

Research article

Open Access

## Evidence of a chimeric genome in the cyanobacterial ancestor of plastids

Jeferson Gross<sup>1</sup>, Jörg Meurer<sup>2</sup> and Debashish Bhattacharya<sup>\*1</sup>

Address: <sup>1</sup>University of Iowa, Department of Biological Sciences and the Roy J. Carver Center for Comparative Genomics, 446 Biology Building, Iowa City, Iowa 52242, USA and <sup>2</sup>Department Biology I, Botany, Ludwig-Maximilians-University Munich, Menzinger Str. 67, 80638 Munich, Germany

Email: Jeferson Gross - [jeferson-gross@uiowa.edu](mailto:jeferson-gross@uiowa.edu); Jörg Meurer - [joerg.meurer@lrz.uni-muenchen.de](mailto:joerg.meurer@lrz.uni-muenchen.de); Debashish Bhattacharya<sup>\*</sup> - [debashibhattacharya@uiowa.edu](mailto:debashibhattacharya@uiowa.edu)

<sup>\*</sup> Corresponding author

Published: 23 April 2008

Received: 3 March 2008

BMC Evolutionary Biology 2008, 8:117 doi:10.1186/1471-2148-8-117

Accepted: 23 April 2008

This article is available from: <http://www.biomedcentral.com/1471-2148/8/117>

© 2008 Gross et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Horizontal gene transfer (HGT) is a vexing fact of life for microbial phylogeneticists. Given the substantial rates of HGT observed in modern-day bacterial chromosomes, it is envisaged that ancient prokaryotic genomes must have been similarly chimeric. But where can one find an ancient prokaryotic genome that has maintained its ancestral condition to address this issue? An excellent candidate is the cyanobacterial endosymbiont that was harnessed over a billion years ago by a heterotrophic protist, giving rise to the plastid. Genetic remnants of the endosymbiont are still preserved in plastids as a highly reduced chromosome encoding 54 – 264 genes. These data provide an ideal target to assess genome chimericism in an ancient cyanobacterial lineage.

**Results:** Here we demonstrate that the origin of the plastid-encoded gene cluster for menaquinone/phyloquinone biosynthesis in the extremophilic red algae Cyanidiales contradicts a cyanobacterial genealogy. These genes are relics of an ancestral cluster related to homologs in Chlorobi/Gammaproteobacteria that we hypothesize was established by HGT in the progenitor of plastids, thus providing a 'footprint' of genome chimericism in ancient cyanobacteria. In addition to *menB*, four components of the original gene cluster (*menF*, *menD*, *menC*, and *menH*) are now encoded in the nuclear genome of the majority of non-Cyanidiales algae and plants as the unique tetra-gene fusion named *PHYLL*. These genes are monophyletic in Plantae and chromalveolates, indicating that loci introduced by HGT into the ancestral cyanobacterium were moved over time into the host nucleus.

**Conclusion:** Our study provides unambiguous evidence for the existence of genome chimericism in ancient cyanobacteria. In addition we show genes that originated via HGT in the cyanobacterial ancestor of the plastid made their way to the host nucleus via endosymbiotic gene transfer (EGT).

### Background

Bacterial chromosomes are subject to a constant influx of genes through horizontal gene transfer (HGT) that often occurs from phylogenetically distant organisms [1]. When

foreign genes confer a selective advantage they are maintained by the host organism and vertically transmitted to descendants. Consequently, over evolutionary time the continual introduction of foreign genes into bacterial

chromosomes is expected to result in a highly chimeric genome. Estimates of the frequency of HGT in prokaryotes are variable depending on the bioinformatic approach used, the organism(s) analysed, and inherent limits to the detection of intraphylum HGT among closely related taxa [1-4]. Overall, percentages of genes acquired by HGT in a typical prokaryote tend to average 2–25% of the total genome [1-4] but exceptional cases are known with values as low as 1.6% in *Mycoplasma genitalium* or as high as 32.6% in *Treponema pallidum* [1]. Regardless of the actual number that is cited, the magnitude of observed HGTs demonstrates that a large fraction of genes in extant prokaryotes originated via lateral transfer. The question remains open whether HGT might have had a similar, or even more dramatic impact on early bacterial evolution. The genomes of ancient prokaryotes could potentially differ substantially in terms of gene content and phylogenetic diversity from their present-day counterparts [4,5]. These considerations are particularly intriguing when formulated in the context of the endosymbiotic origin of the plastid that is derived from the capture of a cyanobacterium ca. 1.5 billion of years ago [6,7]. This endosymbiotic occurred in the founding lineage of photosynthetic eukaryotes, the Plantae [8]. Here we ask the question, is there evidence of chimerism in plastid genes that would provide insights into ancient cyanobacterial genomes? The plastid is a 'living fossil' ideally suited for this purpose because it has largely been protected from HGT within the eukaryotic host [9] (for exceptions see below), thereby conserving its ancient genome characteristics.

The plastid genome is a highly reduced version of the cyanobacterial endosymbiont [8], retaining between 54–264 genes [10]. Phylogenetic surveys of conserved plastid genes convincingly demonstrate their cyanobacterial ancestry [11,12]. Furthermore, phylogenomic analyses indicate that ca. 500–1500 cyanobacterial genes were transferred to the nucleus (endosymbiotic gene transfer, EGT) and their gene products are largely targeted back to the plastid [13-15]. Three known cases contradict the notion of a strictly cyanobacterial ancestry for plastid genes. The first two are the *rbcL* gene in red algae and chromalveolates and the *rpl36* gene shared uniquely by cryptophytes and haptophytes that are phylogenetically most closely related to homologs within the Alphaproteobacteria and to a Planctomycete related to *Rhodospirillum rubrum*, respectively [9,16]. The third case is a firmicute-derived gene for the tau/gamma subunit of DNA polymerase III (*dnaX*) encoded in the plastid genome of the cryptophyte *Rhodomonas salina* [17]. *DnaX* is absent from all other sequenced plastid genomes including that of another cryptophyte (*Guillardia theta*) [17]. Their limited distribution (*rpl36*, *dnaX*) and nested relationship to Alphaproteobacteria (*rbcL*) suggest that these HGT events are unlikely to be examples of genome chimerism that was

