



**HAL**  
open science

**The Ectocarpus IMMEDIATE UPRIGHT gene encodes a member of a novel 1 family of cysteine-rich proteins that have an unusual distribution across the eukaryotes**

Nicolas N. Macaisne, Fuli Liu, Delphine Scornet, Akira F Peters, Agnieszka Lipinska, Marie-Mathilde Perrineau, Antoine Henry, Martina Strittmatter, Susana Coelho, J. Mark Cock

► **To cite this version:**

Nicolas N. Macaisne, Fuli Liu, Delphine Scornet, Akira F Peters, Agnieszka Lipinska, et al.. The Ectocarpus IMMEDIATE UPRIGHT gene encodes a member of a novel 1 family of cysteine-rich proteins that have an unusual distribution across the eukaryotes. *Development*, 2016, 144, pp.409-418. 10.1242/dev.141523 . hal-01925145

**HAL Id: hal-01925145**

**<https://hal.science/hal-01925145>**

Submitted on 16 Nov 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **The *Ectocarpus IMMEDIATE UPRIGHT* gene encodes a member of a novel**  
2 **family of cysteine-rich proteins that have an unusual distribution across the**  
3 **eukaryotes**

4 **Nicolas Macaisne<sup>1</sup>, Fuli Liu<sup>1,†</sup>, Delphine Scornet<sup>1</sup>, Akira F. Peters<sup>2</sup>, Agnieszka Lipinska<sup>1</sup>,**  
5 **Marie-Mathilde Perrineau<sup>1</sup>, Antoine Henry<sup>1</sup>, Martina Strittmatter<sup>1</sup>, Susana M. Coelho<sup>1</sup>,**  
6 **J. Mark Cock<sup>1,\*</sup>**

7 <sup>1</sup>CNRS, Sorbonne Université, UPMC University Paris 06, Algal Genetics Group, UMR 8227,  
8 Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688,  
9 Roscoff, France, <sup>2</sup>Bezhin Rosko, 29250, Santec, France

10 \* Author for correspondance (cock@sb-roscoff.fr)

11

12 †Current address: Key Laboratory of Sustainable Development of Marine Fisheries, Ministry  
13 of Agriculture, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences,  
14 Qingdao, 266071, China

15

16 Short title: *Ectocarpus IMM* gene

17

18 **KEY WORDS: Brown algae; *Ectocarpus*; horizontal gene transfer; initial cell division;**  
19 **life cycle; virus**

20

21

22

23

24 **ABSTRACT**

25 The sporophyte generation of the brown alga *Ectocarpus* sp. exhibits an unusual pattern of  
26 development compared to the majority of brown algae. The first cell division is symmetrical  
27 and the apical/basal axis is established late in development. In the *IMMEDIATE UPRIGHT*  
28 mutant the initial cell undergoes an asymmetric division to immediately establish the  
29 apical/basal axis. We provide evidence which suggests that this phenotype corresponds to the  
30 ancestral state of the sporophyte. The *IMM* gene encodes a protein of unknown function,  
31 containing a repeated motif also found in the *EsV-1-7* gene of the *Ectocarpus* virus. Brown  
32 algae possess large families of EsV-1-7 domain genes but these genes are rare in other  
33 stramenopiles suggesting that the expansion of this family may have been linked with the  
34 emergence of multicellular complexity. EsV-1-7 domain genes have a patchy distribution  
35 across eukaryotic supergroups and occur in several viral genomes, suggesting possible  
36 horizontal transfer during eukaryote evolution.

37

38 **INTRODUCTION**

39 Multicellular organisms with haploid-diploid life cycles are found in several major eukaryotic  
40 groups including the green lineage (Archaeplastida) and the red and brown macroalgae  
41 (Rhodophyta and Phaeophyceae, respectively). In these organisms, a single genome provides  
42 the genetic information to deploy two different developmental programs during the course of  
43 the life cycle, leading to the construction of the sporophyte and gametophyte generations,  
44 respectively (Cock et al., 2013; Coelho et al., 2007). One consequence of this type of life cycle  
45 is that the emergence of developmental innovations for one generation of the life cycle can  
46 occur without it being necessary to evolve novel developmental regulatory modules *de novo*.  
47 This is because it is also possible to adapt regulatory modules that have evolved to function  
48 during one generation of the life cycle to carry out related functions during the other generation.  
49 An important objective of the developmental biologists that study these organisms has been to  
50 understand the relative contributions of these two processes - developmental innovation and  
51 trans-generation co-option - to the evolution of multicellularity in these species (Dolan, 2009;  
52 Pires and Dolan, 2012; Shaw et al., 2011). In the green lineage, embryophytes (which have  
53 haploid-diploid life cycles) are thought to have evolved from a green algal ancestor with a  
54 haploid life cycle by the addition of a sporophyte generation (Bower, 1890; Celakovsky, 1874;  
55 Dolan, 2009; Haig and Wilczek, 2006; Niklas and Kutschera, 2010; Qiu, 2008; Qiu et al., 2006).  
56 It has been proposed that the regulatory networks that controlled the development of early  
57 embryophyte sporophytes were recruited to a large extent from the gametophyte generation

58 (Dolan, 2009; Niklas and Kutschera, 2010). Support for this viewpoint has come both from  
59 broad comparisons of gametophyte and sporophyte transcriptomes (Nishiyama et al., 2003;  
60 Szovenyi et al., 2011) and from demonstrations that homologues of key regulatory genes in  
61 embryophyte sporophytes play important roles in gametophyte function in bryophytes (Aoyama  
62 et al., 2012; Kubota et al., 2014; Menand et al., 2007; Nishiyama et al., 2003). There are  
63 however exceptions (Szovenyi et al., 2011). For example, the KNOX family of TALE  
64 homeodomain transcription factors are not expressed during the gametophyte generation in  
65 bryophytes and therefore appear to have evolved as sporophyte developmental regulators (Sano  
66 et al., 2005).

67 To more fully understand the relative contributions of developmental innovation and trans-  
68 generation co-option to the evolution of multicellularity in organisms with haploid-diploid life  
69 cycles it would be of interest to investigate this phenomenon in several lineages that have  
70 independently evolved complex multicellularity. Not only would this allow the generality of  
71 inferences from studies of the green lineage to be accessed but it would also allow an evaluation  
72 of the importance of the ancestral state on subsequent evolutionary events. For example, the  
73 brown algae (Phaeophyceae) most probably evolved from an ancestor that alternated between  
74 simple, filamentous sporophyte and gametophyte generations (Kawai et al., 2003; Silberfeld et  
75 al., 2010). If this was the case, then the evolution of novel regulatory systems may have played  
76 a more important role in the emergence of novel developmental mechanisms than co-option of  
77 regulators across generations in this phylogenetic group. Unfortunately, very little is currently  
78 known about developmental processes in the brown algae and, for example, no developmental  
79 regulatory genes have so far been characterised at the molecular level. However, the recent  
80 emergence of the filamentous alga *Ectocarpus* sp. as a model organism for this group (Cock et  
81 al., 2013; Cock et al., 2015; Coelho et al., 2012a) has created a context in which this type of  
82 question can be addressed.

83 *Ectocarpus* sp. has a haploid-diploid life cycle that involves alternation between two  
84 generations, which both consist of uniserate filaments with a small number of different cell  
85 types and bearing simple reproductive structures (Cock et al., 2015). The morphological  
86 similarity of the two generations has allowed mutants affected both in switching between  
87 generations (Coelho et al., 2011) and in generation-related developmental processes (Peters et  
88 al., 2008) to be isolated. The *immediate upright* (*imm*) mutant is particularly interesting because  
89 it has major effects on the early development of the sporophyte generation but causes no visible  
90 phenotype during the gametophyte generation (Peters et al., 2008). In individuals that carry this  
91 mutation, the initial cell of the sporophyte generation undergoes an asymmetrical rather than a

92 symmetrical cell division and produces an upright filament and a rhizoid rather than the  
93 prostrate filament typical of wild type sporophytes (Peters et al., 2008). Individuals that carry  
94 this mutation therefore fail to implement the typical early sporophyte developmental program  
95 and resemble gametophytes but produce the sexual structures of the sporophyte generation at  
96 maturity. The absence of a phenotype during the gametophyte generation suggests that the  
97 developmental program directed by the *IMM* gene may have been a sporophyte-specific  
98 innovation.

99 Here we describe the positional cloning of the *IMM* locus and show that this gene encodes  
100 a protein of unknown function, which shares a novel, repeated motif with a viral protein. The  
101 *IMM* gene is part of a large, rapidly evolving gene family in *Ectocarpus* sp. and species with  
102 identifiable homologues exhibit an unusual distribution across the eukaryotic tree of life.

