

Selective DNA and Protein Isolation from Marine Macrophyte Surfaces

Marino Korlević^{1*}, Marsej Markovski¹, Zihao Zhao², Gerhard J. Herndl^{2,3,4}, Mirjana Najdek¹

1. Center for Marine Research, Ruđer Bošković Institute, Croatia

2. Department of Functional and Evolutionary Ecology, University of Vienna, Austria

3. NIOZ, Department of Marine Microbiology and Biogeochemistry, Royal Netherlands Institute for Sea Research, Utrecht University, The Netherlands

4. Vienna Metabolomics Center, University of Vienna, Austria

*To whom correspondence should be addressed:

Marino Korlević

G. Paliaga 5, 52210 Rovinj, Croatia

Tel.: +385 52 804 768

Fax: +385 52 804 780

e-mail: marino.korlevic@irb.hr

Running title: DNA and protein isolation from macrophyte surfaces

Summary

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.

Introduction

Surfaces of marine macrophytes are colonized by a diverse microbial community whose structure and function are poorly understood (Egan *et al.*, 2013). As less than 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; Nõges *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 towards 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 and Thomas *et al.*, 2011) and metagenome sequencing (Burke and Peter Steinberg *et al.*, 2011). 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.

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 (Fig. 1) and June 2018 (Fig. 2). Also, no effect of station, settlement or isolation procedure (DNA or protein) on the removal efficiency was observed (Figs. 1 and 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 336,944 sequences after quality curation and exclusion of eukaryotic, mitochondrial and no relative sequences. The number of sequences classified as chloroplasts was 97,334. After excluding these sequences the total number of retrieved reads was 239,610, ranging from 22,596 to 52,930 sequences per sample (Table S1). 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 (Fig. S1). 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 that 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* ($2.9 \pm 1.7\%$) (Fig. 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 ($33.4 \pm 9.4\%$) (Fig. 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 (Ribosomal Database Project) 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 (Fig. 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 (Table S2). 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; Fig. 5) 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 (Fig. 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 (Table S3). 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 % 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 coextraction 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 (Fig. 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 (Fig. 5b). Functional categories containing the highest number of identified proteins were C (Energy production and conversion), G (Carbohydrate transport and metabolism), P (Inorganic ion transport and metabolism), O (Posttranslational 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 ± 0.06 % of proteins associated with Streptophyta, contributing to 2.2 ± 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 comprised 5.2 ± 0.06 % of all identified proteins in *C. cylindracea*, contributing 19.2 ± 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 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 and Thomas *et al.*, 2011; 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). 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 e.g. 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.

Experimental procedures

Sampling

Leaves of *C. nodosa* were sampled in a *C. nodosa* meadow in the Bay of Saline, northern Adriatic Sea ($45^{\circ}7'5''$ N, $13^{\circ}37'20''$ E) and in a *C. nodosa* meadow invaded by *C. cylindracea* in the Bay of Funtana, northern Adriatic Sea ($45^{\circ}10'39''$ N, $13^{\circ}35'42''$ 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 4 December 2017 (16S rRNA analysis and metaproteomics), 14 December 2017 (metagenomics) and 18 June 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 RNA analysis were obtained during the sampling on 4 December 2017, an additional sampling on 14 December 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 this 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 (5) ml 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⁻¹) and the mixture was incubated at 37 °C for 30 min. Subsequently, proteinase K (final concentration 0.5 mg ml⁻¹) 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 resuspended in 100 µl of deionized water.

Illumina 16S rRNA sequencing

An aliquot of isolated DNA was treated with RNase A (final concentration 200 µg ml⁻¹) 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 µl⁻¹. 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'-GGACTACNVGGGTWTCTAAT-3') primers from the Earth Microbiome Project (<http://press.igsb.anl.gov/earthmicrobiome/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 µl

reactions of which each contained $1 \times$ Q5 Reaction Buffer, 0.2 mM of dNTPmix, 0.7 mg ml⁻¹ BSA (Bovine Serum Albumin), 0.2 μ M of 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: 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 μ l of deionized water. Purified PCR products were sent for Illumina MiSeq sequencing (2×250 bp) at IMGM Laboratories, Martinsried, Germany. Before sequencing at IMGM, 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 8 cycles. Beside samples, a positive and negative control were sequenced. A negative control was comprised 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.

