The genome of *Ectocarpus subulatus* highlights unique mechanisms for stress tolerance in brown algae

5   Simon M. Dittami[1][*], Erwan Corre[2], Loraine Brillet-Guéguen[1,2], Noé Pontoizeau[1,2], Meziane Aite[3],

6   Komlan Avia[1], Christophe Caron[2], Chung Hyun Cho[4], Jonas Collén[1], Alexandre Cormier[1], Ludovic

7   Delage[1], Sylvie Doubleau[5], Clémence Frioux[3], Angélique Gobet[1], Irene González-Navarrete[6], Agnès

8   Groisillier[1], Cécile Hervé[1], Didier Jollivet[7], Hetty KleinJan[1], Catherine Leblanc[1], Agnieszka P. Lipinska[1],

9   Xi Liu[2], Dominique Marie[7], Gabriel V. Markov[1], André E. Minoche[6,8], Misharl Monsoor[2], Pierre

10   Pericard[2], Marie-Mathilde Perrineau[1], Akira F. Peters[9], Anne Siegel[3], Amandine Siméon[1], Camille

11   Trottier[3], Hwan Su Yoon[4], Heinz Himmelbauer[6,8,10], Catherine Boyen[1], Thierry Tonon[1,11]

12   [1] Sorbonne Université, CNRS, Integrative Biology of Marine Models (LBI2M), Station Biologique de

13   Roscoff, 29680 Roscoff, France

14   [2] CNRS, Sorbonne Université, FR2424, ABiMS platform, Station Biologique de Roscoff, 29680, Roscoff,

15   France

16   [3] Institute for Research in IT and Random Systems - IRISA, Université de Rennes 1, France

17   [4] Department of Biological Sciences, Sungkyunkwan University, Suwon 16419, Korea

18   [5] IRD, UMR DIADE, 911 Avenue Agropolis, BP 64501, 34394 Montpellier, France

19   [6] Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr.

20   Aiguader 88, Barcelona, 08003 Spain

21   [7] Sorbonne Université, CNRS, Adaptation and Diversity in the Marine Environment (ADME), Station

22   Biologique de Roscoff (SBR), 29680 Roscoff, France

23   [8] Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany

24   [9] Bezhin Rosko, 40 Rue des Pêcheurs, 29250 Santec, France

25   [10] Department of Biotechnology, University of Natural Resources and Life Sciences (BOKU), Vienna,

26   1190 Vienna, Austria

27   [11] Centre for Novel Agricultural Products, Department of Biology, University of York, Heslington, York,

28   YO10 5DD, United Kingdom.

30   [*] Correspondence: simon.dittami@sb-rocoff.fr, phone +33 29 82 92 362, fax +33 29 82 92 324.

Abstract

Brown algae are multicellular photosynthetic organisms belonging to the stramenopile lineage. They are successful colonizers of marine rocky shores world-wide. The genus *Ectocarpus*, and especially strain Ec32, has been established as a genetic and genomic model for brown algae. A related species, *Ectocarpus subulatus* Kützing, is characterized by its high tolerance of abiotic stress. Here we present the genome and metabolic network of a haploid male strain of *E. subulatus*, establishing it as a comparative model to study the genomic bases of stress tolerance in *Ectocarpus*. Our analyses indicate that *E. subulatus* has separated from *Ectocarpus* sp. Ec32 via allopatric speciation. Since this event, its genome has been shaped by the activity of viruses and large retrotransposons, which in the case of chlorophyll-binding proteins, may be related to the expansion of this gene family. We have identified a number of further genes that we suspect to contribute to stress tolerance in *E. subulatus*, including an expanded family of heat shock proteins, the reduction of genes involved in the production of halogenated defense compounds, and the presence of fewer cell wall polysaccharide-modifying enzymes. However, 96% of genes that differed between the two examined *Ectocarpus* species, as well as 92% of genes under positive selection, were found to be lineage-specific and encode proteins of unknown function. This underlines the uniqueness of brown algae with respect to their stress tolerance mechanisms as well as the significance of establishing *E. subulatus* as a comparative model for future functional studies.

Introduction

51 Brown algae (Phaeophyceae) are multicellular photosynthetic organisms that are successful
52 colonizers of rocky shores of the world's oceans, in particular in temperate and polar regions. In many
53 places they constitute the dominant vegetation in the intertidal zone, where they have adapted to
54 multiple stressors including strong variations in temperature, salinity, irradiation, and mechanical
55 stress (wave action) over the tidal cycle (Davison and Pearson, 1996). In the subtidal environment,
56 brown algae form large kelp forests that harbor highly diverse communities. They are also harvested
57 as food or for industrial purposes, such as the extraction of alginates (McHugh, 2003). The worldwide
58 annual harvest of brown algae has reached 10 million tons by 2014 and is constantly growing (FAO,
59 2016). Brown algae share some basic photosynthetic machinery with land plants, but their plastids
60 derived from a secondary or tertiary endosymbiosis event with a red alga, and they belong to an
61 independent lineage of Eukaryotes, the Stramenopiles (Archibald, 2009). This phylogenetic
62 background, together with their distinct habitat, contributes to the fact that brown algae have
63 evolved numerous unique metabolic pathways, life cycle features, and stress tolerance mechanisms.

64 To enable functional studies of brown algae, strain Ec32 of the small filamentous alga *Ectocarpus* sp.
65 has been established as a genetic and genomic model organism (Peters *et al.*, 2004; Cock *et al.*, 2010;
66 Heesch *et al.*, 2010). This strain was formerly described as *Ectocarpus siliculosus*, but has since been
67 shown to belong to an independent clade by molecular methods (Stache-Crain *et al.*, 1997; Peters *et*
68 *al.*, 2015). More recently two additional brown algal genomes, that of the kelp species *Saccharina*
69 *japonica* (Ye *et al.*, 2015) and that of *Cladosiphon okamuranus* (Nishitsuji *et al.*, 2016), have been
70 characterized. Comparisons between these three genomes have allowed researchers to obtain a first
71 overview of the unique genomic features of brown algae, as well as a glimpse of the genetic diversity
72 within this group. However, given the evolutionary distance between these algae, it is difficult to link
73 genomic differences to physiological differences and possible adaptations to their lifestyle. To be
74 able to generate more accurate hypotheses on the role of particular genes and genomic features for
75 adaptive traits, a common strategy is to compare closely related strains and species that differ only
76 in a few genomic features. The genus *Ectocarpus* is particularly well suited for such comparative
77 studies because it comprises a wide range of morphologically similar but genetically distinct strains
78 and species that have adapted to different marine and brackish water environments (Stache-Crain
79 *et al.*, 1997; Montecinos *et al.*, 2017). One species within this group, *Ectocarpus subulatus* Kützing
80 (Peters *et al.*, 2015) has separated from *Ectocarpus* sp. Ec32 approximately 16 million years ago (Mya;
81 Dittami *et al.*, 2012). It comprises isolates highly resistant to elevated temperature (Bolton, 1983)
82 and low salinity. A strain of this species was even isolated from freshwater (West and Kraft, 1996),
83 constituting one of the handful of known marine-freshwater transitions in brown algae (Dittami *et*
84 *al.*, 2017).

85 Here we present the draft genome and metabolic network of a strain of *E. subulatus*, establishing
86 the genomic basis for its use as a comparative model to study stress tolerance mechanisms, and in
87 particular of low salinity tolerance, in brown algae. Similar strategies have previously been
88 successfully employed in terrestrial plants, where "extremophile" relatives of model- or economically
89 relevant species have been sequenced to explore new stress tolerance mechanisms in the green

3

90 lineage (Oh *et al.*, 2012; Dittami and Tonon, 2012; Dassanayake *et al.*, 2011; Amtmann, 2009; Ma *et*
91 *al.*, 2013; Zeng *et al.*, 2015). The study of the *E. subulatus* genome, and subsequent comparative
92 analysis with other brown algal genomes, in particular that of *Ectocarpus* sp. Ec32, provides insights
93 into the dynamics of *Ectocarpus* genome evolution and divergence, and highlights important
94 adaptive processes, such as a potentially retrotransposon driven expansion of the family of
95 chlorophyll-binding proteins with subsequent diversification. Most importantly, our analyses
96 underline that most of the observed differences between the examined species of *Ectocarpus*
97 correspond to lineage-specific proteins with yet unknown functions.

Results

Sequencing and assembly of the *E. subulatus* genome

100 A total of 34.7 Gb of paired-end read data and of 28.8 Gb of mate pair reads (corresponding to 45
101 million non-redundant mate-pairs) were obtained and used to generate an initial assembly with a
102 total length of 350 Mb, an N50 length of 159 kb, and 8% undefined bases (Ns). However, as
103 sequencing was carried out on DNA from algal material that had not been treated with antibiotics, a
104 substantial part of the assembled scaffolds was of bacterial origin. Removal of these sequences from
105 the final assembly resulted in the final 227 Mb genome assembly with an average GC content of 54%
106 (Table 1). After all cleaning and filtering steps, and considering only algal scaffolds, the average
107 sequencing coverage was 67 X for the pair end library and the genomic coverage (number of unique
108 algal mate pairs * span size / assembly size) was 6.9, 14.4, and 30.4 X for the 3 kb, 5 kb, and 10 kb
109 mate pair libraries, respectively. The bacterial sequences corresponded predominantly to
110 *Alphaproteobacteria* (50%, with the dominant genera *Roseobacter* 8% and *Hyphomonas* 5%)
111 followed by *Gammaproteobacteria* (18%) and *Flavobacteria* (13%). RNA-seq experiments yielded a
112 total of 4.2 Gb of sequence data for a culture of *E. subulatus* Bft15b cultivated in seawater.
113 Furthermore, 4.5 Gb and 4.3 Gb were obtained for two libraries of a freshwater strain of *E. subulatus*
114 from Hopkins River Falls after growth in seawater and in diluted medium, respectively. Of these,
115 96.6% (Bft15b strain in seawater), 87.6% (freshwater strain in seawater), and 85.3% (freshwater
116 strain in diluted medium) were successfully mapped against the final genome assembly of the Bft15b
117 strain.

Gene prediction and annotation

119 Gene prediction was carried out following the protocol employed for *Ectocarpus* sp. Ec32 (Cock *et*
120 *al.*, 2010) using Eugene. The number of predicted proteins was 60% higher than that predicted for
121 Ec32 (Table 1), but this difference can be explained to a large part by the fact that mono-exonic genes
122 (many of which corresponding to transposases) were not removed from our predictions, but were
123 manually removed from the Ec32 genome. This is also coherent with the lower mean number of
124 introns per gene observed in the Bft15b strain. For 10,395 (40 %) of these predicted proteins
125 automatic annotations were generated based on BlastP searches against the Swiss-Prot database;
126 furthermore 724 proteins were manually annotated. The complete set of predicted proteins was
127 used to evaluate the completeness of the genome based on the presence of conserved core

128   eukaryote genes using BUSCO (Simão *et al.*, 2015). This revealed the *E. subulatus* genome to be 86%
129   complete using the full set of conserved eukaryotic genes, and 91% when not considering proteins
130   also absent from all sequenced known brown algae.

Repeated elements

132   Using the REPET pipeline, we determined that, similar to results obtained for strain Ec32, the *E.*
133   *subulatus* genome consisted of 30% repeated elements, *i.e.* 10% less than *S. japonica*. The most
134   abundant groups of repeated elements were large retrotransposon derivatives (LARDs), followed by
135   long terminal repeats (LTRs, predominantly Copia and Gypsy), and long and short interspersed
136   nuclear elements (LINEs). The overall distribution of sequence identity levels within superfamilies
137   showed two peaks, one at an identity level of 78-80%, and one at 96-100% (Figure 1), indicating two
138   periods of high transposon activity in the past. Terminal repeat retrotransposons in miniature (TRIM)
139   and LARDs, both non-autonomous groups of retrotransposons, were among the most conserved
140   families (Figure 1B). In line with previous observations carried out in *Ectocarpus* sp. Ec32, no
141   methylation was detected in the *E. subulatus* genomic DNA, an indication that methylation was most
142   likely not a mechanism to silence transposons in this species.

Organellar genomes

144   Plastid and mitochondrial genomes from *E. subulatus* have 95.5% and 91.5% sequence identity with
145   their *Ectocarpus* sp. Ec32 counterparts, respectively, in the conserved regions (Figure 2). The
146   mitochondrial genome of *E. subulatus* differed from that of *Ectocarpus* sp. Ec32 essentially with
147   respect to the presence of three additional maturase genes, as well as one and two introns within
148   the 16S and 23S rRNA genes, respectively. A large structural difference was observed only in the
149   plastid genome where one inversion of *ca.* 50 kb in the small single copy (SSC) region may have
150   occurred. Furthermore, small differences in gene contents of the *E. subulatus* plastid with respect to
151   *Ectocarpus* sp. Ec32 were detected around two inverted repeat (IR) regions concerning the following
152   genes: psbC (gene truncated), psbD (IR region next to gene), rpoB (large gap, frameshift), and tRNA-
153   Arg and tRNA-Glu (duplicated in the tRNA region). Pseudogenization of genes at the edge of IRs is
154   indeed a common phenomenon (Lee *et al.*, 2016).

Global comparison of predicted proteomes

GO-based comparisons

157   OrthoFinder was used to define clusters of predicted orthologs as well as species-specific proteins.
158   As shown in Figure 3, 11,177 predicted Bft15b proteins had no ortholog in Ec32, while the reverse
159   was true for only 3,605 proteins of strain Ec32. Furthermore, among the clusters of genes, we
160   observed differences in copy number for several of the proteins between the two species. Using gene
161   set enrichment analyses, we attempted to automatically identify functional groups of genes that
162   were over-represented either among the proteins specific to one or the other genome, or that were
163   expanded in one of the two genomes. The results of these analyses point towards several functional
164   groups of proteins that were subject to recent variations between *E. subulatus* and *Ectocarpus* sp.
165   Ec32 (Figure 3). Categories identified as over-represented among the genes unique to *E. subulatus*
166   include DNA integration, chlorophyll binding, and DNA binding, but also false positives such as red

5

167 light signaling, which arise from the presence of transposable elements in the genome (see
168 Supporting Information File S1). However, no significantly enriched GO terms were found among
169 protein families expanded in the *E. subulatus* genome. In contrast, several categories were over-
170 represented among the genes and gene families specific to or expanded in the *Ectocarpus* sp. Ec32
171 strain, many of which were related either to signaling pathways or to the membrane and transporters
172 (Figure 3), although differences with respect to membrane and transporters were not confirmed
173 after manual curation.

Domain-based comparisons

175 Domain-based comparisons were carried out to avoid a possible impact of moderate or poor-quality
176 annotations on the genomic comparisons. In total, 5,728 different InterPro domains were detected
177 in both *Ectocarpus* genomes, with 133,448 and 133,052 instances in *E. subulatus* Bft15b and
178 *Ectocarpus* sp. Ec32 strains respectively. The most common domains in *E. subulatus* were Zinc finger,
179 CCHC-type (IPR001878, 3,861 instances), and Ribonuclease H-like (IPR012337, 3,742 instances). Both
180 were present less than 200 times in Ec32. The most common domains in *Ectocarpus* sp. Ec32 were
181 the ankyrin repeat and ankyrin repeat-containing domains (IPR002110, IPR020683: 4,138 and 4,062
182 occurrences vs *ca.* 3,000 in Bft15b). Two hundred and ninety-six domains were specific to Bft15b,
183 while 582 were specific to Ec32 (see Supporting Information Table S2).

