RESEARCH

Open Access

CrossMark

Rapid turnover of life-cycle-related genes in the brown algae

A. P. Lipinska[1], M. L. Serrano-Serrano[2], A. Cormier[5], A. F. Peters[3], K. Kogame[4], J. M. Cock[1] and S. M. Coelho[1*]

Abstract

Background: Sexual life cycles in eukaryotes involve a cyclic alternation between haploid and diploid phases. While most animals possess a diploid life cycle, many plants and algae alternate between multicellular haploid (gametophyte) and diploid (sporophyte) generations. In many algae, gametophytes and sporophytes are independent and free-living and may present dramatic phenotypic differences. The same shared genome can therefore be subject to different, even conflicting, selection pressures during each of the life cycle generations. Here, we analyze the nature and extent of genome-wide, generation-biased gene expression in four species of brown algae with contrasting levels of dimorphism between life cycle generations.

Results: We show that the proportion of the transcriptome that is generation-specific is broadly associated with the level of phenotypic dimorphism between the life cycle stages. Importantly, our data reveals a remarkably high turnover rate for life-cycle-related gene sets across the brown algae and highlights the importance not only of co-option of regulatory programs from one generation to the other but also of a role for newly emerged, lineage-specific gene expression patterns in the evolution of the gametophyte and sporophyte developmental programs in this major eukaryotic group. Moreover, we show that generation-biased genes display distinct evolutionary modes, with gametophyte-biased genes evolving rapidly at the coding sequence level whereas sporophyte-biased genes tend to exhibit changes in their patterns of expression.

Conclusion: Our analysis uncovers the characteristics, expression patterns, and evolution of generation-biased genes and underlines the selective forces that shape this previously underappreciated source of phenotypic diversity.

Background

As a consequence of sexual reproduction, the vast majority of eukaryotes have life cycles involving an alternation between haploid and diploid phases [1, 2]. The proportion of the life cycle spent in each phase varies dramatically depending on the species. In organisms with haplontic cycles, mitosis only occurs in the haploid stage. Haploid mitosis may lead to asexual (clonal) reproduction, as in *Chlamydomonas* for example, or involve somatic growth and cellular differentiation as in *Volvox*. In these organisms, the zygote undergoes meiosis immediately after syngamy without undergoing any mitotic divisions. Conversely, in diplontic life cycles, mitosis only occurs during the diploid phase,

and meiosis takes place immediately before gamete formation. Diploid mitosis leads to asexual reproduction in unicellular lineages (e.g., diatoms) and to somatic growth and differentiation in multicellular organisms such as Metazoans. Finally, in organisms with haploid-diploid life cycles, mitotic cell divisions occur during both the haploid and diploid phases. In land plants and some algae, these mitotic divisions can lead to the development of two distinct multicellular organisms, one haploid and the other diploid. The haploid organism is generally referred to as the gametophyte, because it produces gametes, and the diploid organism as the sporophyte, because it produces spores. Note, however, that the gametophyte and sporophyte developmental programs are not absolutely linked to ploidy because ploidy and life cycle generation have been shown to be uncoupled during variant life cycles [3, 4]. The gametophyte and sporophyte should therefore be thought of as genetically controlled developmental

* Correspondence: coelho@sb-roscoff.fr
[1]Sorbonne Université, UPMC Univ Paris 06, CNRS, Algal Genetics Group, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688 Roscoff, France
Full list of author information is available at the end of the article

programs that are coordinated with, but not absolutely linked to, life cycle progression.

The evolutionary advantages of life cycles with dominant haploid, dominant diploid, or alternation between two phases have been subject to extensive theoretical work [5–10]. Models exploring the evolution of haploidy and diploidy assume an alternation of generations with free-living haploid and diploid phases, where expanding one phase reduces the other phase. These models predict that purging of deleterious mutations favors expansion of the haploid phase when recombination is rare, but that diploids are favored when recombination is common because mutations are masked from selection [6, 11]. In contrast, niche differentiation between haploids and diploids may favor the maintenance of biphasic life cycles, in which development occurs in both phases [12]. For instance, gametophytes have been shown to exploit low-resource environments more efficiently whereas sporophytes are more vigorous when resources are abundant [13]. The interplay between genetic and ecological factors has been recently explored [9] in a model that assumes different effects of mutations on haploids and diploids of competition between individuals within a generation. The model predicts that temporal variations in ecological niches stabilize alternation of generations. Empirical support for these models has come from the brown alga *Ectocarpus* sp., where dimorphism between generations has been linked to the occupation of different spatio-temporal niches [14].

In organisms with complex life cycles, an allele may be relatively beneficial when expressed in one generation but deleterious when expressed in the other generation (generation antagonism), and in this case, selection acts in opposite directions in haploids and diploids [9, 15]. With this type of generation-dependent antagonistic selection, evolution favors the expansion of whichever generation gains the greatest fitness advantage, on average, from the conflicting selection pressures [15]. Generation antagonism is expected to be particularly relevant in multicellular species where there is alternation of generations with morphologically dissimilar gametophytes and sporophytes, as in the case of many plants and algae. When fitness optima differ between the gametophyte and sporophyte generations for a shared trait, dimorphism can allow each generation to express its optimum trait phenotype. Accordingly, the evolution of generation-biased gene expression may be one mechanism that could help to resolve this intra-locus "generation" conflict, in a similar manner to mechanisms that resolve sexual antagonism [16, 17]. Another potential solution to resolve generation conflict is gene duplication, followed by divergence of the two loci towards distinct optima corresponding to each of the two generations. An equivalent process has been shown to be important in the generation of sex-biased gene expression [17, 18]. While the role of sexual selection

in shaping phenotypic diversity and in driving patterns of evolution of gene expression has been studied extensively (e.g., [19, 20]), we have remained so far largely ignorant about the relationships between generation-biased selection, generation-biased gene expression, and phenotypic differentiation.

The brown algae (Phaeophyceae) are a group of complex multicellular eukaryotes that diverged from plants and animals more than a billion years ago [21]. Brown algal life cycles are extraordinarily diverse, exhibiting a broad range of variation in terms of the relative complexities of the gametophyte and sporophyte generations [22, 23]. Here, we selected two pairs of brown algal species from the orders Ectocarpales and Laminariales, which diverged about 95 Mya [24], to trace the evolutionary history of generation-biased gene expression in the brown algal lineage. The selected species exhibit markedly different levels of dimorphism between life cycle generations: the Laminariales species *Macrocystis pyrifera* and *Saccharina japonica* have complex sporophyte but highly reduced gametophyte generations, whereas the Ectocarpales species include *Scytosiphon lomentaria*, which has a reduced sporophyte but a morphologically complex gametophyte, and *Ectocarpus* sp. which has gametophyte and sporophyte generations of similar complexity. We show that a large proportion of the transcriptome of brown algae exhibit generation-biased expression and that the set of life-cycle-biased genes turns over extremely rapidly during evolution due to a combination of two processes: *de novo* birth of genes with generation-biased expression and gain/loss of generation-biased expression by orthologous loci. Our results uncover the characteristics, expression patterns, and evolution of generation-biased genes and underline the selective forces that shape this previously underappreciated source of phenotypic diversity.

Results
Assembly and annotation of reference genomes for four brown algae
The identification and analysis of generation-biased genes carried out in this study was based on an analysis of assembled genome sequences for four brown algae, including two high-quality, published genomes for *Ectocarpus* sp. and *S. japonica* [21, 25] and two draft genome sequences for *S. lomentaria* and *M. pyrifera*. The two draft genomes were assembled de novo with Masurca [26] using publicly available sequence data [27]. The *M. pyrifera* assembly consisted of 160,020 scaffolds corresponding to a total size of 581 Mbp, which is similar to the size of the genome of *S. japonica* [25]. The 27,450 scaffolds of the *S. lomentaria* draft assembly corresponded to 218 Mbp which, again, is in the expected range for a member of the Ectocarpales [21, 28]. PASA [29] was used to generate gene predictions for *S.*

lomentaria and *M. pyrifera* based on mapping of RNA-seq data, de novo transcriptome assembly, and *ab initio* gene prediction (see the "Methods" section for details). Both of the draft genome assemblies recovered about 80% of the BUSCO v3 eukaryotic gene set [30] (including both complete and fragmented matches) (see Additional file 1: Table S1 for detailed genome statistics). Considering that the reference genomes recover 95.0% (*Ectocarpus* sp.) and 91.1% (*S. japonica*) of this BUSCO gene set, the BUSCO scores for *S. lomentaria* and *M. pyrifera* assemblies indicate that these draft genomes are of good quality.

Measurement of phenotypic differentiation between gametophyte and sporophyte generations

The number of different cell types in each generation and the ratios of the sizes of the gametophyte and the sporophyte at maturity were used as proxies to assess the degree of morphological complexity and the level of phenotypic dimorphism between life cycle generations in the four brown algal species studied (Additional file 1: Table S2; Additional file 2: Figure S1). Using these parameters, the Laminariales species *M. pyrifera* and *S. japonica* exhibited the highest level of phenotypic differentiation between generations, with the sporophyte being more complex than the gametophyte, both in terms of the number of cell types and in terms of size. As far as the Ectocarpales species were concerned, dimorphism between generations was also marked in *S. lomentaria*, but with the gametophyte being more complex than the sporophyte. *Ectocarpus* sp. exhibited the lowest level of differentiation between the gametophyte and sporophyte generations (Additional file 1: Table S2; Additional file 2: Figure S1).

Patterns of generation-biased gene expression in gametophytes and sporophytes

Analyses of generation-biased gene expression used published RNA-seq datasets (at least two replicate samples) for gametophytes and sporophytes of the model brown alga *Ectocarpus* sp. [27, 31], *S. japonica* [25, 32] and *M. pyrifera* [27, 33]. For *S. lomentaria*, a published dataset was available for the gametophyte generation [31] and we generated RNA-seq data for duplicate samples of sporophytes (Additional file 1: Table S3 and "Methods" section for details). DEseq2 was used to compare patterns of gene expression in gametophytes and sporophytes for each of the four brown algal species. Note that only two replicates were available for a subset of the samples (Additional file 1: Table S3). Although DEseq2 has been shown to successfully capture the majority of the true differential expression signals for the most strongly changing genes, being largely insensitive to replicate number in this case [34], the detection rate of genes with smaller fold changes will be affected when only duplicate samples are used. Thus, the total number of generation-biased genes may be underestimated.

The proportion of genes that showed generation-biased expression was similar for the four study species (33–36% of genes in each genome), with the highest level of generation-biased gene expression (36%) being detected in *S. japonica* (Fig. 1a). In the Ectocarpales, more transcripts were gametophyte-biased than sporophyte-biased (Fisher exact test, p value $< 2e{-}16$ for both *Ectocarpus* sp. and *S. lomentaria*). This difference was most marked in *S. lomentaria*, where almost twice as many genes were gametophyte-biased (20% of the transcriptome) than were sporophyte-biased. In both *Ectocarpus* sp. and *S. lomentaria*, the fraction of sporophyte-biased transcripts was relatively low (12% and 13%, respectively) but the proportion was higher in species that have a more conspicuous sporophyte generation (i.e., both Laminariales), with 16–19% of the transcriptome being sporophyte-biased (Fig. 1a).