103

## 104 **RESULTS**

### 105 **Positional cloning of the *IMM* locus**

106 Peters *et al.* (2008) showed that the *imm* mutation behaved as a recessive, Mendelian allele and  
107 was located on an autosome. To map this mutation, a backcrossed descendant of strain Ec137  
108 carrying the *imm* mutation (Peters et al., 2008) was crossed with the out-crossing line Ec568  
109 (Heesch et al., 2010) to generate a segregating family of 1,699 haploid progeny. The *IMM* locus  
110 was then mapped genetically by scanning the genome for linked microsatellite markers and fine  
111 mapping the mutation. To scan for linked markers, a subset of the population (between 30 and  
112 75 individuals) was genotyped for 97 microsatellite markers (Heesch et al., 2010) distributed  
113 at approximately 30 cM intervals along the length of the entire genetic map. Additional markers  
114 were then generated for a region on chromosome 27 that exhibited co-segregation with the  
115 *Imm*<sup>+</sup> phenotype and these were tested against the entire population to fine map the *IMM* locus.  
116 Overall, a total of 121 markers were genotyped (Table S1), allowing the *IMM* locus to be  
117 mapped to a region of 43.7 kbp between coordinates 2,299,499-2,343,206 on chromosome 27  
118 (Fig 1A).

119 To identify the *imm* mutation within the 43.7 kbp interval, this region was amplified as a  
120 series of PCR products and the pooled products sequenced on an Illumina HiSeq platform. As  
121 the *Imm*<sup>-</sup> phenotype is the result of a spontaneous mutation that was not originally present in  
122 the parent strain Ec17, we reconstructed reference sequences for the two parental haplotypes of  
123 this region by sequencing equivalent PCR products amplified from eight wild type haploid  
124 siblings of the *imm* strain Ec137. Comparison of the sequence data for the eight siblings with  
125 that from the Ec137 strain allowed polymorphisms inherited from the diploid parent sporophyte

126 to be distinguished from the causal mutation. Sanger resequencing was used to validate  
127 polymorphisms detected by the Illumina sequencing approach and to generate sequence data  
128 for several short regions that were not covered by the Illumina sequence data. This approach  
129 identified a two base pair deletion within exon five of gene Ec-27\_002610 as the causal  
130 mutation of the Imm<sup>-</sup> phenotype (Fig 1A). No other mutations were detected within the mapped  
131 interval.

132

### 133 **The *IMM* gene encodes a protein of unknown function related to a protein encoded by a** 134 **brown algal virus**

135 The *IMM* gene (locus Ec-27\_002610) is predicted to encode an 862 amino acid (91.8 kDa)  
136 protein, which consists of a long N-terminal domain that shares no similarity with other  
137 *Ectocarpus* sp. genes or genes from other species, plus a C-terminal domain which includes  
138 five imperfect tandem repeats of a 38 amino acid cysteine-rich motif (C-X4-C-X16-C-X2-H-  
139 X12, Fig 1B). The 38 amino acid cysteine-rich motif is very similar to a cysteine-rich repeated  
140 motif found in the EsV-1-7 protein of the *Ectocarpus* virus EsV-1 (Delaroque et al., 2001),  
141 which also contains five of these repeated motifs. Based on this similarity, hereafter we will  
142 refer to the 38 amino acid cysteine-rich motif as an EsV-1-7 repeat.

143 The two base pair deletion in the *imm* mutant causes a frame shift in the part of the gene  
144 that encodes the N-terminal domain. The mutation is predicted to lead to the production of a  
145 418 amino acid protein with a truncated N-terminal domain and possessing no EsV-1-7 repeats  
146 (Fig 1B).

147

### 148 **Disruption of *IMM* function using RNA interference**

149 Recent work has demonstrated that injection of double stranded RNA into zygotes of the brown  
150 alga *Fucus* induces an RNA interference response, leading to knockdown of target gene  
151 expression (Farnham et al., 2013). RNA interference therefore represents a potential approach  
152 to investigate gene function in brown algae but a modification to the *Fucus* protocol was  
153 required because microinjection is not feasible for *Ectocarpus* due to the small size of its cells.  
154 We therefore developed an alternative approach in which synthetic siRNA molecules were  
155 introduced into naked gametes using a transfection reagent (see Materials and Methods for  
156 details).

157 Wild type gametes that fail to fuse with a gamete of the opposite sex can develop  
158 parthenogenetically to give rise to partheno-sporophytes. These partheno-sporophytes go  
159 through the same developmental steps as diploid sporophytes derived from zygotes and are

160 morphologically indistinguishable from the latter. In both cases the initial cell undergoes a  
161 symmetrical division that gives rise to two germ tubes, which grow to form a symmetrical,  
162 prostrate basal filament (Peters et al., 2008). The basal filament, which is composed of  
163 characteristic round and elongated cells, adheres strongly to the substratum. Following  
164 simultaneous introduction of three siRNA molecules targeting the *IMM* gene transcript, a small  
165 proportion ( $0.63\% \pm 0.09\%$  mean  $\pm$  standard deviation in six replicates each of 400 individuals)  
166 of the parthenogenetic gametes adopted a pattern of early development that closely resembled  
167 the phenotype of the *imm* mutant (Peters et al., 2008). These gametes underwent an  
168 asymmetrical rather than a symmetrical initial cell division and the two germ tubes of the  
169 developing partheno-sporophyte gave rise to an upright filament and a rhizoid (Fig 2).  
170 Individuals with this phenotype were not observed in parallel samples of gametes treated with  
171 an siRNA directed against a green fluorescent protein gene sequence as a control (six replicates  
172 of 400 individuals) and the difference between the test and control experiments was highly  
173 significant (Pearson's  $\chi^2 = 13.0$ ,  $p = 0.0003$ ). These observations indicated that RNA-  
174 interference-induced knockdown of *IMM* gene expression had the same developmental  
175 consequences as the *imm* mutation, at least in a small proportion of the treated individuals.  
176 Taken together with the results of the mapping of the genetic mutation, this observation  
177 confirmed that the two base pair deletion identified in exon five of gene Ec-27\_002610 was the  
178 causal mutation of the *Imm*<sup>-</sup> phenotype.

179

### 180 **Expression pattern of the *IMM* gene**

181 Quantitative reverse transcriptase PCR analysis indicated that the *IMM* gene transcript was  
182 approximately twice as abundant in diploid sporophytes and partheno-sporophytes as in  
183 gametophytes (Fig 3). The relatively high abundance of the *IMM* transcript during the  
184 gametophyte generation was surprising because no visible phenotype was detected during this  
185 generation in the *imm* mutant (Peters et al., 2008). The transcript was less abundant in *imm*  
186 mutant partheno-sporophytes than in wild type partheno-sporophytes (Fig 3), suggesting that  
187 the mutation has a destabilising effect on the transcript.

188

### 189 **Analysis of gene expression in the *imm* mutant sporophyte**

190 In a previous study we analysed gene expression in the *imm* mutant using a microarray  
191 constructed with sequences from two subtraction libraries enriched in genes differentially  
192 expressed during either the sporophyte or the gametophyte generation (Peters et al., 2008). This  
193 analysis indicated that sporophyte-specific genes were down regulated and gametophyte-

194 specific genes upregulated in the *imm* mutant during the sporophyte generation. Based on this  
195 information, and the morphological resemblance of the *imm* sporophyte to the wild type  
196 gametophyte, the  $Imm^-$  phenotype was interpreted as representing partial homeotic switching  
197 from the sporophyte to the gametophyte developmental program (Peters et al., 2008). Here we  
198 used multiple RNA-seq datasets to compare the *imm* transcriptome with a broader range of  
199 samples, including two microdissected partheno-sporophyte tissue samples corresponding to  
200 the apical upright filaments and the basal system, respectively. Principal component analysis  
201 indicated that, overall, the transcriptome of the *imm* partheno-sporophyte was actually more  
202 similar to the transcriptomes of wild type partheno-sporophyte samples, particularly samples  
203 that included upright filaments, than to wild type gametophyte samples (Fig 4).

204 We therefore reanalysed the expression patterns of sets of genes that had been previously  
205 identified as being significantly upregulated or down regulated in the *imm* mutant partheno-  
206 sporophyte compared to the wild type partheno-sporophyte (denoted as *imm* upregulated or IUP  
207 and *imm* downregulated or IDW genes by Peters et al., 2008). This analysis indicated that IUP  
208 and IDW genes tended to be up and down regulated, respectively, in gametophyte samples but  
209 they also showed very similar patterns of expression in upright filaments isolated from the  
210 sporophyte generation (Fig S1). Bearing in mind that that the gametophyte generation consists  
211 almost entirely of upright filaments, these results suggested that the IUP and IDW genes, rather  
212 than being life-cycle-regulated genes, may correspond to loci that are differentially regulated  
213 in upright filaments compared to basal tissues.

214 To further investigate this possibility, a genome-wide analysis was carried out using RNA-  
215 seq data and DEseq2 (Love et al., 2014) to identify additional genes that were differentially  
216 expressed in the *imm* partheno-sporophyte compared to the wild type partheno-sporophyte. This  
217 analysis identified 1,578 genes that were significantly differentially expressed between the two  
218 samples (1,087 up-regulated in *imm* and 491 down-regulated; Table S2). Again, analysis of  
219 expression patterns across several different samples indicated that the majority of these genes  
220 did not exhibit life-cycle-generation-specific expression patterns but rather were up- or down-  
221 regulated in upright filaments (Fig S2).

