization accounted for uneven sample sizes.

To investigate the relationship between the number of obtained unprocessed and processed high-quality sequence reads per hypervariable region (predictor variable) with the corresponding length of each hypervariable region (independent variable) we performed linear regression modelling using the R function '*lm*' of the R package 'stats' (R base package, R core Team 2022). We used hierarchical clustering based on Euclidean distances to visualize the similarity/dissimilarity of the taxonomic composition of bacterial communities obtained from the seven hypervariable regions of the RC-PCR approach. The distances were calculated based on the taxa abundance per hypervariable gene fragment with the R function 'dist' (R base package) and transformed into a dendrogram (R function 'hclust' of the R base package). To compare beta-diversity patterns of all samples obtained by each individual hypervariable region, all amplicon datasets were subjected to non-metric multidimensional scaling (NMDS) analyses based on Bray–Curtis (BC) dissimilarity matrices ('*metaMDS*' function of the vegan R package). For NMDS analyses of the metagenomics dataset, we have mapped the 16S rRNA gene fragments to their closest relative reference sequence from the SILVA database. We then considered the number of fragments mapped to the same reference sequence as equivalent to the number of sequence reads assigned to an ASV. Taxonomic units that were shared by or exclusive for individual datasets were visualized using Venn diagrams ('*venn.diagram*' function of the VennDiagram R package; Chen & Boutros, 2011).

RESULTS

Sequence data overview

Hypervariable V3–V4 region of the 16S rRNA gene: The total number of unprocessed sequence reads in this dataset was 1,616,064 (each for R1 and R2). After quality filtering and merging we obtained 647,029 V3–V4 high-quality sequences, which grouped into a total of 154,626 ASVs. After elimination of ASVs with <10 sequences, we were left with 355,602 V3–V4 high-quality sequences, which grouped into a total of 6098 ASVs for downstream analyses. Prior to normalization (rarefying) sample EOK05 from the Bay of Biscay had the lowest number of sequences (6815 grouping into 184 ASVs), and the cage edge sample from the LIS aquaculture installation (LIS_CE) had the highest number of sequences (66,587 grouping into 1163 ASVs). Details for all individual samples as well as thresholds for normalization are provided in Supplementary Table S2:. Rarefaction analyses of the V3–V4 datasets indicated saturated sampling for all 10 samples, including EOK05 (Supplementary Figure S1).

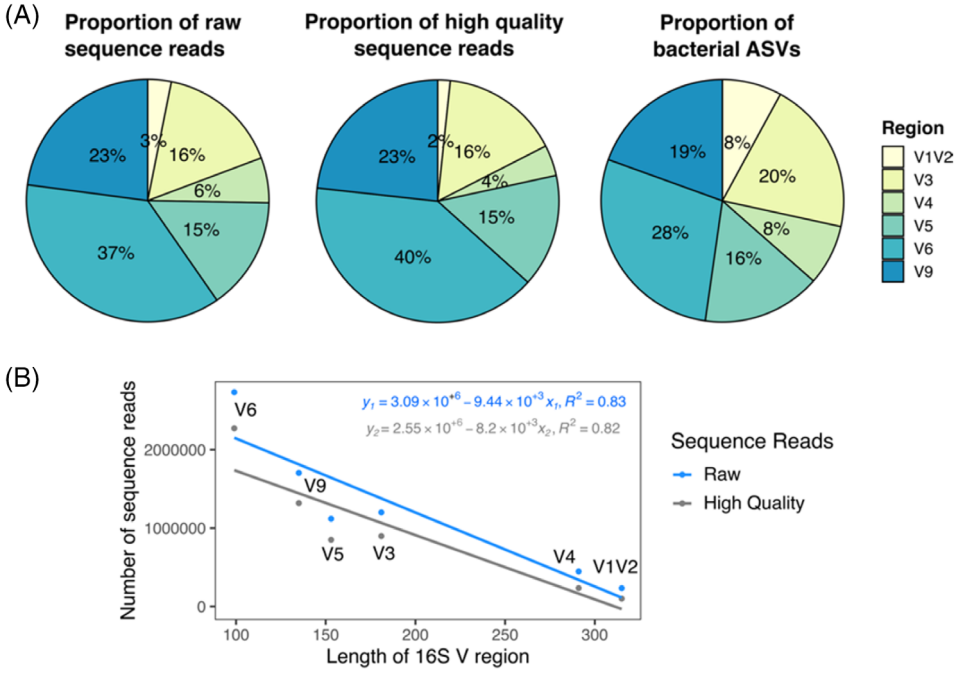

FIGURE 1 (A) Relative distribution of unprocessed and processed sequence reads and ASVs across the different V regions of the RC-PCR approach and (B) relation between *n* sequence reads (unprocessed and processed) and lengths of hypervariable fragments.

Multiple hypervariable regions: The total number of sequence reads obtained for Pool A and Pool B was 8,678,814 (each for R1 and R2). After demultiplexing and quality filtering, we were left with 5,817,980 merged high-quality target reads. Elimination of chimeric sequences and ASVs with <10 sequences resulted in a total of 5,642,220 reads: 94,026 reads for the V1–V2 fragment pool, 892,154 reads for the V3 fragment pool, 231,915 reads for the V4 fragment pool, 850,082 reads for the V5 fragment pool, 2,261,565 reads for the V6 fragment pool and 1,312,478 reads for the V9 fragment pool, which were used for downstream analyses. Plotting the number of obtained unprocessed sequence reads and the number of final high-quality sequence reads per hypervariable region against the corresponding lengths of each hypervariable region, we observed a significant negative linear relationship ($F_{1,4} = 20.1$, $p < 0.05$, $R^2 = 0.83$ for unprocessed sequence reads, and $F_{1,4} = 18.77$, $p < 0.05$, $R^2 = 0.82$ for high-quality sequences reads). Accordingly, the smallest hypervariable region (V6) accounted for most of the obtained unprocessed and high-quality reads (37% and 40%, respectively) while the largest fragment (V1–V2 region) accounted for the lowest proportions (3% and 2%, respectively) (Figure 1A, B). Interestingly, in the final high-quality dataset set used for downstream analyses, the numeric discrepancies sequence numbers obtained for all individual RC-PCR gene fragments are less pronounced compared to the original sequence datasets for the individual fragments prior to quality check. Rarefaction analyses (Supplementary Figure S1:) indicated a saturated sampling for all samples and all individual 16S rRNA fragments. However, one cage edge sample from an aquaculture installation (LIS_CE) stood out from all other samples because of a very low number of sequence reads and ASVs in the V1–V2 and the V5 datasets. Therefore, these two samples were removed before rarefying the individual datasets for alpha-diversity analyses (see below).

TABLE 2 Overview of obtained unprocessed metagenomic reads per sample and the number of extracted fragments mapping to 16S rRNA gene fragments of the SILVA database.

Sample	*n* obtained unprocessed reads	*n* extracted 16S rRNA gene fragments
DUN_CE	267,821,796	373,322
DUN_AZE	208,524,184	260,670
DUN_REF	218,158,529	206,568
LIS_CE	198,017,531	289,020
LIS_AZE	196,380,477	191,982
LIS_REF	215,188,776	131,208
EBI20	251,638,621	287,558
EOK05	253,547,486	658,518
EOK20	301,630,477	331,906
EU08	249,438,850	320,850

Metagenomics dataset: Table 2 provides a summary of the obtained unprocessed metagenomics reads for each of the 10 samples as well as the number of extracted 16S rRNA gene fragments.

