The Known, the New, and a Possible Surprise: A Re-Evaluation of the Nucleomorph-Encoded Proteome of Cryptophytes

Stefan Zauner[1], Thomas Heimerl[2], Daniel Moog[1,2], and Uwe G. Maier[1,2,*]

[1]Department for Cell Biology, Philipps-Universität Marburg, Germany

[2]SYNMIKRO Research Center, Philipps-Universität Marburg, Germany

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

Accepted: May 20, 2019

Abstract

Nucleomorphs are small nuclei that evolved from the nucleus of former eukaryotic endosymbionts of cryptophytes and chlorarachniophytes. These enigmatic organelles reside in their complex plastids and harbor the smallest and most compacted eukaryotic genomes investigated so far. Although the coding capacity of the nucleomorph genomes is small, a significant percentage of the encoded proteins (predicted nucleomorph-encoded proteins, pNMPs) is still not functionally annotated. We have analyzed pNMPs with unknown functions via Phyre[2], a bioinformatic tool for prediction and modeling of protein structure, resulting in a functional annotation of 215 pNMPs out of 826 uncharacterized open reading frames of cryptophytes. The newly annotated proteins are predicted to participate in nucleomorph-specific functions such as chromosome organization and expression, as well as in modification and degradation of nucleomorph-encoded proteins. Additionally, we have functionally assigned nucleomorph-encoded, putatively plastid-targeted proteins among the reinvestigated pNMPs. Hints for a putative function in the periplastid compartment, the cytoplasm surrounding the nucleomorphs, emerge from the identification of pNMPs that might be homologs of endomembrane system-related proteins. These proteins are discussed in respect to their putative functions.

Key words: cryptomonad, nucleomorph, nucleomorph-encoded proteins, periplastid compartment, plastid.

Introduction

Nucleomorphs are small nuclei found in complex plastids of cryptophytes and chlorarachniophytes (Maier et al. 2000). The evolution of both organismal groups started with the uptake of a phototrophic eukaryote by another eukaryotic unicellular cell, respectively. This cellular combination of two eukaryotes led to the reduction of the phototrophic partner, from which the plastid including the plastid envelope as well as the reduced cytoplasm of the symbiont, namely the periplastid compartment (PPC), was retained and separated from the host cytoplasm by two membranes (Zauner et al. 2000). More precisely, in plastid-containing cryptophytes the outermost membrane of the complex plastid is continuous with the nuclear envelope of the host and termed chloroplast ER membrane, followed by the periplastid membrane (PPM), which is directly surrounding the PPC, whereas membranes III (OEM, plastid outer envelope membrane) and IV (IEM, plastid inner envelope membrane) are homologous to the plastid envelope of primary plastids, present, for example, in red or green algae. The PPC of cryptophytes and chlorarachniophytes is a very unusual compartment, as the reduced cell nucleus of the former symbiont, the nucleomorph, is still maintained, making both organismal groups unicellular organisms with two different cytoplasms and nuclei (Moore and Archibald 2009).

Generally, nucleomorph genomes are organized in three little chromosomes encoding less than 1,000 proteins, but not yet all of the encoded proteins could be assigned to a cellular function. For instance, in the first sequenced nucleomorph genome of the cryptophyte *Guillardia theta* (for an ultrastructural view of *G. theta*, see fig. 1) 205 putative open reading frames (ORFs) were identified by Douglas et al. (2001) with the encoding proteins having no predictable function. Although analyzed and published later, three further cryptophyte nucleomorph genome sequences (*Hemiselmis andersenii* CCMP644 [Lane et al. 2007], *Cryptomonas paramecium* [Tanifuji et al. 2011], *Chroomonas mesostigmatica* CCMP1168 [Moore et al. 2012]) also contain many uncharacterized ORFs. Moreover, to further categorize ORFs, Moore et al. (2012) suggested to use either the term "conserved ORF," a protein-coding gene with annotated homologs in other nuclear genomes, or nORFs, hypothetical protein-coding genes present in other nucleomorph genomes only,

or "ORFans," hypothetical protein-coding genes with no significant sequence similarity to any other known gene. Uncovering the functions of ORFans/nORFs-encoded proteins is of special interest and might highlight unknown key mechanisms for the coexistence and cooperation of two merged eukaryotic cells. Therefore, we reanalyzed nucleomorph-encoded proteins with unknown functions of the four cryptophytic nucleomorph genomes, thereby using knowledge on the functional domains and structures, which increased a lot since the first nucleomorph genome was published.

For our analyses, we applied the bioinformatic package provided by Phyre^2 (Kelley et al. 2015), which predicts the molecular structure of proteins, and compares it with already determined protein structures. According to our stringency criteria for interpreting the results obtained from Phyre^2 we could functionally assign the encoded proteins of 67 ORFans/nORFs (66 with unknown functions according to Douglas et al. [2001], and one new description, see later) from *G. theta* (fig. 2, supplementary table 1, Supplementary Material online) and altogether of 215 ORFans/nORFs from all four cryptophyte nucleomorph genomes. Consistent with the already annotated nucleomorph-encoded proteins, no ORFan/nORF-encoded enzyme involved in primary metabolism was detected, but several, cellular protein factors, not known at the time of publishing the nucleomorph sequences, could be predicted. These include factors involved in transcription, RNA processing, RNA degradation, translation, genome modifications and replication, in addition to factors involved in cellular functions such as a symbiont-specific ERAD-like machinery (SELMA), a cytosolic iron–sulfur assembly (CIA) machinery, and protein degradation. Notably, our approach led to the functional assignment of nucleomorph-encoded, putatively plastid-targeted proteins and surprisingly to homologs of factors that might be involved in membrane architecture and vesicle generation (summarized in fig. 2).

Materials and Methods

Data Acquisition and Analyses

Sequence data (*G. theta*: NC_002751.1, NC_002752.1, NC_002753.1; *H. andersenii*: CP000881, CP000882, CP000883; *C. paramecium*: CP002172, CP002173, CP002174; *Ch. mesostigmatica*: CP003680, CP003681, CP003682) were retrieved from NCBI (for accession numbers of the predicted nucleomorph-encoded proteins [pNMPs] see Supplementary Material online) and analyzed by Phyre^2 (Kelley et al. 2015). From the output of the Phyre^2 analyses, hits were inspected according to the criteria described in the text. The accession numbers of the nucleomorph-encoded "coat-like proteins" of chlorarachniophytes are: *Bigelowiella natans*: XP_001712832.1, XP_001712833.1; *Amorphochlora amoebiformis*: BAS01728.1, BAS01729.1; *Lotharella oceanica*: AIB09834.1, AIB09847.1, AIB09848.1, AIB09984.1; *Lotharella vacuolata*: BAS01668.1.

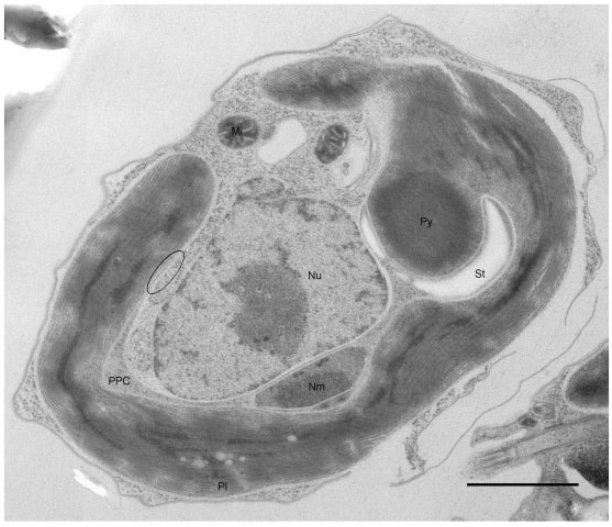


Fig. 1.—Fifty nanometers ultrathin section of a high pressure frozen, freeze substituted, resin embedded *Guillardia theta* cell; Mi, mitochondrion; Nm, nucleomorph; Nu, nucleus; Pl, plastid; PPC, periplastid compartment; Py, pyrenoid; St, starch; circle indicates the position of the transition zone between open form and attached zone of the PPM and the OEM (see text for details). Bar = 1 μm.

