

# Genome-wide analysis of the *Saccharina japonica* sulfotransferase genes and their transcriptional profiles during whole developmental periods

Chang Lu (✉ [ytulc9510@sina.com](mailto:ytulc9510@sina.com))

Institute of Oceanology Chinese Academy of Sciences <https://orcid.org/0000-0002-7053-306X>

Zhanru Shao

Institute of Oceanology Chinese Academy of Sciences

Pengyan Zhang

Chinese Academy of Fishery Science Yellow Sea Fisheries Research Institute

Delin Duan

Institute of Oceanology Chinese Academy of Sciences

---

## Research article

**Keywords:** *Saccharina japonica*; sulfotransferase; fucoidan, transcriptome, genome

**Posted Date:** October 1st, 2019

**DOI:** <https://doi.org/10.21203/rs.2.15456/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background** As a unique sulfated polysaccharide, fucoidan is the major component of cell wall in brown seaweeds. Its biochemical properties are determined by the positions and quantity of sulfate groups. The sulfation process is catalyzed by sulfotransferases (STs), which transfer the sulfuryl groups to carbohydrate backbones and are crucial for fucoidan biosynthesis. Nevertheless, the structures and functions of STs in brown seaweeds are not completely understood. **Results** There are a total of 27 ST genes identified from our genome and transcriptome analysis of *Saccharina japonica* and they were located in the 12 scaffolds and seven contigs. The *S. japonica* ST genes have many introns and alternative splicing sites, and four tandem duplicated gene clusters were identified among this genes family. Generally, the ST genes could be classified into five groups (Group I~V) based on phylogenetic analysis, and the ST proteins, which were encoded by genes within the same group, contained similar conserved motifs. In group I sequences, we found two highly conserved regions (region I: TxPKSGTxW; region IV: RKGxxGDWKxxFT), and Lys 81 and Arg 287 are critical for catalysis and substrate binding. Members of the *S. japonica* ST gene family show various expression patterns in different tissues and developmental stages. Transcriptonal profiles indicate that more than half of ST genes transcriptional levels in kelp basal blades are higher than in distal blades, and decrease with the kelp development stages, while only ST 5, 9, 12, 18 and 25 are increased. The co-down-regulated pathways are related to oxidative phosphorylation, fatty acid biosynthesis and metabolic pathways, while the up-regulated pathways are involved with ribosome, nitrogen and sulfur metabolism. **Conclusion** Various characteristics of the STs allow the feasibilities of *S. japonica* to adapt to the costal environments, meet the needs of *S. japonica* growth and support their giant blades. This also reflects the complexity of fucoidan biosynthesis in different marine environments, tissues, and developmental stages.

## Background

*Saccharina japonica* is an important commercial brown seaweed in Asia. It is rich in crude fibers and carbohydrates and is widely used as a raw material for the extraction of alginate and fucoidan. Moreover, *S. japonica* contains many bioactive substances that are valuable for cosmetics, foods and health [1]. Among all bioactive metabolites, fucoidan is considered highly valuable in the field of medicine.

Fucoidan, a sulfated polysaccharide from non-mammalian sources, is one of the most widely studied sulfated polysaccharides. It mainly exists in echinoderm and cell walls of brown algae [2]. Fucoidan was first discovered by Kylin in brown algae *Laminaria digitata* in 1913 [3]. The main component of the sulfated fucoidan was L-fucose-4-sulfate; galactose, mannose, xylose, glucose, arabinose, glucuronic acid and metal ions exist in small amounts [4, 5]. It was believed that the content and structure of fucoidans in algae vary in different algae species, tissues, age, inhabitation and seasons [6, 7]. For example, the fucoidan content in brown algae growing in intertidal and subtidal zone was 20% and 1-2%, respectively [8]. The complexity in fucoidan content and composition makes the structure-activity studies difficult.

Fucoidan functions as anticoagulant and antithrombotic substances, and exert immunomodulation, anti-inflammation and anti-tumor functions [9-12]. It is also effective in relieving diabetic nephropathy and denine-induced chronic kidney disease [13, 14]. The structural parameters of fucoidan, such as the type of monosaccharide and fucose chain and the molecular weight of polysaccharide, all contribute to its bioactivity, especially the number and position of sulfate groups on the macromolecular skeleton [15-17]. For instance, the 2, 3-disulfated sugar residue is a common structure for anticoagulant activity [18, 19], whereas, the existence of 2-O-sulfation at the C-2 position reduces the anticoagulant activity of fucoidan [20]. Thus, sulfation plays an important role in the function of fucoidan. It has been reported that sulfotransferase (ST) can transfer the sulfuryl groups from the universal donor 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to carbohydrate backbones [21, 22]. It determines the position and quantity of sulfate groups in fucoidan.

The fucoidan biosynthesis pathway was not clear until the release of genome sequences of brown algae *Ectocarpus siliculosus* in 2010 [23]. Based on *E. siliculosus* genome sequencing and analogized with glycosaminoglycan (GAG) biosynthesis, Michel et al. (2010) deduced that fucoidan may first be polymerized into neutral polysaccharides by one or more fucosyltransferases, and then sulfated by specific sulfotransferases [21]. He proposed two routes of GDP-fucose production: 1) fructose-6-phosphate is catalyzed by mannose-6-phosphate isomerase (MPI), phosphomannomutase (PMM) and mannose-1-phosphate guanylyltransferase (MPG) to synthesize GDP-mannose, followed by the production of GDP-fucose, which is catalyzed by GDP-mannose 4, 6-dehydrogenase (GM46D) and the bifunctional enzyme GDP-L-fucoidase synthase (GFS); 2) alternatively, L-fucose is used as the substrate to synthesize GDP-fucose by fucose kinase (FK) and GDP-fucose pyrophosphorylase (GFPP). GDP-fucose is subsequently used to generate fucoidan by fucosyltransferase (FUT) and sulfotransferase (ST).

Previously, many genes involved in fucoidan biosynthesis were studied in *S. japonica*, *Cladosiphon okamuranus* and *Nemacystus decipiens* [24-26]. For example, Chi et al. (2017) worked on the *MPI*, *MPG* and *PMM* in *S. japonica* [27], and Nishitsuji et al. (2019) reported that *FK* and *GFPP* fused in *N. decipiens* genome [26]. Their researches enlighten our further genome-wide analysis on sulfotransferases (STs) in brown algae. Considering few investigation of this key enzyme in *S. japonica*, it therefore needs to explore the biosynthetic mechanism of sulfated fucoidan, especially sulfation modification of fucan catalyzed by ST.

As described above, the structure and content of fucoidan varied with the kelp growth. Therefore, the transcriptional levels of *ST* genes might be fluctuated at different development stages and tissues. Considering *ST* gene family contains large amount of putative members, e.g. 21 in *E. siliculosus* and 9 in *C. okamuranus* [23, 25], it is thus necessary to globally analyze the distinct features of these *STs* in brown algae. Transcriptome sequencing is an effective tool to decipher the transcriptional levels of the total *STs* and the entire metabolic network. This could reflect how the specific *STs* function with the kelp growth at the mRNA level.

In this study, we screened 103 genes (including three *MPIs*, two *PMMs*, three *GM46Ds*, 22 *FUT* and 73 *ST*) involved in fucoidan biosynthesis in *S. japonica*. Specifically, we characterized the *ST* genes by analyzing their sequence features, scaffold locations, phylogenetic relationships, tissue and time specific expression patterns and related enrichment pathways. This is the first study to investigate the characteristics of *ST* family members in *S. japonica*. Our result provides valuable knowledge of the biosynthesis of sulfated fucoidan in brown seaweeds, and have great potential for *in vitro* fucoidan biosynthesis.

## Results

### Identification of fucoidan biosynthetic genes

A total of 103 genes related to fucoidan synthesis were annotated based on the genome and transcriptome databases of *S. japonica* (Table S1). Generally, the ORFs of 81 genes were complete, and the rest of 22 genes were incomplete. The transcriptional levels of *MPI3* (GENE\_013986), *PMM1* (GENE\_007314), *GM46D1* (GENE\_026041) and *ST12* (GENE\_011842) were higher in each catalytic step, and were believed to be essential for fucoidan synthesis during the kelp growth and development.

### Identification and characterization of the *ST* genes

*S. japonica* genome has 73 genes automatically annotated as *ST* genes, among which 27 were further confirmed as *S. japonica* *ST* genes. The *S. japonica* genome is 2.5, 3.58 and 3.83 times larger than that of *E. siliculosus*, *N. decipiens*, and *C. okamuranus*, respectively; the number of annotated *ST* genes in *S. japonica* is approximately 3.17, 7.31 and 8.1 times those in *E. siliculosus*, *N. decipiens* and *C. okamuranus*, respectively (Table 1).

These 27 *ST* genes were named *ST1* to *ST27* (the average FPKM values from high to low). Name, gene ID, scaffold location, ORF length, exon number, amino acid number, molecular weight, and pI of the 27 genes and their corresponding proteins are summarized in Table 2. The *ST* proteins ranged from 117 (*ST17*) to 473 (*ST27*) amino acids in length and 13.48 kDa (*ST17*) to 51.37 kDa (*ST27*) in molecular weight. The predicted pI values of *ST* proteins ranged from 4.77 (*ST17*) to 9.4 (*ST5*).