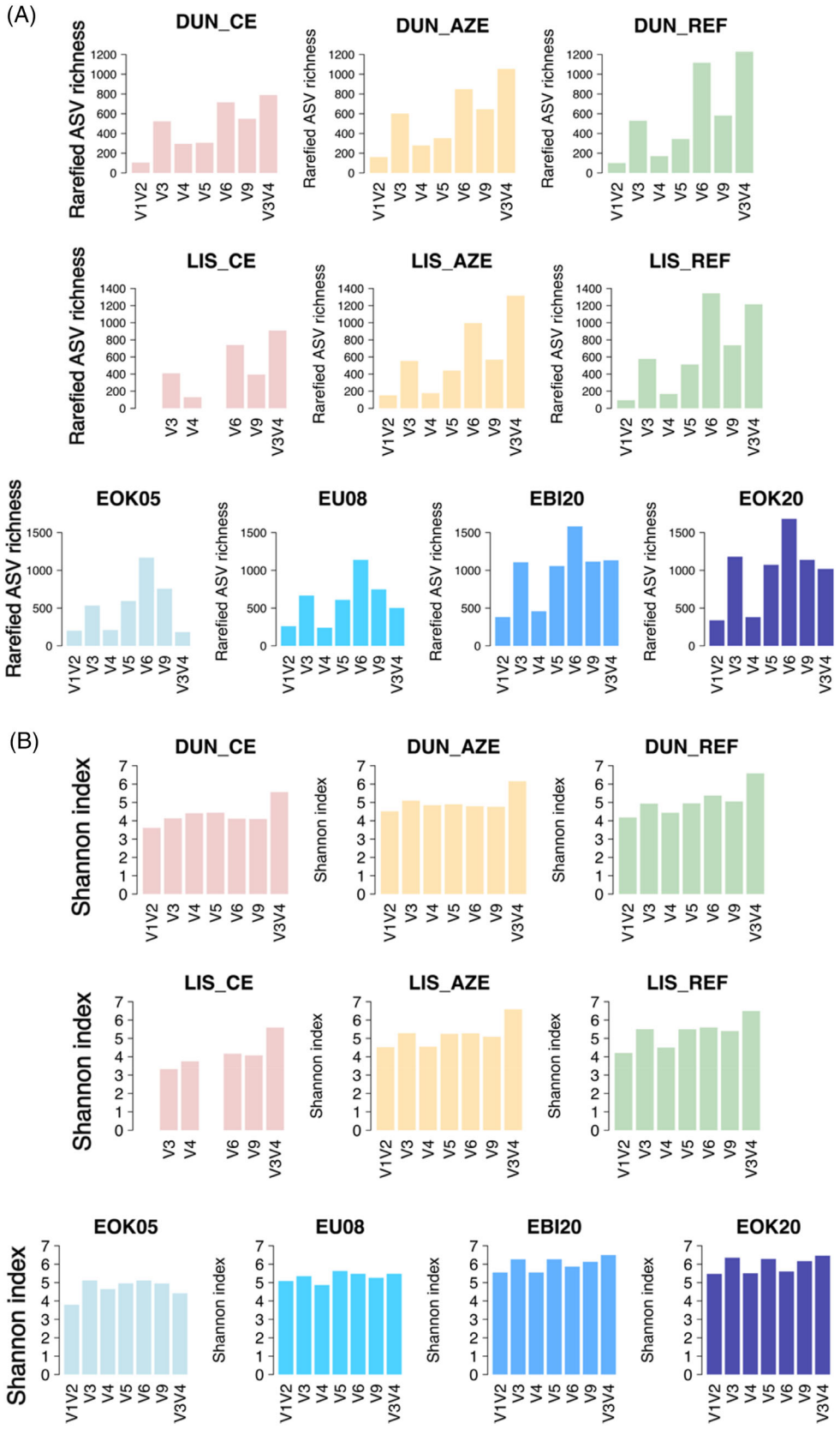

FIGURE 2 Legend on next page.

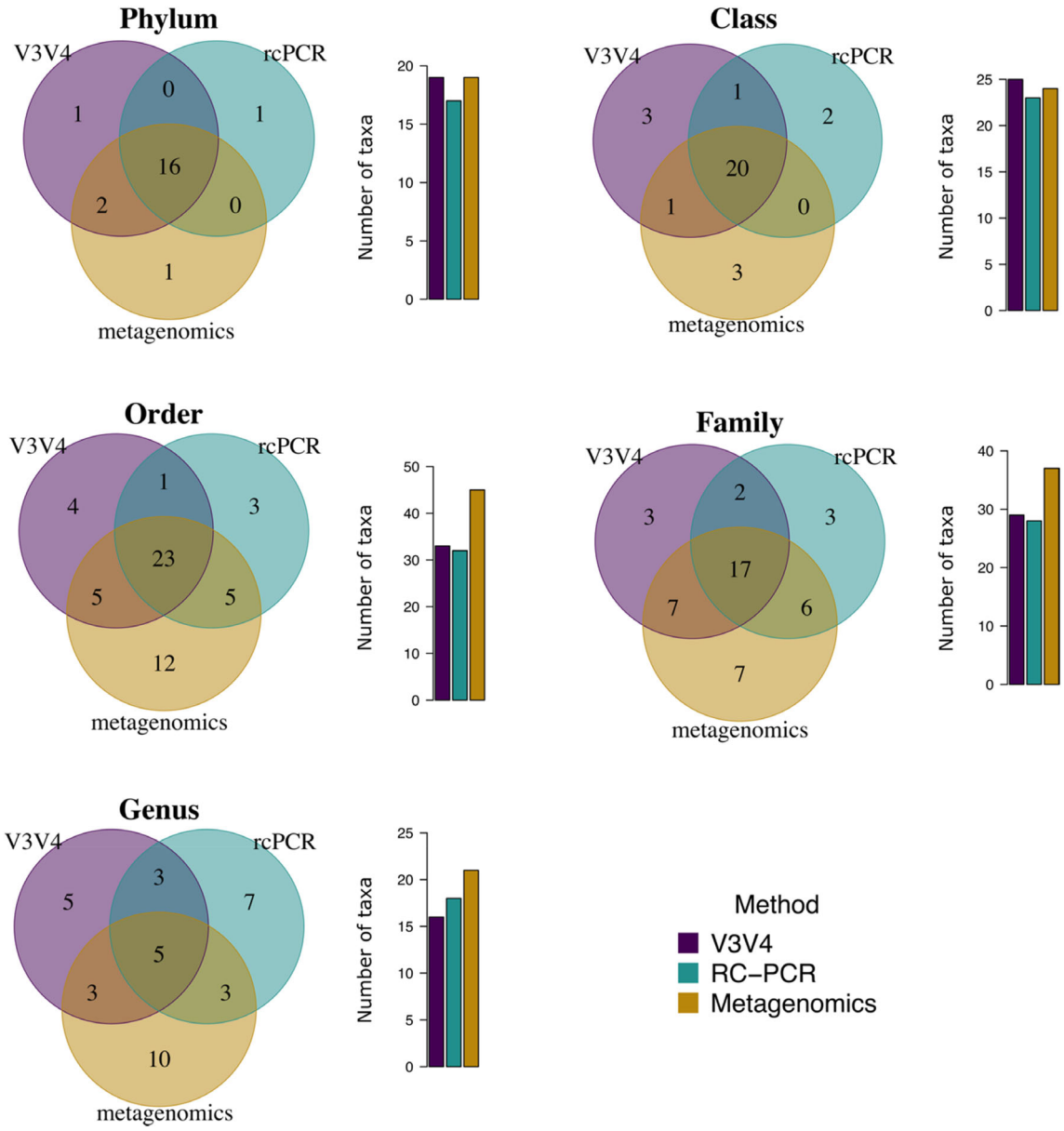

FIGURE 3 Venn diagrams showing the number of unique and shared taxonomic units (contributing at least 0.5% to the total abundance) uncovered in the pooled samples from the RC-PCR approach, the standard V3–V4 approach, and the PCR-independent metagenome reference data at different taxonomic ranks.

Alpha diversity of V3–V4 hypervariable region datasets and multiple hypervariable regions datasets

Different hypervariable gene fragments of the bacterial SSU rRNA reported remarkably different ASV richness values (Figure 2A). In all salmon farm datasets, the V6 region obtained from the RC-PCR approach and the V3–V4 region from the standard approach had the highest values while the V1–V2 and the V4 fragments had the lowest. A similar picture was obtained for the Basque coast samples with the exception that the standard V3–V4 fragment reported a notably lower ASV richness than the V4 fragment, compared to the salmon farm samples. Contrary to rarefied ASV richness, all V fragments reported a very similar Shannon index *H*′

FIGURE 2 Alpha diversity estimated as (A) rarefied ASV richness and (B) Shannon index for the V regions of each sample. Colours indicate the environmental gradient of the samples (red: near salmon cage, yellow: intermediate zone, and green: reference samples for the Scottish salmon farms, and a blue gradient from the poor quality to very good quality samples for Spain samples). Due to low sampling size, LIS_CE samples were removed from the calculation of alpha-diversity measures for the V1–V2 and the V5 fragments to increase rarefied normalization values for all other samples (see results section 'Sequence data overview').

