



**HAL**  
open science

# Candidate genes involved in biosynthesis and degradation of the main extracellular matrix polysaccharides of brown algae and their probable evolutionary history

Lisa Mazéas, Ahlem Bouguerba-Collin, J. Mark Cock, France Denoeud, Olivier Godfroy, Loraine Brillet-Guéguen, Tristan Barbeyron, Agnieszka Lipinska, Ludovic Delage, Erwan Corre, et al.

## ► To cite this version:

Lisa Mazéas, Ahlem Bouguerba-Collin, J. Mark Cock, France Denoeud, Olivier Godfroy, et al.. Candidate genes involved in biosynthesis and degradation of the main extracellular matrix polysaccharides of brown algae and their probable evolutionary history. *BMC Genomics*, 2024, 25 (1), pp.950. 10.1186/s12864-024-10811-3 . hal-04734202

**HAL Id: hal-04734202**

**<https://hal.science/hal-04734202v1>**

Submitted on 14 Oct 2024

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

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

RESEARCH

Open Access



# Candidate genes involved in biosynthesis and degradation of the main extracellular matrix polysaccharides of brown algae and their probable evolutionary history

Lisa Mazéas<sup>1†</sup>, Ahlem Bouguerba-Collin<sup>1†</sup>, J. Mark Cock<sup>1</sup>, France Denoëud<sup>2</sup>, Olivier Godfroy<sup>1</sup>, Loraine Brillet-Guéguen<sup>1,3</sup>, Tristan Barbeyron<sup>1</sup>, Agnieszka P. Lipinska<sup>4</sup>, Ludovic Delage<sup>1</sup>, Erwan Corre<sup>3</sup>, Elodie Drula<sup>5,6,7</sup>, Bernard Henrissat<sup>5,6,8</sup>, Mirjam Czjzek<sup>1</sup>, Nicolas Terrapon<sup>5,6</sup> and Cécile Hervé<sup>1\*</sup>

## Abstract

**Background** Brown algae belong to the Stramenopiles phylum and are phylogenetically distant from plants and other multicellular organisms. This independent evolutionary history has shaped brown algae with numerous metabolic characteristics specific to this group, including the synthesis of peculiar polysaccharides contained in their extracellular matrix (ECM). Alginates and fucose-containing sulphated polysaccharides (FCSPs), the latter including fucans, are the main components of ECMs. However, the metabolic pathways of these polysaccharides remain poorly described due to a lack of genomic data.

**Results** An extensive genomic dataset has been recently released for brown algae and their close sister species, for which we previously performed an expert annotation of key genes involved in ECM-carbohydrate metabolisms. Here we provide a deeper analysis of this set of genes using comparative genomics, phylogenetics analyses, and protein modelling. Two key gene families involved in both the synthesis and degradation of alginate were suggested to have been acquired by the common ancestor of brown algae and their closest sister species *Schizocladia ischiensis*. Our analysis indicates that this assumption can be extended to additional metabolic steps, and thus to the whole alginate metabolic pathway. The pathway for the biosynthesis of fucans still remains biochemically unresolved and we also investigate putative fucosyltransferase genes that may harbour a fucan synthase activity in brown algae.

**Conclusions** Our analysis is the first extensive survey of carbohydrate-related enzymes in brown algae, and provides a valuable resource for future research into the glycome and ECM of brown algae. The expansion of specific families related to alginate metabolism may have represented an important prerequisite for the evolution of developmental complexity in brown algae. Our analysis questions the possible occurrence of FCSPs outside brown algae, notably

<sup>†</sup>Lisa Mazéas and Ahlem Bouguerba-Collin contributed equally to this work.

\*Correspondence:  
Cécile Hervé  
cecile.herve@sb-roscoff.fr

Full list of author information is available at the end of the article



within their closest sister taxon and in other Stramenopiles such as diatoms. Filling this knowledge gap in the future will help determine the origin and evolutionary history of fucan synthesis in eukaryotes.

**Keywords** Extracellular matrix, Brown algae, Carbohydrate-active enzymes, Alginates, Fucans, Glycosyltransferases, Sulfotransferases

## Background

Brown algae correspond to the class Phaeophyceae within the Stramenopile lineage [1]. Many of these organisms are key components of extensive coastal ecosystems that provide high value ecosystem services, including the sequestration of several megatons of carbon per year. They represent a rich but underutilized biomass that will likely provide solutions to address global challenges, such as climate change, food shortages, and rising demands for natural bioactive compounds [2, 3]. These organisms are also impacted by climate change and shifts in their habitats and communities have been reported [4, 5]. They are key components of highly dynamic ecosystems, and to thrive in such habitats they have evolved plasticity in terms of metabolic responses, including at the level of their extracellular matrices (ECMs), or cell walls [6].

Brown algae evolved complex multicellularity independently from animals, fungi and land plants, and their metabolic networks have been shaped by secondary endosymbiosis and a number of horizontal gene transfers (HGTs) [6–8]. This has resulted in the emergence of a range of unique metabolic features, including the generation of specific glycans in their ECMs [9, 10]. Alginates and fucose-containing sulfated polysaccharides (FCSPs) are the main components of the ECM, representing together up to 45% of algal dry weight [11–13]. Alginates are linear polymers made solely of two 1,4-linked epimers:  $\beta$ -D-mannuronic acid (M) and  $\alpha$ -L-guluronic acid (G). FCSPs include both fucans and fucoidans [9]. Fucans are highly sulfated polysaccharides with a backbone structure based on sulfated L-fucose residues, on which additional branches of sulfated fucose, galactose and glucuronate can occur. Fucoidans are a set of heterogeneous polymers with non-fucose backbones (e.g. chains of galactose, mannose or glucuronate), and branches of sulfated fucose [9].

The metabolic pathways for both alginates and sulfated fucans have been predicted [14, 15]. All enzymes biochemically characterized so far are involved either in the synthesis of sugar precursors or in the remodeling of alginates [15]. GDP-mannose is an essential activated sugar used in the alginate biosynthetic pathway, and in the production of fucans through its conversion to GDP-fucose [15]. The three enzymatic steps that lead to the production of GDP-mannose from fructose-6 phosphate are catalysed by mannose-6-phosphate isomerase (MPI), bifunctional phosphomannomutase (PMM)/phosphoglucomutase (PGM) and mannose-1-phosphate

guanylyltransferase (MPG). The later enzyme also has MPI activity [15, 16]. As in other eukaryotes and in most bacteria, the precursor GDP-fucose is likely produced in brown algae from GDP-mannose (the *de novo* pathway) and from fucose (the salvage pathway) [15]. In the *de novo* pathway, GDP-mannose is first converted into GDP-4-keto-6-deoxymannose by GDP-mannose 4,6-dehydratase (GM46D), and subsequently converted to GDP-fucose by GDP-fucose synthase (GFS). In the salvage pathway, cytosolic fucose is first phosphorylated by a fucokinase (FK) and then condensed to GTP by a GDP-fucose pyrophosphorylase (GFPP) to form GDP-fucose. Once GDP-mannose has been synthesized, its oxidation to GDP-mannuronic acid will result in the elongation of an alginate polymer in the form of polymannuronate. Oxidation of GDP-mannose is catalyzed by GDP-mannose dehydrogenase (GMD).

Once these activated sugars (e.g. GDP-mannose, GDP-fucose) have been generated, the elongation or grafting of the glucan chains is the most significant step in ECM synthesis, but, for brown algae, the genes involved have not been identified [15]. Carbohydrate active enzymes (CAZymes) are major actors in polysaccharide metabolism [17, 18]. They are classified into different families in the CAZy database ([www.cazy.org](http://www.cazy.org) [17]), based on their amino acid sequence similarities. CAZy classification includes glycosyltransferases (GTs), glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), enzymes with Auxiliary Activities (AAs) and associated modules such as Carbohydrate-Binding Modules (CBMs). Additional important enzymes involved in the remodeling of glycans in brown algae include mannanuronan C5-epimerases which convert M to G residues in alginates, and the sulfotransferases (STs) and sulfatases (SAs), which mediate the sulfation and desulfation of FCSPs, respectively [7, 15]. The SAs have been classified into families in the SulfAtlas database based on amino acid sequence similarities (<https://sulfatlas.sb-roscoff.fr> [19, 20]), .

Sequencing a large number of genomes of brown algae and their sister groups, allows the emergence and diversification of key biological traits within the brown algae to be investigated, particularly traits associated with the transition to complex multicellularity, such as the elaboration of an adherent ECM. Brown seaweeds have remained poorly described in terms of genome sequencing due, in part, to difficulties with extracting nucleic acids. The Phaeoexplorer project has recently generated

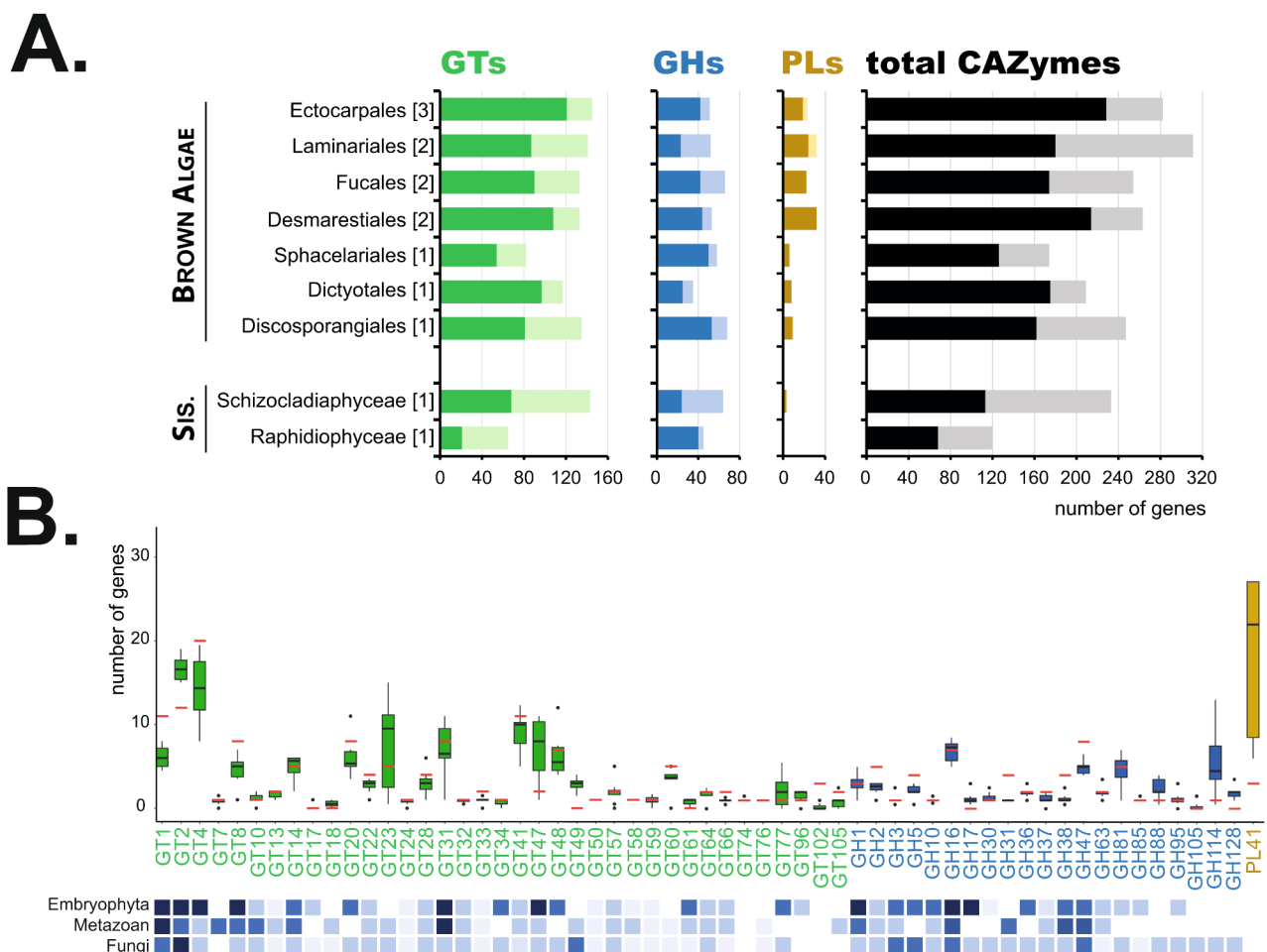
60 new genomes corresponding to 40 species of brown algae and four close sister species (<https://phaeoexplorer.sb-roscoff.fr> [21]), . The 40 brown algal species include representatives of 16 families, spanning all the major orders of the Phaeophyceae. The sister species notably include the sister taxon *Schizocladia ischiensis*, the closest taxon to the brown algae [1, 21, 22]. Here, we have investigated the evolutionary history of carbohydrate metabolism gene families in brown algae by a combination of expert annotations and evolutionary analysis of CAZymes [17], STs and SAs [19], as well as protein modelling of the most promising candidate genes. When relevant, we also performed genomic analyses of additional eukaryotes such as echinoderms or diatoms. This study offers the first representative view of the core CAZome in brown algae, derived from an extensive resource of

sequences from distinct taxonomic origins. Our analysis indicates that the whole alginic metabolic pathway, which includes the abilities to both synthesize and degrade alginate, were gained prior to the diversification of the brown algae. We also discuss the most likely fucan synthase genes in brown algae, an enzymatic step that still remains biochemically unresolved.