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) (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 (<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 IMGM 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 check 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 × 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 contings 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 $< 1 \times 10^{-5}$ 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 (20) ml 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°C . Precipitated proteins were centrifuged at $10,000 \times g$ and 4°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°C for 5 min. Dried pellets were stored at -80°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 a 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 × 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 °C overnight for 18 h. The final filtrate containing peptides was acidified with 1 % (final concentration) trifluoroacetic acid (TFA), freezed at –80 °C, lyophilized and sent to VIME – Vienna Metabolomics Center (University of Vienna) for metaproteomic analysis. Peptides were resuspended 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, USA). Obtained MS/MS spectra were searched against a protein database composed of combined sequenced metagenomes using SEQUEST-HT engines and validated with Percolator in Proteome Discoverer 2.1 (Thermo Fisher Scientific, USA). 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 (Xie, 2014, 2015, 2019, 2020; Neuwirth, 2014; Xie *et al.*, 2018; Wilke, 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_EnvironMicrobiol_2020).

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, nontreated 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 °C. Treated and untreated segments of leaves and thalli were stained in a 2 × solution of SYBR Green I and examined under a Leica TCS SP8 X FLIM confocal microscope (Leica Microsystems, Germany).

Acknowledgements

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

References

- Allaire, J.J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., et al. (2019) rmarkdown: Dynamic documents for R.
- Amann, R.I., 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.
- 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.
- 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.
- Burke, C., Kjelleberg, S., and Thomas, T. (2009) Selective extraction of bacterial DNA from the surfaces of macroalgae. *Appl Environ Microbiol* **75**: 252–256.
- Burke, C., Steinberg, P., Rusch, D., Kjelleberg, S., and Thomas, T. (2011) Bacterial community assembly based on functional genes rather than species. *Proc Natl Acad Sci U S A* **108**: 14288–14293.
- Burke, C., Thomas, T., Lewis, M., Steinberg, P., and Kjelleberg, S. (2011) Composition, uniqueness and variability of the epiphytic bacterial community of the green alga *Ulva australis*. *ISME J* **5**: 590–600.

Cai, X., Gao, G., Yang, J., Tang, X., Dai, J., Chen, D., and Song, Y. (2014) An ultrasonic method for separation of epiphytic microbes from freshwater submerged macrophytes. *J Basic Microbiol* **54**: 758–761.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**: 1621–1624.

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.

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.

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.

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.

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.

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.

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.

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.

Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., and Bork, P. (2017) Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol* **34**: 2115–2122.

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.

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.

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.

Jiang, Y.-F., Ling, J., Dong, J.-D., Chen, B., Zhang, Y.-Y., Zhang, Y.-Z., and Wang, Y.-S. (2015) Illumina-based analysis the microbial diversity associated with *Thalassia hemprichii* in Xincun Bay, South China Sea. *Ecotoxicology* **24**: 1548–56.

Korlević, M., Šupraha, L., Ljubešić, Z., Henderiks, J., Ciglenečki, I., Dautović, J., and Orlić, S. (2016) Bacterial diversity across a highly stratified ecosystem: A salt-wedge Mediterranean estuary. *Syst Appl Microbiol* **39**: 398–408.

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.

Kuo, J. and den Hartog, C. (2001) Seagrass taxonomy and identification key. In *Global Seagrass Research Methods*. Short, F.T. and Coles, R.G. (eds). Amsterdam: Elsevier Science B.V., pp. 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., IV, W.J.H., Lebedev, N., Wang, Z., et al. (2014) Integrated metagenomic and metaproteomic analyses of marine biofilm communities. *Biofouling* **30**: 1211–1223.

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.

Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with BurrowsWheeler transform. *Bioinformatics* **25**: 1754–1760.

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.

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.

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.

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.

Neuwirth, E. (2014) RColorBrewer: ColorBrewer palettes.

Nõges, 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.

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.

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.

R Core Team (2019) R: A language and environment for statistical computing, Vienna, Austria: 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.

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.

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.

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.

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.

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.

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.

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.

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* (Caulerpales, Chlorophyta). *Eur J Phycol* **38**: 325–339.

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.

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.

Wiśniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat Methods* **6**: 359–362.

Xie, Y. (2015) Dynamic Documents with R and knitr, 2nd ed. Boca Raton, Florida: Chapman and Hall/CRC.

Xie, Y. (2014) knitr: A comprehensive tool for reproducible research in R. In *Implementing Reproducible Computational Research*. Stodden, V., Leisch, F., and Peng, R.D. (eds). New York: Chapman and Hall/CRC, pp. 3–32.

Xie, Y. (2019) TinyTeX: A lightweight, cross-platform, and easy-to-maintain LaTeX distribution based on TeX Live. *TUGboat* **40**: 30–32.

Xie, Y. (2020) tinytex: Helper functions to install and maintain 'TeX Live', and compile 'LaTeX' documents.

Xie, Y., Allaire, J.J., and Grolemund, G. (2018) R Markdown: The Definitive Guide, 1st ed. Boca Raton, Florida: 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.

Zhu, H. (2019) kableExtra: Construct complex table with 'kable' and pipe syntax.

Figure Captions

Fig. 1. Confocal microscope images of *C. nodosa* and *C. cylindracea* surfaces from the Bay of Saline and the Bay of Funtana (mixed and monospecific settlements) sampled on 4 December 2017 and stained with SYBR Green I. Scale bar at all images is 60 µm.

Fig. 2. Confocal microscope images of *C. nodosa* and *C. cylindracea* surfaces from the Bay of Funtana (mixed and monospecific settlements) sampled on 19 June 2018 and stained with SYBR Green I. Scale bar at all images is 60 µm.

Fig. 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 (4 December 2017 and 19 June 2018).

Fig. 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 (4 December 2017 and 19 June 2018).

Fig. 5. Relative contribution of each COG category to the total number of annotated coding sequences (A) from metagenomes and identified proteins (B) from metaproteomes associated with 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 (4/14 December 2017 and 19 June 2018). Total number of annotated coding sequences and identified proteins is given above the corresponding bar.

