



HAL
open science

Horizontal and endosymbiotic gene transfer in early plastid evolution

Rafael Isaac Ponce Toledo, Purificacion Lopez-Garcia, David Moreira

► **To cite this version:**

Rafael Isaac Ponce Toledo, Purificacion Lopez-Garcia, David Moreira. Horizontal and endosymbiotic gene transfer in early plastid evolution. *New Phytologist*, 2019, <10.1111/nph.15965>. <hal-02365884>

HAL Id: hal-02365884

<https://hal.science/hal-02365884v1>

Submitted on 15 Nov 2019

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.



HAL Authorization

1 **Horizontal and endosymbiotic gene transfer in early plastid evolution**

2

3

4 **Rafael I. Ponce-Toledo, Purificación López-García and David Moreira**

5

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

8

9 Author for correspondence:

10 *David Moreira*

11 *Tel: +33 69157608*

12 *Email: david.moreira@u-psud.fr*

13

14 Twitter account: [@DEEMteam_Orsay](https://twitter.com/DEEMteam_Orsay)

15

16

17 **Key words:** plastids, endosymbiosis, gene transfer, evolution, *Paulinella*, Cyanobacteria,
18 Archaeplastida.

19 **Summary**

20 Plastids evolved from a cyanobacterium that was engulfed by a heterotrophic eukaryotic host and
21 became a stable organelle. Some of the resulting eukaryotic algae entered into a number of
22 secondary endosymbioses with diverse eukaryotic hosts. These events had major consequences on
23 the evolution and diversification of life on Earth. Although almost all plastid diversity derives from
24 a single endosymbiotic event, analysis of nuclear genomes of plastid-bearing lineages has revealed
25 a mosaic origin of plastid-related genes. In addition to cyanobacterial genes, plastids recruited for
26 their functioning eukaryotic proteins encoded by the host nucleus and also bacterial proteins of non-
27 cyanobacterial origin. Thus, plastid proteins and plastid-localized metabolic pathways evolved by
28 tinkering using gene toolkits from different sources. This mixed heritage seems especially complex
29 in secondary algae containing green plastids, the acquisition of which appears to have been
30 facilitated by many previous acquisitions of red algal genes (the “red carpet hypothesis”).

31

32 **I. Introduction**

33 Oxygenic photosynthesis in eukaryotes appeared >1 billion years ago (Eme *et al.*, 2014) via the
34 endosymbiosis of a close relative of the deep-branching cyanobacterium *Gloeomargarita lithophora*
35 within a phagotrophic eukaryotic host (Ponce-Toledo *et al.*, 2017). Subsequently, the
36 cyanobacterium evolved into a permanent photosynthetic organelle called primary plastid. This
37 endosymbiosis is at the origin of the Archaeplastida, a monophyletic supergroup composed of three
38 primary plastid-bearing lineages: the green algae plus land plants, the red algae, and the
39 glaucophytes (Fig. 1; Adl *et al.*, 2012). Several cercozoan amoebae of the genus *Paulinella* (*P.*
40 *chromatophora*, *P. longichromatophora* and *P. micropora*) (Fig. 2) constitute the only known
41 lineage apart from Archaeplastida in which an independent type of primary photosynthetic
42 organelles evolved (Marin *et al.*, 2005). The *Paulinella* organelles, called "chromatophores", derive
43 from a cyanobacterium of the *Synechococcus/Prochlorococcus* group (α -cyanobacteria), in contrast
44 with the *Gloeomargarita*-like plastids of Archaeplastida. Whereas primary endosymbioses are
45 extremely rare in biological history, with only the Archaeplastida and *Paulinella* cases known,
46 green and red algae have participated as endosymbionts in numerous secondary and tertiary
47 endosymbioses, originating a broad diversity of eukaryotic lineages with complex plastids (Fig. 1;
48 Moreira & Philippe, 2001; Keeling, 2010).

49 The reduced number of primary plastid acquisitions compared to that of eukaryotic lineages
50 with complex plastids suggests that the enslavement of a photosynthetic cyanobacterial
51 endosymbiont (or cyanobiont) is, for unknown reasons, more challenging from an evolutionary
52 point of view and requires specific adaptations to stabilize the cyanobiont as a permanent organelle.
53 Once the primary plastid was fully integrated within the host (through the evolution of metabolite

54 export and protein import systems, transfer of plastid genes to the host nucleus, and evolution of
55 proteins involved in redox regulation) it was easier for algae to become endosymbionts of other
56 eukaryotes. Here, we review the genetic and genomic changes that accompanied the evolution of
57 primary plastids and explore the plastid proteome composition to propose a possible role for host-
58 derived and exogenous genes in the establishment of these plastids. Likewise, we discuss the
59 genetic mosaicism of nuclear genomes in complex plastid-harboring lineages in the light of
60 horizontal gene transfer (HGT) and putative cryptic endosymbioses.