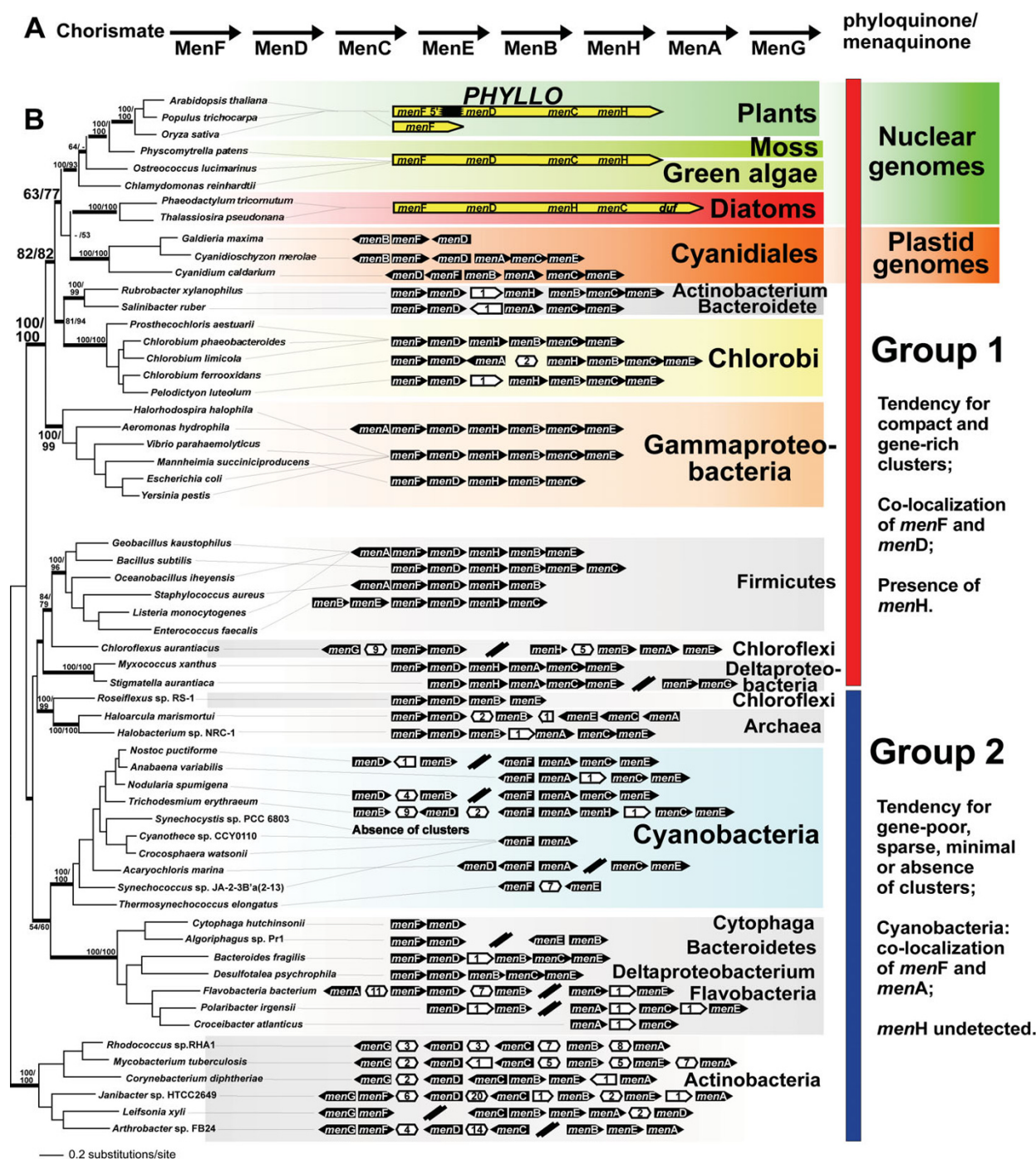
vertically inherited from the captured cyanobacterium followed by differential loss in Plantae [9,16,17]. In contrast, here we provide phylogenetic evidence that plastid and nuclear-encoded proteins for the biosynthesis of phylloquinone (PhQ), a redox cofactor of the Photosystem I complex in algae and plants, may represent surviving traces of a chimeric genome in the captured endosymbiont.

## Results

### **Homologous genes for PhQ biosynthesis are plastid-encoded in Cyanidiales but nuclear-encoded in the majority of photosynthetic eukaryotes**

Biosynthesis of the Photosystem I cofactor PhQ occurs in cyanobacteria, plants, and algae and is analogous to that of menaquinone (MQ), a mobile electron carrier in many bacterial bioenergetic systems [18]. Both biosynthetic pathways have been characterized at the genetic and biochemical level in prokaryotes and eukaryotes and depend on 8 enzymatic steps catalyzed by the so-called 'Men proteins' (Figure 1a) [18-23]. In MQ-using bacteria and cyanobacteria, *men* homologs are present as individual genes usually grouped within operons or in proximally encoded transcripts (Figure 1b). In available genomes of plants, green algae, and stramenopiles *menA*, *menB*, *menE*, and *menG* are nuclear-encoded and constitute independent loci. Genes encoding MenF, MenD, MenC, and MenH tend to be united (with some rearrangements) as a single nuclear-encoded, composite gene termed *PHYLLLO* (Figure 1b and Materials and methods) [18]. In higher plants *menF* (*isochorismate synthase*, ICS) is an independent nuclear-encoded gene that is distal to *PHYLLLO* (Figure 1b) [18].

In contrast to their nuclear location in most algae and plants, *men* genes are plastid-encoded in the early diverging Cyanidiales red algae *Cyanidium caldarium* and *Cyanidioschyzon merolae* (Figure 1b). In these taxa, the *men* homologs (including *menF*, *menD*, and *menC* that are fused in nuclear encoded *PHYLLLO*) are present as individual genes and grouped in a compact cluster that resembles the structure of prokaryotic operons. The nucleotide sequence of the plastid *men* cluster of the Cyanidiales alga *Galdieria maxima* has been partially determined in this study and displays a similar prokaryotic-like genetic organization that appears to be syntenic with the locus in *Cyanidioschyzon merolae* (Figure 1b). Collectively, these observations lead us to postulate that plastid and nuclear-encoded *men* genes in algae and plants are of endosymbiont origin and the nuclear location of most genes is due to EGT.