Figures

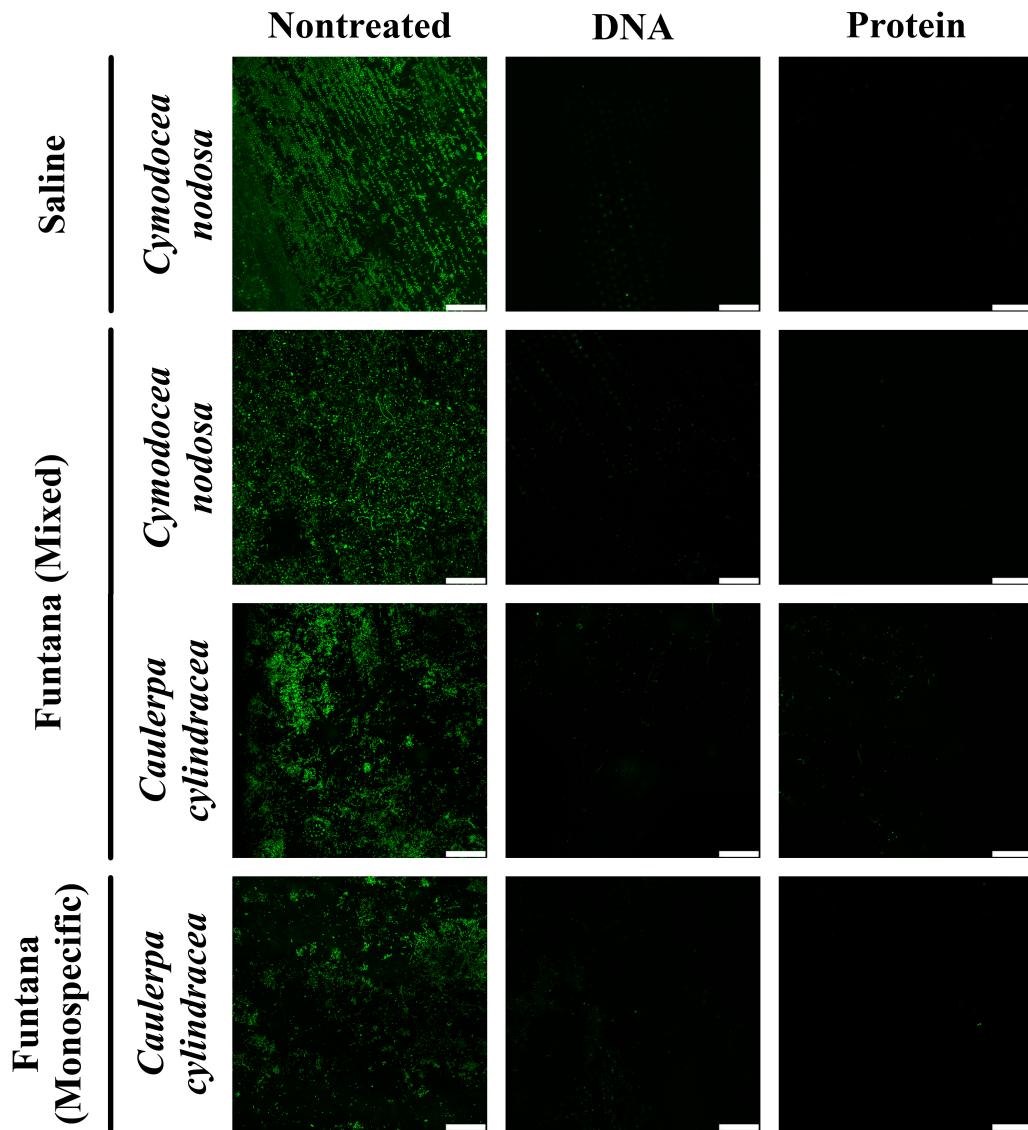


Fig. 1. Confocal microscope images of *C. nodosa* and *C. cylindracea* surfaces from the Bay of Saline and the Bay of Funtana (mixed and monospecific settlements) sampled on 4 December 2017 and stained with SYBR Green I. Scale bar at all images is 60 µm.

