

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

Rafael I. Ponce-Toledo,¹ David Moreira,^{*,1} Purificación López-García,¹ and Philippe Deschamps^{*,1}

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

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

Associate editor: Iñaki Ruiz-Trillo

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 have been detected and subjected to various interpretations. In this context, accurate detection of endosymbiotic gene transfers 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, plastids.

Photosynthesis in eukaryotes takes place in a specialized compartment: the plastid. This organelle first evolved in a common ancestor of Archaeplastida (i.e., Viridiplantae + Rhodophyta + Glaucophyta) through the endosymbiosis of a cyanobacterium inside a eukaryotic host (Moreira and Philippe 2001; Archibald 2009; Keeling 2013). This primary endosymbiotic event entailed massive endosymbiotic gene transfer (EGT) from the cyanobacterial genome to the host nucleus (Weeden 1981; Kleine et al. 2009). Consequently, most proteins required for the proper functioning of primary plastids are encoded in the nuclear genome and addressed to the plastid lumen via specialized signal sequences and a translocation apparatus (Gutensohn et al. 2006). Other photosynthetic eukaryotic phyla obtained their plastids through secondary endosymbiosis, that is, the symbiosis of either green or red algae within another eukaryotic cell, or even through tertiary endosymbiosis (symbiosis of secondary photosynthetic eukaryotes within eukaryotic hosts) (Delwiche 1999; Archibald 2009; Keeling 2013). Euglenida (Excavata) and Chlorarachniophyta (Rhizaria) carry green algal secondary plastids ("green 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 ("red plastids") but so far it has been impossible to retrace a consensual evolutionary history (Lane and Archibald 2008; Archibald 2009; Keeling 2013). Whereas phylogenomic analyses of plastid-encoded genes support the monophyly of all CASH plastids, arguing for a single red algal

secondary endosymbiosis (Yoon et al. 2002; Muñoz-Gómez et al. 2017), most of the phylogenies based on host nuclear genes do not retrieve their monophyly (Baurain et al. 2010; Burki et al. 2016). To reconcile these incongruent results, some authors have proposed the hypothesis that a unique phylum (which may have gone extinct or evolved into one of the extant CASH phyla) acquired a red alga through secondary endosymbiosis and originated the first lineage of red secondary algae. Subsequently, this lineage would have transmitted the secondary red plastid to other CASH phyla via serial tertiary endosymbioses involving different hosts (Larkum et al. 2007; Sanchez-Puerta and Delwiche 2008; Boryl et al. 2009; Baurain et al. 2010; Petersen et al. 2014).

As for the primary endosymbiosis, each secondary or tertiary endosymbiosis was accompanied by numerous EGTs from the nucleus of the endosymbiotic red or green alga to the host nucleus. Consequently, 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 encoded in primary plastid genomes and the EGTs found in the genomes of Archaeplastida 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 found in nuclear genomes of secondary photosynthetic eukaryotes are expected to be useful to determine the red or green algal origin of their plastids. Compared with plastid-encoded genes, EGTs have the additional advantage that they can inform about the presence and identity of past plastids in lineages where plastids have been lost or replaced (cryptic plastid endosymbioses).

However, if EGTs are valuable to track contemporary and cryptic endosymbioses, their detection within whole nuclear genome sequences remains a complex task (Stiller 2011). In the case of primary endosymbiosis, EGT detection is rather straightforward because cyanobacterial-type genes are easily distinguishable from typical eukaryotic nuclear genes. The situation is more difficult in the case of secondary endosymbioses. Indeed, detection of EGT genes transferred from the nucleus of green or red algal endosymbionts can be ambiguous due to the poor resolution often found in single gene phylogenies that hampers distinguishing EGTs from vertically inherited nuclear genes, especially considering the short phylogenetic distance between Archaeplastida and several groups of secondary algae. Two studies on red-plastid-bearing algae, the chromerids (Alveolata) and the diatoms (Stramenopiles), illustrate this issue. Both reported an unexpected high number of genes phylogenetically related to green algal homologs. Whereas in the case of the chromerids the green signal was attributed to probable phylogenetic artifacts and the reduced sampling of red algal genome sequences (Woehle et al. 2011), it was interpreted in diatoms as evidence for a cryptic green algal endosymbiont (Moustafa et al. 2009). However, the subsequent reanalyses of the same genes using richer taxonomic sampling and more robust phylogenetic methods largely erased the evidence for cryptic green endosymbioses in these CASH phyla (Burki et al. 2012; Deschamps and Moreira 2012; Moreira and Deschamps 2014).