61

62 II. Evolution of primary plastids in Archaeplastida

63 Plastids in Archaeplastida derive from an endosymbiotic cyanobacterium that was fully integrated
64 in a heterotrophic eukaryotic host and became an organelle. Nonetheless, plastids differ
65 considerably from free-living cyanobacteria. One of the most remarkable differences is the drastic
66 reduction of plastid genomes, which encode less than 5% of the genes found in typical free-living
67 cyanobacteria (Green, 2011). The number of protein-coding genes in plastid genomes varies across
68 Archaeplastida, with red algae harboring the largest sets (160-235 proteins; Lee *et al.*, 2016),
69 perhaps as a result of the massive gene loss in the nuclear genome likely experienced by the
70 rhodophyte ancestor, which exerted a selective pressure to keep plastid-encoded genes
71 (Bhattacharya *et al.*, 2018).

72 Gene loss in the cyanobiont was a crucial event during early endosymbiosis stages and marked
73 the transformation of the cyanobacterium into an obligatory endosymbiont. Glycogen synthase
74 genes, absent in plastids, were possibly among the first genes to be lost and this may have helped to
75 achieve the early host-endosymbiont metabolic integration (Gavelis & Gile, 2018). Glycogen
76 synthase-defective mutant cyanobacteria release organic carbon outside the cell as a homeostatic
77 response to cope with excessive photosynthate production and the inability to store it as glycogen
78 (Cano *et al.*, 2018). Therefore, it is possible that inefficient glycogen storage allowed the leakage of
79 photosynthates that could be used by the host before the evolution of a specific and more efficient
80 photosynthate export system.

81 It is commonly assumed that endosymbiotic gene transfer (EGT) was critical for plastid
82 endosymbiosis because it allowed the host to gain control over the expression of plastid genes, thus
83 increasing the dependence of the cyanobiont upon the host gene expression machinery. The
84 majority of plastid genes transferred to the host nucleus participate in photosynthesis-related
85 functions (e.g., photosystem subunits, chlorophyll synthesis) and plastid metabolism and
86 maintenance (e.g., amino acid synthesis and plastid division) (Reyes-Prieto *et al.*, 2006). Recent
87 studies have suggested that EGTs contributed also to expand the redox sensing capabilities of the
88 host, sometimes through genetic tinkering that produced new chimeric proteins (Méheust *et al.*,

89 2016). This process was accompanied by the evolution of new redox sensing pathways that helped
90 the host to cope with the increased reactive oxygen species concentration produced by the new
91 organelle (Woehle *et al.*, 2017).

92 Even though plastid-encoded proteins plus plastid-targeted EGT products are essential for the
93 correct functioning of these organelles, they represent less than half of the total plastid proteome in
94 Archaeplastida (Qiu *et al.*, 2013). The fact that the plastid proteome contains a vast proportion of
95 apparently non-cyanobacterial proteins opens interesting questions about the mechanisms
96 underlying plastid acquisition and suggest that plastid symbiogenesis was not as straightforward as
97 it is commonly assumed (Larkum *et al.*, 2007). For instance, recent phylogenetic analyses suggest
98 that the genetic machinery necessary to synthesize the galactolipids of plastid membranes -once
99 thought to derive from the cyanobiont- do not have cyanobacterial origin, opening many questions
100 about the nature of plastid membranes (Sato & Awai, 2017).

101 Host-derived proteins, evolved either from the retargeting of pre-existing proteins or from new
102 gene innovations, participate in a wide range of plastid functions and represent the largest fraction
103 of plastid-targeted proteins (Qiu *et al.*, 2013). For instance, metabolite transporters originated from
104 the retargeting of host membrane proteins are particularly overrepresented in the plastid envelope
105 (58% of plastid transporters appear to derive from host membrane proteins in *Arabidopsis thaliana*;
106 Tyra *et al.*, 2007). These findings suggest that the host drove the integration of the cyanobiont by
107 providing the proteins necessary to connect the cyanobiont metabolism with the energy demands of
108 the host (“host-centric” endosymbiosis model; Karkar *et al.*, 2015).

109 In addition to host-derived genes, non-cyanobacterial bacterial genes were also crucial for plastid
110 evolution and represent 7-15% of the plastid proteome (Qiu *et al.*, 2013). The phylogenetic origin of
111 these non-cyanobacterial prokaryotic genes in Archaeplastida seems to cover a wide range of
112 bacterial phyla (Dagan *et al.*, 2013). After Cyanobacteria, Proteobacteria appear to be the most
113 common contributors to the plastid proteome, particularly alphaproteobacterial proteins that may
114 derive from the mitochondrial ancestor (Dagan *et al.*, 2013; Qiu *et al.*, 2013). Part of these non-
115 cyanobacterial bacterial genes were likely present in the genome of the cyanobacterial plastid
116 ancestor as a consequence of HGT events from various donors (Dagan *et al.*, 2013). Some others
117 may have been subsequently acquired by the Archaeplastida ancestor due to its presumed
118 phagotrophic lifestyle and may have helped to compensate the massive gene loss undergone by the
119 cyanobiont genome.