**Figure 1**

**PhQ/MQ metabolic pathway and phylogeny of MenFD proteins.** (a) Shown are the Men proteins that control the enzymatic steps necessary for the conversion of chorismate to the final product MQ in bacteria and PhQ in cyanobacteria and in photosynthetic eukaryotes. (b) Bayesian majority rule consensus tree of a concatenated alignment of MenF and MenD proteins. The nodal numbers represent bootstrap support values inferred using PHYML (left of slash mark) and neighbor joining (1000 replicates, right of slash mark). Only bootstrap values >50% are shown. For the sake of clarity only support values at nodes of highest interest have been included in the figure. The thick branches have a BPP > 0.95. The branch lengths in this tree are proportional to the number of substitutions per site (see scale in figure). Men genes encoded as clusters in chromosomes of prokaryotes and in the plastid genome of Cyanidiales (black boxes), as well as the architecture of the *PHYLLLO* gene (yellow boxes) in nuclear genomes of photosynthetic eukaryotes, are indicated for each taxon. The break in the 5' terminus of *PHYLLLO* in higher plants indicates a gene-splitting event during evolution [18]. In addition, in plants *menF* is an individual gene distinct from *PHYLLLO*. The *men* gene cluster in the plastid of *Galdieria maxima* was only partially sequenced. White boxes with numbers inside indicate the number of genes with functions unrelated to the menaquinone biosynthesis that separate the *men* genes. Double slashes indicate a large chromosomal separation between *men* genes. According to their structure, gene clusters can be divided in two groups (Group 1 and Group 2) that are correlated with the tree topology. This tree was arbitrarily rooted on the branch leading to the Actinobacteria.

**MenF, menD, menC, menB, menE, and menH were originated via a single HGT event in the ancestor of plastids**

To test the hypothesis of a common endosymbiotic origin of algal and plant *men* genes, we first inferred separate maximum likelihood phylogenies for MenF and MenD. These trees were essentially congruent (results not shown) therefore we concatenated the proteins to gain greater phylogenetic resolution. To generate the alignment, we used the two proteins that are already fused in PHYLO in algae, whereas in plants we concatenated MenF (i.e., the ICS1 paralog in *Arabidopsis*) with MenD in PHYLO. The resulting MenFD tree (Figure 1b) recovers the monophyly of PHYLO and MenFD from the Cyanidiales (Bayesian posterior probability, BPP = 0.98, PHYML bootstrap, PB = 3%, neighbor joining bootstrap, NJB = 77%), consistent with the idea that the tetra-fusion PHYLO arose from an ancestral plastid gene cluster whose remnants are still present in the plastid genome of Cyanidiales. The distribution of operon characteristics (not their overall phylogeny) supports the existence of two superclades that we provisionally named Group 1 and Group 2. Group 1 contains compact, gene-rich *men* gene clusters, the co-localization of *menF* and *menD*, and contains *menH*, whereas Group 2 is defined by relatively gene-poor, dispersed (as in cyanobacteria) clusters. Group 1 contains the unexpected phylogenetic relationship between MenFD in algae and plants with homologs from Chlorobi, Gammaproteobacteria, *Rubrobacter xylanophilus*, and *Salinibacter ruber* (BPP = 1.0, PB = 100%, NJB = 100%). These taxa also share a similar operon structure and gene content. Cyanobacterial MenFD in Group 2 is distantly related to the algal/plant proteins forming a sister group to Bacteroidetes and Flavobacteria. The approximately unbiased (AU) test was used to reposition the cyanobacterial clade on all other basal (20 in total) branches in the tree. Using this approach, the union of cyanobacteria with the algal/plant clade in Group 1 was significantly rejected by the AU-test ( $P < 0.001$ ). Together these results indicate that the phylogenetic history of MenF and MenD proteins is not congruent with a cyanobacterial ancestry and most likely reflects an ancient HGT event.

The topology recovered for the MenC, MenB, MenE, and MenH trees (see Additional files 1, 2, 3) recapitulates the MenFD phylogeny. MenC (part of PHYLO) and MenB are nuclear-encoded in photosynthetic eukaryotes. These proteins and the homologs encoded in the plastid genome of Cyanidiales are phylogenetically most closely related to sequences in the MQ operon of Chlorobi and Gammaproteobacteria. This is also the case for plastid-encoded MenE in Cyanidiales (see Additional file 3) that is nested in a common clade with homologs in Chlorobi/Gammaproteobacteria. In contrast, MenE is encoded in the nuclear genome of *Arabidopsis* and other plants and

algae and branches with Deltaproteobacteria suggesting a phylogenetic origin that is distinct from the Cyanidiales gene (see Additional file 3). Following the trend shown for MenFD (Figure 1b), in the trees of MenC, MenB, and MenE the cyanobacterial homologs are phylogenetically distantly related to the algal/plant clade (see Additional files 1, 2, 3). MenH proteins are almost exclusively found in taxa of Group 1 (for exceptions see below) and the homologs of plants and algae are nested (BPP = 1.0, PB = 96%) among Chlorobi and Gammaproteobacteria homologs (Figure 2). The consistent phylogenetic pattern observed for the six first enzymes of the PhQ biosynthesis in plants and algae (with the exception of the nuclear encoded MenE) suggests that they have the same origin. This implies that the cyanobacterial ancestor of the primary plastid acquired a complete *men* gene cluster (presumably as an operon) via HGT from a single source related to Chlorobi/Gammaproteobacteria. That Men homologs of *Rubrobacter xylanophilus* and *Salinibacter ruber* are also nested within Group 1 representatives and are distantly associated to their taxonomic relatives (Actinobacteria and Bacteroidetes homologs are in the Group 2) suggests that these taxa also gained their *men* operon from a donor related to the source of the algal/plant genes (Figure 1b).