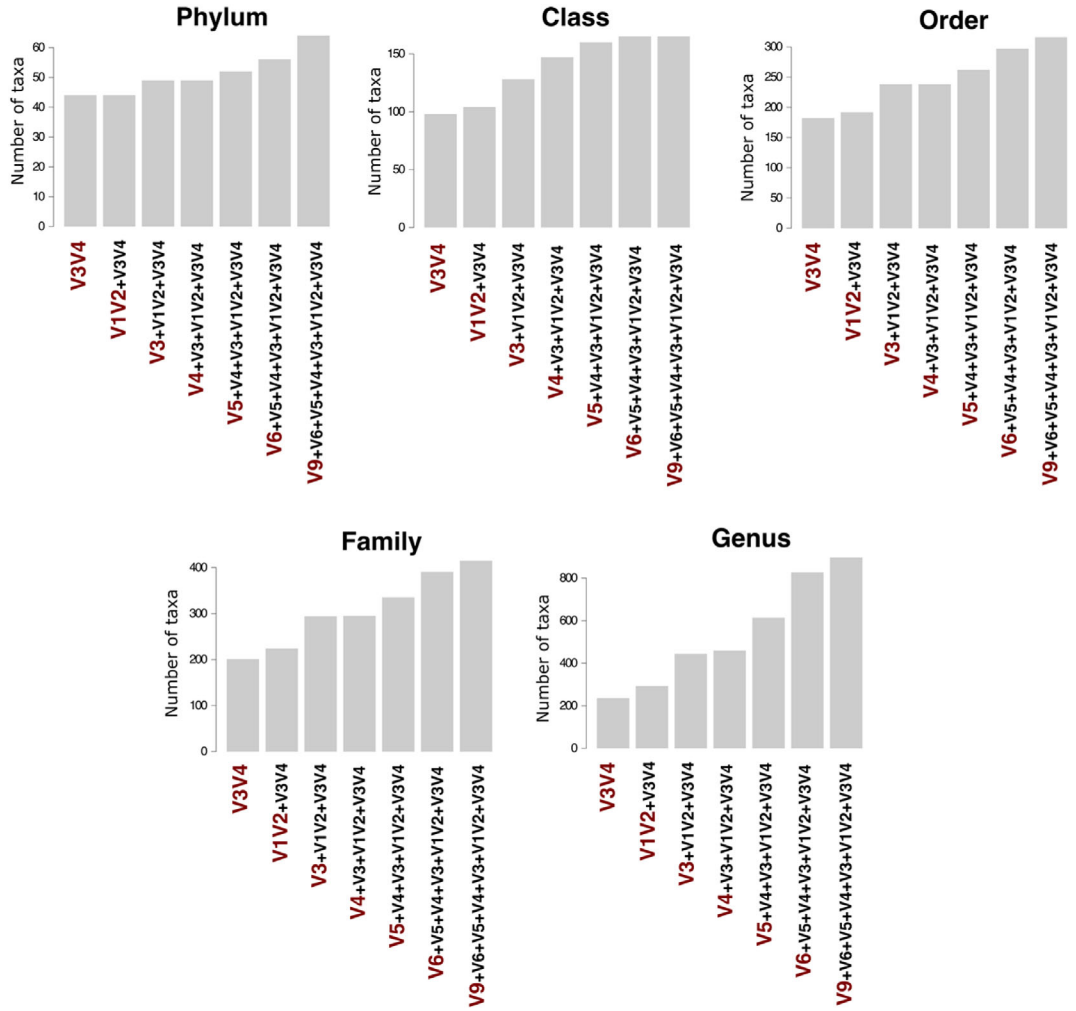

FIGURE 4 Number of taxonomic groups added to the bacterial diversity obtained from the standard V3–V4 approach by each of the six individual hypervariable gene fragments of the RC-PCR approach. Data are shown for different taxonomic ranks.

(Figure 2B), except for the *H′* of the standard V3-V4 fragment at aquaculture installations, which was slightly higher compared to all other V fragments.

Unique and shared taxonomic units reported by the standard V3–V4 16S rRNA gene fragment, the RC-PCR approach, and the metagenomics datasets

To obtain an overview of which taxonomic groups are generally reported by the three different datasets (standard V3–V4 fragment, by the RC-PCR approach, and metagenomics datasets), we pooled all 10 samples per approach. In the first step, we then compared the number of shared and exclusive taxonomic units contributing at least 0.5% to the total abundance among the three datasets on different taxonomic ranks (Figure 3). The metagenome data suggested the presence of 19 bacterial phyla in the pooled samples. Two of these phyla were missed by the RC-PCR approach and one by the V3–V4 approach. At class rank, metagenomes revealed 24 classes, from which four were missed by RC-PCR and three by the V3–V4 amplicons. For lower taxonomic ranks (order, family, genus), neither the standard V3–V4 protocol nor the RC-PCR approach was able to record the taxonomic repertoire of the metagenomics dataset. The discrepancy between what is there (as revealed by metagenomics reference data) and what was recorded with either of the two PCR-based approaches, increased at these lower taxonomic ranks (see also Supplementary Table S3:).

At all analysed taxonomic ranks (phylum—genus) and for all samples, the RC-PCR approach notably expanded the bacterial inventory that we obtained from the standard V3–V4 approach (Figure 4). Thereby, the largest fragment (V1–V2) added the least 'novel' diversity, while the short V6 fragment added the most to the repertoire obtained from the V3-V4 standard approach. For example, the RC-PCR approach detected 2.8 times more bacterial genera compared to the traditional one-fragment standard approach.

