



# Selective DNA and Protein Isolation From Marine Macrophyte Surfaces

Marino Korlević<sup>1\*</sup>, Marsej Markovski<sup>1</sup>, Zihao Zhao<sup>2</sup>, Gerhard J. Herndl<sup>2,3,4</sup> and Mirjana Najdek<sup>1</sup>

<sup>1</sup> Center for Marine Research, Ruđer Bošković Institute, Rovinj, Croatia, <sup>2</sup> Department of Functional and Evolutionary Ecology, University of Vienna, Vienna, Austria, <sup>3</sup> Department of Marine Microbiology and Biogeochemistry, Royal Netherlands Institute for Sea Research (NIOZ), Utrecht University, Den Burg, Netherlands, <sup>4</sup> Vienna Metabolomics Center, University of Vienna, Vienna, Austria

Studies of unculturable microbes often combine methods, such as 16S rRNA sequencing, metagenomics, and metaproteomics. To apply these techniques to the microbial community inhabiting the surfaces of marine macrophytes, it is advisable to perform a selective DNA and protein isolation prior to the analysis to avoid biases due to the host material being present in high quantities. Two protocols for DNA and protein isolation were adapted for selective extractions of DNA and proteins from epiphytic communities inhabiting the surfaces of two marine macrophytes, the seagrass *Cymodocea nodosa* and the macroalga *Caulerpa cylindracea*. Protocols showed an almost complete removal of the epiphytic community regardless of the sampling season, station, settlement, or host species. The obtained DNA was suitable for metagenomic and 16S rRNA sequencing, while isolated proteins could be identified by mass spectrometry. Low presence of host DNA and proteins in the samples indicated a high specificity of the protocols. The procedures are based on universally available laboratory chemicals making the protocols widely applicable. Taken together, the adapted protocols ensure an almost complete removal of the macrophyte epiphytic community. The procedures are selective for microbes inhabiting macrophyte surfaces and provide DNA and proteins applicable in 16S rRNA sequencing, metagenomics, and metaproteomics.

## OPEN ACCESS

### Edited by:

Catarina Magalhães,  
University of Porto, Portugal

### Reviewed by:

Frank O'Neill Aylward,  
Virginia Tech, United States  
Olga Lage,  
University of Porto, Portugal

### \*Correspondence:

Marino Korlević  
marino.korlevic@irb.hr

### Specialty section:

This article was submitted to  
Aquatic Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 09 February 2021

**Accepted:** 30 March 2021

**Published:** 24 May 2021

### Citation:

Korlević M, Markovski M, Zhao Z,  
Herndl GJ and Najdek M (2021)  
Selective DNA and Protein Isolation  
From Marine Macrophyte Surfaces.  
*Front. Microbiol.* 12:665999.  
doi: 10.3389/fmicb.2021.665999

**Keywords:** selective isolation, DNA, proteins, marine macrophytes, *Cymodocea nodosa*, *Caulerpa cylindracea*

## INTRODUCTION

Surfaces of marine macrophytes are colonized by a diverse microbial community whose structure and function are poorly understood (Egan et al., 2013). As <1% of all prokaryotic species are culturable, molecular methods, such as 16S rRNA sequencing, metagenomics, and metaproteomics are indispensable to study these organisms (Amann et al., 1995; Su et al., 2012). Applying these techniques requires an initial isolation step with the purpose of obtaining high-quality DNA and proteins.

Biological material (i.e., proteins and DNA) from pelagic microbial communities is usually isolated by collecting cells onto filters and subsequently isolating the target organisms or communities (Gilbert et al., 2009). If a specific microbial size fraction is aimed, sequential filtration is applied (Massana et al., 1997; Andersson et al., 2010). In contrast, obtaining microorganisms associated to surfaces require either a cell detachment procedure prior to isolation or the host

material is co-extracted with the target material. Methods for separating microbial cells from the host include shaking of host tissue (Gross et al., 2003; Nöges et al., 2010), scraping of macrophyte surfaces (Uku et al., 2007), or applying ultrasonication (Weidner et al., 1996; Cai et al., 2014). It was shown that shaking alone is not sufficient to remove microbial cells from surfaces, at least not from plant root surfaces (Richter-Heitmann et al., 2016). Manual separation methods, such as scraping and brushing are time consuming and subjective, as the detachment efficiency depends on host tissue and the person performing the procedure (Cai et al., 2014). Ultrasonication was proposed as an alternative method as it is providing better results in terms of detachment efficiency (Cai et al., 2014; Richter-Heitmann et al., 2016). The downside of this procedure is that complete cell removal is still not obtained and tissue disruption was observed especially after the application of probe ultrasonication (Richter-Heitmann et al., 2016). An alternative to these cell detachment procedures is the isolation of target epiphytic compounds together with host material (Staufenberger et al., 2008; Jiang et al., 2015). This procedure can lead to problems in the following processing steps, such as mitochondrial and chloroplast 16S rRNA sequence contaminations from the host (Longford et al., 2007; Staufenberger et al., 2008). In addition, when performing metagenomics and metaproteomics host material can cause biased results toward more abundant host DNA and proteins.

An alternative to these procedures is a direct isolation of the target material by incubating macrophyte tissues in an extraction buffer. After the incubation, the undisrupted host tissue is removed followed by the isolation procedure, omitting host material contaminations. To our knowledge, the only procedure describing a direct and selective epiphytic DNA isolation from the surfaces of marine macrophytes was described by Burke et al. (2009). In contrast to previously described methods, this protocol enables an almost complete removal of the surface community. It was used for 16S rRNA gene clone library construction (Burke et al., 2011b) and metagenome sequencing (Burke et al., 2011a). This method, although providing a selective isolation procedure, uses a rapid multi-enzyme cleaner (3M) that is not available worldwide and the chemical constituents are unknown (Burke et al., 2009). Also to our knowledge, no selective isolation protocol to perform (meta)proteomics of epiphytic communities associated with marine macrophytes has been developed yet.

In the present study, we adapted a protocol widely used for DNA isolation from filters (Massana et al., 1997) and a protocol used for protein isolation from soils (Chourey et al., 2010; Hultman et al., 2015). These two adapted methods allowed for a selective extraction of DNA and proteins from epiphytic communities inhabiting the surfaces of two marine macrophytes, the seagrass *Cymodocea nodosa*, and the macroalga *Caulerpa cylindracea*. In addition, we tested the removal efficiency of the protocol and the suitability of obtained DNA and proteins for 16S rRNA sequencing, metagenomics, and metaproteomics.

## MATERIALS AND METHODS

### Sampling

Leaves of *C. nodosa* were sampled in a *C. nodosa* meadow in the Bay of Saline, northern Adriatic Sea (45°7'5" N, 13°37'20" E) and in a *C. nodosa* meadow invaded by *C. cylindracea* in the Bay of Funtana, northern Adriatic Sea (45°10'39" N, 13°35'42" E). Thalli of *C. cylindracea* were sampled in the same *C. nodosa* invaded meadow in the Bay of Funtana and at a locality of only *C. cylindracea* located in the proximity of the invaded meadow. Leaves and thalli for 16S rRNA analysis, metagenomics, and metaproteomics were collected in two contrasting seasons, on December 4, 2017 (16S rRNA analysis and metaproteomics), December 14, 2017 (metagenomics), and June 18, 2018 (16S rRNA analysis, metagenomics, and metaproteomics). During spring 2018, the *C. nodosa* meadow in the Bay of Saline decayed to an extent that no leaves could be retrieved (Najdek et al., 2020). In addition, as not enough DNA for both metagenomic and 16S rRNA analysis were obtained during the sampling on December 4, 2017, an additional sampling on December 14, 2017 was carried out in the Bay of Funtana. Leaves and thalli were collected by diving and transported to the laboratory in containers placed on ice and filled with seawater from the site. Upon arrival to the laboratory, *C. nodosa* leaves were cut into sections of 1–2 cm, while *C. cylindracea* thalli were cut into 5–8 cm long sections. Leaves and thalli were washed three times with sterile artificial seawater (ASW) to remove loosely attached microbial cells.

