



HAL
open science

Origins and early evolution of photosynthetic eukaryotes

Rafael Isaac Ponce Toledo

► **To cite this version:**

Rafael Isaac Ponce Toledo. Origins and early evolution of photosynthetic eukaryotes. Populations and Evolution [q-bio.PE]. Université Paris-Saclay, 2018. English. NNT : 2018SACLS047 . tel-01760725

HAL Id: tel-01760725

<https://theses.hal.science/tel-01760725v1>

Submitted on 6 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



NNT : 2018SACLS047

Origins and early evolution of photosynthetic eukaryotes

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Université Paris-Sud

École doctorale n°577 Structure et
Dynamique des Systèmes Vivants (SDSV)
Spécialité de doctorat: Science de la Vie et de la Santé

Thèse présentée et soutenue à Orsay, le 5 mars 2018, par :

Rafael Isaac Ponce Toledo

Composition du Jury :

Pierre Capy

Professeur, Université Paris-Sud (UMR 9191)

Président

Line Le Gall

Maître de Conférences, MNHN (UMR 7205)

Rapportrice

Marc-André Selosse

Professeur, MNHN (UMR 7205)

Rapporteur

Philippe Lopez

Professeur, UPMC (UMR 7138)

Examineur

David Moreira

Directeur de recherche, Université Paris-Sud (UMR 8079)

Directeur de thèse

Philippe Deschamps

Chargé de recherche, Université Paris-Sud (UMR 8079)

Co-encadrant

ACKNOWLEDGEMENTS

When I arrived in France I knew that my stay in this country would bring me wonderful experiences and it did. I want to thank all the people that have been with me in this professional and personal journey.

Quiero agradecer a mi familia que desde siempre me ha apoyado en todo y ha sido una fuente inagotable de amor y cariño. A mi **madre**, una mujer para quien no tengo palabras para agradecer todo lo que ha hecho por mí y por mis hermanos. A ti que siempre has sido mi mejor aliada, mi mayor motor. Gracias por enseñarme el valor de la libertad. Te amo.

Gracias a mis hermanos **Nancy, Blanca y Rubén**, que aunque nuestro camino juntos no siempre ha sido fácil, nunca he dudado del amor inmenso que sienten por mí, el mismo que yo siento por ustedes. Siempre han creído en mí y junto a ustedes he aprendido a ser una mejor persona. Los amo.

A mi **tía Muñe**, una parte muy importante en mi vida, por haber estado ahí en todo momento.

Gracias **papá** por tus consejos, siempre te recordaré.

Un grand merci à mes directeurs de thèse, **David et Philippe**, pour votre aide constante, pour m'avoir donné de votre temps chaque fois que j'en avais besoin et pour m'avoir accompagné vers le monde de la protistologie et les endosymbioses : deux sujets passionnants même s'ils n'ont pas toujours été faciles à étudier. Merci pour ces échanges et d'avoir partagé avec moi vos grandes connaissances et expériences tout au long de ma thèse, ils ont été très fructueux pour ma formation scientifique.

Merci à tous les gens de mon bureau, y compris ceux qui ne sont plus là : **Marie, Kamil et Aurélien** avec qui j'ai partagé de bons moments et qui m'ont appris à survivre à la bureaucratie française. Gràcies a **Guifré**, mi stalker favorito, por todas las charlas cafeinadas sobre la política catalana. Merci à la « nouvelle » génération du labo : **Jodie** (bonne chance avec Dallol !), **Gwendoline** (désolé pour tout le bruit dans le labo !),

Guillaume (le bio-informaticien multi-tâche !) et **Luis** (creo que ya es tarde para pedir una orden de restricción para ti :P).

Gracias a **Puri** por sus críticas acertadas para mejorar el trabajo en el laboratorio, los artículos y las presentaciones. Je veux remercier également **Paola, Ludwig et Hélène** avec qui je n'ai pas directement travaillé mais qui m'ont toujours aidé, et ont fait du laboratoire un espace plus agréable.

Una parte fundamental de estos años en Francia y en el doctorado han sido los amigos que he hecho por el camino. Juntos hemos vivido momentos maravillosos que siempre recordaré y espero sigamos contruyendo más. Quiero agradecer por orden de aparición a:

Anthony, contigo conocí más de Perú y aprendí de la riqueza latinoamericana. También aprendí que puedo ser consejero amoroso :) .Por cierto, ¡todavía me debes un buen ceviche!

Joan, nuestra amistad es de las cosas más lindas que me ha dejado Europa, nuestras aventuras, nuestros viajes (¡aunque nos peleemos!). Gracias por todo.

Juan, eres una fuente interminable de luz, gracias por todos los momentos que hemos pasado juntos.

Diego y Mario, el dúo dinámico que hizo de Barcelona una de mis ciudades preferidas. Si el Eixample pudiera hablar... no, mejor que no hable.

Merci à **Aimé** pour toutes nos discussions philosophiques autour d'une ou plusieurs bières. Nous n'avons pas pu changer le monde mais nous nous sommes bien amusés.

Tomás, gracias por formar parte de mi vida, y aunque el amor te llevó a tierras celtas, hiciste de París una ciudad menos gris.

Alex, gracias por tu amistad y el tiempo que hemos pasado juntos. Es bueno encontrar alguien con quien compartir la curiosidad.

Gracias a todas y cada una de las chicas del gueto hispanohablante: **Magda, Raquel, Laura, Adriana e Irene**. Fueron una parte importante de las amistades que hice durante el doctorado y con las que compartí más que la comida. Cantine? 12h45?

Merci à **Pierre Capy, Philippe Lopez, Line Le Gall et Marc-André Selosse** d'avoir accepté faire partie de mon jury de thèse.

Merci à tous les membres du **laboratoire d'Écologie, Systématique et Évolution**. Vous avez vraiment été tous très sympas avec moi pendant ces dernières années, et vous avez fait de mon séjour une belle expérience.

Finalmente, un agradecimiento especial a **Richard**. Gracias por el apoyo en los malos momentos, gracias por todas las risas que hemos compartido en los buenos. Gracias por tu (enorme) paciencia conmigo y por todo el amor que me has dado desde el primer día. Gracias por ser mi compañero de aventuras y locuras. Gracias especialmente por estos últimos meses, estoy muy contento de haberte conocido.

INDEX

1. INTRODUCTION	1
1.1 Endosymbiotic theory: a historical perspective.....	2
1.1.1 Pre-molecular era.....	2
1.1.1.1 Symbiosis: the origin of an idea.....	2
1.1.1.2 Constantin Mereschkowsky's symbiogenetic theory.....	3
1.1.1.2.1 The origin of the nucleus.....	5
1.1.1.2.2 Mitochondria.....	7
1.1.2 Molecular era.....	8
1.1.2.1 Lynn Margulis and the renaissance of the symbiogenetic theory.....	8
1.2. Primary photosynthetic eukaryotes.....	10
1.2.1 Origin of the chloroplast: from cyanobacterium to organelle.....	10
1.2.1.1 Endosymbiotic Gene Transfer.....	12
1.2.1.2 Bacterial genes in primary photosynthetic eukaryotes.....	14
1.2.1.3 Host-derived genes in the chloroplast proteome.....	15
1.2.1.4 TIC/TOC complex.....	16
1.2.1.5 <i>Ménage-à-trois</i> hypothesis.....	16
1.2.1.5.1 Was there a chlamydioplast?.....	19
1.2.1.6 Early- or late-branching cyanobacterium?.....	19
1.2.2 Archaeplastida lineages.....	21
1.2.2.1 Viridiplantae.....	21
1.2.2.2 Rhodophyta.....	23
1.2.2.3 Glaucophyta.....	24
1.2.3 <i>Paulinella chromatophora</i> , a second primary plastid-bearing lineage.....	24
1.3 Origin and evolution of eukaryotic lineages with complex plastids.....	26
1.3.1 Secondary green plastids.....	27
1.3.1.1 Chlorarachniophytes.....	27
1.3.1.1.1 Nucleomorph.....	27
1.3.1.1.2 Plastid.....	28
1.3.1.2 Euglenophytes.....	29

1.3.1.2.1 Plastid.....	30
1.3.2 Eukaryotic lineages with red algal-derived plastids.....	30
1.3.2.1 The Chromalveolate hypothesis.....	30
1.3.2.2 Common red algal endosymbiont.....	32
1.3.2.3 Protein import system in red algal-derived plastids.....	33
1.3.2.4 One host or multiple hosts?.....	35
1.3.2.5 Serial endosymbioses.....	36
1.3.2.5.1 Tertiary endosymbiosis in dinoflagellates.....	38
1.3.2.5.2 Kleptoplasty: an early step in endosymbiosis?.....	39
2. OBJECTIVES.....	43
3. MATERIALS AND METHODS.....	46
3.1 Selection of markers.....	47
3.2 Phylogenetic analyses	48
4. ORIGIN AND EVOLUTION OF PRIMARY PLASTIDS.....	50
4.1 Context and objectives.....	51
4.2 Results.....	52
4.3 Manuscript of article 1.....	53
5. EVOLUTION OF SECONDARY-PLASTID HARBORING LINEAGES.....	60
5.1 Context and objectives.....	61
5.2 Results.....	62
5.3 Manuscript of article 2.....	63
6. ORIGIN OF THE SELMA TRANSLOCON IN SECONDARY PLASTIDS OF RED ALGAL ORIGIN.....	77
6.1 Context and objectives.....	78
6.2 Results.....	79
6.3 Manuscript of article 3.....	81
7. DISCUSSION.....	106
7.1 Primary plastids evolved from an unicellular early-branching cyanobacterium.....	107
7.1.1 Primary photosynthetic eukaryotes appeared in a freshwater habitat.....	111
7.2 Evolution of secondary plastid-harboring eukaryotes.....	112
7.2.1 Was the rhizarian ancestor photosynthetic?.....	116
7.2.2 Evolutionary history of the SELMA translocon.....	118
7.3 Perspectives.....	123

8. CONCLUSIONS	126
9. SUPPLEMENTARY MATERIAL	129
9.1 Supplementary material of manuscript 1.....	130
9.2 Supplementary material of manuscript 2.....	142
9.3 Supplementary material of manuscript 3.....	246
RÉSUMÉ EN FRANÇAIS	267
REFERENCES	278

1. INTRODUCTION

"Life did not take over the globe by combat, but by networking"

Lynn Margulis and Dorion Sagan. *Origins of Sex: Three Billion Years of Genetic Recombination* (1986).

1. INTRODUCTION

In this introductory part, I provide the theoretical background and up-to-date knowledge on the origin and evolution of photosynthetic eukaryotes as well as some elements of the current debate that I addressed during my PhD research project. First, I describe the history of the endosymbiotic theory of plastids (section 1.1): how it started, the precursors and the problems it faced before it was widely accepted and became one of the most important contributions to evolutionary thinking of the twentieth century. Then, in section 1.2 and section 1.3, I describe the diversity of primary and secondary photosynthetic eukaryotes, respectively, as well as the evolutionary models aiming at understanding the origin and evolution of these lineages.

1.1 Endosymbiotic theory: a historical perspective

1.1.1 Pre-molecular era

1.1.1.1 Symbiosis: the origin of an idea

The history of symbiosis (from the latin “living together”) is tightly connected to the study of lichens. It is well known that lichens are symbiotic associations between a fungus (usually an ascomycete) and a photosynthesizer (green algae or cyanobacteria), although recent findings have unveiled a third partner, a basidiomycete yeast, that went undetected for over 150 years (Spribille et al., 2016).

The first to propose a composite origin of lichens was the Swiss botanist Simon Schwendener during the annual meeting of the Swiss Natural History Society in 1867 (cited by Honegger, 2000). Schwendener considered lichens to have a dual antagonistic nature, where an enslaved photosynthesizer was exploited by a fungal master (Perru, 2006). However, his ideas did not have a good reception and were firmly rejected by most of his contemporary lichenologists, in part because of the slave-master analogy, which was fiercely ridiculed, but also due to the belief of the “individuality of species” that prevailed during most part of the nineteenth century (Honegger, 2000; Sapp, 1994). At the time when Schwendener exposed his ideas, lichens were thought to be an intermediate group between algae and fungi (Perru, 2006).

Based on the works of Schwendener on lichens, the German biologist Anton de Bary published in 1879 the article “*Die Erscheinung der Symbiose*” where he defined symbiosis as “a phenomenon in which dissimilar organisms live together” (de Bary, 1879; Oulhen et al., 2016), giving to this term the modern biological meaning to describe the

continuum of ecological relationships, from parasitism to mutualism (cited by Oulhen et al., 2016). As a consequence, the role of symbiosis in evolution started to be discussed at the end of the nineteenth century.

Andrei Famintsyn, a Russian plant physiologist, studied the adaptive role of symbiosis using as a model the symbiotic relationship between the photosynthetic algae of the genus *Zoochlorella* and their invertebrate partners (Famintsyn, 1892, 1889). His research led him to suggest that if symbiotic zoochlorellae can grow independently from their animal hosts, it could be also the case for chloroplasts cultured outside the plant cell. (Sapp et al., 2002). However, when he tried to cultivate chloroplasts from the alga *Vaucheria*, his attempts were unsuccessful (Famintsyn, 1907; Provorov, 2016). Famintsyn believed that complex organisms can appear through the symbiotic unification of simpler forms via the transformation of their aggregation into a living entity of higher order (Khakhina, 1992). During a large part of his scientific career, Famintsyn attempted to frame the symbiotic interactions in the context of Darwinian evolution; to him, symbiosis and symbiotic unification of organisms could have impacted the origin of life on Earth and its subsequent diversification (Carrapiço, 2015; Khakhina, 1992).

1.1.1.2 Constantin Mereschowsky's symbiogenetic theory

When Lynn Margulis published her milestone paper "*On the origin of mitosing cells*" (Sagan, 1967), she brought into the scientific debate the controversial claims of the symbiotic origin of several eukaryotic organelles (i.e. mitochondria, chloroplasts and flagella were once free-living bacteria). She acknowledged the Russian botanist Constantin Mereschowsky as the first to suggest in the early twentieth century that chloroplasts had originated from photosynthetic microorganisms (Cyanophyceae). Although disputable, the first proposal of the symbiotic nature of chloroplast was probably made by Andreas Schimper, a German biologist whose paper "*Über die Entwicklung der Chlorophyllkörner und Farbkörper*" (Schimper, 1883) showed that chloroplasts (a term coined by himself) do not appear *de novo* but replicate autonomously within the plant cell. This led him to suggest a symbiotic origin of plastids, though vaguely and without further development than a short footnote: "*If it can be conclusively confirmed that plastids do not arise de novo in egg cells, the relationship between plastids and the organisms within which they are contained would be somewhat reminiscent of a symbiosis. Green plants may in fact owe their origin to the unification of a colourless organism with one uniformly tinged with chlorophyll*" (William Martin & Kowallik, 1999).

Mereschkowsky was au courant of Andreas Schimper's work on chloroplast division, he used his discoveries and his own research on chloroplasts to challenge the autogeneous hypothesis of the origin of the chromatophore (as he called the chloroplast), which was the privileged model at the time. positing that plastids emerged through the "*differentiation of the protoplasmic substance*" (Martin & Kowallik, 1999; Wilson, 1902). Instead, Mereschkowsky proposed that chloroplasts were once free-living Cyanophyceae that became permanent symbionts within nucleated cells giving rise to new organisms (Mereschkowsky, 1905).

Mereschkowsky's theory on the symbiogenetic origin of chloroplasts was heavily inspired by the dual nature of lichens. Lichens were a recurrent example on his publications, showing how simpler organisms living together can associate resulting in a sort of more complex entity (Mereschkowsky, 1905, 1910, 1920). He thought rightly that lichens were polyphyletic, suggesting that symbiotic associations between algae and fungi had happened independently at least ten times (Mereschkowsky, 1920). The reasoning behind his theory was based on a comparative study of the physiological similarities of chloroplasts and cyanobacteria. In his paper "*Über Natur und Ursprung der Chromatophoren im Pflanzenreiche*" (On the nature and origin of chromatophores in the plant kingdom) (Martin & Kowallik, 1999) he enunciated his five arguments to consider chromatophores as symbionts:

1. *The continuity of chromatophores.* Plastids proliferate through division of pre-existing plastids, therefore, it is necessary to postulate that the first chromatophore came from outside to integrate into a non-photosynthetic organism.
2. *Chromatophores are highly independent of the nucleus.* Mereschkowsky remarked the high independence of the plastid in respect to the nucleus control machinery as expected for a symbiont. He also noticed the differences in lipid composition between the chromatophore and the cytoplasm.
3. *The complete analogy of chromatophores and zoochlorellae.* The case of the photosynthetic zoochlorellae is an example of how an independent organism can establish a close symbiotic relationship with another one. Then a similar event could have originated the chromatophore.
4. *There are organisms that we can regard as free-living chromatophores.* Mereschkowsky suggested that the morphological and physiological resemblances of cyanophytes to chromatophores point out that the latter originated from the former.

5. *Cyanophytes actually live as symbionts in cell protoplasm.* The symbiosis of cyanophytes seems to be common in nature.

Mereschkowsky thought that the emergence of new organisms through symbiosis was a common phenomenon in the history of life on Earth. Given the great diversity of algae and plants, he suggested that the plant kingdom was polyphyletic: different “animal cells” would have been invaded by cyanophytes independently in multiple symbiogenetic events (Martin & Kowallik, 1999; Mereschkowsky, 1920). Although only two primary cyanobacterial endosymbioses are recognized so far to have originated photosynthetic organisms (McFadden, 2014), a number much lower than the 15 events initially proposed by Mereschkowsky (Mereschkowsky, 1920). It is worth highlighting the importance of the endosymbiosis in the evolution of secondary photosynthetic eukaryotes (Archibald, 2015).

1.1.1.2.1 The origin of the nucleus

Mereschkowsky is recognized mainly for his contribution on the origin of the chloroplast, however, he also proposed a theory for the origin of the first microorganisms, an event that would have been followed by the symbiogenesis of nucleated cells (Mereschkowsky, 1905, 1910). Nevertheless, the symbiotic origin of the nucleus had already been suggested in 1893 by the Japanese zoologist Shosaburo Watase who speculated that small organisms living together gave rise to the cytoplasm and the nucleus (Sapp, 2009). Likewise, in the early twentieth century, the German biologist Theodor Boveri better known for his research on cancer, enounced briefly a possible symbiogenetic origin of nucleated cells: *“It might be possible that what we call a cell, and for which our mind demands simpler preliminary stages, originated from a symbiosis of two kinds of simple plasmatic structures, such that a number of smaller ones, the chromosomes, settled within a larger one, which we now call the cell body”* (Maderspacher, 2008).

According to Mereschkowsky, life on Earth probably appeared twice, giving rise to organisms with different plasma characteristics (Mereschkowsky, 1910). First, small microorganisms resembling extant micrococci would have emerged during the Earth’s early hot period. Cyanobacteria may have originated from this type of organisms. Then, when the temperature on Earth descended below 50 °C and with a greater availability of organic material, the second appearance of organisms took place, a sort of amoeboid protoplasm. The nucleated cells would be the result of the merging of these two types of cells. The amoeboplasm was invaded by a colony of bacteria that ultimately formed a

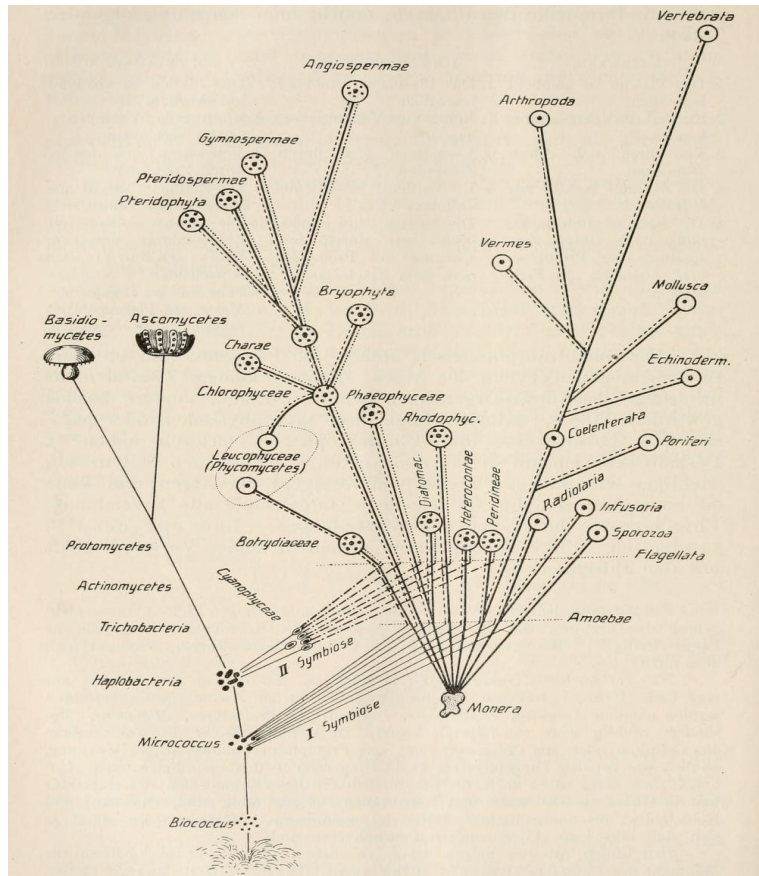


Figure 1. Multiple symbiogenetic origins of nucleated cells and plastids (adapted from Mereschkowsky, 1920).

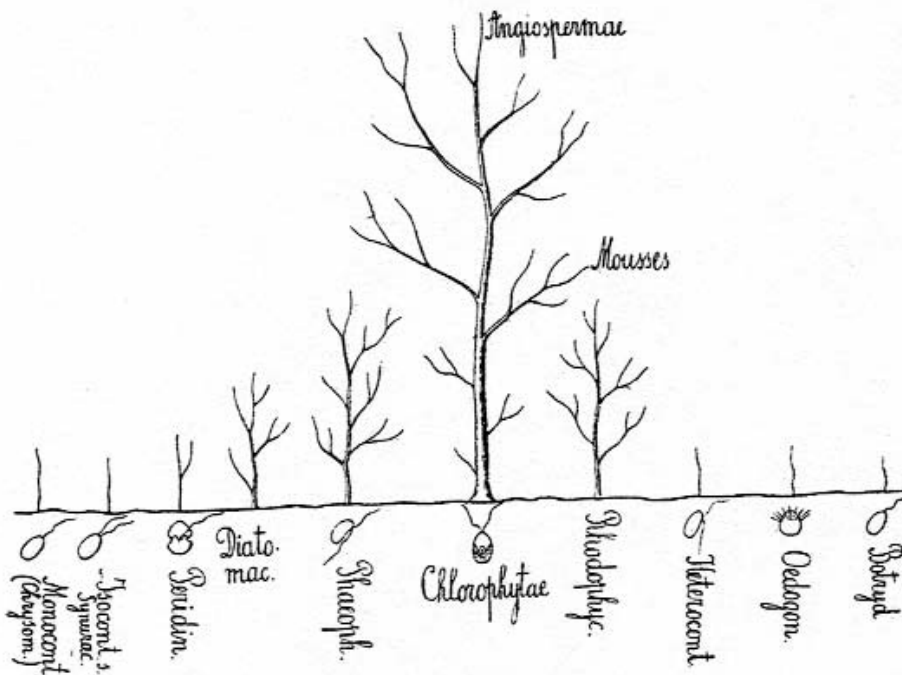


Figure 2. Polyphyletic origin of plants (adapted from Mereschkowsky, 1920)

membrane and gave rise to the nucleus. This process would have occurred multiple times producing independent origins of nucleated organisms (Fig. 1). Likewise, Mereschkowsky thought that plants originated from multiple secondary symbiogenetic events via the acquisition of cyanophycean symbionts (Fig. 1). He imagined the phylogeny of plants not like a tree but like a grove (Fig. 2) (Mereschkowsky, 1920). However, the idea of a polyphyletic origin of the nucleus in eukaryotes and the plastids in plants proved to be wrong (López-García & Moreira, 2015; McFadden & Van Dooren, 2004).

1.1.1.2.2 Mitochondria

Mereschkowsky was opposed to the hypothesis that the chloroplast and mitochondrion shared a same origin, where the former would derive from differentiation of the latter (Guilliermond, 1919; Mereschkowsky, 1920). The ability of chloroplasts to synthesize organic matter from inorganic compounds represented a clear physiological distinction between both organelles that argued for independent origins. Interestingly, although he was uncertain about the origin of mitochondria, he rejected the idea that mitochondria were symbionts, considering it to be incompatible with the symbiogenetic origin of chloroplasts (Mereschkowsky, 1920).

It was Paul Portier who in his book "*Les symbiotes*" published in 1918 analyzed the similarities between bacteria and mitochondria and proposed that mitochondria were in fact "cellular symbionts" derived from free-living bacteria (Perru, 2006; Portier, 1918). He went a step further claiming he was able to culture mitochondria outside the cell, confirming the bacterial origin of mitochondria; however, his results were strongly contested and were considered, rightfully, contaminations (as cited in Perru, 2006).

One decade later, the American Ivan Wallin, aware of the work and criticism of Portier's work, continued the efforts to cultivate mitochondria. However, similar to Paul Portier, his work and proposal of the symbiotic nature of mitochondria were rejected by the scientific community of his time. His book "*Symbiogenesis and the origin of species*" (Wallin, 1927) was one of the latest publications about the endosymbiotic origin of organelles before the theory ceased to be widely discussed for over 40 years, until it was revitalized in the mid-sixties by Lynn Margulis.

1.1.2 Molecular era

1.1.2.1 Lynn Margulis and the renaissance of the symbiogenetic theory

The period between the 30s and the early 50s of the twentieth century produced great advances in the knowledge and description of the cell, in part due to technological innovations such as the electron microscope, invented by Max Knoll and Ernst Ruska in 1931 (Haguenau et al., 2003). The electron microscope allowed to reach levels of resolution physically impossible for the light microscopes, therefore achieving an unprecedented detailed description of the cellular world.

The description of the ultrastructure of algal and plant cells showed the similarities between the chloroplasts and blue-green algae (cyanobacteria) (e.g. presence of sac-like thylakoids) (Gibbs, 1962; Menke, 1960); additionally, ultrastructural observations combined with improved histological techniques revealed the presence of both protein synthesis machinery and packed DNA in chloroplasts and mitochondria (Nass & Nass, 1963; Ris & Plaut, 1962; Schatz et al., 1964), reopening the debate on the origin of the organelles.

Lynn Margulis combined the accumulated cytological and molecular knowledge about mitochondria and chloroplasts with Mereschkowsky's symbiogenetic ideas and published her updated theory of the origin of eukaryotic cells in the now-famous paper "*On the origin of mitosing cells*" (Sagan, 1967); although initially rejected by about fifteen journals (Knoll, 2012). She proposed that mitochondria, chloroplasts and eukaryotic flagella derived from free-living bacterial cells. Similar to the response obtained by earlier proponents of the endosymbiotic origin of organelles, her ideas were received with much hesitation by her contemporaries. The evolutionary model proposed by Margulis offered a sequence of serial endosymbiotic events to explain the origin of organelles within the eukaryotic cell, in clear contrast with the explanations provided by the autogenous models that were popular back then (Raff & Mahler, 1972; Uzzell & Spolsky, 1974). The symbiogenetic model was also criticized due to the prevailing vision of graduality in evolution, where endosymbiosis was thought to be an abrupt event incompatible with the continuity of slow evolutionary change (O'Malley, 2015).

According to Lynn Margulis, the origin of mitochondria was connected to a change in environmental conditions: the increase in oxygen concentration observed during the Mesoproterozoic (Holland, 2006) favored the integration of an aerobic prokaryote (future mitochondrion) into the cytoplasm of a heterotrophic anaerobe to cope with the new oxygen-containing atmosphere giving rise to an obligate symbiosis (Sagan, 1967).

Margulis thought that the selective force behind the symbiogenesis of mitochondria was oxygen utilization, therefore, she argued against a common ancestry between mitochondria and mitochondria-like anaerobic organelles (e.g. hydrogenosomes and mitosomes) which according to her, must have independent origins (Margulis et al., 2006). Within the framework of serial endosymbioses proposed by Margulis, after the acquisition of mitochondria, a second symbiogenetic event with a spirochete-like symbiont originated the flagellum (Bermudes et al., 1987; Margulis & Bermudes, 1985; Sagan, 1967). Interestingly, an alternative symbiotic model posits that the eukaryotic flagellum derived from an epixenosome-like bacterium from the phylum Verrucromicrobia that started as defensive ectosymbiont of the eukaryotic cell (Li & Wu, 2005). However, there is no clear evidence for the prokaryotic past of the eukaryotic flagella (e.g. no associated genome) and the autogenous origin is favored (Moran et al., 2014).

When Linus Pauling and Emile Zuckerkandl published their paper "*Molecules as Documents of Evolutionary History*" where they proposed that nucleic acids and proteins can be used to infer the evolutionary history of organisms (Zuckerkandl & Pauling, 1965), a new revolution in evolutionary studies started. However, they were not the first to propose this idea. In 1892, Ernst Haeckel hypothesized that to know the phylogenetic relationships among the Kingdom Monera, it was necessary to study the atomic composition of their albumen (proteins), clearly suggesting the informative potential of molecules (Sapp, 2009). Nonetheless, it was Pauling and Zuckerkandl who created a theoretical framework for molecular phylogenetics.

Although the symbiogenetic model of organelle evolution encountered strong opposition, early gene sequencing and the development of molecular phylogeny offered the opportunity to test the model at the molecular level. In 1974, Robert M. Schwartz and Margaret O. Dayhoff offered the first phylogenetic evidence of the prokaryotic ancestry of mitochondria and chloroplasts (based on ferredoxin, c-type cytochromes and 5S ribosomal RNA sequences) (Schwartz & Dayhoff, 1978) (Fig. 3). They demonstrated for the first time at the molecular level the prokaryotic ancestry of these organelles. As predicted by Margulis, the chloroplast emerged from a symbiotic event with blue-green algae. Further molecular evidence of the endosymbiotic origin of these organelles will accumulate throughout the 1980s (Gray & Doolittle, 1982).

The description of the third membrane of the chloroplast of unicellular algae of the genus *Euglena* made by electronic microscope observations by Gibbs (Gibbs, 1978) led her to propose that this chloroplast evolved from an endosymbiotic event with a green

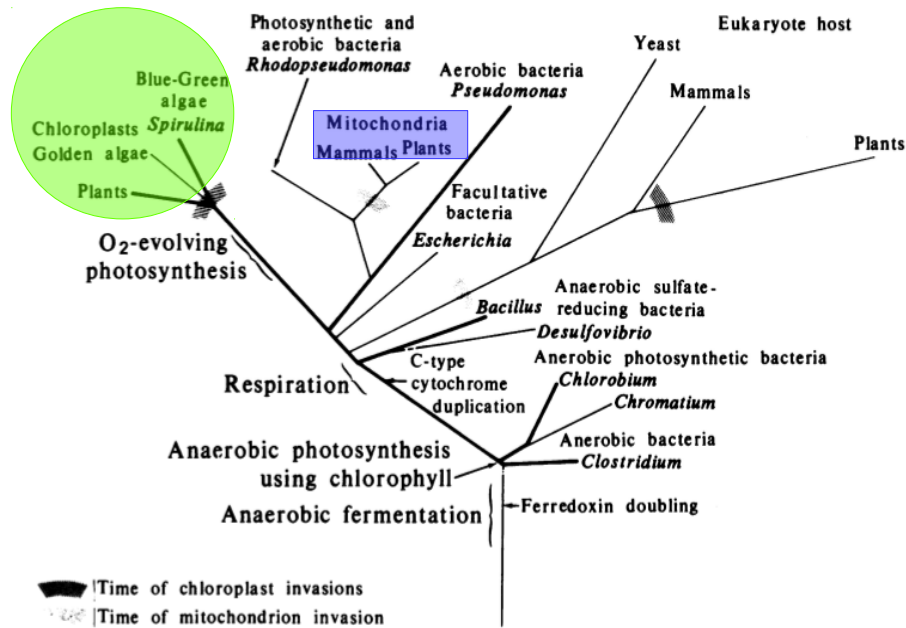


Figure 3. Composite evolutionary tree based on ferredoxine, c-type cytochromes and 5S ribosomal RNA sequences, depicting the prokaryotic ancestry of mitochondria and chloroplasts (adapted from Schwartz and Dayhoff, 1978)

alga. This was a significant change in the endosymbiotic theory and opened the possibility that some algae could have obtained their plastids through eukaryote-into-eukaryote endosymbiosis. This added a layer of complexity to the evolution of photosynthetic eukaryotes where simpler parsimonious models may not reflect the degree of reticulation and recurrence of endosymbiotic events. For instance, the parsimonious Cabozoa hypothesis proposed by Cavalier-Smith (1999) posited a common secondary origin of the green plastids in euglenids and chlorarachniophytes. However, phylogenetic analyses helped to elucidate the origin of the chloroplasts in both green lineages and it is now clear that they were acquired independently (Rogers et al., 2007). This shows how the endosymbiotic theory continues to change, assimilating new information, new methods and new techniques but always working with the experience and knowledge that have been accumulated (Fig. 4).

1.2 Primary photosynthetic eukaryotes

1.2.1 Origin of the chloroplast: from cyanobacterium to organelle

Oxygenic photosynthesis appeared in Cyanobacteria possibly after their divergence from their closest non-photosynthetic relatives, the recently described Melainabacteria, approximately 2.6 bya (Shih et al., 2017; Soo et al., 2017). This was a major event in the

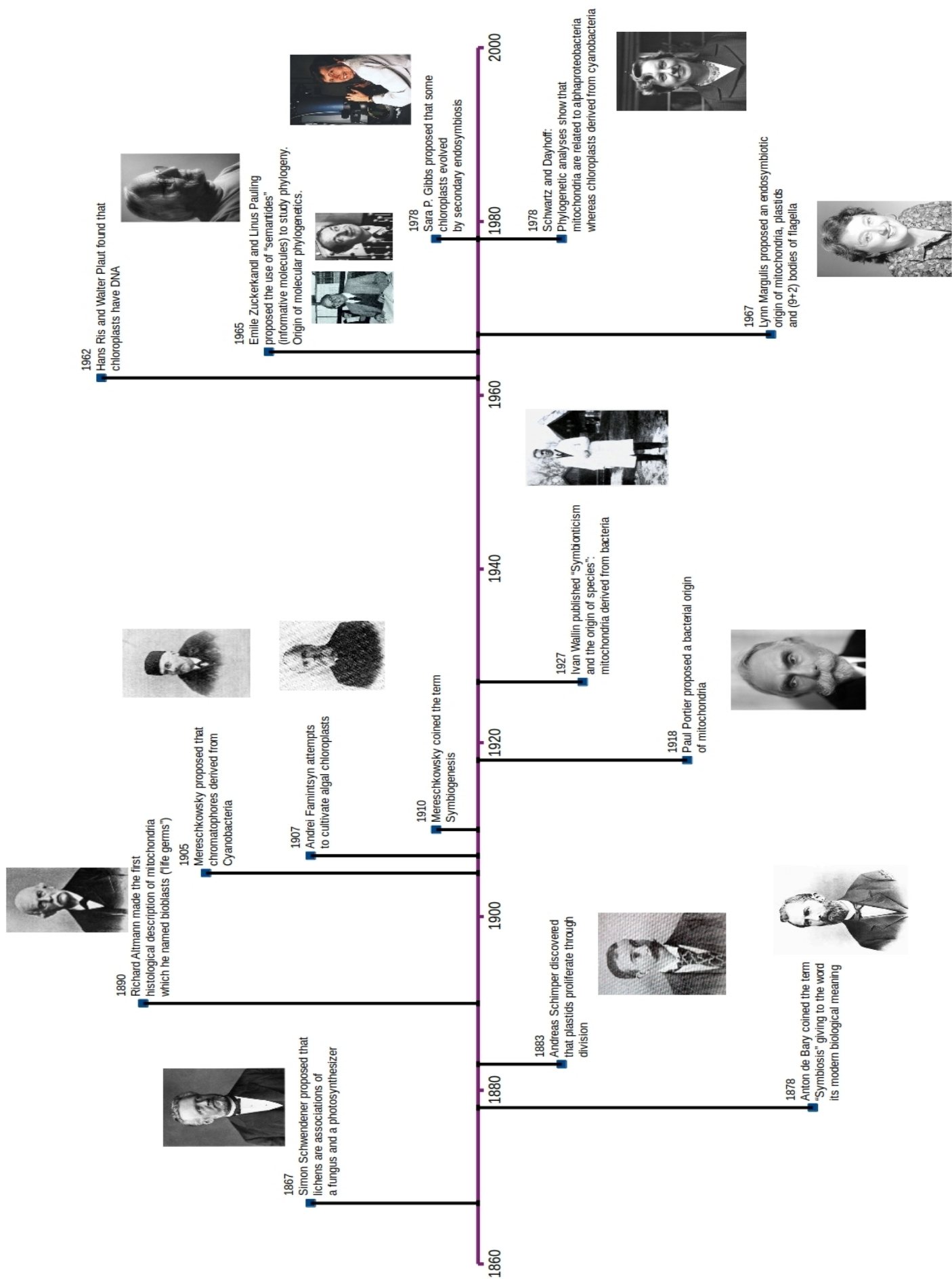


Figure 4. Timeline of landmark events and publications in the history of the endosymbiotic theory

history of Earth that started the rise of oxygen and modified drastically the biogeochemical cycles and whole ecosystems. Then, around 1.6 bya a cyanobacterium spread the ability to photosynthesize through the establishment of an endosymbiotic relationship with a protist giving rise to Archaeplastida, a monophyletic supergroup of photosynthetic eukaryotes that comprises glaucophytes, red algae, green algae and land plants (Rodríguez-Ezpeleta et al., 2005; Yoon et al., 2004).

1.2.1.1 Endosymbiotic Gene Transfer

Chloroplast genomes commonly encode 60-150 proteins whereas free-living cyanobacteria harbor between 2,500-7,000 proteins, with the genomes of N₂-fixing cyanobacteria from the order Stigonematales reaching more than 12,000 encoded proteins (Dagan et al., 2013; Shi & Falkowski, 2008). This dramatic reduction in the coding capacity of chloroplast genomes compared with cyanobacteria is associated to massive gene loss but also to the relocation of cyanobacterial genes into the nuclear genome of the host, a process termed Endosymbiotic Gene Transfer (EGT) (Timmis et al., 2004). Mitochondria, the other bacterial-derived organelles in eukaryotic cells, carry also extremely reduced genomes. It seems to be an adaptation to life within the host cell, also observed in intracellular parasites (Khachane et al., 2007; Sakharkar et al., 2004). Estimations of the contribution of cyanobacterial EGTs to the nuclear genome of primary plastid-bearing eukaryotes can considerably differ among lineages but usually represent between 5 to 18 % of the protein-coding genes of the nucleus, with land plants having the largest set of these genes (Dagan et al., 2013; Makai et al., 2015; Martin et al., 2002; Reyes-Prieto et al., 2006).

EGTs are considered to have played a major role in the establishment of plastids. Initially, the host who fed on cyanobacteria would have integrated random pieces of DNA after digestion (Fig. 5). This continuous supply of cyanobacterial genes increased the host gene repertoire providing a new set of genes that could be functionally active after the addition of promoter sequences (Stegemann & Bock, 2006). After EGTs are efficiently transcribed and imported into the plastid, their homologs in the plastid genome can undergo pseudogenization followed by complete gene loss (Fig. 5).

The genes transferred to the host nucleus created a dependence of the symbiont that together with host-derived proteins targeted to the plastid increased the control of the host over the symbiont autonomy, for instance, in plastid division (Miyagishima, 2011). To

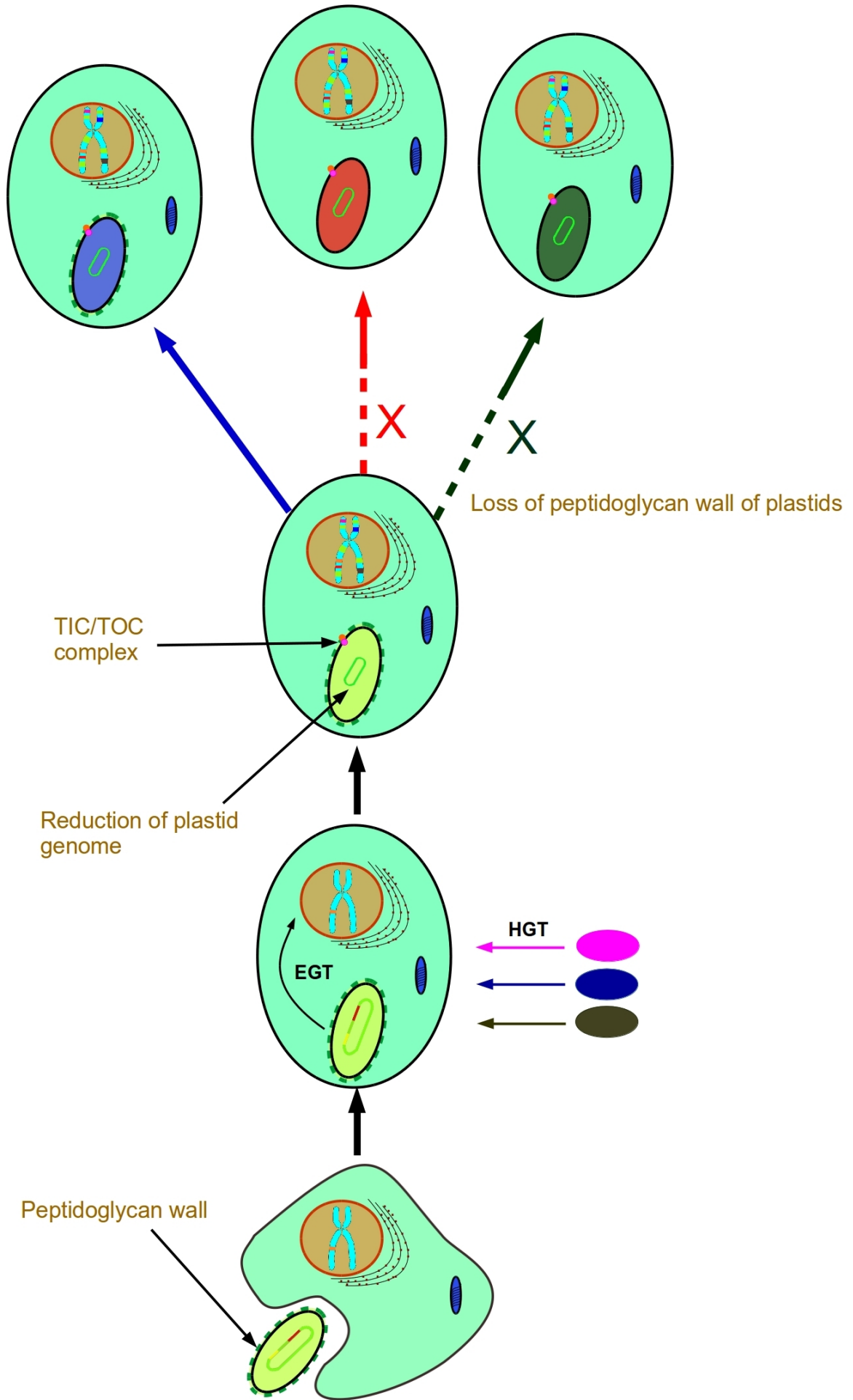


Figure 5. Acquisition of genes by the host via EGT and HGT during the plastid establishment.

import proteins into the plastid, it was necessary to set up an import system which derived from recycled cyanobacterial transporters and host genes (Shi & Theg, 2013) (see below). Although most of the proteins of cyanobacterial origin are targeted back to the plastid (Reyes-Prieto et al., 2006), some EGTs have been neofunctionalized to play new roles in the cytosol expanding the metabolic capacities of the host (Makai et al., 2015).

It is not completely clear why some genes are transferred while others continue to be encoded in the plastid (or mitochondrial) genome. Hydrophobicity of proteins has been proposed as a possible explanation to whether a protein is likely to be transferred or not, according to this, hydrophobic proteins tend to be encoded in the organelles (von Heijne, 1986). Other hypothesis suggests that organelle-encoded proteins continue to be encoded *in situ* because they require a tight and continuous redox regulation to respond to changes on the physical environment (Allen, 2015).

1.2.1.2 Bacterial genes in primary photosynthetic eukaryotes

Although primary plastids derived from a cyanobacterial endosymbiont, it was not the only prokaryotic contributor to the plastid proteome. For instance, Suzuki & Miyagishima (2010), showed that more of 20% of plastid-targeted proteins in the thermoacidophilic red alga *Cyanidioschyzon merolae* derived from bacteria different from cyanobacteria likely acquired via horizontal gene transfer (HGT). Likewise, 7% of proteins in the cyanelle of *Cyanophora paradoxa* had bacterial origins (Price et al., 2012; Qiu et al., 2013). Interestingly, it has been proposed that horizontally transferred genes to freshwater algae may have facilitated the colonization of land (Yue et al., 2012).

Differences in proportion of proteins of bacterial origin are probably due to the combination of differential loss of ancient HGTs acquired by the host before or during the plastid establishment (Fig 5.) coupled with lineage-specific transfers after diversification of the Archaeplastida lineages. For instance, the plastid-encoded ribulose-1,5-bisphosphate carboxylase/oxygenase of red algae was transferred to the rhodophyte ancestor from a proteobacterial source (Delwiche & Palmer, 1996).

Several bacterial genes were likely transferred to the ancestor of Archaeplastida before the loss of phagotrophy (Doolittle, 1998) . These genes could have been retained to counter the massive gene loss (more than 95%) occurred in the endosymbiont genome

during primary endosymbiosis. It is possible, however, that some HGT genes were present in the genome of the cyanobacterial endosymbiont and subsequently passed as EGT into the nuclear genome of the host (Fig. 5). Interestingly, phylogenetic analyses of cyanobacterial genomes have shown high frequency of horizontal gene transfer resulting in chimeric genomes (Gross et al., 2008; Zhaxybayeva et al., 2006). Although the HGT donors belong to multiple bacterial phyla, Proteobacteria and Chlamydiae appeared to be the main contributors after cyanobacteria (Qiu et al., 2013). Some authors have explained the apparent overrepresentation of chlamydial-like proteins in Archaeplastida with a tripartite symbiotic model of plastid acquisition where a chlamydial symbiont helped in the early steps of the plastid establishment by protecting the cyanobiont from host defenses in addition to supply the transporters needed to export the photosynthate as well as enzymes to the host to metabolize it (Brinkman et al., 2002; Facchinelli et al., 2013) (see below).

1.2.1.3 Host-derived genes in the chloroplast proteome

Most of the cyanobacterial proteins that are encoded in the nucleus of primary photosynthetic eukaryotes are targeted to the plastid (Reyes-Prieto et al., 2006). However, they only represent about 60% of the plastid proteome; the other 40 % derive from pre-existing host genes or were acquired from non-cyanobacterial prokaryotes, probably via HGT (Suzuki & Miyagishima, 2010).

It has been suggested that host genes could have helped to the host-symbiont integration by providing a set of transporters that facilitated the exchange of metabolites and import of proteins (Karkar et al., 2015; Shi & Theg, 2013; Suzuki & Miyagishima, 2010). Interestingly, host-derived proteins account for more than half of metabolite transporters in the chloroplast membranes (Tyra et al., 2007). However, the host contribution to the plastid proteome is not only restricted to transporter-related proteins, it also participates in carbohydrate metabolism, RNA processing, heme biosynthesis and protein translation in the chloroplast. Thus, these proteins together with bacterial ones have created mosaic plastid proteomes (Karkar et al., 2015; Oborník & Green, 2005).

Endosymbiotic gene replacement with host-derived genes, either by duplication of host genes or dual targeting of nucleus-encoded proteins (e.g. aminoacyl t-RNA synthetases) (Duchene et al., 2005) increased the dependency of the symbiont hindering the shift to a free-living lifestyle. The mosaicism of plastid functions together with the

contribution to the endosymbiont division machinery (Pyke, 2010) may have facilitated that the enslaved cyanobiont became a permanent organelle.

1.2.1.4 TIC/TOC complex

The vast majority of plastid-located proteins in Archaeplastida lineages are encoded in the nuclear genome and need to be targeted and imported into the plastid (Suzuki & Miyagishima, 2010; Zybailov et al., 2008). To assure that requirement, primary plastids evolved a translocation machinery at the outer and inner membranes, the TOC and TIC complexes, respectively. These two multiprotein complexes have a hybrid nature: some transporters and receptors derive from the endosymbiont (e.g. Toc75, Tic22 and Tic55) while other proteins have eukaryotic origin (e.g. Tic110, Toc159 and Toc34) (Shi & Theg, 2013). Detection of core proteins of the TIC-TOC complex in all algae and plants suggests that the import system was already established in last common ancestor of Archaeplastida (McFadden & Van Dooren, 2004).

To target nucleus-encoded proteins to the plastid, these proteins carry an N-terminal transit peptide (TP) enriched in hydroxylated serine residues (Bhushan et al., 2006) that is recognized at the outer chloroplast membrane by the TOC machinery. Then, the protein is translocated across the inner membrane via the TIC complex in an ATP-dependent manner. Once inside the stroma, the TP is cleaved off by a stromal processing peptidase (Kovács-Bogdán et al., 2010; Paila et al., 2015).

1.2.1.5 The *ménage-à-trois* hypothesis

Primary photosynthetic eukaryotes have chimeric nuclear genomes with genes derived from diverse sources (Deusch et al., 2008; Martin et al., 2002). Most of the bacterial genes come from alpha-proteobacteria and cyanobacteria, the clades that gave rise to mitochondria and primary plastids, respectively (Deusch et al., 2008; Ku et al., 2015). Surprisingly, Chlamydiae seem to have transferred between 20-60 genes to the nuclear genome of the ancestor of Archaeplastida (Ball et al., 2013; Becker et al., 2008; Brinkman et al., 2002; Huang & Gogarten, 2007). These genes participate in a broad spectrum of functions such as fatty acid biosynthesis, membrane transporters, carbohydrate metabolism, rRNA processing and aminoacid biosynthesis, but it was the

enzymes involved in carbohydrate (glycogen) metabolism which led Ball et al. (2013) to propose a major role of Chlamydiae during the early establishment of the cyanobiont. This model is called the *ménage-à-trois* hypothesis and suggests that the cyanobiont might have been phagocytized by a host infected with a chlamydia-like parasite. Then, the latter would have supplied the enzymes to transform the photosynthate produced by the cyanobiont (ADP-glucose) in a form available for the host. In this way, the chlamydial symbiont would have functioned as a mediator between the host and the cyanobiont facilitating its integration. It is noteworthy that Chlamydiales are common parasites and can infect a wide range of eukaryotic hosts, from animals to protists, but they are not known to parasitize Archaeplastida (Horn, 2008).

According to the initial proposal of the *ménage-à-trois* hypothesis (Ball et al., 2013) (Fig. 6A), the cyanobiont recruited early during endosymbiosis a nucleotide-sugar transporter (NST) from the host endomembrane system to the inner membrane of the future plastid. This transporter would have helped to translocate the ADP-glucose produced by the cyanobiont into the cytoplasm of the host. However, eukaryotes use UDP-glucose as precursor for glycogen synthesis and would not have been able to use the exported ADP-glucose. Thus, the chlamydial glycogen synthase GlgA would have catalyzed the storage of ADP-glucose in the glycogen reservoir of the host, working as a bridge between carbohydrate metabolism of the cyanobiont and the host. Then, additional effectors such as the glycogen phosphorylase GlgP and the glycogen debranching enzyme GlgX would catabolize the host glycogen releasing glucose-1-P and maltotetraose, respectively; while maltotetraose would be imported into the chlamydial symbiont, glucose-1-P would be metabolized by the host.

However, although the plastid-targeted NST transporter is monophyletic in green and red algae (Weber et al., 2006), none of the NST proteins in *Cyanophora paradoxa* seems to be located at the plastid membrane, suggesting that glaucophytes do not use the same transporter system to export the photosynthate than the other two lineages (Price et al., 2012); therefore, it seems that the ADP-glucose export system was not established in the common ancestor of Archaeplastida but possibly after the divergence of glaucophytes.

Cyanophora paradoxa exports the photosynthate in the form of glucose-6-phosphate through an UhpC-like transporter of apparent chlamydial origin (Price et al., 2012). The putative ancestral state of the cyanelle together with the discovery of UhpC homologs in the nuclear genomes of red and green algae (Price et al., 2012) updated the

ménage-à-trois model (Facchinelli et al., 2013) (Fig. 6B) suggesting that UhpC was the transporter that was present in the ancestor of Archaeplastida whereas the NST transporter was secondly acquired in the common ancestor of Rhodophyta and Viridiplantae.

In this updated *ménage-à-trois* model, the cyanobiont inhabited the same phagosome vesicle that the Chlamydia-like bacteria, in a composite organelle called “chlamydioplast” (Facchinelli et al., 2013). Here, the cyanobiont would have exported glucose-6-phosphate via the chlamydial-derived UhpC transporter, to be converted to ADP-glucose and exported into the host cytoplasm using the host-derived NST transporter (Fig. 6B). Once outside the chlamydioplast, the ADP-glucose would have participated in the glycogen biosynthesis of the host with the help of chlamydial effectors in a similar manner as proposed in the initial model (Ball et al., 2013). During this period of cyanobiont-Chlamydiae cohabitation, it is hypothesized that several genes would have been horizontally transferred from the chlamydial partner to the cyanobiont and the eukaryotic host (e.g. carbohydrate metabolism-related genes), thereby breaking the tripartite interdependency which ended with the complete loss of the chlamydial symbiont. Then, while the early-branching Glaucophyta kept the peptidoglycan wall from the cyanobiont, it was lost in the common ancestor of Rhodophyta and Viridiplantae, opening the possibility to evolve or relocate other membrane transporters such as the NST transporter, considered to be initially in the chlamydioplast membrane (Facchinelli et al., 2013).

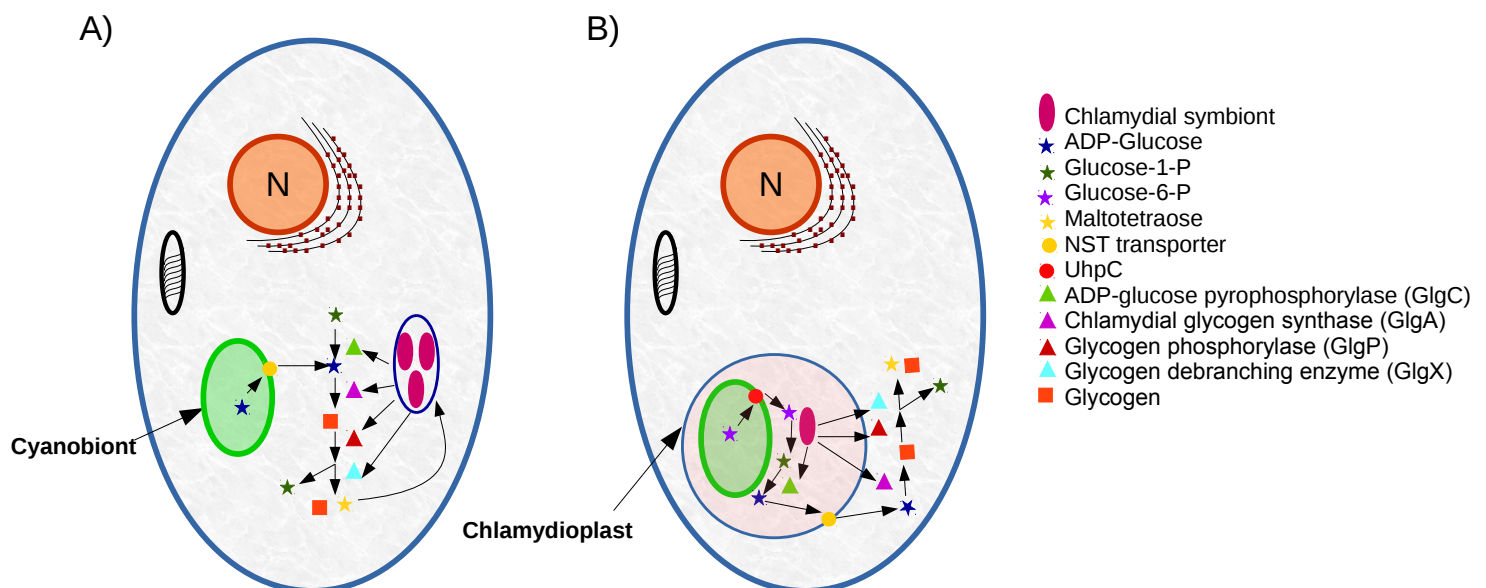


Figure 6. *Ménage-à-trois* hypothesis of primary plastid endosymbiosis. A) Initial proposal B) Updated model. N=Nucleus.

1.2.1.5.1 Was there a chlamydioplast?

The proposal of the *ménage-à-trois* hypothesis created an interesting metabolism-based model to explain the genes of apparent chlamydial origin detected in the nuclear genomes of Archaeplastida (Ball et al., 2013; Becker et al., 2008; Brinkman et al., 2002; Huang & Gogarten, 2007). However, it seems that the role of Chlamydiae in shaping the nuclear genome of primary photosynthetic algae could have been overestimated (Dagan et al., 2013; Deschamps, 2014; Domman et al., 2015). For instance, bacterial phyla different to Chlamydiae (i.e. Gammaproteobacteria, Actinobacteria, Deltaproteobacteria, Bacilli, Bacteroidetes, Betaproteobacteria) outnumber the quantity of genes transferred to nuclear genomes of Archaeplastida (Dagan et al., 2013).

Recent reanalysis of putative chlamydial genes have casted doubt on the contribution of this bacterial group, weakening the major role proposed for a chlamydial symbiont on the onset of primary plastids. The contribution of chlamydia seems to have been greatly overestimated due to a combination of factors such as the reconstruction of phylogenetic trees with insufficient taxon sampling and the use of poor-fitted models of protein evolution (Deschamps, 2014; Domman et al., 2015). For instance, according to the *ménage-à-trois* hypothesis most of chlamydial-derived genes would have been transferred during the existence of the chlamydioplast, before the diversification of Archaeplastida; however, Deschamps (2014) showed that less than 30 % of genes strictly support a HGT from Chlamydiae to the ancestor of Archaeplastida. Therefore, some chlamydial-derived genes in Archaeplastida could be lineage-specific acquisitions via punctual HGTs.

Moreover, the genes involved in the metabolism of carbohydrates, which are thought to be the origin of the metabolic association with a chlamydial symbiont, seem to have a mix origin with genes from the cyanobiont, the eukaryotic host and HGT from bacteria but not necessarily Chlamydiae (Domman et al., 2015). Taken together, it seems that there is no compelling evidence to support that the establishment of primary plastids was mediated by a chlamydial symbiont.

1.2.1.6 Early- or late-branching cyanobacterium?

The monophyly of Archaeplastida is supported by the phylogenomic analyses of nuclear, plastid and mitochondrial genes (Jackson & Reyes-Prieto, 2014; Rodríguez-Ezpeleta et al., 2005). However, the precise identity of the host and the cyanobacterial

endosymbiont remains uncertain. Due to the difficulties to resolve deep phylogenetic relationships within the eukaryotic tree, the description of the host is limited. Regarding the cyanobacterial endosymbiont, early sequencing of plastid and cyanobacterial genes produced the first studies on the origin of the plastid (Nelissen et al., 1995; Turner et al., 1999). Then, the increase of genomic data improved and expanded the phylogeny of cyanobacteria opening the possibility to identify the closest cyanobacterial lineage of the plastid ancestor. Publications that addressed this question can be divided in two types depending on the phylogenetic position of the plastid ancestor, that is, whether the endosymbiont was an early- or a late-branching cyanobacterium (Table 1). Early-branching models suggest that the endosymbiont that entered into a symbiotic association with the eukaryotic ancestor of Archaeplastida belongs to a cyanobacterial clade near the base of the cyanobacterial tree whereas late-branching models support that it emerged during the divergence of cyanobacteria.

It has been suggested that the N_2 -fixation capability common in some late-branching cyanobacterial clades could have been the initial motor of endosymbiosis, providing to the host an element that was scarce in the oceans during the Proterozoic (Anbar, 2002). Thus, diverse late-branching N_2 -fixing cyanobacterial lineages (unicellular or filamentous) have been proposed as putative sister groups of the plastid ancestor (Table 1). Interestingly, nitrogen-dependent symbioses are common between cyanobacteria and different groups of eukaryotes such as plants, fungi and protists (Kneip et al., 2007; Lesser, 2004).

Reference	Year	Markers	Dataset	Approach	Model	Plastid origin
Nelissen et al.	1995	1	16S rRNA	Phylogeny	Early-branching	After divergence of <i>Gloeobacter</i> and <i>Pseudoanabaena</i>
Turner et al.	1999	1	16S rRNA	Phylogeny	Early-branching	After divergence of <i>Gloeobacter</i> and <i>Pseudoanabaena</i>
Rodríguez-Ezpeleta et al.	2005	50	concatenation of plastid proteins	Phylogeny	Early-branching	After divergence of <i>Gloeobacter</i> and <i>Thermosynechococcus elongatus</i> (No <i>Pseudoanabaena</i>)
Schirmeister et al.	2011	1	16S rRNA	Phylogeny	Early-branching	After divergence of <i>Gloeobacter</i> and before <i>Pseudoanabaena</i>
Criscuolo and Gribaldo	2011	191	plastid and nuclear genes	Phylogeny	Early-branching	After divergence of <i>Gloeobacter</i> and <i>Synechococcus</i> (No <i>Pseudoanabaena</i>)
Shih et al.	2013	25	Concatenation of plastid proteins	Phylogeny	Early-branching	After divergence of <i>Gloeobacter</i> and <i>Pseudoanabaena</i>
Li et al.	2014	75	plastid and nuclear genes	Phylogeny	Early-branching	After divergence of <i>Gloeobacter</i> and <i>Synechococcus</i> (No <i>Pseudoanabaena</i>)
Morden and Golden	1989	1	psbA gene (photosystem II protein D1)	Phylogeny	Late-branching	Related to the filamentous cyanobacteria <i>Prochlorothrix hollandica</i>
Marin et al.	2005	4	SSU rRNA TRNA-Ile TRNA-Ala LSU rRNA	Phylogeny	Late-branching	After divergence of <i>Synechococcus/Prochlorococcus</i> clade
Deusch et al.	2008	11569	Nuclear-encoded proteins	Presence/absence Sequence similarity	Late-branching	Related to filamentous, heterocystous-forming Cyanobacteria (section IV)
Deschamps et al.	2008	NA	NA	comparative	Late-branching	similar storage of starch in plastids and unicellular Diazotrophic cyanobacterium (section I)
Falcón et al.	2010	2	16S rDNA RbcL	Phylogeny	Late-branching	Related to unicellular N_2 -fixing cyanobacteria (section I)
Dagan et al.	2012	611	Nuclear-encoded proteins	Presence/absence Sequence similarity	Late-branching	Related to N_2 -fixing cyanobacteria (section IV and V)
deAlda et al.	2014		concatenated 16S+23S rRNA Concatenated 33 plastid proteins	Phylogeny	Late-branching	Related to N_2 fixing filamentous cyanobacteria (section III, IV, V)

Table 1. Characteristics of studies that have proposed either an early or late plastid branching within the phylogeny of Cyanobacteria. NA: not applicable.

1.2.2 Archaeplastida lineages

1.2.2.1 Viridiplantae

Viridiplantae is a monophyletic group composed of two lineages: Chlorophyta and Streptophyta. Chlorophytes include a broad diversity of freshwater, terrestrial and marine green algae; it includes the early-diverging prasinophytes and the core chlorophytes (Leliaert et al., 2012). Streptophyta comprises the charophytes, a group of freshwater green algae, and the embryophytes or land plants (Fig. 7). The divergence between Streptophyta and Chlorophyta is estimated to have occurred more than 1 billion years ago, followed by the adaptive radiation of both lineages (Leliaert et al., 2011; Moczyłowska, 2011). Chlorophytes diversified in the oceans whereas streptophytes evolved in freshwater habitats (Becker & Marin, 2009; Fang et al., 2017)

The prasinophytes is a paraphyletic group within Chlorophyta that inhabits mostly marine environments with some freshwater clades (Leliaert et al., 2012). Phylogenetic analyses of nuclear and plastid multigene data have proved difficult to resolve the phylogenetic relationships among the different clades of prasinophytes because of the rapid early radiation of the group and possibly due to multiple extinctions of ancient lineages (Cocquyt et al., 2010; Fang et al., 2017).

The core chlorophytes includes ecologically and morphologically diverse green algae. This assemblage contains three species-rich classes, Ulvophyceae, Trebouxiophyceae and Chlorophyceae (UTC), and two smaller lineages, Pedinophyceae and Chlorodendrophyceae (Leliaert et al., 2012). The class Ulvophyceae includes mostly multicellular coastal seaweeds (e.g. *Ulva*) but also unicellular organisms like the macroscopic single-celled *Acetabularia*. The Trebouxiophyceae is a group of green algae with a broad range of morphological character (unicellular/multicellular) and ecological environments (marine, freshwater and terrestrial). It is well-known that several species establish symbiotic relationships with diverse group of eukaryotes. For instance, some trebouxiophycean algae participate in symbiotic relationships with fungi to form lichens (Friedl & Bhattacharya, 2002). Other species are photosynthetic symbionts of ciliates (Summerer et al., 2008) and marine invertebrates (Letsch et al., 2009).

Streptophyta is divided in two groups: the charophytes (a paraphyletic group of freshwater green algae) plus the land plants (Leliaert et al., 2012). Recent phylogenomic analyses of chloroplast genomes point to the charophyte class Zygnematophyceae as the closest living relatives of embryophytes (land plants) (Zhong et al., 2015). The colonization

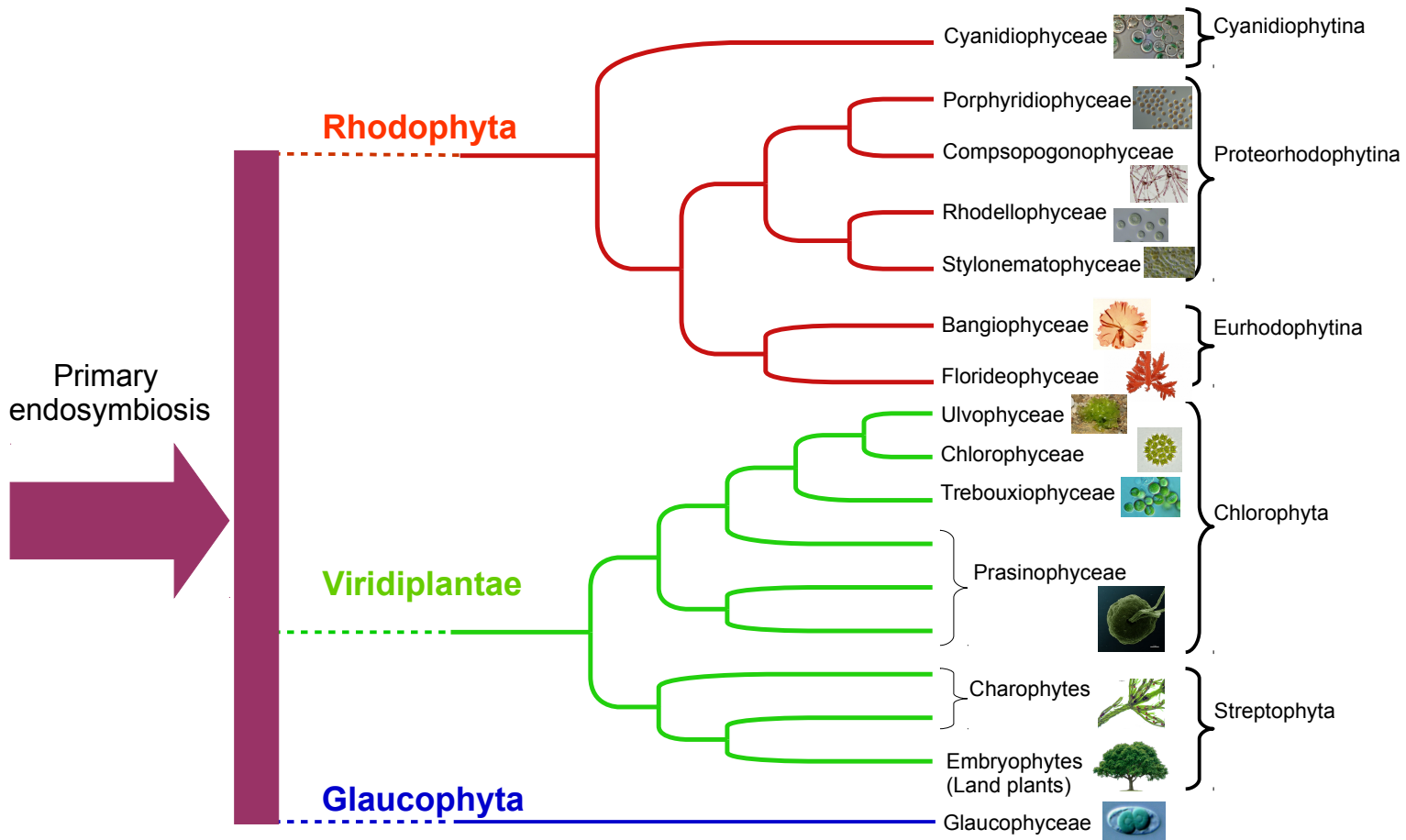


Figure 7. Primary plastid-bearing lineages.

of terrestrial habitats was a gradual change from a freshwater lifestyle passing to moist habitats and eventually to dry land (Becker & Marin, 2009). The evolution of land plants had important ecological consequences: it impacted positively global primary production, increased the atmospheric levels of oxygen, changed the biogeochemical cycles and reshaped the terrestrial ecosystems (Bateman et al., 1998; Lenton, 2001; Lenton et al., 2016).

Chloroplasts of green algae and land plants contain genomes usually in the range of 120 -200 kbp in size, with some exceptions like the non-photosynthetic plastid in the Trebouxioophyceae parasite *Helicosporidium* sp. (37.4 kbp) and in the non-photosynthetic orchid *Epipogium roseum* (19 kbp) (Schelkunov et al., 2017). At the upper limit, the largest green algal chloroplast genome sequence so far is from *Volvox carteri* (525 kbp) (Smith & Lee, 2010) However, the macroscopic *Acetabularia* is thought to have a chloroplast genome up to 2000 kbp, although it has never been completely sequenced (Palmer, 1985).

1.2.2.2 Rhodophyta

Rhodophyta (red algae) is a major lineage of primary plastid-bearing eukaryotes with ~7,100 identified species (Guiry & Guiry, 2017). Red algae have overwhelming representation in marine, coastal and brackish habitats with more than 95% of species living in these environments. The phylum Rhodophyta is divided in three subphyla: the early-branching subphylum Cyanidiophytina which is composed of freshwater unicellular algae living in high-temperature environments (e.g. *Galdieria sulphuraria*); the recently proposed subphylum Proteorhodophytina that comprises mesophilic non-seaweed algae (e.g. *Bulboplastis apyrenoidosa*); and the subphylum Eurhodophytina with mesophilic seaweeds (e.g. *Pyropia perforata*) (Muñoz-Gómez et al., 2017) (Fig. 7).

Within Rhodophyta, cyanidiophycean algae diverged first, prior to the secondary endosymbiotic event that gave rise to secondary red plastids (Kim et al., 2017; Ševčíková et al., 2015). These organisms thrive in high-temperature and acid environments near volcanic areas such as hot springs (Toplin et al., 2008). Interestingly, it is thought that the common ancestor of cyanidiophytes underwent an important genome reduction probably due to the adaptation to the extremophilic lifestyle (Qiu et al., 2015).

The subphylum Eurhodophytina is a monophyletic assemblage composed of two classes: Bangiophyceae and Florideophyceae. These two lineages harbor the vast majority of rhodophyte diversity. Approximately 99 % of the red algal species that have hitherto been described are placed within these classes (mainly in Florideophyceae) (Guiry & Guiry, 2017). Interestingly, the oldest eukaryotic fossil (widely accepted as such) is a bangiophycean species, *Bangiomorpha pubescens*, estimated to be 1.2 billion years old (Butterfield, 2000; Yang et al., 2016).

The subphylum Proteorhodophytina comprises unicellular and filamentous mesophilic algae divided in four classes: Porphyridiophyceae, Compsopogonophyceae, Rhodellophyceae and Stylonematophyceae. Species of this subphylum have shown an impressive variability in the genomic architecture and size of plastid genomes (Muñoz-Gómez et al., 2017). For instance, *Corynoplastis japonica* (Rhodellophyceae) harbors the largest known plastid genome (1127 kbp) completely sequenced. This unusual size is due to a high proportion of intronic sequences accounting for more than 60% of the total plastid DNA (Muñoz-Gómez et al., 2017).

1.2.2.3 Glaucophyta

Glaucophytes are a relatively rare group of unicellular freshwater algae with primary plastids called cyanelles. This group of algae is species-poor with only 22 described species grouped within one class (Glaucophyceae) and 4 genera (*Cyanophora*, *Cyanoptycha*, *Glaucocystis* and *Gloeochaete*); a very small number in comparison to more than 7000 rhodophyte species and almost 11,000 species of green algae and land plants (Guiry & Guiry, 2017). However, there might be cryptic species diversity difficult to identify because of culture biases (Jackson et al., 2015).

To harvest light, cyanelles use phycobilisomes which are complexes of phycobiliproteins (e.g. C-phycocyanin and allophycocyanin) attached to the thylakoids that together with the chlorophyll a give to glaucophyte cyanelles their characteristic blue color and occur only in cyanobacteria, glaucophytes and red algae (Green & Parson, 2003). Cyanelles retained the peptidoglycan wall and carboxysomes from the cyanobacterial ancestor, in contrast to Rhodophyta and Viridiplantae where these components were lost (Fathinejad et al., 2008; Pfanzagl et al., 1996).

This primitive-like characters of the cyanelle in addition to some phylogenetic reconstructions of nuclear and plastid genes suggest that Glaucophyta was the first lineage to diverge within Archaeplastida (Burki et al., 2008; Reyes-Prieto & Bhattacharya, 2007; Rodríguez-Ezpeleta et al., 2005). Nonetheless, phylogenetic and phylogenomic studies supporting “Rhodophyta-first” or “Viridiplantae-first” models are also common in the scientific literature. Thus, the branching order of Archaeplastida lineages continues to be an unresolved question (Deschamps & Moreira, 2009; Hagopian et al., 2004; Mackiewicz & Gagat, 2014).

1.2.3 *Paulinella chromatophora*, a second primary plastid-bearing lineage

There are only two recognized primary endosymbioses: the first originated the plastids of Archaeplastida whereas the second one took place in the heterotrophic ancestor of the filose amoeba *Paulinella chromatophora* (Marin et al., 2005). In the case of *Paulinella*, the chromatophore (chloroplast) is derived from an endosymbiosis with a cyanobacterium closely related to the marine *Synechococcus/Prochlorococcus* (Syn/Pro) clade (Marin et al., 2005). Interestingly, Syn/Pro members are known to be common

cyanobionts of sponges, and can be transmitted vertically (Burgsdorf et al., 2015; Steindler et al., 2005).

Similar to cyanelles in glaucophytes, chromatophores are still surrounded by a peptidoglycan cell wall and harbor carboxysomes, two plesiomorphic characters attesting the cyanobacterial ancestry (Kies, 1974). Most photosynthetic *Paulinella* species live in freshwater environments, however, a recently described species named *Paulinella longichromatophora* showed a transition from freshwater to marine habitats (Kim & Park, 2016).

According to molecular clock estimations, photosynthetic *Paulinella* appeared within the Rhizaria lineage somewhere between 60-140 million years ago (Delaye, Valadez-cano, & Pérez-zamorano, 2016; Nowack, Melkonian, & Glöckner, 2008). Nevertheless, other recent studies suggest that the chromatophore is more ancient and appeared approximately 500 Mya (Sánchez-Baracaldo et al., 2017). Regardless of the age estimation, it is clear that the primary endosymbiosis in *Paulinella* is a more recent event than the one of Archaeplastida which occurred 1.6 bya (Yoon et al., 2004). The younger age of the *Paulinella* chromatophore correlates well with the larger size of its genome and the number of encoded proteins compared to the much more reduced plastids of Archaeplastida. While in *Paulinella* the size of the chromatophore genome is approximately 1 Mb (with about 860 protein-coding genes) (Lhee et al., 2017), green and red algae commonly harbor chloroplasts in the range of 100-250 kbp (~80-220 proteins) (Lee et al., 2016). Nonetheless, the chromatophore genome is also heavily reduced. It encodes only about 25 % of proteins in comparison with free-living cyanobacteria of the Syn/Pro clade. Moreover, it seems that the genome reduction (through pseudogenization and gene loss) is still in progress (Nowack et al., 2008).

Gene content in chromatophores has also been affected by the relocation of chromatophore genes to the host nucleus by EGT. More than 50 putative EGTs have been detected in the nuclear genome of *Paulinella chromatophora* which after transcription and synthesis in the cytoplasm are targeted back to the chromatophore. Most of these proteins participate in photosynthesis and photoacclimatation-related functions (e.g photosystem I subunits PsaE and PsaK) (Nowack et al., 2011, 2016; Nowack & Grossman, 2012).

Interestingly, many genes horizontally transferred from non-cyanobacterial prokaryotes also function in the chromatophore, probably compensating for the loss of their ancestral chromatophore homologs (Nowack et al., 2016). However, the higher contributor of chromatophore-targeted proteins in *Paulinella chromatophora* is neither the

cyanobacterial endosymbiont (EGT) nor other bacteria (HGT) but the host (Singer et al., 2017).

1.3 Origin and evolution of eukaryotic lineages with complex plastids

After primary endosymbiosis, other eukaryotic cells acquired the ability to photosynthesize via the uptake of a red or green algal endosymbiont, a symbiogenetic event known as secondary endosymbiosis. Plastids in chlorarachniophytes and euglenids originated from two independent secondary endosymbioses with green algae whereas cryptophytes, alveolates, stramenopiles and haptophytes (the CASH lineage) acquired a red algal-derived plastid. Secondary endosymbiotic events have had an important impact in the distribution of photosynthetic lineages across the eukaryotic tree (Fig. 8).

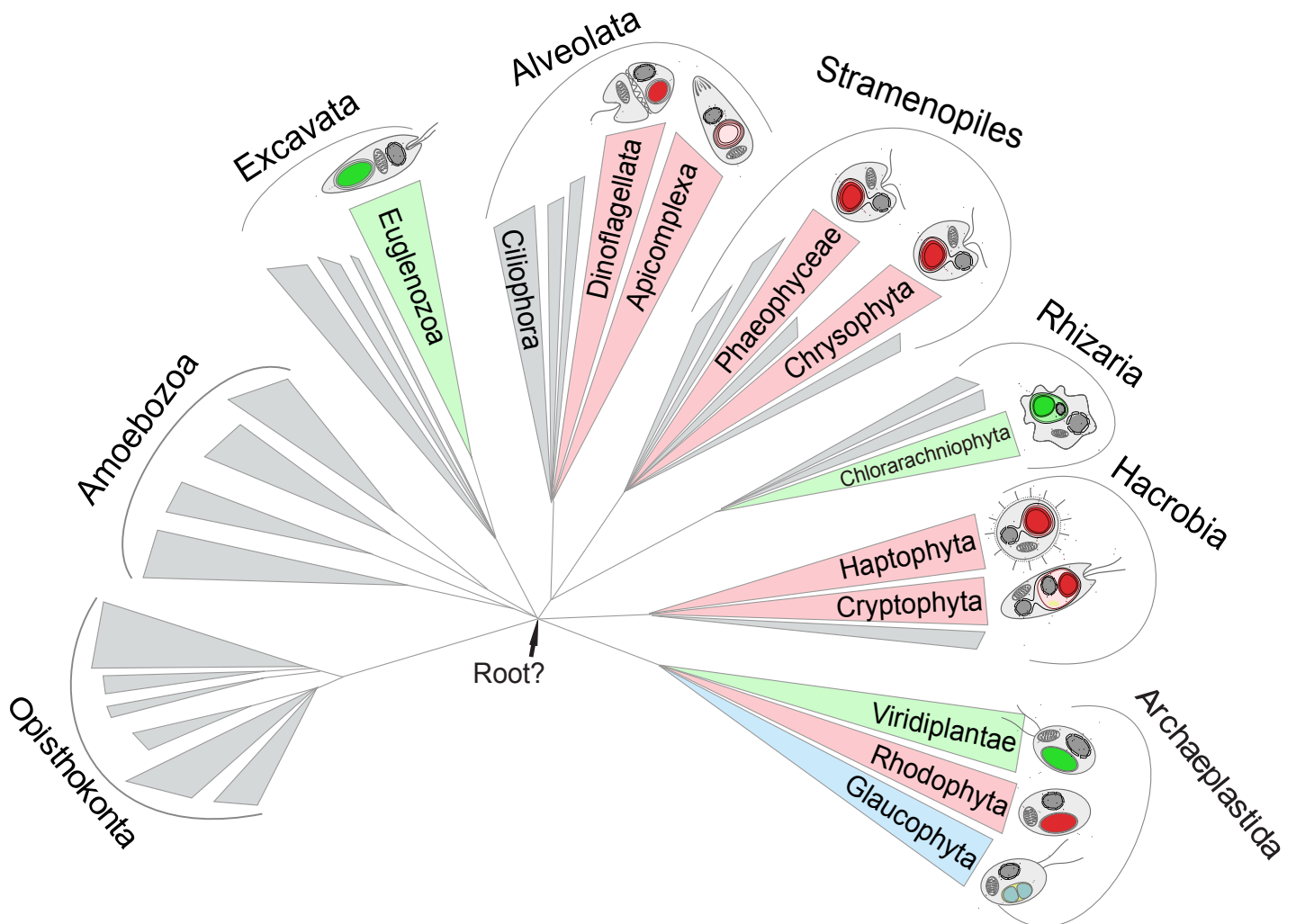


Figure 8. Distribution of photosynthetic lineages across the eukaryotic tree.

1.3.1 Secondary green plastids

1.3.1.1 Chlorarachniophytes

Chlorarachniophytes are marine photosynthetic amoeboflagellates with plastids that originated from a secondary endosymbiosis with a green alga (Ludwig & Gibbs, 1989). They belong to the phylum Cercozoa within the supergroup Rhizaria (Burki & Keeling, 2014) which is widely present in marine plankton with important roles in biogeochemical cycles (Biard et al., 2016).

Similar to cryptophytes, the four membrane-bound plastid in chlorarachniophytes carry a nucleomorph within the periplastid compartment (i.e. the space between the second and third membranes of the plastid); they are the remnant of the nucleus and cytosol of the green algal endosymbiont, respectively (Keeling, 2010). During secondary endosymbiosis, a filose amoeba engulfed an ulvophycean green alga (Suzuki et al., 2016); although the phagotrophic capability was retained in some mixotrophic species (e.g. *Chlorarachnion reptans*), other members have secondarily lost this feeding behavior (e.g. *Lotarella globosa*) (Ishida et al., 1999).

1.3.1.1.1 Nucleomorph

As expected for an organelle that has been completely lost in many secondary plastid-harboring algae, the nucleomorph is heavily reduced. The nucleomorph genomes consist of three linear chromosomes ranging from 330 to 1000 kbp (Hirakawa, 2017). The nucleomorph genome of *Bigelowiella natans* was the first to be completely sequenced. It showed a high density of genes, with extremely small introns, probably due to selective pressure that resulted in the genome reduction (Gilson et al., 2006). Interestingly, similar genome architecture has been found in all nucleomorph genomes sequenced to date, suggesting that the massive genome reduction took place before the diversification of extant species (Suzuki et al., 2015; Tanifuji et al., 2014).

Nucleomorph genomes encode mostly housekeeping genes necessary to the maintenance of the nucleomorph and the expression of 17 proteins targeted to the plastid. This set of proteins is shared among all nucleomorph genomes and are thought to be the reason why chlorarachniophytes still harbors a nucleomorph (Hirakawa, 2017; Suzuki et al., 2015). Moreover, nucleomorph-encoded genes have high rate of evolutionary change (Hirakawa & Ishida, 2014; Patron et al., 2006a) probably due to the loss of genes involved in DNA replication and repair (Curtis et al., 2012).

Although the plastid in chlorarachniophytes and cryptophytes originated from two independent secondary endosymbioses (with a red alga in the case of cryptophytes), both nucleomorphs share important features such as the number of chromosomes, similarities in gene content (Tanifuji et al., 2014), conserved telomeric regions (Hirakawa, 2017), accelerated evolution of nucleomorph-encoded genes (Patron et al., 2006a) and polyploidization (Hirakawa & Ishida, 2014). Convergent evolution in the genomic architecture of the nucleomorphs of chlorarachniophytes and cryptophytes shows that both organelles have apparently experienced similar evolutionary pressures during secondary endosymbiosis.

1.3.1.1.2 Plastid

Plastids in chlorarachniophytes are surrounded by four membranes (Fig. 9A). The first membrane (from outside to inside) correspond to the phagosomal membrane that encapsulated the algal endosymbiont; the second membrane corresponds to the plasma membrane of the green algal endosymbiont whereas the two innermost membranes correspond to the chloroplast membranes where photosynthesis takes place. Similar to chlorophytes, chlorarachniophyte plastids use chlorophyll a and b (Hibberd & Norris, 1984).

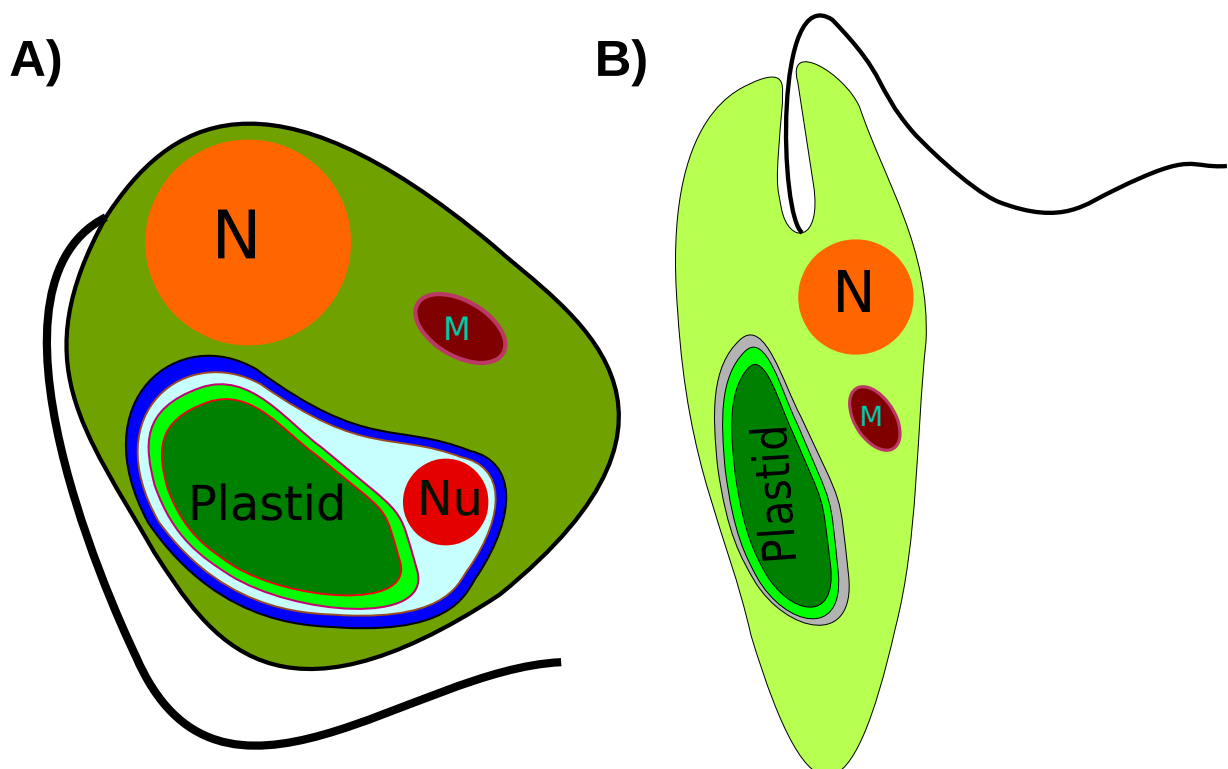


Figure 9. Photosynthetic lineages with secondary green plastids. A) Chlorarachniophytes. B) Euglenophytes. Nu: nucleomorph; M: Mitochondria; N: nucleus.

Chlorarachniophyte plastids have circular genomes that encode approximately 60 proteins (Suzuki et al., 2016b). Phylogenetic analyses of 38 concatenated plastid proteins showed that chlorarachniophytes and euglenids (both with secondary green plastids) acquired independently the ability to photosynthesize (Rogers et al., 2007). The green endosymbiont in chlorarachniophytes derived from a member of the UTC clade (Ulvophyceae-Trebouxiophyceae-Chlorophyceae). Recent analysis with enlarged taxon sampling and larger number of concatenated markers (55 plastid-encoded proteins) improved the placement of the endosymbiont within the UTC clade and showed that the chlorarachniophyte plastids derived from a green alga from the order Bryopsidales (Ulvophyceae) (Suzuki et al., 2016b).