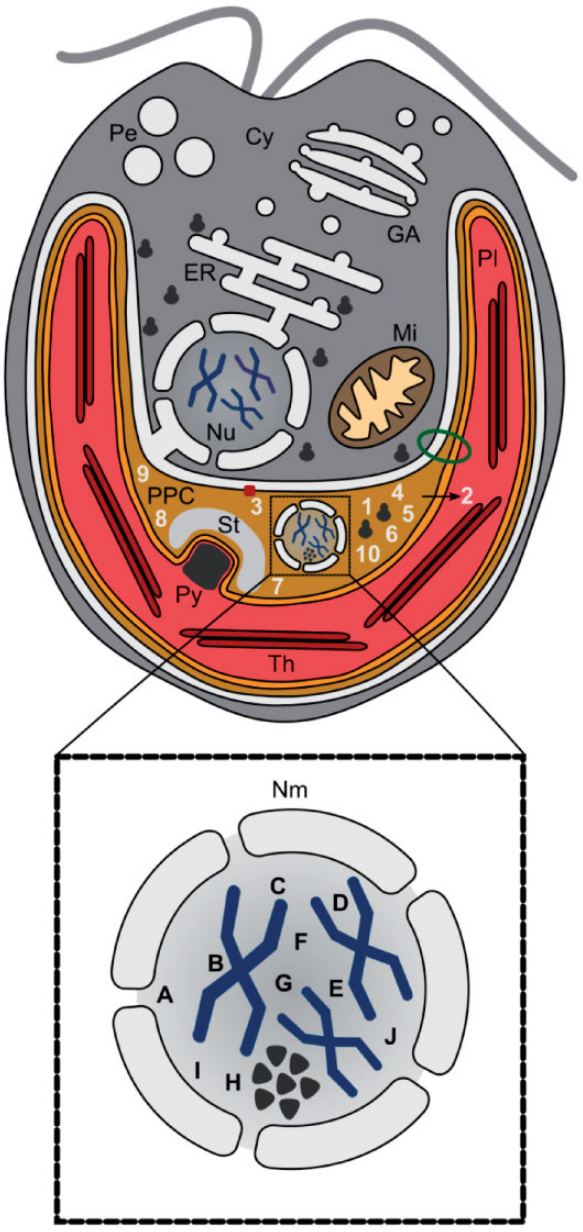
Electron Microscopy

Concentrated cells were high pressure frozen (HPF Compact 02; Wohlwend, CH), freeze substituted (in acetone containing 0.25% osmium tetroxide, 0.2% uranyl acetate, and 5% water) (AFS2; Leica, Wetzlar, Germany) and embedded in Epon 812 substitute resin. Freeze-substitution was done as follows: −90 °C for 20 h, from −90 °C to −60 °C for 1 h, −60 °C for 8 h, −60 °C to −30 °C for 1 h, −30 °C for 8 h, and −30 °C to 0 °C for 1 h. At 0 °C, samples were washed three times with acetone before a 1:1 mixture of Epon 812 substitute resin (Fluka Chemical, Buchs, Switzerland) and acetone was applied at room temperature. After 2 h the 1:1 mixture was substituted with pure resin overnight. After another substitution with fresh Epon, samples were polymerized at 60 °C for 2 days. Using an ultramicrotome (UC7; Leica) and a diamond knife (Diatome, Biel, Switzerland) 50 nm sections were cut from the Epon blocks and transferred to 100-mesh copper grids coated with pioloform. For additional contrast, sections were poststained with 2% uranyl acetate for 20–30 min and subsequently with lead citrate for another 1–2 min. Sections were finally analyzed and imaged using a JEM-2100 transmission electron microscope (JEOL, Tokyo, Japan) equipped with a 2k × 2k F214 fast-scan CCD camera (TVIPS, Gauting, Germany).

Results and Discussion

In order to increase the knowledge on the functions of ORFans/nORFs-encoded proteins, we reanalyzed the protein sequences encoded in the four available nucleomorph

A



B

Abbr.	Classification	CMESO	CPARA	Gt	HAN	Σ
1	translation	5	3	5	5	18
2	plastid proteins	5	0	8	5	18
3	SELMA	1	1	2	2	6
4	CIA	2	2	3	2	9
5	protein modification	3	2	1	2	8
6	protein degradation	2	2	1	6	11
7	plastid division/put. vesicle forming	3	7	4	3	17
8	biochemistry	1	1	2	2	6
9	signaling	0	0	1	0	1
10	unknown	0	0	1	0	1
A	import/export	0	0	1	0	1
B	centromere/kinetochore	2	0	2	2	6
C	telomere	0	0	1	0	1
D	chromatin modification/packaging	1	1	6	1	9
E	transcription	9	6	6	7	28
F	cell cycle, replication, mitosis	13	3	8	9	33
G	RNA processing, splicing	6	7	6	7	26
H	exosome	0	0	4	0	4
I	ribosome biogenesis.	1	2	2	1	6
J	repair/signaling	1	1	2	1	5
	Σ	55	38	66	55	214

Fig. 2.—Scheme of *Guillardia theta* and identified cellular processes and structures. Numbers and letters in (*A*) (left) indicate cellular processes/structures in the PPC/plastid (numbers) or within the nucleomorph (letters), in which pNMPs identified in this study might be involved. Both, letters and numbers refer to (*B*) (right), indicating the statistics on and classification of the identified pNMPs. Abbreviations in (*A*) are as follows: Pe, peroxisome; Cy, cytoplasm; GA, Golgi apparatus; ER, endoplasmic reticulum; Mi, mitochondrion; Nm, nucleomorph; Nu, nucleus; Pl, plastid; PPC, periplastid compartment; Py, pyrenoid; St, starch; Th, thylakoids. The green circle indicates the position of the transition zone between open form and attached zone of the PPM and the OEM (see text for details).

genomes using Phyre^2 (Kelley et al. 2015). In consequence of a new functional annotation of the encoded proteins, their genes lost the "status" of an ORFan/nORF; thus, we termed the encoded proteins pNMPs.

Phyre^2 Predictions

For assigning a protein function based on sequence similarity, BLAST algorithms are powerful tools. Here we applied the tool Phyre^2 (Kelley et al. 2015) to those ORFs, for which no

significant BLAST hits (according to our criteria an *E*-value of at least 10^{-5}) were determined at the time of publishing the respective nucleomorph genomes. Therefore, our data set includes a smaller set of uncharacterized ORFs described, for example, in Douglas et al. (2001), for which in the latest genome project (Moore et al. 2012) a function of the encoded proteins could be predicted by BLAST, but not by Phyre^2. Phyre^2 predicts the tertiary structure of the input sequence and compares it with already determined protein structures. We considered those predictions of Phyre^2 as "significant" in which a hit showed a confidence of at least 80% and alignment coverage of 35% or higher (in the further the "80–35% rule"). The reason for the latter is that nucleomorph-encoded proteins are known to be "compacted" in some cases (e.g., Moore and Archibald 2009) and therefore might lack a strict overall amino acid sequence conservation. The results of the here used semiautomatic approach led to outputs for 215 new pNMPs according to our criteria. From the individual outputs, the first and probably best match, unless otherwise stated in the text, was used for annotating a putative function of the pNMPs, based on the Phyre^2-results generated in January 2019 (supplementary table 1, Supplementary Material online).

