

RESEARCH

Open Access



Phylogenetic analysis of methionine synthesis genes from *Thalassiosira pseudonana*

Mariusz A Bromke* and Holger Hesse

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. The sulfate assimilation and methionine synthesis pathways provide S-containing amino acids for the synthesis of proteins and a range of metabolites such as dimethylsulfoniopropionate. To obtain an insight into the localization and organization of the sulfur metabolism pathways we surveyed the genome of *Thalassiosira pseudonana*—a model organism for diatom research. We have identified and annotated genes for enzymes involved in respective pathways. Protein localization was predicted using similarities to known signal peptide motifs. We performed detailed phylogenetic analyses of enzymes involved in sulfate uptake/reduction and methionine metabolism. Moreover, we have found in up-stream sequences of studied diatoms methionine biosynthesis genes a conserved motif, which shows similarity to the Met31, a cis-motif regulating expression of methionine biosynthesis genes in yeast.

Keywords: Methionine synthesis, Diatoms, Phylogenetics

Background

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 CO₂ 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). Though, 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 CO₂ input to the atmosphere. Bacterial enzymes are mostly responsible of the formation of DMS from dimethylsulfoniopropionate (DMSP). The amino acid methionine is the precursor for the synthesis of DMSP and underlines the dependence on sulfur and nitrogen metabolism (Gage et al. 1997; Keller et al. 1999) of marine algae.

The carbon backbone of methionine derives from aspartate, which in two steps is metabolized to homoserine. In plants homoserine is activated for further processes in a reaction catalyzed by homoserine kinase (HSK), which phosphorylates the γ -hydroxyl group. In some bacteria activation of homoserine is done by *O*-succinylation catalyzed by the enzyme homoserine *O*-succinyltransferase (Zubieta et al. 2007). Phosphorylated or succinylated homoserine is condensed with cysteine by cystathionine γ -synthase to form cystathionine. Subsequently, cystathionine β -lyase hydrolyzes cystathionine to form homocysteine (Hesse and Hoefgen 2003). In fungi and bacteria *O*-acetylhomoserine and *O*-succinylhomoserine, respectively, can be directly sulfhydrylated to form homocysteine, the precursor of methionine

*Correspondence: bromke@mpimp-golm.mpg.de
Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1,
14476 Potsdam-Golm, Germany

(Fogliano et al. 1995; Kerr and Flavin 1970; Park et al. 1998; Rowbury and Woods 1964). In the final step of methionine synthesis, homocysteine is methylated by two non-homologous enzymes: the cobalamin-dependent methionine synthase (METH-type) in mammals, protists and most bacteria (Evans et al. 2004; Goulding and Matthews 1997; Krungkrai et al. 1989) or the cobalamin-independent methionine synthase (METE-type) as found in bacteria (including cyanobacteria), green plants and fungi (Eichel et al. 1995; Kacprzak et al. 2003; Whitfield et al. 1970; Zeh et al. 2002). Many marine algae require external cobalamin supply to maintain methionine synthesis, though the cobalamin auxotrophy is not restricted to specific phyla (Croft et al. 2005). This suggests multiple events of METE-type enzyme loss in the evolution of algae. Methionine not only is a proteogenic amino acid but served a substrate for synthesis of many sulfur-containing metabolites, such as a very important cofactor in methylation reactions, S-adenosylmethionine (SAM) (McQueney et al. 2000).

Research on diatoms advanced significantly with publication of the whole genome sequences of *Thalassiosira pseudonana* (Armbrust et al. 2004) and *Phaeodactylum tricornerutum* (Bowler et al. 2008). Diatoms are amazing organisms derived from a non-photosynthetic eukaryote that domesticated a photoautotrophic eukaryotic cell phylogenetically close to a red alga (Parker et al. 2008). This resulted in an extensive gene transfer events 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. There is still little knowledge on the sulfur uptake and assimilation (Bromke and Hesse 2013; Kopriva et al. 2008). To shed more light on the biosynthesis of methionine in the model diatom *T. pseudonana*, we have performed a detailed phylogenetic analysis of each enzyme from the methionine biosynthesis pathway in this organism. In addition, we have addressed a subcellular localization question in order to present possible models of the pathway.

The compartmentation of sulfate assimilation and methionine synthesis 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), the same as the synthesis of cystathionine and homocysteine. *Arabidopsis* contains three functional isoforms of the cobalamin-independent (CIMS) enzyme. One of them is located in chloroplasts and is devoted to the de novo synthesis of methionine. The two other methionine synthase isoforms in *Arabidopsis* are located in the cytosol and most probably are responsible for the regeneration of methionine by re-methylation of homocysteine (Raveland et al. 2004).

To shed more light on the structure of methionine synthesis genes in *T. pseudonana* as well as to put them in the evolutionary context, we have performed a phylogenetic analysis on eight sequences. These sequences and additional enzyme sequences from homoserine and glutathione synthesis pathways were analyzed to investigate their subcellular localization.

Results and discussion

Identification of methionine biosynthesis pathway proteins in *Thalassiosira pseudonana*

Methionine belongs to the aspartate-family of amino acids. Hence, we have taken for analysis not only enzymes of sulfur transfer and methionine synthesis, but also enzymes responsible for the synthesis of methionine's carbon backbone from aspartic acid. In this study plant, bacterial and fungal genes known to be involved in the methionine biosynthesis pathway were used as queries to identify homologous genes in the genome of *T. pseudonana*. The sequences were analyzed to predict subcellular localization of the proteins. Additionally to the methionine synthesis genes, we have analyzed the localization of two enzymes of the glutathione synthesis pathway. Identified genes with predicted localization of their products are presented in Table 1, which gives additional information about used abbreviations and database entry numbers (<http://genome.jgi-psf.org/Thaps3/Thaps3.home.html>) as well.

A Maximum Likelihood algorithm implemented in the PhyML software was used to investigate phylogenetic relationships between *T. pseudonana* protein sequences and their eukaryotic and prokaryotic homologues. A limited species number as representatives of different phyla was selected (listed in Methods). For generation of phylogenetic trees homoserine acetyltransferase, cystathionine γ/β -synthase-like proteins, methionine synthase, and S-adenosylmethionine synthetase orthologues were used. In case of pyridoxal phosphate-dependent enzymes (cystathionine metabolism enzymes) the respective phylogenetic tree was used to predict enzymatic activity of analyzed proteins.

Prediction of intracellular localization

On the basis of the obtained amino acid sequences, the subcellular localization of proteins involved in sulfur

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

| Protein | Protein ID | Length | Loc | RC | Mito | Secr | Other | TP length | Cleavage site | CD start |
|---------------------------------------|------------|--------|-----|----|-------|-------|-------|-----------|---------------|----------|
| Aspartate kinase | 29008 | 936 | S | 5 | 0.437 | 0.54 | 0.034 | 19 | VTS-FAP | 90 |
| Aspartate kinase | 269371 | 619 | S | 1 | 0.033 | 0.887 | 0.078 | 26 | IGA-FSP | 144 |
| Aspartate semialdehyde dehydrogenase | 32577 | 345 | – | 2 | 0.074 | 0.121 | 0.859 | – | | 6 |
| Homoserine dehydrogenase | 40655 | 435 | – | 1 | 0.048 | 0.096 | 0.926 | – | | 5 |
| Homoserine dehydrogenase | 24587 | 502 | – | 4 | 0.271 | 0.15 | 0.552 | – | | 128 |
| Homoserine dehydrogenase | 29144 | 433 | – | 3 | 0.249 | 0.045 | 0.775 | – | | 1 |
| Homoserine acetyltransferase | 12141 | 375 | – | 1 | 0.081 | 0.062 | 0.938 | – | | 65 |
| Homoserine kinase | 28592 | 358 | M | 5 | 0.439 | 0.082 | 0.412 | 36 | VPA-SSA | 28 |
| Cystathionine β -lyase-like | 269520 | 456 | M | 5 | 0.49 | 0.065 | 0.459 | 66 | SAH-DGS | 1 |
| Cystathionine γ -synthase-like | 9830 | 381 | – | 1 | 0.037 | 0.114 | 0.95 | – | | 61 |
| Methionine γ -lyase-like | 11993 | 487 | M | 3 | 0.765 | 0.025 | 0.257 | 39 | ENK-HAT | 45 |
| Methionine synthase | 693 | 1,359 | – | 2 | 0.123 | 0.066 | 0.878 | – | | 85 |
| SAM synthetase | 21815 | 465 | – | 2 | 0.098 | 0.076 | 0.886 | – | | 4 |
| SAM synthetase | 39946 | 450 | – | 3 | 0.247 | 0.059 | 0.741 | – | | 65 |
| Glutamylcysteine synthetase | 13064 | 636 | M | 5 | 0.57 | 0.053 | 0.445 | 36 | NRV-KDL | 231 |
| Glutathione synthetase | 29212 | 533 | – | 3 | 0.044 | 0.298 | 0.826 | – | | 6 |

The first two columns "ProteinID" and "Length" contain entry numbers of protein models and its amino acids length. In the "Loc" column is the predicted localization: S secreted, M mitochondrial, _ any other localization. RC stands for Reliability Class of the prediction (scale: 1 to 5, where 1 is the best). Columns titled "mito", "secr", "other" contain final NN scores on which the prediction is based for mitochondrial, secretion and other localization; respectively. The "TP length" and "cleavage site" columns present predicted presequence lengths and the respective cleavage site motif. The "CD Start" column contains lengths of N-terminal sequences preceding Conserved Catalytical Domains of respective proteins.

assimilation and methionine synthesis was predicted. In the C-terminal parts of analyzed proteins no motifs responsible for retention in the endoplasmic reticulum (KDEL, DEEL or DDEL) were found. Location assignment was performed with TargetP software and it is based on the predicted presence of N-terminal mitochondrial targeting peptide (mTP) or secretory pathway signal peptide (SP). Additionally, the protein sequences were analyzed for any N-terminal sequences longer than 15 amino acids in front of conserved catalytic domains. These pre-sequences might form specific diatomal targeting peptides, which are not recognized by the used software. The results of the prediction are presented in Table 1. The prediction reliability values are a scale from 1 to 5, where 1 indicates the strongest prediction.

