Identification and Localization of Peroxisomal Biogenesis Proteins Indicates the Presence of Peroxisomes in the Cryptophyte *Guillardia theta* and Other "Chromalveolates"

Ann-Kathrin Mix[1], Ugo Cenci[2], Thomas Heimerl[3], Pia Marter[1], Marie-Louise Wirkner[1], and Daniel Moog[1,*]

[1]Laboratory for Cell Biology, Philipps University Marburg, Germany

[2]Unité de Glycobiologie Structurale et Fonctionnelle, UMR 8576 CNRS-USTL, Université des Sciences et Technologies de Lille, Villeneuve d'Ascq Cedex, France

[3]LOEWE Center for Synthetic Microbiology (Synmikro), Philipps University Marburg, Germany

*Corresponding author: E-mail: daniel.moog@biologie.uni-marburg.de.

Accepted: September 21, 2018

Abstract

Peroxisomes are single-membrane-bound organelles with a huge metabolic versatility, including the degradation of fatty acids (β-oxidation) and the detoxification of reactive oxygen species as most conserved functions. Although peroxisomes seem to be present in the majority of investigated eukaryotes, where they are responsible for many eclectic and important spatially separated metabolic reactions, knowledge about their existence in the plethora of protists (eukaryotic microorganisms) is scarce. Here, we investigated genomic data of organisms containing complex plastids with red algal ancestry (so-called "chromalveolates") for the presence of genes encoding peroxins—factors specific for the biogenesis, maintenance, and division of peroxisomes in eukaryotic cells. Our focus was on the cryptophyte *Guillardia theta*, a marine microalga, which possesses two phylogenetically different nuclei of host and endosymbiont origin, respectively, thus being of enormous evolutionary significance. Besides the identification of a complete set of peroxins in *G. theta*, we heterologously localized selected factors as GFP fusion proteins via confocal and electron microscopy in the model diatom *Phaeodactylum tricornutum*. Furthermore, we show that peroxins, and thus most likely peroxisomes, are present in haptophytes as well as eustigmatophytes, brown algae, and alveolates including dinoflagellates, chromerids, and noncoccidian apicomplexans. Our results indicate that diatoms are not the only "chromalveolate" group devoid of the PTS2 receptor Pex7, and thus a PTS2-dependent peroxisomal import pathway, which seems to be absent in haptophytes (*Emiliania huxleyi*) as well. Moreover, important aspects of peroxisomal biosynthesis and protein import in "chromalveolates" are highlighted.

Key words: chromalveolates, peroxisome, peroxin, PTS1, PTS2, peroxisomal protein import.

Introduction

Peroxisomes are the last identified, so far most neglected and underestimated eukaryotic organelles. Besides their most prevalent functions in fatty acid degradation and cellular detoxification, peroxisomal metabolism can be extremely versatile (Smith and Aitchison 2013; Gabaldon et al. 2016; Deb and Nagotu 2017). It seems that these genome-less, single membrane-bound, dynamic compartments can be individually adapted to specific cellular needs by integrating whole or partial biochemical pathways in a "reaction chamber" that is spatially separated from the remaining cellular processes. Nevertheless, peroxisomes are usually—together with other compartments and organelles, such as the cytoplasm, mitochondria, and plastids—entangled into an intricate cellular network (Hu et al. 2012; Cross et al. 2016; Fransen et al. 2017). The latter is not only based on the frequent exchange of metabolites and intermediates of important biochemical pathways but also on physical contacts between the involved subcellular structures (Shai et al. 2016).

Although at least some eukaryotes—most of them parasites—have lost peroxisomes (Gabaldon et al. 2006, 2016; Schlüter et al. 2006; Zarsky and Tachezy 2015) studies over the last decades have made it clear that peroxisomes play important roles in the majority of eukaryotic cells. In humans, malfunctions in these organelles often related to a loss of function of components of their biogenesis machinery are

known to cause severe disorders. Many of these have dramatic phenotypic effects or are even lethal, as for example in case of the Zellweger syndrome (Smith and Aitchison 2013; Waterham et al. 2016). The biogenesis, maintenance, and proliferation of peroxisomes, which at least in mammals are formed by fusion of vesicles buddying from the ER and mitochondria (Sugiura et al. 2017), is conducted by a set of proteins known as peroxins (Pex) (Pieuchot and Jedd 2012; Erdmann 2016). The presence of certain core peroxins in a eukaryotic genome can be taken as an indicator for the existence of peroxisomes within the individual organisms, as these factors are specific to peroxisomal cell biology (Gabaldon 2010). Besides executing peroxisomal de novo biogenesis and division, the majority of peroxins is involved in the import of proteins into peroxisomes (Rucktäschel et al. 2011). Peroxisomal protein import is a complex process that, although widely conserved in its basic principles, can differ between peroxisome bearing eukaryotes (Galland and Michels 2010; Smith and Aitchison 2013).

One considerable difference of peroxisomal import to many other cellular protein import systems is that matrix proteins are imported post-translationally in a folded state, or even as protein complexes, across the peroxisomal membrane (Walton et al. 1995; Baker et al. 2016). There are two major routes for soluble proteins to reach the peroxisomal matrix. Proteins containing a peroxisomal targeting signal of type 1 (PTS1), a C-terminal tripeptide usually composed of the amino acids [SAC][KRH][LM], are recognized and imported via the Pex5 receptor, whereas proteins containing a PTS2—an N-terminal nona-peptide with the consensus sequence [RK][LVI]X_5[HQ][LA]—are transported in a Pex7-dependent manner (Lametschwandtner et al. 1998; Platta and Erdmann 2007; Rucktäschel et al. 2011; Pieuchot and Jedd 2012; Kim and Hettema 2015). Whereas Pex5 alone is sufficient as a receptor for PTS1 protein import, Pex7 is depending on a coreceptor for importing PTS2-proteins into peroxisomes via a "piggyback-like" mechanism (Kunze et al. 2015). In plants and mammals, this coreceptor is Pex5, or rather a longer form of the protein called Pex5L, which contains a Pex7-binding motif, whereas, for example, yeasts use either Pex20p or Pex18p, and Pex21p as Pex7 coreceptors (Braverman et al. 1998; Montilla-Martinez et al. 2015; Meinecke et al. 2016). Pex5 not only participates in substrate recognition and transport, it also plays a significant role in the translocation process of soluble proteins across the peroxisomal membrane by forming a most likely transient pore together with the membrane protein Pex14, at least (Erdmann and Schliebs 2005; Meinecke et al. 2010; Dias et al. 2017). After release of the cargo into the peroxisomal matrix, Pex5 becomes ubiquitinated by the so-called RING-complex consisting of the ubiquitin ligases Pex2, Pex10, and Pex12 as well as the ubiquitin conjugating enzyme Pex4 and its membrane anchor Pex22 and gets extracted from the peroxisomal membrane via energy-consumptive action of two cytosolic AAA-ATPases, Pex1, and Pex6 (Pieuchot and Jedd 2012; Smith and Aitchison 2013; Hettema et al. 2014). Once back in the cytosol Pex5 (and Pex7) can enter into a new translocation cycle as soon as an ubiquitin moiety has been removed by a deubiquitination enzyme. Alternatively, poly-ubiquitinated Pex5 is removed from the cycle and delivered to the proteasome for degradation (see e.g., Francisco et al. 2014 for review).

In contrast, the import machinery for peroxisomal membrane proteins (PMPs), which are classified into class I and II PMPs depending on their import route, is less understood and involves other peroxins. The membrane peroxisomal targeting signals (mPTS) of class I PMPs are recognized by the Pex19 receptor whereupon the receptor–cargo complex binds to Pex3, an integral protein of the peroxisomal membrane, followed by release and membrane integration of the cargo (Fujiki et al. 2006; Rucktäschel et al. 2011). Class II PMPs are inserted into the ER membrane probably via involvement of another peroxin, Pex16, or in a Sec61- or Get3-dependent manner and reach the peroxisome via a vesicular connection (Kim and Hettema 2015; Mayerhofer 2016). Pex16 is also known to play a role in peroxisomal biogenesis as is another peroxin, Pex11, which is essential for the division of existing peroxisomal structures. However, the latter factor seems not to be involved in peroxisomal protein import (Thoms and Erdmann 2005).

Together these peroxisomal biogenesis and protein import factors ensure functionality of the compartmentalized, but highly interconnected, metabolic networks that are a consequence of the complexity of eukaryotic cells. Some of the most complex organized life-forms on Earth with respect to their subcellular structure are represented by organisms with so-called complex plastids. These go back to an endosymbiosis in which a eukaryote has taken up another, photosynthetic eukaryote in form of a red or green alga that has become an organelle (Keeling 2010; Moog and Maier 2017; Gould 2018). Organisms with complex plastids of red algal origin comprise many evolutionarily and ecologically, but also medically relevant groups such as the cryptophytes, stramenopiles (including diatoms), haptophytes, and alveolates, the latter containing dinoflagellates, chromerids as well as parasitic apicomplexans (Archibald 2015; Maier et al. 2015; Gentil et al. 2017). Together they have been termed "chromalveolates" (Cavalier-Smith 1999, 2000, 2003) although a monophyletic origin of these organisms is doubted (Baurain et al. 2010; Petersen et al. 2014; Stiller et al. 2014). The intricate cellular organization of "chromalveolates" is reflected by an enormous, and individual, metabolic compartmentalization that is so far only partly understood for merely a handful of species. Studies on peroxisomes of "chromalveolates" have been undertaken mainly in diatoms and alveolates (Gonzalez et al. 2011; Moog et al. 2017; Ludewig-Klingner et al. 2018). Diatoms are reported to harbor peroxisomes with the metabolic capacity for fatty acid degradation, the glyoxylate cycle and detoxification of reactive oxygen species (ROS), which use

exclusively PTS1 for protein import (Gonzalez et al. 2011). In alveolates, the situation is somewhat more complicated. Whereas peroxisomes have been identified in dinoflagellates, chromerids, and apicomplexans, most of the latter seem to have lost peroxisomes probably due to adaptations to their parasitic lifestyle (Moog et al. 2017). Moreover, based on the identification of relevant *pex* genes in genomic data, all of the peroxisome bearing organisms investigated seem to be capable of both PTS1- and PTS2-dependent peroxisomal protein import (Moog et al. 2017; Ludewig-Klingner et al. 2018).

In contrast to diatoms and alveolates, for other important groups of organisms containing complex plastids such as cryptophytes, haptophytes, and eustigmatophytes—the latter belonging to the stramenopiles—the presence and functions of peroxisomes is still enigmatic. Here, we show via bioinformatic screenings that genes for peroxins are present in all major groups of organisms with complex red plastids. Our results allow us to deduce the existence of peroxisomes in cryptophytes, haptophytes, eustigmatophytes, and phaeophytes (brown algae), expand current knowledge on peroxisomes in diatoms and alveolates and provide important insights into the composition of potential targeting systems for peroxisomal proteins in "chromalveolates." Moreover, our in silico data for cryptophytes are supported by extensive heterologous localization studies via confocal and electron microscopy.

Materials and Methods

In Silico Analyses/Bioinformatics

To identify peroxins in the genome data of the cryptophyte *Guillardia theta*, the following in silico analyses were performed: 1) Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.genome.jp/kegg/; last accessed June 5, 2018) analysis with *G. theta* proteins (Guith1, gene catalog, best model proteins, 24,840; Curtis et al. 2012) and screening of classification category "peroxisome" for annotated peroxins; 2) BlastP with known Pex protein sequences from yeast (*Saccharomyces cerevisiae*) and *Arabidopsis thaliana* (NCBI) against the JGI *G. theta* all model proteins database (https://genome.jgi.doe.gov/Guith1/Guith1.home.html; e-value cutoff: e-4; last accessed June 5, 2018) and for some peroxin queries (Pex3, Pex13, Pex24): local HMMER search (Eddy 1998) with HMMER profiles build from "seed" and "full" Stockholm alignments extracted from the Pfam database (http://pfam.xfam.org/; last accessed June 5, 2018) against a *G. theta* protein database consisting of sequences extracted from NCBI (51,636; https://www.ncbi.nlm.nih.gov/protein/?term=Guillardia+theta; last accessed June 5, 2018); 3) screening of the *G. theta* JGI genome database via keyword search for "peroxin," "pex," "peroxisomal," and "peroxisome"; 4) extracting the genomic and transcript information from the individual gene models and BlastN versus all_model_transcripts and the three EST databases included in the *G. theta* JGI genome database (see above) as well as the *G. theta* CCMP2712 transcriptome reads (SRR747855) (https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR747855; last accessed June 5, 2018) and the *G. theta* EST databases at NCBI (rna_refseq and est) and the MMETSP (Keeling et al. 2014); 5) manual inspection of each individual gene model using the software tool Sequencher (Genecodes); 6) validation of hits via reciprocal BlastP analysis of sequences against the NCBI nr database (https://blast.ncbi.nlm.nih.gov/Blast.cgi; last accessed June 5, 2018) and identification of conserved domains with NCBI Conserved Domain Search (CDS; https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi; last accessed June 5, 2018).