222 Taken together, these observations suggest an alternative interpretation of the  $Imm^-$   
223 phenotype. Rather than representing a mutation that causes switching between life cycle  
224 generations, we propose that abrogation of the *IMM* gene leads to failure to correctly implement  
225 the early developmental program of the sporophyte. In the absence of a functional *IMM* gene  
226 the initial cell does not divide symmetrically and there is no deployment of a system of basal  
227 filaments before the establishment of the apical/basal axis (Fig 2A). Rather, an asymmetrical



228 division of the initial cell directly produces a basal rhizoid cell and an apical thallus cell. We  
229 suggest that the resemblance with the gametophyte, in terms of gene expression, is not due to  
230 switching to the gametophyte developmental program but rather due to the sporophyte adopting  
231 an alternative developmental program which is more similar to that of the gametophyte  
232 (immediate production of an upright filament).

233 The 1,578 genes that were identified as being significantly differentially expressed between  
234 the *imm* partheno-sporophyte and the wild type partheno-sporophyte were analysed for enriched  
235 gene ontology categories. One significant (FDR < 5%) category was found for the upregulated  
236 genes (G-protein signalling) and two (photosynthesis-related and RNA polymerase II activity)  
237 for the downregulated genes (Tables S3 and S4).

238

### 239 **The *IMM* gene is a member of a large gene family in *Ectocarpus* sp.**

240 A search of the *Ectocarpus* sp. genome identified a large family of 91 genes that encode proteins  
241 with at least one EsV-1-7 domain, indicating that *IMM* is part of a large gene family in this  
242 species. The 91 predicted proteins contain between 1 and 19 copies of the cysteine-rich motif  
243 (6 on average), with multiple motifs being organised as adjacent tandem repeats in almost all  
244 cases. Eighty-seven of the 91 predicted proteins (including *IMM*) did not consist solely of EsV-  
245 1-7 domains but contained at least one additional polypeptide region (of at least 25 contiguous  
246 amino acids and often considerably longer) but only two of these additional polypeptide regions  
247 contained known structural domains (a heavy-metal-associated domain and an ABC transporter  
248 ATP-binding domain; Table S5). Clustering analysis indicated that in most cases these  
249 additional domains were unique. Only 19 proteins could be clustered based on similarities  
250 between these additional polypeptide regions (five clusters of 2 to 9 proteins) and most of these  
251 clusters appeared to have arisen as a result of local gene duplications (eight linked duplicated  
252 gene pairs).

253 In addition to the loci described above, five putative pseudogenes were associated with the  
254 EsV-1-7 family (Table S5). Classification as pseudogenes was based either on the presence of  
255 a stop codon within the predicted coding region (one gene) or on failure to detect any evidence  
256 of gene expression (TPM<1) across multiple tissue samples and life cycle stages (four genes).  
257 In addition, we noted that the 91 EsV-1-7 domain genes described above included three bi- or  
258 monoexonic loci with short open reading frames, which are often located near to or within the  
259 untranslated regions of neighbouring genes and may correspond to gene fragments.

260 Estimates of transcript abundances based on RNA-seq data indicated that the members of  
261 the *Ectocarpus* sp. EsV-1-7 domain gene family have diverse expression patterns (Fig S3),

262 suggesting that they carry out diverse functions at different stages of the life cycle and in  
263 different organs.

264

### 265 **IMM orthologues and EsV-1-7 domain proteins in other species**

266 Searches of the recently published *Saccharina japonica* genome (Ye et al., 2015) and multiple  
267 brown algal transcriptomes produced by the onekp project  
268 (<https://sites.google.com/a/ualberta.ca/onekp/>) identified predicted proteins similar to IMM in  
269 a broad range of brown algal species, including members of the Ectocarpales, the Laminariales  
270 and the Fucales. Reciprocal blasts indicated that a number of these proteins were IMM  
271 orthologues and this conclusion was supported by the observation that best reciprocal blast  
272 matches to Ec-17\_002150 (which is the member of the *Ectocarpus* sp. EsV-1-7 family that is  
273 most similar to IMM) formed a distinct cluster in a phylogenetic tree (Fig S4).

274 Analysis of IMM orthologues from five brown algal species (*Ectocarpus* sp., *Scytosiphon*  
275 *lomentaria*, *S. japonica*, *Macrocystis pyrifera*, *Sargassum muticum*, *Sargassum hemiphyllum*  
276 and *Sargassum thunbergi*) using the paired nested site models (M1a, M2a; M7, M8)  
277 implemented in PAML5 (CODEML; Yang, 2007) did not detect any evidence for positive  
278 selection acting on this protein. Pairwise dN/dS values ranged between 0.0229 and 0.2021.

279 The large size of the EsV-1-7 domain family in *Ectocarpus* sp. and the presence of several  
280 pseudogenes and putative gene fragments suggested that this gene family may be undergoing a  
281 process of gene gain and gene loss over time. We therefore searched for orthologues of the  
282 *Ectocarpus* sp. EsV-1-7 domain genes in the *S. japonica* genome (Ye et al., 2015) to investigate  
283 the evolutionary dynamics of the family. The last common ancestor of *Ectocarpus* sp. (order  
284 Ectocarpales) and *S. japonica* (order Laminariales) existed between 80 and 110 mya (Kawai et  
285 al., 2015; Silberfeld et al., 2010). An extensive search, which included *de novo* annotation of  
286 19 new EsV-1-7 domain genes in *S. japonica*, identified orthologues for only 34 of the 91  
287 *Ectocarpus* sp. genes (Table S6), and no orthologues were found for the five *Ectocarpus* sp.  
288 pseudogenes. Moreover, in general when a *S. japonica* orthologue was identified, it was  
289 considerably diverged from its *Ectocarpus* sp. counterpart. These results suggested that the  
290 EsV-1-7 gene families have diverged relatively rapidly since the divergence of the two species.  
291 To evaluate whether the EsV-1-7 genes are evolving more rapidly than other genes in the  
292 genome, we compared the set of percentage identities between orthologous *Ectocarpus* sp. and  
293 *S. japonica* EsV-1-7 proteins with the percentage identities for a set of 9,845 orthologous  
294 protein pairs obtained by reciprocal Blastp comparison of the predicted proteomes of the two  
295 species. Median identity for the orthologous pairs of EsV-1-7 proteins (31.1% identity) was

296 significantly lower (Wilcoxon test  $p = 9.2 \times 10^{-7}$ ) than the genome-wide average (47.9%). Sex-  
297 biased genes have been shown to evolve rapidly in several, diverse species, including brown  
298 algae (Lipinska et al., 2015). The median identity for the EsV-1-7 proteins was also  
299 significantly lower (Wilcoxon test  $p = 7.6 \times 10^{-5}$ ) than that for 905 orthologous pairs of sex-  
300 biased proteins from *Ectocarpus* sp. and *S. japonica* (44.6%). Taken together, these analyses  
301 indicated that the EsV-1-7 family is evolving not only more rapidly than the average gene in  
302 the genome but also more rapidly than the set of sex-biased genes.

303 A more detailed analysis of sequence conservation at the domain level, indicated that the  
304 EsV-1-7 domains (i.e. the regions of the proteins containing the EsV-1-7 repeats) tended to be  
305 markedly more conserved than the other parts of the proteins (Fig S5). The conservation of the  
306 EsV-1-7 domains at the sequence level suggests that these regions correspond to the main  
307 functional features of the proteins and it is possible that the regions outside these domains do  
308 not have specific functional roles in most cases. Note that the IMM orthologues from diverse  
309 brown algal species also exhibited this characteristic, with the EsV-1-7 repeat region being  
310 more strongly conserved across species than the other parts of the protein (e.g. Fig S5).

311 A search of available stramenopile genomes identified single EsV-1-7 domain genes in  
312 each of the *Nannochloropsis gaditana*, *Nannochloropsis oceanica* and *Pythium ultimum*  
313 genomes, although the *P. ultimum* gene contained only one, poorly conserved repeat (Table  
314 S6). No EsV-1-7 domain genes were found in other stramenopile genomes, including those of  
315 the diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricorutum*. The small number of  
316 EsV-1-7 domain genes in stramenopiles other than the brown algae (zero or one gene per  
317 genome), suggests that the diversification of the brown algal EsV-1-7 domain genes occurred  
318 within the brown algal lineage.

319 Searches for homologues in other eukaryotic groups revealed a similar, patchy distribution  
320 of EsV-1-7 domain proteins across the eukaryotic tree (Fig 5A, Table S6). The only other  
321 species in which EsV-1-7 domain proteins were detected were the cryptophyte *Guillardia theta*  
322 (37 genes), the green algae *Coccomyxa subellipsoidea* (four genes) and *Monoraphidium*  
323 *neglectum* (11 genes), and a single gene with one poorly conserved EsV-1-7 domain in the  
324 fungus *Rhizophagus irregularis*. The genome of the *Ectocarpus* virus EsV-1 contains a single  
325 EsV-1-7 domain gene and single homologues were also found in two distantly related viruses,  
326 *Pithovirus sibericum* (which infects *Acanthamoeba*) and *Emiliana huxleyi* virus PS401.

327 The sets of stramenopile, cryptophyte and chlorophyte EsV-1-7 domain genes each  
328 exhibited characteristic features in terms of the conserved residues within the repeated cysteine-  
329 rich motif (Fig 5B), indicating that each group of genes diversified independently.