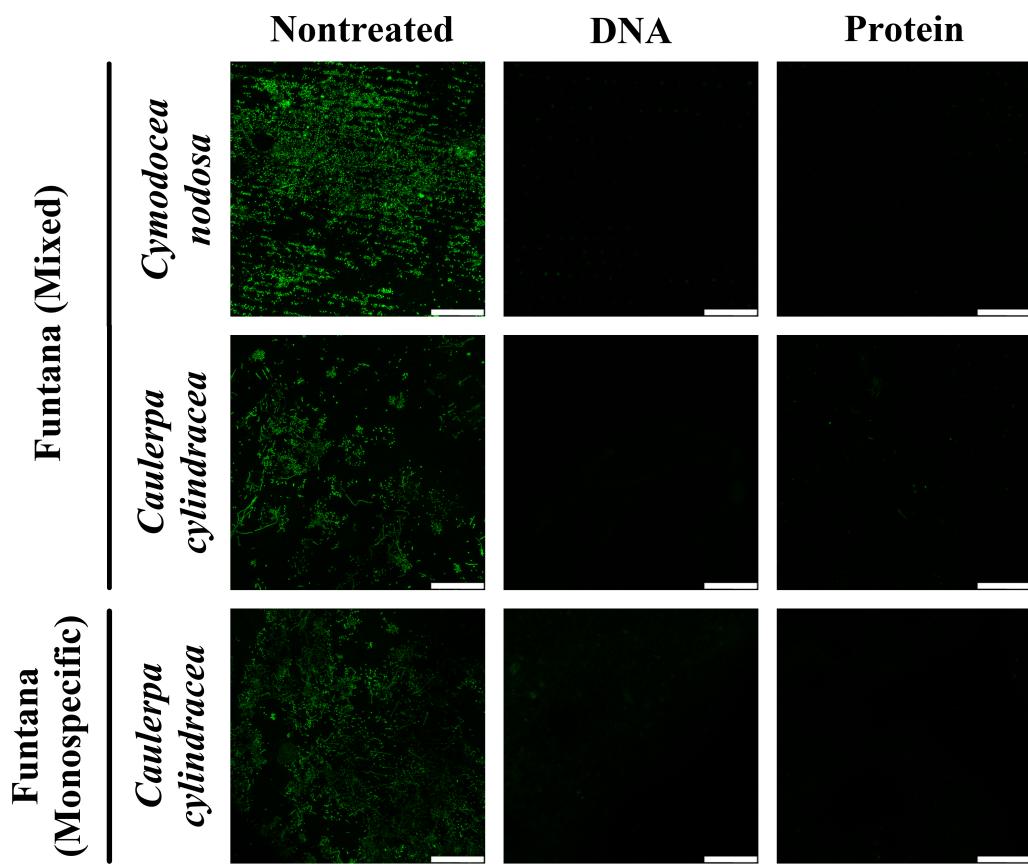


Fig. 2. Confocal microscope images of *C. nodosa* and *C. cylindracea* surfaces from the Bay of Funtana (mixed and monospecific settlements) sampled on 19 June 2018 and stained with SYBR Green I. Scale bar at all images is 60 μ m.

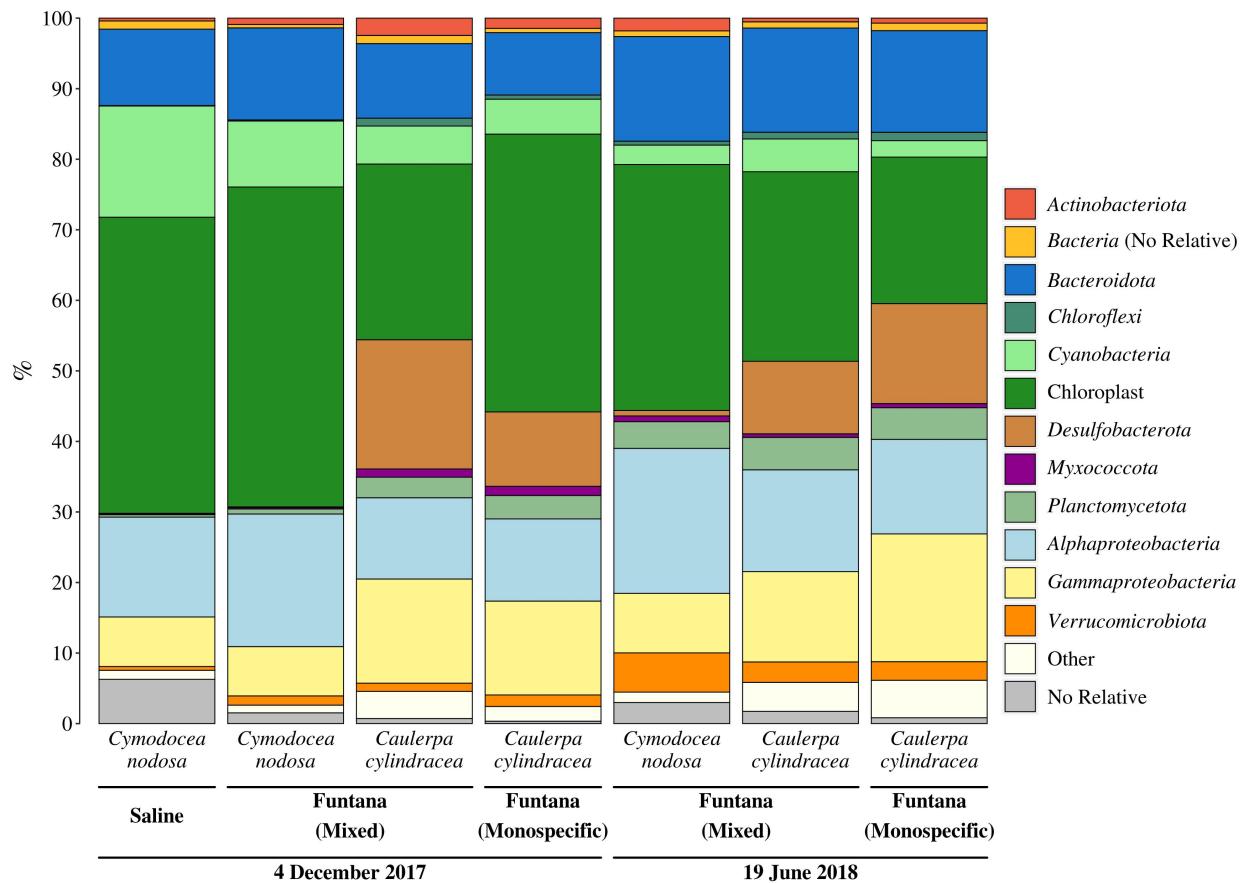


Fig. 3. Taxonomic classification and relative contribution of the most abundant ($\geq 1\%$) bacterial and archaeal sequences from surfaces of two marine macrophytes (*C. nodosa* and *C. cylindracea*) sampled in the Bay of Saline and the Bay of Funtana (mixed and monospecific settlements) and in two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