## Results

### CAZyme profiles in brown algae

Expert annotation of carbohydrate metabolism gene families was performed for all the major orders of brown algae (Phaeophyceae) and their close sister groups. Regarding CAZymes, we have already reported a quite constant number of genes (237 genes on average) for the 7 orders of brown algae (Fig. 1A) [21]. This number is



**Fig. 1** Overview of CAZyme families in brown algae. **(A)** Counts of numbers of genes predicted to encode glycosyltransferases (GTs), glycoside hydrolases (GHs), polysaccharide lyases (PLs) and for all CAZymes (GTs, GHs, PLs, Carbohydrate Esterases CEs, Auxiliary Activities AAs, Carbohydrate Binding Modules CBMs), showing numbers of both full-length proteins (dark colours) and fragments (light colours). The data is averaged by taxonomic order with the number of species analysed per order in brackets. **(B)** Number of genes for selected CAZyme families in brown algae (box plots) and the sister species *S. ischiensis* (red lines). Counts include non-fragmentary proteins and fragments. The species analysed are the same as in Supplementary Table S1. A schematic overview regarding abundance of the corresponding families is indicated below the plots for other groups (the reader is directed to the CAZY database for a more accurate counting in these other groups)

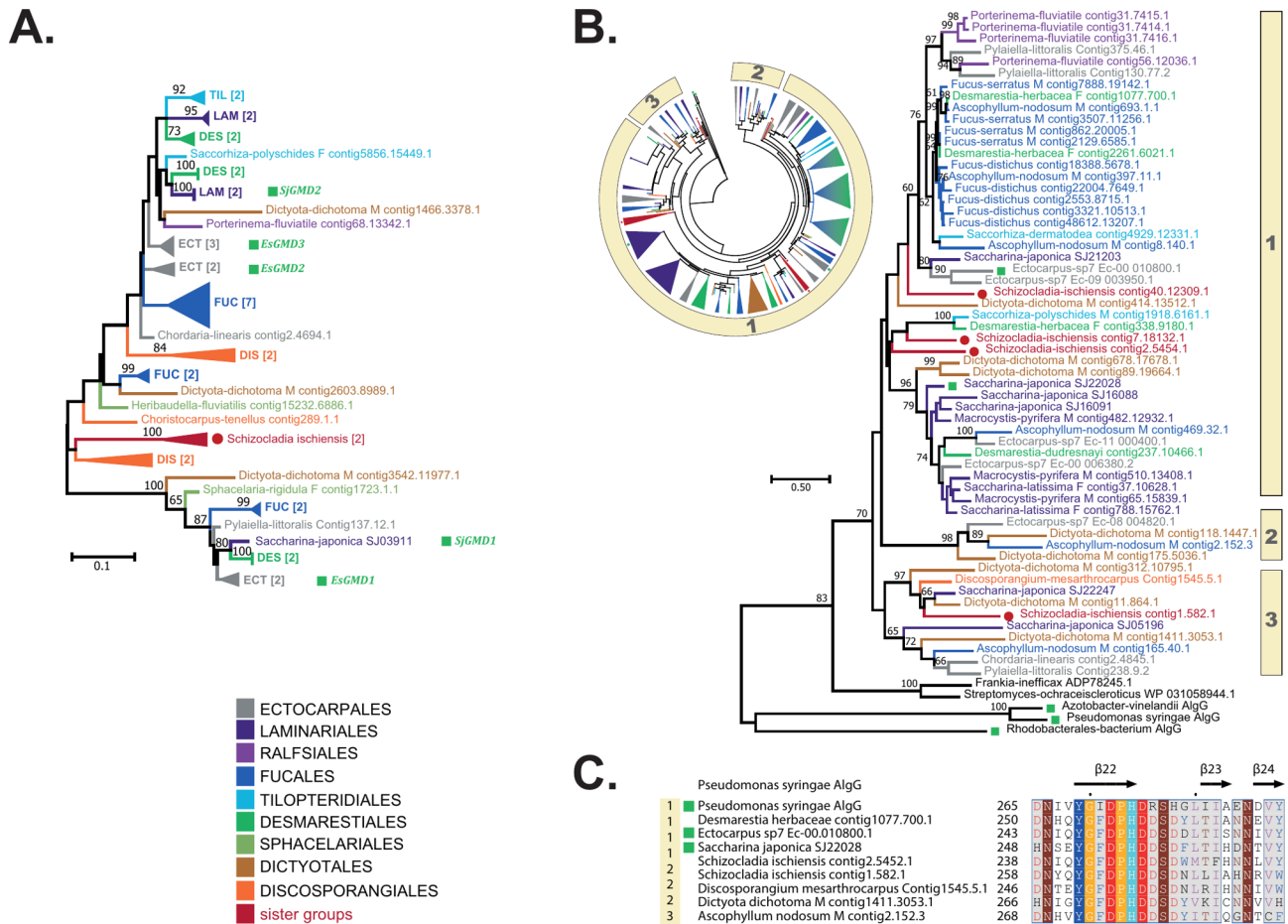
similar in the closest sister taxon *S. ischiensis* (233 genes) while it is twice more than the more distantly related unicellular alga *Heterosigma akashiwo* (120 genes). The brown algal and sister taxa CAZymes are distributed across 69 families (including 37 GT, 21 GH and 1 PL families) with, on average, three genes per family (while reaching up to 16 genes for GT2 and GT4 and 19 genes for the PL41) (Fig. 1B, Supplementary Table S1). The GT31 and GT47 families, which are abundant in Viridiplantae (essentially  $\beta$ -galactosyl- or  $\beta$ -glucuronyltransferases [23, 24]), are also among the largest GT families in brown algae (average of 9 genes per genome for both families). We also detected in brown algae the occurrence of unexpected CAZyme families such as GT23 ( $\alpha$ -1,6-L-fucosyltransferase [25, 26]), usually found in metazoan, GT49 ( $\alpha$ -glucuronyltransferase), found in sac fungi, GT60 ( $\alpha$ -N-acetyl-glucosaminyltransferase), found in amoeba plus chlorophytes. Regarding glycosyl hydrolases, the most populated families were GH16, GH47, GH81 and GH114 (with 6, 5, 4 and 5 genes per genome respectively). The GH16 and GH81 enzymes are likely involved in the metabolism of the storage polysaccharide laminarin in brown algae, with  $\beta$ -1,3-glucanase activities already referenced in those families. GH47 genes are frequently found in eukaryotic genomes and are associated with  $\alpha$ -1,2-mannosidase activity. Outside brown algae, GH114 are found in chlorophytes, fungi and bacteria, with a reported  $\alpha$ -1,4-galactosaminidase activity [27, 28]. Finally PL41, which is, so far, the only PL family reported in brown algae and which has alginate lyase activity [29], contains a high number of genes. In the following lines we discussed the genes that we have annotated and that are possibly involved in ECM-carbohydrate metabolism in brown algae (Supplementary Table S2).

#### **The genes involved in alginate metabolism in brown algae are also present in its closest sister taxon *S. ischiensis***

The alginate and the *de novo* fucan biosynthetic pathways share common steps leading to the generation of GDP-mannose as a precursor, prior to its conversion to GDP-fucose in the case of fucans. The two enzymes involved in these enzymatic conversions, MPI/MPG and PMM/PGM, have been characterized for *Saccharina japonica* [16, 30]. Analysis of their phylogeny indicated that these enzymes are widely present in brown algae but also more generally within the Stramenopiles, with several occurrences in the closest sister taxa and diatoms (Supplementary Fig. S1A, B). GDP-mannose dehydrogenase (GMD) is essential to provide GDP-mannuronic acid for the initial assembly of alginate in its mannuronan form by a mannuronan synthase. The latter putatively belongs to the large and multispecific GT2 family, although no functional validation has been obtained so far to identify specific candidate genes, making phylogenetic studies

of this family highly speculative. The alginate obtained undergoes further modelling through the action of mannuronan C5-epimerases (ManC5-E). Both GMD and ManC5-E activities have been described, including in brown algae [16, 31–34].

GMD belongs to a protein superfamily that also includes uridine diphosphate-glucose dehydrogenases (UGD). In order to explore the sequence diversity within this family, we gathered GMD and UGD protein sequences, using representative sets of brown algal sequences and additional characterized members from public databases. After clustering their BLASTp+Score Ratio (BSR), we generated heatmaps that clearly illustrates that UGD and GMD form two distinct clusters (Supplementary Fig. S2A). Within the first cluster, the UGD sequences from brown algae and the sister taxon *S. ischiensis*, share more than 90% pairwise protein sequence identity. In the second cluster, GMD sequences from brown algae and the sister taxon *S. ischiensis*, share more than 77% pairwise protein sequence identity, while they share less than 25% identity with the UGD sequences, indicating a large evolutive distance between these two families. The BSR-based heatmap indicates a distant cluster of sequences from *Actinobacteria*, which share around 30% identity with both GMD and UGD brown algal sequences. An additional BLAST search indicated that the brown algal GMDs share stronger sequence identities (40%) with a protein belonging to *Candidatus woesebacteriota* (MBP7967557.1), an uncultivated bacterium from a putative novel taxon [35]. The phylogeny suggests that the common ancestor of *S. ischiensis* and brown algae probably gained the GMD sequence via horizontal gene transfer (HGT), and that the gene was duplicated in all brown algal groups (Fig. 2A, Supplementary Fig. S2B). The protein alignment of representative sequences is further indicative of the strong similarity within the GMD family (Supplementary Fig. S3). The mannuronan C5 epimerases are a large multigenic family in brown algae, with an average of 27 members identified per genome. Eleven homologues were identified in *S. ischiensis* (Fig. 2B, Supplementary Fig. S4). Unlike the aforementioned families, screening the other Stramenopile genomes available (i.e. other brown algal sister taxa, diatoms and oomycetes), did not identify any ManC5-E sequences. Three ManC5-E orthologue groups were identified (Fig. 2B, Supplementary Fig. S4). All the ManC5-E sequences possess the DPHD motif, typical of the ManC5-E catalytic site (Fig. 2C, Supplementary Fig. S5), indicating that these sequences are genuine ManC5-Es. Marked amplifications of the gene family were detected in the Fucales, Desmarestiales and Laminariales. The closest bacterial representatives are from *Actinobacteria* with ~30% protein pairwise identity (Fig. 2B, Supplementary Fig. S4).