During secondary endosymbiosis, genes from the plastid and nuclear genomes of the green endosymbiont were relocated to the host nucleus. As a result, most of the plastid proteins in chlorarachniophytes were transferred. For instance, the plastid genome of *Bigelowiella natans* encodes only for 57 proteins (mostly involved in photosynthesis and protein biosynthesis) (Rogers et al., 2007), whereas around 780 plastid-targeted proteins are encoded in the nucleus (Suzuki et al., 2016a). Interestingly, phylogenetic analyses have revealed an important contribution of genes from algae with red algal-derived plastids to the *Bigelowiella* plastid proteome (Curtis et al., 2012).

1.3.1.2 Euglenophytes

The phylum Euglenozoa (Excavata) comprises kinetoplastids, diplomonids and euglenids (Adl et al., 2012). The latter includes photosynthetic and colorless protists. Photosynthetic euglenids such as the model species *Euglena gracilis* belong to the class Euglenophyceae, a monophyletic assemblage that also includes members that secondarily lost the ability to photosynthesize (Bicudo & Menezes, 2016). The plastid in euglenophytes derived from an endosymbiotic event with a *Pyramimonas*-like alga (Prasinophyceae) approximately 100 Mya, possibly in a marine environment (Hrdá et al., 2012; Parfrey et al., 2011).

Euglenophytes are divided in two orders: Eutreptiales, a paraphyletic assemblage composed mostly of marine species; and Euglenales, a monophyletic group widely present in freshwater habitats (Vanclová et al., 2017). *Rapaza viridis*, the sister lineage to all euglenophytes, is an obligatory mixotroph that feeds on green algae from the genus *Tetraselmis*. Interestingly, it is thought that *Rapaza* represents the intermediate stage in plastid acquisition within euglenophycean algae (Yamaguchi et al., 2012).

1.3.1.2.1 Plastid

Once thought to have originated from the same primary endosymbiosis that Viridiplantae (Cavalier-Smith, 1982), or from a common secondary endosymbiosis with chlorarachniophytes (Cavalier-Smith, 1999), it is known today that euglenophytes possess a secondary plastid that was acquired independently in a single endosymbiotic event with a green alga related to *Pyramimonas* (Hrdá et al., 2012). The plastid has three membranes (Fig. 9B), two from the ancestral chloroplast and the origin of the third one is still unclear; it has been suggested to be the vestige of the plasma membrane of the green endosymbiont (Gibbs, 1981) or the food vacuole membrane that encapsulated the alga (Cavalier-Smith, 1999).

Euglenid plastid genomes have a surprising similarity in gene content, approximately 90 genes, mostly involved in photosynthesis and housekeeping functions (Vanclová et al., 2017). An exception is the non-photosynthetic plastid in *Euglena longa* which encodes only 57 genes (Gockel & Hachtel, 2000). *E. longa* is an osmotrophic species phylogenetically close to *E. gracilis*, although its plastid is no longer photosynthetic, it is compulsory for the survival of the protist (Hadariová et al., 2017). The plastid has lost all photosynthetic genes except the *rbcL* gene (large subunit of the enzyme RuBisCO). The *rbcL* protein is extremely divergent and although it is thought to be the reason why *Euglena longa* still retain the non-photosynthetic plastid, its function is still unknown (Záhonová et al., 2016).

1.3.2 Eukaryotic lineages with red algal-derived plastids

1.3.2.1 The Chromalveolate hypothesis

The description of algal cells boosted by the improvement of microscope resolution revealed the hidden complexity of plastid morphology and structure. Although early phylogenetic analyses showed that all plastids derived from cyanobacteria (Delwiche et al., 1995; Nelissen et al., 1995), a single primary endosymbiotic event was not sufficient to explain the structural diversity of plastids, making necessary to propose that some photosynthetic lineages could have evolved from eukaryote-into-eukaryote endosymbioses (Cavalier-Smith, 1982; Gibbs, 1981).

The Chromalveolate hypothesis posits that photosynthetic lineages with chlorophyll c-containing plastids (i.e. cryptophytes, haptophytes, stramenopiles and dinoflagellates)

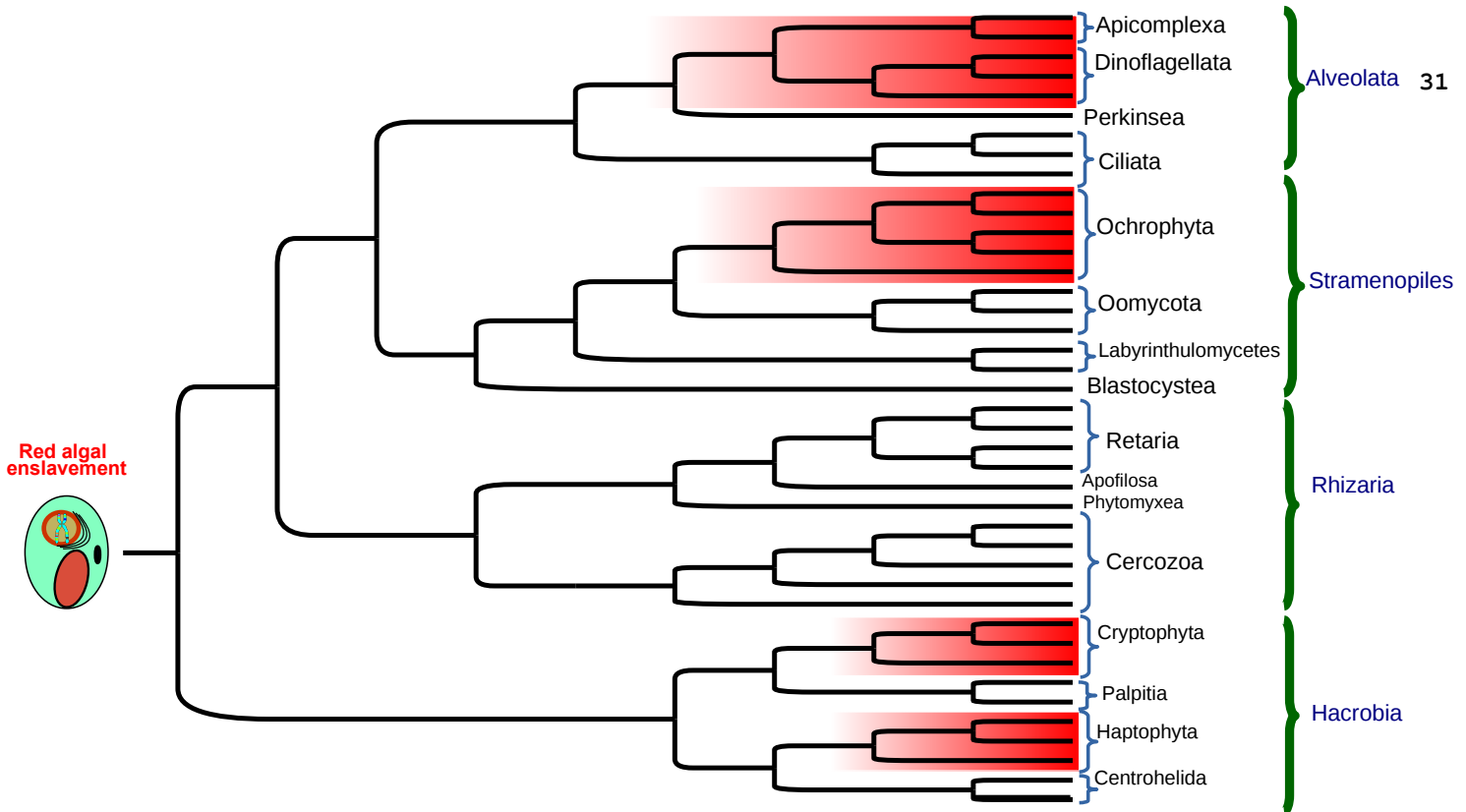


Figure 10. Single red algal enslavement at the origin of secondary red plastids (adapted from Cavalier-Smith, 2017).

evolved from a single endosymbiotic event with a red alga deep in the evolutionary history of eukaryotes (Cavalier-Smith, 1999) (Fig. 10).

These chromalveolates include heterotrophic lineages that, according to the hypothesis, lost their plastids (e.g. ciliates) or the ability to photosynthesize (e.g. apicomplexans). Interestingly, the discovery of non-photosynthetic plastids in *Toxoplasma gondii* and *Plasmodium falciparum*, two alveolate parasites, strongly suggested a photosynthetic past (McFadden et al., 1996). Within the chromalveolates there are also lineages that have replaced their original red plastid such as the dinoflagellates from the genus *Lepidodinium* that replaced their peridinin-containing plastid by a green-colored one (Matsumoto et al., 2011).

Attesting of the secondary origin of the plastid, the ancestor of all chromalveolate plastids is thought to have been surrounded by four membranes. From outside to the inside of the plastid: 1) the phagosomal membrane that encapsulated the red alga; 2) the plasma membrane of the red endosymbiont; 3) and 4) are homologous to the outer and inner membranes of the chloroplast, respectively (Cavalier-Smith, 1999). The four-membrane-bound plastid is still observed in ochrophytes, cryptophytes, haptophytes, chromerids and apicomplexans whereas in peridinin-containing dinoflagellates, the plastid has only three membranes. Although the extra membrane was once considered to be the

residual phagosomal membrane that engulfed the cyanobacterium during primary endosymbiosis (Cavalier-Smith, 1982), today, the secondary nature of the dinoflagellate plastids is widely accepted (Janouskovec et al., 2010). According to the chromalveolate hypothesis, the third outermost membrane of dinoflagellate plastids corresponds to the food vacuole membrane that was retained whereas the periplastid membrane was lost in the common ancestor of all dinoflagellates (Cavalier-Smith, 1999).

Other peculiarity in secondary red algae besides the number of membranes is that in the case of cryptophytes, haptophytes and ochrophytes the outermost membrane (membrane 1) is fused with the nuclear envelope. It was initially suggested that it occurred in the common ancestor of stramenopiles and Hacrobia (cryptophytes plus haptophytes) but phylogenomic analyses did not support the monophyly of the group (Burki et al., 2016). Thus, if a single red algal enslavement took place there was at least two independent membrane fusions, one in the ancestral Hacrobia and the other in stramenopiles (Cavalier-Smith, 2017).

Perhaps the most relevant assumption of the chromalveolate hypothesis is the large number of independent plastid losses, at least 18 (Cavalier-Smith, 2002), necessary to explain the diversity of heterotrophic lineages and their distribution within the eukaryotic tree. A photosynthetic ancestor of chromalveolates implies that plastid-less protist lineages such as Ciliophora (Alveolata), the human gut inhabitant *Blastocystis* (Heterokonta) (Beghini et al., 2017) or the phyla Telonemia and Centroheliozoa which are closely related to Hacrobia (Burki et al., 2010), must have lost their plastids at sometime in the course of evolution. Interestingly, some nuclear genes of ciliates seem to have a red algal origin and have been suggested to be residual EGTs from the red endosymbiont of the common ancestor of all alveolates (Reyes-Prieto et al., 2008). On the other hand, there is no evidence that the ancestor of stramenopiles harbored a plastid; it seems rather that the plastid was acquired after the divergence of stramenopiles in the ancestor of ochrophytes (Leyland et al., 2017; Wang et al., 2017).

1.3.2.2 Common red algal endosymbiont

Although not sufficient, the monophyly of secondary red plastids is a necessary condition to support a single origin of all chromalveolates. Early phylogenies of ribosomal RNA sequences suggested that the red plastids in cryptophytes, haptophytes and stramenopiles did not share a common ancestry (Müller et al., 2001; Oliveira & Bhattacharya, 2000). However, these results were probably due to phylogenetic artifacts.

Improved taxon sampling coupled with phylogenetic analyses of concatenated plastid genes have consistently retrieved the monophyly of chlorophyll *c*-containing plastids (Janouskovec et al., 2010; Sanchez-Puerta et al., 2007; Yoon et al., 2002).

Within the phylogeny of red algal-derived plastids, cryptophytes seem the first to diverge and appeared either sister to all other red lineages (Lida et al., 2007; Shalchian-Tabrizi et al., 2006) or sister to haptophytes (with cryptophytes and haptophytes as sister group of alveolates and stramenopiles) (Kim et al., 2017; Sanchez-Puerta et al., 2007). The closer phylogenetic proximity between cryptophytes and haptophytes is also supported by the shared bacterial *rp136* gene in their plastid genomes, likely transferred to their common ancestor (Rice & Palmer, 2006).

1.3.2.3 Protein import system in red algal-derived plastids

Secondary plastids in cryptophytes, haptophytes, stramenopiles and apicomplexans are surrounded by four membranes whereas the peridinin-containing plastids of dinoflagellates have only three probably due to the loss of the periplastid membrane of the red algal endosymbiont (Cavalier-Smith, 1999). Another important difference among red algal-derived plastids is that while in haptophytes, cryptophytes and stramenopiles the outermost membrane of the plastid is fused with the endoplasmic reticulum (ER) of the host, this is not the case for the plastids in alveolates (e.g. apicomplexans and dinoflagellates).

To import proteins across the first membrane of four membrane-bound plastids, the sec61 complex of the host ER is used to cross co-translationally proteins carrying a signal peptide (SP). In the case of chromerids and apicomplexans, whose plastids are not fused with the host ER, proteins targeted to the plastid enter the secretory pathway of the host and are transported within vesicles thought to fuse with the outermost membrane of the plastid (Bouchut et al., 2014; Waller et al., 2000).

Then, to cross the second outermost membrane, it is thought that the ERAD (Endoplasmic Reticulum-Associated protein Degradation) machinery of the red algal endosymbiont was recycled and relocated to this membrane (Sommer et al., 2007) (Fig. 11). Plastid-targeted and nucleomorph-encoded ERAD-like proteins have been identified in all red lineages carrying plastids surrounded by four-membranes (Felsner et al., 2011). This ERAD-derived system specific to the symbiont has been named SELMA which stands for Symbiont-specific ERAD-like Machinery (Hempel et al., 2009). Interestingly, phylogenetic analyses of some SELMA components have shown to be monophyletic and

derived from the red algal endosymbiont. SELMA is considered to provide strong evidence for the existence of a single secondary endosymbiosis with a red alga at the origin of all complex red plastids (Gould et al., 2015).

The two innermost membranes of complex red plastids correspond to the membranes of primary plastids. Similar to protein import in Archaeplastida lineages, complex red plastids use proteins recycled from the TIC/TOC complex of the red endosymbiont (Stork et al., 2013) (Fig. 11).

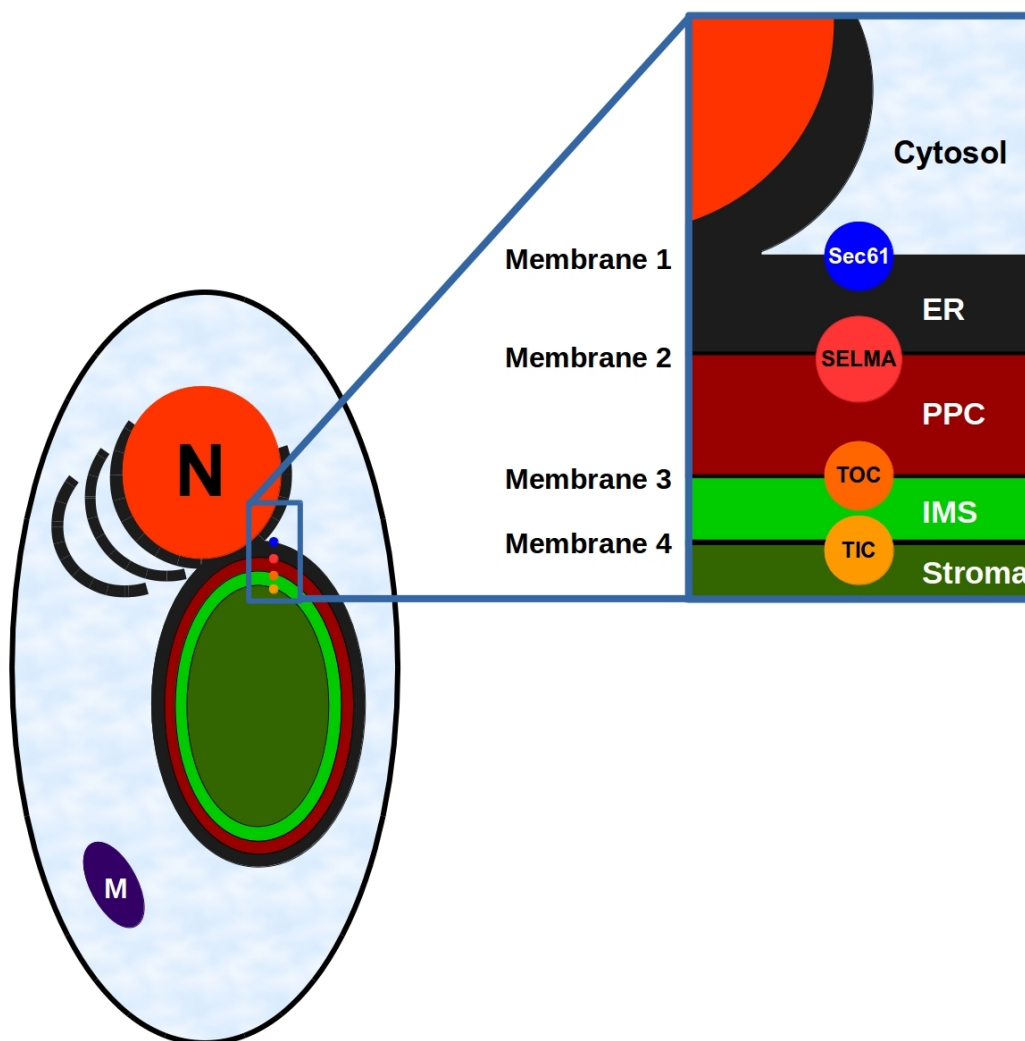


Figure 11. Distribution of translocons in secondary red plastids surrounded by four membranes. The figure depicts the architecture of secondary red plastids in which the outermost membrane is fused with the host ER such as in stramenopiles, haptophytes and cryptophytes. However, cryptophytes harbor a nucleomorph, the vestigial nucleus of the red algal endosymbiont, within the PPC. ER: endoplasmic reticulum; PPC: periplastidial compartment; IMS: Intermembrane space.

1.3.2.4 One host or multiple hosts?

The monophyly of secondary red plastids is widely accepted. However, whether it represents a single symbiogenetic event followed by vertical descent or a single acquisition followed by horizontal transfer via serial endosymbiosis is still unclear. The symbiogenetic model of a single red algal secondary endosymbiosis at the origin of all chlorophyll c-containing plastids assumes that establishing a functional organelle is much more difficult than losing it, because of all the steps required to fully integrate a symbiont (e.g. transfer of genes to the host nucleus, development of plastid targeting system for nucleus-encoded proteins and evolution of a protein import system) (Cavalier-Smith, 1999). According to this parsimonious model, a single event is at the origin of the emergence of all red plastid-harboring algae.

Phylogenomic analyses of recently sequenced new protist lineages, together with the use of complex models of sequence evolution, have improved, although not completely resolved, the relationships among major eukaryotic supergroups (Burki et al., 2010; Yabuki et al., 2014). For instance, analyses of host-derived genes revealed that alveolates and stramenopiles, both with heterotrophic and photosynthetic lineages, are closely related to rhizarian protists, an ecologically and phylogenetically diverse supergroup of single-cell eukaryotes (Burki & Keeling, 2014). The assemblage of Stramenopiles, Alveolates and Rhizaria form a well-supported group named the SAR clade (Burki et al., 2010). Therefore, to propose a single red secondary endosymbiosis in the new context of the global eukaryotic tree, it is necessary to propose the loss of the red plastid at a certain point in the evolution of Rhizaria and a new secondary endosymbiosis with a green alga in an ancestor leading to chlorarachniophytes.

Almost all recent phylogenomic analyses have failed to recover the monophyly of chromalveolates (Baurain et al., 2010; Brown et al., 2013; Burki et al., 2008, 2012). For instance, the putative common origin of the host lineage in Cryptophyta and Haptophyta, once thought to be part of the monophyletic supergroup Hacrobia (Okamoto et al., 2009), is strongly questioned in recent phylogenomic analyses; instead, the nuclear genome of cryptophytes appears to have an unexpected proximity to Archaeplastida lineages (Burki et al., 2016). Models of serial transfer of red algal-derived plastids have been proposed as workarounds to reconcile the evolutionary history of nuclear and plastid genomes of secondary red algae.

1.3.2.5 Serial endosymbioses

Models of serial endosymbioses propose that a heterotrophic ancestor of cryptophytes enslaved a red alga that became a secondary plastid. Then, the cryptophyte plastid was spread across the eukaryotic tree through a series of tertiary, quaternary or even quinary endosymbioses (Fig. 12).

As a consequence, serial endosymbioses would have created a sort of “Russian doll” cell in eukaryotic lineages with red algal-derived plastids (Petersen et al., 2014). Besides that the chromalveolates is likely not a monophyletic assemblage (Baurain et al., 2010), the proposal of serial endosymbioses accounts for the fact that there are many heterotrophic lineages within the chromalveolates that seem to have never been photosynthetic. For instance, although several genes with algal affiliation in the genome of the plant pathogen *Phytophthora sojae* were originally interpreted as remnant genes of a photosynthetic past (Tyler, 2006), recent re-analyses showed that many genes were misinterpreted as algal-derived due to poor taxon sampling, thereby weakening the support for a common photosynthetic ancestry for all stramenopiles (Wang et al., 2017).

Additional support for serial endosymbiosis models comes from recent analyses of plastid-derived genes of groups thought to have inherited the plastid vertically such as the chromerids, peridinin-containing dinoflagellates (PCD) and apicomplexans. These alveolate lineages carry a red algal-derived plastid and early phylogenetic analyses of 23S rRNA and psbA genes showed that the plastid of PCD and the apicoplast appeared to share a common ancestor (Zhang et al., 2000). Likewise, *Chromera velia*, a photosynthetic symbiont of corals closely related to apicomplexans (Moore et al., 2008), shared the same type II Rubisco with dinoflagellates which was likely transferred horizontally from Proteobacteria, suggesting that the HGT took place in the common ancestor of both lineages (Janouskovec et al., 2010). However, Petersen et al. (2014) performed phylogenetic analyses of Calvin cycle markers known to have a common origin in red algal-derived plastids in addition to a handful of proteins necessary for the protein import in complex plastids. They showed that none of the plastid-located proteins they investigated supports a common ancestry between the plastid in PCD and the apicoplast. Instead, they proposed two independent acquisitions of secondary red plastids and proposed the “rhodoplex hypothesis” where red algal-derived plastids in the CASH (Cryptophytes-Alveolates-Stramenopiles-Haptophytes) lineages have a reticulate history as a result of independent eukaryote-eukaryote endosymbioses (secondary, tertiary or quaternary), although the path of plastid donors and acceptors remains unclear. Nonetheless, other

publications have suggested possible scenarios of serial plastid acquisitions (Bodyl, 2017; Bodyl et al., 2009; Kim et al., 2017; Ševčíková et al., 2015) (Fig. 12).

Using regression analyses, Stiller et al., (2014) propose that a red alga became the cryptophyte plastid followed by the enslavement of a cryptophyte cell by the phagotrophic ancestor of ochrophytes via tertiary endosymbiosis. Finally, a quaternary uptake of an ochrophyte gave rise to the haptophyte plastid possibly after haptophytes diverged from centrohelids (Fig. 12).

An alternative sequence of endosymbioses was proposed by Bodyl et al., (2009). In their model, after the symbiogenesis of cryptophytes via the enslavement of a red alga, two independent tertiary endosymbioses took place: the first one in the ancestor of ochrophytes by the uptake of a cryptophyte plastid before the insertion of a bacterial *rpl36* gene in the plastid genome of cryptophytes; the second tertiary endosymbiosis would have

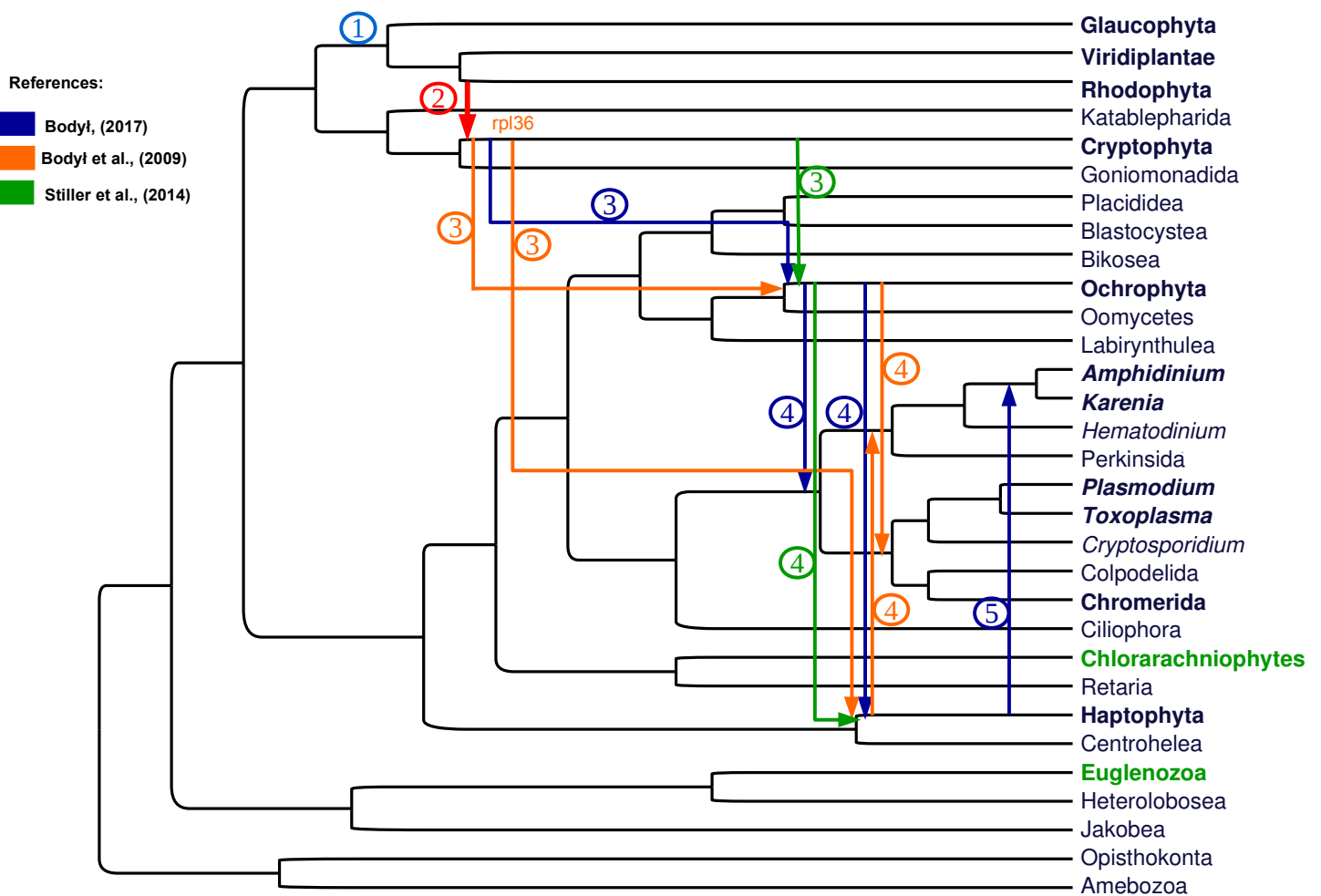


Figure 12. Models of serial endosymbioses according to different studies. Numbers represent the sequence of endosymbiotic events, see text for a detailed description. Present-day plastid-harboring lineages are represented in bold. Chlorarachniophytes and euglenids have both secondary green plastids but their endosymbioses were omitted for clarity.

occurred between a member of the cryptophyte lineage that laterally acquired the bacterial *rp136* gene and the ancestor of haptophytes, which explained the presence of the same bacterial HGT in both lineages (Rice & Palmer, 2006). Within the alveolates, the chromerids and apicoplasts would have acquired their plastids from an ochrophyte in a quaternary endosymbiotic event whereas the peridinin-containing plastid in dinoflagellates would have originated from an haptophyte (Fig. 12).

Models of serial endosymbioses generally assume that whole cells are encapsulated by phagotrophic cells, followed by a drastic reduction of the plastid genome and the loss of cytosolic compartments of the symbiont, such as mitochondria and the nucleus (except in chlorarachniophytes and cryptophytes which still carry a highly reduced nucleomorph). Nonetheless, Bodył, (2017) proposed that some protist lineages with red algal-derived plastids could have acquired their photosynthetic organelles not only from whole cells but also by kleptoplasty. He proposed that the extant 3-membrane bound plastid in peridinin-containing dinoflagellates could have been obtained by myzocytosis from an haptophyte after the loss of the former ochrophyte-derived plastid gained by the ancestor of dinoflagellates and apicomplexans (Fig. 12). The suggestion that photosynthetic eukaryotes with plastids surrounded by three membranes such as euglenophytes or dinoflagellates were acquired from isolated chloroplasts is, however, not new (Whatley & Whatley, 1981).

1.3.2.5.1 Tertiary endosymbiosis in dinoflagellates

Models of serial endosymbiosis consider that the chlorophyll c-containing plastids originated from a single secondary endosymbiosis with a red alga that was horizontally spread to other eukaryotic groups via tertiary and higher order endosymbioses (Petersen et al., 2014; Stiller et al., 2014). However, the presence of tertiary plastids has hitherto been proved only in dinoflagellates (Janouškovec et al., 2017).

Dinoflagellates have a complex and poorly understood evolutionary history. They have undergone major changes such as the complete loss of plastids in the crustacean parasite *Hematodinium* (Gornik et al., 2015) or replacement of the former red plastid in *Lepidodinium chlorophorum* with a green one through another secondary endosymbiotic event. (Kamikawa et al., 2015). Dinoflagellates possess some genomic particularities that differentiate them from other protists possibly due to particular selective pressures. For instance, their nuclear genomes are extremely large, usually in the range of 3,000 –

215,000 Mb (Hackett et al., 2004); for comparison, the genome of *Chromera velia* (Alveolata), a coral-associated alga, is approximately 194 Mb in size (Woo et al., 2015). Moreover, the nuclear DNA of dinoflagellates is greatly methylated (between 12-70% of thymine is replaced by 5-hydroxymethyluracil) (Lin, 2011).

Besides the intricate evolution of the nucleus, the plastid of dinoflagellates shows a similar pattern. Dinoflagellates harbor different types of tertiary plastids. Molecular phylogenies of plastid-derived proteins as well as the presence of accessory pigments typical of haptophytes (e.g. 19'-hexanoyloxyfucoxanthin) (Zapata et al., 2012) in members of the Gymnodiniales lineage (e.g. *Karlodinium veneficum* and *Karenia brevis*) strongly support the replacement of the former peridinin-containing plastid by an haptophyte-derived one via a tertiary endosymbiosis. Akin to other complex red plastids, the cytosol and the nucleus of the haptophyte endosymbiont were lost (Burki et al., 2014; Hwan et al., 2005; Patron et al., 2006b).

Another example of tertiary endosymbiosis in dinoflagellates is the diatom-derived plastid found in dinoflagellates (dinotoms) from the order Peridinales (e.g. *Durinskia baltica*) (Janouškovec et al., 2017). The plastid is surrounded by five membranes and retained cellular components of the diatom endosymbiont such as the nucleus, mitochondria and endoplasmic reticulum (Dodge, 1971; Imanian & Keeling, 2007; Waller & Kořený, 2017). Interestingly, plastids in dinotoms are not monophyletic, these dinoflagellates have taken up different diatom endosymbionts through independent endosymbiotic events (Horiguchi & Takano, 2006; Yamada et al., 2017).

It has been suggested that dinoflagellates belonging to the genus *Dynophysis* have tertiary plastids of cryptomonad origin, possibly closely related to *Teleaulax amphioxeia* (Garcia-Cuetos et al., 2010). Nonetheless, the discussion whether the cryptophyte-derived plastid is a permanent plastid or a kleptoplast (i.e. a "stolen" chloroplast obtained from food capable of carrying out photosynthesis but eventually eliminated) is still unclear (Feinstein et al., 2002; Garcia-Cuetos et al., 2010; Waller & Kořený, 2017).

1.3.2.5.2 Kleptoplasty: an early step in endosymbiosis?

Photosymbiosis is widely spread across the eukaryotic tree, from unicellular protists to metazoans with a variable degree of symbiont integration from temporal acquired phototrophy to permanent plastids (Dorrell & Howe, 2012; Stoecker et al., 2009) (Fig. 13). To create a permanent endosymbiont, it is necessary at least: 1) an export system for the photosynthate which needs to be available to the host metabolic machinery; 2) the transfer

of plastid-encoded genes into the nucleus of the host and retargeting system to the plastid; 3) the development of an import system of nuclear-encoded proteins; 4) the host must take control over the plastid division. To fulfill these requirements, major evolutionary innovations are needed. Therefore, fully integrated endosymbionts are rather scarce in the natural world. For instance, only two primary plastid endosymbiotic events have been hitherto identified (Keeling, 2010).

Kleptoplasts are the preserved chloroplasts of algal prey retained by some eukaryotic lineages. Although the host needs to continue feeding on algae, the lifespan of functional kleptoplasts can be from short periods (hours) up to months (Stoecker et al., 2009). Interestingly, it has been suggested that kleptoplasty might represent an intermediate stage of endosymbiosis (Leliaert et al., 2012).

Mesodinium rubrum, a red tide-forming ciliate, preys on cryptophytes and can retain their plastids (kleptoplasts) and nuclei which remain transcriptionally active for up to 30 days (Johnson et al., 2007; Qiu et al., 2016). *M. rubrum* is an obligate mixotroph that requires regular uptake of cryptophytes from the *Teleaulax* or *Geminigera* genera to survive (Myung et al., 2011; Nishitani et al., 2010). There are important differences in

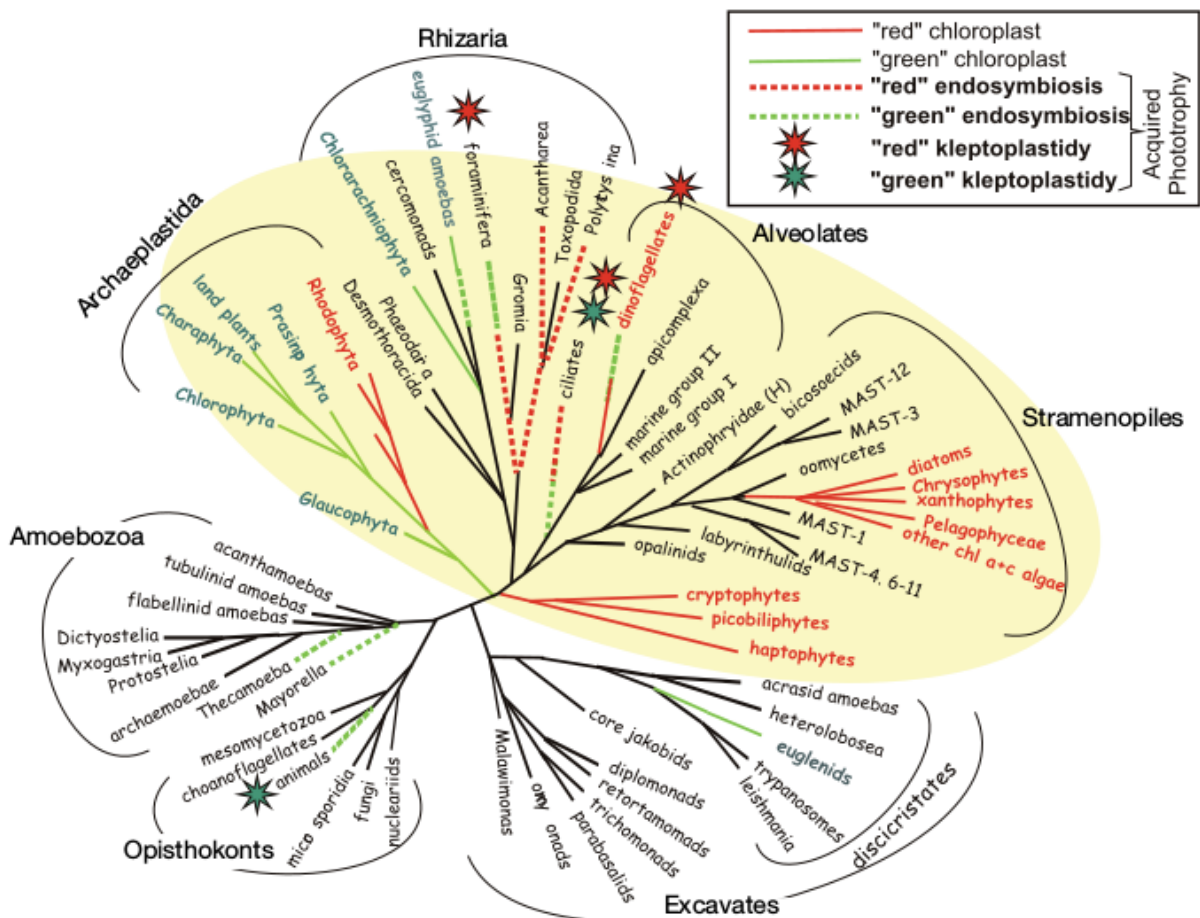


Figure 13. Distribution of permanent plastids and kleptoplasts in the eukaryotic tree (modified from Stoecker et al., 2009)

plastid retention and the preference of cryptophyte species among species of the genus *Mesodinium*. For instance, *Mesodinium chamaeleon* has a higher plastid retention capacity and the prey preference is less strict than in *Mesodinium rubrum* suggesting a different pattern of host-symbiont evolution within the genus (Moeller & Johnson, 2017).

The long term retention of photosynthetic chloroplasts by the sarcoglossan sea slug *Elysia* represents an example of kleptoplasty in metazoans. The stolen chloroplasts reside within the cells of their digestive gland giving them a green-colored pigmentation (de Vries et al., 2014; Händeler et al., 2009). Although the kleptoplast can perform photosynthesis, it is not strictly necessary for the survival of the slug (Christa et al., 2013). Interestingly, it is thought that *Elysia chlorotica* horizontally acquired several genes from its stramenopile symbiont *Vaucheria litorea*. These genes encode plastid-located proteins that participate in the Calvin cycle, chlorophyll biosynthesis, light-harvesting complex and assemblage of the photosystem II (Pierce et al., 2009, 2012, Rumpho et al., 2008, 2009). Nonetheless, recent transcriptomic and genomic analyses have heavily reduced the estimation of the contribution of HGT of algal genes to the nuclear genome of the host, suggesting that if any member of the polyphyletic photosynthetic sarcoglossans is on the track of establishing a permanent symbiont, it is still in an early stage (Bhattacharya et al., 2013; Wägele et al., 2011). Moreover, photosynthetic *Elysia* cannot inherit vertically their symbionts.

On the other hand, *Hatena arenicola* (Katablepharida) can pass the green endosymbiont *Nephroselmis* (Prasinophyceae) to one of the two daughter cells during division which seemingly represent an ongoing tertiary endosymbiosis (Okamoto & Inouye, 2006). The colorless *Hatena* develops a feeding apparatus *de novo* during a predatory phase that allows the uptake of a new *Nephroselmis* cell and subsequently induces the loss of the feeding apparatus thereby returning to a phototrophic lifestyle (Okamoto & Inouye, 2006). As a consequence of continuous algal replacement, same strains of *Hatena arenicola* can harbor different *Nephroselmis* species (Yamaguchi et al., 2014).

As mentioned above, HGT of plastid genes to the host nucleus is necessary to establish a permanent chloroplast but there has not been a systematic search of algal HGTs in *Hatena arenicola* to estimate the degree of integration of the endosymbiont. Nevertheless, the green alga is always inherited by the right-hand daughter cell during division (when the parent is observed from the dorsal side) which suggest a tight control of the endosymbiont inheritance (Okamoto & Inouye, 2006; Yamaguchi et al., 2014).

After having exposed the diversity of photosynthetic eukaryotes as well as the certainties and doubts concerning their evolution, I present our results of the origin and evolution of primary (Chapter 4) and secondary plastids (Chapter 5 and 6). The discussion and the conclusions of my PhD are presented in the chapters 7 and 8, respectively.

2. OBJECTIVES

“Le hasard, pourrait-on dire, engendre un vaste nombre d’individus ; une petite partie de ceux-ci se sont retrouvés organisés de telle sorte que leurs organes pouvaient satisfaire à leurs besoins. Le plus grand nombre n’avait ni adaptation, ni organisation : ces derniers ont tous péri. Ainsi, les espèces que nous voyons aujourd’hui ne sont rien d’autre qu’une petite partie de toutes celles que le destin aveugle a produit”

Pierre Louis Moreau de Maupertuis. *Essai de cosmologie* (1750).

2. OBJECTIVES

As I presented in the Introduction (section 1.1), the endosymbiotic theory of the origin of plastids (and mitochondria) had a very complex and long history of facing constant rejection that only finished when molecular phylogenetic analyses undoubtedly showed the prokaryotic nature of these organelles. Today, symbiogenesis is widely accepted and considered to be a major evolutionary process that reshaped the tree of eukaryotes. However, there are still unresolved questions regarding the evolution of photosynthetic eukaryotes, including the phylogenetic identity of symbionts and the number and type of endosymbiotic events.

In order to better understand the evolution and spread of photosynthesis among eukaryotes, we established three main objectives for my PhD work:

1. Identification of the living cyanobacterial lineage closest to primary plastids.

The cyanobacterial lineage that gave rise to primary plastids remained unknown. Previous studies aiming at identifying the plastid ancestor can be split in two groups depending on whether they support an early-branching or late-branching cyanobacterium as the symbiont that established the endosymbiotic relationship with a phagotrophic host thereby originating the Archaeplastida lineage, the first group of photosynthetic eukaryotes. Thus, the first objective of my PhD was to study the origin of primary plastids taking advantage of the increased sampling of cyanobacterial and plastid genome sequences. These recent sequencing efforts allowed us to use phylogenomic approaches in order to point out the closest extant lineage to the plastid ancestor. With this purpose, we analyzed both plastid- and nuclear-encoded genes of plastid origin (endosymbiotic gene transfers -EGTs-).

2. Study of the evolution of secondary plastid-harboring algae through the analysis of nuclear-encoded genes.

Photosynthetic eukaryotes harbor composite nuclear genomes with genes derived from diverse sources, including vertically inherited eukaryotic genes and EGTs acquired from their plastids. It is assumed that secondary photosynthetic eukaryotes contain a majority of “red” or “green” genes depending on the type of secondary plastid they harbor (red or green plastid, respectively) as a result of endosymbiotic gene transfers. In order to

study the evolution of secondary photosynthetic eukaryotes, we measured the contribution of red and green algal genes to their nuclear genomes. These analyses may help to identify possible cryptic endosymbioses, as it has been proposed for some secondary plastid-harboring lineages.

3. Phylogenetic analysis of the SELMA translocon.

SELMA (Symbiont-specific ERAD-like machinery) is an assemblage of proteins considered to be responsible of protein import through the second outermost membrane of secondary red plastids with four membranes (found in cryptophytes, haptophytes, alveolates and stramenopiles). It is widely accepted that SELMA has a unique origin so that its monophyly is often used as a major argument to support the hypothesis of a single origin of all secondary red plastids. Nonetheless, there has not been a systematic phylogenetic study of their components. The last part of my PhD focused on the phylogenetic analyses of SELMA proteins to shed light on the evolution of the protein import machinery of red algal-derived plastids.

3. MATERIALS AND METHODS

“Don’t compete! — competition is always injurious to the species, and you have plenty of resources to avoid it!”

Pëtr Kropotkin. *Mutual Aid: A Factor of Evolution* (1902)

3. MATERIALS AND METHODS

We studied the origin and evolution of photosynthetic eukaryotes using phylogenomic approaches. Although every part of my PhD project had its particularities depending on the aim of every subproject, we followed a similar methodological approach in all our studies. For a detailed description of the methodology, see the corresponding chapter.

3.1 Selection of markers

The initial step in all our studies was the selection of the most suitable queries to be used in BLASTP searches against our local database to select the phylogenetic markers. To study the origin of primary plastids (chapter 4) (Fig. 14), we used the 149 plastid-encoded proteins of the glaucophyte *Cyanophora paradoxa*. In chapter 5 (Fig. 15), we used the whole proteomes of *Guillardia theta* and *Bigeloviella natans*, two photosynthetic eukaryotes with red and green algal-derived secondary plastids, respectively, to evaluate the contribution of red and green algal genes to the mosaicism observed in secondary plastid-harboring lineages. Finally, to study the evolution of the SELMA translocon (chapter 6) (Fig. 16), the set of protein queries used in this study was retrieved from the largest collection of putative SELMA components reported so far Stork et al., (2012).

The number of genomes and transcriptomes in our local database varied from one study to the other. Before the start of a new study, we incorporated recently published sequences if they were relevant to our analyses to have a more comprehensive taxon sampling. See the materials and methods section in every article manuscript to consult the complete list of species used in every study. We performed reciprocal BLASTP searches against our local database using a homemade python script based on the BLASTP algorithm (Altschul et al., 1997) to increase the number of retrieved proteins. Multiple alignments of protein sequences were performed using MAFFT (Katoh & Standley, 2013) with default parameters. Poorly aligned regions of multiple sequences alignments were removed using BMGE (Crisuolo & Gribaldo, 2010) to keep only phylogenetic informative regions for posterior phylogenetic analyses.

3.2 Phylogenetic analyses

Two cycles of phylogenetic tree reconstruction were performed in trimmed sequenced alignments. First, preliminary phylogenetic trees were constructed using an approximately-ML approach, which is implemented in the program FastTree and is known to be a faster method for ML estimation (Price et al., 2010). In all our phylogenetic analyses, trees were visually inspected to avoid potentially paralogous proteins and identify protein markers fulfilling the criteria of selection previously established (see the corresponding chapter).

After selection of markers, maximum likelihood trees were constructed initially using PhyML (Guindon et al., 2009) (Chapter 4) but the new program IQTREE for maximum likelihood-based inference of phylogenetic trees has shown to outperform PhyML (Nguyen et al., 2015) and it was preferred in ulterior studies (chapter 5 and 6). In both programs, we used the LG substitution matrix (Le & Gascuel, 2008) and a four-category gamma distribution to reconstruct the phylogenetic trees. Bayesian inference of phylogenetic trees was conducted using PhyloBayes-MPI (Lartillot et al., 2013) with the CAT-GTR model (Lartillot & Philippe, 2004) and a gamma distribution with four rate categories.

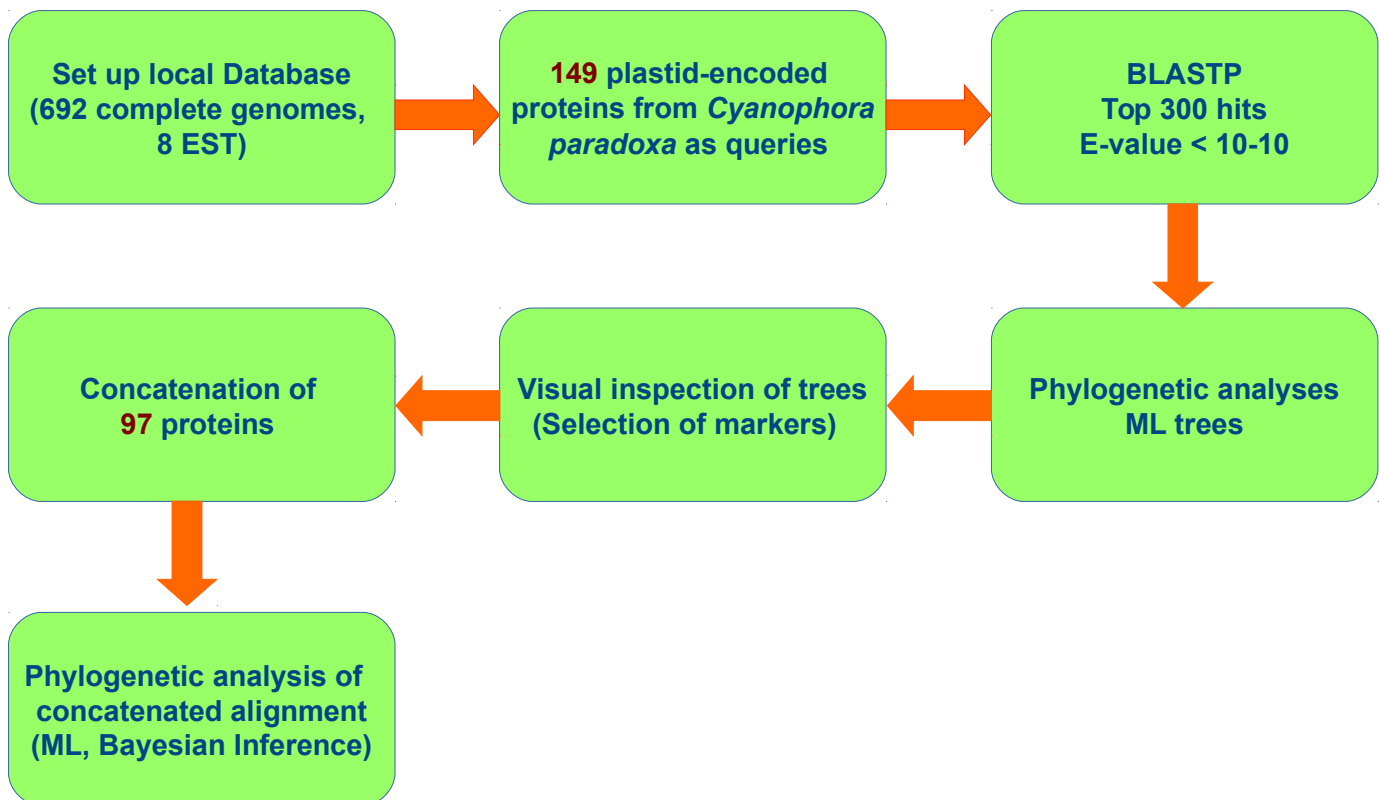


Figure 14. Diagram of the methodology used to study the origin of primary plastids (Chapter 4)

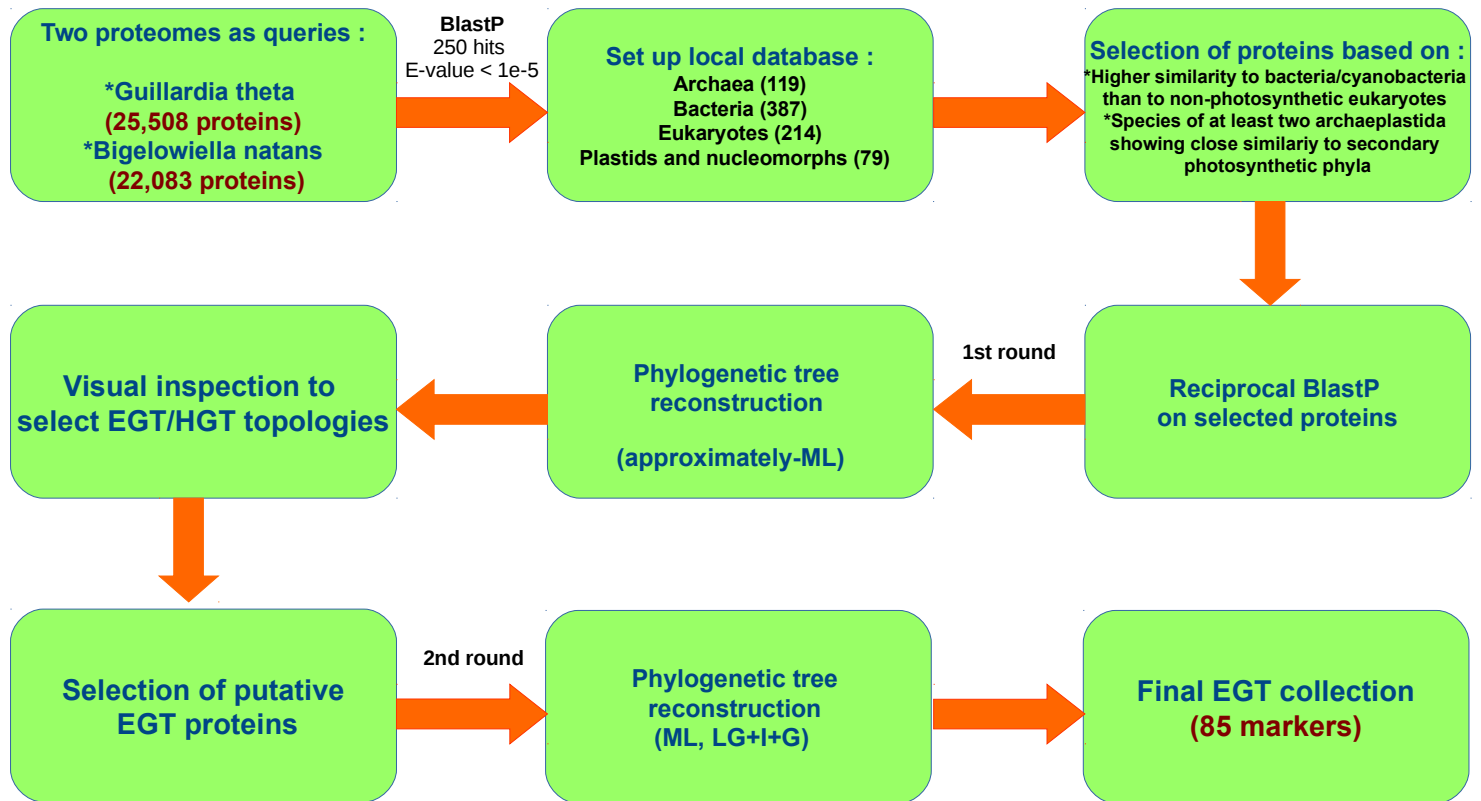


Figure 15. Diagram of the methodology used to study the evolution of eukaryotic lineages with secondary plastids (Chapter 5)

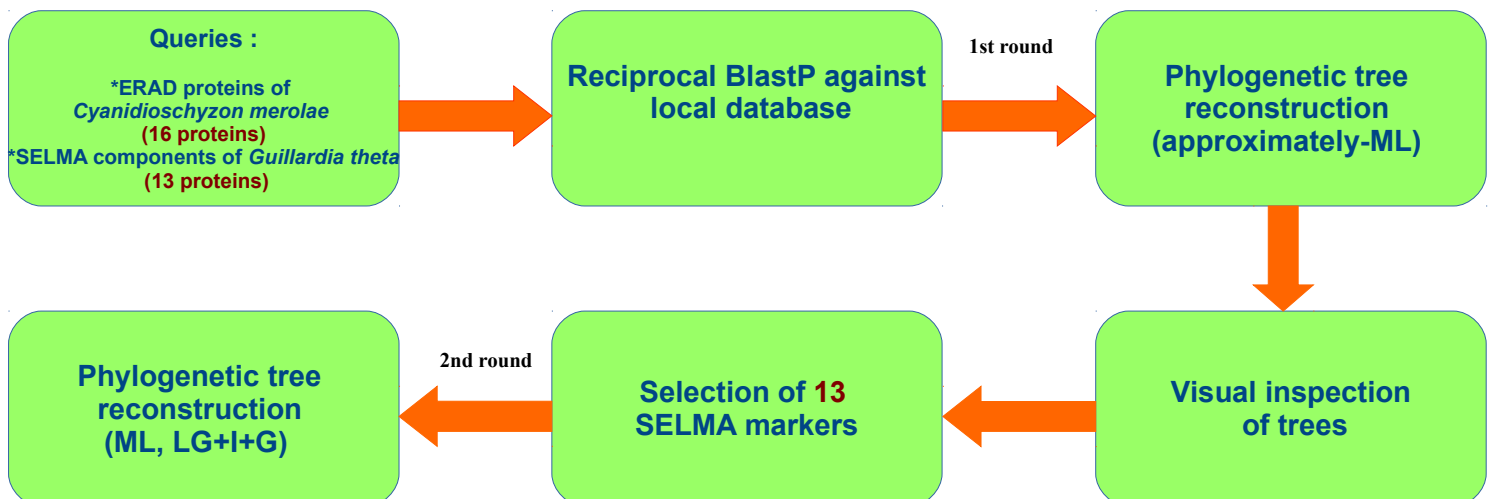


Figure 16. Diagram of the methodology used to study the origin of the SELMA system (Chapter 6)

4. ORIGIN AND EVOLUTION OF PRIMARY PLASTIDS

“It is a rather startling proposal that bacteria, the organisms which are popularly associated with disease, may represent the fundamental causative factor in the origin of species. Evidence of the constructive activities of bacteria has been at hand for many years, but popular conceptions of bacteria have been colored chiefly by their destructive activities as represented in disease. This destructive conception has become so fixed in the popular mind that the average person considers bacteria and disease as synonymous. Bacteria occupy a fundamental position in the world as it is constituted today. It is impossible to conceive of this world as existing without bacteria”

Ivan E. Wallin. *Symbiogenesis and the origin of species* (1927).

4. ORIGIN AND EVOLUTION OF PRIMARY PLASTIDS

4.1. Context and objectives

As described in the Introduction (section 1.2.1.6), the present-day cyanobacterial lineage closest to primary plastids remains unidentified. Despite that studies have supported either an early- or late-branching origin of primary plastids within the phylogeny of cyanobacteria, none of them have successfully identified their cyanobacterial sister lineage. This was probably due to an overlook or insufficient taxon sampling around key cyanobacterial clades.

Remarkably, our team and other groups have recently undertaken sequencing projects of cyanobacterial genomes that have increased the coverage of cyanobacterial species from a broad range of lifestyles, morphologies and phylogenetic affiliation. For instance, Shih et al., (2013) sequenced 54 cyanobacterial genomes giving as a result a twofold increase in the number of sequenced genomes within the phylum Cyanobacteria. The availability of new genomic data allowed us to readdress the longstanding question on the origin of primary plastids.

To study the origin of primary plastids, we established the following objectives:

1. To perform robust phylogenomic analyses using heterogeneous models of protein evolution on concatenated plastid markers.
2. To create a dataset of strictly selected nuclear-encoded genes of cyanobacterial origin (EGT) and perform phylogenomic analyses. As it is shown in section 1.2, plastid genomes are extremely reduced and encode only around 10% of genes found in free-living cyanobacteria. Therefore, EGTs are another source of information to study the origin of primary plastids. Although some studies have remarked the potential of EGTs and have identified some of them (Criscuolo & Gribaldo, 2011; Dagan et al., 2013), there has not been a systematic search and phylogenomic analysis of these genes.
3. To study whether the plastid ancestor was capable of nitrogen fixation, to test the hypothesis that the establishment of primary plastids was based on the nitrogen dependency of the host as has been proposed by some authors (Deusch et al., 2008; Falcón et al., 2010).

4.2 Results

Our phylogenomic analyses of concatenated plastid genes of Archaeplastida (see Figure 17) as well as the concatenated dataset of nuclear-encoded genes of cyanobacterial origin (EGTs), strongly support that the closest present-day relative of plastids is *Gloeomargarita lithophora*, an early-branching cyanobacterium that was recently sequenced by our team and was initially found in an alkaline crater lake in Mexico (Couradeau et al., 2012). This cyanobacterium belongs to a large cyanobacterial clade that has a cosmopolitan distribution as shown by environmental 16S rDNA sequencing studies (Ragon et al., 2014). However, it seems to be restricted to terrestrial habitats. Our results show for the first time a living cyanobacterial clade that is sister to primary plastids in phylogenomic analyses.

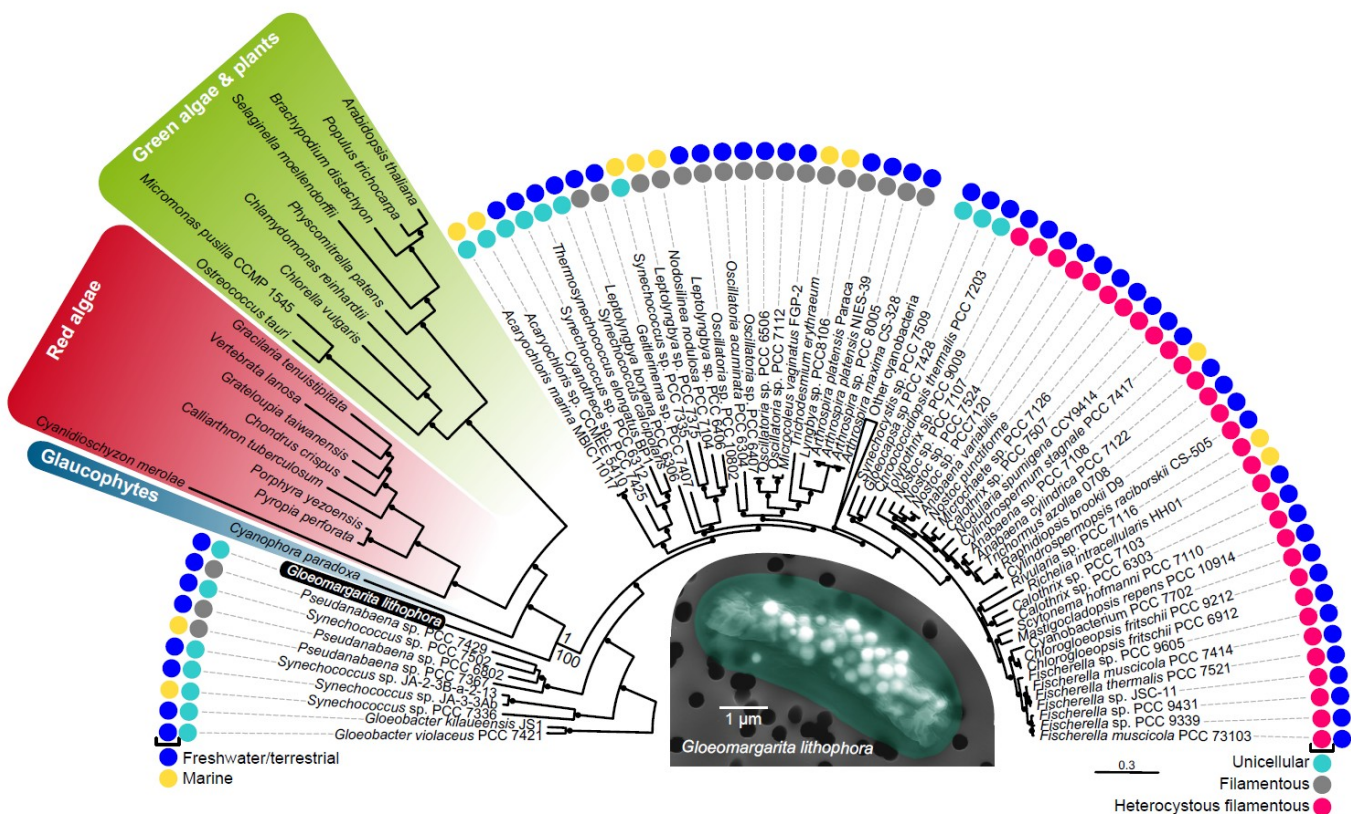


Figure 17. Bayesian phylogenetic tree based on the concatenation of 97 plastid-encoded proteins and their cyanobacterial homologs. Information about the habitat and morphology of the cyanobacterial species is provided. Branches supported by posterior probability 1 are labeled with black circles.

4.3 Manuscript of article 1

An Early-Branching Freshwater Cyanobacterium at the Origin of Plastids

(*Current Biology* 27:386–391)

Rafael I. Ponce-Toledo,¹ Philippe Deschamps,¹ Purificación López-García,¹ Yvan Zivanovic,² Karim Benzerara,³ and David Moreira^{1,4,*}

¹Unité d'Ecologie, Systématique et Evolution, CNRS UMR 8079, Université Paris-Sud/Paris-Saclay, AgroParisTech, 91400 Orsay, France.

²Institut de Génétique et Microbiologie, CNRS UMR 8621, Université Paris-Sud/Paris-Saclay, 91405 Orsay, France.

³Institut de Minéralogie, de Physique des Matériaux, et de Cosmochimie, Sorbonne Universités, UPMC Université Paris 06, CNRS UMR 7590, Muséum National d'Histoire Naturelle, IRD UMR 206, 75005 Paris, France.

⁴Lead Contact

*Correspondence: david.moreira@u-psud.fr

An Early-Branching Freshwater Cyanobacterium at the Origin of Plastids

Rafael I. Ponce-Toledo,¹ Philippe Deschamps,¹ Purificación López-García,¹ Yvan Zivanovic,² Karim Benzerara,³ and David Moreira^{1,4,*}

¹Unité d'Ecologie, Systématique et Evolution, CNRS UMR 8079, Université Paris-Sud/Paris-Saclay, AgroParisTech, 91400 Orsay, France

²Institut de Génétique et Microbiologie, CNRS UMR 8621, Université Paris-Sud/Paris-Saclay, 91405 Orsay, France

³Institut de Minéralogie, de Physique des Matériaux, et de Cosmochimie, Sorbonne Universités, UPMC Université Paris 06, CNRS UMR 7590, Muséum National d'Histoire Naturelle, IRD UMR 206, 75005 Paris, France

⁴Lead Contact

*Correspondence: david.moreira@u-psud.fr

<http://dx.doi.org/10.1016/j.cub.2016.11.056>

SUMMARY

Photosynthesis evolved in eukaryotes by the endosymbiosis of a cyanobacterium, the future plastid, within a heterotrophic host. This primary endosymbiosis occurred in the ancestor of Archaeplastida, a eukaryotic supergroup that includes glaucophytes, red algae, green algae, and land plants [1–4]. However, although the endosymbiotic origin of plastids from a single cyanobacterial ancestor is firmly established, the nature of that ancestor remains controversial: plastids have been proposed to derive from either early- or late-branching cyanobacterial lineages [5–11]. To solve this issue, we carried out phylogenomic and super-network analyses of the most comprehensive dataset analyzed so far including plastid-encoded proteins and nucleus-encoded proteins of plastid origin resulting from endosymbiotic gene transfer (EGT) of primary photosynthetic eukaryotes, as well as wide-ranging genome data from cyanobacteria, including novel lineages. Our analyses strongly support that plastids evolved from deep-branching cyanobacteria and that the present-day closest cultured relative of primary plastids is *Gloeomargarita lithophora*. This species belongs to a recently discovered cyanobacterial lineage widespread in freshwater microbialites and microbial mats [12, 13]. The ecological distribution of this lineage sheds new light on the environmental conditions where the emergence of photosynthetic eukaryotes occurred, most likely in a terrestrial-freshwater setting. The fact that glaucophytes, the first archaeplastid lineage to diverge, are exclusively found in freshwater ecosystems reinforces this hypothesis. Therefore, not only did plastids emerge early within cyanobacteria, but the first photosynthetic eukaryotes most likely evolved in terrestrial-freshwater settings, not in oceans as commonly thought.

RESULTS AND DISCUSSION

Phylogenomic Evidence for the Early Origin of Plastids among Cyanobacteria

To address the question of whether plastids derive from an early- or late-branching cyanobacterium, we have carried out phylogenomic analyses upon a comprehensive dataset of conserved plastid-encoded proteins and the richest sampling of cyanobacterial genome sequences used to date. Plastids have highly reduced genomes (they encode only between 80 and 230 proteins) compared with free-living cyanobacterial genomes, which encode between 1,800 and 12,000 proteins [10]. Most genes remaining in these organelle's genomes encode essential plastid functions (e.g., protein translation and photosystem structure), and their sequences are highly conserved. Therefore, despite the fact that plastid-encoded proteins are relatively few, they are good phylogenetic markers because they (1) are direct remnants of the cyanobacterial endosymbiont and (2) exhibit remarkable sequence and functional conservation.

To have a broad and balanced representation of primary plastids and cyanobacterial groups in our analyses, we mined a large sequence database containing all ribosomal RNAs (rRNAs) and proteins encoded in 19 plastids of Archaeplastida (one glaucophyte, eight green algae and plants, and nine red algae) and 122 cyanobacterial genomes, including recently sequenced members [7, 10, 14] and the genome of *Gloeomargarita lithophora*, sequenced for this work (see the [Experimental Procedures](#)). Sequence similarity searches in this genome dataset allowed the identification of 97 widespread conserved proteins, after exclusion of those only present in a restricted number of plastids and/or cyanobacteria and those showing evidence of horizontal gene transfer (HGT) among cyanobacterial species (see the [Supplemental Experimental Procedures](#) and [Data S1](#) for more information). Preliminary phylogenetic analysis of the dataset of 97 conserved proteins by a supermatrix approach (21,942 concatenated amino acid sites) using probabilistic phylogenetic methods with a site-heterogeneous substitution model (CAT-GTR) retrieved a tree with a deep-branching position of the plastid sequences ([Figure S1A](#)). Also noticeable in this tree was the very long branch of the *Synechococcus-Prochlorococcus* (SynPro) cyanobacterial clade, which reflected its very high evolutionary rate. Since

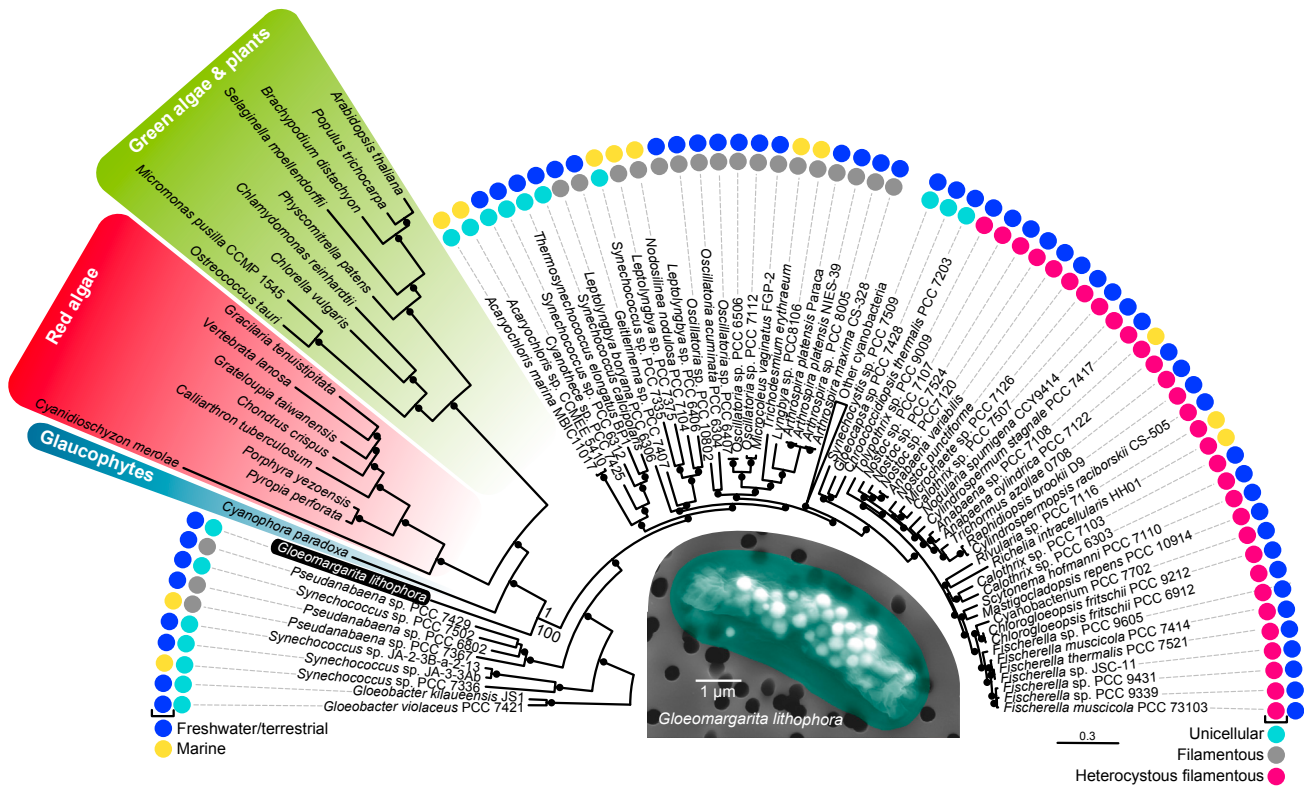


Figure 1. The Position of Plastids in the Cyanobacterial Phylogeny

This Bayesian phylogenetic tree is based on the concatenation of 97 plastid-encoded proteins and their cyanobacterial homologs. Branches supported by posterior probability 1 are labeled with black circles. ML bootstrap value is also indicated for the branch uniting plastids with the cyanobacterium *Gloeomargarita lithophora*. A false-colored scanning electron microscopy image of this cyanobacterium is shown in the center of the tree. Information about the habitat and morphology of the cyanobacterial species is provided. For the complete tree, see [Figure S1](#).

the SynPro cyanobacteria have never been found to be related to plastids, as confirmed by our own analysis, we decided to exclude them from subsequent analyses because of their accelerated evolutionary rate and because their sequences have a strong compositional bias known to induce errors in phylogenomic studies [8].

After removing the fast-evolving SynPro clade, we analyzed the dataset of 97 proteins as well as an rRNA dataset containing plastid and cyanobacterial 16S+23S rRNA concatenated sequences using maximum likelihood (ML) and Bayesian phylogenetic inference. Phylogenetic trees reconstructed for both datasets provide full support for the early divergence of plastids among the cyanobacterial species ([Figures 1](#) and [S1B](#), respectively). Moreover, our trees show that the closest present-day relative of plastids is the recently described deep-branching cyanobacterium *Gloeomargarita lithophora* [12, 13]. This biofilm-forming benthic cyanobacterium has attracted attention by its unusual capacity to accumulate intracellular amorphous calcium-magnesium-strontium-barium carbonates. Phylogenetic analysis based on environmental 16S rRNA sequences has shown that *G. lithophora* belongs to a diverse early-branching cyanobacterial lineage, the *Gloeomargaritales* [13], for which it is the only species isolated so far. Although it was initially found in an alkaline crater lake in Mexico, environmental studies have revealed that *G. lithophora* and

related species have a widespread terrestrial distribution ranging from freshwater alkaline lakes to thermophilic microbial mats [13]. Interestingly, it has never been observed in marine samples [13].

During the long endosymbiotic history of plastids, many genes necessary for plastid function were transferred from the cyanobacterial endosymbiont into the host nuclear genome (endosymbiotic gene transfer, EGT [15]). In a previous work, we carried out an exhaustive search of EGT-derived proteins in Archaeplastida genomes with two strict criteria: (1) the proteins had to be present in more than one group of Archaeplastida and (2) they had to be widespread in cyanobacteria and with no evidence of HGT among cyanobacterial lineages [16]. We have updated our EGT dataset with the new genome sequences available and retained 72 highly conserved proteins (see [Data S1](#)) that we analyzed applying similar approaches to those used for the plastid-encoded proteins. ML and Bayesian inference phylogenetic analyses of the resulting 28,102-amino-acid-long concatenation of these EGT proteins yield similar phylogenetic trees to those based on the plastid-encoded proteins: *G. lithophora* always branches as sister group of the archaeplastid sequences with maximal statistical support ([Figure S1C](#)). Therefore, three different datasets (plastid encoded proteins, plastid rRNA genes, and EGT proteins encoded in the nucleus) converge to support the same result.

Robustness of the Evolutionary Relationship between *Gloeomargarita* and Plastids

To test the robustness of the phylogenetic relation between *Gloeomargarita* and plastids, we investigated possible biases that might lead to an artifactual placement of plastids in our trees. We specifically focused on the dataset of plastid-encoded proteins, as these proteins are better conserved than the EGT proteins. Changes induced by the adaptation of endosymbiotically transferred genes to the new genetic environment (the eukaryotic nuclear genome) most likely impacted the evolutionary rate of EGT proteins (notice the much longer distance between *Gloeomargarita* and archaeplastid sequences for the EGT protein dataset than for the plastid-encoded protein dataset, 20% longer on average; see [Figures S1C and S1E](#)). We first recoded the amino acid sequences by grouping amino acids of similar physicochemical characteristics into four families, a procedure that is known to alleviate possible compositional biases [17]. Phylogenetic trees based on the recoded alignment still retrieve the *Gloeomargarita*-plastids sister relation ([Figure S1D](#)). It has been shown that fast-evolving sites in plastids have a very poor fit to evolutionary models [18]. Therefore, we tested whether the *Gloeomargarita*-plastids relation could be due to the accumulation of fast-evolving sites leading to sequence evolution model violation and long-branch attraction (LBA) artifacts. For that, we calculated the evolutionary rate for each of the 21,942 sites of the 97-protein concatenation and divided them into ten categories, from the slowest- to the fastest-evolving ones. We then reconstructed phylogenetic trees with a progressive inclusion of fast-evolving sites (the so-called slow-fast method, aimed at increasing the signal/noise ratio of sequence datasets [19]). All of the trees show the *Gloeomargarita*-plastids sister relation with full statistical support ([Figures S1E–S1M](#)). Interestingly, the trees based on the slowest-evolving positions exhibit a remarkable reduction of the branch length of the plastid sequences (especially those of green algae and land plants; [Figures S1K–S1M](#)), which become increasingly longer with the addition of fast-evolving positions. This reflects the well-known acceleration of evolutionary rate that plastids have experienced, in particular in Viridiplantae [16]. These results argue against the possibility that the *Gloeomargarita*-plastids relation arises from an LBA artifact due to the accumulation of noise in fast-evolving sites. Finally, we tested whether the supermatrix approach might have generated artifactual results due to the concatenation of markers with potentially incompatible evolutionary histories (because of HGT, hidden paralogy, etc.) that might have escaped our attention. We addressed this issue through the application of a phylogenetic network approach, which can cope with those contradictory histories [20], to analyze the set of 97 phylogenetic trees reconstructed with the individual proteins. The supernetwork based on those trees again confirms *G. lithophora* as the closest cyanobacterial relative of plastids ([Figure 2](#)), in agreement with the phylogenomic results ([Figure 1](#)).

A deep origin of plastids within the cyanobacterial phylogeny was inferred in past studies based on 16S rRNA gene sequences, plastid proteins, and nucleus-encoded proteins of plastid origin [5–8]. However, those studies did not retrieve any close relationship between plastids and any extant cyanobacterial lineage, a result that could be attributed either to lack

of phylogenetic resolution or to incomplete taxonomic sampling [6]. Our results, after inclusion of the new species *G. lithophora*, clearly advocate for the latter and stress the importance of exploration of undersampled environments such as many freshwater and terrestrial ones. Nevertheless, some studies have alternatively proposed that plastids emerged from the apical part of the cyanobacterial tree, being closely related to late-branching filamentous (Nostocales and Stigonomatales) [9–11] or unicellular (Chroococcales) [21] N_2 -fixing cyanobacteria. However, in agreement with our results, it has been shown that those phylogenetic results can be explained by similarities in G+C content due to convergent nucleotide composition between plastids and late-branching cyanobacteria and that the use of codon recoding techniques suppresses the compositional bias and recovers a deep origin of plastids [8].

Studies supporting a late plastid origin did not rely on phylogenomic analyses alone, but also used other methods, such as quantifying the number and sequence similarity of proteins shared between Archaeplastida and different cyanobacterial species. They showed that the heterocyst-forming filamentous cyanobacterial genera *Nostoc* and *Anabaena* appear to possess the largest and most similar set of proteins possibly present in the plastid ancestor [9, 10]. However, these approaches have two important limitations: (1) the number of proteins is highly dependent on the genome size of the different cyanobacteria, which can vary by more than one order of magnitude [7], and (2) it is well known that sequence similarity is a poor proxy for evolutionary relationships [22]. Indeed, if we apply a similar procedure on our set of sequences but include several plastid representatives in addition to cyanobacteria, we observe that individual plastids can have sequences apparently more similar to certain cyanobacterial homologs than to those of other plastids ([Figure S2](#)). If sequence similarity were a good indicator of evolutionary relationship, it would be difficult to deduce from those data that plastids are monophyletic, as they do not always resemble other plastids more than cyanobacteria. Actually, plastid proteins, especially those that have been transferred to the nucleus [16], have evolved much faster than their cyanobacterial counterparts, such that sequences of one particular primary photosynthetic eukaryote can be more similar to short-branching cyanobacterial ones than to those of other distantly related long-branching photosynthetic eukaryotes. Thus, similarity-based approaches have clear limitations and can lead to artifactual results. Phylogenomic analysis is therefore more suitable than crude sequence similarity or single-gene phylogenies for studying the cyanobacterial origin of plastids.

Metabolic Interactions and the Environmental Setting at the Origin of Plastids

Proponents of the recent origin of primary plastids from N_2 -fixing filamentous cyanobacteria suggest that the dearth of biologically available nitrogen during most of the Proterozoic and the ability of these organisms to fix nitrogen played a key role in the early establishment of the endosymbiosis [9, 10]. However, there is no trace of such past ability to fix nitrogen in modern plastids. Furthermore, this metabolic ability is also widespread in many cyanobacterial lineages, including basal-branching clades, which has led to propose that the cyanobacterial ancestor was able to fix nitrogen [23]. Therefore, if N_2 fixation

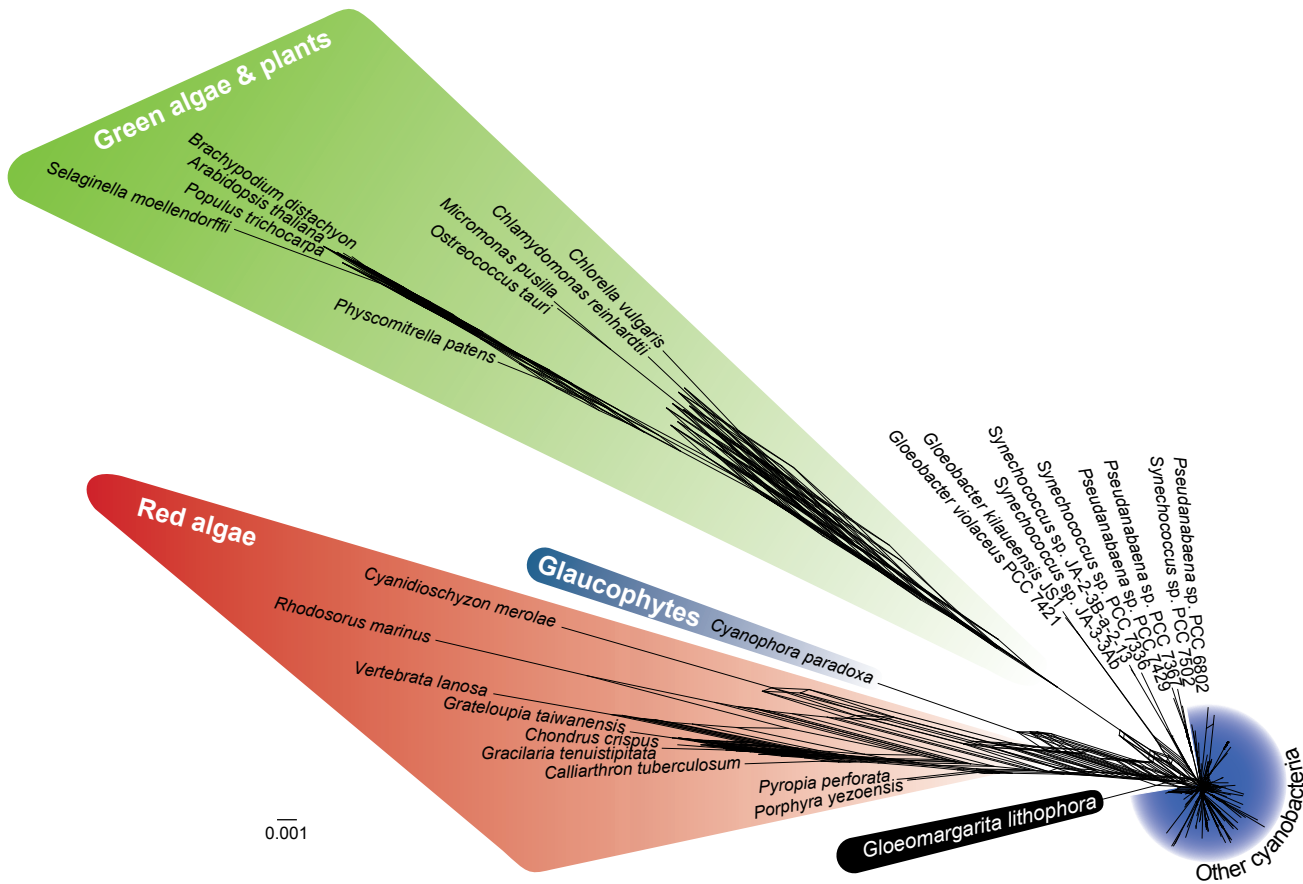


Figure 2. Supernetwork Analysis of Plastid-Encoded Proteins and Cyanobacterial Homologs

This phylogenetic supernetwork is based on the individual ML trees of 97 plastid-encoded proteins. For a distance-based analysis of these individual proteins, see [Figure S2](#).

did actually play a role in the establishment of the plastid, it cannot discriminate in favor of a late- versus early-branching cyanobacterial endosymbiont. The closest modern relative of plastids, *G. lithophora*, lacks the genes necessary for N_2 fixation, which further argues against the implication of this metabolism in the origin of plastids. N_2 fixation is also absent in the cyanobacterial endosymbiont of *Paulinella chromatophora*, which is considered to be a recent second, independent case of primary plastid acquisition in eukaryotes [3, 4, 24]. Other hypotheses have proposed a symbiotic interaction between the cyanobacterial ancestor of plastids and its eukaryotic host based on the metabolism of storage polysaccharides. In that scenario, the cyanobacterium would have exported ADP-glucose in exchange for the import of reduced nitrogen from the host [25].

Although the nature of the metabolic exchanges between the two partners at the origin of primary photosynthetic eukaryotes remains to be elucidated, the exclusive distribution of the *Gloeomargarita* lineage in freshwater and terrestrial habitats [13] provides important clues about the type of ecosystem where the endosymbiosis of a *Gloeomargarita*-like cyanobacterium within a heterotrophic host took place. Like *G. lithophora*, and similar to most basal-branching cyanobacterial lineages, the first cyanobacteria most likely thrived in terrestrial or freshwater hab-

itats [26, 27]. Consistent with this observation, the colonization of open oceans and the diversification of marine planktonic cyanobacteria, including the SynPro clade, occurred later on in evolutionary history, mainly during the Neoproterozoic (1000–541 mya), with consequential effects on ocean and atmosphere oxygenation [27]. Notwithstanding their large error intervals, molecular-clock analyses infer the origin of Archaeplastida during the mid-Proterozoic [28, 29], well before the estimated Neoproterozoic cyanobacterial colonization of oceans. Interestingly, Glaucophyta, the first lineage to diverge within the Archaeplastida (Figures 1 and S1), has been exclusively found in freshwater habitats [30], which is also the case for the most basal clade of red algae, the Cyanidiales, commonly associated to terrestrial thermophilic mats [31]. In contrast with the classical idea of a marine origin of eukaryotic alga, these data strongly support that plastids, and hence the first photosynthetic eukaryotes, arose in a freshwater or terrestrial environment [32]. This most likely happened on a Proterozoic Earth endowed with low atmospheric and oceanic oxygen concentrations [33].

Conclusions

We provide strong phylogenomic evidence from plastid- and nucleus-encoded genes of cyanobacterial ancestry for a deep

origin of plastids within the phylogenetic tree of Cyanobacteria and find that the *Gloeomargarita* lineage represents the closest extant relative of the plastid ancestor. The ecological distribution of both this cyanobacterial lineage and extant early-branching eukaryotic algae suggests that the first photosynthetic eukaryotes evolved in a terrestrial environment, probably in freshwater biofilms or microbial mats. Microbial mats are complex communities in which metabolic symbioses between different microbial types, including cyanobacteria and heterotrophic eukaryotes, are common. They imply physical and genetic interactions between mutualistic partners that may have well facilitated the plastid endosymbiosis. In that sense, it will be especially interesting to study the interactions that the cyanobacterial species of the *Gloeomargarita* lineage establish with other microorganisms, looking for potential symbioses with heterotrophic protists. Our work highlights the importance of environmental exploration to characterize new organisms that can, in turn, be crucial to resolve unsettled evolutionary questions.

EXPERIMENTAL PROCEDURES

Selection of Phylogenetic Markers

We created a local genome database with 692 completely sequenced prokaryotic and eukaryotic genomes, complemented with expressed sequenced tags (ESTs) from eight photosynthetic eukaryotes, all downloaded from GenBank. In particular, the database included 122 cyanobacteria and 20 plastids of primary photosynthetic eukaryotes (primary plastids, found in Archaeplastida). Among the cyanobacteria, we added the genome sequence of *Gloeomargarita lithophora* strain CCAP 1437/1 (GenBank: NZ_CP017675; see the [Supplemental Experimental Procedures](#)). To retrieve plastid-encoded proteins and their cyanobacterial orthologs, we used the 149 proteins encoded in the gene-rich plastid of the glaucophyte *Cyanophora paradoxa* as queries in BLASTP [34] searches against this local database. Plastid and cyanobacterial protein sequences recovered among the top 300 hits with a cut-off E value $<10^{-10}$ were retrieved and aligned to reconstruct preliminary ML single-protein phylogenetic trees (details can be found in the [Supplemental Experimental Procedures](#)). These trees were visually inspected to identify potentially paralogous proteins or HGT among cyanobacteria (see the [Supplemental Experimental Procedures](#)). The problematic proteins detected were discarded. Among the remaining proteins, we selected those present in at least two of the three Archaeplastida lineages (glaucophytes, red algae, and green algae and plants; if absent in one lineage, they were replaced by missing data). Thereby, our final dataset comprised 97 plastid-encoded proteins (see [Data S1](#)). In addition, we updated the dataset of EGT proteins previously published by Deschamps and Moreira [16], including sequences from new available genomes. As in the previous case, ML single-protein phylogenetic trees were reconstructed and inspected to discard proteins with paralogs and HGT cases. All datasets are available at <http://www.ese.u-psud.fr/article909.html?lang=en>.