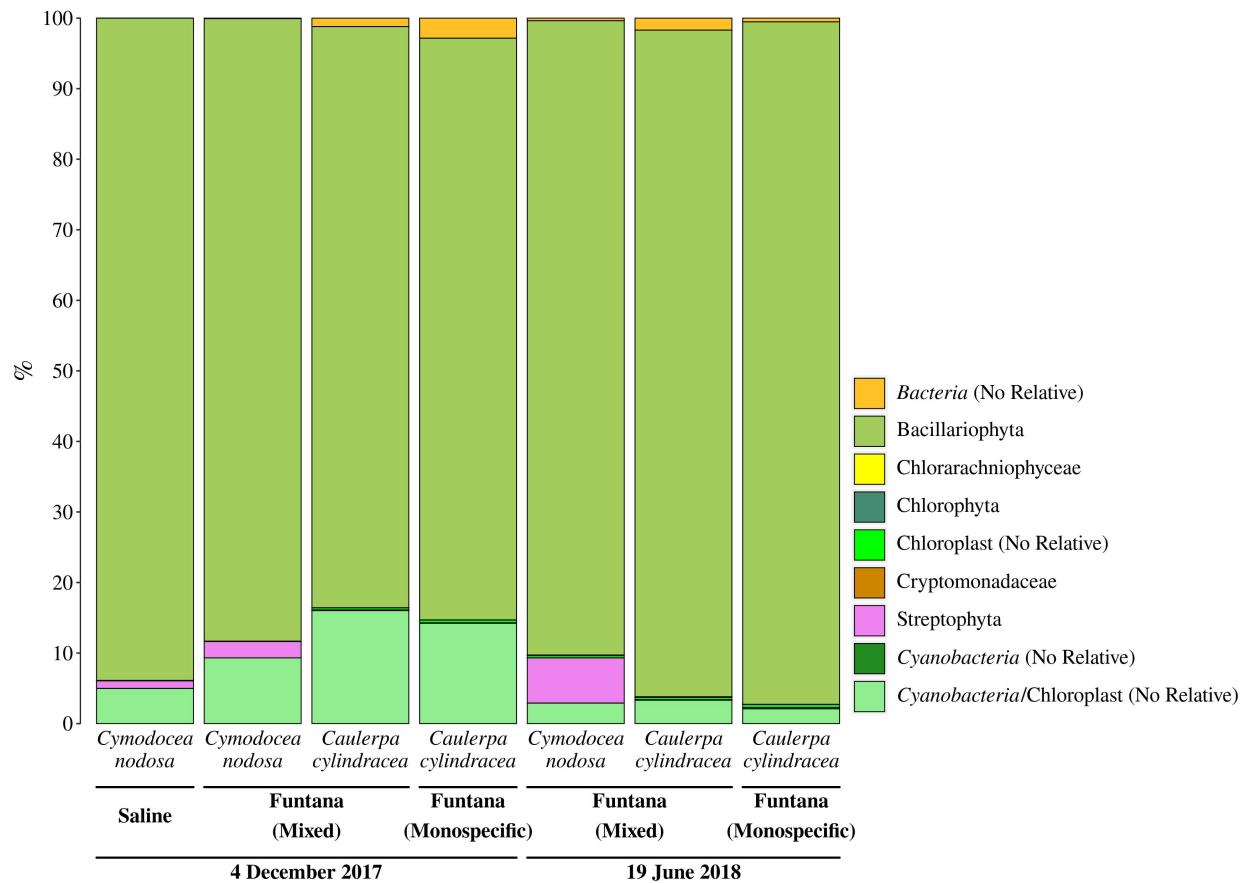


Fig. 4. Taxonomic classification and relative contribution of chloroplast sequences from surfaces of two marine macrophytes (*C. nodosa* and *C. cylindracea*) sampled in the Bay of Saline and the Bay of Funtana (mixed and monospecific settlements) and in two contrasting seasons (4 December 2017 and 19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

