Philipps Universität Marburg

# Genetic compartmentalization in the complex plastid of *Amphidinium carterae* and The endomembrane system (ES) in *Phaeodactylum tricornutum*

Dissertation zur Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

Vorgelegt dem Fachbereich Biologie
der Philipps-Universität Marburg
FB 17 – Biologie, Zellbiologie
von

**Xiaojuan Liu**

aus Fujian, China
Marburg/Lahn 2015

Vom Fachbereich Biologie der Philipps-Universität Marburg

als Dissertation angenommen am: 2015

Erstgutachter: Prof. Dr. Uwe-G. Maier

Zweitgutachterin: Prof. Dr. Ralf Jacob
Prof. Dr. Andrea Maisner
Prof. Dr. Susanne Önel

Tag der Disputation am: 2015

## Publications

**X. liu,** F. Hempel, S. Stork, K. Bolte, D. Moog, T. Heimerl, UG. Maier, S. Zauner (2015). Addressing various compartments via sub-cellular marker proteins of the diatom model organism *Phaeodactylum tricornutum*. Algal Research. In preparation.

**In preparation:**

**X. liu,** C. Grosche, UG. Maier, S. Zauner (2015). Isolation of individual minicircles via a novel transposon-insertion based approach in *Amphidinium carterae*. Forthcoming.

**Contents**


Summary ..... 1

Zusammenfassung ..... 2

Abbreviations ..... 3

Figures and Tables ..... 4

1 Introduction ..... 6

1.1 The evolution of complex plastids ..... 6

1.1.1 The primary endosymbiosis ..... 6

1.1.2 The secondary endosymbiosis ..... 7

1.1.3 Plastid genome and gene transfer ..... 10

1.3 Endomembrane system ..... 11

1.3.1 The endomembrane system ..... 11

1.3.2 Vacuolar protein transport within endomembrane system ..... 12

1.3.2.1 Vacuolar sorting determinants ..... 13

1.3.2.2 Vacuolar protein transport via vesicles ..... 14

1.3.3 The biosynthesis of N-Glycoproteins on endomembrane system ..... 17

1.3.4 Tonoplast intrinsic proteins (Tips) ..... 18

1.3.5 Vacuolar-type $H^+$-ATPases (V-ATPase) ..... 19

2. Aim ..... 21

3 Results ..... 22

3.1 Genetic compartmentalization in the complex plastid of *Amphidinium carterae* ..... 22

3.1.1 The enrichment and isolation of minicircles ..... 22

3.1.2 Analysis of individual minicircles ..... 23

3.1.2.1 The overview of all individual minicircles ..... 24

3.1.2.2 The core regions of minicircles in *A. carterae* CCAM0512 ..... 27

3.1.2.3 Transcription and RNA editing analyses of individual minicircles ..... 28

3.1.3 Evolution analyses of four *A. carterae* strains' minicircles ..... 29

3.1.3.1 Overall genome characteristics and open reading frames ..... 29

3.1.3.2 Phylogenetic analysis of *psbA* genes for 15 minicircles of dinoflagellates ..... 31

3.1.3.3 Alignment analysis of core regions from four *A. carterae* strains ..... 32

3.1.3.4 Phylogenetic analysis of partial LSU/SSU rDNA ..... 34

3.2 The endomembrane system (ES) in *Phaeodactylum tricornutum* ..... 37
3.2.1 Identification of marker proteins ..... 37
3.2.2 Tonoplast intrinsic proteins (Tips) ..... 37
3.2.2.1 *In vivo*-localization of Tip1 ..... 38
3.2.2.2 *In vivo*-localization of Tip2 ..... 39
3.2.2.3 *In vivo*-localization of Tip3 and Tip5 ..... 40
3.2.2.4 *In vivo*-localization of Tip4 ..... 42
3.2.3 *In vivo* localization of Golgi proteins ..... 44
3.2.4 *In vivo* localization of retromer complex ..... 45
3.2.5 *In vivo* localization of vacuolar $H^+$-ATPase proteins ..... 47
4 Discussion ..... 50
4.1 Genetic compartmentalization of peridinin-containing dinoflagellates ..... 50
4.1.1 Minicircles of the peridinin-containing dinoflagellate *Amphidinium carterae* CCAM0512 ..... 50
4.1.1.1 Minicircles with coding genes ..... 51
4.1.1.2 Empty minicircles ..... 54
4.1.2 The evolutional relationship of minicircles ..... 55
4.2 "The endomembrane system (ES) in *Phaeodactylum tricornutum*" ..... 57
4.2.1 Identification of tonoplast intrinsic proteins (Tips) ..... 57
4.2.2 Identification of Golgi-marker proteins ..... 61
4.2.3 Identification of retromer complex proteins ..... 62
4.2.4 Identification of vacuolar type $H^+$-ATPases ..... 64
5 Materials and Methods ..... 65
5.1 Materials ..... 65
5.1.1 Instruments ..... 65
5.1.2 Membranes and filters ..... 66
5.1.3 Antibodies ..... 66
5.1.4 Chemicals ..... 66
5.1.5 Enzymes ..... 66
5.1.6 Software and bioinformatic applications ..... 67
5.1.7 DNA and protein markers ..... 67
5.1.8 Oligonucleotide primers ..... 67

5.1.9 Vectors ..... 68
5.1.10 organisms ..... 68
5.2 Methods ..... 68
5.2.1 Culture of *E.coli* TOP10 ..... 68
5.2.2 Culture of *Phaeodactylum tricornutum* CCAP 1055/1 ..... 69
5.2.3 Culture of *Amphidinium carterae* CCAM0512 ..... 69
5.2.4 Nucleic acid analytics ..... 70
5.2.4.1 Plasmid isolation from *E.coli* ..... 70
5.2.4.2 DNA and RNA isolation from *P. tricornutum* ..... 70
5.2.4.3 cDNA synthesis via reverse transcription (RT) ..... 71
5.2.4.4 Polymerase chain reaction (PCR) ..... 71
5.2.4.5 Agarose gel electrophoresis ..... 72
5.2.4.6 Sequencing ..... 73
5.2.4.7 Restriction and ligation ..... 73
5.2.4.8 Transformation of *E.coli* ..... 74
5.2.4.10 Transfection of *P. tricornutum* ..... 74
5.2.4.11 Minicircles enrichment and isolation from *A. carterae* CCAM0512 ..... 75
5.2.4.12 Transposon-insertion based approach ..... 75
5.2.4.13 DNA extraction, amplification, and sequencing of LSU rDNA domain D1-D6 and SSU rDNA for *A.carterae* ..... 77
5.2.4.14 Sequence alignment and phylogenetic analyses ..... 77
5.2.5 Protein analytics ..... 78
5.2.5.1 Protein isolation from *P. tricornutum* ..... 78
5.2.5.2 Protein extraction fractionation via carbonate extraction ..... 78
5.2.5.3 TCA protein precipitation ..... 79
5.2.5.4 Determination of protein concentration via Amido black ..... 79
5.2.5.5 SDS-polyacrylamide gel electrophoresis (PAGE) ..... 80
5.2.5.6 Western blot analysis ..... 80
5.2.5.7 Self-assembling GFP ..... 81
5.2.5.8 Construction of eGFP fusion proteins ..... 82
5.2.5.9 Confocal laser scanning microscopy ..... 82
5.2.5.10 Electron microscopy ..... 82

6 References ..... 84
7 Supplements ..... 107
7.1 Open reading frames with no known homology ..... 107
7.2 Identified marker proteins in *P. tricornutum* ..... 108
7.3 The Tip2 eGFP fusion protein ..... 108
7.4 Sequences of all used oligonucleotides ..... 109
8 Acknowledgements ..... 112
9 Curriculum vitae ..... 113
10 Erklärung ..... 114

## Summary

Peridinin-containing dinoflagellates are important members of single-celled eukaryotic algae, which arose from an engulfment of an ancient red alga by a so far undefined host cell, a process called secondary endosymbiosis. Their plastids feature a unique membrane architecture and are surrounded by only three membranes. As the reduction of the endosymbiont's genome and gene transfer from the plastid to the nucleus, the whole plastid genome was reorganized into minicircles coding for genes normally coded on the plastid genome. In order to isolate individual minicircles from one representative peridinin-containing dinoflagellate *Amphidinium carterae* CCAM0512 a novel transposon-based approach was carried out within this thesis. 89 minicircles were therefore isolated from *A. carterae*, 18 (20.2 %) are gene-containing minicircles, 71 (79.8 %) are empty minicircles. The 18 gene-containing minicircles are divided into three groups of minicircles, six single-gene minicircles, one two-genes minicircle and one three-genes minicircle. The 71 empty minicircles are divided into six groups. The characteristics of these minicircles and unique features were analyzed in this thesis. In contrast to previously reported organellar RNA editing in peridinin-containing dinoflagellates, no RNA editing was observed on transcripts of minicircles of *A. carterae* based on the analysis of coding genes. Additionally, the transcription of open reading frames was shown in so-called empty minicircles. Finally, based on the comparison with minicircles and rDNA sequences of three other *A. carterae* strains, it was speculated that minicircles undergo a rapid evolutionary diversification.

The mechanisms of protein (e.g. vacuolar proteins) transport and sorting have been well-studied in plants, yeast and animals. However, little is known in the diatom *P. tricornutum*. In order to investigate the protein transport and sorting in *P. tricornutum*, essential marker proteins have to be established. In this work, the identification of marker proteins in the endomembrane system was based on a combination of *in silico* search for homologous proteins of *P. tricornutum* to proteins with known localizations in plants and subsequent *in vivo* localization studies in *P. tricornutum*. Several markers for different subcellular compartments were identified including the plasma membrane, two vacuolar-like structures, cER, hER, the nuclear envelope and the second outermost membrane of the complex plastid (PPM). Furthermore, the three parts of the Golgi apparatus and the cytosol could also be marked. These useful subcellular marker proteins are a very important prerequisite for studying the mechanisms of protein transport and sorting in *P. tricornutum*.

## Zusammenfassung

Peridinin-haltige Dinoflagellaten sind wichtige Vertreter einzelliger eukaryoter Algen, welche durch die Aufnahme einer anzestralen Rotalge durch eine bislang unbekannte Wirtszelle entstanden sind, ein als sekundäre Endosymbiose bezeichneter Prozess. Sie besitzen Plastiden mit einer einzigartigen Membranarchitektur und sind nur von drei Membranen umgeben. Durch die Reduktion des Endosymbionten-Genoms und den Transfer von plastidären Genen in den Zellkern des Wirtes ist das plastidäre Genom in Form von sogenannten „minicirclen" organisiert, auf welchen Gene kodiert sind, die normalerweise im Plastidengenom kodiert sind. Im Rahmen dieser Arbeit wurden individuelle „minicircle" aus dem repräsentativen peridinin-haltigen Dinoflagellaten *Amphidinium carterae* CCAM0512 mittels einer neuartigen Transposon-basierten Methode isoliert. Insgesamt konnten dadurch 89 „minicircle" isoliert werden, davon kodierten 18 (20.2 %) genetische Informationen wohingegen 71 (79.8 %) sogenannte „leere minicircle" waren. Diese 18 kodierenden „minicircle" ließen sich in drei Gruppen unterteilen. Sechs „minicircle" kodierten ein einzelnes Gen, ein „minicircle" welcher zwei Gene kodiert sowie ein „minicircle" welcher drei Gene kodiert. Die 71 „leeren minicircle" ließen sich in sechs Gruppen einteilen, deren Charakteristika und Eigenschaften im Rahmen dieser Arbeit analysiert wurden. Analysen der kodierenden „minicircle" zeigten, dass auf Transkriptebene keine RNA Edierung in A. *carterae* beobachtet werden konnte, im Gegensatz zu „minicirclen" anderer peridinin-haltiger Dinoflagellaten, bei denen RNA Edierung nachgewiesen wurde. Im Falle von „leeren minicirclen" konnte die Transkription von offenen Leserahmen gezeigt werden. Basierend auf einem Vergleich von „minicirclen" und rDNA Sequenzen von drei weiteren A. *carterae* Stämmen, wurde spekuliert, dass „minicircle" einer rapiden evolutionären Diversifikation ausgesetzt sind.

Mechanismen zum Transport von Proteinen und deren Sortierung (z.B. von vakuolären Proteinen) sind in Pflanzen-, Hefe- und tierischen Zellen gut untersucht, jedoch in der Diatomee *P. tricornutum* größtenteils unbekannt. Um den Transport und die Sortierung von Proteinen in *P. tricornutum* untersuchen zu können, müssen initial essentielle Markerproteine etabliert werden. Anhand von *in silico* Analysen konnten solche Markerproteine, welche homolog zu pflanzlichen Proteinen mit bekannter Lokalisation sind, in *P. tricornutum* identifiziert und auf ihre subzelluläre Lokalisation hin untersucht werden. Mehrere Markerproteine für verschiedene subzelluläre Kompartimente, einschließlich der Plasmamembran, zwei vakuolen-ähnlicher Strukturen, des cERs, des hERs, der Kernhülle, der zweitäußersten Membran der komplexen Plastide, der verschiedenen Teile des Golgi-Apparats und des Cytosols, wurden identifiziert. Diese nützlichen Markerproteine stellen eine wichtige Voraussetzung für Studien am Mechanismus von Protein Transport und Sortierung in *P. tricornutum* dar.

# Abbreviations

| | | | |
|---|---|---|---|
| aa | Amino acid | Da | Dalton |
| BLAST | Basic Local Alignment Search Tool | GnTI | N-acetylglucosaminyltransferase I |
| bps | basepairs | PCR | Polymerase chain reaction |
| BTS | bipartite Targeting Sequence | FucT | α1,3-fucosyltransferase |
| CCV | Clathrin-coated vesicle | DV | Dense vesicle |
| cDNA | complementary DNA | ATPase1-3 | vacuolar type $H^+$-ATPase 1-3 |
| cER | chloroplast ER | mRFP | monomeric red fluorescent protein |
| COPI/II | Coat protein I/II | psVSD | Protein structure-dependent VSD |
| EE | early endosome | VSD | Vacuolar sorting determinant |
| EGT | Endosymbiotic gene transfer | HGT | Horizontal gene transfer |
| EM | endomembrane | PM | Plasma membrane |
| ER | endoplasmic reticulum | LV | Lytic vacuole |
| hER | host ER | PPM | The second outermost membrane of the complex plastid |
| EST | expressed sequence tag | Vps26/29 | vacuolar protein sorting 26/29 |
| LE | late endosome | ssVSD | Sequence-specific VSD |
| MC | minicircle | MVB | Multi vesicular body |
| PPM | periplastidal membrane | Tip1-5 | tonoplast intrinsic protein 1-5 |
| Pt | *P. tricornutum* | eGFP | enhanced green fluorescent protein |
| PVC | prevacuolar compartment | ctVSD | C-terminal VSD |
| rpm | rotations per minute | CLSM | Confocal laser scanning microscope |
| RT | room temperature | SA-GFP | Self-assembly GFP |
| SP | signal peptide | XylT | β1,2-xylosyltransferase |
| TGN | trans Golgi network | PSV | Protein storage vacuole |

# Figures

Fig. 1-1: Model of the evolution of complex plastids of green and red algal origin. ..... 9

Fig. 1-2: A working model for protein sorting to vacuoles. ..... 15

Fig. 1-3: N-glycosylation pathway during the synthesis of glycoproteins in plants. ..... 18

Fig. 1-4: Subcellular localization of vacuolar type $H^+$-ATPase (VHA) in plant. ..... 20

Fig. 3-1: The isolated minicircles of *A. carterae* CCAM0512. ..... 23

Fig. 3-2: GC content for genes-containing minicircles of *A. carterae* CCAM0512. ..... 27

Fig. 3-3: Alignment of the core regions of fourteen minicircles of *A. carterae* CCAM0512. ..... 27

Fig. 3-4. Transcription of *psbA* gene and ORFs of three empty minicircles. ..... 28

Fig. 3-5: GC content of *psbA* and 23S rRNA-containing minicircles in four *A. carterae* strains. ... 30

Fig. 3-6. Phylogenetic tree analysis based on amino acid sequences encoded by *psbA* gene from 15 minicircles of dinoflagellates. ..... 31

Fig. 3-7. Alignment of the core region of different *A. carterae* strains' minicircles. ..... 32

Fig. 3-8. Alignment of the non-coding regions of four different *A. carterae* strains' minicircles. 33

Fig. 3-9. Phylogenetic tree analysis based on LSU rDNA sequences covering domains D1-D6. ... 35

Fig. 3-10. Phylogenetic tree analysis based on SSU rDNA sequences. ..... 36

Fig. 3-11: *In vivo* localization of Tip1-eGFP. ..... 39

Fig. 3-12: *In vivo* localization of Tip2-eGFP. ..... 40

Fig. 3-13: *In vivo* localization of Tip3- and Tip5-eGFP. ..... 41

Fig. 3-14: *In vivo* co-staining and co-expression analyses of Tip3 and Tip5 proteins. ..... 42

Fig. 3-15: *In vivo* localization of Tip4-eGFP. ..... 43

Fig. 3-16: *In vivo* localization of proteins located in the Golgi apparatus. ..... 45

Fig. 3-17: *In vivo* localization of retromer complex proteins. ..... 47

Fig. 3-18: Co-expression of retromer complex proteins with Golgi markers. ..... 47

Fig. 3-19: *In vivo* localization of V-ATPase proteins. ..... 49

Fig. 4-1: Schematic overview of identified marker proteins in diatom *P. tricornutum*. ..... 60

Fig. 5-1: Schematic depiction of transposon-insertion based isolation of minicircles. ..... 77

Fig. S1: The alignment of Tip2 nucleotide sequences in *P. tricornutum*. ..... 109

# Tables

Table 3-1: Statistic analysis of isolated individual minicircles. ..... 23

Table 3-2: Properties of minicircles containing chloroplast genes in *A.carterae* CCAM0512. ..... 24

Table 3-3: Properties of empty minicircles in *A.carterae* CCAM0512. ..... 25

Table 3-4: Identical open reading frames in *A. carterae* CCAM0512. ..... 26

Table 3-5: Reported minicircular sequences in four *A.carterae* strains. ..... 29

Table 3-6: The open reading frames of empty minicircles in *A. carterae* strains. ..... 31

Table 3-7: Predicted subcellular localized Tonoplast intrinsic proteins in *P. tricornutum*. ..... 38

Table 3-8: Predicted subcellular localized Golgi marker proteins in *P. tricornutum*. ..... 44

Table 3-9: Predicted subcellular localization of retromer complex proteins in *P. tricornutum*. ..... 46

Table 3-10: Predicted subcellular localization of V-ATPase proteins in *P. tricornutum*. ..... 48

Table S1: All open reading frames mentioned in this study. ..... 107

Table S2: Proteins localized as eGFP/ mRFP fusions in this study. ..... 108

# 1 Introduction

## 1.1 The evolution of complex plastids

### 1.1.1 The primary endosymbiosis

One of the important steps in the origin of life was the evolution of oxygenous photosynthesis by ancestral cyanobacteria. The oxygenous photosynthesis in eukaryotes occurs on a photosynthetic organelle called plastid which arose via a process known as primary endosymbiosis (Marin, Nowack et al. 2005).

Primary endosymbiosis describes the process in which an ancestral free-living photosynthetic cyanobacterium was engulfed by a heterotrophic eukaryotic cell. During the co-evolution of the host cell and the endosymbiont the endosymbiont became an organelle called primary plastid (Deusch, Landan et al. 2008, Ochoa de Alda, Esteban et al. 2014). Primary plastids are surrounded by two membranes. The outer membrane acquired prokaryotic and eukaryotic characteristics during its evolution (Maier, Douglas et al. 2000, Stoebe and Maier 2002). This primary endosymbiosis gave rise to the three major groups glaucophytes, rhodophytes (red algae), and the chlorophytes (green algae and land plants) (Fig.1-1) (Cavalier-Smith 1998, Stoebe and Maier 2002, Adl, Simpson et al. 2005, Reyes-Prieto, Moustafa et al. 2008). Phylogenetic and gene cluster analysis suggests that these primary plastids are of monophyletic origin (Cavalier-Smith 2000, Martin, Rujan et al. 2002, Reyes-Prieto, Hackett et al. 2006, Rockwell, Lagarias et al. 2014). Recently, there is new evidence suggesting that photosynthetic chromotophores in the cercozoan amoeba *Paulinella chromatophora* derived from a different cyanobacterium (Bodył, Mackiewicz et al. 2010). This suggests that those chromotophores originated from an independent endosymbiosis (Bodył, Mackiewicz et al. 2010, Bodyl, Mackiewicz et al. 2012, Nowack and Grossman 2012).

During their evolution plastid genomes were greatly reduced. Based on comparison with the genome of free-living cyanobacteria (≥1.6 Mb in size) primary plastids have only about 100 - 200 kbp (Martin and Herrmann 1998, Stegemann, Hartmann et al. 2003, Reyes-Prieto, Moustafa et al. 2008). Through a natural and omnipresent process called endosymbiotic gene transfer (EGT) many genes of the symbiont's genome were transferred into the host nucleus (Martin and Herrmann 1998, Stegemann, Hartmann et al. 2003, Reyes-Prieto, Hackett et al. 2006, Reyes-Prieto, Moustafa et al. 2008). Martin et al. found that about 18% of the protein-coding genes in the nucleus originated from a cyanobacterium (Martin, Rujan et al. 2002). This is a hint that genes were transferred from the plastid to the host nucleus. Due to the EGT important host nuclear–encoded proteins have to be retargeted back to the plastids to assist important machineries and metabolic functions. For

example more than four hundred nuclear-encoded proteins were transferred back to the plastid in the glaucophytes *Cyanophora paradoxa* (Facchinelli, Pribil et al. 2013).

### 1.1.2 The secondary endosymbiosis

Primary plastids subsequently spread by secondary endosymbiosis (Deusch, Landan et al. 2008, Gould, Waller et al. 2008, Lane and Archibald 2008). Secondary or complex plastids arose by the uptake of a photosynthetic eukaryotic cell that evolved by primary endosymbiosis into a second eukaryotic cell (Cavalier-Smith 1998, Cavalier-Smith 2002).

Both green and red algae have been involved in secondary endosymbiotic events (Keeling 2009, Keeling 2013). The plastid evolution of the secondary emdosymbiosis is strongly debated. The Cabozoa hypothesis explains that secondary plastids of green algal origin trace back to a single common endosymbiosis (Cavalier-Smith 1999). However, this hypothesis was contradicted by other studies analyzing the phylogeny of plastid-encoded proteins. They suggested that these green plastids were acquired twice independently (Cavalier-Smith 2002, Yang, Elamawi et al. 2005, Green 2011). All red algal derived taxa contain chlorophyll c and are usually referred to as “chromalveolates”. Haptophytes, heterokonts, apicomplexans, cryptophytes and dinoflagellates are the major lineages in the chromalveolate group (Fig. 1-1). It has been proposed that one single secondary endosymbiosis with a red alga gave rise to the common ancestor of all chromaveolates (Cavalier-Smith 1999, Green 2011). This proposal is known as the chromalveolate hypothesis. Two chromalveolate genes (fructose bisphosphate aldolase (FBA) and glycerol-3-phosphate dehydrogenase (GAPDH)) have their unique evolutionary history, which supported the monophyletic origin of chromalveolates (Fast, Kissinger et al. 2001, Harper and Keeling 2003, Patron, Rogers et al. 2004). The single origin proposal is also supported by the plastid pigmentation in photosynthetic members, plastid gene and genome relationships, large multigene phylogenies of a concatenated 16-protein data set, an analysis using a 143-protein data set and the unity of SELMA-dependent protein import (Rodriguez-Ezpeleta, Brinkmann et al. 2005, Hackett, Yoon et al. 2007, Teich, Zauner et al. 2007, Burki, Inagaki et al. 2009, Hampl, Hug et al. 2009, Zimorski, Ku et al. 2014, Gould, Maier et al. 2015). But the chromalveolate hypothesis remains very controversial. Some studies based on genome-analysis are not consistent with this suggested monophyletic origin of the chromalveolates (Lane and Archibald 2008, Dagan and Martin 2009, Moustafa, Beszteri et al. 2009, Dorrell and Smith 2011). There are different datas indicate that the host cells are not monophyletic. Such as it was shown that heterokonts are more closely related to alveolates than to haptophytes and cryptophytes on trees, gene replacement in the plastid genomes of the defined monophyletic groups, one-third of the proteins in the biosynthetic pathway of carotenoids in chromists are closely

related to green algal homologs via phylogenetic analyses (Frommolt, Werner et al. 2008). The origin of red complex plastids is still debating. Recently, Petersen et al. suggested the rhodophycean origin for the complex plastid *Chromera velia*, which departs from a single origin of the red complex plastid (Petersen, Ludewig et al. 2014).

In comparison to primary plastids, secondary complex plastids are surrounded by additional membranes (Maier, Douglas et al. 2000, Gould, Waller et al. 2008). In the case of organisms with a green algal endosymbiont, the plastids can be surrounded by three (euglenophytes) or four membranes (chlorarachniophytes). Important members with a red algal endosymbiont are haptophytes, heterokonts, cryptophytes and apicomplexa, all of them with secondary plastids surrounded by four membranes (Stork, Lau et al. 2013). Melkonian suggested that the outermost membrane is an autophagosomal membrane (Melkonian 1996). The outermost membrane of complex plastids is also thought to be homologous to a phagocytotic membrane of the host (Cavalier-Smith 2000, Bolte, Bullmann et al. 2009). After fusion with the endoplasmic reticulum (ER), the phagocytotic membrane became the chloroplast ER membrane (cER) in heterokonts, haptophytes and cryptophytes (Lemgruber, Kudryashev et al. 2013). The second outermost membrane of complex plastids, also called periplastidal membrane (PPM), is thought to be homologous to the cytoplasmic membrane of the endosymbiont (Cavalier-Smith and Chao 2003). It was also suggested that the host ER membrane was involved to form the two outermost membranes (Zimorski, Ku et al. 2014, Gould, Maier et al. 2015). The two innermost membranes of complex plastids appear to originate from the inner and outer envelope of primary plastids (Schleiff and Becker 2011, Stork, Lau et al. 2013). However, a different situation is found in peridinin-containing dinoflagellates the complex plastids of red algal origin are surrounded by only three membranes (Cavalier-Smith 2000). The evolution of dinoflagellates is described in detail in the next paragraph. In chromalveolate lineages cryptophytes are the only member demonstrated to have retained the nucleomorph (the former nucleus of the engulfed red alga) between the outer and inner chloroplast membrane pair (Archibald 2007).

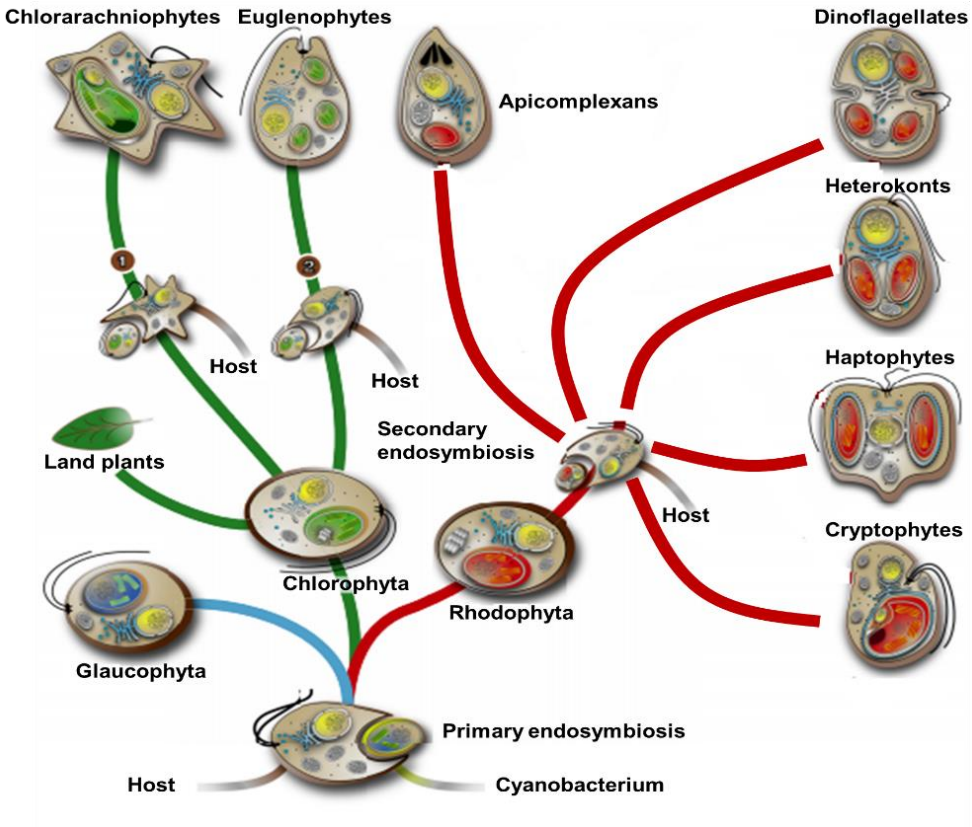


**Fig. 1-1: Model of the evolution of complex plastids of green and red algal origin.**

During primary endosymbiosis an ancestral free-living photosynthetic cyanobacterium was engulfed by a eukaryotic cell and established as an endosymbiont in the three lineages glaucophyta, rhodophyta and chlorophyta. Engulfment of a green or red alga by another unicellular eukaryote and subsequent reduction of the symbiont to an organelle led to the development of so-called secondary or complex plastids. Two independent secondary endosymbiotic events involving algae of the chlorophyta and different eukaryotic hosts resulted in the chlorarachniophytes (1) and euglenophytes (2). The red algal derived taxa collectively are usually referred to as “chromalveolates”. Haptophytes, heterokonts, apicomplexans, cryptophytes and dinoflagellates are the major lineages in the chromalveolate group, but this is strongly discussed. It is unknown if red secondary plastids are mono- or polyphyletic. (Modified after Zimorski (Zimorski, Ku et al. 2014)).

Dinoflagellates are an extremely diverse group of photosynthetic species and non-photosynthetic species. Peridinin-containing dinoflagellates are distinguished from the other species in dinoflagellates by the pigment peridinin and three-membrane surrounded plastids (Dodge and Lee 2000). During the evolution one of the ancestral four membranes must have been lost in dinoflagellates (Zhang, Green et al. 1999, Barbrook and Howe 2000, Zhang, Green et al. 2000, Stoebe and Maier 2002). Gould et al. suggested the plastids have probably lost the second outermost of four membranes (Gould, Maier et al. 2015). It was shown that the membranes of the peridinin-containing plastid are not connected to the host ER and the outermost membrane does not have ribosomes (Gould, Maier et al. 2015).

The evolutionary origin of peridinin-containing plastids in dinoflagellates is also highly debated because of the reduced number of plastid membranes. Previous studies suggested that the peridinin-containing plastids arose from secondary endosymbiosis (Mcfadden 2001, Grosche, Hempel et al. 2014, Gruber, Rocap et al. 2015). However, several hypothesizes indicate that the plastids of dinoflagellates evolved via a more complicated tertiary endosymbiosis, where a secondary plastid-containing alga was engulfed and reduced to an organelle (Hackett, Yoon et al. 2004, Keeling 2010, Gabrielsen, Minge et al. 2011, Burki, Imanian et al. 2014). Dinoflagellates have been shown to contain heterkontophyte-, cryptophyte- and haptophyte-derived tertiary plastids in the so-called “Dinotoms”, Dinophysis and Kareniaceae groups, respectively (Burki, Imanian et al. 2014).

### 1.1.3 Plastid genome and gene transfer

More than 25 years ago the first complete sequence of a plastid genome was reported in the liverwort *Marchantia polymorpha* (Ohyama, Fukuzawa et al. 1986). There are more and more plastid genomes available now, including those of land plants and algae. The plastid genomes of algae and land plants usually are single circular DNA molecules with a size of 100 - 200 kbp and contain approximately 100 - 250 genes (Barbrook and Howe 2000, Nisbet, Hiller et al. 2008). Encoded on these genomes are many components of the photosynthesis, chloroplast replication and protein synthesis machinery (Hiller 2001).

In contrast to the common genome organization found in most eukaryotic groups harboring plastids, peridinin-containing dinoflagellates have a unique degenerated and rearranged plastid genome (Zhang, Green et al. 1999, Barbrook and Howe 2000, Laatsch, Zauner et al. 2004, Takishita, Ishida et al. 2004). After the disintegration and reconfiguration of their plastid genome, the genes are located to plasmid-like molecules of a size of about 0.4 - 10kbp instead of a conventional plastid chromosome genome. These molecules are called minicircles (Zhang, Green et al. 1999, Barbrook and Howe 2000, Howe, Nisbet et al. 2008). Each minicircle contains one to three genes as well as a non-coding region. Most encoded proteins are related to photosynthesis. Additionally many ‘empty’ minicircles have been identified. Although they contain a number of open reading frames of different lengths no significant homology could be detected, thus leaving their purpose unknown (Hiller 2001, Nisbet, Koumandou et al. 2004). Based on the alignment of the non-coding regions of all known minicircles within species a highly conserved region called core region was found. The core region is thought to be responsible for the initiation of replication, the maintenance of the copy number and/or the transcription of minicircles (Zhang, Green et al. 1999, Barbrook and Howe 2000, Hiller 2001, Zhang, Cavalier-Smith et al. 2002, Barbrook, Dorrell et al. 2012). Additionally to the core

and coding region minicircles contain a non-coding region. It has been shown that the non-coding regions possess a tripartite conserved sequence that could be folded *in silico* into secondary structures (Howe, Nisbet et al. 2008). Only about 15 genes have been found on dinoflagellate minicircles (varying on species) until now (Zhang, Green et al. 1999, Barbrook and Howe 2000, Hiller 2001, Nisbet, Koumandou et al. 2004, Howe, Nisbet et al. 2008). Compared to the size of other plastid genomes the plastid genome of dinoflagellates is undoubtedly the smallest (Green 2004, Green 2011). This shrunken plastid genome might result from the missing of numerous genes. Some authors suggested that these missing genes were either deleted or transferred to the host genomes during the process of plastid acquisition (Bachvaroff, Concepcion et al. 2004, Green 2004, Hackett, Yoon et al. 2004, Bachvaroff, Sanchez-Puerta et al. 2006).

As already mentioned during the previous endosymbiosis as well as the following (secondary or even tertiary) endosymbiosis there was a massive transfer of endosymbiotic genes to the host nucleus known as EGT.

However, the reduced plastid genome is insufficient to meet the wide variety of plastid functions such as photosynthesis, lipid and protein biosynthesis. Because of this, the proteins had to be somehow targeted to the plastids. It was shown that nucleus-encoded proteins required for plastidal functions were transported back to the plastids with the aid of targeting signals and a protein-import machinery (Martin and Herrmann 1998, Weber, Linka et al. 2006).

## 1.3 Endomembrane system

### 1.3.1 The endomembrane system

In eukaryotic cells the endomembrane system is made up of several functionally different organellar membranes, including the nuclear envelope, ER, Golgi apparatus, lysosomes or vacuoles, different vesicles (more detail in 1.3.2.2) and the plasma membrane (PM) (Fig.1-2). These organelles are located in the cytoplasm and interconnected by vesicular transport (Schellmann and Pimpl 2009). This membrane system divide the cell into functional and structural compartments (Galili 2001, Gautreau, Oguievetskaia et al. 2014) . It is necessary for proteins to be transported to the right destinations.

The endomembrane compartments fulfill specific functions in the transport of proteins. The endoplasmic reticulum (ER) has a central role in producing, processing and transporting proteins. The Golgi apparatus is another important organelle. The Golgi consists of three networks, cis-Golgi network mainly receive proteins from ER, medial-Golgi transport proteins from cis-Golgi to trans Golgi, while the trans-Golgi or trans Golgi network (TGN) send proteins to next organelles. During

the transport of proteins from the ER to the Golgi apparatus, proteins are modified, sorted and finally transported to destinations such as the vacuole, the extracellular space (so called anterograde transport) or recycled back to the Golgi apparatus (so called retrograde transport) (Bolte, Brown et al. 2004). Endosomes are single membrane-surrounded compartments and represent a major and important sorting compartment in the endomembrane system. Endosomes can be classified into early endosomes, late endosomes and recycling endosomes depending on their main functions (Reyes, Buono et al. 2011). But in most cases it is difficult to distinguish these endosomes (Otegui and Spitzer 2008). The trans-Golgi network (TGN) is believed to serve as an early endosome (Mallard, Tang et al. 2002, Dettmer, Hong-Hermesdorf et al. 2006, Viotti, Bubeck et al. 2010). The main function of TGN/early endosomes is to mediate vacuolar protein transport, receive internalized proteins from the plasma membrane or retransport endocytosed proteins back to the plasma membrane and recycle protein sorting receptors (here also called recycling endosomes) (Mallard, Tang et al. 2002, Otegui and Spitzer 2008). Subsequently, early and recycling endosomes are believed to mature into late endosomes or multi vesicular bodies (MVBs), also called prevacuolar compartments (PVCs) (Reyes, Buono et al. 2011). Late endosomes have two important functions, recycling of protein sorting receptors back to the TGN and transporting of newly synthesized proteins from the Golgi apparatus to vacuoles (Piper and Katzmann 2007, Otegui and Spitzer 2008). It is obvious that the endosomal network is a key compartment that mediates the concentration of proteins in the Golgi apparatus, vacuoles and plasma membrane (Tse, Mo et al. 2004, Otegui and Spitzer 2008).

The most visible compartment surrounded by only one membrane is the vacuole (Maruyama, Mun et al. 2006, Zouhar and Rojo 2009). Vacuoles are involved in maintaining of internal hydrostatic pressure and storage of water, ions, nutrients and secondary metabolites. Similar to animal lysosomes, they also act in intracellular digestion of various waste products and toxic substances (Matsuoka 1993, Otegui and Spitzer 2008, Pereira, Pereira et al. 2013, Ebine, Inoue et al. 2014). In plant cells, at least two different types of vacuoles can be found, the central lytic vacuole (LV) for protein degradation and the protein storage vacuole (PSV) (Swanson, Bethke et al. 1998, Frigerio, Jolliffe et al. 2001, Frigerio, Hinz et al. 2008).

### 1.3.2 Vacuolar protein transport within endomembrane system

Proteins that are secreted to the extracellular space or retained in the endosomal membrane system are commonly called secretory proteins (Vitale and Hinz 2005). Most secretory proteins contain an ER signal peptide or a hydrophobic transmembrane domain for insertion into the ER membrane (Johnson and van Waes 1999, Xiang, Etxeberria et al. 2013). When the signal peptide is cleaved off

from the proteins and then other specific targeting signals are used for transporting proteins to the destinations (Jürgens 2004, Xiang, Etxeberria et al. 2013, Xiang and Van den Ende 2013). Therefore, the targeting signals and correct transport routes are important for the targeting of the vacuolar protein to the final destination.

#### 1.3.2.1 Vacuolar sorting determinants

Protein sorting to the precise destinations depends on many targeting information from the protein. The identification of vacuolar sorting signals/vacuolar sorting determinants (VSS/VSDs) is necessary for better understanding of protein targeting to the vacuoles at the post-Golgi level. Three distinct sorting determinants are known in plant cells. Sequence-specific VSDs (ssVSDs), C-terminal VSDs (ctVSDs) and protein structure-dependent VSDs (psVSDs) have been reported (Hwang 2008, Hegedus, Coutu et al. 2015). Without these important signals vacuolar proteins will be targeted to the wrong subcellular compartments or degraded.

Many studies have shown that sequence-specific VSDs are usually located at the N-terminus of the proteins and mainly sort them to the lytic vacuoles, such as barley aleurain (Koide, Matsuoka et al. 1999, Sanmartin, Ordonez et al. 2007). In some cases sequence-specific VSDs are also recognized by vacuolar sorting receptors and guide proteins to the PSVs, such as seed storage protein 2S albumin and toxin ricin (Frigerio, Jolliffe et al. 2001, Jolliffe, Brown et al. 2004).

C-terminal VSDs are always located at the C-terminal part of the protein. They generally lead proteins target to the protein storage vacuoles (Nishizawa, Maruyama et al. 2003, Vitale and Hinz 2005, Maruyama, Mun et al. 2006). Finally psVSDs are based on the tertiary (physical) structure of proteins and usually transport proteins to the protein storage vacuole (Neuhaus and Rogers 1998, Jolliffe, Brown et al. 2004, Jolliffe, Craddock et al. 2005, Neuhaus 2007, Zouhar and Rojo 2009). Recently some studies have shown that some proteins contain two different VSDs at the same time. For example, the existence of a sequence-specific VSD and a ctVSD at the C-terminal region of α' subunit of soybean β-conglycinin directs the protein to the PSVs (Nishizawa, Maruyama et al. 2003). The protein cardosin A with a characteristic ctVSD and an untraditional vacuolar sorting domain (PSI) was transported to the vacuole. Plant specific domain (PSI) is an additional protein domain of about 100 amino acids present in the protein precusors and is normally deleted after maturation (Simoes and Faro 2004). It was also demonstrated that any one of these two signals alone is sufficient to sort proteins to the vacuole (Pereira, Pereira et al. 2013).

However, Pompa et al shown that the vacuolar phaseolin is secreted to the apoplast without the C-terminal tetrapeptide AFVY. When a cysteine residue was added to the phaseolin for forming interchain disulfide bond, the phaseolin polypeptides connected by engineered disulfide bonds are

transported to vacuoles again (Pompa, De Marchis et al. 2010). Nishizawa et al. proposed that the proportion between the size of the protein and the copy number of vacuolar sorting determinants is also very important for the sorting efficiency (Nishizawa, Maruyama et al. 2003). Based on these data in protein sorting to the vacuole, it is clear that the three well-known types of VSDs identified so far is not thoroughly sufficient to explain all the vacuolar targeting mechanisms.

