Amino Acids
DOI 10.1007/s00726-013-1462-8

INVITED REVIEW

Phylogenetic aspects of the sulfate assimilation genes from *Thalassiosira pseudonana*

Mariusz A. Bromke · Rainer Hoefgen · Holger Hesse

Received: 3 January 2013 / Accepted: 5 January 2013

Abstract Diatoms are unicellular algae responsible for approximately 20 % of global carbon fixation. Their evolution by secondary endocytobiosis resulted in a complex cellular structure and metabolism compared to algae with primary plastids. In the last years the interest on unicellular algae increased. On the one hand assessments suggest that diatom-mediated export production can influence climate change through uptake and sequestration of atmospheric CO_2. On the other hand diatoms are in focus because they are discussed as potential producer of biofuels. To follow the one or other idea it is necessary to investigate the diatoms biochemistry in order to understand the cellular regulatory mechanisms. The sulfur assimilation and methionine synthesis pathways provide S-containing amino acids for the synthesis of proteins and a range of metabolites such as dimethylsulfoniopropionate (DMSP) in order to provide basic metabolic precursors needed for the diatoms metabolism. To obtain an insight into the localization and organization of the sulfur metabolism pathways, the genome of *Thalassiosira pseudonana*—a model organism for diatom research—might help to understand the fundamental questions on adaptive responses of diatoms to dynamic environmental conditions such as nutrient availability in a broader context.

Keywords Sulfate · Cysteine · Homocysteine · *Thalassiosira pseudonana*

M. A. Bromke · R. Hoefgen · H. Hesse (✉)
Max Planck Institute of Molecular Plant Physiology, Wissenschaftspark Potsdam/Golm, Am Mühlenberg 1, 14476 Potsdam, Germany
e-mail: holhesse@gmx.de

Introduction

Diatoms are abundant unicellular algae in aquatic habitats. They produce massive amounts of biomass and are thought to contribute to about 20 % of global carbon fixation (Falkowski et al. 1998; Field et al. 1998). Phytoplankton affects the climate on the global scale, not only by sequestration of CO_2 but also by production of S-containing volatile compounds. According to the hypothesis proposed in 1987 by Charlson, Lovelock, Andreae and Warren (the CLAW hypothesis) phytoplankton is able to regulate the Earth's climate through the generation of clouds by producing the precursor of dimethylsulfide (DMS). This volatile compound is oxidized in the atmosphere to give non-sea salt sulfate, which acts as cloud condensation nuclei, which in turn modifies both the degree of cloudiness and the albedo of clouds. Increased cloudiness and albedo would reduce the extent to which the surface of the planet is warmed by the sun (Charlson et al. 1987). It is not clear whether the system acts as a negative feedback loop, in which higher temperature and higher availability of carbon for phytoplankton and consecutive DMS release may modulate the greenhouse effect of increased anthropogenic CO_2 input to the atmosphere. Dimethylsulfide is created from dimethylsulfoniopropionate (DMSP), which synthesis depends on S-containing compounds.

Sea water contains approximately 28 mM sulfate, providing an unlimited availability of that compound to generate reduced sulfur. Sulfur is found in all organisms as a constituent of the proteogenic amino acids cysteine and methionine and in many coenzymes and other metabolites. In most of these compounds sulfur is present in a reduced form of organic thiols or sulfides. The major form of sulfur in nature, however, is inorganic sulfate. Sulfate is taken up into cells, reduced, and incorporated into cysteine or

Published online: 26 January 2013

homocysteine via the sulfate assimilation pathway to form the first organic S-containing metabolite (Hesse and Hoefgen 2008).

There are differences in organization of sulfur assimilation between sulfur-assimilating organisms. The constituent steps, the enzymes catalysing reduction and the sulfur transfer are not the same among all organisms. The process of sulfur assimilation starts with sulfate uptake by the sulfate transporter being either member of a group of facilitating transporters (Kertesz 2001) or by active transport by high and low affinity sulfate transporter (Buchner et al. 2004; Marzluf 1997). Since sulfate is an inert molecule, for reduction it has to be activated via adenylation to adenosine 5′phosphosulfate (APS) catalysed by ATP sulfurylase (ATPS). In some organisms, such as fungi and many bacteria, a second activation step is required—the phosphorylation of APS with use of APS kinase to form 3′phosphoadenosine 5′phosphosulfate (PAPS) (Kopriva and Koprivova 2004). The reduction of activated sulfate consists of two steps. First, APS or PAPS is reduced to sulfite by APS or PAPS reductases, respectively. In the successive step, sulfite is reduced to sulfide by a sulfite reductase (SiR). In plants and bacteria, *O*-acetylserine-(thiol)lyase (OASTL) transfers sulfide to the amino acid skeleton of *O*-acetylserine (OAS) to form cysteine. In fungi and bacteria, *O*-acetylhomoserine and *O*-succinylhomoserine, respectively, can be directly sulfhydrylated to form homocysteine, the precursor of methionine (Rowbury and Woods 1964; Foglino et al. 1995; Park et al. 1998; Kerr and Flavin 1970).

The compartmentation of sulfate assimilation is an intriguing aspect in understanding S-assimilation and methionine synthesis. While sulfate activation takes place in the cytosol and in plastids, OAS and cysteine are synthesized in all three compartments: cytosol, plastid, and mitochondria (Hesse et al. 2004; Krueger et al. 2009). In photosynthetic eukaryotes, sulfate reduction is confined to plastids (Brunold and Schiff 1976; Hesse et al. 2004).

Research on diatoms advanced significantly with publication of the whole genome sequences of *Thalassiosira pseudonana* (Armbrust et al. 2004) and *Phaeodactylum tricornutum* (Bowler et al. 2008). Diatoms are amazing organisms derived from a nonphotosynthetic eukaryote that domesticated a photoautotrophic eukaryotic cell phylogenetically close to a red alga (Parker et al. 2008). This resulted in an extensive gene transfer event and genomic reorganization. Secondary endocytobiosis also increased the complexity of diatom cell structure, with implications on physiology and biochemistry. An important aspect of diatom cell morphology is that plastids are surrounded by four rather than two membranes with the consequence that all nuclear-encoded plastid proteins have to cross four membranes (Gibbs 1979). The same is true for the exchange of metabolites between plastids and the cytoplasm. Still, little is known about sulfur uptake, assimilation, metabolism and compartmentation in diatoms. Here, we summarize our current understanding of the regulatory network to understand the diatoms sulfur metabolism.

Prediction of intracellular localization

Diatoms differ from land plants or green algae in their cellular organization. Due to the evolutionary history of diatoms, the cells of these organisms contain plastids surrounded by additional two layers of plasma membrane. The outermost membrane is continuous with the endoplasmic reticulum (Gibbs 1979). On the basis of the obtained amino acid sequences, the subcellular localization of proteins involved in sulfur assimilation and methionine synthesis was predicted. Location assignment is based mostly on the predicted presence of N-terminal mitochondrial targeting peptide (mTP) or secretory pathway signal peptide (SP). However, in *T. pseudonana* the used signal peptides are not known. Thus, further N-terminal sequences longer than 15 amino acids in front of conserved catalytic domains have to be included that might form specific diatomal targeting peptides, which are not recognized by the used software (Table 1). Most of the analyzed sequences did not show pre-sequence on which basis the proteins could be sorted into the respective compartment. Only the enzyme gamma-glutamylcysteine synthetase contains a peptide that might direct into mitochondria (Table 1). It is known that in plants the synthesis of the glutathione precursor γ-glutamylcysteine takes place in plastids (Wachter et al. 2005). The incomplete (lack of starting M) sequence of γ-glutamylcysteine synthetase (GSH1) shows high similarity to fungal and mammalian sequences, suggesting that, despite its TargetP prediction of mitochondrial localization, the enzyme might be localized to the cytosol. In the protein model the conserved catalytic domain starts from 231 amino acid of the analyzed sequence. This is in good agreement with known GSH1 proteins showing γ-glutamylcysteine synthetase activity. Therefore, the obtained prediction results concerning mitochondrial localization of those proteins are rather weak and do not agree with common knowledge. Two out of three permeases contain a sequence that, according to the TargetP software, traffics them to the secretory pathway. The other membrane transport protein (ST 1) showed no such sequence. A set of six proteins such as sulfurylase-kinase-pyrophosphatase, APS reductase 3, sulfite reductase (SiR), and OAS (thiol)lyase 4 are predicted to be secreted according to their pre-sequence motifs. In diatoms the outermost membrane of plastids is continuous with the endoplasmic reticulum, through which the secreted proteins are transported. The

