The plastid of *RM12*, the photosynthetic ancestor of Apicomplexa

Jan Janouškovec

master thesis, 2008

University of South Bohemia, Czech Republic

Supervisor: Ing. Miroslav Oborník, PhD.

Janouškovec J. (**2008):** The plastid of *RM12*, the photosynthetic ancestor of Apicomplexa. Master thesis. University of South Bohemia, Faculty of Sciece, Czech Republic.

Annotation: The discovery of a novel alveolate *RM12* enabled to trace origin of parasites of the group Apicomplexa and particularly their plastid called the apicoplast. In this work, I sequenced four genes encoded in *RM12*'s plastid genome and inferred their phylogenies. Further, I used the method of Southern hybridization to assess size of the plastid genome. A close relationship between the apicoplast and *RM12*'s plastid was found. The size of the plastid genome was estimated as 130 kb. The results confirm the apicoplast originated from photosynthetic plastid similar to the plastid of *RM12*. This finding may have significant consequences in research of apicoplast evolution and function.

Acknowledgements: I am greatful to my supervisor Míra for the chance to work on this project and attend intenational conferences, for his support and advices and for his friendly attitude. I thank all the superb people in my laboratory namely Ogar, Brabčák, Luďa, Katcha, Tomík, Dáša, Káťa, Anička and others. I thank to Jula, Milan and all the other people from the academy who did not hesitate to advice me. I thank to all my friends at home as well as in České Budějovice. I especially thank to my mom, dad, Ráďa and the rest of my family, who have understood my scientific deviation and given me the greatest support of all. Finally, I thank my Evelyn for her loving heart.

Prohlášení: Prohlašuji, že svoji diplomovou jsem vypracoval/a samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury. Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své diplomové práce, v úpravě vzniklé vypuštěním vyznačených částí archivovaných Přírodovědeckou fakultou, elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách.

Declaration: I hereby declare that I have worked out this master thesis all by myself using the cited literature. I hereby declare that I agree with publication of my master thesis in electronic form in reduced version (deleted parts will be archived by BF JCU) according to the § 47b of the law no. 111/1998 Sb. Thesis will be placed in the public part of the STAG database run by the University of South Bohemia in České Budějovice on its web site.

....., České Budějovice .....

datum (date) podpis studenta (signature)

Contents

Introduction

1/ Primary endosymbiosis

Many unicellular eukaryotes contain a photosynthetic organelle termed the plastid. The plastid is, similarly to mitochondria, of prokaryotic ancestry and originates through an engulfment of a cyanobacterium by a heterotrophic eukaryote in the event called the primary endosymbiosis. The event resulted in establishment of a photosynthetic eukaryote diverging in three groups with a primary plastid: Chloplastida = Viridiplantae, Rhodophyta and Glaucocystophyta, all referred to as Archaeplastida (Rodriguez-Ezpelenta, 2005). The plastid of these groups is pigmented with chlorophyll *a* and *b*, surrounded by two membranes and contains a reduced multiplied genome. Recent discovery showed that 'cyanelle' of the thecate amoeba *Paulinella chromatophora* has risen from an independent primary endosymbiosis (Marin et al., 2005). Moreover, it seems the event happened relatively recently (Yoon et al., 2006) giving us a unique image of a plastid genesis – an example of a primary endosymbiosis being under way (Rodríguez-Ezpeleta and Philippe, 2006).

Plastids retain their genomes, which encode part of the genes associated with its metabolism. Many genes in extant plastids have been translocated to the cell nucleus. These genes are expressed under the control of the host in cytoplasm and retargeted to the plastid as translated products. The obstacle to get through the plastid membranes is solved out by creation of specific targeting sequences presented at the 5'end of maturing RNAs.

2/ Secondary endosymbiosis

Secondary plastids are of green and red algal ancestry

Plastid of many eukaryotes is of secondary origin. They arose through phagocytosis of a primary plastid-containing alga by another eukaryote. The secondary endosymbiosis has happened at least three times during the organism evolutionary history resulting in independent plastid acquisitions in chromalveolates, euglenophytes and chlorarachniophytes (Bhattacharya et al., 2004; Yoon et al., 2004. The latter two groups represented by *Euglena gracilis* and *Bigelowiella natans* plastids are both of green algal origin and therefore were hypothesized to originate in the same endosymbiotic event (the cabozoan theory of their origin, see Cavalier-Smith 00). However, it was showed upon the completion of both plastid genomes sequencing that they arose in independent endosymbioses and are not specifically related (Rogers et al. 07). Both plastid are, similarly to chlorophytes, chlorophyll *a* and *b*-pigmented and surrounded with 3 (*E.gracilis*) and 4 (*B.natans*) membranes. All the other known photosynthetic complex plastids shown red-algal ancestry. Their hosts involve Chromista (Cryptophyta, Haptophyta, Heterokonta=Stramenopila), Apicomplexa and Dinophyta. Their plastids are surrounded by four membranes (three in the plastids of the peridinin dinoflagellates) and contain chlorophyll *a* and c_2 besides additional pigments (Delwiche, 1999)

Katablepharids, an early stage of secondary endosymbiosis?

A newly described endosymbiont of Katablepharidophyta seems to be a case of an early stage

of the secondary endosymbiosis. Katablepharids are distant sisters of Cryptophyta (Okamoto and Inouye., 2005a), which lack plastid and instead, contain a photosynthetic symbiont of prasinophyte (Chlorophyta) origin (Okamoto and Inouye, 2005b). The endosymbiont is incapable of independent existence: it has lost the cytoskeleton and the endomembrane system and its plastid is about ten times larger that of prasinophytes. However, it has not yet become an organelle remaining only in one of the daughter cells originating after the binary cell division (Okamoto and Inouye, 2005b).

The discovery of the apicoplast

The apicoplast is a small non-pigmented secondary plastid found in many apicomplexan parasites. The organelle was first observed in 1960's as an unknown multimembrane structure in *Plasmodium* and other apicomplexans and got many different names. In 1975, Kilejan et al**.** identified previously unknown circular DNA in *Plasmodium*, the apicoplast genome and later its inverted rRNA gene repeats were discovered. It was also shown some apicomplexan species were sensitive to antibiotics, but this was explained as inhibition of the mitochondrion. Although no other protozoan species have been studied so extensively as Apicomplexa the apicoplast was discover only decades after its first observation. It started with the discovery of the DNA, which showed to be extracellular and inherit maternally, but was surprisingly purified separately from the mitochondrion. In situ hybridization showed it may be a relict plastid. Sequencing of the apicoplast genome revealed several plastid genes, but the most intriguing questions about its origin and function remained unresolved. Sequencing of the nuclear genome of *P. falciparum* (Gardner et al., 2002) enabled to identify nuclear-encoded plastid genes and their signal targeting sequences. Several metabolic pathways have been localized (at least partially) in the apicoplast: fatty acid biosynthesis, isoprenoid biosynthesis, heme biosynthesis. These pathways have proved to be essential for apicomplexan parasites. The organelle has become a promising specific drug target, because plastids are obviously missing in cells of animals and human, who suffer from diseases caused by apicomplexan parasites (reviewed in Waller and McFadden, 2004).

The apicoplast characteristics and its phylogeny

Apicoplast is present in haemosporidians, piroplasmids and coccideans. There is no evidence of a relict plastid or its genome in *Cryptosporidium* using PCR and DNA hybridization techniques (Zhu et al., 2000). Gregarines may also lack both the plastid and its genome. Several methods including DNA hybridization, PCR amplification, plastid antibiotics and transmission electron microscopy were used in an attempt to identify a plastid/plastid genome in *Gregarina niphrandrodes* with null result (Toso and Otomo, 2007). Single apicoplast with low DNA content is present in the cells of other apicomplexans. It is clearly bound by four membranes in *Toxoplasma gondii*, but in *P. falciparum* 3-5 membrane numbers have been observed. Lang-Unnash and her collective (1998) found out apicoplasts are widespread in Apicomplexa (other than gregarines and *Cryptosporidium*) and that they are strongly monophyletic. Regarding number of its membranes the ancient apicoplast has been considered to result from a secondary endosymbiotic event. The issue of its origin became a matter of a great debate. Several works proposed a green algal origin of the apicoplast based on phylogenetic analyses of different plastid genes (Kohler et al., 1997; Blanchard, 1999). However, these analyses suffered from poor sampling or lacked an additional analysis of branch supports. Many other features point out the red-algal origin of the apicoplast (e.g. structure of ribosomal operon, analyses of ribosomal protein genes and gene loss and rearrangement analyses). The current consensus is that the

apicoplast originated from a secondary plastid of red-algal ancestry and is close to the peridinin plastid of the dinoflagellates (Keeling, 2004b).

The peridinin and fucoxanthin plastids

There are several types of plastids in dinoflagellates. The most characteristic for dinoflagellates is the peridinin plastid, pigmented with chlorophyll *c* and xanthophyl peridinin as a major karotenoid. It is presumably of red algal origin and arose through secondary endosymbiosis (Cavalier-Smith, 1999). The plastid is bound by three membranes and is outside the endoplasmatic reticulum. Proteins are shuttled to the plastid owing to the tripartite targeting sequence. Several dinoflagellates lack the peridinin pigment and contain a plastid with three membranes, chlorophylls c_1 and c_2 and 19-hexanoyloxy-fucoxanthin as a major carotenoid. Such pigmentation is typical for haptophyte algae. Phylogenetic analysis based on the plastid genes showed the plastid is derived from tertiary endosymbiosis (Tengs, 2000). That view was soon challenged by Yoon et al., who suggested a model in which the unarmored fucoxanthin dinoflagellates were primitive whereas species with peridinin plastid derived taxa within Dinophyta (Yoon et al., 2002). This scenario assumes tertiary endosymbiosis with the haptophyte plastid happened before the radiation of the group. Their analyses based on phylogeny of only two genes were nevertheless shown mislead by codon usage heterogeneity in DNA minicirle gene sequences (Inagaki, 2004). The position of dinoflagellates with fucoxanthin plastid is not yet completely resolved, but they are currently considered as derived from the peridinin dinoflagellates in which the original plastid was replaced by the haptophyte one. The peridinin plastid is currently considered to be the original secondary plastid of all Dinophyta, suggesting the ancestral dinoflagellate was rather photosynthetic. Evidence of plastid genes in the nuclear genome of a non-photosynthetic dinoflagellate supports this hypothesis (Sanchez-Puerta, 2007).

Organization of the dinoflagellate plastid genome

Dinoflagellates has been known to possess many outstanding features, but the discovery of a unique organization of their plastid genomes was yet an unexpected surprise. Zhang et al. (1999) showed plastid genes of a peridinin dinoflagellate *Heterocapsa triquetra* are organized in gene circles (minicircles). One (or two) genes for polypeptides or a single ribosomal RNA gene are present on one minicirle. Currently, 2 ribosomal RNA genes and 16 proteins, all of them present in other photosynthetic plastids, have been found on dinoflagellate minicirles. The protein genes are limited to genes for the core photosystem subunits, ATP synthase, cytochrome b_6f, two ribosomal proteins and 2 other proteins (Hackett, 2004). Many other plastid genes have been moved to the nucleus as reveal by EST sequencing of three dinoflagellates with peridinin plastid, *Lingulodinium polyedr*um, *Amphinidium carterae* (Bachvaroff, 2004) and *Alexandrium tamarense* (Hackett, 2004). 15 genes presented in plastid genome of all other photosynthetic organisms are nuclear-encoded in *A. catenella* (Hackett, 2004). EST sequencing supported the hypothesis that the peridinin plastid retains a minimal genome after massive transfer of photosynthetic genes to the nucleus (Bachvaroff, 2004).

Phylogeny of the peridinin plastid

The chromalveolate hypothesis suggests red algal ancestry for the peridinin plastid. In order to verify this possibility, plastid-encoded minicircle genes were phylogenetically tested. (Zhang et al., 2000). The genes are, however, the fastest-evolving plastid genes known and therefore may subject to various phylogenetic artefacts (Zhang et al., 1999). Phylogenetic analyses of the first identified minicircle genes (23S+16S rDNA) showed their extremely divergent sequences branch with or within

Apicomplexa (Zhang et al., 2000; Takishita et al., 1999). Regardless the branch support, the results are uneasy to interpret: apicomplexans have the second longest branches in the trees. The grouping with dinoflagellates may be a result of the long branch attraction (LBA) artifact (Inagaki, 2004). The nuclear-encoded plastid targeted genes are not as divergent as the minicircle genes and their analyses has led to the current consensus about sistership of the peridinin plastid with the apicoplast (Fast, 2001).

Perkinsida are likely to contain a remnant plastid

Perkinsids have always been considered as lacking the plastid. It was a surprise when several recent studies have challenged this view showing evidence of a non-photosynthetic plastid in the genus *Perkinsus*. First, Stelter et al. (2007) showed *P. marinus* expresses plant-type ferredoxin with N-terminal plastid targeted sequence similar to those of dinoflagellates. *P. marinus* also proved to be sensitive to inhibitors of plastid-localized fatty acid and isoprenoid biosynthesis and the research team identified 3 genes encoding proteins associated with the latter pathway. The second study concentrated on detailed morphology of *Perkinsus atlanticus* zoospores revealing a structure with four membranes (Teles-Grilo et al., 2007). Multi-membrane structures were also found in enriched plastid fractions The structure is usually located at the apical end of the cell and close to mitochondria, which remarkably resembles situation in Apicomplexa. In addition to that growth tests showed *P. atlanticus* is sensitive to thiostrepton, an inhibitor of protein synthesis in cyanobacteria and plastids (Teles-Grilo et al., 2007). Lastly, Grauvogel et al. (2007) found a set of nuclear genes for proteins involved in the plastid-type isoprenoid synthesis, a metabolic pathway known from many plastids including Apicomplexa and dinoflagellates. No genes involved in the cytosolic-type of the pathway known from non-photosynthetic protists were found in the study.

3/ Tertiary endosymbiosis

All known cases of the tertiary endosymbiosis belong to the clade of Dinophyta. In the tertiary endosymbiosis a heterotrophic eukaryote phagocytised another eukaryote, which already contained a secondary plastid. Tertiary endosymbioses involve few known species, but seem to emerge at least 4 times independently on one another (see the chapter Dinoflagellates). The plastid of *Lepidodinium viride* lacks peridinin and fucoxantin, but contains chlorophyll *a*, *b* and other pigments, which suggest prasinophyte (Chlorophyta) origin (Watanabe et al., 1991). It seems the plastid is in a late stage of organelle endosymbiosis – it lacks all the components of original endosymbiont except ribosomes. It is surrounded by four membranes. *Kryptoperidininium* (*Peridininium) foliaceum* contains a three-membrane-bound plastid as a result of tertiary endosymbiosis with a diatom. Nuclear genes of the endosymbiont genome and rbcLS operon analyses unequivocally supported the endosymbiont came from family Bacilariaceae of pennate diatoms (Chesnick 1996, 1997). *Kryptoperidininium* also has an eyespot (stigma), an unpigmented structure bound by three membranes which may be the remnant of the original peridinin plastid. The diatom endosymbiont has retained a nucleus, mitochondria and ribosomes. There are currently four species in three genera recognized as diatom-plastid-bearing dinoflagellates (Pienaar, 2007). Lastly, some members of the genus *Dinophysis* may represent the earliest stage of plastid acquisition through tertiary endosymbiosis. Ultrastructural and pigment features as well as analyses of molecular data indicate these dinoflagellates contain a 4-membranes

plastid of cryptophyte origin (Takishita, 2002; Hackett, 2003). However, significant discrepancy is connected to *Dinophysis* plastid: there is no nucleomorph inside the dinoflagellate endosymbion. Many photosynthetic genes are encoded in nucleomorph of free-living cryptophytes and their products targeted to the plastid. Moreover, *Dinophysis* does not survive long in cell cultures and therefore it's possible it contains a temporarily kleptoplast rather than a permanent plastid (Hackett 03). Kleptoplasty, temporary retained plastids from prey, is relatively common in dinoflagellates. Interestingly, there is another known case of kleptoplasty in the genus *Dinophysis*: Koike et al. (2005) showed *D. mitra* contains a haptophyte plastid of significant sequence heterogeneity concluding it is probably a temporary kleptoplast. Frequent kleptoplasty in heterotrophic dinoflagellates might explain numerous tertiary endosymbioses. Recent analysis of nuclear genes shows all the dinoflagellate species with tertiary endosymbionts are derived from an ancestor with peridinin plastid originating independently on each other (Shalchian-Tabrizi, 2006).

