Article

Chinese Science Bulletin

Marine Biology

July 2011 Vol.56 No.20: 2119–2130
doi: 10.1007/s11434-011-4546-4

Profiling of the transcriptome of *Porphyra yezoensis* with Solexa sequencing technology

YANG Hui, MAO YunXiang*, KONG FanNa, YANG GuanPin, MA Fei & WANG Li

College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China

Received January 15, 2011; accepted April 15, 2011

With high-throughput Solexa sequencing technology, we profiled *Porphyra yezoensis* transcriptomes from 8 different samples. More than 1200 megabases from 13333334 quality paired-end reads were generated, which were assembled into 31538 unigenes. Blast analysis showed that 56.7% unigenes were novel, which represented the specific genes of *Porphyra* and/or rhodophytes. Several hundreds of unigenes related to stress tolerance were discovered, including genes related to desiccation- (211) and high light-tolerance (31), flavonoid biosynthesis (10), reactive oxygen scavenging (48) and other stress-tolerance processes (208), which indicated there existed complex and diversity modes of stress tolerance in this species. A complete set of essential genes involved in the C3-(57) and C4-(44) carbon fixation pathway (except pyruvate phosphate dikinase) were discovered, which not only proved that they were actively transcribed but also clearly outlined the panoptic view of carbon fixation in *Porphyra*. Moreover, by statistically analyzing the types, proportions and frequencies of the interspersed repeats (TEs) and simple sequence repeats (SSRs), we discovered that the top three types of TEs were all retrotransposons and the trinucleotide was the absolute predominant type among SSRs, promoting our understanding of structural characteristics of the transcriptome. This study substantially improved the global view of the *Porphyra* genome and provided a valuable resource for future research.

***Porphyra yezoensis*, red algae, transcriptome, Solexa sequencing technology, expressed sequence tags, functional analysis, transposable elements, microsatellites**

Citation: Yang H, Mao Y X, Kong F N, et al. Profiling of the transcriptome of *Porphyra yezoensis* with Solexa sequencing technology. Chinese Sci Bull, 2011, 56: 2119–2130, doi: 10.1007/s11434-011-4546-4

Porphyra is a genus of marine red algae with considerable economic value. For hundreds of years, it has been cultivated in east Asian countries, such as China, Korea and Japan, where it remains one of the largest aquaculture industries today [1]. *Porphyra* has important fundamental and applied researching values [2]. The species in this genus reproduce both sexually and asexually. They perform a heteromorphic life cycle, in which a macroscopic foliose haploid thallus phase (gametophyte) and a microscopic diploid filamentous conchocelis phase (sporophyte) alter each other [3]. The entire life cycle can be completed in the laboratory within 2–3 months. The haploid genome of *Porphyra* is relatively small in size (2.7×10^8–5.3×10^8 bp) [4]. These features make it a good material for studying genetics and genomics. Moreover, *Porphyra* is a model organism of the intertidal zone. It has been undergoing profound environmental changes including desiccation, osmotic shock, intense sunlight exposure, high and/or freezing temperatures and other changes on a continuing basis [5]. Therefore, this alga is an ideal research tool to investigate the mechanisms of environmental tolerance.

The whole genome sequence of *P. yezoensis* is not available currently, making it difficult to investigate the alga integratively. Expressed sequence tags (ESTs) are important for studying the structure and function of functional genome. ESTs have played significant roles in accelerating gene discovery, improving genome annotation, elucidating

*Corresponding author (email: yxmao@ouc.edu.cn)

© The Author(s) 2011. This article is published with open access at Springerlink.com

phylogenetic relationships, facilitating breeding programs by providing SSR and SNP markers, analyzing large-scale gene expression and rapidly identifying transcripts involved in specific biological processes [6]. During the last few years, EST analysis of *P. yezoensis* has been conducted in many laboratories [7–9], and approximately 22000 ESTs have been deposited into the dbEST database at NCBI. However, most of the ESTs in the dbEST database are redundant.

Next-generation high-throughput RNA sequencing technology is a powerful and cost-effective tool for increasing the transcript complexity. The sequencing platforms, such as the Illumina/Solexa Genome Analyzer, GS FLX Titanium System and ABI/SOLiD Gene Sequencer, can sequence in parallel massive amounts of cDNAs [10]. They have been successfully used to investigate the transcriptomes of a wide range of organisms including *Pseudosciaena crocea*, zebrafish, *Aspergillus oryzae*, rice, mouse and human [11–16]. The next-generation sequencing technologies have also been used in other functional genomics researches including genome annotation, small ncRNA profiling and aberrant transcript detection, which are areas previously covered by microarrays [17]. In this study, we profiled the transcriptome of *P. yezoensis* using Illumina/Solexa sequencing technology with the aim to characterize its functional genome.

1 Materials and methods

1.1 RNA extraction

Sporophytes and gametophytes of *P. yezoensis* were cultured under different conditions (Table 1), harvested and weighed separately. Equal amount (in fresh weight) were mixed and frozen at −80°C immediately. Total RNA was extracted using Trizol (Invitrogen) according to the manufacturer's instructions. Residual DNA was removed away with RNase-free DNase I (TaKaRa) at 37°C for 30 min.

1.2 Construction of cDNA library

The cDNA library was constructed following the manufacturer's instructions (Illumina). In briefly, Sera-mag Magnetic oligo(dT) Beads were used to isolate poly(A) mRNA from total RNA. To avoid priming bias, the mRNA was fragmented before cDNA synthesis. The first-strand cDNA and the second-strand cDNA were synthesized using the fragments as templates and random hexamer primers, purified with QiaQuick PCR Purification Kit (Qiagen) and resolved in EB buffer, with them poly(A) and sequencing adaptors were ligated after end reparation. The fragments in appropriate lengths were retrieved from agarose gel, PCR amplified sequenced using Illumina HiSeq™ 2000 (Beijing Genomics Institute, Shenzhen, China).

1.3 Assembly and annotation

Low quality reads and only sequencing adaptors containing reads were removed ahead of assembly. The quality reads were assembled into unigenes with SOAPdenovo [18]. The unigenes were annotated by aligning with the deposited ones in diverse protein databases including NCBI non-redundant protein (Nr), UniProt/Swiss-Prot, KEGG and COG. The best one was selected from the matches with an *E*-value of less than 10^{-5}. The unigenes were functionally classified with Blast2GO program [19] and WEGO software [20]. To identify potential novel genes, non-annotated unigenes were predicted by ESTScan [21] and translated into peptide sequences.

