



HAL
open science

Secondary Plastids of Euglenids and Chlorarachniophytes Function with a Mix of Genes of Red and Green Algal Ancestry

Rafael I Ponce-Toledo, David Moreira, Purificacion Lopez-Garcia, Philippe Deschamps

► **To cite this version:**

Rafael I Ponce-Toledo, David Moreira, Purificacion Lopez-Garcia, Philippe Deschamps. Secondary Plastids of Euglenids and Chlorarachniophytes Function with a Mix of Genes of Red and Green Algal Ancestry. *Molecular Biology and Evolution*, 2018, 35 (9), pp.2198-2204. 10.1093/molbev/msy121 . hal-02108901

HAL Id: hal-02108901

<https://hal.science/hal-02108901>

Submitted on 24 Apr 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.

1 **Secondary Plastids of Euglenids and Chlorarachniophytes Function with a Mix**
2 **of Genes of Red and Green Algal Ancestry**

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

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

6 ***Corresponding authors:** E-mails: philippe.deschamps@u-psud.fr; david.moreira@u-psud.fr.

7

8 **Abstract**

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

21

22 **Key words:** Chlorarachniophyta, Euglenida, endosymbiotic gene transfer, phylogenomics, plastids.

23 Photosynthesis in eukaryotes takes place in a specialized compartment: the plastid. This organelle
24 first evolved in a common ancestor of Archaeplastida (*i.e.*, Viridiplantae + Rhodophyta +
25 Glaucophyta) through the endosymbiosis of a cyanobacterium inside a eukaryotic host (Moreira and
26 Philippe 2001; Archibald 2009; Keeling 2013). This primary endosymbiotic event entailed massive
27 endosymbiotic gene transfer (EGT) from the cyanobacterial genome to the host nucleus (Weeden
28 1981; Kleine et al. 2009). Consequently, most proteins required for the proper functioning of
29 primary plastids are encoded in the nuclear genome and addressed to the plastid lumen via
30 specialized signal sequences and a translocation apparatus (Gutensohn et al. 2006). Other
31 photosynthetic eukaryotic phyla obtained their plastids through secondary endosymbiosis, *i.e.* the
32 symbiosis of either green or red algae within another eukaryotic cell, or even through tertiary
33 endosymbiosis (symbiosis of secondary photosynthetic eukaryotes within eukaryotic hosts)
34 (Delwiche 1999; Archibald 2009; Keeling 2013). Euglenida (Excavata) and Chlorarachniophyta
35 (Rhizaria) carry green algal secondary plastids ('green plastids') acquired through two independent
36 endosymbioses involving Prasinophyceae and Ulvophyceae green algae, respectively (Rogers et al.
37 2007; Hrdá et al. 2012; Suzuki et al. 2016). Photosynthetic species in the Cryptophyta, Alveolata,
38 Stramenopiles and Haptophyta (CASH) lineages have plastids derived from red algae ('red plastids')
39 but so far it has been impossible to retrace a consensual evolutionary history (Lane and Archibald
40 2008; Archibald 2009; Keeling 2013). Whereas phylogenomic analyses of plastid-encoded genes
41 support the monophyly of all CASH plastids, arguing for a single red algal secondary
42 endosymbiosis (Yoon et al. 2002; Muñoz-Gómez et al. 2017), most of the phylogenies based on
43 host nuclear genes do not retrieve their monophyly (Baurain et al. 2010; Burki et al. 2016). To
44 reconcile these incongruent results, some authors have proposed the hypothesis that a unique
45 phylum (which may have gone extinct or evolved into one of the extant CASH phyla) acquired a
46 red alga through secondary endosymbiosis and originated the first lineage of red secondary algae.
47 Subsequently, this lineage would have transmitted the secondary red plastid to other CASH phyla

48 via serial tertiary endosymbioses involving different hosts (Larkum et al. 2007; Sanchez-Puerta and
49 Delwiche 2008; Bodył et al. 2009; Baurain et al. 2010; Petersen et al. 2014).

