Intracellular metabolic pathway distribution in diatoms and tools for genome-enabled experimental diatom research

Ansgar Gruber and Peter G. Kroth

Fachbereich Biologie, Universität Konstanz, 78457 Konstanz, Germany

AG, 0000-0002-5876-4391; PGK, 0000-0003-4734-8955

Diatoms are important primary producers in the oceans and can also dominate other aquatic habitats. One reason for the success of this phylogenetically relatively young group of unicellular organisms could be the impressive redundancy and diversity of metabolic isoenzymes in diatoms. This redundancy is a result of the evolutionary origin of diatom plastids by a eukaryote–eukaryote endosymbiosis, a process that implies temporary redundancy of functionally complete eukaryotic genomes. During the establishment of the plastids, this redundancy was partially reduced via gene losses, and was partially retained via gene transfer to the nucleus of the respective host cell. These gene transfers required re-assignment of intracellular targeting signals, a process that simultaneously altered the intracellular distribution of metabolic enzymes compared with the ancestral cells. Genome annotation, the correct assignment of the gene products and the prediction of putative function, strongly depends on the correct prediction of the intracellular targeting of a gene product. Here again diatoms are very peculiar, because the targeting systems for organelle import are partially different to those in land plants. In this review, we describe methods of predicting intracellular enzyme locations, highlight findings of metabolic peculiarities in diatoms and present genome-enabled approaches to study their metabolism.

This article is part of the themed issue 'The peculiar carbon metabolism in diatoms'.

Subject Areas:
physiology, bioinformatics, biochemistry, genomics, plant science, evolution

Keywords:
secondary plastids, gene transfer, protein targeting, chloroplast, algae, metabolism

Author for correspondence:
Ansgar Gruber
e-mail: ansgar.gruber@uni-konstanz.de

1. Introduction

The larger part of all organic biomass on Earth was produced either directly or indirectly by oxygenic photosynthesis, a process that evolved in cyanobacteria earlier than 2.35 billion years ago [1]. At this time oxygenic photosynthesis led to the slow accumulation of molecular oxygen (O_2) in the water or atmosphere and consequently to the availability of O_2 in most habitats, paving the way for aerobic heterotrophic organisms [1,2]. Today, a large fraction of the global photosynthetic carbon/oxygen turnover in the atmosphere takes place in eukaryotes [3,4], which perform photosynthesis in specialized organelles, the plastids. These organelles originate from cyanobacteria, which were taken up by a heterotrophic ancestor of photosynthetic eukaryotes and, during an endosymbiotic relationship with their eukaryotic host, evolved into plastids [5,6]. Cyanobacteria contributed to the organelle evolution in two ways: one is structural, the other is genetic. The structural contribution of cyanobacteria is evident in the membranes/compartments that define the plastid, and in its requirement for semiautonomous replication (rather than de novo formation) within the eukaryotic cell [6,7]. The genetic contribution of cyanobacteria is evident in the cyanobacterial genes that, for instance, encode the proteins required for photosynthesis. The exact extent of this genetic contribution is hard to determine, because only a small part of the plastid proteins are encoded on the plastid genome (which is a direct descendent of the cyanobacterial genome). The majority of the plastid proteins are nucleus encoded and targeted to the plastid only after translation in the cytosol [6,8,9].

Among the different groups of eukaryotes, only one clade, the Archaeplastida, contains plastids that directly arose from cyanobacterial ancestors [10]. These plastids are called primary plastids and are found in green algae, red algae and glaucophytes, which apparently form a monophyletic group [10–12]. However, the Archaeplastida represent only a small fraction of the extant photosynthetic eukaryotes. The majority of photosynthetic eukaryotes contain plastids that evolved by eukaryote–eukaryote endosymbioses; the exact phylogenetic relationships of these plastids and their hosts are currently debated [13]. Also the eukaryotic ancestors of complex plastids contributed structurally as well as genetically to the emerging organisms. Structurally, the eukaryotic ancestry of these plastids is evidenced in the presence of additional membranes and compartments surrounding them [7]. Therefore, plastids derived from eukaryotic algae are also called complex plastids [13,14]. Even though this term initially referred to structural complexity, the evolution of complex plastids also considerably increased the genetic complexity of the resulting cells.

In the following, we discuss the genetic background and the cellular physiology resulting from the acquisition of complex plastids on the example of diatoms, a group of high ecological relevance among the organisms with complex plastids. Diatoms are abundant unicellular algae in aquatic habitats, producing an enormous amount of biomass. Up to 16 gigatonnes of the organic carbon produced by marine phytoplankton per year, or about one-third of total ocean production, is thought to sink into the ocean interior preventing re-entrance of this carbon into the atmosphere for centuries [15]. The role diatoms play in mitigating atmospheric CO_2 concentrations is of special interest, especially now with the rising levels of this greenhouse gas and consequent global warming [16]. A significant fraction of the organic carbon generated by diatoms remains in the upper ocean and supports production by higher trophic levels and bacteria [4]. Despite the important role of diatoms for the global carbon cycle, relatively little is known about the carbohydrate metabolism of these algae [17]. For example, the exact mode of CO_2 fixation is largely unsolved (see [18,19] in this issue); furthermore, although most of the Calvin cycle enzymes in diatoms are very similar to those in land plants, their regulation by light via thioredoxins is rather different [20]. Diatoms can produce and secrete vast amounts of carbohydrates that play important roles in phototrophic biofilms [21], but which can also be used for driving cell motility [22], used as a storage compound [23], or simply be the basic compound of holdfasts [24]. Very little is known about the synthesis and secretion of these carbohydrates [25].

In recent years, a set of diatom genomes have been published, including the centric diatoms *Thalassiosira pseudonana* [26] and *T. oceanica* [27], as well as the pennate diatoms *Fragilariopsis cylindrus* [28], *Fistulifera solaris* [29], *Synedra acus* [30] and *Phaeodactylum tricornutum* [31]. Additionally, a large set of transcriptomic data from different diatoms is available. Diatoms contain three different genomes, the nuclear genome, the plastid genome and the mitochondrial genome. While the physical location and properties and their genome sequences are known for an increasing number of diatoms, the phylogenetic origin of individual genes, the intracellular targeting of individual proteins and the resulting cellular metabolism are much more elusive, and currently cannot be resolved routinely at a genome wide scale.

Already with the beginning of large-scale nuclear genome sequencing projects for diatoms, the fragmented nature of the diatom genomes became apparent. First evidence came from the discovery of unexpected biochemical pathways (such as the urea cycle identified in *T. pseudonana* [26]). Then, well-known biochemical pathways were found to be targeted to different compartments compared to animals, plants or yeasts (such as the partially mitochondrial glycolysis discovered in *P. tricornutum* [17]). Last but not least, diatom genes revealed surprisingly diverse phylogenetic affinities to other organisms (such as the genes of apparent bacterial [31] or green algal [32] origin identified in *T. pseudonana* and *P. tricornutum*). This intra-genome diversity is the result of a number of possible processes that are summarized in figure 1. Generally, genes that did not evolve in a linear way within the nuclear genome of an organism might have been acquired from organisms that had been taken up as endosymbionts (this process is called endosymbiotic gene transfer), or from non-symbiotic relationships between unrelated organisms such as phagotrophy of bacteria or eukaryotes, or virus- or bacterial infections (such processes are called horizontal or lateral gene transfers). While all these gene transfer processes are conceptually possible and cannot be excluded based on hard evidence, there is quite some debate on the questions regarding: (i) which of the depicted processes are more important than others (i.e. endosymbiotic versus lateral gene transfer [33]), (ii) how often certain processes occurred (i.e. the number and order of endosymbioses (cf. views in [34,35]) or (iii) whether they took place at all (such as the proposed chlamydial endosymbioses/infections in the ancestor of Archaeplastida (cf. views in [36,37]).

