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

Simon M. Dittami[1*], Erwan Corre[2], Loraine Brillet-Guéguen[1,2], Noé Pontoizeau[1,2], Meziane Aite[3], Komlan Avia[1], Christophe Caron[2], Chung Hyun Cho[4], Jonas Collén[1], Alexandre Cormier[1], Ludovic Delage[1], Sylvie Doubleau[5], Clémence Frioux[3], Angélique Gobet[1], Irene González-Navarrete[6], Agnès Groisillier[1], Cécile Hervé[1], Didier Jollivet[7], Hetty KleinJan[1], Catherine Leblanc[1], Agnieszka P. Lipinska[1], Xi Liu[2], Dominique Marie[7], Gabriel V. Markov[1], André E. Minoche[6,8], Misharl Monsoor[2], Pierre Pericard[2], Marie-Mathilde Perrineau[1], Akira F. Peters[9], Anne Siegel[3], Amandine Siméon[1], Camille Trottier[3], Hwan Su Yoon[4], Heinz Himmelbauer[6,8,10], Catherine Boyen[1], Thierry Tonon[1,11]

[1] Sorbonne Université, CNRS, Integrative Biology of Marine Models (LBI2M), Station Biologique de Roscoff, 29680 Roscoff, France

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

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

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

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

[6] Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona, 08003 Spain

[7] Sorbonne Université, CNRS, Adaptation and Diversity in the Marine Environment (ADME), Station Biologique de Roscoff (SBR), 29680 Roscoff, France

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

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

[10] Department of Biotechnology, University of Natural Resources and Life Sciences (BOKU), Vienna, 1190 Vienna, Austria

[11] Centre for Novel Agricultural Products, Department of Biology, University of York, Heslington, York, YO10 5DD, United Kingdom.

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

Abstract

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

Introduction

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

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

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

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

Results

Sequencing and assembly of the *E. subulatus* genome

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

Gene prediction and annotation

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

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

Repeated elements

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

Organellar genomes

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

Global comparison of predicted proteomes

GO-based comparisons

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

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

Domain-based comparisons

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

Metabolic network-based comparisons

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

Genes under positive selection

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

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

Manual examination of lineage-specific and of expanded genes and gene families

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

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

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

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

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

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

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

Targeted manual annotation of specific pathways

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

Cell wall metabolism

Cell walls are key components of both plants and algae and, as a first barrier to the surrounding environment, important for many processes including development and the acclimation to environmental changes. Synthesis and degradation of cell wall oligo- and polysaccharides is facilitated by carbohydrate-active enzymes (CAZymes) (http://www.cazy.org/; Cantarel et al. 2009). These comprise several families including glycoside hydrolases (GHs) and polysaccharide lyases (PLs), both involved in the cleavage of glycosidic linkages, glycosyltransferases (GTs), which create glycosidic linkages, and additional enzymes such as the carbohydrate esterases (CEs) which remove methyl or acetyl groups from substituted polysaccharides.

The genome of the brown alga *E. subulatus* encodes 37 GHs (belonging to 17 GH families), 94 GTs (belonging to 28 GT families), nine sulfatases (family S1-2), and 13 sulfotransferases, but lacks genes homologous to known PLs and CEs (Figure 5). In particular, the consistent lack of known alginate lyases and cellulases in the *E. subulatus* and the other brown algal genomes suggests that other, yet unknown genes, may be responsible for cell wall modifications during development. Overall, the gene content of *E. subulatus* is similar to *Ectocarpus* sp. Ec32 and *S. japonica* in terms of the number of CAZY families, but slightly lower in terms of absolute gene number (Cock, *et al.* 2010; Ye *et al.* 2015; Figure 5). Especially *S. japonica* features an expansion of certain CAZY families probably related to the establishment of more complex tissues in this kelp (*i.e.* 82 GHs belonging to 17 GH families, 131 GTs belonging to 31 GT families).

E. subulatus is frequently found in brackish- and even freshwater environments (West and Kraft, 1996) where its cell wall exhibits little or no sulfation (Torode *et al.*, 2015). Hence, we also assessed

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

Central and storage carbohydrate metabolism

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

Sterol metabolism

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

Algal defense: metabolism of phenolics and halogens

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

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

Transporters

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

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

Abiotic stress-related genes

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

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

Discussion

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

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

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

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

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

Few classical stress response genes but no transporters involved in adaptation

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

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

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

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

Genes related to cell-cell communication are under positive selection

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

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

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

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

Conclusion and future work

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

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

Materials and Methods

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Acknowledgements

We would like to thank Philippe Potin, Mark Cock, Susanna Coelho, Florian Maumus, and Olivier Panaud for helpful discussions, as well as Gwendoline Andres for help setting up the Jbrowse instance. This work was funded partially by ANR project IDEALG (ANR-10-BTBR-04) “Investissements d’Avenir, Biotechnologies-Bioressources”, the European Union’s Horizon 2020 research and innovation Programme under the Marie Sklodowska-Curie grant agreement number 624575 (ALFF), and the CNRS Momentum call. Sequencing was performed at the Genomics Unit of the Centre for Genomic Regulation (CRG), Barcelona, Spain.

