

Selective DNA and Protein Isolation from Marine Macrophyte Surfaces

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

*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 Summary

2 Studies of unculturable microbes often combine methods based on DNA, such as 16S rRNA
3 sequencing and metagenomics, with methods that allow insight into the metabolic status, such as
4 metaproteomics. To apply these techniques to the microbial community inhabiting the surfaces of
5 marine macrophytes it is advisable to perform, prior to the analysis, a selective DNA and protein
6 isolation so that the host material, present in higher quantities, is not hampering the analysis.
7 Two protocols, for DNA and protein isolation, were adapted for selective extractions of DNA
8 and proteins from epiphytic communities inhabiting the surfaces of two marine macrophytes: the
9 seagrass *Cymodocea nodosa* and the macroalga *Caulerpa cylindracea*. Protocols showed an almost
10 complete removal of the epiphytic community regardless of the sampling season, station, settlement
11 or host species. The obtained DNA was suitable for metagenomic and 16S rRNA sequencing, while
12 isolated proteins could be identified by mass spectrometry, showing that protocols can be used in 16S
13 rRNA, metagenomic and metaproteomic analysis. Low presence of host DNA and proteins, observed
14 in isolated samples, indicated a selective nature of the protocols. Furthermore, the procedures are
15 based on universally available laboratory chemicals making the protocols widely applicable. Taken
16 together, the adapted protocols ensure an almost complete removal of the macrophyte epiphytic
17 community, are selective for microbes inhabiting macrophyte surfaces and are providing DNA and
18 proteins applicable in 16S rRNA sequencing, metagenomics and metaproteomics.

19 **Introduction**

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

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

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

48 An alternative to these procedures is a direct isolation of the targeted material by incubating
49 macrophyte tissues in an extraction buffer. After the incubation, the undisrupted host tissue is
50 removed and the isolation procedure continues, omitting host material contaminations. To our
51 knowledge the only procedure describing a direct and selective epiphytic DNA isolation from the
52 surfaces of marine macrophytes was provided by Burke *et al.* (2009). In contrast to previously
53 described methods this protocol enables an almost complete removal of the surface community and
54 it was used for 16S rRNA gene clone library construction (Burke and Thomas *et al.*, 2011) and
55 metagenome sequencing (Burke and Peter Steinberg *et al.*, 2011). This method, though providing a
56 selective isolation procedure, uses a rapid multienzyme cleaner (3M) that is not available worldwide
57 and without a known composition (Burke *et al.*, 2009). Also to our knowledge, no selective isolation
58 protocol for proteins from epiphytic communities inhabiting marine macrophytes was established.

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

65 **Results**

66 To assess the removal efficiency of DNA and protein isolation procedures leaves and thalli were
67 examined under a confocal microscope before and after treatments were performed. Developed
68 procedures showed an almost complete removal of the surface community of both, *C. nodosa* and
69 *C. cylindracea*. In addition, a similar removal efficiency was observed for communities sampled
70 in contrasting seasons, December 2017 (Fig. 1) and June 2018 (Fig. 2). Also, no effect of station,
71 settlement or isolation procedure (DNA or protein) on the removal efficiency was observed (Figs. 1
72 and 2).

73 To evaluate if the obtained DNA is suitable to determine the microbial community structure
74 an Illumina sequencing of the V4 16S rRNA region was performed. Sequencing yielded a total
75 of 336,937 sequences after quality curation and exclusion of eukaryotic, mitochondrial and no
76 relative sequences. The number of sequences classified as chloroplast was 97,328. After excluding
77 these sequences the total number of retrieved reads dropped to 239,609, ranging from 22,587 to
78 52,958 sequences per sample (Table S1). Even when the highest sequencing effort was applied the
79 rarefaction curves did not level off which is commonly observed in high-throughput 16S rRNA
80 amplicon sequencing procedures (Fig. S1). Sequence clustering at a similarity level of 97 % yielded a
81 total of 8,360 different OTUs. Taxonomic classification of reads allowed for a macrophyte associated
82 epiphytic community determination that was mainly composed of: *Alphaproteobacteria* (14.9 ± 3.5
83 %), *Bacteroidota* (12.5 ± 2.4 %), *Gammaproteobacteria* (11.6 ± 4.3 %), *Desulfobacterota* (7.8 ±
84 7.4 %), *Cyanobacteria* (6.5 ± 4.7 %) and *Planctomycetota* (2.9 ± 1.7 %) (Fig. 3).

85 Primers used in this study, as in many other 16S rRNA amplicon procedures, also amplified
86 chloroplast 16S rRNA genes. We observed a high proportion of chloroplast sequences in all
87 analyzed samples (33.4 ± 9.4 %) (Fig. 3). To determine if chloroplast sequences originate from
88 hosts or eukaryotic epiphytic organisms, we exported SILVA classified chloroplast sequences
89 and reclassified them using the RDP (Ribosomal Database Project) training set that allows for a

more detailed chloroplast classification. The largest proportion of sequences were classified as Bacillariophyta ($89.7 \pm 5.7\%$) indicating that the DNA removal procedure did not coextract larger 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 if 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 from 207,149,524 to 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 successfully 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 was coextracted with epiphytic it should be detected in larger proportions in sequenced metagenomes. Indeed, no higher proportions of coding sequences classified into phylum Streptophyta or Chlorophyta were detected (Table S3). Sequenced metagenomic DNA originating from the surface of *C. nodosa* contained 1.3 % of coding sequences classified into the phylum 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 into Chlorophyta in December 2017 and 0.1 % in June 2018. A relatively higher proportion of these sequences'

¹¹⁷ RPKM than in the case of *C. nodosa* was observed, indicating a higher coextraction of host DNA in
¹¹⁸ *C. cylindracea*. In December, the proportion of RPKM of sequences classified into Chlorophyta
¹¹⁹ was 8.2 %, while in June 2018 it reached 13.6 %.

¹²⁰ To evaluate if the procedure for protein extraction is suitable for metaproteomic analysis,
¹²¹ obtained proteins were trypsin digested and sequenced using a mass spectrometer. Obtained
¹²² MS/MS spectra were searched against a protein database from sequenced metagenomes. From
¹²³ 14,219 to 16,449 proteins were identified in isolated protein samples (Fig. 5b). In addition,
¹²⁴ successful identification of proteins allowed for an assessment of the relative contribution of
¹²⁵ each COG functional category to the total number of identified proteins (Fig. 5b). Functional
¹²⁶ categories containing the highest number of identified proteins were: C (Energy production and
¹²⁷ conversion), G (Carbohydrate transport and metabolism), P (Inorganic ion transport and metabolism),
¹²⁸ O (Posttranslational modification, protein turnover, chaperones) and E (Amino acid transport and
¹²⁹ metabolism). Isolated proteins could derive from epiphytic organisms inhabiting the macrophyte
¹³⁰ surface or from macrophyte tissue underlying them. The contribution of proteins originating from
¹³¹ host tissues was evaluated by identifying all the proteins predicted to belong to any taxonomic group
¹³² within the phyla Streptophyta and Chlorophyta and by calculating their contribution to the number
¹³³ and abundance of all identified proteins. On average, proteins isolated from the surface of *C. nodosa*
¹³⁴ contained 1.8 ± 0.06 % of proteins associated with Streptophyta, contributing to 2.2 ± 0.8 % of
¹³⁵ total proteins. Similar to metagenomes, proteins associated with Chlorophyta contributed more to
¹³⁶ *C. cylindracea* than proteins associated with Streptophyta to *C. nodosa*. Chlorophyta associated
¹³⁷ proteins comprised 5.2 ± 0.06 % of all identified proteins in *C. cylindracea*, contributing to $19.2 \pm$
¹³⁸ 1.5 % of all protein abundances.