Metabolic network-based comparisons

185 In total, the *E. subulatus* metabolic network reconstruction comprised 2,445 genes associated with
186 2,074 metabolic reactions and 2,173 metabolites in 464 pathways, 259 of which were complete
187 (Figure 3). These results are similar to data previously obtained for *Ectocarpus* sp. Ec32 (Prigent *et*
188 *al.*, 2014; see http://gem-aureme.irisa.fr/ectogem for the most recent version; 1,977 reactions,
189 2,132 metabolites, 2,281 genes, 459 pathways, 272 complete pathways). Comparisons between both
190 networks were carried out on a pathway level (Supporting Information Table S3), focusing on
191 pathways present (*i.e.* complete to more than 50%) in one of the species, but with no reactions in
192 the other. This led to the identification of 16 pathways potentially specific to *E. subulatus* Bft15b,
193 and 11 specific to *Ectocarpus* sp. Ec32, which were further manually investigated. In all of the
194 examined cases, the observed differences were due to protein annotation, but not due to the
195 presence/absence of proteins associated with these pathways in both species. For instance, the
196 pathways "spermine and spermidine degradation III" (PWY-6441) was only found in *E. subulatus*
197 because the corresponding genes had been manually annotated in this species, while this was not
198 the case in *Ectocarpus* sp. Ec32. On the other hand, three pathways related to methanogenesis (PWY-
199 5247, PWY-5248, and PWY-5250) were falsely included in the metabolic network of *E. subulatus* due
200 to an overly precise automatic GO annotation of the gene Bft140_7. All in all, based on our network
201 comparisons, we confirmed no differences regarding the presence or absence of known metabolic
202 pathways in the two examined species of *Ectocarpus*.

Genes under positive selection

204 In total, 7,147 pairs of orthologs were considered to search for genes under positive selection
205 between the two examined strains of *Ectocarpus*, and we identified 83 gene pairs (1.2%) that
206 exhibited dN/dS ratios > 1 (Supporting Information Table S4). This proportion was low compared to

207 the 12% of genes under positive selection found in a study comprising also kelp and diatom species
208 (Teng *et al.*, 2017). Note however, that our analysis focused on the global dN/dS ratio per gene,
209 rather than the local dN/dS ratio per codon site (implemented in codeml, PAML) used by Teng *et al.*
210 (2017). The gene pairs under positive selection may be related to the adaptation to the different
211 environmental niches occupied by the strains investigated. These gene pairs were examined
212 manually, but only one of them (Ec-11_002330, EsuBft305_15) could be assigned a function, *i.e.* a
213 putative mannosyl-oligosaccharide 1,2-alpha-mannosidase activity, possibly involved in glycoprotein
214 modification. Twelve additional pairs contained known protein domains (two Zinc finger domains,
215 one TIP49 domain, one DnaJ domain, one NADH-ubiquinone oxidoreductase domain, one SWAP
216 domain, and six ankyrin repeat domains). Ankyrin repeat domains were significantly over-
217 represented among the genes under positive selection (p < 0.05, Fisher exact test), and the
218 corresponding genes were manually examined by best reciprocal blast search to ensure that they
219 corresponded to true orthologs. Only one pair was part of a protein family that had undergone recent
220 expansion (Ec-27_003170, EsuBft1157_2), and in this case phylogenetic analysis including the other
221 members of the family (EsuBft255_4, EsuBft2264_2, Ec-05_004510, Ec-08_002010) showed Ec-
222 27_003170 and EsuBft1157_2 to form a branch with 100% bootstrap support (data not shown). The
223 remaining 70 pairs of proteins had entirely unknown functions, although four genes were located in
224 the pseudoautosomal region of the sex chromosome of *Ectocarpus* sp. Ec32. Out of the 83 genes
225 under positive selection 72 were found only in brown algae and another four only in stramenopiles
226 (e-value cutoff of 1e-10 against the nr database). They can thus be considered as taxonomically
227 restricted genes. Furthermore, 75 of these genes were expressed in at least one of the two
228 *Ectocarpus* species, and only 10 of the 83 genes encoded short proteins with less than 100 amino
229 acid residues, suggesting that the majority of these genes may be functional. None of them were
230 highly variable, as indicated by the fact that the dN/dS ratio exhibited a weak negative correlation
231 with the rate of synonymous mutations dS (Pearson Correlation coefficient r=-0.05, p < 0.001; Figure
232 4). This suggests that the split of *Ectocarpus* sp. Ec32 and *E. subulatus* was the result of allopatric
233 separation with subsequent speciation due to gradual adaptation to the local environment. Indeed,
234 in cases of sympatric or parapatric speciation, genes under positive selection are predominant
235 among rapidly evolving genes (Swanson and Vacquier, 2002). There was no trend for positively
236 selected genes to be located in specific regions of the genome (dispersion index of genes under
237 positive selection close to a random distribution with values ranging between 0.7 and 0.8 depending
238 on the window size).

239 **Manual examination of lineage-specific and of expanded genes and gene families**
240 The focus of our work is on the genes specific to and expanded in *E. subulatus* and we only give a
241 brief overview of the situation regarding *Ectocarpus* sp. Ec32. It is important to consider that the *E.*
242 *subulatus* Bft15b genome is likely to be less complete than the Ec32 genome, which has been curated
243 and improved for over 10 years now (Cormier *et al.*, 2017). Hence, regarding genes that are present
244 in Ec32 but absent in Bft15b, it is difficult to distinguish between the effects of a potentially
245 incomplete genome assembly and true gene losses in Bft15b. To further reduce this bias during the
246 manual examination of lineage-specific genes, the list of genes to be examined was reduced by
247 additional restrictions. First, only genes that did not have orthologs in *S. japonica* were considered.

7

248 This eliminated several predicted proteins that may have appeared to be lineage-specific due to
249 incomplete genome sequencing, but also proteins that have been recently lost in one of the
250 *Ectocarpus* species. Secondly, the effect of possible differences in gene prediction, notably the
251 manual removal of monoexonic gene models in *Ectocarpus* sp. Ec32, was minimized by including an
252 additional validation step: only proteins without corresponding nucleotide sequences (tblastn, e-
253 value < 1e-10) in the other *Ectocarpus* genome were considered for manual examination. Thirdly,
254 only proteins with a length of at least 50 aa were retained. This reduced the number of lineage-
255 specific proteins to be considered in strain Bft15b to 1,629, and in strain Ec32 to 689 (Supporting
256 Information Table S5).

257 In *E. subulatus*, among the 1,629 lineage-specific genes, 1,436 genes had no homologs (e-value < 1e-
258 5) in the UniProt database: they are thus truly lineage-specific and have unknown functions. Among
259 the remaining 193 genes, 145 had hits (e-value < 1e-5) in *Ectocarpus* sp. Ec32. The majority
260 corresponds to multi-copy genes that had diverged prior to the separation of *Ectocarpus* and *S.
261 japonica*, and for which the *Ectocarpus* sp. Ec32 and *S. japonica* orthologs were probably lost. The
262 remaining 48 genes were manually examined (genetic context, GC content, EST coverage); 18 of
263 them corresponded to probable bacterial contaminations and the corresponding scaffolds were
264 removed. Finally, the remaining 30 genes were manually annotated and classified: 13 had homology
265 only with uncharacterized proteins or were too dissimilar from characterized proteins to deduce
266 hypothetical functions; another eight probably corresponded to short viral sequences integrated into
267 the algal genome (EsuBft1730_2, EsuBft4066_3, EsuBft4066_2, EsuBft284_15, EsuBft43_11,
268 EsuBft551_12, EsuBft1883_2, EsuBft4066_4), and one (EsuBft543_9) was related to a
269 retrotransposon. Two adjacent genes (EsuBft1157_4, EsuBft1157_5) were also found in diatoms and
270 may be related to the degradation of cellobiose and the transport of the corresponding sugars.
271 Furthermore, two genes, EsuBft1440_3 and EsuBft1337_8, contained conserved motifs (IPR023307
272 and SSF56973) typically found in toxin families. Finally, two additional proteins, EsuBft36_20 and
273 EsuBft440_20, consisted almost exclusively of short repeated sequences of unknown function
274 ("ALEW" and "GAAASGVAGGAVVVNG", respectively).

275 In *Ectocarpus* sp. Ec32, 97 proteins corresponded to the *E. siliculosus* virus-1 inserted into the Ec32
276 genome – no similar insertion was detected in *E. subulatus*. The large majority of proteins (511)
277 corresponded to proteins of unknown function without matches in public databases. The remaining
278 81 proteins were generally poorly annotated, usually only via the presence of a domain. Examples
279 are ankyrin repeat-containing domain proteins (12), Zinc finger domain proteins (6), proteins
280 containing wall sensing component (WSC) domains (3), protein kinase-like proteins (3), and Notch
281 domain proteins (2) (see Supporting Information Table S5).

282 Regarding expanded gene families, OrthoFinder indicated 232 clusters of orthologous genes
283 (corresponding to 4,064 proteins) expanded in the genome of *E. subulatus*, and 450 expanded in
284 *Ectocarpus* sp. Ec32 (corresponding to 1,685 proteins; Supporting Information Table S5). Manual
285 examination of the *E. subulatus* expanded gene clusters revealed 48 of them (2,623 proteins) to be
286 false positives, which can be explained essentially by split gene models or gene models associated
287 with transposable elements predicted in the *E. subulatus* but not in the *Ectocarpus* sp. Ec32 genome.

288 The remaining 184 clusters (1,441 proteins) corresponded to proteins with unknown function (139
289 clusters, 1,064 proteins), 98% of which were found only in both *Ectocarpus* genomes. Furthermore,
290 nine clusters (202 proteins) represented sequences related to transposons predicted in both
291 genomes, and eight clusters (31 proteins) were similar to known viral sequences. Only 28 clusters
292 (135 proteins) could be roughly assigned to biological functions (Table 2). They comprised proteins
293 potentially involved in modification of the cell-wall structure (including sulfation), in transcriptional
294 regulation and translation, in cell-cell communication and signaling, as well as a few stress response
295 proteins, notably a set of HSP20s, and several proteins of the light-harvesting complex (LHC)
296 potentially involved in non-photochemical quenching.

297 Among the most striking examples of expansion in *Ectocarpus* sp. Ec32, we found different families
298 of serine-threonine protein kinase domain proteins present in 16 to 25 copies in Ec32 compared to
299 only 5 or 6 (numbers of different families) in *E. subulatus*, Kinesin light chain-like proteins (34 vs. 13
300 copies), two clusters of Notch region containing proteins (11 and 8 vs. 2 and 1 copies), a family of
301 unknown WSC domain containing proteins (8 copies vs. 1), putative regulators of G-protein signaling
302 (11 vs. 4 copies), as well as several expanded clusters of unknown and of viral proteins.

Targeted manual annotation of specific pathways

304 Based on the results of automatic analysis but also on literature studies of genes that may be able to
305 explain physiological differences between *E. subulatus* and *Ectocarpus* sp. Ec32, several gene families
306 and pathways were manually examined and annotated.

Cell wall metabolism

308 Cell walls are key components of both plants and algae and, as a first barrier to the surrounding
309 environment, important for many processes including development and the acclimation to
310 environmental changes. Synthesis and degradation of cell wall oligo- and polysaccharides is
311 facilitated by carbohydrate-active enzymes (CAZymes) (http://www.cazy.org/; Cantarel et al. 2009).
312 These comprise several families including glycoside hydrolases (GHs) and polysaccharide lyases (PLs),
313 both involved in the cleavage of glycosidic linkages, glycosyltransferases (GTs), which create
314 glycosidic linkages, and additional enzymes such as the carbohydrate esterases (CEs) which remove
315 methyl or acetyl groups from substituted polysaccharides.
316 The genome of the brown alga *E. subulatus* encodes 37 GHs (belonging to 17 GH families), 94 GTs
317 (belonging to 28 GT families), nine sulfatases (family S1-2), and 13 sulfotransferases, but lacks genes
318 homologous to known PLs and CEs (Figure 5). In particular, the consistent lack of known alginate
319 lyases and cellulases in the *E. subulatus* and the other brown algal genomes suggests that other, yet
320 unknown genes, may be responsible for cell wall modifications during development. Overall, the
321 gene content of *E. subulatus* is similar to *Ectocarpus* sp. Ec32 and *S. japonica* in terms of the number
322 of CAZY families, but slightly lower in terms of absolute gene number (Cock, *et al.* 2010; Ye *et al.*
323 2015; Figure 5). Especially *S. japonica* features an expansion of certain CAZY families probably related
324 to the establishment of more complex tissues in this kelp (*i.e.* 82 GHs belonging to 17 GH families,
325 131 GTs belonging to 31 GT families).
326 *E. subulatus* is frequently found in brackish- and even freshwater environments (West and Kraft,
327 1996) where its cell wall exhibits little or no sulfation (Torode *et al.*, 2015). Hence, we also assessed

9

328 whether *E. subulatus* had reduced the gene families responsible for this process. Its genome encodes
329 only eight sulfatases and six sulfotransferases compared to ten and seven, respectively, in *Ectocarpus*
330 sp. Ec32. We also documented variations in the GT families, some being present in one or two of the
331 brown algal genomes considered, while absent in other(s) (*e.g.* GH30, GT15, GT18, GT24, GT25,
332 GT28, GT50, GT54, GT65, GT66, GT74, GT77). However, as gene numbers for these families are very
333 low (*e.g.* the GT24 family has one member in *Ectocarpus* sp. Ec32, two in *E. subulatus*, and none in *S.*
334 *japonica*), the results must be taken with caution. Finally, *Ectocarpus* sp. Ec32 has previously been
335 reported to possess numerous proteins with WSC domains (Cock *et al.*, 2010; Michel *et al.*, 2010).
336 These were initially found in yeasts (Verna *et al.*, 1997) where they act as cell surface
337 mechanosensors and activate the intracellular cell wall integrity signaling cascade in response to
338 hypo-osmotic shock (Gualtieri *et al.*, 2004). In brown algae, these WSC domains may also regulate
339 wall rigidity, through the control of the activity of appended enzymes, such as mannuronan C5-
340 epimerases, which act on alginates (Hervé *et al.*, 2016). Surprisingly, the total number of WSC
341 domains is reduced in *E. subulatus* compared to *Ectocarpus* sp. Ec32 with around 320 vs. 444
342 domains, respectively, based on InterProScan (Supporting Information Table S2). Additional
343 information regarding *E. subulatus* CAZYmes can be found in Supporting Information File S1.

Central and storage carbohydrate metabolism

345 A characteristic feature of brown algae is that they store carbohydrates not as glycogen or starch,
346 like most animals and plants, but as laminarin (Read *et al.*, 1996). Brown algae also have the
347 particularity of using the photoassimilate D-fructose 6-phosphate to produce the alcohol sugar D-
348 mannitol instead of sucrose like land plants. The *E. subulatus* genome contains similar sets of genes
349 for carbon storage compared to *Ectocarpus* sp. Ec32: all the genes encoding enzymes involved in
350 sucrose metabolism and starch biosynthesis are completely absent while all genes necessary for
351 trehalose synthesis, as well as laminarin synthesis and recycling were found. Also, three copies of
352 M1PDH genes were found in both *Ectocarpus* species compared to two in *S. japonica*, probably due
353 to a recent duplication of M1PDH1/M1PDH2 in the Ectocarpales (Tonon *et al.*, 2017) (Supporting
354 Information File S1).

