

## Origin and evolution of *alginate-c5-mannuronan-epimerase* gene based on transcriptomic analysis of brown algae

WANG Ren<sup>1</sup>, WANG Xumin<sup>2,3</sup>, ZHANG Yalan<sup>1</sup>, YU Jun<sup>2,3</sup>, LIU Tao<sup>1\*</sup>, CHEN Shengping<sup>4</sup>, CHI Shan<sup>1\*</sup>

<sup>1</sup> College of Marine Life Science, Ocean University of China, Qingdao 266003, China

<sup>2</sup> CAS Key Laboratory of Genome Sciences and Information, Beijing Key Laboratory of Genome and Precision Medicine Technologies, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China

<sup>3</sup> Beijing Key Laboratory of Functional Genomics for Dao-di Herbs, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China

<sup>4</sup> Guangdong Province Key Laboratory of Functional Molecules in Oceanic Microorganism, Zhong Shan School of Medicine, Sun Yat-Sen University, Guangzhou 510080, China

Received 27 March 2013; accepted 30 July 2013

©The Chinese Society of Oceanography and Springer-Verlag Berlin Heidelberg 2014

### Abstract

The coding product of alginate-c5-mannuronan-epimerase gene (*algG* gene) can catalyze the conversion of mannuronate to guluronate and determine the M/G ratio of alginate. Most of the current knowledge about genes involved in the alginate biosynthesis comes from bacterial systems. In this article, based on some algal and bacterial *algG* genes registered on GenBank and EMBL databases, we predicted 94 *algG* genes open reading frame (ORF) sequences of brown algae from the 1 000 Plant Transcriptome Sequencing Project (OneKP). By method of transcriptomic sequence analysis, gene structure and gene localization analysis, multiple sequence alignment and phylogenetic tree construction, we studied the algal *algG* gene family characteristics, the structure modeling and conserved motifs of AlgG protein, the origin of alginate biosynthesis and the variation incidents that might have happened during evolution in algae. Although there are different members in the algal *algG* gene family, almost all of them harbor the conserved epimerase region. Based on the phylogenetic analysis of *algG* genes, we proposed that brown algae acquired the alginate biosynthesis pathway from an ancient bacterium by horizontal gene transfer (HGT). Afterwards, followed by duplications, chromosome disorder, mutation or recombination during evolution, brown algal *algG* genes were divided into different types.

**Key words:** transcriptomic sequencing, alginate-c5-mannuronan-epimerase gene, gene family, alginate, phylogenetic analysis

**Citation:** Wang Ren, Wang Xumin, Zhang Yalan, Yu Jun, Liu Tao, Chen Shengping, Chi Shan. 2014. Origin and evolution of *alginate-c5-mannuronan-epimerase* gene based on transcriptomic analysis of brown algae. Acta Oceanologica Sinica, 33(2): 73–85, doi: 10.1007/s13131-014-0443-4

### 1 Introduction

Alginate is a linear polysaccharide molecule composed of  $\alpha$ -L-guluronate (G) and  $\beta$ -D-mannuronate (M), linked by  $\beta$ -1,4 glycosidic bonds. M residues and G residues are distributed continuously or alternately as blocks (GG blocks, MM blocks, MG blocks, etc.) (Panikkar and Brasch, 1996). Its biological sources are mainly from Phaeophyta algae (for example, *Saccharina japonica*, *Macrocystis pyrifera*, etc.) and a variety of opportunistic pathogens (for example, bacteria from genus *Pseudomonas* and *Azotobacter*) (Gorin and Spencer, 1966; Linker and Jones, 1966; Govan et al., 1981). Alginate, which is rich in guluronic acid, is also present in some calcareous red algae of family Coralliaceae, where it is involved in the deposition of calcium carbonate (Okasaki et al., 1982, 1984; Usov et al., 1995).

The main biological function of algae is to supply strength and flexibility to tissues (Smidsrød and Draget, 1996). Alginate

is the major matrix component of brown algal cell walls which weighs up to 45% in dry weight. M and G distributions vary according to the algal living environment, season and tissue age (Kloareg and Quatrano, 1988). For bacteria, alginate constitutes their sheath. In *Pseudomonas aeruginosa*, alginate is an important virulence factor, particularly in patients with the genetic disorder cystic fibrosis (CF) because the conversion of the strain to the alginate-overproducing mucoid phenotype can often result in chronic pulmonary infections (Pedersen et al., 1992). In *A. vinelandii*, it is thought to be linked with the formation of dormant cysts and the vegetative state capsule (Sadoff, 1975).

Alginate from various sources has shown significantly different gel properties such as the diversities of molecular size, M/G ratio, etc. This has formed the basis for numerous applications in biomaterial industry, medical applications and advanced biotechnology (Skjåk-Bræk and Espevik, 1996). In recent years,

Foundation item: The National High Technology Research and Development Program of China under contract No. 2012AA10A406; the National Natural Science Foundation of China under contract Nos 41206116, 31140070 and 31271397; Technology Project of Ocean and Fisheries of Guangdong Province under contract No. A201201E03; the Fundamental Research Funds for the Central Universities under contract No. 201262003; China Postdoctoral Science Foundation under contract No. 2011M501167; the algal transcriptome sequencing was supported by 1KP Project (www.onekp.com).

\*Corresponding author, E-mail: liutao@ouc.edu.cn, mgbl\_24@ouc.edu.cn

the need to produce alginate with controllable M/G ratio and controllable mechanical properties (elasticity and strength) is increasing. Alginate-c5-mannuronan-epimerase (AlgG protein) is the key enzyme which decides the M/G ratio of alginate. In the final step of alginate biosynthesis, AlgG protein (encoded by *algG* gene) can catalyze the conversion of  $\beta$ -D-mannuronate residues in the polymeric substrate polymannuronan to  $\alpha$ -L-guluronate residues. Synthetic efficiency and enzyme activity of AlgG are vital factors that influence the conversion of M to G.

Most of the current knowledge about the genes involved in the alginate biosynthesis comes from bacterial systems, mainly on genera *Pseudomonas* (Franklin et al., 1994) and *Azotobacter* (Rehm and Valla, 1997). As for brown algae, the study mainly focuses on the transcriptome and genome sequencing analysis.

*P. aeruginosa* has only one *algG* gene in its genome. The protein it encodes is located in the periplasmic, which functions without  $\text{Ca}^{2+}$  ion, converting mannuronate to guluronate (Chitnis and Ohman, 1990; Stephanie et al., 2005). The study also indicates that AlgG protein of *P. aeruginosa* can protect the alginate from degradation by alginate lyase in addition to its epimerization function (Gimmetstad et al., 2003). Meanwhile, two types of *algG* genes are discovered in *A. Vinelandii* - the *algE* genes, a family with seven members, which are named *algE1-7* according to gene cloning order (Ertesvåg et al., 1995; Svanem et al., 1999; Ertesvåg et al., 1994); and the *algG* gene, which exists as single copy (Rehm et al., 1996). The *algG* genes from both genera *Pseudomonas* and *Azotobacter* are located in the gene cluster involved in the alginate biosynthesis pathway. The protein sequence homology connection between *algG* gene and *algE* genes is weak, but the protein sequence homology within the *algE* gene family members is quite high. Besides, in *A. Vinelandii*, Svanem et al. (1999) cloned and expressed a highly homologous to *algE* family protein *algY*, but experiments proved no enzymatic activity.

As for the algal *algG* genes' study, Nyvall et al. (2003) separated 6 different *algG* gene fragments from the EST library of *Laminaria digitata*, two of which are full-length. Furthermore, expression of the *algG* genes was examined by northern-blot analysis and reverse transcriptase-polymerase chain reaction in *L. digitata* throughout a year. Also, Enzyme activity measurements were carried out in protoplasts from young sporophytes. Many *algG* genes are annotated in the recently published *Ectocarpus siliculosus* genome (Cook et al., 2010) and cDNA library. To a certain extent, this result indicates that in algae, a gene family of *algG* genes homologous to bacterial genes might exist. At present, studies on *algG* gene family and its encoded AlgG protein structure is very limited.

Sequence analysis and homology modeling of the *P. aeruginosa* AlgG protein suggest that the epimerase domain of this protein is a right-handed  $\beta$ -helix (RH $\beta$ H), a characteristic of proteins with a carbohydrate-binding and sugar hydrolase (CASH) domain (Stephanie et al., 2005). Similar structure is found in algal *algG* genes. All 26 *E. siliculosus* *algG* genes show significant differences in their amino acid sequence and gene structure, and uneven distribution in the chromosome. All of these raise the importance of more in-depth studies of the epimerase domain and the formation of the gene family.

Moreover, recently in the OneKP, the transcriptomes of more than 218 species of Chromista, including about 19 marine phaeophytes, 22 marine rhodophytes, 171 chlorophytes, 5 cryptophytes, 4 haptophytes and 5 glaucophytes, were sequenced. In

this study, by analyzing these RNA-sequencing data, we confirmed the existence of the predicted full-length *algG* gene ORF in the transcriptomic sequencing data of the brown algae. We carried out further the epimerase domain and phylogenetic analyses for the above mentioned sequences to understand the origin and evolution of *algG* genes, which can provide a theoretical basis for *algG* gene study in brown algae.

## 2 Materials and methods

### 2.1 Data access

The full-length known-cDNA sequences of *algG* genes were obtained from GenBank (<http://www.ncbi.nlm.nih.gov/guide/>) and EMBL database (<http://www.ebi.ac.uk/index.html>), which includes amino acid and nucleotide sequences of *algG* from bacterial genus *Pseudomonas* and brown algae *L. digitata*, *S. japonica*, *E. siliculosus*, and *algG*, *algE*, and *algY* from bacterial genus *Azotobacter*. Twenty-six *algG* gene sequences of *E. siliculosus* were named from *esi-algG1* to *esi-algG26*. The specific information of nucleic acid and amino acid sequences used in this paper are listed in Table S1.

### 2.2 Transcriptomic sequence analysis

The full-length known-cDNA sequences of *algG* acquired from GenBank and EMBL were used to search brown and red algae transcriptome sequences from the transcriptomic sequencing data of OneKP ([www.onekp.com](http://www.onekp.com)) for genes with similarity to *algG* genes with the BLASTN and TBLASTN algorithms. Matching sequence was manually checked for accuracy with the corresponding known-cDNA sequences. Protein coding sequences were deduced from the matching sequence and identified by comparing the translated sequence with known AlgG protein sequences. All the annotated cDNA open reading frame (ORF) sequences encoding *algG* are provided in Table 1.

### 2.3 Bioinformatics analysis

The 3D protein structures were predicted with the online analysis tool (<http://swissmodel.expasy.org/>) based on homology modeling (Arnold et al., 2006; Schwede et al., 2003; Guex and Peitsch, 1997). Gene structure is obtained by Gene Structure Display Server (<http://gsds.cbi.pku.edu.cn/>), which needs both submitting the genome sequence and its cds sequence. Gene localization is obtained by blasting the downloaded cds sequence with its corresponding genome data (<http://www.ncbi.nlm.nih.gov/genome/>) using the local Blast sequence alignment programs (PuTTY, secureFX).