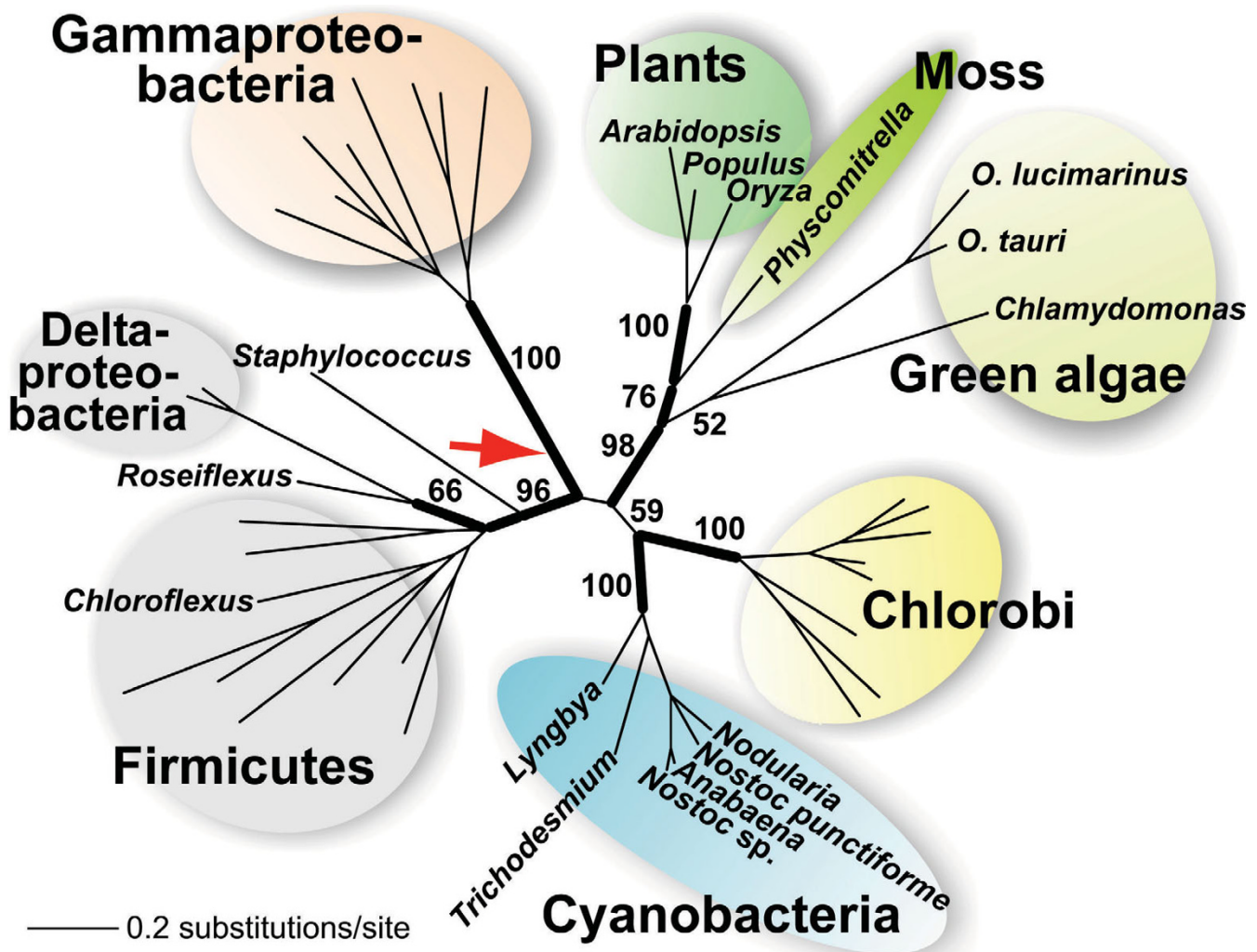
**Independent acquisition of menH by cyanobacteria from a source related to Chlorobi**

In contrast to the general absence of a detectable *menH* gene from taxa of Group 2 (Figure 1b), we were able to identify this gene in members of the cyanobacterial families Oscillatoriales and Nostocales. In *Trichodesmium erythraeum* (Figure 1b) *menH* is encoded within the *men* gene cluster. Surprisingly, phylogenetic analysis of MenH (Figure 2) demonstrates that cyanobacteria have acquired this gene from Chlorobi-related bacteria. This tree shows that algal and plant MenH is not closely related to cyanobacterial homologs. Therefore, the *menH* genes in Oscillatoriales and Nostocales are not vestiges of the ancient gene cluster that was putatively transferred via primary endosymbiosis into eukaryotes but rather is a more recent independent HGT event in some groups of cyanobacteria. This result illustrates the feasibility of *men* gene transfer into cyanobacteria and supports our model of an ancient gene cluster acquisition by the plastid ancestor.

**The MenA phylogeny supports additional HGT events during plastid evolution**

A *menA* gene is also present in the *men* cluster in the plastid genome of Cyanidiales (Figure 1b), whereas it is nuclear encoded in other algae and plants. The phylogeny of MenA reveals however that the Cyanidiales gene is not closely related to homologs in cyanobacteria, Chlorobi, or to the majority of Gammaproteobacteria (Figure 3). Instead, Cyanidiales MenA is monophyletic (BPP = 1.0)