Regardless of how any particular gene was acquired, all gene transfers to the nuclear genome have two immediate consequences for the function of the newly acquired gene. One is that its transcription is affected and most likely altered compared with transcription of the gene in its original genome. The other is that, provided that the gene is transcribed, the transcripts are translated in the cytosol, and the newly synthesized protein, due to the lack of a corresponding targeting signal, is not targeted to the compartment in which the gene product was originally active. While the first effect is especially important in the case of bacterial genes compared with genes that are acquired from other eukaryotes, the latter effect generally affects gene products of newly acquired nuclear genes. The evolution/acquisition of a targeting signal is therefore a crucial step in the reclamation of a gene's function. While it has been proposed for a number of organelle derived nuclear genes that they acquired their targeting pre-sequences via a molecular process called exon shuffling [38,39], this does not explain which kind of targeting (or non-targeting) signal is 'fitted' to a certain gene [9]. Consequently, pre-sequence acquisition is also linked to high substitution rates in the parts of the gene encoding the pre-sequence, and sequence conservation within pre-sequences is rather low [38]. Furthermore, genome wide analyses suggest that only a fraction of the products of genes acquired from the precursors of plastids or mitochondria during the establishment of these organelles, are

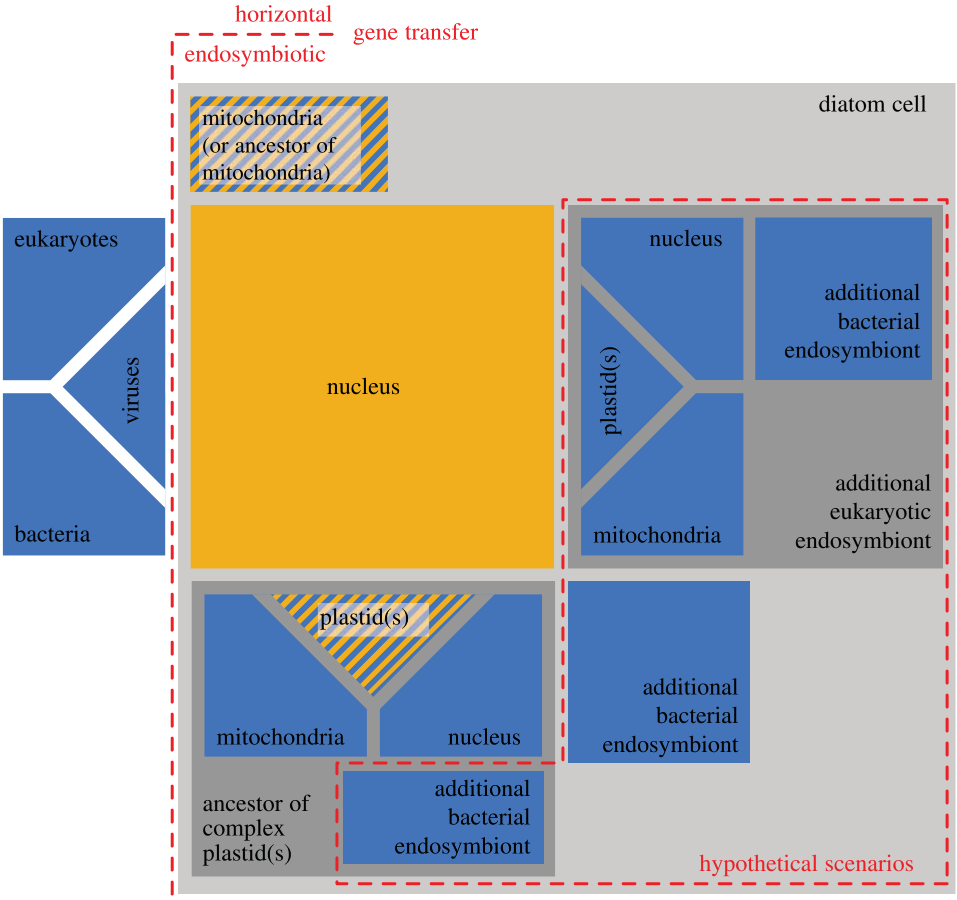
horizontal
gene transfer
endosymbiotic
diatom cell
mitochondria (or ancestor of mitochondria)
eukaryotes
viruses
bacteria
nucleus
nucleus
plastid(s)
mitochondria
additional bacterial endosymbiont
additional eukaryotic endosymbiont
plastid(s)
mitochondria
nucleus
ancestor of complex plastid(s)
additional bacterial endosymbiont
additional bacterial endosymbiont
hypothetical scenarios

Figure 1. Sources of genes and genetic redundancy in diatoms. Blue boxes illustrate the possible sources of genes eventually found in the genomes of diatoms. Mitochondria and plastids (hatched yellow and blue) contain their own genomes and also act as sources for genes transferred to the nucleus (yellow box). Genes are either transferred horizontally or via endosymbiotic gene transfer; however, the exact route of the genes might be more complicated. Various endosymbioses (in addition to the ones that gave rise to plastids and mitochondria) have been proposed on the basis of individual gene phylogenies; these are marked here as hypothetical scenarios.

actually targeted to the organelle from which the gene was acquired [9]. These findings (and many of the findings presented later in this review; §§2–4) are best explained by seeing gene transfer to the nucleus and acquisition of targeting signals as two independent processes.

In addition to gene transfers, another process that may enhance the gene repertoire of diatoms was recently discovered in *F. cylindrus*. This diploid diatom's genome is characterized by a high divergence of haplotypes, which are regulated distinctly from each other in their transcription and therefore may enhance the overall coding as well as the overall regulatory capacity of the genome [28].

All the processes described above lead to (apparent) partial redundancy in the nuclear genomes and hence also in the cellular metabolism of diatoms. For a number of enzymes, this can be seen in the presence of several isoforms. For instance, *P. tricornutum* possesses isoenzymes of fructose-bisphosphate aldolase (FBA) that originate from cyanobacteria, the original host cell, a second host cell, other bacteria and also from duplication of existing genes [40,41]. Genes encoding the Calvin cycle enzymes fructose-1,6-bisphosphatase (FBP) are also present in several copies [42]. Four plastidial FBPases have been identified, and one cytosolic enzyme. Interestingly two of these enzymes contain cysteines that may form regulatory disulfide bridges, while the other two enzymes do not [42].

2. Methods for predicting intracellular enzyme locations in diatom cells

(a) General considerations

Bioinformatic prediction of the intracellular location of a protein has become an important tool in genome annotations and has been summarized in a number of reviews or book chapters [43–46]. In the following, we give a brief summary of the general considerations and then concentrate on the methods/resources available for diatoms. Any prediction result is only meaningful when its predictive power is known. The two important parameters to know are sensitivity (the ability to recognize the true positives) and specificity (the ability to recognize true negatives) [47–49]. There is a trade-off between these two, and it should be kept in mind that all meaningful prediction methods will result in a certain number of false predictions. This can be exemplified by looking at an (absurd) extreme: if a 'prediction' method would simply declare all tested cases as 'positive' this would result in 'complete' recognition of the true positives (highest possible sensitivity). However, this would also result in false prediction of all the negatives (lowest possible specificity). Knowledge of the expected rate of false predictions is therefore more important than 'complete' detection of a certain subset of sequences. The parameters sensitivity, specificity and the Matthews correlation coefficient (MCC), which is an overall measure for the prediction quality [47–49], are in most cases provided in the publications on the respective prediction methods. These parameters, however, may vary depending on the organism for which the method has been developed, compared with the organism with which it is used. Checking back with experimental results obtained with the organism under investigation is therefore advisable. Large-scale proteomic studies, when combined with cell or membrane fractionation (such as [50,51]) often provide data that can be used to evaluate prediction quality; however, also the experimental methods are prone to a certain level of error (the experimental equivalent to 'false positives' are 'contaminations'). Therefore, in the foreseeable future, all -omics scale data on intracellular protein locations, experimental or bioinformatic, will contain a residual proportion of erroneous information.

In addition to these technical limits in protein targeting prediction, there is also a biological effect that might lead to problematic predictions: a significant number of proteins are transported to more than one intracellular compartment. Such proteins are called 'dually targeted' [8]. Generally, a number of mechanisms are known to be responsible here, including the synthesis of distinct pre-proteins (encoded by the same gene), which differ in their targeting signals (so called twin pre-sequences), or the synthesis of a single pre-protein, which is recognized by several protein transport systems (so called ambiguous pre-sequences) [52]. Also diatom genomes contain protein sequences that obtain positive prediction results both in ER-signal peptide predictions and in mitochondrial transit peptide predictions [43], and hence might represent ambiguous targeting signals. Furthermore, dual targeting of proteins to mitochondria and plastids has been shown experimentally to occur for a number of amino acid-tRNA transferases [53], so there are likely more dually targeted sequences to be discovered. As a side note, one interesting hypothesis to explain the maintenance of the cellular metabolism after transfer of a gene from the organelle genome to the nucleus is that a low amount of 'minor mistargeting' might be sufficient to rescue the cell after an organelle-encoded copy of a certain gene has been lost [54]. This proposed mechanism is also in line with the recent discovery that proteins without a recognizable mitochondrial transit peptide may be transported into hydrogenosomes of *Trichomonas vaginalis* [55]. For practical purposes, the findings on dual targeting and the hypotheses on the evolution of targeting signals illustrate that the above-mentioned erroneous information in large-scale datasets, apart from its technical unavoidability, may also be of biological significance.