### 2.4 Multiple sequence alignment and phylogenetic tree construction

The sequences were aligned using ClustalX 1.83 software (Thompson et al., 1997) and Bayesian analyses were performed using MrBayes3.1.2 software (Ronquist and Huelsenbeck, 2003) to investigate evolutionary relationships based on *algG* gene sequences. *algG* gene alignment includes 2878 amino acids from 137 taxa and uses sequences of *A. vinelandii* *algE1-E7* and *algY* as out-groups. Nucleotide substitution models were determined separately for each gene using Modeltest v. 3.7 (Posada and Crandall, 1998) in order to determine the best substitution model. Four separate sequence analyses used four Markov chains (using default heating values), which were run for 4 000 000 generations. Trees were sampled every 1 000 genera-

tions. The first 25% of trees were discarded as burn-in. The remaining trees were used to build a 50% majority rule consensus tree, accompanied with posterior probability values. Tree visualization was carried out using Tree View v. 1.6.5 (Page, 1996).

### 3 Results and analysis

#### 3.1 Transcriptomic sequence analysis

In the present study, we analyzed the transcriptomic sequencing data of OneKP Project. In total, 94 cDNA open reading frame (ORF) sequences, encoding AlgG protein, were annotated in brown algae where 6 are from *Colpomenia sinuosa*, 9 from *Desmarestia viridis*, 6 from *Dictyopteris undulate*, 4 from *Sargassum fusiforme*, 3 from *Ishige okamura*, 8 from *Saccharina sculpera*, 10 from *Petalonia fascia*, 6 from *Saccharina japonica*, 3 from *Sargassum hemiphyllum* var. *chinense*, 5 from *Sargassum henslowianum*, 1 from *Sargassum integerrimum*, 7 from *Sargassum thunbergii*, 4 from *Sargassum vachellianum*, 4 from *Sargassum horneri*, 4 from *Sargassum muticum*, 10 from *Scytosiphon lomentaria*, and 4 from *Scytosiphon dotyi*. Blast searches against the data of the OneKP Project also revealed that some red algae sequences have high similarity to brown algal algG genes but were incomplete in length (data not shown). More detailed information is listed in Table S1. It showed that GC contents varied from 49.9% to 62.9%, and amino acids varied from 289 to 899. These results reveal the same diversity of algG genes in *E. siliculosus* and other brown algae, which indicate that in brown algae, the epimerase function might be completed by these genes together.

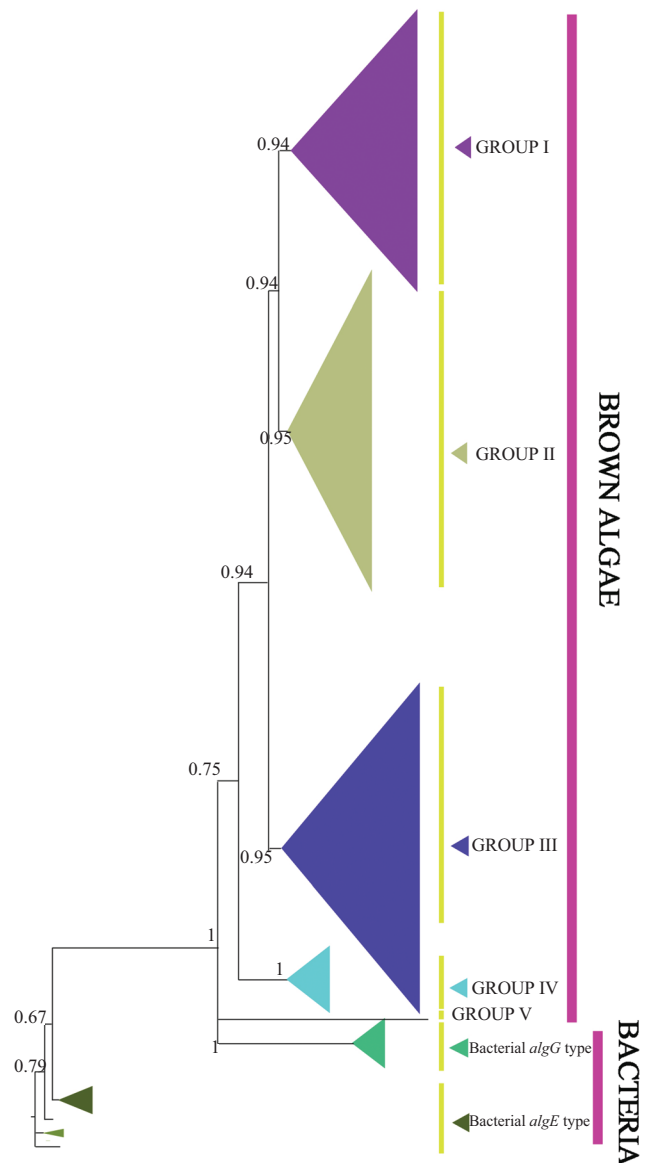
#### 3.2 Multiple sequence alignment and phylogenetic tree construction

AlgG is an important enzyme in the alginate biosynthesis pathway. Alignment analyses at the DNA and protein level showed that the *P. aeruginosa* algG gene (and its product) was quite different from its counterparts (*algE1* to *-E5*) in *A. vinelandii*. Even within the same species, the 26 algG gene members of *E. siliculosus* were quite different from each other. This can further prove the vast diversity of algG genes. Using NCBI/EMBL downloaded AlgG/AlgE and the predicted AlgG amino acids of OneKP, we established a phylogenetic tree based on the Bayesian analysis (schematic summary in Fig. 1, details in Fig. 2).

In the phylogenetic tree (Fig. 1), *A. Vinelandii* AlgE1-7 and AlgY were clustered outside the algG gene type, which indicated that AlgG and AlgE of *A. Vinelandii* may evolve independently. All algal AlgG sequences formed a cluster with bacterial AlgG (including AlgG of *A. Vinelandii* and *Pseudomonas*), which indicated algal AlgG may be more related to bacterial AlgG. These algal AlgG are divided into five groups (Fig. 2), named as Group I-V. Each species included more than one AlgG types. For instance, *E. siliculosus* contained five types and *S. japonica* contained four types. Group IV and V are clustered at the base of the algae branch, which indicate that they may be the original type of algG gene. Most of the predicted AlgG ORF are clustered in Group I, II and III, and included 24, 31 and 32 sequences, respectively. The three known full-length AlgG of *L. digitata* and *S. japonica* were clustered in Group I.

#### 3.3 Gene structure and gene localization

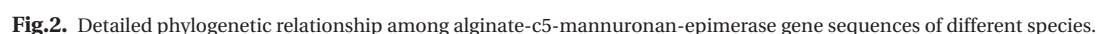
Characteristics of gene structure and gene distribution on the corresponding chromosome are important basis for the



**Fig.1.** Schematic summary of phylogenetic relationship among alginate-c5-mannuronan-epimerase gene sequences of different species.

analysis of the multi-genetic gene family members and the prediction for evolutionary events during the family forming process. Similarities and differences of the gene structure characteristics, including gene length span, homology of the exon sequences, intron fragments insert location and length, can provide useful information for preliminary study of the gene family members and quickly judge whether it is conserved or varied during evolution. Meanwhile, localization of the genes on different chromosomes can provide us clues about the variation, such as duplication, and chromosome disorder that happens during gene family evolution.

Twenty-six *algG* genes are annotated in the recently published *E. siliculosus* genome and cDNA library, with a big length span from 1 kb to 17 kb (Cook et al., 2010). Detailed analysis for these 26 *algG* gene structure and gene distribution will pro-





vide us further information about how the multi-genetic gene family form and about the evolutionary events that can happen during this process. Except for 8 genes (*esi-algG4*, *esi-algG5*, *esi-algG7*, *esi-algG15*, *esi-algG16*, *esi-algG25*, *esi-algG26*), the other 18 genes were located separately on 11 different chromosome linkage group, with about 1 to 3 genes per linkage group (Fig. 3b). These 8 genes, which are located on chromosome FN649760 and are constituted of redundant fragments, cannot mosaic. Genes located on different chromosomes reveal a highly diverse similarity in their sequence. On the contrary, genes located on the same chromosome show a high sequence homology. With the help of Gene Structure Display Server, we found that *E. siliculosus* *algG* gene structure are very complicated. The insertion number of intron vary from 3 to 20, and the inserted fragment length and inserted position are different as well (Fig. 3a). Structure diagram of *esi-algG15*, *esi-algG26*, *esi-algG9* and *esi-algG10* are very similar since they have a high homology in their exon sequences, introns insert location and length. However, the other *algG* genes differ with each other in gene structure in all respects.

### 3.4 Prediction of three dimensional structures for AlgG

Three dimensional structure of AlgG proteins were predicted by SWISS-MODEL (Arnold et al., 2006; Schwede et al., 2003; Guex and Peitsch, 1997). The homology model of SWISS-MODEL comprises of four main steps: identification of structural template(s), alignment of target sequence and template structure(s), model building, and model quality evaluation. These four steps can be repeated until a satisfying modelling result is achieved. Each of the steps requires specialized software and access to up-to-date protein sequence and structure databases, and can be invoked from within the workspace. We can find out whether all these AlgG proteins' homology model are

modeled on 2pyhB (A-module of *A. Vinelandii* *algE4*) or 1ru4A (chain A for pectate lyase *pel9A*). In the SwissModel Workspace, the QMEAN4 score is used to evaluate the generated models. The QMEAN4 score for the sequences we studied ranges from 0.286 to 0.399 (data not shown). Except for *esi-algG20*, the other AlgG proteins listed here have a similar three dimensional structure to its model (left in Fig. 4a), and this indicates a highly conserved structure during evolution. The differences of *esi-algG20* gene structure indirectly prove *algG* gene diversity in evolution.

## 4 Discussion

### 4.1 Alginate-c5-mannuronan-epimerases are encoded by a gene family in algae

As mentioned previously, the *algE* genes of *A. Vinelandii* form a gene family with seven members (Ertesvåg et al., 1995; Svanem et al., 1999; Ertesvåg et al., 1994), while the *algG* gene exists as a single copy (Rehm et al., 1996). The latter is homologous with *algG* gene in *P. aeruginosa*, and both of them are located in the gene cluster involved in the alginate biosynthesis pathway (Chitnis and Ohman, 1993; Penalzoza-Vazquez et al., 1997). Nyvall et al. (2003) isolated and characterized six *algG* genes from *L. digitata*, which exhibited significant similarities with bacterial *algG* genes. Also, we obtained 26 annotated *algG* genes from the recently published *E. siliculosus* genome (Cook et al., 2010). Additionally, 94 putative *algG* gene sequences were obtained from the transcriptomic sequencing data of brown algae in OneKP. All the forward evidences indicate that there might exist an *algG* gene family in brown algae.