Out of analyzed sequences six show a pre-sequence that could be used by the cellular machinery to traffic the proteins to the respective compartment. In two aspartate kinase genes we have detected sequences which according to the used software, would direct these proteins into secretion via endoplasmic reticulum. Due to the evolutionary history of diatoms, the cells of these organisms contain plastids surrounded by additional two layers of plasma membrane, from which the outermost membrane is continuous with the endoplasmic reticulum (Gibbs 1979). The result suggests that these enzymes are directed to the plastid. In fact, we have detected in both aspartate kinase sequences N-terminal

bipartite pre-sequences, which consist of a signal and a transit peptide-like domain responsible for directing the nuclear-encoded protein into diatomal chloroplast. As demonstrated experimentally by means of a heterogeneous protein expression system in diatom *P. tricornutum*, presence of a conserved amino acid motif at the SP cleavage site (ASA-FAP) is associated with plastid-localized proteins (Gruber et al. 2007; Kilian and Kroth 2005). The aspartate kinase sequences do not have the exact ASA-FAP sequence, but contain phenylalanine in the 4th position of the predicted cleavage 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). Phenylalanine and valine were found in pre-sequences of two genes of the sulfur reduction pathway in diatoms (Bromke and Hesse 2013). One should note that this process in phototrophs takes place exclusively in plastids.

There are four sequences that are predicted to be directed into mitochondria: cystathionine β -lyase-like, homoserine kinase, γ -glutamylcysteine synthetase and methionine γ -lyase-like (Table 1). The prediction reliability values are as low as 5, with an exception for methionine γ -lyase-like, in which case the prediction reaches the class 3. In plants reactions of homoserine phosphorylation by homoserine kinase as well as the downstream

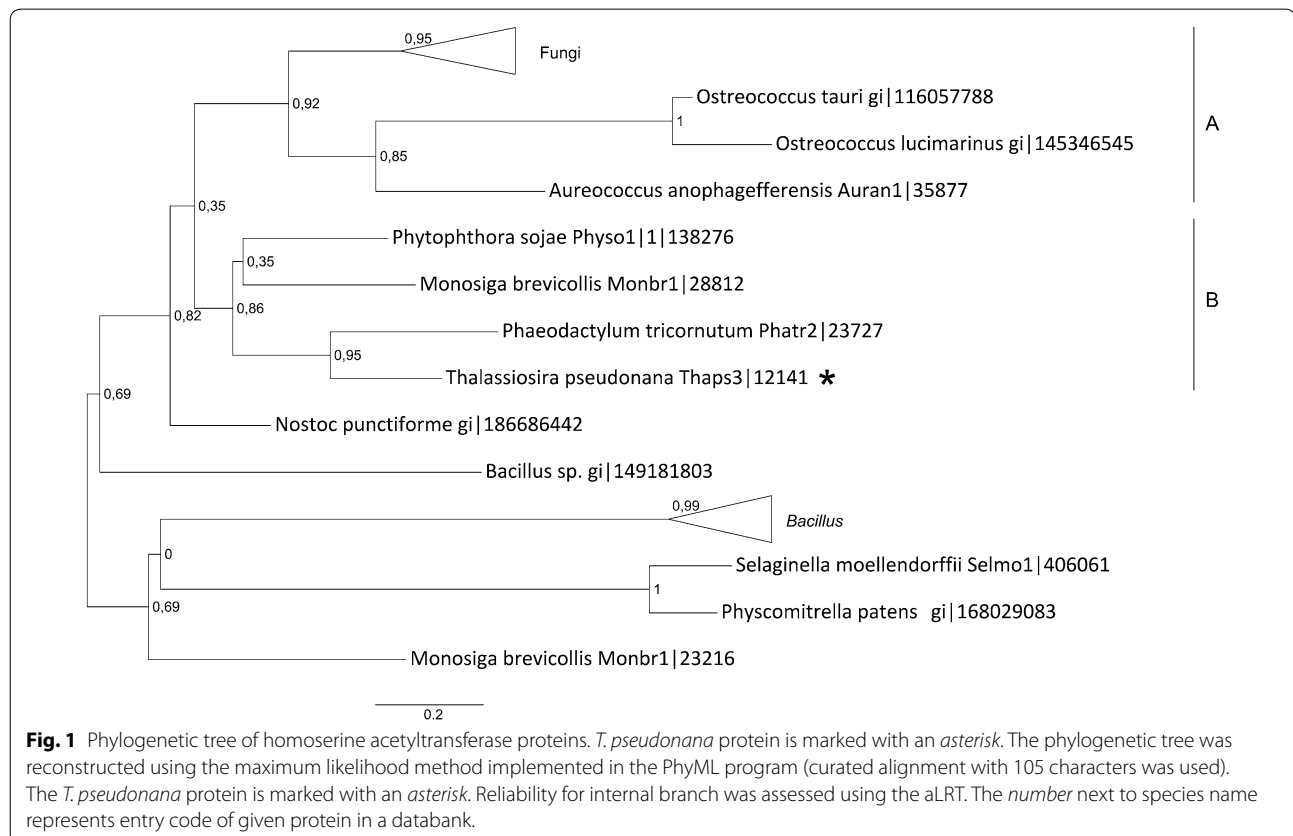
synthesis of cystathionine and homocysteine are localized to plastids (Hesse and Hoefgen 2008), whereas these processes take place in the cytosol of fungi (Marzluf 1997). Moreover, methionine- γ -lyase in *Arabidopsis thaliana* is a cytosolic enzyme (Rebeille et al. 2006). Thus, the prediction of localization of these proteins seems not to be accurate. In animals and fungi γ -glutamylcysteine synthetase and glutathione synthetase gene products are localized to the cytosol (Griffith 1999; Pócsi et al. 2004). In plants, due to the transcript heterogeneity, the synthesis of the glutathione can take place in chloroplast and in cytosol depending on the length of the respective gene transcript (Wachter et al. 2005). Database searches revealed the presence of putative TP sequences for maize γ -glutamylcysteine synthetase and glutathione synthetase proteins (Wachter et al. 2005), despite absence of plastidic TP in cDNA clones reported by Gómez et al. (2004). There is only one sequence of γ -glutamylcysteine synthetase in the genome of *T. pseudonana*, but it is incomplete as it lacks a starting ATG codon. The gene model of γ -glutamylcysteine synthetase shows high similarity to heterokonts and less to animal sequences. Taking this into consideration, we suggest that the enzyme is localized to the cytosol although TargetP predicts a mitochondrial localization (weak, class 5 prediction). On the other

hand, the analysis of the identified glutathione synthetase sequence resulted in a strong score suggesting its cytosolic location of the enzyme. Thus, the glutathione synthesis takes place in the cytosol and is subsequently transported to other cell compartments for biochemical functions.

Analysis of phylogenetic relationship of *Thalassiosira pseudonana* genes

Phylogenetic analysis of enzymes involved in the transacetylation of homoserine

An interesting feature of *T. pseudonana* genome is that it contains next to the identified homoserine kinase a gene encoding a protein with high similarity to homoserine acetyltransferase (HAT; Thaps3: 12141). This is interesting as finding as in some bacteria (Park et al. 1998) and fungi (Kerr and Flavin 1970) the activation of homoserine depends on transacetylation—formation of *O*-acetylhomoserine using acetyl-CoA as substrate. According to the prediction (Table 1), HAT in cells of *T. pseudonana* is cytosolic localized. The phylogenetic tree inferred for the enzyme shows that the putative homoserine acetyltransferase of *T. pseudonana* can be found in one cluster (Fig. 1, cluster B) together with predicted protein sequences from diatom *P. tricornutum*, oomycete *Phytophthora sojae* (both heterokonts) and a