Generation-biased genes were defined as being generation-specific when the TPM for one of the two generations was below the fifth percentile (see the "Methods" section, Additional file 1: Table S4) (Fig. 1a). Between 20 and 44% of gametophyte-biased genes exhibited gametophyte-specific expression patterns and 3 to 24% of sporophyte-biased genes exhibited sporophyte-specific expression patterns, depending on the species. The proportion of generation-specific genes was larger for the gametophyte than for the sporophyte generation in *Ectocarpus* sp. and *S. lomentaria*. This trend was particularly marked for *S. lomentaria* (which has a dominant gametophyte generation) where nearly half of the gametophyte-biased genes were gametophyte-specific. In contrast, in the two Laminariales species, similar proportions of generation-specific genes were observed in both generations (Fig. 1a).

To examine the relationship between the degree of generation-biased expression and transcript abundance (expression level), the generation-biased genes were grouped according to the fold change (FC) difference between gametophyte and sporophyte samples, and the mean expression levels in gametophytes and sporophytes ($\log_2$TPM) were plotted for each group (Fig. 1b). This analysis indicated that, overall, the most marked levels of generation-biased expression (high fold changes) were the result of downregulation of genes in the generation where they were expressed more weakly, rather than strong upregulation in the generation where they were expressed more strongly. However, for gametophyte-biased genes, the expression in sporophytes reached the lower threshold (about $\log_2$TPM < 0) much faster than the expression of sporophyte-biased genes in gametophytes. In other words, when genes exhibited a moderate to high degree of gametophyte-biased expression, this was predominantly
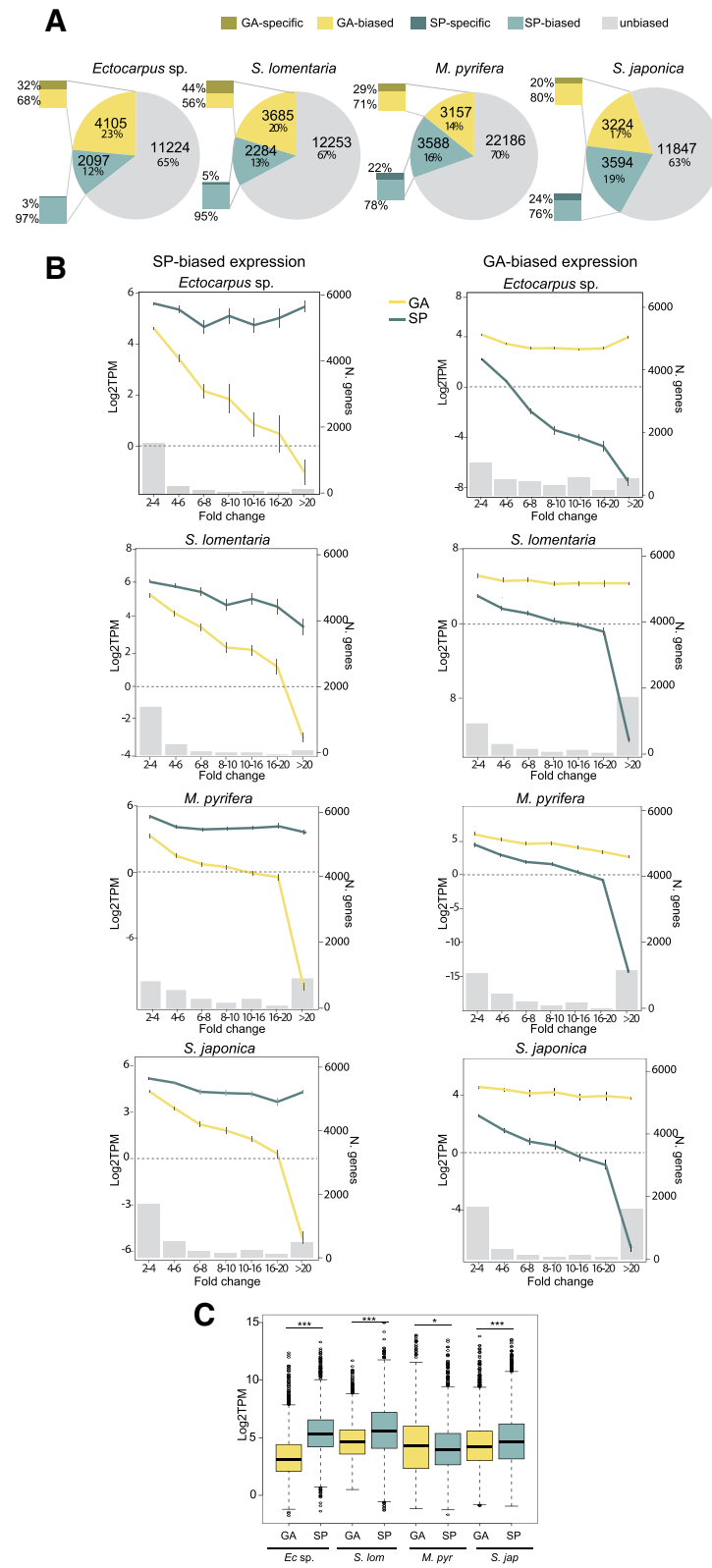
Fig. 1 (See legend on next page.)

(See figure on previous page.)
Fig. 1 Generation-biased gene expression across the four brown algal species. **a** Proportions of unbiased, gametophyte- and sporophyte-biased genes across the four studied species. Bar inserts represent the proportion of generation-specific genes among the generation-biased genes in each species. **b** Mean gene expression levels (log2TPM) at several degrees of generation bias (fold change, FC, represented by gray histograms) for gametophyte-biased (yellow) and sporophyte-biased (blue) genes in the four studied species. The number of genes in each category of FC is represented on the right side of the graph. Error bars represent standard errors. GA gametophyte, SP sporophyte. **c** Boxplot showing the mean expression levels (log2TPM) of gametophyte- and sporophyte-biased genes

due to strong downregulation (silencing) of these genes in the sporophyte generation (Fig. 1b).

Interestingly, in the Ectocarpales species, more than 80% of the sporophyte-biased genes exhibited fold changes of between 2 and 6, whereas in the Laminariales species (which have a dominant sporophyte generation), a greater proportion of the sporophyte-biased genes exhibited very high fold changes between generations, with between 13 and 29% in *S. japonica* and *M. pyrifera*, respectively, exhibiting fold changes of more than 20 (Fig. 1b; Additional file 1: Table S5). Nevertheless, in all four species, the majority of the generation-biased genes with very strong bias (FC > 20) were gametophyte-biased (Fig. 1b; Additional file 1: Table S5), with as many as 48% of the gametophyte-biased genes in *S. lomentaria* belonging to this group (15% in *Ectocarpus* sp., 34% in *M. pyrifera*, and 44% in *S. japonica*).

We also noted that, on average, sporophyte-biased genes were expressed at significantly higher levels than gametophyte-biased genes in all four species (Wilcoxon test, *p* value < 0.02 in all pairwise tests) (Fig. 1c). There was an overall tendency for the number of gametophyte-specific genes to be positively correlated with more conspicuous and complex gametophytes (Spearman's rank correlation rho = 0.949, *p* value = 0.051). For instance, *S. lomentaria*, which has a dominant gametophyte generation, possessed a significantly higher proportion of gametophyte-biased and gametophyte-specific genes than sporophyte-biased genes (Fisher test, *p* value < 2e−16) (Additional file 1: Table S6, Additional file 2: Figure S1).

High turnover of generation-biased gene sets in the brown algae

Orthology relationships were predicted using Orthofinder [35] to assess the conservation of generation-biased genes across the four species. Orthofinder assigned genes to 12,891 orthogroups (OGs). Each OG contained a set of homologous proteins (orthologs and/or paralogs) that were present in one or more of the algal species studied. Comparisons of OGs containing generation-biased genes showed that they were poorly conserved between pairs of species (Fig. 2a). Only 29% and 26% of the OGs containing generation-biased genes were shared by the Ectocarpales and Laminariales species pairs respectively, and conservation between pairs of species from different orders was even lower (17 to 18%). Only 3% of the generation-biased

OGs were conserved across all four of the study species (Fig. 2a). A proportion of the generation-biased genes (up to 18%) did not have orthologs in the genome of any of the other three study species nor in the genomes of four other distant Stramenopile species (Fig. 2b; Additional file 1: Table S7). We refer to these taxonomically restricted genes hereinafter as "orphan" genes. The orphan genes were not included in the OG analysis described above (Fig. 2a) because most orphans are not members of an OG. The analysis of OGs therefore actually overestimated the degree to which generation-biased gene sets were conserved across species. Orphan genes were not over-represented in the generation-biased gene set but, interestingly, generation-biased orphan genes exhibited higher levels of fold change overall compared with the generation-biased genes that are members of OGs (Wilcox test, *p* value < 10e−3; Fig. 2b). This was particularly pronounced in *S. lomentaria,* where about half of the gametophyte-biased orphan genes had expression levels at least 30 times higher ($\log_2$FC > 4.9) in the gametophyte than in the sporophyte. In *S. japonica,* conversely, it was the sporophyte-biased orphan genes that presented overall higher fold changes (Fig. 2b). In other words, orphan genes showed a stronger magnitude of generation bias than older genes.

Evolution of generation-biased expression of orthologous genes across brown algal species

To further analyze the evolutionary history of the generation-biased genes, we focused on genes for which there was a clear orthologous relationship across the four species. The large set of 12,891 OGs, identified using Orthofinder, was screened to identify 6656 single copy orthologous genes with either 1:1:1:1 or 1:1:1:0 occurrence across the four brown algal species (see the "Methods" section for details). We will refer to this set of OGs as "all single orthologs" (ASOs).