Although applying the "80–35% rule" led to a significant increase in the prediction of the putative functions of *ORFan/nORF*-encoded proteins, limitations of this approach exist. First, the tool depends on known but sometimes quantitatively still limited protein structures, which might explain why many ORFans/nORFs still cannot be assigned to a cellular function. Second, we note that counting a Phyre^2-output according to our "80–35% rule" is stringent, but might have overlooked structural homologies with lower prediction. This became obvious in rare cases in which the "80–35% rule" highlighted a pNMP of one specific organism, but not the homologous ones (determined by BLAST) of another cryptophyte. In addition, several pNMPs lack homologs in other nucleomorph genomes (determined by Phyre^2 and BLAST). This might indicate that either the function of a pNMP is not necessary in every cryptophyte PPC/nucleomorph/plastid and therefore can be lost, or their genes were transferred into the cell nucleus, or the sequences are too divergent thereby precluding annotation of their encoded products by the algorithms used. Loss of genes was expected in case of splicing factors in *H. andersenii* (Lane et al. 2007), which is devoid of introns in nucleomorph-specific mRNAs, or in the case of factors acting in the photosynthesis apparatus in the nonphotosynthetic cryptophyte *C. paramecium* (Tanifuji et al. 2011).

In order to make the sequences processed by Phyre^2 informative, we used an acronym of the species name (Gt for *G. theta*, HAN for *H. andersenii*, CPARA for *C. paramecium*, CMESO for *Ch. mesostigmatica*) followed by the number/letters of the ORFans/nORFs in the respective nucleomorph data set (see Materials and Methods for accession numbers). For example, CMESO_443 is a hypothetical protein encoded in the nucleomorph of the cryptophyte *Ch. mesostigmatica*, specified by the number 443 in the NCBI database. For specifying a gene, the identifier is written in italics (for NCBI protein accession numbers and further information on the individual pNMPs see supplementary fig. 1, Supplementary Material online).

Nucleomorph Structure and Its Genetic Apparatus

Nucleomorph: Transport across the Pores, Nuclear Exosome

Early ultrastructural investigations of the PPC of cryptophytes already showed that a nucleomorph is surrounded by a typical nuclear envelope including nuclear pores (e.g., Morrall and Greenwood 1982). However, identification of proteins of the nucleomorph pores is challenging (Irwin and Keeling 2019) and our analyses did not add any new factors situated within the pores of the nucleomorph envelope. Nevertheless, one could expect that beside structural components of the envelope, the shuttling of molecules and complexes from the PPC into a nucleomorph and vice versa is regulated and needs accessory factors for functionality. In this respect Gt2_236 might be acting as the yeast homolog Npa3 which has, beside other described functions, GTPase activity and might be involved in nucleomorph localization of the RNA polymerase II (e.g., Staresincic et al. 2011) in cryptophytes.

From the output of our analyses we postulate that, as in other nuclei, a nuclear exosome might be present in the nucleomorph of cryptophytes. However, we cannot exclude that the postulated exosome is PPC localized. In this, the exosome complex exonucleases Rrp41, Rrp42, and Rrp46 predicted to be encoded by *Gt3_228* (Rrp41), *Gt3_272a* (Rrp42), and *Gt3_209* (Rrp46), as well as a dis3-like exonuclease 2 (*Gt1_443*) might be operating.

Nucleomorph Chromosomes and Chromatin

The nucleomorph genomes are organized in three chromosomes (e.g., Maier et al. 2000). For their structure and maintenance many factors are necessary, and several pNMPs homologous to chromosome-organization factors were identified in our analyses. These include pNMPs acting in the Ndc80 complex of the kinetochore (Gt3_479, HAN_1g89; Gt2_338, CMESO_513, CMESO_279, HAN_3g409), a telomerase associated protein p45 (Gt2_196) and, surprisingly, homologs to Apaf1 (Gt1_755, CMESO_448, CPARA_2gp223, HAN_1g130) known to be not only directly involved in apoptosis but also in regulating centrosome morphology and function (Ferraro et al. 2011). Beside chromosome associated proteins such as histones, already known to be encoded by the nucleomorphs or predicted here as in the case of CPARA_2gp268 (homology to nucleosome core histones), four likely chromatin modifying factors were identified: a N-methyltransferase (Gt1_365a and CMESO_337), a

chromatin remodeling Ruvb1-like protein (Gt1_391), a histone acyltransferase (Gt1_313b), and the Plus3 protein for chromatin modification (HAN_1g96). In addition, Gt1_153 showed structural homology to the antisilencing protein 1 (Asp1) and histone H3 chimera, in which Asp1 might be involved in nucleosome assembly and disassembly. Furthermore, a homolog to the DNA topoisomerase 2 is encoded by *Gt1_183*. Whether this enzyme is acting in the nucleomorph or in the plastid remains to be determined.

We additionally identified two pNMPs, which might be involved in DNA repair: CPARA_2gp260, homologous to the DNA damage-binding protein 1 and CMESO_34, predicted to act as a DNA photolyase. If Gt1_272 and HAN_3g486, homologous to the cryptochrome/photolyase, *N*-terminal domain, are acting in DNA repair or in signaling has to be investigated.

Cell Cycle, Replication, and Mitosis

Analyses of the cell cycle of cryptophytes showed that it is highly controlled (e.g., Onuma et al. 2017). As a result of our analyses, several nucleomorph-encoded pNMPs can now be assigned to cell cycle functions. This includes a homolog to the S-phase specific p45 (CMESO_443, HAN_1g125). Further factors likely acting in the S-phase at the replication complex were also identified: the probable DNA replication complex GINS proteins Psf1, Psf2, and Psf3 (Psf1: CMESO_529; Pfs2: Gt3_177b, HAN_1g168, CMESO_465; Psf3: CMESO_187, CMESO_200, CMESO_380, CMESO_393, CMESO_574, CMESO_7, HAN_2g232; Sld5: HAN_1g33, CPARA_1gp133, CMESO_543, Gt3_187), a PSF1 *N*-terminal domain-like protein (HAN_1g43), subunits of the minichromosome maintenance protein complex, Mcm proteins (Gt1_451, Gt1_532, Gt1_670, Gt1_744, HAN_1g119, Gt2_673, HAN_1g121), the ATP-dependent DNA helicase q4 (CMESO_102, CPARA_3gp393), and a homolog to replication factor A (HAN_2g356). After replication initiation, CMESO_355, a predicted single strand DNA-binding domain protein, SSB, might also be active. In addition, homology to the SKP1-like protein 1a was identified in HAN_3g521, involved in the ubiquitination and subsequent proteasomal degradation of proteins.

We expect no meiosis in the former eukaryotic endosymbiont. Thus, the homolog to the chromosome segregation in meiosis protein 2, encoded by *Gt2_160*, might be involved in recombinational repair and function in the protection of the genome from spontaneous or induced DNA damage (e.g., Huang et al. 2003).

Transcription, RNA Maturation, and Processing

Transcription and RNA processing of nucleomorph genes is, at least due to overlapping coding regions in nucleomorph genomes (e.g., Zauner et al. 2000), complicated and has been studied in more detail recently (Wong et al. 2018). Phyre^2 identified several pNMPs possibly involved in transcription. First, homologs to the subunits either for the RNA polymerase II or III: RNA polymerase II subunit RBP4 (RpoF) (Gt3_132, Han_3g391) and the DNA-directed RNA polymerase III subunit Rpc3 (Gt3_456, CMESO_493, CPARA_2gp188, HAN_1g161). In the process of transcriptional initiation several pNMPs might be involved: the transcription initiation factor IIa large chain (Gt3_266) and subunit 1 (CMESO_83, Han_3g392), the transcription initiation factor TFIID, subunit 9 (TAF 9) (HAN_3g497), as well as the transcription initiation factor IIe subunit beta (CPARA_2gp283, CMESO_424). The nucleomorph genome of cryptophytes is A/T rich (e.g., Moore and Archibald 2009; Grosche et al. 2014). Therefore, defining a TATA-box in silico is complicated. Nevertheless, a TATA-box factor apparently acts in transcriptional regulation as seen by a predicted TATA-binding protein-associated phosphoprotein (CMESO_458) and associated factors of the TATA-binding protein, TAFs (CMESO_214, CPARA_1gp024) and the TFIID subunit 9 (CMESO_441). In addition to a TATA-box, CCAAT boxes can be important cis-acting sequences in promoter regions. This might also be true for nucleomorph-encoded genes, as with HapE a potential protein was identified that perhaps binds in a limited number of gene promoter sequences CCAAT boxes (CMESO_551, HAN_1g41).