#### 1.3.2.2 Vacuolar protein transport via vesicles

Vacuolar protein transport is mediated by different intermediate vesicles. Previous studies have shown that most of newly synthesized proteins are transported from ER to the Golgi apparatus via a coat protein II (COPII)-vesicle-dependent manner (Ritzenthaler, Nebenführ et al. 2002, Takeuchi, Ueda et al. 2002, daSilva, Snapp et al. 2004, Yang, Elamawi et al. 2005). Different types of ER export motifs have already been identified for exiting the ER (e.g. a di-basic motif RKR in tobacco) (Barlowe 2003, Hanton, Renna et al. 2005, Lee and Miller 2007, Lee and Miller 2007, Schoberer, Vavra et al. 2009). ER export motifs are recognized by the known cargo-binding site on Sec24 component of COPII complex, and then the protein will be captured into COPII vesicles and transported to the next organelle (Miller, Beilharz et al. 2003, Mossessova, Bickford et al. 2003, Lee and Miller 2007). COPI was identified in mediating retrograde traffic from the cis-Golgi to the ER. Proteins with ER retention signals are mistargeted to the Golgi. These proteins can be retransported back to the ER in COPI vesicles (Pimpl, Taylor et al. 2006). It was also shown that COPI adjusts the vesicular traffic pathway within the Golgi apparatus (Movafeghi, Happel et al. 1999, Contreras, Ortiz-Zapater et al. 2000, Pimpl, Movafeghi et al. 2000, Pimpl, Hanton et al. 2003).

In plants different routes were used for the vacuolar protein transport after the ER (Fig. 1-2). One of the possibilities involves the passage through the Golgi apparatus. Previous studies have shown that transport of some important lytic vacuolar proteins from the Golgi apparatus to the lytic vacuoles (LVs) is mediated by clathrin-coated vesicles (CCVs) with a diameter of 50 to 70 nm (Hohl, Robinson et al. 1996, Robinson, Hinz et al. 1998, De Marchis, Bellucci et al. 2013). Specific vacuolar targeting signals are recognized by protein sorting receptors in the membrane of the TGN (Rouillé, Rohn et al. 2000). Subsequently the protein sorting receptors are recognized by adaptor proteins (Xiang, Etxeberria et al. 2013, Xiang and Van den Ende 2013). The ligand-receptor complexes are recruited into the clathrin-coated vesicles (Ahmed, Rojo et al. 2000, Kalthoff, Groos et al. 2002, Happel, Höning et al. 2004, Song, Lee et al. 2006). CCVs containing proteins bud from the trans Golgi network, then deliver their cargoes to LVs after fusion with multi vesicular bodies (Fig. 1-2 marked by black arrow) (Robinson and Bonifacino 2001, Tse, Mo et al. 2004, Van Damme, Gadeyne et al. 2011). Clathrin-coated vesicles are one major style of vesicles in plant. Clathrin-coated vesicles can also bud

from the plasma membrane for the uptake of nutrients, the delivery and regulation of signaling components as well as the recycling of proteins (Maldonado-Mendoza and Nessler 1996, Holstein 2002, McMahon and Boucrot 2011, Reynolds, August et al. 2014).

Transport of lytic vacuolar proteins to the LV can also take an alternative pathway from the ER bypassing the Golgi, as shown in Fig. 1-2 (marked by purple arrow). This process is mediated by intermediate compartments, called ER bodies (Matsushima R 2003, Hara-Nishimura, Matsushima et al. 2004). When the vacuolar sorting motifs are recognized by protein sorting receptors, ligand-receptor complexes are recruited into the ER bodies. The ER bodies (diameter of less than 1000 nm) are oil, protein or rubber containing ER-derived membrane structures (Herman and Schmidt 2004, Herman 2008, Xiang, Etxeberria et al. 2013). Therefore, the lytic vacuole sorting pathway could be defined as ER → Golgi → CCV → multi vesicular body → LV and ER → ER body → LV pathways.

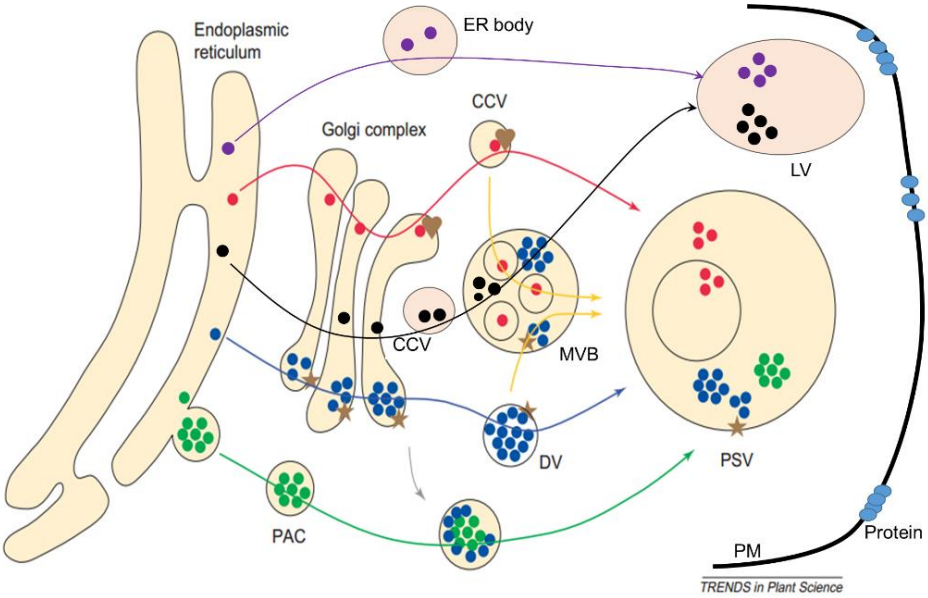


**Fig. 1-2: A working model for protein sorting to vacuoles.**

Many different sorting routes are known for proteins transport to vacuoles. During the transport of lytic vacuolar proteins, proteins are directly targeted from the ER to their destination in ER bodies bypassing the Golgi apparatus (marked by purple arrow). Proteins are transported from the Golgi apparatus to the lytic vacuole (LV) via prevacuolar compartments or multivesicular bodies (MVBs). This pathway is mediated in clathrin-coated vesicles (CCVs) (indicated by black arrow). During the transport of storage proteins, proteins are sorted to protein storage vacuole (PSV) by a receptor-mediated way (brown heart) in clathrin-coated vesicles (red arrow). Some proteins are transported to PSV by a receptor-mediated pathway (brown star) via dense vesicles (DVs) (blue arrow). The multi vesicular body (MVB) might be an alternative intermediate compartment during the protein transport from CCVs and DVs to PSV (yellow arrows). Some proteins bypass the Golgi apparatus using precursor-accumulating (PAC) vesicles (green arrow). These vesicles also receive proteins from the Golgi apparatus. PM: plasma membrane. (Modified after Vitale and Hinz 2005)

However, the targeting of proteins from the ER to the protein storage vacuoles is different and more complicate. It is majorly mediated by dense vesicles (DVs). Different from the lytic protein transport the interaction of ligands and receptors is calcium-dependent in storage vacuolar protein transport (Shimada, Fuji et al. 2003). DVs are small vesicles with a diameter of 150 - 200 nm. Vacuolar proteins are recruited into the mature DVs released from the TGN. Mature DVs are not involved in a protein coat (Hohl, Robinson et al. 1996). Finally the DVs deliver their cargoes to the PSVs (marked by blue arrow) (Vitale and Hinz 2005). All available evidences indicate that some seed storage proteins, such as most 2S ablumins and the toxin ricin, would also be delivered to protein storage vacuoles via clathrin-coated vesicles (marked by red arrow) (Vitale 2001, Jolliffe, Brown et al. 2004, Vitale and Hinz 2005). MVBs or PVCs are the alternative intermediate target of CCVs and DVs during the transport route from the Golgi apparatus to PSVs (marked by yellow arrow) (Robinson, Hinz et al. 1998, Jiang, Phillips et al. 2000, Tse, Mo et al. 2004) (Fig. 1-2).

Studies have also shown that vacuole residing proteins are sorted directly from the ER to the PSVs using precursor-accumulating (PAC) vesicles and are therefore transported in a Golgi-independent way (Hara-Hishimura, Takeuchi et al. 1993, Vitale 2001, Michaeli, Avin-Wittenberg et al. 2014). There is an alternative route. Some vacuolar proteins could be transported from the Golgi apparatus to the PAC vesicles and then target to the PSV (marked by green arrow) (Watanabe, Shimada et al. 2004). Therefore, the protein storage vacuole sorting pathway could be mainly defined as ER → Golgi → DV/CCV → (multi vesicular body) → PSV and ER → PAC → PSV pathways.

Receptor-mediated sorting pathways for secretory proteins in eukaryotic cells depend on mechanisms to recycle the receptors after the dissociating of receptor-ligand complexes (Niemes, Langhans et al. 2010). The recycling of receptors are mediated by different complexes such as retromer. The retromer is a coat complex that locates on the cytosolic face of the TGN/early endosome (Vergés, Sebastián et al. 2007, Schellmann and Pimpl 2009, Seaman, Harbour et al. 2009). In yeast, plants and mammals the retromer complex contains two subcomplexes, a large subcomplex formed by three cargo selective adaptor subunits (Vps26, Vps29 and Vps35) and a small subcomplex formed by membrane deforming sorting nexin proteins (SNXs) (Seaman 2004, Bonifacino and Hurley 2008, Cullen and Korswagen 2012). Most newly synthesized proteins are transported via a protein sorting receptors-dependent parthway. The protein sorting receptors have been well-characterized in some organisms such as the seed storage protein receptors in *A. thaliana*, vacuolar protein sorting 10 (Vps10) in yeast *S. cerevisiae* and mannose 6-phosphate receptors (MPR) in mammals (Horazdovsky, Davies et al. 1997, Nothwehr, Bruinsma et al. 1999, Arighi, Hartnell et al. 2004, Carlton, Bujny et al. 2004, Seaman 2004, Niemes, Langhans et al. 2010, McGough and Cullen 2011, Robinson, Pimpl et al. 2012). In order to maintain the forward transport of proteins the

efficient retrograde transport of the protein sorting receptor is a critical important step. This recycling of receptors is mediated by some complexes such as the retromer. Moreover, the retromer is also responsible for many other physiological and developmental processes such as mediating the transport of internalized toxin and auxin efflux carriers (more diverse functions of retromer see Bonifacino and Hurley 2008).

### 1.3.3 The biosynthesis of N-Glycoproteins on endomembrane system

Glycosylation is a very important translational modification for proteins. There are different glycosylations such as N-glycosylation, O-glycosylation and phosphor-glycosylation. N-glycosylation is a major co- and post-translational modification of proteins in eukaryotic cells. It has been discussed that N-glycosylation plays a crucial role in the folding, assembly, structural formation and stability of several important proteins (Rayon, Lerouge et al. 1998, Baïet, Burel et al. 2011, Mathieu-Rivet, Kiefer-Meyer et al. 2014). Moreover previous studies have shown that the glycosylation plays a role in the sorting of proteins to the vacuoles, especially in the Golgi-independent secretory pathway (Rayon, Lerouge et al. 1998, Paris, Saint-Jean et al. 2010, Pereira, Pereira et al. 2013, Pereira, Pereira et al. 2014).

It has been discussed that the majority of secretory proteins are N-glycosylated in the endoplasmic reticulum (ER) and the Golgi apparatus (Fig. 1-3) (Strasser 2014). The process of N-glycosylation can be mainly divided into three stages: during the initial phase of the process a precursor oligosaccharide is synthesized in the ER and transferred en bloc to the protein by a oligosaccharyltransferase, which is part of a translocation complex (Baiet, Burel et al. 2011, Kajiura, Okamoto et al. 2012, Bosch, Castilho et al. 2013, Mathieu-Rivet, Kiefer-Meyer et al. 2014), the resulting N-glycoproteins are transported to the Golgi apparatus and further catalysed and modified by a large number of highly conserved membrane-bound enzymes, such as glycosyltransferases and glycosylhydrolases, such as N-acetylglucosaminyltransferase I (GnTI), β1, 2-xylosyltransferase (β1, 2-XylT), α1, 3-fucosyltransferase (α1, 3-FucT), α1, 4-fucosyltransferase (α1, 4-FucT) and so on (Kornfeld and Kornfeld 1985, Strasser, Mucha et al. 2000, Wilson 2002, Bondili, Castilho et al. 2006, Strasser, Bondili et al. 2007, Strasser 2014), finally the mature glycoproteins are either secreted or targeted into the plasma membrane.

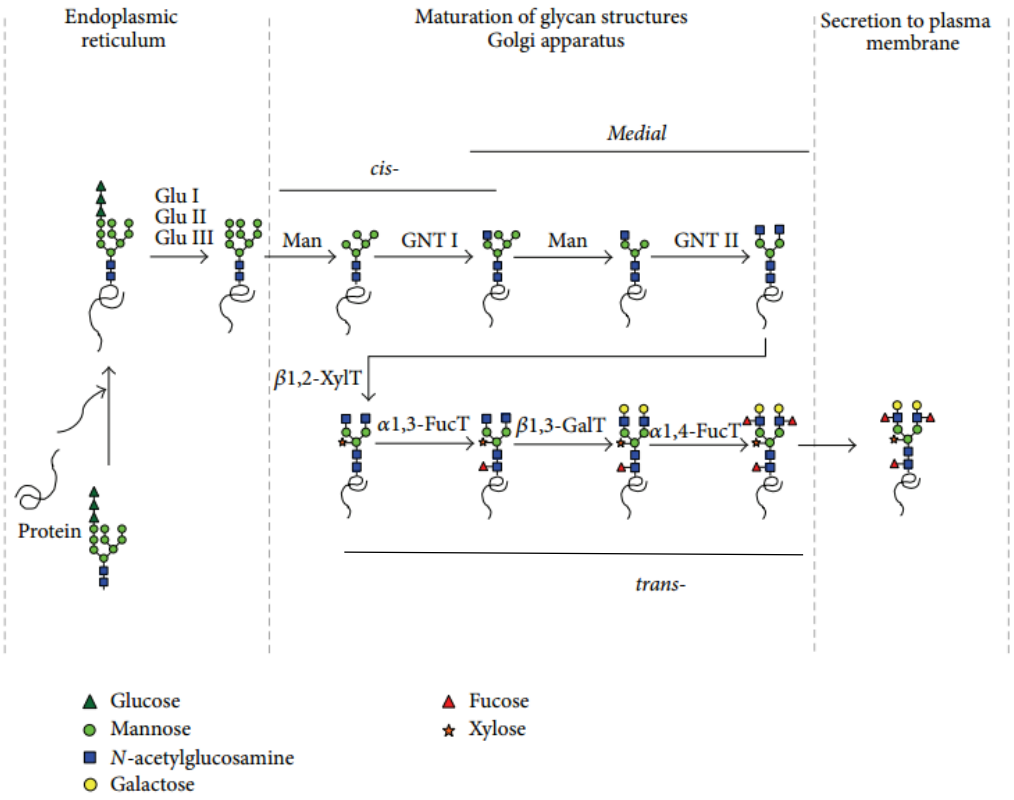


**Fig. 1-3: N-glycosylation pathway during the synthesis of glycoproteins in plants.**

The synthesis, en bloc transfer and initial modification of precursor oligosaccharide occur in the ER under the catalytic action of Glu I/II/III enzymes. Subsequently, the processing and modification of the oligosaccharide chain is performed in the cis- medial- and trans-Glogi apparatus. At last, the mature glycoproteins are either targeted into the plasma membrane or secreted. Glu I: glucosidase I, Glu II: glucosidase II, Glu III: glucosidase III, Man: mannosidase, GnTI: N-acrtylglucosaminyltransferase I, GnTII: N-acrtylglucosaminyltransferase II, β1, 2-XylT: β1, 2-xylosyltransferase, α1, 3-FucT: α1, 3-fucosyltransferase, α1, 4-FucT: α1, 4-fucosyltransferase, β1, 3-GalT: β1, 3-galactosidase. (Modified after Hyun-Soon K. Jae-Heung J. et al 2014)

### 1.3.4 Tonoplast intrinsic proteins (Tips)

Some membrane proteins do not have a direct effect on the protein sorting and transport, but they are important for the cellular metabolism. Aquaporins are channel proteins belong to the the major intrinsic proteins (MIPs) superfamily. Aquaporin pore can selectively mediate the transport of water, gases and small neutral solutes such as urea, glycerol as well as silicic acid (Maurel 1997, Loque, Ludewig et al. 2005, Lienard, Durambur et al. 2008, Uehlein and Kaldenhoff 2008, Wudick, Luu et al. 2014). In most plant species aquaporins can be subdivided into four groups, the plasma membrane intrinsic proteins (PIPs), nodulin 26-like intrinsic proteins (NIPs) localized in the plasma membrane and ER, the tonoplast intrinsic proteins (TIPs) and small basic intrinsic proteins (SIPs) localized in the

ER (Johansson, Karlsson et al. 2000, Baiges, Schaffner et al. 2002, Quigley, Rosenberg et al. 2002, Pandey, Sharma et al. 2013).

Aquaporins have been well-studied on the structure and function in plants, yeast and animals (Kaldenhoff, Ribas-Carbo et al. 2008, Maurel and Plassard 2011, Li, Santoni et al. 2014). Aquaporin proteins contain six conserved transmembrane-spanning α-helices linked by three extra- and two hydrophobic intracellular loops (Kaldenhoff and Fischer 2006, Fischer and Kaldenhoff 2008, Chaumont and Tyerman 2014). The N-terminus and C-terminus of aquaporin proteins are located at the cytosolic side of the membrane (Johansson, Karlsson et al. 2000, Uehlein and Kaldenhoff 2008). It has been shown that the first three transmembrane domains and the rest three transmembrane domains form an inversely repeat (Kaldenhoff, Ribas-Carbo et al. 2008, Chevalier and Chaumont 2014). Previous studies have already shown that the loops play a critical role in the forming of transmembrane channels and its stability (Werner, Uehlein et al. 2001, Kaldenhoff, Bertl et al. 2007, Kaldenhoff, Ribas-Carbo et al. 2008). Two conserved asparagine-proline-alanine (NPA) motifs were located in the different loop regions. The NPA motif and adjacent residues were thought to be critically important for water transport activity (Chaumont, Barrieu et al. 2001, Park, Scheffler et al. 2010).

### 1.3.5 Vacuolar-type $H^+$-ATPases (V-ATPase)

The ATPase is an important multi-subunit transmembrane complex (Clarke, Köhler et al. 2002, Dettmer, Hong-Hermesdorf et al. 2006). There are different classes of ATPases (F-ATPases, V-ATPases, E-ATPases, P-ATPases and A-ATPases), which can be distinguished by their function, structure and the transport of ions (McKersie and Bruce 2009, Xi and Wu 2011, Islam, Patwary et al. 2014). It has been shown that V-ATPases are mainly accumulated at vacuoles, ER, Golgi apparatus, plasma membrane and endosomes (Sze, Schumacher et al. 2002, Krebs, Beyhl et al. 2010)(Fig. 1-4). V-ATPase complexes contain a cytosolic subcomplex V1 (subunits A-H) and hydrophobic subcomplex V0 (subunits a, c, c', c'', d and e) (Dettmer, Hong-Hermesdorf et al. 2006, Seidel, Schnitzer et al. 2008, Ma, Qian et al. 2012). The major function of this complex is to pump protons from the cytoplasm into the lumen of organelles or to the outside of the cell. The V-ATPase catalyzes the hydrolysis of ATP into ADP and a free phosphate ion (Ma, Qian et al. 2012). During the catalytic process a lot of energy is released which is used for the cell metabolism including endosomal transports (Dettmer, Liu et al. 2010, Zhou, Bu et al. 2015). In *Puccinellia tenuiflora* decreasing of the V-ATPase activity obstructs endosomal trafficking (Zhou, Bu et al. 2015). It was also shown that the inhibition of VHA interrupt the transport from TGN to the vacuole as inhibitor concanamycin prevent the TGN formation and release fron the Golgi apparatus (Scheuring, Viotti et al. 2011).

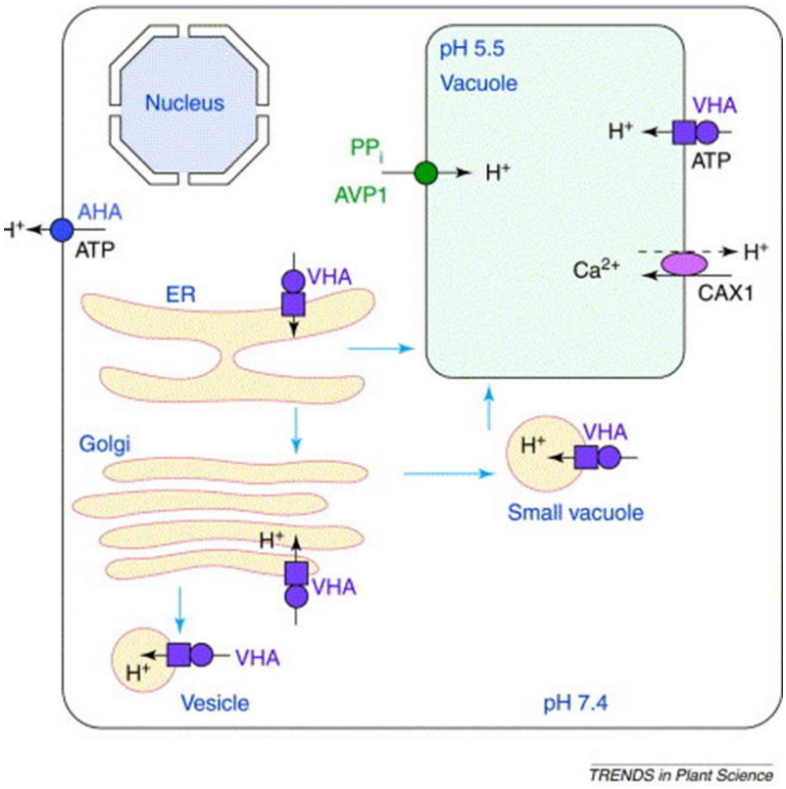


**Fig. 1-4: Subcellular localization of vacuolar type $H^+$-ATPase (VHA) in plant.**

It was shown that vacuolar-type $H^+$-ATPases (VHAs) are mainly distributed to the ER, Golgi apparatus, vesicles, vacuoles and plasma membrane. The major role of VHAs is to pump $H^+$ into the lumen of organelles or out of the cell and provide energy for cell metabolites. Abbreviations: AHA, *A. thaliana* plasma membrane H+-pumping ATPase; AVP1, vacuolar $H^+$-pumping ppase; CAX1, $Ca2^+$/ $H^+$ antiporter. (Modified from Sze, Schumacher et al. 2002)

# 2. Aim

Peridinin-containing dinoflagellates are unicellular alveolates which evolved via secondary endosymbiosis event. The reduction of the endosymbiont genome in addition to gene transfer from the plastid to the nucleus resulted in a unique endosymbiont genome organization in peridinin-containing dinoflagellates. Genes normally found on plastid genomes have been organized into so-called minicircles. Previous studies have already reported several minicircle sequences from different dinoflagellates (Barbrook and Howe 2000, Hiller 2001, Zhang, Cavalier-Smith et al. 2002, Moore 2003, Laatsch, Zauner et al. 2004, Nisbet, Koumandou et al. 2004, Barbrook, Santucci et al. 2006, Howe, Nisbet et al. 2008). However, it remains limit on the isolation of the single minicircle molecule. The aim of this work was to isolate individual minicircle molecules via novel transposon-based approach from the dinoflagellate *A. carterae* CCAM0512. Based on these individual minicircle molecules the characteristics and evolutionary relationship of minicircles were analysed. In order to investigate the localization of minicircles individual minicircle molecules were manipulated for retransfection of *A. carterae* CCAM0512.

Proteins are usually synthesized on the rough endoplasmic reticulum (rER), and are then sorted to different destinations (e.g. vacuole) through the endomembrane system by the secretory pathway. The protein transport and sorting mechanisms (such as the sorting routes, targeting signals and vacuolar protein sorting receptors) have already been well-studied in plant cells, yeast and animals (Geuze 1995, Bonifacino and Traub 2003, Pereira, Pereira et al. 2013, Xiang, Etxeberria et al. 2013, Pereira, Pereira et al. 2014, Zhang, Hicks et al. 2014). Little is known about the protein transport and sorting through the endosomal compartments in diatom *P. tricornutum*. The aim of this work was to gain insight into the sorting mechanisms of protein with targeting signals. Defining appropriate marker proteins for subcellular organelles are essential for studying the mechanisms. However, the known marker proteins in *P. tricornutum* are less enough. To extend the dataset of marker proteins homologous proteins to known localization proteins in plants were identified in *P. tricornutum*. To investigate the subcellular localization of candidate marker proteins eGFP- or mRFP-fusion proteins were studied *in vivo* localization by confocal laser-scanning microscopy.

# 3 Results

## 3.1 Genetic compartmentalization in the complex plastid of *Amphidinium carterae*

Peridinin-containing dinoflagellates are unicellular alveolates that evolved via secondary endosymbiosis (Mcfadden 2001, Shalchian-Tabrizi, Skånseng et al. 2006). The reduction of the endosymbiontic genome in addition to gene transfer from the plastid to the nucleus led to a unique endosymbiont genome organization in peridinin-containing dinoflagellates (Zhang, Green et al. 1999, Barbrook and Howe 2000, Howe, Barbrook et al. 2003). In these organisms genes normally found on the conventional plastid genome have been organized into so-called minicircles (Barbrook and Howe 2000, Hiller 2001, Zhang, Cavalier-Smith et al. 2002, Howe, Nisbet et al. 2008). In order to retrieve their full-length DNA sequence a novel transposon-based approach was used to isolate individual minicircle molecules from the peridinin-containing dinoflagellate *A. carterae* CCAM0512.

### 3.1.1 The enrichment and isolation of minicircles

It was already shown by Laatsch et al. that a high quantity of minicircles could be enriched and isolated by the alkaline lysis (Laatsch, Zauner et al. 2004). Here in order to enrich and isolate minicircles from *A. carterae* CCAM0512, additionally to the alkaline lysis different suppliers for alkaline lysis based kits were tried (see material and method 5.2.4.11). It was found that a significant quantity of minicircles could only be enriched and isolated via the alkaline lysis but not alkaline lysis based kits. Koumandou and Howe have already shown that the copy number of different minicircles per cell is low during the exponential growth stage, but the number is increasing during the later growth phase (Koumandou and Howe 2007). It was found that the significant minicircles could only be visible on an agarose gel via the enrichment and isolation from the old cultures about four to five weeks but not younger cultures.

The isolation of minicircles was shown on Fig. 3-1. A strong signal was detectable at a size of about 2000 bps (lane 1 and lane 2), this resulted from the isolated minicircles. The gel of red box area was retrieved and used for the isolation of individual minicircles by a transposon-insertion based approach (see material and method 5.2.4.12). An additional signal was marked by red stars, representing the gDNA was detectable in all lanes (Fig. 3-1).

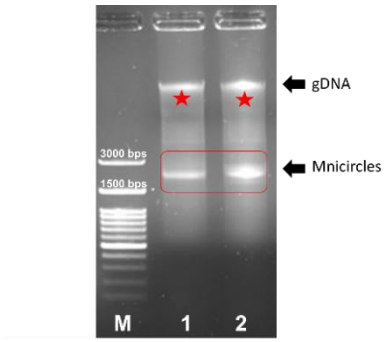


**Fig. 3-1: The isolated minicircles of *A. carterae* CCAM0512.**

**M:** 100 bps DNA ladder H3 RTU marker. **Lane 1 and 2:** Isolated minicircles. Lane 1 and 2 contain a signal in the range of about 2000 bps. These are the isolated minicircles. The gDNA of *A. carterae* CCAM0512 was marked by red stars. The gel of red box area was retrieved and used for the isolation of individual minicircles via a transposon-insertion based approach.

### 3.1.2 Analysis of individual minicircles

By using a transposon-insertion based approach and several times electroporations (see materials and methods 5.2.4.12) it was found that twenty two colonies grew on lysogeny broth (LB) agar plate 1 containing kanamycin antibiotic (Table 3-1) and hundreds of thousands of colonies grew on plate 2 and 3. Subsequently, about 150 colonies were picked out and cultured in liquid LB medium containing kanamycin antibiotic for plasmid preparation (see materials and methods 5.2.4.1). Based on the digestions by enzymes and analysis on the agarose gel 107 potential positive minicircles were sequenced via transposon sequencing forward and reverse primers (see supplements 7.3). The 107 potential positive minicircles are named Juan 1 to Juan 107, shortly J1 - J107. Finally, it was found that 18 out of 89 plasmids (20.2%) are gene-containing minicircles. 71 out of 89 plasmids (79.8%) are empty minicircles. The number of empty minicircles is about four times the number of gene-containing minicircles. While the remainder 18 plasmids are false positive colonies. The more detail will be shown on the next texts.

**Table 3-1: Statistic analysis of isolated individual minicircles.**

(Plate 1, 2 and 3 are three different transposon insertions and electroporations.)

| Plate | Colonies on plate | Sequencing number | name | False positive minicircles | Gene containing minicircles | Empty minicircles |
|---|---|---|---|---|---|---|
| 1 | 22 | 10 | J1-J6, J21-J22, J70, J71 | 2 | 2 | 6 |
| 2 | Hundreds of | 27 | J7-J17, J23-J33, J34-J37, J72 | 4 | 8 | 15 |
| 3 | Hundreds of | 70 | J18-J20, J38-J69, J72-J107 | 12 | 8 | 50 |
| Total number | | 107 | | 18 | 18(20.2%) | 71(79.8%) |

### 3.1.2.1 The overview of all individual minicircles

Within this 18 gene-containing minicircles, it was found that 12 out of 18 minicircles belong to different minicircle molecules. The 12 minicircles include three different *psbA*-containing minicircle molecules, two different *petB*/*atpA*-containing minicircle molecules, two different *23S rRNA*-containing minicircle molecules and one *psaB*-, *psbC*-, *psbD/E/I*-, *petD*- and *atpB*-containing minicircle molecule (Table. 3-2). 11 different genes were identified on these individual minicircles of 2333 bps (J2) – 2664 bps (J33) length from *A. carterae* CCAM0512 (Table. 3-2). The average length of all these gene-containing minicircles is 2474 bps. All genes identified represent the core components of the chloroplast genome of all other photosynthetic organisms, as the encoded subunits of the complexes (photosystem I and photosystem II, the ATP synthase and the cytb6/f complex) are involved in the light reactions of photosynthesis as well as 23S rRNA (Table. 3-2). As reported from minicircles of other species (Hiller 2001, Nisbet, Koumandou et al. 2004), two minicircles containing more than one gene were isolated, namely J29 (including *psbD*, *psbE* and *psbI* genes) and J37 (including *petB* and *atpA* genes). It was also found that these genes on the same minicircle are separated by only about 100 bps to 600 bps nucleotide sequences. However, in conventional plastid genome these genes are located much far from each other (Ohyama, Fukuzawa et al. 1986, Shinozaki, Ohme et al. 1986). In contrast to minicircles from other species (Barbrook and Howe 2000, Hiller 2001, Zhang, Cavalier-Smith et al. 2002), all genes encoded on minicircles in *A. carterae* CCAM0512 have the same start codon (ATG), while TAG, TGA or TAA are used as stop codon in minicircle genes (Table 3-2).

**Table 3-2: Properties of minicircles containing chloroplast genes in *A.carterae* CCAM0512.**

(Minicircles' number: the number of isolated minicircles; Individual minicircles' number: the number of different minicircle molecules; PS I: Photosystem I; PS II: Photosystem II; bps: base pairs)

| Circle' name | Minicircles' number | Individual minicircles' number | Coding gene | Minicircle length (bps) | Gene length (bps) | Start codon | Stop codon | Product |
|---|---|---|---|---|---|---|---|---|
| J12 | 1 | 1 | *psaB* | 2469 | 1875 | ATG | TGA | CP47 PS I |
| J2 | 5 | 3 | *psbA* | 2333 | 1023 | ATG | TAG | D1 PS II |
| J11 | 5 | 1 | *psbC* | 2343 | 1251 | ATG | TAA | CP43 PS II |
| J29 | 1 | 1 | *psbD/E/I* | 2354 | 1068/243 /108 | ATG | TAA | D2/Cytochrome b599 alpha/ PS II |
| J37 | 2 | 2 | *petB/atpA* | 2596 | 660/1041 | ATG/ ATG | TAA/ TGA | Cytochrome b6/ alpha-subunit ATP synthase |
| J30 | 1 | 1 | *petD* | 2481 | 567 | ATG | TAA | Subunit IV ATP synthase |
| J28 | 1 | 1 | *atpB* | 2555 | 1662 | ATG | TAG | Beta-subunit ATP synthase |
| J33 | 2 | 2 | *23S rRNA* | 2664 | _ | _ | _ | 23S rRNA |
| Total number | 18 | 12 | | | | | | |

In addition to minicircles with coding genes, 71 empty minicircles have been isolated (Table 3-3). Of the total 71 empty minicircles, 26 (36.6%) belong to different minicircle molecules. The 26 different molecules were divided into six groups (Table 3-3). For example, the group of J8 empty minicircle contains three different minicircle molecules in 17 minicircles. These empty minicircles include different numbers of open reading frames (ORFs) for which no homology could be detected. Minicircle J9 has nine ORFs>150 bps compared with only two ORFs>150 bps on minicircle J36. Their lengths are different ranging from minicircle J22 (1449 bps) to minicircle J24 (2493 bps) and sequences are not conserved. The average length of these empty minicircles is 1986 bps. These ORFs without known function may be unique to dinoflagellates. It might also be possible that they have specific functions for minicircles.

**Table 3-3: Properties of empty minicircles in *A.carterae* CCAM0512.**

(Minicircles' number: the number of the isolated minicircles; Individual minicircles' number: the number of different minicircle molecules; ORF: open reading frame; bps: base pairs)

| Name/Type | Minicircles' number | Individual minicircles' number | Minicircle length (bps) | ORF>150 bps | Largest ORF(bps) |
|---|---|---|---|---|---|
| J8 | 17 | 3 | 1853 | 4 | 249 |
| J9 | 21 | 7 | 1892 | 9 | 243 |
| J13 | 9 | 6 | 2209 | 3 | 492 |
| J22 | 1 | 1 | 1449 | 3 | 195 |
| J24 | 19 | 6 | 2493 | 6 | 264 |
| J36 | 4 | 3 | 2020 | 2 | 198 |
| Total number | 71 | 26 | | 27 | |

As shown in Table 3-3, these empty minicircles contain several open reading frames with larger than 150 bps without known homology. The lengths of the largest open reading frames in these empty minicircles are different from 195 bps to 492 bps. In order to detect the relationship between the ORFs > 150 bps of gene-containing minicircles and empty minicircles the comparison was conducted. It was found that empty minicircle J9 and *atpB*-containing minicircle have a completely identical open reading frame and a partially identical open reading frame (Table 3-4). At the same time, it was shown that empty minicircles J13, J22 and J36 have an identical open reading frame, empty minicircles J22 and J24 have an identical open reading frame.

**Table 3-4: Identical open reading frames in *A. carterae* CCAM0512.**

(J28 is *atpB* containing minicircle, the rest of minicircles including J9, J13, J22, J24 and J36 are empty minicircles. Identical open reading frames: the minicircles contain the identical open reading frames or partially identical open reading frame. The potential products of the open reading frames are without any known homology. Only the open reading frames > 150 bps were analyzed here. )

| Minicircles | Identical open reading frames |
|---|---|
| J9,J28 | MIYFYLVQCNIRDFETQRGFPPPPCRSTLEDPRVPSSNSSQYARTPEKIHR. |
| J9,J28 | ……..LSLIHISTIIDEFSRVFSHIGSNSSSVPGDPLESTCRGGGESHVVSQNL. |
| J13,J22,J36 | MSPQRSIAPFQVICLSAPLPSIEGLSLSFHQLSLHSFVYSLVEILTSRHI. |
| J22,J24 | MSHDIITTTPNPLSFIGGGLIKVKSLWPMRGAIHQLQYLVHSPSYRRSSEH. |

For the purpose of distinguishing the difference on the nucleotide sequences of gene-containing minicircles and empty minicircles the GC contents of these minicircles were compared (Fig. 3-2). It was shown that the GC contents of the overall minicircles generally appear to be lower than the GC contents of the coding regions, but higher than the GC content of the non-coding regions except the same GC content in *petB*-containing minicircle. A special case is that in *psbE* and *psbI* genes-containing minicircle the overall minicircle presents to be more GC-rich than the coding regions and the non-coding region, the non-coding region shows higher GC content than the coding regions. Except the GC content of *psbA*-containing minicircle is 45%, which is the same as or lower than the GC content of the overall empty minicircles, all the other gene-containing minicircles present to be more GC-rich than the empty minicircles. It was also presented that the GC contents of the overall minicircles, the coding regions and the non-coding regions are lower than the AT content.

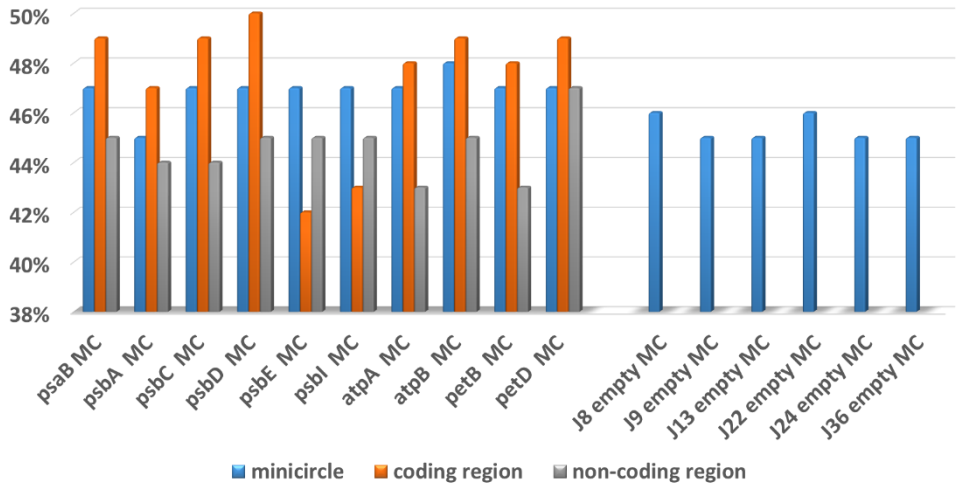


**Fig. 3-2: GC content for minicircles of *A. carterae* CCAM0512.**

The figure shows that the GC content of the overall minicircles, the coding regions and the non-coding regions in genes-containing minicircles and empty minicircles. It was shown that the overall GC contents of the minicircles are generally lower than the GC content of the coding regions, but higher than the GC content of the non-coding regions except the same GC content in *petD* gene minicircle. However, the overall *psbE* and *psbI*-containing minicircles appear to be more GC-rich than the coding regions and the non-coding regions. Except the *psbA*-containing minicircle all the other gene-containing minicircles present to be more GC-rich than the empty minicircles. The more detailed description see text. GC content is calculated on the website http://emboss.bioinformatics.nl/.

### 3.1.2.2 The core regions of minicircles in *A. carterae* CCAM0512

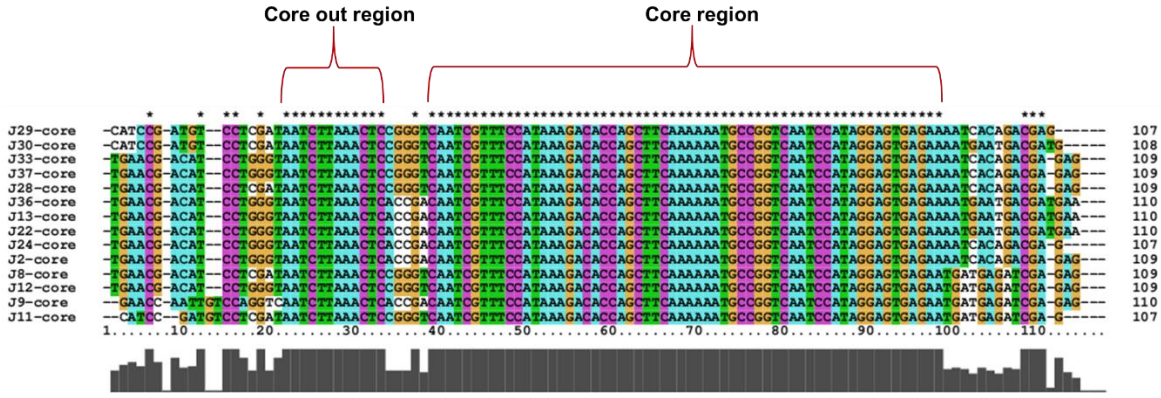


**Fig. 3-3: Alignment of the core regions of fourteen minicircles of *A. carterae* CCAM0512.**

It was shown that all fourteen minicircles have a highly conserved sequence about 60 bps called core region, and a 12 bps conserved sequence called core out region. Only the conserved regions were shown in this figure. Consensus regions of all fourteen minicircles of *A. carterae* CCAM0512 were aligned by using ClustalX 2.1.

In order to compare the non-coding regions of all minicircles the sequences of consensus minicircles were aligned. It was shown that there is a highly conserved core sequence of 60 bps, called core region. At the same time, there is a 12 bps conserved sequence called core out region (Fig. 3-3).

Based on the comparison of the sequences between the core region and the core out region it was found that the nucluotides of minicircle J8, J11, J12, J28, J29, J30, J33 and J37 are the same, but different from that of the remainder six minicircles. To predict the function of the core region different algorithms were used to predict a putative promoter sequence, whereby the sequence "CACCAGCTTCAAAAAATGCCGGTCAATCCATAGGAGTGAGAAAATCACAG/TGATGAGA" was predicted to have a putative function for initial transcription, which has to be confirmed in the future.

