First posted online on 3 January 2017 as 10.1242/dev.141523
Access the most recent version at http://dev.biologists.org/lookup/doi/10.1242/dev.141523

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

Nicolas Macaisne[1], Fuli Liu[1,†], Delphine Scornet[1], Akira F. Peters[2], Agnieszka Lipinska[1], Marie-Mathilde Perrineau[1], Antoine Henry[1], Martina Strittmatter[1], Susana M. Coelho[1], J. Mark Cock[1,*]

[1]CNRS, Sorbonne Université, UPMC University Paris 06, Algal Genetics Group, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688, Roscoff, France, [2]Bezhin Rosko, 29250, Santec, France

[*]Author for correspondance (cock@sb-roscoff.fr)

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

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

ABSTRACT

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

INTRODUCTION

Multicellular organisms with haploid-diploid life cycles are found in several major eukaryotic groups including the green lineage (Archaeplastida) and the red and brown macroalgae (Rhodophyta and Phaeophyceae, respectively). In these organisms, a single genome provides the genetic information to deploy two different developmental programs during the course of the life cycle, leading to the construction of the sporophyte and gametophyte generations, respectively (Cock et al., 2013; Coelho et al., 2007). One consequence of this type of life cycle is that the emergence of developmental innovations for one generation of the life cycle can occur without it being necessary to evolve novel developmental regulatory modules *de novo*. This is because it is also possible to adapt regulatory modules that have evolved to function during one generation of the life cycle to carry out related functions during the other generation. An important objective of the developmental biologists that study these organisms has been to understand the relative contributions of these two processes - developmental innovation and trans-generation co-option - to the evolution of multicellularity in these species (Dolan, 2009; Pires and Dolan, 2012; Shaw et al., 2011). In the green lineage, embryophytes (which have haploid-diploid life cycles) are thought to have evolved from a green algal ancestor with a haploid life cycle by the addition of a sporophyte generation (Bower, 1890; Celakovsky, 1874; Dolan, 2009; Haig and Wilczek, 2006; Niklas and Kutschera, 2010; Qiu, 2008; Qiu et al., 2006). It has been proposed that the regulatory networks that controlled the development of early embryophyte sporophytes were recruited to a large extent from the gametophyte generation (Dolan, 2009; Niklas and Kutschera, 2010). Support for this viewpoint has come both from broad comparisons of gametophyte and sporophyte transcriptomes (Nishiyama et al., 2003; Szovenyi et al., 2011) and from demonstrations that homologues of key regulatory genes in embryophyte sporophytes play important roles in gametophyte function in bryophytes (Aoyama et al., 2012; Kubota et al., 2014; Menand et al., 2007; Nishiyama et al., 2003). There are however exceptions (Szovenyi et al., 2011). For example, the KNOX family of TALE homeodomain transcription factors are not expressed during the gametophyte generation in bryophytes and therefore appear to have evolved as sporophyte developmental regulators (Sano et al., 2005).

To more fully understand the relative contributions of developmental innovation and trans-generation co-option to the evolution of multicellularity in organisms with haploid-diploid life cycles it would be of interest to investigate this phenomenon in several lineages that have independently evolved complex multicellularity. Not only would this allow the generality of inferences from studies of the green lineage to be accessed but it would also

allow an evaluation of the importance of the ancestral state on subsequent evolutionary events. For example, the brown algae (Phaeophyceae) most probably evolved from an ancestor that alternated between simple, filamentous sporophyte and gametophyte generations (Kawai et al., 2003; Silberfeld et al., 2010). If this was the case, then the evolution of novel regulatory systems may have played a more important role in the emergence of novel developmental mechanisms than co-option of regulators across generations in this phylogenetic group. Unfortunately, very little is currently known about developmental processes in the brown algae and, for example, no developmental regulatory genes have so far been characterised at the molecular level. However, the recent emergence of the filamentous alga *Ectocarpus* sp. as a model organism for this group (Cock et al., 2013; Cock et al., 2015; Coelho et al., 2012a) has created a context in which this type of question can be addressed.

Ectocarpus sp. has a haploid-diploid life cycle that involves alternation between two generations, which both consist of uniserate filaments with a small number of different cell types and bearing simple reproductive structures (Cock et al., 2015). The morphological similarity of the two generations has allowed mutants affected both in switching between generations (Coelho et al., 2011) and in generation-related developmental processes (Peters et al., 2008) to be isolated. The *immediate upright* (*imm*) mutant is particularly interesting because it has major effects on the early development of the sporophyte generation but causes no visible phenotype during the gametophyte generation (Peters et al., 2008). In individuals that carry this mutation, the initial cell of the sporophyte generation undergoes an asymmetrical rather than a symmetrical cell division and produces an upright filament and a rhizoid rather than the prostrate filament typical of wild type sporophytes (Peters et al., 2008). Individuals that carry this mutation therefore fail to implement the typical early sporophyte developmental program and resemble gametophytes but produce the sexual structures of the sporophyte generation at maturity. The absence of a phenotype during the gametophyte generation suggests that the developmental program directed by the *IMM* gene may have been a sporophyte-specific innovation.

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

RESULTS

Positional cloning of the *IMM* locus

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

To identify the *imm* mutation within the 43.7 kbp interval, this region was amplified as a series of PCR products and the pooled products sequenced on an Illumina HiSeq platform. As the Imm^{-} phenotype is the result of a spontaneous mutation that was not originally present in the parent strain Ec17, we reconstructed reference sequences for the two parental haplotypes of this region by sequencing equivalent PCR products amplified from eight wild type haploid siblings of the *imm* strain Ec137. Comparison of the sequence data for the eight siblings with that from the Ec137 strain allowed polymorphisms inherited from the diploid parent sporophyte to be distinguished from the causal mutation. Sanger resequencing was used to validate polymorphisms detected by the Illumina sequencing approach and to generate sequence data for several short regions that were not covered by the Illumina sequence data. This approach identified a two base pair deletion within exon five of gene Ec-27_002610 as the causal mutation of the Imm^{-} phenotype (Fig 1A). No other mutations were detected within the mapped interval.

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

The *IMM* gene (locus Ec-27_002610) is predicted to encode an 862 amino acid (91.8 kDa) protein, which consists of a long N-terminal domain that shares no similarity with other *Ectocarpus* sp. genes or genes from other species, plus a C-terminal domain which includes

five imperfect tandem repeats of a 38 amino acid cysteine-rich motif (C-X4-C-X16-C-X2-H-X12, Fig 1B). The 38 amino acid cysteine-rich motif is very similar to a cysteine-rich repeated motif found in the EsV-1-7 protein of the *Ectocarpus* virus EsV-1 (Delaroque et al., 2001), which also contains five of these repeated motifs. Based on this similarity, hereafter we will refer to the 38 amino acid cysteine-rich motif as an EsV-1-7 repeat.

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

Disruption of *IMM* function using RNA interference

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

Wild type gametes that fail to fuse with a gamete of the opposite sex can develop parthenogenetically to give rise to partheno-sporophytes. These partheno-sporophytes go through the same developmental steps as diploid sporophytes derived from zygotes and are morphologically indistinguishable from the latter. In both cases the initial cell undergoes a symmetrical division that gives rise to two germ tubes, which grow to form a symmetrical, prostrate basal filament (Peters et al., 2008). The basal filament, which is composed of characteristic round and elongated cells, adheres strongly to the substratum. Following simultaneous introduction of three siRNA molecules targeting the *IMM* gene transcript, a small proportion (0.63% ± 0.09% mean ± standard deviation in six replicates each of 400 individuals) of the parthenogenetic gametes adopted a pattern of early development that closely resembled the phenotype of the *imm* mutant (Peters et al., 2008). These gametes underwent an asymmetrical rather than a symmetrical initial cell division and the two germ tubes of the developing partheno-sporophyte gave rise to an upright filament and a rhizoid (Fig 2). Individuals with this phenotype were not observed in parallel samples of gametes treated with an siRNA directed against a green fluorescent protein gene sequence as a control

(six replicates of 400 individuals) and the difference between the test and control experiments was was highly significant (Pearson's χ^2 = 13.0, p = 0.0003). These observations indicated that RNA-interference-induced knockdown of *IMM* gene expression had the same developmental consequences as the *imm* mutation, at least in a small proportion of the treated individuals. Taken together with the results of the mapping of the genetic mutation, this observation confirmed that the two base pair deletion identified in exon five of gene Ec-27_002610 was the causal mutation of the Imm$^-$ phenotype.