1.4 Coverage determination

All ESTs of *P. yezoensis* were downloaded from the dbEST database at NCBI and assembled with TGICL and PHRAP (overlap≥100 bp, identity≥96%) in order to obtain EST unigenes from GenBank. Those longer than 200 bp were

Table 1 Different samples of *P. yezoensis*

Phase	Medium	Cultivation conditions
Sporophytes		
Free-living conchocelis filamentoas	Natural seawater with the Provasoli's enrichment solution medium (PES)	18°C, 24 μmol photos m^{-2} s^{-1}, light:dark=12:12
Free-living conchosporangium	Natural seawater with the Provasoli's enrichment solution medium (PES)	24°C, 24 μmol photos m^{-2} s^{-1}, light:dark=12:12
Gametophytes		
Thallus	Natural seawater	9°C, 24 μmol photos m^{-2} s^{-1}, 4 h
Thallus	Natural seawater added with deionized water of equal volume	9°C, 24 μmol photos m^{-2} s^{-1}, 4 h
Thallus	Natural seawater added with 33 g NaCl per litre	9°C, 24 μmol photos m^{-2} s^{-1}, 4 h
Thallus	Natural seawater	9°C, 1500 μmol photos m^{-2} s^{-1}, 4 h
Thallus	Natural seawater	9°C, dark , 4 h
Thallus	Drought	4°C, dark , 4 h

searched with BLASTN 2.2.21 [22] ($E \leqslant 1.00\times10^{-7}$) against unigenes from this study and used for calculating coverage.

1.5 Identification of transposable elements (TE)

To identify the putative transcriptional TE, the ESTs from this study and those from GenBank were screened with RepeatMasker (http://www.repeatmasker.org/) using consensus sequences in Repbase as queries. The unigenes were also aligned with those in TIGR Plant Repeat Databases (http://plantrepeats.plantbiology.msu.edu/index.html) with BLASTN at an *E*-value cut off of 1×10^{-5}; TEs with amino acid homology to known TE ORFs were identified through searching against Repbase database and TIGR Plant Repeat Databases with *E*-value cut off of 1×10^{-5} and also with Protein-based RepeatMasking (http://www.repeatmasker.org/cgi-bin/RepeatProteinMaskRequest).

1.6 Identification of SSRs

Microsatellites (SSRs) with 2–6 bp repeats were identified from both ESTs from this study and ESTs from GenBank with MISA program (MIcroSAtellite identification tool) [23]. Only perfect repeats were collected. The minimum repeat number was 7 for di-, 5 for tri- and 4 for tetra-, penta- and hexanucleotides, respectively [24,25].

2 Results

2.1 Transcriptome assembly

A total of 13333334 quality (with a score of >20) paired-end reads were obtained with an overall length of more than 1200 megabase, which accounted for 81.0% of the raw data. The GC content of the transcriptome was 63.2%. Quality reads were assembled into 197227 contigs, and further clustered into 40181 scaffolds and grouped into 31538 unigenes with a total length of 13211994 nt and a mean length of 419 nt (Table 2, Figure 1).

2.2 Coverage determination

Before this study, a total of 22074 ESTs of *P. yezoensis* had been deposited in the GenBank. After downloading and assembling, a data set consisting of 5475 unigenes was obtained. Of these unigenes, 5013 were longer than 200 bp in length, which were used for further analysis. The comparison showed that 4507 unigenes from GenBank matched with 7246 unigenes from this study, which represented 89.9% of the ESTs obtained from GenBank. The total length of the identical sequences of the two data sets was 2134453 nt accounting for 70.5% of the ESTs from GenBank (Figure 2). It indicated that the pool of transcripts sequenced well covered the whole transcriptome of *P. yezoensis*.

2.3 Unigene annotation and classification

Of all the unigenes obtained, 12815 (40.6%) significantly matched with 66209 in the Nr database. Among them, 311

Table 2 Statistics of *P. yezoensis* reads

Item	Data
Number of total reads	13333334
Total nucleotides (megabase)	>1200
Contigs	197227
Scaffolds	40181
Unigenes	31538
Mean length of contigs (nt)	138
Mean length of scaffolds (nt)	355
Mean length of unigenes (nt)	419

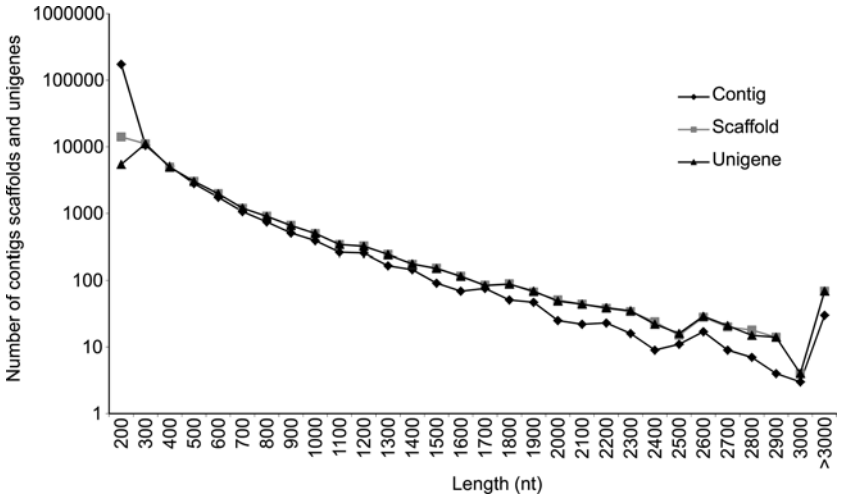

Figure 1 Length distribution of contigs, scaffolds and unigenes.

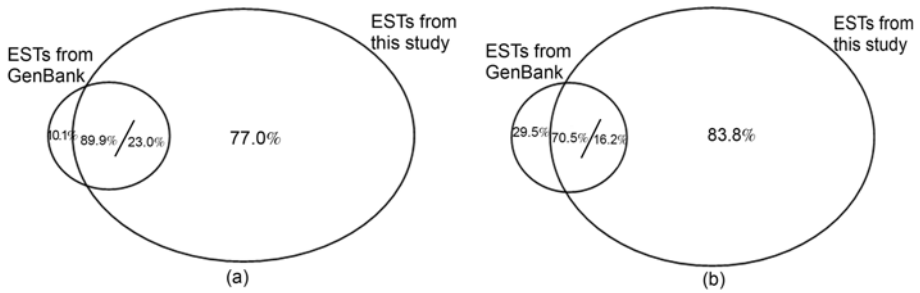

Figure 2 Comparison of the proportion of identical sequences in ESTs from this study and ESTs from GenBank. (a) Represents the proportion of identical unigene number; (b) represents the proportion of the total length of identical unigenes.