The phylogenetic analysis showed that all the algal *algG* genes are clustered into five groups (Fig. 1 and Fig. 2), and all of them are genetically related to the bacterial *algG* type. The

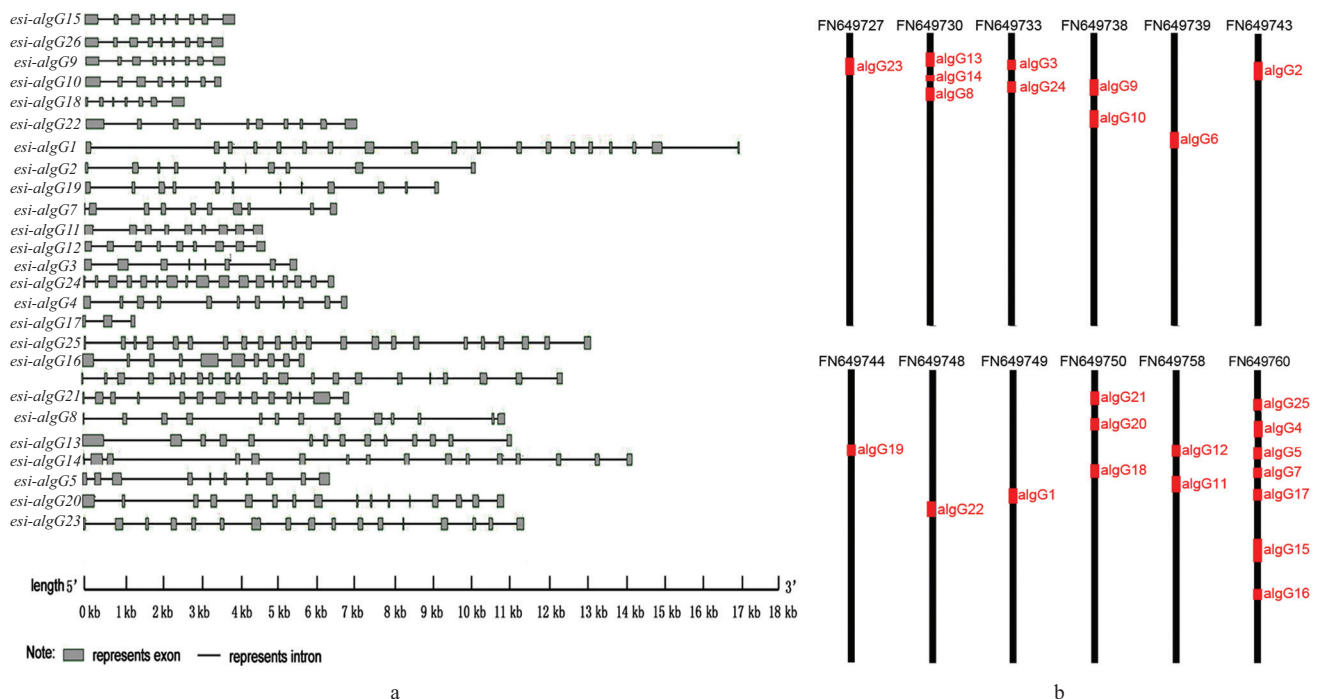
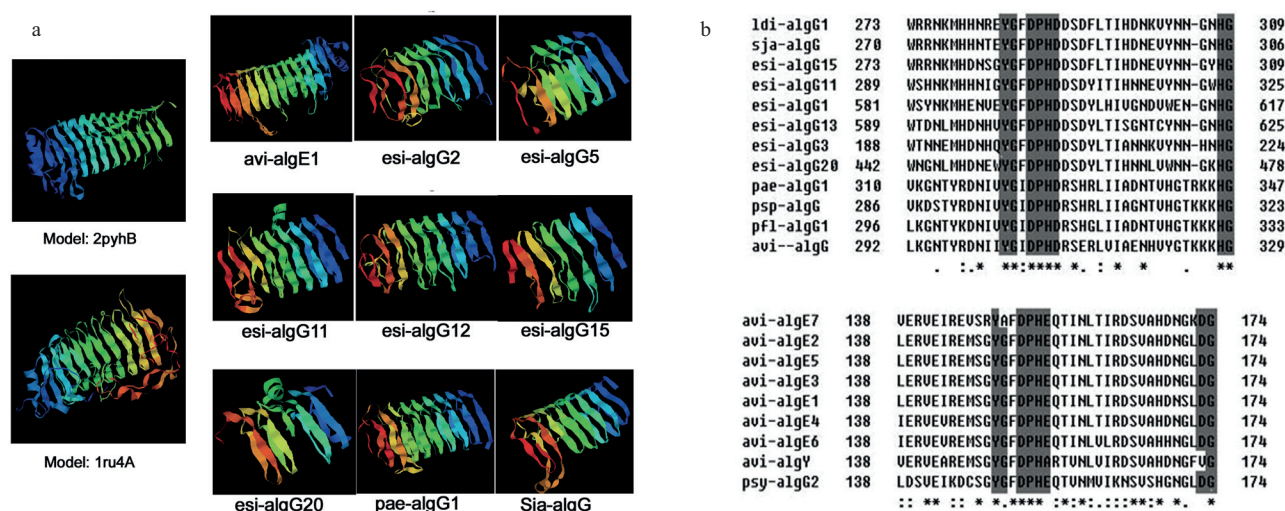


Fig.3. Structure diagram (a) and *algG* gene localization (b) of *Ectocarpus siliculosus*.



**Fig. 4.** Three dimensional structure for AlgG (a) and multiple sequence alignment of conservative region from different species (b). \* represents amino acid sequences that are exactly the same, : more conserved amino acid sequence and . less conserved amino acid sequence. Region in grey represents conserved regions.

differences in the gene structure (Fig. 3a) in terms of intron number, intron length and inserted position, and the uneven distribution of *algG* genes on chromosomes of *E. siliculosus* (Fig. 3b) indicate the *algG* gene diversity in *E. siliculosus*. The coexistence of the complex family members, to a certain extent, has proven that the direction of differentiation of *algG* gene in brown algae is variable. However, most of the members maintained their conserved region (similar with the 324-DPHD motif from bacterial genus *Pseudomonas* which is vital for epimerization) (Fig. 4b), which indicates some conservation also exists. Morch et al. (2008) proved that seven epimerases, *AlgE1-7* of *A. Vinelandii*, possess differences in their epimerization pattern varying from their alginate formation to their mechanical properties. Different members of algal *algG* gene family might act in the same mode. The epimerase function might be completed by these genes effects together. We will discuss the origin of the *algG* gene family and the variation in incidents that happen during evolution in the following paragraphs.

#### 4.2 Structure modeling of AlgG and conserved motifs of epimerase

Alginate biosynthesis mode is proposed to be a multicomponent scaffold for polymerization, modification and export (Jain et al., 2003). Several studies were carried out to support this model (Jain et al., 2003; Jain and Ohman, 1998; Robles-Price et al., 2004; Gimmestad et al., 2003). Stephanie et al. (2005) proposed that AlgG protein contains a repeating sequence in its C-terminal that is predicted to be a right-handed  $\beta$ -helix (RH $\beta$ H), a characteristic of proteins with a carbohydrate-binding and sugar hydrolase (CASH) domain. To further characterize AlgG protein and its role in alginate biosynthesis, we performed several sequence analyses and structure modeling for it. We predicted the three dimensional structure of AlgG protein by SWISS-MODEL (Arnold et al., 2006; Schwede et al., 2003; Guex and Peitsch, 1997). It turns out the C-terminus of AlgG protein is folded into a right-handed  $\beta$ -helix (RH $\beta$ H), and the most matchable models are *A. Vinelandii* AlgE4 A module (Ertesvåg

et al., 1998; Ertesvåg and Valla, 1999; Rozeboom et al., 2008) and pectate lyase C of *Erwinia chrysanthem* (Jenkins et al., 2004). The predicted three dimensional structure is shown in Fig. 4a. Similar with other CASH proteins, AlgG showed the presence of a shallow groove, which might accommodate the linear polysaccharide. Although AlgG of *P. aeruginosa*, *A. Vinelandii* and brown algae have little amino acid sequence identity, at least one motif have been identified that showed a high sequence similarity. This motif corresponds to the *P. Aeruginosa* 324-DPHD motif, and is predicted to be a substrate binding groove (Fig. 4b). Experiments proved that point mutations in this motif disrupt epimerase activity in *P. Aeruginosa*, suggesting that this conserved region might be vital in all AlgG epimerase activity (Stephanie et al., 2005). Particularly, the multiple sequence alignment of the NCBI downloaded sequences of *algG* for both bacteria and algae proved that there exist a similar conserved region with the *P. aeruginosa* 324-DPHD motif, indicating that this region might also be vital in algae epimerase activity (Fig. 4b). This region of brown algae and bacteria is located in the similar position, and shares similar amino acid composition, which provides us with useful knowledge to deeply study the AlgG protein catalytic mechanism in brown algae.

#### 4.3 Origin of alginate biosynthesis in brown algae

In the present study, it is universally believed that Stramenopile plastids arose via a secondary endosymbiosis, in which a unicellular red alga was engulfed by an ancestral host (Reyes-Prieto et al., 2007). The red algae and green plants have originated from one single primary endosymbiosis, between a colorless eukaryotic host and a photosynthetic cyanobacteria (Douglas, 1998; Lang et al., 1999; Moreira et al., 2000). In order to explore the origin of alginate evolution, Nyvall et al. (2003) proposed that red algae acquired the alginate pathway from cyanobacteria via primary plastid endosymbiosis and that, subsequently, this metabolism was passed on to brown algae via secondary endosymbiosis. Existence of alginate in red algae (Okasaki et al., 1982, 1984; Usov et al., 1995) might support this view. However, this evolutionary scenario suffers from several

weaknesses. Firstly, no alginate-biosynthetic genes were found in extant cyanobacteria (<http://wit.integratedgenomics.com/GOLD>), which suggest the alginate-biosynthesis pathway could not be of endosymbiotic origin. Secondly, if this scenario is true, during the evolution, it requires multiple losses of the alginate pathway in green algae, plants, the majority of red algae, diatoms and Oomycetes which belong to Stramenopiles (Gurvan et al., 2010).

In our phylogenetic analysis (Fig. 1 and Fig. 2), the algal *algG* gene sequences firstly formed a monophyletic cluster which showed a close relationship with the *algG* type of bacteria, which is quite distant from the *algE1-7*. Then, the algae *algG* genes were divided into five groups named Group I–V (Fig. 2). Sequences were distributed unevenly in different groups, where most sequences are distributed in Groups I, II and III compared to Groups IV and V. Although gene structure analysis showed high variety in *E. siliculosus algG*, *esi-algG9*, *esi-algG18*, *esi-algG15* and *esi-algG26*, which all belonged to Group I, they have significant similarity in their inserted intron number, fragment length and inserted position (Fig. 3a). This might indicate that Group I is the main functional *algG* genes. The three dimensional structure and the existence of epimerase conserved region also support this hypothesis. Therefore, we proposed that brown algal *algG* or even the whole alginate metabolic pathway would have arisen from ancient bacteria (of which is related to both *A. vinelandii* and *P. aeruginosa*) via horizontal gene transfer (HGT). Then, it diverged into a gene family during evolution. Altogether the bacterial and brown algal *AlgG* display a similar structure and share the same catalytic domain, which indicate that they belong to the same structural family and have arisen from a common ancestor. The alginate pathway in red algae remains uncharacterized at both the biochemical and gene level, so the origin of alginate in red algae is not discussed here.