4/ The chromalveolate hypothesis

Six protist supergroups are currently distinguished in the eukaryotic tree of life: Amoebozoa, Opistokonta, Rhizaria, Archaeplastida and Chromalveolata (Adl, 2007). The chromalveolate hypothesis was originally proposed in 1999 by Thomas Cavalier-Smith and has redefined our view on evolution of eukaryotes. It suggest suggesting a common ancestry of chromists and alveolates. The hypothesis is based on the common presence of a secondary plastid, chlorophyll *c* plastid pigment and tubular cristae in mitochondria (except for the cryptophytes). The apicoplast, a remnant plastid in Apicomplexa, is proposed to be related to the peridinin plastid of dinoflagellates and originated from the same red algal acquisition (Cavalier-Smith, 1999). The hypothesis lies on the idea the acquisition of a secondary endosymbiont is an unlikely process, which has happened rarely and counts on subsequent plastid losses in many eukaryotic lineages involved. In recent molecular analyses grouping of Alveolata with heterokont algae is usually well supported (Nozaki, 2003; Harper, 2005). In some analyses haptophytes group with stramenopiles whereas exact position of Cryptophyta is unclear (Yoon et al., 2004; Harper, 2005). Evidence for monophyly of Cryptophyta and Haptophyta have also been published (Patron, 2007).

Chromalveolata are greatly numerous in species and variable in life strategies accounting for about half of recognized protist diversity (Cavalier-Smith, 2004). Their ancestor should be, according to the theory, photosynthetic, but many known representatives are not. These species are supposed to lose their secondary plastid during their evolution. Most species contain two unequal flagella, which often differ in morphology and function. Four main lineages are found in chromalveolates: Heterokonta, Hapthophyta, Cryptophyta and Alveolata.

Heterokonta is a diverse group involving various phototrophs such as kelps, diatoms, Xanthophyceae and Chrysophyceae as well as several colorless lineages of heterotrophs and parasites, Oomycota, Labyrinthomorpha and Opalinea. According to chromalveolate hypothesis these lineages arose from a photosynthetic ancestor. Phylogenetic analyses reveal some genes of the parasitic genus *Phytophthora* (Oomycota) are likely to be of plastid origin (Tyler et al., 2006; Jiroutova et al., 2007). A reduced nucleus (called nucleomorph) found within plastid membranes of the cryptophyte *Guillardia theta* has retained the genome of the phagocytised red alga (Douglas et al., 2001)

.Investigating gene translocations between the cryptophyte plastid, nucleomorph and nucleus may give valuable data about process of secondary endosymbiont acquisition (Zauner, 2000).

5/ Alveolates and their phylogeny

The members of the group are defined on the presence of cortical alveoli, flattened vesicles forming a single layer under their plasma membrane. Micropores, distinct openings on the surface of their cells, are another synapomorphy of alveolates (Paterson, 1999). The group comprises Ciliophora, Dinophyta and Apicomplexa, which represent great diversity and several different life strategies – heterotrophy, obligatory parasitism and phototrophy (Leander and Keeling, 2003). Alveolate is a monophyletic group (Gajadhar et al., 2001; Fast et al., 2002, Harper, 2005). Apicomplexa are closer to Dinophyta than to Ciliophora forming together a group called Myzozoa (Cavalier-smith and Chao, 2004). A novel phylum has been recently described.

6/ Apicomplexa, the most studied protists

Typical morphological features

Apicomplexa (Levine, 1970) is a strictly parasitic group of alveolate prostists characterized by presence of the apical complex, an assemblage of structures at the anterior end of a cell. The complex enables parasites to penetrate cell and live intracellulary thereby evading much of the host immunity surveillance. It is generally found in sporozoit, merozoit and gamont stages and consist of polar rings, a spiral conoid, pair 2-8 rhoptries and numerous micronemes. The group comprises about 6000 species in more than 300 genera, with bulk of species probably not yet described (Adl et al., 2007). Additional typical features like formation of the parasitophorous vacuole inside host cells and complicated life-cycles are found in some species. The cell membrane of the Apicomplexa is supported by flattened alveoli sac similar to that of dinoflagellates and ciliates. Cells contain mitochondria with tubular cristae. Flagellar stages are restricted to microgametes (for details about Apicomplexa see Perkins et al., 2000).

Phylogeny of Apicomplexa

Apicomplexa contain many medically and economically important species, such as genus *Plasmodium* causing malaria, *T. gondii* causing toxoplasmosis, *Cryptosporidium* sp. causing diseases of human and animals and *Babesia*, *Theileria* and *Eimeria* species causing diseases of domestic animals. The group is adapted to parasitic way of life and possess many derived characteristics, which hamper looking for phylogenetic synapomorphies with other protists. Studies of molecular sequence data, however succeeded in resolving apicomplexan position within the eukarytic tree: they congruently show Apicomplexa is a sister clade of Dinophyta within Alveolata. Apicomplexa includes several distinctive lineages, Coccidia, Haemosporidia, Piroplasmida, Gregarinea and *Cryptosporidium* and is supposed to be monophyletic. Gregarines are traditionally considered as the most primitive apicomplexans. Among the crown groups, Haemosporidians are close to Piroplasmida and their common sister group is Coccidia. Position of Cryptosporidium is not entirely clear, but it either constitutes an early-branching separate lineage in the apicomplexan tree or it is a sister group of gregarines.

7/ Relationships of Colpodellida and Perkinsida

Colpodellids, apicomplexan sisters

Little is known about colpodellids, predatory biflagellate alveolates. They resemble *Perkinsus* species in some features to such as an apical open conoid and rhoptries. Morphological features are obviously insufficient to reveal colpodellid relationships and therefore several 18S rRNA gene sequences of the genus *Colpodella* have been obtained and included to phylogenetic analyses. Siddall et al. (2001) revealed colpodellids are sisters of ciliates based on a concatenated actin+18S rRNA gene dataset, but it was later revealed the actin sequence belonged to *Bodo* sp. rather than to *Colpodella* sp. (Saldarriaga, 2002). Kuvardina et al. (2002) performed analyses in which *Colpodella* sp. forms a sister clade to Apicomplexa, even though this position has obtained only medium bootstrap support.

Morphologyand phylogenetic position of Perkinsus sp.

Perkinsida are protozoan parasites of mollusks and agents of infamous 'Dermo' disease devastating populations of oysters. The genus *Perkinsus* were excluded from Apicomplexa in 1990s based on the overall morphological and molecular evidence. Unlike Apicomplexa, the apical conoid-like collar of *Perkinsus* is side opened and lacks the polar ring complex. *Perkinsus* differs from Apicomplexa also in the structure of micropores and character of its flagella (Siddall et al., 1997). The decisive moment of *Perkinsus* exclusion from Apicomplexa was sequencing of its actin and small ribosomal subunit genes (Reece et al., 1997, Siddall et al., 1997). Further molecular data have congruently shown perkinsids form a sister group of Dinophyta (de la Herran 2000; Figueras, 2001; Saldarriaga 2001; Saldarriaga, 2003; Leander and Keeling, 2004).

New species of Myzozoa

Newly described genera *Parviluciphera* (Noren et al., 1999) and *Cryptophagus* (Brugeolle, 2002) are relatives of Perkinsida. Both are parasites of other eukaryotes (dinoflagellates and cryptophytes respectively) and have zoospores with two unequal flagella. Likewise *Perkinsus* and *Colpodella*, *Cryptophagus* and *Parviluciphera* possess an open side conoid and rhoptries. These structures may be homological to the apical complex of Apicomplexa and therefore ancestral to all Myzozoa (Leander and Keeling, 2003). However, a few other alveolates such as *Acrocoelus*, and *Voromonas* have been described and assigned to Myzozoa (Cavalier-Smith and Chao, 2004) and new species are still discovered (Leander and Hoppenrath, 2007). A lack of morphological or molecular data for many of the new myzozoans hinder us from finding out their evolutionary relationships.

8/ Dinoflagellates

Morphological and ecological features

The group Dinophyta involves single-celled protists of great morphological and ecological variability. Their life strategies comprise heterotrophy (roughly 50% of known species), photoautotrophy and parasitism, with ecological valence stretching vastly from freshwater to marine environments. Dinoflagellates are important primary producers. Some species produce toxins causing infamous 'red tides', other are capable of bioluminiscence or life inside cells of corals. They usually

have two unequal flagella, which enable them to carry out complex movements. Some species possess theca, an outside armor protecting them against predators, whereas other dinoflagellates are 'naked'.

Evolutionary history of dinoflagellates

On the paleontological record, a massive expanse of the group came in Mesozoic: the first non-controversial fossils come from the early Triassic period, but the main divergence of the group came in the Jurassic and Cretaceous (208-66 mya) (Fensome et al., 1999). Following the fossil diversity, the Mesozoic radiation seems to have given rise to most of the recent groups, with total number of species declining until today. This might explain why it is so uneasy to recover relationships between the major dinoflagellate groups by using both, the morphological synapomorphies and molecular matrices (Fensome et al., 1999). Parasites of the group Perkinsida are universally though to be sister group of the dinoflagellates. To better understand the evolution of the group, extensive effort has been put to identify dinoflagellates with plesiomorphic features of the group. *Oxyrrhis marina* is based on the molecular data currently considered as the earliest branch of the dinnoflagelate phylogenetic tree (Saldarriada, 2003; Leander and Keeling, 2004). *Oxyrrhis* is a peculiar heterotroph with many derived features. The inner phylogeny of the dinoflagellates has traditionally collapsed into several lineages with unknown interrelationships (Saldarriaga et al., 2001). Only recent phylogenetic studies based on multi-gene datasets obtained better resolution in inferred dinoflagellate trees (Murray, 2005; Zhang, 2005, 2007; Gribble, 2006) with basal position of *Amphinidium*, *Heterocapsa* and Suessidales.

9/ The status of current knowledge of *RM12*

RM12 is a photosynthetic alveolate living in the reef coral *Plesiastrea versipora*. The organism was co-isolated with endosymbiotic dinoflagellates of the genus *Symbionidium* from samples of the corals collected in Sydney Harbour, Australia in 2001 (Moore, 2006). Morphological distinction from dinoflagellates of the genus *Symbiodinium* resulted in the isolation of the *RM12*'s total DNA and sequencing of the 18S rRNA gene (Moore, 2006). Phylogenetic analysis suggested *RM12* is not a dinoflagellate, but rather has connections to apicomplexans, dinoflagellate sisters. Such a relationship would be of great evolutionary interest, because Apicomplexa are supposed to arise from photosynthetic organisms and they contain a relict plastid (see previous chapters).

Further molecular and morphological data to test this relationship have been collected (Moore et al., 2007, *in press*). Almost entire sequences of the small and large ribosomal subunit RNA genes, a partial sequence of the plastid small nuclear subunit rRNA gene and a sequence of the gene encoding photosystem II D1 protein (*psbA*) were amplified and phylogenetically tested. In all trees of the nuclear ribosomal genes *RM12* branched next to Apicomplexa. The analysis of partial plastid 16S rDNA sequence revealed close proximity of RM12's plastid and the apicoplast. Fast evolving sequences of the peridinin dinoflagellates were omitted from this analysis. In the *psbA* tree computed from the aminoacid dataset, *RM12* appeared as a sister taxon of peridin dinoflagellates, although this branching was not well supported. Apicomplexa lacks this gene and their sequences were therefore missing in the tree.

Strikingly, an exception from the standard genetic code was found in the plastid of *RM12*. An alignment of *psbA* sequences revealed 7 stop codons in the *RM12*'s sequence. All the seven positions are conservative in other plastids and encode tryptophan (UGG). *RM12* has UGA triplets at these positions and because *psbA* product is an essential photosynthesis-associated protein the presence of stop codons in the sequence is extremely unlikely. Alternative UGA-Trp coding is known from mitochondria and the only type of a plastid – the apicoplast of some Coccidea. Considering such alternation in the plastid genetic code is known to happen only once, the presence of UGA-Trp coding in *RM12* is considered to be a strong synapomorphy of both plastids.

Pigment analysis revealed presence of chlorophyll a, violaxanthin, and a novel carotenoid as the major components, and β,β-carotene as a minor component indicating a unique pigment composition of *RM12*'s plastid. Detailed ultrastructure using transmission and scanning electron microscopy showed *RM12*'s plastid is surrounded by 4 membranes, which is the currently hypothesized ancestral number of membranes for both the apicoplast and the plastid of peridinin dinoflagellates. RM12's possesses tubular cristae in its mitochondria.

10/ Summary and the aims of my project

The data suggest, *RM12* is the closest photosynthetic relative of Apicomplexa and bears a pigmented plastid, which is ancestral to the apicoplast. The organism may help to understand the origin of parasitism in Apicomplexa and its plastid may be useful in medical research of the apicoplast.

In this work, I characterize the plastid of the photosynthetic alveolate *RM12*. The experimental part of the work consists of two approaches: **a) inferring the plastid origin and evolution using further molecular markers** and **b) assessing the plastid genome size**. Four genes encoded in *RM12*'s plastid (16S and 23S rRNA genes, elongation factor TU gene = *tufA* and photosystem II protein D1 gene = *psbA*) were sequence and their phylogenies inferred. Radioactively labeled *PsbA* probe was hybridized with the total DNA separated by pulse-field agars gel electrophoresis and the plastid genome size was estimated. The resulting phylogenetic relationships of *RM12*'s plastid are critically discussed in the context of current knowledge of plastid evolution and phylogeny of the alveolates.

Methods

1/ DNA isolation, PCR amplification and gene sequencing

Cultures of *RM12* cells were kindly provided by Robert B. Moore, University of Sydney, Australia. Cells were grown in f/2 seawater medium (see details in Chrudimsky, 2007) at 18°C in 14 hour light cycle.

DNA was successfully isolated by several methods including the standard phenol-chloroform extraction, the chelex extraction and the extraction by using commercial kits Tissue DNA Spin Kit (Jetquick) and DNeasy Plant Mini Kit (Qiagen). The *psbA* gene sequenced in this study was amplified from the DNA extracted by the phenol-chloroform extraction (for details see Chrudimsky, 2007). The other three genes were amplified from the DNA isolated by using DNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions. The extractions yielded between 30-300 ng/µl of DNA, which was stored in a freezer.

