Marine Genomics 32 (2017) 49–59

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Genome-wide computational analysis of the secretome of brown algae (Phaeophyceae)

Makoto Terauchi [a,*], Takahiro Yamagishi [b,1], Takeaki Hanyuda [b], Hiroshi Kawai [b]

[a] Organization for Advanced and Integrated Research, Kobe University, Kobe 657–8501, Japan
[b] Research Center for Inland Seas, Kobe University, Kobe 657–8501, Japan

ARTICLE INFO

Article history:
Received 6 September 2016
Received in revised form 6 December 2016
Accepted 16 December 2016
Available online 5 January 2017

Keywords:
Alginate
Carbohydrate binding module (CBM)
Extracellular matrix (ECM)
Laminariales
Metalloproteinase (Metzincin)
RNA-seq

ABSTRACT

Brown algae have evolved complex multicellularity in the heterokont lineage. They are phylogenetically distant to land plants, fungi and animals. Especially, the members of Laminariales (so-called kelps) have developed highly differentiated tissues. Extracellular matrix (ECM) plays pivotal roles in a number of essential processes in multicellular organisms, such as cell adhesion, cell and tissue differentiations, cell-to-cell communication, and responses to environmental stimuli. In these processes, a set of extracellular secreted proteins called the secretome operates remodeling of the physicochemical nature of ECM and signal transduction by interacting with cell surface proteins and signaling molecules. Characterization of the secretome is a critical step to clarify the contributions of ECM to the multicellularity of brown algae. However, the identity of the brown algal secretome has been poorly understood. In order to reveal the repertory of the brown algal secretome and its involvement in the evolution of Laminariales, we conducted a genome-wide analysis of the brown algal secretome utilizing the published complete genome data of *Ectocarpus siliculosus* and *Saccharina japonica* as well as newly obtained RNA-seq data of seven laminarialean species (*Agarum clathratum*, *Alaria crassifolia*, *Aureophycus aleuticus*, *Costaria costata*, *Pseudochorda nagaii*, *Saccharina angustata* and *Undaria pinnatifida*) largely covering the laminarialean families. We established the *in silico* pipeline to systematically and accurately detect the secretome by combining multiple prediction algorithms for the N-terminal signal peptide and transmembrane domain within the protein sequence. From 16,189 proteins of *E. siliculosus* and 18,733 proteins of *S. japonica*, 552 and 964 proteins respectively were predicted to be classified as the secretome. Conserved domain analysis showed that the domain repertory were very similar to each other, and that of the brown algal secretome was partially common with that of the secretome of other multicellular organisms (land plants, fungi and animals). In the laminarialean species, it was estimated that the gene abundance and the domain architecture of putative ECM remodeling-related proteins were altered compared with those of *E. siliculosus*, and that the alteration started from the basal group of Laminariales. These results suggested that brown algae have developed their own secretome, and its functions became more elaborated in the more derived members in Laminariales.

1. Introduction

Brown algae (Phaeophyceae) are photosynthetic eukaryotes dominant in many coastal ecosystems, phylogenetically distant to animals, fungi, and land plants. They have evolved complex, multicellular thalli consisted of branched uniseriate and multiseriate filaments, and parenchymatous tissues (Bold and Wynne, 1985; Graham and Wilcox, 2000). The order Laminariales (so-called kelps) is one of the most derived lineages in the brown algae, with elaborate thalli and heteromorphic life histories (Kawai et al., 2015; Silberfeld et al., 2010). Laminariales consists of eight families considered basal (Pseudochordaceae, Akkesiphycaceae and Chordaceae) or derived (Aureophycaceae, Alariaceae, Agaraceae, Lessoniaceae and Laminariaceae) (Kawai et al., 2013; Kawai, 2014). The members of the derived families have characteristic features such as differentiation between blade and stipe, occurrence of trumpet-shaped hyphae, pneumatocysts (gas bladders), mucilaginous organs (mucilaginous cells, mucilaginous gland cells and ducts), and meristematic haptera (Bold and Wynne, 1985; Schmitz, 1990). Akkesiphycaceae and Pseudochordaceae lack most of the above-mentioned morphological features (Kawai, 1986; Kawai and Kurogi, 1985), while Chordaceae has trumpet-shaped hyphae (Kawai et al., 2001). On the other hand, even within the derived families, Aureophycaceae, which is most basal in the derived families, and Agaraceae lack pneumatocysts and mucilaginous organs (Kawai et al., 2013), and there appears to be some gradient in morphological complexity correlated with evolution.

* Corresponding author at: Organization for Advanced and Integrated Research, Kobe 657–8501, Japan.
E-mail address: terauchi1117@gmail.com (M. Terauchi).
[1] Present address: Center for Health and Environmental Risk Research, National Institute for Environmental Studies, Onogawa, Tsukuba, Ibaraki 305-8506, Japan.

http://dx.doi.org/10.1016/j.margen.2016.12.002
1874-7787/

In multicellular organisms, extracellular matrix (ECM) is associated with essential roles in providing structural support, cell and tissue differentiation, cell-to-cell communication, and responses to stimuli from the outer environment (Hynes, 2009). ECM is a composite material of macromolecules (proteins and polysaccharides) with specific physicochemical properties. For example, the ECM of animals contains collagen and glycosaminoglycan (Frantz et al., 2010; Hynes, 2009). The brown algal ECM is composed of alginate, fucose-containing sulfated polysaccharides and cellulose (Kloareg and Quatrano, 1988; Deniaud-Bouët et al., 2014; Terauchi et al., 2016). The alginate-containing ECM has been found in few lineages, only brown algae and their phylogenetically related groups, and bacteria (Michel et al., 2010; Popper et al., 2011). The unique brown algal ECM is considered to play a significant role in the evolution of multicellularity in the lineage.

The physicochemical properties of the ECM are spatiotemporally regulated by a set of ECM proteins called the secretome. The secretome has been defined as the proteins secreted into the extracellular space by any cells and developmental stages (Agrawal et al., 2010). In animals, matrix metalloproteinases degrade ECM proteins involved in a variety of developmental processes (Lu et al., 2011). In land plants, expansins loosen the interaction between cellulose microfibrils and hemicelluloses during the cell wall expansion (Sampedro and Cosgrove, 2005). Identifying the diversities of the secretomes in certain organisms is a crucial step for understanding the function and evolution of ECM in these lineages. The secretome can be systematically predicted by searching the subcellular protein localization motifs using prediction algorithms (Lum and Min, 2011a,b; Lum et al., 2014; Meinken et al., 2014, 2015; Min, 2010). In animals and land plants, the repertory of secretomes and their evolutional origins have been predicted from complete genome data (Agrawal et al., 2010; Lum and Min, 2011a; Özbek et al., 2010). The analysis of the secretome in unicellular relatives of animals, choanoflagellates, proposed the scenario of the evolutionary origin of the animal ECM proteins (Williams et al., 2014). In oomycetes, the comparative secretome analysis detected the expansion of the secretome by the horizontal gene transfer, and the functional alteration of the secretome depending on their lifestyle (Misner et al., 2014). We expected that the secretome could provide insight into the contributions of the ECM to the evolution of brown algae, especially Laminariales. However, the identity and function of the brown algal secretome has been poorly characterized.

In brown algae, complete genome data have been published in three species, the uniseriate filamentous species *Ectocarpus siliculosus* (Cock et al., 2010) the kelp species *Saccharina japonica* (Ye et al., 2015) and Cladosiphon okamuranus (Nishitsuji et al., 2016). In order to uncover the identity of the brown algal secretome and its correlation with the evolution of Laminariales, we conducted a genome-wide prediction of the brown algal secretome using the complete genome data of E. siliculosus and S. japonica and the newly obtained RNA-seq data of 7 laminarialean species of varied morphological complexity from five families (*Pseudochorda nagaii* of the basal family, and *Agarum clathratum*, *Alaria crassifolia*, *Aureophycus aleuticus*, *Costaria costata*, *Saccharina angustata* and *Undaria pinnatifida* of derived families).

2. Materials and methods

2.1. Sample collection

Field samples of six laminarialean species were collected for RNA-seq. Mature sporophytes of *Pseudochorda nagaii* (Pseudochordaceae) were collected at Hanasaki, Hokkaido, Japan (43°17′N, 145°36′E) in August 2014. Clean juvenile sporophytes of *Alaria crassifolia, Undaria pinnatifida* (Alariaceae), *Agarum clathratum, Costaria costata* (Agaraceae) and *Saccharina angustata* (Laminariaceae) were collected in February (*A. crassifolia, U. pinnatifida, C. costata* and *S. augustata)* and June (*A. clathratum*), 2015 on the coast at Charatsunai, Muroran, Hokkaido (42°19′N, 140°59′E).

2.2. Culture

Unialgal gametophytes of *Aureophycus aleuticus* (Aureophycaceae, KU-3181), housed in the Kobe University Macro-Algal Culture Collection (KU-MACC) were cultured in PESI medium (Tatewaki, 1966) at 5 °C in long day conditions (16: 8 h light: dark) under cool-white-type fluorescent illumination of approximately 30 μmol photons m^{-2} s^{-1}.

2.3. RNA-seq

Total RNA was extracted from the six field-collected and one culture samples as described above. Total RNA was extracted in a combined protocol following Le Bail et al. (2008) and that in the RNA extraction kit instructions. The extraction was performed with a cetyltrimethylammonium bromide (CTAB) extraction buffer and purified with ethanol and chloroform. After precipitation with LiCl, RNA was further purified with an RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). Contaminating DNA was removed with an RNase-Free DNase Set (Qiagen), according to the manufacturer's instructions. The quality of extracted total RNA was assessed using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, *CA*, USA).

cDNA library preparation, sequencing of the library and *de novo* assembly were performed at the Beijing Genomic Institute (BGI, Shenzhen, China). mRNA was isolated using magnetic beads with Oligo (dT) for the six species except *A. clathratum,* or by removing rRNA from the total RNA for *A. clathratum*. Then the mRNA was fragmented into short fragments and cDNA was synthesized using the mRNA fragments as templates. Short fragments were purified and connected with adapters, then amplified by PCR. The resultant paired-end library (2 × 100 bp reads) was sequenced using an Illumina HiSeq 2000.