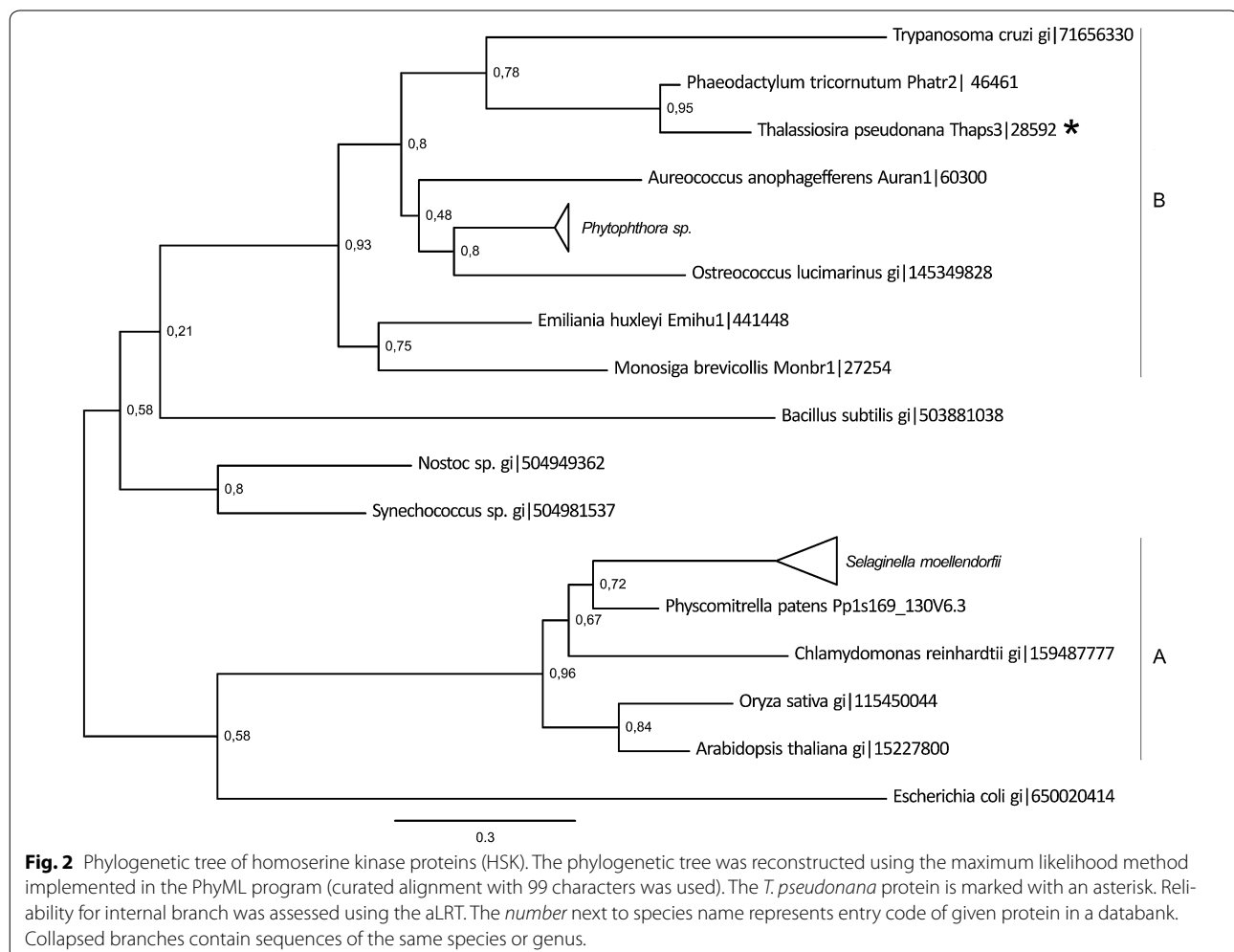


choanoflagellate *Monosiga brevicollis* (a metazoan). The exact topology of this cluster is not resolved. Although, a close relationship between diatomal proteins and the HAT from *M. brevicollis* was reported previously (Sun et al. 2010). These also seem to be related with homoserine acetyltransferases from fungi (Fig. 1, cluster A), in which the direct sulphydrylation of *O*-acetylhomoserine was experimentally confirmed (Wiebers and Garner 1967). *O*-acetylhomoserine (OAH) is also used by bacterial and cyanobacterial cells to synthesize homocysteine in the process of sulphydrylation, in which sulfide is bound with OAH yielding homocysteine (Hwang et al. 2002; Kerr 1971). It is likely that *T. pseudonana* is able to perform a similar reaction in order to synthesize homocysteine from *O*-acetylhomoserine. If this is the case, *O*-phosphohomoserine (OPH) would function mainly as substrate for threonine synthesis as has been shown for bacteria and fungi (Chassagnole et al. 2001; Farfan and Calderon 2000) and not, as in plants, for the synthesis of cystathionine (Curien et al. 1996, 1998; Zeh et al. 2001). This hypothesis has to be verified experimentally.

Phylogenetic analysis of enzymes involved in the phosphorylation of homoserine

The activation of homoserine for synthesis of homocysteine can be done by the phosphorylation of the gamma-hydroxyl group as well. This reaction is catalyzed by the special enzyme homoserine kinase (HSK). The *T. pseudonana* genome contains only one copy of the HSK (Thaps3: 28592). In the genome of *Cyanidioschyzon merolae*, the red algae model organisms, no homoserine kinase gene model could be found. Nevertheless, the genome of *C. merolae* contains genes annotated as aspartate kinase/homoserine dehydrogenase as well as threonine synthase—enzymes producing substrate for HSK and using *O*-phosphohomoserine (OPH). The synthesis of OPH is the only way of homoserine activation in plants and green algae (Hesse and Hoefgen 2003).

This fact is visible on the phylogenetic tree generated from HSK sequences. All plant proteins cluster together (Fig. 2a). The HSK from *T. pseudonana* is found in one cluster with other diatomal sequence and the one from *Trypanosoma cruzi* (Fig. 2b). This observation might



suggest that the HSK in diatoms is evolutionary distant from the plant isoform and thus the OPH and threonine biosynthesis in diatoms is probably differently regulated. Unlike in plants where OPH is the branching point between biosynthesis pathways of methionine and threonine, the *T. pseudonana* HSK might be delivering the substrate for Thr synthesis and is competing for homoserine with HAT. Other sequences from oomycetes and heterokonts can be found in the cluster B (Fig. 2), though its final topology could not be defined.

Phylogenetic analysis of enzymes catalyzing the formation of methionine precursors

Due to their high similarity it is impossible to distinguish cystathionine γ/β -synthases from cystathionine β/γ -lyases solely on basis of their sequences. Both types of enzymes utilize pyridoxal 5'-phosphate (PLP) as cofactor in reactions of cystathionine synthesis and/or lysis. A similar catalytic domain can be found in methionine γ -lyase (MgL), also a PLP-dependent enzyme. MgL is a promiscuous enzyme, which catalyses a number of α/γ - and α/β -elimination reactions on methionine and cysteine derivatives. It also catalyses the γ -replacement reaction of the thiomethyl group of substrates such as methionine, homocysteine and ethionine to yield the corresponding S-substituted homocysteine derivative (Lockwood and Coombs 1991). The enzyme has been found and characterized in various bacteria (Dias and Weimer 1998; Faleev et al. 1996; Inoue et al. 1995; Manukhov et al. 2005), protozoans like *Trichomonas vaginalis* (McKie et al. 1998), and *Entamoeba histolytica* (Tokoro et al. 2003), and in *Arabidopsis* (Rebeille et al. 2006). The predicted plant protein has 28–32% identity (48% similarity) with methionine γ -lyase from *Pseudomonas putida* and *E. histolytica*, respectively (Rebeille et al. 2006).

Therefore, three sequences found in the genome of *T. pseudonana*—cystathionine synthase-like (CgS-like; Thaps3: 9830), cystathionine lyase-like (CbL-like; Thaps3: 269520) and methionine γ -lyase-like (MgL-like; Thaps3: 11993)—were analyzed together (Fig. 3). The cystathionine synthase-like protein clusters together with proteins from the marine algae coccolithophorid *Emiliania huxleyi*, the harmful algal blooms-causing *Aureococcus anophagefferens* and the parasitic eukaryote *T. cruzi* (Fig. 3). Protozoa, such as *Trypanosoma*, synthesize cysteine from methionine (Walker and Barrett 1997), which suggests that this diatomal enzyme might catalyze the same reaction and therefore could be described as putative cystathionine γ -lyase. Still, without biochemical characterization, there are no means to correctly predict whether the enzyme functions as cystathionine synthase or lyase. The methionine γ -lyase-like protein (MgL-like) and the putative cystathionine

lyase (CbL-like) from *T. pseudonana* are found in one cluster together with proteins from marine eukaryotic algae, *Escherichia coli* and yeast (Fig. 3, cluster A). The whole clade is very weakly supported (aLRT = 0.26). Little is known about the catalytic character of the enzymes from photosynthesizing algae. One can assume that green algae *Ostreococcus* catalyze synthesis of homocysteine from cystathionine in the plant manner (other sequences of *Ostreococcus* proteins are found neighboring with plant enzymes suggesting conservation) and thus, the MgL-like enzyme from *T. pseudonana* (Thaps3: 11993) would catalyze the same reaction. The third analyzed PLP-dependent enzyme from the diatom (Thaps3: 269520) clusters together with a sequence from *E. huxleyi*. Its function remains unknown. Noteworthy, the analyzed diatomal proteins do not form common clusters with plant-derived sequences (Fig. 3).