### DNA Isolation

The DNA was isolated according to the protocol for isolation from filters described in Massana et al. (1997). This protocol was modified and adapted for microbial DNA isolation from macrophyte surfaces as described below. Five milliliter of lysis buffer (40 mM EDTA, 50 mM Tris-HCl, 0.75 M sucrose; pH 8.3) was added to 1 g wet weight of leaves or 2 g wet weight of thalli. For every sample, duplicate isolations were performed. Lysozyme was added (final concentration 1 mg ml<sup>-1</sup>) and the mixture was incubated at 37°C for 30 min. Subsequently, proteinase K (final concentration 0.5 mg ml<sup>-1</sup>) and SDS (final concentration 1%) were added and the samples were incubated at 55°C for 2 h. Following the incubation, tubes were vortexed for 10 min and the mixture containing lysed epiphytic cells was separated from host leaves or thalli by transferring the solution into a clean tube. The lysate was extracted twice with a mixture of phenol:chloroform:isoamyl alcohol (25:24:1; pH 8) and once with chloroform:isoamyl alcohol (24:1). After each addition of an organic solvent mixture, tubes were slightly vortexed and centrifuged at 4,500 × g for 10 min. Following each centrifugation, the aqueous phases were retrieved. After the final extraction, 1/10 of chilled 3 M sodium acetate (pH 5.2) was added. DNA was precipitated by adding 1 volume of chilled isopropanol, incubating the mixtures overnight at –20°C and centrifuging at 16,000 × g and 4°C for 20 min. The pellet was washed twice with 1 ml of chilled 70% ethanol and centrifuged after each washing step at 20,000 × g and 4°C for 10 min. After the first washing step, duplicate pellets from the same sample were pooled

and transferred to a clean 1.5 ml tube. The dried pellet was re-suspended in 100  $\mu$ l of deionized water.

## Illumina 16S rRNA Sequencing

An aliquot of isolated DNA was treated with RNase A (final concentration 200  $\mu$ g ml<sup>-1</sup>) for 2 h at 37°C. The DNA concentration was determined using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, USA) according to the manufacturer's instructions and diluted to 1 ng  $\mu$ l<sup>-1</sup>. The V4 region of the 16S rRNA gene was amplified using a two-step PCR procedure. In the first PCR, the 515F (5'-GTGYCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACNVTGGTWTCTAAT-3') primers from the Earth Microbiome Project (<https://earthmicrobiome.org/protocols-and-standards/16s/>) were used to amplify the target region (Caporaso et al., 2012; Apprill et al., 2015; Parada et al., 2016). These primers contained on their 5' end a tagged sequence. Each sample was amplified in four parallel 25  $\mu$ l reactions, in which each contained 1  $\times$  Q5 reaction buffer, 0.2 mM dNTPmix, 0.7 mg ml<sup>-1</sup> BSA (bovine serum albumin), 0.2  $\mu$ M forward and reverse primers, 0.5 U of Q5 High-Fidelity DNA Polymerase (New England Biolabs, USA), and 5 ng of DNA template. Cycling conditions were as follows: initial denaturation at 94°C for 3 min, 20 cycles of denaturation at 94°C for 45 s, annealing at 50°C for 60 s and elongation at 72°C for 90 s, finalized by an elongation step at 72°C for 10 min. The four parallel reactions volumes were pooled and PCR products were purified using the GeneJET PCR Purification Kit (Thermo Fisher Scientific, USA) according to the manufacturer's instructions and following the protocol that included isopropanol addition for better small DNA fragment yield. The column was eluted in 30  $\mu$ l of deionized water. Purified PCR products were sent for Illumina MiSeq sequencing (2  $\times$  250 bp) at IMG M Laboratories, Martinsried, Germany. Before sequencing at IMG M, the second PCR amplification of the two-step PCR procedure was performed using primers targeting the tagged region incorporated in the first PCR. In addition, these primers contained adapter and sample-specific index sequences. The second PCR was carried out for eight cycles. Beside samples, a positive and negative control were sequenced. A negative control was composed of four parallel PCR reactions without DNA template, while for a positive control a mock community composed of evenly mixed DNA material originating from 20 bacterial strains (ATCC MSA-1002, ATCC, USA) was used. Partial 16S rRNA sequences obtained in this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession numbers SAMEA6786270, SAMEA6648792–SAMEA6648794, SAMEA6648809–SAMEA6648811, and SAMEA6648824.

Obtained sequences were analyzed on the computer cluster Isabella (University Computing Center, University of Zagreb) using mothur (version 1.43.0) (Schloss et al., 2009) according to the MiSeq Standard Operating Procedure (MiSeq SOP; [https://mothur.org/wiki/MiSeq\\_SOP](https://mothur.org/wiki/MiSeq_SOP)) (Kozich et al., 2013) and recommendations given from the Riffomonas project to enhance data reproducibility (<http://www.riffomonas.org/>). For alignment and classification of sequences, the SILVA SSU Ref NR 99 database (release 138; <http://www.arb-silva.de>) was used (Quast et al., 2013; Yilmaz et al., 2014). Sequences classified

as chloroplasts by SILVA were exported and reclassified using mothur and the RDP (Ribosomal Database Project; <http://rdp.cme.msu.edu/>) training set (version 16) reference files adapted for mothur (Cole et al., 2014). In comparison to SILVA, RDP allows a more detailed classification of chloroplast sequences. Based on the ATCC MSA-1002 mock community included in the analysis, a sequencing error rate of 0.009% was determined, which is in line with previously reported values for next-generation sequencing data (Kozich et al., 2013; Schloss et al., 2016). In addition, the negative control processed together with the samples yielded only 2 sequences after sequence quality curation.

## Metagenomics

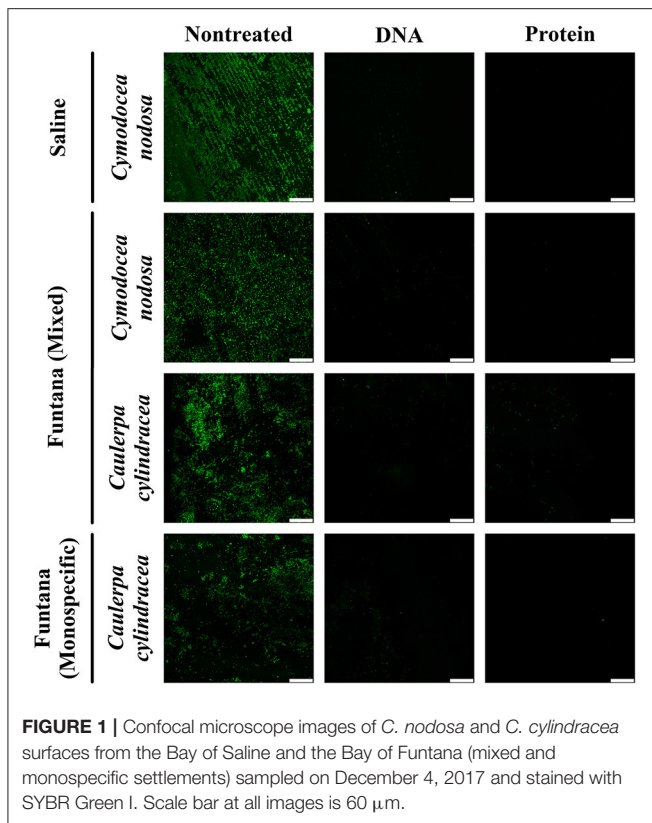
Four DNA samples were selected and sent on dry ice to IMG M Laboratories, Martinsried, Germany for metagenomic sequencing. DNA was purified using AMPure XP Beads (Beckman Coulter, USA) applying a bead:DNA ratio of 1:1 (v/v), quantified with a Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, USA) and checked for integrity on a 1% agarose gel. Metagenomic sequencing libraries were generated from 300 ng of input DNA using a NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs, USA) according to the manufacturer's instructions. Fragments were selected (500–700 bp) using AMPure XP Beads, PCR enriched for 3–5 cycles and quality controlled. Libraries generated from different DNA samples were pooled and sequenced on an Illumina NovaSeq 6000 sequencing system (2  $\times$  250 bp).