Phylogenetic Analyses

ML trees of individual markers were reconstructed using PhyML [35] with the GTR model for rRNA gene sequences and the LG model for protein sequences, in both cases with a four-substitution-rate-category gamma distribution and estimation of invariable sites from the alignments. Branch lengths, tree topology, and substitution rate parameters were optimized by PhyML. Bootstrap values were calculated with 100 pseudoreplicates. Individual ML trees were used to reconstruct a supernetwork [20] with the program Splitstree [36] with default parameters.

Tree reconstruction of concatenated protein datasets under Bayesian inference was carried out using two independent chains in PhyloBayes-MPI [37] with the CAT-GTR model and a four-category gamma distribution. Sequences were also recoded into four amino acid families [17] using a Python script and analyzed with PhyloBayes-MPI in the same way. Chain convergence was monitored by the maxdiff parameter (chains were run until maxdiff < 0.1).

The slow-fast method was applied using the program SlowFaster [38] to produce ten alignments with an increasing proportion of fast-evolving sites included. These alignments were used for phylogenetic reconstruction using PhyloBayes-MPI as with the complete concatenation.

Estimation of Protein Sequence Distances and Heatmap Construction

Distances among protein sequences were calculated using two different approaches: (1) measuring patristic distances between taxa in phylogenetic trees and (2) from pairwise comparisons of aligned protein sequences. Patristic distances are the sum of the lengths of the branches that connect two taxa in a phylogenetic tree and can account for multiple substitutions events to a certain level. We used the glaucophyte *C. paradoxa* as reference, as it has the shortest branch among the plastid sequences, and measured the patristic distances between this taxon and the other taxa (cyanobacterial species and plastids of red algae and Viridiplantae) on the ML trees of the 97 individual plastid-encoded proteins reconstructed as previously described. ProtDist [39] was used to estimate pairwise distances between the plastid proteins of *C. paradoxa* and the proteins of the rest of taxa. Patristic distances and sequence similarities were displayed as heatmaps using the library Matplotlib [40].

ACCESSION NUMBERS

The accession number for the genome sequence of the *Gloeomargarita lithophora* strain CCAP 1437/1 reported in this paper is GenBank: NZ_CP017675.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, two figures, and one dataset and can be found with this article online at <http://dx.doi.org/10.1016/j.cub.2016.11.056>.

AUTHOR CONTRIBUTIONS

D.M., P.L.-G., and K.B. characterized *Gloeomargarita lithophora*. K.B. performed electron microscopy. Y.Z. carried out genome annotation. R.I.P.-T., P.D., and D.M. performed phylogenetic analyses. R.I.P.-T., D.M., P.D., and P.L.-G. wrote the manuscript with input from all co-authors.

ACKNOWLEDGMENTS

This study was supported by European Research Council grants ProtistWorld (P.L.-G.; grant agreement no. 322669) and CALCYAN (K.B.; grant agreement no. 307110) under the European Union's Seventh Framework Program and the RTP CNRS Génomique Environnementale project MetaStrom (D.M.).

Received: September 23, 2016

Revised: November 4, 2016

Accepted: November 29, 2016

Published: January 26, 2017

REFERENCES

- Moreira, D., Le Guyader, H., and Philippe, H. (2000). The origin of red algae and the evolution of chloroplasts. *Nature* 405, 69–72.
- Rodríguez-Ezpeleta, N., Brinkmann, H., Burey, S.C., Roure, B., Burger, G., Löffelhardt, W., Bohnert, H.J., Philippe, H., and Lang, B.F. (2005). Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr. Biol.* 15, 1325–1330.
- Archibald, J.M. (2009). The puzzle of plastid evolution. *Curr. Biol.* 19, R81–R88.
- Keeling, P.J. (2013). The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annu. Rev. Plant Biol.* 64, 583–607.
- Turner, S., Pryer, K.M., Miao, V.P., and Palmer, J.D. (1999). Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J. Eukaryot. Microbiol.* 46, 327–338.

6. Criscuolo, A., and Gribaldo, S. (2011). Large-scale phylogenomic analyses indicate a deep origin of primary plastids within cyanobacteria. *Mol. Biol. Evol.* **28**, 3019–3032.
7. Shih, P.M., Wu, D., Latifi, A., Axen, S.D., Fewer, D.P., Talla, E., Calteau, A., Cai, F., Tandeau de Marsac, N., Rippka, R., et al. (2013). Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc. Natl. Acad. Sci. USA* **110**, 1053–1058.
8. Li, B., Lopes, J.S., Foster, P.G., Embley, T.M., and Cox, C.J. (2014). Compositional biases among synonymous substitutions cause conflict between gene and protein trees for plastid origins. *Mol. Biol. Evol.* **31**, 1697–1709.
9. Deusch, O., Landan, G., Roettger, M., Gruenheit, N., Kowallik, K.V., Allen, J.F., Martin, W., and Dagan, T. (2008). Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. *Mol. Biol. Evol.* **25**, 748–761.
10. Dagan, T., Roettger, M., Stucken, K., Landan, G., Koch, R., Major, P., Gould, S.B., Goremykin, V.V., Rippka, R., Tandeau de Marsac, N., et al. (2013). Genomes of Stigonematalean cyanobacteria (subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids. *Genome Biol. Evol.* **5**, 31–44.
11. Ochoa de Alda, J.A., Esteban, R., Diago, M.L., and Houmard, J. (2014). The plastid ancestor originated among one of the major cyanobacterial lineages. *Nat. Commun.* **5**, 4937.
12. Couradeau, E., Benzerara, K., Gérard, E., Moreira, D., Bernard, S., Brown, G.E., Jr., and López-García, P. (2012). An early-branching microbialite cyanobacterium forms intracellular carbonates. *Science* **336**, 459–462.
13. Ragon, M., Benzerara, K., Moreira, D., Tavera, R., and López-García, P. (2014). 16S rDNA-based analysis reveals cosmopolitan occurrence but limited diversity of two cyanobacterial lineages with contrasted patterns of intracellular carbonate mineralization. *Front. Microbiol.* **5**, 331.
14. Benzerara, K., Skouri-Panet, F., Li, J., Féraud, C., Gugger, M., Laurent, T., Couradeau, E., Ragon, M., Cosmidis, J., Menguy, N., et al. (2014). Intracellular Ca-carbonate biomineralization is widespread in cyanobacteria. *Proc. Natl. Acad. Sci. USA* **111**, 10933–10938.
15. Martin, W., Stoebe, B., Goremykin, V., Hapsmann, S., Hasegawa, M., and Kowallik, K.V. (1998). Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* **393**, 162–165.
16. Deschamps, P., and Moreira, D. (2009). Signal conflicts in the phylogeny of the primary photosynthetic eukaryotes. *Mol. Biol. Evol.* **26**, 2745–2753.
17. Rodríguez-Ezpeleta, N., Brinkmann, H., Roure, B., Lartillot, N., Lang, B.F., and Philippe, H. (2007). Detecting and overcoming systematic errors in genome-scale phylogenies. *Syst. Biol.* **56**, 389–399.
18. Sun, L., Fang, L., Zhang, Z., Chang, X., Penny, D., and Zhong, B. (2016). Chloroplast phylogenomic inference of green algae relationships. *Sci. Rep.* **6**, 20528.
19. Philippe, H., Lopez, P., Brinkmann, H., Budin, K., Germot, A., Laurent, J., Moreira, D., Müller, M., and Le Guyader, H. (2000). Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. *Proc. Biol. Sci.* **267**, 1213–1221.
20. Huson, D.H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267.
21. Falcón, L.I., Magallón, S., and Castillo, A. (2010). Dating the cyanobacterial ancestor of the chloroplast. *ISME J.* **4**, 777–783.
22. Koski, L.B., and Golding, G.B. (2001). The closest BLAST hit is often not the nearest neighbor. *J. Mol. Evol.* **52**, 540–542.
23. Latysheva, N., Junker, V.L., Palmer, W.J., Codd, G.A., and Barker, D. (2012). The evolution of nitrogen fixation in cyanobacteria. *Bioinformatics* **28**, 603–606.
24. Nowack, E.C., Melkonian, M., and Glöckner, G. (2008). Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by eukaryotes. *Curr. Biol.* **18**, 410–418.
25. Deschamps, P., Colleoni, C., Nakamura, Y., Suzuki, E., Putaux, J.L., Buléon, A., Haebel, S., Ritte, G., Steup, M., Falcón, L.I., et al. (2008). Metabolic symbiosis and the birth of the plant kingdom. *Mol. Biol. Evol.* **25**, 536–548.
26. Battistuzzi, F.U., and Hedges, S.B. (2009). A major clade of prokaryotes with ancient adaptations to life on land. *Mol. Biol. Evol.* **26**, 335–343.
27. Sánchez-Baracaldo, P. (2015). Origin of marine planktonic cyanobacteria. *Sci. Rep.* **5**, 17418.
28. Parfrey, L.W., Lahr, D.J., Knoll, A.H., and Katz, L.A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc. Natl. Acad. Sci. USA* **108**, 13624–13629.
29. Eme, L., Sharpe, S.C., Brown, M.W., and Roger, A.J. (2014). On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. *Cold Spring Harb. Perspect. Biol.* **6**, 8.
30. Kies, L., and Kremer, B.P. (1990). Phylum Glaucocystophyta. In *Handbook of Protoctista*, L. Margulis, J.O. Corliss, M. Melkonian, and D.J. Chapman, eds. (Jones and Barlett), pp. 152–166.
31. Yoon, H.S., Müller, K.M., Sheath, R.G., Ott, F.D., and Bhattacharya, D. (2006). Defining the major lineages of red algae (Rhodophyta). *J. Phycol.* **42**, 482–492.
32. Blank, C.E. (2013). Origin and early evolution of photosynthetic eukaryotes in freshwater environments: reinterpreting proterozoic paleobiology and biogeochemical processes in light of trait evolution. *J. Phycol.* **49**, 1040–1055.
33. Lyons, T.W., Reinhard, C.T., and Planavsky, N.J. (2014). The rise of oxygen in Earth's early ocean and atmosphere. *Nature* **506**, 307–315.
34. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
35. Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704.
36. Huson, D.H. (1998). SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14**, 68–73.
37. Lartillot, N., Rodrigue, N., Stubbs, D., and Richer, J. (2013). PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* **62**, 611–615.
38. Kostka, M., Ualikova, M., Cepicka, I., and Flegr, J. (2008). SlowFaster, a user-friendly program for slow-fast analysis and its application on phylogeny of *Blastocystis*. *BMC Bioinformatics* **9**, 341.
39. Felsenstein, J. (1999). PHYLIP - phylogeny inference package, version 3.6 (University of Washington).
40. Hunter, J. (2007). Matplotlib: a 2d graphics environment. *Comput. Sci. Eng.* **9**, 90–95.

5. EVOLUTION OF SECONDARY PLASTID-HARBORING LINEAGES

“Grey, dear friend, is all theory, and green the golden tree of life”

Johann Wolfgang von Goethe. *Faust* (1804).

5. EVOLUTION OF SECONDARY PLASTID-HARBORING LINEAGES

5.1 Context and objectives

After we addressed the origin of primary plastids (previous chapter), we were interested in the study of secondary plastids. Chlorarachniophytes and euglenids both have secondary plastids that originated from two independent endosymbioses with green algae (Rogers et al., 2007). Nonetheless, multiple nuclear-encoded genes that participate in plastid-located biosynthetic pathways seem to have been transferred from red algal sources in both secondary green lineages (Markunas & Triemer, 2016; Yang et al., 2014). Similarly, the finding that chromalveolates seem to encode thousands of green algal-derived genes led Moustafa et al., (2009) to suggest a cryptic endosymbiosis with a green alga in diatoms. However, the reanalysis of these genes by Deschamps & Moreira, (2012) reduced the number of green genes to less than 3% of the initial claim. Hence, it seems that secondary algae have mosaic genomes and the extent of red- or green-algal derived genes remains unclear and can modify substantially the evolutionary hypotheses that are proposed to explain them.

Considering that EGTs were very helpful in the study of primary plastids (previous chapter), we used secondary EGTs to evaluate the contribution of red and green algal-derived genes in secondary plastid-bearing algae. We worked on a specific fraction of secondary EGTs: those that were transferred from cyanobacteria to Archaeplastida during primary endosymbiosis, and then from a red or green alga to secondary photosynthetic eukaryotes. These secondary EGTs are important for multiple reasons: 1. They are easy to spot in phylogenetic trees which allows the differentiation between genes that were likely acquired via EGT or HGT from genes that were vertically inherited; 2. The corresponding proteins are targeted to the plastid where they play important functions; 3. Secondary EGTs may help to identify cryptic endosymbiosis. Thus, to study the evolution of photosynthetic eukaryotes with secondary plastids we established the following objectives:

1. Identification of genes acquired from red and green algae in secondary plastid-harboring lineages that were compatible with a secondary EGT scenario. This was achieved by using robust phylogenetic analyses with a broad taxon sampling as well as strict criteria of selection of markers.

2. Functional annotation of red and green algal genes to identify whether there is a correlation between the gene function and its origin that may explain a bias in gene acquisition.

5.2 Results

Our results show that approximately half of the putative secondary EGTs of chlorarachniophytes and euglenids that we identified in our analyses, seem to have been transferred from red algal sources (See Figure 18). These results are surprising given that both photosynthetic lineages harbor secondary plastids derived from green algae. By contrast, the vast majority of secondary EGTs in eukaryotic lineages with secondary red plastids derive, as expected, from red algae. The quantification of secondary EGTs allowed us to propose different plausible explanations to account for the differences observed in the origin of this type of genes among secondary algae.

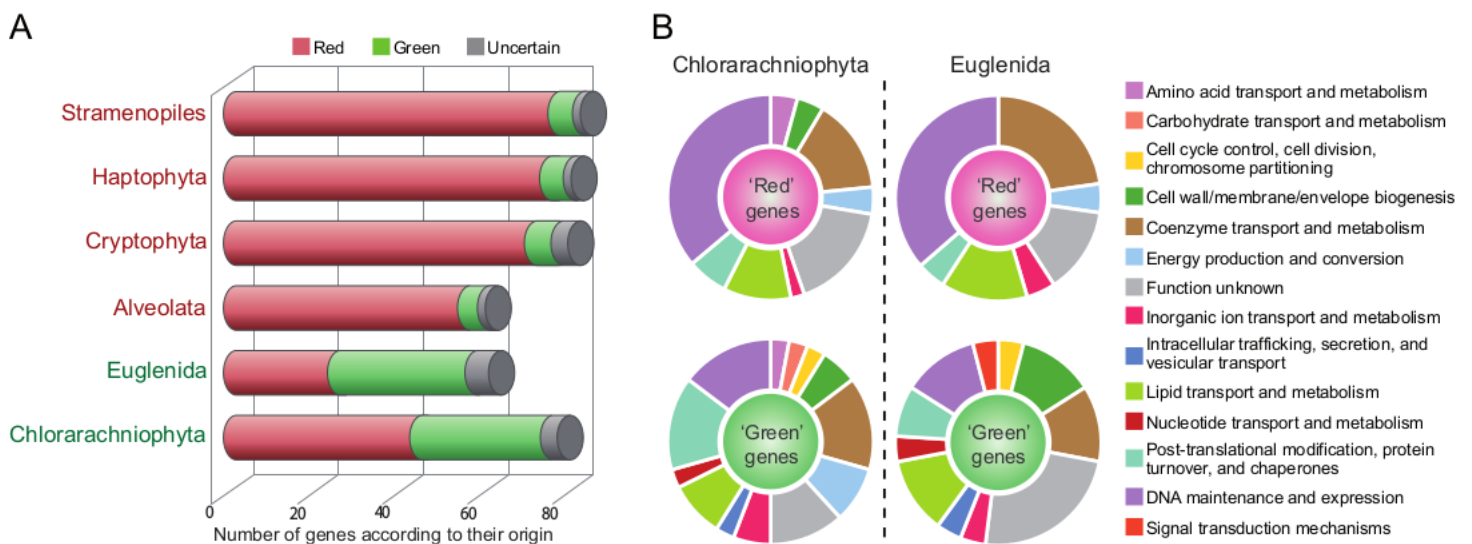


Figure 18. Genes transferred from red or green algae to secondary photosynthetic eukaryotes. A) Number of red or green algal-like genes for each lineage among the 85 genes analyzed, gray blocks correspond to genes for which the origin could not be unambiguously determined in the corresponding phylogenetic trees. B) Gene functions annotated using EggNOG 4.5 (Huerta-Cepas et al., 2016) for “red” and “green” genes detected in transcriptomes and nuclear genomes of euglenids and chlorarachniophytes.

5.3 Manuscript of article 2

Functioning of Secondary Green Plastids is Assured by a Mix of Genes of Red and Green Algal Ancestry

(Manuscript in preparation)

Rafael I. Ponce-Toledo¹, David Moreira¹, Purificación López-García¹ and Philippe Deschamps^{1,*}

¹Unité d'Ecologie Systématique et Evolution, CNRS, Université Paris-Sud, AgroParisTech, Université Paris-Saclay, 91400, Orsay, France

*Corresponding author: philippe.deschamps@u-psud.fr

Abstract

Endosymbiosis has been common all along eukaryotic evolution, providing opportunities for genomic and organellar innovation. Plastids are a prominent example. After the primary endosymbiosis of the cyanobacterial plastid ancestor, photosynthesis spread in many eukaryotic lineages via secondary endosymbioses involving red or green algal endosymbionts and diverse heterotrophic hosts. However, the number of secondary endosymbioses and how they occurred remain poorly understood. In particular, contrasting patterns of endosymbiotic gene transfer (EGT) have been detected and subjected to various interpretations. In this context, accurate detection of EGTs is essential to avoid wrong evolutionary conclusions. We have assembled a strictly selected set of markers that provides robust phylogenomic evidence suggesting that nuclear genes involved in the function and maintenance of green secondary plastids in chlorarachniophytes and euglenids have unexpected mixed red and green algal origins. This mixed ancestry contrasts with the clear red algal origin of most nuclear genes carrying similar functions in secondary algae with red plastids.

Key words: Chlorarachniophyta, Euglenida, endosymbiotic gene transfer, phylogenomics.

Photosynthesis in eukaryotes occurs in specialized organelles: the plastids. They first evolved in the Archaeplastida (Viridiplantae + Rhodophyta + Glaucophyta) through the endosymbiosis of a cyanobacterium inside a eukaryotic host (Archibald 2009; Keeling 2013). This primary endosymbiosis entailed massive endosymbiotic gene transfer (EGT) from the cyanobacterium to the host nucleus (Weeden 1981; Kleine, et al. 2009). Consequently, most proteins required for primary plastid functioning are encoded in the nuclear genome and addressed to the plastid via specialized signal sequences and a

translocation apparatus (Gutensohn, et al. 2006). Other eukaryotes obtained plastids through secondary endosymbiosis (symbiosis of green or red algae within other eukaryotic cells) or even through tertiary endosymbiosis (symbiosis of secondary photosynthetic eukaryotes within other eukaryotes) (Delwiche 1999; Archibald 2009; Keeling 2013). Euglenida (Excavata) and Chlorarachniophyta (Rhizaria) have green secondary plastids acquired through two independent endosymbioses involving Prasinophyceae and Ulvophyceae green algae, respectively (Rogers, et al. 2007; Hrdá, et al. 2012; Suzuki, et al. 2016). Photosynthetic species in the Cryptophyta, Alveolata, Stramenopiles and Haptophyta (CASH) lineages have plastids derived from red algae but their evolutionary history remains controversial (Lane and Archibald 2008; Archibald 2009; Keeling 2013). Whereas phylogenomic analyses of plastid-encoded genes support CASH monophyly, arguing for a single red algal secondary endosymbiosis (Yoon, et al. 2002; Muñoz-Gómez, et al. 2017), most host nuclear gene phylogenies do not support it (Baurain, et al. 2010; Burki, et al. 2016). Some authors tried to reconcile these incongruent results by proposing that a unique phylum (that become extinct or evolved into one of the extant CASH phyla) acquired a red alga through secondary endosymbiosis and, subsequently, transmitted it via serial tertiary endosymbioses with other CASH phyla (Larkum, et al. 2007; Sanchez-Puerta and Delwiche 2008; Bodył, et al. 2009; Baurain, et al. 2010; Petersen, et al. 2014).

As in the primary endosymbiosis, secondary or tertiary endosymbioses entail numerous EGTs from the endosymbiotic red or green algal nucleus to the host nucleus. Thus, secondary photosynthetic eukaryotes possess two types of genes that can inform about the phylogenetic identity of their plastids: plastid-encoded genes and nucleus-encoded genes acquired via EGT. Genes of primary plastid genomes and EGTs in Archaeplastida genomes are related to cyanobacteria and have helped to identify the cyanobacterial lineage at the origin of the first plastid (Ponce-Toledo, et al. 2017). Similarly, plastid-

encoded genes and EGTs in nuclear genomes of secondary photosynthetic eukaryotes are expected to reveal the red or green algal origin of their plastids. Compared to plastid-encoded genes, EGTs have the additional advantage of informing about possible past plastids secondarily lost or replaced (cryptic plastids). However, if EGTs are valuable to track both contemporary and cryptic endosymbioses, their detection within nuclear genome sequences remains complex (Stiller 2011). EGT detection is rather straightforward in Archaeplastida because cyanobacterial-like genes are easily distinguishable from typical eukaryotic genes. It is more difficult in secondary endosymbioses because the poor resolution of many single-gene phylogenies may hamper distinguishing EGTs from vertically inherited nuclear genes, especially considering the short phylogenetic distance between donors (red and green algae) and acceptors (secondary algae). Two recent studies reporting a high number of genes apparently related to green algal homologs in red-plastid-bearing algae, chromerids (Alveolata) and diatoms (Stramenopiles) are illustrative. Whereas in chromerids the green signal was attributed to phylogenetic artifacts and insufficient sampling of red algal genome sequences (Woehle, et al. 2011), in diatoms it was interpreted as evidence for a cryptic green algal endosymbiont (Moustafa, et al. 2009). However, reanalyses of the same genes using improved taxonomic sampling and phylogenetic methods largely erased the evidence of potential cryptic green endosymbiosis (Burki, et al. 2012; Deschamps and Moreira 2012; Moreira and Deschamps 2014).

The impact of horizontal gene transfer (HGT) on eukaryotic evolution remains controversial. HGTs may help to infer the history of genomes and lineages (Abby, et al. 2012) but they can also introduce phylogenetic noise, notably in the study of EGTs (Stiller 2011). Through time, secondary photosynthetic eukaryotes may have accumulated HGTs from various sources, including non-endosymbiotic red and green algae. Such HGTs may

display phylogenies comparable to those of real EGTs, making them difficult to set apart. Thus, anomalous phylogenetic signal in certain secondary photosynthetic groups has been interpreted as HGT rather than EGT from cryptic endosymbionts. For example, the analysis of the nuclear genome of the green-plastid-containing chlorarachniophyte *Bigelowiella natans* indicated that 22% of the genes potentially acquired via HGT had red algal origin (Curtis, et al. 2012). Because of the phagotrophic ability of chlorarachniophytes, these genes were considered as the result of progressive accumulation of HGTs from red algae or from red-plastid-containing CASH lineages (Archibald, et al. 2003; Yang, et al. 2011; Yang, et al. 2014). Likewise, studies on euglenids suggested a similar trend for several genes involved in central metabolic pathways (Maruyama, et al. 2011; Yang, et al. 2011; Markunas and Triemer 2016). If the unexpected presence of these 'red' genes in chlorarachniophytes and euglenids was first considered as evidence of multiple HGTs, the increasing number of reported cases prompted some authors to propose putative cryptic red algal endosymbioses (Maruyama, et al. 2011; Markunas and Triemer 2016). Nevertheless, a systematic investigation of HGT/EGT is still missing in euglenids and chlorarachniophytes but, as mentioned above, it can be difficult to distinguish in the context of secondary endosymbioses among HGT, EGT, and just unresolved trees on the basis of single-gene phylogenies (Deschamps and Moreira 2012).

Here, we have focused on a particular group of genes to reduce uncertainty: genes transferred from the original cyanobacterial plastid endosymbiont into the nuclear genome of Archaeplastida and, subsequently, from Archaeplastida into the genomes of complex secondary algae. In Archaeplastida, these genes are often involved in essential plastid functions and tend to be highly conserved (Reyes-Prieto, et al. 2006; Deschamps and Moreira 2009), we thus expected that they provide strong phylogenetic signal. To identify them, we queried by BLAST the whole predicted proteomes of *Guillardia theta* and

Bigelowiella natans against a local genome database containing representatives of the three domains of life, in particular a comprehensive collection of genomes and transcriptomes of photosynthetic protists (detailed methods are described in the Supplementary Material). *Guillardia* and *Bigelowiella* proteins with hits in other photosynthetic eukaryotes and in cyanobacteria were selected for maximum likelihood (ML) phylogenetic analysis. Trees for these proteins were manually filtered to retain those fulfilling two criteria: i) trees have to support the monophyly of Archaeplastida and a clear separation of Viridiplantae and Rhodophyta (with secondary lineages branching within them), and ii) proteins have to be shared by at least three secondary photosynthetic lineages. We identified in this way 85 genes most likely acquired by secondary photosynthetic eukaryotes from Archaeplastida. 73 were cyanobacterial genes likely transferred sequentially through primary and secondary endosymbioses, and 12 were genes likely transferred from diverse bacterial groups to Archaeplastida and subsequently to secondary photosynthetic groups (supplementary table S2 and figs. S1-S85, Supplementary Material online). Most of them were absent in non-photosynthetic eukaryotes, supporting that they were not misinterpreted vertically-inherited genes.

The 85 ML phylogenies were highly resolved and enabled to unambiguously determine, for each secondary lineage, whether the gene had green or red algal origin. As expected, in most trees (~90%, fig. 1A), the red-plastid-endowed CASH lineages had genes derived from red algae (e.g., fig. 2A and 2B). We expected the opposite situation, namely a majority of 'green' genes, in the green-plastid-endowed chlorarachniophytes and euglenids. However, 47 of the 81 trees containing chlorarachniophytes (58 %, fig. 1A) and 22 of the 51 trees containing euglenids (43 %, fig. 1A) supported a 'red' origin of the corresponding genes (e.g., fig 2A and B). These surprisingly high values were in sharp contrast with the mere 9 trees (10%, fig. 1A) showing CASH phyla branching with green

algae. Interestingly, in 7 of these trees, all CASH were monophyletic, arguing for a common origin of the corresponding genes. Almost all of the 85 genes identified here encode plastid-targeted proteins involved in essential plastid functions (fig. 1B and supplementary table S4, Supplementary Material online). In both chlorarachniophytes and euglenids, these nuclear-encoded 'red' genes participate in plastid genome expression (e.g., elongation factors, aminoacyl-tRNA ligases, ribosomal proteins), light harvesting, chlorophyll biosynthesis, and photosystem II assembly. Keeping these important genes implies a plastid-related selective pressure, which makes improbable that they could have accumulated in the heterotrophic ancestors of green secondary photosynthetic eukaryotes before plastid acquisition.

The disproportion of unexpected gene sources in green *versus* red secondary photosynthetic lineages (~50% and 10%, respectively) is intriguing and may be interpreted in different ways. First, the green algal ancestors of chlorarachniophyte and euglenid plastids may have accumulated many red algal genes by HGT. However, such rampant

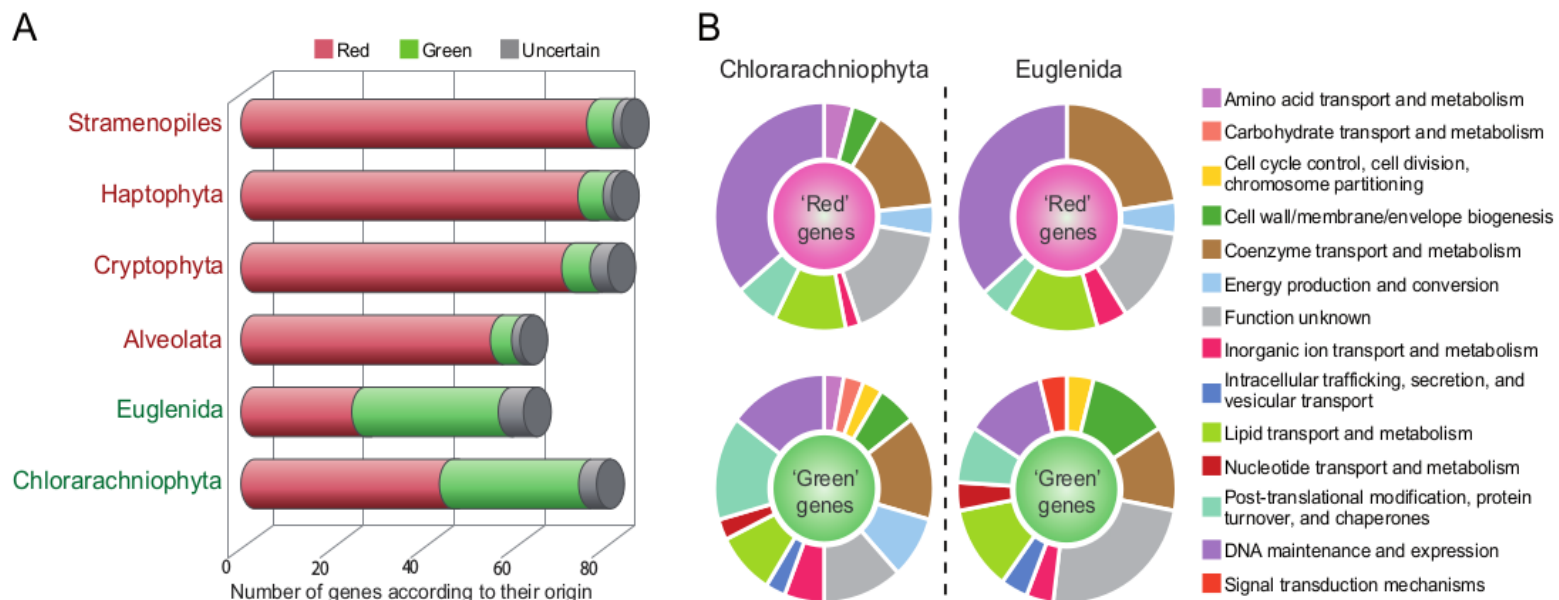


FIG 1. Genes of red and green algal ancestry in secondary photosynthetic eukaryotes. (A) Number of red or green algal-like genes for each lineage among the 85 genes analyzed, gray blocks correspond to genes for which the origin could not be unambiguously determined in the corresponding phylogenetic trees. (B) Gene functions annotated using EggNOG 4.5 (Huerta-Cepas, et al. 2016) for 'red' and 'green' genes detected in transcriptomes and nuclear genomes of euglenids and chlorarachniophytes.

HGT involving essential plastid genes has not been reported so far in any green alga. Second, 'red' genes may have accumulated in chlorarachniophyte and euglenid nuclear genomes by numerous HGTs, for example from food sources. This would imply a long-lasting feeding preference towards red prey in both lineages and also that, for unknown reasons, HGT is much more frequent in secondary green lineages than in red ones. Moreover, the 'red' genes are often shared by all species of the relatively rich taxon sampling available for chlorarachniophytes (fig. 2B), indicating that their acquisition predated the chlorarachniophyte diversification but stopped afterwards (no tree supported recent HGTs involving only chlorarachniophyte subgroups). Despite a poorer taxonomic sampling, a similar trend was observed in euglenids. In addition, chlorarachniophytes and euglenids often branched deep among the red lineages, arguing for an ancient timing of 'red' gene acquisitions comparable to that of the CASH phyla themselves. These observations allow a third interpretation: 'red' genes could have been acquired from ancient cryptic red plastids in chlorarachniophytes and euglenids. The 'red' genes found in chlorarachniophytes are also shared with plastid-endowed stramenopiles and alveolates. Would it be possible that they were transferred from a single red algal endosymbiont ancestral to the whole SAR supergroup (Stramenopiles, Alveolata, and Rhizaria)? This original red plastid would have been lost in many phyla and replaced by a green alga in chlorarachniophytes. However, this scenario poses several problems. First, traces of past red algal plastids, in the form of EGTs, in most non-photosynthetic SAR lineages are scarce and controversial (Elias and Archibald 2009; Stiller, et al. 2009; Stiller 2011). Second, chlorarachniophytes constitute a relatively late-emerging SAR branch (Sierra, et al. 2016); if their present-day green plastid replaced a former red one, this red plastid would have had to be present until recently and been lost in all other rhizarian lineages, which may seem unparsimonious. The case of euglenids is even more difficult to interpret

as these excavates are not closely related with any other photosynthetic lineage. In addition, massive sequence data remain much more limited for euglenids than for chlorarachniophytes (only a few transcriptomes available, see supplementary table S1, Supplementary Material online), making difficult to infer the relative age of possible gene transfers. Nonetheless, 'red' genes were often shared by several euglenids, suggesting a similar pattern of ancient acquisition as in chlorarachniophytes (supplementary figs. S1-S85, Supplementary Material online).

Our results support the presence of an unexpectedly high number of genes of red algal affinity in secondary green algae, euglenids and chlorarachniophytes, which is significantly higher than the frequency of 'green' genes in algae with secondary red plastids, the CASH lineages. Since we have focused on a subset of genes selected because of their strong phylogenetic signal, it is uncertain whether this conclusion can be generalized to the rest of HGTs potentially present in the genomes of these algae. However, we did not identify any particular bias in our gene selection process that could have artificially enriched the 'red' gene frequency observed in euglenids and chlorarachniophytes. Despite the methodological problems inherent to global genome analyses, including the highly unbalanced representation of red and green algal genomes in sequence databases (Deschamps and Moreira 2012), the study of the chlorarachniophyte *B. natans* genome already pointed in that direction, with 22% of EGT genes of apparent red algal ancestry (Curtis, et al. 2012). The origin of 'red' genes in euglenids and chlorarachniophytes, either by cumulative HGT or by EGT from cryptic red algal plastids, remains mysterious but our work indicates that they were acquired early in both groups and that they fulfill functions critical for plastid activity and maintenance. Interestingly, indisputable evidence supports that in the dinoflagellate genus *Lepidodinium* -a third group of complex algae with green plastids- a former red plastid was replaced by the current green one, leading to a mosaic

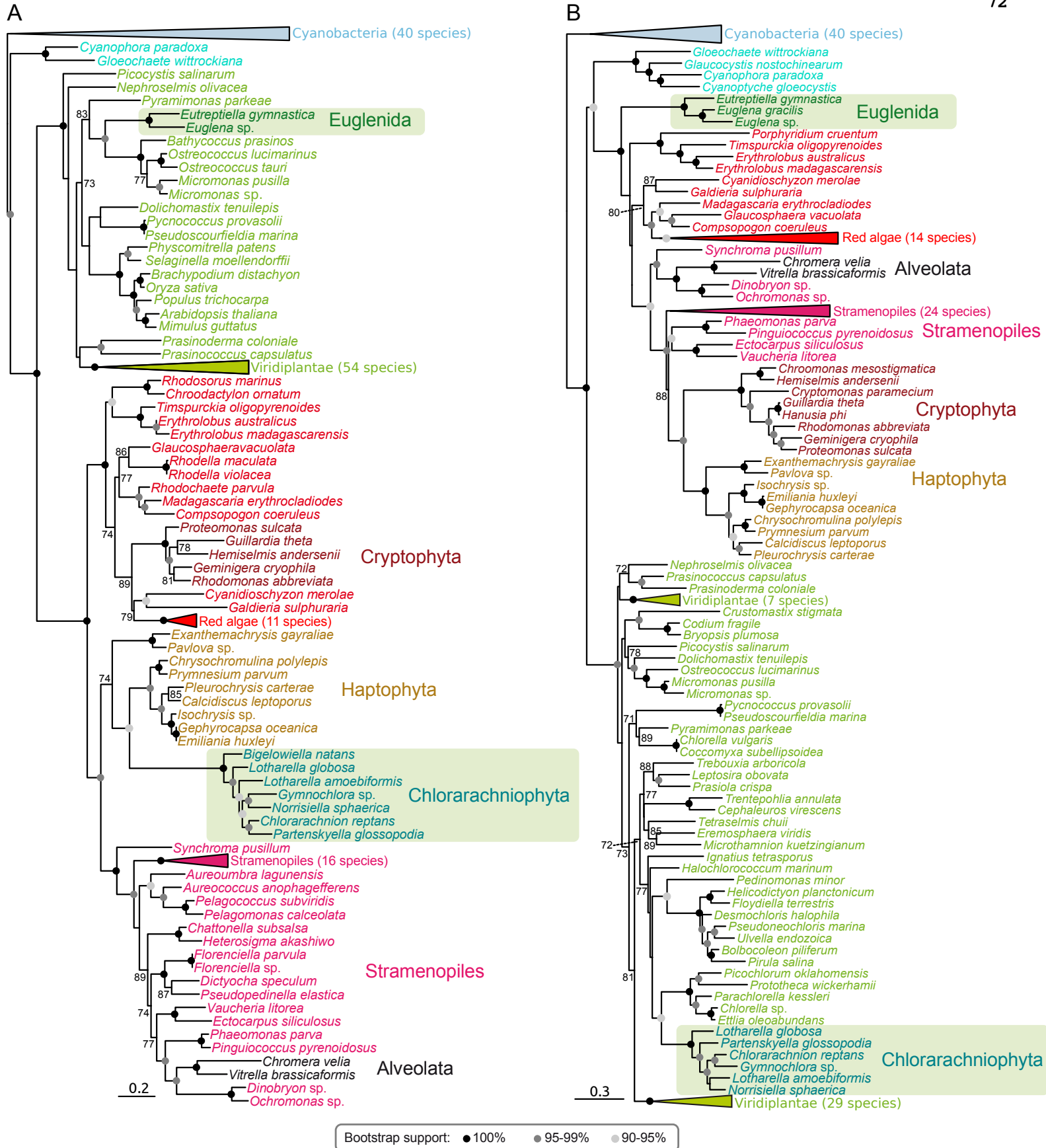


FIG 2. Examples of maximum likelihood phylogenetic trees of nucleus-encoded genes of red and green algal origin in secondary photosynthetic eukaryotes. (A) Protein involved in photosystem II (PSII) assembly (inherited from green algae in euglenids and from a red lineage in chlorarachniophytes). (B) Protein required for thylakoid membrane formation (inherited from green algae in chlorarachniophytes and from a red lineage in euglenids). Bootstrap support values are indicated by black (100%), dark grey (95-99%), and light grey (90-95%) circles. Actual values are provided if between 70 and 90%. Scale bars indicate the number of substitutions per site.

plastid proteome encoded by a mix of red and green algal genes (Minge, et al. 2010) reminiscent of those that we found in euglenids and chlorarachniophytes. It has been proposed that, since they retain more gene-rich genomes than green ones, red plastids have increased capacity for autonomous metabolism that could explain why they are more widespread as secondary plastids across the diversity of eukaryotes (the "portable plastid" hypothesis (Grzebyk, et al. 2003)). It is thus tempting to speculate a similar case for euglenids and chlorarachniophytes as for *Lepidodinium*, with green plastids that replace initial red ones. Even if this hypothesis turns out to be wrong and these cryptic endosymbioses never existed, the ancient acquisition by an unknown mechanism of a significant number of red algal genes in both groups before their diversification and, especially, their maintenance in contemporary species through millions of years of evolution, suggest that the 'red' genes were instrumental in the establishment of the secondary green plastids. Sequencing and analysis of additional genomes of euglenids, chlorarachniophytes, and their non-photosynthetic relatives will help to refine the inventory of 'red' genes and learn more about their timing and, eventually, mechanism of acquisition.

Acknowledgments

This study was supported by European Research Council grant ProtistWorld (P.L.-G., agreement no. 322669), the Université Paris-Sud program "Attractivité" (P.D.) and the Agence Nationale de la Recherche (D.M., project ANR-15-CE32-0003 "ANCESSTRAM").

References

- Abby SS, Tannier E, Gouy M, Daubin V. 2012. Lateral gene transfer as a support for the tree of life. *Proc Natl Acad Sci U S A* 109:4962-4967.
- Archibald JM. 2009. The puzzle of plastid evolution. *Curr Biol.* 19:R81-88.
- Archibald JM, Rogers MB, Toop M, Ishida KI, Keeling PJ. 2003. Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigeloviella natans*. *Proc Natl Acad Sci U S A* 100:7678-7683.
- Baurain D, Brinkmann H, Petersen J, Rodriguez-Ezpeleta N, Stechmann A, Demoulin V, Roger AJ, Burger G, Lang BF, Philippe H. 2010. Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Mol Biol Evol* 27:1698-1709.
- Bodył A, Stiller JW, Mackiewicz P. 2009. Chromalveolate plastids: direct descent or multiple endosymbioses? *Trends Ecol Evol.* 24:119-121.
- Burki F, Flegontov P, Obornik M, Cihlar J, Pain A, Lukes J, Keeling PJ. 2012. Re-evaluating the green versus red signal in eukaryotes with secondary plastid of red algal origin. *Genome Biol Evol* 4:626-635.
- Burki F, Kaplan M, Tikhonenkov DV, Zlatogursky V, Minh BQ, Radaykina LV, Smirnov A, Mylnikov AP, Keeling PJ. 2016. Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc Biol Sci* 283:1823.
- Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, Arias MC, Ball SG, Gile GH, Hirakawa Y, et al. 2012. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* 492:59-65.
- Delwiche CF. 1999. Tracing the thread of plastid diversity through the tapestry of life. *Am. Nat.* 154:S164-S177.
- Deschamps P, Moreira D. 2012. Reevaluating the green contribution to diatom genomes. *Genome Biol Evol* 4:683-688.
- Deschamps P, Moreira D. 2009. Signal conflicts in the phylogeny of the primary photosynthetic eukaryotes. *Mol Biol Evol.* 26:2745-2753.
- Elias M, Archibald JM. 2009. Sizing up the genomic footprint of endosymbiosis. *Bioessays.* 31:1273-1279.
- Grzebyk D, Schofield O, Vetriani C, Falkowski PG. 2003. The mesozoic radiation of eukaryotic algae: the portable plastid hypothesis. *J Phycol* 39:259–267.
- Gutensohn M, Fan E, Frielingsdorf S, Hanner P, Hou B, Hust B, Klosgen RB. 2006. Toc, Tic, Tat et al.: structure and function of protein transport machineries in chloroplasts. *J Plant Physiol.* 163:333-347.
- Hrdá Š, Fousek J, Szabová J, Hampl V, Vlček Č. 2012. The plastid genome of *Eutreptiella* provides a window into the process of secondary endosymbiosis of plastid in euglenids. *PLoS One* 7:e33746.
- Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, et al. 2016. eggNOG 4.5: a hierarchical orthology framework

- with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* 44:D286-293.
- Keeling PJ. 2013. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annu Rev Plant Biol* 64:583-607.
- Kleine T, Maier UG, Leister D. 2009. DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. *Annu Rev Plant Biol*. 60:115-138.
- Lane CE, Archibald JM. 2008. The eukaryotic tree of life: endosymbiosis takes its TOL. *Trends Ecol Evol*. 23:268-275.
- Larkum AW, Lockhart PJ, Howe CJ. 2007. Shopping for plastids. *Trends Plant Sci*. 12:189-195.
- Markunas CM, Triemer RE. 2016. Evolutionary history of the enzymes involved in the Calvin-Benson cycle in euglenids. *J Eukaryot Microbiol* 63:326-339.
- Maruyama S, Suzaki T, Weber AP, Archibald JM, Nozaki H. 2011. Eukaryote-to-eukaryote gene transfer gives rise to genome mosaicism in euglenids. *BMC Evol Biol* 11:1471-2148.
- Minge MA, Shalchian-Tabrizi K, Torresen OK, Takishita K, Probert I, Inagaki Y, Klaveness D, Jakobsen KS. 2010. A phylogenetic mosaic plastid proteome and unusual plastid-targeting signals in the green-colored dinoflagellate *Lepidodinium chlorophorum*. *BMC Evol Biol* 10:191.
- Moreira D, Deschamps P. 2014. What was the real contribution of endosymbionts to the eukaryotic nucleus? Insights from photosynthetic eukaryotes. *Cold Spring Harb Perspect Biol* 6:a016014.
- Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D. 2009. Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science*. 324:1724-1726.
- Muñoz-Gómez SA, Mejía-Franco FG, Durnin K, Colp M, Grisdale CJ, Archibald JM, Slamovits CH. 2017. The new red algal subphylum *Proteorhodophytina* comprises the largest and most divergent plastid genomes known. *Curr Biol* 27:1677-1684.
- Petersen J, Ludewig AK, Michael V, Bunk B, Jarek M, Baurain D, Brinkmann H. 2014. *Chromera velia*, endosymbioses and the rhodoplex hypothesis--plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages). *Genome Biol Evol* 6:666-684.
- Ponce-Toledo RI, Deschamps P, López-García P, Zivanovic Y, Benzerara K, Moreira D. 2017. An early-branching freshwater cyanobacterium at the origin of plastids. *Curr Biol* 27:386-391.
- Reyes-Prieto A, Hackett JD, Soares MB, Bonaldo MF, Bhattacharya D. 2006. Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. *Curr Biol* 16:2320-2325.
- Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ. 2007. The complete chloroplast genome of the chlorarachniophyte *Bigelowiella natans*: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. *Mol Biol Evol*. 24:54-62.

- Sanchez-Puerta MV, Delwiche CF. 2008. A hypothesis for plastid evolution In chromalveolates. *J Phycol.* 44:1097-1107.
- Sierra R, Canas-Duarte SJ, Burki F, Schwelm A, Fogelqvist J, Dixelius C, Gonzalez-Garcia LN, Gile GH, Slamovits CH, Klopp C, et al. 2016. Evolutionary origins of rhizarian parasites. *Mol Biol Evol* 33:980-983.
- Stiller JW. 2011. Experimental design and statistical rigor in phylogenomics of horizontal and endosymbiotic gene transfer. *BMC Evol Biol* 11:1471-2148.
- Stiller JW, Huang J, Ding Q, Tian J, Goodwillie C. 2009. Are algal genes in nonphotosynthetic protists evidence of historical plastid endosymbioses? *BMC Genomics.* 10:484.
- Suzuki S, Hirakawa Y, Kofuji R, Sugita M, Ishida K. 2016. Plastid genome sequences of *Gymnochlora stellata*, *Lotharella vacuolata*, and *Partenskyella glossopodia* reveal remarkable structural conservation among chlorarachniophyte species. *J Plant Res* 129:581-590.
- Weeden NF. 1981. Genetic and biochemical implications of the endosymbiotic origin of the chloroplast. *J Mol Evol.* 17:133-139.
- Woehle C, Dagan T, Martin WF, Gould SB. 2011. Red and problematic green phylogenetic signals among thousands of nuclear genes from the photosynthetic and apicomplexa-related *Chromera velia*. *Genome Biol Evol* 3:1220-1230.
- Yang Y, Maruyama S, Sekimoto H, Sakayama H, Nozaki H. 2011. An extended phylogenetic analysis reveals ancient origin of "non-green" phosphoribulokinase genes from two lineages of "green" secondary photosynthetic eukaryotes: Euglenophyta and Chlorarachniophyta. *BMC Res Notes* 4:330.
- Yang Y, Matsuzaki M, Takahashi F, Qu L, Nozaki H. 2014. Phylogenomic analysis of "red" genes from two divergent species of the "green" secondary phototrophs, the chlorarachniophytes, suggests multiple horizontal gene transfers from the red lineage before the divergence of extant chlorarachniophytes. *PLoS One* 9:e101158.
- Yoon HS, Hackett JD, Pinto G, Bhattacharya D. 2002. The single, ancient origin of chromist plastids. *Proc Natl Acad Sci U S A.* 99:15507-15512.

6. ORIGIN OF THE SELMA TRANSLOCON IN SECONDARY PLASTIDS OF RED ALGAL ORIGIN

“About 30 years ago there was much talk that geologists ought only to observe and not theorise; and I well remember some one saying that at this rate a man might as well go into a gravel-pit and count the pebbles and describe the colours. How odd it is that anyone should not see that all observation must be for or against some view if it is to be of any service!”

Charles Darwin. Letter to Henry Fawcett (September 18, 1861)

6. ORIGIN OF THE SELMA TRANSLOCON IN SECONDARY PLASTIDS OF RED ALGAL ORIGIN

6.1 Context and objectives

As we have seen in the previous chapter, although secondary plastid-harboring lineages have been widely studied, there are still many questions that remain unresolved. In regard to the evolution of secondary red plastids, the major debate concerns the number of endosymbiotic events at the origin of all chlorophyll c-containing algae (section 1.3.2). The monophyly of the SELMA translocon is considered to provide strong evidence of the single origin of secondary red plastids. This protein import system is thought to be shared by all lineages with secondary red plastids surrounded by four membranes. However, the monophyly of SELMA has been inferred from phylogenetic analyses of very few proteins (Felsner et al., 2011). The claim for a SELMA system shared by all secondary red algae has been mostly based on the identification using similarity searches of SELMA-like homologs containing transit peptide (TP) sequences (Stork et al., 2012).

However, TP sequences usually have lineage-specific characteristics that make difficult their identification in oversimplified TP prediction models. Although several programs have been developed to account for the variability of TP sequences among secondary plastid-harboring lineages (Cilingir et al., 2012; Gruber et al., 2015; Gschloessl et al., 2008), prediction programs may fail (false positives or false negatives) or some proteins may lack TP sequences if they were not completely sequenced. Above all, it is important to consider that similarity searches and TP detection do not give any information regarding the evolutionary origin of proteins, and phylogenetic analyses must be performed in order to determine the monophyly of the different components.

To continue with our study of the evolution of secondary photosynthetic eukaryotes, we conducted a comprehensive phylogenetic study of the SELMA translocon aiming to provide key insights into the acquisition of secondary red plastids and their evolution: do all SELMA components derive from the red algal endosymbiont as proposed by the standard model? Are all the SELMA proteins monophyletic?

In this last part of my PhD research project, we performed phylogenetic analyses on the largest collection of putative SELMA components reported so far (Stork et al.,

2012) and analyzed the best-fit scenario for the evolution of this translocon under the chromalveolate hypothesis or a model of serial endosymbioses.

6.2 Results

Our phylogenetic analyses of putative SELMA components show that most of these proteins do not have a monophyletic origin (Table 2). The composition and origin of the SELMA translocon in cryptophytes is extremely different from the import machinery established in stramenopiles, haptophytes and alveolates. In some cases, these red lineages seem to have recruited a different set of paralogous proteins from the ERAD system of the red algal endosymbiont. We show that the evolutionary history of SELMA is not as straightforward as it is commonly assumed in the standard models that have been proposed to explain the origin and evolution of secondary red algae. In the next section I show a preliminary draft of the article manuscript for a detailed description of our results.

Function of the Group	Protein	Cryptophyta	Haptophyta	Stramenopiles	Apicomplexa
Ubiquitination machinery responsible of the ubiquitination of pre-proteins necessary for their import into the PPC	Uba1 (E1)				
	Ubc_new1 (E2)				
	Ubc1 (E2)				
	Ubc4 (E2)				
	Ubc6 (E2)				
	Ubc7 (E2)				
	Ubc8 (E2)				
	Hrd1 (E3)				
	ptE3p (E3)				
	ptDUP				
Translocation of ubiquitinated proteins in a ATP-dependent manner	Ufd1				
	Npl4				
	Cdc48				
Pore-forming translocon	DerA				
	DerA or DerB				
	DerB				
Chaperone proteins	Hsp70				
	Rad23				

Table 2. The different putative components of SELMA and their respective inferred phylogenetic origins in CASH lineages. Empty spaces in the table imply that the corresponding protein is not present or was not detected in our analyses. R: gene derived from the red algal endosymbiont; X: Unknown origin; E: gene derived from the eukaryotic host; C: Cryptophytes; H: Haptophytes; S: Stramenopiles; A: Apicomplexa; Nm: gene encoded in the nucleomorph.

6.3 Manuscript of article 3

Molecular phylogeny disputes the monophyly of the SELMA translocation machinery in complex secondary plastids.

(Manuscript in preparation)

Rafael I. Ponce-Toledo¹, David Moreira¹, Purificación López-García¹ and Philippe Deschamps^{1,*}

¹Unité d'Ecologie Systématique et Evolution, CNRS, Université Paris-Sud, AgroParisTech, Université Paris-Saclay, 91400, Orsay, France

*To whom correspondence should be addressed.

E-mail: philippe.deschamps@u-psud.fr

INTRODUCTION

Eukaryotes have acquired photosynthesis through the engulfment of a cyanobacterium symbiont more than 1.5 billion year ago (Yoon et al. 2004). Three extant lineages are known to have diversified after this primary endosymbiosis: Glaucophytes, Viridiplantae and Rhodophytes, composing the Archaeplastida supergroup (Adl et al. 2005). All photosynthetic Archaeplastida carry a plastid surrounded by two membranes and containing a reduced genome that derived from the cyanobacterium endosymbiont. Depending on the lineage, these small circular genomes encode 60 to 250 genes involved in their own maintenance and expression, as well as in the building of the photosystem (Barbrook et al. 2006).

All other proteins acting in the plastid are encoded in the nuclear genome, transcribed and translated via the canonical machinery and relocated to the plastid compartment using a specialized addressing pathway. This pathway requires a N-terminal tag sequence called the transit peptide. Unfolded proteins are taken in charge by a cytosolic guidance complex that will deliver them to the the outer plastid membrane. Two multi-protein integral translocators of the inner and outer chloroplast membranes, shortly named Tic and Toc, will then recognize the transit peptide and allow proteins to enter the lumen of the plastid where they can either stay soluble or be integrated into the chloroplast or thylakoid membranes (Sjuts et al. 2017).

A survey of the genomes of Archaeplastida has shown that many of the genes they encode are not phylogenetically related to eukaryotes but to cyanobacteria (Martin et al. 2002). These genes were acquired from the endosymbiont soon after primary endosymbiosis via a process called Endosymbiotic Gene Transfer (EGT) (Martin et al. 1998). Many of these genes are addressed to the plastid in extant Archaeplastida (Reyes-Prieto et al. 2006), but they could also have found a role in other compartment of the cell, making primary endosymbiosis an incredible source of metabolic reshuffling and

innovation. The Tic/Toc machinery is one example, as it is composed of proteins from both the host and the symbiont of primary endosymbiosis (Shi and Theg 2013). Moreover, the composition of Tic/Toc being identical in all Archaeplastida species, it has been considered, combined with several phylogenetic evidences (Moreira et al. 2000; Rodríguez-Ezpeleta et al. 2005), as a strong indication that all Archaeplastida share a common ancestor that derived from a single primary endosymbiotic event (McFadden and van Dooren 2004; Steiner et al. 2005).

Although the nature of the common eukaryotic ancestor of Archaeplastida is undefined, a recent study points the ancestor of the primary plastid to an extinct cyanobacterium related to the “early diverging” freshwater Gloeomargaritales (Ponce-Toledo et al. 2017). There is only one other evolutionary event comparable to primary endosymbiosis, involving a marine cyanobacterium related to the Prochlorococcus/Synechococcus clade and a phagotrophic cercozoan, that gave rise to the poorly diversified thecate amoebas of the genus *Paulinella* (Marin et al. 2005; Yoon et al. 2009). These species show a lower degree of integration of the organelle, with a less reduced genome, a moderate amount of EGTs and few identified cases of protein addressing (Nowack et al. 2011; Gagat and Mackiewicz 2014).

Photosynthesis in eukaryotes is not restricted to Archaeplastida and *Paulinella*. Many other distant phyla also carry plastids that derive from so-called ‘complex’ endosymbioses. Chlorarachniophytes (Rhizaria) and Euglenids (Excavata) both carry plastids with pigments and photosystems similar to green algae and surrounded by three and four membranes, respectively. Four phyla contain species with plastids similar to red algae, containing chlorophyll c and surrounded by four membranes: Cryptophytes, Alveolates, Stramenopiles and Haptophytes, with the exception of the Dinoflagellates (Alveolates) that only have three membranes (Keeling 2010). The presence of additional membranes and the similarities with plastid features of Archaeplastida early led to the

hypothesis that all these complex lineages evolved by secondary endosymbioses involving heterotrophic eukaryote hosts and green or red algae (Taylor 1974). Phylogenetic analyses of plastid encoded genes repeatedly proved this point and even determined that plastid of Euglenids and Chlorarachniophytes derive from two independent secondary endosymbioses involving a Prasinophyceae and a Ulvophyceae respectively (Rogers et al. 2007; Hrdá et al. 2012; Suzuki et al. 2016). Similarly, many old and recent phylogenetic analyses have concluded that all rhodophyte-like secondary plastids are monophyletic, and may have derived from a single secondary endosymbiosis (Rodríguez-Ezpeleta et al. 2005; Baurain et al. 2010).

Already in 1982 had Cavalier-Smith intuitively proposed that all eukaryotes with complex plastids containing chlorophyll c derived from a single secondary endosymbiosis, and proposed to call the entire group Chromalveolates (Cavalier-Smith 1982). He himself recently renamed them simply Chromista (Cavalier-Smith 2017), while other authors have proposed to call the same super-group CASH (Petersen et al. 2014). However, efforts aiming at reconstructing the phylogeny of eukaryotes nuclei have failed to retrieve the monophyly of the CASH super-group (Baurain et al. 2010). Moreover, all lineages member of the CASH comprise a large part of non-photosynthetic species, including early diverging branches, with no detectable traces of plastids, nor leftover genes related to Archaeplastida, leaving open the possibility of several independent secondary endosymbioses, in each CASH lineage, with very similar red algae.

The incongruence between plastid phylogeny and nuclear phylogeny, and the existence of heterotrophic chromists is interpreted in two ways: Cavalier-Smith advocates that artifacts in molecular phylogenies are to blame and that the history told by cellular features as well as genes presence/absence patterns should prevail. For him, these latter elements are compatible with the monophyly of Chromists as a whole, who emerged from a unique founder secondary endosymbiosis followed by many (at least 16) secondary

losses (Cavalier-Smith 2017). Other authors consider that the monophyly of all secondary red plastids is compatible with the polyphyly of secondary photosynthetic lineages if the red plastid was transmitted via a chain of higher order endosymbioses. Sometimes termed the “rhodoplex” hypothesis (Bodył et al. 2009; Petersen et al. 2014; Stiller et al. 2014) this scenario depicts a single first secondary endosymbiosis setting up the secondary rhodoplast in one CASH lineage. This rhodoplast is then latter transmitted by tertiary endosymbiosis to another CASH lineage which itself gave it to another lineage by quaternary endosymbiosis. In such a scenario, phylogenies reconstructed using plastid genes will depict a monophyletic ancestry of the plastid in all CASH, each lineage nested within one another. At the same time, phylogenies of nuclear genes of the corresponding lineages may support the polyphyly of CASH lineages. Finally, given the observed acceleration of molecular evolution of plastid encoded sequences (Deschamps and Moreira 2009), it is also possible that the monophyly of secondary red plastids is an artifact, and that multiple secondary endosymbioses involving closely related red algae happened in distant parts of the eukaryotic tree.

Like in the case of Archaeplastida, nuclear genomes of secondary photosynthetic eukaryotes are complex mosaics composed of genes inherited from their eukaryote ancestors, as well as genes transferred by EGT from their former endosymbionts (Lane and Archibald 2008). These later being already eukaryote-prokaryote assemblages, the final mosaic in secondary lineages presents a higher level of complexity, with genes from cyanobacteria and two types of genes from eukaryotes.

In Chlorarachniophytes and Cryptophytes, an additional genetic compartment exists corresponding to the former nucleus of the secondary symbiont (Greenwood 1974; Hibberd and Norris 1984). This nucleomorph is located in the periplastidial space (PPS), between the second outermost membrane (or Periplastidial Membrane, PPM) and the plastid double membrane (PM) and contains a vestigial genome encoding 300 to 450

genes involved in its own replication and expression as well as in some metabolic functions of the plastid (Maier et al. 2000).

Similarly to Archaeplastida, a great proportion of nucleus-encoded proteins in secondary photosynthetic lineages needs to be addressed to the plastid. If the Tic/Toc machinery also exists in the two innermost membrane of secondary plastids (Stork et al. 2013), proteins shall also pass through one (euglenids and dinoflagellates) or two (chlorarachniophytes and all other CASH) additional outer membranes. Concerning the outermost membrane, two observations have helped understanding how it is crossed. First, in Cryptophytes, Stramenopiles and Haptophytes this membrane is fused with the nuclear envelope and populated with 80S ribosomes, making it an equivalent of the reticulum membrane (Cavalier-Smith 2002). Second, in every secondary lineage, the heading sequence of nuclear-encoded plastid-located proteins contains a bipartite transit signal (BTS), a combination of a signal peptide and a transit peptide indicating that plastid targeted proteins are translocated into the reticulum using the Sec61 secretion pathway (Bhaya and Grossman 1991). In the special case of Alveolates, where the plastid is not located in the reticulum, it has been shown that proteins are packaged into vesicles to be transported toward the the plastid where they fuse and deliver proteins in the PPS (Agrawal and Striepen 2010). While the existence of a dedicated translocation machinery at the PPM was theorized long ago, its actual functioning in CASH phyla has only been recently unveiled (Stork et al. 2013). Conversely, for chlorarachniophytes, the story stops here, and no one knows how targeted proteins cross the PPM.

The understanding of how the PPM is crossed in CASH species came from the analysis of the gene content of the nucleomorph genome of the cryptophyte *Guillardia theta* where Sommer et al. identified components of the former red-algal ERAD translocon machinery (Sommer et al. 2007). In the same work, putative homologs of the components found in nucleomorphs were also detected in the genome of alveolates and stramenopiles,

with additional BTS, suggesting the relocation of their protein product to the plastid. ERAD (Endoplasmic Reticulum Associated Degradation) is a ubiquitin dependent pathway used by all eukaryotic cells to extract misfolded proteins from the reticulum back into the cytoplasm where they will be degraded by the proteasome (Meusser et al. 2005). In the context of a minimized cytoplasm like the PPC, where it is not yet clear if there is a endomembrane network, it was speculated that this copy of ERAD could have been tinkered to create a translocon to import proteins across the PPM. This putative new complex was named SELMA for Symbiont-specific Erad-Like MAchinery and the first experimental evidence supporting its function was provided by fluorescence localization as well as preprotein transit peptide interaction experiments of plastid-targeted Derlins (the putative constituents of the translocation pore) (Hempel et al. 2009) and a conditional null mutant of the same Derlin in *Toxoplasma gondii* (Agrawal et al. 2009).

The structure of the SELMA translocon has been further dissected both by *in silico* (Felsner et al. 2011; Moog et al. 2011; Stork et al. 2012) and by functional analyses using mainly two models: *Toxoplasma gondii* and *Phaeodactylum tricornutum* (Hempel et al. 2010; Agrawal et al. 2013; Lau et al. 2015; Lau et al. 2016; Fellows et al. 2017). As to now, SELMA is hypothesized to function using about 15 proteins classified in four functional categories (Table 1). (1) a pore made of a complex of Derlin (Der1) proteins through which proteins enter the PPC. (2) An ubiquitylation machinery to transfer ubiquitin moieties to translocated pre-protein so they can be recognized by the cdc48 complex. It is composed of a SELMA-specific polyubiquitin, a ubiquitin activating enzyme E1, a conjugating enzyme E2, a ubiquitin ligase E3. To avoid a potential degradation of translocated proteins by the proteasome of the PPC, and because the pre-proteins need to be devoid of ubiquitin to cross the Tic/Toc complex, there is also an hypothesized de-ubiquitylation enzyme (Hempel et al. 2010). (3) A cdc48 complex consuming ATP to pull the ubiquitylated pre-proteins through the Derlin pore, and composed of Cdc48, Ufd1 and Npl4. (4) A set of

accessory chaperones. Stork et al. have produced the most comprehensive genomic survey of SELMA components in all available (as of 2012) genomes of red secondary lineages. This survey was conducted using similarity Blast searches starting with ERAD proteins of *Saccharomyces* and SELMA components previously identified in cryptophytes, apicomplexans and stramenopiles as references to query other genomes. The classification of all detected proteins as either part of ERAD or SELMA was based on similarity and on *in silico* detection of BTS (Stork et al. 2012). Using this survey, the authors argued that all CASH phyla possess a symbiont-derived SELMA machinery. However, apart from the Derlins (Hirakawa et al. 2012; Petersen et al. 2014; Cavalier-Smith 2017), cdc48 (Agrawal et al. 2009; Felsner et al. 2011; Kienle et al. 2016), and Uba1 (Felsner et al. 2011), no systematic phylogenetic analysis of the components of the SELMA complex was ever done.

The various phylogenetic analysis of the Cdc48 protein family have concluded that all CASH have inherited their SELMA copies from a red alga. From this single evidence, the community has taken for granted that the whole SELMA complex has a common ancestry across all CASH, the cryptophytes nucleomorph-encoded version being the “original setup” because it is still partially encoded in the former symbiont nucleus (Gould et al. 2015; Cavalier-Smith 2017). This needs, however, to be validated. In this work, we propose to determine the phylogeny of all described components of the SELMA complex and to assess if they support the hypothesis of a monophyletic origin of the complex in the CASH super-group.

MATERIAL AND METHODS

Dataset construction and Phylogenetic inferences

For all protein families analyzed in this work, the starting material is the the set of sequences accession numbers reported in table 1 and supplementary table S1 provided in Stork et al. 2012 (Stork et al. 2012). The corresponding protein sequences were used to query a local custom genomes and transcriptomes database using psi-blast (Altschul et al. 1997), allowing for 5 rounds of sequences searches and an e-value threshold of $1e^{-5}$. The composition of this database is available in supplementary table 1. All hits retrieved using psi-blast were collected in a first multi-protein file and were aligned using MAFFT v7.205 with default parameters (Kato and Toh 2008). Gaps and poorly informative sites were removed using BMGE (Criscuolo and Gribaldo 2010). Some of the SELMA component proteins are rather short, depending on the size of each alignments, we used either the BLOSUM62 or BLOSUM30 substitution matrix in BMGE to maximize the number of recovered sites for phylogenetic tree reconstruction. A first set of phylogenetic trees were reconstructed using FastTree v2.1.7, default parameters (Price et al. 2010). These first trees were manually inspected and pruned to remove duplicate sequences as well as groups of similar proteins that were not related to the orthologous groups of interest. The remaining sequences were realigned and trimmed using the same parameters, and final trees reconstructed with Iqtree v1.5.5 using two combinations of parameters : LG+G(4 cat) +I and PMSF : LG+C60+F+G(4 cat) using as guide trees the ones produced with the first parameters (Minh et al. 2013; Nguyen et al. 2014).

RESULTS

Stork et al. (2012) identified ERAD and SELMA components in various protists using Blast similarity searches. In their model, it is expected that all secondary

endosymbiotic lineages will carry at least 2 copies of each functional component of an ERAD-like machinery; one involved in ERAD, the other in SELMA.

We carried out phylogenetic analyses of 14 components shared by ERAD and SELMA. We gathered all previously reported protein identifiers and used them to retrieve as many similar proteins as possible using psi-blast against a custom genome database. We then used these collections of similar proteins to reconstruct phylogenetic trees. Previous reports argued about the difficulty to reconstruct reliable phylogenetic trees, due to a putative high degree of divergence of the SELMA proteins. Though we also had difficulties to obtain well resolved trees, for most cases, we found sufficient phylogenetic signal to infer the evolution of SELMA components.

Orthologs classification

In the following part, the terminology of previous studies will be used: ERAD components in secondary endosymbiotic lineages will have the prefix h (for host), and SELMA components will be prefixed s (for symbiont). However, we must stress out that host or symbiont prefixes are only a commodity to differentiate ERAD and SELMA, but do not indicate, as we will show below, that the corresponding proteins is actually phylogenetically derived from the former host ERAD or from the red algal symbiont ERAD. In the case of cryptophytes, all components encoded in the nucleomorph genome are *de facto* part of SELMA and evolved from the red algal symbiont. For all other phyla, the classification of each protein as part of cytosolic ERAD or plastidial SELMA was determined by Stork *et al.* using similarity as a proxy of their distance with ERAD genes from *Saccharomyces* and ERAD genes from red algae. Additionally, if a BTS was detected in a ERAD-like protein, in addition to the cytosolic copy, it was considered to be potentially part of the symbiont-derived translocation machinery. However, plastid-targeted does not

necessarily means symbiont-derived. Moreover, As it has been repetitively reported, the similarity ordered hits list produced by blast searches does not reflect the actual phylogenetic relationships between the query and the hits (Koski and Golding 2001). Thus, when trying to circumscribe groups of orthologs in large protein families, it is common to mix up members between different paralogous groups. As a matter of fact, from our phylogenetic trees we determined that some Ubc paralogs were misclassified.

The Der1 translocon

Phylogenetic analyses of the SELMA pore-forming Der1 proteins are scarce and uncertain about their shared ancestry (Hirakawa et al. 2012; Petersen et al. 2014). Cavalier-Smith recently published a set of phylogenies of the Derlin proteins where he redefined two paralogous groups namely DerA and DerB corresponding respectively to Der1 and Dfm1, the two ERAD members found in *Saccharomyces cerevisiae* (Cavalier-Smith 2017). In his interpretation, SELMA DerA is only present in the nucleomorph of cryptophytes and is the only copy in this lineage. In all other CASH, he found two SELMA copies that would have derived from DerB and were duplicated in the ancestor of Chromists.

When looking at the many versions of the phylogenetic trees of derlin proteins in the supplementary information of Cavalier-Smith (2017), it appears that the phylogenetic signal is very unstable. We had similar issues with our analyses and were unable to recover his results. It seems from our tree (supplementary fig. S16) that cryptophytes only have one SELMA Derlin gene derived from DerA in their nucleomorph. We could also recover one monophyletic putative SELMA DerB shared by alveolates, stramenopiles and haptophytes, but for the second putative SELMA Derlin shared by the same lineages, we were unable to define if it is closer to DerA or DerB. It is worth mentioning that the position

Function of the Group	Protein	Cryptophyta	Haptophyta	Stramenopiles	Apicomplexa
Ubiquitination machinery responsible of the ubiquitination of pre-proteins necessary for their import into the PPC	Uba1 (E1)				
	Ubc_new1 (E2)				
	Ubc1 (E2)				
	Ubc4 (E2)				
	Ubc6 (E2)				
	Ubc7 (E2)				
	Ubc8 (E2)				
	Hrd1 (E3)				
	ptE3p (E3)				
	ptDUP				
Translocation of ubiquitinated proteins in a ATP-dependent manner	Ufd1				
	Npl4				
	Cdc48				
Pore-forming translocon	DerA				
	DerA or DerB				
	DerB				
Chaperone proteins	Hsp70				
	Rad23				

Table 1. The different putative components of SELMA and their respective inferred phylogenetic origins in CASH lineages. Empty spaces in the table imply that the corresponding protein is not present or was not detected in our analyses. R: gene derived from the red algal endosymbiont; X: Unknown origin; E: gene derived from the eukaryotic host; C: Cryptophytes; H: Haptophytes; S: Stramenopiles; A: Apicomplexa; Nm: gene encoded in the nucleomorph.

of Der1 and Dfm1 of *Saccharomyces* in our trees were very unstable, and almost never clustered with similar proteins of other fungi, questioning their validity as reference to classify the two paralogous groups.

The ubiquitination machinery: Uba1, Ubc, Hrd1, ptDUP

Uba1, the ubiquitin-activating enzymes, or simply enzymes E1 catalyze the attachment of ubiquitin moieties to the ubiquitin conjugating protein (E2). The gene phylogeny of Uba1 shows that haptophytes, stramenopiles and alveolates encode a sUba1 gene that derive from the same red algal hUba1 ancestor gene. However in cryptophytes, contrary to what was previously published (Felsner et al. 2011), the putative protein corresponding to sUba1 is likely a duplication of the host's ERAD hUba1 (table 1, supplementary fig. S1). The ubiquitin-conjugating (Ubc) enzymes E2 transiently carry ubiquitin for a future transfer on targeted pre-proteins. In their report, Stork et al. detected that different secondary lineages have a putative SELMA copy derived from various isoforms of Ubc. However, a general tree of all isoforms clearly clusters all of them in a single subfamily, that we called Ubc_new in Table 1, because it did not group with any of the eight ERAD isoforms identified so far (Figure 1A). The topology of the phylogenetic tree restricted to Ubc_new displays a shared origin of sUbc protein in the Alveolates-Stramenopiles-Haptophytes assemblage likely by EGT from a red algal ancestor of their plastid (Table 1, Figure 1B). In cryptophytes, we detected an early duplication of the Ubc7 gene of the host. One copy has a putative BTS, suggesting that one of host ERAD-like duplicates might be directed to the plastid as part of the SELMA machinery, but this would need functional confirmation. Interestingly, cryptophytes possess a higher number of Ubcs, including two encoded in the nucleomorph, with no detected homologs in other lineages.

Ubiquitin ligases E3 (Hrd1) transfer ubiquitin to plastid pre-proteins. As reported previously (Hempel et al. 2010), only cryptophytes possess a sHrd1 protein, encoded in the nucleomorph. Hempel et al. used relaxed similarity searches to try to find a protein supporting the same function in stramenopiles. They identified ptE3P, a protein that has a measurable ubiquitin transferase activity, and is located in the PPC. Our similarity searches show that this protein exists in stramenopiles and haptophytes but with no apparent phylogenetic relationships (the corresponding gene tree is unresolved, (see supplementary fig. S10), indicating that this function is probably taken in charge by another type of enzyme in alveolates. Altogether, the recruitment of different proteins to catalyze the same function testifies that there is no shared ancestry of sHrd1 in CASH.

For the SELMA model to be correct, targeted proteins need to be deubiquitinated before passing through the two innermost membranes of the plastid, and to avoid their destruction by the proteasome of the PPC. Hempel et al. (2010) detected in stramenopiles a protein putatively supporting this function. Similar to our results of Uba1 and some Ubc proteins of cryptophytes, we found that the plastid-targeted deubiquitinases derive from a duplication of a host gene (see supplementary fig. S11). Likewise, we found that the

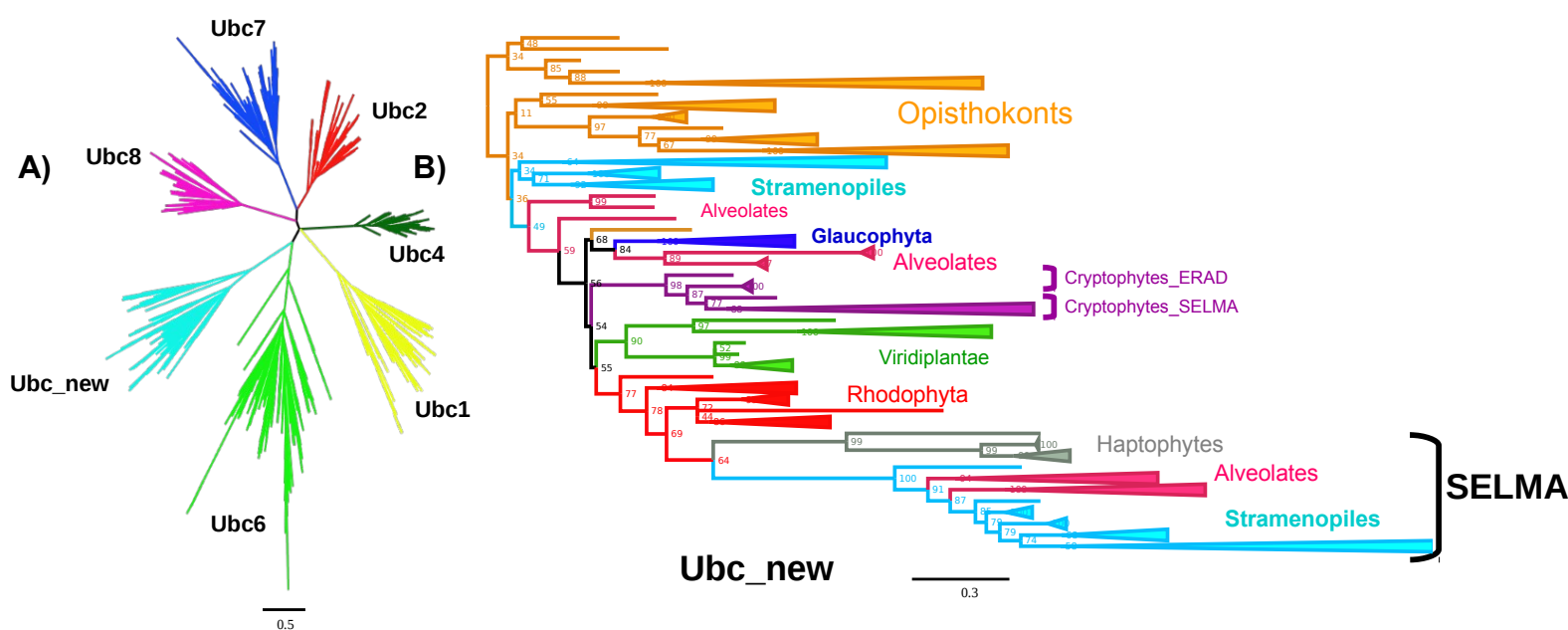


Figure 1. Phylogenies of Ubc paralogs. A) Unrooted phylogenetic tree of all similarity-based detected Ubc proteins. B) ML phylogenetic tree restricted to the Ubc_new isoform. Numbers represent bootstrap values. The scale bars represent the number of substitutions per a unit branch length.

symbiont-specific deubiquitinase reported in haptophytes by Stork et al. (2012), is also a paralogous gene of the host. However, we could not detect a bipartite targeting sequence in the haptophyte proteins. The origin of the plastid-targeted deubiquitinases in stramenopiles is unclear.

The Cdc48-Ufd1-Npl4 complex.

Cdc48, is an ATPase promoting the transport of ubiquitinated proteins through the Derlin complex. We found, like other authors, two copies of sCdc48 (sCdc48-1 and sCdc48-2) in cryptophytes, stramenopiles and haptophytes, and only one in plastid-bearing alveolates (Table 1). In cryptophytes, one copy is encoded in the nucleomorph while the other was transferred to the nucleus. All trees reported so far strongly suggests an early duplication of a single red algal hCdc48 gene in the ancestor of all CASH, one copy having likely been lost in the alveolate ancestor. We also found a clear red algal ancestry for all sCdc48 genes. In our tree reconstructed with the LG model (supplementary fig. S14), we observed that the nucleomorph sCdc48 copy branches between red algae and all other CASH sCdc48, including the nucleus encoded cryptophyte one. However, the tree topology sparsely reconcile with the relayed idea that this nucleomorph copy is the ancestor of the other, eventually by higher level endosymbioses. Indeed, some duplication are lineage specific and do not correspond to each cryptophytes isoform, as one could expect if they were transmitted via EGT. Additionally, when using the PMSF site heterogeneous model of IQ-tree, the nucleomorph sCdc48 copy branched in between different red algae than all other CASH sCdc48, suggesting the possibility that these phylogenies suffer from hidden paralogy (supplementary fig. S15).

Concerning the cofactors sUfd1 and sNpl4, we could determine that both are of red algal origin in cryptophytes but are not related to other CASH genes (table 1,

supplementary figures S12-S13). Instead, our trees show that the PPC-targeted sUfd1 and sNpl4 in the haptophytes-stramenopiles-alveolates assemblage both share a common ancestor gene, but we could not properly determine its origin. These results suggest an independent recruitment of different ancestral genes in cryptophytes compared to other CASH lineages.

Accessory proteins

An Hsp70 chaperone might be involved in the correct folding of translocated proteins in the PPC. In cryptophytes, the symbiont-specific sHsp70 is encoded in the nucleomorph, a clear evidence of its red algal ancestry. In our tree, the PPC-located Hsp70 homologs of stramenopiles group with potential symbiont copies of haptophytes, but could not clearly identify whether all these proteins are derived from the red endosymbiont or from an early duplicated gene from the host. Anyway, they are definitely not related to the nucleomorph-encoded Hsp70 of cryptophytes, once again supporting the idea of an alternative recruitment of ancestral genes in cryptophytes compared to other CASH.

DISCUSSION

It is well established that primary and higher order endosymbioses have shaped the evolution of photosynthetic eukaryotes. When Cavalier-Smith proposed the chromalveolates hypothesis, arguing for a single secondary endosymbiosis between a red alga and a common eukaryote ancestor for alveolates and chromists (*i.e.* cryptophytes, haptophytes and stramenopiles), he was exposing what he believes is the most parsimonious explanation for all the shared traits between lineages carrying plastids with chlorophyll *c* (Cavalier-Smith 1999). This opened two decades of intense debate as

phylogenetics and later phylogenomics were unable to confirm or infirm a molecular support for the chromalveolates. Indeed, the evolutionary history of these lineages is likely more complex and tangled than Cavalier-Smith initially envisioned. Plastid encoded genes of all chromalveolates were shown to derive from the same red alga. However, phylogenies of nuclear-encoded proteins do not recover the monophyly of the host cell (Petersen et al. 2006; Teich et al. 2007; Baurain et al. 2010; Burki et al. 2016). When this incongruence became clear, new hypotheses were proposed, including the possibility of serial higher-order endosymbioses (Bodył et al. 2009; Petersen et al. 2014) or kleptoplasty (Bodył 2017).

When genes of the SELMA translocon were identified in the nucleomorph of *Guillardia theta*, the idea of a preexisting protein transport machinery, originally encoded by the symbiont, and ready to import proteins from the REG lumen across the PPM with moderate tinkering was immediately charming. The observation that all four membrane plastids carrying chromalveolates seemed to have a similar SELMA translocon at the second outermost membrane was considered as a strong evidence for the monophyly of the secondary red plastid. The underlying idea, similarly to Tic/Toc for primary plastids, is the unlikelihood that similar transport systems could have been assembled independently multiple times from ERAD machineries of similar red algal endosymbionts. However, and contrary to Tic/Toc, the phylogenetic support for the monophyly of SELMA has never been systematically tested and rely on only two proteins upon a dozen, Cdc48 and Uba1 (Felsner et al. 2011).

In this study, we performed phylogenetic analyses of a collection of ERAD/SELMA proteins in order to better understand the evolution of the PPM translocation machinery. Except (and only perhaps) for the sCdc48 proteins, we could not retrieve the monophyly of the putative SELMA components in CASH (Table 1). The general trend is that the nucleomorph-encoded copy of cryptophytes is not related to homologous proteins found in

other CASH phyla. Even in the case of Cdc48, depending on the chosen evolution model (LG vs site-heterogeneous), the nuclear-encoded gene shared by all CASH phyla does not always derive from the same red alga than the nucleomorph-encoded gene of cryptophytes (supplementary figures S14-S15). Even the cryptophytes nuclear copy is not necessarily related to the nucleomorph one, indicating a possible case of hidden paralogy that would dismiss this gene from supporting the acknowledged view of the evolution of SELMA. For almost all other proteins, our phylogenetic analyses showed that SELMA components, when detected, have the same origin in stramenopiles, alveolates and haptophytes whereas symbiont-specific ERAD homologs from cryptophytes seem to have been independently acquired.

The fact that SELMA components likely have a common origin in haptophytes, stramenopiles, chromerids and apicomplexa (Table 1) recalls the tree published by Burki et al. using 258 nuclear genes that pointed out a close relationship between haptophytes and the SAR (Stramenopiles-Alveolates-Rhizaria) group (Burki et al. 2012). We also found out that some nuclear-encoded proteins of the ubiquitination chain (Uba1 and Ubcs), reported as potentially derived from the red endosymbiont in cryptophytes, are in fact ERAD-like paralogs from the host. Some of these copies carry a BTS, contrary to their cytosolic counterparts, suggesting that they might be targeted to the PPC/plastid.

It is well known that, in the aftermath of endosymbioses, genes from the host can receive transit peptides and start to relocate their product to the plastid (Archibald 2015). Thus, our results suggest that at least in cryptophytes, some proteins of the translocation machinery are not derived from the ERAD proteins of the red alga but from duplicated genes of the host. Interestingly, the nuclear genome of the cryptophyte *Guillardia theta* revealed that paralogs from the host participate in the control and maintenance of the plastid (Curtis et al. 2012). Therefore, it seems quite possible that ERAD-like host paralogous genes might have also played a role in the establishment of the import system

in the second-outermost membrane of cryptophytes plastids by being duplicated and subfunctionalized during the early steps of the host-symbiont integration. This would be comparable to some subunits of the TIC/TOC complex that were also recruited from pre-existing host nuclear genes (Shi and Theg 2013).

An alternative scenario would be that the original symbiont-derived genes might have been lost and were replaced by host homologs to compensate the loss of function. In any case, it seems that the translocation machinery in cryptophytes has a hybrid nature with genes derived from the host and the red algal endosymbiont. One can argue that EGT genes, like the ones encoding SELMA proteins in haptophytes, stramenopiles, chromerids and apicomplexans, might have evolved differently than nucleomorph-encoded genes, like the SELMA components of cryptophytes. These differences might have promoted phylogenetic artifacts explaining why we do not retrieve the monophyly of these proteins in all CASH. Nevertheless, even nucleus-encoded SELMA components in cryptophytes appear to be different from their homologs in the other red lineages, supporting the idea that these proteins might have been independently acquired.

Evolutionary hypotheses invoking serial higher order endosymbioses to reconcile the monophyly of secondary red plastids and the polyphyly of the hosts posit that one CASH lineage may have been the recipient of the original secondary endosymbiosis with the red algal ancestor of the plastid (Stiller et al. 2014). It has been proposed that this lineage may have been the cryptophytes, because keeping the nucleomorph is seen as a transient state. The fact that they encode a large part of their SELMA genes in the nucleomorph has been interpreted as an additional evidence that the secondary red plastid evolved first in cryptophytes where the red-algal ERAD was converted into SELMA to allow the import of proteins from the REG lumen (Gould et al. 2015). Then, the already functional secondary red plastid would have been transmitted to other CASH lineages by serial endosymbioses followed by the transfer of SELMA genes to the nucleus of the host.

On the other hand, in a model arguing for a single red secondary endosymbiosis, followed by changes and losses, like the one proposed by Cavalier-Smith for the Chromists (Cavalier-Smith 2017), cryptophytes would have diverged very early, keeping a part of the original host/symbiont genome setup thanks to the nucleomorph. In both cases, trees of genes coding for the SELMA machinery should recover all CASH as related, monophyletic with a red algal ancestor. However, our phylogenetic analyses do not support this configuration.

In this work, we show that the SELMA translocation machinery has been assembled differently in cryptophytes compared to haptophyte+alveolates+stramenopiles, and most importantly that genes of the latter assemblage are never related to nucleomorph genes of cryptophytes. This opens an alternative and provocative hypothesis where cryptophytes and other CASH have build up their SELMA transporter independently, and maybe even at the onset of two different secondary endosymbioses with related red algae. As mentioned above, the very fast rate of molecular evolution of plastid genes, together with insufficient taxon sampling, may lead to a false phylogenetic trees where all red secondary plastids appear monophyletic. Still, one could argue that such an hypothesis is not parsimonious with respect to the assemblage of similar complex solutions to the same evolutionary problem, but many cases of comparable convergent evolution exist. Interestingly, in chlorarachniophytes, where the four membrane secondary plastid derives from a green alga, another unknown solution has been developed to import proteins across the second outermost membrane. Our main conclusion is that, if SELMA components are similar in cryptophytes and in other CASH phyla, they are probably not orthologous. This observation is crucial, because it enforces a strong revision of all evolutionary hypotheses that rely on the shared ancestry of SELMA in the CASH super-group.