Phylogenetic analysis of methionine synthase and SAM synthetases

Methionine synthase catalyzes the reaction of transmethylation, where a methyl group is transferred from 5'-methyltetrahydropteroyl glutamates onto the sulfur atom of homocysteine. Methylation of homocysteine can be catalyzed by two non-homologous enzymes: a cobalamin-dependent (METH) and cobalamin-independent methionine synthase (METE). Vitamin B₁₂ (cobalamin) is not needed for the enzymatic reaction catalyzed by enzymes from some bacteria, plants and fungi (Eichel et al. 1995; Kacprzak et al. 2003; Whitfield et al. 1970; Zeh et al. 2002). Cobalamin-dependent methionine synthase is found in mammals, protists and most bacteria (Evans et al. 2004; Goulding and Matthews 1997; Krungkrai et al. 1989). Both enzymes share no sequence similarity. In the genome of *T. pseudonana* only one sequence was found which could be assigned as encoding a methionine synthase enzyme. Analysis of the sequence showed similarity to the eukaryotic METH-type of the enzyme and thus is most likely localized to the cytosol. This is also supported by the TargetP analysis (Table 1). The phylogenetic analysis of cobalamin-dependent methionine synthase proteins shows two main clusters (Fig. 4). The smaller cluster contains cyanobacterial methionine synthases and sequences of proteins from two *Bacillus* species. The other cluster contains sequences of enzymes from diverse organisms—eukaryotes (Fig. 4, clusters A and B) with one prokaryotic exception—MethH from *E. coli*. Methionine synthase from *T. pseudonana* is found in clade A (Fig. 4) with homologues from *P. tricornutum*, *A. anophagefferens* and *Phytophthora sojae*. These four organisms share common ancestor (Parker et al. 2008). Noteworthy, in the genome of other diatom *P. tricornutum* a

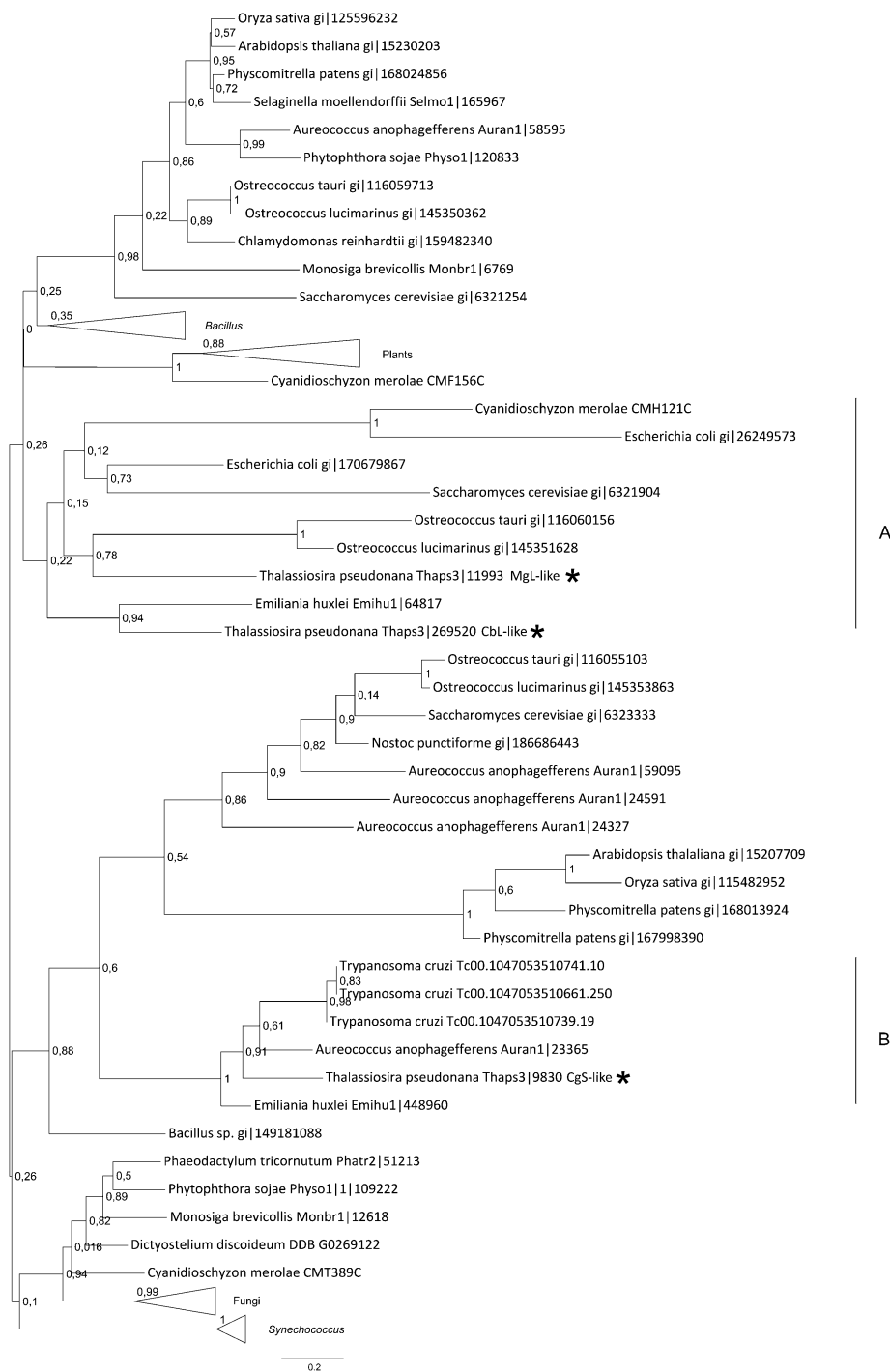
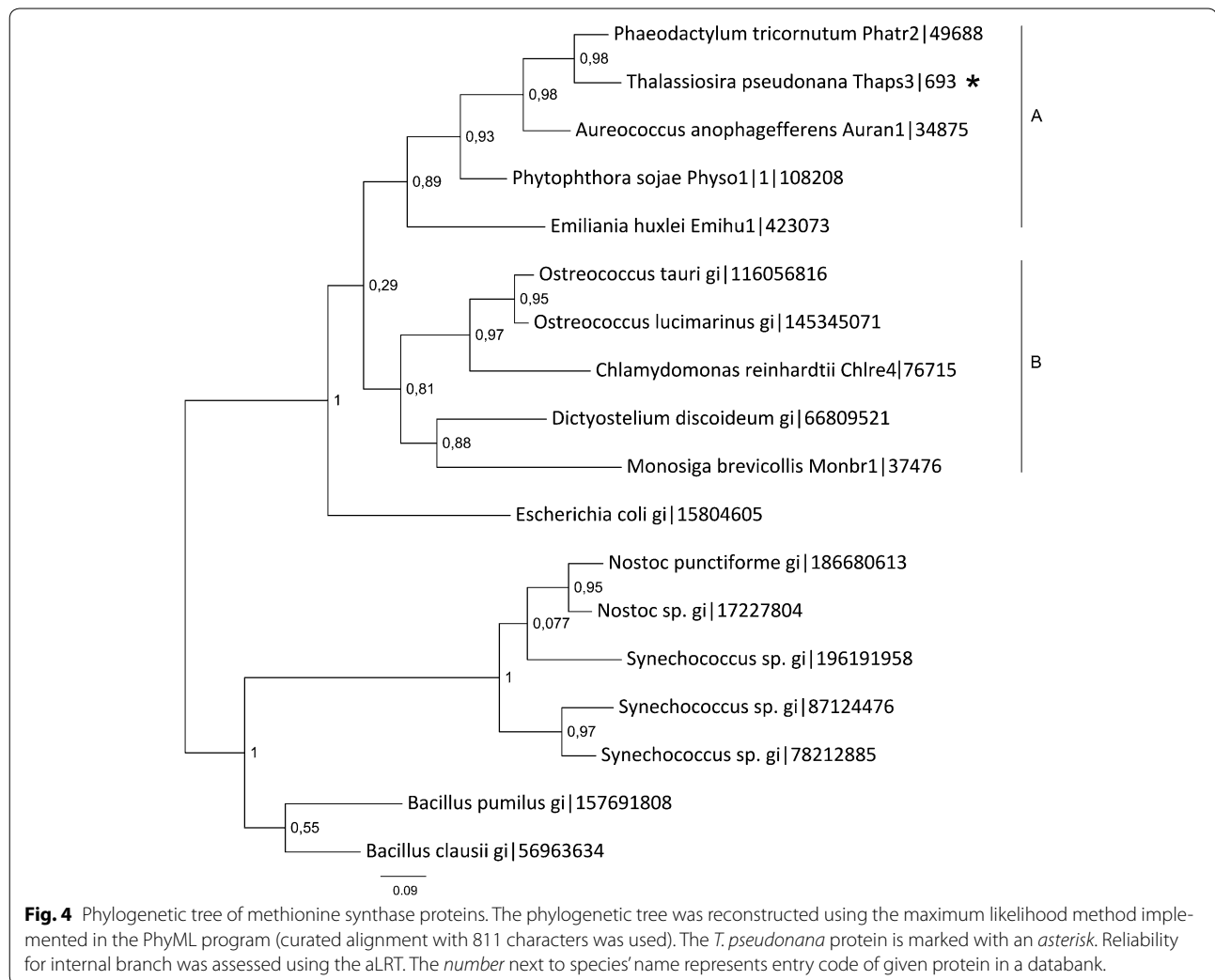