120 An issue that has attracted much attention is the presence of a number of genes apparently
121 transferred from Chlamydiales bacteria to the Archaeplastida (Huang & Gogarten, 2007; Becker *et al.*,
122 2008), which has even led to propose a tripartite model of plastid origin (Facchinelli *et al.*,
123 2013). This model suggests that chlamydial cells infected the Archaeplastida ancestor. This

124 infection would have helped in the early steps of plastid acquisition by i) protecting the cyanobiont
125 from host defenses, ii) supplying multiple enzymes to integrate photosynthates produced by the
126 cyanobiont into the host carbohydrate metabolism (Facchinelli *et al.*, 2013) and/or iii) allowing the
127 cyanobiont to cope with ATP starvation resulting from the hypoxic environment of the host cytosol
128 (Cenci *et al.*, 2018). However, this model encounters several problems. First, there is no report of
129 any *Chlamydia* species able to infect Archaeplastida, which suggests that the Archaeplastida
130 ancestor was not a likely host for this family of pathogenic bacteria. Second, and more importantly,
131 phylogenetic re-analyses of putative *Chlamydia*-derived genes have detected various phylogenetic
132 artefacts and reduced considerably the number of genes compatible with a putative chlamydial
133 ancestry (Moreira & Deschamps, 2014; Domman *et al.*, 2015). Hence, the tripartite model continues
134 to be hotly debated.

135

136 **III. Chromatophore evolution in *Paulinella***

137 Several cercozoan amoebae of the genus *Paulinella* (e.g., *P. chromatophora*, Fig. 2) adopted a
138 primary photosynthetic organelle (called chromatophore) by the endosymbiosis of a
139 cyanobacterium from the *Synechococcus/Prochlorococcus* (Syn/Pro) clade (Marin *et al.*, 2005).
140 *Paulinella* is a good model to study the early evolution of primary plastids since the divergence of
141 the chromatophore from its Syn/Pro ancestor is relatively recent, only 90-140 Mya (Delaye *et al.*,
142 2016). Remarkably, there are important similarities between the primary endosymbioses in
143 Archaeplastida and *Paulinella*, likely due to convergent evolution in the process of plastid
144 acquisition (Table 1).

145 The chromatophore genome is highly reduced, encoding 867 proteins that represent about 1/3 of
146 proteins of its free-living counterparts (Nowack *et al.*, 2008). Similar to the EGTs found in
147 Archaeplastida, *P. chromatophora* has relocated more than 70 chromatophore genes into the nuclear
148 genome (mostly involved in photosynthesis-related functions) (Nowack *et al.*, 2016; Zhang *et al.*,
149 2017). By contrast, these genes represent less than 1% of the *Paulinella* nuclear genome, while in *A.*
150 *thaliana* some reports suggest that the genes of cyanobacterial origin can account up to 18% of the
151 nuclear genes (Martin *et al.*, 2002). Nonetheless, the chromatophore genome reduction is most
152 likely still ongoing and more genes will still be possibly transferred to the host nucleus. At any rate,
153 based on the reduced number of chromatophore-derived genes in the nuclear genome, it seems that
154 EGT may have been less important in the establishment of the chromatophore in *P. chromatophora*
155 than HGT from other bacteria, as about 170 genes of bacterial origin encode proteins likely targeted
156 to the chromatophore (Nowack *et al.*, 2016). Interestingly, the largest contribution of identified
157 chromatophore-targeted proteins derive from the ancestral host genetic repertoire (Singer *et al.*,
158 2017), a similar pattern as the one observed in Archaeplastida proteomes (Qiu *et al.*, 2013),

159 suggesting that in both primary endosymbioses the host played a crucial role to mediate the
160 cyanobiont integration.

161

162 **IV. Complex plastids**

163 Eukaryotic lineages with complex plastids evolved by the engulfment of red or green algae by
164 different hosts. This type of event, called secondary endosymbiosis, originated plastids with three or
165 four membranes (in contrast with the two membranes of primary plastids) and introduced a high
166 degree of reticulation within the eukaryotic global phylogeny that is not yet fully understood
167 (Archibald, 2015). All or a considerable fraction of members from four ecologically diverse
168 eukaryotic lineages have red alga-derived plastids: cryptophytes, alveolates, stramenopiles, and
169 haptophytes (the “CASH” assemblage). Plastids are particularly diverse in alveolates, including the
170 3-membrane peridinin-containing dinoflagellate plastids and the 4-membrane non-photosynthetic
171 ('apicoplasts') and photosynthetic ones found in parasitic Apicomplexa and their relatives. While the
172 phylogeny of plastid genes supports that a single red alga is at the origin of all complex red plastids
173 (Yoon *et al.*, 2002; Muñoz-Gómez *et al.*, 2017), most phylogenetic analyses of nuclear genes
174 suggest that CASH lineages are not monophyletic. This has been interpreted as an indication that
175 complex red plastids were acquired through an undetermined number of events that may have
176 included serial tertiary endosymbioses and kleptoplastidy in different hosts (Bodył *et al.*, 2009;
177 Baurain *et al.*, 2010; Petersen *et al.*, 2014; Bodył 2017). Conversely, the origin of complex green
178 plastids is much clearer: euglenids and chlorarachniophytes acquired their plastids from two
179 independent secondary endosymbioses involving two distantly related green algal endosymbionts
180 (Jackson *et al.*, 2018).