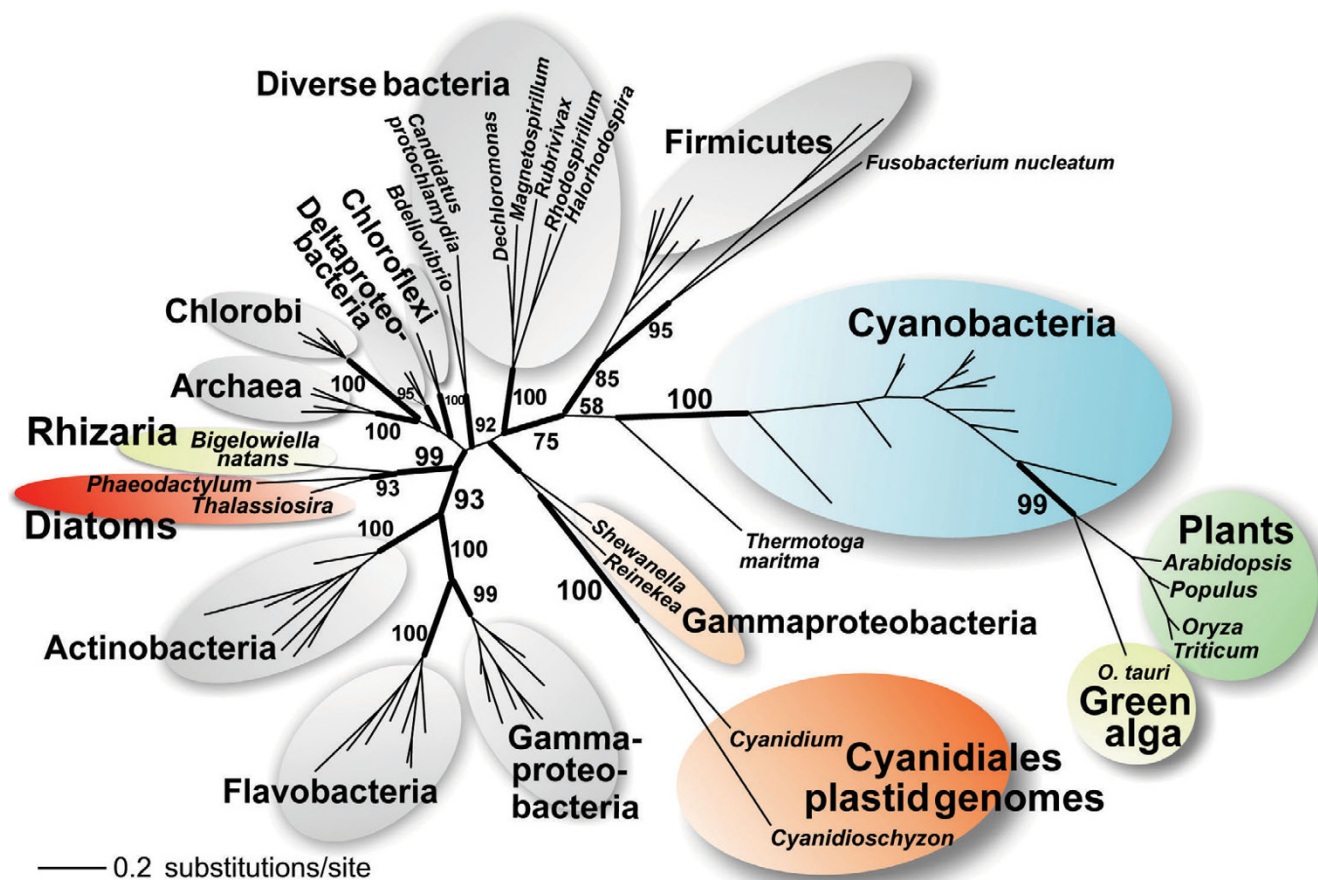


**Figure 2**

**Unrooted Bayesian majority rule consensus tree of MenH.** Chlorobi and the cyanobacterial groups Oscillatoriales and Nostocales form sister clades, suggesting a HGT between these taxa. The position of diatoms (i.e., *T. pseudonana* and *P. tricornutum*) MenH is indicated by the red arrow. These highly diverged long branched sequences were excluded from the final analysis. The numbers at the nodes are PHYML bootstrap values (500 replicates). Only bootstrap values >50% are shown. For the sake of clarity only support values at nodes of highest interest have been included in the figure. The thick branches have a BPP > 0.95. The branch lengths in this tree are proportional to the number of substitutions per site (see scale in figure).

with two Gammaproteobacteria (*Shewanella woodyi* and *Reinekea* sp. MED297) that are distantly related to the other members of this bacterial clade. The AU-test in which the Cyanidiales MenA clade was moved to 20 alternate basal branches in the tree significantly rejects the sister group relationship of this clade to cyanobacteria ( $P = 0.017$ ), Chlorobi ( $P = 0.009$ ), or Gammaproteobacteria ( $P = 0.001$ ). These results imply that the *menA* gene and the *men* gene cluster were obtained in different HGT events, involving distinct donor taxa. The *menA* gene has been incorporated into the gene cluster and exclusively maintained in the plastid genome of Cyanidiales with the

remainder of the *men* genes. Furthermore, the phylogeny of MenA demonstrates that the homologs encoded in the genome of green algae and plants are of cyanobacterial descent (BPP = 1.0, PB = 100%). One possible explanation of our data is that two *menA* genes were present in the prokaryotic endosymbiont, one of cyanobacterial provenance and other introduced by HGT from another distantly related source. These genes were differentially lost from Cyanidiales and members of the green lineage with the former retaining the plastid-encoded non-cyanobacterial gene and the latter retaining the cyanobacterial copy that was transferred to the nucleus. An alternative expla-

**Figure 3**

**Unrooted Bayesian majority rule consensus tree of MenA.** Green algae and plants have a cyanobacterial form of MenA, whereas plastid-encoded MenA in Cyanidiales is unrelated to cyanobacteria and likely originated from a HGT event from a distantly related source that introduced (and thereby duplicated) this gene in the genome of the plastid ancestor. The numbers at the branches are PHYML bootstrap values (500 replicates). Only bootstrap values >50% are shown. For the sake of clarity only support values at nodes of highest interest have been included in the figure. The thick branches have a BPP > 0.95. The branch lengths in this tree are proportional to the number of substitutions per site (see scale in figure).

nation is that the *menA* gene in the green lineage was directly acquired from cyanobacteria via HGT into the nucleus of these taxa rather than by EGT from the endosymbiont. Under this scenario, the HGT event was followed by loss of the plastid encoded *menA* gene in the ancestor of green algae. *MenA* is not plastid-encoded in other non-Cyanidiales red algae such as *Gracilaria tenuitipitata* therefore it is unknown which gene copy is nuclear-encoded in these taxa.

*MenA* in diatoms and the Rhizarian *Bigelowiella natans* are nuclear-encoded and form a monophyletic group (BPP = 1.0, PB = 99%) that is not phylogenetically closely related to homologs in Cyanidiales or the green lineage (Figure 3). These genes define another HGT event that likely occurred in the nucleus of these taxa involving an unknown bacterial source. Chromalveolates (e.g., dia-

toms) and Rhizaria have recently been proposed to be monophyletic [24] and the *MenA* tree is consistent with this idea, showing a unique HGT event shared by these taxa. *MenG* shows a similar phylogenetic history to *MenA*, except that *MenG* is no longer plastid-encoded. This gene is present in the nucleus of all photosynthetic lineages and the phylogeny of *MenG* provides strong support for a cyanobacterial provenance in the green lineage and an origin in Cyanidiales and diatoms from Deltaproteobacteria (see Additional file 4).

## Discussion

### HGT in the plastid ancestor

We propose that *menF*, *menD*, *menH*, *menC*, *menB*, and *menE* that comprise an operon or a gene cluster, and the *menA* gene was introduced via two independent HGT events into the genome of the cyanobacterial ancestor of