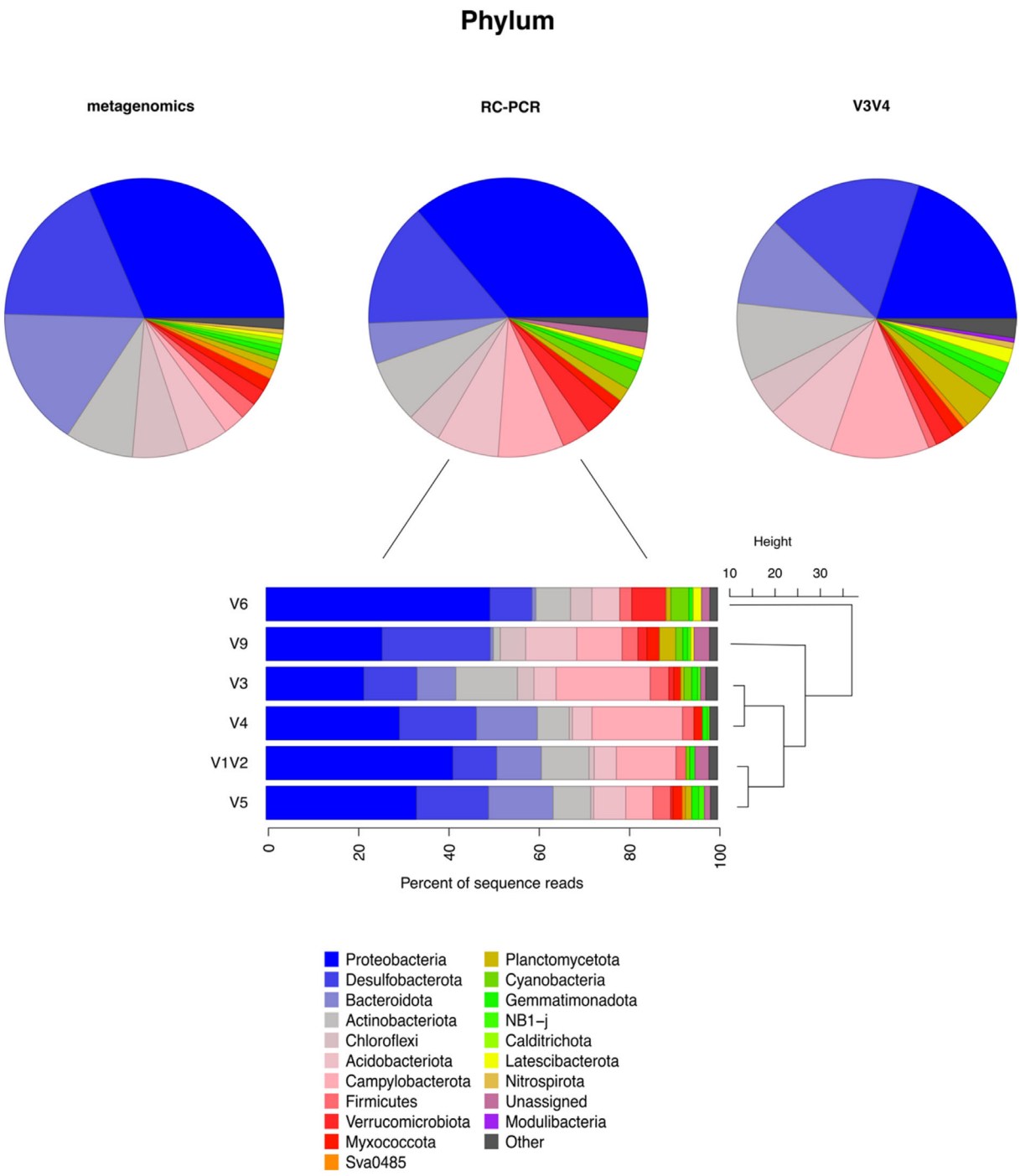

FIGURE 5 Detected bacterial phyla with the standard V3–V4 approach, the RC-PCR approach, and the PCR-free metagenomics data. In each case, data were pooled across all 10 samples. For the RC-PCR approach, the detection of bacterial phyla is also shown for each individual gene fragment. Euclidean distances as a measure of similarity between the individual gene fragments are visualized in the dendrogram. Phyla with less than 0.5% relative abundance by each method are grouped under 'Other' taxa. For the full list of phyla see Supplementary Table S4:.

Qualitative taxonomic repertoire reported by the standard V3–V4 16S rRNA gene fragment, the RC-PCR approach, and the metagenomics datasets

The most abundant phyla (inferred from ASVs that account for at least 0.5% of the sequence abundance within a sample) detected by the RC-PCR approach and the standard V3–V4 approach were congruent (Figure 5). The V5 fragment had the highest similarity to the community structure of the PCR-independent metagenomics dataset (BC similarity = 0.66). In the standard V3–V4 approach, Proteobacteria were underrepresented and Campylobacterota were overrepresented

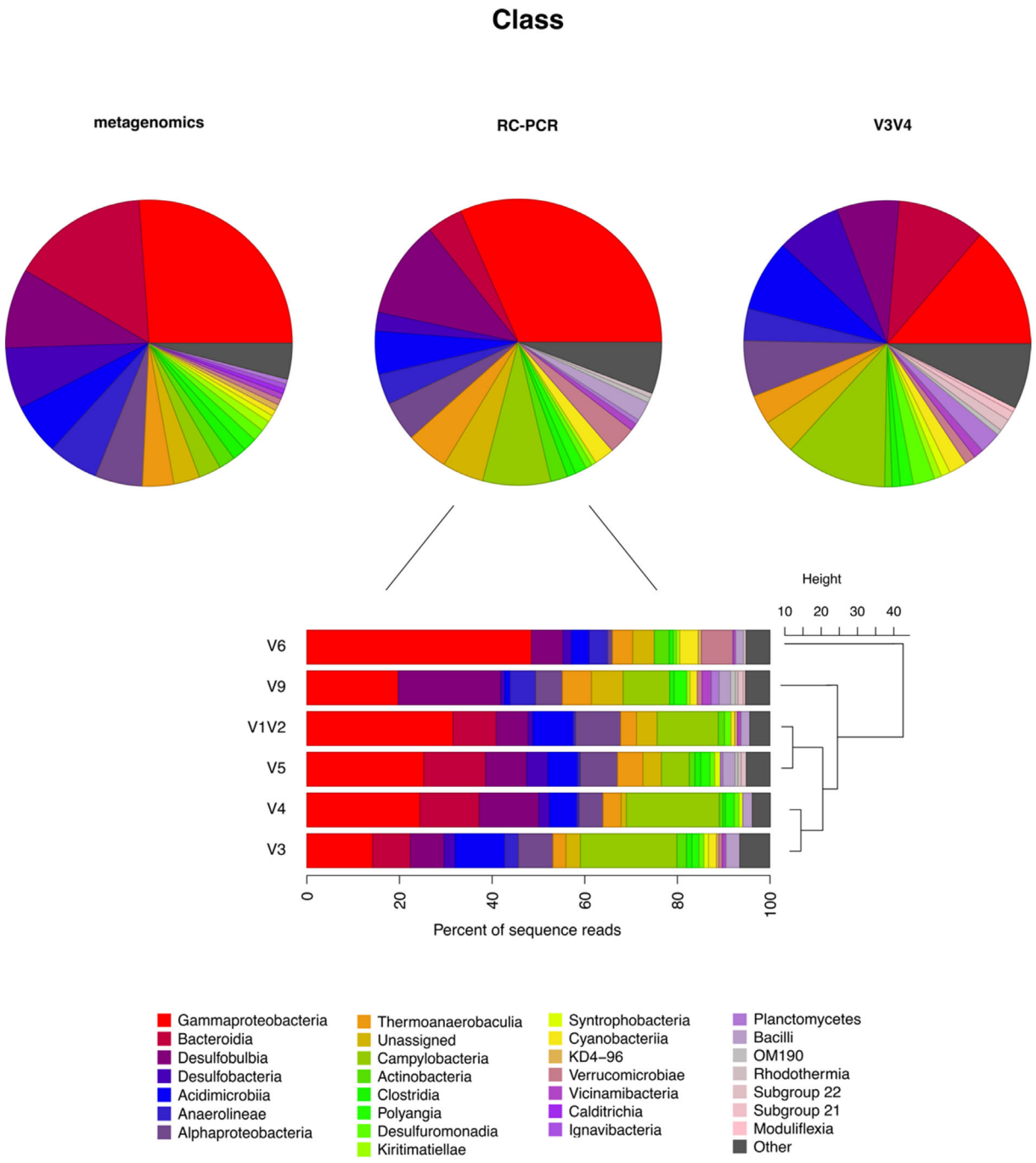

FIGURE 6 Detected bacterial classes with the standard V3–V4 approach, the RC-PCR approach, and the PCR-free metagenomics data. In each case, data were pooled across all 10 samples. For the RC-PCR approach, the detection of bacterial classes is also shown for each individual gene fragment. Euclidean distances as a measure of similarity between the individual gene fragments are visualized in the dendrogram. Classes with less than 0.5% relative abundance by each method are grouped under 'Other' taxa. For the full list of classes see Supplementary Table S4:.

compared to the metagenomics dataset. Among all V fragments of the RC-PCR approach, the dominant phyla were largely the same (Proteobacteria, Campylobacterota, Actionobacteria, Desulfobacter, Bacteroidota, Acidobacteriota, Firmicutes; see Figure 5). However, the proportion of sequences assigned to these phyla was notably different between fragments. In particular, the shorter V fragments, V6 and V9, stood out from the longer fragments in this respect (see for example the proportion of sequence reads assigned to Bacteroidota).

At the class level, the V3–V4 dataset missed two classes which were detected with the RC-PCR fragments (Bacilli and Rhodothermia, Figure 6). And both