181 Similar to primary endosymbiosis, during secondary endosymbiotic events genes were
182 transferred from the algal endosymbiont to the host nucleus (secondary EGTs). This was probably
183 accompanied by transfers from bacterial sources (e.g., Záhonová *et al.*, 2018). Unexpectedly, the
184 analysis of nuclear genes of algal origin in eukaryotic lineages with complex plastids has revealed a
185 wide range of putative algal donors different from the algal endosymbionts that exist today as
186 secondary plastids (Curtis *et al.*, 2012). For instance, despite diatoms contain plastids clearly
187 derived from red algae, phylogenetic analysis of diatom nuclear genomes suggested that more than
188 1,700 nucleus-encoded genes were apparently transferred from green algae (Moustafa *et al.*, 2009).
189 Likewise, recent estimations suggest that ~25% of nucleus-encoded plastid-targeted proteins in the
190 ancestor of ochrophytes (photosynthetic stramenopiles) derive from green algae (Dorrell *et al.*,
191 2017). This apparent massive genetic mosaicism may be explained either as the result of high
192 frequency of eukaryote-to-eukaryote HGT or as the consequence of putative cryptic endosymbioses.
193 Nevertheless, although it is possible that the “green” signal observed in ochrophyte genomes,

194 particularly in diatoms, might attest for a former endosymbiosis with a green alga (see Moustafa *et*
195 *al.*, 2009), it seems more likely that the green contribution to diatom genomes has been largely
196 overestimated because of several undetected tree reconstruction artefacts (Deschamps & Moreira,
197 2012).

198

199 **V. The “red carpet” hypothesis**

200 For some secondary EGTs, phylogenetic analysis allows to trace back the full sequence of
201 endosymbiotic gene transfer from cyanobacteria to the Archaeplastida ancestor and then from red or
202 green algal endosymbionts to the secondary photosynthetic lineages. Phylogenetic inspection of
203 these EGTs in the CASH lineages has shown that -as expected- the majority of these genes were
204 transferred from the red algal secondary endosymbiont (Deschamps & Moreira, 2012; Ponce-
205 Toledo *et al.*, 2018). By contrast, about 30% and 50% of this type of cyanobacteria-derived
206 secondary EGTs appear also to have red algal ancestry in euglenids and chlorarachniophytes,
207 respectively (Ponce-Toledo *et al.*, 2018). Thus, these two green plastid-harboring lineages have an
208 unexpected mix of red and green plastid-targeted proteins that creates a high mosaicism in their
209 plastid metabolic pathways, including biosynthetic pathways, photosynthesis-related functions, and
210 plastid biogenesis (Yang *et al.*, 2011). Interestingly, the dinoflagellate *Lepidodinium chlorophorum*,
211 which is a clear case of recent replacement of an original red algal endosymbiont by a pedinophyte-
212 like green algal one (Kamikawa *et al.*, 2015), harbors several nucleus-encoded genes transferred
213 from the former red plastid that were retargeted to the new green plastid. Thus, the current green
214 algal plastid functions with a mix of red and green genes (Minge *et al.*, 2010; Matsuo & Inagaki,
215 2018).

216 However, in the case of euglenids and chlorarachniophytes it is difficult to know if the presence
217 of this large amount of genes of red algal origin is due to the fact that these lineages also
218 experienced former endosymbioses with red algae. Cryptic endosymbiosis scenarios have to be
219 considered cautiously when interpreting the chimerism observed in the nuclear genomes of
220 photosynthetic eukaryotes (Deschamps & Moreira, 2012). Molecular clock analyses suggest that
221 chlorarachniophytes acquired their green plastid 578-318 Mya (Jackson *et al.*, 2018) while the
222 chlorarachniophyte host lineage appears to have diverged from heterotrophic cercozoa ~1000 Mya
223 (Parfrey *et al.*, 2011). Therefore, there was a long period (>400 million years) during which a
224 putative secondary endosymbiosis with a red alga might have taken place before the acquisition of
225 the present green algal plastid. What is certain is that red algae (or lineages containing red algal
226 secondary plastids) provided many plastid-related genes to both euglenids and chlorarachniophytes.
227 These genes appear to be present in all known species of each of these two phyla, indicating that
228 they correspond to ancient gene acquisitions before the diversification of these two phyla (Ponce-

229 Toledo *et al.*, 2018). We introduce here the “red carpet hypothesis” to propose that these red algal
230 genes, transferred to the host nucleus before and/or during the early steps of endosymbiosis with
231 green algae, provided important plastid-related functions and acted as a sort of “red carpet” to
232 facilitate the subsequent adoption of the new green algal endosymbionts.

233

234 **VI. Conclusions**

235 Phylogenetic analyses of nucleus-encoded plastid-targeted proteins have revealed the massive
236 contribution of non-cyanobacterial proteins to the plastid proteomes of Archaeplastida and
237 *Paulinella*, the only two lineages known to harbor primary plastids (Qiu *et al.*, 2013; Nowack *et al.*,
238 2016). Gene transfers from bacteria other than the cyanobiont and their retargeting to the early
239 plastid seem to have been very frequent in both primary photosynthetic lineages, many possibly
240 replacing genes that were lost in the cyanobiont genome. Nonetheless, the eukaryotic host seems to
241 have been the largest contributor of plastid-targeted proteins, with most of them involved in plastid
242 maintenance and transport of metabolites, which supports the idea that the host drove the early steps
243 of plastid endosymbiosis (Karkar *et al.*, 2015).