REFERENCES

- Adl SM, Simpson AGBB, Farmer MA, Andersen RA, Anderson OR, Barta JR, Bowser SS, Brugerolle G, Fensome RA, Fredericq S, et al. 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J. Eukaryot. Microbiol.* 52:399–451.
- Agrawal S, Chung DWD, Potts N, van Dooren GG, Prudhomme J, Brooks CF, Rodrigues EM, Tan JC, Ferdig MT, Striepen B, et al. 2013. An Apicoplast Localized Ubiquitylation System Is Required for the Import of Nuclear-encoded Plastid Proteins. *PLoS Pathog.* 9.
- Agrawal S, van Dooren GG, Beatty WL, Striepen B. 2009. Genetic evidence that an endosymbiont-derived endoplasmic reticulum-associated protein degradation (ERAD) system functions in import of apicoplast proteins. *J. Biol. Chem.* 284:33683–33691.
- Agrawal S, Striepen B. 2010. More Membranes, more Proteins: Complex Protein Import Mechanisms into Secondary Plastids. *Protist* 161:672–687.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
- Archibald JM. 2015. Genomic perspectives on the birth and spread of plastids: Fig. 1. *Proc. Natl. Acad. Sci.* 112:10147–10153.
- Barbrook AC, Howe CJ, Purton S. 2006. Why are plastid genomes retained in non-photosynthetic organisms? *Trends Plant Sci* 11:101–108.
- Baurain D, Brinkmann H, Petersen J, Rodríguez-Ezpeleta N, Stechmann A, Demoulin V, Roger AJ, Burger G, Lang BF, Philippe H. 2010. Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Mol Biol Evol* 27:1698–1709.
- Bhaya D, Grossman A. 1991. Targeting proteins to diatom plastids involves transport through an endoplasmic reticulum. *Mol Gen Genet* 229:400–404.
- Bodył A. 2017. Did some red alga-derived plastids evolve via kleptoplastidy? A hypothesis. *Biol. Rev.*
- Bodył A, Stiller JW, Mackiewicz P. 2009. Chromalveolate plastids: direct descent or multiple endosymbioses? *Trends Ecol. Evol.* 24:119-21-2.
- Burki F, Kaplan M, Tikhonenkov D V, Zlatogursky V, Minh BQ, Radaykina L V, Smirnov A, Mylnikov P, Keeling PJ, Keeling PJ. 2016. Untangling the early diversification of eukaryotes : a phylogenomic study of the evolutionary origins of Centrohelida , Haptophyta and Cryptista.
- Burki F, Okamoto N, Pombert J-F, Keeling PJ. 2012. The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. *Proc Biol Sci* 279:2246–2254.

- Cavalier-Smith T. 1982. The origins of plastids. *Biol. J. Linn. Soc.* 17:289–306.
- Cavalier-Smith T. 1999. Principles of protein and lipid targeting in secondary symbiogenesis : euglenoid, dinoflagellate, and sporozoan plastids origins and the eukaryote family tree. *J. Eukaryot. Microbiol.* 46:347–366.
- Cavalier-Smith T. 2002. Chloroplast evolution: secondary symbiogenesis and multiple losses. *Curr Biol* 12:R62–R64.
- Cavalier-Smith T. 2017. Kingdom Chromista and its eight phyla: a new synthesis emphasising periplastid protein targeting, cytoskeletal and periplastid evolution, and ancient divergences. *Protoplasma*
- Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* 10:210.
- Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, Arias MC, Ball SG, Gile GH, Hirakawa Y, et al. 2012. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* 492:59–65.
- Deschamps P, Moreira D. 2009. Signal conflicts in the phylogeny of the primary photosynthetic eukaryotes. *Mol. Biol. Evol.* 26:2745–2753.
- Fellows JD, Cipriano MJ, Agrawal S, Striepen B. 2017. A Plastid Protein That Evolved from Ubiquitin and Is Required for Apicoplast Protein Import in *Toxoplasma gondii*. *MBio* 8:1–18.
- Felsner G, Sommer MS, Gruenheit N, Hempel F, Moog D, Zauner S, Martin W, Maier UG. 2011. ERAD components in organisms with complex red plastids suggest recruitment of a preexisting protein transport pathway for the periplastid membrane. *Genome Biol. Evol.* 3:140–150.
- Gagat P, Mackiewicz P. 2014. Protein translocons in photosynthetic organelles of *Paulinella chromatophora*. *Acta Soc. Bot. Pol.* 83:399–407.
- Gould SB, Maier U-G, Martin WF. 2015. Protein Import and the Origin of Red Complex Plastids. *Curr. Biol.* 25:R515–R521.
- Greenwood A. 1974. The Cryptophyta in relation to phylogeny and photosynthesis. In: Eighth international congress of electron microscopy, Australian Academy of Sciences.
- Hempel F, Bullmann L, Lau J, Zauner S, Maier UG. 2009. ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. *Mol. Biol. Evol.* 26:1781–1790.
- Hempel F, Felsner G, Maier UG. 2010. New mechanistic insights into pre-protein transport across the second outermost plastid membrane of diatoms. *Mol. Microbiol.* 76:793–801.

- Hibberd DJ, Norris RE. 1984. Cytology and ultrastructure of *Chlorarachnion reptans* (*Chlorarachniophyta* divisio nova, *Chlorarachniophyceae* classis nova). *J. Phycol.* 20:310–330.
- Hirakawa Y, Burki F, Keeling PJ. 2012. Genome-based reconstruction of the protein import machinery in the secondary plastid of a chlorarachniophyte alga. *Eukaryot. Cell* 11:324–333.
- Hrdá Š, Fousek J, Szabová J, Hampl V, Vlček Č. 2012. The Plastid Genome of *Eutreptiella* Provides a Window into the Process of Secondary Endosymbiosis of Plastid in Euglenids. Badger JH, editor. *PLoS One* 7:e33746.
- Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Br. Bioinform* 9:286–298.
- Keeling PJ. 2010. The endosymbiotic origin, diversification and fate of plastids. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 365:729–748.
- Kienle N, Kloepper TH, Fasshauer D. 2016. Shedding light on the expansion and diversification of the Cdc48 protein family during the rise of the eukaryotic cell. *BMC Evol. Biol.* 16:215.
- Koski LB, Golding GB. 2001. The closest BLAST hit is often not the nearest neighbor. *J Mol Evol* 52:540–542.
- Lane CE, Archibald JM. 2008. The eukaryotic tree of life: endosymbiosis takes its TOL. *trends Ecol. Evol.* 23:268–275.
- Lau JB, Stork S, Moog D, Schulz J, Maier UG. 2016. Protein-protein interactions indicate composition of a 480 kDa SELMA complex in the second outermost membrane of diatom complex plastids. *Mol. Microbiol.* 100:76–89.
- Lau JB, Stork S, Moog D, Sommer MS, Maier UG. 2015. N-terminal lysines are essential for protein translocation via a modified ERAD system in complex plastids. *Mol. Microbiol.* 96:609–620.
- Maier UG, Douglas SE, Cavalier-Smith T. 2000. The nucleomorph genomes of cryptophytes and chlorarachniophytes. *Protist* 151:103–109.
- Marin B, Nowack ECM, Melkonian M. 2005. A plastid in the making: evidence for a second primary endosymbiosis. *Protist* 156:425–432.
- Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D. 2002. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci.* 99:12246–12251.
- Martin W, Stoebe B, Goremykin V, Hapsmann S, Hasegawa M, Kowallik K V. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393:162–165.
- McFadden GI, van Dooren GG. 2004. Evolution: red algal genome affirms a common origin of all plastids. *Curr Biol* 14:R514–R516.

- Meusser B, Hirsch C, Jarosch E, Sommer T. 2005. ERAD: the long road to destruction. *Nat. Cell Biol.* 7.
- Minh BQ, Nguyen MAT, Von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30:1188–1195.
- Moog D, Stork S, Zauner S, Maier UG, Biology G, Stork S, Moog D, Stork S, Zauner S, Maier UG. 2011. In silico and in vivo investigations of proteins of a minimized eukaryotic cytoplasm. *Genome Biol. Evol.* 3:375–382.
- Moreira D, Guyader H Le, Philippe H. 2000. The origin of red algae and the evolution of chloroplasts. *Nature* 405:69–72.
- Nguyen LTL-T, Schmidt H a., von Haeseler A, Minh BQ. 2014. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32:268–274.
- Nowack ECM, Vogel H, Groth M, Grossman AR, Melkonian M, Glöckner G. 2011. Endosymbiotic gene transfer and transcriptional regulation of transferred genes in *Paulinella chromatophora*. *Mol. Biol. Evol.* 28:407–422.
- Petersen J, Ludewig A-K, Michael V, Bunk B, Jarek M, Baurain D, Brinkmann H. 2014. *Chromera velia*, endosymbioses and the rhodoplex hypothesis--plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages). *Genome Biol. Evol.* 6:666–684.
- 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, and dinoflagellates. *J Mol Evol* 62:143–157.
- Ponce-Toledo RI, Deschamps P, López-García P, Zivanovic Y, Benzerara K, Moreira D. 2017. An Early-Branching Freshwater Cyanobacterium at the Origin of Plastids. *Curr. Biol.* 27:386–391.
- Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490.
- Reyes-Prieto A, Hackett JD, Soares MB, Bonaldo MF, Bhattacharya D. 2006. Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. *Curr. Biol.* 16:2320–2325.
- Rodríguez-Ezpeleta N, Brinkmann H, Burger G, Burey SC, Lo W, Bohnert HJ, Lang BF, Roure B, Löffelhardt W, Philippe H. 2005. Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr Biol* 15:1325–1330.
- Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ. 2007. The complete chloroplast genome of the chlorarachniophyte *Bigeloviella natans*: Evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. *Mol. Biol. Evol.* 24:54–62.

- Shi LX, Theg SM. 2013. The chloroplast protein import system: From algae to trees. *Biochim. Biophys. Acta - Mol. Cell Res.* 1833:314–331.
- Sjuts I, Soll J, Bölder B. 2017. Import of Soluble Proteins into Chloroplasts and Potential Regulatory Mechanisms. *Front. Plant Sci.* 8:1–15.
- Sommer MS, Gould SB, Lehmann P, Gruber A, Przyborski JM, Maier UG. 2007. Der1-mediated preprotein import into the periplastid compartment of chromalveolates? *Mol. Biol. Evol.* 24:918–928.
- Steiner JM, Yusa F, Pompe JA, Löffelhardt W. 2005. Homologous protein import machineries in chloroplasts and cyanelles. *Plant J.* 44:646–652.
- Stiller JW, Schreiber J, Yue J, Guo H, Ding Q, Huang J. 2014. The evolution of photosynthesis in chromist algae through serial endosymbioses. *Nat. Commun.* 5:5764.
- Stork S, Lau J, Moog D, Maier UG, Article R. 2013. Three old and one new: Protein import into red algal-derived plastids surrounded by four membranes. *Protoplasma* 250:1013–1023.
- Stork S, Moog D, Przyborski JM, Wilhelmi I, Zauner S, Maier UG. 2012. Distribution of the SELMA translocon in secondary plastids of red algal origin and predicted uncoupling of ubiquitin-dependent translocation from degradation. *Eukaryot. Cell* 11:1472–1481.
- Suzuki S, Hirakawa Y, Kofuji R, Sugita M, Ishida K ichiro. 2016. Plastid genome sequences of *Gymnochlora stellata*, *Lotharella vacuolata*, and *Partenskyella glossopodia* reveal remarkable structural conservation among chlorarachniophyte species. *J. Plant Res.* 129:581–590.
- Taylor FJR. 1974. Implications and Extensions of the Serial Endosymbiosis Theory of the Origin of Eukaryotes. *Taxon* 23:229–258.
- Teich R, Zauner S, Baurain D, Brinkmann H, Petersen J. 2007. Origin and distribution of Calvin cycle fructose and sedoheptulose bisphosphatases in plantae and complex algae: a single secondary origin of complex red plastids and subsequent propagation via tertiary endosymbioses. *Protist* 158:263–276.
- Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. 2004. A Molecular Timeline for the Origin of Photosynthetic Eukaryotes. *Mol. Biol. Evol.* 21:809–818.
- Yoon HS, Nakayama T, Reyes-Prieto A, Andersen RA, Boo SM, Ishida K, Bhattacharya D. 2009. A single origin of the photosynthetic organelle in different *Paulinella* lineages. *BMC Evol. Biol.* 9:98.

7. DISCUSSION

“Y en efecto, uno de los grandes obstáculos con que luchó la teoría de la evolución, fue la oposición de 3 o 4,000 naturalistas dedicados a buscar las especies nuevas, que calificaban a veces con sus propios apellidos, y consideraban como realidades existentes y no como formas transitorias en vía de continuada evolución: por lo mismo, perdían su tradicional prestigio, disminuyendo mucho la importancia de los clasificadores y de sus inservibles colecciones de curiosidades. Otra hubiera sido la situación de la humanidad y de la ciencia si desde Lineo hubiéranse visto los seres como problemas que explicar y no como especies que clasificar”

Alfonso L. Herrera. *Nociones de Biología* (1904).

7. DISCUSSION

The evolution of oxygenic photosynthesis in an ancestor of cyanobacteria ~2.6 bya (Sánchez-Baracaldo, 2015) represents one of the most important evolutionary events in the history of life on Earth. This event changed drastically the atmosphere composition, affecting the biogeochemical cycles and the biological diversity of ecosystems (Hamilton et al., 2016; Lalonde & Konhauser, 2015). Moreover, the contribution of cyanobacteria did not finish there, they are also at the origin of photosynthetic eukaryotes through a primary endosymbiotic event between a cyanobacterium and a phagotrophic eukaryote. This event originated the supergroup Archaeplastida, an assemblage composed of glaucophytes, red algae, green algae and land plants (Rodríguez-Ezpeleta et al., 2005). After primary endosymbiosis, higher order endosymbioses (i.e. an eukaryotic cell engulfs a photosynthetic eukaryote) spread the oxygenic photosynthetic capability across the eukaryotic tree and generated a high degree of reticulation in the evolution of photosynthetic eukaryotes.

Today, the importance of endosymbiotic events in the origin of photosynthetic eukaryotes is widely accepted. Nonetheless, there are still multiple unresolved questions regarding the number and identity of the partners that originated all the diversity of plastid-bearing eukaryotes. Following the objectives we established for my PhD (section II), we profited of the recent increase in genomic data for cyanobacteria and photosynthetic eukaryotes to address some of these open questions. We obtained results with interesting implications for our understanding of the early evolution of photosynthetic eukaryotes.

7.1 Primary plastids evolved from an unicellular early-branching cyanobacterium

When the cyanobacterial origin of chloroplasts became widely accepted, a next step was to identify the type of cyanobacterial endosymbiont that gave rise to the first photosynthetic eukaryotes. Cyanobacteria are a highly diversified prokaryotic phylum with more than 4,600 described species divided in 10 orders (Guiry & Guiry, 2017), with unicellular and multicellular species living in a very broad range of environments including hot springs, hypersaline environments, deserts, oceans and freshwater habitats (Bahl et al., 2011; Garcia-Pichel et al., 1998; Six et al., 2007; Ward et al., 2012). Some cyanobacteria are able to establish symbiotic relationships with a broad range of

organisms, from protists to multicellular species (e.g. sponges and plants) (Adams et al., 2013; Carpenter, 2002). The diversity and the long evolutionary history of cyanobacteria correlate with the low similarity in gene content among distant species: only 9 to 40% of genes were shared between cyanobacteria with large genomes (e.g. *Nostoc punctiforme*) and small genomes (e.g. *Prochlorococcus marinus*), respectively, when compared to other cyanobacterial species in a study performed by Shi & Falkowski (2008), which attests of the remarkable divergent evolution within cyanobacteria.

Considering the differences among cyanobacteria, one legitimate question would be which one of the extant cyanobacterial lineages is phylogenetically closest to the plastid ancestor. The identification of the phylogenetic ascription of the cyanobacterial endosymbiont can shed light on interesting questions regarding the very poorly understood endosymbiotic event that originated the first photosynthetic eukaryotes. Was the plastid ancestor an early- or late-branching cyanobacterium? Unicellular or filamentous? Did early photosynthetic eukaryotes emerge in a marine or terrestrial environment?

As it is described in section 2.1.6, studies on plastid origins can be divided roughly in two groups depending on whether they support an early- or a late-branching endosymbiont within the phylogeny of cyanobacteria. Our phylogenomic analyses of plastid- and nuclear-encoded genes of cyanobacterial origin (EGTs) strongly supported that the early-branching cyanobacterium *Gloeomargarita lithophora* is the present-day closest relative to primary plastids. This species belongs to a recently described new cyanobacterial clade present in freshwater microbialites and microbial mats, including high-temperature ones (Ragon et al., 2014). Additionally, *G. lithophora* is an interesting example of biomineralization due to its capability to form amorphous intracellular carbonates (Couradeau et al., 2012).

Our results contradict the late-branching models that suggest that the plastid ancestor would be related to filamentous heterocyst-forming cyanobacteria (Dagan et al., 2013; Deusch et al., 2008) or unicellular diazotrophic cyanobacteria (Falcón et al., 2010). These models propose that the ability to fix nitrogen found in some cyanobacterial clades may have favored the early establishment of the primary endosymbiosis. According to these models, the cyanobacterial endosymbiont would have supplied fixed nitrogen to the host in addition to the photosynthate. However, the genome of *G. lithophora* does not encode the proteins required for nitrogen fixation which suggests that the primary endosymbiosis was not based on a nitrogen dependence. Indeed, contemporary

photosynthetic eukaryotes are unable to fix nitrogen, which makes the nitrogen fixation hypothesis highly speculative.

Plastids in Archaeplastida have a single origin deeply nested within the phylogeny of cyanobacteria. However, as shown in section 1.2, nuclear genomes and plastid proteomes of Archaeplastida species have a mosaic nature. The apparent overrepresentation of proteins similar to those of nitrogen-fixing filamentous cyanobacteria of section IV and V in the nuclear genomes of Archaeplastida led Dagan et al., (2013) to propose that the plastid ancestor was related to these clades. Nonetheless, these results are probably biased considering that these cyanobacterial lineages have much larger genomes than other cyanobacteria, reaching up to 12 Mb, whereas basal-branching unicellular cyanobacteria (section I) harbor much smaller genomes between 3-5 Mb (Shih et al., 2013). For instance, the genome of *G. lithophora* is a small one, of 3 Mb.

Other possible explanation for the mosaic nature of Archaeplastida genomes has been proposed by the shopping bag hypothesis of plastid endosymbiosis (Larkum et al., 2007). This hypothesis posits that while the final stable plastid originated from a single endosymbiont (a *Gloeomargarita*-like relative according to our phylogenomic analyses), diverse organisms (transient symbionts or not) may have transferred some genes to the host, resulting in a chimeric nature of the nucleus as well as of the proteins targeted to the plastid. Nonetheless, our phylogenomic analyses of EGTs showed, as expected, that the *G. lithophora* lineage is the biggest contributor of nuclear-encoded genes of cyanobacterial origin in Archaeplastida.

Given that our phylogenetic analyses of all type of markers (plastid-encoded genes, ribosomal RNAs, EGTs) point out that plastids evolved from a cyanobacterium related to *G. lithophora*, it is interesting to ask why a *Gloeomargarita* relative, an unicellular early-branching cyanobacterium living in freshwater environments, was the successful partner in a very rare symbiogenetic event that originated the first photosynthetic eukaryotes. As Lynn Margulis noticed, “Symbiogenesis is not ever random” (Margulis, 2010) and despite that it is not possible to have a complete answer regarding the origin of primary plastids, it may be possible to bring some elements of discussion from models of plastid endosymbiosis as well as the ecological setting where the primary endosymbiosis took place.

Firstly, the fact that the plastid ancestor was an early-branching cyanobacterium instead of a filamentous (heterocyst-forming) late-branching one is not a matter of timing of divergence within the phylogeny of cyanobacteria, that is, the primary endosymbiosis with

a *Gloeomargarita* relative 1.6-1.9 bya cannot be simply explained by the possible late evolution of filamentous clades, posterior to the endosymbiotic event that originated Archaeplastida. Evolution of multicellularity in cyanobacteria occurred relatively shortly after the emergence of cyanobacteria ~2.6 bya (Sánchez-Baracaldo, 2015). Furthermore, it has been proposed that the Great Oxygenation Event that occurred ~2.4 bya may have been favored by the evolution of multicellular cyanobacterial lineages (Schirromeister et al., 2015). Microfossils of nostocalean akinetes (i.e. resting cells formed by filamentous heterocyst-forming cyanobacteria) have been observed in stromatolites that have been dated to around 2 billion years old (Amard & Bertrand-Sarfati, 1997). Likewise, evidence of multicellularity in fossil cyanobacteria has been reported in dolomite rocks from Western Australia with similar age (circa 2000 Mya) (Knoll et al., 1988). Relaxed molecular clock phylogenetic analyses also support an early transition to multicellularity in the evolutionary history of cyanobacteria (Schirromeister et al., 2015).

In order to establish a permanent endosymbiosis between two free-living organisms, it is necessary a high degree of genetic integration between the host and the endosymbiont, which eventually will reach a turning point when the endosymbiont can be vertically transmitted. However, the size of the partners can impose physical constraints that may limit the possibilities to establish a permanent endosymbiosis. For instance, endosymbionts in foraminifera (e.g. diatoms, chlorophytes, dinoflagellates and rhodophytes) can be transmitted vertically via asexual reproduction of the host. However, due to the larger size of endosymbionts compared to the foraminiferal gametes, the endosymbionts need to be engulfed again after sexual reproduction, originating only transient endosymbioses (Nowack & Melkonian, 2010).

Therefore, is it possible that the small size of *Gloeomargarita* (less than 3 μm) may have facilitated its integration as a permanent endosymbiont to become an organelle. Interestingly, the other known example of cyanobacterial primary endosymbiosis, the chromatophore in *Paulinella* derives from a cyanobacterial ancestor closely related to the *Synechococcus/Prochlorococcus* clade, an unicellular marine group of microcyanobacteria that underwent a cell size reduction during evolution ending up with cell diameters less than 2 μm (Sánchez-Baracaldo, 2015). Endosymbiotic nitrogen-fixing cyanobacteria seem to follow the same trend. For example, the spheroid bodies which are considered to fix nitrogen within the diatom *Epithemia turgida* derive from a cyanobacterial endosymbiont closely related to the unicellular diazotrophic cyanobacterium *Cyanothece* spp., and they are obligate endosymbionts that have lost the ability to photosynthesize and are vertically

transmitted during cell division (Nakayama et al., 2014; Prechtel et al., 2004). Interestingly, bacterivorous prasinophyte algae, which possibly retained an ancestral phagotrophic capability of Archaeplastida, have a preference for small-sized bacteria (McKie-Krisberg & Sanders, 2014).

By contrast, filamentous cyanobacteria have larger cells surrounded by a continuous outer membrane forming a multicellular organism that can reach several dozens of micrometers (Sánchez-Baracaldo, 2015; Schirromeister et al., 2016). For comparison, the cell size of *Cyanophora paradoxa*, an algal species that belongs to Glaucophyta, possibly the earliest-branching lineage within Archaeplastida (see section 1.2.2.3), is 9-15 µm in length (Korshikov, 1924).

Moreover, filamentous cyanobacteria have complex regulatory mechanisms of cell communication and differentiation that may be difficult to overcome in order to create a permanent endosymbiont (Flores & Herrero, 2010). Taken together, it seems that unicellular small organisms may be better partners when it comes to create permanent endosymbionts.

Other possibility to explain the success of the ancestral *Gloeomargarita*-like cyanobacterium as primary endosymbiont might be that it would have had a genetic/genomic pre-adaptation to live in symbiosis. Ohbayashi et al. (2016) proposed that the plastid ancestor possibly had a DNA replication mechanism independent from the initiation factor DnaA which is highly conserved in prokaryotes but absent in plastids. All cyanobacterial genomes encode the *dnaA* gene but the degree of DnaA dependency to perform DNA replication varies among cyanobacterial clades (Ohbayashi et al., 2016). Therefore, a DnaA-independent mechanism may have facilitated the early reorganization of the replication machinery in the cyanobacterial endosymbiont.

7.1.1 Primary photosynthetic eukaryotes appeared in a freshwater habitat

The possible evolution of DnaA-independent replication mechanisms in cyanobacteria has been suggested to be related to their natural habitat: cyanobacteria living in environments where nutrient abundances tend to be stable, such as many oceanic regions, may be more dependent on DnaA whereas in freshwater habitats where nutrient availability can fluctuate very rapidly, different selective pressures may act favoring cyanobacteria with more flexible DNA replication mechanisms (Ohbayashi et al., 2016).

Remarkably, *Gloeomargarita* has only been found in freshwater environments (Ragon et al., 2014). An interesting perspective of future research would be to study the degree of dependency of replication on the DnaA protein in *G. lithophora*, in particular, to determine whether DNA replication can occur in DnaA-deleted mutants.

The fact that primary plastids most likely evolved in a freshwater environment constitutes an interesting change of paradigm as it is often assumed that the origin of plastids took place in a marine habitat, probably because the ocean is usually seen as the environmental setting where all major evolutionary events have occurred such as the origin of life or the origin of animals (Rota-Stabelli et al., 2013; Sojo et al., 2016). However, oxygenic photosynthesis, both bacterial and eukaryotic, likely appeared on land. The ancestral state reconstruction of morphology and habitat showed that cyanobacterial clades evolved from an unicellular freshwater cyanobacterium (Uyeda et al., 2016). These results are also supported by the fact that basal cyanobacterial lineages such as the thylakoid-less *Gloeobacter*, are restricted to terrestrial environments (Mareš et al., 2013). Additionally, stochastic mapping of trait evolution supports a freshwater origin of primary plastids and has also confirmed our results that *Gloeomargarita* is the present-day cyanobacterial relative closest to the plastid lineage (Sánchez-Baracaldo et al., 2017).

7.2 Evolution of secondary plastid-harboring eukaryotes

Several nuclear-encoded plastid-targeted proteins of red algal origin have been detected in secondary algae with green plastids (i.e. chlorarachniophytes and euglenids) (Archibald et al., 2003; Yang et al., 2011). For instance, the majority of genes of the Calvin-Benson cycle in euglenids seem to derive from red algae instead of the green algal endosymbiont as could be expected (Markunas & Triemer, 2016). Likewise, the tetrapyrrole biosynthetic pathway has been almost completely transferred from a red algal source in chlorarachniophytes and euglenids (Cihlář et al., 2016; Lakey & Triemer, 2016). On the other hand, the CASH lineage (cryptophytes, alveolates, stramenopiles and haptophytes) appears to have acquired multiple genes from green algae. The nuclear genomes of diatoms and other red plastid-harboring groups seem to encode many genes from green algae which has been interpreted as remnants from a possible former secondary endosymbiosis with a green alga that was replaced by a red one (Allen et al., 2012; Frommolt et al., 2008; Moustafa et al., 2009). Likewise, the green origin of some nuclear-encoded proteins of *Chromera velia* and other alveolates opens the possibility of a

cryptic green algal endosymbiosis (Woehle et al., 2011). However, these results have been challenged by subsequent re-analyses of the same data (see below).

In order to gain a deeper understanding of the apparent conflict between the type of present-day secondary plastids and the apparent origin of some of the nuclear-encoded genes of algal origin, we focused on the phylogenetic analysis of secondary EGTs, that is, the subset of nuclear-encoded genes that through the inspection of phylogenetic trees for which it is possible to trace a complete history of endosymbiotic gene transfers: first, from the cyanobacterial endosymbiont to the ancestor of Archaeplastida via primary endosymbiosis; and secondly, from the green or red algal endosymbiont to other eukaryotic hosts through secondary EGT. Detection of secondary EGTs is useful to identify cryptic endosymbiotic events. For instance, analyses of transcriptomic sequences data revealed that the nuclear genome of the parasitic dinoflagellate *Hematodinium* encodes some EGTs from the former red plastid that was secondarily lost (Gornik et al., 2015). Similar remaining secondary EGTs from previous endosymbioses have been observed in other dinoflagellate species such as *Karenia brevis* and *Durinskia baltica* where the former peridinin-containing plastids have been replaced by an haptophyte-derived and a diatom-derived plastid, respectively (Burki et al., 2014; Waller & Kořený, 2017).

This type of secondary EGTs with a complete transmission history are particularly helpful because of the large phylogenetic distance between prokaryotes (cyanobacteria) and photosynthetic eukaryotes makes relatively easy the identification of tree topologies compatible with an EGT scenario and distinguishing these genes from those vertically inherited. However, it is important to consider that, although the phylogenetic tree of a gene may show a topology compatible with an EGT scenario, this does not necessarily imply that the gene was transferred in an endosymbiotic context. It is also possible that the gene was acquired horizontally by punctual HGT from red or green algae, or even that the gene was already part of the endosymbiont genome by more ancient HGTs.

After a strict protocol of identification of genes compatible with an EGT/HGT scenario in photosynthetic eukaryotes with secondary plastids, we recovered a final set of 85 genes. Regarding the genes in the CASH lineage, the vast majority (90%) had a “red” origin, in agreement with their current red algal endosymbionts. The rest of the genes (10%) have ambiguous phylogenies but some genes seem to have been transferred from green algae. Our results do not support previous studies that reported a great number of green genes in diatoms, suggesting a cryptic green algal endosymbiosis (Moustafa et al., 2009). Stramenopiles, including diatoms, had the same proportion of green algal-derived

genes as other lineages with red plastids. The scarce number of green genes that we found in CASH lineages is in line with the results of Deschamps & Moreira, (2012), who reanalyzed the set of genes allegedly acquired from a cryptic green endosymbiont in diatoms and found that the vast majority of genes were misidentified as having a green origin due to a combination of poor taxon sampling and the use of an oversimplistic automated software to identify the phylogenetic origin of proteins. Therefore, given the small proportion of green algal-derived genes in CASH lineages in our analyses, it seems that there is no compelling evidence for a former secondary endosymbiosis with a green alga. The few green algal genes found in these lineages may be the result of classical horizontal transfers.

By contrast, regarding the origin of secondary EGTs in chlorarachniophytes and euglenids, we expected a vast majority of genes of green algal origin as a result of the endosymbiotic gene transfer from their green endosymbionts (similar to the red algal genes in the CASH lineages). However, we found 58% and 45% of red algal genes in chlorarachniophytes and euglenids, respectively. This represents a striking difference when compared to the ~10% of green algal genes in secondary red lineages. There are two ways to explain the high amount of red algal genes in both secondary green lineages: either there was a huge amount of transferred genes from red algae via HGT that created a high degree of chimerism in the plastid proteome of chlorarachniophytes and euglenids or these lineages may have had a former endosymbiosis with a red alga. Our results do not particularly favor any evolutionary scenario but show that chlorarachniophytes and euglenids have independently recruited a large amount of plastid-targeted red algal genes that participate in varied and important functions such as plastid translation (e.g. ribosomal proteins and aminoacyl-tRNA ligases), biosynthetic pathways (e.g. chlorophyll and fatty acid biosynthesis) and photosynthesis-related functions (e.g. light harvesting and photosystem assembly).

The way to interpret the red algal contribution depends on which event is considered to have predominated in the evolution of chlorarachniophytes and euglenids: HGT or EGT. According to Martin (2017), HGT from prokaryotes to eukaryotes as well as among eukaryotes has been overestimated. He posits that although HGT in eukaryotes can occur, it is mostly a rare event. Thus, the vast majority of acquisitions of foreign genes in eukaryotes had occurred in endosymbiotic contexts. However, these claims are not widely accepted and the relevance of HGT in eukaryotic evolution is an ongoing controversial debate (Husnik & McCutcheon, 2017; Keeling, 2009; Koutsovoulos et al.,

2016; Martin, 2017). Nonetheless, if the red algal genes in euglenids and chlorarachniophytes were acquired horizontally and there has been massive acquisitions of foreign genes during their evolution, perhaps due to a mixotrophic lifestyle (Calderon-Saenz & Schnetter, 1989), it would be expected that some genes had dissimilar origins, while some species would have kept the original green algal-derived gene, others could have replaced it later by a red algal one. However, all the red algal genes that we identified are shared by all species within the euglenids or the chlorarachniophytes, implying that if the gene was transferred via HGT, in all cases the transfer was ancient and already present in the last common ancestors of these groups, and that HGT stopped in both groups -by unknown reasons- so that we do not observe any recent case of gene acquisition. This seems clear in chlorarachniophytes, with a rich taxon sampling already available, but requires confirmation in euglenids, for which we only have two transcriptome sequences available.

The alternative possibility to explain the large amount of red algal genes in the secondary green lineages is that previous to the secondary endosymbiosis with green algae, these lineages might have had red plastids that were subsequently replaced by green ones. Although the cryptic plastid hypothesis may be suitable for both chlorarachniophytes and euglenids, there could be two different scenarios: 1) chlorarachniophytes and euglenids both had a former red plastid; 2) only one of these lineages established a red secondary endosymbiosis. Although the red algal genes in chlorarachniophytes and euglenids could have been acquired in different ways, under the cryptic red plastid hypothesis a similar reasoning can be used to explain the huge impact of plastid-targeted red algal-derived genes in these lineages.

Chlorarachniophytes belong to the phylum Cercozoa within the supergroup Rhizaria, a highly diverse group that comprises photosynthetic organisms such as chlorarachniophytes and the primary plastid-bearing *Paulinella* as well as heterotrophic lineages such as radiolarians, foraminiferans and many cercozoans (Cavalier-Smith & Chao, 2003). Importantly, Rhizaria is part of a monophyletic supergroup together with Stramenopiles and Alveolata forming the SAR clade (Burki et al., 2007). Interestingly, stramenopiles and alveolates both have photosynthetic representatives endowed with secondary red plastids.

7.2.1 Was the rhizarian ancestor photosynthetic?

Although the great diversity of Rhizaria and the important contribution to the planktonic biomass, rhizarian protists have received little attention in genomic and transcriptomic analyses until recently (Sierra et al., 2016). Only three complete nuclear genomes have been sequenced so far for: the chlorarachniophyte *Bigelowiella natans* (Curtis et al., 2012), the foraminiferan *Reticulomyxa filosa* (Glöckner et al., 2014) and the plant pathogen *Plasmodiophora brassicae* (Rolfe et al., 2016). The scarcity of genomic data hinders the research of a putative photosynthetic past of Rhizaria. Nonetheless, the genome of *Reticulomyxa filosa* does not encode photosynthesis-related genes, which was interpreted as evidence for the lack of photosynthetic capability in the ancestor of Rhizaria (Glöckner et al., 2014). Similarly, the genome of the rhizarian pathogen *Plasmodiophora*, which has been reduced due to the parasitic lifestyle, does not show any hint of a photosynthetic past (Rolfe et al., 2016). These results together with the very large number of independent plastid losses necessary to explain the vast diversity of heterotrophic lineages within Rhizaria argue against a photosynthetic past of Rhizaria. Nevertheless, there is still a possibility that a red algal plastid was ancestral to SAR and had been kept until shortly before the secondary endosymbiosis with a green alga (Fig. 19A).

Although the branching order of rhizarian clades remains unclear (being deeper nodes more difficult to resolve) (Howe et al., 2011; Pawlowski & Burki, 2009; Sierra et al., 2016), other possibility under the cryptic endosymbiosis hypothesis is that the endosymbiosis took place after the diversification of Rhizaria between a phagotrophic cercozoan and a red plastid-containing endosymbiont, such as a photosynthetic alveolate or ochrophyte (Fig. 19B).

Molecular clocks, chemical biomarkers and fossil evidence support that the SAR clade diversified during the Neoproterozoic (1000- 541 Mya) and red plastid-harboring lineages within stramenopiles and alveolates appeared in this era (Brocks et al., 2017; Butterfield, 2004; Meng et al., 2005; Parfrey et al., 2011) which is temporally congruent for a “red” period also within a putative cercozoan lineage, followed by the loss of the red plastid and a more recent endosymbiotic event between the chlorarachniophyte ancestor and a green alga some 100-150 Mya (Delaye et al., 2016). Importantly, if red algal genes come from this putative alveolate or ochrophyte endosymbiont, it would be expected that phylogenetic trees of these red algal genes show chlorarachniophytes nested within one of these two groups. However, given the poor resolution frequently found in single-gene trees, it can be difficult to distinguish between this result and a tree of a vertically inherited

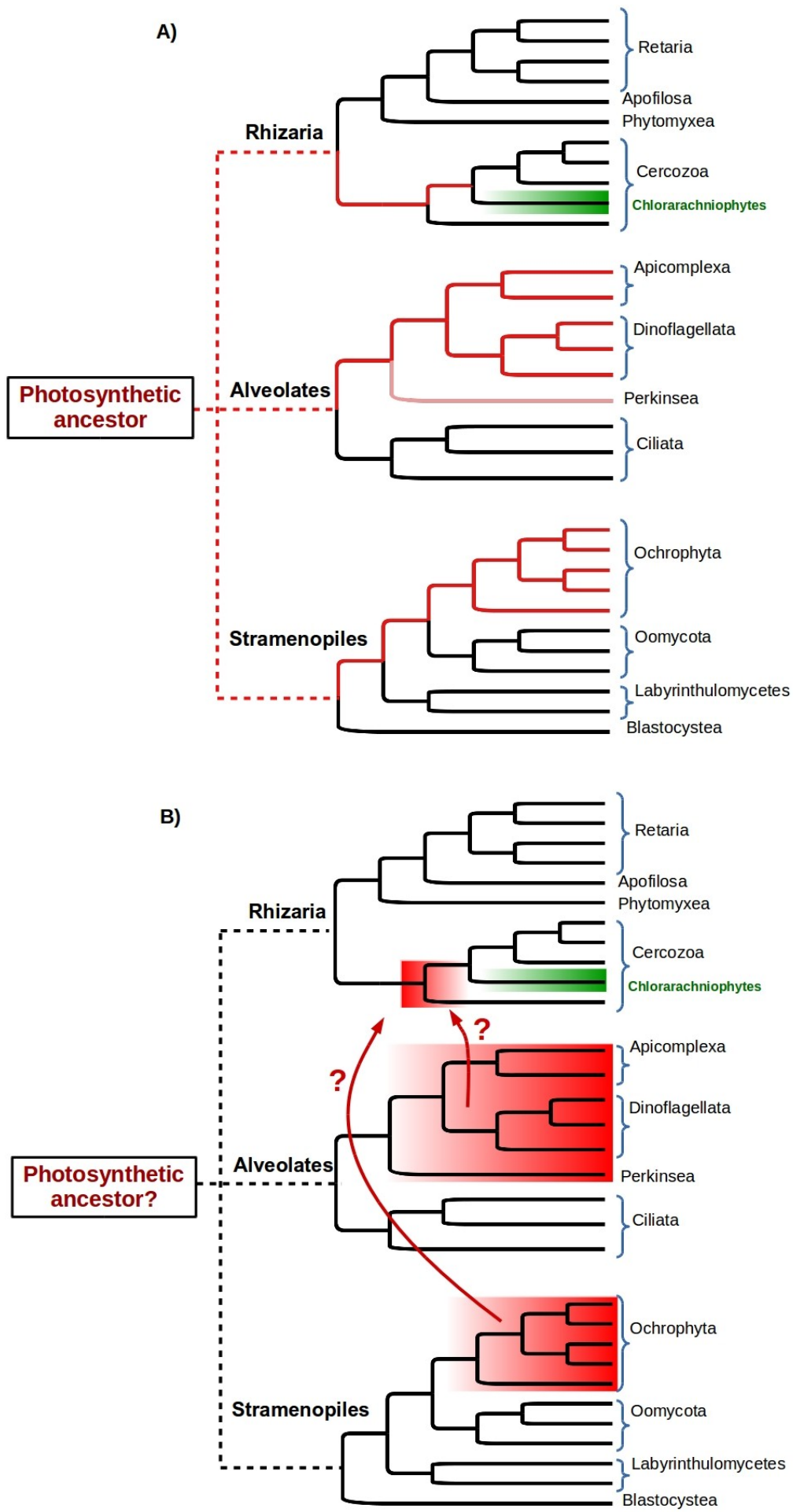


Figure 19. Cryptic red plastid in chlorarachniophytes. A) The last common ancestor of the SAR clade harbored a secondary red plastid. B) An early cercozoan engulfed a red plastid-bearing eukaryote.

gene due to the phylogenetic proximity of the different lineages within the SAR clade. Thus, if the putative red plastid was not ancestral in Rhizaria but particular to a cercozoan lineage that gave rise to the chlorarachniophytes, this may explain the lack of photosynthesis-related genes in other rhizarian lineages.

Phylogenetic analysis of the 18S rDNA shows that Cercozoa is a highly diversified group with deep-branching clades that have been widely overlooked (Howe et al., 2011). To test the hypothesis of a cryptic red algal endosymbiosis, it would be necessary to have a better understanding of cercozoan evolution and generate genomic data for deep-branching cercozoan species that could shed some light on the origin of the red algal genes in chlorarachniophytes. Nonetheless, it is possible that all plastid-derived genes of the putative photosynthetic rhizarian/cercozoan ancestor may have been lost in present-day heterotrophic lineages. Interestingly, massive loss of symbiont-derived genes seems to be very common in secondarily heterotrophic organisms (Huang et al., 2004; Smith & Lee, 2014).

Although the cryptic endosymbiosis hypothesis is difficult to test in chlorarachniophytes and euglenids, we know that this type of plastid replacement can occur in nature. The dinoflagellate *Lepidodinium chlorophorum* represents an interesting case of a red plastid that was replaced by a green one. This species currently possesses a green algal plastid but its genome harbors a mix of red and green algal genes that were recruited from the former red plastid and the present-day green algal endosymbiont, respectively (Sastre et al., 2010). Instead of activating new genes and retarget new plastid proteins from the green endosymbiont, it is possible that under a context of serial endosymbiotic events, the favored evolutionary strategy may be the recycling of remaining EGTs. Thus, the red algal genes in *Lepidodinium* akin to the red genes in chlorarachniophytes and euglenids might have facilitated the acquisition of the new green plastids.

7.2.2 Evolutionary history of the SELMA translocon

Similar to primary plastids in which the monophyly of the multiprotein complex involved in the translocation of proteins across their double membrane, the TIC/TOC complex, strongly supports a single origin of this protein import system (Shi & Theg, 2013), the apparent monophyly of SELMA (i.e. the protein import machinery used to cross the second outermost membrane of secondary red plastids with four membranes, see section 1.3.2.1) would suggest that there was a single red algal secondary endosymbiotic event. In

this scenario (the Chromalveolate hypothesis), SELMA would have evolved vertically in the different secondary red lineages. In the alternative scenario involving higher order endosymbioses spreading secondary red plastids across different lineages, SELMA would have been transferred concomitantly with the plastids as it was the case for the TIC/TOC complex during secondary endosymbioses (Hirakawa et al., 2012; Sheiner & Striepen, 2013). Interestingly, although the protein import system used in chlorarachniophytes to cross the second outermost membrane remains unknown, it seems that it does not derive from the ERAD machinery of its green algal endosymbiont, suggesting that the SELMA complex is an evolutionary innovation of red algal-derived plastids (Hirakawa et al., 2012).

Several putative SELMA proteins have been detected using bioinformatic analyses (Sommer et al., 2007; Stork et al., 2012), however, very few phylogenetic reconstructions of these proteins have been published. The monophyly of this system has been mostly based on phylogenetic analyses of the protein Cdc48, an ATPase considered to translocate proteins in a ubiquitin-dependent manner (Bolte et al., 2011; Felsner et al., 2011; Petersen et al., 2014). On the contrary, phylogenetic trees of derlin proteins have provided unclear results (Petersen et al., 2014). Consequently, strict phylogenetic evidence for the monophyly of SELMA, a system in which very strong evolutionary conclusions are regularly taken for granted, is still missing.

To study the origin and evolution of SELMA, we performed phylogenetic analyses of the multiple SELMA components. Our trees show that the translocation machinery in cryptophytes (in which many SELMA proteins remain encoded in the nucleomorph) differs considerably from that established in the other red lineages (i.e. alveolates, stramenopiles and haptophytes) due to a combination of differentially recruited ERAD paralogs from the red algal endosymbiont and a possible contribution of host genes. The derlin proteins provide a good example.

The ERAD system has commonly two paralogs: derlin1-1 and derlin1-2. Following the nomenclature proposed by Cavalier-Smith (2017) to avoid confusion, derlin1-1 and derlin1-2 will be called hereafter derlin A and derlin B, respectively. This author recently showed that cryptophytes recruited the nucleomorph-encoded derlin A while haptophytes, stramenopiles and alveolates use the paralogous derlin B, which possibly experienced an early duplication in the chromist ancestor generating two paralogs, derlin B1 and derlin B2, which would be targeted to the periplastidial compartment. Likewise, our results also support a differential recruitment of derlin paralogs into the SELMA complex. Interestingly, although Cavalier-Smith proposed that alveolates lost one subparalogous derlin B protein,

we identified the two derlin B1 and derlin B2 in alveolates which have escaped former detection.

As previously reported, the phylogenetic analysis of Cdc48 shows that this protein derives from the ERAD machinery of the red alga at the origin of secondary red plastids. Nonetheless, this protein seems to have a more complex history than usually assumed. It seems that it has undergone lineage-specific duplications and losses that could lead to a hidden paralogy scenario. Cryptophytes, haptophytes and stramenopiles have two copies of symbiont-specific Cdc48 while alveolates only retained one. In cryptophytes one protein is still encoded in the nucleomorph whereas the other is encoded in the nuclear genome. If the two copies present in cryptophytes, haptophytes and stramenopiles had resulted from an early duplication, similar to derlin B1 and derlin B2, we would expect two differentiated paralogs but our results showed that several Cdc48 proteins seem to correspond to late-branching paralogs suggesting that the two copies present in haptophytes and stramenopiles may have been duplicated after diversification of the chromist ancestor probably from the nuclear-encoded *cdc48* gene while the nucleomorph copy was lost.

In the case of the proteins that participate in the ubiquitination pathway in SELMA (necessary to import proteins into the periplastidial compartment), we found a similar pattern: cryptophytes use different ubiquitin activating enzymes (E1), ubiquitin conjugating enzymes (E2) and ubiquitin ligases (E3) when compared to those in haptophytes, alveolates and stramenopiles. Interestingly, some of these proteins remain encoded in the nucleomorph. It is noteworthy that some CASH lineages may have recruited some genes from the host: the ubiquitin activating enzyme in cryptophytes is likely a paralogous gene from the host and at least one E3-ubiquitin ligase that has been localized in the apicoplast as part of the ubiquitination machinery (Agrawal et al., 2013), derives from the host in apicomplexans, suggesting a hybrid nature of the protein import machinery. Interestingly, the TIC/TOC complex in primary plastids was also built from a set of cyanobacterial and eukaryotic proteins (Shi & Theg, 2013).

As pointed out previously by Cavalier-Smith (2017), the fact that the symbiont-specific derlin in alveolates, haptophytes and stramenopiles had evolved from the derlin B paralog whereas cryptophytes use the nucleomorph-encoded derlin A paralog is difficult to reconcile with the hypothesis of multiple putative tertiary and quaternary rounds of endosymbiosis where a cryptophyte symbiont would be the donor of the SELMA translocon to one or more heterotrophic hosts in alveolates, stramenopiles and haptophytes. Likewise, our phylogenetic analyses of derlin proteins as well as for almost

all other SELMA components show that the translocation machinery in cryptophytes does not use the same set of proteins as these other lineages. Therefore, if we assume that there were an undetermined number of serial endosymbioses that would have initiated with a cryptophyte endosymbiont that participated in a tertiary endosymbiotic event, it is necessary to postulate that at the time of that putative endosymbiosis, the SELMA translocon in cryptophytes was more complex and had more symbiont-derived ERAD-like proteins than the SELMA in present-day cryptophytes. Moreover, during the tertiary endosymbiosis, the new would-be red lineage would have recruited a subset of SELMA components from the initial cryptophyte pool of SELMA genes. This subset would be different from the one retained in contemporary cryptophytes. This would imply that SELMA experienced a secondary simplification in cryptophytes during evolution. This evolutionary scenario where differential loss and retention of proteins took place during tertiary symbiogenetic events seems very unparsimonious.

Interestingly, there is an example of endosymbiotic transfer of the SELMA translocon in dinoflagellates of the family Kareniaceae. It is well established that members of this family replaced the ancestral peridinin-containing plastid by an haptophyte-derived one via tertiary endosymbiosis (Gabrielsen et al., 2011). Recent studies showed that these dinoflagellates encode bipartite-targeting sequences with clear haptophyte origin for all SELMA components (Kořený et al., unpublished, cited by Waller & Kořený, 2017). This is a remarkable case showing evidence that SELMA can be endosymbiotically transferred but, as mentioned before, it is not completely surprising considering that the TIC/TOC import machinery has been transferred in every secondary and tertiary plastid endosymbiosis. However, in this example, the Kareniaceae SELMA proteins are clearly related to those of haptophytes, as expected, which seems not to be the case in haptophytes, stramenopiles and alveolates if their red plastids were actually derived from a cryptophyte endosymbiont.

Hence, according to the origin of SELMA components in cryptophytes and the other red lineages, the endosymbiotic transfer of SELMA from a putative cryptophyte endosymbiont seems unlikely. One other possibility is that a more complex SELMA system was present in the last common ancestor of all chlorophyll *c*-containing algae and it was probably composed of more proteins (Fig. 20), but would have been simplified during evolution (loss of paralogous proteins in some cases). SELMA may have been restructured perhaps with retargeting of host-derived proteins in some lineages. This model speculates a differential construction of the system from the same initial set of proteins which is highly dubious (similar to the cryptophyte endosymbiont hypothesis).

Nonetheless, the loss of the nucleomorph and SELMA-related nucleomorph-encoded genes that did not make their way into the nuclear genome may have exerted strong selective pressures to restructure the system from the remaining set of genes.

Although our results do not particularly favor any scenario, we showed that SELMA has a much more complex evolutionary history than initially thought and the system needs to be thoroughly studied in future research.

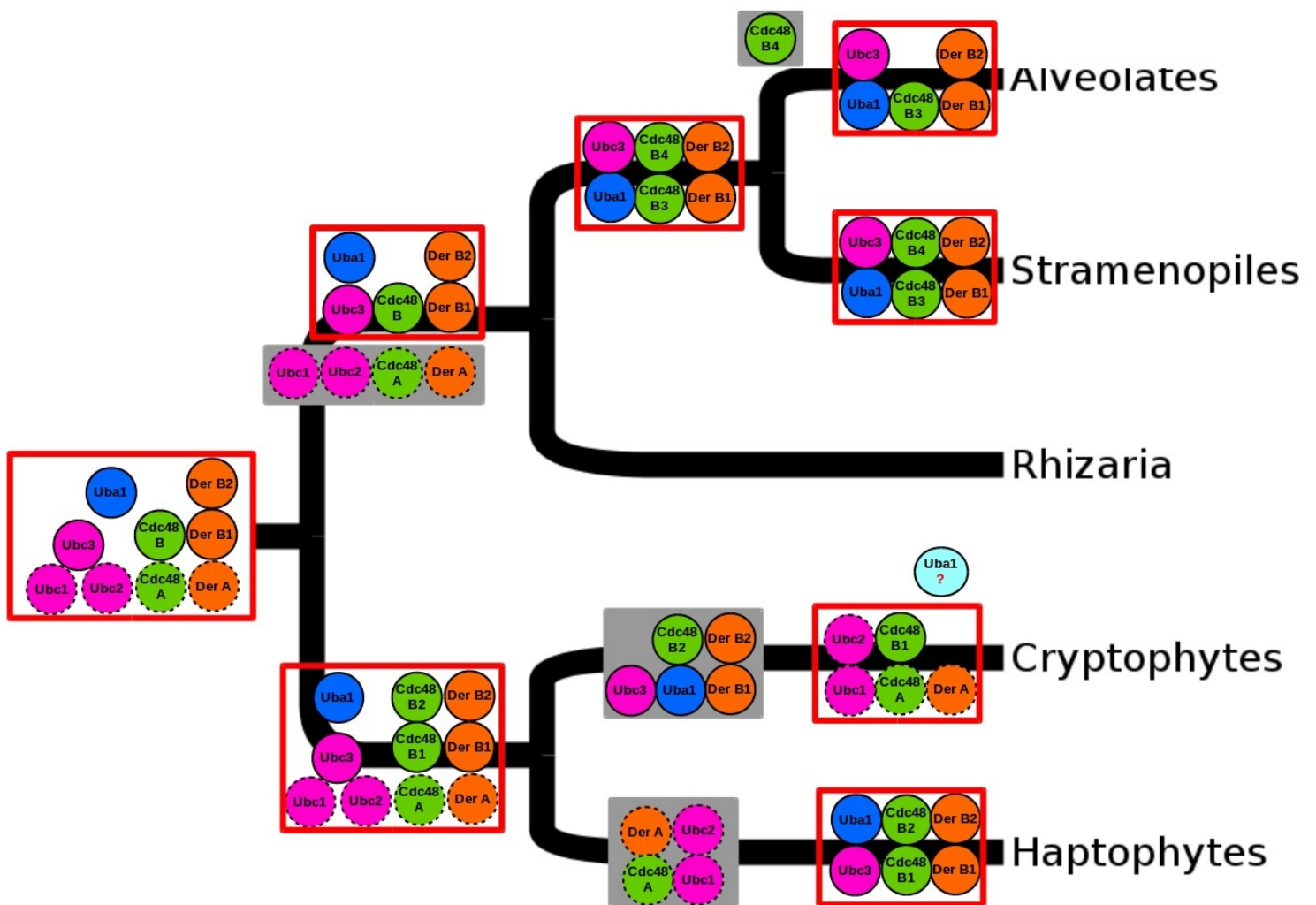


Figure 20. Putative distribution of SELMA components from the chromist ancestor to present-day secondary red lineages. Proteins within red rectangles derive from the ERAD system of the red algal endosymbiont. Proteins within gray rectangles represent proteins that were lost. Dash line circles are nucleomorph-encoded proteins whereas continuous line circles represent proteins encoded in the nuclear genome.

7.3 Perspectives

The results of my PhD have opened interesting questions regarding the evolution of photosynthetic eukaryotes that would be worth considering for future research:

- **DNA replication mechanism of *Gloeomargarita lithophora*.** It has been proposed that a DNA replication mechanism independent from the replication initiator protein DnaA might have facilitated the reorganization of the replication machinery facilitating primary endosymbiosis (Ohbayashi et al., 2016). Some cyanobacteria can replicate their DNA without the initiation factor, showing that the degree of dependency strongly varies among cyanobacterial clades. However, it is not known whether DNA replication can occur without the DnaA protein in *G. lithophora*. Hence, a next step in the study of primary endosymbiosis would be to evaluate the relevance of the DnaA gene in *G. lithophora*, the free-living cyanobacterium closest to primary plastids.
- **Branching order within Archaeplastida:** Although our phylogenomic analyses of plastid and EGT genes support that Glaucophyta was the earliest branching lineage of Archaeplastida and these results are also supported by the ancestral retention of carboxysomes and the peptidoglycan wall of cyanelles (Jackson et al., 2015), the branching order is still a matter of debate (Mackiewicz & Gagat, 2014). To address this problem, one possibility would be to sequence more glaucophyte genomes considering that only the plastid genome of *Cyanophora paradoxa* has been completely sequenced so far (Stirewalt et al., 1995), and perform phylogenomic analyses of plastid genes with a rich taxon sampling, including basal clades from the different Archaeplastida lineages, such as the basal green alga *Verdigellas peltata* (Leliaert et al., 2016).
- **Ancestral gene repertoire of primary plastids.** It would be interesting to reconstruct the probable ancestral gene repertoire of primary plastids through a comparative analysis of plastid and EGT genes among Archaeplastida lineages, in order to understand the process of primary endosymbiosis and the genome reorganization underwent by the cyanobacterial endosymbiont: What happened during the genome reduction of the cyanobacterial endosymbiont? Which genes

were preferentially retained or replaced? How were the plastid-located pathways affected during primary endosymbiosis? Is there any functional pattern between the genes that were transferred to the nucleus and the genes that continue to be encoded in the plastid?

- **Genomic study of basal Cercozoa:** Our results showed that chlorarachniophytes and euglenids encode several genes of red algal origin despite that both photosynthetic lineages carry secondary green plastids. In order to study the origin of these red algal genes, it is necessary to have a better understanding of the genomic evolution and gene content of closer heterotrophic lineages of chlorarachniophytes and euglenids. This may shed some light of the role of eukaryovory in the mosaicism observed in nuclear genomes and plastid proteomes of these green lineages. For instance, the deep-branching cercozoan *Micrometopion nutans* may be a suitable candidate species for further study. It is an eukaryote-eating flagellate deeply nested within the phylogeny of Cercozoa that seem to have diverged earlier than chlorarachniophytes (Howe et al., 2011).
- **Mechanistic model of the protein import in secondary red plastids.** Despite that a model of the protein import into the periplastidial compartment (PPC) mediated by the SELMA translocon has been proposed (mostly developed in diatoms), and has been widely accepted (Bolte et al., 2011; Hempel et al., 2009, 2010, Stork et al., 2012, 2013), we have shown that the evolutionary history of SELMA is more complex than generally assumed and the system needs to be studied carefully. An alternative model has been proposed recently by Cavalier-Smith, (2017) in which SELMA may be localized in the periplastid reticulum within the PPC instead of the second outermost membrane. This model would change previous assumptions and need to be considered in future studies. Thus, it is needed a better microscopic description of the PPC of secondary red plastids as well as of the mechanistic model of protein import. It is necessary to understand from a global perspective the similarities and differences of the PPC among secondary red plastid-harboring lineages, the structure of the SELMA machinery and the origin of their components.

- **To Improve the phylogenetic tree of eukaryotes.** To study the evolution of secondary red plastids it is necessary to have a very good understanding of the phylogenetic tree of eukaryotes. Evolutionary hypotheses regarding the evolution of secondary red plastids are heavily influenced by the resolution of the eukaryotic tree, specially the phylogenetic relationships among plastid-harboring lineages. The likelihood that all secondary red plastids arose by vertical descent from a chromist ancestor as proposed by the chromalveolate hypothesis or by serial endosymbioses as some models argue (Petersen et al., 2014; Stiller et al., 2014) depends on the logic behind the phylogenetic relationships among photosynthetic groups. Likewise, the number of probable plastid losses and the timing of morphological evolution of certain characteristics depend on the branching order suggested by the phylogenetic trees. Hence, to reconcile the evolution of secondary red plastids with the evolution of their hosts is mandatory to improve the resolution of the phylogenetic tree of eukaryotes.

8. CONCLUSIONS

“The knowledge of the laws of nature, whether we can trace them in the alternate ebb and flow of the ocean, in the measured path of comets, or in the mutual attractions of multiple stars, alike increases our sense of the calm of nature, whilst the chimera so long cherished by the human mind in its early and intuitive contemplations, the belief in a ‘discord of the elements’, seems gradually to vanish in proportion as science extends her empire”

Alexander von Humboldt. *Cosmos: a sketch of a physical description of the universe*
(1849)

8. CONCLUSIONS

My PhD project centered on the study of the early evolution of photosynthetic eukaryotes, from the primary endosymbiosis that gave rise to the first group of plastid-bearing eukaryotes up to the acquisition of secondary plastids. They together originated all the extant diversity of eukaryotic organisms capable to photosynthesize. As I showed all along this work, there are still several doubts concerning the origin and evolution of photosynthetic eukaryotes. My PhD project addressed some of these questions. The major outcomes of our work are listed below:

- 1. Primary plastids are closely related to the unicellular early-branching *Gloeomargarita* clade.** Our phylogenomic analyses of plastid and nuclear-encoded genes of cyanobacterial origin of Archaeplastida lineages clearly showed that primary plastids have a deep origin within the phylogeny of cyanobacteria. The plastid ancestor originated from a relative closely related to the recently described cyanobacterium *Gloeomargarita lithophora*. Our results contradict other studies in which primary plastids would have originated from a late-branching filamentous nitrogen-fixing cyanobacterium.
- 2. Primary plastids likely arose in a freshwater habitat.** Our results coupled with the ecological distribution of *Gloeomargarita* strongly support that primary photosynthetic eukaryotes appeared in a freshwater habitat.
- 3. The plastid ancestor was not capable of fixing atmospheric nitrogen.** Primary endosymbiosis was not driven by the dependency of the host on fixed nitrogen supplied by the cyanobacterial endosymbiont, as shown by the lack of genes necessary to fix nitrogen in *G. lithophora*.
- 4. Plastid proteomes of chlorarachniophytes and euglenids have recruited several proteins from red algae.** We found that approximately half of nuclear-encoded plastid-targeted proteins in chlorarachniophytes and euglenids have been acquired from a red algal source. These results leave open the possibility that these green lineages may have carried a red plastid that was replaced by a green one.

5. No evidence for a cryptic green plastid endosymbiosis in chromalveolates. Our results did not support previous claims that chromalveolates encode a large number of genes that would have been transferred from green algae. We argue that the low percentage (10%) of plastid-targeted proteins with a green algal ancestry may be better explained by horizontal gene transfer events.

6. The SELMA translocon has a more complex evolutionary history than previously thought. Our phylogenetic analyses of SELMA components showed that most of these proteins are not monophyletic. The SELMA translocon in cryptophytes is formed from a different set of red algal-derived proteins than that in the other red lineages (i.e. haptophytes, alveolates and stramenopiles).

7. SELMA has seemingly a hybrid nature. Besides the recycling of genes from the ERAD machinery of the red algal endosymbiont, it seems that some red lineages appear to have recruited genes from the host to be part of the SELMA translocation machinery.

1. SUPPLEMENTARY MATERIAL

9.1 Supplementary material of manuscript 1

An Early-Branching Freshwater Cyanobacterium at the Origin of Plastids

(*Current Biology* 27:386–391)

Rafael I. Ponce-Toledo,¹ Philippe Deschamps,¹ Purificación López-García,¹ Yvan Zivanovic,² Karim Benzerara,³ and David Moreira^{1,4,*}

¹Unité d'Ecologie, Systématique et Evolution, CNRS UMR 8079, Université Paris-Sud/Paris-Saclay, AgroParisTech, 91400 Orsay, France.

²Institut de Génétique et Microbiologie, CNRS UMR 8621, Université Paris-Sud/Paris-Saclay, 91405 Orsay, France.

³Institut de Minéralogie, de Physique des Matériaux, et de Cosmochimie, Sorbonne Universités, UPMC Université Paris 06, CNRS UMR 7590, Muséum National d'Histoire Naturelle, IRD UMR 206, 75005 Paris, France.

⁴Lead Contact

DATA S1. Proteins used in the phylogenomic analyses		
Protein name	Reference sequence	Location
allophycocyanin alpha subunit	NP_043159.1	Chloroplast-encoded
allophycocyanin gamma subunit	NP_043184.1	Chloroplast-encoded
allophycocyanin beta 18 subunit	NP_043276.1	Chloroplast-encoded
allophycocyanin beta subunit	NP_043158.1	Chloroplast-encoded
anthranilate synthase component 2	NP_043265.1	Chloroplast-encoded
ATP synthase CF0 B chain	NP_043224.1	Chloroplast-encoded
ATP synthase CF0 B' subunit	NP_043225.1	Chloroplast-encoded
ATP synthase CF0 C subunit	NP_043226.1	Chloroplast-encoded
ATP synthase CF1 alpha subunit	NP_043222.1	Chloroplast-encoded
ATP synthase CF1 beta subunit	NP_043241.1	Chloroplast-encoded
ATP synthase CF1 delta subunit	NP_043223.1	Chloroplast-encoded
ATP synthase CF1 epsilon subunit	NP_043242.1	Chloroplast-encoded
chaperonin GroEL	NP_043141.1	Chloroplast-encoded
cytochrome b6	NP_043175.1	Chloroplast-encoded
cytochrome b6/f complex subunit IV	NP_043174.1	Chloroplast-encoded
cytochrome c biogenesis protein	NP_043267.1	Chloroplast-encoded
cytochrome f	NP_043244.1	Chloroplast-encoded
elongation factor Tu	NP_043207.1	Chloroplast-encoded
glutamine amidotransferase	NP_043221.1	Chloroplast-encoded
glutamyl-tRNA reductase	NP_043163.1	Chloroplast-encoded
heat shock protein 70	NP_043140.2	Chloroplast-encoded
hypothetical protein CypaCp002	NP_043139.1	Chloroplast-encoded
hypothetical protein CypaCp007	NP_043144.1	Chloroplast-encoded
hypothetical protein CypaCp020	NP_043157.1	Chloroplast-encoded
hypothetical protein CypaCp031	NP_043168.1	Chloroplast-encoded
hypothetical protein CypaCp040	NP_043177.1	Chloroplast-encoded
hypothetical protein CypaCp064	NP_043201.1	Chloroplast-encoded
hypothetical protein CypaCp074	NP_043211.1	Chloroplast-encoded
hypothetical protein CypaCp134	NP_043271.1	Chloroplast-encoded
hypothetical protein CypaCp141	NP_043278.1	Chloroplast-encoded
hypothetical protein CypaCp145	NP_043282.1	Chloroplast-encoded
hypothetical protein CypaCp147	NP_043284.1	Chloroplast-encoded
Mg-protoporphyrin IX chelatase	NP_043214.1	Chloroplast-encoded
photochlorophyllide reductase subunit B	NP_043149.1	Chloroplast-encoded
photochlorophyllide reductase subunit L	NP_043285.1	Chloroplast-encoded
photochlorophyllide reductase subunit N	NP_043286.1	Chloroplast-encoded
photosystem I assembly protein Ycf3	NP_043154.1	Chloroplast-encoded
photosystem I assembly protein Ycf4	NP_043147.1	Chloroplast-encoded
photosystem I P700 chlorophyll a apoprotein A1	NP_043150.1	Chloroplast-encoded
photosystem I P700 chlorophyll a apoprotein A2	NP_043151.1	Chloroplast-encoded
photosystem I subunit III	NP_043153.1	Chloroplast-encoded

photosystem I subunit VII	NP_043270.1	Chloroplast-encoded
photosystem II 44 kDa protein	NP_043248.3	Chloroplast-encoded
photosystem II 47 kDa protein	NP_043167.1	Chloroplast-encoded
photosystem II cytochrome c550	NP_043260.1	Chloroplast-encoded
photosystem II protein H	NP_043169.1	Chloroplast-encoded
photosystem II protein I	NP_043236.1	Chloroplast-encoded
photosystem II protein J	NP_043181.1	Chloroplast-encoded
photosystem II protein K	NP_043249.1	Chloroplast-encoded
photosystem II protein N	NP_043165.1	Chloroplast-encoded
photosystem II protein V	NP_043178.1	Chloroplast-encoded
photosystem II protein W	NP_043233.1	Chloroplast-encoded
phycobilisome linker protein	NP_043160.1	Chloroplast-encoded
prenyl transferase	NP_043186.1	Chloroplast-encoded
preprotein translocase subunit SecY	NP_043187.1	Chloroplast-encoded
putative ABC transporter	NP_043219.1	Chloroplast-encoded
putative OmpR transcriptional regulator	NP_043176.1	Chloroplast-encoded
ribosomal protein L1	NP_043172.1	Chloroplast-encoded
ribosomal protein L11	NP_043171.1	Chloroplast-encoded
ribosomal protein L12	NP_043173.1	Chloroplast-encoded
ribosomal protein L14	NP_043193.1	Chloroplast-encoded
ribosomal protein L16	NP_043195.1	Chloroplast-encoded
ribosomal protein L19	NP_043156.1	Chloroplast-encoded
ribosomal protein L2	NP_043199.1	Chloroplast-encoded
ribosomal protein L20	NP_043162.1	Chloroplast-encoded
ribosomal protein L21	NP_043287.1	Chloroplast-encoded
ribosomal protein L22	NP_043197.1	Chloroplast-encoded
ribosomal protein L3	NP_043200.1	Chloroplast-encoded
ribosomal protein L33	NP_043202.1	Chloroplast-encoded
ribosomal protein L35	NP_043161.1	Chloroplast-encoded
ribosomal protein L5	NP_043192.1	Chloroplast-encoded
ribosomal protein L6	NP_043190.1	Chloroplast-encoded
ribosomal protein S10	NP_043206.1	Chloroplast-encoded
ribosomal protein S11	NP_043257.1	Chloroplast-encoded
ribosomal protein S12	NP_043209.1	Chloroplast-encoded
ribosomal protein S13	NP_043258.1	Chloroplast-encoded
ribosomal protein S14	NP_043148.1	Chloroplast-encoded
ribosomal protein S16	NP_043231.1	Chloroplast-encoded
ribosomal protein S17	NP_043194.1	Chloroplast-encoded
ribosomal protein S18	NP_043203.1	Chloroplast-encoded
ribosomal protein S19	NP_043198.1	Chloroplast-encoded
ribosomal protein S2	NP_043227.1	Chloroplast-encoded
ribosomal protein S20	NP_043217.1	Chloroplast-encoded
ribosomal protein S3	NP_043196.1	Chloroplast-encoded
ribosomal protein S4	NP_043212.1	Chloroplast-encoded

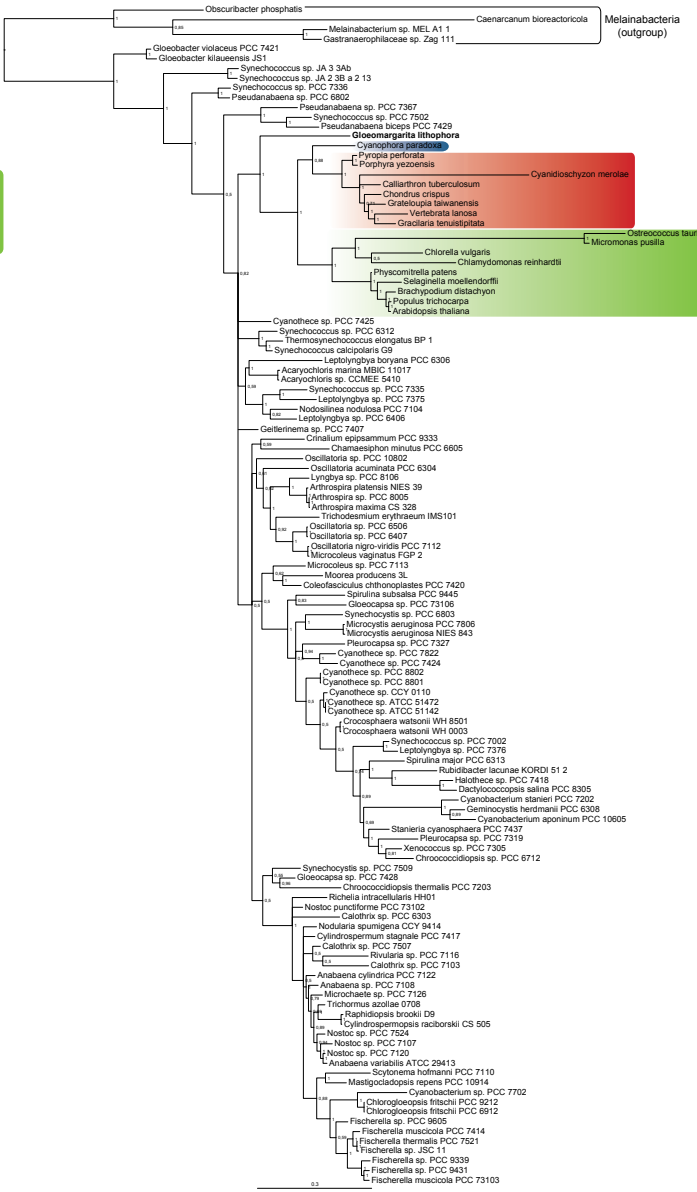
ribosomal protein S5	NP_043188.1	Chloroplast-encoded
ribosomal protein S6	NP_043268.1	Chloroplast-encoded
ribosomal protein S7	NP_043208.1	Chloroplast-encoded
ribosomal protein S8	NP_043191.1	Chloroplast-encoded
ribosomal protein S9	NP_043255.1	Chloroplast-encoded
ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	NP_043240.1	Chloroplast-encoded
RNA polymerase alpha subunit	NP_043256.1	Chloroplast-encoded
RNA polymerase beta subunit	NP_043230.1	Chloroplast-encoded
RNA polymerase beta' subunit	NP_043229.1	Chloroplast-encoded
RNA polymerase beta" subunit	NP_043228.1	Chloroplast-encoded
sulfate ABC transporter protein	NP_043218.1	Chloroplast-encoded
Ycf9	NP_043204.1	Chloroplast-encoded
3-isopropylmalate dehydrogenase	ABA22597.1	Nucleus-encoded
ABC transporter-like protein MlaF	ABA22286.1	Nucleus-encoded
ABC transporter-like protein EcfA2	ABA22129.1	Nucleus-encoded
Activator of bc1 complex ABC-1	ABA22988.1	Nucleus-encoded
Activator of bc1 complex ABC-1	ABA19944.1	Nucleus-encoded
Activator of bc1 complex ABC-1	ABA21853.1	Nucleus-encoded
acetyl-CoA carboxylase carboxyltransferase subunit alpha	ABA20140.1	Nucleus-encoded
AMP-dependent synthetase and ligase	ABA22781.1	Nucleus-encoded
arginine decarboxylase	ABA21777.1	Nucleus-encoded
aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B	ABA23557.1	Nucleus-encoded
ATP synthase F1 subcomplex gamma subunit	ABA22225.1	Nucleus-encoded
ATPase	ABA21953.1	Nucleus-encoded
ATP-dependent DNA helicase RecG	ABA21680.1	Nucleus-encoded
bacterial peptide chain release factor 2 (bRF-2)	ABA22397.1	Nucleus-encoded
bacterial peptide chain release factor 3 (bRF-3)	ABA22906.1	Nucleus-encoded
biopterin transport-related protein BT1	ABA20861.1	Nucleus-encoded
chorismate synthase	ABA20213.1	Nucleus-encoded
cobalt transport protein	ABA22508.1	Nucleus-encoded
cobaltochelate CobN subunit	ABA22919.1	Nucleus-encoded
conserved hypothetical protein	ABA21625.1	Nucleus-encoded
conserved hypothetical protein	ABA24174.1	Nucleus-encoded
conserved hypothetical protein	ABA22776.1	Nucleus-encoded
conserved hypothetical protein	ABA20101.1	Nucleus-encoded
conserved hypothetical protein	ABA22911.1	Nucleus-encoded
coproporphyrinogen III oxidase, anaerobic	ABA21841.1	Nucleus-encoded
C-terminal processing peptidase-2, Serine peptidase, MEROPS	ABA23047.1	Nucleus-encoded
Cys/Met metabolism pyridoxal-phosphate-dependent enzymes	ABA23812.1	Nucleus-encoded
dihydrolipoamide dehydrogenase	ABA21549.1	Nucleus-encoded
DNA gyrase subunit A	ABA24062.1	Nucleus-encoded
DNA gyrase subunit B	ABA22132.1	Nucleus-encoded
FAD dependent oxidoreductase	ABA22720.1	Nucleus-encoded
ferrochelate	ABA21197.1	Nucleus-encoded

geranylgeranyl reductase	ABA21120.1	Nucleus-encoded
glucose-inhibited division protein A subfamily	ABA23146.1	Nucleus-encoded
glutamate synthase (ferredoxin)	ABA20918.1	Nucleus-encoded
glutamate-1-semialdehyde 2,1-aminomutase	ABA24517.1	Nucleus-encoded
GrpE protein	ABA20003.1	Nucleus-encoded
GTPase EngC	ABA20007.1	Nucleus-encoded
GTP-binding protein LepA	ABA20066.1	Nucleus-encoded
GTP-binding protein, HSR1-related protein	ABA24222.1	Nucleus-encoded
heat shock protein DnaJ	ABA20005.1	Nucleus-encoded
heat shock protein DnaJ-like protein	ABA23874.1	Nucleus-encoded
HesB/YadR/YfhF	ABA19829.1	Nucleus-encoded
homogentisate phytyltransferase	ABA23079.1	Nucleus-encoded
iron-regulated ABC transporter permease protein SufD	ABA20052.1	Nucleus-encoded
isoleucyl-tRNA synthetase	ABA23379.1	Nucleus-encoded
lysyl-tRNA synthetase, class-2	ABA22758.1	Nucleus-encoded
LytB protein	ABA22560.1	Nucleus-encoded
membrane protease FtsH catalytic subunit	ABA20199.1	Nucleus-encoded
metallo peptidase, MEROPS family M50B	ABA21352.1	Nucleus-encoded
methionyl-tRNA synthetase	ABA22336.1	Nucleus-encoded
MutS 2 protein	ABA22264.1	Nucleus-encoded
N-acetylglutamate kinase	ABA20178.1	Nucleus-encoded
NnrU	ABA21370.1	Nucleus-encoded
peptidase M48, Ste24p	ABA21057.1	Nucleus-encoded
peptidase M50	ABA20694.1	Nucleus-encoded
peptidyl-prolyl cis-trans isomerase, cyclophilin type	ABA21934.1	Nucleus-encoded
phospho-N-acetylmuramoyl-pentapeptide-transferase	ABA20891.1	Nucleus-encoded
phosphoribulokinase	ABA20405.1	Nucleus-encoded
photosystem II manganese-stabilizing protein PsbO	ABA21465.1	Nucleus-encoded
phytoene synthase	ABA24391.1	Nucleus-encoded
protein of unknown function DUF1092	ABA21779.1	Nucleus-encoded
protoporphyrin IX magnesium-chelatase	ABA22693.1	Nucleus-encoded
Ras superfamily GTP-binding protein YlqF	ABA24244.1	Nucleus-encoded
ribonuclease II	ABA20946.1	Nucleus-encoded
ribosome recycling factor	ABA20253.1	Nucleus-encoded
septum site-determining protein MinD	ABA23085.1	Nucleus-encoded
signal recognition particle subunit FFH/SRP54 (srp54)	ABA23956.1	Nucleus-encoded
SSU ribosomal protein S1P	ABA21128.1	Nucleus-encoded
sulfite reductase (ferredoxin)	ABA24640.1	Nucleus-encoded
transcription-repair coupling factor	ABA22380.1	Nucleus-encoded
UDP-N-acetylglucosamine 1-carboxyvinyltransferase	ABA21058.1	Nucleus-encoded

A



B



C



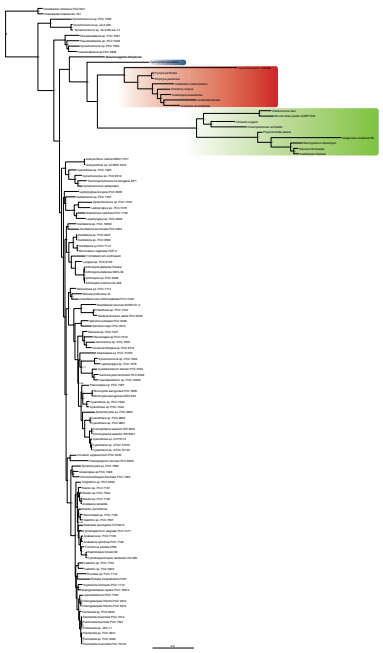
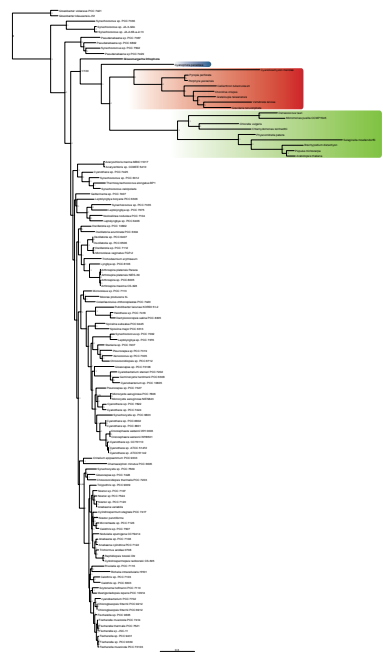
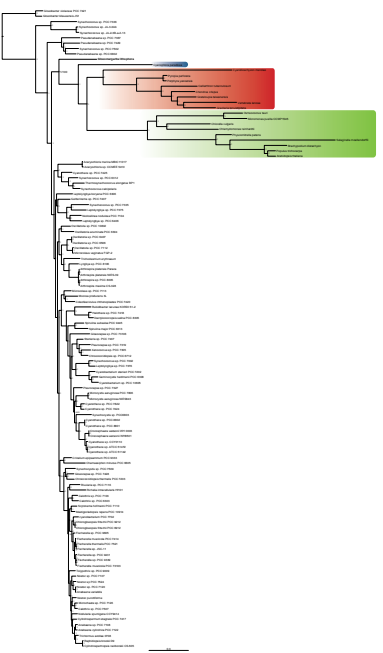
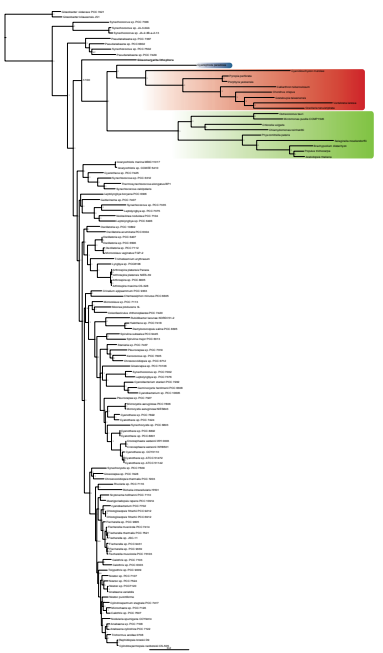
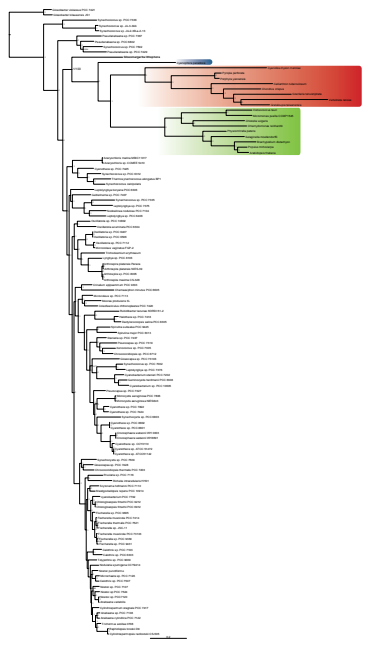
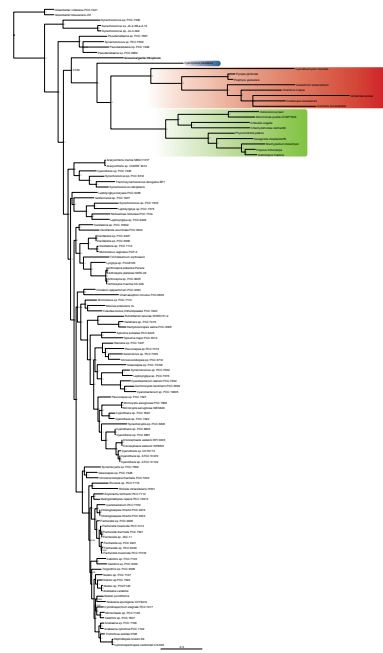
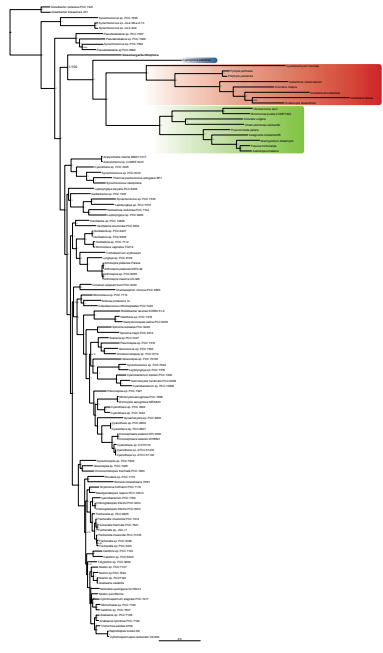
D**E****F****G****H****I****J****K****L****M**

Figure S1. Related to Figure 1. The Position of Plastids in the Cyanobacterial Phylogeny

(A) Bayesian phylogenetic tree based on the concatenation of 97 plastid-encoded proteins and their cyanobacterial homologues (21942 sites), including the fast-evolving cyanobacterial SynPro clade. Numbers indicate posterior probabilities.

(B) Bayesian phylogenetic tree based on the concatenation of 16S + 23S rRNA sequences of plastids and cyanobacteria (3427 nucleotides). The tree is rooted on the sister group of Cyanobacteria, the Melainabacteria. Numbers indicate posterior probabilities.

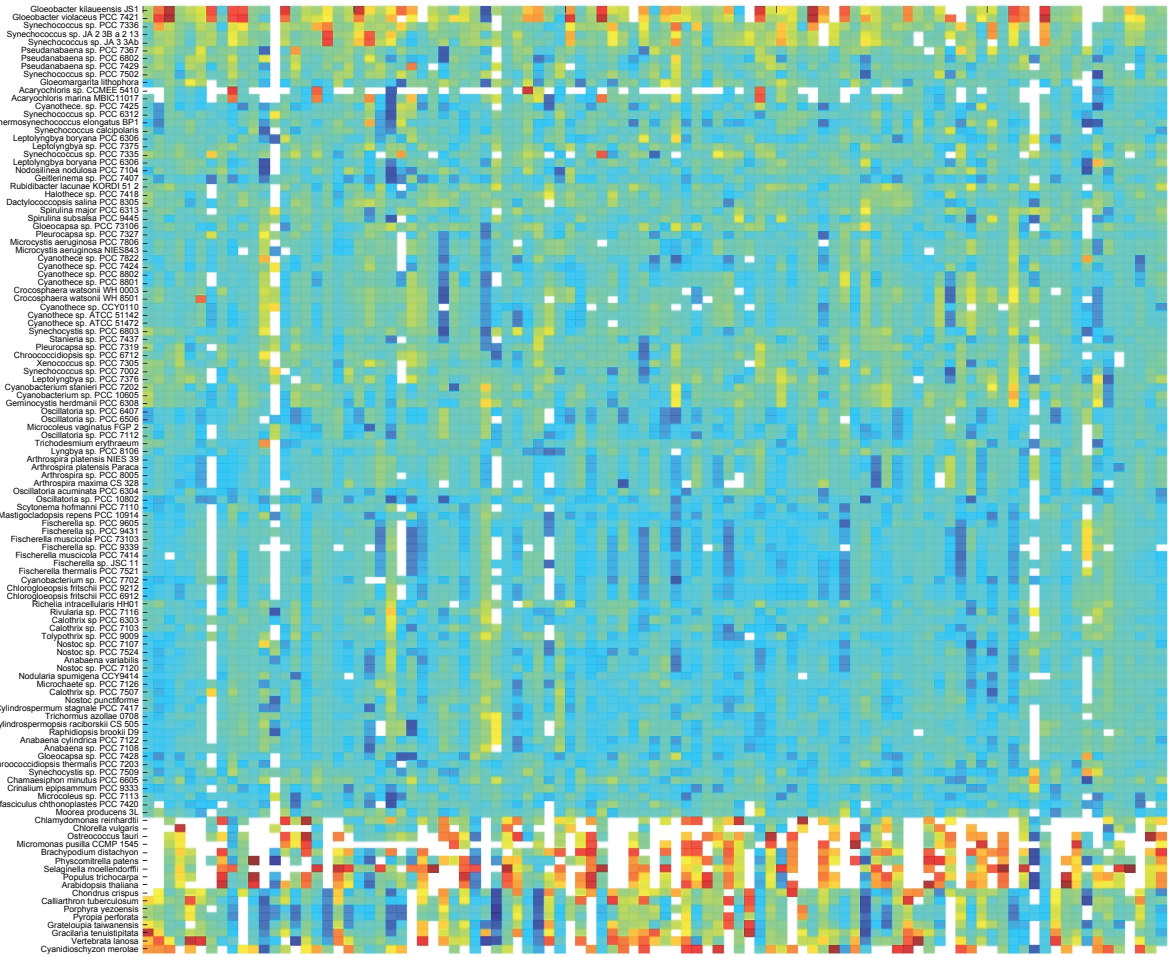
(C) Bayesian phylogenetic tree based on the concatenation of 72 nucleus-encoded proteins of plastid origin (28102 sites). Numbers indicate posterior probabilities. Maximum likelihood bootstrap support is also shown for the *Gloeomargarita*+Archaeplastida group.

(D) Bayesian phylogenetic tree based on the concatenation 97 plastid-encoded proteins and their cyanobacterial homologues recoded with the Dayhoff recoding scheme (21942 sites). Numbers indicate posterior probabilities.

(E) Bayesian phylogenetic tree based on all sites of the concatenation of 97 plastid-encoded proteins and their cyanobacterial homologues. Numbers indicate posterior probabilities.

(F-M) Bayesian phylogenetic trees based on the concatenation 97 plastid-encoded proteins and their cyanobacterial homologues after removal of a proportion of the fastest-evolving sites. (F) 10% fastest-evolving sites removed; (G) 20% fastest-evolving sites removed; (H) 30% fastest-evolving sites removed; (I) 40% fastest-evolving sites removed; (J) 50% fastest-evolving sites removed; (K) 60% fastest-evolving sites removed; (L) 70% fastest-evolving sites removed; and (M) 80% fastest-evolving sites removed. Numbers indicate posterior probabilities.

