

# 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>, Mirjana Najdek<sup>1</sup>

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

## 1 Abstract

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

17 **Introduction**

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

24 Biological material (i.e., proteins and DNA) from pelagic microbial communities is usually  
25 isolated by collecting cells onto filters and subsequently isolating the target organisms or  
26 communities (Gilbert et al., 2009). If a specific microbial size fraction is aimed sequential filtration  
27 is applied (Massana et al., 1997; Andersson et al., 2010). In contrast, obtaining microorganisms  
28 associated to surfaces require either a cell detachment procedure prior to isolation or the host  
29 material is co-extracted with the target material. Methods for separating microbial cells from the  
30 host include shaking of host tissue (Gross et al., 2003; Nõges et al., 2010), scraping of macrophyte  
31 surfaces (Uku et al., 2007) or applying ultrasonication (Weidner et al., 1996; Cai et al., 2014). It  
32 was shown that shaking alone is not sufficient to remove microbial cells from surfaces, at least  
33 not from plant root surfaces (Richter-Heitmann et al., 2016). Manual separation methods, such as  
34 scraping and brushing are time consuming and subjective, as the detachment efficiency depends  
35 on host tissue and the person performing the procedure (Cai et al., 2014). Ultrasonication was  
36 proposed as an alternative method as it is providing better results in terms of detachment efficiency  
37 (Cai et al., 2014; Richter-Heitmann et al., 2016). The downside of this procedure is that complete  
38 cell removal is still not obtained and tissue disruption was observed especially after the application  
39 of probe ultrasonication (Richter-Heitmann et al., 2016). An alternative to these cell detachment  
40 procedures is the isolation of target epiphytic compounds together with host material (Staufenberger  
41 et al., 2008; Jiang et al., 2015). This procedure can lead to problems in the following processing  
42 steps such as mitochondrial and chloroplast 16S rRNA sequence contaminations from the host

43 (Longford et al., 2007; Staufenberger et al., 2008). In addition, when performing metagenomics  
44 and metaproteomics host material can cause biased results towards more abundant host DNA and  
45 proteins.

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

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

65 **Materials and methods**

66 **Sampling**

67 Leaves of *C. nodosa* were sampled in a *C. nodosa* meadow in the Bay of Saline, northern  
68 Adriatic Sea (45°7'5" N, 13°37'20" E) and in a *C. nodosa* meadow invaded by *C. cylindracea* in  
69 the Bay of Funtana, northern Adriatic Sea (45°10'39" N, 13°35'42" E). Thalli of *C. cylindracea*  
70 were sampled in the same *C. nodosa* invaded meadow in the Bay of Funtana and at a locality of  
71 only *C. cylindracea* located in the proximity of the invaded meadow. Leaves and thalli for 16S  
72 rRNA analysis, metagenomics and metaproteomics were collected in two contrasting seasons, on 4  
73 December 2017 (16S rRNA analysis and metaproteomics), 14 December 2017 (metagenomics) and  
74 18 June 2018 (16S rRNA analysis, metagenomics and metaproteomics). During spring 2018, the *C.*  
75 *nodosa* meadow in the Bay of Saline decayed to an extent that no leaves could be retrieved (Najdek  
76 et al., 2020). In addition, as not enough DNA for both metagenomic and 16S RNA analysis were  
77 obtained during the sampling on 4 December 2017, an additional sampling on 14 December 2017  
78 was carried out in the Bay of Funtana. Leaves and thalli were collected by diving and transported to  
79 the laboratory in containers placed on ice and filled with seawater from the site. Upon arrival to the  
80 laboratory, *C. nodosa* leaves were cut into sections of 1 – 2 cm, while *C. cylindracea* thalli were  
81 cut into 5 – 8 cm long sections. Leaves and thalli were washed three times with sterile artificial  
82 seawater (ASW) to remove loosely attached microbial cells.

83 **DNA isolation**

84 The DNA was isolated according to the protocol for isolation from filters described in Massana  
85 et al. (1997). This protocol was modified and adapted for microbial DNA isolation from macrophyte  
86 surfaces as described below. Five (5) ml of lysis buffer (40 mM EDTA, 50 mM Tris-HCl, 0.75 M  
87 sucrose; pH 8.3) was added to 1 g wet weight of leaves or 2 g wet-weight of thalli. For every sample,

88 duplicate isolations were performed. Lysozyme was added (final concentration 1 mg ml<sup>-1</sup>) and the  
89 mixture was incubated at 37 °C for 30 min. Subsequently, proteinase K (final concentration 0.5  
90 mg ml<sup>-1</sup>) and SDS (final concentration 1 %) were added and the samples were incubated at 55 °C  
91 for 2 h. Following the incubation, tubes were vortexed for 10 min and the mixture containing lysed  
92 epiphytic cells was separated from host leaves or thalli by transferring the solution into a clean tube.  
93 The lysate was extracted twice with a mixture of phenol:chloroform:isoamyl alcohol (25:24:1; pH 8)  
94 and once with chloroform:isoamyl alcohol (24:1). After each addition of an organic solvent mixture,  
95 tubes were slightly vortexed and centrifuged at 4,500 × g for 10 min. Following each centrifugation  
96 the aqueous phases were retrieved. After the final extraction 1/10 of chilled 3 M sodium acetate  
97 (pH 5.2) was added. DNA was precipitated by adding 1 volume of chilled isopropanol, incubating  
98 the mixtures overnight at -20 °C and centrifuging at 16,000 × g and 4 °C for 20 min. The pellet  
99 was washed twice with 1 ml of chilled 70 % ethanol and centrifuged after each washing step at  
100 20,000 × g and 4 °C for 10 min. After the first washing step duplicate pellets from the same sample  
101 were pooled and transferred to a clean 1.5 ml tube. The dried pellet was re-suspended in 100 µl of  
102 deionized water.

### 103 Illumina 16S rRNA sequencing

104 An aliquot of isolated DNA was treated with RNase A (final concentration 200 µg ml<sup>-1</sup>) for 2  
105 h at 37 °C. The DNA concentration was determined using the Quant-iT PicoGreen dsDNA Assay  
106 Kit (Invitrogen, USA) according to the manufacturer's instructions and diluted to 1 ng µl<sup>-1</sup>. The V4  
107 region of the 16S rRNA gene was amplified using a two-step PCR procedure. In the first PCR the  
108 515F (5'-GTGYCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACNVGGGTWTCTAAT-3')  
109 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  
110 al., 2016). These primers contained on their 5' end a tagged sequence. Each sample was amplified  
111 in four parallel 25 µl reactions of which each contained 1 × Q5 Reaction Buffer, 0.2 mM dNTPmix,

113 0.7 mg ml<sup>-1</sup> BSA (Bovine Serum Albumin), 0.2 µM forward and reverse primers, 0.5 U of Q5  
114 High-Fidelity DNA Polymerase (New England Biolabs, USA) and 5 ng of DNA template. Cycling  
115 conditions were: initial denaturation at 94 °C for 3 min, 20 cycles of denaturation at 94 °C for 45 s,  
116 annealing at 50 °C for 60 s and elongation at 72 °C for 90 s, finalized by an elongation step at 72 °C  
117 for 10 min. The four parallel reactions volumes were pooled and PCR products were purified using  
118 the GeneJET PCR Purification Kit (Thermo Fisher Scientific, USA) according to the manufacturer's  
119 instructions and following the protocol that included isopropanol addition for better small DNA  
120 fragment yield. The column was eluted in 30 µl of deionized water. Purified PCR products were  
121 sent for Illumina MiSeq sequencing (2 × 250 bp) at IMGM Laboratories, Martinsried, Germany.  
122 Before sequencing at IMGM, the second PCR amplification of the two-step PCR procedure was  
123 performed using primers targeting the tagged region incorporated in the first PCR. In addition, these  
124 primers contained adapter and sample-specific index sequences. The second PCR was carried out  
125 for 8 cycles. Beside samples, a positive and negative control were sequenced. A negative control  
126 was comprised of four parallel PCR reactions without DNA template, while for a positive control a  
127 mock community composed of evenly mixed DNA material originating from 20 bacterial strains  
128 (ATCC MSA-1002, ATCC, USA) was used. Partial 16S rRNA sequences obtained in this study have  
129 been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession numbers  
130 SAMEA6786270, SAMEA6648792 – SAMEA6648794, SAMEA6648809 – SAMEA6648811 and  
131 SAMEA6648824.

