Acta Oceanol. Sin., 2018, Vol. 37, No. 4, P. 89–101
DOI: 10.1007/s13131-018-1178-4
http://www.hyxb.org.cn
E-mail: hyxbe@263.net

Phylogenetic analyses of the genes involved in carotenoid biosynthesis in algae

WANG Shanshan[1], ZHANG Lei[1], CHI Shan[1], WANG Guoliang[2], WANG Xumin[2], LIU Tao[1]*, TANG Xuexi[1]

[1] College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China

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

Received 17 January 2017; accepted 2 June 2017

Abstract

Carotenoids play a crucial role in absorbing light energy for photosynthesis, as well as in protecting chlorophyll from photodamage. In contrast to the Streptophyta, few studies have examined carotenoid biosynthetic pathways in algae, owing to a shortage of datasets. As part of the 1000 Plants Project, we sequenced and assembled the transcriptomes of 41 marine macroalgal species, including 22 rhodophytes and 19 phaeophytes, and then combined the datasets with publicly available data from GenBank (National Center for Biotechnology Information) and the U.S. Department of Energy Joint Genome Institute. As a result, we identified 68 and 79 full-length homologs in the Rhodophyta and Phaeophyceae, respectively, of seven inferred carotenoid biosynthetic genes, including the genes for phytoene synthase (PSY), phytoene desaturase (PDS), ζ-carotene desaturase (ZDS), ζ-carotene isomerase (Z-ISO), prolycopene isomerase (crtISO), lycopene β-cyclase (LCYB), and lycopene ε-cyclase (LCYE). We found that the evolutionary history of the algal carotenoid biosynthetic pathway was more complex than that of the same pathway in the Streptophyta and, more specifically, that the evolutionary history involved endosymbiotic gene transfer, gene duplication, and gene loss. Almost all of the eukaryotic algae that we examined had inherited the seven carotenoid biosynthesis genes *via* endosymbiotic gene transfer. Moreover, *PSY*, *crtISO*, and the ancestral lycopene cyclase gene (*LCY*) underwent duplication events that resulted in multiple gene copies, and the duplication and subsequent divergence of *LCYB* and *LCYE* specialized and complicated the cyclization of lycopene. Our findings also verify that the loss of *LCYE* in both the microphytic rhodophytes and phaeophytes explains the differences in their carotenoid patterns, when compared to the macrophytic rhodophytes. These analyses provide a molecular basis for further biochemical and physiological validation in additional algal species and should help elucidate the origin and evolution of carotenoid biosynthetic pathways.

Key words: carotenoid biosynthesis, algae, phylogenetic analysis

Citation: Wang Shanshan, Zhang Lei, Chi Shan, Wang Guoliang, Wang Xumin, Liu Tao, Tang Xuexi. 2018. Phylogenetic analyses of the genes involved in carotenoid biosynthesis in algae. Acta Oceanologica Sinica, 37(4): 89–101, doi: 10.1007/s13131-018-1178-4

1 Introduction

Carotenoids are a large family of lipophilic isoprenoid compounds that are synthesized by all photoautotrophs, as well as some nonphotosynthetic organisms (Walter and Strack, 2011; Vílchez et al., 2011). Approximately 750 kinds of natural carotenoids have been reported to date, and these compounds play a variety of roles in light harvesting and photoprotection and as stress hormones and signaling apocarotenoids (Lichtenthaler, 1987; Nisar et al., 2015; Gruszecki and Strzałka, 2005). Despite their great diversity, most of the known carotenoids are tetraterpenoids, which are composed of eight isoprenoid units (Takaichi, 2011), and the initial steps of carotenoids biosynthesis are common among algae and land plants.

To form a 40-carbon backbone, 15-cis-phytoene is the first step in the carotenoid biosynthetic pathway (Cunningham and Gantt, 1998; Sandmann, 2002). This step is finished by phytoene synthase (PSY, EC 2.5.1.32), which catalyzes the head-to-head condensation of two geranylgeranyl diphosphate molecules in a two-step reaction, with pre-phytoene diphosphate as a reaction intermediate (Dogbo et al., 1988). Then, the colorless 15-cis-phytoene is desaturated by plant-type phytoene desaturase phytoene desaturase (PDS, EC 1.3.5.5) and ζ-carotene desaturase (ZDS, EC 1.3.5.6; Sandmann, 2009; Klassen, 2010; Matthews et al., 2003; Breitenbach and Sandmann, 2005) and isomerized by ζ-carotene isomerase (Z-ISO, EC 5.2.1.12) and prolycopene isomerase (crtISO, EC 5.2.1.13; Chen et al., 2010; Chai et al., 2011; Masamoto et al., 2001; Isaacson et al., 2004; Li et al., 2007; Bartnikas et al., 1997; Yu et al., 2011) to form pink or red all-trans-lycopene. Nevertheless, Z-ISO and crtISO may be dispensable since their functions can be partially compensated by photoisomerization (Giuliano et al., 2002; Park et al., 2002; Frigaard et al., 2004). Next, cyclization at the termini of the all-trans-lycopene results in the first branching point of the carotenoid biosynthetic pathway, which either (1) introduces a β-ring to both sides of lycopene, thereby forming β-carotene, or (2) introduces an ε-ring to one side and β-ring to the other side, thereby form-

Foundation item: The Leading Talents Program in Taishan Industry of Shandong Province under contract No. LJNY2015010; the China Agriculture Research System under contract No. CARS-50; the Regional Demonstration Project of Marine Economic Innovation and Development under contract No. 12PYY001SF08-ZGHYDX-2; the China-ASEAN Maritime Cooperation Fund.
*Corresponding author, E-mail: liutao@ouc.edu.cn

ing α-carotene. The β- and ε-rings are formed by lycopene β-cyclase (LCYB) and lycopene ε-cyclase (LCYE), respectively (Cunningham et al., 1996; Cui et al., 2011; Krubasik and Sandmann, 2000). Subsequently, different α- and β-carotene derivatives are generated *via* specialized enzymatic reactions (Ladygin, 2000; Chen et al., 2007; Ruiz-Sola and Rodríguez-Concepción, 2012).

The genetic, biochemical, and molecular mechanisms of this process have been described in great detail for the Streptophyta (land plants) and various microorganisms (Ruiz-Sola and Rodríguez-Concepción, 2012; Sandmann, 1994). Nonetheless, relatively few studies have examined the process of carotenoid biosynthesis in algae, and most of the existing work has focused on microalgae, especially model organisms. Lohr et al. (2005), for example, identified most of the carotenoid biosynthetic genes of *Chlamydomonas reinhardtii* and conducted phylogenic analysis in order to characterize the carotenoid biosynthetic genes of the Chlorophyta, and Cunningham et al. (2007) identified 11 carotenoid biosynthetic genes in the unicellular rhodophyte *Cyanidioschyzon merolae* and confirmed the function of lycopene β-cyclase and β-carotene hydroxylase in order to elucidate the evolutionary history of the species' carotenoid biosynthetic pathway. Meanwhile, Li et al. (2016) reported the carotenoid biosynthetic genes from *Rhodomonas* sp. Nonetheless, a full-scale phylogenetic analysis of algae has yet to be completed, since the genomes and transcriptomes and some algae, such as the macrophytic rhodophytes and the phaeophytes, have not been sequenced.

In the present study, we sequenced and assembled the transcriptomes of 22 rhodophytes and 19 phaeophytes part of the 1 000 Plants (1KP) Project (http://www.onekp.com). Then, by combining our datasets with public data from GenBank (National Center for Biotechnology Information) and the U.S. Department of Energy Joint Genome Institute (JGI), we identified seven dominant candidate genes that are involved in the early reactions of the carotenoid biosynthesis pathway (up to carotene). Our results provide unequivocal molecular evidence that most of the carotenoid biosynthetic genes are actively transcribed in algae, especially in marine rhodophytes and phaeophytes. Next, we conducted comprehensive phylogenetic analyses using publicly available sequences from cyanobacteria, plants, and other algae. These analyses should help elucidate the origin and evolution of carotenoid biosynthetic pathways.

2 Materials and methods

2.1 *Sample collection and RNA extraction*

Marine rhodophyte and phaeophyte samples were collected from the coast of China from October 2010 to March 2012 (Table S1). No specific permission was required for these locations, and the study did not involve any endangered or protected species. Upon collection, the fresh samples were rinsed briefly with sterilized seawater and, then, either stored at -80°C or promptly subjected to RNA extraction.

After the algae samples were immersed in liquid nitrogen and ground into a fine powder using a chilled mortar and pestle, total RNA was extracted using either an improved CTAB method or an improved TRIzol method for the phaeophyte and rhodophyte samples, respectively (Li et al., 2012; Johnson et al., 2012). The quality and quantity of the extracted RNA were assessed using a Nanodrop ND 1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and an Agilent 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

2.2 *Transcriptome sequencing and de novo assembly*

cDNA library construction and sequencing were performed by the BGI (Shenzhen, China), using Illumina HiSeq instruments (San Diego, CA, USA). Strict read filtering was performed before the assembly. Paired-end reads with primer or adaptor sequences were removed, and reads with more than 10% of bases with quality scores of below Q20 or with more than 5% of unknown bases (Ns) were excluded. *De novo* assembly was performed using SOAPdenovo-Trans (Li et al., 2008a; Li et al., 2010), and Gapcloser was used fill in gaps in the scaffolds. All the transcriptomes were sequenced within the framework of the 1KP Project.