Furthermore, several pNMPs, homologous to transcription factors, were identified by Phyre^2: MafG (Gt1_176, CPARA_1g129), a subunit of the transcription factor TFIIe (Gt2_207), the RNA polymerase II transcription factor b subunit 2 (Gt3_180, HAN_3g523) and subunit 4 (CMESO_78, CPARA_1gp125, HAN_2g261), and a so-called POU-specific domain protein (CPARA_2gp297). Especially the latter is of interest, as POU-specific domain proteins were not identified in plants and fungi (http://pfam.xfam.org/family/PF00157?tab=pdbBlock#tabview=tab0; Last accessed May 27, 2019). Once transcription has been initiated, the transcription elongation factor Spt5, identified to be encoded by *CMESO_307*, might be involved.

After termination of transcription, the mRNA has to be polyadenylated. For this, a predicted homolog to the subunit 5 of the cleavage and polyadenylation specificity factor was identified (CPARA_3gp337, Han_3g380). Interestingly, *Gt3_590* is predicted to encode a dosage compensation regulator. If so, a regulation to equalize the expression of nucleomorph proteins including pNMPs is present.

Finally, we identified two further putative RNA binding proteins, a dsRNA binding protein (canonical RBD: HAN_1g62) and CMESO_233 as a homolog of synaptotagmin-binding, cytoplasmic RNA-interacting protein, isoform k.

For rRNA modifications/processing four putative factors were identified: a homolog to the rRNA-processing proteins Fcf1 (Gt1_127) and Utp23, (HAN_1g94), a putative dimethyladenosine transferase (CMESO_357), known to dimethylate two adjacent adenosines in the loop of a conserved hairpin near the 3′-end of 18S rRNA, and the predicted U3 small

nucleolar RNA-associated protein 21 (Gt1_859), involved in rRNA processing.

As mentioned above, the nucleomorph genome of *H. andersenii* does not show any introns in protein encoding genes (Lane et al. 2007). Thus, we did not expect any splicing factors to be encoded, but mRNA processing should be necessary. This assumption seems to be correct because with HAN_2g281 a homolog to the pre-mRNA-processing ATP-dependent RNA helicase Prp5 was identified, which might in case of overlapping transcripts (expressed from complimentary strands of the DNA) avoid the creation of partially dsRNA regions between two transcripts. Double stranded RNAs might furthermore be the target of CMESO_235, a predicted homolog to the HkH motif-containing C2H2 finger protein.

Several factors putatively involved in mRNA splicing were identified by Phyre^2. These are predicted homologs to the U1 small nuclear ribonucleoprotein 70 kDa homolog (Gt3_112), the U1 small nuclear ribonucleoprotein C (CPARA_2gp273), and the pre-mRNA-splicing factors Rse1 (CPARA_2gp261) and Snu114 (CPARA_3gp442). Further factors identified to be encoded in nucleomorph genomes and putatively involved in mRNA maturation are homologs of the splicing factor 3b subunit 1 (CPARA_2gp291), a factor homologous to the like-Sm ribonucleoprotein core (HAN_2g262), and the protein Cwc16 (Gt3_72, CMESO_128). Additionally, *CMESO_298* encodes a predicted homolog to the small nuclear ribonucleoprotein-associated protein B, which might have a functional role in pre-mRNA splicing or in snRNP structure. Finally, CPARA_2gp262 was predicted to be homologous to a putative pre-mRNA splicing protein.

Some years ago, we identified unusual introns in nucleomorph-encoded tRNAs of *G. theta* (Kawach et al. 2005). Predicted endonucleases putatively involved in the maturation of nucleomorph-encoded tRNAs were found to be encoded in all of the four sequenced nucleomorph genomes (*Gt2_171*, *CMESO_311*, *CPARA_1gp091*, *HAN_2g237*, *HAN_2g218*).

Translation

Chlorarachniophytes and cryptophytes express two sets of 80S ribosomes, and in case of cryptophytes several ribosomal proteins of the large and small subunit of the PPC-located ribosomes were already identified to be encoded in the nucleomorph genome (e.g., Douglas et al. 2001) and further analyzed (Maier et al. 2013). We add the predicted homologs of 60S ribosomal proteins L23-a (HAN_2g304, CMESO_148, CPARA_3gp416), the 40S ribosomal protein S20 (Gt1_556), L1 (CMESO_438), the 40S ribosomal protein S15a (Gt2_245), the 60S ribosomal protein L3 (Gt1_224) as well as the 60S acidic ribosomal protein P1 (CPARA_3gp428, HAN_2g319) to this list by analyzing nucleomorph-encoded ORFans and nORFs. In addition, Gt2_125b, CMESO_512, and HAN_3g410 were identified with homology to WD40 domain proteins and might be proteins of the translation machinery as well.

For their assembly ribosomes need several biogenesis/assembly factors. Phyre^2 identified predicted homologs to the ribosome assembly protein 4 (CPARA_2gp311) and to UTP21 (CPARA_1gp082, HAN_2g269, CMESO_524), as well as a homolog to BRX1 (Gt3_199). In addition, Gt2_186 might be part of the ribosome or involved in ribosome biogenesis.

For the initiation of the translation within the PPC the suppressor of stem-loop protein 1 might be important, which is predicted to be encoded by three cryptophytes (*CMESO_485*, *CPARA_2gp198*, *HAN_1g156*), although other functions of this protein might be possible. In addition, predicted homologs to the nascent polypeptide-associated complex subunit alpha (Gt1_68, CMESO_47, HAN_2g229) were identified, known to act at the ribosomal exit tunnel.

Protein Modification, Signaling, and Protein Degradation

One would expect that signaling to, within and trough the PPC is an important issue, for example to coordinate the needs of the stroma or to communicate with the host. To process cellular information, the signal transduction protein CBL (a homolog encoded by *Gt1_446*), an E3 ubiquitin-protein ligase involved in cell signaling and protein ubiquitination, might be necessary.

As soon as the nucleomorph-encoded mRNAs are translated the proteins have to be folded and in some cases post-translationally modified. For protein folding, chaperone functions are necessary. It remains to be proven if homologs of the DnaJ homolog subfamily A member 3, encoded by three nucleomorph genomes (*CMESO_340*, *CPARA_1gp057, Han_3g476*), are involved in protein folding.

According to the discussion presented later, posttranslational modifications might be important in the PPC of cryptophytes, and two potential factors involved in such modifications were identified: a predicted palmitoyltransferase (HAN_1g163, Gt3_272b, CMESO_490, CPARA_2gp192) as well as a predicted Rab geranylgeranyltransferase alpha subunit encoded by *CMESO_407*.

We have already noticed that in the PPC of *G. theta* a 26S proteasome is present (Stork et al. 2012; Maier et al. 2015). Nucleomorph-encoded subunits, hidden in ORFs, might be Rpn6 (CPARA_2gp244, HAN_1g98), the non-ATPase regulatory subunit 1 (HAN_3g401) and subunit 12 (HAN_2g268). In addition, a homolog to a zinc finger protein with an Ufm1-specific peptidase domain was identified when analyzing Gt2_270, CMESO_497, CPARA_3gp345, and HAN_3g402. However, a prediction of the cellular machineries, in which this putative deubiquitinating enzyme acts, is not possible at the moment. A further nucleomorph-encoded protein acting as a putative modification factor might be a predicted tail-specific protease (CMESO_239 and HAN_1g57). Whether this

predicted protease is acting in the PPC/nucleomorph or the plastid has to be determined.

