

EEF2 Analysis Challenges the Monophyly of Archaeplastida and Chromalveolata

Eunsoo Kim^{1*}, Linda E. Graham

Department of Botany, University of Wisconsin-Madison, Madison, Wisconsin, United States of America

Abstract

Background: Classification of eukaryotes provides a fundamental phylogenetic framework for ecological, medical, and industrial research. In recent years eukaryotes have been classified into six major supergroups: Amoebozoa, Archaeplastida, Chromalveolata, Excavata, Opisthokonta, and Rhizaria. According to this supergroup classification, Archaeplastida and Chromalveolata each arose from a single plastid-generating endosymbiotic event involving a cyanobacterium (Archaeplastida) or red alga (Chromalveolata). Although the plastids within members of the Archaeplastida and Chromalveolata share some features, no nucleocytoplasmic synapomorphies supporting these supergroups are currently known.

Methodology/Principal Findings: This study was designed to test the validity of the Archaeplastida and Chromalveolata through the analysis of nucleus-encoded eukaryotic translation elongation factor 2 (EEF2) and cytosolic heat-shock protein of 70 kDa (HSP70) sequences generated from the glaucophyte *Cyanophora paradoxa*, the cryptophytes *Goniomonas truncata* and *Guillardia theta*, the katablepharid *Leucocryptos marina*, the rhizarian *Thaumatomonas* sp. and the green alga *Mesostigma viride*. The HSP70 phylogeny was largely unresolved except for certain well-established groups. In contrast, EEF2 phylogeny recovered many well-established eukaryotic groups and, most interestingly, revealed a well-supported clade composed of cryptophytes, katablepharids, haptophytes, rhodophytes, and Viridiplantae (green algae and land plants). This clade is further supported by the presence of a two amino acid signature within EEF2, which appears to have arisen from amino acid replacement before the common origin of these eukaryotic groups.

Conclusions/Significance: Our EEF2 analysis strongly refutes the monophyly of the Archaeplastida and the Chromalveolata, adding to a growing body of evidence that limits the utility of these supergroups. In view of EEF2 phylogeny and other morphological evidence, we discuss the possibility of an alternative eukaryotic supergroup.

Citation: Kim E, Graham LE (2008) EEF2 Analysis Challenges the Monophyly of Archaeplastida and Chromalveolata. PLoS ONE 3(7): e2621. doi:10.1371/journal.pone.0002621

Editor: Rosemary Jeanne Redfield, University of British Columbia, Canada

Received: March 5, 2008; **Accepted:** June 2, 2008; **Published:** July 9, 2008

Copyright: © 2008 Kim et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by grant MCB-9977903 from the National Science Foundation. The grant was used to purchase supplies for molecular sequencing.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: eunsookim@dal.ca

† Current address: Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Canada

Introduction

Eukaryotes constitute one of the three domains of life, distinguished from bacteria and archaeobacteria by their greater molecular, cellular, and reproductive complexity. About 1.5 million species of eukaryotes have been recognized and named thus far, with at least several times that number remaining to be catalogued [1]. Much of eukaryotic diversity occurs among protists, whose high-level classification remains uncertain in spite of the need for a reliable, phylogeny-based classification in ecological, medical, and industrial research.

Eukaryotes can be conservatively classified into about 60 robust lineages based primarily on ultrastructural features [2,3]. Alternatively, eukaryotes have been grouped into only two major clades—unikonts and bikonts—based largely on a single gene fusion event, under the assumption that parallel fusions would be improbable [4–6]. However, this assumption is refuted by evidence that gene fusion events do occur independently in different eukaryotic

groups [7]. A fundamental unikont-bikont dichotomy is also questioned by the phylogenetic position of the bikont Apusozoa among the “unikonts”, as well as other data [8,9]. Other recent authors have classified eukaryotes into 5 or 6 major supergroups: Amoebozoa, Opisthokonta, Archaeplastida (or Plantae), Chromalveolata, Rhizaria, and Excavata, with the first two grouped as ‘unikonts’ by some authors [10–12]. However, the validity of some of these supergroups, notably Excavata, Archaeplastida and Chromalveolata, is controversial [13–21].

The present study was designed to test monophyly of the Archaeplastida and Chromalveolata, each defined by a single primary- or secondary-plastid generating endosymbiotic event [10]. The Archaeplastida is composed of three well characterized monophyletic groups: the Glaucophyta, Rhodophyta (i.e., red algae), and Viridiplantae (i.e., green algae plus land plants) [10]. All members of the Archaeplastida possess double membrane-bound plastids (i.e., primary plastids), which are believed to have been derived directly from a cyanobacterial endosymbiont by primary

endosymbiosis [22]. It should be noted that the rhizarian *Paulinella chromatophora* (which does not group with the Archaeplastida) independently acquired photosynthetic bodies directly from a cyanobacterium [23,24], although there is debate whether to designate these entities as ‘plastids’ or ‘endosymbionts’ [25–27]. The Chromalveolata comprises four monophyletic groups—Alveolata, Cryptophyta (plus Katablepharidae) [8,28], Haptophyta, and Stramenopiles, each group containing at least some members harboring plastids thought to be derived from a red alga by secondary endosymbiosis [10]. Cryptophytes, haptophytes, and stramenopiles were grouped as ‘chromists,’ based on shared features of their plastids. These shared features include the presence of four-bounding membranes and, with some exceptions [29], confluence of the outermost plastid membrane with the nuclear envelope [30]. It should be noted, however, that several “early-diverging” clades within the stramenopiles, and the cryptophyte genus *Goniomonas* (plus katablepharid species), do not possess plastids. Alveolates include ciliates (plastid-less), apicomplexans, and dinoflagellates, the latter including many plastid-less as well as plastid-bearing members [31]. Most plastid-bearing dinoflagellates have peridinin as a major carotenoid fraction and their plastids are generally enclosed by three membranes [32], whereas plastids that are present in the majority of apicomplexans (i.e., apicoplasts) are exclusively non-photosynthetic and are bound by 2–4 membranes [33–36]. Recently, a photosynthetic relative of the apicomplexans has been identified which harbors plastids with four membranes that are related to apicoplasts [37]. In contrast to most ‘chromist’ plastids, the outermost plastid membranes of alveolate plastids are not connected to the nuclear membrane [32,35].

The major issues surrounding the endosymbiotic origins of plastid-bearing eukaryotes can be summarized by the following questions: a) Did the plastids of the Archaeplastida taxa arise from a single or multiple source(s) of cyanobacteria [38–40]? ; b) Are all these plastids derived from a single endosymbiotic event [18,41]? ; c) Can the plastids of the Chromalveolata taxa be traced back to a single red algal type [42]? ; d) Were chromalveolate plastids acquired once or on multiple occasions [43,44]?

As a means of addressing some of these evolutionary concerns, we carefully targeted molecular phylogenetic markers and taxa to the specific issue of the monophyly of the Archaeplastida and Chromalveolata. More focused analyses such as ours can reveal strong gene-specific evidence for or against phylogenetic relationships, which might be overlooked or unrecognizable in concatenation analyses [45]. We chose two conserved, nuclear protein-coding genes that have been widely used to evaluate eukaryotic diversification: EEF2 (eukaryotic translation elongation factor 2) and cytosolic HSP70 (heat-shock protein of 70 kDa) genes. We generated sequence data from representatives of major eukaryotic phyla, including cryptophytes, a glaucophyte, a green alga, a katablepharid, and a rhizarian and analyzed these new sequences together with existing available database sequences for other eukaryotic taxa.