Obtained sequences were analyzed on the Life Science Compute Cluster (LiSC) (CUBE–Computational Systems Biology, University of Vienna). Individual sequences were assembled using MEGAHIT (version 1.1.2) (Li et al., 2015) under default settings. Putative genes were predicted from contigs longer than 200 bp using Prodigal (version 2.6.3) (Hyatt et al., 2010) in metagenome mode (-p meta). Abundances of predicted genes were expressed as Reads Per Kilobase Million (RPKM) and calculated using the BWA algorithm (version 0.7.16a) (Li and Durbin, 2009). All predicted genes were functionally annotated using the eggNOG-mapper (Huerta-Cepas et al., 2017) and eggNOG database (version 5.0) (Huerta-Cepas et al., 2019). Sequence taxonomy classification was determined using the lowest common ancestor algorithm adapted from DIAMOND (version 0.8.36) (Buchfink et al., 2015) and by searching against the NCBI non-redundant database (NR). To determine the phylogeny, the top 10% hits with an e-value < 1  $\times$  10<sup>-5</sup> were used (--top 10). Sequence renaming, coverage information computing, and metagenomic statistics calculations were performed using software tools from BBTools (<https://jgi.doe.gov/data-and-tools/bbtools>). Metagenomic sequences obtained in this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession numbers SAMEA6648795, SAMEA6648797, SAMEA6648809, and SAMEA6648811.

## Protein Isolation

Proteins were isolated according to the protocol for protein isolation from soil described in Chourey et al. (2010) and modified by Hultman et al. (2015). This protocol was further

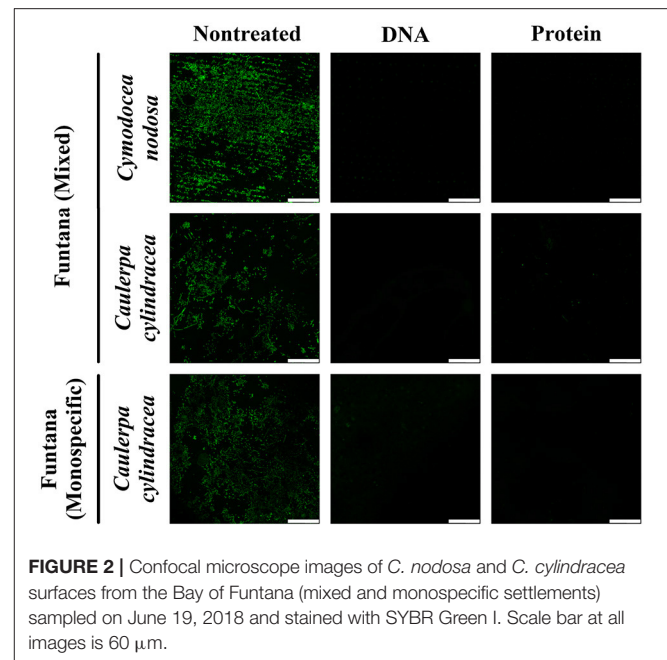




modified and adapted for microbial protein isolation from macrophyte surfaces as described below. Twenty milliliters of protein extraction buffer (4% SDS, 100 mM Tris-HCl [pH 8.0]) was added to 5 g wet weight of leaves or 10 g wet weight of thalli. The mixture was incubated in boiling water for 5 min, vortexed for 10 min, and incubated again in boiling water for 5 min. After a brief vortex, the lysate was transferred to a clean tube separating the host leaves or thalli from the mixture containing lysed epiphytic cells. Dithiothreitol (DTT; final concentration 24 mM) was added and proteins were precipitated with chilled 100% trichloroacetic acid (TCA; final concentration 20%) overnight at  $-20^{\circ}\text{C}$ . Precipitated proteins were centrifuged at  $10,000 \times g$  and  $4^{\circ}\text{C}$  for 40 min. The obtained protein pellet was washed three times with chilled acetone. During the first washing step, the pellet was transferred to a clean 1.5-ml tube. After each washing step, samples were centrifuged at  $20,000 \times g$  and  $4^{\circ}\text{C}$  for 5 min. Dried pellets were stored at  $-80^{\circ}\text{C}$  until further analysis.

## Metaproteomics

Isolated proteins were whole trypsin digested using the FASP (filter-aided sample preparation) Protein Digestion Kit (Expedeon, UK) according to the manufacturer's instructions with small modifications (Wiśniewski et al., 2009). Prior to loading the solution onto the column, protein pellets were solubilized in urea sample buffer included in the kit amended with DTT (final concentration 100 mM) for 45 min at room temperature and centrifuged at  $20,000 \times g$  for 2–5 min at room temperature to remove larger particles. The first washing step



after protein solution loading was repeated twice. In addition, the centrifugation steps were prolonged if the column was clogged. Trypsin digestion was performed on column filters at  $37^{\circ}\text{C}$  overnight for 18 h. The final filtrate containing peptides was acidified with 1% (final concentration) trifluoroacetic acid (TFA), frozen at  $-80^{\circ}\text{C}$ , lyophilized, and sent to VIME–Vienna Metabolomics Center (University of Vienna) for metaproteomic analysis. Peptides were re-suspended in 1% (final concentration) TFA, desalted using the Pierce C18 Tips (Thermo Fisher Scientific, USA) according to the manufacturer's instructions, and sequenced on a Q Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher Scientific). Obtained MS/MS spectra were searched against a protein database composed of combined sequenced metagenomes obtained in this study using SEQUEST-HT engines and validated with Percolator in Proteome Discoverer 2.1 (Thermo Fisher Scientific). The target-decoy approach was used to reduce the probability of false peptide identification. Results whose false discovery rate at the peptide level was  $<1\%$  were kept. For protein identification, a minimum of two peptides and one unique peptide were required. For protein quantification, a chromatographic peak area-based free quantitative method was applied.

## Data Processing and Visualization

Processing and visualization of 16S rRNA, metagenomic, and metaproteomic data were done using R (version 3.6.0) (R Core Team, 2019), package tidyverse (version 1.3.0) (Wickham et al., 2019), and multiple other packages (Neuwirth, 2014; Xie, 2014, 2015, 2019a,b; Wilke, 2018; Xie et al., 2018; Allaire et al., 2019; Zhu, 2019; Bengtsson, 2020). The detailed analysis procedure including the R Markdown file for this paper are available as a GitHub repository ([https://github.com/MicrobesRovinj/Korlevic\\_SelectiveRemoval\\_FrontMicrobiol\\_2021](https://github.com/MicrobesRovinj/Korlevic_SelectiveRemoval_FrontMicrobiol_2021)).