Expression pattern of the *IMM* gene

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

Analysis of gene expression in the *imm* mutant sporophyte

In a previous study we analysed gene expression in the *imm* mutant using a microarray constructed with sequences from two subtraction libraries enriched in genes differentially expressed during either the sporophyte or the gametophyte generation (Peters et al., 2008). This analysis indicated that sporophyte-specific genes were down regulated and gametophyte-specific genes upregulated in the *imm* mutant during the sporophyte generation. Based on this information, and the morphological resemblance of the *imm* sporophyte to the wild type gametophyte, the Imm$^-$ phenotype was interpreted as representing partial homeotic switching from the sporophyte to the gametophyte developmental program (Peters et al., 2008). Here we used multiple RNA-seq datasets to compare the *imm* transcriptome with a broader range of samples, including two microdissected partheno-sporophyte tissue samples corresponding to the apical upright filaments and the basal system, respectively. Principal component analysis indicated that, overall, the transcriptome of the *imm* partheno-sporophyte was actually more similar to the transcriptomes of wild type partheno-sporophyte samples, particularly samples that included upright filaments, than to wild type gametophyte samples (Fig 4).

We therefore reanalysed the expression patterns of sets of genes that had been previously identified as being significantly upregulated or down regulated in the *imm* mutant partheno-

sporophyte compared to the wild type partheno-sporophyte (denoted as *imm* upregulated or IUP and *imm* downregulated or IDW genes by Peters et al., 2008). This analysis indicated that IUP and IDW genes tended to be up and down regulated, respectively, in gametophyte samples but they also showed very similar patterns of expression in upright filaments isolated from the sporophyte generation (Fig S1). Bearing in mind that that the gametophyte generation consists almost entirely of upright filaments, these results suggested that the IUP and IDW genes, rather than being life-cycle-regulated genes, may correspond to loci that are differentially regulated in upright filaments compared to basal tissues.

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

Taken together, these observations suggest an alternative interpretation of the Imm$^-$ phenotype. Rather than representing a mutation that causes switching between life cycle generations, we propose that abrogation of the *IMM* gene leads to failure to correctly implement the early developmental program of the sporophyte. In the absence of a functional *IMM* gene the initial cell does not divide symmetrically and there is no deployment of a system of basal filaments before the establishment of the apical/basal axis (Fig 2A). Rather, an asymmetrical division of the initial cell directly produces a basal rhizoid cell and an apical thallus cell. We suggest that the resemblance with the gametophyte, in terms of gene expression, is not due to switching to the gametophyte developmental program but rather due to the sporophyte adopting an alternative developmental program which is more similar to that of the gametophyte (immediate production of an upright filament).

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

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

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

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

Estimates of transcript abundances based on RNA-seq data indicated that the members of the *Ectocarpus* sp. EsV-1-7 domain gene family have diverse expression patterns (Fig S3), suggesting that they carry out diverse functions at different stages of the life cycle and in different organs.

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

Searches of the recently published *Saccharina japonica* genome (Ye et al., 2015) and multiple brown algal transcriptomes produced by the onekp project (https://sites.google.com/a/ualberta.ca/onekp/) identified predicted proteins similar to IMM in a broad range of brown algal species, including members of the Ectocarpales, the Laminariales and the Fucales. Reciprocal blasts indicated that a number of these proteins were IMM orthologues and this conclusion was supported by the observation that best reciprocal

blast matches to Ec-17_002150 (which is the member of the *Ectocarpus* sp. EsV-1-7 family that is most similar to IMM) formed a distinct cluster in a phylogenetic tree (Fig S4).

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

The large size of the EsV-1-7 domain family in *Ectocarpus* sp. and the presence of several pseudogenes and putative gene fragments suggested that this gene family may be undergoing a process of gene gain and gene loss over time. We therefore searched for orthologues of the *Ectocarpus* sp. EsV-1-7 domain genes in the *S. japonica* genome (Ye et al., 2015) to investigate the evolutionary dynamics of the family. The last common ancestor of *Ectocarpus* sp. (order Ectocarpales) and *S. japonica* (order Laminariales) existed between 80 and 110 mya (Kawai et al., 2015; Silberfeld et al., 2010). An extensive search, which included *de novo* annotation of 19 new EsV-1-7 domain genes in *S. japonica,* identified orthologues for only 34 of the 91 *Ectocarpus* sp. genes (Table S6), and no orthologues were found for the five *Ectocarpus* sp. pseudogenes. Moreover, in general when a *S. japonica* orthologue was identified, it was considerably diverged from its *Ectocarpus* sp. counterpart. These results suggested that the EsV-1-7 gene families have diverged relatively rapidly since the divergence of the two species. To evaluate whether the EsV-1-7 genes are evolving more rapidly than other genes in the genome, we compared the set of percentage identities between orthologous *Ectocarpus* sp. and *S. japonica* EsV-1-7 proteins with the percentage identities for a set of 9,845 orthologous protein pairs obtained by reciprocal Blastp comparison of the predicted proteomes of the two species. Median identity for the orthologous pairs of EsV-1-7 proteins (31.1% identity) was significantly lower (Wilcoxon test $p = 9.2\text{x}10^{-7}$) than the genome-wide average (47.9%). Sex-biased genes have been shown to evolve rapidly in several, diverse species, including brown algae (Lipinska et al., 2015). The median identity for the EsV-1-7 proteins was also significantly lower (Wilcoxon test $p = 7.6\text{x}10^{-5}$) than that for 905 orthologous pairs of sex-biased proteins from *Ectocarpus* sp. and *S. japonica* (44.6%). Taken together, these analyses indicated that the EsV-1-7 family is evolving not only more rapidly than the average gene in the genome but also more rapidly than the set of sex-biased genes.

A more detailed analysis of sequence conservation at the domain level, indicated that the EsV-1-7 domains (i.e. the regions of the proteins containing the EsV-1-7 repeats) tended to be

markedly more conserved than the other parts of the proteins (Fig S5). The conservation of the EsV-1-7 domains at the sequence level suggests that these regions correspond to the main functional features of the proteins and it is possible that the regions outside these domains do not have specific functional roles in most cases. Note that the IMM orthologues from diverse brown algal species also exhibited this characteristic, with the EsV-1-7 repeat region being more strongly conserved across species than the other parts of the protein (e.g. Fig S5).

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

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

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

DISCUSSION

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

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

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

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

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

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

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

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

MATERIALS AND METHODS

Ectocarpus strains and growth conditions

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

Genetic mapping of the *IMM* locus

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

Quantitative RT-PCR analysis of mRNA abundance

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

Ec17) and wild type gametophytes (strain Ec32). These samples were not the same as those used for the RNA-seq analysis. DNase treatment using the TURBO DNA-free kit (Applied Biosystems) was carried out to eliminate any contaminating genomic DNA. RNA concentration and quality was then determined by spectrophotometry and agarose gel electrophoresis. Between 0.2 and 2.0 µg of total RNA was reverse-transcribed to cDNA with the SuperScript First-Strand Synthesis System (Invitrogen).

IMM cDNA was amplified with primers E657Q4F (GGGGTTTGGGTGGAAGAGGACC) and E657Q4R (CGGCGTGGAAGCTGCCTGGTAT) and *ELONGATION FACTOR 1α* (*EF1α*) cDNA was amplified as an internal reference with primers EF1adeL (CAAGTCCGTCGAGAAGAAGG) and EF1autrL (CCAGCAACACCACAATGTCT). Quantitative RT-PCR was carried out using the ABsolute QPCR SYBR Green ROX Mix (ThermoScientific) in a Chromo4 thermocycler (Bio-Rad Laboratories) and data were analysed with the Opticon monitor 3 software (Bio-Rad Laboratories). Amplification specificity was checked using a dissociation curve. The amplification efficiency was tested using a genomic dilution series and was always between 90% and 110%. A standard curve was established for each gene using a range of dilutions of *Ectocarpus* sp. genomic DNA (between 80 and 199,600 copies) and gene expression level was normalized against the *EF1α* reference gene. Two technical replicates were carried out for the standard curves and three technical replicates for the samples. The data shown correspond to the mean ± the standard deviation for the biological replicates.