**Fig. 2** Phylogenetic trees of key enzymes involved in the synthesis of alginates. **(A)** Phylogenetic tree of GDP-mannose 6-dehydrogenases (GMD). The uncompressed tree is shown in Supplementary Fig. S2B. **(B)** Phylogenetic trees of mannuronan C5-epimerases (ManC5-E). The circle shown as an inset indicates a global phylogeny of a large collection of ManC5-Es from various genomes, in which 3 main orthologous groups can be found. The phylogeny shown on the right side is representative of this global view, the 3 groups being indicated. The uncompressed tree of the inset is shown in Supplementary Fig. S4. **(C)** Extract of an alignment of selected ManC5-E sequences from brown algae and the sister species *S. ischiensis*, the bacterial ManC5-E sequence from *Pseudomonas syringae* (pdbcode: 4NK6). The region shown corresponds to the catalytic site as identified in 4NK6, and the secondary structure assignments shown above the sequences are those determined for 4NK6. The red arrowheads below the alignment indicate the conserved catalytic residues His274 and Asp275. The numbering on the left refers to the phylogenetic clusters indicated in B). The full alignment is shown in Supplementary Fig. S5

In addition to alginate synthesis, brown algae also possess a metabolic pathway for alginate degradation, the main actors being the PL41s. This is one of the largest degrading CAZyme families in brown algae (Fig. 1B), with an average of 19 genes per genome, and can be classified into at least four orthologous groups. PL41 duplication events have occurred in several brown algal orders including the Ectocarpales and the Laminariales. *S. ischiensis* displays three PL41 sequences while no homologues were found in other Stramenopiles (Fig. 3A, Supplementary Fig. S6, S7). PL41s are however not limited to brown algae, with bacterial members in some *Gamma-proteobacteria*, *Actinobacteria* and *Alphaproteobacteria*, but with an average of one gene copy per genome (34–40% pairwise identify with the PL41 domain of brown

algae). PL41 degrades alginates to unsaturated oligomannuronic acids, the smallest reaction product being the unsaturated monosaccharide, which is spontaneously converted to 4-deoxy-1-erythro-5-hexoseulose uronate (DEHU) [29]. The generated DEHU has been shown to be further degraded into 2-keto-3-deoxy-gluconate (KGD) by a DEHU reductase in the brown alga *S. japonica*, which thus acts downstream of PL41 [36]. DEHU homologues were found in most brown algal genomes but not all sister taxa. This family of genes is found in other eukaryote genomes (Supplementary Fig. S8). This enzyme is part of the aldo-keto reductase (AKR) superfamily [36] which is found in all life domains and contains enzymes with broad substrate specificities against endogenous substrates and xenobiotics (carbohydrates,

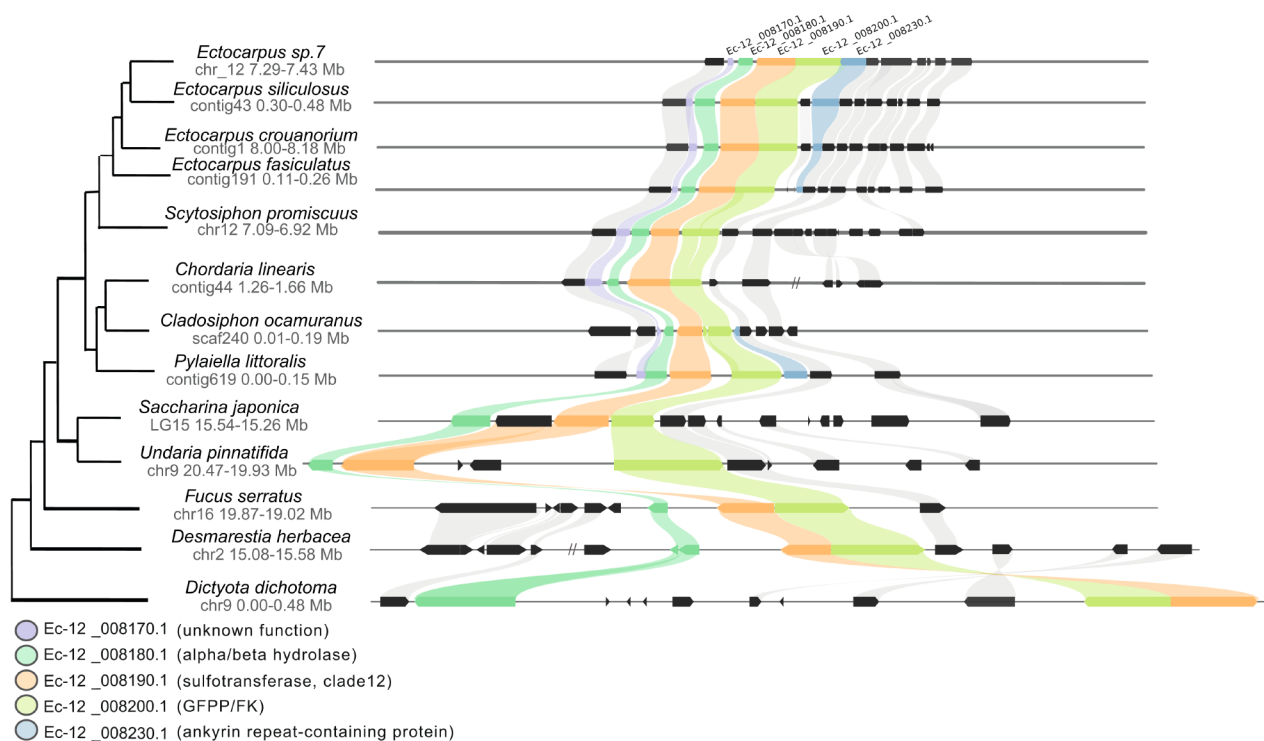


This microsyntenic region was found in other members of the Ectocarpales (Fig. 4) and is conserved in other brown algal orders (Laminariales, Fucales, Desmarestiales), albeit with a greater distance between the genes near the root (*Dictyota dichotoma*). The synteny was not observed with the Discosporangiales and more distant taxa (sister taxa), however this could be due to the less contiguous genome assemblies of *Choristocarpus tenellus* (N50=6.62 Kbp) and *Discosporangium mesarthrocarpum* (N50=12.28 Kbp) [21].

The most important set of enzymes involved in the synthesis of fucans are the fucosyltransferases (FucTs), which transfer GDP-fucose to oligofucans to elongate the chains. Their exact nature is still unknown. The CAZY database includes several FucTs that have been identified and these are classified into a variety of GT families. Our previous analyses of a restricted number of brown algal genomes, indicated that FucT candidate genes for fucan synthesis may belong to either the GT10, GT23, GT41 and GT74 families [15]. The extensive Phaeoexplorer dataset allowed us to perform a more in-depth analyses of these CAZyme families in brown algae, and to refine the identification of putative fucan synthases (Supplementary Tables S1, S2).

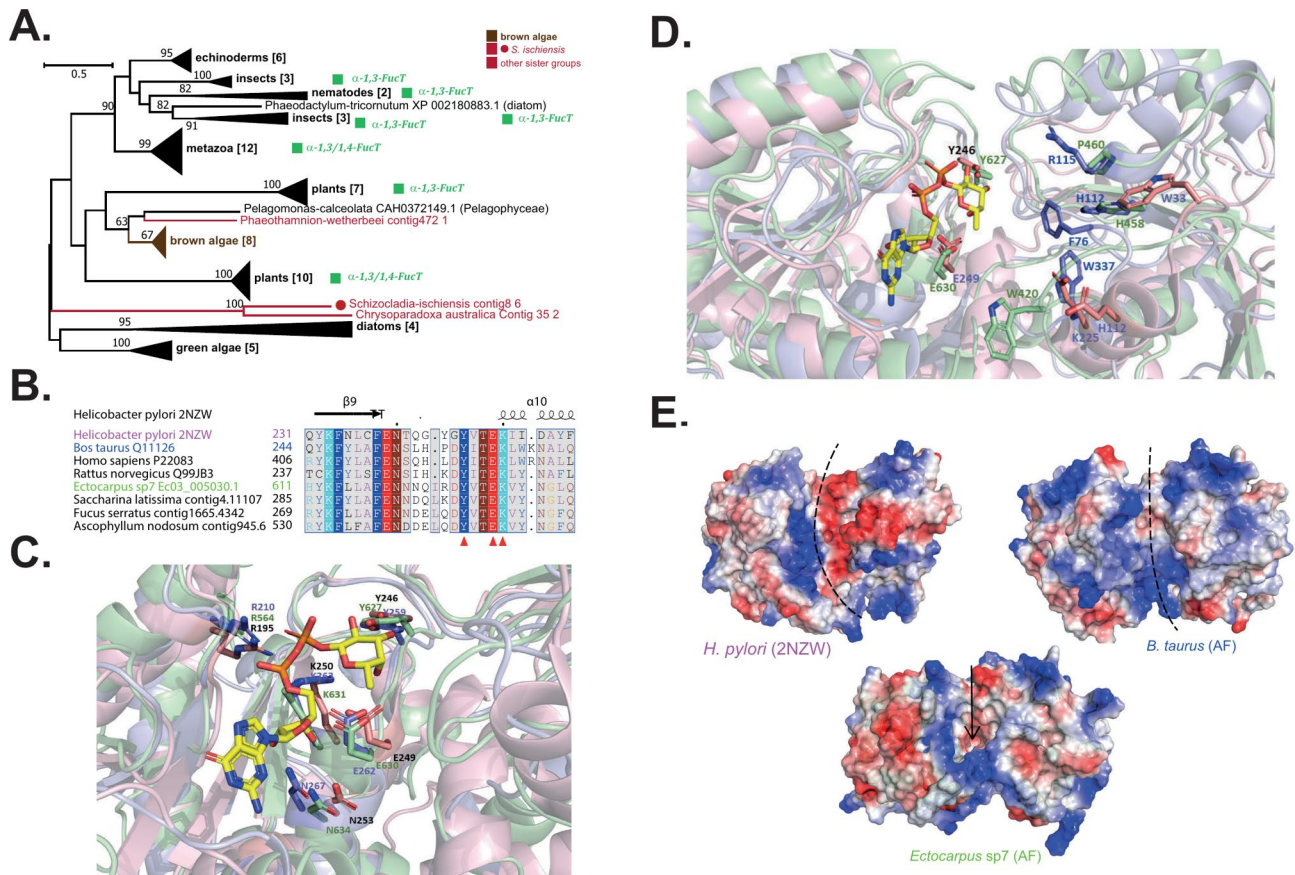
GT10s are referenced as  $\alpha$ -1,3 or  $\alpha$ -1,4-FucTs in the CAZY database, and the activities described relate to the transfer of an N-acetyl glucosamine (GlcNAc) residue

[41, 42]. The fucose residues in fucans are  $\alpha$ -1,3 and/or  $\alpha$ -1,4-linked. It is thus tempting to hypothesise that the GT10 family is involved in their synthesis. However, there is only one GT10 gene per brown algal genome on average (Fig. 5A, Supplementary Fig. S13, Supplementary Table S1). Considering the importance of sulfated fucans in ECMs of all brown algae, we would expect the corresponding FucT gene to be present in multiple copies. As such this GT10 family is not the most favorable candidate for their biosynthesis. Moreover, a close homologue of brown algal GT10s is found in the Pelagophyceae, which are known not to contain fucans. The phylogeny further indicates that the brown algal sequences are closely related to plant GT10s, which are involved in protein glycosylation, where the  $\alpha$ -1-3 FucT activity takes place on a GlcNAc residue, directly linked to an asparagine [43]. While the activity of brown algal GT10s on oligofucans is unlikely, the GDP-fucose binding site is nonetheless highly conserved (Fig. 5B, C), indicative of FucT activity. Comparisons were made with the only known structure of a GT10 from bacteria [44], and additional animal GT10 sequences for which the donor binding site has been studied [45, 46]. Both the alignment (Fig. 5B, Supplementary Fig. S14) and the structure predictions (Fig. 5C) indicate the conservation of key residues within the donor site in the brown algal sequences, while the acceptor site exhibits considerable variation (Fig. 5D, E).