References

Abascal, F., Zardoya, R. and Telford, M.J. (2010) TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.*, **38**, W7–W13. Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkq291.

Aite, M., Chevallier, M., Frioux, C., et al. (2018) Traceability, reproducibility and wiki-exploration for “à-la-carte” reconstructions of genome-scale metabolic models. *PLoS Comput. Biol.*, **in press**.

Amtmann, A. (2009) Learning from evolution: *Thellungiella* generates new knowledge on essential and critical components of abiotic stress tolerance in plants. *Mol. Plant*, **2**, 3–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19529830.

Archibald, J.M. (2009) The puzzle of plastid evolution. *Curr. Biol.*, **19**, R81-8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19174147.

Avia, K., Coelho, S.M., Montecinos, G.J., et al. (2017) High-density genetic map and identification of QTLs for responses to temperature and salinity stresses in the model brown alga *Ectocarpus*. *Sci. Rep.*, **7**, 43241. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28256542.

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

765 **Davison, I.R. and Pearson, G.A.** (1996) Stress tolerance in intertidal seaweeds. *J. Phycol.*, **32**, 197–
766 211.
767 **Delage, L., Leblanc, C., Nyvall Collén, P., Gschloessl, B., Oudot, M.-P., Sterck, L., Poulain, J., Aury,**
768 **J.-M. and Cock, J.M.** (2011) In silico survey of the mitochondrial protein uptake and
769 maturation systems in the brown alga *Ectocarpus siliculosus* E. Newbigin, ed. *PLoS One*, **6**,
770 e19540. Available at: http://dx.plos.org/10.1371/journal.pone.0019540.
771 **Delaroque, N., Müller, D.G., Bothe, G., Pohl, T., Knippers, R. and Boland, W.** (2001) The complete
772 DNA sequence of the *Ectocarpus siliculosus* virus EsV-1 genome. *Virology*, **287**, 112–132.
773 **Desmond, E. and Gribaldo, S.** (2009) Phylogenomics of sterol synthesis: insights into the origin,
774 evolution, and diversity of a key eukaryotic feature. *Genome Biol. Evol.*, **1**, 364–81. Available
775 at: http://www.ncbi.nlm.nih.gov/pubmed/20333205.
776 **Dittami, S.M. and Corre, E.** (2017) Detection of bacterial contaminants and hybrid sequences in the
777 genome of the kelp *Saccharina japonica* using Taxoblast. *PeerJ*, **5**, e4073. Available at:
778 https://peerj.com/articles/4073.
779 **Dittami, S.M., Duboscq-Bidot, L., Perennou, M., Gobet, A., Corre, E., Boyen, C. and Tonon, T.**
780 (2016) Host–microbe interactions as a driver of acclimation to salinity gradients in brown algal
781 cultures. *ISME J.*, **10**, 51–63. Available at:
782 http://www.nature.com/doifinder/10.1038/ismej.2015.104.
783 **Dittami, S.M., Eveillard, D. and Tonon, T.** (2014) A metabolic approach to study algal-bacterial
784 interactions in changing environments. *Mol. Ecol.*, **23**, 1656–60. Available at:
785 http://www.ncbi.nlm.nih.gov/pubmed/24447216.
786 **Dittami, S.M., Gravot, A., Goulitquer, S., Rousvoal, S., Peters, A.F., Bouchereau, A., Boyen, C. and**
787 **Tonon, T.** (2012) Towards deciphering dynamic changes and evolutionary mechanisms
788 involved in the adaptation to low salinities in *Ectocarpus* (brown algae). *Plant J.*, **71**, 366–377.
789 Available at: http://doi.wiley.com/10.1111/j.1365-313X.2012.04982.x.
790 **Dittami, S.M., Heesch, S., Olsen, J.L. and Collén, J.** (2017) Transitions between marine and
791 freshwater environments provide new clues about the origins of multicellular plants and algae
792 O. DeClerck, ed. *J. Phycol.*, **53**, 731–745. Available at:
793 http://doi.wiley.com/10.1111/jpy.12547.
794 **Dittami, S.M., Michel, G., Collén, J., Boyen, C. and Tonon, T.** (2010) Chlorophyll-binding proteins
795 revisited--a multigenic family of light-harvesting and stress proteins from a brown algal
796 perspective. *BMC Evol. Biol.*, **10**, 365. Available at:
797 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3008699&tool=pmcentrez&rend
798 ertype=abstract.
799 **Dittami, S.M., Proux, C., Rousvoal, S., Peters, A.F., Cock, J.M., Coppee, J.-Y., Boyen, C. and Tonon,**
800 **T.** (2011) Microarray estimation of genomic inter-strain variability in the genus *Ectocarpus*
801 (Phaeophyceae). *BMC Mol. Biol.*, **12**, 2. Available at:
802 http://www.ncbi.nlm.nih.gov/pubmed/21226968.
803 **Dittami, S.M., Scornet, D., Petit, J.-L., et al.** (2009) Global expression analysis of the brown alga
804 *Ectocarpus siliculosus* (Phaeophyceae) reveals large-scale reprogramming of the
805 transcriptome in response to abiotic stress. *Genome Biol.*, **10**, R66. Available at:
806 http://genomebiology.com/2009/10/6/R66.
807 **Dittami, S.M. and Tonon, T.** (2012) Genomes of extremophile crucifers: new platforms for
808 comparative genomics and beyond. *Genome Biol.*, **13**, 166. Available at:
809 http://www.ncbi.nlm.nih.gov/pubmed/22897896.
810 **Dong, H.-P., Dong, Y.-L., Cui, L., Balamurugan, S., Gao, J., Lu, S.-H. and Jiang, T.** (2016) High light
811 stress triggers distinct proteomic responses in the marine diatom *Thalassiosira pseudonana*.
812 *BMC Genomics*, **17**, 994. Available at:

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

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

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

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

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

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

Foissac, S., Gouzy, J., Rombauts, S., Mathe, C., Amselem, J., Sterck, L., Peer, Y. V. de, Rouze, P. and Schiex, T. (2008) Genome annotation in plants and fungi: EuGene as a model platform. *Curr. Bioinform.*, **3**, 11. Available at: http://www.ingentaconnect.com/content/ben/cbio/2008/00000003/00000002/art00003?token=004f12e9145a666f3a7b6c7a407b425b6b414c7d666749264f655d375c6b68763050211ce9d7c2a.

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

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

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

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

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

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

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

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

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

Heesch, S., Cho, G.Y., Peters, A.F., 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. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20456050.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Oliveira Dal'Molin, C.G. de, Quek, L.-E., Palfreyman, R.W., Brumbley, S.M. and Nielsen, L.K. (2010) AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. *Plant Physiol.*, **152**, 579–89. Available at: http://www.plantphysiol.org/content/152/2/579.full.

Pavia, H., Cervin, G., Lindgren, A. and Åberg, P. (1997) Effects of UV-B radiation and simulated herbivory on phlorotannins in the brown alga *Ascophyllum nodosum*. *Mar. Ecol. Prog. Ser.*, **157**, 139–146. Available at: http://www.int-res.com/abstracts/meps/v157/p139-146/.

Peers, G., Truong, T.B., Ostendorf, E., Busch, A., Elrad, D., Grossman, A.R., Hippler, M. and Niyogi, K.K. (2009) An ancient light-harvesting protein is critical for the regulation of algal photosynthesis. *Nature*, **462**, 518–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19940928.

Peters, A.F., Coucerio, L., Tsiamis, K., Küpper, F.C. and Valero, M. (2015) Barcoding of cryptic stages of marine brown algae isolated from incubated substratum reveals high diversity. *Cryptogam. Algol.*, **36**, 3–29.

Peters, A.F., Marie, D., Scornet, D., Kloareg, B. and Cock, J.M. (2004) Proposal of *Ectocarpus siliculosus* (Ectocarpales, Phaeophyceae) as a model organism for brown algal genetics and genomics. *J. Phycol.*, **40**, 1079–1088.

Popper, Z.A., Michel, G., Hervé, C., Domozych, D.S., Willats, W.G.T., Tuohy, M.G., Kloareg, B. and Stengel, D.B. (2011) Evolution and diversity of plant cell walls: from algae to flowering plants. *Annu. Rev. Plant Biol.*, **62**, 567–90. Available at: http://www.annualreviews.org/doi/abs/10.1146/annurev-arplant-042110-103809.

Prigent, S., Collet, G., Dittami, S.M., et al. (2014) The genome-scale metabolic network of Ectocarpus siliculosus (EctoGEM): a resource to study brown algal physiology and beyond. *Plant J.*, **80**, 367–381. Available at: http://doi.wiley.com/10.1111/tpj.12627.

Prigent, S., Frioux, C., Dittami, S.M., et al. (2017) Meneco, a Topology-Based Gap-Filling Tool Applicable to Degraded Genome-Wide Metabolic Networks C. Kaleta, ed. *PLOS Comput. Biol.*, **13**, e1005276. Available at: http://dx.plos.org/10.1371/journal.pcbi.1005276.