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

2011) but the increasing number of reported cases has prompted some authors to speculate on putative cryptic red algal endosymbioses in both lineages (Maruyama et al. 2011; Markunas and Triemer 2016). A systematic investigation of HGT/EGT is still missing in euglenids and chlorarachniophytes but, as mentioned earlier, in the context of secondary endosymbioses it can be difficult to distinguish among HGT, EGT, and just unresolved trees on the basis of single-gene phylogenies (Deschamps and Moreira 2012).

In this work, we have focused on a particular group of genes to reduce this uncertainty: genes transferred from the original cyanobacterial plastid endosymbiont into the nuclear genome of Archaeplastida and subsequently transferred from Archaeplastida into the genomes of complex secondary algae. In Archaeplastida, these genes are known to be involved in essential plastid functions and tend to be highly conserved (Reyes-Prieto et al. 2006; Deschamps and Moreira 2009), so we expected that they can 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 (supplementary table S1, Supplementary Material online). *Guillardia* and *Bigelowiella* proteins with hits in other photosynthetic eukaryotes and in cyanobacteria were selected for phylogenetic analysis. Maximum likelihood (ML) phylogenetic trees for these proteins were constructed and manually filtered to retain those fulfilling two criteria: 1) trees have to support a clear separation of Viridiplantae and Rhodophyta (with secondary lineages branching within them), and 2) proteins have to be shared by at least three secondary photosynthetic lineages. We identified in this way 82 genes most likely acquired by secondary photosynthetic eukaryotes from Archaeplastida. 70 were cyanobacterial genes likely transferred sequentially through primary and secondary endosymbioses, and 12 were derived from diverse bacterial groups likely transferred to a common ancestor of Archaeplastida and subsequently transferred to secondary photosynthetic groups (supplementary table S3 and figs. S1–S82, Supplementary Material online). Interestingly, most of these genes were absent in nonphotosynthetic eukaryotes, supporting that they were not misinterpreted vertically inherited ones.

Most of the 82 ML phylogenies were well resolved and enabled us to unambiguously determine, for each secondary lineage, whether the source of the gene was a green or a red alga. As expected, in the great majority of our trees (between 84% and 90%, fig. 1A) the genes of red-plastid-endowed CASH lineages derived from red algae (e.g., fig. 2A and B). Because of their secondary green plastids, we expected the opposite situation in chlorarachniophytes and euglenids, namely a majority of “green” genes. However, 42 of the 78 trees where chlorarachniophytes were present (54%, fig. 1A) supported a “red” origin of the corresponding genes (e.g., fig. 2A). Similarly, 22 of the 61 trees containing euglenids (36%, fig. 1A) also supported a “red” ancestry (e.g., fig. 2B). These surprisingly high values were in sharp contrast with the small number of