330

331 **DISCUSSION**

332 In most brown algal species the first cell division is asymmetrical and defines not only the  
333 apical/basal axis but also establishes the separate apical and basal lineages that will give rise,  
334 respectively, to the thallus and to anchoring structures such as a rhizoid or a holdfast (Fritch,  
335 1959). The *Ectocarpus* sporophyte is unusual in that the first cell division is symmetrical and a  
336 system of basal filaments is deployed before the apical/basal axis is established. A functional  
337 *IMM* locus is necessary for this developmental program to be implemented. When *IMM* is  
338 mutated a more canonical program of early development is observed involving an asymmetrical  
339 initial cell division and formation of a basal rhizoid and an apical upright filament. We  
340 originally proposed, based on phenotypic and gene expression analyses, that the *Imm*<sup>-</sup>  
341 phenotype represented partial switching from the sporophyte to the gametophyte developmental  
342 program but the more detailed analysis carried out in this study suggests rather that the *Imm*<sup>-</sup>  
343 phenotype corresponds to a modified version of the sporophyte developmental program. We  
344 suggest, based on these results, that the developmental program of the *Ectocarpus* sporophyte  
345 evolved through a modification of an ancestral developmental program which more closely  
346 resembled that of the majority of brown algae, i.e. early establishment of the apical/basal axis  
347 and limited deployment of basal structures.

348       Recent work on the life cycles of *Ectocarpus* species under field conditions suggests that  
349 the sporophyte generation tends to persist for longer and to be the stage of the life cycle that  
350 allows the species to overwinter (Couceiro et al., 2015). The gametophyte generation, in  
351 contrast, tends to be ephemeral during the spring months and to have principally a reproductive  
352 role. Based on these observations, the developmental program of the sporophyte could be  
353 interpreted as an adaptation to this generation's role in the field. Delaying establishment of the  
354 apical/basal axis will tend to delay the transition to reproductive maturity because the  
355 reproductive structures develop on the apical upright filaments. Delayed reproduction may be  
356 advantageous, for overwintering for example. Moreover, the establishment of a network of  
357 strongly attached basal filaments comprised of cells with thick cell walls could also contribute  
358 to survival during seasons that are non-optimal for growth and reproduction.

359       Interestingly, no mutant phenotype was observed in *imm* gametophytes (Peters et al.,  
360 2008), suggesting that the processes regulated by *IMM* evolved specifically within the context  
361 of the sporophyte generation. The unusual early development of the sporophyte therefore  
362 represents an example of a major developmental innovation (modified timing of major axis  
363 formation, transition from asymmetrical to symmetrical initial cell division) that has evolved

364 specifically for one generation of the life cycle. Surprisingly, although the *IMM* gene was  
365 expressed most strongly in the sporophyte, transcripts also accumulated, albeit at a lower level,  
366 during the gametophyte generation of the life cycle. Therefore, it is not clear at present how the  
367 effects of this gene are restricted to one generation.

368 The *IMM* protein sequence does not contain any regions that are similar to previously  
369 defined protein domains but the C-terminal end of the protein contains several copies of a motif  
370 that we refer to as an EsV-1-7 repeat based on the presence of similar repeats in the *EsV-1-7*  
371 gene of the *Ectocarpus* virus EsV-1. The large phenotypic changes and the extensive  
372 transcriptome reprogramming observed in the *imm* mutant indicate that *IMM* is likely to have  
373 a regulatory role, presumably mediated by this conserved domain, but further work will be  
374 required to determine how the protein functions at the molecular level. The four highly  
375 conserved cysteine and histidine residues in each EsV-1-7 repeat are reminiscent of several  
376 classes of zinc finger motif and one interesting possibility that would merit further investigation  
377 is that the motif represents a new class of zinc finger motif.

378 The origin of the *IMM* gene itself is also of interest. All the brown algae analysed possess  
379 multiple EsV-1-7 domain genes and detailed comparison of the EsV-1-7 domain gene families  
380 in *Ectocarpus* sp. and *S. japonica* indicated that the family has been evolving rapidly with  
381 considerable gene gain and gene loss during the diversification of the Ectocarpales and  
382 Laminariales. In contrast, EsV-1-7 domain genes are rare or absent from other stramenopile  
383 lineages (a single gene in eustigmatophytes and some oomycete species, none in diatoms). This  
384 observation suggests that the expansion of this family occurred after the brown algal lineage  
385 diverged from other major stramenopiles and was perhaps linked with important events within  
386 this lineage such as the emergence of complex multicellularity. The presence of *IMM*  
387 orthologues in diverse brown algae indicates that this gene evolved early during the  
388 diversification of this group but presumably *IMM* was only later co-opted to direct the unusual  
389 pattern of early sporophyte development observed in *Ectocarpus* sp. (i.e. during the  
390 diversification of the Ectocarpales).

391 Outside the stramenopiles, EsV-1-7 domain proteins were only found in two chlorophyte  
392 species and one cryptophyte species, with a possible single gene in the fungus *Rhizophagus*  
393 *irregularis*. This patchy distribution of EsV-1-7 domain genes across the eukaryotic tree is  
394 difficult to reconcile with vertical inheritance from a common ancestor. Given that 1) the short  
395 EsV-1-7 repeat is the only motif that is conserved across major lineages, 2) that this motif does  
396 show some sequence variation and 3) that the repeat motifs of each lineage have their own  
397 distinct, conserved characteristics (Fig 5B), it is possible that the gene families of each major

398 eukaryotic lineage evolved independently. However, this process would have had to involve  
399 remarkable, multiple convergences towards the use of several highly conserved residues in a  
400 specific configuration in each distinct lineage. In this context, it is interesting to note that EsV-  
401 1-7 domain genes were found in three diverse viral genomes and it is tempting to speculate that  
402 the highly unusual distribution of this gene family in extant eukaryotic lineages is the result, at  
403 least in part, of ancient horizontal transfers of EsV-1-7 domain genes due to cross-species viral  
404 infections.

405

## 406 **MATERIALS AND METHODS**

### 407 ***Ectocarpus* strains and growth conditions**

408 The *Ectocarpus* strains used to map the *imm* mutation are described in the next section. The  
409 *imm* mutant strain Ec419 was generated by crossing the original *imm* mutant strain (Ec137;  
410 (Peters et al., 2008) with a wild type sister Ec25 and selecting gametophyte descendants that  
411 produced Imm<sup>-</sup> partheno-sporophytes. The *imm* gametophyte generation was generated from  
412 these *imm* partheno-sporophytes by inducing the production of meio-spores. The wild type  
413 individuals used in this study were strain Ec32, a brother of the original *imm* mutant strain  
414 Ec137. *Ectocarpus* was cultivated as described previously (Coelho et al., 2012b).

415

### 416 **Genetic mapping of the *IMM* locus**

417 The *imm* mutation was originally detected in strain Ec137, a male descendant of a diploid  
418 sporophyte Ec17 isolated at San Juan de Marcona, Peru (Peters et al., 2008). Ec137 was crossed  
419 with a sister, Ec25, to generate the diploid sporophyte Ec372, which gave rise to a male  
420 gametophyte Ec420. To map the *imm* mutation, Ec420 was crossed with Ec568 and the  
421 resulting diploid sporophyte Ec700 gave rise to a population of 1,699 gametophyte progeny, all  
422 derived from independent meiotic events (i.e. isolated from independently micro-dissected  
423 unilocular sporangia). Initially, 30 individuals from this segregating population were genotyped  
424 with 97 microsatellite markers distributed at approximately 30 cM intervals along the length of  
425 the entire genetic map (Heesch et al., 2010). Additional microsatellite markers were developed  
426 based on the Ec32 genome sequence to fine map the mutant locus (Table S1).

427

### 428 **Quantitative RT-PCR analysis of mRNA abundance**

429 Total RNA was extracted as previously described (Coelho et al., 2011) from five or six  
430 biological replicates each of *imm* mutant partheno-sporophytes (strain Ec419), wild type  
431 partheno-sporophytes (strain Ec32), wild type, heterozygous diploid sporophytes (strain Ec17)

432 and wild type gametophytes (strain Ec32). These samples were not the same as those used for  
433 the RNA-seq analysis. DNase treatment using the TURBO DNA-free kit (Applied Biosystems)  
434 was carried out to eliminate any contaminating genomic DNA. RNA concentration and quality  
435 was then determined by spectrophotometry and agarose gel electrophoresis. Between 0.2 and  
436 2.0  $\mu\text{g}$  of total RNA was reverse-transcribed to cDNA with the SuperScript First-Strand  
437 Synthesis System (Invitrogen).

438 *IMM* cDNA was amplified with primers E657Q4F  
439 (GGGGTTTGGGTGGAAGAGGACC) and E657Q4R (CGGCGTGGAAGCTGCCTGGTAT)  
440 and *ELONGATION FACTOR 1 $\alpha$*  (*EF1 $\alpha$* ) cDNA was amplified as an internal reference with  
441 primers EF1adeL (CAAGTCCGTCGAGAAGAAGG) and EF1autrL  
442 (CCAGCAACACCACAATGTCT). Quantitative RT-PCR was carried out using the ABsolute  
443 QPCR SYBR Green ROX Mix (ThermoScientific) in a Chromo4 thermocycler (Bio-Rad  
444 Laboratories) and data were analysed with the Opticon monitor 3 software (Bio-Rad  
445 Laboratories). Amplification specificity was checked using a dissociation curve. The  
446 amplification efficiency was tested using a genomic dilution series and was always between  
447 90% and 110%. A standard curve was established for each gene using a range of dilutions of  
448 *Ectocarpus* sp. genomic DNA (between 80 and 199,600 copies) and gene expression level was  
449 normalized against the *EF1 $\alpha$*  reference gene. Two technical replicates were carried out for the  
450 standard curves and three technical replicates for the samples. The data shown correspond to  
451 the mean  $\pm$  the standard deviation for the biological replicates.