Biochemistry

In contrast to discovering numerous new molecular biological functions to the proteins encoded by nucleomorph-encoded ORFs, very few biochemical functions could be assigned. This includes a homolog to the protein Dph2, probably involved in diphthamide biosynthesis (Gt1_414). In addition, a putative NADH-cytochrome b5 reductase (HAN_1g84) is encoded in the nucleomorph of *H. andersenii* CCMP644. Considering that thioredoxins are involved in redox regulation, we can now include pNMPs homologous to thioredoxin (Gt2_139, CMESO_341, CPARA_3gp387, HAN_3g458) in the biochemistry repertoire. Although the localization of these putative thioredoxins is not determined, the identification of a thioredoxin in the PPC of a nucleomorph-lacking diatom (Weber et al. 2009; Moog et al. 2011) might indicate the importance of redox regulation in PPCs in general.

Other Cellular Functions

Beside the machineries involved in storing genetic information and its expression, only minor additional PPC-specific machineries were identified. The reanalysis of the nucleomorph genome sequence of *G. theta* led to the identification of two *orf*-encoded factors of the SELMA complex some years ago (Sommer et al. 2007). An ERAD-associated E3 ubiquitin-protein ligase Hrd1 is encoded in all four nucleomorph genomes (Gt2_477, CMESO_94, CPARA_3gp400, HAN_2g286). The limitations of a Phyre^2 search can be seen in case of a second SELMA factor, a Derlin protein. Phyre^2 identified it as a rhomboid-like protein in Gt1_201 and HAN_1g47, because no Derlin structure is present in the database. In addition, BLAST searches indicated Derlin homologs with highest confidence encoded in the remaining nucleomorphs as well (annotated as Der1; CPARA_2gp263, CMESO_260). Additionally, we recently published our findings on the CIA machinery (Grosche et al. 2018). ORFs detected in our analyses, which might encode the CIA factors Cia2 and Cia1 are *Gt3_143, CPARA_2gp182*, *CMESO_129*, *HAN_1g169*, *Gt3_357*, *CMESO_410*, *CPARA_3gp362*, and *HAN_2g329*. Whether *Gt1_312* encodes a further Cia1 remains to be elucidated.

Nucleomorph-Encoded Plastid Proteins

One major reason for the existence of a nucleomorph in cryptophytes and chlorarachniophytes might be the presence of nucleomorph-localized genes encoding plastid proteins. These are, as far as we know, expressed with an *N*-terminal transit peptide-like (TPL) sequence. Although an indicative phenylalanine is present in the proximate *N*-terminal region in several but not in all already identified nucleomorph-encoded plastid proteins, a TPL sequence can be overlooked easily.

By analyzing the nucleomorph genome of *G. theta*, Douglas et al. (2001) identified 30 *orfs* encoding putative plastid proteins. From those, 19 could be assigned to a plastid function whereas the remaining 11 *orf*s were postulated to encode plastid factors according to their homology to cyanobacterial proteins. Our reanalysis with Phyre^2 indicated that several of the encoded pNMPs can now be functionally assigned (see below). The same is true for some candidates of the other three cryptophyte nucleomorph genomes. Since the publication of Douglas et al. (2001), two pNMPs were already annotated by us in independent studies: *Gt1_467*, encoding a putative subunit of the SUF (sulfur formation) system (Hjorth et al. 2005) as well as a phycocyanobilin lyase cpcT (Bolte et al. 2008) encoded by *Gt1_222*. Interestingly, *Gt1_225* encodes a predicted phycocyanobilin lyase (cpcT) as well.

Reanalyzing ORFs with homology to cyanobacterial proteins via Phyre^2 led to two corrections according to Douglas et al. (2001). First, Gt1_176 was identified to be homologous to the transcription factor Maf(G) (see above). And second the *orf*s *Gt3_81* and *Gt3_116* were wrongly annotated and due to a missing nucleotide in the sequence encode together a methyltransferase of the plastidal rRNA small subunit (GidB) (see also Moore et al.2012).

The annotated *G. theta* ORFs with homology to cyanobacterial factors in Douglas et al. (2001), for which a putative function was identified via Phyre^2, are: Gt2_227 (homologs CMESO_304 and HAN_3g442, vkorc1/thioredoxin domain protein, Gt2_249, CMESO_411, and HAN_2g330 [upf0603 protein at1g54780, chloroplastic, involved in photosystem II repair], Gt1_228, HAN_1g126, and CMESO_444 [DNA processing chain A, DprA], Gt2_496, the GTPase Der, involved in ribosome biogenesis, HAN_2g321 [Glucose-inhibited division protein B, GidB], whose encoded protein was renamed into RsmG which might act as a methyltransferase, apparently specific to 16S rRNA, and Gt1_1019 [a chromosome partition protein SMC]). However, a PPC or plastid localization has to be determined for the latter.

For Gt2_163, annotated as a putative plastid protein by Douglas et al. (2001) with homologs encoded in two other nucleomorph genomes as well (CMESO_301 and HAN_2g361), Phyre^2 found homologies a to an uncharacterized protein. For completeness, analyses of Gt2_323, Gt1_773, and Gt2_125, previously annotated as putative cyanobacterial homologs (Douglas et al. 2001), led to no results according to our search criteria. Finally, we identified with (CMESO_38) a putative nucleomorph-encoded plastid protein, being homologous to the DEAD-box ATP-dependent RNA helicase CshA.

Unknown

Enigmatic and perhaps wrongly predicted is Gt1_197, showing a hit to a 55 kDa erythrocyte membrane protein. The

reason for this prediction might be solely a PDZ domain, found in both proteins.

The Surprises

Although putative membranous structures in the PPC of cryptophytes have been described (e.g., Gibbs 1981), it is difficult to actually classify these structures, as a discrimination between a detached membrane-surrounded compartment and, for example, an invagination of the PPM or extension of the OEM into the PPC on the basis of plain ultrastructural investigations is impossible. Based on our protein analyses, the presence of vesicle-like structures in the PPC might be possible. Firstly, Phyre^2 identified a homolog to Sey1 (CPARA_1gp126, CPARA_3gp415, CPARA_1gp064, Gt3_261, Gt1_532a, Gt1_534), a dynamin-like GTPase, involved, for example, in proper ER formation (e.g., Anwar et al. 2012), but also physically associated with other ER components that are involved in nuclear pore complex related functions (Casey et al. 2015). Further hints for components possibly involved in the formation of vesicular structures in the PPC were gained by the Phyre^2 outputs of several pNMPs, which seem to be homologous to components of the COPI-coat, the coatomer subunit alpha (HAN_1g132, CPARA_2gp222, CMESO_220, HAN_1g117, CMESO_450, CPARA_1gp092) and the beta′-subunit (Gt1_365, CPARA_2gp219, HAN_1g74, CPARA_2gp322, CMESO_288). As in the first impression two putative coat proteins were predicted, which in other organisms directly interact, we investigated them further. Both pNMPs have a lower molecular mass than the coatomer subunit alpha and beta′-subunit of eukaryotic homologs. In principle, this can be explained by the fact that nucleomorphs harbor compacted genomes encoding compacted proteins (e.g., Lane et al. 2007). However, by analyzing the nucleomorph-encoded putative coatomer proteins for coat protein-indicative domains via BLASTp, we noticed that only WD40 domains are predicted in these nucleomorph-encoded proteins with the exception of CPARA_2gp219 with no predicted domains at all.