50 As for the primary endosymbiosis, each secondary or tertiary endosymbiosis was accompanied
51 by numerous EGTs from the nucleus of the endosymbiotic red or green alga to the host nucleus.
52 Consequently, secondary photosynthetic eukaryotes possess two types of genes that can inform
53 about the phylogenetic identity of their plastids: plastid-encoded genes and nucleus-encoded genes
54 acquired via EGT. Genes encoded in primary plastid genomes and the EGTs found in the genomes
55 of Archaeplastida are related to cyanobacteria and have helped to identify the cyanobacterial lineage
56 at the origin of the first plastid (Ponce-Toledo et al. 2017). Similarly, plastid-encoded genes and
57 EGTs found in nuclear genomes of secondary photosynthetic eukaryotes are expected to be useful
58 to determine the red or green algal origin of their plastids. Compared to plastid-encoded genes,
59 EGTs have the additional advantage that they can inform about the presence and identity of past
60 plastids in lineages where plastids have been lost or replaced (cryptic plastid endosymbioses).
61 However, if EGTs are valuable to track contemporary and cryptic endosymbioses, their detection
62 within whole nuclear genome sequences remains a complex task (Stiller 2011). In the case of
63 primary endosymbiosis, EGT detection is rather straightforward because cyanobacterial-type genes
64 are easily distinguishable from typical eukaryotic nuclear genes. The situation is more difficult in
65 the case of secondary endosymbioses. Indeed, detection of EGT genes transferred from the nucleus
66 of green or red algal endosymbionts can be ambiguous due to the poor resolution often found in
67 single gene phylogenies that hampers distinguishing EGTs from vertically inherited nuclear genes,
68 especially considering the short phylogenetic distance between Archaeplastida and several groups
69 of secondary algae. Two studies on red-plastid-bearing algae, the chromerids (Alveolata) and the
70 diatoms (Stramenopiles), illustrate this issue. Both reported an unexpected high number of genes
71 phylogenetically related to green algal homologs. Whereas in the case of the chromerids the green
72 signal was attributed to probable phylogenetic artifacts and the reduced sampling of red algal

73 genome sequences (Woehle et al. 2011), it was interpreted in diatoms as evidence for a cryptic
74 green algal endosymbiont (Moustafa et al. 2009). However, the subsequent reanalyses of the same
75 genes using richer taxonomic sampling and more robust phylogenetic methods largely erased the
76 evidence for cryptic green endosymbioses in these CASH phyla (Burki et al. 2012; Deschamps and
77 Moreira 2012; Moreira and Deschamps 2014).

78 The extent and impact of horizontal gene transfer (HGT) on eukaryotic evolution remain
79 controversial topics (Leger et al. 2018). HGTs might be valuable to infer the history of genomes and
80 lineages (Abby et al. 2012) but they can also introduce inconvenient noise in phylogenomic
81 analyses, in particular for the study of EGTs (Stiller 2011). Through time, secondary photosynthetic
82 eukaryotes may have accumulated HGTs in their nuclear genomes from various sources, perhaps
83 even including non-endosymbiotic red or green algae. Unfortunately, gene phylogenies of such
84 HGTs may display topologies comparable to those of EGTs, making them difficult to set apart. In
85 this context, anomalous phylogenetic signal in certain secondary photosynthetic groups has been
86 interpreted as HGT rather than EGT from cryptic endosymbionts. This is the case of the nuclear
87 genome sequence of the green-plastid-containing chlorarachniophyte alga *Bigelowiella natans*, in
88 which 22% of the genes potentially acquired via HGT appeared to have a red algal origin (Curtis et
89 al. 2012). Because of the phagotrophic ability of chlorarachniophytes, the presence of these genes
90 was considered to be the result of progressive accumulation of HGTs from red algae or from red-
91 plastid-containing CASH lineages, some eventually substituting original 'green' EGTs (Archibald et
92 al. 2003; Yang et al. 2011; Yang et al. 2014). Analogous studies on euglenid species suggested a
93 similar trend for several genes involved in central metabolic pathways (Maruyama et al. 2011; Yang
94 et al. 2011; Markunas and Triemer 2016). The unexpected presence of those 'red' genes in
95 chlorarachniophytes and euglenids was first considered as the result of multiple HGTs (e.g.,
96 Archibald et al. 2003; Maruyama et al. 2011) but the increasing number of reported cases has
97 prompted some authors to speculate on putative cryptic red algal endosymbioses in both lineages

98 (Maruyama et al. 2011; Markunas and Triemer 2016). A systematic investigation of HGT/EGT is
99 still missing in euglenids and chlorarachniophytes but, as mentioned above, in the context of
100 secondary endosymbioses it can be difficult to distinguish among HGT, EGT, and just unresolved
101 trees on the basis of single-gene phylogenies (Deschamps and Moreira 2012).