HSP70 is a molecular chaperone which assists in assembly and folding of proteins and occurs universally in all organisms [46]. Because of its highly conserved sequence, HSP70 has been extensively surveyed to address some of the most ancient evolutionary events such as early bacterial, archaeobacterial, and eukaryotic diversification patterns [46,47]. Like HSP70, EEF2 (and its prokaryotic homolog EF-G) is also highly conserved. EEF2 constitutes an essential component of the translational machinery, where it is involved in the protein elongation step, specifically in the translocation of tRNAs and mRNA [48,49]. EEF2 is valued as a phylogenetic marker because of its large size (~800 amino acids) and consequently its potential to retain more phylogenetic signal

than smaller proteins. In a previous study, EEF2 phylogeny strongly suggested a sister relationship between rhodophytes and Viridiplantae; this observation was argued as nucleocytoplasmic evidence in support of a single endosymbiotic origin for primary plastids [50]. Although EEF2 sequences from glaucophytes could critically test Archaeplastida monophyly, such sequences are so far not available.

In this study, we determined six EEF2 and four cytosolic HSP70 sequences from diverse eukaryotic groups. Importantly, our EEF2 and cytosolic HSP70 phylogenies included, for the first time, nearly full-length sequences of glaucophytes, katablepharids, or cryptophytes. The results of our study, specifically the EEF2 analysis, strongly refute the monophyly of the Archaeplastida and Chromalveolata.

Results

EEF2 and cytosolic HSP70 analyses

EEF2 phylogenetic trees inferred from maximum likelihood (RAxML, PhyloBayes) and distance (FastME) methods were more or less similar, although some deep branching patterns that had low bootstrap support, differed (data not shown). In some cases posterior probabilities (PhyloBayes) were very high (>0.95) even when respective bootstrap values (RAxML, FastME) were low (<70%), although these numbers are not directly comparable as they have different statistical interpretations. For this reason, we interpreted a given relationship as being well supported only when all three supporting values (bootstrap and posterior probability) were high (>90% or >0.90). In all three analyses, a number of well-established eukaryotic groups, including rhodophytes, alveolates, opisthokonts, euglenozoans, and Viridiplantae were recovered with >90% bootstrap support and high posterior probabilities (Figure 1). Most interestingly, unlike what is expected from the Archaeplastida proposal, the glaucophyte *C. paradoxa* did not branch with rhodophytes or Viridiplantae in the EEF2 tree. Instead, EEF2 analysis identified a well-supported clade composed of cryptophytes, katablepharids, haptophytes, rhodophytes, and Viridiplantae (Figure 1). Within this large clade, relationships among the five eukaryotic groups were poorly resolved. Although Viridiplantae and rhodophytes were each other’s closest sister group, bootstrap values supporting this relationship were low (49% for RAxML and 66% for FastME) (Figure 1).

Support for the monophyly of the EEF2 of cryptophytes, katablepharids, haptophytes, rhodophytes, and Viridiplantae is further provided in the form of a two amino acid signature (Figure 2). In these EEF2, two consecutive amino acids, serine (S) followed by alanine (A), occur at positions 212 and 213 whereas most other taxa encode the highly conserved amino acid sequences, glycine (G) and serine (S), at these positions (Figure 2). This suggests that the SA amino acids arose via amino acid replacement of the ancestral GS residues.

In contrast to EEF2, major eukaryotic relationships were largely unresolved in the cytosolic HSP70 phylogeny (Figure S1). Of well-established eukaryotic groups, only the alveolates, stramenopiles, euglenozoans, and rhizaria were recovered with >50% bootstrap support and monophyletic cryptophytes, opisthokonts, rhodophytes, and Viridiplantae were not recovered in the ML tree (Figure S1). As most nodes in the ML tree were poorly supported, it is not clear whether some abnormal branching patterns in the cytosolic HSP70 tree are simply due to the lack of informative sites or other factors (e.g., incomplete lineage sorting, horizontal gene transfer, paralogy) that can lead to discordance between the gene and the species trees. HSP70 phylogeny is also known to be susceptible to the long branch attraction (LBA) artifact [51].

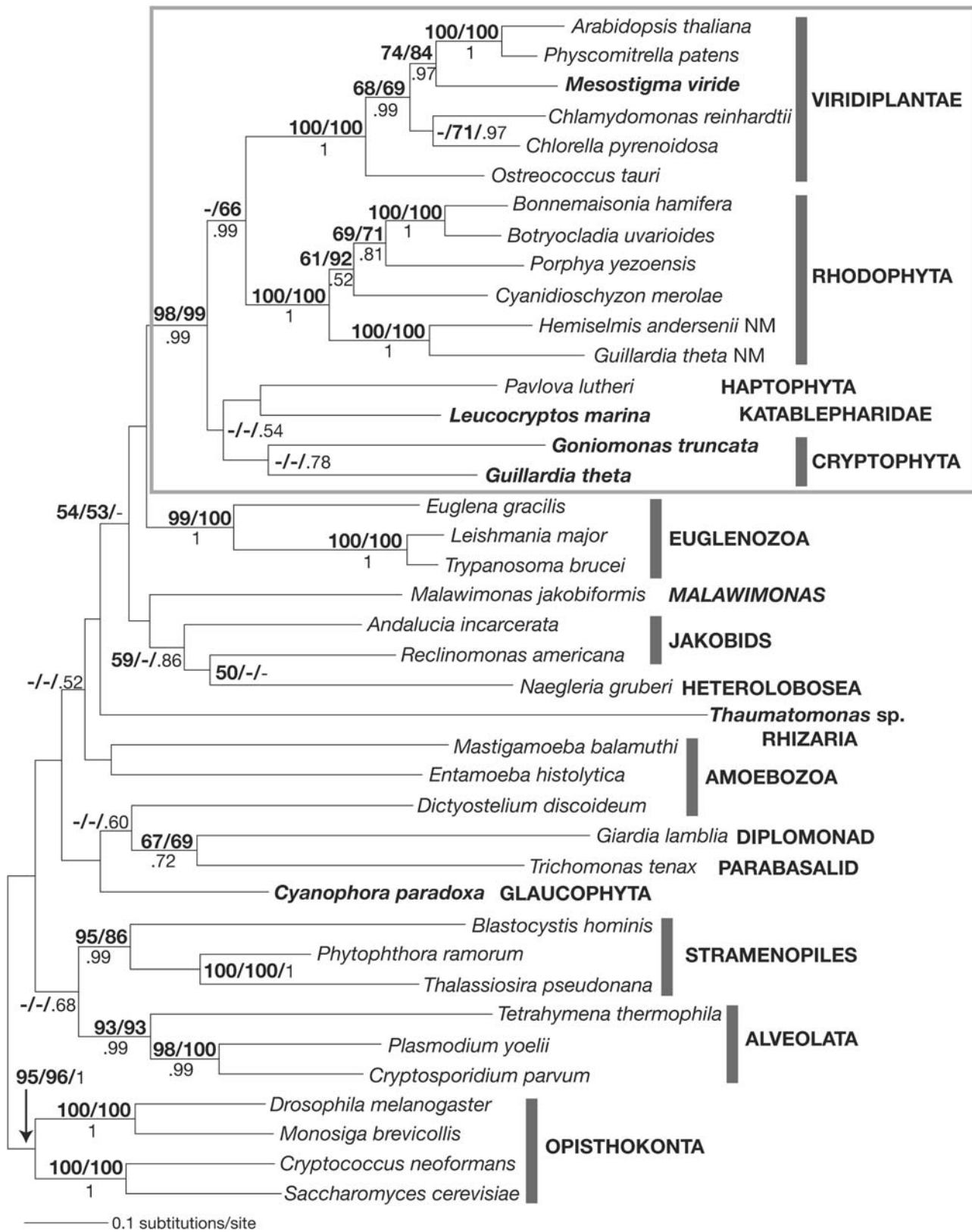



Figure 1. Maximum likelihood tree based on the EEF2 alignment, under the WAG+Γ+I+F model of protein evolution (RAxML). Bootstrap support values >50% (RAxML/FastME) and posterior probabilities >0.50 are indicated at the corresponding nodes. Sequences newly obtained in this study are labeled in bold. Note that the Viridiplantae, Rhodophyta, Haptophyta, Katablepharidae, and Cryptophyta formed a well-supported clade. NM stands for nucleomorph. doi:10.1371/journal.pone.0002621.g001