139 **Discussion**

140 The study of marine macrophyte epiphytic communities using culture independent techniques,
141 such as 16S rRNA analysis, metagenomics and metaproteomics, requires an initial step of biological
142 material isolation. Methods that have been developed for selective isolation of epiphytic biological
143 material can be divided into two groups: (i) procedures involving, prior to extraction, a cell
144 detachment step such as shaking (Gross *et al.*, 2003; Nōges *et al.*, 2010), scraping (Uku *et al.*, 2007)
145 and ultrasonication (Weidner *et al.*, 1996; Cai *et al.*, 2014) and (ii) procedures involving a host
146 tissue incubation aiming at direct lysis of epiphytic microbial cells (Burke *et al.*, 2009). Protocols
147 that include a cell detachment procedure do not provide complete cell removal (Cai *et al.*, 2014;
148 Richter-Heitmann *et al.*, 2016) and in the case of probe ultrasonication tissue disruption can also be
149 observed (Richter-Heitmann *et al.*, 2016). Though, host tissue incubation procedures provide an
150 almost complete cell removal and selective isolation, existing protocols like the one developed for
151 DNA isolation by Burke *et al.* (2009) use in the extraction buffer a rapid multienzyme cleaner (3M)
152 not available worldwide and without a known composition (Burke *et al.*, 2009). To circumvent
153 these problems we developed and tested two protocols for selective DNA and protein isolation from
154 marine macrophyte epiphytic communities.

155 To test if the developed DNA and protein isolation procedures could be applied on a variety
156 of macrophyte species we selected *C. nodosa* and *C. cylindracea* as representatives of seagrass
157 and macroalgal species, on which the procedures were tested. These species especially differ
158 morphologically. While *C. nodosa* leaves are flat, *C. cylindracea* thallus is composed of uneven
159 surfaces (Kuo and den Hartog, 2001; Verlaque *et al.*, 2003). The developed procedures showed an
160 almost complete removal of epiphytic cells from the surfaces of both species comparable to the
161 result of Burke *et al.* (2009) and indicating that structural differences do not impact the removal
162 efficiency. In addition, isolation protocols were tested in two contrasting season as it is known that
163 macrophytes are harboring more algal epiphytes during autumn and winter (Reyes and Sansón,
164 2001). No differences in the removal efficiency was observed between seasons suggesting that

165 protocols can be used on macrophyte samples retrieved throughout the year. Also, no removal
166 differences were observed on samples derived from the same host but from different localities or
167 settlements demonstrating the stability of the protocol in cell removal efficiency.

168 Successful amplification and sequencing of the V4 16S rRNA gene region proved that the
169 isolated DNA can be used to estimate the microbial epiphytic diversity. Taxonomic groups detected
170 in this step can also be often found in epiphytic communities associated with other macrophytes
171 (Burke and Thomas *et al.*, 2011; Morrissey *et al.*, 2019). A problem often encountered in studies
172 focusing on epiphytic communities is the presence of large proportions of chloroplast 16S rRNA
173 sequences in the pool of amplified molecules, especially if the epiphytic DNA was isolated without
174 prior selection (Staufenberger *et al.*, 2008). These sequences can derive from host chloroplasts or
175 from eukaryotic epiphytic chloroplast DNA. Although the proportion of obtained chloroplast 16S
176 rRNA sequences in our samples was substantial they derived almost exclusively from eukaryotic
177 epiphytes. High proportion of chloroplast 16S rRNA sequences in studies applying selective
178 procedures that include direct cellular lysis on host surfaces were observed before (Michelou *et*
179 *al.*, 2013). It is possible that chloroplast specific sequences even in these studies are originating
180 from eukaryotic epiphytic cells and not host chloroplast. Indeed, it is common during 16S rRNA
181 profiling of pelagic microbial communities to observe high proportions of chloroplast sequences
182 (Gilbert *et al.*, 2009; Korlević *et al.*, 2016). In addition, a very low proportion of chloroplast 16S
183 rRNA sequences in samples originating from *C. cylindracea* in comparison to *C. nodosa* could be
184 explained by the presence of three introns in the gene for 16S rRNA in some members of the genus
185 *Caulerpa* that could hamper the amplification process (Lam and Lopez-Bautista, 2016).

186 Beside 16S rRNA sequencing high quality DNA is needed for metagenomics. The obtained
187 number of metagenomic sequences and assembly statistics were comparable to metagenomes and
188 metatranscriptomes derived from similar surface associated communities (Crump *et al.*, 2018; Cúcio
189 *et al.*, 2018). In addition, functional annotation of predicted coding sequences to COG functional
190 categories showed that the obtained metagenomes can be used to determine the metabolic capacity

191 of surface associated communities (Leary *et al.*, 2014; Cúcio *et al.*, 2018). The proportion of
192 coding sequences, including their RPKM, originating from *C. nodosa* metagenomes and classified
193 as Streptophyta was low indicating a good selectivity of the isolation procedure towards epiphytic
194 cells. In the case of DNA samples isolated from the surface of *C. cylindracea* the proportion of
195 Chlorophyta coding sequences was also low but their RPKM was higher than in the case of *C.*
196 *nodosa*. One of the causes for this elevated RPKM of Chlorophyta sequences in *C. cylindracea*
197 could lay in tissue structure differences between these two host species. While *C. nodosa* leaves
198 are composed of individual cells, the thallus of *C. cylindracea* is, like in other siphonous algal
199 species, composed of a single large multinucleate cell (Coneva and Chitwood, 2015). The absence
200 of individual cells in *C. cylindracea* could cause a leakage of genetic material into the extraction
201 buffer causing an elevated presence of host sequences in metagenomic data.

202 To obtain an insight into the metabolic status of uncultivated members, a metaproteomic
203 approach is required (Saito *et al.*, 2019). The applied protocol for epiphytic protein isolation
204 followed by a metaproteomic analysis identified between 14,219 and 16,449 proteins, which is
205 higher than previously reported for e.g. soils (Chourey *et al.*, 2010; Hultman *et al.*, 2015), seawater
206 (Williams *et al.*, 2012) and biofilms (Leary *et al.*, 2014). The functional annotation of identified
207 proteins into COG functional categories showed that the protein isolation procedure can be used to
208 assess the metabolic status of the macrophyte epiphytic community (Leary *et al.*, 2014). Similar
209 to metagenomes, the number of identified proteins, including their abundances, associated with
210 Streptophyta in *C. nodosa* samples were low indicating that the procedure is selective for epiphytic
211 cell proteins. In addition, a higher number of identified proteins, and especially their abundances,
212 associated with Chlorophyta was observed in *C. cylindracea* samples. The cause of this elevated
213 presence of Chlorophyta associated proteins can be, as in the case of the DNA isolation procedure,
214 explained by the absence of individual cells in this siphonous alga (Coneva and Chitwood, 2015).

215 In conclusion, the developed protocols for DNA and protein isolation from macrophyte surfaces
216 are providing an almost complete removal of the epiphytic community and are shown to ensure

217 removal from both, *C. nodosa* and *C. cylindracea*, in different seasons. Also, the obtained DNA
218 and proteins are suitable for 16S rRNA sequencing, metagenomics and metaproteomics, while the
219 obtained material contains low quantities of host DNA or proteins making the procedures epiphyte
220 selective. Furthermore, the procedures are based on universally available laboratory chemicals
221 making the protocols widely applicable.