102 In this work, we have focused on a particular group of genes to reduce this uncertainty: genes
103 transferred from the original cyanobacterial plastid endosymbiont into the nuclear genome of
104 Archaeplastida and subsequently transferred from Archaeplastida into the genomes of complex
105 secondary algae. In Archaeplastida, these genes are known to be involved in essential plastid
106 functions and tend to be highly conserved (Reyes-Prieto et al. 2006; Deschamps and Moreira 2009),
107 so we expected that they can provide strong phylogenetic signal. To identify them, we queried by
108 BLAST the whole predicted proteomes of *Guillardia theta* and *Bigeloviella natans* against a local
109 genome database containing representatives of the three domains of life, in particular a
110 comprehensive collection of genomes and transcriptomes of photosynthetic protists (supplementary
111 table S1, Supplementary Material online). *Guillardia* and *Bigeloviella* proteins with hits in other
112 photosynthetic eukaryotes and in cyanobacteria were selected for phylogenetic analysis. Maximum
113 likelihood (ML) phylogenetic trees for these proteins were constructed and manually filtered to
114 retain those fulfilling two criteria: i) trees have to support a clear separation of Viridiplantae and
115 Rhodophyta (with secondary lineages branching within them), and ii) proteins have to be shared by
116 at least three secondary photosynthetic lineages. We identified in this way 82 genes most likely
117 acquired by secondary photosynthetic eukaryotes from Archaeplastida. 70 were cyanobacterial
118 genes likely transferred sequentially through primary and secondary endosymbioses, and 12 were
119 derived from diverse bacterial groups likely transferred to a common ancestor of Archaeplastida and
120 subsequently transferred to secondary photosynthetic groups (supplementary table S3 and figs. S1-
121 S82, Supplementary Material online). Interestingly, most of these genes were absent in non-
122 photosynthetic eukaryotes, supporting that they were not misinterpreted vertically-inherited ones.

123 Most of the 82 ML phylogenies were well resolved and enabled us to unambiguously determine,
124 for each secondary lineage, whether the source of the gene was a green or a red alga. As expected,
125 in the great majority of our trees (between 84 and 90%, fig. 1A) the genes of red-plastid-endowed
126 CASH lineages derived from red algae (e.g., fig. 2A and 2B). Because of their secondary green
127 plastids, we expected the opposite situation in chlorarachniophytes and euglenids, namely a
128 majority of 'green' genes. However, 42 of the 78 trees where chlorarachniophytes were present
129 (54%, fig. 1A) supported a 'red' origin of the corresponding genes (e.g., fig. 2A). Similarly, 22 of the
130 61 trees containing euglenids (36%, fig. 1A) also supported a 'red' ancestry (e.g., fig. 2B). These
131 surprisingly high values were in sharp contrast with the small number of trees (<10%, fig. 1A)
132 showing CASH phyla embedded within green algae. Interestingly, the CASH phyla were
133 monophyletic in 7 of these trees, arguing for a common evolutionary origin of the corresponding
134 'green' genes. Almost all of the 82 genes identified here encode plastid-targeted proteins involved in
135 essential plastid functions (fig. 1B and supplementary table S4, Supplementary Material online). For
136 instance, in both chlorarachniophytes and euglenids, these nuclear-encoded 'red' genes participate in
137 plastid genome expression (e.g., elongation factors and aminoacyl-tRNA synthetases), light
138 harvesting, chlorophyll biosynthesis, and photosystem II assembly. Keeping these important genes
139 implies a plastid-related selective pressure, which excludes that they could have accumulated in the
140 heterotrophic ancestors of green secondary photosynthetic eukaryotes prior to plastid acquisition.

141 The marked disproportion of unexpected gene sources in green *versus* red secondary
142 photosynthetic lineages is intriguing and may be interpreted in different ways. First, the green algal
143 ancestors of chlorarachniophyte and euglenid plastids may have had a high proportion of red algal
144 HGT genes in their genomes. However, such a high HGT proportion involving essential genes has
145 not been reported so far in any green alga. Second, these 'red' genes may have accumulated in
146 chlorarachniophyte and euglenid nuclear genomes by numerous HGTs, for example from food
147 sources. This would imply that, for unknown reasons, HGT is much more frequent in secondary