Most successful protein targeting prediction tools mimic the biological signal recognition in a way that those domains of the pre-protein (N-terminal, internal, or C-terminal), which are relevant for the biological process, are also used as (protein sequence-) input data (although there are also other approaches [44]: for instance, the methods used by Wolf PSORT (https://wolfpsort.hgc.jp/) [56], or Euk-ECC-mPLoc (http://www.csbio.sjtu.edu.cn/bioinf/euk-multi-2/) [57]). For the success of the prediction, it is therefore crucial that the features used by the prediction program are also included in the gene model (the coordinates of translation/transcription start and stop and the exact intron splicing sites) from which the input protein sequence has been derived. Sequence conservation in targeting pre-sequences is often rather low, so homology based gene-modelling algorithms often exclude them. In this case, manual checking of the gene models is necessary, following the steps described in [43]. For the diatoms *P. tricornutum* and *T. pseudonana*, alternative gene catalogues have been re-curated with criteria that specifically emphasize N-terminal completeness of the gene models; these gene catalogues have been published and are freely available for further analyses [58].

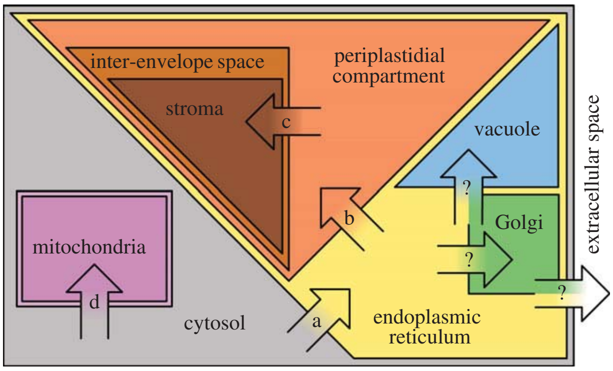

Figure 2. Intracellular transport steps for diatom proteins. Some transport steps occur concurrently, while others are completed sequentially. In this scheme, arrows symbolize the transport steps between the different compartments and do not correspond to the actual structure of the organelles. The transport steps are discussed §§2b–e according to the small letters.

(b) Diatom cell structure and sequential protein targeting

A significant difference between plastids of diatoms and those of higher plants is that diatom plastids are surrounded by four membranes instead of two like all primary plastids. The outermost membrane is continuous with the endoplasmic reticulum (ER). This means that importation of all nuclear-encoded plastid proteins into the plastid stroma and the associated compartments (including the periplastidial compartment and the thylakoid lumen) is preceded by importation of the proteins into the ER [59,60]. Similarly, secreted or vacuole-targeted proteins are also transported through the ER, and many protein targeting pathways comprise a number of sequentially completed steps (figure 2). The targeting signals are usually cleaved after the import reaction [61,62], so that targeting signals for consecutive targeting steps, which are often encoded downstream of preceding targeting signals, become accessible for the recognition at the N-terminus of the protein.

As mentioned in §1, diatoms contain three distinct genomes, the nuclear, the plastidial and the mitochondrial. So far, there have been no reports on proteins that are exported from the plastids or mitochondria; thus for plastid or mitochondria encoded genes it can be safely assumed that the gene product functions within the respective compartment. Nevertheless, proteins encoded by the mitochondrial genome may be targeted to the inner mitochondrial membrane, and plastid genome encoded proteins may be further targeted to the thylakoid membrane or lumen. For such subsequent targeting steps, the organelle-encoded proteins are recognized by the same targeting signals as their nuclear genome encoded counterparts (minus the already cleaved targeting signals from preceding targeting steps). In the following (§2c–e), we discuss the prediction of targeting signals for the steps illustrated in figure 2, assuming that the proteins have are already been targeted into the starting compartment and that the targeting signals are either cleaved off, or were not required in the case of organelle genome encoded proteins.

(c) Endoplasmic reticulum

Proteins are imported into the ER by signal peptide-mediated, co-translational transport through the Sec translocon. Diatom signal peptides share the same features as signal peptides from other eukaryotes (described in [63]), and it has been shown that they are interchangeable between ER proteins and plastid targeted proteins [61]. The prediction program SignalP (http://www.cbs.dtu.dk/services/SignalP/) [46] is suitable for detecting diatom ER-signal peptides, although its performance depends on the version. SignalP 3.0 NN is the most suitable version for diatom sequences, whereas SignalP 4.0 is less sensitive when tested with diatom sequences [58]. Interestingly, although the bulk of ER-targeted diatom proteins possess N-terminal ER-signal peptides, recently also a membrane protein without such an N-terminal signal (the nucleotide transporter NTT5 in *P. tricornutum*) has been shown to be targeted to the ER [64].

Once transported to the ER lumen (arrow marked ‘a’ in figure 2), proteins can be further transported to the Golgi, the vacuole or the extracellular space (arrows marked with ‘?’ in figure 2). Vacuolar location was recently confirmed experimentally for a number of proteins in *P. tricornutum* [65–67]. Interestingly, not all of the identified vacuolar *P. tricornutum* proteins contain an N-terminal signal peptide [65,66]. A di-leucine-based motif found in signal peptide containing vacuolar *P. tricornutum* proteins has experimentally been shown to be important for successful targeting of the protein to the vacuole [65]. Although a number of proteins that are targeted to the Golgi [66] or to the extracellular space [21] are known, the targeting signals that initiate these transport steps in diatoms are unknown, and currently large-scale predictions of protein targeting to these compartments are not routinely possible for diatoms.

(d) Plastid and associated compartments

From the ER lumen, proteins can also be further transported to the plastid via the periplastidial compartment (arrows ‘b’ and ‘c’ in figure 2). Transport to the periplastidial compartment is thought to occur via an ERAD-derived translocon called SELMA [68,69]; further transport to the stroma then occurs via systems similar to the Toc and Tic translocon known from higher plants [70,71]. Plastid targeted proteins as well as proteins that are residing in the periplastidial compartment both have an additional transit peptide-like domain directly downstream of the ER-signal peptide [59,61,72]. These pre-sequences are therefore often referred to as bipartite targeting signals. Plastid proteins additionally show a conserved sequence motif at the signal peptide cleavage site (the ‘ASAFAP’-motif) [61,73], a positive net charge [74], and a conserved sequence motif at the transit peptide cleavage site [62]. Two dedicated prediction methods are available for plastid protein prediction in diatoms (and other organisms with ‘ASAFAP’-motifs in their nucleus encoded plastid targeted proteins): HECTAR (http://webtools.sb-roscoff.fr/) [75] and ASAFind (http://rocaplab.ocean.washington.edu/tools/asafind) [58]. HECTAR applies a hierarchical combination of prediction modules consisting of pre-existing prediction methods for the respective transport steps and a newly developed screen for ‘ASAFAP’-motifs; the results of these modules are then combined via support vector machines. Owing to the large number of pre-existing modules, HECTAR can only be run on the web server and the source code is not available (so it is difficult to customize the method). HECTAR prediction results for diatom plastid proteins are fairly specific (0.94) but not sensitive (0.71) [58]. ASAFind builds on the prediction results of SignalP, which need to be provided to ASAFind in addition to the input protein sequences. In addition to the web service, the Python source code is available (https://bitbucket.org/rocaplab/asafind), so that the method may be customized or further developed. The results of ASAFind are split into ‘high confidence’ (sensitivity of 0.80 at a specificity of 0.99) and ‘low confidence’ (sensitivity of 0.93 at a specificity of 0.82) plastid protein predictions [58].