In any case, as we wanted to get more data for nucleomorph-encoded factors that might play a role in the generation of putative vesicular structures, we additionally screened the Phyre^2-negative ORFans/nORFs of *G. theta* having at least one predicted membrane spanning region via BLAST searches. This led to the notion that coding regions, present in the nucleomorph genomes (*Gt3_160a*, *Gt3_160b*, *Gt3_160c*, *HAN_2g253*, *CPARA_2gp179*, *CPARA_3gp341*, *CMESO_133*), are homologous to Erv14/15/Cornichon protein notably known to be involved in COPII vesicle formation and incorporation of specific secretory cargos.

The Nucleomorph and PPC of Cryptophytes

We have identified a putative function for 215 previously unannotated pNMPs, although at least for one (Gt1_197) the functional description presented here is based on a PDZ domain solely. Several of the functions highlighted by the new annotations were not unexpected. This holds true for proteins involved in chromosome structure maintenance and DNA replication as well as in the functions converting the information stored in the DNA into proteins. In some ways, the outcome of our studies indicates the "history in the knowledge" in molecular and cellular biology. So in the case of ribosomal proteins, which were investigated in early times of cellular research and therefore already identified to be encoded in the nucleomorph genome of *G. theta* in Douglas et al. (2001). However, other predicted protein functions not known at this time, such as specialized factors of the kinetochore, the centromere or factors involved in translation were identified here and increased knowledge on the nucleomorph and the PPC of cryptophytes. Interesting is the identification of a homolog to Dph2. This enzyme is involved in the biosynthesis of diphthamide, a posttranslationally modified histidine found in eEF-2 (e.g., Zhang et al. 2010). In any case, at this point of analysis, it became obvious that maintaining a nucleomorph led to several mechanisms, which have to be expressed in the PPC (e.g., Maier et al. 2013; Grosche et al. 2014). This includes redox regulation in the PPC of cryptophytes as indicated by the presence of a nucleomorph-encoded putative thioredoxin. Interestingly a putative thioredoxin was identified in a PPC version in diatoms as well, indicating the importance of redox regulation in PPCs in general (Weber et al. 2009; Moog et al. 2011). However, enzymes involved in a primary metabolism are, according to the here presented functional description of pNMPs, not nucleomorph-encoded.

Some, or even all subunits of molecular machines acting in the PPC of cryptophytes can be encoded in the cell nucleus of the host. Those proteins are encoded with a PPC-targeting signal (Gruber et al. 2007), which is very helpful for an in silico prediction of the protein localization. By contrast, different components of PPC-specific cellular machineries can be encoded either completely or in parts by a nucleomorph genome as well. In these cases, their identification and localization deduced by an nucleomorph-encoded ORF might be comfortable due to the high gene density with no or very small intergenic regions in nucleomorph genomes of cryptophytes. The strategy to inspect the nucleomorph genome for factors of interest already led to important results: the identification of SELMA, an ERAD-derived system proposed to act as a protein transporter (Sommer et al. 2007), and of two putative factors of a CIA machinery (Grosche et al. 2018).

The results presented here warrant investigation of the possible function of PPC factors for generating vesicular structures or shaping the architecture of the PPC-surrounding membranes. Vesicular structures in the PPC of cryptophytes were already postulated in early ultrastructural investigations (see e.g., Gibbs 1981; Ludwig and Gibbs 1985). However, due to the use of traditional preparation techniques

(susceptible to artifacts) of cells as well as the analysis of 2D projection images, we are not completely convinced about the existence of these structures (deduced from ultrastructural observations only). Thus, a 3D view of the PPC is needed to avoid potential miss-interpretations of visualized structures. To strengthen ultrastructural data, the identification of proteins involved in vesicle generation and fusion as well as membrane shaping is needed and was enabled here by investigating nucleomorph-encoded ORFs via Phyre^2 and BLAST.

COPI vesicles are known to be formed at the Golgi apparatus (e.g., Jackson 2014). According to ultrastructural data (see also fig. 1), a typical Golgi apparatus is not present in the PPC of cryptophytes. With respect to the ER or "ER-derived" structures that might be another potential source for vesicle generation, the nuclear envelope of the nucleomorph is a prominent element, but not known to be continuous with an ER-like compartment. With respect to an ER-like compartment within PPCs, Gibbs summarized in 1981 new and existing data indicating the putative presence of a periplastid reticulum (PR) in several algae by the exclusion of cryptophytes having scattered vesicles and tubules instead (Gibbs 1981). An additional, putative host-derived ER-like membrane might be the membrane that surrounds the PPC. This membrane, the PPM, was interpreted by many researchers to be derived from the cytoplasmic membrane of the secondary endosymbiont (e.g., Cavalier-Smith 2000). However, in a recent publication, the origin of the PPM was discussed as an ER-derived membrane (Gould et al. 2015). In this hypothesis, the eukaryotic endosymbiont was subsequently surrounded by the host's ER, which ultimately closed to form the outer surrounding envelope of the later complex plastid, whereas the plasma membrane of the eukaryotic endosymbiont was lost.

Regardless of the nature and origin of a membrane, proteins are necessary not only for vesicle generation but also for vesicle fusion. Candidates with putative roles in vesicle generation in the PPC of cryptophytes were identified by Phyre^2. Among these are nucleomorph-encoded proteins with homology to COPI alpha and beta' subunits. In COPI-coat generation both proteins interact with each other and help to build a coat (e.g., Lee and Goldberg 2010). In the nucleomorph-encoded COPI protein candidates WD40 repeats were recognized (with the exception of CPARA_2gp322, see above), but we obtained no convincing evidence for further domains present in bona fide COPI alpha and beta' subunits of other organisms. Thus, although a protein pair was identified having sequence similarities to COPI-related proteins, participation of the nucleomorph-encoded ORFs in vesicle generation is in question. Nevertheless, we cannot exclude that the "missing" domains are provided by other proteins as discussed in Zaremba-Niedzwiedzka (2017). In any case, searching for "incomplete COPI factors" encoded by the nucleomorphs of chlorarachniophytes (*B. natans* [Gilson et al. 2006]; *L. oceanica* [Tanifuji et al. 2014]; *A. amoebiformis* and *L. vacuolata* [Suzuki et al. 2015]) via Phyre^2 led to the identification of putative homologs (according to the "80–35% rule") to COPI alpha and one beta' subunit as well (alpha: BNATCHR1110, BNATCHR1111, BAS01728.1, BAS01729.1, AIB09834.1_M951_chr2139, AIB09847.1_M951_chr2153, AIB09848.1_M951_chr2154, BAS01668.1. beta: AIB09984.1_M951_chr380). Although not all of these proteins are predicted to harbor a WD40 repeat, their presence might indicate the importance of these factors in PPCs in general.

In respect of vesicular fusion with a target membrane, SNAREs as well as Rab proteins and their interactors can be major players (e.g. Baker and Hughson 2016). Though, we did not find any evidence for the presence of such proteins while reanalyzing nucleomorph-encoded ORFs. Many SNAREs are proteins with a *C*-terminal single transmembrane domain (e.g., Beilharz et al. 2003) and are integrated into their target membrane via a C-tail anchor insertion mechanism (e.g., Denic et al. 2013). The genes encoding the components for that, as for example Get3, were also not detected to be encoded in the nucleomorph genome. However, some SNAREs do not harbor a *C*-terminal transmembrane domain, like for example SNAP-25, which is a palmitoylated membrane protein (e.g., Gonzalo and Linder 1998). Interestingly, a palmitoyl-transferase was identified to be encoded by nucleomorph-specific *orfs* (*HAN_1g163*, *CPARA_2gp192*, *CMESO_490*, *Gt3_272b*). A slightly better, but also indirect evidence was obtained for the presence of Rab proteins. This class of G-proteins is coupled to a target membrane via a prenyl moiety, posttranslationally added to the *C*-terminus of the protein (Seabra 1998). A geranylgeranyltransferase was already identified (*Gt1_ggt*, *CMESO_364* (ggt), *HAN_3g517* (ggt), *CPARA_1gp044* (ggt)) in the first description of nucleomorph sequences from cryptophytes (Douglas et al. 2001; Lane et al. 2008; Tanifuji et al. 2011; Moore et al. 2012). However, in our analyses *CMESO_407* was predicted to be homologous to a Rab geranylgeranyltransferase alpha subunit. In any case, we cannot exclude the possibility that the predicted palmitoyl-transferase and the geranylgeranyltransferase might have other functions than Rab/SNARE proteins in the PPC of cryptophytes.

