Plant Mol Biol Rep (2012) 30:421–432
DOI 10.1007/s11105-011-0352-z

Genome-Wide Analysis of Biotin Biosynthesis in Eukaryotic Photosynthetic Algae

Hongli Cui · Yipeng Wang · Hongyu Zhang · Yinchu Wang · Song Qin

Published online: 14 September 2011
© Springer-Verlag 2011

Abstract Biotin is a cofactor responsible for carbon dioxide transfer in several carboxylase enzymes, which play a significant role in various metabolic reactions such as fatty acid synthesis, branched chain amino acid catabolism, and gluconeogenesis. Biotin is also involved in citric acid cycle, which is the process of biochemical energy generation during aerobic respiration. Though the function of biotin in the growth of algae has been extensively investigated, little is known about the biosynthetic routes of biotin in the algal kingdom. In the present study, 44 biotin biosynthesis-related genes were identified from 14 eukaryotic photosynthetic algal genomes by BLASTP and TBLASN programs. A comprehensive analysis was performed to characterize distribution, phylogeny, structure domains, and coevolution patterns of those genes. Forty-four biotin biosynthesis-related enzymes (BBREs) were found to be distributed in three groups: 7-keto-8-aminopelargonic acid synthase, diaminopelargonic acid synthase/dethiobiotin synthetase, and biotin synthase. Structure domains were considerably conserved among the subfamilies of BBREs. The intramolecular coevolutionary sites are widely distributed in biotin synthase. The present study provides new insights into the origin and evolution of biotin biosynthetic pathways in eukaryotic photosynthetic algae. Furthermore, the characterization of biotin biosynthesis-related genes from algae will promote the identification and functional studies of BBREs.

H. Cui · Y. Wang · H. Zhang · Y. Wang · S. Qin (✉)
Yantai Institute of Coastal Zone Research,
Chinese Academy of Sciences,
Yantai 264003, People's Republic of China
e-mail: sqin@yic.ac.cn

H. Cui · H. Zhang · Y. Wang
Graduate University of the Chinese Academy of Sciences,
Beijing 100049, People's Republic of China

Keywords Biotin biosynthesis · Molecular evolution · Coevolution · Comparative genomics · Algae

Abbreviations

BBRE	Biotin biosynthesis related enzyme
BIOF	7-Keto-8-aminopelargonic acid synthase
BIOA	Diaminopelargonic acid synthase
BIOD	Dethiobiotin synthase
BIOB	Biotin synthase

Introduction

Algae, which lack roots, stems, leaves, conducting vessels, and complex sex organs (Round 1973), are a highly diverse group of photosynthetic organisms that are ubiquitously distributed on Earth. Algae are critical for terrestrial and atmospheric condition maintenance, and can be broadly divided into unicellular (microalgae) and multicellular (macroalgae) subgroups (Grossman 2005). They are responsible for up to 50% of the planet's atmospheric carbon fixation (Field et al. 1998) and play an important role in major biogeochemical processes, primary productivity, and food webs, especially in oligotrophic waters (Chi et al. 2008). *Chlorophyta*, *Glaucocystophyta*, and *Rhodophyta* are derived from the primary endosymbiotic event (Lewis and McCourt 2004; Moreira et al. 2000), while dominant bloom-forming eukaryotic phytoplankton found in the ocean, such as *Heterokonphyta* and *Haptophyta*, are derived by secondary endosymbiosis (Stiller and Hall 1997; Yoon et al. 2004). It makes algae the best material to study the origin and evolution of physiology and metabolism.

Biotin is also called vitamin H or vitamin B_7 and is a cofactor responsible for carbon dioxide transfer in several carboxylase enzymes, which play significant role in various

metabolic reactions such as fatty acid synthesis, branched chain amino acid catabolism, and gluconeogenesis (De Clercq 1997; Zempleni and Mock 1999). For example, acetyl-CoA carboxylase (ACCase) catalyzes the biotin-dependent carboxylation of acetyl-CoA to produce malonyl-CoA, which is the essential first step of the biosynthesis of long chain fatty acids (Chi et al. 2011; Li et al. 2010; Gu et al. 2011; Stolf-Moreira et al. 2011). Besides, biotin plays a role in the citric acid cycle, which is the process by which biochemical energy is generated during aerobic respiration (De Clercq 1997; Zhang et al. 2010). Biotin is also helpful in maintaining a steady blood sugar level (Zempleni et al. 2009). More recently, evidence emerged that biotin also plays unique roles in cell signaling, epigenetic regulation of genes, and chromatin structure (Demetriou et al. 2010; Zempleni 2005).

The biotin biosynthesis pathway described in bacteria, which is from the precursor pimeloyl-CoA, relies on four enzymes (BIOF, BIOA, BIOD, and BIOB) and is well conserved in plant and other microorganisms (Entcheva et al. 2002; Pinon et al. 2005; Picciocchi et al. 2003; Poirier et al. 2011) (Fig. 1a). Although the enzymes involved in this pathway has been well studied in bacteria, particularly in *E. coli* and *B. subtilis* (Streit and Entcheva 2003), it is interesting to note that the source of pimeloyl-CoA remains enigmatic in most organisms (Webb et al. 2007). Moreover, compared to bacteria, little was known about biotin synthesis in eukaryotic photosynthetic algae. Recently, genome sequences of a number of different algae became available from the Department of Energy (DOE) Joint Genome Institute (http://www.jgi.doe.gov). The complete genome sequences allowed us to obtain a comprehensive dataset of genes encoding enzymes and provided a new and comprehensive insight into the biotin biosynthetic pathways in eukaryotic photosynthetic alga. In the present study, genome sequence data of 14 algae species available were used to investigate the biosynthesis of biotin. Emphasis was centered on the distribution, phylogeny, structure domain, and coevolution pattern of BBRE genes in eukaryotic algae.

Materials and Methods

Data Sources

The genomes of 13 eukaryotic photosynthetic algae included *Chlamydomonas reinhardtii*, *Chlorella* sp. NC64A, *Chlorella vulgaris*, *Coccomyxa* sp. C-169, *Volvox cartei*, *Micromonas pusiua*, *Micromonas* sp. RCC299, *Ostreococcus* sp. RCC809, *Ostreococcus tauri*, *Ostreococcus lucimarinus*, *Phaeodactylum tricornutum*, *Thalassiosira pseudonana* and *Emiliania huxleyi* were obtained from the website of the DOE Joint Genome Institute (Walnut Creek, CA, USA; http://www.jgi.doe.gov). The genomes of the red alga *Cyanidioschyzon merolae* was obtained from the *C. merolae* Genome Project (http://merolae.biol.s.u-tokyo.ac.jp). Each genome was fed into the program makeblastdb to create an organism–species database (Altschul et al. 1990).