132 Obtained sequences were analysed on the computer cluster Isabella (University Computing  
133 Center, University of Zagreb) using mothur (version 1.43.0) (Schloss et al., 2009) according to the  
134 MiSeq Standard Operating Procedure (MiSeq SOP; [https://mothur.org/wiki/MiSeq\\_SOP](https://mothur.org/wiki/MiSeq_SOP)) (Kozich et  
135 al., 2013) and recommendations given from the Riffomonas project to enhance data reproducibility  
136 (<http://www.riffomonas.org/>). For alignment and classification of sequences the SILVA SSU Ref  
137 NR 99 database (release 138; <http://www.arb-silva.de>) was used (Quast et al., 2013; Yilmaz et  
138 al., 2014). Sequences classified as chloroplasts by SILVA were exported and reclassified using  
139 mothur and the RDP (<http://rdp.cme.msu.edu/>) training set (version 16) reference files adapted for

140 mothur (Cole et al., 2014). In comparison to SILVA, RDP allows a more detailed classification of  
141 chloroplast sequences. Based on the ATCC MSA-1002 mock community included in the analysis a  
142 sequencing error rate of 0.009 % was determined, which is in line with previously reported values  
143 for next-generation sequencing data (Kozich et al., 2013; Schloss et al., 2016). In addition, the  
144 negative control processed together with the samples yielded only 2 sequences after sequence  
145 quality curation.

## 146 Metagenomics

147 Four DNA samples were selected and sent on dry ice to IMGM Laboratories, Martinsried,  
148 Germany for metagenomic sequencing. DNA was purified using AMPure XP Beads (Beckman  
149 Coulter, USA) applying a bead:DNA ratio of 1:1 (v/v), quantified with a Qubit dsDNA HS Assay  
150 Kit (Thermo Fisher Scientific, USA) and check for integrity on a 1 % agarose gel. Metagenomic  
151 sequencing libraries were generated from 300 ng of input DNA using a NEBNext Ultra II FS  
152 DNA Library Prep Kit for Illumina (New England Biolabs, USA) according to the manufacturer's  
153 instructions. Fragments were selected (500 – 700 bp) using AMPure XP Beads, PCR enriched for 3  
154 – 5 cycles and quality controlled. Libraries generated from different DNA samples were pooled and  
155 sequenced on an Illumina NovaSeq 6000 sequencing system (2 × 250 bp).

156 Obtained sequences were analysed on the Life Science Compute Cluster (LiSC) (CUBE –  
157 Computational Systems Biology, University of Vienna). Individual sequences were assembled using  
158 MEGAHIT (version 1.1.2) (Li et al., 2015) under default settings. Putative genes were predicted  
159 from contigs longer than 200 bp using Prodigal (version 2.6.3) (Hyatt et al., 2010) in metagenome  
160 mode (-p meta). Abundances of predicted genes were expressed as Reads Per Kilobase Million  
161 (RPKM) and calculated using the BWA algorithm (version 0.7.16a) (Li and Durbin, 2009). All  
162 predicted genes were functionally annotated using the eggNOG-mapper (Huerta-Cepas et al., 2017)  
163 and eggNOG database (version 5.0) (Huerta-Cepas et al., 2019). Sequence taxonomy classification  
164 was determined using the lowest common ancestor algorithm adapted from DIAMOND (version

165 0.8.36) (Buchfink et al., 2015) and by searching against the NCBI non-redundant database (NR). To  
166 determine the phylogeny, the top 10 % hits with an e-value  $< 1 \times 10^{-5}$  were used (--top 10). Sequence  
167 renaming, coverage information computing and metagenomic statistics calculations were performed  
168 using software tools from BBTools (<https://jgi.doe.gov/data-and-tools/bbtools>). Metagenomic  
169 sequences obtained in this study have been deposited in the European Nucleotide Archive (ENA) at  
170 EMBL-EBI under accession numbers SAMEA6648795, SAMEA6648797, SAMEA6648809 and  
171 SAMEA6648811.

## 172 Protein isolation

173 Proteins were isolated according to the protocol for protein isolation from soil described in  
174 Chourey et al. (2010) and modified by Hultman et al. (2015). This protocol was further modified  
175 and adapted for microbial protein isolation from macrophyte surfaces as described below. Twenty  
176 (20) ml of protein extraction buffer (4 % SDS, 100 mm Tris-HCl [pH 8.0]) was added to 5 g wet  
177 weight of leaves or 10 g wet weight of thalli. The mixture was incubated in boiling water for 5  
178 min, vortexed for 10 min and incubated again in boiling water for 5 min. After a brief vortex, the  
179 lysate was transferred to a clean tube separating the host leaves or thalli from the mixture containing  
180 lysed epiphytic cells. Dithiothreitol (DTT; final concentration 24 mM) was added and proteins were  
181 precipitated with chilled 100 % trichloroacetic acid (TCA; final concentration 20 %) overnight at  
182  $-20^{\circ}\text{C}$ . Precipitated proteins were centrifuged at  $10,000 \times g$  and  $4^{\circ}\text{C}$  for 40 min. The obtained  
183 protein pellet was washed three times with chilled acetone. During the first washing step the pellet  
184 was transferred to a clean 1.5 ml tube. After each washing step, samples were centrifuged at 20,000  
185  $\times g$  and  $4^{\circ}\text{C}$  for 5 min. Dried pellets were stored at  $-80^{\circ}\text{C}$  until further analysis.

186 **Metaproteomics**

187 Isolated proteins were whole trypsin digested using the FASP (filter-aided sample preparation)  
188 Protein Digestion Kit (Expedeon, UK) according to the manufacturer's instructions with small  
189 modifications (Wiśniewski et al., 2009). Prior to loading the solution onto the column, protein pellets  
190 were solubilized in urea sample buffer included in the kit amended with DTT (final concentration  
191 100 mM) for 45 min at room temperature and centrifuged at 20,000 × g for 2 – 5 min at room  
192 temperature to remove larger particles. The first washing step after protein solution loading was  
193 repeated twice. In addition, the centrifugation steps were prolonged if the column was clogged.  
194 Trypsin digestion was performed on column filters at 37 °C overnight for 18 h. The final filtrate  
195 containing peptides was acidified with 1 % (final concentration) trifluoroacetic acid (TFA), freezed  
196 at –80 °C, lyophilized and sent to VIME – Vienna Metabolomics Center (University of Vienna) for  
197 metaproteomic analysis. Peptides were re-suspended in 1 % (final concentration) TFA, desalted  
198 using the Pierce C18 Tips (Thermo Fisher Scientific, USA) according to the manufacturer's  
199 instructions and sequenced on a Q Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer  
200 (Thermo Fisher Scientific, USA). Obtained MS/MS spectra were searched against a protein database  
201 composed of combined sequenced metagenomes obtained in this study using SEQUEST-HT engines  
202 and validated with Percolator in Proteome Discoverer 2.1 (Thermo Fisher Scientific, USA). The  
203 target-decoy approach was used to reduce the probability of false peptide identification. Results  
204 whose false discovery rate at the peptide level was < 1 % were kept. For protein identification  
205 a minimum of two peptides and one unique peptide were required. For protein quantification, a  
206 chromatographic peak area-based free quantitative method was applied.