Rao, A.Q., ud Din, S., Akhtar, S., et al. (2016) Genomics of salinity tolerance in plants. In I. Y. Abdurakhmonov, ed. *Plant Genomics*. InTech, pp. 273–299. Available at: http://www.intechopen.com/books/plant-genomics/genomics-of-salinity-tolerance-in-plants.

Read, S.M., Currie, G. and Bacic, A. (1996) Analysis of the structural heterogeneity of laminarin by electrospray-ionisation-mass spectrometry. *Carbohydr. Res.*, **281**, 187–201.

Ritter, A., Dittami, S.M., Goulitquer, S., Correa, J.A., Boyen, C., Potin, P. and Tonon, T. (2014) Transcriptomic and metabolomic analysis of copper stress acclimation in *Ectocarpus siliculosus* highlights signaling and tolerance mechanisms in brown algae. *BMC Plant Biol.*, **14**, 116. Available at: http://www.biomedcentral.com/1471-2229/14/116.

Rival, A., Ilbert, P., Labeyrie, A., et al. (2013) Variations in genomic DNA methylation during the long-term in vitro proliferation of oil palm embryogenic suspension cultures. *Plant Cell Rep.*, **32**, 359–368. Available at: http://link.springer.com/10.1007/s00299-012-1369-y.

Roeder, V., Collén, J., Rousvoal, S., Corre, E., Leblanc, C. and Boyen, C. (2005) Identification of stress gene transcripts in *Laminaria digitata* (Phaeophyceae) protoplast cultures by expressed sequence tag analysis. *J. Phycol.*, **41**, 1227–1235.

Saier, M.H., Reddy, V.S., Tsu, B. V, Ahmed, M.S., Li, C. and Moreno-Hagelsieb, G. (2016) The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Res.*, **44**, D372-9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26546518.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V. and Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210–3212. Available at:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Figures