were homologous with those of *P. yezoensis* and 55 to those of other species in genus *Porphyra*. Searching against Swiss-prot database found 11364 unigenes (36.0%) hit 51387 deposited. Searching against COG database found 7067 unigenes were assigned to 25 functional categories. The categories of "translation, ribosomal structure and biogenesis", "general function prediction only" and "cell cycle control, cell division, chromosome partitioning" were the first three largest (Figure 3). In addition, 166 unigenes were found to be involved in "defense mechanisms" including ABC-type bacteriocin/lantibiotic exporters, ABC-type multidrug transport system, restriction endonuclease, Na^+-driven multidrug efflux pump, uncharacterized vancomycin resistance protein, beta-lactamase class C and other penicillin binding proteins. A total of 8308 unigenes were assigned to at least one pathway of 120 pathways in total. Of these pathways, metabolic (23.8%) and spliceosome (21.3%) contained the most unigenes. A total of 2390 unigenes (7.6%) were assigned to signed to at least one of 10923 available GO terms. Among the unigenes, 1387 were assigned to at least one GO term in the biological process category, 1741 in the molecular function category and 1517 in the cellular component category. Enriched GO categories involving more than 1000 unigenes belonged to the subcategories of cell, cell part, catalytic activity, cellular process, binding and metabolic process (Figure 4).

In total, 13671 unigenes (43.3%) were annotated, while 17867 (56.7%) were not. Of the failure in annotation, 5045 unigenes were predicted to be possibly novel in *P. yezoensis* according to the results of ESTscan program.

2.4 Gene tags related to the stress tolerance

In plant, the adaptation or tolerance to the environmental stress includes complex mechanisms. In this study, several hundreds of gene tags related to the stress tolerance were

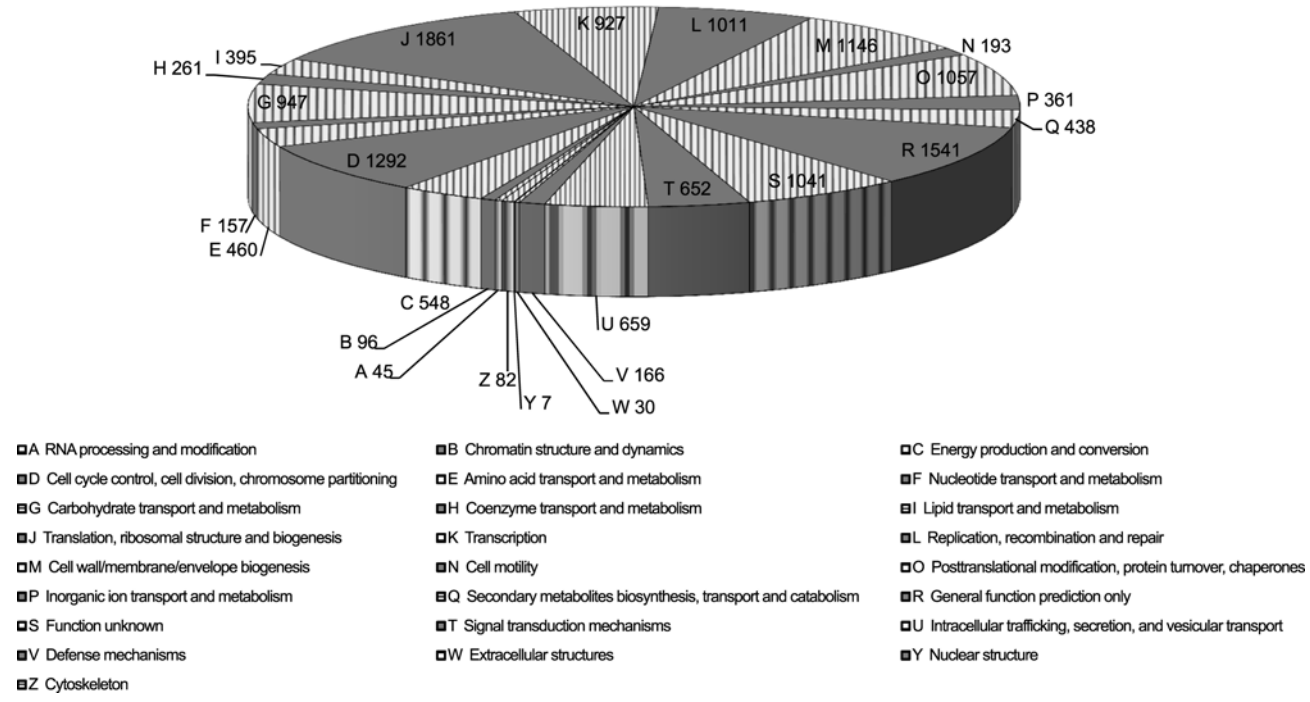

Figure 3 COG functional categorization of unigenes. The numbers represent the amount of unigenes included in each functional category.

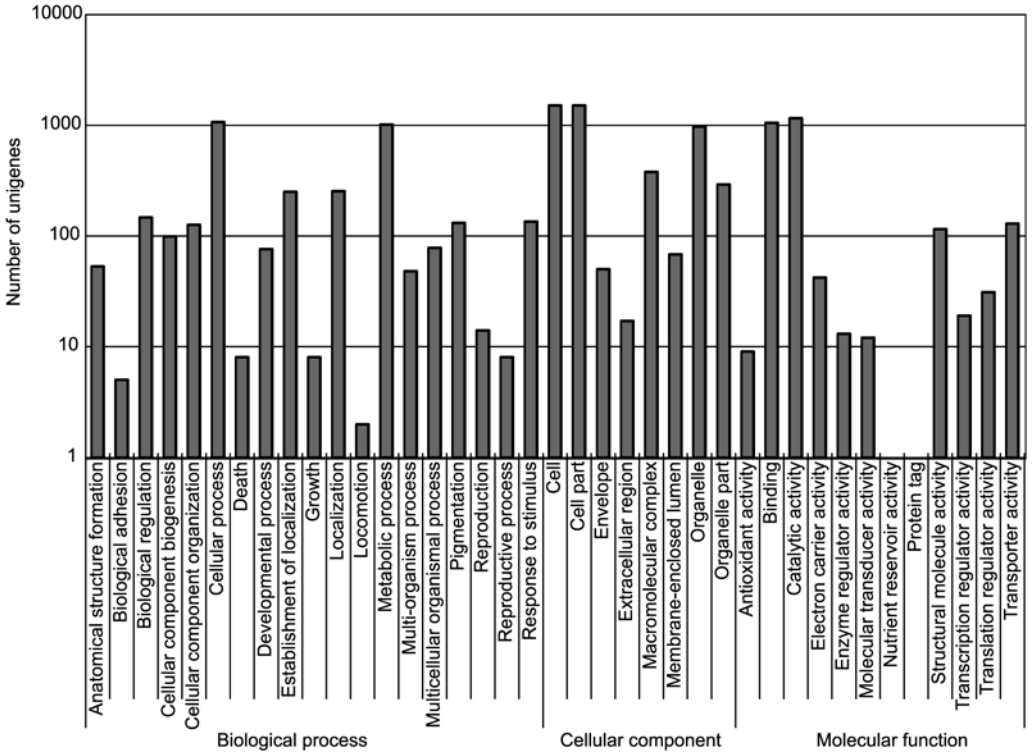

Figure 4 GO classification for unigenes.

discovered, including desiccation-tolerance related genes (211), high light-tolerance related genes (31), genes involved in flavonoid biosynthesis pathway (10), genes related to other stress-tolerance processes (208) and genes encoded a series of reactive oxygen scavenging enzymes (48) (Tables 3 and 4). The results not only indicated that there existed diversity modes of stress tolerance in *P. yezoensis* but also provided a batch of gene candidates for deciphering the molecular mechanisms of stress tolerance.