2.3 *Functional annotation and identification of carotenoid biosynthetic genes*

To reconstruct the algal carotenoid biosynthetic pathway, all assembled algal sequences were assigned to the Kyoto Encyclopedia of Genes and Genomes (KEGG) Automatic Annotation Server (KAAS) for functional annotation (Moriya et al., 2007). Briefly, we individually submitted the algal sequences to the server in FASTA format; selected the representative set for eukaryotes, as recommended by KAAS, and all algae and plants as the reference data; and selected the bi-directional best hit (BBH) information method for our analysis.

According to the results of the KEGG pathway analysis, seven dominant carotenoid biosynthetic genes (*PSY*, *PDS*, *Z-ISO*, *ZDS*, *crtISO*, *LCYB*, and *LCYE*) were analyzed further. Related nucleotide sequences were downloaded from the GenBank and JGI databases and used to search for our target sequences by means of local BLASTn, with an E-value less than 10^{-5}. Matching sequences were manually checked for accuracy using known cDNA sequences and MEGA 5 (Tamura et al., 2011). The resulting output was filtered to exclude exact duplicates, and each sequence was analyzed independently. Afterward, the online BLASTX tool from NCBI was used to examine the homology of cDNA open reading frame sequences from the transcriptome data. All full-length sequences were uploaded to the GenBank database (Table S2).

2.4 *Sequence alignments and phylogenetic analyses*

Selected full-length sequences were aligned using ClustalX 2.1 (Thompson et al., 1997; Sievers et al., 2011), and Bayesian analysis was performed using MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). The analyses were conducted as two independent runs, each of which involved four incrementally heated Metropolis-coupled Monte-Carlo Markov Chains running for 5 000 000 generations, and trees were generated every 100 generations. The first 25% of the trees were discarded during the burn-in phase, and the remaining trees were used to build a 50% majority rule consensus tree, accompanied with posterior probability values. The average standard deviation of split frequencies at the end of the run was below 0.01, indicating stationary conditions. Analysis parameters were as follows: printfreq=1 000, samplefreq=100, nchain=4, and temp=0.2. Tree visualization was implemented in FigTree v. 1.3.1 (http://tree.bio.ed.ac.uk/software/ figtree/).

3 Results

3.1 *Sequencing and unigene annotation*

A total of 503 310 608 raw reads (89.2 Gb), with an average length of 180 bp, were generated from the transcriptomes of 22 rhodophytes and 19 phaeophytes, and the reads were assembled into 2 161 986 scaffolds, with an average length of 717 bp and an

N50 of 1 751 bp. All the sequences were aligned against the local nr protein database, which was downloaded from NCBI, and when the E-value cutoff was set to 10^{-5}, a total of 585 247 unigenes produced significant BLAST matches.

3.2 *Carotenoid biosynthetic pathway analysis*

All 41 algal transcriptome datasets were subject to KEGG pathway analysis. The numbers of partial key putative genes in the carotenoid biosynthetic pathways of marine rhodophytes and phaeophytes are listed in Table 1.

According to the results of the KEGG pathway analysis, seven dominant carotenoid biosynthetic genes (*PSY*, *PDS*, *Z-ISO*, *ZDS*, *crtISO*, *LCYB*, and *LCYE*) were selected, and their sequences were BLASTed against the algal transcriptomes. As a result of BLAST analysis, we identified 68 and 79 full-length homologs in the 22 rhodophyte and 19 phaeophyte transcriptomes, respectively, and we found that *PSY*, *PDS*, *ZDS*, *crtISO*, and *LCYB* were expressed in all 41 rhodophyte and phaeophyte species, which suggests that they are indispensable in the production of carotenes.

Notably, *Z-ISO* was absent in the rhodophyte transcriptomes, whereas *LCYE* was absent in the phaeophyte transcriptomes, and we confirmed this trend by examining published rhodophyte and ochrophyte genomes, respectively (Table S3). Therefore, it is possible that the *Z-ISO* gene was lost during the evolutionary history of the Rhodophyta. However, the absence of *Z-ISO* did not lead to the interruption of carotenoid synthesis, which indicated that the

Table 1. Partial key putative genes in the carotenoid biosynthetic pathway of marine rhodophytes and phaeophytes

Taxonomy	Organism	*PSY*	*PDS*	*Z-ISO*	*ZDS*	*crtISO*	*LCYB*	*LCYE*	*ZEP*	*VDE*
Rhodophyta	*Pyropia yezoensis*	1	2	0	1	1	1	0	0	0
Rhodophyta	*Ceramium kondoi*	1	1	0	1	1	1	0	0	0
Rhodophyta	*Heterosiphonia pulchra*	1	1	0	1	1	1	1	0	0
Rhodophyta	*Symphyocladia latiuscul*	1	1	0	1	1	1	0	1	0
Rhodophyta	*Neosiphonia japonica*	1	1	0	1	1	1	0	0	0
Rhodophyta	*Dumontia simplex*	1	1	0	1	1	1	0	0	0
Rhodophyta	*Gloiopeltis furcata*	1	1	0	1	1	1	1	0	0
Rhodophyta	*Mazzaella japonica*	1	1	0	1	1	1	0	0	0
Rhodophyta	*Chondrus crispus*	1	1	0	1	1	1	0	1	0
Rhodophyta	*Ahnfeltiopsis flabelliformis*	1	1	0	1	1	1	1	0	0
Rhodophyta	*Eucheuma denticulatum*	1	1	0	1	1	1	1	0	0
Rhodophyta	*Betaphycus gelatinus*	1	2	0	1	1	1	1	0	0
Rhodophyta	*Kappaphycus alvarezii*	3	1	0	2	2	3	0	0	0
Rhodophyta	*Gracilaria vermiculophylla*	1	1	0	1	1	1	0	1	0
Rhodophyta	*Gracilaria chouae*	2	1	0	1	2	1	0	0	0
Rhodophyta	*Gracilaria blodgettii*	1	1	0	1	1	1	0	1	0
Rhodophyta	*Gracilariopsis lemaneiformis*	1	1	0	1	1	1	0	0	0
Rhodophyta	*Grateloupia livida*	1	1	0	1	1	1	1	0	0
Rhodophyta	*Grateloupia turuturu*	1	1	0	1	1	1	1	0	0
Rhodophyta	*Grateloupia catenata*	1	1	0	1	1	1	1	0	0
Rhodophyta	*Grateloupia chiangii*	1	1	0	1	1	1	0	0	0
Rhodophyta	*Grateloupia filicina*	2	1	0	1	1	1	1	1	1
Phaeophyceae	*Desmarestia viridis*	1	1	1	1	2	1	0	1	1
Phaeophyceae	*Dictyopteris undulata*	1	1	2	1	3	1	0	1	1
Phaeophyceae	*Ishige okamurai*	1	1	1	1	1	1	0	1	0
Phaeophyceae	*Saccharina japonica*	1	2	1	2	2	1	0	2	1
Phaeophyceae	*Saccharina sculpera*	1	1	1	1	3	1	0	1	1
Phaeophyceae	*Undaria pinnatifida*	1	1	1	2	3	1	0	1	2
Phaeophyceae	*Punctaria latifolia*	1	1	2	1	3	1	0	1	2
Phaeophyceae	*Colpomenia sinuosa*	1	1	1	1	3	1	0	0	0
Phaeophyceae	*Petalonia fascia*	1	1	2	1	3	1	0	1	1
Phaeophyceae	*Scytosiphon lomentaria*	1	1	1	1	3	1	0	1	1
Phaeophyceae	*Scytosiphon dotyi*	1	1	1	1	2	1	0	1	1
Phaeophyceae	*Sargassum fusiforme*	2	1	1	1	2	1	0	1	1
Phaeophyceae	*Sargassum hemiphyllum* var. *chinense*	2	1	2	1	1	1	0	1	1
Phaeophyceae	*Sargassum henslowianum*	1	1	1	1	2	1	0	2	1
Phaeophyceae	*Sargassum horneri*	1	2	3	1	1	1	0	1	1
Phaeophyceae	*Sargassum integerrimum*	1	1	1	1	2	1	0	1	1
Phaeophyceae	*Sargassum muticum*	1	1	2	1	2	1	0	1	1
Phaeophyceae	*Sargassum thunbergii*	1	2	3	1	2	1	0	1	1
Phaeophyceae	*Sargassum vachellianum*	2	1	1	1	2	1	0	1	1

Note: *PSY*, *PDS*, *Z-ISO*, *ZDS*, *crtISO*, *LCYB*, *LCYE*, *ZEP*, *VDE* represent phytoene synthase gene, phytoene desaturase gene, ζ-carotene isomerase gene, ζ-carotene desaturase gene, prolycopene isomerase gene, lycopene beta-cyclase gene, lycopene epsilon-cyclase gene, zeaxanthin epoxidase gene, and violaxanthin de-epoxidase gene, respectively.

relevant isomerization may be compensated by other mechanisms, such as those involving light, as in cyanobacteria and plants (Chen et al., 2010; Masamoto et al., 2001). On the other hand, the phaeophytes were deficient in α-carotenoids, owing to the absence of *LCYE*, which explains why both α-carotenoids and β-carotenoids are found in macrophytic rhodophytes but only β-carotenoids are found in phaeophytes.

3.3 *Duplication of the phytoene synthase gene (PSY) in algae*

Homology analysis identified 13 and 11 putative *PSY* sequences in the rhodophyte and phaeophyte 1KP transcriptome data, respectively, and the full-length *PSY* cDNAs of the rhodophytes and phaeophytes ranged from 1 197 to 1 515 bp and shared 40.68%–52.23% identity.

When combined with the *PSY* sequences of cyanobacteria, other algae, and plants, the translated *PSY* amino acid sequences were used to construct a Bayesian phylogenetic tree (Fig. 1). The *PSY* sequences from the cyanobacteria and Glaucophyta were the first to diverge at the base of the tree, followed by two large and strongly supported (PP=1) clades that included other eukaryotic algal *PSY*s. This topology suggests that algae inherited *PSY* from cyanobacteria *via* endosymbiotic gene transfer (EGT).