## Confocal Microscopy

Host leaves and thalli from DNA and protein isolation steps were washed seven times in deionized water and fixed with formaldehyde (final concentration ~3%). In addition, non-treated leaves and thalli, washed three times in ASW to remove loosely attached microbial cells, were fixed in the same concentration of formaldehyde and used as a positive control. For long-term storage, fixed leaves and thalli were immersed in a mixture of phosphate-buffered saline (PBS) and ethanol (1:1) and stored at  $-20^{\circ}\text{C}$ . Treated and untreated segments of leaves and thalli were stained in a  $2 \times$  solution of SYBR Green I and examined under a Leica TCS SP8 X FLIM confocal microscope (Leica Microsystems, Germany).

## RESULTS

To assess the removal efficiency of the DNA and protein isolation procedures, leaves and thalli were examined under a confocal microscope before and after treatments were performed. The modified procedures resulted in an almost complete removal of the surface community of both, *C. nodosa* and *C. cylindracea*. In addition, a similar removal efficiency was observed for communities sampled in contrasting months, December 2017 (Figure 1) and June 2018 (Figure 2). Also, no effect of station, settlement, or isolation procedure (DNA or protein) on the removal efficiency was observed (Figures 1, 2).

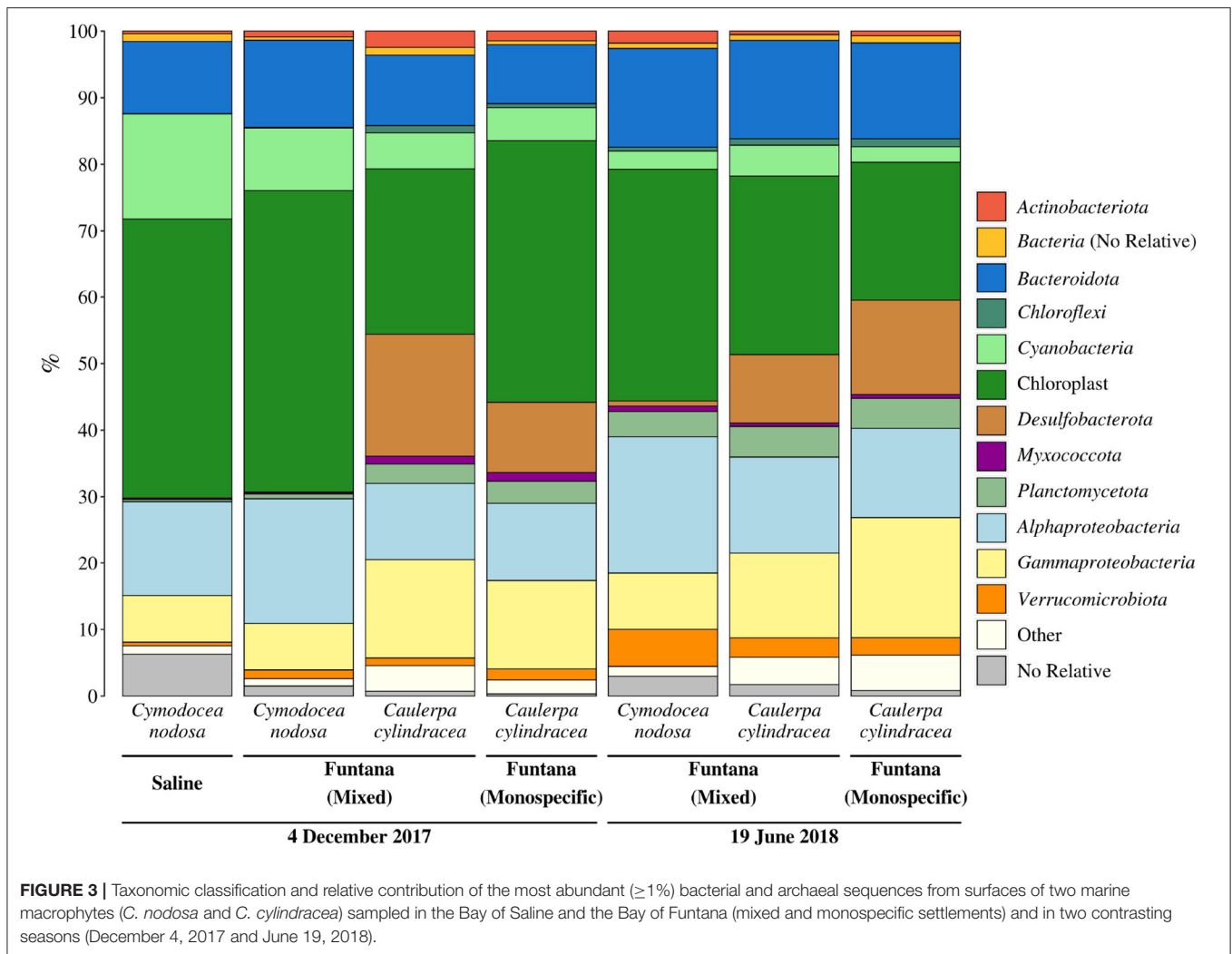
To evaluate whether the obtained DNA is suitable to determine the composition of the microbial community, Illumina sequencing of the V4 region of the 16S rRNA was performed. Sequencing yielded a total of 292,888 sequences after quality curation and exclusion of eukaryotic, mitochondrial, and no relative sequences. The number of sequences classified as chloroplasts was 97,331. After excluding these sequences, the total number of retrieved reads was 195,557, ranging from 13,667 to 41,842 sequences per sample (Supplementary Table 1). Even when the highest sequencing effort was applied, the rarefaction curves did not level off which is commonly observed in high-throughput 16S rRNA amplicon sequencing (Supplementary Figure 1). Sequences clustering at a similarity level of 97% yielded a total of 8,355 different OTUs. Taxonomic classification of reads revealed a macrophyte-associated epiphytic community mainly composed of *Alphaproteobacteria* ( $14.9 \pm 3.5\%$ ), *Bacteroidota* ( $12.5 \pm 2.4\%$ ), *Gammaproteobacteria* ( $11.6 \pm 4.3\%$ ), *Desulfobacterota* ( $7.8 \pm 7.5\%$ ), *Cyanobacteria* ( $6.5 \pm 4.7\%$ ), and *Planctomycetota* ( $2.9 \pm 1.7\%$ ) (Figure 3).

Primers used in this study, as in many other 16S rRNA amplicon procedures, also amplified chloroplast 16S rRNA genes. We observed a high proportion of chloroplast sequences in all analyzed samples ( $33.4 \pm 9.4\%$ ) (Figure 3). To determine whether chloroplast sequences originate from the host or eukaryotic epiphytic organisms, we exported SILVA-classified chloroplast sequences and reclassified them using the RDP training set that allows for a more detailed chloroplast classification. The largest proportion of sequences was classified as Bacillariophyta ( $89.7 \pm 5.7\%$ ) indicating that the DNA

removal procedure resulted in only minor co-extracted quantities of host DNA (Figure 4). Chloroplast sequences classified as Streptophyta constituted  $3.3 \pm 2.8\%$  of all chloroplast sequences originating from *C. nodosa* samples, while sequences classified as Chlorophyta comprised only  $0.02 \pm 0.01\%$  of all chloroplast sequences associated with *C. cylindracea* samples.

