

Plastid origin: who, when and why?

Chuan Ku, Mayo Roettger, Verena Zimorski, Shijulal Nelson-Sathi, Filipa L. Sousa, William F. Martin*

Institute of Molecular Evolution, Heinrich-Heine-University Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany

Abstract

The origin of plastids is best explained by endosymbiotic theory, which dates back to the early 1900s. Three lines of evidence based on protein import machineries and molecular phylogenies of eukaryote (host) and cyanobacterial (endosymbiont) genes point to a single origin of primary plastids, a unique and important event that successfully transferred two photosystems and oxygenic photosynthesis from prokaryotes to eukaryotes. The nature of the cyanobacterial lineage from which plastids originated has been a topic of investigation. Recent studies have focused on the branching position of the plastid lineage in the phylogeny based on cyanobacterial core genes, that is, genes shared by all cyanobacteria and plastids. These studies have delivered conflicting results, however. In addition, the core genes represent only a very small portion of cyanobacterial genomes and may not be a good proxy for the rest of the ancestral plastid genome. Information in plant nuclear genomes, where most genes that entered the eukaryotic lineage through acquisition from the plastid ancestor reside, suggests that heterocyst-forming cyanobacteria in Stanier's sections IV and V are most similar to the plastid ancestor in terms of gene complement and sequence conservation, which is in agreement with models suggesting an important role of nitrogen fixation in symbioses involving cyanobacteria. Plastid origin is an ancient event that involved a prokaryotic symbiont and a eukaryotic host, organisms with different histories and genome evolutionary processes. The different modes of genome evolution in prokaryotes and eukaryotes bear upon our interpretations of plastid phylogeny.

Keywords: cyanobacteria; endosymbiosis; evolution; gene transfer; genomics; organelle; photosynthesis; phylogeny

Plastid origin: 110 years since Mereschkowsky

Plastids are eukaryotic metabolic compartments responsible for photosynthesis [1] and a variety of metabolic functions including the biosynthesis of amino acids [2], nucleotides [3], lipids [4], and cofactors [5]. The importance of plastids to photosynthetic lineages of eukaryotes cannot be overstated. In 1905, Mereschkowsky proposed a fully articulated version of endosymbiotic theory positing that plastids originated from cyanobacteria that came to reside as symbionts in eukaryotic cells [6,7]. The theory did not make its way into mainstream biological thinking until it was revived in a synthesis by Margulis [8,9] that also incorporated Wallin's [10] – and Paul Portier's (published in French, cited in Sapp [11]) – ideas about endosymbiotic theory for mitochondrial origin, yet adorned by Margulis' own suggestion of a spirochaete origin of flagella. The spirochaete story never took hold, leaving endosymbiotic theory with three main players: the plastid, the mitochondrion, and its host, which is now understood to be an archaeon [12]. Well into the 1970s, resistance to the concept of endosymbiosis for the origin of plastids (and mitochondria) was stiff [13–15].

With the availability of protein and DNA sequences, of which Mereschkowsky knew nothing, strong evidence had accumulated by the early 1980s that plastids originated through endosymbiosis from cyanobacteria, rather than autogenously [16]. However, there have been recurrent suggestions, starting with Mereschkowsky [6] and tracing into the 1970's [17], that there were several independent origins of plastids from cyanobacteria. Today it is widely, but not universally [18,19], accepted that plastids had a single origin. The strongest evidence for that view is that the protein import machinery, a good marker for endosymbiotic events [20], in the three lineages of Archaeplastida – eukaryotes with primary plastids [21] – consists of homologous host-originated components [22–24], which would not be the case had plastids in those lineages arisen from independent cyanobacterial symbioses.

Genomics has enriched our understanding of plastid origin. We now know that the plastid genome has undergone extreme reduction while leaving its imprints in the nuclear genome through endosymbiotic gene transfer [25,26] and that plastids have spread across eukaryotes through symbiosis and become secondary and tertiary plastids [27]. Genomic data also supports the case for a single plastid origin. Despite some concerns about incomplete sampling and phylogenetic artifacts [18], recent analyses from different groups, incorporating improved phylogenetic algorithms and sampling of cyanobacteria, provide two lines of evidence, in addition to

* Corresponding author. Email: bill@hhu.de

Handling Editor: Andrzej Bodyl

the protein import machinery, for a monophyletic origin of plastids involving a single host lineage and a single cyanobacterial lineage. (i) All plastids are monophyletic and nested within the cyanobacterial clade [28–34]. (ii) Eukaryotes with primary plastids are monophyletic based on both nuclear [29,35] and mitochondrial [36] genomes.

However, much is still left unknown about the origin of plastids. What is the cyanobacterial lineage most closely related to plastids? What was the ancestor of plastids like as it entered into the symbiotic relationship? What might be the reasons that triggered this far-reaching event? Here we consider these aspects and recent advances on issues concerning the when, who and why of plastid origin.

When: early or late branching for the plastid lineage within cyanobacteria?

Cyanobacteria are traditionally classified into five sections according to their morphological and developmental patterns [37]. Section I are unicellular cocci, section II are cocci that aggregate, section III are filamentous, section IV are filamentous with heterocysts, and section V are filamentous with true branching and heterocysts. That taxonomy does not, however, correspond to molecular phylogenetic groupings in any phylogeny reported so far. Although different molecular phylogenies are often themselves mutually inconsistent,

some general trends regarding the relationships of extant cyanobacteria have emerged from studies based on rDNA or multiple protein-coding genes [30,32–34,38–43], which are summarized as the backbone tree shown in Fig. 1. There is little dispute that the thylakoid-lacking genus *Gloeobacter* (with which *Aphanothece* is possibly synonymous [40,44]) is the most basal lineage within cyanobacteria, whose closest relatives discovered so far are the nonphotosynthetic melainabacteria [43]. The other cyanobacteria can be further divided into a basal clade of *Synechococcus* strains (e.g., JA-2-3B), a clade of *Synechococcus* sp. PCC 7502 and *Pseudanabaena* strains (e.g., PCC 7367), and a core clade consisting of the other taxa (Fig. 1). The core cyanobacteria contain the majority of the described species and strains and can be further divided into three groups: a clade of section I taxa (e.g., *Thermosynechococcus*; this clade’s basal position is not recovered in some analyses [31,38,39,41,43]), a clade of mixed section III and I taxa (including the fast-evolving SynPro clade of *Synechococcus* and *Prochlorococcus*) and finally a clade where all sections are represented, including heterocyst-forming and true-branching taxa.