244 It is still not completely clear why secondary endosymbioses have been much more recurrent
245 than primary ones in the evolution of contemporary eukaryotic photosynthetic lineages. However, it
246 seems that nucleus-encoded genes acquired from previous endosymbioses can be helpful to regain a
247 plastid, likely because the expression of these genes is already under the control of the host and
248 because they carry targeting signals that can be more easily reused to target proteins towards a new
249 endosymbiont (Matsuo & Inagaki, 2018), thus speeding up the host-endosymbiont integration.
250 Although cryptic endosymbioses are difficult to prove, we postulate that plastid-targeted secondary
251 EGTs are critical markers to test these evolutionary scenarios.

252

253

254 **Acknowledgements**

255 We thank Marc-André Selosse for the invitation to contribute this article, Eva Nowack for the
256 *Paulinella chromatophora* picture, and Philippe Deschamps for discussion on plastid evolution
257 topics. Our work was supported by ERC Advanced Grants “Protistworld” and “Plast-Evol” (322669
258 and 787904, respectively) and ANR grant ANCESSTRAM (ANR-15-CE32-0003).

259

260

261 **ORCID**

262 Rafael I. Ponce-Toledo <https://orcid.org/0000-0002-6194-8845>

263 Purificación López-García <https://orcid.org/0000-0002-0927-0651>

264 David Moreira <https://orcid.org/0000-0002-2064-5354>

265

266

267 **References**

268 **Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F,**

269 **Dunthorn M, Hampl V et al. 2012.** The revised classification of eukaryotes. *Journal of*
270 *Eukaryotic Microbiology* **59**: 429–493.

271 **Archibald JM. 2015.** Genomic perspectives on the birth and spread of plastids. *Proceedings of the*
272 *National Academy of Sciences, USA* **112**: 10147–10153.

273 **Baurain D, Brinkmann H, Petersen JJ, Rodríguez-Ezpeleta N, Stechmann A, Demoulin V,**

274 **Roger AJ, Burger G, Lang BF, Philippe H. 2010.** Phylogenomic evidence for separate
275 acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Molecular Biology and*
276 *Evolution* **27**: 1698–1709.

277 **Becker B, Hoef-Emden K, Melkonian M. 2008.** Chlamydial genes shed light on the evolution of
278 photoautotrophic eukaryotes. *BMC Evolutionary Biology* **8**: 203.

279 **Bhattacharya D, Qiu H, Lee JM, Su Yoon H, Weber APM, Price DC. 2018.** When less is more:
280 red algae as models for studying gene loss and genome evolution in eukaryotes. *Critical*
281 *Reviews in Plant Sciences* **37**: 81–99.

282 **Bodyl A. 2017.** How to lose the plasmalemma? Lessons from ciliates, dinoflagellates and
283 euglenids. *BioEssays* **39**: 1700149.

284 **Bodyl A, Stiller JW, Mackiewicz P. 2009.** Chromalveolate plastids: direct descent or multiple
285 endosymbioses? *Trends in Ecology and Evolution* **24**: 119–121.

286 **Cano M, Holland SC, Artier J, Burnap RL, Ghirardi M, Morgan JA, Yu J. 2018.** Glycogen
287 synthesis and metabolite overflow contribute to energy balancing in cyanobacteria. *Cell*
288 *Reports* **23**: 667–672.

289 **Cenci U, Qiu H, Pillonel T, Cardol P, Remacle C, Colleoni C, Kadouche D, Chabi M, Greub G,**
290 **Bhattacharya D et al. 2018.** Host-pathogen biotic interactions shaped vitamin K metabolism in
291 Archaeplastida. *Scientific Reports* **8**: 15243.

292 **Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, Arias MC, Ball SG, Gile**
293 **GH, Hirakawa Y et al. 2012.** Algal genomes reveal evolutionary mosaicism and the fate of
294 nucleomorphs. *Nature* **492**: 59–65.

295 **Dagan T, Roettger M, Stucken K, Landan G, Koch R, Major P, Gould SB, Goremykin V V.,**
296 **Rippka R, De Marsac NT et al. 2013.** Genomes of stigonematalean cyanobacteria (subsection
297 V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids. *Genome Biology*
298 *and Evolution* **5**: 31–44.

299 **Delaye L, Valadez-cano C, Pérez-zamorano B. 2016.** How really ancient is *Paulinella*
300 *chromatophora*? *PLoS Currents Tree of life* **8**:
301 ecurrents.tol.e68a099364bb1a1e129a17b4e06b0c6b.

302 **Deschamps P, Moreira D. 2012.** Reevaluating the green contribution to diatom genomes. *Genome*
303 *Biology and Evolution* **4**: 683–688.

304 **Domman D, Horn M, Embley TM, Williams TA. 2015.** Plastid establishment did not require a
305 chlamydial partner. *Nature Communications* **6**: 6421.