Amino acid sequence analysis showed that most proteins were non-secretory, and only five proteins (*ST2*, *ST3*, *ST13*, *ST19*, and *ST25*) have transmembrane helices, and were predicted to have signal peptide or signal anchor. In addition, five proteins without the transmembrane domain (*ST1*, *ST4*, *ST6*, *ST7*, and *ST14*) contained signal peptides. Only *ST2* and *ST16*, were predicted to target the chloroplast, whereas *ST3*, *ST7*, *ST10*, *ST17*, *ST19* and *ST22*, were predicted to be located in the mitochondria (Table S2).

### Sequence analysis of the *ST* genes

The 27 *S. japonica* *ST* proteins could be classified into 5 groups (I - V) (Fig. 1A). A total of 20 conserved motifs were identified (Fig. 1B and Table S3). Group I, II and V contained *Sulfotransfer\_1* domain (PF00685), group IV contained *Sulfotransfer-2* domain (PF03567), and group III contained *Gal-O-sulfotr*

domain (PF06990), which was only present in algae [28]. The ST sequences in the same group contain similar conserved motifs (Fig. 1A, B).

Gene structures analysis showed that most *ST* genes (26, %) had more than three introns. Only one gene had less than three introns (ST17, 2 introns). The longest intron identified in the *ST* genes was nearly 15 kb (Fig. 1C).

We analyzed the types and numbers of all alternative splicing sites in *S. japonica* *ST* genes in different tissues and developmental stages. A total of 368 sites were identified in all samples. The most abundant alternative splicing site type was the ES type (127). Some types were centrally detected in several specific genes, for example, p5\_splice (*ST1*), p3\_splice (*ST3* and *ST22*), ALS (*ST11* and *ST22*) and AFS (*ST10* and *ST20*). In addition, *ST* genes expressed in the basal blade contained more alternative splicing sites than those expressed in the distal blade. Details of these sites are listed in Table S4.

The PCR products of the cDNA sequences of three selected *STs* appeared as a single band, except for GENE\_014314, as revealed by 1.5% agarose gel electrophoresis (Fig. S1). The results of these four selected *STs* in plasmid sequencing data of recombinant vector and RNA-Seq analysis were consistent. The coding sequences of the four *STs* are provided in Table S5.

### **Scaffold location and gene duplication of the *STs***

The 27 *ST* genes loci distributed randomly on 12 scaffolds and 7 contigs in *S. japonica* (Fig. 2). Only scaffold 3, 4, 6, 14 and 23, contained two or more *ST* loci. The *ST* family only contained tandem duplication which covered 33.3% of whole *ST* gene family. Duplicate *ST* gene pairs were found on scaffold 6 (*ST11* and *ST15*), 14 (*ST9* and *ST17*) and 23 (*ST25* and *ST19*). A group of three tandem duplicates, *ST1*, *ST4* and *ST7*, was identified on scaffold 4, and no collinearity among *ST* family members was observed.

### **Sequence alignment and phylogenetic analysis**

As shown in Fig. 3, the secondary structure of *S. japonica* *ST* protein contained abundant alpha-helix (on average 56.38%), followed by beta-bridge (on average 27.44%). Although there were some differences in amino acid sequences, the sequences in the same group displayed similar secondary structures. In group I, we found two highly conserved motifs (regions I (TxPKSGTxW) and IV (RKGxxGDWKxxFT)) and two conserved active sites (Lys<sup>59</sup> and Arg<sup>276</sup>) (Fig. 4).

A phylogenetic tree was constructed based on the *ST* amino acid sequences: 27 *S. japonica*, 21 *E. siliculosus*, six *C. okamuranus* and nine *P. tricornutum* (Fig. 5). Five *ST* clades in our phylogenetic tree were determined based on the classification result of 15 *E. siliculosus* *ST* sequences [21]. Fifty-four of the 63 *ST* proteins were grouped into four subfamilies and the *E. siliculosus* *STs* previously clustered into two different clades (clade A and clade B) fell into one group, which contained 15 *STs*. In addition, seven, 13 and 19 *STs* clustered into clades C, D and E, respectively, and the remaining nine *STs* were unclassified.

## Transcript profiles of *STs* in different tissues and developmental stages

A heatmap of the transcript levels of the 27 *ST* genes represented by FPKM values in different tissues and developmental stages was established by TBtools according to the RNA-Seq data (Fig. 6). In the tested samples, *ST1* had a relatively high expression level, *ST27* was barely detected. Most *STs* expressed more strongly in the basal than in the distal blade.

The trends of transcriptional levels of the 27 *ST* genes in *S. japonica* basal blade within all developmental stages are shown in Table 3. The *STs* exhibited several major expressed patterns: profile0 (*ST8*, *ST20*, *ST21* and *ST24*), profile2 (*ST6* and *ST26*), profile3 (*ST1*, *ST10* and *ST27*), profile6 (*ST15* and *ST23*), profile25 (*ST14*), profile28 (*ST25*) and profile29 (*ST5*, *ST12* and *ST18*). Profile 0 and 29 are the two representative profiles, the former contains genes with a down-regulated expression pattern from January to June while the latter had an up-regulated trend. The most enriched pathways in profile 0 included oxidative phosphorylation, photosynthesis - antenna proteins, photosynthesis, carbon fixation and metabolic pathway. Genes involved in ribosome, nitrogen metabolism, sulfur metabolism and inositol phosphate metabolism were enriched in profile 29 (Table S6).

We analyzed the transcriptional levels of the *STs* in distal blade, 1/3, 2/3 and basal blade of *S. japonica* collected in April (Table 4). According to our data, the *STs* exhibited five major expression patterns: profile0 (*ST17*, *ST18* and *ST25*), profile1 (*ST8*), profile4 (*ST12* and *ST19*), profile9 (*ST2*, *ST11*, *ST14*, *ST22* and *ST24*) and profile17 (*ST23*). The transcriptional levels of most *STs* were decreased, as observed for profiles 0, 1, 4 and 9, and genes related to basal metabolism and photosynthesis were enriched in these profiles, (Table S7).

## Discussion

The bioactivity of fucoidan in algae is affected by the number and position of sulfate groups, which vary in different species, locations, seasons, etc. In brown algae, sulfotransferase (*ST*) catalyzes the sulfation reaction in fucoidan synthesis. In this study, we selected 27 *ST* genes by screening the *S. japonica* genome and transcriptome databases and analyzed their structure, scaffold locations, phylogeny, duplication patterns and expression profiles in different tissues and developmental stages. This study provides valuable information of the *ST* gene family and facilitates future studies on their functional divergence in brown algae.

A large number of introns and diverse alternative splicing events existed in the *ST* sequences. Alternative splicing may have helped the kelp synthesize various *ST* enzymes with different functions, which has been proposed by previous studies [29]. Another observation is that similar motif organization only occurred in the same subgroup of *S. japonica* *ST* genes, and conserved motifs varied among different subgroups. This result indicates a close evolutionary relationship of *STs* within each subgroup.

*S. japonica* contains more *ST* genes than other sequenced brown algae. The number of annotated *ST* sequences in *S. japonica* genome is approximately 3- to 8-fold higher than that in *E. siliculosus*, *N.*

*decipiens*, and *C. okamuranus* (Table 1). Tandem duplication events may be significant in *S. japonica* *ST* gene family expansion. Considering that *S. japonica* exhibited morphological larger blade and obvious tissue differentiation, we presumed that the more abundant *ST* genes might enable the production of different sulfated fucoidans in various environmental conditions and growth or developmental stages.

Two highly conserved regions I (TYPKSGTxW) and region IV (RKGxxGDWKxxFT) for *ST* proteins from a broad spectrum of species were found [30]. Compared with early studies, the consensus sequences in group I of *S. japonica* *ST* proteins were different (region I: TxVKSGETTW, 78-86; region IV: LRKGGARRWRDVFT, 286-299). Both region I and region IV are involved in the binding of the cosubstrate PAPS, and Lys<sup>81</sup> and Arg<sup>287</sup> may be critical for catalysis and cosubstrate binding, respectively [30]. Despite the change of basic groups, the conservativeness of key sites indicated the similarity of the catalytic mechanism between *S. japonica* group I *ST* proteins and *ST* from plants and bacteria. In the other four groups, we did not find these highly conserved motifs. Their active sites need to be further explored with methods of site-directed mutation etc.

While distal blades contain higher fucoidan content [8], most *STs* are more highly expressed in the basal than in the distal blades. The basal blades of *S. japonica* have a strong ability to divide and produce more new cells than other tissues [31]. Considering fucoidan is a component of cell wall [2], the higher expression pattern of *ST* genes at basal blades may synthesize more fucoidan to meet the needs of new cells at the basal blades. From January to June, seawater temperature in the northern hemisphere gradually increases. High temperature leads to slower growth, development and metabolism of *S. japonica*. Some *STs* (eg. *ST8*, *20*, *21*, *24*), which enriched in the same profile as those related to oxidative phosphorylation and metabolic pathways, therefore showed a down-regulated trend. In addition, in mid-April, kelp enters the breeding season. Honya et al. (1999) found that *S. japonica* synthesized fucoidans with a higher sulfated content during its reproduction period, which played an important role in the formation and setting of spores [32, 33]. We speculated that those *STs* which showed an up-regulated trend (eg. *ST5*, *12*, *18*, *25*) can meet the reproductive needs of kelp. At the same time, protein degradation in kelp accelerated [34]. Correspondingly, nitrogen and sulfur metabolism were strengthened, and genes related to nitrogen and sulfur metabolism were also enriched in the profile of up-regulated *ST* gene enrichment. The different transcriptions profiles indicated *STs* various functions in synthesizing fucoidan with different sulfate positions and sulfated degree, and this will be beneficial to adaptation of *S. japonica* during the growth and development.