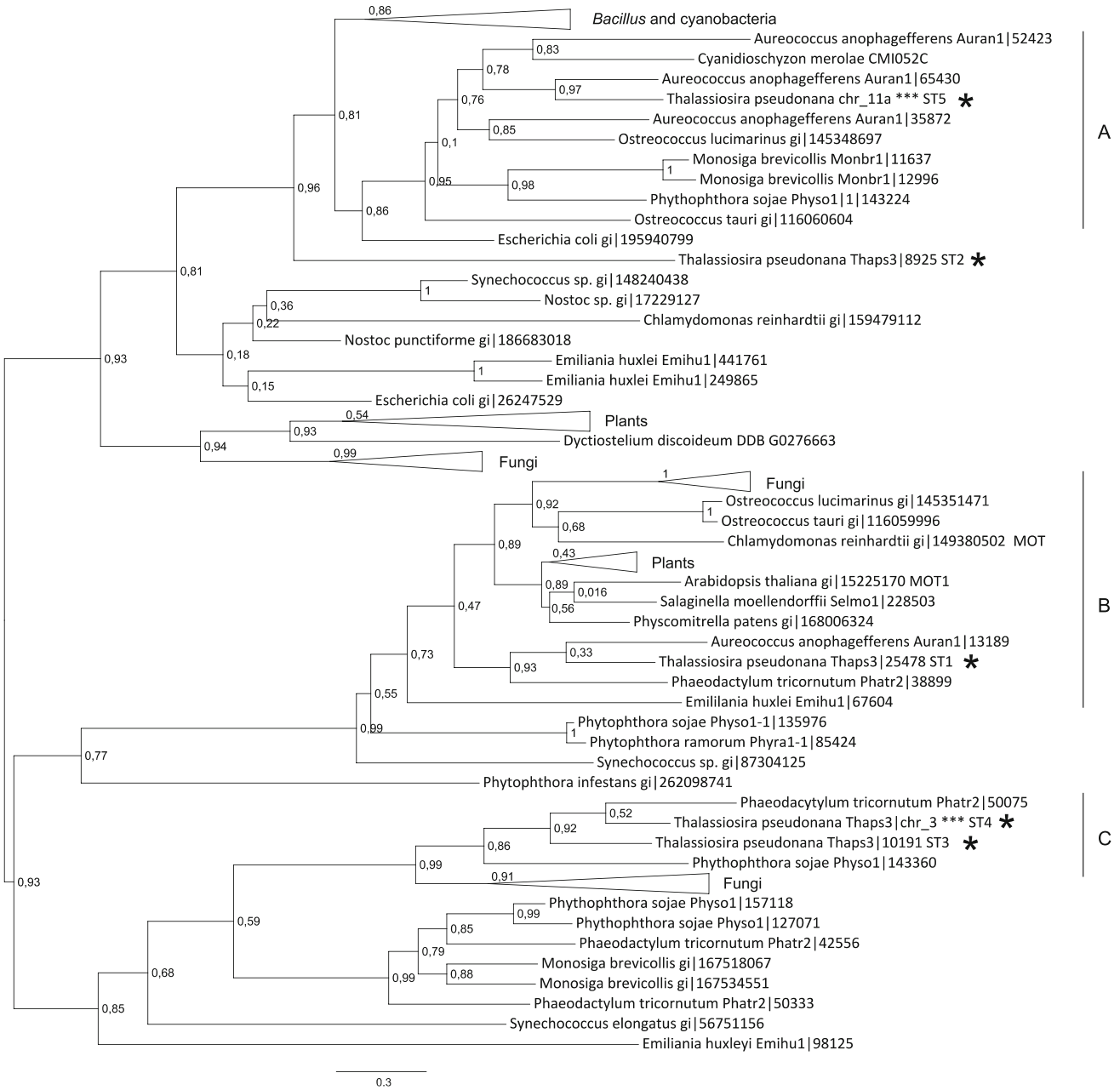

Fig. 1 Evolutionary relationship of sulfate transporter/permease of proteins from *T. pseudonana* and *E. coli, Bacillus* sp., *Synechococcus* sp., *Nostoc* sp., *Phaeodactylum tricornutum, Emiliania huxleyi, Aureococcus anophagefferens, Ostreococcus* sp., *Chlamydomonas reinhardtii, Phytophthora* sp., *Selaginella moellendorffii, Physcomitrella patens, A. thaliana, O. sativa, Porphyra purpurea, Cyanidioschyzon merolae, Monosiga brevicollis, Dictyostelium discoideum,* and *Trypanosoma cruzi*. Amino acid sequences of annotated sulfur metabolism gene models of *T. pseudonana* were identified by protein–protein BLAST (Altschul et al. 1997) in order to identify the orthologous genes from other organisms. The obtained results from blastp were by use of NCBI Conserved Domain Search (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Marchler-Bauer et al. 2005) for presence of conserved domains. Additionally, InParanoid7 (http://inparanoid.sbc.su.se/) was used to verify orthology of eukaryotic sequences (Ostlund et al. 2010). The protein sequences of *C. merolae* were obtained from http://merolae.biol.s.u-tokyo.ac.jp. Sequences of *T. pseudonana*, *P. tricornutum*, *S. moellendorffii*, *E. huxleyi*, *P. sojae, A. anophagefferens, M. brevicollis* proteins were acquired from http://genome.jgi-psf.org. The protein sequences from *D. discoideum* were obtained from http://dictybase.org. Sequences of proteins from *T. cruzi* were obtained from http://www.genedb.org. Sequences from other organisms were retrieved from http://blast.ncbi.nlm.nih.gov/Blast.cgi. The phylogenetic tree was reconstructed using the maximum likelihood method implemented in the PhyML program (http://www.phylogeny.fr. Sequences were aligned by MUSCLE 3.7 (Edgar 2004) and curation was done by Gblocks 0.92 (Castresana 2000). The phylogenetic trees were reconstructed using the maximum likelihood method implemented in the PhyML program (v3.0 aLRT) (Anisimova and Gascuel 2006; Guindon and Gascuel 2003). *T. pseudonana* proteins are marked with *asterisks*. Reliability for internal branch was assessed using the aLRT test as bootstrapping procedure. Number next to species name represents entry code of given protein in a databank. FigTree 1.3.1 software (http://tree.bio.ed.ac.uk/software/figtree/) was used to generate graphics

N-terminal bipartite pre-sequences consist of a signal and a transit peptide-like domain directs the nuclear-encoded protein into plastids. Moreover, presence of a conserved amino acid motif at the signal peptide's predicted cleavage site (ASAFAP) was demonstrated as a trait of the proteins localized to diatomal plastids (Gruber et al. 2007; Kilian and Kroth 2005). Although none of the protein sequences contain the exact conserved ASAFAP-motif, SiR e.g. contains phenylalanine in the forth position of the motif. It has been shown that phenylalanine in this position is crucial for successful transfer to the plastid. In some cases phenylalanine (F in the motif sequence) can be exchanged by other bulky amino acid—tryptophan, tyrosine or leucine (Gruber et al. 2007). Despite the lack of the typical motif in the signal peptides, the other three sequences (sulfurylase-kinase-pyrophosphatase, APS reductase 3 and OAS (thiol)lyase 4) might be as well localised in plastids. No motifs responsible for retention in the endoplasmic reticulum (KDEL, DEEL or DDEL) were found in C-terminal parts of analyzed proteins.

Phylogenetic analysis of sulfate transporters/permeases

In diatoms, sulfate from the environment passes the plasma membrane through channels made by sulfate transporters or permeases, which probably control the influx of the anion into the cells. In the *T. pseudonana* genome five sequences were found, which by their similarity could be annotated as sulfate transporters or permeases. Four of

Table 1 Predicted target peptides of selected proteins of *Thalassiosira pseudonana*