Gene Retrieval and Annotation

Genomes of these algae were examined for the presence of biotin biosynthesis-related genes (*BIOF*, *BIOA*, *BIOD*, and *BIOB*). Firstly, we followed JGI's or the *C. merolae* Genome Project's annotation to determine the number of biotin biosynthesis-related genes present in each algal genome. Then, an initial set of query proteins including well characterized and putative BBRE identified from *C. reinhardtii* and *A. thaliana* were obtained from GenBank

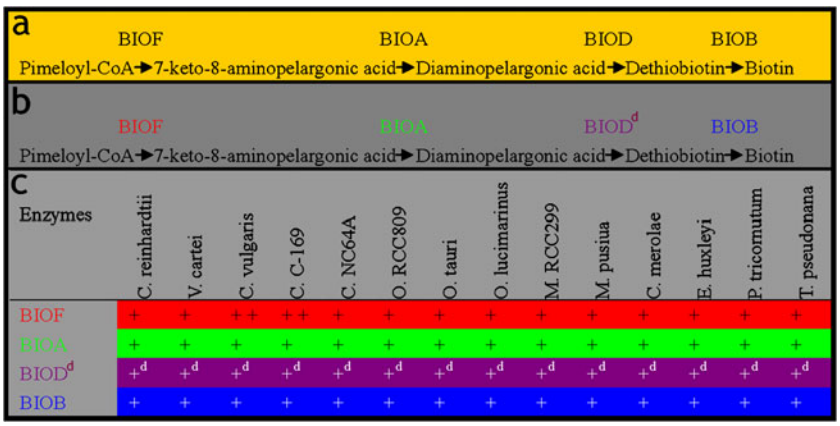

Fig. 1 The scheme of biotin biosynthetic pathway and distribution of presumed enzymes involved in the biosynthetic pathway of biotin among 14 eukaryotic photosynthetic alga. **a** The biotin biosynthesis pathway described in bacteria (Pinon et al. 2005). **b** The putative biotin biosynthetic pathway in algal kingdom. Some enzymes have been functionally identified, while others are just suggested by sequence homology. Database searches were carried out with the BLASTP program. **c** A list of enzymes involved in biotin biosynthesis. Presence or absence of putative orthologs in a corresponding genome is indicated by '+' or '−', respectively. Note: BIOD[d], which stands for *BIOD* gene, is well present in algal genomes but is fused upstream of *BIOA* gene

database (National Center for Biotechnology Information). Each protein in this query dataset was used to search the potential novel sequences in 14 eukaryotic photosynthetic algal genomes sequences were available by BLASTP and TBLASTN programs, with an *E* value <1e–10. The searches were iterated until convergence. The detailed information for *BBRE*, which was included in the query, was summarized in Table 1. BBRE protein sequences from other distinct organisms, analyzed in this study, were summarized in Table 2.

Multiple Sequence Alignment and Phylogenetic Analysis

All protein sequences from our dataset were multiply aligned using ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/; Thompson et al. 1994). The final alignment was further refined after excluding the relatively poorly conserved regions at the protein ends and consisted of sequences spanning the conserved domains. Phylogeny trees were constructed by the neighbor-joining (NJ) method, using the program MEGA 4.0 (http://www.megasoftware.net/; Tamura et al. 2007) and the maximum-likelihood (ML) method, using the program PHYML (http://atgc.lirmm.fr/phyml/; Guindon and Gascuel 2003), with bootstrap support values deriving from 1,000 randomized, replicate datasets.

Structure Domain Predict and Coevolution Analysis

The Simple Modular Architecture Research Tool (SMART and http://smart.embl-heidelberg.de/) was applied to predict the structure domains of these BBRE proteins sequences (Letunic et al. 2009). Coevolution Analysis using Protein Sequences (CAPS and http://bioinf.gen.tcd.ie/caps/) was used to measure the intramolecular coevolution between amino acid sites belonging to BIOB. We used a recently developed parametric model (Fares and Travers 2006), briefly, significance test: random sampling=10,000, threshold alpha value=0.001; weight correlation by divergence time between sequences using synonymous distances by minimum R=0.1 GrSize=5. The sample size of ten was sufficient for accurate results by CAPS analysis (Fares and McNally 2006). The structural PDB file for BIOB (Berkovitch et al. 2004) was used to identify the coevolving amino acid positions in the structure (for example, all the amino acid positions in this study refer to their location in the BIOB 3-dimensional structure). InterMap3D (http://www.cbs.dtu.dk/services/Inter Map3D/) was also applied to test the intramolecular coevolution positions (alignment methods: CLUSTAL; the minimum length of the query protein that has to be matched when looking for homologues (70%); analysis methods: RCW MI, MI/Entropy, and Dependency) between the protein sequences (Gouveia-Oliveira et al. 2009).

Results

General Comparison of *BBRE* Genes of Eukaryotic Photosynthetic Algae

Forty-four *BBRE* genes were predicted and annotated from 14 eukaryotic photosynthetic algal genomes using BLASTP and TBLASN programs with the query sequence. The candidate genes identified in this study were listed in Table 3. The distribution of *BBRE* genes involved in biotin biosynthesis was summarized in Fig. 1. According to the BLASTP results, homologues of *BIOF*, *BIOA*, and *BIOB* are present in algae while it failed to discover homologues of *BIOD*. Moreover, we predicted that BIOA was a bifunctional enzyme (diaminopelargonic acid synthase/dethiobiotin synthase) by the online BLASTP program (i.e., the *BIOD* gene was positioned immediately upstream of *BIOA*). There were two copies of *BIOF* in *C. vulgaris*, and *Coccomyxa* sp. C-169, respectively. The eukaryotic photosynthetic alga shares the same biotin biosynthetic pathway, which is similar within bacteria (Fig. 1b). *BIOF*, *BIOA*, *BIOD*^{d}, and *BIOB* genes were widely distributed among all the species (Fig. 1c).

Multiple Sequence Alignment and Phylogeny Analysis of *BBRE* Genes

Multiple alignments and phylogeny analysis of the predicted homologous *BBRE* genes from database were constructed. The NJ tree topology matches our ML tree (data not shown), and the high branch support values

Table 1 List of organisms and biotin biosynthetic-related enzymes analyzed in query

Gene name	Species	Accession no.	Length	Type
BIOF	*A. thaliana*	NP_196082.2	343	*Streptophyta*
BIOF	*C. reinhardtii*	XP_001693053.1	328	*Chlorophyta*
BIOA	*A. thaliana*	NP_200567.2	833	*Streptophyta*
BIOA	*C. reinhardtii*	XP_001690674.1	819	*Chlorophyta*
BIOD	*S. cerevisiae*	EDN62863.1	237	*Eukaroyta*
BIOB	*A .thaliana*	NP_181864.1	387	*Streptophyta*
BIOB	*C. reinhardtii*	XP_001696322.1	165	*Chlorophyta*

Table 2 List of organisms and biotin biosynthetic enzymes analyzed in this study (except for the sequences from 14 eukaryotic photosynthetic alga genomes)