## Conclusions

A total of 27 *ST* genes were identified through analyzing the genome and transcriptome data of *S. japonica*. The large number of alternative splicing events, gene expansions, and the variable expression patterns of the *ST* genes had contributed to the complexity of fucoidan biosynthesis in *S. japonica*. These *ST* genes give rise to various fucoidan molecules that function in different tissues and developmental stages. To our knowledge, this is the first in-depth analysis of the *ST* gene family in algae, which lays a

foundation for future functional studies on these genes and informs the improvement of fucoidan production through genetic engineering.

## Methods

### Algal sample collection

*S. japonica* sporophytes were collected from January to June in 2014 from cultivated rafts in Gaolv Aquatic Company in Rongcheng, Shandong, China. Prof. Duan had undertaken the formal identification of the plant material used in your study. Three robust individuals were harvested for each time point, rinsed with filtered seawater several times and cut into three tissue sites. Each sectioned tissue sample was snap frozen in liquid nitrogen, and stored at -80 °C until total RNA isolation. Detailed information of the kelp samples is summarized in Table 5.

### RNA extraction and cDNA synthesis

Total RNA was isolated from the samples using a FastPure Plant Total RNA Isolation Kit (Vazyme, China). The extracted RNA was quantified using a Nanodrop 2000 Spectrophotometer (Thermo Scientific, USA). RNA quality was assessed by 1% agarose gel electrophoresis. First-strand cDNA was synthesized using HiScript II Q Select RT SuperMix (Vazyme, China) and stored at -20 °C for subsequent gene cloning and quantitative real-time PCR (qRT-PCR) analysis. All manipulations were operated following the manufacturers' instructions.

### Identification of sulfotransferase family members in *S. japonica* genome

Based on our previously sequenced *S. japonica* genome (NCBI: MEHQ00000000) and recently released *S. japonica* transcriptome data (NCBI: PRJNA512328), we identified 73 genes automatically annotated as *sulfotransferase* (*ST*) genes that catalyze the last step of fucoidan biosynthesis. These 73 genes were submitted to SMART (<http://smart.embl-heidelberg.de/>) [35, 36] and Pfam (<http://pfam.xfam.org/search>) [37] to confirm the presence of the conserved sulfotransferase domain. The predicted genes without the sulfotransferase domain and monthly average FPKM < 0.5 were excluded.

### Sequence analysis and chromosomal localization

The open reading frames (ORFs) of high-confidence *ST* genes were predicted using ORF finder (<https://www.ncbi.nlm.nih.gov/orffinder/>). The ExPASy ProtParam tool (<https://web.expasy.org/protparam/>) was used to analyze the physical and chemical properties of the deduced *ST* proteins, including the theoretical isoelectric point (pI), molecular weight (MW), and amino acid composition. The SignalP-5.0 Server (<http://www.cbs.dtu.dk/services/SignalP/>) was used to predict signal peptides [38], and TMHMM (v2.0; <http://www.cbs.dtu.dk/services/TMHMM/>) was employed for predicting the transmembrane helices in the proteins. Possible localization to the chloroplast, mitochondrion and cytoplasm was predicted by HECTAR ([http://webtools.sb-roscoff.fr/root?tool\\_id=abims\\_hectar](http://webtools.sb-roscoff.fr/root?tool_id=abims_hectar)) [39] and Target-P (<http://www.cbs.dtu.dk/services/TargetP/>) [40].

Structure of the *ST* genes was analyzed by Gene Structure Display Server (GSDS v2.0; <http://gsds.cbi.pku.edu.cn/>) [41]. The online software MEME (server v5.0.5, <http://meme-suite.org/>) was used to identify conserved motifs in the *ST* proteins with the following parameters: any number of repetitions, maximum of 20 motifs and an optimum motif width of 6 – 200 amino acid residues [42]. The chromosomal positions of the *ST* genes were acquired by aligning the full-length *ST* nucleic acid sequences to the *S. japonica* genome. We used MCScanX to search for duplicate genes in the *ST* family [43]. TBtools was used to display the chromosomal positions of *STs* and their relative physical distances [44].

### Identification of alternative splicing events

TopHat 2.1.1 was used to analyze alternative splicing events in the 27 *ST* genes [45]. Alternative splicing sites supported by less than five reads were filtered out, and the remaining were mapped to known alternative splicing sites, allowing for 1 bp error. The known alternative splicing sites were identified and the unmapped new alternative splicing sites were classified again. Refer to Fig. S2 for details.

### PCR amplification and sequencing of the *ST* genes

We randomly selected four genes (GENE\_011842, GENE\_026617, GENE\_013439 and GENE\_014314) for PCR amplification. Primers used to amplify the full-length cDNA sequences of these four genes are listed in Table 6. PCR amplification was performed using the synthesized cDNA as the template. The 20  $\mu$ L reaction mixture contained 10  $\mu$ L 2  $\times$  Phanta Max Master Mix (Vazyme, China), 3  $\mu$ L template, 1  $\mu$ L of each of the forward and reverse primer (10  $\mu$ M) and 5  $\mu$ L ddH<sub>2</sub>O. The reaction mixtures were briefly centrifuged and placed in a thermal cycler (Takara, Japan). The conditions used for PCR were as follows: 95 °C for 5 min, followed by five cycles of 95 °C for 15 s, 65 °C for 15 s, 72 °C for 90 s, 30 cycles of 95 °C for 15 s, 60 °C for 15 s, 72 °C for 90 s, and a final extension at 72 °C for 10 min. The PCR products were purified using the gel-cutting recovery kit (Insight, China) and inserted into TOPO cloning vector using a 5 min TA/Blunt-Zero Cloning Kit (Vazyme, China). The 5  $\mu$ L ligation mixture contained 1  $\mu$ L 5 $\times$ TA/Blunt-Zero Cloning Mix and 4  $\mu$ L purified PCR products (40 ng/ $\mu$ L). The mixture was briefly centrifuged and incubated at 37 °C for 10 min.

For transformation, 5  $\mu$ L of each recombinant plasmids was mixed with 50  $\mu$ L Trans1-T1 Phage Resistant Chemically Competent Cell (TRANS, China). The mixture was left on ice for 30 min, heat-shocked at 42°C for 30 s, and left on ice for another 2 min. The transformed cells were shaken in LB medium at 37°C for 1 h, spread on LB plates containing 100  $\mu$ g/mL ampicillin and incubated overnight at 37 °C. Colony PCR was carried out using M13 primers and plasmids were sanger sequenced (Sangon, Shanghai, China).

### Sequence alignment and phylogenetic analysis

All the 27 *S. japonica* *ST* proteins were aligned by Clustal W (<https://www.genome.jp/tools-bin/clustalw>) with the default parameters [46]. To analyze the evolutionary relationships among the 27 *STs* in *S.*

*japonica*, a neighbor-joining (NJ) phylogenetic tree was constructed based on their full-length amino acid sequences with MEGA 7.0.26 using the p-distance method with 1000 bootstrap replications, Gamma 3, partial deletion, and 50% site coverage as the cutoff value [47].

The amino acid sequences of STs derived from *E. siliculosus*, *C. okamuranus*, *Phaeodactylum tricornutum* and *S. japonica* were subjected to phylogenetic analysis (Table S8). The maximum likelihood (ML) phylogenetic tree was constructed by MEGA 7.0.26 using the full-length amino acid sequences of 63 ST proteins with 1000 bootstrap replications, the WAG+F+G model, Gamma 3, partial deletion, and 50% site coverage as the cutoff value [47].

### **Transcript profiling of the *ST* genes in different tissues and developmental stages**

Differentially expressed genes (DEGs) across samples were identified by the edgeR package (<http://www.r-project.org/>) [48]. Genes with a fold change  $\geq 2$  and a false discovery rate (FDR)  $< 0.05$  were considered as significant DEGs.

The transcriptional profiles of the *S. japonica* *ST* genes in different tissues and developmental stages were determined. Transcript data of the *ST* genes in each sample were obtained, normalized, and clustered by the Short Time-series Expression Miner software (STEM) [49]. Significant clustered profiles had  $p$ -value  $\leq 0.05$ . The DEGs in all profile were subjected to Gene Ontology (GO) and KEGG pathway enrichment analysis, and enriched GO terms or pathways with  $Q$  value  $\leq 0.05$  were considered to be significant. The heatmap of *ST* gene expression was drawn by TBtools [44].