The ASO dataset was used to assess the conservation of generation-biased gene expression across the four species. Of the 6656 ASOs, 5027 (76%) included genes that were generation-biased in at least one of the species. However, only 22 gametophyte-biased genes and 10 sporophyte-biased genes consistently exhibited patterns of generation-biased expression across all four species (Additional file 2: Figure S3A, Additional file 1: Table S11). The number of genes with conserved generation-biased
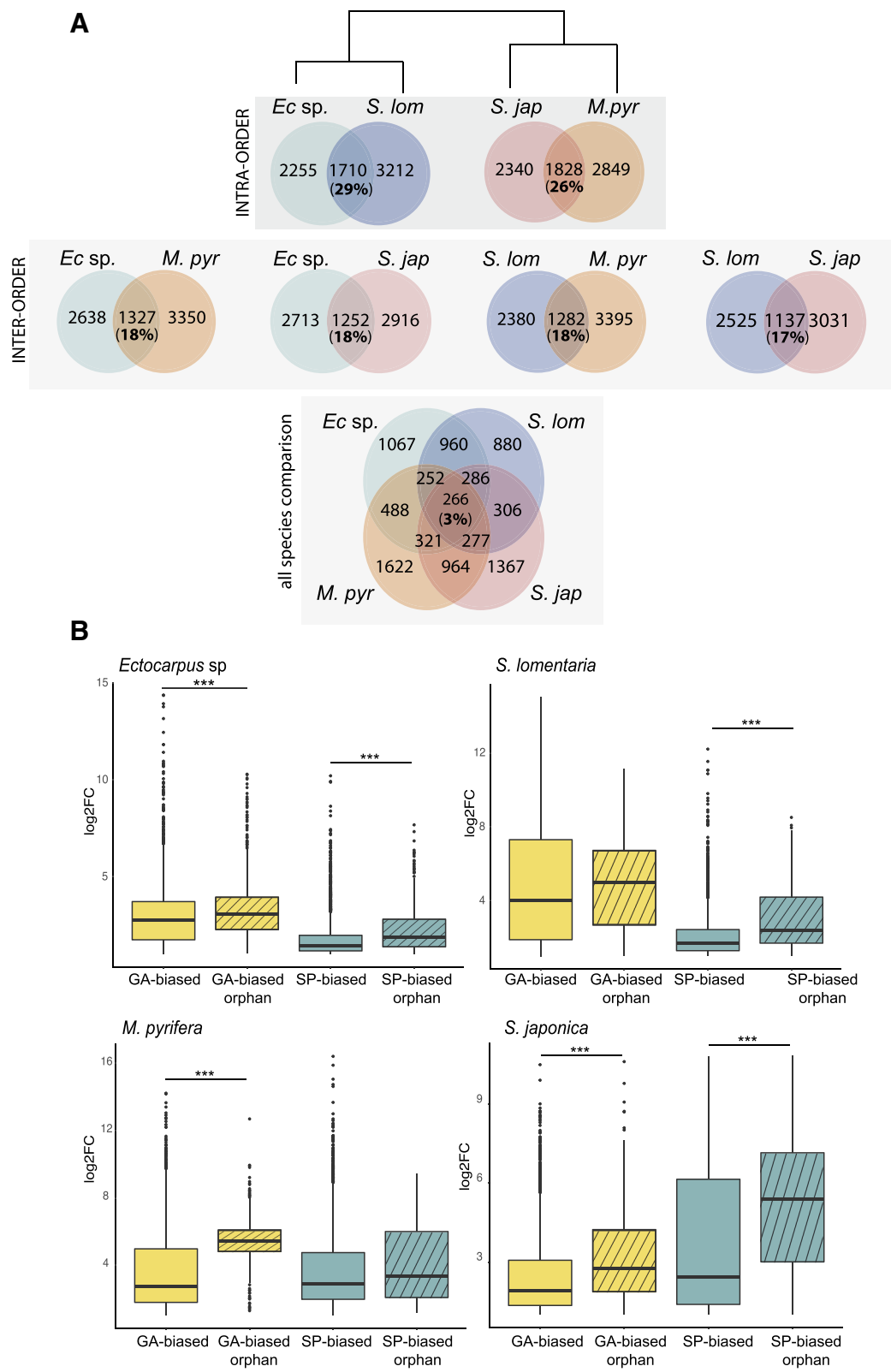
Fig. 2 (See legend on next page.)

expression increased to 167 gametophyte-biased and 116 sporophyte-biased when we took into account orthologous genes with generation bias in three species (with the ortholog missing or unbiased in the fourth species) (Fig. 3a). Eighteen percent of the ASOs that included generation-biased genes (898 of the 5027) exhibited discordant generation-biased expression patterns, so that, for example, the ortholog of a gene that was sporophyte-biased in one species was gametophyte-biased in at least one of the other three species (Fig. 3a).

We used hierarchical clustering of expression levels for all the members of the 1:1:1:1 subset of the ASO dataset with at least one generation-biased member in one of the studied species to visualize global transcription patterns within and among the four species. In this analysis, samples clustered primarily by species and not according to life cycle stage (Fig. 3b), reflecting the low level of conservation of generation-biased expression patterns of gene expression across the lineages.

Taken together, these analyses indicated that overall, generation-biased expression of the ASO dataset was extremely poorly conserved across the four brown algal species.

Generation-biased gene expression within the Ectocarpales and Laminariales

To analyze divergences of generation-biased expression patterns within orders, we used the Orthofinder analysis to identify single copy (1:1) orthologs shared either by the two Ectocarpales (6644 OGs) or by the two Laminariales (7128 OGs) species. These sets of 1:1 OGs were termed "pairwise single orthologs" (PSOs) (Additional file 1: Table S8).

Between 22 and 32% (Ectocarpales) and 34 and 45% (Laminariales) of the generation-biased genes were PSOs. Note however that the numbers of PSOs may be slightly underestimated as benchmarking of the S. lomentaria and M. pyrifera genome assemblies using BUSCO v3 indicated that they were not fully complete (Additional file 1: Table S1). Between 29 and 44% (for both Ectocarpales and Laminariales) of the PSOs gained either sporophyte- or gametophyte-biased expression in one of the
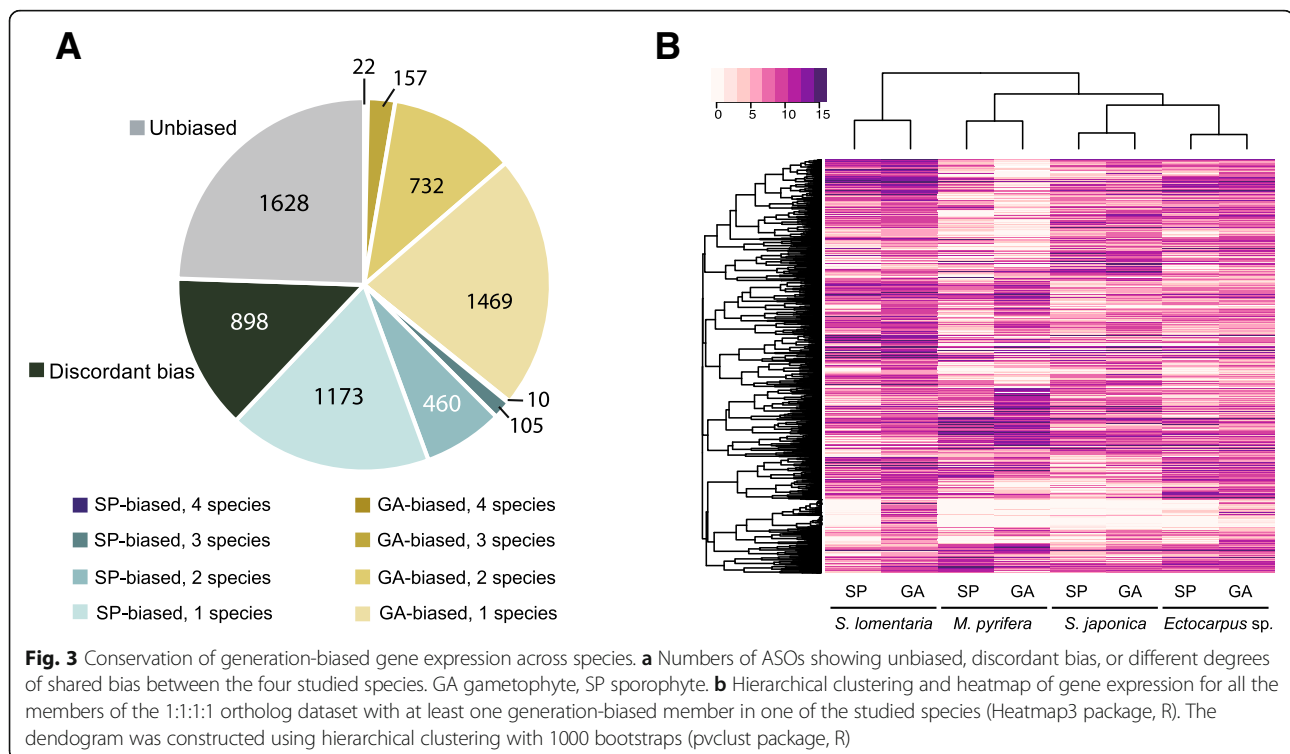

Fig. 3 Conservation of generation-biased gene expression across species. **a** Numbers of ASOs showing unbiased, discordant bias, or different degrees of shared bias between the four studied species. GA gametophyte, SP sporophyte. **b** Hierarchical clustering and heatmap of gene expression for all the members of the 1:1:1:1 ortholog dataset with at least one generation-biased member in one of the studied species (Heatmap3 package, R). The dendogram was constructed using hierarchical clustering with 1000 bootstraps (pvclust package, R)
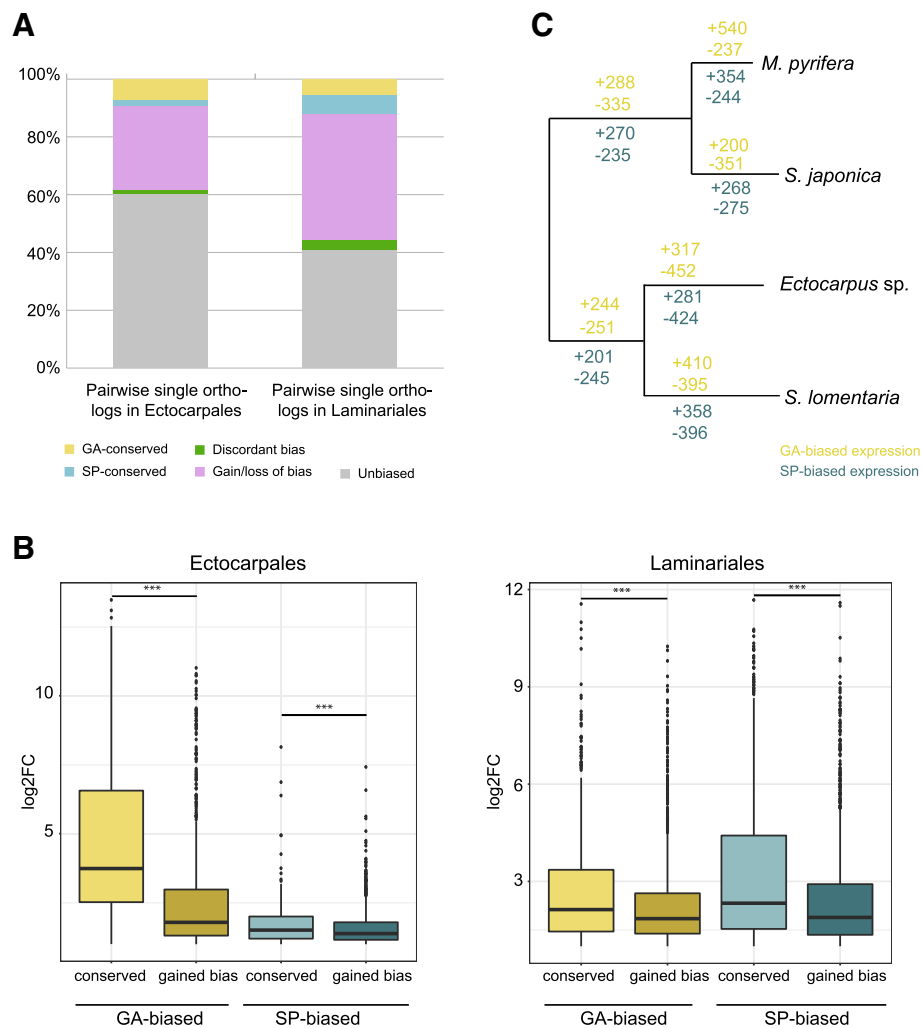
Fig. 4 Conservation of generation bias across the Ectocarpales (*Ectocarpus* sp. and *S. lomentaria*) and Laminariales (*S. japonica* and *M. pyrifera*). **a** Pairwise single orthologs (PSOs) in Ectocarpales and Laminariales. Genes with conserved gametophyte (yellow) or sporophyte bias (blue) exhibited the same bias in the same generation in the two species. Genes with "discordant bias" (green) were gametophyte-biased in one species and sporophyte-biased in the other species. "Gain/loss of bias" genes (pink) were generation-biased in one species but not in the other species. Unbiased PSOs are shown in gray. **b** Overall level of generation-biased expression (log2FC) for PSOs that are conserved versus PSOs that gained bias in Ectocarpales and Laminariales. **c** Representations of generation-biased gene gain/loss events across the branches of the Ectocarpales and Laminariales phylogeny. Expected numbers of events are based on multiple stochastic mappings (see the "Methods" section for detail)