Neither HECTAR nor ASAFind detect proteins that might be targeted to the periplastidial compartment, and currently, no ready-to-use prediction method for protein targeting to this compartment is available. Nevertheless, candidate sequences can be investigated manually. Although the transit peptide-like domains of bipartite pre-sequences for plastid or periplastidial compartment targeting in diatoms are hard to detect, they can be identified following the methods described in [43]. For a complementary approach, a large dataset of experimentally tested or candidate proteins suggested to be targeted to the periplastidial compartment of *P. tricornutum* has been published [72] and can be screened for proteins of interest.

(e) Mitochondrial transit peptides

Mitochondrial proteins in diatoms are imported into the organelle in a similar way as in other eukaryotes, based on similar N-terminal mitochondrial targeting peptides (arrow ‘d’ in figure 2). The program TargetP (http://www.cbs.dtu.dk/services/TargetP/) [76] is a suitable tool for the prediction of these targeting peptides. TargetP optionally also detects chloroplast transit peptides for importation into primary plastids (which do not exist in diatoms); therefore while working with diatom sequences, checking of the ‘non-plant’ organism group box is required. TargetP, in parallel to identifying mitochondrial transit peptides also identifies signal peptides (using the previously mentioned SignalP), so when used in preliminary screens it also identifies those diatom proteins that might be further targeted to the plastid or associated compartments via the ER.

3. Metabolic peculiarities in diatom cells

The strong structural and phylogenetic distance between diatoms and green algae/land plants based on the different endosymbiotic history (see §1) is reflected by a number of differences regarding the location and organization of metabolic pathways. One essential difference is the storage of carbohydrates: while green algae and land plants store the α-glucan starch within the plastids, diatoms accumulate the ß-glucan chrysolaminarin in cytosolic vacuoles, thus not in the compartment where photosynthesis takes place. There is only scarce information available on the localization, synthesis and storage processes of chrysolaminarin. The structure of chrysolaminarin is fundamentally based on a β-1,3-linked glucan backbone, which is infrequently branched with mainly β-1,6-linkages [65]. The synthetic

pathway	*Chlamydomonas reinhardtii*	*Phaeodactylum tricornutum*
Calvin cycle	plastid	plastid
oxidative pentose phosphate pathway	plastid, cytosol	cytosol
glycolysis first half	plastid	cytosol, plastid
glycolysis second half	cytosol	cytosol, plastid, mitochondria
Entner–Doudoroff pathway	—	mitochondria
urea cycle	—	mitochondria
carbohydrate storage	plastid	vacuole

Figure 3. Metabolic peculiarities in the diatom *Phaeodactylum tricornutum* compared with the green alga *Chlamydomonas reinhardtii*. Intracellular locations of pathways discussed in the text (§3) are indicated '—' means that there is no evidence of this pathway in *C. reinhardtii*. Drawings: Marc Halder.

pathway of chrysolaminarin is still unclear. Based on enzyme activity assays of *Cylindrotheca cryptica*, UDP-glucose likely serves as the substrate for chrysolaminaran synthesis [77], while a membrane bound glucan synthase is presumably involved in the transport of the glucan chain into the vacuole (Huang W, Río Bártulos C, Kroth PG 2017, unpublished data). Accordingly, vacuolar transglycosylases have been identified recently in *P. tricornutum* that likely are involved in the modification and addition of chrysolaminarin side chains [78].

The lack of the possibility of intermediate storage of carbohydrates in the plastids combined with the presence of four membranes surrounding the diatom plastids requires additional transport capacities to transport all carbohydrates from the plastids to the cytosol and the vacuoles. A recent report accordingly demonstrated a higher number of plastidial triose-phosphate translocators in *P. tricornutum*, allowing the passage of triose phosphates across all four plastid membranes [79]. By contrast, nucleotides in diatoms require a net import into the plastids. Furthermore, some metabolic pathways such as the oxidative pentose phosphate pathway (OPP) [42,80] delivering riboses for nucleotide synthesis (see below), and the nucleotide synthesis pathways [81], are missing in diatom plastids and have been re-allocated into the cytosol, requiring the permanent importation of nucleotides across the plastid membranes [81]. Similar to the triose phosphate translocators, diatoms here possess a couple of nucleotide translocators (NTTs) [64,81]. Interestingly, these translocators possess different pre-sequences for targeting into the various plastid membranes [64,81] and might have been recruited from different sources. While NTT1 of *P. tricornutum* and *T. pseudonana* are homologous to NTTs found in higher plants, NTT2 of the two diatoms and NTT5 of *P. tricorntum* are more similar to NTTs of parasitic intracellular bacteria [81].

As already mentioned in §1, sequencing of the *T. pseudonana* genome initially revealed that diatoms possess genes for enzymes that are involved in an ornithine–urea cycle [26]. This pathway allows metazoans to discard excess nitrogen (a by-product of the degradation of proteins in cellular catabolism) in the form of urea. Photosynthetic organisms, however, have a constant demand for nitrogen for their anabolism and the ornithine–urea cycle is absent in green algae and plants [26,82]. RNAi-mediated knock-down of the mitochondrial carbamoyl phosphate synthase by Allen *et al*. [84] led to delayed recovery from nitrogen starvation compared with wild type cells; the authors therefore concluded that, in contrast with its function in metazoans, the urea cycle in diatoms 'serves as a distribution and repackaging hub for inorganic carbon and nitrogen'. In line with these results, it is noteworthy that diatoms also possess a urease which is normally not found in organisms containing a urea cycle. Knocking out the urease in *P. tricornutum* results in an increase of intracellular urea, arginine and ornithine [83].

Photosynthetic carbon fixation in plants and in algae is based on the Calvin–Benson–Bassham cycle, or simply the Calvin cycle or reductive pentose phosphate pathway. Plastids from land plants and green algae (including the model green alga *Chlamydomonas reinhardtii*) additionally contain an OPP (figure 3). This ubiquitous process is responsible for the generation of NADPH and pentose phosphates in the dark, for biosynthesis of nucleotides, fatty acids and amino acids, via decarboxylation of glucose-6-phosphate. In contrast with land plants, in diatoms there is apparently no complete OPP within the plastids, suggesting that diatom plastids lack this pathway [82]. Two essential enzymes, glucose-6-phosphate dehydrogenase (G6PDH) and 6-phosphogluconate dehydrogenase (6PGDH) were found to be cytosolic enzymes, further indicating that the complete OPP is only functional in the cytosol [17,42].

One very striking finding is that in diatoms, the genes of the enzymes of the second half of the glycolysis reaction chain (from glyceraldehyde-3-phosphate to pyruvate) have been duplicated, and the respective isoenzymes now are imported into the mitochondria [17]. This essentially allows the organisms to generate pyruvate within the mitochondria, without importation from the cytosol. The reason for this unusual distribution might be the presence of a functional Entner–Doudoroff pathway in the mitochondria of diatoms [84], an ancient alternative for the glycolytic Embden–Meyerhof–Parnas pathway. The involved genes encode 6-phosphogluconate dehydratase (EDD) and 2-keto-3-deoxyphosphogluconate aldolase (EDA), producing pyruvate as well as glyceraldehyde-3-phosphate from glucose, which may substitute the first steps of glycolysis within the mitochondria. This might also explain why the OPP has been removed from the plastids, as it generates 6-phosphogluconate, which is the substrate for the Entner–Doudoroff pathway. The utilization and potential advantage of this pathway within the mitochondria so far is unknown, but degrading 6-phosphogluconate via this way in the mitochondria might result in different NADH/ATP ratios when compared with cytosolic glycolysis. In *C. reinhardtii*, the first half of glycolysis takes place in the plastid, while the second half of glycolysis takes place exclusively in the cytosol [85] (figure 3).

Finally, as mentioned in §2b, diatoms possess an additional compartment compared with plants: the periplastidial space between the two innermost and two outermost envelope membranes of the diatom plastids. This compartment represents the former cytosol of the secondary endosymbiont. Apparently, all visible functional structures have been removed from this compartment, although some reports propose the presence of vesicular structures in this compartment [86,87]. Interestingly, there is evidence that among the various proteins targeted to this compartment (see [72] for an overview), metabolic enzymes are also found, such as a thioredoxin [20] and a 6PGDH [42]. The role of this extra compartment and why it has been retained in diatoms is still unknown. One manuscript proposes that this compartment could be acidic and thus may function as a CO_2 trap [88]; however, so far there is no experimental proof for this hypothesis, and the finding of potentially active metabolic enzymes in the periplastidial compartment makes this assumption rather unlikely.