### **Verification of target genes by quantitative real-time PCR (qRT-PCR)**

qRT-PCR was used to verify the transcript profiles of the four target genes (GENE\_011842, GENE\_021484, GENE\_009777 and XLOC\_011209) involved in fucoidan biosynthesis. Gene-specific primers used for qRT-PCR were designed using Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) (Table 6)

qRT-PCR was performed on a Takara Thermal Cycle Dice<sup>TM</sup> Real Time System (Takara, Japan). A 10  $\mu$ L qRT-PCR reaction contained 5  $\mu$ L 2 $\times$  ChamQ SYBR Color qPCR Master Mix (Vazyme, China), 0.5  $\mu$ L template, 0.5  $\mu$ L of each of the forward and reverse primers (10  $\mu$ M), and 3.5  $\mu$ L ddH<sub>2</sub>O. Conditions used for qRT-PCR were as follows: 95 °C for 30 s, followed by 40 cycles of 95 °C for 10 s and 60 °C for 30 s; and one cycle of 95 °C for 15 s, 60 °C for 60 s and 72 °C for 15 s. Three biological repeats and two technical replicates were performed. The relative transcriptional levels of the genes were calculated by the  $2^{-\Delta\Delta C_t}$  method [50], and  *$\beta$ -actin* was used as the internal reference [51].

## **Abbreviations**

ST: sulfotransferase; PAPS: 3'-phosphoadenosine 5'-phosphosulfate; MPI: mannose-6-phosphate isomerase; PMM: phosphomannomutase; MPG: mannose-1-phosphate guanylyltransferase; GM46D: GDP-mannose 4, 6-dehydrogenase; GFS: GDP-L-fucoidase synthase; FK: fucose kinase; GFPP: GDP-

fucose pyrophosphorylase; FUT: fucosyltransferase; RT-qPCR: Real-time quantitative polymerase chain reaction; NCBI: National Center for Biotechnology Information; FPKM: Fragments per kilobase of transcript per million mapped reads; ORFs: open reading frames; ML: maximum likelihood; NJ: neighbor-joining; AA: amino acid; MW: molecular weight; UTR: untranslated Region; DEGs: Differentially expressed genes.

## Declarations

### Ethics approval and consent to participate

*S. japonica* sporophytes “Zhongke 2” were collected from cultivated rafts in Gaolv Aquatic Company in Rongcheng, Shandong, China. Prof. Delin Duan had undertaken the formal identification of the plant material used in your study. The remaining samples are kept in our laboratory -80 refrigerator. The authors declare the compliance with institutional, national and international guidelines. The experiment was environmentally friendly and had no negative effects on the local ecosystem

### Consent for publication

Not applicable

### Availability of data and material

All of the datasets supporting the results of this article are included within the article and its Additional files.

### Competing interests

The authors declare that they have no competing interests.

### Funding

This work was supported by the National Natural Science Foundation of China (no. 41806175), the Shandong Key Research and Development Program (2018GHY115023), Open Foundation of the State Key Laboratory of Bioactive Seaweed Substances (SKL-BASS1702) and Joint Research Project between China and Japan (No. 2017YFE0130900). These funders provided financial support for the purchase of instruments and reagents for research.

### Authors' contributions

CL, ZS and DD conceived and designed research. CL and ZS conducted experiments. PZ contributed to the cultivation and the collection of kelp samples. CL performed the bioinformatics analysis with the supports of Gene Denovo Biotechnology Co. CL interpreted the data and wrote the manuscript. ZS and DD gave critical revision of the manuscript. All authors have read and approved the manuscript.

## Acknowledgements

We would like to thank Yaya Li and Tenglong Jiao, two technicians in Gene Denovo Biotechnology Co. for their help with bioinformatics analysis.

## Authors' information

<sup>1</sup>Key Laboratory of Experimental Marine Biology, Center for Ocean Mega-Science, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, 266071, P. R. China

<sup>2</sup>Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, P. R. China

<sup>3</sup>Division of Mariculture Ecology and Technology, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, 266071, China

<sup>4</sup>University of Chinese Academy of Sciences, Beijing 100049, P. R. China

<sup>5</sup>State Key Laboratory of Bioactive Seaweed Substances, Qingdao Bright Moon Seaweed Group Co Ltd, Qingdao 266400, P. R. China

## References

1. Wijesinghea WAJP, Jeona YJ. Biological activities and potential industrial applications of fucose rich sulfated polysaccharides and fucoidans isolated from brown seaweeds: a review. *Carbohydr Polym.* 2012; 88:13-20.
2. Berteau O, Mulloy B. Sulfated fucans, fresh perspectives: structures, functions, and biological properties of sulfated fucans and an overview of enzymes active toward this class of polysaccharide. *Glycobiology.* 2003; 13: 29R-40R.
3. Kylin H. Zur biochemie der meersalgen. *Hoppe-Seyler's Zeitschrift fur Physiologische Chemie.* 1913; 83:171–197.
4. Pomin VH, Pereira MS, Valente A, Tollefsen DM, Pavão MSG, Mourão PAS. Selective cleavage and anticoagulant activity of a sulfated fucan: stereospecific removal of a 2-sulfate ester from the polysaccharide by mild acid hydrolysis, preparation of oligosaccharides, and heparin cofactor II-dependent anticoagulant activity. *Glycobiology.* 2005; 15:369–381.
5. Fernando IPS, Sanjeewa KKA, Samarakoon KW, Lee WW, Kim HS, Kang N, Ranasinghe P, Lee HS, Jeon YJ. A fucoidan fraction purified from *Chnoospora minima*; a potential inhibitor of LPS-induced inflammatory responses. *Int J Biol Macromol.* 2005; 104:1185-1193.
6. Skriptsova AV, Shevchenko NM, Zvyagintseva TN, Imbs TI. Monthly changes in the content and monosaccharide composition of fucoidan from *Undaria pinnatifida* (Laminariales, Phaeophyta). *J Appl Phycol.* 2010; 22:79–86.

7. Li B, Lu F, Wei X, Zhao R. Fucoïdan: structure and bioactivity. *Molecules*. 2008; 13:1671-1695.
8. Ji MH. *Seaweed Chemistry*. 2nd ed. Beijing: Science Press; 1997. (In Chinese)
9. Cumashi A, Ushakova NA, Preobrazhenskaya ME, D'Incecco A, Piccoli A, Totani L, Tinari N, Morozevich GE, Berman AE, Bilan MI, Usov AI, Ustyuzhanina NE, Grachev AA, Sanderson CJ, Kelly M, Rabinovich GA, Iacobelli S, Nifantiev NE . A comparative study of the anti-inflammatory, anticoagulant, antiangiogenic, and antiadhesive activities of nine different fucoïdians from brown seaweeds. *Glycobiology*. 2007; 17:541–542.
10. Damonte EB, Matulewicz MC, Cerezo AS. Sulfated seaweed polysaccharides as antiviral agents. *Curr Med Chem*. 2004; 11:2399–419.
11. Azevedo TCG, Bezerra MEB, Santos MGL, Souza LA, Marques CT, Benevides NMB, Leite EL. Heparinoids algal and their anticoagulant, hemorrhagic activities and platelet aggregation. *Biomed Pharmacother*. 2009; 63:477–483.
12. Thomes P, Rajendran M, Pasanban B, Rengasamy R. Cardioprotective activity of *Cladosiphon okamuranus* against isoproterenol induced myocardial infarction in rats. *Phytomedicine*. 2010; 18:52-57.
13. Wang J, Wang F, Yun H, Zhang H, Zhang QB. Effect and mechanism of fucoïdan derivatives from *Laminaria japonica* in experimental adenine-induced chronic kidney disease. *J Ethnopharmacol*. 2012; 139:807-813.
14. Xu YJ, Zhang QB, Luo DL, Wang J, Duan DL. Low molecular weight fucoïdan modulates P-selectin and alleviates diabetic nephropathy. *Int J Biol Macromol*. 2016; 91:233-240.
15. Morya VK, Kim J, Kim E. Algal fucoïdan: structure and size-dependent bioactivities and their perspectives. *Appl Microbiol Biotechnol*. 2012; 93:71-82.
16. Nishino T, Nagumo T. Anticoagulant and antithrombin activities of over sulfated fucans. *Carbohyd Res*. 1992; 229:355-362.
17. Chevolot L, Foucault A, Chauber F, Kervarec N, Siquin C, Fisher AM, Boisson-Vidal C. Further data on the structure of brown seaweed fucans: Relationships with anticoagulant activity. *Carbohyd Res*. 1999; 319:154–165.
18. Chevolot L, Mulloy B, Ratiskol J, Foucault A, Collic-Jouault Sylvia. A disaccharide repeat unit is the major structure in fucoïdians from two species of brown algae. *Carbohyd Res*. 2001; 330:529-535.
19. Yoon SJ, Pyun YR, Hwang JK, Mourão PAS. A sulfated fucan from the brown alga *Laminaria cichorioides* has mainly heparin cofactor II-dependent anticoagulant activity. *Carbohyd Res*. 2007; 342:2326–2330.
20. Pereira MS, Melo FR, Mourão PAS. Is there a correlation between structure and anticoagulant action of sulfated galactans and sulfated fucans? *Glycobiology*. 2002; 12:573-580.
21. Michel G, Tonon T, Scornet D, Cock JM, Kloareg B. The cell wall polysaccharide metabolism of the brown alga *Ectocarpus siliculosus*. Insights into the evolution of extracellular matrix polysaccharides in Eukaryotes. *New Phytol*. 2010; 188:82-97.