Gene name	Species	Accession no.	Type
BIOF	*Cyanothece* sp. PCC 8802	ACU99035.1	*Cyanophyta*
	Trichodesmium erythraeum IMS101	ABG49822.1	*Cyanophyta*
	Nostoc punctiforme PCC 73102	ACC81963.1	*Cyanophyta*
	Gloeobacter violaceus PCC 7421	NP_923343.1	*Cyanophyta*
	Synechocystis sp. PCC 6803	NP_442395.2	*Cyanophyta*
	Synechococcus elongatus PCC 7942	YP_399046.1	*Cyanophyta*
	Arabidopsis thaliana	CAB85568.1	*Streptophyta*
	Capnocytophaga sputigena Capno	ZP_03389997.1	*Bacteria*
	Capsaspora owczarzaki ATCC 30864	EFW39876.1	*Eukaroyta*
	Geobacter bemidjiensis Bem	YP_002137650.1	*Bacteria*
BIOA	*Prochlorococcus marinus* NATL1A	ABM76428.1	*Cyanophyta*
	Synechococcus sp. WH 8102	NP_896722.1	*Cyanophyta*
	Prochlorococcus marinus MIT 9303	YP_001016470.1	*Cyanophyta*
	Sclerotinia sclerotiorum 1980 UF-70	DAA33960.1	*Eukaroyta*
	Capsaspora owczarzaki ATCC 30864	EFW46213.1	*Eukaroyta*
	Desulfovibrio salexigens DSM 2638	YP_002992569.1	*Bacteria*
	Arabidopsis thaliana	NP_200567.2	*Streptophyta*
BIOB	*Nostoc punctiforme* PCC 73102	YP_001868973.1	*Cyanophyta*
	Cyanothece sp. PCC 8802	YP_003135947.1	*Cyanophyta*
	Trichodesmium erythraeum IMS101	ABG52621.1	*Cyanophyta*
	Synechococcus elongatus PCC 7942	ABB56451.1	*Cyanophyta*
	Gloeobacter violaceus PCC 7421	NP_924958.1	*Cyanophyta*
	Arabidopsis thaliana	AAO41898.1	*Streptophyta*
	Prochlorococcus marinus NATL2A	YP_291849.1	*Cyanophyta*
	Saccharomyces cerevisiae FostersB	EGA58537.1	*Eukaroyta*
	Capnocytophaga gingivalis ATCC 33624	ZP_04058295.1	*Bacteria*
	Capsaspora owczarzaki ATCC 30864	EFW43202.1	*Eukaroyta*

coincide with high neighbor joining bootstrap values, suggesting that the *BBRE* gene phylogeny is robust in different tree reconstruction methods.

Pimeloyl-CoA is converted to 7-keto-8-aminopelargonic acid by BIOF coded by *BIOF* gene. To elucidate BIOF phylogeny, 16 proteins from the database were analyzed using NJ methods. Observation of the tree revealed that all the BIOF fell into three distinct subfamilies (Fig. 2): BIOF from organisms of *Chlorophyta*, *Streptophyta* (*A. thaliana*), *Haptophyta* (*E. huxleyi*) and bacteria (*Geobacter bemidjiensis Bem*); BIOF from organisms belonging to *Cyanophyta* (except for *Synechococcus elongatus* PCC 7942) and BIOF from the other organisms. According to the results of BLASTP and phylogeny tree, close relationship exists between *Chlorophyta* algae and higher plant. The phylogeny tree shows that *Haptophyta* (*E. huxleyi*) clustered into *Chlorophyta* and *S. elongatus* PCC 7942 was out of the *Cyanophyta* group, suggesting lateral gene transfer. It is worth considering that two copies of *BIOF* with high similarity (Identities=81% and Identities=83%, Table 4) exist in *C. vulgaris* and *Chlorella* sp. C-169, respectively.

Diaminopelargonic acid synthase is the enzyme that catalyzes the formation of diaminopelargonic acid from 7-keto-8-aminopelargonic acid. The evolutionary relationship of *BIOA* genes were demonstrated by the phylogenic tree in Fig. 3. The phylogeny tree showed that there were two distinct evolutionary routines from cyanobacterial to eukaryotic photosynthetic algae. It was interesting that *Chlorophyta* and *Rhodophyta* (*C. merolae*) fell into the same clade, which suggested that these algae acquired *BIOA* genes after primary endosymbiosis. It seems that close evolutionary relationship exists among bacteria (*Desulfovibrio salexigens* DSM 2638), *Haptophyta* (*E. huxleyi*) and *Heterokonphyta* (*T. pseudonana*), suggesting that they obtained the *BIOA* genes from bacteria through lateral gene transfer. It was worth considering that bifunctional BIOA from fungi (*Sclerotinia sclerotiorum* 1980 UF-70 and *Capsaspora owczarzaki* ATCC 30864), named dethiobiotin synthetase/adenosylmethionine-

Table 3 Candidate genes for the enzymes involved in biotin biosynthesis in 14 eukaryotic photosynthetic alga genomes (locations of genes are indicated by positions on either chromosomes or scaffolds). Note: BIOA[d] stands for a bifunctional enzyme (diaminopelargonate synthase/dethiobiotin synthetase)

Species	Type	Gene locations
C. reinhardtii	BIOF	jgi|Chlre4|410225|pasa_Sanger_mRNA17906
	BIOA[d]	jgi|Chlre4|108892|e_gwH.100.1.1
	BIOB	jgi|Chlre4|402991|pasa_Sanger_mRNA27712
V. cartei	BIOF	jgi|Volca1|78219|estExt_Genewise1.C_1190002
	BIOA[d]	jgi|Volca1|40597|gw1.1.404.1
	BIOB	jgi|Volca1|102782|estExt_fgenesh4_pg.C_10400
C. C-169	BIOF	jgi|Coc_C169_1|34890|fgenesh1_pm.1_#_231
	BIOF	jgi|Coc_C169_1|55900|Genemark1.4929_g
	BIOA[d]	jgi|Coc_C169_1|810|gw1.1.95.1
	BIOB	jgi|Coc_C169_1|45922|estExt_fgenesh1_pg.C_10396
C. vulgaris	BIOF	jgi|Chlvu1|34895|estExt_Genewise1.C_10210
	BIOF	jgi|Chlvu1|30868|e_gw1.12.526.1
	BIOA[d]	jgi|Chlvu1|24859|e_gw1.1.275.1
	BIOB	jgi|Chlvu1|40260|estExt_Genewise1Plus.C_10217
C. NC64A	BIOF	jgi|ChlNC64A_1|22749|e_gw1.9.103.1
	BIOA[d]	jgi|ChlNC64A_1|145418|IGS.gm_9_00225
	BIOB	jgi|ChlNC64A_1|22943|e_gw1.9.107.1
O. RCC809	BIOF	jgi|OstRCC809_1|12230|gw1.17.302.1
	BIOA[d]	jgi|OstRCC809_1|43144|fgenesh1_pm.C_scaffold_13000082
	BIOB	jgi|OstRCC809_1|49198|estExt_Genewise1Plus.C_170172
O. tauri	BIOF	jgi|Ostta4|5827|gw1.10.00.319.1
	BIOA[d]	jgi|Ostta4|8959|fgenesh1_pm.C_Chr_13.0001000025
	BIOB	jgi|Ostta4|26432|estExt_genewise1.C_Chr_10.00010220
O. lucimarinus	BIOF	jgi|Ost9901_3|3070|gwEuk.10.177.1
	BIOA[d]	jgi|Ost9901_3|597|gwEuk.21.17.1
	BIOB	jgi|Ost9901_3|46799|estExt_GenewiseEukaryote.C_Chr_100247
M. RCC299	BIOF	jgi|MicpuN2|69048|gw2.09.240.1
	BIOA[d]	jgi|MicpuN2|59263|EuGene.0600010265
	BIOB	jgi|MicpuN2|97893|fgenesh2_pm.C_Chr_09000085
M. pusiua	BIOF	jgi|MicpuC2|49296
	BIOA[d]	jgi|MicpuC2|34980
	BIOB	jgi|MicpuC2|23693
C. merolae	BIOF	gnl|CMER|CML225C
	BIOA[d]	gnl|CMER|CMG023C
	BIOB	gnl|CMER|CML210C
E. huxleyi	BIOF	jgi|Emihu1|72721|e_gw1.190.15.1
	BIOA[d]	jgi|Emihu1|456336|estExtDG_Genemark1.C_1770059
	BIOB	jgi|Emihu1|423761|estExtDG_fgenesh_newKGs_pm.C_1670005
P. tricornutum	BIOF	jgi|Phatr2|36165|fgenesh1_pg.C_chr_9000280
	BIOA[d]	jgi|Phatr2|19762|estExt_gwp_gw1.C_chr_60266
	BIOB	jgi|Phatr2|21296|estExt_gwp_gw1.C_chr_120108
T. pseudonana	BIOF	jgi|Thaps3|31216|e_gw1.1.382.1
	BIOA[d]	jgi|Thaps3|260744|thaps1_ua_kg.chr_1000004
	BIOB	jgi|Thaps3|34585|e_gw1.6.318.1