For the identification of peroxins in other "chromalveolates," sequences of *G. theta* and the diatom *Phaeodactylum tricornutum* (Gonzalez et al. 2011) were used in addition to queries from yeast and *Arabidopsis* (NCBI) for BlastP analyses (e-value cutoff: e-4) against the genome data of: *Emiliania huxleyi* (https://genome.jgi.doe.gov/Emihu1/Emihu1.home.html; last accessed June 5, 2018; Read et al. 2013), *Chrysochromulina tobin* CCMP291 (NCBI; Hovde et al. 2015), *P. tricornutum* CCAP 1055/1 (https://genome.jgi.doe.gov/Phatr2/Phatr2.home.html; last accessed June 5, 2018; Bowler et al. 2008), *Thalassiosira pseudonana* CCMP1335 (https://genome.jgi.doe.gov/Thaps3/Thaps3.home.html; last accessed June 5, 2018; Armbrust et al. 2004), *Aureococcus anophagefferens* clone 1984 (https://genome.jgi.doe.gov/Auran1/Auran1.home.html; last accessed June 5, 2018; Gobler et al. 2011), *Nannochloropsis gaditana* strain: B-31 (NCBI, Corteggiani Carpinelli et al. 2014), *Nannochloropsis oceanica* CCMP1779 (https://genome.jgi.doe.gov/Nanoce1779/Nanoce1779.home.html; last accessed June 5, 2018; Vieler et al. 2012), *Ectocarpus siliculosus* Ec 32 CCAP 1310/04 (NCBI; Cock et al. 2010), and *Symbiodinium microadriaticum* CCMP2467 (NCBI; Aranda et al. 2016). Similar to the procedure for *G. theta* peroxin candidates, hits were analyzed for validation via reciprocal BlastP against NCBI nr and conserved domain prediction using NCBI Conserved Domain Search (see above). Targeting prediction for the detected protein sequences was carried out as described previously (Moog et al. 2017).

Culture Conditions

The diatom *P. tricornutum* (Bohlin, UTEX646) was cultured at 21°C in Erlenmeyer flasks under agitation (150–200 rpm) in constant light (24 h; 8,000–10,000 Lux) in f/2 medium (pH 7.0) containing 1.66% (w/v) Tropic Marin (Dr. Biener GmbH) salt, 2 mM Tris/HCl (pH 8.0), and 1.5 mM NH_4Cl as a nitrogen source, or on solid f/2 plates with agar agar (1.5% w/v). Zeocin (InvivoGen) was added to a final concentration of 75 μg/ml for selection of positively transformed clones (see below). For induction of protein overexpression, cells were

transferred from solid f/2 agar plates to 1.5-ml reaction tubes containing 50–75 µl liquid f/2 medium with 0.9 mM $NaNO_3$ instead of NH_4Cl and incubated for 24 h as described earlier.

The cryptophyte *G. theta* (CCAM2327/CCMP2712) was cultured at 21°C in stationary Erlenmeyer flasks in a daily 14 h light and 10 h dark cycle (ca. 750 Lux) in f/2 liquid medium (pH 7.2) containing 3.0% (w/v) Tropic Marin (Dr. Biener GmbH) salt, 5 mM Tris/HCl (pH 8.0), and 500 mM NH_4Cl as a nitrogen source.

DNA-/RNA-Isolation and cDNA Synthesis

For DNA-/RNA-isolation, *G. theta* or *P. tricornutum* were inoculated in a volume of 150 ml f/2 and grown for 7 days, respectively, as described earlier. *Phaeodactylum tricornutum* cells were harvested by centrifugation at 1.500 × g and 21°C for 5 min., whereas *G. theta* was centrifuged at 3.000 × g and 21°C for 10 min. DNA was isolated from the pelletized cells via the CTAB method. *Guillardia theta* RNA was extracted using the RNeasy Mini kit from Qiagen. About 1 µg of RNA was treated with DNAseI (Thermo Fisher Scientific) to remove potential genomic DNA contamination followed by cDNA synthesis via the RevertAid First Strand cDNA Synthesis kit (Thermo Fisher Scientific).

Gene Sequence Amplification, Plasmid Construction, and Transformation

Genes for localization studies were amplified via standard polymerase chain reaction (PCR) using the PHUSION High Fidelity DNA polymerase (Finnzymes) and gene-specific oligonucleotides synthesized by Sigma–Aldrich usually equipped with terminal restriction sites (see Supplementary Material online for primer sequences). PCR amplified fragments were cloned into pJet1.2/blunt via the CloneJET PCR Cloning Kit (Thermo Fisher Scientific), sequenced (Macrogen), and cloned together with the gene encoding the enhanced green fluorescent protein (*egfp*) into the pPha-NR plasmid (GenBank: JN180663). Prior to biolistic transformation of diatoms, generated plasmids were isolated from a 50-ml *E. coli* culture using the Qiagen QIAfilter Midi/Maxi Kit and again verified by sequencing. Transformation of *P. tricornutum* cells then was performed with plasmids coupled to Tungsten M10 particles in a Biolistic PDS-1000/He Particle Delivery System (Biorad) using 1,350 psi rupture discs as described previously (Apt et al. 1996; Sommer et al. 2007).

Confocal Laser Scanning Microscopy

To analyze the subcellular localization of an individual GFP fusion construct, gene expression was induced for 24 h after growth medium shift of f/2 with 1.5 mM NH_4Cl (uninduced condition; no expression) to f/2 containing 0.9 mM $NaNO_3$ (induction of nitrate reductase promoter in pPha-NR; expression). Fusion protein localization was then monitored with a Leica TCS SP2 confocal laser scanning microscope and a HCX PL APO 40_/1.25–0.75 Oil CS objective. eGFP and chlorophyll/autofluorescence of the diatom plastid was excited at 488 nm using a 65-mW Argon laser, whereas emission spectra were detected at a bandwidth of 500–520 nm for eGFP and 625–720 nm for the plastid autofluorescence, respectively, via two photomultiplier tubes. Image processing was carried out with Leica LAS AF lite and ImageJ (Schneider et al. 2012) software.

Transmission Electron Microscopy

Selected diatom clones expressing either *P. tricornutum* or *G. theta* Pex16-GFP were further analyzed via transmission electron microscopy (TEM). To this end, clones were grown for 7 days in 50 ml f/2 medium under noninducing conditions (NH_4Cl as nitrogen source) followed by another day of growth under inducing conditions in 50 ml f/2 containing 0.9 mM $NaNO_3$. Cells were then harvested by centrifugation (1.500 × g, 21°C, 5 min.) and subjected to high-pressure freezing. Subsequent freeze substitution (with acetone, containing 0.25% osmium tetroxide, 0.2% uranyl acetate, 0.05% ruthenium red, and 5% water), resin embedding, and sectioning was done as previously described (Renicke et al. 2017).

Labeling of GFP(-tagged proteins) on ultrathin sections was performed as described previously (Liu et al. 2016). Analysis of the samples was conducted with a JEOL JEM2100 TEM equipped with a fast-scan 2k CCD TVIPS (Gauting, Germany) F214 camera (Department for Cell Biology, Philipps University Marburg) or a JEOL JEM1400 TEM with a 4k TVIPS TemCam F416 camera (Institute for Virology, Philipps University Marburg).

Phylogeny

For phylogenetic analyses, we used a set of the protein sequences from cryptophytes, apicomplexans, and dinoflagellates identified here and previously to perform BLAST analyses against a local database containing sequences compiled from different sources (e.g., MMETSP, NCBI, JGI) without applying any filters or clade preferences. We aligned all sequences selected with an E-value less than 1e-10 using MAFFT with high speed settings (Katoh and Standley 2013). Preliminary trees were generated with Fasttree (Price et al. 2010) after using BMGE (Criscuolo and Gribaldo 2010) with a block size of 4 and the BLOSUM30 similarity matrix. We applied a "dereplication" step with TreeTrimmer (Maruyama et al. 2013), sizing down robustly supported monophyletic clades in order to reduce sequence redundancy. The final sets of sequences were selected manually. When we considered a set of sequences to be of sufficient high quality, we realigned protein sequences with MUSCLE, carried out a block selection using BMGE with a block size of four and the matrix BLOSUM30. Finally, trees were generated using Phylobayes-4.1 (Lartillot et al. 2009) under the catfix C20 + Poisson model with the two chains stopped when convergence was reached

(maxdiff < 0.05) after at least 900 cycles, discarding 200 burn-in trees. Bootstrap support values were estimated from 100 replicates using IQ-TREE (Nguyen et al. 2015) under the LG4X model (Le et al. 2012) and mapped onto the Bayesian tree.

Results

Peroxins in *G. theta* and Other "Chromalveolates"

The presence of genes encoding peroxins was first investigated in the genome of the cryptophyte *G. theta* via BLAST and HMMER searches as well as by screening genome annotation data using the KEGG (Kyoto Encyclopedia of Genes and Genomes) pipeline (see Materials and Methods). Our survey led to the identification of a set of 16 peroxins (including two isoforms of two components, respectively; table 1) most likely sufficient for peroxisome biogenesis, division, and maintenance in the cryptophyte. The set includes the two receptors for PTS1 and PTS2, Pex5 and Pex7, the docking complex component Pex14, the RING-finger complex factors Pex2, Pex10, and Pex12 as well as the ubiquitin-conjugating enzyme Pex4, its membrane anchor Pex22, and the two AAA-type ATPases Pex1 and Pex6 for receptor recycling (table 1). In addition, we detected three peroxins, Pex3, Pex16, and Pex19, with possible involvement in the transport of peroxisomal membrane proteins. Interestingly, Pex19 seems to be present in two copies in *G. theta*, a situation similar to the peroxisomal division factor Pex11 (also two isoforms detected). Not detected was Pex13, which in other organisms, besides Pex14, constitutes an important component of the peroxisomal docking complex (Smith and Aitchison 2013) (table 1).

Haptophytes are a further group of organisms with complex plastids of red-algal origin for which the presence of peroxisomes is enigmatic. Screening the genomic data of the cosmopolitan alga *E. huxleyi* led to the identification of 9 conserved classical peroxin genes encoding Pex1 to Pex6, Pex10, Pex12, and Pex22, whereas the PTS2 receptor Pex7, Pex11 (peroxisome division), the docking complex components Pex13 and Pex14 as well as Pex16 and Pex19 (peroxisomal membrane protein import) were not detected (table 1). It is unclear whether the rather fragmentary occurrence of *pex* genes in *E. huxleyi* is authentic or due to a high divergence of the not detected factors. Even the quality of the genome data cannot be excluded as a reason for this result. Thus, we have screened additional genomic data for another haptophyte, *C. tobin*, to obtain comparative data for this group of organisms. As table 1 shows with five components, the resulting list of detected *pex* genes in *C. tobin* is even shorter as for *E. huxleyi*. However, whereas in both data sets, the ones for *C. tobin* and *E. huxleyi*, a gene encoding Pex7 seems to be absent, Pex19, the putative receptor for peroxisomal membrane proteins of class I was identified in the genome of *C. tobin*, indicating the general presence of this factors in haptophytes (table 1).