222 **Experimental procedures**

223 **Sampling**

224 Leaves of *C. nodosa* were sampled in a *C. nodosa* meadow in the Bay of Saline (45°7'5''
225 N, 13°37'20'' E) and in a *C. nodosa* meadow invaded by *C. cylindracea* in the proximity of the
226 village of Funtana (Bay of Funtana; 45°10'39'' N, 13°35'42'' E). Thalli of *C. cylindracea* were
227 sampled in the same *C. nodosa* invaded meadow in the Bay of Funtana and on a locality of
228 only *C. cylindracea* located in the proximity of the invaded meadow. Leaves and thalli for 16S
229 rRNA analysis, metagenomics and metaproteomics were collected in two contrasting seasons, on 4
230 December 2017 (16S rRNA analysis and metaproteomics), 14 December 2017 (metagenomics) and
231 18 June 2018 (16S rRNA analysis, metagenomics and metaproteomics). During spring 2018 the *C.*
232 *nodosa* meadow in the Bay of Saline decayed to an extent that no leaves could be retrieved (Najdek
233 *et al.*, 2020). In addition, as not enough DNA for both metagenomic and 16S RNA analysis were
234 obtained during the sampling on 4 December 2017, an additional sampling on 14 December 2017
235 was carried out in the Bay of Funtana. Leaves and thalli were collected by diving and transported to
236 the laboratory in containers placed on ice and filled with site seawater. Upon arrival to the laboratory,
237 *C. nodosa* leaves were cut into sections of 1 – 2 cm, while *C. cylindracea* thalli were cut into 5 – 8
238 cm long sections. Leaves and thalli were washed three times with sterile artificial seawater (ASW)
239 to remove loosely attached microbial cells.

240 **DNA isolation**

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

245 duplicate isolations were performed. Lysozyme was added (final concentration 1 mg ml⁻¹) and the
246 mixture was incubated at 37 °C for 30 min. Subsequently, proteinase K (final concentration 0.5
247 mg ml⁻¹) and SDS (final concentration 1 %) were added and the samples were incubated at 55 °C
248 for 2 h. Following the incubation, tubes were vortexed for 10 min and the mixture containing lysed
249 epiphytic cells was separated from host leaves or thalli by transferring the solution into a clean tube.
250 The lysate was extracted twice with a mixture of phenol:chloroform:isoamyl alcohol (25:24:1; pH
251 8) and once with chloroform:isoamyl alcohol (24:1). After each organic solvent mixture addition
252 tubes were slightly vortexed and centrifuged at 4,500 × g for 10 min. Following each centrifugation
253 aqueous phases were retrieved. After the final extraction 1/10 of chilled 3 M sodium acetate (pH
254 5.2) was added. DNA was precipitated by adding 1 volume of chilled isopropanol, incubating
255 the mixtures overnight at -20 °C and centrifuging at 16,000 × g and 4 °C for 20 min. The pellet
256 was washed twice with 1 ml of chilled 70 % ethanol and centrifuged after each washing step at
257 20,000 × g and 4 °C for 10 min. After the first washing step duplicate pellets from the same sample
258 were pooled and transferred to a clean 1.5 ml tube. The dried pellet was resuspended in 100 µl of
259 deionized water.

260 Illumina 16S rRNA sequencing

261 An aliquot of isolated DNA was treated with RNase A (final concentration 200 µg ml⁻¹) for 2
262 h at 37 °C. The DNA concentration was determined using the Quant-iT PicoGreen dsDNA Assay
263 Kit (Invitrogen, USA) according to the manufacturer's instructions and diluted to 1 ng µl⁻¹. The V4
264 region of the 16S rRNA gene was amplified using a two-step PCR procedure. In the first PCR the
265 515F (5'-GTGYCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACNVGGGTWTCTAAT-3')
266 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.*,
267 2015; Parada *et al.*, 2016). These primers contained on their 5' end a tagged sequence. Each
268 sample was amplified in four parallel 25 µl reactions of which each contained: 1 × Q5 Reaction
269

270 Buffer , 0.2 mM of dNTPmix, 0.7 mg ml⁻¹ BSA (Bovine Serum Albumin), 0.2 µM of forward
271 and reverse primers, 0.5 U of Q5 High-Fidelity DNA Polymerase (New England Biolabs, USA)
272 and 5 ng of DNA template. Cycling conditions were: initial denaturation at 94 °C for 3 min, 20
273 cycles of denaturation at 94 °C for 45 s, annealing at 50 °C for 60 s and elongation at 72 °C for
274 90 s, finalized by an elongation step at 72 °C for 10 min. The four parallel reactions volumes
275 were pooled and PCR products were purified using the GeneJET PCR Purification Kit (Thermo
276 Fisher Scientific, USA) according to the manufacturer's instructions and following the protocol
277 that included isopropanol addition for better small DNA fragment yield. The column was eluted
278 in 30 µl of deionized water. Purified PCR products were sent for Illumina MiSeq sequencing
279 (2 × 250 bp) at IMGM Laboratories, Martinsried, Germany. Before sequencing at IMGM, the
280 second PCR amplification of the two-step PCR procedure was performed using primers targeting
281 the tagged region incorporated in the first PCR. In addition, these primers contained adapter and
282 sample-specific index sequences. The second PCR was carried out for 8 cycles. Beside samples, a
283 positive and negative control were sequenced. A negative control was comprised of four parallel
284 PCR reactions without DNA template, while for a positive control a mock community composed
285 of evenly mixed DNA material originating from 20 bacterial strains (ATCC MSA-1002, ATCC,
286 USA) was used. Partial 16S rRNA sequences obtained in this study have been deposited in the
287 European Nucleotide Archive (ENA) at EMBL-EBI under accession numbers SAMEA6786270,
288 SAMEA6648792 – SAMEA6648794, SAMEA6648809 – SAMEA6648811.

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

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

303 **Metagenomics**

304 Four DNA samples were selected and sent on dry ice to IMGM Laboratories, Martinsried,
305 Germany for metagenomic sequencing. DNA was purified using AMPure XP Beads (Beckman
306 Coulter, USA) applying a bead:DNA ratio of 1:1 (v/v), quantified with a Qubit dsDNA HS Assay
307 Kit (Thermo Fisher Scientific, USA) and check for integrity on a 1 % agarose gel. Metagenomic
308 sequencing libraries were generated from 300 ng of input DNA using a NEBNext Ultra II FS
309 DNA Library Prep Kit for Illumina (New England Biolabs, USA) according to the manufacturer's
310 instructions. Fragments were selected (500 – 700 bp) using AMPure XP Beads, PCR enriched for 3
311 – 5 cycles and quality controlled. Libraries generated from different DNA samples were pooled and
312 sequenced on an Illumina NovaSeq 6000 sequencing system (2 × 250 bp).

313 Obtained sequences were analyzed on a Life Science Compute Cluster (LiSC) (CUBE –
314 Computational Systems Biology, University of Vienna). Individual sequences were assembled using
315 MEGAHIT (version 1.1.2) (Li *et al.*, 2015) under default settings. Putative genes were predicted
316 from contings longer than 200 bp using Prodigal (version 2.6.3) (Hyatt *et al.*, 2010) in metagenome
317 mode (-p meta). Abundances of predicted genes were expressed as Reads Per Kilobase Million
318 (RPKM) and calculated using the BWA algorithm (version 0.7.16a) (Li and Durbin, 2009). All
319 predicted genes were functionally annotated using the eggNOG-mapper (Huerta-Cepas *et al.*, 2017)
320 and eggNOG database (version 5.0) (Huerta-Cepas *et al.*, 2019). Sequence taxonomy classification
321 was determined using the lowest common ancestor algorithm adapted from DIAMOND (version

322 0.8.36) (Buchfink *et al.*, 2015) and by searching against the NCBI non-redundant database (NR). To
323 determine phylogeny the top 10 % hits with an e-value $< 1 \times 10^{-5}$ were used (--top 10). Sequence
324 renaming, coverage information computing and metagenomic statistics calculations were performed
325 using software tools from BBTools (<https://jgi.doe.gov/data-and-tools/bbtools>). Metagenomic
326 sequences obtained in this study have been deposited in the European Nucleotide Archive (ENA) at
327 EMBL-EBI under accession numbers SAMEA6648795, SAMEA6648797, SAMEA6648809 and
328 SAMEA6648811.