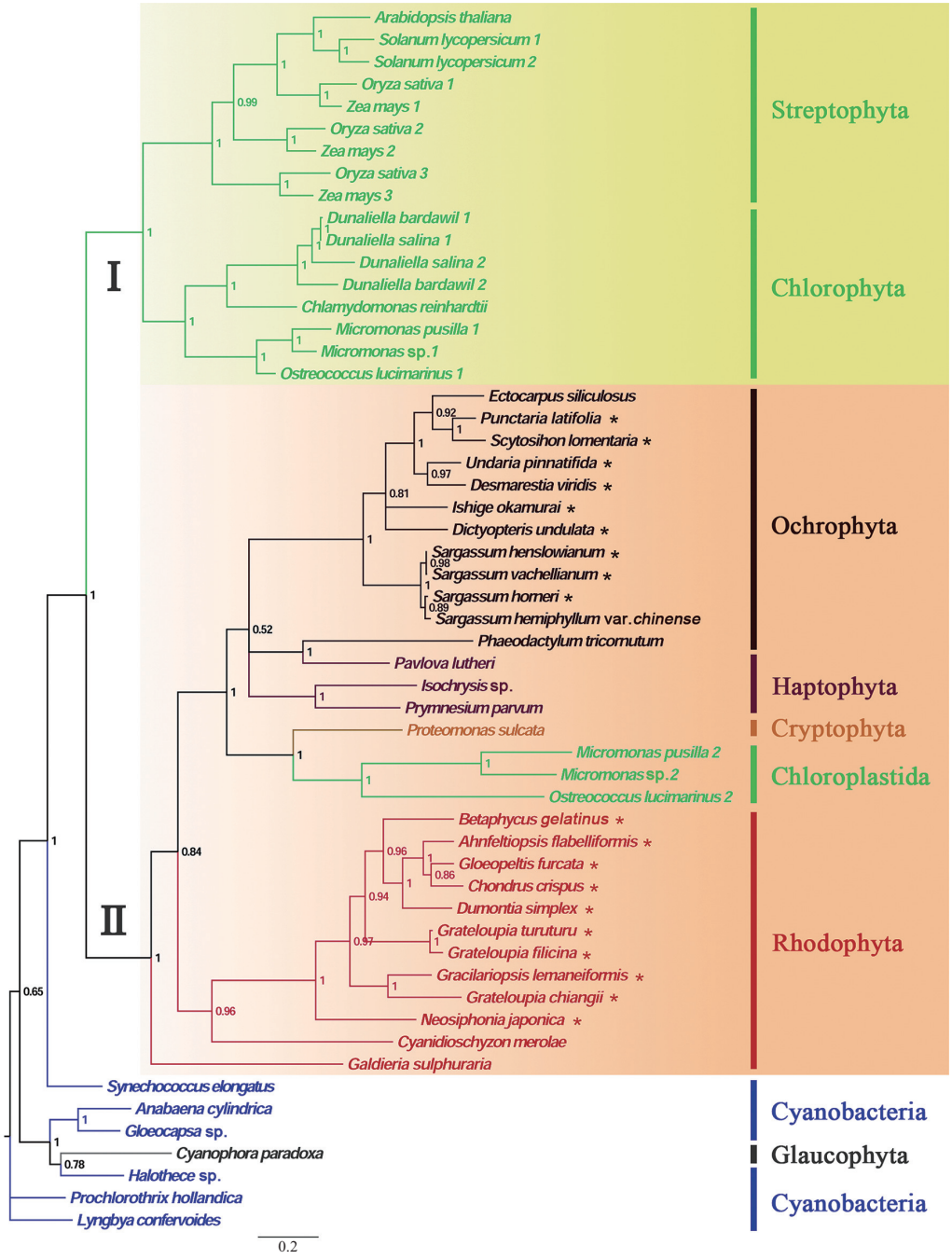

Fig. 1. Bayesian phylogenetic tree of phytoene synthase genes (*PSY*s). Posterior probabilities (PP) of >50% are indicated at the nodes. Cyanobacterial *PSY* is included as the outgroup. Asterisks indicate putative *PSY*s from the rhodophyte and phaeophyte species in the 1KP Project. Roman numerals (I and II) denote two classes of *PSY*.

Furthermore, the tree contained two separate gene duplication events. In aglae (expected Glaucophyta), an ancient gene duplication gave rise to two distinct *PSY* classes, followed by a gene loss, and both *PSY* classes were only retained by the Prasinophyceae. In contrast, the other chlorophytes retained only Class I *PSY*, and both the rhodophytes and secondary symbiotic algae (Ochrophyta, Haptophyta, and Cryptophyta) retained only Class II *PSY*. Then, a subsequent gene duplication event generated multiple paralogs in various chlorophytes. Although different organisms possess different *PSY* classes, all *PSY* versions share a similar substrate-Mg^{2+}-binding site and catalytic residues, and the major differences in the genes occur in regions that are unessential for the enzymatic function (Fig. 2).

3.4 *Phylogeny of common-origin algal desaturase genes: PDS and ZDS*

Twenty-eight putative *PDS* genes and 28 putative *ZDS* genes were identified in the 22 rhodophyte and 19 phaeophyte transcriptomes. The shared identity (28.70%–36.56%) and similar N- and C-terminal regions of the putative *PDS* and *ZDS* genes suggests that the two phylogenetically related (Fig. 3). The Bayesian phylogenetic tree of the *PDS* and *ZDS* sequences (Fig. S1) in-

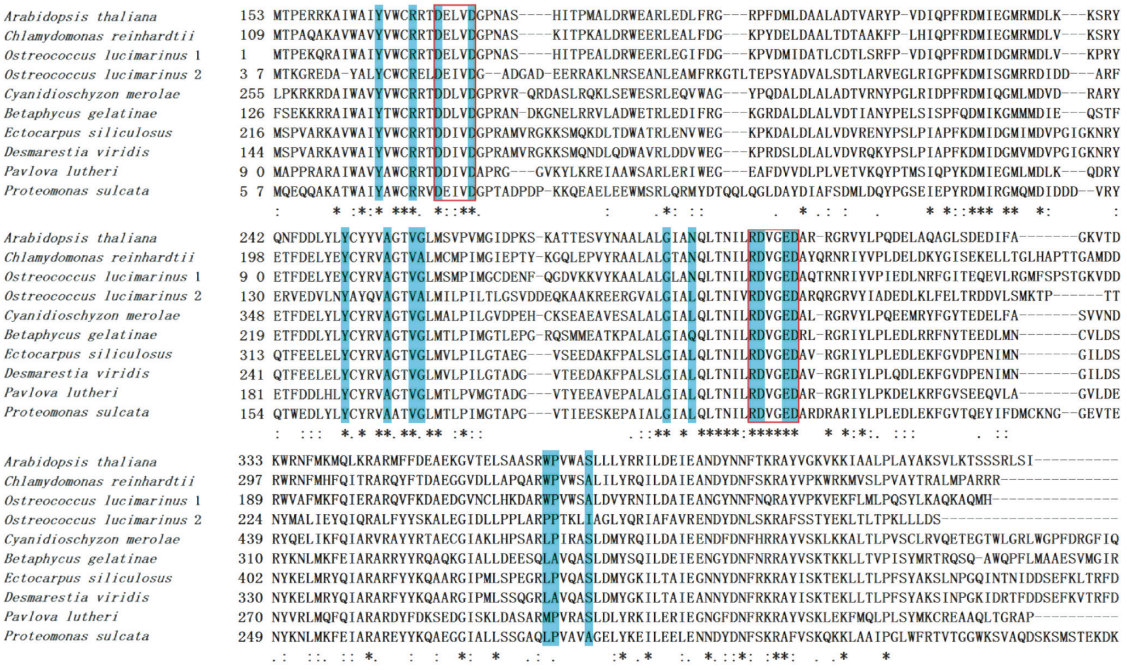

Fig. 2. Amino acid sequences of phytoene synthase (PSY) proteins from the Rhodophyta, Ochrophyta, Haptophyta, Cryptophyta, Chlorophyta, and Streptophyta. The alignment indicates partail sequences of PSY Classes I and II. The predicted PSY sequences were aligned using ClustalX 2.1. Numbers indicate the position of the first residue in each aligned sequence. Red rectangles indicate the substrate-Mg^{2+}-binding sites, and the substrate-binding pocket and catalytic residues are highlighted in blue.

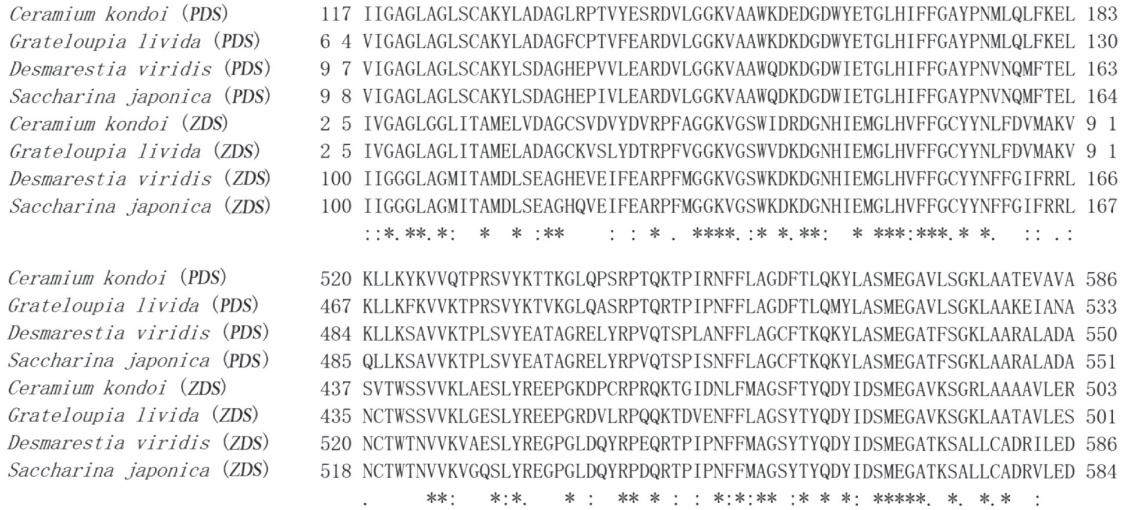

Fig. 3. N- and C-terminal regions of the phytoene and ζ-carotene desaturase genes (*PDS* and *ZDS*) from 22 rhodophytes algae and 19 phaeophytes. The predicted *PDS* and *ZDS* amino acid sequences were aligned using ClustalX 2.1. Numbers indicate the position of the first and last residue in each aligned sequence.

cludes two separate clusters, which prompted us to construct the Bayesian phylogenetic trees of *PDS* and *ZDS* separately.

