

Monophyly of Archaeplastida supergroup and relationships among its lineages in the light of phylogenetic and phylogenomic studies. Are we close to a consensus?

Paweł Mackiewicz*, Przemysław Gagat

Department of Genomics, Faculty of Biotechnology, University of Wrocław, Fryderyka Joliot-Curie 14a, 50-383 Wrocław, Poland

Abstract

One of the key evolutionary events on the scale of the biosphere was an endosymbiosis between a heterotrophic eukaryote and a cyanobacterium, resulting in a primary plastid. Such an organelle is characteristic of three eukaryotic lineages, glaucophytes, red algae and green plants. The three groups are usually united under the common name Archaeplastida or Plantae in modern taxonomic classifications, which indicates they are considered monophyletic. The methods generally used to verify this monophyly are phylogenetic analyses. In this article we review up-to-date results of such analyses and discussed their inconsistencies. Although phylogenies of plastid genes suggest a single primary endosymbiosis, which is assumed to mean a common origin of the Archaeplastida, different phylogenetic trees based on nuclear markers show monophyly, paraphyly, polyphyly or unresolved topologies of Archaeplastida hosts. The difficulties in reconstructing host cell relationships could result from stochastic and systematic biases in data sets, including different substitution rates and patterns, gene paralogy and horizontal/endosymbiotic gene transfer into eukaryotic lineages, which attract Archaeplastida in phylogenetic trees. Based on results to date, it is neither possible to confirm nor refute alternative evolutionary scenarios to a single primary endosymbiosis. Nevertheless, if trees supporting monophyly are considered, relationships inferred among Archaeplastida lineages can be discussed. Phylogenetic analyses based on nuclear genes clearly show the earlier divergence of glaucophytes from red algae and green plants. Plastid genes suggest a more complicated history, but at least some studies are congruent with this concept. Additional research involving more representatives of glaucophytes and many understudied lineages of Eukaryota can improve inferring phylogenetic relationships related to the Archaeplastida. In addition, alternative approaches not directly dependent on phylogenetic methods should be developed.

Keywords: Archaeplastida; Glaucophyta; monophyly; phylogenomics; phylogenetic analyses; primary plastid; Rhodophyta; Viridiplantae

Introduction

Two-membrane plastids are characteristic of three eukaryotic lineages, glaucophytes (Glaucophyta), red algae (Rhodophyta) and green plants (Viridiplantae or Chloroplastida) [1]. Glaucophytes comprise a small group of freshwater unicellular algae of only 13 species [2]. Rhodophytes constitute a large assemblage of both unicellular and multicellular algae, mainly sea inhabitants, with about 5000–6000 species [3]. The green clade is the largest, with 350 000 widespread species in both aquatic (mainly green algae) and terrestrial environments (land plants) [4]. It has highly diversified into unicellular, colonial and multicellular forms, including land plants with composite tissue organization. These three major lineages are commonly classified together in

the Archaeplastida or Plantae, one of several evolutionary supergroups of the domain Eukaryota [5–7].

There is a general consensus that Archaeplastida plastids are of prokaryotic origin, the result of endosymbiosis between a heterotrophic eukaryotic host and a photosynthetic cyanobacterium [8–10]. In support of this, Archaeplastida plastid genomes share similar gene contents and conserved gene arrangement with cyanobacterial genomes [11,12], including an unusual tRNA-Leu group I intron [13]. Glaucophyte plastids (called cyanelles) retain more cyanobacterial characteristics than do red algae and green plant plastids. The most remarkable is the presence of peptidoglycan, a typical component of bacterial cell walls located between the inner and outer membrane [14], as well as carboxysomes with enzymes involved in carbon fixation [15]. In addition, glaucophytes and rhodophytes still possess phycobilisomes, typical cyanobacterial photosystem II light-harvesting antennae [16].

* Corresponding author. Email: pamac@smorfland.uni.wroc.pl

Handling Editor: Andrzej Bodyl

There are many features, beside those mentioned above, suggesting a common origin of Archaeplastida plastids. For example, they all are bound by two membranes, corresponding to the outer and inner membranes of their cyanobacterial ancestor. This distinguishes them from complex plastids surrounded by three or four membranes present in many other photosynthetic (and also heterotrophic) eukaryotes. In contrast to the Archaeplastida plastids, complex plastids did not evolve directly from a cyanobacterium but arose by endosymbioses of green or red algae with other eukaryotes [8,9,17,18]. As a result, Archaeplastida plastids are called primary and the latter secondary or tertiary plastids. Moreover, primary plastids from all Archaeplastida representatives as well as their descendants (higher-order plastids) share the same unique *atpA* gene cluster in the genome [11] and translocase supercomplex (Toc/Tic apparatus) responsible for protein import of nuclear-encoded proteins. The complex is composed of a conserved set of cyanobacterium-derived homologs and subunits that are presumed to have arisen de novo in the common host [19–22]. The Archaeplastida also evolved a common mosaic feature of nuclear plastid-targeted genes of Calvin cycle enzymes [23]. At least three of the original cyanobacterial genes were replaced by host homologs in their common ancestor.

Transformation of an endosymbiont into a true organelle is considered to be a very complicated process requiring many modifications and inventions both in the host and endosymbiont [24,25]. Therefore, it generally is assumed that the primary endosymbiosis happened only once in the common ancestor of all Archaeplastida members [10,26,27]. Such a view requires monophyly of both primary plastids and their hosts, which is, however, still controversial because the clear similarity of plastids and their monophyly are also consistent with alternative scenarios for the evolution of the Archaeplastida supergroup [28–33] (Fig. 1). A good way to test the monophyly of Archaeplastida plastids and hosts is to carry out phylogenetic analyses of genes present in plastids, compared to those inferred from nuclear and mitochondrial genomes.

Testing the monophyly of Archaeplastida based on plastid genes

In support of a common origin of primary plastids, the first phylogenetic analyses involving plastid SSU (16S) rRNA genes grouped all representatives of Archaeplastida together [34–36]. In agreement with earlier assumptions based on structural similarities of plastids and cyanobacteria, the three eukaryotic lineages branched from the cyanobacterial clade in phylogenetic trees. These results were also strongly confirmed by analyses of LSU (23S) rRNA, tRNA genes (for example, see [37,38]) as well as numerous studies of single [39–41] and concatenated sets of plastid and/or nuclear-encoded plastid genes (see [42–46] and Tab. 1 for other references).

The most extensive phylogenetic studies of plastid evolution to date involved 191 genes with a total of more than 90 000 sites [47] and 75 genes with almost 40 000 sites [48]. The results agree with those from phylogenetic analyses of

21 complete genomes using correlation of compositional vectors calculated on frequency of amino acid strings [49]. In all these studies, representatives of the Glaucophyta, Rhodophyta and Viridiplantae clustered together with significant support. However, such a grouping does not exclude the possibility that each or at least two of these groups acquired their plastids independently from closely related cyanobacterial lineages that may now be extinct, or were passed from one Archaeplastida lineage to others via a secondary endosymbiosis (Fig. 1). This could lead to the same phylogenetic tree topology as in the case of a single plastid origin [28,32,33]. Finding a cyanobacterial lineage that clearly breaks up the monophyly of Archaeplastida plastids would refute the single plastid origin concept. So far, however, all Archaeplastida plastids cluster exclusive of cyanobacteria with significant support, even in protein [50] and RNA gene [38,51] phylogenies that are most rich in representatives (up to 127) of known cyanobacteria lineages.

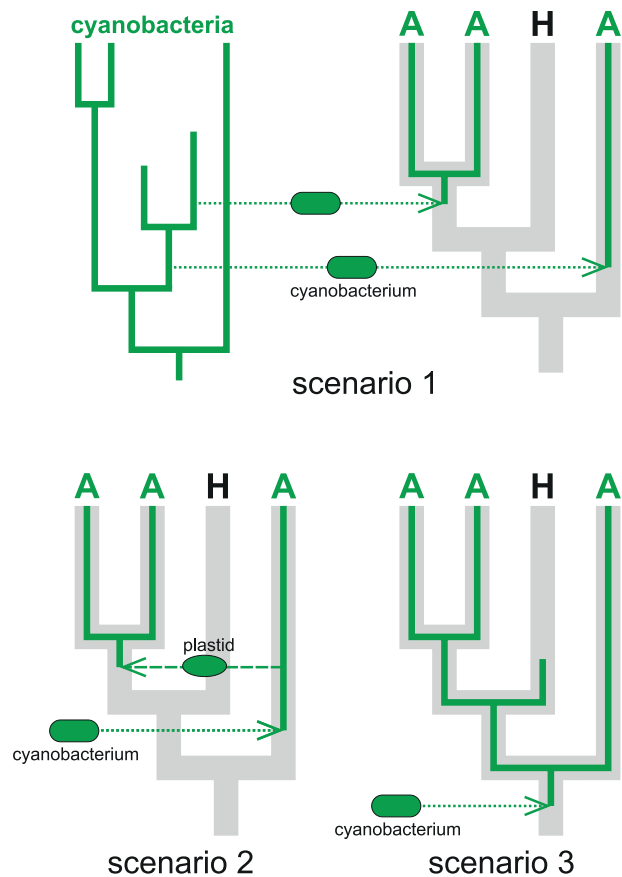


Fig. 1 Three alternative scenarios for primary plastid origin assuming a non-monophyly of Archaeplastida lineages (A) separated by a heterotrophic lineage (H). The first scenario assumes that independent primary endosymbioses involved closely related cyanobacteria and ancestors of separated Archaeplastida lineages. The cyanobacteria finally became extinct and the resulted plastids became similar by convergent evolution. The second hypothesis supposes that a single primary endosymbiosis occurred in one Archaeplastida lineage and next the primary plastid was transferred into the remaining lineages via secondary endosymbioses. The third scenario presumes an ancient primary endosymbiosis to have occurred very early in the eukaryote evolution and losses of primary plastids in lineages that are now heterotrophic.

Tab. 1 Characteristics of phylogenetic studies based on plastid genes and nuclear genes for plastid-related proteins (with “n” prefix) including three lineages of Archaeplastida. Additional notes were added only at the same references describing a different variant of studies.

Reference	Markers	Taxa	Sites	Data type	Tree topology
Adachi et al. 2000 [43]	45	10	9957	p	[G,[R,V]]
Bachvaroff et al. 2005 [183]	9	13	3452	p	[V,(G,R)]
Bachvaroff et al. 2005 [183]	9	13	6904	g	[V,[G,R]]
Bhattacharya and Medlin 1995 [54] ¹	1	40	1403	S	[G,(R,V)]
Bhattacharya and Medlin 1995 [54] ²	1	40	512	S	(V,(G,R))
Criscuolo and Gribaldo 2011 [47] ³	134	74	22 019	np	[G,(R,V)]
Criscuolo and Gribaldo 2011 [47]	134	74	22 019	np	[V,[G,R]]
Criscuolo and Gribaldo 2011 [47] ⁴	134	74	66 057	ng	[G,(R,V)]
Criscuolo and Gribaldo 2011 [47]	127	83	18 934	p	[R,(G,V)]
Criscuolo and Gribaldo 2011 [47] ³	127	83	18 934	p	[R,[G,V]]
Criscuolo and Gribaldo 2011 [47] ³	191	83	30 149	p+np	[R,[G,V]]
Criscuolo and Gribaldo 2011 [47]	191	83	30 149	p+np	[V,(G,R)]
Criscuolo and Gribaldo 2011 [47] ⁴	127	83	56 802	g	[R,(G,V)]
Criscuolo and Gribaldo 2011 [47] ⁴	191	83	90 447	g+ng	[R,(G,V)]
Cuvelier et al. 2010 [184]	22	23	4425	p	[G,(R,V)]
Deschamps and Moreira 2009 [46]	124	23	28 603	np	[V,[G,R]]
Deschamps and Moreira 2009 [46] ⁵	55	21	9177	p	[G,(R,V)]
Deschamps and Moreira 2009 [46] ³	55	21	9177	p	[V,(G,R)]
Deschamps and Moreira 2009 [46] ⁶	55	21	9177	p	[V,(G,R)]
Douglas and Turner 1991 [34] ⁷	1	18	862	S	[V,(G,R)]
Douglas and Turner 1991 [34]	1	18	1118	S	(V,(G,R))
Falcon et al. 2010 [167]	1	65	1255	S	[N,[G,(R,V)]]
Giovannoni et al. 1993 [35]	1	17	1184	S	[G,(R,V)]
Hagopian et al. 2004 [185]	17	25	3347	p	[G,(R,V)]
Hagopian et al. 2004 [185]	41	25	8810	p	[R,(G,V)]
Helmchen et al. 1995 [164]	1	27	1403	S	(G,(R,V))
Helmchen et al. 1995 [164] ⁸	1	27	1403	S	[V,[G,R]]
Janouskovec et al. 2010 [186]	34	24	7599	p	[R,[G,V]]
Janouskovec et al. 2010 [186]	68	24	15 736	p	[R,[G,V]]
Janouskovec et al. 2010 [186]	23	24	4438	p	[V,[G,R]]
Janouskovec et al. 2012 [187]	1	158	1273	S	[G,[R,V]]
Khan et al. 2007 [188] ³	45	17	9081	p	[G,(R,V)]
Khan et al. 2007 [188]	45	17	9081	p	[R,(G,V)]
Kim et al. 2011 [189]	4	38	4062	t+S+L	[V,(G,R)]
Le Corguille et al. 2009 [190]	44	20	8652	p	[R,[G,V]]
Li et al. 2014 [48] ⁶	75	42	12 956	p+np	[G,[R,V]]
Li et al. 2014 [48] ⁹	75	42	12 956	p+np	[G,(R,V)]
Li et al. 2014 [48] ⁵	75	42	12 956	p+np	[V,[G,R]]
Li et al. 2014 [48] ⁵	75	42	38 868	g+ng	[V,[G,R]]
Li et al. 2014 [48] ⁹	75	42	38 868	g+ng	[V,[G,R]]
Li et al. 2014 [48] ⁴	75	42	38 868	g+ng	[V,(G,R)]
Marin et al. 2005 [37]	4	60	4104	t+S+L	[V,[G,R]]
Marin et al. 2007 [51]	4	101	4126	t+S+L	[V,[G,R]]
Martin et al. 1998 [191]	45	10	11 039	p	G,[R,V]
Martin et al. 2002 [192]	41	16	8303	p	G,[R,V]
Nelissen et al. 1995 [193]	1	36	1277	S	[V,[G,R]]
Ochoa et al. 2014 [38] ¹⁰	2	93	935	S+L	(G,(R,V))
Ochoa et al. 2014 [38]	2	99	1029	S+L	[G,[R,V]]
Ochoa et al. 2014 [38] ¹¹	2	69	1029	S+L	[V,(G,R)]
Ochoa et al. 2014 [38]	18	38	2906	p	[V,[G,R]]
Ohta et al. 1997 [42]	7	10	NA	p	(R,(G,V))
Ohta et al. 2003 [194]	41	17	8308	p	(G,[R,V])
Olsen et al. 1994 [36]	1	253	NA	S	(G,(R,V))
Price et al. 2012 [20]	NA	61	17 049	p	[R,(G,V)]
Qui et al. 2012 [163]	30	48	7288	p	[G,[R,V]]
Reyes-Prieto and Bhattacharya 2007 [162]	19	16	5136	np	[G,[R,V]]
Reyes-Prieto et al. 2010 [133]	NA	NA	NA	S	[G,[R,V]]
Rodriguez-Ezpeleta et al. 2005 [45]	50	31	10 334	p	[G,(R,V)]