8-amino-7-oxononanoate aminotransferase, is evolutionary close with eukaryotic alga (Identities=26–41%; Table 4).

Biotin synthase is the last enzyme involved in the formation of biotin in eukaryotic photosynthetic algae. It

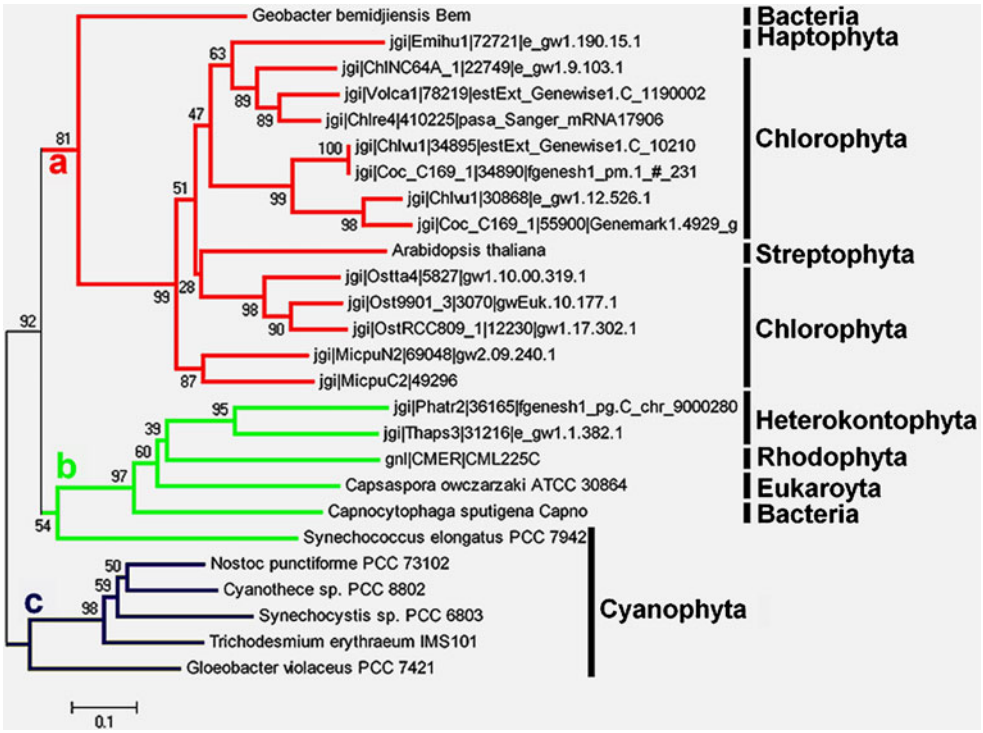

Fig. 2 Phylogenetic tree for the *BIOF* from distinct organisms including cyanobacteria, bacteria, algae, fungi, and higher plants. Major groups of organisms are labeled to allow comparison between the phylogeny of *BIOF* and algae evolution. Note that the overall phylogeny of *BIOF* follows the currently accepted system of classification of algae (Brinkman and Philippe, 2007)

plays a key role in converting dethiobiotin to biotin. According to the results of BLASTP and phylogeny tree (Fig. 4), the *BIOB* genes were conserved in the evolutionary history. BIOB from different organisms shared 51–97% similarities (Table 4). As the protein sequences' phylogeny tree showed that *BIOB* from algae, eukaryota (*Saccharomyces cerevisiae FostersB*), and higher plant (*A. thaliana*) clustered into one group. It is interesting that *BIOB* from *Nostoc punctiforme* PCC 73102 is different from other *Cyanophyta*, may be due to the function of diazotroph.

Structure Domain Prediction and Coevolutionary Analysis

In order to analyze the evidence in the coevolutionary relationship between biotin biosynthetic lineages and other biochemical structures, particularly proteorhodopsins and photosynthetic reaction center, the structure domains of BBREs were predicted by SMART.

Structure domain analysis of BBRE proteins showed that not only Pfam: Aminotran_1_2 existed in BIOF of all organisms but also Pfam: Aminotran_5, Pfam: Beta_elim_-lyase, and Pfam: Cys_Met_Meta_PP existed in BIOF of some organisms. Pfam: Aminotran_3 and CbiA were distributed on BIOA and Pfam: BATS and Elp3 existed in BIOB (Fig. 5). Pfam: Aminotran_1_2, Aminotran_5, and Aminotran_3 belonged to aminotransferases, which share certain mechanistic features with other pyridoxal-phosphate-dependent enzymes, such as the covalent binding of the pyridoxal-phosphate group to a lysine residue. Pfam: Beta_elim_lyase was found in many tryptophanases (tryptophan indole-lyase, TNase), tyrosine phenol-lyases (TPL) and threonine aldolases. It is involved in the degradation of amino acids, which suggested that biotin played a key role in amino acid metabolism. Pfam: Cys_Met_Meta_PP belonged to pyridoxal-phosphate, which is the active form of vitamin B6 (pyridoxine or pyridoxal). PLP is a versatile catalyst, acting as a coenzyme in a multitude of reactions, including decarboxylation, deamination and transamination. Pfam: BATS is biotin and thiamin synthesis-associated domain, biotin synthase (BIOB), catalyzes the last step of the biotin biosynthetic pathway. Pfam: Elp3 is a member of superfamily that contains MoaA, NifB, PqqE, coproporphyrinogen III oxidase, biotin synthase and MiaB families, and includes a representative in the eukaryotic elongator subunit, Elp-3. Coevolutionary analysis showed a large group including a great number of coevolving sites on BIOB (Fig. 6). Accordingly, it can be concluded that intramolecular coevolution contributed to the adaptive evolution of BIOB.