452

### 453 **RNA interference**

454 Small interfering RNAs (siRNAs) directed against the *IMM* gene transcript were designed using  
455 version 3.2 of E-RNAi (Horn and Boutros, 2010). The specificity of the designed siRNAs was  
456 determined by comparing the sequence (Blastn) with complete genome and transcriptome  
457 sequences. Candidates that matched, even partially, genomic regions or transcripts in addition  
458 to *IMM* were rejected. Three siRNAs with predicted high specificity corresponding to different  
459 positions along the *IMM* transcript were selected (Table S7). The control siRNA directed  
460 against jellyfish GFP (Caplen et al., 2001) was obtained from Eurogentec. siRNAs were  
461 introduced into *Ectocarpus* sp. strain Ec32 gametes using the transfection reagent HiPerFect  
462 (Qiagen, Valencia, USA). One microlitre each of 0.5  $\mu\text{g}/\mu\text{l}$  solutions of HPLC purified siRNAs  
463 in 1X Universal siMAX<sup>®</sup> siRNA Buffer (MWG Eurofins, Ebersberg, Germany) was mixed  
464 with 12  $\mu\text{l}$  of HiPerFect transfection reagent in a final volume of 100  $\mu\text{l}$  of natural seawater,  
465 vortexed to mix and incubated for 10 min at room temperature before being added dropwise to

466 100 µl of freshly released gametes (Coelho et al., 2012b) in natural seawater in a Petri dish.  
467 After rotating gently to mix, the Petri dish was incubated overnight at 13°C. The following day  
468 10 ml of Provasoli-enriched natural seawater was added and incubation continued at 13°C.  
469 RNAi-induced phenotypes were observed under a light microscope and the number of  
470 individuals that resembled the *imm* mutant were scored in lots of 400 individuals each for six  
471 experimental replicates. Control treatments were carried out in the same manner using a siRNA  
472 directed against a transcript that is not found in *Ectocarpus* sp. gametes (green fluorescent  
473 protein).

474

### 475 **Transcriptome analysis**

476 RNA-seq analysis was carried out to compare the transcriptome of the *imm* mutant with wild  
477 type individuals corresponding to either the sporophyte or gametophyte generations of the life  
478 cycle. Biological replicates (duplicates) were cultivated in 90 mm Petri dishes in natural sea  
479 water supplemented with 0.01% Provasoli enrichment (Starr and Zeikus, 1993) under a 12 hour  
480 light:12 hour dark cycle of white fluorescent light (10-30 mol m<sup>-2</sup>s<sup>-1</sup> photon fluence rate).  
481 Total RNA was extracted from between 0.05 and 0.1 g of tissue as described previously (Coelho  
482 et al., 2011). The total RNA was quantified and cDNA, synthesised from an oligo-dT primer  
483 for each replicate, was independently fragmented, prepared and sequenced by Fasteris (CH-  
484 1228 Plan-les-Ouates, Switzerland) using an Illumina HiSeq 2000 platform to generate 100 bp  
485 single-end reads. The quality of the sequence data was assessed using the FASTX toolkit  
486 ([http://hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html)) and the reads were trimmed and filtered  
487 using a quality threshold of 25 (base calling) and a minimal size of 60 bp. Only reads in which  
488 more than 75% of the nucleotides had a minimal quality threshold of 20 were retained. Filtered  
489 reads were mapped to the *Ectocarpus* sp. genome (Cock et al., 2010b), available at ORCAE  
490 (Sterck et al., 2012), using TopHat2 with the Bowtie2 aligner (Kim et al., 2013) and the mapped  
491 sequencing data were processed with HTSeq (Anders et al., 2014) to count the numbers of  
492 sequencing reads mapped to exons. Expression values were represented as Transcripts Per  
493 Kilobase Million (TPM). Genes with TPMs of <1 in all samples were considered not to be  
494 expressed.

495 Heat maps schematically representing gene transcript abundances were generated using the  
496 log<sub>2</sub> transformed TPM values centred by gene in Cluster 3.0 (de Hoon et al., 2004) and  
497 Treeview 1.1.6 (Saldanha, 2004).

498 Principal component analysis (PCA) of transcript abundances was carried using log<sub>2</sub>  
499 transformed TPM expression values in R (prcomp) and visualized with the rgl package for R



500 (R Development Core Team, 2009). Genes with TPMs <1 were considered not expressed and  
501 their log<sub>2</sub> values were set to 10e-5 to remove noise from the data. Only genes with divergent  
502 expression patterns were used for the whole transcriptome PCA analysis (4,381 genes). These  
503 genes were chosen based on a calculation that estimated expression variance across all libraries  
504 ( $\Delta\text{TPM} = \max \text{TPM} - \min \text{TPM}$ ) and the third quartile (harbouring genes with  $\Delta\text{TPM} > 32$ ) was  
505 chosen for the PCA calculation.

506 Analysis of differential gene expression was carried out using the DESeq2 package  
507 (Bioconductor; Anders and Huber, 2010) using an adjusted P-value cut-off of 0.1 and a minimal  
508 fold-change of 2. Differentially expressed genes are listed in Table S2.

509

### 510 **Identification and analysis of the *Ectocarpus* sp. EsV-1-7 domain gene family**

511 *Ectocarpus* sp. EsV-1-7 domain proteins were identified by iteratively blasting IMM and other  
512 EsV-1-7 domain protein sequences against the predicted proteome (Blastp). Clusters of closely  
513 related genes within the *Ectocarpus* sp. EsV-1-7 domain family were identified by generating  
514 a similarity network based on comparisons of the non-repeat regions of the proteins with the  
515 EFI-EST similarity tool (<http://efi.igb.illinois.edu/efi-est/index.php>).

516

### 517 **Searches for *IMM* orthologues and members of the EsV-1-7 family in other genomes**

518 *Saccharina japonica* EsV-1-7 domain genes were identified either by blasting *Ectocarpus* sp.  
519 protein sequences against the predicted proteome (Blastp) or against the *S. japonica* genome  
520 sequence (tBlastn) (Ye et al., 2015). The coding regions of novel EsV-1-7 domain genes  
521 detected in the genome (19 genes) were assembled using GenomeView (Abeel et al., 2012) and  
522 the publically available genome and RNA-seq sequence data (Ye et al., 2015). The same  
523 approach was used to improve the gene models for five of the EsV-1-7 domain genes previously  
524 reported by Ye et al. (2015; indicated by adding "mod" for modified to the protein identifier).  
525 Orthology between *Ectocarpus* sp. and *S. japonica* EsV-1-7 family proteins was determined  
526 using reciprocal blast analysis combined with manual comparisons of protein alignments. To  
527 analyse the rate of evolution of the EsV-1-7 family of proteins, a set of 9,845 orthologues pairs  
528 was first identified by comparing the complete predicted proteomes of *Ectocarpus* sp. and *S.*  
529 *japonica* and retaining reciprocal best Blastp matches. Global percentage identities based on  
530 alignments of the full protein sequences were then calculated for pairs of orthologous proteins  
531 using EMBOSS Needle (Li et al., 2015) and the set of percentage identities for the 34  
532 orthologous pairs of EsV-1-7 proteins was compared with that of the complete set of 9,845  
533 orthologue pairs. To compare the EsV-1-7 proteins with known fast-evolving proteins, a similar

534 analysis was then carried out with a set of 905 *Ectocarpus* sp. sex-biased genes that have one-  
535 to-one orthologues in *S. japonica* (Lipinska et al., 2015). The statistical significance of  
536 differences between datasets was evaluated with a Wilcoxon test and bootstrap resampling with  
537 replacement (10,000 replicates). EMBOSS Needle was also used to calculate shared identity  
538 between individual domains of orthologous EsV-1-7 family proteins from *Ectocarpus* sp. and  
539 *S. japonica*.

540 Searches for proteins related to IMM from additional species were carried out against the  
541 NCBI, Uniprot and onekp (<https://sites.google.com/a/ualberta.ca/onekp/>) databases using both  
542 Blastp and HMMsearch, the latter using both an alignment of *Ectocarpus* IMM homologues  
543 and an alignment of brown algal IMM orthologues. In addition, HMMsearch and tBlastn or  
544 Blastp searches were carried out against the following genomes and complete deduced  
545 proteomes: *Thalassiosira pseudonana* (diatom; Thaps3 assembled and unmapped scaffolds,  
546 <http://genome.jgi-psf.org/Thaps3/Thaps3.download.ftp.html>) (Armbrust et al., 2004),  
547 *Phaeodactylum tricornutum* (diatom; Phatr2 assembled and unmapped scaffolds,  
548 <http://genome.jgi-psf.org/Phatr2/Phatr2.download.ftp.html>) (Bowler et al., 2008), *Aureococcus*  
549 *anophagefferens* (Pelagophyceae; [http://genome.jgi-  
550 psf.org/Auran1/Auran1.download.ftp.html](http://genome.jgi-psf.org/Auran1/Auran1.download.ftp.html)) (Gobler et al., 2011), *Emiliana huxleyi*  
551 (haptophyte; <http://genome.jgi.doe.gov/Emihu1/Emihu1.download.ftp.html>) (Read et al., 2013),  
552 *Chlorella variabilis* NC64A (chlorophyte; JGI  
553 [http://genome.jgi.doe.gov/pages/dynamicOrganismDownload.jsf?organism=ChlNC64A\\_1](http://genome.jgi.doe.gov/pages/dynamicOrganismDownload.jsf?organism=ChlNC64A_1))  
554 (Blanc et al., 2010), *Monoraphidium neglectum* SAG 48.87 (chlorophyte; NCBI RefSeq  
555 assembly GCF\_000611645.1) (Bogen et al., 2013) and *Bathycoccus prasinus* (chlorophyte;  
556 <http://bioinformatics.psb.ugent.be/orcae/overview/Bathy>) (Moreau et al., 2012). Finally,  
557 iterative searches using EsV-1-7 domain proteins from phylogenetically distant species were  
558 carried out against the NCBI and Uniprot databases to identify distantly related members of  
559 this gene family.