The raw reads (60,000,000 on average) were filtered using the filter_fq software of BGI to obtain high quality "clean" reads (52,000,000 on average) by discarding the reads with adaptor contamination, those with unknown nucleotides comprising >5%, and those with the low quality (the rate of reads with quality value ≤10 was >20%). The clean reads were assembled using Trinity software (Grabherr et al., 2011). For each sample (species), around 60,000 transcripts were obtained. Potential contaminating bacterial sequences (<1% on average) were removed by BLASTn search using the local BLAST+ software package from NCBI (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/). The open reading frames (ORFs) of at least 70 amino acids within the final transcripts were selected using the Transdecoder software (http://transdecoder.github.io/). The clean reads and the assembled transcripts were deposited in DDBJ Sequence Read Archive (Accession: DRA005083) and DDBJ Transcriptome Shotgun Assembly database. The detailed RNA-seq statistics including the accession number for each dataset are in Table 6.

2.4. Prediction of the secretome

The secretome was predicted from the proteome of *E. siliculosus* (16,189 proteins from the Uniprot database after excluding 145 chloroplast genome-encoded proteins) (Cock et al., 2010) and *S. japonica* (18,733 proteins) (Ye et al., 2015). The proteins were categorized into four types according to the presence/absence of an N-terminal signal peptide (SP) (von Heijne, 1990) and a transmembrane domain (TM) (Wallin and von Heijne, 1998): Type 0 (SP absent and TM absent), Type 1 (SP present and TM absent), Type 2 (SP absent and TM present) and Type 3 (SP present and TM present). Combining prediction algorithms for SP and TM has been shown to considerably improve the prediction accuracy for the secretome compared with the single use of each algorithm (Meinken et al., 2014, 2015). SP was analyzed using three algorithms, HECTAR (Gschloessl et al., 2008), Philius (Reynolds et al., 2008) and SignalP 4.1 (Petersen et al., 2011). TM was examined using four algorithms, SOSUI (Hirokawa et al., 1998), TMHMM (Krogh et al., 2001), Phobius (Käll et al., 2004) and Philius.

The performance of the individual algorithms and the combined methods (cut-off threshold) was benchmarked using four positive/negative test datasets. The three of them, designated here as SP606, TM939 and Cy_Nu1000, were prepared for comparing the performance of SP prediction algorithms (Petersen et al., 2011). The datasets were downloaded from http://www.cbs.dtu.dk/services/SignalP/. SP606 contained 606 proteins with SP. TM939 had 939 membrane proteins with TM within the first 70 amino acids and downstream sequence. Cy_Nu1000 included 1000 cytoplasmic and nuclear proteins without SP and TM. Another negative test dataset, designated here as CP66, harbored 66 chloroplastic proteins with the N-terminal bipartite targeting peptide. It consisted of 11 proteins from brown algae and cryptophyte that was used for testing the performance of HECTAR (Gschloessl et al., 2008), and 55 proteins from diatoms that was used for evaluating the performance of ASAFind (Gruber et al., 2015). The prediction sensitivity, specificity and Matthews' Correlate Coefficient (MCC) (Matthews, 1975) was calculated using following equations: Sensitivity (%) = TP/(TP + FN) × 100, Specificity (%) = TN/(TN + FP) × 100, $MCC = (TP \times TN - FP \times FN)/((TP + FP)(TP + FN)(TN + FP)(TN + FN))^{1/2}$ where TP, FP, TN and FN indicate true positive, false positive, true negative and false negative, respectively.

The proteins predicted to be SP positives by all the algorithms ("all positive threshold") were taken as those with highly likely SP. The proteins with SP absent on the "all positive threshold" but SP present on another threshold ("conditional threshold") were regarded as those with weak SP. SignalP 4.1 consists of two neural networks: SignalP-noTM and SignalP-TM. The latter network acts when the N-terminal TM was detected in the input sequence. On the "conditional threshold", if SignalP-noTM was activated, the proteins predicted to be SP positives by at least two algorithms were considered as those with SP present, whereas if SignalP-TM was run, the proteins predicted to be SP positives by all algorithms were judged as those with SP present. The proteins predicted to be TM positives by at least two algorithms ("two positive threshold") were assigned as those with TM present. The N-terminal TM overlapped with the highly likely SP was assumed to be a false positive. The N-terminal TM overlapped with the weak SP was not counted and marked as "unclear". The Type 1 proteins were further filtered to remove the intracellular proteins by manually searching the C-terminal ER retention signal ([KRHQSA][DENQ]EL motif) (Munro and Pelham, 1986; Sigrist et al., 2002), and by annotation in combination with BLASTp search against the NCBI protein database and the conserved domain analysis described below. Those probable intracellular proteins were marked as "less likely secretome". The remaining proteins were predicted to belong to the secretome and marked as "highly likely secretome".

2.5. Bioinformatic analysis

The proteins in the predicted secretome were blasted (BLASTp, cut-off *E*-value: $1E^{-4}$) against all the ORFs from the transcriptome of the laminarialean species and proteomes of *E. siliculosus* and *S. japonica* using the local BLAST+ software. The proteins of interest were compared with those of other organisms using NCBI BLAST algorithms. Conserved domains within each protein were determined with InterProScan v5-55.0 (Jones et al., 2014) standalone mode against the InterPro databases, and the batch web CD-search tool against conserved domain database (CDD) (Marchler-Bauer et al., 2015). The proteins acting on carbohydrates were searched using dbCAN (Yin et al., 2012).

3. Results and discussion

3.1. Benchmark of prediction methods of the secretome

In order to establish the prediction method appropriate for the brown algal secretome, the performance of the prediction algorithms SignalP 4.1, Philius and HECTAR for signal peptide (SP) was examined against four positive/negative test datasets (Table 1; Positive dataset: SP606; Negative dataset: Cy_Nu1000, TM939 and CP66). The prediction for SP often suffers from false positives of SP confused with the N-terminal transmembrane domain (TM) (Petersen et al., 2011). SignalP 4.0 (4.1) and Philius were developed for discriminating SP from the N-terminal TM (Petersen et al., 2011; Reynolds et al., 2008). Among the three algorithms, SignalP 4.1 showed the highest specificity (94.4%) and MCC (0.859) followed by Philius (Table 1). However, the individual use of SignalP 4.1 or Philius would not be enough for predicting the brown algal secretome, because the majority of chloroplastic proteins of heterokonts have an N-terminal bipartite target peptide consisting of an N-terminal SP followed by a chloroplast transit peptide (Gschloessl et al., 2008). They indeed showed the considerably high rate of false positive of SP (SignalP 4.1: 92.4%; Philius: 95.5%) against the negative test dataset, CP66 that contained proteins with the bipartite target peptide (Table 1). HECTAR can separately recognize SP and the bipartite target peptide (Gschloessl et al., 2008). The algorithm achieved the much lower rate of false positive of SP (24.2%) against CP66 (Table 1). Nonetheless, the lowest specificity (85.9%) and MCC (0.743) were shown for HECTAR due to the highest rate of false positive of SP against the negative test dataset, TM939 that contained proteins with the N-terminal TM (Table 1). We thus considered that HECTAR alone was not sufficient for predicting the brown algal secretome. It was expected that the combined use of SignalP 4.1, Philius and HECTAR could improve the prediction accuracy for the brown algal secretome by complementarily enabling sorting of the SP, the N-terminal TM and the bipartite target peptide.

The prediction results for SP by the three algorithms were evaluated using four cut-off thresholds in order to decide whether a given protein had SP or not (the detailed descriptions of the thresholds are provided in Materials and methods). The "conditional threshold" and the "all positive threshold" resulted in the better specificity (97.0 and 97.5%, respectively) and MCC (0.904 and 0.888, respectively) than those of single use of each algorithm, "one positive threshold" and "two positive threshold" (Table 1). We thought that the specificity was important in spite of the slightly worse sensitivity, since the secretome prediction primarily depended on the presence of SP. Based on the specificity and MCC, we decided to accept the proteins predicted to be SP positives on the "all positive threshold".

The prediction performance for TM was evaluated using the same datasets (Positive dataset: TM939; Negative dataset: SP606, Cy_Nu1000 and CP66). The false positive of TM confused with SP and the bipartite target peptide was not a serious problem since it could be corrected using the prediction result for SP. As expected, the higher specificity and MCC were generally shown for the data corrected using the result of prediction for SP on the "all positive threshold" compared with those of the data not corrected (Table 2). In the corrected data, the "two positive threshold" resulted in the highest MCC (0.897, Table 2). We thus decided to take the proteins predicted to be TM positives on the "two positive threshold".

3.2. Genome-wide prediction of the brown algal secretome

The brown algal secretome was predicted from the two proteomes of *E. siliculosus* (16,189 proteins) and *S. japonica* (18,733 proteins) by integrating the SP/TM prediction modules described above (Fig. 1). The 842 and 1292 proteins were assigned to Type 1 (SP present/TM absent) in *E. siliculosus* and *S. japonica* (Table 3, Supplementary Data-1). After removing the probable intracellular proteins, 552 and 964 proteins were obtained in *E. siliculosus* and *S. japonica* (Table 3, Supplementary Data-1).

A total of 206 and 200 conserved domains were found in the secretome of *E. siliculosus* and *S. japonica*, respectively (Table 4, Supplementary Data-1). The repertory of the domains in the secretome was similar to each other, but they differed in that the number of proteins containing each domain increased in *S. japonica*. This suggested that the expansion of the content of the secretome in *S. japonica* in

Table 1
Benchmark of prediction algorithms for signal peptide (SP).

