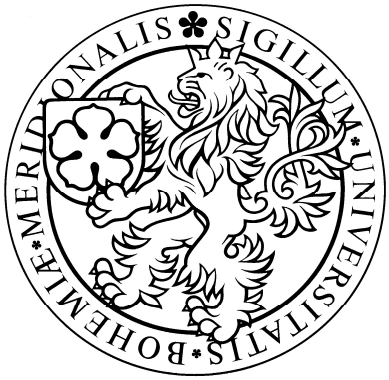
University of South Bohemia
Faculty of Science
Department of Molecular Biology



Ph.D. Thesis

The origin and localization of selected metabolic pathways in marine diatoms

Mgr. Kateřina Jiroutová

Supervisor: Doc. ing. Miroslav Oborník, PhD

University of South Bohemia, Faculty of Science, Department of Molecular Biology &
Biology Centre, Academy of Sciences of the Czech Republic, Institute of Parasitology

České Budějovice, 2009

Jiroutová K., 2009: The origin and localization of selected metabolic pathways in marine diatoms. Ph.D. Thesis, in English – 98 p., Faculty of Sciences, University of South Bohemia, České Budějovice, Czech Republic.

Annotation:

Sequenced diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* belong to the chromist algae harboring secondary plastids, which display distinct evolutionary history when compared to photosynthetic organelles from rhodophytes, green algae and plants. Via secondary endosymbiosis, heterotrophic eukaryotic ancestor of diatoms engulfed red alga, and in addition to the new organelle, it obtained fitness increasing peculiarities in the chimerical cell metabolism and lifestyle. We examined phylogeny and *in silico* localization of the nuclear-encoded but plastid located enzymes of tryptophan biosynthesis. We suggest that the diatom tryptophan pathway represents an extreme in the trend of plastid (cyanobacterial) enzymes to be replaced by eukaryotic isoforms. In addition, the gene napped during the endosymbiotic gene transfer from the diatom plastid genome to the diatom nucleus (*psb28*) was described.

Financial support:

This work was supported by the Grant Agency of the Academy of Sciences of the Czech Republic, Project No. IAA500220502 and Research Plan No. z60220518, and Ministry of Education of the Czech Republic, project no: 6007665801 (M. Oborník). This study is a part of the research project of the Institute of Parasitology, Academy of Sciences of the Czech Republic (Z60220518), Research Centre (LC06009) and Research Program (524/03/H133) from the Ministry of Education, Youth and Sports of the Czech Republic, and of the Faculty of Biological Sciences, University of South Bohemia, České Budějovice (MSM 6007665801)

I hereby declare that I did this work on my own, in collaboration with co-authors of the presented papers and using only the cited literature.

I declare that in accordance with the Czech legal code § 47b law No. 111/1998 in its valid version, I consent to the publication of my PhD. thesis (in an edition made by removing marked parts archived by the Faculty of Science) in an electronic way in the public access to the STAG database run by the University of South Bohemia in České Budějovice on its web pages.

České Budějovice, September, 2009

……..

Acknowledgements:

First of all I would like to thank my supervisor Miroslav Oborník for giving me excellent topic of my thesis. At least, I found the protists and algae as very interesting organisms for research. He also gave me many good advices even he caused my personal growth during the last 5 years.

My thanks belong to Chris Bowler the head of Laboratory of Plant Molecular Biology in Paris for a great opportunity of staying in his laboratory and support in the diatom research. I should thank to the former members of the Laboratory of molecular taxonomy: Aleš Horák, Jan Janouškovec and Tomaš Chrudimský for a substantial help at my start of the studies. I also thank to Luděk Kořený, Jan Brabec and other members of our laboratory for creating a good background for research. I sent many thanks to all my friends, who make my life more colorful and exciting. Specially, I would thank to my unborn baby to be kind and do not problems during the writing of PhD. thesis.

Preface

This PhD. thesis is based upon two publications and one manuscript:

1.) On the first publication, Kateřina Jiroutová is the first author with the significant contribution:

Jiroutová K., Horák A., Bowler C., Oborník M., 2007. Tryptophan biosynthesis in stramenopiles: eukaryotic winners in the diatom complex chloroplast. *Journal of Molecular Evolution* 65, 496–511.

2) The second publication originated as the conclusion of genome project of the sequencing diatom *Phaeodactylum tricornutum*. Kateřina Jiroutová participates in the annotation team and verified the gene models encoding the enzymes for aromatic amino acid biosynthesis and isopentenyl diphosphate synthesis.

Bowler C., Allen A.E., Badger J.H., Grimwood J., Jabbari K., Kuo A., Maheswari U., Martens C., Maumus F., Otillar R.P., Rayko E., Salamov A., Vandepoele K., Beszteri B., Gruber A., Heijde M., Katinka M., Mock T., Valentin K., Verret F., Berges J.A., Brownlee C., Cadoret J.P., Chiovitti A., Choi C.J., Coesel S., De Martino A., Detter J.C., Durkin C., Falciatore A., Fournet F., Haruta M., Huysman M.J.J., Jenkins B.D., **Jiroutová K.**, Jorgensen R.E., Joubert Y., Kaplan A., Kröger N., Kroth P.G., La Roche J., Lindquist E., Lommer M., Martin–Jézéquel V., Lopez P.J., Lucas S., Mangogna M., McGinnis K., Medlin L.K., Montsant A., Oudot–Le Secq M.P., Napoli C., Oborník M., Schnitzler-Parker M., Petit J.L., Porcel B.M., Poulsen N., Robison M., Rychlewski L., Rynearson T.A., Schmutz J., Shapiro H., Siaut M., Stanley M., Sussman M.R., Taylor A.R., Vardi A., von Dassow P., Vyverman W., Willis A., Wyrwicz L.S., Rokhsar D.S., Weissenbach J., Armbrust V.E., Green B.R., Van de Peer Y., Grigoriev I.V., 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456:239-244.

3) Enclosed manuscript is the version prepared for submission with the author primacy of K. Jiroutová:

Jiroutová K., Bowler C., Oborník M., 2009. A gene in the process of endosymbiotic gene transfer. (unpublished manuscript)

…….

Miroslav Oborník (supervisor)

Table of contents

1. Introduction

Diatoms (Bacillariophycae) are unicellular eukaryotic photoautotrophs with cell no longer than 200 µm. They display extreme diversity mostly as photosynthetic autotrophs, sometimes as the colorless heterotrophs or photosynthetic symbionts. Generally, diatoms can be found in almost any aquatic environment. The advanced ability of carbon dioxide uptake and innovation of the photosystem protective pigments placed these algae on the top of recent diversity and abundance in the waters. About one fifth of carbon fixation in the world‘s ocean goes right though the pelagic diatom species. They are also good savers of trace elements soluble in the waters, and are able to recycle critical nutrients many times (Kooistra et al., 2007).

Diatom plastid is derived from a secondary endosymbiotic event in which a heterotrophic eukaryote engulfed a photosynthetic eukaryote possessing primary plastid (alga). Diatom ancestor in particular had acquired a red algal symbiont; however a trace of green endosymbioses has also been proposed (Moustafa et al., 2009). The red origin of the diatom plastid leads to unique ultrastructural and metabolic features. Together with other eukaryotes bearing secondary red plastid, they were proposed to form supergroup called Chromalveolata (Cavalier-Smith, 1999), which reflects a fundamentally different evolutionary history from the terrestrial plants. Diatoms belong to the smaller particular group of so called heterokont algae. However, diatoms cannot typically represent the heterokont features such as two flagella, which are occurred just in the part of diatom species, and only during the sexual reproduction. (Chepurnov et al., 2004)

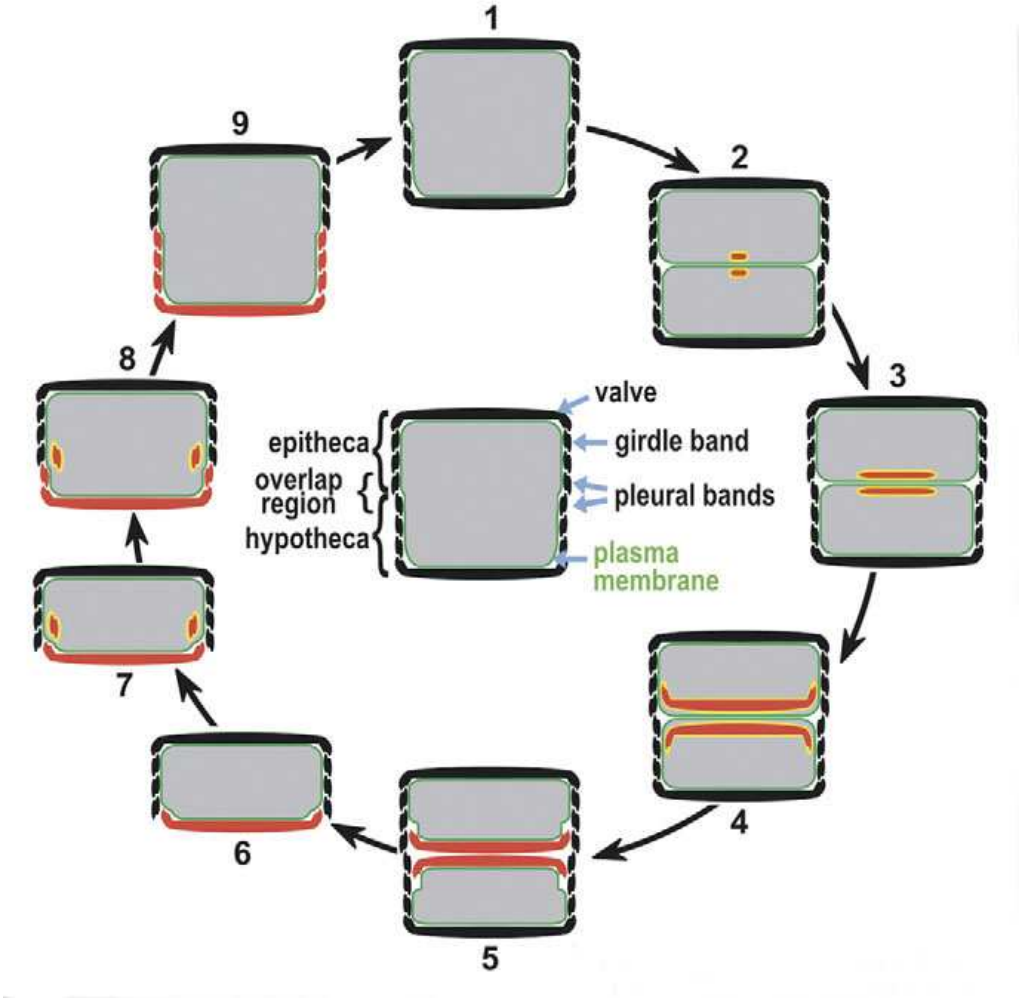
The fact that diatoms can be easily maintained and cultivated in the laboratory conditions predicts them to be the model for physiological, genetic and molecular biological studies. Recent conspicuous progress made in diatom research mostly come from the two sequenced genomes, a wide expressed sequence tags (EST) library and transfection system of diatom (Apt et al., 1996; Armbrust et al., 2004; Bowler et al., 2008; Maheswari et al., 2009). Using this genomic knowledge I would like to contribute to unresolved questions of gene transfer from plastid to nucleus, as a consequence of secondary endosymbiosis. Together with the origin of metabolic pathways in complex plastid it brings a light into the ecological and photosynthetic success of these ubiquitous organisms.

2. The frustule of diatoms

The most privilege feature that is advantageous for the living in the ocean is an external cell wall of diatoms, which is stiffened by silica. Since diatoms do not escape as flagellates, the cell wall protects them from predation. Moreover, the building up the silica-organic shells costs less energy than to make it from carbohydrates. In fact, the organisms with carbonated cell wall are endangered by decreasing pH of the worlds` ocean, and some chromalveolates such as cocolithophores possessing calcium carbonate shell will probably soon become extinct (Doney, 2006). The silica shell was able to retain over million years in sediments, which can perfectly traits the Earth`s history for many decades. At least, no other tiny organism can create so amazing ornaments and cell wall architecture. The ability of diatoms to form silica structures in the regular temperatures of the see water attracts the attention of nanotechnologists (Kröger, 2007).

The name “diatoms” is derived from Greek “*diatomos”*, which means “cut in half”; because of their rigid cell wall consisting of two parts called valves (Armbrust, 2009). The valves are of unequal size and resemble a Petri dish system of closure. The overlapping valve called epitheca attaches the smaller hypotheca by series of girdle bands, which create the mantle. The whole flusture encloses protoplasm, but permits growth by expansion (Kooistra et al., 2007). Both valves are patterned and consist of silica bound to organic matter. The uptake of soluble silica ions $Si(OH)_4$ is mediated by silica transporters in plasma membrane. Silica deposition vesicles cumulate hydrated silica inside. The levels of silica can saturate inside the cells, therefore the concentration of soluble silicic acid during the seasonal bloom can be limited in the ocean (Martin-Jezequel et al., 2000). Inside silica deposition vesicles, silicic acid is bounded at organic matrix of silaffins and long chain polyamines, which cause the silica polymerization to result in very fine nanostructure. It displays species-specific nanostructure of such fine details that diatoms have long been used to test the resolution of optical microscopes (Kooistra et al., 2007). Construction of valves follows immediately upon mitotic nuclear division, and girdle band (see figure 1).

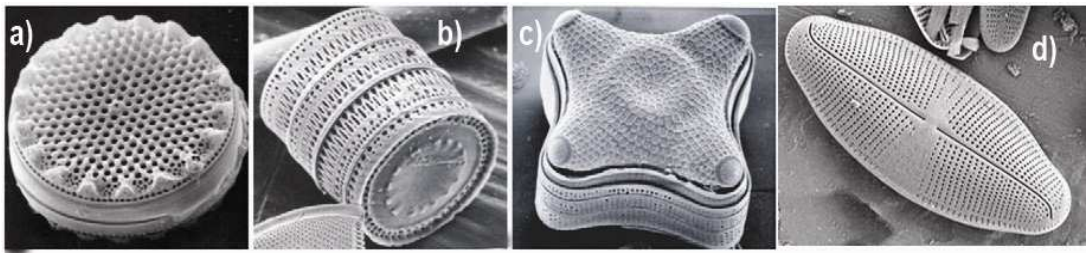
Figure 1. Diatom flusture: division and formation (Kroger, 2007). In the center the schematic structure of a diatom cell wall is shown in cross section. The gray area represents the protoplast, and silica is represented in black. For simplicity intracellular organelles other than the silica depositing vesicles (SDV) are not shown. Cells arranged in the circle show different stages of the cell cycle: (1) after chromosome duplication the cell wall contains the maximum number of girdle bands; (2) immediately after cytokinesis new biosilica (red) is formed in each sibling cell inside a valve SDV (yellow); (3) expansion of the valve SDVs as more and more silica is deposited; (4) at the final stage each SDV contains a fully developed valve; (5) the newly formed valves are deposited by SDV exocytosis in the cleavage furrow on the surface of each protoplast; (6) the sibling cells have separated; (7 and 8) expansion of the protoplast in interphase requires the synthesis of new silica (red) inside girdle band SDVs (yellow); each girdle band is synthesized in a separate SDV, and after SDV exocytosis is added only to the newly formed valve (hypovalve); (9) after synthesis of the final hypovalve girdle band (pleural band) cell expansion stops, and DNA replication is initiated



According to different shapes of valves, diatoms themselves are divided into several groups, which approximately correspond to their evolution (see figure 2). Their fascinating biodiversity with approximately 10,000 species has been recently taxonomically revised giving evidence that the diatoms form two clades: the Coscinodiscophytina and the Bacillariophytina, the latter containing two subclades Bacillariophyceae and Mediophyceae (Medlin and Kaszmarska 2004). Diatoms species diversity evolved during the last 250 million years ago and there is likely no phylogenetic group in which the evolutionary rate has occurred faster. However, the earliest well-preserved diatom fossils come from the later era of

190 million years ago (Sims et al., 2006). Before the expansion of diatoms and other Chromists, the phytoplankton mostly had consisted of cyanobacteria and green algae (Falkowski et al., 2004). Based on the frustule morphology, radial centrics appear to be the most basal group of diatoms. Their features are radialy organized valves with a ring, ribs radiating from center and labiate processes, which are usually in a ring, in the valve mantle or secondarily repositioned near the center. The similarities are found in early fossils, and those heavy silicified organisms probably lived at the bottom of shallow seas. The other radial centrics seems to be adapted more to the planctonic lifestyle; many of them form chain colonies. Lately, the second major lineage of diatoms diverged: bipolar and multipolar centrics. Apparently, multipolarity was a progressive step, because multipolar centrics also show a series of well supported lineages that emerge from a basal sprawl. Their ancestral habitat was probably epiphytic or epilithic. However, several species adapted to planctonic lifestyle, for e.g. Thalassiosirales. The mucilage pads excreted from the apical pore field allows to make connection to the cell surface or to another cell. Multipolar centrics also show a radial pore organization, but their cell outline is usually elongated, triangular or star-like. Based on phylogenetic data, radial centric Thalassiosirales belong to multipolar centrics, their shape probably reverse to radial centricity (Sims et al., 2006, Kooistra et al., 2007). The third group of diatoms rapidly evolved probably after cretaceous mass extinction. Pennate diatoms are elongated with a midrib, from which ribs extend out more or less perpendicularly. They are subdivided into older araphid pennates, which have apical pore fields and apical labiates processes, and evolutionary younger raphid pennates. Distinguishing feature is a slit; a raphe located in cell walls and allows diatoms to glide along surfaces (Kooistra et al., 2007). The evolution of the raphe greatly expanded and probably had as profound an impact on diatom diversification. The raphid pennates dominate the community in the Southern Ocean, the largest region of diatom-based carbon export (Zielinski and Gersode, 1997).

Figure 2. Scaning electron microscopy of cell walls from different diatom species. (Kroger, 2007 and website: http://www.ansp.org/). The shape variability represents the diatom groups: a) polar centric b) side view of polar centric c) multipolar centric d) pennate with raphe



Living with the frustule is associated to interesting life-cycle attributes. Physical and developmental constraints associated with replication of the cell wall mean that in each mitotic division, one daughter cell is slightly smaller than the other. Over successive divisions, cells of dramatically different sizes are found within a population. However, cell size is ultimately restored through sexual reproduction, which occurs differently in centric and pennate diatoms. In centric diatoms, only small cells are receptive to an environmental trigger and become either microgamete resembling "a sperm", which break free of the wall, or egg-like macrogamete, which remains encased within the wall. Microgamete swims to an egg-like macrogamete, gains entry past the frustule, and fertilizes the egg nucleus. The resulted zygote swells into form of specialized cell known as the auxospore. Unlike the old cell wall, the auxospore produces a much larger wall, which restores the original cell size. Similarly to centrics, pennates have a size requirement for the initiation of sexual reproduction, but produce morphologically identical isogametes, which are unable to swim and instead move towards one another. Their fusion creates the zygote and auxospore, which breaks free of the old cell wall. Although centric diatom "sperms" risk finding of "eggs" in the dilute ocean, pennates seem to form isogametes only when they find an appropriate mate of the opposite sex (Armbrust, 2009).

3. Origin of the diatom complex plastid

3.1. Primary endosymbiosis

Mitochondrion and primary plastids are descendants of prokaryotic organisms engulfed by ancient eukaryote during the process called primary endosymbiosis. Adaptation of formerly free-living prokaryotic cell to the environment inside the eukaryotic cell has lead to novelties in their metabolism and lifestyle (Hoffmeister and Martin, 2003; McFadden, 2000). Until this event, the photosynthesis was only present in prokaryotes. Primary endosymbiosis seems to be rare and valuable process. It was suggested that since it is very complicated and includes also establishment of protein import machinery, primary endosymbiosis very probably happened only once in evolution (Fast et al., 2001; Cavalier-Smith, 2003; Obornik and Green, 2005). During the time, when the plastid became fully established, many genes of both photosynthetic and non-photosynthetic functions were transferred to the nucleus of primary host or had been lost. Products of these nuclear encoded genes are expressed in the cytosol and targeted to the place of action thanks to the targeting sequences, kind of cellular address, which is usually placed at the N-terminus of the protein.

Therefore, the recent plastids contain less than 10% of the genome of free-living cyanobacteria (Martin and Hermann, 1998).

Eukaryotic protoalga appears as cellular chimera and rapidly diversified into lineages of Archeplastida (formerly Plantae) group (Adl et al., 2005). First branch is represented by glaucophytes with photosynthetic organelle called cyanelle, which still retains the cyanobacterial peptidoglycan cell wall within its envelope. The second branch is formed by green algae and land plants, the most successful terrestrial photosynthetic eukaryotes. The last group contains red algae-rhodophytes that retained the original cyanobacterial phycobilisome pigments while chlorophytes replaced phycobilisomes by chlorophyll b. However, the main pigment chlorophyll *a* and two surrounding membranes are shared by all primary plastids (Bhattacharya et al., 2003).

The expected monophyly of primary plastids is supported by several common features, such as similar gene content in plastid genomes, the presence of plastid-specific gene clusters, and the conservation of the plastid-protein import machinery (Palmer 2003). Recently, a large amount of nuclear, mitochondrial and plastid sequences was tested under several molecular phylogenetic methods to recover this relationship (Baldauf et al., 2000; Moreira et al., 2000; Rodriguez-Ezpeleta et al. 2005). However, subsequent analyses of such data showed that the tree-building signal is based on more rapidly evolving sequences. This suggests that the result could be affected by phylogenetic artifacts caused by rate variation in sequence evolution (Nozaki et al., 2007). On the other hand phylogenomic analyses provided strong support for monophyly of organisms with primary plastid (Huang and Gogarten 2007). According to molecular clock, the event of primary endosymbiosis was recently estimated to happen about 1.5 billion years ago (Yoon et al., 2004).