Discussion

It is well known that higher plants, green algae, red algae, and *Glaucocystophyta* algae are derived from a primary endosymbiotic event in which a nonphotosynthetic eukaryote acquired a chloroplast by engulfing (or being invaded by) a prokaryotic cyanobacterium (Falkowski et al. 2004; Lewis

Table 4 The percentage of amino acid sequence similarity for BBRG proteins from different organisms. The *red* rectangle presents BIOF, the *green* rectangle presents BIOB, and the *blue* rectangle presents for BIOA

	C. rei	V. car	C. NC64A	C. vul	C. C-169	O. luc	O. tau	M. pus	M. RCC299	O. RCC809	C. mer	T. pse	P. tri	E. hux
C. rei		76	74	62	62	58	60	58	63	60	53	55	55	51
V. car	63		84	79	81	68	69	63	69	69	60	60	60	56
C. NC64A	57	55		84	81	72	72	63	72	72	60	61	62	59
C. vul-1	56	60	60		97	69	71	63	69	68	60	60	61	56
C. C-169	55	60	60	99		72	69	65	71	71	62	62	63	58
O. luc	46	48	50	57	57		93	78	83	97	65	62	65	64
O. tau	50	46	54	60	59	76		75	82	92	63	63	66	63
M. pus	48	46	54	50	50	50	55		83	79	61	58	59	60
M. RCC299	44	40	54	46	45	53	54	51		83	62	61	62	61
O. RCC809	47	47	52	56	57	77	76	50	50		66	62	63	63
C. mer	32	30	35	34	34	33	32	29	33	33		59	61	60
T. pse	35	30	36	33	33	30	35	29	29	30	42		88	73
P. tri	30	28	32	29	29	26	28	27	29	28	42	46		74
E. hux	47	50	48	55	54	50	58	42	44	48	30	28	26	
S. 1980UF-70	31	31	30	34	35	33	34	32	33	32	26	26	31	33
C. ATCC30864	31	32	35	41	39	34	33	32	34	33	32	31	30	34

and McCourt 2004; Moreira et al. 2000). *C. merolae*, a unicellular *Rhodophyta*, is one of the most primitive red alga and probably diverged from near the root of the red lineage (Sato and Moriyama 2007). In contrast, dominant bloom-forming eukaryotic phytoplankton, such as *Heterokonphyta* and *Haptophya* found in the ocean are derived by secondary endosymbiosis, whereby nonphotosynthetic eukaryote acquired a chloroplast by engulfing a photosynthetic eukaryote, probably a red algal endosymbiont (Croft et al. 2006; Falkowski et al. 2004; Grzebyk et al. 2004; Keeling 2004).

Therefore, 14 photosynthetic algae analyzed in this study should be clustered into three groups: ① *Chlorophyta* (10); ② *Rhodophyta* (1); ③ *Haptophyta* (1) and *Heterokonphyta* (2). The phylogeny trees of *BBRE* are largely in agreement with the organism tree. However, some special events occurred. For example, *BIOF* (*E. huxleyi*) belonged to *Chlorophyta*, suggesting lateral gene transfer phenomena

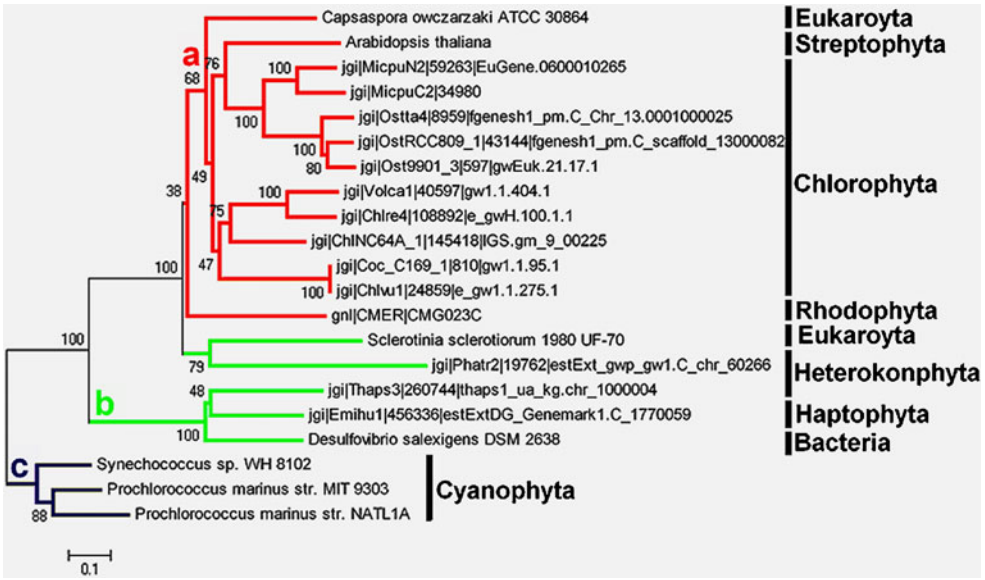

Fig. 3 Phylogenetic tree for the *BIOA* from distinct organisms including cyanobacteria, bacteria, algae, fungi, and higher plants. Major groups of organisms are labeled to allow comparison between the phylogeny of *BIOA* and algae evolution. Note that the overall phylogeny of *BIOA* follows the currently accepted system of classification of algae (Brinkman and Philippe, 2007)

play a significant role in the evolution of the eukaryotic photosynthetic algae (Andersson et al. 2003; Huang et al. 2004; Nixon et al. 2002). Besides, it has already been identified that lateral gene transfer was a potentially important evolutionary mechanism in eukaryotic organisms. An example of a similar situation as *Cyanophyta* is the *BIOF* (*S. elongatus* PCC 7942). Referring to Tran et al. (2009) and Cui et al. (2011), who found that an ancient gene duplication event produced two classes of phytoene synthase and lycopene cyclase, respectively, in the algal kingdom. It is easy to speculate that the same event may have occurred producing two copies of *BIOF* with high similarity in *C. vulgaris* and *Chlorella* sp. C-169, respectively. It was interesting that a close relationship exists among *Haptophyta* (*E. huxleyi*), *Heterokonphyta* (*P. tricornutum* and *T. pseudonana*), and bacteria in the evolution of *BIOF* and *BIOA*, which suggests that they obtained these genes from bacteria through lateral gene transfer.

The analysis of structure domain of BBREs protein sequences from different algae shows multiple protein families. They may be particularly evident in the intermolecular coevolutionary relationships displayed by some