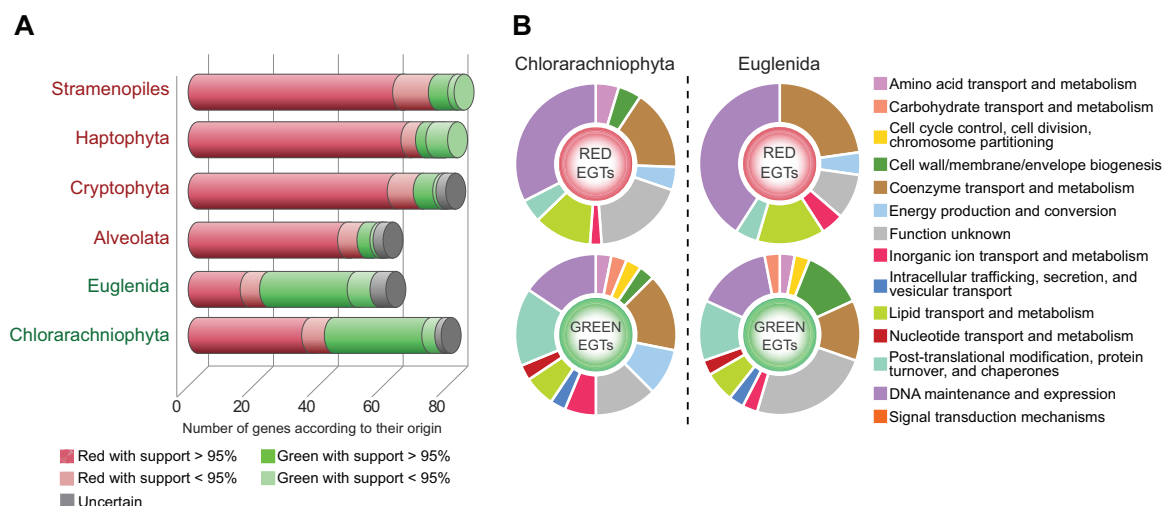


Fig. 1. Genes of red and green algal ancestry in secondary photosynthetic eukaryotes. (A) Number of red or green algal-like genes in each lineage among the 82 genes analyzed classified according to their origin and statistical support in phylogenetic trees ([supplementary figs. S1–S82](#),

trees (<10%, [fig. 1A](#)) showing CASH phyla embedded within green algae. Interestingly, the CASH phyla were monophyletic in seven of these trees, arguing for a common evolutionary origin of the corresponding “green” genes. Almost all of the 82 genes identified here encode plastid-targeted proteins involved in essential plastid functions ([fig. 1B](#) and [supplementary table S4, Supplementary Material](#) online). For instance, in both chlorarachniophytes and euglenids, these nuclear-encoded “red” genes participate in plastid genome expression (e.g., elongation factors and aminoacyl-tRNA synthetases), light harvesting, chlorophyll biosynthesis, and photosystem II assembly. Keeping these important genes implies a plastid-related selective pressure, which excludes that they could have accumulated in the heterotrophic ancestors of green secondary photosynthetic eukaryotes prior to plastid acquisition.

The marked disproportion of unexpected gene sources in green versus red secondary photosynthetic lineages is intriguing and may be interpreted in different ways. First, the green algal ancestors of chlorarachniophyte and euglenid plastids may have had a high proportion of red algal HGT genes in their genomes. However, such a high HGT proportion involving essential genes has not been reported so far in any green alga. Second, these “red” genes may have accumulated in chlorarachniophyte and euglenid nuclear genomes by numerous HGTs, for example, from food sources. This would imply that, for unknown reasons, HGT is much more frequent in secondary green lineages than in red ones, as well as a long-lasting feeding preference toward red prey in both secondary green lineages. Moreover, the “red” genes are shared by all the species of the relatively rich taxon sampling available for chlorarachniophytes ([fig. 2A](#)), indicating that their acquisition predated the diversification of this group and stopped afterward (we did not retrieve any tree supporting a recent HGT involving only a subgroup of chlorarachniophytes). Our data therefore argue for an ancient timing of “red” gene

acquisition. These observations may support a third interpretation: the “red” genes are shared by all SAR lineages (Stramenopiles, Alveolata, and Rhizaria) because they were acquired from a single common secondary red algal endosymbiosis ancestral to the whole SAR supergroup. 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. On the one hand, traces of past presence of red algal plastids, in the form of EGTs, in nonphotosynthetic SAR lineages are very often controversial ([Elias and Archibald 2009](#); [Stiller et al. 2009](#); [Stiller 2011](#)). On the other hand, plastid-bearing chlorarachniophytes constitute a relatively late-emerging branch within SAR ([Sierra et al. 2016](#)), implying that 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 this group of excavates has no close phylogenetic relationship 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 it difficult to infer the relative age of possible gene transfers. Nonetheless, “red” genes were often shared by several euglenids in our trees, suggesting a similar pattern of ancient acquisition as in chlorarachniophytes ([supplementary figs. S1–S82, Supplementary Material](#) online).

Our results show the presence of an unexpectedly high number of genes of red algal affinity in the two groups of eukaryotic algae with secondary green plastids, the euglenids and chlorarachniophytes, which is significantly higher than the frequency of “green” genes in algae with secondary red plastids, the CASH lineages. To address this question, we have focused on a subset of genes selected because of their strong phylogenetic signal and their implication in plastid-related



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 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 gray (95–99%), and light gray (85–95%) circles. Scale bars indicate the number of substitutions per site. Complete trees can be seen, respectively, in [supplementary figures S74 and S62, Supplementary Material online](#).

activities. It is therefore uncertain whether this conclusion can be applied to the rest of HGTs/EGTs potentially present in the genomes of all these algae. In fact, in addition to the problems inherent to the accurate detection of EGTs, our focus on these specific genes may explain, at least partly, the different results obtained in recent analyses of all potential EGTs in some CASH lineages, not only those of ultimate cyanobacterial origin (Dorrell et al. 2017).