If we take the last common ancestor of the core cyanobacteria as a reference point, much of the recent debate on the closest extant cyanobacterial neighbor of the plastids is about “early” (point “a” in Fig. 1 [30–33]) versus “late” (points “b–e” in Fig. 1 [29,34,38,39,45,46]) plastid branching. Here instead of “early origin” [47] or “deep origin” [30], the term

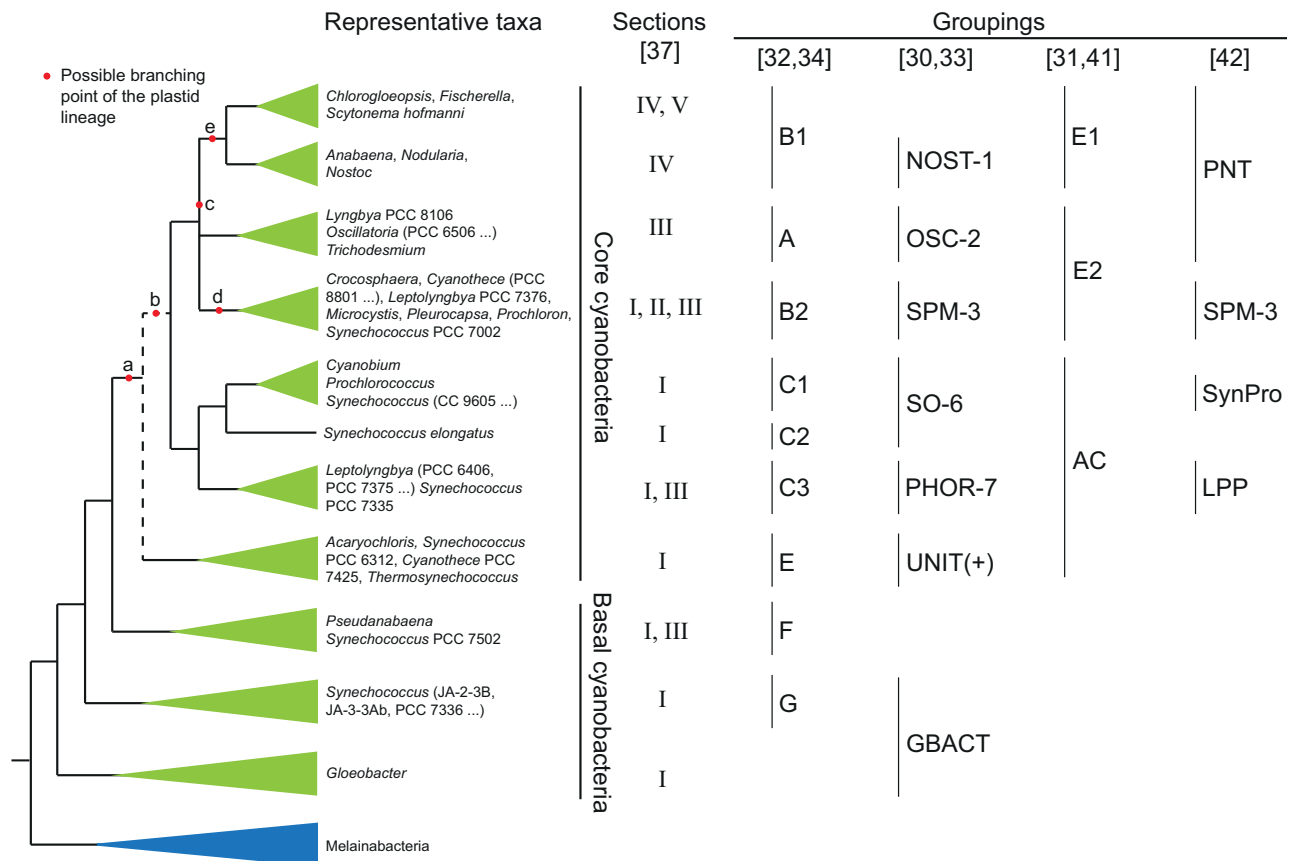


Fig. 1 Possible branching points of the plastid lineage in a consensus cyanobacterial tree based on molecular phylogenetic studies [29–34,38–40,42,43], showing section distribution [37] and groupings in different references. Dotted lines indicate a branching pattern not recovered in some references [31,38,39,41–43]. Branching points: a [30–33]; b [29]; c [34]; d [45] (or within group B2) and [38]; e [39,46] (or within group B1). Numbers in brackets indicate reference numbers for sections and groupings.

“early vs. late branching” is used, because the point where the plastid lineage (the lineage that leads to plastids and does not include other sampled cyanobacteria) branches off the tree might not correspond temporally to the origin of plastids, as the endosymbiosis event itself could have taken place much later than the branching point. Importantly, and as always, discussions of “early” or “late” branching are contingent upon the taxon sample underpinning the tree.

Is there any way to correlate plastid origin to geological time? If a universal molecular clock existed for the cyanobacterial tree, the depth of the branching point could be estimated from sequence divergence. But plastids apparently have a much higher substitution rate than free-living cyanobacteria, among which some taxa such as the group C1 (Fig. 1) also have a higher rate than others. Fossil records indicate that eukaryotes originated no earlier than 1800 million years ago (mya) [48] and complex red algae already existed 1200 mya [49]. These two benchmarks set the upper and lower bounds for the origin of Archaeplastida and plastids, which, according to fossil-calibrated trees that take those benchmarks into account (and is thus not independent of them), is estimated to have taken place around 1500–1600 mya [50,51]. These estimates, however, do not by themselves directly discriminate early- or late-branching scenarios because the branching position of the lineage and the event of plastid origin are two different things.

If the estimated time of the plastid origin is so ancient that it predates the diversification of major cyanobacterial lineages in the core clade (Fig. 1), then a late branching from within the core clade can be ruled out. However, molecular clock estimates suggest the last common ancestor of the core clade existed around 2500–3000 mya [38,41,42]. This makes both the early- and late-branching hypotheses possible. If the former is correct, it would mean that plastids originated from a relatively basal lineage about 1 billion years after the lineage branched off from the main stem of the cyanobacterial tree. If a late-branching hypothesis (points “c”, “d” or “e” in Fig. 1) is correct, this would set the branching time to the range 2.3–1.6 billion years ago [38,41,42], much closer in time to the minimum date of plastid origin (1.2 billion years ago) set by *Bangiomorpha* [49,51].

One possible explanation for the hitherto unsettled dispute is that the topologies of trees including genes from plastids (or plastid-derived nuclear genes) and free-living cyanobacteria are highly sensitive to the differences between evolutionary models used in the phylogenetic analyses. In one recently published study, for example, when the model is changed from LG+16 Γ to LG+d+CAT, the sister group of plastids changes from the core cyanobacterial clade to the B2 (Fig. 1) clade (supplementary Fig. 3A and 3C in Ochoa de Alda et al. [34]). Additionally, plastid genes and plastid-derived nuclear genes have higher substitution rates than cyanobacterial homologs and this can cause the plastid lineage to branch off at a more basal position [39]. Nucleotide compositional biases of protein-coding genes were also suggested to be misleading in tree reconstruction [33]. These factors and the intrinsic uncertainty associated with deep phylogenetics, especially for prokaryotes, where only some “core” genes can be used for analyses, will likely continue to plague the plastid evolution issue. Indeed, even

the earliest studies on large datasets of concatenated plastid proteins revealed that fully resolved, but conflicting trees are obtained [52,53], such that the choice of models determines the result. However, core gene phylogenetics need not be the only way to investigate early plastid evolution.