Algorithm	SP606 false negative (%)	Cy_Nu1000 false positive (%)	TM939 false positive (%)	CP66 false positive (%)	Sensitivity (%)	Specificity (%)	MCC
Single use							
SignalP 4.1	4.8	0.4	5.0	92.4	95.2	94.4	0.859
Philius	6.3	3.6	14.0	95.5	93.7	88.5	0.754
HECTAR	3.0	0.8	27.5	24.2	97.0	85.9	0.743
Combined use (SignalP 4.1 + Philius + HECTAR)							
Positive in at least one algorithm **(one positive threshold)**	1.7	3.9	30.1	24.2	98.3	83.1	0.718
Positive in at least two algorithms **(two positive threshold)**	3.3	0.6	12.8	22.7	96.7	93.0	0.845
Positive in all algorithms **(all positive threshold)**	8.9	0.3	3.5	21.2	91.1	97.5	0.888
Conditional threshold	5.1	0.6	4.3	21.2	94.9	97.0	0.904

comparison to *E. siliculosus* (Table 3) derived from gene duplication and/or domain shuffling in each protein family rather than gene gain in the laminarialean lineage. The content of the secretome of brown algae (*E. siliculosus*: 3.4%; *S. japonica*: 5.1%) was similar to those of the chlorophytes (average: 4.8%) (Lum et al., 2014), fungi (average: 4.4%) (Meinken et al., 2014) and animals (average: 8.1%) (Meinken et al., 2015). On the other hand, although the domain repertory of the brown algal secretome was in part common with that of the above three groups, it was quite different (Table 5). For each lineage, the similarity of the domain repertory within the lineage was significantly higher than that across lineages, and that to brown algae was lower than other combinations (Table 5). The result may show that the domain repertory of the brown algal secretome is much different from that of the other groups. Thereby, we consider that it reflects the phylogenetically independent evolution of multicellularity of brown algae from other multicellular lineages. The presence of the domains shared with other lineages probably suggests that some proteins in the brown algal secretome have the function similar to that of proteins in the secretome of other multicellular lineages.

3.3. Alteration of the secretome in the laminarialean lineage

It has been suggested that the gene family expansion and innovation of the domain architecture of ECM proteins contributed to multicellularity and adaptive evolution in animals and green plants (Lespinet et al., 2002; Özbek et al., 2010; Prochnik et al., 2010). The protein families in the brown algal secretome were compared with regard to those features among *E. siliculosus*, *S. japonica* and the seven laminarialean species using the RNA-seq data (Table 6). We identified 21 families in the secretome of *S. japonica* of which members were abundant more than twice compared with that of *E. siliculosus* (Table 7). The members of 20 families except family19 were significantly enriched in the secretome of *E. siliculosus* or *S. japonica* (Table 7, Fisher's exact test, $p < 0.05$), indicating that they have important roles in ECM. Although some members of the families were categorized into Type 2 or 3, the Type 0 proteins may be also secreted as leaderless secreted proteins (LSPs) (Agrawal et al., 2010). In *E. siliculosus*, for example, one Type 0 mannuronan C5-epimerase (MEPs, family9, Table 7) was detected by the proteomic analysis against the ECM protein fraction (Terauchi

Table 2
Benchmark of prediction algorithms for transmembrane domain (TM).

Algorithm	Not corrected			Corrected		
	Sensitivity (%)	Specificity (%)	MCC	[a]Sensitivity (%)	[a]Specificity (%)	[a]MCC
Single use						
TMHMM	90.8	87.3	0.763	88.6	97.5	0.876
Phobius	84.9	96.1	0.826	84.8	96.2	0.827
SOSUI	87.2	70.1	0.550	84.5	96.2	0.824
Philius	85.8	97.3	0.851	84.8	97.5	0.846
Combined use (TMHMM + Phobius + SOSUI + Philius)						
Positive in at least one algorithm **(one positive threshold)**	97.4	66.7	0.621	95.6	93.4	0.878
Positive in at least two algorithms **(two positive threshold)**	93.3	87.1	0.782	92.4	96.8	0.897
Positive in at least three algorithms **(three positive threshold)**	85.1	97.8	0.852	85.1	98.1	0.857
Positive in all algorithms **(all positive threshold)**	76.3	99.0	0.803	76.3	99.1	0.805

[a] These data were corrected using the results of prediction for SP against SP606, Cy_Nu1000, TM939 and CP66. The N-terminal TM was regarded as a false positive if it overlapped with the SP detected on the "all positive threshold" or the bipartite target peptide. In SP606 and CP66, only the N-terminal region of the protein sequences corresponding to SP or the bipartite target peptide was concerned.

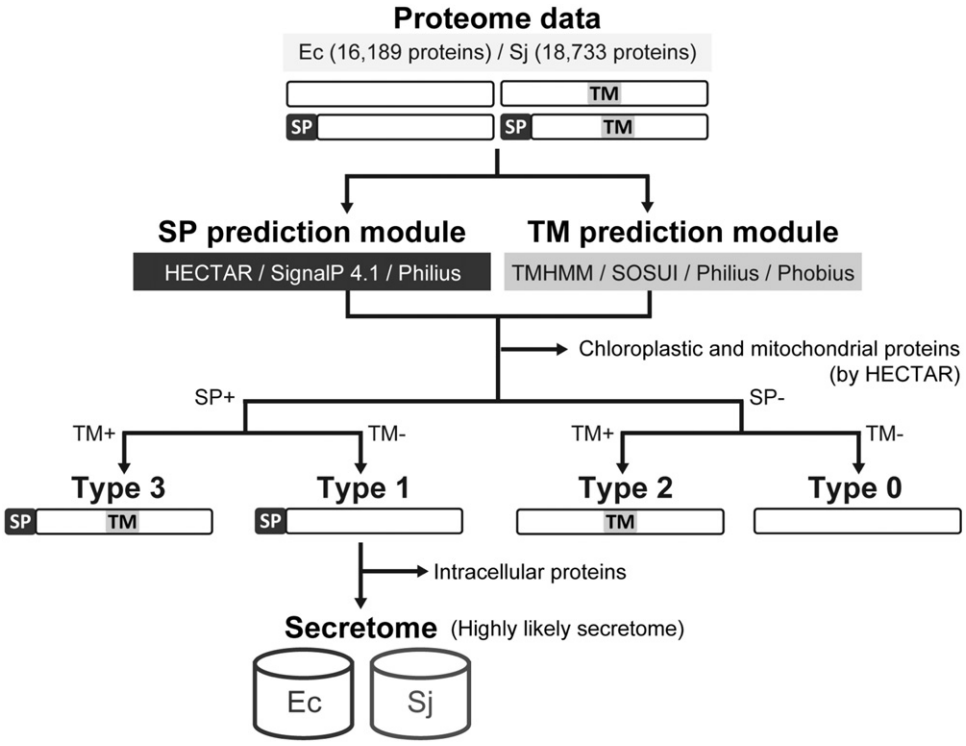

Fig. 1. Overview of the *in silico* pipeline to predict the brown algal secretome. The symbols "+" or "−" indicate that the signal peptide or the transmembrane domain is present or absent, respectively. Ec: *E. siliculosus*; SP: signal peptide; Sj: *S. japonica*; TM: transmembrane domain.

et al., 2016). Among the 20 families, 17 families were significantly expanded in the genome of *S. japonica* compared with that of *E. siliculosus* (Table 7, Fisher's exact test, $p < 0.05$). Additionally, the number of members of six (family1, 2, 8, 22, 32, 37), one (family29) and four (family9, 26, 30, 38) families was increased at least in Pseudochordaceae, Aureophycaceae and Alariaceae, respectively (Table 7). Considering the abundance of the members of these families in the genome of *S. japonica* (Table 7) and the limitation of the RNA-seq that only expressed genes were detected, it is assumed that those species also harbor more members in their genome. These data infer that the expansion of the ECM protein families started from the basal group of Laminariales and repeated during the evolution of laminarialean species. It is likely to be correlated with the elaboration of the ECM remodeling system as discussed in the following sections.

3.3.1. Catalytic ECM remodeling-related proteins acting on polysaccharides

MEPs (family9, Table 7) catalyze the final step of alginate synthesis (*i.e.* epimerization of mannuronic acid (M) to guluronic acid (G) within the mannuronan chain) (Fischl et al., 2016; Inoue et al., 2016). The present study found that the domain architecture of MEPs was slightly altered in the laminarialean species. MEPs of *E. siliculosus* had two putative alginate-binding WSC domains (InterPro domain ID: IPR002889) (Michel et al., 2010) per protein at maximum (Fig. 2). Some MEPs of the laminarialean species may have three or four WSC domains (Fig. 2). Since increase in the number of WSC domains was found in *P. nagaii*, the alteration of the domain architecture of MEPs probably started before the derived group of Laminariales diverged. The expansion of MEPs and the alteration of the domain architecture may result in their varied spatiotemporal expression (Fischl et al., 2016; Tonon et al., 2008), differentiation of the enzyme activity (Fischl et al., 2016; Inoue et al., 2016) and localization in the ECM, correlated with the diverse MG sequence of alginate (Haug et al., 1974; Miller, 1996).

Alginates are structurally modified by cross-link to phenolic compounds by Vanadium-dependent haloperoxidase (vHPO, family8, Table 7) in the presence of hydrogen peroxide (Deniaud-Bouët et al., 2014; Salgado et al., 2009). Strong induction in the expression of vHPO genes by biotic and abiotic stresses has been reported in *E. siliculosus* and laminarialean species for reinforcement of ECM (Cosse et al., 2009; Strittmatter et al., 2016; Dittami et al., 2012). Furthermore, it was reported that the vHPOs of *S. japonica* were differentially expressed depending on generations (sporophyte and gametophyte), developmental stages and tissues (Ye et al., 2015). Hydrogen peroxide required for the activity of vHPO might be supplied by Cupin (family5, Table 7) similar to germin-like proteins of land plants. Some germin-like proteins were shown to convert superoxide to hydrogen peroxide as superoxide dismutase (Dunwell et al., 2004).