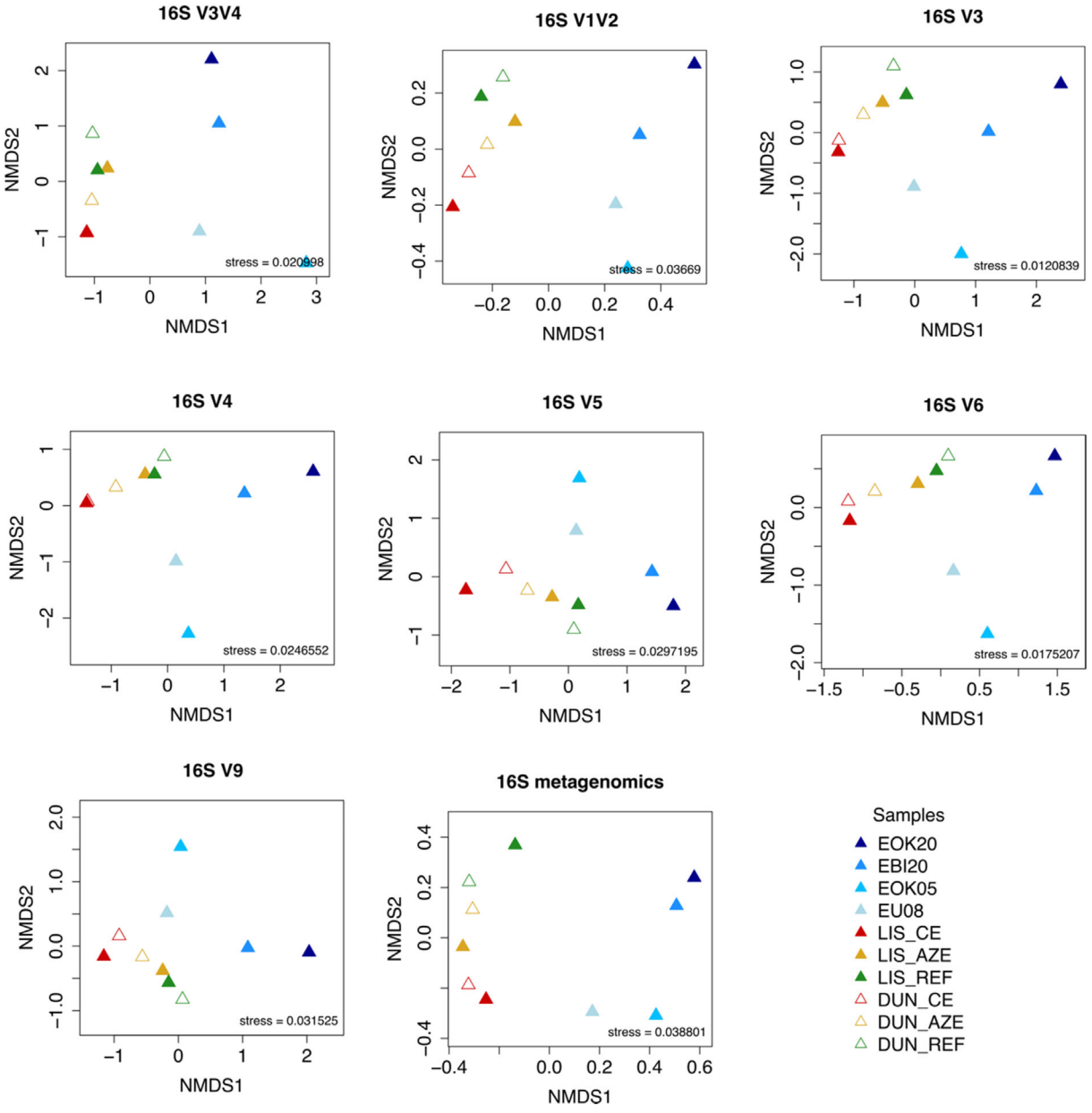

FIGURE 7 Non-metric multidimensional scaling (NMDS) based on Bray Curtis dissimilarity. The dissimilarity values are calculated using the number of sequences assigned to Amplicon Sequence Variants for the 16S hypervariable regions (for each RC-PCR obtained region and the standard V3–V4 region) and the number of 16S rRNA gene fragments that mapped on individual reference sequences in the SILVA database in the case of the metagenomics datasets.

PCR-based approaches missed several classes that were recorded with the PCR-independent metagenome dataset. Once more, it was the Vgment that had the highest similarity to the bacterial community structure inferred from the PCR-independent metagenomics dataset (BC similarity = 0.61). In the standard V3–V4 dataset, gamma-proteobacteria were underrepresented and Campylobacteria were overrepresented compared to the metagenomics dataset. When comparing the class-level inventory of the individual V fragments of the RC-PCR with each other (Figure 6), we found a high congruency among the longer V fragments. In general, gamma-proteobacteria, Campylobacteria, Desulfobulbia, Acidimicrobiia, alpha-proteobacteria, Bacteroidia, Bacilli, and Actinobacteria subgroup 22 were the most abundant classes. In accordance with the observations on phylum level, once more the two shortest fragments, V6 and V9, were most different to the longer V fragments when considering the proportion of sequence reads assigned to the dominant classes.

Differences in dissimilarity patterns

The NMDS of all 10 samples showed a clear separation of the six samples from the aquaculture installations

and the four samples from the Bay of Biscay along axis 1 in the standard V3–V4 approach (Figure 7). Axis 2 mirrored the gradient of environmental impact. Thus, the bacterial community structures inferred from the standard V3–V4 approach allowed us to distinguish habitat types (aquaculture installations versus Bay of Biscay, axis 1) as well as environmental impact (high to low, axis 2). Likewise, all six individual fragments of the RC-PCR approach separated habitat types and environmental impact. However, the V1–V2 fragment indicated a higher resolution, because it succeeded to separate the two cage edge samples of the DUN and LIS farms in the multidimensional space, and also showed a clearer clustering pattern comparing the LIS_AZE and LIZ_REF samples (however, please note the 10-fold lower scale of the axes, which places this observation into perspective—BC distance DUN_CE and LIS_CE was 0.1 for the V3–V4 fragment and 0.34 for the V1–V2 fragment). The V1–V2 fragment most closely mirrored the patterns of the habitat type gradient (axis 1) and environmental impact (axis 2) of the standard V3–V4 fragment, while for most other RC-PCR fragments these gradients were not as strongly associated with axis 1 and 2, respectively.

Among both PCR-based datasets, the V1–V2 fragment of the RC-PCR approach and the standard V3–V4 fragment were most congruent with the beta-diversity pattern resulting from the metagenome dataset.

DISCUSSION

Improved characterisation of bacterial communities across environmental gradients can be achieved by molecular technologies which can help us gain a better understanding of microbial diversity changes and potentially infer emerging environmental changes (Milke et al., 2022). Here, we discuss the performance of a new approach based on RC-PCR that employs amplicon sequencing (partial 16S rRNA amplified products from six gene fragments) to characterise bacterial communities at different coastal marine environments. Our results show that this approach provides an added value to the traditional amplicon sequencing (one-fragment amplification) by increasing the taxonomic bacterial diversity. Considering the cost-efficiency of the RC-PCR approach, six hypervariable 16S rRNA gene fragments are obtained for nearly the same price as the traditional V3–V4 fragment of the traditional approach: while Illumina sequencing costs are identical, it is only the difference between the first-stage PCR of the traditional one-fragment PCR versus the costs for an RC-PCR kit that slightly increases the overall costs for the RC-PCR approach. The time needed to prepare the RC-PCR samples for Illumina sequencing is about the same as is needed for the traditional one-fragment PCR approach.

It was reported previously that different hypervariable 16S rRNA gene regions obtained from the same bacterial communities resulted in largely congruent beta-diversity patterns (Guo et al., 2016; Kerrigan et al., 2019). Thus, any of these V regions could in theory suffice to infer the general partitioning of bacterial diversity. This could question the necessity of analysing several hypervariable 16S rRNA gene regions in microbial diversity studies (Fuks et al., 2018). However, in spite of similar beta-diversity patterns obtained from all individual V regions, several other measures inferred from these V regions differed remarkably from each other. It is well known that the taxonomic annotation may vary substantially when using different fragments of the bacterial SSU rRNA (Logares et al., 2014), and thus the selection of the 16S region is crucial to associate taxonomic identities with specific environments (Bukin et al., 2019; Knight et al., 2018).

One explanation is the varying levels of the conservation of the target gene and the length of the amplicon product (Peterson et al. 2021). For example, Kerrigan et al. (2019) found the V6 region to be notably more variable compared to the V4 region. Yang et al. (2016) found V4–V6 to be the closest to the full-length 16S rRNA sequences, while the V2 and V8 regions were the least reliable. The correlation between the taxonomic resolution and the hypervariable gene region analysed can be a major limitation for the comparison between studies that use different primer sets (Kerrigan et al., 2019; Tremblay et al., 2015). Thus, several studies have aimed to evaluate the most suitable region and primer set for specific surveys, by performing in silico analyses or preliminary studies of specific types of samples or environments (Bukin et al., 2019). Others have suggested a combination of data derived from different amplified regions as a suitable approach to increase taxonomic coverage and approximate the resolution of a full-length approach (Fuks et al., 2018; Klindworth et al., 2013). To this end, the RC-PCR approach presented here, which maps seven of the nine hypervariable regions, is an efficient alternative, employing multiple primers in only two PCR reactions, providing information from a range of different region sizes (6%–20% of the full 16S rRNA gene), which can be combined to obtain more reliable results.

In addition to taxonomic resolution limitations when using DNA markers, the amplification step can introduce several biases, resulting from primer mismatches or degenerate bases (Guo et al., 2016; Klindworth et al., 2013; Tremblay et al., 2015). For example, Guo et al. (2016) reported a severe underestimation of a whole taxonomic entity (Verrucomicrobia) when targeting the V6–V8 region of the 16S rRNA. Thus, under- and overestimations of individual bacterial taxon groups could very well link to amplification biases.