#### 4.4 Gene variation

Algal *AlgG* enzymes are encoded by a multiple gene family, and the family members vary from each other. Algae must survive in complex ocean environment, thus algae would benefit from having a large number of diverse genes to enable it to sense and respond to this correspondingly diverse and ever-changing set of environmental conditions. Multiple different mechanisms, including neofunctionalization (Lynch and Force, 2000), subfunctionalization (Hughes and Liberles, 2007), pseudogenization and dosage balance, may account for this functional divergence of *algG* gene. Gene duplication is an important evolutionary mechanism which provides genomes with the genetic raw material for the emergence of genes with new or altered functions (Flagel and Wendel, 2009). The large size of *algG* gene family in algae may result from either an increased rate of gene duplication or a low rate of loss. In this article, we proposed a bold assumption for its variation during evolution. All these thoughts were based on *E. siliculosus algG* gene structure and gene localization. The original type was introduced into the genome of brown algal ancestor, and then went through a chromosome variation, which the chromosome may replicate at least four times to create 16 chromosome copies. As most of the duplicated genes were initially redundant in function, four copies were lost. The remaining 12 copies went through gene duplication once or twice before any gene losses happened. The rest of the genes diverged into different types and may derive new structural or regulatory functions allowing genetic inno-

vation to adapt to the new environment. So on average, 1 to 3 *algG* genes per chromosome linkage group is found in *E. siliculosus*. As we assumed, much variation has happened during the brown algal ancestor phase. It was proven with the phylogenetic tree that the predicted ORF of *algG* genes for brown algae were separately clustered into its four groups (Fig. 1). Gene duplication events are important sources of novel gene functions. However, more often than not, a duplicated gene may lose its function and become a pseudogene. The same situation occurs to *esi-algG23*, *esi-algG5* and *esi-algG17*. *Esi-algG3* and *esi-algG24* (*esi-algG9* and *esi-algG10*, *esi-algG11* and *esi-algG12*) are located on the same chromosome group FN649733 (FN649738 and FN649758), and were clustered together in smaller groups in the phylogenetic tree (Fig. 2). This might be the result of a single gene duplication. However, *esi-algG8*, *esi-algG13* and *esi-algG14* have the same signature, which might be the result of two gene duplication. Going through gene duplication twice can later on lead to gene loss, resulting in only 3 copies left on the same chromosome. This gene linkage phenomenon can be sufficient evidence for gene duplication incidents. However, it should be noted that this is a preliminary estimation, and further analysis of the general rate of protein evolution at the genomic level will be required for a more precise calculation.

#### References

- Arnold K L, Bordoli J, Kopp, et al. 2006. The SWISS-MODEL Workspace: a web-base environment for protein structure homology modeling. *Bioinformatics*, 22: 195–201
- Chitnis C E, Ohman D E. 1990. Cloning of *Pseudomonas aeruginosa algG*, which controls alginate structure. *Bacteriol*, 172(6): 2894–2900
- Chitnis C E, Ohman D E. 1993. Genetic analysis of the alginate biosynthetic gene cluster of *Pseudomonas aeruginosa* shows evidence of an operonic structure. *Mol Microbiol*, 8: 583–590
- Cook J M, Sterck L, Rouzé P, et al. 2010. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature*, 465: 617–621
- Douglas S E. 1998. Plastid evolution: origins, diversity, trends. *Curr Opin Genet Dev*, 8: 655–661
- Ertesvåg H, Doseth B, Larsen B, et al. 1994. Cloning and expression of an *Azotobacter vinelandii* mannuronan C-5-epimerase gene. *J Bacteriol*, 176: 2846–2853
- Ertesvåg H, Høidal H K, Hals I K, et al. 1995. A family of modular type mannuronan C-5-epimerase genes controls alginate structure in *Azotobacter vinelandii*. *Mol Microbiol*, 9: 719–731
- Ertesvåg H, Høidal H K, Skjåk-Bræk G, et al. 1998. The *Azotobacter vinelandii* mannuronan C-5-epimerase *AlgE1* consists of two separate catalytic domains. *J Biol Chem*, 273: 30927–30932
- Ertesvåg H, Valla S. 1999. The A modules of the *Azotobacter vinelandii* mannuronan-C-5-epimerase *AlgE1* are sufficient for both epimerization and binding of Ca<sup>2+</sup>. *J Bacteriol*, 181: 3033–3038
- Flagel L, Wendel J. 2009. Gene duplication and evolutionary novelty in plants. *New Phytologist*, 183: 557–564
- Franklin M J, Chitnis C E, Gacesa P, et al. 1994. *Pseudomonas aeruginosa AlgG* is a polymer level alginate C5-mannuronan epimerase. *J Bacteriol*, 176: 1821–1830
- Gimmestad M, Sletta H, Ertesvåg H, et al. 2003. The *Pseudomonas fluorescens* *AlgG* protein, but not its mannuronan C-5-epimerase activity, is needed for alginate polymer formation. *J Bacteriol*, 185: 3515–3523
- Gorin P A, Spencer J F. 1966. Exocellular alginic acid from *Azotobacter vinelandii*. *Can J Chem*, 44: 993–998
- Govan J R W, Fyfe J A M, Jarman T R. 1981. Isolation of alginate producing mutants of *Pseudomonas fluorescens*, *Pseudomonas putida* and *Pseudomonas mendocina*. *J Gen Microbiol*, 125: 217–220



- Guex N, Peitsch M C. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modelling. *Electrophoresis*, 18: 2714–2723
- Gurvan M, Thierry T, Delphine S, et al. 2010. The cell wall polysaccharide metabolism of the brown alga *Ectocarpus siliculosus*. Insights into the evolution of extracellular matrix polysaccharides in Eukaryotes. *New Phytologist*, 188: 82–97
- Rozeboom H J, Bjerkkan T M, Kalk K H, et al. 2008. Structural and Mutational Characterization of the Catalytic A-module of the Mannuronan C-5-epimerase AlgE4 from *Azotobacter vinelandii*. *J Biol Chem*, 283: 23819–23828
- Hughes T, Liberles D A. 2007. The pattern of evolution of smaller-scale gene duplicates in mammalian genomes is more consistent with neo-than subfunctionalisation. *J Mol Evol*, 65(5): 574–588
- Jain S, Franklin M J, Ertesvag H, et al. 2003. The dual roles of AlgG in C-5-epimerization and secretion of alginate polymers in *Pseudomonas aeruginosa*. *Mol Microbiol*, 47: 1123–1133
- Jain S, Ohman D E. 1998. Deletion of algK in mucoid *Pseudomonas aeruginosa* blocks alginate polymer formation and results in uronic acid secretion. *J Bacteriol*, 180: 634–641
- Jenkins J, Shevchik V E, Hugouvieux-Cotte-Pattat N, et al. 2004. The crystal structure of pectate lyase Pel9A from *Erwinia chrysanthemi*. *J Biol Chem*, 279: 9139–9145
- Kloareg B, Quatrano R S. 1988. Structure of the cell walls of marine algae and ecophysiological functions of the matrix polysaccharides. *Oceanogr Mar Biol Annu Rev*, 1988: 259–315
- Lang B F, Seif E, Gray M W, et al. 1999. A comparative genomics approach to the evolution of eukaryotes and their mitochondria. *J Eukaryot Microbiol*, 46: 320–326
- Linker A, Jones R S. 1966. A new polysaccharide resembling alginic acid isolated from pseudomonas. *J Biol Chem*, 241: 3845–385
- Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics*, 154(1): 459–473
- Morch Y A, Holtan S, Donati I, et al. 2008. Mechanical properties of C-5 epimerized alginates. *Biomacromolecules*, 9: 2360–2368
- Moreira D, Le Guyader H, Philippe H. 2000. The origin of red algae and the evolution of chloroplasts. *Nature*, 405: 69–72
- Nyvall P, Erwan C, Claire B, et al. 2003. Characterization of Mannuronan C-5-Epimerase Genes from the Brown Alga *Laminaria digitata*. *Plant Physiology*, 133(2): 726–735
- Okasaki M, Furuya K, Tsukayama K, et al. 1982. Isolation and identification of alginic acid from a calcareous red alga *Serraticardia maxima*. *Botanica Marina*, 25: 123–131
- Okasaki M, Shiroto C, Furuya K. 1984. Relationship between the location of polyuronides and calcification sites in the calcareous red algae *Serraticardia maxima* and *Lithothamnion japonica* (Rhodophyta, Corallinaceae). *Jap J Phycol*, 32: 364–372
- Page R D. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*, 12: 357–358
- Panikkar R, Brasch D J. 1996. Composition and block structure of alginates from New Zealand brown seaweeds. *Carbohydr Res*, 293: 119–132
- Pedersen S S, Hoiby N, Espersen F, et al. 1992. Role of alginate in infection with mucoid *Pseudomonas aeruginosa* in cystic fibrosis. *Thorax*, 47: 6–13
- Penaloza-Vazquez A, Kidambi S P, Chakrabarty A M, et al. 1997. Characterization of the alginate biosynthetic gene cluster in *Pseudomonas syringae* pv. *syringae*. *J Bacteriol*, 179: 4464–4472
- Posada D, Crandall K A. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics*, 14: 817–818
- Rehm B H A, Ertesvåg H, Valla S. 1996. A new *Azotobacter vinelandii* mannuronan C-5-epimerase gene (*algG*) is part of an alg gene cluster physically organized in a manner similar to that in *Pseudomonas aeruginosa*. *Bacteriol*, 178: 5884–5889
- Rehm B H A, Valla S. 1997. Bacterial alginates: biosynthesis and applications. *Appl Microbiol Biotechnol*, 48: 281–288
- Reyes-Prieto A, Weber A P, Bhattacharya D. 2007. The origin and establishment of the plastid in algae and plants. *Annual Review of Genetics*, 41: 147–168
- Robles-Price A, Wong T Y, Sletta H, et al. 2004. *AlgX* is a periplasmic protein required for alginate biosynthesis in *Pseudomonas aeruginosa*. *J Bacteriol*, 186: 7369–7377
- Ronquist F, Huelsenbeck J P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19: 1572–1574
- Sadoff H L. 1975. Encystment and germination of *Azotobacter vinelandii*. *Bacteriol Rev*, 39: 516–539
- Schwede T, Kopp J, Guex N, et al. 2003. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research*, 31: 3381–3385
- Skjåk-Bræk G, Espevik T. 1996. Application of alginate gels in biotechnology and biomedicine. *Carbohydr Eur*, 14: 19–25
- Smidsrød O, Draet K I. 1996. Chemistry and physical properties of alginates. *Carbohydr Eur*, 14: 6–13
- Stephanie A, Douthit D, Mensur E O, et al. 2005. Epimerase Active Domain of *Pseudomonas aeruginosa* AlgG, a Protein That Contains a Right-Handed  $\beta$ -Helix. *Journal of Bacteriology*, 187(13): 4573–4583
- Svanem B I, Skjåk-Bræk G, Ertesvåg H, et al. 1999. Cloning and expression of three new *Azotobacter vinelandii* genes closely related to a previously described gene family encoding mannuronan C-5-epimerases. *Bacteriol*, 181: 68–77
- Thompson J D, Gibson T J, Plewniak F, et al. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*, 25: 4876–4882
- Usov A I, Bilan M I, Klochova N G. 1995. Polysaccharides of algae. 48. Polysaccharide composition of several calcareous red algae: isolation of alginate from *Corallina pilulifera* P. et R. (Rhodophyta, Corallinaceae). *Botanica Marina*, 38: 43–52