### 3.1.2.3 Transcription and RNA editing analyses of individual minicircles

RNA-editing can be commonly observed in land plants, although the frequency varies in different lineages. In the algae, the RNA editing remains undetectable until now, aside from the editing that occurred in the dinoflagellates *Ceratium horridum* and *Heterocapsa triquetra* (Zauner, Greilinger et al. 2004, Dang and Green 2009). In order to study the RNA editing on minicircles of *A. carterae* CCAM0512 mRNA derived sequences were compared with genomic sequences of the *psbA* gene. It was shown that no RNA editing was observed in the *psbA* gene. The observed RNA products support the previous study that the transcription of coding genes on minicircles (Barbrook and Howe 2000, Barbrook, Symington et al. 2001, Zhang, Cavalier-Smith et al. 2002, Nisbet, Hiller et al. 2008).

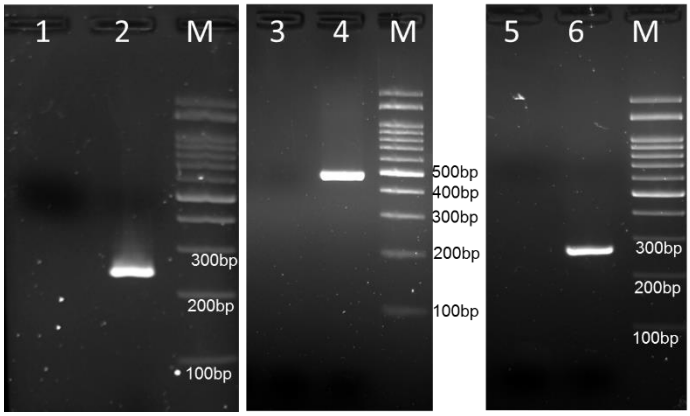


**Fig. 3-4. Transcription of ORFs of three empty minicircles.**

It was indicated that the three ORFs were transcribed in *A. carterae* CCAM0512. The figure showed the results of RT-PCR by using the primers designed to amplify the regions of three largest ORFs of different empty minicircles (J9, J13 and J24). Lane 1, lane 3 and lane 5: control PCR without reverse transcriptase added. Lane 2, lane 4 and lane 6: The transcriptional products of ORFs on empty minicicle J9, J13 and J24. The lengths are 243 bps, 492 bps and 264 bps, respectively. M: DNA markers.

It was already shown that transcription occurs over a large part of the empty minicircle including the core region (Nisbet, Hiller et al. 2008). Here, the reverse transcriptase-PCR (RT-PCR) was carried out in order to analyse the transcription on the largest ORFs of empty minicircles (J9, J13 and J24, see

table 3-2). Products were observed for all three ORFs (Fig. 3-4) and comfirmed by sequencing, which indicated that these ORFs of empty minicircles might be transcribed.

### 3.1.3 Evolution analyses of four *A. carterae* strains' minicircles

Until now, minicircles have been identified in four different *Amphidinium carterae* strains, including *A. carterae* CCAM0512 presented here and *A. carterae* CS-21, *A. carterae* CCAP1102/6 and *A. carterae* CCMP1314. The protein-coding genes identified on minicircles of the different *A. carterae* strains are listed in table 3-5 (Barbrook and Howe 2000, Hiller 2001, Zhang, Cavalier-Smith et al. 2002, Barbrook, Santucci et al. 2006). All these genes are important for the function of plastid. However, the reported genes on minicircles remain very limited. To compare the minicircles' relationship of these four *A. carterae* strains phylogenetic analysis were performed for the coding sequences and non-coding sequences in this chapter.

**Table 3-5: Reported minicircular sequences in four *A.carterae* strains.**

(LSU: large subunit; SSU: small subunit; rRNA: ribosome RNA)

| strains | *atpA* | *atpB* | *petB* | *petD* | *psaA* | *psaB* | *psbA* | *psbB* | *psbC* | *psbD* | *psbE* | *psbI* | *LSU-rRNA* | *SSU-rRNA* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *A. carterae* CCAM0512 | ✓[a] | ✓ | ✓[a] | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓[b] | ✓[b] | ✓[b] | ✓ | |
| *A. carterae* CCAP1102/6 | ✓[a] | ✓ | ✓[a] | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓[b] | ✓[b] | ✓[b] | ✓ | ✓ |
| *A. carterae* CS21 | ✓[a] | ✓ | ✓[a] | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓[b] | ✓[b] | ✓[b] | ✓ | ✓ |
| *A. carterae* CCMP1314 | | | | | | | ✓ | | | | | | ✓ | |

Genes located on the same minicircles are indicated by superscript lettering a and b. Minicircles in *A. carterae* CCAP 1102/6, *A. carterae* CS21 and *A. carterae* CCMP1314 see references (Barbrook and Howe 2000, Barbrook, Symington et al. 2001, Hiller 2001, Zhang, Cavalier-Smith et al. 2002, Nisbet, Koumandou et al. 2004, Barbrook, Santucci et al. 2006). PsbB mincircle was isolated by Groche C. (Grosche 2012).

#### 3.1.3.1 Overall genome characteristics and open reading frames

Based on the comparison of the GC content of the *psbA* gene-containing minicircles, it was shown that the GC contents of the coding regions are the same in these four *A. carterae* strains and higher than the GC content of the overall minicircles, the non-coding regions and core regions (Fig. 3-5a). It generally appears to be more GC-rich in *A. carterae* CCMP1314 than the other *A. carterae* strains. The GC contents of the overall minicircles in *A. carterae* CCAP1102/6 and *A. carterae* CCMP1314 are higher than that of *A. carterae* CCAM0512 and *A. carterae* CS21. The GC content of A. carterae CS21

non-coding regions is 43%, which is lower than that in the other three species. In the core regions, the GC content increases progressively from *A. carterae* CCAM0512, *A. carterae* CCAP1102/6, *A. carterae* CS21 to *A. carterae* CCMP1314 (Fig. 3-5a). For the 23S rRNA containing minicircles the GC contents of the overall minicircles are the same, as shown in Fig. 3-5b.

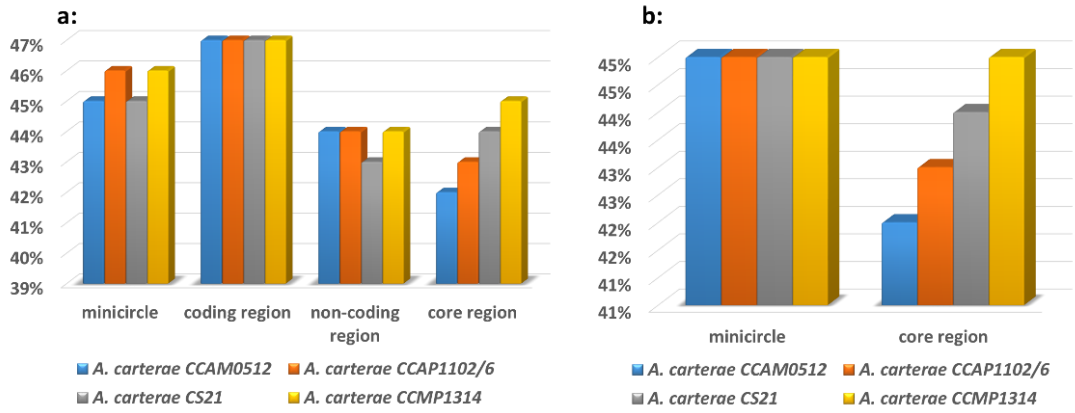


**Fig. 3-5: GC content of *psbA* and 23S rRNA-containing minicircles in four *A. carterae* strains.**

The figure shows that the GC contents of the overall minicircles, the coding regions, the non-coding regions and the core regions in four *A. carterae* strains. **a:** GC content for *psbA*-containing minicircles. It was shown that the GC contents of the coding regions are nearly the same and are higher than the GC contents of the overall minicircles, the non-coding regions and the core regions. The more detailed description see text. **b:** GC content for 23S rRNA containing minicircles. It was shown that the GC contents of the overall minicircles are the same on these four *A. carterae* strains. GC content is calculated on the website http://emboss.bioinformatics.nl/.

Like shown in table 3-4, several identical open reading frames (the lengths > 150 bps) of the gene-containing minicircles and empty minicircles were found in *A. carterae* CCAM0512, they do not have known homology. The open reading frames (the lengths > 150 bps) were also compared in four *A. carterae* strains. It was shown that 7 out of 27 pairs of open reading frames have a high query cover, low e-value and high identity (Table 3-6). Among these three pairs of open reading frames have a 100% query cover, very low e-value and considerable high identity (83%, 98% and 96%) (marked by red).

**Table 3-6: The open reading frames of empty minicircles in *A. carterae* strains.**

(The name is consist of the empty minicircle name and open reading frame number in *A. carterae* CCAM0512---accession number and open reading frame number in *A. carterae* CCAP1102/6 and CS21 strains. Accession numbers of sequences in *A. carterae* CCAP1102/6: [AJ582641] [AF401630]. Accession numbers of sequences in *A. carterae* CS21: [AJ307015] [DQ507216] [AJ318067]. The sequences of the open reading frames see supplements 7.1. The comparison was performed on the website http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=tblastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome )

| Name | Max score | total score | query cover | e-value | identity |
|---|---|---|---|---|---|
| J8 ORF2---AJ307015 ORF1 | 67.4 | 90.9 | 70% | 5e-16 | 100% |
| J8 ORF3---AJ582641 ORF1 | 38.5 | 38.5 | 29% | 3e-06 | 100% |
| J13 ORF2---DQ507216 ORF1 | 251 | 251 | 100% | 2e-80 | 83% |
| J22 ORF1---AF401630 ORF1 | 103 | 103 | 100% | 5e-29 | 98% |
| J24 ORF2---AF401630 ORF2 | 65.5 | 65.5 | 61% | 2e-15 | 91% |
| J24 ORF2---AJ318067 ORF1 | 56.6 | 56.6 | 56% | 2e-12 | 88% |
| J24 ORF6---AJ318067 ORF2 | 112 | 112 | 100% | 7e-32 | 96% |

### 3.1.3.2 Phylogenetic analysis of *psbA* genes for 15 minicircles of dinoflagellates

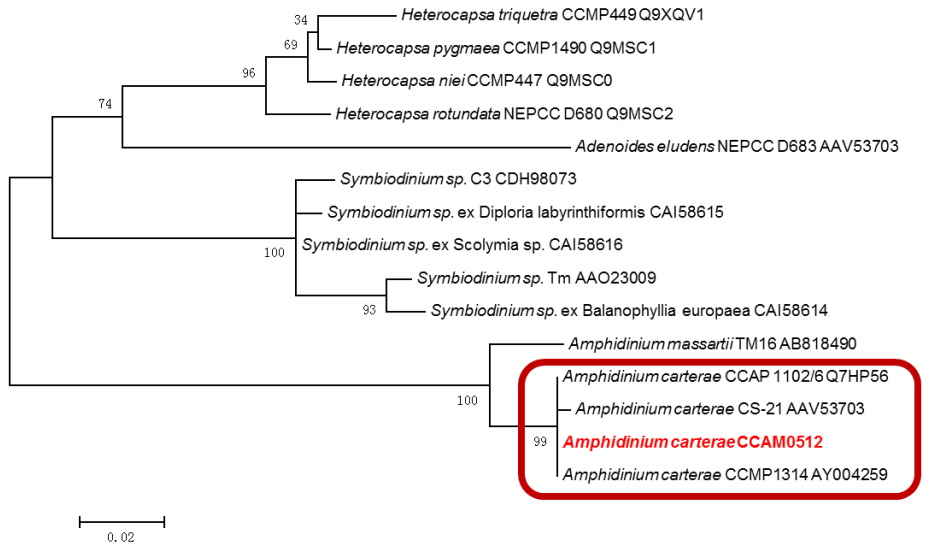


**Fig. 3-6. Phylogenetic tree analysis based on amino acid sequences encoded by *psbA* gene from 15 minicircles of dinoflagellates.**

It was shown that the four *A. carterae* strains form a tightly clade with a strong statistical support (bootstrap support values: 99%, marked by red box). It was indicated that the coding regions of *psbA* gene for the four *A. carterae* strains are almost identical. All the other species was found to form an independent clade from the clade consisting of *A. carterae* and *A. massartii*. The phylogenetic tree was built by maximum likelihood analysis using Mega 6.0. The tree with the highest log

likelihood (-1923.2228) is shown. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 15 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 336 positions in the final dataset. The bootstrap consensus tree inferred from 1000 replicates. Scale bar represents 0.02 expected substitutions per site in the aligned regions.

Phylogenetic analysis of PsbA encoded on minicircles from different peridinin-containing dinoflagellate species showed that *A. carterae* CCAM0512 forms an independent and tight clade with the other three different *A. carterae* strains, thus indicating that the four strains are closely related to each other with strong statistical support (bootstrap support values: 99%) (Fig. 3-6). The psbA coding sequences of these four *A. carterae* strains are almost identical. At the same time, these results could be confirmed by the phylogenetic analysis for amino acid sequences of other proteins encoded on minicircles (data not shown). As shown in Fig. 3-6, *Heterocapsa* and *Symbiodinium* species form an independent clade, respectively separated from the clade containing *A. carterae*.

### 3.1.3.3 Alignment analysis of core regions from four *A. carterae* strains

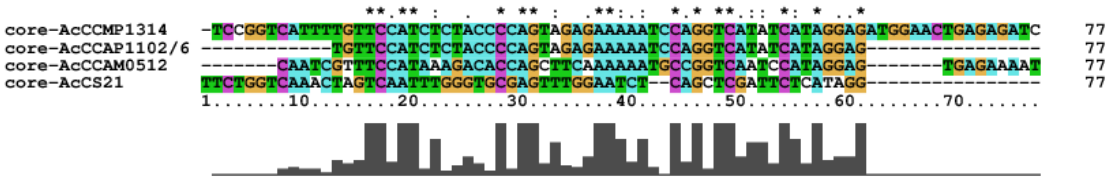


**Fig. 3-7. Alignment of the core region of different *A. carterae* strains' minicircles.**

It was shown that the core regions are highly diverse and unrelated between these four strains. The core regions of four *A. carterae* strains (*A. carterae* CCMP1314, *A. carterae* CCAP1102/6, *A. carterae* CCAM0512 and *A. carterae* CS21) were compared by using ClustalX 2.1.

It was already shown that the psbA coding sequences encoded on minicircles of four *A. carterae* strains are highly identical and these four strains form a tightly sister-relationship in a clade. Therefore, it is essential to compare the core regions of these four strains. Surprisingly, the core regions of strain CCAM0512 and strain CS-21 are highly diverse in comparison to each other and to the core regions of the strains CCMP1314 and CCAP1102/6 (Fig. 3-7).

The coding sequences and core regions of the four *A. carterae* strains have already been compared. Finally the remainder non-coding regions of psbA minicircles were also compared for the four *A. carterae* strains (Fig. 3-8). For *psbA* there is a 13 bps conserved sequence downstream the stop codon, upstream the start codon there are four longer conserved regions. However, the sequences are highly variable in the rest of the non-coding regions. It was shown that the core regions are located in a more variable region. To predict the function of the conserved sequence upstream the

start codon of *psbA* different algorithms were used to predict a putative promoter sequence, whereby a conserved sequence of 50 bps might have a function for the initial transcription, which has to be confirmed in the future.

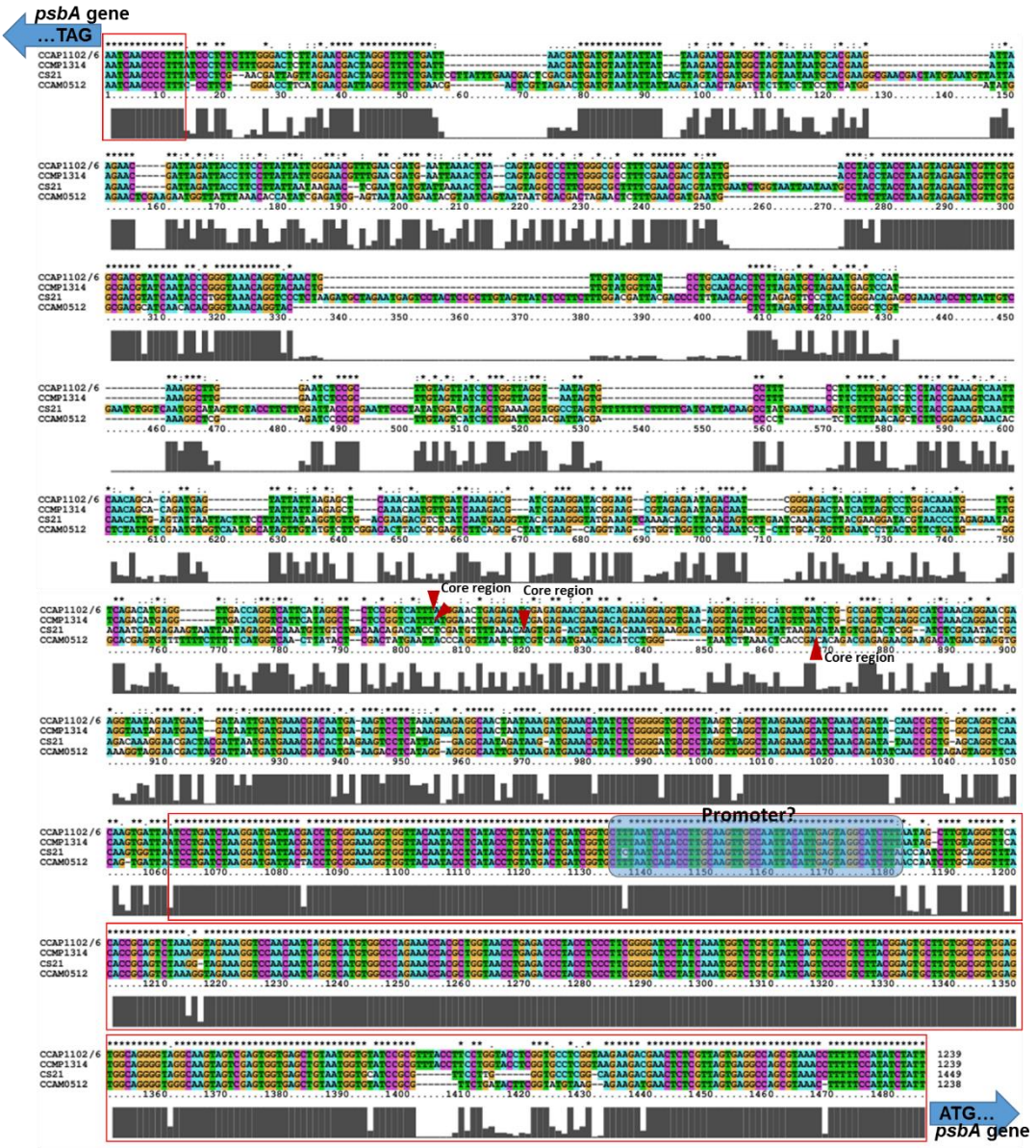


**Fig. 3-8. Alignment of the non-coding regions of four different *A. carterae* strains' minicircles.**

"TAG" is the stop codon of the *psbA* gene, while "ATG" is the start codon. The core regions of four *A. carterae* strains' minicircles were truncated and marked by red triangles. It was shown that the core regions are located in a more variable region. Upstream the start codon and downstream the stop codon of *psbA* gene (marked by the red boxes) of the non-coding sequences are conserved. The putative promoter region was marked by blue box. The non-coding regions without core regions of four *A. carterae* strains' *psbA* minicircles were compared by using ClustalX 2.1.

According to these analyses, the sequences of the *A.carterae* strain CCAP1102/6 and the *A.carterae* strain CCMP1314 are almost 99% identical therefore it was speculated that they are probably the same strains.

#### 3.1.3.4 Phylogenetic analysis of partial LSU/SSU rDNA

With the purpose of comparing the evolutionary relationship of these four *A. carterae* strains further phylogenetic analyses based upon two nuclear genes (partial LSU rDNA and SSU rDNA) were carried out. The *A. carterae* strains formed clades that were unambiguously separated from some other *Amphidinium* species (Fig. 3-9 and 3-10). The tree obtained for LSU rDNA sequences showed that all *A. carterae* species formed a tightly sister relationship within the same clade (Fig. 3-9). This clade was divided into three subclades, containing the strain *A. carterae* CCMP1748 and another two larger clades (clade 1 and clade 2). Additionally, it was shown that the strains of *A. carterae* CS-21 and *A. carterae* CCAM0512 are in the same clade, forming a tight neighbor relationship with strong statistical support (bootstrap support values: 100%) (Fig. 3-9). The phylogenetic position of *A. carterae* species based upon its SSU rDNA sequences was shown in Fig. 3-10. All the *A. carterae* species were grouped in the same clade. The species *A. carterae* CCMP1314, *A. carterae* CCAP1102 and *A. carterae* CCAM0512 were found to form a sister relationship, which was confirmed by the high bootstrap support values.

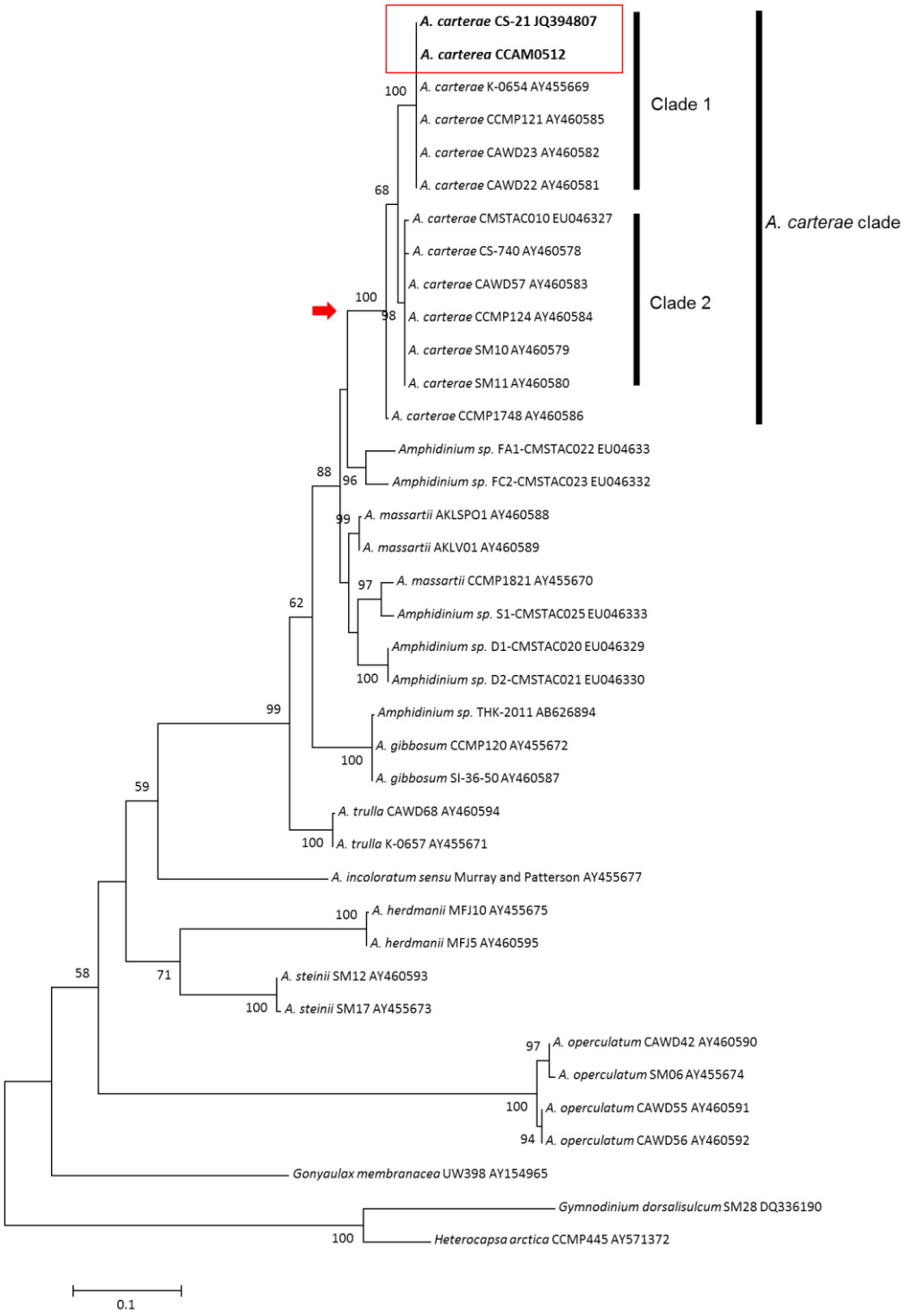


**Fig. 3-9. Phylogenetic tree analysis based on LSU rDNA sequences covering domains D1-D6.**

It was shown that all the *A. carterae* species form an independent clade (marked by red arrow at the node). This *A. carterae* clade was divided into three subclades, clade 1, clade2 and *A. carterae* CCMP1748. *A. carterae* CS-21 and *A. carterae* CCAM0512 were contained in the clade1 and form the closest relationship (marked by the red box). The corresponding GenBank accession numbers of sequences are after the each taxa. The phylogenetic tree was made using Maximum Likelihood method based on the large subunit rDNA sequences covering domain 1 to domain 6. The tree with the highest log likelihood (-5458.4502) is shown. Bootstrap values (>50%) are given at each node. Evolutionary analyses were conducted in MEGA6 (Tamura, Stecher et al. 2013).

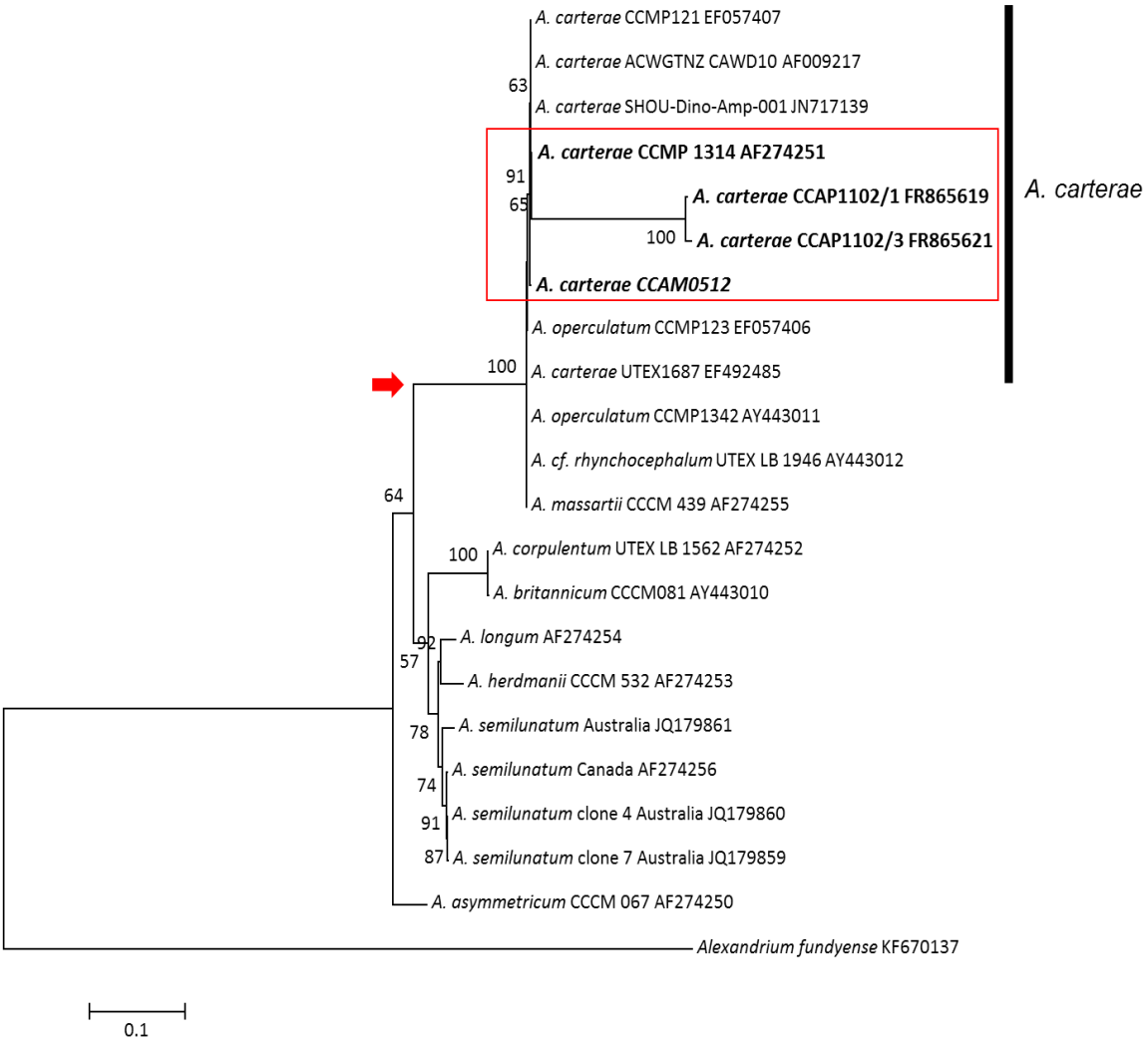


**Fig. 3-10. Phylogenetic tree analysis based on SSU rDNA sequences.**

All the *A. carterae* species form an independent clade together with *A. operculatum* CCMP1342, *A. cf. rhynchocephalum* UTEX and *A. massartii* CCCM 439 (marked by red arrow at the node). *A. carterae* CCAM0512, A. carterae CCMP1314 as well as A. carterae CCAP1102 form a nearest phylogenetic neighbours (marked by the red box). The corresponding GenBank accession numbers of sequences are after the each taxa. The phylogenetic tree was carried out as described in Fig. 3-9.

## 3.2 The endomembrane system (ES) in *Phaeodactylum tricornutum*

The localization of proteins on a subcellular level is a very important technique for cell biology research. In recent years, the most common used method for protein localization is to fuse the proteins of interest to a reporter, whereby examples include the enhanced green fluorescent protein (eGFP) and monomeric red fluorescent protein (mRFP) and express these fusion proteins in the cells. The already known marker proteins in diatoms are not enough for studying the mechanisms of vacuolar protein sorting and transport, therefore establishing further subcellular-specific marker proteins in endomembrane system is the aim of this work.

In this project, to study subcellular-specific localization of interest proteins eGFP- and mRFP-fusion proteins were generated and the confocal laser scanning microscope was used to identify the subcellular localization of proteins *in vivo* in *P. tricornutum*.

### 3.2.1 Identification of marker proteins

Marker proteins with known localization are not enough for the investigation of the mechanisms of vacuolar protein sorting and transport in *P. tricornutum*. In order to identify more marker proteins proteins with a known localization in plants were collected from the literature. All homologous genes were extracted from the *P. tricornutum* genome database.

In order to correct the gene models retrieved from the database, the predicted gene model sequences have been compared with expressed sequence tags (EST). Simultaneously, the start codons upstream the predicted proteins were screened on the genome browser. If it was necessary, the gene models were corrected. If there was no EST available, the corresponding cDNAs were amplified and sequenced. In addition, for each protein the presence of targeting signals and transmembrane domains (TMDs) were predicted with different servers (see material and methods chapter). Subsequently, putative marker proteins were selected in order to analyze their *in vivo*-localization. An overview of all proteins can be found in the following chapters.

### 3.2.2 Tonoplast intrinsic proteins (Tips)

The first identified proteins are tonoplast intrinsic proteins (Tips). Tips are one of the aquaporins family groups. Aquaporins are channel proteins belong to the major intrinsic proteins (MIPs) superfamily. The tonoplast intrinsic proteins form transmembrane channels that selectively mediate the transport of water, gases and small neutral solutes such as glycerol and urea (Gattolin, Sorieul et al. 2009). This protein family was found in a variety of plants, animals and bacteria and have a high

sequence similarity to each other (Johnson and Chrispeels 1992, Jauh, Phillips et al. 1999, Liu, Ludewig et al. 2003, Moriyasu, Hattori et al. 2003, Zardoya 2005). Tips were widely used as markers for lytic vacuole and protein storage vacuole in higher plants (Frigerio, Hinz et al. 2008, Gattolin, Sorieul et al. 2009). The localization and function of several Tip isoforms (including three gamma-Tip (Tip1), three delta-Tip (Tip2), the seed specific alpha- and beta-Tip (Tip3:1 and Tip3:2), one epsilon-Tip (Tip4:1) and one zeta-Tip (Tip5:1)) have already been well-studied or predicted in plant species, but it remains unknown about the localization of Tip proteins in diatoms.

Finally, five Tip proteins (Tip1-5) homologous to Tip proteins of *A. thaliana* were identified and selected for *in vivo*-localization studies in *P. tricornutum* (Table 3-7). All these Tip proteins have six putative transmembrane domains (called TMD1-TMD6). Based on the observation on the amino acid sequence of the five Tip proteins it was shown that each Tip protein has one (in case of Tip3 and Tip4) or two (in case of Tip1, Tip2 and Tip5) conserved Asn-Pro-Ala (NPA) motifs between two TMDs to form extra-cytosolic loop helices. The prediction of targeting signals showed that only Tip4 has a signal peptide. All the other Tips also comprise the specific domain structure, but they are lack of a predictable signal peptide.

**Table 3-7: Predicted subcellular localized Tonoplast intrinsic proteins in *P. tricornutum*.**

(Pt: *P. tricornutum*, P: plastid (envelope), PPM: periplastidal membrane, EM: endosomal membrane, SP: signal peptide; TMD: transmembrane domain; loc: localization)

| protein family | name | BLAST hits/specific function | Pt ID | SP | TMD | Predicted-loc | eGFP-loc |
|---|---|---|---|---|---|---|---|
| Tonoplast intrinsic protein (TIP) | Tip1 | aquaporin | 31553 | - | 6 | EM | Plasma membrane |
| | Tip2 | glycerol uptake facilitator | - | - | 6 | EM | Vacuole |
| | Tip3 | glycerol uptake facilitator | 20755 | - | 6 | EM | cER/nuclear envelope |
| | Tip4 | glycerol uptake facilitator | 19409 | + | 6 | P/PPM | PPM |
| | Tip5 | aquaporin | 43157 | - | 6 | PPM | cER/nuclear envelope |

### 3.2.2.1 *In vivo*-localization of Tip1

In order to study the subcellular localization of Tip1, eGFP was attached to the C-terminus of the protein and were expressed in *P. tricornutum*. The fluorescence pattern of the Tip1-eGFP fusion protein surrounded the cell. Therefore it was speculated that Tip1 localized to the cytoplasmic membrane (Fig. 3-11a). However, besides the plasma membrane, one or several dots inside the cells could be observed in some clones (Fig. 3-11b). This intracellular fluorescence might stem from endocytosed compartments. To test this, co-staining with FM4-64 was carried out. This dye is used

visualize the plasma membrane and endomembrane system-dependent internalization processes. The co-localization of Tip1-eGFP with FM4-64 indicates that the internal Tip1-eGFP fluorescence belongs to the endosomal system (Fig. 3-11c). Based upon the prediction, Tip1 protein is a membrane protein. In order to confirm this carbonate extraction was performed. As shown in Fig. 3-11d, the Tip1-eGFP fusion protein was inserted into the membrane, although a weak signal could be detected in the fraction of soluble proteins.

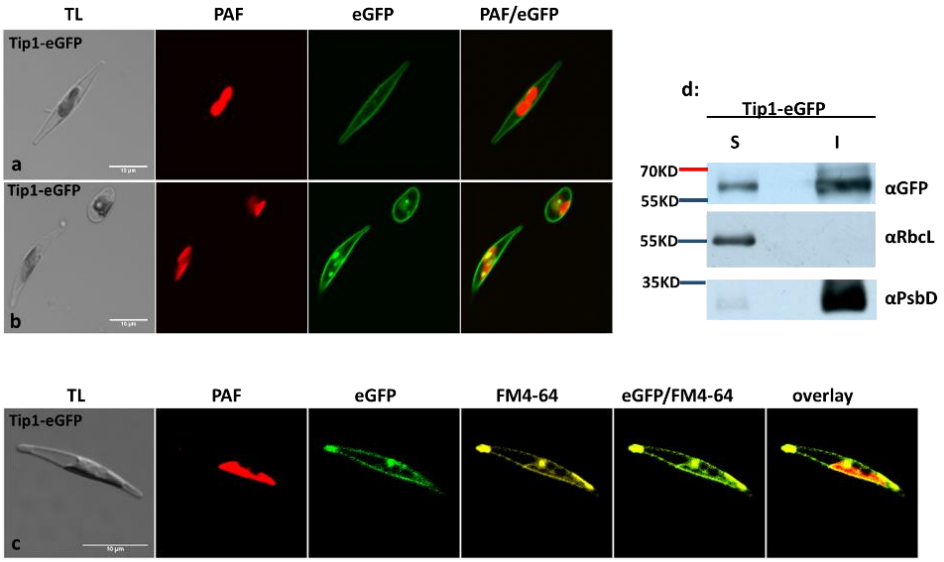


**Fig. 3-11: *In vivo* localization of Tip1-eGFP.**

**a:** The Tip1-eGFP fusion protein showed a plasma membrane fluorescence; **b:** An additional dot-like fluorescence could be observed in some clones. **c:** The intracellular dots of Tip1-eGFP overlapped with FM4-64. **d:** Carbonate extraction of Tip1-eGFP. Although a weak signal could be observed in the fraction of soluble, it was shown that Tip1 is membrane protein as the signal was very strong in fraction of integral membrane. The thylakoid membrane protein PsbD (25 kDa) and the stromal protein RbcL (55 kDa) were used as makers for the fraction of soluble and integral membrane proteins. The expected molecular weight of Tip1-eGFP protein is 60 kDa. For a detailed description see text. TL (gray): transmitted light; PAF (red): plastid autofluorescence; eGFP (green): enhanced green fluorescent protein; FM4-64 (yellow): a dye used to visualize the plasma membrane and endomembrane system-dependent internalization processes, stained for 25min; overlay: the overlay of PAF, GFP and FM4-64.

### 3.2.2.2 *In vivo*-localization of Tip2

In plant cells the vacuole is the largest subcellular compartment. Until now no vacuolar proteins have been reported in diatoms. In order to visualize this single membrane surrounded structure, Tip2 was selected for *in vivo* localization studies in *P. tricornutum.*

After *in vivo* localization the Tip2-eGFP fusion protein was targeted to the vacuolar-like structure (Fig. 3-12a). By carbonate extraction, it was shown that Tip2-eGFP is an integral membrane protein (Fig.3-12b).

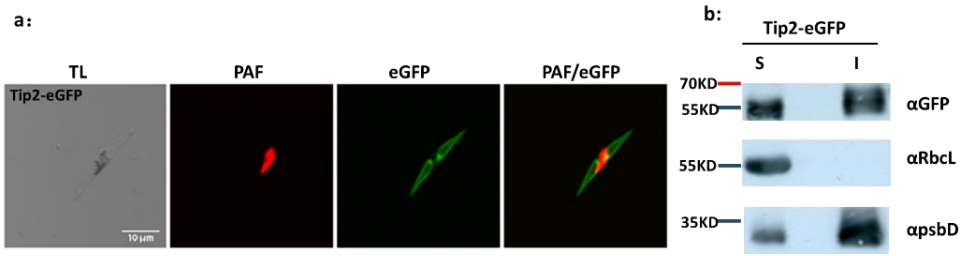


**Fig. 3-12: *In vivo* localization of Tip2-eGFP.**

**a:** Tip2-eGFP fusion protein was expressed. The fusion protein was targeted to a vacuolar-like structure. **b:** Carbonate extraction of the Tip2-eGFP fusion protein. It was shown that Tip2 is a membrane protein. The thylakoid membrane protein PsbD (25 kDa) and the stromal protein RbcL (55 kDa) were used as makers for the fraction of soluble and integral membrane proteins. The expected molecular weight of Tip2-eGFP is 55 kDa. For a detailed description see text. TL: transmitted light; PAF (red): plastid autofluorescence; eGFP (green): enhanced green fluorescent protein.

### 3.2.2.3 *In vivo*-localization of Tip3 and Tip5

Co-evolution of the host cell and the endosymbiont gave rise to the establishment of the secondary or complex platids. The secondary or complex plastids are surrounded by four membranes (the outermost membrane, cER; the second outermost membrane, PPM; the outer and inner envelope membranes). The membranes of the host ER, the nuclear envelope and the outermost plastidal membrane (cER membrane) are connected to each other. To better study the ER function, it is necessary to have additional marker proteins. Here, two homologous tonoplast intrinsic proteins Tip3 and Tip5 were selected for localization studies in *P. tricornutum*.

As shown in Fig.3-13a and b, the fluorescence of Tip3-eGFP and Tip5-eGFP could be detected in the cER-membrane and the nuclear envelope. However, in most cases fluorescence could also be observed in additional structures, which needs to be confirmed in the future. By the carbonate extraction the signal was mainly observed in the fraction of integral membrane (Fig. 3-13c and d). It indicated that Tip3 and Tip5 were targeted to the membrane, which is consistent with the prediction *in silico*.

To verify the localization of Tip3-eGFP immunogold labelling and electron microscopy was performed. As shown in Fig. 3-13e, the immunogold particles predominantly label the outermost membrane of the complex plastid (cER membrane) and the nuclear envelope. At the same time, several additional immunogold particles could be detected (marked by red arrow), which might hint that Tip3-eGFP was also located in the host ER.