Table 1

Peroxins in *Guillardia theta* and Other "Chromalveolates"

	Cr.	Ha.		Stramenopiles					Alveolates[e]			
Peroxin	**Gt[a]**	**Eh[a]**	**Ct[a]**	**Pt[a]**	**Tp[a]**	**Aa[a]**	**Ng[a]**	**Es[a]**	**Cv[a]**	**Tg[a]**	**Pf[a]**	**Sm[a]**
Pex1	+	+	+	+	+	nd	+	+	+	+	nd	+
Pex2	+	+	nd	+	nd	nd	+	+	+	+	nd	nd[c]
Pex3	+	+	+	+	+	nd	+	+	+	+	nd	nd[c]
Pex4	+	+	+	+	+	+	+	+	+	+	+	+
Pex5	+	+	nd[d]	+	+	nd	+	+	+	+	nd	+
Pex6	+	+	+	+	+	nd	+	+	+	+	nd	+
Pex7	+	nd	nd	nd	nd	nd	+	+	+	+	nd	+
Pex10	+	+	?	+	+	+	+	+	+	+	?	+
Pex11	2	nd	nd	+	+	nd	+	+	+	+	nd	+
Pex12	+	+	nd	+	+	nd	+	+	+	+	nd	nd[c]
Pex13	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
Pex14	+	nd	nd	nd	nd	nd	+	+	+	+	nd	? [c]
Pex16	+	nd	nd	+	+	nd	+	+	+	+	nd	+
Pex19	2	nd	+	+	+	+	+	+	nd[b]	nd	nd	+
Pex22	+	+	nd	nd	nd	nd	nd	+	+	+	+	nd[c]

[a]Organism name (and strain) abbreviations: Gt, *Guillardia theta* CCMP2712; Eh, *Emiliania huxleyi* CCMP1516; Ct, *Chrysochromulina tobin* CCMP291; Pt, *Phaeodactylum tricornutum* CCAP 1055/1; Tp, *Thalassiosira pseudonana* CCMP1335; Aa, *Aureococcus anophagefferens* clone 1984; Ng, *Nannochloropsis gaditana*; Es, *Ectocarpus siliculosus* Ec 32 CCAP 1310/04; Cv, *Chromera velia* CCMP2878; Tg, *Toxoplasma gondii* ME49; Pf, *Plasmodium falciparum* 3D7; Sm, *Symbiodinium microadriaticum* CCMP2467.

[b]As shown in Moog et al. (2017) a Pex19 candidate was detected in a related chromerid species, *Vitrella brassicaformis* CCMP3155 (NCBI accession number: CEL92557.1).

[c]According to Ludewig-Klingner et al. (2018) present in other dinoflagellates.

[d]According to transcript data (MMETSP database; Keeling et al. 2014) present in other haptophytes: *Chrysochromulina rotalis* (CAMPEP_0115884546) and *Phaeocystis antarctica* (CAMPEP_0172982558).

[e]Data for Cv, Tg, and Pf were already published in Moog et al. (2017), in which the presence of peroxins in apicomplexans was studied in more detail.

Cr., cryptophytes; Ha., haptophytes; +, Ortholog present; 2, two potential orthologs identified; ?, Orthology status unclear; nd, not detected. See supplementary table S1, Supplementary Material online, for protein identification numbers.

Studies concerning the presence of peroxins in the diatom *P. tricornutum* have been conducted previously (Gonzalez et al. 2011; Gabaldon et al. 2016; Gentekaki et al. 2017). Here, we reinvestigated the genome of *P. tricornutum* for genes encoding peroxins, which led to the so far most comprehensive set of identified peroxins involved in peroxisomal protein import and biogenesis in the diatom (table 1). In addition, we screened the genomes of other closely and more distantly related stramenopiles including the diatom *T. pseudonana*, the pelagophyte *A. anophagefferens* the eustigmatophyte *N. gaditana* (and *N. oceanica*; see supplementary table S1, Supplementary Material online) as well as the multicellular brown alga *E. siliculosus*. As known from previous studies, diatoms lack the machinery for PTS2-mediated peroxisomal protein import (Gonzalez et al. 2011). Interestingly, this does not seem to be the case in all stramenopiles, as we were able to identify genes encoding the PTS2 receptor Pex7 in two *Nannochloropsis* spp. and *E. siliculosus* (table 1). The determined set of peroxins is comparably comprehensive in *P. tricornutum* and the closely related diatom *T. pseudonana*,

with the exception that we could not identify Pex2 in the latter. In contrast to the other stramenopiles listed in table 1, *N. gaditana* and *E. siliculosus* seem to possess Pex14, whereas *E. siliculosus* is the only organism in this group for which Pex22 could be identified (membrane anchor for Pex4); a Pex4 gene, however, is present in all of the investigated stramenopiles (table 1). The patchiest list of peroxins emerged for *A. anophagefferens*. Only genes encoding the ubiquitin conjugating enzyme Pex4, the putative ubiquitin ligase Pex10 and the receptor for peroxisomal membrane proteins of type 1, Pex19, were detected in the genome of the pelagophyte (table 1). The macroalga *E. siliculosus*, on the other hand, has the most complete set of Pex factors (14) with only Pex13 (similar to the other stramenopiles) having not been identified in its genomic data (table 1).

The presence and distribution of *pex* genes in alveolates has been investigated by us and others in earlier studies (Moog et al. 2017; Ludewig-Klingner et al. 2018). Nonetheless, we have included an excerpt of these results into table 1 for completeness. The data show that chromerids (*Chromera velia*) as well as certain apicomplexan parasites (coccidians, example: *Toxoplasma gondii*) contain a full set of *pex* genes most likely capable of expressing factors for peroxisome biogenesis, maintenance, and (PTS1- and PTS2-mediated) protein import, whereas other apicomplexans, such as the malaria causing agent *Plasmodium falciparum*, lack the majority of *pex* genes, and thus most likely peroxisomes (table 1; Moog et al. 2017; Ludewig-Klingner et al. 2018). Here, we additionally screened genomic data for a dinoflagellate, *S. microadriaticum*, a coral symbiont, which revealed that *pex* genes are also present in this diverse branch of the alveolates (Ludewig-Klingner et al. 2018). As for all organisms with complex red plastids screened during this study, we did not detect a Pex13 gene in *S. microadriaticum* (table 1). However, the dinoflagellate seems to possess genes encoding the three peroxisomal receptors Pex5, Pex7, and Pex19 as well as six additional peroxins (9 in total; table 1) probably sufficient to synthesize and maintain functional peroxisomes.

Although not applicable for the investigated alveolates, one apparent peculiarity resulting from our screening was that in organisms in which no gene for Pex7 could be identified, also the gene encoding Pex14 seemed to be absent. This was the case in the two diatoms, *A. anophagefferens* and the two haptophytes (table 1). Moreover, *G. theta* appears to be the only organisms in this in silico analysis for which isoforms of certain peroxins (Pex11 and Pex19) were detected.

Conserved Domains of *G. theta* Peroxins

BLAST analyses revealed that the *G. theta* Pex16, similar to the *P. tricornutum* homolog (Pt_Pex16), possesses a conserved Pex16 domain, which is, however, located much more N-terminal in comparison to the one detected for Pt_Pex16 (domain C-terminal). At least one putative TMD has been determined for Gt_Pex16 via TMHMM and TOPCONS, which is in contrast to Pt_Pex16 for which no putative transmembrane domains (TMDs) could be identified (see supplementary table S2, Supplementary Material online). The two Pex11 isoforms of *G. theta* have a high degree of conservation throughout the entire primary sequence, respectively, and for both a conserved Pex11 domain was detected via NCBI conserved domain search as well as one single TMD by TOPCONS (see Materials and Methods). The three putative peroxisomal ubiquitin ligases of *G. theta*—Pex2, Pex10, and Pex12—all possess a conserved Pex2_Pex12 superfamily domain, but only for Pex2 and Pex12 a C-terminal RING finger domain, typically present in these E3 ligase proteins, was identified via conserved domain search (NCBI). With respect to the presence of putative TMDs in these proteins, between zero and four TMDs could be predicted via TOPCONS and TMHMM (see supplementary table S2, Supplementary Material online).

Pex14, the only component of the peroxisomal docking complex identified in *G. theta*, as per NCBI conserved domain detection, has an N-terminal conserved Pex14_N domain and TOPCONS predicted one TMD in the protein primary sequence (see supplementary fig. S1, Supplementary Material online). Whereas the highly conserved Gt_Pex4 possesses a typical ubiquitin-conjugating enzyme E2 catalytic (UBCc) domain (NCBI conserved domain search), its potential membrane anchor Pex22 lacks any functional domain prediction. Though a putative TMD was identified by TOPCONS and TMHMM (see supplementary table S2, Supplementary Material online). The two Pex19 proteins found in *G. theta* both contain a conserved Pex19 domain within their C-terminal sequence part. TMHMM and TOPCONS predicted a putative TMD for one of the isoforms (Gt_Pex19-1). For the putative Gt_Pex3, which was initially identified with HMMER, a potential Pex3 domain was predicted via NCBI conserved domain search, although clear database hits (NCBI nr) were scarce when using the primary sequence as a query. In contrast, the putative PTS2 receptor Gt_Pex7 was found to possess a highly conserved primary sequence with the typical WD40 repeats (six in total) known to play a role in protein–protein interaction (see supplementary fig. S1, Supplementary Material online). Gt_Pex5, the putative PTS1 receptor, contains five typical TPR repeats within its C-terminal part and two Pex14 interaction sites (WxxxF motifs) as well as a N-terminal Pex5 motif are located in the N-terminal half of the primary sequence (see fig. 1 and supplementary fig. S1, Supplementary Material online). In addition to that, a region most likely important for interaction of Pex5 with the PTS2 receptor Pex7 was identified in the Gt_Pex5 sequence (see below).

Pex7-Binding Domain in Pex5 Proteins

As we have identified orthologs of Pex5 and Pex7 in *G. theta* and many other "chromalveolates," as expected, a

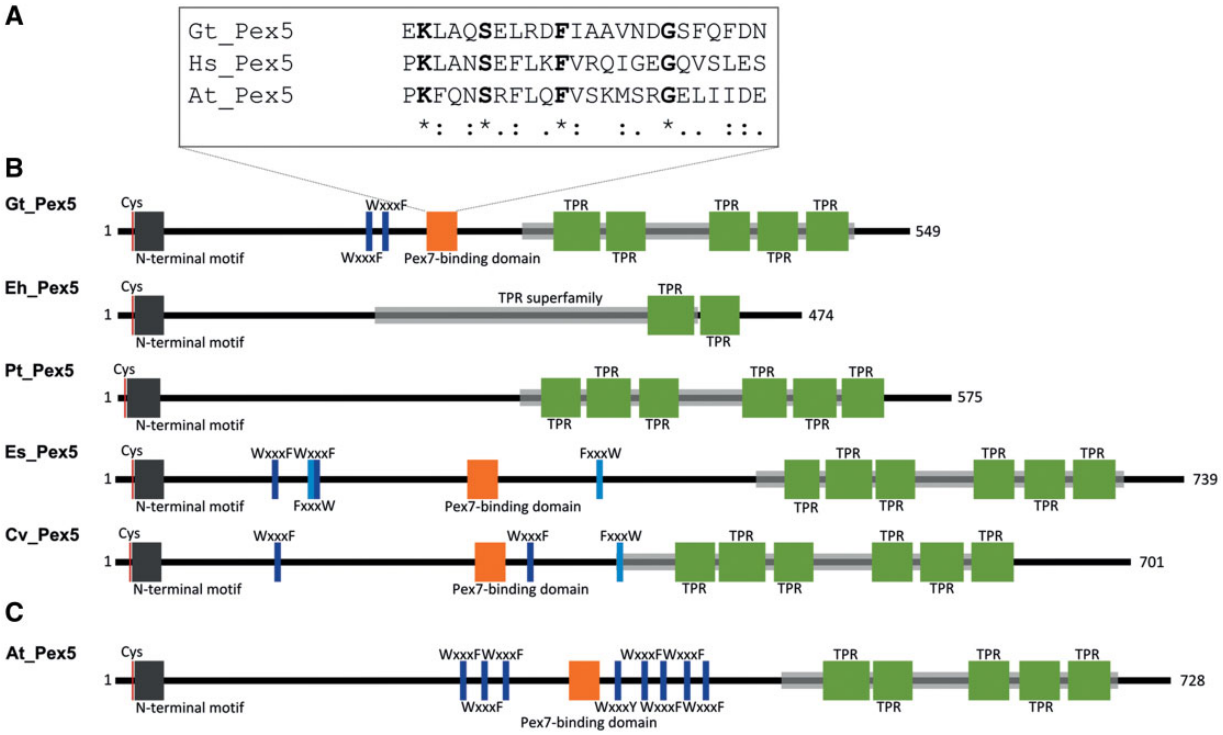

Fig. 1.—Pex7-binding site and domain structure of Pex5. (*A*) Alignment of the Pex7-binding domain identified in *Guillardia theta* with corresponding sequences from *Homo sapiens* and *Arabidopsis thaliana*. Each motif contains four highly conserved amino acids: Lys2, Ser6, Phe11, and Gly18. (*B*) Structural alignment of identified "chromalveolate" Pex5 sequences with indicated functional domains detected via NCBI conserved domain search. (*C*) Structural alignment of Pex5 from the model plant *Arabidopsis thaliana*. Numbers indicate amino acid positions. Abbreviations: Gt, *Guillardia theta*; Hs, *Homo sapiens*; At, *Arabidopsis thaliana*; Eh, *Emiliania huxleyi*; Pt, *Phaeodactylum tricornutum*; Es, *Ectocarpus siliculosus*, Cv, *Chromera velia*; WxxxF/FxxxW, putative Pex14/Pex13 interaction sites; TPR, tetratricopeptide repeat.