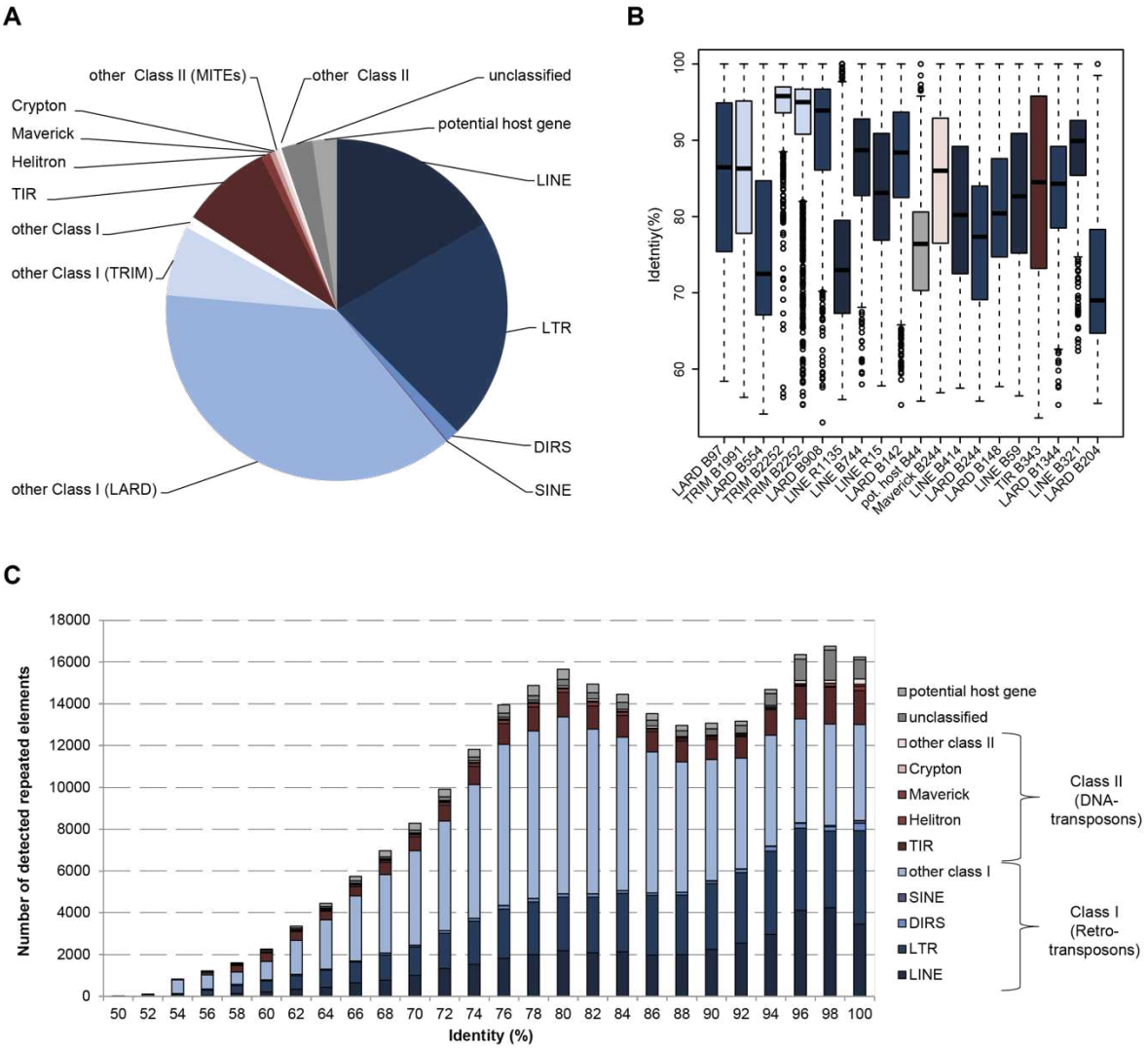


Figure 1: Repeated elements identified within the genome of *E. subulatus*. A) Different superfamilies detected; B) Boxplot of sequence identity levels for the 20 most abundant transposon families; C) Distribution of sequence identities in the superfamilies.