RNA interference

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

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

Transcriptome analysis

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

Heat maps schematically representing gene transcript abundances were generated using the log2 transformed TPM values centred by gene in Cluster 3.0 (de Hoon et al., 2004) and Treeview 1.1.6 (Saldanha, 2004).

Principal component analysis (PCA) of transcript abundances was carried using log2 transformed TPM expression values in R (prcomp) and visualized with the rgl package for R

(R Development Core Team, 2009). Genes with TPMs <1 were considered not expressed and their log2 values were set to 10e-5 to remove noise from the data. Only genes with divergent expression patterns were used for the whole transcriptome PCA analysis (4,381 genes). These genes were chosen based on a calculation that estimated expression variance across all libraries (ΔTPM = max TPM – min TPM) and the third quartile (harbouring genes with ΔTPM>32) was chosen for the PCA calculation.

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

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

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

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

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

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

Searches for proteins related to IMM from additional species were carried out against the NCBI, Uniprot and onekp (https://sites.google.com/a/ualberta.ca/onekp/) databases using both Blastp and HMMsearch, the latter using both an alignment of *Ectocarpus* IMM homologues and an alignment of brown algal IMM orthologues. In addition, HMMsearch and tBlastn or Blastp searches were carried out against the following genomes and complete deduced proteomes: *Thalassiosira pseudonana* (diatom; Thaps3 assembled and unmapped scaffolds, http://genome.jgi-psf.org/Thaps3/Thaps3.download.ftp.html) (Armbrust et al., 2004), *Phaeodactylum tricornutum* (diatom; Phatr2 assembled and unmapped scaffolds, http://genome.jgi-psf.org/Phatr2/Phatr2.download.ftp.html) (Bowler et al., 2008), *Aureococcus anophagefferens* (Pelagophyceae; http://genome.jgi-psf.org/Auran1/Auran1.download.ftp.html) (Gobler et al., 2011), *Emiliana huxleyi* (haptophyte; ttp://genome.jgi.doe.gov/Emihu1/Emihu1.download.ftp.html) (Read et al., 2013), *Chlorella variablis* NC64A (chlorophyte; JGI http://genome.jgi.doe.gov/pages/dynamicOrganismDownload.jsf?organism=ChlNC64A_1) (Blanc et al., 2010), *Monoraphidium neglectum* SAG 48.87 (chlorophyte; NCBI RefSeq assembly GCF_000611645.1) (Bogen et al., 2013) and *Bathycoccus prasinos* (chlorophyte; http://bioinformatics.psb.ugent.be/orcae/overview/Bathy) (Moreau et al., 2012). Finally, iterative searches using EsV-1-7 domain proteins from phylogenetically distant species were carried out against the NCBI and Uniprot databases to identify distantly related members of this gene family.

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

Statistics

Sample sizes were chosen to allow adequate downstream statistical analysis.

Data Availability

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

Acknowledgements

We thank Toshiki Uji for providing RNA-seq data.

Competing interests

The authors declare no competing or financial interests.

Author Contributions

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

Funding

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

Supplementary information

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

References

Abeel, T., Van Parys, T., Saeys, Y., Galagan, J. and Van de Peer, Y. (2012). GenomeView: a next-generation genome browser. *Nucleic Acids Res* **40**, e12.

Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol* **11**, R106.

Anders, S., Pyl, S. T. and Huber, W. (2014). HTSeq — A Python framework to work with high-throughput sequencing data. *BioRxiv Prepr.*

Aoyama, T., Hiwatashi, Y., Shigyo, M., Kofuji, R., Kubo, M., Ito, M. and Hasebe, M. (2012). AP2-type transcription factors determine stem cell identity in the moss Physcomitrella patens. *Development* **139**, 3120–9.

Armbrust, E., Berges, J., Bowler, C., Green, B., Martinez, D., Putnam, N., Zhou, S., Allen, A., Apt, K., Bechner, M., et al. (2004). The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. *Science* **306**, 79–86.

Blanc, G., Duncan, G., Agarkova, I., Borodovsky, M., Gurnon, J., Kuo, A., Lindquist, E., Lucas, S., Pangilinan, J., Polle, J., et al. (2010). The Chlorella variabilis NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell* **22**, 2943–55.

Bogen, C., Al-Dilaimi, A., Albersmeier, A., Wichmann, J., Grundmann, M., Rupp, O., Lauersen, K. J., Blifernez-Klassen, O., Kalinowski, J., Goesmann, A., et al. (2013). Reconstruction of the lipid metabolism for the microalga Monoraphidium neglectum from its genome sequence reveals characteristics suitable for biofuel production. *BMC Genomics* **14**, 926.

Bower, F. O. (1890). On antithetic as distinct from homologous alternation of generations in plants. *Ann Bot* **4**, 347–370.

Bowler, C., Allen, A., Badger, J., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otillar, R., et al. (2008). The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239–44.

Caplen, N. J., Parrish, S., Imani, F., Fire, A. and Morgan, R. A. (2001). Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 9742–9747.

Celakovsky, L. (1874). Ueber die verschiedenen Formen und die Bedeutung des Generationwechsels der Pflanzen. *Sitzungsberichte Koeniglichen Boehmischen Ges. Wiss. Prague* **2**, 21–61.

Cock, J. M. and Coelho, S. M. (2011). Algal models in plant biology. *J. Exp. Bot.* **62**, 2425–2430.

Cock, J. M., Sterck, L., Rouzé, P., Scornet, D., Allen, A. E., Amoutzias, G., Anthouard, V., Artiguenave, F., Aury, J.-M., Badger, J. H., et al. (2010a). The Ectocarpus genome and the independent evolution of multicellularity in brown algae. *Nature* **465**, 617–621.

Cock, J. M., Sterck, L., Rouzé, P., Scornet, D., Allen, A. E., Amoutzias, G., Anthouard, V., Artiguenave, F., Aury, J., Badger, J., et al. (2010b). The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* **465**, 617–21.

Cock, J. M., Godfroy, O., Macaisne, N., Peters, A. F. and Coelho, S. M. (2013). Evolution and regulation of complex life cycles: a brown algal perspective. *Curr Opin Plant Biol* **17**, 1–6.

Cock, J. M., Godfroy, O., Strittmatter, M., Scornet, D., Uji, T., Farnham, G., Peters, A. F. and Coelho, S. M. (2015). Emergence of *Ectocarpus* as a model system to study the evolution of complex multicellularity in the brown algae. In *Evolutionary transitions to multicellular life* (ed. Ruiz-Trillo, I.) and Nedelcu, A. M.), pp. 153–162. Dordrecht: Springer.

Coelho, S., Peters, A., Charrier, B., Roze, D., Destombe, C., Valero, M. and Cock, J. (2007). Complex life cycles of multicellular eukaryotes: new approaches based on the use of model organisms. *Gene* **406**, 152–70.

Coelho, S. M., Godfroy, O., Arun, A., Le Corguillé, G., Peters, A. F. and Cock, J. M. (2011). *OUROBOROS* is a master regulator of the gametophyte to sporophyte life cycle transition in the brown alga *Ectocarpus*. *Proc Natl Acad Sci U A* **108**, 11518–11523.

Coelho, S. M., Scornet, D., Rousvoal, S., Peters, N., Dartevelle, L., Peters, A. F. and Cock, J. M. (2012a). *Ectocarpus*: A model organism for the brown algae. *Cold Spring Harb. Protoc* **2012**, 193–198.

Coelho, S. M., Scornet, D., Rousvoal, S., Peters, N. T., Dartevelle, L., Peters, A. F. and Cock, J. M. (2012b). How to cultivate *Ectocarpus*. *Cold Spring Harb Protoc* **2012**, 258–261.