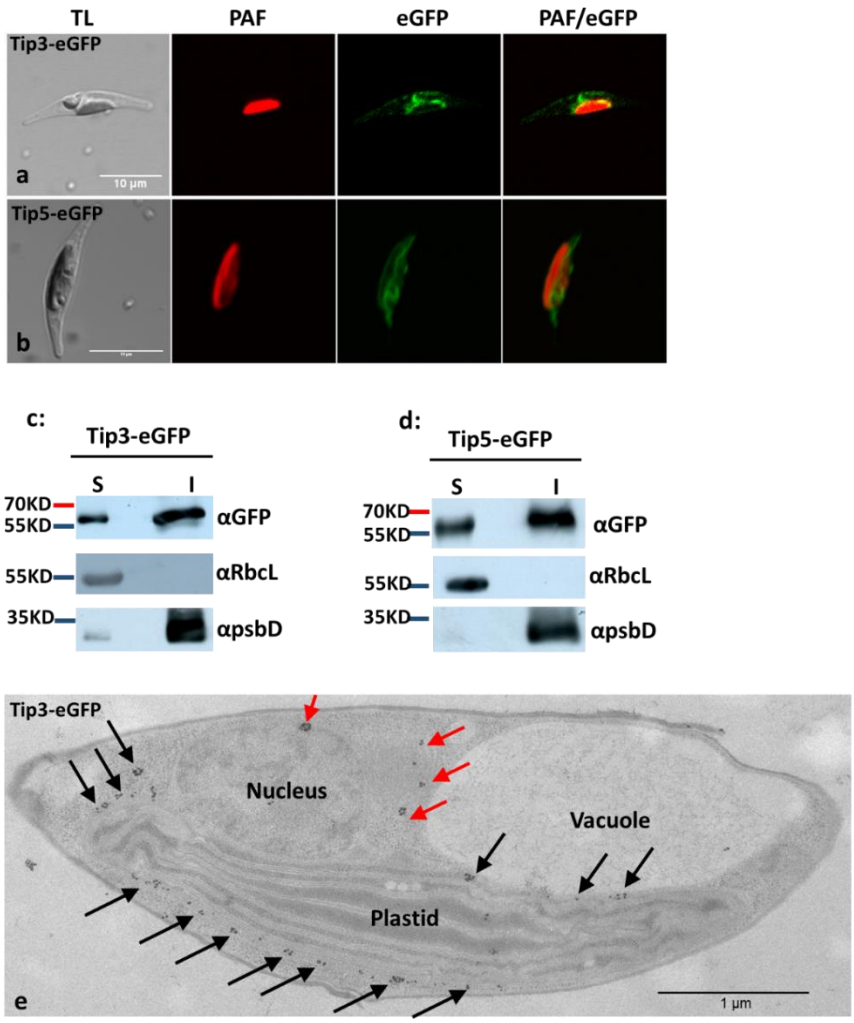


**Fig. 3-13: *In vivo* localization of Tip3- and Tip5-eGFP.**

**a-b:** Tip3- and Tip5-eGFP fusion proteins were expressed under the control of a nitrate inducible promoter. The fluorescence of Tip3-eGFP and Tip5-eGFP fusion proteins were targeted to the cER membrane and nuclear envelope. In most cases additional fluorescence extended from the cER membrane and nuclear envelope was also observed. **c and d:** Carbonate extraction of Tip3- and Tip5- eGFP. It indicated that Tip3 and Tip5 are integral membrane proteins. The thylakoid membrane protein PsbD (25 kDa) and the stromal protein RbcL (55 kDa) were used as makers for the fraction of soluble and integral membrane proteins. The expected molecular weight of Tip3/ Tip5-eGFP proteins are 56 kDa and 60 kDa, respectively. **e:** Immunoelectron microscopic analyses on ultra-thin cuts of *P. tricornutum*, indicating that Tip3 was significantly targeted into the outermost membrane of the complex plastid (cER membrane) (marked by black arrows). Immunogold labeling was partially accumulated in the nuclear envelope and might be on the host ER (marked by red arrows). For a detailed description see text. TL: transmitted light; PAF (red): plastid autofluorescence; eGFP (green): enhanced green fluorescent protein.

In order to verify the subcellular localization of the Tip5-eGFP, co-localization was performed between Tip5-eGFP and DAPI. Partial fluorescence of Tip5 surrounded the DAPI staining, which further confirmed the fluorescence of Tip5 was targeted to the nuclear envelope (Fig. 3-14a). Co-expression of Tip5-eGFP and Tip3-mRFP was carried out and verified the co-localization of these

proteins (Fig. 3-14b). However, a strong dot-like fluorescence could be detected on the mRFP channel in some cases (data not shown), which might result from the overexpression of the mRFP fusion protein. One of the important components of the host ERAD machinery hDer1 has already been identified to enrich in the cER membrane, nuclear envelope and the host ER (Hempel, Bullmann et al. 2009). The co-expression of hDer1-eGFP and Tip3-mRFP indicated that these proteins have the same subcellular localization, but the fluorescence was distributed differently (Fig. 3-14c). To combine all the results Tip3 and Tip5 might localize in the membrane of cER, nuclear envelope and the host ER.

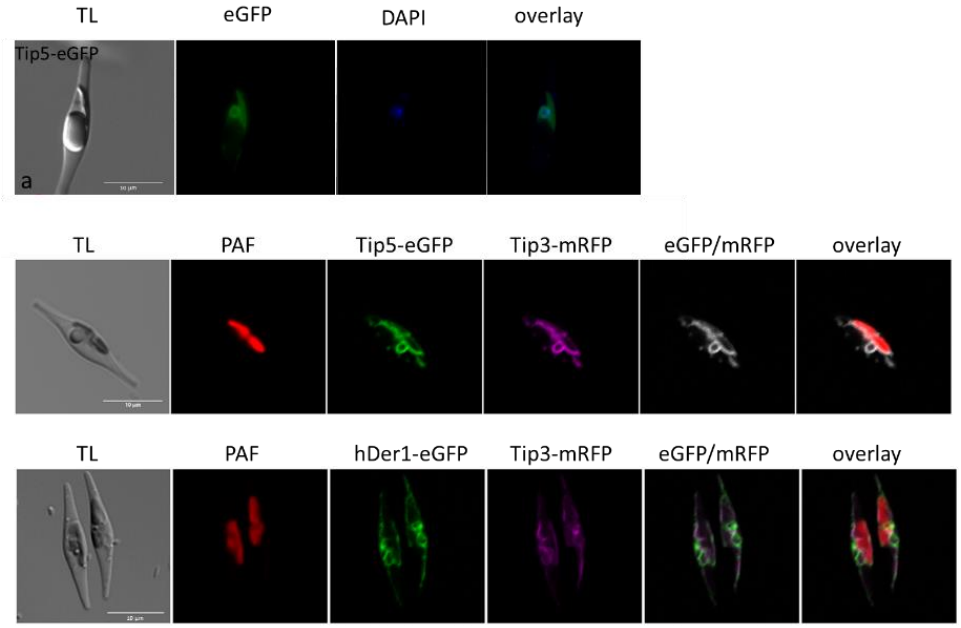


**Fig. 3-14: *In vivo* co-staining and co-expression analyses of Tip3 and Tip5 proteins.**

**a:** Co-localization of Tip5-eGFP with DAPI showed that partial fluorescence of the Tip5-eGFP was targeted to the nuclear envelope. **b:** Co-expression of Tip3-mRFP and Tip5-eGFP showed that these two proteins colocalize in the cER membrane and nuclear envelope. An additional fluorescence extended from the cER membrane and nuclear envelope was also observed. **c:** Co-expression of hDer1-eGFP and Tip3-mRFP showed that these two proteins are targeted to the same compartments (cER membrane, nuclear envelope and host ER), but the fluorescence was distributed differently. TL: transmitted light; PAF (red): plastid autofluorescence; eGFP (green): enhanced green fluorescent protein; DAPI (blue): 4',6-Diamidin-2-phenylindol; mRFP (Magenta): monomeric red fluorescent protein; overlay: the overlay of PAF, eGFP and mRFP.

### 3.2.2.4 *In vivo*-localization of Tip4

Likewise, the subcellular localization of Tip4 was analyzed in *P. tricornutum*. It was known that the complex plastid is surrounded by four membranes. The periplastidal membrane (PPM) is the second outermost membrane of the complex plastid which is located between the periplastidal

compartment (PPC) and the chloroplast ER (cER) space. Previous study showed that N-terminal bipartite targeting sequences (BTS) are specific signals for importing proteins into the complex plastid (PPC or PPM) (Gruber, Vugrinec et al. 2007). Interestingly, Tip4 is the only Tip which possesses a potential signal peptide.

Subcellular localization studies of Tip4-eGFP showed that the fluorescence pattern results in a classical "blob-like" structure known to represent a PPC/PPM localization, as shown in Fig. 3-15a. To further confirm this localization, the self-assembly GFP assay was performed. It was shown that the C-terminus of Tip4 was located in the PPC, as fluorescence signals could be observed upon co-expression of Tip4-S11 with a GFPS1-10 fused to a PPC marker but not fused to an ER marker (Fig. 3-15b). Compared with the subcellular localization of sDer1-2 (Gruber, Vugrinec et al. 2007, Hempel, Bullmann et al. 2009) Tip4 does not have an aromatic amino acid at the +1 position of the predicted transit peptide for plastid import. The PPM localization of Tip4 was concluded.

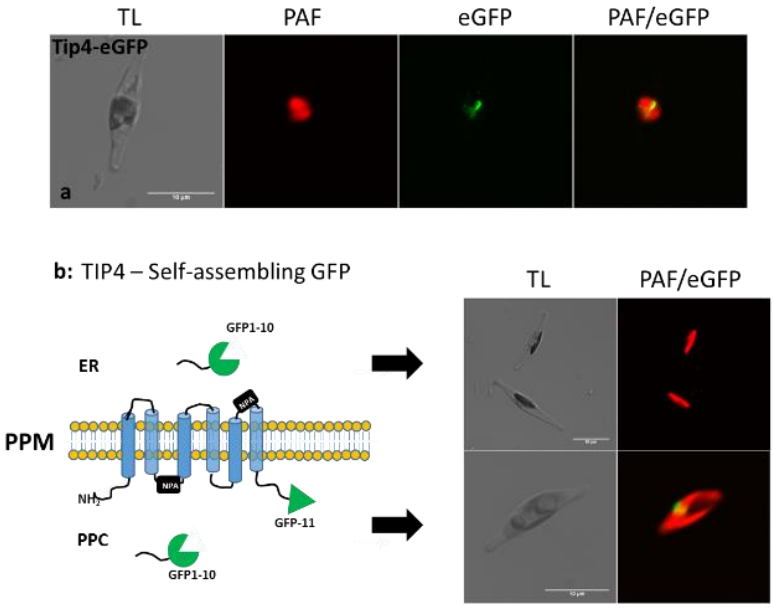


**Fig. 3-15: *In vivo* localization of Tip4-eGFP.**

**a:** Subcellular localization of Tip4-eGFP indicated that it was targeted a characteristic PPM/PPC-like compartments. **b:** The self-assembly GFP assay was used to confirm the localization of Tip4. GFP-11 was fused to the C-terminus of full-length of Tip4 and co-expressed together with Hsp70BTS-S1-10 (a PPC marker) and PDI-S1-10 (an ER marker), respectively. Only after co-expressed of Tip4-S11 with the PPC marker (lower pannel) but not after co-expression with the ER marker (upper pannel) a GFP signal was detectable. Self-assembly GFP assay results showed that the C-terminus of Tip4 is located on the PPC. TL: transmitted light; PAF (red): plastid autofluorescence; eGFP (green): enhanced green fluorescent protein.

### 3.2.3 *In vivo* localization of Golgi proteins

The Golgi apparatus consists of three main networks (the cis-Golgi network, the medial Golgi and the trans-Golgi network). In order to identify marker proteins for the slightly different Golgi networks, a large-scale search for proteins homologous to known Golgi localization proteins was conducted in the genome of *P. tricornutum*. Finally three proteins homologous to known enzymes involved in the N-Glycosylation pathway were selected for *in vivo*-localization in *P. tricornutum* (Table. 3-8).

**Table 3-8: Predicted subcellular localized Golgi marker proteins in *P. tricornutum*.**

(Pt: *P. tricornutum*; SP: signal peptidase; TMD: transmembrane domain; loc: localization). The identifications and cloning of GnT1 and FucT was conducted by Clément Ovide.

| name | BLAST hits/specific function | Pt ID | SP | TMD | Predicted-loc | eGFP-loc |
|---|---|---|---|---|---|---|
| N-acetylglucosaminyltransferase I (GnT1) | catalyse N-glycoproteins transported from ER to Golgi | 54844 | - | 1 | Cis-Golgi | Cis-Golgi |
| α-mannosidase I (α-Man I) | enzyme involved in N-glycan biosynthesis; Transfers an alpha-D-mannosyl residue from dolichyl-phosphate D-mannose into membrane lipid-linked oligosaccharide | 44425 | + | 6 | Cis-Golgi | unknown |
| β1,2-xylosyltransferase (XylT) | glycosyltransferase involved in N-glycan biosynthetic pathway | 45496 | + | 1 | Medial-Golgi | Medial-Golgi |
| α1,3-fucosyltransferase (FucT) | glycosyltransferase involved in N-glycan biosynthetic pathway | 54599 | + | 1 | Trans-Golgi/TGN | Trans-Golgi/TGN |
| α1,3-Mannosyltransferase (α1,3-Man) | enzyme involved in N-glycan biosynthesis; Transfers an alpha-D-mannosyl residue from dolichyl-phosphate D-mannose into membrane lipid-linked oligosaccharide | 22554 | + | - | Medial-Golgi | unknown |

The first one GnT1 is known to be localized in the cis-Golgi (Kajiura, Okamoto et al. 2012). The second protein XylT is a type II membrane proteins that belongs to the glycosyltransferase family 61 and has been confirmed to localize on medial cisternae of the Golgi apparatus (Pagny, Bouissonnie et al. 2003, Kajiura, Okamoto et al. 2012). The third one FucT is involved in the transfer of alfa-1,3-linked fucose residues to N-glycans and is expected to be targeted to the TGN (Fitchette-Lainé, Gomord et al. 1994, Breton, Mucha et al. 2001). By expressing all three as eGFP fusion proteins a dot-like structure or a long strip fluorescence pattern was observed (Fig. 3-16a/b/c). By carbonate extraction it could be shown that XylT is an integral membrane protein, as shown in Fig. 3-16d.

To confirm whether the localizations of these three Golgi marker proteins are slightly different, co-expression of XylT-mRFP/ GnT1-eGFP and XylT-mRFP/ FucT-eGFP was performed. It was shown that

the eGFP and mRFP fluorescence partially overlapped for both co-expressions (Fig. 3-16e/f). All these results showed that these three proteins were targeted into different cisternae of the Golgi apparatus.

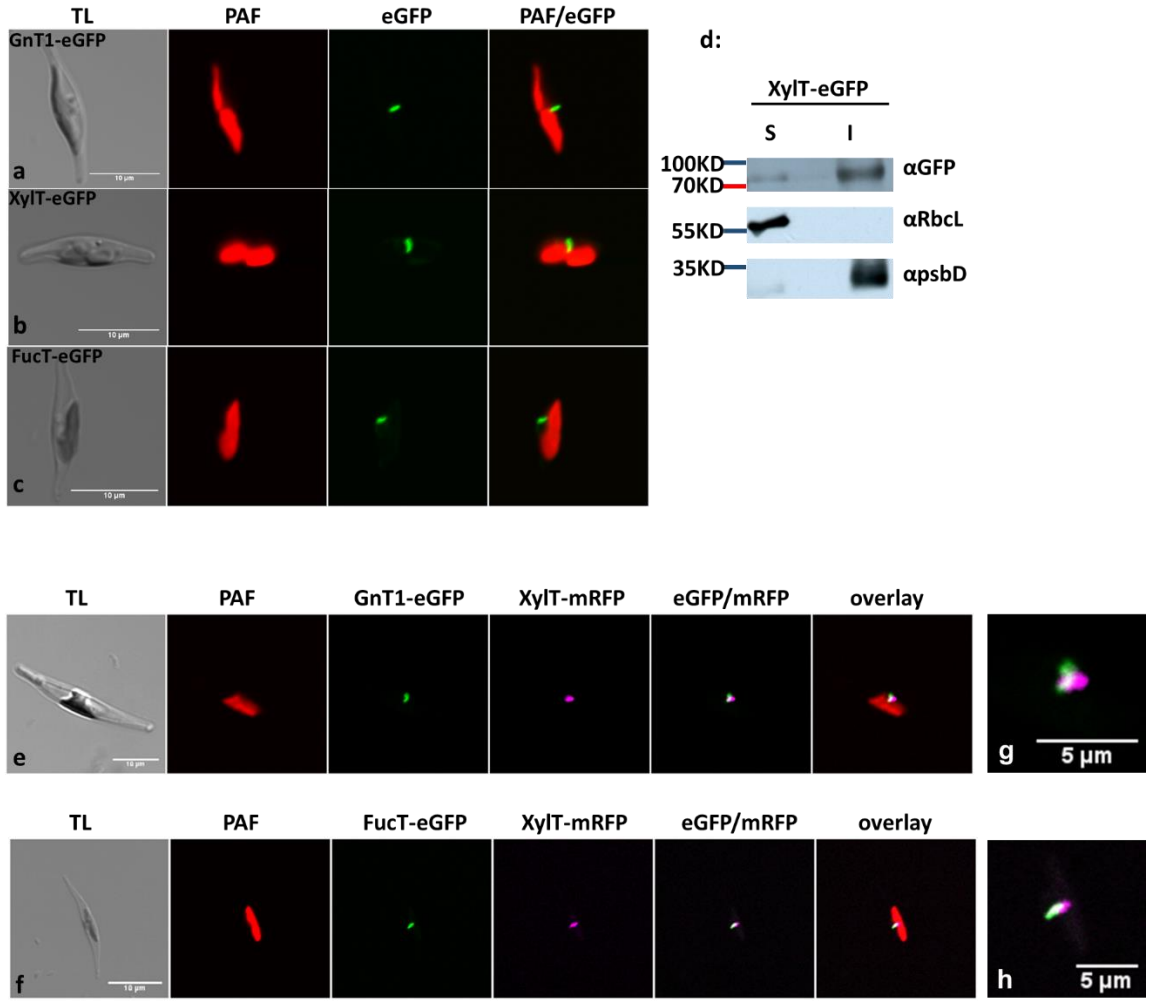


**Fig. 3-16: *In vivo* localization of proteins located in the Golgi apparatus.**

**a-c:** The GnT1-, XylT- and FucT-eGFP fusion proteins showed a similar dot-like or long strip-like fluorescence pattern. **d:** Carbonate extraction of XylT-eGFP. The thylakoid membrane protein PsbD (25 kDa) and the stromal protein RbcL (55 kDa) were used as makers for the fraction of soluble and integral membrane proteins. The expected molecular weight of XylT-eGFP proteins is 83 kDa. **e and f:** Co-expression of XylT-mRFP/ GnT1-eGFP and XylT-mRFP/ FucT-eGFP showed eGFP and mRFP fluorescence partially overlapped. **g and h**: The overlay of XylT-mRFP/ GnT1-eGFP and XylT-mRFP/ FucT-eGFP was enlarged, respectively. TL: transmitted light; PAF (red): plastid autofluorescence; eGFP (green): enhanced green fluorescent protein; mRFP (Magenta): monomeric red fluorescent protein; overlay: overlay of PAF, eGFP and mRFP.

### 3.2.4 *In vivo* localization of retromer complex

The structure and function of the retromer complex have already been well identified in yeast, animals and some plants, while the retromer complex in diatoms is still not much known. In *Arabidopsis thaliana*, the retromer complex comprises a small subcomplex (three sorting nexins) and

a larger subcomplex (Vps26, Vps29 and Vps35). In order to localize the putative retromer subunits in *P. tricornutum*, several proteins homologous to retromer subunits in *A. thaliana* were analyzed in *P. tricornutum* (Table 3-9).

**Table 3-9: Predicted subcellular localization of retromer complex proteins in *P. tricornutum*.**

(Pt: *P. tricornutum*; SP: signal peptidase; TMD: transmembrane domain; loc: localization)

| Retromer subunits | BLAST hits/specific function | Pt ID | SP | TMD | Predicted-loc | eGFP-loc |
|---|---|---|---|---|---|---|
| Vacuolar protein sorting-associated protein 26 (Vps26) | Inform multi-protein complexes, Intracellular protein transport | 41962 | - | - | TGN/EE | Golgi/EE |
| Vacuolar protein sorting-associated protein 29 (Vps29) | Inform multi-protein complexes , Intracellular trafficking, secretion and vesicular transport | 17936 | - | - | TGN/EE | Golgi/EE |
| Vacuolar protein sorting-associated protein 35 (Vps35) | Inform multi-protein complexes that facilitate retrograde transport of lytic vacuolar-targeting receptors back to the trans-Golgi network | 43830 | + | - | TGN/EE | Golgi/EE |
| Sorting nexin dimer (SNX1) | Membrane coat complex retromer, subunit VPS5/SNX1, sorting nexins, and related PX domain-containing proteins | 3137 | - | - | TGN/EE | unknown |

Homologs of the potential subunits Vps26 and Vps29 were selected for *in vivo* localization. The fluorescence patterns of these eGFP fusion proteins were similar and showed a dot or small diamond-like structure (Fig. 3-17a and b). In order to make sure that they are colocalization as predicted, the co-expression of Vps26-eGFP and Vps29-mRFP was performed. It was shown that these the fluorescence pattern overlapped completely (Fig. 3-17c and d). Taking together, it was speculated that two potential homologs (Vps26 and Vps29) of proteins in *A. thaliana* might also locate in one complex in *P. tricornutum*.

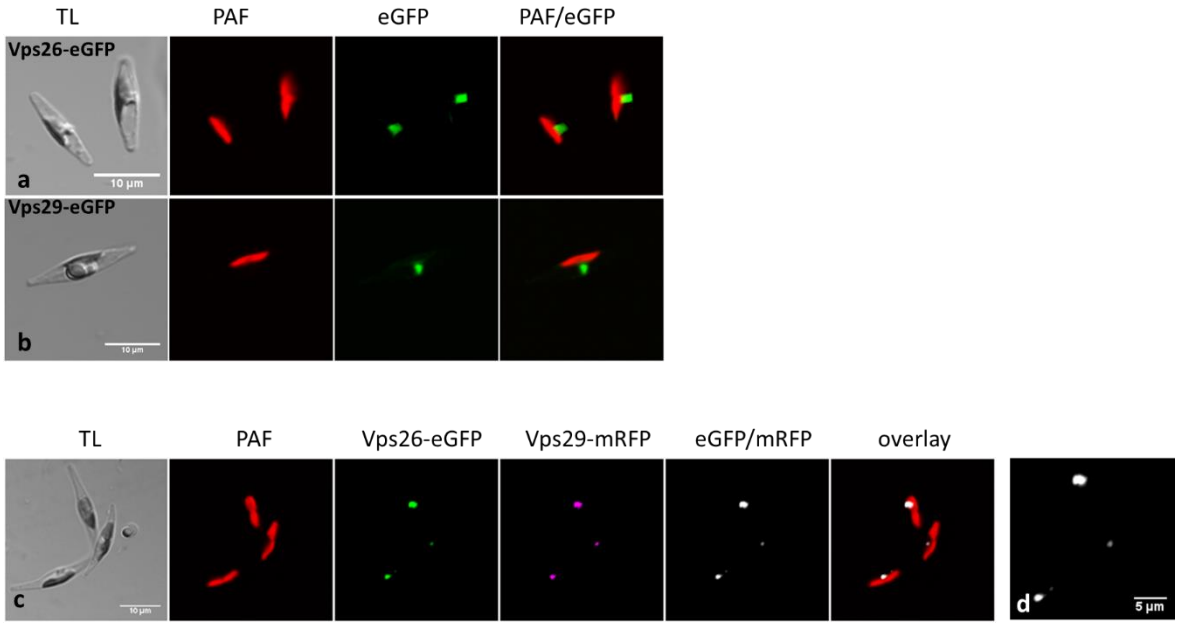

**Fig. 3-17: *In vivo* localization of retromer complex proteins.**

**a-b:** The Vps26- and Vps29-eGFP fusion proteins were expressed in *P. tricornutum*. It was shown that the fluorescence patterns are similar, a dot or a small diamond-like structure. **c:** The co-expression of these two potential retromer complex subunits showed that the fluorescence pattern overlapped. **d:** The overlay of Vps26-eGFP/ Vps29-mRFP was enlarged. TL: transmitted light; PAF (red): plastid autofluorescence; eGFP (green): enhanced green fluorescent protein; mRFP (Magenta): monomeric red fluorescent protein; overlay: the overlay of PAF, eGFP and mRFP.

In order to distinguish the Golgi marker proteins and retromer co-expression of one retromer subunit (Vps29) and the medial Golgi marker protein XylT was carried out. Only partial overlapped fluorescence was observed (Fig. 3-18a and enlarged picture b). However, the co-expression of the Vps29 together with the trans-Golgi marker protein FucT showed that their fluorescence overlapped completely (Fig. 3-18c and d). These co-localizations further indicated that the XylT and FucT was located on different cisternaes of the Golgi apparatus and the retromer subunit was targeted to the trans-Golgi network.

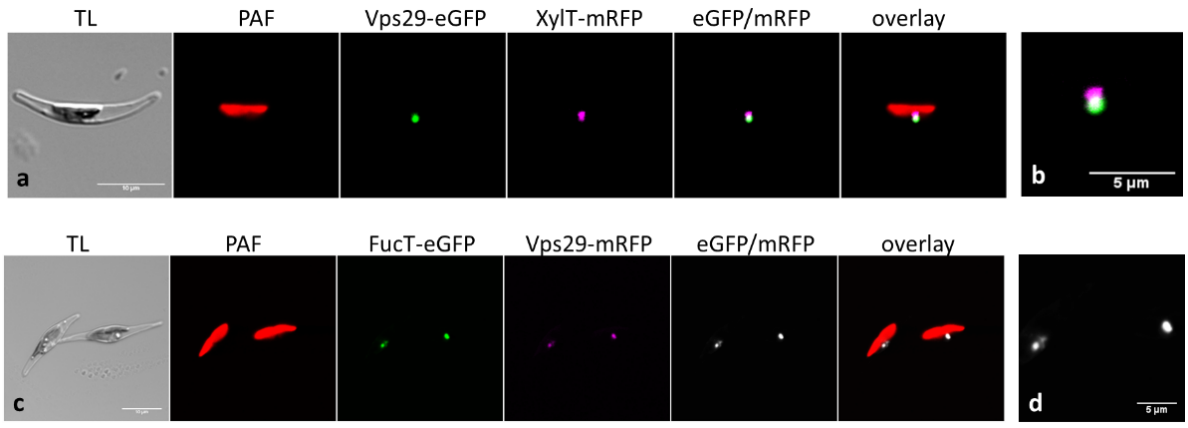


**Fig. 3-18: Co-expression of retromer complex proteins with Golgi markers.**

**a:** Co-expression of the medial Golgi marker XylT and the retromer subunit Vps29 showed that their fluorescence overlapped partially. **b:** The overlapped fluorescence of XylT-mRFP with Vps29-eGFP was enlarged. **c:** Co-expression of the trans-Golgi network marker FucT and the retromer subunit Vps29 showed the fluorescence overlapped completely. **d:** The overlapped fluorescence of FucT-eGFP with Vps29-mRFP was enlarged. TL: transmitted light; PAF (red): plastid autofluorescence; eGFP (green): enhanced green fluorescent protein; mRFP (Magenta): monomeric red fluorescent protein; overlay: the overlay of PAF, eGFP and mRFP.

### 3.2.5 *In vivo* localization of vacuolar $H^+$-ATPase proteins

Previous studies have already shown that two different vacuoles exist in plant cells, namely the central lytic vacuole and the protein storage vacuole (Marty 1999, Jin, Kim et al. 2001, Park, Kim et al. 2004). In order to mark these different types of vacuoles, more putative vacuolar candidates were

analyzed *in vivo* localization. It has already shown that vacuolar $H^+$-ATPases (VHAs) are localized to vacuoles and other endosomal membranes, including the ER, Golgi apparatus and vesicles as well as the plasma membrane (Sze, Schumacher et al. 2002). Therefore proteins homologous to VHAs of *A. thaliana* were analyzed *in silico* in *P. tricornutum* (Table 3-10).

**Table 3-10: Predicted subcellular localization of V-ATPase proteins in *P. tricornutum*.**

(Pt: *P. tricornutum*; SP: signal peptidase; TMD: trans-membrane domain; PM: plasma membrane; loc: localization)

| name | BLAST hits/specific function | Pt ID | SP (TargetP RC) | TMD | Predicted-loc | eGFP-loc |
|---|---|---|---|---|---|---|
| ATPase1 | Vacuolar H+-transporting two-sector ATPase, C subunit | 21882 | + | 4 | PM/Vacuole | Lytic vacuole? |
| ATPase2 | Vacuolar ATP synthase subunit G2; hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances; proton transport | 44050 | - | - | ATPase complex/ soluble | Soluble protein |

After *in silico* analyses, PtATPase1 with four potential TMDs and PtATPase2 without TMD were selected for *in vivo*-localization analyses. The fluorescence pattern of PtATPase1-eGFP showed a dot-like structure (Fig. 3-19a). In some clones several dot-like fluorescence could also be observed (Fig. 3-19b). By carbonate extraction it was shown that ATPase1 is an integral membrane protein (Fig. 3-19d). It was suggested that ATPase1 might be localized on the small vesicles or vacuoles in *P. tricornutum*. The second protein homologous to vacuolar $H^+$-ATPase in *A. thaliana* is called ATPase2 in *P. tricornutum*. Based on the prediction the N-terminus of ATPase2 amino acid sequence does not contain a hydrophobic signal peptide. When expressing the eGFP fusion construct it showed a cytoplasmic localization, the fluorescence was distributed in the cell (Fig. 3-19c). By carbonate extraction it was confirmed that ATPase2-eGFP is a soluble protein (Fig. 3-19e).

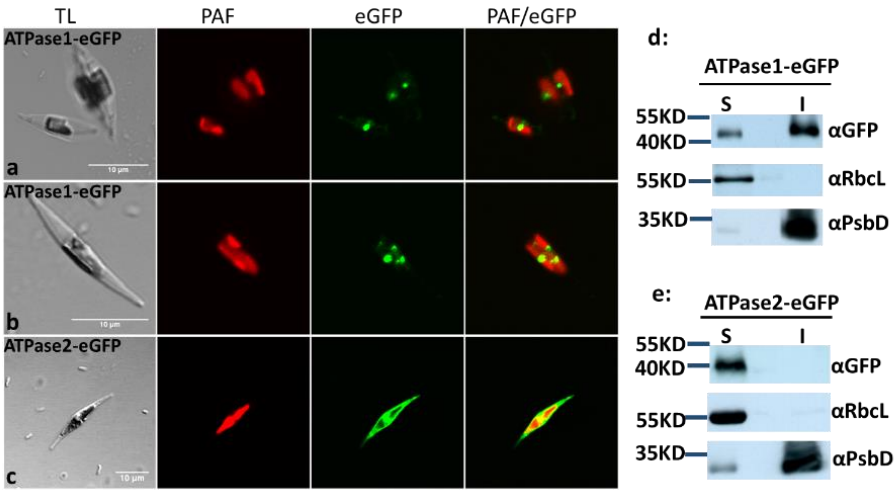


**Fig. 3-19: *In vivo* localization of V-ATPase proteins.**

**a/b:** ATPase1-eGFP was expressed under the control of a nitrate inducible promoter. Slightly different fluorescence pattern were observed, either one (a) or several (b) dot-like structures. ATPase2-eGFP showed a cytoplasmic localization. The fluorescence was distributed in the cell. **d/e:** Carbonate extractions of ATPase1-eGFP and ATPase2-eGFP proteins, respectively. The thylakoid membrane protein PsbD (25 kDa) and the stromal protein RbcL (55 kDa) were used as makers for the fraction of soluble and integral membrane proteins. The expected molecular weight of ATPase1-eGFP and ATPase2-eGFP are 45 kDa and 41 kDa, respectively. For a detailed description see text. TL: transmitted light; PAF (red): plastid autofluorescence; eGFP (green): enhanced green fluorescent protein.

# 4 Discussion

## 4.1 Genetic compartmentalization of peridinin-containing dinoflagellates

Peridinin-containing dinoflagellates are important members of phytoplankton and, as stated in the introduction, arose from secondary endosymbiosis of a red alga by a so far undefined host cell (Mcfadden 2001). In contrast to other groups, their complex plastid is surrounded by only three membranes (Dodge and Lee 2000). Apart from that peridinin-containing dinoflagellates stand out by additional unusual features with regard to their genome organization (Zhang, Cavalier-Smith et al. 2002). Besides their extraordinary large nuclear genome which is organized in permanently condensed, para-crystalline chromosomes (Bodansky, Mintz et al. 1979, Gautier, Michel-Salamin et al. 1986, Bouligand and Norris 2001, Chow, Yan et al. 2010) they possess so called minicircles. Within this thesis a new method transposon-insertion based approach for minicircle isolation from one representative of the peridinin-containing dinoflagellates was used and finally isolated individual minicircles were characterized.

### 4.1.1 Minicircles of the peridinin-containing dinoflagellate *Amphidinium carterae* CCAM0512

Since the discovery of minicircles of peridinin-containing dinoflagellats in 1999 (Zhang, Green et al. 1999) several minicircles of diverse peridinin-containing dinoflagellates were isolated and described such as in *A. operculatum*, *A. carterae* CS21, *A. carterae* CCAP1102/6, *Ceratium horridum*, *Adenoides eludens*, *Heterocapsa* species as well as *Symbiodinium* species (Zhang, Green et al. 1999, Barbrook and Howe 2000, Barbrook, Symington et al. 2001, Hiller 2001, Zhang, Cavalier-Smith et al. 2002, Moore 2003, Laatsch, Zauner et al. 2004, Nisbet, Koumandou et al. 2004, Nelson and Green 2005, Barbrook, Santucci et al. 2006). Coding for genes normally found in conventional plastid genomes were located on these small plasmid-like minicircles. Previous studies have already shown that the minicircle contains one to three coding genes (coding region), the remainder of the minicircle was called non-coding region (Barbrook and Howe 2000). It was also shown that several minicircles without coding region do not have potential known function (also called empty minicircles) (Hiller 2001). Within this thesis a transposon-based method was used to isolate minicircles. So far, transposons were predominantly used in bacteria, plants and animals for mutagenesis or some other molecular biology studies (Kim, Vanguri et al. 1998, Balciunas and Ekker 2005, Carlson, Frandsen et al. 2005, Schnable, Ware et al. 2009, Hackett, Largaespada et al. 2010). The transposon-based approach is in so far different, that it is not based on known sequences and thereby has the capacity to isolate minicircles with deviant or unknown sequence. The structural nature of plasmid-

like minicircles provides the basis for transposon insertion and subsequent usage of bacterial proliferation.

Within this thesis, by using the transposon-insertion based approach hundreds of thousands of potential positive colonies grew on plastes. Based on the plasmid preparation and sequencing it was found that 89 out of 107 (83.2%) sequencing samples are positive minicircles, 18 out of 89 (20.2%) minicircles are gene-containing minicircles, 71 out of 89 (79.8%) minicircles are empty minicircles (Table 3-1). The number of empty minicircles is about four times the number of gene-containing minicircles. A possible explanation for this could be that the empty minicircles are present at higher copy number than the gene-containing minicircles in *A. carterae* CCAM0512 cells. But this can not rule out the possibility that the different division ratio of the *E. coli* cells harboring positive minicircles in liquid LB medium without antibiotic (see materials and methods 5.2.4.12). 83.2% sequencing sameples are positive minicircles, it was indicated that the transposon-insertion based approach is a considerable efficient method. The isolated individual minicircles could work as a vector and modified by different reporter gene (e.g. eGFP gene) for further research.

In the following points the implications deduced from the analysis of these different minicircles are discussed.

#### 4.1.1.1 Minicircles with coding genes

Of the isolated minicircles 18 out of 89 minicircles are gene-containing minicircles. Within these 18 minicircles it was found that 12 minicircles belong to different minicircle molecules. The 12 minicircles include three different *psbA*-containing minicircle molecules, two different *petB*/*atpA*-containing minicircle molecules, two different *23S rRNA*-containing minicircle molecules and one *psaB*-, *psbC*-, *psbD/E/I*-, *petD*- and *atpB*-containing minicircle molecule (Table 3-2). These different minicircle molecules exhibit open reading frames of considerable length which encode for known proteins PsbA, PsaB, PsbC, PsbD, PsbE, PsbI, PetB, PetD, AtpA and 23S rRNA(Table 3-2). The lengths of these individual minicircle molecules (2333 bps – 2664 bps) are similar to the sizes of minicircles reported before such as *A. carterae* CS21 (2327 bps – 2713 bps), *A. operculatum* (2311 bps – 2713 bps) and *H. triquetra* (2151 bps – 3121 bps) (Zhang, Green et al. 1999, Zhang, Cavalier-Smith et al. 2002, Barbrook, Santucci et al. 2006). Compared to the minicircles in these species the length of minicircles in *Ceratium horridum* are much larger 5200 bps - 6700 bps (Laatsch, Zauner et al. 2004). The average length of eight different groups of gene-containing minicircles is 2474 bps. The variable lengths of minicircle might result from the different evolutions and rearrangements of minicircles in dinoflagellates. Most of the isolated minicircles are single-gene minicircles while the *petB*/*atpA* and *psbD*/*E*/*I* gene pairs are encoded on one single minicircle of 2596 bps and 2354 bps, respectively.

This is in line with the already reported two- or three- genes minicircles in other species (Barbrook, Santucci et al. 2006). Previous studies showed that the *petB*/*atpA* genes, *psbD*/*E*/*I* genes on a single minicircle are not normally adjacent to each other in reported conventional plastid genomes in other plants and algae (Ohta, Matsuzaki et al. 2003, Puerta, Bachvaroff et al. 2005). The observed synteny suggests that multiple genes on a single minicircle were generated by tandem rearrangement of two or three single-gene minicircles, rather than by fragmentation of a multi-gene molecule (Howe, Nisbet et al. 2008).

From the point of the gene transcription, the same as the transcriptions of genes that have been studied previously (Zhang, Green et al. 1999, Barbrook and Howe 2000, Barbrook, Symington et al. 2001, Takishita, Ishida et al. 2004), it was shown here that the *psbA* gene might be also transcribed in *A. carterae* CCAM0512. As for multiple genes minicircles previous studies suggested that these genes (*petB*/*atpA* and *psbD*/*E*/*I* genes) on a single minicircle are transcribed separately as the sizes of transcripts observed were the same as the single genes (Howe, Nisbet et al. 2008, Nisbet, Hiller et al. 2008), latter a polycistronic transcript was suggested which is produced first and subsequently cleaved rapidly into two or three separate transcripts (Dang and Green 2010, Barbrook, Dorrell et al. 2012). It was demonstrated that the the minicircles were transcribed via a rolling circle model in the dinoflagellates *Hetercapsa triquetra* and *A. carterae* CCAP1102/6 (Dang and Green 2010, Barbrook, Dorrell et al. 2012). It was shown that the transcripts contain several ORFs which were not known previously to be expressed (Barbrook, Dorrell et al. 2012). Simultaneously, Dang and Green showed that the transcripts are even larger than the minicircle itself (Dang and Green 2010).

There is no doubt that the plastid genomes of dinoflagellates are the smallest compared to other conventional plastid genomes. Previous studies suggested that this shrunken plastid genome might stem from the transfer of the normally plastid-located genes to the host genome and dramatically reducing and deleting of the chloroplast genome during the course of plastid acquisition (Bachvaroff, Concepcion et al. 2004, Green 2004, Hackett, Yoon et al. 2004, Tanikawa, Akimoto et al. 2004, Patron, Waller et al. 2005). Compared with all about 15 reported protein-coding genes on minicircles psaA, psbB and small subunit ribosomal ribonucleic acid (SSU-rRNA) minicircles are still missing in this work. Compared with 50 – 200 genes in other typical plastid genomes not only the tRNA but genes encoding ribosomal proteins and several small proteins for photosynthesis and the ATPases are still absence on known minicircles. A possible explanation for this could be that these genes have already transferred into the nucleus. It's also possible that these genes are not as abundant as the other genes in *A. carterae* cells, thus it is hard to be isolated here. In initial studies putative f-Met tRNA gene was identified in *A. carterae* CS21 and *A. carterae* CCAP1102/6 and putative proline and tryptophan tRNA genes were found in *Heterocapsa* species (Barbrook, Santucci et al. 2006).

Whereas tRNA gene was not found so far in *A. carterae* CCAM0512. It was shown that 15 retained genes on minicircles encode subunits of the main complexes that involve in the light reactions of photosynthesis such as photosystem I (PSI), PSII, the ATP synthase and the cytb6/f complex, as well as rRNAs proteins (23S rRNA and 16S rRNA) and tRNA (Hiller 2001, Koumandou and Howe 2007, Howe, Nisbet et al. 2008). All genes encoded on minicircles in *A. carterae* CCAM0512 use the conventional genetic code with the same start codon (ATG) and three different stop codons (TAG, TGA or TAA). In *A. operculatum* the *psaA* and *psbB* genes on minicircles do not have standard start codon in the expected positions, while utilize GTA as an initiation codon (Barbrook and Howe 2000). In terms of *Heterocapsa triquetra* ATA, TTG and ATG were used for the initial coding (Zhang, Green et al. 1999). GTG becomes a potential start codon in *petB* gene containing minicircle in *A. carterae* CS-21 (Hiller 2001). The variety of genetic codons might provide an important hint for evolutionary relationship of minicircles.