306 **Dorrell RG, Gile G, McCallum G, Méheust R, Baptiste EP, Klinger CM, Brillet-Guéguen L,**
307 **Freeman KD, Richter DJ, Bowler C. 2017.** Chimeric origins of ochrophytes and haptophytes
308 revealed through an ancient plastid proteome. *eLife* **6**: e23717.

309 **Eme L, Sharpe SC, Brown MW, Roger AJ. 2014.** On the age of eukaryotes: Evaluating evidence
310 from fossils and molecular clocks. *Cold Spring Harbor Perspectives in Biology* **6**: a016139.

311 **Facchinelli F, Colleoni C, Ball SG, Weber APM. 2013.** Chlamydia, cyanobiont, or host: Who was
312 on top in the ménage à trois? *Trends in Plant Science* **18**: 673–679.

313 **Gagat P, Mackiewicz P. 2017.** *Cymbomonas tetramitiformis* - a peculiar prasinophyte with a taste
314 for bacteria sheds light on plastid evolution. *Symbiosis* **71**:1–7.

315 **Gavelis GS, Gile GH. 2018.** How did cyanobacteria first embark on the path to becoming
316 plastids?: lessons from protist symbioses. *FEMS Microbiology Letters* **365**: fny209.

317 **Green BR. 2011.** Chloroplast genomes of photosynthetic eukaryotes. *Plant Journal* **66**: 34–44.

318 **Huang J, Gogarten JP. 2007.** Did an ancient chlamydial endosymbiosis facilitate the
319 establishment of primary plastids? *Genome Biology* **8**: R99.

320 **Jackson C, Knoll AH, Chan CX, Verbruggen H. 2018.** Plastid phylogenomics with broad taxon
321 sampling further elucidates the distinct evolutionary origins and timing of secondary green
322 plastids. *Scientific Reports* **8**: 1523.

323 **Kamikawa R, Tanifuji G, Kawachi M, Miyashita H, Hashimoto T, Inagaki Y. 2015.** Plastid
324 genome-based phylogeny pinpointed the origin of the green-colored plastid in the
325 dinoflagellate *Lepidodinium chlorophorum*. *Genome Biology and Evolution* **7**: 1133-1140.

326 **Karkar S, Facchinelli F, Price DC, Weber APM, Bhattacharya D. 2015.** Metabolic connectivity
327 as a driver of host and endosymbiont integration. *Proceedings of the National Academy of*
328 *Sciences, USA* **112**: 10208–10215.

329 **Keeling PJ. 2010.** The endosymbiotic origin, diversification and fate of plastids. *Philosophical*
330 *Transactions of the Royal Society B: Biological Sciences* **365**: 729–748.

331 **Larkum AW, Lockhart PJ, Howe CJ. 2007.** Shopping for plastids. *Trends in Plant Science* **12**:
332 189-195.

- 333 **Lee J, Cho CH, Park SI, Choi JW, Song HS, West JA, Bhattacharya D, Yoon HS. 2016.**
334 Parallel evolution of highly conserved plastid genome architecture in red seaweeds and seed
335 plants. *BMC Biology* **14**: 75.
- 336 **Mackiewicz P, Bodyl A, Gagat P. 2012.** Protein import into the photosynthetic organelles of
337 *Paulinella chromatophora* and its implications for primary plastid endosymbiosis. *Symbiosis*
338 **58**:99–107.
- 339 **Marin B, Nowack ECM, Melkonian M. 2005.** A plastid in the making: Evidence for a second
340 primary endosymbiosis. *Protist* **156**: 425–432.
- 341 **Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa**
342 **M, Penny D. 2002.** Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast
343 genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus.
344 *Proceedings of the National Academy of Sciences, USA* **99**: 12246–12251.
- 345 **Matsuo E, Inagaki Y. 2018.** Patterns in evolutionary origins of heme, chlorophyll a and
346 isopentenyl diphosphate biosynthetic pathways suggest non-photosynthetic periods prior to
347 plastid replacements in dinoflagellates. *PeerJ* **6**: e5345.
- 348 **Méheust R, Zelzion E, Bhattacharya D, Lopez P, Bapteste E. 2016.** Protein networks identify
349 novel symbiogenetic genes resulting from plastid endosymbiosis. *Proceedings of the National*
350 *Academy of Sciences, USA* **113**: 3579–3584.
- 351 **Minge MA, Shalchian-Tabrizi K, Tørresen OK, Takishita K, Probert I, Inagaki Y, Klaveness**
352 **D, Jakobsen KS. 2010.** A phylogenetic mosaic plastid proteome and unusual plastid-targeting
353 signals in the green-colored dinoflagellate *Lepidodinium chlorophorum*. *BMC Evolutionary*
354 *Biology* **10**: 191.
- 355 **Moreira D, Deschamps P. 2014.** What was the real contribution of endosymbionts to the
356 eukaryotic nucleus? Insights from photosynthetic eukaryotes. *Cold Spring Harbor Perspectives*
357 *in Biology* **6**: a016014.
- 358 **Moreira D, Philippe H. 2001.** Sure facts and open questions about the origin and evolution of
359 photosynthetic plastids. *Research in Microbiology* **152**: 771-780.
- 360 **Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D. 2009.** Genomic
361 footprints of a cryptic plastid endosymbiosis in diatoms. *Science* **324**: 1724–1726.
- 362 **Muñoz-Gómez SA, Mejía-Franco FG, Durnin K, Colp M, Grisdale CJ, Archibald JM,**
363 **Slamovits CH. 2017.** The new red algal subphylum Proteorhodophytina comprises the largest
364 and most divergent plastid genomes known. *Current Biology* **27**: 1677-1684.
- 365 **Nowack ECM, Melkonian M, Glöckner G. 2008.** Chromatophore genome sequence of *Paulinella*
366 sheds light on acquisition of photosynthesis by eukaryotes. *Current Biology* **18**: 410–418.