207 **Data processing and visualization**

208 Processing and visualization of 16S rRNA, metagenomic and metaproteomic data were done  
209 using R (version 3.6.0) (R Core Team, 2019), package tidyverse (version 1.3.0) (Wickham et

210 al., 2019) and multiple other packages (Neuwirth, 2014; Xie, 2014, 2015, 2019a, 2019b; Wilke,  
211 2018; Xie et al., 2018; Allaire et al., 2019; Zhu, 2019; Bengtsson, 2020). The detailed analysis  
212 procedure including the R Markdown file for this paper are available as a GitHub repository  
213 ([https://github.com/MicrobesRovinj/Korlevic\\_SelectiveRemoval\\_FrontMicrobiol\\_2021](https://github.com/MicrobesRovinj/Korlevic_SelectiveRemoval_FrontMicrobiol_2021)).

214 **Confocal microscopy**

215 Host leaves and thalli from DNA and protein isolation steps were washed seven times in  
216 deionized water and fixed with formaldehyde (final concentration ~3 %). In addition, non-treated  
217 leaves and thalli, washed three times in ASW to remove loosely attached microbial cells, were fixed  
218 in the same concentration of formaldehyde and used as a positive control. For long term storage,  
219 fixed leaves and thalli were immersed in a mixture of phosphate-buffered saline (PBS) and ethanol  
220 (1:1) and stored at -20 °C. Treated and untreated segments of leaves and thalli were stained in a 2  
221 × solution of SYBR Green I and examined under a Leica TCS SP8 X FLIM confocal microscope  
222 (Leica Microsystems, Germany).

223 **Results**

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

231 To evaluate whether the obtained DNA is suitable to determine the composition of the microbial  
232 community Illumina sequencing of the V4 region of the 16S rRNA was performed. Sequencing  
233 yielded a total of 292,888 sequences after quality curation and exclusion of eukaryotic, mitochondrial  
234 and no relative sequences. The number of sequences classified as chloroplasts was 97,331. After  
235 excluding these sequences, the total number of retrieved reads was 195,557, ranging from 13,667  
236 to 41,842 sequences per sample (Table S1). Even when the highest sequencing effort was applied,  
237 the rarefaction curves did not level off which is commonly observed in high-throughput 16S rRNA  
238 amplicon sequencing (Fig. S1). Sequences clustering at a similarity level of 97 % yielded a total of  
239 8,355 different OTUs. Taxonomic classification of reads revealed a macrophyte-associated epiphytic  
240 community mainly composed of *Alphaproteobacteria* ( $14.9 \pm 3.5 \%$ ), *Bacteroidota* ( $12.5 \pm 2.4 \%$ ),  
241 *Gammaproteobacteria* ( $11.6 \pm 4.3 \%$ ), *Desulfobacterota* ( $7.8 \pm 7.5 \%$ ), *Cyanobacteria* ( $6.5 \pm 4.7 \%$ )  
242 and *Planctomycetota* ( $2.9 \pm 1.7 \%$ ) (Fig. 3).

243 Primers used in this study, as in many other 16S rRNA amplicon procedures, also amplified  
244 chloroplast 16S rRNA genes. We observed a high proportion of chloroplast sequences in all analysed  
245 samples ( $33.4 \pm 9.4 \%$ ) (Fig. 3). To determine whether chloroplast sequences originate from  
246 the host or eukaryotic epiphytic organisms, we exported SILVA-classified chloroplast sequences  
247 and reclassified them using the RDP (Ribosomal Database Project) training set that allows for

248 a more detailed chloroplast classification. The largest proportion of sequences was classified as  
249 Bacillariophyta ( $89.7 \pm 5.7\%$ ) indicating that the DNA removal procedure resulted in only minor  
250 co-extracted quantities of host DNA (Fig. 4). Chloroplast sequences classified as Streptophyta  
251 constituted  $3.3 \pm 2.8\%$  of all chloroplast sequences originating from *C. nodosa* samples, while  
252 sequences classified as Chlorophyta comprised only  $0.02 \pm 0.01\%$  of all chloroplast sequences  
253 associated with *C. cylindracea* samples.

254 To determine whether the extracted DNA can be used for metagenomic sequencing, four  
255 samples containing epiphytic DNA were selected and shotgun sequenced using an Illumina platform.  
256 Metagenomic sequencing yielded between 207,149,524 and 624,029,930 sequence pairs (Table S2).  
257 Obtained sequences were successfully assembled into contigs whose L50 ranged from 654 to 1,011  
258 bp. In addition, predicted coding sequences were functionally annotated (9,066,667 – 20,256,215  
259 annotated sequences; Fig. 5a) and taxonomically classified. Functional annotation allowed for an  
260 assessment of the relative contribution of each COG (Clusters of Orthologous Groups) functional  
261 category to the total number of annotated coding sequences (Fig. 5a). Functional categories  
262 containing the highest number of sequences were C (Energy production and conversion), E (Amino  
263 acid transport and metabolism), M (Cell wall/membrane/envelope biogenesis), L (Replication,  
264 recombination and repair) and P (Inorganic ion transport and metabolism). If host DNA is  
265 co-extracted with epiphytes it should be detected in large proportions in sequenced metagenomes.  
266 However, no large proportions of coding sequences classified as Streptophyta and Chlorophyta  
267 were detected (Table S3). Sequenced metagenomic DNA originating from the surface of *C. nodosa*  
268 contained 1.3 % of coding sequences classified as Streptophyta in December 2017 and 0.7 % in June  
269 2018. Furthermore, the summed RPKM (Reads Per Kilobase Million) of these sequences constituted  
270 1.7 % of total RPKM of all successfully classified sequences in December 2017 and 1.1 % in June  
271 2018. Similar low proportions of host coding sequences were detected in metagenomic samples  
272 originating from the surfaces of *C. cylindracea*. Of all successfully classified coding sequences 0.2  
273 % were classified as Chlorophyta in December 2017 and 0.1 % in June 2018. A relatively higher  
274 proportion of RPKM of these sequences than in the case of *C. nodosa* was observed, indicating

275 a higher co-extraction of host DNA in *C. cylindracea*. In December, the proportion of RPKM of  
276 sequences classified as Chlorophyta was 8.2 %, while in June 2018 it reached 13.6 %.

277 To evaluate whether the procedure for protein extraction is suitable for metaproteomic analysis,  
278 obtained proteins were trypsin digested and sequenced using a mass spectrometer. Obtained  
279 MS/MS spectra were searched against a protein database from sequenced metagenomes. From  
280 14,219 to 16,449 proteins were identified in isolated protein samples (Fig. 5b). In addition,  
281 successful identification of proteins allowed for an assessment of the relative contribution of  
282 each COG functional category to the total number of identified proteins (Fig. 5b). Functional  
283 categories containing the highest number of identified proteins were C (Energy production and  
284 conversion), G (Carbohydrate transport and metabolism), P (Inorganic ion transport and metabolism),  
285 O (Posttranslational modification, protein turnover, chaperones) and E (Amino acid transport and  
286 metabolism). Isolated proteins could originate from epiphytic organisms inhabiting the macrophyte  
287 surface and/or from macrophyte tissue underlying them. The contribution of proteins originating  
288 from host tissues was evaluated by identifying all the proteins predicted to belong to any taxonomic  
289 group within the phyla Streptophyta and Chlorophyta and by calculating their contribution to the  
290 number and abundance (NAAF – Normalized Abundance Area Factor) of all identified proteins.  
291 On average, proteins isolated from the surface of *C. nodosa* contained  $1.8 \pm 0.06$  % of proteins  
292 associated with Streptophyta, contributing to  $2.2 \pm 0.8$  % of total proteins. Similar to metagenomes,  
293 proteins associated with Chlorophyta contributed more to *C. cylindracea* than proteins associated  
294 with Streptophyta to *C. nodosa*. Chlorophyta associated proteins comprised  $5.2 \pm 0.06$  % of all  
295 identified proteins in *C. cylindracea*, contributing  $19.2 \pm 1.5$  % to the total protein abundance.