In the Bayesian phylogenetic tree of *PDS* (Fig. 4), the sequences were grouped into four main clades: (1) cyanobacteria; (2) Glaucophyta; (3) Rhodophyta; and (4) Ochrophyta, Haptophyta, Cryptophyta, Chlorophyta, and Streptophyta. The topological structure indicated that the primary endosymbiotic algae (Glaucophyta, Rhodophyta, and Chlorophyta) inherited *PDS* from cyanobacteria during primary endosymbiosis *via* EGT. Nevertheless, the *PDS* sequences from the Ochrophyta, Haptophyta, and Cryptophyta are clearly distinct from the sequences from Rhodophyta and are strongly supported (PP=1) as a sister taxon of Chlorophyta. These data suggest that secondary symbiotic algae may have shared a common *PDS* origin (from Chlorophyta) before their divergence.

In the Bayesian phylogenetic tree of *ZDS* (Fig. 5), cyanobacterial sequences clustered at the base, whereas the photosynthetic eukaryotic sequences clustered into two large and well-separated clades. Accordingly, the topology supports the hypothesis that algal *ZDS* has a cyanobacterial origin. In contrast to *PDS* (Fig. 4), the *ZDS* from the secondary symbiotic algae is more closely related to the rhodophyte gene than to the chlorophyte gene, which suggests that *ZDS* from the secondary symbiotic algae was acquired from a rhodophyte endosymbiont during the secondary endosymbiosis *via* EGT.

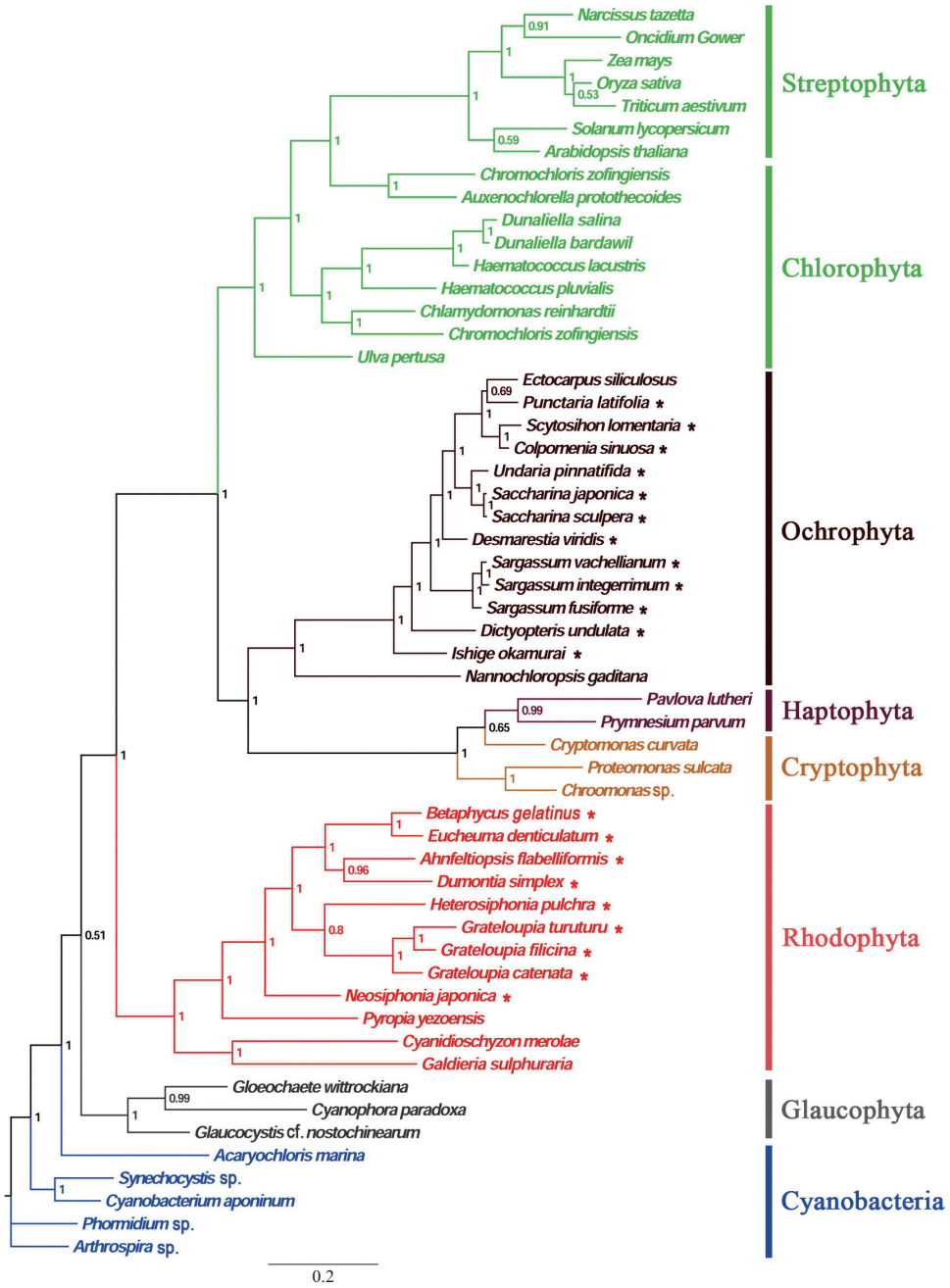

Fig. 4. Bayesian phylogenetic tree of phytoene desaturase genes (*PDS*s). Posterior probabilities of >50% are indicated at the nodes. Cyanobacterial *PDS* is included as the outgroup. Asterisks indicate putative *PDS*s from the rhodophyte and phaeophyte species in the 1KP Project.

3.5 *Phylogeny of different-origin algal isomerase genes: Z-ISO and crtISO*

No rhodophyte *Z-ISO* genes were detected in the 1KP transcriptomic data or published databases. Nonetheless, 11 full-length putative rhodophyte *Z-ISO*s were obtained from the 1KP Project, and each sequence had an unabridged C-terminal region, which is responsible for the ζ-carotene isomerase activity. The number of *crtISO* copies among the marine rhodophytes and phaeophytes was different, with only a single copy of *crtISO* obtained from the rhodophytes and three obtained from the phaeophytes (Table S2).