560 Phylogenetic trees were constructed based on ClustalW-generated alignments using the  
561 Maximum likelihood approach (PhyML) implemented in Seaview (Gouy et al., 2010).

562

## 563 **Statistics**

564 Sample sizes were chosen to allow adequate downstream statistical analysis.

565

## 566 **Data Availability**

567 Sequence data from this article can be found in the Sequence Read Archive under accession  
568 number SRP037532 (see Table S8 for details).

569

### 570 **Acknowledgements**

571 We thank Toshiki Uji for providing RNA-seq data.

572

### 573 **Competing interests**

574 The authors declare no competing or financial interests.

575

### 576 **Author Contributions**

577 J.M.C., and S.M.C. designed the research. N.M., A.F.P., F.L. and D.S. performed the positional  
578 cloning and characterisation of the *IMM* gene. D.S., S.M.C., M.S., and A.H. developed and  
579 applied the RNAi knockdown approach. A.L., M.M.P. and J.M.C., analysed data. J.M.C. wrote  
580 the article.

581

### 582 **Funding**

583 This work was supported by the Centre National de la Recherche Scientifique, the Agence  
584 Nationale de la Recherche (project Bi-cycle ANR-10-BLAN-1727 and project Idealg ANR-10-  
585 BTBR-04-01), the Interreg program France (Channel)-England (project Marinexus), the  
586 University Pierre et Marie Curie and the European Research Council (grant agreement 638240).  
587 F.L. and M.M.P. were supported by grants from the China Scholarship Council and the Brittany  
588 Region (SAD program), respectively.

589

### 590 **Supplementary information**

591 Supplementary information available online at  
592 <http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.133678/-/DC1>

593

### 594 **References**

- 595 **Abeel, T., Van Parys, T., Saeys, Y., Galagan, J. and Van de Peer, Y. (2012).**  
596 GenomeView: a next-generation genome browser. *Nucleic Acids Res* **40**, e12.
- 597 **Anders, S. and Huber, W. (2010).** Differential expression analysis for sequence count  
598 data. *Genome Biol* **11**, R106.
- 599 **Anders, S., Pyl, S. T. and Huber, W. (2014).** HTSeq — A Python framework to work  
600 with high-throughput sequencing data. *BioRxiv Prepr.*

- 601 **Aoyama, T., Hiwatashi, Y., Shigyo, M., Kofuji, R., Kubo, M., Ito, M. and Hasebe, M.**  
602 (2012). AP2-type transcription factors determine stem cell identity in the moss  
603 *Physcomitrella patens*. *Development* **139**, 3120–9.
- 604 **Armbrust, E., Berges, J., Bowler, C., Green, B., Martinez, D., Putnam, N., Zhou, S.,**  
605 **Allen, A., Apt, K., Bechner, M., et al.** (2004). The genome of the diatom  
606 *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**, 79–  
607 86.
- 608 **Blanc, G., Duncan, G., Agarkova, I., Borodovsky, M., Gurnon, J., Kuo, A., Lindquist, E.,**  
609 **Lucas, S., Pangilinan, J., Polle, J., et al.** (2010). The *Chlorella variabilis* NC64A  
610 genome reveals adaptation to photosymbiosis, coevolution with viruses, and  
611 cryptic sex. *Plant Cell* **22**, 2943–55.
- 612 **Bogen, C., Al-Dilaimi, A., Albersmeier, A., Wichmann, J., Grundmann, M., Rupp, O.,**  
613 **Lauersen, K. J., Blifernez-Klassen, O., Kalinowski, J., Goesmann, A., et al.**  
614 (2013). Reconstruction of the lipid metabolism for the microalga *Monoraphidium*  
615 *neglectum* from its genome sequence reveals characteristics suitable for biofuel  
616 production. *BMC Genomics* **14**, 926.
- 617 **Bower, F. O.** (1890). On antithetic as distinct from homologous alternation of  
618 generations in plants. *Ann Bot* **4**, 347–370.
- 619 **Bowler, C., Allen, A., Badger, J., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U.,**  
620 **Martens, C., Maumus, F., Otilar, R., et al.** (2008). The *Phaeodactylum* genome  
621 reveals the evolutionary history of diatom genomes. *Nature* **456**, 239–44.
- 622 **Caplen, N. J., Parrish, S., Imani, F., Fire, A. and Morgan, R. A.** (2001). Specific  
623 inhibition of gene expression by small double-stranded RNAs in invertebrate and  
624 vertebrate systems. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 9742–9747.
- 625 **Celakovsky, L.** (1874). Ueber die verschiedenen Formen und die Bedeutung des  
626 Generationwechsels der Pflanzen. *Sitzungsberichte Koeniglichen Boehmischen Ges.*  
627 *Wiss. Prague* **2**, 21–61.
- 628 **Cock, J. M. and Coelho, S. M.** (2011). Algal models in plant biology. *J. Exp. Bot.* **62**, 2425–  
629 2430.
- 630 **Cock, J. M., Sterck, L., Rouzé, P., Scornet, D., Allen, A. E., Amoutzias, G., Anthouard,**  
631 **V., Artiguenave, F., Aury, J.-M., Badger, J. H., et al.** (2010a). The *Ectocarpus*  
632 genome and the independent evolution of multicellularity in brown algae. *Nature*  
633 **465**, 617–621.
- 634 **Cock, J. M., Sterck, L., Rouzé, P., Scornet, D., Allen, A. E., Amoutzias, G., Anthouard,**  
635 **V., Artiguenave, F., Aury, J., Badger, J., et al.** (2010b). The *Ectocarpus* genome  
636 and the independent evolution of multicellularity in brown algae. *Nature* **465**,  
637 617–21.
- 638 **Cock, J. M., Godfroy, O., Macaisne, N., Peters, A. F. and Coelho, S. M.** (2013). Evolution  
639 and regulation of complex life cycles: a brown algal perspective. *Curr Opin Plant*  
640 *Biol* **17**, 1–6.

- 641 **Cock, J. M., Godfroy, O., Strittmatter, M., Scornet, D., Uji, T., Farnham, G., Peters, A. F.**  
642 **and Coelho, S. M.** (2015). Emergence of *Ectocarpus* as a model system to study  
643 the evolution of complex multicellularity in the brown algae. In *Evolutionary*  
644 *transitions to multicellular life* (ed. Ruiz-Trillo, I.) and Nedelcu, A. M.), pp. 153–  
645 162. Dordrecht: Springer.
- 646 **Coelho, S., Peters, A., Charrier, B., Roze, D., Destombe, C., Valero, M. and Cock, J.**  
647 (2007). Complex life cycles of multicellular eukaryotes: new approaches based on  
648 the use of model organisms. *Gene* **406**, 152–70.
- 649 **Coelho, S. M., Godfroy, O., Arun, A., Le Corguillé, G., Peters, A. F. and Cock, J. M.**  
650 (2011). *OUROBOROS* is a master regulator of the gametophyte to sporophyte life  
651 cycle transition in the brown alga *Ectocarpus*. *Proc Natl Acad Sci U A* **108**, 11518–  
652 11523.
- 653 **Coelho, S. M., Scornet, D., Rousvoal, S., Peters, N., Dartevelle, L., Peters, A. F. and**  
654 **Cock, J. M.** (2012a). *Ectocarpus*: A model organism for the brown algae. *Cold*  
655 *Spring Harb. Protoc* **2012**, 193–198.
- 656 **Coelho, S. M., Scornet, D., Rousvoal, S., Peters, N. T., Dartevelle, L., Peters, A. F. and**  
657 **Cock, J. M.** (2012b). How to cultivate *Ectocarpus*. *Cold Spring Harb Protoc* **2012**,  
658 258–261.
- 659 **Couceiro, L., Le Gac, M., Hunsperger, H. M., Mauger, S., Destombe, C., Cock, J. M.,**  
660 **Ahmed, S., Coelho, S. M., Valero, M. and Peters, A. F.** (2015). Evolution and  
661 maintenance of haploid-diploid life cycles in natural populations: The case of the  
662 marine brown alga *Ectocarpus*. *Evol. Int. J. Org. Evol.* **69**, 1808–1822.
- 663 **de Hoon, M. J. L., Imoto, S., Nolan, J. and Miyano, S.** (2004). Open source clustering  
664 software. *Bioinforma. Oxf. Engl.* **20**, 1453–1454.
- 665 **Delaroque, N., Müller, D., Bothe, G., Pohl, T., Knippers, R. and Boland, W.** (2001).  
666 The complete DNA sequence of the *Ectocarpus siliculosus* Virus EsV-1 genome.  
667 *Virology* **287**, 112–32.
- 668 **Dolan, L.** (2009). Body building on land: morphological evolution of land plants. *Curr*  
669 *Opin Plant Biol* **12**, 4–8.
- 670 **Farnham, G., Strittmatter, M., Coelho, S. M., Cock, J. M. and Brownlee, C.** (2013). Gene  
671 silencing in *Fucus* embryos: developmental consequences of RNAi-mediated  
672 cytoskeletal disruption. *J Phycol* **49**, 819–829.
- 673 **Fritch, F. E.** (1959). *The Structure and Reproduction of the Algae*. Cambridge.
- 674 **Gobler, C. J., Berry, D. L., Dyhrman, S. T., Wilhelm, S. W., Salamov, A., Lobanov, A. V.,**  
675 **Zhang, Y., Collier, J. L., Wurch, L. L., Kustka, A. B., et al.** (2011). Niche of  
676 harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. *Proc*  
677 *Natl Acad Sci U A* **108**, 4352–7.