A



B

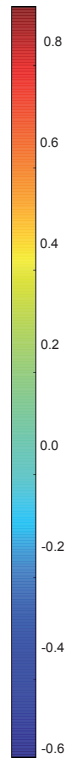
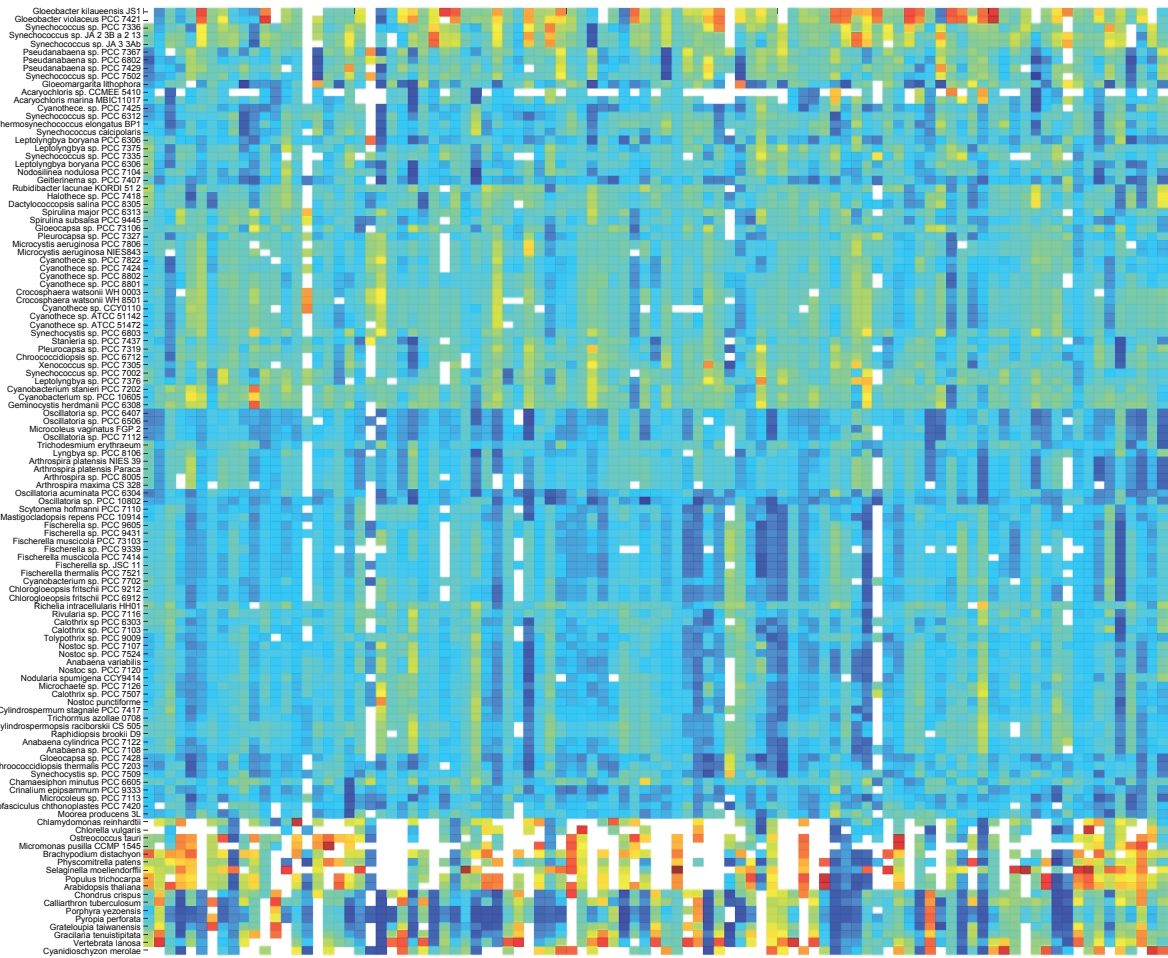


Figure S2. Related to Figure 2. Heatmaps Showing the Distance Between the Plastid Proteins of *Cyanophora paradoxa* and Those of Other Taxa for the 97 Plastid-encoded Proteins Used in our Analyses.

(A) Distances calculated from pairwise alignments.

(B) Patristic distances.

Supplemental Experimental Procedures

Sequencing and Annotation of the *Gloeomargarita lithophora* Genome

Gloeomargarita lithophora (strain CCAP 1437/1) was grown in liquid BG11 medium (<https://www.ccap.ac.uk/media/documents/BG11.pdf>) at 20°C and natural sunlight illumination. Cells were harvested by centrifugation at 10000 g for 10 minutes. DNA was extracted from the pelleted cells using the MoBio PowerSoil DNA Isolation kit (MoBio, Carlsbad, CA) following the manufacturer's instructions. DNA was sequenced using both the Illumina Genome Analyzer II and the 454 Titanium technologies (Eurofins Genomics). The mix of sequences produced with these two methods was combined into a single genomic scaffold (size=3049282 bp) using the Newbler assembler (<http://454.com/products/analysis-software/index.asp>).

ORFs were predicted using Prodigal [S1]. 2990 ORFs were validated as highly probable protein coding genes by BLAST [S2] comparison with the cyanobacterial gene sequences available in GenBank. All *G. lithophora* protein sequences were included in our local database used for the identification of plastid and cyanobacterial conserved markers (see main text). The annotated *G. lithophora* genome sequence has been deposited in GenBank under accession number NZ_CP017675.

Selection of Phylogenetic Markers

In addition to the information provided in the main text, our procedure to exclude potentially problematic markers based on the manual inspection of the phylogenetic trees reconstructed with each individual marker was as follows: We first identified trees containing species with two or more distantly related sequences, which indicated the presence of paralogs. If these were not issued from recent duplications affecting only single species of small groups of species within a lineage, we excluded the marker. We also detected ancient paralogs that were duplicated before the diversification of cyanobacteria and plastids. If the two paralogs were well separated in our trees, we kept both as two different

individual markers (e.g., apoproteins A1 and A2 of the photosystem I P700 chlorophyll *a*). Otherwise, we eliminated them. In addition, we compared our individual trees with the multigene cyanobacterial phylogeny recently published by Shih et al. (2013) [S3]. These authors identified a set of well-supported clearly monophyletic cyanobacterial groups. When those groups were retrieved non-monophyletic with strong support in our single-marker trees, the corresponding markers were discarded. We therefore followed a highly conservative approach to reduce as much as possible any noise due to erroneous marker selection.

Phylogenetic Analysis

Multiple sequence alignments were carried out with MAFFT [S4] using default parameters. Selection of conserved regions for phylogenetic analyses was done with trimAL [S5] with automatic selection of parameters (-automated1 option).

Substitution models for phylogenetic analysis of individual protein markers were selected using ProtTest 3 [S6].

Supplemental References

- S1. Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 1471-2105.
- S2. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389-3402.
- S3. Shih, P.M., Wu, D., Latifi, A., Axen, S.D., Fewer, D.P., Talla, E., Calteau, A., Cai, F., Tandeau de Marsac, N., Rippka, R., et al. (2013). Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci U S A* 110, 1053-1058.
- S4. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30, 772-780.
- S5. Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 25, 1972-1973.
- S6. Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27, 1164-1165.

9.2 Supplementary material of manuscript 2

Functioning of Secondary Green Plastids is Assured by a Mix of Genes of Red and Green Algal Ancestry

(Manuscript in preparation)

Rafael I. Ponce-Toledo¹, David Moreira¹, Purificación López-García¹ and Philippe Deschamps^{1*}

¹Unité d'Ecologie Systématique et Evolution, CNRS, Université Paris-Sud, AgroParisTech, Université Paris-Saclay, 91400, Orsay, France

*Corresponding author: philippe.deschamps@u-psud.fr

Functioning of Secondary Green Plastids is Assured by a Mix of Genes of Red and Green Algal Ancestry

Supplementary Material

Rafael I. Ponce-Toledo, David Moreira, Purificación López-García, and Philippe Deschamps

Unité d'Ecologie Systématique et Evolution, CNRS, Université Paris-Sud, AgroParisTech, Université Paris-Saclay, 91400, Orsay, France

Supplementary Materials and Methods

Data availability

Protein sequences used for inferring all phylogenetic trees presented in this work are available for download at <http://www.ese.u-psud.fr/article950.html?lang=en>. They include nonaligned sequences and trimmed alignments.

Sequence analysis

A local database was constructed to host the proteomes predicted for various nuclear genomes and transcriptomes as well as plastid genomes (see Table S1 for the complete list).

All proteins in the *Bigeloviella natans* (Chlorarachniophyta) and *Guillardia theta* (Cryptophyta) predicted proteomes were used as queries for BLASTp sequence similarity searches {Camacho, 2009 #4853} against the local database. We retained up to 350 top hits retained with an e-value threshold $<1e-05$. BLASTp outputs were parsed to identify proteins shared by diverse photosynthetic eukaryotes and that are more similar to cyanobacteria or other bacteria than to non-photosynthetic eukaryotes.

For protein sequences that passed this pre-selection, reciprocal BLASTp were done against the database to collect a set of up to 600 similar proteins. We then aligned each set using Mafft v7.123b {Katoh, 2013 #4577} with default parameters. Non conserved

regions of the alignments were trimmed with BMGE v1.0 {Criscuolo, 2010 #4578} allowing a maximum of 50% gaps per position. Preliminary phylogenetic trees were inferred from trimmed alignments using FastTree v2.1.7 {Price, 2010 #4665} with default parameters. This first set of trees was then manually inspected to determine those showing a topology compatible with an EGT/HGT scenario. For all positive topologies, only sequences corresponding to the portion of interest of each draft phylogenetic tree (the part showing the photosynthetic eukaryotes and the closest outgroup) were saved for the remaining steps. A second round of alignment/trimming/inference was then processed with these subsets of similar proteins. Some alignments were then improved by removing duplicates and very short partial sequences. To speed up subsequent calculations, several outgroup sequences were also removed from all alignments (see Table S2). The final sequence datasets were realigned, trimmed using TrimAL v1.4.rev15 with “gappy-out” parameter {Capella-Gutierrez, 2009 #4702} and used to infer ML trees using IQtree v1.5.1 with the CAT 20+1-Gamma model of sequence evolution, plus 1000 pseudo-bootstrap replicates {Nguyen, 2015 #4854}. A final selection of these trees done by manual inspection to keep those fulfilling the following two requirements: 1) The protein has to be shared by cyanobacteria (or other bacteria), Archaeplastida and at least three secondary photosynthetic lineages, and 2) the corresponding phylogenetic trees have to support the monophyly of Archaeplastida and a clear separation of Viridiplantae and Rhodophyta. Finally, the 85 trees passing this final filter (figs S1-S85) were inspected to infer the phylogenetic origin of the corresponding genes in the secondary photosynthetic lineages (Table S3).

Gene functional annotation

We annotated the functions of the 85 proteins in the final selection (see above) through the EggNOG 4.5 {Huerta-Cepas, 2016 #4852} web portal (<http://eggnogdb.embl.de>). For each protein dataset we used as queries the ortholog sequences of *Guillardia theta* and *Bigeloviella natans*.

Table S1. List of species in the local genome database**Archaea**

<i>Acidianus hospitalis</i> W1	<i>Methanocaldococcus jannaschii</i> DSM 2661
<i>Acidilobus saccharovorans</i> 345-15	<i>Methanocaldococcus</i> sp. FS406-22
<i>Aciduliprofundum boonei</i> T469	<i>Methanocaldococcus vulcanius</i> M7
<i>Aeropyrum pernix</i> K1	<i>Methanocella paludicola</i> SANAE
<i>Archaeoglobus fulgidus</i> DSM_4304	<i>Methanococcoides burtonii</i> DSM_6242
<i>Archaeoglobus profundus</i> DSM 5631	<i>Methanococcus aeolicus</i> Nankai_3
<i>Archaeoglobus veneficus</i> SNP6	<i>Methanococcus maripaludis</i> C5
<i>Caldiarchaenum subterraneum</i>	<i>Methanococcus vannielii</i>
<i>Caldivirga maquilingensis</i> IC_167	<i>Methanococcus voltae</i> A3
<i>Candidatus Methanoregula boonei</i> 6A8	<i>Methanocorpusculum labreanum</i> Z
<i>Candidatus Micrarchaeum acidiphilum</i>	<i>Methanoculleus marisnigri</i> JR1
<i>Candidatus Nanosalinarum</i> sp. J07AB56	<i>Methanohalobium evestigatum</i> Z-7303
<i>Candidatus Nanosalina</i> sp. J07AB43	<i>Methanohalophilus mahii</i> DSM 5219
<i>Candidatus Nitrosoarchaeum koreensis</i> MY1	<i>Methanoplanus petrolearius</i> DSM11571
<i>Candidatus Nitrosoarchaeum limnia</i> SFB1	<i>Methanopyrus kandleri</i> AV19
<i>Candidatus Nitrososphaera gargensis</i> Ga9.2	<i>Methanosaeta concilii</i> GP6
<i>Candidatus Nitrosotenuis uzonensis</i>	<i>Methanosaeta harundinacea</i> 6Ac
<i>Candidatus Parvarchaeum acidiphilum</i>	<i>Methanosaeta thermophila</i> PT
<i>Candidatus Parvarchaeum acidophilum</i>	<i>Methanosalsum zhilinae</i> DSM 4017
<i>Cenarchaeum symbiosum</i> A	<i>Methanosarcina acetivorans</i> C2A
<i>Crenarchaeota</i> sp.	<i>Methanosarcina barkeri</i> str. Fusaro
<i>Desulfurococcus kamchatkensis</i>	<i>Methanosarcina mazei</i> Go1
<i>Desulfurococcus mucosus</i> DSM2162	<i>Methanosphaera stadtmanae</i> DSM_3091
<i>Euryarchaeota</i> sp.	<i>Methanosphaerula palustris</i> E1-9c
<i>Evoldeep Eury fosmid</i>	<i>Methanospirillum hungatei</i> JF_1
<i>Evoldeep Thaum fosmid</i>	<i>Methanothermobacter marburgensis</i> . Marburg
<i>Ferroglobus placidus</i> DSM 10642	<i>Methanothermobacter thermautotrophicus</i>
<i>Ferroplasma acidarmanus</i> fer1	<i>Methanothermococcus okinawensis</i> IH1
<i>Halalkalicoccus jeotgali</i> B3	<i>Methanothermus fervidus</i>
<i>Haloarcula hispanica</i> ATCC 33960	<i>Methanotorris igneus</i> Kol 5
<i>Haloarcula marismortui</i> ATCC_43049	<i>Nanoarchaeum equitans</i> Kin4_M
<i>Halobacterium salinarum</i> R1	<i>Natrialba magadii</i> ATCC 43099
<i>Haloferax volcanii</i> DS2	<i>Natronomonas pharaonis</i> DSM_2160
<i>Halogeometricum borinquense</i> DSM 11551	<i>Nitrosopumilus maritimus</i> SCM1
<i>Halomicrobium mukohataei</i> DSM 12286	<i>Picrophilus torridus</i> DSM_9790
<i>halophilic archaeon</i> DL31	<i>Pyrobaculum aerophilum</i> IM2
<i>Halopiger xanaduensis</i> SH-6	<i>Pyrobaculum arsenaticum</i> DSM 13514
<i>Haloquadratum walsbyi</i> DSM_16790	<i>Pyrobaculum calidifontis</i> JCM 11548
<i>Halorhabdus utahensis</i> DSM 12940	<i>Pyrobaculum islandicum</i> DSM 4184
<i>Halorubrum lacusprofundi</i> ATCC 49239	<i>Pyrobaculum</i> sp. 1860
<i>Haloterrigena turkmenica</i> DSM 5511	<i>Pyrococcus abyssi</i> GE5
<i>Hyperthermus butylicus</i> DSM_5456	<i>Pyrococcus furiosus</i> DSM 3638
<i>Ignicoccus hospitalis</i> KIN4_I	<i>Pyrococcus horikoshii</i> OT3
<i>Ignisphaera aggregans</i> DSM17230	<i>Pyrococcus</i> sp. NA2
<i>Korarchaeum cryptofilum</i> OPF8	<i>Pyrococcus yayanosii</i> CH1
<i>Marine group II euryarchaeote</i> SCGC AB629J06	<i>Pyrolobus fumarii</i> IA
<i>Metallosphaera cuprina</i> Ar-4	<i>Staphylothermus hellenicus</i> DSM 12710
<i>Metallosphaera sedula</i> DSM_5348	<i>Staphylothermus marinus</i> F1
<i>Methanobacterium</i> sp. AL-21	<i>Sulfolobus acidocaldarius</i> DSM_639
<i>Methanobrevibacter ruminantium</i> M1	<i>Sulfolobus islandicus</i> L.S.2.15
<i>Methanobrevibacter smithii</i> ATCC_35061	<i>Sulfolobus solfataricus</i> P2
<i>Methanocaldococcus fervens</i> AG86	<i>Sulfolobus tokodaii</i> str 7
<i>Methanocaldococcus infernus</i> ME	<i>Thermococcus barophilus</i> MP

Thermococcus gammatolerans EJ3
Thermococcus kodakarensis KOD1
Thermococcus onnurineus NA1
Thermococcus sibiricus MM 739
Thermococcus sp. AM4
Thermofilum pendens Hrk_5
Thermoplasma acidophilum DSM_1728
Thermoplasma volcanium GSS1

Bacteria

Acaryochloris marina MBIC11017
Acaryochloris sp. CCMEE 5410
Acholeplasma laidlawii PG_8A
Acidiphilium cryptum JF_5
Acidithiobacillus ferrooxidans ATCC 2327
Acidobacteria bacterium Ellin345
Acidothermus cellulolyticus 11B
Acinetobacter baumannii AB0057
Actinobacillus pleuropneumoniae L20
Aeromonas hydrophila ATCC_7966
Agrobacterium tumefaciens C58
Akkermansia muciniphila ATCC BAA-835
Alcanivorax borkumensis SK2
Aliivibrio salmonicida LFI1238
Alkalilimnicola ehrlichei MLHE_1
Alkaliphilus metalliredigens QYMF
Alteromonas macleodii Deep ecotype
Amoebophilus asiaticus 5a2
Anabaena cylindrica PCC 7122
Anabaena sp. PCC 7108
Anabaena variabilis
Anaerocellum thermophilum DSM 6725
Anaeromyxobacter sp. Fw109_5
Anaplasma marginale StMaries
Anoxybacillus flavithermus WK1
Aquifex aeolicus VF5
Arcobacter butzleri RM4018
Aromatoleum aromaticum EbN1
Arthrobacter aureescens TC1
Arthrospira maxima CS-328
Arthrospira platensis NIES-39
Arthrospira platensis Paraca
Arthrospira sp. PCC 8005
Azoarcus sp. BH72
Azorhizobium caulinodans ORS_571
Bacillus amyloliquefaciens FZB42
Bacteroides fragilis NCTC_9343
Bacteroides fragilis YCH46
Bartonella bacilliformis KC583
Baumannia cicadellinicola Hc
Bdellovibrio bacteriovorus HD100
Beijerinckia indica subsp. *Indica* ATCC 9039
Bifidobacterium adolescentis ATCC_15703
Blochmannia floridanus
Bordetella bronchiseptica RB50

Thermoproteus neutrophilus V24Sta
Thermoproteus tenax Kra 1
Thermoproteus uzoniensis 768- 20
Thermosphaera aggregans DSM 11486
Uncultured Marine group II euryarchaeote
uncultured Methanogenic archaeon RC-I
Vulcanisaeta distributa DSM14429
Vulcanisaeta moutnovskia 768-28

Borrelia afzelii PKo
Brachyspira hyodysenteriae WA1
Bradyrhizobium sp. BTAi1
Brucella abortus 9_941
Buchnera aphidicola APS
Burkholderia sp. 383
Caldicellulosiruptor saccharolyticus DSM_8903
Calothrix sp. PCC 6303
Calothrix sp. PCC 7103
Calothrix sp. PCC 7507
Campylobacter concisus 13826
Candidatus Phytoplasma asteris AYWB
Carboxydotherrnus hydrogenoformans Z_2901
Carsonella ruddii PV
Caulobacter crescentus CB15
Cellvibrio japonicus Ueda107
Chamaesiphon minutus PCC 6605
Chlamydia muridarum Nigg
Chlamydia trachomatis 70
Chlamydomydia abortus S26_3
Chlamydomydia pneumoniae CWL029
Chlorobaculum parvum NCIB 8327
Chlorobium chlorochromatii CaD3
Chloroflexus aurantiacus J_10_fl
Chlorogloeopsis fritschii PCC 6912
Chlorogloeopsis fritschii PCC 9212
Chloroherpeton thalassium ATCC 35110
Chromobacterium violaceum ATCC_12472
Chromohalobacter salexigens DSM 3043
Chroococcidiopsis sp. PCC 6712
Chroococcidiopsis thermalis PCC 7203
Citrobacter koseri ATCC_BAA895
Clavibacter michiganensis NCPPB_382
Clostridium acetobutylicum ATCC_824
Coleofasciculus chthonoplastes PCC 7420
Colwellia psychrerythraea 34H
Coprothermobacter proteolyticus DSM 5265
Corynebacterium diphtheriae NCTC_13129
Coxiella burnetii RSA_493
Criblamydia sequanensis CRIB-18
Crinalium epipsammum PCC 9333
Crocospaera watsonii WH 3
Crocospaera watsonii WH501
Cronobacter sakazakii ATCC BAA-894
Cupriavidus taiwanensis

Cyanobacterium sp. PCC 10605
cyanobacterium sp. PCC 7702
Cyanobacterium stanieri PCC 7202
Cyanobacterium Yellowstone A
Cyanobacterium Yellowstone B
Cyanobium gracile PCC 6307
Cyanothece sp. ATCC51142
Cyanothece sp. ATCC 51472
Cyanothece sp. CCY0110
Cyanothece sp. PCC 7424
Cyanothece sp. PCC 7425
Cyanothece sp. PCC 7822
Cyanothece sp. PCC 8801
Cyanothece sp. PCC 8802
Cylindrospermopsis raciborskii CS-505
Cylindrospermum stagnale PCC 7417
Cytophaga hutchinsonii ATCC_33406
Dactylococcopsis salina PCC 8305
Dechloromonas aromatica RCB
Dehalococcoides sp. BAV1
Deinococcus geothermalis DSM_11300
Delftia acidovorans SPH_1
Desulfitobacterium hafniense Y51
Desulfococcus oleovorans Hxd3
Desulfotalea psychrophila Lsv54
Desulfotomaculum reducens MI_1
Desulfovibrio desulfuricans G20
Desulfovibrio vulgaris DP4
Desulfovibrio vulgaris Hildenborough
Diaphorobacter sp. TPSY
Dichelobacter nodosus VCS1703A
Dictyoglomus thermophilum H-6-12
Dinoroseobacter shibae DFL_12
Ehrlichia canis Jake
Elusimicrobium minutum Pei191
Enterobacter sp. 638
Enterococcus faecalis V583
Erythrobacter litoralis HTCC2594
Escherichia coli 536
Exiguobacterium sibiricum 255-15
Fervidobacterium nodosum Rt17_B1
Finegoldia magna ATCC_29328
Fischerella muscicola PCC 73103
Fischerella muscicola PCC 7414
Fischerella sp. JSC-11
Fischerella sp. PCC 9339
Fischerella sp. PCC 9431
Fischerella sp. PCC 9605
Fischerella thermalis PCC 7521
Flavobacterium johnsoniae UW101
Flavobacterium psychrophilum JIP02_86
Francisella novicida U112
Frankia alni ACN14a
Fusobacterium nucleatum ATCC_25586
Geitlerinema sp. PCC 7407
Geminocystis herdmannii PCC 6308
Geobacillus kaustophilus HTA426
Geobacter metallireducens GS_15
Geobacter sulfurreducens PCA
Geobacter uraniumreducens Rf4
Gloeobacter kilauensis JS1
Gloeobacter violaceus PCC7421
Gloeocapsa sp. PCC 73106
Gloeocapsa sp. PCC 7428
Gloeomargarita lithophora
Gluconacetobacter diazotrophicus PAI_5
Gramella forsetii KT0803
Granulibacter bethesdensis CGDNIH1
Haemophilus ducreyi 35000HP
Hahella chejuensis KCTC_2396
Haliangium ochraceum DSM 14365
Halorhodospira halophila SL1
Halothece sp. PCC 7418
Halothermothrix orenii H 168
Helicobacter acinonychis Sheeba
Heliobacterium modesticaldum Ice1
Herminiimonas arsenicoxydans
Herpetosiphon aurantiacus ATCC 23779
Hydrogenobaculum sp. Y04AAS1
Hyphomonas neptunium ATCC_15444
Idiomarina loihiensis L2TR
Jannaschia sp. CCS1
Janthinobacterium sp. Marseille
Kineococcus radiotolerans SRS30216
Klebsiella pneumoniae MGH_78578
Kocuria rhizophila DC2201
Lactobacillus acidophilus NCFM
Lactococcus lactis I11403
Laptolyngbya sp. PCC 7376
Lawsonia intracellularis PHE_MN1_00
Legionella pneumophila Corby
Leifsonia xyli CTCB07
Leptolyngbya boryana PCC 6306
Leptolyngbya sp. PCC 6406
Leptolyngbya sp. PCC 7375
Leptospira borgpetersenii JB197
Leptothrix cholodnii SP-6
Leuconostoc citreum KM20
Listeria innocua Clip11262
Lyngbya sp. PCC8106
Lysinibacillus sphaericus C3_41
Macrococcus caseolyticus JCSC5402
Magnetococcus sp. MC_1
Magnetospirillum magneticum AMB_1
Mannheimia succiniciproducens MBEL55E
Maricaulis maris MCS10
Marinobacter aquaeolei VT8
Marinomonas sp. MWYL1
Mastigocladopsis repens PCC 10914
Mesoplasma florum LI
Mesorhizobium sp. BNC1
Methylidium petroleiphilum PM1

Methylobacillus flagellatus KT
Methylobacterium extorquens PA1
Methylocella silvestris BL2
Methylococcus capsulatus Bath
Microchaete sp. PCC 7126
Microcoleus sp. PCC 7113
Microcoleus vaginatus FGP-2
Microcystis aeruginosa NIES843
Microcystis aeruginosa PCC 7806
Moorea producens 3L
Moorella thermoacetica ATCC 39073
Mycobacterium abscessus
Mycoplasma agalactiae PG2
Myxococcus xanthus DK_1622
Natranaerobius thermophilus JWNM-WN-LF
Nautilia profundicola AmH
Neisseria gonorrhoeae FA_1090
Neorickettsia sennetsu Miyayama
Nitratiruptor sp. SB155_2
Nitrobacter hamburgensis XI4
Nitrosococcus oceani ATCC_19707
Nitrosomonas europaea ATCC_19718
Nitrospira multififormis ATCC_25196
Nocardia farcinica IFM_10152
Nocardioides sp. JS614
Nodosilinea nodulosa PCC 7104
Nodularia spumigena CCY9414
Nostoc punctiforme 11
Nostoc sp. PCC 7107
Nostoc sp. PCC7120
Nostoc sp. PCC 7524
Novosphingobium aromaticivorans DSM_12444
Oceanobacillus iheyensis HTE831
Ochrobactrum anthropi ATCC_49188
Oenococcus oeni PSU_1
Oligotropha carboxidovorans OM5
Onion yellows *Phytoplasma* sp. OY-M
Opitutus terrae PB90-1
Oscillatoria acuminata PCC 6304
Oscillatoria sp. PCC 10802
Oscillatoria sp. PCC 6407
Oscillatoria sp. PCC 6506
Oscillatoria sp. PCC 7112
Parabacteroides distasonis ATCC_8503
Parachlamydia acanthamoebae
Paracoccus denitrificans PD1222
Parvibaculum lavamentivorans DS1
Pasteurella multocida Pm70
Pectobacterium atrosepticum SCRI1043
Pediococcus pentosaceus ATCC_25745
Pelagibacter ubique HTCC1062
Pelobacter carbinolicus DSM_2380
Pelobacter propionicus DSM_2379
Pelodictyon luteolum DSM_273
Pelotomaculum thermopropionicum SI
Petrogla mobilis SJ95
Phenylobacterium zucineum HLK1
Photobacterium profundum SS9
Photorhabdus luminescens TTO1
Phytoplasma australiense
Plesiocystis pacifica SIR-1
Pleurocapsa sp. PCC 7319
Pleurocapsa sp. PCC 7327
Polaromonas sp. JS666
Polynucleobacter sp. QLW PIDMWA 1
Porphyromonas gingivalis W83
Prochlorococcus marinus AS9601
Prochlorococcus marinus CCMP1375
Prochlorococcus marinus CCMP1986
Prochlorococcus marinus MIT9215
Prochlorococcus marinus MIT9301
Prochlorococcus marinus MIT9303
Prochlorococcus marinus MIT9312
Prochlorococcus marinus MIT9313
Prochlorococcus marinus MIT9515
Prochlorococcus marinus NATL1A
Prochlorococcus marinus NATL2A
Propionibacterium acnes KPA171202
Prosthecochloris vibrioformis DSM_265
Proteus mirabilis HI4320
Protochlamydia amoebophila UWE25
Pseudanabaena sp. PCC 6802
Pseudanabaena sp. PCC 7367
Pseudanabaena sp. PCC 7429
Pseudoalteromonas atlantica T6c
Pseudomonas aeruginosa PAO1
Psychrobacter arcticus 273_4
Psychromonas ingrahamii 37
Ralstonia eutropha H16
Raphidiopsis brookii D9
Renibacterium salmoninarum ATCC_33209
Rhizobium etli CFN_42
Rhodobacter sphaeroides 2_4_1
Rhodococcus sp. RHA1
Rhodoferax ferrireducens T118
Rhodopirellula baltica SH_1
Rhodopseudomonas palustris BisA53
Rhodospirillum rubrum ATCC 11170
Richelia intracellularis HH01
Rickettsia akari Hartford
Rivularia sp. PCC 7116
Roseiflexus castenholzii DSM_13941
Roseobacter denitrificans OCh_114
Rubidibacter lacunae KORDI 51-2
Rubrobacter xylanophilus DSM_9941
Ruegeria sp. TM1040
Saccharophagus degradans 2_40
Saccharopolyspora erythraea NRRL_2338
Salinibacter ruber DSM_13855
Salinispora arenicola CNS_205
Salmonella enterica arizonae
Scytonema hofmanni PCC 7110

Serratia proteamaculans 568
Shewanella amazonensis SB2B
Shigella boydii Sb227
Silicibacter pomeroyi DSS_3
Simkania negevensis Z
Sinorhizobium medicae WSM419
Sodalis glossinidius
Solibacter usitatus Ellin6076
Sorangium cellulosum
Sphingomonas wittichii RW1
Sphingopyxis alaskensis RB2256
Spirulina major PCC 6313
Spirulina subsalsa PCC 9445
Stanieria sp. PCC 7437
Staphylococcus aureus MRSA252
Stenotrophomonas maltophilia K279a
Streptococcus agalactiae 2603V_R
Streptomyces avermitilis MA_4680
Sulcia muelleri GWSS
Sulfurihydrogenibium sp. YO3AOP1
Sulfurimonas denitrificans DSM 1251
Sulfurovum sp. NBC37_1
Symbiobacterium thermophilum IAM_14863
Synechococcus calcipolaris
Synechococcus elongatus PCC6301
Synechococcus elongatus PCC7942
Synechococcus sp. CC9311
Synechococcus sp. CC9605
Synechococcus sp. CC9902
Synechococcus sp. PCC 6312
Synechococcus sp. PCC 7002
Synechococcus sp. PCC 7335
Synechococcus sp. PCC 7336
Synechococcus sp. PCC 7502
Synechococcus sp. RCC307
Synechococcus sp. WH5701
Synechococcus sp. WH7803
Synechococcus sp. WH8102
Synechocystis sp. PCC6803
Synechocystis sp. PCC 7509
Syntrophobacter fumaroxidans MPOB
Syntrophomonas wolfei. *Wolfei* str. Goettingen
Syntrophus aciditrophicus SB
Thermoanaerobacter pseudethanolicus
Thermobifida fusca YX
Thermodesulfobivrio yellowstonii DSM 11347
Thermomicrobium roseum DSM 5159
Thermosiphon melanesiensis BI429
Thermosynechococcus elongatus BP1
Thermotoga lettingae TMO
Thermus thermophilus HB8
Thioalkalivibrio sp. HL-EbGR7
Thiobacillus denitrificans ATCC_25259
Thiomicrospira crumogena XCL_2
Tolypfthrix sp. PCC 9009
Treponema denticola ATCC_35405
Trichodesmium erythraeum
Trichormus azollae 708
Tropheryma whipplei TW08_27
Ureaplasma parvum ATCC 700970
Verminephrobacter eiseniae EF01_2
Vibrio cholerae N16961
Waddlia chondrophila WSU 86-1044
Wigglesworthia glossinidia
Wolbachia endosymbiont TRS
Wolinella succinogenes DSM_1740
Xanthobacter autotrophicus Py2
Xanthomonas campestris ATCC_33913
Xenococcus sp. PCC 7305
Xylella fastidiosa 9a5c
Yersinia enterocolitica 8081
Zymomonas mobilis ZM4

Eukaryotes

Acanthamoeba castellanii str Neff
Ahnfeltiopsis flabelliformis
Alexandrium minutum CCMP113
Allomyces macrogynus
Alveolata sp. CCMP3155
Ammonia sp.
Anophryoides haemophila AH6
Aphanochaete repens
Aplanochytrium sp. PBS07
Aplanochytrium stocchinoi GSBS06
Arabidopsis thaliana
Aspergillus niger
Asterionella formosa
Asterionellopsis glacialis
Asteromonas gracilis
Aurantiochytrium limacinum ATCCMYA1381
Aureococcus anophagefferens
Aureoumbra lagunensis CCMP1507
Aureoumbra lagunensis CCMP1510
Babesia bigemina
Babesia microti
Bangia atropurpurea
Bathycoccus prasinus
Batrachochytrium dendrobatidis JAM81
Betaphycus philippinensis
Bicosoecid sp. ms1
Bigelowiella natans
Blastocystis hominis
Bodo saltans
Bolbocoleon piliferum
Bolidomonas pacifica CCMP1866
Botryococcus braunii
Botryosphaerella sudetica

Botrytis cinerea
Brachiomonas submarina
Brachypodium distachyon
Bryopsis plumosa
Caenorhabditis elegans
Cafeteria roebergensis E4-10
Calcidiscus leptoporus RCC1130
Calliarthron tuberculosum
Capsaspora owczarzaki
Carteria crucifera
Cephaleuros virescens
Ceramium kondoi
Cerataulina daemon
Chaetoceros curvisetus
Chaetoceros neogracile CCMP1317
Chaetoceros simplex
Chaetomium globosum
Chaetopeltis orbicularis
Chattonella subsalsa CCMP2191
Chlamydomonas reinhardtii
Chlorarachnion reptans CCCM449
Chlorella sp. NC64A
Chlorella vulgaris C-169
Chlorococcum oleofaciens
Chondrus crispus
Chromera velia
Chromera velia
Chroodactylon ornatum
Chroomonas mesostigmatica CCMP1168
Chrysochromulina polylepis CCMP1757
Cladophora glomerata
Coccomyxa subellipsoidea
Codium fragile
Compsopogon coeruleus
Condyllostoma magnum COL2
Coprinopsis cinerea
Corethron pennatum L29A3
Crustomastix stigmata CCMP3273
Cryptococcus neoformans JEC21
Cryptoglana skujai
Cryptomonas paramecium CCAP977
Cryptosporidium hominis TU502
Cryptosporidium parvum
Cyanidioschyzon merolae
Cyanidium caldarium
Cyanophora paradoxa
Cyanoptycha gloeocystis
Cylindrocapsa geminella
Cylindrotheca closterium
Daphnia pulex
Debaryomyces hansenii CBS767
Desmochloris halophila
Dictyocha speculum CCMP1381
Dictyostelium discoideum
Dictyostelium purpureum QSDP1
Didymosphenia geminata
Dinobryon sp. UTEXLB2267
Diplonema papillatum
Dolichomastix tenuilepis CCMP3274
Dumontia simplex
Dunaliella salina
Ectocarpus siliculosus
Eimeria mitis
Elphidium margaritaceum
Emiliana huxleyi CCMP1516
Entamoeba dispar SAW760
Entamoeba histolytica
Entamoeba invadens IPI
Entamoeba nuttalli P19
Eremosphaera viridis
Erythrolobus australicus CCMP3124
Erythrolobus madagascarensis CCMP3276
Ettlia oleoabundans
Eucampia antarctica CCMP1452
Eucheuma denticulatum
Eudorina elegans
Euglenaformis proxima
Euglena gracilis
Euglena longa
Euglenaria anabaena
Euglena sp.
Eunotia naegelii
Euplotes focardii TN1
Eutreptiella gymnastica
Exanthemachrysis gayraliae RCC1523
Filamoeba nolandi
Fistulifera solaris
Florenciella parvula RCC1693
Florenciella sp. RCC1587
Floydiella terrestris
Fragilariopsis cylindrus
Fragilariopsis kerguelensis L26-C5
Frittschiella tuberosa
Fucus vesiculosus
Furcellaria lumbricalis
Fusarium graminearum
Galdieria sulphuraria
Gelidium elegans
Gelidium vagum
Geminigera cryophila CCMP2564
Gephyrocapsa oceanica RCC1303
Giardia lamblia
Glaucocystis nostochinearum
Glaucosphaera vacuolata
Gloeochaete wittrockiana
Gloiopeltis furcata
Golenkinia longispicula
Goniomonas sp.
Gonium pectorale
Gracilaria changii
Gracilaria chilensis
Gracilaria chouae

Gracilaria salicornia
Gracilaria tenuistipitata
Gracilariopsis lemaneiformis
Grammatophora oceanica CCMP 410
Grateloupia filicina
Grateloupia taiwanensis
Gregarina niphandroides
Griffithsia okiensis
Guillardia theta
Gymnochlora sp. CCMP2014
Gymnochlora stellata
Gymnodinium catenatum GC744
Haematococcus lacustris
Hafniomonas reticulata
Halochlorococcum marinum
Hanusia phi CCMP325
Helicodictyon planctonicum
Helobdella robusta
Hemiselmis andersenii CCMP439
Heterochlamydomonas inaequalis
Heterosigma akashiwo CCMP2393
Heterosigma akashiwo NIES293
Heterosiphonia pulchra
Histoplasma capsulatum
Homo sapiens
Ignatius tetrasporus
Isochrysis sp. CCMP1324
Laccaria bicolor
Leishmania infantum
Leishmania major strain Friedlin
Leptocylindrus danicus CCMP1856
Leptosira obovata
Lithodesmium undulatum
Litonotus pictus P1
Lobochlamys segnis
Lobomonas rostrata
Lotharella amoebiformis
Lotharella globosa
Lotharella vacuolata
Madagascaria erythrocladiodes CCMP3234
Mayorella sp.
Mazzaella japonica
Melampsora larici-populina
Micromonas pusilla CCMP1545
Micromonas sp. RCC299
Microthamnion kuetzingianum
Mimulus guttatus
Mitosporidium daphniae
Monomorphina aenigmatica
Monosiga brevicollis
Mucor circinelloides
Mycosphaerella graminicola
Naegleria gruberi
Nannochloropsis gaditana
Nannochloropsis granulata
Nematostella vectensis
Neobodo designis CCAP1951
Neoparamoeba aestuarina
Neosiphonia japonica
Neospora caninum
Nephroselmis olivacea
Neurospora crassa OR74A
Norrisiella sphaerica BC52
Ochromonas sp. CCMP1393
Odontella aurita isolate 1302-5
Odontella sinensis
Oedogonium cardiacum
Oryza sativa
Ostreococcus lucimarinus
Ostreococcus tauri
Palpitomonas bilix NIES-2562
Pandorina morum
Parachlorella kessleri
Paramecium tetraurelia
Paramoeba atlantica
Paraphysomonas imperforata PA2
Partenskyella glossopodia RCC365
Pavlova lutheri
Pavlova sp. CCMP459
Pedinomonas minor
Pelagococcus subviridis CCMP1429
Pelagomonas calceolata CCMP1756
Percolomonas cosmopolitus AE-1
Percursaria percursa
Perkinsus marinus
Pessonella sp.
Phacotus lenticularis
Phaeodactylum tricornutum
Phaeomonas parva CCMP2877
Phanerochaete chrisosporium
Phycomyces blakesleeanus
Physcomitrella patens
Phytophthora capsici
Phytophthora sojae
Picochlorum oklahomensis CCMP2329
Picocystis salinarum CCMP1897
Pinguicoccus pyrenoidosus CCMP2078
Pirula salina
Planophila laetevirens
Plasmodium falciparum 3D7
Plasmodium vivax North Korea
Plasmodium yoelii 17XNL
Pleurochrysis carterae CCMP645
Polarella glacialis CCMP1383
Polysphondylium pallidum
Polytomella parva SAG 63-3
Populus trichocarpa
Porphyra pulchra
Porphyra purpurea
Porphyra sp.
Porphyra yezoensis
Porphyridium cruentum

Porphyridium purpureum
Postia placenta
Prasinococcus capsulatus CCMP1194
Prasinoderma coloniale CCMP1413
Prasiola crispa
Proboscia alata PI-D3
Proboscia inermis CCAP1064-1
Procryptobia sorokini
Proteomonas sulcata CCMP704
Prototheca wickerhamii
Prymnesium parvum Texoma1
Pseudoneochloris marina
Pseudopedinella elastica CCMP716
Pseudoscourfieldia marina
Pteridomonas danica PT
Pteromonas angulosa
Puccinia graminis
Pycnococcus provasolii RCC2336
Pycnococcus sp. CCMP1998
Pyramimonas parkeae CCMP726
Pyropia haitanensis
Pyropia perforata
Rhizopus oryzae
Rhizosolenia imbricata
Rhizosolenia setigera CCMP 1694
Rhodella maculata CCMP736
Rhodella violacea
Rhodochaete parvula
Rhodomonas abbreviata
Rhodomonas salina
Rhodorus marinus CCMP769
Rhynchomonas nasuta
Rosalina sp.
Roundia cardiophora
Rozella allomycis CSF55
Saccharina japonica
Saccharomyces cerevisiae RM11-1a
Salpingoeca rosetta
Sargassum horneri
Schizochytrium aggregatum ATCC28209
Schizophyllum commune
Schizosaccharomyces pombe
Scrippsiella trochoidea CCMP3099
Selaginella moellendorffii
Sorites sp.
Spermatozopsis exsultans

Spizellomyces punctatus
Sporobolomyces roseus
Sporolithon durum
Stagonospora nodorum
Stephanosphaera pluvialis
Stereomyxa ramosa
Stichococcus bacillaris
Stigeoclonium helveticum
Striatella unipunctata CCMP2910
Stygamoeba regulata BSH-02190019
Synchroma pusillum CCMP3072
Tetrademus dimorphus
Tetrahymena thermophila
Tetraselmis chuii PLY429
Thalassiosira pseudonana
Thalassiothrix antarctica L6-D1
Thecamonas trahens
Theileria orientalis Shintoku
Theileria parva Muguga
Thraustochytrium sp. LLF1b
Timspurckia oligopyrenoides CCMP3278
Toxoplasma gondii GT1
Trachelomonas volvocina
Trebouxia arboricola
Trentepohlia annulata
Trichoderma reesei
Trichomonas vaginalis G3
Trichoplax adhaerens Grell-BS-1999
Trichosphaerium sp.
Triparma laevis
Trypanosoma brucei TREU927
Trypanosoma cruzi strain CL Brener
Ulnaria acus
Ulvella endozoica
Undaria pinnatifida
Unidentified eukaryote sp. NY0313808BC1
Uronema belkai
Ustilago maydis
Vannella robusta
Vannella sp.
Vaucheria litorea CCMP2940
Vertebrata lanosa
Vexillifera sp.
Vitrella brassicaformis CCMP3346
Volvox carteri
Yarrowia lipolytica

Plastids and nucleomorphs

Alveolata sp. CCMP3155
Arabidopsis thaliana
Asterionella formosa
Asterionellopsis glacialis
Aureococcus anophagefferens
Aureoumbra lagunensis CCMP1507
Bangia atropurpurea

Bigelowiella natans
Brachypodium distachyon
Calliarthron tuberculatum
Cerataulina daemon
Chaetoceros simplex
Chlamydomonas reinhardtii
Chlorella vulgaris C-169
Chondrus crispus

Chromera velia
Cryptoglana skujai
Cryptomonas paramecium CCAP977
Cyanidioschyzon merolae
Cyanidium caldarium
Cyanophora paradoxa
Cylindrotheca closterium
Didymosphenia geminata
Ectocarpus siliculosus
Emiliana huxleyi CCMP1516
Euglenaformis proxima
Euglena gracilis
Euglena longa
Euglenaria anabaena
Eunotia naegeli
Eutreptiella gymnastica
Fistulifera solaris
Fucus vesiculosus
Gelidium elegans
Gelidium vagum
Gracilaria chilensis
Gracilaria salicornia
Gracilaria tenuistipitata
Gracilariopsis lemaneiformis
Grateloupia taiwanensis
Guillardia theta
Gymnochlora stellata
Heterosigma akashiwo NIES293
Lithodesmium undulatum
Lotharella vacuolata
Micromonas pusilla CCMP1545
Micromonas sp. RCC299
Monomorphina aenigmatica
Nannochloropsis gaditana
Nannochloropsis granulata
Odontella sinensis
Oryza sativa
Ostreococcus tauri
Partenskyella glossopodia RCC365
Pavlova lutheri
Phaeodactylum tricorutum
Physcomitrella patens
Populus trichocarpa
Porphyra pulchra
Porphyra purpurea
Porphyra yezoensis
Porphyridium purpureum
Pyropia haitanensis
Pyropia perforata
Rhizosolenia imbricata
Rhodomonas salina
Roundia cardiophora
Saccharina japonica
Sargassum horneri
Selaginella moellendorffii
Sporolithon durum
Thalassiosira pseudonana
Toxoplasma gondii GT1
Trachelomonas volvocina
Triparma laevis
Ulnaria acus
Undaria pinnatifida
Vaucheria litorea CCMP2940
Vertebrata lanosa

Table S2. List of outgroup cyanobacterial species excluded from final alignments

<i>Acaryochloris marina</i> MBIC11017	<i>Fischerella thermalis</i> PCC 7521
<i>Anabaena</i> sp. PCC 7108	<i>Geminocystis herdmanii</i> PCC 6308
<i>Arthrospira maxima</i> CS-328	<i>Gloeocapsa</i> sp. PCC 73106
<i>Arthrospira platensis</i> NIES-39	<i>Laptolyngbya</i> sp. PCC 7376
<i>Arthrospira</i> sp. PCC 8005	<i>Leptolyngbya</i> sp. PCC 6406
<i>Calothrix</i> sp. PCC 7103	<i>Leptolyngbya</i> sp. PCC 7375
<i>Chamaesiphon minutus</i> PCC 6605	<i>Lyngbya</i> sp. PCC8106
<i>Chlorogloeopsis fritschii</i> PCC 6912	<i>Mastigocladopsis repens</i> PCC 10914
<i>Chroococcidiopsis</i> sp. PCC 6712	<i>Microchaete</i> sp. PCC 7126
<i>Coleofasciculus chthonoplastes</i> PCC 7420	<i>Microcoleus vaginatus</i> FGP-2
<i>Crocospaera watsonii</i> WH8501	<i>Microcystis aeruginosa</i> NIES843
<i>Cyanobacterium</i> sp. PCC 10605	<i>Moorea producens</i> 3L
<i>Cyanobacterium</i> sp. PCC 7702	<i>Nodosilinea nodulosa</i> PCC 7104
<i>Cyanobacterium stanieri</i> PCC 7202	<i>Nostoc punctiforme</i> 11
<i>Cyanobacterium Yellowstone B</i>	<i>Nostoc</i> sp. PCC 7120
<i>Cyanothece</i> sp. ATCC51142	<i>Nostoc</i> sp. PCC 7524
<i>Cyanothece</i> sp. CCY0110	<i>Oscillatoria</i> sp. PCC 10802
<i>Cyanothece</i> sp. ATCC 51472	<i>Oscillatoria</i> sp. PCC 7112
<i>Cyanothece</i> sp. PCC 7424	<i>Oscillatoria</i> sp. PCC 6506
<i>Cyanothece</i> sp. PCC 7425	<i>Pleurocapsa</i> sp. PCC 7319
<i>Cyanothece</i> sp. PCC 7822	<i>Pseudanabaena</i> sp. PCC 6802
<i>Cyanothece</i> sp. PCC 8801	<i>Pseudanabaena</i> sp. PCC 7429
<i>Cylindrospermopsis raciborskii</i> CS-505	<i>Raphidiopsis brookii</i> D9
<i>Cylindrospermum stagnale</i> PCC 7417	<i>Richelia intracellularis</i> HH01
<i>Dactylococcopsis salina</i> PCC 8305	<i>Rubidibacter lacunae</i> KORDI 51-2
<i>Fischerella muscicola</i> PCC 7414	<i>Spirulina major</i> PCC 6313
<i>Fischerella</i> sp. JSC-11	<i>Stanieria</i> sp. PCC 7437
<i>Fischerella</i> sp. PCC 9339	<i>Thermosynechococcus elongatus</i> BP1
<i>Fischerella</i> sp. PCC 9431	<i>Tolypfthrix</i> sp. PCC 9009
<i>Fischerella</i> sp. PCC 9605	<i>Trichormus azollae</i> 708

Table S3. Phylogenetic affinity of genes in secondary photosynthetic eukaryotes inferred from single-gene phylogenies (see Figs. S1-S85)

Colors and letters (R,G,U,A) indicate the position of secondary photosynthetic eukaryotes in individual gene trees: close to red algae (R), green algae (G), unresolved (U) or absent (A).

Tree	Outgroup	Cryptophyta	Haptophyta	Alveolata	Stramenopiles	Euglenida	Chlorarach.
RG_002	Cyanobacteria	R	R	A	R	A	R
RG_003	Cyanobacteria	R	A	R	R	R	R
RG_005	Cyanobacteria	U	U	U	U	U	U
RG_006	Cyanobacteria	R	R	R	R	G	R
RG_007	Cyanobacteria	R	R	R	R	G	R
RG_008	Cyanobacteria	R	A	R	R	G	G
RG_010	Cyanobacteria	R	R	R	R	G	R
RG_013	Cyanobacteria	R	R	R	R	R	R
RG_014	Cyanobacteria	R	R	R	R	U	R
RG_015	Cyanobacteria	R	R	R	R	R	G
RG_016	Cyanobacteria	R	R	R	R	U	R
RG_017	Cyanobacteria	R	R	R	R	R	G
RG_018	Cyanobacteria	R	R	R	R	A	R
RG_019	Cyanobacteria	R	R	U	R	G	R
RG_020	Cyanobacteria	R	R	R	R	G	G
RG_022	Cyanobacteria	R	R	A	R	A	R
RG_024	Cyanobacteria	R	R	A	R	G	A
RG_025	Cyanobacteria	R	R	R	R	U	G
RG_026	Cyanobacteria	R	R	A	R	A	G
RG_029	Cyanobacteria	R	R	R	R	A	G
RG_030	Cyanobacteria	R	R	R	R	G	A
RG_031	Cyanobacteria	R	R	A	R	R	R
RG_032	Cyanobacteria	R	R	R	R	G	G
RG_033	Cyanobacteria	R	R	R	R	R	G
RG_034	Cyanobacteria	R	R	U	R	R	R
RG_035	Cyanobacteria	R	R	A	R	A	U
RG_036	Bacteria	R	R	A	R	A	R
RG_037	Cyanobacteria	R	R	A	R	A	R
RG_038	Cyanobacteria	U	R	A	R	A	R
RG_039	Cyanobacteria	R	R	R	R	G	G
RG_040	Cyanobacteria	R	R	R	R	R	G
RG_041	Cyanobacteria	R	R	R	R	G	R
RG_042	Cyanobacteria	U	R	R	R	G	R
RG_047	Cyanobacteria	R	R	A	R	R	R
RG_048	Cyanobacteria	G	G	G	G	G	G
RG_049	Cyanobacteria	R	R	R	R	U	R
RG_050	Cyanobacteria	R	R	A	R	G	R
RG_053	Bacteria	R	R	A	R	R	G
RG_057	Cyanobacteria	R	R	R	R	G	R
RG_058	Cyanobacteria	R	R	R	R	R	R
RG_059	Cyanobacteria	U	R	R	R	G	R

RG_060	Cyanobacteria	R	R	R	R	A	R
RG_061	Cyanobacteria	R	R	R	R	R	A
RG_062	Cyanobacteria	U	U	A	U	R	U
RG_064	Cyanobacteria	U	U	U	U	A	U
RG_066	Cyanobacteria	R	G	G	G	G	G
RG_067	Cyanobacteria	R	R	R	R	G	G
RG_070	Bacteria	R	R	A	R	R	R
RG_071	Cyanobacteria	A	R	R	R	A	R
RG_072	Bacteria	R	R	R	R	A	R
RG_073	Bacteria	R	R	R	R	R	G
RG_074	Cyanobacteria	G	G	G	G	G	A
RG_075	Cyanobacteria	R	R	R	R	A	G
RG_076	Cyanobacteria	G	G	G	G	G	G
RG_079	Bacteria	R	R	R	R	A	R
RG_081	Cyanobacteria	R	R	R	R	R	R
RG_082	Cyanobacteria	G	R	A	R	R	R
RG_083	Cyanobacteria	R	R	R	R	A	R
RG_084	Cyanobacteria	R	U	A	U	A	U
RG_085	Cyanobacteria	R	R	R	R	R	G
RG_086	Cyanobacteria	R	R	R	R	A	G
RG_087	Cyanobacteria	R	R	R	R	G	G
RG_088	Cyanobacteria	R	R	R	R	G	G
RG_089	Cyanobacteria	R	R	R	R	R	G
RG_090	Cyanobacteria	R	R	R	R	R	G
RG_092	Cyanobacteria	R	R	R	R	U	G
RG_093	Cyanobacteria	R	R	R	R	R	R
RG_095	Cyanobacteria	R	R	A	R	G	G
RG_096	Cyanobacteria	R	R	U	R	G	U
RG_097	Cyanobacteria	R	R	A	R	A	R
RG_101	Cyanobacteria	R	R	A	R	U	R
RG_102	Cyanobacteria	R	R	R	R	G	R
RG_104	Cyanobacteria	A	R	R	R	G	R
RG_105	Cyanobacteria	R	R	U	R	R	G
RG_106	Cyanobacteria	R	R	A	R	A	G
RG_107	Bacteria	R	R	R	R	R	R
RG_108	Cyanobacteria	R	R	R	R	G	R
RG_110	Cyanobacteria	R	R	R	R	U	R
RG_111	Cyanobacteria	R	R	R	R	G	R
RG_112	Bacteria	R	R	A	R	A	R
RG_113	Bacteria	R	R	R	R	R	G
RG_114	Bacteria	R	R	R	R	G	G
RG_115	Bacteria	A	R	R	R	G	A
RG_116	Cyanobacteria	G	G	A	G	G	G

Table S4. EggNOG functional annotation of genes transferred from Archaeplastida to secondary photosynthetic eukaryotes

Code	Reference sequence	Functional class	Function description
RG_002	AT1G24360.1	Lipid transport and metabolism	reductase
RG_003	AT4G14070.1	Lipid transport and metabolism	long-chain-fatty-acid--CoA ligase
RG_005	AT3G51820.1	Coenzyme transport and metabolism	bacteriochlorophyll chlorophyll a synthase
RG_006	AT5G38660.1	Function unknown	Protein of unknown function (DUF2854)
RG_007	AT3G62910.1	Translation, ribosomal structure and biogenesis	Peptide chain release factor 1 directs the termination of translation in response to the peptide chain termination codons UAG and UAA
RG_008	AT5G15450.1	Post-translational modification, protein turnover, and chaperones	Part of a stress-induced multi-chaperone system, it is involved in the recovery of the cell from heat-induced damage, in cooperation with DnaK, DnaJ and GrpE. Acts before DnaK, in the processing of protein aggregates. Protein binding stimulates the ATPase activity
RG_010	AT3G08010.1	Function unknown	Protein of unknown function (DUF1092)
RG_013	AT5G64840.1	Function unknown	ABC transporter
RG_014	AT1G32500.1	Post-translational modification, protein turnover, and chaperones	feS assembly protein SufD
RG_015	AT1G15700.1	Energy production and conversion	Produces ATP from ADP in the presence of a proton gradient across the membrane. The gamma chain is believed to be important in regulating ATPase activity and the flow of protons through the CF(0) complex
RG_016	AT3G60620.1	Lipid transport and metabolism	Phosphatidate cytidyltransferase
RG_017	AT4G25080.3	Coenzyme transport and metabolism	Mg-protoporphyrin IX methyl transferase
RG_018	AT1G74470.1	Coenzyme transport and metabolism	geranylgeranyl reductase
RG_019	AT4G34350.1	Lipid transport and metabolism	Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP)
RG_020	AT4G35440.1	Inorganic ion transport and metabolism	Cl-channel, voltage gated
RG_022	CMB059C	Inorganic ion transport and metabolism	aluminum resistance protein
RG_024	AT4G17740.2	Cell wall/membrane/envelope biogenesis	Protease
RG_025	AT5G21920.1	Function unknown	integral membrane protein
RG_026	AT5G22620.1	Carbohydrate transport and metabolism	phosphoglycerate mutase

RG_029	AT3G16950.2	Energy production and conversion	Dihydrolipoyl dehydrogenase
RG_030	AT3G10270.1	Replication, recombination and repair	DNA gyrase negatively supercoils closed circular double-stranded DNA in an ATP-dependent manner and also catalyzes the interconversion of other topological isomers of double-stranded DNA rings, including catenanes and knotted rings (By similarity)
RG_031	AT3G10690.1	Replication, recombination and repair	DNA gyrase negatively supercoils closed circular double-stranded DNA in an ATP-dependent manner and also catalyzes the interconversion of other topological isomers of double-stranded DNA rings, including catenanes and knotted rings (By similarity)
RG_032	AT2G22360.1	Post-translational modification, protein turnover, and chaperones	ATP binding to DnaK triggers the release of the substrate protein, thus completing the reaction cycle. Several rounds of ATP-dependent interactions between DnaJ, DnaK and GrpE are required for fully efficient folding. Also involved, with DnaK and GrpE, in the DNA replication of plasmids through activation of initiation proteins
RG_033	AT1G55930.1	Inorganic ion transport and metabolism	Domain of unknown function DUF21
RG_034	AT5G62790.1	Lipid transport and metabolism	Catalyzes the NADP-dependent rearrangement and reduction of 1-deoxy-D-xylulose-5-phosphate (DXP) to 2-C-methyl-D-erythritol 4-phosphate (MEP) (By similarity)
RG_035	AT3G08740.1	Translation, ribosomal structure and biogenesis	Involved in peptide bond synthesis. Stimulates efficient translation and peptide-bond synthesis on native or reconstituted 70S ribosomes in vitro. Probably functions indirectly by altering the affinity of the ribosome for aminoacyl-tRNA, thus increasing their reactivity as acceptors for peptidyl transferase (By similarity)
RG_036	AT5G16715.1	Translation, ribosomal structure and biogenesis	amino acids such as threonine, to avoid such errors, it has a posttransfer editing activity that hydrolyzes mischarged Thr-tRNA(Val) in a tRNA-dependent manner (By similarity)
RG_037	CMO352C	Replication, recombination and repair	helicase
RG_038	AT2G04842.1	Translation, ribosomal structure and biogenesis	threonyl-tRNA synthetase
RG_039	AT3G12080.2	Nucleotide transport and metabolism	GTPase that plays an essential role in the late steps of ribosome biogenesis (By similarity)
RG_040	AT2G30390.1	Coenzyme transport and metabolism	Catalyzes the ferrous insertion into protoporphyrin IX (By similarity)
RG_041	AT5G42270.1	Post-translational modification, protein turnover, and chaperones	Acts as a processive, ATP-dependent zinc metallopeptidase for both cytoplasmic and membrane proteins. Plays a role in the quality control of integral membrane proteins
RG_042	AT4G36810.1	Coenzyme transport and metabolism	Polyprenyl synthetase
RG_047	AT1G56050.1	Translation, ribosomal structure and biogenesis	gtp-binding protein
RG_048	AT3G59400.1	Function unknown	GUN4 domain protein

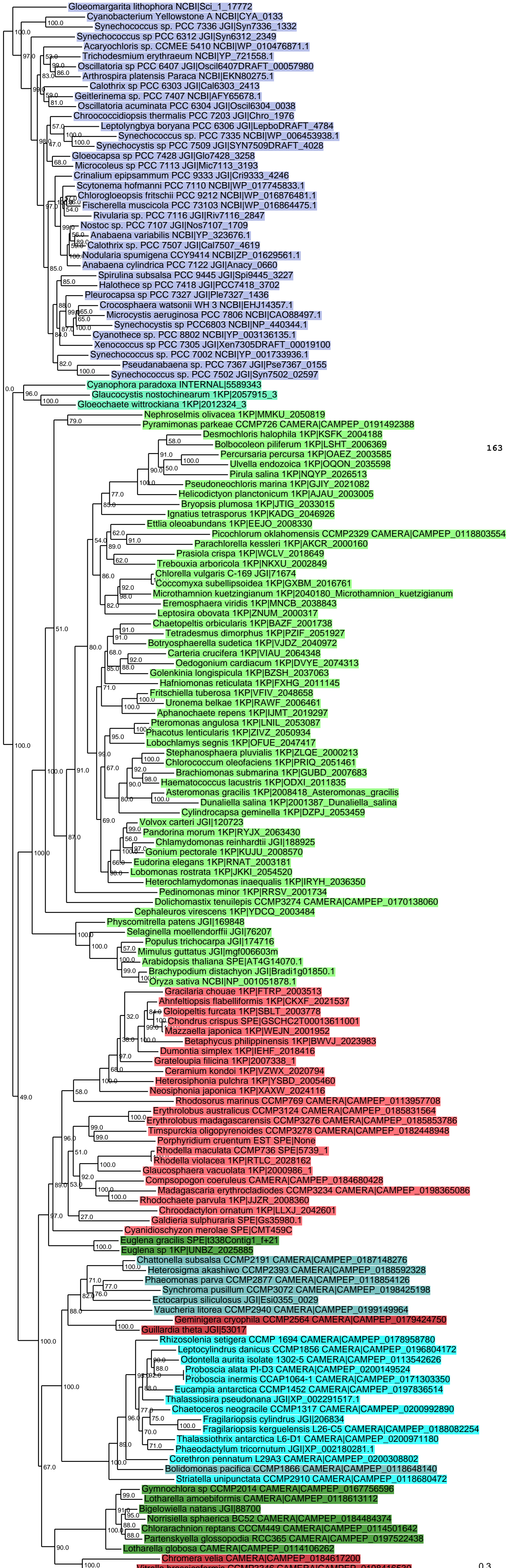
RG_049	AT5G13630.2	Coenzyme transport and metabolism	cobaltochelataase, cobn subunit
RG_050	AT1G58290.1	Coenzyme transport and metabolism	Catalyzes the NADPH-dependent reduction of glutamyl-tRNA(Glu) to glutamate 1-semialdehyde (GSA)
RG_053	CMR268C	Post-translational modification, protein turnover, and chaperones	Redox regulated molecular chaperone. Protects both thermally unfolding and oxidatively damaged proteins from irreversible aggregation. Plays an important role in the bacterial defense system toward oxidative stress
RG_057	AT3G17930.1	Function unknown	Protein of unknown function (DUF3007)
RG_058	AT2G18950.1	Coenzyme transport and metabolism	Tocopherol phytyltransferase
RG_059	AT3G11945.1	Coenzyme transport and metabolism	Tocopherol phytyltransferase
RG_060	AT2G43945.1	Function unknown	NA
RG_061	AT1G78620.1	Function unknown	Integral membrane protein
RG_062	AT5G49030.1	Translation, ribosomal structure and biogenesis	isoleucyltRNA synthetase
RG_064	AT1G62640.1	Lipid transport and metabolism	Catalyzes the condensation reaction of fatty acid synthesis by the addition to an acyl acceptor of two carbons from malonyl-ACP. Catalyzes the first condensation reaction which initiates fatty acid synthesis and may therefore play a role in governing the total rate of fatty acid production. Possesses both acetoacetyl-ACP synthase and acetyl transacylase activities. Its substrate specificity determines the biosynthesis of branched-chain and or straight-chain of fatty acids
RG_066	AT5G08650.1	Cell wall/membrane/envelope biogenesis	Required for accurate and efficient protein synthesis under certain stress conditions. May act as a fidelity factor of the translation reaction, by catalyzing a one-codon backward translocation of tRNAs on improperly translocated ribosomes. Back-translocation proceeds from a post-translocation (POST) complex to a pre-translocation (PRE) complex, thus giving elongation factor G a second chance to translocate the tRNAs correctly. Binds to ribosomes in a GTP-dependent manner
RG_067	AT3G24590.1	Intracellular trafficking, secretion, and vesicular transport	Signal peptidase i
RG_070	AT1G62750.1	Translation, ribosomal structure and biogenesis	Catalyzes the GTP-dependent ribosomal translocation step during translation elongation. During this step, the ribosome changes from the pre-translocational (PRE) to the post-translocational (POST) state as the newly formed A-site-bound peptidyl-tRNA and P-site-bound deacylated tRNA move to the P and E sites, respectively. Catalyzes the coordinated movement of the two tRNA molecules, the mRNA and conformational changes in the ribosome
RG_071	WP_010473488.1	Translation, ribosomal structure and biogenesis	Increases the formation of ribosomal termination complexes and stimulates activities of RF-1 and RF-2. It binds guanine nucleotides and has strong preference for UGA stop codons. It may interact directly with the ribosome. The stimulation of RF-1 and RF-2 is significantly reduced by GTP and GDP, but not by GMP

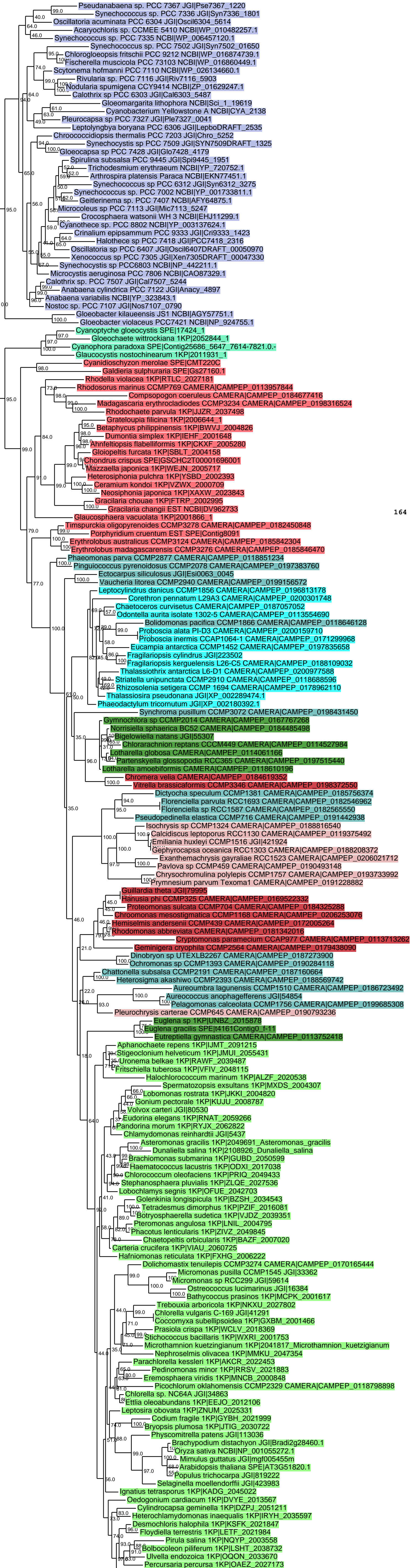
RG_072	AT5G08280.1	Coenzyme transport and metabolism	Tetrapolymerization of the monopyrrole PBG into the hydroxymethylbilane pre-uroporphyrinogen in several discrete steps
RG_073	AT4G15560.1	Coenzyme transport and metabolism	Catalyzes the acyloin condensation reaction between C atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield 1-deoxy-D-xylulose-5-phosphate (DXP)
RG_074	AT1G10830.1	Function unknown	NnrU family
RG_075	AT5G08740.1	Energy production and conversion	faD-dependent pyridine nucleotide-disulfide oxidoreductase
RG_076	AT1G08490.1	Amino acid transport and metabolism	Cysteine desulfurase
RG_079	AT5G18570.1	Energy production and conversion	An essential GTPase which binds GTP, GDP and possibly (p)ppGpp with moderate affinity, with high nucleotide exchange rates and a fairly low GTP hydrolysis rate. It may play a role in control of the cell cycle, stress response, ribosome biogenesis and in those bacteria that undergo differentiation, in morphogenesis control
RG_081	AT2G25840.2	Translation, ribosomal structure and biogenesis	Tryptophanyl-tRNA synthetase
RG_082	AT3G13490.1	Translation, ribosomal structure and biogenesis	lysyl-tRNA synthetase
RG_083	AT2G24200.1	Amino acid transport and metabolism	Presumably involved in the processing and regular turnover of intracellular proteins. Catalyzes the removal of unsubstituted N-terminal amino acids from various peptides
RG_084	AT5G36170.1	Translation, ribosomal structure and biogenesis	Peptide chain release factor 2 directs the termination of translation in response to the peptide chain termination codons UGA and UAA
RG_085	AT1G08520.1	Coenzyme transport and metabolism	magnesium chelatase
RG_086	AT3G27110.1	Post-translational modification, protein turnover, and chaperones	peptidase, M48
RG_087	AT2G20890.1	Cell cycle control, cell division, chromosome partitioning	May be involved in photosynthetic membrane biogenesis
RG_088	AT3G50820.1	Cell wall/membrane/envelope biogenesis	photosystem II manganese-stabilizing
RG_089	AT5G17230.1	Lipid transport and metabolism	phytoene synthase
RG_090	AT1G65260.1	Transcription, Signal transduction mechanisms	Phage shock protein A
RG_092	AT4G02790.1	Transcription	Required for a late step of 50S ribosomal subunit assembly. Has GTPase activity
RG_093	AT3G25920.1	Translation, ribosomal structure and biogenesis	Binds to the 23S rRNA

RG_095	AT5G30510.1	Translation, ribosomal structure and biogenesis	30S ribosomal protein S1
RG_096	AT3G63190.1	Translation, ribosomal structure and biogenesis	Responsible for the release of ribosomes from messenger RNA at the termination of protein biosynthesis. May increase the efficiency of translation by recycling ribosomes from one round of translation to another
RG_097	AT5G24490.1	Translation, ribosomal structure and biogenesis	ribosomal subunit Interface protein
RG_101	AT5G04590.1	Energy production and conversion	Sulfite reductase
RG_102	AT5G01220.1	Cell wall/membrane/envelope biogenesis	Glycosyl transferase (Group 1
RG_104	AT4G26500.1	Function unknown	Fe-S metabolism associated SufE
RG_105	AT5G52970.1	Function unknown	Inherit from NOG: 15.0 kDa protein
RG_106	AT5G55220.1	Post-translational modification, protein turnover, and chaperones	Involved in protein export. Acts as a chaperone by maintaining the newly synthesized protein in an open conformation
RG_107	AT5G64050.1	Translation, ribosomal structure and biogenesis	Catalyzes the attachment of glutamate to tRNA(Glu) in a two-step reaction glutamate is first activated by ATP to form Glu-AMP and then transferred to the acceptor end of tRNA(Glu)
RG_108	AT5G23120.1	Function unknown	The ortholog in <i>A.thaliana</i> is involved in photosystem II (PSII) assembly, but knockout of the corresponding gene in <i>Synechococcus</i> PCC 7002 has no effect on PSII activity
RG_110	AT3G04870.2	Function unknown	desaturase
RG_111	AT2G32480.1	Cell wall/membrane/envelope biogenesis	Membrane-associated zinc metalloprotease
RG_112	AT2G22650.1	Amino acid transport and metabolism	oxidoreductase
RG_113	AT2G25830.1	Transcription	transcriptional regulatory protein
RG_114	AT2G02500.1	Lipid transport and metabolism	Catalyzes the formation of 4-diphosphocytidyl-2-C- methyl-D-erythritol from CTP and 2-C-methyl-D-erythritol 4- phosphate (MEP)
RG_115	AT5G13650.1	Signal transduction mechanisms	gtp-binding protein tyra
RG_116	AT3G26710.1	Function unknown	Protein of unknown function (DUF3529)