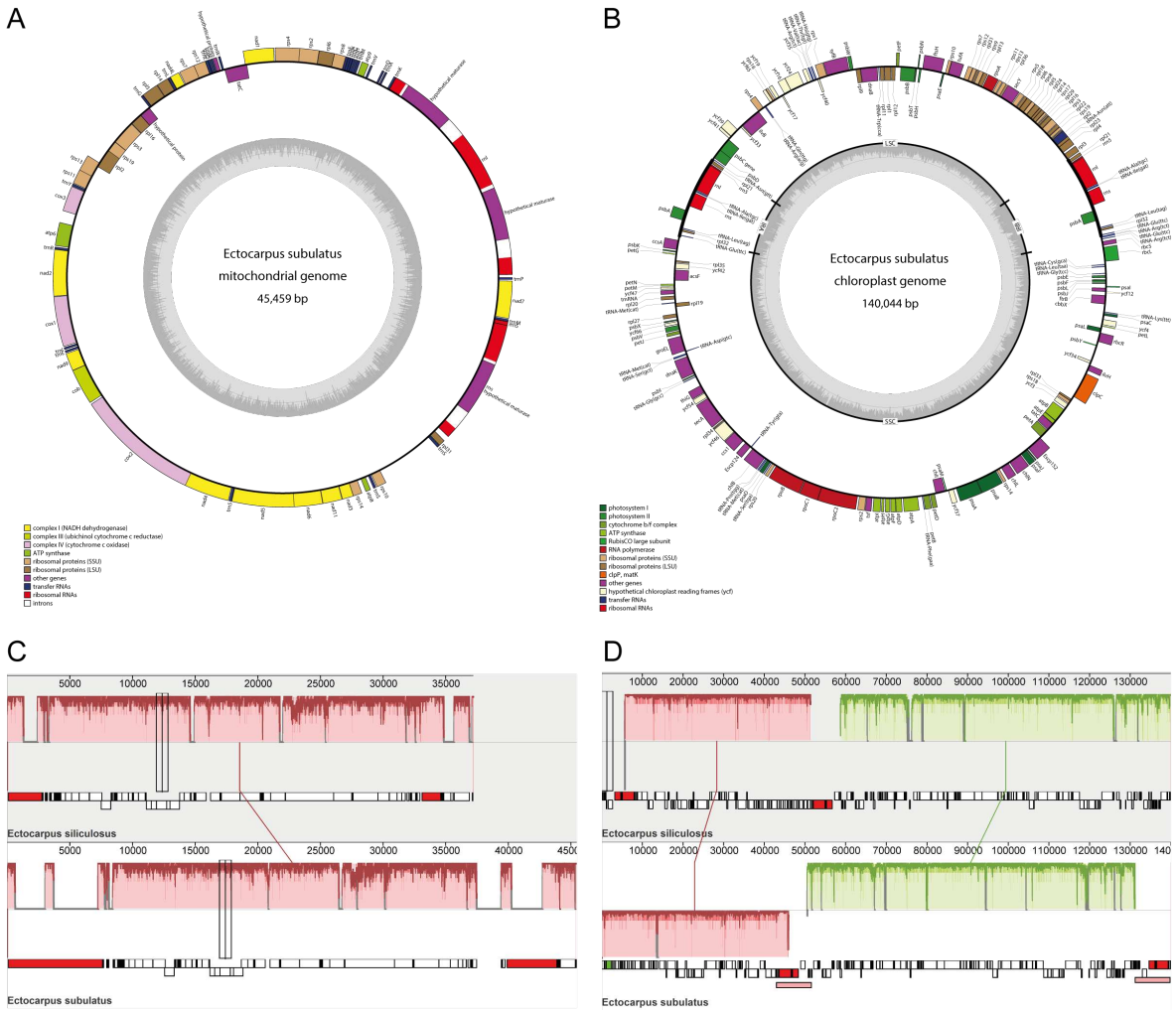


Figure 2: Mitochondrial and plastidial genomes of *E. subulatus* Bft15b visualized using OrganellarGenomeDRAW (Lohse *et al.*, 2013) (panels A and B), and aligned with the *Ectocarpus* sp. Ec32 organelles using Mauve (panels C and D). In panels C and D blocks of the same color correspond to orthologous sequences.

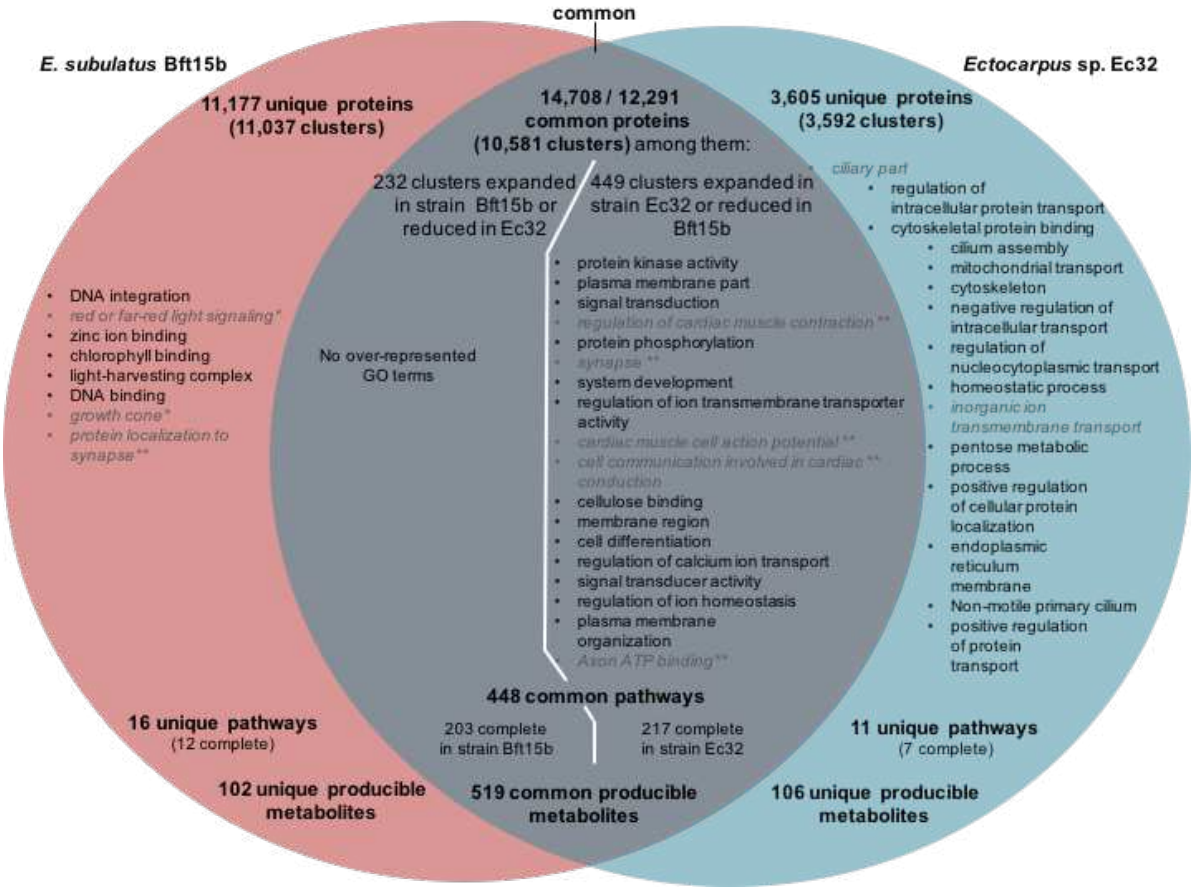
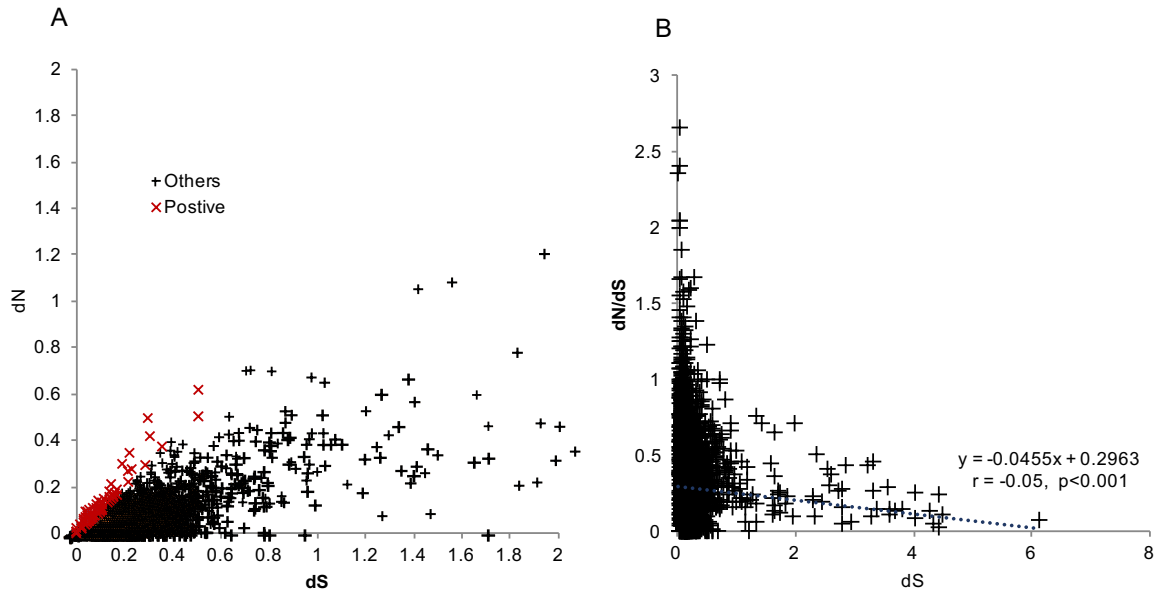