Sterol metabolism

356 Sterols are important modulators of membrane fluidity among eukaryotes, and provide the
357 backbone for signaling molecules (Desmond and Gribaldo, 2009). Fucosterol, cholesterol, and
358 ergosterol are the most abundant sterols in *Ectocarpus* sp. Ec32, where their relative abundance
359 varies according to sex and temperature (Mikami *et al.*, 2018). All three molecules are thought to be
360 synthesized from squalene by a succession of 12 to 14 steps, relying on a roughly conserved set of
361 twelve enzymes (Desmond and Gribaldo, 2009). The *E. subulatus* and *Ectocarpus* sp. Ec32 genomes
362 each encode homologs of twelve of them (SQE, CAS, CYP51, FK, SMO, HSD3B, EBP, CPI1, DHCR7,
363 SC5DL, and two SMTs). The remaining two, a delta-24-reductase (DHCR24) and a C22 desaturase
364 (CYP710), were probably lost secondarily. In land plants, these latter enzymes are involved in the two
365 steps transforming fucosterol into stigmasterol. Fucosterol is the main sterol in brown algae, and
366 provides a substrate for saringosterol, a brown-alga specific C24-hydroxylated fucosterol-derivative
367 with antibacterial activity (Wächter *et al.*, 2001).

Algal defense: metabolism of phenolics and halogens

Polyphenols are a group of defense compounds in brown algae that are likely to be important both for abiotic (Pavia *et al.*, 1997) and biotic stress tolerance (Geiselman and McConnell, 1981). Brown algae produce specific polyphenols called phlorotannins, which are analogous to land plant tannins. These products are polymers of phloroglucinol, which are synthesized via the activity of a phloroglucinol synthase, a type III polyketide synthase characterized in *Ectocarpus* sp. Ec32 (Meslet-Cladière *et al.*, 2013). In analogy to the flavonoid pathway of land plants, the further metabolism of phlorotannins is thought to be driven by members of chalcone isomerase-like (CHIL), aryl sulfotransferase (AST), flavonoid glucosyltransferase (FGT), flavonoid O-methyltransferase (OMT), polyphenol oxidase (POX), and tyrosinase (TYR) families (Cock *et al.*, 2010). While copy numbers between the two *Ectocarpus* species and *S. japonica* are identical for PKS III, CHIL, FGT, OMT and POX, *E. subulatus* encodes fewer ASTs and TYRs (Figure 5). In the case of ASTs, this may be related to the lower concentration of sulfate in low salinity environments frequently colonized by *E. subulatus*.

A second important and original defense mechanism in brown algae is the production of halogenated compounds via the activity of halogenating enzymes, *e.g.* the vanadium-dependent haloperoxidase (vHPO). While *S. japonica* has recently been reported to possess 17 potential bromoperoxidases (vBPO) and 59 putative iodoperoxidases (vIPO) (Ye *et al.*, 2015), *Ectocarpus* sp. Ec32 and *E. subulatus* possess only a single vBPO each and no vIPO, but have in turn slightly expanded a haloperoxidase family closer to vHPO characterized in several marine bacteria (Fournier *et al.*, 2014) (Figure 5). One difference between the two *Ectocarpus* species is that *E. subulatus* Bft15b possesses only three vHPO genes compared to the five copies found in the genome of Ec32. In addition, homologs of thyroid peroxidases (TPOs) may also be involved in halide transfer and stress response. Again, Ec32 and Bft15b show a reduced set of these genes compared to *S. japonica*, and Ec32 contains more copies than Bft15b. Finally, a single haloalkane dehalogenase (HLD) was found exclusively in *Ectocarpus* sp. Ec32.

Transporters

Transporters are key actors driving salinity tolerance in terrestrial plants (Volkov, 2015). We therefore carefully assessed potential differences in this group of proteins that may explain physiological differences between Ec32 and Bft15b based on the five main categories of transporters described in the Transporter Classification Database (TCDB) (Saier *et al.*, 2016): channels/pores, electrochemical potential-driven transporters, primary active transporters, group translocators, and transmembrane electron carriers. A total of 292 genes were identified in *E. subulatus* (Supporting Information Table S1). They consist mainly of transporters belonging to the three first categories listed above. All 27 annotated transporters of the channels/pores category belong to the alpha-type channel (1.A.) and are likely to be involved in movements of solutes by energy-independent processes. One hundred and forty-five proteins were found to correspond to the second category (electrochemical potential-driven transporters) containing transporters using a carrier-mediated process to catalyze uniport, antiport, or symport. The most represented superfamilies are APC (Amino Acid-Polyamine-Organocation, 24), DMT (Drug/Metabolite Transporter, 16), MFS (Major Facilitator Superfamily, 32), and MC (Mitochondrial Carrier, 34). Primary active transporters (third

11

408 category) use a primary source of energy to drive the active transport of a solute against a
409 concentration gradient. Eighty proteins representing this category were found in the *E. subulatus*
410 genome, including 59 ABC transporters and 15 belonging to the P-type ATPase superfamily. No
411 homologs of group translocators or transmembrane electron carriers were identified, but 14
412 transporters were classified as category 9, which is poorly characterized. A 1:1 ratio of orthologous
413 genes coding for all of the transporters described above was observed between both *Ectocarpus*
414 genomes, except for EsuBft583_3, an anion-transporting ATPase, which is also present in diatoms
415 and *S. japonica*, but may have been recently lost in *Ectocarpus* sp. Ec32.

Abiotic stress-related genes

417 Reactive oxygen species (ROS) scavenging enzymes, including ascorbate peroxidases, superoxide
418 dismutases, catalases, catalase peroxidases, glutathione reductases, (mono)dehydroascorbate
419 reductases, and glutathione peroxidases are important for the redox equilibrium of organisms (see
420 Das and Roychoudhury 2014 for a review). An increased reactive oxygen scavenging capacity has
421 been correlated with stress tolerance in brown algae (Collén and Davison, 1999). In the same vein,
422 chaperone proteins including heat shock proteins (HSPs), calnexin, calreticulin, T-complex proteins,
423 and tubulin-folding co-factors are important for protein re-folding under stress. The transcription of
424 these genes is very dynamic and generally increases in response to stress in brown algae (Roeder *et*
425 *al.*, 2005; Mota *et al.*, 2015). In total, 104 genes encoding members of the protein families listed
426 above were manually annotated in the *E. subulatus* Bft15b genome (Supporting Information Table
427 1). However, with the exception of HSP20 proteins which were present in three copies in Bft15b vs.
428 one copy in Ec32 and had already been identified in the automatic analysis, no clear difference in
429 gene number was observed between the two *Ectocarpus* species.

430 Different families of chlorophyll-binding proteins (CBPs), such as the LI818/LHCX family, have been
431 suspected to be involved in non-photochemical quenching (Peers *et al.*, 2009). CBPs have been
432 reported to be up-regulated in response to abiotic stress in stramenopiles (*e.g.* Zhu and Green 2010;
433 Dong *et al.* 2016), including *Ectocarpus* (Dittami *et al.*, 2009), probably as a way to deal with excess
434 light energy when photosynthesis is affected. They have also previously been shown to be among
435 the most variable functional groups of genes between *Ectocarpus* sp. Ec32 and *E. subulatus* by
436 comparative genome hybridization experiments (Dittami *et al.*, 2011). We have added the putative
437 *E. subulatus* CBPs to a previous phylogeny of *Ectocarpus* sp. Ec32 CBPs (Dittami *et al.*, 2010) and
438 found both a small group of LHCX CBPs as well as a larger group belonging to the LHCF/LHCR family
439 that have probably undergone a recent expansion (Figure 6). Although some of the proteins
440 appeared to be truncated (marked with asterisks), all of them were associated with at least some
441 RNA-seq reads, suggesting that they may be functional. A number of LHCR family proteins were also
442 flanked by LTR-like sequences as predicted by the LTR-harvest pipeline (Ellinghaus *et al.*, 2008).

Discussion

444 Here we present the draft genome and metabolic network of *E. subulatus* strain Bft15b, a brown alga
445 which, compared to *Ectocarpus* sp. Ec32, is characterized by high abiotic stress tolerance (Bolton,
446 1983; Peters *et al.*, 2015). Based on time-calibrated molecular trees, both species separated roughly

447   16 Mya (Dittami *et al.*, 2012), *i.e.* slightly before *e.g.* the split between *Arabidopsis thaliana* and
448   *Thellungiella salsuginea* 7-12 Mya (Wu *et al.*, 2012). According to our analysis, the split between
449   *Ectocarpus* sp. Ec32 and *E. subulatus* was probably due to allopatric separation with subsequent
450   adaptation of *E. subulatus* to highly fluctuating and low salinity habitats leading to speciation.

Genome evolution of *Ectocarpus* species driven by transposons and viruses

452   Compared to the extremophile plant models *T. salsuginea* or *Arabidopsis lyrata* which have almost
453   doubled in genome size with respect to *A. thaliana*, the *E. subulatus* genome is only approximately
454   23% larger than that of *Ectocarpus* sp. Ec32. In *T. salsuginea* and *A. lyrata*, the observed expansion
455   was attributed mainly to the activity of transposons (Wu *et al.*, 2012; Hu *et al.*, 2011). In the case of
456   *Ectocarpus*, we also observed traces of recent transposon activity, especially from LTR transposons,
457   which is in line with the absence of DNA methylation, and bursts in transposon activity have indeed
458   been identified as one potential driver of local adaptation and speciation in other model systems
459   such as salmon (de Boer *et al.*, 2007). Furthermore, LTRs are known to mediate the
460   retrotransposition of individual genes, leading to the duplication of the latter (Tan *et al.*, 2016). In
461   the *E. subulatus* genome, only a few cases of gene duplication were observed since the separation
462   from *Ectocarpus* sp. Ec32, and in most of them no indication of the involvement of LTRs was found.
463   The only exception was a recent expansion of the LHCR family, in which proteins were flanked by a
464   pair of LTR-like sequences. These elements lacked both the group antigen (GAG) and reverse
465   transcriptase (POL) proteins, which implies that, if retro-transposition was the mechanism underlying
466   the expansion of this group of proteins, it would have depended on other active transposable
467   elements to provide these activities.

468   The second major factor that impacted the *Ectocarpus* genomes were viruses. Viral infections are a
469   common phenomenon in Ectocarpales (Müller *et al.*, 1998), and a well-studied example is the
470   *Ectocarpus siliculosus* virus-1 (EsV-1) (Delaroque *et al.*, 2001). It was found to be present latently in
471   host cells of several strains of *Ectocarpus* sp. closely related to strain Ec32, and has also been found
472   integrated in the genome of the latter strain, although it is not expressed (Cock *et al.*, 2010). As
473   previously indicated by comparative genome hybridization experiments (Dittami *et al.*, 2011), the *E.*
474   *subulatus* genome does not contain a complete EsV-1 like insertion, although a few shorter EsV-1-
475   like proteins were found. Thus, the EsV-1 integration observed in *Ectocarpus* sp. Ec32 has likely
476   occurred after the split with *E. subulatus*. This, together with the presence of other viral sequences
477   specific to *E. subulatus*, indicates that, in addition to transposable elements, viruses have shaped the
478   *Ectocarpus* genomes over the last 16 million years.

Few classical stress response genes but no transporters involved in adaptation

481   A main aim of this study was to identify gene functions that may potentially be responsible for the
482   high abiotic stress and salinity tolerance of *E. subulatus*. Similar studies on genomic adaptation to
483   changes in salinity or to drought in terrestrial plants have previously highlighted genes generally
484   involved in stress tolerance to be expanded in "extremophile" organisms. Examples are the
485   expansion of catalase, glutathione reductase, and heat shock protein families in desert poplar (Ma

486 *et al.*, 2013), arginine metabolism in jujube (Liu *et al.*, 2014), or genes related to cation transport,
487 abscisic acid signaling, and wax production in *T. salsuginea* (Wu *et al.*, 2012). In our study, we found
488 a few genomic differences that match these expectations. *E. subulatus* possesses two additional
489 HSP20 proteins and has an expanded family of CBPs probably involved in non-photochemical
490 quenching, which may contribute to its high stress tolerance. It also has a slightly reduced set of
491 genes involved in the production of halogenated defense compounds which may be related to its
492 habitat preference: *E. subulatus* is frequently found in brackish and even freshwater environments
493 with low availability of halogens. It also specializes in highly abiotic stressful habitats for brown algae
494 and may thus invest less energy in halogen-based defense.

495 Another anticipated adaptation to life in varying salinities lies in modifications of the cell wall.
496 Notably, the content of sulfated polysaccharides is expected to play a crucial role as these
497 compounds are present in all marine plants and algae, but absent in their freshwater relatives
498 (Kloareg and Quatrano, 1988; Popper *et al.*, 2011). The fact that we found only small differences in
499 the number of encoded sulfatases and sulfotransferases indicates that the absence of sulfated cell-
500 wall polysaccharides previously observed in *E. subulatus* in low salinities (Torode *et al.*, 2015) is
501 probably a regulatory effect or simply related to the availability of sulfate depending on the salinity.
502 This is also coherent with the wide distribution of *E. subulatus*, which comprises marine, brackish
503 water, and freshwater environments.

504 Finally, transporters have previously been described as a key element in plant adaptation to different
505 salinities (see Rao *et al.*, 2016 for a review). Similar results have also been obtained for *Ectocarpus* in
506 a study of quantitative trait loci (QTLs) associated with salinity and temperature tolerance (Avia *et*
507 *al.*, 2017). In our study, however, we found no indication of genomic differences related to
508 transporters between the two species. This observation corresponds to previous physiological
509 experiments indicating that *Ectocarpus*, unlike many terrestrial plants, responds to strong changes
510 in salinity as an osmoconformer rather than an osmoregulator, *i.e.* it allows the intracellular salt
511 concentration to adjust to values close to the external medium rather than keeping the intracellular
512 ion composition constant (Dittami *et al.*, 2009).

Genes related to cell-cell communication are under positive selection

513
514 In addition to genes that may be directly involved in the adaptation to the environment, we found
515 several gene clusters containing domains potentially involved in cell-cell signaling that were
516 expanded in the *Ectocarpus* sp. Ec32 genome (Table 2), notably a family of ankyrin repeat-containing
517 domain proteins (Mosavi *et al.*, 2004) was more abundant in Ec32. Furthermore, we identified six
518 ankyrin repeat-containing domain proteins among the genes under positive selection between the
519 two species. The exact function of these proteins, however, is still unknown. The only well-annotated
520 gene under positive selection, a mannosyl-oligosaccharide 1,2-alpha-mannosidase, is probably
521 involved in the modification of glycoproteins which are also important for cell-cell interactions
522 (Tulsiani *et al.*, 1982). Although these genes are not rapidly evolving in *Ectocarpus*, these observed
523 differences may be, in part, responsible for the existing pre-zygotic reproductive barrier between the
524 two examined species of *Ectocarpus* (Lipinska *et al.*, 2016).

14

Genes of unknown function and lineage-specific genes are likely to play a dominant role in adaptation

Despite the gene functions identified as potentially involved in adaptation and speciation above, it is important to keep in mind that the vast majority of genomic differences between the two species of *Ectocarpus* corresponds to proteins of entirely unknown functions. Among the 83 gene pairs under positive selection, 84% were also entirely unknown, and 92% represented genes taxonomically restricted to brown algae. In addition, we identified 1,629 lineage-specific genes, of which 88% were entirely unknown. These genes were for the most part expressed and are thus likely to correspond to true genes. For the lineage-specific genes, their absence from the *Ectocarpus* sp. Ec32 and *S. japonica* genomes was also confirmed on the nucleotide level. A large part of the mechanisms that underlie the adaptation to different ecological niches in *Ectocarpus* may, therefore, lie in these genes of unknown function. This can be explained in part by the fact that still only few brown algal genomes are available and that currently most of our knowledge on the functions of their proteins is based on studies in model plants, animals, yeast, or bacteria. Brown algae, however, are part of the stramenopile lineage that has evolved independently from the former for over 1 billion years (Yoon *et al.*, 2004). They differ from land plants even in otherwise highly conserved aspects, for instance in their life cycles, their cell walls, and their primary metabolism (Charrier *et al.*, 2008). Furthermore, substantial contributions of lineage-specific genes to the evolution of organisms and the development of innovations have also been described for animal models (see Tautz and Domazet-Lošo, 2011 for a review) and studies in basal metazoans furthermore indicate that they are essential for species-specific adaptive processes (Khalturin *et al.*, 2009).