Pex7-binding domain/motif/interaction site was detected in the Gt_Pex5 primary sequence. This motif is known from human, plant (fig. 1*A* and *C*), and yeast Pex7 coreceptors and is composed of a stretch of 24 amino acids at the end of the first half of the protein in front of the TPR repeats/motifs (Dodt et al. 2001; Galland and Michels 2010). Figure 1*A* shows an alignment of the potential Pex7-binding motif found in the *G. theta* Pex5 (Gt_Pex5) compared with the similar sequence areas of the *Homo sapiens* and *Arabidopsis thaliana* Pex5(L) versions. This motif can also be found in nearly all other detected Pex5 sequences of chromalveolates that contain Pex7 (fig. 1*B* gives some examples; see also supplementary fig. S2*C*, Supplementary Material online); among these the stramenopile *E. siliculosus*, the chromerid *C. velia*, the apicomplexan *Toxoplasma* and, although much less conserved, the dinoflagellate *S. microadriaticum* (not included into the alignment).

Although Pex7 seems to be absent from *P. tricornutum*, we similarly screened Pt_Pex5 (32173) for the presence of a potential Pex7 interaction site (supplementary fig. S2*B*, Supplementary Material online). Not surprisingly, a motif containing the highly conserved amino acids K2, S6, F11, and G18 could not be found in the diatom Pex5 sequence. However, we detected a stretch of amino acids showing at least some general sequence similarities to a Pex7 interaction site via a primary sequence alignment. The corresponding sequence part of Pt_Pex5 is highly divergent in terms of sequence conservation when compared with the *G. theta*, human and plant Pex7 interaction motifs (supplementary fig. S2*B*, Supplementary Material online). All four of the usually highly conserved motif positions (K2, S6, F11, and G18) are divergent in the respective part of the *P. tricornutum* sequence aligning to this motif and the diatom sequence contains an additional amino acid (insertion) in the middle. Thus, it is very likely that the detected degenerate motif in the Pt_Pex5 primary sequence is nonfunctional.

As mentioned earlier and similar to *P. tricornutum*, a Pex7 homolog could not be identified in the genome of *E. huxleyi* (see table 1). This finding pointing to the absence of a PTS2-dependent import of peroxisomal proteins in the haptophyte could be supported by the complete absence of a potential Pex7 interaction site in the Eh_Pex5 primary sequence (fig. 1*B* and supplementary fig. S2*A*, Supplementary Material online). Although the sequence shows a high degree of conservation at the N-terminus and especially the C-terminal part (TPR motifs) when aligned with Gt_Pex5, a whole region of ~70

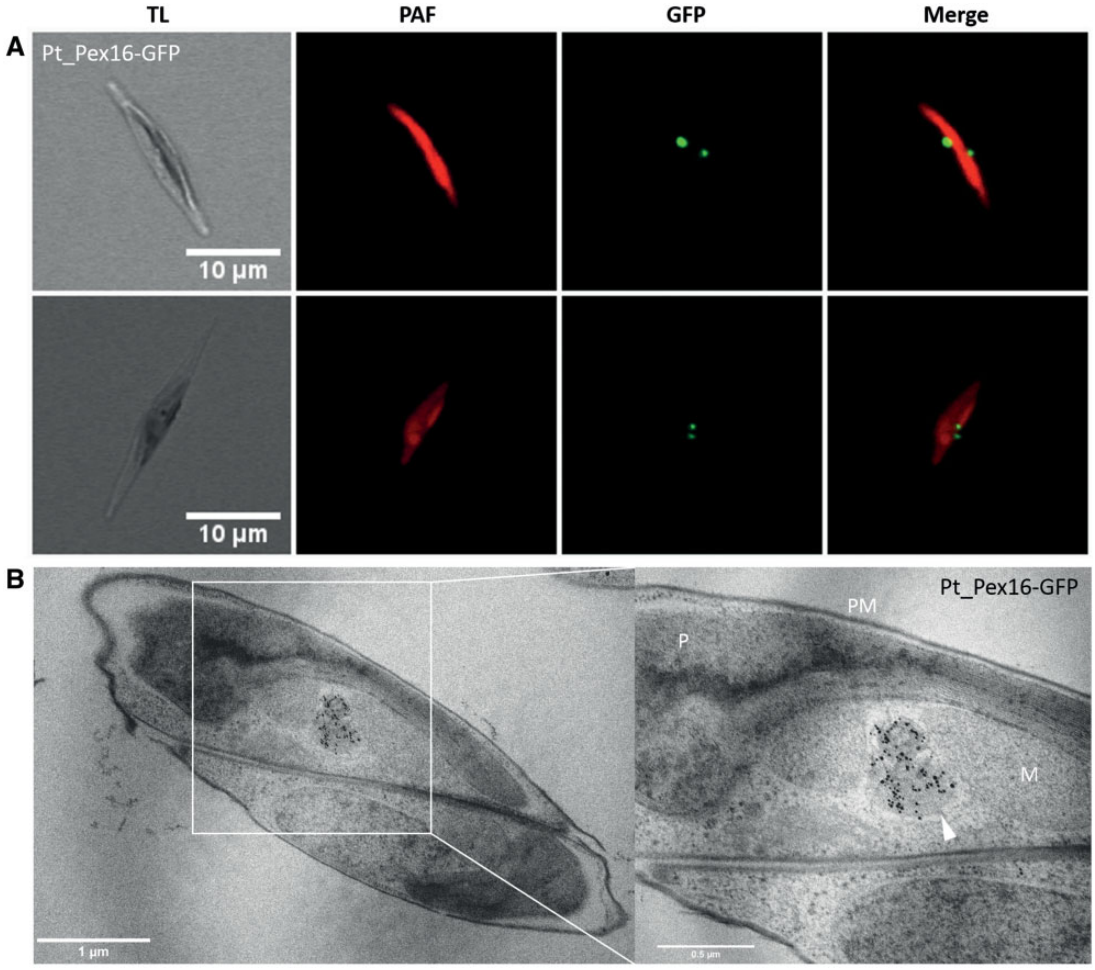

Fig. 2.—Localization studies of *Phaeodactylum tricornutum* Pex16-GFP via confocal and electron microscopy. (*A*) A confocal microscopic analysis of the Pt_Pex16-GFP expression in *P. tricornutum* (two clones) revealed that the fusion protein localized in small dot-like structures near the complex plastid, which are typical for a peroxisomal localization. TL, transmitted light; PAF, plastid autofluorescence; GFP, green fluorescent protein; Merge, overlay of PAF and GFP. (*B*) Transmission electron microscopy with a Pex16-GFP overexpressing *P. tricornutum* clone showed a specific labeling of two electron dense circular structures (arrowhead) near the mitochondrion and complex plastid of the diatom. Pt, *Phaeodactylum tricornutum*; M, mitochondrion; P, plastid; PM, plasma membrane.

amino acids is missing in Eh_Pex5 (ID: 449555) that has been identified via BlastP in the *E. huxleyi* genome database. In the comparable sequence part of Gt_Pex5, this is exactly the region where the potential Pex7 interaction site is located (supplementary fig. S2*A*, Supplementary Material online). It has to be noticed though, that the EST coverage for *E. huxleyi pex5* gene is incomplete, which does not allow us to preclude that the analyzed gene model detected in the genome database of the haptophyte is completely authentic. However, even after including transcript data information of several other *E. huxleyi* strains from the MMETSP project (Keeling et al. 2014) to clarify the *eh_pex5* gene model situation, a potential Pex7-binding motif could not be identified in the resulting translated Eh_Pex5 amino acid sequences (not shown). In addition to that, screening of another haptophyte, *C. tobin* CCMP291, genomic data also did not result in the detection of a Pex7 ortholog, which further speaks for the absence of a PTS2 pathway in haptophytes. As several other Pex factors (including Pex5) have also not been detected in this particular screening (see table 1), the genome data of *C. tobin* might be rather fragmentary. Thus, the observations made for this haptophyte have to be interpreted with caution and may not be entirely representative for haptophytes in general (see *E. huxleyi*, table 1). However, we detected putative Pex5 sequences in the transcriptomes of two other haptophytes, *Chrysochromulina rotalis* (CAMPEP_0115884546) and *Phaeocystis antarctica* (CAMPEP_0172982558) (MMETSP database; Keeling et al. 2014), during our phylogenetic analysis, respectively (see below). Both Pex5 candidates contain the typical TPR repeats known from other PTS1 receptors and, similar to the Eh_Pex5, lack a potential Pex7 interaction site, although it seems that at least the *C. rotalis* sequence might be incomplete at the N-terminus.

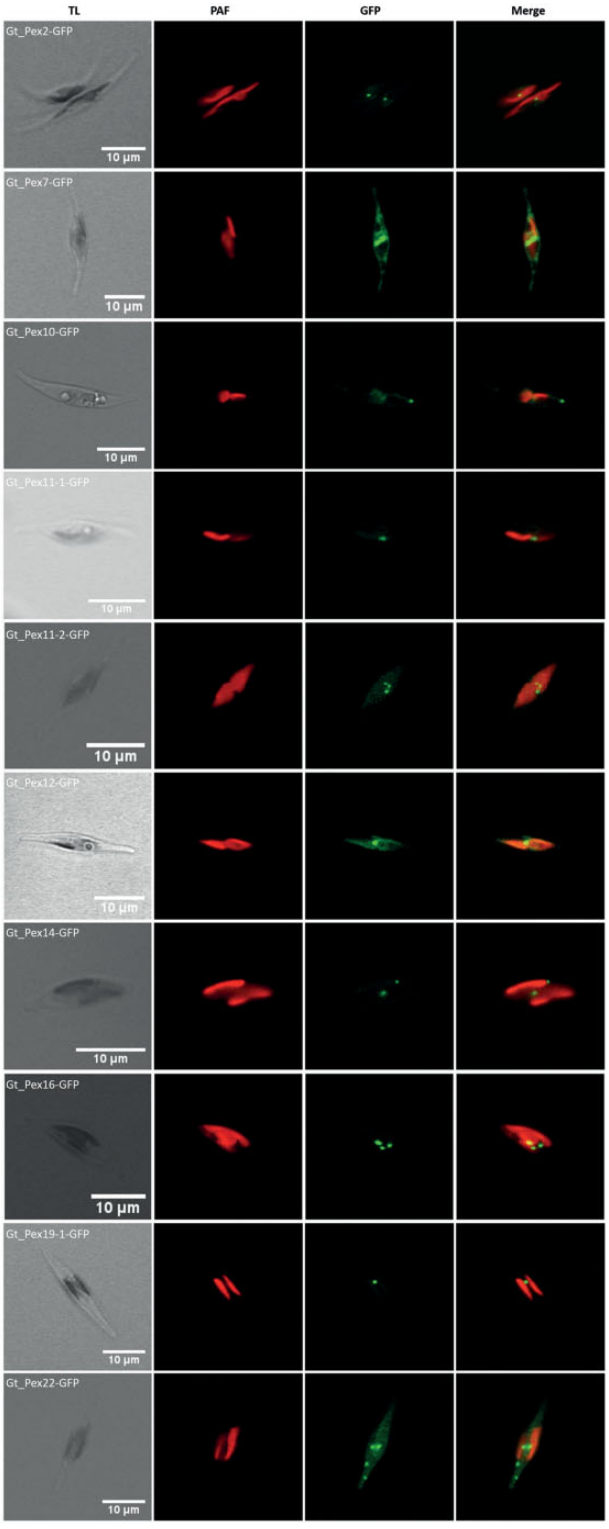

Fig. 3.—Heterologous localization studies of *Guillardia theta* peroxin-GFP fusion proteins in *Phaeodactylum tricornutum*. With the exception of the cytosolic PTS2 receptor Pex7, all *G. theta* peroxins expressed as fusion proteins with C-terminal GFP localized to small dot-like structures within *P. tricornutum*, most likely representing peroxisomal structures. Please see supplementary figures S3–S5, Supplementary Material online, for further information about observed localization patterns of *G. theta* peroxins in *P. tricornutum* and the effect of GFP positioning on targeting. TL, transmitted light; PAF, plastid autofluorescence; GFP, green fluorescent protein; Merge, overlay of PAF and GFP.

Interestingly, when Pex7 is absent and a potential Pex7-binding domain is lacking (or nonfunctional/degenerate) in Pex5, as it is the case for diatom and haptophyte Pex5 sequences, the proteins also seem to lack putative Pex14 interaction sites (WxxxF/Y motifs). The latter are highly conserved in other Pex5 primary sequences of organisms with complex plastids and other eukaryotes (fig. 1*B*). As shown in table 1, the lack of these binding motifs in Pex5 moreover correlates with the absence of Pex7 and Pex14 in case of diatoms and haptophytes.