148 green lineages than in red ones, as well as a long-lasting feeding preference towards red prey in
149 both secondary green lineages. Moreover, the 'red' genes are shared by all the species of the
150 relatively rich taxon sampling available for chlorarachniophytes (fig. 2A), indicating that their
151 acquisition predated the diversification of this group and stopped afterwards (we did not retrieve
152 any tree supporting a recent HGT involving only a subgroup of chlorarachniophytes). Our data
153 therefore argue for an ancient timing of 'red' gene acquisition. These observations may support a
154 third interpretation: the 'red' genes are shared by all SAR lineages (Stramenopiles, Alveolata, and
155 Rhizaria) because they were acquired from a single common secondary red algal endosymbiosis
156 ancestral to the whole SAR supergroup. This original red plastid would have been lost in many
157 phyla and replaced by a green alga in chlorarachniophytes. However, this scenario poses several
158 problems. On the one hand, traces of past presence of red algal plastids, in the form of EGTs, in
159 non-photosynthetic SAR lineages are very often controversial (Elias and Archibald 2009; Stiller et
160 al. 2009; Stiller 2011). On the other hand, plastid-bearing chlorarachniophytes constitute a relatively
161 late-emerging branch within SAR (Sierra et al. 2016), implying that if their present-day green
162 plastid replaced a former red one, this red plastid would have had to be present until recently and
163 been lost in all other rhizarian lineages, which may seem unparsimonious. The case of euglenids is
164 even more difficult to interpret as this group of excavates has no close phylogenetic relationship
165 with any other photosynthetic lineage. In addition, massive sequence data remain much more
166 limited for euglenids than for chlorarachniophytes (only a few transcriptomes available, see
167 supplementary table S1, Supplementary Material online), making it difficult to infer the relative age
168 of possible gene transfers. Nonetheless, 'red' genes were often shared by several euglenids in our
169 trees, suggesting a similar pattern of ancient acquisition as in chlorarachniophytes (supplementary
170 figs. S1-S82, Supplementary Material online).

171 Our results show the presence of an unexpectedly high number of genes of red algal affinity in
172 the two groups of eukaryotic algae with secondary green plastids, the euglenids and

173 chlorarachniophytes, which is significantly higher than the frequency of 'green' genes in algae with
174 secondary red plastids, the CASH lineages. To address this question, we have focused on a subset of
175 genes selected because of their strong phylogenetic signal and their implication in plastid-related
176 activities. It is therefore uncertain whether this conclusion can be applied to the rest of HGTs/EGTs
177 potentially present in the genomes of all these algae. In fact, in addition to the problems inherent to
178 the accurate detection of EGTs, our focus on these specific genes may explain, at least partly, the
179 different results obtained in recent analyses of all potential EGTs in some CASH lineages, not only
180 those of ultimate cyanobacterial origin (e.g., Dorrell et al. 2017).

181 However, we could not identify any particular bias in our gene selection process that could have
182 artificially enriched the observed 'red' gene frequency in euglenids and chlorarachniophytes.
183 Despite the methodological problems inherent to global genome analyses cited above, including a
184 highly unbalanced representation of red and green algal genomes in sequence databases
185 (Deschamps and Moreira 2012), the study of the chlorarachniophyte *B. natans* genome already
186 pointed in that direction, with 22% of EGT genes of apparent red algal ancestry (Curtis et al. 2012).
187 The origin of the 'red' genes in euglenids and chlorarachniophytes, either by cumulative HGT or by
188 EGT from cryptic red algal endosymbionts, remains mysterious but our work indicates that they
189 were acquired early in both groups and that they fulfill essential functions for plastid activity and
190 maintenance. Interestingly, indisputable evidence supports that in a third group of complex algae
191 with green plastids, the dinoflagellate genus *Lepidodinium*, a former red plastid was replaced by the
192 current green one, leading to a mosaic plastid proteome encoded by a mix of red and green algal
193 genes (Minge et al. 2010), reminiscent of those found in euglenids and chlorarachniophytes. It has
194 been proposed that, since they retain more gene-rich genomes than green ones, red plastids have
195 increased capacity for autonomous metabolism that could explain why they are more widespread
196 across the diversity of eukaryotes as secondary plastids (the "portable plastid" hypothesis (Grzebyk
197 et al. 2003)). It is thus tempting to speculate for euglenids and chlorarachniophytes a similar case as

198 for *Lepidodinium*, with initial red plastids subsequently replaced by green ones. Even if this
199 hypothesis turns out to be wrong and these cryptic red endosymbioses did not exist, the ancient
200 acquisition by another mechanism of a significant number of red algal genes in both groups before
201 their diversification and, especially, their maintenance in the contemporary species through millions
202 of years of evolution, suggest that the 'red' genes were instrumental in the establishment and
203 maintenance of the secondary green plastids. Sequencing and analysis of additional genomes of
204 euglenids, chlorarachniophytes, and their non-photosynthetic relatives will help to refine the
205 inventory of 'red' genes in these lineages and their timing and, eventually, mechanism of acquisition.