296 **Discussion**

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

309 Successful amplification and sequencing of the V4 region of the 16S rRNA gene proved that the  
310 isolated DNA can be used to estimate the microbial epiphytic diversity. Taxonomic groups detected  
311 in this step can also be often found in epiphytic communities associated with other macrophytes  
312 (Burke et al., 2011b; Morrissey et al., 2019). A problem often encountered in studies focusing on  
313 epiphytic communities is the presence of large proportions of chloroplast 16S rRNA sequences in  
314 the pool of amplified molecules, especially if the epiphytic DNA was isolated without prior selection  
315 (Staufenberger et al., 2008). These sequences can derive from host chloroplasts or from eukaryotic  
316 epiphytic chloroplast DNA. Although the proportion of obtained chloroplast 16S rRNA sequences  
317 in our samples was substantial, they derived almost exclusively from eukaryotic epiphytes. High  
318 proportion of chloroplast 16S rRNA sequences in studies applying selective procedures that include  
319 direct cellular lysis on host surfaces were observed before (Michelou et al., 2013). It is possible that  
320 chloroplast-specific sequences even in these studies originated from eukaryotic epiphytic cells and  
321 not from host chloroplasts. Indeed, it is common during 16S rRNA profiling of pelagic microbial

322 communities to observe high proportions of chloroplast sequences (Gilbert et al., 2009; Korlević et  
323 al., 2016). One of the solutions to further reduce chloroplast 16S rRNA sequence contamination is to  
324 use primers that minimise the amplification of these reads if the sequencing and study design allow  
325 it (Hanshew et al., 2013). In addition, a very low proportion of chloroplast 16S rRNA sequences  
326 in samples originating from *C. cylindracea* in comparison to *C. nodosa* could be explained by the  
327 presence of three introns in the gene for 16S rRNA in some members of the genus *Caulerpa* that  
328 could hamper the amplification process (Lam and Lopez-Bautista, 2016).

329 High quality DNA is also needed for metagenomics. The obtained number of metagenomic  
330 sequences and assembly statistics were comparable to metagenomes and metatranscriptomes derived  
331 from similar surface associated communities (Crump et al., 2018; Cúcio et al., 2018). In addition,  
332 functional annotation of predicted coding sequences to COG functional categories showed that  
333 the obtained metagenomes can be used to determine the metabolic capacity of surface associated  
334 communities (Leary et al., 2014; Cúcio et al., 2018). The proportion of coding sequences, including  
335 their RPKM, originating from *C. nodosa* metagenomes and classified as Streptophyta was low  
336 indicating that the isolation procedure was specific for epiphytic cells. DNA samples isolated from  
337 the surface of *C. cylindracea* exhibited a low proportion of Chlorophyta coding sequences, however,  
338 their RPKM was higher than in the samples originating from *C. nodosa*. One of the reasons for this  
339 elevated RPKM of Chlorophyta sequences in *C. cylindracea* could be the differences in the tissue  
340 structure between these two host species. While *C. nodosa* leaves are composed of individual cells,  
341 the thallus of *C. cylindracea* is, like in other siphonous algal species, composed of a single large  
342 multinucleate cell (Coneva and Chitwood, 2015). The absence of individual cells in *C. cylindracea*  
343 could cause a leakage of genetic material into the extraction buffer causing an elevated presence of  
344 host sequences in the samples for metagenome analyses.

345 To obtain insight into the metabolic status of uncultivated prokaryotes, a metaproteomic  
346 approach is required (Saito et al., 2019). The applied protocol for epiphytic protein isolation  
347 followed by a metaproteomic analysis identified between 14,219 and 16,449 proteins, which is

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

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

## 364 Acknowledgements

365 This work was funded by the Croatian Science Foundation through the MICRO-SEAGRASS  
366 project (project number IP-2016-06-7118). ZZ and GJH were supported by the Austrian Science  
367 Fund (FWF) project ARTEMIS (project number P28781-B21). We would like to thank Margareta  
368 Buterer for technical support, Paolo Paliaga for help during sampling, Lucija Horvat for technical  
369 support in confocal microscopy, Dušica Vujaklija for help in accessing the confocal microscope  
370 and Sonja Tischler for technical support in mass spectrometry. In addition, we would like to thank  
371 the University Computing Center of the University of Zagreb for access to the computer cluster

<sup>372</sup> Isabella and CUBE – Computational Systems Biology of the University of Vienna for access to the  
<sup>373</sup> Life Science Compute Cluster (LiSC).

374 **References**

- 375 Allaire, J. J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., et al. (2019).
- 376 *remarkdown: Dynamic documents for R.*
- 377 Amann, R. I., Ludwig, W., and Schleifer, K. H. (1995). Phylogenetic identification and *in situ*  
378 detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59, 143–169.
- 379 Andersson, A. F., Riemann, L., and Bertilsson, S. (2010). Pyrosequencing reveals contrasting  
380 seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *ISME J.* 4, 171–181.  
381 doi:10.1038/ismej.2009.108.
- 382 Apprill, A., McNally, S., Parsons, R., and Weber, L. (2015). Minor revision to V4 region SSU  
383 rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb.*  
384 *Ecol.* 75, 129–137. doi:10.3354/ame01753.
- 385 Bengtsson, H. (2020). *matrixStats: Functions that apply to rows and columns of matrices (and*  
386 *to vectors).*
- 387 Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using  
388 DIAMOND. *Nat. Methods* 12, 59–60. doi:10.1038/nmeth.3176.
- 389 Burke, C., Kjelleberg, S., and Thomas, T. (2009). Selective extraction of bacterial DNA from  
390 the surfaces of macroalgae. *Appl. Environ. Microbiol.* 75, 252–256. doi:10.1128/AEM.01630-08.
- 391 Burke, C., Steinberg, P., Rusch, D., Kjelleberg, S., and Thomas, T. (2011a). Bacterial  
392 community assembly based on functional genes rather than species. *Proc. Natl. Acad. Sci.*  
393 *U. S. A.* 108, 14288–14293. doi:10.1073/pnas.1101591108.

394 Burke, C., Thomas, T., Lewis, M., Steinberg, P., and Kjelleberg, S. (2011b). Composition,  
395 uniqueness and variability of the epiphytic bacterial community of the green alga *Ulva australis*.  
396 *ISME J.* 5, 590–600. doi:10.1038/ismej.2010.164.

397 Cai, X., Gao, G., Yang, J., Tang, X., Dai, J., Chen, D., et al. (2014). An ultrasonic method for  
398 separation of epiphytic microbes from freshwater submerged macrophytes. *J. Basic Microbiol.* 54,  
399 758–761. doi:10.1002/jobm.201300041.

400 Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al.  
401 (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq  
402 platforms. *ISME J.* 6, 1621–1624. doi:10.1038/ismej.2012.8.

403 Chourey, K., Jansson, J., VerBerkmoes, N., Shah, M., Chavarria, K. L., Tom, L. M., et al.  
404 (2010). Direct cellular lysis/protein extraction protocol for soil metaproteomics. *J. Proteome Res.* 9,  
405 6615–6622. doi:10.1021/pr100787q.

406 Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., et al. (2014). Ribosomal  
407 Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 42,  
408 D633–D642. doi:10.1093/nar/gkt1244.