Couceiro, L., Le Gac, M., Hunsperger, H. M., Mauger, S., Destombe, C., Cock, J. M., Ahmed, S., Coelho, S. M., Valero, M. and Peters, A. F. (2015). Evolution and maintenance of haploid-diploid life cycles in natural populations: The case of the marine brown alga *Ectocarpus*. *Evol. Int. J. Org. Evol.* **69**, 1808–1822.

de Hoon, M. J. L., Imoto, S., Nolan, J. and Miyano, S. (2004). Open source clustering software. *Bioinforma. Oxf. Engl.* **20**, 1453–1454.

Delaroque, N., Müller, D., Bothe, G., Pohl, T., Knippers, R. and Boland, W. (2001). The complete DNA sequence of the *Ectocarpus siliculosus* Virus EsV-1 genome. *Virology* **287**, 112–32.

Dolan, L. (2009). Body building on land: morphological evolution of land plants. *Curr Opin Plant Biol* **12**, 4–8.

Farnham, G., Strittmatter, M., Coelho, S. M., Cock, J. M. and Brownlee, C. (2013). Gene silencing in *Fucus* embryos: developmental consequences of RNAi-mediated cytoskeletal disruption. *J Phycol* **49**, 819–829.

Fritch, F. E. (1959). *The Structure and Reproduction of the Algae*. Cambridge.

Gobler, C. J., Berry, D. L., Dyhrman, S. T., Wilhelm, S. W., Salamov, A., Lobanov, A. V., Zhang, Y., Collier, J. L., Wurch, L. L., Kustka, A. B., et al. (2011). Niche of harmful alga Aureococcus anophagefferens revealed through ecogenomics. *Proc Natl Acad Sci U A* **108**, 4352–7.

Gouy, M., Guindon, S. and Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**, 221–4.

Haig, D. and Wilczek, A. (2006). Sexual conflict and the alternation of haploid and diploid generations. *Philos Trans R Soc Lond B Biol Sci* **361**, 335–43.

Heesch, S., Cho, G. Y., Peters, A. F., Le Corguillé, G., Falentin, C., Boutet, G., Coëdel, S., Jubin, C., Samson, G., Corre, E., et al. (2010). A sequence-tagged genetic map for the brown alga *Ectocarpus siliculosus* provides large-scale assembly of the genome sequence. *New Phytol* **188**, 42–51.

Horn, T. and Boutros, M. (2010). E-RNAi: a web application for the multi-species design of RNAi reagents--2010 update. *Nucleic Acids Res* **38**, W332-9.

Kawai, H., Maeba, S., Sasaki, H., Okuda, K. and Henry, E. C. (2003). *Schizocladia ischiensis*: a new filamentous marine chromophyte belonging to a new class, Schizocladiophyceae. *Protist* **154**, 211–28.

Kawai, H., Hanyuda, T., Draisma, S. G. A., Wilce, R. T. and Andersen, R. A. (2015). Molecular phylogeny of two unusual brown algae, *Phaeostrophion irregulare* and *Platysiphon glacialis*, proposal of the Stschapoviales ord. nov. and Platysiphonaceae fam. nov., and a re-examination of divergence times for brown algal orders. *J. Phycol.* **51**, 918–928.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36.

Kubota, A., Kita, S., Ishizaki, K., Nishihama, R., Yamato, K. T. and Kohchi, T. (2014). Co-option of a photoperiodic growth-phase transition system during land plant evolution. *Nat Commun* **5**, 3668.

Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., Park, Y. M., Buso, N. and Lopez, R. (2015). The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res.* **43**, W580-584.

Lipinska, A., Cormier, A., Luthringer, R., Peters, A. F., Corre, E., Gachon, C. M. M., Cock, J. M. and Coelho, S. M. (2015). Sexual dimorphism and the evolution of sex-biased gene expression in the brown alga *Ectocarpus*. *Mol. Biol. Evol.* **32**, 1581–1597.

Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550.

Menand, B., Yi, K., Jouannic, S., Hoffmann, L., Ryan, E., Linstead, P., Schaefer, D. and Dolan, L. (2007). An ancient mechanism controls the development of cells with a rooting function in land plants. *Science* **316**, 1477–80.

Moreau, H., Verhelst, B., Couloux, A., Derelle, E., Rombauts, S., Grimsley, N., Van Bel, M., Poulain, J., Katinka, M., Hohmann-Marriott, M. F., et al. (2012). Gene functionalities and genome structure in Bathycoccus prasinos reflect cellular specializations at the base of the green lineage. *Genome Biol* **13**, R74.

Niklas, K. J. and Kutschera, U. (2010). The evolution of the land plant life cycle. *New Phytol* **185**, 27–41.

Nishiyama, T., Fujita, T., Shin, I. T., Seki, M., Nishide, H., Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K., et al. (2003). Comparative genomics of Physcomitrella patens gametophytic transcriptome and Arabidopsis thaliana: implication for land plant evolution. *Proc Natl Acad Sci U A* **100**, 8007–12.

Peters, A. F., Scornet, D., Ratin, M., Charrier, B., Monnier, A., Merrien, Y., Corre, E., Coelho, S. M. and Cock, J. M. (2008). Life-cycle-generation-specific developmental processes are modified in the *immediate upright* mutant of the brown alga *Ectocarpus siliculosus*. *Development* **135**, 1503–12.

Pires, N. D. and Dolan, L. (2012). Morphological evolution in land plants: new designs with old genes. *Philos Trans R Soc Lond B Biol Sci* **367**, 508–18.

Qiu, Y.-L. (2008). Phylogeny and evolution of charophytic algae and land plants. *J Syst Evol* **46**, 287–306.

Qiu, Y.-L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrovska, O., Lee, J., Kent, L., Rest, J., et al. (2006). The deepest divergences in land plants inferred from phylogenomic evidence. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 15511–15516.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*.

Read, B. A., Kegel, J., Klute, M. J., Kuo, A., Lefebvre, S. C., Maumus, F., Mayer, C., Miller, J., Monier, A., Salamov, A., et al. (2013). Pan genome of the phytoplankton Emiliania underpins its global distribution. *Nature* **499**, 209–13.

Saldanha, A. J. (2004). Java Treeview--extensible visualization of microarray data. *Bioinforma. Oxf. Engl.* **20**, 3246–3248.

Sano, R., Juarez, C. M., Hass, B., Sakakibara, K., Ito, M., Banks, J. A. and Hasebe, M. (2005). KNOX homeobox genes potentially have similar function in both diploid unicellular and multicellular meristems, but not in haploid meristems. *Evol Dev* **7**, 69–78.

Shaw, A. J., Szovenyi, P. and Shaw, B. (2011). Bryophyte diversity and evolution: windows into the early evolution of land plants. *Am J Bot* **98**, 352–69.

Silberfeld, T., Leigh, J. W., Verbruggen, H., Cruaud, C., de Reviers, B. and Rousseau, F. (2010). A multi-locus time-calibrated phylogeny of the brown algae (Heterokonta, Ochrophyta, Phaeophyceae): Investigating the evolutionary nature of the "brown algal crown radiation". *Mol Phylogenet Evol* **56**, 659–74.

Starr, R. C. and Zeikus, J. A. (1993). UTEX-The culture collection of algae at the University of Texas at Austin. *J Phycol* **29 (Suppl.)**, 1–106.

Sterck, L., Billiau, K., Abeel, T., Rouzé, P. and Van de Peer, Y. (2012). ORCAE: online resource for community annotation of eukaryotes. *Nat Methods* **9**, 1041.

Szovenyi, P., Rensing, S. A., Lang, D., Wray, G. A. and Shaw, A. J. (2011). Generation-biased gene expression in a bryophyte model system. *Mol Biol Evol* **28**, 803–12.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586–91.

Ye, N., Zhang, X., Miao, M., Fan, X., Zheng, Y., Xu, D., Wang, J., Zhou, L., Wang, D., Gao, Y., et al. (2015). *Saccharina* genomes provide novel insight into kelp biology. *Nat Commun* **6**, 6986.