Table S1. Information of the *algG* genes used in the article

Abbreviation	Species name	Taxonomy	Database number of the amino acid	Database number of the nucleotide	GC content of cDNA sequence/%	Number of the coded amino acid
avi- <i>algE1</i>	<i>Azotobacter vinelandii</i>	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Azotobacter	Q44494 <sup>e</sup>	AAA87311 <sup>e</sup>	66.9	1402
avi- <i>algE2</i>	<i>Azotobacter vinelandii</i>	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Azotobacter	Q44495 <sup>e</sup>	AAA87312 <sup>e</sup>	67.5	996
avi- <i>algE3</i>	<i>Azotobacter vinelandii</i>	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Azotobacter	Q44496 <sup>e</sup>	AAA87313 <sup>e</sup>	67.6	1838
avi- <i>algE4</i>	<i>Azotobacter vinelandii</i>	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Azotobacter	Q44493 <sup>e</sup>	AAA87310 <sup>e</sup>	67.9	552
avi- <i>algE5</i>	<i>Azotobacter vinelandii</i>	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Azotobacter	Q44492 <sup>e</sup>	AAA87309 <sup>e</sup>	67.5	996
avi- <i>algE6</i>	<i>Azotobacter vinelandii</i>	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Azotobacter	Q9ZFHO <sup>e</sup>	AAD04920 <sup>e</sup>	69.3	873
avi- <i>algE7</i>	<i>Azotobacter vinelandii</i>	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Azotobacter	Q9ZFG9 <sup>e</sup>	AAD04921 <sup>e</sup>	64.3	885
avi- <i>algY</i>	<i>Azotobacter vinelandii</i>	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Azotobacter	Q9ZFG8 <sup>e</sup>	AAD04922 <sup>e</sup>	65.8	531
pae- <i>algG1</i>	<i>Pseudomonas aeruginosa</i>	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas	Q51371 <sup>e</sup>	CAA61231 <sup>e</sup>	63	524
pae- <i>algG2</i>	<i>Pseudomonas aeruginosa</i>	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas	AEV1P9 <sup>e</sup>	ABR82097 <sup>e</sup>	66.4	538
pae- <i>algG3</i>	<i>Pseudomonas aeruginosa</i>	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas	E2ZXU0 <sup>e</sup>	EFQ39969 <sup>e</sup>	65.9	520
psp- <i>algG</i>	<i>Pseudomonas</i> sp.	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas	Q6YB90 <sup>e</sup>	AA63149 <sup>e</sup>	61.5	518
psy- <i>algG2</i>	<i>Pseudomonas syringae</i>	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas	EGH13063.1 <sup>n</sup>	AEA01000436.1 <sup>n</sup>	61.2	619
pfl- <i>algG1</i>	<i>Pseudomonas fluorescens</i>	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas	EGH13064.1 <sup>n</sup>	AEA01000436.1 <sup>n</sup>	60.8	984
pfl- <i>algG2</i>	<i>Pseudomonas fluorescens</i>	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas	Q4KHY3 <sup>n</sup>	.....	-----	528
pfl- <i>algG3</i>	<i>Pseudomonas fluorescens</i>	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas	Q3KHQ8 <sup>n</sup>	.....	-----	522
pfl- <i>algG4</i>	<i>Pseudomonas fluorescens</i>	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas	E2XLX8 <sup>n</sup>	.....	-----	521
pbr- <i>algG</i>	<i>Pseudomonas brassicacearum</i>	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas	C3KDDZ <sup>n</sup>	.....	-----	528
pai- <i>algG</i>	<i>Pseudomonas alcityphenolia</i>	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas	YP_004333933.1 <sup>n</sup>	.....	-----	1870
ppu- <i>algG</i>	<i>Pseudomonas putida</i>	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas	D5KU48 <sup>e</sup>	.....	-----	521
esi- <i>algG1</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	Q88NC9 <sup>e</sup>	AA66907 <sup>e</sup>	61.9	518
esi- <i>algG2</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBJ25887 <sup>n</sup>	FN648214 <sup>n</sup>	57.4	723
esi- <i>algG3</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBJ26770 <sup>n</sup>	FN648652 <sup>n</sup>	57.9	362
esi- <i>algG4</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBJ28911 <sup>n</sup>	FN647882 <sup>n</sup>	59.6	368
esi- <i>algG5</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBJ30859 <sup>n</sup>	FN647882 <sup>n</sup>	56.1	447
esi- <i>algG6</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBJ31264 <sup>n</sup>	FN648422 <sup>n</sup>	59.6	466
esi- <i>algG7</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBJ32036 <sup>n</sup>	FN648491 <sup>n</sup>	63.1	845
esi- <i>algG8</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBJ32091 <sup>n</sup>	FN648495 <sup>n</sup>	55.2	447
esi- <i>algG9</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBJ32281 <sup>n</sup>	FN648516 <sup>n</sup>	60.9	505
esi- <i>algG10</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBJ32282 <sup>n</sup>	FN648516 <sup>n</sup>	61.8	512
esi- <i>algG11</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBJ32775 <sup>n</sup>	FN648561 <sup>n</sup>	58.6	559
esi- <i>algG12</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBJ32776 <sup>n</sup>	FN648561 <sup>n</sup>	60.6	552
esi- <i>algG13</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBJ33360 <sup>n</sup>	FN648670 <sup>n</sup>	65	785
esi- <i>algG14</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBJ33364 <sup>n</sup>	FN648670 <sup>n</sup>	62.5	712
esi- <i>algG15</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBJ33482 <sup>n</sup>	FN648702 <sup>n</sup>	61.8	508
esi- <i>algG16</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBJ34033 <sup>n</sup>	FN649120 <sup>n</sup>	61	632
esi- <i>algG17</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBJ48575 <sup>n</sup>	FN648586 <sup>n</sup>	57.2	112
esi- <i>algG18</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBN74322 <sup>n</sup>	FN648586 <sup>n</sup>	58.6	305
esi- <i>algG19</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBN74601 <sup>n</sup>	FN648487 <sup>n</sup>	58	390
esi- <i>algG20</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBN77791 <sup>n</sup>	FN648532 <sup>n</sup>	62.7	674
esi- <i>algG21</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBN77791 <sup>n</sup>	FN648916 <sup>n</sup>	57.9	679
esi- <i>algG22</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBN78412 <sup>n</sup>	FN647822 <sup>n</sup>	62.1	593
esi- <i>algG23</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBN78838 <sup>n</sup>	FN648032 <sup>n</sup>	59.2	745
esi- <i>algG24</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBN79256 <sup>n</sup>	FN647683 <sup>n</sup>	60.1	885
esi- <i>algG25</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBN80173 <sup>n</sup>	FN647849 <sup>n</sup>	60.8	906
esi- <i>algG26</i>	<i>Ectocarpus siliculosus</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Ectocarpales; Ectocarpaceae; Ectocarpus	CBJ33483 <sup>n</sup>	FN648702 <sup>n</sup>	61.7	508
ldi- <i>algG1</i>	<i>Laminaria digitata</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Laminariales; Laminariaceae; Laminaria	CAD42945.1 <sup>n</sup>	AJ496449.1 <sup>n</sup>	57	502
ldi- <i>algG6</i>	<i>Laminaria digitata</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Laminariales; Laminariaceae; Laminaria	CAD42950.1 <sup>n</sup>	AJ496454.1 <sup>n</sup>	56.2	499
sia- <i>algG</i>	<i>Saccharina japonica</i>	Eukaryota; Stramenopiles; PX clade; Phaeophyceae; Laminariales; Laminariaceae; Saccharina	BAF80877.1 <sup>n</sup>	AB299380.1 <sup>n</sup>	58.6	499