2.5 Gene tags related to the CO_2 fixation

Carbon fixation is an important biological process to all photosynthetic organisms. With the help of bioinformatic techniques all the essential genes related to C3-pathway (57 unigenes) were discovered from the transcriptome (Table 5), which provided the unequivocal molecular evidence that all the C3-pathway genes were actively transcribed and there existed C3-pathway in *P. yezoensis* (Figure 5). It was interesting to note that 44 enzyme-encoding unigenes invovled in C4-pathway were found, covering almost all enzymes needed for C4-carbon fixation except the pyruvate phosphate dikinase (EC 2.7.9.1) (Table 5, Figure 5). The results helped us to clearly outline the panoptic view of carbon fixation of this species.

2.6 Active transposable elements (TE)

The presence of TEs in EST collections indicates they are actively transcribed. A total of 614 and 262 transposon-like

Table 3 Some stress-tolerance related genes in *P. yezoensis*

Stress-tolerance related genes	Unigenes
Desiccation-tolerance related genes	
Actin related proteins	164
Peptidyl-prolyl *cis-trans* isomerases	23
Tubulin related proteins	20
TCP-1 subunit genes	3
Late embryogenesis abundant(LEA)protein-related	1
High light-tolerance related genes	
Phycoerythrins	9
Phycocyanins	4
Light-harvest proteins	4
Light inducible protein/nucleases	4
3-dehydroquinate synthase/*O*-methyltransferase fusion	3
Blue-light photoreceptors	2
Beta-carotene biosynthetic proteins	2
Chlorophyll binding proteins	2
Protochlorophyllide reductase chloroplast precursor	1
Flavonoid biosynthesis	
P-coumarate 3-hydroxylase	2
Caffeoyl-CoA *O*-methyltransferase	2
Flavonoid 3′-monooxygenase	2
Anthocyanidin reductase	2
Chalcone synthase	1
Trans-cinnamate 4-monooxygenase	1
Other stress-tolerance processes	
Heat shock proteins & cochaperones	96
Ubiquitin related proteins	84
Carbonic anhydrases	12
Glutathione *S*-transferases	12
Haloacid dehalogenase-like hydrolases	2
Vanadium-dependent bromoperoxidase 1	2

Table 4 Key enzymes involved in the antioxidant system in *P. yezoensis*

Reactive oxygen scavenging enzymes	Unigenes
Superoxide dismutase	10
Mn-SOD	5
Cu/Zn-SOD	4
Fe-SOD	1
Dehydroascorbate reductase	7
Peroxiredoxin	6
Glutathione peroxidase	5
Catalase	5
Thioredoxin reductase	4
Ascorbate peroxidase	4
Glutathione reductase	3
Heme oxygenase	2
Haem peroxidase	2
Total	48

sequences were found in ESTs from this study (31538 unigenes) and those from GenBank (5013 unigenes), respectively (Figure 6). In both ESTs, the top three types were related to retrotransposons, of which Gypsy accounted for the richest. In ESTs from this study, R1 was the second abundant, followed by Copia and CACTA. However, in the ESTs from GenBank, Copia was the second abundant, followed by R1 and Ngaro. As a whole, the majority of TEs were retrotransposons and the remaining were DNA transposons. Only one putative miniature inverted-repeat transposable element (MITE) was found, implying its scarceness in *P. yezoensis*.

2.7 Characterization and comparison of SSRs

A total of 1047 perfect SSRs were found in ESTs from this study, while 287 SSRs were from those of the GenBank (Table 6). On an average, there existed one SSR every 11.6 and 10.6 kb, respectively, in our data set and in that of GenBank. The abundance of SSR types from more to less was tri-, di-, tetra-, hexa- and pentanucleotide repeats in order. Among all types, the trinucleotide was absolutely the predominant type with the ratio of 78.8% on average. It reconfirmed the fact that trinucleotide repeats in the coding regions would not cause frameshift mutation that could silence genes, but only would increase the variation of genes. The most abundant motifs were CCG in tri-, CCCG in tetra-,

Table 5 Statistics of C3/C4-pathway related enzymes

Enzyme codes	Enzyme names	Unigenes
C3-pathway		57
EC 1.2.1.13/59	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) (GAPDH)	2
EC 2.2.1.1	Transketolase	8
EC 2.7.1.19	Phosphoribulokinase	2
EC 2.7.2.3	Phosphoglycerate kinase (PGK)	3
EC 3.1.3.11	Fructose-1,6-bisphosphatase (FBPase)	20
EC 3.1.3.37	Sedoheptulose-bisphosphatase (SBPase)	3
EC 4.1.1.39	Ribulose bisphosphate carboxylase/oxygenase (Rubisco)	3
EC 4.1.2.13	Fructose-bisphosphate aldolase	9
EC 5.1.3.1	Ribulose-phosphate 3-epimerase	4
EC 5.3.1.1	Triose-phosphate isomerase (TIM)	2
EC 5.3.1.6	Ribose-5-phosphate isomerase	1
C4-pathway		44
EC 1.1.1.37	Malate Dehydrogenase (NAD^+) (MD)	1
EC 1.1.1.39	Malate Dehydrogenase (decarboxylating) (NAD^+) (MD)	4
EC 1.1.1.40	Malic enzyme (ME)	7
EC 1.1.1.82	Malate dehydrogenase ($NADP^+$)	1
EC 2.6.1.1	Aspartate aminotransferase (AST)	10
EC 2.6.1.2	Alanine aminotransferase (ALT)	4
EC 2.7.1.40	Pyruvate kinase	10
EC 4.1.1.31	Phosphoenolpyruvate carboxylase (PEPC)	4
EC 4.1.1.49	Phosphoenolpyruvate carboxykinase (PEPCK)	3
Total		101