The existence of one more primary endosymbiosis is exceptional curiosity. It was clearly shown that *Paullinea chromatophora* (Rhizaria) possess cyanelle-like endosymbiont inside the cell (Theissen and Martin 2006). The advanced level of integration of the *Paulinella* endosymbiont with its host indicates the presence of well-developed metabolite exchange and, probably, import of at least some proteins such as membrane antiporters (Weber et al., 2006). Recently, the genome of *P. chromatophora* plastid-like organelle (called chromatophore) was sequenced (Nowack et al., 2008).

3.2. Secondary and tetriary endosymbiosis

However, not all plastids are products of primary endosymbiosis, which gave rise unicellular green and red algae. Secondary endosymbiosis is a few hundred million years younger process that include acquisition of the photosynthetic organelle through the swallowing primary plastid possessing eukaryotic alga, which was in the case of diatoms rhodophyte. As a result, eukaryotic alga became the photosynthetic organelle surrounded by more than two membranes, usually three or four, and termed secondary or complex plastid. The widely accepted hypothesis is that rhodophytes and chlorophytes subsequently gave rise to many other algal plastids through secondary endosymbiosis (Martin and Hermann, 1998). It is believed that secondary endosymbiosis has happened once in the red lineage of primary plastids, and twice in the green lineage. Complex plastids derived from red algae can be found in chromalveolates (Yoon et al., 2002; Bhattacharya et al., 2003), the large group of protists containing stramenopiles (including diatoms), haptophytes, cryptophytes, dinoflagellates and other alveolates such as ciliates and apicomplexan parasites (figure 3). As mentioned above, two separated branches of photosynthetic eukaryotes is result of two secondary endosymbioses with green algae. Chlorarachniophyta represent distinct group inside the supergroup Rhizaria and Euglenophyta belong to Excavates. However, the symbiotic evolution has not finished by secondary endosymbiotic event. Some dinoflagellates replaced their secondary plastid by tertiary one by engulfing other chromalveolate (diatom, cryptophyte, haptophyte), in the process called tertiary endosymbiosis. (Chesnick et al., 1997; Ishida and Green, 2002; Tengs et al., 2000; Inagaki et al., 2000; Yoon et al., 2002; Obornik et al., 2009).

Thomas Cavalier-Smith postulated in his chromalveolate hypothesis that single endosymbiotic event and subsequent loses of plastid in several branches are more probable than several independent secondary endosymbioses with rhodophyte cell. As already mentioned, particularly dinoflagellates display the highest diversity among plastids found in secondary algae. The majority of dinoflagellates plastids are derived from the red lineage, bound by three membranes, and pigmented by peridinin as the main carotenoid. The dinoflagellate peridinin plastid does not contain a typical circular genome known from other plastids. This genome has been fragmented to the set of minicircles, small circular molecules that mostly holds one gene. Only about fourteen plastid protein genes have been found in peridinin dinoflagellates (Green, 2004), leaving in question the location of the sequences that encode the remaining components of the photosynthetic apparatus (Ishida and Green, 2002;

Hackett et al., 2004; Bachvaroff et al., 2004). Given the ancestral origin of the red algal secondary plastid in dinoflagellates, tertiary replacement has been used to explain the plastid in species such as *Karenia* spp. and *Karlodinium micrum*, which share characters with the haptophyte algae (Chesnick et al., 1997). Other tertiary endosymbiosis also would account for tertiary plastid origin in other species such as *Dinophysis* spp. (swallowed cryptophyte) (Hackett et al. 2003), or *Kryptoperidinium foliaceum* (engulfed diatom, which still retains mitochondrion) (McEvan and Keeling, 2004). In rare cases, such as *Lepidodinium viride* and *Lepidodinium chlorophorum* the original red alga-derived plastid was replaced by prasinophyte, green alga with primary plastid (Watanabe et al., 1991).

However the validity of eukaryotic supergroup Chromalveolata, which is supposed to appear thanks to this single endosymbiotic event, is not spectacular. Supergroup Chromalveolata was postulated on the basis of molecular phylogenetic analyses that unite particular members of these morphologically disparate lineages among alveolates and chromists (figure 3 and 4). For example, distribution and phylogeny of the plastid-targeted glyceraldehyde-3- phosphate dehydrogenase gene (GAPDH) provided evidence that the plastids of Alveolata and Chromista appear to be the result of a single secondary endosymbiotic event, suggesting a common ancestor for these two lineages (Fast et al. 2001; Harper and Keeling, 2003). On the other hand, the six nuclear-encoded protein genes phylogeny provides low support of the branch hatophytes and cryptophytes, but strongly supports existence and mutual relationship of stramenopiles and alveolates (Harper et al., 2005). Recent phylogenomic studies even show the sisterhood of stramenopiles and alveolates with Rhizaria (Burki et al., 2007; Hampl et al., 2009). This results lead to destabilization of chromists, while cryptophytes + haptophytes and alveolates + stramenopiles are obviously related groups (Patron et al., 2007, Frommolt et al., 2008).

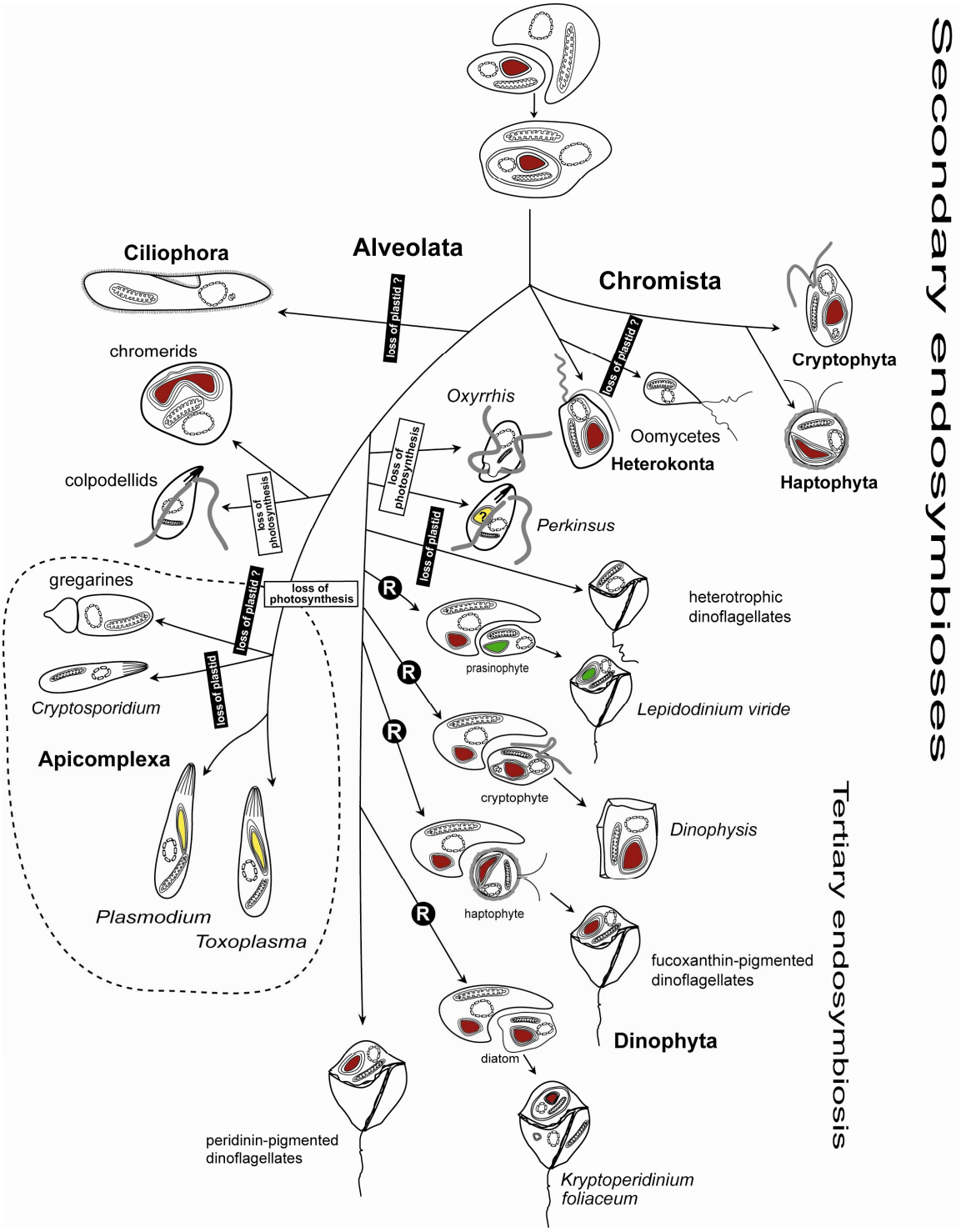


Figure 3. Evolutionary history of Apicomplexa and other chromalveolates (Oborník et al., 2009). It is expected that chromalveolates appeared due to a single secondary endosymbiotic event. Their evolution is accompanied by numerous losses and replacements of the plastid, including tertiary endosymbioses by engulfing other photosynthetic chromalveolates. According to our phylogenetic analyses, chromerids branch as early apicomplexans, closely related to the heterotrophic colpodellids. Gregarines appear on the root of the parasitic apicomplexans. R in a black circle indicates plastid replacement..

The other part of dinoflagellate diversity includes organisms that had lost photosynthesis and are thus secondarily heterotrophic (Sanchez-Puerta et al., 2007). Similarly to those, the complete phylum Ciliophora (Alveolata) constitutes another branch without photosynthesis. No ultrastructural evidence of plastid being present in ciliates was found in the cell. However, sixteen genes of plastid origin were identified in ciliate genomes, which suggest that even ciliates could have photosynthetic ancestors in agreement with the chromalveolate hypothesis (Reyes-Prieto et al., 2008). These free living protists are characterized by two differently sized nuclei and cilia on the cortex surface. It is believed that ciliates lost their plastid early after acquisition through secondary endosymbiosis. Another evidence supporting this possibility lies in the carotenoid synthesis: the presence of zeaxantin epoxidase (ZEP) in the genome of ciliate *Tetrahymena thermophila*. Product of this gene works exclusively for the photosynthetic pigment synthesis (Coesel et al., 2008).

Obligate intracellular parasites from the phylum Apicomplexa represent in addition to ciliates, dinoflagellates and chromerids the fourth group belonging to alveolates. They are responsible for a number of serious diseases affecting a broad range of animal hosts, including humans. One of the most lethal infectious diseases today, malaria, is caused by the apicomplexan *Plasmodium falciparum* (Carlton et al., 2008). Other common infections include toxoplasmosis, cryptosporidiosis, and babesiosis (Gubbels et al., 2008; Šlapeta, 2006; Hunfeldet al., 2008). It was found that most of apicomplexans possess a relic non-photosynthetic plastid, seemingly homologous to the photosynthetic organelle. This plastid, called the apicoplast, was first identified in *Plasmodium* and *Toxoplasma* and closely related species (McFadden et al. 1996; Wilson et al. 1996; Kohler et al. 1997), but no significant evidence for the plastid presence has been found in gregarines and genus *Cryptosporidium* (figure 4). Apicomplexan plastid genome has reduced to a 35 kb circle that does not contain any photosynthetic genes. Ultrastructural and molecular data have confirmed that this organelle originated from secondary endosymbiotic event (Waller et al. 1998). Recently, evidence for photosynthetic ancestry of Apicomplexa was discovered. The alveolate symbiont isolated from stony corals in Australia and named *Chromera velia* is according to molecular phylogeny closely related to apicomplexan parasites, and chromeran plastid shares common origin with the apicoplast (Moore et al., 2008; Oborník et al., 2009).

3.3. Plastid of Chromists

To explore evolution of complex plastids, plastids of Chromista in particular provide several highly informative characters. The most impressive one is the subsequent reduction of a secondary endosymbiont within a eukaryotic cell. Several intermediates can be found that illustrate the particular steps: plastids of cryptophytes contains, similarly to chlorarachniophytes with green algal plastid, remnant of eukaryotic nucleus called nucleomorph. This small nucleus is placed in the periplastidal space between the primary and secondary membranes, exactly in the place of cytosol of the former eukaryotic symbiont (Archibald 2007). The nucleomorph of cryptophytes contain three small chromosomes and the translation in the periplastid space is still running (Hansmann, 1988). Moreover, nucleomorph encodes 18S ribosomal DNA (rDNA) genes, which are evolutionarily distinct from those present in the host nuclear genome (Douglas et al., 1991) and are similar to those found in rhodophytes (Matsuzaki et al., 2004). In other chromalveolates such as, haptophytes, stramenopiles and apicomplexans, nucleomorph is not present.

The origin of multiple membranes of chromalveolate plastids has been intensively studied. It was suggested that outermost membrane originated from fagocytic vacuole surrounding the algal prey (latter endosymbiont), the second inner membrane is a remnant of the plasma membrane of the endosymbiont that the inner two membranes are a remnants of the primary plastids membranes (Gibbs 1981; Cavalier-Smith 2000; Ishida et al., 2000) The outermost membrane of the plastid is connected to the endoplasmic reticulum (ER) of the host in haptophytes and stramenopiles, but does not appear to be such in Apicomplexa. In peridinin-containing dinoflagellates, only three membranes surround the plastid, the outermost of which is not connected to any internal host membranes (Gibbs, 1979; van Dooren et al., 2001).

The loss of photoautotrophy was also noted in Stramenopiles, which contain lineages of heterotrophs and parasites (figure 3): Oomycota, Labyrinthomorpha and Opalinea (Nishi et al., 2005; Tsui et al., 2009). With respect to chromalveolate hypothesis they arose from photosynthetic progenitor, but had lost photosynthesis or even the plastid. Phylogenetic analyses reveal that some genes of parasitic genus *Phytophthora* (Oomycota) are likely to be of plastid origin (Tyler et al., 2006; Jiroutová et al., 2007). Also the photosynthetic plastids has dramatically modified during evolution. The chromist plastids lost rhodophyte phycobilins, bur retained chlorophyll *a* and β-carotene, and acquired various forms of chlorophyll *c* and carotenoid fucoxantin (Jeffrey and Wright, 2005). These pigments

permitted light harvesting outside the absorption spectrum of rhodophytes and chlorophytes. The xanthophylls diadinoxantin and diantoxantin protect the photosystem of the chromist lineages from harmful effects of light saturation (Falkowski and Raven, 2007) (*see chapter 6.2*).

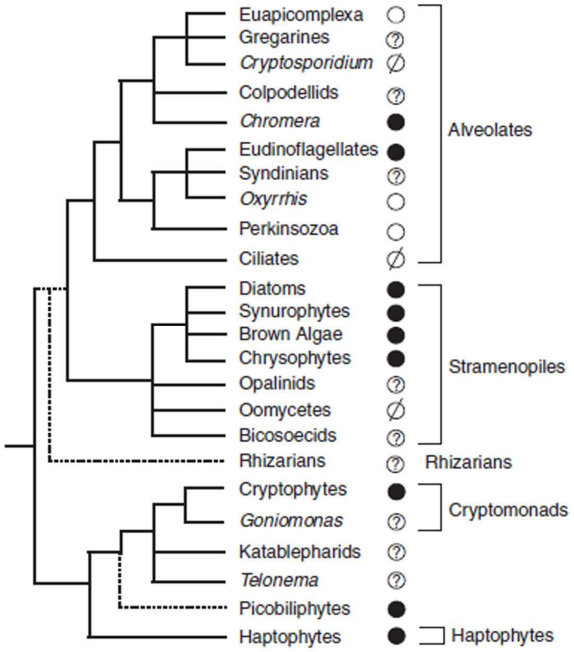


Figure 4. Schematic tree outlining the current hypotheses of chromalveolate relationships (Keeling, 2009; modified). Regions of the tree for which no consistent relationships have emerged are indicated by polytomies (e.g. lineages at the base of apicomplexa and dinoflagellates, and the branching order of most subgroups of stramenopiles). Other more tenuous relationships are indicated by dashed lines. The monophyly of alveolates and stramenopiles is consistently found, but needs further evidence. The relationship between Rhizaria and subgroups of chromalveolates is an emerging observation of great interest that needs to be further refined and would be much stronger if other supporting characters were found.
black circle = photosynthesis; white circle = loss of photosynthesis; crossed circle = loss of plastid; question mark = not finisfed results for subgroup

Extensive sequencing of the chromist plastids genomes allows their genomic comparisons. Plastid genomes from cryptophyte *Guilardia theta* (Douglas and Penny, 1999), haptophyte *Emiliania huxleyi* (Sánchez puerta et al., 2005) and three diatoms *Thalassiosira pseudonana* (Armbrust et al., 2004), *Odontella sinensis* (Kowallik et al., 1995) and *Phaeodactylum tricornutum* (Oudot-Le Secq et al., 2007) share the compact structure and complete lack of introns. When taking into account that contrary to rhodophyte plastid genomes which carry more genes than any of the chromistan plastid, the plastid gene transfer to the host nucleus had obviously to continue during secondary endosymbiosis (Oudot-Le Secq et al., 2007). There is the small difference in gene content among these chromist groups. Surprisingly, *G. theta* has 20 genes that have not been found in all three diatoms; however most of those genes are on the plastid genome of rhodophytes (Reith and Munholland, 1995). Also the presence of the nucleomorph in *G. theta* suggests that the integration of the secondary endosymbiont has not reached the same state as in the heterokonts or haptophytes (Oudot-Le Secq et al., 2007). In contrast, the haptophyte *E. huxleyi* has the smallest number of plastid genes of the chromists plastid genomes sequenced so far (Sánchez puerta et al., 2005). It appears that gene transfer has progressed further in the haptophyte lineage than in

the heterokonts. On the other hand, the *E. huxleyi* genome has nine genes not found in any of the diatom genomes (Oudot-Le Secq et al., 2007).

It is obvious that the process of endosymbiotic gene transfer also occurred during secondary endosymbiosis, from both, the nuclear genome of the engulfed alga that contained the genes encoding plastid targeted proteins required for plastid function and from the secondary plastid genome (McFadden 1999). During annotation of both diatom genomes the gene napped during endosymbiotic gene transfer coding for *psb28* was found in *T. pseudonana*. In this diatom two versions of the gene - plastid and nuclear encoded one were found. Moreover, they seem to be closely related to each other (Armbrust et al., 2004). Predicted ER signal sequence was detected at the N-terminus of the nuclear encoded *psb28* which suggests that this protein is putatively targeted to the diatom complex plastid. Such targeting of this protein was also experimentally confirmed (Jiroutová et al., unpublished results). Thus it was suggested that the *psb28* gene is in the in progress of endosymbiotic gene transfer, where the plastid copy had not been deleted yet. On the other hand, more derived pennate *P. tricornutum* possesses only plastid copy of the gene (Oudot-Le Secq et al., 2007), which suggests that the gene transfer event can be very distinct between even related lineages.

4. Protein import into the complex plastid

It has already been mentioned that many genes are transferred from the endosymbiont to the host nucleus and their products have to be targeted to the place of action. This represents a critical step for plastid establishment and the process is called endosymbiotic gene transfer (EGT) (Delwiche 1999; Martin and Herrmann 1998). It has been postulated that plastid-derived EGT played a significant role in plant nuclear-genome evolution, with 18% of all nuclear genes in *Arabidopsis thaliana* displaying a cyanobacterial origin with about one-half of these recruited for non-plastid functions (Martin et al., 2002). This gene transfer took place over a timescale of hundreds millions years, and multiple evidence suggests that it is still an ongoing process (Millen et al., 2001; Adams et al., 2002). Moreover, process of gene transfer from plastid to nucleus was experimentally reconstructed in tobacco plant (Huang et al., 2004; Stegemann and Bock 2006).

The pathways for intracellular protein transport across membranes have been the focus of interest in recent years. Contrary to animal and plant cells, chromalveolates have developed unique protein transport mechanisms for targeting of nuclear-encoded proteins over multiple membrane of the complex plastid (Lang et al., 1998). As mentioned above, there is a major

structural difference between primary and secondary plastids. All primary plastids are surrounded by two envelope membranes, while secondary plastids are surrounded by more than two membranes, usually three or four (e.g. Achibald and Keeling 2002; Battacharya et al., 2003). Owing to this complexity, the mechanism of targeting to secondary plastids requires more signals to cross the boundary membranes and is less understood (Ishida 2005).

Since the protein targeting in secondary endosymbiotic algae is actually modification of the targeting mechanism to primary photosynthetic organelles, we have to mention the primary targeting first. In other words, secondary algae utilize the targeting to primary organelles, because the secondary plastid shares origin with the primary one which represents one of the components of the complex plastids. There are two known protein-import apparatuses (translocons) in plants: the Toc and Tic complexes that are in the outer and inner membranes of the primary plastid envelope, respectively (Soll and Schleiff 2004). When the protein precursors reach the surface of the outer membrane of the primary plastid, the transit peptides (TP) are recognized by and interaction with one of the Toc components. The precursors are transported across the membrane through Toc75, the protein channel in the Toc complex, and then the transit peptides make contact with the Tic complex of the inner membrane. The precursors are subsequently imported through the Tic complex into the stroma of the plastids, where the transit peptides are cleaved off by stromal processing peptidases (Jarvis and Robinson 2004). Therefore, secondary endosymbiosis is coming with at least one important evolutionary innovation: combination of the targeting to primary plastid (Tic and Toc) with the targeting to the endoplasmic reticulum. The targeting signal thus consist of at least two parts, ER signal peptide (SP), which target the protein into ER and over the outermost membranes and conventional chloroplast transit peptide that get the protein over inner membranes of the complex plastid (McFadden 1999).