206

207 **Materials and Methods**

208 **Sequence Analysis**

209 A local database was constructed to host the predicted proteomes from various nuclear genomes and
210 transcriptomes as well as plastid genomes (for the complete list, see supplementary table S1,
211 Supplementary Material online). All proteins of the *Bigelowiella natans* (Chlorarachniophyta) and
212 *Guillardia theta* (Cryptophyta) predicted proteomes were used as queries for BLASTp sequence
213 similarity searches (Camacho et al. 2009) against the local database. We retained up to 350 top hits
214 with an e-value threshold of 1e-05. BLASTp outputs were parsed with a custom Python script to
215 identify the proteins having hits in diverse photosynthetic eukaryotes and that were more similar to
216 cyanobacteria or other bacteria than to non-photosynthetic eukaryotes.

217 For these proteins, reciprocal BLASTp searches were done against the database to collect up to
218 600 similar sequences. We then aligned each set using Mafft v7.123b (Katoh and Standley 2013)
219 with default parameters. Non-conserved alignment regions were trimmed with BMGE v1.0
220 (Criscuolo and Gribaldo 2010) with the BLOSUM62 matrix and allowing less than 50% gaps per
221 position. Preliminary phylogenetic trees were inferred from trimmed alignments using FastTree
222 v2.1.7 (Price et al. 2010) with default parameters. These trees were then manually inspected to

223 identify those compatible with an EGT/HGT scenario. For all positive cases, only the sequences
224 corresponding to the portion of interest of each phylogenetic tree (the part showing the
225 photosynthetic eukaryotes and the closest outgroup) were retained for the remaining steps. We then
226 removed very short partial sequences and, to speed up subsequent calculations, several outgroup
227 sequences from all alignments (see supplementary table S2, Supplementary Material online). The
228 final sequence datasets were realigned and trimmed using TrimAL v1.4.rev15 with “gappy-out”
229 parameter (Capella-Gutierrez et al. 2009). ML phylogenetic trees were inferred using IQtree v1.5.1
230 with the PMSF model of sequence evolution (Wang et al. 2018) parameterized using guided trees
231 constructed with the LG+G+I model. Statistical support was calculated with 1000 ultrafast
232 bootstrap replicates (Minh et al. 2013; Nguyen et al. 2015; Hoang et al. 2018).

233 Final selection of trees was done by manual inspection to keep those fulfilling the following
234 two requirements: i) the protein had to be shared by Cyanobacteria (or other bacteria),
235 Archaeplastida and at least three secondary photosynthetic lineages, and ii) the corresponding
236 phylogenetic trees had to support the clear separation of Viridiplantae and Rhodophyta (plus the
237 lineages with secondary green and red plastids nested within them). Finally, the 82 trees passing this
238 final filter (supplementary figs. S1-S82, Supplementary Material online) were inspected to infer the
239 phylogenetic origin of the corresponding genes in the secondary photosynthetic lineages
240 (supplementary table S3, Supplementary Material online).

241

242 **Gene Functional Annotation**

243 We annotated the functions of the 82 proteins from the final selection (see above) through the
244 EggNOG 4.5 (Huerta-Cepas et al. 2016) web portal (<http://eggnogdb.embl.de>). For each protein we
245 used as queries the ortholog sequences of *Guillardia theta* and *Bigeloviella natans*. Functional
246 annotations are shown in supplementary table S4, Supplementary Material online.

247

248 **Data Availability**

249 Protein sequence datasets used in this work are available for download at <http://www.esse.u-psud.fr/article950.html?lang=en>. They include nonaligned sequences and trimmed alignments.

251

252 **Supplementary Material**

253 Supplementary figures S1–S82 and tables S1-S4 are available at Molecular Biology and Evolution
254 online (<http://www.mbe.oxfordjournals.org/>).

255

256 **Acknowledgments**

257 This study was supported by European Research Council grant ProtistWorld (P.L.-G., agreement no.
258 322669), the Université Paris-Sud program “Attractivité” (P.D.) and the Agence Nationale de la
259 Recherche (D.M., project ANR-15-CE32-0003 "ANCESSTRAM"). We thank the Associated Editor
260 and two anonymous reviewers for constructive comments.