Viridiplantae	<i>Arabidopsis thaliana</i>	GTVAFSAGLHGWA
	<i>Mesostigma viride</i>	GTVAFSAGLHGWA
	<i>Ostreococcus tauri</i>	GTVCFSAGLHNWA
	<i>Chlamydomonas reinhardtii</i>	STVSFSAGLHGWA
Rhodophyta	<i>Bonnemaisonia hamifera</i>	GTVAFSAGLHGWA
	<i>Botryocladia uvarioides</i>	GTVAFSAGLHGWA
	<i>Porphyra yezoensis</i>	GTIAFSAGLHGWA
	<i>Cyanidioschyzon merolae</i>	GTVAFSAGLHGWA
	<i>Guillardia theta</i> NM	GNVAFSAGLHGWA
	<i>Hemiselmis andersenii</i> NM	NTVSFSAGLHGWA
Haptophyta	<i>Emiliana huxleyi</i>	GTVAFSAGLHGWA
	<i>Pavlova lutheri</i>	GTVSFSAGLQGW
Cryptophyta	<i>Guillardia theta</i>	GTVSFSAGLHGWA
	<i>Goniomonas truncata</i>	GTVAFSAGIHAWA
Katablepharidae	<i>Leucocryptos marina</i>	GTVSFSAGLHGWA
Glaucomphyta	<i>Cyanophora paradoxa</i>	GTVCFGSGYHQWG
Stramenopiles	<i>Blastocystis hominis</i>	GTVAFGSLQQWG
	<i>Phytophthora ramorum</i>	GTVAFGSLHQWG
	<i>Thalassiosira pseudonana</i>	GSVAFGSLHQWA
Rhizaria	<i>Thaumatomonas</i> sp.	GTVGFGLHGWG
Alveolata	<i>Cryptosporidium parvum</i>	GTVSFGSGLHGWA
	<i>Plasmodium yoelii</i>	GTVSFGSLQGW
	<i>Tetrahymena thermophila</i>	GSVAFGSGKECWA
Opisthokonta	<i>Drosophila melanogaster</i>	GSVGFGLHGWA
	<i>Monosiga brevicollis</i>	GTVGFGLHGWA
	<i>Saccharomyces cerevisiae</i>	GTVAFGSLHGWA
Amoebozoa	<i>Dictyostelium discoideum</i>	GTVAFGSLHGWG
	<i>Entamoeba histolytica</i>	GTVAFGSLHGWA
	<i>Mastigamoeba balamuthi</i>	GTVAFGSLHRWA
Euglenozoa	<i>Euglena gracilis</i>	GTVAFGSLQGW
	<i>Leishmania major</i>	GTVAIGSLQAWA
Malawimonas	<i>Malawimonas jakobiformis</i>	GDVGMGSLIGWA
Heterolobosea	<i>Naegleria gruberi</i>	GTVAFGSLHCWG
Diplomonad	<i>Giardia lamblia</i>	GNVAFGSLHGWG
Parabasalid	<i>Trichomonas tenax</i>	GTVAFGSLHSFG
Jakobids	<i>Andalucia incarcerationata</i>	GTVAFGSLHGWG
	<i>Reclinomonas americana</i>	GTVSFASGLMGWG

Figure 2. Two amino acid signature within the EEF2. Note that EEF2 of Viridiplantae, Rhodophyta, Haptophyta, Katablepharidae, and Cryptophyta have amino acids serine and alanine at positions 212 and 213, whereas most other eukaryotes have glycine and serine residues instead. doi:10.1371/journal.pone.0002621.g002

Concatenated protein phylogeny

As single-gene trees are generally poorly resolved due to the presence of limited phylogenetic signal [52], a combined analysis of six proteins— α -tubulin, β -tubulin, actin, cytosolic HSP70, cytosolic HSP90, and EEF2—was performed in an attempt to improve the resolution of the tree by increasing the number of informative characters. Up to about 40% of missing data for a particular taxon was permitted for increasing taxonomic sampling, important for the accuracy of phylogenetic inference [52–54]. The final alignment included 2,797 amino acids with 278 constant sites

and had in total 5.23% missing data. Well-established groups including the alveolates, cryptophytes, euglenozoa, haptophytes, rhodophytes, opisthokonts, stramenopiles, and Viridiplantae were recovered with strong bootstrap support and >0.95 posterior probabilities (Figure 3). In addition, higher level-groupings such as the Opisthokonta-Amoebozoa and the Euglenozoa-Heterolobosea clades received strong support (Figure 3). Cryptophyta, Katablepharidae, and Haptophyta formed a clade with moderate to strong support, which is consistent with recent multiple-gene phylogenies that suggested a close relationship between Cryptophyta and

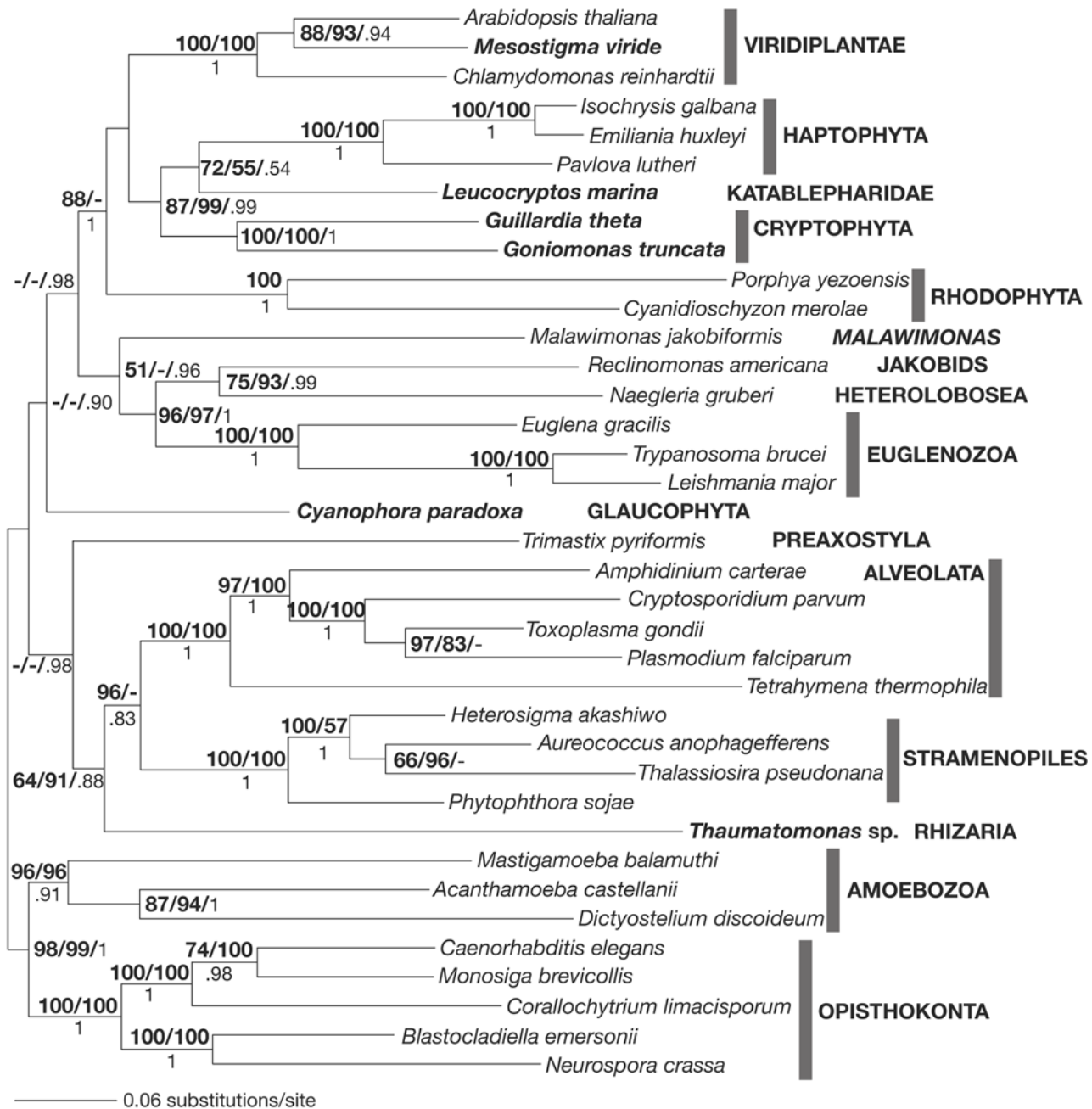


Figure 3. Maximum likelihood tree for the concatenated protein data set, under the WAG+ Γ +I model of protein evolution (RAxML) with the unlinked option. Included proteins are EEF2, actin, cytosolic HSP70, cytosolic HSP90, and α -tubulin, and β -tubulin. Bootstrap support values >50% (RAxML/FastME) and posterior probabilities >0.50 are shown. doi:10.1371/journal.pone.0002621.g003