In Apicomplexa and peridinin-pigmented dinoflagellates, the outermost plastid membrane is not connected to the endomembrane system of the cell. The proteins are then transported to the plastid in vesicles derived from the secretory system that afterwards fuse with the outermost membrane. Apicomplexan protein transport might be direct, whereas in peridinin-pigmented dinoflagellates the Golgi apparatus is probably involved. In Chromalveolates, with the possible exception of peridinin-pigmented dinoflagellates, transport across the second outermost membrane could be mediated by an ER-associated degradation (ERAD)-derived translocon machinery (Hempel et al., 2007). Transport across the inner membrane pair might involve a Toc/Tic-based translocon system (Kroth and

Strotmann 1999). However, no Toc components have been identified yet in any of these organisms, indicating unknown mechanism of protein transport across the third membrane.

In the case of cryptophytes, stramenopiles and haptophytes, the ER is continuous with the outermost membrane of the complex plastid, and contains 80S ribosomes on the plasma face. The protein transport across this membrane thus occurs co-translationally. In the ER lumen the SP is cleaved off, exposing the TP for further targeting. The next step, how proteins get across the remaining three membranes to reach the stroma of plastids, is similarly to alveolates less understood. However, diatoms as the members of stramenopiles and heterokonts became excellent models for protein-import experiments. For this purpose, *in vitro* isolation of intact plastids has been done for two heterokont algae: the diatom *Odontella sinensis* (Wittpoth et al._1998) and the raphidophyte *Heterosigma akashiwo* (Chaal et al. 2003). Interestingly, *in vitro* import of heterokont precursors into higher plant plastids demonstrated that both diatom and raphidophyte precursors without the SP sequences were successfully imported into and sometimes correctly processed in isolated pea plastids (Lang et al. 1998; Chaal and Green 2005). So the protein-import mechanism may, at least in a part, be similar between higher plant and heterokont plastids (Lang et al. 1998; Chaal et al. 2003).

Furthermore, *in vivo* experiments showed the native bipartite targeting sequence itself directs green fluorescent protein (GFP) into the plastid of *Phaeodactylum tricornutum* (Apt et al., 2002; Kilian and Kroth, 2005). Interestingly, also heterologous presequences from the diatom *Odontella sinensis* (Kilian and Kroth 2005) or from the dinoflagellate *Symbiodinium* sp. (Lang, 2000 in Gruber et al., 2007) and the cryptophyte *G. theta* (Gould et al. 2006a) were able to direct GFP into the plastid of *P. tricornutum*, indicating similarities between the plastid protein import machineries in cryptophytes, dinoflagellates and diatoms.

Nuclear-encoded proteins are translated directly into the lumen between the first and second outermost membranes of the secondary plastid in diatoms, where they are N-terminally processed by cutting out SP. The further experiments of protein import into the plastid of *P. tricornutum* revealed a crucial role of the first amino acid of the transit peptide after the SP cleavage. According to genome database of sequenced *P. tricornutum* and *T. pseudonana* most proteins targeted to the plastid have SP cleavage site between alanine (A) and phenylalanine (F). Signal peptide consist about 20 amino acids, mostly with hydrophobic character. Pheneylalnine (F) at the first possiton of was verified by *in vivo* experiments, in which bipartite transit peptide was fused with GFP (Apt et al., 2002). When the first amino acid of TP was changed, the transit peptide is also functional with the other two aromatic amino acids: tyrosine (Y) and tryptophan (W), but also with leucine (L) (Kilian and Kroth,

2005). This procedure demonstrates that final localization is determined by the presence or absence of this aromatic amino acid in the transit peptide (Gruber et al., 2007). Furthermore the sequence consensus “ASAFAP” was proposed as the most likely abundant amino acids at the cleaveage site of SP and TP. However, the motif itself without the leading TP does not target the protein into plastid domain and creates the “blob-like” structures instead (Kilian and Kroth, 2005; Gruber et al., 2007).

This could imply that the hypothesis in which Cavalier Smith (1999) proposed a relocated Toc complex as a translocon within the second outermost membrane is correct. However, screening of genomic data for cryptophytes, diatoms and apicomplexans did not lead to the identification of any Toc components encoded by these organisms (McFadden and van Dooren, 2004, Hempel et al., 2007). It is possible that the protein-import apparatuses and mechanisms have been largely modified in the heterokont lineage since the secondary endosymbiotic plastid acquisition (Ishida 2005).

5. Diatom genomics and molecular biology approaches

Various aspects connected to molecular mechanisms in such important ecological players as diatoms were largely unexplored for a long time. Substantial progress has been made when the whole genome of the centric diatom *Thalassiosira pseudonana* (Armbrust et al., 2004) was sequenced, and recently the completion of the genome of the pennate diatom *Phaeodactylum tricornutum* and comparison of both genomes was finished (Bowler et al., 2008). To understand the information within diatom genomes Expressed Sequence Tag (EST) database was generated; over 130, 000 ESTs were sequenced from *P. tricornutum* (Maheswari et al., 2005; Montsant et al., 2005). Lately, the EST library for *T. pseudonana* was updated (Mock et al., 2008; Maheswari et al 2009). This significantly improves the identification of predicted genes, their structural annotation and analyses of expression. Moreover, the molecular tools required for functional genomics, such as genetic transformation, or gene silencing are available (Apt et al., 1996; Falciatore et al., 1999, De riso et al., 2009).

5.1. Sequencing of diatom genomes

The nuclear, plastid and mitochondrial genomes of marine centric diatom *Thalassiosira pseudonana* was sequenced by whole-genome shotgun method. The 34.5 Mb nuclear genome of the alga was predicted to contain over 11,000 genes distributed on 24

diploid chromosomes. Functions of about half of the genes in this genome could not be assigned solely on the basis of sequence similarity to homologues in other organisms, and these genes are supposed to encode proteins involved in diatom-specific processes. The other half of the predicted genes has similar alignment scores to their closest homologs in other studied eukaryotic organisms (Armbrust et al., 2004). Due to the low similarity the diatom specific genes cannot be compared with any other sequenced organisms so far. Genes coding for transporters for uptake of all inorganic nutrients and genes involved in silica metabolism seem to be the most valuable. Considerable progress has been made with silicon uptake by special proteins from the surrounding water, the storage and processing of silicon before cell division was also studied. On the other hand, the discovery and characterization of remarkable biomolecules such as silaffins, polyamines and silacidins in the siliceous cell walls of diatoms strongly impacts the growing field of materials synthesis (Sumper and Brunner, 2008). *T. pseudonana* is equipped with a fully silicified cell wall exhibiting porous nanopatterns, a feature that is characteristic for most diatom species, while pennate *P. tricornutum* can survive without silica (Round et al. 1990).

Unlike photosynthetic plants, diatoms possess a complete urea cycle (Armbrust et al., 2004). The urea cycle was previously thought to be restricted to organisms that consume complex organic nitrogen compounds and excrete nitrogenous waste products. But diatoms seem to use it for nitrogen, polyamine and energy storage, or for cell signalizing pathway (Allen et al., 2006). Diatoms also generate chemical energy from the breakdown of fat trough metabolizing fatty acids in mitochondrion during the whole cell cycle, which is typical feature for animals. The combination of this ability with “a diatom starch” chrysolaminaran metabolism probably allows diatoms to survive for long periods of darkness, which occurs at the seas near the poles, where they live in high density (Armbrust et al., 2004). For this reason, alga containing fat is also excellent food for the zooplankton; therefore diatoms are crucial component of the food webs in the ocean (Wichard et al., 2007).

Contrary to *T. pseudonana*, which is a centric diatom with radial cell symmetry and non-motile planctonic lifestyle, the second sequenced diatom *Phaeodactylum tricornutum* belongs to pennates. They display a range of different features, including their bilateral symmetry. Although they probably diverge 90 million years ago, the pennates represents the most diversified branch, which constitutes the major component of phytoplankton in pelagic and benthic habitats. In spite of this fact, genome structures of *T. pseudonana* and *P. tricornutum* are quite different and substantial fractions of genes are not shared by these two representatives (Figure 5a). The completed *P. tricornutum* genome is approximately 27.4

megabases (Mb) in size, which is slightly smaller than *T. pseudonana*. *P. tricornutum* is predicted to contain fewer genes and shares 57% of its genes with *T. pseudonana*, of which 1,328 are absent from other sequenced eukaryotes (Bowler et al., 2008). Compared to other eukaryotes, huge diatom diversity has been generated by unexpected mechanisms, different from known genome duplication events. It seems to be expansion of gene families and gain of introns in centric diatoms, and huge expansion of diatom-specific retrotransposable elements in the *P. tricornutum* genome. These mobilization elements also appear to have expanded in other pennate diatoms, so they may have been a significant driving force in the generation of pennate diatom diversity through transpositional duplications and subsequent genome fragmentation (Bowler et al., 2008). Interestingly, the genomes of diatoms consist of many genes of bacterial origin (Figure 5b). The source is of course endosymbiotic gene transfer from mitochondria and plastid, but also from the bacteria, who cooperates and lives with diatoms in the seas for long time. Horizontal gene transfer creates together with secondary endosymbiosis genuine mixture of genes suitable for successful living in the ocean (Bowler et al., 2008). Recently, it has been shown that diatom genomes contain high fraction of genes displaying origin in green plastids, although the diatom plastid had obviously originated from engulfing rhodophyte (Moustafa et al., 2009).

a) **b)**

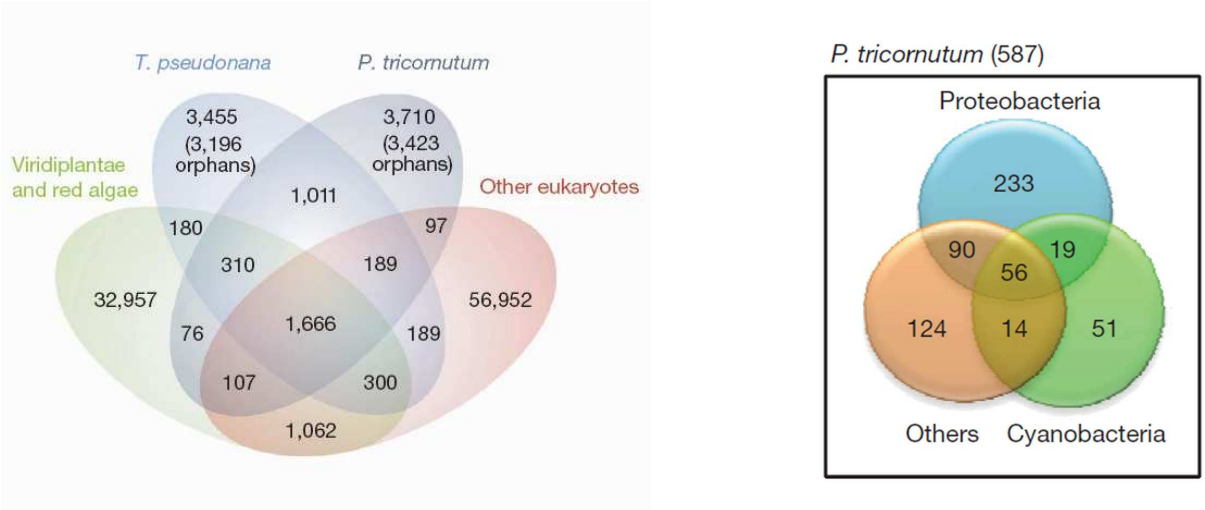


Figure 5. a) Venn diagram on the left showed shared and unique gene families (Bowler et al., 2008). Representation of shared/unique gene families in *P. tricornutum*, *T. pseudonana*, Viridiplantae (plants and green algae) and red algae, and other eukaryotes (that is, other chromalveolates and Opisthokonta (fungi and metazoa). In addition to the total number of gene families specific to *P. tricornutum* and *T. pseudonana*, the number of families consisting of a single gene (denoted 'orphans') is also indicated. For example, of the 3,710 gene families that are only found in *P. tricornutum*, 3,423 consist of single copy genes whereas 287 gene families have at least two members.

Figure 5. b) Venn diagram on the right display summary, which bacterial classes are most related to the bacterial genes identified in *P. tricornutum.* (Bowler et al., 2008). In this case, the Venn diagram indicates the number of trees in which the designated taxa occur within the same clade or in a sister clade of *P. tricornutum.*

The availability of diatom genome sequences provides opportunities for large scale gene expression analyses. Rapidly growing number of sequence information requires accurate and precise gene annotation, especially in the case of model organisms. Generally, expressed sequence tags (ESTs) help with gene identification, their structural and functional analyses. For *P. tricornutum* with high ecological and evolutionary importance, sixteen cDNA libraries constructed in different growing conditions has been made and covered with over 130,000 ESTs. A smaller library was also generated for *T. pseudonana*, where seven libraries were derived from cells grown in different nutrient and stress regimes and 77 000 ESTs were sequenced. This would facilitate exploration of diatom responses to ecologically relevant conditions. Cells grown in different conditions can provide dataset for comparative analyses. It also aided the study of key signaling and regulatory pathways, nutrient and silica metabolism and prediction of targeting signal into plastid (Montsant et al., 2007; Mock et al., 2008; Lopez et al., 2005; Kilian and Kroth 2005).

5.2. Molecular biology approaches

Although *T. pseudonana* has been a model organism for physiological studies of widely distributed species, a range of reverse genetics tools have been generated mainly for pennate *P. tricornutum* (Siaut et al., 2007; De riso et al., 2009), therefore making it a good model for functional genomic studies. Recently, research in diatom biology has entered the postgenomic era. Before the sequence data were available, studies had been performed in diatoms by northern blotting, RNA dot blot (Oeltjen et al., 2004) and RT-PCR (Leblanc et al., 1999). However, in last ten years, quantitative RT-PCR has become the method of choice because it combines high-throughput procedures with the accurate expression profiling of selected genes that can accurate the EST of microarray information (Siaut et al., 2007).

The distinctive biological features of diatoms cannot be analyzed using other model organisms. To uncover the molecular basis of the diatom's exceptional physiological capabilities, genetic transformation need to be applied. Previously, methods for stable genetic transformation have been established for the diatom *P. tricornutum* (Apt et al. 1996) and more recently also for *T. pseudonana*, promising more studies of silica flusture formation were made (Poulsen et al., 2006). Among species that have not been completely sequenced are stabily transformable pennates such as *Navicula saprophila* (Dunahay et al. 1995), *Cylindrotheca fusiformis* (Fischer et al. 1999), and centric diatom *Cyclotella cryptica* (Dunahay et al. 1995). For another centric diatom, *Thalassiosira weissflogii*, transient

transformation could be performed (Falciatore et al., 1999). In all these species the transformation system is based on helium accelerated particle bombardment of exogenous DNA followed by selection, with transformation efficiencies still low, on the order of 10^{-5}–10^{-6}/g of plasmid DNA. In each case from one to a few copies of the plasmid are integrated randomly into the diatom genome, with integrations being stable even in the absence of selection pressure (Montsant et al. 2005b). However, these methods have mainly been applied to functional studies with *P. tricornutum* (Falciatore et al., 1999; Zaslavskaia et al. 2000, Apt et al. 2002, Kilian and Kroth 2005). Attention was particularly paid to the role of specific domains of the sequence associated with nuclear-encoded plastid polypeptides in the multi-step transport process that delivers these proteins to the plastid. By this study, the very exact motif of amino acid sequence of transit peptide had been identified. By mutations of the cleavage site of SP, the other aromatic amino acids can substitute postulated phenylalanine (F) without the change of function (see chapter protein import for details). Experimental proteins are fused with reported fluorescence protein and can be easily distinguished by fluorescence or confocal microscopy (Kilian and Kroth 2005, Gruber et al., 2007). Moreover, other address sequences can be intensively studied to specify their intracellular localization in mitochondria, ER, nucleus, transport vesicles or diatom surface (Siaut et al., 2007).

Another powerful tool in functional genetics is with no doubt gene silencing via RNA interference (RNAi). Although first search in diatom genomes had not found the canonical members of RNAi machinery, experiments show that gene silencing in *P. tricornutum* is possible. Successful silencing was done on GUS reporter, expressed in transgenic line, as well as knockdown of endogenous phytochrome and cryptochrome (De riso et al., 2009).

The diatom genomic repository is rapidly expanding with several sequencing projects. For example, the genomes of two additional pennate diatoms, *Pseudo-nitzschia multiseries* and *Fragilariopsis cylindrus*, are currently nearing completion at JGI, together with accompanying EST collections (Armbrust, 2009). For instance, *P. multiseries* belongs to small fraction of this genus, which can produce a toxic domoic acid (Bates and Trainer, 2006). On the other hand, *F. cylindrus* is adapted for living in polar sea near the poles (Mock and Hoch, 2005). The common feature that makes connection of these pennates with the *P. tricornutum* is production of ferritin, an iron-storage molecule that protects the cell against oxidative stress (Marchetti, et al., 2008). The enhanced iron storage provided by ferritin in *P. multiseries* probably underlies its numerical dominance in the massive diatom blooms that result from iron fertilization and helps to explain the importance of raphid diatoms in

regulating the flux of CO_2 into surface waters (Marchetti, et al., 2008; Armbrust, 2009, Bowler et al., 2009).

6. Evolution of metabolic pathways in diatom plastid

Composition of metabolic pathways located in the secondary plastid is tightly associated to the process of secondary endosymbiosis. This process increase the complexity of intracellular localization and phylogenetic distribution of particular enzymes and provides information concerning former symbiotic participants. Accumulating knowledge reffering to the targeting of nuclear encoded proteins helps to improve bioinformatic tools for prediction of signal and transit peptides. In addition, the localization of specific enzymes can be confirmed by *in vivo* assay in diatom by using the protein fused with fluorescent marker. To trace the origin of the protein, molecular phylogeny approaches are applied and place of action of metabolic pathways can be discussed and compared.

6.1. Carbohydrate metabolism

Carbon fixation through the Calvin cycle and synthesis of carbohydrates represent the most fundamental metabolic pathways for the photosynthetic organisms. Despite their diversity and abundance, the carbon fixation in diatoms is extreemely effective, possibly due to secondary plastid acquisition. Carbon export can influence climate change through uptake and sequestration of atmospheric CO_2 (Granum et al., 2005). Recently, the role of diatoms in mitigating atmospheric CO_2 concentrations is of special interest, particularly thanks to the rising levels of this “greenhouse gas” and consequent global climate changes. A significant fraction of the organic carbon generated by diatoms remains in the upper ocean and supports production by higher trophic levels and bacteria. Although the typical sea water contains about 10 mM CO_2 (Reibesel et al., 1993), ribulose-1,5- bisphosphate carboxylase/oxygenases (Rubisco) from diatoms have half-saturation constants for CO_2 of 30–60 mM (Badger et al., 1998). Surprisingly, the carbon concentration is not CO_2 limited in diatoms, so CO_2 concentrating mechanism could resemble C4 acid pathway in land plants (Figure 6.). Unlike the higher plants were the C4 mechanism is divided into the leaf tissue, the diatoms have only single cell available. On the other hand, the enzymes working in the C4 pathway have been found in many copies in diatom genomes and can represent modification of efficient carbon fixation (Kroth et al., 2008; Bowler et al., 2008). The putative pathway is probably distributed in membranous system inside the diatom cell.

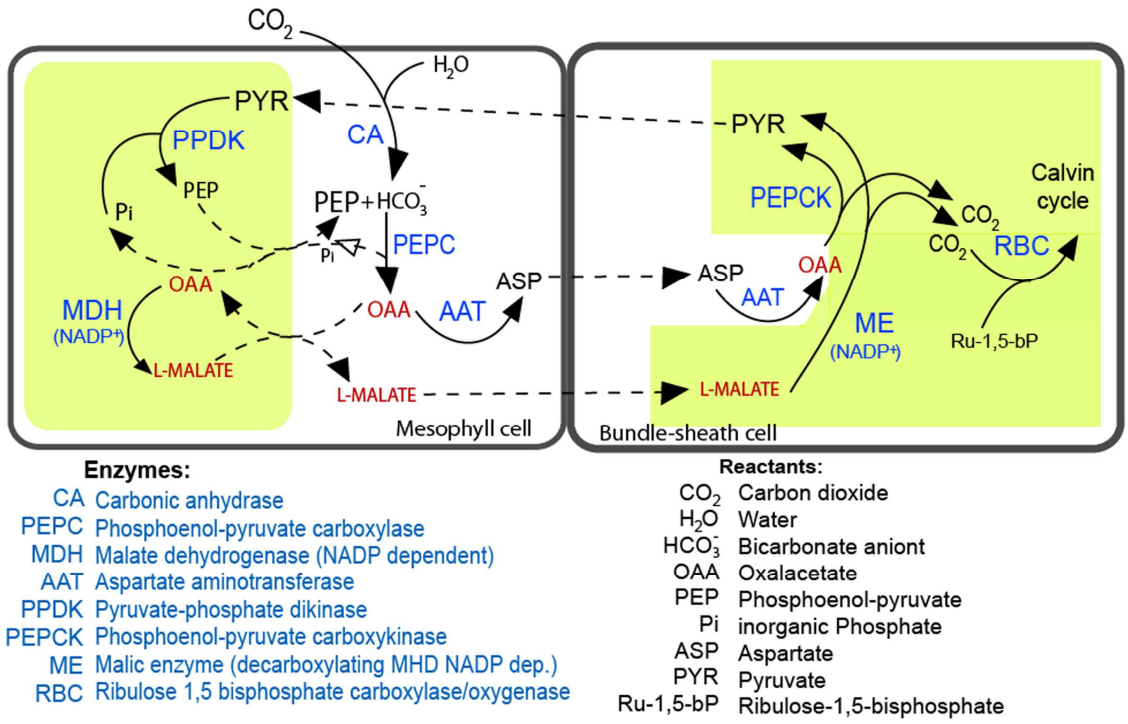


Figure 6. C4-dikarboxylic acid pathway in plants (KEEG pathway, modified). In mesophyl cell connected to stomata CO_2 gas is turned to bicarbonate ion by CA, to be bounded to PEP, when the phosphate is envolved and C4 Oxalacetate is rised done by PEPC. OAA is moved by membrane transporter into plastid to be turned to L-malate by NADP dependent MDH or alternatively transaminated to Aspartate. L-malate or Aspartate are able to transport to the Bundle-shelth cell, were L-malete is decarboxylated to pyruvate by malic enzyme. Similarly Aspartate is deaminated to oxalacetate to be decarboxylated by PEPCK. By process of decarboxylation CO_2 is envoleved into the plastid stroma and rising concentration of CO_2 turnes the reaction constant of RBC for carboxylation of Ru-1,5-bP. The pyruvate is regenerated by returnig back to mesophyll cell, and by bounding of phosphate PEP is rised done by PPDK.