Figure 3: Comparison of gene content and metabolic capacities of *E. subulatus* and *Ectocarpus* sp. Ec32. The top part of the Venn diagram displays the number of predicted proteins and protein clusters unique and common to both genomes in the OrthoFinder analysis. The middle part shows GO annotations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



1099

1100 **Figure 4**: Rate of non-synonymous (dN) to synonymous (dS) mutations in 7,147 orthologous gene
1101 pairs with one copy in each examined species of *Ectocarpus*. Gene pairs with dN/dS > 1 are
1102 considered to be under positive selection and displayed in red panel A. The resulting ratio was plotted
1103 against the rate of synonymous mutations (speed of evolution) and dotted line corresponds to a
1104 linear regression (panel B).

GENES:	*E. sp.*	*E. su*	*S. ja*
Glycosyltransferases			
GT1	1	1	3
GT2	11	9	24
GT4	13	14	13
GT7	1	1	1
GT8	3	4	5
GT10	1	1	2
GT13	2	2	1
GT14	5	5	1
GT15	1	0	0
GT18	0	1	0
GT20	6	8	4
GT22	3	3	2
GT23	2	2	17
GT24	1	2	0
GT25	3	0	1
GT27	0	0	1
GT28	0	3	4
GT31	3*	5	4
GT33	1	1	0
GT34	1	1	1
GT41	3	3	4
GT47	8	5	14
GT48	3	4	2
GT49	3	2	1
GT50	1	0	1
Halogen metabolism			
vBPO	1	1	17
vIPO	0	0	59
vHPO	5	3	2
TPO	7	5	19
HLD	1	0	0

GENES:	*E. sp.*	*E. su*	*S. ja*
GT54	1	0	0
GT57	2	2	2
GT58	1	1	1
GT59	1	1	0
GT60	3	4	1
GT64	2	2	3
GT65	1	0	0
GT66	1	0	0
GT68	0	0	2
GT74	1	0	1
GT76	1	1	1
GT77	1	0	11
GT90	0	0	1
GT92	0	0	2
GTnc	0	6	0
Alginate metabolism			
MPI	4	4	5
PMM	1	0	1
MPG	0	0	?
GMD	3	2	3
MS	2	1	1
ManC5-E	28	24	105
Fucan metabolism			
FK	2	1	1
GFPP	2	1	1
GM46D	2	2	1
GFS	1	1	1
Sulfatases & sulfotransferases (ST)			
Sulfatases	9	9	n. i.
ST	15	13	24

GENES:	*E. sp.*	*E. su*	*S. ja*
Glycoside hydrolases			
GH1	3	2	1
GH2	2	3	3
GH3	1	1	1
GH5	2	2	2
GH10	1	1	1
GH13	0	0	1
GH16	6	7	3
GH17	1	1	2
GH18	0	0	1
GH30	1	0	1
GH31	1	1	0
GH36	2	2	0
GH37	1	1	1
GH38	1	1	0
GH47	5	5	2
GH63	1	1	0
GH81	5	4	53
GH85	1	1	1
GH88	6	3	1
GH95	1	1	0
GH114	0	0	7
GH128	0	0	1
Phenolics			
PKS III	3	3	2
CHIL	2	2	2
AST	5	3	7
FGT	1	1	1
OMT	1	1	1
POX	1	1	2
TYR	28	20	22