Targeting Prediction of *G. theta* and *P. tricornutum* Peroxins

In preparation of in vivo localization studies, the primary sequences of the identified set of peroxins in *G. theta* given in table 1 were analyzed for the presence of specific targeting signals as described in the Materials and Methods section. We found that for most of the putative soluble (cytosolic) peroxins (Pex1, Pex6, Pex4, Pex5, Pex7, and Pex19) no obvious targeting signal or TMD was predicted (see supplementary table S2, Supplementary Material online). The only exception was Pex19-1, one of the two isoforms of the peroxisomal membrane protein receptor in *G. theta* (see table 1), for which a putative signal peptide or transmembrane domain was detected at the N-terminus. All remaining components of the peroxisomal importomer (Pex2, Pex10, Pex12, Pex14, Pex22, Pex3, and Pex16) as well as the division machinery (Pex11) are known to be membrane integral, which could be mostly confirmed via our prediction pipeline (see supplementary table S2, Supplementary Material online). Several peroxins seem to possess putative Pex19 binding sites (Pex2, Pex11-1, Pex12, and Pex16). Whereas the majority of the potentially membrane-integral peroxins of *G. theta* lacked an N-terminal targeting signal prediction, a putative mitochondrial targeting peptide (mTP) was predicted for the Pex3 candidate whereas a potential signal peptide was identified for Pex22 (see supplementary table S2, Supplementary Material online).

Similar to the *G. theta* peroxins, Pex16 from *P. tricornutum* was analyzed for the presence of potential targeting signals. With exception of a putative TMD identified via TOPCONS (rather weak prediction) no further relevant elements for targeting were detected within the Pt_Pex16 primary sequence (see supplementary table S2, Supplementary Material online).

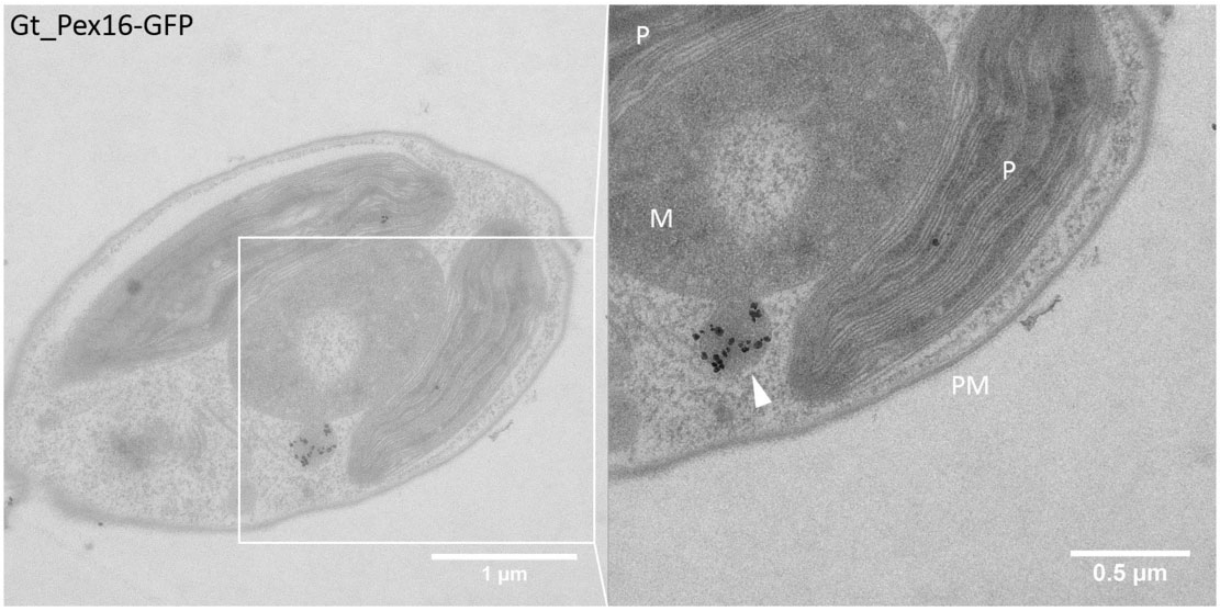

Fig. 4.—Transmission electron microscopic investigation of the *Guillardia theta* Pex16-GFP localization in *Phaeodactylum tricornutum*. Cross-section of a *P. tricornutum* cell expressing Gt_Pex16-GFP. The white arrowhead indicates a specifically immunogold-labeled electron dense structure in proximity to the nucleus and complex plastid, most likely representing a peroxisome (see fig. 2*B* for comparison). Gt, *Guillardia theta*; M, mitochondrion; P, plastid; PM, plasma membrane.

Localization Studies of *P. tricornutum* and *G. theta* Peroxins

Although the cryptophyte is an outstanding model system to study the evolution of complex red plastids via secondary endosymbiosis as well as mechanisms of eukaryotic compartmentalization, one drawback of working with *G. theta* is that methods for genetic manipulation are not available. As reported earlier, the diatom *P. tricornutum* is a suitable model system to perform heterologous protein localization studies of *G. theta* proteins because the two organisms share a similar cellular architecture/compartmentalization (Gould et al. 2006; Gile et al. 2015). Thus, as a control for the heterologous localization studies of *G. theta* peroxins in *P. tricornutum*, we analyzed the targeting and localization of the so far not localized peroxin Pt_Pex16 in the diatom *P. tricornutum* via in vivo localization studies of GFP fusion proteins. Figure 2*A* shows the results of the expression of Pt_Pex16 with GFP fused to its C-terminus visualized via confocal microscopy. GFP fluorescence concentrates in small, dot-like structures in direct vicinity to, but not overlaying with, the autofluorescence of the complex plastid, which is typical for a peroxisomal localization in *P. tricornutum* (Gonzalez et al. 2011). In addition, we have analyzed a Pex16-GFP expressing *P. tricornutum* clone via transmission electron microscopy (TEM) using immunogold-labeling on ultrathin sections. As figure 2*B* shows, gold particles specifically mark two circularly shaped electron dense structures with ~200–400 nm in diameter, most likely representing *P. tricornutum* peroxisomes, very similar to the observation made by McCarthy et al. (2017).

Similar to the experiments performed for *P. tricornutum* Pex16, we then wanted to study the localization of peroxins detected in the genomic data of the cryptophyte *G. theta*. Thus, we have generated fusion constructs of *G. theta pex* genes and the gene encoding the enhanced green fluorescent protein (*egfp*) and expressed them in the diatom. As shown by figure 3, the majority of heterologously expressed *G. theta* peroxin-GFP fusion proteins localized to one or more small dot-like structures that were located mostly in close proximity to the plastid of *P. tricornutum*—among these the predicted membrane proteins Pex2, Pex10, Pex11-1 and Pex11-2, Pex12, Pex14, Pex16, Pex19-1, and Pex22. A clear exception to this observation, as expected, was determined for the putative soluble PTS2 receptor Gt_Pex7-GFP, which showed a most likely cytosolic localization in the diatom (GFP fluorescence distributed throughout the cell; fig. 3). Sometimes, the Gt_Pex7-GFP fusion protein even concentrated in several most likely cytosolic dots in size relatively similar to labeled peroxisomal structures (supplementary fig. S3, Supplementary Material online). Especially, in case of membrane integral *G. theta* peroxins, the positioning of GFP at one of the termini had an effect on the localization of the fusion protein after heterologous expression in *P. tricornutum*. This was revealed for Gt_Pex10, which, when expressed with C-terminal GFP, showed a dot-like, most likely peroxisomal localization, whereas expression of the protein with an N-terminal GFP in front of the mature protein led to a distributed cytosolic GFP signal of the fusion construct in the diatom (supplementary fig. S4, Supplementary Material online).

Fig. 5.—Phylogenetic analysis of Pex4. The tree shown is the consensus tree obtained with Phylobayes 4.1 with ML bootstrap values (left) and Bayesian posterior probabilities (right) mapped onto the nodes. Bootstrap values >50% are shown, while only posterior probabilities >0.6 are indicated. The tree is midpoint rooted and the scale bar shows the inferred number of amino acid substitutions per site. Sequences are colored according to their taxonomic affiliation (see box).

It has to be noted that the typical dot-like peroxisomal localization pattern could not be observed for all *P. tricornutum* clones expressing Gt_Pex12-GFP and Gt_Pex22-GFP. Here, we often detected a labeling of most likely the ER/nuclear envelope and plasma membrane instead of a clear and clean peroxisomal GFP localization (supplementary fig. S5, Supplementary Material online). Moreover, a weak cytosolic signal was sometimes observed for Gt_Pex22-GFP in addition to concentrated, typical peroxisomal fluorescence spots (fig. 3).

In addition to the confocal microscopy experiments, we have analyzed the localization of a representative *G. theta* peroxin fused to GFP—Gt_Pex16-GFP—via TEM. Figure 4 shows that the obtained immunogold-labeling was restricted to a small (200–400 nm in diameter) electron dense circular structure near the mitochondrion and the complex plastid very similar in structure, size, and labeling to the results obtained for Pt_Pex16-GFP (see fig. 2*B*) and work recently published by McCarthy et al. (2017).

Fig. 6.—Phylogenetic analysis of Pex5. The tree shown is the consensus tree obtained with Phylobayes 4.1 with ML bootstrap values (left) and Bayesian posterior probabilities (right) mapped onto the nodes. Bootstrap values >50% are shown, while only posterior probabilities >0.6 are indicated. The tree is midpoint rooted and the scale bar shows the inferred number of amino acid substitutions per site. Sequences are colored according to their taxonomic affiliation (see box).

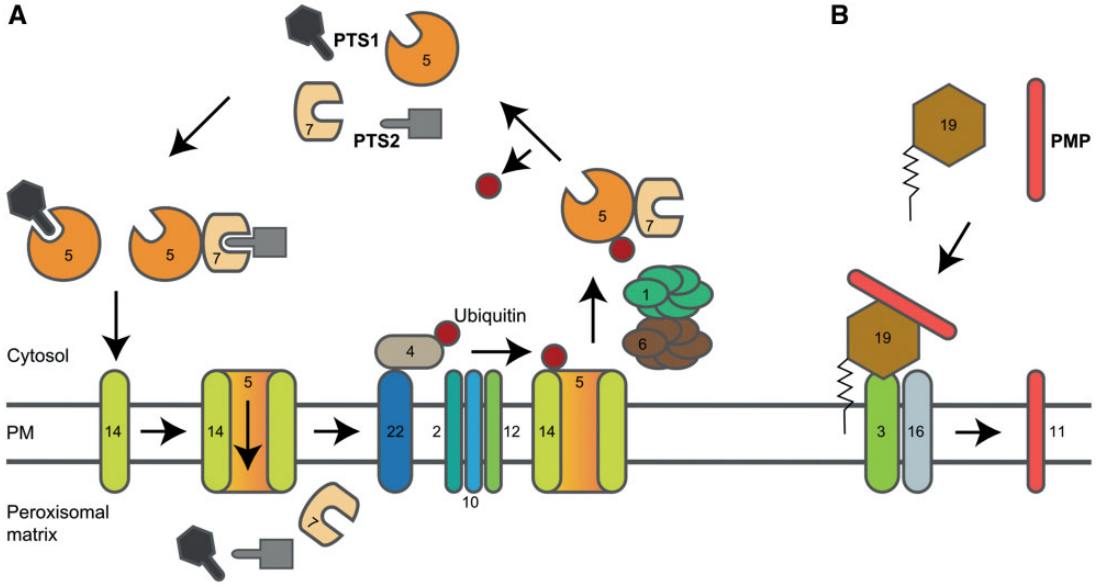

Fig. 7.—Model for peroxisomal protein import in *Guillardia theta* and other "chromalveolates." (*A*) Soluble peroxisomal PTS1 or PTS2 proteins are recognized by receptor proteins—Pex5 or Pex7—in the cytoplasm. Whereas the Pex5-PTS1-cargo-complex is able to bind the docking complex (Pex14) by means of internal Pex14 interaction sites (WxxxF/Y) to generate a transient pore to transfer the cargo into the peroxisomal matrix, Pex7 requires a coreceptor (Pex5 including a Pex7-binding domain) for cargo transport. After release of the cargo into the peroxisomal matrix, Pex5 is ubiquitinated by the RING complex, consisting of the E3 ubiquitin ligases Pex2, Pex10, and Pex12 together with the ubiquitin conjugating enzyme Pex4 (which might also play further roles aside from peroxisomal protein import) and its membrane anchor Pex22, and extracted into the cytoplasm. The latter step is energy-dependent and performed by the action of the two homo-hexameric AAA-ATPases Pex1 and Pex6. Back in the cytoplasm, the PTS1 receptor Pex5 is either degraded by the proteasome (polyubiquitinated Pex5, not shown), or deubiquitinated (monoubiquitinated Pex5) and recycled to enter a new round of peroxisomal protein import. Note that some "chromalveolates" (diatoms and haptophytes) do not contain a PTS2 import pathway (Pex7 absent, Pex14 lacking as well or highly derived). (*B*) Model for peroxisomal membrane protein import: The receptor for membrane proteins of peroxisomes might bind its substrates (PMP) in the cytoplasm to transfer the proteins into the peroxisomal membrane by an unknown mechanism after docking to the internal membrane protein Pex3 (and perhaps Pex16, membrane integration unclear). The farnesylation of *G. theta* and/or "chromalveolate" Pex19 is hypothetical. Numbers indicate peroxins (Pex). Note that *G. theta* isoforms of Pex19 and Pex11 were omitted from the model for simplicity. Abbreviations: PTS1/2, peroxisomal targeting signal type 1/2; PM, peroxisomal membrane; PMP, peroxisomal membrane protein. Parts of the figure were adapted from Kim and Hettema (2015).