Tab. 1 (continued)

Reference	Markers	Taxa	Sites	Data type	Tree topology
Rogers et al. 2007 [195]	38	28	9108	p	[R,[G,V]]
Sanchez-Puerta et al. 2007 [170]	62	14	11 986	p	[G,[R,V]]
Sanchez-Puerta et al. 2007 [170]	62	14	23 972	g	[V,(G,R)]
Sato 2006 [196]	27	35	NA	p	[G,[R,V]]
Sato 2006 [196]	11	35	NA	p	[R,[G,V]]
Sato 2006 [196]	27	35	NA	g	[V,[G,R]]
Shih et al. 2013 [50]	25	164	NA	p	[G,[R,V]]
Tengs et al. 2000 [197]	1	72	1798	S	[G,(R,V)]
Turner et al. 1999 [161]	1	63	1377	S	[G,[R,V]]
Vogl et al. 2003 [174]	48	7	14 821	p	(R,(G,V))
Vogl et al. 2003 [174] ¹²	48	7	5404	p	(V,(G,R))
Wang et al. 2008 [198]	20	20	2160	np	[V,[G,R]]
Wang et al. 2008 [198]	10	22	4088	p	[V,[G,R]]
Yoon et al. 2004 [44]	6	46	7111	g+S	[G,[R,V]]
Yoon et al. 2004 [44]	5	46	1934	p	[G,(R,V)]
Yoon et al. 2005 [165]	5	47	1490	p	[R,[G,V]]
Yoon et al. 2009 [166]	2	31	NA	S+L	[R,[G,V]]

¹ Maximum likelihood method was used. ² Neighbor joining and logDet methods were used. ³ Amino acids were recoded. ⁴ Codons were degenerated. ⁵ Stationary composition model was used. ⁶ CAT model was used. ⁷ Alignment gaps were excluded. ⁸ Neighbor joining and maximum parsimony methods were used. ⁹ Nonstationary (tree-heterogeneous) composition model was used. ¹⁰ Data complexity was reduced. ¹¹ Compositional bias was reduced. ¹² Fast evolving sites were excluded. g – genes; n – nuclear; p – proteins; S – SSU rRNA gene; L – LSU rRNA gene; t- Ile-tRNA + Ala-tRNA genes; G – Glaucophyta; N – *Nephroselmis olivace*; R – Rhodophyta; V – Viridiplantae; Square brackets indicate a significant grouping with a posterior probability >0.9 or a bootstrapp support >90%.

Testing the monophyly of Archaeplastida based on nuclear genes

In contrast to the general concordance of phylogenetic trees based on plastid genetic markers, different nuclear gene data sets produced monophyly, paraphyly, polyphyly or unresolved topologies of Archaeplastida host lineages (Tab. 2). In total, 26 data sets we gathered favor glaucophyte, red alga and green plant monophyly, whereas 50 do not recover the monophyly of these groups. Because their sequences were available and useful as a marker across comparable studies [52,53], initial analyses of representatives of all three Archaeplastida lineages were based on nuclear SSU (18S) rRNA gene [54–56]. However, the trees obtained were inconclusive and had very poorly resolved deeper branches. The three Archaeplastida lineages did not group together and, surprisingly, glaucophytes clustered with cryptophytes, a group of algae bearing a secondary plastid of rhodophyte origin. The lack of Archaeplastida monophyly was also confirmed in subsequent studies based on the same marker, and joined with LSU (28S) rRNA gene, but including many more taxa, even 2551 species ([57,58]; see Tab. 2 for other examples). Interestingly, glaucophytes still clustered with cryptophytes in these trees [59,60] and other recently discovered related lineages, such as heterotrophic katablepharids and picozoans [61–63]; the latter were formerly called picobiliphytes and considered photosynthetic but available data indicate that they are heterotrophic [64,65]. Several later analyses recovered a Rhodophyta-Viridiplantae monophyly [58,62,66,67] and even a clade containing all three Archaeplastida lineages, but still mixed with cryptophytes [67], katablepharids,

picozoans and the newly discovered heterotrophic flagellate *Palpitomonas bilix* [68]. However, the relationships were poorly supported or, at best, had moderate support. A relatively high value (Bayesian PP = 1) was obtained only by Kim et al. [69] for the clade including glaucophytes with cryptophytes and katablepharids in a tree based on genes encoding large and small subunit rRNA.

The monophyly of Archaeplastida was recovered by initial global phylogenetic studies of eukaryotes based on several nuclear-encoded proteins [58,70,71], but not with very strong support (Tab. 2). The inclusion of 143 protein sequences with a total length of more than 30 000 sites [45] and the application of a more advanced site-heterogeneous mixed-model (CAT) [72], as well as other approaches, recovered monophyly of primary-plastid host lineage with significant support [46,73]. However, the analyses did not consider representatives of cryptophytes or their relatives, which grouped with Archaeplastida in rRNA gene trees. In fact, when these taxa and other eukaryotic lineages were taken into account, different results were obtained depending on the data set and method used (Tab. 2).

In some cases almost all deep branches, including Archaeplastida lineages, remained completely unresolved [74–76]. Interestingly, in analyses involving the largest number of taxa (451), major eukaryotic supergroups, including the Opisthokonta, Amoebozoa, Excavata and SAR (Stramenopiles, Alveolates and Rhizaria) obtained significant or some support, but not the Archaeplastida. As in rRNA gene phylogenies, they were mixed with cryptophytes, haptophytes and katablepharids [77]. Many other trees inconsistent with the monophyly of Archaeplastida hosts were also obtained

Tab. 2 Characteristics of phylogenetic studies based on nuclear genetic markers including three lineages of Archaeplastida. Additional notes were added only at the same references describing a different variant of studies.

Reference	Markers	Taxa	Sites	Data type	Tree topology
Baldauf et al. 2000 [71]*	4	61	1528	p	(G,(R,V))
Baldauf et al. 2000 [71]*	6	61	NA	p+S+L	(G,[R,V])
Baurain et al. 2010 [199]	108	57	15 392	p	[G,(R,(C,V)),(H,SA)]
Bhattacharya and Medlin 1995 [54]	1	47	1612	S	((C,G),(V,(R,(H+SAR))))
Bhattacharya and Weber 1997 [70]*	1	48	748	actin gene	(G,(R,V))
Bhattacharya et al. 1995 [55]	1	47	1621	S	((C,G),(V,(R,(H+SAR))))
Brown et al. 2013 [93]	159	68	43 615	p	[G,[H,T],[C,(P,(R,V))]]
Burki et al. 2007 [160]	123	49	29 908	p	(R,[G,V])
Burki et al. 2008 [85]	135	65	31 921	p	[G,(R,V)]
Burki et al. 2009 [86]	127	72	29 235	p	[R,(G,V)]
Burki et al. 2012 [81]	258	68	55 881	p	[[P,G],[R,V]]
Burki et al. 2012 [81] ¹	258	65	55 881	p	[G,[R,V]]
Cavalier-Smith 2002 [59]	1	53	NA	S	((((C,G),V),(R,(H,(E+SAR))))))
Cavalier-Smith 2004 [60]	1	193	1044	S	(C,G),others,R,others,V
Cavalier-Smith et al. 2014 [80] ²	173	73	45 194	p	[G,[R,V]]
Cavalier-Smith et al. 2014 [80] ³	173	98	45 194	p	(R,(G,V))
Cavalier-Smith et al. 2014 [80] ⁴	178	122	47 510	p	(R,(G,V))
Cavalier-Smith et al. 2014 [80]	192	75	52 824	p	[G,[C,K],[R,V]]
Cuvelier et al. 2008 [62]	1	124	NA	S	((SA,(R,V)),(G,(P,[C,K])))
Deschamps and Moreira 2009 [46]*	143	28	23 238	p	[G,[R,V]]
Hackett et al. 2007 [84] ^{4,5}	16	44	NA	p	[G,[R,V]]
Hackett et al. 2007 [84] ⁴	16	46	NA	p	[G,[R,V]]
Hackett et al. 2007 [84]	16	46	6735	p	[R,(G,V)]
Hampl et al. 2009 [79] ⁵	143	34	35 584	p	((G,R),(H,V))
Hampl et al. 2009 [79]	143	48	35 584	p	[G,[V,(H,R)]]
Hampl et al. 2009 [79] ⁶	143	49	35 584	p	[(G,R),(V,(C,H))]
Hampl et al. 2009 [79] ⁷	NA	47	NA	p	(G,(V,(H,R)))
Ishida et al. 2010 [63]	1	270	1020	S	((R,((G,(P,(C,K))),(H,(Ce+E))))),(SAR,(E,V)))
Kim et al. 2006 [69]	2	48	3287	S+L	[[G,[C,K],[R,V]]
Kim and Graham 2008 [78]	1	40	736	EF2 p	((G,(Am,E)),(Am,(Ri,(E[(C,(H,K)),[R,V]]))))
Kim and Graham 2008 [78]	6	37	2797	p	(G,[E,[R,(V,[C,(H,K)])]])
Minge et al. 2009 [200]	78	37	17 283	p	[(R,V),(G,(C,H))]
Moreira et al. 2000 [201]*	6	12	1938	p	(G,(R,V))
Moreira et al. 2007 [66]	2	91	2574	S+L	((R,V),(H,(Ce,(C,G))))
Nikolaev et al. 2004 [75]	2	121	1405	actin p+S	(R,V),others,(C,(Ce,G))
Nozaki et al. 2003 [83]*	4	53	1525	p	(R,(G,(SA,(E,V))))
Nozaki et al. 2003 [83]* ²	4	34	1525	p	(R,[G,[SA,(E,V)])]
Nozaki 2005 [29]*	4	36	1525	p	(R,[G,(SA,(E,V))])
Nozaki et al. 2007 [98] ⁵	19	31	5216	p	[R,(G,(V,(H,SA)))]
Nozaki et al. 2007 [98]	19	33	5216	p	[(E,R),((G,V),(H,SA))]
Nozaki et al. 2009 [31] ⁵	19	29	5216	p	[R,((G,[V,(H,SA)])))]
Nozaki et al. 2009 [31]	19	35	5216	p	[(E,R),((G,(V,(H,SA)))]
Nozaki et al. 2009 [31] ⁵	21	29	6048	p	[R,[G,[V,(H,SA)]]]
Nozaki et al. 2009 [31]	21	35	6048	p	[(E,R),(G,[V,(H,SA)])]
Okamoto and Inouye 2005 [61]	1	69	1109	S	((R,(Ce,(G,(C,K))),(V,(H,SAR))))
Okamoto et al. 2009 [159]	1	67	NA	Hsp90 p	(V,(G,R))
Parfrey et al. 2010 [77] ²	16	88	6578	p+S	(G,(R,(V,(C,H))))
Parfrey et al. 2010 [77]	16	451	6578	p+S	((G,V),(R,(H,(C,K))))
Patron et al. 2007 [202]	102	38	18 425	p	(R,(G,V))
Reeb et al. 2009 [74]	1	64	NA	EF2 p	[[G,Am],[SAR,[E,[(C,[H,K]),[R,V]]]]]
Reeb et al. 2009 [74]	3	108	1088	p	[R,(C,[H,K]),(G,(V,(E+SAR+T)))]
Reeb et al. 2009 [74]	6	75	2302	p	((G,R),(V,(E,SAR)))
Reeb et al. 2009 [74]	7	75	NA	p (+ EF2)	(G,(SAR,(E,((V,T),[R,(C,[H,K])]]))))
Reeb et al. 2009 [74]	5	112	4197	p+S+L	[[G,[K,C]],[[R,V],[E+SAR+T]]]
Rodriguez-Ezpeleta et al. 2005 [45]*	143	34	30 113	p	[R,(G,V)]
Rodriguez-Ezpeleta et al. 2007 [73]*	143	37	30 244	p	(R,(G,V))
Rodriguez-Ezpeleta et al. 2007 [172]* ⁵	143	51	31 604	p	(G,(R,V))
Rodriguez-Ezpeleta et al. 2007 [172]*	143	64	31 604	p	(R,(G,V))

Tab. 2 (continued)

Reference	Markers	Taxa	Sites	Data type	Tree topology
Tekle et al. 2007 [203]	4	74	2120	p+S	(G,[R,V])
Tekle et al. 2008 [204]	4	100	2291	p+S	[R,[G,[V,((C,[H,E]),[E,SAR])]]]
van de Peer and de Wachter 1997 [57]	1	500	NA	S	((R,(Am,(C,G))),((SAR,(H,V))))
van de Peer et al. 1996 [56]	1	107	NA	S	((SR,(R,(A,Am))),((H,(V,(C,G)))))
van de Peer et al. 2000 [58]*	4	61	1528	p	(G,(R,V))
van de Peer et al. 2000 [58]	1	2551	1574	S	((R,V),((Am,(C,G)),(H,SAR)))
Yabuki et al. 2010 [68]	5	18	4840	p+S+L	[R,[SAR,(G,(H,(V,[Pa,[C,K]]))]]]
Yabuki et al. 2010 [68]	6	18	5559	p+S+L	(G,[(H,[C,K]),(Pa,(R,V))])
Yabuki et al. 2010 [68]	1	67	1335	S	(V,(R,Pa,(P,(G,(C,K))))
Yabuki et al. 2014 [87] ⁸	157	64	41 372	p	((G,R),((H,V),(Pa,[C,K])))
Yabuki et al. 2014 [87] ⁹	157	64	41 372	p	[R,G,V]
Yoon et al. 2008 [76] ⁵	4	92	NA	p+S	[R,[G,((C,V),(E,(H,SAR)))]]
Yoon et al. 2008 [76]	4	105	NA	p+S	[(Ri,R),E,(G,[C,V]),(H,SAR)]
Yoon et al. 2011 [64]	7	78	2594	p	[R,[(V,(G,((H,[P,T]),(C,K)))),[E,SAR]]]
Zhao et al. 2012 [67] ¹⁰	124	77	27 638	p	[G,[R,V]]
Zhao et al. 2012 [67]	124	79	27 638	p	[G,[R,V]]
Zhao et al. 2012 [67]	2	80	NA	S+L	((R,V),(C,G))
Zhao et al. 2012 [67]	1	142	NA	S	((R,V),(Ce,(C,G)))

* Cryptophyta, Haptophyta or related taxa were not considered. ¹ Telonemid, centrohelid and picobiliphyte (picozoan) were excluded. ² Impact of missing data was decreased. ³ Impact of missing data for more taxa was decreased. ⁴ Fast evolving sites were excluded. ⁵ Long-branched taxa were excluded. ⁶ Cryptophyta was included. ⁷ Long-branched sequences were excluded. ⁸ Maximum likelihood method was used. ⁹ Bayesian method was used. ¹⁰ Two *Malawimonas* taxa were excluded. p – proteins; S – SSU rRNA gene; L – LSU rRNA gene; A – Alveolata; Am – Amoebozoa; P – Picozoa; C – Cryptophyta; Ce – Centrohelida; E – Excavata; G – Glaucophyta; H – Haptophyta; K – Katablepharida; Pa – *Palpitomonas*; R – Rhodophyta; Ri – Rhizaria; SA – Stramenopila + Alveolata; SAR – Stramenopila + Alveolata + Rhizaria; T – *Telonemia*; V – Viridiplantae. Square brackets indicate a significant grouping with a posterior probability >0.9 or a bootstrapp support >90%.