Protein	Abbreviation	Protein ID	Length	mTP	SP	Other	Loc	RC	TPlen	Cleavage site	CD start
Sulfate transporter/permease	ST 1	25478	587	0.206	0.088	0.743	_	3	–		ND
Sulfate transporter/permease	ST 2	8925	541	0.035	0.946	0.222	S	2	26	EAI-SYA	ND
Sulfate transporter/permease	ST 3	10191	1088	0.035	0.426	0.862	_	3	–		
Sulfate transporter/permease	ST 4	3373	942	0.715	0.026	0.313	M	3	8	RPN-AHI	94
Sulfate transporter/permease	ST 5	8725	637	0.191	0.626	0.13	S	3	16	AFA-HPA	81
APS kinase	APK	35055	199	0.41	0.044	0.553	_	5	–		31
ATP sulfurylase	ATPS 2	31173	406	0.417	0.081	0.423	_	5	–		55
ATP sulfurylase complex (sulfurylase-kinase-pyrophosphatase)	SKP	269714	968	0.053	0.943	0.039	S	1	20	GQA-GAV	152
APS reductase	APR 1	24887	387	0.052	0.055	0.96	_	1	–		1
APS reductase	APR 2	270415	394	0.025	0.94	0.075	S	1	26	LND-VPV	20
Sulfite reductase	SiR	22293	631	0.255	0.805	0.017	S	3	20	AAA-FVT	43
Serine acetyltransferase	SAT 1	31984	313	0.092	0.081	0.905	_	1	–		39
Serine acetyltransferase	SAT 2	270416	524	0.165	0.127	0.562	_	4	–		1
Serine acetyltransferase	SAT 3	37497	381	0.037	0.114	0.95	_	1	–		2
OAS (thiol)lyase	OASTL 1	31829	358	0.34	0.176	0.36	_	5	–		44
OAS (thiol)lyase	OASTL 2	1247	311	0.258	0.073	0.746	_	3	–		1
OAS (thiol)lyase	OASTL 3	267987	355	0.267	0.025	0.792	_	3	–		15
OAS (thiol)lyase	OASTL 4	264585	449	0.181	0.744	0.036	S	3	28	VLS-AVR	60
Glutamylcysteine synthetase	ECS	13064	636	0.57	0.053	0.445	M	5	36	NRV-KDL	231
Glutathione synthetase	GSHS	29212	533	0.044	0.298	0.826	_	3	–		6

Columns "mTP", "SP", "other" contains final NN scores on which the prediction is based. "Loc", is the predicted localization: "S", secreted; "M", mitochondrial; "_", any other localization. "RC" stands for reliability class. "TPlen" is the predicted presequence length and the respective cleavage site motif is presented in next column. "CD Start" column contains lengths of N-terminal sequences preceding conserved catalytical domains of respective proteins. "ND" stands for "not detected"

Identification of N-terminal extensions from the conserved regions of the target peptides was performed with use of NCBI Conserved Domain Search (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Marchler-Bauer et al. 2005). Prediction of subcellular localization was performed with TargetP 1.1 online software (http://www.cbs.dtu.dk/services/TargetP/) as described in Emanuelsson et al. (2007). The scores are not probabilities but the most likely according to TargetP. Each prediction has its reliability class (RC) expressed in number from 1 to 5, where 1 indicates the strongest prediction. Presence of ER retention motifs in analyzed sequences (KDEL, DEEL or DDEL) was verified manually

those sequences fall into three distinct clusters (Fig. 1; clusters A, B and C) and the fifth sulfate transporter is basal to the clade A (Fig. 1).

Sulfate transporter 1 (Fig. 1, cluster B; ST 1) shows highest similarity to proteins from diatom *Phaeodactylum tricornutum* and *Aureococcus anophagefferens*. This clade, although not strongly supported (aLRT = 0.47), is found clustering together with plant and green algae sequences (Fig. 1, cluster B). Two sequences from that cluster represent two known molybdenum ion transporters—MOT 1 (known as sulfate transporter 5;2) from *A. thaliana* (Tomatsu et al. 2007) and the *Chlamydomonas reinhardtii* molybdate transporter, MOT (Tejada-Jimenez et al. 2007). Both transporters, Arabidopsis MOT 1 and *T. pseudonana* ST1 do not have STAS domains. Molybdenum is an essential component of nitrate assimilatory reduction. The molybdate anion is nearly identical to the sulfate anion in terms of charge distribution, effective size (230 vs. 240 pm), structure and stereochemistry (Cole et al. 1986). It has been shown that molybdate is transported into cells of *T. pseudonana* and sulfate acts as a competitive inhibitor of the uptake (Cole et al. 1986). The sequence homology makes it likely that ST 1 might function in *T. pseudonana* as molybdate transporter.

Sulfate transporters 3 and 4 (Fig. 1, cluster C) are very similar to each other and are found in one cluster together with another diatomal protein, a protein from *Phytophthora sojae,* and, more distantly, with fungal transporters. In contrast to ST 1 and ST2, sulfate transporters 3, 4 and 5 posses the STAS domain (sulfate transporter/anti-sigma factor antagonist). Mutational analyses of the STAS domain in Arabidopsis support hypothesis that this domain is involved in protein–protein interactions that might control sulfate transport (Rouached et al. 2005). Apart from having a STAS domain, the C-terminal fragments of ST 3

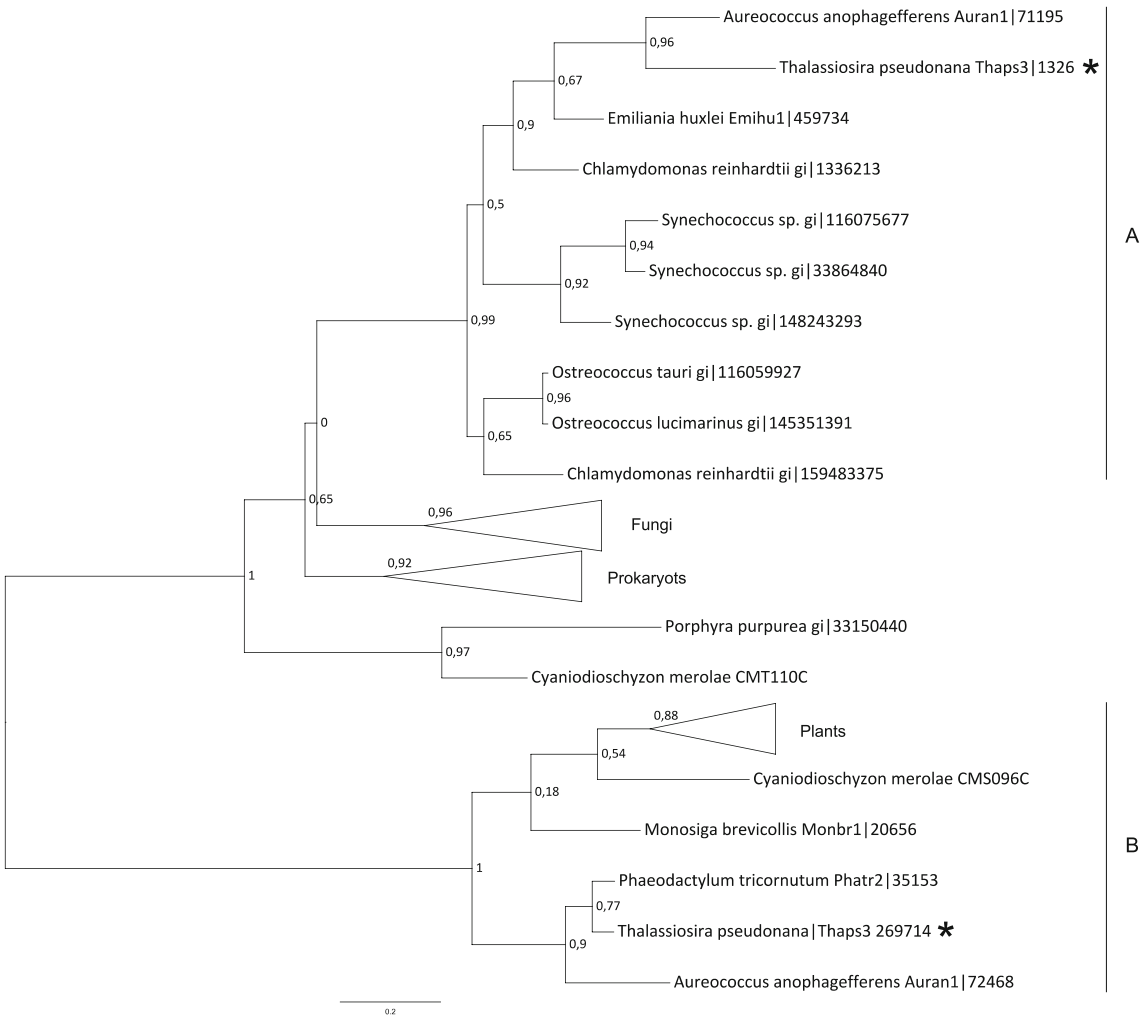

Fig. 2 Phylogenetic tree of ATP sulfurylase proteins. The phylogenetic tree was reconstructed using the maximum likelihood method implemented in the PhyML program. *T. pseudonana* proteins are marked with *asterisks*. Reliability for internal branch was assessed using the aLRT test and is presented on the graph. *Number next to species name* represents entry code of given protein in a databank

and ST 4 are structurally homologous with that of the catabolite gene-activator protein (CAP), a soluble *Escherichia coli* transcription factor. A similar domain is found also in cAMP-activated protein kinases and in eukaryotic cyclic nucleotide-gated ion channels, membrane proteins activated by cytoplasmic cGMP or cAMP (Li and Lester 1999). This domain might function as regulatory element of the sulfate transporter by connecting cellular cyclic nucleotide levels with sulfur metabolism.