4. Genome-enabled approaches to study diatom metabolism

Sequence information from genomic or RNAseq data can be a great help for better understanding metabolic processes. On the one hand, it allows the elucidation of the primary structure of metabolic enzymes, whereas on the other hand, these data enable studies on the function of a gene or a gene product by disruption or modification of the gene and subsequent characterization of the obtained mutants. Furthermore, the respective gene of interest can be transferred, expressed, purified and studied *in vitro*. For a number of organisms, simple random mutagenesis, including ultraviolet radiation or chemical treatment that induces mutations, proves to be an easy technique. However, as most random mutations here will instead lead to the inactivation of one allele, this method is limited to those algal species that have at least one stable haploid stage and/or that are offering crossbreeding opportunities, which is required for the creation of homozygous mutants. For diatoms, which are diploid organisms and which rarely, if at all, cross in the laboratory [89], a heterozygous mutation in one allele (while the second allele is still wild type) may not necessarily lead to a clear phenotype. There is even evidence that a single inactivated allele may result in upregulation of the second allele abolishing a phenotype [90].

An alternative to random mutagenesis is the targeted genetic modification of algae, including the introduction/overexpresssion of new genes, or the knock-down or knockout of specific genes. Such approaches require the transfer and integration of foreign DNA via species-specific transformation protocols [91]. Significant progress in the development of gene transfer systems for diatoms has been made within the past 25 years. Such techniques allow the modification of algae, either to obtain strains that produce certain compounds of commercial interest (referred to as 'metabolic engineering'), or to gain information about cellular, physiological or biochemical mechanisms by switching off, downregulating or overexpressing existing or foreign genes (referred to as 'reverse genetics' or 'genome-enabled experimental biology'). Up to now, successful genetic manipulation of more than 40 different eukaryotic microalgae species has been reported, including model species such as the green algae *C. reinhardtii* [92], *Dunaliella salina* [93], *Chlorella vulgaris* [94] and *Haematococcus pluvialis* [95], the eustigmatophyte *Nannochloropsis* sp. [96], as well as the diatoms *T. pseudonana* [97], *P. tricornutum* [98] and *Cylindrotheca fusiformis* [99].

Targeting of transgenic DNA specifically into the respective organelles is mostly impossible; however, organelle-specific expression can be achieved by simply using promoters that are active in the respective compartments or by using organelle-specific selection markers. Most transformation methods rely on the fact that, as soon as DNA can be successfully imported into cells there is a chance that it can be integrated into the nuclear genome. A number of genetic transformation techniques have been tested for diatoms. While biolistics has long been the method of choice [91], recently successful genetic transformations have been reported based on electroporation [100,101] as well as on incubation with conjugating *Escherichia coli* [102].

For efficient expression of nuclear transgenes, strong promoters are required. For nuclear expression in diatoms these often are based on light-driven promoters of light harvesting antennae (FCP proteins) [91], and nitrate-inducible promoters from the diatom nitrate reductase [103]. In addition to the overexpression of genes, silencing (knock-down) or disruption (knockout) of key genes has been demonstrated to be an important tool to modify cell lines. Silencing approaches take advantage of the RNAi mechanism in eukaryotic cells that results in the degradation of double-stranded mRNA [90,104]. By expressing artificial genes that encode inverted repeats of a gene or its antisense sequence, such double-stranded mRNAs are generated in the cells, which then result in a degradation of the respective endogenous mRNA and a decreased amount of the respective gene product.

Knocking out genes in diatoms is especially difficult as the haploid stages (gametes) are very short-lived. Therefore, different genome editing approaches that target both alleles such as meganucleases [105], TALENs (transcription activator-like effector nucleases) [105,106] or CRISPR (clustered regularly interspaced short palindromic repeats) [107,108] have recently become very popular approaches and can be used in diatoms. All these protocols are based on nucleases that bind specifically to distinct DNA sequences, allowing the targeting of single genes within a full genome, and inducing modification of the target gene such as by insertions or mutations. Such genome editing approaches are extremely helpful for targeting all copies of a gene in a genome, for instance in diatoms, which are diploid organisms for which no crossing options have so far been available. Some approaches such as CRISPR/Cas9 even allow targeting of several genes at the same time. Nevertheless, different measures have to be taken to avoid off-target activity of the nuclease; here mostly the transient expression or temporary presence of the nucleases in the cells are the preferred ways to ensure targeted modification of the cells.

Meanwhile genetic modification has been successfully applied to modify the metabolism of diatoms in a number of cases. One very early example of metabolic engineering is the trophic conversion of *P. tricornutum* by expressing a human glucose translocator targeted to the plasma membrane, allowing the cells to grow on glucose [109]. Other examples are the induced change of the lipid composition of *P. tricornutum* upon expression of a hetereologous Δ5-elongase [110] or the increase in intracellular lipid amounts obtained by either overexpression of a diacylglycerol acyltransferase [111], or by silencing of the UDP-glucose pyrophosphorylase, which shifts the metabolism from carbohydrate to lipid accumulation [112]. Examples of genome-enabled experimental diatom biology are the silencing of the *P. tricornutum* pyruvate phosphate dikinase (PPDK), which did not result in reduced photosynthetic capacity of the transformed cells, implying that the PPDK in *P. tricornutum* does not function in the immediate fixation of inorganic carbon [113]. Further examples are the RNAi-mediated knock-down of the *P. tricornutum* carbamoyl phosphate synthase [82] or the TALEN mediated knockout of the urease [83] of the same diatom.

5. Conclusion

Diatoms show a very peculiar repertoire and intracellular distribution of metabolic enzymes, which is a result of their specific evolutionary origin, combining at least two bacteria with at least two eukaryotic organisms. The rearrangements of the intracellular distribution of metabolic pathways presented in this review also underline that the transfer of genes to the nucleus and the acquisition of targeting signals by these genes are independent processes that, taken together, may sometimes lead to surprising results.

Data accessibility. This article has no additional data.

Competing interests. We declare we have no competing interests.

Funding. This work has been supported by the University of Konstanz and by grants of the German Research Foundation DFG (grant KR 1661/8-2 to P.G.K.), and the Gordon and Betty Moore Foundation GBMF 4966 (grant DiaEdit to P.G.K.).

Acknowledgement. We thank Marc Halder (Universität Konstanz) for the preparation of the drawings in figure 3.