(indicating paraphyly or polyphyly) but with more significant support (Tab. 2). In one of them, glaucophytes formed a separate clade, whereas red algae and green plants grouped significantly with cryptophytes, haptophytes and katablepharids [78] (see also [74] for another interpretation). In other studies, a clade with the three Archaeplastida lineages was indeed recovered, but they still clustered significantly with haptophytes [79], cryptophytes and katablepharids [80], *Palpitomonas* [68], as well as with picozoans [81]. The latter study was based on the most sequence-rich data set to date, 258 nuclear-encoded proteins with 55 881 sites in total. In addition, other representatives of the main eukaryotic supergroups, such as Excavata and SAR, grouped with the Archaeplastida with high support in other phylogenetic analyses [29,31,64,68,74,76,82,83]. Such relationships were also inferred from many other data sets but with weaker or no support (Tab. 2). It should be emphasized, however, that many other phylogenies based on both smaller [84] and much larger data sets [67,80,81,85–87] have high or very high support values for monophyly of glaucophytes, red algae and green algae/land plants.

Testing the monophyly of Archaeplastida based on mitochondrial genes

Because of the inconsistencies in nuclear multigene phylogenies of main eukaryotic lineages, including glaucophytes,

red algae and green plants, mitochondrial genes have been explored as alternative phylogenetic markers to trace host cell history. Successful sequencing of glaucophyte mitochondrial genomes enabled testing the monophyly of Archaeplastida. Studies based on 42 mitochondrion-encoded and nucleus-encoded, mitochondrion-targeted proteins (11 384 sites) and 84 taxa (including two glaucophytes) supported the monophyly of the primary plastid-bearing eukaryotes with 0.99 and 1.0 posterior probabilities [88]. Inclusion of seven glaucophyte sequences confirmed this result for the concatenated amino acid alignment of 14 proteins encoded in mitochondrial genomes (3267 sites) from 49 taxa [89]. However, topology tests of the protein tree showed that placing red algae as a sister branch to haptophytes outside the Archaeplastida, as well as a topology uniting red algae and haptophytes within the Archaeplastida, were not statistically worse than the best tree. In addition to that, the Archaeplastida clade obtained very weak (51%) bootstrap support in maximum-likelihood analysis using the corresponding nucleotide alignment, whereas in the Bayesian tree the clade of Hacrobia (katablepharid, cryptophytes and haptophytes) branched within the Archaeplastida as sister to the red algae with 0.99 posterior probabilities for both the Archaeplastida + Hacrobia and the Hacrobia + red algae clades [89]. Interestingly, this topology corresponds to results obtained from rRNA genes and some nuclear protein data.

The results based on mitochondrial genes should be considered with caution because it was demonstrated that

the clustering of mitochondrial sequences from red algae and green plants can be artificial, resulting from similarly slow evolutionary rates of these sequences compared to many other eukaryotic taxa [90].

Influence of different methodologies on discrepancies in Archaeplastida phylogenies

The striking differences in inferring phylogenetic positions and relationships of Archaeplastida hosts could result from many methodological limitations. On one side there is stochastic error caused by poor phylogenetic signal in the data, which makes it difficult to resolve relationships, especially at deep levels; for example, among very early diverging major eukaryotic lineages [91]. One way to overcome this problem is to increase the number of genes analyzed and use phylogenomic approaches based on huge data sets on a genomic scale [92]. However, even consideration of more than 40 000 sites does not always result in deep robust phylogenies for the Archaeplastida [80,87,93]. Moreover, analyses of many genetic markers are vulnerable to systematic errors, which can lead to false phylogenetic reconstructions, reflecting unreal evolutionary relationships but with high statistical support.

The best-known systematic bias is long-branch attraction (LBA) [94] and the correlated effect that has been called short-branch exclusion (SBE) [95]. These biases cause, respectively, artificial clustering of highly diverged sequences, as well as of sequences with less than average substitution rates. The potential influence of LBA on the Archaeplastida was nicely shown by Rodriguez-Ezpeleta et al. [73]. The supergroup was moderately supported (64%) when two representatives of red algae, *Porphyra* and *Cyanidioschyzon*, were considered. Interestingly, the inclusion of only *Porphyra* increased the support for the whole Archaeplastida clade to 99%, whereas the tree with only *Cyanidioschyzon* moved this taxon to kinetoplastids with 100% support. Both kinetoplastids and *Cyanidioschyzon* are characterized by fast evolving sequences and their clustering is most consistent with LBA. In turn, the SBE effect concerns some sequences of glaucophytes [95] and could be responsible for grouping of short branches leading to glaucophytes and katablepharids in some tested tree topologies [81]. This effect could also explain common clustering of red algal and green plant in phylogenetic trees based on slowly evolving mitochondrial sequences [90].

In addition to that, gene-rich data sets can suffer from inadequate taxon sampling because it is not always possible to collect the full sets of sequences from all species. If, however, including many taxa is required to test their relationships, final alignments can lack many sequences from understudied species. Some of the discrepancies observed in above-described phylogenies likely result from such missing data [96]. For example, the exclusion of missing data and a more distant excavate outgroup changed the polyphyly of Archaeplastida (grouped with a katablepharid and a cryptophyte) to monophyly [80]. However, further removing missing data resulted in weak or lack of support for the clade. It clearly shows that topologies obtained are susceptible to

different data selection, and therefore, conclusions based on phylogenomic data should be drawn with caution.

It was shown that increased taxon-sampling could improve phylogenetic accuracy even at the cost of including data that are missing for some organisms [97]. Including new eukaryotic lineages, especially haptophytes, cryptophytes, katablepharids, picozoans and *Palpitomonas* [79–81] broke up the monophyly of Archaeplastida that had been obtained previously using less taxon-rich data sets [45,46,68]. The analyses based on the largest number of proteins (258 with 55 881 sites), represented by 68 species, grouped Archaeplastida lineages, but also included picozoans in this cluster, which significantly affiliated with glaucophytes [81]. As could be expected, exclusion of picozoans from the data recovered Archaeplastida as the monophyletic clade. The influence of taxon-sampling on Archaeplastida relationships with other eukaryotes was also shown by Jackson and Reyes-Prieto [89] using the concatenated alignment of mitochondrial proteins. Archaeplastida monophyly was obtained when all glaucophyte representatives were included. However, after removing glaucophyte taxa other than those previously published (i.e. *Cyanophora paradoxa* and *Glaucocystis nostochinearum*), two phylogenetic methods failed to recover the common origin of Archaeplastida lineages, because the glaucophytes grouped with the katablepharid *Leucocryptos marina* (significantly using the Bayesian approach), whereas red algae clustered with cryptophytes and haptophytes. The whole clade including Archaeplastida and the other eukaryotes obtained significant support by the Bayesian method. Interestingly, the authors recovered the monophyly of Archaeplastida only when excavate taxa were excluded. When they were present, they grouped with glaucophytes and green algae [89]. It indicates that the relationships obtained strongly depend on selection of taxa and sequences, and thus should be treated with caution.

Another problem with the multi-gene data is gene sampling. Its influence on the incongruity of Archaeplastida monophyly between results of Nozaki et al. [98] and Hackett et al. [84] was studied by Inagaki et al. [99], who evaluated the significance of Rhodophyta-Viridiplantae monophyly by analyzing multi-gene data sets of varied sizes [99]. The authors showed that recovery of this relationship depends on gene sampling in phylogenetic inferences with fewer than 10 000 amino acid positions. The sampled data sets consisted of different genes and supported various topologies but the tree based on the full multi-gene data set recovered the monophyly of red and green lineages. However, no representative of glaucophytes was considered in these studies.

Comparing data gathered from nuclear genes (Tab. 2) with more than two markers, we found that the monophyly of the Archaeplastida was, on average, obtained more often with greater numbers of genes (mean 103 vs. 52), longer alignments (26 694 vs. 14 624) but smaller numbers of taxa (59 vs. 69). The first two differences were statistically significant but not the last (Mann–Whitney test, *P*-value: 0.04, 0.04 and 0.75, respectively). These comparisons, however, should be considered with caution because the data compared are not completely independent.

Phylogenomic data sets consist of concatenated alignments of many genes, which most probably evolved subject to

different substitution rates and patterns. Therefore, application of improper, usually simple, evolutionary models can also lead to inaccurate phylogenies. Model violation can also result from saturation of mutations, across-site rate variation, heterogeneous substitutions across sites, rate variation across sites through time (heterotachy), site-interdependent substitutions as well as compositional heterogeneity and nonstationarity of substitutions. It was shown in the case of the Archaeplastida that model misspecification could reinforce systematic errors such as LBA of the red alga *Cyanidioschyzon* with non-Archaeplastida taxa, giving an incorrect topology even with a strong support [73].

Several approaches are applied to overcome or at least reduce the impacts of these errors; for example, removing rapidly evolving sites, genes and taxa [100], recoding amino acids into functional categories [101], and using a covarion [102] and site-heterogeneous mixture (CAT) models [72]. Although initial studies involving Archaeplastida were based on simpler phylogenetic models, many later analyses were usually carried out with advanced approaches using more realistic models. However, they still resulted in varied phylogenetic positions and relationships of the Archaeplastida. On one hand, the application of some of the modern methods increased support for the monophyly of Archaeplastida and alleviated the influence of some systematic errors [46,73,84] in comparison to earlier studies. On the other hand, Hampl et al. [79] obtained the non-monophyly of Archaeplastida (specifically placing of haptophytes within this supergroup) even after application of the CAT model and recoding of amino acids by functional categories, as well as progressive exclusion of rapidly evolving gene sequences and long-branch taxa. Therefore, the authors concluded that these relationships do not appear to be due to long-branch attraction artifacts. Similarly, Nozaki et al. used the CAT model and only slowly evolving genes and recovered polyphyly of Archaeplastida lineages [31]. Similarly, the application of a covarion model in rRNA gene-based phylogenies still grouped glaucophytes with cryptophytes but did not improve the resolution of Archaeplastida [67]. There also are differences depending on the phylogenetic methods used. Yabuki et al. [87] obtained significant for monophyly of the Archaeplastida in a Bayesian tree but not using maximum-likelihood. In contrast, Jackson and Reyes-Prieto [89], using mitochondrial genes, obtained some bootstrap support (although very weak) for Archaeplastida monophyly with maximum-likelihood but very strong confidence for separation of Archaeplastida lineages by katablepharid, cryptophytes and haptophytes using Bayesian inference.

To avoid problems with concatenated alignments, Chan et al. [103], used novel genomic data from two mesophilic red algae, *Porphyridium cruentum* and *Calliarthron tuberosum*, considering each gene phylogeny separately and found that about 50% of the examined protein phylogenies support the monophyly of red and green algae. However, to resolve monophyly of the whole Archaeplastida inclusion of complete glaucophyte genomes is necessary.

It seems that there is no clear interdependence between the tree topology obtained with respect to Archaeplastida monophyly, and the methodological approaches used in phylogenetic analyses.

Inconsistency of Archaeplastida phylogenies in the context of gene and plastid evolution

The disagreement between plastid gene trees supporting the monophyly of Archaeplastida plastids, and many phylogenies based on nuclear (and some mitochondrial) markers that often favor other topologies, is very intriguing. Besides methodological problems with phylogenetic analyses, differences in the inferred trees can result from disparate gene histories that are incongruent with species trees. One of them is gene paralogy [104,105]. When different paralogous gene families are retained across lineages, trees based on these genes will reproduce gene histories and duplication events but not accurate phylogenetic relationships among taxa. Although such genes should be excluded from phylogenetic data sets, it can be difficult to recognize hidden paralogs and separate them from real orthologs. The same problem concerns genes subjected to horizontal gene transfer (HGT), which seems to be a common process, not only in prokaryotic organisms, but also eukaryotes including Archaeplastida lineages [106–111]. Such genes are most likely also present in many alignments analyzed and could cause artificial clustering. One of the genes possibly subjected to this process is *EF2*, which encodes translation elongation factor. Phylogenies based on this marker gave a very strong signal supporting the non-monophyly of Archaeplastida: the significant grouping of rhodophytes and viridiplants with katablepharid, cryptophytes and haptophytes but excluding glaucophytes [78]. Therefore, it was proposed that the gene was likely horizontally transferred [74]. The close relationship of Archaeplastida with katablepharid, cryptophytes and haptophytes was also reproduced by some trees of mitochondrial genes. If these results are reliable (e.g. do not reflect a compositional bias) and we insist on Archaeplastida monophyly, we should also assume HGT at least for some of these genes from Archaeplastida to the other photosynthetic lineages. Widespread horizontal transfer of mitochondrial genes has been reported in plants [112–115].

Another process that can complicate inferred phylogenies of the Archaeplastida is endosymbiotic gene transfer (EGT). During this process, genes from an endosymbiont, including those from both plastid and nuclear genomes, are moved and integrated into the host nuclear genome [107,116–118]. The presence of these genes in an alignment would mask phylogenetic signal from true (endogenous) orthologs (Fig. 2). Interestingly, in many trees based on nuclear genes the primary plastid-containing lineages intermix with other eukaryotes, such as cryptophytes, haptophytes and picozoans, all of which contain genes most likely from EGT during secondary (or tertiary) endosymbioses involving the red algal lineage [8,9,17,18]. In agreement with that, Leigh et al. [119] found an influence of EGT in phylogenetic analyses on attraction of the red alga *Porphyra* to stramenopiles, which possess plastids derived from an ancient secondary endosymbiosis with a red alga (Fig. 2). The interfering effects of EGT could also explain the poor resolution of deep branches and a failure to recover the monophyly of Archaeplastida [77]. However, the supergroup remained polyphyletic in phylogenetic analyses after exclusion of cryptophyte and haptophyte taxa [77]. It should be noted that Archaeplastida