primary plastids. The remnants of the *men* gene cluster and *menA* are still present in the plastid genome of the Cyanidiales in a prokaryotic-like gene arrangement. In contrast, the components of the horizontally transferred *men* cluster were moved to the nucleus in the majority of algal lineages. An alternative interpretation of our data is that the *men* genes were introduced into the plastid genome of the Plantae ancestor by a HGT event that post-dates primary endosymbiosis but precedes the split of at least the red and green algae (i.e., glaucophyte genes have not yet been described). HGT into plastids is the most likely explanation for the origin of the *rbcl* gene in red and chromalveolate algae, the *rpl36* gene shared by cryptophytes and haptophytes, and the unique occurrence of the *dnaX* gene in *R. salina* (see above) [9,16,17,25]. We suggest that several lines of (albeit) circumstantial evidence argue against this scenario for origin of the *men* gene cluster. The first is that recent comprehensive phylogenomic analyses of plastid genomes uncovered only the three cases of HGT described above [9,17]. This result, when magnified over the >1 billion years of plastid evolution leads to the conclusion that plastid genomes are generally resistant to HGT [9]. Second, the three known plastid HGT events involve a perfect gene replacement (*rpl36*) [9], a novel single gene insertion (*dnaX*) [17], and the replacement of an intact operon (*rbclS*) [16]. From our perspective, these mechanisms are unlikely to be the explanation for the introduction of the horizontally transferred *men* gene cluster in the plastid lineage for two reasons. First, the plastid genome is essentially free of large new DNA integrants. And second, a sparse, gene-poor *men* cluster of cyanobacterial origin that putatively existed in the endosymbiont would provide a dispersed and poorly conserved substrate for a single replacement involving homologous recombination with a highly dissimilar DNA segment containing six genes of Chlorobi/Gammaproteobacteria provenance. We propose that it is more likely that the introduction of the original *men* gene cluster occurred in the more dynamic genomes of freeliving cyanobacteria, where foreign genes are commonly co-integrated with cyanophages in chromosomal genomic islands in events not involving homologous recombination and that often introduce redundant alleles that duplicate native cyanobacterial genes [26,27]. Following the original integration we suggest that the native cyanobacterial *men* pathway in the plastid forerunner was lost in favour of the foreign gene cluster. The gain of an additional *men* cluster directly into the newly established organelle, thereby creating allelic redundancy is unlikely giving the known tendency of endosymbionts (e.g., plastids, *Carsonella*) to rapidly streamline their genomes by gene loss [28].

Comparative genomic analyses demonstrate that the acquisition of foreign genes by HGT is common among cyanobacteria [29]. In line with this idea, we provide clear

evidence for the ancient acquisition by cyanobacteria of a foreign *menH* from a Chlorobi source (Figure 2). Similarly, our phylogeny of *MenC* (see Additional file 1) shows that this gene in *Crocospaera watsoni* and *Cyanotheca* sp. was introduced by HGT from a Firmicute source. Together, these examples underline the case we make for *men* gene cluster transfer into the once free-living plastid ancestor. In a parallel example, the thecate amoeba *Paulinella chromatophore* has a plastid-like photosynthetic organelle, the cyanelle, which was also derived from a cyanobacterium. The genomes of the cyanelle and the cyanobacteria *Prochlorococcus* and *Synechococcus* that are the closest relatives of the cyanelle share common components of a carboxysomal operon (including the *rbcl* gene) that is phylogenetically most closely related to Gammaproteobacteria. This implies that the carboxysomal operon was introduced by HGT into the common ancestor of *Prochlorococcus* and *Synechococcus* species and the *Paulinella* cyanelle [30]. This constitutes a recent example of the type of HGT event we postulate occurred for the *men* gene cluster ca. 1.5 billion years ago.

#### **Relics of endosymbiont genome chimerism may now reside in the nucleus**

With regard to interphylum HGT, genome analysis of cyanobacteria suggests that 23% of the tested genes provide conflicting phyletic patterns that include cases of HGTs from donors outside the cyanobacterial clade [29]. In contrast, the majority of phylogenetic studies have found a consistent signal of cyanobacterial ancestry in the plastid genome [e.g., [11,12]]. If HGT was similarly prevalent at the time of endosymbiosis (and there is no obvious reason to think otherwise) then one would expect to find many more examples of plastid genes that are not phylogenetically affiliated with cyanobacteria. The current data suggest however that the contribution of interphylum HGT to the cyanobacterial ancestor of plastids was rare and the *men* gene cluster represents an unlikely survivor.

On the other hand, the apparent absence of additional HGT events in the plastid could reflect a lack of thorough analysis of all genes (not just the phylogenetically conserved set) in all plastid genomes to assign their origins. To assess this possibility, we did an additional phylogenomic analysis in which all of the plastid protein-encoding genes in *Cyanidioschyzon*, the glaucophyte *Cyanophora paradoxa*, and the chlorophyte *Nephroselmis olivacea* were used to query our large bacterial plus eukaryotic protein database. The trees were sorted to identify any clear examples of non-cyanobacterial genes in these plastids [15]. No other cases of interphylum HGT were found (consistent with Rice and Palmer [9] and Khan *et al.*, [17]) which leads us to postulate that within the limits of these bioin-

formatic approaches, the remaining set of plastid genes is of cyanobacterial ancestry.

The absence of endosymbiont chimericism that is found in modern-day plastids may reflect the fact that this organelle encodes only 1/10 (ca. 200 out of 2,500 ancestral genes) of the original cyanobacterial genome. Therefore, the evidence for ancient HGT may have been lost over evolutionary time. Genes preserved in plastids generally have conserved roles in transcription and translation (i.e., informational) or encode subunits of multiprotein complexes involved in photosynthesis [31]. These sequences may be more resistant to HGT than the set of metabolic and house-keeping genes (i.e., operational) that were lost. Operational genes have been demonstrated in cyanobacteria to be relatively more frequently transferred than informational genes, although genes in all functional categories are subject to HGT [29]. Many operational genes once encoded in the cyanobacterial ancestor of plastids have been relocated to the nucleus since primary endosymbiosis [31]. The *menF*, *menD*, *menC*, *menH*, and *menB* genes that are derived from original plastid gene cluster, and possibly *menA* and *menG* (of cyanobacterial origin), followed this path and are now nuclear-encoded in the vast majority of algae and plants. From our perspective, a remarkable exception to this rule is the ancestral *men* gene cluster that was, for unknown reasons, 'frozen' in the plastid genome of a single algal group, the Cyanidiales. Finally, phylogenetic studies of nuclear encoded plastid targeted proteins often show a phylogenetic history that is inconsistent with a cyanobacterial ancestry [32-35]. Although these cases can be interpreted as examples of HGT into the nuclear genome of photosynthetic eukaryotes, which is prevalent in some algae (e.g., chromalveolates) [e.g., [36-38]], our study at least raises the possibility that some may have arisen from a putative chimeric cyanobacterial ancestor and trace their origin to EGT [4,5].

In our phylogenomic analysis we failed to detect significant departure from the cyanobacterial origin of plastid-encoded genes apart from the known cases of *rbcL* [16] and MQ/PhQ. This however should not be interpreted as an absence of HGT in the plastid forerunner. In fact, comparative genomic analyses demonstrate that HGT is common in cyanobacteria, affecting ca. 61% of the genes [29]. These transfers are however largely intraphylum and beyond the limits of detection of our study. Therefore it is entirely possible (and probable) that plastid genomes contain a highly reticulated pattern of gene ancestry that predates primary endosymbiosis. In line with this idea, results of a recent analysis of cyanobacteria-derived genes in the nuclear genome of *Plantae* [39] showed that in terms of presence/absence and sequence similarity, the highest proportion of these proteins could be traced back

to *Nostoc* sp. PCC7120 and *Anabaena variabilis* ATCC29143, identifying them as potential donor lineages. Nonetheless a substantial fraction was found to be more similar to homologs in seven other studied cyanobacteria. These data suggest that even if the plastid ancestor was related to heterocyst-forming *Anabaena/Nostoc*, the captured cell also contained genes from a phylogenetically broad spectrum of cyanobacterial lineages presumably as a consequence of ancient intraphylum HGT.