Haptophyta (Katablepharidae was not examined in these studies) [55,56]. As in the EEF2 analysis, a clade comprising the Cryptophyta, Katablepharidae, Haptophyta, Rhodophyta, and Viridiplantae was recovered in the combined protein tree. Although the clade received the highest posterior probability of 1.0 in both analyses, bootstrap support values for the clade decreased from 98 or 99% in the EEF2 tree to 88 or <50% in the combined analysis (compare Figures 1 and 3). The glaucophyte *C. paradoxa* did not branch close to rhodophytes or Viridiplantae and its phylogenetic position to other eukaryotic groups was unresolved. Lastly, the rhizarian *Thaumatomonas* sp. branched with

alveolates and stramenopiles with weak to moderate support values, consistent with a previous study based upon >100 concatenated protein sequences [14].

Discussion

EEF2 phylogeny refutes the monophyly of Archaeplastida and Chromalveolata

Moreira et al. [50] showed in their EEF2 tree that rhodophytes and Viridiplantae were closely related to each other and that a sister relationship was strongly supported, with a 100% ML bootstrap

value. Prior to that, the hypothesis that a single plastid-generating endosymbiotic event occurred at the origin of glaucophytes, rhodophytes and Viridiplantae was primarily, if not entirely, based on plastid-related features, because no nucleocytoplasmic data in support of the hypothesis were available. Therefore, the Moreira et al. EEF2 result was regarded as the first strong nucleocytoplasmic evidence favoring a monophyletic origin of the Archaeplastida [18,57], although glaucophytes were not examined in their study. Subsequently, an analysis based on >100 concatenated nucleus-encoded proteins indicated that glaucophytes branched closely to rhodophytes and Viridiplantae [58]. Together with plastid-related evidence (see below for details), these results have convinced many researchers in the field that the controversy surrounding the origin of primary plastids was settled (e.g. [11]). However, both the Moreira et al. [50] and Rodriguez-Ezpeleta et al. [58] studies suffered from inadequate taxonomic sampling, notably lacking sequences of cryptophytes, katablepharids, and haptophytes, which appear to be critical in evaluating the validity of the supergroup Archaeplastida as well as the Chromalveolata (see discussion below). In our EEF2 analysis, which included glaucophytes, cryptophytes, and katablepharids, the monophyly of the Archaeplastida and Chromalveolata was strongly refuted. In addition, the specific affiliation of rhodophytes and Viridiplantae is no longer significantly supported, although they still form a well-supported clade together with cryptophytes, katablepharids, and haptophytes (Figure 1).

In recent studies, the strong associations among haptophytes, rhodophytes, and Viridiplantae in EEF2 phylogenies were interpreted as evidence for lateral gene transfer from a red algal endosymbiont to the haptophyte nucleus [56,59]. However, with the addition of our new cryptophyte and katablepharid EEF2 sequences, it is now clear that the earlier proposal is no longer tenable (Figure 1). Cryptophytes have a copy of EEF2 gene in the nucleomorph genome, in addition to one or more nucleus-encoded copies [60,61]. In our study, as is predicted from its red algal ancestry, the cryptophyte nucleomorph-encoded EEF2 branched close to the red algal EEF2 (Figure 1). In contrast, the nucleus-encoded EEF2 of cryptophytes, katablepharids, and haptophytes did not show specific affiliation to the nucleomorph or red algal copies (Figure 1). These branching patterns suggest that the EEF2 gene residing in the nuclei of haptophytes, cryptophytes, and katablepharids was not obtained through endosymbiotic gene transfer from the red algal endosymbiont and probably descended vertically from their ancestors. In addition, the hypothesis of an endosymbiotic gene transfer of EEF2 gene requires the *a priori* assumption that katablepharids and the cryptophyte genus *Goniomonas* once possessed plastids, although there is no molecular or ultrastructural evidence of plastids in these lineages [62,63].

Concatenated protein analysis

Neither the monophyly of the Archaeplastida nor the Chromalveolata were recovered in our concatenated six protein phylogeny (Figure 3). It should be noted, however, that the clade comprising cryptophytes, katablepharids, haptophytes, rhodophytes, and Viridiplantae is no longer significantly supported by bootstrap values. One possible reason might be a long-branch effect of red algal-derived sequences, especially as their actin and β -tubulin sequences are relatively quite divergent (data not shown). On the other hand, it cannot be completely ruled out that compared to other molecular markers, EEF2 has discordant phylogenetic signal, although EEF2 phylogeny does not show any obvious signs of conflict with the five other protein markers examined in this study (data not shown). Nevertheless, it is difficult to differentiate between these two possibilities given the fact that not many other nucleus-encoded

molecular markers have been examined at a similar level of taxonomic sampling. Finally, it is worth mentioning that in a study of >100 concatenated nucleus-encoded protein sequences (albeit with more than 50% of sequence data missing for cryptophytes and haptophytes), the phylogenetic relationships of cryptophytes & haptophytes, rhodophytes, or Viridiplantae to other eukaryotic groups remained unresolved [55]. This suggests that use of markers selected specifically for their information value may be an effective alternative to inferring deep phylogenies by the concatenation approach (or total evidence approach). Given that individual molecular markers can have differing histories due to lateral gene transfer, hidden paralogy, and deep coalescence [64], a concatenation approach can potentially hide strong local phylogenetic signal.

Evaluation of the supergroup Archaeplastida

Over the years, the origin of the plastids in glaucophytes, rhodophytes and Viridiplantae has drawn considerable attention [18,19,38,40,65]. These plastids are known as primary plastids as they are thought to have arisen directly from a cyanobacterial ancestor that was engulfed by an eukaryotic host. Several molecular and genomic data support the notion that these primary plastids arose from a single cyanobacterial endosymbiont. Two particularly compelling pieces of evidence supporting this hypothesis are the presence of an inner plastid membrane translocon Tic110 protein [40,66] and a unique *atpA* gene cluster [67]. These features are common to the plastids of the Viridiplantae, glaucophytes, and rhodophytes, but not found in the cyanobacteria examined thus far. These features may represent post-endosymbiotic inventions that occurred prior to the diversification of the three 'primary' plastids [40,67], although it is also possible that they may have been characteristic of ancestral cyanobacteria of a type so far undiscovered among modern taxa. Triple-helix chlorophyll-binding, light-harvesting antenna complexes (LHCs) have been suggested as another case of post-endosymbiotic innovation [68]. Because LHC homologs have not been identified in glaucophytes [69], such LHCs may have evolved after divergence of the glaucophyte plastid. Plastid genome content and gene phylogenies suggest a single origin of glaucophyte, rhodophyte and Viridiplantae plastids [58,70], although such results do not completely rule out alternative hypotheses [38]. We also note that some other features once considered specific to plastids, such as inverted repeats in rRNA and the *psbB* gene cluster [40,67,71] have been subsequently identified in cyanobacteria, and thus no longer support (nor refute) a single origin hypothesis [39]. In summary, although the hypothesis of a common ancestry for red, green, and glaucophyte plastids is best supported by current data, additional genomic and molecular data for cyanobacteria are needed to further test the hypothesis.