species (Fig. 4a), whereas discordant generation-biased expression was observed for 1.7% (Ectocarpales) and 3.6% (Laminariales) of the PSOs (Fig. 4a). Therefore, high turnover (gain/loss) of generation-biased gene expression patterns was also observed at the order level.

A correlation was observed between the level of bias and the conservation of generation-biased genes within each lineage (Ectocarpales and Laminariales), i.e., the mean fold change in expression for genes that were conservatively generation-biased (both gametophyte-biased or both sporophyte-biased) across the species pairs was significantly higher than that for genes that were

generation-biased in one of the species but unbiased in the other (Fig. 4b; Wilcoxon test, *p* value < 5e−07).

Taken together, these analyses suggested that the rapid turnover of generation-biased gene sets involves not only the emergence of new generation-biased genes but also the emergence, in a species-specific fashion, of novel generation-biased expression patterns associated with existing orthologous genes.

Evolutionary history of generation-biased gene sets

We used a phylogenetic stochastic mapping approach to investigate the evolution of generation-biased gene

expression. Phylogenetic stochastic mapping allows reconstruction of the history of trait changes (in our case, generation bias) based on the estimation of the probabilities and expectations of gain and loss events of bias for each branch of an underlying phylogenetic tree [36]. Rates of gain and loss of expression bias were equal for both gametophyte and sporophyte bias, as determined by a likelihood ratio test between the ER and ARD models (all *p* values > 0.05). The stochastic mapping results highlighted widespread and rapid turnover of generation-biased gene expression during the evolution of the Laminariales and Ectocarpales (Fig. 4c). Specifically, more events of gain of bias were observed for gametophyte-biased genes, and, conversely, sporophyte-biased genes presented more events of loss of bias. Overall, gametophyte-biased genes presented a higher total number of events compared with sporophyte-biased (3469 versus 3008; chi-squared test = 16.281, df = 1, *p* value = 5.462e−5). However, the mean of the inferred transition rates across genes, calculated using the maximum likelihood and fitMk functions in the phytools R package, was 1.301 for gametophyte bias and 1.317 for sporophyte bias, indicating that, overall, the rate of turnover was similar for both generations.

Duplicated generation-biased genes

Gene duplication and generation-specific co-option of paralogs may be a mechanism to resolve potential generation antagonism due to evolutionary divergence between the two generations. Analysis of in-paralogs identified by Orthofinder indicated that the generation-biased gene sets of *S. lomentaria* and *M. pyrifera* were not enriched in members of duplicated gene pairs compared with the rest of the genes in each genome (Fisher test, p value = 0.4 and p value = 0.8 respectively). However, duplicated genes constituted 24 and 27% of gametophyte- and sporophyte-biased genes in *S. japonica* and 17% of gametophyte- and sporophyte-biased genes in *Ectocarpus* sp. (Additional file 1: Table S7, Additional file 2: Figure S4), which was significantly more than expected by chance (Fisher test, p value < 2e−16 for both comparisons).

Discordant bias was observed for 28% and 21% of generation-biased in-paralog pairs in *S. japonica* and *Ectocarpus* sp., respectively. The set of in-paralogs with discordant bias was completely different for each species, indicating that duplication of genes followed by acquisition of two, opposite generation-biased expression patterns by the resulting in-paralogs occurred independently in each of the species.

Predicted functions of generation-biased genes

An analysis of gene ontology (GO) terms associated with the generation-biased genes was carried out using Blast2GO [37] to search for enrichment in particular functional groups. First, Blast2GO analysis was carried out for each species, in order to relate gene function to the phenotypic generation dimorphisms specific to each species. Note that only 61% of the genes in the *Ectocarpus* sp. genome have a predicted function [38], reflecting the large evolutionary distances (more than a billion years) between brown algae and classical plant and animal models [21].

The GO terms associated with *M. pyrifera* and *S. japonica* sporophyte-biased genes were enriched in biological processes related to reproduction, carbohydrate metabolism, protein modification, growth and development, signaling, cell communication, response to external stimulus, and homeostasis (Fisher exact test, *p* value < 0.05, Additional file 1: Table S9). Interestingly, a similar set of GO terms was enriched for the gametophyte-biased genes of *S. lomentaria*, in addition to categories related to sexual reproduction and cilium motility (Additional file 2: Figure S2, Additional file 1: Table S9). This result suggests that similar genetic processes are at work in the morphologically complex, long-lived, dominant generations of these three species, despite the large morphological differences between gametophytes of *S. lomentaria* and sporophytes of Laminariales and the limited number of shared generation-biased genes.

Analysis of the gametophyte-biased genes from all the studied species identified four GO terms related to biological processes that were consistently significantly enriched in all species (Fisher exact test *p* value < 0.05). These terms included microtubule and flagellar movement-related categories and corresponded to between 10 and 50% of gametophyte-biased genes with assigned ontologies (Additional file 1: Table S10). Eleven GO terms were consistently enriched for the sporophyte-biased genes of all the studied species. These terms, which were related to carbohydrate metabolism and small GTPase signaling processes, corresponded to between 10 and 20% of the sporophyte-biased genes in each species. GO terms related to signal transduction and protein-protein interactions were also enriched in the gametophyte-biased genes (Additional file 1: Table S10). Among the OGs conservatively gametophyte-biased across all the studied species, OG 0001298 containing orthologs of Ec-03_003430 was particularly interesting because its predicted product shares 35% amino acid identity with the human sperm flagellar protein 2 (SPEF2). *SPEF2*, which has orthologs in a range of species including vertebrates, *Drosophila*, and protozoans with motile cilia or flagella [39], is predominantly expressed in the testis and spermatocytes and has been shown to be involved in male germ cell differentiation and sperm motility in animals [40, 41].

Structural characteristics of the generation-biased genes

Several structural characteristics (GC and GC3 content, coding region size, and intron number) were compared

between sporophyte-biased, gametophyte-biased, and unbiased genes (Additional file 2: Figure S3; Additional file 1: Table S12). Gametophyte-biased genes tended to have longer coding regions, to possess more introns, and to have a lower GC3 content than unbiased genes in all four species (Wilcoxon test, p value < 3e−05). In contrast, sporophyte-biased genes did not present a consistent trend in relation to the unbiased genes in any of the species studied.

Evolutionary features of the generation-biased genes

The evolutionary dynamics of generation-biased genes was investigated by calculating ratios of pairwise synonymous to non-synonymous substitution rates and comparing these data with gene expression divergence (see the "Methods" section). This analysis, which was applied to all the PSOs that could be examined for each order (Ectocarpales and Laminariales), indicated that gametophyte-biased genes evolve faster (i.e., had higher dN/dS ratios) than unbiased or sporophyte-biased genes in both Ectocarpales species (Fig. 5a; Wilcoxon test, p value < 1e−05) and in *M. pyrifera* (Wilcoxon test p value < 4e−04). The kelp *S. japonica* was the only exception; no significant difference was observed between the rates of evolution of gametophyte-biased and unbiased genes for this species (Fig. 5a).

Sex-biased genes have been shown to exhibit accelerated rates of evolution in *Ectocarpus* sp. [31] and many of the gametophyte-biased genes also exhibit a sex-biased pattern of expression (809 genes or about 20%) because this is the sexual generation of the life cycle. However, when these sex-biased genes were removed from the dN/dS analysis, the gametophyte-biased genes still showed faster evolutionary rates than unbiased or sporophyte-biased genes (Additional file 2: Figure S5A; Wilcoxon test p value = 6.3e−13 for the gametophyte versus unbiased comparison, p value = 2e−09 for the gametophyte versus sporophyte-biased comparison) indicating that the faster evolutionary rates of gametophyte-biased genes were not solely due to the presence of sex-biased genes. Similar results were obtained when we subdivided the generation-biased genes into conserved bias and species-specific bias (i.e., genes that showed generation bias in one species but were unbiased in the other) (Additional file 2: Figure S5B, C). This latter analysis suggested that the faster evolutionary rates of gametophyte-biased genes were not correlated with the degree of conservation of expression across the lineages.