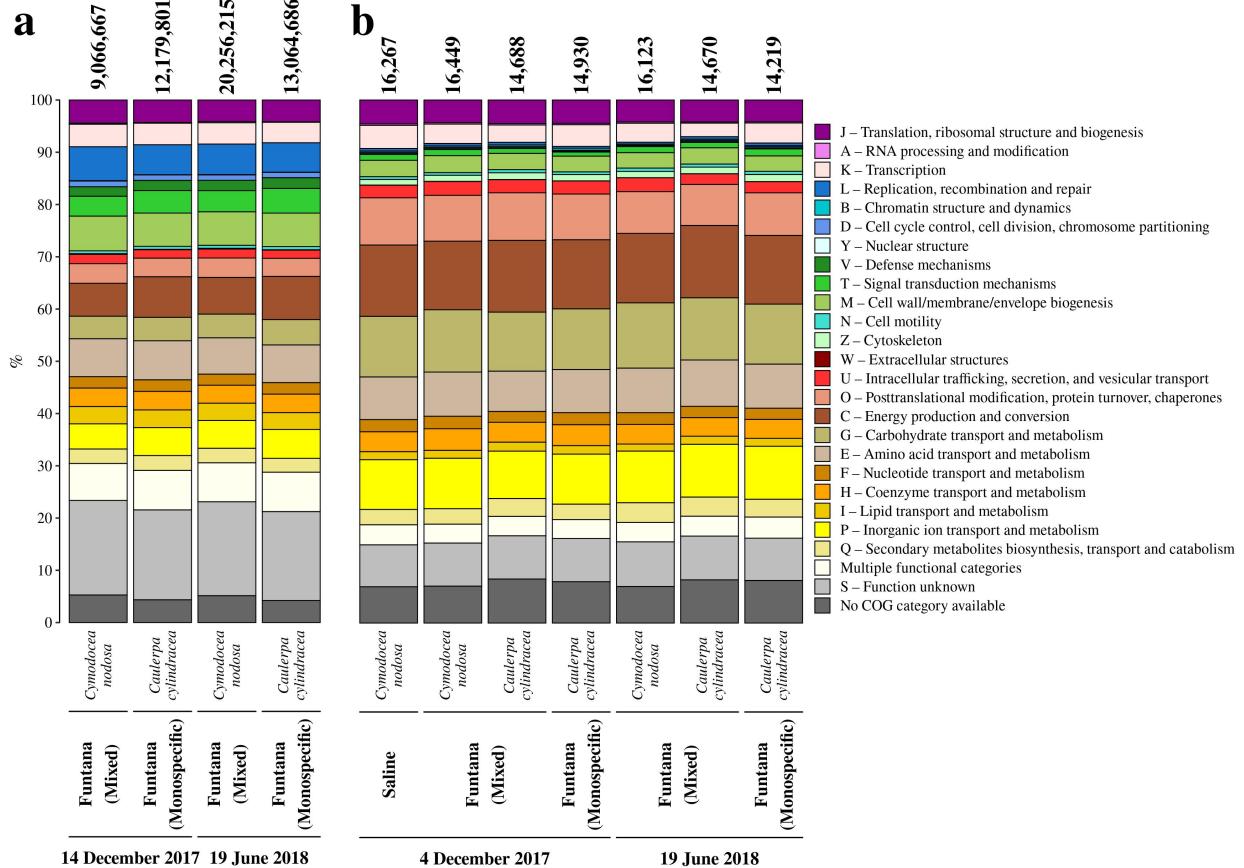


Fig. 5. Relative contribution of each COG category to the total number of annotated coding sequences (A) from metagenomes and identified proteins (B) from metaproteomes associated with surfaces of two marine macrophytes (*C. nodosa* and *C. cylindracea*) sampled in the Bay of Saline and the Bay of Funtana (mixed and monospecific settlements) and in two contrasting seasons (4/14 December 2017 and 19 June 2018). Total number of annotated coding sequences and identified proteins is given above the corresponding bar.