In contrast to their plastids, little or no evidence supports an hypothesis of a common ancestry for the host (i.e., nucleocytoplasmic) component of Viridiplantae, glaucophytes, and rhodophytes. These three lineages differ in ultrastructure and biochemistry [72]. In addition, nucleus-encoded gene phylogenies have often been inconclusive [13,19,73,74]. Although in large-scale phylogenies based on concatenated databases the monophyly of Viridiplantae, glaucophytes, and rhodophytes was initially recovered with strong support [58], the addition of cryptophyte or haptophyte sequences significantly lowered or eliminated support for monophyly of the three lineages [14,55]. Likewise, mitochondrion-encoded gene phylogenies remain largely inconclusive as to the relationship between Viridiplantae and red algae [75–77] (mitochondrial genome data for the glaucophyte taxa are not publicly available for analysis).

Furthermore, although the mechanism of plastid origin by primary endosymbiosis is widely accepted [78], this concept is

primarily based on the presence of two plastid membranes, which may not be a reliable marker if membranes have been lost over time [41]. Some dinoflagellates, for example, have plastids with two membranes that clearly are not of primary origin [32]. Another example is provided by the transient plastids (i.e., kleptoplastids) of the sea slug *Elysia chlorotica*, which have only two membranes, despite their origin from the stramenopiles *Vaucheria litorea*. Such kleptoplastids apparently lost two of the four original plastid membranes [79]. These observations suggest that loss of plastid membranes can occur during or after the engulfment of algal endosymbionts, potentially masking secondary or tertiary origin.

In summary, current data do not provide strong evidence for monophyly of the host lineage of the Viridiplantae, glaucophytes, and rhodophytes, whereas our EEF2 data strongly refute the concept of the Archaeplastida. The observed discrepancy between the nucleocytoplasmic and the plastid genealogy might be better explained by postulating multiple acquisitions of plastids in these eukaryotic lineages. If so, at least one of the ‘primary’ plastids may actually be of secondary origin.

Evaluation of the supergroup Chromalveolata

The chromalveolate hypothesis, namely that cryptophytes, haptophytes, stramenopiles, and alveolates arose from a common ancestor via a secondary endosymbiotic event [30], continues to be debated [15,17,20,43]. The presence of many, early-diverging plastid-less taxa within stramenopiles and alveolates [80–82], and accumulating molecular data, generally conflict with the chromalveolate hypothesis or require massive plastid losses, despite the value of plastids in amino acid, fatty acid and heme biosynthesis, as well as photosynthesis [13,15]. Lack of any sort of evidence from nucleus-encoded gene phylogenies casts further doubt on the chromalveolate hypothesis [8,13]. Although the nucleus-encoded, plastid targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) phylogeny has been presented as evidence for the chromalveolate hypothesis, cytosolic GAPDH sequences among ‘chromalveolate’ taxa did not form a clade, indicating that homologs have discordant evolutionary histories [44,83]. In addition, the plastid-targeted GAPDH tree [84] is inconsistent with accepted organismal relationships; the apicomplexan *Toxoplasma gondii* is a sister to haptophytes with strong support, to the exclusion of peridinin-type dinoflagellates [85]. Overall, the GAPDH phylogenies seem to be more consistent with multiple occasions of plastid acquisition among ‘chromalveolate’ taxa. Plastid-encoded gene phylogenies vary in their level of support for the chromalveolate hypothesis, depending on taxonomic sampling and types and number of analyzed genes [42,86–88]. Even when monophyly of the ‘chromalveolate’ plastids is recovered, it is also consistent with the “serial hypothesis”, which postulates serial transfer of red algal-derived plastids among ‘chromalveolates’ [88,89]. Finally, recent molecular phylogenies showing that rhizaria are closely related to stramenopiles and alveolates [14,56], together with EEF2 evidence presented here appear to deal a fatal blow to the chromalveolate hypothesis.

Evidence for a new eukaryotic supergroup

Based on EEF2 and some morphological data, we propose an alternative eukaryotic supergroup that includes cryptophytes, katablepharids, haptophytes, rhodophytes, and Viridiplantae. We suggest the name Plastidophila (“friendly to plastid”) for the potential clade, because most subclades, except for katablepharid species and one cryptophyte genus (*Goniomonas*), are dominated by plastid-bearing members. Although genomic evidence for Plastidophila is yet limited, some morphological features shared among katablepharids, cryptophytes and Viridiplantae, especially “early-

diverging” prasinophyte green algae, are consistent with this new concept [62,90–92]. For instance, ejectisomes (i.e., ejective organelles) of katablepharids are similar to those of the prasinophyte green alga *Pyramimonas*. Although ejectisomes of *Pyramimonas* form as a spirally coiled ribbon and those of katablepharids take the shape of an elongated tube with a single straight slit, both types discharge into a linear structure [90]. Further, two central flagellar microtubules that do not penetrate into the flagellar insertion area occur in both prasinophytes and katablepharids-cryptophytes [91]. In addition, both the katablepharid *Kathablepharis ovalis* and the prasinophyte *Pyramimonas* possess electron-dense material below the flagellar terminal plate [91]. The striated root that occurs in katablepharids has been suggested to be homologous to the system I fibrous roots found in Viridiplantae [91]. Finally, cell surfaces consisting of a basal fibrous layer and an upper scaly layer is common to katablepharids [90] and scaly green algae such as the “basal” streptophyte green alga *Mesostigma* [93] and the prasinophyte *Tetraselmis* [94]. Based on comparative morphology, Lee and Kugrens [90] and Lee et al. [91] suggested that katablepharids represent evolutionary intermediates between cryptophytes and Viridiplantae. A close relationship between katablepharids and cryptophytes is supported by SSU and LSU rRNA phylogenies [8,28]. Recent analyses based on concatenated protein data sets also suggest a sister relationship between the haptophytes and cryptophytes (katablepharids were unexamined), although the phylogenetic position of this clade relative to other eukaryotes remained unresolved [55,56]. Consistent with these results, our analyses also suggested that cryptophytes, katablepharids, and haptophytes are closely related to each other, although a specific relationship between cryptophytes and katablepharids was not recovered. If the cryptophyte-katablepharids-haptophyte clade and the Plastidophila supergroup suggested by EEF2 phylogeny are indeed correct, it follows that morphological traits common to katablepharids and “early-diverging” green algae might represent features that were shared by the common ancestor of Plastidophila. Hence, katablepharids (plus the cryptophyte *Goniomonas*) may be useful models of the heterotrophic flagellate that was ancestral to the photosynthetic lineage that led to land plants and other algae within the Plastidophila. Genomic analysis of katablepharids and the cryptophyte *Goniomonas* may illuminate nucleocytoplasmic traits of the plant lineage that existed prior to the massive invasion of genes from a cyanobacterial precursor to the plastid [86].

Conclusion

The concepts of Archaeplastida and Chromalveolata do provide a simple way to explain the distribution of primary and secondary plastids by minimizing the number of plastid-generating endosymbiotic events required. However, our EEF2 data add to a growing body of evidence that refutes the Archaeplastida and Chromalveolata. By fostering inaccurate assumptions of relationships, continued use of these supergroup concepts may be deleterious to progress in studies of ecologically, medically, and industrially important protists. Given the lack of support for the monophyly of the Archaeplastida and Chromalveolata, it is sensible to consider alternative evolutionary models. Based on EEF2 analysis and some ultrastructural traits, we suggest testing the concept of a supergroup Plastidophila that links katablepharids-cryptophytes, haptophytes, rhodophytes, and Viridiplantae.