22. Coughtrie MWH. Function and organization of the human cytosolic sulfotransferase (SULT) family. *Chem-Biol Interact.* 2016; 259:2-7.
23. Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, Anthouard V, Artiguenave F, Aury JM, Badger JH, Beszteri B, Billiau K, Bonnet E, Bothwell JH, Bowler C, Boyen C, Brownlee C, Carrano CJ, Charrier B, Cho GY, Coelho SM, Collén J, Corre E, Silva CD, Delage L, Delaroque N, Dittami SM, Doulebeau S, Elias M, Farnham G, Gachon CMM, Gschloessl B, Heesch S, Jabbari K, Jubin C, Kawai H, Kimura K, Kloareg B, Küpper FC, Lang D, Bail AL, Leblanc C, Lerouge P, Lohr M, Lopez PJ, Martens C, Maumus F, Michel G, Miranda-Saavedra D, Morales J, Moreau H, Motomura T, Nagasato C, Napoli CA, Nelson DR, Nyvall-Collén P, Peters AF, Pommier C, Potin P, Poulain J, Quesneville H, Read B, Rensing SA, Ritter A, Rousvoal S, Samanta M, Samson G, Schroeder DC, Ségurens B, Strittmatter M, Tonon T, Tregear JW, Valentin K, Dassow PV, Yamagishi T, Peer YV, Wincker P. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature.* 2010; 465:617-621.
24. Ye NH, Zhang XW, Miao M, Fan X, Zheng Y, Xu D, Wang JF, Zhou L, Wang DS, Gao Y, Wang YT, Shi WY, Ji PF, Li DM, Guan Z, Shao CW, Zhuang ZM, Gao ZQ, Qi J, Zhao FQ. *Saccharina* genomes provide novel insight into kelp biology. *Nature Commun.* 2015; 6:6986.
25. Nishitsuji K, Arimoto A, Iwai K, Sudo Y, Hisata K, Fujie M, Arakaki N, Kushiro T, Konishi T, Shinzato C, Satoh N, Shoguchi E. A draft genome of the brown alga, *Cladosiphon okamuranus*, S-strain: a platform for future studies of 'mozuku' biology. *DNA Res.* 2016; 23:561-570.
26. Nishitsuji K, Arimoto A, Higa Y, Mekaru M, Kawamitsu M, Satoh N, Shoguchi E. Draft genome of the brown alga, *Nemacystus decipiens*, Onna-1 strain: duplication of genes involved in the sulfated fucan biosynthesis pathway. *Scientific Rep.* 2019; 9:4607.
27. Chi S, Liu T, Wang XM, Wang R, Wang SS, Wang GL, Shan GL, Liu C. Functional genomics analysis reveals the biosynthesis pathways of important cellular components (alginate and fucoindan) of *Saccharina*. *Curr Genet.* 2018; 64:259-273.
28. Hirschmann F, Krause F, Papenbrock J. The multi-protein family of sulfotransferases in plants: composition, occurrence, substrate specificity, and functions. *front Plant Sci.* 2014; 5:1-13.
29. Lareau LF, Green RE, Bhatnagar RS, Brenner SE. The evolving roles of alternative splicing. *Curr Opin Struct Biol.* 2004; 14:273–282.
30. Marsolais F, Varin L. Identification of amino acid residues critical for catalysis and cosubstrate binding in the flavonol 3-sulfotransferase. *J Biol Chem.* 1995; 270: 30458-30463.
31. Duan DL, Miu GR, Wang XL. *Mariculture Biology of Laminaria japonica*. 1<sup>st</sup> Beijing: Science Press; 2015. (in Chinese)
32. Honya M, Mori H, Anzai M, Araki Y, Nishizawa K. Monthly changes in the content of fucans, their constituent sugars and sulphate in cultured *Laminaria japonica*. *Hydrobiologia.* 1999; 398:411–416.
33. Qu YY, Cao ZM, Wang WW, Wang N, Li XJ, Pan JH. Monthly variations of fucoindan content and its composition in the farmed brown alga *Saccharina sculpera* (Laminariales, Phaeophyceae). *J Appl Phycol.* 2019; 31:2623-2628.

34. Wu HY, Wang XY, Zhu AC. Analysis of component changes of *Saccharina japonica* at mushroom-adult stage. *Marine Science*. 2015; 39:35-38. (in Chinese)
35. Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res*. 2015; 43:D257-D260.
36. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res*. 2018; 46:D493-D496.
37. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. The Pfam protein families database in 2019. *Nucleic Acids Res*. 2019; 47:D427-D432.
38. Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, Heijne GV, Nielsen H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnol*. 2019; 37:420-423.
39. Gschloessl B, Guermeur Y, Cock JM. HECTAR: a method to predict subcellular targeting in heterokonts. *BMC Bioinformatics*. 2008; 9:393.
40. Emanuelsson O, Brunak S, Heijne GV, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc*. 2007; 2:953-971.
41. Hu B, Jin JP, Guo AY, Zhang H, Luo JC, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*. 2015; 31:1296-1297.
42. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren JY, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*. 2009; 37:W202-W208.
43. Wang YP, Tang HB, DeBarry JD, Tan X, Li JP, Wang XY, Lee TH, Jin HZ, Marler B, Guo H, Kissinger JC, Paterson AH. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012; 40:e49.
44. Chen CJ, Chen H, He YH, Xia R. TBtools, a Toolkit for Biologists integrating various biological data handling tools with a user-friendly interface. *BioRxiv*. 2018; 14:48.
45. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009; 25:1105-1111.
46. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res*. 2003; 31:3497-3500.
47. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*. 2016; 33:1870-1874.
48. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010; 26:139-140.
49. Ernst J, Bar-Joseph Z. STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics*. 2006; 7:191.
50. Schmittgen TD, Zakrajsek BA, Mill AG, Gorn V, Singer MJ, Reed MW. Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time

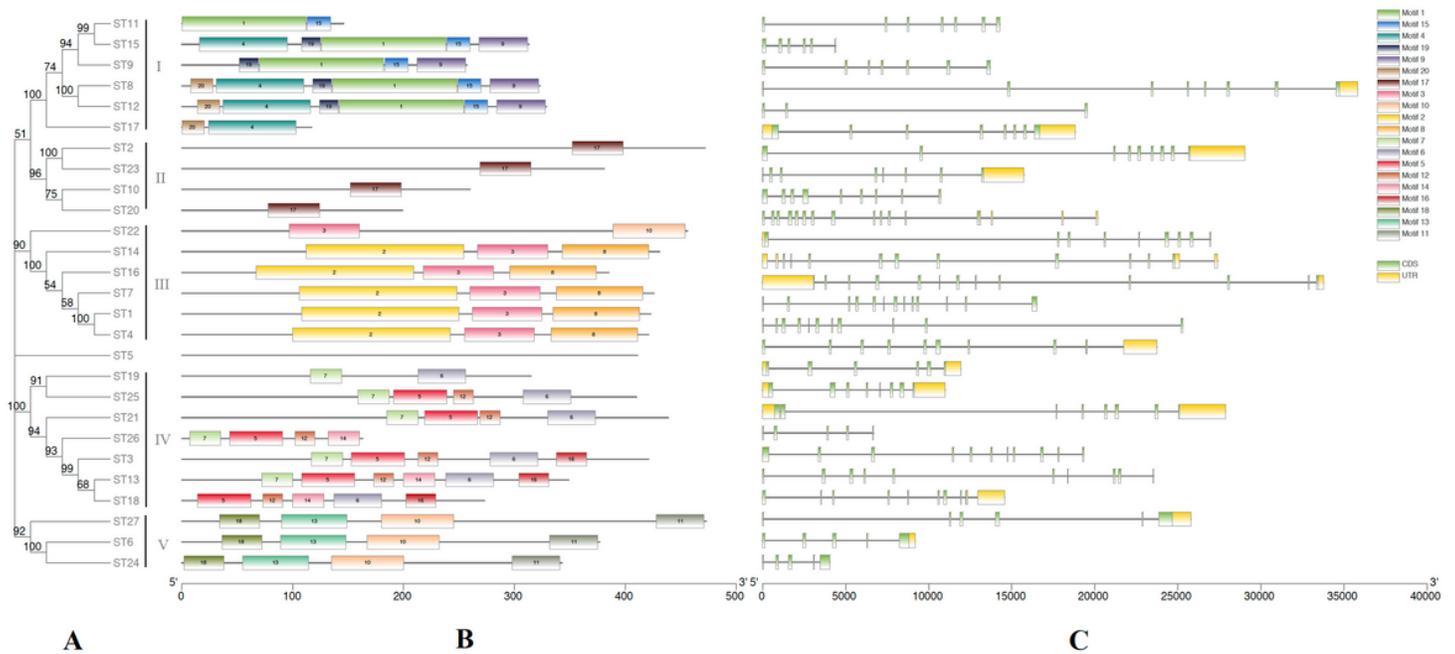
methods. Anal Biochem. 2000; 285: 194-204.

51. Shao ZR, Zhang PY, Li QY, Wang XL, Duan DL. Characterization of mannitol-2-dehydrogenase in *Saccharina japonica*: evidence for a new polyol-specific long-chain dehydrogenases/reductase. PLoS ONE. 2014; 9: e97935.