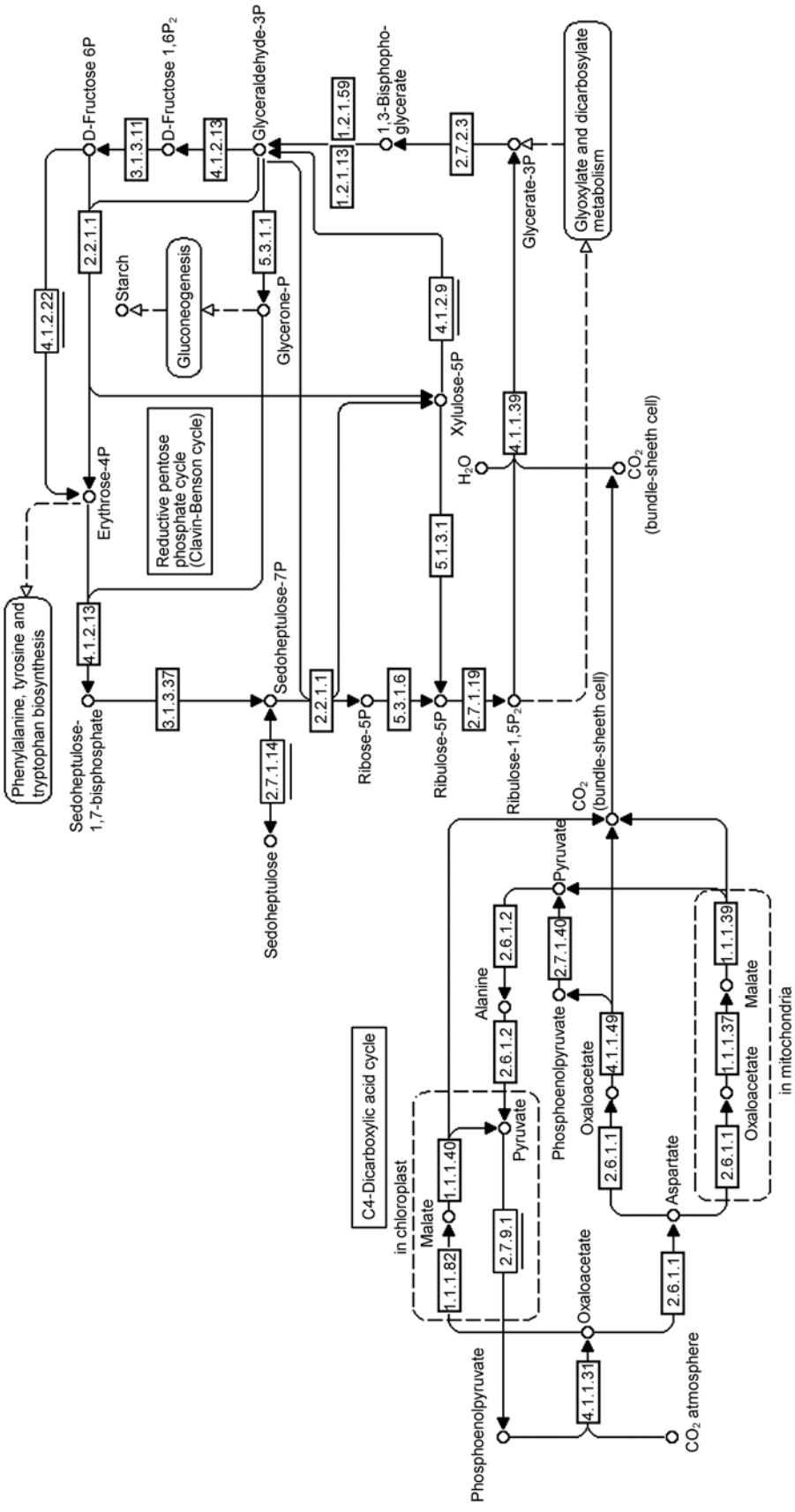

Figure 5 Putative C3/C4-pathway in *P. yezoensis* generated by KEGG. The numbers within the small boxes are enzyme codes. The boxes without underlines indicate enzymes identified in ESTs from this study; The boxes with underlines, not identified. The names and numbers of enzymes are listed in Table 5.

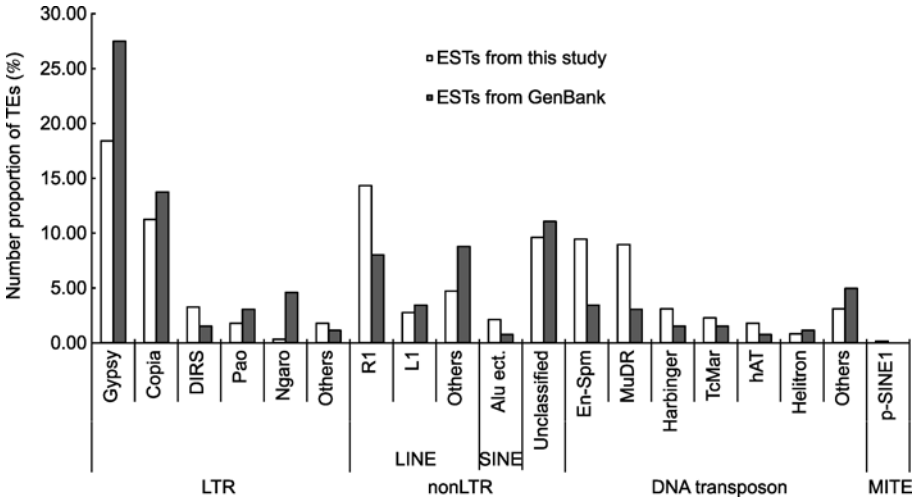

Figure 6 Distribution of TEs with transcriptional activity in *P. yezoensis*.