to be continued

Continued from Table S1

Abbreviation	Species name	Taxonomy	Database number of the amino acid	Database number of the nucleotide	GC content of cDNA sequence/%	Number of the coded amino acid	Tissue description	Voucher data (indicate if pending)
scaffold-QLMZ-2008026-Colpomenia sinuosa	<i>Colpomenia sinuosa</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Colpomenia	scaffold-QLMZ-2008027		56.2	426	leaves	2012020101
scaffold-QLMZ-2008025-Colpomenia sinuosa	<i>Colpomenia sinuosa</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Colpomenia	scaffold-QLMZ-2008026		60.5	426	leaves	2012020101
scaffold-QLMZ-2008027-Colpomenia sinuosa	<i>Colpomenia sinuosa</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Colpomenia			60.2	446	leaves	2012020101
scaffold-QLMZ-2013947-Colpomenia sinuosa	<i>Colpomenia sinuosa</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Colpomenia			58.4	494	leaves	2012020101
scaffold-QLMZ-2017500-Colpomenia sinuosa	<i>Colpomenia sinuosa</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Colpomenia			59.9	501	leaves	2012020101
scaffold-QLMZ-2003153-Colpomenia sinuosa	<i>Colpomenia sinuosa</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Colpomenia			61.4	636	leaves	2012020101
scaffold-FSQE-2008317-Desmarestia viridis	<i>Desmarestia viridis</i>	Ochrophyta; Phaeophyceae; Desmarestiales; Desmarestiaceae; Desmarestia			57.8	426	branches and leaves	201203022
scaffold-FSQE-2001029-Desmarestia viridis	<i>Desmarestia viridis</i>	Ochrophyta; Phaeophyceae; Desmarestiales; Desmarestiaceae; Desmarestia			52.4	438	branches and leaves	201203022
scaffold-FSQE-2001027-Desmarestia viridis	<i>Desmarestia viridis</i>	Ochrophyta; Phaeophyceae; Desmarestiales; Desmarestiaceae; Desmarestia			50.6	440	branches and leaves	201203022
scaffold-FSQE-2001028-Desmarestia viridis	<i>Desmarestia viridis</i>	Ochrophyta; Phaeophyceae; Desmarestiales; Desmarestiaceae; Desmarestia			53.8	441	branches and leaves	201203022
scaffold-FSQE-2000750-Desmarestia viridis	<i>Desmarestia viridis</i>	Ochrophyta; Phaeophyceae; Desmarestiales; Desmarestiaceae; Desmarestia			53.8	441	branches and leaves	201203022
scaffold-FSQE-2000752-Desmarestia viridis	<i>Desmarestia viridis</i>	Ochrophyta; Phaeophyceae; Desmarestiales; Desmarestiaceae; Desmarestia			54.7	491	branches and leaves	201203022
scaffold-FSQE-2004718-Desmarestia viridis	<i>Desmarestia viridis</i>	Ochrophyta; Phaeophyceae; Desmarestiales; Desmarestiaceae; Desmarestia			55.3	498	branches and leaves	201203022
scaffold-FSQE-2002614-Desmarestia viridis	<i>Desmarestia viridis</i>	Ochrophyta; Phaeophyceae; Desmarestiales; Desmarestiaceae; Desmarestia			53.8	498	branches and leaves	201203022
scaffold-LRFE-2006001-Dictyopteris undulata	<i>Dictyopteris undulata</i>	Ochrophyta; Phaeophyceae; Dictyotales; Dictyotaceae; Dictyopteris			56.3	338	branches and leaves	201203022
scaffold-LRFE-2006000-Dictyopteris undulata	<i>Dictyopteris undulata</i>	Ochrophyta; Phaeophyceae; Dictyotales; Dictyotaceae; Dictyopteris			55.7	420	leaves	201203023
scaffold-LRFE-2016070-Dictyopteris undulata	<i>Dictyopteris undulata</i>	Ochrophyta; Phaeophyceae; Dictyotales; Dictyotaceae; Dictyopteris			56.1	472	leaves	201203023
scaffold-LRFE-2015641-Dictyopteris undulata	<i>Dictyopteris undulata</i>	Ochrophyta; Phaeophyceae; Dictyotales; Dictyotaceae; Dictyopteris			54.7	474	leaves	201203023
scaffold-LRFE-2012622-Dictyopteris undulata	<i>Dictyopteris undulata</i>	Ochrophyta; Phaeophyceae; Dictyotales; Dictyotaceae; Dictyopteris			54.7	484	leaves	201203023
scaffold-LRFE-2006206-Dictyopteris undulata	<i>Dictyopteris undulata</i>	Ochrophyta; Phaeophyceae; Dictyotales; Dictyotaceae; Dictyopteris			55.8	488	leaves	201203023
scaffold-LDRY-2019956-Hizikia fusiforme	<i>Hizikia fusiforme</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; <i>Sargassum</i>			54.7	424	branches and leaves	2011040004
scaffold-LDRY-2020267-Hizikia fusiforme	<i>Hizikia fusiforme</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; <i>Sargassum</i>			53.1	439	branches and leaves	2011040004
scaffold-LDRY-2019221-Hizikia fusiforme	<i>Hizikia fusiforme</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; <i>Sargassum</i>			54.1	513	branches and leaves	2011040004
scaffold-LDRY-2011976-Hizikia fusiforme	<i>Hizikia fusiforme</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; <i>Sargassum</i>			53.1	531	branches and leaves	2011040004
scaffold-ATPT-2002334-Ishige okamurai	<i>Ishige okamurai</i>	Ochrophyta; Phaeophyceae; Ishigeales; Ishigeaceae; Ishige			53.3	409	branches	201202108
scaffold-ATPT-2014676-Ishige okamurai	<i>Ishige okamurai</i>	Ochrophyta; Phaeophyceae; Ishigeales; Ishigeaceae; Ishige			50.5	486	branches	201202108
scaffold-ATPT-2003397-Ishige okamurai	<i>Ishige okamurai</i>	Ochrophyta; Phaeophyceae; Ishigeales; Ishigeaceae; Ishige			52.8	494	branches	201202108
scaffold-RAPY-2016442-Kjellmaniella crassifolia	<i>Kjellmaniella crassifolia</i>	Ochrophyta; Phaeophyceae; Laminariales; Laminariaceae; Kjellmaniella			55	421	branches	201202108
scaffold-RAPY-2014805-Kjellmaniella crassifolia	<i>Kjellmaniella crassifolia</i>	Ochrophyta; Phaeophyceae; Laminariales; Laminariaceae; Kjellmaniella			59.8	438	branches	201202108
scaffold-RAPY-2016200-Kjellmaniella crassifolia	<i>Kjellmaniella crassifolia</i>	Ochrophyta; Phaeophyceae; Laminariales; Laminariaceae; Kjellmaniella			58.6	467	branches	201202108
scaffold-RAPY-20088005-Kjellmaniella crassifolia	<i>Kjellmaniella crassifolia</i>	Ochrophyta; Phaeophyceae; Laminariales; Laminariaceae; Kjellmaniella			59.9	486	branches	201202108
scaffold-RAPY-2002705-Kjellmaniella crassifolia	<i>Kjellmaniella crassifolia</i>	Ochrophyta; Phaeophyceae; Laminariales; Laminariaceae; Kjellmaniella			56.5	500	branches	201202108
scaffold-RAPY-2012031-Kjellmaniella crassifolia	<i>Kjellmaniella crassifolia</i>	Ochrophyta; Phaeophyceae; Laminariales; Laminariaceae; Kjellmaniella			60.7	501	branches	201202108
scaffold-RAPY-2016425-Kjellmaniella crassifolia	<i>Kjellmaniella crassifolia</i>	Ochrophyta; Phaeophyceae; Laminariales; Laminariaceae; Kjellmaniella			58.4	557	branches	201202108
scaffold-RAPY-2007960-Kjellmaniella crassifolia	<i>Kjellmaniella crassifolia</i>	Ochrophyta; Phaeophyceae; Laminariales; Laminariaceae; Kjellmaniella			62.9	895	branches	201202108
scaffold-VRGZ-2004699-Petalonia fascia	<i>Petalonia fascia</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Petalonia			56.2	404	leaves	201203013
scaffold-VRGZ-2023009-Petalonia fascia	<i>Petalonia fascia</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Petalonia			56.7	404	leaves	201203013
scaffold-VRGZ-2011481-Petalonia fascia	<i>Petalonia fascia</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Petalonia			57.8	424	leaves	201203013
scaffold-VRGZ-2002152-Petalonia fascia	<i>Petalonia fascia</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Petalonia			58.9	434	leaves	201203013
scaffold-VRGZ-2009530-Petalonia fascia	<i>Petalonia fascia</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Petalonia			56	435	leaves	201203013
scaffold-VRGZ-2016723-Petalonia fascia	<i>Petalonia fascia</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Petalonia			58.8	504	leaves	201203013
scaffold-VRGZ-2012524-Petalonia fascia	<i>Petalonia fascia</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Petalonia			60.2	515	leaves	201203013
scaffold-VRGZ-2002468-Petalonia fascia	<i>Petalonia fascia</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Petalonia			58.9	550	leaves	201203013
scaffold-VRGZ-2002469-Petalonia fascia	<i>Petalonia fascia</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Petalonia			58.3	557	leaves	201203013
scaffold-VRGZ-2012966-Petalonia fascia	<i>Petalonia fascia</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Petalonia			59.1	759	leaves	201203013
scaffold-SRSQ-2019594-SRSQ-Saccharina japonica	<i>Saccharina japonica</i>	Ochrophyta; Phaeophyceae; Laminariales; Laminariaceae; Saccharina			58	398	leaves	2011040101

to be continued

Continued from Table S1

Abbreviation	Species name	Taxonomy	Database number of the amino acid	Database number of the amino nucleotide	GC content of cDNA sequence/%	Number of the coded amino acid	Tissue description	Voucher data (indicate if pending)
scaffold-SRSQ-2017321-SRSQ-Saccharina japonica	<i>Saccharina japonica</i>	Ochrophyta; Phaeophyceae; Laminariales; Laminariaceae; Saccharina			60.4	438	leaves	2011040101
scaffold-SRSQ-2009816-SRSQ-Saccharina japonica	<i>Saccharina japonica</i>	Ochrophyta; Phaeophyceae; Laminariales; Laminariaceae; Saccharina			60.2	486	leaves	2011040101
scaffold-SRSQ-2017100-SRSQ-Saccharina japonica	<i>Saccharina japonica</i>	Ochrophyta; Phaeophyceae; Laminariales; Laminariaceae; Saccharina			58.4	498	leaves	2011040101
scaffold-SRSQ-2016424-SRSQ-Saccharina japonica	<i>Saccharina japonica</i>	Ochrophyta; Phaeophyceae; Laminariales; Laminariaceae; Saccharina			59.8	502	leaves	2011040101
scaffold-SRSQ-2009404-SRSQ-Saccharina japonica	<i>Saccharina japonica</i>	Ochrophyta; Phaeophyceae; Laminariales; Laminariaceae; Saccharina			62.8	899	leaves	2011040101
scaffold-VYER-2001762-Sargassum hemiphyllum var. chinense	<i>Sargassum hemiphyllum</i> var. <i>chinense</i>	Phaeophyceae; Fucales; Sargassaceae; Sargassum			57.4	289	branches and leaves	201120023
scaffold-VYER-2001763-Sargassum hemiphyllum var. chinense	<i>Sargassum hemiphyllum</i> var. <i>chinense</i>	Phaeophyceae; Fucales; Sargassaceae; Sargassum			56.3	440	branches and leaves	201120023
scaffold-VYER-2006531-Sargassum hemiphyllum var. chinense	<i>Sargassum hemiphyllum</i> var. <i>chinense</i>	Phaeophyceae; Fucales; Sargassaceae; Sargassum			53.5	531	branches and leaves	201120023
scaffold-FIKG-2003118-Sargassum henslowianum	<i>Sargassum henslowianum</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			54	437	branches and leaves	201120023
scaffold-FIKG-2003119-Sargassum henslowianum	<i>Sargassum henslowianum</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			54	437	branches and leaves	201120024
scaffold-FIKG-2012557-Sargassum henslowianum	<i>Sargassum henslowianum</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			56.7	440	branches and leaves	201120024
scaffold-FIKG-2009988-Sargassum henslowianum	<i>Sargassum henslowianum</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			51.5	443	branches and leaves	201120024
scaffold-FIKG-2015471-Sargassum henslowianum	<i>Sargassum henslowianum</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			53.8	497	branches and leaves	201120024
scaffold-FOMH-2018075-Sargassum in integrinum	<i>Sargassum integrinum</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			56.9	440	branches and leaves	201120026
scaffold-YRMA-2003280-Sargassum thunbergii	<i>Sargassum thunbergii</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			53.1	433	branches and leaves	201104041
scaffold-YRMA-2106544-Sargassum thunbergii	<i>Sargassum thunbergii</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			50.8	443	branches and leaves	201104041
scaffold-YRMA-2003281-Sargassum thunbergii	<i>Sargassum thunbergii</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			53.7	443	branches and leaves	201104041
scaffold-YRMA-2003885-Sargassum thunbergii	<i>Sargassum thunbergii</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			53.7	446	branches and leaves	201104041
scaffold-YRMA-2106508-Sargassum thunbergii	<i>Sargassum thunbergii</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			52	501	branches and leaves	201104041
scaffold-YRMA-2003481-Sargassum thunbergii	<i>Sargassum thunbergii</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			54.3	513	branches and leaves	201104041
scaffold-YRMA-2011471-Sargassum thunbergii	<i>Sargassum thunbergii</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			53.6	531	branches and leaves	201104041
scaffold-HFHK-2008735-Sargassum vachellianum	<i>Sargassum vachellianum</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			51.6	443	branches and leaves	201102023
scaffold-HFHK-2012631-Sargassum vachellianum	<i>Sargassum vachellianum</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			53.2	514	branches and leaves	201102023
scaffold-HFHK-2008339-Sargassum vachellianum	<i>Sargassum vachellianum</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			53.2	530	branches and leaves	201102023
scaffold-HFHK-2013676-Sargassum vachellianum	<i>Sargassum vachellianum</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			51.3	814	branches and leaves	201102023
scaffold-RWXW-2016212-Sargassum horneri	<i>Sargassum horneri</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			55.9	315	branches and leaves	2011050003
scaffold-RWXW-2007799-Sargassum horneri	<i>Sargassum horneri</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			51.9	439	branches and leaves	2011050003
scaffold-RWXW-2010628-Sargassum horneri	<i>Sargassum horneri</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			51.3	500	branches and leaves	2011050003
scaffold-RWXW-2003705-Sargassum horneri	<i>Sargassum horneri</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			51.6	818	branches and leaves	2011050003
scaffold-JGCD-2003061-Sargassum muticum	<i>Sargassum muticum</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			49.9	442	branches and leaves	201203003
scaffold-JGCD-2008685-Sargassum muticum	<i>Sargassum muticum</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			54.9	482	branches and leaves	201203003
scaffold-JGCD-2001666-Sargassum muticum	<i>Sargassum muticum</i>	Ochrophyta; Phaeophyceae; Fucales; Sargassaceae; Sargassum			53.1	531	branches and leaves	201203003
scaffold-JCXF-2000336-Scytosiphon lomentarinari	<i>Scytosiphon lomentarinari</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Scytosiphon			59.2	426	leaves	2012020100
scaffold-JCXF-2000335-Scytosiphon lomentarinari	<i>Scytosiphon lomentarinari</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Scytosiphon			58.4	435	leaves	2012020100
scaffold-JCXF-2014297-Scytosiphon lomentarinari	<i>Scytosiphon lomentarinari</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Scytosiphon			59.2	466	leaves	2012020100
scaffold-JCXF-2000337-Scytosiphon lomentarinari	<i>Scytosiphon lomentarinari</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Scytosiphon			57.6	470	leaves	2012020100
scaffold-JCXF-2001227-Scytosiphon lomentarinari	<i>Scytosiphon lomentarinari</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Scytosiphon			58	491	leaves	2012020100
scaffold-JCXF-2002095-Scytosiphon lomentarinari	<i>Scytosiphon lomentarinari</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Scytosiphon			59.2	505	leaves	2012020100
scaffold-JCXF-2002093-Scytosiphon lomentarinari	<i>Scytosiphon lomentarinari</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Scytosiphon			60.1	505	leaves	2012020100
scaffold-JCXF-201118-Scytosiphon lomentarinari	<i>Scytosiphon lomentarinari</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Scytosiphon			58.1	585	leaves	2012020100
scaffold-JCXF-2014707-Scytosiphon lomentarinari	<i>Scytosiphon lomentarinari</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Scytosiphon			61.2	759	leaves	2012020100
scaffold-JCXF-2010169-Scytosiphon lomentarinari	<i>Scytosiphon lomentarinari</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Scytosiphon			54.9	821	leaves	2012020100
scaffold-JCXF-2016819-Scytosiphon doyo	<i>Scytosiphon doyo</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Scytosiphon			55.7	435	leaves	201203012
scaffold-ULXR-2017937-Scytosiphon doyo	<i>Scytosiphon doyo</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Scytosiphon			60.2	460	leaves	201203012
scaffold-ULXR-2069029-Scytosiphon doyo	<i>Scytosiphon doyo</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Scytosiphon			59.1	461	leaves	201203012
scaffold-ULXR-2018189-Scytosiphon doyo	<i>Scytosiphon doyo</i>	Ochrophyta; Phaeophyceae; Ectocarpales; Scytosiphonaceae; Scytosiphon			61.7	760	leaves	201203012