Figures

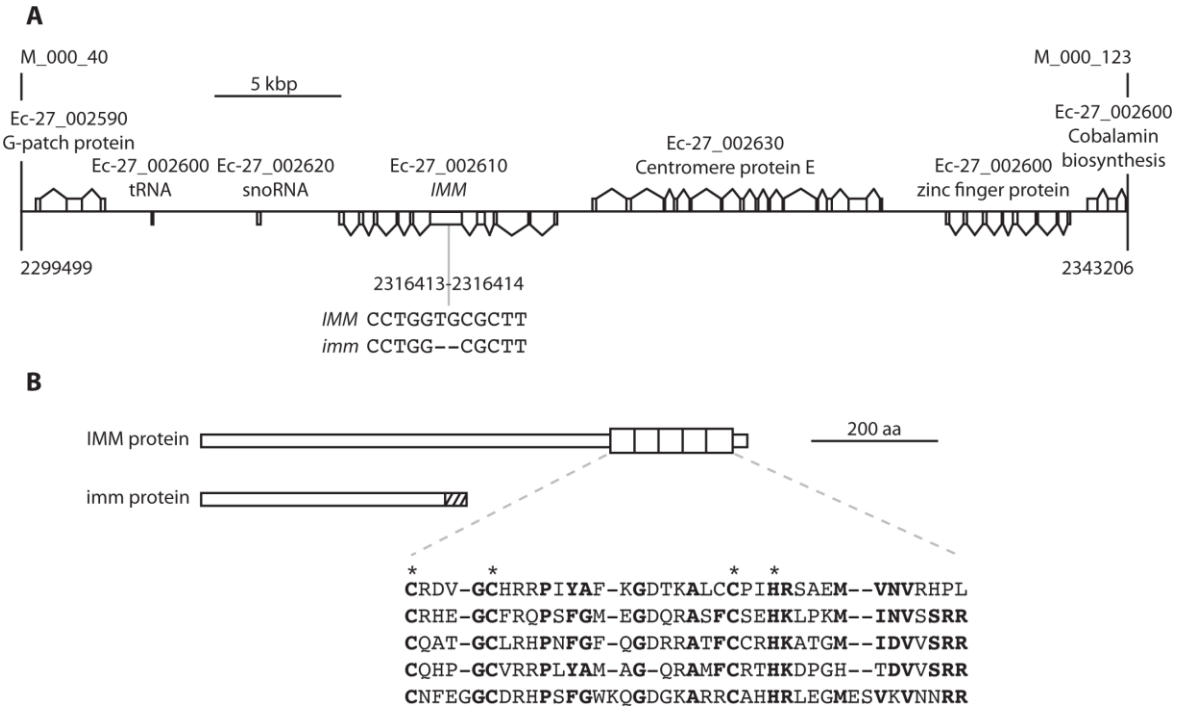

Fig. 1. **Positional cloning of the *IMM* gene and features of the encoded protein.**

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

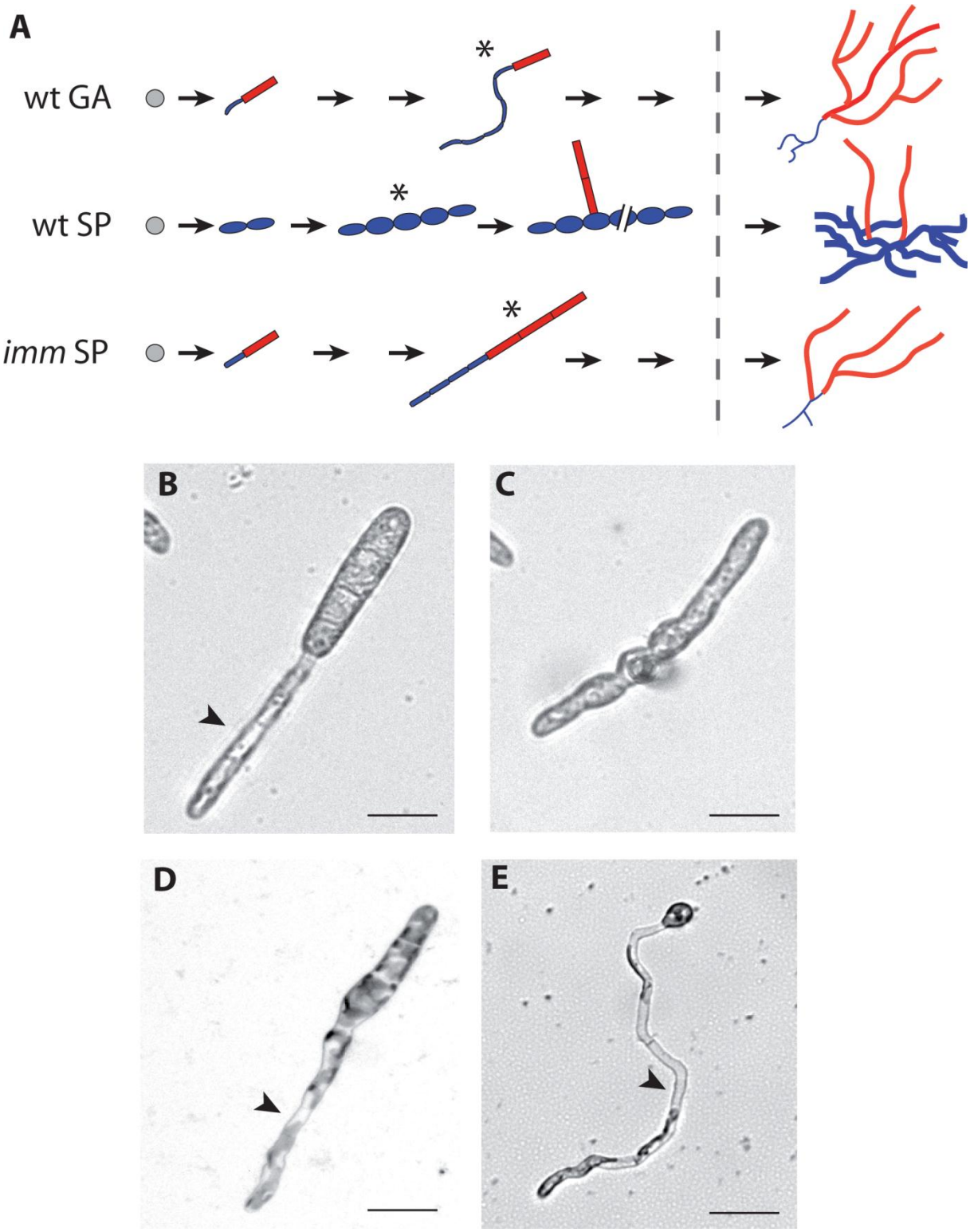

Fig. 2. Introduction of siRNAs that target the *IMM* gene induces a phenotype that mimics the *imm* mutation.

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

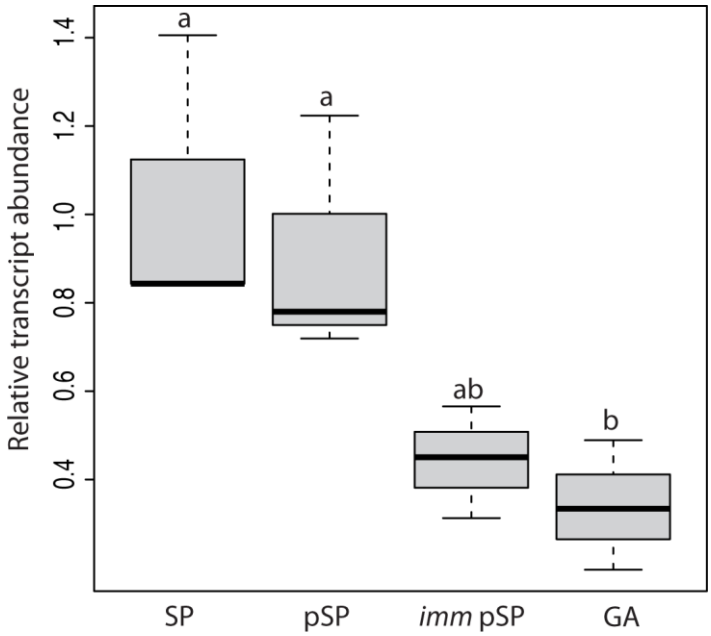

Fig. 3. Boxplot showing a quantitative reverse transcriptase polymerase chain reaction analysis of *IMM* transcript abundance at different stages of the *Ectocarpus* sp. life cycle. The dark bar indicates the median. SP, wild type diploid sporophyte; pSP, wild type partheno-sporophyte; *imm* pSP, partheno-sporophyte of the *imm* mutant; GA wild type gametophyte; a and b indicate statistically different samples (t test, $P<0.05$).