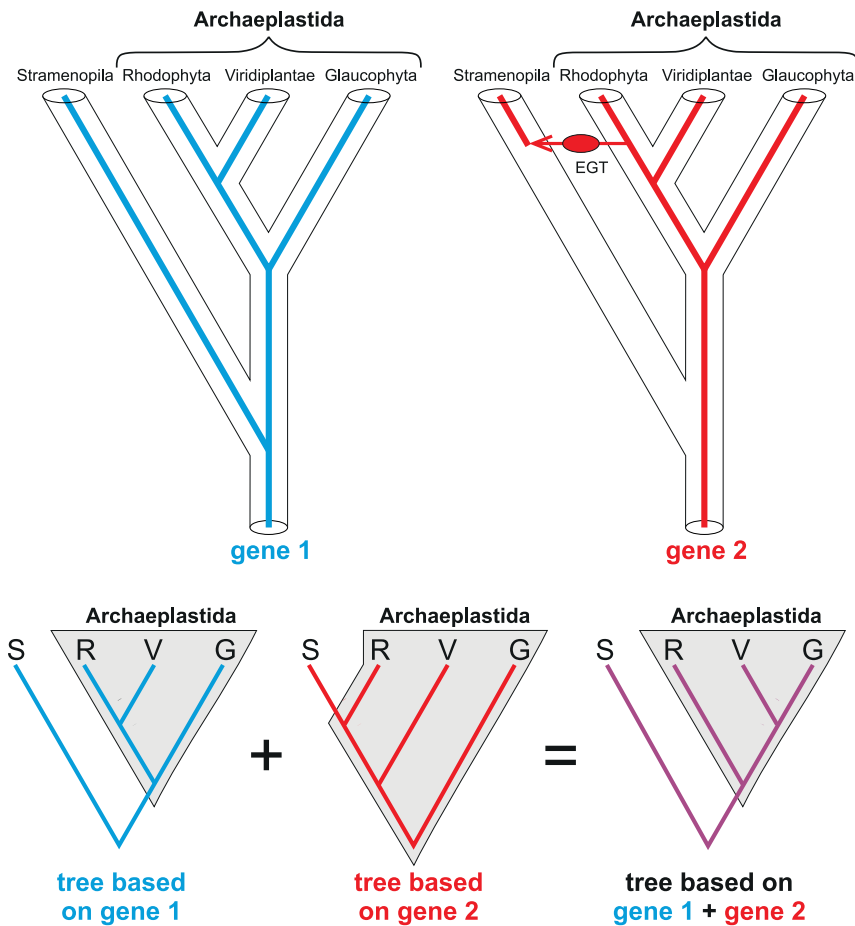


Fig. 2 Influence of endosymbiotic gene transfer (EGT) associated with a secondary endosymbiosis between a rhodophyte and a stramenopile on the inference of phylogenetic relationships between Archaeplastida and other eukaryotic lineages. The gene 1 is a true ortholog (vertically inherited) therefore the tree based on this gene reflects real host relationships, i.e. the monophyly of Archaeplastida. However, the tree based on the gene 2, which was acquired by the stramenopile from the red algae during the secondary endosymbiosis, shows a close affiliation of Stramenopila to Rhodophyta, thereby indicates the paraphyly of Archaeplastida. The simultaneous use of these genes in a tree reconstruction could mask the phylogenetic signal from the true ortholog leading to the attraction of rhodophytes to stramenopiles and its basal placement in respect to other Archaeplastida.

lineages also join other photosynthetic lineages in trees based on conserved nuclear rRNA genes, which are not considered to be involved in the EGT process. Inferences about EGT, especially in aplastidic eukaryotes, should be performed with caution and use appropriate controls; otherwise invoking it could lead to overestimation of EGT events and its impact on eukaryotic phylogenies [120].

It is remarkable that Archaeplastida members show affiliation to cryptophytes, haptophytes and related taxa in rRNA, nuclear and some mitochondrial gene trees. If these groupings are not a result of some compositional bias or another systematic error, but rather reflect true host cell history, it is worthwhile considering alternative views [77,79]. Three scenarios have been proposed (Fig. 1) that are compatible with the unquestionable monophyly of plastid genes, and explain both similar and different plastid features of Archaeplastida members (for wide discussion see [27–33,41,83,121,122]). One of them assumes that multiple independent primary endosymbioses occurred involving closely related cyanobacteria and separate Archaeplastida lineages. The resulting plastids were next subjected

to convergent evolution, whereas the intervening cyanobacterial lineages became extinct. The second possibility supposes that a single primary endosymbiosis occurred in one Archaeplastida lineage and the primary plastid was next transferred via secondary endosymbioses into the remaining lineages. The third hypothesis claims that an ancient primary endosymbiosis occurred very early in eukaryote evolution, before divergence of some major lineages, and was followed by subsequent losses of the primary plastids in ancestors of groups that now contain no primary plastid.

To accept or reject these scenarios we should determine if they are consistent with available data on key processes involved in plastid evolution: (i) multiple endosymbiotic events followed by transformation of the endosymbionts into true organelles and (ii) multiple plastid losses. The problem is still hotly debated [9,17,18,123]. Plastids in the reduced or vestigial form are still present in some eukaryotes that have lost photosynthesis, and are involved in vital non-photosynthetic functions (e.g. amino acid, heme, isoprenoid, and fatty acid biosynthesis); therefore, their loss seems unfavorable [124–127]. The transformation of endosymbiont to organelle

is usually considered a complicated process but successful endosymbiotic events might not be as rare as once assumed [128]. In addition to examples of higher-order endosymbiosis [8,9,17,18,129], there is also at least one case of a successful primary endosymbiosis independent of Archaeplastida, namely between a freshwater thecate amoeba, *Paulinella chromatophora*, and a cyanobacterium from a phylogenetic lineage different from the ancestor of Archaeplastida plastids [37,51,130]. Two photosynthetically active chromatophores retained by *Paulinella* fulfill all criteria to be considered true organelles [131]. Their deep integration involved a significant reduction of the endosymbiont's genome [132,133], transfer of many endosymbiont genes to the host nucleus [134,135] and evolution of a machinery for import of host-encoded proteins into the chromatophores [136–139]. Moreover, apart from *Paulinella*, there are many other eukaryotes in endosymbiotic relationships with cyanobacteria [140–143]. However, the first scenario assuming independent primary plastid endosymbioses in Archaeplastida lineages encounters a problem with an unparsimonious convergent evolution of several specific and complex plastidal features, e.g. Toc/Tic import apparatus [20] and composition of Calvin cycle enzymes [23].

At present, a single factor influencing inferred relationship of Archaeplastida with other eukaryotic supergroups in phylogenetic studies cannot be pinpointed and, therefore, be used to remove doubts about the monophyly or non-monophyly of glaucophytes, red algae and green algae/plants.

Testing the relationships among Archaeplastida lineages

If we assume the monophyly of Archaeplastida despite its controversies, we should consider relationships among its three lineages. Glaucophytes are usually considered to be the earliest branch of the Archaeplastida because their plastids retain more ancestral features typical of cyanobacteria than are present in red algae or green plants; for example, peptidoglycan and carboxysomes [14,15]. Other characteristics that distinguish the Glaucophyta from the Rhodophyta and Viridiplantae include (i) the presence of a cyanobacterial-like fructose-1,6-bisphosphate aldolase, which was replaced by a duplicated cytosolic copy in the case of the other Archaeplastida lineages [144], (ii) lack of triple-helix chlorophyll-binding, light harvesting antenna complexes [145,146] and (iii) lack of plastidal phosphate-translocator [20]. All these differences support the hypothesis that glaucophytes are sister to red algae and green plants.

On the other hand, glaucophytes and rhodophytes share the presence of unstacked thylakoid membranes with phycobilisomes also characteristic of cyanobacteria [16]. Moreover, these two Archaeplastida groups lack chlorophyll b present in Viridiplantae and both synthesize starch in the cytoplasm contrary to green plants, which relocated the pathway to plastids [147–150]. The close affiliation of these two groups was noticed by Cavalier-Smith who united them under the subkingdom Biliphyta, non-phagotrophic and phycobilisome-containing algae [26,151]. However, the features shared by Glaucophyta and Rhodophyta represent

characteristics of the cyanobacterial ancestor, and therefore should not be used as evidence for the monophyly of the Glaucophyta-Rhodophyta clade.

There is also evidence for the third possibility, i.e. a closer affiliation of glaucophytes and green plants. The Glaucophyta and Viridiplantae retain cyanobacterial RuBisCO genes; in contrast, red algae acquired RuBisCO from a proteobacterium via horizontal gene transfer [152]. Moreover, N-terminal sequences of two plastid proteins of photosystems I and II in the glaucophyte *Cyanophora* show significant similarity to those in green plants but not in red algae [153,154]. It was also found that the moss *Physcomitrella*, representing the “green” lineage, has nine homologous genes related to peptidoglycan biosynthesis that are essential for plastid division but are absent from red algae [155,156].

The first phylogenetic studies based on single genes that included all three representatives of the Archaeplastida were inconclusive and contradictory (see for example [39,40,157,158]), but subsequent analyses significantly resolved relationships within Archaeplastida (Tab. 1 and Tab. 2). In fact, all three possible topologies were proposed; however, phylogenetic analyses based on nuclear-encoded proteins (sometimes concatenated with rRNA genes) more often indicated the early-branching of the Glaucophyta (in 15 data sets) than the early divergence of the Rhodophyta (in 9 data sets) – Tab. 2. A basal Viridiplantae was proposed only from an Hsp90 tree, but without significant support [159]. What is more, the glaucophyte-first hypothesis was significantly supported by the larger number of studies, e.g. [46,67,84], whereas the rhodophyte-first concept obtained significant support only in one investigation, by Burki et al. [160], and with a simultaneous weak support for the whole Archaeplastida. In analyses by Rodriguez-Ezpeleta et al. [73], the glaucophytes + green clade received 90% support with 64% support for the whole Archaeplastida; however, when the highly diverged red alga *Cyanidioschyzon* was removed, and only *Porphyra* retained, these support values changed to 46% and 99%, respectively. It should be noticed that the earlier separation of glaucophytes was obtained using the data set with the largest number of alignment sites (55 881) among those tested [81]. Although, the earlier branching of the Rhodophyta was recovered using data sets with, on average, greater numbers of markers (128 vs. 83), alignment sites (29 885 vs. 22 748) and taxa (62 vs. 57), the differences were not statistically significant (Mann–Whitney or *t*-Student tests, $P > 0.12$).

Analyses of gains and losses of nuclear genes encoding transcription-associated proteins (TAPs, comprising transcription factors and other transcriptional regulators) were also used to assess support for the alternative branching orders in Archaeplastida [20]. The Viridiplantae-first scenario required 48 such events, the Rhodophyta-first scenario 46 events, whereas the Glaucophyta-first scenario 47 changes. However, when the pattern of TAP evolution typical of Viridiplantae (a greater probability of gains than losses) was taken into account, an earlier divergence of glaucophytes was favored. All these results based on nuclear markers tend to support the earlier separation of glaucophytes from the “red-green” lineage.

In contrast, phylogenies based on plastid genes and/or nuclear genes derived from a cyanobacterial endosymbiont are much less conclusive. Thirty data sets (15 significantly) supported the earlier divergence of glaucophytes (e.g. [44,161–163]), 27 (15 significantly) – green plants (e.g. [46,51,164]) and 17 (6 significantly) – rhodophytes (e.g. [47,165,166]) – Tab. 1. In one case involving SSU rRNA gene data, a fourth strange topology was obtained: the prasinophyte *Nephroselmis olivacea* took a significant basal position to glaucophytes and a clade of rhodophytes with the rest of green algae and land plants [167]. Considering cumulative results, support for the three possible topologies does not seem to depend on the type of data sets used, that is, plastid-encoded genes vs. nuclear-encoded genes or rRNA vs. protein genes. As for nuclear gene data, the rhodophyte-first concept was recovered by data sets with larger on average numbers of markers (74 vs. 52 and 48), alignment sites (21 972 vs. 14 715 and 12 301) and taxa (45 vs. 37 and 39) compared to topologies favoring the earliest separation of green plants and glaucophytes, respectively. The differences were, however, not statistically significant in Kruskal–Wallis test ($P > 0.28$). Accordingly, the most site-rich data set (more than 90 000 nucleotide sites) indicated the Rhodophyta-first hypothesis [47].

Other studies based on plastid genomes also produced inconsistent results. Comparisons of conserved gene arrangements showed either the earlier divergence of red algae [42] or glaucophytes [11], using 6 and 10 representatives of Archaeplastida, respectively. On the other hand, general comparisons of 50 plastid genomic sequences using BLAST results in distance phylogenies [168], as well as phylogenetic analysis of 21 complete plastid genomes using correlation analysis of compositional vectors calculated based on frequency of amino acid strings [49], recovered the Viridiplantae as the earliest-diverging lineage. In turn, studies of the presence and absence of 261–277 genes in 17–20 plastid genomes indicated the early-branching of Rhodophyta [169,170] or Viridiplantae [171]. Bayesian and maximum-likelihood methods applied to mitochondrial proteins showed the earlier divergence of glaucophytes with rather weak support [88,89], whereas the maximum-likelihood phylogenetic tree based on aligned nucleotides indicated the earlier separation of red algae [89]. Alternative topologies positioning rhodophytes or green plants as the earliest branch within the Archaeplastida were not rejected by the approximately unbiased test in the protein tree.

Reasons behind inconsistencies in inferring relationships between Archaeplastida lineages

Disagreements over the phylogeny of Archaeplastida may result from different methodological limitations of phylogenetic inferences and complex gene histories, especially those related to plastids as discussed above. The influence of fast-evolving sites was tested by Rodriguez-Ezpeleta et al. [172] and Hackett et al. [84]. Their trees based on complete alignments of nuclear-encoded proteins placed red algae at the basal position of the clade containing glaucophytes and green plants. However, the removal of fast-evolving sites

recovered glaucophytes as the basal clade to the sister-group of red algae and green plants with a stronger support.

An important process that may influence Archaeplastida phylogenies is heterotachy, or lineage-specific rate variations across sites within a protein through time. It results from structural and functional constraints acting on proteins [173]. Vogl et al. [174] showed that many plastid genes have significantly different phylogenies under standard substitution models and proposed that this incongruence could indeed result from heterotachy. In agreement with that, the application of heterotachy models found that many plastid genes of red and green plants are, in fact, subject to such evolution [175–177], and this process can cause errors in inference of phylogenetic trees leading to the LBA effect [178,179].

Deschamps and Moreira [46] studied conflicts in Archaeplastida relationships related to nuclear-encoded proteins in great detail. Following the findings of Leigh et al. [119], who noticed an attraction between rhodophytes and stramenopiles (containing red alga-derived plastid), Deschamps and Moreira assumed that the basal position of rhodophytes in relation to other Archaeplastida lineages [45,84,160,172] could result from endosymbiotic gene transfers linked to the secondary plastid endosymbiosis (Fig. 2). In agreement with that, the exclusion of eukaryotic representatives involved in such endosymbioses shifted the position of rhodophytes in phylogenetic trees from the base of the Archaeplastida clade to sister to the “green” lineage.

Comparable EGT-related effects could be responsible for the basal position of the Viridiplantae lineage, as its members were also engaged in secondary plastid endosymbioses. Such a position was obtained in trees using proteins of plastid origin encoded in nuclear genomes. However, the exclusion of fast-evolving proteins in the Viridiplantae did not change this result. Conserved plastid-encoded proteins also favored this topology, even under methods designed to decrease LBA effects and are more robust against compositional and evolutionary rate biases (i.e. amino acid recoding and a site-heterogeneous mixture CAT model). In contrast, a tree based on these proteins, but constructed using a simpler model that is more sensitive to systematic error, produced a topology with the basal placement of the Glaucophyta. This would seem to suggest that this topology could be artificial and result from (i) an attraction of glaucophyte sequences to a cyanobacterial outgroup due to a compositional bias, and (ii) the affiliation of “green” and “red” lineages because of a LBA artefact [46].

Criscuolo and Gribaldo [47] studying more plastid-encoded proteins at the amino acid, degenerate codon, and recoded amino acid levels obtained the earlier divergence of Rhodophyta. The same topology was favored using a joint set of cyanobacterial homologous plastid- and nuclear-encoded proteins for degenerated codons and recoded amino acids, but the Viridiplantae-first scenario was supported by the combined set and the nuclear-encoded proteins alone at the amino acid level. The early-branching Glaucophyta topology was in turn recovered from the nuclear-encoded proteins using degenerated codons and recoded amino acids. These results clearly indicate an influence of data types and methods on inferences of relationships within the Archaeplastida.