Phylogeny of "Chromalveolate" Peroxins Pex4 and Pex5

Next, we wanted to assess the phylogenetic history of at least a subset of the detected peroxins in *G. theta* and other chromalveolates to gain more knowledge about their evolutionary past. We used the MMETSP as well as the NCBI (nr) database as a basis for our calculations (see Materials and Methods). Particularly, the MMETSP database is a resource of transcriptome data for many organisms for which a genome sequence is not yet present (Keeling et al. 2014). However, since the MMETSP data is not free of contamination, the results have to be evaluated with caution.

Two peroxins were in the focus of our phylogenetic analyses: (i) the putative receptor of the prevailing PTS1 import pathway, Pex5, which was identified in all major "chromalveolate" groups (table 1), and (ii) the potential ubiquitin conjugating E2 enzyme Pex4, which is present in all "chromalveolates" investigated in this study, even in those most likely lacking peroxisomes (noncoccidian apicomplexan parasites such as *Plasmodium*), making this factor particularly interesting. For the results of our inferred Pex4 phylogeny, although branching topologies often lack a clear statistical support, we would like to highlight three observations. First, Pex4 proteins of *G. theta* and other cryptophytes group together with sequences of Viridiplantae and kinetoplastids (Kinetoplastida, Euglenozoa) (fig. 5). Second, Pex4 sequences of organisms apparently lacking peroxisomes, such as the apicomplexans *Plasmodium*, *Cryptosporidium*, and *Gregarina*, group together with Pex4 of peroxisome-containing apicomplexans (coccidians) as well as other alveolates including chromerids, dinoflagellates (low posterior probabilities of 0.93), but not the included ciliates (fig. 5). Third, Pex4 of the *A. anophagefferens*, for which only a small subset of peroxisomal biogenesis factors could be detected (table 1), and Pex4 sequences of other pelagophytes are positioned within a subclade separated from other stramenopiles and ciliate Pex4 sequences, although stramenopiles, alveolates, and rhizarians (SAR) show a common phylogenetic affiliation and form a clade, which also contains haptophytes (fig. 5).

With respect to the phylogeny for Pex5, figure 6 suggests that cryptophyte and haptophyte sequences are sisters,

however with statistical support under the accepted threshold of 0.95 posterior probabilities (pp = 0.86). The latter clade groups together with a mixed clade consisting of alveolates on the one side and green algae and plants, red algae, amoebozoans, and rhizarians on the other side, again with lacking statistical support for the different nodes. It has to be noted that stramenopile Pex5 sequences form a separate clade at the base of the former two described ones (low posterior probabilities of 0.78; fig. 6).

Discussion

Organisms with complex red plastids, together termed "chromalveolates" (Cavalier-Smith 1999), comprise many ecologically important and medically relevant life forms including photosynthetic algae and parasitic protists. Cryptophytes, from an evolutionary standpoint, are one of the most interesting groups among these, because their secondary plastids still contain the remnant of the former red-algal endosymbiont nucleus (nucleomorph) (Curtis et al. 2012). Although peroxisomes are metabolically highly variable, single membrane-bound subcellular compartments with huge significance to the cellular metabolism, knowledge about these organelles in cryptophytes is absent. For some other "chromalveolates," it is at least known that peroxisomes exist, although information about their metabolic capabilities is scarce.

Here, we set out to explore the presence of peroxisomes in the cryptophyte *G. theta* and other "chromalveolates" via an initial bioinformatic screening for peroxins—specific factors for peroxisome biogenesis—followed by subcellular localization studies. We detected a total of 16 peroxins in *G. theta* most likely adequate to generate and maintain fully functional peroxisomes in the cryptophyte. All detected peroxins possess the expected functional domains known from human, yeast, and plant orthologs (NCBI CDS; e.g., see fig. 1 and supplementary fig. S1, Supplementary Material online). Among these 16 peroxins are two isoforms for the division factor Pex11 and PMP class I receptor Pex19, respectively. These factor duplications do not seem to be unusual and have been reported, for example, for the plant *Arabidopsis thaliana*, which contains five Pex11 and two Pex19 isoforms, and other eukaryotes (Orth et al. 2007; McDonnell et al. 2016). In contrast to diatoms, *G. theta* possesses the gene encoding the PTS2 receptor Pex7 and thus most likely operates a PTS2-dependent matrix protein import. This deduction is supported by the fact that the potential Pex7 coreceptor Pex5 of the cryptophyte harbors a conserved Pex7-binding domain that aligns with known human and plant Pex5(L) motifs (fig. 1*A*) essential for the binding of and interaction with Pex7 during PTS2-mediated peroxisomal protein import (Matsumura et al. 2000; Dodt et al. 2001; Woodward and Bartel 2005; Galland and Michels 2010). Based on the findings of our in silico analyses, a mitochondrial targeting peptide was identified for the putative Pex3 of *G. theta* (supplementary table S2, Supplementary Material online). This is highly interesting as recent studies in mammals suggest an involvement of mitochondria-derived outer membrane vesicles (in addition to ER-derived ones) in peroxisome biogenesis (Sugiura et al. 2017). Although nothing is known about this process in "chromalveolates," these protists might constitute a valuable source for the study of peroxisome biogenesis and evolution, especially with respect to a potential mitochondrial contribution. However, as we have so far not been able to amplify the putative *pex3* gene via PCR from *G. theta* DNA, experimental information on the localization of this factor is absent. Future studies will address the actual localization of Gt_Pex3 and reveal a potential involvement of the mitochondrion in peroxisomal biogenesis in *G. theta* and other "chromalveolates."

Due to the lack of methods for genomic manipulation of *G. theta*, we have used the diatom *P. tricornutum* as a surrogate in order to determine the subcellular localization of cyrptophyte peroxins (see above). Heterologous localization studies with Gt_Pex7 fused to GFP (C-terminally) in *P. tricornutum* revealed that the fusion protein localizes to the cytosol. This observation was expected for the PTS2 receptor, which is supposed to enter the peroxisome only transiently during PTS2-mediated protein import and otherwise is a cytosolic protein (Nair et al. 2004; Rodrigues et al. 2015). Sometimes overexpressed Gt_Pex7-GFP even concentrated in several cytosolic spots, most likely representing inclusion bodies, in *P. tricornutum* (supplementary fig. S3, Supplementary Material online) probably due to the high protein–protein interaction potential derived from the many WD40 domains of Pex7 (see supplementary fig. S1, Supplementary Material online).

In contrast, the majority of the other localized *G. theta* peroxin-GFP constructs resulted in one or more small, dot-like fluorescence patterns typical for a peroxisomal localization in *P. tricornutum* (Gonzalez et al. 2011). Exceptions were Pex12-GFP and Pex22-GFP from *G. theta* for which at least some clones showed a localization of the fusion proteins to the ER/nuclear envelope and plasma membrane, respectively (see supplementary fig. S5, Supplementary Material online). These observations are probably due to the use of a heterologous system for localization, strong overexpression, or the possibility that these components have adapted some additional functions in subcellular compartments beside, or in addition to, the peroxisome. The latter might be particularly the case for Gt_Pex22 for which prediction of an N-terminal signal peptide was relatively high (see supplementary table S2, Supplementary Material online). Supplemental to these findings, our localization studies revealed that positioning of the GFP moiety has a significant effect on the targeting of a fusion construct (supplementary fig. S4, Supplementary Material online). This might be especially important in case of potential membrane integral peroxins, such as Pex10, for which a putative peroxisomal localization was observed when expressed

as fusion protein with C-terminal GFP (fig. 3 and supplementary fig. S4, Supplementary Material online). In contrast, positioning of GFP at the N-terminus of Gt_Pex10 and expression in *P. tricornutum* led to a localization of the fusion construct in the cytoplasm of the diatom (supplementary fig. S4, Supplementary Material online). This effect can possibly be explained by a sterical hindrance of GFP either preventing binding of relevant receptor/transport proteins or disrupting a peroxisomal membrane integration of the fusion construct in general. To verify the confocal microscopic data for peroxisomal localizations of heterologously expressed *G. theta* peroxins fused to GFP in *P. tricornutum* (see above), we have compared them to the localization results for the so far not localized Pex16 from *P. tricornutum* (overexpressed endogenous vs foreign protein). As our localization studies with Pt_Pex16-GFP revealed (see fig. 2) the labeling pattern observed via confocal microscopy and via TEM is similar to the results obtained for the *G. theta* peroxin-GFP fusion proteins expressed and localized in the diatom system (figs. 2–4). Particularly, the TEM results for the two *P. tricornutum* and *G. theta* Pex16-GFP localizations are highly similar and are also in accordance to already published work (McCarthy et al. 2017). Here, the labeling of small electron dense circular structures, most likely representing peroxisomes, were clearly visualized with the TEM and nicely correlate with the confocal microscopy results for the diatom and cryptophyte Pex16-GFP constructs. Thus, *P. tricornutum* is a suitable model organism for localization studies of *G. theta* peroxins and, most likely, other peroxisomal proteins fused to GFP in vivo. Moreover, these results suggest that Pex16 is exclusively localized to the peroxisome in *P. tricornutum* and *G. theta* because it was not detectable in the ER (membrane). Therefore, a function of Pex16 in recruitment of peroxisomal membrane proteins to both, the ER and peroxisomes, as has been reported for human and plant Pex16 factors (Kim and Mullen 2013), is rather unlikely for Gt_Pex16 and Pt_Pex16. It also remains unclear whether Pex16 in *G. theta* and *P. tricornutum* is a class II peroxisomal membrane protein inserted in the ER, or rather a class I PMP that is directly imported via Pex19 from the cytosol into peroxisomes, or whether it is an integral membrane protein at all. At least in the yeast *Yarrowia lipolytica* Pex16 seems to be an intraperoxisomal peripheral membrane protein with a role in peroxisomal fission (Kim and Mullen 2013). Although there are some more or less reliable in silico predictions for the presence of putative TMDs (supplementary table S2, Supplementary Material online) for both proteins, it is possible that a similar situation applies to Pex16 of *P. tricornutum* and *G. theta*.

In addition to the cryptophyte peroxisomal biogenesis factors, we found in silico evidence for the presence of peroxins in the haptophytes *E. huxleyi* and *C. tobin*, although the list was much more incomplete when compared with *G. theta*. In addition, the results suggest that a peroxisomal targeting of soluble proteins via PTS2 is, in contrast to cryptophytes but similar to diatoms, absent in haptophytes due to the lack of a Pex7 encoding gene (table 1). Moreover, peroxins have been identified in stramenopiles, including diatoms, the eustigmatophyte *Nannochloropsis* and the brown alga *Ectocarpus*. For the latter two, the list of detected peroxins was almost as comprehensive as for the cryptophyte, whereas for diatoms certain Pex factors seemed to be absent, or have so far not been identified (Gonzalez et al. 2011; Gabaldon et al. 2016; Gentekaki et al. 2017). We reinvestigated the genomes of two diatoms, *P. tricornutum* and *T. pseudonana*, for *pex* genes and present the thitherto most complete list of peroxins, which might play an important role in peroxisomal biogenesis and maintenance in these organisms. As peroxins are supposed to be specific to peroxisomal cell biology, their genomic presence can be considered as an indicator for the existence of peroxisomes in a certain organism (Gabaldon et al. 2006; Schlüter et al. 2006; Gabaldon 2010). Thus, our results suggest that peroxisomes are indeed present in the mentioned crytophytes, haptophytes, and stramenopiles.