**Fig. 4** Syntenic region of putative fucan-related genes in brown algal genomes. This region contains five conserved genes encoding: a protein with an unknown function, an alpha/beta hydrolase, a sulfotransferase (ST), a GDP-fucose pyrophosphorylase/L-fucokinase (GFPP/FK) and an Ankyrin repeat-containing protein



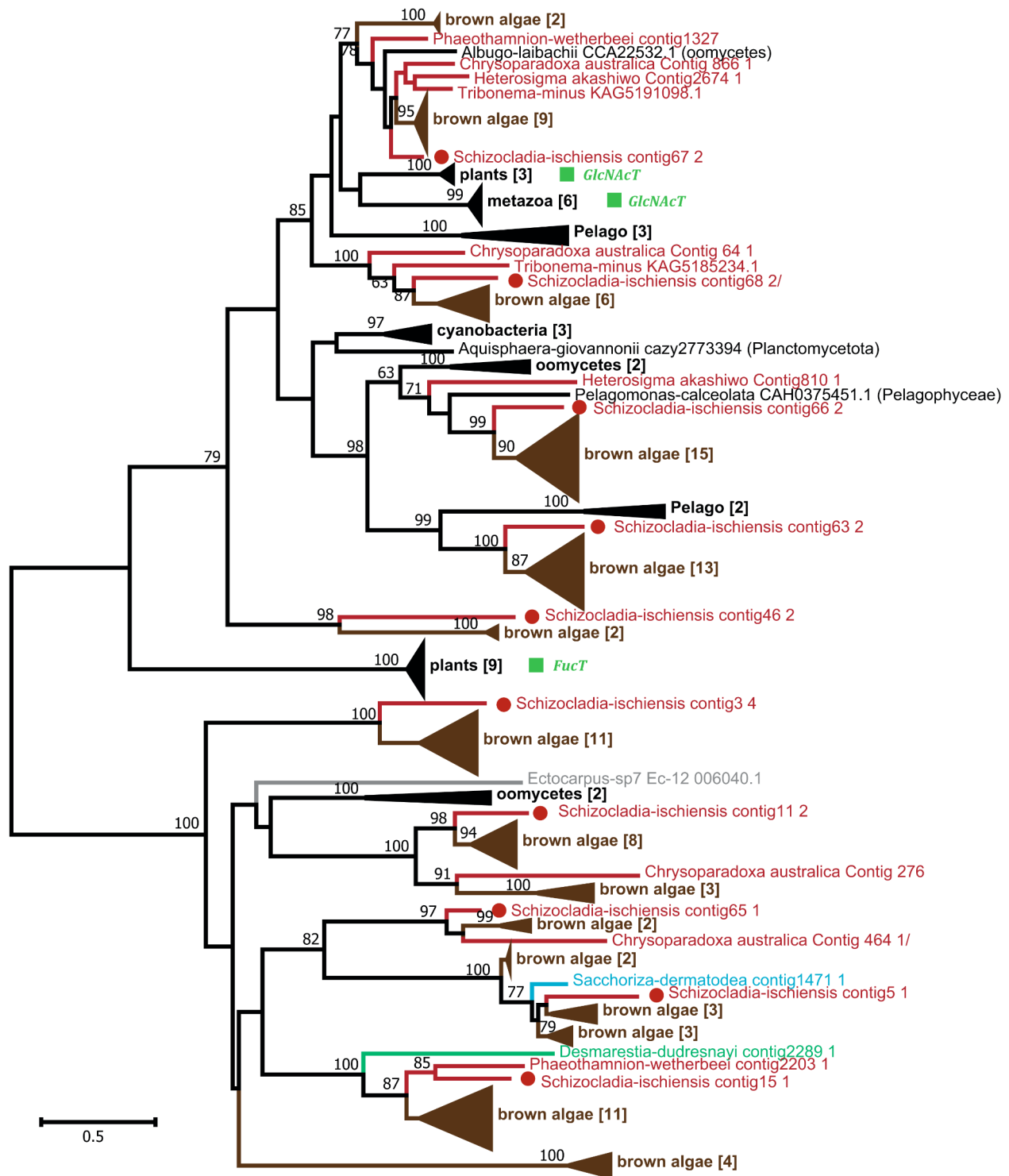


**Fig. 5** Comparison of brown algal GT10s with biochemically characterised orthologues. **(A)** Phylogenetic tree of the GT10 family. This family is known to contain FucT activities for species outside brown algae. The uncompressed tree is shown in Supplementary Fig. S13. **(B)** Extract of an alignment of selected GT10 sequences from brown algae against the bacterial GT10 sequence from *Helicobacter pylori* (pdbcode: 2NZW) focusing on the GDP-fucose donor site. Additional animal sequences are also aligned and secondary structure information is shown for 2NZW. The red arrowheads below the alignment indicate the conserved catalytic residues Tyr246, Glu249 and Lys250, which interact with the donor substrate GDP-fucose. The 2NZW sequence is shown in pink, a *Bos taurus* sequence in blue and the *Ectocarpus* species 7 sequence in green, and the same color code is used in C, D and E. The full alignment is shown in Supplementary Fig. S14. **(C)** Overlay of the *H. pylori* GT10 crystallized structure (2NZW; pink) and the best models generated by AlphaFold2 for the *B. taurus* (blue) and brown algal (green) sequences. The cartoon representation focuses on the GDP-fucose binding site where GDP-fucose is apparent in yellow/orange colors. Some conserved catalytic residues, including those discussed in B), are highlighted. **(D)** Larger view of the overlay presented in B) which allows visualization of the acceptor binding site, in addition to the GDP-fucose binding site. Despite apparent structural conservation, the nature of the residues at the pocket surface differ drastically between the three sequences. **(E)** Molecular surfaces of 2NZW and the AlphaFold2 predictions for the *B. taurus* and algal sequences based on their electrostatic potentials. While the substrate binding cleft is clearly apparent in 2NZW and in the *B. taurus* sequence (dotted lines), it appears more enclosed in the brown algal sequence (arrow), although the N-terminal domain may form a flexible loop to allow accommodation of a large substrate

GT41 is a multispecific family in the CAZy database, with N-glucosyltransferase, N-acetylglucosaminyl-transferase and protein O-fucosyltransferase activity, the latter being restricted, so far, to the green lineage [47]. Brown algal genomes harbor ten GT41 genes on average, with homologues found in all ochrophytes, usually with several copies per genome. The GT41 phylogeny indicates that brown algal GT41s are organized into numerous orthologous groups, with *S. ischiensis* frequently found at the base of these algal groups. None of these orthologues are closely related to plant O-FucT (Fig. 6, Supplementary Fig. S15). Both the alignment and protein modelling of some selected algal sequences indicate a donor site which favours the binding of a UDP-sugar rather than a

GDP-sugar such as GDP-fucose (Supplementary Fig. S16, S17). A significant number of the orthologous groups are distant from the orthologous groups that contain the GlcNAc characterized activity, and could indicate new activities in the family. Overall, the phylogeny shows that the evolution of O-FucT activity has been validated only within the green lineage so far. Finally, the presence of multiple GT41 homologues in ochrophytes other than brown algae, does not favour this class of enzymes as strong candidates for fucan synthesis.

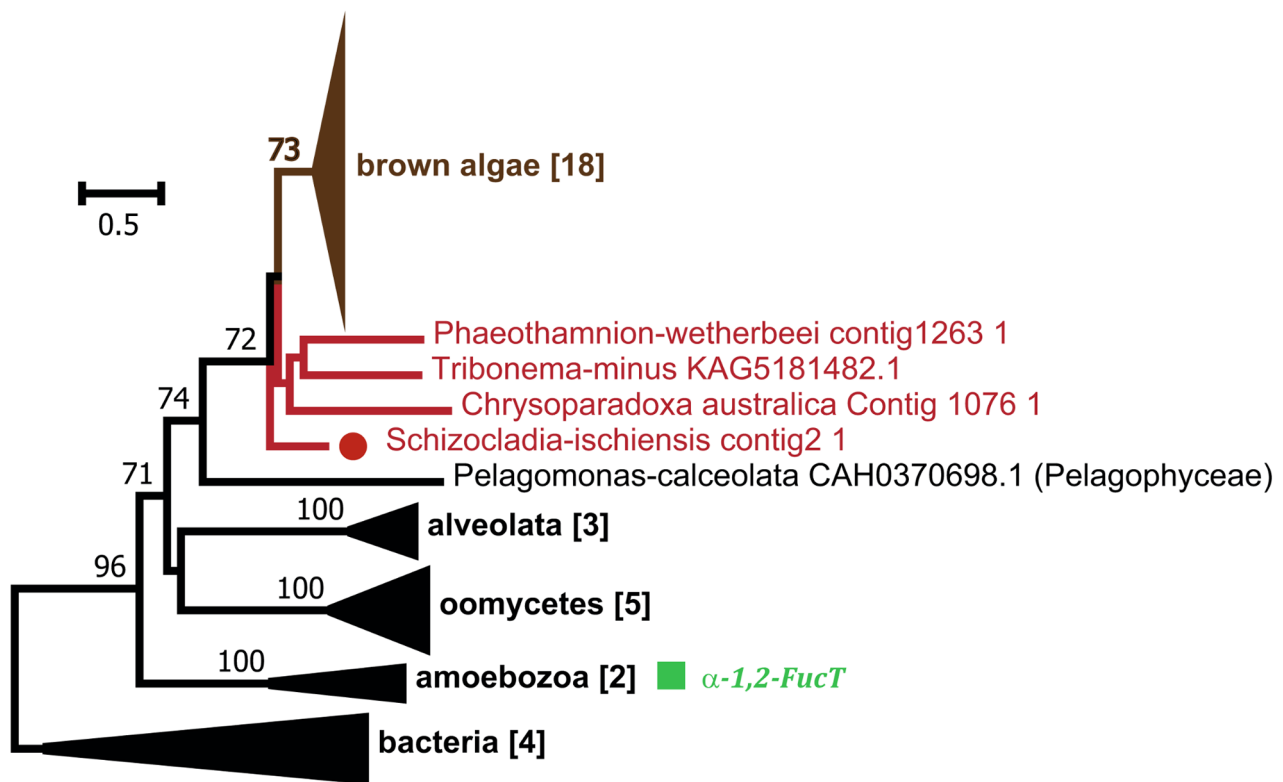
$\alpha$ -1-2 fucosyltransferase activity has been reported in the GT74 family, which is also represented in brown algal genomes. This family has a restricted occurrence in eukaryotes, being only characterized in the amoeba



**Fig. 6** Phylogenetic tree of the GT41 family. Some GT41 family members in non-brown-algal species are known to have FucT activities (green boxes). The uncompressed tree is shown in Supplementary Fig. S15

*Dictyostelium discoideum*, where it participates in the glycosylation of the Skp-1 protein, involved in the ubiquitination regulatory system [48]. Phylogenetic analysis indicated a common origin of GT74 genes from brown

algae and other Stramenopiles (Fig. 7, Supplementary Fig. S18), although some genomes lacked this gene (i.e. some diatoms and brown algal sister taxa). The GT74 protein alignment indicates large insertions between conserved



**Fig. 7** Phylogenetic tree of the GT74 family. Some GT74 family members in non-brown-algal species are known to have FucT activities (green box). The phylogeny is based uniquely on the catalytic GT74 domain. The uncompressed tree is shown in Supplementary Fig. S18

motifs in the brown algal sequences, as compared to the amoeba sequence (Supplementary Fig. S19). The *D. discoideum* enzyme has a bimodular structure, the GT74 module being appended to a GT2 domain with  $\beta$ -1,3-galactosyltransferase activity [48]. Bimodularity with a GT2 was also observed in the brown algal sequences. Most brown algae have only one GT74 gene. Again, considering the preponderance of sulfated fucans in ECMs of brown algae, and that this gene occurs in Stramenopiles known not to produce fucans, the GT74 family is not the best candidate to contain fucan synthases.

Finally, the GT23 family is present in all brown algae and is known to contain fucosyltransferases in other eukaryotes. On average, seven GT23 genes were found per genome in brown algae (Supplementary Tables S1, S2). GT23 genes are usually found in metazoa, where  $\alpha$ -1,6 FucT activity has been reported and the structure of the human isoform FUT8 has been characterised [25, 26]. Uncharacterized homologues have also been found in bacteria, fungi and the Chlorophyta but this family is very rare in these lineages. Despite bioinformatics signals of primary sequence conservation, the GT23 from brown algae are only distantly related to other sequences in the family. A BSR-based heatmap indicated two main clusters (Fig. 8A), with the brown algal sequences representing an independent cluster, sharing on average less

than 30% pairwise identity with the second cluster that contains the metazoan sequences. The phylogeny of the brown algal GT23 sequences shows that they do not strictly distribute in relation to taxonomy and this may indicate the presence of orthologue groups (Fig. 8B, Supplementary Fig. S20). All these GT23 sequences share, on average, 37% pairwise protein identity. Some duplication events have occurred in the Fucales and Laminariales. Overall, this analysis indicated that brown algae contain GT23 sequences that may belong to one or to several sub-classes within the GT23 family. The diversity of brown algal GT23 proteins may indicate functionally and/or structurally distinct proteins. This diversity of proteins could potentially reflect the variety of positions to which a fucose is transferred in fucans and more generally in FCSPs (variety of monosaccharides, carbon position, linkage type). As such, the diverse GT23 proteins may reflect the FCSP diversity and the complex enzymatic machinery required for their synthesis.