Fig. 3 Phylogenetic tree of PLP-dependent enzymes. The phylogenetic tree was reconstructed using the maximum likelihood method implemented in the PhyML program (curated alignment with 149 characters was used). *T. pseudonana* proteins are marked with asterisks. Reliability for internal branch was assessed using the aLRT test. The number next to species' name represents entry code of given protein in a databank.

long methionine synthase gene could be found (Phatr2, ProteinID 23399). This gene model contains a 2,280-long intron, which in the same reading frame as both

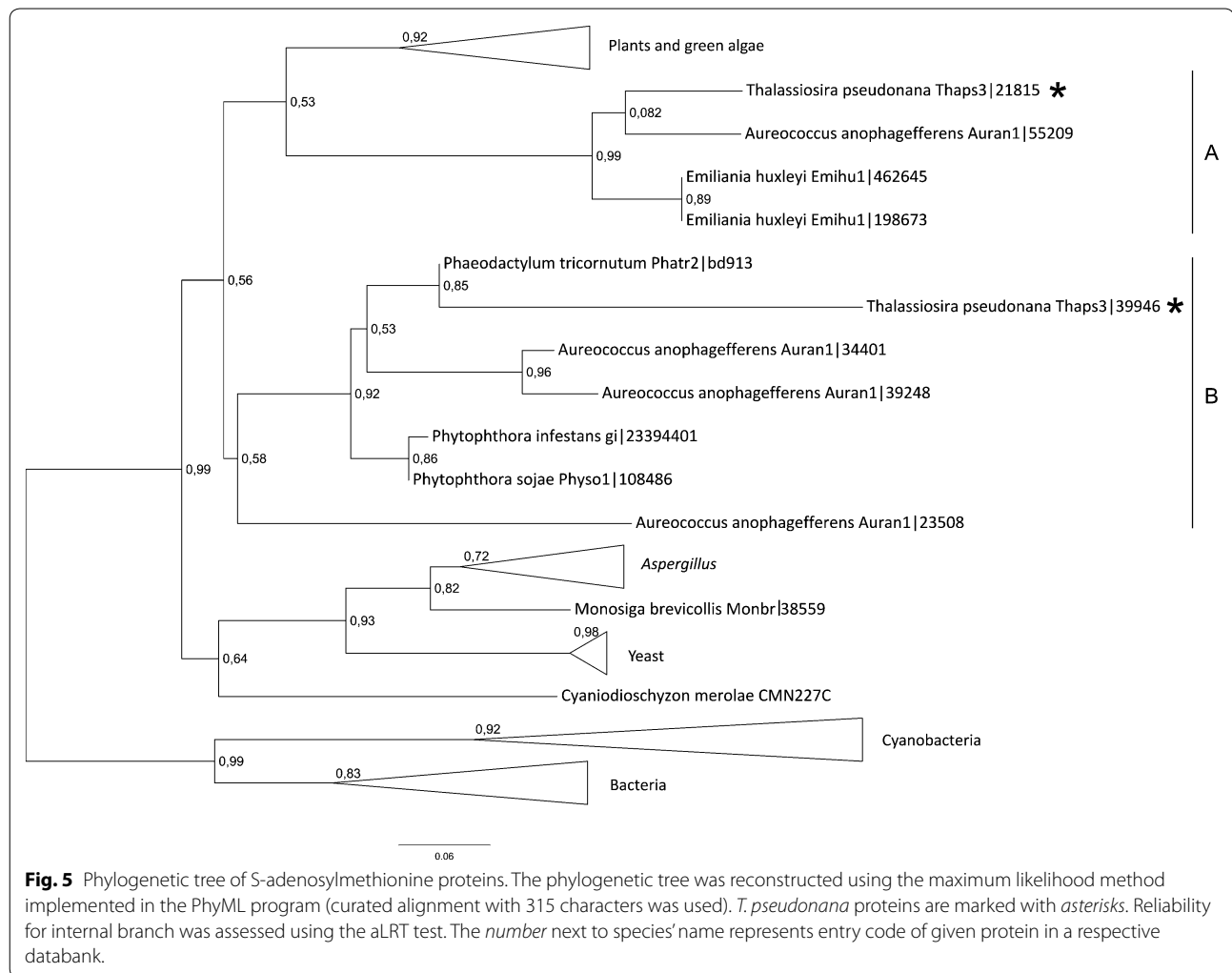
exons encodes a conserved domain of MetH and pterin-binding domain. This is probably a result of a mutation-duplication which led to doubling of the domain. The



P. tricornutum genome repository at Joint Genome Institute (<http://genome.jgi-psf.org/Phatr2>) contains also EST's of the intron suggesting the presence the intron-bearing mRNA in cells and validates the genome sequencing outcome. The presence of two MetH domains has to be verified experimentally. Moreover, the organism contains also a METE-type methionine synthase, which shows similarity to prokaryotic cobalamin-independent enzymes and to the one found in *Chlamydomonas reinhardtii*. The clade B contains MS sequences from green algae and primitive eukaryotes—choanoflagellate *Monosiga brevicollis* and amoeba *Dictyostelium dyscoideum* (Fig. 4, cluster B). We can conclude from the obtained data that cobalamin-dependent MS of *T. pseudonana* might originate from an enzyme of the non-photosynthetic eukaryotic ancestor, while the METE-type of enzyme coming from the genome of a phototrophic eukaryote (ancestor of the diatomal plastid) was lost during evolution.

S-adenosylmethionine synthetase (SAMS) catalyzes the reaction forming SAM from methionine and ATP.

Multiple S-adenosylmethionine synthetase isoenzymes can be found in divergent organisms. Single isoforms of SAM synthetase exist in *Plasmodium falciparum* (Chiang et al. 1999) and in *Drosophila melanogaster* (Larsson and Rasmuson-Lestander 1997). Higher plants like *Cantharanthus roseus* have three isoforms of SAMS (Schröder et al. 1997), whereas in the rice genome two genes could be identified encoding SAMS (Lee et al. 1997). Despite evolutionary divergence, the SAMS enzyme is higher conserved. Between species approximately 30% of S-adenosylmethionine proteins are identical in almost every species (Sánchez-Pérez et al. 2004). The genome of *T. pseudonana* contains two sequences of S-adenosylmethionine synthetase. The absence of a predictable SP at the N-terminal part of both SAM synthetase isoforms suggests that these enzymes are located in the cytosol. The phylogenetic tree of SAMS sequences shows separation of eukaryotic and prokaryotic enzymes (Fig. 5). SAM synthetases from *T. pseudonana* are



found in the clusters A and B (Fig. 5). The first cluster, apart from SAMS1 (Thaps3:21815) contains two *E. huxleyi* sequences and one *A. anophagefferens*. The second cluster (Fig. 5, cluster B) with SAMS2 (Thaps3:39946) contains sequences from organisms belonging to the *Heterokontophyta* clade. Fungal and plant sequences are found in well defined clusters. The sequence comparison suggest that diatomal SAM synthases might originate from two lineages of organisms, which reflects the history of diatoms (Parker et al. 2008): SAM synthetase 2 originates from the heterotrophic host (Thaps3:39946) while the ancestor of SAMS1 derives from the engulfed phototrophic algal cell (Thaps3:21815).

Analysis of the up-stream DNA sequences of methionine metabolism genes

To gain insight on the regulation of the gene expression we have analyzed DNA sequences of putative promoter regions of the methionine metabolism genes.

We have compared sequences of 16 genes in search for conserved motifs. Additionally a set of ten up-stream sequences of ten genes not-related with sulfur amino acids metabolism was used as a negative control. The analysis resulted in three conserved sites present in the pool of studied sequences and absent in the negative control set (see [Methods](#) chapter). The best result was a 11-nt-long motif G[TA]T[GT]G[AT]GT[TG]G[GTA] (consensus sequence with variable nucleotides in square brackets), which was found in 15 out of analyzed 16 sequences (Table 2). The E value that is the measure of statistical significance calculated on motifs log likelihood ratio, its width, number of sites and the background letter frequencies in the training set. For the best match discussed here the E value was $4.2e-3$. Subsequently the motif was compared using TOMTOM with a database of prokaryotic, yeast and plant motifs. Matching with plants and prokaryotic motifs produced ambiguous results, where as the comparison