We also screened the genome of the pelagophyte *A. anophagefferens* as another representative of stramenopiles with complex plastids. The resulting set of only three potential peroxin candidates (Pex4, Pex10, Pex19), however, makes it difficult to estimate whether peroxisomes are present or not in *A. anophagefferens*. An explanation for the scattered list of peroxins in the pelagophyte could be that the genes encoding these factors are too divergent to be detected by our screening or that the genome data of *A. anophagefferens* is fragmentary. Another reason for this observation might be that the three detected peroxins are the evolutionary remains of peroxisomal biogenesis factors documenting the former presence of peroxisomes in a pelagophyte ancestor, but these structures might have been lost in *A. anophagefferens*. The gene products could have acquired new functions, similar to the situation in nonperoxisome bearing apicomplexans, which contain Pex4 and Pex22 (Moog et al. 2017; Ludewig-Klingner et al. 2018), from which at least the first one is also present in *A. anophagefferens*. However, the identification of a Pex19 candidate, a protein specific to the import mechanisms of peroxisomal membrane proteins of class I, if not a genomic contamination, is an indicator for the existence of peroxisomes in the pelagophyte or at least that the ancestors of *A. anophagefferens* probably maintained such organelles in the past.

Together with the data obtained for alveolates, which show that a comprehensive set of peroxins is present in dinoflagellates (*S. microadriaticum* most likely has a PTS1-, PTS2-, and mPTS-mediated import pathway; table 1; Ludewig-Klingner et al. 2018), photosynthetic chromerids, coccidians but most likely no other apicomplexans (e.g., *Plasmodium*, *Cryptopsoridium*; Moog et al. 2017), our results suggest that peroxisomes are generally present in "chromalveolate" groups. One peculiarity that our bioinformatic results imply is that diatoms seem to be not the only group of peroxisome-

containing "chromalveolates" that do not possess a PTS2 import pathway (Gonzalez et al. 2011). The PTS2 receptor Pex7 also seems to be absent in haptophytes (table 1). The lack of a peroxisomal PTS2 import pathway in diatoms is certainly no singularity as also nematodes, such as *Caenorhabditis elegans*, the fruit fly *Drosophila* and probably the red alga *Cyanidioschyzon merolae*—organisms from different branches of the eukaryotic tree—have been suggested to have lost this pathway (Motley et al. 2000; Shinozaki et al. 2009; Gonzalez et al. 2011; Faust et al. 2012; Zarsky and Tachezy 2015; Reumann et al. 2016). Therefore, with the haptophyte *E. huxleyi* (and most likely haptophytes in general) our study revealed a further important entry on the list of PTS2 pathway-lacking organisms.

Interestingly, no conserved Pex7-binding site could be detected in Pex5 proteins of Pex7-lacking "chromalveolates" (diatoms and haptophytes). Likewise no Pex14 interaction sites (WxxxF/Y) were identified in the PTS1 receptor primary sequences of these organisms, while also Pex14 was lacking or at least was not detectable (table 1). This observation pattern suggests that "chromalveolates" devoid of a PTS2 import pathway, that is Pex7, may also lack Pex14 or have replaced the central docking complex component by another more divergent one. The situation is in contrast to other PTS2 pathway lacking organisms such as *C. elegans*, *C. merolae* (Pex7 absent), and potentially *Drosophila* (contains a Pex7 homolog), which contain homologs of Pex14 and maintain conserved Pex14 interaction sites (WxxxF/Y) within their PTS1 receptor (Pex5) primary sequences, respectively (see supplementary material S3, Supplementary Material online). The concomitant lack of Pex7 and Pex14 in "chromalveolates" might have implications on the Pex5-mediated PTS1 protein import pathway, as it is hypothesized that the PTS1 receptor builds up a transient pore together with Pex14 in the peroxisomal membrane during matrix protein import (Erdmann and Schliebs 2005; Meinecke et al. 2016).

In contrast to diatoms, other stramenopiles including eustigmatophytes and brown algae (*Ectocarpus*) seem to possess Pex7 and therefore most likely a PTS2-dependent import pathway. Notably, these organisms also harbor genes encoding Pex14. For *A. anophagefferens*, neither Pex7 nor Pex5 or Pex14 could be identified in the course of our genomic screening.

While the absence of Pex14 seems to be specific for those "chromalveolates" that do not operate a PTS2-dependent matrix protein import mechanism, Pex13, another component of the peroxisomal docking complex known from other eukaryotes including plants, yeast, and animals (Smith and Aitchison 2013; Cross et al. 2016), could not be identified in any of the here investigated species (table 1). It is thus very likely that "chromalveolates" have entirely lost Pex13 or replaced the factor by another, so far unknown component—a finding that has already been indicated by us and others when studying peroxins in alveolates (Moog et al. 2017; Ludewig-Klingner et al. 2018).

Another peculiarity that emerged during this and earlier studies was that the peroxin Pex4 is present in all investigated "chromalveolates" regardless whether they contain peroxisomes or not. This is not only the case in peroxisome-lacking apicomplexans such as, for example, *Plasmodium* and *Cryptopsoridium* (Moog et al. 2017; Ludewig-Klingner et al. 2018) but maybe also in the stramenopile *A. anophagefferens* (see above). In addition to that, another stramenopile, the parasite *Blastocystis*, contains a gene encoding a putative Pex4 homolog in its genome although further genes for peroxisomal biogenesis and maintenance seem to be absent (Gentekaki et al. 2017). It is thus very likely that Pex4 plays not only a role in ubiquitination during peroxisomal biogenesis and maintenance but also has an important function (most likely essential) as a potential ubiquitin conjugating enzyme aside from peroxisome biology in these organisms. One lead into this direction might be that the Pex4 membrane anchor Pex22 from *G. theta* when expressed as a GFP fusion protein in the diatom not only localized to peroxisomes but sometimes also to the ER/nuclear envelope and the plasma membrane (see above).

To analyze the evolutionary history of Pex4 proteins in "chromalveolates" with and without peroxisomes, we have inferred a phylogeny with the sequences detected by our in silico screening using Phylobayes (see Materials and Methods). Indeed our calculations suggest a common origin (monophyly) of Pex4 sequences of apicomplexans, chromerids, and dinoflagellates, which contain peroxisomes and those probably devoid of such organelles (fig. 5). Moreover, Pex4 of the pelagophyte *A. anophagefferens*, which is so far unclear to possess peroxisomes, groups with Pex4 proteins of other stramenopiles, including those of diatoms, *Ectocarpus* and *Nannochloropsis*, but also ciliates and rhizarians. These results support our hypothesis that Pex4 is a putative ubiquitin conjugating E2 enzyme that has adopted additional responsibilities before the loss of peroxisomes in certain "chromalveolates" (noncoccidian apicomplexans and perhaps *Aureococcus*) (see above). Studies addressing the role of Pex4 in "chromalveolates" in more detail are required in the future.

As for the cryptophyte Pex4 sequences, we have observed a positioning within members of the Viridiplantae, which group together with Pex4 proteins of Euglenozoa (kinetoplastids). However, at least the close phylogenetic relationship of nucleus-encoded cryptophyte and Archaeplastida (here Viridiplantae) protein sequences has been observed repeatedly in recent phylogenies (Burki et al. 2016; Kim et al. 2017; Brown et al. 2018).

With respect to our second phylogenetic analysis, Pex5 of "chromalveolates," monophyly of cryptophyte and haptophyte Pex5 is suggested, whereas stramenopiles and alveolate Pex5 sequences group separately from each other. Although

the tree supports monophyly of all of the major individual "chromalveolate" groups, that is alveolates (ciliates, apicomplexans, chromerids, and dinoflagellates), cryptophytes, haptophytes, and stramenopiles, respectively, with the exception of the potential cryptophyte-haptophyte affiliation, there is no sign for a particular phylogenetic relationship of the host-specific Pex5 sequences of organisms with complex red plastids. Moreover, the phylogeny suggests that loss of the Pex7- and Pex14-binding/interaction sites in Pex5, as it has most likely been the case for the here included diatom, haptophyte, and *C. merolae* (only loss of the Pex7-binding motif) sequences, occurred several times independently in the eukaryotic tree of life (fig. 6).

As both of our phylogenies (figs. 5 and 6) represent single protein analyses, the number of phylogenetically informative positions included as a basis for inference has been restricted (125 for Pex4 and 217 for Pex5, respectively). Thus, such analyses are particularly prone to phylogenetic signal erosion, often leading to rather low statistical support values at individual nodes, as could be seen in case of those at the split of several major taxa included here (figs. 5 and 6). Although determining the exact phylogenetic position of an individual clade might be difficult then based on our analyses, the trees are still informative with respect to the clade–internal relationships and evolutionary histories (e.g., see cryptophyte Pex4 clade in fig. 5, or cryptophyte as well as stramenopile Pex5 clades in fig. 6, respectively).

We would like to summarize the results of our present and earlier studies and include them into a model for peroxisomal protein import in *G. theta* and other "chromalveolates" in which soluble PTS1- or PTS2-bearing peroxisomal matrix proteins as well as PMPs of class I are translocated (fig. 7). Pex7-lacking "chromalveolates" are devoid of a PTS2-mediated import pathway, harbor a shorter version of the PTS1 receptor Pex5 (without Pex7 and Pex14 binding/interaction sites) and may possess a more divergent Pex14 or replaced the docking complex factor by another so far unknown component.

Conclusions

Based on the identified peroxins—factors specific for peroxisome biogenesis, maintenance, and division—we conclude that "chromalveolates" are generally in possession of peroxisomes with some exceptions mainly comprising noncoccidian apicomplexans. Our results indicate that diatoms are not the only group of organisms devoid of PTS2-mediated peroxisomal protein import, but also the haptophyte *E. huxleyi* seems to lack Pex7 and thus a PTS2 pathway. In "chromalveolates" in which Pex7 is absent also Pex14 seems missing or is possibly highly derived, whereas Pex13 is generally not present in any of the investigated "chromalveolates." Furthermore, the PTS1 receptor Pex5 contains a highly conserved Pex7-binding domain in "chromalveolates" with Pex7, whereas this site is absent from Pex5 of organisms devoid of Pex7, in which the PTS1 receptor is also lacking any recognizable Pex14 binding sites. As our phylogeny suggests this loss has probably occurred several times independently within "chromalveolate" groups and the eukaryotic tree of life in general. In addition, our results suggest a significant role for the putative ubiquitin conjugating enzyme Pex4, probably even aside from peroxisomal mechanisms, in "chromalveolates" and provide important insights into the cell biology of peroxisomes and the phylogenetic history of certain peroxins in these organisms. Next, it will be important to study peroxisomal protein import in more detail and to analyze the metabolic capacities of peroxisomes in the diversity of "chromalveolates" and other eukaryotic organisms.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We would like to thank Uwe Maier for the broad support of the work, helpful discussions, and critically reading of the manuscript. We are grateful to Marion Debus for technical assistance with sample preparation for electron microscopy. We also like to thank Larissa Kolesnikova from the EM facility of the Institute of Virology, Philipps University Marburg, for access to their electron microscopic facility and for taking pictures for us, which are part of this work, during times we had technical issues with our system. This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—409750538.

Literature Cited

Apt KE, Kroth-Pancic PG, Grossman AR. 1996. Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. Mol Gen Genet. 252(5):572–579.

Aranda M, et al. 2016. Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. Sci Rep. 6:39734.

Archibald JM. 2015. Genomic perspectives on the birth and spread of plastids. Proc Natl Acad Sci U S A. 112(33):10147–10153.

Armbrust EV, et al. 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 306(5693):79–86.

Baker A, Lanyon-Hogg T, Warriner SL. 2016. Peroxisome protein import: a complex journey. Biochem Soc Trans. 44(3):783–789.

Baurain D, et al. 2010. Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. Mol Biol Evol. 27(7):1698–1709.

Bowler C, et al. 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature 456(7219):239–244.

Braverman N, Dodt G, Gould SJ, Valle D. 1998. An isoform of pex5p, the human PTS1 receptor, is required for the import of PTS2 proteins into peroxisomes. Hum Mol Genet. 7(8):1195–1205.

Brown MW, et al. 2018. Phylogenomics places orphan protistan lineages in a novel eukaryotic super-group. Genome Biol Evol. 10(2):427–433.