- 678 **Gouy, M., Guindon, S. and Gascuel, O.** (2010). SeaView version 4: A multiplatform  
679 graphical user interface for sequence alignment and phylogenetic tree building.  
680 *Mol Biol Evol* **27**, 221–4.
- 681 **Haig, D. and Wilczek, A.** (2006). Sexual conflict and the alternation of haploid and  
682 diploid generations. *Philos Trans R Soc Lond B Biol Sci* **361**, 335–43.
- 683 **Heesch, S., Cho, G. Y., Peters, A. F., Le Corguillé, G., Falentin, C., Boutet, G., Coëdel, S.,  
684 Jubin, C., Samson, G., Corre, E., et al.** (2010). A sequence-tagged genetic map for  
685 the brown alga *Ectocarpus siliculosus* provides large-scale assembly of the  
686 genome sequence. *New Phytol* **188**, 42–51.
- 687 **Horn, T. and Boutros, M.** (2010). E-RNAi: a web application for the multi-species  
688 design of RNAi reagents--2010 update. *Nucleic Acids Res* **38**, W332-9.
- 689 **Kawai, H., Maeba, S., Sasaki, H., Okuda, K. and Henry, E. C.** (2003). *Schizocladia*  
690 *ischiensis*: a new filamentous marine chromophyte belonging to a new class,  
691 Schizocladiphyceae. *Protist* **154**, 211–28.
- 692 **Kawai, H., Hanyuda, T., Draisma, S. G. A., Wilce, R. T. and Andersen, R. A.** (2015).  
693 Molecular phylogeny of two unusual brown algae, *Phaeostrophion irregulare* and  
694 *Platysiphon glacialis*, proposal of the Stschapoviales ord. nov. and  
695 Platysiphonaceae fam. nov., and a re-examination of divergence times for brown  
696 algal orders. *J. Phycol.* **51**, 918–928.
- 697 **Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S. L.** (2013).  
698 TopHat2: accurate alignment of transcriptomes in the presence of insertions,  
699 deletions and gene fusions. *Genome Biol* **14**, R36.
- 700 **Kubota, A., Kita, S., Ishizaki, K., Nishihama, R., Yamato, K. T. and Kohchi, T.** (2014).  
701 Co-option of a photoperiodic growth-phase transition system during land plant  
702 evolution. *Nat Commun* **5**, 3668.
- 703 **Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., Park, Y. M., Buso,  
704 N. and Lopez, R.** (2015). The EMBL-EBI bioinformatics web and programmatic  
705 tools framework. *Nucleic Acids Res.* **43**, W580-584.
- 706 **Lipinska, A., Cormier, A., Luthringer, R., Peters, A. F., Corre, E., Gachon, C. M. M.,  
707 Cock, J. M. and Coelho, S. M.** (2015). Sexual dimorphism and the evolution of  
708 sex-biased gene expression in the brown alga *Ectocarpus*. *Mol. Biol. Evol.* **32**,  
709 1581–1597.
- 710 **Love, M. I., Huber, W. and Anders, S.** (2014). Moderated estimation of fold change and  
711 dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550.
- 712 **Menand, B., Yi, K., Jouannic, S., Hoffmann, L., Ryan, E., Linstead, P., Schaefer, D. and  
713 Dolan, L.** (2007). An ancient mechanism controls the development of cells with a  
714 rooting function in land plants. *Science* **316**, 1477–80.
- 715 **Moreau, H., Verhelst, B., Couloux, A., Derelle, E., Rombauts, S., Grimsley, N., Van Bel,  
716 M., Poulain, J., Katinka, M., Hohmann-Marriott, M. F., et al.** (2012). Gene

- 717 functionalities and genome structure in *Bathycoccus prasinus* reflect cellular  
718 specializations at the base of the green lineage. *Genome Biol* **13**, R74.
- 719 **Niklas, K. J. and Kutschera, U.** (2010). The evolution of the land plant life cycle. *New*  
720 *Phytol* **185**, 27–41.
- 721 **Nishiyama, T., Fujita, T., Shin, I. T., Seki, M., Nishide, H., Uchiyama, I., Kamiya, A.,**  
722 **Carninci, P., Hayashizaki, Y., Shinozaki, K., et al.** (2003). Comparative  
723 genomics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis*  
724 *thaliana*: implication for land plant evolution. *Proc Natl Acad Sci U A* **100**, 8007–  
725 12.
- 726 **Peters, A. F., Scornet, D., Ratin, M., Charrier, B., Monnier, A., Merrien, Y., Corre, E.,**  
727 **Coelho, S. M. and Cock, J. M.** (2008). Life-cycle-generation-specific  
728 developmental processes are modified in the *immediate upright* mutant of the  
729 brown alga *Ectocarpus siliculosus*. *Development* **135**, 1503–12.
- 730 **Pires, N. D. and Dolan, L.** (2012). Morphological evolution in land plants: new designs  
731 with old genes. *Philos Trans R Soc Lond B Biol Sci* **367**, 508–18.
- 732 **Qiu, Y.-L.** (2008). Phylogeny and evolution of charophytic algae and land plants. *J Syst*  
733 *Evol* **46**, 287–306.
- 734 **Qiu, Y.-L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrowska, O.,**  
735 **Lee, J., Kent, L., Rest, J., et al.** (2006). The deepest divergences in land plants  
736 inferred from phylogenomic evidence. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 15511–  
737 15516.
- 738 **R Development Core Team** (2009). *R: A Language and Environment for Statistical*  
739 *Computing*.
- 740 **Read, B. A., Kegel, J., Klute, M. J., Kuo, A., Lefebvre, S. C., Maumus, F., Mayer, C.,**  
741 **Miller, J., Monier, A., Salamov, A., et al.** (2013). Pan genome of the  
742 phytoplankton *Emiliana* underpins its global distribution. *Nature* **499**, 209–13.
- 743 **Saldanha, A. J.** (2004). Java Treeview--extensible visualization of microarray data.  
744 *Bioinforma. Oxf. Engl.* **20**, 3246–3248.
- 745 **Sano, R., Juarez, C. M., Hass, B., Sakakibara, K., Ito, M., Banks, J. A. and Hasebe, M.**  
746 (2005). KNOX homeobox genes potentially have similar function in both diploid  
747 unicellular and multicellular meristems, but not in haploid meristems. *Evol Dev* **7**,  
748 69–78.
- 749 **Shaw, A. J., Szovenyi, P. and Shaw, B.** (2011). Bryophyte diversity and evolution:  
750 windows into the early evolution of land plants. *Am J Bot* **98**, 352–69.
- 751 **Silberfeld, T., Leigh, J. W., Verbruggen, H., Cruaud, C., de Reviers, B. and Rousseau,**  
752 **F.** (2010). A multi-locus time-calibrated phylogeny of the brown algae  
753 (Heterokonta, Ochrophyta, Phaeophyceae): Investigating the evolutionary nature  
754 of the “brown algal crown radiation”. *Mol Phylogenet Evol* **56**, 659–74.

- 755 **Starr, R. C. and Zeikus, J. A.** (1993). UTEX-The culture collection of algae at the  
756 University of Texas at Austin. *J Phycol* **29 (Suppl.)**, 1–106.
- 757 **Sterck, L., Billiau, K., Abeel, T., Rouzé, P. and Van de Peer, Y.** (2012). ORCAE: online  
758 resource for community annotation of eukaryotes. *Nat Methods* **9**, 1041.
- 759 **Szovenyi, P., Rensing, S. A., Lang, D., Wray, G. A. and Shaw, A. J.** (2011). Generation-  
760 biased gene expression in a bryophyte model system. *Mol Biol Evol* **28**, 803–12.
- 761 **Yang, Z.** (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*  
762 **24**, 1586–91.
- 763 **Ye, N., Zhang, X., Miao, M., Fan, X., Zheng, Y., Xu, D., Wang, J., Zhou, L., Wang, D., Gao,**  
764 **Y., et al.** (2015). *Saccharina* genomes provide novel insight into kelp biology. *Nat*  
765 *Commun* **6**, 6986.
- 766
- 767



768 **Figure legends**

769

770 **Fig. 1. Positional cloning of the *IMM* gene and features of the encoded protein.**

771 (A) Schematic representation of the 43,708 bp interval on chromosome 27 between the closest  
772 recombining markers to the *IMM* locus. Protein coding exons, a tRNA and a snoRNA locus are  
773 shown as boxes. Genes above the line are transcribed to the right, genes below the line to the  
774 left. The position of the single mutation within the mapped interval (within exon five of the  
775 gene Ec-27\_002610) is indicated. (B) Schematic representation of the wild type *IMM* protein  
776 and the predicted product of the *imm* mutant allele including an alignment of the five EsV-1-7  
777 repeats at the C-terminal end of the *IMM* protein. Conserved residues are shown in bold and  
778 four highly conserved cysteine and histidine residues are indicated with an asterisk.