Table 2 Predicted conserved motifs in 15 putative promoters of *T. pseudonana* genes

| Gene | Protein ID | Length | Start | p value | Sites |
|---------------------------------------|------------|--------|-------|----------|------------------------------------|
| Aspartate kinase | 269371 | 500 | 232 | 1.31e-05 | GATGCTCACG GTTGAAGTTGG TCTCGCCATG |
| Aspartate semialdehyde dehydrogenase | 32577 | 500 | 114 | 5.43e-05 | GGCCATTGTG ATTTGAGTAGT GGTA AAAAAG |
| Cystathionine β -lyase-like | 269520 | 1,000 | 780 | 1.32e-07 | GGAGAAGAGG GTTGGAGTTGG AAGGTCGTGT |
| Cystathionine γ -synthase-like | 9830 | 500 | 45 | 3.82e-05 | GAGGGCGGAG GAAGGAGTGA GCCCTCTCTC |
| Homoserine acetyltransferase | 12141 | 700 | 64 | 8.59e-06 | AAGAGGCTTC GTTGGTGTGA CGACGACAGA |
| Homoserine dehydrogenase | 29144 | 300 | 60 | 4.12e-05 | ACAACCTAAC GTTGGTGTGG CTGGTGTGC |
| Homoserine dehydrogenase | 40655 | 500 | 332 | 1.72e-06 | TCCTTTGCGT GATGGAGTTGT CTGATGTAT |
| Homoserine dehydrogenase | 24587 | 300 | 154 | 3.92e-06 | GTTAGTTGTT GTTAGAGTTGT TTGGTTTCT |
| Homoserine kinase | 28592 | 1,000 | 81 | 2.85e-07 | CCATATCTTT GTTGGAGTTGT TTTGATCGA |
| Methionine synthase | 693 | 1,000 | 36 | 4.39e-07 | CGTCGACAGA GTTTGAGTTGG CTGAAACAGA |
| Methionine γ -lyase-like | 11993 | 700 | 323 | 6.40e-06 | GCGTTGACAT GATTGTGTTGT GCCTCATGTC |
| SAM synthetase | 21815 | 500 | 336 | 9.81e-06 | AACGTGCCTC GTTTGTGTAGG CTGGTATGGT |
| SAM synthetase | 39946 | 300 | 152 | 1.04e-05 | ACCATTGTTT GATTGTGTTGT GGGTAACAGC |
| Glutamylcysteine synthetase | 13064 | 1,000 | 4 | 3.29e-05 | GTG TTTTGAGTTGA TGCTTTTCGG |
| Glutathione synthetase | 29212 | 1,000 | 700 | 3.41e-05 | GCTTCATGAG GAAAGAGTTGG GTGTTTGAG |

The columns "Protein ID" and "Length" contain entry numbers of protein models and length of is DNA sequence up-stream of 5'-UTR. The following columns present start position, random letters probability and the sequence of the conserved motif with ten nucleotide padding.

with the collection of yeast cis-regulatory motifs resulted in identification of a MET31 binding site (p value 0.0487645 and E value 5.9724) (Fig. 6). Only one sequence from the analyzed pool that didn't contained said motif belonged to one of two aspartate kinases. The phosphorylation of aspartate is a regulatory point of the whole methionine and threonine synthesis pathway in yeast (Ramos et al. 1991) and plants (Curien et al. 2008). These data can be interpreted that at least one

isoform is differently regulated and does not belong to the Met/Thr-regulatory network.

Presence of different transcription controllers enables integration of various stimuli and proper response on the level of gene expression. This includes binding of different transcription factors to different DNA-binding sites within a promoter and binding of different members of a transcription factors family to one site in a promoter. In budding yeast, *Saccharomyces cerevisiae*, organic sulfur metabolism is regulated by the Met4 transcriptional system (Lee et al. 2010). The Met4 cannot bind to DNA but through protein-protein interaction with either of two highly similar DNA-binding proteins, Met31p and Met32p, or with a homodimer of Cbf1 proteins. Met31p and Met32p are C2H2-type zinc finger proteins which bind to consensus motif AACTGTGG found in 5'-upstream of several genes of the sulfur metabolism in yeast (Blaiseau et al. 1997). The Cbf1 (centromere-binding factor 1) is a basic helix-loop-helix protein, which forms homodimers to bind to sites with consensus CACGTG core present at MET gene promoters (Kuras et al. 1996). This makes it to one of main activators of methionine metabolism and glutathione synthesis genes in yeast. Very little is known about the transcriptional regulation of gene expression in diatoms (Upper 1974). In genomes of *T. pseudonana* and *P. tricornutum* as much as 46 and 40 zinc finger transcription factors were identified, respectively. These belong to three types: C2H2, CCCH and TAZ type (Rayko et al. 2010). Neither, Met4p nor Met31p/Met32p yeast orthologs were found in the genome of *T. pseudonana*. Though, it can not be excluded

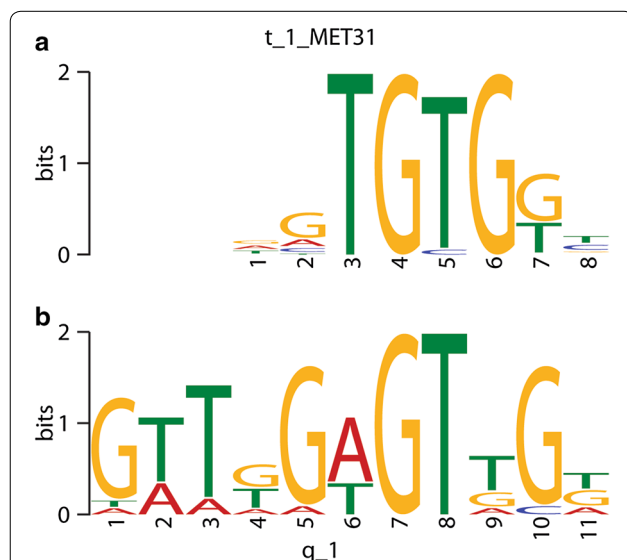


Fig. 6 The sequence-logo comparison of conserved motifs between **a** yeast MET31 binding site and **b** the motif found in putative promoters of *T. pseudonana* methionine metabolism genes.

that the regulation of the sulfur metabolism in this diatom might rely on similar conserved mechanism, including conserved protein binding sites in the promoter regions of the involved genes, as in unicellular organisms like yeast.

Conclusions

Our goal in this study was to identify core features enabling to predict potential exchanges between species and localisation of proteins in order to get insight into the localization of cysteine and methionine pathways in diatoms. Due to the fact that relatively few sequences are known from diatom species, current subcellular localization prediction algorithms are not capable to assign some of the tested protein sequences to their compartments. In most cases it was possible to identify the signal and transit peptide-like domains in sequences of proteins localized to plastids, which was based on the previous observations made in other organisms. However, in other cases the prediction of the localization of proteins known to be targeted to plastids was not possible, which might be caused by modified and/or adapted signal peptides allowing crossing four membranes into the lumen of plastids.

The analysis of gene sequences encoding enzymes of the methionine synthesis pathway revealed that the predicted proteins are specific for algae of the chromalveates or heterokont type. The homology of the tested proteins drops down comparing the algae proteins with orthologous proteins from plants and even more when compared to bacterial or cyanobacterial proteins. In consequence, the results mirror the distinct evolutionary path of diatoms.

The homology search made it possible to identify two enzymes, HSK and HAT, utilizing homoserine as precursor to direct this compound into pathways forming either threonine or homocysteine, respectively. The analysis does not allow to rule out that similar to yeast acetylated homoserine is the acceptor for reduced sulfur from sulfur assimilation pathway. To proof this finding and specify the true catalytic properties of the PLP-enzymes (CbL-like, CgS-like, and MgL-like) further biochemical studies have to be done. Moreover, the identification of a common motif in promoter regions of 15 genes involved in methionine metabolism of *T. pseudonana* suggests evolutionary conservation of the regulatory mechanism of the pathway. This finding, together with the homoserine acetyltransferase gene supports the hypothesis of the residual “fungal-type” of sulfur amino acids biosynthesis pathway in diatoms (Bromke and Hesse 2013). Phylogenetic consideration revealed that in *T. pseudonana* methionine synthase represents the Meth type of identified methionine synthases requiring cobalamin as

cofactor. The cobalamin auxotrophy in *T. pseudonana* is probably the basis for a symbiotic relationship with cobalamin-producing bacteria.

The intensive analysis of the N-terminal site of methionine synthase protein could not identify a transit peptide and thus the result strongly suggests that the synthesis of methionine takes place in the cytosol, which would be in concordance with methionine synthesis in plants. It is likely that during the evolution *T. pseudonana* lost its cobalamin-independent enzyme. On the other hand SAM synthetase 1 seems to be retained from the phototrophic ancestor algae, whereas the SAMS 2 is a legacy of the heterotrophic ancestor-host cell.

Diatoms are fascinating organisms and studying their biology will bring many interesting observations and conclusions. Sequencing more and different diatoms will help increase our knowledge and to refine the prediction algorithms in order to get insight into diatoms metabolism.