The genes encoding ribosomal subunit RNAs were amplified as 2 (16S rRNA) and 3 (23S rRNA) overlapping parts. with organism specific primers. The partial sequence of the plastid 16S rRNA gene (Moore et al., 2008) was used to design organism specific primers (pSR950 and pSF800). The first part of 23S rDNA was amplified with non-specific primers and enabled me to design two specific primers (pLSU_RX and pLSU_RX2). *PsbA* was amplified in two parts each of them covering most of the gene; the gene amplification has been previously described (Moore et al., 2007). *TufA* was amplified with a single primer pair (for primer details see Table 1). The genes or their parts were amplified in 25 µl reaction mixes containing 400 µM nucleotides, 0.25 U of *Taq* polymerase (Topbio), PCR buffer (10mM Tris-Cl pH 8.8, 1% Triton X-100, 1.5 mM $MgCl_2$, 50mM KCl), 0.4 pM primers (0.7 pM of *psbA* gene primers), 35-60 ng of DNA (75-90 ng in *psbA* amplification) and autoclaved milipore H_2O. PCR cycles were: 5mins at 95°C (denaturation) followed with 30 cycles of [95 °C for 1min (denaturation), xx °C for 1 min (primer annealing), 72 °C for 1:30 min (elongation)]; 72 °C for 8 mins (final elongation), where xx is the annealing temperature for the particular primer pairs described in Table 1.

PCR products were electrophoretically separated on agarose gel, cut out of the gel and purified with Quigen Gel Extraction Kit. The purified products were cloned using TOPO TA Cloning Kit, pCR®2.1-TOPO® vector (16S and 23S rDNA, *tufA*) and Quiagen PCR Cloning Kit, pDRIVE vector (*psbA*) according to the manufacturers' protocols. One to five clones of each gene or gene part were sequenced (both strands; see Table 1). Gene sequences were assembled in Seqman (DNASTAR).

2/ Sequence alignments and phylogenetic analyses

Nucleotide sequences of ribosomal RNA genes and aminoacid sequences (*tufA*, *psbA*) were aligned in MAFFT (Katoh et al. 2002, 2005) using the L-INS-I algorithm at default conditions.

Nucleotide sequences of *tufA* and *psbA* were reverse-translated from aminoacids aligned using the Clustal W algorithm in Bioedit 7.0 (Hall, 1999). Q-INS-I algorithm (MAFFT) was used to align the '23S-SHORT' dataset (see bellow) and the results compared to the L-INS-I alignments. Q-INS-I aligns ribosomal sequences according to their secondary structure, but is currently available only for limited datasets; see the program website, http://align.bmr.kyushu-u.ac.jp/mafft/online/server/). Two alignments of the 16S and 23S rRNA genes containing/lacking divergent sequences of the peridinin dinoflagellates were created: "16S+DINO"/"16S' and "23S+DINO"/"23S" respectively. An experimental short dataset of 23S rDNA containing partial sequences of 8 additional dinoflagellate and 2 apicomplexan species ("23S SHORT+DINO) was prepared. In protein genes, three dataset were created based on amino acid, nucleotide, and 1. + 2. nucleotide coding positions (psbA_aa, psbA_n, psbA_n12 and tufA_aa, tufA_n, tufA_n12). The aligned sequences were manually edited in Bioedit 7.0 (Hall 99). Unaligned and ambiguously aligned positions as well as sequences of fast-evolving species from outside Alveolata were omitted from the datasets.

Nucleotide datasets were analysed using the maximum likelihood (ML) method in PHYML 2.4.2 (Guindon and Gascuel, 2003; Guindon et al., 2005; both, the computer-based and the online version of the program were used) and the Logdet paralinear distances (Logdet) method as implemented in PAUP* 4.10b (Swofford 2001). Aminoacid datasets were analyzed using ML in PHYML 2.4.2. The maximum likelihood models for nucleotide and aminoacide matrices were GTR+Γ+I and WAG+Γ+I respectively, with gamma distribution in 4 categories and all parameters estimated from datasets. Logdet distances were calculated for the nucleotide datasets by using TBR branch swapping. Bootstrap branch supports were computed for all derived trees (300, 500 or 1000 replications).

3/ Pulse-field agarose gel electrophoresis and Southern hybridization

Pulse-field agarose gel electrophoresis (PFGE) was used to separate large DNA molecules. Living cells of RM12 were slowly centrifuged, washed in SET buffer and embedded in 1.5% low-melting agarose ($2{,}6x10^8$ cells/1 ml). Plugs were digested in a lysis buffer (0.5 M EDTA pH=9.5, 1% N-Lauroylsarcosine and Proteinase K (20 mg/ml)) at 50 °C for 28 h. Digested plugs were stored in SET buffer in 4 °C. PFGE was carried out in 0.5x TBE buffer using the CHEF-DR® III system (BioRad) with PULSE MARKER™ 0.1-200 kb (SIGMA) as a control. PFGE run at the following conditions: current = 6 V/cm; runtime = 20 h; initial/final switch time = 0.5/25 s; temperature = 14 ºC.

The PFGE gel was stained in EtBr to check up the separation of the total DNA. The standard Southern blot protocol (Sambrook and Russell 2001) was followed to transfer the DNA to BioBondTM-Plus NYLON MEMBRANE (Sigma) via alkaline transfer. HexaLabel™ DNA Labeling Kit (Fermentas) and MicroSpinTM G-25 columns (Amersham) were used to make and purify radioactively labeled DNA probe from the cloned *psbA* gene. Southern hybridization was performed in phosphate buffers according to the standard protocol at 65°C (Sambrook and Russell 2001).

Table 1: A scheme of amplification of four genes used in this study with lengths of the consensus sequences. PCR primer pairs, primer sequences, primer anneling temperatures (T_m) and number of sequenced clones for particular genes/gene parts are indicated. All the parts were sequenced from both DNA strands.

Gene	Amplification scheme	Primers		Primer sequences		T_m	Clones seq.
		Forw.	Rev.	Forw. (5 '- 3')	Rev. (5 '- 3')		
Plastid SSU rRNA gene		pS_BrF2 pSF800	pSR950 pSRcoc	GATCCTGGCTCYGRAWRAACG TGGGATTAGATACCCCAGTAGTCTTA	CACYCTTGCGAGCGTACTCCT TATTATCCAGCCGCACCTTCCAG	52 ˚C 50 ˚C	2 3
Plastid LSU rRNA gene		pLSU1F pLF400 pLSU3F	pLSU_RX2 pLSU_RX pLR3100	CTGAATCATCTTAGTACTCAAAG ACAAYKATTYYCTTAGTAGTGGCGA GCGAAATTCCTGTCGGGTAAGTCC	ATTGAGTGACAAGCACTGTATCTA TCTCCCACTTTAATTGTATAATGAG TTTTCYYACTTATATGCTTTCAGYA	45 ˚C 45 ˚C 48 ˚C	2 1 2
Elongation factor TU gene *(tufA)*		EfTUF	Ef-tu1R	GGDCATGTAGATCATGGDAAAAC	ATAAAATTGAGGTYWATATCC	42 ˚C	5
Photosystem II subunit D1 gene *(psbA)*		psbA_U140Forw psbA_U140Forw	psbA_micRev psbAL1_Rev	ACCGTYTYTAYATCGGTTGGTTCGG ACCGTYTYTAYATCGGTTGGTTCGG	GATTTGCACGATTTAAGATATCGGCT CRTGCATWACTTCCATWCC	45 ˚C 45 ˚C	1 3

Results

Amplification of the genes

The amplification of *RM12* plastid genes turned out to be difficult with respect to numerous bacterial contaminations. Specific primers for these genes are uneasy to design, because Apicomplexa sequences are highly divergent and conservative parts of gene alignments are often similar to bacterial homologs. Our cultures of *RM12* are not axenic, which results in sequencing of many bacterial genes. All in all, 12, 15, 4 and 12 gene primers were tested to amplify 16S rDNA, 23S rDNA, *psbA* and *tufA* respectively. Organism-specific primers based on already sequenced parts had to be designed in the case of ribosomal RNA genes. The length of the amplified sequences is shown in Table 1. The almost complete plastid 16S rDNA sequence was aligned to the previously published partial sequence (accession number EU106871; Moore et al., 2008). One transitional (G-T) and one transversional (G-A) mutation differ between the sequences in their overlapping parts. The *psbA* gene was amplified as two sequences (with two different primer pairs), both of which are identical and differ only slightly in length. The longer of the *psbA* sequences (amplified with psbA_U140Forw and psbAL1_Rev primers) has already been published (EU106869, Moore et al., 2008).

Plastid 16S rDNA analyses (Fig. 2-3)

Sistership of *RM12* and Apicomplexa is strongly recovered (100% trees) in ML and LogDet analyses of the "16S" dataset. Apicomplexa is well differentiated to subgroups in accordance with their standard classification in the ML tree. The backbone of the tree is unsupported, but several eukaryotic clades are well recovered (Stramenopila, Cryptophyta, Viridiplantae). Alveolates appear as sisters of stramenopiles, but this branching lacks bootstrap support. *Euglena gracilis* branches artificially together with two *Pavlova* species (Haptophyta). Several dinoflagellate tertiary endosymbionts branch within their ancestral families: *Dinophysis acuta* and *D. norvegica* plastids with cryptophytes, *D.mitra* kleptoplast with haptophytes and fucoxanthin-containing dinoflagellates (*Karenia and Karlodinium*) with haptophytes. In the LogDet tree, the bootstrap support for the split between Coccidea and other Apicomplexa remarkably decreases. The LogDet tree also reveals unsupported branching of *B. natans* and 4 red algae of the Bangiophycea family next to the alveolates.

The peridinin dinoflagellates form extraordinary long branches in both analyses of the "16S+DINO" dataset (Fig. 2). They group with Piroplasmida in the ML tree, but this branching is not supported. Interestingly, they are excluded from Apicomplexa in the LogDet tree, while *RM12* retains its sister position to the parasitic group. However, the LogDet tree lacks any branching support at this level. Several species appear next to or within the alveolates in the tree among which are remarkable the fucoxathin-containing dinoflagellates (in the "16S"ML tree reliably settled within Haptophyta).

Plastid 23S rDNA analyses (Fig. 4-5)

23S rDNA dataset contains significantly less sequences, but has better support of the tree backbone than the 16S rDNA dataset in the ML analyses. *RM12* forms a sister clade to Apicomplexa in these trees with highly significant support (100% of the ML and 98% of the LogDet trees). All the main groups of photosynthetic eukaryotes are recovered in the ML trees with variable supports. The long branch of non-photosynthetic green alga *Helicosporidium* appears next to the alveolate clade. This branching is surprisingly supported in the ML trees (bootstrap 87 and 88), but is missing in the shortest LogDet tree (Fig. 6). In the bootstrap LogDet tree (not shown) *Helicosporidium* appears together with *B. natans* as a second sister branch to the alveolates after the euglenophytes. Euglenophyta are stably recovered close to alveolates appearing in 35-60 % of trees in particular analyses.

If the peridinin dinoflagellate sequences are add they group with the fastest evolving sequences of the genus *Plasmodium*. This grouping is preserved in the LogDet trees.

The short experimental dataset "23S SHORT+DINO" was aligned using two different algorithms in MAFFT: Q-INS-I and L-INS-I. The former aligns ribosomal sequences according to their secondary structure, whereas the latter is a standard algorithm. All trees derived from these two alignments were significantly less supported than the analyses of the complete 23S rDNA sequences. Addition of more species did not bring about better resolution among alveolates. Dinoflagellates always appeared next to Haemosporidia and *RM12*'s sister position to the clade of apicomplexans and dinoflagellates was revealed in only 84% of trees. The trees derived from Q-INS-I were slightly better resolved than the L-INS-I trees.

PsbA analyses (Fig. 6-8)

Apicomplexans lack the *psbA* gene and therefore were not included in *psbA* analyses. Particular lineages of photosynthetic eukaryotes are usually recovered in this gene, but their relationships remain unresolved. The longest branches in the trees belong to the peridinin dinoflagellates, which are always associated with haptophytes and the fucoxanthin-containing dinoflagellates. *RM12* is found next to Stramenopila in the tree derived from the aminoacid dataset, but this position has no support. Nucleotide trees give a bit different topology – Stramenopila are replaced among other eukaryotes, whereas *RM12* appears as a sister of the peridinin dinoflagellates in the ML trees and as sister of the peridinin+fucoxanthin dinoflagellates in the LogDet trees. In all of the cases, support for this branching is very low. Peridinin dinoflagellates are monophyletic in all trees and *Amphinidium* and *Heterocapsa* genera branch at the basis of the group.

TufA analyses (Fig. 9-10)

No *tufA* sequences were found in dinoflagellates. *RM12* has a surprisingly long branch and always groups with Apicomplexa. This is very well supported in both the aminoacid and nucleotid-derived ML trees and even in LogDet analyses. The main lineages of photosynthetic eukaryotes are recovered in many cases, but their relationships are not supported. Euglenophytes form a sister clade

of the alveolates in the trees derived from the nucleotide dataset, but are found among Viridiplantae in the aminoacid-derived tree.

The size of *RM12*'s plastid genome

Large amount of *RM12*'s DNA was separated by using PFGE forming a long smear across the whole gel. Two bands of unsuspected sizes corresponding to the nuclear genome were visible at the top of the gel. The smear got brighter roughly between 30-60 kbs. DNA was blotted and cross-linked onto a nylon membrane.

Southern hybridization revealed an apparent band about 130kb in size corresponding to the size of a plastid genome. Interestingly, a strong DNA hybridization appeared from about 60 kbs to the bottom of the membrane (the shortest DNA molecules transferred to the membrane were about 8 kb in size). DNA transferred to the membrane from around the well hybridized at the top of the membrane.

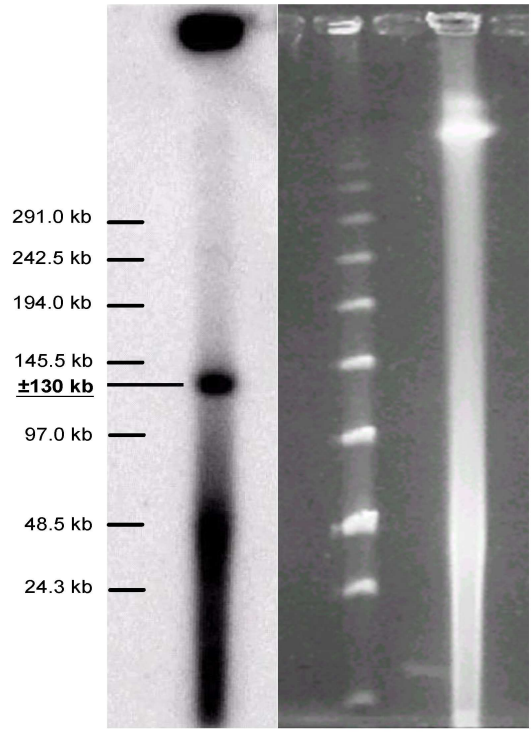


Fig. 1: Pictures of the hybridization membrane (left) and the PFGE gel (right) Hybridization with *psbA* probe revealed a 130kb band and an unexpected smear at the bottom of the gel.