261

262 **References**

- 263 Abby SS, Tannier E, Gouy M, Daubin V. 2012. Lateral gene transfer as a support for the tree of life. *Proc*
264 *Natl Acad Sci U S A* 109:4962-4967.
- 265 Archibald JM. 2009. The puzzle of plastid evolution. *Curr Biol.* 19:R81-88.
- 266 Archibald JM, Rogers MB, Toop M, Ishida KI, Keeling PJ. 2003. Lateral gene transfer and the evolution of
267 plastid-targeted proteins in the secondary plastid-containing alga *Bigeloviella natans*. *Proc Natl Acad Sci*
268 *U S A* 100:7678-7683.
- 269 Baurain D, Brinkmann H, Petersen J, Rodriguez-Ezpeleta N, Stechmann A, Demoulin V, Roger AJ, Burger
270 G, Lang BF, Philippe H. 2010. Phylogenomic evidence for separate acquisition of plastids in
271 cryptophytes, haptophytes, and stramenopiles. *Mol Biol Evol.* 27:1698-1709.
- 272 Bodył A, Stiller JW, Mackiewicz P. 2009. Chromalveolate plastids: direct descent or multiple
273 endosymbioses? *Trends Ecol Evol.* 24:119-121.
- 274 Burki F, Flegontov P, Obornik M, Cihlar J, Pain A, Lukes J, Keeling PJ. 2012. Re-evaluating the green
275 versus red signal in eukaryotes with secondary plastid of red algal origin. *Genome Biol Evol.* 4:626-635.
- 276 Burki F, Kaplan M, Tikhonenkov DV, Zlatogursky V, Minh BQ, Radaykina LV, Smirnov A, Mylnikov AP,
277 Keeling PJ. 2016. Untangling the early diversification of eukaryotes: a phylogenomic study of the
278 evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc Biol Sci.* 283:1823.

- 279 Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, Arias MC, Ball SG, Gile GH, Hirakawa Y
280 et al. 2012. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* 492:59-
281 65.
- 282 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+:
283 architecture and applications. *BMC Bioinformatics*. 10:421.
- 284 Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming
285 in large-scale phylogenetic analyses. *Bioinformatics*. 25:1972-1973.
- 286 Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for
287 selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol*.
288 10:1471-2148.
- 289 Delwiche CF. 1999. Tracing the thread of plastid diversity through the tapestry of life. *Am Nat*. 154:S164-
290 S177.
- 291 Deschamps P, Moreira D. 2012. Reevaluating the green contribution to diatom genomes. *Genome Biol Evol*.
292 4:683-688.
- 293 Deschamps P, Moreira D. 2009. Signal conflicts in the phylogeny of the primary photosynthetic eukaryotes.
294 *Mol Biol Evol*. 26:2745-2753.
- 295 Dorrell RG, Gile G, McCallum G, Méheust R, Baptiste EP, Klinger CM, Brillet-Guéguen L, Freeman KD,
296 Richter DJ, Bowler C. 2017. Chimeric origins of ochrophytes and haptophytes revealed through an
297 ancient plastid proteome. *Elife*. 6.
- 298 Elias M, Archibald JM. 2009. Sizing up the genomic footprint of endosymbiosis. *Bioessays* 31:1273-1279.
- 299 Grzebyk D, Schofield O, Vetriani C, Falkowski PG. 2003. The mesozoic radiation of eukaryotic algae: the
300 portable plastid hypothesis. *J Phycol*. 39:259-267.
- 301 Gutensohn M, Fan E, Frielingsdorf S, Hanner P, Hou B, Hust B, Klosgen RB. 2006. Toc, Tic, Tat et al.:
302 structure and function of protein transport machineries in chloroplasts. *J Plant Physiol*. 163:333-347.
- 303 Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the ultrafast
304 bootstrap approximation. *Mol Biol Evol*. 35:518-522.
- 305 Hrdá Š, Fousek J, Szabová J, Hampl V, Vlček Č. 2012. The plastid genome of *Eutreptiella* provides a
306 window into the process of secondary endosymbiosis of plastid in euglenids. *PLoS One* 7:e33746.
- 307 Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa
308 S, Kuhn M, Jensen LJ, von Mering C, Bork P. 2016. eggNOG 4.5: a hierarchical orthology framework
309 with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res*.
310 44:D286-293.
- 311 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in
312 performance and usability. *Mol Biol Evol*. 30:772-780.
- 313 Keeling PJ. 2013. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annu
314 Rev Plant Biol*. 64:583-607.
- 315 Kleine T, Maier UG, Leister D. 2009. DNA transfer from organelles to the nucleus: the idiosyncratic genetics
316 of endosymbiosis. *Annu Rev Plant Biol*. 60:115-138.
- 317 Lane CE, Archibald JM. 2008. The eukaryotic tree of life: endosymbiosis takes its TOL. *Trends Ecol Evol*.
318 23:268-275.
- 319 Larkum AW, Lockhart PJ, Howe CJ. 2007. Shopping for plastids. *Trends Plant Sci*. 12:189-195.
- 320 Leger MM, Eme L, Stairs CW, Roger AJ. 2018. Demystifying eukaryote lateral gene transfer.
321 *Bioessays*.40:e1700242.
- 322 Markunas CM, Triemer RE. 2016. Evolutionary history of the enzymes involved in the Calvin-Benson cycle
323 in euglenids. *J Eukaryot Microbiol*. 63:326-339.
- 324 Maruyama S, Suzuki T, Weber AP, Archibald JM, Nozaki H. 2011. Eukaryote-to-eukaryote gene transfer
325 gives rise to genome mosaicism in euglenids. *BMC Evol Biol*. 11:1471-2148.
- 326 Minge MA, Shalchian-Tabrizi K, Torresen OK, Takishita K, Probert I, Inagaki Y, Klaveness D, Jakobsen KS.
327 2010. A phylogenetic mosaic plastid proteome and unusual plastid-targeting signals in the green-colored
328 dinoflagellate *Lepidodinium chlorophorum*. *BMC Evol Biol*. 10:191.
- 329 Minh BQ, Nguyen MA, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol
330 Evol*. 30:1188-1195.
- 331 Moreira D, Deschamps P. 2014. What was the real contribution of endosymbionts to the eukaryotic nucleus?
332 Insights from photosynthetic eukaryotes. *Cold Spring Harb Perspect Biol*. 6:a016014.