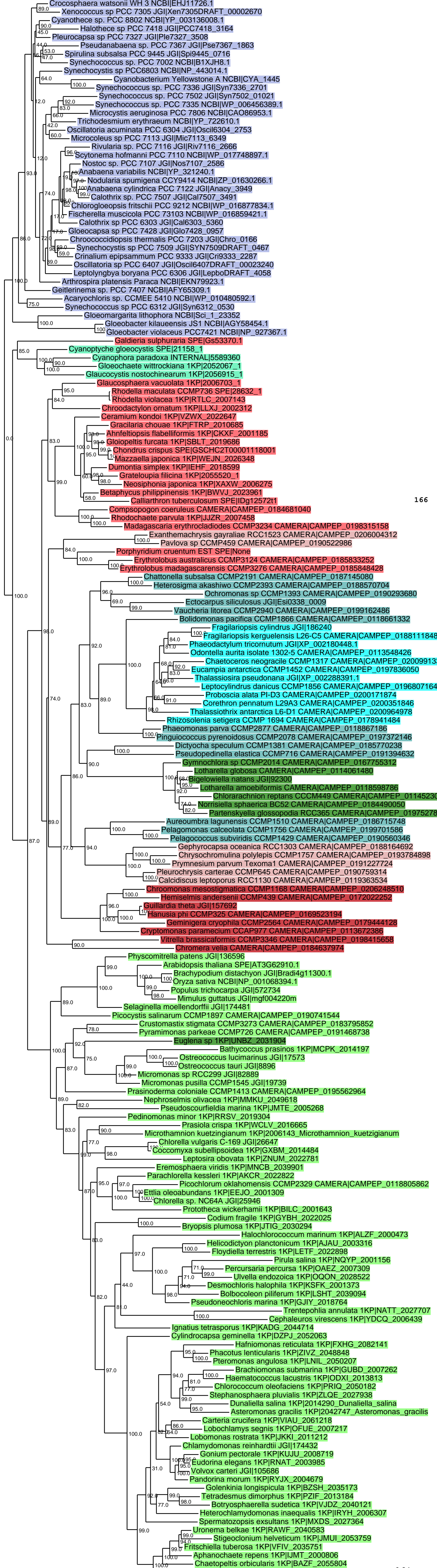
162





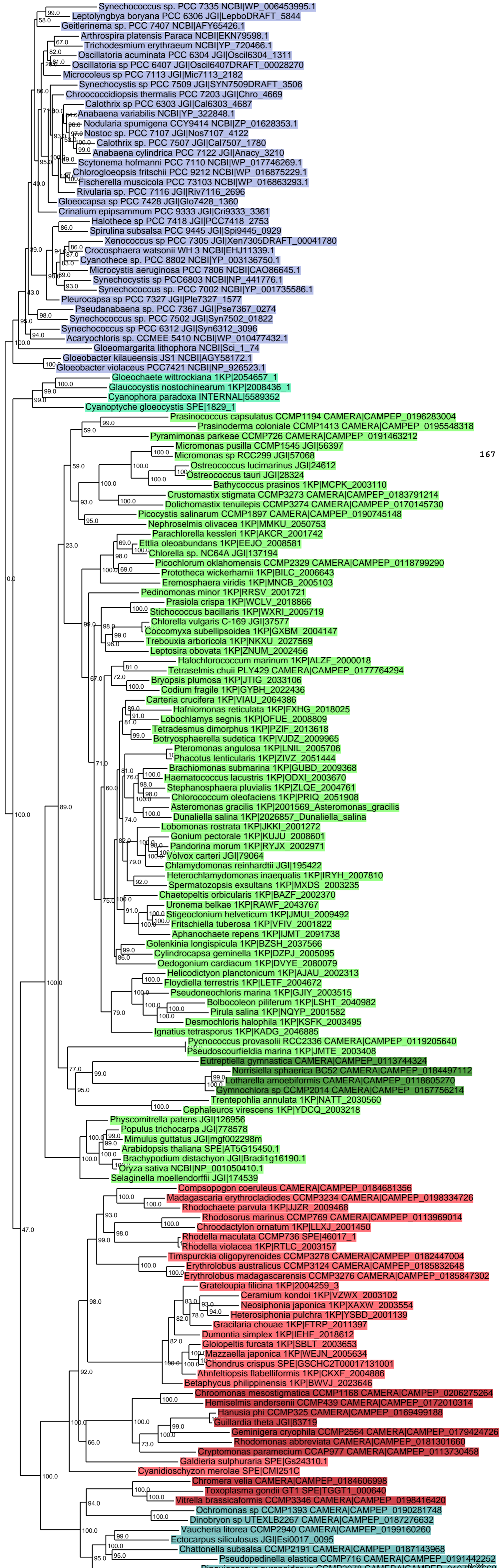


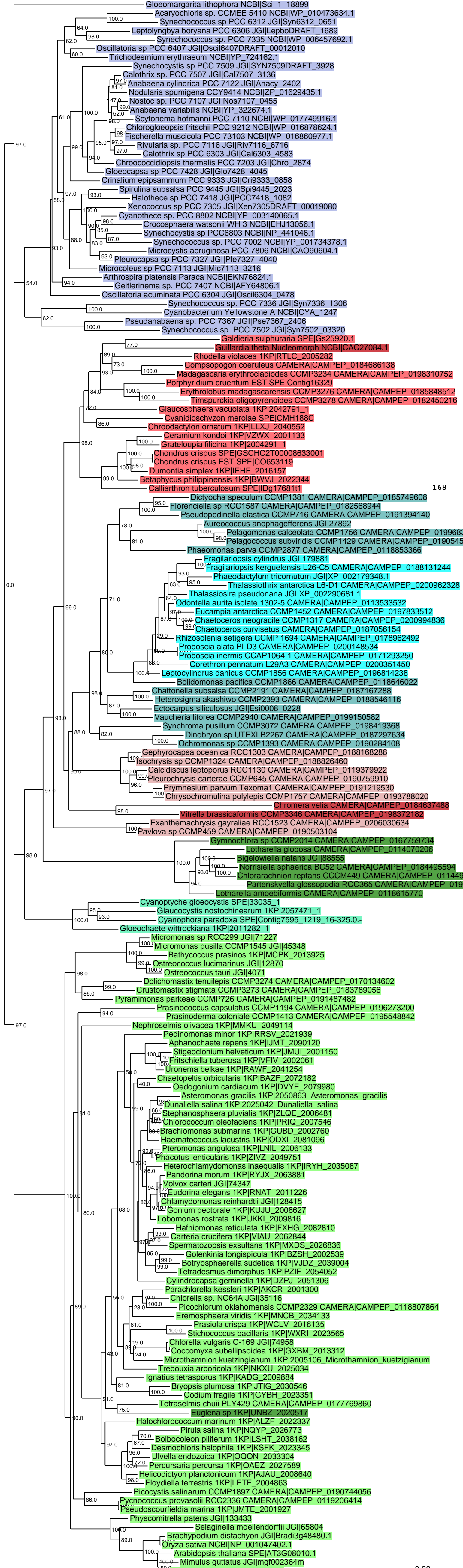
165



166

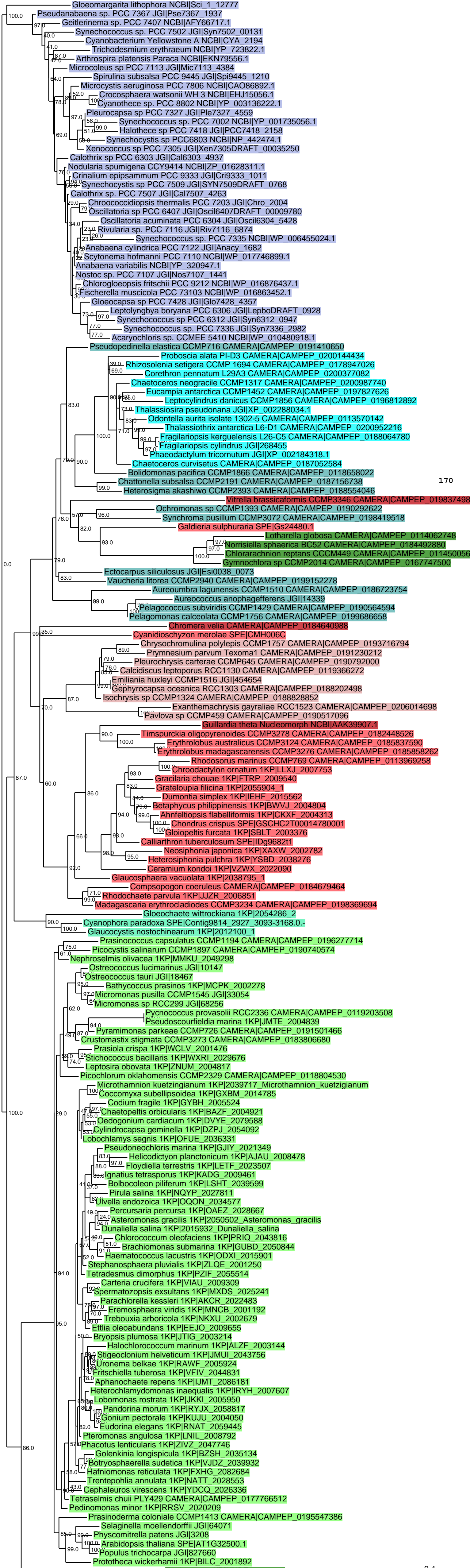
0.24



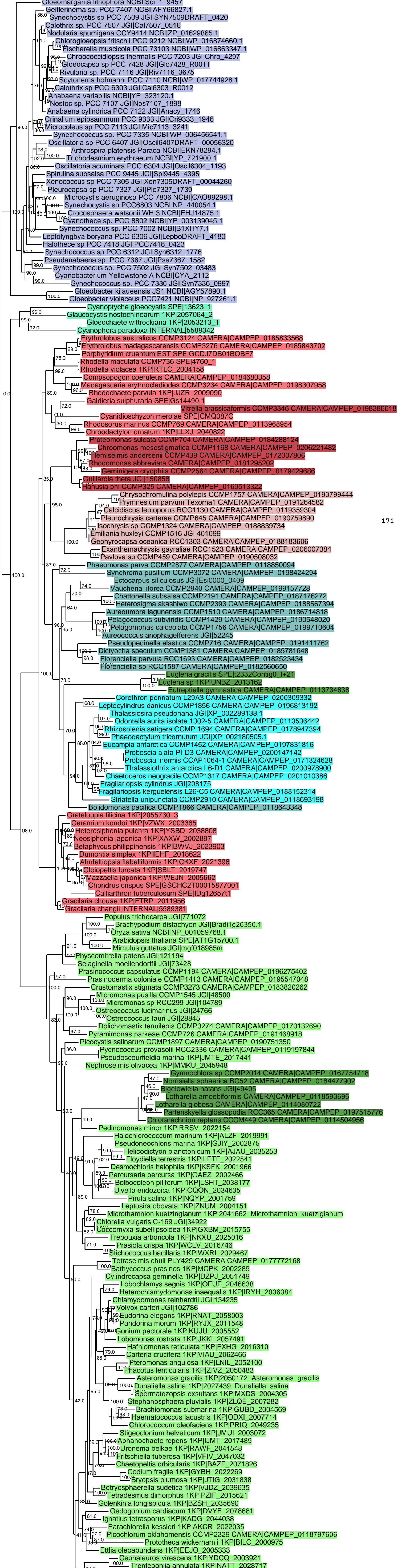


168





170



171

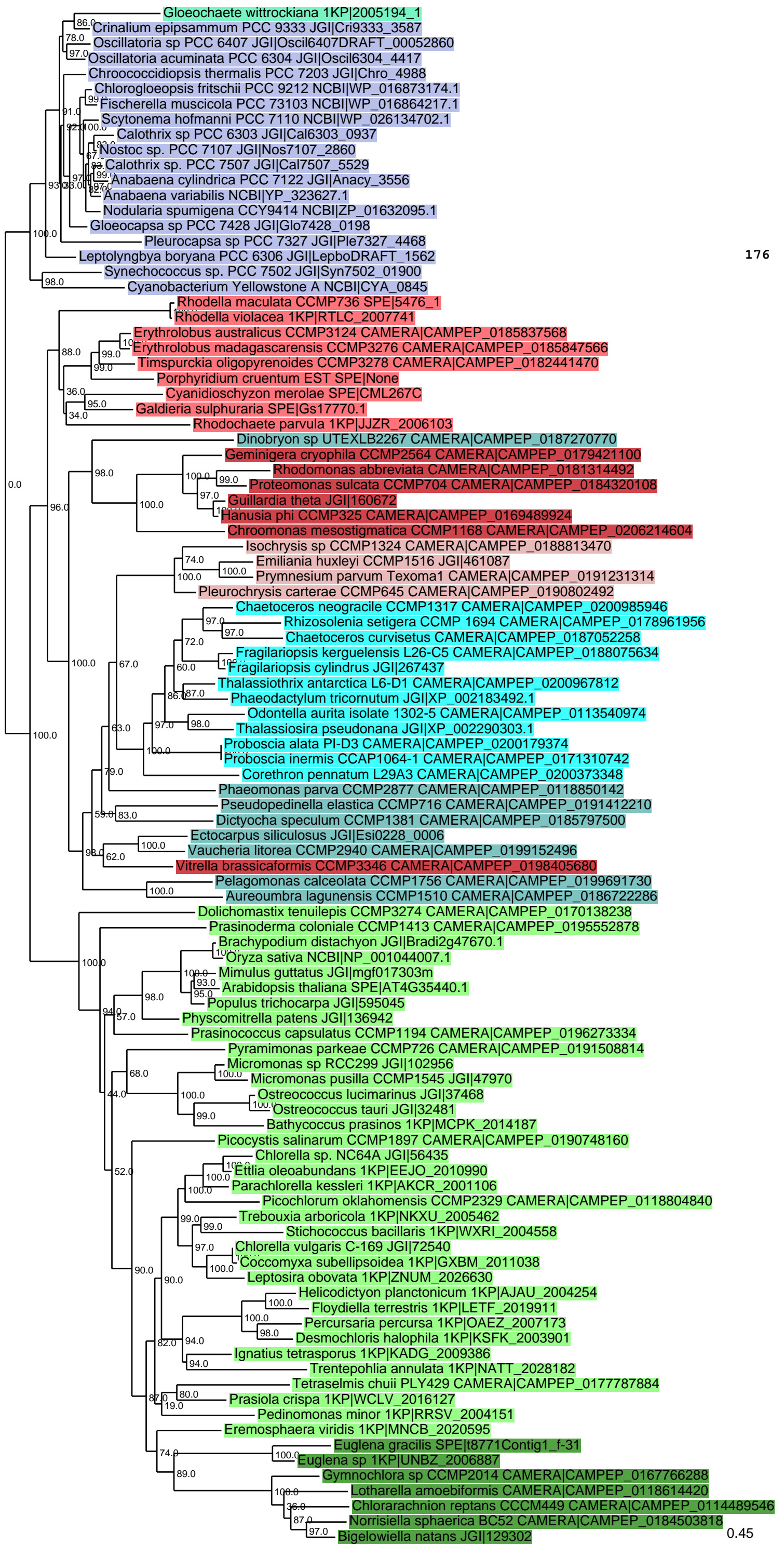


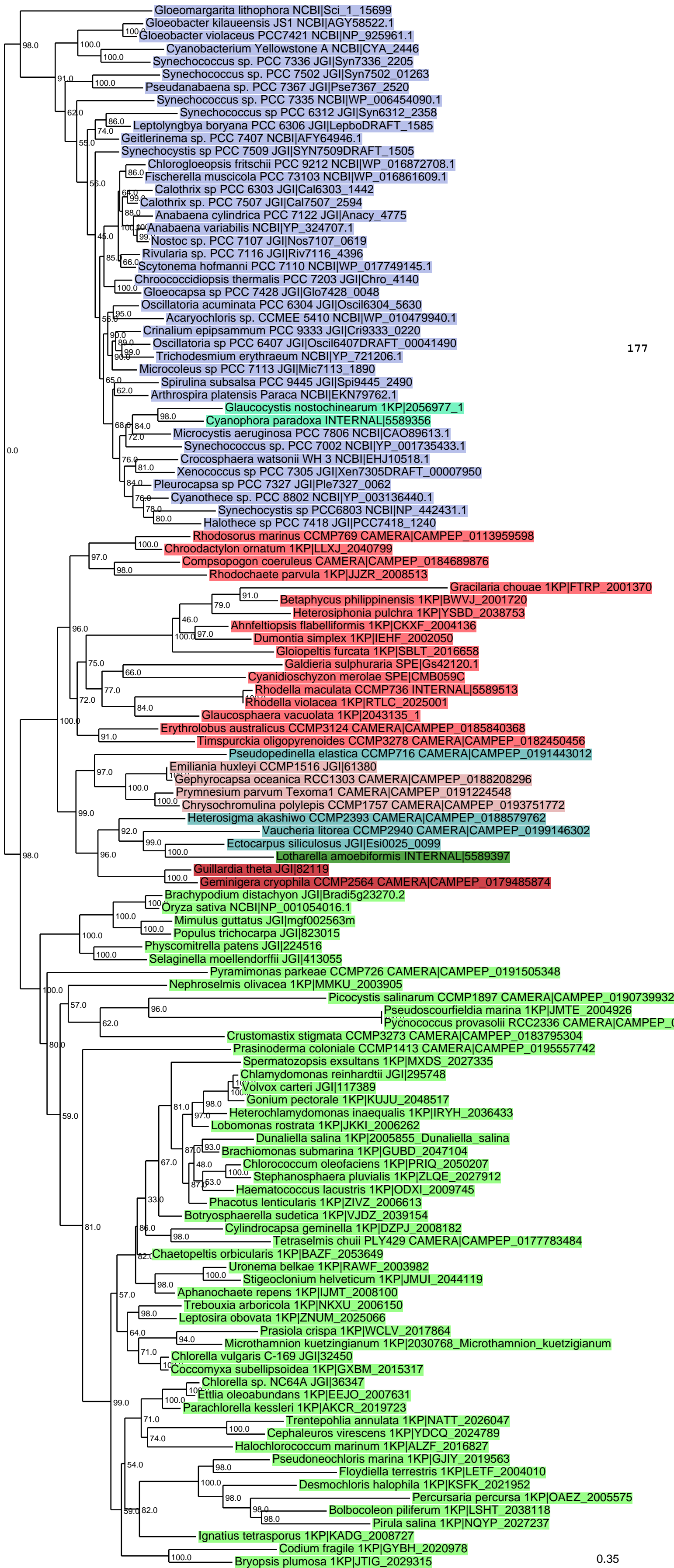
172

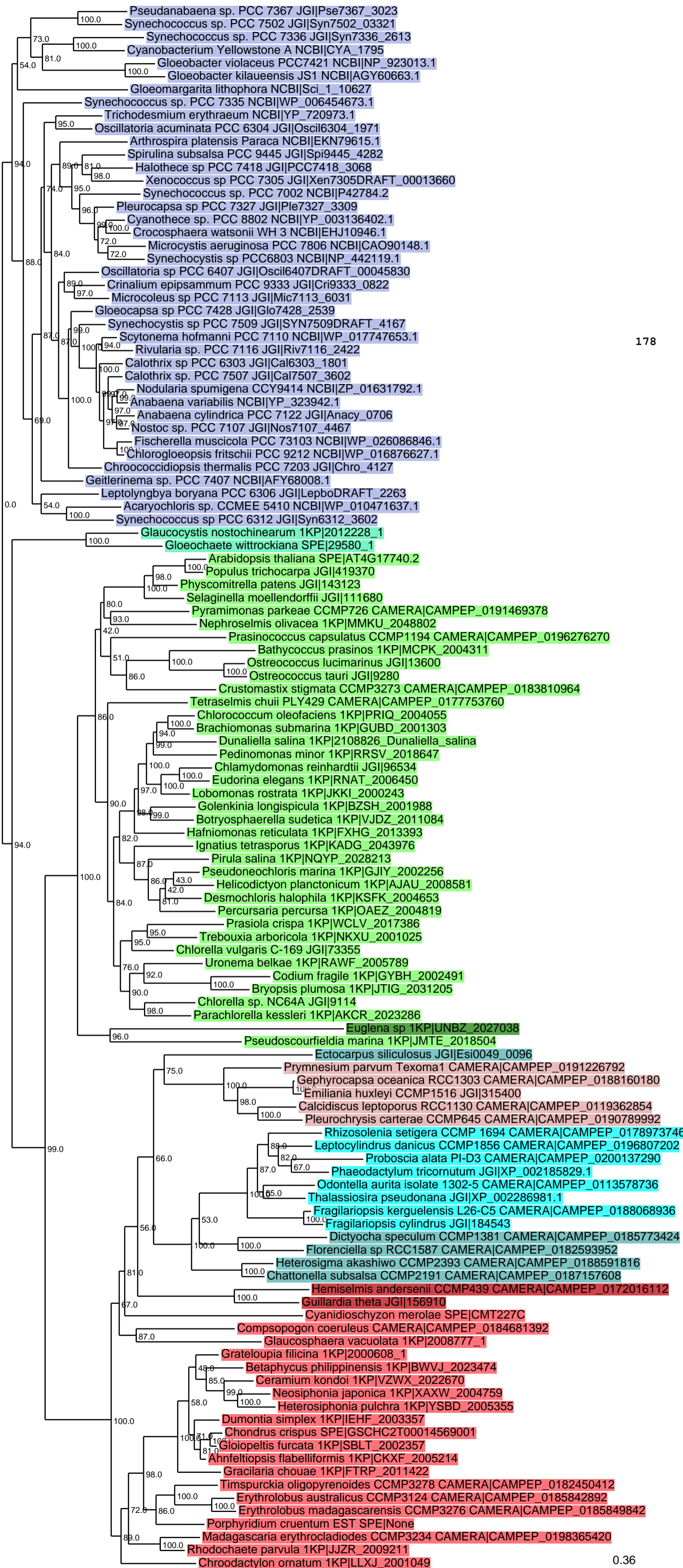


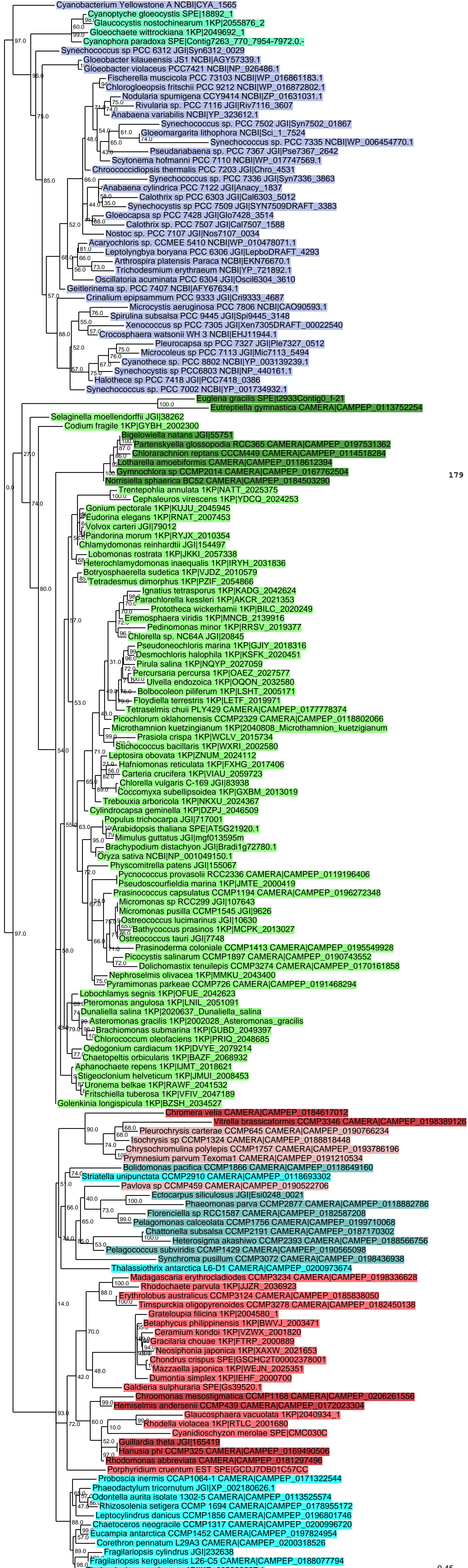


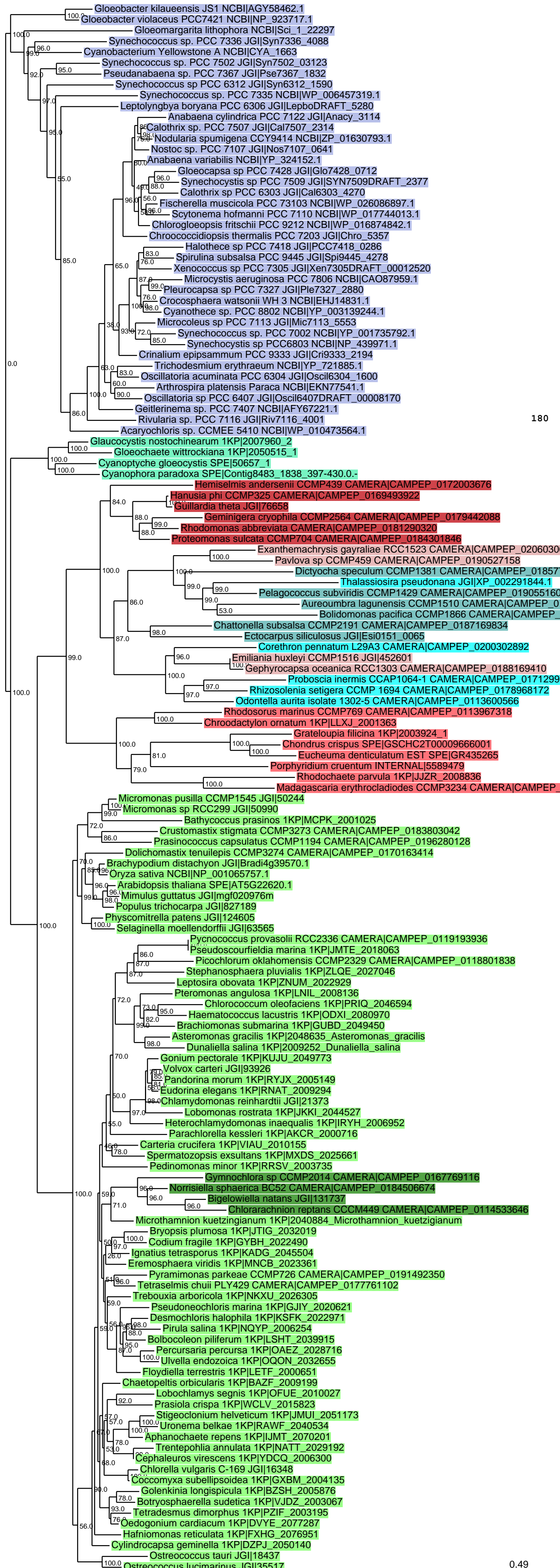


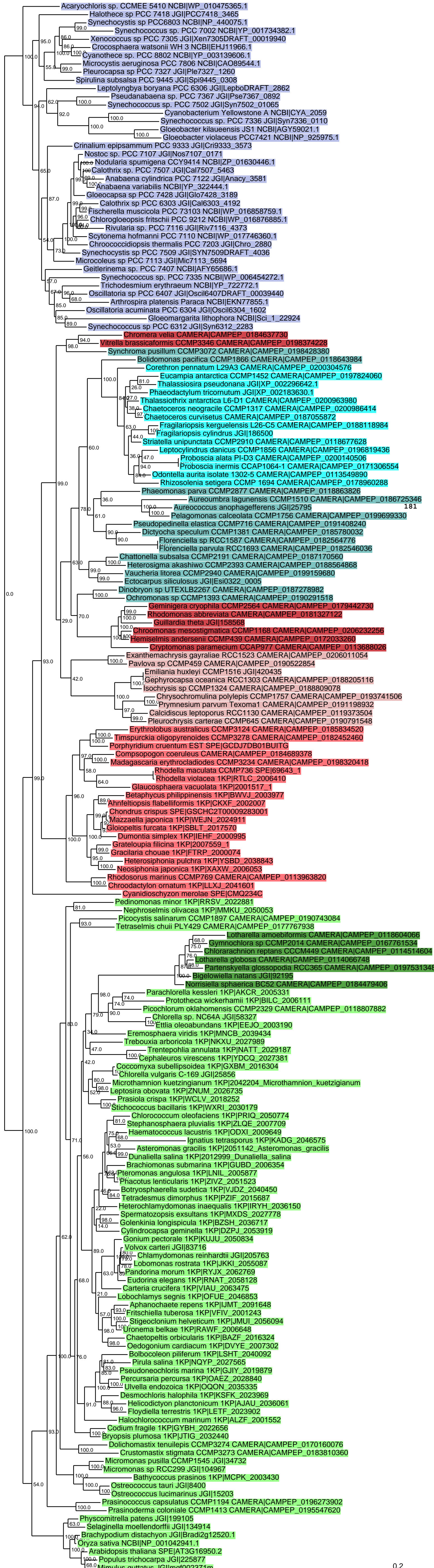


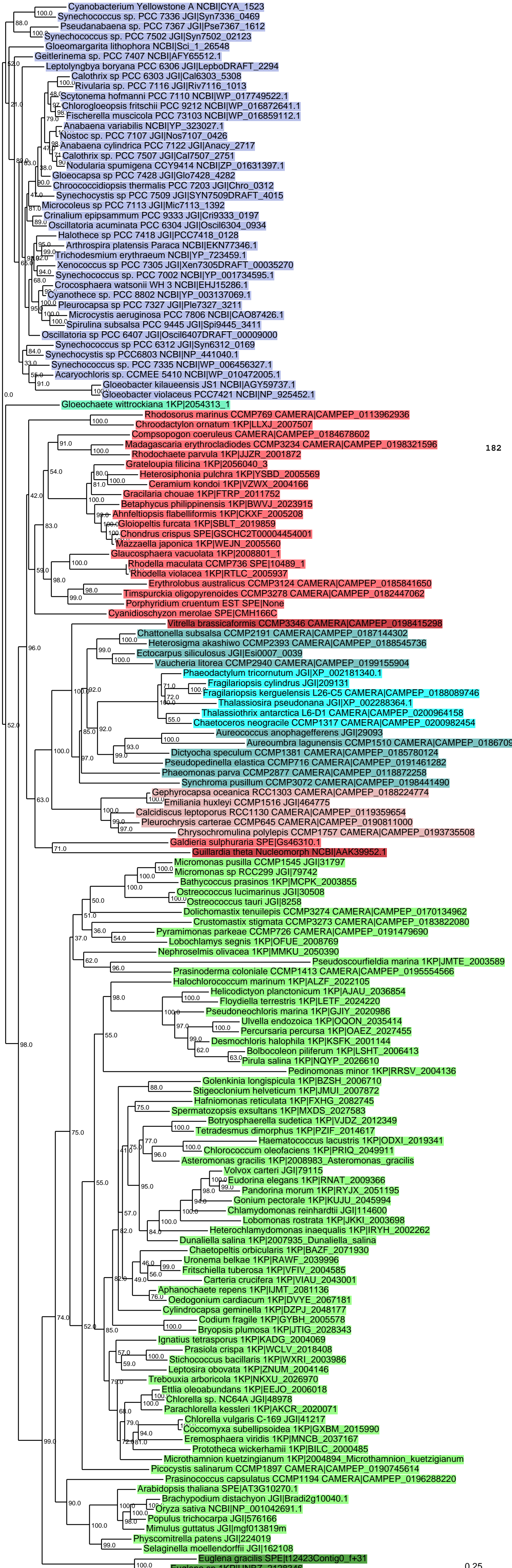


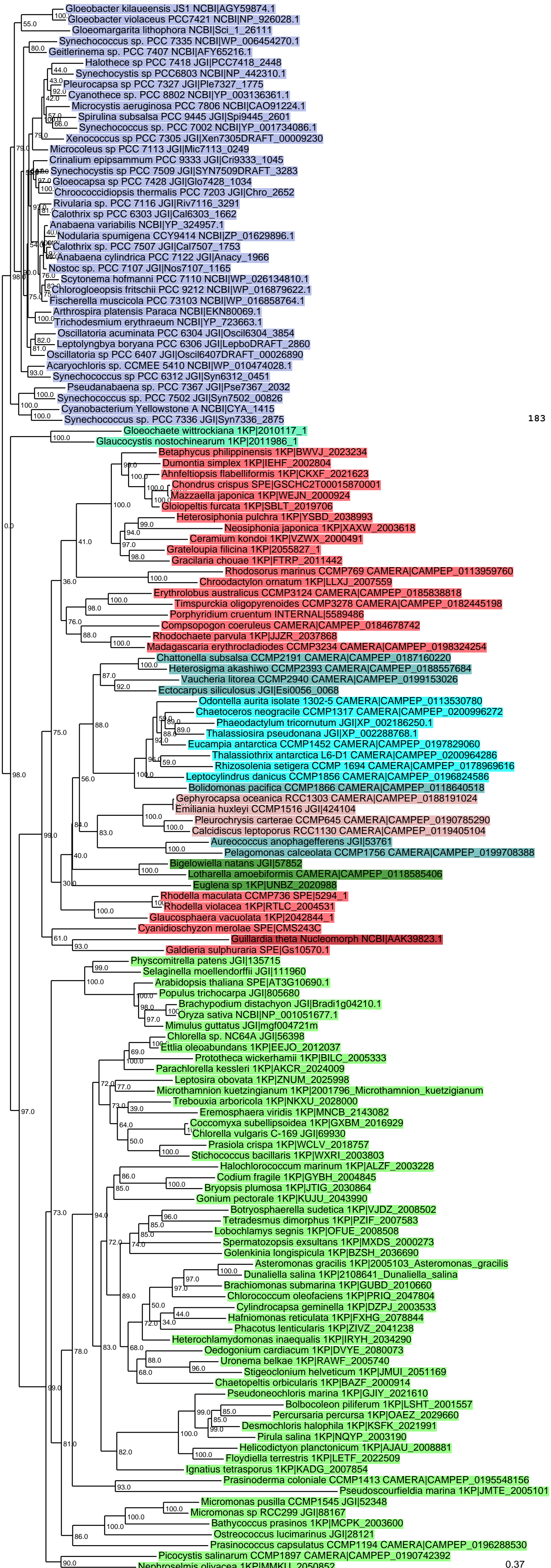






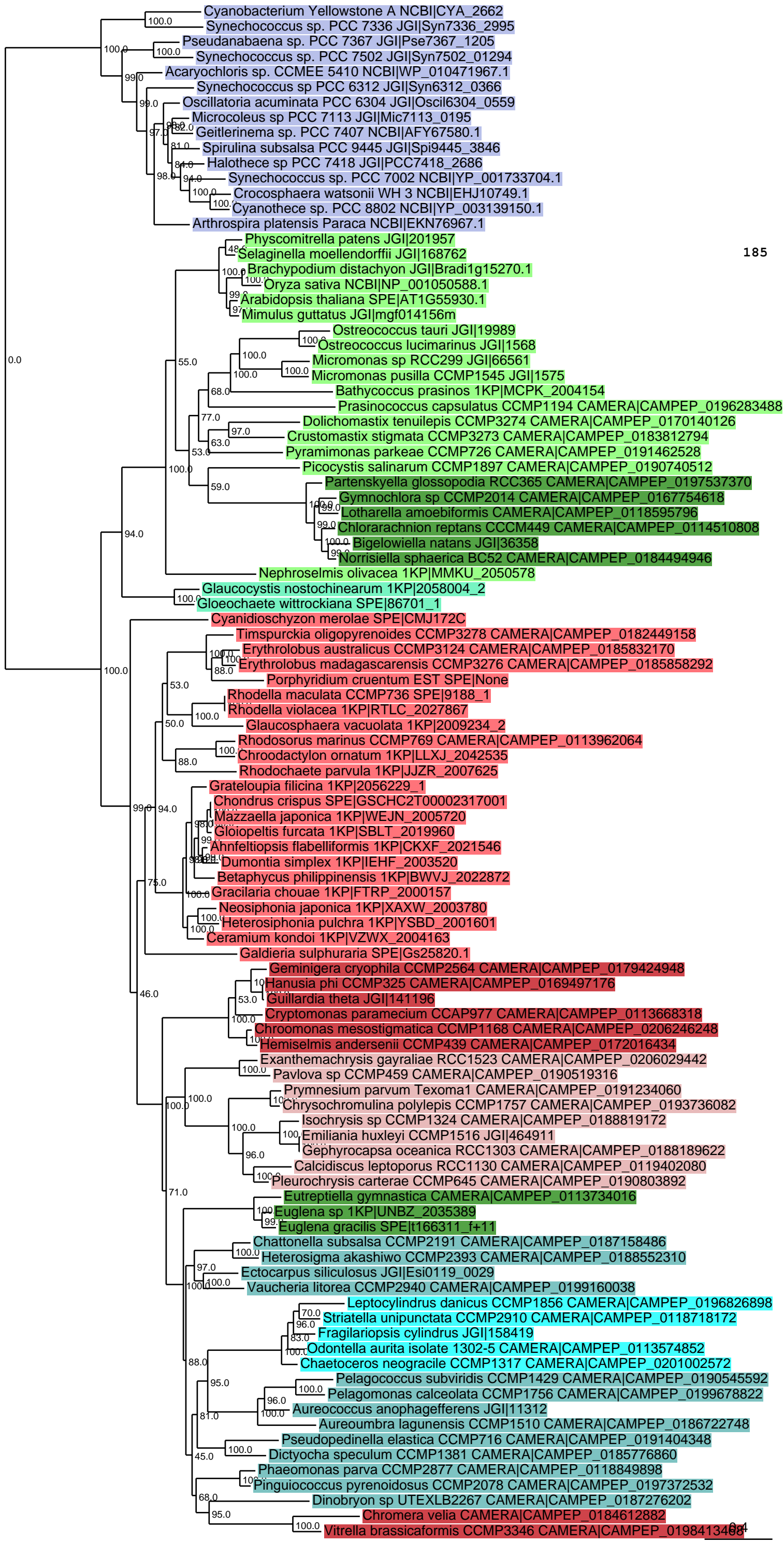






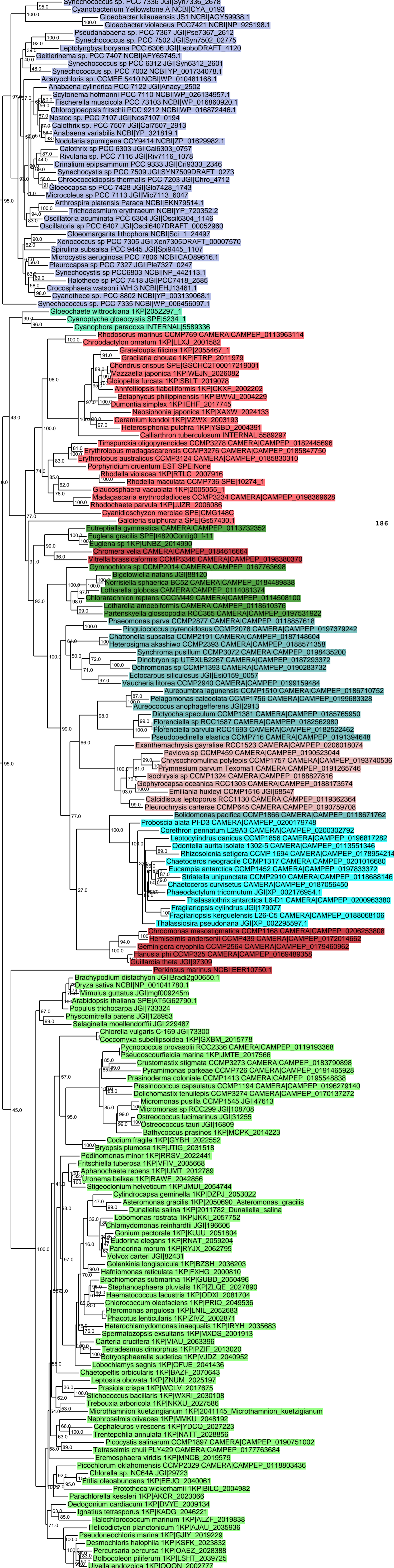


184

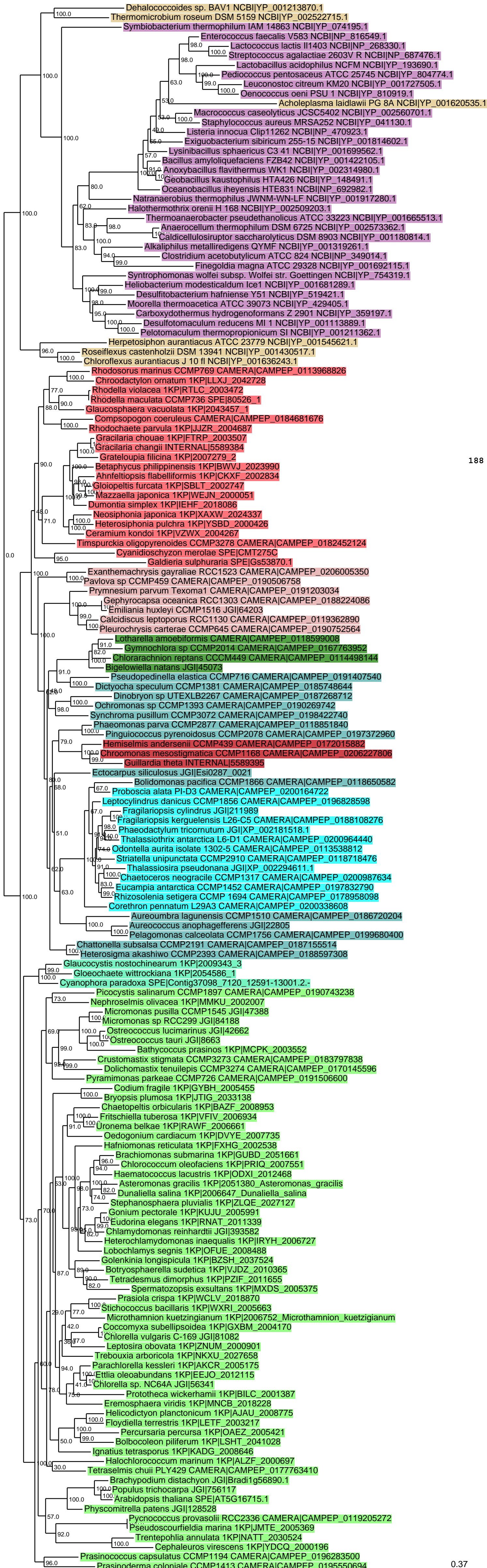


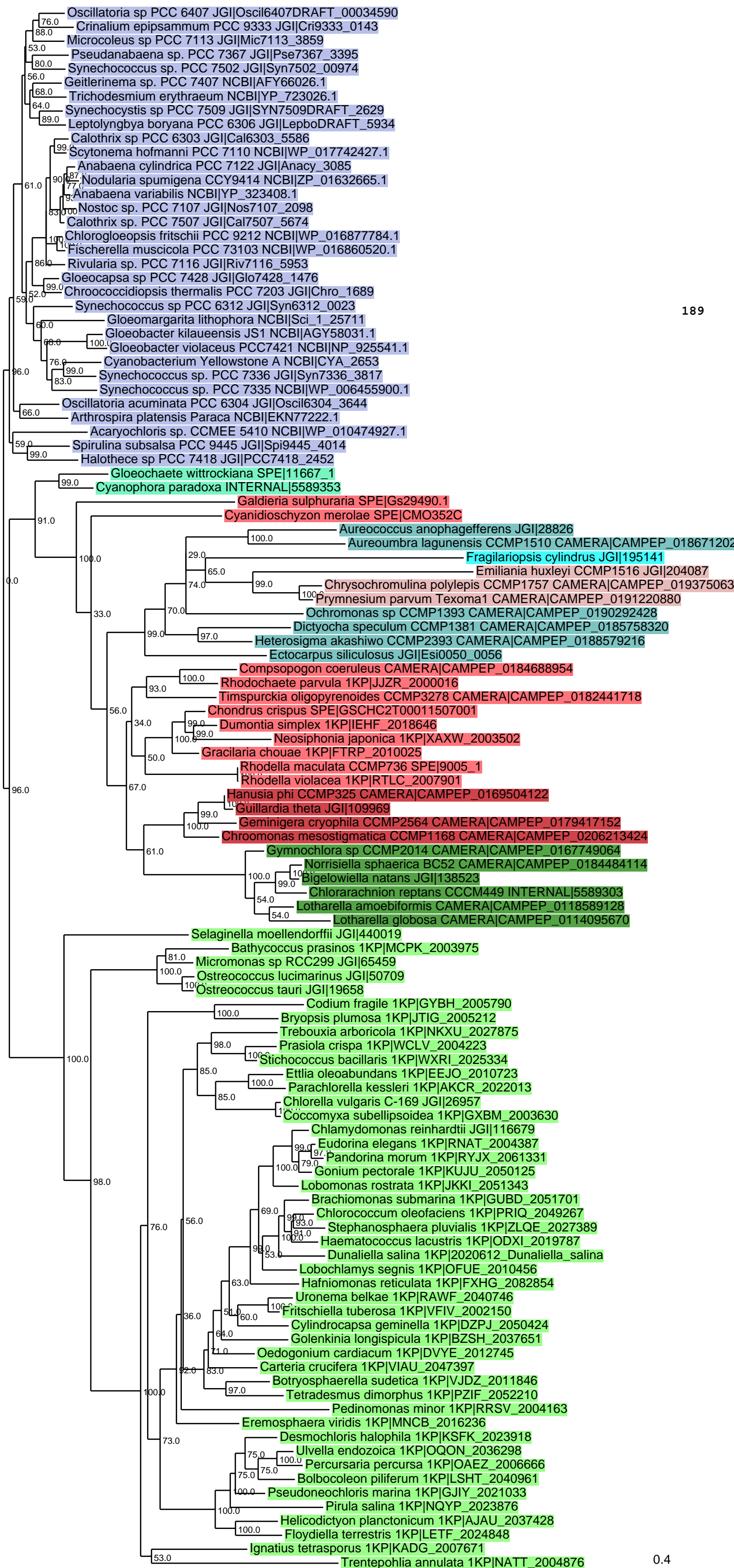
0.0

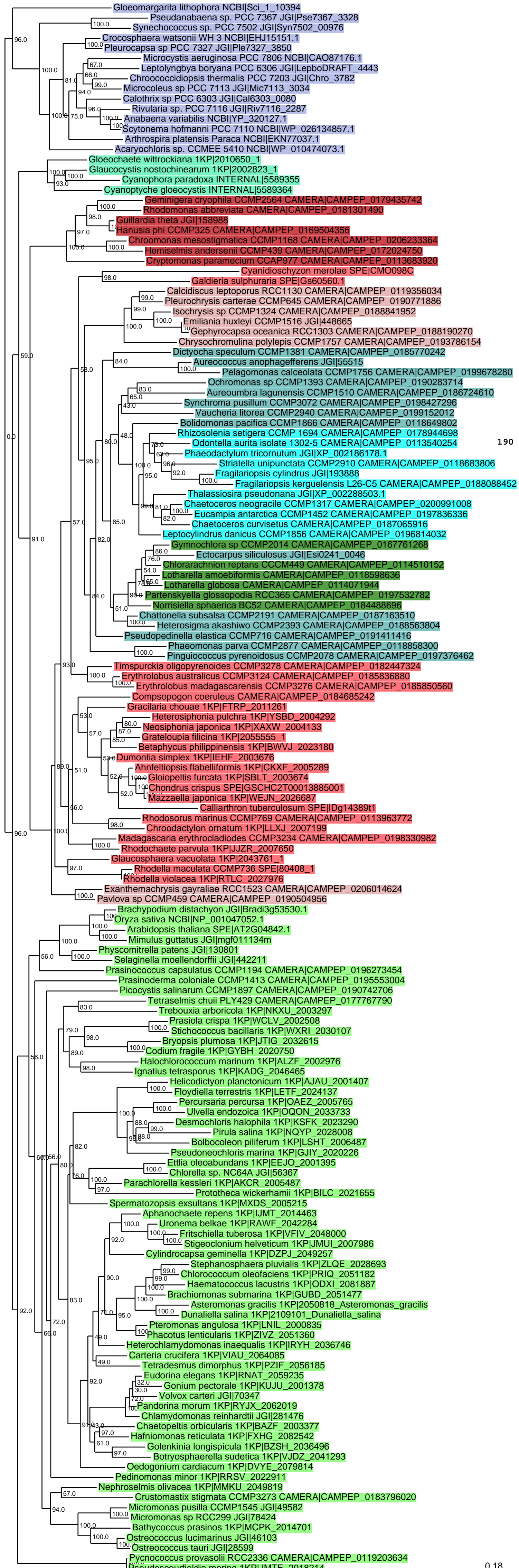
- Cyanobacterium Yellowstone A NCBI|CYA_2662
- Synechococcus sp. PCC 7336 JGI|Syn7336_2995
- Pseudanabaena sp. PCC 7367 JGI|Pse7367_1205
- Synechococcus sp. PCC 7502 JGI|Syn7502_01294
- Acaryochloris sp. CCME5 5410 NCBI|WP_010471967.1
- Synechococcus sp. PCC 6312 JGI|Syn6312_0366
- Oscillatoria acuminata PCC 6304 JGI|Oscil6304_0559
- Microcoleus sp. PCC 7113 JGI|Mic7113_0195
- Geitlerinema sp. PCC 7407 NCBI|AFY67580.1
- Spirulina subsalsa PCC 9445 JGI|Spi9445_3846
- Halotheca sp. PCC 7418 JGI|PCC7418_2686
- Synechococcus sp. PCC 7002 NCBI|YP_001733704.1
- Crocospaera watsonii WH 3 NCBI|EHJ10749.1
- Cyanothece sp. PCC 8802 NCBI|YP_003139150.1
- Arthrospira platensis Paraca NCBI|EKN76967.1
- Physcomitrella patens JGI|201957
- Selaginella moellendorffii JGI|168762
- Brachypodium distachyon JGI|Bradi1g15270.1
- Oryza sativa NCBI|NP_001050588.1
- Arabidopsis thaliana SPE|AT1G55930.1
- Mimulus guttatus JGI|mgf014156m
- Ostreococcus tauri JGI|19989
- Ostreococcus lucimarinus JGI|1568
- Micromonas sp. RCC299 JGI|66561
- Micromonas pusilla CCMP1545 JGI|1575
- Bathycoccus prasinos 1KP|MCPK_2004154
- Prasinococcus capsulatus CCMP1194 CAMERA|CAMPEP_0196283488
- Dolichomastix tenuilepis CCMP3274 CAMERA|CAMPEP_0170140126
- Crustomastix stigmata CCMP3273 CAMERA|CAMPEP_0183812794
- Pyramimonas parkeae CCMP726 CAMERA|CAMPEP_0191462528
- Picocystis salinarum CCMP1897 CAMERA|CAMPEP_0190740512
- Partenskyella glossopodia RCC365 CAMERA|CAMPEP_0197537370
- Gymnochloa sp. CCMP2014 CAMERA|CAMPEP_0167754618
- Lotharella amoebiformis CAMERA|CAMPEP_0118595796
- Chlorarachnion reptans CCCC449 CAMERA|CAMPEP_0114510808
- Bigelowiella natans JGI|36358
- Norrisiella sphaerica BC52 CAMERA|CAMPEP_0184494946
- Nephroselmis olivacea 1KP|MMKU_2050578
- Glaucocystis nostochinearum 1KP|2058004_2
- Gloeochaete wittrockiana SPE|86701_1
- Cyanidioschyzon merolae SPE|CMJ172C
- Timspurckia oligopyrenoides CCMP3278 CAMERA|CAMPEP_0182449158
- Erythrolobus australicus CCMP3124 CAMERA|CAMPEP_0185832170
- Erythrolobus madagascarensis CCMP3276 CAMERA|CAMPEP_0185858292
- Porphyridium cruentum EST SPE|None
- Rhodella maculata CCMP736 SPE|9188_1
- Rhodella violacea 1KP|RTL_C_2027867
- Glaucosphaera vacuolata 1KP|2009234_2
- Rhodorus marinus CCMP769 CAMERA|CAMPEP_0113962064
- Chroodactylon ornatum 1KP|LLXJ_2042535
- Rhodochaete parvula 1KP|JJZR_2007625
- Grateloupia filicina 1KP|2056229_1
- Chondrus crispus SPE|GSCHC2T00002317001
- Mazzaella japonica 1KP|WEJN_2005720
- Gloiopeltis furcata 1KP|SBLT_2019960
- Ahnfeltiopsis flabelliformis 1KP|CKXF_2021546
- Dumontia simplex 1KP|IEHF_2003520
- Betaphycus philippinensis 1KP|BWVJ_2022872
- Gracilaria chouae 1KP|FTRP_2000157
- Neosiphonia japonica 1KP|XAXW_2003780
- Heterosiphonia pulchra 1KP|YSBD_2001601
- Ceramium kondoi 1KP|VZWX_2004163
- Galdieria sulphuraria SPE|Gs25820.1
- Geminigera cryophila CCMP2564 CAMERA|CAMPEP_0179424948
- Hanusia phi CCMP325 CAMERA|CAMPEP_0169497176
- Guillardia theta JGI|141196
- Cryptomonas paramecium CCAP977 CAMERA|CAMPEP_0113668318
- Chroomonas mesostigmatica CCMP1168 CAMERA|CAMPEP_0206246248
- Hemiselmis andersenii CCMP439 CAMERA|CAMPEP_0172016434
- Exanthemachrysis gayraliae RCC1523 CAMERA|CAMPEP_0206029442
- Pavlova sp. CCMP459 CAMERA|CAMPEP_0190519316
- Prymnesium parvum Texoma1 CAMERA|CAMPEP_0191234060
- Chrysochromulina polylepis CCMP1757 CAMERA|CAMPEP_0193736082
- Isochrysis sp. CCMP1324 CAMERA|CAMPEP_0188819172
- Emiliania huxleyi CCMP1516 JGI|464911
- Gephyrocapsa oceanica RCC1303 CAMERA|CAMPEP_0188189622
- Calcidiscus leptoporus RCC1130 CAMERA|CAMPEP_0119402080
- Pleurochrysis carterae CCMP645 CAMERA|CAMPEP_0190803892
- Eutreptiella gymnastica CAMERA|CAMPEP_0113734016
- Euglena sp. 1KP|UNBZ_2035389
- Euglena gracilis SPE|t166311_f+11
- Chattonella subsalsa CCMP2191 CAMERA|CAMPEP_0187158486
- Heterosigma akashiwo CCMP2393 CAMERA|CAMPEP_0188552310
- Ectocarpus siliculosus JGI|Esi0119_0029
- Vaucheria litorea CCMP2940 CAMERA|CAMPEP_0199160038
- Leptocylindrus danicus CCMP1856 CAMERA|CAMPEP_0196826898
- Striatella unipunctata CCMP2910 CAMERA|CAMPEP_0118718172
- Fragilariopsis cylindrus JGI|158419
- Odontella aurita isolate 1302-5 CAMERA|CAMPEP_0113574852
- Chaetoceros neograticus CCMP1317 CAMERA|CAMPEP_0201002572
- Pelagococcus subviridis CCMP1429 CAMERA|CAMPEP_0190545592
- Pelagomonas calceolata CCMP1756 CAMERA|CAMPEP_0199678822
- Aureococcus anophagefferens JGI|11312
- Aureoumbra lagunensis CCMP1510 CAMERA|CAMPEP_0186722748
- Pseudopedinella elastica CCMP716 CAMERA|CAMPEP_0191404348
- Dictyocha speculum CCMP1381 CAMERA|CAMPEP_0185776860
- Phaeomonas parva CCMP2877 CAMERA|CAMPEP_0118849898
- Pinguicoccus pyrenoidosus CCMP2078 CAMERA|CAMPEP_0197372532
- Dinobryon sp. UTEXLB2267 CAMERA|CAMPEP_0187276202
- Chromera velia CAMERA|CAMPEP_0184612882
- Vitrella brassicaformis CCMP3346 CAMERA|CAMPEP_01984134084