RNA editing is a post-transcriptional process that can insert, delete and substitute nucleotides in mRNA prior to translation to proteins. RNA editing can be widely observed on transcripts from nuclear or organellar genomes (including mitochondrial and plastid) in different species (Miyata and Sugita 2004, Wang and Morse 2006, Grosche, Funk et al. 2012, Takenaka, Zehrmann et al. 2013, Mungpakdee, Shinzato et al. 2014). However, RNA editing is very rare to be observed in plastid-encoded RNAs in algae. Only three editing events were found so far in the minicircle-encoded plastid genes in *Ceratium horridum*, *Lingulodinium polyedrum* and *Heterocapsa triquetra* (Zauner, Greilinger et al. 2004, Wang and Morse 2006, Dang and Green 2009). In contrast to previously reported organellar RNA editing in peridinin-containing dinoflagellates, RNA editing was not observed on transcriptions of minicircles of *A. carterae* CCAM0512 based on the analysis of nine coding minicircles genes (eight genes were analysed by Grosche C.(Grosche 2012)).

Previous studies have already shown that the minicircles of dinoflagellates contain a highly conserved core region (Zhang, Green et al. 1999, Barbrook, Symington et al. 2001, Zhang, Cavalier-Smith et al. 2002). The core regions are species-specific and are unstable in length (Koumandou, Nisbet et al. 2004). The highly diverse core regions of the four *A. carterae* strains indicated that the core regions are only identical within the strain but considerable variation between different strains (Fig. 3-3/7). The function of the core region remains uncertain. Based on the prediction the core region has a putative promoter sequence in *A. carterae* CCAM0512 (Fig. 3-3). Previous studies suggested that the initiation of replication usually possess multiple direct and inverted repeats in plastids (Sears, Stoike et al. 1996, Kunnimalaiyaan, Shi et al. 1997). It was speculated that the core region is responsible for the maintenance of the copy number, the initiation of replication and/or the transcription of minicircles (Zhang, Green et al. 1999, Barbrook and Howe 2000, Barbrook,

Symington et al. 2001, Hiller 2001, Zhang, Cavalier-Smith et al. 2002). It was also inferred to have a function on the membrane attachment (Howe, Nisbet et al. 2008). As mentioned above, no obvious direct or inverted repeat was observed in *A. carterae* CCAM0512. Therefore it is hard to speculate the function of the core region.

#### 4.1.1.2 Empty minicircles

In addition to the minicircles with obvious coding region, seventy-one empty minicircles were also isolated (Table 3-3). Of the total 71 empty minicircles, 26 (36.6%) belong to the different empty minicircle molecules. The 26 different empty minicircles were divided into six distinct groups according to the conserved sequence. For example, the group of J8 empty minicircle contains three different minicircle molecules in 17 minicircles. These empty minicircles have recognizable core regions and include short open reading frames but potential products are without any known homology. The average length of empty minicircle is 1986 bps, which is much smaller than gene-containing minicircles (2474 bps). A direct reason is that the empty minicircles do not contain the coding genes. The big difference of the lengths could also give a valuable hint for the regular rearrangement of these minicircles during the evolution. Previous analysis showed that the transcription of empty minicircles occurred in *A. carterae* CCAP1102/6 and *A. carterae* CS21 and this transcription cover the large part of the minicircle including the core region (Nisbet, Hiller et al. 2008). As the transcriptions of three largest open reading frames in empty minicircles were observed in *A. carterae* CCAM0512 (Fig. 3-4), the open reading frames of empty minicircles could be transcribed and might be functional. Based on the comparison of the open reading frames (> 150 bps) it was found that empty minicircle J9 and *atpB*-containing minicircle have a completely identical open reading frame and a partially identical open reading frame (Table 3-4). This could be explained by that the empty minicircles were generated by the rearrangement between the gene-containing minicircles. At the same time, it was shown that empty minicircles J13, J22 and J36 have an identical open reading frame, empty minicircles J22 and J24 have an identical open reading frame. A possible explanation for these identical open reading frames could be that the minicircles were generated by tandem rearrangement of several different fragments during the evolutionary process, and these fragments might be abundant in the cells. It could be speculated that these identical open reading frames could be translated to functional proteins, but these proteins remain unknown so far.

A summary of GC content is shown in Fig. 3-5. Except the GC content of *psbA*-containing minicircle is 45%, which is the same as or lower than the GC content of the overall empty minicircles, all the other gene-containing minicircles present to be more GC-rich than the empty minicircles. This discrepancy might be important for the stability and expression of the coding regions.

These empty minicircles may be unique to dinoflagellates. There remains many questions about the empty minicircles: 'how did they evolve? Where do they come from?' and 'What is the function of the ORFs on empty minicircles? Do they translate?' Nisbet et al. explained that the small empty minicircles might generate from a homologous recombination and internal deletion event between the minicircles with coding region (Nisbet, Koumandou et al. 2004). It was also speculated that empty minicircles may come from 'parasitic' DNA elements (Howe, Nisbet et al. 2008). As the empty minicircles do not have any identifiable coding characteristics, the functional analysis of these empty minicircles remains hard.

### 4.1.2 The evolutional relationship of minicircles

So far, fourteen gene-containing minicircles were isolated from four *A. carterae* strains (Tab. 3-5), *A. carterae* CCAM0512 (in this work), *A. carterae* CS21 (Hiller 2001, Barbrook, Santucci et al. 2006), *A. carterae* CCAP1102/6 (Barbrook and Howe 2000, Nisbet, Koumandou et al. 2004) as well as *A. carterae* CCMP1314 (Zhang, Cavalier-Smith et al. 2002).

Our present results revealed that the core regions are highly identical in *A. carterae* CCAM0512 (Fig. 3-3), but the core regions of four *A. carterae* strains are apparently unrelated and cannot be mutually aligned (Fig. 3-7). An alignment of the non-coding regions of the four psbA minicircles showed that upstream and downstream the *psbA* gene are conserved. The conserved regions might be useful for the stability of the gene expression (Fig. 3-8). Together with this, a conserved sequence of 50 bps upstream of the *psbA* gene and downstream of the core regions might have a function on the transcription initial based on the prediction. It was shown that the core regions of these four *A. carterae* strains are located in the variable region. On the contrary, the coding regions (e.g. psbA minicircles) of these four *A. carterae* strains showed a considerable high identity (more than 97%) based on the molecular phylogenetic analysis (Fig. 3-6). From the phylogenetic analyses of LSU rDNA and SSU rDNA sequences, these four *A. carterae* strains were found to be a sister relationship in a clade containing *A. carterae* with high statistical support (Fig. 3-9 and Fig. 3-10).

Heretofore, several empty minicircles were identified including 10 empty minicircles in *A. carterae* CS21 and 5 empty minicircles in *A. carterae* CCAP1102/6 (Barbrook, Santucci et al. 2006). Within this thesis, six distinct groups of empty minicircles were identified in 26 different empty minicircle molecules in *A. carterae* CCAM0512. Among all small open reading frames 27 open reading frames (the length > 150 bps) were compared with that in *A. carterae* CS21 and *A. carterae* CCAP1102/6 (Table 3-6). It was found that 7 out of 27 pairs of open reading frames have a high query cover, low e-value and high identity. Among these three pairs of open reading frames have a 100% query cover, very low e-value and considerable high identity (83%, 98% and 96%). It was speculated that these

open reading frames might have a specific function in dinoflagellates, but it has to be comfirmed in the future. The presence of conserved open reading frames among these different *A. carterae* strains suggested that these open reading frames are evolutionary conserved and are related to each other.

Taken all these analyses together, it was shown that nuclear encoded LSU and SSU and minicircle encoded psbA are highly identical in these four *A. carterae* strains, while the core regions are completely different. A hypothesis was put forward to suggest that the core regions of minicircles evolved at a very fast speed. As the core regions are highly diverse within the different strains, this could be used to efficiently distinguish different strains in dinoflagellates such as the toxic and non-toxic strains.

## 4.2 "The endomembrane system (ES) in *Phaeodactylum tricornutum*"

As already mentioned the endomembrane system is made up of the different organellar membranes including the nuclear envelope, ER, Golgi apparatus, lysosomes or vacuoles, vesicles, endosomes and the plasma membrane. In eukaryotic cells, proteins are encoded in the nuclear genome and synthesized in the cytoplasm, some of them must be transported to the different subcellular compartments such as the plasma membrane, lysosomes/ vacuoles or the extracellular. Only if the proteins are targeted to their appropriate final destinations, they can perform required function. However, the mechanisms for trafficking of proteins during the endomembrane system *in vivo* in the diatom *P. tricornutum* remain poorly understood. The known marker proteins in *P. tricornutum* are not enough. Consequently, the aim of this project was to identify different marker proteins localized in the different subcellular compartments and therefore provide an essential condition for further detailed researches about the protein trafficking on endomembrane system.

### 4.2.1 Identification of tonoplast intrinsic proteins (Tips)

Water and other small molecules across the membrane is largely controlled by membrane channels called tonoplast intrinsic proteins (Tips) (Höfte, Hubbard et al. 1992). Tips together with another four groups the plasma membrane intrinc proteins (PIPs), nodulin 26-like intrinsic proteins (NIPs) and small basic intrinsic proteins (SIPs) belong to the aquaporins family. The aquaporins are channel proteins belong to the major intrinsic protein family (MIPs) (Johanson and Gustavsson 2002). Subcellular localization of several Tip isoforms (including three gamma-Tip (Tip1), three delta-Tip (Tip2), the seed specific alpha- and beta-Tip (Tip3:1 and Tip3:2), one epsilon-Tip (Tip4:1) and one zeta-Tip (Tip5:1)) have already been identified or predicted *in vivo* in *A. thaliana* (Höfte, Hubbard et al. 1992, Johanson, Karlsson et al. 2001). Previous study has already shown that these channel proteins are distributed to specific developmental stages and tissue types in plant cells (Rivera-Serrano, Rodriguez-Welsh et al. 2012). It was identified that Tip1:1 (gamma-Tip) and Tip2:1 (delta-Tip) proteins are targeted to the tonoplast of the central vacuole in transgenic *A. thaliana* mature roots, root tips and leaves (Hunter, Craddock et al. 2007). While in embryos of seeds Tip1:1 and

Tip2:1 are transported to the protein storage vacuoles (Karlsson, Johansson et al. 2000, Saito, Ueda et al. 2002, Gillespie, Rogers et al. 2005, Hunter, Craddock et al. 2007, Schüssler, Alexandersson et al. 2008, Gattolin, Sorieul et al. 2009). Subcellular localization of eGFP-fused Tip protein indicated that AtTip4:1 and AtTip1:2 were localized mainly on the tonoplast membrane and other uncharacteristic endomembranes (Liu, Ludewig et al. 2003).

Within this thesis an initial localisation for five Tip proteins homologous to Tips of *A. thaliana* is present first in *P. tricornutum*. The use of eGFP and mRFP fusions to these Tips cDNA sequences allowed us to investigate the subcellular localization *in vivo*. It was shown that Tip1-eGFP was obviously targeted to the plasma membrane in *P. tricornutum*, and additional dot-like fluorescence was also observed in some clones as shown in Fig. 3-11. However, the fluorescence pattern of the Tip1-eGFP was not completely consistent with the fluorescence of plasma membrane protein-PDZ2, which was solely and equally distributed to the plasma membrane (Stork 2013). In order to explain the localization of these dot-like structures co-expression of Tip1-eGFP with FM4-64 was performed. FM4-64 is specific used for staining the plasma membrane and to follow endomembrane system-dependent internalization processes (Rigal, Doyle et al. 2015). The overlapping of these dots with dots stained by the dye FM4-64 indicated that the dot-like structures belong to the endomembrane system. Based on these results, Tip1-eGFP fusion was speculated to localize on the plasma membrane and endosomal vesicles. This is important and the first time to show the recycling of plasma membrane proteins in diatoms. It is still unknow whether the endocytosis happens in diatoms or not. The identification of the recycling plasma membrane in diatoms is very useful for the endocytosis research. The subcellular localization of Tip2 showed that Tip2-full length-eGFP fusion protein was typically targeted to the vacuolar-like membrane, as shown in Fig. 3-12. As the model of Tip2 is not completely supported by expressed sequence tags (ESTs), the Tip2 gene was amplified from cDNA. The alignment of Tip2 nucleotide sequences see supplements 7.3. Tip2 is the first identified vacuolar-like marker protein in diatoms in *P. tricornutum*. Thus, it is very important for investigating the mechanisms of vacuolar protein transport. At the same time, according to the predictions the N-terminus of Tip2 protein sequence is lack of signal peptide, and Tip2 contains six conserved transmembrane domains, three extra loops and two-hydrophobic intracellular loops. Therefore, it will be very meaningful and interesting to mutate several potential targeting signal for studing the secretory pathway of Tip2, such as the mutations of the N-terminal amino acid, C-terminal amino acid and the amino aicd on the their loops.

Endosymbiotic events gave rise to the formation of different groups of organisms with the so-called complex or secondary plastids. The complex or secondary plastid is surrounded by four membrane structures in *P. tricornutum*. The outermost membrane of the complex plastid called chloroplast ER

(cER membrane) membrane, it is continuous with the outer membrane of the nuclear envelope and was thought to connect with the host ER (Gibbs 1979, Cavalier-Smith and Chao 2003). Subcellular localization of eGFP-fused Tip3 and Tip5 indicated that they localized on the outermost membrane of the complex plastid, nuclear envelope and host ER (Fig. 3-13). It was observed that the immuno gold particles are predominantly present in the outermost membrane of the plastid (cER membrane) and partially present in the nuclear envelope and host ER by applying the electron microscopy. Based on the absence of N-terminal targeting signal sequence (BTS) in both Tip proteins, it was further indicated that Tip3 and Tip5 locate on the cER membrane. In comparison with an important component of the host ERAD machinery (cER membrane, nuclear envelope and host ER marker-hDer1) (Hempel, Bullmann et al. 2009), it was observed that to some extent the fluorescence pattern of three proteins are similar. Co-expression of Tip3-mRFP with hDer1-eGFP showed the overlapping of the fluorescence but the distribution of the fluorescence is different. Taken all these results together, Tip3- and Tip5-eGFP fusions were transported to the outermost membrane of the complex plastid (cER membrane), nuclear envelope and host ER in *P. tricornutum* (Fig. 3-13/14). All these makers are important for investigating protein transport across these ER membranes. These abundant ER membrane markers provide a valuable insight for distinguishing different structures of ER membranes and better investigating the function of these ER structures.

Subcellular localization of eGFP-fused Tip4 indicated that Tip4 was inserted into the second outermost membrane (PPM). Several facts support this speculation. Firstly, the fluorescence pattern is typical PPM localization, which is similar to the fluorescence distribution of the identified PPM marker PtE3P (Hempel, Felsner et al. 2010). Secondly, Tip4-eGFP is a membrane protein by the carbonate extraction, as shown in Fig. 3-15. Thirdly, the self-assembly GFP assay indicated that the C-terminus of Tip4 is localized to the PPC. Last but not the least, Tip4 is the only one identified aquaporin protein of these five Tip homologous proteins with a predicted signal peptide and an N-terminal extension, which could lead the protein transport into the PPC or the PPM. Previous study showed that only at the +1 position of the potential signal peptidase cleavage site is an aromatic amino acid, the plastid protein is allowed to import (Gruber, Vugrinec et al. 2007). However, at the +1 position of the Tip4 potential signal peptidase cleavage site is not an aromatic amino acid. Taken all these evidences together, Tip4 should be targeted to the PPM of the complex plastid.

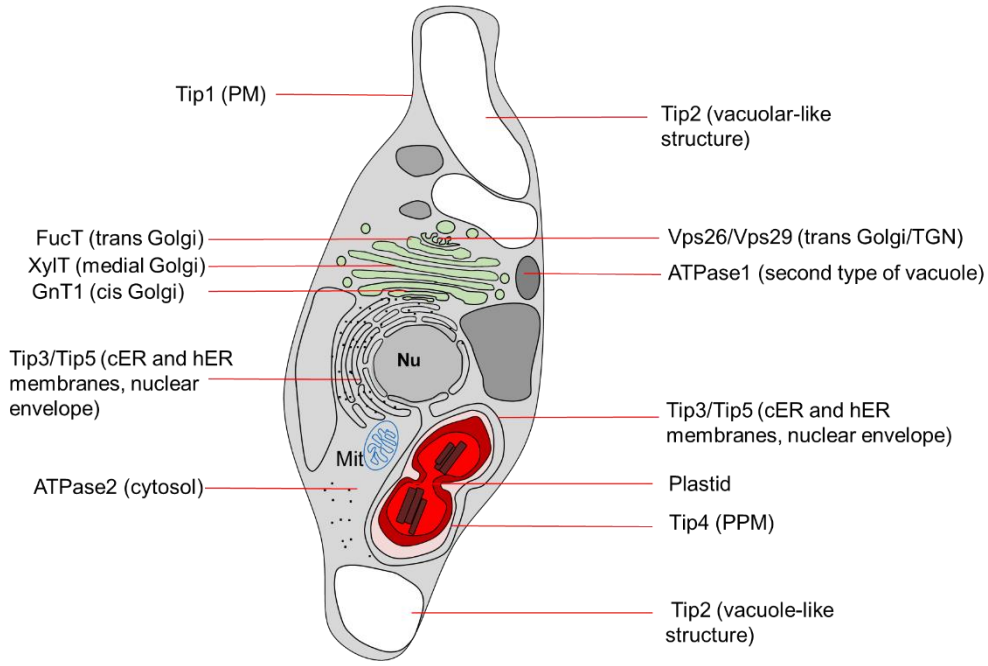


**Fig. 4-1: Schematic overview of identified marker proteins in diatom *P. tricornutum*.**

Proteins which have been identified within this thesis were shown in this figure. The markers of different subcellular compartments include five Tips, Tip1 (plasma membrane (PM) and endosomal compartments), Tip2 (vacuolar-like structure), Tip3 and Tip5 (nuclear envelope, cER and host ER (hER) membranes) and Tip4 (second outermost membrane of the complex plastid, PPM), two subunits of the putative retromer Vps26 and Vps29 (trans Golgi network), three Golgi apparatus markers GnTI (cis Golgi network), XylT (medial Golgi network) and FucT (trans Golgi network), and two V-ATPases ATPase1 (probably in a second type of vacuole) and ATPase2 (cytosol). Nu: nucleus, Mit: Mitochondria.

All in all, the subcellular localizations of five Tips (Tip1-5) were addressed in *P. tricornutum* (Fig. 4-1). Contrary to the Tips in higher plants five Tips was not only targeted to the tonoplast in *P. tricornutum*. Tip1 is targeted to the plasma membrane and endosomal vesicles, Tip2 is located on the vacuolar-like structure, Tip3 and Tip5 are enriched on the cER membrane, nuclear envelope and host ER, Tip4 is a PPM protein. These results indicated that the subcellular localizations of the protein candidates are not always matched with the predictions, the localizations could be completely different in different organisms. A possible explanation for this could be that not all these five Tips are really homologous to tonoplast intrinsic proteins in higher plants, such as Tip1 might be homologous to one of the plasma membrane intrinsic proteins (PIPs) or the nodulin 26-like intrinsic proteins (NIPs) which localized in the plasma membrane and ER in higher plants, Tip3 and Tip5 might be homologous to the small basic intrinsic proteins which localized in the ER in higher plants (Johansson, Karlsson et al. 2000, Quigley, Rosenberg et al. 2002, Pandey, Sharma et al. 2013). Another possible explanation for this could be that the expression of these Tips relies on the specific developmental stages and tissue types in higher plants (Rivera-Serrano, Rodriguez-Welsh et al. 2012). The proteins expressed in different conditions might effect their localizations. Thus, it can not be ruled out the reason from the different organisms. About four hundred aquaporins were identified

in recent study in ten different plants containing monocots and dicots (Regon, Panda et al. 2014). Among them, *Citrus sinensis* (57), *Fagaria vesca* (42), *Sorghum bicolor* (40) and *Zea mays* (43) have higher aquaporin genes than the others. 38 aquaporin genes and 11 Tips were identified in *A. thaliana* (Regon, Panda et al. 2014). Regarding to the diatom *P. tricornutum*, 42 major intrinsic proteins were found in the database. So far, only one Tip protein (Tip2) homologous to Tip of *A. thaliana* is a tonoplast intrinsic protein in *P. tricornutum*.

Previous studies have already shown that the localizations and amount of Tips in higher plants could be changed by stressed conditions such as the salt exposure and dark adaptation (Boursiac, Chen et al. 2005, Uenishi, Nakabayashi et al. 2014). Thus, the identified five Tips in *P. tricornutum* could also be used to study the expression and localization in different environmental stress conditions. Compared with most aquaporins in higher plants, it was shown that these five Tips also have either one or two NPA motifs in their sequences and six transmembrane domains (Table 3-7). To some extent, this suggested that all the Tips are closely related and conserved during the process of evolution. The mutation of these NPA motifs will be a method to detect its function. The function of these five Tips remains unknown in *P. tricornutum*. It is still unknown whether they could form an aquaporin pore to selectively mediate the transport of water, gases and small neutral solutes. Therefore, the identification of the localizations of these candidate proteins will be very important for the studying of these proteins' function.

### 4.2.2 Identification of Golgi-marker proteins

In most eukaryotes, the Golgi apparatus is divided into the cis-Golgi apparatus, the medial Golgi apparatus and the trans-Golgi apparatus. As already mentioned in the introduction the N-glycosylation pathway mainly occurs in the ER and the Golgi apparatus (Rayon, Lerouge et al. 1998, Mathieu-Rivet, Kiefer-Meyer et al. 2014). This process is catalyzed and modified by a large number of important and highly conserved membrane-bound glycosylhydrolases and glycosyltransferases such as GnT1, XylT and FucT. The N-glycosylation pathway is mainly-studied in yeast and higher eukaryotes, but the data regarding the glycosyltransferases localization in this pathway remain unknown in diatoms. Three proteins homologous to AtGnT1, AtXylT and AtFucT were fused to eGFP and expressed *in vivo P. tricornutum*. It was shown that the sequence of PtXylT shares 24% identity with the AtXylT (Baïet, Burel et al. 2011). As expected these three eGFP fusion proteins showed similar dot-like or long strip-like fluorescence pattern (Fig. 3-16). This could be explained by the Golgi localization of the three candidate proteins via the known localization in plant. In *A. thaliana* GnT1 is known to be localized in the cis-Golgi apparatus (Kajiura, Okamoto et al. 2012). AtXylT was targeted to the medial cisternae of Golgi (Pagny, Bouissonnie et al. 2003, Kajiura, Okamoto et al. 2012). FucT

is localized to the trans-Golgi or TGN in *A. thaliana* (Fitchette-Lainé, Gomord et al. 1994). The N-glycosylation in mammalian is different. The GnT1 is substituted by an alfa-1, 3-fucose. The XylT and FucT transferases locate on the trans-Golgi apparatus (Kim, Jeon et al. 2014).

In some cases banana-like fluorescence pattern of XylT-eGFP was observed, as shown in Fig. 3-16b. The fluorescence pattern matches with the shape of the Golgi apparatus. Based on the co-expressions of XylT-mRFP with GnT1-eGFP and XylT-mRFP with FucT-eGFP, it was observed that the fluorescence overlapped only partially. Taken together, it was indicated that the localizations of these three Golgi marker proteins are slightly different. Previous study indicated that AtXylT acts at multiple stages of the plant N-glycosylation pathway, especially in the medial-Golgi (Kajiura, Okamoto et al. 2012). This could be a reason to explain why the XylT could partially overlap with the cis-Golgi marker GnT1 and the trans-Golgi marker FucT in *P. tricornutum*. Therefore, it was speculated that GnT1 might be localized on the cis Golgi apparatus, XylT is majorly targeted to the medial Golgi apparatus, while FucT is more possible enriched in the trans Golgi apparatus (Fig. 4-1). However, it is still unknown whether these three potential Golgi markers could also work in the course of N-glycosylation in *P. tricornutum* or not. In any case, the identification of the three Golgi marker proteins is important for the studying on their function and the pathway of N-glycosylation in *P. tricornutm*. N-glycosylation is an important and ubiquitous modification in the synthesis of new proteins in eukaryotes. This modification occurs in the course of protein secretion. It is crucial for the right folding, structural formation and assembly of the secreted proteins and the precisely targeting of glycoproteins to outside the cell or the membranes. Thus, the mutation of the glycosylation sites will be a new insight to investigate the protein transport in diatoms. As β1, 2-xylosylated N-glycans might have a function on inducing immune-reponses as pollen allergens which indicates that β1, 2-xylosylated N-glycans are equivalently interest in the algal-produced biopharmaceuticals.

### 4.2.3 Identification of retromer complex proteins

In addition to the already described coat proteins the retromer complex is mainly responsible for the retrograde transport of protein sorting receptors (Seaman 2005, Bonifacino and Hurley 2008, Schellmann and Pimpl 2009, Reyes, Buono et al. 2011). The retromer complex contains two subcomplexes, a large subcomplex is formed by three core subunits (Vps26, Vps29 and Vps35) for cargo recognition and a small subcomplex is formed by membrane deforming sorting nexin proteins (SNXs) in yeast, plants and mammal cells (Bonifacino and Hurley 2008, Otegui and Spitzer 2008, Cullen and Korswagen 2012). The structure of retromer has been well-studied in yeast, plants and mammal cells, but the localization of the core subunits and sorting nexins of retromer is still debated.

Within this thesis the subcellular localization of two homologous proteins to large retromer subunits of *A. thaliana* Vps26 and Vps29 were investigated *in vivo* in *P. tricornutum*. As expected the fluorescence patterns of the Vps26- and Vps29-eGFP fusion protein are similar, a dot or small diamond-like structure (Fig. 3-17). The co-localization of the Vps26-eGFP and Vps29-mRFP indicated that these two homologous proteins located in the same structure and might also belong to the retromer complex in *P. tricornutum*. In the present investigation, where the co-localization of the Vps29-mRFP with the TGN marker FucT but only partially overlap of the Vps29-mRFP with the medial Golgi marker XylT was observed. In agreement with recent studies in *Arabidopsis* and tobacco roots (Niemes, Langhans et al. 2010, Stierhof, Viotti et al. 2013) Vps26 and Vps29 subunits are localized to the TGN in *P. tricornutum* (Fig. 4-1). This provides an important prerequisite for the investigation of the function of retromer complex and the relationship between the retromer and three Golgi maker proteins. The receptor-dependent protein transport between different intracellular compartments is essential for many physiological activity in plant. The continuous recycling of receptors via retromer evade the degradation and can be used for the next rounds of protein transport. Previous studies have already idendified that trans Golgi network is the starting point for receptor-ligand interaction and package the receptor-ligand complexes into the clathrin-coated vesicles, subsequently, the cargo will be delivered into the next comparments such as the prevacuolar compartments/ late endosomal compartments (Tse, Mo et al. 2004, Detter, Hong-Hermesdorf et al. 2006, Robinson, Jiang et al. 2008). To some extent, it is contradicts to the current localization of the retromer.

The localization of retromer in higher plant cells is debated. The localization of retromer subunits seems not to be restricted to the TGN. Previous studies showed that the retromer in mammals localized to the tubular extensions of early and recycling endosomes (Carlton, Bujny et al. 2004, Bonifacino and Hurley 2008). Published data showed that the sorting nexins and the components of the large retromer subunit proteins were localized to the pre-vacuolar compartment (PVC) in *A. thaliana* and tobacco roots (Geldner, Anders et al. 2003, Tse, Mo et al. 2004, Oliviusson, Heinzerling et al. 2006, Jaillais, Santambrogio et al. 2007, Kleine-Vehn, Leitner et al. 2008, Yamazaki, Shimada et al. 2008). It was also suggested that the SNX2b and SNX1 are localized to the TGN, PVC and an endosomal compartment in plant (Phan, Kim et al. 2008, Robinson, Jiang et al. 2008). However, the colocalization of SNX2a with TGN markers but not with Golgi or PVC marker indicated that the sorting nexins are exclusively localized to the TGN rather than the pre-vacuolar compartment in Arabidopsis and tobacco roots (Niemes, Langhans et al. 2010). Previous studies showed that the retromer complex consists of sorting nexins subcomplex and the large Vps26/29/35 core subunits, but later it was suggested that the two subcomplexes may not always bind together and might have different functions in *A. thaliana* (Harbour, Breusegem et al. 2010, Pourcher, Santambrogio et al. 2010). This makes the localization of retromer subunits even more difficult.

### 4.2.4 Identification of vacuolar type $H^+$-ATPases

In order to identify more vacuolar marker protein, the vacuolar type $H^+$-ATPases (VHAs) were analyzed *in silico*. This membrane-bound multisubunit complex contain at least 26 genes encoding subunits (Sze, Schumacher et al. 2002). VHAs are essential for mediating the pH of intracellular compartments. Simultaneously, VHAs are important for protein transport, plant growth, development and adaptation to the changing environmental conditions and maintaining metabolite and ion balance in plant cells (Sze, Schumacher et al. 2002, Kluge, Lahr et al. 2003). Inhibition of the vacuolar type $H^+$-ATPase effect the secretion and results in the mistargeting of vacuolar proteins in plant (Matsuoka, Higuchi et al. 1997).

The fluorescence pattern of PtATPase1-eGFP fusion protein showed one or in some clones three dot-like structures in *P. tricornutum* (Fig. 3-19a/b). Based on the hypothesis VHA-c is targeted to the ER, TGN and the vacuole-specific subsector in *A. thaliana* (Seidel, Schnitzer et al. 2008). PtATPase1 homologous to hydrophobic subunit C of *A. thaliana* is a membrane protein. Thus, it was speculated that these dot-like fluorescence might be localized on the small vesicles or small vacuoles.

Another protein homologous to VHA-G2 subunit of *A. thaliana* is PtATPase2. PtATPase2-eGFP fusion protein showed a cytosolic fluorescence (Fig. 3-19c). The best match protein in *A. thaliana* is vacuolar ATPase protein 10 (TAIR: AT3g01390.2). It belongs to the subunit G of vacuolar type ATPase cytosolic V1 sector. The identified cytosolic PtATPase2 is in agreement with the published data from *A. thaliana* (Aviezer-Hagai, Nelson et al. 2000, Endler, Meyer et al. 2006). As already shown in the introduction the vacuolar type $H^+$-ATPase contains at least 26 gene encoding subunits, therefore, the localization of the two subunits is only the starting point in *P. tricornutum*. The identification of the subcellular localizations of these VHA subunits might be very useful for further studies on the localization of the other VHA subunits and for researches on the function of VHAs in the diatom *P. tricornutum*.

# 5 Materials and Methods

## 5.1 Materials

### 5.1.1 Instruments

PCR-Thermo-Cycler:

| | |
|---|---|
| Mastercycler gradient | Eppendorf, Hamburg |
| Mastercycler personal | Eppendorf, Hamburg |

Centrifuges:

| | |
|---|---|
| Centrifuge 5415 D | Eppendorf, Hamburg |
| Centrifuge 5417 R | Eppendorf, Hamburg |
| Centrifuge 5810 R | Eppendorf, Hamburg |
| Mikro 22 R | Hettich Zentrifugen, Tuttlingen |
| MiniSpin® Plus | Eppendorf, Hamburg |
| PicoFuge® | Stratagene, La Jolla, USA |
| L755 Ultracentrifuge | Beckman Coulter |

Biolistic transfection:

| | |
|---|---|
| FrenchPress MiniZelle FA-003 | G. Heinemann ULT |
| Biolistic PDS-1000/He Particle Delivery System | Biorad, Munich |
| Rupture Discs 1350 psi | Biorad, Munich |
| Macrocarrier | Biorad, Munich |
| M 10 (Ø 0.7 µm) Tungsten-Particles | Biorad, Munich |
| Frenchpress SLM-AMINCO 4-3399 | SLM-AMINCO Instruments |

Confocal Laser Scanning Microscope:

| | |
|---|---|
| CLSM Leica TCS SP2 | Leica, Wetzlar |

Incubation:

| | |
|---|---|
| Incubator | Heraeus Instruments, Hanau |
| Climate Chamber MLR-350 SANYO | Ewald Gmbh |
| Thermocycler 60 | Biomed, Oberschleißheim |
| Thermomixer comfort | Eppendorf, Hamburg |
| Thermomixer compact | Eppendorf, Hamburg |

Other instruments:

| | |
|---|---|
| Nanodrop ND-1000 photometer | peqlab |
| ABI Prism 377 | Applied Biosystems |

### 5.1.2 Membranes and filters

| | |
|---|---|
| Nitrocellulose membrane | Macherey-Nagel |
| Whatman 3MM | Schleicher & Schuell, Dassel |
| FP 30/ 0.2 CA-S - 0.2 µm sterile filter | Schleicher& Schuell |
| Fuji-Medical-X-ray-Film, 30 x 40 cm | Fuji Film |
| X-ray film developer and replenisher | Kodak |

### 5.1.3 Antibodies

| Primary antibodies | Dilution | Manufacturer |
|---|---|---|
| α GFP | 1:3000 | Biomol |
| α Rubisco | 1:5000 | Agrisera |
| α psbD(D2 protein of PSII) | 1:5000 | Agrisera |
| α psbO | 1:1000 | Agrisera |

| Secondary antibodies | Dilution | Manufacturer |
|---|---|---|
| αgoat (HRP coupled) | 1:10000 | Sigma-Aldrich, Munich |
| αrabbit (HRP coupled) | 1:10000 | Sigma-Aldrich, Munich |

### 5.1.4 Chemicals

Unless otherwise noted, all chemicals used in this work were obtained from Roth GmbH, Sigma Aldrich or Merck and stored and used according to the manufacturer's instructions.

### 5.1.5 Enzymes

| | |
|---|---|
| DNAseI | Thermo Scientific/Fermentas, St. Leon-Rot |
| Phusion High Fidelity DNA-Polymerase (5 U/µl) | Thermo Scientific/Fermentas, St. Leon-Rot |
| Restriction endonucleases (10 U/µl) | Invitrogen, Karlsruhe |
| RNase A (70 U/µl) | Thermo Scientific/Fermentas, St. Leon-Rot |
| T4-DNA-Ligase (1 U/µl) | Thermo Scientific/Fermentas, St. Leon-Rot |
| Taq-DNA-Polymerase | Biotools, Madrid |

### 5.1.6 Software and bioinformatic applications

| | |
|---|---|
| Sequencher 5.1 | GeneCodes |
| LCS Lite 2.5 | Leica |
| ClustalW and ClustalX | Alignment |
| Mega4.0 and Mega6.0 | Phylogenetic analyses |
| ImageJ | Tony Collins, McMaster Biophotonics Facility |

The following websites were commonly used:

| | |
|---|---|
| SOSUI | http://harrier.nagahama-i-bio.ac.jp/sosui/sosui_submit.html |
| TMHMM Server v.2.0 | http://www.cbs.dtu.dk/services/TMHMM-2.0/ |
| ΔG prediction server v1.0 | http://dgpred.cbr.su.se/index.php?p=fullscan |
| TOPCONS | http://topcons.cbr.su.se/ |
| SignalP 4.1 Server | http://www.cbs.dtu.dk/services/SignalP/ |
| TargetP 1.1 Server | http://www.cbs.dtu.dk/services/TargetP/ |
| ChloroP Server | http://www.cbs.dtu.dk/services/ChloroP/ |
| Psortb | http://www.psort.org/psortb/ |
| Phaeodactylum digital gene expression | http://www.diatomics.biologie.ens.fr/EST/est.htm |
| P. tricornutum data base v2.0 | http://genome.jgi-psf.org/pages/blast.jsf?db=Phatr2 |
| Tm calculator | http://www.thermoscientificbio.com/webtools/tmc/ |
| Pubmed | http://www.ncbi.nlm.nih.gov/entrez |
| ClustlW2 | http://www.ebi.ac.uk/Tools/clustalw2/index.html |
| Clustal Omega | http://www.ebi.ac.uk/Tools/msa/clustalo/ |
| TransportDB | http://www.membranetransport.org/ |
| Arabidopsis thaliana | http://www.arabidopsis.org |

### 5.1.7 DNA and protein markers

For agarose gel electrophoresis, the GeneRuler™ 1 kb Plus DNA Ladder or Lambda DNA/EcoRI+HindIII Marker from Thermo Scientific/MBI Fermentas were used as DNA length marker. As for SDS-PAGE marker, the PageRuler™ Prestained Protein Ladder from Thermo Scientific/Fermentas was used.

### 5.1.8 Oligonucleotide primers

All oligonucleotide primers used in this work were supplied from Sigma. All oligonucleotide primers used for amplifying genes, colony PCR and sequencing in this work were listed in supplemental data 7.3. The working concentration of primers was 5 pmol/µl.

### 5.1.9 Vectors

| | | |
|---|---|---|
| pJet1.2/blunt cloning vector | Ampr,PT7, eco47IR | MBI-Fermentas |
| pPha-T1 (provided from Peter Kroth, University Konstanz, Germany) | Ampr, Zeor,PfcpA +TfcpA | Zaslavskaia et al.2000 |
| pPha-NR | Ampr, Zeor, PNR+ TNR | Acc. JN180663 |
| pMOD™-3 <R6Kγori/MCS> | Ampr, R6Kγori | Cat.No.MOD1503 |
| pPha-Dual 2xNR | Ampr, Zeor, PNR+TNR(both MCS I and II) | Acc. JN180664 |

### 5.1.10 organisms

***Escherichia coli* TOP10** *(F-, mcrA, Δ(mrrhsdRMS-mcrBC), φ80lacZΔM15, ΔlacX74, nupG, recA1, araD139, Δ(ara-leu)7697, galE15, galK16, rpsL(Str$^R$), endA1, λ)* from Invitrogen.

***Phaeodactylum tricornutum***, Strain CCAP 1055/1

***Amphidinium carterae*** strain CCAM0512

**TransforMax$^{TM}$EC100D$^{TM}$ Pir$^{-}$116 Electroncompetent *E.coli***

F– *mcr*A Δ(*mrr*-hsdRMS-*mcr*BC) φ80d*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ– *rps*L *nup*G *pir*-116(DHFR). Maintains plasmids at ~250 copies per cell.

**TransforMax EC100D *pir*+**

F– *mcr*A Δ(*mrr*-hsdRMS-*mcr*BC) φ80d*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ– *rps*L *nup*G *pir*+(DHFR). Maintains plasmids at ~15 copies per cell.

## 5.2 Methods

### 5.2.1 Culture of *E.coli* TOP10

*E.coli* TOP10 cells were obtained from Invitrogen, Karlsruhe. For long-termstorage, *E.coli* TOP10 cells were stored at -80$^0$C in a LB/glycerin (1:1) mixture. Short-termcultivation was carried out overnight at 37°C and constant shaking (200 rpm) insterile LB liquid medium or on 1.5% LB-agar plates with the appropriate antibiotic (50 µg/ml ampicillin or 25 µg/ml kanamycin).

LB-medium: (pH=7.0)

| | |
|---|---|
| 1% (w/v) | Bacto-Tryptone |
| 0.5% (w/v) | Yeast extract |
| 1% (w/v) | NaCl |
| 50 µg/ml | ampicillin or 25ug/ml kanamycin |

Add 1.5% (w/v) agar for LB solid medium.

### 5.2.2 Culture of *Phaeodactylum tricornutum* CCAP 1055/1

*P. tricornutum* cells were cultured under 24 h light condition (8000-11000 1x) at 22°C, constantly shaken at 225 rpm in f/2- liquid medium or on f/2 agar plates. The cultures were inoculated each two to three weeks for strain maintenance. For the selection of transformed clones, a concentration of 75 µg/mL zeozin™ (InvivoGen) was added to the medium. For strain maintenance cultures were transferred to new f/2 agar plates every four weeks.

| f/2 medium (pH 7.0) | |
|---|---|
| Tropic marine sea salt | 16.6 g |
| Tris (2 M, pH 8.0) | 1 mL |
| $NaH_2PO_4 \cdot H_2O$ (0.1 M) | 360 µL |
| $NaNO_3$ (1 M)/ $NH_4Cl$ | 890 µL/ 1 mL |
| f/2 vitamin solution | 1 mL |
| Trace elements | 1 mL |
| | ad 1 L dH2O |

| f/2 trace elements | | f/2 vitamin solution | |
|---|---|---|---|
| $FeCl_3$ | 11.65 mM | Biotin | 2 µM |
| $Na_2EDTA$ | 11.71 mM | Cyanocobalamine | 0.37 µM |
| $CuSO_4$ | 39 µM | Thiamine-HCl | 297 µM |
| $ZnSO_4$ | 77 µM | | |
| $CoCl_2$ | 42 µM | | |
| $MnCl_2$ | 910 µM | | |
| $Na_2MoO_4$ | 26 µM | | |

### 5.2.3 Culture of *Amphidinium carterae* CCAM0512

*Amphidinium carterae* (Hulburt) (strain CCAM0512) comes from the algae collection department of Marburg University and were cultured in f/2-medium at 12 h/ 12 h light/ dark photoperiod, 20 °C. For putative minicircles isolation, the cells were cultured to the middle of stationary phase.

| f/2 medium (pH 7.0) | |
|---|---|
| Tropic marine sea salt | 30 g |
| Tris (1 M, pH 8.0) | 5 ml |
| $NaH_2PO_4 \cdot H_2O$ (0,1 M) | 360 µl |
| $NaNO_3$ (1 M) | 900 µl |
| f/2 vitamin solution | 1 mL |
| Trace elements | 1 mL |
| ad 1 L $dH_2O$ | |

### 5.2.4 Nucleic acid analytics

#### 5.2.4.1 Plasmid isolation from *E.coli*

Plasmid isolation from *E.coli* was carried out by alkaline extraction according to (Bimboim and Doly 1979). Previous to plasmid preparation, liquid cultures were inoculated from colonies grown on LB agar plates and incubated overnight at 37°C. The next day, 1.5ml of the liquid culture was used and centrifuged full speed for 1min. After discarding the supernatant by tilt or tap, the pellet was resuspended in 200 µl of P1 buffer. 200 µl of buffer P2 were added and mixed with the suspension by inverting the Eppendorf cup 5-10 times. Following an incubation for 5 min at RT, 200 µl of P3 and 20 µl chloroform were added and mixed with the lysate. After an incubation of 5 min on ice the lysate was centrifuged for 5 min (4°C/ 20,000 g). The resulting supernatant was then transferred to a new Eppendorf cup and 400 µl of isopropyl alcohol was added and mixed with the supernatant. Following a centrifugation (20 min/ 4°C/ 20,000 g) the supernatant was discarded and the pellet washed with 70 % ethanol (5 min/ 4°C/ 20,000 g). The supernatant was again removed and the dried pellet was finally dissolved in 40 µl dd$H_2O$.

| Buffer P1 | | Buffer P2 | |
|---|---|---|---|
| Tris/HCl (pH 8.0) | 50 mM | NaOH | 200 mM |
| EDTA | 10 mM | SDS | 1 % (w/v) |
| RNase A | 100 µg/mL | | |

| Buffer P3 | |
|---|---|
| KAc (pH 5.5) | 3 M |

#### 5.2.4.2 DNA and RNA isolation from *P. tricornutum*

The isolation of genomic DNA from *P. tricornutum* was carried out via the CTAB-method described (Doyle and Doyle 1990). *P. tricornutum* cells were collected from 150 mL liquid culture by centrifugation (3000 g/ 5 min/ RT). The supernatant was discarded and pellet was resuspended in 1600 µL of 2x CTAB containing buffer B. Cell lysis, then the suspensions was transferred into two new Eppendorf cups (EP) and incubated at 70 °C for 30 min, after that cell debris was centrifuged (20000 g/ 5 min/ RT). The upper phase was moved to a new EP and mixed with one volume of PCI (DNA isolation, saturated with EDTA) gently, then centrifuged at 20000 g/ 10 min/ RT. Subsequently, the upper phase was moved to a new EP again, mixed with 1/10 volume of NaAc (pH 4.5) and 2/3 volume of isopropyl alcohol gently and centrifuged (20000 g/ 20 min/ 4 °C). After that, the supernatant was removed and the pellets was washed with 500 µL of 70 % ethanol and centrifuged (20000 g/ 20 min/ 4 °C). At last, the supernatant was discarded, while the pellets were dried in a

desiccator. After drying, the pellet was dissolved in 15 µL of dd$H_2O$, after which the DNA solutions were transferred into one EP.