The higher evolutionary rates of gametophyte-biased genes were due to the accumulation of non-synonymous changes (Additional file 2: Figure S6A, B). There was no correlation between dN/dS and fold change of generation-biased gene expression between generations (Spearman's rho = − 0.145). Codon usage bias (CUB), measured as the effective number of codons (ENC), indicated that gametophyte-biased genes had significantly lower codon usage bias (i.e., higher ENC) than sporophyte-biased and unbiased genes in both Ectocarpales and Laminariales (pairwise Wilcoxon test p value < 1e−09) (Additional file 2: Figure S6C).
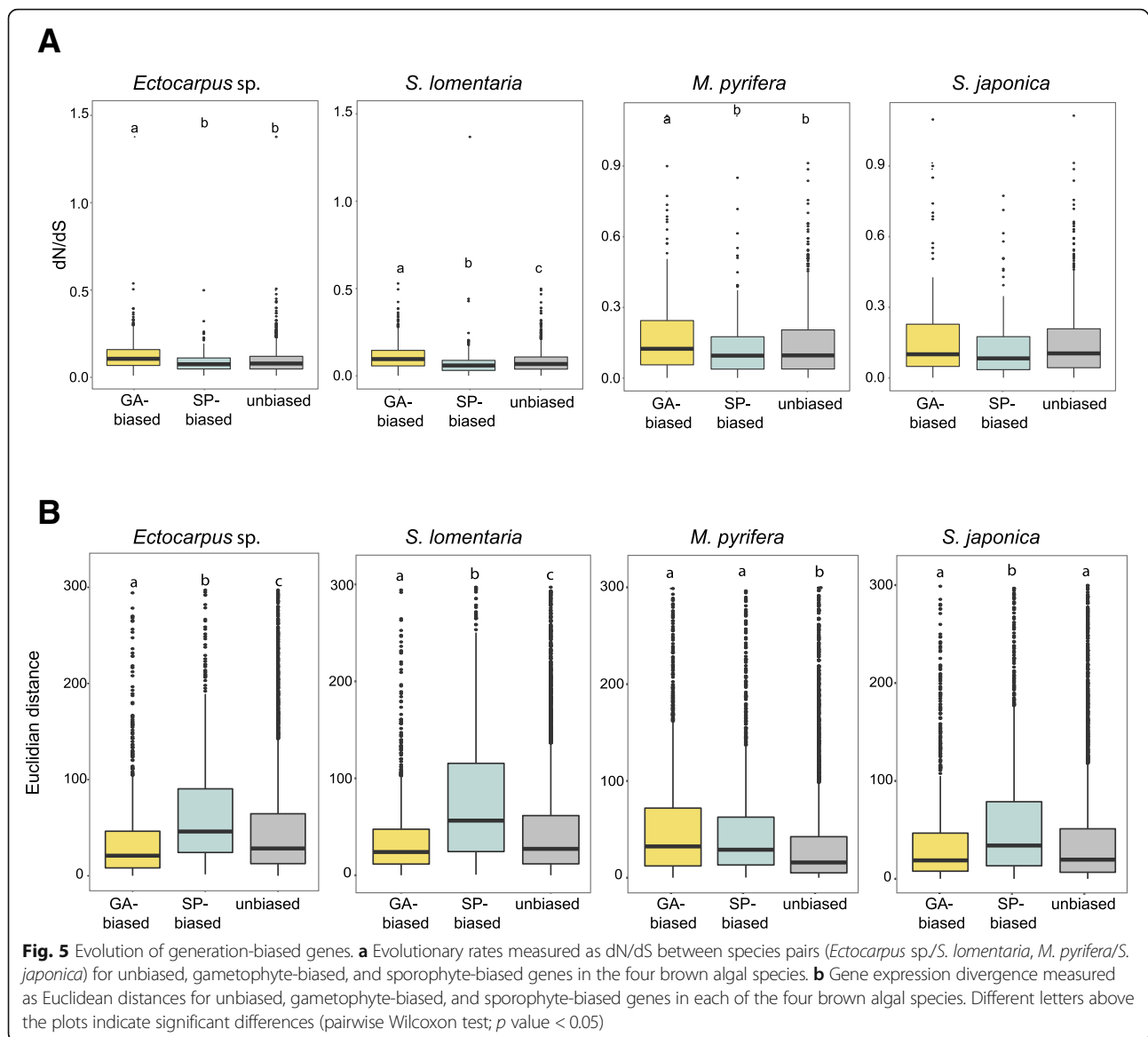
To assess whether increased protein divergence rates were due to increased positive selection or relaxed purifying selection, we performed a maximum likelihood analysis using codeml in PAML4. In addition to the four study species, we searched for orthologs of the ASOs in published transcriptome data for three additional *Ectocarpus* species (*E. fasciculatus*, an unnamed *Ectocarpus* species from New Zealand, and *Ectocarpus siliculosus*) and the recently published genome of *Cladosiphon okamuranus* [28]. The 400 conserved orthologs identified by this analysis included 292 orthologs that exhibited generation bias in at least one of the studied species. For 29 of these comparisons, both pairs of models (M1a-M2a, M7-M8) suggested positive selection and a total of 81 genes were predicted to be evolving under positive selection as indicated by the model M7-M8 alone (Additional file 1: Table S13). Among the 81 genes identified by the M7-M8 model, 64 exhibited generation bias in at least one species. Taken together, our analysis is consistent with the idea that a subset of the generation-biased genes exhibit signatures of positive selection, although the set of generation-biased genes was not significantly enriched in genes that were predicted to be under positive selection (Fisher's exact test, p value = 0.2074).

When patterns of gene expression were considered, measured as Euclidian distances for PSOs within each order, sporophyte-biased genes showed overall significantly larger divergence than unbiased or gametophyte-biased genes with the exception of *M. pyrifera* where both gametophyte- and sporophyte-biased genes showed higher divergence than unbiased genes (Fig. 5b; Wilcoxon test p value < 2e−16). Furthermore, in the Ectocarpales, the expression patterns of gametophyte-biased genes tended to be even more conserved than those of unbiased genes (Fig. 5b).

Taken together, our data suggests that gametophyte- and sporophyte-biased genes have distinct patterns of evolution: gametophyte-biased genes tend to exhibit rapid evolution of their coding sequence whereas sporophyte-biased genes tend to exhibit changes to their patterns of expression.

Discussion

Here, we have used four phylogenetically diverse brown algal species with different levels of generation dimorphism and complexity to investigate genome-wide generation-biased gene expression patterns and to assess

Fig. 5 Evolution of generation-biased genes. **a** Evolutionary rates measured as dN/dS between species pairs (*Ectocarpus* sp./*S. lomentaria*, *M. pyrifera*/*S. japonica*) for unbiased, gametophyte-biased, and sporophyte-biased genes in the four brown algal species. **b** Gene expression divergence measured as Euclidean distances for unbiased, gametophyte-biased, and sporophyte-biased genes in each of the four brown algal species. Different letters above the plots indicate significant differences (pairwise Wilcoxon test; *p* value < 0.05)

the potential role of generation-specific selection in shaping these patterns of gene expression. This study used two high-quality genome assemblies and two draft genome assemblies. Estimations of generation-biased gene expression based on the latter could potentially be biased due to missing genes or gene model fragmentation, and this could, in turn, affect the detection of orphans, orthologs, and duplicated genes. We therefore tested several alternative methods to optimize the assembly and annotation of the draft genomes. The versions used for the study, which employed Masurca for the assembly and PASA for the annotation, were deemed to be of sufficient quality, based on several quality and completeness tests, to minimize any bias. We would like to underline the importance of evaluating draft genome quality when carrying out comparative genomic analyses.

Differential gene expression underlies phenotypic dimorphism between life cycle generations

Between 30 and 36% of the genomes of the brown algae species studied here was differentially regulated during the gametophyte and sporophyte generations of the life cycle. This is substantially more than in the moss *Funaria hygrometrica*, where 24% of the genome is differentially expressed [42], and even exceeds the situation in *Arabidopsis*, where 23–27% of the genome is generation-biased [43]. Comparative analysis between these two land plants showed that the relative proportion of generation-biased genes assigned to the two life cycle generations was lower in the moss than in *Arabidopsis* (proportion of generation-biased genes for each species and generation: *Funaria* gametophyte 10%, *Funaria* sporophyte 13%, *Arabidopsis* gametophyte 0.5–4%, *Arabidopsis*

sporophyte 23–24%) [43], consistent with the lower level of phenotypic dimorphism between generations in the former. Likewise, we found that for the brown algae studied here the relative numbers of generation-specific genes during each generation was broadly correlated with differences in size and complexity between the two generations. Despite this tendency, however, the absolute number of gametophyte-biased genes was relatively high in Laminariales, where the gametophyte is much less complex, morphologically, than the sporophyte. This tendency is difficult to explain but we note that previous analyses using *Ectocarpus* sp. also indicated that the gametophyte developmental program deployed more genes than the sporophyte program [4, 44].

Generation-conflict and generation-biased gene expression

In many brown algal species, free-living gametophytes and sporophytes display extensive morphological and physiological dimorphisms (reviewed in [1, 45, 46]) and this phenotypic diversity reflects in many cases different ecological niche preferences for gametophytes and sporophytes (e.g., [14, 47]). Gametophyte and sporophyte development and function are controlled by a common genome, with a large number of genes carrying out functions during both generations. When there are marked morphological and physiological differences between the two generations, as is the case for most of the species studied here, this can lead to conflict due to genes being submitted to different selection pressures during the different generations of the life cycle. Generation-biased gene expression is one mechanism to reduce inter-generational conflict, allowing gene products to be targeted specifically to one generation (although this does not necessarily mean that every generation-biased gene arose due to generation antagonism). Gene duplication, followed by acquisition of generation bias through neo-functionalization, can play an important role in the resolution of generation conflict, and we found some evidence for this, at least for two of the brown algal species analysed. Note that gene duplication followed by neo-functionalization has also been proposed as one of the mechanisms that allow resolution of sexual antagonism (reviewed in [17]).

Generation-biased genes turned over rapidly during the evolution of the brown algae

Perhaps the most striking result of our analysis is the remarkably limited number of generation-biased genes that were shared by all the four of the studied species, indicating a rapid turnover of life-cycle-biased genes in brown algae. This turnover appears to be due to a combination of two processes: emergence of new genes with strong generation bias and gain/loss of bias for existing, orthologous genes. Phylogenetic stochastic mapping results were consistent with rapid loss and gain of expression bias in orthologous genes.

The ancestor of brown algae is thought to have alternated between multicellular, isomorphic gametophyte and sporophyte generations without a clearly dominant generation [24]. From this morphologically simple ancestor, there was a tendency, in most brown algal lineages, to evolve towards increased complexity of either the gametophyte or the sporophyte generation [24]. Our data indicate that this increase in size and developmental complexity was accompanied by an overall increase in the proportion of the transcriptome that become gametophyte- or sporophyte-biased, depending on the lineage.

Recent analysis of *Ectocarpus* sp. developmental mutants has indicated that the evolution of the sporophyte and gametophyte genetic programs involved both co-option of genetic programs from one generation to the other and generation-specific innovations [48, 49]. We observed that a subset of the expressed genes in the four brown algal species exhibited switching of bias between life cycle generations in the different lineages, in line with the idea that the evolution of the generation-specific developmental programs in the brown algae has, to some extent, involved sharing of genes between generations. However, note that the existence of generation-biased orphan genes indicates that the evolution of brown algal gametophyte and sporophyte developmental programs has also involved generation-specific innovations during the evolution of the sporophyte and gametophyte genetic programs.

The evolutionary origins of the sporophyte and gametophyte developmental programs in land plants have been intensively studied, particularly with regard to the question of whether each generation has independently evolved its own developmental pathways or, alternatively, whether there has been recruitment of developmental programs from one generation to the other during evolution [50–52]. It is currently thought that the developmental networks that implement land plant sporophyte programs were mainly recruited from the gametophyte generation, which was initially the dominant generation [42, 50, 53] although there is also evidence that there have been sporophyte-specific innovations [42, 54]. We suggest however, based on the observations presented here for the brown algae, that it may be an oversimplification to think in terms of one generation gradually recruiting programs from the other generation and that a more extensive sampling of generation-biased gene sets in land plants may reveal a more dynamic situation involving important amounts of both lineage-specific gene evolution and lineage-specific switching of generation-biased expression patterns.