Sequencing error rates and chimeras can also skew the results of amplicon datasets and read filtering is

necessary in these datasets (Guo et al., 2016; Knight et al., 2018). This explains the higher loss of information during data cleaning of the longer V fragments obtained via RC-PCR compared to the smaller V fragments.

To overcome the fundamental limitations of the partial 16S hypervariable regions related to amplicon length and conservation level, and to avoid PCR biases, we used metagenome data from which we extracted and assembled the complete 16S rRNA gene. A known advantage of the shotgun approach is an increase in taxonomic resolution (Perez-Cobas et al., 2020) which corroborates with the findings of our study. However, the overall bacterial coverage was more limited with metagenomics. This might be due to the assembly step that introduces a potential bias in which low-abundance species might remain unassembled (Thomas et al., 2012). And/or it is indicated that the sequencing depth might not have been sufficient for the detection of some rare taxa which could be detected with both our PCR-based approaches, (Tessler et al., 2017; Zaheer et al., 2018). The overall cost for preparing, sequencing in a sufficient sequencing depth, and analysing the metagenomics samples is higher than amplicon analysis (Knight et al., 2018) and thus metagenomics is more often chosen when aiming to go beyond the taxonomic profiles of a microbial community (provided efficiently also by the amplicon sequencing) to the functional ones. Notably, the direct comparison between metagenomics with PCR-based approaches has in some cases led to controversial results, revealing significant differences between the two methods (Clooney et al., 2016; Tessler et al., 2017). Our goal when using metagenomics data here was only to utilize the full-length 16S rRNA gene as a reference in our comparisons between RC-PCR with the traditional one-fragment amplicon analysis. The use of third-generation platforms that provide the full-length 16S V1–V9 regions (ca. 1600 bp) is currently being assessed (Santos et al., 2020; Tedersoo et al., 2021) as a future strategy for sequencing microbial communities.

CONCLUSIONS

We evaluated the performance of a RC-PCR-based approach for bacterial identification in marine environmental samples. This method can simultaneously provide data deriving from seven out of nine hypervariable 16S rRNA gene regions. We demonstrated that each of the individual gene fragments of the RC-PCR approach increased the number of bacterial taxa that we obtained from the standard V3–V4 protocol at all taxonomic ranks (phylum, class, order, family, genus). Since the seven targeted V regions of the RC-PCR protocol can be amplified in only two PCR reactions, it is a time-efficient solution for a multiple primer approach. Additionally, the RC-PCR approach is fully automatable, and, thus, ideal for routine applications such as compliance biomonitoring. The data deriving from the Illumina sequencing can be demultiplexed and analysed for each gene fragment and subsequently combined to obtain a high fraction of taxonomic information. This strategy makes the microbial community analysis using seven V regions only slightly more expensive than the traditional one-fragment approach. In particular, compared to metagenomics, the RC-PCR-based approach is a more cost-efficient method. Overall, we suggest that combining the information from the different gene fragments targeted by the RC-PCR-based approach would enable a time, cost-efficient, and informative taxonomic profiling of microbial communities in environmental samples.

AUTHOR CONTRIBUTIONS

Kleopatra Leontidou: Data curation (lead); formal analysis (lead); investigation (equal); methodology (equal); software (lead); visualization (lead); writing – original draft (equal). **Ion L. Abad-Recio:** Data curation (equal); formal analysis (equal); investigation (equal); software (equal); writing – review and editing (equal). **Verena Rubel:** Resources (equal); writing – review and editing. **Sabine Filker:** Funding acquisition (equal); methodology (equal); resources (equal); writing – review and editing (equal). **Martin Däumer:** Investigation (equal); methodology; writing – review and editing (equal). **Alexander Thielen:** Investigation (equal); methodology (equal); writing – review and editing (equal). **Anders Lanzén:** Conceptualization (equal); investigation (equal); methodology (equal); supervision (equal); validation (equal); writing – review and editing (equal). **Thorsten Stoeck:** Conceptualization (lead); funding acquisition (equal); methodology; project administration (equal); supervision (lead); validation; writing – original draft.

ACKNOWLEDGEMENTS

This study was supported by a grant of the Deutsche Forschungsgemeinschaft for a joined project between the RPTU Kaiserslautern-Landau, awarded to TS and SF (grants STO 414/19-1 and Fi 2089/3-1) and AZTI. Furthermore, we thank Thomas A. Wilding of SAMS (Scottish Association for Marine Science), Oban, UK, for collecting and providing samples from the salmon aquaculture installation and Hans-Werner Breiner (RPTU) for his contribution to lab work. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data and the R code used in the manuscript for the production of the figures shown in the results

section are shared in figshare [DOI: 10.6084/m9.figshare.22276948]. https://figshare.com/s/c574b7358fc7d2e49ac1

ORCID

Thorsten Stoeck https://orcid.org/0000-0001-5180-5659

REFERENCES

Andrews, S. (2010) FastQC: a quality control tool for high throughput sequence data [online]. Available Online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Aylagas, E., Atalah, J., Sanchez-Jerez, P., Pearman, J.K., Casado, N., Asensi, J. et al. (2021) A step towards the validation of bacteria biotic indices using DNA metabarcoding for benthic monitoring. *Molecular Ecology Resources*, 21(6), 1889–1903. Available from: https://doi.org/10.1111/1755-0998.13395

Brandt, L.D. & House, C.H. (2016) Marine subsurface microbial community shifts across a hydrothermal gradient in Okinawa trough sediments. *Archaea*, 2016, 2690329. Available from: https://doi.org/10.1155/2016/2690329

Bukin, Y.S., Galachyants, Y.P., Morozov, I.V., Bukin, S.V., Zakharenko, A.S. & Zemskaya, T.I. (2019) The effect of 16S rRNA region choice on bacterial community metabarcoding results. *Scientific Data*, 6, 190007. Available from: https://doi.org/10.1038/sdata.2019.7

Bus, M.M., de Jong, E.A., King, J.L., van der Vliet, W., Theelen, J. & Budowle, B. (2021) Reverse complement-PCR, an innovative and effective method for multiplexing. *BioTechniques*, 71(3), 484–489. Available from: https://doi.org/10.2144/btn-2021-0031

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J. & Holmes, S.P. (2016) DADA2: high-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581–583. Available from: https://doi.org/10.1038/nmeth.3869

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J. et al. (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America*, 108(Suppl 1), 4516–4522. Available from: https://doi.org/10.1073/pnas.1000080107

Chen, H. & Boutros, P.C. (2011) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*, 12, 35. Available from: https://doi.org/10.1186/1471-2105-12-35

Chen, S., Zhou, Y., Chen, Y. & Gu, J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890. Available from: https://doi.org/10.1093/bioinformatics/bty560

Clark, D.E., Pilditch, C.A., Pearman, J.K., Ellis, J.I. & Zaiko, A. (2020) Environmental DNA metabarcoding reveals estuarine benthic community response to nutrient enrichment—evidence from an in-situ experiment. *Environmental Pollution*, 267, 115472. Available from: https://doi.org/10.1016/j.envpol.2020.115472

Clooney, A.G., Fouhy, F., Sleator, R.D., O' Driscoll, A., Stanton, C., Cotter, P.D. et al. (2016) Comparing apples and oranges?: next generation sequencing and its impact on microbiome analysis. *PLoS One*, 11(2), e0148028. Available from: https://doi.org/10.1371/journal.pone.0148028

Coelho, F.J., Louvado, A., Domingues, P.M., Cleary, D.F., Ferreira, M., Almeida, A. et al. (2016) Integrated analysis of bacterial and microeukaryotic communities from differentially active mud volcanoes in the Gulf of Cadiz. *Scientific Reports*, 6, 35272. Available from: https://doi.org/10.1038/srep35272

Coolen, J.P.M., Wolters, F., Tostmann, A., van Groningen, L.F.J., Bleeker-Rovers, C.P., Tan, E. et al. (2021) SARS-CoV-2 whole-genome sequencing using reverse complement PCR: for easy, fast and accurate outbreak and variant analysis. *Journal of Clinical Virology*, 144, 104993. Available from: https://doi.org/10.1016/j.jcv.2021.104993