As the brown algal GT23 sequences seem rather distant from other GT23s, and knowing that some of the latter have been biochemically and structurally characterized, we examined further the potential catalytic sites of the brown algal sequences. A sequence alignment was made for a selection of GT23 proteins from brown algae and the sister taxon *S. ischiensis*, against the human

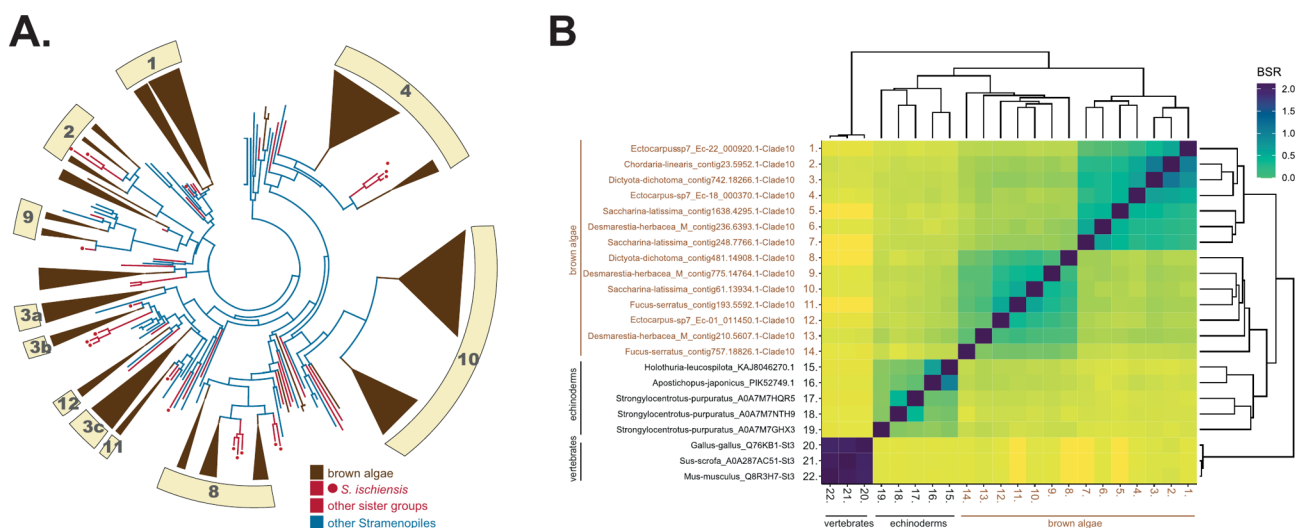


mammalian GT23s. One could speculate that brown algal GT23s catalyse transfer onto oligosaccharidic precursors of FCSPs (e.g. FCSOs).

Sulfated fucans are glycans largely restricted to brown algae, exception made of some marine echinoderms (i.e. body wall of sea cucumbers, egg jelly coat of sea urchins) where these molecules are known to be transiently produced and to occur with more linear and regular structures [9, 51, 52]. We therefore also explored the CAZyme content of publicly available echinoderm genomes. We report the presence of the GT10, GT23, GT41 families (Supplementary Table S1). GT10 members are particularly abundant in these organisms, with on average 79 members, compared to 4 GT23 members. The GT23 members in these systems are not very similar to brown algal members of the family and are more closely associated with animal orthologues (Fig. 8A). We also screened diatom genomes as recent literature suggest the presence of fucans in these organisms [53]. We did not detect any GT23 members in this case and only an average of 3 GT10 members per genome (Supplementary Table S1).

FCSPs, and especially fucans, are known to be heavily sulfated in brown algae. The FCSPs are likely initially polymerized in the state of neutral glycans by GTs, prior to their sulfation by STs. GTs and STs form large multigenic families in brown algae and brown algal STs can be clustered into at least 15 clades [15]. In order to investigate the origin of brown algal STs within the Stramenopiles, we analysed the ST gene content of all the Phaeoexplorer brown algal and sister taxa genomes,

along with genomes of other Stramenopiles. A phylogenetic tree built using these sequences (Fig. 9A, Supplementary Fig. S22) confirmed the occurrence of ~10 major orthologous groups in brown algae. All of these clades could include members with activities on polysaccharides, apart from clade 1, which is expected to be active on non-carbohydrate substrates [7, 14, 15]. Clade 12 includes genes that are located within the syntenic region that also contains the dual FK-GFPP enzymes for the synthesis of the GDP-fucose precursor (Fig. 4). This ST and FK-GFPP gene cluster is present in most brown algal genomes, including early diverging ones, suggesting a role for the STs in the sulfation of fucans or fucoidans. In brown algae, clades 4 and 10 include a high number of duplicated genes. Clade 4 includes members from other ochrophyta species but clade 10 appears to be specific to brown algae. In order to study the possible relationship of echinoderm STs with those of brown algae, especially those belonging to clade 10, we clustered these protein sequences using the BLASTP+Score Ratio (BSR) approach (Fig. 9B). The BSR-based heatmap indicates a closer relationship of the brown algal ST sequences from group 10 to echinoderm counterparts than to vertebrate STs, albeit with a pairwise identity score of below 30%. A protein alignment of some of these sequences can be found in Supplementary Fig. S23. Note that no echinoderm STs have been biochemically characterized so far, thus the conclusion of their possible involvement in fucan sulfation is highly speculative. Regarding the sulfatase genes, as previously reported, brown algae and *S.*



**Fig. 9** Phylogenetic position of STs from brown algae and other eukaryotes. **(A)** Phylogenetic tree of STs from brown algae and other Stramenopiles. The phylogenetic tree is shown in circle format with collapsed clades. Brown algal sequences are indicated in brown. Sequences belonging to sister groups of brown algae Chrysosporadoxophyceae and Schizocladiphyceae are indicated by red lines, with red circles for the latter. Other Stramenopile sequences are shown in blue. Numbers indicate the brown algal clades as identified in [15]. The corresponding uncompressed and fully-labelled tree is shown in Supplementary Fig. S22. **(B)** Heatmap representing BLASTP + Score Ratios (BSRs) for ST protein sequence pairwise alignments. The colour code correlates BSRs. Four groups are apparent: two harbouring brown algal STs (orthologues from group 10), one containing metazoan sequences and the last group containing STs from echinoderms, which is more closely related to brown algal sequences than the metazoan group

*ischiensis* contain 7 members per genome on average and they all belong to the S1\_2 family [21]. As for the STs, no biochemical activities have been determined so far for the corresponding enzymes.

## Discussion

### CAZyme numbers may not correlate with developmental complexity in brown algae

Our analyses provide the first representative view of the core CAZome in brown algae, derived from an exhaustive resource of sequences of distinct taxonomic origins, covering 16 Phaeophyceae families. Prior to the acquisition of large genomic datasets for these organisms, it had been suggested that some CAZyme families, especially GTs, were amplified in algae with more complex tissues (e.g. the kelp *Saccharina japonica*) relative to filamentous forms (e.g. *Ectocarpus* species 7) [54]. However, this statement has been gradually invalidated as more genomes have become available [15], and the analysis reported here confirms that the number of total CAZymes in individual brown algal genomes is not correlated with either tissue complexity or taxonomic origin. The filamentous sister taxon *S. ischiensis* only harbors a slightly lower number of CAZymes and the other sister taxa analyzed possessed half as many as *S. ischiensis*. The latter case may indicate the absence of a dense and carbohydrate-rich ECM surrounding these unicellular organisms. One difficulty regarding 'exotic' genomes as compared to referenced/model species, lays in the automatic structural annotation of the genes, which may lead to the prediction of fragmented genes. In brown algae, the accounting of domain duplicates in predicted proteins and/or gene fragments by different authors, sometimes in the same organism, can lead to distinct conclusions [15, 54, 55]. One should proceed with caution when drawing conclusions from the analysis of datasets coming from distinct sources. In our specific case, while we cannot address the issue regarding prediction of gene fragments at the whole genome scale, all species have been treated equally from sample preparation to structural and expert annotation, thus limiting the risks in giving biased statements.

Regarding the CAZyme families shared by all brown algae, GT2, GT23, GT31, GT47 and PL41 have a relatively high number of genes per genome, which would be coherent with them being involved in the assembly and remodelling of cell wall polysaccharides. The GT2 family is a large and multispecific family (e.g. includes proteins with different biochemical activities). It is generally one of the largest GT families in eukaryotes, and this is also the case in brown algae. Around two thirds of the GT2 sequences in brown algae are putative cellulose synthase (CESAs) and cellulose synthase-like (CSL) proteins [14], with some being possibly involved in the synthesis of mixed linkage glucans [56]. Other GT2s related to

dolichyl-phosphate  $\beta$ -glucosyltransferases are involved in protein N-linked glycosylation. The remaining GT2 members do not relate to any specific GT2 homologue with a known activity, and may include putative mannan synthases [14]. Note that brown algal GT2 genes do not share homology with the Alg8 gene, a GT2 member known to be involved in alginate synthesis in alginate-producing bacteria [14].

GT31 is also a large family in other eukaryotes (e.g. 26 genes in *Homo sapiens* and 33 genes in *Arabidopsis thaliana*) with essentially glycoprotein galactosyltransferase and N-acetylglucosaminyltransferase activities and glycolipid activity in metazoans [57], and galactosyltransferase activity during the synthesis of arabinogalactan proteins (AGPs) in plants [23]. Similarly, GT47 members are involved in the synthesis of glycosaminoglycans in metazoans and have various activities in the synthesis of cell-wall polysaccharides in plants [24]. Brown algae have been shown to express chimeric AGPs [58] and it is tempting to speculate that the GT31 members may be involved in the synthesis of these molecules, although a definite chemical and structural analysis of those glycoproteins still needs to be carried out for brown algae. The PL41 family may be the largest, probably monospecific, CAZyme family in brown algae, indicating the importance of alginate metabolism in these organisms.

### The abilities to synthesize and degrade alginate in brown algae were gained prior to their diversification

Alginate is a major component of brown algal cell walls where its rheological properties plays important roles in determining the wall's mechanical strength. Our extended analyses confirmed a previous statement [21] that several key gene families involved in alginate synthesis (GMD, ManC5-E) and degradation (PL41, and possibly GH88) were acquired by the common ancestor of brown algae and *S. ischiensis*. This event was followed by a marked amplification of some key families (ManC5-E, PL41) in brown algae relative to its sister taxon.

The ManC5-Es catalyze the final step of alginate synthesis in brown algae: epimerization of mannuronic acid (M) residues into guluronic acid (G) residues, which occurs directly within the polymannuronate polymer. Brown algae are known to contain a large number of ManC5-E genes [15, 33, 55, 59] and this diversity of enzymes presumably yields polymers with a variety of either random or block-wise distribution of G residues, and thus alginates of distinct rheological properties. These modifications are thought to be regulated in response to stress [59], seasonal variations [60] and developmental processes [33]. However, this is a one-way process, as the G residues cannot be modified further once generated, thus locking the alginate in a definitive chemical and rheological state. PL41 activity would be

relevant in this context, to further modulate the length and the properties of the alginate chains during cell elongation or development. The large size of the PL41 family in brown algae indicates a variety of lytic activities correlated with alginate patterns. As a consequence of these activities, alginates, mannuronan C5-epimerases and alginate lyases, are likely to determine the porosity and mechanical strength of the wall, and may therefore be seen as functional analogues to pectins, pectin methyl-esterases and pectin lyases in plants, respectively [61].

Alginate is also synthesized by bacteria belonging to the *Azotobacter* and *Pseudomonas* genera (*Gamma-proteobacteria*) [62, 63]. The multigenic mannuronan C5-epimerase family has been extensively studied in these organisms, including at the biochemical level [64]. These bacteria also produce alginate lyases, exclusively of the PL5 [65, 66] and PL7 families [67, 68]. *Pseudomonas aeruginosa* is an opportunistic human pathogen, and a leading cause of chronic infection in the lungs of immunocompromised individuals. *Azotobacter vinelandii* is a nitrogen-fixing bacterium found in soil. The ability of these two bacteria to synthesize alginate gives them several selective advantages, the main one being the implementation of a protective shield against adverse environmental conditions [68, 69]. With the exception of PMM (AlgC), all the genes required for alginate biosynthesis are clustered in a single operon, including PL5-alginate lyase (AlgL) [70, 71]. The products derived from the lyase action are not used as a carbon source by the bacteria, thus several functions have been putatively attributed to these lyases [69, 70]. In *P. aeruginosa* they contribute to cell viability by clearing the periplasmic space of trapped alginate [69, 71]. In *A. vinelandii* they trigger coat rupture during cyst germination [67]. Additional alginate lyase activities have been described in other PL families, essentially from bacteria (PL6, PL15, PL17, PL18, PL31, PL32, PL36, PL39) and some eukaryotes (PL8, PL14) [15]. Screening of public databases indicates uncharacterized PL41 homologues in bacteria, essentially from the *Streptomyces* within the *Actinomycetota* (Fig. 3A, Supplementary Fig. S6) but these organisms are not known to contain an alginate biosynthetic pathway. Indeed, in contrast to the *Azotobacter* and *Pseudomonas* genera, most of these bacteria use the products of alginate lyase action as a carbon source [72].

The occurrence of alginate metabolic pathways in *S. ischiensis* correlates with the immunodetection of alginate in this species [22]. There is strong evidence that the ManC5-E, PL41 and GH88 genes have been acquired by *S. ischiensis* via horizontal gene transfer (HGT). We have previously proposed that about 1% of brown algal genes are derived from HGTs, with these genes having been principally acquired from bacterial genomes [21].