Lipinska *et al. Genome Biology*   (2019) 20:35

Page 13 of 18

Interestingly, despite the marked differences between the generation-biased gene sets of the four studied brown algae, the enriched GO terms associated with genes expressed during the more morphologically complex, long-lived, dominant generation tended to be similar. The predicted functions of both the sporophyte-biased genes of kelps and the gametophyte-biased genes of *S. lomentaria* were enriched in GO terms associated with polysaccharide and cell wall biosynthesis, developmental processes, cell signaling, and cell communication. These conserved, enriched GO terms could reflect developmental and morphological processes common to dominant life cycle generations, such as extended multicellular growth.

Rapid evolution of the coding regions of gametophyte-biased genes

On average, gametophyte-biased genes were found to be evolving significantly more rapidly (higher dN/dS) than sporophyte-biased and unbiased genes in all the species studied except *S. japonica*. This was surprising because purifying selection is expected to be more efficient for genes expressed during the haploid phase of the life cycle due to the absence of masking of recessive and partially recessive mutations [55, 56]. However, accelerated evolution of gametophyte-biased genes has also been previously reported in land plant systems [43, 57], and a number of hypotheses have been put forward to explain this phenomenon. It has been suggested, for example, that gametophyte-biased genes are under relaxed constraint because of lower expression breadth and low level of tissue complexity [42, 57–59]. This hypothesis is unlikely to explain our observations because in *S. lomentaria* the gametophyte generation is the dominant phase of the life cycle and is larger and more complex than the sporophyte. It has also been proposed that strong selection on reproductive traits during gametogenesis (i.e., during the gametophyte generation) may explain the faster rates of evolution of gametophyte-biased genes in plants [57]. However, when sex-biased genes were excluded from the *Ectocarpus* sp. gametophyte-biased gene dataset, the remaining genes still exhibited a significantly higher rate of evolution than those of unbiased or sporophyte-biased rates. Finally, land plant gametophyte-biased gene sets are enriched in young genes [57] and this may affect evolution rate as young genes are known to evolve more rapidly. However, rapid evolution of young genes is unlikely to explain the faster evolutionary rates of brown algal gametophyte-biased genes because the gametophyte- and sporophyte-biased gene sets contained almost exclusively genes that were ancestral to the two orders (Ectocarpales and Laminariales).

We did note, however, that gametophyte-biased genes have overall lower levels of expression than sporophyte-biased genes, and expression levels have been negatively correlated with evolutionary rates [60–62]. Moreover, since gametophyte-biased genes present a higher number of gain/loss of bias events than sporophyte-biased genes, one interesting possibility is that gametophyte-biased genes may be less associated with complex gene interaction networks and therefore be more dispensable and thus under less constraint [63–65]. More information about gene interaction networks will be needed for the brown algae in order to test this hypothesis.

Sporophyte-biased and gametophyte-biased genes exhibit different patterns of evolution

In contrast to the gametophyte-biased genes, sporophyte-biased genes did not exhibit overall accelerated rates of evolution of their coding sequences but they did exhibit significantly higher levels of diversification of expression patterns (measured as Euclidean distance), compared to both unbiased and gametophyte-biased genes. Therefore, while the gametophyte-biased genes exhibited accelerated evolution of their coding regions, the sporophyte-biased genes appeared to have experienced accelerated evolution of their regulatory sequences. Decoupling of protein sequence evolution and expression pattern evolution has been observed in other eukaryotes (e.g., [66, 67] but see [68, 69]), but it is not clear why the mechanisms of evolution should differ between gametophyte-biased and sporophyte-biased genes in these brown algal species. It has been suggested that mutations that change protein sequences and mutations affecting gene regulation play different roles during evolution, with genes involved in physiological traits tending to exhibit the former and genes involved in morphological traits evolving primarily in terms of gene expression [66, 70]. An in-depth functional analysis of brown algae genes using experimental approaches will be crucial to understand if the different modes of evolution of gametophyte- and sporophyte-biased genes are associated with different functions of the gene networks underlying each generation.

Conclusions

This study afforded the first comparative analysis of generation-biased gene expression across several species with complex life cycles to understand the role of generation-specific selection in shaping patterns of gene expression and divergence. Our analyses revealed that an extensive proportion of the genome exhibits generation-biased expression in the brown algae and the relative proportion of genes that are generation-biased is correlated with the degree of phenotypic dimorphism between generations. Life-cycle-biased genes turn over very rapidly during evolution due to a combination of two processes: gain/loss of generation-biased expression by orthologous loci and the emergence, de novo, of genes with generation-biased expression. Our results are consistent with the idea that the

evolution of the genetic program associated with each generation appears to have involved cross-generation recruitment of genes but the existence of generation-biased orphan genes emphasizes an important role for generation-specific developmental innovations in each lineage. Finally, our analysis indicates that the gametophyte and the sporophyte have distinct modes of evolution, with gametophyte-biased genes evolving rapidly predominantly at the level of their sequence and sporophyte-biased genes diverging mostly at the level of their patterns of expression.

Methods

Biological material and generation of genomic and transcriptomic sequence data

The algal strains used, sequencing statistics, and accession numbers are listed in Additional file 1: Table S1 and S3. We used published RNA-seq datasets for gametophytes and sporophytes of the model brown alga *Ectocarpus* sp. [27, 31], *S. japonica* [25, 32], and *M. pyrifera* [27] and for gametophytes of *S. lomentaria* [31] (duplicate samples for the Ectocarpales and triplicate samples for the Laminariales). Duplicate samples of *S. lomentaria* sporophytes (strain Zy2) were derived from a controlled laboratory cross between the Asari6 female gametophyte and the Asari9 male gametophyte. Both gametophytes were field collected. Sporophyte clones were grown in 20 °C 14 h:10 h light:dark conditions, in half-strength Provasoli-enriched seawater [71] which allowed them to be maintained as immature thalli (absence of meiotic structures). For the extraction of RNA from gametophytes and sporophytes of *Ectocarpus* sp., gametophytes and sporophytes of *S. lomentaria*, and gametophytes of both kelps, we used whole thallus, containing all of the cell types described in Additional file 1: Table S2 except spores. For sporophytes of the kelps *M. pyrifera* and *S. japonica*, RNA was extracted from whole fronds. In these tissues, all of the cell types described in Additional file 1: Table S2 are expected to be present, except spores and haptera. Total RNA was extracted using the Qiagen Mini kit (http://www.qiagen.com) as previously described [72]. RNA was sequenced with Illumina HiSeq 2000 paired-end technology with a read length of 125 bp (Fasteris, Switzerland) and is available under the accession numbers detailed in Additional file 1: Table S3.

For the sporophytes of kelps, we used RNA-seq data produced from replicate samples of adult individuals from natural populations (SRA references provided in Additional file 1: Table S3).

Genome assemblies were generated for *S. lomentaria* and *M. pyrifera* using Masurca [26] with default parameters and filtered against the NCBI nucleotide database using Blobtools [73] with a minimum cutoff *e* value of 10e−20 to remove potential bacterial contamination.

Assembly statistics are presented in Additional file 1: Table S1. Sets of reference genes for each species were derived from the published genomes of *Ectocarpus* sp. and *S. japonica* [25, 38] or from draft genome assemblies for *S. lomentaria* and *M. pyrifera* [27] (Additional file 1: Table S1). Gene prediction for *S. lomentaria* and *M. pyrifera* was performed with PASA (https://github.com/PASApipeline/PASApipeline/wiki). Gene prediction took into account *ab initio* prediction with Augustus [74], mapping of RNA-seq data (see Additional file 1: Table S3 for sample details) onto genome assemblies using Stringtie with default settings and minimum junction coverage of 3, and de novo transcriptome assemblies with Trinity [75] using default parameters and normalized mode. To perform the Augustus analysis, the assembled genomes were annotated using BRAKER2 [76] as follows: (i) protein evidence from *Ectocarpus* sp. (downloaded from Orcae http://bioinformatics.psb.u-gent.be/orcae/) was aligned to the genomes using Genome-Threader [77], (ii) RNA-seq reads were mapped to the genomes using TopHat2 [78], (iii) gene predictors AUGUSTUS v3.3 [74] and GeneMark-ET v 4.33 [79] were trained using BRAKER2 and external evidences from (i) and (ii), and finally protein coding genes models were predicted using AUGUSTUS with the trained model obtained in (iii).

The completeness of the annotated genomes was assessed using the BUSCO v.3 [30] eukaryote gene set as the reference (Additional file 1: Table S1). Quality filtering of the raw reads was performed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc), and adapter sequences were trimmed using Trimmomatic (leading and trailing bases with quality below 3 and the first 12 bases were removed, minimum read length 50 bp [80]. *Ectocarpus sp.* and *S. japonica* reads were aligned to the reference genomes [21, 25, 38] using Tophat2 [78]. Protein sequences were predicted for *S. lomentaria* and *M. pyrifera* using Transdecoder (https://github.com/TransDecoder/TransDecoder/wiki) based on the PASA predictions generated for these two species. Gene expression levels were represented as TPMs. Genes with expression values below the fifth percentile of all TPM values calculated per species were considered not to be expressed and were removed from the analysis.

Identification of generation-biased and generation-specific genes

The filtering steps described above yielded a set of expressed genes in the transcriptome that were then classified based on their generation-expression patterns. Genes were considered to be gametophyte-biased or sporophyte-biased if they exhibited at least a twofold difference in expression between generations with a false discovery rate (FDR) of < 0.05. Generation-biased genes

Lipinska *et al. Genome Biology*   (2019) 20:35

Page 15 of 18

were defined as generation-specific when the TPM was below the fifth percentile for one of the generations.

Gene orthology

Orthofinder [81] was used to assess orthologous relationships between the genes of the four studied species (blastp, *e* value < 1e−5). Orthofinder identified a total of 12,891 orthogroups (OGs), of which 4, 043 contained only one gene per species and therefore represented the set of 1:1:1:1 OGs. An additional 2613 OGs had only one member in three of the studied species but no ortholog (i.e., the gene was missing) in the fourth species (1:1:1:0 OGs). We considered that these 1:1:1:0 OGs, which most likely represent single copy ancestral genes that were lost in one of the species, also provided useful information about conservation of generation-biased gene expression because they consisted of members from two different orders (Ectocarpales and Laminariales). We therefore combined the two sets of OGs (1:1:1:1 and 1:1:1:0) to create the "all single orthologs" (ASO) dataset, which was composed of a total of 6656 OGs. Note that the 1:1:1:0 OGs could also represent OGs where one of the genes is missing from one of the genome assemblies, particularly the draft genome assemblies. The ASO dataset was employed to assess conservation of generation-biased gene expression across the four studied species.

For pairwise comparisons within orders, we selected OGs that contained only one member in each of the two species (6644 OGs for the Ectocarpales and 7128 OGs for the Laminariales). We refer to the OGs in these datasets as "pairwise single orthologs" (PSOs). Note that some pairwise orthologs may have been missed in these comparisons due to the use of the two draft genomes for *S. lomentaria* and *M. pyrifera*.