The carbon concentrating mechanism that facilitates an ample flux of CO_2 to photosynthesis under CO_2 limitation is known to occur in the few marine diatoms (Johnston and Raven, 1996; Badger et al., 1998). The membrane abundance of carbonic anhydrase (CA), the enzyme with high affinity to dissolved inorganic carbon HCO_3^-, playes the main role in carbon concentrating mechanism in aquatic environment. As the dissolved carbon dioxide, bicarbonate ions are more common form for water system, but require active transport inside the cell compared to inert CO_2. CAs efficiently fixate CO_2 in a CO_2-limited environment by catalytically dissolving of HCO_3^- at the very near of cell membrane (Karlsson et al., 1998). The other active CAs are intracellular and serve for the reactions for rising CO_2 or HCO_3^- for the catalytic needs. The overall sequence similarities among CAs are rather low and they are commonly identified by the presence of conserved domains and by their biochemical properties such as the requirement of Co, Cd or Zn elements (Park et al., 2007). Seven CAs are predicted for *P. tricornutum*. Two CAs are related to the β-type and show similarity to CAs found in both plants and prokaryotes (Kroth et al., 2008). One of these proteins is located in the plastid. This localization was demonstrated by GFP fusion proteins,

the other CAs possesses an ER signal peptide (Tanaka et al., 2005). The other identified CAs belong to the α-family and all are putatively targeted to the ER or periplasmic space of *P. tricornutum*.

The C4 dikarboxylic pathway is represented by phosphoenolpyruvate (PEP) processing enzymes: phosphoenolpyruvate carboxylase (PEPC), phosphoenolpyruvate carboxykinase (PEPCK) and pyruvate phosphate dikinase (PPDK) (Figure 6.). Each of these enzymes is required for C4-metabolism, but also for C3 metabolism, and was identified in the genomes of both sequenced diatoms *T. pseudonana* and *P. tricornutum*. The analysis of gene transcription for the putative C4- related genes in *T. pseudonana* indicated that the gene encoding PEPCK was up-regulated about 1.5 fold under reduced CO_2 concentrations, whereas expression of genes encoding PEPC and PPDK were unaffected (Granum et al., 2005). On the other hand, PPDK from *P. tricornutum* was identified and includes bipartite signal peptide, so PEP is putatively generated in the plastid (Kroth et al., 2008). Two genes encoding PEPC have been identified in *P. tricornutum*. The first one has a high degree of similarity with the PEPCs from chlorophytes and possesses a signal peptide at the N-terminus of the enzyme, but a transit peptide was not found. This suggests that the protein is targeted either to the ER or to the periplastidic space of the complex plastids (Gruber et al., 2007). The second PEPC is similar to PEPCs from bacteria and contains a predicted mitochondrial targeting presequence. Surprisingly, *in silico* analysis of both *P. tricornutum* and *T. pseudonana* indicated that the decarboxylating enzymes PEPCK, malate dehydrogenase (MDH) and NADH dependent malic enzyme (ME) do not possess plastid targeting sequences, but rather mitochondrial ones. Moreover, there is no evidence that malate or oxalacetate (OAA) transporters in these organisms are localized to plastid membranes (Kroth et al., 2008).

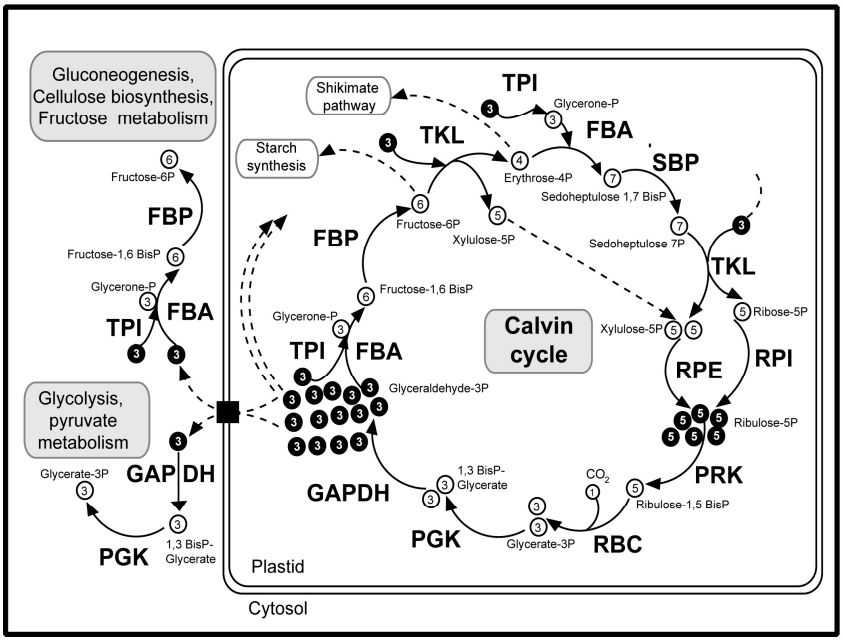
These results in *P. tricornutum* and *T. pseudonana* suggest that generation of CO_2 requires subsequent decarboxylation steps that do not occur in the plastid. Decarboxylation of malate to generate CO_2 may occur within mitochondria, which are often in close proximity with plastids. However, any CO_2 molecules released from the mitochondrion must get across six membranes to enter the plastid stroma. CO_2 would be converted to HCO_3^- by CA activity during movement between the mitochondria and plastids, thereby reducing at least part of the elevated CO_2 concentration. Together with the evidence of bicarbonate transporters in the plastid membrane, HCO_3^- is first fixed and then formed again, thereby dissipating ATP for PEP formation. In *P. tricornutum* and *T. pseudonana* PEP might be exported from the plastid and carboxylated in the ER by PEPC leading to a micro-compartmentalization of the enzyme to the outer chloroplast membranes. Rising OAA is transported to mitochondria, where

decarboxylation of oxalacetate and malate occur, but CO_2 is transported for utilization by Rubisco in the plastid (Kroth et al., 2008). Although the most C4-plants carry out the first carboxylation step in the cytosol, in diatoms it might be in an extra membranous system of ER and periplastidic space.

In photosynthetic eukaryotes, carbon fixation is exclusively done by enzyme Rubisco involved in the Calvin cycle, which is a typical plastid pathway. However, many enzymes of the Calvin cycle have isoforms working in glycolysis and gluconeogenesis not restricted to plastid. Generally, the carbohydrate pathway provides the precursors for starch, lipid and amino acid biosynthesis. Although most of the Calvin cycle enzymes in diatoms are very similar to those in land plants (Figure 7), there are indications that they may be differently regulated by light (Michels et al, 2005).

The small and the large subunit of Rubisco are the only plastid genome encoded Calvin cycle enzymes in *P. tricornutum* and *T. pseudonana* (Oudot-Le Secq et al., 2007). The rest of Calvin cycle enzymes are encoded in the nucleus and targeted to the diatom plastid. Although the cyanobacterial origin for this enzymes was proposed, several Calvin cycle enzymes are of different origin than the endosymbiont`s genome (Patron et al., 2004). Reduction of carbohydrates catalyzed by the plastid targeted phosphoglycerate kinase (PGK) and glyceraldehyde- 3-phosphate dehydrogenase (GAPDH) represents next two enzymatic steps after Rubisco carbon fixation.

Figure 7. Calvin cycle (Martin and Schnarrenberger, 1997; modified)



Calvin cycle is also known as reducing pentose phosphate pathway. The figures in circles mean number of bouding carbons in the molecule. Full line arrows mean enzymatic step, while dashed line arrows explain transport. The starting compound is the pentose Ribulose-5-phosphate Ru5P (5), which is phosphorylated to Ribulose-1,5 bisphoshate. The utilization od CO_2 is mediated by Rubisco (RBC) to give a rise to two trioses. Thereafter, they are phosphorylated by PGK and reduced by GAPDH to Glyceraldehyde-3-phosphate (here marked as 3). GA3P is accumulated within the plastid and part of this product is transported to cytosol to synthesize hexose. From six Ru5Ps 12 GA3Ps are made, two of them are needed for synthesis of 1 molecule of glucose. However, 10 GA3Ps are going back under carbon donor regeneretion through the conversion of monosacharides to give rise of new 6 Ru5Ps.

Enzymes of Calvin cycle

PRK	Phosphoribulokinase
RBC	Ribulose-1,5-bisphosphate carboxylase/oxygenase
PGK	3-phopsphoglycerate kinase
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
FBA	Fructose-1,6-bisphosphate aldolase
TPI	Triosephosphate isomerase
FBP	Fructose-1,6-bisphosphatase
SBP	Sedoheptulose-1,7-bisphosphatase
TKL	Transketolase
RPI	Ribose-5-phosphate isomerase
RPE	Ribulose-5-phosphate-3-epimerase

In these two cases, a single isoenzyme is targeted to the plastids of both sequenced diatoms. However, *P. tricornutum* encodes another 5 isoforms of GAPDH, two of them are targeted to the mitochondrion and the three are probably cytosolic. Similarly, higher plant isoforms of GAPDH in the cytosol are involved in glycolysis. Recent analysis using plastid-targeted GAPDH genes from diatoms and other heterokont organisms indicate a common origin of all chromalveolates. While the members of Archaeplastida group has surely plastid targeted GAPH originated from cyanobacteria, all Chromalveolates form a robust monophyletic clade indicated a single GAPDH duplication in this supergroup (Harper and Keeling, 2003; Oborník et al., 2009). However, in more detailed phylogenic tree diatoms are sister clade to cryptophytes, whereas the other heterokonts such pelagophytes, phaeophytes and raphidiophytes group with dinoflagellates. In this phylogenetic tree Chromista seem to bee paraphyletic, moreover, haptophytes with chromerids and apicomplexans form distant group from rest of Chromalveolates (Takishita et al., 2009). However, since GAPDH seems to be frequently duplicated during evolution, this complex phylogenetic pattern can be result of such duplications and gradual losses of particlular duplicant (Oborník et al., 2009).

Enzymes for the regeneration of the ribulose-1,5-bisphosphate are nuclear encoded as multiple isoforms in both diatoms. In *P. tricornutum* there are three fructose-1,6-bisphosphate aldolase (FBA) isoforms, whereas only two have been found in *T. pseudonana*. Thereafter, four fructose-1,6-bisphosphatase FBPase and two triosephosphate isomerase (TPI) are encoded in *P. tricornutum*, but *T. pseudonana* possess only two isoforms of FBPase and one TPI. Transaldolase (TAL) is in three isoforms in *T. pseudonana* and two in *P. tricornutum* (Kroth et al., 2008). In both sequenced diatom genomes single plastid targeted ribose-5-phosphate isomerases (RPI) was identified together with one ribulose-phosphate epimerase (RPE) and transketolase (TKL).

Fructose bisphosphate aldolase (FBA) is divided into two classes of phylogeneticaly and structurally unrelated enzymes. Class I FBAs are heterotetramers found primarily in eukaryotes, where they are widespread (Gross et al., 1999). In contrast, class II FBAs are homodimers found mostly in eubacteria. FBAs of class II are subsequently distributed in group A and group B. Archaeplastida posses mostly class I FBAs with the evidence of ancestral duplication (Patron et al., 2004; Rogers and Keeling, 2004). However, *P. tricornutum* encodes two FBAs class I, where one isoform group together with single FBA I from *T. pseudonana* and comes probably from bacteria through horizontal gene transfer. Unlike these FBAs I are cytosolic, the second FBA I from pennate diatom, which is related to

FBAs from *Karenia brevis* and rhodophytes, is plastid targeted (Kroth et al., 2005). FBAs II span a greater diversity in eukaryotes and are found in cytosol of ascomycetes, oomycetes, *Euglena* and *Bigelowiella* (Patron et al., 2004; Rogers and Keeling 2004). The rest of plastid targeted FBAs are from diatoms and together with cryptophytes and dinoflagellates they belong to group A of class FBAs II. Additionally, cytosolic FBAs II of both sequenced diatoms also belong to group A. They seem to be more related to non photosynthetic oomycetes and ascomycetes (Kroth et al., 2005). However, the apicomplexans possess only FBAs of class I (Patron et al., 2004; Rogers and Keeling 2004). The presence of several plastid targeted FBA and TPI isoforms might be explained by the possible presence of glycolytic pathway in diatom plastids (Kroth et al., 2008). However, that would include FBA and TPI reactions in the reverse direction when compared to the Calvin cycle. Chlorophytes such as *Chlamydomonas renhardtii* keep the glycolytic pathway in the plastid, similarly to diatoms (Schnarrenberger et al. 1994).

Fructose bisphosphatase (FBPase) and sedoheptulose bisphosphatase (SBPase) are related enzymes that catalyze similar reactions in Calvin cycle: the dephosphorylation of fructose-1,6-bisphosphate to fructose-6-phosphate (Zimmermann et al. 1976) and dephosphorylation of sedoheptulose-1,7-bisphosphate to sedoheptulose-7-phosphate (Cadet and Meunier 1988). Plastid FBPases of Archaeplastida are not of cyanobacterial origin, but they are instead of it related to cytosolic forms, suggesting that they originated through nuclear gene duplication (Martin et al. 1996; Rogers and Keeling 2004). The phylogenetic tree enriched by sequences from heterokonts shows all eukaryotic plastid-targeted FBPases in a single clade (Teich et al., 2007). While SBP is universally found in photosynthetic lineages; its presence in apicomplexans, ciliates, trypanosomes (Rogers and Keeling, 2004), and ascomycetes is surprising given that no metabolic function beyond the one in the plastid Calvin cycle is described so far (Teich et al., 2007). However, the SBPase sequences from other Chromalveolates strongly group together in the phylogenetic tree, and the same assemblage is recovered for plastid-targeted FBP sequences, although this is less supported. Both SBP and plastid-targeted FBP are most likely of rhodophyte origin (Teich et al., 2007). Search within the diatom genomes has not led to finding of plastid targeted SBPase; just cytosolic form has been detected. It is possible that role of plastidic SBPase in this case is substituted by FBPase (Gruber et al. 2008).

6.2 Biosynthesis of photosynthetic pigments

Photosynthesis requires synthesis of essential pigments that are crucial for harvesting of light energy. While chlorophytes use the chlorophylls (Chls) *a* and *b* as the major light-harvesting pigments, and the carotenoids: lutein, neoxanthin, violaxanthin, and β-carotene contribute to light-harvesting only to a limited extent because their main function is photoprotection (Bassi and Caffarri 2000), diatoms posses *c*-type Chls which lack the aliphatic phytol side chain, and thus are much more polar. In general, diatoms contain two different Chl *c* species: Chl c_1 and Chl c_2, althouth a survey of many diatom strains showed that Chl c_1 is sometimes substituted by Chl c_3 (Jeffrey 1989). In addition, the carotenoid pigments are more involved in light-harvesting with no restriction like in green algae.

All photosynthetic organisms synthesize chlorophylls *de novo* by common pathway with heme being also precursor for cytochrome prosthetic groups, and vitamin B_{12}. Photosynthetic eukaryotes including diatoms begin heme synthesis by glutamate being ligated to tRNA-Glu by glutamyl-tRNA synthase (GluRS) (Beale 1999) (see Figure 8.). The two subsequent enzymatic steps take place in plastid and give rise to 5- aminolevulinate (ALA). Contrary to that non-photosynthetic eukaryotes use single step condensation of succinyl-CoA and glycine that is derived from and located in mitochondrion.

Subsequent six enzymatic steps are leading to formation of protoporphyrin IX, which is the branching point for the pathways synthetizing heme and chlorophyll respectively. The mechanism of this synthesis is similar in all organisms, but in case of photosynthetic eukaryotes the pathway is located in plastid (for details see Figure 8.). In plants and algae, subcellular localization of the reactions common to heme and chlorophyll take part in chloroplast with the exception of protoporphyrinogen oxidase (PPOX), which was suggested to be dually targeted to chloroplasts and mitochondria (Watanabe et al., 2001). Small amount of ferrochelatase activity was found in mitochondria as well (Cornah et al., 2003).

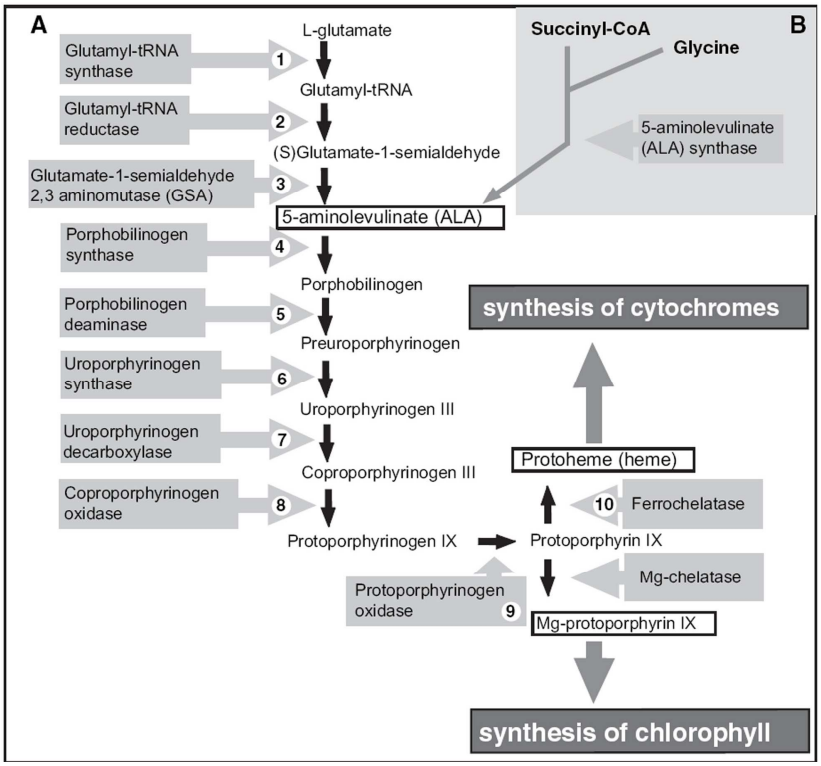


Figure 8a). General scheme of the heme biosynthesis pathway (Oborník and Green, 2005). The first part of the pathway, synthesis of ALA, differs in various organisms. In photosynthetic eukaryotes, ALA is synthesized in the plastid via the ''glutamate'' or C5 pathway (A), while in animals, fungi, and apicomplexans, ALA is produced in the mitochondrion using the Shemin or ''succinyl-CoA'' pathway (B). Thereafter, the condensation of 2 molecules ALA yields in porphobilinogen and is catalyzed by porphobilinogen synthase (PBGS). Then two-step condensation of 4 porphobilinogen molecules results in formation of uroporpyrinogen III, and is mediated by porphobilinogen deaminase (PBGD) and uroporpyrinogen III synthase (UPGS). Enzyme uroporphyrinogen decarboxylase (UPGD) gives rise to coproporphyrinogen III, which is transported back to mitochondrion in non photosynthetic eukaryotes. By oxidative decarboxylation of side chains, coproporphyrinogen III is converted to protoporphyrinogen IX. This enzymatic step is mediated by coproporphyrinogen oxidase (CPOX). Finally protoporphyrin IX is made by side chain oxidation of protoporohyrinogen oxidase (PPOX). Protoporphyrin IX is the branching point for the pathways leading to heme and to chlorophyll. Afterwards, heme is formed by insertion of iron ion (Fe^{2+}) done by ferrochelatase (Fch).

The C-5 pathway that starts from glutamate to synthesize ALA is strictly used by photosynthetic eukaryotes, where GluRSs form two isoforms, the organellar and cytosolic. In plants, rhodophytes and diatoms, GluRSs constitute two distinct clusters. It is obvious that first cluster contains one gene from diatom *T. pseudonana* and two *C. merolae* genes that originated in eubacteria; it is expected that it has been transfered to the diatom nucleus thanks to EGT from organellar genomes. Sequences from non-photosynthetic eukaryotes and plants constitue a sister group to Archaean GluRSs. This suggests that these genes originate from the secondary host nuclear genes coding for cytosolic enzymes (Oborník and Green, 2005). In three enzymatic steps following the ALA synthesis, the enzymes from photosynthetic eukaryotes show the expected plastid (cyanobacterial) origin (Oborník and Green 2005). The porphobilinogen deaminases (PBGD) from photosynthetic eukaryotes undoubtedly affiliate with α-proteobacteria, however the PBGDs of non-photosynthetic eukaryotes form distinct

group probably originating in eukaryotic nucleus. The mitochondrial origin of PBGD in photosynthetic eukaryotes can be explained by gene acquisition from mitochondria that happened at early time of Archaeplastida evolution, before radiation of primary plastid lineages. This enzyme can serve as a evidence for single origin of primary plastids (Obornik and Green 2005). Rhodophyte and diatom UPGD and PPOX are of probable cyanobacterial origin as the consequence of primary endosymbiotic event (Oborník and Green 2005). Two unrelated coproporphyrinogen oxidases (CPOX) exist in photosynthetic eukaryotes: an oxygen-dependent and an oxygen-independent (Dailey 1997). The oxygen–independent form is obviously derived from plastid, while oxygen-dependent CPOX comes from eukaryotic nucleus. The additional isoform of oxygen-dependent CPOX has been found in diatoms. Contrary to first CPOX, which was transferred from primary host's nucleus, the origin of second one remains unclear (Oborník and Green 2005).