Despite the probable importance of unknown and lineage-specific genes for local adaptation, *Ectocarpus* may still heavily rely on classical stress response genes for abiotic stress tolerance. Many of the gene families known to be related to stress response in land plants (including transporters and genes involved in cell wall modification) for which no significant differences in gene contents were observed, have previously been reported to be strongly regulated in response to environmental stress in *Ectocarpus* (Dittami *et al.*, 2009; Dittami *et al.*, 2012; Ritter *et al.*, 2014). This high transcriptomic plasticity is probably one of the features that allow *Ectocarpus* to thrive in a wide range of environments and may form the basis for its capacity to further adapt to "extreme environments" such as freshwater (West and Kraft, 1996).

Conclusion and future work

We have shown that *E. subulatus* has separated from *Ectocarpus sp.* Ec32 probably via a mechanism of allopatric speciation. Its genome has since been shaped mainly by the activity of viruses and transposons, particularly large retrotransposons. Over this period of time, *E. subulatus* has adapted to environments with high abiotic variability including brackish water and even freshwater. We have identified a number of genes that likely contribute to this adaptation, including HSPs, CBPs, a reduction of genes involved in halogenated defense compounds, or some changes in cell wall polysaccharide modifying enzymes. However, the vast majority of genes that differ between the two examined *Ectocarpus* species or that have recently been under positive selection are lineage-specific and encode proteins of unknown function. This underlines the fundamental differences that exist

15

565 between brown algae and terrestrial plants or other lineages of algae. Studies as the present one,
566 *i.e.* without strong *a priori* assumptions about the mechanisms involved in adaptation, are therefore
567 essential to start elucidating the specificities of this lineage as well as the various functions of the
568 unknown genes. Finally, *E. subulatus* has become an important brown algal model to study the role
569 of algal-bacterial interactions in response to environmental changes. This is due mainly to its
570 dependence on specific bacterial taxa for freshwater tolerance (KleinJan *et al.*, 2017; Dittami *et al.*,
571 2016). The presented algal genome and metabolic network are indispensable tools in this context as
572 well, as they will allow for the separation of algal and bacterial responses in culture experiments, and
573 facilitate the implementation of global approaches based on the use of metabolic network
574 reconstructions (Dittami *et al.*, 2014; Levy *et al.*, 2015).

Materials and Methods

576 **Biological material.** Haploid male parthenosporophytes of *E. subulatus* strain Bft15b (Culture
577 Collection of Algae and Protozoa CCAP accession 1310/34), isolated in 1978 by Dieter G. Müller in
578 Beaufort, North Carolina, USA, were grown in 14 cm (ca. 100 ml) Petri Dishes in Provasoli-enriched
579 seawater (Starr and Zeikus, 1993) under a 14/10 daylight cycle at 14°C. Approximately 1 g fresh
580 weight of algal culture was dried on a paper towel and immediately frozen in liquid nitrogen. For
581 RNA-seq experiments, in addition to Bft15b, a second strain, the diploid freshwater strain CCAP
582 1310/196 isolated from Hopkins River Falls, Australia (West and Kraft, 1996), was included. One
583 culture was grown as described above for Bft15b, and for a second culture, seawater was diluted 20-
584 fold with distilled water prior to the addition of Provasoli nutrients (Dittami *et al.*, 2012).

585 **Flow cytometry** experiments to measure nuclear DNA contents were carried out as described
586 (Bothwell *et al.*, 2010), except that young sporophyte tissue was used instead of gametes. Samples
587 of the genome-sequenced *Ectocarpus* sp. strain Ec32 (CCAP accession 1310/4 from San Juan de
588 Marcona, Peru), were run in parallel as a size reference.

589 **Nucleic acid extraction and sequencing**. DNA and RNA were extracted using a phenol-chloroform-
590 based method according to Le Bail *et al.* (2008). For DNA sequencing, four Illumina libraries were
591 prepared and sequenced on a HiSeq 2000: one paired-end library (Illumina TruSeq DNA PCR-free LT
592 Sample Prep kit #15036187, sequenced with 2x100 bp read length), and three mate-pair libraries
593 with span sizes of 3kb, 5kb, and 10kb respectively (Nextera Mate Pair Sample Preparation Kit;
594 sequenced with 2x50bp read length). One poly-A enriched RNA-seq library was generated for each
595 of the three aforementioned cultures according to the Illumina TruSeq Stranded mRNA Sample Prep
596 kit #15031047 protocol and sequenced with 2x50 bp read length.

597 **Methylation.** The degree of DNA methylation was examined by HPLC on CsCl-gradient purified DNA
598 (Le Bail *et al.*, 2008) from three independent cultures per strain as previously described (Rival *et al.*,
599 2013).

600 **Sequence assembly**. Redundancy of mate pairs (MPs) was reduced by mapping MPs to a preliminary
601 assembly, to mitigate the negative effect of redundant chimeric MPs during scaffolding. Clean DNA

602   reads were assembled using SOAPDenovo2 (Luo *et al.*, 2012). Scaffolding was then carried out using
603   SSPACE basic 2.0 (Boetzer *et al.*, 2011) (trim length up to 5 bases, min 3 links to scaffold contigs, min
604   15 reads to call a base during an extension) followed by a run of GapCloser (part of the SOAPDenovo
605   package, default settings). Alternative assemblers (CLC and Velvet) were also tested but yielded
606   significantly lower final contig and scaffold lengths. RNA-seq reads were cleaned using Trimmomatic
607   (default settings), first assembled *de novo* using Trinity 2.1.1 (Grabherr *et al.*, 2011) and filtered by
608   coverage with an FPKM cutoff of 1. Later, a second genome-guided assembly was performed with
609   Tophat2 and with Cufflinks.

610   **Removal of bacterial sequences:** As cultures were not treated with antibiotics prior to DNA
611   extraction, bacterial scaffolds were removed from the final assembly using the taxoblast pipeline
612   (Dittami and Corre, 2017). Every scaffold was cut into fragments of 500 bp, and these fragments were
613   aligned (blastn, e-value cutoff 0.01) against the GenBank non-redundant nucleotide (nt) database.
614   Scaffolds for which more than 90% of their 500 bp-fragments had bacterial sequences as best blast
615   hits were removed from the assembly (varying this threshold between 30 and 95% resulted in only
616   very minor differences in the final assembly). "Bacterial" scaffolds were submitted to the MG-Rast
617   server to obtain an overview of the taxa present in the sample (Meyer *et al.*, 2008).

618   **Repeated elements** were searched for *de novo* using TEdenovo and annotated using TEannot with
619   default parameters. Both tools are part of the REPET pipeline (Flutre *et al.*, 2011), of which version
620   2.5 was used for our dataset.

621   **Assessment of genome completeness:** BUSCO 2.0 analyses (Simão *et al.*, 2015) were run on the
622   servers of the IPlant Collaborative (Goff *et al.*, 2011) with the general eukaryote database as a
623   reference and default parameters. BUSCO internally uses Augustus (Stanke *et al.*, 2004) to predict
624   protein coding sequences. As the latter tool performed poorly on both *Ectocarpus* strains in
625   preliminary tests, predicted proteins were used as input instead of DNA sequences.

626   **Organellar genomes**, *i.e.* plastid and mitochondrion, were manually assembled based on scaffolds
627   416 and 858 respectively, using the published genome of *Ectocarpus* sp. Ec32 as a guide (Delage *et*
628   *al.*, 2011; Le Corguillé *et al.*, 2009; Cock *et al.*, 2010). In the case of the mitochondrial genome, the
629   correctness of the manual assembly was verified by PCR where manual and automatic assemblies
630   diverged. Both organellar genomes were visualized using OrganellarGenomeDRAW (Lohse *et al.*,
631   2013) and aligned with the *Ectocarpus* sp. Ec32 organelles using Mauve 2.3.1 (Darling *et al.*, 2004).

632   **Gene prediction.** Putative protein-coding sequences were identified using Eugene 4.1c (Foissac *et*
633   *al.*, 2008). RNA-seq reads were mapped against the assembled genome using GenomeThreader
634   1.6.5, and all available proteins from the Swiss-Prot database (Dec. 2014) as well as predicted
635   proteins from the *Ectocarpus* sp. Ec32 genome (Cock *et al.*, 2010) were aligned to the genome using
636   KLAST (Nguyen and Lavenier, 2009). Both aligned *de novo*-assembled transcripts and proteins were
637   provided to Eugene for gene prediction, which was run with the parameter set previously optimized
638   for the *Ectocarpus* sp. Ec32 genome (Cock *et al.*, 2010).

639   **Functional annotation**. Predicted proteins were compared to the Swiss-Prot database by BlastP
640   search (e-value cutoff 1e-5), and the results imported to Blast2GO (Götz *et al.*, 2008), which was used
641   to run InterPro domain searches and automatically annotate proteins with a description, GO
642   numbers, and EC codes. The genome and all automatic annotations were imported into Apollo (Lee
643   *et al.*, 2013; Dunn *et al.*, 2017) for manual curation.

644   **Metabolic network reconstruction**. The *E. subulatus* genome-scale metabolic model (GEM)
645   reconstruction was carried out as previously described by Prigent *et al.* (2014) by merging an
646   annotation-based reconstruction obtained with Pathway Tools (Karp *et al.*, 2016) and an orthology-
647   based reconstruction based on the *Arabidopsis thaliana* metabolic network AraGEM (de Oliveira
648   Dal'Molin *et al.*, 2010) using Pantograph (Loira *et al.*, 2015). A final step of gap-filling was then carried
649   out using the Meneco tool (Prigent *et al.*, 2017). The entire reconstruction pipeline is available via
650   the AuReMe workspace (Aite *et al.*, 2018; http://aureme.genouest.org/). For pathway-based
651   analyses, pathways that contained only a single reaction or that were less than 50% complete were
652   not considered.

653   **Genome comparisons.** Functional comparisons of gene contents were based primarily on
654   orthologous clusters of genes shared with version 2 of the *Ectocarpus* sp. Ec32 genome (Cormier *et*
655   *al.*, 2017) as well as the *Saccharina japonica* (Areschoug) genome (Ye *et al.*, 2015). They were
656   determined by the OrthoFinder software version 0.7.1 (Emms and Kelly, 2015). For any predicted
657   proteins that were not part of a multi-species cluster, we verified the absence in the other two
658   genomes also by tblastn searches. Proteins without hit (threshold e-value of 1e-10) were considered
659   lineage-specific proteins. Blast2GO 3.1 (Götz *et al.*, 2008) was then used to identify significantly
660   enriched GO terms among the lineage-specific genes or the expanded gene families (Fischer's exact
661   test with FDR correction FDR<0.05). In parallel, a manual examination of these genes was carried
662   out. Furthermore, we compared both *Ectocarpus* genomes with respect to the presence or absence
663   of Interpro domain annotations. A third approach consisted in identifying clusters of genes that were
664   expanded in either of the two *Ectocarpus* genomes. All protein families expanded in the *E. subulatus*
665   genome were manually examined.

666   **Genes under positive selection**. We examined clusters of orthologous genes with one homolog in *E.*
667   *subulatus* and one in *Ectocarpus* sp. Ec32 to search for genes potentially under positive selection. To
668   this means, pairwise alignments of protein-coding nucleotide sequences were performed using
669   TranslatorX (Abascal *et al.*, 2010) and Muscle (Edgar, 2004). The aligning regions were then analyzed
670   in the yn00 package of PaML4.4 (Yang, 2007), and all proteins with a ratio of non-synonymous to
671   synonymous mutations (dN/dS) > 1 were manually examined. The distribution of these genes across
672   the genome was examined by calculating variance to mean ratios based on window sizes of 50 to
673   500 genes.

674   **Phylogenetic analyses.** Phylogenetic analyses were carried out for gene families of particular
675   interest. For chlorophyll-binding proteins (CBPs), reference sequences were obtained from a
676   previous study (Dittami *et al.*, 2010), and aligned together with *E. subulatus* and *S. japonica* CBPs
677   using MAFFT (G-INS-i) (Katoh *et al.*, 2002). Alignments were then manually curated, conserved

18

678 positions selected in Jalview (Waterhouse *et al.*, 2009), and maximum likelihood analyses carried out
679 using PhyML 3.0 (Guindon and Gascuel, 2003), the LG substitution model, 100 bootstrap replicates,
680 and an estimation of the gamma distribution parameter. The resulting phylogenetic tree was
681 visualized using MEGA7 (Kumar *et al.*, 2016). The same procedure was also used in the case of
682 selected Ankyrin Repeat domain-containing proteins.

683 **Data availability:** Raw sequence data (genomic and transcriptomic reads) as well as assembled
684 scaffolds and predicted proteins and annotations were submitted to the European Nucleotide
685 Archive (ENA) under project accession number PRJEB25230 using the EMBLmyGFF3 script (Dainat
686 and Gourlé, 2018). A JBrowse (Skinner *et al.*, 2009) instance comprising the most recent annotations
687 is available via the server of the Station Biologique de Roscoff (http://mmo.sb-
688 roscoff.fr/jbrowseEsu/?data=data/public/ectocarpus/subulatus_bft). The reconstructed metabolic
689 network of *E. subulatus* is available at http://gem-aureme.irisa.fr/sububftgem. Additional resources
690 and annotations including a blast server are available at http://application.sb-
691 roscoff.fr/project/subulatus/index.html. The complete set of manual annotations is provided in
692 Supporting Information Table S1.

Acknowledgements

References

702 **Abascal, F., Zardoya, R. and Telford, M.J.** (2010) TranslatorX: multiple alignment of nucleotide
703 sequences guided by amino acid translations. *Nucleic Acids Res.*, **38**, W7–W13. Available at:
704 https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkq291.
705 **Aite, M., Chevallier, M., Frioux, C., et al.** (2018) Traceability, reproducibility and wiki-exploration
706 for "à-la-carte" reconstructions of genome-scale metabolic models. *PLoS Comput. Biol.*, **in
707 press**.
708 **Amtmann, A.** (2009) Learning from evolution: *Thellungiella* generates new knowledge on essential
709 and critical components of abiotic stress tolerance in plants. *Mol. Plant*, **2**, 3–12. Available at:
710 http://www.ncbi.nlm.nih.gov/pubmed/19529830.
711 **Archibald, J.M.** (2009) The puzzle of plastid evolution. *Curr. Biol.*, **19**, R81-8. Available at:
712 http://www.ncbi.nlm.nih.gov/pubmed/19174147.
713 **Avia, K., Coelho, S.M., Montecinos, G.J., et al.** (2017) High-density genetic map and identification
714 of QTLs for responses to temperature and salinity stresses in the model brown alga
715 *Ectocarpus*. *Sci. Rep.*, **7**, 43241. Available at:
716 http://www.ncbi.nlm.nih.gov/pubmed/28256542.