- 333 Moreira D, Philippe H. 2001. Sure facts and open questions about the origin and evolution of photosynthetic
334 plastids. *Res Microbiol.* 152:771-780.
- 335 Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D. 2009. Genomic footprints of a
336 cryptic plastid endosymbiosis in diatoms. *Science* 324:1724-1726.
- 337 Muñoz-Gómez SA, Mejía-Franco FG, Durnin K, Colp M, Grisdale CJ, Archibald JM, Slamovits CH. 2017.
338 The new red algal subphylum Proteorhodophytina comprises the largest and most divergent plastid
339 genomes known. *Curr Biol.* 27:1677-1684.
- 340 Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic
341 algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32:268-274.
- 342 Petersen J, Ludewig AK, Michael V, Bunk B, Jarek M, Baurain D, Brinkmann H. 2014. *Chromera velia*,
343 endosymbioses and the rhodoplex hypothesis--plastid evolution in cryptophytes, alveolates,
344 stramenopiles, and haptophytes (CASH lineages). *Genome Biol Evol.* 6:666-684.
- 345 Ponce-Toledo RI, Deschamps P, López-García P, Zivanovic Y, Benzerara K, Moreira D. 2017. An early-
346 branching freshwater cyanobacterium at the origin of plastids. *Curr Biol.* 27:386-391.
- 347 Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large
348 alignments. *PLoS One.* 5:0009490.
- 349 Reyes-Prieto A, Hackett JD, Soares MB, Bonaldo MF, Bhattacharya D. 2006. Cyanobacterial contribution to
350 algal nuclear genomes is primarily limited to plastid functions. *Curr Biol.* 16:2320-2325.
- 351 Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ. 2007. The complete chloroplast genome of the
352 chlorarachniophyte *Bigeloviella natans*: evidence for independent origins of chlorarachniophyte and
353 euglenid secondary endosymbionts. *Mol Biol Evol.* 24:54-62.
- 354 Sanchez-Puerta MV, Delwiche CF. 2008. A hypothesis for plastid evolution in chromalveolates. *J Phycol.*
355 44:1097-1107.
- 356 Sierra R, Canas-Duarte SJ, Burki F, Schwelm A, Fogelqvist J, Dixelius C, Gonzalez-Garcia LN, Gile GH,
357 Slamovits CH, Klopp C et al. 2016. Evolutionary origins of rhizarian parasites. *Mol Biol Evol.* 33:980-
358 983.
- 359 Stiller JW. 2011. Experimental design and statistical rigor in phylogenomics of horizontal and endosymbiotic
360 gene transfer. *BMC Evol Biol.* 11:1471-2148.
- 361 Stiller JW, Huang J, Ding Q, Tian J, Goodwillie C. 2009. Are algal genes in nonphotosynthetic protists
362 evidence of historical plastid endosymbioses? *BMC Genomics.* 10:484.
- 363 Suzuki S, Hirakawa Y, Kofuji R, Sugita M, Ishida K. 2016. Plastid genome sequences of *Gymnochloa*
364 *stellata*, *Lotharella vacuolata*, and *Partenskyella glossopodia* reveal remarkable structural conservation
365 among chlorarachniophyte species. *J Plant Res.* 129:581-590.
- 366 Wang HC, Minh BQ, Susko E, Roger AJ. 2018. Modeling site heterogeneity with posterior mean site
367 frequency profiles accelerates accurate phylogenomic estimation. *Syst Biol.* 67:216-235.
- 368 Weeden NF. 1981. Genetic and biochemical implications of the endosymbiotic origin of the chloroplast. *J*
369 *Mol Evol.* 17:133-139.
- 370 Woehle C, Dagan T, Martin WF, Gould SB. 2011. Red and problematic green phylogenetic signals among
371 thousands of nuclear genes from the photosynthetic and apicomplexa-related *Chromera velia*. *Genome*
372 *Biol Evol.* 3:1220-1230.
- 373 Yang Y, Maruyama S, Sekimoto H, Sakayama H, Nozaki H. 2011. An extended phylogenetic analysis reveals
374 ancient origin of "non-green" phosphoribulokinase genes from two lineages of "green" secondary
375 photosynthetic eukaryotes: Euglenophyta and Chlorarachniophyta. *BMC Res Notes* 4:330.
- 376 Yang Y, Matsuzaki M, Takahashi F, Qu L, Nozaki H. 2014. Phylogenomic analysis of "red" genes from two
377 divergent species of the "green" secondary phototrophs, the chlorarachniophytes, suggests multiple
378 horizontal gene transfers from the red lineage before the divergence of extant chlorarachniophytes. *PLoS*
379 *One* 9:e101158.
- 380 Yoon HS, Hackett JD, Pinto G, Bhattacharya D. 2002. The single, ancient origin of chromist plastids. *Proc*
381 *Natl Acad Sci U S A.* 99:15507-15512.