Methods

Sequence selection

Amino acid sequences of annotated sulfur metabolism gene models of *T. pseudonana* were used as queries to perform protein–protein BLAST (Altschul et al. 1997) in order to identify the orthologous genes from other organisms. The search was restricted to a set of organisms representing different prokaryotic and eukaryotic organisms: *E. coli*, *Bacillus sp.*, *Synechococcus sp.*, *Nostoc sp.*, *P. tricornutum*, *E. huxleyi*, *A. anophagefferens*, *Ostreococcus sp.*, *C. reinhardtii*, *Phytophthora sp.*, *Selaginella moellendorffii*, *Physcomitrella patens*, *A. thaliana*, *Oryza sativa*, *Porphyra purpurea*, *C. merolae*, *Monosiga brevicollis*, *Dictyostelium discoideum*, and *T. cruzi*. Hit-results with E value $>1e-10$ were discarded. The protein sequences obtained from blastp were then analyzed 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 ‘*C. merolae* Genome Project’ (<http://merolae.biol.s.u-tokyo.ac.jp>). Sequences of *T. pseudonana*, *P. tricornutum*, *S. moellendorffii*, *E. huxleyi*, *P. sojiae*, *A. anophagefferens*, *M. brevicollis* proteins were acquired from Joint Genome Institute sequencing projects (<http://genome.jgi-psf.org>). The protein sequences from *D. discoideum* were obtained from dictyBase (<http://dictybase.org>). Sequences of proteins from *T. cruzi* were obtained from GeneDB (<http://www.genedb.org>). Sequences from other organisms were retrieved from GeneBank (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Analysis of DNA and protein sequences

Target peptides were sought within amino acid sequences upstream of conserved catalytical domains of proteins of interest. To identify N-terminal extensions from the conserved regions of the proteins, sequences were analyzed 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/>). The analysis of sequences longer than 15 amino acids for presence of N-terminal motifs was performed as described in Emanuelsson et al. (2007). Settings were as follows: non plant organisms; prediction of cleavage site and without cut-offs. The output contains final Neuronal Networks scores (NN) (Nielsen et al. 1997) on which the prediction is based. The scores are not probabilities. However, the location with the highest score is 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. Results of the analysis as well as conserved domain starting position are presented in Table 1. Presence of ER retention motifs in analyzed sequences (KDEL, DEEL or DDEL) was verified manually.

The detection of conserved motifs in DNA sequences of the up-stream regions of the *T. pseudonana* methionine metabolism genes was performed using an online tool MEME (<http://meme-suite.org/meme/>). The motif search was defined by following parameters: three best results, expected length 6–12 nt, occurrence once or none, ten sequences as a negative set. The sequences used in the analysis were obtained from *T. pseudonana* genome repository (<http://genome.jgi-psf.org/Thaps3/Thaps3.home.html>). The length of the DNA promoter region sequence of a gene was chosen to be longer than the half of a region between two ORFs (including UTRs). 5'-UTR, as defined by the repository, was included in the input data. In cases of longer stretches of DNA between genes 1,000 nt-long sequences were retrieved. The lengths, not including UTRs, are given in Table 2. The negative set, which served to verify the MEME results, composed of 5'-regions (500 nt + 5'-UTR) of ten non-sulfur-related proteins: actin-related protein, nitrate reductase, GAPDH, HSF60, citrate synthase, LHCr1, galactose kinase, MAP kinase, ubiquitin, tubulin, lipase. The results of the MEME analysis were applied to TOMTOM (<http://meme-suite.org/tools/tomtom>), which is the online software performing comparison of given motifs with available databases. Comparisons with plant (Jaspar 2014), prokaryotic (ProDoric 8.9) and yeast (MacIsaac v.1) databases were performed using default settings. The sequence-logo graphics was generated by the TOMTOM.

Phylogenetic analysis

Phylogenetic trees were inferred using Phylogeny.fr (<http://www.phylogeny.fr>)—an on-line software package for sequence analysis (Dereeper et al. 2008). Sequences were aligned with MUSCLE 3.7 (Edgar 2004) and curation of the alignments was done with Gblocks 0.92 (Castresana 2000) using following settings:(a) conserved position: in at least half the number of sequences + 1; (b) flank position: at least 85% of the number of sequences; (c) maximum number of contiguous nonconserved positions: 8; (d) 10 amino acids as minimum length of a block; (e) now gaps allowed. The phylogenetic trees were reconstructed using the maximum likelihood method implemented in the PhyML program (v3.0 aLRT) (Guindon and Gascuel 2003). WAG substitution method was used as default. The bootstrapping procedure was replaced by an approximate likelihood-ratio test (aLRT) (Anisimova and Gascuel 2006). The test's accuracy and power in comparison to the bootstrapping was discussed by Anisimova et al. (2011). Support values are presented at the nodes with a good statistics threshold set to 0.8. FigTree 1.3.1 software (<http://tree.bio.ed.ac.uk/software/figtree/>) was used to generate graphics.

Authors' contributions

MAB conduct the analysis, prepared the manuscript; HH prepared the manuscript. Both authors read and approved the final manuscript.

Acknowledgements

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 and is greatly acknowledged.

Compliance with ethical guidelines

Competing interests

This work was supported by Max Planck Society and Deutsche Forschungsgemeinschaft (Grant number HE 3088/4). The authors declare that they have no competing interests.

Received: 7 May 2015 Accepted: 17 July 2015

Published online: 04 August 2015

References

- Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W et al (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
- Anisimova M, Gil M, Dufayard J-F, Dessimoz C, Gascuel O (2011) Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol*. doi:10.1093/sysbio/syr041
- Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH et al (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* 306(5693):79–86

- Blaiseau PL, Isnard AD, Surdin-Kerjan Y, Thomas D (1997) Met31p and Met32p, two related zinc finger proteins, are involved in transcriptional regulation of yeast sulfur amino acid metabolism. *Mol Cell Biol* 17(7):3640–3648
- Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A et al (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456(7219):239–244
- Bromke M, Hesse H (2013) Phylogenetic aspects of the sulfate assimilation genes from *Thalassiosira pseudonana*. *Amino Acids* 44(5):1253–1265. doi:10.1007/s00726-013-1462-8
- 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
- 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
- Chassagnole C, Fell DA, Rais B, Kudla B, Mazat JP (2001) Control of the threonine-synthesis pathway in *Escherichia coli*: a theoretical and experimental approach. *Biochem J* 356(Pt 2):433–444
- Chiang PK, Chamberlin ME, Nicholson D, Soubes S, Su X, Subramanian G et al (1999) Molecular characterization of *Plasmodium falciparum* S-adenosylmethionine synthetase. *Biochem J* 344(Pt 2):571–576
- Croft MT, Lawrence AD, Raux-Deery E, Warren MJ, Smith AG (2005) Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* 438(7064):90–93
- Curien G, Dumas R, Ravanel S, Douce R (1996) Characterization of an *Arabidopsis thaliana* cDNA encoding an S-adenosylmethionine-sensitive threonine synthase. Threonine synthase from higher plants. *FEBS Lett* 390 (1):85–90
- Curien G, Job D, Douce R, Dumas R (1998) Allosteric activation of *Arabidopsis* threonine synthase by S-adenosylmethionine. *Biochemistry* 37(38):13212–13221
- Curien G, Biou V, Mas-Droux C, Robert-Genthon M, Ferrer J-L, Dumas R (2008) Amino acid biosynthesis: new architectures in allosteric enzymes. *Plant Physiol Biochem* 46(3):325–339. doi:10.1016/j.plaphy.2007.12.006
- Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F et al (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36(Web Server issue):W465–W469
- Dias B, Weimer B (1998) Purification and characterization of L-methionine gamma-lyase from *Brevibacterium linens* BL2. *Appl Environ Microbiol* 64(9):3327–3331
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797
- Eichel J, Gonzalez JC, Hotze M, Matthews RG, Schroder J (1995) Vitamin B₁₂-independent methionine synthase from a higher-plant (*Catharanthus roseus*)—molecular characterization, regulation, heterologous expression, and enzyme properties. *Eur J Biochem* 230(3):1053–1058
- 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
- Evans JC, Huddler DP, Hilgers MT, Romanchuk G, Matthews RG, Ludwig ML (2004) Structures of the N-terminal modules imply large domain motions during catalysis by methionine synthase. *Proc Natl Acad Sci USA* 101(11):3729–3736. doi:10.1073/pnas.0308082100
- Faleev NG, Troitskaya MV, Paskonova EA, Saporovskaya MB, Belikov VM (1996) L-methionine- γ -lyase in *Citrobacter intermedius* cells: Stereochemical requirements with respect to the thiol structure. *Enzym Microb Technol* 19(8):590–593
- Falkowski PG, Barber RT, Smetacek V (1998) Biogeochemical controls and feedbacks on ocean primary production. *Science* 281(5374):200–206
- Farfan M, Calderon IL (2000) Enrichment of threonine content in *Saccharomyces cerevisiae* by pathway engineering. *Enzym Microb Technol* 26(9–10):763–770
- 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
- Fogliano M, Borne F, Bally M, Ball G, Patte JC (1995) A direct sulphydrylation pathway is used for methionine biosynthesis in *Pseudomonas aeruginosa*. *Microbiol UK* 141:431–439
- Gage DA, Rhodes D, Nolte KD, Hicks WA, Leustek T, Cooper AJ et al (1997) A new route for synthesis of dimethylsulphoniopropionate in marine algae. *Nature* 387(6636):891–894
- 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
- Gómez LD, Vanacker H, Buchner P, Noctor G, Foyer CH (2004) Intercellular Distribution of Glutathione Synthesis in Maize Leaves and Its Response to Short-Term Chilling. *Plant Physiol* 134(4):1662–1671. doi:10.1104/pp.103.033027
- Goulding CW, Matthews RG (1997) Cobalamin-dependent methionine synthase: a modular protein and a zinc metalloenzyme. *FASEB J* 11(9):249
- Griffith OW (1999) Biologic and pharmacologic regulation of mammalian glutathione synthesis. *Free Radic Biol Med* 27(9–10):922–935. doi:10.1016/S0891-5849(99)00176-8
- 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
- Hesse H, Hoefgen R (2003) Molecular aspects of methionine biosynthesis. *Trends Plant Sci* 8(6):259–262
- Hesse H, Hoefgen R (2008) Metabolism of methionine in plants and phototrophic bacteria. In: Hell R, Dahl C, Knaff D, Leustek T (eds) Sulfur metabolism in phototrophic organisms. *Advances in photosynthesis and respiration*, vol 27. Springer, Netherlands, pp 93–110. doi:10.1007/978-1-4020-6863-8_5
- 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
- Hwang BJ, Yeom HJ, Kim Y, Lee HS (2002) *Corynebacterium glutamicum* utilizes both transsulfuration and direct sulphydrylation pathways for methionine biosynthesis. *J Bacteriol* 184(5):1277–1286. doi:10.1128/jb.184.5
- Inoue H, Inagaki K, Sugimoto M, Esaki N, Soda K, Tanaka H (1995) Structural analysis of the L-methionine gamma-lyase gene from *Pseudomonas putida*. *J Biochem* 117(5):1120–1125
- Kacprzak MM, Lewandowska J, Matthews RG, Paszewski A (2003) Transcriptional regulation of methionine synthase by homocysteine and choline in *Aspergillus nidulans*. *Biochem J* 376:517–524. doi:10.1042/bj20030747
- Keller MD, Kiene RP, Matrai PA, Bellows WK (1999) Production of glycine betaine and dimethylsulphoniopropionate in marine phytoplankton. II. N-limited chemostat cultures. *Mar Biol* 135(2):249–257
- Kerr DS (1971) O-acetylhomoserine sulphydrylase from *Neurospora*. Purification and consideration of its function in homocysteine and methionine synthesis. *J Biol Chem* 246(1):95–102
- Kerr DS, Flavin M (1970) Regulation of methionine synthesis and nature of cystathionine gamma-synthase in *Neurospora*. *J Biol Chem* 245(7):1842–1855
- 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
- Kopriva S, Patron NJ, Keeling P, Leustek T (2008) Phylogenetic analysis of sulfate assimilation and cysteine biosynthesis in phototrophic organisms. In: Hell R, Dahl C, Knaff DB, Leustek T (eds) Sulfur metabolism in phototrophic organisms. *Advances in photosynthesis and respiration*. Springer, The Netherlands, pp 31–58
- Krueger S, Niehl A, Martin MC, Steinhäuser D, Donath A, Hildebrandt T et al (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
- Krungkrai J, Webster HK, Yuthavong Y (1989) Characterization of cobalamin-dependent methionine synthase purified from the human malarial parasite, *Plasmodium falciparum*. *Parasitol Res* 75(7):512–517
- Kuras L, Cherest H, Surdin-Kerjan Y, Thomas D (1996) A heteromeric complex containing the centromere binding factor 1 and two basic leucine zipper factors, Met4 and Met28, mediates the transcription activation of yeast sulfur metabolism. *EMBO J* 15(10):2519–2529
- Larsson J, Rasmuson-Lestander A (1997) Cloning, mapping and mutational analysis of the S-adenosylmethionine decarboxylase gene in *Drosophila melanogaster*. *Mol Gen Genet* 256(6):652–660
- Lee JH, Chae HS, Lee JH, Hwang B, Hahn KW, Kang BG et al (1997) Structure and expression of two cDNAs encoding S-adenosyl-L-methionine synthetase of rice (*Oryza sativa* L.). *Biochim Biophys Acta* 1354(1):13–18
- Lee TA, Jorgensen P, Bogner AL, Peyraud C, Thomas D, Tyers M (2010) Dissection of combinatorial control by the Met4 transcriptional complex. *Mol Biol Cell* 21(3):456–469. doi:10.1091/mbc.E09-05-0420