779

780 **Fig. 2. Introduction of siRNAs that target the *IMM* gene induces a phenotype that mimics**  
781 **the *imm* mutation.**

782 (A) Schematic representation of the early development of the wild type gametophyte (wt GA),  
783 the wild type sporophyte (wt SP) and the *imm* mutant sporophyte (*imm* SP). Grey, initial cell;  
784 blue, basal tissues consisting of narrow rhizoid cells or round filament cells; red, apical  
785 (upright) filaments consisting of cylindrical cells. Note the immediate asymmetrical division of  
786 the wild type gametophyte and the *imm* sporophyte germlings and the production of a  
787 characteristic straight rhizoid by the *imm* mutant sporophyte. In the wild type sporophyte apical  
788 organs develop only once an extensive basal system has been deployed (indicated by the double  
789 diagonal lines). Asterisks indicate the stages shown in panels B-E. (B) Three-day-old wild type  
790 partheno-sporophyte germling after introduction of siRNAs targeting the *IMM* transcript. The  
791 straight rhizoid and the immediate emergence of an upright filament are phenotypes typical of  
792 the *imm* mutant. (C) Three-day-old wild type partheno-sporophyte germling after introduction  
793 of control *GFP* siRNA. Note the absence of a rhizoid in the wild type. (D) Three-day-old *imm*  
794 partheno-sporophyte showing the straight rhizoid typical of this mutant. (E) Six-day-old wild  
795 type gametophyte showing the wavy rhizoid typical produced by this generation. Arrow heads  
796 indicate rhizoids. Size bar = 20  $\mu$ M.

797

798 **Fig. 3. Boxplot showing a quantitative reverse transcriptase polymerase chain reaction**  
799 **analysis of *IMM* transcript abundance at different stages of the *Ectocarpus* sp. life cycle.**

800 The dark bar indicates the median. SP, wild type diploid sporophyte; pSP, wild type partheno-  
801 sporophyte; *imm* pSP, partheno-sporophyte of the *imm* mutant; GA wild type gametophyte; a  
802 and b indicate statistically different samples (t test, P<0.05).

803

804 **Fig. 4. Comparison of the *imm* partheno-sporophyte transcriptome with those of different**  
805 **tissues and life cycle stages of wild type *Ectocarpus* sp..**

806 Principal component analysis was used to compare gene transcription patterns across samples  
807 (see Methods for details). The three dimensions PC1, PC2 and PC3 represent 51.9%, 15.4%  
808 and 12.2% of the variance respectively. *imm* pSP, *imm* partheno-sporophyte; pSP, wild type  
809 partheno-sporophyte; up\_pSP, upright filaments of wild type partheno-sporophyte; ba\_pSP,  
810 basal system of wild type partheno-sporophyte; iGA\_Female, immature wild type female  
811 gametophyte; iGA\_Male, immature wild type male gametophyte; mGA\_Female, mature wild  
812 type female gametophyte; mGA\_Male, mature wild type male gametophyte.

813

814 **Fig. 5. Distribution of EsV-1-7 domain genes across the eukaryotic supergroups.**

815 (A) Schematic tree of the eukaryotes showing the phylogenetic position of species that possess  
816 EsV-1-7 domain genes. (B) Amino acid sequence logos or individual sequences of EsV-1-7  
817 repeats from diverse eukaryotic species. For *Ectocarpus* sp. and *Guillardia theta*, divergent  
818 repeats were eliminated to generate more representative logos. Note that the *Rhizophagus*  
819 *irregularis* homologue was the only gene on a short sequence scaffold (2,177 bp), so its  
820 assignment to this species requires further verification.

821

822

823 **Supplementary Information**

824

825 **Fig. S1. Heat map representing abundances of transcripts of the *immediate upright up-***  
826 **regulated (IUP) and *immediate upright down-regulated (IDW) genes identified by Peters***  
827 **et al. (2008) at different stages of the life cycle and in different tissues.**

828 Expression levels are based on log<sub>2</sub> transformed TPMs. Gametophyte samples are on the left  
829 of the *imm* mutant partheno-sporophyte sample and sporophyte samples on the right. Note the  
830 similarity between transcript abundances in gametophyte samples and in sporophyte upright  
831 filaments. wt, wild type; *imm*, *imm* mutant; GA, gametophyte; pSP, partheno-sporophyte.  
832 Samples are whole organism unless otherwise stated. Gene identifiers (LocusIDs) have the  
833 format Ec-00\_000000.

834

835 **Fig. S2. Heat map representing transcript abundances at different stages of the life cycle**  
836 **and in different tissues for 1,087 genes that were significantly differentially expressed**  
837 **between the *imm* and wild type partheno-sporophytes.**

838 Expression levels are based on log<sub>2</sub> transformed TPMs. Gametophyte samples are on the left  
839 and sporophyte samples on the right. Note that genes that are upregulated in the gametophyte  
840 generation also tend to be upregulated in sporophyte upright filaments. Gene names were  
841 omitted for clarity. wt, wild type; *imm*, *imm* mutant; GA, gametophyte; pSP, partheno-  
842 sporophyte. Samples are whole organism unless otherwise stated.

843

844 **Fig. S3. Heat map representing abundances of transcripts of the members of the**  
845 ***Ectocarpus* sp. EsV-1-7 domain gene family at different stages of the life cycle and in**  
846 **different tissues.**

847 Expression levels are based on log<sub>2</sub> transformed TPMs. wt, wild type; *imm*, *imm* mutant; GA,  
848 gametophyte; pSP, partheno-sporophyte. Samples are whole organism unless otherwise stated.  
849 Gene identifiers (LocusIDs) have the format Ec-00\_000000.

850

851 **Fig. S4. Unrooted maximum likelihood tree based on an alignment of IMM and Ec-**  
852 **17\_002150 orthologues from diverse brown algal species.**

853 The tree was generated from a Muscle alignment with the MEGA7 maximum likelihood  
854 algorithm using the JTT+G model and 1000 bootstrap replicates. Bootstrap support values are  
855 indicated at each node. Ec-17\_002150 is the most similar gene to *IMM* in the *Ectocarpus* sp.  
856 genome. IMM, *IMM* orthologue; 17\_2150, Ec-17\_002150 orthologue; Esp, *Ectocarpus* sp.;

857 Efas, *Ectocarpus fasciculatus*; Slo, *Scytosiphon lomentaria*; Pfa, *Petalonia fascia*; Pla,  
858 *Punctaria latifolia*; Sja, *Saccharina japonica*; Kcr, *Kjellmaniella crassifolia*; Sho, *Sargassum*  
859 *horneri*; Sth, *Sargassum thunbergi*; She, *Sargassum hemiphyllum*; Smu, *Sargassum muticum*;  
860 Csi, *Colpomenia sinuosa*; Iok, *Ishige okamurai*.

861

862 **Fig. S5. Comparison of eight representative orthologous pairs of *Ectocarpus* sp. and *S.***  
863 ***japonica* EsV-1-7 family proteins showing the weak conservation of regions outside the**  
864 **EsV-1-7 repeat domains.**

865 *Ectocarpus* sp. (Ec) and *S. japonica* (SJ) predicted proteins are represented by blue cylinders,  
866 with the EsV-1-7 repeat regions indicated by a thicker, green cylinder. Grey lines delineate  
867 equivalent domains in pairs of orthologues. Global percent amino acid identities between  
868 domains are indicated.

869

870 **Table S1. List of the genetic markers used to map the *IMM* locus.**

871

872 **Table S2. List of the genes that were significantly differentially expressed between the**  
873 ***imm* and wild type sporophyte generations.**

874

875 **Table S3. Results of a GOslim (p<0.05) search for enriched gene ontology categories in**  
876 **the list of genes that were significantly upregulated in the *imm* sporophyte compared with**  
877 **the wild type sporophyte (TPM>1, fold change>2).**

878

879 **Table S4. Results of a GOslim (p<0.05) search for enriched gene ontology categories in**  
880 **the list of genes that were significantly downregulated in the *imm* sporophyte compared**  
881 **with the wild type sporophyte (TPM>1, fold change>2).**

882

883 **Table S5. Characteristics of the 91 members of the *Ectocarpus* sp. EsV-1-7 domain gene**  
884 **family.**

885

886 **Table S6. EsV-1-7 domain genes in other species.**

887

888 **Table S7. siRNA molecules used for the RNA interference experiments.**

889

890 **Table S8. *Ectocarpus* RNA-seq data used in this study.**