## Tables

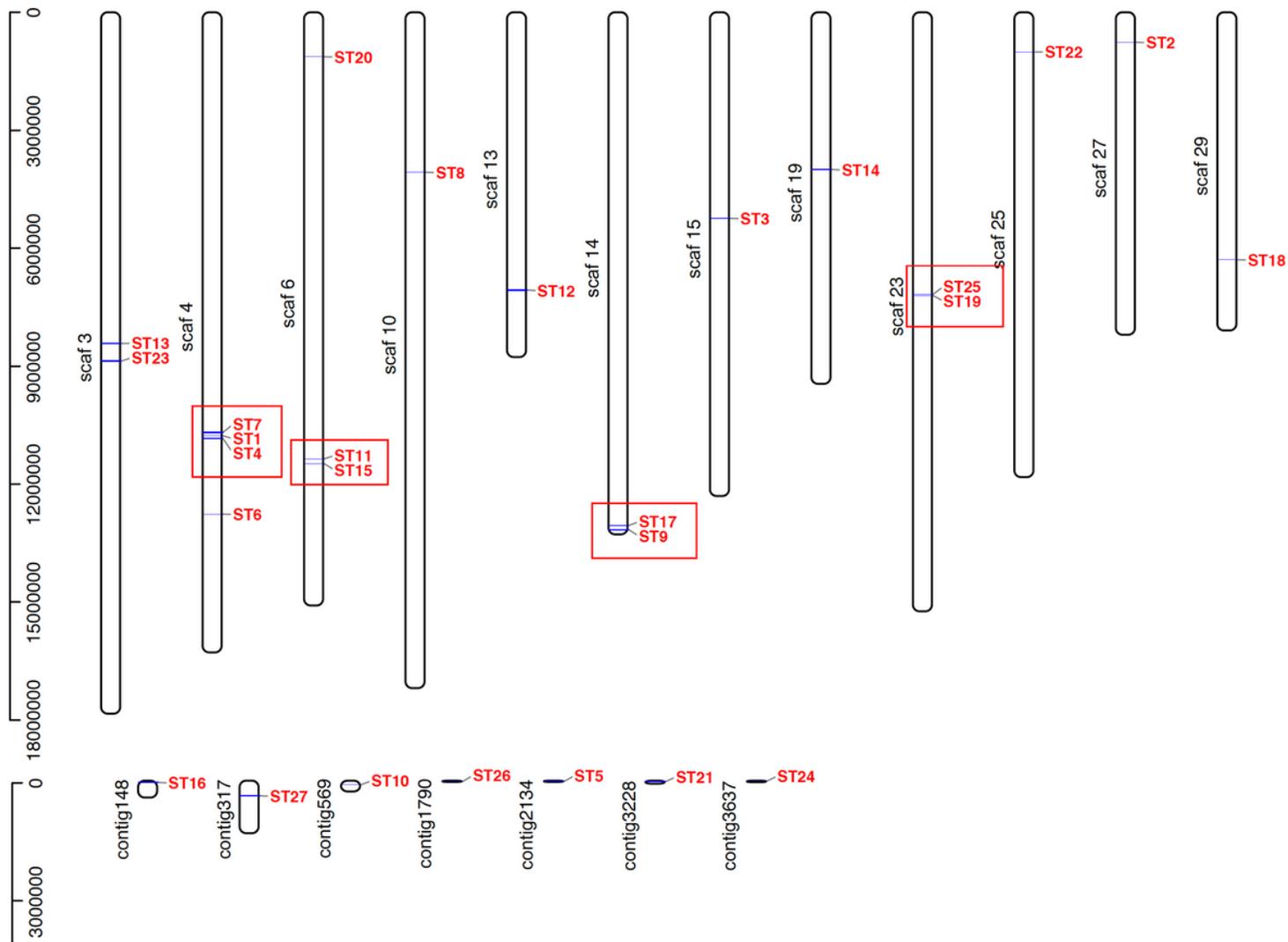
Due to technical limitations, the tables have been placed in the Supplementary Files section.

## Figures



**Figure 1**

Phylogenetic relationship, conserved motifs and gene structure of the ST genes and their corresponding proteins in *S. japonica*. A. Phylogenetic tree of the 27 *S. japonica* ST proteins. The neighbor-joining phylogenetic tree was constructed using MEGA 7.0.26, with 1000 bootstrap replications, P-distance method, Gamma 3, partial deletion and 50% site coverage as the cutoff. B. Conserved motifs identified in the 27 ST proteins. Twenty putative motifs were indicated by boxes of different color. Details on the motifs are listed in Table S3. C. Structures of the ST genes. Exons, introns and UTRs are indicated by green boxes, black lines and yellow boxes, respectively.



**Figure 2**

Scaffold locations of the ST genes and identification of duplicate genes. Tandem duplicated genes are indicated by red rectangles.

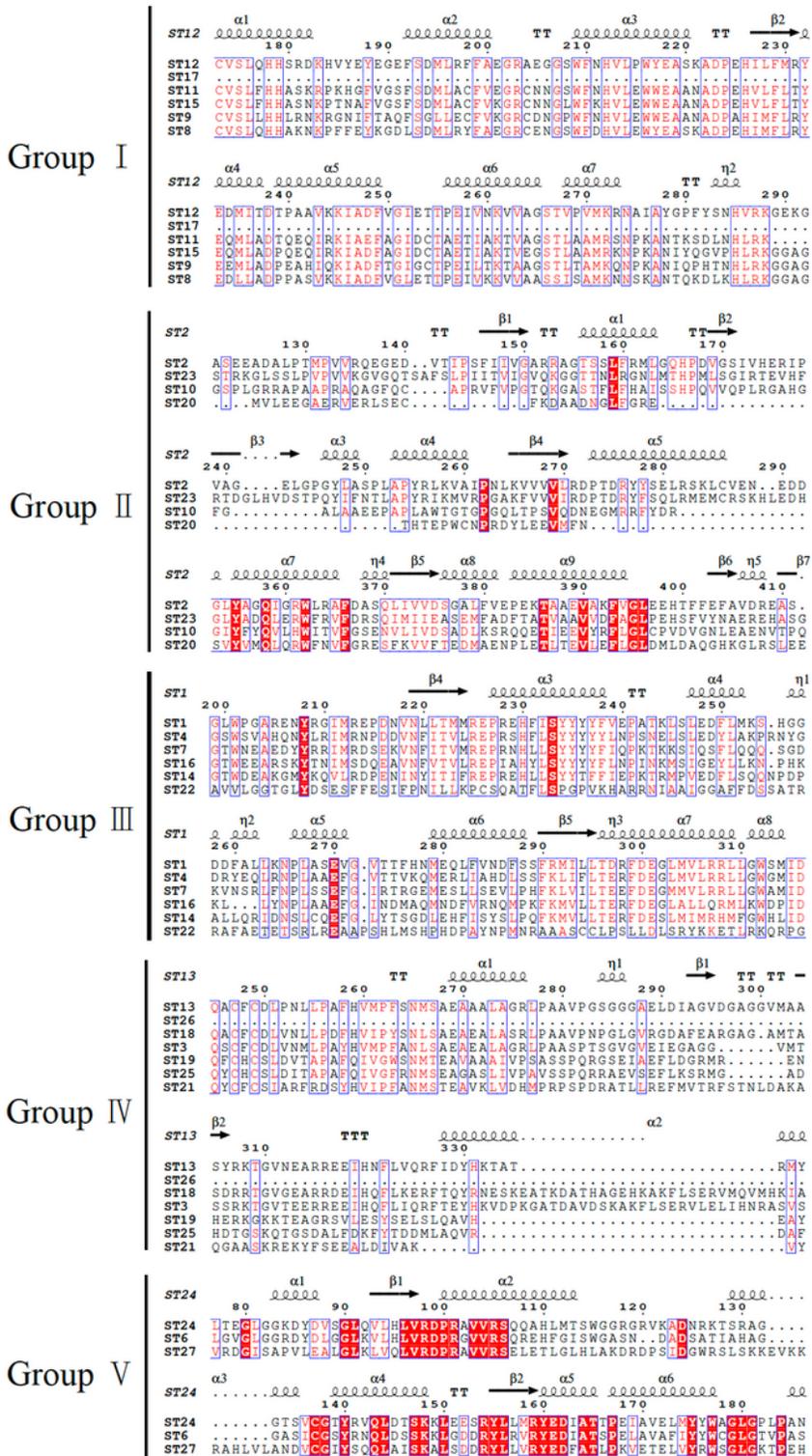


Figure 3

The multiple sequence alignment and secondary structures of 27 ST *S. japonica* amino acids (partial).

ST12 1 10 20 30 40 50  
 ST12 MAMWKREHVHKCTSTPFRELPVRPEAALPKWTPSKEE.EVKYAVKLCRGVQPQNVTDQDEAL  
 ST17 .....MPFRELPTRPRMIVPDWLPQAQDKEIQYTLSLVNGVYQMHKQEEQA  
 ST11 .....  
 ST15 .....MMVVPWELPTQKGDKTEVIYPLHLVNGTYQGPHQEEKA  
 ST9 .....  
 ST8 .....MPSSSVTSTPFRELPGRPKVVLPKWTPAKEEEVITYSLKTCRQVQNGVHDEAV

ST12 60 70 80 90 100 110  
 ST12 LLDLIDTFETREDDVFVCTFVKS<sup>▼</sup>GTTW<sup>—</sup>TQ<sup>—</sup>QIVALLN<sup>—</sup>GDQGEKSYRELVPWLEAYLG.  
 ST17 TNDLIDNFETRDDDVFVCTYVKS<sup>▼</sup>GTTW<sup>—</sup>TQ<sup>—</sup>QIISLLN<sup>—</sup>KGDQSTKNYTEVVPWLES<sup>—</sup>LAFS.  
 ST11 .....  
 ST15 TNDLIRTFETRSDSDVFVCTFVKS<sup>▼</sup>GTTW<sup>—</sup>TQ<sup>—</sup>HIINLLN<sup>—</sup>GGEPSHKNYTEVIPWLES<sup>—</sup>LTFN.  
 ST9 .....MKS<sup>▼</sup>GTTW<sup>—</sup>VQ<sup>—</sup>HIINLLN<sup>—</sup>KGEPSTKSYAEAVPWLEYLFTFG  
 ST8 LNDLIDNFETRDDDVFVCTYVKS<sup>▼</sup>GTTW<sup>—</sup>TQ<sup>—</sup>QIL<sup>—</sup>LLMN<sup>—</sup>KGEQGEKTYAEAVPWLE<sup>—</sup>TA<sup>—</sup>YFG.