To determine whether the extracted DNA can be used for metagenomic sequencing, four samples containing epiphytic DNA were selected and shotgun sequenced using an Illumina platform. Metagenomic sequencing yielded between 207,149,524 and 624,029,930 sequence pairs (Supplementary Table 2). Obtained sequences were successfully assembled into contigs whose L50 ranged from 654 to 1,011 bp. In addition, predicted coding sequences were functionally annotated (9,066,667–20,256,215 annotated sequences; Figure 5A) and taxonomically classified. Functional annotation allowed for an assessment of the relative contribution of each COG (Clusters of Orthologous Groups) functional category to the total number of annotated coding sequences (Figure 5A). Functional categories containing the highest number of sequences were C (energy production and conversion), E (amino acid transport and metabolism), M (cell wall/membrane/envelope biogenesis), L (replication, recombination and repair), and P (inorganic ion transport and metabolism). If host DNA is co-extracted with epiphytes, it should be detected in large proportions in sequenced metagenomes. However, no large proportions of coding sequences classified as Streptophyta and Chlorophyta were detected (Supplementary Table 3). Sequenced metagenomic DNA originating from the surface of *C. nodosa* contained 1.3% of coding sequences classified as Streptophyta in December 2017 and 0.7% in June 2018. Furthermore, the summed RPKM (reads per kilobase million) of these sequences constituted 1.7% of total RPKM of all successfully classified sequences in December 2017 and 1.1% in June 2018. Similar low proportions of host coding sequences were detected in metagenomic samples originating from the surfaces of *C. cylindracea*. Of all successfully classified coding sequences, 0.2% sequences were classified as Chlorophyta in December 2017 and 0.1% in June 2018. A relatively higher proportion of RPKM of these sequences than in the case of *C. nodosa* was observed, indicating a higher co-extraction of host DNA in *C. cylindracea*. In December, the proportion of RPKM of sequences classified as Chlorophyta was 8.2%, while in June 2018 it reached 13.6%.

To evaluate whether the procedure for protein extraction is suitable for metaproteomic analysis, obtained proteins were trypsin digested and sequenced using a mass spectrometer. Obtained MS/MS spectra were searched against a protein database from sequenced metagenomes. From 14,219 to 16,449 proteins were identified in isolated protein samples (Figure 5B). In addition, successful identification of proteins allowed for an assessment of the relative contribution of each COG functional category to the total number of identified proteins (Figure 5B). Functional categories containing the highest number of identified proteins were C (energy production and conversion), G (carbohydrate transport and metabolism), P (inorganic ion

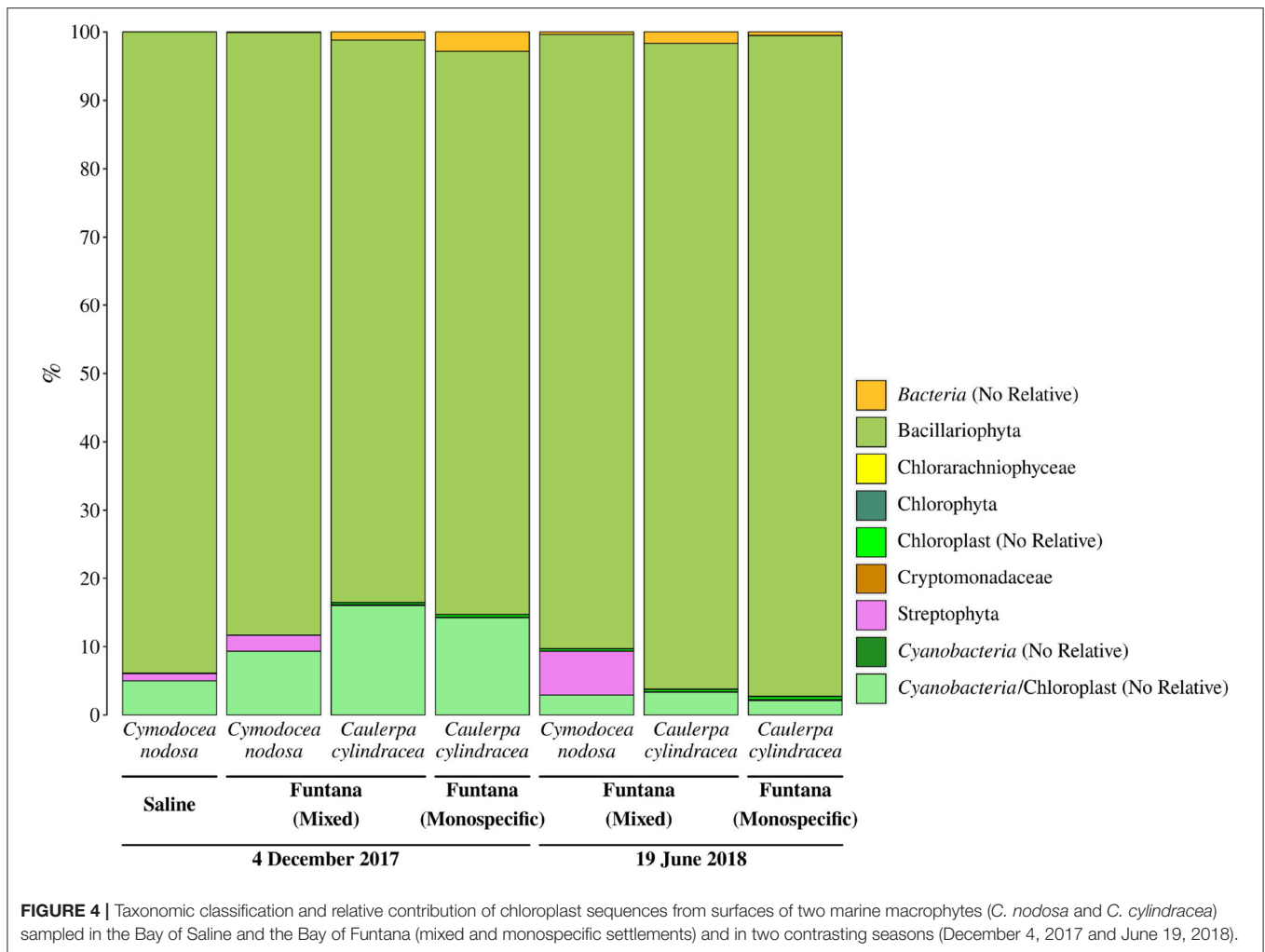


**FIGURE 3 |** Taxonomic classification and relative contribution of the most abundant ( $\geq 1\%$ ) bacterial and archaeal sequences from surfaces of two marine macrophytes (*C. nodosa* and *C. cylindracea*) sampled in the Bay of Saline and the Bay of Funtana (mixed and monospecific settlements) and in two contrasting seasons (December 4, 2017 and June 19, 2018).

transport and metabolism), O (post-translational modification, protein turnover, chaperones), and E (amino acid transport and metabolism). Isolated proteins could originate from epiphytic organisms inhabiting the macrophyte surface and/or from macrophyte tissue underlying them. The contribution of proteins originating from host tissues was evaluated by identifying all the proteins predicted to belong to any taxonomic group within the phyla Streptophyta and Chlorophyta and by calculating their contribution to the number and abundance (NAAF [normalized abundance area factor]) of all identified proteins. On average, proteins isolated from the surface of *C. nodosa* contained  $1.8 \pm 0.06\%$  of proteins associated with Streptophyta, contributing to  $2.2 \pm 0.8\%$  of total proteins. Similar to metagenomes, proteins associated with Chlorophyta contributed more to *C. cylindracea* than proteins associated with Streptophyta to *C. nodosa*. Chlorophyta-associated proteins composed  $5.2 \pm 0.06\%$  of all identified proteins in *C. cylindracea*, contributing  $19.2 \pm 1.5\%$  to the total protein abundance.