Phyre^2 identified a homolog to Sey1, a putative ER-protein involved in maintaining the structure of the tubular endoplasmic reticulum network (e.g., Park and Blackstone 2010). We cannot exclude that the Sey1 candidate might be involved in nuclear pore formation functions of the nucleomorph envelope or act at the PPM as discussed later. However, another function for the nucleomorph-encoded Sey1 is more likely. We already identified a homolog to Gt3_ORF261 encoded by the nucleus of a diatom and targeted to the PPC when expressed as a GFP fusion protein (Moog et al. 2011). This protein is not only highly conserved in organisms with complex plastids of red algal origin, a homolog of this ORF is also present in *Cyanidioschyzon merolae* (CMD101C) and was

detected in a proteome study of the plastid-dividing machineries (Yoshida et al. 2010). Thus, as the complex plastid of cryptophytes (and diatoms) is of red algal origin, the most plausible explanation for the function of a nucleomorph-encoded Sey1 is that this protein is involved in the division of the complex plastid in cryptophytes.

At that point the question arises what functions vesicles or alleged vesicle generation/fusion proteins, respectively, in the PPC of cryptophytes might have. In this regard trafficking but also structural functions might be possible. In case of a function in molecule trafficking, lipid transfer from the PPM to membrane III (OEM), which is homologous to the outer envelope membrane of primary plastids, is possible. Although the lipid composition of membrane III is not known, vesicular transfer might provide lipids present in the PPM to membrane III. Very likely is lipid transfer to the envelope of the nucleomorph. As mentioned, the nucleomorph is surrounded by a nuclear envelope, but most likely not continuous with a PR/ER. Thus, to compensate an ER-localized lipid synthesis machinery, vesicles might provide lipids to the nucleomorph envelope, especially in the cell cycle in which two nucleomorph-envelopes are formed finally. In any case, for the generation of putative vesicles providing lipids to the nucleomorph envelope, the here via BLAST searches identified Erv14/15/Cornichon protein, encoded in the compacted nucleomorph genome of all analyzed cryptophytes, might be essential.

In addition to a "lipid transfer-only" function, putative PPC vesicles might be involved in vesicular protein transport. For that, a scrutiny of models explaining transport of nucleus-encoded proteins, either into the PPC or the stroma of "chromists" (a eukaryotic, probably polyphyletic kingdom which includes heterokontophytes, cryptophytes, haptophytes), might be useful (see e.g., Gibbs 1979; Kilian and Kroth 2005). In the case of diatoms, it was discussed that vesicles can originate at the PPM and fuse with a PR, from which again vesicles bud and fuse with the OEM (Kilian and Kroth 2005). Partial support for such a model was recently obtained by examination of a diatom at the ultrastructural level (Flori et al. 2016). These data and others were implemented recently in a further model on protein transport into the symbiont of chromists (Cavalier-Smith 2018). In short, ultrastructural data on different chromists were interpreted in a way that a PR exists, which is not connected to the PPM, and that proteins, transported via periplastid endocytic vesicles to the PR, use the proposed protein translocator SELMA (Stork et al. 2012) for protein delivery from the PR into the PPC. Although structures potentially resembling membranes were identified in PPCs in several cases (but not in our inspection presented here for *G. theta*, fig. 1), it is, according to the methods used, not possible to decide if in diatoms (and other organisms housing a plastid surrounded by four membranes) a PR exists in the PPC or if the membranous structures are artifacts from invaginations or disintegrating membranes, as especially in disrupted cells (a nightmare for studying ultrastructure) putative internal membranous structures are abundant (Flori et al. 2016).

Although the function of the two putative COPI proteins in bending activity remains speculative, they might be involved in the coupling of membranes in the PPC. The PPC, surrounded by membrane II (PPM) and the OEM, is not uniform. Instead, both PPC-surrounding membranes are either in near contact (attached zone) or the PPC expands (open form), the latter by trend in the middle of the plastid (Figs. 1 and 2). Thus, the transition zone between both, the attached zone and open form, needs to be stabilized and the here identified candidate proteins for vesicle generation and membrane bending might be involved in this function. Unfortunately, no transformation protocol for cryptophytes exists and no further statement can be made on the function(s) of the ER-related proteins encoded by the nucleomorph genome. In any case the most plausible and compelling function of the Sey1 homologous pNMP is in participating in plastid division, whereas the role of all other proteins with a putative membrane/vesicle function identified here is speculative. However, our reanalyses of nucleomorph-encoded ORFs led to the description of many new cellular players in the PPC of cryptophytes, which provide new data for explaining the functions needed in a nucleomorph-containing PPC. As in future more protein structures will be available, Phyre^2 should be used then for a re-evaluation of the ORFans/nORFs not characterized so far, combined with synteny analyses indicating conserved genomic localizations of essential ORFs (see Moore et al. 2012).

Conclusions

The re-evaluation of nucleomorph-encoded ORFs presented here indicates the enormous efforts, which have to be provided for maintaining a reduced cell nucleus and the machineries for the expression and modification of the encoded proteins within the PPC. Building on the many annotations and structures of proteins, which were published after the first publications on the nucleomorph genome sequences, we identified by the use of Phyre^2 putative components for genome organization, the expression of coding regions, mRNA processing and degradation as well as the transport of the RNA polymerase II into the nucleus. Additionally, we increased knowledge on PPC located functions such as ribosome assembly or modifications and degradation of proteins in addition to the annotation of predicted plastid proteins encoded by ORFs of nucleomorph genomes. Although most of these functions were not unexpected, we provide here evidence on the nature and variety of the involved factors. An intriguing finding is ORF-encoded proteins that might be involved in membrane organization. However, a closer inspection of these factors indicated that with the exception of a homolog to Sey1, probably involved in plastid division, the in vivo function of the other here identified factors are still in question.

Supplementary Material

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

Acknowledgments

We thank Marion Schön for excellent technical assistance. Financial support for this study was provided by the Deutsche Forschungsgemeinschaft (Ma 1232/16-1).

Literature Cited

Anwar K, et al. 2012. The dynamin-like GTPase Sey1p mediates homotypic ER fusion in *S. cerevisiae*. J Cell Biol. 197(2):209–217.

Baker RW, Hughson FM. 2016. Chaperoning SNARE assembly and disassembly. Nat Rev Mol Cell Biol. 17:465–479. doi: 10.1038/nrm.2016.65.

Beilharz T, Egan B, Silver PA, Hofmann K, Lithgow T. 2003. Bipartite signals mediate subcellular targeting of tail-anchored membrane proteins in *Saccharomyces cerevisiae*. J Biol Chem. 278(10):8219–8223.

Bolte K, et al. 2008. Complementation of a phycocyanin-bilin lyase from *Synechocystis* sp. PCC 6803 with a nucleomorph-encoded open reading frame from the cryptophyte *Guillardia theta*. BMC Plant Biol. 8:56.

Casey AK, Chen S, Novick P, Ferro-Novick S, Wente SR. 2015. Nuclear pore complex integrity requires Lnp1, a regulator of cortical endoplasmic reticulum. Mol Biol Cell. 26(15):2833–2844.