329 Protein isolation

330 Proteins were isolated according to the protocol for protein isolation from soil described in
331 Chourey *et al.* (2010) and modified by Hultman *et al.* (2015). This protocol was further modified
332 and adapted for microbial protein isolation from macrophyte surfaces as described below. 20 ml
333 of protein extraction buffer (4 % SDS, 100 mM Tris-HCl [pH 8.0]) was added to 5 g wet weight
334 of leaves or 10 g wet weight of thalli. The mixture was incubated in boiling water for 5 min,
335 vortexed for 10 min and incubated again in boiling water for 5 min. After a brief vortex the lysate
336 was transferred to a clean tube separating the host leaves or thalli from the mixture containing
337 lysed epiphytic cells. Dithiothreitol (DTT; final concentration 24 mM) was added and proteins were
338 precipitated with chilled 100 % trichloroacetic acid (TCA; final concentration 20 %) overnight at
339 -20°C . Precipitated proteins were centrifuged at $10,000 \times g$ and 4°C for 40 min. The obtained
340 protein pellet was washed three times with chilled acetone. During the first washing step the pellet
341 was transferred to a clean 1.5 ml tube. After each washing step samples were centrifuged at 20,000
342 $\times g$ and 4°C for 5 min. Dried pellets were stored at -80°C until further analysis.

343 **Metaproteomics**

344 Isolated proteins were whole trypsin digested using the FASP (filter-aided sample preparation)
345 Protein Digestion Kit (Expedeon, UK) according to the manufacturer's instructions with small
346 modifications (Wiśniewski *et al.*, 2009). Before the solution was loaded to the column, protein
347 pellets were solubilized in a urea sample buffer included in the kit amended with DTT (final
348 concentration 100 mM) for 45 min at room temperature and centrifuged at 20,000 × g for 2 – 5 min
349 at room temperature to remove larger particles. The first washing step after protein solution loading
350 was repeated twice. In addition, centrifugation steps were prolonged if the column was clogged.
351 Trypsin digestion was performed on column filters at 37 °C overnight for 18 h. The final filtrate
352 containing peptides was acidified with 1 % (final concentration) trifluoroacetic acid, freezed at –80
353 °C for 15 min, lyophilized and sent to the Vienna Metabolomics Center (University of Vienna) for
354 metaproteomic analysis. Peptides were resuspended in 1 % (final concentration) trifluoroacetic
355 acid (TFA), desalted using the Pierce C18 Tips (Thermo Fisher Scientific, USA) according to the
356 manufacturer's instructions and sequenced on a Q Exactive Hybrid Quadrupole-Orbitrap Mass
357 Spectrometer (Thermo Fisher Scientific, USA). Obtained MS/MS spectra were searched against a
358 protein database composed of combined sequenced metagenomes using SEQUEST-HT engines
359 and validated with Percolator in Proteome Discoverer 2.1 (Thermo Fisher Scientific, USA). The
360 target-decoy approach was used to reduce the probability of false peptide identification. Results
361 whose false discovery rate at the peptide level was <1 % were kept. For protein identification a
362 minimum of two peptides and one unique peptide were required. For protein quantification, a
363 chromatographic peak area-based free quantitative method was applied.

364 **Data processing and visualization**

365 Processing and visualization of 16S rRNA, metagenomic and metaproteomic data was done
366 using R (version 3.6.0) (R Core Team, 2019), package tidyverse (version 1.3.0) (Wickham *et*

³⁶⁷ *al.*, 2019) and multiple other packages (Xie, 2014, 2015, 2019, 2020; Neuwirth, 2014; Xie *et*
³⁶⁸ *al.*, 2018; Wilke, 2018; Allaire *et al.*, 2019; Zhu, 2019; Bengtsson, 2020). The detailed analysis
³⁶⁹ procedure including the R Markdown file for this paper are available as a GitHub repository
³⁷⁰ (https://github.com/mkorlevic/Korlevic_SelectiveRemoval_EnvironMicrobiol_2020).

³⁷¹ **Confocal microscopy**

³⁷² Host leaves and thalli from DNA and protein isolation steps were washed seven times in
³⁷³ deionized water and fixed with formaldehyde (final concentration ~ 3 %). In addition, nontreated
³⁷⁴ leaves and thalli, washed three times in ASW to remove loosely attached microbial cells, were fixed
³⁷⁵ in the same concentration of formaldehyde and used as a positive control. For long therm storage,
³⁷⁶ fixed leaves and thalli were immersed in a mixture of phosphate-buffered saline (PBS) and ethanol
³⁷⁷ (1:1) and stored at –20 °C. Treated and untreated segemnts of leaves and thalli were stained in a 2
³⁷⁸ × solution of SYBR Green I and examined under a Leica TCS SP8 X FLIM confocal microscope
³⁷⁹ (Leica Microsystems, Germany).

³⁸⁰ **Acknowledgements**

³⁸¹ This work was founded by the Croatian Science Foundation through the MICRO-SEAGRASS
³⁸² project (IP-2016-06-7118). We would like to thank Margareta Buterer for technical support, Paolo
³⁸³ Paliaga for help during sampling, Lucija Horvat for technical support in confocal microscopy and
³⁸⁴ Dušica Vujaklija for help in accessing the confocal microscope. In addition, we would like to thank
³⁸⁵ the University Computing Center of the University of Zagreb for access to the computer cluster
³⁸⁶ Isabella and CUBE – Computational Systems Biology of the University of Vienna for access to the
³⁸⁷ Life Science Compute Cluster (LiSC).

388 **References**

- 389 Allaire, J.J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., et al. (2019)
- 390 rmarkdown: Dynamic documents for R.
- 391 Amann, R.I., Ludwig, W., and Schleifer, K.H. (1995) Phylogenetic identification and *in situ*
392 detection of individual microbial cells without cultivation. *Microbiol Rev* **59**: 143–169.
- 393 Andersson, A.F., Riemann, L., and Bertilsson, S. (2010) Pyrosequencing reveals contrasting
394 seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *ISME J* **4**: 171–181.
- 395 Apprill, A., McNally, S., Parsons, R., and Weber, L. (2015) Minor revision to V4 region SSU
396 rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat Microb*
397 *Ecol* **75**: 129–137.
- 398 Bengtsson, H. (2020) matrixStats: Functions that apply to rows and columns of matrices (and
399 to vectors).
- 400 Buchfink, B., Xie, C., and Huson, D.H. (2015) Fast and sensitive protein alignment using
401 DIAMOND. *Nat Methods* **12**: 59–60.
- 402 Burke, C., Kjelleberg, S., and Thomas, T. (2009) Selective extraction of bacterial DNA from
403 the surfaces of macroalgae. *Appl Environ Microbiol* **75**: 252–256.
- 404 Burke, C., Steinberg, P., Rusch, D., Kjelleberg, S., and Thomas, T. (2011) Bacterial
405 community assembly based on functional genes rather than species. *Proc Natl Acad Sci U S A* **108**:
406 14288–14293.
- 407 Burke, C., Thomas, T., Lewis, M., Steinberg, P., and Kjelleberg, S. (2011) Composition,
408 uniqueness and variability of the epiphytic bacterial community of the green alga *Ulva australis*.
409 *ISME J* **5**: 590–600.