Region I

ST12 120 130 140 150 160 170  
 ST12 .....GFLSDLEAKGWTFEKIVANPERFFFKTHASLD<sup>—</sup>DLP<sup>—</sup>SGKAPGVKVIYVCRNPKDV  
 ST17 .....GAKDKPAASQHE.....  
 ST11 .....MKT<sup>—</sup>HATL<sup>—</sup>KDLP<sup>—</sup>AGSATGLKVIYVARNPKDV  
 ST15 .DVRDEAPDLHREAKHWTLDALKSTANRRFMKTHANLKDLPAGSAKGLKVIYVARNPKDV  
 ST9 VKGQPASSPSNVEAQSWTLD<sup>—</sup>ELRSTPNRRFMKTHANLKDLPAGAAKGLKVIYVARNPKDV  
 ST8 .....GFLAGIEATGWTFEKIVANPDRRFFKSHANLKDLP<sup>—</sup>CGKAPGVKVIYVCRNPKDV

η1 → β1 →

ST12 α1 α2 α3 α4 α5 α6 α7 η2  
 ST12 180 190 200 210 220 230  
 ST12 CVSLQHHSRDKH<sup>—</sup>HYEYEGEFS<sup>—</sup>DMLRF<sup>—</sup>FAEGR<sup>—</sup>AEGS<sup>—</sup>WFNHVLPWYEA<sup>—</sup>SKADPEHILFMR<sup>—</sup>Y  
 ST17 .....  
 ST11 CVSLFHHASKR<sup>—</sup>PKHGFVGSFS<sup>—</sup>DMLACFVEGR<sup>—</sup>CNNGS<sup>—</sup>WFNHVLEWWEA<sup>—</sup>ANADPEHVLFLTY  
 ST15 CVSLFHHASNK<sup>—</sup>PTNAFVGSFS<sup>—</sup>DMLACFV<sup>—</sup>KGR<sup>—</sup>CNNGL<sup>—</sup>WFKH<sup>—</sup>VLEWWEA<sup>—</sup>ANADPEHVLFLTY  
 ST9 CVSLHHLRNK<sup>—</sup>RGNI<sup>—</sup>FAQF<sup>—</sup>SGLLECFV<sup>—</sup>KGR<sup>—</sup>CDNGP<sup>—</sup>WFNHVLEWWEA<sup>—</sup>ANADPEHIMFLRY  
 ST8 CVSLQHAKN<sup>—</sup>KPF<sup>—</sup>FERGDL<sup>—</sup>S<sup>—</sup>DMLRY<sup>—</sup>FAEGR<sup>—</sup>CENGS<sup>—</sup>WFNHVLEWYEA<sup>—</sup>SKADPEHIMFLRY

ST12 α4 α5 α6 α7 η2  
 ST12 240 250 260 270 280 290  
 ST12 EDMITDTPAAVK<sup>—</sup>KIADFV<sup>—</sup>G<sup>—</sup>LETPEIVN<sup>—</sup>KVVAG<sup>—</sup>STVP<sup>—</sup>VMKRN<sup>—</sup>AIAYGPFYSN<sup>—</sup>HVRK<sup>—</sup>GEKG  
 ST17 .....  
 ST11 EQMLADTQEQIR<sup>—</sup>KIAEFAG<sup>—</sup>IDCTAETIAKT<sup>—</sup>VAG<sup>—</sup>STLAMR<sup>—</sup>SNPKANTKSDLN<sup>—</sup>HLRK  
 ST15 EQMLADPQEQIR<sup>—</sup>KIADFAG<sup>—</sup>IDCTAETIAKT<sup>—</sup>VEG<sup>—</sup>STLAMR<sup>—</sup>KNPKANIYQGVPHLRKGGAG  
 ST9 EEMLADPEAHIQ<sup>—</sup>KIADF<sup>—</sup>TG<sup>—</sup>ICTPEILT<sup>—</sup>KTAA<sup>—</sup>AGSTLTAMKQ<sup>—</sup>NPKANIQPHNHLRKGAG  
 ST8 EDLLADPPASVK<sup>—</sup>KIADFV<sup>—</sup>GLETPEIVK<sup>—</sup>VVAASSI<sup>—</sup>SAMKNN<sup>—</sup>SKANTQKDLKHLRKGAG

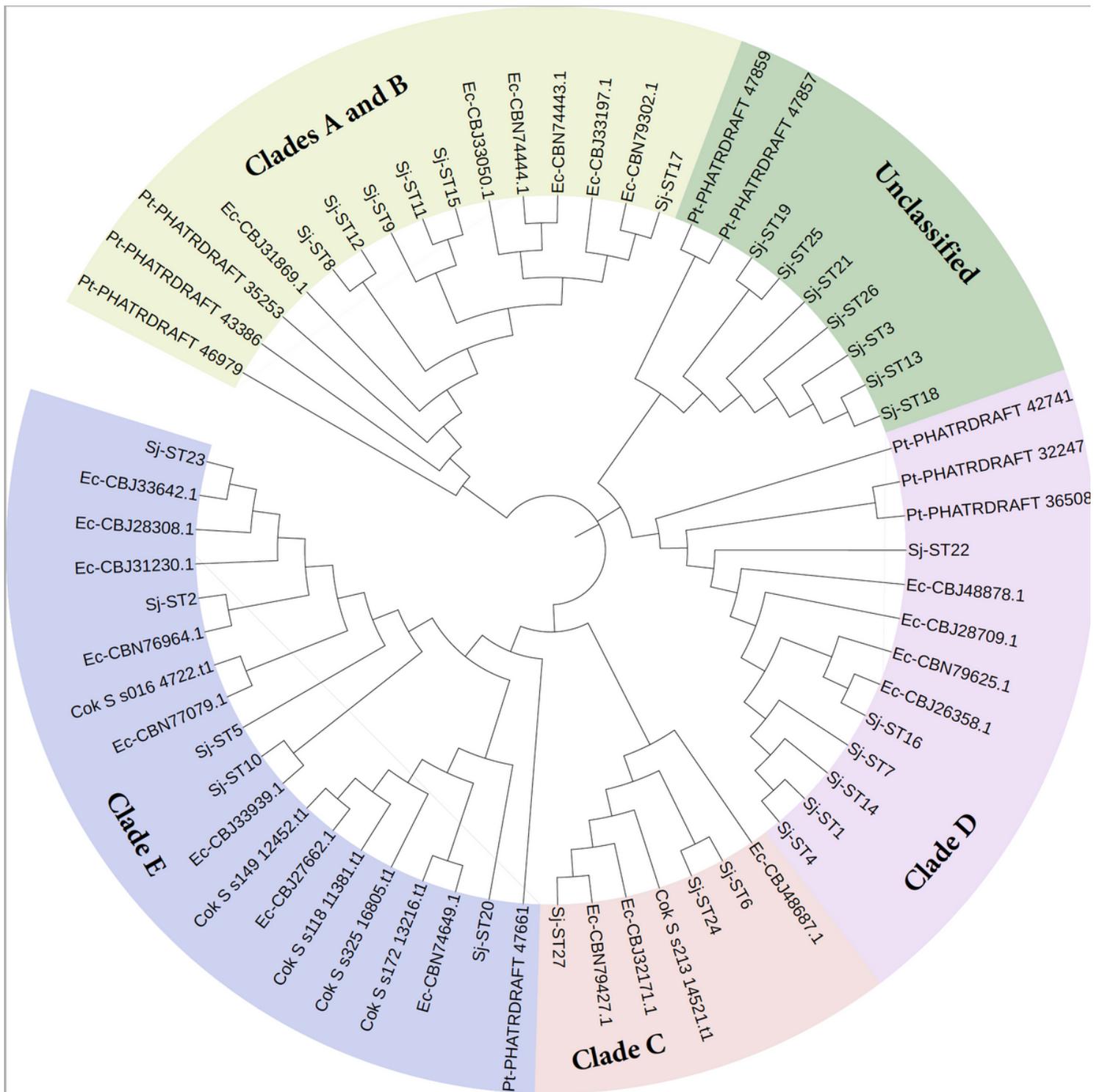
Region IV

ST12 300 310 320  
 ST12 GWRDYFTV<sup>—</sup>QGSEQFDALYRQ<sup>—</sup>QMAGSGLTMD<sup>—</sup>FGGGVEM  
 ST17 .....  
 ST11 .....  
 ST15 GWRDIFTVRESEAFDDIYRQHMKGSGLKMD<sup>—</sup>FGEGLVM  
 ST9 GWRDSFTVRESEAFDKIYRHEMKGSGLTMD<sup>—</sup>FGEGLIM  
 ST8 GWRDVFTVRQSENF<sup>—</sup>DALYRE<sup>—</sup>QMAGSGL<sup>—</sup>EMDFGEGLAM