The last enzyme of the heme pathway - ferrochelatase shows more complex evolutionary pattern. This enzyme catalyzes the first committed step to the synthesis of heme and bilin pigments. In this case, two rhodophyte sequences do not cluster either with other plastid sequences or with cyanobacterial sequences and appear to originate in proteobacteria. The same origin displays ferrochelatase from apicomplexan parasites (Sato and Wilson 2003). Although heterokonts have also acquired their plastid via secondary endosymbiosis from rhodophyte, the diatom has a typical plastid-cyanobacterial ferrochelatase (Obornik and Green 2005).

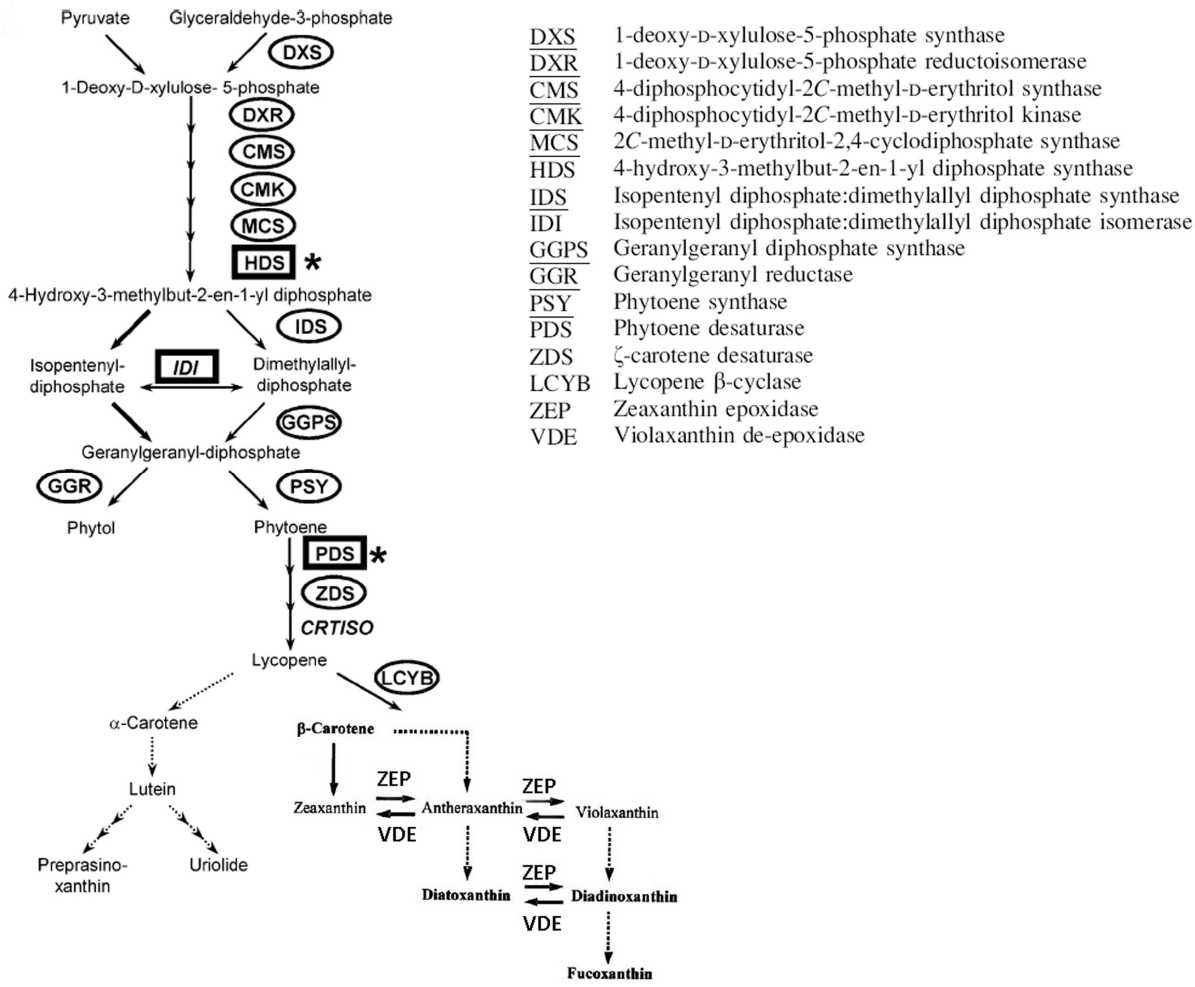
On the other hand, the Mg-chelatase, enzyme of the first step in chl synthesis, has subunit ChlI encoded on the plastid genome of the diatoms (Kowallik et al. 1995; Armbrust et al. 2004), similarly to chlorophytes (with the exception of *C. reinhardtii*), glaucophytes, red algae, and other heterokont algae examined so far. Interestingly, *T. pseudonana* posses two related copies of divinyl-protochlorophyllide reductase; these may be involved not only in reduction of divinyl-protochlorophyllide *a*, but also in formation of Chl c_1 from Chl c_2 (Wilhelm et al., 2006). In the case of protochlorophyllide oxidoreductase, which is light and NADPH dependent enzyme, diatoms probably do not share the plastid encoded forms from rhodophytes and chlorophytes. Instead of it, transformants of *P. tricornutum* were able to synthezise Chl *a* not only in light but also in the dark (Zaslavskaia et al. 2001). This implies that diatoms might contain another light-independent protochlorophyllide oxidoreductase which is unrelated to the light-dependent forms (Wilhelm et al., 2006).

Carotenoids are diterpenoids, essential components of the photosynthetic apparatus, serve as both light-harvesting and photoprotective proteins in plants and algae (Morosinotto et al. 2003; Grossman et al. 2004). Interstingly, diatoms do not use α-carotene pathway and contain xanthophylls, which are oxygen containing compounds derived from β-carotene (Pennington et al. 1988). Fucoxanthin is the dominant and major light-harvesting xanthophyll in diatoms. Other carotenoids present in significant amounts are β-carotene and the acetylenic xanthophylls diadinoxanthin and diatoxanthin (Jeffrey and Wright, 2005). They bring the advantage of absorbtion of light in the blue range of spectrum and are responsible for the characteristic brown color of chromist algae.

Generally, carotenoids are derived from the isoprene units, which are synthesized in all living organisms by condensations of the five-carbon compound isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (see Figure 9). Two distinct and independent biosynthetic routes to synthetize IPP exist: cytosolic mevalonate (MVA) and plastid located 2-C-methyl-d-erythritol 4-phosphate (MEP) pathway. Unlike non photosynthetic eukaryotes use only the MVA as single pathway, land plants harbor both essential and completely different metabolic routes for isoprenoid synthesis. Although it was shown that common ancestor of Archaeplastida bears both pathways, several green algal lineages such as Prasinophyceae, Ulvophyceae or Chlorophyceae lost cytosolic MVA pathway (Grauvogel and Petersen, 2007). In eukaryotes, MEP pathway is restricted to plastid-bearing organisms and the carotenoid biosynthesis is done exclusively through the plastid located MEP pathway. Additionally, the MEP pathway is essential for apicomplexan parasites *Plasmodium falciparum*, *Theileria parva* and *T. annulata*, and considered as the target for antimalarial drugs (Jomaa et al., 1999). The redundancy of IPP biosynthesis in land plants could be explained by increased requirement of isoprenoid compounds for growth and regulation in highly organized multi-cellular organisms.

The majority of enzymes involved in the carotenoid pathway have been found in chromists, including diatoms. According to chromalveolate hypothesis (Cavalier-Smith, 1999), plastid-targeted nucleus-encoded proteins from diatoms and other related chromists should be closely related to their red algal homologs. This hypothesis was correct for nine enzymes of carotenoid biosynthesis (see Figure 9). The tree based on concatened alignment of the enzymes of interest shows close relationship among diatoms, haptophytes and cryptophytes forming a sister group to rhodophytes (Frommolt et al., 2008). Enzymes 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthese (HDS), isopentenyl diphosphate: dimethylallyl diphosphate isomerase (IDI) and phytoene desaturase (PDS) make an exception.

Figure 9. Generalized scheme of carotenoid biosynthesis in plants and algae summarizing major results of publication of Frommolt et al. (2008; modified). Encircled enzyme names represent chromist proteins most closely related to their red algal counterparts, whereas framed names are chromist proteins most closely related to their green algal counterparts; asterisks designate the chromist proteins that share diagnostic indels specifically with homologs from basal green algae; filled triangles label the pair of genes that are syntenic in the genomes of a diatom, a pelagophyte, and two prasinophyte algae.



For HDS involved in MEP pathway, a significant evolutionary separation is found between the cyanobacterial and red algal cluster on one side and the proteins from plants, green algae, chromists, and also from alveolates on the other. This topology was previously interpreted as result of a replacement of cyanobacterial gene by HGT after the primary endosymbiosis (Lange et al., 2000). More recent analyses, however, show the relationships of plant - Chromalveolate clade with bacterial group of Chlamydia. This suggests that Chlamydia-like bacterium may have donated HDS to eukaryotic phototrophs at the level of the primary endosymbiosis (Huang and Gogarten 2007; Moustafa et al. 2008). On the other hand, phylogenic tree inferred from phytoene desaturase PDS places diatoms, pelagophytes

and haptophytes together with basal Chloroplastida of the class Prasinophyceae, whereas rhodophytes form distinct clade. However, the distribution of other chromists is more disturbed and the explanation could be due to two different horizontal gene transfers from basal Chloroplastida (Frommolt et al., 2008). In the case of isopentenyl diphosphate: dimethylallyl diphosphate isomerase (IDI), two phylogeneticaly unrelated classes of enzymes have been described: Type 1 IDI is present in rhodophytes, chromalveolates, and the Chloroplastida, whereas cyanobacteria contain a type 2 IDI (Bouvier et al. 2005). The diatoms, pelagophytes and haptophytes have two or three isoforms of IDI, which seems to originate in gene duplication. Similarly to PDS, genes from chromist algae are related to prasinophytes (Frommolt et al., 2008).

Violaxanthin de-epoxidase (VDE) and zeaxanthin epoxidase (ZEP) represent the enzymes involved in xanthophyll cycle (Figure 9). In diatoms, more copies of the genes encoding these enzymes were found when compared with other photosynthetic eukaryotes. It seems that both enzymes originate as the eukaryotic innovation of ancient Archaeplastida and very probably diversified before the secondary endosymbiosis had occurred (Coesel et al., 2008; Frommolt et al., 2008). Rhodophyte different capture mechanism and phycobilisomes would have led to the loss or replacement of most “green” genes, whereas prasinophytes retain those enzymes (Frommolt et al., 2008). Consequently, the diatom chromist lineage likely obtained all paralogues of ZEP and VDE genes during the process of secondary endosymbiosis by gene transfer from the nucleus of the algal endosymbiont to the host nucleus (Coesel et al., 2008). VDE and ZEP in diatoms form sister clade to prasinophytes, which may indicate horizontal gene transfer as the alternative explanation. However, the higher numbers of such transfer are proposed very improbable (Frommolt et al., 2008). Furthermore, the presence of ZEP gene in *Tetrahymena thermophila* provides the first evidence for a plastid gene encoded in a heterotrophic ciliate, providing support for the chromalveolate hypothesis (Coesel et al., 2008).

6.3. Aromatic Amino acid biosynthesis

Aromatic amino acids are abundant components of all living organisms. In land plants they play a direct role in development regulation, pathogen defense response, and environmental interactions. In addition, a number of indole alkaloids have important pharmacological uses (Radwanski and Last 1995). The aromatic amino acid biosynthesis is divided to four pathways. First part, the shikimate pathway is common for biosynthesis of all three aromatic amino acids and is connected to carbohydrate metabolism. Products derived from last intermediate - chorismate (chorismic acid) are distributed toward particular pathways of tryptophan, phenylalanine, and tyrosine synthesis (Figure 10.). Besides this, chorismate enters to synthesis of ubiquinones, *p*-aminobenzoate and folates (Mustafa and Verpoorte 2005; Sahr et al., 2006).

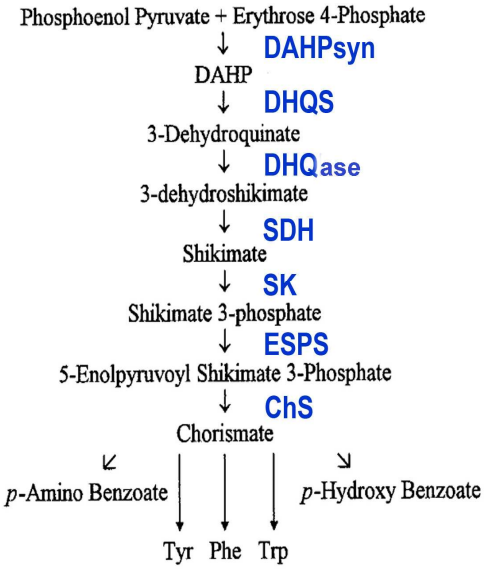


Figure 10. Shikimate pathway consists of seven enzymatic steps (Hermann, 1999; modified). It starts by condensation of phosphoenol pyruate and erythrose-4-phosthate, mediated by 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (DAHPsyn). Second step is a formation of quinate circle catalyzed by 3-dehydroquinate synthase (DHQS) and the third step is catalyzed by 3-dehydroquinate dehydratase (DHQase). Following steps 4 to 7 are catalyzed by shikimate 5-dehydrogenase (SDH), shikimate kinase (SK), 5-enolpyruvylshikimate-3-phosphate synthase (ESPS), and chorismate synthase (ChS), respectively.

The aromatic amino acid pathways are present in most of prokaryotic lineages, in fungi and in photosynthetic eukaryotes. More recently, the shikimate pathway enzymes has been identified in alveolates; all enzymes of shikimate pathway were found in ciliate *Tetrahymena thermophila* (Richards et al., 2006) and parasite *Toxoplasma gondii* (Cambell et al., 2004). Several *Plasmodium* species possess incomplete enzyme set in their genomes (McConkey et al., 2004). Interestingly, the inhibition of shikimate pathway (particularly 5-enolpyruvylshikimate-3-phospate synthase) is a potential therapeutic target and has been demonstrated to curtail the growth of *Plasmodium falciparum in vitro* (McConkey 1999). It is generally known that animals are not able to synthesize aromatic AAs *de novo* (only synthesis

of tyrosine by hydroxylation F) and require them in their diet. Moreover, they preferably require aromatic AAs as precursors for biogenic amines, and hormones. Shikimate pathway is conserved and its biochemical reactions do not differ among eukaryotes and prokaryotes. The pathway is found in archaea, bacteria, fungi, and plants. Most prokaryotes examined to date encode shikimate pathway enzymes as separate polypeptides; although in some bacterial lineages fusions have occurred, usually involving two genes encoding particular enzymes (Herrmann and Weaver 1999).

Based on low sequence similarity and molecular weight, 3-deoxy-D-arabino-heptulosonate 7-phosphate synthases (DAHPsyn) can be classified into two non-identical families (Walker et al., 1996). However, identical function was confirmed by analysis of X-ray crystal structures (Webby et al., 2005). In addition, the DAHPsyn I family is divided into subfamilies DAHPsyn Iα and DAHPsyn Iβ. While the α subfamily is widely represented in gram-negative bacteria, the members of the DAHPsyn Iβ subfamily are more closely related to 3-deoxy-D-*manno*-octulosonate 8-phosphate synthases (Subramaniam et al., 1998), which are probably not found in eukaryotes. Higher plants encode two isoforms belonging to DAHPsyn II family. One isoform is located in the cytoplasm and other in plastid. They are differently regulated (Rubin and Jensen 1985). Similarly, DAHPsyn II has been found in *T. gondii* and diatoms, in contrast to cyanobacteria, which entirely posses DAHP family I. Interestingly, non-photosynthetic oomycetes and fungi encompass both families of DAHPsyn. Richards et al. (2006) propose that the common eukaryotic ancestor retained both variants of DAHP synthase. Confirming this theory, the sequences encoding DAHP I type have been found in photosynthetic eukaryotes (haptophytes and cryptophytes). However, phylogeny of DAHPsyn families appears to be more complex. While eukaryotes encoding DAHPsyn I are monophyletic, organisms bearing DAHPsyn II are scattered through the phylogenetic tree together with several procaryotic groups. This is explained by a high number of horizontal gene transfers that is proposed to happen there (Richards et al. 2006).

Non-photosynthetic fungi, oomycetes and alveolates *T. thermophila* and *T. gondii* possess pentafunctional gene the *arom*, which includes conserved domains homologous to all genes involved in the fusion. The *arom* contains enzymes codin for steps from 2 to 6. Eventually, the reduced forms of this multiprotein have been found in diatom *T. pseudonana* 3-dehydroquinate dehydratase: shikimate 5-dehydrogenase (DHQase:SDH) and zygomycete *R.oryzeae* 3-dehydroquinate synthase: 3-dehydroquinate dehydratase (DHQS:DHQase). They form monophyletic group together with sequences from fungi, oomycetes and alveolates. The origin of the *arom* polypeptide is explained as ancient in eukaryotes (Richards et al., 2006).

Contrary to the *arom*, plants' shikimate genes display different evolutionary origins. The plant shikimate enzymes are encoded as separate proteins, with the exception DHQase and SDH enzymes, which are fused to form a bifunctional enzyme (Mousdale et al., 1987). However, the phylogeny shows the plant DHQase:SDH fusion being affiliated to bacterial species not to *arom* polypeptide, as in the case of diatoms (Richards et al., 2006). The other genes in plants and diatoms are bacterial in origin too, and result in mosaic character of the shikimate pathway.

While shikimate kinase (SK) and chorismate synthase (ChS) appear to be of cyanobacterial origin, DHQS and 5-enolpyruvylshikimate-3-phosphate synthase (ESPS) originate in the proteobacterial lineages probably by horizontal gene transfer (Richards et al., 2006). The Shikimate pathway enzymes from photosynthetic eukaryotes are nuclear encoded, but they are assumed to function within the plastid. However, only DAHPsyn, ESPS and SK have been experimentally demonstrated to be imported to plastid (Zhao et al., 2002; della-Cioppa et al. 1986; Schmid et al., 1992). Searching for Signal peptide encoded at N-termini of enzymes concludes that shikimate pathway is putatively located in plastid of diatoms (Richards et al., 2006).

Biosynthesis of tryptophan (Trp) involves also seven enzymatic steps, but is done by five enzymes (Figure 11.). Tryptophan synthezing enzymes are also nuclear-encoded, with the exception of plastid encoded α-subunit tryptophan synthase (TS) in rhodophytes (Ohta et al. 1993). The localization of this plastid pathway has been experimentally confirmed by activity of enzymes in plastid extracts (Radwanski and Last 1995). Fungi, oomycetes and diatoms contain gene fusions encoding tryptophan pathway enzymes. The complex called *trp1* (tryptophan biosynthesis protein) involves component II (β-subunit) of anthranilate synthase (AS), fused with indole glycerol phosphate synthase (InGPS) and phosphoribosyl anthranilate isomerase (PRAI) and is found in most fungal species. However, in yeasts (Endomycetales), component II (β-subunit) of AS is fused with InGPS only (DeMoss and Wegman 1965; Matchett and DeMoss 1975). The fusion of PRAI:InGPS is present in genomes of diatoms and oomycetes, where it indicates the origin from heterokont nucleus (Jiroutová et al., 2007). The other eukaryote-specific fusion is that of α + β subunits of tryptophan synthase, which has been found in fungi, including yeasts, and in diatoms (Burns et al., 1990; Jiroutová et al., 2007). TS subunits of oomycetes are encoded separately (Jiroutová et al., 2007). Diatoms are the only eukaryotes that possess fused subunits of anthranilate synthase (AS). This fusion is almost certainly of bacterial origin and form the distinct group from the other ASs, which are

not fused (Bae and Crawford 1990). Interestingly, the anthranilate phosphoribosyl transferase (APRT) gene from oomycetes display origin in the rhodophyte nucleus and has likely been transferred to heterokont's nucleus during secondary endosymbiosis. This suggests possible retention of the complex plastid in the ancestor of stramenopiles and later loss of this organelle in oomycetes (Jiroutová et al., 2007).