These discrepancies among Archaeplastida phylogenies may be caused by conflicts between protein-coding genes and their amino acid translations into proteins, which persists even when sophisticated and better-fitting models are used, i.e. the CAT model and the nonstationary (tree-heterogeneous) composition model [48]. Since convergent composition biases are induced at synonymous codon positions, analyses based on amino acids should be preferred. Such analyses using better-fitting models on amino acid alignments showed the early divergence of Glaucophyta, whereas the ancient (and probably artificial) separation of the Viridiplantae was recovered using standard homogeneous models at the amino acid level as well as standard and advanced models at the nucleotide level [48].

Another reason for the inconsistency in inferring relationships within Archaeplastida could be ancient gene paralogy [163]. Comparisons of the reduced genome of *Paulinella plastids* (acquired independently from those of the Archaeplastida via a primary cyanobacterial endosymbiosis) with free-living strains of the picocyanobacterium *Prochlorococcus* showed differential gene loss and concerted evolution among paralogous gene families. If similar events occurred after the primary endosymbiosis, they could mislead phylogenetic analyses based on plastid genes (Fig. 3). In fact, phylogenetic analyses of serially duplicated genes for photosystem II (*psbA*, *psbB*, *psbC* and *psbD*) indicated the basal emergence of Viridiplantae and a strong sisterhood of the Rhodophyta and Glaucophyta. In contrast, trees inferred from single-copy orthologs and anciently duplicated single-copy genes (*atpA*, *atpB*, *psaA* and *psaB*) favored the early-branching of glaucophytes, with significant support for a Rhodophyta-Viridiplantae clade. It should be noted that the tree built using the concatenated alignment of all these genes supported the Viridiplantae-first scenario, which suggests that phylogenetic signal from unrecognized paralogs dominated this data set [163].

A similar effect also could be caused by horizontal gene transfer into plastid genomes, which cannot be excluded. Although such transfers are less numerous compared to mitochondrial genomes [117], at least seven have been reported so far. The best known is the ribosomal protein *rpl36* gene, which was transferred from a proteobacterium or a planctomycete bacterium to plastid genomes of cryptophytes and haptophytes [180]. The other examples are genes encoding the large and small subunits of RuBisCO form I, acquired by the primary plastid genome of the common ancestor of red algae from a proteobacterium [152] (for additional cases, see [181,182] and references therein).

The difficulties in resolving relationships among the Archaeplastida lineages through phylogenetic methods suggest that early Archaeplastida members, just after the origin of primary plastids, underwent a rapid radiation including many molecular evolutionary processes, such as diversification (often increase) in substitution rate, and replacements, transfers, duplications and losses of genes. It seems that the much smaller subset of plastid genes is in general more influenced by these phenomena than nuclear genes, which quite unanimously indicate the earlier divergence of Glaucophyta from Rhodophyta and Viridiplantae lineages.

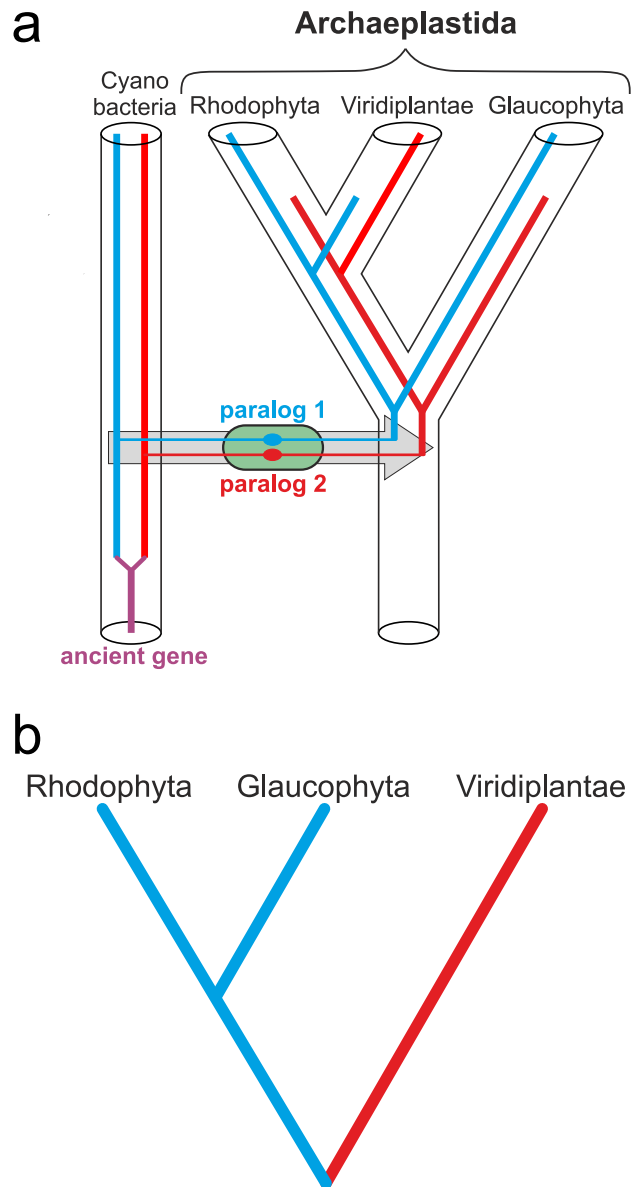


Fig. 3 Influence of differential loss of duplicated genes on the inference of phylogenetic relationships among Archaeplastida lineages. An ancient gene was duplicated in a cyanobacteria lineage into two paralogous genes, which were subsequently transferred via the primary endosymbiosis to the common ancestor of three Archaeplastida lineages (a). The genes were next subjected to differential losses. The paralog 1 stayed in rhodophytes and glaucophytes but was lost in viridiplants, which retained the paralog 2 that in turn decayed in rhodophytes and glaucophytes. This process misleads phylogenetic inference about the branching order of Archaeplastida lineages because the phylogenetic tree based on these genes indicates the earlier divergence of Viridiplantae (b) but, in fact, Glaucophyta separated as the first lineage (a). The Viridiplantae branch of the tree coalesces with Rhodophyta-Glaucophyta branch to the duplication event but not to the lineage separation.

Conclusions

Although it is commonly accepted that primary plastids of glaucophytes, red algae and green plants are of cyanobacterial origin, the monophyly of these groups is not unquestionably supported by all phylogenetic studies. Phylogenies based

on plastid genes clearly recover monophyly of the three Archaeplastida plastid lineages; however, trees based on nuclear markers, rRNA and protein coding genes, vary substantially and have yet to resolve clear evolutionary positions of Archaeplastida host lineages. When non-monophyly is recovered, the primary plastid-containing eukaryotes often cluster with cryptophytes, haptophytes and related taxa. The new results based on mitochondrial genes suggest the monophyly of glaucophytes, red algae and green plants but do not exclude the other possibilities. The discrepancies among tree topologies obtained could result from many methodological limitations, for example stochastic and systematic errors involving horizontal/endosymbiotic gene transfers and hidden paralogy. However, even if these problems are taken into account, it is difficult to definitively exclude alternative scenarios for the origin of primary plastids. Nevertheless, trees based on nuclear genes that support the

monophyly of Archaeplastida generally point to an earlier separation of glaucophytes in contrast to phylogenies based on mitochondrial and plastid genes. The smaller sets in the latter two cases appear more disturbed by complicated evolutionary phenomena. To conclusively solve the problem of the monophyly and the branching order of Archaeplastida plastids, more sequence data are required, especially those including glaucophytes and other poorly studied eukaryotic lineages. Giving the problems in phylogenetic reconstruction, other methods are desirable. EGT/HGT events can be identified by diverse composition-based methods. Other approaches alternative to phylogenetic analyses can include phylogenetic profiles, genomic contexts of studied genes, and presence or absence of genes in genomes. Analyses of rare mutation events called molecular signatures (e.g. conserved indels and gene fusion or splits) can also be helpful.

Acknowledgments

We are very grateful to the Reviewers for their excellent comments and insightful remarks that significantly improved the paper. This work was supported by Polish National Science Centre grant No. 2011/01/N/NZ8/00150 to P.G. and P.M.

Authors' contributions

The following declarations about authors' contributions to the research have been made: writing the manuscript and collecting data: PM, PG.

Competing interests

No competing interests have been declared.

References

- Delwiche CF, Timme RE. Plants. *Curr Biol*. 2011;21(11):R417–R422. <http://dx.doi.org/10.1016/j.cub.2011.04.021>
- Schenk HE. Glaucocystophytes. In: *Encyclopedia of life sciences*. Chichester: John Wiley & Sons; 2001. <http://dx.doi.org/10.1038/npg.els.0003061>
- Thomas DN. *Seaweeds*. London: Natural History Museum; 2002.
- Lewis LA, McCourt RM. Green algae and the origin of land plants. *Am J Bot*. 2004;91(10):1535–1556. <http://dx.doi.org/10.3732/ajb.91.10.1535>
- Adl SM, Simpson AGB, Farmer MA, Andersen RA, Anderson OR, Barta JR, et al. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol*. 2005;52(5):399–451. <http://dx.doi.org/10.1111/j.1550-7408.2005.00053.x>
- Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, Bowser SS, et al. The revised classification of eukaryotes. *J Eukaryot Microbiol*. 2012;59(5):429–514. <http://dx.doi.org/10.1111/j.1550-7408.2012.00644.x>
- Keeling PJ, Burger G, Durnford DG, Lang BF, Lee RW, Pearlman RE, et al. The tree of eukaryotes. *Trends Ecol Evol*. 2005;20(12):670–676. <http://dx.doi.org/10.1016/j.tree.2005.09.005>
- Gould SB, Waller RF, McFadden GI. Plastid evolution. *Ann Rev Plant Biol*. 2008;59(1):491–517. <http://dx.doi.org/10.1146/annurev.arplant.59.032607.029215>
- Keeling PJ. The endosymbiotic origin, diversification and fate of plastids. *Phil Trans R Soc Lond B*. 2010;365(1541):729–748. <http://dx.doi.org/10.1098/rstb.2009.0103>
- Reyes-Prieto A, Weber APM, Bhattacharya D. The origin and establishment of the plastid in algae and plants. *Annu Rev Genet*. 2007;41(1):147–168. <http://dx.doi.org/10.1146/annurev.genet.41.110306.130134>
- Stoebe B, Kowallik KV. Gene-cluster analysis in chloroplast genomics. *Trends Genet*. 1999;15(9):344–347. [http://dx.doi.org/10.1016/S0168-9525\(99\)01815-6](http://dx.doi.org/10.1016/S0168-9525(99)01815-6)
- Stoebe B, Martin W, Kowallik KV. Distribution and nomenclature of protein-coding genes in 12 sequenced chloroplast genomes. *Plant Mol Biol Rep*. 1998;16(3):243–255. <http://dx.doi.org/10.1023/A:1007568326120>
- Besendahl A, Qiu YL, Lee J, Palmer JD, Bhattacharya D. The cyanobacterial origin and vertical transmission of the plastid tRNA(Leu) group-I intron. *Curr Genet*. 2000;37(1):12–23.
- Pfanzagl B, Zenker A, Pittenauer E, Allmaier G, Martinez-Torrecuadrada J, Schmid ER, et al. Primary structure of cyanelle peptidoglycan of *Cyanophora paradoxa*: a prokaryotic cell wall as part of an organelle envelope. *J Bacteriol*. 1996;178(2):332–339.
- Burey SC, Fathi-Nejad S, Poroyko V, Steiner JM, Löffelhardt W, Bohnert HJ. The central body of the cyanelles of *Cyanophora paradoxa*: a eukaryotic carboxysome? *Can J Bot*. 2005;83(7):758–764. <http://dx.doi.org/10.1139/b05-060>
- Kies L, Kremer BP. Phylum Glaucocystophyta. In: Margulis L, editor. *Handbook of protocista*. Boston, MA: Jones and Bartlett Publishers; 1990. p. 152–166.
- Archibald JM. The puzzle of plastid evolution. *Curr Biol*. 2009;19(2):R81–R88. <http://dx.doi.org/10.1016/j.cub.2008.11.067>
- Bodył A, Stiller JW, Mackiewicz P. Chromalveolate plastids: direct descent or multiple endosymbioses? *Trends Ecol Evol*. 2009;24(3):119–121. <http://dx.doi.org/10.1016/j.tree.2008.11.003>
- Yusa F, Steiner JM, Löffelhardt W. Evolutionary conservation of dual Sec translocases in the cyanelles of *Cyanophora paradoxa*. *BMC Evol Biol*. 2008;8(1):304. <http://dx.doi.org/10.1186/1471-2148-8-304>
- Price DC, Chan CX, Yoon HS, Yang EC, Qiu H, Weber APM, et al. *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. *Science*. 2012;335(6070):843–847. <http://dx.doi.org/10.1126/science.1213561>
- Chan CX, Gross J, Yoon HS, Bhattacharya D. Plastid origin and evolution: new models provide insights into old problems. *Plant Physiol*. 2011;155(4):1552–1560. <http://dx.doi.org/10.1104/pp.111.173500>
- McFadden GI, van Dooren GG. Evolution: red algal genome affirms a common origin of all plastids. *Curr Biol*. 2004;14(13):R514–R516. <http://dx.doi.org/10.1016/j.cub.2004.06.041>
- Reyes-Prieto A, Bhattacharya D. Phylogeny of Calvin cycle enzymes supports Plantae monophyly. *Mol Phylogenet Evol*. 2007;45(1):384–391. <http://dx.doi.org/10.1016/j.ympev.2007.02.026>
- Cavalier-Smith T, Lee JJ. Protozoa as hosts for endosymbioses and the conversion of symbionts into organelles. *J Eukaryot Microbiol*. 1985;32(3):376–379. <http://dx.doi.org/10.1111/j.1550-7408.1985.tb04031.x>
- Cavalier-Smith T. Membrane heredity and early chloroplast evolution. *Trends Plant Sci*. 2000;5(4):174–182. [http://dx.doi.org/10.1016/S1360-1385\(00\)01598-3](http://dx.doi.org/10.1016/S1360-1385(00)01598-3)