References

1. Fischer WW, Hemp J, Johnson JE. 2016 Evolution of oxygenic photosynthesis. *Annu. Rev. Earth Planet. Sci.* **44**, 647–683. (doi:10.1146/annurev-earth-060313-054810)
2. Holland HD. 2006 The oxygenation of the atmosphere and oceans. *Phil. Trans. R. Soc. B* **361**, 903–915. (doi:10.1098/rstb.2006.1838)
3. Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O, Taylor FJ. 2004 The evolution of modern eukaryotic phytoplankton. *Science* **305**, 354–360. (doi:10.1126/science.1095964)
4. Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. 1998 Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237–240. (doi:10.1126/science.281.5374.237)
5. Archibald JM. 2009 The puzzle of plastid evolution. *Curr. Biol.* **19**, R81–R88. (doi:10.1016/j.cub.2008.11.067)
6. McFadden GI. 2014 Origin and evolution of plastids and photosynthesis in eukaryotes. *Cold Spring Harb. Perspect. Biol.* **6**, a016105. (doi:10.1101/cshperspect.a016105)
7. Cavalier-Smith T. 2000 Membrane heredity and early chloroplast evolution. *Trends Plant Sci.* **5**, 174–182. (doi:10.1016/S1360-1385(00)01598-3)
8. Millar AH, Whelan J, Small I. 2006 Recent surprises in protein targeting to mitochondria and plastids. *Curr. Opin. Plant Biol.* **9**, 610–615. (doi:10.1016/j.pbi.2006.09.002)
9. Martin W, Herrmann RG. 1998 Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiol.* **118**, 9–17. (doi:10.1104/pp.118.1.9)
10. Adl SM *et al.* 2012 The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* **59**, 429–493. (doi:10.1111/j.1550-7408.2012.00644.x)
11. Rodriguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Loffelhardt W, Bohnert HJ, Philippe H, Lang BF. 2005 Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr. Biol.* **15**, 1325–1330. (doi:10.1016/j.cub.2005.06.040)
12. Moreira D, Le Guyader H, Philippe H. 2000 The origin of red algae and the evolution of chloroplasts. *Nature* **405**, 69–72. (doi:10.1038/35011054)
13. Archibald JM. 2015 Endosymbiosis and eukaryotic cell evolution. *Curr. Biol.* **25**, R911–R921. (doi:10.1016/j.cub.2015.07.055)
14. Gould SB, Waller RF, McFadden GI. 2008 Plastid evolution. *Annu. Rev. Plant Biol.* **59**, 491–517. (doi:10.1146/annurev.arplant.59.032607.092915)
15. Falkowski PG, Barber RT, Smetacek VV. 1998 Biogeochemical controls and feedbacks on ocean primary production. *Science* **281**, 200–207. (doi:10.1126/science.281.5374.200)
16. Granum E, Raven JA, Leegood RC. 2005 How do marine diatoms fix 10 billion tonnes of inorganic carbon per year. *Can. J. Bot.* **83**, 898–908. (doi:10.1139/b05-077)
17. Kroth PG *et al.* 2008 A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS ONE* **3**, e1426. (doi:10.1371/journal.pone.0001426)
18. Matsuda Y, Hopkinson BM, Nakajima K, Dupont CL, Tsuji Y. 2017 Mechanisms of carbon dioxide acquisition and CO_2 sensing in marine diatoms: a gateway to carbon metabolism. *Phil. Trans. R. Soc. B* **372**, 20160403. (doi:10.1098/rstb.2016.0403)
19. Raven JA, Giordano M. 2017 Acquisition and metabolism of carbon in the Ochrophyta other than diatoms. *Phil. Trans. R. Soc. B* **372**, 20160400. (doi:10.1098/rstb.2016.0400)
20. Weber T, Gruber A, Kroth PG. 2009 The presence and localization of thioredoxins in diatoms, unicellular algae of secondary endosymbiotic origin. *Mol. Plant* **2**, 468–477. (doi:10.1093/mp/ssp010)
21. Bruckner CG, Rehm C, Grossart HP, Kroth PG. 2011 Growth and release of extracellular organic compounds by benthic diatoms depend on interactions with bacteria. *Environ. Microbiol.* **13**, 1052–1063. (doi:10.1111/j.1462-2920.2010.02411.x)
22. Heintzelman MB. 2006 Cellular and molecular mechanics of gliding locomotion in eukaryotes. *Int. Rev. Cytol.* **251**, 79–129. (doi:10.1016/S0074-7696(06)51003-4)
23. Chiovitti A, Molino P, Crawford SA, Teng R, Spurck T, Wetherbee R. 2004 The glucans extracted with warm water from diatoms are mainly derived from intracellular chrysolaminaran and not extracellular polysaccharides. *Eur. J. Phycol.* **39**, 117–128. (doi:10.1080/0967026042000201885)

24. Bahulikar RA, Kroth PG. 2007 Localization of EPS components secreted by freshwater diatoms using differential staining with fluorophore-conjugated lectins and other fluorochromes. *Eur. J. Phycol.* **42**, 199 – 208. (doi:10.1080/09670260701289779)

25. Buhmann MT, Schulze B, Forderer A, Schleheck D, Kroth PG. 2016 Bacteria may induce the secretion of mucin-like proteins by the diatom *Phaeodactylum tricornutum*. *J. Phycol.* **52**, 463 – 474. (doi:10.1111/jpy.12409)

26. Armbrust EV *et al.* 2004 The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**, 79 – 86. (doi:10.1126/science.1101156)

27. Lommer M *et al.* 2012 Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biol.* **13**, R66. (doi:10.1186/gb-2012-13-7-r66)

28. Mock T *et al.* 2017 Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature* **541**, 536 – 540. (doi:10.1038/nature20803)

29. Tanaka T *et al.* 2015 Oil accumulation by the oleaginous diatom *Fistulifera solaris* as revealed by the genome and transcriptome. *Plant Cell* **27**, 162 – 176. (doi:10.1105/tpc.114.135194)

30. Galachyants YP *et al.* 2015 Sequencing of the complete genome of an araphid pennate diatom *Synedra acus* subsp. *radians* from Lake Baikal. *Dokl. Biochem. Biophys.* **461**, 84 – 88. (doi:10.1134/S1607672915020064)

31. Bowler C *et al.* 2008 The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239 – 244. (doi:10.1038/nature07410)

32. Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D. 2009 Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* **324**, 1724 – 1726. (doi:10.1126/science.1172983)

33. Ku C, Martin WF. 2016 A natural barrier to lateral gene transfer from prokaryotes to eukaryotes revealed from genomes: the 70% rule. *BMC Biol.* **14**, 89. (doi:10.1186/s12915-016-0315-9)

34. 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**, 347 – 366. (doi:10.1111/j.1550-7408.1999.tb04614.x)

35. Baurain D *et al.* 2010 Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Mol. Biol. Evol.* **27**, 1698 – 1709. (doi:10.1093/molbev/msq059)

36. Becker B, Hoef-Emden K, Melkonian M. 2008 Chlamydial genes shed light on the evolution of photoautotrophic eukaryotes. *BMC Evol. Biol.* **8**, 203. (doi:10.1186/1471-2148-8-203)

37. Martin WR, Kloesges T, Thiergart T, Woehle C, Gould S, Dagan T. 2012 Modern endosymbiotic theory: getting lateral gene transfer into the equation. *Endocytobiosis Cell Res.* **23**, 1 – 5.

38. Long M, de Souza SJ, Rosenberg C, Gilbert W. 1996 Exon shuffling and the origin of the mitochondrial targeting function in plant cytochrome c1 precursor. *Proc. Natl Acad. Sci. USA* **93**, 7727 – 7731. (doi:10.1073/pnas.93.15.7727)

39. Kilian O, Kroth PG. 2004 Presequence acquisition during secondary endocytobiosis and the possible role of introns. *J. Mol. Evol.* **58**, 712 – 721. (doi:10.1007/s00239-004-2593-z)

40. 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**, 389 – 400. (doi:10.1007/s00294-005-0033-2)

41. Allen AE, Moustafa A, Montsant A, Eckert A, Kroth PG, Bowler C. 2012 Evolution and functional diversification of fructose bisphosphate aldolase genes in photosynthetic marine diatoms. *Mol. Biol. Evol.* **29**, 367 – 379. (doi:10.1093/molbev/msr223)

42. Gruber A, Weber T, Bartulos CR, Vugrinec S, Kroth PG. 2009 Intracellular distribution of the reductive and oxidative pentose phosphate pathways in two diatoms. *J. Basic Microbiol.* **49**, 58 – 72. (doi:10.1002/jobm.200800339)

43. Gruber A, Kroth PG. 2014 Deducing intracellular distributions of metabolic pathways from genomic data. In *Plant metabolism—methods and protocols* (ed. G Sriram), pp. 187 – 211. New York, NY: Humana Press.

44. Nakai K, Horton P. 2007 Computational prediction of subcellular localization. In *Protein targeting protocols* (ed. M van der Giezen), pp. 429 – 465. Totowa, NJ: Humana Press.