The sulfate transporter/permease 5 is found in cluster A (Fig. 1, cluster A). It shows close relationship with a sequence found in genomes of marine unicellular organisms: *Aureococcus anophagefferens. Cyanidioschyzon merolae, Ostreococcus* sp. and *Monosiga brevicollis.* Additionally, cluster A contains a sequence from another heterokont, *P. sojae*. The sulfate transporter 2 (Fig. 1) is found separated from the other sequences and does not show close relationship with analyzed proteins.

Phylogenetic analysis of enzymes of the sulfate reduction pathway: ATP sulfurylase, APS reductase, and sulfite reductase

The sulfate reduction pathway is initiated by coupling sulfate with ATP to form adenosine-5′-phosphosulfate (APS). This reaction is catalysed by the enzyme ATP sulfurylase (ATPS). Plant ATPS forms a homotetramer of 52–54 kDa (Murillo and Leustek 1995) while bacterial ATPS consists of four heterodimers composed from 35 to 53 kDa CysN subunits (Leyh et al. 1988). The reaction generates inorganic pyrophosphate that has to be removed by a respective phosphatase to promote the forward reaction. In fungi, ATPS and APS kinase are fused into a single 59–64 kDa subunit that assemble into a homohexamer (MacRae et al. 2001). But also other combinations of ATPS fusion proteins could be identified such as the fusion with APS reductase in *Heterocapsa triquetra* or with an

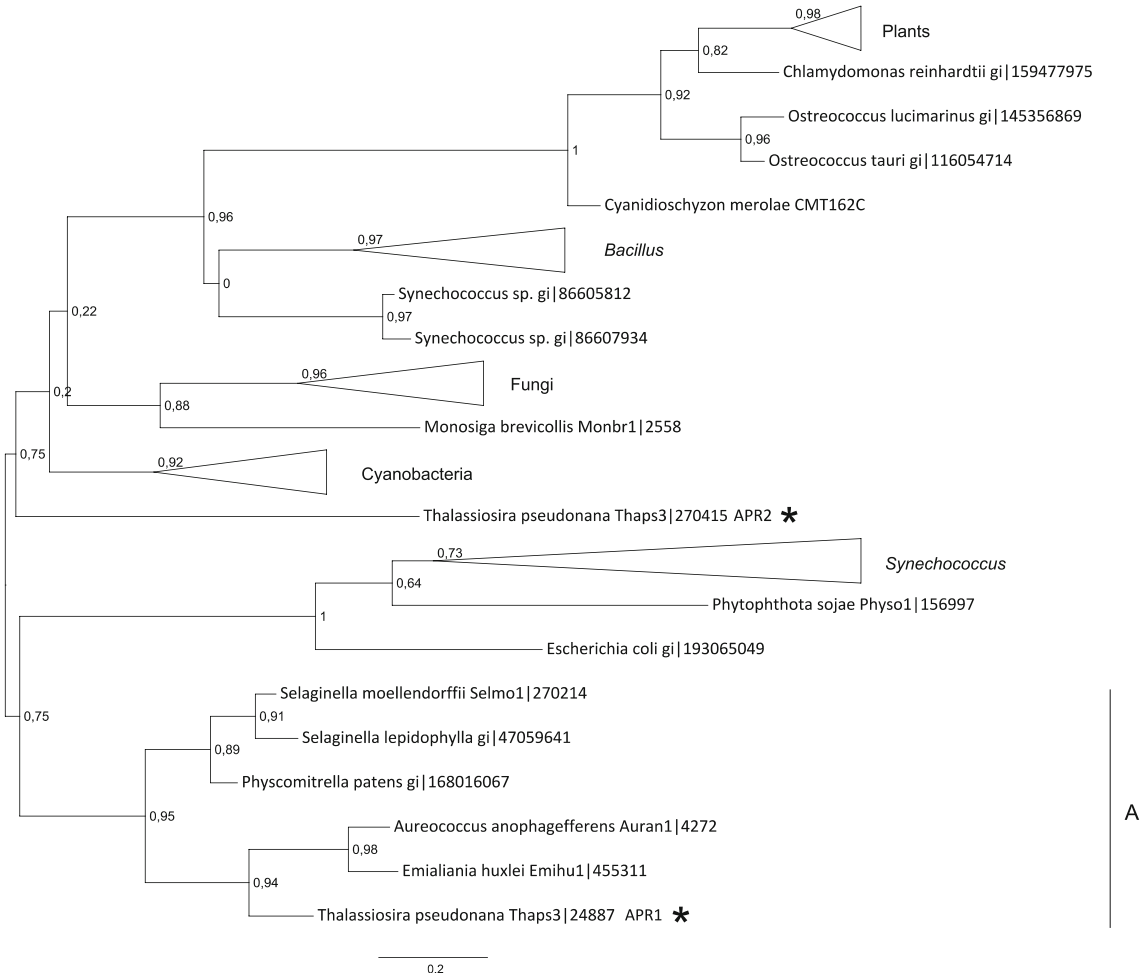

Fig. 3 Phylogenetic tree of APS reductase proteins. The phylogenetic tree was reconstructed using the maximum likelihood method implemented in the PhyML program. *T. pseudonana* proteins are marked with *asterisks*. Reliability for internal branch was assessed using the aLRT test and is presented on the graph. *Number next to species name* represents entry code of given protein in a databank

inorganic pyrophosphatase e.g. in *Emiliania huxleyi* (Patron et al. 2008; Bradley et al. 2009). The genome of *T. pseudonana* contains two sequences that could be annotated as ATP sulfurylase. One ORF encodes for the single enzyme (ATPS 2; Thaps3:1326) and the second isoform encodes a composite enzyme consisting of ATPS fused to APS kinase and pyrophosphatase (named sulfurylase-kinase-pyrophosphatase, SKP; Thaps3:269714). The sequence of the ATPS 2 enzyme clusters together with ATP sulfurylase sequences from *Emiliania huxleyi, A. anophagefferens, and C. reinhardtii*. These sequences contribute to the formation of a well-separated cluster consisting of eukaryotic algal and cyanobacterial enzyme sequences (Fig. 2, cluster A). The ATPS2 isoform contains no N-terminal sequences and thus being predicted to be located to the cytosol (Table 1).

The composite enzyme SKP from *T. pseudonana* clusters together with homologous genes from *P. tricornutum* and *A. anophagefferens*. (Fig. 2, cluster B). Like *T. pseudonana*, genomes of both organisms contain the enzyme composed of three catalytic domains: ATP sulfurylase, APS kinase and inorganic pyrophosphatase. The pre-sequence analysis suggests the presence of a targeting sequence in the *T. pseudonana* SKP protein, which would direct the protein into plastids; hence, sulfate in plastids could be used to synthesize 3′-phospho-adenosine-5′-phosphosulfate (PAPS). The genome of *T. pseudonana* encodes an additional APS kinase that is not bound with any other enzyme (Table 1). PAPS, synthesized by APS kinases, is also a donor of sulfate groups for the synthesis of important sulfated compounds for algae such as sulfolipids (Harwood 2004) and sulfated carbohydrates (Hoagland et al. 1993).

Adenosine-5′-phosphosulfate or PAPS are reduced to sulfite by APS or PAPS reductase, respectively. APR exists as a homodimer in prokaryotes but it is fused to thioredoxin in plants and green algae (Kopriva et al. 2002; Carroll et al. 2005a). Most APR enzymes contain a FeS cluster; however, a variant of the enzyme from early branching streptophyte lineages (such as *Physcomitrella* and *Selaginella*) has been shown to catalyse APS reduction without the FeS chemistry (Weber et al. 2000; Kopriva and Koprivova 2004; Kopriva et al. 2001, 2007; Carroll et al. 2005b). PAPS reductase is similar to bacterial APR but does not bind the FeS cluster (Kopriva et al. 2002; Kopriva and Koprivova 2004). Sequences similar to PAPS reductases (or novel moss-like form of APR) are found in genomes of marine algae like *T. pseudonana*—both enzymes lack a conserved pair of cysteines in thioredoxin-like domains at their C-termini. In vitro experiments have shown that this