Unlike the algal *PDS* and *ZDS* genes, *Z-ISO* is not a paralog of

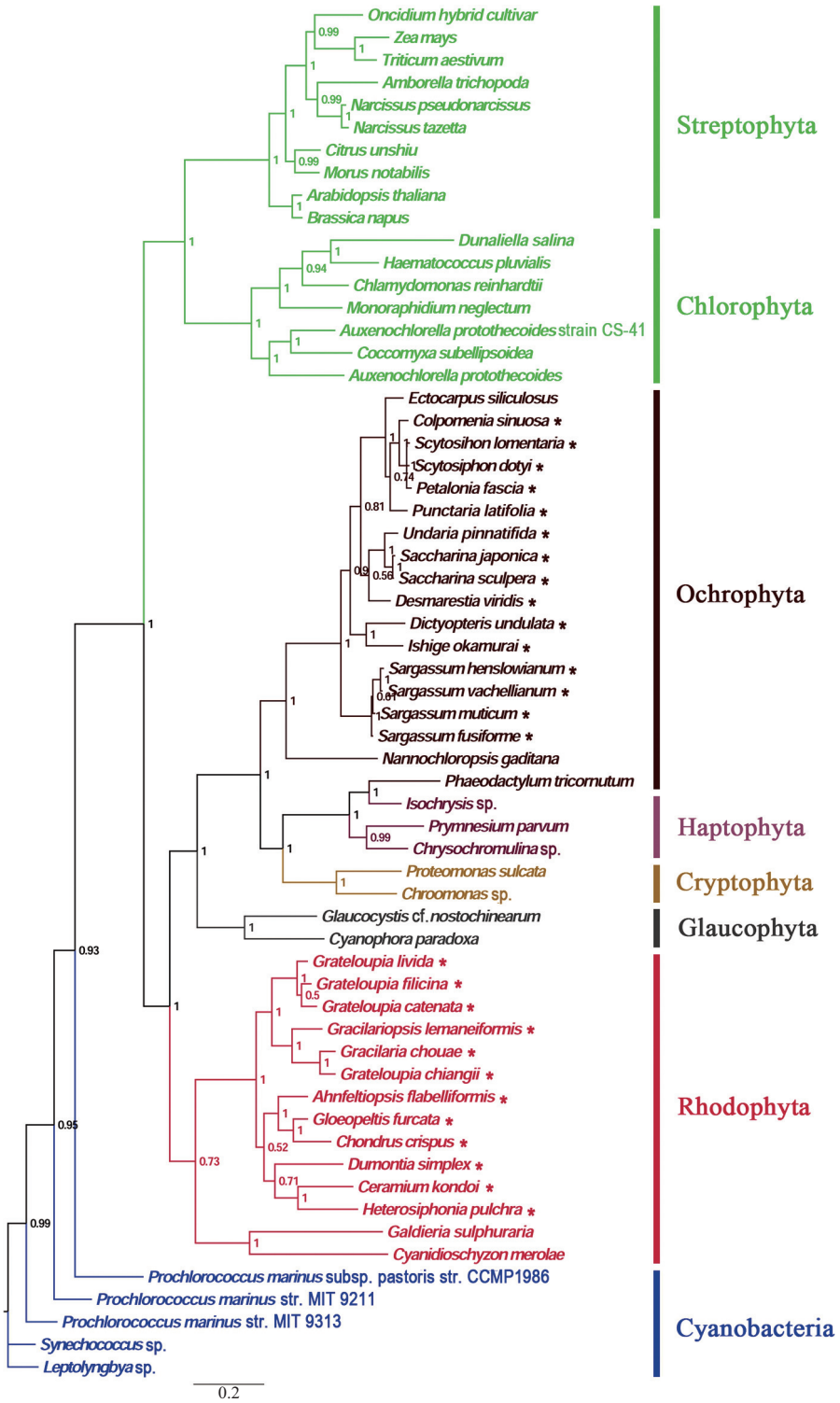

Fig. 5. Bayesian phylogenetic tree of ζ-carotene desaturase genes (*ZDS*s). Posterior probabilities of >50% are indicated at the nodes. Cyanobacterial *ZDS* is included as the outgroup. Asterisks indicate putative *ZDS*s from the rhodophyte and phaeophyte species in the 1KP Project.

crtISO; and *Z-ISO* shares 21.46%–25.11% identity with the nitrite and nitric oxide reductase U gene (*NnrU*) from denitrifying bacteria, whereas *crtISO* only shares 16.73%–23.71% identity with the bacterial phytoene desaturase gene (*crtI*). Furthermore, Z-ISO mediates isomerization *via* electron transfer activity, which may be inherent in NnrU (Chen et al., 2010). On the other hand, the isomerization executed by crtISO is a reversible desaturation reaction that is followed by reoxidation, thereby yielding a trans-configuration of the newly formed double bond (Giuliano et al., 2002). Together, this suggests that *Z-ISO* has a *NnrU* progenitor and that *crtISO* is an evolutionary descendant of *crtI*.

The topology of the *Z-ISO* phylogenetic tree was relatively simple, when compared to the topologies of the other trees (Fig. 6), and indicated that *Z-ISO* originated from cyanobacteria. Although the grouping of sequences from the Ochrophyta, Cryptophyta, Haptophyta, and Chlorophyta was strongly supported (PP=1), the loss of the rhodophyte *Z-ISO* makes the origin of the secondary symbiotic algae uncertain.

The phylogenetic tree of *crtISO* also validated the cyanobacterial provenance of the gene in algae and plants (Fig. 7). Nonetheless, the analysis of *crtISO* is a little complicated, owing to its duplication in the Rhodophyta and Ochrophyta and its loss in the Rhodophyta. The Glaucophyta, Rhodophyta, and Chlorophyta obtained this gene from cyanobacteria. A prior gene duplication produced two types of *crtISO* genes in the Rhodophyta, followed by a gene loss in the macrophytic rhodophytes. The Ochrophyta, Cryptophyta, and Haptophyta inherited one type of *crtISO* from the Rhodophyta, and subsequent duplication of the gene occurred at least three times in the Ochrophyta.

3.6 *Independent duplications of the lycopene cyclase gene (LCY) in algae*

We found only 11 putative rhodophyte *LCYB*s in the 1KP transcriptomic data. Nevertheless, 14 putative *LCYB*s and eight putative *LCYE*s were identified in the rhodophyte species. It is exciting that this is the first detailed report of *LCYE* in the Rhodophyta. Two lycopene cyclase genes from the marine rhodophytes were highly conserved, with shared identities of 39.47%–44.04%, and all the putative rhodophyte *LCYB* and *LCYE* sequences contained three conserved motifs that were similar to those found in the *LCY* from Chlorophyta, Streptophyta, and cyanobacteria.

The main difference among these sequences is the three amino acid residues inserted into the putative rhodophyte *LCYB* (Fig. 8). The nucleotide sequences of the rhodophyte *LCYE* sequences were used for further searches in the genomes of *Cyanidioschyzon merolae* and *Galdieria sulphuraria*, but nothing was found. Therefore, we assumed that only macrophytic rhodophytes possess both *LCYB* and *LCYE*, whereas microphytic rhodophytes possess only *LCYB*. This may explain why both α- and β-carotenoids are present in macrophytic rhodophytes, whereas only β-carotenoids are present in microphytic rhodophytes.

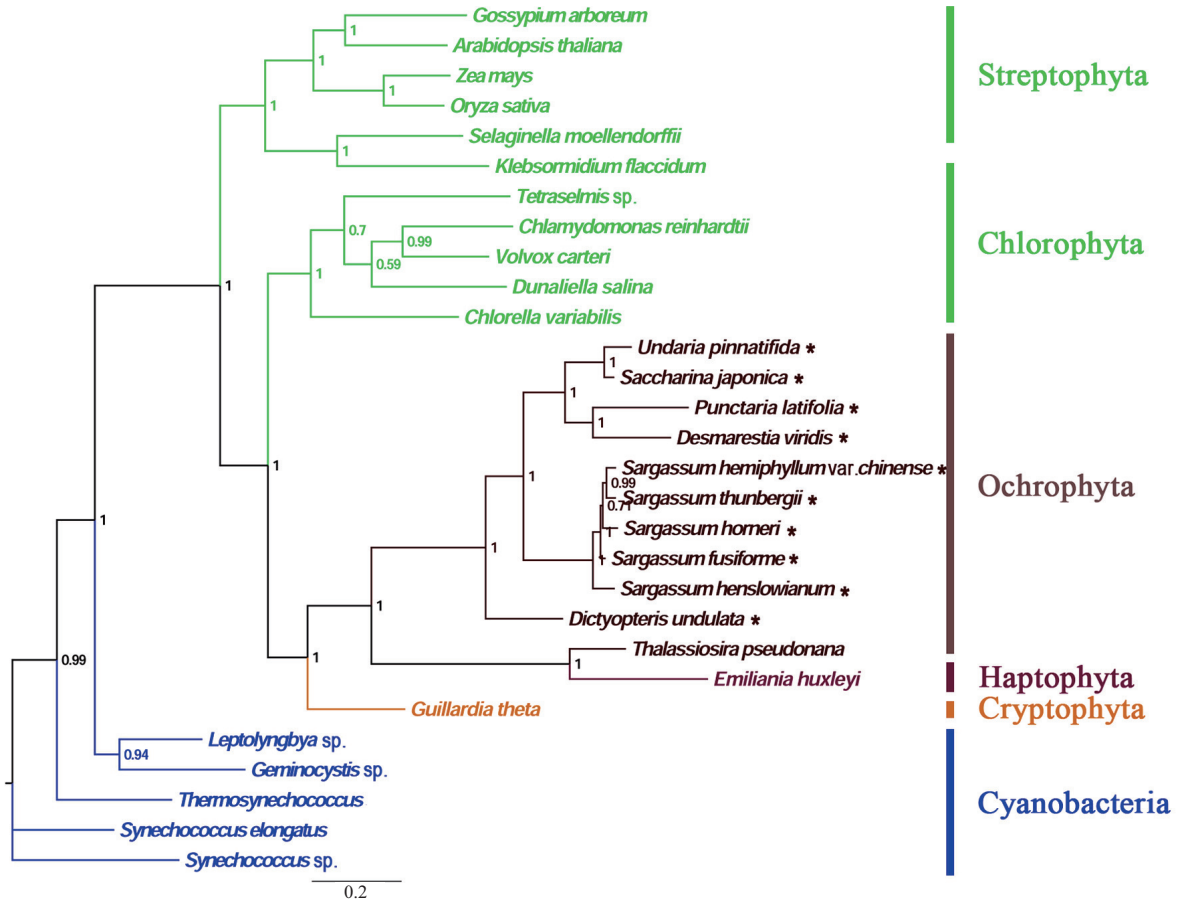

Fig. 6. Bayesian phylogenetic tree of ζ-carotene isomerase genes (*Z-ISO*s). Posterior probabilities of >50% are indicated at the nodes. Cyanobacterial *Z-ISO* is included as the outgroup. Asterisks indicate putative *PDS*s from the rhodophyte and phaeophyte species in the 1KP Project.