45. Nielsen H, Brunak S, von Heijne G. 1999 Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Eng.* **12**, 3 – 9. (doi:10.1093/protein/12.1.3)

46. Emanuelsson O, Brunak S, von HG, Nielsen H. 2007 Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2**, 953 – 971. (doi:10.1038/nprot.2007.131)

47. Brown CD, Davis HT. 2006 Receiver operating characteristics curves and related decision measures: a tutorial. *Chemometr. Intell. Lab. Syst.* **80**, 24 – 38. (doi:10.1016/j.chemolab.2005.05.004)

48. Baldi P, Brunak S, Chauvin Y, Andersen CA, Nielsen H. 2000 Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics* **16**, 412 – 424. (doi:10.1093/bioinformatics/16.5.412)

49. Fawcett T. 2006 An introduction to ROC analysis. *Pattern Recognit. Lett.* **27**, 861 – 874. (doi:10.1016/j.patrec.2005.10.010)

50. Lepetit B, Volke D, Szabo M, Hoffmann R, Garab G, Wilhelm C, Goss R. 2007 Spectroscopic and molecular characterization of the oligomeric antenna of the diatom *Phaeodactylum tricornutum*. *Biochemistry* **46**, 9813 – 9822. (doi:10.1021/bi7008344)

51. Grouneva I, Rokka A, Aro EM. 2011 The thylakoid membrane proteome of two marine diatoms outlines both diatom-specific and species-specific features of the photosynthetic machinery. *J. Proteome Res.* **10**, 5338 – 5353. (doi:10.1021/pr200600f)

52. Peeters N, Small I. 2001 Dual targeting to mitochondria and chloroplasts. *Biochim. Biophys. Acta* **1541**, 54 – 63. (doi:10.1016/S0167-4889(01)00146-X)

53. Gile GH, Moog D, Slamovits CH, Maier UG, Archibald JM. 2015 Dual organellar targeting of aminoacyl-tRNA synthetases in diatoms and cryptophytes. *Genome Biol. Evol.* **7**, 1728 – 1742. (doi:10.1093/gbe/evv095)

54. Martin W. 2010 Evolutionary origins of metabolic compartmentalization in eukaryotes. *Phil. Trans. R. Soc. B* **365**, 847 – 855. (doi:10.1098/rstb.2009.0252)

55. Garg S, Stolting J, Zimorski V, Rada P, Tachezy J, Martin WF, Gould SB. 2015 Conservation of transit peptide-independent protein import into the mitochondrial and hydrogenosomal matrix. *Genome Biol. Evol.* **7**, 2716 – 2726. (doi:10.1093/gbe/evv175)

56. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. 2007 WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* **35**, W585 – W587. (doi:10.1093/nar/gkm259)

57. Chou KC, Shen HB. 2010 A new method for predicting the subcellular localization of eukaryotic proteins with both single and multiple sites: Euk-mPLoc 2.0. *PLoS ONE* **5**, e9931. (doi:10.1371/journal.pone.0009931)

58. Gruber A, Rocap G, Kroth PG, Armbrust EV, Mock T. 2015 Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant J.* **81**, 519 – 528. (doi:10.1111/tpj.12734)

59. Kroth PG. 2002 Protein transport into secondary plastids and the evolution of primary and secondary plastids. *Int. Rev. Cytol.* **221**, 191 – 255. (doi:10.1016/S0074-7696(02)21013-X)

60. Patron NJ, Waller RF. 2007 Transit peptide diversity and divergence: a global analysis of plastid targeting signals. *Bioessays* **29**, 1048 – 1058. (doi:10.1002/bies.20638)

61. Kilian O, Kroth PG. 2005 Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. *Plant J.* **41**, 175 – 183. (doi:10.1111/j.1365-313X.2004.02294.x)

62. Huesgen PF, Alami M, Lange PF, Foster LJ, Schroder WP, Overall CM, Green BR. 2013 Proteomic amino-termini profiling reveals targeting information for protein import into complex plastids. *PLoS ONE* **8**, e74483. (doi:10.1371/journal.pone.0074483)

63. Nielsen H, Engelbrecht J, Brunak S, von Heijne G. 1997 Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng. Des. Sel.* **10**, 1 – 6. (doi:10.1093/protein/10.1.1)

64. Chu L, Gruber A, Ast M, Schmitz-Esser S, Altensell J, Neuhaus HE, Kroth PG, Haferkamp I. 2017 Shuttling of (deoxy-) purine nucleotides between compartments of the diatom *Phaeodactylum tricornutum*. *New Phytol.* **213**, 193 – 205. (doi:10.1111/nph.14126)

65. Schreiber V *et al.* 2017 The central vacuole of the diatom *Phaeodactylum tricornutum*: identification of new vacuolar membrane proteins and of a

functional di-leucine-based targeting motif. *Protist* **168**, 271–282. (doi:10.1016/j.protis.2017.03.001)