Plastid origin and the origin of oxygenic photosynthesis

A topic usually circumvented in the origin of plastids issue is the origin of cyanobacteria themselves and the advent of oxygenic photosynthesis. Oxygenic photosynthesis using two photosystems in series only occurs among the cyanobacteria (including plastids). Other phototrophic prokaryotes use only one photosystem (PS), either a homolog of PSII, as in alpha- and gammaproteobacteria and chloroflexi, or a homolog of PSI, as in chlorobia, acidobacteria, and firmicutes (heliobacteria), to carry out anaerobic photosynthesis. Several lines of evidence strongly implicate lateral gene transfer (LGT) in photosystem evolution: (i) the photosystems have a very patchy and restricted distribution across distant prokaryotic phyla, (ii) the photosystems are found in combination with three unrelated CO₂-fixing pathways (the Calvin cycle, the reverse citric acid cycle and the 3-hydroxypropionate pathway), and (iii) the complete machinery for photosystem biogenesis, including chlorophyll biosynthesis, is found on large mobile plasmids in proteobacteria [54]. The main question regarding the origin of oxygenic photosynthesis was whether the two photosystems, which clearly share a common ancestor [55,56], arose within the same genome and were exported as single PSs to other lineages [57] or whether the photosystems had long independent evolutionary histories in bacterial lineages and were merged via LGT in the cyanobacterial ancestor (reviewed in [58]).

These alternatives can be discriminated by studying chlorophyll biosynthesis evolution [59], because the two distinct photosystem types corresponding to reaction center I (RCI) in PSI and reaction center II (RCII) in PSII are clearly related and RC evolution cannot proceed without chlorophyll. Thus, if the divergence of the two RC types reflects lineage divergence [60,61], then chlorophyll biosynthetic pathways supporting RCI- and RCII-based bacterial photosynthesis should reflect a deep dichotomy similar to that observed between the RCs themselves. Conversely, if the two RC types evolved via gene duplication within the same prokaryote [57] – a protocyanobacterium – and were subsequently exported to other lineages, then there should be no deep dichotomy among chlorophyll biosynthetic pathways. Investigation of chlorophyll biosynthesis evolution provided evidence in favor of Allen’s protocyanobacterial hypothesis [59], which posits that the first organism that possessed two photosystems was an anaerobe that first used them alternatively, growing either like *Chlorobium* in the presence of sulfide (H₂S photosynthesis using PSI) or like *Rhodobacter* in the absence of sulfide (cyclic electron transport using PSII). A redox-dependent regulator, or redox switch, controlling the mutually exclusive expression of the two photosystems would have allowed the protocyanobacterium to use light in the presence or absence of sulfide. A similar, but not identical, situation is observed

today for *Oscillatoria* [57], which can either express only PSI or both photosystems [62].

The transition from having two photosystems to having oxygenic photosynthesis required the invention of a water splitting apparatus. Blankenship [63,64] has favored a model in which a Mn-dependent enzyme such as manganese catalase was the precursor to the water splitting complex. Other models entail environmentally available Mn^{II} to start [65–67]. Mn^{II} is known to undergo photooxidation to Mn^{III} in the presence of uv light [68], hence in the context of Allen's protocyanobacterial hypothesis, were a mutation to occur in the redox switch, allowing both photosystems to be expressed in an environment where sulfide was lacking but Mn^{II} was available, a flux of electrons from environmental (aqueous) Mn^{II} to ferredoxin would have ensued. The final step would have been the transition from oxidizing environmental Mn^{II} (possibly with the help of uv) one at a time as a substrate, to oxidizing a portable Mn^{III} reserve in the cluster of the water splitting apparatus at the periplasmic side of PSII [69]. Recently, geochemical evidence was reported that is in agreement with both that scenario [70] and the model of Dismukes and colleagues [65,71], which also suggested a role for environmental Mn^{II} while also pointing out a role for high CO₂ in the origin of water splitting. The Mn-oxidizing abilities of RCI from *Rhodobacter* [72] are also compatible with the models deriving water splitting complex from environmental Mn^{II}.

An understanding of the evolution of photosynthesis further highlights the uniqueness and importance of the origin of plastids. Although photosynthetic genes have been exported from ancient cyanobacteria to other prokaryotic lineages, none of them have received and retained both photosystems. The only known case of successful transfer of the two photosystems is the origin of photosynthetic eukaryotes, which involved endosymbiosis of cyanobacterial cells. This made plastid-bearing eukaryotes the only group of organisms other than cyanobacteria that are capable of oxygenic photosynthesis.

Who: the 1% and the 99% of the ancestral plastid genome

There are a number of cyanobacterial genomes sequenced, but only a small fraction of that information is typically used to investigate cyanobacterial or plastid evolution. Studies that aim to infer the backbone phylogeny of plastids and free-living cyanobacteria either use single gene trees (16S rDNA) [31,38,41] or, more commonly, sets of protein-coding genes that included 23 [39], 25 [32], 33 [34], 42 [52], 50 [29] or 75 [33] genes. The number of protein-coding genes in sequenced free-living cyanobacteria ranges from 1716 in *Prochlorococcus marinus* MED4 [73] to 12 356 in *Scytonema hofmanni* PCC 7110, the most gene-rich prokaryote known to date [39]. In other words, the cyanobacterial phylogenies are typically based on about 0.1–1% of the genes present in cyanobacterial genomes.

Based on these phylogenies, we can divide the genomes into clades/lineages (Fig. 1). But what can the lineage labels based on the 1% tell us about the other 99% of the genome?

Because of the importance of lateral gene transfer in shaping prokaryotic genomes [74,75], no universal classification can be applied to prokaryotic genomes as a whole [76,77] such that the tree of the shared 1% tells us little, if anything, about how many and what other genes are present in the rest of the genome. Consider, for example, the lineage formed by cyanobacterial sections IV and V (Fig. 1). *Fischerella thermalis* PCC 7521 has 5340 protein-coding genes, not even half as many as *S. hofmanni* in the same lineage [39]. Similarly, different ecotypes of the marine cyanobacterium *Prochlorococcus marinus* show variation in the number of open-reading frames by as much as 40% across strains [78], although their core genes are very similar. Traits such as the filamentous morphology that are determined by multiple genes also cannot be predicted from the core gene phylogeny [32]. In view of these, the core genes can provide a useful taxonomy and convenient genome labels, which however are poor proxies for the number and nature of the genes comprising the other 99%. The debate about the plastid branching point is thus one on how we can best label the plastid ancestor when we try to fit it onto the backbone tree based on 1% of the genome.

Many genes that were present in the cyanobacterial plastid ancestor have been transferred to the nucleus [26,28]. That is, a substantial component of the “other 99%” resides in nuclear chromosomes, but these genes can also be used to address plastid origin. Identification of eukaryote-cyanobacterial homologous genes and sequence similarity comparison suggest that present-day cyanobacteria in the sections IV and V (point “e” in Fig. 1) tend to harbor the most homologs and that they have a higher similarity with those in plastid-bearing eukaryotes [39,46]. This suggests that, in terms of overall genome similarity, extant section IV and V taxa should be most similar to the ancestral plastid, but this does not imply that we would find the plastid ancestor among them. As in the case of mitochondrial endosymbiosis [79], no present-day cyanobacteria would contain the same complement of genes as the plastid ancestor due to the cumulative effect of lateral gene transfer and gene loss.