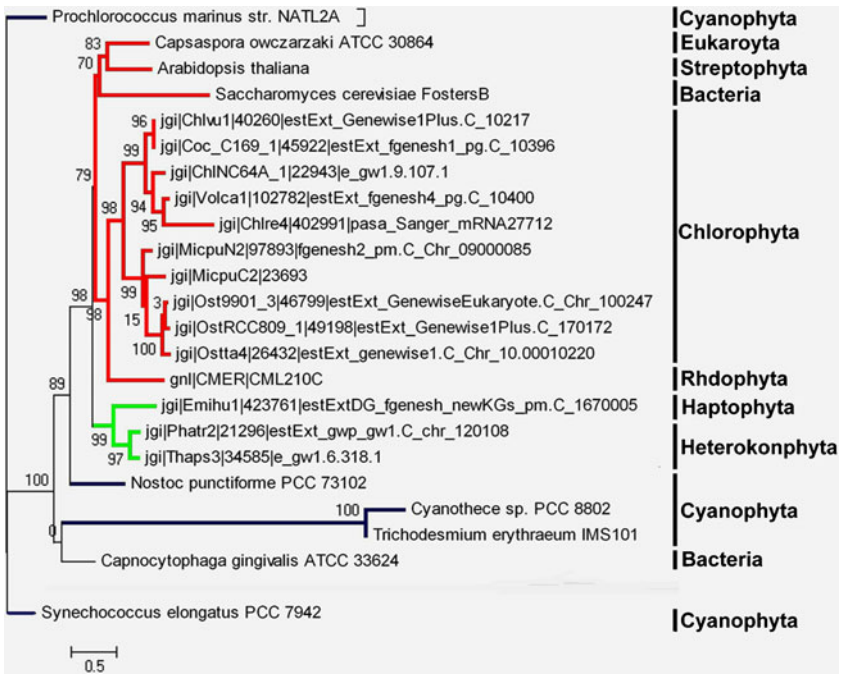

Fig. 4 Phylogenetic tree for the *BIOB* from distinct organisms including cyanobacteria, bacteria, algae, fungi, and higher plants. Major groups of organisms are labeled to allow comparison between the phylogeny of *BIOB* and algae evolution. Note that the overall phylogeny of *BIOB* follows the currently accepted system of classification of algae (Brinkman and Philippe, 2007)

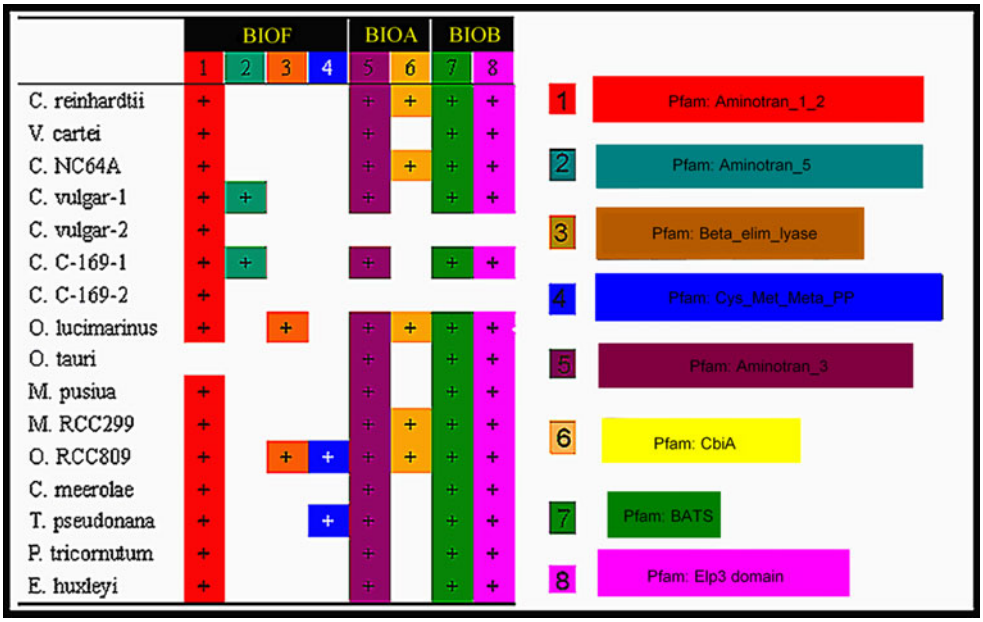

Fig. 5 Analysis of the structure domains in BIOF, BIOA, and BIOB protein sequences from different algae produced with SMART (Simple Modular Architecture Research Tool) program by http://smart.embl-heidelberg.de/. The shading and numbers reflect correspondence to specific structure domains. Presence of putative orthologs in a corresponding genome is indicated by '+'

biotin biosynthetic lineages with other biochemical structures, particularly biosynthesis of long chain fatty acids and citric acid cycle (De Clercq 1997; Gu et al. 2011). It was Klassen (2010) who found that intermolecular coevolutionary relationships occurred among carotenoids biosynthetic lineages, proteorhodopsins, and the photosynthetic reaction center. All these exchanges of catalysts occurred in coevolution with oxygenic photosynthesis, photosynthesis and were inherited to eukaryotes including plants, along with oxygenic photosynthesis (Sandmann 2002). Moreover, there are large groups including a great number of intramolecular coevolving sites on BIOB, which may be responsible for the highly conserved evolution of BIOB. These results suggest that coevolution as a potential important evolutionary mechanism is widespread in eukaryotic organisms. Further research is necessary for delineating the coevolution relationship in and among these structure domains.

It is widely accepted that biotin biosynthesis in bacteria is catalyzed by four enzymes (BIOF, BIOA, BIOD, and BIOB) respectively, although some detailed information about this pathway is still incomplete. In the budding yeast *S. cerevisiae*, homologues of *BIOA*, *BIOD*, and *BIOB*, but not *BIOF*, are present; as a result, the source of 7-keto-8-aminopelargonic acid remains unknown (Giaever et al. 2002; Phalip et al. 1999; Zhang et al. 1994). Initial information on biotin synthesis and transport in plants came from analysis of the enzyme BIOA (*bio1* biotin auxotrophic mutant of *A. thaliana*) (Schneider et al. 1989; Shellhammer and Meinke 1990). Then, the first biochemical characterization of BIOB (encoded by the *bio2* gene in *A. thaliana*) was described (Baldet et al. 1997; Picciocchi et al. 2001 and 2003). The enzyme, BIOF, involved in the first step of biotin biosynthetic pathway, was cloned and functionally invested (Pinon et al. 2005). In this paper, homologous genes of *BIOF*, *BIOA*,

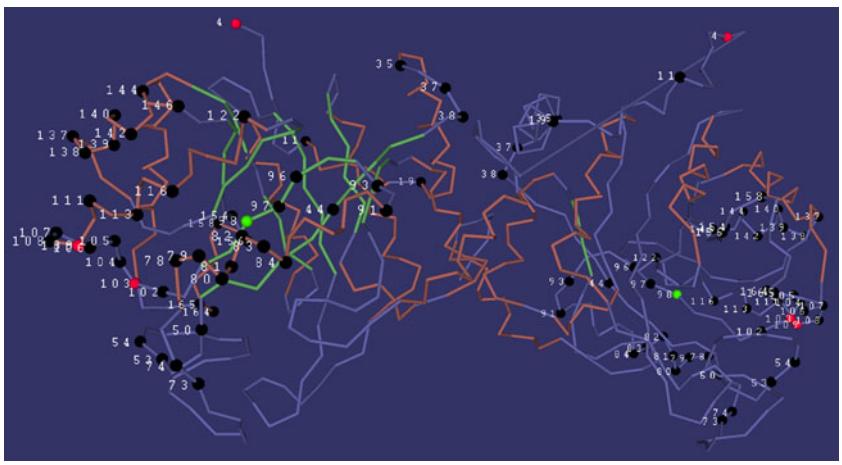

Fig. 6 The 3D displays of intramolecular coevolutionary sites on the BIOB. Balls of the same color stand for the coevolving pairs of sites in the same group

and *BIOB* were present in eukaryotic algae according to our BLASP results, which is in accordance with previously researches.