- Lockwood BC, Coombs GH (1991) Purification and characterization of methionine gamma-lyase from *Trichomonas vaginalis*. *Biochem J* 279(Pt 3):675–682
- Manukhov IV, Mamaeva DV, Rastorguev SM, Faleev NG, Morozova EA, Demidkina TV et al (2005) A gene encoding L-methionine gamma-lyase is present in *Enterobacteriaceae* family genomes: identification and characterization of *Citrobacter freundii* L-methionine gamma-lyase. *J Bacteriol* 187(11):3889–3893. doi:10.1128/jb.187.11.3889-3893.2005
- Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWweese-Scott C, Geer LY, Gwadz M et al (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. *Annu Rev Microbiol* 51:73–96
- McKie AE, Edlind T, Walker J, Mottram JC, Coombs GH (1998) The primitive protozoan *Trichomonas vaginalis* contains two methionine gamma-lyase genes that encode members of the gamma-family of pyridoxal 5'-phosphate-dependent enzymes. *J Biol Chem* 273(10):5549–5556
- McQueney MS, Anderson KS, Markham GD (2000) Energetics of S-adenosyl-methionine synthetase catalysis. *Biochemistry* 39(15):4443–4454
- Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int J Neural Syst* 8(5–6):581–599
- Ostlund G, Schmitt T, Forslund K, Kostler T, Messina DN, Roopra S et al (2010) InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res* 38(Database issue):D196–D203
- 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
- Pócsi I, Prade RA, Penninckx MJ (2004) Glutathione, Altruistic Metabolite in Fungi. In: *Advances in Microbial Physiology*, vol 49, Academic Press, pp 1–76. doi:10.1016/S0065-2911(04)49001-8
- Ramos C, Delgado MA, Calderon IL (1991) Inhibition by different amino acids of the aspartate kinase and the homoserine kinase of the yeast *Saccharomyces cerevisiae*. *FEBS Lett* 278(1):123–126. doi:10.1016/0014-5793(91)80098-N
- Ravanel S, Block MA, Rippert P, Jabrin S, Curien G, Rebeille F (2004) Methionine metabolism in plants: chloroplasts are autonomous for de novo methionine synthesis and can import S-adenosylmethionine from the cytosol. *J Biol Chem* 279(21):22548–22557
- Rayko E, Maumus F, Maheswari U, Jabbari K, Bowler C (2010) Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *New Phytol* 188(1):52–66. doi:10.1111/j.1469-8137.2010.03371.x
- Rebeille F, Jabrin S, Bligny R, Loizeau K, Gambonnet B, Van Wilder V et al (2006) Methionine catabolism in *Arabidopsis* cells is initiated by a gamma-cleavage process and leads to S-methylcysteine and isoleucine syntheses. In: *Proceedings of the national academy of sciences of the United States of America* 103(42):15687–15692. doi:10.1073/pnas.0606195103
- Rowbury RJ, Woods DD (1964) O-succinylhomoserine as intermediate in synthesis of cystathionine by *Escherichia coli*. *J Gen Microbiol* 36(3):341–358
- Sánchez-Pérez GF, Bautista JM, Pajares MA (2004) Methionine adenosyltransferase as a useful molecular systematics tool revealed by phylogenetic and structural analyses. *J Mol Biol* 335(3):693–706
- Schröder G, Eichel J, Breinig S, Schroder J (1997) Three differentially expressed S-adenosylmethionine synthetases from *Catharanthus roseus*: molecular and functional characterization. *Plant Mol Biol* 33(2):211–222
- Sun G, Yang Z, Ishwar A, Huang J (2010) Algal genes in the closest relatives of animals. *Mol Biol Evol* 27(12):2879–2889. doi:10.1093/molbev/msq175
- Tokoro M, Asai T, Kobayashi S, Takeuchi T, Nozaki T (2003) Identification and characterization of two isoenzymes of methionine gamma-lyase from *Entamoeba histolytica*: a key enzyme of sulfur-amino acid degradation in an anaerobic parasitic protist that lacks forward and reverse trans-sulfuration pathways. *J Biol Chem* 278(43):42717–42727
- Upper D (1974) The unsuccessful self-treatment of a case of "writer's block". *J Appl Behav Anal* 7(3):497. doi:10.1901/jaba.1974.7-497a
- 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
- Walker J, Barrett J (1997) Parasite sulphur amino acid metabolism. *Int J Parasitol* 27(8):883–897
- Whitfield C, Steers EJ, Weissbach H (1970) Purification and properties of 5-methyltetrahydropteroyltri-glutamate-homocysteine transmethylase. *J Biol Chem* 245(2):390–401
- Wiebers JL, Garner HR (1967) Acyl derivatives of homoserine as substrates for homocysteine synthesis in *Neurospora crassa*, yeast, and *Escherichia coli*. *J Biol Chem* 242(23):5644–5649
- Zeh M, Casazza AP, Kreft O, Roessner U, Bieberich K, Willmitzer L et al (2001) Antisense inhibition of threonine synthase leads to high methionine content in transgenic potato plants. *Plant Physiol* 127(3):792–802
- Zeh M, Leggewie G, Hoefgen R, Hesse H (2002) Cloning and characterization of a cDNA encoding a cobalamin-independent methionine synthase from potato (*Solanum tuberosum* L.). *Plant Mol Biol* 48(3):255–265
- Zubieta C, Krishna SS, Daniel M, Mitchell DM, Polat A, Sanjay A et al (2007) Crystal structure of homoserine O-succinyltransferase from *Bacillus cereus* at 2.4 Å resolution. *Proteins* 68(4):999–1005

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com