367 **Nowack ECM, Price DC, Bhattacharya D, Singer A, Melkonian M, Grossman AR. 2016.** Gene
368 transfers from diverse bacteria compensate for reductive genome evolution in the
369 chromatophore of *Paulinella chromatophora*. *Proceedings of the National Academy of*
370 *Sciences, USA* **113**: 12214–12219.

371 **Parfrey LW, Lahr DJG, Knoll AH, Katz LA. 2011.** Estimating the timing of early eukaryotic
372 diversification with multigene molecular clocks. *Proceedings of the National Academy of*
373 *Sciences, USA* **108**: 13624–13629.

374 **Petersen J, Ludewig AK, Michael V, Bunk B, Jarek M, Baurain D, Brinkmann H. 2014.**
375 *Chromera velia*, endosymbioses and the rhodoplex hypothesis - Plastid evolution in
376 cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages). *Genome Biology*
377 *and Evolution* **6**: 666–684.

378 **Ponce-Toledo RI, Deschamps P, López-García P, Zivanovic Y, Benzerara K, Moreira D. 2017.**
379 An early-branching freshwater cyanobacterium at the origin of plastids. *Current Biology* **27**:
380 386–391.

381 **Ponce-Toledo RI, Moreira D, López-García P, Deschamps P. 2018.** Secondary plastids of
382 euglenids and chlorarachniophytes function with a mix of genes of red and green algal ancestry.
383 *Molecular Biology and Evolution* **35**: 2198–2204.

384 **Qiu H, Price DC, Weber APM, Facchinelli F, Yoon HS, Bhattacharya D. 2013.** Assessing the
385 bacterial contribution to the plastid proteome. *Trends in Plant Science* **18**: 680–687.

386 **Reyes-Prieto A, Hackett JD, Soares MB, Bonaldo MF, Bhattacharya D. 2006.** Cyanobacterial
387 contribution to algal nuclear genomes is primarily limited to plastid functions. *Current Biology*
388 **16**: 2320–2325.

389 **Sato N, Awai K. 2017.** “Prokaryotic pathway” is not prokaryotic: Noncyanobacterial origin of the
390 chloroplast lipid biosynthetic pathway revealed by comprehensive phylogenomic analysis.
391 *Genome Biology and Evolution* **9**: 3162–3178.

392 **Singer A, Poschmann G, Mühlich C, Valadez-Cano C, Hänsch S, Hüren V, Rensing SA,**
393 **Stühler K, Nowack ECM. 2017.** Massive protein import into the early-evolutionary-stage
394 photosynthetic organelle of the amoeba *Paulinella chromatophora*. *Current Biology* **27**: 2763–
395 2773.

396 **Tyra HM, Linka M, Weber AP, Bhattacharya D. 2007.** Host origin of plastid solute transporters
397 in the first photosynthetic eukaryotes. *Genome Biology* **8**: R212.

398 **Woehle C, Dagan T, Landan G, Vardi A, Rosenwasser S. 2017.** Expansion of the redox-sensitive
399 proteome coincides with the plastid endosymbiosis. *Nature Plants* **3**: 17066.

400 **Yang Y, Maruyama S, Sekimoto H, Sakayama H, Nozaki H. 2011.** An extended phylogenetic
401 analysis reveals ancient origin of “non-green” phosphoribulokinase genes from two lineages of

402 “green” secondary photosynthetic eukaryotes: Euglenophyta and Chlorarachniophyta. *BMC*
403 *Research Notes* **4**: 330.

404 **Yoon HS, Hackett JD, Pinto G, Bhattacharya D. 2002.** The single, ancient origin of chromist
405 plastids. *Proceedings of the National Academy of Sciences, USA* **99**: 15507–15512.

406 **Záhonová K, Füssy Z, Birčák E, Novák Vanclová AMG, Klimeš V, Vesteg M, Krajčovič J,**
407 **Oborník M, Eliáš M. 2018.** Peculiar features of the plastids of the colourless alga *Euglena*
408 *longa* and photosynthetic euglenophytes unveiled by transcriptome analyses. *Scientific Reports*
409 **8**: 17012.