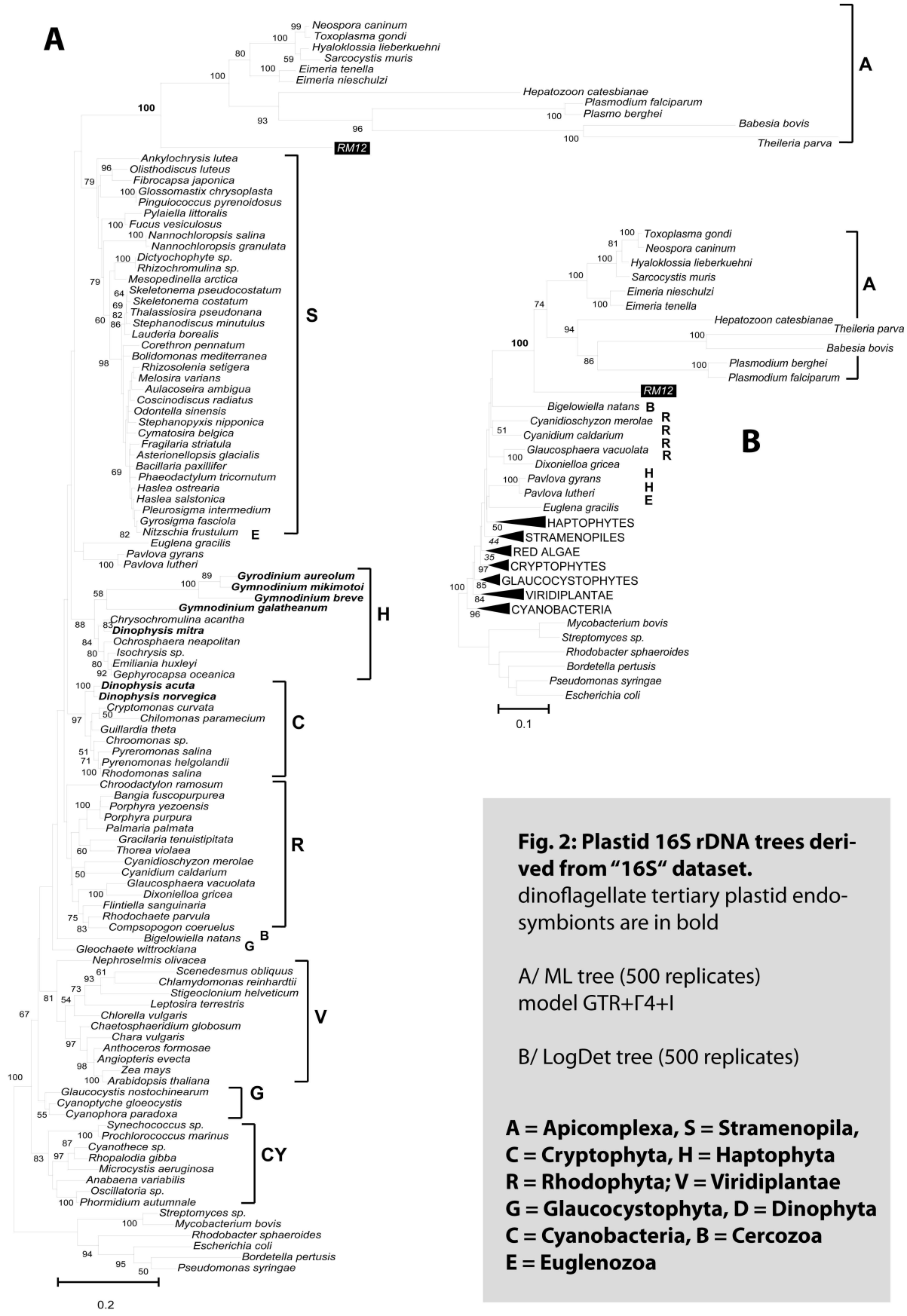


Fig. 2: Plastid 16S rDNA trees derived from "16S" dataset.
dinoflagellate tertiary plastid endosymbionts are in bold

A/ ML tree (500 replicates)
model GTR+Γ4+I

B/ LogDet tree (500 replicates)

A = Apicomplexa, S = Stramenopila,
C = Cryptophyta, H = Haptophyta
R = Rhodophyta; V = Viridiplantae
G = Glaucocystophyta, D = Dinophyta
C = Cyanobacteria, B = Cercozoa
E = Euglenozoa

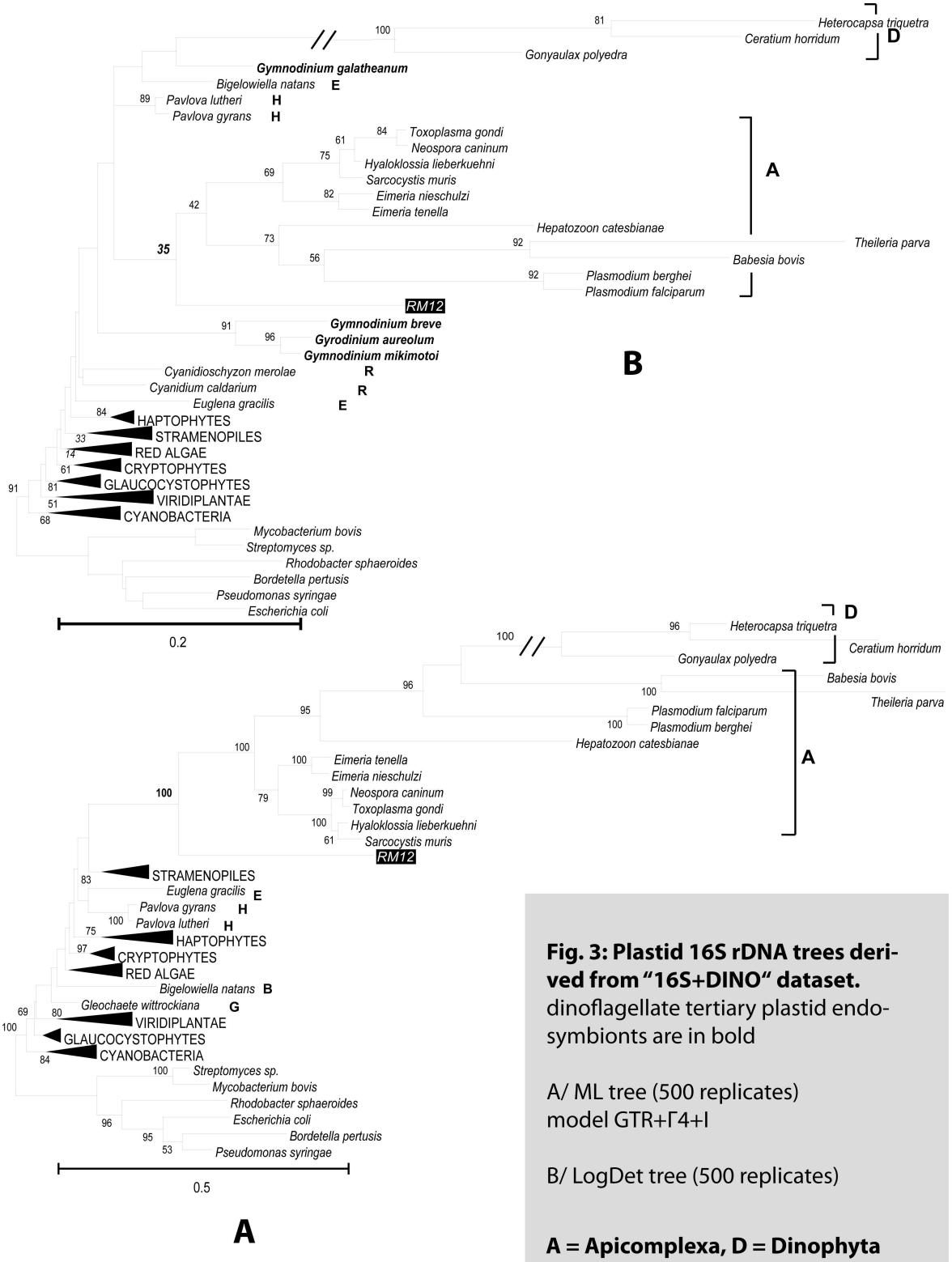


Fig. 3: Plastid 16S rDNA trees derived from "16S+DINO" dataset. dinoflagellate tertiary plastid endosymbionts are in bold

A/ ML tree (500 replicates) model GTR+Γ4+I

B/ LogDet tree (500 replicates)

A = Apicomplexa, D = Dinophyta

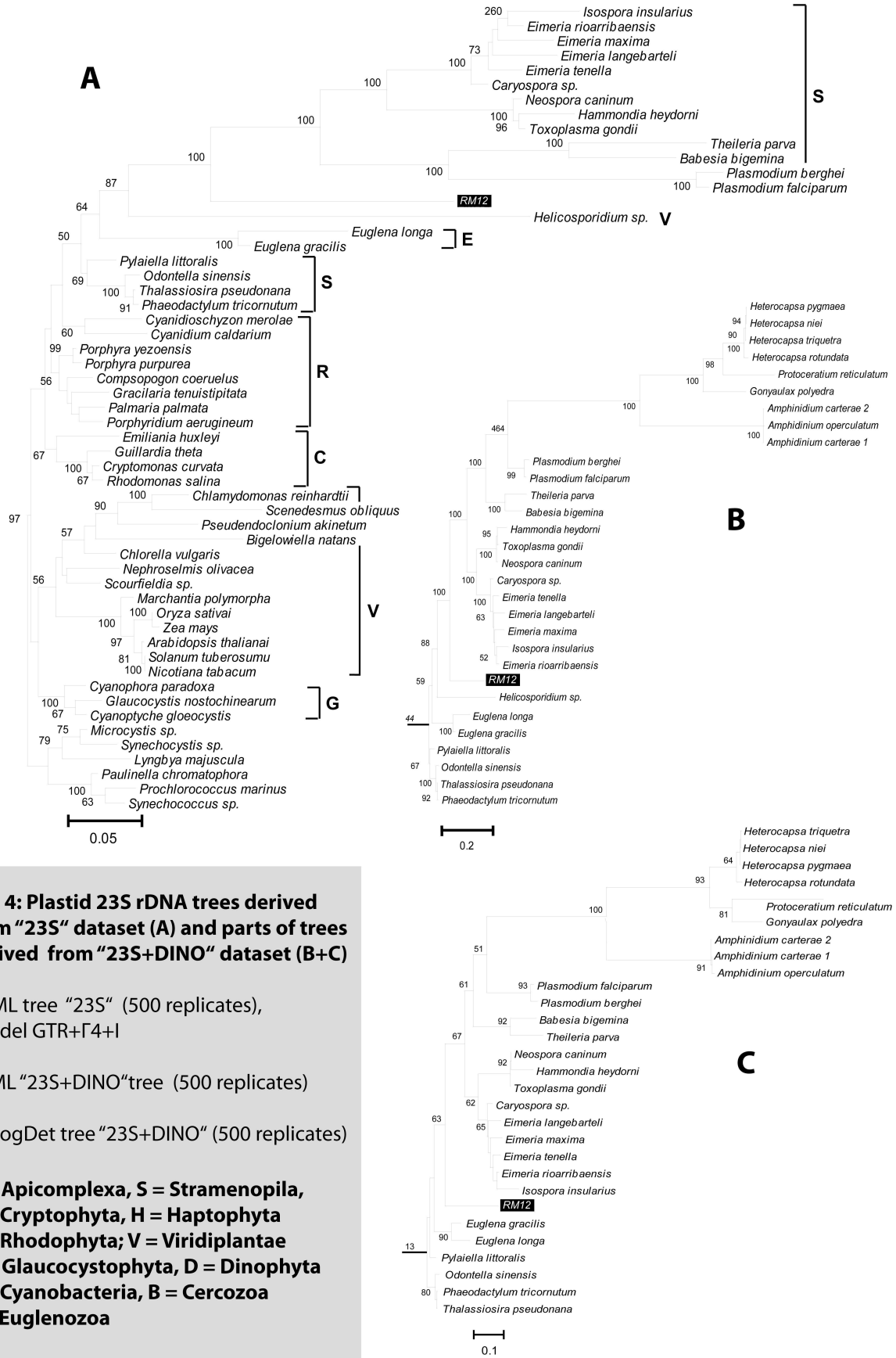


Fig. 4: Plastid 23S rDNA trees derived from "23S" dataset (A) and parts of trees derived from "23S+DINO" dataset (B+C)

A/ ML tree "23S" (500 replicates), model GTR+Γ4+I

B/ ML "23S+DINO"tree (500 replicates)

C/ LogDet tree "23S+DINO" (500 replicates)

A = Apicomplexa, S = Stramenopila,
C = Cryptophyta, H = Haptophyta
R = Rhodophyta; V = Viridiplantae
G = Glaucocystophyta, D = Dinophyta
C = Cyanobacteria, B = Cercozoa
E = Euglenozoa

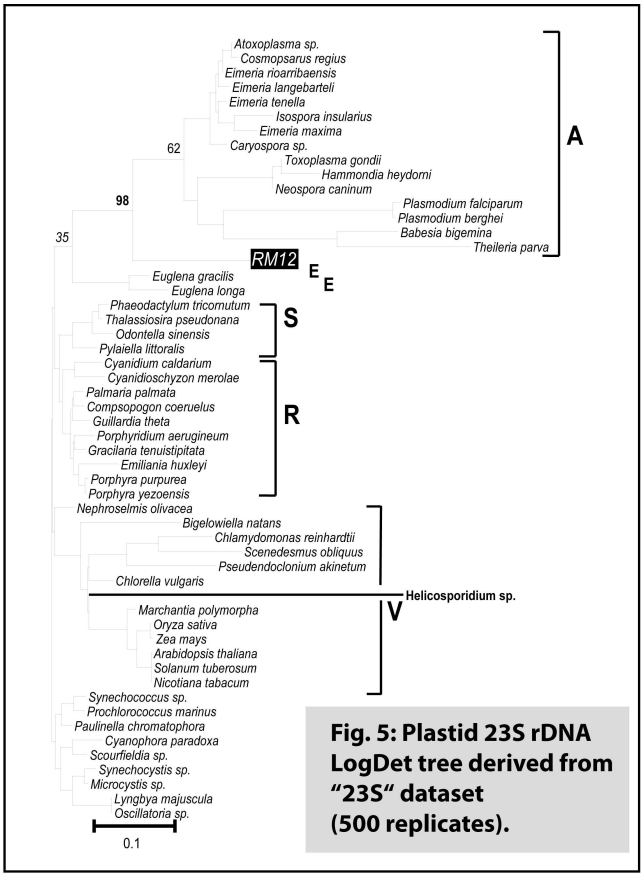


Fig. 5: Plastid 23S rDNA LogDet tree derived from "23S" dataset (500 replicates).