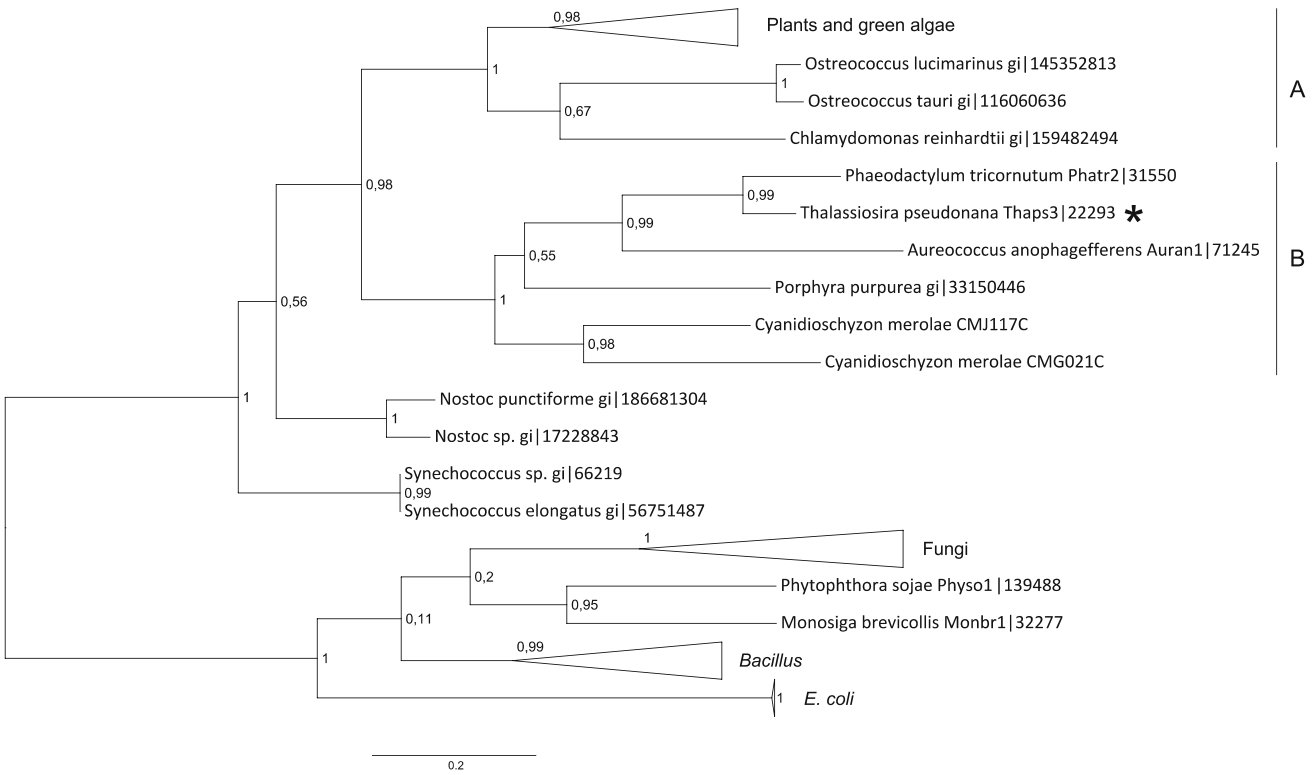

Fig. 4 Phylogenetic tree of sulfite reductase proteins. The phylogenetic tree was reconstructed using the maximum likelihood method implemented in the PhyML program. The *T. pseudonana* protein is marked with an *asterisk*. Reliability for internal branch was assessed using the aLRT test and is presented on the graph. *Number next to species name* represents entry code of given protein in a databank

organism is able to reduce APS with 100-fold higher activity than respective isoforms in plants (Gao et al. 2000; Kopriva et al. 2008). In our investigation, the phylogenetic tree was reconstructed using both APS and PAPS reductase sequences from other organisms. In the genome of *T. pseudonana* two sequences encoding putative adenosyl-5′-phosphosulfate reductase (APR) could be identified, while in the marine chlorophyte *Enteromorpha intestinalis* (Gao et al. 2000) and *C. reinhardtii* (Ravina et al. 2002) only one APR coding sequence was found. The APR 1 enzyme of *T. pseudonana* clusters together with the sequence from *E. huxleyi* and *A. anophagefferens* (Fig. 3, cluster A). These three sequences form a branch in the cluster which contains also sequences of APR from the moss *Physcomitrella patens* and the lycophyte *Selaginella moellendorffii.* The APR enzyme from *P. patens* was shown to be a representative of a novel class of APS reductases—APR-B. This enzyme shows similarity to PAPS reductase (characterized by the lack of C-terminal thioredoxin domain) but is able to reduce both APS and PAPS (Kopriva et al. 2007).

The other *T. pseudonana* APR isoform (APR2) is probably localized to plastids (Table 1) and is found at the basis of a clade, which contains sequences from many distinct organisms (Fig. 3).

Sulfite produced by *T. pseudonana* putative APR reductases is further reduced by the enzyme sulfite reductase (SiR). Plant SiR functions as monomer containing a siroheme and a FeS cluster and utilizes ferredoxin as source of electrons for the reduction (Krueger and Siegel 1982; Nakayama et al. 2000). In contrast, the bacterial NADPH-dependent SiR is an oligomer of eight flavoprotein subunits (CysJ), four 64 kDa siroheme and [4Fe-4S] cluster binding hemoproteins (CysI) (Crane et al. 1995). SiRs in fungi are composed from two different subunits requiring siroheme, FAD and FMN as co-factors (Kobayashi and Yoshimoto 1982). In the genome of *T. pseudonana* only one sequence of SiR was identified. The predicted *T. pseudonana* sulfite reductase protein clusters together with the *P. tricornutum* and *A. anophagefferens* SiR protein sequences forming a heterokont subcluster within cluster C (Fig. 4). Both diatomal enzymes cluster close to rhodophytes sequences. This is in agreement with the origin of plastids in diatoms, which were acquired through secondary endosymbiosis with an ancient eukaryotic red alga (Parker et al. 2008). SiR sequences from eukaryotic photosynthesizing organisms (Fig. 4, cluster A) and cyanobacteria form two clusters, which are well separated from the cluster of heterotrophic sulfate-assimilating organisms (Fig. 4). The sequence of SiR from *Thalassiosira pseudonana* contains a leader sequence that directs the protein to plastids of the diatom (Table 1).

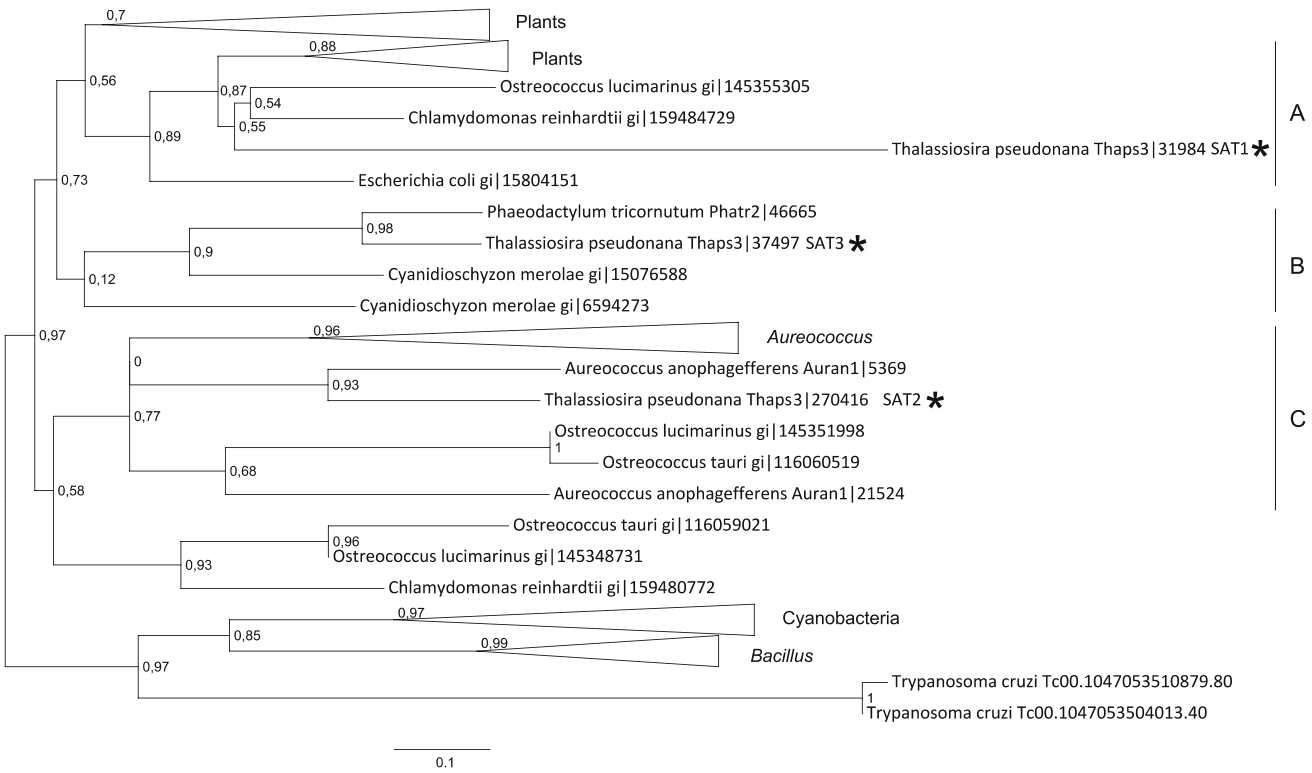

Fig. 5 Phylogenetic tree of serine acetyltransferase proteins. The phylogenetic tree was reconstructed using the maximum likelihood method implemented in the PhyML program. *T. pseudonana* proteins are marked with *asterisks*. Reliability for internal branch was assessed using the aLRT test and is presented on the graph. *Number next to species name* represents entry code of given protein in a databank