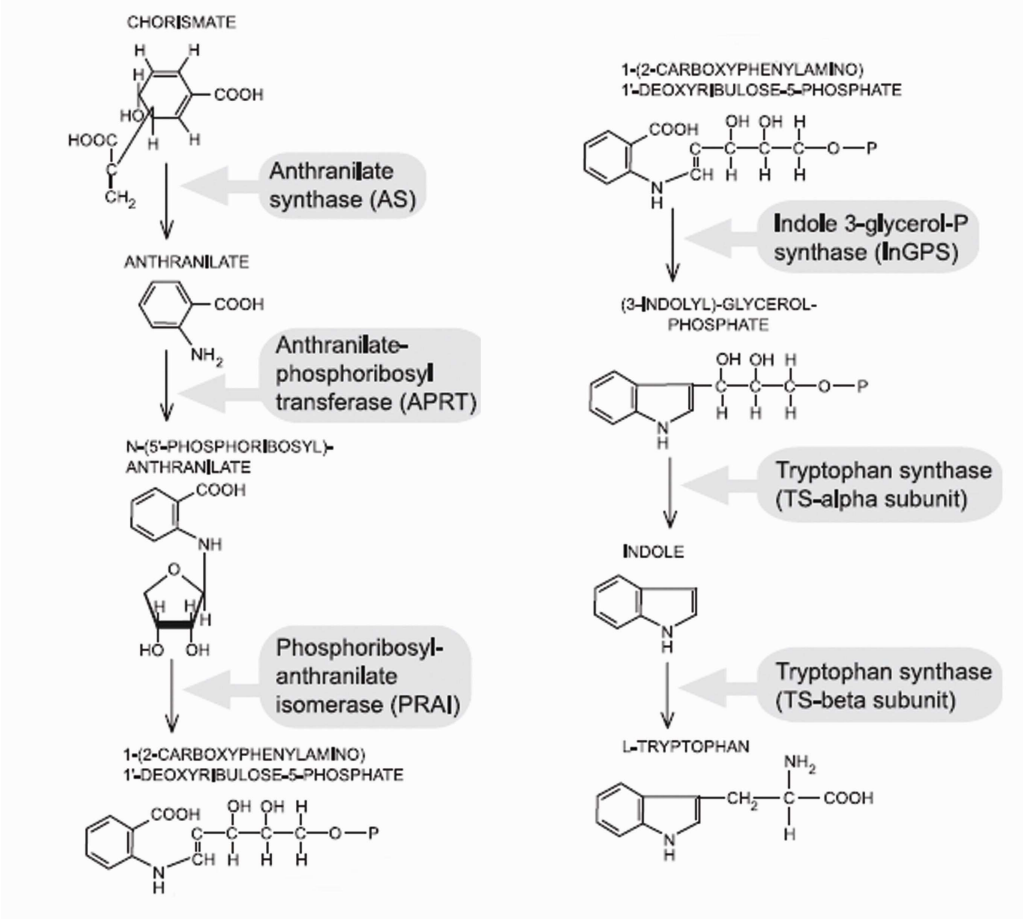


Figure 7. Tryptophan pathway (Crawford 1989; Jiroutová et al., 2007).The first reaction is catalyzed by the bifunctional enzyme anthranilate synthase (AS). AS β-subunit is an aminotransferase and cleaves glutamine, which is utilized by α-subunit AS to convert chorismate to anthranilate. The second step in Trp synthesis is the addition of phosphoribosyl pyrophosphate to the anthranilate, catalyzed by anthranilate phosphoribosyl transferase (APRT). The third step is arrangement of phosphoribosyl done by phosphoribosyl anthranilate isomerase (PRAI) and is followed by the fourth step, a decarboxylation and ring closure catalyzed by indole glycerol phosphate synthase (InGPS). The final step, the removing of glycerol phosphate side chain and its replacement by the alanyl part of L-serine, is catalyzed by the complex enzyme tryptophan synthase (TS).

The trend to replace plastid enzymes by eukaryotic homologues is most evident particulalrly in the tryptophan pathway in diatoms. Although the complete pathway is putatively located in the complex plastid, only one of involved enzymes, indole-3- glycerol phosphate synthase (InGPS), displays a possible cyanobacterial origin. Additionaly, *P. tricornutum* possess the unique fusion of InGPS with the cyanobacteria-derived hypothetical

protein COG4398. The other two diatom InGPSs form the clade with the *C. merolae*, but also with the proteins from heterotrophic and parasitic apicomplexans (Jiroutová et al., 2007).

Since an extremely large amount of ATP (78 mol) is required to synthesize 1 mol of tryptophan (Atkinson 1977), such location for tryptophan synthesis is advantageous for diatoms. Tryptophan synthesis represents the pathway with the highest known level of replacement of originaly plastid, now nuclear encoded genes, by eukaryotic cytosolic homologues that obtained plastid targeted presequences.

7. List of Abbrevations

A alanine

AA amino acid

ALA 5- aminolevulinate

APRT anthranilate phosphoribosyl transferase

AS anthranilate synthase

CA carbonic anhydrase

CPOX coproporphyrinogen oxidase

DAHPsyn 3-deoxy-D-arabino-heptulo-sonate 7-phosphate synthase

DHQase 3-dehydroquinate dehydratase

DHQS 3-dehydroquinate synthase

EGT endosymbiotic gene transfer

ER endoplasmic reticulum

ERAD endoplasmic reticulum-associated degradation

ESPS 5-enolpyruvylshikimate-3-phosphate synthase

EST expressed sequence tag

F phenylalanine

GAPDH glyceraldehyde- 3-phosphate dehydrogenase

GFP green fluorescent protein

GluRS glutamyl-tRNA synthase

HDS 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase

HGT horizontal gene transfer

Chl chlorophyll

ChS chorismate synthase

IDI isopentenyl diphosphate: dimethylallyl diphosphate isomerase

InGPS indole glycerol phosphate synthase

OAA oxalacetate

P proline

PBGD porphobilinogen deaminase

PDS phytoene desaturase

PEP phosphoenolpyruvate

PEPC phosphoenolpyruvate carboxylase

PEPCK phosphoenolpyruvate carboxykinase

PGK 3-phosphoglycerate kinase

PPDK pyruvate phosphate dikinase

PPOX protoporphyrinogen oxidase

PRAI phosphoribosyl anthranilate isomerase

S serine

SDH shikimate 5-dehydrogenase

SDV silica depositing vesicle

SK shikimate kinase

SP signal peptide (ER)

Tic translocon of the inner chloroplast membrane

Toc translocon of the outer chloroplast membrane

TP transit peptide

TS tryptophan synthase

UPGD uroporphyrinogen decarboxylase

VDE violaxanthin de-epoxidase

ZEP zeaxanthin epoxidase

8. References

Adams K.L., Daley D.O., Whelan J., Palmer J.D. (2002): Genes for two mitochondrial ribosomal proteins in flowering plants are derived from their chloroplast or cytosolic counterparts. *Plant Cell* 14(4):931-43.

Adl S.M., Simpson A.G., Farmer M.A., Andersen R.A., Anderson O.R., Barta J.R., Bowser S.S., Brugerolle G., Fensome R.A., Fredericq S., James T.Y., Karpov S., Kugrens P., Krug J., Lane C.E., Lewis L.A., Lodge J., Lynn D.H., Mann D.G., McCourt R.M., Mendoza L., Moestrup O., Mozley-Standridge S.E., Nerad T.A., Shearer C.A., Smirnov A.V., Spiegel F.W., Taylor M.F. (2005): The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J. Eukaryot. Microbiol*. 52, 399–451.

Allen A.E., Vardi A., Bowler C. (2006): An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms. *Curr. Opin. Plant Biol.* 9: 264–273.

Apt K.E., Zaslavkaia L., Lippmeier J.C., Lang M., Kilian O., Wetherbee R., Grossman A.R., Kroth P.G. (2002): In vivo characterization of diatom multipartite plastid targeting signals. *J. Cell Sci.* 115:4061-4069.

Apt K., P.G. Kroth-Pancic, A.R. Grossman (1996): Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. *Mol. Gen. Genet*. 252:572–9

Archibald J.M., Keeling P.J. (2002): Recycled plastids: a 'green movement' in eukaryotic evolution. *Trends Genet.* 18(11):577-84.

Archibald J.M., Rogers M.B., Toop M., Ishida K-I., Keeling P.J. (2003): Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigelowiella natans*. *P.N.A.S.* 100:7678–7683

Armbrust E.V., Berges J.A., Bowler C., Green B.R., Martinez D., Putnam N.H., Zhou S., Allen A.E., Apt K.E., Bechner M., Brzezinski M.A., Chaal B.K., Chiovitti A., Davis A.K., Demarest M.S., Detter J.C., Glavina T., Goodstein D., Hadi M.Z., Hellsten U., Hildebrand M., Jenkins B.D., Jurka J., Kapitonov V.V., Kröger N., Lau W.W.Y., Lane T.W., Larimer F.W., Lippmeier J.C., Lucas S., Medina M., Montsant A., Obornik M., Parker M.S., Palenik B., Pazour G.J., Richardson P.M., Rynearson T.A., Saito M.A., Schwartz D.C., Thamatrakoln K., Valentin K., Vardi A., Wilkerson F.P., Rokhsar D.S. (2004) The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism. *Science* 306:79–86

Armbrust V.E. (2009): The life of diatoms in the worlds` oceans. *Nature* 459:185-193

Atkinson D.E. (1977): Cellular energy metabolism and its regulation. Academic Press, New York

Badger M.R., Andrews TJ., Whitney S.M., Ludwig M., Yellowlees D.C., Leggat W., Price G.D. (1998): The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO_2-concentrating mechanisms in algae. *Can. J. Bot.* 76: 1052–1071.

Bae Y.M., Crawford, I.P. (1990): The *Rhizobium meliloti* TrpE(G) gene is regulated by attenuation, and its product Anthranilate synthase is regulated by feedback inhibition. *J. Bacteriol.* 172(6):3318–3327

Bachvaroff, T.R., Concepcion, G.T., Rogers, C.R., Herman, E.M., Delwiche, C.F. (2004): Dinoflagellate expressed sequence tag data indicate massive transfer of chloroplast genes to the nuclear genome. *Protist* 155, 65–78.

Bassi R., Caffarri S. (2000): Lhc proteins and the regulation of photosynthetic light-harvesting function by xanthophylls. *Photosynth. Res.* 64: 243—256

Bates S.S., Trainer V.L. (2006): in Ecology of Harmful Algae (eds. Granuli E., Turner J.T.) Ch. 7, Springer.

Beale S. I. (1999): Enzymes of chlorophyll biosynthesis. *Photosynth.Res*. 60:43–73.

Bhattacharya D., Yoon H.S., Hackett J.D. (2003): Photosynthetic eukaryotes unite: endosymbiosis connects the dots. *BioEssays* 26:50–60, 2004.

Bouvier F, Rahier A, Camara B. (2005): Biogenesis, molecular regulation and function of plant isoprenoids. *Prog Lipid Res*. 44:357–429.

Bowler C., Allen A.E., Badger J.H., Grimwood J., Jabbari K., Kuo A., Maheswari U., Martens C., Maumus F., Otillar R.P., Rayko E., Salamov A., Vandepoele K., Beszteri B., Gruber A., Heijde M., Katinka M., Mock T., Valentin K., Verret F., Berges J.A., Brownlee C., Cadoret J.P., Chiovitti A., Choi C.J., Coesel S., De Martino A., Detter J.C., Durkin C., Falciatore A., Fournet F., Haruta M., Huysman M.J.J., Jenkins B.D., Jiroutova K., Jorgensen R.E., Joubert Y., Kaplan A., Kröger N., Kroth P.G., La Roche J., Lindquist E., Lommer M., Martin–Jézéquel V., Lopez P.J., Lucas S., Mangogna M., McGinnis K., Medlin L.K., Montsant A., Oudot–Le Secq M.P., Napoli C., Obornik M., Schnitzler-Parker M., Petit J.L., Porcel B.M., Poulsen N., Robison M., Rychlewski L., Rynearson T.A., Schmutz J., Shapiro H., Siaut M., Stanley M., Sussman M.R., Taylor A.R., Vardi A., von Dassow P., Vyverman W., Willis A., Wyrwicz L.S., Rokhsar D.S., Weissenbach J., Armbrust V.E., Green B.R., Van de Peer Y., Grigoriev I.V. (2008): The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature 456*: 239-244.

Bowler C., Karl D.M., Colwell R.R. (2009): Microbial oceanography in a sea of opportunity. *Nature* 459(7244):180-4.

Burki F, Shalchian-Tabrizi K, Minge M, Skjaeveland A, Nikolaev SI, Jakobsen KS, Pawlowski J. (2007): Phylogenomics reshuffles the eukaryotic supergroups. PLoS One. 2(8):e790

Burns D.M., Horn V., Paluh J., Yanofsky C. (1990): Evolution of the tryptophan synthetase of fungi. Analysis of experimentally fused *Escherichia coli* tryptophan synthetase alpha and beta chains. *J Biol. Chem. 265*(4):2060-9.

Cadet, F., Meunier J.C. (1988): pH and kinetic studies of chloroplast sedoheptulose-1,7-bisphosphatase from spinach (Spinacia oleracea). *Biochem J 253*:249–254

Campbell, S.A., Richards T.A., Mui E.J., Samuel B.U., Coggins J.R., McLeod R., Roberts C.W. (2004): A complete shikimate pathway in *Toxoplasma gondii*: an ancient eukaryotic innovation. *Int. J. Parasitol.* **34**:5–13.

Carlton J.M., Escalante A.A., Neafsey D., Volkman S.K. (2008): Comparative evolutionary genomics of human malaria parasites. *Trends Parasitol.* 24(12):545-50.

Cavalier-Smith T. (1999): Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J. Eukaryot. Microbiol*. 46(4):347-66

Cavalier-Smith T. (2000): Membrane heredity and early chloroplast evolution. *Trends Plant Sci*. 5(4):174-82.

Cavalier-Smith T. (2002): Chloroplast evolution: secondary symbiogenesis and multiple losses. *Curr. Biol.* 12(2):R62-4.

Cavalier-Smith T. (2003): Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae). Philos Trans R Soc Lond B Biol Sci. 358(1429):109-33

Coesel S., Oborník M., Varela J., Falciatore A., Bowler C. (2008): Evolutionary Origins and Functions of the Carotenoid Biosynthetic Pathway in Marine Diatoms. *PLoS ONE* 3(8): e2896.

Cornah J.E., Terry M.J., Smith A.G. (2003). Green or red: what stops the traffic in the tetrapyrrole pathway? *Trends Plant Sci. 8:*224–230.

Crawford I.P. (1989): Evolution of biosynthetic pathway: tryptophan paradigm. *Annu Rev Microbiol* 43:567–600

Dailey H.A. (1997): Enzymes of heme biosynthesis. *J. Biol. Inorg. Chem*. 2:411–417.

De Riso V., Raniello R., Maumus F., Rogato A., Bowler C., Falciatore A. (2009): Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucleic Acids Res.* 37(14):e96

della-Cioppa G., Bauer S.C., Klein B.K., Shah D.M., Fraley R.T., Kishore G.M. (1986): Translocation of the precursor of 5-enolpyruvylshikimate-3-phosphate synthase into chloroplasts of higher plants in vitro. *PNAS* 83:6873–6877

Delwiche C.F. (1999): Tracing the thread of plastid diversity through the tapestry of life. *Am. Nat.* 154:S164–S177

DeMoss J.A., Wegman J. (1965): An eznyme aggregate in tryptophan pathway of *Neurospora crassa*. *PNAS* 54:241–247

Doney, S.C. (2006): The dangers of ocean acidification. *Sci. Am.* 294: 58-65

Douglas S.E., Murphy C.A., Spencer D.F., Gray M.W. (1991): Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. *Nature* 350:148–151

Douglas S.E., Penny S.L. (1999): The plastid genome of the cryptophyte alga *Guillardia theta*: complete sequence and conserved synteny groups confirm its common ancestry with red algae. *J Mol Evol* 48, 236– 244.

Dunahay T., Jarvis E., Roessler P. (1995): Genetic transformation of the diatoms *Cyclotella cryptica* and *Navicula saprophita*. *J. Phycol.* 31, 1004–1012.

Falciatore A., Casotti R., Leblanc C., Abrescia C., Bowler C. (1999): Transformation of Nonselectable Reporter Genes in Marine Diatoms. *Mar Biotechnol* 1, 239-251.

Falkowski P.G., Katz M.E., Knoll A.H., Quigg A., Raven J.A., Schofield O., Taylor F.J. (2004): The evolution of modern eukaryotic phytoplankton. *Science*. Jul 16;305(5682):354-60

Falkowski, P.G., J.A., Raven (2007): Aquatic photosynthesis. Princeton, NJ, Princenton University Press.

Fast N.M., Kissinger J.C., Roos D.S., Keeling P.J. (2001): Nuclear-encoded, plastid targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol. Biol. Evol.*;18:418–426.

Fischer H., Robl I., Sumper M., Kröger N. (1999): Targeting and covalent modification of cell wall and membrane proteins heterologously expressed in the diatom Cylindrotheca fusiformis (Bacillariophyceae). *J. Phycol.* 35, 113–120.

Frommolt R., Werner S., Paulsen H., Goss R., Wilhelm C., Zauner S., Maier U.G., Grossman A.R., Bhattacharya D., Lohr M. (2008): Ancient recruitment by

chromists of green algal genes encoding enzymes for carotenoid biosynthesis. *Mol Biol Evol.* 25(12):2653-67.

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

Gibbs S.P. (1981): The chloroplasts of some algal groups may have evolved from endosymbiotic eukaryotic algae. Ann N Y Acad Sci.;361:193-208.

Gould S.B., Sommer M.S., Hadfi K., Zauner S., Kroth P.G., Maier U.G.. (2006a): Protein targeting into the complex plastid of cryptophytes. *J. Mol. Evol.* 62(6):674-81.

Gould S.B., Sommer M.S., Kroth P.G., Gile G.H., Keeling P.J., Maier U.G. (2006b): Nucleus-to-nucleus gene transfer and protein retargeting into a remnant cytoplasm of cryptophytes and diatoms. *Mol. Biol. Evol.* 23(12):2413-22

Granum E., Raven J.A., Leegood R.C. (2005): How do marine diatoms fix 10 billion tonnes of anorganic carbon per year. *Can. J. Bot.* 83: 898–908

Grauvogel C., Petersen J. (2007): Isoprenoid biosynthesis authenticates the classification of the green alga Mesostigma viride as an ancient streptophyte. *Gene* 396(1):125-33.

Green B.R. (2004): The chloroplast genome of dinoflagellates — a reduced instruction set? *Protist* 155, 23–31.

Greenwood A.D., Griffiths H.B., Santore U.J. (1977): Chloroplasts and cell compartments in Cryptophyceae. Br Phycol J 12:119

Gross W., Lenze D., Nowitzki U., Weiske J., Schnarrenberger C. (1999): Characterization, cloning, and evolutionary history of the chloroplast and cytosolic class I aldolases of the red alga *Galdieria sulphuraria*. *Gene* 230:7–14

Grossman A.R., Lohr M., Im C.S. (2004): *Chlamydomonas reinhardtii* in the landscape of pigments. *Annu Rev. Genet.* 38:119–173.

Gruber A., Vugrinec S., Hempel F., Gould S.B., Maier U.G., Kroth P.G. (2007): Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. *Plant Mol. Biol.*64(5):519-30.

Gruber A., Weber T., Bártulos C.R., Vugrinec S., Kroth P.G. (2009): Intracellular distribution of the reductive and oxidative pentose phosphate pathways in two diatoms. *J. Basic Microbiol.* 49(1):58-72

Gubbels M.J., White M., Szatanek T. (2008): The cell cycle and Toxoplasma gondii cell division: tightly knit or loosely stitched? *Int J Parasitol.*38(12):1343-58.

Hackett J.D., Maranda L., Yoon H.S., Bhattacharya D. (2003): Phylogenetic evidence for the cryptophyte origin of the plastid of *Dinophysis* (Dinophysiales, Dinophyceae). *J. Phycol.* 39:440–448.

Hackett J.D., Yoon H.S., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Nosenko T., Bhattacharya D. (2004): Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. *Curr. Biol.* 14, 213–218.

Hampl V., Hug L., Leigh J.W., Dacks J.B., Lang B.F., Simpson A.G., Roger A.J. (2009): Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". *PNAS* 106(10):3859-64.

Hansmann P. (1988). Untrastructural localization of RNA in cryptomonads. *Protoplasma* 146:81–88.

Harper J.T., Keeling P.J. (2003): Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. *Mol Biol. Evol*. 20(10):1730-5.

Harper J.T., Waanders E., Keeling P.J. (2005): On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. *Int J Syst Evol Microbiol.* 55: 487-96.

Hempel F., Bozarth A., Sommer M.S., Gauner S., Przyborski J.M., Maier U.G (2007): Transport of nuclear-encoded proteins into secondarily evolved plastids. *Biol. Chem.* 388: 899–906,

Herrmann K.M., (1995): The Shikimate Pathway as an Entry to Aromatic Secondary Metabolism. *Plant Physiol.* 107: 7-12

Herrmann K.M., Weaver L.M., (1999): The Shikimate pathway. *Annu. Rev. Plant Physiol. Plant Mol. Biol*. 50:473-503.

Hoffmeister M., Martin W. (2003): Interspecific evolution: microbial symbiosis, endosymbiosis and gene transfer. *Environ Microbiol.* 2003 Aug;5(8):641-9.

Huang CY, MA Ayliffe, JN Timmis (2004). Simple and complex nuclear loci created by newly transferred chloroplast DNA in tobacco. *PNAS* 29;101(26):9710-5

Huang, J., Gogarten, P. (2007): Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? *Genome Biology* 8: R99

Hunfeld K.P., Hildebrandt A., Gray J.S. (2008): Babesiosis: recent insights into an ancient disease. *Int. J. Parasitol.* 38(11):1219-37

Chaal B.K., Green B.R. (2005): Protein import pathways in “complex” chloroplasts derived from secondary endosymbiosis involving a red algal ancestor. *Plant Mol. Biol.* 57:333–342

Chaal B.K., Ishida K., Green B.R. (2003): A thylakoidal processing peptidase from the heterokont alga *Heterosigma akashiwo*. *Plant Mol. Biol.* 52(2):463-72.

Chepurnov V.A., Mann D.G., Sabbe K., Vyverman W. (2004): Experimental studies on sexual reproduction in diatoms. *Int. Rev. Cytol.* 237:91-154

Chesnick J.M., Kooistra W.H., Wellbrock U., Medlin L.K. (1997). Ribosomal RNA analysis indicates a benthic pennate diatom ancestry for the endosymbionts of the dinoflagellates *Peridinium foliaceum* and *Peridinium balticum* (Pyrrhophyta). *J. Eukaryot. Microbiol.* 44(4):314-20

Inagaki Y., Dacks J.B., Doolittle W.F., Watanabe K.I., Ohama T. (2000): Evolutionary relationship between dinoflagellates bearing obligate diatom endosymbionts: insight into tertiary endosymbiosis. *Int. J. Syst. Evol. Microbiol.* 6:2075-81.