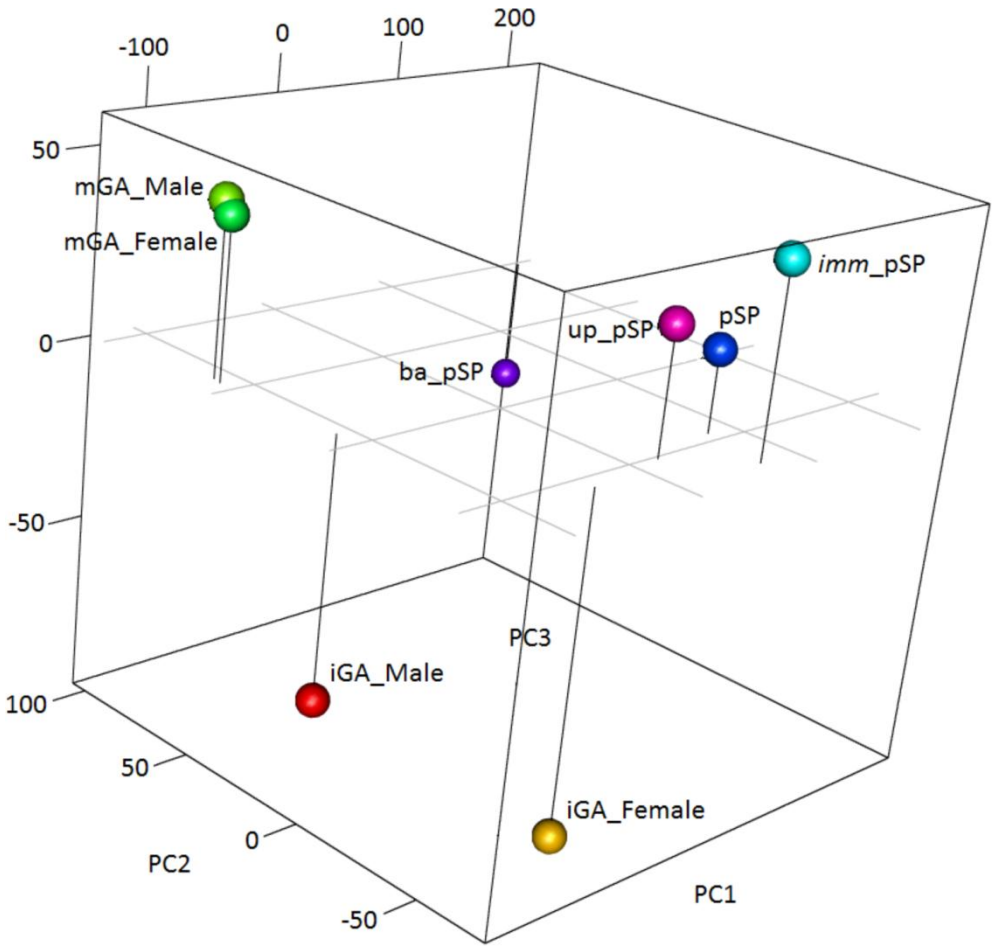

Fig. 4. Comparison of the *imm* partheno-sporophyte transcriptome with those of different tissues and life cycle stages of wild type *Ectocarpus* sp..

Principal component analysis was used to compare gene transcription patterns across samples (see Methods for details). The three dimensions PC1, PC2 and PC3 represent 51.9%, 15.4% and 12.2% of the variance respectively. *imm* pSP, *imm* partheno-sporophyte; pSP, wild type partheno-sporophyte; up_pSP, upright filaments of wild type partheno-sporophyte; ba_pSP, basal system of wild type partheno-sporophyte; iGA_Female, immature wild type female gametophyte; iGA_Male, immature wild type male gametophyte; mGA_Female, mature wild type female gametophyte; mGA_Male, mature wild type male gametophyte.

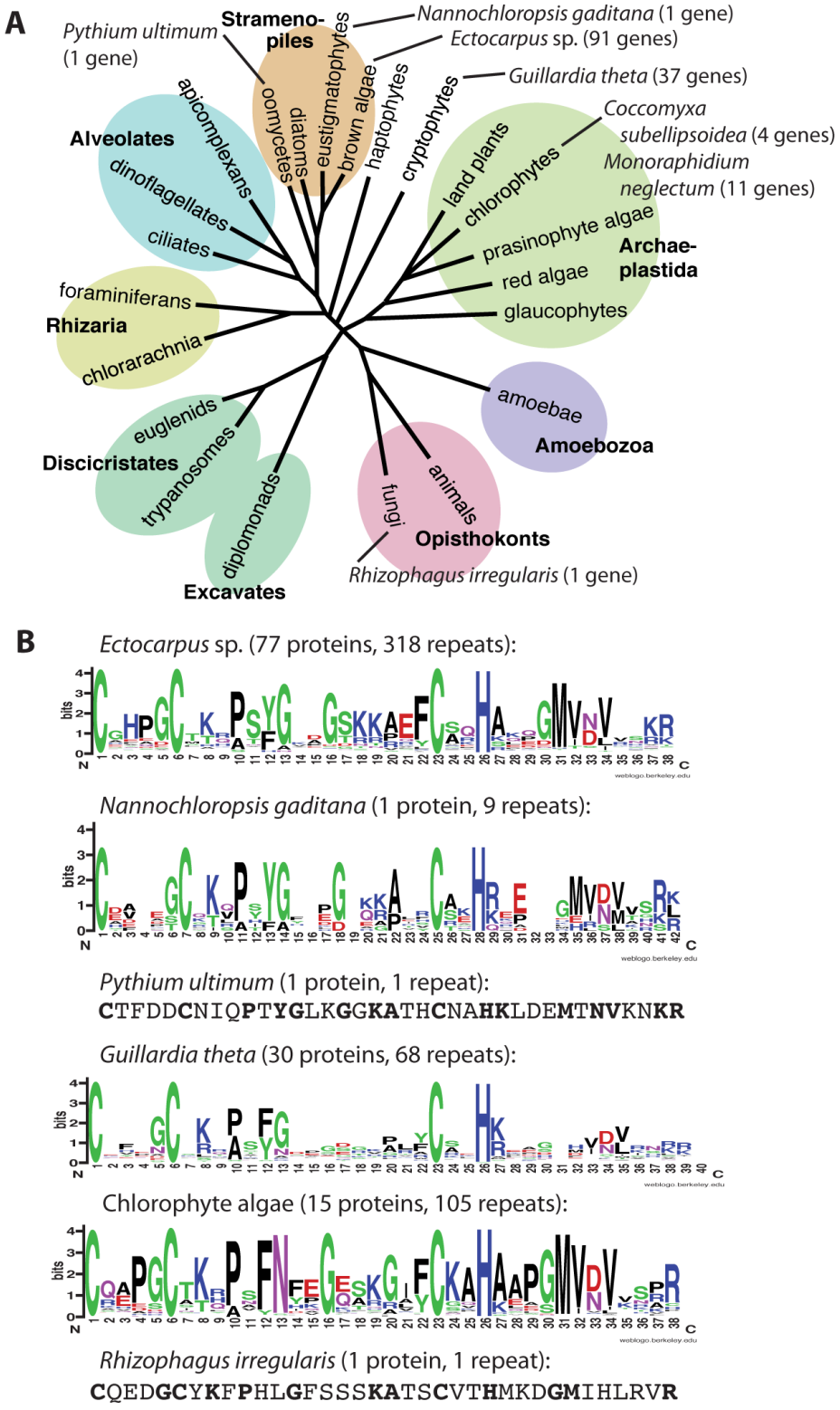

Fig. 5. Distribution of EsV-1-7 domain genes across the eukaryotic supergroups.