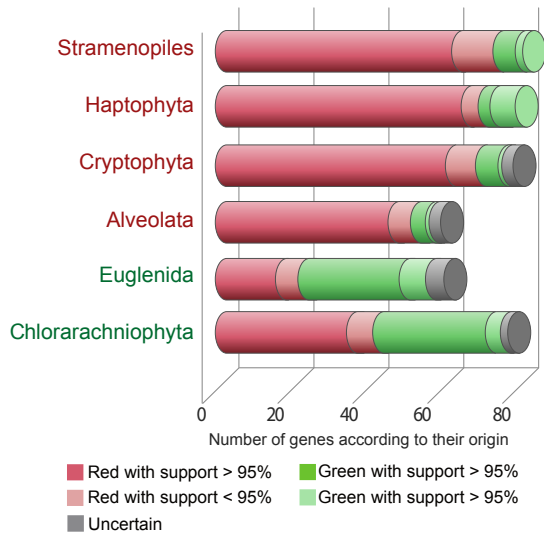
382 **Figure captions**

383 FIG 1. Genes of red and green algal ancestry in secondary photosynthetic eukaryotes. (A) Number
384 of red or green algal-like genes in each lineage among the 82 genes analyzed classified according to
385 their origin and statistical support in phylogenetic trees (supplementary figs. S1-S82,
386 Supplementary Material online). (B) Gene functions of the 'red' and 'green' genes detected in
387 transcriptomes and nuclear genomes of chlorarachniophytes and euglenids.

388

389 FIG 2. Examples of maximum likelihood phylogenetic trees of nucleus-encoded genes of red and
390 green algal origin in secondary photosynthetic eukaryotes. (A) Protein involved in photosystem II
391 assembly (inherited from green algae in euglenids and from a red lineage in chlorarachniophytes).
392 (B) Protein required for thylakoid membrane formation (inherited from green algae in
393 chlorarachniophytes and from a red lineage in euglenids). Bootstrap support values are indicated by
394 black (100%), dark grey (95-99%), and light grey (85-95%) circles. Scale bars indicate the number
395 of substitutions per site. Complete trees can be seen, respectively, in supplementary figs. S74 and
396 S62, Supplementary Material online.

A



B

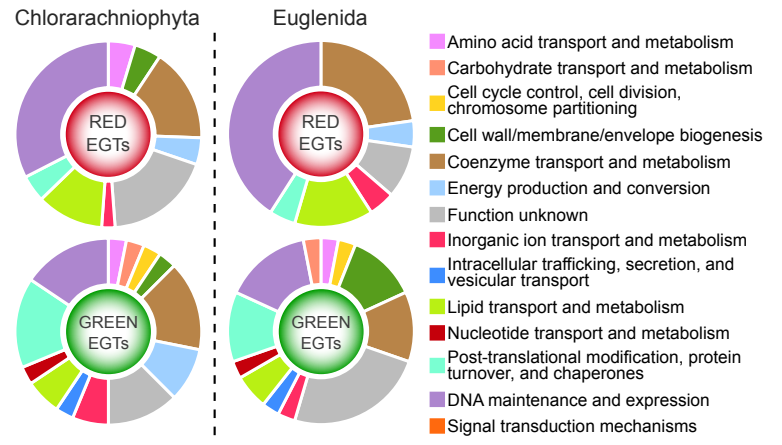


Figure 1



Figure 2