## DISCUSSION

To test whether the developed DNA and protein isolation protocols efficiently detach microbes from the macrophyte surface, we selected *C. nodosa* and *C. cylindracea* as representatives of seagrass and macroalgal species. These species differ morphologically. While *C. nodosa* leaves are flat, *C. cylindracea* thalli are characterized by an uneven surface (Kuo and den Hartog, 2001; Verlaque et al., 2003). The developed protocol led to an almost complete removal of epiphytic cells from the surfaces of both species comparable to the result of Burke et al. (2009), indicating that structural differences do not impact the removal efficiency. In addition, isolation protocols were tested in two contrasting seasons, as it is known that macrophytes are harboring more algal epiphytes during autumn and winter (Reyes and Sansón, 2001). No differences in the removal efficiency was observed between seasons, suggesting that these protocols can be used on macrophyte samples retrieved throughout the year. Also, no removal differences



**FIGURE 4 |** Taxonomic classification and relative contribution of chloroplast sequences from surfaces of two marine macrophytes (*C. nodosa* and *C. cylindracea*) sampled in the Bay of Saline and the Bay of Funtana (mixed and monospecific settlements) and in two contrasting seasons (December 4, 2017 and June 19, 2018).

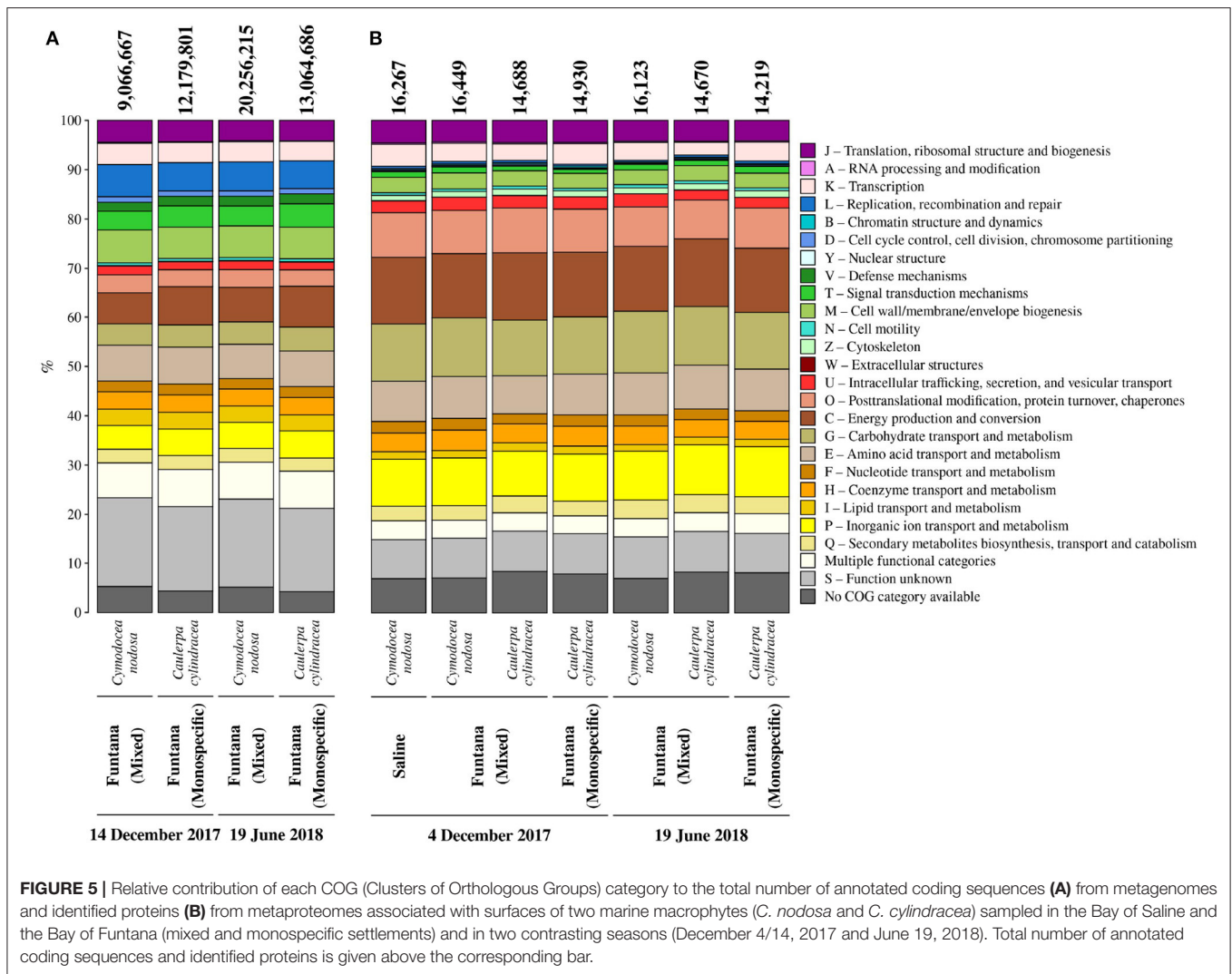
were observed on samples derived from the same host but from different locations.

Successful amplification and sequencing of the V4 region of the 16S rRNA gene proved that the isolated DNA can be used to estimate the microbial epiphytic diversity. Taxonomic groups detected in this step can also be often found in epiphytic communities associated with other macrophytes (Burke et al., 2011b; Morrissey et al., 2019). A problem often encountered in studies focusing on epiphytic communities is the presence of large proportions of chloroplast 16S rRNA sequences in the pool of amplified molecules, especially if the epiphytic DNA was isolated without prior selection (Staufenberger et al., 2008). These sequences can derive from host chloroplasts or from eukaryotic epiphytic chloroplast DNA. Although the proportion of obtained chloroplast 16S rRNA sequences in our samples was substantial, they derived almost exclusively from eukaryotic epiphytes. High proportion of chloroplast 16S rRNA sequences in studies applying selective procedures that include direct cellular lysis on host surfaces were observed before (Michelou et al., 2013). It is possible that chloroplast-specific sequences even in these studies originated from eukaryotic epiphytic cells and

not from host chloroplasts. Indeed, it is common during 16S rRNA profiling of pelagic microbial communities to observe high proportions of chloroplast sequences (Gilbert et al., 2009; Korlević et al., 2016). One of the solutions to further reduce chloroplast 16S rRNA sequence contamination is to use primers that minimize the amplification of these reads if the sequencing and study design allow it (Hanshew et al., 2013). In addition, a very low proportion of chloroplast 16S rRNA sequences in samples originating from *C. cylindracea* in comparison to *C. nodosa* could be explained by the presence of three introns in the gene for 16S rRNA in some members of the genus *Caulerpa* that could hamper the amplification process (Lam and Lopez-Bautista, 2016).

High-quality DNA is also needed for metagenomics. The obtained number of metagenomic sequences and assembly statistics were comparable to metagenomes and metatranscriptomes derived from similar surface-associated communities (Crump et al., 2018; Cúcio et al., 2018). In addition, functional annotation of predicted coding sequences to COG functional categories showed that the obtained metagenomes can be used to determine the metabolic capacity





of surface-associated communities (Leary et al., 2014; Cúcio et al., 2018). The proportion of coding sequences, including their RPKM, originating from *C. nodosa* metagenomes and classified as Streptophyta was low indicating that the isolation procedure was specific for epiphytic cells. DNA samples isolated from the surface of *C. cylindracea* exhibited a low proportion of Chlorophyta coding sequences; however, their RPKM was higher than in the samples originating from *C. nodosa*. One of the reasons for this elevated RPKM of Chlorophyta sequences in *C. cylindracea* could be the differences in the tissue structure between these two host species. While *C. nodosa* leaves are composed of individual cells, the thallus of *C. cylindracea* is, like in other siphonous algal species, composed of a single large multinucleate cell (Coneva and Chitwood, 2015). The absence of individual cells in *C. cylindracea* could cause a leakage of genetic material into the extraction buffer causing an elevated presence of host sequences in the samples for metagenome analyses.