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

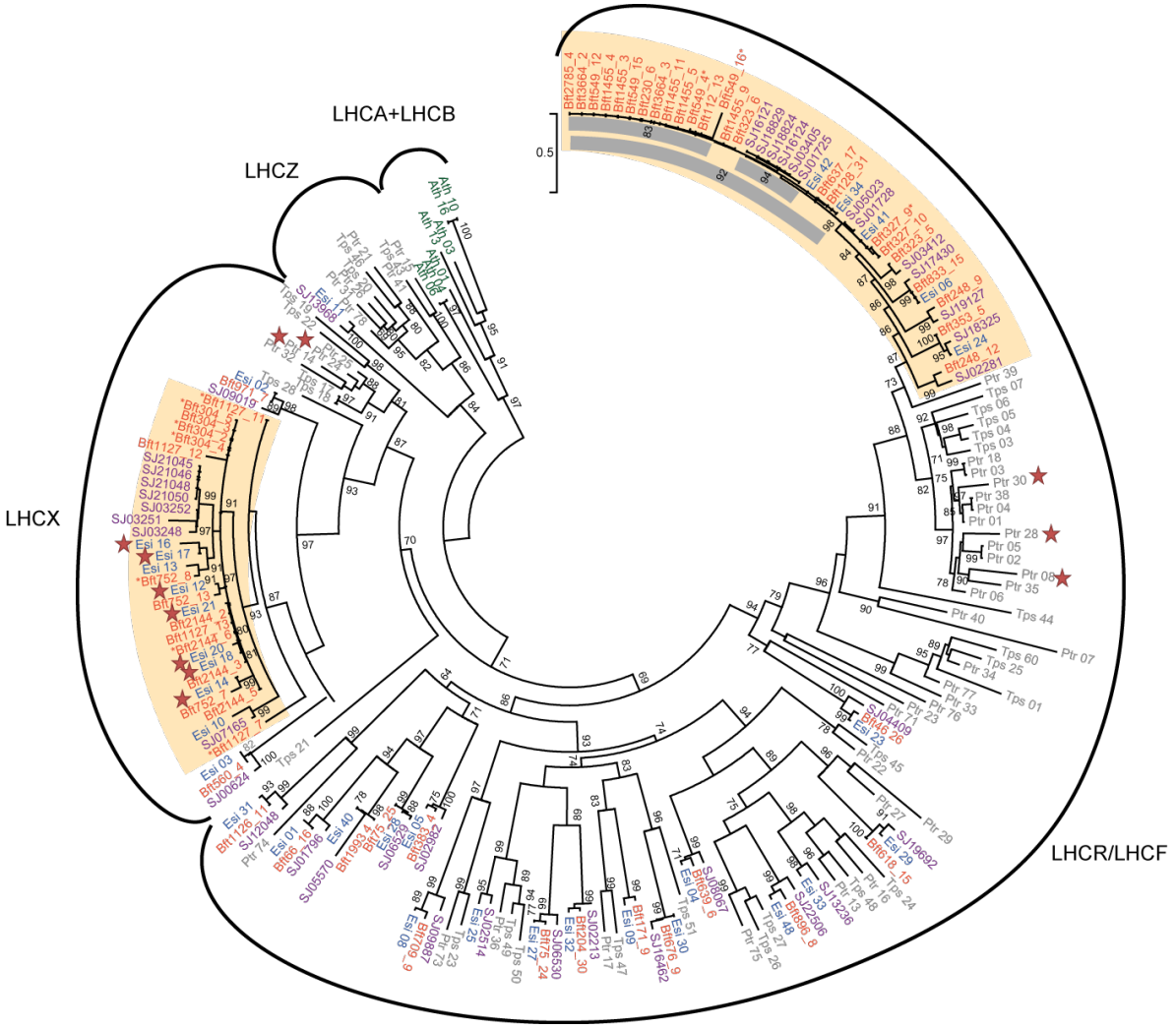


Figure 6: Maximum likelihood tree of *E. subulatus* Bft15b (orange) *Ectocarpus* sp. Ec32 (blue), *S. latissima* (purple), and diatom (*Thalassiosira pseudonana* and *Phaeodactylum tricornutum*, grey) CBP sequences. Support values correspond to bootstrap values from 100 replicate runs. *A. thaliana*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

Tables

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

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

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

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

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

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

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

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

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

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

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

Supporting Information

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

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

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

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

Supporting Information Table S5: Complete list of OrthoFinder results.

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