Fig. 6: PsbA ML tree LogDet tree derived from "psbA_aa" dataset.
Dinofalgellate tertiary plastid endosymbionts are higlighted in bold
model WAG+Γ4+I (500 replicates)

A = Apicomplexa, S = Stramenopila, C = Cryptophyta, H = Haptophyta, R = Rhodophyta; V = Viridiplantae, G = Glaucocystophyta, D = Dinophyta, C = Cyanobacteria, B = Cercozoa, E = Euglenozoa

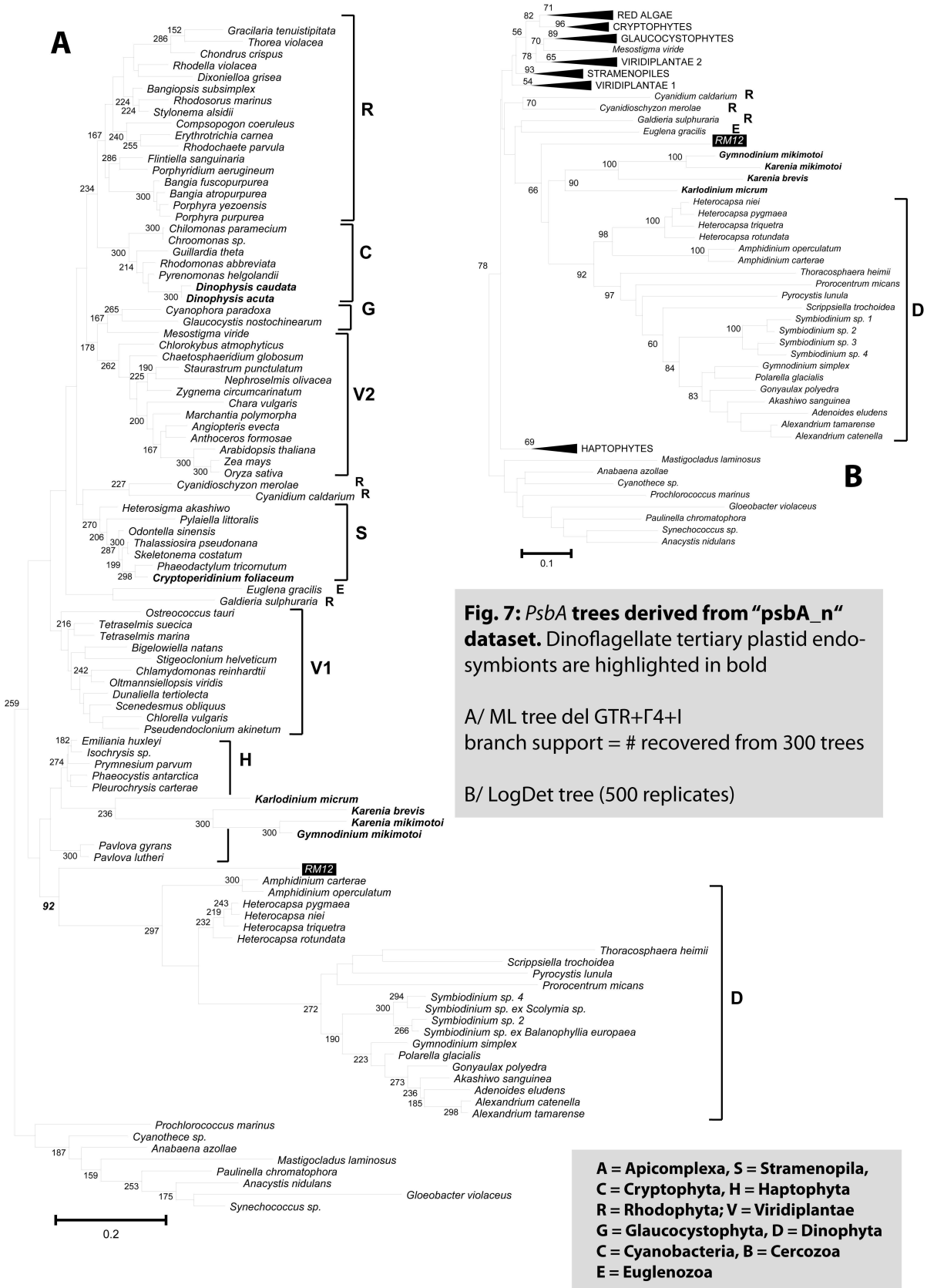
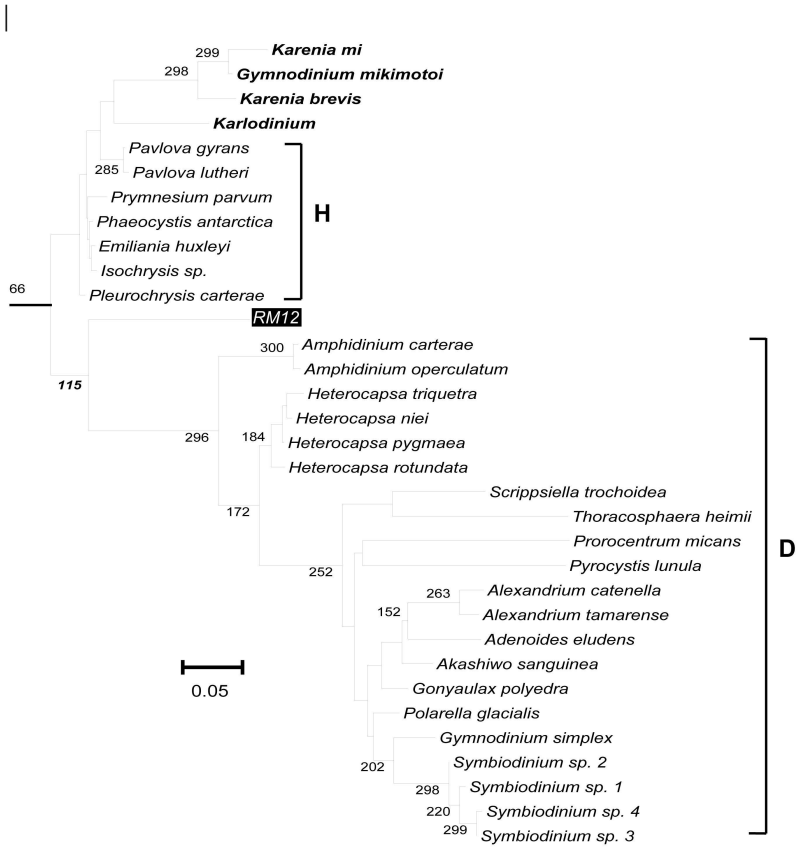


Fig. 7: *PsbA* **trees derived from "psbA_n" dataset.** Dinoflagellate tertiary plastid endosymbionts are highlighted in bold

A/ ML tree del GTR+Γ4+I
branch support = # recovered from 300 trees

B/ LogDet tree (500 replicates)

A = Apicomplexa, S = Stramenopila,
C = Cryptophyta, H = Haptophyta
R = Rhodophyta; V = Viridiplantae
G = Glaucocystophyta, D = Dinophyta
C = Cyanobacteria, B = Cercozoa
E = Euglenozoa



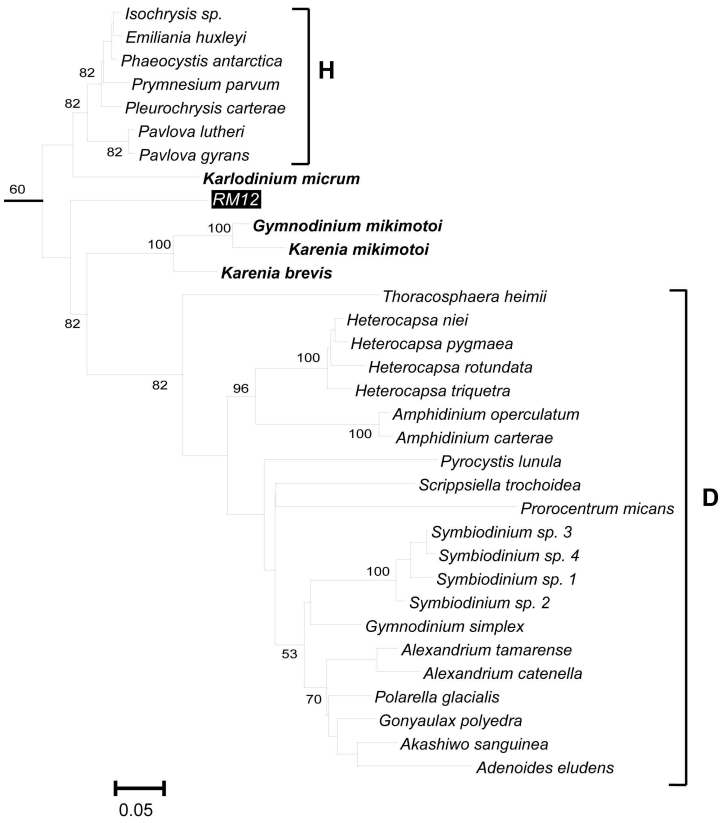


Fig. 8: Partial *PsbA* trees derived from "psbA_n12" dataset. Dinoflagellate tertiary plastid endosymbionts are highlighted in bold

A/ ML tree del GTR+Γ4+I branch support = # recovered from 300 trees

B/ LogDet tree (500 replicates)

H = Haptophyta

D = Dinophyta

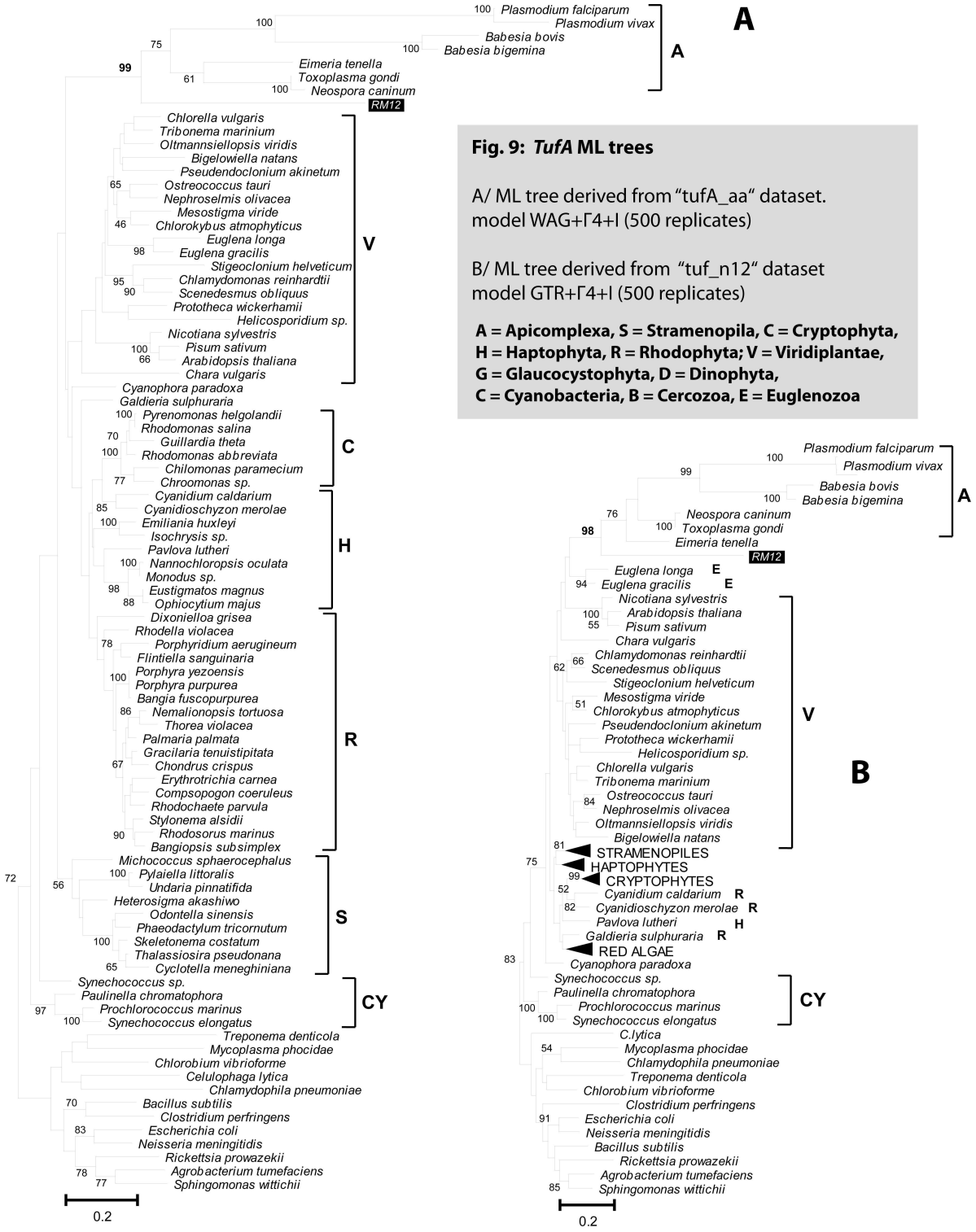


Fig. 9: *TufA* ML trees

A/ ML tree derived from "tufA_aa" dataset.
model WAG+Γ4+I (500 replicates)

B/ ML tree derived from "tuf_n12" dataset
model GTR+Γ4+I (500 replicates)

A = Apicomplexa, S = Stramenopila, C = Cryptophyta, H = Haptophyta, R = Rhodophyta; V = Viridiplantae, G = Glaucocystophyta, D = Dinophyta, C = Cyanobacteria, B = Cercozoa, E = Euglenozoa

Fig. 10: *TufA* trees derived from the "tufA_n" dataset

A/ ML tree model GTR+Γ4+I (500 replicates)

B/ LogDet tree (500replicates)

A = Apicomplexa, S = Stramenopila,
C = Cryptophyta,
H = Haptophyta, R = Rhodophyta;
V = Viridiplantae,
G = Glaucocystophyta,
D = Dinophyta,
C = Cyanobacteria, B = Cercozoa,
E = Euglenozoa

Discussion

What have the analyses revealed?

Highly divergent sequences of plastid 16S and 23S rDNA from peridinin dinoflagellates proved to have little phylogenetic use. LogDet analysis, which may resolve phylogenetic artifacts connected with different nucleotide composition in particular sequences, was able to replace them out of Apicomplexa in 16S rDNA tree, but not in 23SrDNA tree (compare to Zhang, 2001, who managed to it in both trees). Interestingly in the former of those trees, RM12 still retains its sistership with Apicomplexa. Use of minicircle gene sequences may however result also in phylogenetic artefacts (Inagaki, 2004). Therefore only results of "16S"and "23S"of are taken into account. In the tree derived from both these datasets, *RM12* specifically groups with Apicomplexa and this grouping is very strongly supported. LogDet analysis is also strongly supportive for this branching, which is very important, because both branches of RM12 and Apicomplexa are yet long compared to the other species. Similarly, in all *tufA* ML trees *RM12* robustly groups with Apicomplexa. Therefore, analyses of the 16S rDNA, 23S rDNA and *tufA* gene unequivocally prove *RM12*'s plastid and the apicoplast are of the same origin and closely related to each other. Relationship between the *RM12*'s plastid and plastids of peridinin dinoflagellates is unclear. Both *psbA* nucleotide analyses show their sistership, but *RM12* branches close to stramenopiles in the *psbA t*ree derived from amino acids. The aminoacid-derived trees of both protein genes (*tufA, psbA*) turned out to be slightly better supported with *E. gracilis* branching within green algae in both cases. *PsbA* is not a good marker to infer the plastid origin: the trees are never supported at the deeper branch level. The position of *RM12*'s sequence is not much stable, but it always branches next to the dinoflagellate, stramenopile and haptophyte sequences suggesting its plastid is of a red algal origin. These results confirm previous findings (Moore et al. 2008) about the *RM12*'s plastid evolutionary relationships.