26. Cavalier-Smith T. The origins of plastids. *Bot J Linn Soc.* 1982;17(3):289–306. <http://dx.doi.org/10.1111/j.1095-8312.1982.tb02023.x>
27. Palmer JD. The symbiotic birth and spread of plastids: how many times and whodunit? *J Phycol.* 2003;39(1):4–12. <http://dx.doi.org/10.1046/j.1529-8817.2003.02185.x>
28. Howe C, Barbrook A, Nisbet RE, Lockhart P, Larkum AW. The origin of plastids. *Philos Trans R Soc Lond B Biol Sci.* 2008;363(1504):2675–2685. <http://dx.doi.org/10.1098/rstb.2008.0050>
29. Nozaki H. A new scenario of plastid evolution: plastid primary endosymbiosis before the divergence of the “Plantae”, emended. *J Plant Res.* 2005;118(4):247–255. <http://dx.doi.org/10.1007/s10265-005-0219-1>
30. Larkum AWD, Lockhart PJ, Howe CJ. Shopping for plastids. *Trends Plant Sci.* 2007;12(5):189–195. <http://dx.doi.org/10.1016/j.tplants.2007.03.011>
31. Nozaki H, Maruyama S, Matsuzaki M, Nakada T, Kato S, Misawa K. Phylogenetic positions of Glaucophyta, green plants (Archaeplastida) and Haptophyta (Chromalveolata) as deduced from slowly evolving nuclear genes. *Mol Phylogenet Evol.* 2009;53(3):872–880. <http://dx.doi.org/10.1016/j.ympev.2009.08.015>
32. Stiller JW, Reel DC, Johnson JC. A single origin of plastids revisited: convergent evolution in organellar genome content. *J Phycol.* 2003;39(1):95–105. <http://dx.doi.org/10.1046/j.1529-8817.2003.02070.x>
33. Stiller JW, Hall BD. The origin of red algae: implications for plastid evolution. *Proc Natl Acad Sci USA.* 1997;94(9):4520–4525.
34. Douglas SE, Turner S. Molecular evidence for the origin of plastids from a cyanobacterium-like ancestor. *J Mol Evol.* 1991;33(3):267–273. <http://dx.doi.org/10.1007/BF02100678>
35. Giovannoni SJ, Wood N, Huss V. Molecular phylogeny of oxygenic cells and organelles based on small-subunit ribosomal RNA sequences. In: Lewin RA, editor. *Origins of plastids.* New York, NY: Chapman and Hall; 1993. p. 159–170. http://dx.doi.org/10.1007/978-1-4615-2818-0_10
36. Olsen GJ, Woese CR, Overbeek R. The winds of (evolutionary) change: breathing new life into microbiology. *J Bacteriol.* 1994;176(1):1–6.
37. Marin B, Nowack EC, Melkonian M. A plastid in the making: evidence for a second primary endosymbiosis. *Protist.* 2005;156(4):425–432. <http://dx.doi.org/10.1016/j.protis.2005.09.001>
38. Ochoa de Alda JAG, Esteban R, Diago ML, Houmar J. The plastid ancestor originated among one of the major cyanobacterial lineages. *Nat Commun.* 2014;5:4937. <http://dx.doi.org/10.1038/ncomms5937>
39. Delwiche C. Phylogenetic analysis of tufa sequences indicates a cyanobacterial origin of all plastids. *Mol Phylogenet Evol.* 1995;4(2):110–128. <http://dx.doi.org/10.1006/mpev.1995.1012>
40. Morden CW, Delwiche CF, Kuhnel M, Palmer JD. Gene phylogenies and the endosymbiotic origin of plastids. *Biosystems.* 1992;28(1–3):75–90. [http://dx.doi.org/10.1016/0303-2647\(92\)90010-V](http://dx.doi.org/10.1016/0303-2647(92)90010-V)
41. Palmer JD, Delwiche CF. The origin and evolution of plastids and their genomes. In: Soltis DE, Soltis PS, Doyle JJ, editors. *Molecular systematics of plants II.* Boston, MA: Kluwer Academic Publishers; 1998. p. 375–409. http://dx.doi.org/10.1007/978-1-4615-5419-6_13
42. Ohta N, Sato N, Nozaki H, Kuroiwa T. Analysis of the cluster of ribosomal protein genes in the plastid genome of a unicellular red alga *Cyanidioschyzon merolae*: translocation of the str cluster as an early event in the rhodophyte-chromophyte lineage of plastid evolution. *J Mol Evol.* 1997;45(6):688–695. <http://dx.doi.org/10.1007/PL00006273>
43. Adachi J, Waddell PJ, Martin W, Hasegawa M. Plastid genome phylogeny and a model of amino acid substitution for proteins encoded by chloroplast DNA. *J Mol Evol.* 2000;50(4):348–358.
44. Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. A molecular timeline for the origin of photosynthetic eukaryotes. *Mol Biol Evol.* 2004;21(5):809–818. <http://dx.doi.org/10.1093/molbev/msh075>
45. Rodríguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Löffelhardt W, et al. Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr Biol.* 2005;15(14):1325–1330. <http://dx.doi.org/10.1016/j.cub.2005.06.040>
46. Deschamps P, Moreira D. Signal conflicts in the phylogeny of the primary photosynthetic eukaryotes. *Mol Biol Evol.* 2009;26(12):2745–2753. <http://dx.doi.org/10.1093/molbev/msp189>
47. Criscuolo A, Gribaldo S. Large-scale phylogenomic analyses indicate a deep origin of primary plastids within cyanobacteria. *Mol Biol Evol.* 2011;28(11):3019–3032. <http://dx.doi.org/10.1093/molbev/msr108>
48. Li B, Lopes JS, Foster PG, Embley TM, Cox CJ. Compositional biases among synonymous substitutions cause conflict between gene and protein trees for plastid origins. *Mol Biol Evol.* 2014;31(7):1697–1709. <http://dx.doi.org/10.1093/molbev/msu105>
49. Chu KH, Qi J, Yu ZG, Anh V. Origin and phylogeny of chloroplasts revealed by a simple correlation analysis of complete genomes. *Mol Biol Evol.* 2003;21(1):200–206. <http://dx.doi.org/10.1093/molbev/msh002>
50. Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, et al. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci USA.* 2013;110(3):1053–1058. <http://dx.doi.org/10.1073/pnas.1217107110>
51. Marin B, Nowack EC, Glöckner G, Melkonian M. The ancestor of the *Paulinella chromatophore* obtained a carboxysomal operon by horizontal gene transfer from a *Nitrococcus*-like γ -proteobacterium. *BMC Evol Biol.* 2007;7(1):85. <http://dx.doi.org/10.1186/1471-2148-7-85>
52. Sogin ML. The phylogenetic significance of sequence diversity and length variations in eukaryotic small subunit ribosomal RNA coding regions. In: Warren L, Koprowski H, editors. *New perspectives on evolution.* New York, NY: Wiley-Liss; 1991. p. 175–188. (Wistar symposium series; vol 4).
53. Sogin ML, Elwood HJ, Gunderson JH. Evolutionary diversity of eukaryotic small-subunit rRNA genes. *Proc Natl Acad Sci USA.* 1986;83(5):1383–1387.
54. Bhattacharya D, Medlin L. The phylogeny of plastids: a review based on comparisons of small-subunit ribosomal RNA coding regions. *J Phycol.* 1995;31(4):489–498. <http://dx.doi.org/10.1111/j.1529-8817.1995.tb02542.x>
55. Bhattacharya D, Helmchen T, Bibeau C, Melkonian M. Comparisons of nuclear-encoded small-subunit ribosomal RNAs reveal the evolutionary position of the Glaucocystophyta. *Mol Biol Evol.* 1995;12(3):415–420.
56. van de Peer Y, Rensing SA, Maier UG, de Wachter R. Substitution rate calibration of small subunit ribosomal RNA identifies chlorarachniophyte endosymbionts as remnants of green algae. *Proc Natl Acad Sci USA.* 1996;93(15):7732–7736.
57. van de Peer Y, de Wachter R. Evolutionary relationships among the eukaryotic crown taxa taking into account site-to-site rate variation in 18S rRNA. *J Mol Evol.* 1997;45(6):619–630.
58. van de Peer Y, Baldauf SL, Doolittle WF, Meyer A. An updated and comprehensive rRNA phylogeny of (crown) eukaryotes based on rate-calibrated evolutionary distances. *J Mol Evol.* 2000;51(6):565–576.
59. Cavalier-Smith T. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int J Syst Evol Microbiol.* 2002;52(2):297–354.
60. Cavalier-Smith T. Only six kingdoms of life. *Proc Biol Sci.* 2004;271(1545):1251–1262. <http://dx.doi.org/10.1098/rspb.2004.2705>
61. Okamoto N, Inoué I. The katablepharids are a distant sister group of the Cryptophyta: a proposal for Katablepharidophyta *divisio nova*/katablepharida phylum *novum* based on SSU rDNA and beta-tubulin phylogeny. *Protist.* 2005;156(2):163–179. <http://dx.doi.org/10.1016/j.protis.2004.12.003>
62. Cuvelier ML, Ortiz A, Kim E, Moehlig H, Richardson DE, Heidelberg JF, et al. Widespread distribution of a unique marine protistan lineage. *Environ Microbiol.* 2008;10(6):1621–1634. <http://dx.doi.org/10.1111/j.1462-2920.2008.01580.x>
63. Ishida K, Inagaki Y, Sakaguchi M, Oiwa A, Kai A, Suzuki M, et al. Comprehensive SSU rRNA phylogeny of eukaryota. *Endocytobiosis Cell Res.* 2010;20:81–88.
64. Yoon HS, Price DC, Stepanauskas R, Rajah VD, Sieracki ME, Wilson WH, et al. Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science.* 2011;332(6030):714–717. <http://dx.doi.org/10.1126/science.1203163>
65. Seenivasan R, Sausen N, Medlin LK, Melkonian M. *Picomonas judraskeda* gen. et sp. nov.: the first identified member of the Picozoa

- phylum nov., a widespread group of Picoeukaryotes, formerly known as “picobiliphytes”. *PLoS ONE*. 2013;8(3):e59565. <http://dx.doi.org/10.1371/journal.pone.0059565>
66. Moreira D, von der Heyden S, Bass D, López-García P, Chao E, Cavalier-Smith T. Global eukaryote phylogeny: combined small- and large-subunit ribosomal DNA trees support monophyly of Rhizaria, Retaria and Excavata. *Mol Phylogenet Evol*. 2007;44(1):255–266. <http://dx.doi.org/10.1016/j.ympev.2006.11.001>
 67. Zhao S, Burki F, Brate J, Keeling PJ, Klaveness D, Shalchian-Tabrizi K. *Colodictyon* – an ancient lineage in the tree of eukaryotes. *Mol Biol Evol*. 2012;29(6):1557–1568. <http://dx.doi.org/10.1093/molbev/mss001>
 68. Yabuki A, Inagaki Y, Ishida K. *Palpitomonas bilix* gen. et sp. nov.: a novel deep-branching heterotroph possibly related to Archaeplastida or Hacrobia. *Protist*. 2010;161(4):523–538. <http://dx.doi.org/10.1016/j.protis.2010.03.001>
 69. Kim E, Simpson AGB, Graham LE. Evolutionary relationships of apusomonads inferred from taxon-rich analyses of 6 nuclear encoded genes. *Mol Biol Evol*. 2006;23(12):2455–2466. <http://dx.doi.org/10.1093/molbev/msl120>
 70. Bhattacharya D, Weber K. The actin gene of the glaucocystophyte *Cyanophora paradoxa*: analysis of the coding region and introns, and an actin phylogeny of eukaryotes. *Curr Genet*. 1997;31(5):439–446.
 71. Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science*. 2000;290(5493):972–977. <http://dx.doi.org/10.1126/science.290.5493.972>
 72. Lartillot N. A bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol*. 2004;21(6):1095–1109. <http://dx.doi.org/10.1093/molbev/msh112>
 73. Rodríguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang BF, Philippe H. Detecting and overcoming systematic errors in genome-scale phylogenies. *Syst Biol*. 2007;56(3):389–399. <http://dx.doi.org/10.1080/10635150701397643>
 74. Reeb VC, Peglar MT, Yoon HS, Bai JR, Wu M, Shiu P, et al. Interrelationships of chromalveolates within a broadly sampled tree of photosynthetic protists. *Mol Phylogenet Evol*. 2009;53(1):202–211. <http://dx.doi.org/10.1016/j.ympev.2009.04.012>
 75. Nikolaev SI, Berney C, Fahrni JF, Bolivar I, Polet S, Mylnikov AP, et al. The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. *Proc Natl Acad Sci USA*. 2004;101(21):8066–8071. <http://dx.doi.org/10.1073/pnas.0308602101>
 76. Yoon HS, Grant J, Tekle YI, Wu M, Chaon BC, Cole JC, et al. Broadly sampled multigene trees of eukaryotes. *BMC Evol Biol*. 2008;8(1):14. <http://dx.doi.org/10.1186/1471-2148-8-14>
 77. Parfrey LW, Grant J, Tekle YI, Lasek-Nesselquist E, Morrison HG, Sogin ML, et al. Broadly sampled multigene analyses yield a well-resolved eukaryotic tree of life. *Syst Biol*. 2010;59(5):518–533. <http://dx.doi.org/10.1093/sysbio/syq037>
 78. Kim E, Graham LE. EEF2 analysis challenges the monophyly of Archaeplastida and Chromalveolata. *PLoS ONE*. 2008;3(7):e2621. <http://dx.doi.org/10.1371/journal.pone.0002621>
 79. Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AGB, et al. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic “supergroups”. *Proc Natl Acad Sci USA*. 2009;106(10):3859–3864. <http://dx.doi.org/10.1073/pnas.0807880106>
 80. Cavalier-Smith T, Chao EE, Snell EA, Berney C, Fiore-Donno AM, Lewis R. Multigene eukaryote phylogeny reveals the likely protozoan ancestors of opisthokonts (animals, fungi, choanozoans) and Amoebozoa. *Mol Phylogenet Evol*. 2014;81:71–85. <http://dx.doi.org/10.1016/j.ympev.2014.08.012>
 81. Burki F, Okamoto N, Pombert JF, Keeling PJ. The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. *Proc Biol Sci*. 2012;279(1736):2246–2254. <http://dx.doi.org/10.1098/rspb.2011.2301>
 82. Tekle YI, Grant J, Anderson OR, Nerad TA, Cole JC, Patterson DJ, et al. Phylogenetic placement of diverse amoebae inferred from multigene analyses and assessment of clade stability within “Amoebozoa” upon removal of varying rate classes of SSU-rDNA. *Mol Phylogenet Evol*. 2008;47(1):339–352. <http://dx.doi.org/10.1016/j.ympev.2007.11.015>
 83. Nozaki H, Matsuzaki M, Takahara M, Misumi O, Kuroiwa H, Hasegawa M, et al. The phylogenetic position of red algae revealed by multiple nuclear genes from mitochondria-containing eukaryotes and an alternative hypothesis on the origin of plastids. *J Mol Evol*. 2003;56(4):485–497. <http://dx.doi.org/10.1007/s00239-002-2419-9>
 84. Hackett JD, Yoon HS, Li S, Reyes-Prieto A, Rummele SE, Bhattacharya D. Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates. *Mol Biol Evol*. 2007;24(8):1702–1713. <http://dx.doi.org/10.1093/molbev/mms089>
 85. Burki F, Shalchian-Tabrizi K, Pawlowski J. Phylogenomics reveals a new “megagroup” including most photosynthetic eukaryotes. *Biol Lett*. 2008;4(4):366–369. <http://dx.doi.org/10.1098/rsbl.2008.0224>
 86. Burki F, Inagaki Y, Brate J, Archibald JM, Keeling PJ, Cavalier-Smith T, et al. Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, telonemia and centroheliozoa, are related to photosynthetic chromalveolates. *Genome Biol Evol*. 2009;1:231–238. <http://dx.doi.org/10.1093/gbe/evp022>
 87. Yabuki A, Kamikawa R, Ishikawa SA, Kolisko M, Kim E, Tanabe AS, et al. *Palpitomonas bilix* represents a basal cryptist lineage: insight into the character evolution in Cryptista. *Sci Rep*. 2014;4:4641. <http://dx.doi.org/10.1038/srep04641>
 88. Zhao S, Shalchian-Tabrizi K, Klaveness D. Sulcozoa revealed as a paraphyletic group in mitochondrial phylogenomics. *Mol Phylogenet Evol*. 2013;69(3):462–468. <http://dx.doi.org/10.1016/j.ympev.2013.08.005>
 89. Jackson CJ, Reyes-Prieto A. The mitochondrial genomes of the Glaucophytes *Gloeochaete wittrockiana* and *Cyanoptylche gloeocystis*: multilocus phylogenetics suggests a monophyletic archaeplastida. *Genome Biol Evol*. 2014;6(10):2774–2785. <http://dx.doi.org/10.1093/gbe/evu218>
 90. Stiller JW, Riley J, Hall BD. Are red algae plants? A critical evaluation of three key molecular data sets. *J Mol Evol*. 2001;52(6):527–539. <http://dx.doi.org/10.1007/s002390010183>
 91. Roger AJ, Hug LA. The origin and diversification of eukaryotes: problems with molecular phylogenetics and molecular clock estimation. *Philos Trans R Soc Lond B Biol Sci*. 2006;361(1470):1039–1054. <http://dx.doi.org/10.1098/rstb.2006.1845>
 92. Delsuc F, Brinkmann H, Philippe H. Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet*. 2005;6(5):361–375. <http://dx.doi.org/10.1038/nrg1603>
 93. Brown MW, Sharpe SC, Silberman JD, Heiss AA, Lang BF, Simpson AGB, et al. Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. *Proc Biol Sci*. 2013;280(1769):20131755. <http://dx.doi.org/10.1098/rspb.2013.1755>
 94. Felsenstein J. Cases in which parsimony or compatibility methods will be positively misleading. *Syst Zool*. 1978;27(4):401. <http://dx.doi.org/10.2307/2412923>
 95. Stiller JW, Harrell L. The largest subunit of RNA polymerase II from the Glaucocystophyta: functional constraint and short-branch exclusion in deep eukaryotic phylogeny. *BMC Evol Biol*. 2005;5(1):71. <http://dx.doi.org/10.1186/1471-2148-5-71>
 96. Roure B, Baurain D, Philippe H. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Mol Biol Evol*. 2013;30(1):197–214. <http://dx.doi.org/10.1093/molbev/mss208>
 97. Philippe H, Snell EA, Bapteste E, Lopez P, Holland PW, Casane D. Phylogenomics of eukaryotes: impact of missing data on large alignments. *Mol Biol Evol*. 2004;21(9):1740–1752. <http://dx.doi.org/10.1093/molbev/msh182>
 98. Nozaki H, Iseki M, Hasegawa M, Misawa K, Nakada T, Sasaki N, et al. Phylogeny of primary photosynthetic eukaryotes as deduced from slowly evolving nuclear genes. *Mol Biol Evol*. 2007;24(8):1592–1595. <http://dx.doi.org/10.1093/molbev/mms091>
 99. Inagaki Y, Nakajima Y, Sato M, Sakaguchi M, Hashimoto T. Gene sampling can bias multi-gene phylogenetic inferences: the relationship between red algae and green plants as a case study. *Mol Biol Evol*. 2009;26(5):1171–1178. <http://dx.doi.org/10.1093/molbev/msp036>
 100. Brinkmann H, van der Giezen M, Zhou Y, de Raucourt GP, Philippe