Table 6 Characterization of SSRs in transcribed sequences of *P. yezoensis*

Dinucleotide				Trinucleotide				Tetranucleotide				Pentanucleotide				Hexanucleotide			
ESTs from this study		ESTs from GenBank		ESTs from this study		ESTs from GenBank		ESTs from this study		ESTs from GenBank		ESTs from this study		ESTs from GenBank		ESTs from this study		ESTs from GenBank	
AC[a)]	57[b)]	AG	15	CCG	417	CCG	108	CCCG	9	CCCG	6	AGGGG	4	AGGGG	1	ACCGCC	7	ACCGCC	2
AG	29	CG	8	AGC	143	AAC	46	AGGG	7	ACGC	3	AAGGG	2			ACGGCG	7	ACCGCG	1
CG	28	AC	5	ACC	87	AGC	44	ACAT	6	CCGG	2	CCGCG	2			ACCGTC	3	ACCGTC	1
AT	1			ACG	65	ACC	11	ACGC	6	ACAT	1	AAAAC	1			ACGCCG	3	ACTCGG	1
				AGG	59	ACG	10	AGCC	5	ACCG	1	ACAGT	1			AGCGGC	3	AGCGCC	1
				AAG	18	AGG	9	CCGG	5			ACCCC	1			AGGCGC	3	AGCGGC	1
				AAC	17	AAG	5	ACCG	3			ACGCG	1			AACAGC	2		
				ATC	6	ACT	4	ACCT	2			ACGGG	1			AATAGC	2		
				AAT	1	ATC	1	ACGG	2			CCCCG	1			ACCGGG	2		
								AGGC	2							AGCGCC	2		
								AAAC	1							AGGATG	2		
								AAAG	1							AAAAAC	1		
								AACC	1							AACCCT	1		
								AAGC	1							AACGCC	1		
								ACAG	1							AAGGAC	1		
								ACCC	1							AAGGCC	1		
								ACTC	1							AAGGGG	1		
								ACTG	1							ACAGCG	1		
								AGAT	1							ACCCGC	1		
																ACCGCG	1		
																ACGCGC	1		
																ACTCGG	1		
																ACTGAG	1		
																CCCCCG	1		
Total	115 (11.0%)[c)]		28 (9.8%)	813 (77.7%)		238 (82.9%)		56 (5.4%)		13 (4.5%)		14 (1.3%)		1 (0.4%)		49 (4.7%)		7 (2.4%)	

a) Repeat motif type. b) Total number of a given motif type. c) Percentage of a given motif type in ESTs from this study or ESTs from GenBank.

AGGGG in penta- and ACCGCC in hexanucleotide repeats, which were obviously rich in GC. Among the total of 143 dinucleotide repeats, AC, AG and CG were the top three motifs while AT motif was very scare. As a whole, the transcribed SSRs showed significantly high GC content (78.8% in this study and 74.3% in GenBank). Comparison of the SSRs from *P. yezoensis* with those from other plants showed that the order of the proportion of EST-SSR types was the same among different species, however, the frequency was quite different (Table 7).

3 Discussion

3.1 cDNA library and coverage of the transcriptome

The utilization of Solexa sequencing technology resulted in nearly 120 times more nucleotides and 6 times more unigenes than those from *P. yezoensis* ESTs registered in the GenBank. Considering the temporal and spatial specificity of the pattern of gene expression, in order to reveal as many genes of *P. yezoensis* as possible the cDNA library for sequencing was constructed from the mixed samples consisting of both gametophyte and sporophyte at different developmental stages and under different stress conditions. The deep sequence coverage added with a wide range of expressed RNA under various conditions make it possible to detect very rare transcripts and find the maximal genes in the genome. Though the actual gene number of *P. yezoensis* still remains unknown, as many as 31538 unigenes provide us considerable genomic information for this species. Despite a variety of ways were used for annotating the transcriptome, more than half of the unigenes (56.7%) still could not be matched with any known genes, which might represent the specific genes of *P. yezoensis* and/or rhodophyta.

3.2 Cabon fixation pathway

The calvin cycle (C3-pathway) is a well known metabolic pathway for carbon fixation in plants, algae and cyanobacteria. The identification of some genes related to C3- pathway of *P. yezoensis* were reported in the past few years [26,27]. In this study, with the help of bioinformatic techniques all the essential genes related to C3-pathway were discovered from the transcriptome, which not only provided the unequivocal molecular evidence that all the C3-pathway genes were actively transcribed and there existed C3-pathway in *P. yezoensis* but also helped us to clearly outline the panoptic view of calvin cycle of this species.

Hatch-Slack pathway, also known as C4 carbon-fixation pathway, has long been studied in higher plants (C4 plants) with much higher photosynthetic rate than C3 plants. By genome sequencing, all the required genes involved in C4 photosynthesis were presented in unicellular green algae *Ostreococcus tauri*, the smallest free-living eukaryote yet described [28]. In addition, some intermediate products of C4-pathway had been detected in diatoms, brown alga, euglenoids, and dinoflagellates, and C4-like photosynthetic characteristics were also detected in green algae and diatom. In rhodophytes, genes of the key enzymes in the C4 carbon-fixation pathway, such as PEPCK and aspartate aminotransferase (AST) were abundantly transcribed in *P. haitanensis* sporophytes by analyzing the ESTs while PEPC was not detected [29]. In this study, except for pyruvate phosphate dikinase, all genes involved in C4-pathway were identified including PEPC, the key enzyme responsible for accepting CO_2 from air. The metabolic map including all detected enzyme genes involved in carbon fixation drawed by KEGG showed a nearly complete pathway of carbon fixation that was similar to higher plants. Hence, a C4-like carbon-fixation pathway might play a special role in fixing inorganic carbon dioxide in *P. yezoensis*.

3.3 Stress-tolerance mechanisms

As a model of intertidal species, *P. yezoensis* has evolved an outstanding ability to cope with environmental stress including desiccation, high light and nutrient limitation caused by the tidal changes. Although the mechanisms that

Table 7 Comparison among EST-SSRs from different plants [a)]

Source	Di-	Tri-	Tetra-	Pentanucleotide	Total length of EST (kb)	Frequency (kb)
Rice	2	1	3	4	17304	3.4
Soybean	2	1	3	4	3675	7.4
Maize	2	1	3	4	6411	8.1
Tomato	2	1	3	4	4444	11.1
Porphyra yezoensis	2	1	3	4	13212	11.6
Arabidopsis thaliana	2	1	3	4	14808	13.8
Poplar	2	1	3	4	1880	14.0
Cotton	2	1	3	4	4788	20.0

a) Proportion of unit is 1 > 2 > 3 >4. Frequency is the average distance of ESTs containing one SSR.

protect membranes, photosystems and other cellular structures remain poorly understood in *P. yezoensis* [30], some clues are revealed by the gene expression when they are under stress conditions. The photosynthetic rates are increased in *P. yezoensis* from exposure to the air at low tide [31]. However, with the extension of exposure time under high light, excess light will lead to photosystem damage. A large number of genes related to light absorbing were found in transcribed sequences of *P. yezoensis* including genes of phycoerythrin, phycocyanin, beta-carotene biosynthetic protein, chlorophyll binding protein and the like. Beta-carotene may do good to transfer excitation energy to heat that will be dissipated [32], and the damage done such as the accumulation of reactive oxygen species (ROS) are cleaned up by reactive oxygen scavenging enzymes [33]. In this study, a variety of reactive oxygen scavenging enzymes were identified including glutathione peroxidase, catalase, superoxide dismutase and the like. Superoxide dismutases are important antioxidant defense in nearly all cells exposed to oxygen. Four known metalloforms of SOD enzymes have been identified: they contain either Fe, Mn, Cu and Zn, or Ni, and the first three types (Fe-, Mn-, Cu/Zn-SOD) have been reported in eukaryotic algae, whereas, only MnSOD has been reported in rhodophytes [34]. In this study, a total of ten SOD genes were identified including 5 for Mn-SOD, 4 for Cu/Zn-SOD, and 1 for Fe-SOD types. The finding of new SOD types in *P. yezoensis* provide us with new clues in understanding the reactive oxygen metabolism in rhodophytes further.