Region IV

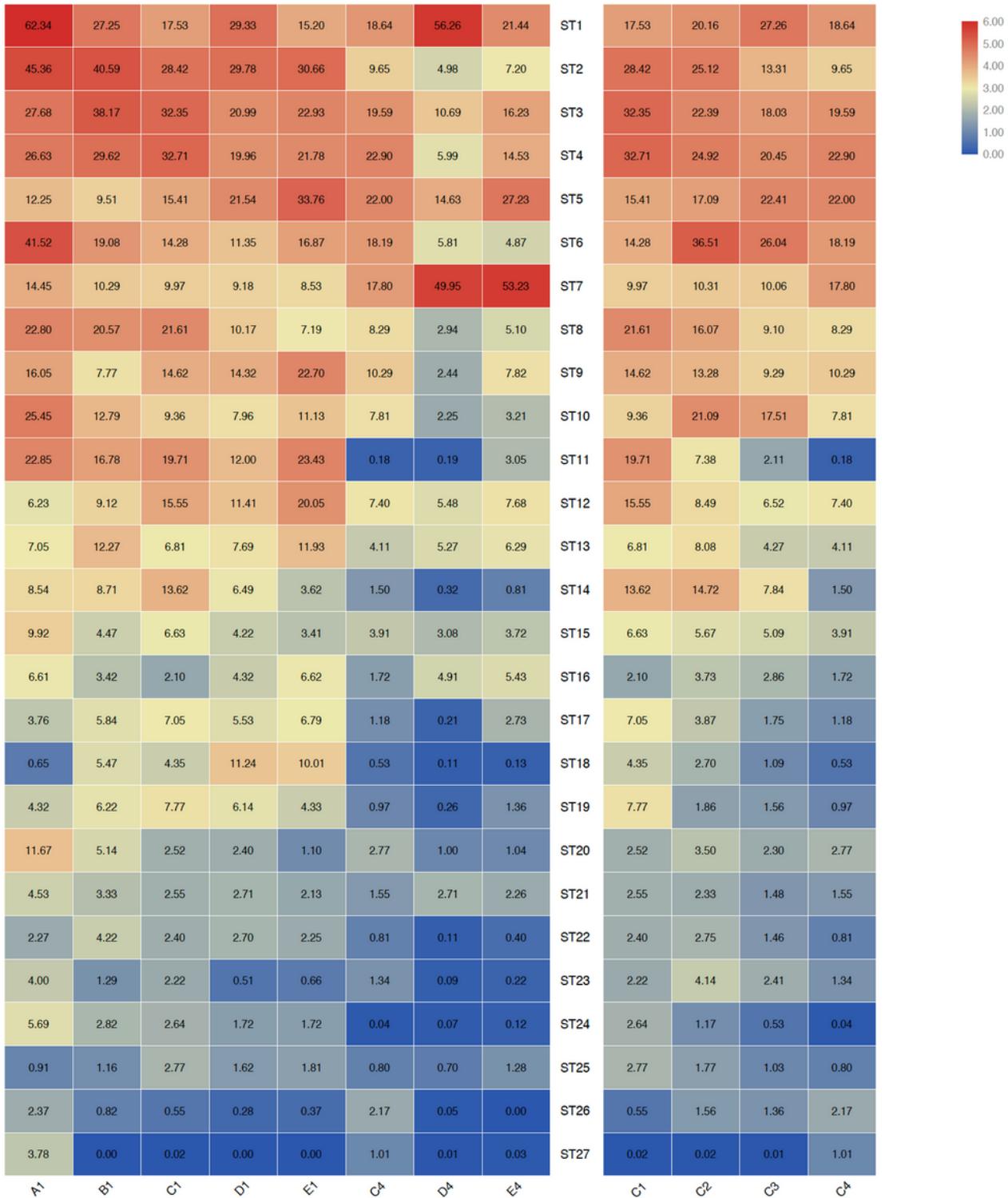
Figure 4

Two highly conserved motifs and sites in group I. They were respectively marked with black lines and triangles.



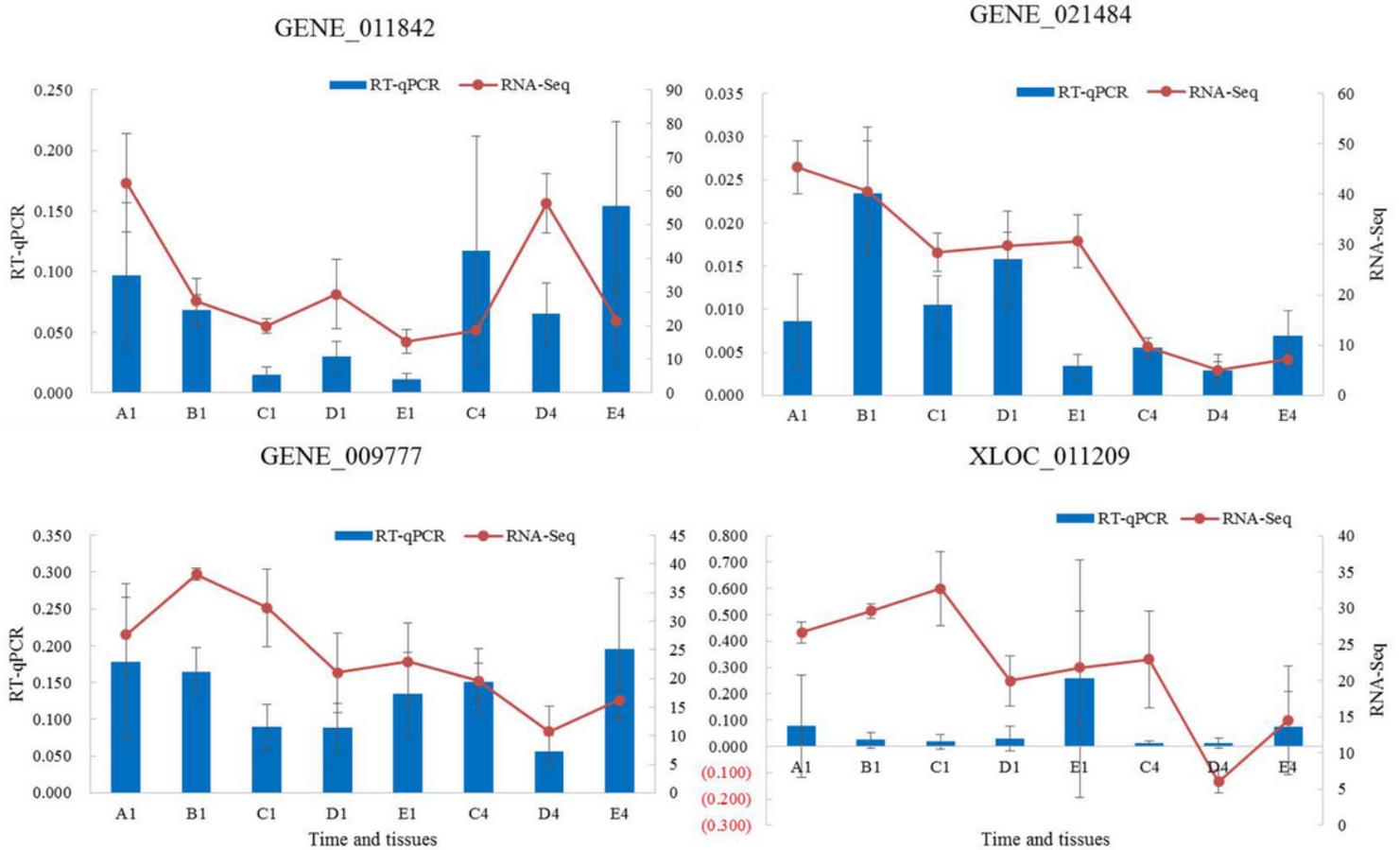
**Figure 5**

Phylogenetic tree of the ST proteins from *E. siliculosus*, *C. okamuranus*, *P. tricornutum* and *S. japonica*. A maximum likelihood phylogenetic tree was constructed based on full-length amino acid sequences of the 63 STs, with 1000 bootstrap replications, the WAG+F+G model, Gamma 3, partial deletion and 50% site coverage as the cutoff. These 63 ST proteins were clustered into four subfamilies. Unclassified proteins were colored in green.



**Figure 6**

Transcript profiles of the *S. japonica* ST genes in different tissues and developmental stages. The heatmap was constructed by TBtools.



**Figure 7**

Expression profiles of the four selected ST genes. Quantitative RT-PCR was used to determine the transcript levels of four STs, GENE\_011842, GENE\_021484, GENE\_009777 and XLOC\_011209. Data are represented by mean  $\pm$  standard deviation of two replicates. The relative transcript levels of selected four ST genes were calculated using the  $2^{-\Delta\Delta C_t}$  method with  $\beta$ -actin as the internal reference gene.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS3.xlsx](#)
- [TableS4.xlsx](#)
- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [FigS1.pdf](#)
- [FigS2.pdf](#)
- [TableS6.xlsx](#)
- [TableS7.xlsx](#)

- [TableS8.xlsx](#)
- [TableS5.xlsx](#)
- [Tables.docx](#)