- 410 Cai, X., Gao, G., Yang, J., Tang, X., Dai, J., Chen, D., and Song, Y. (2014) An ultrasonic
411 method for separation of epiphytic microbes from freshwater submerged macrophytes. *J Basic*
412 *Microbiol* **54**: 758–761.
- 413 Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., et al.
414 (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq
415 platforms. *ISME J* **6**: 1621–1624.
- 416 Chourey, K., Jansson, J., VerBerkmoes, N., Shah, M., Chavarria, K.L., Tom, L.M., et al.
417 (2010) Direct cellular lysis/protein extraction protocol for soil metaproteomics. *J Proteome Res* **9**:
418 6615–6622.
- 419 Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., et al. (2014) Ribosomal
420 Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Res* **42**:
421 D633–D642.
- 422 Coneva, V. and Chitwood, D.H. (2015) Plant architecture without multicellularity: Quandaries
423 over patterning and the soma-germline divide in siphonous algae. *Front Plant Sci* **6**: 287.
- 424 Crump, B.C., Wojahn, J.M., Tomas, F., and Mueller, R.S. (2018) Metatranscriptomics and
425 amplicon sequencing reveal mutualisms in seagrass microbiomes. *Front Microbiol* **9**: 388.
- 426 Cúcio, C., Overmars, L., Engelen, A.H., and Muyzer, G. (2018) Metagenomic analysis shows
427 the presence of bacteria related to free-living forms of sulfur-oxidizing chemolithoautotrophic
428 symbionts in the rhizosphere of the seagrass *Zostera marina*. *Front Mar Sci* **5**: 171.
- 429 Egan, S., Harder, T., Burke, C., Steinberg, P., Kjelleberg, S., and Thomas, T. (2013) The
430 seaweed holobiont: Understanding seaweed-bacteria interactions. *FEMS Microbiol Rev* **37**:
431 462–476.
- 432 Gilbert, J.A., Field, D., Swift, P., Newbold, L., Oliver, A., Smyth, T., et al. (2009) The

433 seasonal structure of microbial communities in the Western English Channel. *Environ Microbiol* **11**:
434 3132–3139.

435 Gross, E.M., Feldbaum, C., and Graf, A. (2003) Epiphyte biomass and elemental composition
436 on submersed macrophytes in shallow eutrophic lakes. *Hydrobiologia* **506-509**: 559–565.

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

440 Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forsslund, S.K., Cook, H.,
441 et al. (2019) eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology
442 resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* **47**: D309–D314.

443 Hultman, J., Waldrop, M.P., Mackelprang, R., David, M.M., McFarland, J., Blazewicz, S.J., et
444 al. (2015) Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature*
445 **521**: 208–212.

446 Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010)
447 Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC*
448 *Bioinformatics* **11**: 119.

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

452 Korlević, M., Šupraha, L., Ljubešić, Z., Henderiks, J., Ciglenečki, I., Dautović, J., and Orlić, S.
453 (2016) Bacterial diversity across a highly stratified ecosystem: A salt-wedge Mediterranean estuary.
454 *Syst Appl Microbiol* **39**: 398–408.

455 Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. (2013)

456 Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon
457 sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* **79**: 5112–5120.

458 Kuo, J. and den Hartog, C. (2001) Seagrass taxonomy and identification key. In *Global*
459 *Seagrass Research Methods*. Short, F.T. and Coles, R.G. (eds). Amsterdam: Elsevier Science B.V.,
460 pp. 31–58.

461 Lam, D.W. and Lopez-Bautista, J.M. (2016) Complete chloroplast genome for *Caulerpa*
462 *racemosa* (Bryopsidales, Chlorophyta) and comparative analyses of siphonous green seaweed
463 plastomes. *Cymbella* **2**: 23–32.

464 Leary, D.H., Li, R.W., Hamdan, L.J., IV, W.J.H., Lebedev, N., Wang, Z., et al. (2014)
465 Integrated metagenomic and metaproteomic analyses of marine biofilm communities. *Biofouling*
466 **30**: 1211–1223.

467 Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015) MEGAHIT: An ultra-fast
468 single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph.
469 *Bioinformatics* **31**: 1674–1676.

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

472 Longford, S., Tujula, N., Crocetti, G., Holmes, A., Holmström, C., Kjelleberg, S., et al. (2007)
473 Comparisons of diversity of bacterial communities associated with three sessile marine eukaryotes.
474 *Aquat Microb Ecol* **48**: 217–229.

475 Massana, R., Murray, A.E., Preston, C.M., and DeLong, E.F. (1997) Vertical distribution and
476 phylogenetic characterization of marine planktonic *Archaea* in the Santa Barbara Channel. *Appl*
477 *Environ Microbiol* **63**: 50–56.

478 Michelou, V.K., Caporaso, J.G., Knight, R., and Palumbi, S.R. (2013) The ecology of microbial

- 479 communities associated with *Macrocystis pyrifera*. *PloS One* **8**: e67480.
- 480 Morrissey, K.L., Çavas, L., Willems, A., and De Clerck, O. (2019) Disentangling the influence
481 of environment, host specificity and thallus differentiation on bacterial communities in siphonous
482 green seaweeds. *Front Microbiol* **10**: 717.
- 483 Najdek, M., Korlević, M., Paliaga, P., Markovski, M., Ivančić, I., Iveša, L., et al. (2020)
484 Dynamics of environmental conditions during the decline of a *Cymodocea nodosa* meadow.
485 *Biogeosciences* **17**: 3299–3315.
- 486 Neuwirth, E. (2014) RColorBrewer: ColorBrewer palettes.
- 487 Nõges, T., Luup, H., and Feldmann, T. (2010) Primary production of aquatic macrophytes and
488 their epiphytes in two shallow lakes (Peipsi and Võrtsjärv) in Estonia. *Aquat Ecol* **44**: 83–92.
- 489 Parada, A.E., Needham, D.M., and Fuhrman, J.A. (2016) Every base matters: Assessing small
490 subunit rRNA primers for marine microbiomes with mock communities, time series and global field
491 samples. *Environ Microbiol* **18**: 1403–1414.
- 492 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013) The SILVA
493 ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic
494 Acids Res* **41**: D590–D596.
- 495 R Core Team (2019) R: A language and environment for statistical computing, Vienna, Austria:
496 R Foundation for Statistical Computing.
- 497 Reyes, J. and Sansón, M. (2001) Biomass and production of the epiphytes on the leaves of
498 *Cymodocea nodosa* in the Canary Islands. *Bot Mar* **44**: 307–313.
- 499 Richter-Heitmann, T., Eickhorst, T., Knauth, S., Friedrich, M.W., and Schmidt, H. (2016)
500 Evaluation of strategies to separate root-associated microbial communities: A crucial choice in
501 rhizobiome research. *Front Microbiol* **7**: 773.