RNA isolation was carried out by collecting 150 ml liquid culture (centrifuge at 10000 g/ 30 s/ 4 °C) from exponential *P. tricornutum* cells. After discarding the supernatant, the pellet was resuspended in six ml NAES in six Eppendorf cups, each cup including 1 ml. At the same time, 100 beads and 1ml water-saturated PCI was added into each cup. After vortex them 3x20 s, centrifuged for 10 min at 20000 g RT. Upper aqueous phase was moved into a new EP, 1 volume of water saturated PCI was added into the aqueous phase, mixed gently and centrifuged 15 min, 20000 g RT again. The upper aqueous phase was transferred to a new EP again and added 1 volume of chloroform, centrifuged 15 min, 20000 g RT. Subsequently, the aqueous supernatant was transferred to a new EP and added 0.7 volume of isopropanol to precipitate RNA, incubated the sample overnight at -20°C. At the next day, centrifuged it full speed for 30 min and used 500 µL 70% EtOH to wash RNA. At last, the pellets were dried in a desiccator and dissolved it in 15 µl of DEPC $H_2O$.

| Buffer B | | NAES | |
|---|---|---|---|
| Tris/HCl (pH 8.0) | 0.1 M | NaAc (pH 5,1) | 50 mM |
| $Na_2EDTA$ | 0.02 M | EDTA | 10 mM |
| NaCl | 1.4 M | SDS | 1 % (w/v) |
| 2-Mercaptoethanol | 0.2 % (v/v) | | |

**5.2.4.3 cDNA synthesis via reverse transcription (RT)**

Before cDNA synthesis, RNA isolation sample including potential genomic DNA had to be treated by DNaseI in order to remove genomic DNA. Each 500 ng of RNA, 1 µl DNAseI (1 U/µl) and 2 µl 10x reaction buffer were mixed in 20 µl total, and was incubated at RT for 20 min. After DNA removal, RNA sample was checked on the agarose gel (selection). The cDNA synthesis was followed directly with the SuperScript™ II RT according to the manufacturer's instructions. 1.3 µl random hexamer primer and 1.7 µl DEPC $H_2O$ were added to the samples, and incubated at 70°C for 5 min. Then the samples were kept on ice and 2 µl 10mM dNTPs and 4 µl reverse transcriptase buffer were added. Following an incubation for 5 min at 25°C 1 µl reverse transciptase was added. After incubation for 60 min at 42°C the reaction was eventually stopped by incubating the samples at 70°C for 10 min.

**5.2.4.4 Polymerase chain reaction (PCR)**

The Phusion High Fidelity PCR Kit was used to amplify the DNA sequences from *P. tricornutum* gDNA or cDNA. The concentration of the used primers (see supplementary material 7.3) was 5 pmol/µl and

the annealing temperature depended on the melting temperature of the primers. The reaction mix consisted of the following components:

| | |
|---|---|
| Phusion-HF-Buffer (4mM $MgCl_2$) | 5 µl |
| $MgCl_2$ (50 mM) | 2 µl |
| dNTPs (10 mM) | 1 µl |
| Forward-Primer (5 pmol/µL) | 1 µl |
| Reverse-Primer (5 pmol/µL) | 1 µl |
| gDNA/cDNA solution (250 ng) | 0.2 µl/2µl |
| Phusion High Fidelity DNA-Polymerase (5 U/µL) | 0.25 µl |
| | ad 25 µl $ddH_2O$ |

| Step | Duration | Temperature |
|---|---|---|
| 1.Denaturation | 1 min | 98°C |
| 2.Denaturation | 20 s | 98°C |
| 3.Annealing | 30 s | Tm-3°C |
| 4.Elongation | Depends on sequence length | 72°C |
| 5.Final Elongation | 4 min | 72°C |
| 6.Cool down | 20 s | 20°C |
| Step 2-4 were repeated 29 times (gDNA) or 40 times (cDNA). | | |

Colony PCR was performed with Taq DNA Polymerase (Biotools) according to the manufacturer's instructions. A *P. tricornutum* colony, grown on selective medium after transfection, was resuspended in 20 µl of $ddH_2O$ and boiled for 10 min at 95°C. After centrifugation (2000 g/ 10 min/ RT), the supernatant was used for colony PCR reaction. The reaction mix consisted of the following components:

| | |
|---|---|
| 10xstandard Buffer with $MgCl_2$ (Biotools) | 2.5 µl |
| $MgCl_2$ (50 mM) | 1.25 µl |
| dNTPs (10 mM) | 0.5 µl |
| For-Primer (5 pmol/µL) | 0.5 µl |
| Rev-Primer (5 pmol/µL) | 0.5 µl |
| Boiled colony suspension supernatant | 5 µl |
| Taq- DNA-Polymerase (Biotools) | 0.25 µl |
| | ad 25 ul $ddH_2O$ |

#### 5.2.4.5 Agarose gel electrophoresis

For separation of DNA or RNA on an agarose gel, 10 µl sample and 2 µl loading buffer mix was loaded onto a 1% to 2% agarose gel which was based on 1xTBE buffer. Roti®-GelStain (Roth) was

used to stain nucleic acid. For the elution of DNA fragments, the JetSorb DNA-Ectraction-Kit or PCR purify kit was used according to manufacturer's instructions.

| DNA loading buffer (pH 7.0) | | 10xTBE | |
|---|---|---|---|
| Urea | 4 M | Tris/HCl (pH 8.8) | 1 M |
| EDTA | 0.05 M | Boric acid | 0.83 M |
| Sucrose | 50% (w/v) | EDTA | 0.01 M |
| Xylene cyanol | 0.1 (w/v) | | |
| Bromphenol blue | 0.1 (w/v) | | |

**5.2.4.6 Sequencing**

The sequencing of DNA was performed by the method described in Sanger, Nicklen et al. (Sanger, Nicklen et al. 1977). The sequence reaction mixture was consisted of the following components:

| | |
|---|---|
| Sequencing primer | 1 µl |
| ABI mix | 2 µl |
| Plasmid solution | 4 µl |
| $ddH_2O$ | 3 µl |

The following PCR reaction:

| Step | Duration | Temperature |
|---|---|---|
| 1.Denaturation | 3 min | 95°C |
| 2.Denaturation | 30 s | 95°C |
| 3.Annealing | 30 s | 50-55°C |
| 4.Elongation | 90 s | 60°C |
| 5.Final Elongation | 4 min | 60°C |
| 6.Cool down | 20 s | 20°C |
| Step 2-4 were repeated 30 times. | | |

After the PCR program, the reaction mix was moved to a new 1.5 ml EP and mixed with 64 µl 100% ethanol and 26 µl $ddH_2O$. The mix was saved in the dark for 30 min. After that, the mix was centrifuged for 30 min at 2000 g, 4°C. The pellet was washed by 70% EtOH once, then was dried and dissolved in 3 µl for amide loading dye. The ABI PRISM 377 DNA Sequencher was used to separate the sequencing reaction. The nucleotides sequences were analyzed on Sequencher 5.1.)

**5.2.4.7 Restriction and ligation**

To prepare DNA fragment for ligation, normally, the mix is made up of the following components:

| | |
|---|---|
| vector | 0.3 µl |
| DNA | 1 µl |
| restriction enzyme 1 | 0.25 µl |

| | |
|---|---|
| restriction enzyme 2 | 0.25 µl |
| 10x SDB buffer | 1 µl |
| | ad 10 µl dd$H_2O$ |

The restriction mix was incubated 30 min to 1h at 37°C. Then each sample was added 2ul loading dye and loaded on agarose gel electrophoresis.

The eluted PCR products were ligated into pJet1.2/blunt cloning vector with the ClonJetTM PCT Cloning Kit from Thermo Scientific/Fermentas. The ligation of DNA fragments into the expression vector pPha-NR, the ratio of insert and vector was 3:1 and added 0.5-1 µl T4 ligase from Thermo Scientific, and 1 µl 10xligase buffer in 10 µl volume. The ligation mix was incubated at RT for 30 min.

#### 5.2.4.8 Transformation of *E.coli*

Before the transformation of *E.coli*, chemically RbCl-competent cells were prepared and stored at -80°C. For the transformation of *E.coli* TOP10, 50 µl of RbCl-competent *E. coli* suspension were added to the ligation mix and incubated on ice for 20-30 min. Then, the cells were heat shocked at 42°C for 45 s. After a short recovering time on ice the cells were plated on 1.5 % agar-LB plates containing 50 µg/ml ampicillin. The plate was incubated at 37°C overnight.

*E. coli* TOP10 cells were cultured overnight in 100 ml LB medium with streptomycin (50 µg/ml), a 1 L culture with antibiotic was inoculated at 37°C, the starting OD600 is 0.1, at the same time, Sterile $MgCl_2$ and $MgSO_4$ were added into the culture, a final concentration of them is 10 mM each. When the cells concentration increased to OD600 of about 0.6, then centrifuged and used 33 ml RF1 to resuspend it, incubated 30min on ice. After the second centrifugation, the pellet was resuspended in 50 ml RF2 and incubated 30 min on ice again. At last, aliquots of 100 µl were prepared in 4°C room and frozen in liquid nitrogen and stored them at -80°C.

| RF1 | | RF2 | |
|---|---|---|---|
| $RbCl_2$ | 100 mM | MOPS | 10 mM |
| $MnCl_2$x $4H_2O$ | 50 mM | $RbCl_2$ | 10 mM |
| KAc | 30 mM | $CaCl_2$x$2H_2O$ | 75 mM |
| $CaCl_2$x$2H_2O$ | 10 mM | Glycerin | 15% (w/v) |
| Glycerin | 15% (w/v) | Adjust pH to 5.8 with NaOH | |
| Adjust pH to 5.8 with acetic acid | | | |

#### 5.2.4.10 Transfection of *P. tricornutum*

Before transfection, the wild type *P. tricornutum* cells were cultured on f/2 agar plate (without zeozin) for overnight. The concentration of cells on each plate is $10^8$ cells in 100 µl f/2 liquid medium. Normally, three plates' wide type cells were prepared for one construct. The next day, the biolistic

transfection was carried out as described by Apt, Grossman et al. and Zaslavskaia, Lippmeier et al. (Apt, Grossman et al. 1996, Zaslavskaia, Lippmeier et al. 2000). At the beginning, 50 µl M10 for each construct, 5 µg plasmid (5000/plasmid concentration), 50 µl 2.5 M CaCl2 and 20 µl 0.1 M SP enzyme was mixed using vortex for 1min. Then incubated 10 min at room temperature. After centrifugation (full speed/ 5 min), the supernatant was threw away and then 250 µl 100% ETOH HPLC-Quality was used to wash the pellet once. At last, 50 µl 100% ETOH HPLC-Quality was used to re-suspend the pellet for transfection.

The transfection was carried out by using Biolistic PDS-1000/He Particle Delivery System. Before transfection, a cleaning work was conducted with 100% ethanol (HPLC grade). After the cleaning, the transfection was started. Firstly, 15 µl of the DNA bound microcarrier suspension was added on a microcarrier membrane, then the components of the particle gun was assembled according to manufacturer's instructions. Secondly, when the vacuum increased to 25 psi, turn on the pressure until to 1350 psi, then release immediately. Lastly, the transfected plates were sealed by Parafilm and cultured at 22°C under the continuous light overnight. The next day, 1 ml f/2 medium was used to wash each transfected plate to three f/2 ammonium agar plates including selective zeocin. The plates were cultured under the continuous light at 22°C until colonies were visible.

#### 5.2.4.11 Minicircles enrichment and isolation from *A. carterae* CCAM0512

*A. carterae* CCAM0512 was harvested from four to five weeks cultures by centrifugation (3000 g, 3 min). Alkaline extraction according to Bimboim and Doly for bacterial plasmid preparation was used to enrich minicircles (Bimboim and Doly 1979). Additionally, different alkaline lysis based kits were also used to enrich the minicircles. The following kits were used:

| | Company |
|---|---|
| PeqGOLD Plasmid Miniprep KitI | Peq Lab |
| PureYieldTM Plasimid Miniprep System, 50 preps | Promega |
| UltraClean 6 minute Mini Plasmid Prep Kit | MoBio |
| High Yield Plasmid Mini Kit | SLG |

#### 5.2.4.12 Transposon-insertion based approach

A transposon-insertion based approach was used to isolate individual minicircles from *A. carterae* CCAM0512 in high copy number. The natural structure of plasmid-like minicircles provides a good basis for transposon insertion and subsequent usage of bacterial proliferation. Firstly, a kanamycin antibiotic resistance gene was inserted into the multi-cloning site of an EZ-Tn5 (Epicentre) transposon in EZ-Tn5 pMOD-3 vector (EZ-Tn5™ pMOD™ <R6K$\gamma$ori/MCS> Transposon Construction

Vectors, EPICENTRE Company). EZ-Tn5 (Epicentre) transposon contains two mosaic ends (5'-AGATGTGTATAAGAGACAG-3'), a multiple cloning site and R6K$\gamma$*ori* of replication. R6K$\gamma$*ori*-dependent replication needs the *pir* gene product produced by TransforMax EC100D *pir*$^{+}$ *E.coli* cells. Secondly, the EZ-Tn5 (epicentre) transposon including the kanamycin antibiotic resistance gene was amplified by transposon primers (see supplement 7.3). Subsequently, the transposon was inserted into the target minicircles in *vitro* by using the reaction mixture (37°C, 2 hours) which is made up of the following components:

| | |
|---|---|
| Minicircles | 6.5 µl/ 0.2 µg |
| Transposon | 2 µl |
| EZ-Tn5 10xReaction buffer | 1 µl |
| EZ-Tn5 Transposase 1 U/ µl | 0.5 µl |
| | ad 10 µl dd$H_2O$ |

The 10xreaction buffer and transposase were provided by the EZ-Tn5$^{TM}$ Custom Transposome Construction Kits. They were used according to the manufacturer's instructions.

After transposon insertion reaction, as the EZ-Tn5 10xreaction buffer contains $Mg^{2+}$, the reaction mixture was not stopped by EZ-Tn5 10xstop solution as manufactors' protocol. The reaction mixture was purified by the next steps. Firstly, Phenol: Chloroform: Isoamyl Alcohol (25:24:1, v/v) was used to purify the minicircles, then chloroform was used for the second step purification of transposon inserted minicircles, and subsequently 2.5 volume 100% Ethanol and 0.1 volume NaAc, pH 4.8 was used to the precipitation of transposon inserted minicircles. At last, electroporation procedures for bacterial transformation (TransforMax$^{TM}$ EC100D$^{TM}$ *pir*$^{+}$ electrocompetent *E.coli*, 2.5 kV, fast charge rate, 2 mm cuvette) was used. Each transformation mixture was cultured in 300 ml LB liquid medium without antibiotic for 1 hour, and subsequently plated the culture on the plate. The plasmids (potential individual minicircles) were isolated via standard plasmid preparation (Alkaline lysis). The transposon insertion is schematically visualized in figure 5-1.

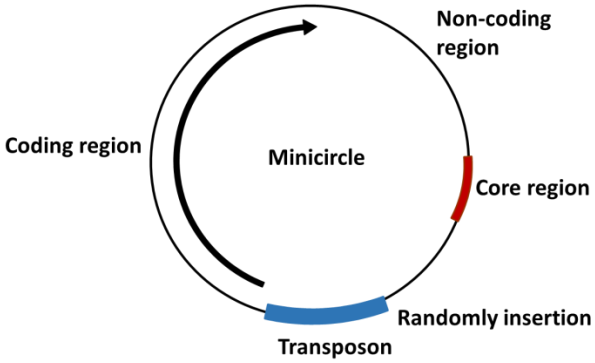


**Fig. 5-1: Schematic depiction of transposon-insertion based isolation of minicircles.**

It was shown that the minicircle provides a vector standard for the transposon insertion. The red region is the core region of minicircles. The arrow region is the coding region of minicircles. The remainder of minicircles is called the non-coding region. The transposon (marked by blue) contains an origin of replication and a kanamycin antibiotic resistence gene. The transposon was inserted into the minicircles randomly in the reaction mix, so it was also possible inserted into the core region, coding region and non-coding region.

### 5.2.4.13 DNA extraction, amplification, and sequencing of LSU rDNA domain D1-D6 and SSU rDNA for *A.carterae*

Genomic DNA was isolated from 100 ml of exponentially growing *A. carterae* cells according to the CTAB method described in Doyle and Doyle (Doyle and Doyle 1990). Extracted DNA was used as a template, the Phusion High Fidelity PCR Kit was used to amplify approximately 1400 bp of the LSU rDNA gene covering the variable domains D1-D6, using the primers (LSU-for: ACCCGCTGAATTTAAGCATA; LSU-rev: CCACCATGCCCTCCTACTCA); and approximately 1700 bp of the SSU rDNA gene, using the primers(SSU-for:GTCTCAAAGATTAAGCCATGCATGTC;SSU-rev:CTTCTCCTTCCTCTAAGTGATAAGGTTC). Primers concentration is 5 pmol/µl and the applied annealing temperature was calculated with the $T_m$ calculator. Sequencing was done by Macrogen Europe Company.

### 5.2.4.14 Sequence alignment and phylogenetic analyses

Sequences were aligned by clustalx2.0 and Mega 6.0 software as described in Tamura (Tamura, Stecher et al. 2013). As outgroup species, the dinoflagellates *Gonyaulax membranacea, Gymnodinium dorsalisulcum, Heterocapsa arctica, Alexandrium fundyense* were used. The data matrix of LSU comprised 1183 aligned positions totally, which has excluded the hypervariable D2 domain region; for SSU data matrix, 1687 aligned positions were included. At the alignment, the phylogenetic trees were constructed with maximum likelihood algorithm on the software of Mega6.0. The optimal parameters were as follows: bootstrap replications: 1000; substitution model:

general time reversible model; ML heuristic method: nearest-neighbor-interchange (NNI); Branch swap filter: very strong. All species included in the molecular analyses with their corresponding GenBank accession numbers are shown at the behind of the name on the phylogenetic trees.

### 5.2.5 Protein analytics

#### 5.2.5.1 Protein isolation from *P. tricornutum*

To isolate protein from *P. tricornutum*, two different methods were used. One is the alkaline lysis, which can screen the expression of protein from small amount of culture samples. The other one is using French press to extract protein for a subsequent fractionating.

Alkaline lysis: *P. tricornutum* cells were collected from 5-10 ml liquid medium by centrifugation at 1500 g/ 10 min/ 4°C. After discarding the supernatant, the pellet was resuspended in 1 ml ddH2O. 150 µl of lysis buffer was added to the pellet, vortexed and incubated on ice for 10 min for cell lysis.French press passage: *P. tricornutum* cells were harvested by centrifugation at 1500 g/ 10 min/ 4°C. From now on, every step was done on ice. After discarding the supernatant, the pellet was resuspended in 2985 µl SolA buffer and 15 µl PIC was added into the suspension in a ratio of 1:200 in order to inactivate proteases. Then the cells were disrupted through the pressure cell under a constant pressure of 1000 psi. After the French press, centrifugation (1500 g/ 10 min/ 4°C) was carried out to pellet the intact cells, while supernatant was used for the next experiments.

| Lysis buffer | | Protease inhibitor cocktail (PIC) | |
|---|---|---|---|
| NaOH | 1.85 M | Antipain | 200 µg/mL |
| 2-Mercaptoethanol* | 7.5 % (v/v) | Aprotinin | 200 µg/mL |
| *2-Mercaptoethanol was added directly before an alkaline lysis was carried out. | | Chymostatin | 200 µg/mL |
| | | Elastatinal | 200 µg/mL |
| | | Leupeptin | 200 µg/mL |
| | | Pepstatin | 200 µg/mL |
| | | Trypsin-Inhibitor | 200 µg/mL |
| | | $Na_2EDTA$ | 200 µg/mL |
| | | in 280 mM Hepes/KOH buffer (pH 7.9) | |

#### 5.2.5.2 Protein extraction fractionation via carbonate extraction

Carbonate extraction was used to separate different kind of proteins, soluble, membrane associated and integral membrane proteins from the whole protein extraction. The supernatant from the French press was used to do carbonate extraction. Firstly, supernatant was centrifuged (120,000 g/ 45 min/ 4°C) in order to pelletize the associated and integral membrane proteins, then the soluble proteins (2 ml) in supernatant was transferred into two new Eppendorf cups and stored on ice. To

separate associated membrane proteins, the pellet was resuspended in carbonate buffer containing PIC (1:200), incubated on ice for 30min and mixed gently by plastic head dropper. Subsequently, in supernatant the associated membrane proteins was separated from integral membrane proteins by centrifugation at 120000 g/ 45 min/ 4°C. Again, the associated membrane proteins in supernatant was transferred into two new Eppendorf cups and stored on ice. While the integral membrane proteins in pellet were resuspended in 2388 µl SolA buffer containing 12 µl PIC (1:200) and transferred into another two new Eppendorf cups for the precipitation of proteins in next step.

| Solubilization buffer A (SolA) (pH 7.5) | | Carbonate buffer (pH 11.5) | |
|---|---|---|---|
| Imidazole | 50 mM | $NaHCO_3$ | 100 mM |
| NaCl | 50 mM | EDTA | 1 mM |
| 6-aminohexanoic acid | 2 mM | | |
| EDTA | 1 mM | | |
| Sucrose | 8.5 % | | |

#### 5.2.5.3 TCA protein precipitation

To precipitate proteins the TCA precipitation method was used. TCA was added to the samples to a final concentration of 12.5%. After an incubation for 30 min on ice the samples were centrifuged (20,000 x g/ 10 min/ 4°C). To remove residual TCA, the pellet was washed at least three times with 80% (v/v) acetone. After the last washing step the pellet was dried and dissolved in an appropriate volume of urea loading buffer by incubated at 55°C for 10 min, then centrifuged for 30 s full speed and save them in -20°C for next step analysis.

| Urea loading buffer (pH 6.8) | |
|---|---|
| Urea | 8 M |
| Tris/Hcl | 200 mM |
| EDTA | 0.1 mM |
| SDS | 5% (w/v) |
| Bromphenol blue | 0.03% (w/v) |
| 2-Mercaptoethanol* | 1% (v/v) |
| *2-Mercaptoethanol was added to samples before using directly | |

#### 5.2.5.4 Determination of protein concentration via Amido black

To determinate the concentration of proteins, Amido black method was used as described in Popov (Popov, Schmitt et al. 1974). 3 µl protein samples were filled with 97 µl $ddH_2O$ to an end 100 µl. Then the proteins were added 400 µl amido black staining solution. After mixing, the samples were centrifuged at 20000 g/ 15 min/ 4°C. The protein pellet was washed by 500 µl washing solution and

centrifuged again. At last, the pellet was dried in dessicator and dissolved in 1 ml 200 mM NaOH. After transferring sample to a cuvette, its absorption was measured at 615 nm wavelength by a photometer. Before measuring, a blanking measurement was carried out by using 1 ml 200 mM NaOH. Based on a standard curve, the calculation of the protein concentration was performed.

| Amido black staining solution | | Amido black washing solution | |
|---|---|---|---|
| Acetic acid | 10 % (v/v) | Methanol | 90 % (v/v) |
| Methanol | 90 % (v/v) | Acetic acid | 10 % (v/v) |
| a pinch of amido black | | | |

**5.2.5.5 SDS-polyacrylamide gel electrophoresis (PAGE)**

SDS-PAGE was used to separate proteins based on a discontinuous buffer system described by Laemmli (Laemmli 1970).

Before loading onto the gel, the protein samples with urea loading dye were boiled up at 55°C for 10 min, the protein marker was PageRulerTM Prestained Protein Ladder (MBI Fermentas). The gel was run at 140 V and 20 mA current in the stacking gel, 30 mA in the resolving gel,

| 4xStacking gel buffer | | 4xResolving gel buffer | |
|---|---|---|---|
| Tris/HCl (pH 6.8) | 500 mM | Tris/HCl (pH 8.8) | 1.5 M |
| SDS | 0.4% (w/v) | SDS | 0.4% (w/v) |
| 10XSDS running buffer | | | |
| Tris | 250 mM | | |
| Glycine | 2 M | | |
| SDS | 1% (w/v) | | |

| The SDS-PAGE gels consisted of the following components: | | |
|---|---|---|
| | Resolving gel (12.5 %)* | Resolving gel (12.5 %)* |
| Acrylamide 30 % (v/v) | 4.1 mL | 0.9 mL |
| $dH_2O$ | 3.2 mL | 2.8 mL |
| 4x Resolving gel buffer | 2.5 mL | - |
| 4x Stacking gel buffer | - | 1.25 mL |
| TEMED | 20 µl | 15 µl |
| APS 10 % (v/v) | 150 µl | 85 µl |
| *for separation of proteins with a molecular weight between 25-100 kDa. | | |

**5.2.5.6 Western blot analysis**

Specific proteins separated by SDS-PAGE and was detected by Western blot. Firstly, the SDS-gel and nitrocellulose was incubated on the transfer buffer for about 5 min. Whatman filter papers were

also incubated on the transfer buffer for about 1 min. Secondly, the semi-dry transfer system was built up precisely from the direction of the anode to the cathode: three Whatman filter papers, the blotting membrane (nitrocellulose the SDS-gel and then three Whatman filter papers. Protein transfer was carried out at 50 V and an electric current of 1 mA/cm$^2$ for 75 min.

After blotting, the membrane was transferred into a blocking bottle and incubated with blocking solution for 1 h at RT to block unspecific binding sites. The membrane was incubated in the blocking solution including the first antibody in an appropriate dilution over night at 4°C on a rotator. The next day, the first antibody solution was removed and the membrane was washed thrice with TBS-T buffer for 10 min. After the washing steps the membrane was incubated with the secondary antibody for 1 h at RT. The membrane was washed once again thrice with TBS-T buffer and eventually one time with TBS buffer.

The detection of the antibody-bound proteins was performed by the chemiluminescence, which produced from the reaction catalyzed by the horseradish peroxidase when incubated with $H_2O_2$ and the luminol containing ECL solution. Before the immunodetection, 30 % $H_2O_2$ was added to the ECL solution at a ratio of 1:1000, which then was mixed and poured out on the top of membrane, incubated for about 5 min. After incubation, the ECL solution was removed, the chemiluminescence was detected by placing a photographic film on the membrane immediately.

| Transfer buffer | | Blocking solution | |
|---|---|---|---|
| Tris/HCl (pH 8.4) | 25 mM | Milk powder or BSA In TBS-T buffer | 5 % (w/v) |
| Glycine | 192 mM | | |
| Methanol | 20 % (v/v) | | |
| | | | |
| TBS | | TBS-T | |
| Tris/HCl (pH 7.5) | 100 mM | Tris/HCl (pH 7.5) | 100 mM |
| NaCl | 150 mM | NaCl | 150 mM |
| | | Tween 20 | 0.1 % (v/v) |
| | | | |
| ECL solution | | | |
| Luminol 250 mM* | 400 µl | | |
| Coumaric acid 90 mM* | 178 µL | | |
| Tris/HCl 1 M (pH 8.5) | 4 mL | | |
| | ad 20 mL dH2O | | *in DMSO |

**5.2.5.7 Self-assembling GFP**

The self-assembling GFP method was used to detect the topology of TIP4 protein in *P. tricornutum*. In this approach GFP was split into two parts: long fragment GFPs1-10 and short fragment GFPs11.

For topology analyses of TIP4, GFPs11 was fused to the C-terminus of the protein, whereas the long fragment GFPs1-10 was fused to the C-terminus of specific markers for the ER lumen (PDI protein) and the PPC marker (Hsp70BTS), respectively. Only when both parts are targeted to the same cellular compartment can assemble to produce green fluorescence.

After transfection, the colonies were checked by colony PCR via amplification of the first cloning site (*SpeI/SacII*) with an NR promoter primer and an internal primer from the compartment marker genes.

#### 5.2.5.8 Construction of eGFP fusion proteins

All the C-terminus of predicted protein were fused to eGFP or mRFP and cloned into pPha-NR vector (GenBank accession no. JN180663) or pPha-dual-NR vectors, transfected into the diatom *P. tricornutum* and expressed via the nitrate reductase promoter. The sequences of genes containing introns or predicted gene model was not confirmed by EST data were amplified from cDNA, the rest of genes were amplified from gDNA. For further information about the protein sequences and primer sequences used for transfection and PCR, respectively, see the supplemental material. Biolistic transfection was used as described by Sommer (Sommer, Gould et al. 2007).

#### 5.2.5.9 Confocal laser scanning microscopy

All *P. tricornutum* transfectants were analyzed with a confocal laser scanning microscope. Leica TCS SP2 using a HCX PL APO 40× /1.25 to 0.75 oil CS objective after fixing with 4% paraformaldehyde–0.0075% glutaraldehyde in 1× phosphate buffered saline buffer. The fluorescence of enhanced green fluorescent protein (eGFP) and plastid autofluorescence was excited with an argon laser at 488 nm and detected with two photomultiplier tubes at a bandwidths of 500 to 520 nm and 625 to 720 nm for eGFP and plastid autofluorescence, respectively. The fluorescence of mRFP was excited with a HeNe 1.2 mW laser at 543 nm and detected with a photomultiplier tubes at a bandwidths of 580 to 600 nm for mRFP.

#### 5.2.5.10 Electron microscopy

For transmission electron microscopic analyses *P. tricornutum* cells were harvested (5 min, 2000 g), high pressure frozen (Wohlwend HPF Compact 02) and freeze-substituted (Leica AFS2) with a medium on acetone basis containing 0.25% osmium tetroxide, 0.2% uranyl acetate and 5% water. Samples were then embedded in Epon 812 resin (Fluka) and cut to 50 nm ultrathin sections. Details of the procedure are described in Peschke et al. For immunolabeling on ultrathin sections a primary

antibody against GFP (Rockland) (dilution 1:500 and 1:1000) and a secondary antibody coupled to ultrasmall gold particles (Aurion) (dilution 1:100) were used, followed by silver enhancement. The procedure of immunolabeling and silver enhancement are described in Rachel and Danscher, respectively (Danscher 1981, Rachel, Meyer et al. 2010). After silver enhancement sections were post-stained with 2% uranyl acetate and 0.5% lead citrate. Samples were analysed using a JEOL 2100 electron microscope.

# 6 References

Adl, S. M., A. G. Simpson, M. A. Farmer, R. A. Andersen, O. R. Anderson, J. R. Barta, S. S. Bowser, G. Brugerolle, R. A. Fensome, S. Fredericq, T. Y. James, S. Karpov, P. Kugrens, J. Krug, C. E. Lane, L. A. Lewis, J. Lodge, D. H. Lynn, D. G. Mann, R. M. McCourt, L. Mendoza, O. Moestrup, S. E. Mozley-Standridge, T. A. Nerad, C. A. Shearer, A. V. Smirnov, F. W. Spiegel and M. F. Taylor (2005). "The new higher level classification of eukaryotes with emphasis on the taxonomy of protists." J Eukaryot Microbiol **52**(5): 399-451.

Ahmed, S. U., E. Rojo, V. Kovaleva, S. Venkataraman, J. E. Dombrowski, K. Matsuoka and N. V. Raikhel (2000). "The plant vacuolar sorting receptor AtELP is involved in transport of NH2-terminal propeptide-containing vacuolar proteins in *Arabidopsis thaliana*." The Journal of cell biology **149**(7): 1335-1344.

Apt, K. E., A. Grossman and P. Kroth-Pancic (1996). "Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*." Molecular and General Genetics MGG **252**(5): 572-579.

Archibald, J. M. (2007). "Nucleomorph genomes: structure, function, origin and evolution." Bioessays **29**(4): 392-402.

Arighi, C. N., L. M. Hartnell, R. C. Aguilar, C. R. Haft and J. S. Bonifacino (2004). "Role of the mammalian retromer in sorting of the cation-independent mannose 6-phosphate receptor." The Journal of cell biology **165**(1): 123-133.

Aviezer-Hagai, K., H. Nelson and N. Nelson (2000). "Cloning and expression of cDNAs encoding plant V-ATPase subunits in the corresponding yeast null mutants." Biochimica et Biophysica Acta (BBA)-Bioenergetics **1459**(2): 489-498.

Baïet, B., C. Burel, B. Saint-Jean, R. Louvet, L. Menu-Bouaouiche, M.-C. Kiefer-Meyer, E. Mathieu-Rivet, T. Lefebvre, H. Castel and A. Carlier (2011). "N-glycans of *Phaeodactylum tricornutum* diatom and functional characterization of its N-acetylglucosaminyltransferase I enzyme." Journal of Biological Chemistry **286**(8): 6152-6164.

Bachvaroff, T. R., G. T. Concepcion, C. R. Rogers, E. M. Herman and C. F. Delwiche (2004). "Dinoflagellate expressed sequence tag data indicate massive transfer of chloroplast genes to the nuclear genome." Protist **155**(1): 65-78.

Bachvaroff, T. R., M. V. Sanchez-Puerta and C. F. Delwiche (2006). "Rate variation as a function of gene origin in plastid-derived genes of peridinin-containing dinoflagellates." J Mol Evol **62**(1): 42-52.

Baiet, B., C. Burel, B. Saint-Jean, R. Louvet, L. Menu-Bouaouiche, M. C. Kiefer-Meyer, E. Mathieu-Rivet, T. Lefebvre, H. Castel, A. Carlier, J. P. Cadoret, P. Lerouge and M. Bardor (2011). "N-glycans

of *Phaeodactylum tricornutum* diatom and functional characterization of its N-acetylglucosaminyltransferase I enzyme." J Biol Chem **286**(8): 6152-6164.

Baiges, I., A. R. Schaffner, M. J. Affenzeller and A. Mas (2002). "Plant aquaporins." Physiol Plant **115**(2): 175-182.

Balciunas, D. and S. C. Ekker (2005). "Trapping fish genes with transposons." Zebrafish **1**(4): 335-341.

Barbrook, A., H. Symington, R. Nisbet, A. Larkum and C. Howe (2001). "Organisation and expression of the plastid genome of the dinoflagellate *Amphidinium operculatum*." Molecular Genetics and Genomics **266**(4): 632-638.

Barbrook, A. C., R. G. Dorrell, J. Burrows, L. J. Plenderleith, R. E. Nisbet and C. J. Howe (2012). "Polyuridylylation and processing of transcripts from multiple gene minicircles in chloroplasts of the dinoflagellate *Amphidinium carterae*." Plant Mol Biol **79**(4-5): 347-357.

Barbrook, A. C. and C. J. Howe (2000). "Minicircular plastid DNA in the dinoflagellate *Amphidinium operculatum*." Mol Gen Genet **263**(1): 152-158.

Barbrook, A. C., N. Santucci, L. J. Plenderleith, R. G. Hiller and C. J. Howe (2006). "Comparative analysis of dinoflagellate chloroplast genomes reveals rRNA and tRNA genes." BMC Genomics **7**: 297.

Barlowe, C. (2003). "Signals for COPII-dependent export from the ER: what's the ticket out?" Trends in cell biology **13**(6): 295-300.

Bimboim, H. and J. Doly (1979). "A rapid alkaline extraction procedure for screening recombinant plasmid DNA." Nucleic acids research **7**(6): 1513-1523.

Boczar, B. A., J. Liston and R. A. Cattolico (1991). "Characterization of satellite DNA from three marine dinoflagellates (dinophyceae): *Glenodinium sp.* and two members of the toxic genus, Protogonyaulax." Plant physiology **97**(2): 613-618.

Bodansky, S., L. B. Mintz and D. S. Holmes (1979). "The mesokaryote *Gyrodiniumcohnii* lacks nucleosomes." Biochemical and biophysical research communications **88**(4): 1329-1336.

Bodyl, A., P. Mackiewicz and P. Gagat (2012). "Organelle evolution: *Paulinella* breaks a paradigm." Curr Biol **22**(9): R304-306.

Bodył, A., P. Mackiewicz and J. Stiller (2010). "Comparative genomic studies suggest that the cyanobacterial endosymbionts of the amoeba *Paulinella chromatophora* possess an import apparatus for nuclear-encoded proteins." Plant Biology **12**(4): 639-649.

Boevink, P., K. Oparka, S. S. Cruz, B. Martin, A. Betteridge and C. Hawes (1998). "Stacks on tracks: the plant Golgi apparatus traffics on an actin/ER network†." The Plant Journal **15**(3): 441-447.

Bolte, K., L. Bullmann, F. Hempel, A. Bozarth, S. Zauner and U. G. MAIER (2009). "Protein Targeting into Secondary Plastids1." Journal of Eukaryotic Microbiology **56**(1): 9-15.

Bolte, S., S. Brown and B. Satiat-Jeunemaitre (2004). "The N-myristoylated Rab-GTPase m-Rabmc is involved in post-Golgi trafficking events to the lytic vacuole in plant cells." Journal of cell science **117**(6): 943-954.

Bondili, J. S., A. Castilho, L. Mach, J. Glössl, H. Steinkellner, F. Altmann and R. Strasser (2006). "Molecular cloning and heterologous expression of β1, 2-xylosyltransferase and core α1, 3-fucosyltransferase from maize." Phytochemistry **67**(20): 2215-2224.

Bonifacino, J. S. and J. H. Hurley (2008). "Retromer." Current opinion in cell biology **20**(4): 427-436.

Bonifacino, J. S. and L. M. Traub (2003). "Signals for sorting of transmembrane proteins to endosomes and lysosomes." Annu Rev Biochem **72**: 395-447.

Bosch, D., A. Castilho, A. Loos, A. Schots and H. Steinkellner (2013). "N-Glycosylation of Plant-produced Recombinant Proteins." Current Pharmaceutical Design **19**(31): 5503-5512.

Bouligand, Y. and V. Norris (2001). "Chromosome separation and segregation in dinoflagellates andbacteria may depend on liquid crystalline states." Biochimie **83**(2): 187-192.

Boursiac, Y., S. Chen, D.-T. Luu, M. Sorieul, N. van den Dries and C. Maurel (2005). "Early effects of salinity on water transport in Arabidopsis roots. Molecular and cellular features of aquaporin expression." Plant physiology **139**(2): 790-805.

Breton, C., J. Mucha and C. Jeanneau (2001). "Structural and functional features of glycosyltransferases." Biochimie **83**(8): 713-718.

Burki, F., B. Imanian, E. Hehenberger, Y. Hirakawa, S. Maruyama and P. J. Keeling (2014). "Endosymbiotic gene transfer in tertiary plastid-containing dinoflagellates." Eukaryot Cell **13**(2): 246-255.

Burki, F., Y. Inagaki, J. Brate, J. M. Archibald, P. J. Keeling, T. Cavalier-Smith, M. Sakaguchi, T. Hashimoto, A. Horak, S. Kumar, D. Klaveness, K. S. Jakobsen, J. Pawlowski and K. Shalchian-Tabrizi (2009). "Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, telonemia and centroheliozoa, are related to photosynthetic chromalveolates." Genome Biol Evol **1**: 231-238.

Carlson, C. M., J. L. Frandsen, N. Kirchhof, R. S. McIvor and D. A. Largaespada (2005). "Somatic integration of an oncogene-harboring Sleeping Beauty transposon models liver tumor development in the mouse." Proceedings of the National Academy of Sciences of the United States of America **102**(47): 17059-17064.

Carlton, J., M. Bujny, B. J. Peter, V. M. Oorschot, A. Rutherford, H. Mellor, J. Klumperman, H. T. McMahon and P. J. Cullen (2004). "Sorting nexin-1 mediates tubular endosome-to-TGN transport through coincidence sensing of high-curvature membranes and 3-phosphoinositides." Current biology **14**(20): 1791-1800.