We know LGT is more frequent among prokaryotes [77] and that the three main mechanisms responsible for prokaryote LGT (transduction, conjugation and transformation) have not been reported to play an important role in bringing genes to eukaryotes from prokaryotes or other eukaryotes. Gene transfers to plastids are also rare [19]. Such findings suggest that 1.5 billion years after plastid origin, even if we know the cyanobacterial lineage most closely related to plastids, it cannot faithfully represent the plastid ancestor in terms of gene content, and cyanobacterial pangenomes figure into this issue. By contrast, the present-day plastid genome, together with plastid-derived nuclear genes, has been more or less “frozen” from the time of endosymbiosis. Therefore, the best reconstruction of the plastid ancestor we can get is probably to identify all nuclear genes of plastid origin by removing the genes contributed by the archaeal host and the mitochondrial endosymbiont [80–82] in addition to the eukaryote-specific gene inventions.

Why: the physiological context of plastid origin

When addressing the ancestral state of microbial physiology that led to the initial advantageous association between the founder endosymbiont and its host, the variety of plastid functions within a eukaryotic cell readily gives rise to different hypotheses. Since Mereschkowsky's initial endosymbiotic theory [6], the production of carbohydrates by the cyanobacterial endosymbiont was thought to be a crucial key for the establishment of the plastids [45]. Additionally, the possible scarcity of freely available oxygen at the time and place where the endosymbiotic event occurred [83–85] led Martin and Müller [80] to propose a syntrophic association between a cyanobacterial symbiont with its heterotrophic, eukaryotic host. In this case, the advantage to the host would have been the coupling of the photosynthetic waste (oxygen) produced by the cyanobacterium with the aerobic respiration that occurred at the host mitochondria. Yet the theory best supported by observations from modern cyanobacterial symbioses is that nitrogen fixation played a crucial role in the establishment of the association between the host and its endosymbiont. In nature, highly diverse taxonomic hosts form symbiotic or endosymbiotic associations with photosynthesizing and N₂-fixing cyanobacteria in various environments. They range from autotrophic algae and plants [86] to heterotrophic fungi where over 1500 species represent cases of lichen symbioses with cyanobacteria [87]. Specifically in plants, cyanobacterial symbioses are present in the four main groups – gymnosperms, angiosperms, pteridophytes and bryophytes [86]. The first confirmed example of a nonfilamentous intracellular endosymbiont with the potential to offer fixed nitrogen to its host is the spheroid body in *Rhopalodia gibba*, a diatom alga [88]. Symbiosis is not restricted to unicellular endosymbionts. The vertically transmitted endosymbiont (*Anabaena*) of the pteridophyte *Azolla*, which was reduced to an organism devoted to nitrogen fixation, is a multicellular cyanobacterium [89]. Another example is the angiosperm *Gunnera manicata* that uses the nitrogen provided by the filamentous symbiont (*Nostoc punctiforme*) and continues growing under N-limited conditions in the presence of the symbiont [90]. In gymnosperms, a nitrogen-fixing cyanobacterial symbiont (*Nostoc*) was found in the roots of most known cycad species [91,92].

It has been shown that nitrogen deprivation stimulates modern symbiotic associations [86,90] and that nitrogen-rich conditions on the contrary facilitate the dissipation of pre-established symbiotic relationships [86]. If nitrogen was the key factor for the host and symbiont to form the initial association, it is congruent with the finding that present-day members of filamentous, heterocyst-forming and N₂-fixing cyanobacterial sections IV and V have a collection of genes most similar to that possessed by the plastid ancestor [39,46]. However, the ability to fix nitrogen was lost over time and modern plant plastids do not perform nitrogen fixation anymore [93]. Nitrogen fixation, including the splitting of the stable triple electron pair bond between the two nitrogen atoms is highly energy expensive. Each mol of fixed nitrogen requires 16 mol of hydrolyzed ATP [94]. The oxidation state of the environment and, as a consequence thereof, the increased availability of nitrate [95] result in a highly

decreased need to acquire fixed nitrogen via symbiosis today and shape present-day N₂-fixing cyanobacterial symbioses into a niche solution in N-poor areas [96], whereas at the origin of plastids, it might have played a crucial role.

Interpreting trees for plastid origin

In recent years, there have been repeated claims in the literature that a chlamydial infection had something to do with plastid origin as a kind of a helper symbiont. According to the most recent version of the hypothesis [97], this hypothetical chlamydial endosymbiont, which is found neither in any extant plastid-bearing lineages nor in any contemporary cyanobacterial symbioses, had contributed a significant number of genes to the nuclear genome. Some problems with the chlamydial hypothesis and reasons why it is unlikely to be true have been discussed previously [20]. Among them is the circumstance that gene transfer from endosymbionts to their host has become very popular topics these days, and people continuously find the idea of gene transfer interesting, be it the human genome, where many claims for gene transfer turned out to be artefacts [98], the case of the plastid-bearing slug *Elysia* [99], where the claims for gene transfer also turned out to be unfounded [100,101], the case of trypanosomes [102,103], where claims for gene transfer also turned out not to be true [104], or the case of ciliates, where claims for gene transfer in the context of secondary endosymbiosis also turned out not to be true [105]. Of those examples, the chlamydia case is most similar to the trypanosome and ciliate cases, because the chlamydia story involves small numbers of genes with odd branching patterns that (*i*) do not stand out above the background signals to be expected in such analyses [105] and that (*ii*) were identified as such in earlier studies of plant genomes [28]. Indeed, none of the examples of cyanobacterial symbioses outlined in the previous section involve chlamydia or any other helper bacteria, cyanobacteria do just fine by themselves when it comes to establishing stable symbioses, both intracellular and extracellular.

Since the chlamydial hypothesis is solely based on trees, rather than cellular evidence as in the endosymbiotic theory for plastids, it is important to know what the trees can really tell us. In addition to potential phylogenetic errors in the trees showing transfers from chlamydiae [106], it is notable that many other lineages of prokaryotes also appear to have donated genes to eukaryotes. Anyone interested in proposing a hypothesis similar to the chlamydial one can readily single out another lineage and a eukaryotic compartment or pathways where more proteins have been apparently donated from that lineage. This appearance, however, comes from the circumstance that genes donated to eukaryotes by plastids and mitochondria continue to be transferred among free-living prokaryotes long after the organelle had its origin [79,107–109]. Chlamydiae and cyanobacteria even specifically share a sufficiently large number of genes that they form a common cluster (a module) in gene sharing networks (Fig. 1C in Dagan et al. [75]). Genes forming that module can generate trees in which chlamydiae look as if they donated genes to eukaryotes [75]. Though proponents of the

chlamydial helper symbiont hypothesis tend to overlook this effect, cognoscenti increasingly appreciate how gene transfer among prokaryotes affects inferences of gene origin through the endosymbiotic origin of organelles [109].