Notes on future analyses

LogDet method is known to be useful to deal with LBA. Aminoacid trees also included long branching species and therefore an analysis dealing with different codon usage and different site rates in different sequences would be useful. I suggest two possibilities to improve resolution of the trees. The first is to analyze the aminoacid dataset using the LogDet method (LdDist program), which is currently available on Linux platform. The other is to use the CAT model, a complicated ML model, which takes into account rate heterogeneity across sites of a dataset. The model in implemented in PhyloBayes, a program which runs on MacOS/Linux.

A stop codon in the tufA gene

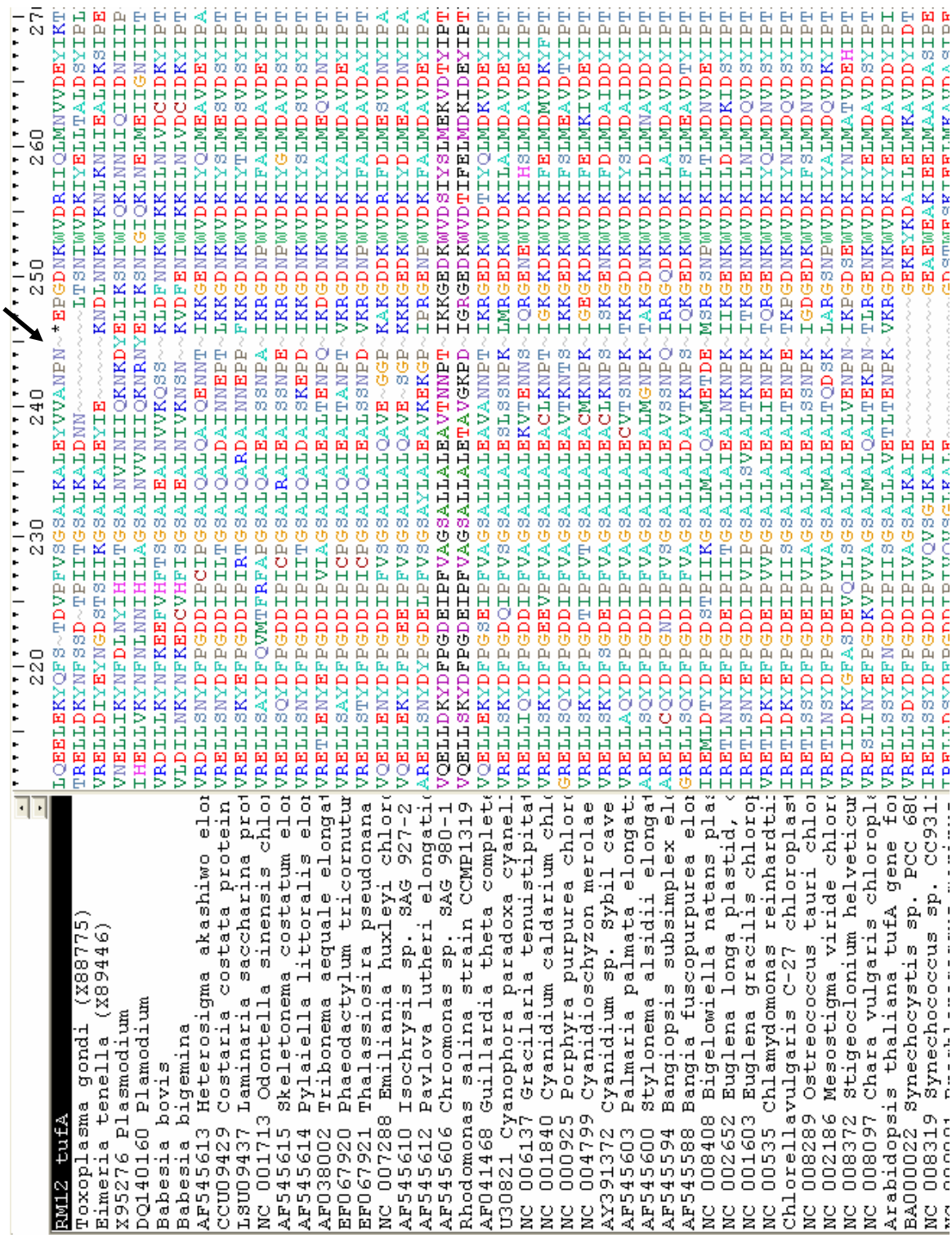
Another suggestion of *RM12's* plastid affinity to the apicoplast comes from the aminoacid sequences of *tufA* (Fig.11). An unsuspected stop codon appears in the middle of RM12's *tufA* sequence. The stop codon is UGA encoded. Both strands of five *tufA* clones were sequenced to reject the possibility of wrong base presence at one of the triplet positions. The presence of a UGA stop

codon suggests it may be the same case of an non-canonical genetic code as in the *psbA* gene. The aminoacid alignment is nevertheless not conservative at the site and sequence of no other apicomplexan is disrupted similarly. Interestingly, there is a UGG position encoding tryptophan in *RM12*'s sequence. There is no UGG-Trp position in the *psbA* dataset and therefore this is the first prove *RM12* encodes tryptophan with both UGG and UGA codons. The presence of non-canonical code in the gene also means *tufA* is encoded in the plastid genome. To prove this, DNA hybridization of the *RM12*'s *tufA* gene probe with the total DNA separated using PFGE was performed. The experiment was unsuccessful, providing only a weak and disputable signal (data not shown).

The tufA analysis of Hackett et al. (2004)

The most promising for future is the *tufA* analysis, because a homolog from a peridinin dinoflagellate may be identified. Hackett et al. (2004) have already published a *tufA* gene analysis with a putative homolog of the gene from *Alexandrium catenella* obtained by EST sequencing. *A. catenella* sequence groups with a coccidian sequence and authors conclude the analysis reflects evolutionary sistership of apicoplast and peridinin plastid (because of grouping of coccidian and. The sequence, which is supposed to be stored in the *A. catenella* EST database (GenBank) was extensively searched. No alignable homolog of *T. gondi*, *E. tenella* and *RM12*'s tufA sequences has been identified despite extensive search. Another serious issue refers to the analysis. *TufA* is a slowly evolving gene with easily identified homologs across bacteria and bacteria-derived organelles. Nuclear-encoded mitochondrial homologs of *tufA* are known from eukaryotes including e.g. ciliates (*Tetrahymena thermophila*, *Paramecium tetraurelia*, *Monoeuplotes crassus*) and red algae (*Cyanidioschyzon merolae*). The sequence of *A.catenella* is the longest sequence in the discussed tree. Considering the analysis contained only sequences from photosynthetic eukaryotes, there is no evidence of *A. catenella* sequence origin – it may be a mitochondrial homolog or a result of bacterial contamination. In such a case of a very distant sequence, the grouping of the *A.catenella* with Apicomplexa, the second longest branches in the tree, would not be surprising. Preliminary analyses of the dataset tested in this diploma showed artificial branching of ciliate mitochondrial *tufA* sequences with the apicoplast homologs in poorly sampled datasets (see the chapter Results). The analysis of Hackett et al. (2004) is rooted with a glaucocystophyte *C. paradoxa* giving no chance to distinguish between these possibilities. Moreover, the tree backbone is completely unresolved and the dinoflagellate-apicomplexan clade is not supported in two of the three branch support analyses made (neighbor joining and Bayesian analysis). Insignificant bootstrap support (62) for the clade appears in the maximum parsimony analysis, which is an unlikely method to reveal such kind of a phylogenetic artifact. I conclude, the published *tufA* analysis gives null evidence of relationship between the peridinin plastid and the apicoplast.

Fig. 11: *TufA* dataset with a stop codon in *RM12*'s sequence depicted with an arrow. The stop codon is UGA encoded and very probably substituted with an aminoacid through a non-canonical code usage. The aminoacid may be tryptophan as in the *psbA* gene. UGG encodes for tryptophan at the position 253 of the alignment.



The RM12's plastid genome

Size of the *RM12*'s plastid genome, ±130 kb, is comparable to sizes of other secondary plastid genomes (Table 2). Reduced plastids of Apicomplexa are extremely small ±35kb), similarly to the plastid of parasitic green alga *Helicosporidium*. *Ostreococcus tauri* and *B. natans* have usual number of plastid genes, but extraordinarily compacted genomes.

The results of the hybridization gave inexplicable smear at the bottom of gel. It is not known what the smear is, but one explanation might be presence of additional minicirles in the plastid. The plastid minicirles are known only from dinoflagellates. All the other plastid genomes are organised on a circular molecule. It is necessary to further test this hypothesis.

Table 2: Plastid genomes in eukaryotes and their respective nucleotide lengths (bp).

Parasite genomes are highlighted in bold; extraordinary compacted genomes of phototrophs are in red. Switch in organism life-style to parasitism or heterotrophy may result in complete elimination or in reduction of the plastid genome. Remnant plastids of non-photosynthetic organisms retain important metabolic functions. Part of the genes encoding the plastid protein machinery remains in the plastid genome whereas the rest is tranlocated to the nucleus and re-targeted as translated gene products.

SPECIES	GROUP	PL. GENOME SIZE	LIFE STRATEGY
Plasmodium falciparum	Apicomplexa	**34,682 bp**	**parasite**
Eimeria tenella	Apicomplexa	**34,750 bp**	**parasite**
Toxoplasma gondi	Apicomplexa	**34,996 bp**	**parasite**
Theileria parva	Apicomplexa	**39,579 bp**	**parasite**
Babesia bovis	Apicomplexa	**33,351 bp**	**parasite**
Odontella sinensis	Stramenopila	119,704 bp	phototroph
Phaeodactylum tricornutum	Stramenopila	117,369 bp	phototroph
Thalassiosira pseudonana	Stramenopila	128,814 bp	phototroph
Guillardia theta	Cryptophyta	121,524 bp	phototroph
Rhodomonas salina	Cryptophyta	135,854 bp	phototroph
Emiliania huxleyi	Haptophyta	105,309 bp	phototroph
Cyanidium caldarium	Rhodophyta	164,921 bp	phototroph
Cyanidioschyzon merolae	Rhodophyta	149,987 bp	phototroph
Gracilaria tenuistipitata	Rhodophyta	183,883 bp	phototroph
Porphyra purpurea	Rhodophyta	191,028 bp	phototroph
Chlamydomonas reinhardtii	Viridiplantae	203,828 bp	phototroph
Chlorella vulgaris	Viridiplantae	150,613 bp	phototroph
Helicosporidium sp.	Viridiplantae	**37,454 bp**	**parasite**
Leptosira terrestris	Viridiplantae	195,081 bp	phototroph
Nephroselmis olivacea	Viridiplantae	200,799 bp	phototroph
Oltmannsiellopsis viridis	Viridiplantae	151,933 bp	phototroph
Ostreococcus tauri	Viridiplantae	71,666 bp	phototroph
Pseudendoclonium akinetum	Viridiplantae	195,867 bp	phototroph
Scenedesmus obliquus	Viridiplantae	161,452 bp	phototroph
Stigeoclonium helveticum	Viridiplantae	223,902 bp	phototroph
Arabidopsis thaliana	Viridiplantae	154,478 bp	phototroph
Zea mays	Viridiplantae	140,384 bp	phototroph
Angiopteris evecta	Viridiplantae	153,901 bp	phototroph
Anthoceros formosae	Viridiplantae	161,162 bp	phototroph
Marchantia polymorpha	Viridiplantae	121,024 bp	phototroph
Cyanophora paradoxa	Glaucocystophyta	135,599 bp	phototroph
Euglena gracilis	Euglenophyta	143,171 bp	phototroph
Euglena longa	Euglenophyta	**73,345 bp**	**heterotroph**
Bigelowiella natans	Cercozoa	69,166 bp	phototroph

Conclusions

Phylogenetic analyses of four genes sequenced in this study showed *RM12's* plastid is closely related to the apicoplast. Relationship to the plastid of peridinin dinoflagellates is not clear from these analyses. *RM12*'s plastid genome size is about 130kb. The results strongly support previous findings that the apicoplast originated from a photosynthetic plastid. The apicoplast is in the spotlight of drug research against apicomplexan parasites. Sequencing of the *RM12*'s plastid would reveal its similarities to the apicoplast and could help to understand its origin and function more in the detail.

References

ADL, S. M., LEANDER, B. S., SIMPSON, A. G. B., ARCHIBALD, J. M., ANDERSON, O. R., BASS, D., BOWSER, S. S., BRUGEROLLE, G., FARMER, M. A., KARPOV, S., KOLISKO, M., LANE, C. E. , LODGE, D. J., MANN, D. G., MEISTERFELD, R., MENDOZA, L., MOESTRUP, O., MOZLEY STANDRIDGE, S.E., SMIRNOV, A. V. & SPIEGEL, F. (2007) Diversity, Nomenclature, and Taxonomy of Protists**.** *Syst. Biol.* 56(4), 684–689.

BEN ALI, A., DE BAERE, R., VAN DER AUWERA, G., DE WACHTER, R. & VAN DE PEER, Y. (2001). Phylogenetic relationships among algae based on complete large-subunit rRNA sequences. *International Journal of Systematic and Evolutionary Microbiology* 51, 737–749.

ARCHIBALD, J. M. & KEELING, P. J. (2002). Recycled plastids: a green movement in eukaryotic evolution. *Trends Genet* 18, 577–584.

AZEVEDO, C. (1989). Fine Structure of *Perkinsus atlanticus n. sp.* (Apicomplexa, Perkinsea) Parasite of the Clam Ruditapes decussatus from Portugal. *The Journal of Parasitology*, Vol. 75, No. 4., 627-635.

BACHVAROFF, T. R., CONCEPCION, G. T., ROGERS, C. R., HERMAN, E. M., & DELWICHE, C. F. (2004). Dinoflagellate expressed sequence tag data indicate
massive transfer of chloroplast genes to the nuclear genome. *Protist* 155, 65–78.

BACHVAROFF, T. R., PUERTA, M. V. S. & DELWICHE, CH. F. (2005). Chlorophyll c–Containing Plastid Relationships Based on Analyses of a Multigene Data Set with All Four Chromalveolate Lineages. *Mol. Biol. Evol*. 22(9), 1772–1782.

BALDAUF, S. L., ROGER, A. J., WENK-SIEFERT, I. & DOOLITTLE, W. F. (2000). A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290, 972–977.

BARBROOK, A. C. & HOWE, C. J. (2000). Minicircular plastid DNA in the
dinoflagellate *Amphidinium operculatum*. *Molecular and General Genetics* 263, 152–158.

BARBROOK, A. C., HOWE, C. J. & PURTON, S. (2005). Why are plastid genomes retained in non-photosynthetic organisms? *Trends in plant science* Vol. 11 No.2

BARTA, J. R. & THOMPSON R. C. A. (2006) What is *Cryptosporidium*? Reappraising its biology and phylogenetic affinities. *Trends in parasitology* Vol. 22 No.10

BEN ALI, A., DE BAERE, R., VAN DER AUWERA, G., DE WACHTER, R. & VAN DE PEER, Y. (2001). Phylogenetic relationships among algae based on complete large-subunit rRNA sequences. *Int J Syst Evol Microbiol* 51, 737–749.

BHATTACHARYA, D., YOON, H. S. & HACKETT, J. D. (2004). Photosynthetic
eukaryotes unite: endosymbiosis connects the dots. *Bioessays* 26, 50–60.

BLANCHARD, J. L. & HICKS, J. S. (1999). The non-photosynthetic plastid in malarial parasites and other apicomplexans is derived from outside the green plastid lineage. *J Eukaryot Microbiol* 46, 367–375.