Cavalier-Smith, T. (1998). "A revised six-kingdom system of life." Biol Rev Camb Philos Soc **73**(3): 203-266.

Cavalier-Smith, T. (1999). "Principles of Protein and Lipid Targeting in Secondary Symbiogenesis: Euglenoid, Dinoflagellate, and Sporozoan Plastid Origins and the Eukaryote Family Tree1, 2." Journal of Eukaryotic Microbiology **46**(4): 347-366.

Cavalier-Smith, T. (2000). "membrane heredity and early chloroplast evolution." trends in plant science **5**: 1360-1385.

Cavalier-Smith, T. (2002). "The phagotrophic origin of eukaryotes and phylogenetic classification of protozoa." international journal of systematic and evolutionary microbiology **52**: 297-354.

Cavalier-Smith, T. and E. E.-Y. Chao (2003). "Phylogeny and classification of *phylum Cercozoa* (Protozoa)." Protist **154**(3): 341-358.

Chaumont, F., F. Barrieu, E. Wojcik, M. J. Chrispeels and R. Jung (2001). "Aquaporins constitute a large and highly divergent protein family in maize." Plant Physiology **125**(3): 1206-1215.

Chaumont, F. and S. D. Tyerman (2014). "Aquaporins: highly regulated channels controlling plant water relations." Plant Physiol **164**(4): 1600-1618.

Chevalier, A. S. and F. Chaumont (2014). "Trafficking of Plant Plasma Membrane Aquaporins: Multiple Regulation Levels and Complex Sorting Signals." Plant Cell Physiol.

Chow, M. H., K. T. Yan, M. J. Bennett and J. T. Wong (2010). "Birefringence and DNA condensation of liquid crystalline chromosomes." Eukaryotic cell **9**(10): 1577-1587.

Clarke, M., J. Köhler, Q. Arana, T. Liu, J. Heuser and G. Gerisch (2002). "Dynamics of the vacuolar H+-ATPase in the contractile vacuole complex and the endosomal pathway of Dictyostelium cells." Journal of cell science **115**(14): 2893-2905.

Contreras, I., E. Ortiz-Zapater, L. M. Castilho and F. Aniento (2000). "Characterization of Cop I coat proteins in plant cells." Biochem Biophys Res Commun **273**(1): 176-182.

Cullen, P. J. and H. C. Korswagen (2012). "Sorting nexins provide diversity for retromer-dependent trafficking events." Nature Cell Biology **14**(1): 29-37.

Dagan, T. and W. Martin (2009). "Getting a better picture of microbial evolution en route to a network of genomes." Philos Trans R Soc Lond B Biol Sci **364**(1527): 2187-2196.

Dang, Y. and B. R. Green (2009). "Substitutional editing of Heterocapsa triquetra chloroplast transcripts and a folding model for its divergent chloroplast 16S rRNA." Gene **442**(1-2): 73-80.

Dang, Y. and B. R. Green (2010). "Long transcripts from dinoflagellate chloroplast minicircles suggest "rolling circle" transcription." J Biol Chem **285**(8): 5196-5203.

Danscher, G. (1981). "Localization of gold in biological tissue." Histochemistry **71**(1): 81-88.

daSilva, L. L., E. L. Snapp, J. Denecke, J. Lippincott-Schwartz, C. Hawes and F. Brandizzi (2004). "Endoplasmic reticulum export sites and Golgi bodies behave as single mobile secretory units in plant cells." Plant Cell **16**(7): 1753-1771.

De Marchis, F., M. Bellucci and A. Pompa (2013). "Traffic of human alpha-mannosidase in plant cells suggests the presence of a new endoplasmic reticulum-to-vacuole pathway without involving the Golgi complex." Plant Physiol **161**(4): 1769-1782.

Dettmer, J., A. Hong-Hermesdorf, Y.-D. Stierhof and K. Schumacher (2006). "Vacuolar H+-ATPase activity is required for endocytic and secretory trafficking in *Arabidopsis*." The Plant Cell **18**(3): 715-730.

Dettmer, J., T. Y. Liu and K. Schumacher (2010). "Functional analysis of Arabidopsis V-ATPase subunit VHA-E isoforms." Eur J Cell Biol **89**(2-3): 152-156.

Deusch, O., G. Landan, M. Roettger, N. Gruenheit, K. V. Kowallik, J. F. Allen, W. Martin and T. Dagan (2008). "Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor." Mol Biol Evol **25**(4): 748-761.

Dodge, J. and J. Lee (2000). "Phylum dinoflagellata." Illustrated Guide to the Protozoa. Society of Protozoologists. Lawrence, KA: 656-689.

Dorrell, R. G. and A. G. Smith (2011). "Do red and green make brown?: perspectives on plastid acquisitions within chromalveolates." Eukaryot Cell **10**(7): 856-868.

Doyle, J. and J. Doyle (1990). "Isolation of DNA from small amounts of plant tissues." BRL focus **12**: 13-15.

Ebine, K., T. Inoue, J. Ito, E. Ito, T. Uemura, T. Goh, H. Abe, K. Sato, A. Nakano and T. Ueda (2014). "Plant vacuolar trafficking occurs through distinctly regulated pathways." Curr Biol **24**(12): 1375-1382.

Endler, A., S. Meyer, S. Schelbert, T. Schneider, W. Weschke, S. W. Peters, F. Keller, S. Baginsky, E. Martinoia and U. G. Schmidt (2006). "Identification of a vacuolar sucrose transporter in barley and Arabidopsis mesophyll cells by a tonoplast proteomic approach." Plant Physiology **141**(1): 196-207.

Facchinelli, F., M. Pribil, U. Oster, N. J. Ebert, D. Bhattacharya, D. Leister and A. P. Weber (2013). "Proteomic analysis of the Cyanophora paradoxa muroplast provides clues on early events in plastid endosymbiosis." Planta **237**(2): 637-651.

Fast, N. M., J. C. Kissinger, D. S. Roos and P. J. Keeling (2001). "Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids." Molecular Biology and Evolution **18**(3): 418-426.

Fischer, M. and R. Kaldenhoff (2008). "On the pH regulation of plant aquaporins." J Biol Chem **283**(49): 33889-33892.

Fitchette-Lainé, A. C., V. Gomord, A. Chekkafi and L. Faye (1994). "Distribution of xylosylation and fucosylation in the plant Golgi apparatus." The Plant Journal **5**(5): 673-682.

Frigerio, L., G. Hinz and D. G. Robinson (2008). "Multiple vacuoles in plant cells: rule or exception?" Traffic **9**(10): 1564-1570.

Frigerio, L., N. A. Jolliffe, A. Di Cola, D. H. Felipe, N. Paris, J.-M. Neuhaus, J. M. Lord, A. Ceriotti and L. M. Roberts (2001). "The internal propeptide of the ricin precursor carries a sequence-specific determinant for vacuolar sorting." Plant physiology **126**(1): 167-175.

Frommolt, R., S. Werner, H. Paulsen, R. Goss, C. Wilhelm, S. Zauner, U. G. Maier, A. R. Grossman, D. Bhattacharya and M. Lohr (2008). "Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis." Mol Biol Evol **25**(12): 2653-2667.

Gabrielsen, T. M., M. A. Minge, M. Espelund, A. Tooming-Klunderud, V. Patil, A. J. Nederbragt, C. Otis, M. Turmel, K. Shalchian-Tabrizi, C. Lemieux and K. S. Jakobsen (2011). "Genome evolution of a tertiary dinoflagellate plastid." PLoS One **6**(4): e19132.

Galili, A. V. a. G. (2001). "The endomembrane system and the problem of protein sorting." plant Physiol **125**: 115-118.

Gattolin, S., M. Sorieul and L. Frigerio (2011). "Mapping of tonoplast intrinsic proteins in maturing and germinating Arabidopsis seeds reveals dual localization of embryonic TIPs to the tonoplast and plasma membrane." Mol Plant **4**(1): 180-189.

Gattolin, S., M. Sorieul, P. R. Hunter, R. H. Khonsari and L. Frigerio (2009). "In vivo imaging of the tonoplast intrinsic protein family in Arabidopsis roots." BMC Plant Biol **9**: 133.

Gautier, A., L. Michel-Salamin, E. Tosi-Couture, A. McDowall and J. Dubochet (1986). "Electron microscopy of the chromosomes of dinoflagellates in situ: confirmation of Bouligand's liquid crystal hypothesis." Journal of ultrastructure and molecular structure research **97**(1): 10-30.

Gautreau, A., K. Oguievetskaia and C. Ungermann (2014). "Function and regulation of the endosomal fusion and fission machineries." Cold Spring Harb Perspect Biol **6**(3).

Geldner, N., N. Anders, H. Wolters, J. Keicher, W. Kornberger, P. Muller, A. Delbarre, T. Ueda, A. Nakano and G. Jürgens (2003). "The Arabidopsis GNOM ARF-GEF mediates endosomal recycling, auxin transport, and auxin-dependent plant growth." Cell **112**(2): 219-230.

Geuze, W. H. a. H. J. (1995). "Intracellular trafficking of lysosomal membrane proteins." Bioessays **18**: 379-389.

Gibbs, S. P. (1979). "The route of entry of cytoplasmically synthesized proteins into chloroplasts of algae possessing chloroplast ER." Journal of cell science **35**(1): 253-266.

Gillespie, J., S. W. Rogers, M. Deery, P. Dupree and J. C. Rogers (2005). "A unique family of proteins associated with internalized membranes in protein storage vacuoles of the Brassicaceae." The Plant Journal **41**(3): 429-441.

Gould, S. B., U. G. Maier and W. F. Martin (2015). "Protein Import and the Origin of Red Complex Plastids." Curr Biol **25**(12): R515-R521.

Gould, S. B., R. F. Waller and G. I. McFadden (2008). "Plastid evolution." Annu. Rev. Plant Biol. **59**: 491-517.

Green, B. R. (2004). "The chloroplast genome of dinoflagellates–a reduced instruction set?" Protist **155**(1): 23-31.

Green, B. R. (2011). "After the primary endosymbiosis: an update on the chromalveolate hypothesis and the origins of algae with Chl c." Photosynth Res **107**(1): 103-115.

Grosche, C. (2012). "Spezielle Leistungen der plastide: RNA-edierung in Landpflanzen, Genomreduktion und Proteinimport in Peridinin-haltigen Dinoflagellaten." Dissertation.

Grosche, C., H. T. Funk, U. G. Maier and S. Zauner (2012). "The chloroplast genome of Pellia endiviifolia: gene content, RNA-editing pattern, and the origin of chloroplast editing." Genome Biol Evol **4**(12): 1349-1357.

Grosche, C., F. Hempel, K. Bolte, S. Zauner and U. G. Maier (2014). "The periplastidal compartment: a naturally minimized eukaryotic cytoplasm." Curr Opin Microbiol **22**: 88-93.

Gruber, A., G. Rocap, P. G. Kroth, E. V. Armbrust and T. Mock (2015). "Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage." Plant J **81**(3): 519-528.

Gruber, A., S. Vugrinec, F. Hempel, S. B. Gould, U. G. Maier and P. G. Kroth (2007). "Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif." Plant Mol Biol **64**(5): 519-530.

Höfte, H., L. Hubbard, J. Reizer, D. Ludevid, E. M. Herman and M. J. Chrispeels (1992). "Vegetative and seed-specific forms of tonoplast intrinsic protein in the vacuolar membrane of *Arabidopsis thaliana.*" Plant Physiology **99**(2): 561-570.

Hackett, J. D., H. S. Yoon, S. Li, A. Reyes-Prieto, S. E. Rummele and D. Bhattacharya (2007). "Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates." Mol Biol Evol **24**(8): 1702-1713.

Hackett, J. D., H. S. Yoon, M. B. Soares, M. F. Bonaldo, T. L. Casavant, T. E. Scheetz, T. Nosenko and D. Bhattacharya (2004). "Migration of the plastid genome to the nucleus in a peridinin dinoflagellate." Curr Biol **14**(3): 213-218.

Hackett, P. B., D. A. Largaespada and L. J. Cooper (2010). "A transposon and transposase system for human application." Molecular Therapy **18**(4): 674-683.

Hampl, V., L. Hug, J. W. Leigh, J. B. Dacks, B. F. Lang, A. G. Simpson and A. J. Roger (2009). "Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups"." Proc Natl Acad Sci U S A **106**(10): 3859-3864.

Hanton, S. L., L. Renna, L. E. Bortolotti, L. Chatre, G. Stefano and F. Brandizzi (2005). "Diacidic motifs influence the export of transmembrane proteins from the endoplasmic reticulum in plant cells." Plant Cell **17**(11): 3081-3093.

Happel, N., S. Höning, J. M. Neuhaus, N. Paris, D. G. Robinson and S. E. Holstein (2004). "ArabidopsisμA-adaptin interacts with the tyrosine motif of the vacuolar sorting receptor VSR-PS1." The Plant Journal **37**(5): 678-693.

Hara-Nishimura, I., R. Matsushima, T. Shimada and M. Nishimura (2004). "Diversity and formation of endoplasmic reticulum-derived compartments in plants. Are these compartments specific to plant cells?" Plant Physiol **136**(3): 3435-3439.

Hara-Hishimura, I., Y. Takeuchi, K. Inoue and M. Nishimura (1993). "Vesicle transport and processing of the precursor to 2S albumin in pumpkin." The Plant Journal **4**(5): 793-800.

Harbour, M. E., S. Y. Breusegem, R. Antrobus, C. Freeman, E. Reid and M. N. Seaman (2010). "The cargo-selective retromer complex is a recruiting hub for protein complexes that regulate endosomal tubule dynamics." Journal of cell science **123**(21): 3703-3717.

Harper, J. T. and P. J. Keeling (2003). "Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids." Mol Biol Evol **20**(10): 1730-1735.

Hegedus, D. D., C. Coutu, M. Harrington, B. Hope, K. Gerbrandt and I. Nikolov (2015). "Multiple internal sorting determinants can contribute to the trafficking of cruciferin to protein storage vacuoles." Plant molecular biology **88**(1-2): 3-20.

Hempel, F., L. Bullmann, J. Lau, S. Zauner and U. G. Maier (2009). "ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms." Molecular biology and evolution **26**(8): 1781-1790.

Hempel, F., G. Felsner and U. G. Maier (2010). "New mechanistic insights into pre-protein transport across the second outermost plastid membrane of diatoms." Mol Microbiol **76**(3): 793-801.

Herman, E. and M. Schmidt (2004). "Endoplasmic reticulum to vacuole trafficking of endoplasmic reticulum bodies provides an alternate pathway for protein transfer to the vacuole." Plant Physiol **136**(3): 3440-3446.

Herman, E. M. (2008). "Endoplasmic reticulum bodies: solving the insoluble." Current opinion in plant biology **11**(6): 672-679.

Hiller, R. G. (2001). "'Empty' minicircles and petB/atpA and psbD/psbE (cytb559 alpha) genes in tandem in *Amphidinium carterae* plastid DNA." FEBS Lett **505**(3): 449-452.

Hiller, R. G. (2001). "'Empty'minicircles and petB/atpA and psbD/psbE (cytb 559 α) genes in tandem in *Amphidinium carterae* plastid DNA." FEBS letters **505**(3): 449-452.

Hohl, I., D. G. Robinson, M. J. Chrispeels and G. Hinz (1996). "Transport of storage proteins to the vacuole is mediated by vesicles without a clathrin coat." Journal of Cell Science **109**(10): 2539-2550.

Holstein, S. E. (2002). "Clathrin and plant endocytosis." Traffic **3**(9): 614-620.

Horazdovsky, B., B. Davies, M. Seaman, S. McLaughlin, S.-H. Yoon and S. Emr (1997). "A sorting nexin-1 homologue, Vps5p, forms a complex with Vps17p and is required for recycling the vacuolar protein-sorting receptor." Molecular biology of the cell **8**(8): 1529-1541.

Howe, C. J., A. C. Barbrook, V. L. Koumandou, R. E. Nisbet, H. A. Symington and T. F. Wightman (2003). "Evolution of the chloroplast genome." Philos Trans R Soc Lond B Biol Sci **358**(1429): 99-106; discussion 106-107.

Howe, C. J., R. E. Nisbet and A. C. Barbrook (2008). "The remarkable chloroplast genome of dinoflagellates." J Exp Bot **59**(5): 1035-1045.

Hunter, P. R., C. P. Craddock, S. Di Benedetto, L. M. Roberts and L. Frigerio (2007). "Fluorescent reporter proteins for the tonoplast and the vacuolar lumen identify a single vacuolar compartment in Arabidopsis cells." Plant Physiol **145**(4): 1371-1382.

Hwang, I. (2008). "Sorting and anterograde trafficking at the Golgi apparatus." Plant Physiol **148**(2): 673-683.

Islam, M. S., N. I. A. Patwary, N. H. Muzahid, S. M. Shahik, M. Sohel and M. A. Hasan (2014). "A systematic study on structure and function of ATPase of wuchereria bancrofti." Toxicology international **21**(3): 269.

Jürgens, G. (2004). "Membrane trafficking in plants." Annu. Rev. Cell Dev. Biol. **20**: 481-504.

Jaillais, Y., M. Santambrogio, F. Rozier, I. Fobis-Loisy, C. Miège and T. Gaude (2007). "The retromer protein VPS29 links cell polarity and organ initiation in plants." Cell **130**(6): 1057-1070.

Jauh, G.-Y., T. E. Phillips and J. C. Rogers (1999). "Tonoplast intrinsic protein isoforms as markers for vacuolar functions." The Plant Cell **11**(10): 1867-1882.

Jiang, L., T. E. Phillips, S. W. Rogers and J. C. Rogers (2000). "Biogenesis of the protein storage vacuole crystalloid." The Journal of cell biology **150**(4): 755-770.

Jin, J. B., Y. A. Kim, S. J. Kim, S. H. Lee, D. H. Kim, G.-W. Cheong and I. Hwang (2001). "A new dynamin-like protein, ADL6, is involved in trafficking from the trans-Golgi network to the central vacuole in Arabidopsis." The Plant Cell **13**(7): 1511-1526.

Johanson, U. and S. Gustavsson (2002). "A new subfamily of major intrinsic proteins in plants." Molecular biology and evolution **19**(4): 456-461.

Johanson, U., M. Karlsson, I. Johansson, S. Gustavsson, S. Sjövall, L. Fraysse, A. R. Weig and P. Kjellbom (2001). "The complete set of genes encoding major intrinsic proteins in Arabidopsis provides a framework for a new nomenclature for major intrinsic proteins in plants." Plant physiology **126**(4): 1358-1369.

Johansson, I., M. Karlsson, U. Johanson, C. Larsson and P. Kjellbom (2000). "The role of aquaporins in cellular and whole plant water balance." Biochim Biophys Acta **1465**(1-2): 324-342.

Johnson, A. E. and M. A. van Waes (1999). "The translocon: a dynamic gateway at the ER membrane." Annual review of cell and developmental biology **15**(1): 799-842.

Johnson, K. D. and M. J. Chrispeels (1992). "Tonoplast-bound protein kinase phosphorylates tonoplast intrinsic protein." Plant Physiology **100**(4): 1787-1795.

Jolliffe, N., C. Craddock and L. Frigerio (2005). "Pathways for protein transport to seed storage vacuoles." Biochemical Society Transactions **33**(Pt 5): 1016-1018.

Jolliffe, N. A., J. C. Brown, U. Neumann, M. Vicre, A. Bachi, C. Hawes, A. Ceriotti, L. M. Roberts and L. Frigerio (2004). "Transport of ricin and 2S albumin precursors to the storage vacuoles of Ricinus communis endosperm involves the Golgi and VSR-like receptors." Plant J **39**(6): 821-833.

Kajiura, H., T. Okamoto, R. Misaki, Y. Matsuura and K. Fujiyama (2012). "Arabidopsis beta1,2-xylosyltransferase: substrate specificity and participation in the plant-specific N-glycosylation pathway." J Biosci Bioeng **113**(1): 48-54.

Kaldenhoff, R., A. Bertl, B. Otto, M. Moshelion and N. Uehlein (2007). "Characterization of plant aquaporins." Methods Enzymol **428**: 505-531.

Kaldenhoff, R. and M. Fischer (2006). "Aquaporins in plants." Acta Physiol (Oxf) **187**(1-2): 169-176.

Kaldenhoff, R., M. Ribas-Carbo, J. F. Sans, C. Lovisolo, M. Heckwolf and N. Uehlein (2008). "Aquaporins and plant water balance." Plant Cell Environ **31**(5): 658-666.

Kalthoff, C., S. Groos, R. Kohl, S. Mahrhold and E. J. Ungewickell (2002). "Clint: a novel clathrin-binding ENTH-domain protein at the Golgi." Mol Biol Cell **13**(11): 4060-4073.

Karlsson, M., I. Johansson, M. Bush, M. C. McCann, C. Maurel, C. Larsson and P. Kjellbom (2000). "An abundant TIP expressed in mature highly vacuolated cells." The Plant Journal **21**(1): 83-90.

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

Keeling, P. J. (2010). "The endosymbiotic origin, diversification and fate of plastids." Philos Trans R Soc Lond B Biol Sci **365**(1541): 729-748.

Keeling, P. J. (2013). "The number, speed, and impact of plastid endosymbioses in eukaryotic evolution." Annual review of plant biology **64**: 583-607.

Kim, H. S., J. H. Jeon, K. J. Lee and K. Ko (2014). "N-glycosylation modification of plant-derived virus-like particles: an application in vaccines." Biomed Res Int **2014**: 249519.

Kim, J. M., S. Vanguri, J. D. Boeke, A. Gabriel and D. F. Voytas (1998). "Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence." Genome research **8**(5): 464-478.

Kleine-Vehn, J., J. Leitner, M. Zwiewka, M. Sauer, L. Abas, C. Luschnig and J. Friml (2008). "Differential degradation of PIN2 auxin efflux carrier by retromer-dependent vacuolar targeting." Proceedings of the National Academy of Sciences **105**(46): 17812-17817.

Kluge, C., J. Lahr, M. Hanitzsch, S. Bolte, D. Golldack and K.-J. Dietz (2003). "New insight into the structure and regulation of the plant vacuolar H+-ATPase." Journal of bioenergetics and biomembranes **35**(4): 377-388.

Koide, Y., K. Matsuoka, M.-a. Ohto and K. Nakamura (1999). "The N-terminal propeptide and the C terminus of the precursor to 20-kilo-dalton potato tuber protein can function as different types of vacuolar sorting signals." Plant and Cell Physiology **40**(11): 1152-1159.

Kornfeld, R. and S. Kornfeld (1985). "Assembly of asparagine-linked oligosaccharides." Annual review of biochemistry **54**(1): 631-664.

Koumandou, V. L. and C. J. Howe (2007). "The copy number of chloroplast gene minicircles changes dramatically with growth phase in the dinoflagellate *Amphidinium operculatum*." Protist **158**(1): 89-103.

Koumandou, V. L., R. E. R. Nisbet, A. C. Barbrook and C. J. Howe (2004). "Dinoflagellate chloroplasts–where have all the genes gone?" TRENDS in Genetics **20**(5): 261-267.

Krebs, M., D. Beyhl, E. Gorlich, K. A. Al-Rasheid, I. Marten, Y. D. Stierhof, R. Hedrich and K. Schumacher (2010). "Arabidopsis V-ATPase activity at the tonoplast is required for efficient nutrient storage but not for sodium accumulation." Proc Natl Acad Sci U S A **107**(7): 3251-3256.

Kunnimalaiyaan, M., F. Shi and B. L. Nielsen (1997). "Analysis of the tobacco chloroplast DNA replication origin (ori B) downstream of the 23 S rRNA gene." Journal of molecular biology **268**(2): 273-283.

Laatsch, T., S. Zauner, B. Stoebe-Maier, K. V. Kowallik and U. G. Maier (2004). "Plastid-derived single gene minicircles of the dinoflagellate *Ceratium horridum* are localized in the nucleus." Mol Biol Evol **21**(7): 1318-1322.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." nature **227**(5259): 680-685.

Lane, C. E. and J. M. Archibald (2008). "The eukaryotic tree of life: endosymbiosis takes its TOL." Trends Ecol Evol **23**(5): 268-275.

Lee, M. C. and E. A. Miller (2007). Molecular mechanisms of COPII vesicle formation. Seminars in cell & developmental biology, Elsevier.

Lee, M. C. and E. A. Miller (2007). "Molecular mechanisms of COPII vesicle formation." Semin Cell Dev Biol **18**(4): 424-434.

Lemgruber, L., M. Kudryashev, C. Dekiwadia, D. T. Riglar, J. Baum, H. Stahlberg, S. A. Ralph and F. Frischknecht (2013). "Cryo-electron tomography reveals four-membrane architecture of the Plasmodium apicoplast." Malaria journal **12**(1): 25.

Li, G., V. Santoni and C. Maurel (2014). "Plant aquaporins: roles in plant physiology." Biochim Biophys Acta **1840**(5): 1574-1582.

Lienard, D., G. Durambur, M. C. Kiefer-Meyer, F. Nogue, L. Menu-Bouaouiche, F. Charlot, V. Gomord and J. P. Lassalles (2008). "Water transport by aquaporins in the extant plant Physcomitrella patens." Plant Physiol **146**(3): 1207-1218.

Liu, L.-H., U. Ludewig, B. Gassert, W. B. Frommer and N. von Wirén (2003). "Urea transport by nitrogen-regulated tonoplast intrinsic proteins in *Arabidopsis*." Plant physiology **133**(3): 1220-1228.

Loque, D., U. Ludewig, L. Yuan and N. von Wiren (2005). "Tonoplast intrinsic proteins AtTIP2;1 and AtTIP2;3 facilitate NH3 transport into the vacuole." Plant Physiol **137**(2): 671-680.

Ma, B., D. Qian, Q. Nan, C. Tan, L. An and Y. Xiang (2012). "Arabidopsis vacuolar H+-ATPase (V-ATPase) B subunits are involved in actin cytoskeleton remodeling via binding to, bundling, and stabilizing F-actin." J Biol Chem **287**(23): 19008-19017.

Maier, U.-G., S. E. Douglas and T. Cavalier-Smith (2000). "The nucleomorph genomes of cryptophytes and chlorarachniophytes." Protist **151**(2): 103-109.

Maldonado-Mendoza, I. E. and C. L. Nessler (1996). "Cloning and expression of a plant homologue of the small subunit of the Golgi-associated clathrin assembly protein AP19 from *Camptotheca acuminata*." Plant Mol Biol **32**(6): 1149-1153.

Mallard, F., B. L. Tang, T. Galli, D. Tenza, A. Saint-Pol, X. Yue, C. Antony, W. Hong, B. Goud and L. Johannes (2002). "Early/recycling endosomes-to-TGN transport involves two SNARE complexes and a Rab6 isoform." J Cell Biol **156**(4): 653-664.

Marin, B., E. C. Nowack and M. Melkonian (2005). "A plastid in the making: evidence for a second primary endosymbiosis." Protist **156**(4): 425-432.

Martin, W. and R. G. Herrmann (1998). "Gene transfer from organelles to the nucleus: How much, what happens, and why?" Plant Physiology **118**(1): 9-17.

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

Marty, F. (1999). "Plant vacuoles." The Plant Cell **11**(4): 587-599.

Maruyama, N., L. C. Mun, M. Tatsuhara, M. Sawada, M. Ishimoto and S. Utsumi (2006). "Multiple vacuolar sorting determinants exist in soybean 11S globulin." Plant Cell **18**(5): 1253-1273.

Mathieu-Rivet, E., M. C. Kiefer-Meyer, G. Vanier, C. Ovide, C. Burel, P. Lerouge and M. Bardor (2014). "Protein N-glycosylation in eukaryotic microalgae and its impact on the production of nuclear expressed biopharmaceuticals." Front Plant Sci **5**: 359.

Matsuoka, K., T. Higuchi, M. Maeshima and K. Nakamura (1997). "A vacuolar-type H+-ATPase in a nonvacuolar organelle is required for the sorting of soluble vacuolar protein precursors in tobacco cells." The Plant Cell **9**(4): 533-546.

Matsuoka, K. N. a. K. (1993). "protein targeting to the vacuole in plant cells." Plant Physiol. **101**: 1-5.

Matsushima R, H. Y., Yamada K, Shimada T, Nishimura M and Nishimura IH (2003). "The ER body, a novel endoplasmic reticulum-derived structure in arabidopsis." Plant Cell Physiol **44(7)**: 661-666.

Maurel, C. (1997). "Aquaporins and Water Permeability of Plant Membranes." Annu Rev Plant Physiol Plant Mol Biol **48**: 399-429.

Maurel, C. and C. Plassard (2011). "Aquaporins: for more than water at the plant-fungus interface?" New Phytol **190**(4): 815-817.

Mcfadden (2001). "Primary and secondary endosymbiosis and the origin of plastids." J phycol **37**: 951-959.

McGough, I. J. and P. J. Cullen (2011). "Recent advances in retromer biology." Traffic **12**(8): 963-971.

McKersie, B. D. and W. Bruce (2009). Transgenic Plants with Increased Yield, Google Patents.

McMahon, H. T. and E. Boucrot (2011). "Molecular mechanism and physiological functions of clathrin-mediated endocytosis." Nat Rev Mol Cell Biol **12**(8): 517-533.

Melkonian, M. (1996). "Phylogeny of photosynthetic protists and their plastids." VERHANDLUNGEN-DEUTSCHEN ZOOLOGISCHEN GESELLSCHAFT **89**: 71-96.

Michaeli, S., T. Avin-Wittenberg and G. Galili (2014). "Involvement of autophagy in the direct ER to vacuole protein trafficking route in plants." Frontiers in plant science **5**.

Miller, E. A., T. H. Beilharz, P. N. Malkus, M. C. S. Lee, S. Hamamoto, L. Orci and R. Schekman (2003). "Multiple Cargo Binding Sites on the COPII Subunit Sec24p Ensure Capture of Diverse Membrane Proteins into Transport Vesicles." Cell **114**(4): 497-509.

Miyata, Y. and M. Sugita (2004). "Tissue- and stage-specific RNA editing of rps 14 transcripts in moss (Physcomitrella patens) chloroplasts." J Plant Physiol **161**(1): 113-115.

Moore, R. B. (2003). "Highly organized structure in the non-coding region of the psbA minicircle from clade C *Symbiodinium*." International Journal of Systematic and Evolutionary Microbiology **53**(6): 1725-1734.

Moriyasu, Y., M. Hattori, G.-Y. Jauh and J. C. Rogers (2003). "Alpha tonoplast intrinsic protein is specifically associated with vacuole membrane involved in an autophagic process." Plant and cell physiology **44**(8): 795-802.

Mossessova, E., L. C. Bickford and J. Goldberg (2003). "SNARE Selectivity of the COPII Coat." Cell **114**(4): 483-495.

Moustafa, A., B. Beszteri, U. G. Maier, C. Bowler, K. Valentin and D. Bhattacharya (2009). "Genomic footprints of a cryptic plastid endosymbiosis in diatoms." science **324**(5935): 1724-1726.

Movafeghi, A., N. Happel, P. Pimpl, G.-H. Tai and D. G. Robinson (1999). "Arabidopsis Sec21p and Sec23p homologs. Probable coat proteins of plant COP-coated vesicles." Plant Physiology **119**(4): 1437-1446.

Mungpakdee, S., C. Shinzato, T. Takeuchi, T. Kawashima, R. Koyanagi, K. Hisata, M. Tanaka, H. Goto, M. Fujie, S. Lin, N. Satoh and E. Shoguchi (2014). "Massive gene transfer and extensive RNA editing of a symbiotic dinoflagellate plastid genome." Genome Biol Evol **6**(6): 1408-1422.

Nelson, M. J. and B. R. Green (2005). "Double hairpin elements and tandem repeats in the non-coding region of Adenoides eludens chloroplast gene minicircles." Gene **358**: 102-110.

Neuhaus, H. E. (2007). "Transport of primary metabolites across the plant vacuolar membrane." FEBS Lett **581**(12): 2223-2226.

Neuhaus, J.-M. and J. C. Rogers (1998). Sorting of proteins to vacuoles in plant cells. Protein Trafficking in Plant Cells, Springer**:** 127-144.

Niemes, S., M. Langhans, C. Viotti, D. Scheuring, M. San Wan Yan, L. Jiang, S. Hillmer, D. G. Robinson and P. Pimpl (2010). "Retromer recycles vacuolar sorting receptors from the trans-Golgi network." Plant J **61**(1): 107-121.

Nisbet, R. E., R. G. Hiller, E. R. Barry, P. Skene, A. C. Barbrook and C. J. Howe (2008). "Transcript analysis of dinoflagellate plastid gene minicircles." Protist **159**(1): 31-39.

Nisbet, R. E., L. Koumandou, A. C. Barbrook and C. J. Howe (2004). "Novel plastid gene minicircles in the dinoflagellate *Amphidinium operculatum*." Gene **331**: 141-147.

Nishizawa, K., N. Maruyama, R. Satoh, Y. Fuchikami, T. Higasa and S. Utsumi (2003). "AC-terminal sequence of soybean β-conglycinin α' subunit acts as a vacuolar sorting determinant in seed cells." The Plant Journal **34**(5): 647-659.

Nothwehr, S. F., P. Bruinsma and L. A. Strawn (1999). "Distinct domains within Vps35p mediate the retrieval of two different cargo proteins from the yeast prevacuolar/endosomal compartment." Molecular biology of the cell **10**(4): 875-890.

Nowack, E. C. and A. R. Grossman (2012). "Trafficking of protein into the recently established photosynthetic organelles of *Paulinella chromatophora*." Proceedings of the National Academy of Sciences **109**(14): 5340-5345.

Ochoa de Alda, J. A., R. Esteban, M. L. Diago and J. Houmard (2014). "The plastid ancestor originated among one of the major cyanobacterial lineages." Nat Commun **5**: 4937.

Ohta, N., M. Matsuzaki, O. Misumi, S.-y. Miyagishima, H. Nozaki, K. Tanaka, T. Shin-i, Y. Kohara and T. Kuroiwa (2003). "Complete sequence and analysis of the plastid genome of the unicellular red alga *Cyanidioschyzon merolae*." DNA research **10**(2): 67-77.

Ohyama, K., H. Fukuzawa, T. Kohchi, H. Shirai, T. Sano, S. Sano, K. Umesono, Y. Shiki, M. Takeuchi and Z. Chang (1986). "Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA." Nature **322**: 572-574.

Oliviusson, P., O. Heinzerling, S. Hillmer, G. Hinz, Y. C. Tse, L. Jiang and D. G. Robinson (2006). "Plant retromer, localized to the prevacuolar compartment and microvesicles in *Arabidopsis*, may interact with vacuolar sorting receptors." Plant Cell **18**(5): 1239-1252.

Otegui, M. S. and C. Spitzer (2008). "Endosomal functions in plants." Traffic **9**(10): 1589-1598.

Pagny, S., F. Bouissonnie, M. Sarkar, M. Follet-Gueye, A. Driouich, H. Schachter, L. Faye and V. Gomord (2003). "Structural requirements for Arabidopsis β1, 2-xylosyltransferase activity and targeting to the Golgi." The Plant Journal **33**(1): 189-203.

Pandey, B., P. Sharma, D. M. Pandey, I. Sharma and R. Chatrath (2013). "Identification of new aquaporin genes and single nucleotide polymorphism in bread wheat." Evol Bioinform Online **9**: 437-452.

Paris, N., B. Saint-Jean, M. Faraco, W. Krzeszowiec, G. Dalessandro, J.-M. Neuhaus and G. P. Di Sansebastiano (2010). "Expression of a glycosylated GFP as a bivalent reporter in exocytosis." Plant cell reports **29**(1): 79-86.

Park, M., S. J. Kim, A. Vitale and I. Hwang (2004). "Identification of the protein storage vacuole and protein targeting to the vacuole in leaf cells of three plant species." Plant physiology **134**(2): 625-639.

Park, W., B. E. Scheffler, P. J. Bauer and B. T. Campbell (2010). "Identification of the family of aquaporin genes and their expression in upland cotton (*Gossypium hirsutum L.*)." BMC plant biology **10**(1): 142.

Patron, N. J., M. B. Rogers and P. J. Keeling (2004). "Gene replacement of fructose-1,6-bisphosphate aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates." Eukaryot Cell **3**(5): 1169-1175.

Patron, N. J., R. F. Waller, J. M. Archibald and P. J. Keeling (2005). "Complex protein targeting to dinoflagellate plastids." Journal of molecular biology **348**(4): 1015-1024.

Pereira, C., S. Pereira and J. Pissarra (2014). "Delivering of proteins to the plant vacuole--an update." Int J Mol Sci **15**(5): 7611-7623.

Pereira, C., S. Pereira, B. Satiat-Jeunemaitre and J. Pissarra (2013). "Cardosin A contains two vacuolar sorting signals using different vacuolar routes in tobacco epidermal cells." Plant J **76**(1): 87-100.

Petersen, J., A.-K. Ludewig, V. Michael, B. Bunk, M. Jarek, D. Baurain and H. Brinkmann (2014). "Chromera velia, endosymbioses and the rhodoplex hypothesis—plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages)." Genome biology and evolution **6**(3): 666-684.

Phan, N. Q., S.-J. Kim and D. C. Bassham (2008). "Overexpression of Arabidopsis sorting nexin AtSNX2b inhibits endocytic trafficking to the vacuole." Molecular plant **1**(6): 961-976.

Pimpl, P., S. L. Hanton, J. P. Taylor, L. L. Pinto-daSilva and J. Denecke (2003). "The GTPase ARF1p controls the sequence-specific vacuolar sorting route to the lytic vacuole." The Plant Cell **15**(5): 1242-1256.

Pimpl, P., A. Movafeghi, S. Coughlan, J. Denecke, S. Hillmer and D. G. Robinson (2000). "In situ localization and in vitro induction of plant COPI-coated vesicles." The Plant Cell **12**(11): 2219-2235.

Pimpl, P., J. P. Taylor, C. Snowden, S. Hillmer, D. G. Robinson and J. Denecke (2006). "Golgi-mediated vacuolar sorting of the endoplasmic reticulum chaperone BiP may play an active role in quality control within the secretory pathway." Plant Cell **18**(1): 198-211.

Piper, R. C. and D. J. Katzmann (2007). "Biogenesis and function of multivesicular bodies." Annu Rev Cell Dev Biol **23**: 519-547.

Pompa, A., F. De Marchis, A. Vitale, S. Arcioni and M. Bellucci (2010). "An engineered C-terminal disulfide bond can partially replace the phaseolin vacuolar sorting signal." Plant J **61**(5): 782-791.

Popov, N., M. Schmitt, S. Schulzeck and H. Matthies (1974). "Reliable micromethod for determination of the protein content in tissue homogenates." Acta biologica et medica Germanica **34**(9): 1441-1446.

Pourcher, M., M. Santambrogio, N. Thazar, A.-M. Thierry, I. Fobis-Loisy, C. Miège, Y. Jaillais and T. Gaude (2010). "Analyses of sorting nexins reveal distinct retromer-subcomplex functions in development and protein sorting in *Arabidopsis thaliana*." The Plant Cell **22**(12): 3980-3991.

Puerta, M. V. S., T. R. Bachvaroff and C. F. Delwiche (2005). "The complete plastid genome sequence of the haptophyte *Emiliania huxleyi*: a comparison to other plastid genomes." DNA Research **12**(2): 151-156.

Quigley, F., J. M. Rosenberg, Y. Shachar-Hill and H. J. Bohnert (2002). "From genome to function: the Arabidopsis aquaporins." Genome Biol **3**(1): 0001-0001.0017.

Rachel, R., C. Meyer, A. Klingl, S. Guerster, T. Heimerl, N. Wasserburger, T. Burghardt, U. Küper, A. Bellack and S. Schopf (2010). "Analysis of the ultrastructure of archaea by electron microscopy." Methods in cell biology **96**: 47-69.

Regon P., Panda P., Kshetrimayum E., Panda S.K. (2014). "Genome-wide comparative analysis of tonoplast intrinsic protein (TIP) genes in plants" **14**: 617-629.

Rayon, C., P. Lerouge and L. Faye (1998). "The protein N-glycosylation in plants." Journal of Experimental Botany **49**(326): 1463-1472.

Reyes-Prieto, A., J. D. Hackett, M. B. Soares, M. F. Bonaldo and D. Bhattacharya (2006). "Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions." Curr Biol **16**(23): 2320-2325.

Reyes-Prieto, A., A. Moustafa and D. Bhattacharya (2008). "Multiple genes of apparent algal origin suggest ciliates may once have been photosynthetic." Curr Biol **18**(13): 956-962.

Reyes, F. C., R. Buono and M. S. Otegui (2011). "Plant endosomal trafficking pathways." Curr Opin Plant Biol **14**(6): 666-673.

Reynolds, G. D., B. August and S. Y. Bednarek (2014). "Preparation of enriched plant clathrin-coated vesicles by differential and density gradient centrifugation." Methods Mol Biol **1209**: 163-177.

Rigal, A., S. M. Doyle and S. Robert (2015). Live Cell Imaging of FM4-64, a Tool for Tracing the Endocytic Pathways in Arabidopsis Root Cells. Plant Cell Expansion, Springer**:** 93-103.

Ritzenthaler, C., A. Nebenführ, A. Movafeghi, C. Stussi-Garaud, L. Behnia, P. Pimpl, L. A. Staehelin and D. G. Robinson (2002). "Reevaluation of the effects of brefeldin A on plant cells using tobacco Bright Yellow 2 cells expressing Golgi-targeted green fluorescent protein and COPI antisera." The Plant Cell **14**(1): 237-261.