to be continued



Continued from Table S1

Abbreviation	Collector	Sample type (including buffer for RNA)	Concentration/ ng $\mu$ l <sup>-1</sup>	Volume/ $\mu$ l	Total RNA Mass/ $\mu$ g	OD 260/280	OD 260/230	RIN score (if available)	28S:18S (if available)	Additional notes	Locality and date of collection collector
scaffold-QLMZ-2008026-Colpomenia_sinuosa	Liu Tao's lab	total RNA in RNase free water	1680.27	30	48	2.05	2.17	0	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-28.II.2012
scaffold-QLMZ-2008025-Colpomenia_sinuosa	Liu Tao's lab	total RNA in RNase free water	1680.27	30	48	2.05	2.17	0	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-28.II.2012
scaffold-QLMZ-2008027-Colpomenia_sinuosa	Liu Tao's lab	total RNA in RNase free water	1680.27	30	48	2.05	2.17	0	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-28.II.2012
scaffold-QLMZ-2013947-Colpomenia_sinuosa	Liu Tao's lab	total RNA in RNase free water	1680.27	30	48	2.05	2.17	0	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-28.II.2012
scaffold-QLMZ-2017500-Colpomenia_sinuosa	Liu Tao's lab	total RNA in RNase free water	1680.27	30	48	2.05	2.17	0	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-28.II.2012
scaffold-QLMZ-2003153-Colpomenia_sinuosa	Liu Tao's lab	total RNA in RNase free water	1680.27	30	48	2.05	2.17	0	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-28.II.2012
scaffold-FSQE-2008317-Desmarestia viridis	Liu Tao's lab	total RNA in RNase free water	2689.84	70	188	1.98	2.13	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-20.III.2012
scaffold-FSQE-2001026-Desmarestia viridis	Liu Tao's lab	total RNA in RNase free water	2689.84	70	188	1.98	2.13	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-20.III.2012
scaffold-FSQE-2001027-Desmarestia viridis	Liu Tao's lab	total RNA in RNase free water	2689.84	70	188	1.98	2.13	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-20.III.2012
scaffold-FSQE-2001028-Desmarestia viridis	Liu Tao's lab	total RNA in RNase free water	2689.84	70	188	1.98	2.13	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-20.III.2012
scaffold-FSQE-2001029-Desmarestia viridis	Liu Tao's lab	total RNA in RNase free water	2689.84	70	188	1.98	2.13	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-20.III.2012
scaffold-FSQE-2000750-Desmarestia viridis	Liu Tao's lab	total RNA in RNase free water	2689.84	70	188	1.98	2.13	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-20.III.2012
scaffold-FSQE-2000752-Desmarestia viridis	Liu Tao's lab	total RNA in RNase free water	2689.84	70	188	1.98	2.13	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-20.III.2012
scaffold-FSQE-2004718-Desmarestia viridis	Liu Tao's lab	total RNA in RNase free water	2689.84	70	188	1.98	2.13	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-20.III.2012
scaffold-FSQE-2002614-Desmarestia viridis	Liu Tao's lab	total RNA in RNase free water	2689.84	70	188	1.98	2.13	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-20.III.2012
scaffold-LIRF-2006001-Diclyopteris undulata	Liu Tao's lab	total RNA in RNase free water	2689.84	70	188	1.98	2.13	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-20.III.2012
scaffold-LIRF-2006000-Diclyopteris undulata	Liu Tao's lab	total RNA in RNase free water	1344.35	30	40	1.97	1.79	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-20.III.2012
scaffold-LIRF-2016070-Diclyopteris undulata	Liu Tao's lab	total RNA in RNase free water	1344.35	30	40	1.97	1.79	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-20.III.2012
scaffold-LIRF-2015641-Diclyopteris undulata	Liu Tao's lab	total RNA in RNase free water	1344.35	30	40	1.97	1.79	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-20.III.2012
scaffold-LIRF-2012622-Diclyopteris undulata	Liu Tao's lab	total RNA in RNase free water	1344.35	30	40	1.97	1.79	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-20.III.2012
scaffold-LIRF-2006206-Diclyopteris undulata	Liu Tao's lab	total RNA in RNase free water	1344.35	30	40	1.97	1.79	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-20.III.2012
scaffold-LDRY-2019956-Hizkia_fusifrome	Liu Tao's lab	total RNA in RNase free water	832.24	40	32	1.91	1.79	1.5	1.5	trizol-ethanol, -20°C	Zhangzhou, Fujiang (China)-2.IV.2011
scaffold-LDRY-2020267-Hizkia_fusifrome	Liu Tao's lab	total RNA in RNase free water	832.24	40	32	1.91	1.79	1.5	1.5	trizol-ethanol, -20°C	Zhangzhou, Fujiang (China)-2.IV.2011
scaffold-LDRY-2019221-Hizkia_fusifrome	Liu Tao's lab	total RNA in RNase free water	832.24	40	32	1.91	1.79	1.5	1.5	trizol-ethanol, -20°C	Zhangzhou, Fujiang (China)-2.IV.2011
scaffold-LDRY-2019176-Hizkia_fusifrome	Liu Tao's lab	total RNA in RNase free water	832.24	40	32	1.91	1.79	1.5	1.5	trizol-ethanol, -20°C	Zhangzhou, Fujiang (China)-2.IV.2011
scaffold-ATPP-2002334-Ishige_okamurai	Liu Tao's lab	total RNA in RNase free water	1205.34	30	36	1.92	1.49	1.3	1.3	trizol-ethanol, -20°C	Shantou, Guangdong (China)-29.II.2012
scaffold-ATPP-2014676-Ishige_okamurai	Liu Tao's lab	total RNA in RNase free water	1205.34	30	36	1.92	1.49	1.3	1.3	trizol-ethanol, -20°C	Shantou, Guangdong (China)-29.II.2012
scaffold-ATPP-2003397-Ishige_okamurai	Liu Tao's lab	total RNA in RNase free water	1205.34	30	36	1.92	1.49	1.3	1.3	trizol-ethanol, -20°C	Shantou, Guangdong (China)-29.II.2012
scaffold-RAPY-2016442-Kjellmaniella crassifolia	Liu Tao's lab	total RNA in RNase free water	1205.34	30	36	1.92	1.49	1.3	1.3	trizol-ethanol, -20°C	Shantou, Guangdong (China)-29.II.2012
scaffold-RAPY-2014805-Kjellmaniella crassifolia	Liu Tao's lab	total RNA in RNase free water	1205.34	30	36	1.92	1.49	1.3	1.3	trizol-ethanol, -20°C	Shantou, Guangdong (China)-29.II.2012
scaffold-RAPY-2016200-Kjellmaniella crassifolia	Liu Tao's lab	total RNA in RNase free water	1205.34	30	36	1.92	1.49	1.3	1.3	trizol-ethanol, -20°C	Shantou, Guangdong (China)-29.II.2012
scaffold-RAPY-2008005-Kjellmaniella crassifolia	Liu Tao's lab	total RNA in RNase free water	1205.34	30	36	1.92	1.49	1.3	1.3	trizol-ethanol, -20°C	Shantou, Guangdong (China)-29.II.2012
scaffold-RAPY-2002705-Kjellmaniella crassifolia	Liu Tao's lab	total RNA in RNase free water	1205.34	30	36	1.92	1.49	1.3	1.3	trizol-ethanol, -20°C	Shantou, Guangdong (China)-29.II.2012
scaffold-RAPY-2012031-Kjellmaniella crassifolia	Liu Tao's lab	total RNA in RNase free water	1205.34	30	36	1.92	1.49	1.3	1.3	trizol-ethanol, -20°C	Shantou, Guangdong (China)-29.II.2012
scaffold-RAPY-2016425-Kjellmaniella crassifolia	Liu Tao's lab	total RNA in RNase free water	1205.34	30	36	1.92	1.49	1.3	1.3	trizol-ethanol, -20°C	Shantou, Guangdong (China)-29.II.2012
scaffold-RAPY-2007960-Kjellmaniella crassifolia	Liu Tao's lab	total RNA in RNase free water	1205.34	30	36	1.92	1.49	1.3	1.3	trizol-ethanol, -20°C	Shantou, Guangdong (China)-29.II.2012
scaffold-VRGZ-2004699-Petalonia_fascia	Liu Tao's lab	total RNA in RNase free water	3057	25	60	1.88	2.01	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-9.III.2012
scaffold-VRGZ-2023009-Petalonia_fascia	Liu Tao's lab	total RNA in RNase free water	3057	25	60	1.88	2.01	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-9.III.2012
scaffold-VRGZ-2011481-Petalonia_fascia	Liu Tao's lab	total RNA in RNase free water	3057	25	60	1.88	2.01	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-9.III.2012
scaffold-VRGZ-2002152-Petalonia_fascia	Liu Tao's lab	total RNA in RNase free water	3057	25	60	1.88	2.01	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-9.III.2012
scaffold-VRGZ-2009530-Petalonia_fascia	Liu Tao's lab	total RNA in RNase free water	3057	25	60	1.88	2.01	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-9.III.2012
scaffold-VRGZ-2016723-Petalonia_fascia	Liu Tao's lab	total RNA in RNase free water	3057	25	60	1.88	2.01	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-9.III.2012
scaffold-VRGZ-2012524-Petalonia_fascia	Liu Tao's lab	total RNA in RNase free water	3057	25	60	1.88	2.01	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-9.III.2012
scaffold-VRGZ-2002468-Petalonia_fascia	Liu Tao's lab	total RNA in RNase free water	3057	25	60	1.88	2.01	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-9.III.2012
scaffold-VRGZ-2002469-Petalonia_fascia	Liu Tao's lab	total RNA in RNase free water	3057	25	60	1.88	2.01	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-9.III.2012
scaffold-VRGZ-2012966-Petalonia_fascia	Liu Tao's lab	total RNA in RNase free water	3057	25	60	1.88	2.01	1.5	1.5	trizol-ethanol, -20°C	Qingdao, Shandong (China)-9.III.2012
scaffold-SRSQ-2019594-SRSQ-Saccharina japonica	Liu Tao's lab	total RNA in RNase free water	752.6	30	22	2.02	2.05	1.5	1.5	trizol-ethanol, -20°C	Rongcheng, Shandong (China)-16.IV.2011
scaffold-SRSQ-2017321-SRSQ-Saccharina japonica	Liu Tao's lab	total RNA in RNase free water	752.6	30	22	2.02	2.05	1.5	1.5	trizol-ethanol, -20°C	Rongcheng, Shandong (China)-16.IV.2011
scaffold-SRSQ-2098616-SRSQ-Saccharina japonica	Liu Tao's lab	total RNA in RNase free water	752.6	30	22	2.02	2.05	1.5	1.5	trizol-ethanol, -20°C	Rongcheng, Shandong (China)-16.IV.2011