BODYL, A. (2005) Do plastid-related characters support the chromalveolate hypothesis? *J Phycol* 41, 712–719.

BRUGEOLLE, G. (2002). *Cryptophagus subtilis*: a new parasite of cryptophytes affiliated with the Perkinsozoa lineage. *Europ. J. Protistol.* 37, 379–390.

BRUGEROLLE, G. (2002). *Colpodella vorax* : ultrastructure, predation, life-cycle, mitosis, and phylogenetic relationships. *Europ. J. Protistol.* 38, 113–125.

BUSHEK, D., DUNGAN, C. F. & LEWITUS, A. J. (2002). Serological Affinities of the Oyster Pathogen *Perkinsus marinus* (Apicomplexa) with Some Dinoflagellates (Dinophyceae). *J. Eukaryot. Microbiol.,* 49(1), 11–16.

CAI, X., FULLERB, A. J., MCDOUGALD, L. R. & ZHU, G. (2003). Apicoplast genome of the coccidian *Eimeria tenella*. *Gene* 321, 39–46.

CARRENO, R. A., DONALD, S. M., & BARTA, J. R. (1999). Cryptosporidium is more closely related to the gregarines than to coccidia as shown by phylogenetic analysis of apicomplexan parasites inferred using small-subunit ribosomal RNA gene sequences. *Parasitol Res* 85, 899±904.

CAVALIER-SMITH, T. (1999). Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J Eukaryot Microbiol* 46, 347–366.

CAVALIER-SMITH, T. (2000). Membrane heredity and early chloroplast evolution. *Trends in plant science*. Vol. 5, No. 4.

CAVALIER-SMITH, T. (2002). Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote–eukaryote chimaeras (meta-algae). *Phil. Trans. R. Soc. Lond.* B (2003) 358, 109–134.

CAVALIER-SMITH, T. (2002). Chloroplast Evolution: Secondary Symbiogenesis and Multiple LossesDispatch. *Current Biology*, Vol. 12, R62–R64.

CAVALIER-SMITH, T., CHAO, E. E. (2004). Protalveolate phylogeny and systematics and the origins of Sporozoa and dinoflagellates (phylum Myzozoa nom. nov.). *European Journal of Protistology* 40, 185–212.

CAVALIER-SMITH, T. (2003) Protist phylogeny and the high-level classification
of Protozoa. *Europ. J. Protistol.* 39, 338–348.

CATHLEEN A. C., ROBLEDO J. A. F. & VASTA, G. R. (2001). Fine Structure of Clonally Propagated In Vitro Life Stages of a *Perkinsus* sp. Isolated from the Baltic Clam Macoma balthica. *J. Eukaryot. Microbiol.,* 48(1), 38–51.

CHESNICK, J. M., KOOISTRA, W. H., WELLBROCK, U., MEDLIN, L. (1997). Ribosomal RNA analysis indicates a benthic pennate diatom ancestry for the endosymbionts of the dinoflagellates *Peridinium foliaceum* and *Peridinium balticium* (Pyrrhophyta). *J Euk Microbiol* 44, 314–320.

CHRUDIMSKY T. (2007). Molecular characterization of novel symbiotic alveolate isolated from corals. Master thesis. Univ. of South Bohemia

DAUGBJERG, N., HANSEN, G., LARSEN, J., & MOESTRUP, O. (2000). Phylogeny of some of the major genera of dinoflagellates based on ultrastructure and partial LSU rDNA sequence data, including the erection of three new genera of unarmoured dinoflagellates. *Phycologia* 39, 302–317.

DE LA HERRAN, R., GARRIDO-RAMOS, M. A., NAVAS, J., RUIZ REJON, C., RUIZ REJON, M. (1999). Molecular characterization of the ribosomal RNA gene region of *Perkinsus atlanticus*: its use in phylogenetic analysis and molecular diagnostics. P*arasitology* (2000), 120, 345±353.

DE KONING, A. P., KEELING, P. J. (2006). The complete plastid genome sequence of the parasitic green alga *Helicosporidium* sp. is highly reduced and structured. *BMC Biology*, 4:12.

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

DOUGLAS, S. E., PENNY, S. L. (1999). The plastid genome of the cryptophyte alga, *guillardia theta.* Complete sequence and conserved synteny groups confirm its common ancestry with Red Algae. *J Mol Evol* 48, 236–244.

DOUGLAS, S., ZAUNER, S., FRAUNHOLZ, M., BEATON, M., PENNY, S., DENG, L. T., WU, X., REITH, M., CAVALIER-SMITH, T., MAIER, U. G. (2001). The highly reduced genome of an enslaved algal nucleus. *Nature* 410, 1091–1096.

ELLIS, J. T., AMOYAL, G., RYCE, C., HARPER, P. A. W., CLOUGH, K. A., HOMAN, W. L. & BRINDLEY, **P. J. (1998)** Comparison of the large subunit ribosomal DNA of Neospora and Toxoplasma and development of a new genetic marker for their differentiation based on the D2 domain, *Molecular and Cellular Probes* 12, 1–13.

EMBLEY, T. M. & HIRT, R. P. (1998). Early branching eukaryotes? *Curr Opin Genet Dev* 8, 624–629.

ESCALANTE, A. A., AYALA, F. J. (1995). Evolutionary Origin of Plasmodium and Other Apicomplexa Based on rRNA Genes. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 92, No. 13, 5793-5797.

FAST, N. M., KISSINGER, J. C., ROOS, D. S. & KEELING, P. J. (2001). Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol Biol Evol* 18, 418–426.

FAST, N. M., XUE, L., BINGHAM, S. & KEELING, P. J. (2002). Re-examining alveolate evolution using multiple protein molecular phylogenies. *J Eukaryot Microbiol* 49, 30–37.

FENSOME, R. A., SALDARRIAGA, J. F. AND TAYLOR, F. J. R. (1999). Dinoflagellate phylogeny revisited: reconciling morphological and molecular based phylogenies. *Grana* 38, 66–80.

FOTH, B.J., MC FADDEN, G.I. (2003) The apicoplast: a plastid in Plasmodium falciparum and other Apicomplexan parasites. *Int Rev Cytol* 224, 57–110.

FUNES, S., DAVIDSON, E., REYES-PRIETO, A., MAGALLON, S., HERION, P., KING, M. P. & GONZALEZ-HALPHEN, D. (2002). A green algal apicoplast ancestor. *Science* 298, 2155.

FIGUERAS, A., LORENZO, G., ORDA´S, M.C., GOUY, M. AND NOVOA, B. (2000). Sequence of the Small Subunit Ribosomal RNA Gene of Perkinsus atlanticus-like Isolated from Carpet Shell Clam in Galicia, Spain. *Mar. Biotechnol.* 2, 419–428.

GAJADHAR, A. A., MARQUARDT, W. C., HALL, R., GUNDERSON, J., ARIZTIA-CARMONA, E. V., & SOGIN, M. L. (1991). Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata* and C*rypthecodinium cohnii* reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. *Molecular and Biochemical Parasitology* 45, 147–154.

GARDNER, M. J., et al. (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* 41, 498–511.

GAST, R. J., MORAN, D. M., BEAUDOIN, D. J., BLYTHE, J. N., DENNETT, M. R. (2006). Abundance of a novel Dinoflagellate phylotype in the Ross sea, *Antarctica. J. Phycol* 42, 233–242.

GLEESON, M. T. (2000). The plastid in Apicomplexa: What use is it? *International Journal for Parasitology* 30, 1053±1070.

GOCKEL, G. & HACHTEL, W. (2000). Complete Gene Map of the Plastid Genome of the Nonphotosynthetic Euglenoid Flagellate. *Astasia longa. Protist*, Vol. 151, 347–351.

GRAUVOGEL, REECE, K. S., BRINKMANN, H., PETERSEN, J. (2007). Plastid Isoprenoid Metabolism in the Oyster Parasite Perkinsus marinus Connects Dinoflagellates and Malaria Pathogens - New Impetus for Studying Alveolates. *J Mol Evol* 65, 725–729.

GRIBBLE, K. E. & ANDERSON, D. M. (2006). Molecular phylogeny of the heterotrophic dinoflagellates, protoperidinium, diplopsalis and preperidinium (dinophyceae), inferred from large subunit rdna1 j. *Phycol.* 42, 1081–1095.

GUINDON, S. & GASCUEL, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Bio*l 52, 696–704.

GUINDON, S., LETHIEC F., DUROUX P., GASCUEL O. (2005). PHYML Online -a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* 33 (Web Server issue): W557-9.

GUNDERSON, J. H., GOSS, S. H., COATS D. W. (1999). The phylogenetic position of *Amoebophyra* sp. from *Gymnodinium sanguineum*. *Journal of Eukaryotic Microbiology* 46, 195–197.

HALL, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT., *Nuc. Acid. Res*., 41, 95-98.

HACKETT, J. D., MARANDA, L., YOON, H. S., BHATTACHARYA, D. (2003) Phylogenetic evidence for the cryptophyte origin of the plastid of Dinophysis (Dinophysiales, Dinophyceae). *J Phycol* 39, 440–448.

HACKETT, J. D., YOON, H. S., SOARES, M. B., BONALDO, M. F., CASAVANT, T., SCHEETZ, T. E., NOSENKO, T. & BHATTACHARYA, D. (2004). Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. *Current Biology* 14, 213–218.

HACKETT, J. D., ANDERSON, D. M., ERDNER, D. L., BHATTACHARYA, D. (2004). Dinoflagellates: A remarkable evolutionary Experiment. *American Journal of Botany* 91(10), 1523–1534.

HACKETT, J. D., SCHEETZ, T. E., YOON, H. S., SOARES, M. B., BONALDO, M. F., CASAVANT, T. L. & BHATTACHARYA, D. (2005). Insights into a dinoflagellate genome through expressed sequence tag analysis. *BMC Genomics* 6, 80
HARPER, J. T. & KEELING, P. J. (2003). Nucleus-encoded, plastidtargeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. *Mol Biol Evol* 20, 1730–1735.

HOWE, C. J., BARBROOK, A. C., KOUMANDOU, V. L., NISBET, R. E., SYMINGTON, H. A., & WIGHTMAN, T. F. (2003). Evolution of the chloroplast genome. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 358, 99–106.

IIDA, K., TAKISHITA K., OHSHIMA K., INAGAKI Y. (2007). Assessing the monophyly of chlorophyll-c containing plastids by multi-gene phylogenies under the unlinked model conditions. *Molecular Phylogenetics and Evolution* 45, 227–238.

INAGAKI, Y., DACKS, J. B., DOOLITTLE W. F., WATANABE, K. I. & OHAMA T. (2000). Evolutionary relationship between dinoflagellates bearing obligate diatom endosymbionts: Insight into tertiary endosymbiosis. *International Journal of Systematic and Evolutionary Microbiology* 50, 2075–2081.

INAGAKI, Y., SIMPSON A., DACKS J., ROGER A. (2004). Phylogenetic artifacts can be caused by leucine, serine, and arginine codon usage heterogeneity: dinoflagellate plastid origins as a case study. *Syst Biol* 53, 582–593.

INAGAKI, Y. & ROGER, A. J. (2006). Phylogenetic estimation under codon models can be biased by codon usage heterogeneity. *Molecular Phylogenetics and Evolution* 40, 428–434.

ISHIDA, K., & GREEN, B. R. (2002). Second- and third-hand chloroplasts in dinoflagellates: phylogeny of oxygen-evolving enhancer 1 (*Psb*O) protein reveals replacement of a nuclear-encoded plastid gene by that of a haptophyte tertiary endosymbiont. *Proceedings of the National Academy of Sciences, USA* 99, 9294–9299.

JANSON, S., & GRANE´LI, E. (2003). Genetic analysis of the psbA gene from single cells indicates a cryptomonad origin of the plastid in *Dinophysis* (Dinophyceae). *Phycologia* 42, 473–477.

JIROUTOVA, K., HORAK, A., BOWLER, C. OBORNIK, M. (2007). Tryptophan Biosynthesis in Stramenopiles: Eukaryotic Winners in the Diatom Complex Chloroplast. *J Mol E vol* 65, 496–511.

KATOH K., KUMA K., TOH H., MYATA T. (2005). MAFFT version 5: Improvement in acuracy of multiple sequence alignment. *Nuc Acid Res* 33, 511-518.

KATOH K., MISAWA K., KUMA K., MYATA T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nuc Acid Res* 30, 3059-3066.

KEELING, P. J. (2004a). Reduction and Compaction in the Genome of the Apicomplexan Parasite *Cryptosporidium parvum. Developmental Cell* 614.

KEELING, P. J. (2004b). Diversity and evolutionary history of plastids and their hosts. *American Journal of Botany* 91(10), 1481–1493.

KOHLER, S., DELWICHE, C. F., DENNY, P. W., TILNEY, L. G., WEBSTER, P., WILSON, R. J. M., PALMER, J. D. & ROOS, D. S. (1997). A plastid of probable green algal origin in Apicomplexan parasites. *Science* 275, 1485–1489.

KÖHLER, S., DELWICHE, CH. F., DENNY, P. W., TILNEY, L. G., WEBSTER, P., WILSON, R. J. M., PALMER, J. D., ROOS, D. S. (1997). A Plastid of Probable Green Algal Origin in Apicomplexan Parasites. *Science* Vol. 275, No. 5305,1485-1489.

KOIKEA, K., SEKIGUCHIB, H., KOBIYAMAA, A., TAKISHITAC, K., KAWACHID, M., KOIKEA, K., & OGATA, T. (2005). A Novel Type of Kleptoplastidy in Dinophysis (Dinophyceae): Presence of Haptophyte-type Plastid in Dinophysis mitra. *Protist*, Vol. 156, 225-237.

KUVARDINA, O. N., LEANDER, B. S., ALESHIN, V. V., MYL'NIKOV, A. P., KEELING, P. J. & SIMDYANOV, T. G. (2002). The phylogeny of colpodellids (Alveolata) using small subunit rRNA gene sequences suggests they are the free-living sister group to apicomplexans. *J Eukaryot Microbiol* 49, 498–504.

LAATSCH, T., ZAUNER, S., STOEBE-MAIER, B., KOWALLIK, K. V., & MAIER, U. G. (2004). Plastid-Derived Single Gene Minicircles of the Dinoflagellate Ceratium horridum Are Localized in the Nucleus. *Molecular Biology and Evolution* vol. 21 no. 7.

LANG-UNNASCH, N., REITHB, M. E., MUNHOLLAND, J., BARTAC, J. R. (1998). Plastids are widespread and ancient in parasites of the phylum Apicomplexa. *International Journal for Parasitology* 17 632-643.

LANG-UNNASCH, N. & AIELLO D. P. (1999). Sequence evidence for an altered genetic code in the Neospora caninum plastid. *International Journal for Parasitology* 29, 1557±1562.

LEANDER, B. L., CLOPTON, R. E. & KEELING, P. J. (2003). Phylogeny of gregarines (Apicomplexa) as inferred from small-subunit rDNA and b-tubulin. *Int J Syst Evol Microbiol* 53, 345–354.

LEANDER B. S. & KEELING, P. J. (2003). Morphostasis in alveolate evolution. *Trends in Ecology and Evolution* Vol.18, No 8.

LEANDER, B. S. & KEELING, P. J. (2004). Early evolution of dinoflagellates and apicomplexans inferred from HSP90 and actin phylogeny. *J Phycol* 40, 341–350.