Although the gene abundance did not increase in the laminarialean species, the alteration of the domain architecture was detected in another putative alginate degradation-related protein, family 88 glycoside hydrolase (family43, GH88, Fig. S1 in Supplementary data-2). GH88 of brown algae probably participates in alginate degradation based on the unsaturated uronic acid dehydrogenase activity, cooperating with the unidentified alginate lyase (Supplementary data-2). The ECM polysaccharide-degrading enzymes have not been identified in brown

Table 3
Summary of genome-wide secretome prediction from *E. siliculosus* and *S. japonica*.

	E. siliculosus	*S. japonica*
Type	Number of proteins (%)	
Total	16,189	18,733
Type 0	11,551 (71.4)	12,973 (69.3)
Type 1	842 (5.2)	1292 (6.9)
Type 2	2526 (15.6)	2997 (16.0)
Type 3	203 (1.3)	263 (1.4)
Chloroplast	[a]448 (2.8)	[a]585 (3.1)
Mitochondrion	[a]619 (3.8)	[a]623 (3.3)
Highly likely secretome (included in Type 1)	552 (3.4)	964 (5.1)

A full list is in Supplementary Data-1.
[a] Including the proteins with "TM present/absent".

Table 4
Comparison of the composition of conserved domains in the secretome of *E. siliculosus* and *S. japonica*.

		E. siliculosus		*S. japonica*	
InterPro ID	Domain name	Count	%	Count	%
IPR011050	Pectin lyase fold/virulence factor	[a]30	[b]5.4	[a]127	[b]13.2%
IPR013994	Carbohydrate-binding WSC, subgroup	29	5.3	44	4.6%
IPR002889	Carbohydrate-binding WSC	29	5.3	45	4.7%
IPR006626	Parallel beta-helix repeat	18	3.3	66	6.8%
IPR029058	Alpha/Beta hydrolase fold	15	2.7	11	1.1%
IPR022441	Parallel beta-helix repeat-2	11	2.0	37	3.8%
IPR004302	Chitin-binding, domain 3	10	1.8	8	0.8%
IPR012336	Thioredoxin-like fold	9	1.6	6	0.6%
IPR000800	Notch domain	9	1.6	20	2.1%
IPR000742	EGF-like domain	9	1.6	20	2.1%
IPR013111	EGF-like domain, extracellular	9	1.6	1	0.1%
IPR000254	Cellulose-binding domain, fungal	8	1.4	5	0.5%
IPR011051	RmlC-like cupin domain	7	1.3	25	2.6%
IPR007742	Periplasmic copper-binding protein NosD, beta helix domain	7	1.3	29	3.0%
IPR008979	Galactose-binding domain-like	7	1.3	24	2.5%
IPR011990	Tetratricopeptide-like helical domain	7	1.3	4	0.4%
IPR010255	Haem peroxidase	6	1.1	1	0.1%
IPR002016	Haem peroxidase, plant/fungal/bacterial	6	1.1	1	0.1%
IPR006045	Cupin 1	6	1.1	25	2.6%
IPR017853	Glycoside hydrolase superfamily	6	1.1	7	0.7%
IPR032675	Leucine-rich repeat domain, L domain-like	6	1.1	11	1.1%
IPR012938	Glucose/sorbosone dehydrogenase	5	0.9	6	0.6%
IPR011041	Soluble quinoprotein glucose/sorbosone dehydrogenase	5	0.9	5	0.5%
IPR002921	Fungal lipase-like domain	5	0.9	3	0.3%
IPR001305	Heat shock protein DnaJ, cysteine-rich domain	5	0.9	1	0.1%
IPR011009	Protein kinase-like domain	5	0.9	4	0.4%
IPR008922	Uncharacterised domain, di-copper centre	4	0.7	3	0.3%
IPR002227	Tyrosinase copper-binding domain	4	0.7	3	0.3%
IPR002035	von Willebrand factor, type A	4	0.7	23	2.4%
IPR027417	P-loop containing nucleoside triphosphate hydrolase	4	0.7	7	0.7%

Thirty most abundant domains in the secretome of *E. siliculosus* are shown.
A full list and the result against CDD are in Supplementary Data-1.
[a] When the domain was multiply detected in a sequence, it was counted as one.
[b] The percentage (%) was calculated based on a total of 552 (*E. siliculosus*) and 964 (*S. japonica*) proteins.

algae although the enzyme activity of the alginate lyase was detected in several brown algae including *U. pinnatifida* (Shiraiwa et al., 1975). The proteins of family22, 37 and 38 (Table 7) contained Pectin lyase fold/virulence factor domain (InterPro domain ID: IPR011050). Additionally, some family38 proteins of the laminarialean species showed the sequence signature of Parallel beta-helix repeat (InterPro domain ID: IPR006626). These domains are present in a number of carbohydrate-active enzymes (Jenkins et al., 1998) including MEPs (Fig. 2). The sequence features suggest that they could have lytic activities on the ECM polysaccharides of brown algae.

3.3.2. Catalytic ECM remodeling-related proteins acting on proteins

The ECM of brown algae contained proteins tightly bound to other ECM components, indicating that the proteins maintain the structural integrity of the brown algal ECM (Deniaud-Bouët et al., 2014; Mabeau and Kloareg, 1987). In animals, the Zn^{2+}-dependent metalloproteinases called metzincins play central roles in ECM remodeling (Lu et al., 2011). The member of metzincins, reprolysin-like metalloproteinases (family3, Table 7) were composed of the conserved Zn^{2+}-binding catalytic region (HEXXHXXGXXH) and the methionine-containing motif called Met-turn (MXX) (Gomis-Rüth, 2003), and zero to six WSC domains (Fig. 3A). The WSC domain might have ancillary functions for the metalloproteinases by providing the ECM-binding capacity like the glycosaminoglycan-binding domains of the ADAMTS family reprolysin-type metalloproteinases of animals (Apte, 2009). The redundancy of the WSC domain was increased in the metalloproteinases of the laminarialean species compared with *E. siliculosus* (Fig. 3B). Since it was found in *P. nagaii*, it is predicted that the alteration of the domain architecture took place in the basal group of Laminariales.

In animals, the metalloproteinase activity is controlled by endogenous inhibitor proteins, tissue inhibitors of metalloproteinases (TIMPs) (Brew and Nagase, 2010). The TIMP-like proteins of brown algae (family1, Table 7) contained TIMP-like domain (InterPro domain ID: IPR008993), and were similar to the N-terminal domain (N-TIMP) of the animal TIMPs, which alone had an inhibitory effect (Fig. 3C). They presumably have the metalloproteinase inhibitor activities because they retained the cysteine motif (CXC) and the other two cysteines (Fig. 3C), which are important for the proper folding and the inhibitor activity of TIMPs (Meng et al., 1999). The expansion of these families indicates the elaboration of the proteolytic network in the ECM of Laminariales. This idea would be supported by the expansion of Subtilisin-like serine protease (family26, Table 7) which is involved in processing of the ECM proteins in land plants (Srivastava et al., 2008).

3.3.3. Non-catalytic ECM remodeling-related proteins acting on polysaccharides

The family30 proteins (Table 7) are likely to be carbohydrate binding module 32 (CBM32)-containing proteins shown by dbCAN with the highly conserved tryptophan residue for the ligand binding (Abbott et al., 2007) (Fig. 4A), lacking any catalytic module. The proteins containing only CBM called “isolated CBM” have been reported to participate in polysaccharide degradation based on their carbohydrate

Table 5
Similarity of the domain repertory between the brown algal secretome and that of other lineages.

Lineage	Total no. domains	No. species analysed	Average no. domains per species	[b]Shared with Brown algae (%)	[b]Shared with Animals (%)	[b]Shared with Green plants (%)	[b]Shared with Fungi (%)
Brown algae	226	2	156	[c]55.1 ± 0.5*	30.8 ± 1.2	37.2 ± 0.3	34.0 ± 3.0
Animals	[a]633	9	342	12.5 ± 1.0	[c]94.8 ± 5.6*	29.3 ± 0.9**	25.3 ± 1.3**
Green plants	[a]789	8	190	21.9 ± 3.5	40.8 ± 3.6**	[c]79.6 ± 5.5*	54.4 ± 4.5**
Fungi	[a]614	10	72	22.6 ± 9.6	33.5 ± 8.3	62.4 ± 7.7**	[c]84.7 ± 18.7*

For each lineage, the similarity of the domain repertory was statistically evaluated based on Tukey-Kramer test. Asterisk (*) indicates that the similarity of the domain repertory within the lineage was significantly different from that across lineages ($p < 0.01$). Doubleasterisk (**) indicates that the similarity of the domain repertory in the given combination was significantly different from that to brown algae ($p < 0.01$).
[a] The list of conserved domains (Pfam ID) was obtained from Meinken et al. (2015) (animals), Lum et al. (2014) (green plants) and Meinken et al. (2014) (fungi). The cluster domain ID was downloaded from CDD using the Pfam ID.
[b] For each species, the ratio (%) of the number of cluster domain IDs shared with at least one other species to the total number of detected cluster domain IDs was calculated and the average ratio was expressed with SD.
[c] The values indicate the similarity of the domain repertory within the lineage.

Table 6
Summary of RNA-seq.