However, in this paper, we failed to discover the homologous gene of *BIOD* within the eukaryotic photosynthetic alga across the corresponding genomes according to BLASTP program. Many hypotheses responsible for this event have been discussed widely. It was Croft et al. (2006) who speculated that the conversion of 7,8-diaminopelargonic acid to dethiobiotin must be carried out by an as yet unidentified enzyme. The source of dethiobiotin in alga may result from an important and unsuspected symbiosis with bacteria. The source of cobalamin seems to be bacteria, indicating an important and unsuspected symbiosis (Croft et al. 2005). The importance of biotin in the symbiotic interactions between the alga and the fungus in *Peltigera* was discussed (Bednar and Holm-Hansen 1964). It was notable that the protein sequence of BIOA from alga showed high similarity to the bifunctional BIOA (dethiobiotin synthetase/adenosylmethionine-8-amino-7-oxononanoate aminotransferase) from fungi (*S. sclerotiorum* 1980 UF-70 and *C. owczarzaki* ATCC 30864). In fact, a close inspection of gene sequences reveals that the *BIOD* gene is well present in algal genomes but is fused upstream of the *BIOA* gene (Derelle et al. 2006; Matsuzaki et al. 2004; Misumi et al. 2005). Moreover, this was the case in most ascomycete and basidiomycete fungi and in flowering plants, as it was clearly demonstrated by Hall and Dietrich (2007) and Muralla et al. (2008), respectively. Poirier et al. (2011) have identified a novel BIOD from *A. nidulans*, which perform above both activities. A mitochondrial bifunctional enzyme (diaminopelargonate synthase/dethiobiotin synthetase) has been discovered in Yunnan Red Pear (Zhang et al. 2011). Thus, it appears that eukaryotes including algae present a bifunctional enzyme able to catalyze both diaminopelargonic acid synthase (BIOA) and dethiobiotin synthetase (BIOD) activities. These observations suggest that a fusion event between prokaryotic BIOD and BIOA ancestor genes occurred early in the evolution of contemporary eukaryotes (Muralla et al. 2008).

Conclusion

In the present study, a comprehensive analysis including gene annotation, phylogeny, structure domain profiling, and coevolution of *BBRE* genes in 14 eukaryotic photosynthetic algae was performed. A sum of 44 *BBRE* genes from corresponding genomes was identified. According to the results of BLASTP and TBLASTN, homologues of *BIOF*, *BIOA*, and *BIOB*, were present in eukaryotic alga. The *BIOD* gene was present upstream of the *BIOA* gene. The phylogenetic trees of BBRE in algae were consistent with the organism tree (Philippe and Brinkmann 2007) except for some special events that occurred in the evolutionary history. The structure domains of BBRE were highly conserved in each subfamily, indicating their functional conservation. The results of CAPS and Inter Map 3D suggested that coevolution pattern played a key role in sustaining conservation of the evolution of *BIOB*. The comprehensive annotation undertaken in this investigation improved our understanding of the involvement of these genes in biotin synthesis in eukaryotic photosynthesis algae and provided the source for selection of candidate genes for functional validation studies.

Acknowledgments This work was supported by the National Natural Science Foundation of China (40876082), Chinese Academy of Sciences, international partnership program 'Typical environment processes and their effects on resources', and Outstanding Young Scholars Fellowship of Shandong Province (Molecular Phycology) (JQ200914).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Andersson JO, Sjogren AM, Davis LAM, Embley TM, Roger AJ (2003) Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. Curr Biol 13:94–104

Baldet P, Alban C, Douce R (1997) Biotin synthesis in higher plants: purification and characterization of bioB gene product equivalent from *Arabidopsis thaliana* overexpressed in *Escherichia coli* and its subcellular localization in pea leaf cells. FEBS Lett 419:206–210

Bednar TW, Holm-Hansen O (1964) Biotin liberation by the lichen alga *Coccomyxa* sp. and by *Chlorella pyrenoidosa*. Plant Cell Physiol 5:297

Berkovitch F, Nicolet Y, Wan JT, Jarrett JT, Drennan CL (2004) Crystal structure of biotin synthase, an S-adenosylmethionine-dependent radical enzyme. Science 303:76–79

Chi XY, Zhang XW, Guan XY, Ding L, Li YX, Wang MQ, Lin HZ, Qin S (2008) Fatty acid biosynthesis in eukaryotic photosynthetic microalgae: Identification of a microsomal delta 12 desaturase in *Chlamydomonas reinhardtii*. J Microbiol 46:189–201

Chi XY, Yang QL, Lu YD, Wang JY, Zhang QF, Pan LJ, Chen MN, He YA, Yu SL (2011) Genome-wide analysis of fatty acid desaturases in soybean (*Glycine max*). Plant Mol Biol Rep. doi:10.1007/s11105-010-0284-z

Croft MT, Lawrence AD, Raux-Deery E, Warren MJ, Smith AG (2005) Algae acquire vitamin B12 through a symbiotic relationship with bacteria. Nature 438:90–93

Croft MT, Warren MJ, Smith AG (2006) Algae need their vitamins. Eukaryot Cell 5:1175–1183

Cui HL, Wang YC, Qin S (2011) Molecular evolution of lycopene cyclases involved in the formation of carotenoids in eukaryotic algae. Plant Mol Biol Rep. doi:10.1007/s11105-011-0297-2

De Clercq PJ (1997) Biotin: a timeless challenge for total synthesis. Chem Rev 97:1755–1792

Demetriou K, Kapazoglou A, Bladenopoulos K, Tsaftaris AS (2010) Epigenetic chromatin modifiers in barley: II. Characterization and expression analysis of the *HDA1* family of barley histone deacetylases during development and in response to jasmonic acid. Plant Mol Biol Rep. doi:10.1007/s11105-009-0121-4

Derelle E, Ferraz C, Rombauts S, Rouzé P, Worden AZ, Robbens S, Partensky F, Degroeve S, Echeynié S, Cooke R, Saeys Y, Wuyts J, Jabbari K, Bowler C, Panaud O, Piégu B, Ball SG, Ral JP, Bouget FY, Piganeau G, Baets BD, Picard A, Delseny M, Demaille J, Peer YVD, Moreau H (2006) Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. Proc Natl Acad Sci USA 103:11647–11652

Entcheva P, Phillips DA, Streit WR (2002) Functional analysis of Sinorhizobium meliloti genes involved in biotin synthesis and transport. Appl Environ Microbiol 68:2843–2848

Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O, Taylor FJR (2004) The evolution of modern eukaryotic phytoplankton. Science 305:354–360

Fares MA, McNally D (2006) CAPS: coevolution analysis using protein sequences. Bioinformatics 22:2821

Fares MA, Travers SAA (2006) A novel method for detecting intramolecular coevolution: adding a further dimension to selective constraints analyses. Genetics 173:9

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. Science 281:237–240

Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, Dow S, Lucau-Danila A, Anderson K, André B (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. Nature 418:387–391

Gouveia-Oliveira R, Roque FS, Wernersson R, Sicheritz-Ponten T, Sackett PW, Mlgaard A, Pedersen AG (2009) InterMap3D: predicting and visualizing co-evolving protein residues. Bioinformatics 25:1963–1965

Grossman AR (2005) Paths toward algal genomics. Plant Physiol 137:410

Grzebyk D, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O, Taylor FJR, Falkowski PG (2004) Response to comment on "The evolution of modern eukaryotic phytoplankton". Science 306(5705):2191