409 Coneva, V., and Chitwood, D. H. (2015). Plant architecture without multicellularity:  
410 Quandaries over patterning and the soma-germline divide in siphonous algae. *Front. Plant Sci.* 6,  
411 287. doi:10.3389/fpls.2015.00287.

412 Crump, B. C., Wojahn, J. M., Tomas, F., and Mueller, R. S. (2018). Metatranscriptomics  
413 and amplicon sequencing reveal mutualisms in seagrass microbiomes. *Front. Microbiol.* 9, 388.  
414 doi:10.3389/fmicb.2018.00388.

415 Cúcio, C., Overmars, L., Engelen, A. H., and Muyzer, G. (2018). Metagenomic analysis shows  
416 the presence of bacteria related to free-living forms of sulfur-oxidizing chemolithoautotrophic  
417 symbionts in the rhizosphere of the seagrass *Zostera marina*. *Front. Mar. Sci.* 5, 171.  
418 doi:10.3389/fmars.2018.00171.

419 Egan, S., Harder, T., Burke, C., Steinberg, P., Kjelleberg, S., and Thomas, T. (2013). The  
420 seaweed holobiont: Understanding seaweed-bacteria interactions. *FEMS Microbiol. Rev.* 37,  
421 462–476. doi:10.1111/1574-6976.12011.

422 Gilbert, J. A., Field, D., Swift, P., Newbold, L., Oliver, A., Smyth, T., et al. (2009). The  
423 seasonal structure of microbial communities in the Western English Channel. *Environ. Microbiol.*  
424 11, 3132–3139. doi:10.1111/j.1462-2920.2009.02017.x.

425 Gross, E. M., Feldbaum, C., and Graf, A. (2003). Epiphyte biomass and elemental composition  
426 on submersed macrophytes in shallow eutrophic lakes. *Hydrobiologia* 506-509, 559–565.  
427 doi:10.1023/B:HYDR.0000008538.68268.82.

428 Hanshew, A. S., Mason, C. J., Raffa, K. F., and Currie, C. R. (2013). Minimization of  
429 chloroplast contamination in 16S rRNA gene pyrosequencing of insect herbivore bacterial  
430 communities. *J. Microbiol. Methods* 95, 149–155. doi:10.1016/j.mimet.2013.08.007.

431 Huerta-Cepas, J., Forsslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., et al.  
432 (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper.  
433 *Mol. Biol. Evol.* 34, 2115–2122. doi:10.1093/molbev/msx148.

434 Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forsslund, S. K., Cook,  
435 H., et al. (2019). eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated  
436 orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314.  
437 doi:10.1093/nar/gky1085.

- 438 Hultman, J., Waldrop, M. P., Mackelprang, R., David, M. M., McFarland, J., Blazewicz, S.  
439 J., et al. (2015). Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes.  
440 *Nature* 521, 208–212. doi:10.1038/nature14238.
- 441 Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010).  
442 Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf.* 11,  
443 119. doi:10.1186/1471-2105-11-119.
- 444 Jiang, Y.-F., Ling, J., Dong, J.-D., Chen, B., Zhang, Y.-Y., Zhang, Y.-Z., et al. (2015).  
445 Illumina-based analysis the microbial diversity associated with *Thalassia hemprichii* in Xincun Bay,  
446 South China Sea. *Ecotoxicology* 24, 1548–1556. doi:10.1007/s10646-015-1511-z.
- 447 Korlević, M., Šupraha, L., Ljubešić, Z., Henderiks, J., Ciglenečki, I., Dautović, J., et al. (2016).  
448 Bacterial diversity across a highly stratified ecosystem: A salt-wedge Mediterranean estuary. *Syst.*  
449 *Appl. Microbiol.* 39, 398–408. doi:10.1016/j.syapm.2016.06.006.
- 450 Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D. (2013).  
451 Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon  
452 sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* 79,  
453 5112–5120. doi:10.1128/AEM.01043-13.
- 454 Kuo, J., and den Hartog, C. (2001). “Seagrass taxonomy and identification key,” in *Global*  
455 *Seagrass Research Methods*, eds. F. T. Short and R. G. Coles (Amsterdam: Elsevier Science B.V.),  
456 31–58.
- 457 Lam, D. W., and Lopez-Bautista, J. M. (2016). Complete chloroplast genome for *Caulerpa*  
458 *racemosa* (Bryopsidales, Chlorophyta) and comparative analyses of siphonous green seaweed  
459 plastomes. *Cymbella* 2, 23–32.

- 460 Leary, D. H., Li, R. W., Hamdan, L. J., Judson Hervey IV, W., Lebedev, N., Wang, Z., et  
461 al. (2014). Integrated metagenomic and metaproteomic analyses of marine biofilm communities.  
462 *Biofouling* 30, 1211–1223. doi:10.1080/08927014.2014.977267.
- 463 Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: An ultra-fast  
464 single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph.  
465 *Bioinformatics* 31, 1674–1676. doi:10.1093/bioinformatics/btv033.
- 466 Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler  
467 transform. *Bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324.
- 468 Longford, S., Tujula, N., Crocetti, G., Holmes, A., Holmström, C., Kjelleberg, S., et al. (2007).  
469 Comparisons of diversity of bacterial communities associated with three sessile marine eukaryotes.  
470 *Aquat. Microb. Ecol.* 48, 217–229. doi:10.3354/ame048217.
- 471 Massana, R., Murray, A. E., Preston, C. M., and DeLong, E. F. (1997). Vertical distribution  
472 and phylogenetic characterization of marine planktonic *Archaea* in the Santa Barbara Channel. *Appl.*  
473 *Environ. Microbiol.* 63, 50–56.
- 474 Michelou, V. K., Caporaso, J. G., Knight, R., and Palumbi, S. R. (2013). The ecology  
475 of microbial communities associated with *Macrocystis pyrifera*. *PLoS One* 8, e67480.  
476 doi:10.1371/journal.pone.0067480.
- 477 Morrissey, K. L., Çavas, L., Willems, A., and De Clerck, O. (2019). Disentangling the influence  
478 of environment, host specificity and thallus differentiation on bacterial communities in siphonous  
479 green seaweeds. *Front. Microbiol.* 10, 717. doi:10.3389/fmicb.2019.00717.
- 480 Najdek, M., Korlević, M., Paliaga, P., Markovski, M., Ivančić, I., Iveša, L., et al. (2020).  
481 Dynamics of environmental conditions during the decline of a *Cymodocea nodosa* meadow.  
482 *Biogeosciences* 17, 3299–3315. doi:10.5194/bg-17-3299-2020.

- 483 Neuwirth, E. (2014). *RColorBrewer: ColorBrewer palettes*.
- 484 Nõges, T., Luup, H., and Feldmann, T. (2010). Primary production of aquatic macrophytes  
485 and their epiphytes in two shallow lakes (Peipsi and Võrtsjärv) in Estonia. *Aquat. Ecol.* 44, 83–92.  
486 doi:10.1007/s10452-009-9249-4.
- 487 Parada, A. E., Needham, D. M., and Fuhrman, J. A. (2016). Every base matters: Assessing  
488 small subunit rRNA primers for marine microbiomes with mock communities, time series and  
489 global field samples. *Environ. Microbiol.* 18, 1403–1414. doi:10.1111/1462-2920.13023.
- 490 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA  
491 ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic  
492 Acids Res.* 41, D590–D596. doi:10.1093/nar/gks1219.
- 493 R Core Team (2019). *R: A language and environment for statistical computing*. Vienna,  
494 Austria: R Foundation for Statistical Computing.
- 495 Reyes, J., and Sansón, M. (2001). Biomass and production of the epiphytes on the leaves of  
496 *Cymodocea nodosa* in the Canary Islands. *Bot. Mar.* 44, 307–313. doi:10.1515/BOT.2001.039.
- 497 Richter-Heitmann, T., Eickhorst, T., Knauth, S., Friedrich, M. W., and Schmidt, H. (2016).  
498 Evaluation of strategies to separate root-associated microbial communities: A crucial choice in  
499 rhizobiome research. *Front. Microbiol.* 7, 773. doi:10.3389/fmicb.2016.00773.
- 500 Saito, M. A., Bertrand, E. M., Duffy, M. E., Gaylord, D. A., Held, N. A., Hervey, W. J., et  
501 al. (2019). Progress and challenges in ocean metaproteomics and proposed best practices for data  
502 sharing. *J. Proteome Res.* 18, 1461–1476. doi:10.1021/acs.jproteome.8b00761.
- 503 Schloss, P. D., Jenior, M. L., Koumpouras, C. C., Westcott, S. L., and Highlander, S. K. (2016).  
504 Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. *PeerJ* 4,  
505 e1869. doi:10.7717/peerj.1869.