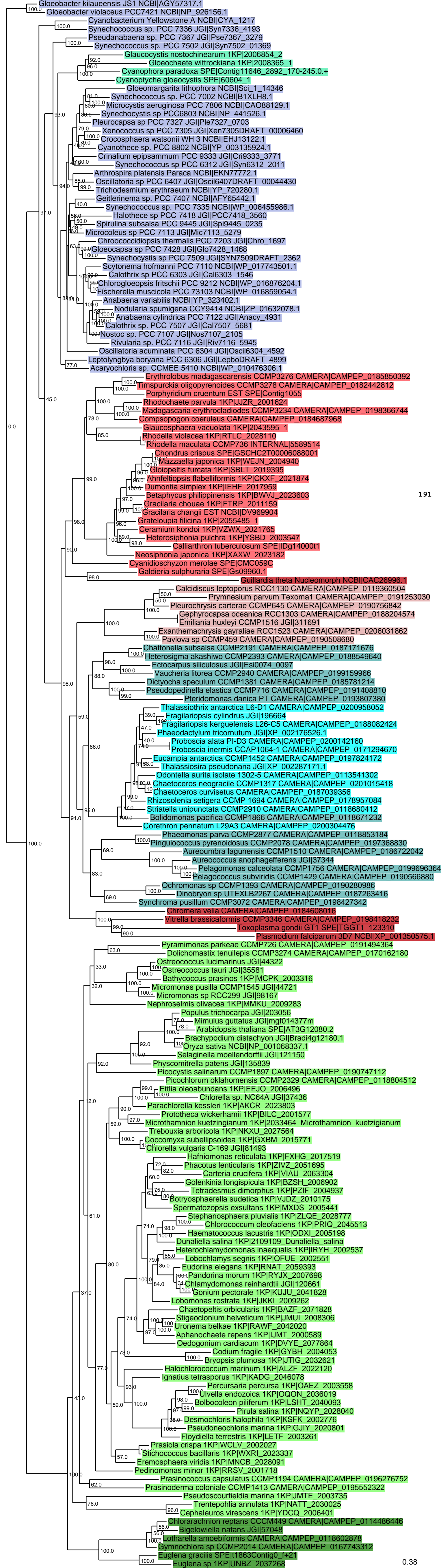




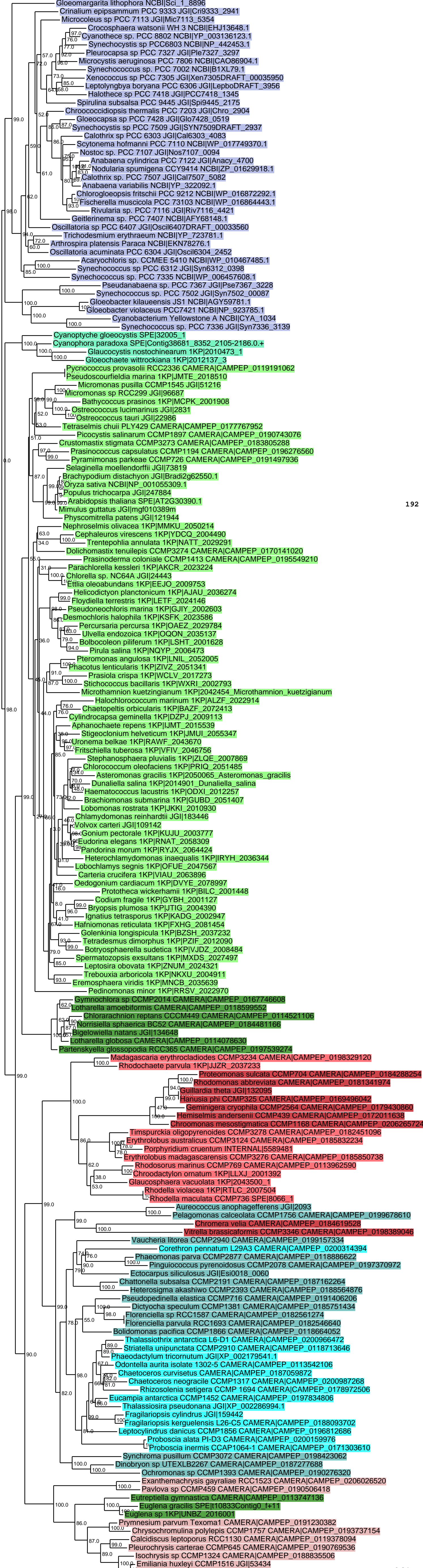


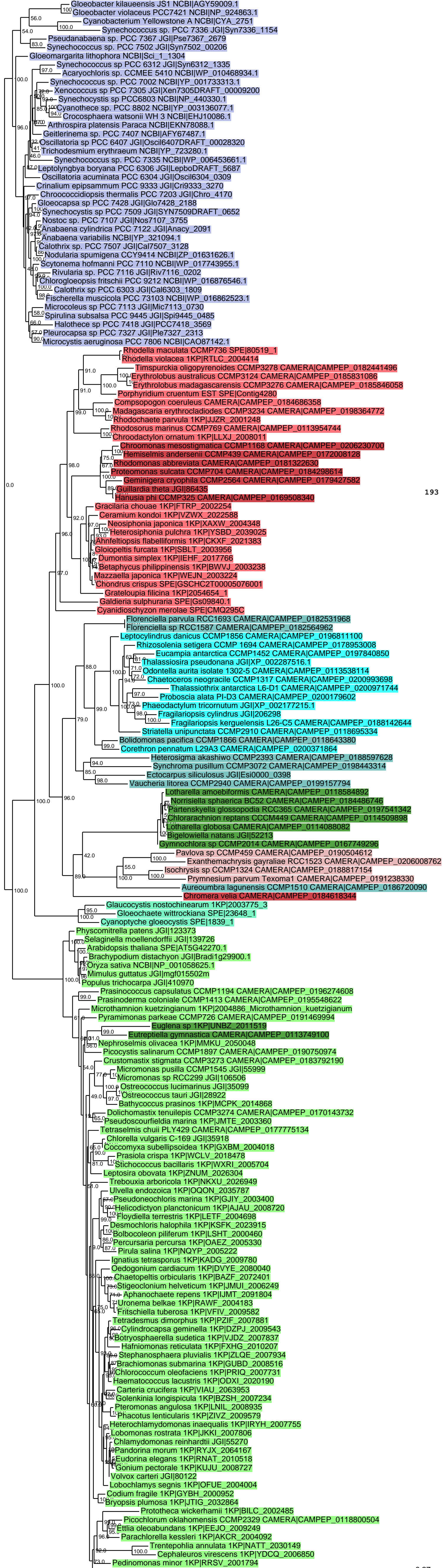


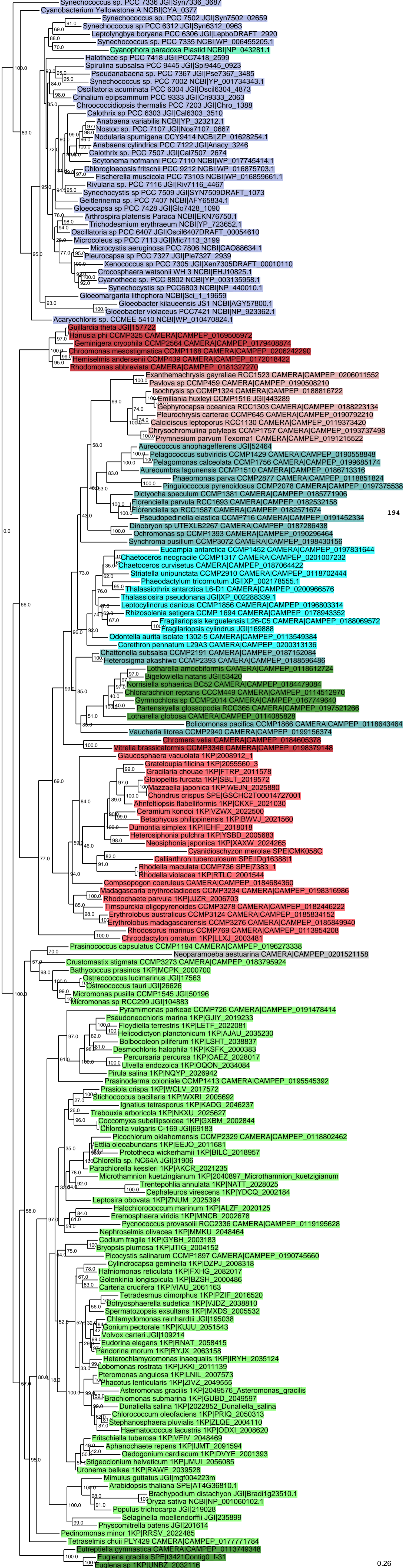
190



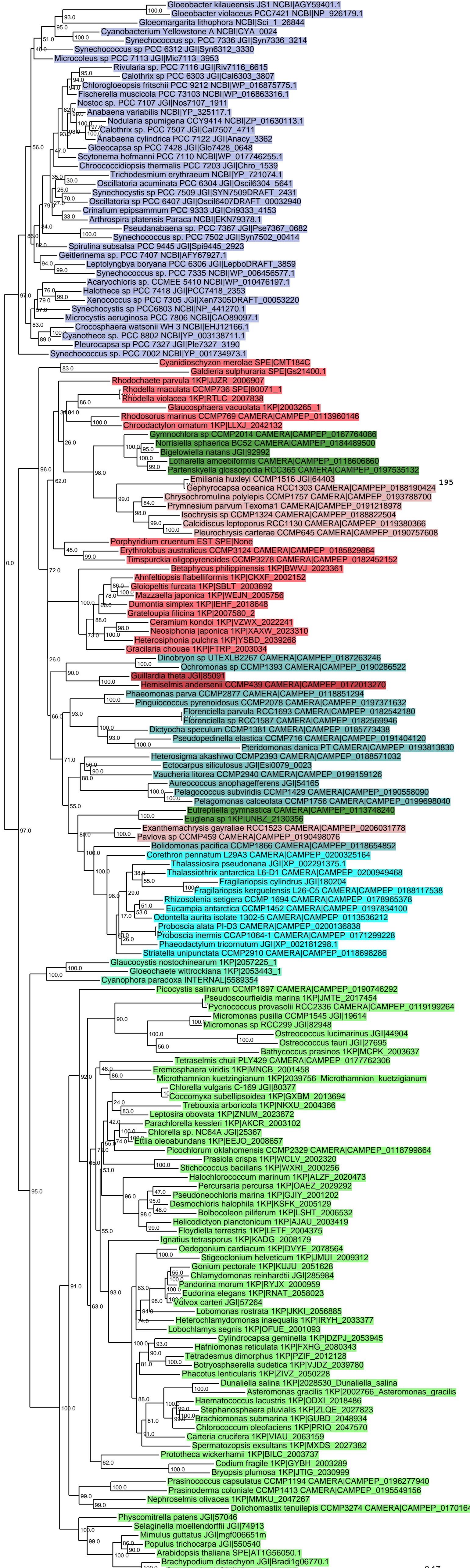
191

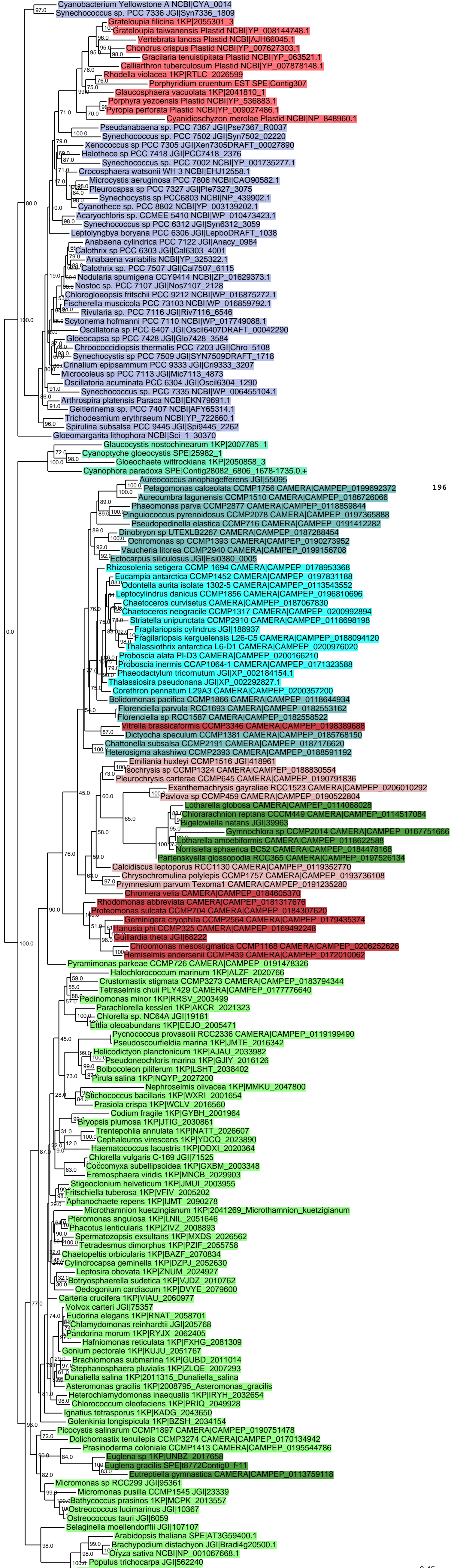




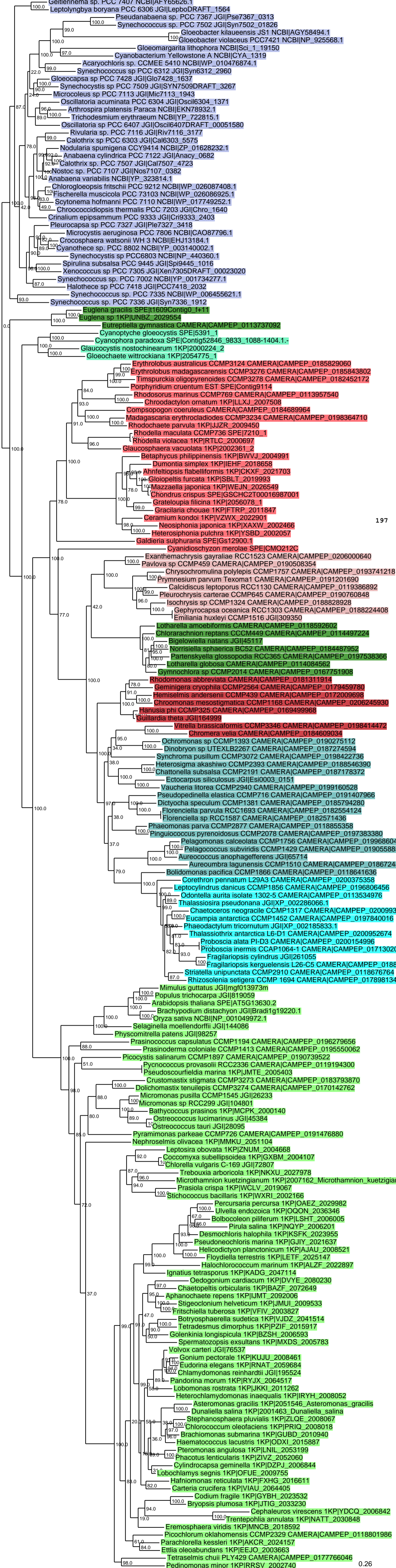


194



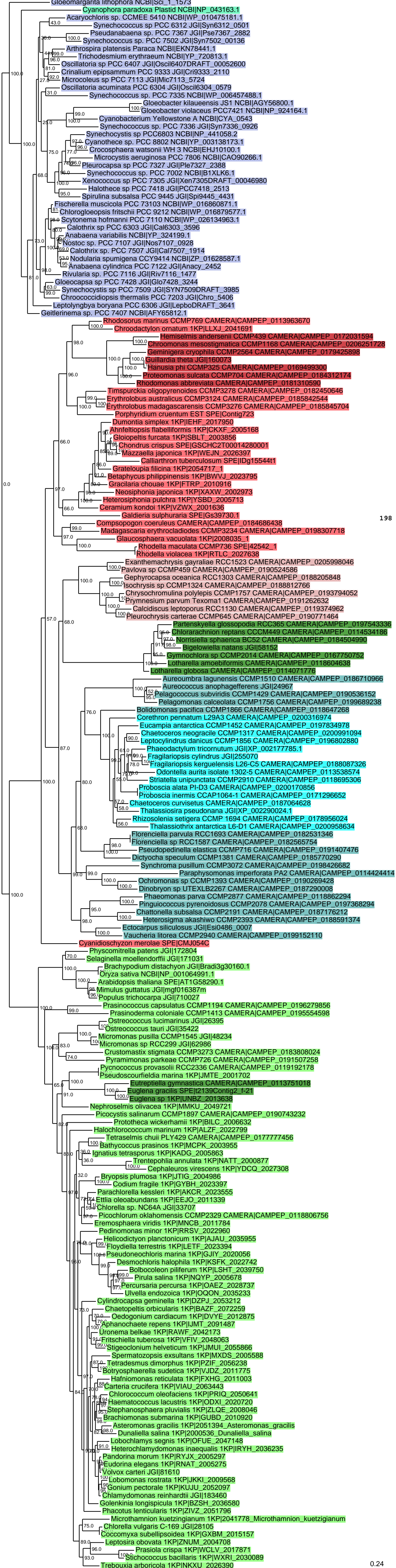


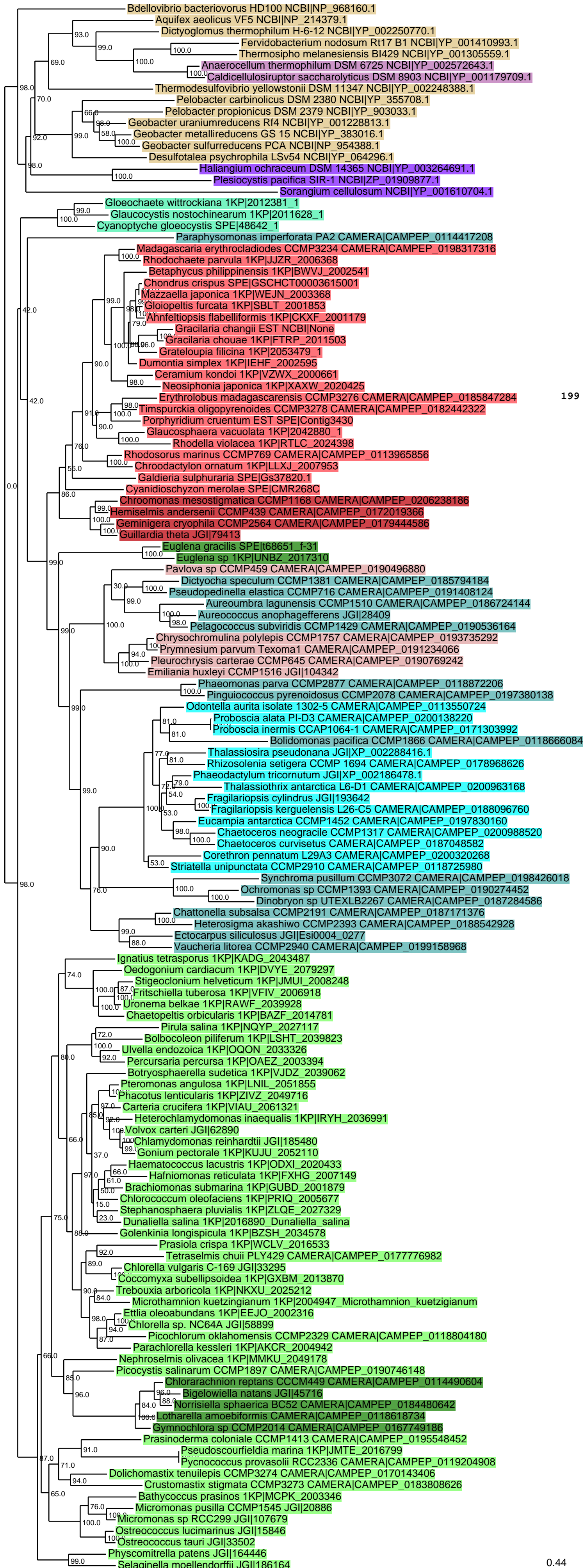
196



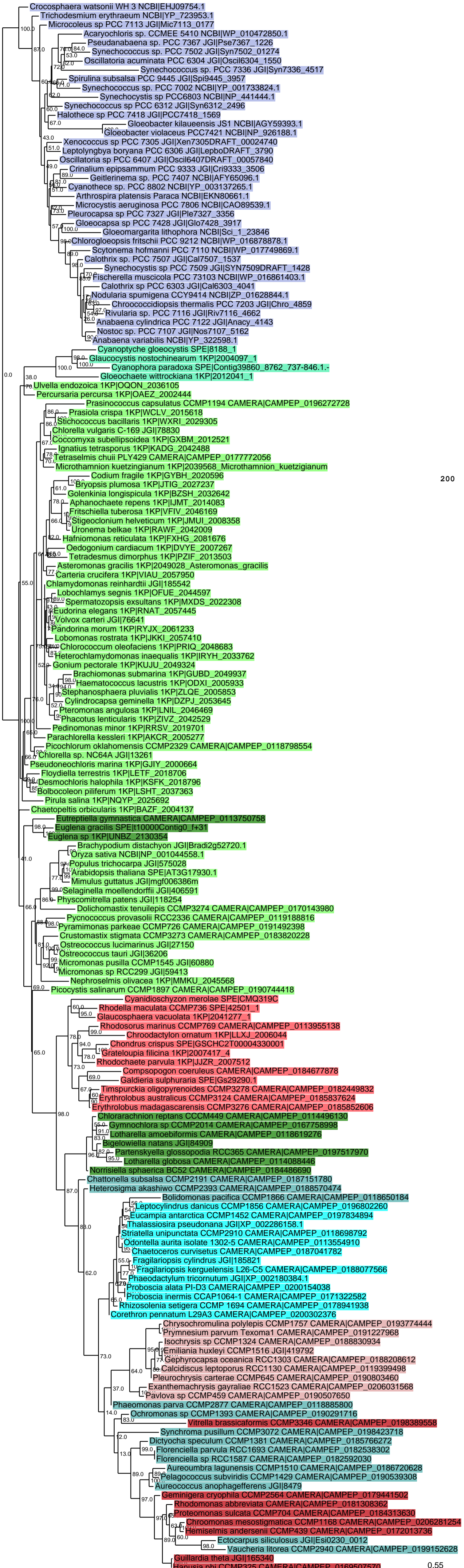
197

0.26

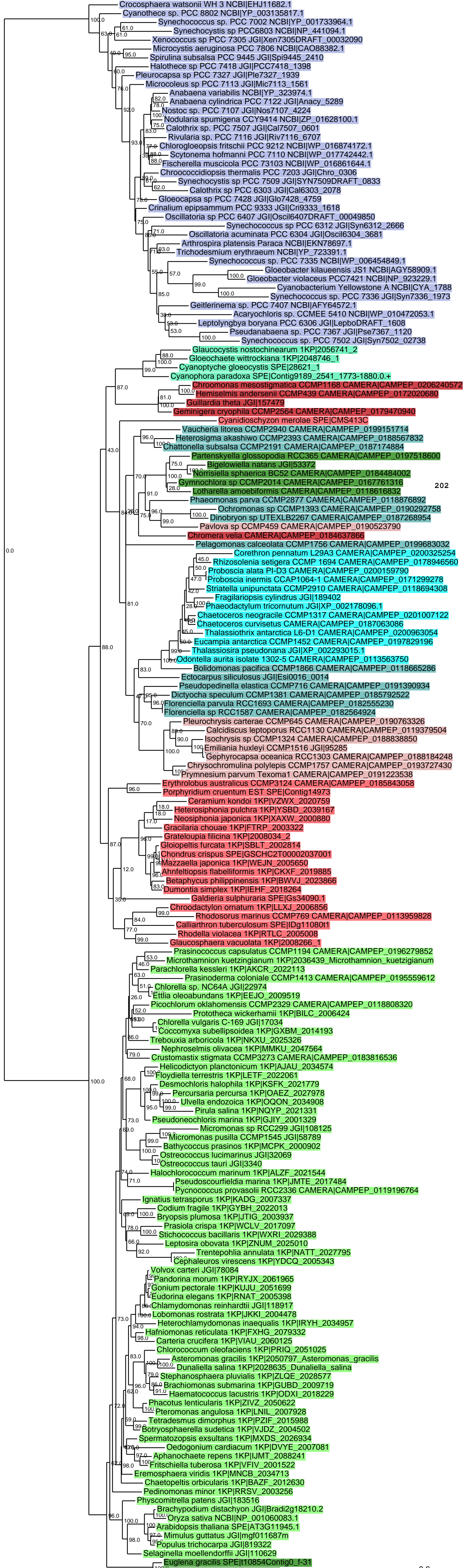




199



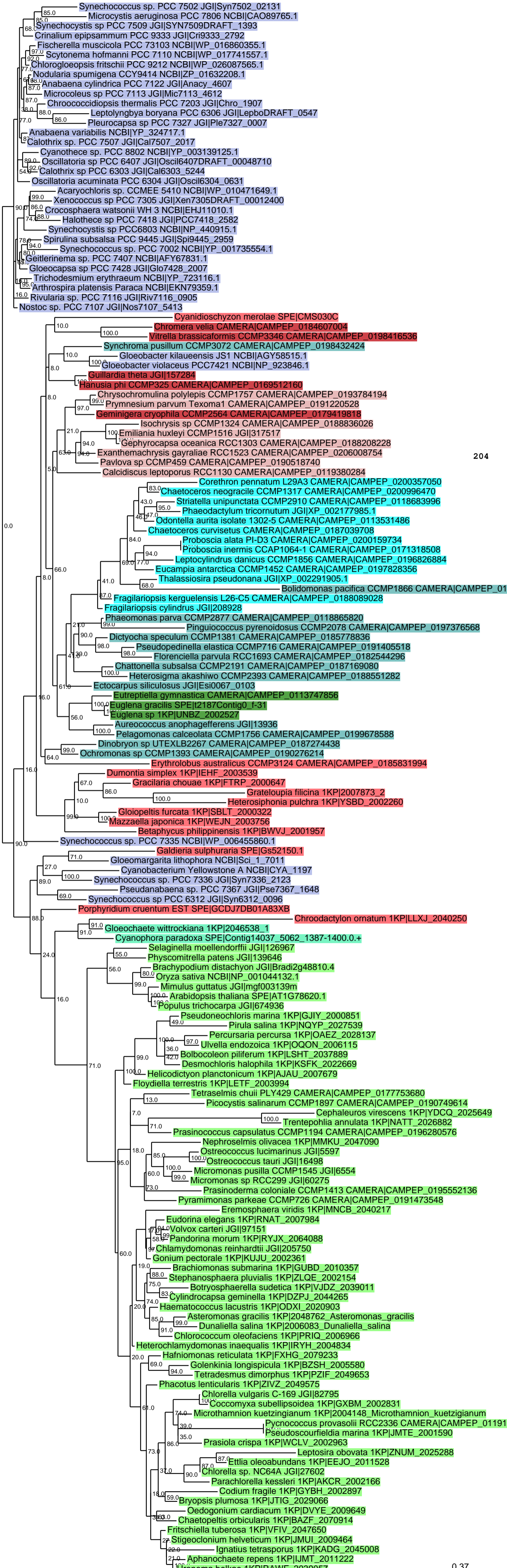




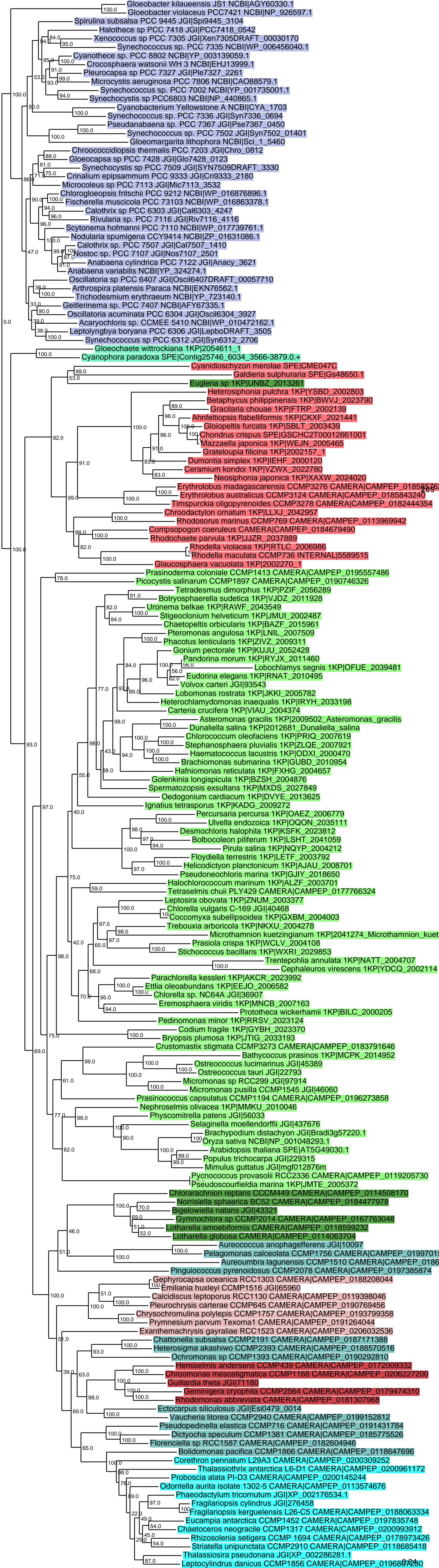
202

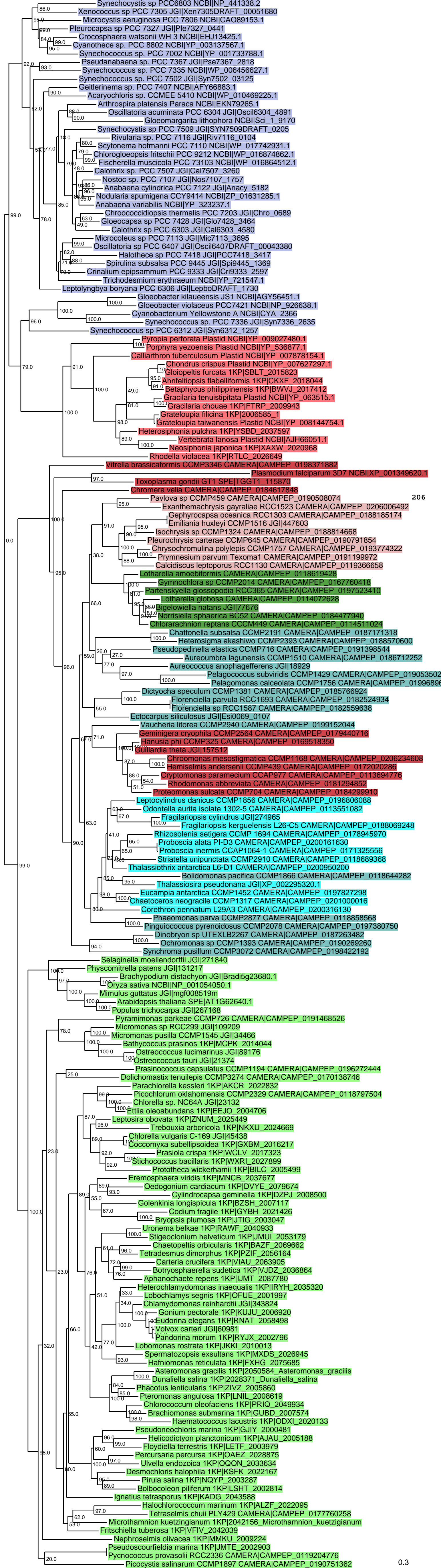


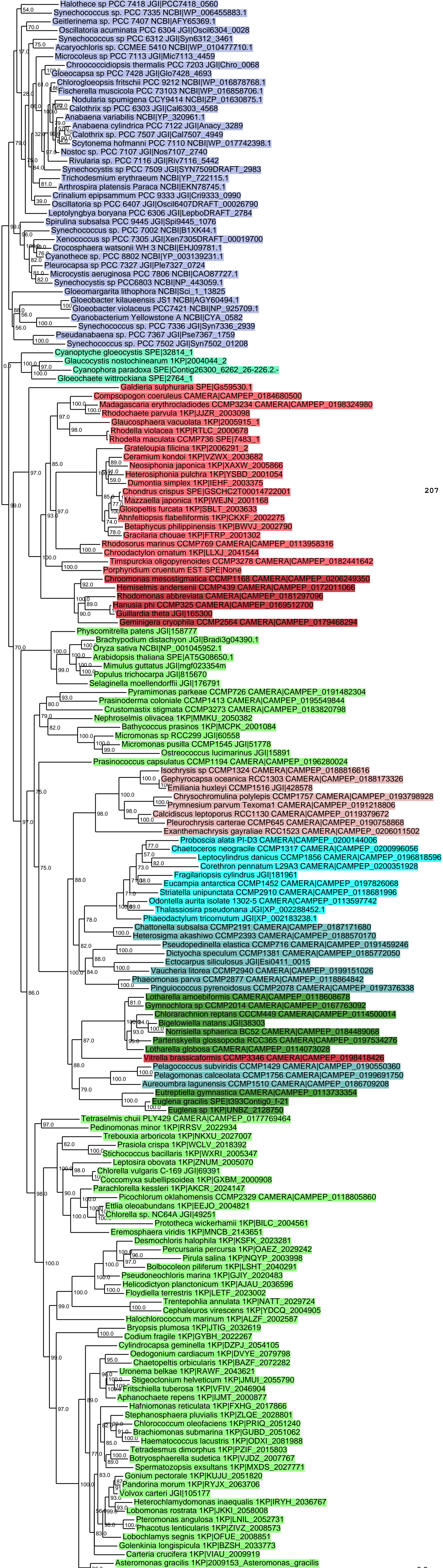
203



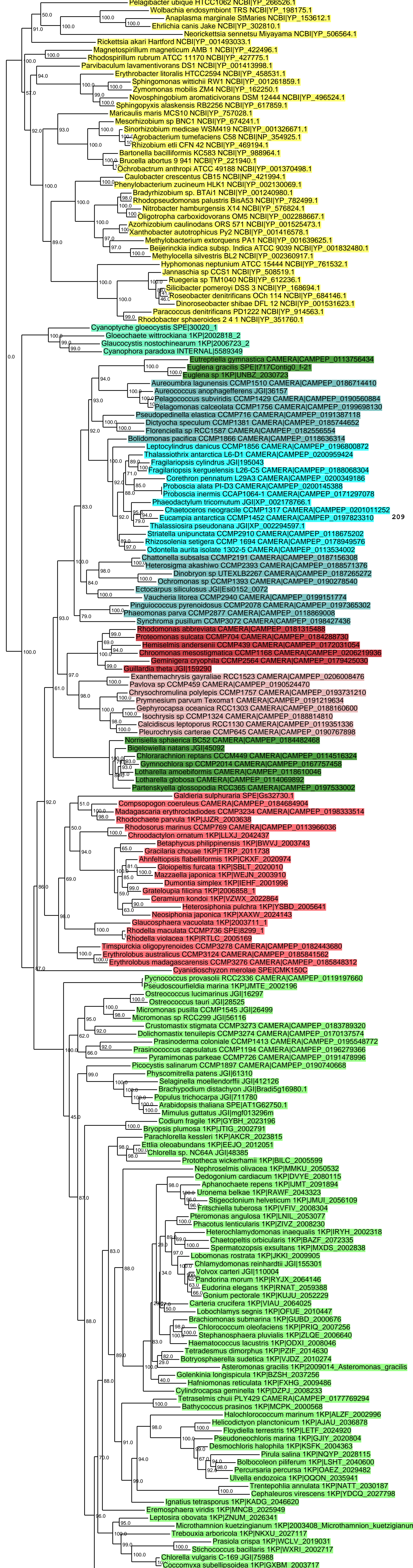
204

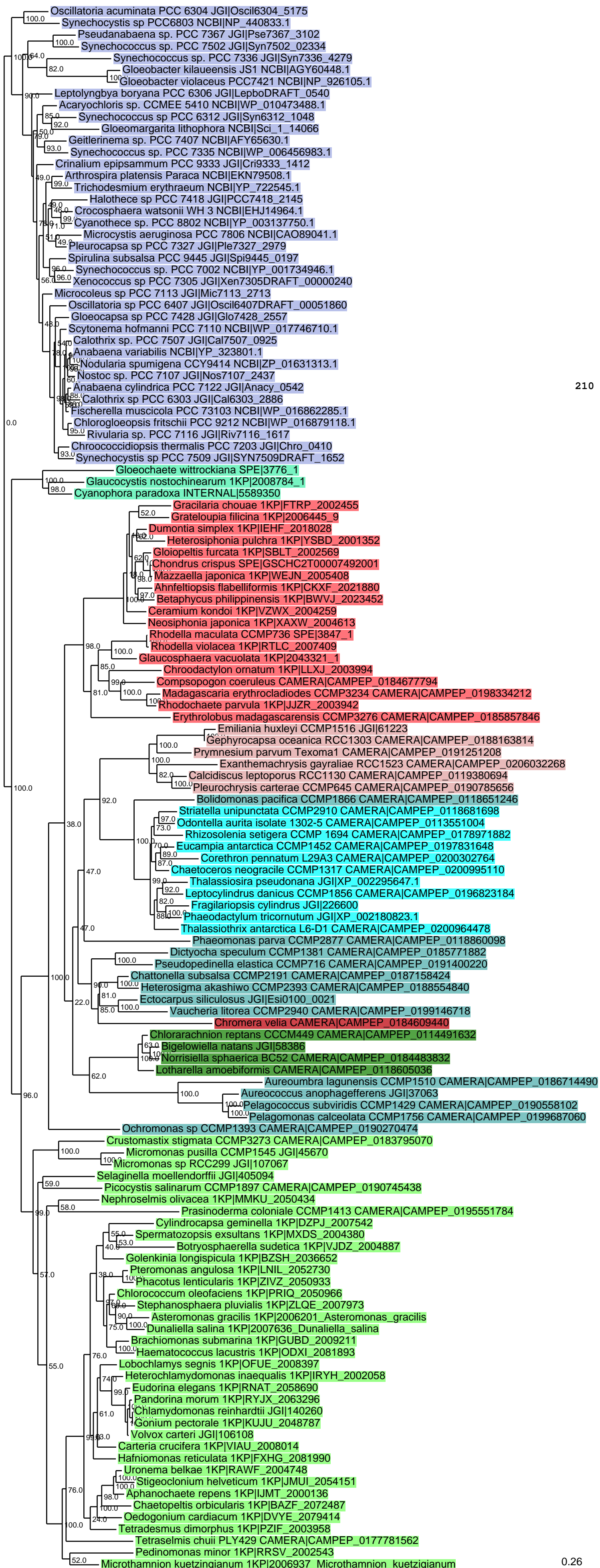


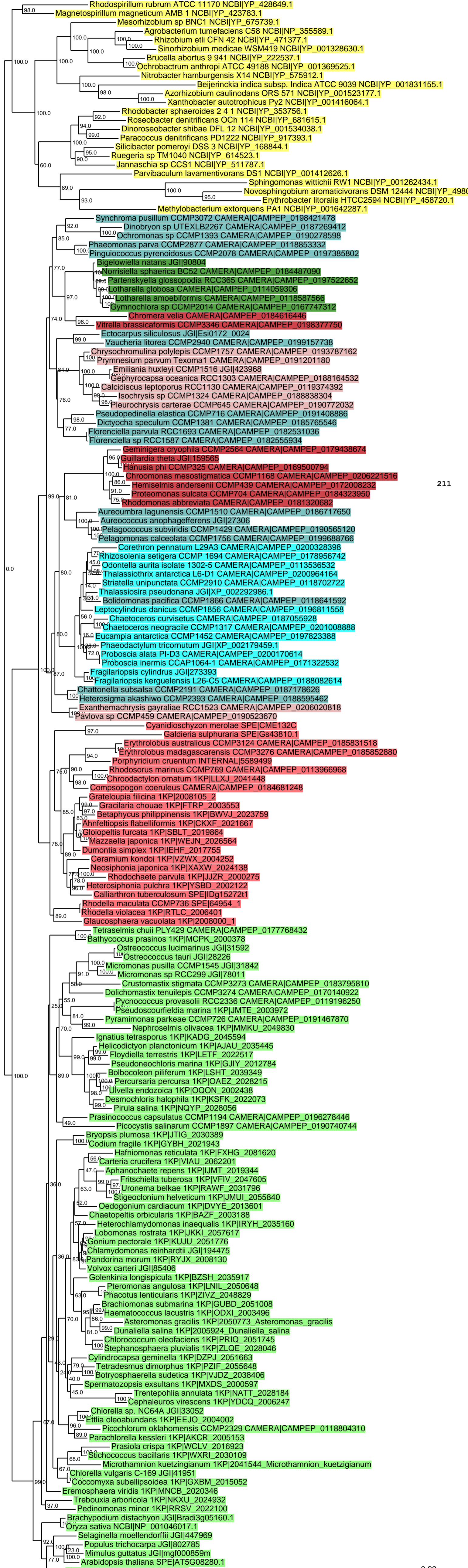










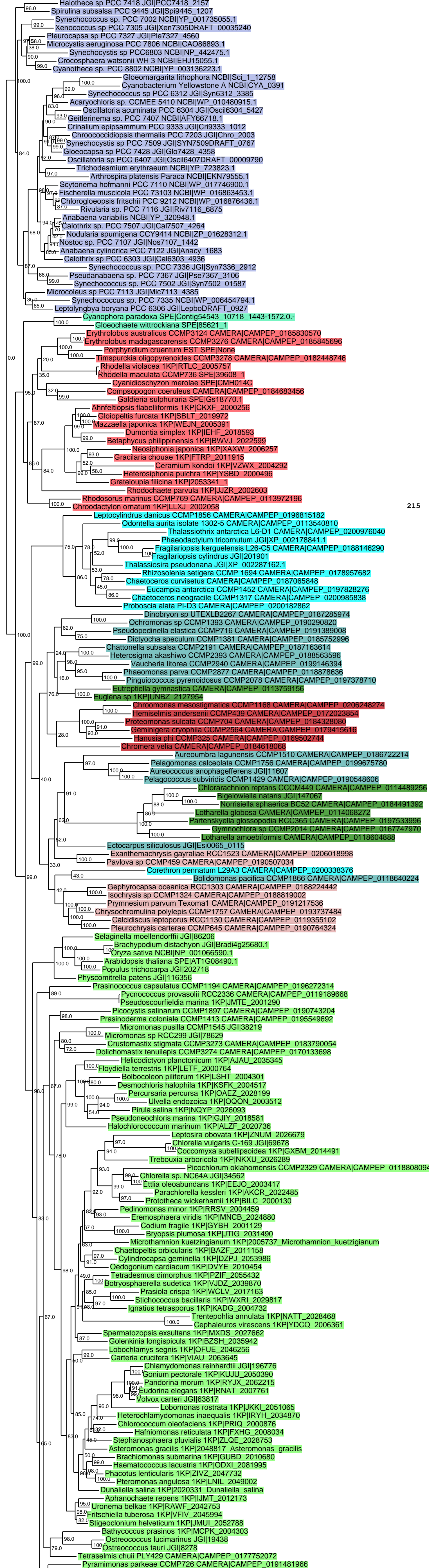


211







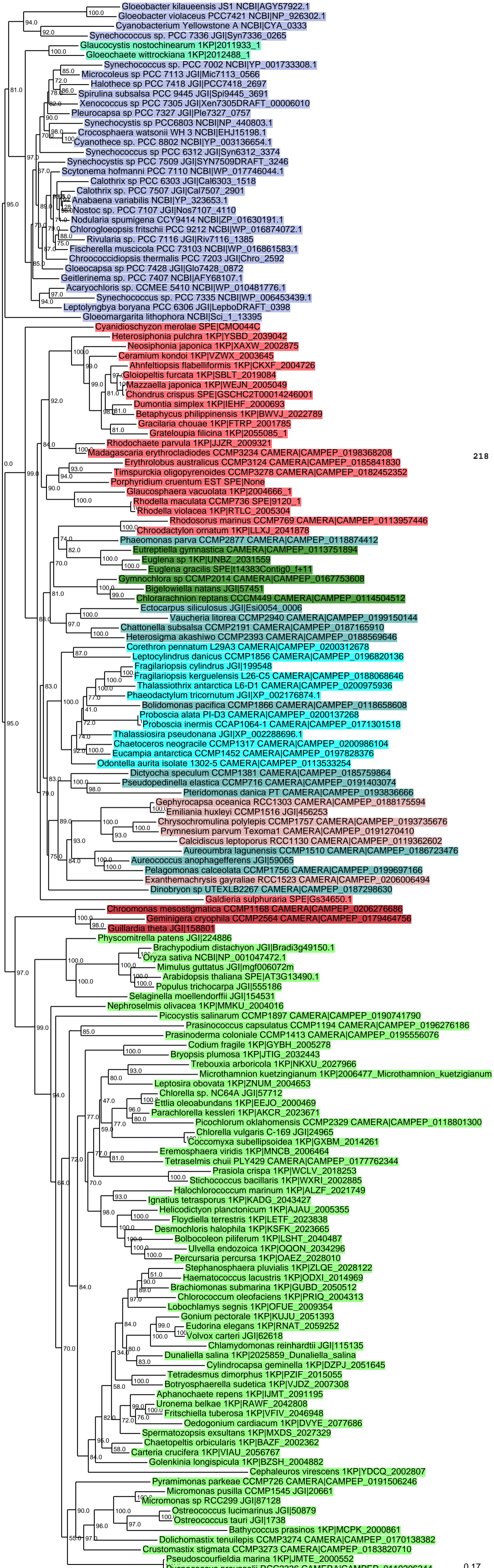


215



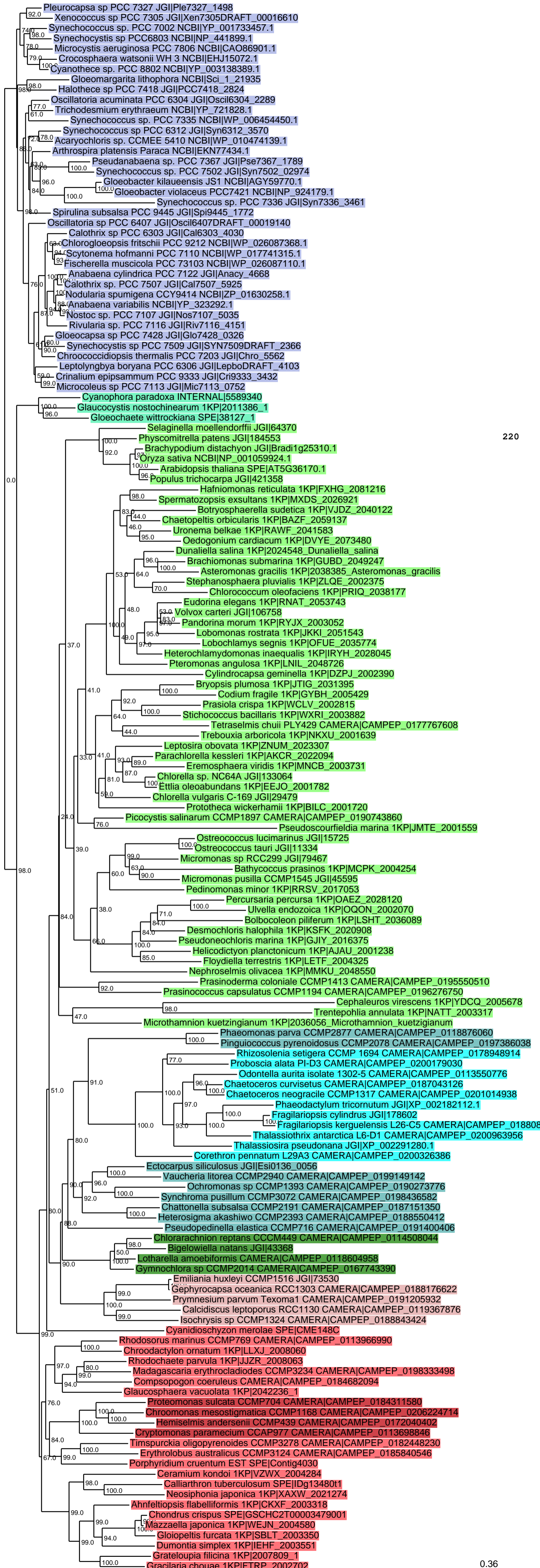


217



218

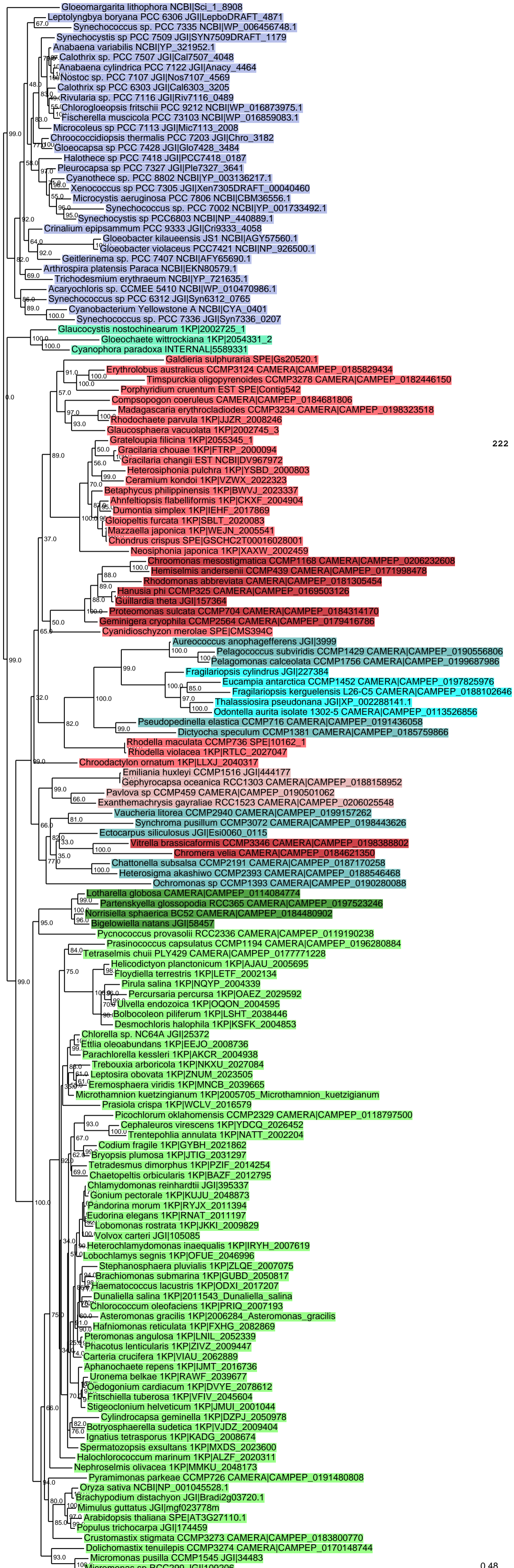




220

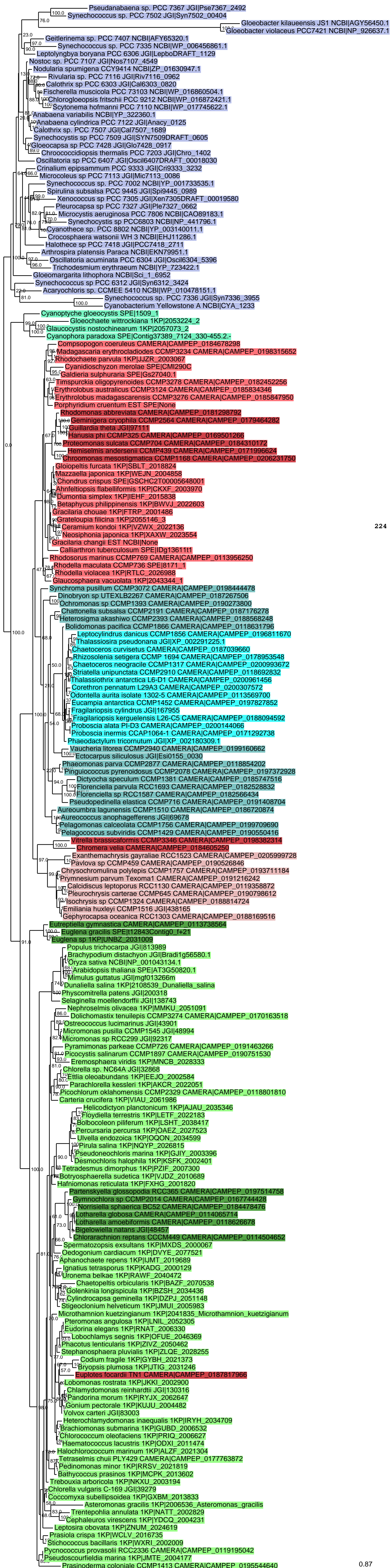


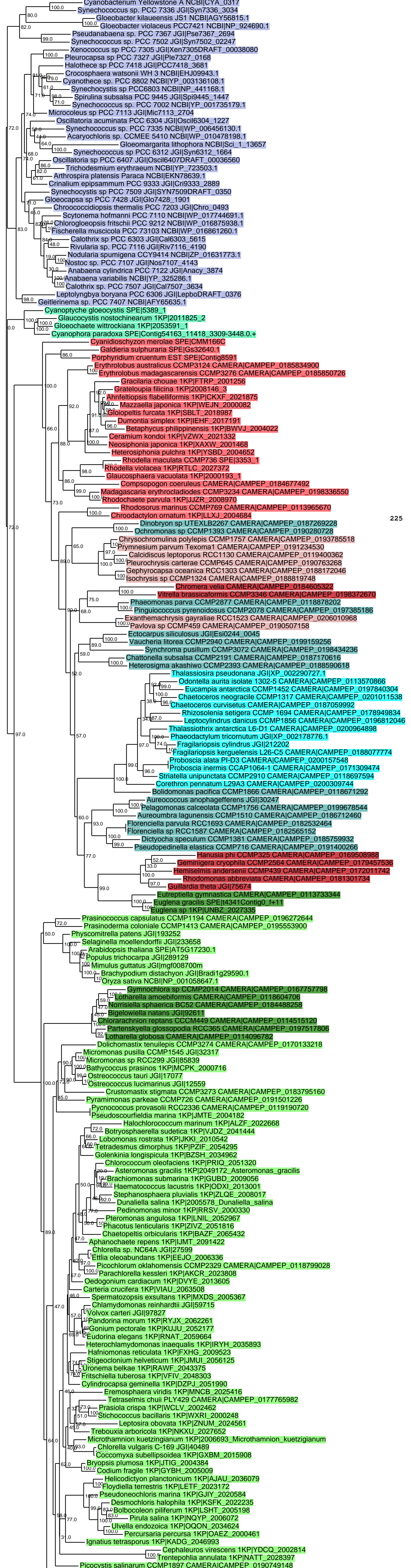
221

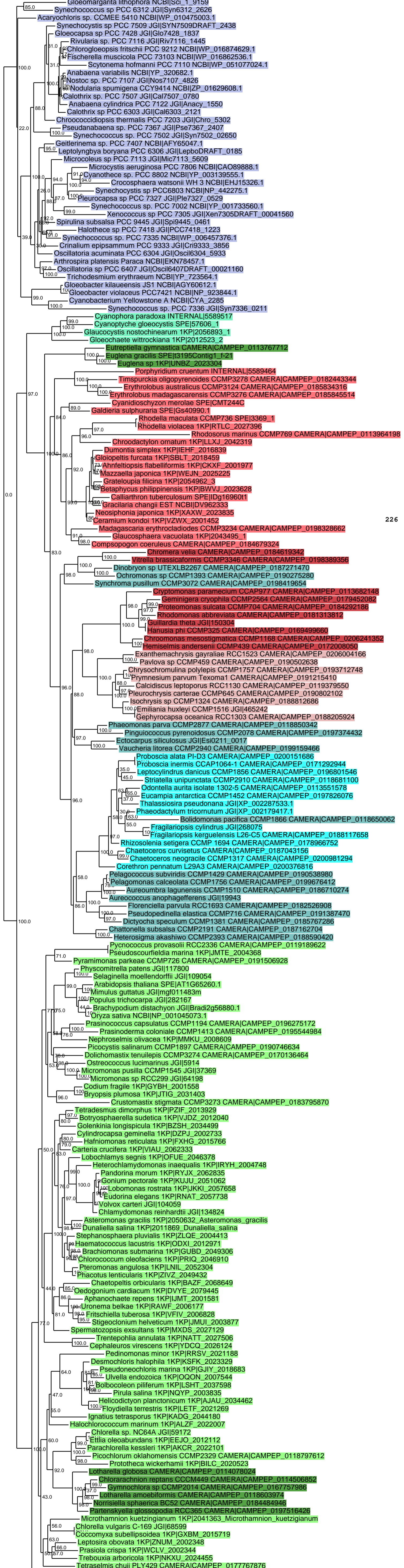




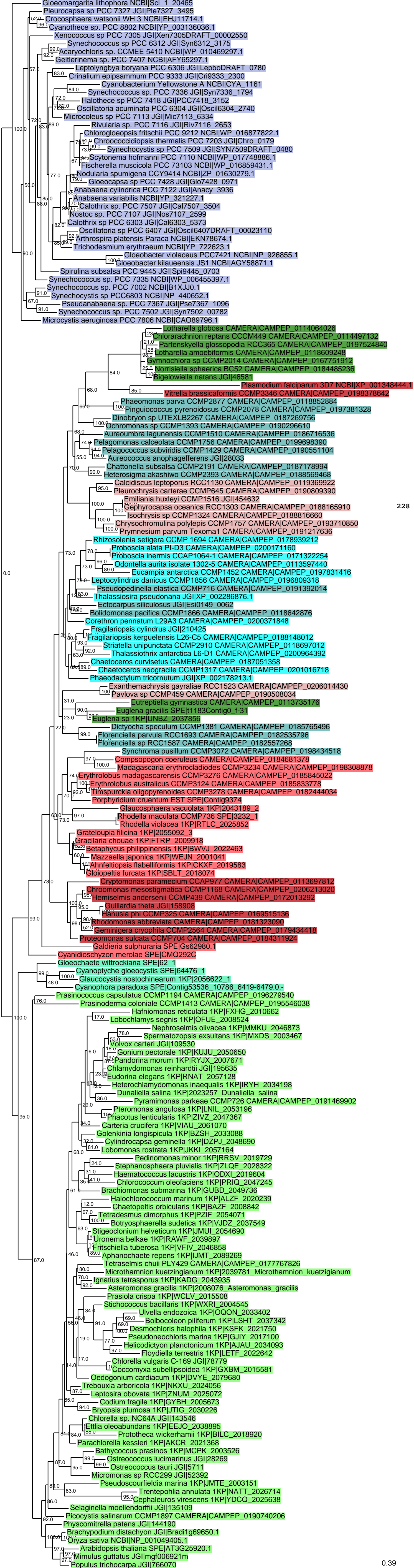
223









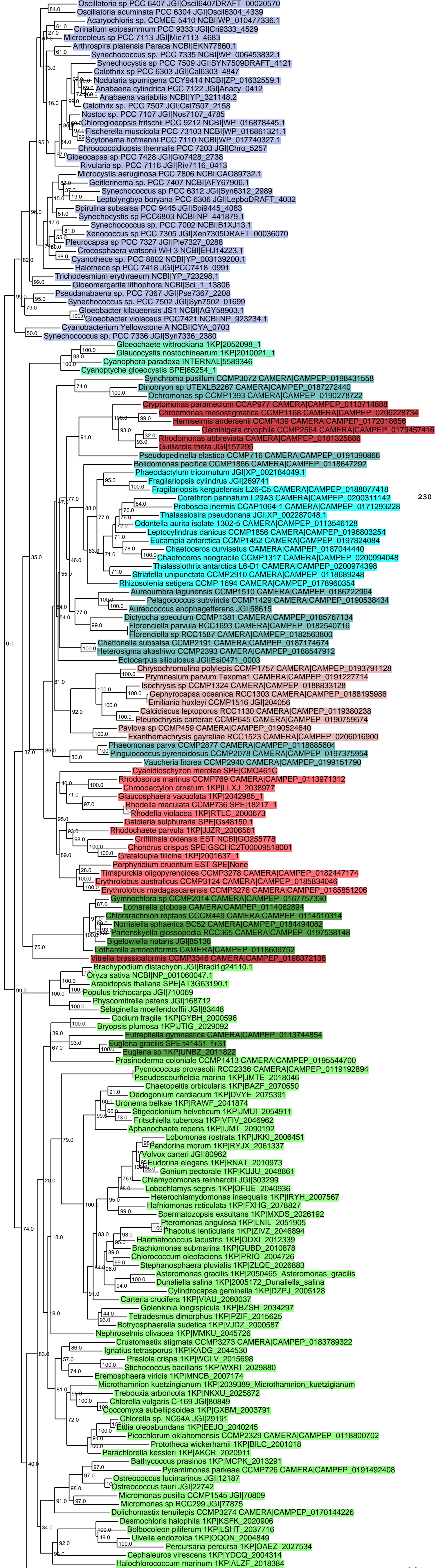


228



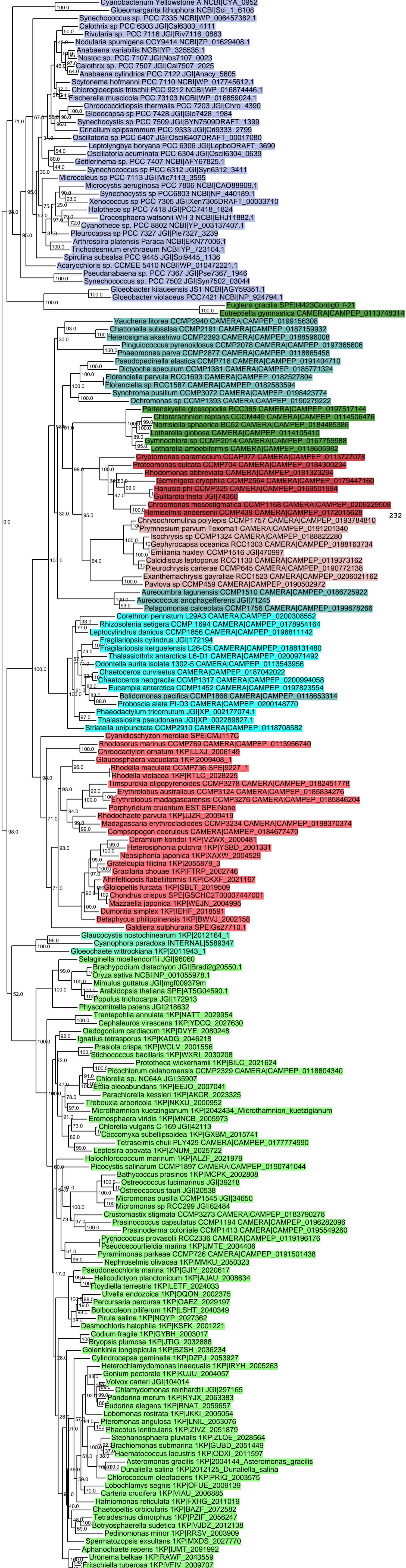
229

0.37



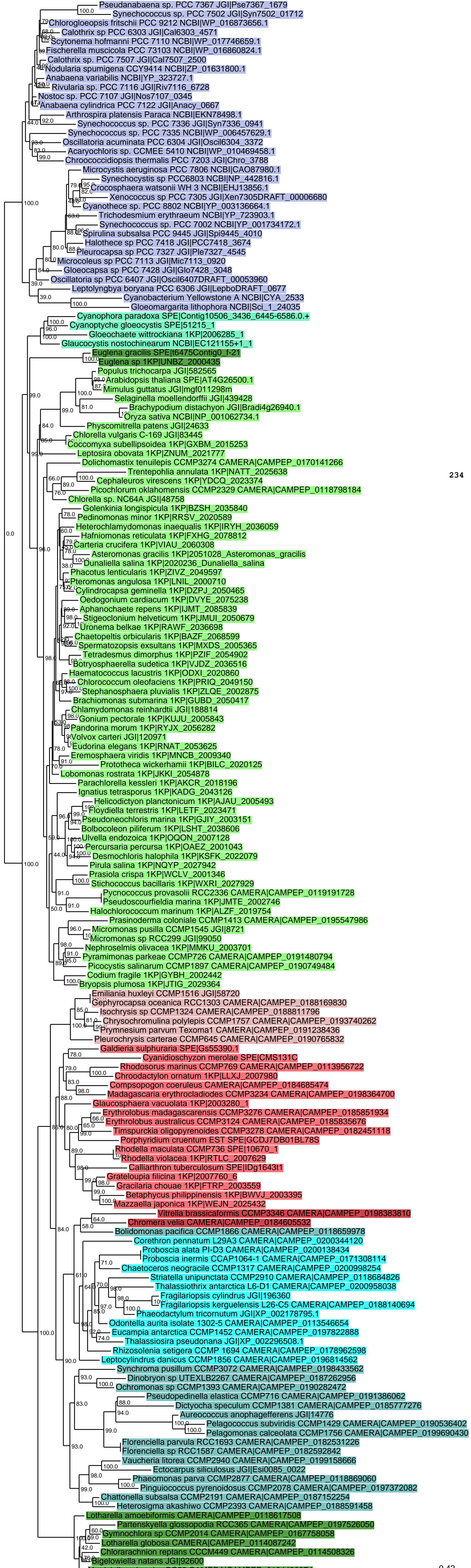
230

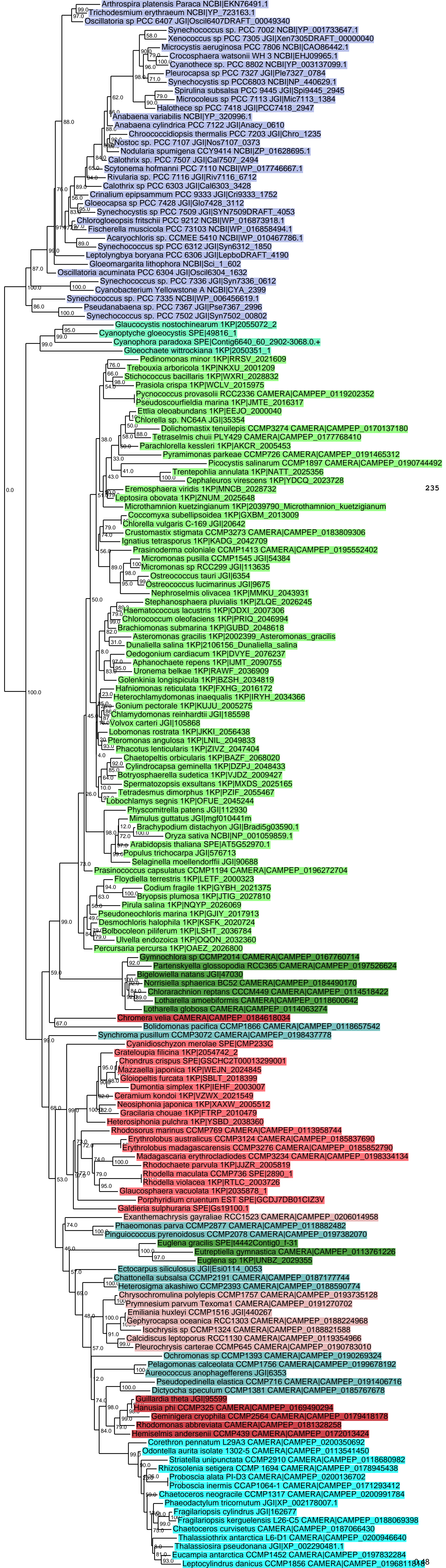


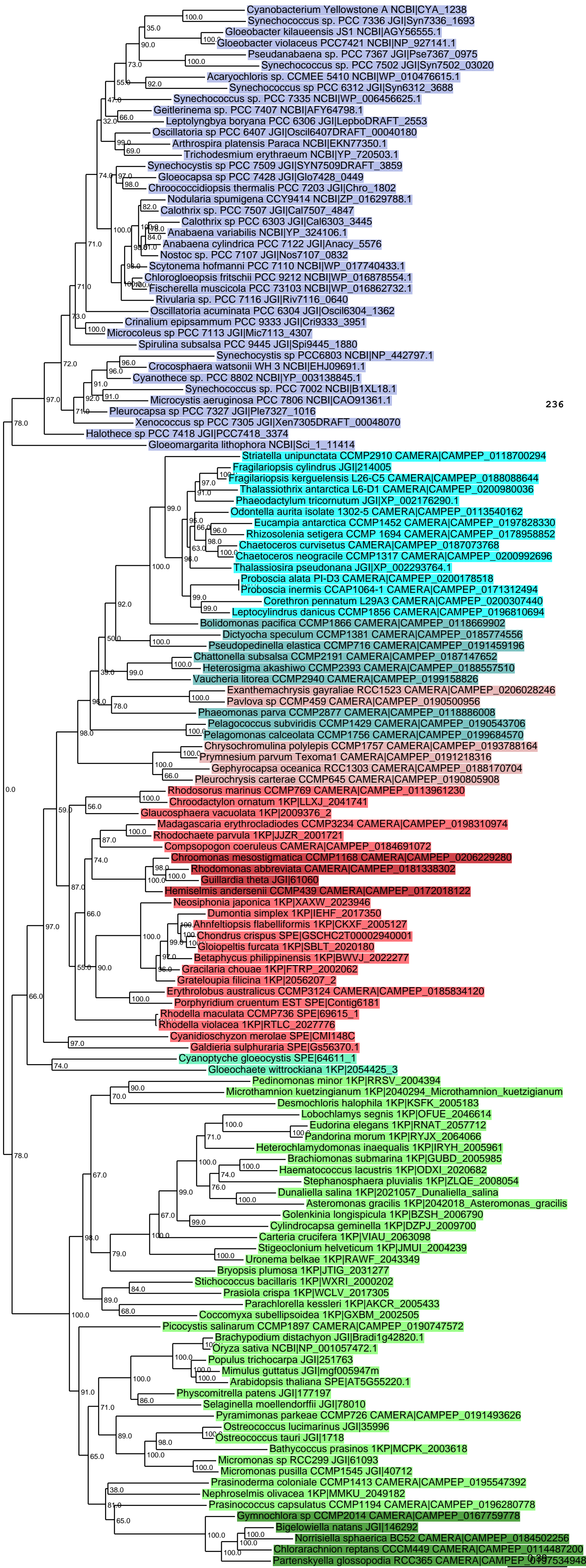




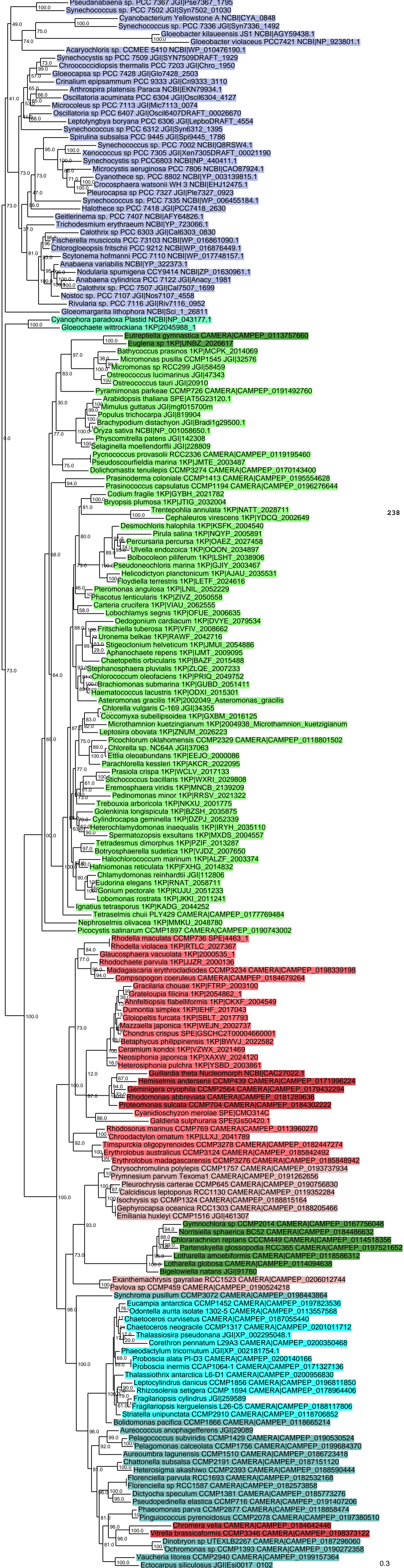
233

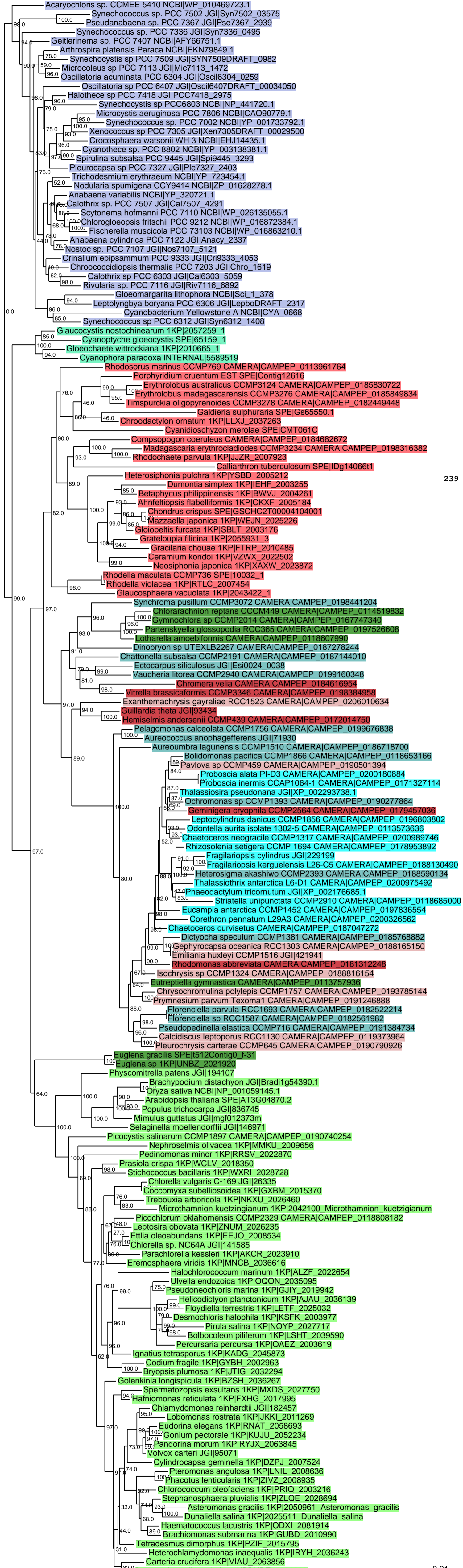


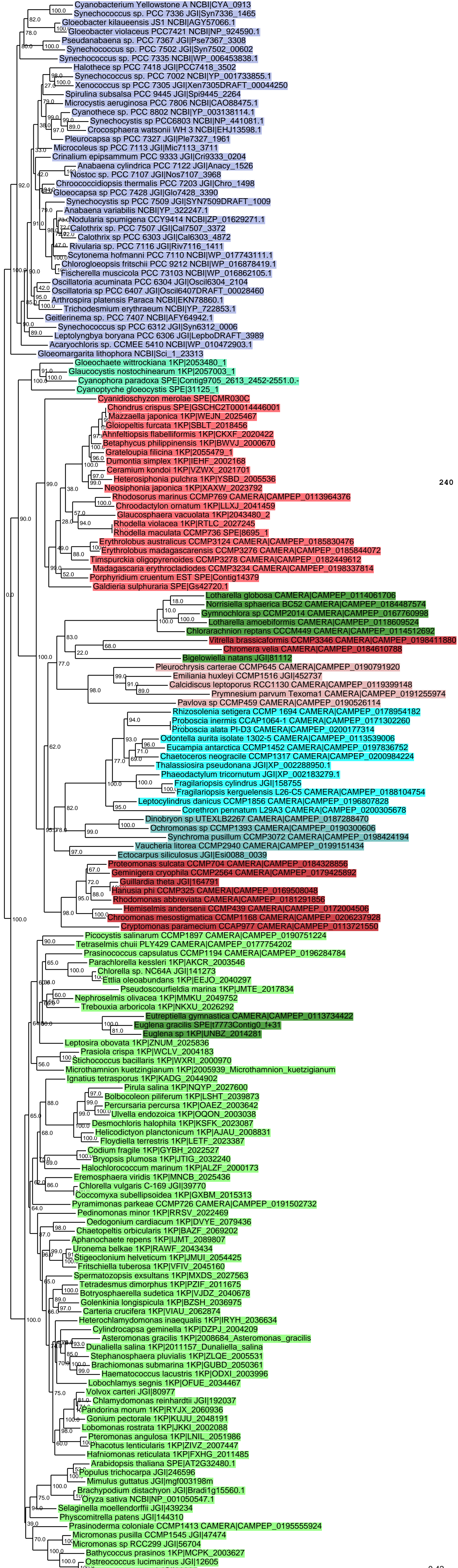




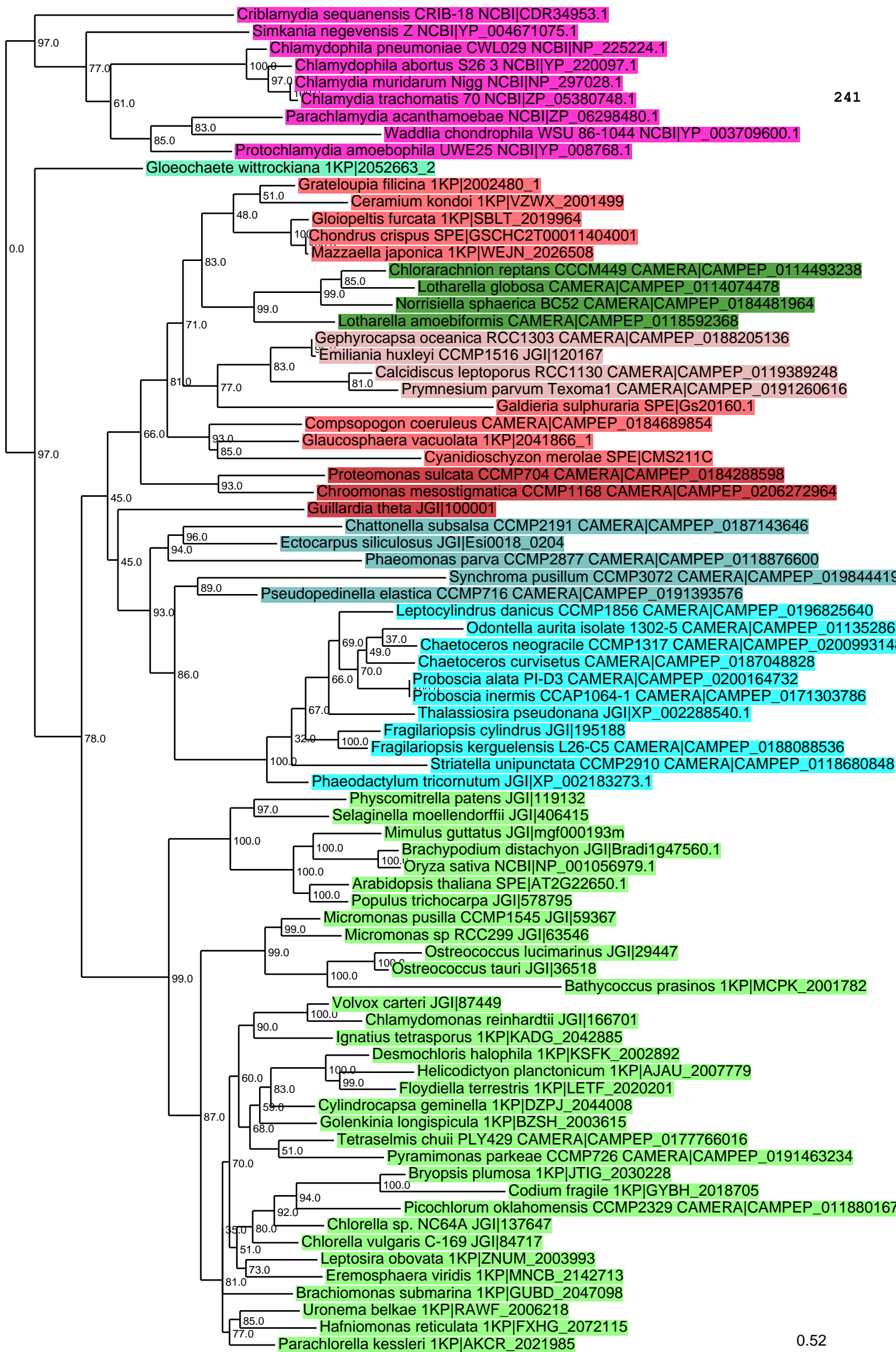


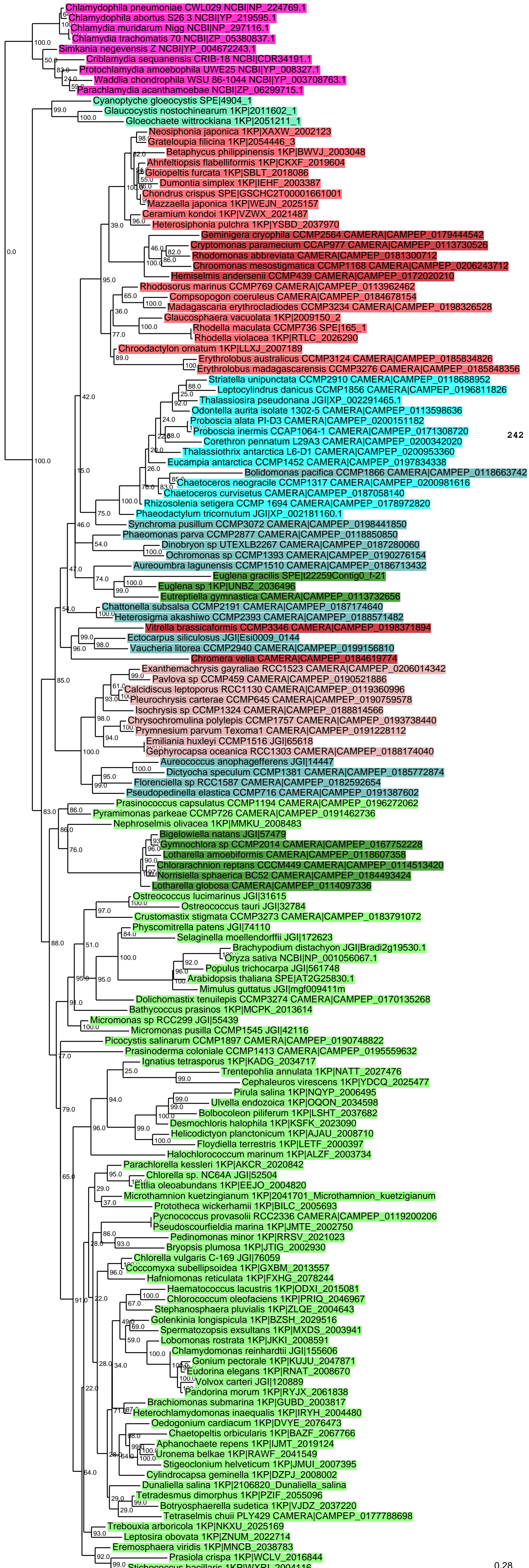




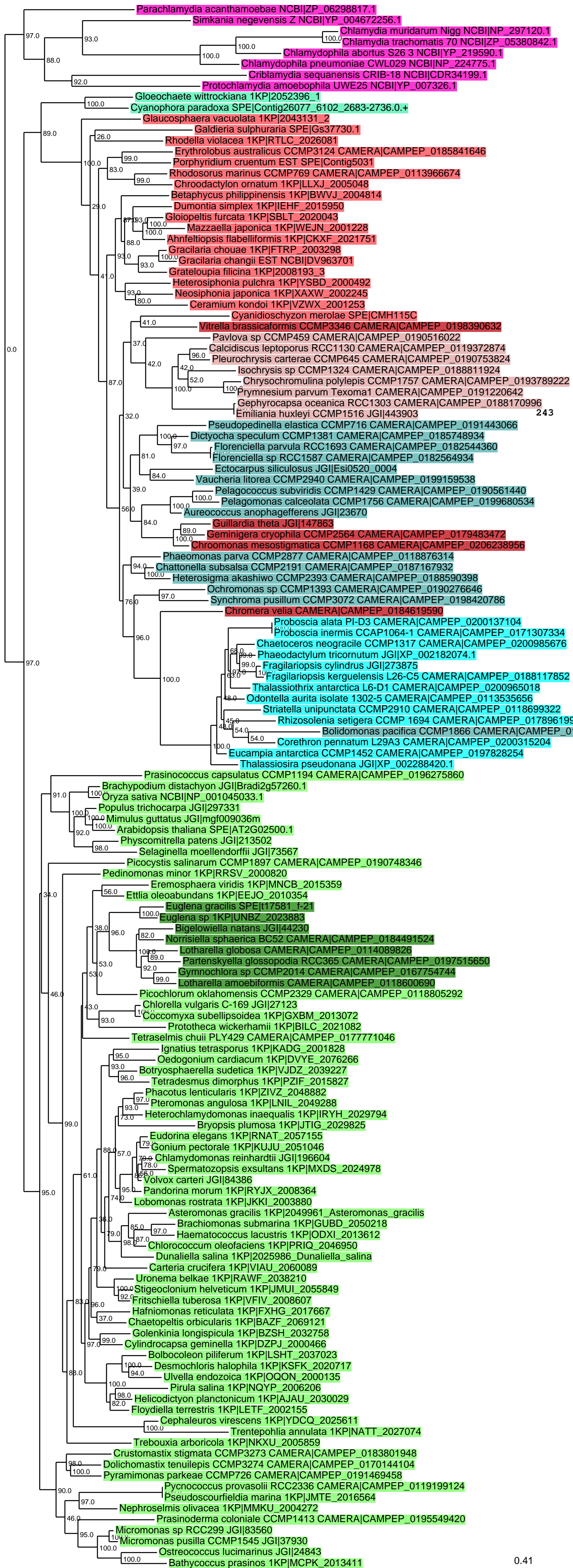


240





242







245

0.57

9.3 Supplementary material of manuscript 3

Molecular phylogeny disputes the monophyly of the SELMA translocation machinery in complex secondary plastids.

(Manuscript in preparation)

Rafael I. Ponce-Toledo¹, David Moreira¹, Purificación López-García¹ and Philippe Deschamps^{1,*}

¹Unité d'Ecologie Systématique et Evolution, CNRS, Université Paris-Sud, AgroParisTech, Université Paris-Saclay, 91400, Orsay, France

*To whom correspondence should be addressed.

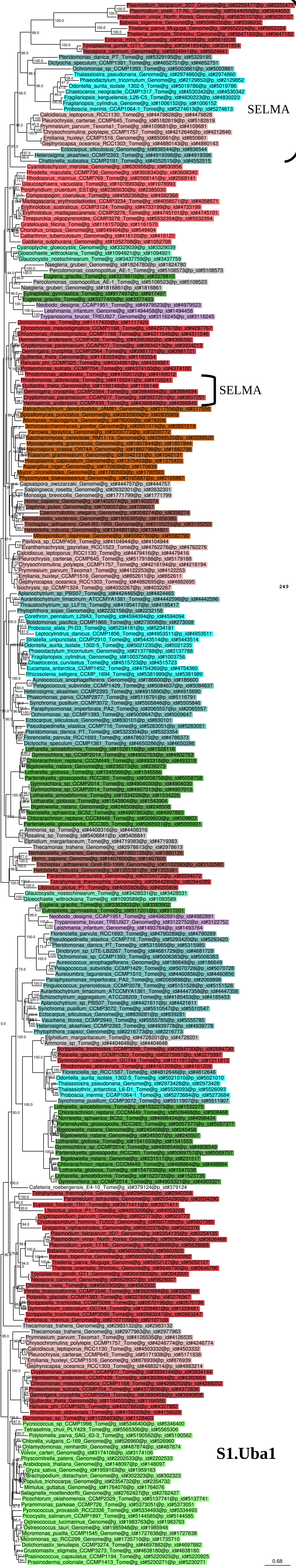
E-mail: philippe.deschamps@u-psud.fr

Supplementary table S1

The composition of the genome local database used in this study is the same as depicted in the Table S1 of section 9.2.

Supplementary figure:

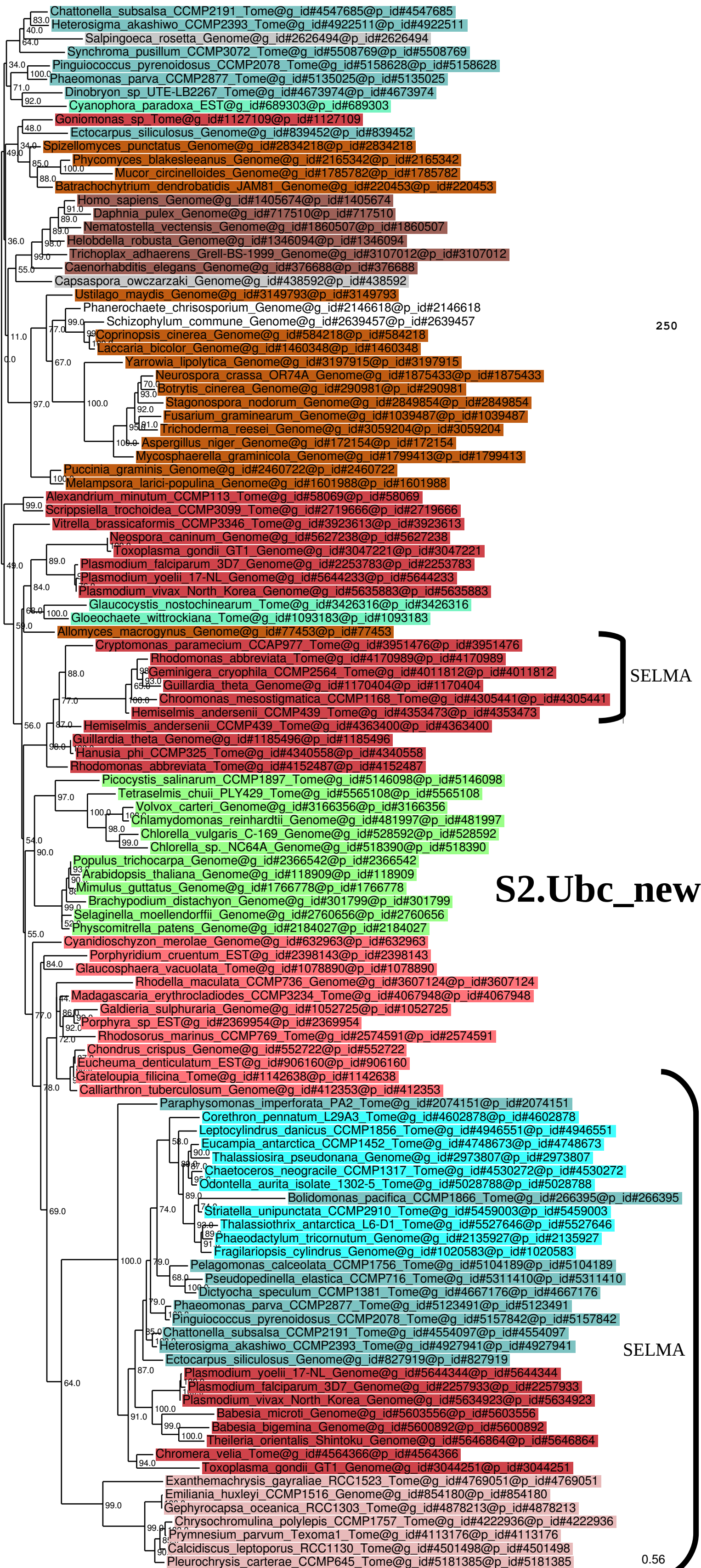
- S1. ML phylogenetic tree of Uba1 proteins
- S2. ML phylogenetic tree of Ubc_new proteins
- S3. ML phylogenetic tree of Ubc1 proteins
- S4. ML phylogenetic tree of Ubc2 proteins
- S5. ML phylogenetic tree of Ubc4 proteins
- S6. ML phylogenetic tree of Ubc6 proteins
- S7. ML phylogenetic tree of Ubc7 proteins
- S8. ML phylogenetic tree of Ubc8 proteins
- S9. ML phylogenetic tree of Hrd1 proteins
- S10. ML phylogenetic tree of ptE3p proteins
- S11. ML phylogenetic tree of ptDUP proteins
- S12. ML phylogenetic tree of Ufd1 proteins
- S13. ML phylogenetic tree of Npl4 proteins
- S14. ML phylogenetic tree of Cdc48 proteins
- S15. ML phylogenetic tree of Cdc48 proteins reconstructed under the PMSF model
- S16. ML phylogenetic tree of derlin proteins
- S17. ML phylogenetic tree of Hsp70 proteins
- S18. ML phylogenetic tree of Rad23 proteins



SELMA

SELMA

S1.Uba1



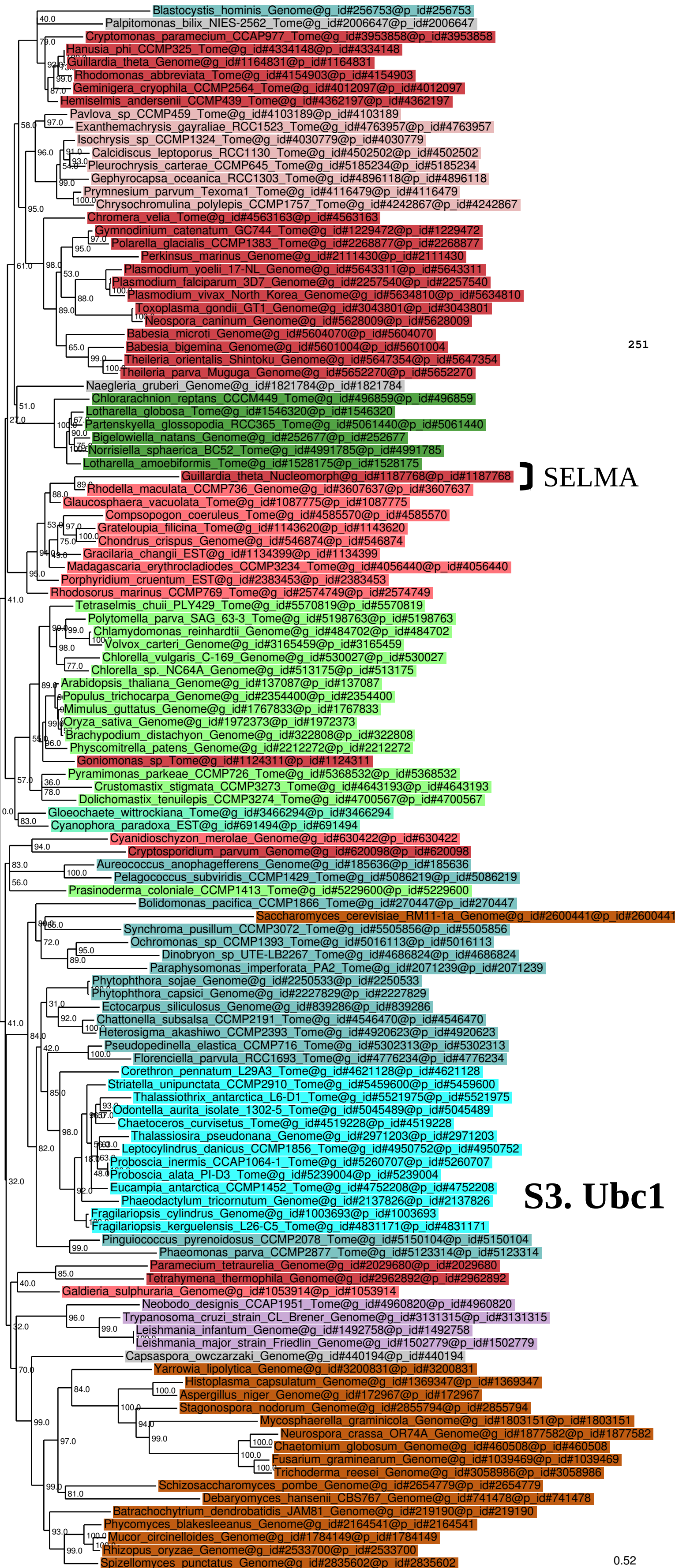
250

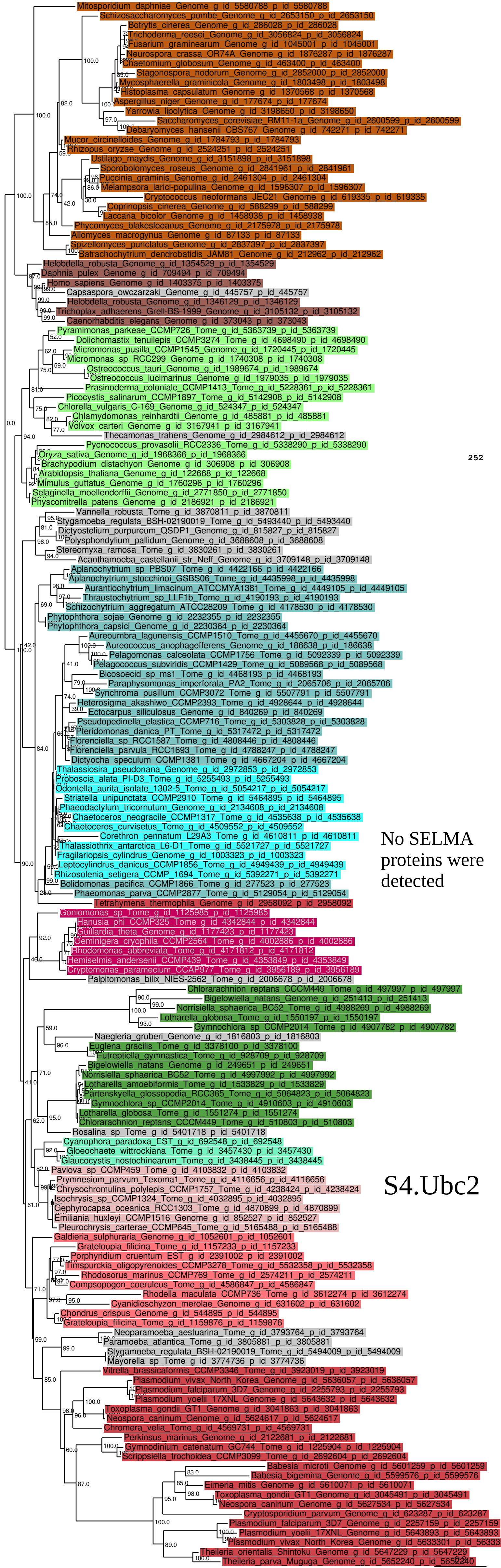
SELMA

S2.Ubc_new

SELMA

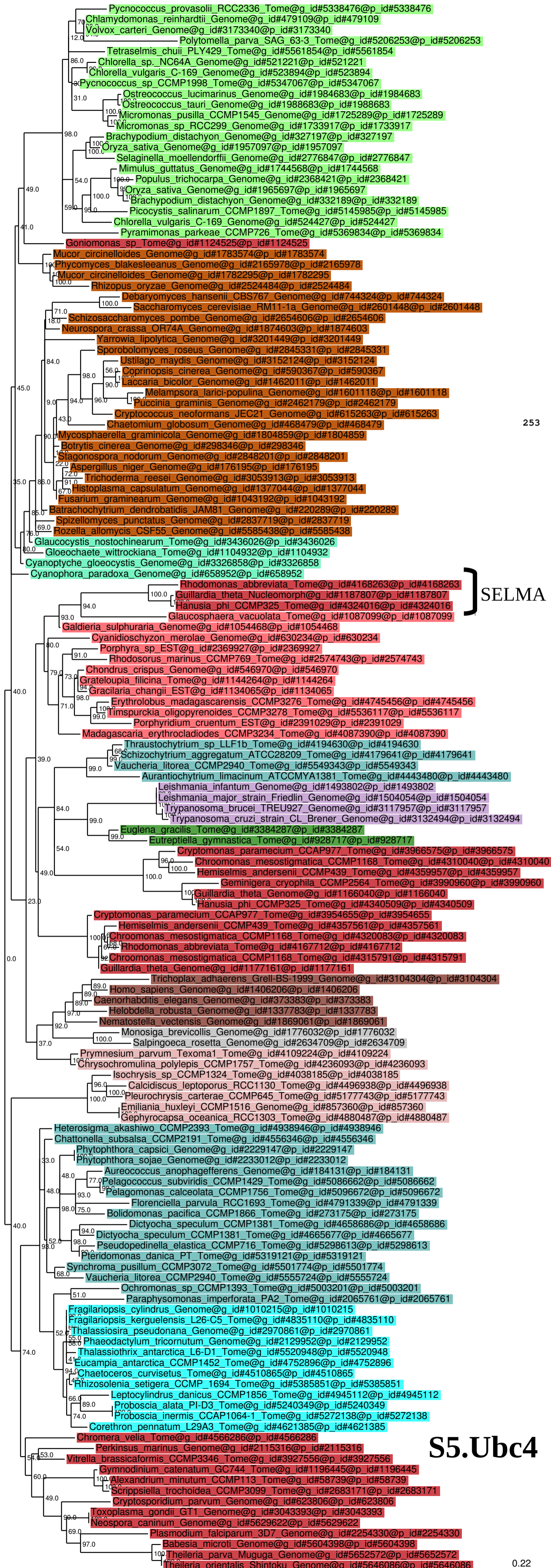
0.56

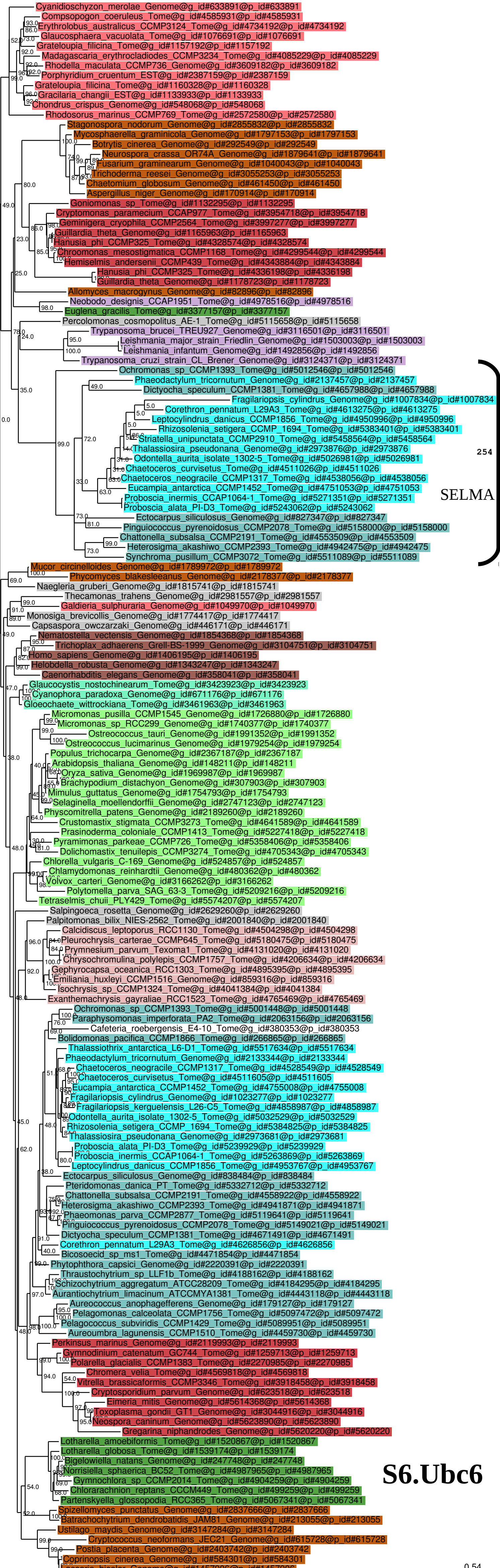




No SELMA proteins were detected

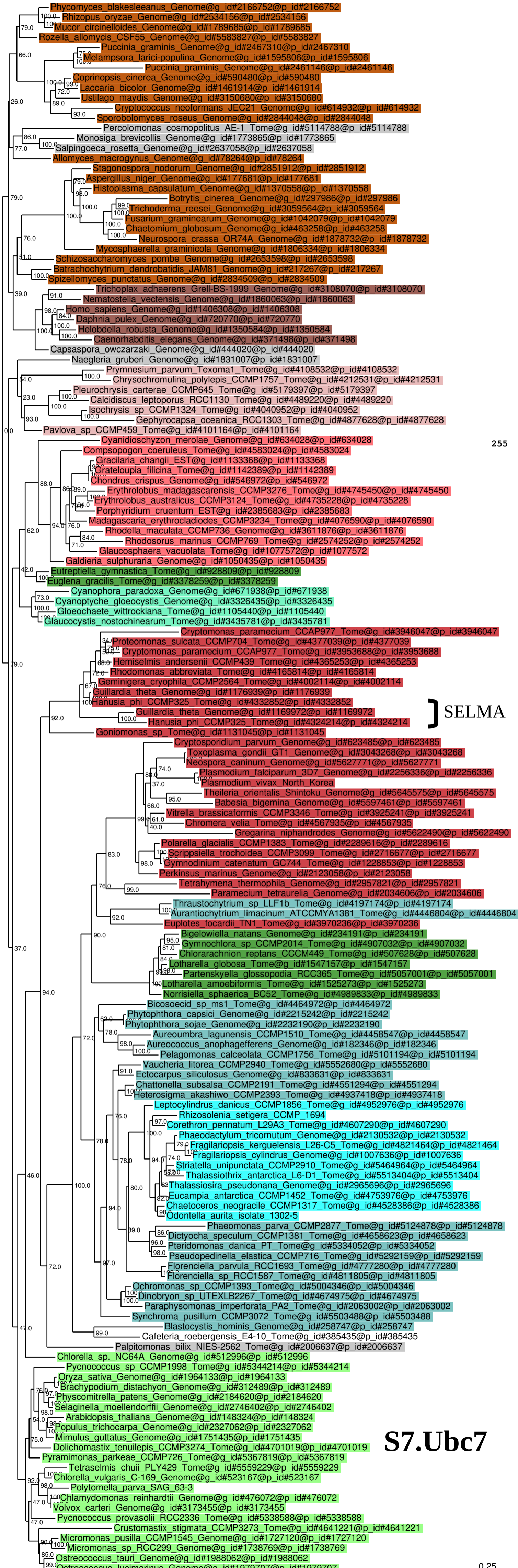
S4.Ubc2





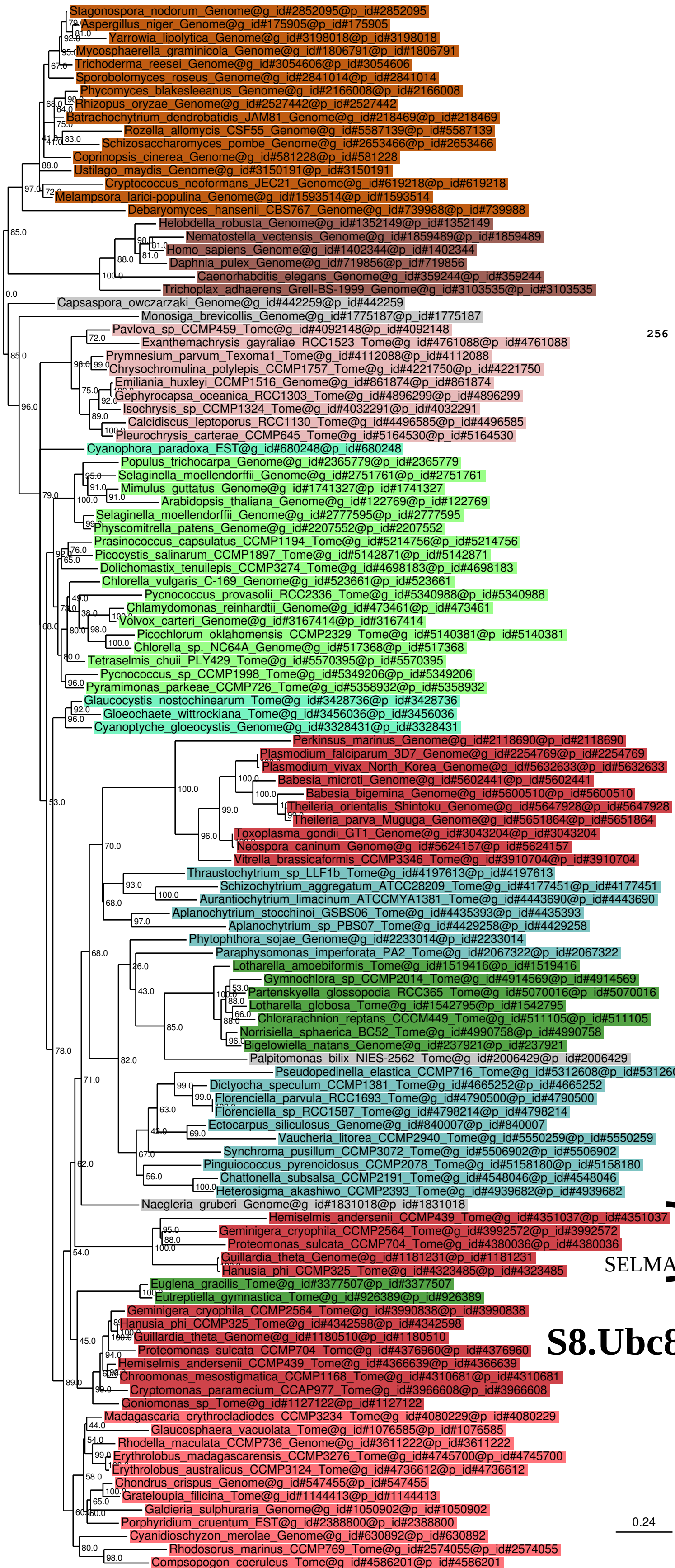
254
SELMA

S6.Ubc6



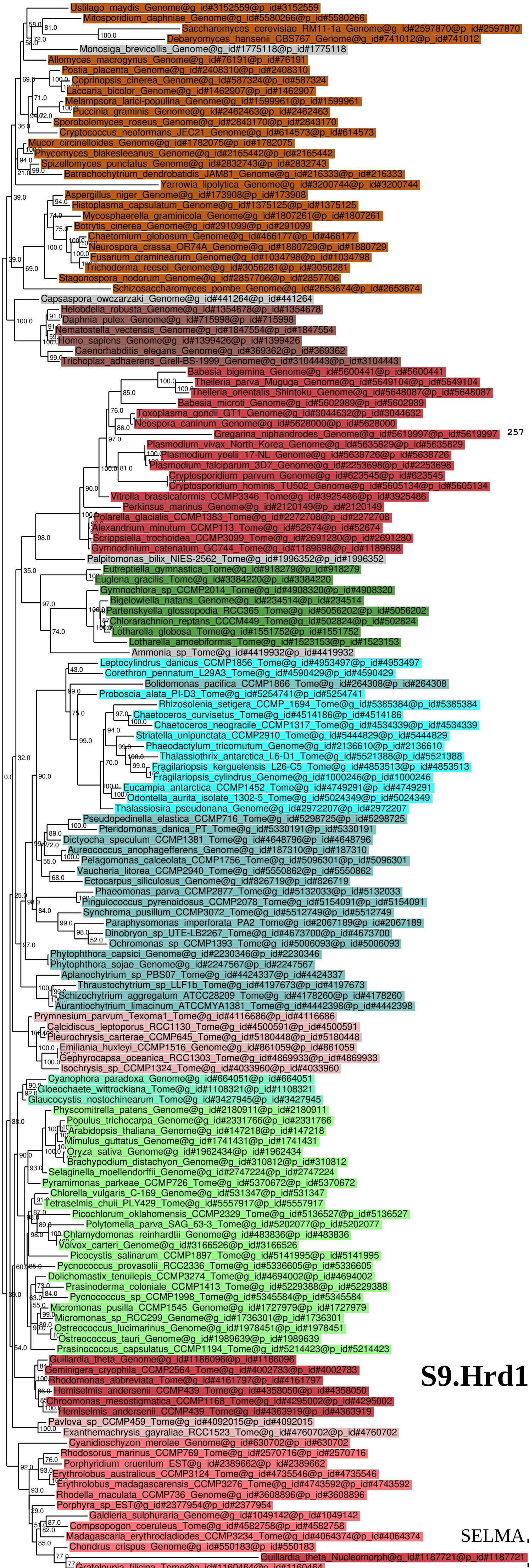
SELMA

S7.Ubc7



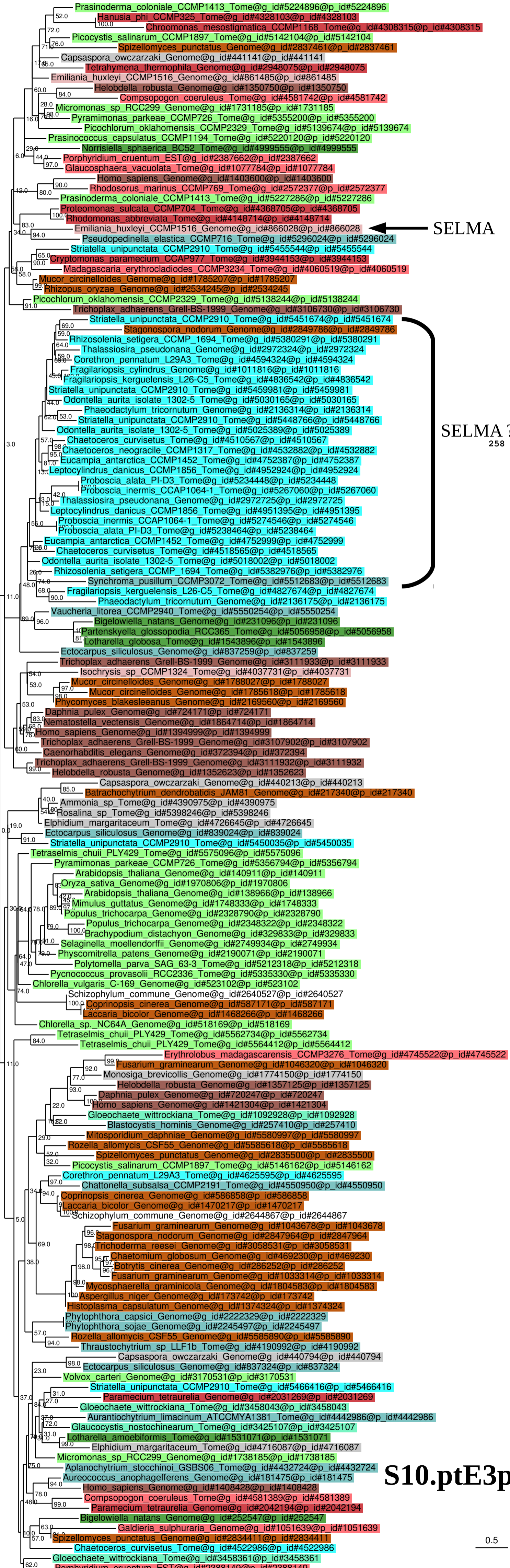
SELMA

S8.Ubc8



S9.Hrd1

SELMA



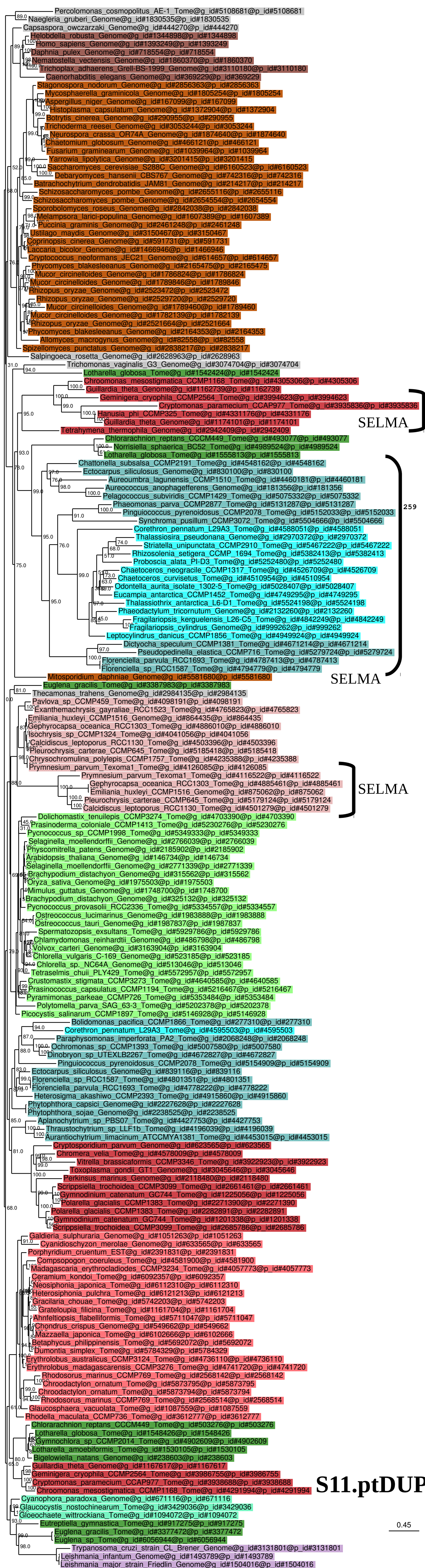
SELMA

SELMA?

258

S10.ptE3p

0.5



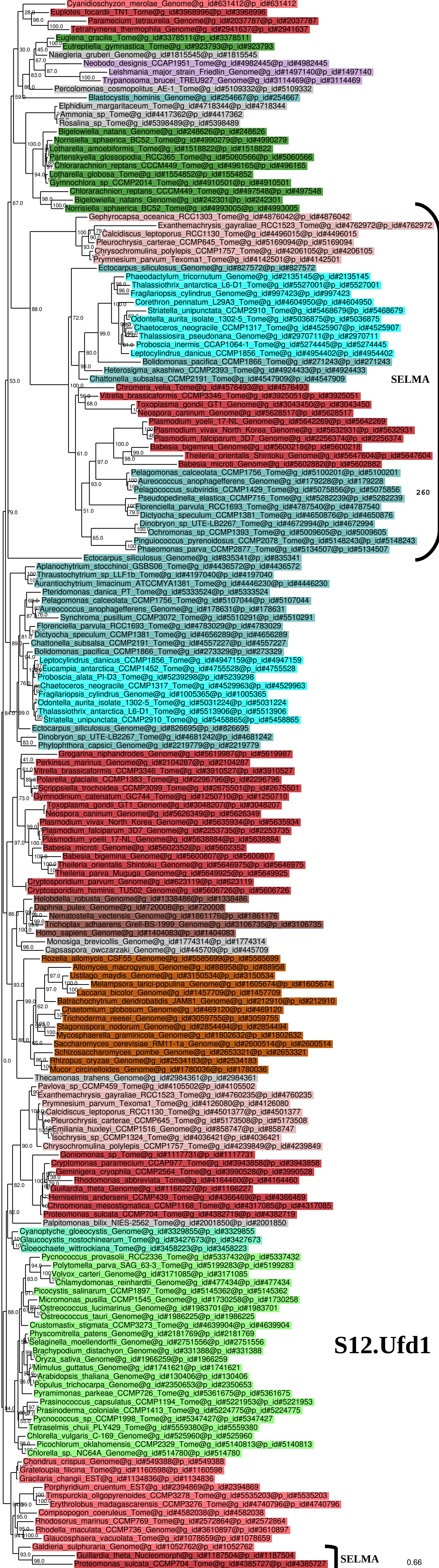
SELMA

259

SELMA

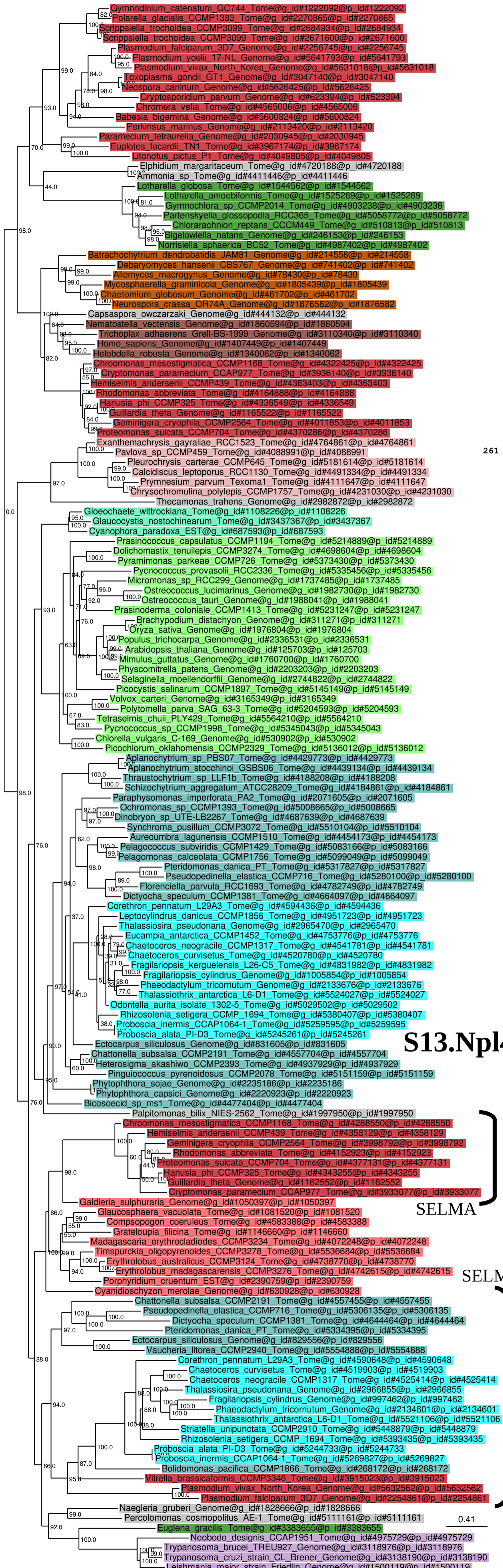
SELMA

S11.ptDUP



S12.Ufd1

SELMA 0.66



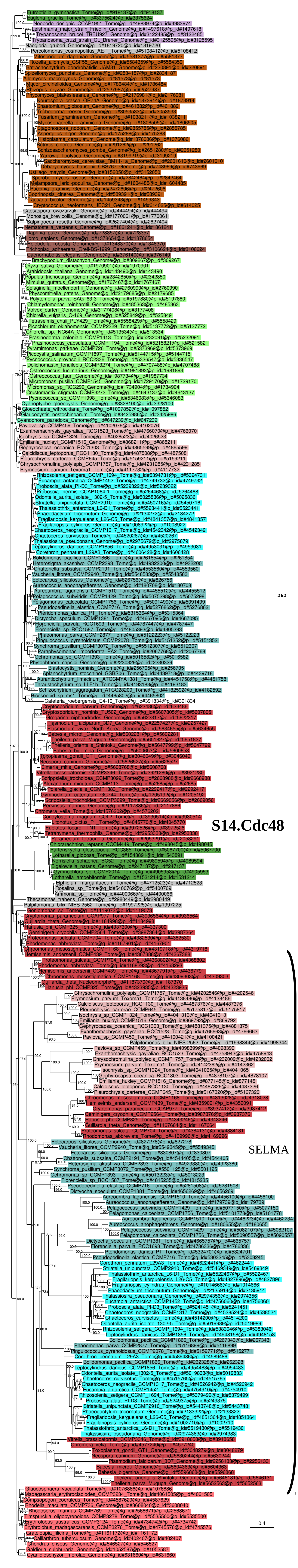
261

S13.Npl4

SELMA

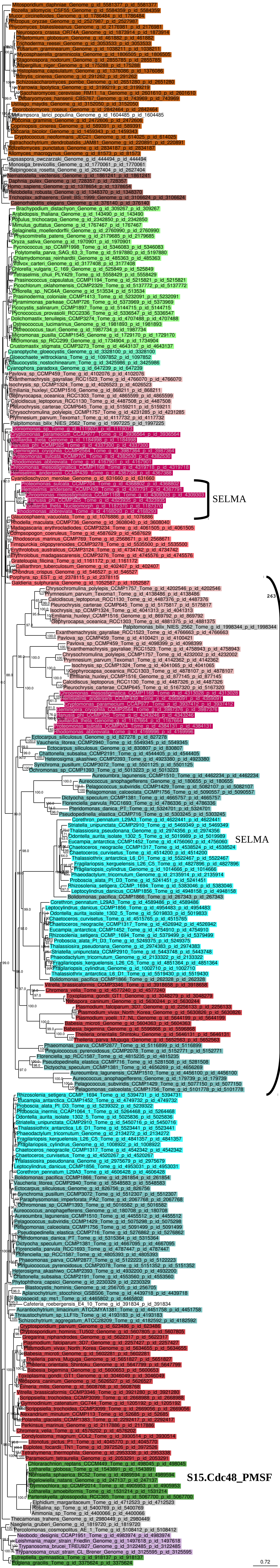
SELMA

0.41



S14.Cdc48

SELMA

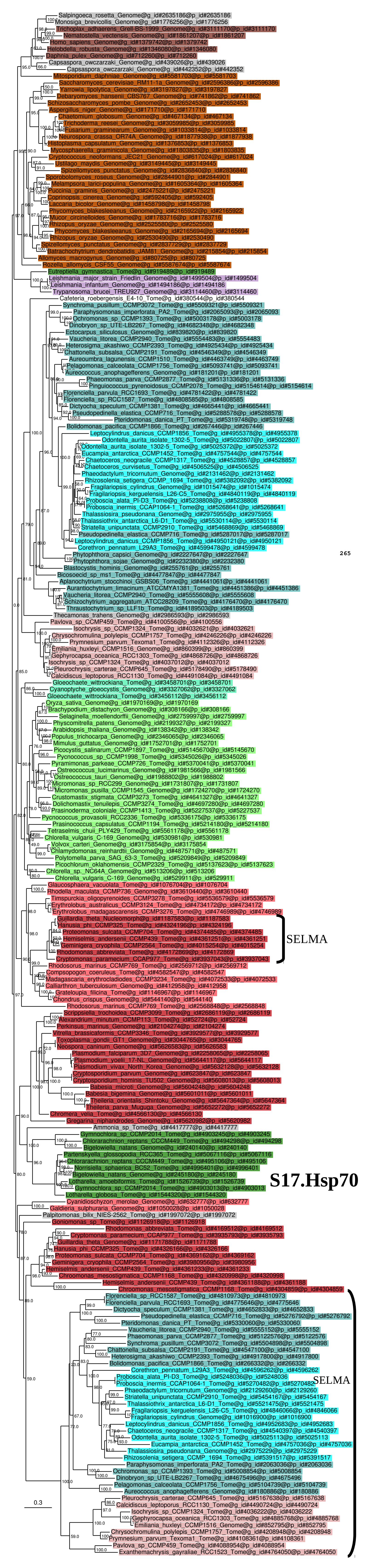


SELMA

SELMA

S15.Cdc48_PMSF

263



SELMA

S17.Hsp70

SELMA

0.3

RÉSUMÉ EN FRANÇAIS

Les eucaryotes photosynthétiques primaires proviennent de l'endosymbiose entre une cyanobactérie et un eucaryote hétérotrophe. Cette endosymbiose a donné naissance au super-groupe des Archaeplastida, un ensemble d'eucaryotes photosynthétiques possédant des plastes entourés de deux membranes et composée des glaucophytes, des algues rouges, des algues vertes et des plantes terrestres (Rodríguez-Ezpeleta et al., 2005; Yoon et al., 2004). Il existe un cas de deuxième endosymbiose primaire avec une cyanobactérie du clade *Synechococcus/Prochlorococcus* qui est à l'origine du chromatophore de *Paulinella chromatophora*, un Cercozoa amiboïde présent dans les écosystèmes d'eau douce (Lhee et al., 2017).

Suite à la diversification des Archaeplastida, d'autres lignées d'eucaryotes eucaryovores ont acquis des plastes secondaires par l'incorporation du plaste d'une algue rouge ou d'une algue verte. D'une part, les chlorarachniophytes et les euglénophytes ont des plastes secondaires verts dérivés de deux endosymbioses indépendantes avec des algues vertes (Rogers et al., 2007). D'autre part, les cryptophytes, haptophytes, stramenopiles et alvéolés contiennent des plastes secondaires rouges. Les événements endosymbiotiques primaires, secondaires et même tertiaires –dans le cas de certains groupes de dinoflagellés (Archibald, 2015; Gabrielsen et al., 2011)– sont à l'origine de toute la diversité des eucaryotes photosynthétiques actuels. Cependant, bien que les recherches de ces dernières années aient largement amélioré nos connaissances sur l'évolution des eucaryotes photosynthétiques, l'histoire évolutive complexe de ces lignées ainsi que l'absence de données génomiques dans certaines lignées laissent persister de nombreuses inconnues.

Pour mon travail de thèse, nous avons profité de la récente disponibilité de données génomiques et transcriptomiques pour un nombre élargi d'eucaryotes photosynthétiques, ainsi que de nouveaux génomes de cyanobactéries, pour appliquer un approche phylogénomique et ainsi approfondir notre appréhension de l'origine et de l'évolution des eucaryotes photosynthétiques et améliorer notre compréhension des endosymbioses primaires et secondaires. Ainsi, nous avons pu identifier la lignée cyanobactérienne la plus proche phylogénétiquement des plastes primaires et proposer quelques hypothèses pour retracer l'évolution des plastes secondaires. Les résultats les plus pertinents de mon travail de thèse sont résumés ci-dessous.

1. Origine des plastes primaires

L'origine unique des plastes primaires des Archaeplastida est largement soutenue par les analyses phylogénomiques des gènes nucléaires, plastidiaux et mitochondriaux (Jackson et Reyes-Prieto, 2014, Rodríguez-Ezpeleta et al., 2005). Cependant, l'identité de la lignée cyanobactérienne qui a donné naissance aux plastes primaires a été longuement débattue. Les études visant à son identification peuvent être divisés en deux groupes selon la position phylogénétique de la cyanobactérie endosymbiote dans la phylogénie de cyanobactéries. Les modèles « early-branching » proposent que la cyanobactérie à l'origine des plastes primaires aurait une origine précoce dans la phylogénie des cyanobactéries alors que les modèles « late-branching » suggèrent que les plastes dérivent d'une lignée cyanobactérienne qui se serait diversifiée plus tardivement.

Pour identifier la lignée cyanobactérienne la plus proche aux plastes primaires, il a été nécessaire d'étendre l'échantillonnage taxonomique des cyanobactéries. Quelques études ont récemment entrepris des projets de séquençage de génomes cyanobactériens qui ont augmenté l'échantillonnage taxonomique des espèces avec différents styles de vie, morphologies et positions phylogénétiques. Cette augmentation de données « omiques » nous a permis d'utiliser des approches phylogénomiques afin de mettre en évidence la lignée la plus proche de l'ancêtre du plaste. Cependant, le séquençage de *Gloeomargarita lithophora* par notre équipe, une cyanobactérie représentant un clade s'étant diversifié précocement a été décisif pour placer l'origine des plastes dans la phylogénie des cyanobactéries. *G. lithophora* appartient à un nouveau clade cyanobactérien récemment décrit dans les microbialites d'eau douce et les tapis microbiens (Ragon et al., 2014).

Nos analyses phylogénomiques des gènes plastidiaux concaténés ont montré clairement que *G. lithophora* est la cyanobactérie actuelle connue la plus proche de l'ancêtre des plastes primaires (Fig. 21). De même, les gènes d'origine cyanobactérien codés dans les noyaux (EGT, *Endosymbiotic Gene Transfer*, transfert endosymbiotique des gènes en anglais) des Archaeplastida corroborent les résultats des gènes des plastes (Fig. 22). Ces résultats montrent pour la première fois le groupe frère des plastes primaires et montrent aussi que le plaste n'est pas apparenté aux cyanobactéries filamenteuses capable de former des hétérocystes ou aux cyanobactéries diazotrophiques unicellulaires comme avait été proposé dans un certain nombre d'études précédentes (Dagan et al., 2013; Deusch et al., 2008; Falcón et al., 2010).



Figure 21. Arbre phylogénétique construit par inférence bayésienne de la concaténation de 97 protéines chloroplastiques (21,942 sites) avec leurs homologues cyanobactériennes. Les numéros indiquent des probabilités postérieures.



Figure 22. Arbre phylogénétique construit par inférence bayésienne de la concaténation de 72 protéines (28,102 sites) d'origine cyanobactérienne codées dans les noyaux des Archaeplastida. Les numéros indiquent des probabilités postérieures.

Un des arguments des études qui proposent une proximité phylogénétique entre les cyanobactéries filamenteuses et les plastes primaires est basé sur la surreprésentation apparente de protéines similaires à celles des cyanobactéries filamenteuses fixatrices d'azote dans les génomes nucléaires des Archaeplastida. Par exemple, Dagan et al., (2013) ont trouvé que les génomes des cyanobactéries filamenteuses des section IV et V codent le plus grand nombre de protéines homologues aux protéines d'origine cyanobactérienne dans les noyaux des espèces d'Archaeplastida qu'ils ont étudiées. Néanmoins, ces résultats sont probablement biaisés étant donné que ces lignées cyanobactériennes ont des génomes beaucoup plus grands que les autres cyanobactéries, atteignant jusqu'à 12 Mb, alors que les cyanobactéries unicellulaires basales (section I) ont des génomes de tailles comprises entre 3 et 5 Mb (Shih et al., 2013). Par exemple, la taille du génome de *G. lithophora* est de 3 Mb.

Il convient de signaler que même dans les analyses phylogénomiques des gènes concaténés de Dagan et ses collègues, les plastes ont une origine précoce mais ceci a été attribué au phénomène d'attraction des longues branches (LBA, *Long Branch Attraction* en anglais). Cependant, nous n'avons pas détecté un effet de LBA dans nos résultats. Les arbres phylogénétiques reconstruits avec des sous-ensembles de sites d'alignements séparés selon leurs taux d'évolution, ont montré que la relation phylogénétique de *G. lithophora* et les plastes n'était pas affectée par les taux d'évolution des sites, ce qui serait attendu si nos résultats étaient dû à un effet de LBA.

La plupart des modèles « late-branching » donne une importance majeure à la capacité de fixer l'azote de plusieurs clades cyanobactériens et suggère que l'établissement de l'endosymbiose primaire pourrait s'être initialement basée sur l'exploitation par l'hôte de cette capacité métabolique de la cyanobactérie symbiote. Cependant, le génome de *G. lithophora* ne code pas les protéines nécessaires à la fixation de l'azote, ce qui suggère que l'endosymbiose primaire n'était pas basée sur un échange de cet élément. De plus, les eucaryotes photosynthétiques ne gardent aucune trace de ce passé supposé des plastes.

Une conséquence importante de nos résultats sur la proximité phylogénétique de *Gloeomargarita* et l'ancêtre des plastes primaires concerne le type d'écosystème qui aurait hébergé les premiers eucaryotes photosynthétiques. Les cyanobactéries issues du nouvel ordre des Gloeomargaritales n'ont été retrouvées que dans des environnements d'eau douce (Ragon et al., 2014) et elles sont généralement très peu abondantes (Couradeau et al., 2011; Sghaï et al., 2015). La distribution écologique de *G. lithophora*

suggère que les premières eucaryotes photosynthétiques sont apparus dans un habitat d'eau douce. Nos résultats sont appuyés par la reconstruction des caractères ancestraux des eucaryotes photosynthétiques, qui conclue également à une émergence dans un environnement terrestre et confirme nos résultats en montrant que les plastes primaires dérivent d'un ancêtre cyanobactérien proche à *G. lithophora*.

2. Évolution des lignées d'eucaryotes avec des plastes secondaires

Dans mon travail de thèse, je me suis intéressé aussi à l'étude des plastes complexes. La diversité des plastes secondaires est large ainsi que leurs morphologies, mais il existe deux grandes divisions : les plastes d'origine « rouges » si le plaste dérive d'une endosymbiose avec une algue rouge, ou « verte » si l'endosymbiote est une algue verte. Les groupes possédant des plastes secondaires verts incluent les chlorarachniophytes et les euglénophytes dont les plastes sont issus de deux endosymbioses indépendantes (Rogers et al., 2007). Cependant, plusieurs protéines codées dans les noyaux des euglénophytes et des chlorarachniophytes semblent avoir été acquies par transfert de gènes depuis des algues rouges. La plupart de ces protéines participent à des fonctions liées à la photosynthèse (Markunas et Triemer, 2016; Yang et al., 2014) mais aussi à l'entretien et la division du plaste comme la protéine FtsZ. Celle-ci est nécessaire à la division du plaste, et dans le cas des chlorarachniophytes, elle dérive d'une algue rouge (Yang et al., 2014).

D'autre part, les noyaux des lignées eucaryotes possédant des plastes secondaires rouges ou chromalvéolés (i.e. les cryptophytes, haptophytes, stramenopiles et alvéolés) semblent coder plusieurs gènes dérivés d'algues vertes, ce qui a été interprété comme la preuve d'une endosymbiose cryptique impliquant une algue verte dans certaines de ces lignées (Moustafa et al., 2009). Pourtant, la réanalyse de ces gènes par Deschamps & Moreira, (2012) a réduit énormément le nombre de gènes qui semblent provenir des algues vertes.

Pour étudier l'évolution des eucaryotes avec des plastes secondaires, nous nous sommes intéressées dans l'analyse phylogénétique des EGT secondaires. Ce type de gènes sont potentiellement dérivés de transferts des gènes d'un endosymbiote vers le noyau de l'hôte. Ils représentent une source très importante d'information sur l'évolution des plastes parce qu'il est possible en principe d'en retracer l'histoire complète de transferts: d'abord, de l'endosymbiote cyanobactérien à l'ancêtre des Archaeplastida pendant l'endosymbiose primaire; ensuite, de l'algue endosymbiote (verte ou rouge) vers

les noyaux d'autres hôtes eucaryotes. La détection des EGT secondaires est utile pour identifier les événements d'endosymbioses cryptiques. Par exemple, bien que le parasite apicomplexe *Cryptosporidium parvum* n'ait pas d'apicoplaste (i.e. le plaste non-photosynthétique dérivé d'une algue rouge), le noyau a gardé un petit nombre de gènes d'origine plastidiale (EGT secondaires), ce qui suggère que *C. parvum* a eu un plaste mais qu'il l'a perdu secondairement (Huang et al., 2004).

Dans notre étude des gènes nucléaires chez les lignées eucaryotes possédant des plastes secondaires, nous avons pu identifier 85 gènes compatibles avec une topologie d'EGT secondaire. La plupart de ces gènes sont d'origine cyanobactérienne mais nous avons aussi trouvé un petit nombre de gènes bactériens non-cyanobactériens, possiblement acquis par HGT chez les Archaeplastida et ensuite transféré comme EGT pendant les endosymbioses secondaires. En ce qui concerne l'origine des EGT secondaires chez les chromalvéolés, la plupart des gènes (90%) a été transféré à partir d'une algue rouge, conformément à l'origine de leurs plastes secondaires. Le reste des gènes (10%) semble provenir des algues vertes. Nos résultats infirment les études précédentes qui ont rapporté un grand nombre de gènes verts dans les diatomées, suggérant une endosymbiose cryptique avec une algue verte (Moustafa et al., 2009). En revanche, ils sont compatibles avec celles de Deschamps & Moreira (2012) indiquant que les EGT secondaires verts avaient été identifiés à tort. De la même façon, nos résultats ne montrent pas de preuves suffisantes pour soutenir l'hypothèse d'une endosymbiose cryptique verte chez les diatomées ou dans d'autres lignées d'eucaryote possédant actuellement un plaste secondaire rouge. Il est possible que les quelques gènes d'algues vertes trouvés dans ces lignées puissent être le résultat de transferts horizontaux.