- 502 Saito, M.A., Bertrand, E.M., Duffy, M.E., Gaylord, D.A., Held, N.A., Hervey, W.J., et al.
503 (2019) Progress and challenges in ocean metaproteomics and proposed best practices for data
504 sharing. *J Proteome Res* **18**: 1461–1476.
- 505 Schloss, P.D., Jenior, M.L., Koumpouras, C.C., Westcott, S.L., and Highlander, S.K. (2016)
506 Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. *PeerJ* **4**:
507 e1869.
- 508 Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., et al.
509 (2009) Introducing mothur: Open-source, platform-independent, community-supported software for
510 describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.
- 511 Staufenberger, T., Thiel, V., Wiese, J., and Imhoff, J.F. (2008) Phylogenetic analysis of bacteria
512 associated with *Laminaria saccharina*. *FEMS Microbiol Ecol* **64**: 65–77.
- 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.
- 516 Uku, J., Björk, M., Bergman, B., and Díez, B. (2007) Characterization and comparison of
517 prokaryotic epiphytes associated with three East African seagrasses. *J Phycol* **43**: 768–779.
- 518 Verlaque, M., Durand, C., Huisman, J.M., Boudouresque, C.-F., and Le Parco, Y. (2003) On the
519 identity and origin of the Mediterranean invasive *Caulerpa Racemosa* (Caulerpales, Chlorophyta).
520 *Eur J Phycol* **38**: 325–339.
- 521 Weidner, S., Arnold, W., and Puhler, A. (1996) Diversity of uncultured microorganisms
522 associated with the seagrass *Halophila stipulacea* estimated by restriction fragment length
523 polymorphism analysis of PCR-amplified 16S rRNA genes. *Appl Environ Microbiol* **62**: 766–771.
- 524 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., et al. (2019)

- 525 Welcome to the tidyverse. *J Open Source Softw* **4**: 1686.
- 526 Wilke, C.O. (2018) cowplot: Streamlined plot theme and plot annotations for 'ggplot2'.
- 527 Williams, T.J., Long, E., Evans, F., DeMaere, M.Z., Lauro, F.M., Raftery, M.J., et al. (2012) A
528 metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal
529 surface waters. *ISME J* **6**: 1883–1900.
- 530 Wiśniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample
531 preparation method for proteome analysis. *Nat Methods* **6**: 359–362.
- 532 Xie, Y. (2015) Dynamic Documents with R and knitr, 2nd ed. Boca Raton, Florida: Chapman
533 and Hall/CRC.
- 534 Xie, Y. (2014) knitr: A comprehensive tool for reproducible research in R. In *Implementing
535 Reproducible Computational Research*. Stodden, V., Leisch, F., and Peng, R.D. (eds). New York:
536 Chapman and Hall/CRC, pp. 3–32.
- 537 Xie, Y. (2019) TinyTeX: A lightweight, cross-platform, and easy-to-maintain LaTeX
538 distribution based on TeX Live. *TUGboat* **40**: 30–32.
- 539 Xie, Y. (2020) tinytex: Helper functions to install and maintain 'TeX Live', and compile
540 'LaTeX' documents.
- 541 Xie, Y., Allaire, J.J., and Grolemund, G. (2018) R Markdown: The Definitive Guide, 1st ed.
542 Boca Raton, Florida: Chapman and Hall/CRC.
- 543 Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., et al. (2014) The
544 SILVA and "All-Species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res* **42**:
545 D643–D648.
- 546 Zhu, H. (2019) kableExtra: Construct complex table with 'kable' and pipe syntax.

547 **Figure Captions**

548 **Fig. 1.** Confocal microscope images of *C. nodosa* and *C. cylindracea* surfaces from the Bay of
549 Saline and the Bay of Funtana (Mixed and Monospecific Settlements) sampled on 4 December 2017
550 and stained with SYBR Green I. Scale bar in all images is 60 μ m.

551 **Fig. 2.** Confocal microscope images of *C. nodosa* and *C. cylindracea* surfaces from the Bay of
552 Funtana (Mixed and Monospecific Settlements) sampled on 19 June 2018 and stained with SYBR
553 Green I. Scale bar in all images is 60 μ m.

554 **Fig. 3.** Taxonomic classification and relative contribution of the most abundant (> 1 %) bacterial
555 and archaeal sequences from surfaces of two marine macrophytes (*C. nodosa* and *C. cylindracea*)
556 sampled in the Bay of Saline and the Bay of Funtana (Mixed and Monospecific Settlements) and in
557 two contrasting seasons (4 December 2017 and 19 June 2018).

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

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

568 **Figures**

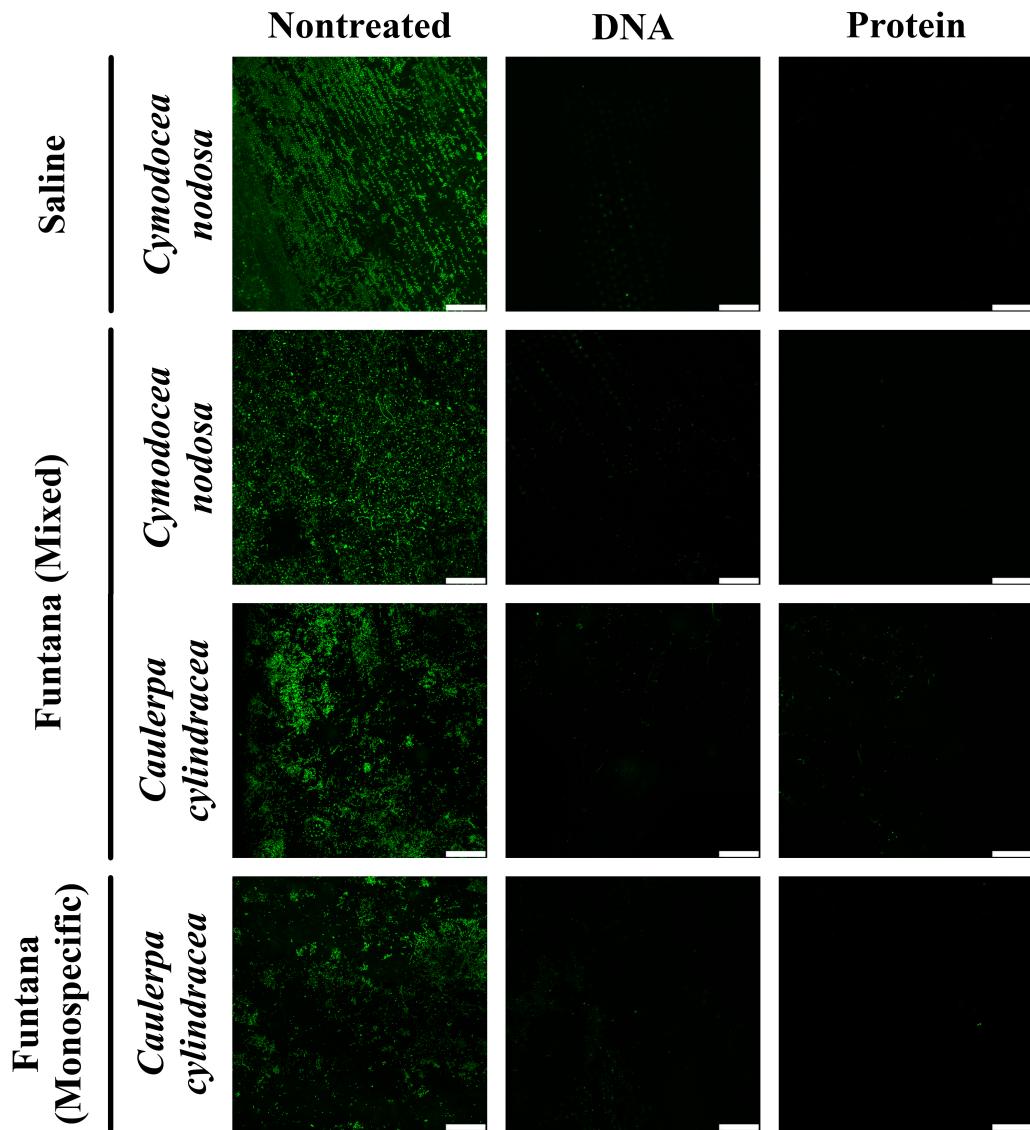


Fig. 1. Confocal microscope images of *C. nodosa* and *C. cylindracea* surfaces from the Bay of Saline and the Bay of Funtana (Mixed and Monospecific Settlements) sampled on 4 December 2017 and stained with SYBR Green I. Scale bar in all images is 60 μ m.