Here, a critic might ask whether apparent chimerism such as that observed for nuclear-encoded genes is also observed for organelle-encoded proteins, and if not, why not? The answer is that it is observed, the problem being that very few studies have ever looked for such (apparent) chimerism. For example, the phylogenies of the proteins in the *Reclinomonas* mitochondrial genome [110], long the largest mitochondrial genome known with 63 protein coding genes, were rarely scrutinized. In 2004, Esser et al. [111], did however investigate the phylogeny of the 55 proteins encoded in the *Reclinomonas americana* mitochondrial genome that are sufficiently well conserved for phylogenetic inference. They found that “the *Reclinomonas* protein branched with homologues from *Rickettsia* species in 5 trees, with homologues from *Wolbachia* in 10 trees, basal to *Rickettsia* and *Wolbachia* in 5 trees, with other α -proteobacteria or groupings thereof in 16 trees, and not with homologues from any α -proteobacterium in 19 trees with bootstrap proportions less than 70% for 53 of the 55 proteins studied” (reference [111] p. 1646). They furthermore surmised: “Recalling that the *Reclinomonas* mitochondrion inherited its genome from proteobacteria, rather than having acquired it through lateral acquisition from various donors, such disparate results could mean (1) that a degree of noise exists in the data (for example, due to poor conservation, as in the case of the twelve proteins that were excluded for lack of good homologues); (2) that the phylogenetic method is producing an imperfect estimation of the phylogeny, producing artifacts in some cases, but getting close to the true position in other cases; (3) that any number of problems inherent to phylogeny reconstruction, such as model misspecification or poor sampling, were present; (4) that the eubacteria sampled might be avidly exchanging these genes over time; or (5) any combination of the above.” (reference [111] p. 1646). That is very much the same thing as we are saying here.

With regard to plastid encoded proteins, similar studies are lacking to our knowledge, but for nuclear encoded genes in plants, Martin et al. [28] did find that phylogenetic trees for three cyanobacterial genomes “suggest at face value the *Arabidopsis* lineage to have acquired genes not from one cyanobacterium, but from all three sampled [even at a bootstrap probability (BP) ≥ 0.95], whereby that view

contradicts independent evidence suggesting a single origin of plastids from one cyanobacterium, not three or more in the *Arabidopsis* lineage”. We note that concatenation, which is very popular for the study of organelle genome phylogeny, condenses the many disparate signals that individual genes contain into a single averaged signal [29,30,32–34], while studies that investigate nuclear genes or (n.b.) seek evidence for gene transfer to eukaryotes are based on individual gene phylogenies [97]. Furthermore, concatenation harbors many pitfalls that are not yet well-understood [112]. Supertrees, which summarize the information from many trees in a single tree, support the participation of a plastid, a mitochondrion, and a host in eukaryote evolution, but no other extra partners [113]. Clearly, there is a need for additional critical studies of the phylogenetics of organelle origins.

Conclusion

Based on evidence from protein import machineries and phylogenies of eukaryotes and prokaryotes, plastids had a single origin involving a eukaryotic host and a cyanobacterial endosymbiont. This event is unique in that it is the only known successful transfer of oxygenic photosynthesis from cyanobacteria to another lineage. The origin of plastids might have involved nitrogen fixation during the initial physiological interactions of the symbiosis, as suggested by the overall similarity between genomes of diazotrophic cyanobacteria and plastid-derived nuclear genes in photosynthetic eukaryotes. Genomics and phylogenetics continue to enrich our understanding of plastid origin in the context of cyanobacterial evolution, but even with the copious amounts of data available, “core” genome phylogenies reported by different groups tend to conflict. This is because phylogenies are heavily model-dependent [12]. There is thus a need to look beyond the “1%” core genome phylogenies when addressing plastid origin. And if we keep lateral gene transfer among prokaryotes in mind – and the resulting fluid nature of prokaryotic chromosomes in the context of organelle origins [107–109] – we can make sense of the phylogenetic patterns observed in plant and algal genomes, without the need to infer spirochaete or chlamydial gene donors or the like. Understanding endosymbiosis in eukaryote genome evolution in the context of prokaryote genome evolution is a challenging exercise of keeping the bigger picture in focus.

Acknowledgments

This work was funded by the European Research Council (grant No. 232975 to WFM). CK is grateful to the Deutscher Akademischer Austauschdienst for a PhD stipend.

Authors' contributions

The following declarations about authors' contributions to the research have been made: compiled the literature: CK, MR, VZ, FS, SN-S, WFM; prepared the figure: CK; wrote the manuscript: CK, MR, VZ, FS, SN-S, WFM. CK and MR contributed equally to this work.

Competing interests

No competing interests have been declared.

References

- Allen JF. Photosynthesis of ATP – electrons, proton pumps, rotors, and poise. *Cell*. 2002;110(3):273–276. [http://dx.doi.org/10.1016/S0092-8674\(02\)00870-X](http://dx.doi.org/10.1016/S0092-8674(02)00870-X)
- Hagelstein P, Sieve B, Klein M, Jans H, Schultz G. Leucine synthesis in chloroplasts: leucine/isoleucine aminotransferase and valine aminotransferase are different enzymes in spinach chloroplasts. *J Plant Physiol*. 1997;150(1–2):23–30. [http://dx.doi.org/10.1016/S0176-1617\(97\)80176-9](http://dx.doi.org/10.1016/S0176-1617(97)80176-9)
- Zrenner R, Stitt M, Sonnewald U, Boldt R. Pyrimidine and purine biosynthesis and degradation in plants. *Annu Rev Plant Biol*. 2006;57(1):805–836. <http://dx.doi.org/10.1146/annurev.arplant.57.032905.105421>