Phylogenetic analysis of the cysteine synthase complex: serine acetyltransferase and *O*-acetylserine(thiol)lyase

To form cysteine, sulfide is incorporated into *O*-acetylserine (OAS) in a reaction catalysed by the enzyme *O*-acetylserine(thiol)lyase (OASTL). The enzyme is regulated by serine acetyltransferase, an enzyme that synthesizes OAS by acetylation of serine. Both enzymes, serine acetyltransferase and *O*-acetylserine(thiol)lyase (OASTL), are necessary to build a functional cysteine synthase complex and their interaction enables an allosteric regulation of the cysteine formation (reviewed by Hell and Wirtz 2008). The cysteine synthase complex is composed of a SAT-hexamer and two OASTL dimers (Feldman-Salit et al. 2009). According to the model of function, under sufficient sulfide supply cysteine synthase complex is fully

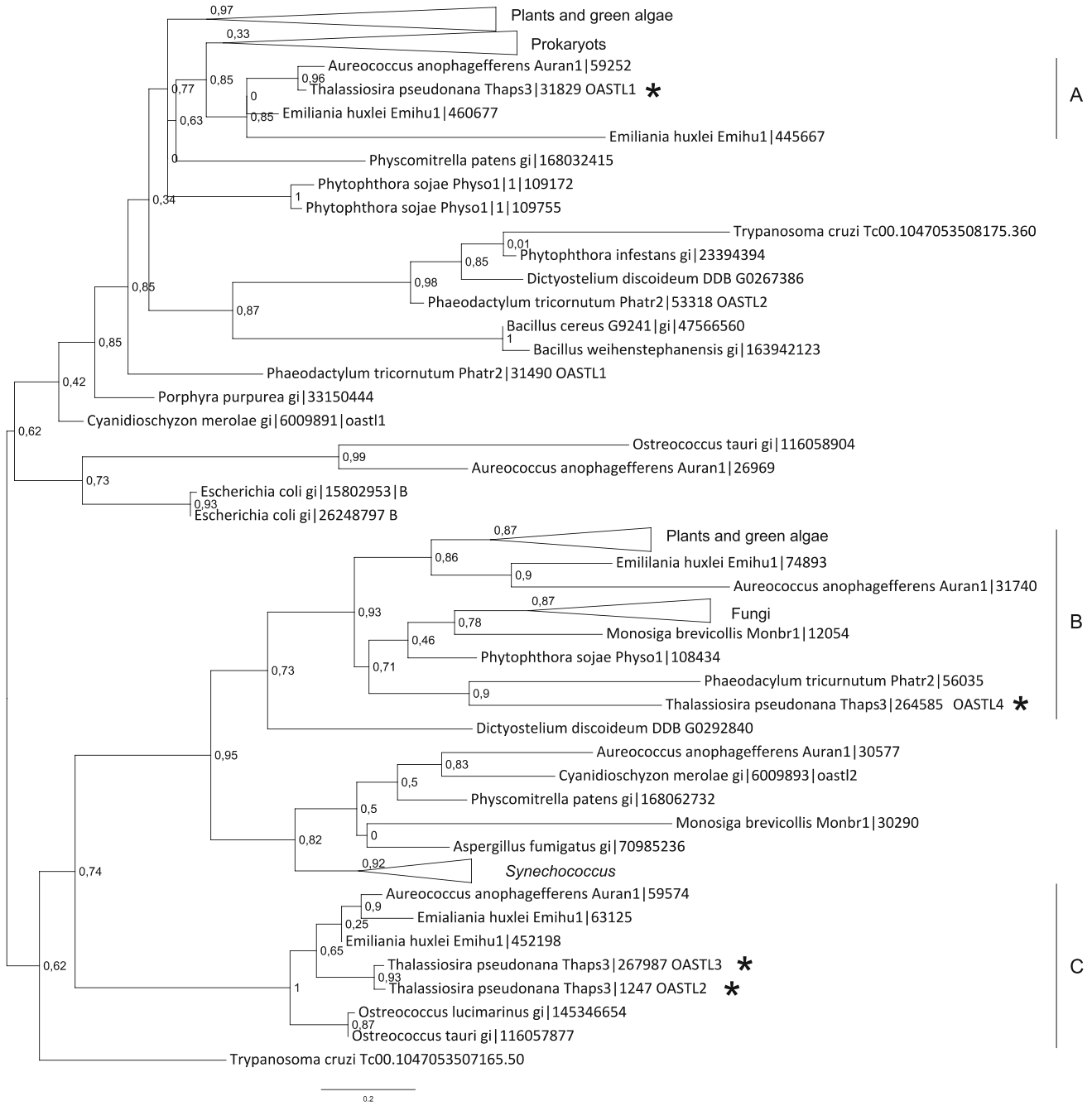

Fig. 6 Phylogenetic tree of OAS-(thiol)lyase proteins. The phylogenetic tree was reconstructed using the maximum likelihood method implemented in the PhyML program. *T. pseudonana* proteins are marked with *asterisks*. Reliability for internal branch was assessed using the aLRT test and is presented on the graph. *Number next to species name* represents entry code of given protein in a databank

associated, meaning that all available serine acetyltransferase is bound to OAS-(thiol)lyase but there is an excess of free and active OASTL dimers. The bound SAT-hexamer is active and produces *O*-acetylserine, while OASTL enzyme does not produce cysteine because SAT blocks the substrate binding site.

In the genome of *T. pseudonana*, three genes encoding serine acetyltransferase (SAT) and four OAS (thiol)lyase sequences (OASTL) were identified. The SAT 2 isoform is located in cluster C (Fig. 5) together with many serine acetyltransferase gene models of marine heterokont *A. anophagefferens* and prasinophytes from genus *Ostreococcus*. The SAT 1 sequence is found in one cluster (Fig. 5, cluster A) together with green algae and plant-derived sequences. The third SAT of *T. pseudonana* forms a cluster together with proteins from *P. tricornutum* and *C. merolae* (Fig. 5, cluster B). The phylogenetic relationship may suggest that this isoform is of red algae origin.

Two genes encoding OASTL (OASTL 2 and OASTL 3; Thaps3:1247 and Thaps3:267987, respectively) seem to be paralogous, as they show high similarity (Fig. 6, cluster C) and are found to be in reversed orientation separated by only 94 nucleotides. This is probably a result of relatively recent gene duplication. The cluster C (Fig. 6) contains sequences from other marine species *E. huxleyi, O. tauri, O. lucimarinus* and *A. anophagefferens*. Therefore, OASTL 2 and OASTL 3 together with other algal sequences might represent an algae-specific class of enzymes catalysing cysteine synthesis. The OASTL 1 isoform (Thaps3:31829) can be found clustering together with similar sequences from *E. huxleyi* and *A. anophagefferens* (Fig. 1, cluster A). Noteworthy, *T. pseudonana* OASTL isoforms 1, 2 and 3 show little similarity to OASTL sequences from *P. tricornutum*.

The fourth OASTL found in the genome of *T. pseudonana* shows low similarity to the other three sequences of OAS (thiol)lyases (Fig. 6, cluster B) but OASTL 4 contains, as the other enzymes, the characteristic conserved cysK domain. Both OASTL 4 and its homolog from *P. tricornutum* are members of a distinct branch, which neighbors a clade with sequences from choanoflagellate *M. brevicollis,* heterokonts of *Phytophthora* genus and fungi. The cluster B contains as well protein sequences from plants, mosses, green algae and the coccolithophorid *E. huxleyi*. The function of those proteins is unknown, although all of them contain the conserved cysK domain and PLP-binding site. OASTL 4 shows very little similarity to fungal *O*-acetylhomoserine(thiol)lyases or to cystathionine gamma-synthase-like proteins (not shown).