- 506 Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al.  
507 (2009). Introducing mothur: Open-source, platform-independent, community-supported software  
508 for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541.  
509 doi:10.1128/AEM.01541-09.
- 510 Staufenberger, T., Thiel, V., Wiese, J., and Imhoff, J. F. (2008). Phylogenetic analysis  
511 of bacteria associated with *Laminaria saccharina*. *FEMS Microbiol. Ecol.* 64, 65–77.  
512 doi:10.1111/j.1574-6941.2008.00445.x.
- 513 Su, C., Lei, L., Duan, Y., Zhang, K.-Q., and Yang, J. (2012). Culture-independent methods for  
514 studying environmental microorganisms: Methods, application, and perspective. *Appl. Microbiol.*  
515 *Biotechnol.* 93, 993–1003. doi:10.1007/s00253-011-3800-7.
- 516 Uku, J., Björk, M., Bergman, B., and Díez, B. (2007). Characterization and comparison  
517 of prokaryotic epiphytes associated with three East African seagrasses. *J. Phycol.* 43, 768–779.  
518 doi:10.1111/j.1529-8817.2007.00371.x.
- 519 Verlaque, M., Durand, C., Huisman, J. M., Boudouresque, C.-F., and Le Parco, Y. (2003). On  
520 the identity and origin of the Mediterranean invasive *Caulerpa racemosa* (Caulerpales, Chlorophyta).  
521 *Eur. J. Phycol.* 38, 325–339. doi:10.1080/09670260310001612592.
- 522 Weidner, S., Arnold, W., and Puhler, A. (1996). Diversity of uncultured microorganisms  
523 associated with the seagrass *Halophila stipulacea* estimated by restriction fragment length  
524 polymorphism analysis of PCR-amplified 16S rRNA genes. *Appl. Environ. Microbiol.* 62, 766–771.
- 525 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., et al. (2019).  
526 Welcome to the tidyverse. *J. Open Source Softw.* 4, 1686. doi:10.21105/joss.01686.
- 527 Wilke, C. O. (2018). *cowplot: Streamlined plot theme and plot annotations for 'ggplot2'*.

- 528 Williams, T. J., Long, E., Evans, F., DeMaere, M. Z., Lauro, F. M., Raftery, M. J., et al. (2012).
- 529 A metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula
- 530 coastal surface waters. *ISME J.* 6, 1883–1900. doi:10.1038/ismej.2012.28.
- 531 Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample
- 532 preparation method for proteome analysis. *Nat. Methods* 6, 359–362. doi:10.1038/nmeth.1322.
- 533 Xie, Y. (2014). “knitr: A comprehensive tool for reproducible research in R,” in *Implementing*
- 534 *Reproducible Computational Research*, eds. V. Stodden, F. Leisch, and R. D. Peng (New York:
- 535 Chapman and Hall/CRC), 3–32.
- 536 Xie, Y. (2015). *Dynamic Documents with R and knitr*. 2nd ed. Boca Raton, Florida: Chapman
- 537 and Hall/CRC.
- 538 Xie, Y. (2019a). TinyTeX: A lightweight, cross-platform, and easy-to-maintain LaTeX
- 539 distribution based on TeX Live. *TUGboat* 40, 30–32.
- 540 Xie, Y. (2019b). *tinytex: Helper functions to install and maintain 'TeX Live', and compile*
- 541 *'LaTeX' documents*.
- 542 Xie, Y., Allaire, J. J., and Grolemund, G. (2018). *R Markdown: The Definitive Guide*. 1st ed.
- 543 Boca Raton, Florida: Chapman and Hall/CRC.
- 544 Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., et al. (2014). The
- 545 SILVA and "All-Species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* 42,
- 546 D643–D648. doi:10.1093/nar/gkt1209.
- 547 Zhu, H. (2019). *kableExtra: Construct complex table with 'kable' and pipe syntax*.

548 **Figure legends**

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

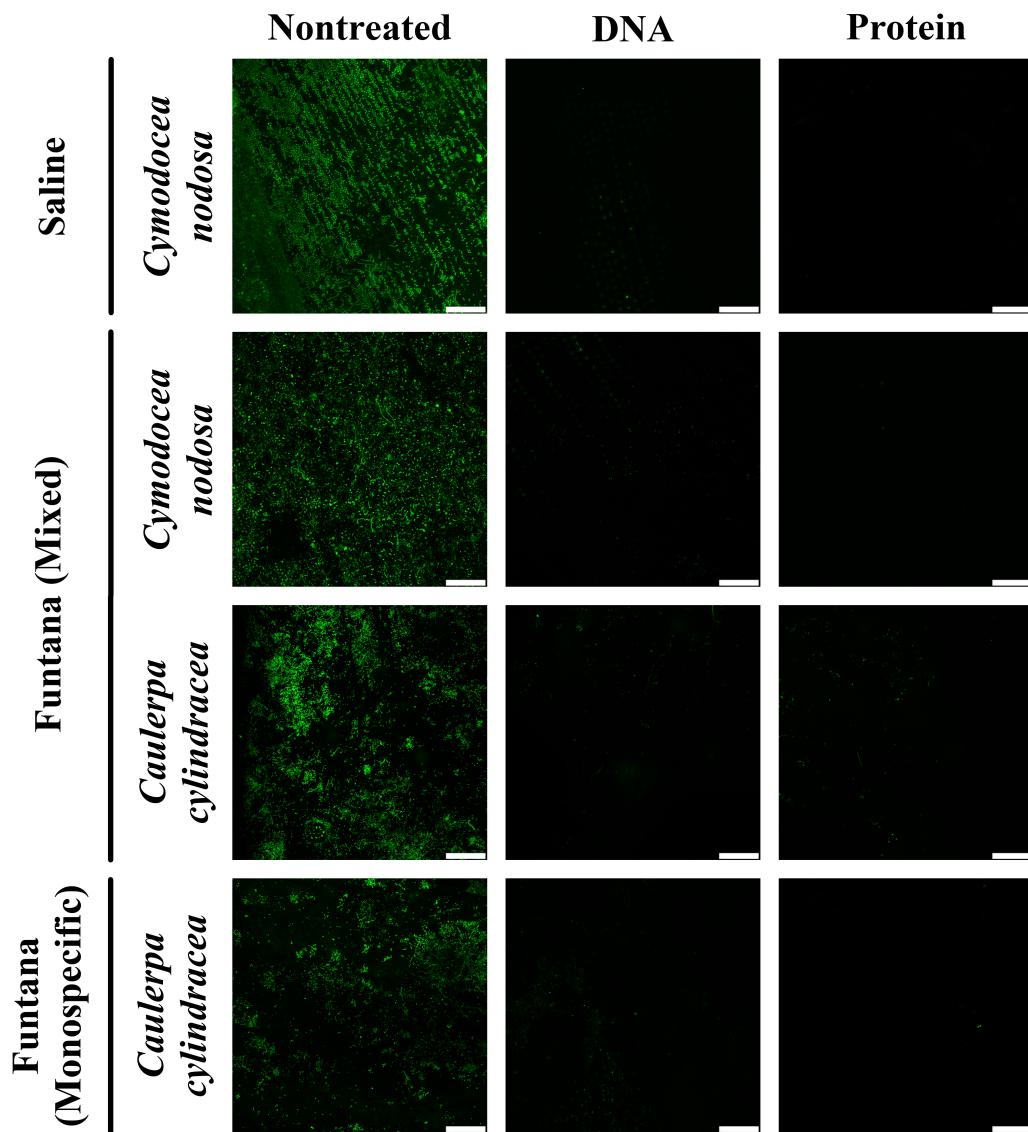
552 **Fig. 2.** Confocal microscope images of *C. nodosa* and *C. cylindracea* surfaces from the Bay of  
553 Funtana (mixed and monospecific settlements) sampled on 19 June 2018 and stained with SYBR  
554 Green I. Scale bar at all images is 60  $\mu$ m.