Rivera-Serrano, E. E., M. F. Rodriguez-Welsh, G. R. Hicks and M. Rojas-Pierce (2012). "A small molecule inhibitor partitions two distinct pathways for trafficking of tonoplast intrinsic proteins in Arabidopsis." PLoS One **7**(9): e44735.

Robinson, D. G., G. Hinz and S. E. Holstein (1998). "The molecular characterization of transport vesicles." Plant molecular biology **38**(1-2): 49-76.

Robinson, D. G., L. Jiang and K. Schumacher (2008). "The endosomal system of plants: charting new and familiar territories." Plant Physiology **147**(4): 1482-1492.

Robinson, D. G., P. Pimpl, D. Scheuring, Y.-D. Stierhof, S. Sturm and C. Viotti (2012). "Trying to make sense of retromer." Trends in plant science **17**(7): 431-439.

Robinson, M. S. and J. S. Bonifacino (2001). "Adaptor-related proteins." Current Opinion in Cell Biology **13**(4): 444-453.

Rockwell, N. C., J. C. Lagarias and D. Bhattacharya (2014). "Primary endosymbiosis and the evolution of light and oxygen sensing in photosynthetic eukaryotes." Front Ecol Evol **2**(66).

Rodriguez-Ezpeleta, N., H. Brinkmann, S. C. Burey, B. Roure, G. Burger, W. Loffelhardt, H. J. Bohnert, H. Philippe and B. F. Lang (2005). "Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes." Curr Biol **15**(14): 1325-1330.

Rouillé, Y., W. Rohn and B. Hoflack (2000). Targeting of lysosomal proteins. Seminars in cell & developmental biology, Elsevier.

Saito, C., T. Ueda, H. Abe, Y. Wada, T. Kuroiwa, A. Hisada, M. Furuya and A. Nakano (2002). "A complex and mobile structure forms a distinct subregion within the continuous vacuolar membrane in young cotyledons of Arabidopsis." The Plant Journal **29**(3): 245-255.

Sanger, F., S. Nicklen and A. R. Coulson (1977). "DNA sequencing with chain-terminating inhibitors." Proceedings of the National Academy of Sciences **74**(12): 5463-5467.

Sanmartin, M., A. Ordonez, E. J. Sohn, S. Robert, J. J. Sanchez-Serrano, M. A. Surpin, N. V. Raikhel and E. Rojo (2007). "Divergent functions of VTI12 and VTI11 in trafficking to storage and lytic vacuoles in *Arabidopsis*." Proc Natl Acad Sci U S A **104**(9): 3645-3650.

Santos, S. R., D. J. Taylor, R. A. Kinzie Iii, M. Hidaka, K. Sakai and M. A. Coffroth (2002). "Molecular phylogeny of symbiotic dinoflagellates inferred from partial chloroplast large subunit (23S)-rDNA sequences." Molecular phylogenetics and evolution **23**(2): 97-111.

Schüssler, M. D., E. Alexandersson, G. P. Bienert, T. Kichey, K. H. Laursen, U. Johanson, P. Kjellbom, J. K. Schjoerring and T. P. Jahn (2008). "The effects of the loss of TIP1; 1 and TIP1; 2 aquaporins in *Arabidopsis thaliana*." The Plant Journal **56**(5): 756-767.

Schellmann, S. and P. Pimpl (2009). "Coats of endosomal protein sorting: retromer and ESCRT." Curr Opin Plant Biol **12**(6): 670-676.

Scheuring, D., C. Viotti, F. Krüger, F. Künzl, S. Sturm, J. Bubeck, S. Hillmer, L. Frigerio, D. G. Robinson and P. Pimpl (2011). "Multivesicular bodies mature from the trans-Golgi network/early endosome in *Arabidopsis*." The Plant Cell **23**(9): 3463-3481.

Schleiff, E. and T. Becker (2011). "Common ground for protein translocation: access control for mitochondria and chloroplasts." Nature Reviews Molecular Cell Biology **12**(1): 48-59.

Schnable, P. S., D. Ware, R. S. Fulton, J. C. Stein, F. Wei, S. Pasternak, C. Liang, J. Zhang, L. Fulton and T. A. Graves (2009). "The B73 maize genome: complexity, diversity, and dynamics." science **326**(5956): 1112-1115.

Schoberer, J., U. Vavra, J. Stadlmann, C. Hawes, L. Mach, H. Steinkellner and R. Strasser (2009). "Arginine/lysine residues in the cytoplasmic tail promote ER export of plant glycosylation enzymes." Traffic **10**(1): 101-115.

Seaman, M. N. (2004). "Cargo-selective endosomal sorting for retrieval to the Golgi requires retromer." The Journal of cell biology **165**(1): 111-122.

Seaman, M. N. (2005). "Recycle your receptors with retromer." Trends Cell Biol **15**(2): 68-75.

Seaman, M. N., M. E. Harbour, D. Tattersall, E. Read and N. Bright (2009). "Membrane recruitment of the cargo-selective retromer subcomplex is catalysed by the small GTPase Rab7 and inhibited by the Rab-GAP TBC1D5." Journal of cell science **122**(14): 2371-2382.

Sears, B. B., L. L. Stoike and W.-L. Chiu (1996). "Proliferation of direct repeats near the Oenothera chloroplast DNA origin of replication." Molecular biology and evolution **13**(6): 850-863.

Seidel, T., D. Schnitzer, D. Golldack, M. Sauer and K. J. Dietz (2008). "Organelle-specific isoenzymes of plant V-ATPase as revealed by in vivo-FRET analysis." BMC Cell Biol **9**: 28.

Shalchian-Tabrizi, K., M. Skånseng, F. Ronquist, D. Klaveness, T. R. Bachvaroff, C. F. Delwiche, A. Botnen, T. Tengs and K. S. Jakobsen (2006). "Heterotachy processes in rhodophyte-derived

secondhand plastid genes: Implications for addressing the origin and evolution of dinoflagellate plastids." Molecular biology and evolution **23**(8): 1504-1515.

Shimada, T., K. Fuji, K. Tamura, M. Kondo, M. Nishimura and I. Hara-Nishimura (2003). "Vacuolar sorting receptor for seed storage proteins in Arabidopsis thaliana." Proc Natl Acad Sci U S A **100**(26): 16095-16100.

Shinozaki, K., M. Ohme, M. Tanaka, T. Wakasugi, N. Hayashida, T. Matsubayashi, N. Zaita, J. Chunwongse, J. Obokata and K. Yamaguchi-Shinozaki (1986). "The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression." The EMBO journal **5**(9): 2043.

Simoes, I. and C. Faro (2004). "Structure and function of plant aspartic proteinases." Eur J Biochem **271**(11): 2067-2075.

Sommer, M. S., S. B. Gould, P. Lehmann, A. Gruber, J. M. Przyborski and U. G. Maier (2007). "Der1-mediated preprotein import into the periplastid compartment of chromalveolates?" Mol Biol Evol **24**(4): 918-928.

Song, J., M. H. Lee, G. J. Lee, C. M. Yoo and I. Hwang (2006). "*Arabidopsis* EPSIN1 plays an important role in vacuolar trafficking of soluble cargo proteins in plant cells via interactions with clathrin, AP-1, VTI11, and VSR1." Plant Cell **18**(9): 2258-2274.

Stegemann, S., S. Hartmann, S. Ruf and R. Bock (2003). "High-frequency gene transfer from the chloroplast genome to the nucleus." Proc Natl Acad Sci U S A **100**(15): 8828-8833.

Stierhof, Y.-D., C. Viotti, D. Scheuring, S. Sturm and D. G. Robinson (2013). "Sorting nexins 1 and 2a locate mainly to the TGN." Protoplasma **250**(1): 235-240.

Stoebe, B. and U. G. Maier (2002). "One, two, three: nature's tool box for building plastids." Protoplasma **219**(3-4): 123-130.

Stork, S. (2013). "Investigation of the proteome of the periplastidal membrane and of Sec61-dependent membrane protein import into the complex plastid of the diatom *Phaeodactylum tricornutum*." Dissertation.

Stork, S., J. Lau, D. Moog and U. G. Maier (2013). "Three old and one new: protein import into red algal-derived plastids surrounded by four membranes." Protoplasma **250**(5): 1013-1023.

Strasser, R. (2014). "Biological significance of complex N-glycans in plants and their impact on plant physiology." Front Plant Sci **5**: 363.

Strasser, R., J. S. Bondili, U. Vavra, J. Schoberer, B. Svoboda, J. Glössl, R. Léonard, J. Stadlmann, F. Altmann and H. Steinkellner (2007). "A unique β1, 3-galactosyltransferase is indispensable for the biosynthesis of N-glycans containing Lewis a structures in *Arabidopsis thaliana*." The Plant Cell **19**(7): 2278-2292.

Strasser, R., J. Mucha, L. Mach, F. Altmann, I. Wilson, J. Glössl and H. Steinkellner (2000). "Molecular cloning and functional expression of β1, 2-xylosyltransferase cDNA from *Arabidopsis thaliana*." Febs Letters **472**(1): 105-108.

Swanson, S. J., P. C. Bethke and R. L. Jones (1998). "Barley aleurone cells contain two types of vacuoles: characterization of lytic organelles by use of fluorescent probes." The Plant Cell **10**(5): 685-698.

Sze, H., K. Schumacher, M. L. Müller, S. Padmanaban and L. Taiz (2002). "A simple nomenclature for a complex proton pump: VHA genes encode the vacuolar H+-ATPase." Trends in plant science **7**(4): 157-161.

Takenaka, M., A. Zehrmann, D. Verbitskiy, B. Hartel and A. Brennicke (2013). "RNA editing in plants and its evolution." Annu Rev Genet **47**: 335-352.

Takeuchi, M., T. Ueda, N. Yahara and A. Nakano (2002). "Arf1 GTPase plays roles in the protein traffic between the endoplasmic reticulum and the Golgi apparatus in tobacco and *Arabidopsis* cultured cells." The Plant Journal **31**(4): 499-515.

Takishita, K., K.-i. Ishida and T. Maruyama (2004). "Phylogeny of nuclear-encoded plastid-targeted GAPDH gene supports separate origins for the peridinin-and the fucoxanthin derivative-containing plastids of dinoflagellates." Protist **155**(4): 447-458.

Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar (2013). "MEGA6: molecular evolutionary genetics analysis version 6.0." Molecular biology and evolution **30**(12): 2725-2729.

Tanikawa, N., H. Akimoto, K. Ogoh, W. Chun and Y. Ohmiya (2004). "Expressed Sequence Tag Analysis of the Dinoflagellate *Lingulodinium polyedrum* During Dark Phase." Photochemistry and photobiology **80**(1): 31-35.

Teich, R., S. Zauner, D. Baurain, H. Brinkmann and J. Petersen (2007). "Origin and distribution of Calvin cycle fructose and sedoheptulose bisphosphatases in plantae and complex algae: a single secondary origin of complex red plastids and subsequent propagation via tertiary endosymbioses." Protist **158**(3): 263-276.

Tse, Y. C., B. Mo, S. Hillmer, M. Zhao, S. W. Lo, D. G. Robinson and L. Jiang (2004). "Identification of multivesicular bodies as prevacuolar compartments in Nicotiana tabacum BY-2 cells." Plant Cell **16**(3): 672-693.

Uehlein, N. and R. Kaldenhoff (2008). "Aquaporins and plant leaf movements." Ann Bot **101**(1): 1-4.

Uenishi, Y., Y. Nakabayashi, A. Tsuchihira, M. Takusagawa, K. Hashimoto, M. Maeshima and K. Sato-Nara (2014). "Accumulation of TIP2; 2 aquaporin during dark adaptation is partially PhyA dependent in roots of *Arabidopsis* seedlings." Plants **3**(1): 177-195.

Van Damme, D., A. Gadeyne, M. Vanstraelen, D. Inze, M. C. Van Montagu, G. De Jaeger, E. Russinova and D. Geelen (2011). "Adaptin-like protein TPLATE and clathrin recruitment during plant somatic cytokinesis occurs via two distinct pathways." Proc Natl Acad Sci U S A **108**(2): 615-620.

Vergés, M., I. Sebastián and K. E. Mostov (2007). "Phosphoinositide 3-kinase regulates the role of retromer in transcytosis of the polymeric immunoglobulin receptor." Experimental cell research **313**(4): 707-718.

Viotti, C., J. Bubeck, Y. D. Stierhof, M. Krebs, M. Langhans, W. van den Berg, W. van Dongen, S. Richter, N. Geldner, J. Takano, G. Jurgens, S. C. de Vries, D. G. Robinson and K. Schumacher (2010). "Endocytic and secretory traffic in *Arabidopsis* merge in the trans-Golgi network/early endosome, an independent and highly dynamic organelle." Plant Cell **22**(4): 1344-1357.

Vitale, A. and G. Hinz (2005). "Sorting of proteins to storage vacuoles: how many mechanisms?" Trends Plant Sci **10**(7): 316-323.

Vitale, H. H. a. A. (2001). "Vacuolar sorting determinants within a plant storage protein trimer act cumulatively " traffic **2**: 737-741.

Wang, Y., L. Jensen, P. Højrup and D. Morse (2005). "Synthesis and degradation of dinoflagellate plastid-encoded psbA proteins are light-regulated, not circadian-regulated." Proceedings of the National Academy of Sciences of the United States of America **102**(8): 2844-2849.

Wang, Y. and D. Morse (2006). "Rampant polyuridylylation of plastid gene transcripts in the dinoflagellate *Lingulodinium*." Nucleic Acids Res **34**(2): 613-619.

Watanabe, E., T. Shimada, K. Tamura, R. Matsushima, Y. Koumoto, M. Nishimura and I. Hara-Nishimura (2004). "An ER-localized form of PV72, a seed-specific vacuolar sorting receptor, interferes the transport of an NPIR-containing proteinase in *Arabidopsis* leaves." Plant and cell physiology **45**(1): 9-17.

Weber, A. P., M. Linka and D. Bhattacharya (2006). "Single, ancient origin of a plastid metabolite translocator family in Plantae from an endomembrane-derived ancestor." Eukaryot Cell **5**(3): 609-612.

Werner, M., N. Uehlein, P. Proksch and R. Kaldenhoff (2001). "Characterization of two tomato aquaporins and expression during the incompatible interaction of tomato with the plant parasite Cuscuta reflexa." Planta **213**(4): 550-555.

Wilson, I. B. (2002). "Glycosylation of proteins in plants and invertebrates." Current opinion in structural biology **12**(5): 569-577.

Wudick, M. M., D. T. Luu, C. Tournaire-Roux, W. Sakamoto and C. Maurel (2014). "Vegetative and sperm cell-specific aquaporins of *Arabidopsis* highlight the vacuolar equipment of pollen and contribute to plant reproduction." Plant Physiol **164**(4): 1697-1706.

Xi, C. and J. Wu (2011). Control of biofilm formation, Google Patents.

Xiang, L., E. Etxeberria and W. Van den Ende (2013). "Vacuolar protein sorting mechanisms in plants." FEBS J **280**(4): 979-993.

Xiang, L. and W. Van den Ende (2013). "Trafficking of plant vacuolar invertases: from a membrane-anchored to a soluble status. Understanding sorting information in their complex N-terminal motifs." Plant Cell Physiol **54**(8): 1263-1277.

Yamazaki, M., T. Shimada, H. Takahashi, K. Tamura, M. Kondo, M. Nishimura and I. Hara-Nishimura (2008). "*Arabidopsis* VPS35, a retromer component, is required for vacuolar protein sorting and involved in plant growth and leaf senescence." Plant and cell physiology **49**(2): 142-156.

Yang, Y. D., R. Elamawi, J. Bubeck, R. Pepperkok, C. Ritzenthaler and D. G. Robinson (2005). "Dynamics of COPII vesicles and the Golgi apparatus in cultured Nicotiana tabacum BY-2 cells provides evidence for transient association of Golgi stacks with endoplasmic reticulum exit sites." Plant Cell **17**(5): 1513-1531.

Zardoya, R. (2005). "Phylogeny and evolution of the major intrinsic protein family." Biology of the Cell **97**(6): 397-414.

Zaslavskaia, L. A., J. C. Lippmeier, P. G. Kroth, A. R. Grossman and K. E. Apt (2000). "Transformation of the diatom *Phaeodactylum tricornutum* (Bacillariophyceae) with a variety of selectable marker and reporter genes." Journal of Phycology **36**(2): 379-386.

Zauner, S., D. Greilinger, T. Laatsch, K. V. Kowallik and U. G. Maier (2004). "Substitutional editing of transcripts from genes of cyanobacterial origin in the dinoflagellate *Ceratium horridum*." FEBS Lett **577**(3): 535-538.

Zhang, C., G. R. Hicks and N. V. Raikhel (2014). "Plant vacuole morphology and vacuolar trafficking." Front Plant Sci **5**: 476.

Zhang, Z., T. Cavalier-Smith and B. R. Green (2002). "Evolution of dinoflagellate unigenic minicircles and the partially concerted divergence of their putative replicon origins." Mol Biol Evol **19**(4): 489-500.

Zhang, Z., B. Green and T. Cavalier-Smith (1999). "Single gene circles in dinoflagellate chloroplast genomes." Nature **400**(6740): 155-159.

Zhang, Z., B. R. Green and T. Cavalier-Smith (2000). "Phylogeny of ultra-rapidly evolving dinoflagellate chloroplast genes: a possible common origin for sporozoan and dinoflagellate plastids." Journal of Molecular Evolution **51**(1): 26-40.

Zhou, A., Y. Bu, T. Takano, X. Zhang and S. Liu (2015). "Conserved V-ATPase c subunit plays a role in plant growth by influencing V-ATPase-dependent endosomal trafficking." Plant biotechnology journal.

Zimorski, V., C. Ku, W. F. Martin and S. B. Gould (2014). "Endosymbiotic theory for organelle origins." Current opinion in microbiology **22**: 38-48.

Zouhar, J. and E. Rojo (2009). "Plant vacuoles: where did they come from and where are they heading?" Curr Opin Plant Biol **12**(6): 677-684.

# 7 Supplements

## 7.1 Open reading frames with no known homology

### Table S1: All open reading frames mentioned in this study.

Only the open reading frames >= 150 bps were listed in this table. J8-, J9-, J13-, J22-, J24-, J36- ORFs are the open reading frames of empty minicircles in *A. carterae* CCAM0512. J28 ORFs are the open reading frames of *atpB* gene containing minicircle in *A. carterae* CCAM0512. The rests are the open reading frames of empty minicircles in *A. carterae* CCAP 1102/6 and *A. carterae* CS21, they are named by the accession numbers plus the ORF number in the table.

| name | Open reading frames |
|---|---|
| J8 ORF1 | MGDPNIDSMAFNHREGLRYQKCCLRVTKNSGLHHLIGSKEDDHIALQLRSQLSN. |
| J8 ORF2 | MHELVVANTTSPLTTFDRGGFALKQAYSIRGPDPAGVEEHTTRIELLKRTIVLISNYSMIANYA. |
| J8 ORF3 | MAVRQIQLPSGTALIGKGRGCNTAQVITFVRRHTTTNIATLNRWRCPCRAQEGRKTWPFYQ. |
| J8 ORF4 | MTRHHYTELVKSKRQHRCTPPNKQGNYSFSIPTKLEGVRGSTFRPLHSYLPQRRGRDRSLILAYPQNEARTIN LEPNHRVLN. |
| J9 ORF1 | MPRHLPSVGASFLQARLDRSLDSRCTMVSLNDVTSNGLRVLAYFPLNDSTMAGNNVSWSIST. |
| J9 ORF2 | MLSLFVVIIRCHAIKEDRTMLALMPFGLRVYTVVRRLPCPVSSDKGGYERLSCSNNSNLVPALRTNF. |
| J9 ORF3 | MSNLLAISVVSVTISRPLGPNWILSLIHISTIIDEFSRVFSHIGSNSSSVPGDPLESTCRGGGESHVVSQNL. |
| J9 ORF4 | MIYFYLVQCNIRDFETQRGFPPPPCRSTLEDPRVPSSNSSQYARTPEKIHR. |
| J9 ORF5 | MVEMCIRDRIQLGPRGRDIVTDTTLIARRLLIHYRQQLSLHRRGNYVGLQTSDCRRHRPHLSMC. |
| J9 ORF6 | MYLPSSTVLYLGAFHLICLSISVVSTFIFVPSSTFLFDPSELSGESPDLCANYHLIPFTSSSSSFSRSHHSHSYGLTG IF. |
| J9 ORF7 | MIVSRSTKNLCVKRVRGWSCCCRIVFHNLPCQKTQGKVDVLQQCRHEDQMALEPTSFYPL. |
| J9 ORF8 | MQTGIEDVYHFLMKVNTCYSVIEITNGLHSIGQEYSWLVLRYQDRYPGVQHTQ. |
| J9 ORF9 | MATSLITCHQTREDIIDPIRTKRTRYRHRHDTDSEEVAHSLSPTTLIASSR. |
| J13 ORF1 | MEYVEWLSSNATGMACWLVWRCRSRVMMPCLQGCYPSHVVAYQSRTINSCE. |
| J13 ORF2 | MLLRSYLAKHSLRRYVNSLGLPISYTEPGTVLVSSSSFLLRKPKYGLVRLASLNPTLALNVVQVNALNHAALLS NSLVRSIEDGCNPTIVFQSVMGTLKRSNRVLGVKLELRALVLAPFARHRVWNYGYGKTNKAVARIKDVGKA SLVTYMGLISVSTAISSG. |
| J13 ORF3 | MSPQRSIAPFQVICLSAPLPSIEGLSLSFHQLSLHSFVYSLVEILTSRHI. |
| J22 ORF1 | MSHDIITTTPNPLSFIGGGLIKVKSLWPMRGAIHQLQYLVHSPSYRRSSEH. |
| J22 ORF2 | MSPQRSIAPFQVICLSAPLPSIEGLSLSFHQLSLHSFVYSLVEILTSRHI. |
| J22 ORF3 | MVLKNIPDCTNKVTDRHSIVPAYQDEYLDPSTKHEPRVPFLTSGVNLTRYKDATSVLWQLARCD. |
| J24 ORF1 | MLIVLTCVPFREVYATYLFQFLYGSHYSLLHGNSVPLVYVISITSVALTEDCLSINI. |
| J24 ORF2 | MNDITNFIRVLNSYSTFSLEAFVPCVELPDPRLNSTNRSFSRHSVFVLVCDFLTPMD. |
| J24 ORF3 | MDNMMSCLVGASAYPHWQLSQSGSPLCVKTTRFLLIPRTVYLNWWYYGTVGL. |
| J24 ORF4 | MSHDIITTTPNPLSFIGGGLIKVKSLWPMRGAIHQLQYLVHSPSYRRSSEH. |
| J24 ORF5 | MNDILGNLKLTDNRFHKDTSFKKCRSIHRSEKITDENEDRMTGERTISRVKTWIWQFYTRHKGLQRERRIRI QDSNKVSYVVHTDRV. |
| J24 ORF6 | MSSDGETELHHVWEDVYDTLHPVVTIQDVPAVAVIHLWQALCNVIRCKAGLGINP. |
| J36 ORF1 | MSPQRSIAPFQVICLSAPLPSIEGLSLSFHQLSLHSFVYSLVEILTSRHI. |
| J36 ORF2 | MGRDTHLGKRVQWNAYPSPCWCNQVCIFRTVQLYESSYSVNCDQRYWSSVSSFPSPYGTVGVVRR. |
| J28 ORF1 | MIYFYLVQCNIRDFETQRGFPPPPCRSTLEDPRVPSSNSSQYARTPEKIHR. |
| J28 ORF2 | MRDGSVLASECEAFVWDLFYMESFVSMLHSVFKGIAPTPNKEALDLISLELEYCASNELSLALASSGLFIKSYA NALIAEVQQIAYGGILRAVALAGTDGLDLVSTYGHLTYQPLVVPVGRVCQGRILNCVGAPMDAYDDIVISAA YSSVESPASVVLNALWYGGTASSSPSKLTTPLAHYDQSFANAAPIHKLSLIHISTIIDEFSRVFSHIGSNSSSVPG DPLESTCRGGGESHVVSQNL. |
| AJ307015 ORF1 | MHELVVANTTSPLTTFDRGGFALKQAYSIRGLLSQRDQRGAYHSH. |
| AJ582641 ORF1 | MAVRQIQLPSGTALIGKGL. |

| | |
|---|---|
| DQ507216 ORF1 | MLLRSYLAKHSLRRYVNSLGLPISFLPGTVLVSSSSFLLRKPKYGLVRLPLSTRLFHPMSVQVNALNHAALLSNS LVRSIEGGCNPTIVFQSVMGTLKRSNRVLGVKLELRALVLAPFARHRVWNYGYGKTNKAVARIKDVGKASLV TYMGLISVSTAISSG. |
| AF401630 ORF1 | MSHDIITTTPIPLSFIGGGLIKVKSLWPMRGAIHQLQYLVHSPSYRRSSEH. |
| AF401630 ORF2 | MNDITNFIRVLNSYFTFLLEAFKPCVELPDPRLNSI. |
| AJ318067 ORF1 | MNDITNFIRVLNSYSTFLLEASEPCVELLDPRS. |
| AJ318067 ORF2 | MSSDSETELHHVWEDVYDTLHPLVTIQDVPAVAVIHLWQALCNVIRCKAGLGINPSISTRPSWALQGSL. |

## 7.2 Identified marker proteins in *P. tricornutum*

### Table S2: Proteins localized as eGFP/ mRFP fusions in this study.

Protein ID from http://genome.jgi-psf.org/Phatr2/Phatr2.home.html. "-" means unknown. **Abbreviations**: **eGFP**, enhanced green fluorescent protein; **mRFP**, monomeric red fluorescent protein; **Tip1-5**, Tonoplast intrinsic protein 1-5; **GnTI**, N-acetylglucosaminyltransferase I; **XylT**, β1,2-xylosyltransferase; **FucT**, α1,3-fucosyltransferase; **Vps26**, Vacuolar protein sorting 26; **Vps29**, Vacuolar protein sorting 29; **ATPase1-2**, Vacuolar type $H^+$-ATPase 1-2.

| Pt-Protein ID | eGFP/ mRFP fusion protein | Potential compartment |
|---|---|---|
| 31553 | Tip1-eGFP | cytoplasmic membrane, endocytic vesicles |
| - | Tip2-eGFP | Vacuolar-like membrane |
| 20755 | Tip3-eGFP/ mRFP | ER membrane (cER, nuclear envelope, host ER) |
| 19409 | Tip4-eGFP | periplastidal membrane |
| 43157 | Tip5-eGFP | ER membrane (cER, nuclear envelope, host ER) |
| 54844 | GnT1-eGFP | cis Golgi |
| 45496 | XylT-eGFP/ mRFP | medial Golgi |
| 54599 | FucT-eGFP | Trans golgi network |
| 41962 | Vps26-eGFP | Trans golgi network |
| 17936 | Vps29-eGFP/ mRFP | Trans golgi network |
| 21882 | ATPase1-eGFP | lytic vacuole? A second type of vacuole |
| 44050 | ATPase2-eGFP | cytosol |

## 7.3 The Tip2 eGFP fusion protein

The gene model of Tip2 is not supported completely by ESTs, so the gene was amplified from cDNA. The gene model and all ESTs were aligned by ClustalX2.1 (Fig. S1). Together with this the protein sequence of Tip2 is different from the genome database (Protein ID: 1370), the sequence was shown [1].

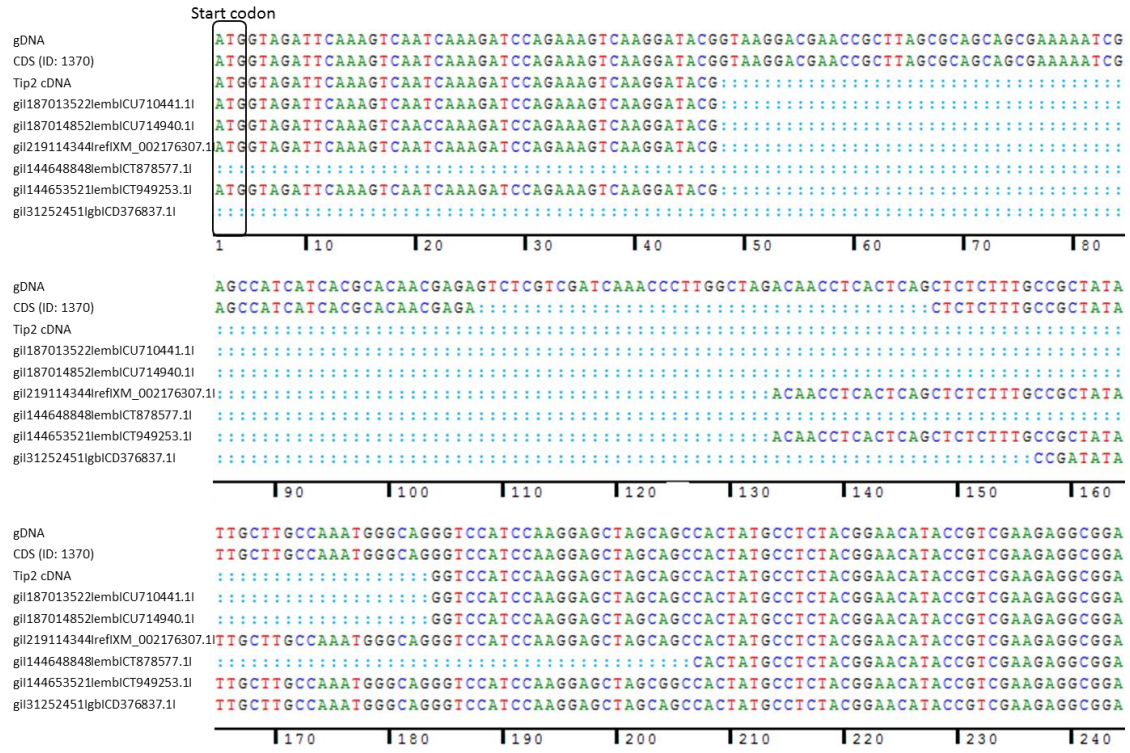


**Fig. S1: The alignment of Tip2 nucleotide sequences in *P. tricornutum*.**

It was shown that the 5' terminus of Tip2 gene model is not completely supported by ESTs. Tip2 gene was amplified from cDNA. Only the part is not completely supported by ESTs was shown in this figure. The protein sequence of Tip2_FL_eGFP was shown [1].

[1] Tip2-eGFP fusion protein sequence used for *in vivo* localization studies in *P. tricornutum*.

>Tip2-eGFP

MVDSKSIKDPESQGYGSIQGASSHYASTEHTVEEADAEEGPFPVDLKSMLIAEVFGTCTFVQIGCAANAVALYTHN
STTMTIDWQVGVVWALAMTVAVFLSAALSGAHLNPAVSFSFALARPADFRFRKLIPYWAAQLGGALLAGIINLFLF
HQAISHYEKKMAIVPGAAGSIQSAAAFGCYWSLNSKYISNGVHAFFIEAFGTGVLVFCIFAATHIKNPLPGVAVPPII
GAMYGILVVTLGPMTGGSFNPVRDMGPRIVSVIGHWGPTALTNFLPYLLGPMIGGPIGAFLADKVLML:eGFP.

## 7.4 Sequences of all used oligonucleotides

Cutting sites of restriction enzymes are marked in red.

| Protein name | Primer name | sequence |
|---|---|---|
| XylT | Golgi pro-MunI-for | CAATTGATGGCGTTTTTGCCGAATCG |
| | Golgi pro-BglII-rev | AGATCTGAATAAGGTACTCTTAATTGTCC |
| Vps26 | Vps26-MunI-for | CAATTGATGAACGTTGGATCTTTACTTGG |
| | Vps26-BglII-rev | AGATCTCCCCAAGTCTTTCCGCCATAAG |
| Vps29 | Vps29-EcoRI-for | GAATTCATGGCCAATTTTGGGGAGCTTG |
| | Vps29-BglII-rev | AGATCTTGTCAAGAGCGAAGCCATAAGG |
| v-ATPases1 | v-ATPases1-for | GAATTCATGAGTGTCGAAATGGAAACTTGC |
| | v-ATPases1-rev | GGATCCGTTGTTCCCCTCGCACACGAA |
| v-ATPases2 | v-ATPases2-EcoRI-for | GAATTCATGGCGGAATCAGGTACCGGC |
| | v-ATPases2-BamHI-rev | GGATCCGACCGTCCCGGCCATCTTTTGCG |

| | | |
|---|---|---|
| TIP1 | TIP1-for1-Xba1 | TCTAGAATGGCTTCCATTATCAACAT |
| | TIP1-rev1-BamHI | GGATCCGGCCTGAGGCGCGGCGTTTTCC |
| TIP2 | EcoRI-TIP2-for | GAATTCATGGTAGATTCAAAGTCAATCA |
| | XbaI-TIP2-rev | TCTAGATAGCATTAGCACTTTGTCGGCC |
| TIP3 | EcoRI-TIP3-for2 | GAATTCATGGTTGAGTACGGTGAGTTCGC |
| | TIP3/Xba1-rev | TCTAGACGCCATGAGCAGGCGGTCCGC |
| TIP4 | Aqua2_3_BglII_neu | AGATCTGGCGCCACCGTACAAAAC |
| | Aqua2_5_EcoRI | GAATTCATGGGGCGCCGTTGGTTG |
| TIP5 | TIP5-for1-Xba1 | TCTAGAATGGTCAAGGACTACGTCGAAGC |
| | TIP5-rev1-BamHI | GGATCCGTTAGTCTTCTTGGCAGTGGTC |
| GnT1 | GnT1-for-MunI | CAATTGATGCGGTTGTGGAAACG |
| | GnT1-rev-BamHI | GGATCCTCTTTTCGGTGACGGAA |
| FucT | FucT-for-SacI | GAGCTCATGTCACTTCGCAAG |
| | FucT-rev-BglII | AGATCTCGGATCGAACTTCCA |
| SA-GFP colony PCR | P.t.MGDG1_Bam_r | GGATCCGGCACTCCCGAGATCAGTG |
| | PDI_rv_BglII | AGATCTCAATTCGCCTTCATCAAAAAGATCC |
| pJet seq | pJet-uni | CTCTCAAGATTTTCAGGCTGTAT |
| | pJet-rev | GCACAAGTGTTAAAGCAGTT |
| pPha-Dual seq | pPha_5'vorNdeI | GCTTAACTATGCGGCATCAG |
| | pPhaDual_seq_MCS_EcoHind_for | GGACATATTGTCGTTAGAACGCGG |
| | pPhaDual_seq_MCS_EcoHind_rev | GTCTTATCCAGGTCCAAACAGATTG |
| | pPhaDual_seq_MCS_SpeSac_rv | CTAACGCAGCTTAGACATAAAC |
| pPha-NR seq | pPha-NR-for | GGTCGGGTTTCGGATCCTTCC |
| | pPha-NR-rev | GATGAACATAAAACGACGATGAG |
| eGFP | eGFP-for-BamHI | GGATCCATGGTGAGCAAGGGCGAG |
| | eGFP-rev-HindIII | AAGCTTTTACCTGTACAGCTCGTCCATG |
| | eGFP-for-XbaI | TCTAGAATGGTGAGCAAGGGCGAG |
| | eGFP-rev-5' | CGTCTCCCATGGCTCTGATTTCCCGATTTGG |
| Sa-GFP | Sa-GFP-1-10-fw-EcoRI | GAATTCATGGGTGGCACTAGTAGC |
| | Sa-GFP-1-10-rv-BamHI | GGATCCGGTACCCTTTTCGTTGGG |
| | Sa-GFP-11-fw-SpeI | ACTAGTCGTACTGGGCGAAAGC |
| | Sa-GFP-11-rv-XbaI | TCTAGAGGATCCGCCACCAGACC |
| Minicircles | Transposon forward PCR | ATTCAGGCTGCGCAACTGT |
| | Transposon reverse PCR | GTCAGTGAGCGAGGAAGCGGAAG |
| | J25-rev2-primer | GCTGCCAATGAGGGTACGCGGA |
| | J25-for2-primer | TCCGCGTACCCTCATTGGCAGC |
| | J30i2 | CAAGATCAATAAATTAGTAATGGTG |
| | J14i2 | GTTGTTTCCTTATTTGAGGGGCAG |
| | J33i2 | CGCAACCAGATATTAGCCTATACG |
| | petD_for AC | CCCTTTTGGATTAATGGTTG |
| | petD_for2 AC | TCAGGACATATTGGTCTTCC |
| | petD_rev AC | AAGTAGCATTACACGAATGG |
| | atpB rev out AC | TGTAGAATAGGTCCCAGACG |
| | atpB for out AC | TGATGGTATCCTTACAGGTC |
| | atpB for2 AC | AACCTTCGGTCTATGACTCC |
| Minicircles | core out AC_for2 | ACTCCGGGTCAATCGTTTCC |
| | petB_for2_AC | TAGGCCAGAAGTATACCAGG |
| | petB_rev2_AC | ATAATGCCTCGGGAGAATAG |
| | petB_for out_AC | TTTCCTAATGATTCGTAAGC |
| | AC_T5_for | ATACCCTTCGTTATCCTTCG |
| | AC_T6_for | TTGCTGCGTTCGTATCTTGC |
| | AC_T10_for | TTTAACACCTTTCGCCCTCG |
| | AC_T3_for | AAACACTCGTGCCCCATCAG |
| | AC_T8_for | AAGAGGATTAGGGGTTGTGG |
| | AC_T16_for | TTACGCGATACAGCTCATCC |

| | | |
|---|---|---|
| | AC_T14_for | ATCGTGGATGAGCTGTATCG |
| | core out AC_rev | TGGAAACGATTGTCGGTGAG |
| | AC_T13_for | TGTGATGAGGTCTGTAGTGG |
| | AC_T13_rev | TGTGATGAGGTCTGTAGTGG |
| | TP sequencing-for | CGTACTATCAACAGGTTGAACTGCC |
| | TP sequencing-rev | GAGCCAATATGCGAGAACACCCGAG |
| | AC_T12_for | ACTCGATGACCTCAACCTTG |
| | AC_T12_rev | TGATGAGGTGCCTGACAAGC |
| | AC_T17_for | ACCTGATTGCACCAACAAGG |
| | AC_T17_rev | ATGGGATTGTATCAGGGGAC |
| | J12-sequencing primer for1 | GTTCCCAGATAAGGGAATTAGGGTTC |
| | J12-sequencing primer rev1 | CTCATCGAAGACAGCGGTAGACG |
| | psaB-rev-AC | TTGAAGGAGAGATCCATACC |
| | J2-sequencing primer for1 | GCAGAGACTGCTGGTTCTGAGTCCCT |
| | J2-sequencing primer rev1 | CCCTTGTATTACTGTTTATGTAAGC |
| | J28i3 primer | AACCTTCGGTCTATGACTCC |
| | J32i primer for | ATAATGCCTCGGGAGAATAG |
| | J32i primer rev | TAGGCCAGAAGTATACCAGG |

# 8 Acknowledgements

First and foremost, I would like to express my deepest gratitude to Dr. Stefan Zauner and Prof. Dr. Uwe G. Maier for giving me the Ph.D position, for their excellent guidance, caring, patience, and providing me with an excellent platform for finishing the Ph.D study. I would never have been able to finish my research and writing without them. Besides my supervisors, I would also like to thank the rest of my committee members Prof. Ralf Jacob and Prof. Andrea Maisner for their insightful comments and encouragement for the past several years.

I gratefully acknowledge financial support from the Deutsche Forschungsgemeischaft in the RTG 1216 “Intra- and Intercellular Transport and Communication” (IITC).

I would like to thank Prof. Muriel Bardor and Clément Ovide for friendly giving me two constructs (GnTI and FucT). I would also like to thank Dr. Thomas Heimerl and Dr. Kathrin Bolte for doing the work on the electron microscope.

Thanks to Heidi Thierfelder for her constant willingness to culture dinoflagellates for me. Thanks to Lucette Claudet for her friendly help in dealing with administrative stuff.

Furthermore, I would like to thank Dr. Christopher Grosche, Dr. Simone Stork and Dr. Julia Lau for giving me help and suggestions in lab. I would also like to thank Dr. Stefan Zauner, Dr. Christopher Grosche, Dr. Julia Lau, Jonny Gentil, Viktoria Schreiber and Christian Mayer who patiently corrected my writing. Without them this work would not have been possible.

Special thanks goes to Angela Zimmer, Kamilla, Jonny Gentil, Dr. Simone Stork, Eike M. Trapp, Dr. Julia Lau and all other students of AG Maier for the help in my German life.

Last but not the least, I would like to thank my family: my parents, my brother and my husband for the unceasing encouragement, attention, love and support.

# 9 Curriculum vitae

**Personal Data:**

**Name:** Xiaojuan Liu

**Born:** 12/07/1985

**In:** Wuping, Longyan, Fujian, China

**Education:**

| | |
|---|---|
| 09/1993-06/1999 | Primary school |
| 09/1999-06/2002 | Middle school |
| 09/2002-06/2005 | High school |
| 09/2005-06/2009 | Fujian Normal University<br>Biology Department<br>Bachelor thesis in the lab of Prof. Yi Zheng |
| 09/2009-06/2012 | Fujian Normal University<br>Biology Department, Aquatic Biology<br>Master thesis in the lab of Prof. Huiru Zhuang |
| 09/2012-10/2015 | Philipps University of Marburg<br>Biology Department, Cell Biology<br>Ph.D. thesis in the lab of Prof. Dr. Uwe G. Maier and Dr. Stefan Zauner |

# 10 Erklärung

Ich versichere, dass ich meine Dissertation

**“Genetic compartmentalization in the complex plastid of Amphidinium carterae and The endomembrane system (ES) in Phaeodactylum tricornutum”**

Selbständig, ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich beyeichneten Quellen und Hilfen bedient habe. Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, 30/08/2015 Xiaojuan Liu