En ce qui concerne les chlorarachniophytes et les euglénophytes, deux lignées photosynthétiques avec des plastes secondaires verts, nous avons trouvé un mélange de gènes rouges et verts (Fig. 23), présents en proportions très semblables. Nos résultats montrent que 58% et 45% des gènes EGT proviennent d'algues rouges chez les chlorarachniophytes et les euglénophytes, respectivement. Cela représente une différence très importante comparé aux ~ 10% de gènes d'algues vertes dans les lignées rouges secondaires. L'importante contribution des algues rouges aux génomes nucléaires des chlorarachniophytes et des euglénophytes a été retrouvée à plusieurs reprises. Par exemple, Curtis et al. (2012) ont identifié que parmi les groupes de gènes provenant des algues et qui sont codés dans le noyau de *Bigeloviella natans*, 22 % des gènes sembleraient avoir été transférés des algues rouges.

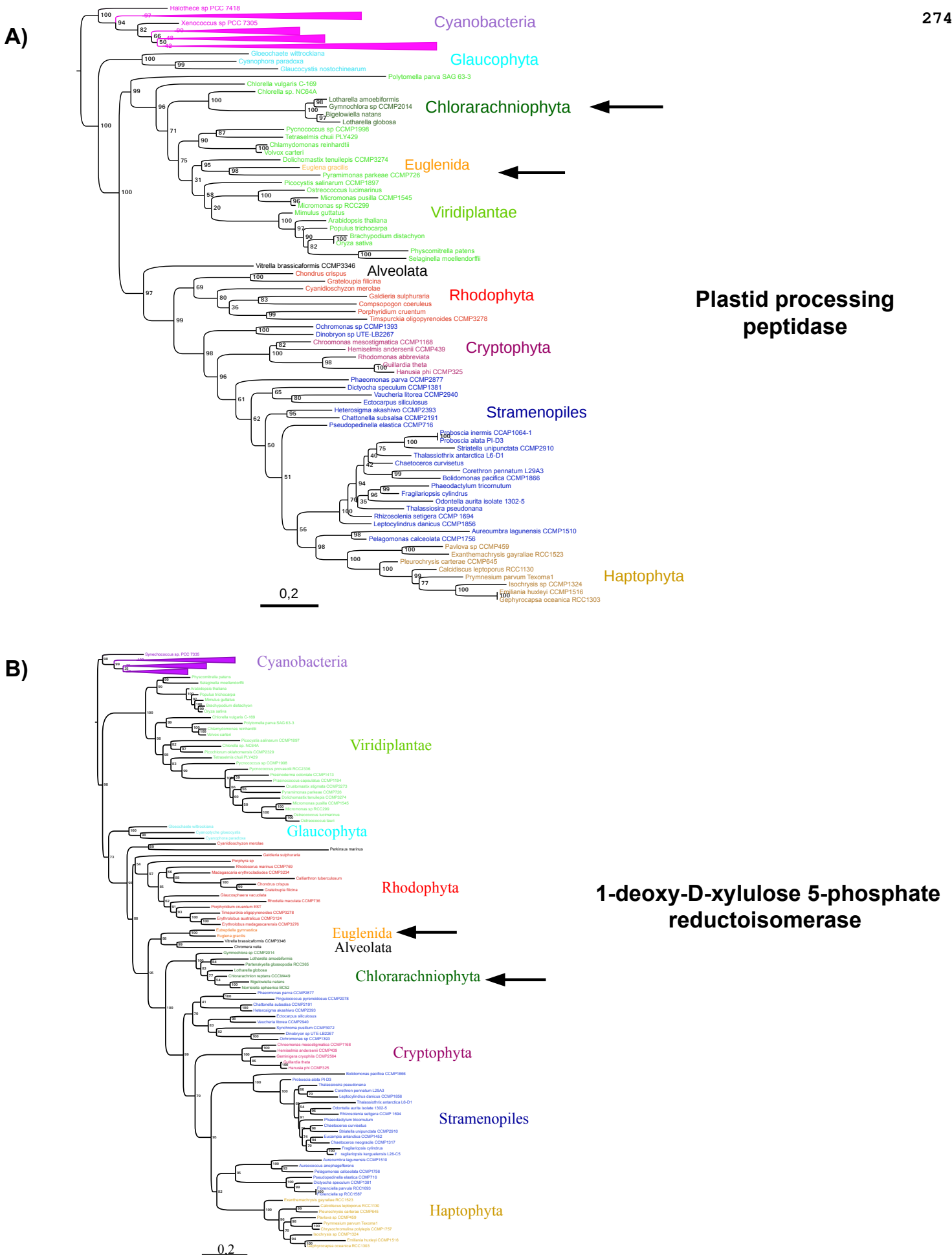


Figure 23. Protéines codées dans les génomes nucléaires de chlorarachniophytes et d'euglénophytes ont des origines diverses. A) Exemple de gène transféré à partir des algues vertes. B) Exemple de gène transféré à partir des algues rouges. Les arbres phylogénétique ont été reconstruits avec une approche de maximum de vraisemblance. Les numéros représentent des valeur de bootstrap.

De même, nos résultats montrent que les chlorarachniophytes et les euglénophytes ont un degré élevé de chimérisme. En plus, ces gènes participent à des fonctions importantes et variées dans le plaste telles que la synthèse des protéines, la biosynthèse de la chlorophylle et des acides gras ainsi que des fonctions liées à la photosynthèse. Il y a deux façons d'expliquer la grande quantité de gènes d'algues rouges dans ces lignées secondaires vertes : il y a eu une énorme quantité de HGT d'algues rouges vers les noyaux des chlorarachniophytes et des euglénophytes ou ces lignées ont eu un plaste secondaire rouge qui a été remplacé par le plaste vert présent actuellement. Pourtant, il est difficile de savoir si les gènes rouges sont effectivement ce qu'il a resté d'une endosymbiose cryptique avec une algue rouge. Il existe des cas où ce type de remplacement de plaste a eu lieu : le dinoflagellé *Lepidodinium chlorophorum* en est un exemple. Cette espèce possède actuellement un plaste secondaire vert mais son génome abrite un mélange de gènes d'algues rouges et vertes qui ont été recrutés à partir de l'ancien plaste rouge et du plaste vert actuel, respectivement (Sastre et al., 2010). Il semblerait que dans le contexte d'événements endosymbiotiques en série, les gènes d'endosymbioses antérieures aient été réutilisés. Ainsi, il est possible que les gènes rouges de *Lepidodinium* comme dans les cas des chlorarachniophytes et des euglénophytes auraient pu faciliter l'acquisition des nouveaux plastes verts.

3. Évolution du système de transport de protéines SELMA

La machinerie d'importation de protéines SELMA, qui se localise dans la deuxième membrane la plus externe des plastes secondaires rouges entourés par quatre membranes, semble dériver du système ERAD de l'algue rouge endosymbiote à l'origine du plaste. L'apparente monophylie de SELMA suggère une seule endosymbiose secondaire avec une algue rouge. Selon l'hypothèse des Chromalvéolés, SELMA aurait évolué dans l'ancêtre des chromistes et aurait été transmis aux lignées rouges secondaires actuelles par descendance verticale. Selon les hypothèses d'endosymbioses en série SELMA aurait été transféré en même temps que les plastes lors d'un nombre indéterminé d'endosymbioses tertiaires ou quaternaires, comme c'était le cas pour le complexe TIC/TOC au cours des endosymbioses secondaires (Hirakawa et al., 2012; Sheiner & Striepen, 2013). Il existe un exemple de transfert endosymbiotique du translocon SELMA chez les dinoflagellés de la famille des Kareniaceae. Il est bien établi que des membres de cette famille ont remplacé le plaste ancestral contenant de la péridinine par un plaste dérivé d'une haptophyte (Gabrielsen et al., 2011). Des études

récentes mais qui ne sont pas encore publiés (Kořený et al., cité par Waller & Kořený, 2017), ont montré que ces dinoflagellés encodent des protéines qui semblent à celles du système SELMA des haptophytes et qu'elles sont adressées au plaste. Ceci est un exemple remarquable montrant que SELMA peut être transféré endosymbiotiquement.

Dans les cas des cryptophytes, haptophytes, alvéolés et stramenopiles, plusieurs protéines faisant possiblement partie du système SELMA ont été détectées en utilisant des analyses bioinformatiques (Sommer et al., 2007, Stork et al., 2012). Cependant, un très faible nombre de reconstructions phylogénétiques permettant d'identifier leurs origines ont été publiées jusqu'à maintenant (Felsner et al., 2011; Petersen et al., 2014). Pour étudier l'origine et l'évolution de SELMA, nous avons effectué plusieurs analyses phylogénétiques des multiples composants du complexe. Nos analyses ont montré que, contrairement à ce qui est généralement accepté, la plupart des protéines du système SELMA ne sont pas monophylétiques. Nos arbres phylogénétiques montrent que la machinerie de translocation des cryptophytes (dans laquelle de nombreuses protéines SELMA restent encodées dans le nucléomorphe) diffère considérablement de la machinerie établie dans les autres lignées rouges (alvéolés, straménopiles et haptophytes). Il semble que le système SELMA des cryptophytes ait été établi à partir d'un set de protéines paralogues du système ERAD de l'algue rouge endosymbiote, différent de celui des autres groupes d'algues secondaires. Ces résultats sont difficiles à réconcilier avec l'hypothèse d'endosymbioses en série où le système SELMA des alvéolés, straménopiles et haptophytes aurait dérivé d'une cryptophyte symbiote qui aurait transmis le système lors d'une ou plusieurs endosymbioses tertiaires (Petersen et al., 2014).

Par conséquent, sous l'hypothèse d'endosymbioses en série, il est nécessaire de postuler que lors de l'endosymbiose tertiaire, la machinerie SELMA chez les cryptophytes était plus complexe et contenait plus de protéines que le système SELMA des cryptophytes actuelles. De plus, au cours de l'endosymbiose tertiaire, la nouvelle lignée rouge aurait recruté un sous-ensemble de composants de SELMA qui seraient différents de ceux retenus dans les cryptophytes contemporains. De même, cette hypothèse impliquerait que SELMA ait subi une simplification secondaire dans les cryptophytes au cours de leur évolution.

Si l'on considère l'hypothèse des Chromalvéolés, le système SELMA de l'ancêtre chromiste de toutes les lignées secondaires rouges aurait été plus complexe et constitué d'un plus grand nombre de protéines, mais aurait été simplifié pendant l'évolution de chaque lignée avec la perte de protéines paralogues dans certains cas. Il est possible que

le système SELMA ait été restructuré, peut-être en utilisant aussi des protéines dérivées de l'hôte dans certaines lignées. Ce modèle spécule une construction différentielle du système à partir du même ensemble initial de protéines. Pourtant, comme dans l'hypothèse d'endosymbioses en série, ces modèles semblent très peu parcimonieux.

Malgré le modèle d'origine de SELMA, il apparaît que la structure et la composition du système ait été très influencé par la perte du nucléomorphe et les composants du système codés par celui-ci qui n'ont pas été transféré vers les noyaux des haptophytes, alvéolés et straménopiles. Bien que nos résultats ne soient pas particulièrement favorables à une des possibles hypothèses, nous avons montré que l'évolution du système SELMA est beaucoup plus complexe qu'on ne le pensait auparavant. Ainsi, il sera nécessaire de continuer l'étude de SELMA de manière approfondie dans les recherches futures.

REFERENCES

- Adams, D. G., Bergman, B., Nierzwicki-Bauer, S. A., Duggan, P. S., Rai, A. N., & Schüßler, A. (2013). Cyanobacterial-plant symbioses. *The Prokaryotes: Prokaryotic Biology and Symbiotic Associations* (pp. 359–400).
- Adl, S. M., Simpson, A. G. B., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., ... Spiegel, F. W. (2012). The revised classification of eukaryotes. *Journal of Eukaryotic Microbiology*, *59*(5), 429–493.
- Agrawal, S., Chung, D. W. D., Ponts, N., van Dooren, G. G., Prudhomme, J., Brooks, C. F., ... Le Roch, K. G. (2013). An Apicoplast Localized Ubiquitylation System Is Required for the Import of Nuclear-encoded Plastid Proteins. *PLoS Pathogens*, *9*(6).
- Allen, A. E., Moustafa, A., Montsant, A., Eckert, A., Kroth, P. G., & Bowler, C. (2012). Evolution and functional diversification of fructose bisphosphate aldolase genes in photosynthetic marine diatoms. *Molecular Biology and Evolution*, *29*(1), 367–379.
- Allen, J. F. (2015). Why chloroplasts and mitochondria retain their own genomes and genetic systems: Colocation for redox regulation of gene expression. *Proceedings of the National Academy of Sciences*, *112*(33), 10231–10238.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–402.
- Amard, B., & Bertrand-Sarfati, J. (1997). Microfossils in 2000 Ma old cherty stromatolites of the Franceville Group, Gabon. *Precambrian Research*, *81*(3–4), 197–221.
- Anbar, A. D. (2002). Proterozoic Ocean Chemistry and Evolution: A Bioinorganic Bridge? *Science*, *297*(5584), 1137–1142.
- Archibald, J. M. (2015). Genomic perspectives on the birth and spread of plastids. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(33), 10147–10153.
- Archibald, J. M., Rogers, M. B., Toop, M., Ishida, K.-I., & Keeling, P. J. (2003). Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigeloviella natans*. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 7678–7683.
- Bahl, J., Lau, M. C. Y., Smith, G. J. D., Vijaykrishna, D., Cary, S. C., Lacap, D. C., ... Pointing, S. B. (2011). Ancient origins determine global biogeography of hot and cold desert cyanobacteria. *Nature Communications*, *2*(1).
- Ball, S. G., Subtil, A., Bhattacharya, D., Moustafa, A., Weber, A. P. M., Gehre, L., ... Dauvillée, D. (2013). Metabolic Effectors Secreted by Bacterial Pathogens: Essential Facilitators of Plastid Endosymbiosis? *The Plant Cell*, *25*(1), 7–21.

- Bateman, R. M., Crane, P. R., DiMichele, W. A., Kenrick, P. R., Rowe, N. P., Speck, T., & Stein, W. E. (1998). Early evolution of land plants: phylogeny, physiology, and ecology of the primary terrestrial radiation. *Annual Review of Ecology and Systematics*. 29(1), 263–292.
- Baurain, D., Brinkmann, H., Petersen, J. J., Rodriguez-Ezpeleta, N., Stechmann, A., Demoulin, V., ... Philippe, H. H. (2010). Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Molecular Biology and Evolution*. 27(7), 1698–1709.
- Becker, B., Hoef-Emden, K., & Melkonian, M. (2008). Chlamydial genes shed light on the evolution of photoautotrophic eukaryotes. *BMC Evolutionary Biology*. 8(1), 203.
- Becker, B., & Marin, B. (2009). Streptophyte algae and the origin of embryophytes. *Annals of Botany*. 103(5), 999–1004.
- Beghini, F., Pasolli, E., Truong, T. D., Putignani, L., Cacciò, S. M., & Segata, N. (2017). Large-scale comparative metagenomics of Blastocystis, a common member of the human gut microbiome. *The ISME Journal*. 11(12), 2848–2863.
- Bermudes, D., Margulis, L., & Tzertzinis, G. (1987). Prokaryotic Origin of Undulipodia: Application of the Panda Principle to the Centriole Enigma. *Annals of the New York Academy of Sciences*. 503(1), 187–197.
- Bhattacharya, D., Pelletreau, K. N., Price, D. C., Sarver, K. E., & Rumpho, M. E. (2013). Genome analysis of *Elysia chlorotica* egg DNA provides no evidence for horizontal gene transfer into the germ line of this kleptoplastic mollusc. *Molecular Biology and Evolution*. 30(8), 1843–1852.
- Bhushan, S., Kuhn, C., Berglund, A. K., Roth, C., & Glaser, E. (2006). The role of the N-terminal domain of chloroplast targeting peptides in organellar protein import and miss-sorting. *FEBS Letters*. 580(16), 3966–3972.
- Biard, T., Stemmann, L., Picheral, M., Mayot, N., Vandromme, P., Hauss, H., ... Not, F. (2016). In situ imaging reveals the biomass of giant protists in the global ocean. *Nature*. 532(7600), 504–507.
- Bicudo, C. E. de M., & Menezes, M. (2016). Phylogeny and Classification of Euglenophyceae: A Brief Review. *Frontiers in Ecology and Evolution*. 4, 17
- Bodył, A. (2017). Did some red alga-derived plastids evolve via kleptoplastidy? A hypothesis. *Biological Reviews*. 93(1), 201–222
- Bodył, A., Stiller, J. W., & Mackiewicz, P. (2009). Chromalveolate plastids: direct descent or multiple endosymbioses? *Trends in Ecology and Evolution*. 24(3), 119–21
- Bolte, K., Gruenheit, N., Felsner, G., Sommer, M. S., Maier, U. G., & Hempel, F. (2011). Making new out of old: Recycling and modification of an ancient protein translocation system during eukaryotic evolution. *BioEssays*, 33(5), 368–376.

- Bouchut, A., Geiger, J. A., Derocher, A. E., & Parsons, M. (2014). Vesicles bearing toxoplasma apicoplast membrane proteins persist following loss of the relict plastid or golgi body disruption. *PLoS ONE*, 9(11).
- Brinkman, F. S. L., Blanchard, J. L., Cherkasov, A., Av-Gay, Y., Brunham, R. C., Fernandez, R. C., ... Jones, S. J. M. (2002). Evidence that plant-like genes in Chlamydia species reflect an ancestral relationship between chlamydiaceae, cyanobacteria, and the chloroplast. *Genome Research*, 12(8), 1159–1167.
- Brocks, J. J., Jarrett, A. J. M., Sirantoine, E., Hallmann, C., Hoshino, Y., & Liyanage, T. (2017). The rise of algae in Cryogenian oceans and the emergence of animals. *Nature*, 548(7669), 578–581.
- Brown, M. W., Sharpe, S. C., Silberman, J. D., Heiss, A. A., Lang, B. F., Simpson, A. G. B., & Roger, A. J. (2013). Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. *Proceedings of the Royal Society B: Biological Sciences*, 280(1769), 20131755–20131755.
- Burgsdorf, I., Slaby, B. M., Handley, K. M., Haber, M., Blom, J., Marshall, C. W., ... Steindler, L. (2015). Lifestyle evolution in cyanobacterial symbionts of sponges. *mBio*, 6(3).
- Burki, F., Imanian, B., Hehenberger, E., Hirakawa, Y., Maruyama, S., & Keeling, P. J. (2014). Endosymbiotic gene transfer in tertiary plastid-containing dinoflagellates. *Eukaryotic Cell*, 13(2), 246–255.
- Burki, F., Inagaki, Y., Brate, J., Archibald, J. M., Keeling, P. J., Cavalier-Smith, T., ... Shalchian-Tabrizi, K. (2010). Large-Scale Phylogenomic Analyses Reveal That Two Enigmatic Protist Lineages, Telonemia and Centroheliozoa, Are Related to Photosynthetic Chromalveolates. *Genome Biology and Evolution*, 1(0), 231–238.
- Burki, F., Kaplan, M., Tikhonenkov, D. V., Zlatogursky, V., Minh, B. Q., Radaykina, L. V., ... Keeling, P. J. (2016). Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proceedings of the Royal Society B: Biological Sciences*, 283(1823), 20152802.
- Burki, F., & Keeling, P. J. (2014). Rhizaria. *Current Biology*, 24(3).
- Burki, F., Okamoto, N., Pombert, J.-F., & Keeling, P. J. (2012). The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. *Proceedings of the Royal Society B: Biological Sciences*, 279(1736), 2246–2254.
- Burki, F., Shalchian-Tabrizi, K., Minge, M., Skjaeveland, A., Nikolaev, S. I., Jakobsen, K. S., & Pawlowski, J. (2007). Phylogenomics reshuffles the eukaryotic supergroups. *PLoS ONE*, 2(8).

- Burki, F., Shalchian-Tabrizi, K., Pawlowski, J., Andersson, J. O., Roger, A. J., Arisue, N., ... Yang, Z. (2008). Phylogenomics reveals a new “megagroup” including most photosynthetic eukaryotes. *Biology Letters*, 4(4), 366–9.
- Butterfield, N. J. (2000). *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology*, 26(3), 386–404.
- Butterfield, N. J. (2004). A vaucheriacean alga from the middle Neoproterozoic of Spitsbergen: implications for the evolution of Proterozoic eukaryotes and the Cambrian explosion. *Paleobiology*, 30(2), 231–252.
- Calderon-Saenz, E., & Schnetter, R. (1989). Morphology, biology, and systematics of *Cryptochlora perforans* (Chlorarachniophyta), a phagotrophic marine alga. *Plant Systematics and Evolution*, 163(3–4), 165–176.
- Carpenter, E. J. (2002). Marine cyanobacterial symbioses. In *Biology and Environment* (Vol. 102, pp. 15–18).
- Carrapiço, F. (2015). Can We Understand Evolution Without Symbiogenesis? In *Reticulate Evolution, Interdisciplinary Evolution Research* (Vol. 3, pp. 81–105).
- Cavalier-Smith, T. (1982). The origins of plastids. *Biological Journal of the Linnean Society*, 17(3), 289–306.
- Cavalier-Smith, T. (1999). Principles of Protein and Lipid Targeting in Secondary Symbiogenesis: Euglenoid, Dinoflagellate, and Sporozoan Plastid Origins and the Eukaryote Family Tree, 2. *The Journal of Eukaryotic Microbiology*, 46(4), 347–366.
- Cavalier-Smith, T. (2002). The phagotrophic origin of eukaryotes and phylogenetic classification on protozoa. *International Journal of Systematic and Evolutionary Microbiology*.
- Cavalier-Smith, T. (2017). Kingdom Chromista and its eight phyla: a new synthesis emphasising periplastid protein targeting, cytoskeletal and periplastid evolution, and ancient divergences. *Protoplasma*, pp. 1–61.
- Cavalier-Smith, T., & Chao, E. E. (2003). Phylogeny and classification of phylum Cercozoa (Protozoa). *Protist*, 154(3–4), 341–358.
- Christa, G., Zimorski, V., Woehle, C., Tielens, A. G. M., Wagele, H., Martin, W. F., & Gould, S. B. (2013). Plastid-bearing sea slugs fix CO₂ in the light but do not require photosynthesis to survive. *Proceedings of the Royal Society B: Biological Sciences*, 281(1774), 20132493–20132493.
- Cihlář, J., Füssy, Z., Horák, A., & Oborník, M. (2016). Evolution of the tetrapyrrole biosynthetic pathway in secondary algae: Conservation, redundancy and replacement. *PLoS ONE*, 11(11).

- Cilingir, G., Broschat, S. L., & Lau, A. O. T. (2012). ApicoAP: The first computational model for identifying apicoplast-targeted proteins in multiple species of apicomplexa. *PLoS ONE*, 7(5).
- Cocquyt, E., Verbruggen, H., Leliaert, F., & De Clerck, O. (2010). Evolution and cytological diversification of the green seaweeds (Ulvophyceae). *Molecular Biology and Evolution*, 27(9), 2052–2061.
- Couradeau, E., Benzerara, K., Gerard, E., Moreira, D., Bernard, S., Brown Jr, G. E., & Lopez-Garcia, P. (2012). An early-branching microbialite cyanobacterium forms intracellular carbonates. *Science*, 336(4), 459–462.
- Couradeau, E., Benzerara, K., Moreira, D., Gérard, E., Kaźmierczak, J., Tavera, R., & López-García, P. (2011). Prokaryotic and eukaryotic community structure in field and cultured microbialites from the alkaline Lake Alchichica (Mexico). *PLoS ONE*, 6(12).
- Criscuolo, A., & Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology*, 10(1), 210.
- Criscuolo, A., & Gribaldo, S. (2011). Large-Scale phylogenomic analyses indicate a deep origin of primary plastids within cyanobacteria. *Molecular Biology and Evolution*, 28(11), 3019–3032.
- Curtis, B. A., Tanifuji, G., Burki, F., Gruber, A., Irimia, M., Maruyama, S., ... Archibald, J. M. (2012). Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature*, 492(7427), 59–65.
- Dagan, T., Roettger, M., Stucken, K., Landan, G., Koch, R., Major, P., ... Martin, W. F. (2013). Genomes of stigonematalean cyanobacteria (subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids. *Genome Biology and Evolution*, 5(1), 31–44.
- de Bary, A. (1879). *Die Erscheinung der Symbiose*. (V. V. K. J. Trubner, Ed.). Strassburg.
- de Vries, J., Christa, G., & Gould, S. B. (2014). Plastid survival in the cytosol of animal cells. *Trends in Plant Science*. 19(6), 347-350.
- Delaye, L., Valadez-cano, C., & Pérez-zamorano, B. (2016). How Really Ancient Is Paulinella Chromatophora ? *PLoS Currents Tree of Life*, 1–12.
- Delwiche, C. F., Kuhsel, M., & Palmer, J. D. (1995). Phylogenetic Analysis of *tufA* sequences indicates a Cyanobacterial origin of all plastids. *Molecular Phylogenetics and Evolution*. 4(2), 110-28.
- Delwiche, C. F., & Palmer, J. D. (1996). Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. *Molecular Biology and Evolution*, 13(6), 873–882.

- Deschamps, P. (2014). Primary endosymbiosis: Have cyanobacteria and Chlamydiae ever been roommates? *Acta Societatis Botanicorum Poloniae*, 83(4), 291–302.
- Deschamps, P., Colleoni, C., Nakamura, Y., Suzuki, E., Putaux, J. L., Buléon, A., ... Ball, S. (2008). Metabolic symbiosis and the birth of the plant kingdom. *Molecular Biology and Evolution*, 25(3), 536–548.
- Deschamps, P., & Moreira, D. (2009). Signal conflicts in the phylogeny of the primary photosynthetic eukaryotes. *Molecular Biology and Evolution*, 26(12), 2745–2753.
- Deschamps, P., & Moreira, D. (2012). Reevaluating the green contribution to diatom genomes. *Genome Biology and Evolution*, 4(7), 683–688.
- Deusch, O., Landan, G., Roettger, M., Gruenheit, N., Kowallik, K. V., Allen, J. F., ... Dagan, T. (2008). Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. *Molecular Biology and Evolution*, 25(4), 748–761.
- Dodge, J. D. (1971). A dinoflagellate with both a mesocaryotic and a eucaryotic nucleus I. Fine structure of the nuclei. *Protoplasma*, 73(2), 145–157.
- Domman, D., Horn, M., Embley, T. M., & Williams, T. A. (2015). Plastid establishment did not require a chlamydial partner. *Nature Communications*, 6, 6421.
- Doolittle, F. W. (1998). You are what you eat: A gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends in Genetics*, 14(8), 307–11.
- Dorrell, R. G., & Howe, C. J. (2012). What makes a chloroplast? Reconstructing the establishment of photosynthetic symbioses. *Journal of Cell Science*, 125(8), 1865–1875.
- Duchene, A.-M., Giritch, A., Hoffmann, B., Cognat, V., Lancelin, D., Peeters, N. M., ... Small, I. D. (2005). Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, 102(45), 16484–16489.
- Facchinelli, F., Colleoni, C., Ball, S. G., & Weber, A. P. M. (2013). Chlamydia, cyanobiont, or host: Who was on top in the ménage à trois? *Trends in Plant Science*, 18(12), 673–9.
- Falcón, L. I., Magallón, S., & Castillo, A. (2010). Dating the cyanobacterial ancestor of the chloroplast. *The ISME Journal*, 4(6), 777–783.
- Famintsyn. (1892). Nochmals die Zoochlorellen. *Biologisches Centralblatt.*, 12, 51–54.
- Famintsyn. (1907). Zapiski Imperatorskoi akademii nauk, fiz.-mat. otd. *Zapiski Imperatorskoi Akademii Nauk, Fiz.-Mat. Otd.*, (20), 1–14.
- Famintsyn, A. (1889). Beitrag zur Symbiose von Algen und Thieren. *Mémoires de l'Académie Impériale Des Sciences, St. Petersbourg*, 7, 26.

- Fang, L., Leliaert, F., Zhang, Z., Penny, D., & Zhong, B. (2017). Evolution of the Chlorophyta: Insights from chloroplast phylogenomic analyses. *Journal of Systematics and Evolution*, 55(4), 322–332.
- Fathinejad, S., Steiner, J. M., Reipert, S., Marchetti, M., Allmaier, G., Burey, S. C., ... Bohnert, H. J. (2008). A carboxysomal carbon-concentrating mechanism in the cyanelles of the “coelacanth” of the algal world, *Cyanophora paradoxa*? In *Physiologia Plantarum* (Vol. 133, pp. 27–32).
- Feinstein, T. N., Traslavina, R., Sun, M. Y., & Lin, S. (2002). Effects of light on photosynthesis, grazing, and population dynamics of the heterotrophic dinoflagellate *Pfiesteria piscicida* (Dinophyceae). *Journal of Phycology*, 38(4), 659–669.
- Felsner, G., Sommer, M. S., Gruenheit, N., Hempel, F., Moog, D., Zauner, S., ... Maier, U. G. (2011). ERAD components in organisms with complex red plastids suggest recruitment of a preexisting protein transport pathway for the periplastid membrane. *Genome Biology and Evolution*, 3(1), 140–150.
- Flores, E., & Herrero, A. (2010). Compartmentalized function through cell differentiation in filamentous cyanobacteria. *Nature Reviews Microbiology*.
- Friedl, T., & Bhattacharya, D. (2002). Origin and Evolution of Green Lichen Algae BT - Symbiosis: Mechanisms and Model Systems. In J. Seckbach (Ed.) (pp. 341–357). Dordrecht: Springer Netherlands.
- Frommolt, R., Werner, S., Paulsen, H., Goss, R., Wilhelm, C., Zauner, S., ... Lohr, M. (2008). Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis. *Molecular Biology and Evolution*, 25(12), 2653–2667.
- Gabrielsen, T. M., Minge, M. A., Espelund, M., Tooming-Klunderud, A., Patil, V., Nederbragt, A. J., ... Jakobsen, K. S. (2011). Genome evolution of a tertiary dinoflagellate plastid. *PLoS ONE*, 6(4).
- Garcia-Cuetos, L., Moestrup, Ø., Hansen, P. J., & Daugbjerg, N. (2010). The toxic dinoflagellate *Dinophysis acuminata* harbors permanent chloroplasts of cryptomonad origin, not kleptochloroplasts. *Harmful Algae*, 9(1), 25–38.
- Garcia-Pichel, F., Nübel, U., & Muyzer, G. (1998). The phylogeny of unicellular, extremely halotolerant cyanobacteria. *Archives of Microbiology*, 169(6), 469–482.
- Gibbs, S. P. (1962). The ultrastructure of the chloroplasts of algae. *Journal of Ultrastructure Research*, 7(5–6), 418–435.
- Gibbs, S. P. (1978). May Have Evolved From Symbiotic Green Algae. *Canadian Journal of Botany*, 56(22), 2883–2889.
- Gibbs, S. P. (1981). The Chloroplasts of some algal groups may have evolved from endosymbiotic eukaryotic algae. *Annals of the New York Academy of Sciences*, 361(1), 193–208.

- Gilson, P. R., Su, V., Slamovits, C. H., Reith, M. E., Keeling, P. J., & McFadden, G. I. (2006). Complete nucleotide sequence of the chlorarachniophyte nucleomorph: Nature's smallest nucleus. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, *103*, 9566–9571.
- Glöckner, G., Hülsmann, N., Schleicher, M., Noegel, A. A., Eichinger, L., Gallinger, C., ... Schliwa, M. (2014). The genome of the foraminiferan reticulomyxa filosa. *Current Biology*, *24*(1), 11–18.
- Gockel, G., & Hachtel, W. (2000). Complete Gene Map of the Plastid Genome of the Nonphotosynthetic Euglenoid Flagellate *Astasia longa*. *Protist*, *151*(4), 347–351.
- Gornik, S. G., Febrimarsa, Cassin, A. M., MacRae, J. I., Ramaprasad, A., Rchiad, Z., ... Waller, R. F. (2015). Endosymbiosis undone by stepwise elimination of the plastid in a parasitic dinoflagellate. *Proceedings of the National Academy of Sciences*, *112*(18), 5767–5772.
- Gould, S. B., Maier, U. G., & Martin, W. F. (2015). Protein Import and the Origin of Red Complex Plastids. *Current Biology*, *25*(12), 515–521.
- Gray, M. W., & Doolittle, W. F. (1982). Has the endosymbiont hypothesis been proven? *Microbiological Reviews*, *46*(1), 1–42.
- Green, B. R., & Parson, W. W. (2003). Light-harvesting antennas in photosynthesis. *Advances in Photosynthesis and Respiration*.
- Gross, J., Meurer, J., & Bhattacharya, D. (2008). Evidence of a chimeric genome in the cyanobacterial ancestor of plastids. *BMC Evolutionary Biology*, *8*(1), 117.
- Gruber, A., Rocap, G., Kroth, P. G., Armbrust, E. V., & Mock, T. (2015). Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant Journal*, *81*(3), 519–528.
- Gschloessl, B., Guermeur, Y., & Cock, J. M. (2008). HECTAR: A method to predict subcellular targeting in heterokonts. *BMC Bioinformatics*, *9*(1), 393.
- Guilliermond, A. (1919). Sur l'origine mitochondriale des plastides. *Annales Des Sc. Nat. Botaniques*, (X), 225.
- Guindon, S., Delsuc, F., Dufayard, J. F., & Gascuel, O. (2009). Estimating maximum likelihood phylogenies with PhyML. *Methods in Molecular Biology*, *537*, 113–137.
- Guiry, M. D., & Guiry, G. M. (2017). AlgaeBase. World-wide electronic publication, National University of Ireland, Galway. <http://www.algaebase.org>; searched on 15 october 2017
- Hackett, J. D., Anderson, D. M., Erdner, D. L., & Bhattacharya, D. (2004). Dinoflagellates: A remarkable evolutionary experiment. *American Journal of Botany*. *91*(10), 1523-34.

- Hadariová, L., Vesteg, M., Birčák, E., Schwartzbach, S. D., & Krajčovič, J. (2017). An intact plastid genome is essential for the survival of colorless *Euglena longa* but not *Euglena gracilis*. *Current Genetics*, 63(2), 331–341.
- Hagopian, J. C., Reis, M., Kitajima, J. P., Bhattacharya, D., & De Oliveira, M. C. (2004). Comparative analysis of the complete plastid genome sequence of the red alga *Gracilaria tenuistipitata* var. *liui* provides insights into the evolution of rhodoplasts and their relationship to other plastids. *Journal of Molecular Evolution*, 59(4), 464–477.
- Haguenau, F., Hawkes, P. W., Hutchison, J. L., Satiat-Jeunemaître, B., Simon, G. T., & Williams, D. B. (2003). Key Events in the History of Electron Microscopy. *Microscopy and Microanalysis*, 9(2), 96–138.
- Hamilton, T. L., Bryant, D. A., & Macalady, J. L. (2016). The role of biology in planetary evolution: Cyanobacterial primary production in low-oxygen Proterozoic oceans. *Environmental Microbiology*, 18(2), 325–40.
- Händeler, K., Grzybowski, Y. P., Krug, P. J., & Wägele, H. (2009). Functional chloroplasts in metazoan cells - a unique evolutionary strategy in animal life. *Frontiers in Zoology*, 6(1), 28.
- Hempel, F., Bullmann, L., Lau, J., Zauner, S., & Maier, U. G. (2009). ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. *Molecular Biology and Evolution*, 26(8), 1781–1790.
- Hempel, F., Felsner, G., & Maier, U. G. (2010). New mechanistic insights into pre-protein transport across the second outermost plastid membrane of diatoms. *Molecular Microbiology*, 76(3), 793–801.
- Hibberd, D. J., & Norris, R. E. (1984). Cytology and ultrastructure of Chlorarachnion reptans (Chlorarachniophyta Divisio Nova, Chlorarachniophyceae Classis Nova). *Journal of Phycology*, 20(2), 310–330.
- Hirakawa, Y. (2017). Chlorarachniophytes With Complex Secondary Plastids of Green Algal Origin. *Advances in Botanical Research*, 84, 359–393.
- Hirakawa, Y., Burki, F., & Keeling, P. J. (2012). Genome-based reconstruction of the protein import machinery in the secondary plastid of a chlorarachniophyte alga. *Eukaryotic Cell*, 11(3), 324–333.
- Hirakawa, Y., & Ishida, K.-I. (2014). Polyploidy of Endosymbiotically Derived Genomes in Complex Algae. *Genome Biology and Evolution*, 6(4), 974–980.
- Holland, H. D. (2006). The Oxygenation of the aAtmosphere and Oceans. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 361(1470), 903–915.
- Honegger, R. (2000). Simon Schwendener (1829–1919) and the Dual Hypothesis of Lichens. *The Bryologist*, 103(2), 307–313.

- Horiguchi, T., & Takano, Y. (2006). Serial replacement of a diatom endosymbiont in the marine dinoflagellate *Peridinium quinquecorne* (Peridinales, Dinophyceae). *Phycological Research*, 54(3), 193–200.
- Horn, M. (2008). *Chlamydiae* as Symbionts in Eukaryotes. *Annual Review of Microbiology*, 62(1), 113–131.
- Howe, A. T., Bass, D., Scoble, J. M., Lewis, R., Vickerman, K., Arndt, H., & Cavalier-Smith, T. (2011). Novel Cultured Protists Identify Deep-branching Environmental DNA Clades of Cercozoa: New Genera *Tremula*, *Micrometopion*, *Minimassisteria*, *Nudifila*, *Peregrinia*. *Protist*, 162(2), 332–372.
- Hrdá, Š., Fousek, J., Szabová, J., Hampl, V., & Vlček, Č. (2012). The plastid genome of eutreptiella provides a window into the process of secondary endosymbiosis of plastid in euglenids. *PLoS ONE*, 7(3).
- Huang, J., & Gogarten, J. P. (2007). Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? *Genome Biology*, 8(6), R99.
- Huang, J., Mullapudi, N., Lancto, C. a, Scott, M., Abrahamsen, M. S., & Kissinger, J. C. (2004). Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in *Cryptosporidium parvum*. *Genome Biology*, 5(11), R88.
- Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., ... Bork, P. (2016). EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research*, 44(D1), 286–293.
- Husnik, F., & McCutcheon, J. P. (2017). Functional horizontal gene transfer from bacteria to eukaryotes. *Nature Reviews Microbiology*. 16, 67-79.
- Hwan, S. Y., Hackett, J. D., Van Dolah, F. M., Nosenko, T., Lidie, K. L., & Bhattacharya, D. (2005). Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. *Molecular Biology and Evolution*, 22(5), 1299–1308.
- Imanian, B., & Keeling, P. J. (2007). The dinoflagellates *Durinskia baltica* and *Kryptoperidinium foliaceum* retain functionally overlapping mitochondria from two evolutionarily distinct lineages. *BMC Evolutionary Biology*, 7(1), 172.
- Ishida, K., Green, B. R., & Cavalier-Smith, T. (1999). Diversification of a Chimaeric Algal Group, the Chlorarachniophytes: Phylogeny of Nuclear and Nucleomorph Small-Subunit rRNA Genes. *Molecular Biology and Evolution*, 16(3), 321–331.
- Jackson, C., Clayden, S., & Reyes-Prieto, A. (2015). The Glaucophyta: The blue-green plants in a nutshell. *Acta Societatis Botanicorum Poloniae*. 84(2), 149-165.
- Jackson, C., & Reyes-Prieto, A. (2014). The mitochondrial genomes of the glaucophytes *gloeochaete wittrockiana* and *cyanoptycha gloeocystis*: Multilocus phylogenetics

suggests monophyletic archaeplastida. *Genome Biology and Evolution*, 6(10), 2774–2785.

- Janouškovec, J., Gavelis, G. S., Burki, F., Dinh, D., Bachvaroff, T. R., Gornik, S. G., ... Saldarriaga, J. F. (2017). Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. *Proceedings of the National Academy of Sciences*, 114(2), 171–180.
- Janouskovec, J., Horak, A., Obornik, M., Lukes, J., & Keeling, P. J. (2010). A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proceedings of the National Academy of Sciences*, 107(24), 10949–10954.
- Johnson, M. D., Oldach, D., Delwiche, C. F., & Stoecker, D. K. (2007). Retention of transcriptionally active cryptophyte nuclei by the ciliate *Myrionecta rubra*. *Nature*, 445(7126), 426–428.
- Kamikawa, R., Tanifuji, G., Kawachi, M., Miyashita, H., Hashimoto, T., & Inagaki, Y. (2015). Plastid genome-based phylogeny pinpointed the origin of the green-colored plastid in the dinoflagellate *Lepidodinium chlorophorum*. *Genome Biology and Evolution*, 7(4), 1133–1140.
- Karkar, S., Facchinelli, F., Price, D. C., Weber, A. P. M., & Bhattacharya, D. (2015). Metabolic connectivity as a driver of host and endosymbiont integration. *Proceedings of the National Academy of Sciences*, 112(33), 10208–10215.
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780.
- Keeling, P. J. (2009). Functional and ecological impacts of horizontal gene transfer in eukaryotes. *Current Opinion in Genetics and Development*, 19(6), 613–619.
- Keeling, P. J. (2010). The endosymbiotic origin, diversification and fate of plastids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1541), 729–748.
- Khachane, A. N., Timmis, K. N., & Martins Dos Santos, V. A. P. (2007). Dynamics of reductive genome evolution in mitochondria and obligate intracellular microbes. *Molecular Biology and Evolution*, 24(2), 449–456.
- Khakhina, L. N. (1992). *Concepts of symbiogenesis: A historical and critical study of the research of Russian botanists* (Yale University).
- Kies, L. (1974). Elektronenmikroskopische Untersuchungen an *Paulinella chromatophora* Lauterborn, einer Thekamöbe mit blau-grünen Endosymbionten (Cyanellen). *Protoplasma*, 80(1), 69–89.
- Kim, S., & Park, M. G. (2016). *Paulinella longichromatophora* sp. nov., a New Marine Photosynthetic Testate Amoeba Containing a Chromatophore. *Protist*, 167(1), 1–12.

- Kim, Moore, C. E., Archibald, J. M., Bhattacharya, D., Yi, G., Yoon, H. S., & Shin, W. (2017). Evolutionary dynamics of cryptophyte plastid genomes. *Genome Biology and Evolution*, 9(7), 1859–1872.
- Kneip, C., Lockhart, P., Voß, C., & Maier, U.-G. (2007). Nitrogen fixation in eukaryotes—New models for symbiosis. *BMC Evolutionary Biology*, 7, 1–12.
- Knoll, A. H. (2012). Lynn Margulis, 1938–2011. *Proceedings of the National Academy of Sciences*, 109(4), 1022–1022.
- Knoll, A. H., Strother, P. K., & Rossi, S. (1988). Distribution and diagenesis of microfossils from the lower proterozoic duck creek dolomite, Western Australia. *Precambrian Research*, 38(3), 257–279.
- Korshikov, A. (1924). Protistologische Beobachtungen. I Cyanophora paradoxa n. g. et sp. *Russ Arch Protistol.*, 3, 57–74.
- Koutsovoulos, G., Kumar, S., Laetsch, D., Stevens, L., Daub, J., & Conlon, C. (2016). No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*. *Pnas*, 113(23), 1–6.
- Kovács-Bogdán, E., Soll, J., & Bölter, B. (2010). Protein import into chloroplasts: The Tic complex and its regulation. *Biochimica et Biophysica Acta - Molecular Cell Research*.
- Ku, C., Nelson-Sathi, S., Roettger, M., Sousa, F. L., Lockhart, P. J., Bryant, D., ... Martin, W. F. (2015). Endosymbiotic origin and differential loss of eukaryotic genes. *Nature*, 524(7566), 427–432.
- Lahey, B., & Triemer, R. (2016). The tetrapyrrole synthesis pathway as a model of horizontal gene transfer in euglenoids. *Journal of Phycology*, 217, 198–217.
- Lalonde, S. V., & Konhauser, K. O. (2015). Benthic perspective on Earth's oldest evidence for oxygenic photosynthesis. *Proceedings of the National Academy of Sciences*, 112(4), 995–1000.
- Larkum, A. W. D., Lockhart, P. J., & Howe, C. J. (2007). Shopping for plastids. *Trends in Plant Science*, 12(5), 189–195.
- Lartillot, N., & Philippe, H. H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution*, 21(6), 1095–1109.
- Lartillot, N., Rodrigue, N., Stubbs, D., & Richer, J. (2013). Phylobayes mpi: Phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Systematic Biology*, 62(4), 611–615.
- Le, S. Q., & Gascuel, O. (2008). An improved general amino acid replacement matrix. *Molecular Biology and Evolution*, 25(7), 1307–1320.

- Lee, J., Cho, C. H., Park, S. I., Choi, J. W., Song, H. S., West, J. A., ... Yoon, H. S. (2016). Parallel evolution of highly conserved plastid genome architecture in red seaweeds and seed plants. *BMC Biology*, *14*(1), 75.
- Leliaert, F., Smith, D. R., Moreau, H., Herron, M. D., Verbruggen, H., Delwiche, C. F., & De Clerck, O. (2012). Phylogeny and Molecular Evolution of the Green Algae. *Critical Reviews in Plant Sciences*, *31*(1), 1–46.
- Leliaert, F., Tronholm, A., Lemieux, C., Turmel, M., Depriest, M. S., Bhattacharya, D., ... Lopez-Bautista, J. M. (2016). Chloroplast phylogenomic analyses reveal the deepest-branching lineage of the Chlorophyta, Palmophyllophyceae class. nov. *Scientific Reports*, *6*.
- Leliaert, F., Verbruggen, H., & Zechman, F. W. (2011). Into the deep: New discoveries at the base of the green plant phylogeny. *BioEssays*, *33*(9), 683–692.
- Lenton, T. M. (2001). The role of land plants, phosphorus weathering and fire in the rise and regulation of atmospheric oxygen. *Global Change Biology*, *7*(6), 613–629.
- Lenton, T. M., Dahl, T. W., Daines, S. J., Mills, B. J. W., Ozaki, K., Saltzman, M. R., & Porada, P. (2016). Earliest land plants created modern levels of atmospheric oxygen. *Proceedings of the National Academy of Sciences*, *113*(35), 9704–9709.
- Lesser, M. P. (2004). Discovery of Symbiotic Nitrogen-Fixing Cyanobacteria in Corals. *Science*, *305*(5686), 997–1000.
- Letsch, M. R., Friedl, T., & Lewis, L. A. (2009). *Elliptochloris marina* sp. nov. (trebouxiophyceae, chlorophyta), symbiotic green alga of the temperate pacific sea anemones *anthopleura xanthogrammica* and *a. elegantissima* (anthozoa, cnidaria) *1*, *45*, 1127–1135.
- Leyland, B., Leu, S., & Boussiba, S. (2017). Are Thraustochytrids algae? *Fungal Biology*, *121*(10), 835–840.
- Lhee, D., Yang, E. C., Kim, J. I., Nakayama, T., Zuccarello, G., Andersen, R. A., & Yoon, H. S. (2017). Diversity of the Photosynthetic Paulinella Species, with the Description of *Paulinella micropora* sp. nov. and the Chromatophore Genome Sequence for strain KR01. *Protist*, *168*(2), 155–170.
- Li, B., Lopes, J. S. J. S., Foster, P. G., Embley, T. M., & Cox, C. J. (2014). Compositional biases among synonymous substitutions cause conflict between gene and protein trees for plastid origins. *Molecular Biology and Evolution*, *31*(7), 1697–1709.
- Li, J. Y., & Wu, C. F. (2005). New symbiotic hypothesis on the origin of eukaryotic flagella. *Naturwissenschaften*, *92*(7), 305–309.
- Lida, 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*(1), 227–238.

- Lin, S. (2011). Genomic understanding of dinoflagellates. *Research in Microbiology*, 162(6), 551–569.
- López-García, P., & Moreira, D. (2015). Open Questions on the Origin of Eukaryotes. *Trends in Ecology and Evolution*. 30(11), 697-708.
- Ludwig, M., & Gibbs, S. P. (1989). Evidence that the nucleomorph of Chlorarachnion reptans (Chlorarachniophyceae) are vestigial nuclei: morphology, division and DNA-DAPI fluorescence. *Journal of Phycology*, 25(2), 385–394.
- Mackiewicz, P., & Gagat, P. (2014). Monophyly of Archaeplastida supergroup and relationships among its lineages in the light of phylogenetic and phylogenomic studies. Are we close to a consensus? *Acta Societatis Botanicorum Poloniae*. 83(4), 263-280.
- Maderspacher, F. (2008). Theodor Boveri and the natural experiment. *Current Biology*. 18(7), 279-286.
- Makai, S., Li, X., Hussain, J., Cui, C., Wang, Y., Chen, M., ... He, G. (2015). A census of nuclear cyanobacterial recruits in the plant kingdom. *PLoS ONE*, 10(3).
- Mareš, J., Hrouzek, P., Kaňa, R., Ventura, S., Strunecký, O., & Komárek, J. (2013). The Primitive Thylakoid-Less Cyanobacterium *Gloeobacter* Is a Common Rock-Dwelling Organism. *PLoS ONE*, 8(6).
- Margulis, L. (2010). Symbiogenesis. A new principle of evolution rediscovery of Boris Mikhaylovich Kozo-Polyansky (1890–1957). *Paleontological Journal*, 44(12), 1525–1539.
- Margulis, L., Chapman, M., Guerrero, R., & Hall, J. (2006). The last eukaryotic common ancestor (LECA): acquisition of cytoskeletal motility from aerotolerant spirochetes in the Proterozoic Eon. *Proceedings of the National Academy of Sciences of the United States of America*, 103(35), 13080–13085.
- Margulis, & Bermudes. (1985). Symbiosis as a mechanism of evolution: status of cell symbiosis theory. *Symbiosis (Philadelphia, Pa.)*, 1, 101–24.
- Marin, B., Nowack, E. C. M., & Melkonian, M. (2005). A plastid in the making: Evidence for a second primary endosymbiosis. *Protist*, 156(4), 425–432.
- Markunas, C. M., & Triemer, R. E. (2016). Evolutionary History of the Enzymes Involved in the Calvin-Benson Cycle in Euglenids. *Journal of Eukaryotic Microbiology*, 63(3), 326–339.
- Martin, W. F. (2017). Too Much Eukaryote LGT. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*. 39(12)

- Martin, W., & Kowallik, K. (1999). Annotated English translation of Mereschkowsky's 1905 paper "Über Natur und Ursprung der Chromatophoren im Pflanzenreiche." *European Journal of Phycology*, 34(3), 287–295.
- Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., ... Penny, D. (2002). Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proceedings of the National Academy of Sciences*, 99(19), 12246–12251.
- Matsumoto, T., Shinozaki, F., Chikuni, T., Yabuki, A., Takishita, K., Kawachi, M., ... Inagaki, Y. (2011). Green-colored Plastids in the Dinoflagellate Genus *Lepidodinium* are of Core Chlorophyte Origin. *Protist*, 162(2), 268–276.
- McFadden, G. I. (2014). Origin and evolution of plastids and photosynthesis in eukaryotes. *Cold Spring Harbor Perspectives in Biology*, 6(4).
- McFadden, G. I., & Van Dooren, G. G. (2004). Evolution: Red algal genome affirms a common origin of all plastids. *Current Biology*, 14(13), 514–516.
- McFadden, Reith, M., Munholland, J., & Lang-Unnasch, N. (1996, June). Plastid in human parasites. *Nature*, 381(6582), 482.
- McKie-Krisberg, Z. M., & Sanders, R. W. (2014). Phagotrophy by the picoeukaryotic green alga *Micromonas*: Implications for Arctic Oceans. *ISME Journal*, 8(10), 1953–1961.
- Meng, F., Zhou, C., Yin, L., Chen, Z., & Yuan, X. (2005). The oldest known dinoflagellates: Morphological and molecular evidence from Mesoproterozoic rocks at Yongji, Shanxi Province. *Chinese Science Bulletin*, 50(12), 1230–1234.
- Menke, W. (1960). Das allgemeine Bauprinzip des Lamellarsystems der Chloroplasten. *Experientia*, 16(12), 537–538.
- Mereschkowsky, C. (1905). Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. *Biologisches Zentralblatt*, 25(July), 593–604.
- Mereschkowsky, C. (1910). Theorie der zwei Plasmaarten als Grundlage der Symbiogenese, einer neuen Lehre von der Entstehung der Organismen. In *Biologisches Centralblatt*. (Vol. 30, pp. 278–288–303–347–367).
- Mereschkowsky, K. (1920). La plante considérée comme un complexe symbiotique. *Bulletin de La Société Des Science Naturelles de l'Ouest de La France*, 6, 17–98.
- Miyagishima, S. -y. (2011). Mechanism of Plastid Division: From a Bacterium to an Organelle. *Plant Physiology*, 155(4), 1533–1544.
- Moczyłowska, M. (2011). The early Cambrian phytoplankton radiation: Acritarch evidence from the Lukati Formation, Estonia. *Palynology*, 35(1), 103–145.
- Moeller, H. V., & Johnson, M. D. (2017). Preferential Plastid Retention by the Acquired Phototroph *Mesodinium chamaeleon*. *Journal of Eukaryotic Microbiology*.

- Moore, R. B., Oborník, M., Janouškovec, J., Chrudimský, T., Vancová, M., Green, D. H., ... Carter, D. A. (2008). A photosynthetic alveolate closely related to apicomplexan parasites. *Nature*, 452(7189), 900–900.
- Moran, J., McKean, P. G., & Ginger, M. L. (2014). Eukaryotic flagella: Variations in form, function, and composition during evolution. *BioScience*, 64(12), 1103–1114.
- Morden, C. W., & Golden, S. S. (1989). psbA genes indicate common ancestry of prochlorophytes and chloroplasts. *Nature*, 337, 382–385.
- Moustafa, A., Beszteri, B., Maier, U. G., Bowler, C., Valentin, K., & Bhattacharya, D. (2009). Genomic Footprints of a Cryptic Plastid Endosymbiosis in Diatoms. *Science*, 324(5935), 1724–1726.
- Müller, K. M., Oliveira, M. C., Sheath, R. G., & Bhattacharya, D. (2001). Ribosomal DNA phylogeny of the Bangiophycidae (Rhodophyta) and the origin of secondary plastids. *American Journal of Botany*, 88(8), 1390–1400.
- Muñoz-Gómez, S. A., Mejía-Franco, F. G., Durnin, K., Colp, M., Grisdale, C. J., Archibald, J. M., & Slamovits, C. H. (2017). The New Red Algal Subphylum Proteorhodophytina Comprises the Largest and Most Divergent Plastid Genomes Known. *Current Biology*, 27(11).
- Myung, G., Kim, H. S., Park, J. S., Park, M. G., & Yih, W. (2011). Population growth and plastid type of *Myrionecta rubra* depend on the kinds of available cryptomonad prey. *Harmful Algae*, 10(5), 536–541.
- Nass, M. M., & Nass, S. (1963). Intramitochondrial fibers with DNA characteristics. I. Fixation and electron staining reactions. *The Journal of Cell Biology*, 19, 593–611.
- Nelissen, B., Van de Peer, Y., Wilmotte, A., & De Wachter, R. (1995). An early origin of plastids within the cyanobacterial divergence is suggested by evolutionary trees based on complete 16S rRNA sequences. *Molecular Biology and Evolution*, 12(6), 1166–1173.
- Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274.
- Nishitani, G., Nagai, S., Baba, K., Kiyokawa, S., Kosaka, Y., Miyamura, K., ... Kamiyama, T. (2010). High-level congruence of *Myrionecta rubra* prey and *Dinophysis* species plastid identities as revealed by genetic analyses of isolates from Japanese coastal waters. *Applied and Environmental Microbiology*, 76(9), 2791–2798.
- Nowack, E. C. M., & Grossman, A. R. (2012). Trafficking of protein into the recently established photosynthetic organelles of *Paulinella chromatophora*. *Proceedings of the National Academy of Sciences of the United States of America*, 109(14), 5340–5.

- Nowack, E. C. M., & Melkonian, M. (2010). Endosymbiotic associations within protists. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 365(1541), 699–712.
- Nowack, E. C. M., Melkonian, M., & Glöckner, G. (2008). Chromatophore Genome Sequence of *Paulinella* Sheds Light on Acquisition of Photosynthesis by Eukaryotes. *Current Biology*, 18(6), 410–418.
- Nowack, E. C. M., Price, D. C., Bhattacharya, D., Singer, A., Melkonian, M., & Grossman, A. R. (2016). Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of *Paulinella chromatophora*. *Proceedings of the National Academy of Sciences*, 113(43), 12214–12219.
- Nowack, E. C. M., Vogel, H., Groth, M., Grossman, A. R., Melkonian, M., & Glöckner, G. (2011). Endosymbiotic gene transfer and transcriptional regulation of transferred genes in *Paulinella chromatophora*. *Molecular Biology and Evolution*, 28(1), 407–422.
- O'Malley, M. A. (2015). Endosymbiosis and its implications for evolutionary theory, *Proceedings of the National Academy of Sciences*. 112(33). 10270-10277.
- Oborník, M., & Green, B. R. (2005). Mosaic origin of the heme biosynthesis pathway in photosynthetic eukaryotes. *Molecular Biology and Evolution*, 22(12), 2343–2353.
- Ochoa de Alda, J. A. G., Esteban, R., Diago, M. L., & Houmard, J. (2014). The plastid ancestor originated among one of the major cyanobacterial lineages. *Nature Communications*, 5, 4937.
- Ohbayashi, R., Watanabe, S., Ehira, S., Kanesaki, Y., Chibazakura, T., & Yoshikawa, H. (2016). Diversification of DnaA dependency for DNA replication in cyanobacterial evolution. *ISME Journal*, 10(5), 1113–1121.
- Okamoto, N., Chantangsi, C., Horák, A., Leander, B. S., & Keeling, P. J. (2009). Molecular phylogeny and description of the novel Katablepharid *Roombia truncata* gen. et sp. nov., and establishment of the Hacrobia taxon nov. *PLoS ONE*, 4(9).
- Okamoto, N., & Inouye, I. (2006). *Hatena arenicola* gen. et sp. nov., a Katablepharid Undergoing Probable Plastid Acquisition. *Protist*, 157(4), 401–419.
- Oliveira, M. C., & Bhattacharya, D. (2000). Phylogeny of the Bangiophycidae (Rhodophyta) and the secondary endosymbiotic origin of algal plastids. *American Journal of Botany*, 87(4), 482–492.
- Oulhen, N., Schulz, B. J., & Carrier, T. J. (2016). English translation of Heinrich Anton de Bary's 1878 speech, "Die Erscheinung der Symbiose" ("De la symbiose"). *Symbiosis*, 69, 131–139.
- Paila, Y. D., Richardson, L. G. L., & Schnell, D. J. (2015). New insights into the mechanism of chloroplast protein import and its integration with protein quality control, organelle biogenesis and development. *Journal of Molecular Biology*. 427(5), 1038-1060.

- Palmer, J. D. (1985). Comparative Organization of Chloroplast Genomes. *Annual Review of Genetics*, 19(1), 325–354.
- Parfrey, L. W., Lahr, D. J. G., Knoll, A. H., & Katz, L. A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences*, 108(33), 13624–13629.
- Patron, N. J., Rogers, M. B., & Keeling, P. J. (2006a). Comparative rates of evolution in endosymbiotic nuclear genomes. *BMC Evolutionary Biology*, 6, 46.
- Patron, N. J., Waller, R. F., & Keeling, P. J. (2006b). A tertiary plastid uses genes from two endosymbionts. *Journal of Molecular Biology*, 357(5), 1373–1382.
- Pawlowski, J., & Burki, F. (2009). Untangling the phylogeny of amoeboid protists. In *Journal of Eukaryotic Microbiology* (Vol. 56, pp. 16–25).
- Perru, O. (2006). Aux origines des recherches sur la symbiose vers 1868-1883. *Revue D'histoire Des Sciences*, 59, 5–27.
- Petersen, J., Ludewig, A. K., Michael, V., Bunk, B., Jarek, M., Baurain, D., & Brinkmann, H. (2014). Chromera velia, endosymbioses and the rhodoplex hypothesis - Plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages). *Genome Biology and Evolution*, 6(3), 666–684.
- Pfanzagl, B., Zenker, A., Pittenauer, E., Allmaier, G., Martinez-Torrecuadrada, J., Schmid, E. R., ... Löffelhardt, W. (1996). Primary structure of cyanelle peptidoglycan of *Cyanophora paradoxa*: A prokaryotic cell wall as part of an organelle envelope. *Journal of Bacteriology*, 178(2), 332–339.
- Pierce, S. K., Curtis, N. E., & Schwartz, J. A. (2009). Chlorophyll a synthesis by an animal using transferred algal nuclear genes. *Symbiosis*, 49(3), 121–131.
- Pierce, S. K., Fang, X., Schwartz, J. A., Jiang, X., Zhao, W., Curtis, N. E., ... Wang, J. (2012). Transcriptomic evidence for the expression of horizontally transferred algal nuclear genes in the photosynthetic sea slug, *Elysia chlorotica*. *Molecular Biology and Evolution*, 29(6), 1545–1556.
- Portier, P. (1918). *Les symbiotes* (Mason). Paris.
- Price, D. C., Chan, C. X., Yoon, H. S., Yang, E. C., Qiu, H., Weber, A. P. M., ... Bhattacharya, D. (2012). *Cyanophora paradoxa* Genome Elucidates Origin of Photosynthesis in Algae and Plants. *Science*, 335(6070), 843–847.
- Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE*, 5(3).
- Provorov, N. A. (2016). K.S. Mereschkowsky and the origin of the eukaryotic cell: 111 years of symbiogenesis theory. *Agricultural Biology*, 51(5), 746–758.
- Pyke, K. A. (2010). Plastid division. *AoB Plants*, 2010, plq016.

- Qiu, D., Huang, L., & Lin, S. (2016). Cryptophyte farming by symbiotic ciliate host detected in situ. *Proceedings of the National Academy of Sciences*, *113*(43), 12208–12213.
- Qiu, H., Price, D. C., Weber, A. P. M., Facchinelli, F., Yoon, H. S., & Bhattacharya, D. (2013). Assessing the bacterial contribution to the plastid proteome. *Trends in Plant Science*, *18*(12), 680–687.
- Qiu, H., Price, D. C., Yang, E. C., Yoon, H. S., & Bhattacharya, D. (2015). Evidence of ancient genome reduction in red algae (Rhodophyta). *Journal of Phycology*, *51*(4), 624–636.
- Raff, R. A., & Mahler, H. R. (1972). The non symbiotic origin of mitochondria. *Science (New York, N.Y.)*, *177*(4049), 575–582.
- Ragon, M., Benzerara, K., Moreira, D., Tavera, R., & López-García, P. (2014). 16S rDNA-based analysis reveals cosmopolitan occurrence but limited diversity of two cyanobacterial lineages with contrasted patterns of intracellular carbonate mineralization. *Frontiers in Microbiology*, *5*(JULY), 1–11.
- Reyes-Prieto, A., & Bhattacharya, D. (2007). Phylogeny of nuclear-encoded plastid-targeted proteins supports an early divergence of glaucophytes within Plantae. *Molecular Biology and Evolution*, *24*(11), 2358–2361.
- Reyes-Prieto, A., Hackett, J. D., Soares, M. B., Bonaldo, M. F., & Bhattacharya, D. (2006). Cyanobacterial Contribution to Algal Nuclear Genomes Is Primarily Limited to Plastid Functions. *Current Biology*, *16*(23), 2320–2325.
- Reyes-Prieto, A., Moustafa, A., & Bhattacharya, D. (2008). Multiple Genes of Apparent Algal Origin Suggest Ciliates May Once Have Been Photosynthetic. *Current Biology*, *18*(13), 956–962.
- Rice, D. W., & Palmer, J. D. (2006). An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. *BMC Biology*, *4*, 31.
- Ris, H., & Plaut, W. (1962). Ultrastructure of DNA-containing areas in the chloroplast of *Chlamydomonas*. *The Journal of Cell Biology*, *13*, 383–391.
- Rodríguez-Ezpeleta, N., Brinkmann, H., Burey, S. C., Roure, B., Burger, G., Löffelhardt, W., ... Lang, B. F. (2005). Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes. *Current Biology*, *15*(14), 1325–1330.
- Rogers, M. B., Gilson, P. R., Su, V., McFadden, G. I., & Keeling, P. J. (2007). The complete chloroplast genome of the chlorarachniophyte *Bigeloviella natans*: Evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. *Molecular Biology and Evolution*, *24*(1), 54–62.
- Rolfe, S. A., Strelkov, S. E., Links, M. G., Clarke, W. E., Robinson, S. J., Djavaheri, M., ... Borhan, M. H. (2016). The compact genome of the plant pathogen *Plasmodiophora*

brassicae is adapted to intracellular interactions with host Brassica spp. *BMC Genomics*, 17(1), 272.

- Rota-Stabelli, O., Daley, A. C., & Pisani, D. (2013). Molecular timetrees reveal a cambrian colonization of land and a new scenario for ecdysozoan evolution. *Current Biology*, 23(5), 392–398.
- Rumpho, M. E., Pochareddy, S., Worful, J. M., Summer, E. J., Bhattacharya, D., Pelletreau, K. N., ... Soule, K. M. (2009). Molecular Characterization of the Calvin Cycle Enzyme Phosphoribulokinase in the Stramenopile Alga *Vaucheria litorea* and the Plastid Hosting Mollusc *Elysia chlorotica*. *Molecular Plant*, 2(6), 1384–1396.
- Rumpho, M. E., Worful, J. M., Lee, J., Kannan, K., Tyler, M. S., Bhattacharya, D., ... Manhart, J. R. (2008). Horizontal gene transfer of the algal nuclear gene psbO to the photosynthetic sea slug *Elysia chlorotica*. *Proceedings of the National Academy of Sciences*, 105(46), 17867–17871.
- Sagan, L. (1967). On the origin of mitosing cells. *Journal of Theoretical Biology*, 14(3), 255–274.
- Saghai, A., Zivanovic, Y., Zeyen, N., Moreira, D., Benzerara, K., Deschamps, P., ... López-García, P. (2015). Metagenome-based diversity analyses suggest a significant contribution of non-cyanobacterial lineages to carbonate precipitation in modern microbialites. *Frontiers in Microbiology*, 6(AUG).
- Sakharkar, K. R., Kumar Dhar, P., & Chow, V. V. T. K. (2004). Genome reduction in prokaryotic obligatory intracellular parasites of humans: A comparative analysis. *International Journal of Systematic and Evolutionary Microbiology*, 54(6), 1937–1941.
- Sánchez-Baracaldo, P. (2015). Origin of marine planktonic cyanobacteria. *Scientific Reports*, 5, 17418.
- Sánchez-Baracaldo, P., Raven, J. A., Pisani, D., & Knoll, A. H. (2017). Early photosynthetic eukaryotes inhabited low-salinity habitats. *Proceedings of the National Academy of Sciences*, 114(37), 7737–7745.
- Sanchez-Puerta, M. V., Bachvaroff, T. R., & Delwiche, C. F. (2007). Sorting wheat from chaff in multi-gene analyses of chlorophyll c-containing plastids. *Molecular Phylogenetics and Evolution*, 44(2), 885–897.
- Sapp, J. (1994). *Evolution by Association: A History of Symbiosis* (Oxford University). New York.
- Sapp, J. (2009). *The New Foundations of Evolution: On the Tree of Life*. New York: OXFORD UNIVERSITY PRESS.
- Sapp, J., Carrapiço, F., & Zolotonosov, M. (2002). Symbiogenesis: the hidden face of Constantin Merezhkowsky. *History and Philosophy of the Life Sciences*, 24(3–4), 413–440.

- Sastre, V., Santinelli, N., & Akselman, R. (2010). A phylogenetic mosaic plastid proteome and unusual plastid-targeting signals in the green-colored dinoflagellate *Lepidodinium chlorophorum*. *Harmful Algae News*, 42, 16–17.
- Schatz, G., Haslbrunner, E., & Tuppy, H. (1964). Deoxyribonucleic acid associated with yeast mitochondria. *Biochemical and Biophysical Research Communications*, 15(2), 127–132.
- Schelkunov, M. I., Shtratnikova, V. Y., Nuraliev, M. S., Selosse, M., Penin, A., & Logacheva, M. D. (2017). Exploring the Limits for Reduction of Plastid Genomes: A Case Study of the Mycoheterotrophic Orchids *Epipogium aphyllum* and *Epipogium roseum*. *Genome Biology and Evolution*, 7(September), 1179–1191.
- Schimper, A. (1883). Über die Entwicklung der Chlorophylkörner und Farbkörner. *Botanische Zeitung*, 41, 105–114.
- Schirromeister, B. E., Antonelli, A., & Bagheri, H. C. (2011). The origin of multicellularity in cyanobacteria. *BMC Evolutionary Biology*, 11(1), 45.
- Schirromeister, B. E., Gugger, M., & Donoghue, P. C. J. (2015). Cyanobacteria and the Great Oxidation Event: evidence from genes and fossils. *Palaeontology*, 58(5), 769–785.
- Schirromeister, B. E., Sanchez-Baracaldo, P., & Wacey, D. (2016). Cyanobacterial evolution during the Precambrian. *International Journal of Astrobiology*, 15(3), 187–204.
- Schwartz, R. M., & Dayhoff, M. O. (1978). Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. *Science (New York, N.Y.)*, 199(4327), 395–403.
- Ševčíková, T., Horák, A., Klimeš, V., Zbránková, V., Demir-Hilton, E., Sudek, S., ... Eliáš, M. (2015). Updating algal evolutionary relationships through plastid genome sequencing: did alveolate plastids emerge through endosymbiosis of an ochrophyte? *Scientific Reports*, 5(1), 10134.
- Shalchian-Tabrizi, K., Skånseng, M., Ronquist, F., Klaveness, D., Bachvaroff, T. R., Delwiche, C. F., ... Jakobsen, K. S. (2006). Heterotachy processes in rhodophyte-derived secondhand plastid genes: Implications for addressing the origin and evolution of dinoflagellate plastids. *Molecular Biology and Evolution*, 23(8), 1504–1515.
- Sheiner, L., & Striepen, B. (2013). Protein sorting in complex plastids. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1833(2), 352–359.
- Shi, L., & Theg, S. (2013). The chloroplast protein import system: From algae to trees. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1833(2), 314–331.
- Shi, T., & Falkowski, P. G. (2008). Genome evolution in cyanobacteria: The stable core and the variable shell. *Proceedings of the National Academy of Sciences*, 105(7), 2510–2515.

- Shih, P. M., Hemp, J., Ward, L. M., Matzke, N. J., & Fischer, W. W. (2017). Crown group Oxyphotobacteria postdate the rise of oxygen. *Geobiology*, *15*(1), 19–29.
- Shih, P. M., Wu, D., Latifi, A., Axen, S. D., Fewer, D. P., Talla, E., ... Kerfeld, C. A. (2013). Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy of Sciences*, *110*(3), 1053–1058.
- Sierra, R., Canas-Duarte, S. J., Burki, F., Schwelm, A., Fogelqvist, J., Dixelius, C., ... Pawlowski, J. (2016). Evolutionary origins of rhizarian parasites. *Molecular Biology and Evolution*, *33*(4), 980–983.
- Singer, A., Poschmann, G., Mühlich, C., Valadez-Cano, C., Hänsch, S., Hüren, V., ... Nowack, E. C. M. (2017). Massive Protein Import into the Early-Evolutionary-Stage Photosynthetic Organelle of the Amoeba *Paulinella chromatophora*. *Current Biology*, *27*(18), 2763–2773.
- Six, C., Thomas, J.-C., Garczarek, L., Ostrowski, M., Dufresne, A., Blot, N., ... Partensky, F. (2007). Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. *Genome Biology*, *8*(12), R259.
- Smith, D. R., & Lee, R. W. (2010). Low Nucleotide Diversity for the Expanded Organelle and Nuclear Genomes of *Volvox carteri* Supports the Mutational-Hazard Hypothesis, *27*(10), 2244–2256.
- Smith, D. R., & Lee, R. W. (2014). A Plastid without a Genome: Evidence from the Nonphotosynthetic Green Algal Genus *Polytomella*. *Plant Physiology*, *164*(4), 1812–1819.
- Sojo, V., Herschy, B., Whicher, A., Camprubí, E., & Lane, N. (2016). The Origin of Life in Alkaline Hydrothermal Vents. *Astrobiology*, *16*(2), 181–197.
- Sommer, M. S., Gould, S. B., Lehmann, P., Gruber, A., Przyborski, J. M., & Maier, U. G. (2007). Der1-mediated preprotein import into the periplastid compartment of chromalveolates? *Molecular Biology and Evolution*, *24*(4), 918–928.
- Soo, R. M., Hemp, J., Parks, D. H., Fischer, W. W., & Hugenholtz, P. (2017). On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. *Science*, *355*(6332), 1436–1440.
- Spribille, T., Tuovinen, V., Resl, P., Vanderpool, D., Wolinski, H., Aime, M. C., ... McCutcheon, J. P. (2016). Basidiomycete yeasts in the cortex of ascomycete macrolichens. *Science*, *353*(6298), 488–492.
- Stegemann, S., & Bock, R. (2006). Experimental Reconstruction of Functional Gene Transfer from the Tobacco Plastid Genome to the Nucleus. *The Plant Cell Online*, *18*(11), 2869–2878.

- Steindler, L., Huchon, D., Avni, A., & Ilan, M. (2005). 16S rRNA phylogeny of sponge-associated cyanobacteria. *Applied and Environmental Microbiology*, *71*(7), 4127–4131.
- Stiller, J. W., Schreiber, J., Yue, J., Guo, H., Ding, Q., & Huang, J. (2014). The evolution of photosynthesis in chromist algae through serial endosymbioses. *Nature Communications*, *5*, 5764.
- Stirewalt, V. L., Michalowski, C. B., Löffelhardt, W., Bohnert, H. J., & Bryant, D. A. (1995). Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. *Plant Molecular Biology Reporter*, *13*(4), 327–332.
- Stoecker, D. K., Johnson, M. D., De Vargas, C., & Not, F. (2009). Acquired phototrophy in aquatic protists. *Aquatic Microbial Ecology*, *57*(3), 279–310.
- Stork, S., Lau, J., Moog, D., & Maier, U. G. (2013). Three old and one new: Protein import into red algal-derived plastids surrounded by four membranes. *Protoplasma*, *250*(5), 1013–1023.
- Stork, S., Moog, D., Przyborski, J. M., Wilhelmi, I., Zauner, S., & Maier, U. G. (2012). Distribution of the SELMA translocon in secondary plastids of red algal origin and predicted uncoupling of ubiquitin-dependent translocation from degradation. *Eukaryotic Cell*, *11*(12), 1472–1481.
- Summerer, M., Sonntag, B., & Sommaruga, R. (2008). Ciliate-symbiont specificity of freshwater endosymbiotic chlorella (trebouxiophyceae, chlorophyta) 1. *Journal of Phycology*, *44*, 77–84.
- Suzuki, K., & Miyagishima, S. Y. (2010). Eukaryotic and Eubacterial Contributions to the Establishment of Plastid Proteome Estimated by Large-Scale Phylogenetic Analyses. *Molecular Biology and Evolution*, *27*(3), 581–590.
- Suzuki, S., Ishida, K. I., & Hirakawa, Y. (2016a). Diurnal transcriptional regulation of endosymbiotically derived genes in the chlorarachniophyte *Bigeloviella natans*. *Genome Biology and Evolution*, *8*(9), 2672–2682.
- Suzuki, Hirakawa, Y., Kofuji, R., Sugita, M., & Ishida, K. (2016b). Plastid genome sequences of *Gymnochlora stellata*, *Lotharella vacuolata*, and *Partenskyella glossopodia* reveal remarkable structural conservation among chlorarachniophyte species. *Journal of Plant Research*, *129*(4), 581–590.
- Suzuki, Shirato, S., Hirakawa, Y., & Ishida, K. I. (2015). Nucleomorph genome sequences of two chlorarachniophytes, *Amorphochlora amoebiformis* and *Lotharella vacuolata*. *Genome Biology and Evolution*, *7*(6), 1533–1545.
- Tanifuji, G., Onodera, N. T., Brown, M. W., Curtis, B. A., Roger, A. J., Ka-Shu Wong, G., ... Archibald, J. M. (2014). Nucleomorph and plastid genome sequences of the

chlorarachniophyte *Lotharella oceanica*: convergent reductive evolution and frequent recombination in nucleomorph-bearing algae. *BMC Genomics*, 15(1), 374.

- Timmis, J. N., Ayliffe, M. A., Huang, C. Y., & Martin, W. (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics*, 5(2), 123–135.
- Toplin, J. A., Norris, T. B., Lehr, C. R., McDermott, T. R., & Castenholz, R. W. (2008). Biogeographic and phylogenetic diversity of thermoacidophilic cyanidiales in Yellowstone National Park, Japan, and New Zealand. *Applied and Environmental Microbiology*, 74(9), 2822–2833.
- Turner, S., Pryer, K. M., Miao, V. P., & Palmer, J. D. (1999). Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *The Journal of Eukaryotic Microbiology*, 46(4), 327–338.
- Tyler, B. M. (2006). Phytophthora Genome Sequences Uncover Evolutionary Origins and Mechanisms of Pathogenesis. *Science*, 313(5791), 1261–1266.
- Tyra, H. M., Linka, M., Weber, A. P., & Bhattacharya, D. (2007). Host origin of plastid solute transporters in the first photosynthetic eukaryotes. *Genome Biology*, 8(10), R212.
- Uyeda, J. C., Harmon, L. J., & Blank, C. E. (2016). A comprehensive study of cyanobacterial morphological and ecological evolutionary dynamics through deep geologic time. *PLoS ONE*, 11(9).
- Uzzell, T., & Spolsky, C. (1974). Mitochondria and plastids as endosymbionts: a revival of special creation? *American Scientist*, 62(3), 334–343.
- Vanclová, A. M. G., Hadariová, L., Hrdá, Š., & Hampl, V. (2017). Chapter Nine - Secondary Plastids of Euglenophytes. In Y. B. T.-A. in B. R. Hirakawa (Ed.), *Secondary Endosymbioses* (Vol. 84, pp. 321–358). Academic Press.
- von Heijne, G. (1986). Why mitochondria need a genome. *FEBS Letters*, 198(1), 1–4.
- Wägele, H., Deusch, O., Händeler, K., Martin, R., Schmitt, V., Christa, G., ... Martin, W. (2011). Transcriptomic evidence that longevity of acquired plastids in the photosynthetic slugs *Elysia timida* and *Plakobranthus ocellatus* does not entail lateral transfer of algal nuclear genes. *Molecular Biology and Evolution*, 28(1), 699–706.
- Waller, R. F., Reed, M. B., Cowman, A. F., & McFadden, G. I. (2000). Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway. *The EMBO Journal*, 19(8), 1794–802.
- Waller, R., & Kořený, L. (2017). Plastid Complexity in Dinoflagellates: A Picture of Gains, Losses, Replacements and Revisions. *Advances in Botanical Research*.
- Wallin, I. E. (1927). *Symbiogenesis and the origin of species*. (T. W. A. W. Company, Ed.). Baltimore.

- Wang, Q., Sun, H., & Huang, J. (2017). Re-analyses of “Algal” Genes Suggest a Complex Evolutionary History of Oomycetes. *Frontiers in Plant Science*, 8.
- Ward, D. M., Castenholz, R. W., & Miller, S. R. (2012). Cyanobacteria in geothermal habitats. In *Ecology of Cyanobacteria II: Their Diversity in Space and Time* (Vol. 9789400738, pp. 39–63).
- Weber, A. P. M., Linka, M., & Bhattacharya, D. (2006). Single, ancient origin of a plastid metabolite translocator family in Plantae from an endomembrane-derived ancestor. *Eukaryotic Cell*, 5(3), 609–612.
- Whatley, J. M., & Whatley, F. R. (1981). Chloroplast Evolution. *New Phytologist*, 87(2), 233–247.
- Wilson, E. (1902). *The cell in development and inheritance*. The Macmillan Company.
- Woehle, C., Dagan, T., Martin, W. F., & Gould, S. B. (2011). Red and problematic green phylogenetic signals among thousands of nuclear genes from the photosynthetic and apicomplexa-related *Chromera velia*. *Genome Biology and Evolution*, 3(1), 1220–1230.
- Woo, Y. H., Ansari, H., Otto, T. D., Linger, C. M. K., Olisko, M. K., Michálek, J., ... Pain, A. (2015). Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. *eLife*, 4(JULY 2015), 1–41.
- Yabuki, A., Kamikawa, R., Ishikawa, S. a, Kolisko, M., Kim, E., Tanabe, A. S., ... Inagaki, Y. (2014). *Palpitomonas bilix* represents a basal cryptist lineage: insight into the character evolution in Cryptista. *Scientific Reports*, 4, 4641.
- Yamada, N., Sym, S. D., & Horiguchi, T. (2017). Identification of highly divergent diatom-derived chloroplasts in dinoflagellates, including a description of *Durinskia kwazulunatalensis* sp. nov. (Peridinales, Dinophyceae). *Molecular Biology and Evolution*, 34(6), 1335–1351.
- Yamaguchi, A., Yubuki, N., & Leander, B. S. (2012). Morphostasis in a novel eukaryote illuminates the evolutionary transition from phagotrophy to phototrophy: description of *Rapaza viridis* n. gen. et sp. (Euglenozoa, Euglenida). *BMC Evolutionary Biology*, 12(1), 29.
- Yamaguchi, H., Nakayama, T., Hongoh, Y., Kawachi, M., & Inouye, I. (2014). Molecular diversity of endosymbiotic *Nephroselmis* (Nephroselmidophyceae) in *Hatena arenicola* (Katablepharidophycota). *Journal of Plant Research*, 127(2), 241–247.
- Yang, E. C., Boo, S. M., Bhattacharya, D., Saunders, G. W., Knoll, A. H., Fredericq, S., ... Yoon, H. S. (2016). Divergence time estimates and the evolution of major lineages in the florideophyte red algae. *Scientific Reports*, 6(1), 21361.
- Yang, Y., Maruyama, S., Sekimoto, H., Sakayama, H., & Nozaki, H. (2011). An extended phylogenetic analysis reveals ancient origin of “non-green” phosphoribulokinase

genes from two lineages of “green” secondary photosynthetic eukaryotes: Euglenophyta and Chlorarachniophyta. *BMC Research Notes*, 4, 330.

- Yang, Y., Matsuzaki, M., Takahashi, F., Qu, L., & Nozaki, H. (2014). Phylogenomic analysis of “red” genes from two divergent species of the “green” secondary phototrophs, the chlorarachniophytes, suggests multiple horizontal gene transfers from the red lineage before the divergence of extant chlorarachniophytes. *PLoS ONE*, 9(6).
- Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G., & Bhattacharya, D. (2004). A Molecular Timeline for the Origin of Photosynthetic Eukaryotes. *Molecular Biology and Evolution*, 21(5), 809–818.
- Yoon, H. S., Hackett, J. D., Pinto, G., & Bhattacharya, D. (2002). The single, ancient origin of chromist plastids. *Proceedings of the National Academy of Sciences of the United States of America*, 99(24), 15507–15512.
- Yue, J., Hu, X., Sun, H., Yang, Y., & Huang, J. (2012). Widespread impact of horizontal gene transfer on plant colonization of land. *Nature Communications*, 3, 1152.
- Záhonová, K., Füssy, Z., Oborník, M., Eliáš, M., & Yurchenko, V. (2016). RuBisCO in non-photosynthetic alga *Euglena longa*: Divergent features, transcriptomic analysis and regulation of complex formation. *PLoS ONE*, 11(7).
- Zapata, M., Fraga, S., Rodríguez, F., & Garrido, J. L. (2012). Pigment-based chloroplast types in dinoflagellates. *Marine Ecology Progress Series*, 465, 33–52.
- Zhang, Z., Green, B. R., & Cavalier-Smith, T. (2000). Phylogeny of Ultra-Rapidly Evolving Dinoflagellate Chloroplast Genes: A Possible Common Origin for Sporozoon and Dinoflagellate Plastids. *Journal of Molecular Evolution*, 51(1), 26–40.
- Zhaxybayeva, O., Gogarten, J. P., Charlebois, R. L., Doolittle, W. F., & Papke, R. T. (2006). Phylogenetic analyses of cyanobacterial genomes: Quantification of horizontal gene transfer events. *Genome Research*, 16(9), 1099–1108.
- Zhong, B., Sun, L., & Penny, D. (2015). The Origin of Land Plants: A Phylogenomic Perspective. *Evolutionary Bioinformatics Online*, 11, 137–41.
- Zuckerlandl, E., & Pauling, L. (1965). Molecules as documents of evolutionary history. *Journal of Theoretical Biology*, 8(2), 357–366.
- Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q., & van Wijk, K. J. (2008). Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS ONE*, 3(4).

Titre : Origine et évolution des eucaryotes photosynthétiques

Mots clés : Photosynthèse, endosymbiose, plastes, cyanobactéries, eucaryotes

Résumé : Les plastes primaires proviennent d'une cyanobactérie qui a établi une relation endosymbiotique avec un hôte eucaryote. Cet événement a donné naissance au super-groupe Archaeplastida qui inclut les Viridiplantae (algues vertes et plantes terrestres), les Rhodophyta (algues rouges) et les Glaucophyta. Suite à l'endosymbiose primaire, les algues rouges et vertes ont étendu la capacité de photosynthèse à d'autres lignées eucaryotes via des endosymbioses secondaires. Bien que des progrès considérables aient été réalisés dans la compréhension de l'évolution des eucaryotes photosynthétiques, d'importantes questions sont restées ouvertes, telles que l'identité de la lignée cyanobactérienne la plus proche des plastes primaires ainsi que le nombre et l'identité des partenaires dans les endosymbioses secondaires.

Ma thèse a consisté à étudier l'origine et l'évolution précoce des eucaryotes photosynthétiques en utilisant des approches phylogénétiques et phylogénomiques. Je montre par mon travail que les plastes primaires ont évolué à partir d'un symbiote phylogénétiquement proche de *Gloeomargarita lithophora*, une cyanobactérie représentant un clade s'étant diversifié précocement et qui a été détectée uniquement dans les milieux terrestres. Ce résultat fournit des pistes nouvelles sur le contexte écologique dans lequel l'endosymbiose primaire a probablement eu lieu. En ce qui concerne l'évolution des lignées eucaryotes avec des plastes secondaires, je montre que les génomes nucléaires des chlorarachniophytes et des euglénophytes, deux lignées photosynthétiques avec des plastes dérivés d'algues vertes, encodent un grand nombre de gènes acquis par transferts depuis des algues rouges. Enfin, je mets en évidence que SELMA, la machinerie de translocation des protéines à travers la seconde membrane externe des plastes rouges secondaires à quatre membranes, a une histoire étonnamment compliquée aux implications évolutives importantes : les cryptophytes ont recruté un ensemble de composants de SELMA différent de ceux des haptophytes, straménopiles et alvéolés. Ainsi, ma thèse a permis d'identifier pour la première fois la lignée cyanobactérienne la plus proche des plastes primaires et apporte de nouvelles pistes pour éclaircir les événements complexes qui ont jalonné l'évolution des eucaryotes photosynthétiques secondaires.

Title : Origins and early evolution of photosynthetic eukaryotes

Keywords : Photosynthesis, endosymbiosis, plastids, cyanobacteria, eukaryotes

Abstract : Primary plastids derive from a cyanobacterium that entered into an endosymbiotic relationship with a eukaryotic host. This event gave rise to the supergroup Archaeplastida which comprises Viridiplantae (green algae and land plants), Rhodophyta (red algae) and Glaucophyta. After primary endosymbiosis, red and green algae spread the ability to photosynthesize to other eukaryotic lineages via secondary endosymbioses. Although considerable progress has been made in the understanding of the evolution of photosynthetic eukaryotes, important questions remained debated such as the present-day closest cyanobacterial lineage to primary plastids as well as the number and identity of partners in secondary endosymbioses.

The main objectives of my PhD were to study the origin and evolution of plastid-bearing eukaryotes using phylogenetic and phylogenomic approaches to shed some light on how primary and secondary endosymbioses occurred. In this work, I show that primary plastids evolved from a close relative of *Gloeomargarita lithophora*, a recently sequenced early-branching cyanobacterium that has been only detected in terrestrial environments. This result provide interesting hints on the ecological setting where primary endosymbiosis likely took place. Regarding the evolution of eukaryotic lineages with secondary plastids, I show that the nuclear genomes of chlorarachniophytes and euglenids, two photosynthetic lineages with green alga-derived plastids, encode for a large number of genes acquired by transfers from red algae. Finally, I highlight that SELMA, the translocation machinery putatively used to import proteins across the second outermost membrane of secondary red plastids with four membranes, has a surprisingly complex history with strong evolutionary implications: cryptophytes have recruited a set of SELMA components different from those present in haptophytes, stramenopiles and alveolates.

In conclusion, during my PhD I identified for the first time the closest living cyanobacterium to primary plastids and provided new insights on the complex evolution that have undergone secondary plastid-bearing eukaryotes.