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

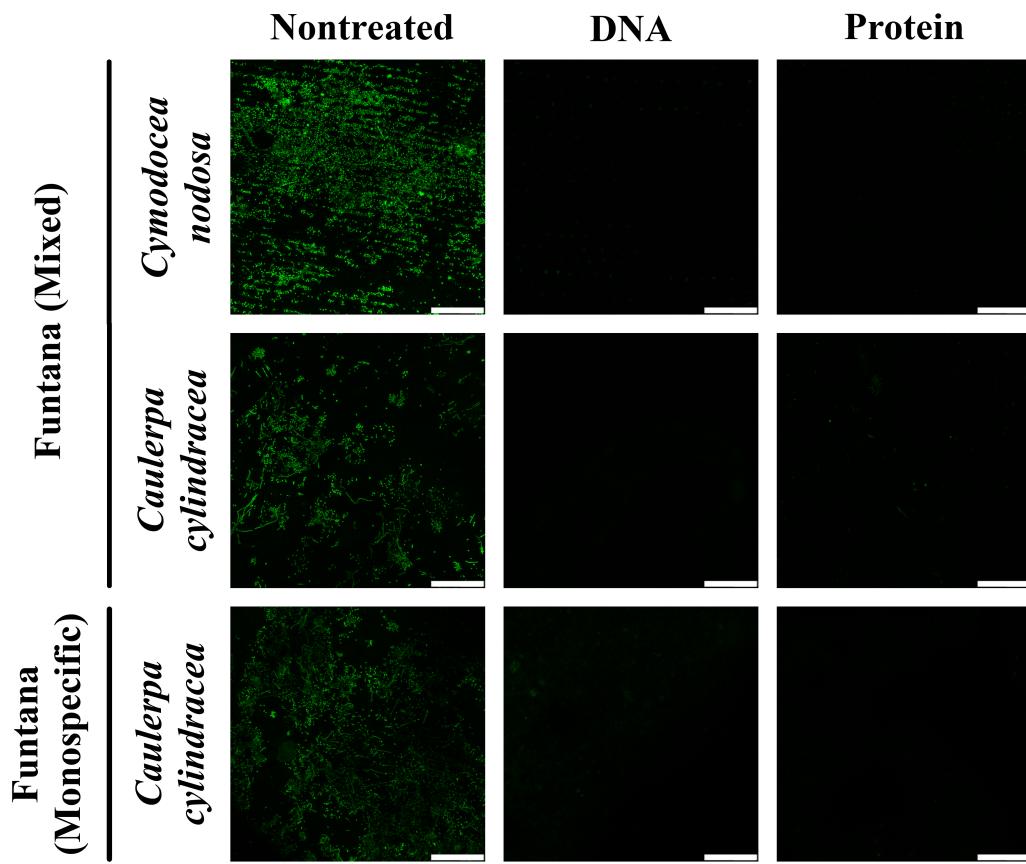
559 **Fig. 4.** Taxonomic classification and relative contribution of chloroplast sequences from surfaces of  
560 two marine macrophytes (*C. nodosa* and *C. cylindracea*) sampled in the Bay of Saline and the Bay  
561 of Funtana (mixed and monospecific settlements) and in two contrasting seasons (4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