66. Liu X, Hempel F, Stork S, Bolte K, Moog D, Heimerl T, Maier UG, Zauner S. 2016 Addressing various compartments of the diatom model organism *Phaeodactylum tricornutum* via sub-cellular marker proteins. *Algal Res.* **20**, 249–257. (doi:10.1016/j.algal.2016.10.018)
67. Huang W, Rio Bartulos C, Kroth PG. 2016 Diatom vacuolar 1,6-beta-transglycosylases can functionally complement the respective yeast mutants. *J. Eukaryot. Microbiol.* **63**, 536–546. (doi:10.1111/jeu.12298)
68. Sommer MS, Gould SB, Lehmann P, Gruber A, Przyborski JM, Maier UG. 2007 Der1-mediated preprotein import into the periplastid compartment of chromalveolates? *Mol. Biol. Evol.* **24**, 918–928. (doi:10.1093/molbev/msm008)
69. Hempel F, Bullmann L, Lau J, Zauner S, Maier UG. 2009 ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. *Mol. Biol. Evol.* **26**, 1781–1790. (doi:10.1093/molbev/msp079)
70. Bullmann L, Haarmann R, Mirus O, Bredemeier R, Hempel F, Maier UG, Schleiff E. 2009 Filling the gap: the evolutionary conserved Omp85 in plastids of chromalveolates. *J. Biol. Chem.* **285**, 6848–6856. (doi:10.1074/jbc.M109.074807)
71. Maier UG, Zauner S, Hempel F. 2015 Protein import into complex plastids: cellular organization of higher complexity. *Eur. J. Cell Biol.* **94**, 340–348. (doi:10.1016/j.ejcb.2015.05.008)
72. Moog D, Stork S, Zauner S, Maier UG. 2011 In silico and in vivo investigations of proteins of a minimized eukaryotic cytoplasm. *Genome Biol. Evol.* **3**, 375–382. (doi:10.1093/gbe/evr031)
73. Gruber A, Vugrinec S, Hempel F, Gould SB, Maier UG, Kroth PG. 2007 Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. *Plant Mol. Biol.* **64**, 519–530. (doi:10.1007/s11103-007-9171-x)
74. Felsner G, Sommer MS, Maier UG. 2010 The physical and functional borders of transit peptide-like sequences in secondary endosymbionts. *BMC Plant Biol.* **10**, 223. (doi:10.1186/1471-2229-10-223)
75. Gschloessl B, Guermeur Y, Cock JM. 2008 HECTAR: a method to predict subcellular targeting in heterokonts. *BMC Bioinformatics* **9**, 393. (doi:10.1186/1471-2105-9-393)
76. Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000 Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–1016. (doi:10.1006/jmbi.2000.3903)
77. Roessler PG. 1987 UDP-glucose pyrophosphorylase activity in the diatom *Cyclotella cryptica*. Pathway of chrysolaminarin synthesis. *J. Phycol.* **23**, 494–498. (doi:10.1111/j.1529-8817.1987.tb02536.x)
78. Chan CX *et al.* 2012 Porphyra (Bangiophyceae) transcriptomes provide insights into red algal development and metabolism. *J. Phycol.* **48**, 1328–1342. (doi:10.1111/j.1529-8817.2012.01229.x)
79. Moog D, Rensing SA, Archibald JM, Maier UG, Ullrich KK. 2015 Localization and evolution of putative triose phosphate translocators in the diatom *Phaeodactylum tricornutum*. *Genome Biol. Evol.* **7**, 2955–2969. (doi:10.1093/gbe/evv190)
80. Michels AK, Wedel N, Kroth PG. 2005 Diatom plastids possess a phosphoribulokinase with an altered regulation and no oxidative pentose phosphate pathway. *Plant Physiol.* **137**, 911–920. (doi:10.1104/pp.104.055285)
81. Ast M, Gruber A, Schmitz-Esser S, Neuhaus HE, Kroth PG, Horn M, Haferkamp I. 2009 Diatom plastids depend on nucleotide import from the cytosol. *Proc. Natl Acad. Sci. USA* **106**, 3621–3626. (doi:10.1073/pnas.0808862106)
82. Allen AE *et al.* 2011 Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* **473**, 203–207. (doi:10.1038/nature10074)
83. Weyman PD, Beeri K, Lefebvre SC, Rivera J, McCarthy JK, Heuberger AL, Peers G, Allen AE, Dupont CL. 2015 Inactivation of *Phaeodactylum tricornutum* urease gene using transcription activator-like effector nuclease-based targeted mutagenesis. *Plant Biotechnol. J.* **13**, 460–470. (doi:10.1111/pbi.12254)
84. Fabris M, Matthijs M, Rombauts S, Vyverman W, Goossens A, Baart GJE. 2012 The metabolic blueprint of *Phaeodactylum tricornutum* reveals a eukaryotic Entner–Doudoroff glycolytic pathway. *Plant J.* **70**, 1004–1014. (doi:10.1111/j.1365-313X.2012.04941.x)
85. Johnson X, Alric J. 2013 Central carbon metabolism and electron transport in *Chlamydomonas reinhardtii*: metabolic constraints for carbon partitioning between oil and starch. *Eukaryot. Cell* **12**, 776–793. (doi:10.1128/EC.00318-12)
86. Gibbs SP. 1981 The chloroplast endoplasmic reticulum: structure, function, and evolutionary significance. *Int. Rev. Cytol.* **72**, 49–99. (doi:10.1016/S0074-7696(08)61194-8)
87. Flori S, Jouneau PH, Finazzi G, Marechal E, Falconet D. 2016 Ultrastructure of the periplastidial compartment of the diatom *Phaeodactylum tricornutum*. *Protist* **167**, 254–267. (doi:10.1016/j.protis.2016.04.001)
88. Lee RE, Kugrens P. 1998 Hypothesis: the ecological advantage of chloroplast ER—the ability to outcompete at low dissolved CO_2 concentrations. *Protist* **149**, 341–345. (doi:10.1016/S1434-4610(98)70040-9)
89. Chepurnov VA, Mann DG, von Dassow P, Vanormelingen P, Gillard J, Inze D, Sabbe K, Vyverman W. 2008 In search of new tractable diatoms for experimental biology. *Bioessays* **30**, 692–702. (doi:10.1002/bies.20773)
90. Huysman MJ *et al.* 2013 AUREOCHROME1a-mediated induction of the diatom-specific cyclin dsCYC2 controls the onset of cell division in diatoms (*Phaeodactylum tricornutum*). *Plant Cell* **25**, 215–228. (doi:10.1105/tpc.112.106377)
91. Zaslavskaia LA, Lippmeier JC, Kroth PG, Grossman AR, Apt KE. 2000 Transformation of the diatom *Phaeodactylum tricornutum* (Bacillariophyceae) with a variety of selectable marker and reporter genes. *J. Phycol.* **36**, 379–386. (doi:10.1046/j.1529-8817.2000.99164.x)
92. Kindle KL. 1990 High-frequency nuclear transformation of *Chlamydomonas reinhardtii*. *Proc. Natl Acad. Sci. USA* **87**, 1228–1232. (doi:10.1073/pnas.87.3.1228)
93. Barzegari A, Hejazi MA, Hosseinzadeh N, Eslami S, Mehdizadeh AE, Hejazi MS. 2009 Dunaliella as an attractive candidate for molecular farming. *Mol. Biol. Rep.* **37**, 3427–30. (doi:10.1007/s11033-009-9933-4)
94. Dawson HN, Burlingame R, Cannons AC. 1997 Stable transformation of *Chlorella*: rescue of nitrate reductase-deficient mutants with the nitrate reductase gene. *Curr. Microbiol.* **35**, 356–362. (doi:10.1007/s002849900268)
95. Steinbrenner J, Sandmann G. 2006 Transformation of the green alga *Haematococcus pluvialis* with a phytoene desaturase for accelerated astaxanthin biosynthesis. *Appl. Environ. Microbiol.* **72**, 7477–7484. (doi:10.1128/AEM.01461-06)
96. Kilian O, Benemann CS, Niyogi KK, Vick B. 2011 High-efficiency homologous recombination in the oil-producing alga *Nannochloropsis* sp. *Proc. Natl Acad. Sci. USA* **108**, 21 265–21 269. (doi:10.1073/pnas.1105861108)
97. Poulsen N, Chesley PM, Kröger N. 2006 Molecular genetic manipulation of the diatom *Thalassiosira pseudonana* (Bacillariophyceae). *J. Phycol.* **42**, 1059–1065. (doi:10.1111/j.1529-8817.2006.00269.x)
98. Apt KE, Kroth-Pancic PG, Grossman AR. 1996 Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. *Mol. Gen. Genet.* **252**, 572–579.
99. Kröger N, Wetherbee R. 2000 Pleuralins are involved in theca differentiation in the diatom *Cylindrotheca fusiformis*. *Protist* **151**, 263–273. (doi:10.1078/1434-4610-00024)
100. Miyahara M, Aoi M, Inoue-Kashino N, Kashino Y, Ifuku K. 2013 Highly efficient transformation of the diatom *Phaeodactylum tricornutum* by multi-pulse electroporation. *Biosci. Biotechnol. Biochem.* **77**, 874–876. (doi:10.1271/bbb.120936)
101. Niu YF, Yang ZK, Zhang MH, Zhu CC, Yang WD, Liu JS, Li HY. 2012 Transformation of diatom *Phaeodactylum tricornutum* by electroporation and establishment of inducible selection marker. *BioTechniques* **52**, 1–3. (doi:10.2144/000113881)
102. Karas BJ *et al.* 2015 Designer diatom episomes delivered by bacterial conjugation. *Nat. Commun.* **6**, 6925. (doi:10.1038/ncomms7925)
103. Poulsen N, Kroger N. 2005 A new molecular tool for transgenic diatoms: control of mRNA and protein biosynthesis by an inducible promoter-terminator cassette. *FEBS J.* **272**, 3413–3423. (doi:10.1111/j.1742-4658.2005.04760.x)
104. 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**, e96. (doi:10.1093/nar/gkp448)

105. Daboussi F *et al.* 2014 Genome engineering empowers the diatom *Phaeodactylum tricornutum* for biotechnology. *Nat. Commun.* **5**, 3831. (doi:10.1038/ncomms4831)

106. Serif M, Lepetit B, Weißert K, Kroth PG, Rio Bartulos C. 2017 A fast and reliable strategy to generate TALEN-mediated gene knockouts in the diatom *Phaeodactylum tricornutum*. *Algal Res.* **23**, 186 – 195. (doi:10.1016/j.algal.2017.02.005)

107. Nymark M, Sharma AK, Sparstad T, Bones AM, Winge P. 2016 A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci. Rep.* **6**, 24951. (doi:10.1038/srep24951)

108. Hopes A, Nekrasov V, Kamoun S, Mock T. 2016 Editing of the urease gene by CRISPR-Cas in the diatom *Thalassiosira pseudonana*. *Plant Methods* **12**, 49. (doi:10.1186/s13007-016-0148-0)

109. Zaslavskaia LA, Lippmeier JC, Shih C, Ehrhardt D, Grossman AR, Apt KE. 2001 Trophic conversion of an obligate photoautotrophic organism through metabolic engineering. *Science* **292**, 2073 – 2075. (doi:10.1126/science.160015)

110. Hamilton ML, Powers S, Napier JA, Sayanova O. 2016 Heterotrophic production of omega-3 long-chain polyunsaturated fatty acids by trophically converted marine diatom *Phaeodactylum tricornutum*. *Mar. Drugs* **14**, 53. (doi:10.3390/md14030053)

111. Dinamarca J, Levitan O, Kumaraswamy GK, Lun DS, Falkowski PG. 2017 Overexpression of a diacylglycerol acyltransferase gene in *Phaeodactylum tricornutum* directs carbon towards lipid biosynthesis. *J. Phycol.* **53**, 405 – 414. (doi:10.1111/jpy.12513)

112. Zhu BH, Shi HP, Yang GP, Lv NN, Yang M, Pan KH. 2016 Silencing UDP-glucose pyrophosphorylase gene in *Phaeodactylum tricornutum* affects carbon allocation. *New Biotechnol.* **33**, 237 – 244. (doi:10.1016/j.nbt.2015.06.003)

113. Haimovich-Dayan M, Garfinkel N, Ewe D, Marcus Y, Gruber A, Wagner H, Kroth PG, Kaplan A. 2013 The role of C4 metabolism in the marine diatom *Phaeodactylum tricornutum*. *New Phytol.* **197**, 177 – 185. (doi:10.1111/j.1469-8137.2012.04375.x)