Ishida K., Cavalier-Smith T., Green B.R. (2000): Endomembrane structure and the chloroplast protein targeting pathway in *Heterosigma akashiwo* (Raphidophyceae, Chromista). *J. Phycol.* 36:1135–1144

Ishida K., Green B.R. (2002): Second-hand and third-hand chloroplasts in dinoflagellates: phylogeny of oxygen-evolving enhancer 1 (PsbO) protein reveals replacement of a nuclear-encoded plastid gene by that of a haptophyte tertiary endosymbiont. *PNAS* 99(14):9294–9299

Ishida K. (2005): Protein targeting into plastids: a key to understanding the symbiogenetic acquisitions of plastids. *J. Plant Res.* 118(4):237-45.

Jarvis P., Robinson C. (2004): Mechanisms of protein import and routing in chloroplasts. *Curr. Biol.* 14:R1064–R1077

Jeffrey S.W. (1989): Chlorophyll *c* Pigments and their distribution in the Chromophyte Algae. In Green J.C., Leadbeater BSC, Diver WL (eds.) The Chromophyte Algae: Problems and Perspectives. Clarendon Press, Oxford, pp 13—36

Jeffrey S.W., Wright S.W (2005): Photosynthetic pigments in marine macroalgae: insights from cultures and the sea. Algal cultures, Analogues of blooms and Applications, Vol. 1. D.V. Subba Rao, ed. Enfield, NH, Science publishers, pp. 33-90.

Jiroutová K., Horák A., Bowler C., Oborník M. (2007): Tryptophan biosynthesis in stramenopiles: eukaryotic winners in the diatom complex chloroplast. *J. Mol. Evol.* 65:496–511.

Johnston A.M., Raven J.A. (1996): Inorganic carbon accumulation by the marine diatom *Phaeodactylum tricornutum*. *Eur. J. Phycol.* 31: 285–290

Jomaa H., Wiesner J., Sanderbrand S., Altincicek B., Weidemeyer C., Hintz M., Türbachova I., Eberl M., Zeidler J., Lichtenthaler H.K., Soldati D., Beck E.

(1999). Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* 285, 1573–1576.

Karlsson J., Clarke A.K., Chen Z.Y., Hugghins S.Y., Park Y.I., Husic H.D., Moroney J.V., Samuelsson G. (1998): A novel a-type carbonic anhydrase associated with the thylakoid membrane in *Chlamydomonas reinhardtii* is required for growth at ambient CO_2. *EMBO J*. 17: 1208–1216

Keeling P.J. (2009): Chromalveolates and the evolution of plastids by secondary endosymbiosis. *J. Eukaryot. Microbiol*. 56(1):1-8.

Kilian O., Kroth P.G. (2005): Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. *Plant. J.* 41(2):175-83.

Kohler S., Delwitche C.F, Denny P.W., Tilney L.G., Webster P., Wilson R.J., Palmer J.D., Roos D. S (1997): A plastid of probable green algal origin in apicomplexan parasites. *Science* 275:1485–1489.

Kooistra W.H., Gersonde C.F., Medlin L.K., Mann D.G. (2007): in *Evolution of Primary Producers in the Sea* (eds. Falkowski, P. G. and Knoll, A. H.) Elsevier 207–249.

Kowallik K.V., Stoebe B., Schaffran I., Kroth-Pancic P., Freier U. (1995): The chloroplast genome of a chlorophyll *a* + *c*-containing alga *Odontella sinensis. Plant. Mol. Biol. Rep.* 13, 336– 342.

Kröger N. (2007): Prescribing diatom morphology: toward genetic engineering of biological nanomaterials. *Curr. Opin. Chem. Biol.* 11(6):662-9.

Kroth P., Strotmann H. (1999): Diatom plastids: secondary endocytobiosis, plastid genome and protein import. *Physiol. Plant.* 107:136–141

Kroth PG, Chiovitti A, Gruber A, Martin-Jezequel V, Mock T, Parker MS, Stanley MS, Kaplan A, Caron L, Weber T, Maheswari U, Armbrust EV, Bowler C. (2008): A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS One* 3(1):e1426.

Kroth PG, Schroers Y, Kilian O. (2005): The peculiar distribution of class I and class II aldolases in diatoms and in red algae. *Curr. Genet.* 48(6):389-400.

Lang M, Apt KE, Kroth PG (1998): Protein transport into “complex” diatom plastids utilizes two different targeting signals. *J. Biol. Chem.* 273:30973–30978

Lang M (2000): Untersuchungen zum Transport kernkodierter Plastiden- Proteine in Kieselalgen. Ph.D. thesis. In **Gruber A., Vugrinec S., Hempel F., Gould S.B., Maier U.G., Kroth P.G.** (2007): Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. *Plant Mol. Biol.* Jul;64(5):519-30.

Lange, B.M., Rujan, T., Martin,W., Croteau, R. (2000): Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *PNAS* 97, 13172–13177.

Lang M., Apt K.E., Kroth P.G. (1998): Protein transport into "complex" diatom plastids utilizes two different targeting signals. *J. Biol. Chem.* 273(47):30973-8.

Leblanc C., Falciatore A., Watanabe M., Bowler C. (1999):. Semi-quantitative RT-PCR analysis of photoregulated gene expression in marine diatoms. Plant Mol. Biol. 40, 1031–1044.

Li S., Nosenko T., Hackett J.D., Bhattacharya D. (2006): Phylogenomic analysis identifies red algal genes of endosymbiotic origin in the chromalveolates. *Mol. Biol. Evol.* 23, 663–674.

Lopez,P.J., Descles J., Allen A.E., Bowler C. (2005): Prospects in diatom research. *Curr. Opin. Biotechnol*, 16, 180–186.

Maheswari U., Montsant A., Goll J., Krishnasamy S., Rajyashri K.R., Patell V.M., Bowler C. (2005): The Diatom EST Database. *Nucleic. Acids. Res.* 33: D344–347.

Maheswari U., Mock T., Armbrust E.V., Bowler C., (2009): Update of the Diatom EST Database: a new tool for digital transcriptomics. *Nucleic. Acids. Res.* 37(Database issue):D1001-5.

Marchetti A., Parker M.S., Moccia L.P., Lin E.O., Arrieta A.L., Ribalet F., Murphy M.E., Maldonado M.T., Armbrust E.V. (2008): Ferritin is used for iron storage in bloom-forming marine pennate diatoms. *Nature* 457, 467–470

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

Martin W., Rujan T., Richly E., Hansen A., Cornelsen S., Lins T., Leister D., Stoebe B., Hasegawa M., Penny D. (2002): Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *PNAS* 99(19):12246-12251

Martin W., Mustafa A.-Z., Henze K., Schnarrenberger C. (1996): Higher- plant chloroplast and cytosolic fructose-1,6-bisphosphatase isoenzymes: origins via duplication rather than prokaryote –eukaryote divergence. *Plant. Mol. Biol.* 32:485–491

Martin W., **Schnarrenberger C.** (1997): The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. *Curr. Genet*. 32(1):1-18

Martin-Jezequel V., Hildenbrand M., Brzezinski M.A. (2000): Silicon metabolism in diatoms: Implications for growth. *J. Phycol*. 36:821:840

Matchett W.H., DeMoss J.A. (1975): The subunit structure of tryptophan synthase from *Neurospora crassa*. *J. Biol. Chem.* 250:2941–2946

Matsuzaki M., Misumi O., Shin-I T., Maruyama S., Takahara M., Miyagishima S.Y., Mori T., Nishida K., Yagisawa F., Nishida K., Yoshida Y., Nishimura Y., Nakao S., Kobayashi T., Momoyama Y., Higashiyama T., Minoda A., Sano M., Nomoto H., Oishi K., Hayashi H., Ohta F., Nishizaka S., Haga S., Miura S., Morishita T., Kabeya Y., Terasawa K., Suzuki Y., Ishii Y., Asakawa S., Takano H., Ohta N., Kuroiwa H., Tanaka K., Shimizu N., Sugano S., Sato N., Nozaki H., Ogasawara N., Kohara Y., Kuroiwa T. (2004): Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* 428(6983):653-7.

Medlin L.K., Kaczmarska I. (2004): Evolution of diatoms: V. Morphological and cytological support for the major clade and taxonomic revision. *Phycologia* 43:245-270

McConkey G.A. (1999): Targeting the shikimate pathway in the malaria parasite *Plasmodium falciparum*. *Antimicrob. Agents Chemother*. 43:175– 177

McConkey G.A., Pinney J.W., Westhead D.R., Plueckhahn K., Fitzpatrick T.B., Macheroux P., Kappes B. (2004): Annotating the Plasmodium genome and the enigma of the shikimate pathway. *Trends Parasitol.* 20:60–65

McEwan M.L., Keeling PJ. (2004): HSP90, tubulin and actin are retained in the tertiary endosymbiont genome of *Kryptoperidinium foliaceum*. *J. Eukaryot. Microbiol.* 51(6):651-9.

McFadden G.I., Reith M.E., Munholland J., Lang-Unnasch N. (1996): Plastid in human parasites. *Nature* 381(6582):482.

McFadden G.I., van Dooren G.G. (2004): Evolution: red algal genome affirms a common origin of all plastids. *Curr. Biol.* 14(13):R514-6.

McFadden G.I. (2000): Mergers and acquisitions: malaria and the great chloroplast heist. *Genome Biol.* 1(4): 1026.

McFadden G.I. (1999): Plastids and protein targeting. *J. Eukaryot. Microbiol*. 46(4):339-46.

Michels A., Wedel N., Kroth P. (2005): Diatom plastids possess a phosphoribulokinase with an altered regulation and no oxidative pentose phosphate pathway. *Plant Physiol*. 137:911–920

Millen R.S., Olmstead R.G., Adams K.L., Palmer J.D., Lao N.T., Heggie L., Kavanagh T.A., Hibberd J.M., Gray J.C., Morden C.W., Calie P.J., Jermiin L.S., Wolfe

K.H. (2001): Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* 13(3):645-58

Mock T., Hoch N. (2005): Long-term temperature acclimation of photosynthesis in steady-state cultures of the polar diatom *Fragilariopsis cylindrus*. *Photosynth Res.* 85(3):307-17.

Mock T., Samanta M.P., Iverson V., Berthiaume C., Robison M., Holtermann K., Durkin C., Bondurant S.S., Richmond K., Rodesch M., Kallas T., Huttlin E.L., Cerrina F., Sussman M.R., Armbrust E.V. (2008): Whole-genome expression profiling of the marine diatom Thalassiosira pseudonana identifies genes involved in silicon bioprocesses. *PNAS* 105(5):1579-84.

Montsant A., Jabbari K., Maheswari U., Bowler C. (2005a): Comparative genomics of the pennate diatom *Phaeodactylum tricornutum*. *Plant. Physiol.* 137: 500–513.

Montsant A., Maheswari U., Bowler C., Lopez P.J. (2005b) Diatomics: Toward diatom functional genomics. *J. Nanosci. Nanotechnol.* 5, 5–14.

Montsant A., Allen A.E., Coesel S., De Martino A., Falciatore A., Mangogna M., Siaut M., Heijde M., Jabbari K., Maheswari U., Rayko E., Vardi A., Apt K.E., Berges J.A., Chiovitti A., Davis A.K., Thamatrakoln K., Hadi M.Z., Lane T.W., Lippmeier J.C., Martinez D., Parker M.S., Pazour G.J., Saito M.A., Rokhsar D.S., Armbrust E.V., Bowler C. (2007): Identification and comparative genomic analysis of signaling and regulatory components in the diatom *Thalassiosira pseudonana*. *J. Phycol.* 43, 585–604.

Moore R.B., Oborník M., Janouskovec J., Chrudimský T., Vancová M., Green D.H., Wright S.W., Davies N.W., Bolch C.J., Heimann K., Slapeta J., Hoegh-Guldberg O., Logsdon J.M., Carter D.A. (2008): A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* 451(7181):959-63

Morosinotto T., Caffarri S., Dall'Osto L., Bassi R. (2003): Mechanistic aspects of the xanthophyll dynamics in higher plant thylakoids. *Physiol. Plant* 119:347–354.

Mousdale D.M., Campbell M.S., Coggins J.R., (1987): Purification and characterisation of a bi-functional dehydroquinase-shikimate: NADP oxidoreductase from pea seedlings. *Phytochemistry* 26:2665–2670.

Moustafa A., Beszteri B., Maier U.G., Bowler C., Valentin K., Bhattacharya D. (2009): Genomic footprints of a cryptic plastid endosymbiosis in diates Science. 324(5935):1724-6.

Mustafa N.R., Verpoorte R. (2005): Chorismate derived C6C1 compounds in plants. *Planta* 222(1):1-5.

Nelson M.J., Dang Y., Filek E., Zhang Z., Yu V.W., Ishida K., Green B.R. (2007): Identification and transcription of transfer RNA genes in dinoflagellate plastid minicircles *Gene.* 392(1-2):291-8.

Nishi A., Ishida K., Endoh H. (2005): Reevaluation of the evolutionary position of opalinids based on 18S rDNA, and alpha- and beta-tubulin gene phylogenies *J. Mol. Evol.* 60(6):695-705

Nowack E.C., Melkonian M., Glöckner G. (2008): Chromatophore genome sequence of Paulinella sheds light on acquisition of photosynthesis by eukaryotes. *Curr. Biol.* 18(6):410-8.

Nozaki, H. Iseki M., Hasegawa M., Misawa K., Nakada T., Sasaki N., Watanabe M. (2007): Phylogeny of primary photosynthetic eukaryotes as deduced from slowly evolving nuclear genes. *Mol. Biol. Evol*. 24(8):1592-5

Oborník M., Green B.R. (2005): Mosaic origin of the heme biosynthesis pathway in photosynthetic eukaryotes. *Mol. Biol. Evol.* 22:2343–2353

Oborník M, Janouškovec J., Chrudimský T., Lukeš J. (2009): Evolution of the apicoplast and its host: From heterotrophy to autotrophy and back again. *Int. J. Parasitol.* 39, 1-12.

Oeltjen A., Marquardt J., Rhiel E. (2004): Differential circadian expression of genes fcp2 and fcp6 in *Cyclotella cryptica*. *Int. Microbiol.* 7, 127–131.

Ohta N., Sato N., Kawano S., Kuroiwa T. (1993): The trpA gene on the plastid genome of *Cyanidium caldarium* strain RK-1. *Curr. Genet.* 25:357–361

Oudot-Le Secq M.P., Grimwood J., Shapiro H., Armbrust E.V., Bowler C., Green B.R. (2007): Chloroplast genomes of the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*: comparison with other plastid genomes of the red lineage. *Mol. Genet. Genom.* 277, 427-439.

Palmer J.D. (2003). The symbiotic birth and spread of plastids: how many times and whodunit? *J. Phycol. 39*, 4–11.

Park H., Song B., Morel F.M. (2007): Diversity of the cadmium-containing carbonic anhydrase in marine diatoms and natural waters. *Environ. Microb.* 9: 403–413.

Patron N.J., Inagaki Y., Keeling P.J. (2007): Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. *Curr. Biol.* 17(10):887-91.

Patron N.J., Rogers M.B., Keeling P.J. (2004): Gene replacement of fructose-1,6-bisphosphate aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates. *Eukaryot. Cell.* 3, 1169–1175.

Pennington F., Guillard R.R.L., Liaaen-Jensen S. (1988): Carotenoid distribution patterns in Bacillariophyceae (diatoms). *Biochem. Syst. Ecol.* 16: 589—592

Pienaar R.N., Sakai H., Horiguchi T. (2007): Description of a new dinoflagellate with a diatom endosymbiont, *Durinskia capensis* sp. nov. (Peridiniales, Dinophyceae) from South Africa. *J. Plant. Res.* 120(2):247-58.

Poulsen N.; Chesley P.M., Kroger N. (2006): Molecular genetic manipulation of the diatom *Thalassiosira pseudonana* (Bacillariophyceae). *J. Phycol.* 42 (5): 1059-1065

Radwanski E.R., Last R.L. (1995): Tryptophan biosynthesis and metabolism: biochemical and molecular genetics. *Plant Cell* 7:921–934

Reyes-Prieto A., Moustafa A., Bhattacharya D. (2008): Multiple genes of apparent algal origin suggest ciliates may once have been photosynthetic. *Curr. Biol.* 18(13):956-62

Reith M.E., Munholland J. (1995): Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. *Plant. Mol. Biol. Rep.* 13, 333–335.

Riebesell U., Wolf-Gladrow D.A., Smetacek V. (1993)???: Carbon dioxide limitation of marine phytoplankton growth rates. *Nature* 361: 249–251.

Richards T.A., Dacks J.B., Campbell S.A., Blanchard J.L., Foster P.G., McLeod R., Roberts C.W. (2006): Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements. *Eukaryot. Cell*. 5(9):1517-31.

Rodriguez-Ezpeleta N., Brinkmann H., Burey S.C., Roure B., Burger G., Löffelhardt W., Bohnert H.J., Philippe H., Lang B.F. (2005): Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr. Biol*. 15:1325–1330

Rogers M., Keeling P.J. (2004): Lateral transfer and recompartmentalization of Calvin cycle enzymes of plants and algae. *J. Mol. Evol.* 58(4):367-75.

Round F.E., Mann D.G., Crawford R.M. (1990): The Diatoms: Biology & Morphology of the Genera. Cambridge University Press;

Rubin J.L., Jensen R.A. (1985): Differentially regulated isozymes of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase from seedlings of *Vigna radiata* (L.) Wilczek. *Plant Physiol* 79:711– 718

Sahr T., Ravanel S., Basset G., Nichols B.P., Hanson A.D., Rébeillé F. (2006): Folate synthesis in plants: purification, kinetic properties, and inhibition of aminodeoxychorismate synthase. *Biochem. J.* 396(1):157-62.

Sánchez-Puerta M.V., Bachvaroff T.R., Delwiche (2005): The complete plastid genome sequence of the haptophyte *Emiliania huxleyi*: a comparison to other plastid genomes. *DNA Res.* (2):151-6

Sanchez-Puerta M.V., Lippmeier J.C., Apt K.E., Delwiche C.F. (2007): Plastid genes in a non-photosynthetic dinoflagellate. *Protist* 158(1):105-17.

Sato S., Wilson R. J. M. (2003): Proteobacteria-like ferrochelatase in the malaria parasite. *Curr. Genet.* 42:292–300.

Schmidt, S., Sunyaev S., Bork P., Dandekar T. (2003): Metabolites: a helping hand for pathway evolution? *Trends Biochem. Sci.* 28:336–341.

Schnarrenberger C., Pelzer-Reith B., Yatsuki H., Freund S., Jacobshagen S., Hori K. (1994): Expression and sequence of the only detectable aldolase in *Chlamydomonas reinhardtii*. *Arch. Biochem. Biophys.* 313:173–178

Siaut M., Heijde M., Mangogna M., Montsant A., Coesel S., Allen A., Manfredonia A., Falciatore A., Bowler C. (2007): Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. *Gene* 406:23-35.

Sims P.A., Mann D.G., Medlin L.K. (2006): Evolution of the diatoms: insights from fossil, biological and molecular data. *Phycologia* 45, 361–402

Slapeta J. (2006): *Cryptosporidium* species found in cattle: a proposal for a new species. *Trends Parasitol.* 22(10):469-74.

Soll J., Schleiff E. (2004): Protein import into chloroplasts. *Nat. Rev.* 5:198–208

Stegemann S., Bock R. (2006): Experimental reconstruction of functional gene transfer from the tobacco plastid genome to the nucleus. *Plant Cell* (11):2869-78.

Subramaniam P.S., Xie G., Xia T., Jensen R.A. (1998): Substrate ambiguity of 3-deoxy-D-*manno*-octulosonate 8-phosphate synthase from *Neisseria gonorrhoeae* in the context of its membership in a protein family containing a subset of 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthases. *J. Bacteriol.* 180:119–127.

Sumper M., Brunner E. (2008): Silica biomineralization in diatoms: the model organism Thalassiosira pseudonana. *Chem.biochem.* 9(8):1187-94.

Takishita K., Yamaguchi H., Maruyama T., Inagaki Y. (2009): A hypothesis for the evolution of nuclear-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase genes in "chromalveolate" members. *PLoS One* 4(3):e4737.

Tanaka Y., Nakatsuma D., Harada H., Ishida M., Matsuda Y. (2005): Localization of soluble beta-carbonic anhydrase in the marine diatom *Phaeodactylum tricornutum*.

Sorting to the chloroplast and cluster formation on the girdle lamellae. *Plant Physiol.* 138: 207–17.

Teich R., Zauner S., Baurain D., Brinkmann H., Petersen J. (2007): Origin and distribution of Calvin cycle fructose and sedoheptulose bisphosphatases in plantae and complex algae: a single secondary origin of complex red plastids and subsequent propagation via tertiary endosymbioses. *Protist* 158(3):263-76.

Tengs T., Dahlberg O.J., Shalchian-Tabrizi K., Klaveness D., Rudi K., Delwiche C.F., Jakobsen K.S. (2000): Phylogenetic analyses indicate that the 19'Hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin**.** *Mol. Biol. Evol.* 17(5):718-29.

Theissen U., Martin W. (2006): The difference between organelles and endosymbionts. *Curr. Biol.* 16, 1016–1017

Tsui C.K., Marshall W., Yokoyama R., Honda D., Lippmeier J.C., Craven K.D., Peterson P.D., Berbee M.L. (2009): Labyrinthulomycetes phylogeny and its implications for the evolutionary loss of chloroplasts and gain of ectoplasmic gliding. *Mol. Phylogenet. Evol.* 50(1):129-40.