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

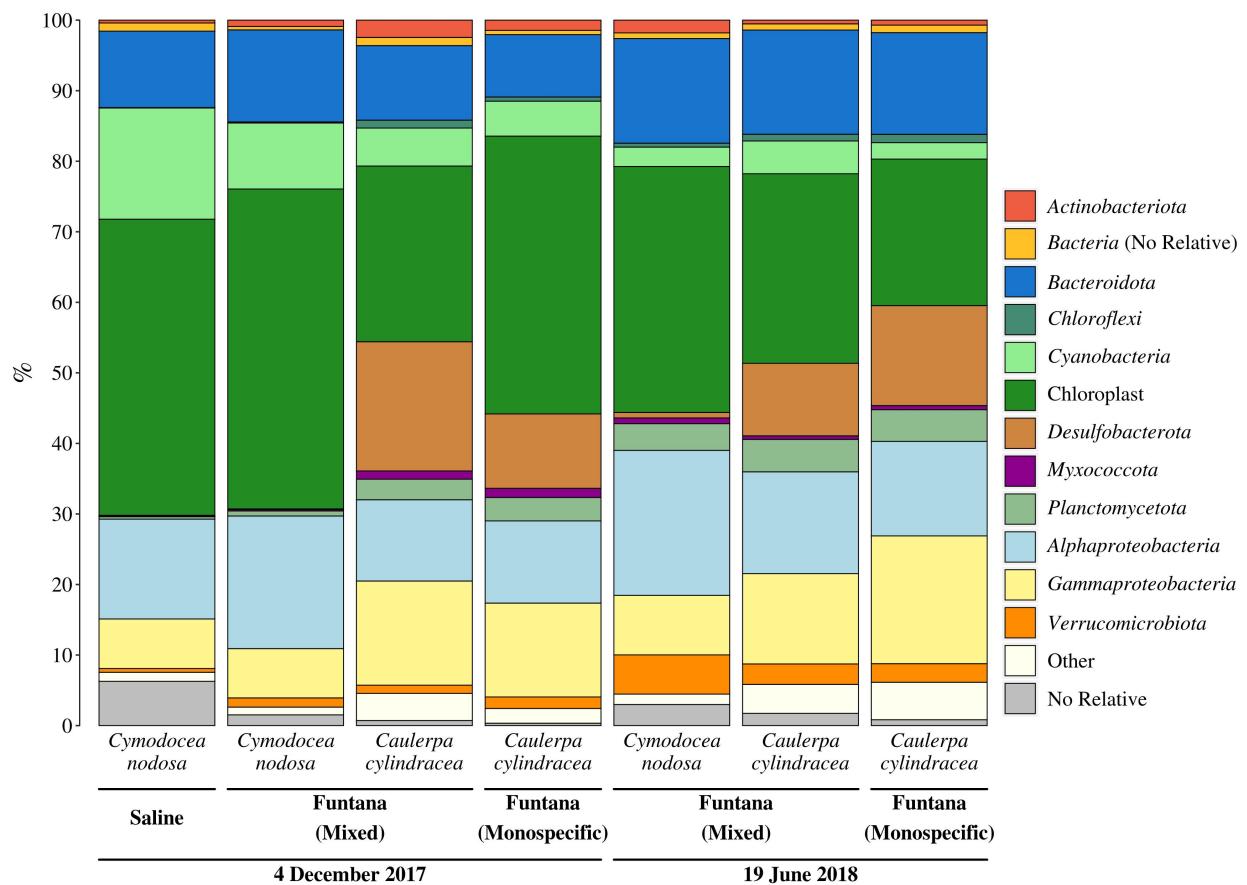
569 **Figures**



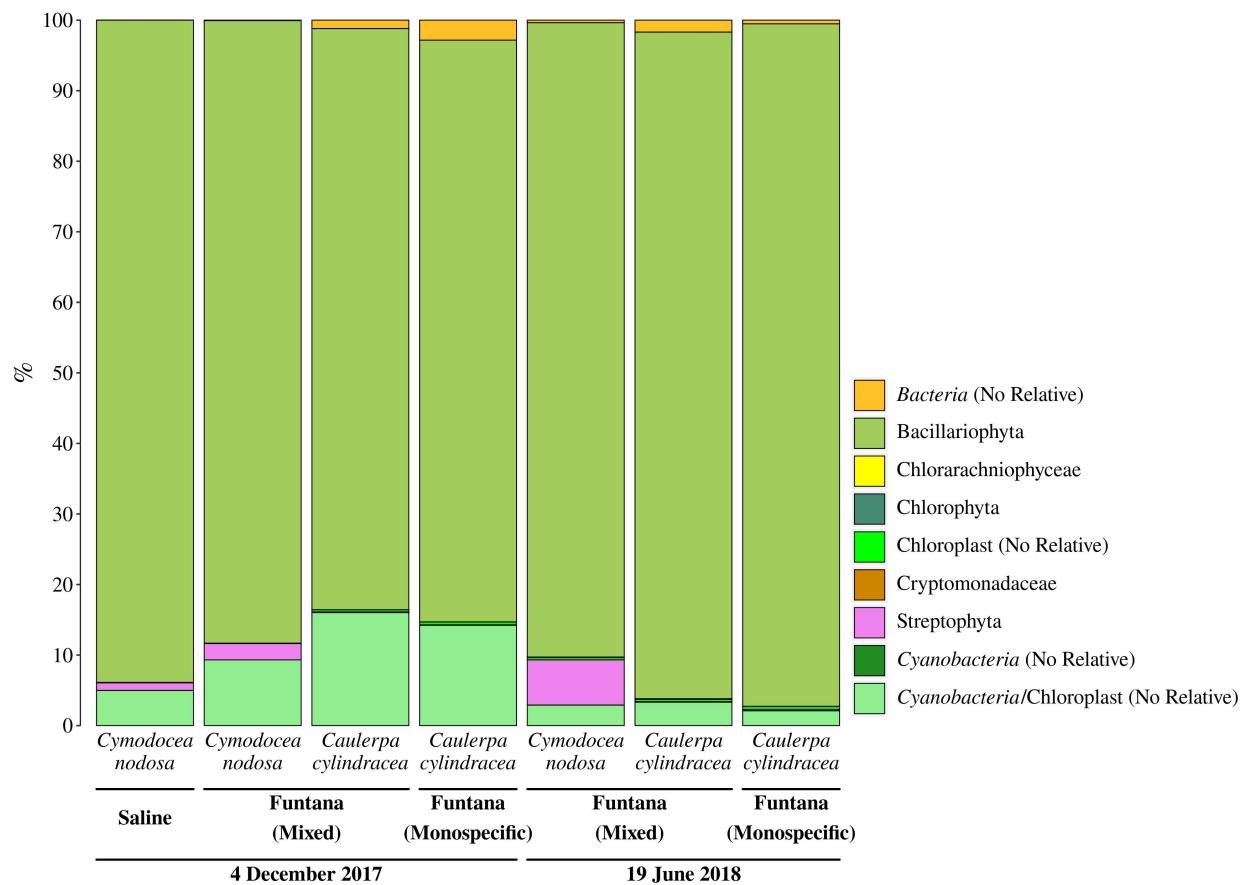
**Fig. 1.** Confocal microscope images of *C. nodosa* and *C. cylindracea* surfaces from the Bay of Saline and the Bay of Funtana (mixed and monospecific settlements) sampled on 4 December 2017 and stained with SYBR Green I. Scale bar at all images is 60  $\mu$ m.



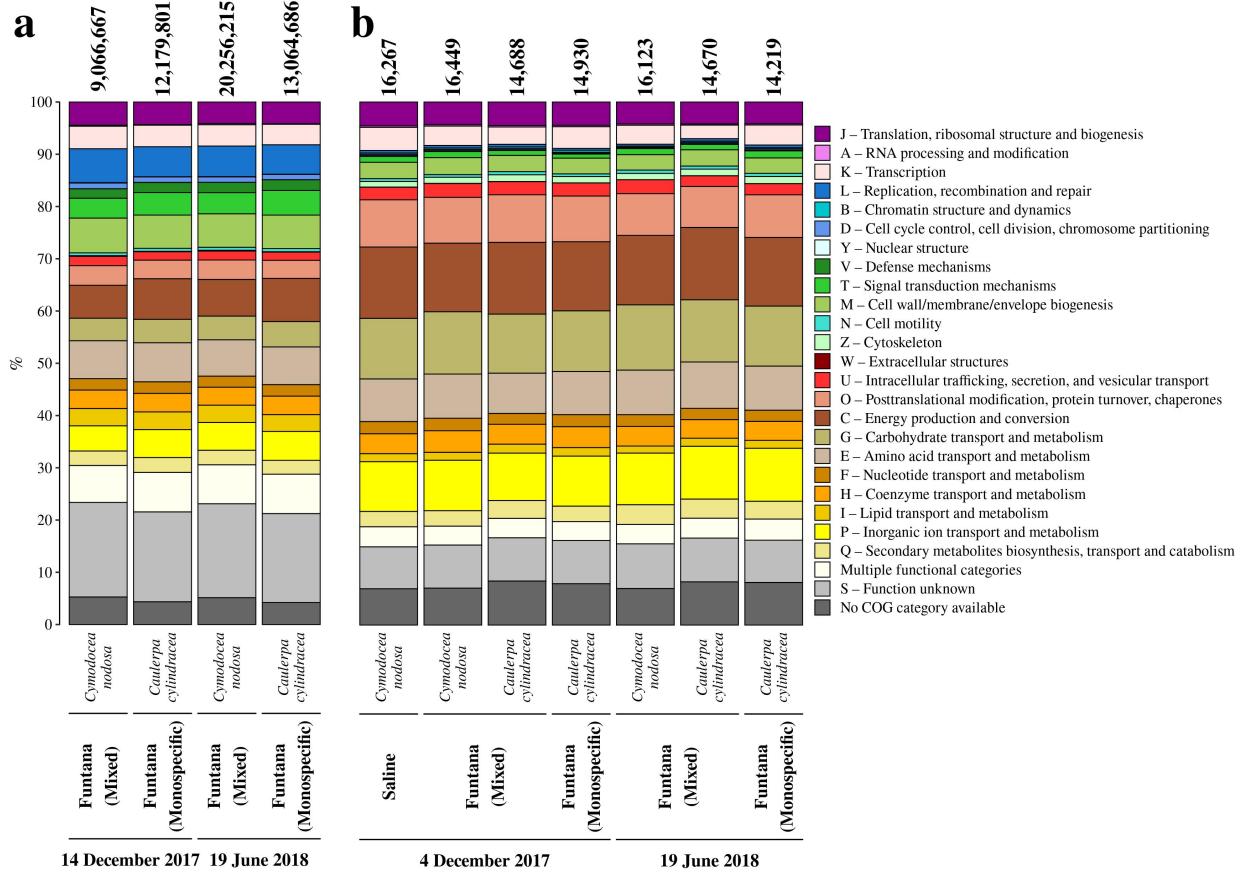
**Fig. 2.** Confocal microscope images of *C. nodosa* and *C. cylindracea* surfaces from the Bay of Funtana (mixed and monospecific settlements) sampled on 19 June 2018 and stained with SYBR Green I. Scale bar at all images is 60  $\mu$ m.



**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.