However, we could not identify any particular bias in our gene selection process that could have artificially enriched the

observed “red” gene frequency in euglenids and chlorarachniophytes. Despite the methodological problems inherent to global genome analyses cited above, including a 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 the “red” genes in euglenids and chlorarachniophytes, either by cumulative HGT or by EGT from cryptic red algal endosymbionts,

remains mysterious but our work indicates that they were acquired early in both groups and that they fulfill essential functions for plastid activity and maintenance. Interestingly, indisputable evidence supports that in a third group of complex algae with green plastids, the dinoflagellate genus *Lepidodinium*, a former red plastid was replaced by the current green one, leading to a mosaic plastid proteome encoded by a mix of red and green algal genes (Minge et al. 2010), reminiscent of those 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 across the diversity of eukaryotes as secondary plastids (the “portable plastid” hypothesis; Grzebyk et al. 2003). It is thus tempting to speculate for euglenids and chlorarachniophytes a similar case as for *Lepidodinium*, with initial red plastids subsequently replaced by green ones. Even if this hypothesis turns out to be wrong and these cryptic red endosymbioses did not exist, the ancient acquisition by another mechanism of a significant number of red algal genes in both groups before their diversification and, especially, their maintenance in the contemporary species through millions of years of evolution, suggest that the “red” genes were instrumental in the establishment and maintenance of the secondary green plastids. Sequencing and analysis of additional genomes of euglenids, chlorarachniophytes, and their nonphotosynthetic relatives will help to refine the inventory of “red” genes in these lineages and their timing and, eventually, mechanism of acquisition.

Materials and Methods

Sequence Analysis

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

For these proteins, reciprocal BLASTp searches were done against the database to collect up to 600 similar sequences. We then aligned each set using Mafft v7.123b (Katoh and Standley 2013) with default parameters. Nonconserved alignment regions were trimmed with BMGE v1.0 (Criscuolo and Gribaldo 2010) with the BLOSUM62 matrix and allowing <50% gaps per position. Preliminary phylogenetic trees were inferred from trimmed alignments using FastTree v2.1.7 (Price et al. 2010) with default parameters. These trees were then manually inspected to identify those compatible with an EGT/HGT scenario. For all positive cases, only the

sequences corresponding to the portion of interest of each phylogenetic tree (the part showing the photosynthetic eukaryotes and the closest outgroup) were retained for the remaining steps. We then removed very short partial sequences and, to speed up subsequent calculations, several outgroup sequences from all alignments (see [supplementary table S2, Supplementary Material](#) online). The final sequence data sets were realigned and trimmed using TrimAL v1.4.rev15 with “gappy-out” parameter (Capella-Gutierrez et al. 2009). ML phylogenetic trees were inferred using IQtree v1.5.1 with the PMSF model of sequence evolution (Wang et al. 2018) parameterized using guided trees constructed with the LG+G+I model. Statistical support was calculated with 1000 ultrafast bootstrap replicates (Minh et al. 2013; Nguyen et al. 2015; Hoang et al. 2018).

Final selection of trees was done by manual inspection to keep those fulfilling the following two requirements: 1) the protein had to be shared by Cyanobacteria (or other bacteria), Archaeplastida and at least three secondary photosynthetic lineages, and 2) the corresponding phylogenetic trees had to support the clear separation of Viridiplantae and Rhodophyta (plus the lineages with secondary green and red plastids nested within them). Finally, the 82 trees passing this final filter ([supplementary figs. S1–S82, Supplementary Material](#) online) were inspected to infer the phylogenetic origin of the corresponding genes in the secondary photosynthetic lineages ([supplementary table S3, Supplementary Material](#) online).

Gene Functional Annotation

We annotated the functions of the 82 proteins from the final selection (see above) through the EggNOG 4.5 (Huerta-Cepas et al. 2016) web portal (<http://eggnogetdb.embl.de>; last accessed June 17, 2018). For each protein we used as queries the ortholog sequences of *Guillardia theta* and *Bigelowiella natans*. Functional annotations are shown in [supplementary table S4, Supplementary Material](#) online.

Data Availability

Protein sequence data sets used in this work are available for download at <http://www.esu-psud.fr/article950.html?lang=en>; last accessed June 17, 2018. They include nonaligned sequences and trimmed alignments.