Orphan (or de novo) genes (i. e., taxonomically restricted genes) were defined as genes present in the genome of only one species and having no BLASTp match ($10^{-4}e$ value cutoff) with a range of other stramenopile genome-wide proteomes from public databases (indicating that they are likely to have evolved since the split from the most recent common ancestor): the brown alga *Cladosiphon okamuranus* [28], the eustigmatophyte *Nannochloropsis gaditana* [82], the pelagophyte *Aureococcus anophagefferens* [83], and the diatom *Thalassiosira pseudonana* [84]. Duplicated genes were identified in each species using the dataset generated by Orthofinder. Note that this definition does not exclude genes that are restricted to a single species but have duplicated in that species. Such duplicated orphan genes will be grouped into species-specific OGs.

Prediction of gene function

InterProScan [85] and BLAST2GO [37] were used to assign protein function annotations to genes in all four studied species. Fisher's exact test with a *p* value cutoff of 0.05 was used to detect enrichment of specific GO terms in various groups of generation-biased genes.

The visualization of gene ontology data used for Additional file 2: Figure S2 was generated using Revigo [86].

Evolutionary analysis

To estimate the evolutionary rates (non-synonymous to synonymous substitutions, dN/dS) for generation-biased and unbiased genes, pairwise analyses were carried out on the PSOs for each order (Ectocarpales and Laminariales). Orthologous protein sequences were aligned with Tcoffee (M-Coffee mode [87]), and the alignments curated with Gblocks [88] and then translated back to nucleotide sequence using Pal2Nal [89] or TranslatorX [90] (Additional file 3). Sequences that produced a gapless alignment exceeding 100 bp in length were retained for pairwise dN/dS (ω) analysis using phylogenetic analysis by maximum likelihood (PAML4, CodeML, F3x4 model of codon frequencies, runmode = − 2) [91]. Genes with saturated synonymous substitution values (dS > 2) were excluded from the analysis. Protein alignments, corresponding Gblocks html file and CDS sequences are presented in Additional file 3.

The positive selection analysis was carried out using CodeML (PAML4, F3x4 model of codon frequencies) using additional orthologs of the 1:1 best ortholog set from Orthofinder found in the transcriptomes of three *Ectocarpus* species (*E. fasciculatus*, an unnamed *Ectocarpus* species from New Zealand, and *E. siliculosus*) and in the genome of *Cladosiphon okamuranus* [28]. The analysis was therefore based on data from seven species in total. Protein alignment and curation was performed as described above. Gapless alignments longer than 100 bp containing sequences from at least three species were retained for subsequent analysis. CodeML paired nested site models (M1a, M2a; M7, M8) [91] of sequence evolution were used, and the outputs compared using the likelihood ratio test. The second model in each pair (M2a and M8) is derived from the first by allowing variable dN/dS ratios between sites to be greater than 1, making it possible to detect positive selection at critical amino acid residues.

The effective number of codons (ENC) was calculated using ENCprime [92] with ribosomal genes as background nucleotide composition.

Euclidean distances

Euclidean distances were estimated for all the PSOs for each of the two orders (Ectocarpales and Laminariales)

following the approach of [93]. The following formula was used:

$$\text{EucD} = \sqrt{\sum\nolimits_{j=1}^{k} \left(x_{1j} - x_{2j}\right)^2}$$

where x_{ij} is the expression level of the gene under consideration (TPM) in species i (i.e., species 1 or species 2) during stage j (i.e., gametophyte or sporophyte) and k is the total number of stages (i.e., two, gametophyte and sporophyte). All statistical analysis was performed using RStudio (R version 3.4.2).

Stochastic mapping approach to assess the evolutionary dynamics of generation-biased expression

We conducted an evolutionary analysis of the presence and absence of generation-biased gene expression as a dynamic between gain and loss of phyletic patterns [94]. To estimate the evolutionary dynamics of each event, we tested whether the rates of gain ($0 \rightarrow 1$) and loss ($1 \rightarrow 0$) of bias were equal (ER model) or different (ARD model), and implemented stochastic mapping for each gene using the *phytools* R package [95]. The number of events on each branch only included those transitions that effectively produced a change of state at the start and end of the specified branch. These changes in state (gain or loss of bias) were mapped separately for sporophyte and gametophyte, respectively.

Additional files

Additional file 1: Table S1. Genome and transcriptome assembly statistics for the four studied species. **Table S2.** Phenotypic differences between gametophytes (GA) and sporophytes (SP) of the four studied species. **Table S3.** Strains used in this study and DNA and RNA sequencing data statistics. **Table S4.** Gene expression levels (measured as TPM) in gametophyte and sporophyte generations in each of the four studied species. **Table S5.** Gametophyte- and sporophyte-biased genes identified for each of the four studied species (DEseq2 FC < 2, padj< 0.05, TPM > 5th percentile). **Table S6.** Generation-biased gene expression and morphological complexity of sporophytes and gametophytes across the four studied species. **Table S7.** Numbers of duplicated, single copy and orphan genes among the generation-biased genes in each of the studied brown algal species. **Table S8.** Orthology statistics based on the Orthofinder analysis. **Table S9.** Enriched gene ontology (GO) terms associated with the generation-biased genes (GBGs) identified for each of the studied species (Fisher test *p*-value< 0.05). **Table S10.** GO terms consistently enriched in either gametophyte-biased or sporophyte-biased genes across all four brown algal species. **Table S11.** List of GBGs conserved across all the studied species and their predicted functions. **Table S12.** Structural characteristics of the generation-biased genes identified for each of the studied species. **Table S13.** PAML codeml analysis with the F3X4 substitution model. (ZIP 9175 kb)

Additional file 2: Figure S1. Brown algal species used in this study. **Figure S2.** Visualisation of GO terms associated with generation-biased genes. **Figure S3.** Structural characteristics of unbiased, gametophyte- and sporophyte-biased genes across brown algal species. **Figure S4.** Proportions of single copy versus duplicated genes and numbers of orphan genes in the generation biased and unbiased genes. **Figure S5.** Evolutionary rates for generation-biased genes and sex-biased genes. **Figure S6.** Non-synonymous substitutions, synonymous substitutions and codon usage bias for unbiased, gametophyte- and sporophyte-biased genes in the four studied species. (ZIP 113664 kb)

Additional file 3: Protein alignments, corresponding Gblocks html file and CDS sequences for the lineage conserved gametophyte-biased genes. (ZIP 2109 kb)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Sorbonne Université, UPMC Univ Paris 06, CNRS, Algal Genetics Group, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688 Roscoff, France. ²Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland. ³Bezhin Rosko, 29250 Santec, France. ⁴Department of Biological Sciences, Faculty of Sciences, Hokkaido University, Sapporo 060-0810, Japan. ⁵Laboratoire Ecologie et Biologie des Interactions, Equipe Ecologie Evolution Symbiose, Université de Poitiers, UMR CNRS 7267, Poitiers, France.

References

1. Coelho SM, Peters AF, Charrier B, Roze D, Destombe C, Valero M, et al. Complex life cycles of multicellular eukaryotes: new approaches based on the use of model organisms. Gene. 2007;406:152–70.
2. Valero M, Richerd S, Perrot V, Destombe C. Evolution of alternation of haploid and diploid phases in life cycles. Trends Ecol Evol. 1992;7:25–9.
3. Arun A, Peters NT, Scornet D, Peters AF, Cock JM, Coelho SM. Non-cell autonomous regulation of life cycle transitions in the model brown alga *Ectocarpus*. New Phytol. 2013;197:503–10.
4. Coelho SM, Godfroy O, Arun A, Le Corguillé G, Peters AF, Cock JM. OUROBOROS is a master regulator of the gametophyte to sporophyte life

cycle transition in the brown alga *Ectocarpus*. Proc Natl Acad Sci U S A. 2011;108:11518–23.

5.   Perrot V, Richerd S, Valero M. Transition from haploidy to diploidy. Nature. 1991;351:315–7.

6.   Otto SP, Goldstein DB. Recombination and the evolution of diploidy. Genetics. 1992;131:745–51.

7.   Jenkins CD. Selection and the evolution of genetic life cycles. Genetics. 1993;133:401–10.

8.   Mable BK, Otto SP. The evolution of life cycles with haploid and diploid phases. BioEssays. 1998;20:453–62.

9.   Rescan M, Lenormand T, Roze D. Interactions between genetic and ecological effects on the evolution of life cycles. Am Nat. 2016;187:19–34.

10. Scott MF, Rescan M. Evolution of haploid-diploid life cycles when haploid and diploid fitnesses are not equal. Evol Int J Org Evol. 2017;71:215–26.

11. Paquin CE, Adams J. Relative fitness can decrease in evolving asexual populations of *S. cerevisiae*. Nature. 1983;306:368–70.

12. Hughes O. Ecology and the evolution of biphasic life cycles. Am Nat. 1999; 154:306–20.

13. Destombe C, Godin J, Nocher M, Richerd S, Valero M. Differences in response between haploid and diploid isomorphic phases of *Gracilaria verrucosa* (Rhodophyta: Gigartinales) exposed to artificial environmental conditions; 1993.

14. Couceiro L, Le Gac M, Hunsperger HM, Mauger S, Destombe C, Cock JM, et al. Evolution and maintenance of haploid-diploid life cycles in natural populations: the case of the marine brown alga *Ectocarpus*. Evol Int J Org Evol. 2015;69:1808–22.

15. Immler S, Otto SP. The evolution of sex chromosomes in organisms with separate haploid sexes. Evolution. 2015;69:694–708.

16. Connallon T, Knowles LL. Intergenomic conflict revealed by patterns of sex-biased gene expression. Trends Genet TIG. 2005;21:495–9.

17. Parsch J, Ellegren H. The evolutionary causes and consequences of sex-biased gene expression. Nat Rev Genet. 2013;14:83–7.

18. Connallon T, Clark AG. The resolution of sexual antagonism by gene duplication. Genetics. 2011;187:919–37.

19. Harrison PW, Wright AE, Zimmer F, Dean R, Montgomery SH, Pointer MA, et al. Sexual selection drives evolution and rapid turnover of male gene expression. Proc Natl Acad Sci U S A. 2015;112:4393–8.

20. Safran RJ, Vortman Y, Jenkins BR, Hubbard JK, Wilkins MR, Bradley RJ, et al. The maintenance of phenotypic divergence through sexual selection: an experimental study in barn swallows *Hirundo rustica*. Evol Int J Org Evol. 2016;70:2074–84.

21. Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, et al. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. Nature. 2010;465:617–21.

22. Luthringer R, Cormier A, Peters AF, Cock JM, Coelho SM. Sexual dimorphism in the brown algae. Perspectives in Phycology. 2015;1:11–25.

23. Cock JM, Godfroy O, Macaisne N, Peters AF, Coelho SM. Evolution and regulation of complex life cycles: a brown algal perspective. Curr Opin Plant Biol. 2014;17:1–6.