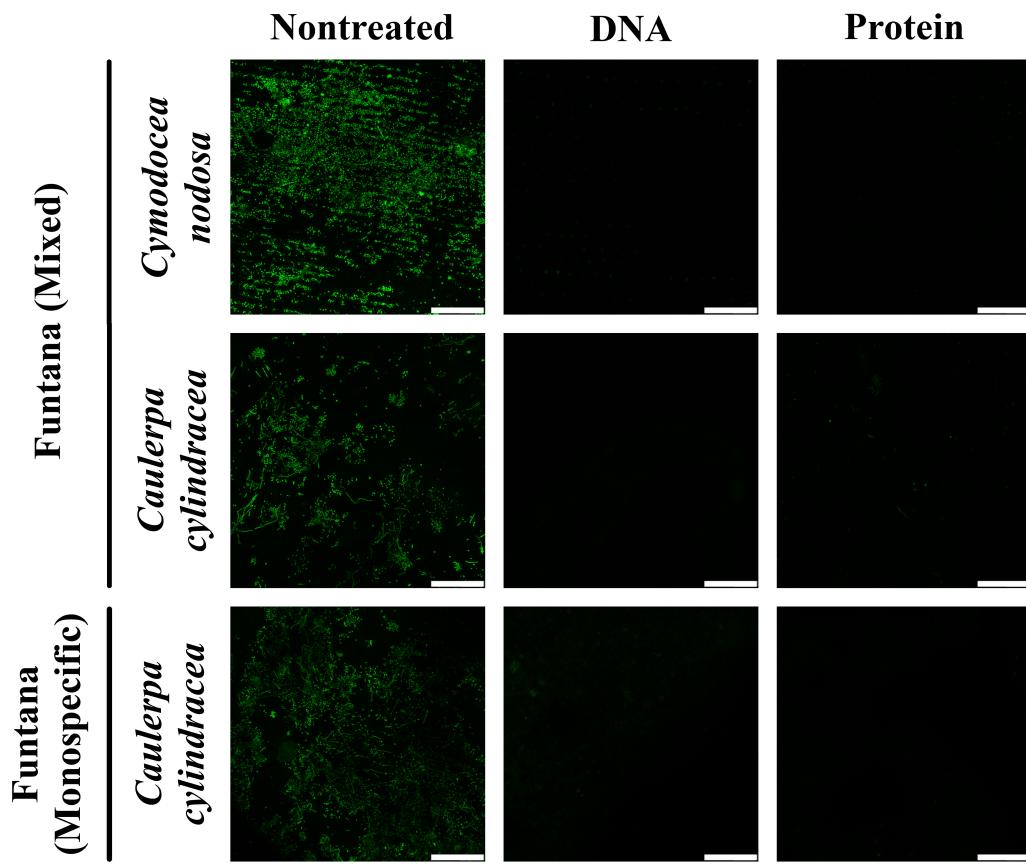


Fig. 2. Confocal microscope images of *C. nodosa* and *C. cylindracea* surfaces from the Bay of Funtana (Mixed and Monospecific Settlements) sampled on 19 June 2018 and stained with SYBR Green I. Scale bar in all images is 60 μ m.

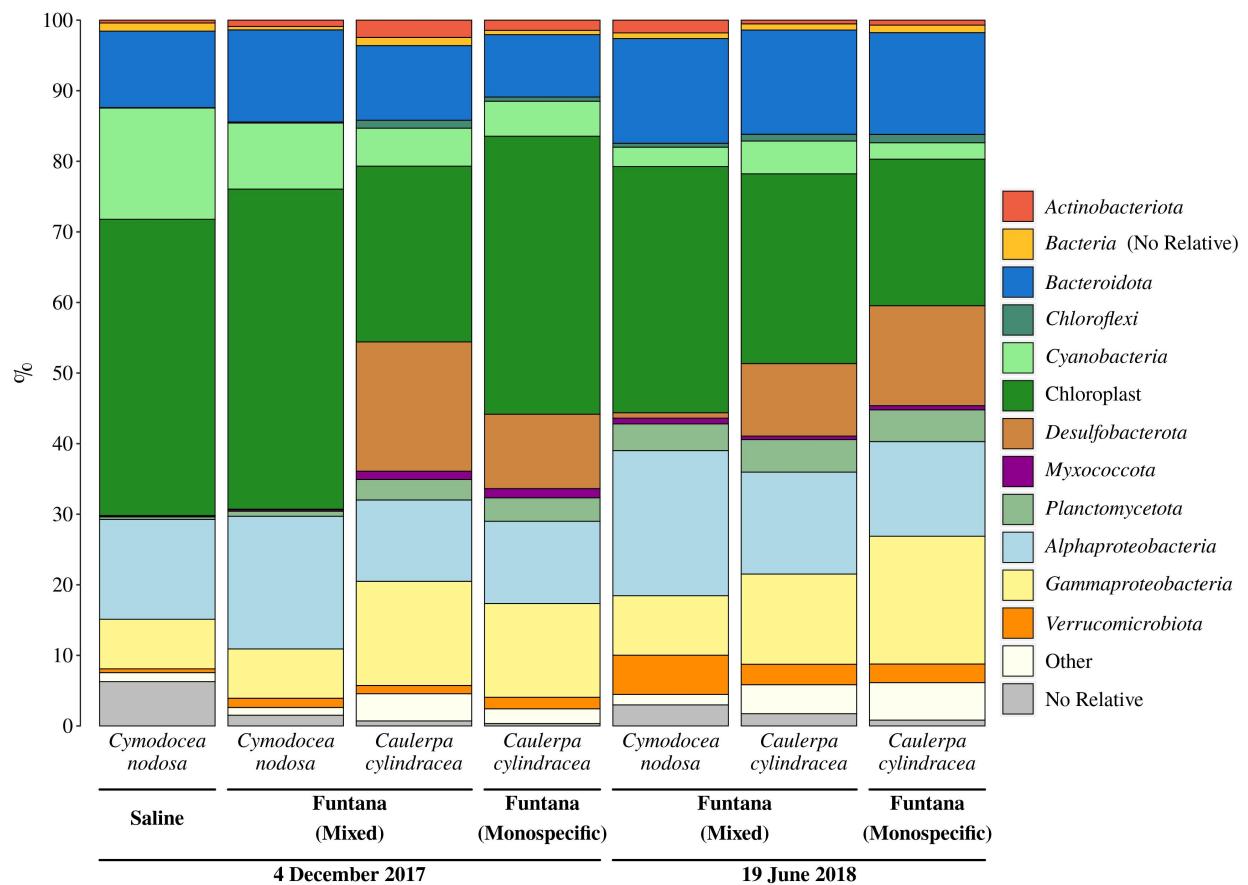


Fig. 3. Taxonomic classification and relative contribution of the most abundant (> 1 %) bacterial and archaeal sequences from surfaces of two marine macrophytes (*C. nodosa* and *C. cylindracea*) sampled in the Bay of Saline and the Bay of Funtana (Mixed and Monospecific Settlements) and in two contrasting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