Supplementary Material

[Supplementary data](#) are available at *Molecular Biology and Evolution* online.

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”). We thank the Associate Editor and two anonymous reviewers for constructive comments.

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(13):4962–4967.
- Archibald JM. 2009. The puzzle of plastid evolution. *Curr Biol*. 19(2):R81–R88.
- 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(13):7678–7683.
- Baurain D, Brinkmann H, Petersen J, Rodriguez-Espeleta 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(7):1698–1709.
- Bodýl A, Stiller JW, Mackiewicz P. 2009. Chromalveolate plastids: direct descent or multiple endosymbioses? *Trends Ecol Evol*. 24(3):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(6):626–635.
- Burki F, Kaplan M, Tikhonenkov DV, Zlatogursky V, Minh BQ, Radaykina LV, Smirnov A, Mlynikov 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):20152802.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973.
- Crisuolo 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(1):210–2148.
- 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(7427):59–65.
- Delwiche CF. 1999. Tracing the thread of plastid diversity through the tapestry of life. *Am Nat*. 154(S4):S164–S177.
- Deschamps P, Moreira D. 2009. Signal conflicts in the phylogeny of the primary photosynthetic eukaryotes. *Mol Biol Evol*. 26(12):2745–2753.
- Deschamps P, Moreira D. 2012. Reevaluating the green contribution to diatom genomes. *Genome Biol Evol*. 4(7):683–688.
- Dorrell RG, Gile G, McCallum G, Méheust R, Baptiste EP, Klinger CM, Brillet-Guéguen L, Freeman KD, Richter DJ, Bowler C. 2017. Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. *Elife* 6:e23717.
- Elias M, Archibald JM. 2009. Sizing up the genomic footprint of endosymbiosis. *Bioessays* 31(12):1273–1279.
- Grzebyk D, Schofield O, Vetriani C, Falkowski PG. 2003. The mesozoic radiation of eukaryotic algae: the portable plastid hypothesis. *J Phycol*. 39(2):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(3):333–347.
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 35(2):518–522.
- Hrdá Š, Fousek J, Szabová J, Hampl V, 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):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(D1):D286–D293.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.
- 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(5):268–275.
- Larkum AW, Lockhart PJ, Howe CJ. 2007. Shopping for plastids. *Trends Plant Sci*. 12(5):189–195.
- Leger MM, Eme L, Stairs CW, Roger AJ. 2018. Demystifying eukaryote lateral gene transfer. *Bioessays* 40(5):e1700242.
- Markunas CM, Triemer RE. 2016. Evolutionary history of the enzymes involved in the Calvin-Benson cycle in euglenids. *J Eukaryot Microbiol*. 63(3):326–339.
- Maruyama S, Suzuki T, Weber AP, Archibald JM, Nozaki H. 2011. Eukaryote-to-eukaryote gene transfer gives rise to genome mosaicism in euglenids. *BMC Evol Biol*. 11(1):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.
- Minh BQ, Nguyen MA, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol*. 30(5):1188–1195.
- 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(7):a016014.
- Moreira D, Philippe H. 2001. Sure facts and open questions about the origin and evolution of photosynthetic plastids. *Res Microbiol*. 152(9):771–780.
- Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D. 2009. Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* 324(5935):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(11):1677–1684.
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 32(1):268–274.
- 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(3):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(3):386–391.
- Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 5(3):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(23):2320–2325.
- 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(1):54–62.
- Sanchez-Puerta MV, Delwiche CF. 2008. A hypothesis for plastid evolution in chromalveolates. *J Phycol*. 44(5):1097–1107.
- Sierra R, Cañas-Duarte SJ, Burki F, Schwelm A, Fogelqvist J, Dixelius C, González-García LN, Gile GH, Slamovits CH, Klopp C, et al. 2016.

- Evolutionary origins of rhizarian parasites. *Mol Biol Evol.* 33(4): 980–983.
- Stiller JW. 2011. Experimental design and statistical rigor in phylogenomics of horizontal and endosymbiotic gene transfer. *BMC Evol Biol.* 11(1):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(4):581–590.
- Wang HC, Minh BQ, Susko E, Roger AJ. 2018. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst Biol.* 67(2):216–235.
- Weeden NF. 1981. Genetic and biochemical implications of the endosymbiotic origin of the chloroplast. *J Mol Evol.* 17(3):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(6):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(24): 15507–15512.