	Pseudochorda nagaii	*Aureophycus aleuticus*	*Undaria pinnatifida*	*Alaria crassifolia*	*Costaria costata*	*Agarum clathratum*	*Saccharina angustata*
Total raw reads	53,927,464	61,447,334	75,232,776	53,720,052	63,393,504	49,905,468	64,722,116
Total clean reads	46,891,474	52,340,454	64,411,058	45,946,438	53,952,954	48,157,710	53,357,020
Q20 percentage	95.59%	96.01%	96.02%	95.94%	95.88%	97.72%	95.74%
[a]DDBJ Sequence Read Archive Accession	DRA005083	DRA005083	DRA005083	DRA005083	DRA005083	DRA005083	DRA005083
BioProject Accession	PRJDB5131	PRJDB5131	PRJDB5131	PRJDB5131	PRJDB5131	PRJDB5131	PRJDB5131
BioSample Accession	SAMD00058090	SAMD00058088	SAMD00058092	SAMD00058087	SAMD00058089	SAMD00058086	SAMD00058091
[b]Total assembled transcripts	78,803	52,562	54,203	55,569	59,788	55,794	58,936
DDBJ Transcriptome Shotgun Assembly database Accession	IABM01000001–IABM01078803	IABK01000001–IABK01052562	IABO01000001–IABO01054203	IABJ01000001–IABJ01055569	IABL01000001–IABL01059788	IABI01000001–IABI01055794	IABN01000001–IABN01058936
[b]Average length for transcripts (nt)	613	732	782	695	655	481	671
[b]Total ORFs	31,469	27,518	28,294	29,001	28,428	22,667	30,854
[b]Average length for ORFs (aa)	242	219	225	213	209	173	213

[a] The clean read data was deposited.
[b] The data was calculated after excluding bacterial contaminant sequences.

binding activity (Guillén et al., 2010). CBM32 binds to galactose, *N*-acetylgalactosamine and polygalacturonic acid (Abbott et al., 2007; Guillén et al., 2010). CBM32 is also present in many bacterial alginate lyases, and it was reported that its enzyme activity was reduced by the truncation of the domain (Badur et al., 2015; Han et al., 2016; Weiner et al., 2008). It indicates that the CBM32 possesses the alginate binding activity. Among 56 proteins of *S. japonica*, SJ19651 showed the similarity with the highest score (*E*-value: 2.3E-09) to CBM32 of the bacterial putative alginate lyase (NCBI accession: AFU99265) (Fig. 4A). These data infer that the family30 proteins are alginate binding proteins. There have been reports that isolated CBMs non-enzymatically alter the structural state of the target polysaccharide (Guillén et al., 2010). In land plants, overexpression of CBM affected the cell wall extensibility, cell wall polysaccharide composition and growth pattern (Nardi et al., 2015; Obembe et al., 2007). Taken together, it is hypothesized that the brown algal isolated CBM32 belongs to the ECM remodeling system. The expansion of the family30 might be correlated with the expansion of MEPs (family9) and diversification of the molecular structure of alginates of laminarialean species. CBM32 was detected in other families (family35, 41) of which proteins had the Pectin lyase fold domain as is the case for the aforementioned family22, 37 and 38. Although they were not significantly expanded in the secretome of *S. japonica* compared with *E. siliculosus*, *co*-occurrence of the two domains implies that they would be additional candidates of the ECM polysaccharides-degrading enzymes (*i.e.* alginate lyase) of brown algae.

The family2 proteins (Table 7) showed the sequence signature of YoaJ (CDD ID: COG4305), a bacterial homologue of land plants' expansins identified in *Bacillus subtilis* (Georgelis et al., 2011). YoaJ and land plants' expansins non-enzymatically disrupt the non-covalent bonds between cellulose and matrix polysaccharides, enabling the cell wall loosening (Georgelis et al., 2011; Sampedro and Cosgrove, 2005). YoaJ consists of the N-terminal RlpA-like double-psi beta-barrel (DPBB) domain (InterPro domain ID: IPR009009) as is the case for land plants' expansins and the C-terminal CBM63. CBM63 was able to bind to cellulose and matrix polysaccharides, and both DPBB domain and CBM63 are necessary for the plant cell wall loosening activity (Georgelis et al., 2011). The family2 proteins of brown algae had the same domain architecture composed of DPBB domain and CBM63 detected by dbCAN (Fig. 4B, C). The two of three aromatic amino acids within CBM63 important for its polysaccharide binding activity were

Table 7
Secretome protein family expanded in the laminarialean lineage.

		No. members in each family									
Family no.	Family name	Ec	Sj	Pn	Aa	Up	Ac	Cc	Aga	Sa	Pattern of expansion
1	TIMP-like protein*	[a]1/2	[a]9/14**	4	3	3	5	5	6	7	expanded at least in Pseudochordaceae
2	YoaJ-like protein*	3/3	13/16**	5	2	3	7	6	1	3	expanded at least in Pseudochordaceae
3	Reprolysin-like metalloproteinase*	3/7	9/19**	6	4	5	3	23	2	6	expanded at least in Agaraceae
5	Cupin*	6/8	25/31**	1	2	3	1	0	1	2	expanded at least in Laminariaceae
8	Vanadium-dependent haloperoxidase*	1/1	32/85**	3	3	7	10	8	6	12	expanded at least in Pseudochordaceae
9	Mannuronan C-5 epimerase*	14/31	58/105**	25	17	43	42	40	19	40	expanded at least in Alariaceae
14	Secreted protein, family14*	1/1	4/4	1	1	1	1	0	0	2	expanded at least in Laminariaceae
16	Secreted protein, family16*	1/3	2/7	1	1	0	1	0	0	0	expanded at least in Laminariaceae
18	Secreted protein, family18*	3/4	7/11	1	3	3	3	5	2	3	expanded at least in Agaraceae
19	Lipase	1/6	2/15	1	4	3	1	8	4	13	expanded at least in Agaraceae
21	Endo-1,3-beta glucanase, GH81*	2/5	8/35**	1	2	5	1	1	1	3	expanded at least in Laminariaceae
22	Pectin lyase fold-containing protein, family22*	2/3	14/27**	5	4	11	10	10	1	9	expanded at least in Pseudochordaceae
24	Notch domain-containing protein*	9/23	19/49**	4	7	10	17	13	4	12	expanded at least in Laminariaceae
25	vWFA-like protein*	2/5	23/38**	2	1	2	4	1	1	10	expanded at least in Laminariaceae
26	Subtilisin-like serine protease*	1/12	7/41**	12	10	14	15	20	14	9	expanded at least in Alariaceae
29	Cysteine-rich repeat-containing protein*	2/4	40/94**	3	5	6	4	14	5	12	expanded at least in Aureophycaceae
30	CBM32-like carbohydrate binding protein*	2/10	21/56**	10	5	20	34	24	11	30	expanded at least in Alariaceae
32	Secreted protein, family32*	1/3	6/18**	7	0	4	9	17	0	10	expanded at least in Pseudochordaceae
33	Leucine rich repeat carbohydrate binding protein*	4/18	8/20	13	6	13	17	20	5	28	expanded at least in Agaraceae
37	Pectin lyase fold-containing protein, family37*	7/7	15/24**	8	7	7	9	22	3	9	expanded at least in Pseudochordaceae
38	Pectin lyase fold-containing protein, family38*	2/8	21/59**	7	3	7	20	9	6	9	expanded at least in Alariaceae

Asterisk (*) indicates that the members of the protein family were significantly enriched in the secretome of *E. siliculosus* or *S. japonica* (Fisher's exact test, $p < 0.05$). Doubleasterisk (**) indicates that the protein family was significantly expanded in the genome of *S. japonica* compared with *E. siliculosus* (Fisher's exact test, $p < 0.05$). A full list of sequence IDs of the families are in Supplementary Data-1.
[a] The number on the left and right indicates the gene abundance in the secretome and the genome, respectively. Aa: *A. aleuticus*; Ac: *A. crassifolia*; Aga: *A. clathratum*; Cc: *C. costata*; Ec: *E. siliculosus*; Pn: *P. nagaii*; Sa: *S. angustata*; Sj: *S. japonica*; Up: *U. pinnatifida*.

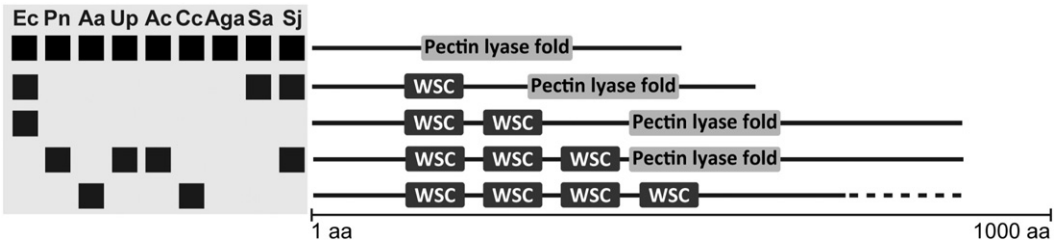

Fig. 2. Catalytic ECM remodeling-related proteins acting on polysaccharides. Domain architecture of Mannuronan C-5 epimerases (MEPs, family9). Characteristic domains are shown: Pectin lyase fold (InterPro domain ID: IPR011050) and WSC (InterPro domain ID: IPR002889). The black box indicates that the corresponding domain architecture on the right was predicted. The diagram is drawn to scale (aa: amino acids). Aa: *A. aleuticus*; Ac: *A. crassifolia*; Aga: *A. clathratum*; Cc: *C. costata*; Ec: *E. siliculosus*; Pn: *P. nagaii*; Sa: *S. angustata*; Sj: *S. japonica*; Up: *U. pinnatifida*.