The localization of SAT and OASTL proteins cannot be precisely predicted (Table 1). From *Arabidopsis thaliana* it is known that both enzymes are encoded by five SAT and nine OASTL-like gene isoforms. These genes encode proteins localized to different compartments such as the cytosol, plastids and mitochondria. In *Arabidopsis,* the main SAT activity is localized to mitochondria, whereas OASTL activity is mainly localised to the cytosol. The main pool of OAS is located in the mitochondria and cytosol, whereas sulfide is present in high amounts in plastids and in the cytosol, to which it presumably diffuses or is transported from the chloroplasts (Haas et al. 2008; Heeg et al. 2008; Watanabe et al. 2008; Krueger et al. 2009). Therefore, we speculate that cysteine synthesis in *T. pseudonana* might be performed in three compartments.

Conclusion

The predictions of genes encoding enzymes of the sulfate uptake, assimilation and methionine synthesis were used to set up a model proposing the diatomal assimilatory pathways of sulfate reduction. In the model, sulfate is

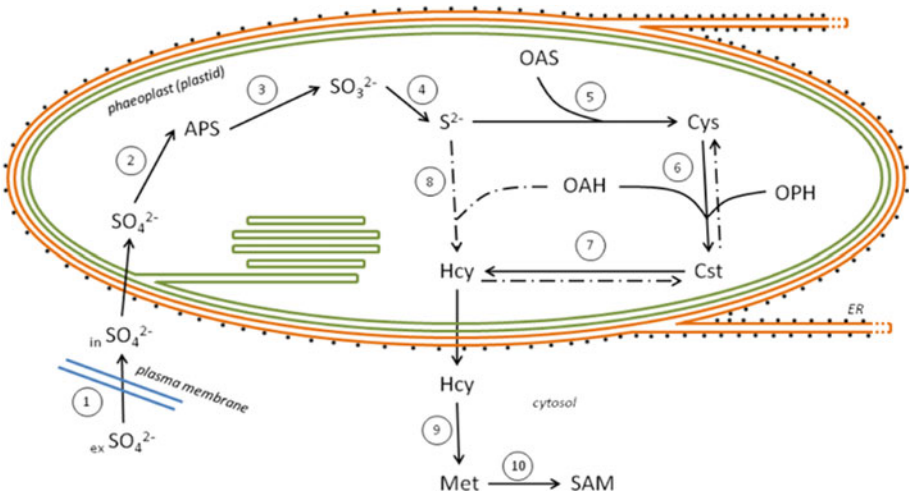

Fig. 7 The model of sulfate assimilation and methionine synthesis in *Thalassiosira pseudonana*. *Dashed lines* are used to mark alternative pathways. The numbers correspond to diatomal enzymes: (1) sulfate transporter/permease; (2) ATP sulfurylase; (3) APS reductase; (4) sulfite reductase; (5) OAS (thiol)lyase; (6) cystathionine γ-synthase and/or -lyase; (7) cystathionine β-lyase and/or –synthase; (8) O-acetylhomoserine (thiol)lyase; (9) methionine synthase; (10) SAM synthetase

transported into the diatomal cell via permease-like transporters (Fig. 7) which is subsequently transported into plastids. Sulfate is activated by adenylation performed by ATP sulfurylase (2AB). Whether the tripartite SKP enzyme catalyses phosphorylation of APS in plastids is not clear. APS is reduced by APS reductase, which contains thioredoxin domain at the C-terminus (3AB). Sulfite is then reduced by sulfite reductase (SiR) yielding sulfide (4AB). Cysteine (Cys) is synthesized by the co-operation of OASTL (5AB) and SAT, which delivers the other substrate, *O*-acetylserine. Sulfide can pass the membranes of plastids and therefore, to protect mitochondria from H_2S toxicity, cysteine is also synthesized in other compartments (e.g. Krueger et al. 2009). From this point onwards our models depict alternative pathways. In plastids, the sulfide moiety of cysteine reacts with a carbon backbone such as *O*-acetylhomoserine (OAH) or *O*-phosphohomoserine (OPH) as used in other organisms to finally form methionine. To prove the model, further biochemical studies have to be performed. Besides understanding the regulatory node of this pathway, the cross talk between different compartments with respect to exchange of metabolites such as pathway intermediates and synthesis of other S-containing compounds needs to be analyzed. Finally, although evidence is provided that in algae comparable sulfur metabolism occurs, the knowledge about the physiology is still unclear and need to be explored further.

Acknowledgments Prof. Dr. Lothar Willmitzer and Dr. Stefanie Hartmann are acknowledged for help and support in conducting the research. Financial support was provided by Max-Planck Society and Deutsche Forschungsgemeinschaft (HE 3088/4) and is greatly acknowledged.

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17):3389–3402

Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst Biol 55(4):539–552

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kroger N, Lau WW, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP, Rokhsar DS (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 306(5693):79–86

Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret F, Berges JA, Brownlee C, Cadoret JP, Chiovitti A, Choi CJ, Coesel S, De Martino A, Detter JC, Durkin C, Falciatore A, Fournet J, Haruta M, Huysman MJJ, Jenkins BD, Jiroutova K, Jorgensen RE, Joubert Y, Kaplan A, Kroger N, Kroth PG, La Roche J, Lindquist E, Lommer M, Martin-Jezequel V, Lopez PJ, Lucas S, Mangogna M, McGinnis K, Medlin LK, Montsant A, Oudot-Le Secq MP, Napoli C, Obornik M, Parker MS, Petit JL, Porcel BM, Poulsen N, Robison M, Rychlewski L, Rynearson TA, Schmutz J, Shapiro H, Siaut M, Stanley M, Sussman MR, Taylor AR, Vardi A, von Dassow P, Vyverman W, Willis A, Wyrwicz LS, Rokhsar DS, Weissenbach J, Armbrust EV, Green BR, Van De Peer Y, Grigoriev IV (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature 456(7219):239–244

Bradley M, Rest J, Li W-H, Schwartz N (2009) Sulfate activation enzymes: phylogeny and association with pyrophosphatase. J Mol Evol 68(1):1–13

Brunold C, Schiff JA (1976) Studies of sulfate utilization of algae: 15. enzymes of assimilatory sulfate reduction in *Euglena* and their cellular localization. Plant Physiol 57(3):430–436

Buchner P, Takahashi H, Hawkesford MJ (2004) Plant sulphate transporters: co-ordination of uptake, intracellular and long-distance transport. J Exp Bot 55(404):1765–1773

Carroll KS, Gao H, Chen HY, Leary JA, Bertozzi CR (2005a) Investigation of the iron-sulfur cluster in Mycobacterium tuberculosis APS reductase: implications for substrate binding and catalysis. Biochemistry 44(44):14647–14657. doi:10.1021/bi051344a

Carroll KS, Gao H, Chen HY, Stout CD, Leary JA, Bertozzi CR (2005b) A conserved mechanism for sulfonucleotide reduction. PLoS Biol 3(8):1418–1435

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17(4):540–552

Charlson RJ, Lovelock JE, Andreae MO, Warren SG (1987) Oceanic phytoplankton, atmospheric sulfur, cloud albedo and climate. Nature 326(6114):655–661

Cole J, Howarth R, Nolan S, Marino R (1986) Sulfate inhibition of molybdate assimilation by planktonic algae and bacteria: some implications for the aquatic nitrogen cycle. Biogeochemistry 2(2):179–196

Crane BR, Siegel LM, Getzoff ED (1995) Sulfite reductase structure at 1.6 A: evolution and catalysis for reduction of inorganic anions. Science 270(5233):59–67

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5):1792–1797

Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP SignalP and related tools. Nat Protoc 2(4):953–971

Falkowski PG, Barber RT, Smetacek V (1998) Biogeochemical controls and feedbacks on ocean primary production. Science 281(5374):200–206

Feldman-Salit A, Wirtz M, Hell R, Wade RC (2009) A mechanistic model of the cysteine synthase complex. J Mol Biol 386(1):37–59

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. Science 281(5374):237–240. doi:10.1126/science.281.5374.237

Foglino M, Borne F, Bally M, Ball G, Patte JC (1995) A direct sulfhydrylation pathway is used for methionine biosynthesis in *Pseudomonas aeruginosa*. Microbiology-Uk 141:431–439

Gao Y, Schofield OM, Leustek T (2000) Characterization of sulfate assimilation in marine algae focusing on the enzyme 5′-adenylylsulfate reductase. Plant Physiol 123(3):1087–1096

Gibbs SP (1979) The route of entry of cytoplasmically synthesized proteins into chloroplasts of algae possessing chloroplast ER. J Cell Sci 35(1):253–266

Gruber A, Vugrinec S, Hempel F, Gould SB, Maier UG, Kroth PG (2007) Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. Plant Mol Biol 64(5):519–530

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52(5):696–704

Haas FH, Heeg C, Queiroz R, Bauer A, Wirtz M, Hell R (2008) Mitochondrial serine acetyltransferase functions as a pacemaker of cysteine synthesis in plant cells. Plant Physiol 148(2):1055–1067. doi:10.1104/pp.108.125237