410 **Zhang R, Nowack ECM, Price DC, Bhattacharya D, Grossman AR. 2017.** Impact of light
411 intensity and quality on chromatophore and nuclear gene expression in *Paulinella*
412 *chromatophora*, an amoeba with nascent photosynthetic organelles. *Plant Journal* **90**: 221–234.
413

414 **Figure legends**

415 **Fig. 1** The distribution of photosynthesis in the global eukaryotic phylogeny. Colored solid
416 branches correspond to photosynthetic lineages endowed with primary plastids and colored dashed
417 branches to lineages with secondary plastids (green and red colors indicate the type of secondary
418 endosymbiont, green or red algae, respectively). Blue arrows show the two known primary
419 endosymbioses (in Archaeplastida and *Paulinella*) and green and red arrows indicate the secondary
420 endosymbioses involving green and red algal endosymbionts. Grey branches correspond to non-
421 photosynthetic eukaryotic phyla. The tree has been largely modified from Adl *et al.*, 2012.

422

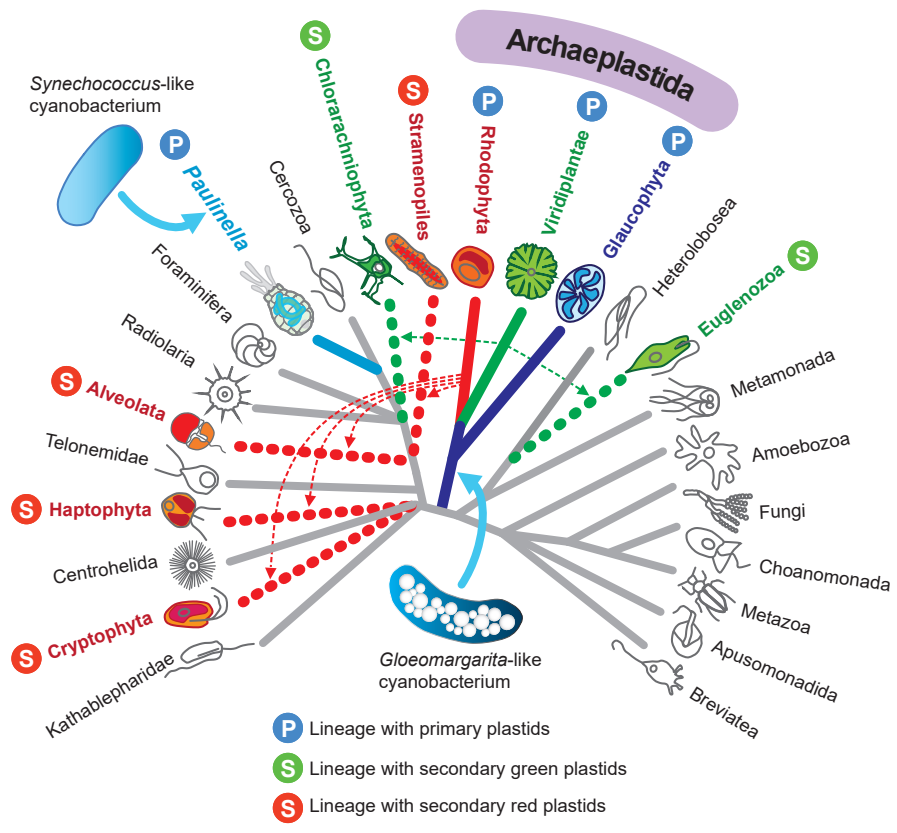
423 **Fig. 2** Light microscopy image of *Paulinella chromatophora*. The pigmented *Synechococcus*-like
424 primary plastids are easily visible within the cytoplasm. Scale bar: 10 μm . Image courtesy of Eva
425 Nowack (Heinrich-Heine-Universität Düsseldorf).

426

427 **Table 1** Comparison of plastid characters between Archaeplastida and *Paulinella chromatophora*

Character	Archaeplastida	<i>Paulinella</i>	Reference
Plastid genome size	100-200 kbp	1021 kbp	Nowack et al., 2008
Number of plastid genes	80-250	911	Nowack et al., 2008
EGTs in plastid proteome	70-390	>70	Qiu et al., 2013; Singer et al., 2017; Nowack et al., 2016; Zhang et al., 2017
Non-cyanobacterial prokaryotic proteins in plastid proteome	40-240	>170	Qiu et al., 2013; Singer et al., 2017; Nowack et al., 2016; Zhang et al., 2017
Host-derived and proteins of uncertain origin in plastid proteome	320-900	>390	Qiu et al., 2013; Singer et al., 2017; Nowack et al., 2016; Zhang et al., 2017
Lineage age estimation	>1000 Myr	90-140 Myr	Eme et al., 2014; Delaye et al., 2016
Import system of nucleus-encoded proteins into the plastid	Translocons at the outer and inner plastid membranes (TOC/TIC complex)	Vesicles of the host endomembrane system fuse with the outer plastid membrane and proteins cross the inner membrane through a simplified TIC translocon	Mackiewicz et al., 2012
Phagotrophic capacity	Lost in Rhodophyta, Glaucophyta, and most Viridiplantae but preserved in some prasinophytes	Lost	Gagat & Mackiewicz, 2017
Peptidoglycan wall	Present in Glaucophyta and some Viridiplantae species	Present	Gagat & Mackiewicz, 2017

428



Ponce-Toledo et al., Fig. 1



Ponce-Toledo et al., Fig. 2