Gu K, Chiam H, Tian D, Yin Z (2011) Molecular cloning and expression of heteromeric ACCase subunit genes from *Jatropha curcas*. Plant Sci. doi:10.1016/j.plantsci.2011.01.007

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704

Hall C, Dietrich FS (2007) The reacquisition of biotin prototrophy in *Saccharomyces cerevisiae* involved horizontal gene transfer, gene duplication, and gene clustering. Genetics 177:2293–2307

Huang JL, Mullapudi N, Sicheritz-Ponten T, Kissinger JC (2004) A first glimpse into the pattern and scale of gene transfer in the Apicomplexa. Int J Parasitol 34:265–274

Keeling PJ (2004) Diversity and evolutionary history of plastids and their hosts. Am J Bot 91:1481–1493

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

Letunic I, Doerks T, Bork P (2009) SMART 6: recent updates and new developments. Nucleic Acids Res 37:D229–D232

Lewis LA, McCourt RM (2004) Green algae and the origin of land plants. Am J Bot 91:1535

Li MJ, Xia H, Zhao CZ, Li AQ, Li CS, Bi YP, Wan SB, Wang XJ (2010) Isolation and characterization of putative acetyl-CoA carboxylases in *Arachis hypogaea* L. Plant Mol Biol Rep 28:58–68

Matsuzaki M, Misumi O, Shin-I T et al (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. Nature 428:653–657

Misumi O, MatsuzakiM NH, Miyagishima SY, Mori T, Nishida K, Yagisawa F, Yoshida Y, Kuroiwa H, Kuroiwa T (2005) *Cyanidioschyzon merolae* genome: a tool for facilitating comparable studies on organelle biogenesis in photosynthetic eukaryotes. Plant Physiol 137:567–585

Moreira D, Le Guyader H, Philippe H (2000) The origin of red algae and the evolution of chloroplasts. Nature 405:69–72

Muralla R, Chen E, Sweeney C, Gray JA, Dickerman A, Nikolau BJ, Meinke D (2008) A bifunctional locus (BIO3-BIO1) required for biotin biosynthesis in Arabidopsis. Plant Physiol 146:60–73

Nixon JEJ, Wang A, Field J, Morrison HG, McArthur AG, Sogin ML, Loftus BJ, Samuelson J (2002) Evidence for lateral transfer of genes encoding ferredoxins, nitroreductases, NADH oxidase, and alcohol dehydrogenase 3 from anaerobic prokaryotes to *Giardia lamblia* and *Entamoeba histolytica*. Eukaryot Cell 1:181–190

Phalip V, Kuhn I, Lemoine Y, Jeltsch JM (1999) Characterization of the biotin biosynthesis pathway in *Saccharomyces cerevisiae* and evidence for a cluster containing BIO5, a novel gene involved in vitamer uptake. Gene 232:43–51

Philippe H, Brinkmann H (2007) The diversity of eukaryotes and the root of the eukaryotic tree. In: Eukaryotic Membranes and Cytoskeleton: Origins and Evolution. Springer, New York, 20–37

Picciocchi A, Douce R, Alban C (2001) Biochemical characterization of the *Arabidopsis* biotin synthase reaction. The importance of mitochondria in biotin synthesis. Plant Physiol 127:1224–1233

Picciocchi A, Douce R, Alban C (2003) The plant biotin synthase reaction. Identification and characterization of essential mitochondrial accessory protein components. J Biol Chem 278:24966–24975

Pinon V, Ravanel S, Douce R, Alban C (2005) Biotin synthesis in plants. The first committed step of the pathway is catalyzed by a cytosolic 7-keto-8-aminopelargonic acid synthase. Plant Physiol 139:1666–1676

Poirier Y, Magliano P, Flipphi M, Sanglard D (2011) Characterization of the *Aspergillus nidulans* biotin biosynthetic gene cluster and use of the bioDA gene as a new transformation marker. Fungal Genet Biol 48:208–215

Round FE (1973) The biology of algae, 2nd edn. Edward Arnold, London

Sandmann G (2002) Molecular evolution of carotenoid biosynthesis from bacteria to plants. Physiol Plant 116:431–440

Sato N, Moriyama T (2007) Genomic and biochemical analysis of lipid biosynthesis in the unicellular rhodophyte *Cyanidioschyzon merolae*: Lack of a plastidic desaturation pathway results in the coupled pathway of galactolipid synthesis. Eukaryot Cell 6:1006–1017

Schneider T, Dinkins R, Robinson K, Shellhammer J, Meinke DW (1989) An embryo-lethal mutant of *Arabidopsis thaliana* is a biotin auxotroph. Dev Biol 131:161–167

Shellhammer J, Meinke D (1990) Arrested embryos from the bio1 auxotroph of *Arabidopsis thaliana* contain reduced levels of biotin. Plant Physiol 93:1162–1167

Stiller JW, Hall BD (1997) The origin of red algae: implications for plastid evolution. Proceedings of the National Academy of Sciences of the United States of America 94: 4520

Stolf-Moreira R, Lemos EGM, Carareto-Alves L et al (2011) Transcriptional profiles of roots of different soybean genotypes subjected to drought stress. Plant Mol Biol Rep 29:19–34

Streit WR, Entcheva P (2003) Biotin in microbes, the genes involved in its biosynthesis, its biochemical role and perspectives for biotechnological production. Appl Microbiol Biotechnol 61:21–31

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. 2007. Mol Biol Evol 24:1596–1599

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673

Tran D, Haven J, Qiu WG, Polle JEW (2009) An update on carotenoid biosynthesis in algae: phylogenetic evidence for the existence of two classes of phytoene synthase. Planta 229:723–729

Webb ME, Marquet A, Mendel RR, Rebeille F, Smith AG (2007) Elucidating biosynthetic pathways for vitamins and cofactors. Nat Prod Rep 24:988–1008

Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D (2004) A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol 21:809

Zempleni J (2005) Uptake, localization, and noncarboxylase roles of biotin. Annu Rev Nutr 25:175–196

Zempleni J, Mock DM (1999) Biotin biochemistry and human requirements* 1. J Nutr Biochem 10:128–138

Zempleni J, Wijeratne SSK, Hassan YI (2009) Biotin. Biofactors 35:36–46

Zhang HJ, Sun LW, Liu LL, Lian J, An SL, Wang X, Zhang J, Jin JL, Li SY, Xi JH (2010) Proteomic analysis of interactions between the generalist herbivore *Spodoptera exigua* (Lepidoptera: Noctuidae) and *Arabidopsis thaliana*. Plant Mol Biol Rep 28:324–333

Zhang S, Sanyal I, Bulboaca GH, Rich A, Flint DH (1994) The gene for biotin synthase from *Saccharomyces cerevisiae*: cloning, sequencing, and complementation of *Escherichia coli* strains lacking biotin synthase. Arch Biochem Biophys 309:29–35

Zhang XD, Allan AC, Yi Q, Chen LM, Li KZ, Shu Q, Su J (2011) Differential gene expression analysis of Yunnan Red Pear, *Pyrus Pyrifolia*, during fruit skin coloration. Plant Mol Biol Rep 29:305–314