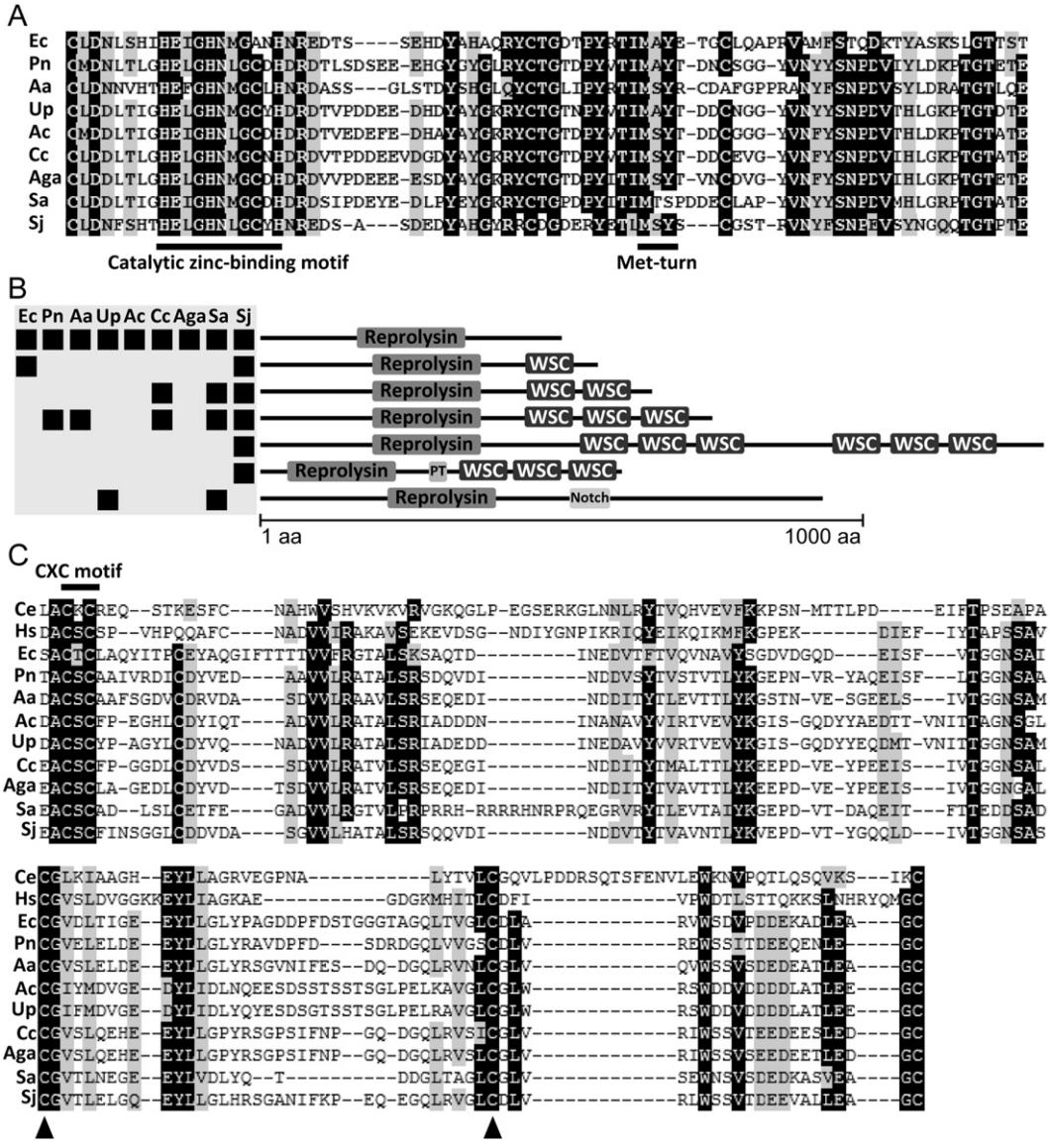

Fig. 3. Catalytic ECM remodeling-related proteins acting on proteins. A) Alignment of the amino acid sequences of the predicted catalytic region of reprolysin-like metalloproteinases (family3). The bars denote the conserved motifs of the metalloproteinases in other organisms. The sequences are from D8LCC0 (Ec: *E. siliculosus*), CL490.Contig3 (Pn: *P. nagaii*), Unigene15318 (Aa: *A. aleuticus*), CL4273.Contig1 (Up: *U. pinnatifida*), Unigene1848 (Ac: *A. crassifolia*), Unigene19273 (Cc: *C. costata*), Unigene12674 (Aga: *A. clathratum*), Unigene21160 (Sa: *S. angustata*) and SJ20417 (Sj: *S. japonica*). B) Domain architecture of reprolysin-like metalloproteinases. Characteristic domains are shown: Notch (CDD ID: cl02419), PT (CDD ID: cl04822), Reprolysin (CDD ID: pfam13582) and WSC (CDD ID: cl02568). The black box indicates that the corresponding domain architecture on the right was predicted. The diagram is drawn to scale (aa: amino acids). Aa: *A. aleuticus*; Ac: *A. crassifolia*; Aga: *A. clathratum*; Cc: *C. costata*; Ec: *E. siliculosus*; Pn: *P. nagaii*; Sa: *S. angustata*; Sj: *S. japonica*; Up: *U. pinnatifida*. C) Alignment of the amino acid sequences of Tissue inhibitor of metalloproteinase (TIMP)-like proteins of brown algae (family1) with TIMPs of animals. The partial sequence similar to the N-TIMP was aligned. The bars denote the highly conserved cysteine (CXC) motifs of TIMPs in animals. The arrowheads point to the conserved cysteines, which form the disulfide bonds to the two cysteines within the CXC motif. The sequences are from AAG50211.1 (Ce: *Caenorhabditis elegans*), AAC50729.1 (Hs: *Homo sapiens*), D8LCN7 (Ec: *E. siliculosus*), CL1076.Contig1 (Pn: *P. nagaii*), Unigene15891 (Aa: *A. aleuticus*), Unigene20341 (Ac: *A. crassifolia*), Unigene16855 (Up: *U. pinnatifida*), Unigene18339 (Cc: *C. costata*), Unigene13088 (Aga: *A. clathratum*), Unigene13966 (Sa: *S. angustata*) and SJ21872 (Sj: *S. japonica*).

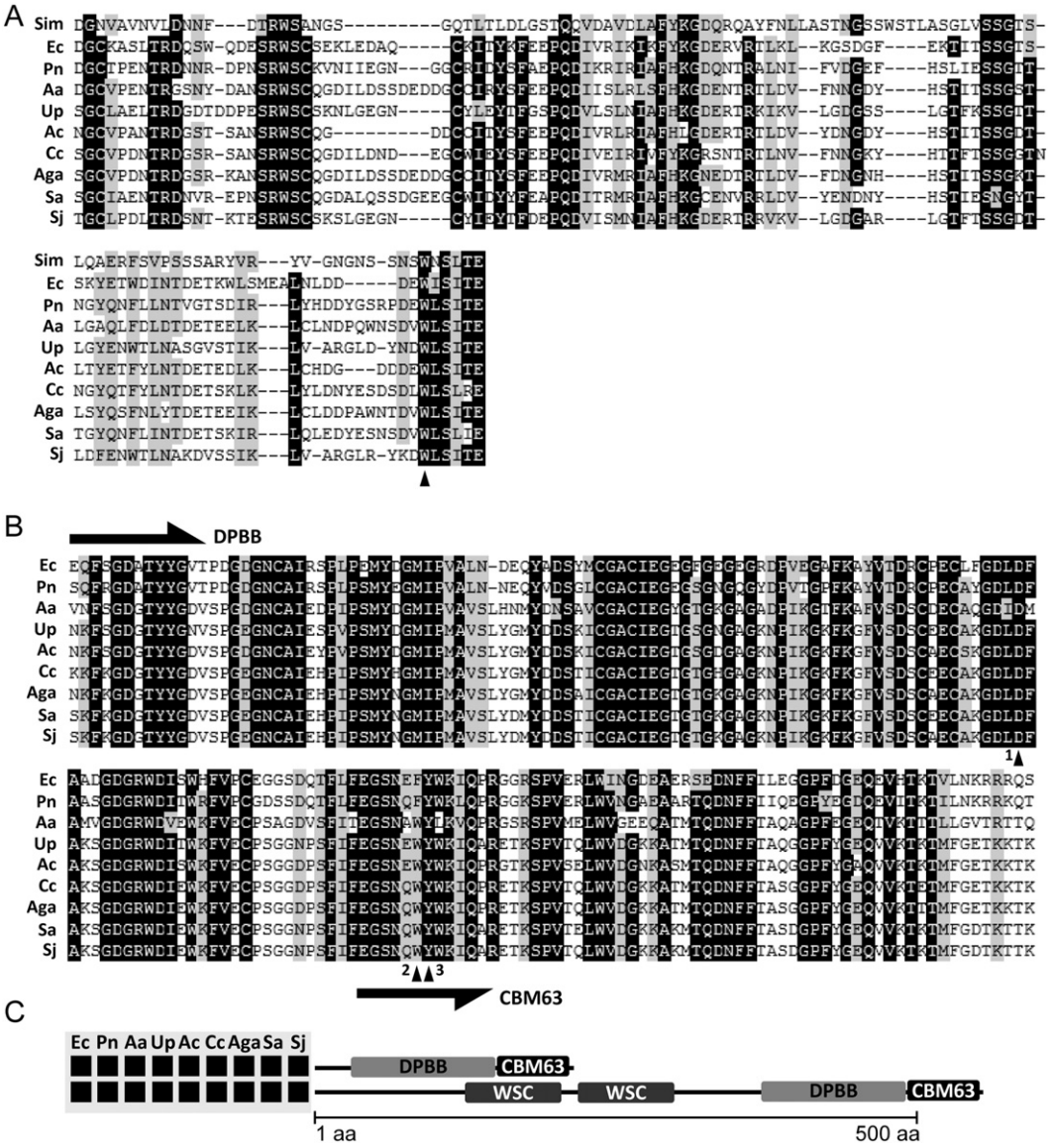

Fig. 4. Non-catalytic ECM remodeling-related proteins acting on polysaccharides. A) Alignment of the amino acid sequences of Carbohydrate binding module 32 (CBM32, family30). The arrowhead points to the conserved tryptophan residue for the ligand binding of CBM32. The sequences are from AFU99265.1 (Sim: *Simiduia agarivorans*), D7G122 (Ec: *E. siliculosus*), Unigene30789 (Pn: *P. nagaii*), CL1391.Contig1 (Aa: *A. aleuticus*), CL570.Contig1 (Ac: *A. crassifolia*), CL2434.Contig1 (Up: *U. pinnatifida*), CL3627.Contig1 (Cc: *C. costata*), CL7001.Contig1 (Aga: *A. clathratum*), Unigene22543 (Sa: *S. angustata*) and SJ19651 (Sj: *S. japonica*). B) Alignment of the amino acid sequences of YoaJ-like proteins (family2). The arrowhead-1 points to the conserved aspartic acid residue within RlpA-like double-psi beta-barrel (DPBB) domain for the plant cell wall loosening activity of YoaJ and land plants' expansins. The arrowhead-2, 3 indicate the aromatic amino acid residues within CBM63 for the polysaccharide-binding activity of YoaJ. C) Domain architecture of YoaJ-like proteins. Characteristic domains are shown: CBM63 (dbCAN), DPBB (InterPro domain ID: IPR009009), WSC (InterPro domain ID: IPR002889). The black box indicates that the corresponding domain architecture on the right was predicted. The diagram is drawn to scale (aa: amino acids). Aa: *A. aleuticus*; Ac: *A. crassifolia*; Aga: *A. clathratum*; Cc: *C. costata*; Ec: *E. siliculosus*; Pn: *P. nagaii*; Sa: *S. angustata*; Sj: *S. japonica*; Up: *U. pinnatifida*.