Tyler B.M., Tripathy S., Zhang X., Dehal P., Jiang R.H., Aerts A., Arredondo F.D., Baxter L., Bensasson D., Beynon J.L., Chapman J., Damasceno C.M., Dorrance A.E., Dou D., Dickerman A.W., Dubchak I.L., Garbelotto M., Gijzen M., Gordon S.G., Govers F., Grunwald N.J., Huang W., Ivors K.L., Jones R.W., Kamoun S., Krampis K., Lamour K.H., Lee M.K., McDonald W.H., Medina M., Meijer H.J., Nordberg E.K., Maclean D.J., Ospina-Giraldo M.D., Morris P.F., Phuntumart V., Putnam N.H., Rash S., Rose J.K., Sakihama Y., Salamov A.A., Savidor A., Scheuring C.F., Smith B.M., Sobral B.W., Terry A., Torto-Alalibo T.A., Win J., Xu Z., Zhang H., Grigoriev I.V., Rokhsar D.S., Boore J.L. (2006): *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 313(5791):1261-6

van Dooren G.G., Schwartzbach S.D., Osafune T., McFadden G.I. (2001): Translocation of proteins across the multiple membranes of complex plastids. *Biochim. Biophys. Acta* 1541(1-2):34-53.

Walker G.E, Dunbar B., Hunter I.S., Nimmo H.G., Coggins J.R. (1996): Evidence for a novel class of microbial 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase in *Streptomyces coelicolor* A3(2), *Streptomyces rimosus* and *Neurospora crassa*. *Microbiology* 142:1973-82.

Waller R.F., Keeling P.J, Donald R.G., Striepen B., Handman N., Lang-Unnasch, Cowman A.F., Besra G.S., Roos D.S., McFadden G.I. (1998): Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum. PNAS* 95:12352–12357.

Watanabe M.M., Takeda Y., Sasa T., Inouye I., Suda S., Sawaguchi T., Chihara M.A. (1987): green dinoflagellate with chlorophylls *a* and *b* morphology fine structure of the chloroplast and chlorophyll composition. *J. Phycol.* 23:382–389.

Watanabe N., Che F.-S., Iwano M., Takayama S., Yoshida S., and Isogai A. (2001): Dual targeting of spinach protoporphyrinogen oxidase II to mitochondria and chloroplasts by alternative use of two in-frame initiation codons. *J. Biol. Chem.* 276: 20474–20481.

Webby C.J., Lott J.S., Baker H.M., Baker E.N., Parker E.J. (2005): Crystallization and preliminary X-ray crystallographic analysis of 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from *Mycobacterium tuberculosis*. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* 1:61:403-6.

Weber A.P., Linka M., Bhattacharya D. (2006): Single, ancient origin of a plastid metabolite translocator family in Plantae from an endomembrane-derived ancestor. *Eukaryot. Cell* 5, 609–612

Wichard T., Gerecht A., Boersma M., Poulet S.A., Wiltshire K., Pohnert G. (2007): Lipid and fatty acid composition of diatoms revisited: rapid wound-activated change of food quality parameters influences herbivorous copepod reproductive success. *Chem.biochem.* 8(10):1146-53

Wilhelm C., **Büchel C.**, **Fisahn J.**, **Goss R.**, **Jakob T.**, **Laroche J.**, **Lavaud J.**, **Lohr M.**, **Riebesell U.**, **Stehfest K.**, **Valentin K.**, **Kroth P.G**. (2006): The regulation of carbon and nutrient assimilation in diatoms is significantly different from green algae. *Protist* 157(2):91-124

Wittpoth C., Kroth P.G., Weyrauch K., Kowallik K.V., Strotmann H. (1998): Functional characterization of isolated plastids from two marine diatoms. *Planta* 260 (1): 79-85

Yoon H.S., Hackett J.D., Ciniglia C., Pinto G., Bhattacharya D. (2004): A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol*. 21: 809–818.

Yoon H.S., Hackett J.D., Pinto G., Bhattacharya D. (2002): The single, ancient origin of chromist plastids. *PNAS* 99(24):15507-12

Zaslavskaia, L. A., Lippmeier, J. C., Kroth, P. G., Grossman, A. R. and Apt, K. E. (2000): Additional selectable marker and reporter genes for use in transforming the diatom *Phaeodactylum tricornutum. J. Phycol.* **36**,379 -386.

Zaslavskaia, L.A., Lippmeier, J.C., Shih, C., Ehrhardt, D., Grossman, A.R., Apt, K.E. (2001): Trophic obligate conversion of an photoautotrophic organism through metabolic engineering. *Science* 292, 2073–2075.

Zhao J., Weaver L.M., Herrmann K.M. (2002): Translocation of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase precursor into isolated chloroplasts. *Planta* 216(1):180-6.

Zielinski U., Gersonde R. (1997): Diatom distribution in Southern Ocean surface sediments (Atlantic sector): implications for paloeenvironmental reconstructions. *Palaeogeogr. Palaeoclimat. Palaeoecol.* 129, 213–250

Zimmermann G, Kelly G.J., Latzko E. (1976): Efficient purification and molecular properties of spinach chloroplast fructose 1,6- bisphosphatase. *Eur. J. Biochem.* 70:361–367

9. Objectives

The main aim of this thesis was to describe the origin and localization of enzymes working in tryptophan biosynthesis in *Thalassiosira pseudonana*. During the PhD studies the new diatom sequenced genome of *Phaeodactylum tricornutum* came out. This brings the opportunity to annotate the genes encoding tryptophan pathway.

Particular objectives:

- Identification of genes encoding tryptophan biosynthesis in diatoms, particularly the verification of the targeting signal encoded on full-length cDNA.
- Identification of genes encoding tryptophan biosynthesis in the other eukaryotes and partly in prokaryotes.
- Functional comparison and phylogenetic analysis of the enzymes synthesizing tryptophan among all organisms so far sequenced.

The second part of this PhD. research refers the technique of verifying the functionality of nuclear-encoded protein targeting into diatom plastid by *in vivo* assay of transformation of *P. tricornutum*.

Particular objectives:

- Characterization of genes encoding Psb28 protein in the photosynthetic eukaryotes.
- Phylogenetic relation of eukaryotic Psb28, with the emphasis on two copies of gene encoding Psb28 in *T. pseudonana*.
- Amplifying the cDNA of nuclear encoded Psb28 from *T. pseudonana* and subsequent cloning into diatom expressing vector. Transformation of model organism *P. tricornutum* by this construct and evaluation of plastid localization in diatoms.

10. Results

Enclosed publications and manuscript:

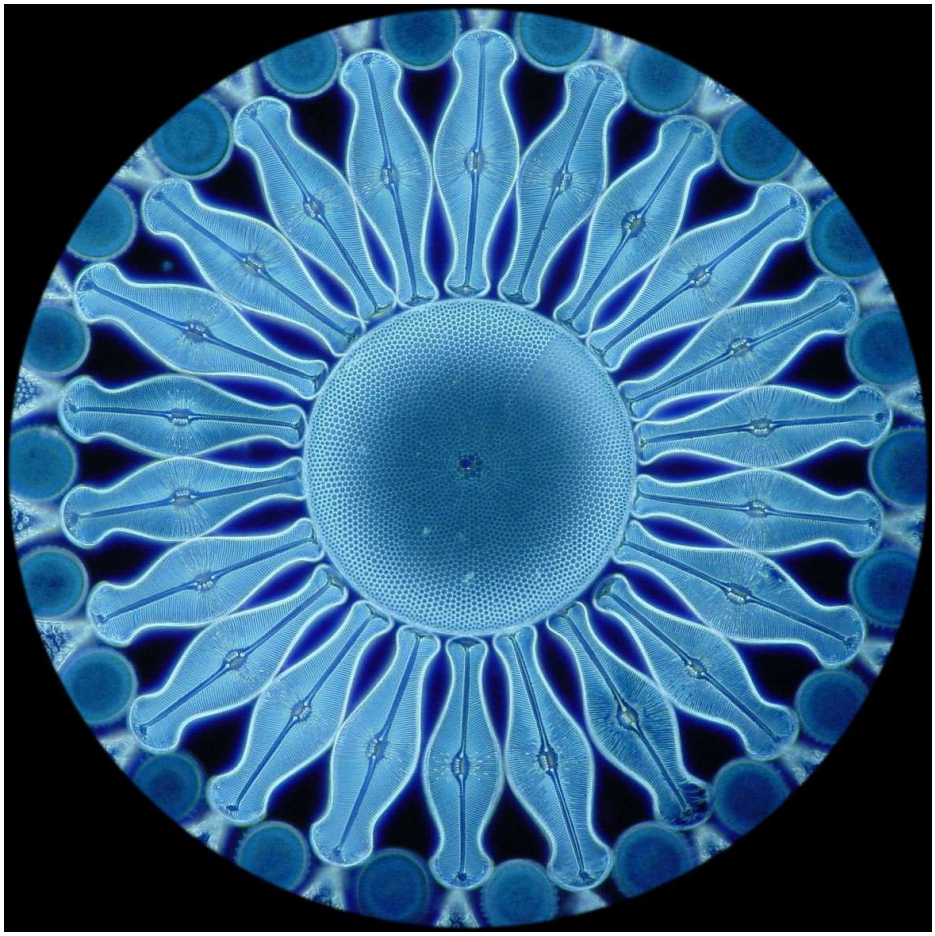
I. Jiroutová K., Horák A., Bowler C., Oborník M., 2007. Tryptophan biosynthesis in stramenopiles: eukaryotic winners in the diatom complex chloroplast. *J. Mol. Evol.* 65:496–511.

II. Bowler C., Allen A.E., Badger J.H., Grimwood J., Jabbari K., Kuo A., Maheswari U., Martens C., Maumus F., Otillar R.P., Rayko E., Salamov A., Vandepoele K., Beszteri B., Gruber A., Heijde M., Katinka M., Mock T., Valentin K., Verret F., Berges J.A., Brownlee C., Cadoret J.P., Chiovitti A., Choi C.J., Coesel S., De Martino A., Detter J.C., Durkin C., Falciatore A., Fournet F., Haruta M., Huysman M.J.J., Jenkins B.D., **Jiroutová K.**, Jorgensen R.E., Joubert Y., Kaplan A., Kröger N., Kroth P.G., La Roche J., Lindquist E., Lommer M., Martin–Jézéquel V., Lopez P.J., Lucas S., Mangogna M., McGinnis K., Medlin L.K., Montsant A., Oudot–Le Secq M.P., Napoli C., Oborník M., Schnitzler-Parker M., Petit J.L., Porcel B.M., Poulsen N., Robison M., Rychlewski L., Rynearson T.A., Schmutz J., Shapiro H., Siaut M., Stanley M., Sussman M.R., Taylor A.R., Vardi A., von Dassow P., Vyverman W., Willis A., Wyrwicz L.S., Rokhsar D.S., Weissenbach J., Armbrust V.E., Green B.R., Van de Peer Y., Grigoriev I.V., 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456:239-244.

III. Jiroutová K., Bowler C., Oborník M., 2009. A gene in the process of endosymbiotic gene transfer. unpublished

I.

Tryptophan Biosynthesis in Stramenopiles: Eukaryotic Winners in the Diatom Complex Chloroplast



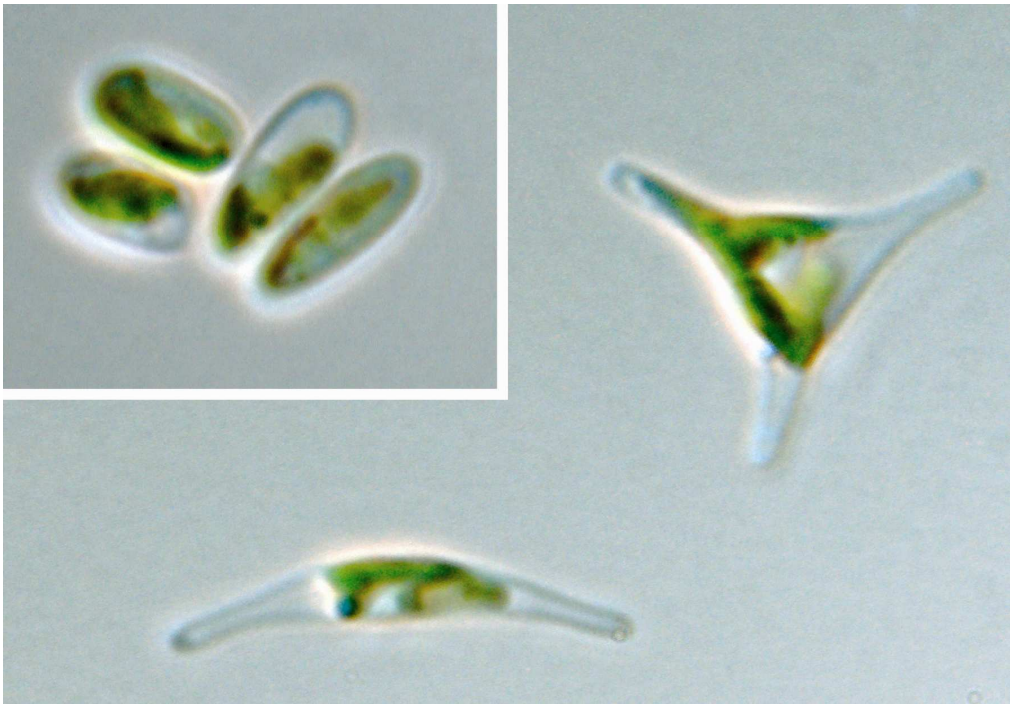
Journal of Molecular Evolution

Jiroutová K, Horák A, Bowler C, Oborník M. 2007. Tryptophan biosynthesis in stramenopiles: eukaryotic winners in the diatom complex chloroplast. *J. Mol. Evol.* 65, 496–511.

Tryptophan is an essential amino acid that, in eukaryotes, is synthesized either in the plastids of photoautotrophs or in the cytosol of fungi and oomycetes. Here we present an *in silico* analysis of the tryptophan biosynthetic pathway in stramenopiles, based on analysis of the genomes of the oomycetes *Phytophthora sojae* and *P. ramorum* and the diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*. Although the complete pathway is putatively located in the complex chloroplast of diatoms, only one of the involved enzymes, indole-3-glycerol phosphate synthase (InGPS), displays a possible cyanobacterial origin. On the other hand, in *P. tricornutum* this gene is fused with the cyanobacteria-derived hypothetical protein COG4398. Anthranilate synthase is also fused in diatoms. This fusion gene is almost certainly of bacterial origin, although the particular source of the gene cannot be resolved. All other diatom enzymes originate from the nucleus of the primary host (red alga) or secondary host (ancestor of chromalveolates). The entire pathway is of eukaryotic origin and cytosolic localization in oomycetes; however, one of the enzymes, anthranilate phosphoribosyl transferase, was likely transferred to the oomycete nucleus from the red algal nucleus during secondary endosymbiosis. This suggests possible retention of the complex plastid in the ancestor of stramenopiles and later loss of this organelle in oomycetes.

II.

The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes



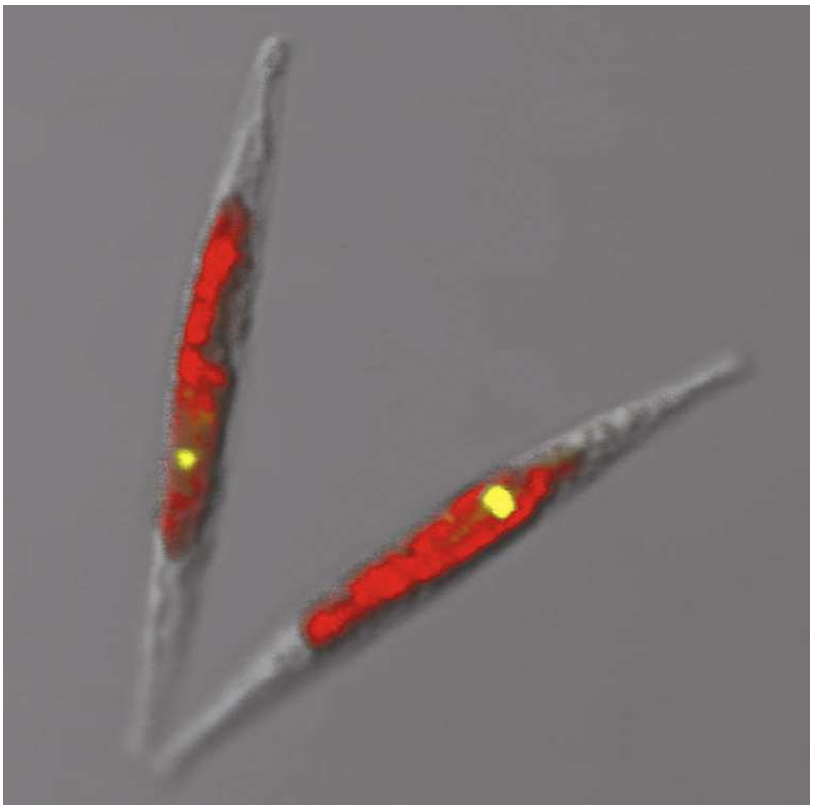
Nature Vol. 456

Bowler C., Allen A.E., Badger J.H., Grimwood J., Jabbari K., Kuo A., Maheswari U., Martens C., Maumus F., Otillar R.P., Rayko E., Salamov A., Vandepoele K., Beszteri B., Gruber A., Heijde M., Katinka M., Mock T., Valentin K., Verret F., Berges J.A., Brownlee C., Cadoret J.P., Chiovitti A., Choi C.J., Coesel S., De Martino A., Detter J.C., Durkin C., Falciatore A., Fournet F., Haruta M., Huysman M.J.J., Jenkins B.D**., Jiroutová K**., Jorgensen R.E., Joubert Y., Kaplan A., Kröger N., Kroth P.G., La Roche J., Lindquist E., Lommer M., Martin–Jézéquel V., Lopez P.J., Lucas S., Mangogna M., McGinnis K., Medlin L.K., Montsant A., Oudot–Le Secq M.P., Napoli C., Oborník M., Schnitzler-Parker M., Petit J.L., Porcel B.M., Poulsen N., Robison M., Rychlewski L., Rynearson T.A., Schmutz J., Shapiro H., Siaut M., Stanley M., Sussman M.R., Taylor A.R., Vardi A., von Dassow P., Vyverman W., Willis A., Wyrwicz L.S., Rokhsar D.S., Weissenbach J., Armbrust V.E., Green B.R., Van de Peer Y., Grigoriev I.V., 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239-244.

Diatoms are photosynthetic secondary endosymbionts found throughout marine and freshwater environments, and are believed to be responsible for around one-fifth of the primary productivity on Earth. The genome sequence of the marine centric diatom Thalassiosira pseudonana was recently reported, revealing a wealth of information about diatom biology. Here we report the complete genome sequence of the pennate diatom Phaeodactylum tricornutum and compare it with that of T. pseudonana to clarify evolutionary origins, functional significance and ubiquity of these features throughout diatoms. In spite of the fact that the pennate and centric lineages have only been diverging for 90 million years, their genome structures are dramatically different and a substantial fraction of genes (40%) are not shared by these representatives of the two lineages. Analysis of molecular divergence compared with yeasts and metazoans reveals rapid rates of gene diversification in diatoms. Contributing factors include selective gene family expansions, differential losses and gains of genes and introns, and differential mobilization of transposable elements. Most significantly, we document the presence of hundreds of genes from bacteria. More than 300 of these gene transfers are found in both diatoms, attesting to their ancient origins, and many are likely to provide novel possibilities for metabolite management and for perception of environmental signals. These findings go a long way towards explaining the incredible diversity and success of the diatoms in contemporary oceans.

III.

A gene in the process of endosymbiotic gene transfer



Jiroutová K., Bowler C., Oborník M., 2009. A gene in the process of endosymbiotic gene transfer. (unpublished manuscript)

The endosymbiotic birth of organelles is accompanied by massive transfer of endosymbiont genes to the eukaryotic host nucleus. In the centric diatom *Thalassiosira pseudonana* the gene encoding Psb28 has likely been duplicated during diatom evolution because one copy is still encoded in the plastid genome while a second nuclear-encoded version possesses a bipartite N-terminal presequence necessary to target the protein into the diatom complex plastid. Both versions of the gene are transcribed. Here we provide experimental evidence for successful targeting of the nuclear-encoded Psb28 to the diatom multimembrane plastid. Extensive phylogenetic analyses demonstrate that nucleotide composition of the analyzed genes deeply influences the tree topology and that appropriate methods designed to deal with nucleotide and amino acid bias need to be used to overcome this obstacle.

11. Conclusions

Tryptophan synthesis in diatoms has extremely mosaic character:

- **Anthranilate synthase in diatoms is bacterial in origin and obtained by LGT.**
- **Stramenopile‘s APRT displays „red“ origin; evidence that Oomycetes lost their secondary plastid.**
- **Tryptophan synthase and Prai/InGP fusion show common origin with fungi**
- **One of IGPS is the only enzyme of plastid (cyanobacterial) origin involved in the diatom tryptophan synthesis. Interestingly, *P. tricornutum* possess the unique fusion of InGPS with the cyanobacteria-derived hypothetical protein COG4398.**
- ***P. tricornutum* and *T. pseudonana* differ in extra encoded beta subunit of TS**
- **Pennate diatom *P. tricornutum* has a single nuclear-encoded Psb28 gene, whereas the *T. pseudonana* possesses two versions of the gene, one encoded in plastome and second in nuclear genome.**
- **Both *T. pseudonana* versions of Psb28 are transcribed, with the nuclear-encoded Psb28 being targeted to the complex plastid.**
- **The localization of *T. pseudonana* nuclear-encoded Psb28 was tested by transformation and in vivo expression of this protein in the pennate diatom *P. tricornutum*.**
- **Phylogenetic analyses show that nuclear encoded Psb28 is a duplicate of the plastid homologue, although this relationship only became apparent when methods designed to deal with phylogenetic artifacts**
- **The nuclear encoded Psb28 version contains many non-homologous mutations and both nucleotide and amino acid sequences show surprisingly high mutual divergence caused probably by BIAS.**