MAAs are a group of UV-protective metabolites in aquatic organisms, which have been found in *Porphyra* spp. [35]. However, very little is known about the genes involved in the biosynthesis of these compounds. In this study, three 3-dehydroquinate synthase/*O*-methyltransferase fusion genes, which were proposed to catalyze the formation of deoxygadusol, the core of all MAAs, were found in *P. yezoensis* [36]. Flavonoids are another group of compounds thought to protect against ultraviolet radiation in terrestrial plants [37]. A pathway of flavonoid biosynthesis including many homologues genes related to the terrestrial plant was identified in *P. yezoensis*, and this pathway was also found in *Ectocarpus* but was completely absent from diatom or green algal genomes [38]. It was worth noting that LI818 family proteins, the light-stress related LHCs, were failed to be identified in the pool of unigenes. Dittami et al. [39] suggested a possibility that LI818 proteins might have originated in an ancient chlorophyll *a*/*c*-containing organism and could have been later transferred to the green lineage, thus, they would not be found in red algae. Results in this study provide favorable evidence for this hypothesis.

Desiccation stress leads to substantial loss of water in *P. yezoensis*, which threatens the membrane integrity of cell. Three TCP-1 subunit genes and a large number of *actin* and *tubulin* genes were found in this study. TCP-1 forms part of cytosolic type II chaperonin CCT that has an important role in the folding of actin and tubulin [40]. Active expression of these proteins plays an important role in maintaining cellular integrity during desiccation/rehydration cycles. Late-embryogenesis-abundant (LEA) proteins have been found in many desiccation-tolerant plant and rehydrating bryophyte tissues, and have been suggested to be associated with desiccation tolerance [41,42]. One gene encoding LEA protein was identified in this study, which implied that drought tolerance in a manner analogous to LEA proteins in plant and bryophyte might also exist in *P. yezoensis*. Besides, many other stress-tolerance related genes were found including genes of heat shock proteins and cochaperones, light inducible proteins, ubiquitin related proteins and the like. The found mechanisms together with many other unfound make *P. yezoensis* a good model for environmental tolerance research.

3.4 TEs and SSRs

Transposable elements are known to play a prominent role in the evolution of genome complexity across eukaryotes by disrupting genes, mediating chromosome rearrangements and providing alternative promoters, exons, terminators and splice junctions [43]. More and more studies show that TEs have a capacity for affecting adjacent genes by altering transcriptional regulation [44–47], and retrotransposon transcription has a key influence upon the transcriptional output of genome [48]. In this study, many different kinds and abundant numbers of TEs were identified in the pool of transcripts, which provided the evidence that these elements not only played key roles in the diversity of transcripts and the complexity of functional genome but also might contribute to the adaptation and tolerance to the environmental stress.

EST-derived SSRs are functional markers for genetic research. They are derived from protein-coding sequences as well as untranslated regions (UTRs), and can be transferred between species since they are related to genes that are relatively well conserved among taxa [49]. The consistency of the types and the proportion was quite well by comparing the SSRs from this study with those from GenBank, which reflected the actual distribution of SSRs in functional genome of *P. yezoensis*.

The characteristics of EST-SSRs have been investigated in many plants based on a variety of distinct mining criteria, which often lead to different results even using the same data. To facilitate the comparative analysis of the type and frequency of EST-SSRs among species, the same motif parameters were adopted for SSR screening in all data. EST-SSRs in different plants showed the same order of proportion from di- to pentanucleotide, implied that it was a general character of EST-SSR distribution in plant. However, the frequency of EST-SSRs was quite different among plants. In *P. yezoensis*, it exhibited a relatively moderate frequency, which was less prevalent than rice, soybean and

maize, similar to tomato, but more prevalent than *A. thaliana*, poplar and cotton. It may suggest that the degree of genetic variability caused by simple sequence repeats in the genes of *P. yezoensis* is moderate.

This work was supported by the National Natural Science Foundation of China (30972247, 30700621), Special Fund for Agro-scientific Research in the Public Interest (200903030), National High Technology Research and Development Program of China (2006AA10A402, 2006AA10A413), Program for New Century Excellent Talents in University (NCET-06-0596), and Promotive Research Fund for Young and Middle-aged Scientisits of Shandong Province (2008BS06002).