4. Wang Z, Benning C. Chloroplast lipid synthesis and lipid trafficking through ER-plastid membrane contact sites. *Biochem Soc Trans.* 2012;40(2):457–463. <http://dx.doi.org/10.1042/BST20110752>
5. Gerdes S, Lerma-Ortiz C, Frelin O, Seaver SMD, Henry CS, de Crecy-Lagard V, et al. Plant B vitamin pathways and their compartmentation: a guide for the perplexed. *J Exp Bot.* 2012;63(15):5379–5395. <http://dx.doi.org/10.1093/jxb/ers208>
6. Mereschkowsky C. Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. *Biol Cent.* 1905;25(18):593–604.
7. Martin W, Kowallik K. Annotated English translation of Mereschkowsky's 1905 paper "Über Natur und Ursprung der Chromatophoren im Pflanzenreiche." *Eur J Phycol.* 1999;34(3):287–295. <http://dx.doi.org/10.1080/09670269910001736342>
8. Sagan L. On the origin of mitosing cells. *J Theor Biol.* 1967;14(3):225–274. [http://dx.doi.org/10.1016/0022-5193\(67\)90079-3](http://dx.doi.org/10.1016/0022-5193(67)90079-3)
9. Margulis L. Origin of eukaryotic cells. New Haven, CT: Yale University Press; 1970.
10. Wallin IE. Symbiogenesis and the origin of species. London: Tindall and Cox; 1927.
11. Sapp J. Evolution by association: a history of symbiosis. New York, NY: Oxford University Press; 1994.
12. Williams TA, Foster PG, Cox CJ, Embley TM. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature.* 2013;504(7479):231–236. <http://dx.doi.org/10.1038/nature12779>
13. Raff RA, Mahler HR. The non symbiotic origin of mitochondria. *Science.* 1972;177(4049):575–582. <http://dx.doi.org/10.1126/science.177.4049.575>
14. Bogorad L. Evolution of organelles and eukaryotic genomes. *Science.* 1975;188(4191):891–898. <http://dx.doi.org/10.1126/science.1138359>
15. Cavalier-Smith T. The origin of nuclei and of eukaryotic cells. *Nature.* 1975;256(5517):463–468. <http://dx.doi.org/10.1038/256463a0>
16. Gray MW, Doolittle WF. Has the endosymbiont hypothesis been proven? *Microbiol Rev.* 1982;46(1):1–42.
17. Raven PH. A multiple origin for plastids and mitochondria: many independent symbiotic events may have been involved in the origin of these cellular organelles. *Science.* 1970;169(3946):641–646. <http://dx.doi.org/10.1126/science.169.3946.641>
18. Howe CJ, Barbrook AC, Nisbet RER, Lockhart PJ, Larkum AWD. 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>
19. Stiller JW. Toward an empirical framework for interpreting plastid evolution. *J Phycol.* 2014;50(3):462–471. <http://dx.doi.org/10.1111/jpy.12178>
20. Zimorski V, Ku C, Martin WF, Gould SB. Endosymbiotic theory for organelle origins. *Curr Opin Microbiol.* 2014;22:38–48. <http://dx.doi.org/10.1016/j.mib.2014.09.008>
21. 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>
22. 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>
23. Steiner JM, Yusa F, Pompe JA, Löffelhardt W. Homologous protein import machineries in chloroplasts and cyanelles. *Plant J.* 2005;44(4):646–652. <http://dx.doi.org/10.1111/j.1365-313X.2005.02559.x>
24. Shi LX, Theg SM. The chloroplast protein import system: from algae to trees. *Biochim Biophys Acta.* 2013;1833(2):314–331. <http://dx.doi.org/10.1016/j.bbamcr.2012.10.002>
25. Martin W, Brinkmann H, Savonna C, Cerff R. Evidence for a chimeric nature of nuclear genomes: eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes. *Proc Natl Acad Sci USA.* 1993;90(18):8692–8696. <http://dx.doi.org/10.1073/pnas.90.18.8692>
26. 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>
27. Gould SB, Waller RF, McFadden GI. Plastid evolution. *Annu Rev Plant Biol.* 2008;59(1):491–517. <http://dx.doi.org/10.1146/annurev.arplant.59.032607.092915>
28. 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>
29. 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>
30. 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>
31. Schirrmeyer BE, Antonelli A, Bagheri HC. The origin of multicellularity in cyanobacteria. *BMC Evol Biol.* 2011;11(1):45. <http://dx.doi.org/10.1186/1471-2148-11-45>
32. 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>
33. 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>
34. Ochoa de Alda JAG, Esteban R, Diago ML, Houmard J. The plastid ancestor originated among one of the major cyanobacterial lineages. *Nat Commun.* 2014;5:4937. <http://dx.doi.org/10.1038/ncomms5937>
35. Katz LA, Grant JR, Parfrey LW, Burleigh JG. Turning the crown upside down: gene tree parsimony roots the eukaryotic tree of life. *Syst Biol.* 2012;61(4):653–660. <http://dx.doi.org/10.1093/sysbio/sys026>
36. 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>
37. Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J Gen Microbiol.* 1979;111(1):1–61. <http://dx.doi.org/10.1099/00221287-111-1-1>
38. 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>
39. Dagan T, Roettger M, Stucken K, Landan G, Koch R, Major P, et al. Genomes of stigonematalean cyanobacteria (subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids. *Genome Biol Evol.* 2013;5(1):31–44. <http://dx.doi.org/10.1093/gbe/evs117>
40. Mareš J, Hrouzek P, Kaňa R, Ventura S, Strunecký O, Komárek J. The primitive thylakoid-less cyanobacterium *Gloeobacter* is a common rock-dwelling organism. *PLoS ONE.* 2013;8(6):e66323. <http://dx.doi.org/10.1371/journal.pone.0066323>
41. Schirrmeyer BE, de Vos JM, Antonelli A, Bagheri HC. Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event. *Proc Natl Acad Sci USA.* 2013;110(5):1791–1796. <http://dx.doi.org/10.1073/pnas.1209927110>
42. Sánchez-Baracaldo P, Ridgwell A, Raven JA. A neoproterozoic transition in the marine nitrogen cycle. *Curr Biol.* 2014;24(6):652–657. <http://dx.doi.org/10.1016/j.cub.2014.01.041>
43. Soo RM, Skennerton CT, Sekiguchi Y, Imelfort M, Paech SJ, Dennis PG, et al. An expanded genomic representation of the phylum Cyanobacteria. *Genome Biol Evol.* 2014;6(5):1031–1045. <http://dx.doi.org/10.1093/gbe/evu073>
44. Mareš J, Komárek J, Compère P, Oren A. Validation of the generic name *Gloeobacter* Rippka et al. 1974, Cyanophyceae. *Cryptogam Algal.* 2013;34(3):255–262. <http://dx.doi.org/10.7872/crya.v34.iss3.2013.255>
45. Deschamps P, Colleoni C, Nakamura Y, Suzuki E, Pataux 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>