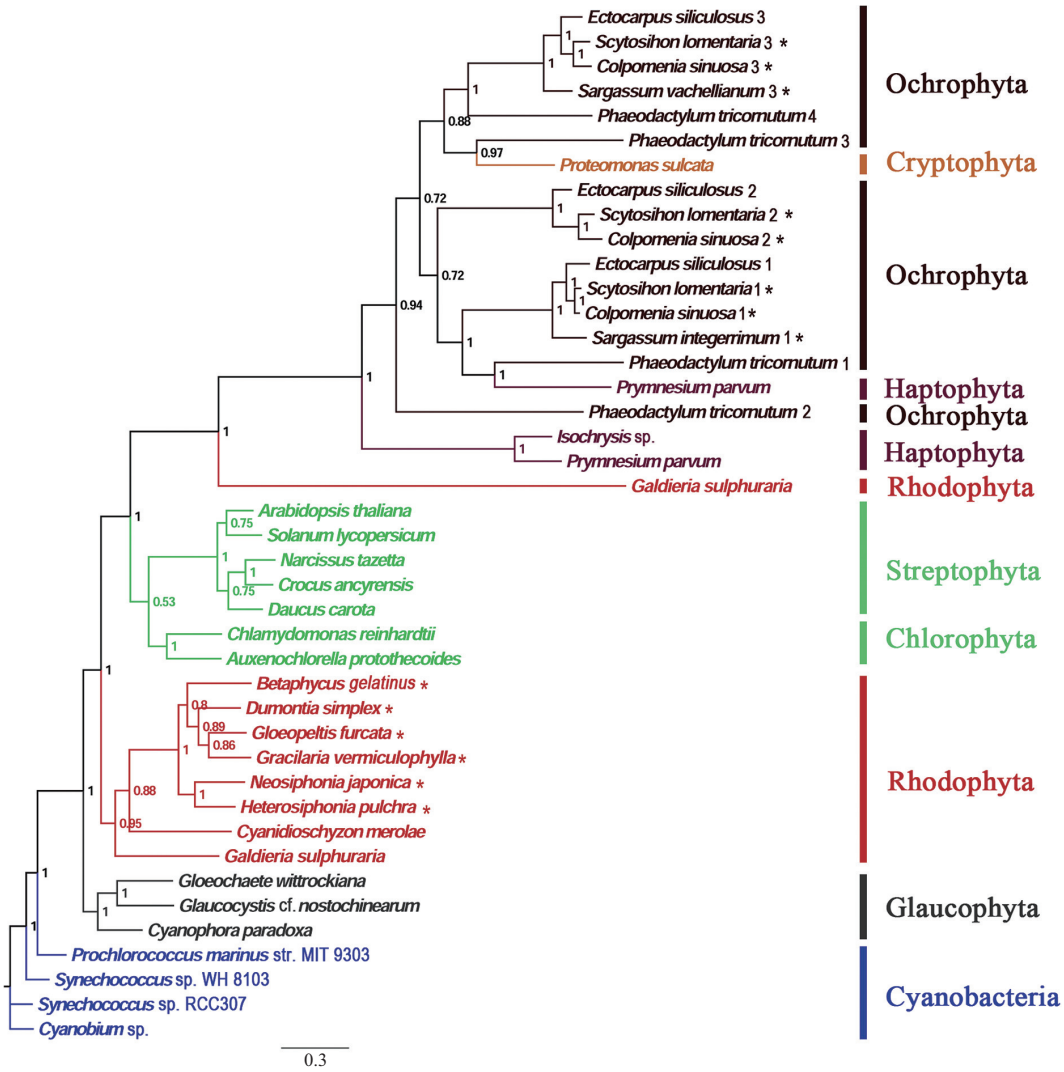

Fig. 7. Bayesian phylogenetic tree of prolycopene isomerase genes (*crtISO*s). Posterior probabilities of >50% are indicated at the nodes. Cyanobacterial *crtISO* is included as the outgroup. Asterisks indicate putative *crtISO*s from the rhodophyte and phaeophyte species in the 1KP Project.

As expected, the topological structure of the lycopene cyclase phylogenetic tree (Fig. 9) indicated that *LCY*s of primary endosymbiotic algae also have a cyanobacterial origin and that the secondary endosymbiotic algae acquired *LCYB* from a rhodophyte-like secondary endosymbiont *via* EGT. In addition, independent gene duplications and subsequent divergence generated *LCYB* and *LCYE* after the differentiation of each algal lineage. Moreover, the gene loss of *LCYE* did occur in the microphytic rhodophytes or in the Ochrophyta, according to the results of multiple alignments and phylogenetic analyses.

4 Discussion

4.1 *Cyanobacterial origin of carotenoid biosynthetic genes in diverse algae*

During primary endosymbiosis, cyanobacteria were engulfed by a eukaryotic host and slowly evolved into plastids (McFadden, 2001a), and subsequently, during the secondary endosymbiosis event, eukaryotic hosts engulfed and retained a red or green alga, which ultimately resulted in the modern diversity of algae (Cavalier-Smith, 1999; Douzery et al., 2004; Bhattacharya and Medlin, 1998). Accordingly, cyanobacteria are regarded as the progenitor of the chloroplast, and since then, EGT has been a ubiquitous and continuous process in modern algae and plants, providing opportunities for gene transfer to algal genomes and the replacement of pre-existing and functionally equivalent host genes. This phenomenon seems to be a prominent mechanism governing the evolution of nucleus-encoded plastid-targeted proteins, such as carotenoid biosynthetic enzymes (Timmis et al., 2004; Martin et al., 2002; Ni et al., 2012; Keeling and Palmer, 2008; Martin and Herrmann, 1998).

The present study is in agreement with other studies and demonstrates that all seven dominant genes involved in the early reactions of carotenoid biosynthesis have cyanobacterial origins (Sandmann, 2002; Cui et al., 2011; Bhattacharya and Medlin, 1998; Martin et al., 2002). The Glaucophyta, Rhodophyta, and Chlorophyta obtained carotenoid biosynthesis genes from cyanobacteria *via* primary endosymbiosis-mediated EGT, whereas the phylogenetic trees of *PSY*, *ZDS*, *crtISO*, and *LCY*, indicate that the Ochrophyta, Haptophyta, and Cryptophyta acquired their genes from a rhodophyte-like organism *via* secondary endosymbiosis-mediated EGT. This transfer breaks down the interspecies

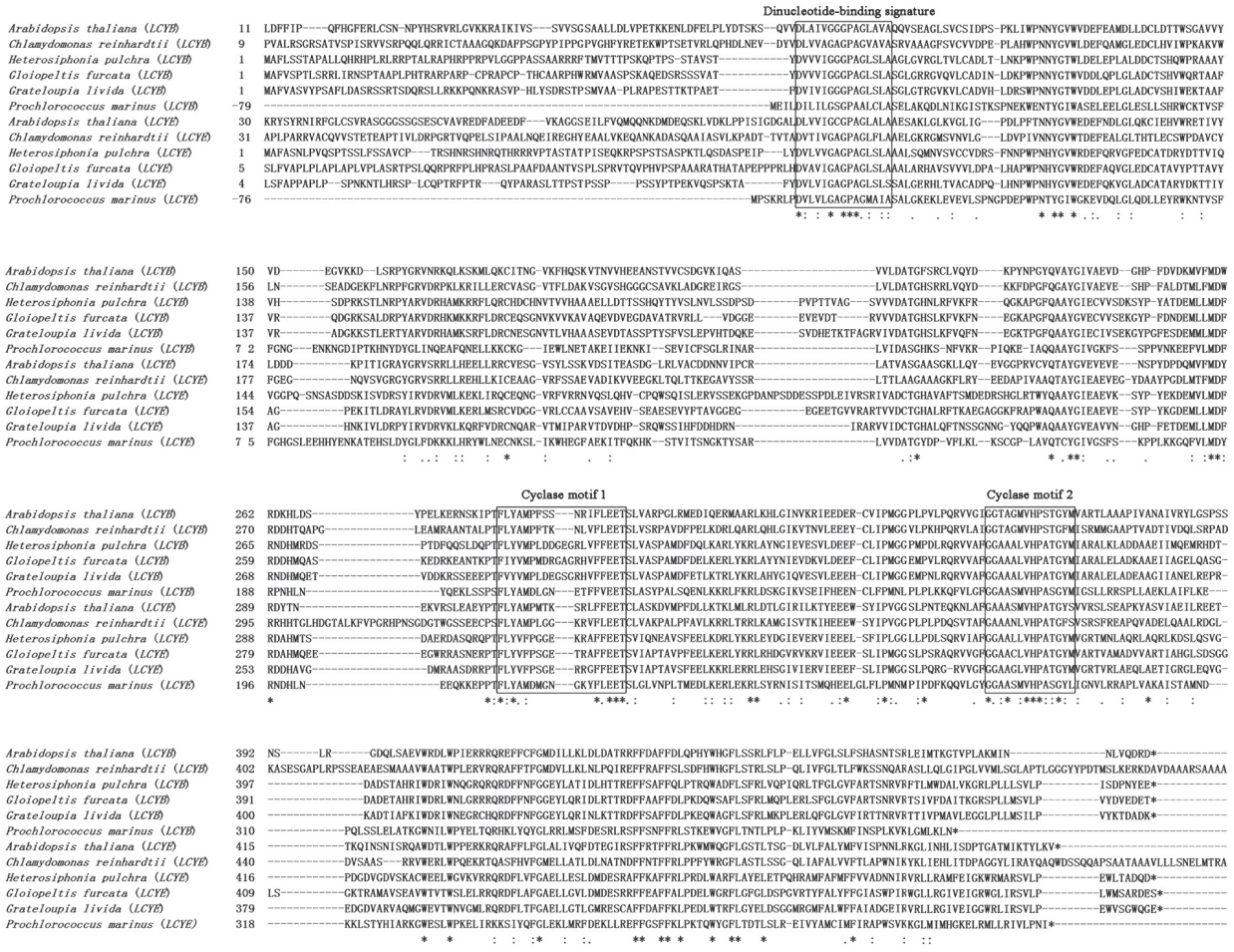

Fig. 8. Amino acid sequences of lycopene cyclase (LCY) proteins from cyanobacteria and the Rhodophyta, Chlorophyta, and Streptophyta. The predicted *LCYB* and *LCYE* sequences were aligned using ClustalX 2.1. Numbers indicate the position of the first residue in each aligned sequence. Black rectangles indicate the three conserved motifs.

barrier between algal lineages and has allowed algae to survive. In the present study, we found that EGT has occurred frequently among algal carotenoid biosynthetic genes and that the mechanism seems to have played an important role in the evolution of eukaryotic algae.