## Conclusion

We asked the question, do plastid genomes still preserve vestiges of HGT that occurred during the free-living phase of the cyanobacterial endosymbiont? For this purpose our phylogenetic study targeted the set of *Men* proteins responsible for the biosynthesis of MQ in eubacteria and PhQ in cyanobacteria and in photosynthetic eukaryotes. These proteins are encoded in the plastid genome of the red algae Cyanidiales and in the nuclear genome of algae and plants, indicating that they were derived from the cyanobacterial endosymbiont that gave rise to the plastids and later on transferred to the nucleus by EGT. This result clearly contradicts the expectation of a cyanobacterial provenance for the MQ/PhQ pathway in *Plantae*, showing an affiliation of these sequences to Chlorobi/Gammaproteobacteria (Figures 1b, 2; see Additional files 1, 2, 3). This suggests that the cyanobacterial ancestor of the primary plastid contained a complete *men* gene cluster (*menF*, *menD*, *menH*, *menB*, *menC*, and *menE*) that originated via HGT. In addition, the *menA* gene was also independently derived by HGT in the plastid ancestor from an unknown donor. The *men* gene cluster in Cyanidiales retains this ancestral prokaryotic feature and thus represents a 'living fossil' providing a view into ancient genome chimericism. In contrast, *Men* homologs are nuclear-encoded in the genomes of green algae, plants, and diatoms and are monophyletic with the plastid-encoded homologs in Cyanidiales. This demonstrates that a chimeric component of the endosymbiont genome was relocated in some lineages from the plastid to the nucleus. The ancestral cyanobacterial lineage that donated the plastid either went extinct or is not yet represented in GenBank because the cyanobacteria in all of our trees analysis are clearly monophyletic and display a sporadic pattern of *men* gene retention typical of Group 2 (Figure 1b). We speculate that the discovery of extant cyanobacteria that contain a *men* operon derived from Chlorobi/Gammaproteobacteria would potentially provide an useful piece of evidence to identify the lineage most closely related to the plastid ancestor.

## Materials and methods

### Sequence data mining

The genes for MQ and PhQ biosynthesis have previously been characterized in *E. coli* and in *Synechocystis* sp. PCC 6803 [19,20], as has *PHYLL*O (with the functional



*menDCH* modules), the *menF* genes (*ICS*), *menG*, *menA*, and *menE* in *Arabidopsis* [18,21-23]. The product of all these genes served as queries to find by BLASTP [40] searches the corresponding homologs in other prokaryotes and eukaryotes. For the collection of proteins and initial analysis of the structure of *men* gene clusters, about 120 taxa of menaquinone-using bacteria were selected among the major prokaryotic phyla. Only a fraction of taxa from this original data set was used for phylogenetic analysis. All retrieved Men proteins were authenticated by the presence of the corresponding coding genes within clusters of *men* genes, as well by comparisons to aligned modules of Men proteins in the Conserved Domain Database [41]. This approach allowed us to identify the majority of Men homologs. The use of positive cases also made it possible to deduce cut-offs of the e-value for each family of proteins to identify homologs by BLASTP in some taxa where Men proteins were not encoded within *men* gene clusters.

The eukaryotic proteins from *Arabidopsis* and rice and the Cyanidiales *Cyanidium caldarium* and *Cyanidioschyzon merolae* were obtained from GenBank [42]. The nucleotide sequence of part of the *men* gene cluster of *Galdieria maxima*, containing *menB*, *menF*, and *menD*, was determined in our laboratory (GenBank accession number [EU281345](#)). The remaining plant sequences of *Physcomitrella patens* ssp. *patens*, *Populus trichocarpa*, the green algae *Chlamydomonas reinhardtii*, *Ostreococcus lucimarinus*, and *Ostreococcus tauri*, and the diatom algae *Phaeodactylum tricornutum*, and *Thalassiosira pseudonana* were retrieved from the DOE joint genome institute database [43]. *PHYLLO* was identified in all these organisms with variations in its modular structure. In green algae and the moss *Physcomitrella patens* *PHYLLO* displays the architecture *menFDCH*. In higher plants the *menF* enzymatic function is provided by an individual gene (*isochorismate synthase*, *ICS*) that originated via duplication of the *menF* module of *PHYLLO* early in vascular plant evolution [18]. This event was followed by fission of the 3' region of the *PHYLLO* *menF* module. As a consequence, the architecture of *PHYLLO* (*menF5'DCH*) in higher plants presents a short inactive *menF* module that is homologous to the 5' region of *menF* genes [18]. In diatoms the *menC* module of the *PHYLLO* follows *menH*. In addition, the 3' module encoding a domain of unknown function (*duf*, COG5637) is also combined to the preceding *men* components, resulting in the pentamodular fusion *men FDHC duf*.

All of the retrieved sequences were manually annotated with help of the comparative prediction algorithms FGENESH+ [44] and Wise2 [45] combined with multiple alignments with known genes of plant and algae. This approach permitted us to deduce exon-intron boundaries

with maximum confidence overcoming common errors resulting from automated annotation pipelines. Sequences were aligned using ClustalW [46] and manually edited. Only conserved regions of proteins that could be unambiguously aligned were used to construct phylogenetic trees. The protein alignments with accession numbers are available from the Debashish Bhattacharya lab web page [47].

### Phylogenetic analysis

The phylogeny of the large alignment of MenFD was inferred using MrBayes [48]. Metropolis-coupled Markov chain Monte Carlo from a random starting tree was used in this analysis with two independent runs (i.e., nrun = 2 command) and 1 cold and 3 heated chains. The Bayesian analysis (aamodel = mixed; rates = invgamma) was run for 2 million generations with trees sampled every 100th generation. To increase the probability of chain convergence, we sampled trees after the standard deviation values of the two runs were < 0.05 to calculate the posterior probabilities. The remaining phylogenies were discarded as burn-in. Bootstrap support for nodes in the majority rule consensus Bayesian tree was determined using maximum likelihood (i.e., PHYML V2.4.3 [49]) and neighbor joining (NJ; MEGA [50]). For the PHYML analysis the model parameters were: model of amino acid substitution = WAG (identified using ProtTest V1.3 [51] with "Fast" optimization and a BIONJ tree); starting tree = NJ; gamma = 1.361; p-inv. = 0.073. For the MEGA analysis: model of amino acid substitution = JTT; gamma = 0.84 (obtained using ProtTest).