conserved in brown algae (Fig. 4B). Additionally, the brown algal sequences contained the highly conserved aspartic acid residue within the DPBB domain (Fig. 4B) crucial for the cell wall loosening activity of YoaJ and land plants' expansins (Nikolaidis et al., 2014). These data strongly suggest that the family2 proteins have the ECM binding and loosening activities. The fusion with the WSC domain in some family2 proteins conserved in *E. siliculosus* and laminarialean species (Fig. 4C) supports that they are parts of the ECM remodeling system.

Although the gene abundance was not expanded in the genome of *S. japonica*, the alteration of the domain architecture occurred in the putative CBM1-containing cellulose binding proteins of the laminarialean species (family42, Fig. S2 in Supplementary data-2). They may take part in ECM remodeling based on the cellulose binding activity or oxidative cellulose lytic activity (Supplementary data-2).

4. Conclusions

We have established the *in silico* pipeline to predict the brown algal secretome. The analyses of the genomes of *E. siliculosus* and *S. japonica* predicted that 552 and 964 proteins belong to the secretome, respectively. It is likely that there are many LSPs encoded in the genome of brown algae as well as animals and land plants. Although the sequence similarity was low, some domains were shared between the secretomes of brown algae and other multicellular organisms. We consider that they are evolutionarily distinct but functionally analogous. The alteration of the gene family abundance and/or the domain architecture were detected in the alginate-modifying MEP and vHPO families and newly found ECM remodeling-related proteins of the laminarialean species. It is assumed that the alteration started from the basal

group of Laminariales and improved the robustness of functions of the ECM in normal development and environmental response, which might be linked to the evolution of multicellularity in Laminariales. The precise function of the novel ECM remodeling-related proteins should be experimentally characterized. The secretome database is expected to be a useful reference for the future studies of the brown algal ECM.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

We are grateful to Dr. Eric C. Henry for valuable comments on the manuscript, and Drs. Taizo Motomura, Chikako Nagasato, Atsuko Tanaka and Hiroki Kawamoto (Muroran Marine Station, Hokkaido University) for their help in collecting the field samples. A part of this study was supported by a scientific research grant from the Japan Society for Promotion of Sciences (Project Number: 25291087) to H.K.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.margen.2016.12.002.

References

Abbott, D.W., Hrynuik, S., Boraston, A.B., 2007. Identification and characterization of a novel periplasmic polygalacturonic acid binding protein from *Yersinia enterolitica*. J. Mol. Biol. 367, 1023–1033.

Agrawal, G.K., Jwa, N.S., Lebrun, M.H., Job, D., Rakwal, R., 2010. Plant secretome: unlocking secrets of the secreted proteins. Proteomics 10, 799–827.

Apte, S.S., 2009. A disintegrin-like and metalloprotease (reprolysin-type) with thrombospondin type 1 motif (ADAMTS) superfamily: functions and mechanisms. J. Biol. Chem. 284, 31493–31497.

Badur, A.H., Jagtap, S.S., Yalamanchili, G., Lee, J.K., Zhao, H., Rao, C.V., 2015. Alginate lyases from alginate-degrading *Vibrio splendidus* 12B01 are endolytic. Appl. Environ. Microbiol. 81, 1865–1873.

Bold, H.C., Wynne, M.J., 1985. Introduction to the Algae. Structure and Reproduction. second ed. Prentice-Hall, Englewood Cliffs, NJ, USA, p. 720.

Brew, K., Nagase, H., 2010. The tissue inhibitors of metalloproteinases (TIMPs): an ancient family with structural and functional diversity. Biochim. Biophys. Acta 1803, 55–71.

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

Cosse, A., Potin, P., Leblanc, C., 2009. Patterns of gene expression induced by oligoguluronates reveal conserved and environment-specific molecular defense responses in the brown alga *Laminaria digitata*. New Phytol. 182, 239–250.

Deniaud-Bouët, E., Kervarec, N., Michel, G., Tonon, T., Kloareg, B., Hervé, C., 2014. Chemical and enzymatic fractionation of cell walls from Fucales: insights into the structure of the extracellular matrix of brown algae. Ann. Bot. 114, 1203–1216.

Dittami, S.M., Gravot, A., Goulitquer, S., Rousvoal, S., Peters, A.F., Bouchereau, A., Boyen, C., Tonon, T., 2012. Towards deciphering dynamic changes and evolutionary mechanisms involved in the adaptation to low salinities in *Ectocarpus* (brown algae). Plant J. 71, 366–377.

Dunwell, J.M., Purvis, A., Khuri, S., 2004. Cupins: the most functionally diverse protein superfamily? Phytochemistry 65, 7–17.

Fischl, R., Bertelsen, K., Gaillard, F., Coelho, S., Michel, G., Klinger, M., Boyen, C., Czjzek, M., Hervé, C., 2016. The cell-wall active mannuronan C5-epimerases in the model brown alga *Ectocarpus*: from gene context to recombinant protein. Glycobiology (cww040).

Frantz, C., Stewart, K.M., Weaver, V.M., 2010. The extracellular matrix at a glance. J. Cell Sci. 123, 4195–4200.

Georgelis, N., Tabuchi, A., Nikolaidis, N., Cosgrove, D.J., 2011. Structure-function analysis of the bacterial expansin EXLX1. J. Biol. Chem. 286, 16814–16823.

Gomis-Rüth, F.X., 2003. Structural aspects of the metzincin clan of metalloendopeptidases. Mol. Biotechnol. 24, 157–202.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q.D., et al., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652.

Graham, L.E., Wilcox, L.W., 2000. Algae. Prentice-Hall, London, p. 640.

Gruber, A., Rocap, G., Kroth, P.G., Armbrust, E., Mock, T., 2015. Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. Plant J. 81, 519–528.

Gschloessl, B., Guermeur, Y., Cock, J., 2008. HECTAR: a method to predict subcellular targeting in heterokonts. BMC Bioinf. 9, 393.

Guillén, D., Sánchez, S., Rodríguez-Sanoja, R., 2010. Carbohydrate-binding domains: multiplicity of biological roles. Appl. Microbiol. Biotechnol. 85, 1241–1249.

Han, W., Gu, J., Cheng, Y., Liu, H., Li, Y., Li, F., 2016. Novel alginate lyase (Aly5) from a polysaccharide-degrading marine bacterium, *Flammeovirga* sp. strain MY04: effects of module truncation on biochemical characteristics, alginate degradation patterns, and oligosaccharide-yielding properties. Appl. Environ. Microbiol. 82, 364–374.

Haug, A., Larsen, B., Smidsrød, O., 1974. Uronic acid sequence in alginate from different sources. Carbohydr. Res. 32, 217–225.

Hirokawa, T., Boon-Chieng, S., Mitaku, S., 1998. SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics 14, 378–379.

Hynes, R.O., 2009. The extracellular matrix: not just pretty fibrils. Science 326, 1216–1219.

Inoue, A., Satoh, A., Morishita, M., Tokunaga, Y., Miyakawa, T., Tanokura, M., Ojima, T., 2016. Functional heterologous expression and characterization of mannuronan C5-epimerase from the brown alga *Saccharina japonica*. Algal Res. 16, 282–291.

Jenkins, J., Mayans, O., Pickersgill, R., 1998. Structure and evolution of parallel b-helix proteins. J. Struct. Biol. 122, 236–246.

Jones, P., et al., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236–1240.

Käll, L., Krogh, A., Sonnhammer, E., 2004. A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. 338, 1027–1036.

Kawai, H., 1986. Life history and systematic position of *Akkesiphycus lubricus* (Phaeophyceae). J. Phycol. 22, 286–291.

Kawai, H., 2014. Recent advances in the phylogeny and taxonomy of Laminariales, with special reference to the newly discovered basal member *Aureophycus*. Perspect. Phycol. 1, 27–40.

Kawai, H., Kurogi, M., 1985. On the life history of *Pseudochorda nagaii* (Pseudochordaceae fam. nov.) and its transfer from the Chordariales to the Laminariales (Phaeophyta). Phycologia 24, 289–296.

Kawai, H., Sasaki, H., Maeda, Y., Arai, S., 2001. Morphology, life history, and molecular phylogeny of *Chorda rigida*, sp. nov.(Laminariales, Phaeophyceae) from the sea of Japan and the genetic diversity of *Chorda filum*. J. Phycol. 37, 130–142.

Kawai, H., Hanyuda, T., Ridgway, L.M., Holser, K., 2013. Ancestral reproductive structure in basal kelp *Aureophycus aleuticus*. Sci. Rep. 3, 2491.