Materials and Methods

Sequencing of EEF2 and cytosolic HSP70 genes

Genomic DNA and/or cDNA were purified from *Cyanophora paradoxa*, *Goniomonas truncata*, *Leucocryptos marina* (NIES 1335),

Mesostigma viride, and *Thaumatomonas* sp. as described in Kim et al. (2006). EEF2 and cytosolic HSP70 genes, typically ~2.5 Kbp and ~2.0 Kbp in size excluding intron regions, are considered relatively large for PCR amplification protocols that employ degenerate primers, so consequently, 2–3 overlapping fragments were PCR amplified and sequenced to obtain nearly full-length sequences of each gene or cDNA. Degenerate primers of about 20–30 bp in size were designed to target conserved sequence regions across diverse eukaryotic taxa within EEF2 and cytosolic HSP70 genes (Table S1). In most cases, the use of these degenerate primer pairs enabled the amplification of only partial regions, hence species-specific primers were subsequently identified from partial sequencing and used to amplify the adjacent fragment(s) (Table S2). EST data for *C. paradoxa* and *M. viride* were utilized to identify species-specific primer sites for EEF2 gene amplifications of these organisms. In many cases, a two-step nested PCR approach was adapted to obtain larger amounts of PCR fragments from very little starting DNA material. PCR amplification, PCR fragment cloning, and sequencing were performed as previously described [8]. As eukaryotes encode 3 or 4 types of HSP70 (i.e., cytosolic, ER, mitochondrial, and plastid forms), each sequenced HSP70 fragment was carefully examined to verify that it contained signature sequence sites for the cytosolic form [95,96]. The EEF2 sequence of *G. theta* was retrieved from the 4× genome assembly, generated by the US Department of Energy Joint Genome Institute (<http://www.jgi.doe.gov/>). Newly obtained EEF2 and cytosolic HSP70 sequences were deposited in GenBank with accession numbers EU812174–812204 (Table S2).

Molecular sequence analysis

Newly obtained EEF2 and HSP70 sequences were manually assembled and aligned to sequences downloaded from GenBank using MacClade ver. 4.08 [97]. Ambiguous regions were excluded. Phylogenetic analysis was performed based on deduced amino acid sequences to minimize phylogenetic artifacts caused by codon usage variations [98]. The final EEF2 and cytosolic HSP70 sequence alignments included 736 and 462 amino acid sites and had 1.08% and 1.28% missing data, respectively. The two alignments were analyzed individually and were combined with α -tubulin, β -tubulin, actin, and cytosolic HSP90 alignments [8] for concatenated protein analysis.

Maximum likelihood analysis of amino acid sequence alignments was performed using RAxML ver. 7.0.4 [99] and PhyloBayes ver. 2.3 [100]. For RAxML analysis, ML trees were inferred with the WAG+I+I+F for the EEF2 data and the WAG+I+I for the concatenated protein data (4 discrete gamma rates), and from 100 distinct randomized maximum parsimony starting trees. The models of protein evolution were selected using ProtTest ver. 1.4 [101]. For the concatenated data set, the ‘-M’

option was applied so that each protein partition had its own branch length. Bootstrap analysis was based on 100 re-samplings. For analysis with PhyloBayes, constant sites were deleted and the CAT+I model of protein evolution with 4 discrete categories for gamma distributed rates was applied [100]. Markov chains were run for 60,000 cycles, the first 5,000 points were discarded as burn-in, and every 10th tree from the remaining points was collected to compute the posterior probabilities for individual nodes. For each analysis, two chains were run in parallel and compared to check for convergence.

Protein distance analysis was performed using TREE-PUZZLE ver. 5.2 [102] and FastME [103]. For TREE-PUZZLE analysis, pairwise maximum likelihood distances were estimated under the WAG+I+I model with 4 and 8 discrete Gamma distribution rates for EEF2 and the concatenated data set, respectively. The resulting distance matrices were then used to construct distance trees using FastME with the initial tree construction option of the Greedy Minimum Evolution algorithm and the tree swapping option of the Balanced Nearest Neighbor Interchanges algorithm. Bootstrap analysis was based on 100 re-samplings using puzzleboot ver. 1.03 (available from www.tree-puzzle.de). Bootstrap datasets were generated using the SEQBOOT program from the PHYLIP package ver. 3.66 [104].

Supporting Information

Figure S1

Found at: doi:10.1371/journal.pone.0002621.s001 (0.37 MB EPS)

Table S1

Found at: doi:10.1371/journal.pone.0002621.s002 (0.27 MB EPS)

Table S2

Found at: doi:10.1371/journal.pone.0002621.s003 (0.28 MB EPS)

Acknowledgments

We thank B. Larget and A. J. Roger for access to computation facility for phylogenetic analysis, A. G. B. Simpson for providing initial sequence alignments, and A. D. Tsousis for Greek translation. J. M. Archibald, D. F. Spencer, and A. J. Roger provided helpful comments on the manuscript. We also thank J. M. Archibald, M.W. Gray, P. J. Keeling, G. I. McFadden and C. E. Lane for providing the nucleus-encoded EEF2 sequence from *Guillardia theta*, which was obtained from preliminary genome sequence data produced by the Joint Genome Institute’s Community Sequencing Program (<http://www.jgi.doe.gov/>).

Author Contributions

Conceived and designed the experiments: EK. Performed the experiments: EK. Analyzed the data: EK. Contributed reagents/materials/analysis tools: EK LEG. Wrote the paper: EK LEG.

References

- Bisby FA, Roskov YR, Ruggiero MA, Orrell TM, Paglinawan LE, et al. (2007) Species 2000 & ITIS catalogue of life: 2007 annual checklist. Species 2000. Retrieved Jan. 21, 2008 <www.catalogueoflife.org/annual-checklist/2007/>.
- Patterson DJ (1999) The diversity of eukaryotes. *Am Nat* 154: S96–S124.
- Patterson DJ (2000) The lineages of eukaryotes. Tree of life web project. Retrieved Jan 21, 2008 <www.tolweb.org/notes/?note_id=49>.
- Stechmann A, Cavalier-Smith T (2002) Rooting the eukaryote tree by using a derived gene fusion. *Science* 297: 89–91.
- Richards TA, Cavalier-Smith T (2005) Myosin domain evolution and the primary divergence of eukaryotes. *Nature* 436: 1113–1118.
- Stechmann A, Cavalier-Smith T (2003) Phylogenetic analysis of eukaryotes using heat-shock protein Hsp90. *J Mol Evol* 57: 408–419.
- Makiuchi T, Nara T, Annoura T, Hashimoto T, Aoki T (2007) Occurrence of multiple, independent gene fusion events for the fifth and sixth enzymes of pyrimidine biosynthesis in different eukaryotic groups. *Gene* 394: 78–86.
- Kim E, Simpson AGB, Graham LE (2006) Evolutionary relationships of apusomonads inferred from taxon-rich analyses of 6 nuclear encoded genes. *Mol Biol Evol* 23: 2455–2466.
- Nozaki H, Matsuzaki M, Misumi O, Kuroiwa H, Higashiyama T, et al. (2005) Phylogenetic implications of the CAD complex from the primitive red alga *Cyanidioschyzon merolae* (Cyanidiales, Rhodophyta). *J Phycol* 41: 652–657.
- Adl SM, Simpson AGB, Farmer MA, Andersen RA, Anderson OR, et al. (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol* 52: 399–451.
- Keeling PJ, Burger G, Durmford DG, Lang BF, Lee RW, et al. (2005) The tree of eukaryotes. *Trends Ecol Evol* 20: 670–676.
- Simpson AGB, Roger AJ (2004) The real ‘kingdoms’ of eukaryotes. *Curr Biol* 14: R693–R696.
- Parfrey LW, Barbero E, Lasser E, Dunthorn M, Bhattacharya D, et al. (2006) Evaluating support for the current classification of eukaryotic diversity. *PLoS Genet* 2: e220.