Around 10% of those HGT-derived genes have predicted functions in carbohydrate transport and metabolism [21].

While previous work indicated that brown algal GMD was derived from a HGT of an ancestral GMD sequence from *Actinobacteria* [14], our analyses based on additional algal sequences indicates that this conclusion is not strongly supported, and other origins of horizontal gene transfer cannot be excluded, in particular from other bacterial groups. Surprisingly, the brown algal ManC5-E and PL41 families seem to derived from distinct HGTs (different donor taxa). Thus, the catalytic steps for the synthesis and the degradation of alginate, may have arisen independently in *S. ischiensis* and brown algae, although within a narrow time-scale. While a global CAZome expansion is not observed in brown algae as compared to *S. ischiensis*, the PL41 family is the exception to this rule, with a greater number observed in brown algae as compared to its sister taxon (i.e. an average of 22 genes vs. 3). The situation is similar for the ManC5-E family which shows great expansion in brown algae (i.e. an average of 27 genes vs. 11). The acquisition of these enzymatic steps, and the expansion of the gene families in brown algae, probably allowed the generation of extensive ECMs with fine-tuned structures and rigidity. While there is no dynamic correlation between the number of CAZyme families and tissue complexity, these extensive ECMs may have represented an important prerequisite for the evolution of developmental complexity, and for the emergence of large, resilient substrate-anchored multicellular organisms, in the hostile coastal environment.

#### **The fucan biosynthetic pathway in brown algae needs to be fully resolved**

The set of genes involved in the synthesis of sulfated glycans remains unknown. Access to the large Phaeoexplorer genome dataset allowed us to identify promising candidate genes. Based on the large size of the GT23 family in all brown algal genomes, the specific phylogenetic position of the brown algal members of this family compared to biochemically characterized members, and information gained from modelling of GT23 protein structure, this family represents a candidate for fucan synthases in brown algae. This hypothesis is also supported by transcriptomic data obtained during the development of the *Fucus* embryo [73], where the expression of GT23 genes is up-regulated during early fucan deposition into the walls. Elongation of the oligosaccharidic moieties of FCSPs is believed to be followed by sulfation, catalysed by STs. The diversity of the brown algal ST orthologue groups may indicate a variety of acceptors, including many monosaccharides and their available hydroxyl positions found in FCSPs. In contrast to the GTs, we found homologues of some brown algal STs in echinoderms and these enzymes are known to synthesize

sulfated fucans of regular patterns [51]. While the origin of such homology might find its roots near the origin of eukaryotes (LECA), this would imply that the ability to perform fucan elongation and its subsequent sulfation, to have been lost in all eukaryotes, except brown algae and a narrow number of taxa within animals. The current available data does not identify taxa donors to favour a scenario involving HGT in the two groups of organisms. Functional convergence could also explain part of the similarity, yet this needs to be biochemically explored. STs belonging to orthologue group 10 are interesting candidates for the sulfation of fucans but other groups cannot be excluded. For example, orthogroup 12 contains a gene that is frequently associated with FK-GFFP in a syntenic region, and groups 2 and 8 contain genes that are also up-regulated during early development of the *Fucus* embryo [73]. These latter two groups contain orthologues from *S. ischiensis*. At this stage we cannot conclude that the GT23 family, the STs, the SAs and the fucan biosynthetic pathway were obtained independently in brown algae as compared to echinoderms. Functional validation of all of those genes is needed to progress further on this question.

In addition to brown algae and echinoderms, immunodetection procedures and fucose assays have provided evidence for the presence of fucans/fucoidans in the diatoms *Chaetoceros* and *Thalassiosira weissflogii* [74]. The composition of these secreted glycans in diatoms differ markedly from those of brown algal fucans/fucoidans [75]. The precise chemical structure of the diatom glycans has not been investigated in detail, the structures reported being essentially limited to glucuronomannans [76]. It is therefore possible that the anti-fucans/fucoidans antibodies cross-react with additional, yet to be described, glycan structures in these organisms. The BAM2 antibody, which recognises  $\alpha$ -(1 $\rightarrow$ 3)-fucans featuring 4-O-sulfate esters based on analysis using synthetic oligofucans [53], detects epitopes in diatoms, but precise structural validation of the compounds detected still needs to be carried out. In addition, glycan content has not been explored in detail in Stramenopiles other than brown algae and diatoms. A few studies have reported analyses of total monosaccharide composition in some of these organisms and these analyses either did not detect fucose (*Tribonema minus*, Xanthophyceae) [77], or only detected fucose at a limited level (*Nannochloropsis oceanica* Eustigmatophyceae) [78]. These results suggest that FCSPs are unlikely to be constitutively produced by these organisms at high levels. In short, both the lack of biochemical validations of any putative fucan synthase genes, and the fragmented knowledge of FCSP structures outside brown algae, impairs progression on tracking the evolutionary history of fucan synthesis in eukaryotes.

The situation of the sister taxon *S. ischiensis* is intriguing. Alginate has been detected immunologically in this species [22] and all the genes involved in alginate biosynthesis and degradation in brown algae are conserved. We have also identified GT23 members highly similar to their brown algal orthologues (~48% protein identity). The ability of *S. ischiensis* to produce sulfated fucans needs to be established. The outcome would contribute towards testing the hypothesis that the GT23 family contains fucan synthase genes. If future research establishes the occurrence of fucans in *S. ischiensis*, this would indicate an ECM with a glycan composition very similar to that of brown algae. One substantial difference between the Phaeophyceae and the Schizocladiophyceae is the absence of plasmodesmata in *S. ischiensis* [22, 79]. Plasmodesmata have been shown to be crucial in cell-to-cell signaling in land plants and may play a role in the evolution of tissue complexity [80–82]. Thus, it is not only the building up of a consistent and dynamic glycan-rich ECM in brown algae that may have contributed to the evolution of complex multicellular organisms, but also some structural features that function in close vicinity to ECMs such as plasmodesmata.

## Conclusion

Our meticulous comparative analyses of a vast genomic macroalgal resource, provides the first large scale view of CAZomes and ECM-polysaccharide metabolisms in brown algae. This work indicates that the gene families involved in both the synthesis and degradation of alginate are detected in brown algae and in their filamentous sister taxon *S. ischiensis*, and were duplicated in all brown algal groups. The expansion of these gene families, in addition to the occurrence of structural features such as plasmodesmata, may have represented an important prerequisite for the evolution of developmental complexity in brown algae. Our phylogenetic and protein modelling analyses also helped to refine glycosyltransferase candidate genes involved in fucan synthesis. However, only future work dedicated to both functional investigation of those genes, and filling the knowledge gap on FCSPs structures, will help to gain a comprehensive understanding of fucan metabolism in these organisms.

## Methods

### Identification of carbohydrate-related proteins

The CAZyme predicted proteins were identified by human curators of the CAZy database (<http://www.cazy.org/> [17]). Curation consisted in the expert validation of the predicted modularity for each protein, using a combination of BLASTp pairwise alignments against both characterized members and previously curated sequences integrated in the CAZy database, as well as HMMER v3.4 ([www.hmm.org](http://www.hmm.org); [83]) comparisons to in-house HMM



profiles of the families, and subfamilies when available. Proteins obtaining 100% coverage without gaps in pairwise alignments against a previously curated sequence were not manually inspected, allowing to decrease the load of human curation by processing genomes sequentially. The ST predicted proteins from various Stramenopiles were identified in the UniProt database based on the presence of a sulfotransferase-specific Pfam signature. Sequences of brown algal genomes not present in Uniprot were obtained by BLASTp searches in the Phaeoexplorer database (<https://phaeoexplorer.sb-roscoff.fr> [21]), using the sequences found in the UniProt database as query. All other genes were identified by BLASTp searches of brown algal enzymes in the Phaeoexplorer database, using characterized brown algal genes as the queries, when known (MPI, PMM, GMD, UGD, ManC5-E, DEHU reductase). The homology threshold was chosen as ~35% pairwise sequence identity. In addition to the identified brown algal sequences, representative and characterized members of the targeted families were retrieved from the CAZy database (for CAZymes) and from public databases.

#### Phylogenetic analyses

Protein sequences were aligned with the MAFFT program using the FFT-NS-i iterative refinement method (<https://mafft.cbrc.jp/alignment/server> [84]). In the case of multimodular proteins, the targeted catalytic domains were extracted prior to their alignment. The multiple alignments were curated manually using the Jalview program (<https://www.jalview.org> [85]), to retain only the informative sites. Prior to carrying out comprehensive phylogenetics, a preliminary analysis was carried out for each protein family to select the optimal model of sequence evolution using the IQ-TREE program (<http://iqtree.cibiv.univie.ac.at/>; [86]). Phylogenetic trees were subsequently constructed using the maximum likelihood approach with the RAxML program ([87], available at <https://usegalaxy.fr>). The reliability of the trees was tested with an MR-based Bootstopping criterion which generated resamplings of the datasets (range of 350 to 1000). The trees were edited in Mega 11.0.13 (<https://www.megasoftware.net/>; [88]) and graphically shaped with Adobe Illustrator CS6. For all phylogenies, only bootstrap values above 60% are shown. The brown algal sequences are indicated with a color code in relation to their taxonomy as indicated by the key legend. Sequences from the closest sister taxa *S. ischiensis* are further indicated by the red circle. In all instances, the green squares indicate sequences which have been biochemically characterized.

#### Heatmaps based on blast score ratios

To calculate the Blast Score Ratio (BSR), all pairs of sequences were compared using the NCBI BLASTp tool.

The BSR, which corresponds to the bitscore normalized to the length of the protein query, was then calculated from the BLAST result following the procedure described in [89]. Heat maps representing these BSR results and associated dendrograms were constructed using the ggheatmap function in the ggplot2 package in RStudio.

#### Extracted alignments and AlphaFold2 structural modeling

Some specific catalytic regions were extracted from alignments (Figs. 2C, 5B and 8C). Alignments of full-length proteins were made using the MAFFT program and edited in the ESPript 3.0 program (<https://esprict.ibcp.fr>; [90]). The color code of the amino acid residues indicates their biochemical character. Regions displaying high sequence similarity are boxed. The green squares indicate sequences which have been biochemically characterized. AlphaFold v2 [91] was used to predict 3D protein structures from primary amino acid sequences. The PyMOLv2.5.5 Molecular Graphics program (<http://www.pymol.org/>) was used to create and display the structural images.

#### Syntenic region analysis

Orthogroups, derived from predicted proteins of the 60 strains sequences in the Phaeoexplorer project (<https://phaeoexplorer.sb-roscoff.fr> [21]), were searched for genes of interest. Microsynteny analysis was carried out on a subset of species where the target orthologues resided on a singular chromosome/scaffold, using McscanX [92], following the pipeline described at <https://github.com/tanghaibao/jcvi/wiki/Mcscan>. Gene annotations were then processed using AGAT (<https://github.com/NBISweden/AGAT>) and LASTZ [93] tools (C-score cutoff=0.7) to infer syntenic blocks between species.

#### Abbreviations

BSR	BLASTp + Score Ratio
CAZyme	Carbohydrate active enzymes
ECM	Extracellular matrix
FCSP	Fucose-containing sulphated polysaccharide
FK	Fucokinase
FucT	Fucosyltransferase
GFPP	GDP-fucose pyrophosphorylase
GFS	GDP-L-fucose synthetase
GM46D	GDP-mannose 4,6-dehydratase
GH	Glycoside hydrolase
GT	Glycosyltransferase
HGT	Horizontal gene transfer
ManC5-E	Mannuronan C5-epimerase
MPI	Mannose-6-phosphate isomerase
MPG	Mannose-1-phosphate guanylyltransferase
PGM	Phosphoglucomutase
PL	Polysaccharide lyase
PMM	Phosphomannomutase
SA	Sulfatase
ST	Sulfotransferase
UGD	Uridine diphosphate-glucose dehydrogenases

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10811-3>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

## Acknowledgements

This work was supported by the ANR project BrownSugar (ANR-20-CE44-0011), the France Génomique National infrastructure project Phaeoexplorer (ANR-10-INBS-09) and the European Research Council (ID: 638240). We are grateful to the Roscoff Bioinformatics platform ABiMS (<http://abims.sb-roscoff.fr>), part of the Institut Français de Bioinformatique (ANR-11-INBS-0013) and BioGenouest network, for providing computing and storage resources.