4.2 *Duplication of carotenoid biosynthetic genes*

In line with previous studies, we found that gene duplication has played an important role in the evolution of the carotenoid biosynthetic pathway (Klassen, 2010; Cui et al., 2011; Hittinger and Carroll, 2007; Tran et al., 2009). In fact, duplication events have occurred constantly among different algal lineages (Figs 1, 7 and 8), and as reported by Tran et al. (2009), the duplication of the phytoene synthase gene is a typical example. An ancient gene duplication also created the two *PSY* classes, and more recent gene duplication events have mainly occurred in the Chlorophyta, thereby generating various paralogs. Another similar example is the duplication of the *crtISO* gene. A prior gene duplication produced two kinds of *crtISO* in the Rhodophyta, whereas a subsequent duplication event produced additional paralogs in the Ochrophyta.

In contrast, the duplication of *LCY* seems more significant. An independent duplication event followed by functional divergence generated *LCYB* and *LCYE* in both cyanobacteria and algae. As a result, algae can synthesize lycopene with both β- and ε-rings, rather than with only a single ring type (Cunningham et al., 1996; Krubasik and Sandmann, 2000; Cunningham et al., 1994; Stickforth et al., 2003). Because gene duplication is necessary for genetic novelty and for environmental adaptation (Hittinger and Carroll, 2007; Tran et al., 2009), the multiple copies of carotenoid biosynthetic genes make the carotenoid biosynthetic pathway specialized and complicated, as well as flexible and adaptable. This state of affairs can also result in differential regulation in response to developmental or environmental cues (Sandmann, 2009; Krubasik and Sandmann, 2000; Li et al., 2008b).

4.3 *Loss of carotenoid biosynthetic genes*

There is no doubt that algae gained more cyanobacterial carotenoid biosynthetic genes *via* EGT and gene duplication (Ni et al., 2012; Tran et al., 2009). Nevertheless, the loss of specific genes is also an important mechanism in the evolution of the carotenoid biosynthetic pathway. The role of gene loss in the evolution of *PSY* is relatively complicated, in that the Chlorophyta (except the Prasinophyceae) and Streptophyta only retained Class I *PSY*, whereas the Rhodophyta and secondary endosymbiotic algae only retained Class II *PSY*. In contrast, the loss of the rhodophyte *Z-ISO* is extraordinarily complete, and the continued synthesis of ζ-carotene suggests that rhodophytes might use light to compensate for the isomerization, as in cyanobacteria and plants (Takaichi, 2011; Breitenbach and Sand-

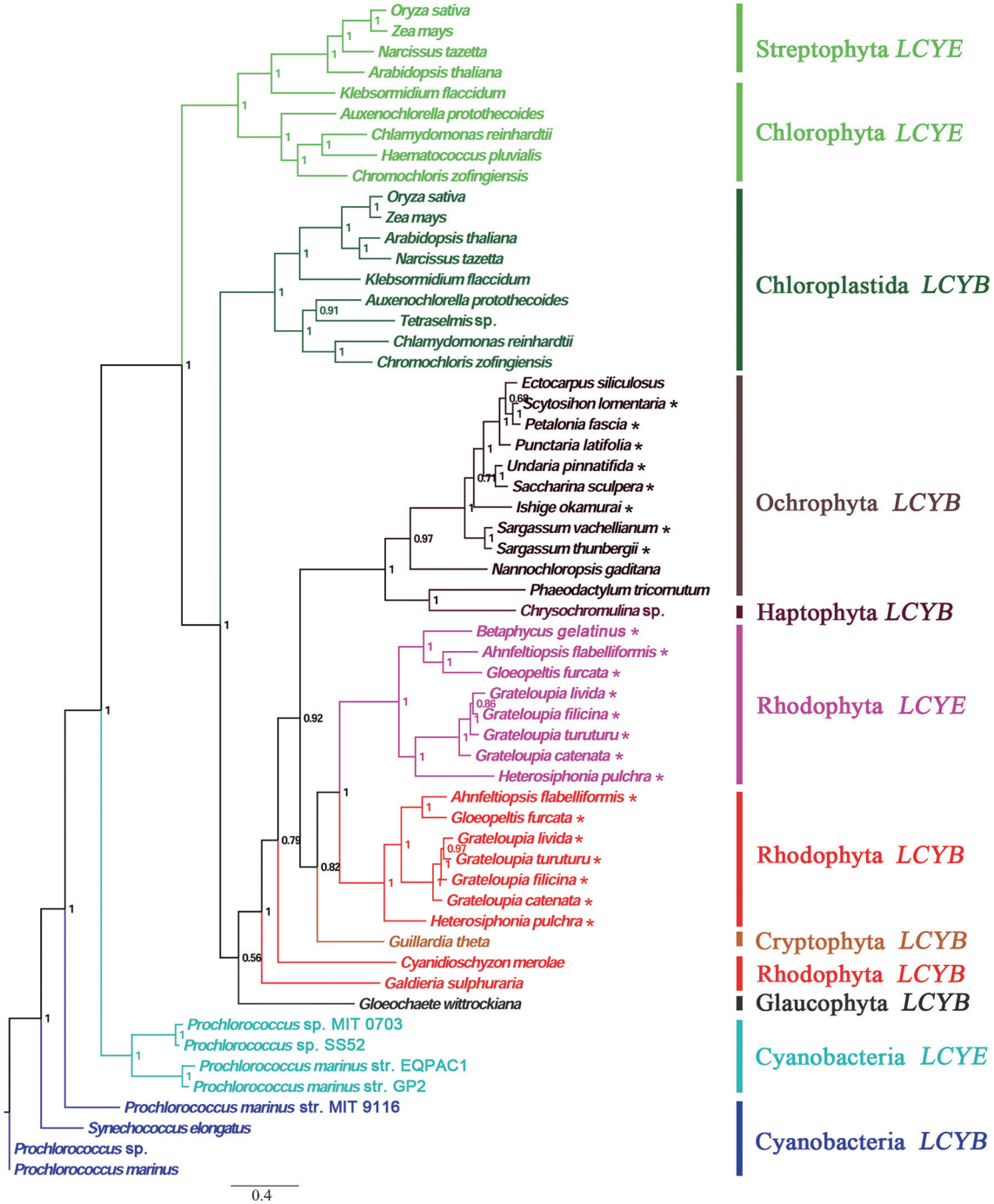

Fig. 9. Bayesian phylogenetic tree of lycopene cyclase genes (*LCY*s). Posterior probabilities of >50% are indicated at the nodes. Cyanobacterial *LCY* is included as the outgroup. Asterisks indicate putative *LCY*s from the rhodophyte and phaeophyte species in the 1KP Project.

mann, 2005; Chen et al., 2010; Masamoto et al., 2001; Takaichi et al., 2016). Meanwhile, both α-carotene and its derivatives are absent in the Cyanidiophyceae, Ochrophyta, Haptophyta, and Cryptophyta, which suggests that the loss of *LCYE* prevents the formation of ε-rings. Taken together, these data indicate that differential gene loss has caused an unbalanced distribution of genes among the algal lineages. It is also interesting that every duplication event was associated with subsequent gene loss. This phenomenon may result from evolutionary adaptation to various environments (Bhattacharya and Medlin, 1998; Lund et al., 2008; McFadden, 2001b; Millen et al., 2001). In summary, the present study provides a comprehensive analysis of carotenoid biosynthetic genes in algae and reveals that EGT, gene duplication, and gene loss have all contributed to the successful evolution of the carotenoid biosynthetic pathway. These findings provide a molecular basis for further biochemical and physiological validation in additional algal species and should help elucidate the origin and evolution of the carotenoid biosynthetic pathway. In addition, the present study analyzes the phylogenetics of algal *Z-ISO* and *crtISO* genes. The study also raises some

new questions, such as the function of multiple gene copies and the mechanism of gene loss. As additional algal omics data are published, more algal carotenoid biosynthetic genes will be explored and will help researchers answer interesting questions in this field.

Acknowledgements

The authors are grateful to Michael Melkonian from Universität zu Köln for providing transcriptomic data for Glaucophyta, Cryptophyta, and Haptophyta.