- H. An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. *Syst Biol.* 2005;54(5):743–757. <http://dx.doi.org/10.1080/10635150500234609>
101. Susko E, Roger AJ. On reduced amino acid alphabets for phylogenetic inference. *Mol Biol Evol.* 2007;24(9):2139–2150. <http://dx.doi.org/10.1093/molbev/msm144>
 102. Tuffley C, Steel M. Modeling the covarion hypothesis of nucleotide substitution. *Math Biosci.* 1998;147(1):63–91. [http://dx.doi.org/10.1016/S0025-5564\(97\)00081-3](http://dx.doi.org/10.1016/S0025-5564(97)00081-3)
 103. Chan CX, Yang EC, Banerjee T, Yoon HS, Martone PT, Estevez JM, et al. Red and green algal monophyly and extensive gene sharing found in a rich repertoire of red algal genes. *Curr Biol.* 2011;21(4):328–333. <http://dx.doi.org/10.1016/j.cub.2011.01.037>
 104. Maddison WP. Gene trees in species trees. *Syst Biol.* 1997;46(3):523–536. <http://dx.doi.org/10.1093/sysbio/46.3.523>
 105. Page R. From gene to organismal phylogeny: reconciled trees and the gene tree/species tree problem. *Mol Phylogenet Evol.* 1997;7(2):231–240. <http://dx.doi.org/10.1006/mpev.1996.0390>
 106. Chan CX, Bhattacharya D. Analysis of horizontal genetic transfer in red algae in the post-genomics age. *Mob Genet Elem.* 2013;3(6):e27669. <http://dx.doi.org/10.4161/mge.27669>
 107. Chan CX, Bhattacharya D, Reyes-Prieto A. Endosymbiotic and horizontal gene transfer in microbial eukaryotes: impacts on cell evolution and the tree of life. *Mob Genet Elem.* 2012;2(2):101–105. <http://dx.doi.org/10.4161/mge.20110>
 108. Huang J, Yue J. Horizontal gene transfer in the evolution of photosynthetic eukaryotes: HGT in plants. *J Syst Evol.* 2013;51(1):13–29. <http://dx.doi.org/10.1111/j.1759-6831.2012.00237.x>
 109. Schönknecht G, Weber APM, Lercher MJ. Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. *BioEssays.* 2014;36(1):9–20. <http://dx.doi.org/10.1002/bies.201300095>
 110. Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet.* 2008;9(8):605–618. <http://dx.doi.org/10.1038/nrg2386>
 111. Keeling PJ. Role of horizontal gene transfer in the evolution of photosynthetic eukaryotes and their plastids. In: Gogarten MB, Gogarten JP, Olendzenski LC, editors. *Horizontal gene transfer*. Totowa, NJ: Humana Press; 2009. p. 501–515. (vol 532). http://dx.doi.org/10.1007/978-1-60327-853-9_29
 112. Richardson AO, Palmer JD. Horizontal gene transfer in plants. *J Exp Bot.* 2006;58(1):1–9. <http://dx.doi.org/10.1093/jxb/erl148>
 113. Bergthorsson U, Richardson AO, Young GJ, Goertzen LR, Palmer JD. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. *Proc Natl Acad Sci USA.* 2004;101(51):17747–17752. <http://dx.doi.org/10.1073/pnas.0408336102>
 114. Bergthorsson U, Adams KL, Thomason B, Palmer JD. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature.* 2003;424(6945):197–201. <http://dx.doi.org/10.1038/nature01743>
 115. Woloszynska M, Bocer T, Mackiewicz P, Janska H. A fragment of chloroplast DNA was transferred horizontally, probably from non-eudicots, to mitochondrial genome of *Phaseolus*. *Plant Mol Biol.* 2004;56(5):811–820. <http://dx.doi.org/10.1007/s11103-004-5183-y>
 116. Timmis JN, Ayliffe MA, Huang CY, Martin W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet.* 2004;5(2):123–135. <http://dx.doi.org/10.1038/nrg1271>
 117. Kleine T, Maier UG, Leister D. DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. *Annu Rev Plant Biol.* 2009;60(1):115–138. <http://dx.doi.org/10.1146/annurev.arplant.043008.092119>
 118. Lane CE, Archibald JM. The eukaryotic tree of life: endosymbiosis takes its TOL. *Trends Ecol Evol.* 2008;23(5):268–275. <http://dx.doi.org/10.1016/j.tree.2008.02.004>
 119. Leigh JW, Susko E, Baumgartner M, Roger AJ. Testing congruence in phylogenomic analysis. *Syst Biol.* 2008;57(1):104–115. <http://dx.doi.org/10.1080/10635150801910436>
 120. Stiller JW. Experimental design and statistical rigor in phylogenomics of horizontal and endosymbiotic gene transfer. *BMC Evol Biol.* 2011;11(1):259. <http://dx.doi.org/10.1186/1471-2148-11-259>
 121. Andersson JO, Roger AJ. A cyanobacterial gene in nonphotosynthetic protists – an early chloroplast acquisition in eukaryotes? *Curr Biol.* 2002;12(2):115–119.
 122. Cavalier-Smith T. The origin, losses and gains of chloroplasts. In: Lewin RA, editor. *Origins of plastids*. New York, NY: Chapman and Hall; 1993. p. 291–348.
 123. Dorrell RG, Smith AG. Do red and green make brown?: perspectives on plastid acquisitions within chromalveolates. *Eukaryot Cell.* 2011;10(7):856–868. <http://dx.doi.org/10.1128/EC.00326-10>
 124. Vesteg M, Vacula R, Krajčovič J. On the origin of chloroplasts, import mechanisms of chloroplast-targeted proteins, and loss of photosynthetic ability – review. *Folia Microbiol Praha.* 2009;54(4):303–321. <http://dx.doi.org/10.1007/s12223-009-0048-z>
 125. Krause K. From chloroplasts to “cryptic” plastids: evolution of plastid genomes in parasitic plants. *Curr Genet.* 2008;54(3):111–121. <http://dx.doi.org/10.1007/s00294-008-0208-8>
 126. Borza T, Popescu CE, Lee RW. Multiple metabolic roles for the nonphotosynthetic plastid of the green alga *Prototheca wickerhamii*. *Eukaryot Cell.* 2005;4(2):253–261. <http://dx.doi.org/10.1128/EC.4.2.253-261.2005>
 127. Mazumdar J, Wilson EH, Masek K, Hunter CA, Striepen B. Apicoplast fatty acid synthesis is essential for organelle biogenesis and parasite survival in *Toxoplasma gondii*. *Proc Natl Acad Sci USA.* 2006;103(35):13192–13197. <http://dx.doi.org/10.1073/pnas.0603391103>
 128. Bodył A, Mackiewicz P, Gagat P. Organelle evolution: *Paulinella* breaks a paradigm. *Curr Biol.* 2012;22(9):R304–R306. <http://dx.doi.org/10.1016/j.cub.2012.03.020>
 129. Gagat P, Bodył A, Mackiewicz P, Stiller JW. Tertiary plastid endosymbioses in dinoflagellates. In: Löffelhardt W, editor. *Endosymbiosis*. Vienna: Springer; 2014. p. 233–290. http://dx.doi.org/10.1007/978-3-7091-1303-5_13
 130. Kies L, Kremer BP. Function of cyanelles in the tecamoeba *Paulinella chromatophora*. *Naturewissenschaften.* 1979;66:578–579.
 131. Bodył A, Mackiewicz P, Stiller JW. The intracellular cyanobacteria of *Paulinella chromatophora*: endosymbionts or organelles? *Trends Microbiol.* 2007;15(7):295–296. <http://dx.doi.org/10.1016/j.tim.2007.05.002>
 132. Nowack ECM, Melkonian M, Glöckner G. Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by eukaryotes. *Curr Biol.* 2008;18(6):410–418. <http://dx.doi.org/10.1016/j.cub.2008.02.051>
 133. Reyes-Prieto A, Yoon HS, Moustafa A, Yang EC, Andersen RA, Boo SM, et al. Differential gene retention in plastids of common recent origin. *Mol Biol Evol.* 2010;27(7):1530–1537. <http://dx.doi.org/10.1093/molbev/msq032>
 134. Nowack ECM, Vogel H, Groth M, Grossman AR, Melkonian M, Glockner G. Endosymbiotic gene transfer and transcriptional regulation of transferred genes in *Paulinella chromatophora*. *Mol Biol Evol.* 2011;28(1):407–422. <http://dx.doi.org/10.1093/molbev/msq209>
 135. Nakayama T, Ishida K. Another acquisition of a primary photosynthetic organelle is underway in *Paulinella chromatophora*. *Curr Biol.* 2009;19(7):R284–R285. <http://dx.doi.org/10.1016/j.cub.2009.02.043>
 136. Nowack ECM, Grossman AR. Trafficking of protein into the recently established photosynthetic organelles of *Paulinella chromatophora*. *Proc Natl Acad Sci USA.* 2012;109(14):5340–5345. <http://dx.doi.org/10.1073/pnas.1118800109>
 137. Mackiewicz P, Bodył A, Gagat P. Protein import into the photosynthetic organelles of *Paulinella chromatophora* and its implications for primary plastid endosymbiosis. *Symbiosis.* 2012;58(1–3):99–107. <http://dx.doi.org/10.1007/s13199-012-0202-2>
 138. Mackiewicz P, Bodył A, Gagat P. Possible import routes of proteins into the cyanobacterial endosymbionts/plastids of *Paulinella chromatophora*. *Theory Biosci.* 2012;131(1):1–18. <http://dx.doi.org/10.1007/s12064-011-0147-7>
 139. Bodył A, Mackiewicz P, Stiller JW. Comparative genomic studies suggest that the cyanobacterial endosymbionts of the amoeba *Paulinella chromatophora* possess an import apparatus for