14. Burki F, Shalchian-Tabrizi K, Minge M, Skjæveland A, Nikolaev SI, et al. (2007) Phylogenomics reshuffles the eukaryotic supergroups. *PLoS ONE* 2: e790.
15. Bodyl A (2005) Do plastid-related characters support the chromalveolate hypothesis? *J Phycol* 41: 712–719.
16. Stiller JW, Riley J, Hall BD (2001) Are red algae plants? A critical evaluation of three key molecular data sets. *J Mol Evol* 52: 527–539.
17. Keeling PJ, Archibald JM, Fast NM, Palmer JD (2004) Comment on “The evolution of modern eukaryotic phytoplankton”. *Science* 306: 2191b.
18. Palmer JD (2003) The symbiotic birth and spread of plastids: How many times and whodunit? *J Phycol* 39: 4–11.
19. Nozaki H, Matsuzaki M, Takahara M, Misumi O, Kuroiwa H, et al. (2003) 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* 56: 485–497.
20. Grzebyk D, Katz ME, Knoll AH, Quigg A, Raven JA, et al. (2004) Response to comment on “The evolution of modern eukaryotic phytoplankton”. *Science* 306: 2191c.
21. Yoon HS, Grant J, Tekle YI, Wu M, Chaon BC, et al. (2008) Broadly sampled multigene trees of eukaryotes. *BMC Evol Biol* 8: 14.
22. Jarvis P, Soll M (2001) Toc, Tic, and chloroplast protein import. *Biochim Biophys Acta* 1541: 64–79.
23. Marin B, Nowack ECM, Melkonian M (2005) A plastid in the making: primary endosymbiosis. *Protist* 156: 425–432.
24. Nowack ECM, Melkonian M, Glockner G (2008) Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by eukaryotes. *Curr Biol* 18: 410–418.
25. Bodyl A, Mackiewicz P, Stiller JW (2007) The intracellular cyanobacteria of *Paulinella chromatophora*: endosymbionts or organelles? *Trends Microbiol* 15: 295–296.
26. Theissen U, Martin W (2006) The difference between organelles and endosymbionts. *Curr Biol* 16: R1016–R1017.
27. Bhattacharya D, Archibald JM (2006) The difference between organelles and endosymbionts - response to Theissen and Martin. *Curr Biol* 16: R1017–R1018.
28. Okamoto N, Inouye I (2005) 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* 156: 163–179.
29. Andersen RA (2004) Biology and systematics of heterokont and haptophyte algae. *Am J Bot* 91: 1508–1522.
30. Cavalier-Smith T (1999) Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J Eukaryot Microbiol* 46: 347–366.
31. Graham LE, Wilcox LW (2000) *Algae*. Upper Saddle River NJ: Prentice Hall.
32. Schnepf E, Elbrachter M (1999) Dinophyte chloroplasts and phylogeny: a review. *Grana* 38: 81–97.
33. Kohler S, Delwiche CF, Denny PW, Tilney LG, Webster P, et al. (1997) A plastid of probable green algal origin in apicomplexan parasites. *Science* 275: 1485–1489.
34. Kohler S (2005) Multi-membrane-bound structures of Apicomplexa: I. the architecture of the *Toxoplasma gondii* apicoplast. *Parasitol Res* 96: 258–272.
35. Hopkins J, Fowler R, Krishna S, Wilson I, Mitchell G, et al. (1999) The plastid in *Plasmodium falciparum* asexual blood stages: a three-dimensional ultrastructural analysis. *Protist* 150: 283–295.
36. Tomova C, Geerts WJC, Muller-Reichert T, Entzeroth R, Humbel BM (2006) New comprehension of the apicoplast of *Sarcocystis* by transmission electron tomography. *Biol Cell* 98: 535–545.
37. Moore RB, Obornik M, Janouskovec J, Chrudimsky T, Vancova M, et al. (2008) A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* 451: 959–963.
38. Stiller JW, Reel DC, Johnson JC (2003) A single origin of plastids revisited: convergent evolution in organellar genome content. *J Phycol* 39: 95–105.
39. Larkum AWD, Lockhart PJ, Howe CJ (2007) Shopping for plastids. *Trends Plant Sci* 12: 189–195.
40. McFadden GI, van Dooren GG (2004) Evolution: red algal genome affirms a common origin of all plastids. *Curr Biol* 14: R514–R516.
41. Stiller JW, Hall BD (1997) The origin of red algae: implications for plasmid evolution. *Proc Natl Acad Sci U S A* 94: 4520–4525.
42. Sanchez-Puerta MV, Bachvaroff TR, Delwiche CF (2007) Sorting wheat from chaff in multi-gene analyses of chlorophyll c-containing plastids. *Mol Phylogenet Evol* 44: 885–897.
43. Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, et al. (2004) The evolution of modern eukaryotic phytoplankton. *Science* 305: 354–360.
44. Fast NM, Kissinger JC, Roos DS, Keeling PJ (2001) Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol Biol Evol* 18: 418–426.
45. Bucknam J, Boucher Y, Baptiste E (2006) Refuting phylogenetic relationships. *Biol Direct* 1: 26.
46. Gupta RS, Golding GB (1993) Evolution of HSP70 gene and its implications regarding relationships between archaeobacteria, eubacteria, and eukaryotes. *J Mol Evol* 37: 573–582.
47. Gupta RS, Singh B (1994) Phylogenetic analysis of 70 kD heat shock protein sequences suggests a chimeric origin for the eukaryotic cell nucleus. *Curr Biol* 4: 1104–1114.
48. Gomez-Lorenzo MG, Spahn CMT, Agrawal RK, Grassucci RA, Penczek P, et al. (2000) Three-dimensional cryo-electron microscopy localization of EEF2 in the *Saccharomyces cerevisiae* 80S ribosome at 17.5 angstrom resolution. *EMBO J* 19: 2710–2718.
49. Jorgensen R, Merrill AR, Andersen GR (2006) The life and death of translation elongation factor 2. *Biochem Soc Trans* 34: 1–6.
50. Moreira D, Le Guyader H, Philippe H (2000) The origin of red algae and the evolution of chloroplasts. *Nature* 405: 69–72.
51. Germet A, Philippe H (1999) Critical analysis of eukaryotic phylogeny: a case study based on the HSP70 family. *J Eukaryot Microbiol* 46: 116–124.
52. Philippe H, Delsuc F, Brinkmann H, Lartillot N (2005) Phylogenomics. *Annu Rev Ecol Evol Syst* 36: 541–562.
53. Wiens JJ (2006) Missing data and the design of phylogenetic analyses. *J Biomed Inform* 39: 34–42.
54. Philippe H, Snell EA, Baptiste E, Lopez P, Holland PWH, et al. (2004) Phylogenomics of eukaryotes: Impact of missing data on large alignments. *Mol Biol Evol* 21: 1740–1752.
55. Patron NJ, Inagaki Y, Keeling PJ (2007) Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. *Curr Biol* 17: 887–891.
56. Hackett JD, Yoon HS, Li S, Reyes-Prieto A, Rummele SE, et al. (2007) Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of Rhizaria with Chromalveolates. *Mol Biol Evol* 24: 1702–1713.
57. McFadden GI (2001) Primary and secondary endosymbiosis and the origin of plastids. *J Phycol* 37: 951–959.
58. Rodriguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, et al. (2005) Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr Biol* 15: 1325–1330.
59. Nosenko T, Bhattacharya D (2007) Horizontal gene transfer in chromalveolates. *BMC Evol Biol* 7: 173.
60. Lane CE, van den Heuvel K, Korera C, Curtis BA, Parsons BJ, et al. (2007) Nucleomorph genome of *Hemiselms anderseni* reveals complete intron loss and compaction as a driver of protein structure and function. *Proc Natl Acad Sci U S A* 104: 19908–19913.
61. Douglas S, Zauner S, Fraunholz M, Beaton M, Penny S, et al. (2001) The highly reduced genome of an enslaved algal nucleus. *Nature* 410: 1091–1096.
62. Vørs N (1992) Ultrastructure and autecology of the marine, heterotrophic flagellate *Leucocryptos marina* (Braaud) Butcher 1967 (Katablepharidaceae/Katablepharidae), with a discussion of the genera *Leucocryptos* and *Katablepharis/Katablepharis*. *Eur J Protistol* 28: 369–389.
63. McFadden GI, Gilson PR, Hill DR (1994) *Goniomonas*: ribosomal RNA sequences indicate that this phagotrophic flagellate is a close relative of the host component of cryptomonads. *Eur J Phycol* 29: 29–32.
64. Maddison WP (1997) Gene trees in species trees. *Syst Biol* 46: 523–536.
65. Stiller JW (2007) Plastid endosymbiosis, genome evolution and the origin of green plants. *Trends Plant Sci* 12: 391–396.
66. Steiner JM, Yusa F, Pompe JA, Löffelhardt W (2005) Homologous protein import machineries in chloroplasts and cyanelles. *Plant J* 44: 646–652.
67. Stoebe B, Kowallik KV (1999) Gene-cluster analysis in chloroplast genomics. *Trends Genet* 15: 344–347.
68. Durnford DG, Deane JA, Tan S, McFadden GI, Gantt E, et al. (1999) A phylogenetic assessment of the eukaryotic light-harvesting antenna proteins, with implications for plastid evolution. *J Mol Evol* 48: 59–68.
69. Rissler HM, Durnford DG (2005) 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* 46: 416–424.
70. Stoebe B, Martin W, Kowallik KV (1998) Distribution and nomenclature of protein-coding genes in 12 sequenced chloroplast genomes. *Plant Mol Biol Rep* 16: 243–255.
71. Löffelhardt W, Bohnert HJ, Bryant DA (1997) The complete sequence of the *Cyanophora paradoxa* cyanelle genome (Glaucocystophyceae). *Plant Syst Evol*. pp 149–162.
72. O’Kelly C (1993) Relationships of eukaryotic algal groups to other protists. In: Berner T, ed (1993) *Ultrastructure of microalgae*. Boca Raton, FL: CRC Press. pp 269–294.
73. Stiller JW, Harrell L (2005) The largest subunit of RNA polymerase II from the Glaucocystophyta: functional constraint and short-branch exclusion in deep eukaryotic phylogeny. *BMC Evol Biol* 5: 71.
74. Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290: 972–977.
75. Burger G, Saint-Louis D, Gray MW, Lang BF (1999) Complete sequence of the mitochondrial DNA of the red alga *Porphyra purpurea*: cyanobacterial introns and shared ancestry of red and green algae. *Plant Cell* 11: 1675–1694.
76. Secq MPO, Goer SL, Stam WT, Olsen JL (2006) Complete mitochondrial genomes of the three brown algae (Heterokonta: Phaeophyceae) *Dictyota dichotoma*, *Fucus vesiculosus* and *Desmarestia viridis*. *Curr Genet* 49: 47–58.
77. Kim E, Lane CE, Curtis BA, Kozera C, Bowman S, et al. (2008) Complete sequence and analysis of the mitochondrial genome of *Hemiselms anderseni* CCMP644 (Cryptophyceae). *BMC Genomics* 9: 215.