Burki F, et al. 2016. Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. Proc Biol Sci. 283:20152802.
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):347–366.
Cavalier-Smith T. 2000. Membrane heredity and early chloroplast evolution. Trends Plant Sci. 5(4):174–182.
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):109–133. discussion 133–104.
Cock JM, et al. 2010. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. Nature 465(7298):617–621.
Corteggiani Carpinelli E, et al. 2014. Chromosome scale genome assembly and transcriptome profiling of *Nannochloropsis gaditana* in nitrogen depletion. Mol Plant 7(2):323–335.
Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol. 10:210.
Cross LL, Ebeed HT, Baker A. 2016. Peroxisome biogenesis, protein targeting mechanisms and PEX gene functions in plants. Biochim Biophys Acta 1863(5):850–862.
Curtis BA, et al. 2012. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature 492(7427):59–65.
Deb R, Nagotu S. 2017. Versatility of peroxisomes: an evolving concept. Tissue Cell 49(2 Pt B):209–226.
Dias AF, et al. 2017. The peroxisomal matrix protein translocon is a large cavity-forming protein assembly into which PEX5 protein enters to release its cargo. J Biol Chem. 292(37):15287–15300.
Dodt G, Warren D, Becker E, Rehling P, Gould SJ. 2001. Domain mapping of human PEX5 reveals functional and structural similarities to *Saccharomyces cerevisiae* Pex18p and Pex21p. J Biol Chem. 276(45):41769–41781.
Eddy SR. 1998. Profile hidden Markov models. Bioinformatics 14(9):755–763.
Erdmann R. 2016. Assembly, maintenance and dynamics of peroxisomes. Biochim Biophys Acta 1863(5):787–789.
Erdmann R, Schliebs W. 2005. Peroxisomal matrix protein import: the transient pore model. Nat Rev Mol Cell Biol. 6(9):738–742.
Faust JE, Verma A, Peng C, McNew JA. 2012. An inventory of peroxisomal proteins and pathways in *Drosophila melanogaster*. Traffic 13(10):1378–1392.
Francisco T, et al. 2014. Ubiquitin in the peroxisomal protein import pathway. Biochimie 98:29–35.
Fransen M, Lismont C, Walton P. 2017. The peroxisome-mitochondria connection. How and why? Int J Mol Sci. 18(6):1126.
Fujiki Y, Matsuzono Y, Matsuzaki T, Fransen M. 2006. Import of peroxisomal membrane proteins: the interplay of Pex3p- and Pex19p-mediated interactions. Biochim Biophys Acta 1763(12):1639–1646.
Gabaldon T. 2010. Peroxisome diversity and evolution. Philos Trans R Soc Lond B Biol Sci. 365(1541):765–773.
Gabaldon T, et al. 2006. Origin and evolution of the peroxisomal proteome. Biol Direct 1:8.
Gabaldon T, Ginger ML, Michels PA. 2016. Peroxisomes in parasitic protists. Mol Biochem Parasitol. 209(1–2):35–45.
Galland N, Michels PA. 2010. Comparison of the peroxisomal matrix protein import system of different organisms. Exploration of possibilities for developing inhibitors of the import system of trypanosomatids for anti-parasite chemotherapy. Eur J Cell Biol. 89(9):621–637.
Gentekaki E, et al. 2017. Extreme genome diversity in the hyper-prevalent parasitic eukaryote *Blastocystis*. PLoS Biol. 15(9):e2003769.
Gentil J, Hempel F, Moog D, Zauner S, Maier UG. 2017. Review: origin of complex algae by secondary endosymbiosis: a journey through time. Protoplasma 254(5):1835–1843.
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(6):1728–1742.
Gobler CJ, et al. 2011. Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. Proc Natl Acad Sci U S A. 108(11):4352–4357.
Gonzalez NH, et al. 2011. A single peroxisomal targeting signal mediates matrix protein import in diatoms. PLoS One 6(9):e25316.
Gould SB. 2018. Membranes and evolution. Curr Biol. 28(8):R381–R385.
Gould SB, et al. 2006. Nucleus-to-nucleus gene transfer and protein retargeting into a remnant cytoplasm of cryptophytes and diatoms. Mol Biol Evol. 23(12):2413–2422.
Hettema EH, Erdmann R, van der Klei I, Veenhuis M. 2014. Evolving models for peroxisome biogenesis. Curr Opin Cell Biol. 29:25–30.
Hovde BT, et al. 2015. Genome sequence and transcriptome analyses of *Chrysochromulina tobin*: metabolic tools for enhanced algal fitness in the prominent order prymnesiales (Haptophyceae). PLoS Genet. 11(9):e1005469.
Hu J, et al. 2012. Plant peroxisomes: biogenesis and function. Plant Cell 24(6):2279–2303.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.
Keeling PJ. 2010. The endosymbiotic origin, diversification and fate of plastids. Philos Trans R Soc Lond B Biol Sci. 365(1541):729–748.
Keeling PJ, et al. 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol. 12(6):e1001889.
Kim JI, et al. 2017. Evolutionary dynamics of cryptophyte plastid genomes. Genome Biol Evol. 9(7):1859–1872.
Kim PK, Hettema EH. 2015. Multiple pathways for protein transport to peroxisomes. J Mol Biol. 427(6 Pt A):1176–1190.
Kim PK, Mullen RT. 2013. PEX16: a multifaceted regulator of peroxisome biogenesis. Front Physiol. 4:241.
Kunze M, et al. 2015. Mechanistic insights into PTS2-mediated peroxisomal protein import: the co-receptor PEX5L drastically increases the interaction strength between the cargo protein and the receptor PEX7. J Biol Chem. 290(8):4928–4940.
Lametschwandtner G, et al. 1998. The difference in recognition of terminal tripeptides as peroxisomal targeting signal 1 between yeast and human is due to different affinities of their receptor Pex5p to the cognate signal and to residues adjacent to it. J Biol Chem. 273(50):33635–33643.
Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25(17):2286–2288.
Le SQ, Dang CC, Gascuel O. 2012. Modeling protein evolution with several amino acid replacement matrices depending on site rates. Mol Biol Evol. 29(10):2921–2936.
Liu X, et al. 2016. Addressing various compartments of the diatom model organism *Phaeodactylum tricornutum* via sub-cellular marker proteins. Algal Res. 20:249–257.
Ludewig-Klingner AK, Michael V, Jarek M, Brinkmann H, Petersen J. 2018. Distribution and evolution of peroxisomes in alveolates (Apicomplexa, dinoflagellates, ciliates). Genome Biol Evol. 10(1):1–3543.
Maier UG, Zauner S, Hempel F. 2015. Protein import into complex plastids: cellular organization of higher complexity. Eur J Cell Biol. 94(7–9):340–348.

Maruyama S, Eveleigh RJ, Archibald JM. 2013. Treetrimmer: a method for phylogenetic dataset size reduction. BMC Res Notes 6:145.
Matsumura T, Otera H, Fujiki Y. 2000. Disruption of the interaction of the longer isoform of Pex5p, Pex5pL, with Pex7p abolishes peroxisome targeting signal type 2 protein import in mammals. Study with a novel Pex5-impaired Chinese hamster ovary cell mutant. J Biol Chem. 275(28):21715–21721.
Mayerhofer PU. 2016. Targeting and insertion of peroxisomal membrane proteins: ER trafficking versus direct delivery to peroxisomes. Biochim Biophys Acta 1863(5):870–880.
McCarthy JK, et al. 2017. Nitrate reductase knockout uncouples nitrate transport from nitrate assimilation and drives repartitioning of carbon flux in a model pennate diatom. Plant Cell 29(8):2047–2070.
McDonnell MM, et al. 2016. The early-acting peroxin PEX19 is redundantly encoded, farnesylated, and essential for viability in *Arabidopsis thaliana*. PLoS One 11(1):e0148335.
Meinecke M, Bartsch P, Wagner R. 2016. Peroxisomal protein import pores. Biochim Biophys Acta 1863(5):821–827.
Meinecke M, et al. 2010. The peroxisomal importomer constitutes a large and highly dynamic pore. Nat Cell Biol. 12(3):273–277.
Montilla-Martinez M, et al. 2015. Distinct pores for peroxisomal import of PTS1 and PTS2 proteins. Cell Rep. 13(10):2126–2134.
Moog D, Maier UG. 2017. Cellular compartmentation follows rules: the Schnepf theorem, its consequences and exceptions: a biological membrane separates a plasmatic from a non-plasmatic phase. Bioessays 39(8):1700030.
Moog D, Przyborski JM, Maier UG. 2017. Genomic and proteomic evidence for the presence of a peroxisome in the apicomplexan parasite *Toxoplasma gondii* and other Coccidia. Genome Biol Evol. 9(11):3108–3121.
Motley AM, Hettema EH, Ketting R, Plasterk R, Tabak HF. 2000. *Caenorhabditis elegans* has a single pathway to target matrix proteins to peroxisomes. EMBO Rep. 1(1):40–46.
Nair DM, Purdue PE, Lazarow PB. 2004. Pex7p translocates in and out of peroxisomes in *Saccharomyces cerevisiae*. J Cell Biol. 167(4):599–604.
Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32(1):268–274.
Orth T, et al. 2007. The PEROXIN11 protein family controls peroxisome proliferation in *Arabidopsis*. Plant Cell 19(1):333–350.
Petersen J, et al. 2014. *Chromera velia*, endosymbioses and the rhodoplex hypothesis–plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages). Genome Biol Evol. 6(3):666–684.
Pieuchot L, Jedd G. 2012. Peroxisome assembly and functional diversity in eukaryotic microorganisms. Annu Rev Microbiol. 66:237–263.
Platta HW, Erdmann R. 2007. The peroxisomal protein import machinery. FEBS Lett. 581(15):2811–2819.
Price MN, Dehal PS, Arkin AP. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One 5(3):e9490.
Read BA, et al. 2013. Pan genome of the phytoplankton *Emiliania* underpins its global distribution. Nature 499(7457):209–213.
Renicke C, Allmann AK, Lutz AP, Heimerl T, Taxis C. 2017. The mitotic exit network regulates spindle pole body selection during sporulation of *Saccharomyces cerevisiae*. Genetics 206(2):919–937.
Reumann S, Chowdhary G, Lingner T. 2016. Characterization, prediction and evolution of plant peroxisomal targeting signals type 1 (PTS1s). Biochim Biophys Acta 1863(5):790–803.
Rodrigues TA, Grou CP, Azevedo JE. 2015. Revisiting the intraperoxisomal pathway of mammalian PEX7. Sci Rep. 5:11806.
Rucktäschel R, Girzalsky W, Erdmann R. 2011. Protein import machineries of peroxisomes. Biochim Biophys Acta 1808(3):892–900.
Schlüter A, et al. 2006. The evolutionary origin of peroxisomes: an ER-peroxisome connection. Mol Biol Evol. 23(4):838–845.
Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9(7):671–675.
Shai N, Schuldiner M, Zalckvar E. 2016. No peroxisome is an island—peroxisome contact sites. Biochim Biophys Acta 1863(5):1061–1069.
Shinozaki A, Sato N, Hayashi Y. 2009. Peroxisomal targeting signals in green algae. Protoplasma 235(1–4):57–66.
Smith JJ, Aitchison JD. 2013. Peroxisomes take shape. Nat Rev Mol Cell Biol. 14(12):803–817.
Sommer MS, et al. 2007. Der1-mediated preprotein import into the periplastid compartment of chromalveolates? Mol Biol Evol. 24(4):918–928.
Stiller JW, et al. 2014. The evolution of photosynthesis in chromist algae through serial endosymbioses. Nat Commun. 5:5764.
Sugiura A, Mattie S, Prudent J, McBride HM. 2017. Newly born peroxisomes are a hybrid of mitochondrial and ER-derived pre-peroxisomes. Nature 542(7640):251–254.
Thoms S, Erdmann R. 2005. Dynamin-related proteins and Pex11 proteins in peroxisome division and proliferation. FEBS J. 272(20):5169–5181.
Vieler A, et al. 2012. Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga *Nannochloropsis oceanica* CCMP1779. PLoS Genet. 8:e1003064.
Walton PA, Hill PE, Subramani S. 1995. Import of stably folded proteins into peroxisomes. Mol Biol Cell 6(6):675–683.
Waterham HR, Ferdinandusse S, Wanders RJ. 2016. Human disorders of peroxisome metabolism and biogenesis. Biochim Biophys Acta 1863(5):922–933.
Woodward AW, Bartel B. 2005. The *Arabidopsis* peroxisomal targeting signal type 2 receptor PEX7 is necessary for peroxisome function and dependent on PEX5. Mol Biol Cell 16(2):573–583.
Zarsky V, Tachezy J. 2015. Evolutionary loss of peroxisomes—not limited to parasites. Biol Direct 10:74.

Associate editor: Bill Martin