Cordier, T., Alonso-Saez, L., Apotheloz-Perret-Gentil, L., Aylagas, E., Bohan, D.A., Bouchez, A. et al. (2021) Ecosystems monitoring powered by environmental genomics: a review of current strategies with an implementation roadmap. *Molecular Ecology*, 30(13), 2937–2958. Available from: https://doi.org/10.1111/mec.15472

Cordier, T., Forster, D., Dufresne, Y., Martins, C.I.M., Stoeck, T. & Pawlowski, J. (2018) Supervised machine learning outperforms taxonomy-based environmental DNA metabarcoding applied to biomonitoring. *Molecular Ecology Resources*, 18(6), 1381–1391. Available from: https://doi.org/10.1111/1755-0998.12926

Dully, V., Rech, G., Wilding, T.A., Lanzen, A., MacKichan, K., Berrill, I. et al. (2021) Comparing sediment preservation methods for genomic biomonitoring of coastal marine ecosystems. *Marine Pollution Bulletin*, 173(Pt B), 113129. Available from: https://doi.org/10.1016/j.marpolbul.2021.113129

Ewing, B., Hillier, L., Wendl, M.C. & Green, P. (1998) Base-calling of automated sequencer traces using Phred I. Accuracy assessment. *Genome Research*, 8(3), 175–185. Available from: https://doi.org/10.1101/gr.8.3.175

Falkowski, P.G., Fenchel, T. & Delong, E.F. (2008) The microbial engines that drive Earth's biogeochemical cycles. *Science*, 320(5879), 1034–1039. Available from: https://doi.org/10.1126/science.1153213

Frühe, L., Dully, V., Forster, D., Keeley, N.B., Laroche, O., Pochon, X. et al. (2021) Global trends of benthic bacterial diversity and community composition along organic enrichment gradients of Salmon farms. *Frontiers in Microbiology*, 12, 637811. Available from: https://doi.org/10.3389/fmicb.2021.637811

Fuks, G., Elgart, M., Amir, A., Zeisel, A., Turnbaugh, P.J., Soen, Y. et al. (2018) Combining 16S rRNA gene variable regions enables high-resolution microbial community profiling. *Microbiome*, 6(1), 17. Available from: https://doi.org/10.1186/s40168-017-0396-x

Guo, J., Cole, J.R., Zhang, Q., Brown, C.T. & Tiedje, J.M. (2016) Microbial community analysis with ribosomal gene fragments from shotgun metagenomes. *Applied and Environmental Microbiology*, 82(1), 157–166. Available from: https://doi.org/10.1128/AEM.02772-15

Harrison, B.K., Myrbo, A., Flood, B.E. & Bailey, J.V. (2018) Abrupt burial imparts persistent changes to the bacterial diversity of turbidite-associated sediment profiles. *Geobiology*, 16(2), 190–202. Available from: https://doi.org/10.1111/gbi.12271

Herlemann, D.P., Labrenz, M., Jurgens, K., Bertilsson, S., Waniek, J.J. & Andersson, A.F. (2011) Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *The ISME Journal*, 5(10), 1571–1579. Available from: https://doi.org/10.1038/ismej.2011.41

Hong, S., Bunge, J., Leslin, C., Jeon, S. & Epstein, S.S. (2009) Polymerase chain reaction primers miss half of rRNA microbial diversity. *The ISME Journal*, 3(12), 1365–1373. Available from: https://doi.org/10.1038/ismej.2009.89

Jones, D.S., Flood, B.E. & Bailey, J.V. (2016) Metatranscriptomic insights into polyphosphate metabolism in marine sediments. *The ISME Journal*, 10(4), 1015–1019. Available from: https://doi.org/10.1038/ismej.2015.169

Keeley, N., Wood, S.A. & Pochon, X. (2018) Development and preliminary validation of a multi-trophic metabarcoding biotic index for monitoring benthic organic enrichment. *Ecological Indicators*, 85, 1044–1057. Available from: https://doi.org/10.1016/j.ecolind.2017.11.014

Kerrigan, Z., Kirkpatrick, J.B. & D'Hondt, S. (2019) Influence of 16S rRNA hypervariable region on estimates of bacterial diversity and community composition in seawater and marine sediment. *Frontiers in Microbiology*, 10, 1640. Available from: https://doi.org/10.3389/fmicb.2019.01640

Kieser, R.E., Bus, M.M., King, J.L., van der Vliet, W., Theelen, J. & Budowle, B. (2020) Reverse complement PCR: a novel one-step PCR system for typing highly degraded DNA for human identification. *Forensic Science International: Genetics*, 44, 102201. Available from: https://doi.org/10.1016/j.fsigen.2019.102201

Kim, M., Morrison, M. & Yu, Z. (2011) Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *Journal of Microbiological Methods*, 84(1), 81–87. Available from: https://doi.org/10.1016/j.mimet.2010.10.020

Kim, Y.H., Yang, I., Bae, Y.S. & Park, S.R. (2008) Performance evaluation of thermal cyclers for PCR in a rapid cycling condition. *Biotechniques*, 44(4), 495–496. Available from: https://doi.org/10.2144/000112705

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. et al. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, 41(1), e1. Available from: https://doi.org/10.1093/nar/gks808

Knight, R., Vrbanac, A., Taylor, B.C., Aksenov, A., Callewaert, C., Debelius, J. et al. (2018) Best practices for analysing microbiomes. *Nature Reviews. Microbiology*, 16(7), 410–422. Available from: https://doi.org/10.1038/s41579-018-0029-9

Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M. & Pace, N.R. (1985) Rapid determination of 16S RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences of the United States of America*, 82, 6955–6959. Available from: https://doi.org/10.1073/pnas.82.20.695

Langmead, B. & Salzberg, S.L. (2012) Fast gapped-read alignment with bowtie 2. *Nature Methods*, 9(4), 357–359. Available from: https://doi.org/10.1038/nmeth.1923

Lanzén, A., Mendibil, I., Borja, Á., & Alonso-Sáez, L. (2021) A microbial mandala for environmental monitoring: Predicting multiple impacts on estuarine prokaryote communities of the Bay of Biscay. *Molecular Ecology*, 30, 2969–2987. https://doi.org/10.1111/mec.15489

Laroche, E., Casiot, C., Fernandez-Rojo, L., Desoeuvre, A., Tardy, V., Bruneel, O. et al. (2018) Dynamics of bacterial communities mediating the treatment of an As-rich acid mine drainage in a field pilot. *Frontiers in Microbiology*, 9, 3169. Available from: https://doi.org/10.3389/fmicb.2018.03169

Liu, A.G. (2011) Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences.

Locey, K.J. & Lennon, J.T. (2016) Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 113(21), 5970–5975. Available from: https://doi.org/10.1073/pnas.1521291113

Logares, R., Sunagawa, S., Salazar, G., Cornejo-Castillo, F.M., Ferrera, I., Sarmento, H. et al. (2014) Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environmental Microbiology*, 16(9), 2659–2671. Available from: https://doi.org/10.1111/1462-2920.12250

Mardis, E.R. (2013) Next-generation sequencing platforms. *Annual Review of Analytical Chemistry*, 6, 287–303. Available from: https://doi.org/10.1146/annurev-anchem-062012-092628

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*, 17(1), 10–12.

McFarlin, K.M., Questel, J.M., Hopcroft, R.R. & Leigh, M.B. (2017) Bacterial community structure and functional potential in the northeastern Chukchi Sea. *Continental Shelf Research*, 136, 20–28. Available from: https://doi.org/10.1016/j.csr.2017.01.018

McMurdie, P.J. & Holmes, S. (2013) Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, 8(4), e61217. Available from: https://doi.org/10.1371/journal.pone.0061217

Milke, F., Sanchez-Garcia, S., Dlugosch, L., McNichol, J., Fuhrman, J., Simon, M. et al. (2022) Composition and biogeography of planktonic pro- and eukaryotic communities in the Atlantic Ocean: primer choice matters. *Frontiers in Microbiology*, 13, 895875. Available from: https://doi.org/10.3389/fmicb.2022.895875

Miller, C.S., Baker, B.J., Thomas, B.C., Singer, S.W. & Banfield, J.F. (2011) EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biology*, 12(5), 1–14. Available from: https://doi.org/10.1016/B978-0-12-407863-5.00017-4

Miller, C.S., Handley, K.M., Wrighton, K.C., Frischkorn, K.R., Thomas, B.C. & Banfield, J.F. (2013) Short-read assembly of full-length 16S amplicons reveals bacterial diversity in subsurface sediments. *PLoS One*, 8(2), 1–11. Available from: https://doi.org/10.1371/journal.pone.0056018

Moorlag, S.J.C.F.M., Coolen, J.P.M., van den Bosch, B., Jin, E.H.-M., Buil, J.B., Wertheim, H.F.L. et al. (2023) Targeting the 16S rRNA gene by reverse complement PCR next-generation sequencing: specific and sensitive detection and identification of microbes directly in clinical samples. *Microbiology Spectrum*, 11(3). Available from: https://doi.org/10.1128/spectrum.04483-22

Moreira, V.A., Cravo-Laureau, C., de Carvalho, A.C.B., Baldy, A., Bidone, E.D., Sabadini-Santos, E. et al. (2022) Microbial indicators along a metallic contamination gradient in tropical coastal sediments. *Journal of Hazardous Materials*, 443, 130244.