78. Gibbs SP (1981) The Chloroplasts of some algal groups may have evolved from endosymbiotic eukaryotic algae. *Ann N Y Acad Sci* 361: 193–208.
79. Rumpho ME, Summer EJ, Manhart JR (2000) Solar-powered sea slugs. Mollusc/algal chloroplast symbiosis. *Plant Physiol* 123: 29–38.
80. Leander BS, Keeling PJ (2003) Morphostasis in alveolate evolution. *Trends Ecol Evol* 18: 395–402.
81. Moriya M, Nakayama T, Inouye I (2002) A new class of the stramenopiles, Placidiea classis nova: description of *Placidia cafeteriopsis* gen. et sp nov. *Protist* 153: 143–156.
82. Kim E, Archibald JM (2008) Diversity and evolution of plastids and their genomes. In: Sandelius AS, Aronsson H, eds (2008) *The Chloroplast: Interactions with the environment*. Heidelberg: Springer. pp In press.
83. Harper JT, Keeling PJ (2003) Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. *Mol Biol Evol* 20: 1730–1735.
84. Takishita K, Ishida KI, Maruyama T (2004) Phylogeny of nuclear-encoded plastid-targeted GAPDH gene supports separate origins for the peridinin- and the fucoxanthin derivative-containing plastids of dinoflagellates. *Protist* 155: 447–458.
85. Takishita K, Kawachi M, Noel MH, Matsumoto T, Kakizoe N, et al. (2008) Origins of plastids and glyceraldehyde-3-phosphate dehydrogenase genes in the green-colored dinoflagellate *Lepidodinium chlorophorum*. *Gene* 410: 26–36.
86. Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, et al. (2002) Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci U S A* 99: 12246–12251.
87. Ohta N, Matsuzaki M, Misumi O, Miyagishima S, Nozaki H, et al. (2003) Complete sequence and analysis of the plastid genome of the unicellular red alga *Cyanidioschyzon merolae*. *DNA Res* 10: 67–77.
88. Bachvaroff TR, Puerta MVS, Delwiche CF (2005) Chlorophyll c-containing plastid relationships based on analyses of a multigene data set with all four chromalveolate lineages. *Mol Biol Evol* 22: 1772–1782.
89. Bodyl A, Moszczynski K (2006) Did the peridinin plastid evolve through tertiary endosymbiosis? A hypothesis. *Eur J Phycol* 41: 435–448.
90. Lee RE, Kugrens P (1991) *Katablepharis ovalis*, a colorless flagellate with interesting cytological characteristics. *J Phycol* 27: 505–513.
91. Lee RE, Kugrens P, Mylnikov AP (1992) The structure of the flagellar apparatus of two strains of *Katablepharis* (Cryptophyceae). *Br Phycol J* 27: 369–380.
92. Clay B, Kugrens P (1999) Systematics of the enigmatic katablepharids, including EM characterization of the type species, *Katablepharis phoenixstoni*, and new observations on *K. remigera* com. nov. *Protist* 150: 43–59.
93. Domozych DS, Wells B, Shaw PJ (1992) Scale biogenesis in the green alga, *Mesosigma viride*. *Protoplasma* 167: 19–32.
94. Domozych DS, Stewart KD, Mattox KR (1981) Development of the cell wall in *Tetraselmis*: role of the Golgi apparatus and extracellular wall assembly. *J Cell Sci* 52: 351–371.
95. Gupta RS (1998) Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaeobacteria, eubacteria, and eukaryotes. *Microbiol Mol Biol Rev* 62: 1435–1491.
96. Boorstein WR, Ziegelhoffer T, Craig EA (1994) Molecular evolution of the HSP70 multigene family. *J Mol Evol* 38: 1–17.
97. Maddison DR, Maddison WP (2001) *MacClade 4: analysis of phylogeny and character evolution*. Sunderland, MA: Sinauer Associates Inc.
98. Inagaki Y, Simpson AGB, Dacks JB, Roger AJ (2004) Phylogenetic artifacts can be caused by leucine, serine, and arginine codon usage heterogeneity: dinoflagellate plastid origins as a case study. *Syst Biol* 53: 582–593.
99. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.
100. Lartillot N, Brinkmann H, Philippe H (2007) Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol* 7: S4.
101. Abascal F, Zardoya R, Posada D (2005) ProfTest: selection of best-fit models of protein evolution. *Bioinformatics* 21: 2104–2105.
102. Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18: 502–504.
103. Desper R, Gascuel O (2002) Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *J Comput Biol* 9: 687–705.
104. Felsenstein J (2005) *PHYLIP (Phylogeny Inference Package) version 3.6*. Seattle: Department of Genome Sciences, University of Washington.