To obtain insight into the metabolic status of uncultivated prokaryotes, a metaproteomic approach is required (Saito et al.,

2019). The applied protocol for epiphytic protein isolation followed by a metaproteomic analysis identified between 14,219 and 16,449 proteins, which is higher than previously reported for soils (Chourey et al., 2010; Hultman et al., 2015), seawater (Williams et al., 2012), and biofilms (Leary et al., 2014). The functional annotation of identified proteins into COG functional categories showed that the protein isolation protocol can be used to assess the metabolic status of the epiphytic community (Leary et al., 2014). Similar to the results of the metagenomic analysis, the number and abundance of identified proteins affiliated to Streptophyta in *C. nodosa* samples were low, indicating that the procedure is selective for epiphytic cell proteins. In addition, a higher number and abundance of identified proteins associated with Chlorophyta were observed in *C. cylindracea* samples. The cause of this elevated presence of Chlorophyta-associated proteins can be, similar to the DNA isolation protocol, explained by the absence of individual cells in this siphonous alga (Coneva and Chitwood, 2015).



In conclusion, the developed protocols for DNA and protein isolation from macrophyte surfaces almost completely remove the epiphytic community from both, *C. nodosa* and *C. cylindracea*, in different seasons. Also, the obtained DNA and proteins are suitable for 16S rRNA sequencing, metagenomics and metaproteomics analyses while the obtained material contains low quantities of host DNA and proteins making the protocols specific for epiphytes. Furthermore, the protocols are based on universally available laboratory chemicals hence, making them widely applicable.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repositories and accession numbers can be found in the article.

## AUTHOR CONTRIBUTIONS

MK designed the work with the intellectual contribution from GH and MN, and prepared the manuscript with editorial help from MM, ZZ, GH, and MN. MK, MM, ZZ, and MN performed the sampling and laboratory analysis. MK and ZZ analyzed the data. All authors contributed to the article and approved the final submitted version.

## REFERENCES

- Allaire, J. J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., et al. (2019). *markdown: Dynamic documents for R*.
- Amann, R. L., Ludwig, W., and Schleifer, K. H. (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59, 143–169.
- Andersson, A. F., Riemann, L., and Bertilsson, S. (2010). Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *ISME J.* 4, 171–181. doi: 10.1038/ismej.2009.108
- Apprill, A., McNally, S., Parsons, R., and Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.* 75, 129–137. doi: 10.3354/ame01753
- Bengtsson, H. (2020). *matrixStats: Functions That Apply to Rows and Columns of Matrices (and to Vectors)*.
- Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176
- Burke, C., Kjelleberg, S., and Thomas, T. (2009). Selective extraction of bacterial DNA from the surfaces of macroalgae. *Appl. Environ. Microbiol.* 75, 252–256. doi: 10.1128/AEM.01630-08
- Burke, C., Steinberg, P., Rusch, D., Kjelleberg, S., and Thomas, T. (2011a). Bacterial community assembly based on functional genes rather than species. *Proc. Natl. Acad. Sci. U.S.A.* 108, 14288–14293. doi: 10.1073/pnas.1101591108
- Burke, C., Thomas, T., Lewis, M., Steinberg, P., and Kjelleberg, S. (2011b). Composition, uniqueness and variability of the epiphytic bacterial community of the green alga *Ulva australis*. *ISME J.* 5, 590–600. doi: 10.1038/ismej.2010.164
- Cai, X., Gao, G., Yang, J., Tang, X., Dai, J., Chen, D., et al. (2014). An ultrasonic method for separation of epiphytic microbes from freshwater submerged macrophytes. *J. Basic Microbiol.* 54, 758–761. doi: 10.1002/jobm.201300041
- Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community

## FUNDING

This work was funded by the Croatian Science Foundation through the MICRO-SEAGRASS project (project number IP-2016-06-7118). ZZ and GH were supported by the Austrian Science Fund (FWF) project ARTEMIS (project number P28781-B21).

## ACKNOWLEDGMENTS

We would like to thank Margareta Buterer for technical support, Paolo Paliaga for help during sampling, Lucija Horvat for technical support in confocal microscopy, Dušica Vujaklija for help in accessing the confocal microscope, and Sonja Tischler for technical support in mass spectrometry. In addition, we would like to thank the University Computing Center of the University of Zagreb for access to the computer cluster Isabella and CUBE–Computational Systems Biology of the University of Vienna for access to the Life Science Compute Cluster (LiSC).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.665999/full#supplementary-material>

- analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624. doi: 10.1038/ismej.2012.8
- Chourey, K., Jansson, J., VerBerkmoes, N., Shah, M., Chavarria, K. L., Tom, L. M., et al. (2010). Direct cellular lysis/protein extraction protocol for soil metaproteomics. *J. Proteome Res.* 9, 6615–6622. doi: 10.1021/pr100787q
- Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., et al. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 42, D633–D642. doi: 10.1093/nar/gkt1244
- Coneva, V., and Chitwood, D. H. (2015). Plant architecture without multicellularity: Quandaries over patterning and the soma-germline divide in siphonous algae. *Front. Plant Sci.* 6:287. doi: 10.3389/fpls.2015.00287
- Crump, B. C., Wojahn, J. M., Tomas, F., and Mueller, R. S. (2018). Metatranscriptomics and amplicon sequencing reveal mutualisms in seagrass microbiomes. *Front. Microbiol.* 9:388. doi: 10.3389/fmicb.2018.00388
- Cúcio, C., Overmars, L., Engelen, A. H., and Muyzer, G. (2018). Metagenomic analysis shows the presence of bacteria related to free-living forms of sulfur-oxidizing chemolithoautotrophic symbionts in the rhizosphere of the seagrass *Zostera marina*. *Front. Mar. Sci.* 5:171. doi: 10.3389/fmars.2018.00171
- Egan, S., Harder, T., Burke, C., Steinberg, P., Kjelleberg, S., and Thomas, T. (2013). The seaweed holobiont: Understanding seaweed-bacteria interactions. *FEMS Microbiol. Rev.* 37, 462–476. doi: 10.1111/1574-6976.12011
- Gilbert, J. A., Field, D., Swift, P., Newbold, L., Oliver, A., Smyth, T., et al. (2009). The seasonal structure of microbial communities in the Western English Channel. *Environ. Microbiol.* 11, 3132–3139. doi: 10.1111/j.1462-2920.2009.02017.x
- Gross, E. M., Feldbaum, C., and Graf, A. (2003). Epiphyte biomass and elemental composition on submersed macrophytes in shallow eutrophic lakes. *Hydrobiologia* 506–509, 559–565. doi: 10.1023/B:HYDR.0000008538.68268.82
- Hanshaw, A. S., Mason, C. J., Raffa, K. F., and Currie, C. R. (2013). Minimization of chloroplast contamination in 16S rRNA gene pyrosequencing of insect herbivore bacterial communities. *J. Microbiol. Methods* 95, 149–155. doi: 10.1016/j.mimet.2013.08.007

- Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* 34, 2115–2122. doi: 10.1093/molbev/msx148
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085
- Hultman, J., Waldrop, M. P., Mackelprang, R., David, M. M., McFarland, J., Blazewicz, S. J., et al. (2015). Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* 521, 208–212. doi: 10.1038/nature14238
- Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Jiang, Y. F., Ling, J., Dong, J. D., Chen, B., Zhang, Y. Y., Zhang, Y. Z., et al. (2015). Illumina-based analysis of the microbial diversity associated with *Thalassia hemprichii* in Xincun Bay, South China Sea. *Ecotoxicology* 24, 1548–1556. doi: 10.1007/s10646-015-1511-z
- Korlević, M., Šupraha, L., Ljubešić, Z., Henderiks, J., Ciglenečki, I., Dautović, J., et al. (2016). Bacterial diversity across a highly stratified ecosystem: a salt-wedge Mediterranean estuary. *Syst. Appl. Microbiol.* 39, 398–408. doi: 10.1016/j.syapm.2016.06.006
- Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* 79, 5112–5120. doi: 10.1128/AEM.01043-13
- Kuo, J., and den Hartog, C. (2001). “Seagrass taxonomy and identification key,” in *Global Seagrass Research Methods*, eds F. T. Short and R. G. Coles (Amsterdam: Elsevier Science B.V.), 31–58.
- Lam, D. W., and Lopez-Bautista, J. M. (2016). Complete chloroplast genome for *Caulerpa racemosa* (Bryopsidales, Chlorophyta) and comparative analyses of siphonous green seaweed plastomes. *Cymbella* 2, 23–32.
- Leary, D. H., Li, R. W., Hamdan, L. J., Judson Hervey, W. IV, Lebedev, N., Wang, Z., et al. (2014). Integrated metagenomic and metaproteomic analyses of marine biofilm communities. *Biofouling* 30, 1211–1223. doi: 10.1080/08927014.2014.977267
- Li, D., Liu, C. M., Luo, R., Sadakane, K., and Lam, T. W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Longford, S., Tujula, N., Crocetti, G., Holmes, A., Holmström, C., Kjelleberg, S., et al. (2007). Comparisons of diversity of bacterial communities associated with three sessile marine eukaryotes. *Aquat. Microb. Ecol.* 48, 217–229. doi: 10.3354/ame048217
- Massana, R., Murray, A. E., Preston, C. M., and DeLong, E. F. (1997). Vertical distribution and phylogenetic characterization of marine planktonic *Archaea* in the Santa Barbara Channel. *Appl. Environ. Microbiol.* 63, 50–56.
- Michelou, V. K., Caporaso, J. G., Knight, R., and Palumbi, S. R. (2013). The ecology of microbial communities associated with *Macrocystis pyrifera*. *PLoS ONE* 8:e67480. doi: 10.1371/journal.pone.0067480
- Morrissey, K. L., Çavas, L., Willems, A., and De Clerck, O. (2019). Disentangling the influence of environment, host specificity and thallus differentiation on bacterial communities in siphonous green seaweeds. *Front. Microbiol.* 10:717. doi: 10.3389/fmicb.2019.00717
- Najdek, M., Korlević, M., Paliaga, P., Markovski, M., Ivančić, I., Iveša, L., et al. (2020). Dynamics of environmental conditions during the decline of a *Cymodocea nodosa* meadow. *Biogeosciences* 17, 3299–3315. doi: 10.5194/bg-17-3299-2020
- Neuwirth, E. (2014). *RColorBrewer: ColorBrewer Palettes*.
- Nöges, T., Luup, H., and Feldmann, T. (2010). Primary production of aquatic macrophytes and their epiphytes in two shallow lakes (Peipsi and Võrtsjärv) in Estonia. *Aquat. Ecol.* 44, 83–92. doi: 10.1007/s10452-009-9249-4
- Parada, A. E., Needham, D. M., and Fuhrman, J. A. (2016). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* 18, 1403–1414. doi: 10.1111/1462-2920.13023
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Reyes, J., and Sansón, M. (2001). Biomass and production of the epiphytes on the leaves of *Cymodocea nodosa* in the Canary Islands. *Bot. Mar.* 44, 307–313. doi: 10.1515/BOT.2001.039
- Richter-Heitmann, T., Eickhorst, T., Knauth, S., Friedrich, M. W., and Schmidt, H. (2016). Evaluation of strategies to separate root-associated microbial communities: a crucial choice in rhizobiome research. *Front. Microbiol.* 7:773. doi: 10.3389/fmicb.2016.00773
- Saito, M. A., Bertrand, E. M., Duffy, M. E., Gaylord, D. A., Held, N. A., Hervey, W. J., et al. (2019). Progress and challenges in ocean metaproteomics and proposed best practices for data sharing. *J. Proteome Res.* 18, 1461–1476. doi: 10.1021/acs.jproteome.8b00761
- Schloss, P. D., Jenior, M. L., Koumpouras, C. C., Westcott, S. L., and Highlander, S. K. (2016). Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. *PeerJ* 4:e1869. doi: 10.7717/peerj.1869
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09
- Staufenberger, T., Thiel, V., Wiese, J., and Imhoff, J. F. (2008). Phylogenetic analysis of bacteria associated with *Laminaria saccharina*. *FEMS Microbiol. Ecol.* 64, 65–77. doi: 10.1111/j.1574-6941.2008.00445.x
- Su, C., Lei, L., Duan, Y., Zhang, K.-Q., and Yang, J. (2012). Culture-independent methods for studying environmental microorganisms: Methods, application, and perspective. *Appl. Microbiol. Biotechnol.* 93, 993–1003. doi: 10.1007/s00253-011-3800-7
- Uku, J., Björk, M., Bergman, B., and Díez, B. (2007). Characterization and comparison of prokaryotic epiphytes associated with three East African seagrasses. *J. Phycol.* 43, 768–779. doi: 10.1111/j.1529-8817.2007.00371.x
- Verlaque, M., Durand, C., Huisman, J. M., Boudouresque, C. F., and Le Parco, Y. (2003). On the identity and origin of the Mediterranean invasive *Caulerpa racemosa* (Caulerpaceae, Chlorophyta). *Eur. J. Phycol.* 38, 325–339. doi: 10.1080/09670260310001612592
- Weidner, S., Arnold, W., and Puhler, A. (1996). Diversity of uncultured microorganisms associated with the seagrass *Halophila stipulacea* estimated by restriction fragment length polymorphism analysis of PCR-amplified 16S rRNA genes. *Appl. Environ. Microbiol.* 62, 766–771.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., et al. (2019). Welcome to the tidyverse. *J. Open Source Softw.* 4:1686. doi: 10.21105/joss.01686
- Wilke, C. O. (2018). *cowplot: Streamlined Plot Theme and Plot Annotations for ggplot2*.
- Williams, T. J., Long, E., Evans, F., DeMaere, M. Z., Lauro, F. M., Raftery, M. J., et al. (2012). A metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters. *ISME J.* 6, 1883–1900. doi: 10.1038/ismej.2012.28
- Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362. doi: 10.1038/nmeth.1322
- Xie, Y. (2014). “knitr: A comprehensive tool for reproducible research in R,” in *Implementing Reproducible Computational Research*, eds V. Stodden, F. Leisch, and R. D. Peng (New York, NY: Chapman and Hall; CRC), 3–32.
- Xie, Y. (2015). *Dynamic Documents With R and knitr, 2nd Edn*. Boca Raton, FL: Chapman and Hall; CRC.
- Xie, Y. (2019a). TinyTeX: a lightweight, cross-platform, and easy-to-maintain LaTeX distribution based on TeX Live. *TUGboat* 40, 30–32.
- Xie, Y. (2019b). *tinytex: Helper Functions to Install and Maintain ‘TeX Live’, and Compile ‘LaTeX’ Documents*.
- Xie, Y., Allaire, J. J., and Grolmund, G. (2018). *R Markdown: The Definitive Guide, 1st Edn*. Boca Raton, FL: Chapman and Hall; CRC.

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., et al. (2014). The SILVA and “All-Species Living Tree Project (LTP)” taxonomic frameworks. *Nucleic Acids Res.* 42, D643–D648. doi: 10.1093/nar/gkt1209

Zhu, H. (2019). *kableExtra: Construct Complex Table With 'kable' and Pipe Syntax*.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Korlević, Markovski, Zhao, Herndl and Najdek. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.