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

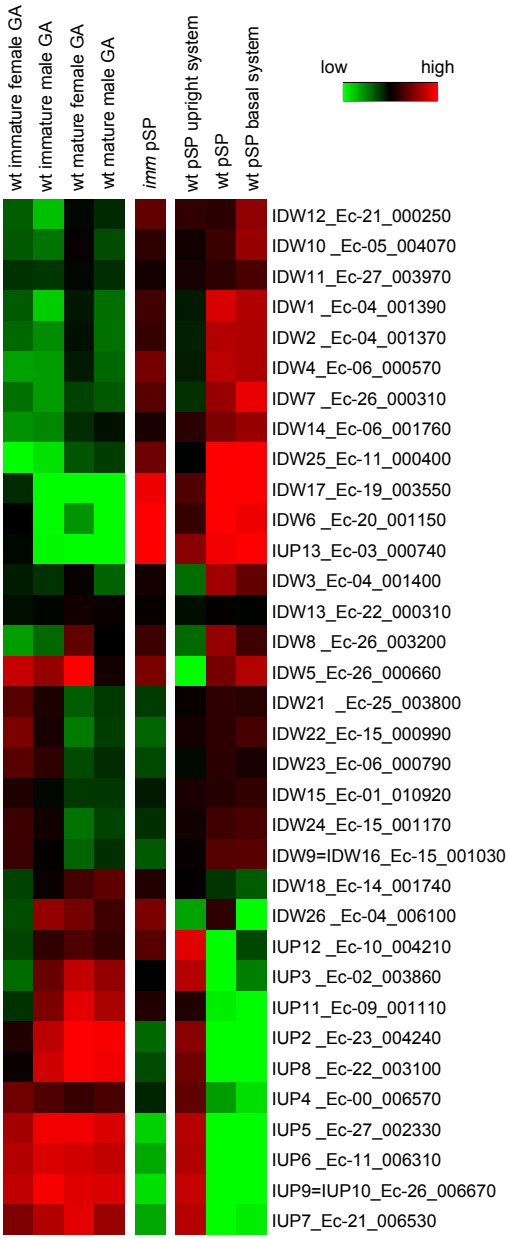

Fig. S1. Heat map representing abundances of transcripts of the *immediate upright* up-regulated (IUP) and *immediate upright* down-regulated (IDW) genes identified by Peters et al. (2008) at different stages of the life cycle and in different tissues.
Expression levels are based on log2 transformed TPMs. Gametophyte samples are on the left of the *imm* mutant partheno-sporophyte sample and sporophyte samples on the right. Note the similarity between transcript abundances in gametophyte samples and in sporophyte upright filaments. wt, wild type; *imm, imm* mutant; GA, gametophyte; pSP, partheno-sporophyte. Samples are whole organism unless otherwise stated. Gene identifiers (LocusIDs) have the format Ec-00_000000.

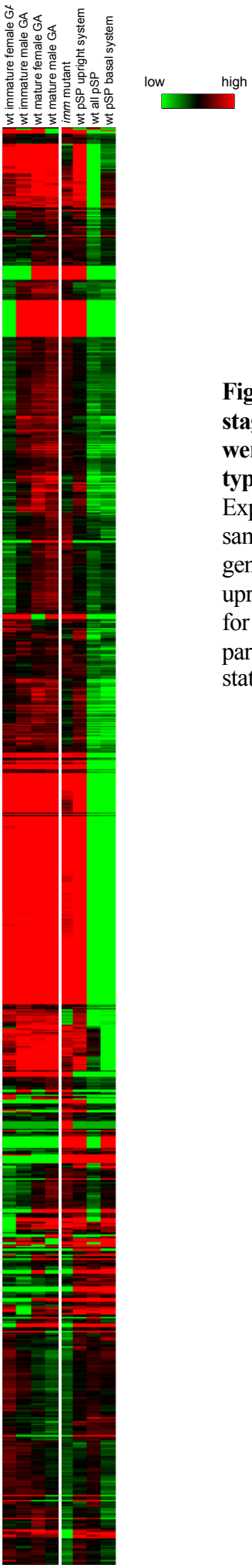

Fig. S2. Heat map representing transcript abundances at different stages of the life cycle and in different tissues for 1,087 genes that were significantly differentially expressed between the *imm* and wild type partheno-sporophytes.

Expression levels are based on log2 transformed TPMs. Gametophyte samples are on the left and sporophyte samples on the right. Note that genes that are upregulated in the gametophyte generation also tend to be upregulated in sporophyte upright filaments. Gene names were omitted for clarity. wt, wild type; *imm, imm* mutant; GA, gametophyte; pSP, partheno-sporophyte. Samples are whole organism unless otherwise stated.

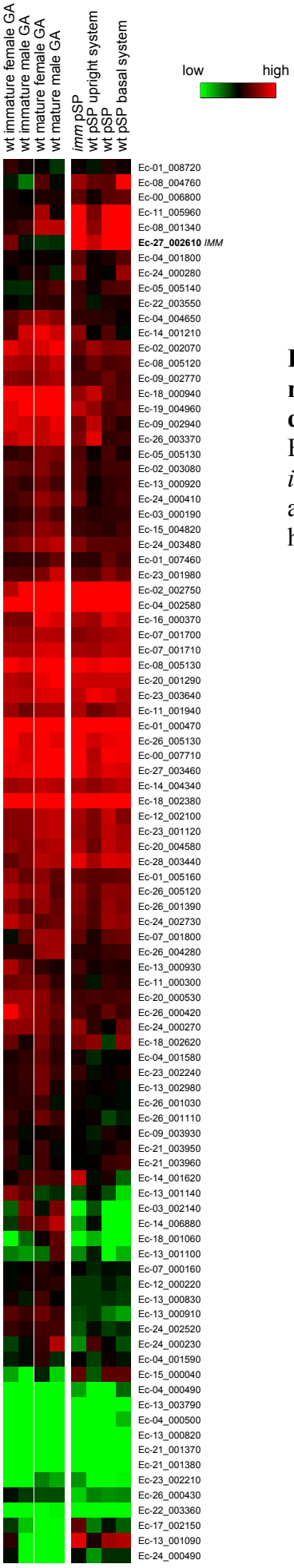

Fig. S3. Heat map representing abundances of transcripts of the members of the *Ectocarpus* sp. EsV-1-7 domain gene family at different stages of the life cycle and in different tissues.
Expression levels are based on log2 transformed TPMs. wt, wild type; *imm, imm* mutant; GA, gametophyte; pSP, partheno-sporophyte. Samples are whole organism unless otherwise stated. Gene identifiers (LocusIDs) have the format Ec-00_000000.