Cavalier-Smith T. 2000. Membrane heredity and early chloroplast evolution. Trends Plant Sci. 5(4):174–182.

Cavalier-Smith T. 2018. Kingdom Chromista and its eight phyla: a new synthesis emphasising periplastid protein targeting, cytoskeletal and periplastid evolution, and ancient divergences. Protoplasma 255(1):297–357.

Denic V, Dötsch V, Sinning I. 2013. Endoplasmic reticulum targeting and insertion of tail-anchored membrane proteins by the GET pathway. Cold Spring Harb Perspect Biol. 5(8):a013334.

Douglas S, et al. 2001. The highly reduced genome of an enslaved algal nucleus. Nature 410(6832):1091–1096.

Ferraro E, et al. 2011. Apaf1 plays a pro-survival role by regulating centrosome morphology and function. J Cell Sci. 124(Pt 20):3450–3463.

Flori S, Jouneau PH, Finazzi G, Maréchal E, Falconet D. 2016. Ultrastructure of the periplastidial compartment of the diatom *Phaeodactylum tricornutum*. Protist 167(3):254–267.

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

Gibbs SP. 1981. The chloroplast endoplasmic reticulum: structure, function, and evolutionary significance. Int Rev Cytol. 72:49–99.

Gilson PR, et al. 2006. Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. Proc Natl Acad Sci U S A. 103(25):9566–9571.

Gonzalo S, Linder ME. 1998. SNAP-25 palmitoylation and plasma membrane targeting require a functional secretory pathway. Mol Biol Cell. 9(3):585–597.

Gould SB, Maier UG, Martin WF. 2015. Protein import and the origin of red complex plastids. Curr Biol. 25(12):R515–R521.

Grosche C, Hempel F, Bolte K, Zauner S, Maier UG. 2014. The periplastidal compartment: a naturally minimized eukaryotic cytoplasm. Curr Opin Microbiol. 22:88–93.

Grosche C, Diehl A, Rensing SA, Maier UG. 2018. Iron–sulfur cluster biosynthesis in algae with complex plastids. Genome Biol Evol. 10(8):2061–2071.

Gruber A, et al. 2007. Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. Plant Mol Biol. 64(5):519–530.

Hjorth E, Hadfi K, Zauner S, Maier UG. 2005. Unique genetic compartmentalization of the SUF system in cryptophytes and characterization of a SufD mutant in *Arabidopsis thaliana*. FEBS Lett. 579(5):1129–1135.

Huang ME, Rio AG, Nicolas A, Kolodner RD. 2003. A genomewide screen in *Saccharomyces cerevisiae* for genes that suppress the accumulation of mutations. Proc Natl Acad Sci U S A. 100(20):11529–11534.

Irwin NAT, Keeling PJ. 2019. Extensive reduction of the nuclear pore complex in nucleomorphs. Genome Biol Evol. 11(3): 678–687.

Jackson LP. 2014. Structure and mechanism of COPI vesicle biogenesis. Curr Opin Cell Biol. 29:67–73.

Kawach O, et al. 2005. Unique tRNA introns of an enslaved algal cell. Mol Biol Evol. 22(8):1694–1670.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 10(6):845–858.

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

Lane CE, et al. 2007. Nucleomorph genome of *Hemiselmis andersenii* reveals complete intron loss and compaction as a driver of protein structure and function. Proc Natl Acad Sci U S A. 104(50):19908–19913.

Lee C, Goldberg J. 2010. Structure of coatomer cage proteins and the relationship among COPI, COPII, and clathrin vesicle coats. Cell 142(1):123–132.

Ludwig M, Gibbs SP. 1985. DNA is present in the nucleomorph of cryptomonads: further evidence that the chloroplast evolved from a eukaryotic endosymbiont. Protoplasma 127(1–2):9–20.

Maier UG, Douglas SE, Cavalier-Smith T. 2000. The nucleomorph genomes of cryptophytes and chlorarachniophytes. Protist 151(2):103–109.

Maier UG, et al. 2013. Massively convergent evolution for ribosomal protein gene content in plastid and mitochondrial genomes. Genome Biol Evol. 5(12):2318–2329.

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.

Moog D, Stork S, Zauner S, Maier UG. 2011. In silico and in vivo investigations of proteins of a minimized eukaryotic cytoplasm. Genome Biol Evol. 3:375–382.

Moore CE, Archibald JM. 2009. Nucleomorph genomes. Annu Rev Genet. 43:251–264.

Moore CE, Curtis B, Mills T, Tanifuji G, Archibald JM. 2012. Nucleomorph genome sequence of the cryptophyte alga *Chroomonas mesostigmatica* CCMP1168 reveals lineage-specific gene loss and genome complexity. Genome Biol Evol. 4(11):1162–1175.

Morrall S, Greenwood AD. 1982. Ultrastructure of nucleomorph division in species of *Cryptophyceae* and its evolutionary implications. J Cell Sci. 54:311–328.

Onuma R, Mishra N, Miyagishima S. 2017. Regulation of chloroplast and nucleomorph replication by the cell cycle in the cryptophyte *Guillardia theta*. Sci Rep. 7(1):2345.

Park SH, Blackstone C. 2010. Further assembly required: construction and dynamics of the endoplasmic reticulum network. EMBO Rep. 11(7):515–521.

Seabra MC. 1998. Membrane association and targeting of prenylated Ras-like GTPases. Cell Signal. 10(3):167–172.

Sommer MS, et al. 2007. Der1-mediated pre-protein import into the periplastid compartment of chromalveolates? Mol Biol Evol. 24(4):918–928.

Staresincic L, Walker J, Dirac-Svejstrup AB, Mitter R, Svejstrup JQ. 2011. GTP-dependent binding and nuclear transport of RNA polymerase II by Npa3 protein. J Biol Chem. 286(41):35553–35561.
Stork S, et al. 2012. Distribution of the SELMA translocon in secondary plastids of red algal origin and uncoupling of ubiquitin-dependent translocation from degradation. Eukaryot Cell 11(12):1472–1481.
Suzuki S, Shirato S, Hirakawa Y, Ishida K-I. 2015. Nucleomorph genome sequences of two chlorarachniophytes, *Amorphochlora amoebiformis* and *Lotharella vacuolata*. Genome Biol Evol. 7:1533–1545.
Tanifuji G, et al. 2011. Complete nucleomorph genome sequence of the nonphotosynthetic alga *Cryptomonas paramecium* reveals a core nucleomorph gene set. Genome Biol Evol. 3:44–54.
Tanifuji G, et al. 2014. Nucleomorph and plastid genome sequences of the chlorarachniophyte *Lotharella oceanica*: convergent reductive evolution and frequent recombination in nucleomorph-bearing algae. BMC Genomics 15:374.
Weber T, Gruber A, Kroth PG. 2009. The presence and localization of thioredoxins in diatoms, unicellular algae of secondary endosymbiotic origin. Mol Plant 2(3):468–477.
Wong DK, Grisdale CJ, Fast NM. 2018. Evolution and diversity of pre-mRNA splicing in highly reduced nucleomorph genomes. Genome Biol Evol. 10(6):1573–1583.
Yoshida Y, et al. 2010. Chloroplasts divide by contraction of a bundle of nanofilaments consisting of polyglucan. Science 329(5994):949–953.
Zaremba-Niedzwiedzka K, et al. 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature 541(7637):353–358.
Zauner S, et al. 2000. Chloroplast protein and centrosomal genes, a tRNA intron, and odd telomeres in an unusually compact eukaryotic genome, the cryptomonad nucleomorph. Proc Natl Acad Sci U S A. 97(1):200–205.
Zhang Y, et al. 2010. Diphthamide biosynthesis requires an organic radical generated by an iron–sulphur enzyme. Nature 465(7300):891–896.

Associate editor: Martin Embley