1 Mumford T F, Miura A. *Porphyra* as food: Cultivation and economics. In: Lemby C A, Walland J R, eds. Algae and Human Affairs. Cambridge: Cambridge University Press, 1988. 87–117
2 Kuwano K, Aruga Y, Saga N. Cryopreservation and clonal gametophytic thalli of *Porphyra* (Rhodophyta). Plant Sci, 1996, 116: 117–124
3 Sahoo D, Tang X R, Yarish C. *Porphyra*—the economic seaweed as a new experimental system. Curr Sci India, 2002, 83: 1313–1316
4 Le Gall Y, Brown S, Marie D, et al. Quantification of nuclear DNA and GC content in marine macroalgae by flow cytometry of isolated nuclei. Protoplasma, 1993, 173: 123–132
5 Davison I R, Pearson G A. Stress tolerance in intertidal seaweeds. J Phycol, 1996, 32: 197–211
6 Guo S, Zheng Y, Joung J G, et al. Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types. BMC Genomics, 2010, 11: 384
7 Nikaido I, Asamizu E, Nakajima M, et al. Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, *Porphyra yezoensis*. DNA Res, 2000, 7: 223–227
8 Asamizu E, Nakajima M, Kitade Y, et al. Comparison of RNA expression profiles between the two generations of *Porphyra yezoensis* (Rhodophyta), based on expressed sequence tag frequency analysis. J Phycol, 2003, 39: 923–930
9 Xu M J, Mao Y X, Zhang X C, et al. Bioinformatic analysis of expressed sequence tags from sporophyte of *Porphyra yezoensis* (Bagiaceae, Rhodophyta). Prog Nat Sci, 2006, 16: 39–49
10 Morozova O, Hirst M, Marra M A. Applications of new sequencing technologies for transcriptome analysis. Annu Rev Genom Hum G, 2009, 10: 135–151
11 Mu Y, Ding F, Cui P, et al. Transcriptome and expression profiling analysis revealed changes of multiple signaling pathways involved in the large yellow croaker during *Aeromonas hydrophila* infection. BMC Genomics, 2010, 11: 506
12 Hegedűs Z, Zakrzewska A, Agoston V C, et al. Deep sequencing of the zebrafish transcriptome response to mycobacterium infection. Mol Immunol, 2009, 46: 2918–2930
13 Wang B, Guo G, Wang C, et al. Survey of the transcriptome of *Aspergillus oryzae* via massively parallel mRNA sequencing. Nucleic Acids Res, 2010, 15: 5075–5087
14 Zhang G, Guo G, Hu X, et al. Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. Genome Res, 2010, 20: 646–654
15 Rosenkranz R, Borodina T, Lehrach H, et al. Characterizing the mouse ES cell transcriptome with Illumina sequencing. Genomics, 2008, 92: 187–194
16 Montgomery S B, Sammeth M, Gutierrez-Arcelus M, et al. Transcriptome genetics using second generation sequencing in a Caucasian population. Nature, 2010, 464: 773–777
17 Morozova O, Marra M A. Applications of next-generation sequencing technologies in functional genomics. Genomics, 2008, 92: 255–264
18 Li R, Zhu H, Ruan J, et al. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res, 2010, 20: 265–272
19 Conesa A, Götz S, García-Gómez J M, et al. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics, 2005, 21: 3674–3676
20 Ye J, Fang L, Zheng H, et al. WEGO: A web tool for plotting GO annotations. Nucleic Acids Res, 2006, 34: W293–297
21 Iseli C, Jongeneel C V, Bucher P. ESTScan: A program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proc Int Conf Intell Syst Mol Biol, 1999, 7: 138–148
22 Altschul S F, Madden T L, Schäffer A A, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res, 1997, 25: 3389–3402
23 Thiel T, Michalek W, Varshney R K, et al. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet, 2003, 106: 411–422
24 Cardle L, Ramsay L, Milbourne D, et al. Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics, 2000, 156: 847–854
25 Wang C B, Guo W Z, Cai C P, et al. Characterization, development and exploitation of EST-derived microsatellites in *Gossypium raimondii* Ulbrich. Chinese Sci Bull, 2006, 51: 557–561
26 Teich R, Zauner S, Baurain D, et al. Origin and distribution of Calvin cycle fructose and sedoheptulose bisphosphatases in plantae and complex algae: A single secondary origin of complex red plastids and subsequent propagation via tertiary endosymbioses. Protist, 2006, 158: 263–276
27 Kitade Y, Asamizu E, Fukuda S, et al. Identification of genes preferentially expressed during asexual sporulation in *Porphyra yezoensis* gametophytes (Bangiales, Rhodophyta). J Phycol, 2008, 44: 113–123
28 Derelle E, Ferraz C, Rombauts S, et al. Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. Proc Natl Acad Sci USA, 2006, 103: 11647–11652
29 Fan X L, Fang Y J, Hu S N, et al. Generation and analysis of 5,318 expressed sequence tags from the filamentous sporophyte of *Porphyra haitanensis* (Phodophyta). J Phycol, 2007, 43: 1287–1294
30 Blouin N A, Brodie J A, Grossman A C, et al. *Porphyra*: A marine crop shaped by stress. Trends Plant Sci, 2011, 16: 29–37
31 Gao K S, Ji Y, Aruga Y, et al. Relationship of CO_2 concentrations to photosynthesis of intertidal macroalgae during emersion. Hydrobiologia, 1999, 398/399: 355–359
32 Herbert S K. Photoinhibition resistance in the red alga *Porphyra perforata*: The role of photoinhibition repair. Plant Physiol, 1990, 92: 514–519
33 Collén J, Davison I R. Stress tolerance and reactive oxygen metabolism in the intertidal red seaweeds *Mastocarpus stellatus* and *Chondrus crispus*. Plant Cell Environ, 1999, 22: 1143–1151
34 Wolfe-Simon F, Grzebyk D, Schofield O, et al. The role and evolution of superoxide dismutases in algae. J Phycol, 2005, 41: 453–465
35 Korbee N, Huovinen P, Figueroa F L, et al. Availability of ammonium influences photosynthesis and the accumulation of mycosporine-like amino acids in two *Porphyra* species (Bangiales, Rhodophyta). Mar Biol, 2005, 146: 645–654
36 Singh S P, Klisch M, Sinha R P, et al. Genome mining of mycosporine-like amino acid (MAA) synthesizing and non-synthesizing cyanobacteria: A bioinformatics study. Genomics, 2010, 95: 120–128
37 Rozema J, Björn L O, Bornman J F, et al. The role of UV-B radiation in aquatic and terrestrial ecosystems–an experimental and functional analysis of the evolution of UV-absorbing compounds. J Photochem Photobiol B, 2002, 66: 2–12
38 Cock1 J M, Sterck L, Rouze P, et al. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. Nature, 2010, 465: 617–621
39 Dittami S M, Michel G, Collén J, et al. Chlorophyll-binding proteins revisited—amultigenic family of light-harvesting and stress proteins from a brown algal perspective. BMC Evol Biol, 2010, 10: 365
40 Llorca O, Martín-Benito J, Ritco-Vonsovici M, et al. Eukaryotic

chaperonin CCT stabilizes actin and tubulin folding intermediates in open quasi-native conformations. EMBO J, 2000, 19: 5971–5979
41 Oliver M J, Dowd S E, Zaragoza J, et al. The rehydration transcriptome of the desiccation-tolerant bryophyte *Tortula ruralis*: Transcript classification and analysis. BMC Genomics, 2004, 5: 89
42 Bartels D. Desiccation tolerance studied in the resurrection plant *Craterostigma plantagineum*. Integr Comp Biol, 2005, 45: 696–701
43 Bennetzen J L. Transposable element contributions to plant gene and genome evolution. Plant Mol Biol, 2000, 42: 251–269
44 Nigumann P, Redik K, Mätlik K, et al. Many human genes are transcribed from the antisense promoter of L1 retrotransposon. Genomics, 2002, 79: 628–634
45 Miyao A, Tanaka K, Murata K, et al. Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon rich regions of the genome. Plant Cell, 2003, 15: 1771–1780
46 Le Q, Melayah D, Bonnivard E, et al. Distribution dynamics of the Tnt1 retrotransposon in tobacco. Mol Genet Genomics, 2007, 278: 639–651
47 Sin H S, Huh J W, Kim D S, et al. Transcriptional control of the HERV-HLTR element of the *GSDML* gene in human tissues and cancer cells. Arch Virol, 2006, 151: 1985–1994
48 Faulkner G J, Kimura Y, Daub C O, et al. The regulated retrotransposon transcriptome of mammalian cells. Nat Genet, 2009, 41: 563–571
49 Gupta P K, Rustgi S. Molecular markers from the transcribed/expressed region of the genome in higher plants. Funct Integr Genomic, 2004, 4: 139–162

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.