to be continued

Continued from Table S1

Abbreviation	Collector	Sample type (including buffer for RNA)	Concentration/ ng $\mu$ l <sup>-1</sup>	Volume/ $\mu$ l	Total RNA Mass/ $\mu$ g	OD 260/280	OD 260/230	RIN score (if available)	28S:18S (if available)	Additional notes	Locality and date of collection collector
scaffold-SRSQ-2017100-SRSQ-Saccharina japonica	Liu Tao's lab	total RNA in RNase free water	752.6	30	22	2.02	2.05	1.5		trizol-ethanol, -20°C	Rongcheng, Shandong (China)-16.IV.2011
scaffold-SRSQ-2016424-SRSQ-Saccharina japonica	Liu Tao's lab	total RNA in RNase free water	752.6	30	22	2.02	2.05	1.5		trizol-ethanol, -20°C	Rongcheng, Shandong (China)-16.IV.2011
scaffold-SRSQ-2009404-SRSQ-Saccharina japonica	Liu Tao's lab	total RNA in RNase free water	752.6	30	22	2.02	2.05	1.5		trizol-ethanol, -20°C	Rongcheng, Shandong (China)-16.IV.2011
scaffold-VYER-2001762-Sargassum hemiphyllum var. chinense	Liu Tao's lab	total RNA in RNase free water	758.6	40	30	1.88	1.57	1.2		trizol-ethanol, -20°C	Xuwen, Guangdong (China)-13.XII.2011
scaffold-VYER-2001763-Sargassum hemiphyllum var. chinense	Liu Tao's lab	total RNA in RNase free water	758.6	40	30	1.88	1.57	1.2		trizol-ethanol, -20°C	Xuwen, Guangdong (China)-13.XII.2011
scaffold-VYER-2006531-Sargassum hemiphyllum var. chinense	Liu Tao's lab	total RNA in RNase free water	758.6	40	30	1.88	1.57	1.2		trizol-ethanol, -20°C	Xuwen, Guangdong (China)-13.XII.2011
scaffold-FIKG-2003118-Sargassum henslowianum	Liu Tao's lab	total RNA in RNase free water	758.6	40	30	1.88	1.57	1.2		trizol-ethanol, -20°C	Xuwen, Guangdong (China)-13.XII.2011
scaffold-FIKG-2003119-Sargassum henslowianum	Liu Tao's lab	total RNA in RNase free water	766.1	35	25	1.85	1.89	1.3		trizol-ethanol, -20°C	Xuwen, Guangdong (China)-13.XII.2011
scaffold-FIKG-2012557-Sargassum henslowianum	Liu Tao's lab	total RNA in RNase free water	766.1	35	25	1.85	1.89	1.3		trizol-ethanol, -20°C	Xuwen, Guangdong (China)-13.XII.2011
scaffold-FIKG-2009989-Sargassum henslowianum	Liu Tao's lab	total RNA in RNase free water	766.1	35	25	1.85	1.89	1.3		trizol-ethanol, -20°C	Xuwen, Guangdong (China)-13.XII.2011
scaffold-FIKG-2015471-Sargassum henslowianum	Liu Tao's lab	total RNA in RNase free water	766.1	35	25	1.85	1.89	1.3		trizol-ethanol, -20°C	Xuwen, Guangdong (China)-13.XII.2011
scaffold-FOMH-2018075-Sargassum integrirum	Liu Tao's lab	total RNA in RNase free water	1114.7	45	45	1.92	1.86	1.2		trizol-ethanol, -20°C	Zhanjiang, Guangdong (China)-13.XII.2011
scaffold-YRMA-2003280-Sargassum thunbergii	Liu Tao's lab	total RNA in RNase free water	1193.6	25	29	1.79	1.76	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-16.IV.2011
scaffold-YRMA-2106544-Sargassum thunbergii	Liu Tao's lab	total RNA in RNase free water	1193.6	25	29	1.79	1.76	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-16.IV.2011
scaffold-YRMA-2003281-Sargassum thunbergii	Liu Tao's lab	total RNA in RNase free water	1193.6	25	29	1.79	1.76	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-16.IV.2011
scaffold-YRMA-2003282-Sargassum thunbergii	Liu Tao's lab	total RNA in RNase free water	1193.6	25	29	1.79	1.76	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-16.IV.2011
scaffold-YRMA-2003885-Sargassum thunbergii	Liu Tao's lab	total RNA in RNase free water	1193.6	25	29	1.79	1.76	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-16.IV.2011
scaffold-YRMA-2106508-Sargassum thunbergii	Liu Tao's lab	total RNA in RNase free water	1193.6	25	29	1.79	1.76	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-16.IV.2011
scaffold-YRMA-2011471-Sargassum thunbergii	Liu Tao's lab	total RNA in RNase free water	1193.6	25	29	1.79	1.76	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-16.IV.2011
scaffold-HFHK-2069735-Sargassum vachellianum	Liu Tao's lab	total RNA in RNase free water	907.91	30	27	1.73	1.55	0		trizol-ethanol, -20°C	Shantou, Guangdong (China)-22.II.2011
scaffold-HFHK-2012631-Sargassum vachellianum	Liu Tao's lab	total RNA in RNase free water	907.91	30	27	1.73	1.55	0		trizol-ethanol, -20°C	Shantou, Guangdong (China)-22.II.2011
scaffold-HFHK-2069339-Sargassum vachellianum	Liu Tao's lab	total RNA in RNase free water	907.91	30	27	1.73	1.55	0		trizol-ethanol, -20°C	Shantou, Guangdong (China)-22.II.2011
scaffold-HFHK-2013676-Sargassum vachellianum	Liu Tao's lab	total RNA in RNase free water	907.91	30	27	1.73	1.55	0		trizol-ethanol, -20°C	Shantou, Guangdong (China)-22.II.2011
scaffold-RWXW-2016212-Sargassum_homeri	Liu Tao's lab	total RNA in RNase free water	625.81	40	24	2.08	1.98	0		trizol-ethanol, -20°C	Wenzhou, Zhejiang (China)-7.V.2011
scaffold-RWXW-2007799-Sargassum_homeri	Liu Tao's lab	total RNA in RNase free water	625.81	40	24	2.08	1.98	0		trizol-ethanol, -20°C	Wenzhou, Zhejiang (China)-7.V.2011
scaffold-RWXW-2010628-Sargassum_homeri	Liu Tao's lab	total RNA in RNase free water	625.81	40	24	2.08	1.98	0		trizol-ethanol, -20°C	Wenzhou, Zhejiang (China)-7.V.2011
scaffold-RWXW-2003705-Sargassum_homeri	Liu Tao's lab	total RNA in RNase free water	625.81	40	24	2.08	1.98	0		trizol-ethanol, -20°C	Wenzhou, Zhejiang (China)-7.V.2011
scaffold-JGGD-2010803-Sargassum_muticum	Liu Tao's lab	total RNA in RNase free water	2475.48	30	70	1.77	1.66	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-7.III.2012
scaffold-JGGD-2003061-Sargassum_muticum	Liu Tao's lab	total RNA in RNase free water	2475.48	30	70	1.77	1.66	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-7.III.2012
scaffold-JGGD-2080685-Sargassum_muticum	Liu Tao's lab	total RNA in RNase free water	2475.48	30	70	1.77	1.66	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-7.III.2012
scaffold-JGGD-2001666-Sargassum_muticum	Liu Tao's lab	total RNA in RNase free water	2475.48	30	70	1.77	1.66	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-7.III.2012
scaffold-JCXF-2000336-Scytosion_homentaritari	Liu Tao's lab	total RNA in RNase free water	3021.58	35	100	1.89	1.89	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-28.II.2012
scaffold-JCXF-2000335-Scytosion_homentaritari	Liu Tao's lab	total RNA in RNase free water	3021.58	35	100	1.89	1.89	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-28.II.2012
scaffold-JCXF-2014297-Scytosion_homentaritari	Liu Tao's lab	total RNA in RNase free water	3021.58	35	100	1.89	1.89	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-28.II.2012
scaffold-JCXF-2000337-Scytosion_homentaritari	Liu Tao's lab	total RNA in RNase free water	3021.58	35	100	1.89	1.89	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-28.II.2012
scaffold-JCXF-2001227-Scytosion_homentaritari	Liu Tao's lab	total RNA in RNase free water	3021.58	35	100	1.89	1.89	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-28.II.2012
scaffold-JCXF-2002095-Scytosion_homentaritari	Liu Tao's lab	total RNA in RNase free water	3021.58	35	100	1.89	1.89	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-28.II.2012
scaffold-JCXF-2002093-Scytosion_homentaritari	Liu Tao's lab	total RNA in RNase free water	3021.58	35	100	1.89	1.89	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-28.II.2012
scaffold-JCXF-2001118-Scytosion_homentaritari	Liu Tao's lab	total RNA in RNase free water	3021.58	35	100	1.89	1.89	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-28.II.2012
scaffold-JCXF-2014707-Scytosion_homentaritari	Liu Tao's lab	total RNA in RNase free water	3021.58	35	100	1.89	1.89	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-28.II.2012
scaffold-JCXF-2010169-Scytosion_homentaritari	Liu Tao's lab	total RNA in RNase free water	3021.58	35	100	1.89	1.89	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-28.II.2012
scaffold-ULXR-20116819-Scytosion_dotoyo	Liu Tao's lab	total RNA in RNase free water	2194.4	30	60	2.01	2.15	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-9.III.2012
scaffold-ULXR-2017937-Scytosion_dotoyo	Liu Tao's lab	total RNA in RNase free water	2194.4	30	60	2.01	2.15	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-9.III.2012
scaffold-ULXR-2069029-Scytosion_dotoyo	Liu Tao's lab	total RNA in RNase free water	2194.4	30	60	2.01	2.15	0		trizol-ethanol, -20°C	Qingdao, Shandong (China)-9.III.2012
scaffold-ULXR-2018189-Scytosion_dotoyo	Liu Tao's lab	total RNA in RNase free water	2194.4	30	60	2.01	2.15	1.5		trizol-ethanol, -20°C	Qingdao, Shandong (China)-9.III.2012

Notes: n represents NCBI access number, e EMBL access number, ..... data that missing cds sequence in NCBI, and ..... data that cannot calculate GC content as cds missing. All the sequences with "scaffold" in its name are from ONEkp (www.ONEkp.com).