## Author contributions

Author contributions: Software: L.B.G., E.C.; Formal analysis: F.D., O.G. and J.M.C.; Investigation: L.M., N.T., A. L. and C.H.; Resources: E.C., J.M.C. and N.T.; Data Curation: T.B. and N.T.; Writing—Original Draft: L.M., A. B.-C. and C.H.; Writing—Review & Editing: all authors; Visualization: L.M., A.B.-C., M.C., A.L. and C.H.; Supervision: C.H.; Project administration: F.D., J.M.C. and C.H.; Funding acquisition: J.M.C. and C.H.; Conceptualization: C.H.

## Funding

This work was supported by to ANR project BrownSugar (ANR-20-CE44-0011), the France Génomique National infrastructure project Phaeoexplorer (ANR-10-INBS-09) and the European Research Council (ID: 638240).

## Data availability

The raw datasets which have been analysed in this current study are publicly available at <https://phaeoexplorer.sb-roscoff.fr>. The corresponding publication of this project is cited within the manuscript.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

## Author details

- <sup>1</sup>Integrative Biology of Marine Models Laboratory, Sorbonne Université, CNRS, Station Biologique de Roscoff, Roscoff, France
- <sup>2</sup>Génomique Métabolique, Institut François Jacob, CEA, CNRS, Université Evry, Université Paris-Saclay, Genoscope, Evry 91057, France
- <sup>3</sup>CNRS, Sorbonne Université, FR2424, ABiMS-IFB, Station Biologique, Roscoff, France
- <sup>4</sup>Department of Algal Development and Evolution, Max Planck Institute for Biology, 72076 Tübingen, Germany
- <sup>5</sup>Aix Marseille Univ, CNRS, UMR 7257 AFMB, Marseille, France
- <sup>6</sup>INRAE, USC 1408 AFMB, Marseille, France
- <sup>7</sup>INRAE, Aix-Marseille Univ, UMR1163 BBF, Marseille, France
- <sup>8</sup>Department of Biotechnology and Biomedicine, Technical University of Denmark, Kgs. Lyngby, Denmark

Received: 7 June 2024 / Accepted: 18 September 2024

Published online: 10 October 2024

## References

1. Bringloe TT, Starko S, Wade RM, Vieira C, Kawai H, De Clerck O, Cock JM, Coelho SM, Destombe C, Valero M, et al. Phylogeny and evolution of the Brown Algae. *CRC Crit Rev Plant Sci.* 2020;39(4):281–321.
2. Farghali M, Mohamed IMA, Osman AI, Rooney DW. Seaweed for climate mitigation, wastewater treatment, bioenergy, bioplastic, biochar, food, pharmaceuticals, and cosmetics: a review. *Environ Chem Lett* 2022.
3. Pessarrodona A, Assis J, Filbee-Dexter K, Burrows MT, Gattuso J-P, Duarte CM, Krause-Jensen D, Moore PJ, Smale DA, Wernberg T. Global seaweed productivity. *Sci Adv.* 2022;8(37):eabn2465.
4. Filbee-Dexter K, Feehan CJ, Smale DA, Krumhansl KA, Augustine S, de Bettignies F, Burrows MT, Byrnes JEK, Campbell J, Davoult D, et al. Kelp carbon sink potential decreases with warming due to accelerating decomposition. *PLoS Biol.* 2022;20(8):e3001702.
5. Arriaga O, Wawrzynkowski P, Ibáñez H, Muguerza N, Díez I, Pérez-Ruzafa I, Gorostiaga JM, Quintano E, Becerro MA. Short-term response of macroalgal communities to ocean warming in the Southern Bay of Biscay. *Mar Environ Res.* 2023;190:106098.
6. Popper ZA, Michel G, Hervé C, Domozych DS, Willats WGT, Tuohy MG, Kloareg B, Stengel DB. Evolution and diversity of plant cell walls: from algae to flowering plants. *Annu Rev Plant Biol.* 2011;62(1):567–90.
7. Kloareg B, Badis Y, Cock JM, Michel G. Role and evolution of the extracellular matrix in the acquisition of complex multicellularity in eukaryotes: a macroalgal perspective. *Genes (Basel)* 2021, 12(7).
8. Blaby-Haas CE, Merchant SS. Comparative and functional Algal Genomics. *Annu Rev Plant Biol.* 2019;70(1):605–38.
9. Deniaud-Bouët E, Hardouin K, Potin P, Kloareg B, Hervé C. A review about brown algal cell walls and fucose-containing sulfated polysaccharides: cell wall context, biomedical properties and key research challenges. *Carbohydr Polym.* 2017;175:395–408.
10. Deniaud-Bouët E, Kervarec N, Michel G, Tonon T, Kloareg B, Hervé C. Chemical and enzymatic fractionation of cell walls from Fucales: insights into the structure of the extracellular matrix of brown algae. *Ann Botany.* 2014;114(6):1203–16.
11. Birgersson PS, Oftebro M, Strand WI, Aarstad OA, Sætrom GI, Sletta H, Arlov Ø, Aachmann FL. Sequential extraction and fractionation of four polysaccharides from cultivated brown algae *Saccharina Latissima* and *Alaria esculenta*. *Algal Res.* 2023;69:102928.
12. Lorbeer AJ, Lahnstein J, Bulone V, Nguyen T, Zhang W. Multiple-response optimization of the acidic treatment of the brown alga *Ecklonia radiata* for the sequential extraction of fucoidan and alginate. *Bioresour Technol.* 2015;197:302–9.
13. Nguyen TT, Mikkelsen MD, Tran VHN, Trang VTD, Rhein-Knudsen N, Holck J, Rasin AB, Cao HTT, Van TTT, Meyer AS. Enzyme-assisted fucoidan extraction from Brown Macroalgae *Fucus distichus* subsp. *evanescens* and *Saccharina Latissima*. *Mar Drugs* 2020, 18(6).
14. Michel G, Tonon T, Scornet D, Cock JM, Kloareg B. The cell wall polysaccharide metabolism of the brown alga *Ectocarpus siliculosus*. Insights into the evolution of extracellular matrix polysaccharides in eukaryotes. *New Phytol.* 2010;188(1):82–97.
15. Mazéas L, Yonamine R, Barbeyron T, Henrissat B, Drula E, Terrapon N, Nagasato C, Hervé C. Assembly and synthesis of the extracellular matrix in brown algae. *Semin Cell Dev Biol.* 2023;134:112–24.
16. Chi S, Liu T, Wang X, Wang R, Wang S, Wang G, Shan G, Liu C. Functional genomics analysis reveals the biosynthesis pathways of important cellular components (alginate and fucoidan) of *Saccharina*. *Curr Genet.* 2018;64(1):259–73.
17. Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res.* 2022;50(D1):D571–7.
18. Carpita NC, McCann MC. The Cell Wall. In: Buchanan B, Gruissem W, Jones R, editors. *Biochemistry and Molecular Biology of plants*. Md.: American Society of Plant Physiologists; 2000. Rockville.
19. Stam M, Lelièvre P, Hoebeke M, Corre E, Barbeyron T, Michel G. SulfAtlas, the sulfatase database: state of the art and new developments. *Nucleic Acids Res.* 2023;51(D1):D647–53.
20. Barbeyron T, Brillet-Guéguen L, Carré W, Carrière C, Caron C, Czjzek M, Hoebeke M, Michel G. Matching the diversity of sulfated biomolecules: creation of a classification database for sulfatases reflecting their substrate specificity. *PLoS ONE.* 2016;11(10):e0164846–0164846.
21. Denoëud F, Godfroy O, Cruaud C, Heesch S, Nehr Z, Tadrent N, Couloux A, Brillet-Guéguen L, Delage L, Mckeown D et al. Evolutionary genomics of the