Nguyen, U.T., Lincoln, S.A., Valladares Juarez, A.G., Schedler, M., Macalady, J.L., Muller, R. et al. (2018) The influence of pressure on crude oil biodegradation in shallow and deep Gulf of Mexico sediments. *PLoS One*, 13(7), e0199784. Available from: https://doi.org/10.1371/journal.pone.0199784

Obi, C.C., Adebusoye, S.A., Ugoji, E.O., Ilori, M.O., Amund, O.O., & Hickey, W.J. (2016). Microbial communities in sediments of Lagos Lagoon, Nigeria: elucidation of community structure and potential impacts of contamination by municipal and industrial wastes. *Frontiers in Microbiology*, 7, 1213. https://doi.org/10.3389/fmicb.2016.01213

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D. et al. (2020) Vegan: community ecology package. R Package Version 2.5-7. https://CRAN.R-project.org/package=vegan

Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apotheloz-Perret-Gentil, L., Beja, P., Boggero, A., et al. (2018) The future of biotic indices in the ecogenomic era: integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. *Science of The Total Environment*, 637, 1295–1310. Available from: https://doi.org/10.1016/j.scitotenv.2018.05.002

Perez-Cobas, A.E., Gomez-Valero, L. & Buchrieser, C. (2020) Metagenomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses. *Microbial Genomics*, 6(8), 1–22. Available from: https://doi.org/10.1099/mgen.0.000409

Peterson, D., Bonham, K.S., Rowland, S., Pattanayak, C.W; RESONANCE Consortium; Klepac-Ceraj, V. (2021) Comparative analysis of 16S rRNA gene and metagenome sequencing in pediatric gut microbiomes. *Frontiers in Microbiology*, 12, 670336. Available from: https://doi.org/10.3389/fmicb.2021.670336

Pinto, A.J. & Raskin, L. (2012) PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PLoS One*, 7(8), e43093. Available from: https://doi.org/10.1371/journal.pone.0043093

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J. et al. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, 35(21), 7188–7196. Available from: https://doi.org/10.1093/nar/gkm864

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

Rosselli, R., Romoli, O., Vitulo, N., Vezzi, A., Campanaro, S., de Pascale, F. et al. (2016) Direct 16S rRNA-seq from bacterial

communities: a PCR-independent approach to simultaneously assess microbial diversity and functional activity potential of each taxon. *Scientific Reports*, 6, 32165. Available from: https://doi.org/10.1038/srep32165

Santos, A., van Aerle, R., Barrientos, L. & Martinez-Urtaza, J. (2020) Computational methods for 16S metabarcoding studies using nanopore sequencing data. *Computational and Structural Biotechnology Journal*, 18, 296–305. Available from: https://doi.org/10.1016/j.csbj.2020.01.005

Schwarzer, R., Freys, S.D., Neuwinger, N., Beikert, N., Eberspacher, B., Edelmann, A. et al. (2021) A single dose of the Biontech/Pfizer BNT162b2 vaccine protected elderly residents from severe COVID-19 during a SARS-coronavirus-2 outbreak in a senior citizen home in Germany. *Immunity, Inflammation and Disease*, 9(4), 1809–1814. Available from: https://doi.org/10.1002/iid3.532

Silverman, J.D., Bloom, R.J., Jiang, S., Durand, H.K., Dallow, E., Mukherjee, S. et al. (2021) Measuring and mitigating PCR bias in microbiota datasets. *PLoS Computational Biology*, 17(7), e1009113. Available from: https://doi.org/10.1371/journal.pcbi.1009113

Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R. et al. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences of the United States of America*, 103(32), 12115–12120. Available from: https://doi.org/10.1073/pnas.0605127103

Stackebrandt, E. & Goebel, B.M. (1994) Taxonomic note: A place for DNA-DNA reassociation and 16s rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology*, 44, 846–849. Available from: https://doi.org/10.1099/00207713-44-4-846

Stoeck, T., Hayward, B., Taylor, G.T., Varela, R. & Epstein, S.S. (2006) A multiple PCR-primer approach to access the microeukaryotic diversity in environmental samples. *Protist*, 157(1), 31–43. Available from: https://doi.org/10.1016/j.protis.2005.10.004

Tara Ocean Foundation, Tara Oceans, European Molecular Biology Laboratory (EMBL) et al. (2022) Priorities for ocean microbiome research. *Nature Microbiology*, 7, 937–947. Available from: https://doi.org/10.1038/s41564-022-01145-5

Tedersoo, L., Albertsen, M., Anslan, S. & Callahan, B. (2021) Perspectives and benefits of high-throughput long-read sequencing in microbial ecology. *Applied and Environmental Microbiology*, 87(17), e0062621. Available from: https://doi.org/10.1128/AEM.00626-21

Tessler, M., Neumann, J.S., Afshinnekoo, E., Pineda, M., Hersch, R., Velho, L.F.M. et al. (2017) Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. *Scientific Reports*, 7(1), 6589. Available from: https://doi.org/10.1038/s41598-017-06665-3

Thomas, T., Gilbert, J. & Meyer, F. (2012) Metagenomics—a guide from sampling to data analysis. *Microbial Informatics and Experimentation*, 2(1), 3. Available from: https://doi.org/10.1186/2042-5783-2-3

Tremblay, J., Singh, K., Fern, A., Kirton, E.S., He, S., Woyke, T. et al. (2015) Primer and platform effects on 16S rRNA tag sequencing. *Frontiers in Microbiology*, 6, 771. Available from: https://doi.org/10.3389/fmicb.2015.00771

Urich, T., Lanzén, A., Qi, J., Huson, D.H., Schleper, C. & Schuster, S.C. (2008) Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One*, 3(6), e2527. Available from: https://doi.org/10.1371/journal.pone.0002527

van de Peer, Y. (1996) A quantitative map of nucleotide substitution. *Nucleic Acids Research*, 24(17), 3381–3391. Available from: https://doi.org/10.1093/nar/24.17.3381

Verhoeven, J.T.P., Salvo, F., Knight, R., Hamoutene, D. & Dufour, S.C. (2018) Temporal bacterial surveillance of Salmon aquaculture sites indicates a long lasting benthic impact with minimal recovery. *Frontiers in Microbiology*, 9, 3054. Available from: https://doi.org/10.3389/fmicb.2018.03054

Walsh, E.A., Kirkpatrick, J.B., Rutherford, S.D., Smith, D.C., Sogin, M. & D'Hondt, S. (2016) Bacterial diversity and community composition from seasurface to subseafloor. *The ISME Journal*, 10(4), 979–989. Available from: https://doi.org/10.1038/ismej.2015.175

Woese, C.R. & Fox, G.E. (1975) Phylogenetic structure of the prokaryotic domain. *Proceedings of the National Academy of Sciences of the United States of America*, 74(11), 5088–5090.

Yang, B., Wang, Y. & Qian, P.Y. (2016) Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics*, 17, 135. Available from: https://doi.org/10.1186/s12859-016-0992-y

Zaheer, R., Noyes, N., Ortega Polo, R., Cook, S.R., Marinier, E. & van Domselaar, G. (2018) Impact of sequencing depth on the characterization of the microbiome and resistome. *Scientific Reports*, 8(1), 5890. Available from: https://doi.org/10.1038/s41598-018-24280-8

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: Leontidou, K., Abad-Recio, I.L., Rubel, V., Filker, S., Däumer, M., Thielen, A. et al. (2023) Simultaneous analysis of seven 16S rRNA hypervariable gene regions increases efficiency in marine bacterial diversity detection. *Environmental Microbiology*, 25(12), 3484–3501. Available from: https://doi.org/10.1111/1462-2920.16530