References

Bartnikas T B, Tosques I E, Laratta W P, et al. 1997. Characterization of the nitric oxide reductase-encoding region in Rhodobacter sphaeroides 2.4.3. J Bacteriol, 179(11): 3534–3540

Bhattacharya D, Medlin L. 1998. Algal phylogeny and the origin of land plants. Plant Physiol, 116(1): 9–15

Breitenbach J, Sandmann G. 2005. ζ-Carotene *cis* isomers as products and substrates in the plant poly-*cis* carotenoid biosynthetic pathway to lycopene. Planta, 220(5): 785–793

Cavalier-Smith T. 1999. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. J Eukaryot Microbiol, 46(4): 347–366

Chai Chenglin, Fang Jun, Liu Yang, et al. 2011. *ZEBRA2*, encoding a carotenoid isomerase, is involved in photoprotection in rice. Plant Mol Biol, 75(3): 211–221

Chen Qian, Jiang Jianguo, Wang Fei. 2007. Molecular phylogenies and evolution of *crt* genes in algae. Crit Rev Biotechnol, 27(2): 77–91

Chen Yu, Li Faqiang, Wurtzel E T. 2010. Isolation and characterization of the *Z-ISO* gene encoding a missing component of carotenoid biosynthesis in plants. Plant Physiol, 153(1): 66–79

Cui Hongli, Wang Yinchu, Qin Song. 2011. Molecular evolution of lycopene cyclases involved in the formation of carotenoids in eukaryotic algae. Plant Mol Biol Rep, 29(4): 1013–1020

Cunningham F X, Gantt E. 1998. Genes and enzymes of carotenoid biosynthesis in plants. Annu Rev Plant Physiol Plant Mol Biol, 49: 557–583

Cunningham F X Jr, Lee H, Gantt E. 2007. Carotenoid biosynthesis in the primitive red alga *Cyanidioschyzon merolae*. Eukaryot Cell, 6(3): 533–545

Cunningham F X, Pogson B, Sun Zairen, et al. 1996. Functional analysis of the β and ε lycopene cyclase enzymes of Arabidopsis reveals a mechanism for control of cyclic carotenoid formation. Plant Cell, 8(9): 1613–1626

Cunningham F X, Sun Zairen, Chamovitz D, et al. 1994. Molecular structure and enzymatic function of lycopene cyclase from the cyanobacterium Synechococcus sp strain PCC7942. Plant Cell, 6(8): 1107–1121

Dogbo O, Laferriére A, D'Harlingue A, et al. 1988. Carotenoid biosynthesis: Isolation and characterization of a bifunctional enzyme catalyzing the synthesis of phytoene. Proc Natl Acad Sci USA, 85(19): 7054–7058

Douzery E J P, Snell E A, Bapteste E, et al. 2004. The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils?. Proc Natl Acad Sci USA, 101(43): 15386–15391

Frigaard N U, Maresca J A, Yunker C E, et al. 2004. Genetic manipulation of carotenoid biosynthesis in the green sulfur bacterium *Chlorobium tepidum*. J Bacteriol, 186(16): 5210–5220

Giuliano G, Giliberto L, Rosati C. 2002. Carotenoid isomerase: a tale of light and isomers. Trends Plant Sci, 7(10): 427–429

Gruszecki W I, Strzałka K. 2005. Carotenoids as modulators of lipid membrane physical properties. Biochim Biophys Acta, 1740(2): 108–115

Hittinger C T, Carroll S B. 2007. Gene duplication and the adaptive evolution of a classic genetic switch. Nature, 449(7163): 677–681

Isaacson T, Ohad I, Beyer P, et al. 2004. Analysis in vitro of the enzyme CRTISO establishes a poly-cis-carotenoid biosynthesis pathway in plants. Plant Physiol, 136(4): 4246–4255

Johnson M T J, Carpenter E J, Tian Zhijian, et al. 2012. Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. PLoS One, 7(11): e50226

Keeling P J, Palmer J D. 2008. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet, 9(8): 605–618

Klassen J L. 2010. Phylogenetic and evolutionary patterns in microbial carotenoid biosynthesis are revealed by comparative genomics. PLoS One, 5(6): e11257

Krubasik P, Sandmann G. 2000. Molecular evolution of lycopene cyclases involved in the formation of carotenoids with ionone end groups. Biochem Soc Trans, 28(6): 806–810

Ladygin V G. 2000. Biosynthesis of carotenoids in the chloroplasts of algae and higher plants. Russ J Plant Physl, 47(6): 796–814

Li Ruiqiang, Li Yingrui, Kristiansen K, et al. 2008a. SOAP: short oligonucleotide alignment program. Bioinformatics, 24(5): 713–714

Li Faqiang, Murillo C, Wurtzel E T. 2007. Maize *Y9* encodes a product essential for 15-cis-ζ-carotene isomerization. Plant Physiol, 144(2): 1181–1189

Li Tianyong, Ren Lei, Zhou Guan, et al. 2012. A suitable method for extracting total RNA from red algae. Transactions of Oceanology and Limnology (in Chinese), (4): 64–71

Li Faqiang, Vallabhaneni R, Wurtzel E T. 2008b. *PSY3*, a new member of the phytoene synthase gene family conserved in the Poaceae and regulator of abiotic stress-induced root carotenogenesis. Plant Physiol, 146(3): 1333–1345

Li Huanqin, Wang Wenlei, Wang Zhaokai, et al. 2016. De novo transcriptome analysis of carotenoid and polyunsaturated fatty acid metabolism in *Rhodomonas* sp. J Appl Phycol, 28(3): 1649–1656

Li Ruiqiang, Zhu Hongmei, Ruan Jue, et al. 2010. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res, 20(2): 265–272

Lichtenthaler H K. 1987. Chlorophylls and carotenoids: pigments of photosynthetic biomembranes. Method Enzymol, 148: 350–382

Lohr M, Im C S, Grossman A R. 2005. Genome-based examination of chlorophyll and carotenoid biosynthesis in *Chlamydomonas reinhardtii*. Plant Physiol, 138(1): 490–515

Lund A, Andersson P, Eriksson J, et al. 2008. Automatic fitting procedures for EPR s pectra of disordered systems: matrix diagonalization and perturbation methods applied to fluorocarbon radicals. Spectrochim Acta Part A, 69(5): 1294–300

Martin W, Herrmann R G. 1998. Gene transfer from organelles to the nucleus: how much, what happens, and why?. Plant Physiol, 118(1): 9–17

Martin W, Rujan T, Richly E, et al. 2002. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc Natl Acad Sci USA, 99(19): 12246–12251

Masamoto K, Wada H, Kaneko T, et al. 2001. Identification of a gene required for *cis*-to-*trans* carotene isomerization in carotenogenesis of the cyanobacterium *Synechocystis* sp. PCC 6803. Plant Cell Physiol, 42(12): 1398–1402

Matthews P D, Luo Ruibai, Wurtzel E T. 2003. Maize phytoene desaturase and ζ-carotene desaturase catalyse a poly-=Z desaturation pathway: implications for genetic engineering of carotenoid content among cereal crops. J Exp Bot, 54(391): 2215–2230

McFadden G I. 2001. Primary and secondary endosymbiosis and the origin of plastids. J Phycol, 37(6): 951–959

McFadden G I. 2001. Chloroplast origin and integration. Plant Physiol, 125(1): 50–53

Millen R S, Olmstead R G, Adams K L, et al. 2001. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell, 13(3): 645–658

Moriya Y, Itoh M, Okuda S, et al. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res, 35(S2): W182–W185

Ni Ting, Yue Jipei, Sun Guiling, et al. 2012. Ancient gene transfer from algae to animals: mechanisms and evolutionary significance.

BMC Evol Biol, 12: 83
Nisar N, Li Li, Lu Shan, et al. 2015. Carotenoid metabolism in plants. Mol Plant, 8(1): 68–82
Park H, Kreunen S S, Cuttriss A J, et al. 2002. Identification of the carotenoid isomerase provides insight into carotenoid biosynthesis, prolamellar body formation, and photomorphogenesis. Plant Cell, 14(2): 321–332
Ronquist F, Huelsenbeck J P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, 19(12): 1572–1574
Ruiz-Sola M Á, Rodríguez-Concepción M. 2012. Carotenoid biosynthesis in Arabidopsis: a colorful pathway. Arabidopsis Book, 10: e0158
Sandmann G. 1994. Carotenoid biosynthesis in microorganisms and plants. Eur J Biochem, 223(1): 7–24
Sandmann G. 2002. Molecular evolution of carotenoid biosynthesis from bacteria to plants. Physiol Plant, 116(4): 431–440
Sandmann G. 2009. Evolution of carotene desaturation: the complication of a simple pathway. Arch Biochem Biophys, 483(2): 169–174
Sievers F, Wilm A, Dineen D, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol, 7: 539
Stickforth P, Steiger S, Hess W R, et al. 2003. A novel type of lycopene ε-cyclase in the marine cyanobacterium *Prochlorococcus marinus* MED4. Arch Microbiol, 179(6): 409–415
Takaichi S. 2011. Carotenoids in algae: distributions, biosyntheses and functions. Mar Drugs, 9(6): 1101–1118
Takaichi S, Yokoyama A, Mochimaru M, et al. 2016. Carotenogenesis diversification in phylogenetic lineages of Rhodophyta. J Phycol, 52(3): 329–338
Tamura K, Peterson D, Peterson N, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol, 28(10): 2731–2739
Thompson J D, Gibson T J, Plewniak F, et al. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res, 25(24): 4876–4882
Timmis J N, Ayliffe M A, Huang C Y, et al. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet, 5(2): 123–135
Tran D, Haven J, Qiu Weigang, et al. 2009. An update on carotenoid biosynthesis in algae: phylogenetic evidence for the existence of two classes of phytoene synthase. Planta, 229(3): 723–729
Vílchez C, Forján E, Cuaresma M, et al. 2011. Marine carotenoids: biological functions and commercial applications. Mar Drugs, 9(3): 319–333
Walter M H, Strack D. 2011. Carotenoids and their cleavage products: biosynthesis and functions. Nat Prod Rep, 28(4): 663–692
Yu Qiuju, Ghisla S, Hirschberg J, et al. 2011. Plant carotene cis-trans isomerase CRTISO: a new member of the FADred-dependent flavoproteins catalyzing non-redox reactions. J Biol Chem, 286(10): 8666–8676

Supplementary information:

Fig. S1. Bayesian phylogenetic tree of the phytoene desaturase (*PDS*) and ζ-carotene desaturase (*ZDS*) genes. Posterior probabilities of >50% are indicated at the nodes. Asterisks indicate putative *PDS* and *ZDS* genes identified in the Rhodophyta and Phaeophyceae species included in the 1KP project.

Table S1. Rhodophyta and Phaeophyceae species included in the 1KP project.

Table S2. List of sequences used in the present study. Bold text indicates the sequences used to reconstruct phylogenetic trees.

Table S3. Published Rhodophyta and Phaeophyceae genomes.

The supplementary information is available online at www.hyxb.org.cn/aosen/ch/index.aspx. The supplementary information is published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.