19

717 **Bail, A. Le, Dittami, S.M., Franco, P.O. de, Rousvoal, S., Cock, J.M., Tonon, T. and Charrier, B.**
718   (2008) Normalisation genes for expression analyses in the brown alga model *Ectocarpus*
719   *siliculosus*. *BMC Mol. Biol.*, **9**, 75. Available at: http://www.biomedcentral.com/1471-
720   2199/9/75.
721 **Boer, J.G. de, Yazawa, R., Davidson, W.S. and Koop, B.F.** (2007) Bursts and horizontal evolution of
722   DNA transposons in the speciation of pseudotetraploid salmonids. *BMC Genomics*, **8**, 422.
723   Available at: http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-8-422.
724 **Boetzer, M., Henkel, C. V, Jansen, H.J., Butler, D. and Pirovano, W.** (2011) Scaffolding pre-
725   assembled contigs using SSPACE. *Bioinformatics*, **27**, 578–9. Available at:
726   http://www.ncbi.nlm.nih.gov/pubmed/21149342.
727 **Bolton, J.J.** (1983) Ecoclinal variation in *Ectocarpus siliculosus* (Phaeophyceae) with respect to
728   temperature growth optima and survival limits. *Mar. Biol.*, **73**, 131–138. Available at:
729   http://www.springerlink.com/index/10.1007/BF00406880.
730 **Bothwell, J.H., Marie, D., Peters, A.F., Cock, J.M. and Coelho, S.M.** (2010) Role of
731   endoreduplication and apomeiosis during parthenogenetic reproduction in the model brown
732   alga *Ectocarpus*. *New Phytol.*, **188**, 111–21. Available at:
733   http://www.ncbi.nlm.nih.gov/pubmed/20618911.
734 **Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B.** (2009) The
735   Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic*
736   *Acids Res.*, **37**, D233–D238. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18838391.
737 **Charrier, B., Coelho, S.M., Bail, A. Le, et al.** (2008) Development and physiology of the brown alga
738   *Ectocarpus siliculosus*: two centuries of research. *New Phytol.*, **177**, 319–32. Available at:
739   http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&l
740   ist_uids=18181960.
741 **Cock, J.M., Sterck, L., Rouzé, P., et al.** (2010) The *Ectocarpus* genome and the independent
742   evolution of multicellularity in brown algae. *Nature*, **465**, 617–21. Available at:
743   http://www.ncbi.nlm.nih.gov/pubmed/20520714.
744 **Collén, J. and Davison, I.R.** (1999) Reactive oxygen metabolism in intertidal *Fucus* spp.
745   (Phaeophyceae). *J. Phycol.*, **35**, 62–69.
746 **Corguillé, G. Le, Pearson, G., Valente, M., et al.** (2009) Plastid genomes of two brown algae,
747   *Ectocarpus siliculosus* and *Fucus vesiculosus*: further insights on the evolution of red-algal
748   derived plastids. *BMC Evol. Biol.*, **9**, 253. Available at:
749   http://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-9-253.
750 **Cormier, A., Avia, K., Sterck, L., et al.** (2017) Re-annotation, improved large-scale assembly and
751   establishment of a catalogue of noncoding loci for the genome of the model brown alga
752   *Ectocarpus*. *New Phytol.*, **214**, 219–232. Available at:
753   http://doi.wiley.com/10.1111/nph.14321.
754 **Dainat, J. and Gourlé, H.** (2018) NBISweden/EMBLmyGFF3: EMBLmyGFF3-1.2.2. *Zenodo*. Available
755   at: https://zenodo.org/record/1208446#.WtSToNNubIE.
756 **Darling, A.C.E., Mau, B., Blattner, F.R. and Perna, N.T.** (2004) Mauve: multiple alignment of
757   conserved genomic sequence with rearrangements. *Genome Res.*, **14**, 1394–1403. Available
758   at: http://www.genome.org/cgi/doi/10.1101/gr.2289704.
759 **Das, K. and Roychoudhury, A.** (2014) Reactive oxygen species (ROS) and response of antioxidants
760   as ROS-scavengers during environmental stress in plants. *Front. Environ. Sci.*, **2**, 53. Available
761   at: http://journal.frontiersin.org/article/10.3389/fenvs.2014.00053/abstract.
762 **Dassanayake, M., Oh, D.-H., Haas, J.S., et al.** (2011) The genome of the extremophile crucifer
763   *Thellungiella parvula*. *Nat. Genet.*, **43**, 913–918. Available at:
764   http://dx.doi.org/10.1038/ng.889.

765   **Davison, I.R. and Pearson, G.A.** (1996) Stress tolerance in intertidal seaweeds. *J. Phycol.*, **32**, 197–
766   211.

767   **Delage, L., Leblanc, C., Nyvall Collén, P., Gschloessl, B., Oudot, M.-P., Sterck, L., Poulain, J., Aury,**
768   **J.-M. and Cock, J.M.** (2011) In silico survey of the mitochondrial protein uptake and
769   maturation systems in the brown alga *Ectocarpus siliculosus* E. Newbigin, ed. *PLoS One*, **6**,
770   e19540. Available at: http://dx.plos.org/10.1371/journal.pone.0019540.

771   **Delaroque, N., Müller, D.G., Bothe, G., Pohl, T., Knippers, R. and Boland, W.** (2001) The complete
772   DNA sequence of the *Ectocarpus siliculosus* virus EsV-1 genome. *Virology*, **287**, 112–132.

773   **Desmond, E. and Gribaldo, S.** (2009) Phylogenomics of sterol synthesis: insights into the origin,
774   evolution, and diversity of a key eukaryotic feature. *Genome Biol. Evol.*, **1**, 364–81. Available
775   at: http://www.ncbi.nlm.nih.gov/pubmed/20333205.

776   **Dittami, S.M. and Corre, E.** (2017) Detection of bacterial contaminants and hybrid sequences in the
777   genome of the kelp *Saccharina japonica* using Taxoblast. *PeerJ*, **5**, e4073. Available at:
778   https://peerj.com/articles/4073.

779   **Dittami, S.M., Duboscq-Bidot, L., Perennou, M., Gobet, A., Corre, E., Boyen, C. and Tonon, T.**
780   (2016) Host–microbe interactions as a driver of acclimation to salinity gradients in brown algal
781   cultures. *ISME J.*, **10**, 51–63. Available at:
782   http://www.nature.com/doifinder/10.1038/ismej.2015.104.

783   **Dittami, S.M., Eveillard, D. and Tonon, T.** (2014) A metabolic approach to study algal-bacterial
784   interactions in changing environments. *Mol. Ecol.*, **23**, 1656–60. Available at:
785   http://www.ncbi.nlm.nih.gov/pubmed/24447216.

786   **Dittami, S.M., Gravot, A., Goulitquer, S., Rousvoal, S., Peters, A.F., Bouchereau, A., Boyen, C. and**
787   **Tonon, T.** (2012) Towards deciphering dynamic changes and evolutionary mechanisms
788   involved in the adaptation to low salinities in *Ectocarpus* (brown algae). *Plant J.*, **71**, 366–377.
789   Available at: http://doi.wiley.com/10.1111/j.1365-313X.2012.04982.x.

790   **Dittami, S.M., Heesch, S., Olsen, J.L. and Collén, J.** (2017) Transitions between marine and
791   freshwater environments provide new clues about the origins of multicellular plants and algae
792   O. DeClerck, ed. *J. Phycol.*, **53**, 731–745. Available at:
793   http://doi.wiley.com/10.1111/jpy.12547.

794   **Dittami, S.M., Michel, G., Collén, J., Boyen, C. and Tonon, T.** (2010) Chlorophyll-binding proteins
795   revisited--a multigenic family of light-harvesting and stress proteins from a brown algal
796   perspective. *BMC Evol. Biol.*, **10**, 365. Available at:
797   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3008699&tool=pmcentrez&rend
798   ertype=abstract.

799   **Dittami, S.M., Proux, C., Rousvoal, S., Peters, A.F., Cock, J.M., Coppee, J.-Y., Boyen, C. and Tonon,**
800   **T.** (2011) Microarray estimation of genomic inter-strain variability in the genus *Ectocarpus*
801   (Phaeophyceae). *BMC Mol. Biol.*, **12**, 2. Available at:
802   http://www.ncbi.nlm.nih.gov/pubmed/21226968.

803   **Dittami, S.M., Scornet, D., Petit, J.-L., et al.** (2009) Global expression analysis of the brown alga
804   *Ectocarpus siliculosus* (Phaeophyceae) reveals large-scale reprogramming of the
805   transcriptome in response to abiotic stress. *Genome Biol.*, **10**, R66. Available at:
806   http://genomebiology.com/2009/10/6/R66.

807   **Dittami, S.M. and Tonon, T.** (2012) Genomes of extremophile crucifers: new platforms for
808   comparative genomics and beyond. *Genome Biol.*, **13**, 166. Available at:
809   http://www.ncbi.nlm.nih.gov/pubmed/22897896.

810   **Dong, H.-P., Dong, Y.-L., Cui, L., Balamurugan, S., Gao, J., Lu, S.-H. and Jiang, T.** (2016) High light
811   stress triggers distinct proteomic responses in the marine diatom *Thalassiosira pseudonana*.
812   *BMC Genomics*, **17**, 994. Available at:

813   http://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-3335-5.

814   **Dunn, N., Diesh, C., Deepak, et al.** (2017) GMOD/Apollo: Apollo2.0.8(JB#d3827c). *Zenodo*.
815   Available at: https://zenodo.org/record/1063658#.WsTyndNubIE.

816   **Edgar, R.C.** (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput.
817   *Nucleic Acids Res.*, **32**, 1792–7. Available at:
818   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=390337&tool=pmcentrez&rende
819   rtype=abstract.

820   **Ellinghaus, D., Kurtz, S. and Willhoeft, U.** (2008) LTRharvest, an efficient and flexible software for
821   de novo detection of LTR retrotransposons. *BMC Bioinformatics*, **9**, 18. Available at:
822   http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-18.

823   **Emms, D.M. and Kelly, S.** (2015) OrthoFinder: solving fundamental biases in whole genome
824   comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.*, **16**, 157.
825   Available at: http://genomebiology.com/2015/16/1/157.

826   **Flutre, T., Duprat, E., Feuillet, C. and Quesneville, H.** (2011) Considering transposable element
827   diversification in de novo annotation approaches Y. Xu, ed. *PLoS One*, **6**, e16526. Available at:
828   http://dx.plos.org/10.1371/journal.pone.0016526.

829   **Foissac, S., Gouzy, J., Rombauts, S., Mathe, C., Amselem, J., Sterck, L., Peer, Y. V. de, Rouze, P.**
830   **and Schiex, T.** (2008) Genome annotation in plants and fungi: EuGene as a model platform.
831   *Curr. Bioinform.*, **3**, 11. Available at:
832   http://www.ingentaconnect.com/content/ben/cbio/2008/00000003/00000002/art00003?tok
833   en=004f12e9145a666f3a7b6c7a407b425b6b414c7d666749264f655d375c6b68763050211ce9
834   d7c2a.

835   **Food and Agriculture Organization of the United Nations, F.** (2016) Global production statistics
836   1950-2014. Available at: http://www.fao.org/fishery/statistics/global-production/en.

837   **Fournier, J.-B., Rebuffet, E., Delage, L., et al.** (2014) The Vanadium Iodoperoxidase from the
838   marine flavobacteriaceae species *Zobellia galactanivorans* reveals novel molecular and
839   evolutionary features of halide specificity in the vanadium haloperoxidase enzyme family.
840   *Appl. Environ. Microbiol.*, **80**, 7561–73. Available at:
841   http://www.ncbi.nlm.nih.gov/pubmed/25261522.

842   **Geiselman, J.A. and McConnell, O.J.** (1981) Polyphenols in brown algae *Fucus vesiculosus* and
843   *Ascophyllum nodosum*: Chemical defenses against the marine herbivorous snail, *Littorina*
844   *littorea*. *J. Chem. Ecol.*, **7**, 1115–1133. Available at:
845   http://link.springer.com/10.1007/BF00987632.

846   **Goff, S.A., Vaughn, M., McKay, S., et al.** (2011) The iPlant collaborative: cyberinfrastructure for
847   plant biology. *Front. Plant Sci.*, **2**, 34. Available at:
848   http://journal.frontiersin.org/article/10.3389/fpls.2011.00034/abstract.

849   **Götz, S., García-Gómez, J.M., Terol, J., et al.** (2008) High-throughput functional annotation and
850   data mining with the Blast2GO suite. *Nucleic Acids Res.*, **36**, 3420–35. Available at:
851   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2425479&tool=pmcentrez&rend
852   ertype=abstract.

853   **Grabherr, M.G., Haas, B.J., Yassour, M., et al.** (2011) Full-length transcriptome assembly from
854   RNA-Seq data without a reference genome. *Nat. Biotechnol.*, **29**, 644–652. Available at:
855   http://www.nature.com/doifinder/10.1038/nbt.1883.

856   **Gualtieri, T., Ragni, E., Mizzi, L., Fascio, U. and Popolo, L.** (2004) The cell wall sensor Wsc1p is
857   involved in reorganization of actin cytoskeleton in response to hypo-osmotic shock in
858   *Saccharomyces cerevisiae*. *Yeast*, **21**, 1107–1120. Available at:
859   http://www.ncbi.nlm.nih.gov/pubmed/15484288.

860   **Guindon, S. and Gascuel, O.** (2003) A simple, fast, and accurate algorithm to estimate large

861   phylogenies by maximum likelihood. *Syst. Biol.*, **52**, 696–704. Available at:
862   http://www.ncbi.nlm.nih.gov/pubmed/14530136.

863 **Heesch, S., Cho, G.Y., Peters, A.F., et al.** (2010) A sequence-tagged genetic map for the brown alga
864   *Ectocarpus siliculosus* provides large-scale assembly of the genome sequence. *New Phytol.*,
865   **188**, 42–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20456050.

866 **Hervé, C., Siméon, A., Jam, M., et al.** (2016) Arabinogalactan proteins have deep roots in
867   eukaryotes: identification of genes and epitopes in brown algae and their role in *Fucus*
868   *serratus* embryo development. *New Phytol.*, **209**, 1428–1441. Available at:
869   http://www.ncbi.nlm.nih.gov/pubmed/26667994.

870 **Hu, T.T., Pattyn, P., Bakker, E.G., et al.** (2011) The *Arabidopsis lyrata* genome sequence and the
871   basis of rapid genome size change. *Nat. Genet.*, **43**, 476–81. Available at:
872   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3083492&tool=pmcentrez&rend
873   ertype=abstract.

874 **Karp, P.D., Latendresse, M., Paley, S.M., et al.** (2016) Pathway Tools version 19.0 update: software
875   for pathway/genome informatics and systems biology. *Brief. Bioinform.*, **17**, 877–890.
876   Available at: http://www.ncbi.nlm.nih.gov/pubmed/26454094.

877 **Katoh, K., Misawa, K., Kuma, K. and Miyata, T.** (2002) MAFFT: a novel method for rapid multiple
878   sequence alignment based on fast Fourier transform. *Nucleic Acids Res.*, **30**, 3059–66.
879   Available at: http://www.ncbi.nlm.nih.gov/pubmed/12136088.

880 **Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R. and Bosch, T.C.G.** (2009) More than just
881   orphans: are taxonomically-restricted genes important in evolution? *Trends Genet.*, **25**, 404–
882   413. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0168952509001450.

883 **KleinJan, H., Jeanthon, C., Boyen, C. and Dittami, S.M.** (2017) Exploring the cultivable *Ectocarpus*
884   microbiome. *Front. Microbiol.*, **8**, 2456. Available at:
885   https://www.frontiersin.org/articles/10.3389/fmicb.2017.02456/abstract.

886 **Kloareg, B. and Quatrano, R.S.** (1988) Structure of the cell-walls of marine-algae and
887   ecophysiological functions of the matrix polysaccharides. *Ocean. Mar Biol*, **26**, 259–315.