- emergence of brown algae as key components of coastal ecosystems. *bioRxiv* 2024:2024.2002.2019.579948.
22. Kawai H, Maebara S, Sasaki H, Okuda K, Henry EC. *Schizocladia ischiensis*: a New Filamentous Marine Chromophyte belonging to a New Class, Schizocladophyceae. *Protist*. 2003;154(2):211–28.
  23. Geshi N, Johansen JN, Diloekpimol A, Rolland A, Belcram K, Verger S, Kotake T, Tsumuraya Y, Kaneko S, Tryfona T, et al. A galactosyltransferase acting on arabinogalactan protein glycans is essential for embryo development in *Arabidopsis*. *Plant J*. 2013;76(1):128–37.
  24. Zhang L, Prabhakar PK, Bharadwaj VS, Bomble YJ, Peña MJ, Urbanowicz BR. Glycosyltransferase family 47 (GT47) proteins in plants and animals. *Essays Biochem*. 2023;67(3):639–52.
  25. Boruah BM, Kadirvelraj R, Liu L, Ramiah A, Li C, Zong G, Bosman GP, Yang JY, Wang LX, Boons GJ, et al. Characterizing human  $\alpha$ -1,6-fucosyltransferase (FUT8) substrate specificity and structural similarities with related fucosyltransferases. *J Biol Chem*. 2020;295(50):17027–45.
  26. García-García A, Ceballos-Laita L, Serna S, Artschwager R, Reichardt NC, Corzana F, Hurtado-Guerrero R. Structural basis for substrate specificity and catalysis of  $\alpha$ 1,6-fucosyltransferase. *Nat Commun*. 2020;11(1):973.
  27. Bamford NC, Le Mauff F, Subramanian AS, Yip P, Millán C, Zhang Y, Zacharias C, Forman A, Nitz M, Codée JDC, et al. Ega3 from the fungal pathogen *aspergillus fumigatus* is an endo- $\alpha$ -1,4-galactosaminidase that disrupts microbial biofilms. *J Biol Chem*. 2019;294(37):13833–49.
  28. Tamura J-I, Hasegawa K, Kadowaki K, Igarashi Y, Kodama T. Molecular cloning and sequence analysis of the gene encoding an endo- $\alpha$ -1,4-polygalactosaminidase of *Pseudomonas* sp. 881. *J Ferment Bioeng*. 1995;80(4):305–10.
  29. Inoue A, Ojima T. Functional identification of alginate lyase from the brown alga *Saccharina Japonica*. *Sci Rep*. 2019;9(1):4937.
  30. Zhang P, Shao Z, Li L, Liu S, Yao J, Duan D. Molecular characterisation and biochemical properties of phosphomannomutase/phosphoglucomutase (PMM/PGM) in the brown seaweed *Saccharina Japonica*. *J Appl Phycol*. 2018;30(4):2687–2687.
  31. Tenhaken R, Voglas E, Cock JM, Neu V, Huber CG. Characterization of GDP-mannose dehydrogenase from the brown alga *Ectocarpus siliculosus* providing the precursor for the alginate polymer. *J Biol Chem*. 2011;286(19):16707–15.
  32. Zhang P, Shao Z, Jin W, Duan D. Comparative characterization of two GDP-mannose dehydrogenase genes from *Saccharina Japonica* (Laminariales, Phaeophyceae). *BMC Plant Biol*. 2016;16(1):62.
  33. Fischl R, Bertelsen K, Gaillard F, Coelho S, Michel G, Klingner M, Boyen C, Czjzek M, Hervé C. The cell-wall active mannanuronic C5-epimerases in the model brown alga *Ectocarpus*: from gene context to recombinant protein. *Glycobiology*. 2016;26(9):973–83.
  34. Inoue A, Satoh A, Morishita M, Tokunaga Y, Miyakawa T, Tanokura M, Ojima T. Functional heterologous expression and characterization of mannanuronic C5-epimerase from the brown alga *Saccharina Japonica*. *Algal Res*. 2016;16:282–91.
  35. Murray RG, Schleifer KH. Taxonomic notes: a proposal for recording the properties of putative taxa of prokaryotes. *Int J Syst Bacteriol*. 1994;44(1):174–6.
  36. Inoue A, Ojima T. Functional identification of the 4-deoxy-L-erythro-5-hexoseulose uronate reductase from a brown alga, *Saccharina Japonica*. *Biochem Biophys Res Commun*. 2021;545:112–8.
  37. Penning TM. The aldo-keto reductases (AKRs): overview. *Chemico-Biol Interact*. 2015;234:236–46.
  38. Cartmell A, Lowe EC, Baslé A, Firbank SJ, Ndeh DA, Murray H, Terrapon N, Lombard V, Henrissat B, Turnbull JE, et al. How members of the human gut microbiota overcome the sulfation problem posed by glycosaminoglycans. *Proc Natl Acad Sci U S A*. 2017;114(27):7037–42.
  39. Hehemann J-H, Boraston AB, Czjzek M. A sweet new wave: structures and mechanisms of enzymes that digest polysaccharides from marine algae. *Curr Opin Struct Biol*. 2014;28:77–86.
  40. Nishitsuji K, Arimoto A, Higa Y, Mekarum M, Kawamitsu M, Satoh N, Shoguchi E. Draft genome of the brown alga, *Nemacystus decipiens*, Onna-1 strain: Fusion of genes involved in the sulfated fucan biosynthesis pathway. *Sci Rep*. 2019;9(1):4607–4607.
  41. Both P, Sobczak L, Breton C, Hann S, Nöbauer K, Paschinger K, Kozmon S, Mucha J, Wilson IBH. Distantly related plant and nematode core  $\alpha$ 1,3-fucosyltransferases display similar trends in structure–function relationships. *Glycobiology*. 2011;21(11):1401–15.
  42. Brito C, Kandzia S, Graça T, Conradt HS, Costa J. Human fucosyltransferase IX: specificity towards N-linked glycoproteins and relevance of the cytoplasmic domain in intra-golgi localization. *Biochimie*. 2008;90(9):1279–90.
  43. Ma B, Simala-Grant JL, Taylor DE. Fucosylation in prokaryotes and eukaryotes. *Glycobiology*. 2006;16(12):r158–84.
  44. Sun HY, Lin SW, Ko TP, Pan JF, Liu CL, Lin CN, Wang AH, Lin CH. Structure and mechanism of *Helicobacter pylori* fucosyltransferase. A basis for lipopolysaccharide variation and inhibitor design. *J Biol Chem*. 2007;282(13):9973–82.
  45. Breton C, Oriol R, Imberty A. Conserved structural features in eukaryotic and prokaryotic fucosyltransferases. *Glycobiology*. 1998;8(1):87–94.
  46. Oulmouden A, Wierinckx A, Petit J-M, Costache M, Palcic MM, Mollicone R, Oriol R, Julien R. Molecular Cloning and expression of a bovine  $\alpha$ (1,3)-Fucosyltransferase gene homologous to a putative ancestor gene of the human FUT3-FUT5-FUT6 Cluster\*. *J Biol Chem*. 1997;272(13):8764–73.
  47. Kumar S, Wang Y, Zhou Y, Dillard L, Li F-W, Sciandra CA, Sui N, Zentella R, Zahn E, Shabanowitz J, et al. Structure and dynamics of the *Arabidopsis* O-fucosyltransferase SPINDLY. *Nat Commun*. 2023;14:1538.
  48. van Der Wel H, Morris HR, Panico M, Paxton T, North SJ, Dell A, Thomson JM, West CM. A non-golgi  $\alpha$ 1,2-fucosyltransferase that modifies Skp1 in the cytoplasm of *Dictyostelium*. *J Biol Chem*. 2001;276(36):33952–63.
  49. Yanagidani S, Uozumi N, Ihara Y, Miyoshi E, Yamaguchi N, Taniguchi N. Purification and cDNA cloning of GDP-L-Fuc:N-acetyl-beta-D-glucosaminide:alpha1-6 fucosyltransferase (alpha1-6 FucT) from human gastric cancer MKN45 cells. *J Biochem*. 1997;121(3):626–32.
  50. Tomida S, Takata M, Hirata T, Nagae M, Nakano M, Kizuka Y. The SH3 domain in the fucosyltransferase FUT8 controls FUT8 activity and localization and is essential for core fucosylation. *J Biol Chem*. 2020;295(23):7992–8004.
  51. Pomin VH, Mourão PAS. Structure, biology, evolution, and medical importance of sulfated fucans and galactans. *Glycobiology*. 2008;18(12):1016–27.
  52. Pereira MS, Mulloy B, Mourão PAS. Structure and anticoagulant activity of sulfated fucans: comparison between the regular, repetitive, and linear fucans from echinoderms with the more heterogeneous and branched polymers from brown algae. *J Biol Chem*. 1999;274(12):7656–67.
  53. Crawford CJ, Schultz-Johansen M, Luong P, Vidal-Melgosa S, Hehemann J-H, Seeberger PH. Automated synthesis of Algal Fucoidan Oligosaccharides. *J Am Chem Soc*. 2024;146(27):18320–30.
  54. Liu T, Wang X, Wang G, Jia S, Liu G, Shan G, Chi S, Zhang J, Yu Y, Xue T, et al. Evolution of complex thallus alga: genome sequencing of *Saccharina Japonica*. *Front Genet*. 2019;10:378.
  55. Ye N, Zhang X, Miao M, Fan X, Zheng Y, Xu D, Wang J, Zhou L, Wang D, Gao Y et al. *Saccharina* genomes provide novel insight into kelp biology. *Nat Commun* 2015, 6.
  56. Salmeán AA, Duffieux D, Harholt J, Qin F, Michel G, Czjzek M, Willats WGT, Hervé C. Insoluble (1  $\rightarrow$  3), (1  $\rightarrow$  4)- $\beta$ -D-glucan is a component of cell walls in brown algae (Phaeophyceae) and is masked by alginates in tissues. *Sci Rep*. 2017;7(1):2880.
  57. Kolbinger F, Streiff MB, Katopodis AG. Cloning of a human UDP-galactose-2-acetamido-2-deoxy-D-glucose 3beta-galactosyltransferase catalyzing the formation of type 1 chains. *J Biol Chem*. 1998;273(1):433–40.
  58. Hervé C, Siméon A, Jam M, Cassin A, Johnson KL, Salmeán AA, Willats WGT, Doblin MS, Bacic A, Kloareg B. 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*. 2016;209(4):1428–41.
  59. Tonon T, Rousvoal S, Roeder V, Boyen C. Expression profiling of the mannanuronic C5-epimerase multigenic family in the brown alga *Laminaria digitata* (Phaeophyceae) under biotic stress conditions. *J Phycol*. 2008;44(5):1250–6.
  60. Nyvall P, Corre E, Boisset C, Barbeyron T, Rousvoal S, Scornet D, Kloareg B, Boyen C. Characterization of mannanuronic C-5-epimerase genes from the brown alga *Laminaria digitata*. *Plant Physiol*. 2003;133(2):726–35.
  61. Sénéchal F, Wattier C, Rustérucci C, Pelloux J. Homogalacturonan-modifying enzymes: structure, expression, and roles in plants. *J Exp Bot*. 2014;65(18):5125–60.
  62. Gawin A, Tietze L, Aarstad OA, Aachmann FL, Brautaset T, Ertesvåg H. Functional characterization of three *Azotobacter chroococcum* alginate-modifying enzymes related to the *Azotobacter vinelandii* AlgE Mannuronan C-5-epimerase family. *Sci Rep*. 2020;10(1):12470.
  63. Linker A, Jones RS. A new polysaccharide resembling alginic acid isolated from pseudomonads. *J Biol Chem*. 1966;241(16):3845–51.
  64. Bjerkan TM, Lillehov BE, Strand WI, Skjåk-Braek G, Valla S, Ertesvåg H. Construction and analyses of hybrid *Azotobacter vinelandii* mannanuronic C-5 epimerases with new epimerization pattern characteristics. *Biochem J*. 2004;381(Pt 3):813–21.
  65. Ertesvåg H, Erlien F, Skjåk-Braek G, Rehm BH, Valla S. Biochemical properties and substrate specificities of a recombinantly produced *Azotobacter vinelandii* alginate lyase. *J Bacteriol*. 1998;180(15):3779–84.

66. Schiller NL, Monday SR, Boyd CM, Keen NT, Ohman DE. Characterization of the *Pseudomonas aeruginosa* alginate lyase gene (aglL): cloning, sequencing, and expression in *Escherichia coli*. *J Bacteriol*. 1993;175(15):4780–9.
67. Gimmestad M, Ertesvåg H, Heggeset TM, Aarstad O, Svanem BI, Valla S. Characterization of three new *Azotobacter vinelandii* alginate lyases, one of which is involved in cyst germination. *J Bacteriol*. 2009;191(15):4845–53.
68. Yamasaki M, Moriwaki S, Miyake O, Hashimoto W, Murata K, Mikami B. Structure and function of a hypothetical *Pseudomonas aeruginosa* protein PA1167 classified into family PL-7: a NOVEL ALGINATE LYASE WITH a  $\beta$ -SANDWICH FOLD\*. *J Biol Chem*. 2004;279(30):31863–72.
69. Jain S, Ohman DE. Role of an alginate lyase for alginate transport in mucoid *Pseudomonas aeruginosa*. *Infect Immun*. 2005;73(10):6429–36.
70. May TB, Chakrabarty AM. *Pseudomonas aeruginosa*: genes and enzymes of alginate synthesis. *Trends Microbiol*. 1994;2(5):151–7.
71. Gheorghita AA, Wolfram F, Whitfield GB, Jacobs HM, Pfoh R, Wong SSY, Guitor AK, Goodyear MC, Berezuk AM, Khursigara CM, et al. The *Pseudomonas aeruginosa* homeostasis enzyme AglL clears the periplasmic space of accumulated alginate during polymer biosynthesis. *J Biol Chem*. 2022;298(2):101560.
72. Thomas F, Barbeyron T, Tonon T, Génicot S, Czjzek M, Michel G. Characterization of the first alginolytic operons in a marine bacterium: from their emergence in marine Flavobacteria to their independent transfers to marine Proteobacteria and human gut Bacteroides. *Environ Microbiol*. 2012;14:2379–94.
73. Linardić M, Cokus SJ, Pellegrini M, Braybrook SA. Growth of the *Fucus* embryo: insights into wall-mediated cell expansion through mechanics and transcriptomics. *bioRxiv* 2020:2020.2001.2029.925107.
74. Vidal-Melgosa S, Sichert A, Francis TB, Bartosik D, Niggemann J, Wichels A, Wilts WGT, Fuchs BM, Teeling H, Becher D, et al. Diatom fucan polysaccharide precipitates carbon during algal blooms. *Nat Commun*. 2021;12(1):1150.
75. Gugi B, Le Costaouec T, Burel C, Lerouge P, Helbert W, Bardor M. Diatom-Specific Oligosaccharide and Polysaccharide structures help to unravel Biosynthetic capabilities in Diatoms. *Mar Drugs*. 2015;13(9):5993–6018.
76. Le Costaouec T, Unamunzaga C, Mantecon L, Helbert W. New structural insights into the cell-wall polysaccharide of the diatom *Phaeodactylum tricornutum*. *Algal Res*. 2017;26:172–9.
77. Huo S, Wang H, Chen J, Hu X, Zan X, Zhang C, Qian J, Zhu F, Ma H, Elshobary M. A preliminary study on polysaccharide extraction, purification, and antioxidant properties of sugar-rich filamentous microalgae *Tribonema minus*. *J Appl Phycol*. 2022;34(6):2755–67.
78. Bernaerts TMM, Gheysen L, Kyomugasho C, Jamsazzadeh Kermani Z, Vandionant S, Foubert I, Hendrickx ME, Van Loey AM. Comparison of microalgal biomasses as functional food ingredients: focus on the composition of cell wall related polysaccharides. *Algal Res*. 2018;32:150–61.
79. Terauchi M, Nagasato C, Kajimura N, Mineyuki Y, Okuda K, Katsaros C, Motomura T. Ultrastructural study of plasmodesmata in the brown alga *Dictyota dichotoma* (Dictyotales, Phaeophyceae). *Planta*. 2012;236(4):1013–26.
80. Sager RE, Lee JY. Plasmodesmata at a glance. *J Cell Sci* 2018, 131(11).
81. German L, Yeshvekar R, Benitez-Alfonso Y. Callose metabolism and the regulation of cell walls and plasmodesmata during plant mutualistic and pathogenic interactions. *Plant Cell Environ*. 2023;46(2):391–404.
82. Cook ME, Graham LE, Botha CEJ, Lavin CA. Comparative ultrastructure of plasmodesmata of *Chara* and selected bryophytes: toward an elucidation of the evolutionary origin of plant plasmodesmata. *Am J Bot*. 1997;84(9):1169–78.
83. Eddy SR. Accelerated Profile HMM searches. *PLoS Comput Biol*. 2011;7(10):e1002195.
84. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*. 2017;20(4):1160–6.
85. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25(9):1189–91.
86. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14(6):587–9.
87. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312–3.
88. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol*. 2021;38(7):3022–7.
89. Hobbs EEM, Gloster TM, Pritchard L. cazy\_webscraper: local compilation and interrogation of comprehensive CAZyme datasets. *Microb Genom* 2023, 9(8).
90. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*. 2014;42(W1):W320–4.
91. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583–9.
92. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and colinearity in plant genomes. *Science*. 2008;320(5875):486–8.
93. Harris R. Improved Pairwise Alignment of Genomic DNA. *Pennsylvania State University* 2007, PhD Thesis.

## Publisher's note

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