24. Silberfeld T, Leigh JW, Verbruggen H, Cruaud C, de Reviers B, Rousseau F. 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. 2010;56:659–74.

25. Ye N, Zhang X, Miao M, Fan X, Zheng Y, Xu D, et al. *Saccharina* genomes provide novel insight into kelp biology. Nat Commun. 2015;6:6986.

26. Zimin AV, Marcais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome assembler. Bioinforma Oxf Engl. 2013;29:2669–77.

27. Lipinska AP, Toda NRT, Heesch S, Peters AF, Cock JM, Coelho SM. Multiple gene movements into and out of haploid sex chromosomes. Genome Biol. 2017;18:104.

28. Nishitsuji K, Arimoto A, Iwai K, Sudo Y, Hisata K, Fujie M, et al. A draft genome of the brown alga, *Cladosiphon okamuranus*, S-strain: a platform for future studies of 'mozuku' biology. DNA Res Int J Rapid Publ Rep Genes Genomes. 2016;23:561–70.

29. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RKJ, Hannick LI, et al. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 2003;31:5654–66.

30. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol. 2018;35:543–8.

31. Lipinska A, Cormier A, Luthringer R, Peters AF, Corre E, Gachon CMM, et al. Sexual dimorphism and the evolution of sex-biased gene expression in the brown alga *Ectocarpus*. Mol Biol Evol. 2015;32:1581–97.

32. Teng L, Han W, Fan X, Xu D, Zhang X, Dittami SM, et al. Evolution and expansion of the prokaryote-like lipoxygenase family in the brown alga *Saccharina japonica*. Front Plant Sci. 2017;8:2018.

33. Salavarría E, Paul S, Gil-Kodaka P, Villena GK. First global transcriptome analysis of brown algae Macrocystis integrifolia (Phaeophyceae) under marine intertidal conditions. 3 Biotech. 2018;8:185.

34. Schurch NJ, Schofield P, Gierlinski M, Cole C, Sherstnev A, Singh V, et al. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA N Y N. 2016;22:839–51.

35. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16:157.

36. Cohen O, Pupko T. Inference of gain and loss events from phyletic patterns using stochastic mapping and maximum parsimony--a simulation study. Genome Biol Evol. 2011;3:1265–75.

37. Conesa A, Götz S. Blast2GO: a comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics. 2008;2008:619832.

38. Cormier A, Avia K, Sterck L, Derrien T, Wucher V, Andres G, et al. Re-annotation, improved large-scale assembly and establishment of a catalogue of noncoding loci for the genome of the model brown alga *Ectocarpus*. New Phytol. 2017;214:219–32.

39. Chan SW, Fowler KJ, Choo KHA, Kalitsis P. Spef1, a conserved novel testis protein found in mouse sperm flagella. Gene. 2005;353:189–99.

40. Lehti MS, Zhang F-P, Kotaja N, Sironen A. SPEF2 functions in microtubule-mediated transport in elongating spermatids to ensure proper male germ cell differentiation. Dev Camb Engl. 2017;144:2683–93.

41. Sironen A, Thomsen B, Andersson M, Ahola V, Vilkki J. An intronic insertion in KPL2 results in aberrant splicing and causes the immotile short-tail sperm defect in the pig. Proc Natl Acad Sci U S A. 2006;103:5006–11.

42. Szovenyi P, Rensing SA, Lang D, Wray GA, Shaw AJ. Generation-biased gene expression in a bryophyte model system. Mol Biol Evol. 2011;28: 803–12.

43. Szovenyi P, Ricca M, Hock Z, Shaw JA, Shimizu KK, Wagner A. Selection is no more efficient in haploid than in diploid life stages of an angiosperm and a moss. Mol Biol Evol. 2013;30:1929–39.

44. Arun A, Coelho SM, Peters AF, Bourdareau S, Pérès L, Scornet D, et al. Convergent recruitment of TALE homeodomain life cycle regulators to direct sporophyte development in land plants and brown algae. eLife. https://doi.org/10.7554/eLife.43101.

45. Clayton MN. Evolution and life histories of brown algae. Bot. 2009;31:379.

46. Thornber CS. Functional properties of the isomorphic biphasic algal life cycle. Integr Comp Biol. 2006;46:605–14.

47. Bessho K, Iwasa Y. Optimal seasonal schedules and the relative dominance of heteromorphic and isomorphic life cycles in macroalgae. J Theor Biol. 2010;267:201–12.

48. Godfroy O, Uji T, Nagasato C, Lipinska AP, Scornet D, Peters AF, et al. DISTAG/ TBCCd1 is required for basal cell fate determination in *Ectocarpus*. Plant Cell. 2017; 29(12):3102–3122. https://doi.org/10.1105/tpc.17.00440.

49. Macaisne N, Liu F, Scornet D, Peters AF, Lipinska A, Perrineau M-M, et al. The Ectocarpus IMMEDIATE UPRIGHT gene encodes a member of a novel family of cysteine-rich proteins with an unusual distribution across the eukaryotes. Dev Camb Engl. 2017;144:409–18.

50. Dolan L. Body building on land — morphological evolution of land plants. Curr Opin Plant Biol. 2009;12:4–8.

51. Pires ND, Dolan L. Morphological evolution in land plants: new designs with old genes. Philos Trans R Soc B Biol Sci. 2012;367:508–18.

52. Shaw AJ, Szovenyi P, Shaw B. Bryophyte diversity and evolution: windows into the early evolution of land plants. Am J Bot. 2011;98:352–69.

53. Niklas KJ, Kutschera U. The evolution of the land plant life cycle. New Phytol. 2010;185:27–41.

54. Sano R, Juarez CM, Hass B, Sakakibara K, Ito M, Banks JA, et al. KNOX homeobox genes potentially have similar function in both diploid unicellular and multicellular meristems, but not in haploid meristems. Evol Dev. 2005;7:69–78.

55. Gerstein AC, Cleathero LA, Mandegar MA, Otto SP. Haploids adapt faster than diploids across a range of environments. J Evol Biol. 2011;24:531–40.

56. Orr HA, Otto SP. Does diploidy increase the rate of adaptation? Genetics. 1994;136:1475–80.

57. Gossmann TI, Saleh D, Schmid MW, Spence MA, Schmid KJ. Transcriptomes of plant gametophytes have a higher proportion of rapidly evolving and young genes than sporophytes. Mol Biol Evol. 2016;33:1669–78.

58. Duret L, Mouchiroud D. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. Mol Biol Evol. 2000;17:68–74.

59. Slotte T, Bataillon T, Hansen TT, St Onge K, Wright SI, Schierup MH. Genomic determinants of protein evolution and polymorphism in *Arabidopsis*. Genome Biol Evol. 2011;3:1210–9.

60. Cherry JL. Expression level, evolutionary rate, and the cost of expression. Genome Biol Evol. 2010;2:757–69.

61. Krylov DM, Wolf YI, Rogozin IB, Koonin EV. Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. Genome Res. 2003;13:2229–35.

62. Pal C, Papp B, Hurst LD. Highly expressed genes in yeast evolve slowly. Genetics. 2001;158:927–31.

63. Alvarez-Ponce D, Fares MA. Evolutionary rate and duplicability in the Arabidopsis thaliana protein-protein interaction network. Genome Biol Evol. 2012;4:1263–74.

64. Alvarez-Ponce D, Feyertag F, Chakraborty S. Position matters: network centrality considerably impacts rates of protein evolution in the human protein-protein interaction network. Genome Biol Evol. 2017;9:1742–56.

65. Hirsh AE, Fraser HB. Protein dispensability and rate of evolution. Nature. 2001;411:1046–9.

66. Liao B-Y, Weng M-P, Zhang J. Contrasting genetic paths to morphological and physiological evolution. Proc Natl Acad Sci U S A. 2010;107:7353–8.

67. Tirosh I, Barkai N. Evolution of gene sequence and gene expression are not correlated in yeast. Trends Genet. 2008;24:109–13.

68. Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, et al. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Science. 2005;309:1850–4.

69. Nuzhdin SV, Wayne ML, Harmon KL, McIntyre LM. Common pattern of evolution of gene expression level and protein sequence in Drosophila. Mol Biol Evol. 2004;21:1308–17.

70. Wray GA. The evolutionary significance of cis-regulatory mutations. Nat Rev Genet. 2007;8:206–16.

71. Coelho SM, Scornet D, Rousvoal S, Peters NT, Dartevelle L, Peters AF, et al. How to cultivate Ectocarpus. Cold Spring Harb Protoc. 2012;2012:258–61.

72. Coelho SM, Scornet D, Rousvoal S, Peters NT, Dartevelle L, Peters AF, et al. Ectocarpus: a model organism for the brown algae. Cold Spring Harb Protoc. 2012;2012:193–8.

73. Laetsch D, Blaxter M. BlobTools: interrogation of genome assemblies [version 1; referees: 2 approved with reservations]. F1000Research. 2017;6:1287.

74. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinforma Oxf Engl. 2008;24:637–44.

75. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–52.

76. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinforma Oxf Engl. 2016;32:767–9.

77. Gremme G, Brendel V, Sparks ME, Kurtz S. Engineering a software tool for gene structure prediction in higher organisms. Inf Softw Technol. 2005;47:965–78.

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

79. Lomsadze A, Burns PD, Borodovsky M. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. Nucleic Acids Res. 2014;42:e119.

80. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinforma Oxf Engl. 2014;30:2114–20.

81. Li L, Stoeckert CJJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2003;13:2178–89.

82. Radakovits R, Jinkerson RE, Fuerstenberg SI, Tae H, Settlage RE, Boore JL, et al. Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropis gaditana*. Nat Commun. 2012;3:686.

83. Gobler CJ, Berry DL, Dyhrman ST, Wilhelm SW, Salamov A, Lobanov AV, et al. Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. Proc Natl Acad Sci U S A. 2011;108:4352–7.

84. Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, et al. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science. 2004;306:79–86.

85. Zdobnov EM, Apweiler R. InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinforma Oxf Engl. 2001;17:847–8.

86. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One. 2011;6:e21800.

87. Notredame C, Higgins DG, Heringa J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol. 2000;302:205–17.

88. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000;17:540–52.

89. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 2006;34:W609–12.

90. Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 2010;38:W7–13.

91. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–91.

92. Satapathy SS, Sahoo AK, Ray SK, Ghosh TC. Codon degeneracy and amino acid abundance influence the measures of codon usage bias: improved Nc (Nc) and ENCprime (N(')c) measures. Genes Cells Devoted Mol Cell Mech. 2017;22:277–83.

93. Pereira V, Waxman D, Eyre-Walker A. A problem with the correlation coefficient as a measure of gene expression divergence. Genetics. 2009;183:1597–600.

94. Cohen O, Ashkenazy H, Belinky F, Huchon D, Pupko T. GLOOME: gain loss mapping engine. Bioinformatics. 2010;26:2914–5.

95. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol. 2012;3:217–23.