888 **Kumar, S., Stecher, G. and Tamura, K.** (2016) MEGA7: Molecular Evolutionary Genetics Analysis
889   Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.*, **33**, 1870–1874. Available at:
890   http://www.ncbi.nlm.nih.gov/pubmed/27004904.

891 **Lee, E., Helt, G.A., Reese, J.T., et al.** (2013) Web Apollo: a web-based genomic annotation editing
892   platform. *Genome Biol.*, **14**, R93. Available at:
893   http://genomebiology.biomedcentral.com/articles/10.1186/gb-2013-14-8-r93.

894 **Lee, J., Cho, C.H., Park, S.I., Choi, J.W., Song, H.S., West, J.A., Bhattacharya, D. and Yoon, H.S.**
895   (2016) Parallel evolution of highly conserved plastid genome architecture in red seaweeds and
896   seed plants. *BMC Biol.*, **14**, 75. Available at:
897   http://bmcbiol.biomedcentral.com/articles/10.1186/s12915-016-0299-5.

898 **Levy, R., Carr, R., Kreimer, A., Freilich, S. and Borenstein, E.** (2015) NetCooperate: a network-
899   based tool for inferring host-microbe and microbe-microbe cooperation. *BMC Bioinformatics*,
900   **16**, 164. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25980407.

901 **Lipinska, A.P., Damme, E.J.M. Van and Clerck, O. De** (2016) Molecular evolution of candidate male
902   reproductive genes in the brown algal model *Ectocarpus*. *BMC Evol. Biol.*, **16**, 5. Available at:
903   http://www.ncbi.nlm.nih.gov/pubmed/26728038.

904 **Liu, M.-J., Zhao, J., Cai, Q.-L., et al.** (2014) The complex jujube genome provides insights into fruit
905   tree biology. *Nat. Commun.*, **5**, 5315. Available at:
906   http://www.nature.com/doifinder/10.1038/ncomms6315.

907 **Lohse, M., Drechsel, O., Kahlau, S. and Bock, R.** (2013) OrganellarGenomeDRAW--a suite of tools
908   for generating physical maps of plastid and mitochondrial genomes and visualizing expression

23

909   data sets. *Nucleic Acids Res.*, **41**, W575-81. Available at:
910   http://www.ncbi.nlm.nih.gov/pubmed/23609545.

911 **Loira, N., Zhukova, A. and Sherman, D.J.** (2015) Pantograph: A template-based method for
912   genome-scale metabolic model reconstruction. *J. Bioinform. Comput. Biol.*, **13**, 1550006.
913   Available at: http://www.worldscientific.com/doi/abs/10.1142/S0219720015500067.

914 **Luo, R., Liu, B., Xie, Y., et al.** (2012) SOAPdenovo2: an empirically improved memory-efficient
915   short-read de novo assembler. *Gigascience*, **1**, 18. Available at:
916   http://gigascience.biomedcentral.com/articles/10.1186/2047-217X-1-18.

917 **Ma, T., Wang, J., Zhou, G., et al.** (2013) Genomic insights into salt adaptation in a desert poplar.
918   *Nat. Commun.*, **4**, 2797. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24256998.

919 **McHugh, D.J.** (2003) A guide to the seaweed industry. *FAO Fish. Tech. Pap. (FAO, Rome, Italy)*.

920 **Meslet-Cladière, L., Delage, L., Leroux, C.J.-J., et al.** (2013) Structure/Function analysis of a type iii
921   polyketide synthase in the brown alga *Ectocarpus siliculosus* reveals a biochemical pathway in
922   phlorotannin monomer biosynthesis. *Plant Cell*, **25**, 3089–103. Available at:
923   http://www.plantcell.org/content/early/2013/08/27/tpc.113.111336.

924 **Meyer, F., Paarmann, D., D'Souza, M., et al.** (2008) The metagenomics RAST server - a public
925   resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC*
926   *Bioinformatics*, **9**, 386. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18803844.

927 **Michel, G., Tonon, T., Scornet, D., Cock, J.M. and Kloareg, B.** (2010) The cell wall polysaccharide
928   metabolism of the brown alga Ectocarpus siliculosus. Insights into the evolution of
929   extracellular matrix polysaccharides in eukaryotes. *New Phytol.*, **188**, 82–97. Available at:
930   http://www.ncbi.nlm.nih.gov/pubmed/20618907 [.

931 **Mikami, K., Ito, M., Taya, K., Kishimoto, I., Kobayashi, T., Itabashi, Y. and Tanaka, R.** (2018)
932   Parthenosporophytes of the brown alga *Ectocarpus siliculosus* exhibit sex-dependent
933   differences in thermotolerance as well as fatty acid and sterol composition. *Mar. Environ. Res.*
934   Available at: http://linkinghub.elsevier.com/retrieve/pii/S0141113617306979.

935 **Montecinos, A.E., Couceiro, L., Peters, A.F., Desrut, A., Valero, M. and Guillemin, M.-L.** (2017)
936   Species delimitation and phylogeographic analyses in the *Ectocarpus* subgroup *siliculosi*
937   (Ectocarpales, Phaeophyceae) M. Cock, ed. *J. Phycol.*, **53**, 17–31. Available at:
938   http://www.ncbi.nlm.nih.gov/pubmed/27454456.

939 **Mosavi, L.K., Cammett, T.J., Desrosiers, D.C. and Peng, Z.** (2004) The ankyrin repeat as molecular
940   architecture for protein recognition. *Protein Sci.*, **13**, 1435–1448. Available at:
941   http://doi.wiley.com/10.1110/ps.03554604.

942 **Mota, C.F., Engelen, A.H., Serrão, E.A. and Pearson, G.A.** (2015) Some don't like it hot:
943   microhabitat-dependent thermal and water stresses in a trailing edge population J. Watling,
944   ed. *Funct. Ecol.*, **29**, 640–649. Available at: http://doi.wiley.com/10.1111/1365-2435.12373.

945 **Müller, D.G., Kapp, M. and Knippers, R.** (1998) Viruses in marine brown algae. In Academic Press,
946   pp. 49–67. Available at:
947   http://www.sciencedirect.com/science/article/pii/S0065352708608052.

948 **Nguyen, V.H. and Lavenier, D.** (2009) PLAST: parallel local alignment search tool for database
949   comparison. *BMC Bioinformatics*, **10**, 329. Available at: http://www.biomedcentral.com/1471-
950   2105/10/329.

951 **Nishitsuji, K., Arimoto, A., Iwai, K., et al.** (2016) A draft genome of the brown alga, *Cladosiphon*
952   *okamuranus* , S-strain: a platform for future studies of "mozuku" biology. *DNA Res.*, dsw039.
953   Available at: http://dnaresearch.oxfordjournals.org/lookup/doi/10.1093/dnares/dsw039.

954 **Oh, D.-H., Dassanayake, M., Bohnert, H.J. and Cheeseman, J.M.** (2012) Life at the extreme: lessons
955   from the genome. *Genome Biol.*, **13**, 241. Available at:
956   http://www.ncbi.nlm.nih.gov/pubmed/22390828.

957   **Oliveira Dal'Molin, C.G. de, Quek, L.-E., Palfreyman, R.W., Brumbley, S.M. and Nielsen, L.K.** (2010)
958   AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis.
959   *Plant Physiol.*, **152**, 579–89. Available at:
960   http://www.plantphysiol.org/content/152/2/579.full.
961   **Pavia, H., Cervin, G., Lindgren, A. and Åberg, P.** (1997) Effects of UV-B radiation and simulated
962   herbivory on phlorotannins in the brown alga *Ascophyllum nodosum*. *Mar. Ecol. Prog. Ser.*,
963   **157**, 139–146. Available at: http://www.int-res.com/abstracts/meps/v157/p139-146/.
964   **Peers, G., Truong, T.B., Ostendorf, E., Busch, A., Elrad, D., Grossman, A.R., Hippler, M. and Niyogi,**
965   **K.K.** (2009) An ancient light-harvesting protein is critical for the regulation of algal
966   photosynthesis. *Nature*, **462**, 518–21. Available at:
967   http://www.ncbi.nlm.nih.gov/pubmed/19940928.
968   **Peters, A.F., Coucerio, L., Tsiamis, K., Küpper, F.C. and Valero, M.** (2015) Barcoding of cryptic
969   stages of marine brown algae isolated from incubated substratum reveals high diversity.
970   *Cryptogam. Algol.*, **36**, 3–29.
971   **Peters, A.F., Marie, D., Scornet, D., Kloareg, B. and Cock, J.M.** (2004) Proposal of *Ectocarpus*
972   *siliculosus* (Ectocarpales, Phaeophyceae) as a model organism for brown algal genetics and
973   genomics. *J. Phycol.*, **40**, 1079–1088.
974   **Popper, Z.A., Michel, G., Hervé, C., Domozych, D.S., Willats, W.G.T., Tuohy, M.G., Kloareg, B. and**
975   **Stengel, D.B.** (2011) Evolution and diversity of plant cell walls: from algae to flowering plants.
976   *Annu. Rev. Plant Biol.*, **62**, 567–90. Available at:
977   http://www.annualreviews.org/doi/abs/10.1146/annurev-arplant-042110-103809.
978   **Prigent, S., Collet, G., Dittami, S.M., et al.** (2014) The genome-scale metabolic network of
979   Ectocarpus siliculosus (EctoGEM): a resource to study brown algal physiology and beyond.
980   *Plant J.*, **80**, 367–381. Available at: http://doi.wiley.com/10.1111/tpj.12627.
981   **Prigent, S., Frioux, C., Dittami, S.M., et al.** (2017) Meneco, a Topology-Based Gap-Filling Tool
982   Applicable to Degraded Genome-Wide Metabolic Networks C. Kaleta, ed. *PLOS Comput. Biol.*,
983   **13**, e1005276. Available at: http://dx.plos.org/10.1371/journal.pcbi.1005276.
984   **Rao, A.Q., ud Din, S., Akhtar, S., et al.** (2016) Genomics of salinity tolerance in plants. In I. Y.
985   Abdurakhmonov, ed. *Plant Genomics*. InTech, pp. 273–299. Available at:
986   http://www.intechopen.com/books/plant-genomics/genomics-of-salinity-tolerance-in-plants.
987   **Read, S.M., Currie, G. and Bacic, A.** (1996) Analysis of the structural heterogeneity of laminarin by
988   electrospray-ionisation-mass spectrometry. *Carbohydr. Res.*, **281**, 187–201.
989   **Ritter, A., Dittami, S.M., Goulitquer, S., Correa, J.A., Boyen, C., Potin, P. and Tonon, T.** (2014)
990   Transcriptomic and metabolomic analysis of copper stress acclimation in *Ectocarpus*
991   *siliculosus* highlights signaling and tolerance mechanisms in brown algae. *BMC Plant Biol.*, **14**,
992   116. Available at: http://www.biomedcentral.com/1471-2229/14/116.
993   **Rival, A., Ilbert, P., Labeyrie, A., et al.** (2013) Variations in genomic DNA methylation during the
994   long-term in vitro proliferation of oil palm embryogenic suspension cultures. *Plant Cell Rep.*,
995   **32**, 359–368. Available at: http://link.springer.com/10.1007/s00299-012-1369-y.
996   **Roeder, V., Collén, J., Rousvoal, S., Corre, E., Leblanc, C. and Boyen, C.** (2005) Identification of
997   stress gene transcripts in *Laminaria digitata* (Phaeophyceae) protoplast cultures by expressed
998   sequence tag analysis. *J. Phycol.*, **41**, 1227–1235.
999   **Saier, M.H., Reddy, V.S., Tsu, B. V, Ahmed, M.S., Li, C. and Moreno-Hagelsieb, G.** (2016) The
1000   Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Res.*, **44**, D372-9.
1001   Available at: http://www.ncbi.nlm.nih.gov/pubmed/26546518.
1002   **Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V. and Zdobnov, E.M.** (2015) BUSCO:
1003   assessing genome assembly and annotation completeness with single-copy orthologs.
1004   *Bioinformatics*, **31**, 3210–3212. Available at:

1005   http://bioinformatics.oxfordjournals.org/lookup/doi/10.1093/bioinformatics/btv351.

1006 **Skinner, M.E., Uzilov, A. V, Stein, L.D., Mungall, C.J. and Holmes, I.H.** (2009) JBrowse: a next-
1007   generation genome browser. *Genome Res.*, **19**, 1630–8. Available at:
1008   http://www.ncbi.nlm.nih.gov/pubmed/19570905.

1009 **Stache-Crain, B., Müller, D.G. and Goff, L.J.** (1997) Molecular systematics of *Ectocarpus* and
1010   *Kuckuckia* (Ectocarpales, Phaeophyceae) inferred from phylogenetic analysis of nuclear- and
1011   plastid-encoded DNA sequences. *J. Phycol.*, **33**, 152–168.

1012 **Stanke, M., Steinkamp, R., Waack, S. and Morgenstern, B.** (2004) AUGUSTUS: a web server for
1013   gene finding in eukaryotes. *Nucleic Acids Res.*, **32**, W309-12. Available at:
1014   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=441517&tool=pmcentrez&rende
1015   rtype=abstract.

1016 **Starr, R.C. and Zeikus, J.A.** (1993) Utex - the culture collection of algae at the University of Texas at
1017   Austin: 1993 list of cultures. *J. Phycol.*, **29**, 1–106.

1018 **Swanson, W.J. and Vacquier, V.D.** (2002) The rapid evolution of reproductive proteins. *Nat. Rev.*
1019   *Genet.*, **3**, 137–44. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11836507.

1020 **Tan, S., Cardoso-Moreira, M., Shi, W., et al.** (2016) LTR-mediated retroposition as a mechanism of
1021   RNA-based duplication in metazoans. *Genome Res.*, **26**, 1663–1675. Available at:
1022   http://www.ncbi.nlm.nih.gov/pubmed/27934698.

1023 **Tautz, D. and Domazet-Lošo, T.** (2011) The evolutionary origin of orphan genes. *Nat. Rev. Genet.*,
1024   **12**, 692–702. Available at: http://www.nature.com/articles/nrg3053.

1025 **Teng, L., Fan, X., Xu, D., Zhang, X., Mock, T. and Ye, N.** (2017) Identification of genes under positive
1026   selection reveals differences in evolutionary adaptation between brown-algal species. *Front.*
1027   *Plant Sci.*, **8**, 1429. Available at:
1028   http://journal.frontiersin.org/article/10.3389/fpls.2017.01429/full.

1029 **Tonon, T., Li, Y. and McQueen-Mason, S.** (2017) Mannitol biosynthesis in algae: more widespread
1030   and diverse than previously thought. *New Phytol.*, **213**, 1573–1579. Available at:
1031   http://doi.wiley.com/10.1111/nph.14358.

1032 **Torode, T.A., Marcus, S.E., Jam, M., Tonon, T., Blackburn, R.S., Hervé, C. and Knox, J.P.** (2015)
1033   Monoclonal antibodies directed to fucoidan preparations from brown algae. *PLoS One*, **10**,
1034   e0118366. Available at:
1035   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4333822&tool=pmcentrez&rend
1036   ertype=abstract.

1037 **Tulsiani, D.R., Hubbard, S.C., Robbins, P.W. and Touster, O.** (1982) alpha-D-mannosidases of rat
1038   liver golgi membranes. Mannosidase II is the GlcNAcMAN5-cleaving enzyme in glycoprotein
1039   biosynthesis and mannosidases Ia and IB are the enzymes converting Man9 precursors to
1040   Man5 intermediates. *J. Biol. Chem.*, **257**, 3660–8. Available at:
1041   http://www.ncbi.nlm.nih.gov/pubmed/7061502.