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

Kloareg, B., Quatrano, R.S., 1988. Structure of the cell walls of marine algae and ecophysiological functions of the matrix of polysaccharides. Oceanogr. Mar. Biol. Annu. Rev. 26, 259–315.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580.

Le Bail, A., Dittami, S.M., De Franco, P.-O., Rousvoal, S., Cock, J.M., Tonon, T., Charrier, B., 2008. Normalisation genes for expression analyses in the brown alga model *Ectocarpus siliculosus*. BMC Mol. Biol. 9, 75.

Lespinet, O., Wolf, Y.I., Koonin, E.V., Aravind, L., 2002. The role of lineage-specific gene family expansion in the evolution of eukaryotes. Genome Res. 12, 1048–1059.

Lu, P., Takai, K., Weaver, V.M., Werb, Z., 2011. Extracellular matrix degradation and remodeling in development and disease. Cold Spring Harb. Perspect. Biol. 3:a005058. http://dx.doi.org/10.1101/cshperspect.a005058.

Lum, G., Min, X.J., 2011a. Plant Secretomes: current status and future perspectives. Plant. OMICS 4, 114–119.

Lum, G., Min, X.J., 2011b. FunSecKB: the fungal secretome knowledgebase. Database (bar001).

Lum, G., Meinken, J., Orr, J., Frazier, S., Min, X.J., 2014. PlantSecKB: the plant secretome and subcellular proteome knowledgebase. Comput. Mol. Biol. 4, 1–17.

Mabeau, S., Kloareg, B., 1987. Isolation and analysis of the cell walls of brown algae: *Fucus spiralis*, *F. ceranoides*, *F. vesiculosus*, *F. serratus*, *Bifurcaria bifurcata* and *Laminaria digitata*. J. Exp. Bot. 38, 1573–1580.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., et al., 2015. CDD: NCBI's conserved domains database. Nucleic Acids Res. 43, D222–D226.

Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim. Biophys. Acta 405, 442–451.

Meinken, J., Asch, D.K., Neizer-Ashun, K.A., Chang, G.H., Cooper Jr., C.R., Min, X.J., 2014. FunSecKB2: a fungal protein subcellular location knowledgebase. Comput. Mol. Biol. 4, 1–17.

Meinken, J., Walker, G., Cooper, C.R., Min, X.J., 2015. MetazSecKB: the human and animal secretome and subcellular proteome knowledgebase. Database http://dx.doi.org/10.1093/database/bav077.

Meng, Q., Malinovskii, V., Huang, W., Hu, Y., Chung, L., Nagase, H., Bode, W., Maskos, K., Brew, K., 1999. Residue 2 of TIMP-1 is a major determinant of affinity and specificity for matrix metalloproteinases but effects of substitutions do not correlate with those of the corresponding P1′ residue of substrate. J. Biol. Chem. 274, 10184–10189.

Michel, G., Tonon, T., Scornet, D., Cock, J.M., 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.

Miller, I.J., 1996. Alginate composition of some New Zealand brown seaweeds. Phytochemistry 41, 1315–1317.

Min, X.J., 2010. Evaluation of computational methods for secreted protein prediction in different eukaryotes. J. Proteomics Bioinforma. 3, 143–147.

Misner, I., Blouin, N., Leonard, G., Richards, T.A., Lane, C.E., 2014. The secreted proteins of *Achlya hypogyna* and *Thraustotheca clavata* identify the ancestral oomycete secretome and reveal gene acquisitions by horizontal gene transfer. Genome Biol. Evol. 7, 120–135.

Munro, S., Pelham, H.R., 1986. An Hsp70-like protein in the ER: identity with the 78 kd glucose-regulated protein and immunoglobulin heavy chain binding protein. Cell 46, 291–300.
Nardi, C.F., Villarreal, N.M., Rossi, F.R., Martínez, S., Martínez, G.A., Civello, P.M., 2015. Overexpression of the carbohydrate binding module of strawberry expansin2 in *Arabidopsis thaliana* modifies plant growth and cell wall metabolism. Plant Mol. Biol. 88, 101–117.
Nikolaidis, N., Doran, N., Cosgrove, D.J., 2014. Plant expansins in bacteria and fungi: evolution by horizontal gene transfer and independent domain fusion. Mol. Biol. Evol. 31, 376–386.
Nishitsuji, K., Arimoto, A., Iwai, K., Sudo, Y., Hisata, K., Fujie, M., Arakaki, N., Kushiro, T., Konishi, T., Shinzato, C., 2016. A draft genome of the brown alga, Cladosiphon okamuranus, S-strain: a platform for future studies of 'mozuku'biology. DNA Res. dsw039.
Obembe, O.O., Jacobsen, E., Timmers, J., Gilbert, H., Blake, A.W., Knox, J.P., Visser, R.G.F., Vincken, J.P., 2007. Promiscuous, non-catalytic, tandem carbohydrate-binding modules modulate the cell-wall structure and development of transgenic tobacco (*Nicotiana tabacum*) plants. J. Plant Res. 120, 605–617.
Özbek, S., Balasubramainian, P., Chiquet-Ehrismann, R., Tucker, R.P., Adams, J.C., 2010. The evolution of extracellular matrix. Mol. Biol. Cell 21, 4300–4306.
Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786.
Popper, Z.A., Michel, G., Hervé, C., Domozych, D.S., Willats, W.G., Tuohy, M.G., Kloareg, B., Stengel, D.B., 2011. Evolution and diversity of plant cell walls: from algae to flowering plants. Annu. Rev. Plant Biol. 62, 567–590.
Prochnik, S.E., Umen, J., Nedelcu, A.M., Hallmann, A., Miller, S.M., Nishii, I., Ferris, P., Kuo, A., Mitros, T., Fritz-Laylin, L.K., et al., 2010. Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. Science 329, 223–226.
Reynolds, S.M., Käll, L., Riffle, M.E., Bilmes, J.A., Noble, W.S., 2008. Transmembrane topology and signal peptide prediction using dynamic bayesian networks. PLoS Comput. Biol. 4, e1000213.
Salgado, L.T., Cinelli, L.P., Viana, N.B., de Carvalho, R.T., de Souza Mourão, P.A., Teixeira, V.L., Farina, M., Filho, G.M.A., 2009. A vanadium bromoperoxidase catalyzes the formation of high-molecular-weight complexes between brown algal phenolic substances and alginates. J. Phycol. 45, 193–202.
Sampedro, J., Cosgrove, D.J., 2005. The expansin superfamily. Genome Biol. 6, 242.
Schmitz, K., 1990. Algae. In: Behnke, H.D., Sjolund, R.D. (Eds.), Sieve elementsComparative Structure Induction and Development. Springer, Berlin, pp. 1–18.
Shiraiwa, Y., Abe, K., Sasaki, S.F., Ikawa, T., Nisizawa, K., 1975. Alginate lyase activities in the extracts from several brown algae. Bot. Mar. 18, 97–104.
Sigrist, C.J.A., Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M., Bairoch, A., Bucher, P., 2002. PROSITE: a documented database using patterns and profiles as motif descriptors. Brief. Bioinform. 3, 265–274.
Silberfeld, T., Leigh, J.W., Verbruggen, H., Cruaud, C., De Reviers, B., Rousseau, F., 2010. A multi-locus time-calibrated phylogeny of the brown algae (Heterokonta, Ochrophyta, Phaeophyceae): investigating the evolutionary nature of the "brown algal crown radiation". Mol. Phylogenet. Evol. 56, 659–674.
Srivastava, R., Liu, J.X., Howell, S.H., 2008. Proteolytic processing of a precursor protein for a growth-promoting peptide by a subtilisin serine protease in *Arabidopsis*. Plant J. 56, 219–227.
Strittmatter, M., Grenville-Briggs, L.J., Breithut, L., Van West, P., Gachon, C.M., Küpper, F.C., 2016. Infection of the brown alga *Ectocarpus siliculosus* by the oomycete *Eurychasma dicksonii* induces oxidative stress and halogen metabolism. Plant Cell Environ. 39, 259–271.
Tatewaki, M., 1966. Formation of a crustaceous sporophyte with unilocular sporangia in *Scytosiphon lomentaria*. Phycologia 6, 62–66.
Terauchi, M., Nagasato, C., Inoue, A., Ito, T., Motomura, T., 2016. Distribution of alginate and cellulose and regulatory role of calcium in the cell wall of the brown alga *Ectocarpus siliculosus* (Ectocarpales, Phaeophyceae). Planta http://dx.doi.org/10.1007/s00425-016-2516-4.
Tonon, T., Rousvoal, S., Roeder, V., Boyen, C., 2008. Expression profiling of the mannuronan C5-epimerase multigenic family in the brown alga *Laminaria digitata* (Phaeophyceae) under biotic stress conditions. J. Phycol. 44, 1250–1256.
von Heijne, G.J., 1990. The signal peptide. Membr. Biol. 115, 195–201.
Wallin, E., von Heijne, G., 1998. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. Protein Sci. 7, 1029–1038.
Weiner, R.M., Taylor II, L.E., Henrissat, B., Hauser, L., Land, M., Coutinho, P.M., Rancurel, C., Saunders, E.H., Longmire, A.G., Zhang, H., et al., 2008. Complete genome sequence of the complex carbohydrate-degrading marine bacterium, *Saccharophagus degradans* strain 2-40T. PLoS Genet. 4, e1000087.
Williams, F., Tew, H.A., Paul, C.E., Adams, J.C., 2014. The predicted secretomes of *Monosiga brevicollis* and *Capsaspora owczarzaki*, close unicellular relatives of metazoans, reveal new insights into the evolution of the metazoan extracellular matrix. Matrix Biol. 37, 60–68.
Ye, N., Zhang, X., Miao, M., Fan, X., Zheng, Y., Xu, D., Wang, J., Zhou, L., Wang, D., Gao, Y., et al., 2015. *Saccharina* genomes provide novel insight into kelp biology. Nat. Commun. 6, 6986.
Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., Xu, Y., 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 40, W445–W451.