MC FADDEN, G. I. & WALLER, R. F. (1997). Plastids in parasites of humans. *Bioessays* 19, 1033–1040.

MARIN, B., NOWACK, E. C. M. & MELKONIAN, M. (2005). A Plastid in the Making: Evidence for a Second Primary Endosymbiosis. *Protist*, Vol. 156, 425–432.

MOORE, R. B. (2006). Molecular Ecology and Phylogeny of Protistan Algal Symbionts from Corals. *Phd thesis*. School of Molecular and Microbial Biosciences, The University of Sydney

MOORE, R. B., OBORNIK, M., JANOUSKOVEC, J., CHRUDIMSKY, T., VANCOVA, M., GREEN, D. H., WRIGHT, S. W., DAVIES, N. W., BOLCH, C. J. S., HEIMANN, K., SLAPETA, J., HOEGH-GULDBERG, O., LOGSDON JR. J. M., CARTER, D. A. (2008). A photosynthetic alveolate closely related to apicomplexan parasites. I*n press*

MORRISON, D. A. & ELLIS, J. T. (1997). Effects of Nucleotide Sequence Alignment on Phylogeny Estimation: A Case Study of 18s rDNAs of Apicomplexa. *Mol. Biol. Evol.* 14(4), 428-441.

MUGRIDGE, N. B., MORRISON, D. A., HECKEROTH, A. R., JOHNSON, A. M., TENTER, A. M. (1999). Phylogenetic analysis based on full-length large subunit ribosomal RNA gene sequence comparison reveals that Neospora caninum is more closely related to Hammondia heydorni than to *Toxoplasma gondii. International Journal for Parasitology* 29, 1545±1556.

MURRAY, S., FLØ-JORGENSEN, M. F., HO, S. Y., PATTERSON, D. J., JERMIIN, L. S. (2005). Improving the analysis of dinoflagellate phylogeny based on rDNA. *Protist* 156, 269–286.

NOREN, F., MOESTRUP, Ø. & REHNSTAM-HOLM, A. S. (1999). Parvilucifera infectans Nore´n et Moestrup gen. et sp. nov. (Perkinsozoa phylum nov.): a parasitic flagellate capable of killing toxic microalgae. *Eur J Protistol* 35, 233–254.

NOZAKI, H., MATSUZAKI, M., TAKAHARA, M., MISUMI, O., KUROIWA, O., HASEGAWA, M., SHIN-I, T., KOHARA, Y., OGASAWARA, N. & KUROIWA, T. (2003). The phylogenetic position of red algae revealed by multiple nuclear genes from mitochondria-containing eukaryotes & an alternative hypothesis on the origin of plastids. *Journal of Molecular Evolution* 56, 485–497.

OKAMOTO, N. & INOUYE, I. (2005b). A Secondary Symbiosis in Progress? *Science* vol. 310, 287.

OKAMOTO, N. & INOUYE, I. (2005a). The Katablepharids are a Distant Sister Group of the Cryptophyta: A Proposal for Katablepharidophyta Divisio Nova/ Kathablepharida Phylum Novum Based on SSU rDNA & Beta-Tubulin Phylogeny. *Protist* 156, 163.

PATRON, N. J., WALLER, R. F., ARCHIBALD, J. M. & KEELING J. P. (2005**).** Complex Protein Targeting to Dinoflagellate Plastids. *J. Mol. Biol.* 348, 1015–1024.

PATRON, N. J., WALLER, R. F., KEELING, P. J. (2006). A tertiary plastid uses genes from two endosymbionts. *J Mol Biol* 357, 1373–1382.

PATTERSON, D. J. (1999). The Diversity of Eukaryotes. *The American naturalist* vol. 65.

PERKINS, F. O., BARTA, J. R., CLOPTON, R. E., PEIRCE, M. A., UPTON, S. J. (2000). Phylum Apicomplexa. *The Illustrated Guide to the Protozoa*, vol I. , second ed., Allen Press, Lawrence, KS 66044 USA.

PETERSEN, J., TEICH, R., BRINKMANN, H., CERFF, R. (2006) A ‘‘green’’ phosphoribulokinase in complex algae with red plastids: evidence for a single secondary endosymbiosis leading to haptophytes, cryptophytes, heterokonts & dinoflagellates. *J Mol Evol* 62, 143–157.

REECE, K. S., SIDDALL, M. E., BURRESON, E. M. & GRAVES, J. E. (1997). Phylogenetic analysis of Perkinsus based on actin gene sequences. *J Parasitol* 83, 417–423.

ROBBENS, S., DERELLE, E., FERRAZ, C., WUYTS, J., MOREAU, H., VAN DE PEER, Y. (2007). The complete chloroplast & mitochondrial DNA sequence of *Ostreococcustauri*: organelle genomes of the smallest eukaryote are examples of compaction. *Mol Biol Evol* 25, 17.

RODRIGUEZ-EZPELETA, N., BRINKMANN, H., BUREY, S. C., ROURE, B., BURGER, G., LÖFFELHARDT, W., BOHNERT, H. J., PHILIPPE, H., LANG, B. F. (2005). Monophyly of primary photosynthetic eukaryotes: green plants, red algae, & glaucophytes. *Curr Biol* 15, 1325–1330.

ROGERS, M. B., GILSON, P. R., VANESSA, S., GEOFFREY, U., MCFADDEN, I., KEELING, P. J. (2007). The Complete Chloroplast Genome of the Chlorarachniophyte Bigelowiella natans. Evidence for Independent Origins of Chlorarachniophyte & Euglenid Secondary Endosymbionts. *Mol. Biol. Evol.* 24, 54–62.

SALDARRIAGA, J. F., TAYLOR, F.J., KEELING, P.J., CAVALIER-SMITH, T. (2001). Dinoflagellate nuclear SSU rRNA phylogeny suggests multiple plastid losses & replacements. *J Mol Evol* 53, 204–213.

SALDARRIAGA, J. F. , MC EWAN, M. L., FAST, N. M., TAYLOR, F. J. R. & KEELING, P. J. (2003). Multiple protein phylogenies show that Oxyrrhis marina & Perkinsus marinus are early branches of the dinoflagellate lineage. *Int J Syst Evol Microbiol* 53, 355–365.

SALDARRIAGA, J. F., TAYLOR, F. J. R., CAVALIER-SMITH, T., MENDEN-DEUER, S., KEELING, P. J. (2004). Molecular data & the evolutionary history of dinoflagellates. *Eur J Protist* 40, 85–111.

SAMBROOK, J., RUSSELL, D. W. (2001). Molecular cloning, third edition. *Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.*

SANCHEZ-PUERTA, M. V., LIPPMEIER, J. C., APT, K. E., DELWICHE, C. F. (2007). Plastid genes in a non-photosynthetic dinoflagellate. *Protist* 158, 105–117.

SHALCHIAN-TABRIZI, K., MINGE, M.A., CAVALIER-SMITH, T., NEDREKLEPP, J.M., KLAVENESS, D., JAKOBSEN, K.S. (2006). Combined heat shock protein 90 & ribosomal RNA sequence phylogeny supports multiple Replacements of Dinoflagellate Plastids *J. Eukaryot. Microbiol.* 53, 217–224.

SIDDALL, M. E., REECE, K. S., GRAVES, J. E. & BURRESON, E. M. (1997). ‘Total evidence’ refutes the inclusion of Perkinsus species in the phylum Apicomplexa. *Parasitology* 115, 165–176.

SIDDALL, M. E., REECE, K. S., NERAD, T. A. & BURRESON, T. M. (2001). Molecular Determination of the Phylogenetic Position of a Species in the Genus *Colpodella* (Alveolata). *American Museum Novitates* 3314, 1–12.

STELTER K., EL-SAYED N. M., SEEBER F. (2007). The expression of a plant type ferredoxin redox system provides molecular evidence for a plastid in the early dinoflagellate Perkinsus marinus. *Protist* 158, 119–130.

SWOFFORD, D. L. (2002). Phylogenetic Analysis Using Parsimony (&other methods). *Sunderl&, MA: Sinauer Associates.*

TAKISHITA, K., KOIKE, K., MARUYAMA, T. AND OGATA, T. (2002). Molecular evidence for plastid robbery (kleptoplastidy) in *Dinophysis*, a dinoflagellate causing diarrhetic shellfish poisoning. *Protist* 153, 293–302.

TAKISHITA K. , UCHIDA A. (1999). Molecular cloning and nucleotide sequence analysis of psbA from the dinoflagellates: Origin of the dinoflagellate plastid Phycological Research 47, 207–216.

TAKISHITA K., ISHIDA K. AND MARUYAMAA T. (2004). Phylogeny of Nuclear-Encoded Plastid-Targeted GAPDH Gene Supports Separate Origins for the Peridinin- and the Fucoxanthin Derivative-Containing Plastids of Dinoflagellates. *Protist,* Vol. 155, 447–458.

TELES-GRILO M. L., TATO-COSTA J., DUARTE S. M., MAIA A., CASAL G., AZEVEDO C. (2007). Is there a plastid in Perkinsus atlanticus (phylum Perkinsozoa)? *Eur J Protistol* 43, 163–167.

TENGS, T., DAHLBERG, O., SHALCHIAN-TABRIZI, K., KLAVENESS, D., RUDI, K., DELWICHE, C., JAKOBSEN, K. (2000). Phylogenetic analyses indicate that the 190-hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Mol Biol Evol* 17, 718–729.

TOSO, M. A. & OMOTO, C. K. (2007). Gregarina niphandrodes may Lack Both a Plastid Genome and Organelle. *J. Eukaryot. Microbiol.* 54(1), 66–72.

TYLER, B. M. ET AL. (2006). Phytophthora Genome Sequences Uncover Evolutionary Origins and Mechanisms of Pathogenesis. *Science* vol 313, 1261-65.

VAN DE PEER, Y. & DE WACHTER, R. (1997). Evolutionary relationships among the eukaryotic crown taxa taking into account site-to-site variation in 18S rRNA. *J Mol Evol* 45, 619–630.
VAN DE PEER, Y., BALDAUF, S. L., DOOLITTLE, W. F. & MEYER, A. (2000). An updated and comprehensive rRNA phylogeny of (crown) eukaryotes based on rate-calibrated evolutionary distances. *J Mol Evol* 51, 565–576.

WALLER R. F. ET AL. (1998). Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum. Proc. Natl. Acad. Sci. USA v*ol. 95, 12352–12357.

WALLER, R. F., KEELING, P. J., VAN DOOREN, G. G. & MCFADDEN, G. I. (2003). Comment on ‘‘A green algal apicoplast ancestor.’’ *Science* 301, 49.

WALLER R. & MCFADDEN G. I. (2004). The Apicoplast: A Review of the Derived Plastid of Apicomplexan Parasites. *Curr. Issues Mol. Biol.* 7, 57-80.

WATANABE, M. M., SASA, T., SUDA, S., INOUYE, I., & TAKICHI, S. (1991). Major carotenoid composition of an endosymbiont is a green dinoflagellate, *Lepidodinium viride*. *Journal of Phycology* 27, 75.

WILLIAMSON, D. H., GARDNER, M. J., PREISER, P., MOORE, D. J., RANGACHARI, K. & WILSON, R. J. (1994). The evolutionary origin of the 35 kb circular DNA of Plasmodium falciparum: new evidence supports a possible rhodophyte ancestry. *Mol Gen Genet* 243, 249–252.

WILSON, R. J. M. , DENNY P., PREISER, P., RANGACHARI, K., ROBERTS, K., ROY, A., WHYTE A., STRATH, M., MOORE, D. J., MOORE, P. W. & WILLIAMSON D. H. (1996). Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J. Mol. Biol.* 261, 155–172.

WILSON, R. J. M. & WILLIAMSON D. H. (1997). Extrachromosomal DNA in the Apicomplexa. *Microbiology & molecular biology reviews*, Mar. 1997, 1–16.

YOON, H. S. , HACKETT, J. D. & BHATTACHARYA, D. (2002). A single origin of the peridinin- & fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis*. Proc Natl Acad Sci USA* 99, 11724–11729.

YOON, H. S. , HACKETT, J. D. , PINTO, G. & BHATTACHARYA, D. (2002). The single, ancient origin of chromist plastids. *Proc Natl Acad Sci USA* 99, 15507–15512.

YOON, H. S., HACKETT, J., CINIGLIA, C., PINTO, G. & BHATTACHARYA, D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. *Molecular Biology & Evolution* 21, 809–818.

YOON H. S. , REYES-PRIETO A. , MELKONIAN M. & BHATTACHARYA D. (2006). Minimal plastid genome evolution in the *Paulinella* endosymbiont. *Current Biology* Vol 16 No 17.

YOON H. S. , HACKETT J. D. , VAN DOLAH F. M., NOSENKO T., LIDIE K. L. & BHATTACHARYA D. (2005). Tertiary Endosymbiosis Driven Genome Evolution in Dinoflagellate Algae. Mol. Biol. Evol. 22(5), 1299–1308.

ZARDOYA R., COSTAS E., LOPEZ-RODA V., GARRIDO-PERTIERRA A., BAUTISTA J. M. (1995). Revised dinoflagellate phylogeny inferred from molecular analysis of large-subunit ribosomal RNA gene sequences. *J Mol Evol* 41, 637–645.

ZAUNER S., FRAUNHOLZ M., WASTL J., PENNY S., BEATON M., CAVALIER-SMITH T., MAIER U. G., DOUGLAS S. (2000). Chloroplast Protein & Centrosomal Genes, a tRNA Intron, & Odd Telomeres in an Unusually Compact Eukaryotic Genome, the Cryptomonad Nucleomorph*. Proceedings of the National Academy of Sciences of the United States of America*, Vol. 97, No. 1, 200-205.

ZHANG, Z., B. R. GREEN, & T. CAVALIER-SMITH. (1999). Single gene circles in dinoflagellate chloroplast genomes. *Nature* 400, 155–159.

ZHANG, Z., GREEN, B. R. & CAVALIER-SMITH, T. (2000). Phylogeny of ultra-rapidly evolving dinoflagellate chloroplast genes: a possible common origin for sporozoan & dinoflagellate plastids. *J Mol Evol* 51, 26–40.

ZHANG, Z., CAVALIER-SMITH, T. & GREEN B. R. (2002). Evolution of dinoflagellate unigenic minicircles & the partially concerted divergence of their putative replicon origins. *Molecular Biology & Evolution* 19, 489–500.

ZHANG H., BHATTACHARYA D., LIN C. S. (2005). Phylogeny of dinoflagellates based on mitochondrial cytochrome b & nuclear small subunit rDNA sequence comparisons. J Phycol 41,411–420.

ZHANG H., BHATTACHARYA C. D., LIN C. S. (2007). A Three-Gene Dinoflagellate Phylogeny Suggests Monophyly of Prorocentrales & a Basal Position for *Amphidinium* & *Heterocapsa. J Mol Evol* (2007) 65, 463–474.

ZHU, G., MARCHEWKA, M. J. & KEITHLY, J. S. (2000A). *Cryptosporidium parvum* appears to lack a plastid genome. *Microbiology* 146, 315±321.

ZHU G., KEITHLY J.S & PHILIPPE H. (2000B). What is the phylogenetic position of *Cryptosporidium* ? *International Journal of Systematic & Evolutionary Microbiology* 50, 1673–1681.