1042 **Verna, J., Lodder, A., Lee, K., Vagts, A. and Ballester, R.** (1997) A family of genes required for
1043   maintenance of cell wall integrity and for the stress response in *Saccharomyces cerevisiae*.
1044   *Proc. Natl. Acad. Sci. U. S. A.*, **94**, 13804. Available at:
1045   http://www.ncbi.nlm.nih.gov/pubmed/9391108.

1046 **Volkov, V.** (2015) Salinity tolerance in plants. Quantitative approach to ion transport starting from
1047   halophytes and stepping to genetic and protein engineering for manipulating ion fluxes. *Front.*
1048   *Plant Sci.*, **6**, 873. Available at:
1049   http://journal.frontiersin.org/Article/10.3389/fpls.2015.00873/abstract.

1050 **Wächter, G.A., Franzblau, S.G., Montenegro, G., Hoffmann, J.J., Maiese, W.M. and Timmermann,**
1051   **B.N.** (2001) Inhibition of *Mycobacterium tuberculosis* growth by saringosterol from *Lessonia*
1052   *nigrescens*. *J. Nat. Prod.*, **64**, 1463. Available at:

1053   https://pubs.acs.org/doi/abs/10.1021/np010101q.

1054   **Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. and Barton, G.J.** (2009) Jalview
1055   Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**,
1056   1189–91. Available at:
1057   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2672624&tool=pmcentrez&rend
1058   ertype=abstract.

1059   **West, J. and Kraft, G.** (1996) *Ectocarpus siliculosus* (Dillwyn) Lyngb. from Hopkins River Falls,
1060   Victoria - the first record of a freshwater brown alga in Australia. *Muelleria*, **9**, 29–33.

1061   **Wu, H.-J., Zhang, Z., Wang, J.-Y., et al.** (2012) Insights into salt tolerance from the genome of
1062   *Thellungiella salsuginea*. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 12219–24. Available at:
1063   http://www.ncbi.nlm.nih.gov/pubmed/22778405.

1064   **Yang, Z.** (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.*, **24**, 1586–
1065   91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17483113.

1066   **Ye, N., Zhang, X., Miao, M., et al.** (2015) *Saccharina* genomes provide novel insight into kelp
1067   biology. *Nat. Commun.*, **6**, 6986. Available at:
1068   http://www.nature.com/ncomms/2015/150424/ncomms7986/full/ncomms7986.html.

1069   **Yoon, H.S., Hackett, J.D., Ciniglia, C., Pinto, G. and Bhattacharya, D.** (2004) A molecular timeline
1070   for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.*, **21**, 809–18. Available at:
1071   http://mbe.oxfordjournals.org/cgi/content/abstract/21/5/809.

1072   **Zeng, X., Long, H., Wang, Z., et al.** (2015) The draft genome of Tibetan hulless barley reveals
1073   adaptive patterns to the high stressful Tibetan Plateau. *Proc. Natl. Acad. Sci. U. S. A.*, **112**,
1074   1095–100. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25583503.

1075   **Zhu, S.-H. and Green, B.R.** (2010) Photoprotection in the diatom *Thalassiosira pseudonana*: Role of
1076   LI818-like proteins in response to high light stress. *Biochim. Biophys. Acta - Bioenerg.*, **1797**,
1077   1449–1457. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0005272810001362.

1078
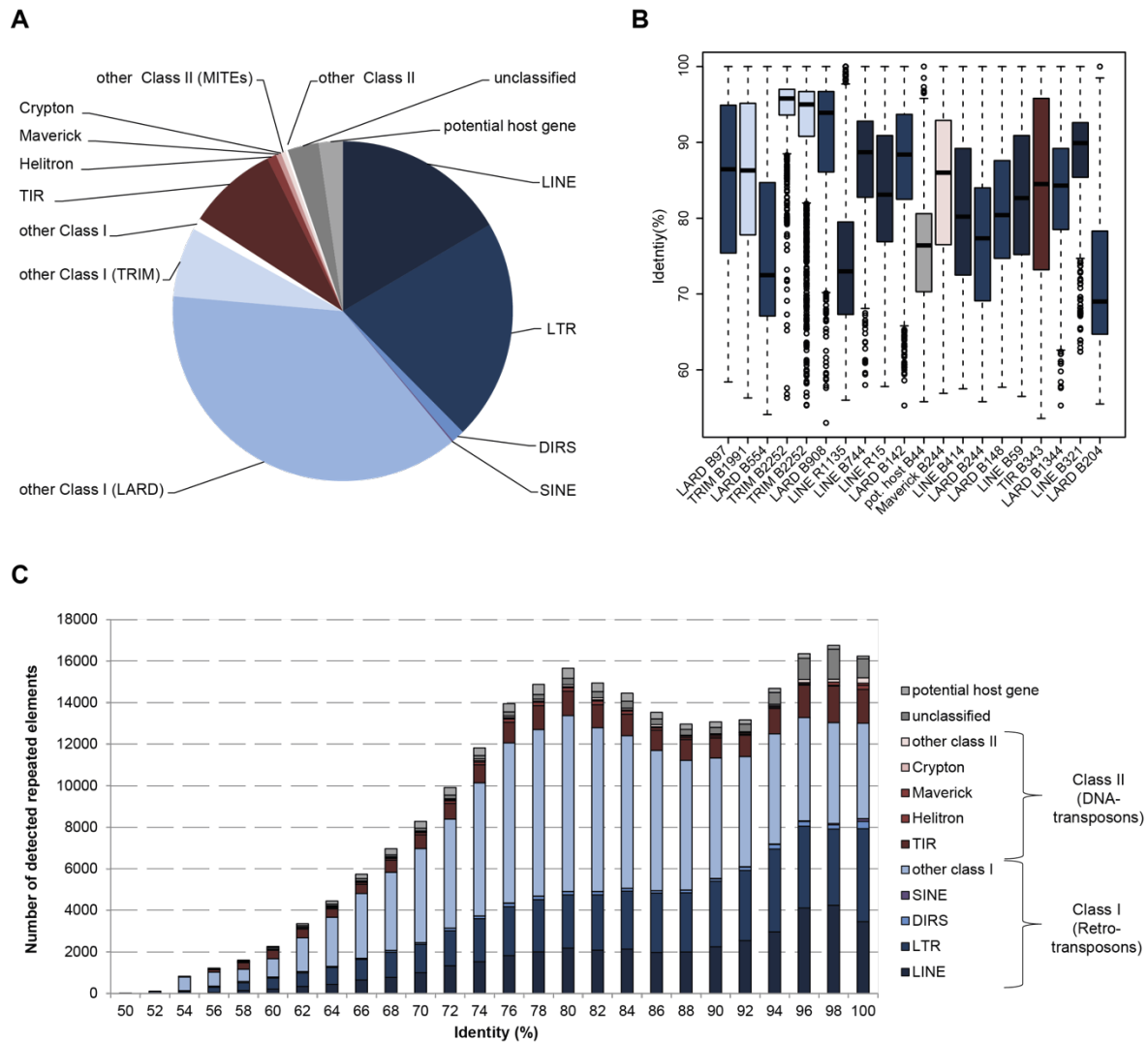
Figures

Figure 1: Repeated elements identified within the genome of *E. subulatus*. A) Different superfamilies detected; B) Boxplot of sequence identity levels for the 20 most abundant transposon families; C) Distribution of sequence identities in the superfamilies.
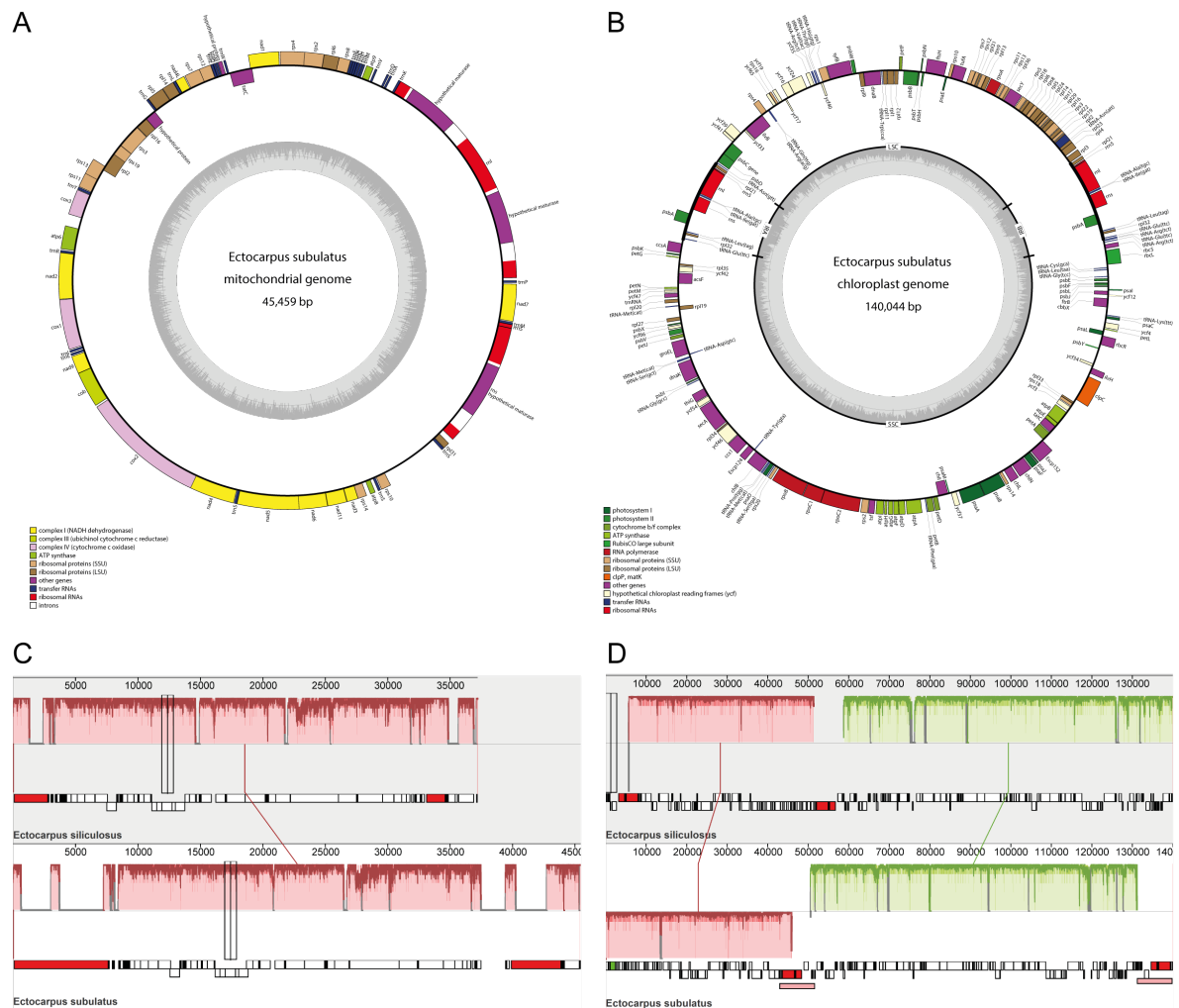
1084
1085 **Figure 2**: Mitochondrial and plastidial genomes of *E. subulatus* Bft15b visualized using
1086 OrganellarGenomeDRAW (Lohse *et al.*, 2013) (panels A and B), and aligned with the *Ectocarpus* sp.
1087 Ec32 organelles using Mauve (panels C and D). In panels C and D blocks of the same color correspond
1088 to orthologous sequences.

29
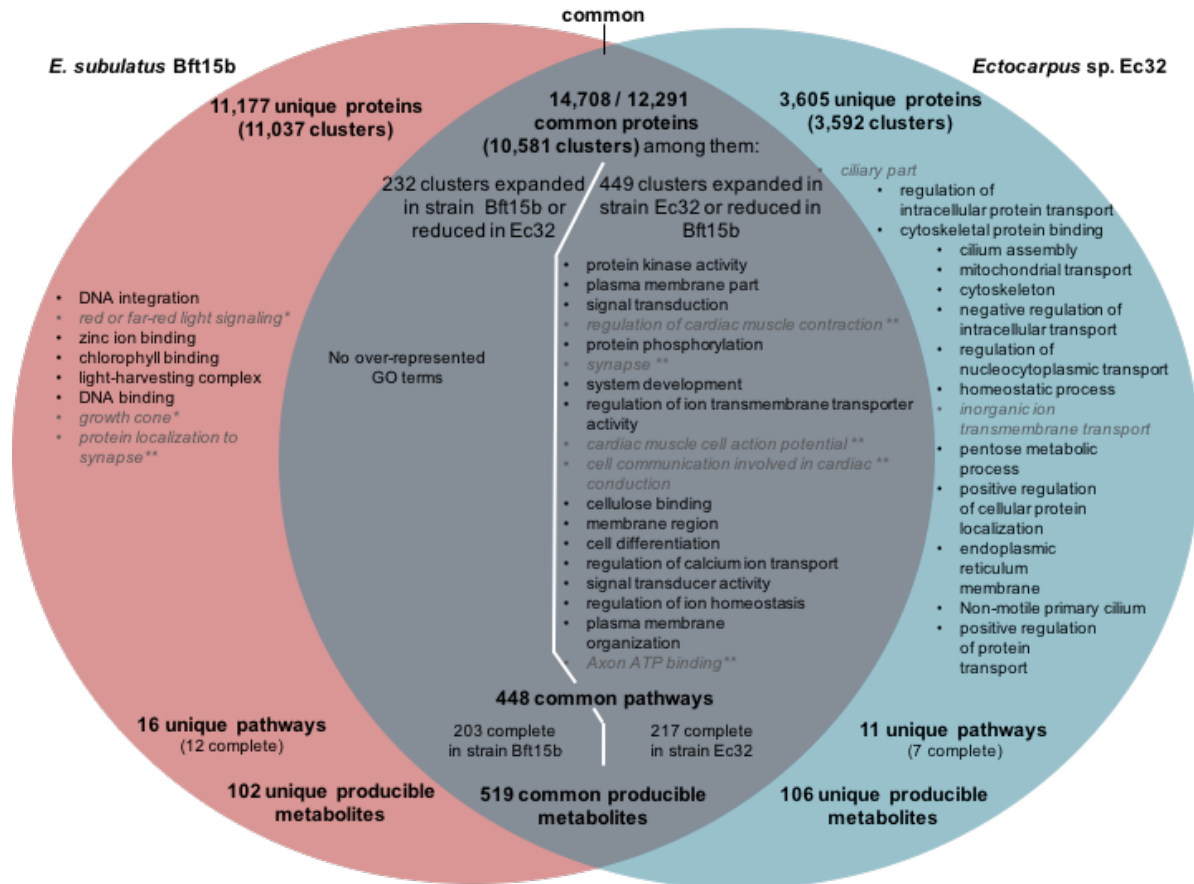
Figure 3: Comparison of gene content and metabolic capacities of *E. subulatus* and *Ectocarpus* sp. Ec32. The top part of the Venn diagram displays the number of predicted proteins and protein clusters unique and common to both genomes in the OrthoFinder analysis. The middle part shows GO annotations significantly enriched (FDR ≤ 0.05) among these proteins. For the common clusters, the diagram also contains the results of gene set enrichment analyses for annotations found among clusters expanded in *E. subulatus* Bft15b and those expanded in *Ectocarpus* sp. Ec32. Functional annotations not directly relevant to the functioning of *Ectocarpus* or shown to be false positives are shown in grey and italics. The bottom part shows the comparison of both genomes in terms of their metabolic pathways.
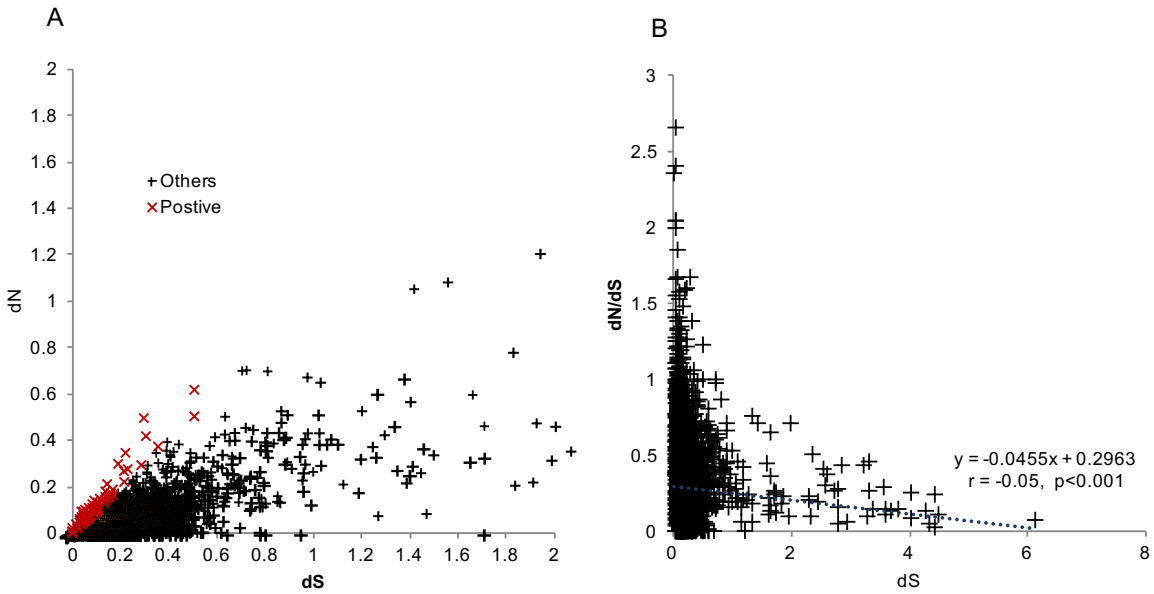
1099