- nuclear-encoded proteins. *Plant Biol.* 2009;12:639–649. <http://dx.doi.org/10.1111/j.1438-8677.2009.00264.x>
140. Carpenter EJ, Foster RA. Marine cyanobacterial symbioses. In: Rai AN, Bergman B, Rasmussen U, editors. *Cyanobacteria in symbiosis*. Dordrecht: Kluwer Academic Publishers; 2002. p. 11–17. http://dx.doi.org/10.1007/0-306-48005-0_2
 141. Raven JA. Evolution of cyanobacterial symbioses. In: Rai AN, Bergman B, Rasmussen U, editors. *Cyanobacteria in symbiosis*. Dordrecht: Kluwer Academic Publishers; 2002. p. 329–346. http://dx.doi.org/10.1007/0-306-48005-0_16
 142. Kneip C, Lockhart P, Voß C, Maier UG. Nitrogen fixation in eukaryotes – new models for symbiosis. *BMC Evol Biol.* 2007;7(1):55. <http://dx.doi.org/10.1186/1471-2148-7-55>
 143. Kneip C, Voß C, Lockhart PJ, Maier UG. The cyanobacterial endosymbiont of the unicellular algae *Rhopalodia gibba* shows reductive genome evolution. *BMC Evol Biol.* 2008;8(1):30. <http://dx.doi.org/10.1186/1471-2148-8-30>
 144. Rogers M, Keeling PJ. Lateral transfer and re compartmentalization of Calvin cycle enzymes of plants and algae. *J Mol Evol.* 2004;58(4):367–375. <http://dx.doi.org/10.1007/s00239-003-2558-7>
 145. Durnford DG, Deane JA, Tan S, McFadden GI, Gantt E, Green BR. A phylogenetic assessment of the eukaryotic light-harvesting antenna proteins, with implications for plastid evolution. *J Mol Evol.* 1999;48(1):59–68.
 146. Rissler HM, Durnford DG. Isolation of a novel carotenoid-rich protein in *Cyanophora paradoxa* that is immunologically related to the light-harvesting complexes of photosynthetic eukaryotes. *Plant Cell Physiol.* 2005;46(3):416–424. <http://dx.doi.org/10.1093/pcp/pci054>
 147. Plancke C, Colleoni C, Deschamps P, Dauvillee D, Nakamura Y, Haebel S, et al. Pathway of cytosolic starch synthesis in the model glaucophyte *Cyanophora paradoxa*. *Eukaryot Cell.* 2008;7(2):247–257. <http://dx.doi.org/10.1128/EC.00373-07>
 148. Deschamps P, Haferkamp I, d’Hulst C, Neuhaus HE, Ball SG. The relocation of starch metabolism to chloroplasts: when, why and how. *Trends Plant Sci.* 2008;13(11):574–582. <http://dx.doi.org/10.1016/j.tplants.2008.08.009>
 149. Deschamps P, Colleoni C, Nakamura Y, Suzuki E, Putaux JL, Buleon A, et al. Metabolic symbiosis and the birth of the plant kingdom. *Mol Biol Evol.* 2008;25(3):536–548. <http://dx.doi.org/10.1093/molbev/msm280>
 150. Ball S, Colleoni C, Cenci U, Raj JN, Tirtiaux C. The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. *J Exp Bot.* 2011;62(6):1775–1801. <http://dx.doi.org/10.1093/jxb/erq411>
 151. Cavalier-Smith T. Eukaryote kingdoms: seven or nine? *Biosystems.* 1981;14(3–4):461–481.
 152. Delwiche CF, Palmer JD. Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. *Mol Biol Evol.* 1996;13(6):873–882.
 153. Shibata M, Kashino Y, Satoh K, Koike H. Isolation and characterization of oxygen-evolving thylakoid membranes and photosystem II particles from a glaucocystophyte, *Cyanophora paradoxa*. *Plant Cell Physiol.* 2001;42(7):733–741. <http://dx.doi.org/10.1093/pcp/pce092>
 154. Koike H, Shibata M, Yasutomi K, Kashino Y, Satoh K. Identification of photosystem I components from a glaucocystophyte, *Cyanophora paradoxa*: the PsdD protein has an N-terminal stretch homologous to higher plants. *Photosynth Res.* 2000;65(3):207–217. <http://dx.doi.org/10.1023/A:1010734912776>
 155. Machida M, Takechi K, Sato H, Chung SJ, Kuroiwa H, Takio S, et al. Genes for the peptidoglycan synthesis pathway are essential for chloroplast division in moss. *Proc Natl Acad Sci USA.* 2006;103(17):6753–6758. <http://dx.doi.org/10.1073/pnas.0510693103>
 156. Takano H, Takechi K. Plastid peptidoglycan. *Biochim Biophys Acta.* 2010;1800(2):144–151. <http://dx.doi.org/10.1016/j.bbagen.2009.07.020>
 157. Lockhart PJ, Howe CJ, Bryant DA, Beanland TJ, Larkum AWD. Substitutional bias confounds inference of cyanelle origins from sequence data. *J Mol Evol.* 1992;34(2):153–162. <http://dx.doi.org/10.1007/BF00182392>
 158. Lockhart PJ, Penny D, Hendy MD, Howe CJ, Beanland TJ, Larkum AWD. Controversy on chloroplast origins. *FEBS Lett.* 1992;301(2):127–131. [http://dx.doi.org/10.1016/0014-5793\(92\)81231-A](http://dx.doi.org/10.1016/0014-5793(92)81231-A)
 159. Okamoto N, Chantangsi C, Horák A, Leander BS, Keeling PJ. Molecular phylogeny and description of the novel katablepharid *Roombia truncata* gen. et sp. nov., and establishment of the hacrobia taxon nov. *PLoS ONE.* 2009;4(9):e7080. <http://dx.doi.org/10.1371/journal.pone.0007080>
 160. Burki F, Shalchian-Tabrizi K, Minge M, Skjæveland Å, Nikolaev SI, Jakobsen KS, et al. Phylogenomics reshuffles the eukaryotic supergroups. *PLoS ONE.* 2007;2(8):e790. <http://dx.doi.org/10.1371/journal.pone.0000790>
 161. Turner S, Pryer KM, Miao VP, Palmer JD. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J Eukaryot Microbiol.* 1999;46(4):327–338.
 162. Reyes-Prieto A, Bhattacharya D. Phylogeny of nuclear-encoded plastid-targeted proteins supports an early divergence of glaucophytes within Plantae. *Mol Biol Evol.* 2007;24(11):2358–2361. <http://dx.doi.org/10.1093/molbev/msm186>
 163. Qiu H, Yang EC, Bhattacharya D, Yoon HS. Ancient gene paralogy may mislead inference of plastid phylogeny. *Mol Biol Evol.* 2012;29(11):3333–3343. <http://dx.doi.org/10.1093/molbev/mss137>
 164. Helmchen TA, Bhattacharya D, Melkonian M. Analyses of ribosomal RNA sequences from glaucocystophyte cyanelles provide new insights into the evolutionary relationships of plastids. *J Mol Evol.* 1995;41(2):203–210.
 165. Yoon HS, Hackett JD, van Dolah FM, Nosenko T, Lidie KL, Bhattacharya D. Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. *Mol Biol Evol.* 2005;22(5):1299–1308. <http://dx.doi.org/10.1093/molbev/msi118>
 166. Yoon HS, Nakayama T, Reyes-Prieto A, Andersen RA, Boo SM, Ishida K, et al. A single origin of the photosynthetic organelle in different *Paulinella* lineages. *BMC Evol Biol.* 2009;9(1):98. <http://dx.doi.org/10.1186/1471-2148-9-98>
 167. Falcón LI, Magallón S, Castillo A. Dating the cyanobacterial ancestor of the chloroplast. *ISME J.* 2010;4(6):777–783. <http://dx.doi.org/10.1038/ismej.2010.2>
 168. Auch AF, Henz SR, Holland BR, Göker M. Genome BLAST distance phylogenies inferred from whole plastid and whole mitochondrion genome sequences. *BMC Bioinformatics.* 2006;7(1):350. <http://dx.doi.org/10.1186/1471-2105-7-350>
 169. Nozaki H, Ohta N, Matsuzaki M, Misumi O, Kuroiwa T. Phylogeny of plastids based on cladistic analysis of gene loss inferred from complete plastid genome sequences. *J Mol Evol.* 2003;57(4):377–382. <http://dx.doi.org/10.1007/s00239-003-2486-6>
 170. Sanchez-Puerta MV, Bachvaroff TR, Delwiche CF. Sorting wheat from chaff in multi-gene analyses of chlorophyll *c*-containing plastids. *Mol Phylogenet Evol.* 2007;44(2):885–897. <http://dx.doi.org/10.1016/j.ympev.2007.03.003>
 171. de Las Rivas J. Comparative analysis of chloroplast genomes: functional annotation, genome-based phylogeny, and deduced evolutionary patterns. *Genome Res.* 2002;12(4):567–583. <http://dx.doi.org/10.1101/gr.209402>
 172. Rodríguez-Ezpeleta N, Brinkmann H, Burger G, Roger AJ, Gray MW, Philippe H, et al. Toward resolving the eukaryotic tree: the phylogenetic positions of jakobids and cercozoans. *Curr Biol.* 2007;17(16):1420–1425. <http://dx.doi.org/10.1016/j.cub.2007.07.036>
 173. Lopez P, Casane D, Philippe H. Heterotachy, an important process of protein evolution. *Mol Biol Evol.* 2002;19(1):1–7.
 174. Vogl C, Badger J, Kearney P, Li M, Clegg M, Jiang T. Probabilistic analysis indicates discordant gene trees in chloroplast evolution. *J Mol Evol.* 2003;56(3):330–340. <http://dx.doi.org/10.1007/s00239-002-2404-3>
 175. Ane C. Covariation structure in plastid genome evolution: a new statistical test. *Mol Biol Evol.* 2005;22(4):914–924. <http://dx.doi.org/10.1093/molbev/msi076>
 176. Whelan S, Blackburne BP, Spencer M. Phylogenetic substitution models for detecting heterotachy during plastid evolution. *Mol Biol Evol.* 2011;28(1):449–458. <http://dx.doi.org/10.1093/molbev/msq215>
 177. Lockhart PJ, Steel MA, Barbrook AC, Huson DH, Charleston MA, Howe CJ. A covariotide model explains apparent phylogenetic

- structure of oxygenic photosynthetic lineages. *Mol Biol Evol.* 1998;15(9):1183–1188.
178. Lockhart P, Steel M. A tale of two processes. *Syst Biol.* 2005;54(6):948–951. <http://dx.doi.org/10.1080/10635150500234682>
 179. Lockhart P. Heterotachy and tree building: a case study with plastids and eubacteria. *Mol Biol Evol.* 2005;23(1):40–45. <http://dx.doi.org/10.1093/molbev/msj005>
 180. Rice DW, Palmer JD. An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. *BMC Biol.* 2006;4(1):31. <http://dx.doi.org/10.1186/1741-7007-4-31>
 181. Mackiewicz P, Bodył A, Moszczyński K. The case of horizontal gene transfer from bacteria to the peculiar dinoflagellate plastid genome. *Mob Genet Elem.* 2013;3(4):e25845. <http://dx.doi.org/10.4161/mge.25845>
 182. Moszczyński K, Mackiewicz P, Bodył A. Evidence for horizontal gene transfer from bacteroidetes bacteria to dinoflagellate minicircles. *Mol Biol Evol.* 2012;29(3):887–892. <http://dx.doi.org/10.1093/molbev/msr276>
 183. Bachvaroff TR, Sanchez-Puerta MV, Delwiche CF. Chlorophyll *c*-containing plastid relationships based on analyses of a multigene data set with all four chromalveolate lineages. *Mol Biol Evol.* 2005;22(9):1772–1782. <http://dx.doi.org/10.1093/molbev/msi172>
 184. Cuvelier ML, Allen AE, Monier A, McCrow JP, Messie M, Tringe SG, et al. Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proc Natl Acad Sci USA.* 2010;107(33):14679–14684. <http://dx.doi.org/10.1073/pnas.1001665107>
 185. Hagopian JC, Reis M, Kitajima JP, Bhattacharya D, de Oliveira MC. Comparative analysis of the complete plastid genome sequence of the red alga *Gracilaria tenuistipitata* var. *liui* provides insights into the evolution of rhodoplasts and their relationship to other plastids. *J Mol Evol.* 2004;59(4):464–477. <http://dx.doi.org/10.1007/s00239-004-2638-3>
 186. Janouškovec J, Horak A, Obornik M, Lukes J, Keeling PJ. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc Natl Acad Sci USA.* 2010;107(24):10949–10954. <http://dx.doi.org/10.1073/pnas.1003335107>
 187. Janouškovec J, Horák A, Barott KL, Rohwer FL, Keeling PJ. Global analysis of plastid diversity reveals apicomplexan-related lineages in coral reefs. *Curr Biol.* 2012;22(13):R518–R519. <http://dx.doi.org/10.1016/j.cub.2012.04.047>
 188. Khan H, Parks N, Kozera C, Curtis BA, Parsons BJ, Bowman S, et al. Plastid genome sequence of the cryptophyte alga *Rhodomonas salina* CCMP1319: lateral transfer of putative DNA replication machinery and a test of chromist plastid phylogeny. *Mol Biol Evol.* 2007;24(8):1832–1842. <http://dx.doi.org/10.1093/molbev/msm101>
 189. Kim E, Harrison JW, Sudek S, Jones MDM, Wilcox HM, Richards TA, et al. Newly identified and diverse plastid-bearing branch on the eukaryotic tree of life. *Proc Natl Acad Sci USA.* 2011;108(4):1496–1500. <http://dx.doi.org/10.1073/pnas.1013337108>
 190. Le Corguillé G, Pearson G, Valente M, Viegas C, Gschloessl B, Corre E, et al. Plastid genomes of two brown algae, *Ectocarpus siliculosus* and *Fucus vesiculosus*: further insights on the evolution of red-algal derived plastids. *BMC Evol Biol.* 2009;9(1):253. <http://dx.doi.org/10.1186/1471-2148-9-253>
 191. Martin W, Stoebe B, Goremykin V, Hansmann S, Hasegawa M, Kowalik KV. Gene transfer to the nucleus and the evolution of chloroplasts. *Nature.* 1998;393(6681):162–165. <http://dx.doi.org/10.1038/30234>
 192. Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, et al. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci USA.* 2002;99(19):12246–12251. <http://dx.doi.org/10.1073/pnas.182432999>
 193. Nelissen B, van de Peer Y, Wilmotte A, de Wachter R. An early origin of plastids within the cyanobacterial divergence is suggested by evolutionary trees based on complete 16S rRNA sequences. *Mol Biol Evol.* 1995;12(6):1166–1173.
 194. Ohta N, Matsuzaki M, Misumi O, Miyagishima SY, Nozaki H, Tanaka K, et al. Complete sequence and analysis of the plastid genome of the unicellular red alga *Cyanidioschyzon merolae*. *DNA Res.* 2003;10(2):67–77. <http://dx.doi.org/10.1093/dnares/10.2.67>
 195. Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ. The complete chloroplast genome of the chlorarachniophyte *Bigelowiella natans*: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. *Mol Biol Evol.* 2007;24(1):54–62. <http://dx.doi.org/10.1093/molbev/msl129>
 196. Sato N. Origin and evolution of plastids: genomic view on the unification and diversity of plastids. In: Wise RR, Hooper JK, editors. *The structure and function of plastids*. Dordrecht: Springer; 2006. p. 75–102. (Advances in photosynthesis and respiration). http://dx.doi.org/10.1007/978-1-4020-4061-0_4
 197. Tengs T, Dahlberg OJ, Shalchian-Tabrizi K, Klaveness D, Rudi K, Delwiche CF, et al. Phylogenetic analyses indicate that the 19'Hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Mol Biol Evol.* 2000;17(5):718–729.
 198. Wang Y, Joly S, Morse D. Phylogeny of dinoflagellate plastid genes recently transferred to the nucleus supports a common ancestry with red algal plastid genes. *J Mol Evol.* 2008;66(2):175–184. <http://dx.doi.org/10.1007/s00239-008-9070-z>
 199. Baurain D, Brinkmann H, Petersen J, Rodriguez-Ezpeleta N, Stechmann A, Demoulin V, et al. Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Mol Biol Evol.* 2010;27(7):1698–1709. <http://dx.doi.org/10.1093/molbev/msq059>
 200. Minge MA, Silberman JD, Orr RJ, Cavalier-Smith T, Shalchian-Tabrizi K, Burki F, et al. Evolutionary position of breviate amoebae and the primary eukaryote divergence. *Proc Biol Sci.* 2009;276(1657):597–604. <http://dx.doi.org/10.1098/rspb.2008.1358>
 201. Moreira D, Le Guyader H, Philippe H. The origin of red algae and the evolution of chloroplasts. *Nature.* 2000;405(6782):69–72. <http://dx.doi.org/10.1038/35011054>
 202. Patron NJ, Inagaki Y, Keeling PJ. Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. *Curr Biol.* 2007;17(10):887–891. <http://dx.doi.org/10.1016/j.cub.2007.03.069>
 203. Tekle YI, Grant J, Cole JC, Nerad TA, Anderson OR, Patterson DJ, et al. A multigene analysis of *Corallomyxa tenera* sp. nov. suggests its membership in a clade that includes Gromia, Haplosporidia and Foraminifera. *Protist.* 2007;158(4):457–472. <http://dx.doi.org/10.1016/j.protis.2007.05.002>
 204. Tekle YI, Parfrey LW, Katz LA. Molecular data are transforming hypotheses on the origin and diversification of eukaryotes. *Bioscience.* 2009;59(6):471–481. <http://dx.doi.org/10.1525/bio.2009.59.6.5>