46. Deusch O, Landan G, Roettger M, Gruenheit N, Kowallik KV, Allen JF, et al. Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. *Mol Biol Evol.* 2008;25(4):748–761. <http://dx.doi.org/10.1093/molbev/msn022>
47. 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.
48. Knoll AH. Paleobiological perspectives on early eukaryotic evolution. *Cold Spring Harb Perspect Biol.* 2014;6(1):a016121. <http://dx.doi.org/10.1101/cshperspect.a016121>
49. Butterfield NJ. *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology.* 2000;26(3):386–404. [http://dx.doi.org/10.1666/0094-8373\(2000\)026<0386:BPNGNS>2.CO;2](http://dx.doi.org/10.1666/0094-8373(2000)026<0386:BPNGNS>2.CO;2)
50. Yoon HS. 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>
51. Parfrey LW, Lahr DJG, Knoll AH, Katz LA. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc Natl Acad Sci USA.* 2011;108(33):13624–13629. <http://dx.doi.org/10.1073/pnas.1110633108>
52. Martin W, Stoebe B, Goremykin V, Hansmann S, Hasegawa M, Kowallik KV. Gene transfer to the nucleus and the evolution of chloroplasts. *Nature.* 1998;393(6681):162–165. <http://dx.doi.org/10.1038/30234>
53. Lockhart PJ, Howe CJ, Barbrook AC, Larkum AWD, Penny D. Spectral analysis, systematic bias, and the evolution of chloroplasts. *Mol Biol Evol.* 1999;16(4):573.
54. Petersen J, Brinkmann H, Bunk B, Michael V, Päuker O, Pradella S. Think pink: photosynthesis, plasmids and the *Roseobacter* clade: plasmids and phototrophy. *Environ Microbiol.* 2012;14(10):2661–2672. <http://dx.doi.org/10.1111/j.1462-2920.2012.02806.x>
55. Schubert WD, Klukas O, Saenger W, Witt HT, Fromme P, Krauß N. A common ancestor for oxygenic and anoxygenic photosynthetic systems. *J Mol Biol.* 1998;280(2):297–314. <http://dx.doi.org/10.1006/jmbi.1998.1824>
56. Sadekar S. Conservation of distantly related membrane proteins: photosynthetic reaction centers share a common structural core. *Mol Biol Evol.* 2006;23(11):2001–2007. <http://dx.doi.org/10.1093/molbev/msl079>
57. Allen JF. A redox switch hypothesis for the origin of two light reactions in photosynthesis. *FEBS Lett.* 2005;579(5):963–968. <http://dx.doi.org/10.1016/j.febslet.2005.01.015>
58. Olson JM, Blankenship RE. Thinking about the evolution of photosynthesis. *Photosynth Res.* 2004;80(1–3):373–386. <http://dx.doi.org/10.1023/B:PRES.0000030457.06495.83>
59. Sousa FL, Shavit-Grievink L, Allen JF, Martin WF. Chlorophyll biosynthesis gene evolution indicates photosystem gene duplication, not photosystem merger, at the origin of oxygenic photosynthesis. *Genome Biol Evol.* 2013;5(1):200–216. <http://dx.doi.org/10.1093/gbe/evs127>
60. Blankenship RE. Molecular evidence for the evolution of photosynthesis. *Trends Plant Sci.* 2001;6(1):4–6. [http://dx.doi.org/10.1016/S1360-1385\(00\)01831-8](http://dx.doi.org/10.1016/S1360-1385(00)01831-8)
61. Hohmann-Marriott MF, Blankenship RE. Evolution of photosynthesis. *Annu Rev Plant Biol.* 2011;62(1):515–548. <http://dx.doi.org/10.1146/annurev-arplant-042110-103811>
62. Oren A, Padan E. Induction of anaerobic, photoautotrophic growth in the cyanobacterium *Oscillatoria limnetica*. *J Bacteriol.* 1978;133(2):558–563.
63. Blankenship RE, Hartman H. The origin and evolution of oxygenic photosynthesis. *Trends Biochem Sci.* 1998;23(3):94–97. [http://dx.doi.org/10.1016/S0968-0004\(98\)01186-4](http://dx.doi.org/10.1016/S0968-0004(98)01186-4)
64. Raymond J, Blankenship R. The origin of the oxygen-evolving complex. *Coord Chem Rev.* 2008;252(3–4):377–383. <http://dx.doi.org/10.1016/j.ccr.2007.08.026>
65. Dismukes GC, Klimov VV, Baranov SV, Kozlov YN, DasGupta J, Tyrshkin A. The origin of atmospheric oxygen on Earth: the innovation of oxygenic photosynthesis. *Proc Natl Acad Sci USA.* 2001;98(5):2170–2175. <http://dx.doi.org/10.1073/pnas.061514798>
66. Sauer K, Yachandra VK. A possible evolutionary origin for the Mn₄ cluster of the photosynthetic water oxidation complex from natural MnO₂ precipitates in the early ocean. *Proc Natl Acad Sci USA.* 2002;99(13):8631–8636. <http://dx.doi.org/10.1073/pnas.132266199>
67. Allen JF, Martin W. Evolutionary biology: out of thin air. *Nature.* 2007;445(7128):610–612. <http://dx.doi.org/10.1038/445610a>
68. Hakala M. Photoinhibition of manganese enzymes: insights into the mechanism of photosystem II photoinhibition. *J Exp Bot.* 2006;57(8):1809–1816. <http://dx.doi.org/10.1093/jxb/erj189>
69. Kupitz C, Basu S, Grotjohann I, Fromme R, Zatsepin NA, Rendek KN, et al. Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser. *Nature.* 2014;513(7517):261–265. <http://dx.doi.org/10.1038/nature13453>
70. Johnson JE, Webb SM, Thomas K, Ono S, Kirschvink JL, Fischer WW. Manganese-oxidizing photosynthesis before the rise of cyanobacteria. *Proc Natl Acad Sci USA.* 2013;110(28):11238–11243. <http://dx.doi.org/10.1073/pnas.1305530110>
71. Khorobrykh A, Dasgupta J, Kolling DRJ, Terentyev V, Klimov VV, Dismukes GC. Evolutionary origins of the photosynthetic water oxidation cluster: bicarbonate permits Mn²⁺ photo-oxidation by anoxygenic bacterial reaction centers. *Chembiochem.* 2013;14(14):1725–1731. <http://dx.doi.org/10.1002/cbic.201300355>
72. Allen JP, Olson TL, Oyala P, Lee WJ, Tufts AA, Williams JC. Light-driven oxygen production from superoxide by Mn-binding bacterial reaction centers. *Proc Natl Acad Sci USA.* 2012;109(7):2314–2318. <http://dx.doi.org/10.1073/pnas.1115364109>
73. Roca G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA, et al. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature.* 2003;424(6952):1042–1047. <http://dx.doi.org/10.1038/nature01947>
74. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. *Nature.* 2000;405(6784):299–304. <http://dx.doi.org/10.1038/35012500>
75. Dagan T, Artzy-Randrup Y, Martin W. Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc Natl Acad Sci USA.* 2008;105(29):10039–10044. <http://dx.doi.org/10.1073/pnas.0800679105>
76. Doolittle WF. Phylogenetic classification and the universal tree. *Science.* 1999;284(5423):2124–2128. <http://dx.doi.org/10.1126/science.284.5423.2124>
77. Doolittle WF, Baptiste E. Pattern pluralism and the Tree of Life hypothesis. *Proc Natl Acad Sci USA.* 2007;104(7):2043–2049. <http://dx.doi.org/10.1073/pnas.0610699104>
78. Paul S, Dutta A, Bag SK, Das S, Dutta C. Distinct, ecotype-specific genome and proteome signatures in the marine cyanobacteria *Prochlorococcus*. *BMC Genomics.* 2010;11(1):103. <http://dx.doi.org/10.1186/1471-2164-11-103>
79. Thiergart T, Landan G, Schenk M, Dagan T, Martin WF. An evolutionary network of genes present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial origin. *Genome Biol Evol.* 2012;4(4):466–485. <http://dx.doi.org/10.1093/gbe/evs018>
80. Martin W, Müller M. The hydrogen hypothesis for the first eukaryote. *Nature.* 1998;392(6671):37–41. <http://dx.doi.org/10.1038/32096>
81. McInerney JO, O'Connell MJ, Pisani D. The hybrid nature of the eukaryota and a consilient view of life on Earth. *Nat Rev Microbiol.* 2014;12(6):449–455. <http://dx.doi.org/10.1038/nrmicro3271>
82. Williams TA, Embley TM. Archaeal “dark matter” and the origin of eukaryotes. *Genome Biol Evol.* 2014;6(3):474–481. <http://dx.doi.org/10.1093/gbe/evu031>
83. Holland HD. The oxygenation of the atmosphere and oceans. *Philos Trans R Soc Lond B Biol Sci.* 2006;361(1470):903–915. <http://dx.doi.org/10.1098/rstb.2006.1838>
84. Johnston DT, Wolfe-Simon F, Pearson A, Knoll AH. Anoxygenic photosynthesis modulated Proterozoic oxygen and sustained Earth's middle age. *Proc Natl Acad Sci USA.* 2009;106(40):16925–16929. <http://dx.doi.org/10.1073/pnas.0909248106>