Figure 4: Rate of non-synonymous (dN) to synonymous (dS) mutations in 7,147 orthologous gene pairs with one copy in each examined species of *Ectocarpus*. Gene pairs with dN/dS > 1 are considered to be under positive selection and displayed in red panel A. The resulting ratio was plotted against the rate of synonymous mutations (speed of evolution) and dotted line corresponds to a linear regression (panel B).

1105

Glycosyltransferases

GENES:	E. sp.	E. su	S. ja
GT1	1	1	3
GT2	11	9	24
GT4	13	14	13
GT7	1	1	1
GT8	3	4	5
GT10	1	1	2
GT13	2	2	1
GT14	5	5	1
GT15	1	0	0
GT18	0	1	0
GT20	6	8	4
GT22	3	3	2
GT23	2	2	17
GT24	1	2	0
GT25	3	0	1
GT27	0	0	1
GT28	0	3	4
GT31	3*	5	4
GT33	1	1	0
GT34	1	1	1
GT41	3	3	4
GT47	8	5	14
GT48	3	4	2
GT49	3	2	1
GT50	1	0	1

Halogen metabolism

	E. sp.	E. su	S. ja
vBPO	1	1	17
vIPO	0	0	59
vHPO	5	3	2
TPO	7	5	19
HLD	1	0	0

GENES:	E. sp.	E. su	S. ja
GT54	1	0	0
GT57	2	2	2
GT58	1	1	1
GT59	1	1	0
GT60	3	4	1
GT64	2	2	3
GT65	1	0	0
GT66	1	0	0
GT68	0	0	2
GT74	1	0	1
GT76	1	1	1
GT77	1	0	11
GT90	0	0	1
GT92	0	0	2
GTnc	0	6	0

Alginate metabolism

	E. sp.	E. su	S. ja
MPI	4	4	5
PMM	1	0	1
MPG	0	0	?
GMD	3	2	3
MS	2	1	1
ManC5-E	28	24	105

Fucan metabolism

	E. sp.	E. su	S. ja
FK	2	1	1
GFPP	2	1	1
GM46D	2	2	1
GFS	1	1	1

Sulfatases & sulfotransferases (ST)

	E. sp.	E. su	S. ja
Sulfatases	9	9	n. i.
ST	15	13	24

Glycoside hydrolases

GENES:	E. sp.	E. su	S. ja
GH1	3	2	1
GH2	2	3	3
GH3	1	1	1
GH5	2	2	2
GH10	1	1	1
GH13	0	0	1
GH16	6	7	3
GH17	1	1	2
GH18	0	0	1
GH30	1	0	1
GH31	1	1	0
GH36	2	2	0
GH37	1	1	1
GH38	1	1	0
GH47	5	5	2
GH63	1	1	0
GH81	5	4	53
GH85	1	1	1
GH88	6	3	1
GH95	1	1	0
GH114	0	0	7
GH128	0	0	1

Phenolics

|	E. sp.	E. su	S. ja
PKS III	3	3	2
CHIL	2	2	2
AST	5	3	7
FGT	1	1	1
OMT	1	1	1
POX	1	1	2
TYR	28	20	22

Figure 5: Prevalence of several genes families related to key metabolic pathways in brown algae. *E. sp.*, *Ectocarpus sp.* Ec32; *E. su*, *Ectocarpus subulatus*; *S. ja*, *Saccharina japonica*; GT, glycosyltransferase; GH, glycosyl hydrolase; vBPO, vanadium-dependent bromoperoxidase; vIPO, vanadium-dependent ioperoxidase; vHPO, vanadium-dependent haloperoxidase; TPO, homolog of thyroid peroxidase; HLD, haloalkane dehalogenase; MPI, mannose-6-phosphate isomerase; PMM, phosphomannomutase; MPG, mannose-1-phosphate guanyltransferase; GMD, GDP-mannose 6-dehydrogenase; MS, mannuronan synthase; ManC5-E, mannuronan C5-epimerase; FK, L-fucokinase; GFPP, GDP-fucose pyrophosphorylase; GM46D, GDP-mannose 4,6-dehydratase; GFS, GDP-L-fucose synthetase; ST, sulfotransferase; PKS III, type III polyketide synthase; CHIL, chalcone isomerase-like; AST, aryl sulfotransferase; FGT, flavonoid glucosyltransferase; OMT, flavonoid O-methyltransferase (OMT); POX, polyphenol oxidase; TYR, tyrosinase. *for GT31, three genes have been identified in *Ectocarpus* sp. Ec32 in the present study that have not been initially annotated in the corresponding genome (Cock *et al.*, 2010) and companion paper (Michel *et al.*, 2010). n. i.: The presence of sulfatases in the *S. japonica* genome was not indicated in the corresponding paper (Ye *et al.*, 2015)

1120
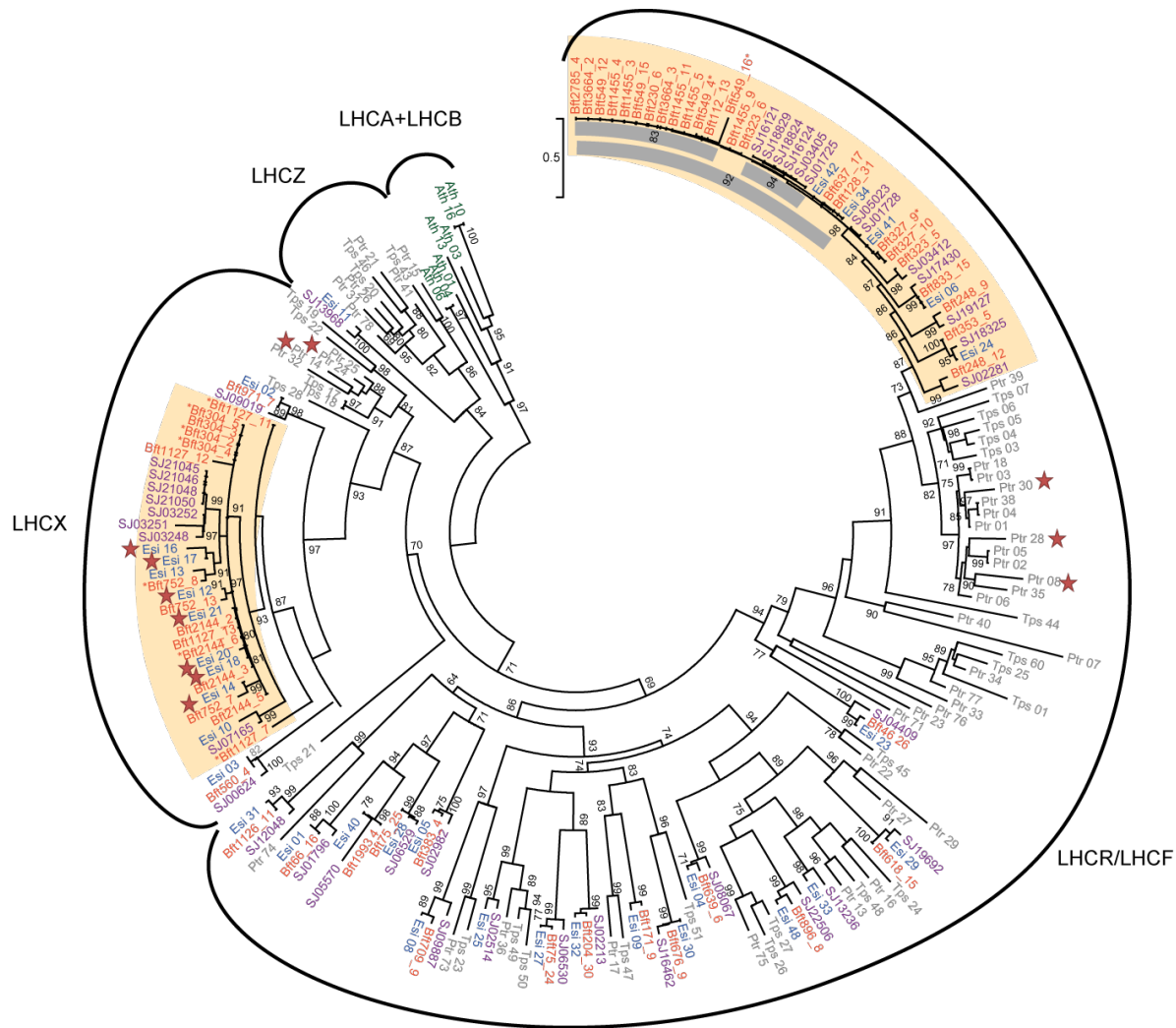
1121

Figure 6: Maximum likelihood tree of *E. subulatus* Bft15b (orange) *Ectocarpus* sp. Ec32 (blue), *S. latissima* (purple), and diatom (*Thalassiosira pseudonana* and *Phaeodactylum tricornutum*, grey) CBP sequences. Support values correspond to bootstrap values from 100 replicate runs. *A. thaliana* sequences (green) were added as outgroup. Accessions for *E. subulatus* are given without the Esu prefix, for *Ectocarpus* sp. Ec32, diatoms and *A. thaliana* see (Dittami *et al.*, 2010). Stars indicate genes that have been previously shown to be stress-induced (Dittami *et al.*, 2010), asterisks next to the protein names indicate incomplete proteins. Probable expansions in *E. subulatus* are indicated by an ocher background.

Tables

Table 1: Assembly statistics of the two available *Ectocarpus* and of the *S. japonica* genomes.

	Ectocarpus Ec32	*E. subulatus* Bft15b	*S. japonica*
Sequencing strategy	Sanger+Bac libraries	Illumina (PE+MP)	Illumina PE+PacBio

33

Genome size (flow cytometry)	214*	227	545**
Genome size (sequenced)	196 Mb	242 Mb	537 Mb**
Sequencing Coverage	11 X[#]	124 X	178 X**
G/C contents	53%	54%	50%
Number of scaffolds >2kb	1,561	5,286	6,985
Scaffold N50 (kb)	497 kb	113 kb	254 kb
Number of predicted genes	17,418	25,939	18,733
Mean number of exons per gene	8.0	5.3	6.5**
Repetitive elements	30%	30%	40%
BUSCO genome completeness	94% (99%*[#])	86% (91%*[#])	91% (96%*[#])
BUSCO Fragmented proteins	7.4%	13.5%	14.2%

1132 * according to (Peters *et al.*, 2004)

1133 ** according to (Ye *et al.*, 2015)

1134 [#] according to (Cock *et al.*, 2010)

1135 [##] 23% according to (Cock *et al.*, 2010), but 30% when re-run with the current version (2.5) of the
1136 REPET pipeline.

1137 *[#] not considering proteins absent from all three brown algal genomes

1138 **Table 2**: Clusters of orthologous genes as identified by OrthoFinder expanded in the genome of *E.*
1139 *subulatus* Bft15b after manual identification of false positives, and removal of clusters without
1140 functional annotation or related to transposon or viral sequences.

Cluster(s)	# Ec32	# Bft15b	Putative annotation or functional domain
Cell-wall related proteins			
OG0000597	1	3	Peptidoglycan-binding domain
OG0000284, -782, -118	6	12	Carbohydrate-binding WSC domain
OG0000889	1	2	Cysteine desulfuration protein
OG0000431	1	3	Galactose-3-O-sulfotransferase (partial)
Transcriptional regulation and translation			
OG0000785	1	2	AN1-type zinc finger protein
OG0000059	4	10	C2H2 zinc finger protein
OG0000884	1	2	Zinc finger domain
OG0000766	1	2	DNA-binding SAP domain
OG0000853	1	2	RNA binding motif protein

34

OG0000171	1	6	Helicase
OG0000819	1	2	Fungal transcriptional regulatory protein domain
OG0000723	1	2	Translation initiation factor eIF2B
OG0000364	2	3	Ribosomal protein S15
OG0000834	1	2	Ribosomal protein S13
Cell-cell communication and signaling			
OG0000967	1	2	Ankyrin repeat-containing domain
OG0000357	2	3	Regulator of G protein signaling domain
OG0000335	2	3	Serine/threonine kinase domain
OG0000291	2	3	Protein kinase
OG0000185	3	4	Octicosapeptide/Phox/Bem1p domain
Others			
OG0000726	1	3	HSP20
OG0000104	1	9	Light harvesting complex protein
OG0000277	3	3	Major facilitator superfamily transporter
OG0000210	2	4	Cyclin-like domain
OG0000721	1	2	Myo-inositol 2-dehydrogenase
OG0000703	1	2	Short-chain dehydrogenase
OG0000749	1	2	Putative Immunophilin
OG0000463	1	3	Zinc-dependent metalloprotease with notch domain

Supporting Information

Supporting Information Table S1: Complete table of manual protein annotations.

Supporting Information Table S2: Domain-based comparisons between *E. subulatus* and *Ectocarpus* sp. Ec32.

Supporting Information Table S3. Comparison of metabolic pathways found in *E. subulatus* and *Ectocarpus* sp. Ec32.

Supporting Information Table S4: Values of dN/dS associated with orthologous gene pairs between *E. subulatus* and *Ectocarpus* sp. Ec32.

Supporting Information Table S5: Complete list of OrthoFinder results.

Supporting Information File S1: Supplementary text concerning the automatic annotation of red-far sensors, CAZymes, polyamine metabolism, and stress response genes in *E. subulatus*.