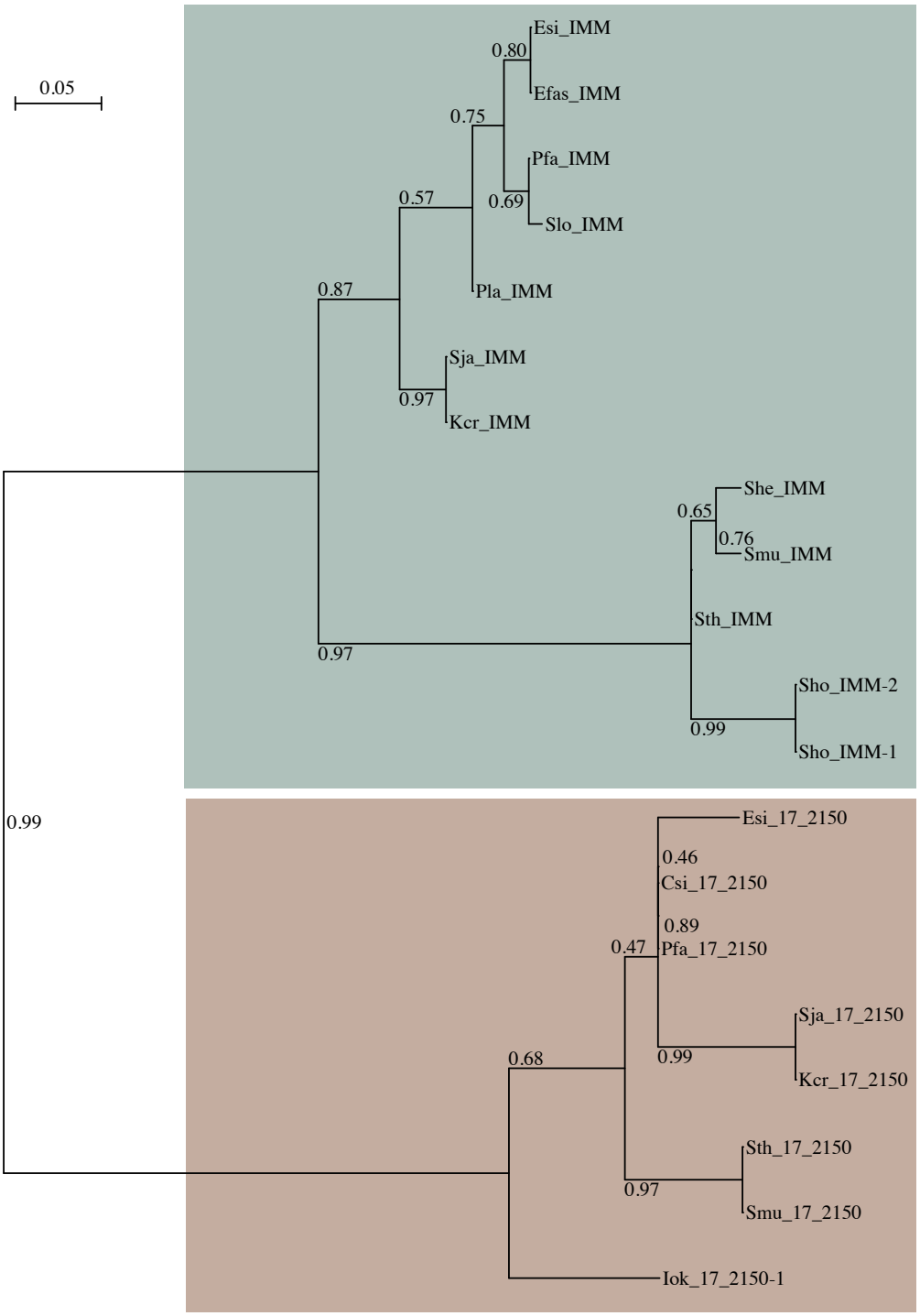

Fig. S4. Unrooted maximum likelihood tree based on an alignment of IMM and Ec-17_002150 orthologues from diverse brown algal species.
The tree was generated from a Muscle alignment with the MEGA7 maximum likelihood algorithm using the JTT+G model and 1000 bootstrap replicates. Bootstrap support values are indicated at each node. Ec-17_002150 is the most similar gene to *IMM* in the *Ectocarpus* sp. genome. IMM, IMM orthologue; 17_2150, Ec-17_002150 orthologue; Esp, *Ectocarpus* sp.; Efas, *Ectocarpus fasiculatus*; Slo, *Scytosiphon lomentaria*; Pfa, *Petalonia fascia*; Pla, *Punctaria latifolia*; Sja, *Saccharina japonica*; Kcr, *Kjellmaniella crassifolia*; Sho, *Sargassum horneri*; Sth, *Sargassum thunbergi*; She, *Sargassum hemiphyllum*; Smu, *Sargassum muticum*; Csi, *Colpomenia sinuosa*; Iok, *Ishige okamurai*.

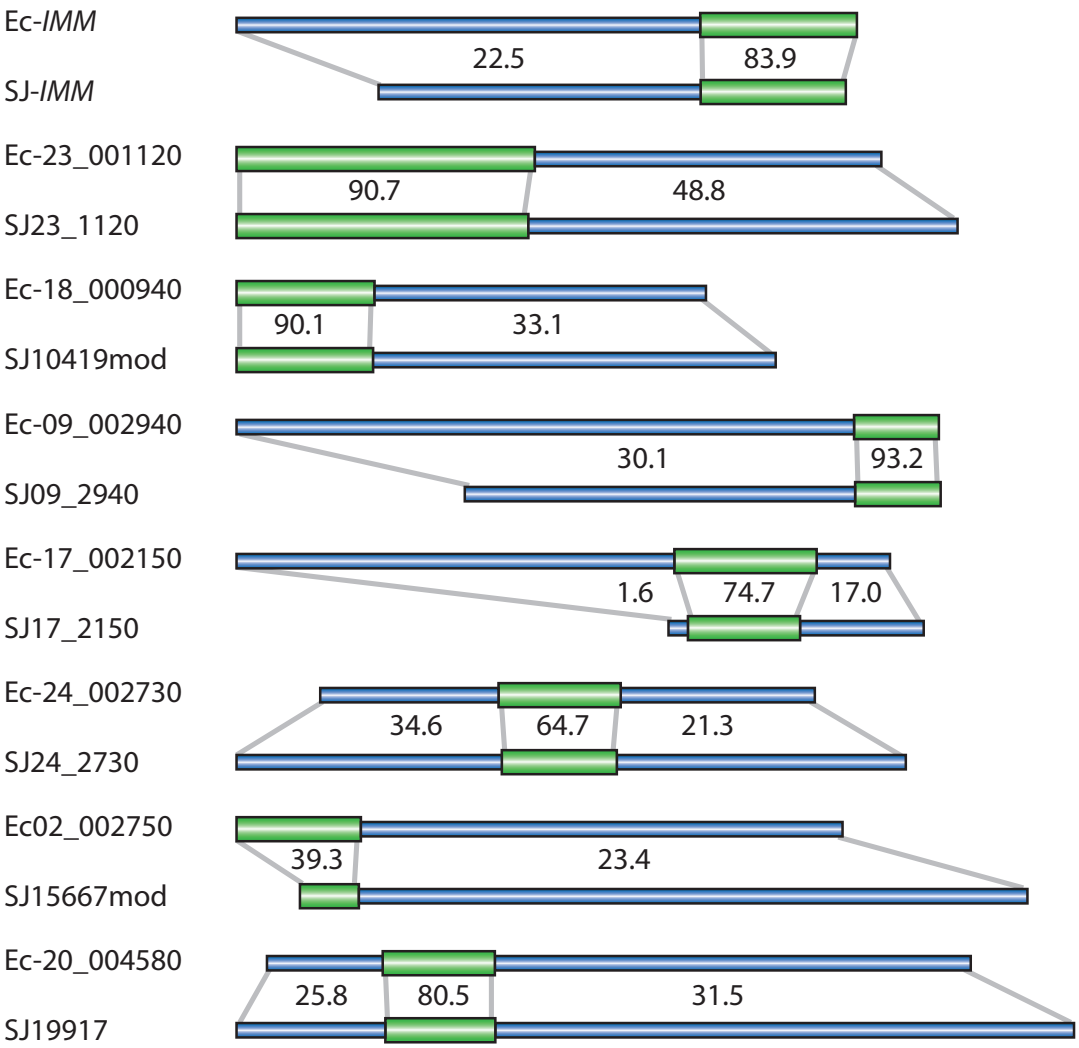

Fig. S5. Comparison of eight representative orthologous pairs of *Ectocarpus* sp. and *S. japonica* EsV-1-7 family proteins showing the weak conservation of regions outside the EsV-1-7 repeat domains. *Ectocarpus* sp. (Ec) and *S. japonica* (SJ) predicted proteins are represented by blue cylinders, with the EsV-1-7 repeat regions indicated by a thicker, green cylinder. Grey lines delineate equivalent domains in pairs of orthologues. Global percent amino acid identities between domains are indicated.

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

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

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

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

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

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

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

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

Click here to Download Tables S1-S8