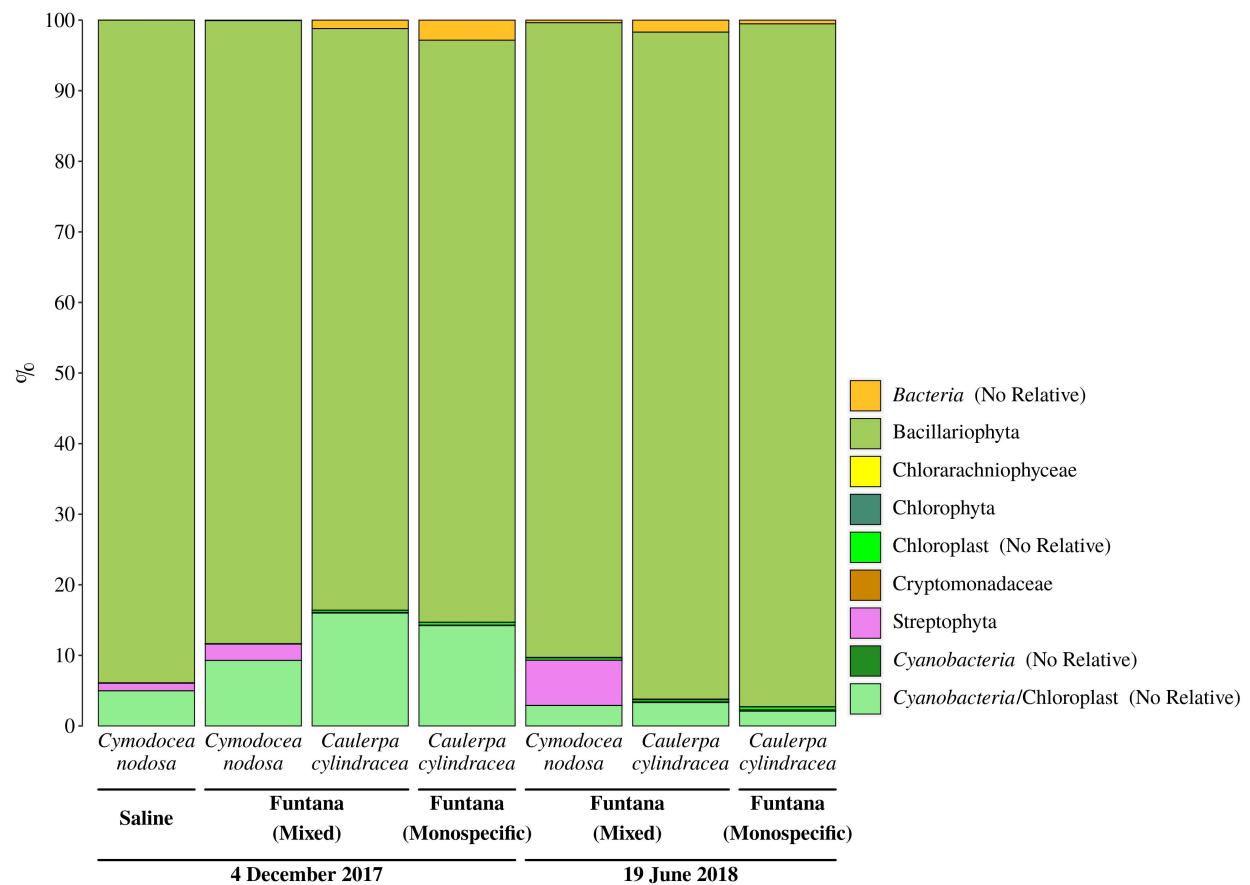


Fig. 4. Taxonomic classification and relative contribution of chloroplast sequences from surfaces of two marine macrophytes (*C. nodosa* and *C. cylindracea*) sampled in the Bay of Saline and the Bay of Funtana (Mixed and Monospecific Settlements) and in two contrasting seasons (4 December 2017 and 19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

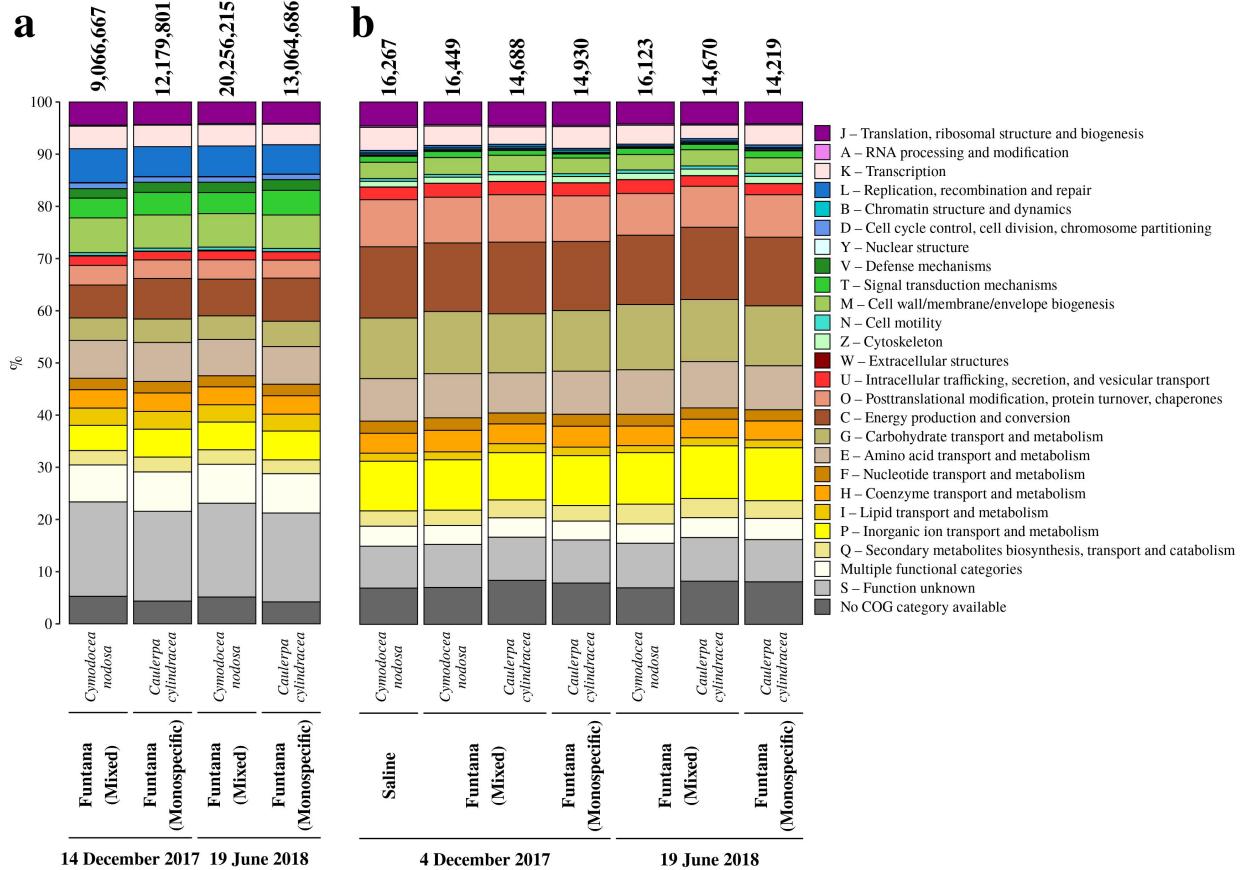


Fig. 5. Relative contribution of each COG category to the total number of annotated coding sequences (A) from metagenomes and identified proteins (B) from metaproteomes associated with surfaces of two marine macrophytes (*C. nodosa* and *C. cylindracea*) sampled in the Bay of Saline and the Bay of Funtana (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.