85. Kasting J. Earth's early atmosphere. *Science*. 1993;259(5097):920–926. <http://dx.doi.org/10.1126/science.11536547>
86. Rai AN, Söderbäck E, Bergman B. Cyanobacterium-plant symbioses. *New Phytol*. 2000;147(3):449–481. <http://dx.doi.org/10.1046/j.1469-8137.2000.00720.x>
87. Rikkinen J. Lichen guilds share related cyanobacterial symbionts. *Science*. 2002;297(5580):357–357. <http://dx.doi.org/10.1126/science.1072961>
88. Prechtel J. Intracellular spheroid bodies of *Rhopalodia gibba* have nitrogen-fixing apparatus of cyanobacterial origin. *Mol Biol Evol*. 2004;21(8):1477–1481. <http://dx.doi.org/10.1093/molbev/msh086>
89. Ran L, Larsson J, Vigil-Stenman T, Nylander JAA, Ininbergs K, Zheng WW, et al. Genome erosion in a nitrogen-fixing vertically transmitted endosymbiotic multicellular cyanobacterium. *PLoS ONE*. 2010;5(7):e11486. <http://dx.doi.org/10.1371/journal.pone.0011486>
90. Chiu WL. Nitrogen deprivation stimulates symbiotic gland development in *Gunnera manicata*. *Plant Physiol*. 2005;139(1):224–230. <http://dx.doi.org/10.1104/pp.105.064931>
91. Rai AN, Bergman B, Rasmussen U, editors. *Cyanobacteria in symbiosis*. Dordrecht: Kluwer Academic Publishers; 2002.
92. Costa JL, Romero EM, Lindblad P. Sequence based data supports a single *Nostoc* strain in individual coralloid roots of cycads. *FEMS Microbiol Ecol*. 2004;49(3):481–487. <http://dx.doi.org/10.1016/j.femsec.2004.05.001>
93. Allen JF, Raven JA. Free-radical-induced mutation vs redox regulation: costs and benefits of genes in organelles. *J Mol Evol*. 1996;42(5):482–492. <http://dx.doi.org/10.1007/BF02352278>
94. 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>
95. Falkowski PG, Fenchel T, Delong EF. The microbial engines that drive Earth's biogeochemical cycles. *Science*. 2008;320(5879):1034–1039. <http://dx.doi.org/10.1126/science.1153213>
96. Usher KM, Bergman B, Raven JA. Exploring cyanobacterial mutualisms. *Annu Rev Ecol Evol Syst*. 2007;38(1):255–273. <http://dx.doi.org/10.1146/annurev.ecolsys.38.091206.095641>
97. Facchinelli F, Colleoni C, Ball SG, Weber APM. Chlamydia, cyanobiont, or host: who was on top in the ménage à trois? *Trends Plant Sci*. 2013;18(12):673–679. <http://dx.doi.org/10.1016/j.tplants.2013.09.006>
98. Salzberg SL. Microbial genes in the human genome: lateral transfer or gene loss? *Science*. 2001;292(5523):1903–1906. <http://dx.doi.org/10.1126/science.1061036>
99. Rumpho ME, Worful JM, Lee J, Kannan K, Tyler MS, Bhattacharya D, et al. Horizontal gene transfer of the algal nuclear gene *psbO* to the photosynthetic sea slug *Elysia chlorotica*. *Proc Natl Acad Sci USA*. 2008;105(46):17867–17871. <http://dx.doi.org/10.1073/pnas.0804968105>
100. Wägele H, Deusch O, Handeler K, Martin R, Schmitt V, Christa G, et al. Transcriptomic evidence that longevity of acquired plastids in the photosynthetic slugs *Elysia timida* and *Plakobranthus ocellatus* does not entail lateral transfer of algal nuclear genes. *Mol Biol Evol*. 2011;28(1):699–706. <http://dx.doi.org/10.1093/molbev/msq239>
101. de Vries J, Christa G, Gould SB. Plastid survival in the cytosol of animal cells. *Trends Plant Sci*. 2014;19(6):347–350. <http://dx.doi.org/10.1016/j.tplants.2014.03.010>
102. Hannaert V, Saavedra E, Duffieux F, Szikora JP, Rigden DJ, Michels PAM, et al. Plant-like traits associated with metabolism of *Trypanosoma parasites*. *Proc Natl Acad Sci USA*. 2003;100(3):1067–1071. <http://dx.doi.org/10.1073/pnas.0335769100>
103. Martin W, Borst P. Secondary loss of chloroplasts in trypanosomes. *Proc Natl Acad Sci USA*. 2003;100(3):765–767. <http://dx.doi.org/10.1073/pnas.0437776100>
104. Berriman M. The genome of the African trypanosome *Trypanosoma brucei*. *Science*. 2005;309(5733):416–422. <http://dx.doi.org/10.1126/science.1112642>
105. Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M, Wortman JR, et al. Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol*. 2006;4(9):e286. <http://dx.doi.org/10.1371/journal.pbio.0040286>
106. Moreira D, Deschamps P. What was the real contribution of endosymbionts to the eukaryotic nucleus? Insights from photosynthetic eukaryotes. *Cold Spring Harb Perspect Biol*. 2014;6(7):a016014. <http://dx.doi.org/10.1101/cshperspect.a016014>
107. Martin W. Mosaic bacterial chromosomes: a challenge en route to a tree of genomes. *Bioessays*. 1999;21(2):99–104. [http://dx.doi.org/10.1002/\(SICI\)1521-1878\(199902\)21:2<99::AID-BIES3>3.0.CO;2-B](http://dx.doi.org/10.1002/(SICI)1521-1878(199902)21:2<99::AID-BIES3>3.0.CO;2-B)
108. Esser C, Martin W, Dagan T. The origin of mitochondria in light of a fluid prokaryotic chromosome model. *Biol Lett*. 2007;3(2):180–184. <http://dx.doi.org/10.1098/rsbl.2006.0582>
109. Richards TA, Archibald JM. Cell evolution: gene transfer agents and the origin of mitochondria. *Curr Biol*. 2011;21(3):R112–R114. <http://dx.doi.org/10.1016/j.cub.2010.12.036>
110. Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, et al. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature*. 1997;387(6632):493–497. <http://dx.doi.org/10.1038/387493a0>
111. Esser C. A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol Biol Evol*. 2004;21(9):1643–1660. <http://dx.doi.org/10.1093/molbev/msh160>
112. Thiergart T, Landan G, Martin WF. Concatenated alignments and the case of the disappearing tree. *BMC Evol Biol*. 2015 (in press).
113. Pisani D, Cotton JA, McInerney JO. Supertrees disentangle the chimerical origin of eukaryotic genomes. *Mol Biol Evol*. 2007;24(8):1752–1760. <http://dx.doi.org/10.1093/molbev/msm095>