Harwood J (2004) Membrane lipids in algae. In: Lipids in photosynthesis: structure, function and genetics. pp 53–64

Heeg C, Kruse C, Jost R, Gutensohn M, Ruppert T, Wirtz M, Hell R (2008) Analysis of the Arabidopsis O-acetylserine(thiol)lyase gene family demonstrates compartment-specific differences in the regulation of cysteine synthesis. Plant Cell 20(1):168–185. doi:10.1105/tpc.107.056747

Hell R, Wirtz M (2008) Metabolism of cysteine in plants and phototrophic bacteria. In: sulfur metabolism in phototrophic organisms. pp 59–91

Hesse H, Hoefgen R (2008) Metabolism of methionine in plants and phototrophic bacteria. In: sulfur metabolism in phototrophic organisms. pp 93–110

Hesse H, Kreft O, Maimann S, Zeh M, Hoefgen R (2004) Current understanding of the regulation of methionine biosynthesis in plants. J Exp Bot 55(404):1799–1808

Hoagland KD, Rosowski JR, Gretz MR, Roemer SC (1993) Diatom extracellular polymeric substances: function, fine structure, chemistry, and physiology. J Phycol 29(5):537–566

Kerr DS, Flavin M (1970) Regulation of methionine synthesis and nature of cystathionine gamma-synthase in *Neurospora*. J Biol Chem 245(7):1842

Kertesz MA (2001) Bacterial transporters for sulfate and organosulfur compounds. Res Microbiol 152(3–4):279–290

Kilian O, Kroth PG (2005) Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. Plant J 41(2):175–183

Kobayashi K, Yoshimoto A (1982) Studies on yeast sulfite reductase. IV. Structure and steady-state kinetics. Biochim Biophys Acta 705(3):348–356

Kopriva S, Koprivova A (2004) Plant adenosine 5′-phosphosulphate reductase: the past, the present, and the future. J Exp Bot 55(404):1775–1783

Kopriva S, Buchert T, Fritz G, Suter M, Weber M, Benda R, Schaller J, Feller U, Schurmann P, Schunemann V, Trautwein AX, Kroneck PM, Brunold C (2001) Plant adenosine 5′-phosphosulfate reductase is a novel iron-sulfur protein. J Biol Chem 276(46):42881–42886

Kopriva S, Buchert T, Fritz G, Suter M, Benda R, Schunemann V, Koprivova A, Schurmann P, Trautwein AX, Kroneck PM, Brunold C (2002) The presence of an iron-sulfur cluster in adenosine 5′-phosphosulfate reductase separates organisms utilizing adenosine 5′-phosphosulfate and phosphoadenosine 5′-phosphosulfate for sulfate assimilation. J Biol Chem 277(24):21786–21791

Kopriva S, Fritzemeier K, Wiedemann G, Reski R (2007) The putative moss 3′-phosphoadenosine-5′-phosphosulfate reductase is a novel form of adenosine-5′-phosphosulfate reductase without an iron-sulfur cluster. J Biol Chem 282(31):22930–22938

Kopriva S, Patron NJ, Keeling P, Leustek T (2008) Phylogenetic analysis of sulfate assimilation and cysteine biosynthesis in phototrophic organisms. In: Sulfur Metabolism in Phototrophic Organisms. pp 31–58

Krueger R, Siegel L (1982) Spinach siroheme enzymes—isolation and characterization of ferredoxin sulfite reductase and comparison of properties with ferredoxin nitrite reductase. Biochemistry 21(12):2892–2904

Krueger S, Niehl A, Martin MC, Steinhauser D, Donath A, Hildebrandt T, Romero LC, Hoefgen R, Gotor C, Hesse H (2009) Analysis of cytosolic and plastidic serine acetyltransferase mutants and subcellular metabolite distributions suggests interplay of the cellular compartments for cysteine biosynthesis in Arabidopsis. Plant Cell Environ 32(4):349–367. doi:10.1111/j.1365-3040.2009.01928.x

Leyh TS, Taylor JC, Markham GD (1988) The sulfate activation locus of *Escherichia coli* K12: cloning, genetic, and enzymatic characterization. J Biol Chem 263(5):2409–2416

Li J, Lester HA (1999) Functional roles of aromatic residues in the ligand-binding domain of cyclic nucleotide-gated channels. Mol Pharmacol 55(5):873–882

MacRae IJ, Segel IH, Fisher AJ (2001) Crystal structure of ATP sulfurylase from *Penicillium chrysogenum*: insights into the allosteric regulation of sulfate assimilation. Biochemistry 40(23):6795–6804

Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWweese-Scott C, Geer LY, Gwadz M, He SQ, Hurwitz DI, Jackson JD, Ke ZX, Lanczycki CJ, Liebert CA, Liu CL, Lu F, Marchler GH, Mullokandov M, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Yamashita RA, Yin JJ, Zhang DC, Bryant SH (2005) CDD: a conserved domain database for protein classification. Nucleic Acids Res 33:D192–D196

Marzluf GA (1997) Molecular genetics of sulfur assimilation in filamentous fungi and yeast. Ann Rev Microbiol 51:73–96

Murillo M, Leustek T (1995) Adenosine-5′-triphosphate-sulfurylase from *Arabidopsis thaliana* and *Escherichia coli* are functionally equivalent but structurally and kinetically divergent: nucleotide sequence of two adenosine-5′-triphosphate-sulfurylase cDNAs from *Arabidopsis thaliana* and analysis of a recombinant enzyme. Arch Biochem Biophys 323(1):195–204

Nakayama M, Akashi T, Hase T (2000) Plant sulfite reductase: molecular structure, catalytic function and interaction with ferredoxin. J Inorg Biochem 82(1–4):27–32

Ostlund G, Schmitt T, Forslund K, Kostler T, Messina DN, Roopra S, Frings O, Sonnhammer EL (2010) In: Paranoid 7: new algorithms and tools for eukaryotic orthology analysis. Nucleic Acids Res 38 (Database issue):D196–203

Park SD, Lee JY, Kim Y, Kim JH, Lee HS (1998) Isolation and analysis of metA, a methionine biosynthetic gene encoding homoserine acetyltransferase in *Corynebacterium glutamicum*. Mol Cells 8(3):286–294

Parker MS, Mock T, Armbrust EV (2008) Genomic insights into marine microalgae. Annu Rev Genet 42:619–645

Patron N, Durnford D, Kopriva S (2008) Sulfate assimilation in eukaryotes: fusions, relocations and lateral transfers. BMC Evol Biol 8(1):39

Ravina CG, Chang CI, Tsakraklides GP, McDermott JP, Vega JM, Leustek T, Gotor C, Davies JP (2002) The sac mutants of *Chlamydomonas reinhardtii* reveal transcriptional and posttranscriptional control of cysteine biosynthesis. Plant Physiol 130(4):2076–2084

Rouached H, Berthomieu P, El Kassis E, Cathala N, Catherinot V, Labesse G, Davidian JC, Fourcroy P (2005) Structural and functional analysis of the C-terminal STAS (sulfate transporter and anti-sigma antagonist) domain of the Arabidopsis thaliana sulfate transporter SULTR1.2. J Biol Chem 280(16):15976–15983

Rowbury RJ, Woods DD (1964) O-succinylhomoserine as intermediate in synthesis of cystathionine by *Escherichia coli*. J Gen Microbiol 36(3):341–358

Tejada-Jimenez M, Llamas A, Sanz-Luque E, Galvan A, Fernandez E (2007) A high-affinity molybdate transporter in eukaryotes. Proc Natl Acad Sci USA 104(50):20126–20130

Tomatsu H, Takano J, Takahashi H, Watanabe-Takahashi A, Shibagaki N, Fujiwara T (2007) An *Arabidopsis thaliana* high-affinity molybdate transporter required for efficient uptake of molybdate from soil. Proc Natl Acad Sci USA 104(47):18807–18812. doi:10.1073/pnas.0706373104

Wachter A, Wolf S, Steininger H, Bogs J, Rausch T (2005) Differential targeting of GSH1 and GSH2 is achieved by multiple transcription initiation: implications for the compartmentation of glutathione biosynthesis in the Brassicaceae. Plant J 41(1):15–30

Watanabe M, Mochida K, Kato T, Tabata S, Yoshimoto N, Noji M, Saito K (2008) Comparative genomics and reverse genetics analysis reveal indispensable functions of the serine acetyltransferase gene family in *Arabidopsis*. Plant Cell 20(9):2484–2496

Weber M, Suter M, Brunold C, Kopriva S (2000) Sulfate assimilation in higher plants characterization of a stable intermediate in the adenosine 5′-phosphosulfate reductase reaction. European J Biochem/FEBS 267(12):3647–3653