The MenH and MenA trees were also inferred using MrBayes as described above. Nodal support was assessed using PHYML bootstrap. The model parameters for these analyses were as follows. MenH PHYML: model of amino acid substitution = WAG (identified using ProtTest); starting tree = NJ; gamma = 2.014; p-inv. = 0.067. MenA PHYML: model of amino acid substitution = Blosum62 (identified using ProtTest); starting tree = NJ; Gamma = 1.591; p-inv. = 0.025. The MenC, MenB, MenE and MenG trees were constructed using MrBayes with specific parameters for each tree as indicated (see Additional files 1, 2, 3, 4).

To address alternative hypotheses about the tree topologies, we generated a backbone phylogeny in each case that was identical to the 'best' PHYML tree but excluded the group of interest. Using this backbone tree, the group of interest was then added individually using MacClade V4.05 [52] to 20 alternative branches that spanned the most likely placements of the clade. The site-by-site likelihoods for the trees in this analysis were calculated using the appropriate data set and model with TREEPUZZLE (V5.2, 31) [53] and the default settings. The approxi-

mately unbiased (AU-) test was then implemented using CONSEL V0.1i [54] to assign the tree probabilities.

### List of Abbreviations

AU-test: approximately unbiased test; BPP: Bayesian posterior probability; EGT: endosymbiotic gene transfer; HGT: horizontal gene transfer; ICS: isochorismate synthase; MQ: menaquinone; NJ: neighbor joining; NJB: neighbor joining bootstrap; PB: PHYML bootstrap; PhQ: phyloquinone.

### Authors' contributions

JG conceived and designed the project, collected and analysed the data and performed the phylogenetic analysis. JM helped to supervise the project. DB helped design the phylogenetic approach, analysed the data, and supervised the project. JG wrote the first draft of the manuscript and DB is responsible for the final version. All authors read and approved the final manuscript.

### Additional material

#### Additional file 1

*Phylogeny of MenC.* This figure indicates that the MenC module of PHYLO and the individual plastid-encoded MenC proteins of Cyanidiales are related to homologs of Chlorobi and Gammaproteobacteria. This is a Bayesian majority rule consensus tree using 58 taxa. Posterior probability support values are only indicated (as percentages) for external nodes of the major clades. Analysis parameters: mcmc ngen = 500,000; startingtree = PHYML; samplefreq = 100; aamodel = mixed; rates = invgamma; burnin = 1,250.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-117-S1.pdf>]

#### Additional file 2

*Phylogeny of MenB.* The tree is juxtaposed to a partial alignment of the corresponding amino acid positions 69 – 116 in the E. coli MenB homolog (GenBank accession number, [NP\\_416765.1](#)). Two deletions are correlated with the tree topology. Note that deletion 2 is shared by all Group 1 taxa in Fig. 1b, including photosynthetic eukaryotes, Chlorobi, and Gammaproteobacteria. This is a Bayesian majority rule consensus tree using MrBayes and the following parameters: mcmc ngen = 500,000; startingtree = PHYML; samplefreq = 100; aamodel = mixed; rates = invgamma; burnin = 1,250. Posterior probability support values are only indicated (as percentages) for external nodes of the major clades.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-117-S2.pdf>]

#### Additional file 3

*Phylogeny of MenE.* The plastid-encoded menE of Cyanidiales has its origin in the same gene cluster that was transferred to the plastid ancestor from a Chlorobi/Gammaproteobacteria source. The MenE proteins in the nuclear genome of the remaining algae and plants have a different phylogenetic provenance and are most likely derived by HGT, presumably directly into the host nucleus. This is a Bayesian majority rule consensus tree using 66 taxa. Posterior probability support values are only indicated (as percentages) for external nodes of the major clades. Analysis parameters are: mcmc ngen = 500,000; startingtree = PhyML generated tree; samplefreq = 100; aamodel = mixed; rates = invgamma; burnin = 1,250.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-117-S3.pdf>]

#### Additional file 4

*Phylogeny of MenG/UbiE.* MenG proteins are responsible for the methylation step of the MQ/PhQ pathway (Lohmann et al. 2006). In eubacteria the same enzymes also catalyse the methylation required for biosynthesis of ubiquinone (Meganathan 2001). In this case MenG homologs are referred to as UbiE. In photosynthetic eukaryotes UbiE functions in mitochondria and MenG is plastid targeted for the synthesis of PhQ. In accordance with its organellar distribution, UbiEs are related to homologs in Alphaproteobacteria, whereas MenG in the nuclear genomes of plants and green algae is derived from the cyanobacterial endosymbiont. MenG encoded in the nuclear genome of Cyanidiales and diatoms are related to Deltaproteobacteria and presumably originated via HGT. This is a Bayesian majority rule consensus tree using 81 taxa. Posterior probability support values are only indicated (as percentages) for external nodes of the major clades. Analysis parameters: ngen = 1,100,000; startingtree = PHYML generated tree; samplefreq = 100; aamodel = mixed; rates = gamma; burnin = 2,750.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-117-S4.pdf>]

### Acknowledgements

This work was partially supported by grants from the Deutsche Forschungsgemeinschaft (SFB TR1 and Me 1794/4) to JM and the National Science Foundation and the National Aeronautics and Space Administration awarded to DB (EF 04-31117, NNG04GM17G). JG was supported by a Ph.D. scholarship from the CAPES Foundation from the Brazilian Ministry of Education. We thank Adrian Reyes-Prieto and Ahmed Moustafa (University of Iowa) for technical assistance and helpful discussions of this work.

### References

1. Koonin EV, Makarova KS, Aravind L: **Horizontal gene transfer in prokaryotes: quantification and classification.** *Annu Rev Microbiol* 2001, **55**:709-742.
2. Ochman H, Lawrence JG, Groisman EA: **Lateral gene transfer and the nature of bacterial innovation.** *Nature* 2000, **405**:299-304.
3. Nakamura Y, Itoh T, Matsuda H, Gojobori T: **Biased biological functions of horizontally transferred genes in prokaryotic genomes.** *Nat Genet* 2004, **36**:760-766.
4. Esser C, Martin W, Dagan T: **The origin of mitochondria in light of a fluid prokaryotic chromosome model.** *Biol Lett* 2007, **3**:180-184.
5. Martin W: **Mosaic bacterial chromosomes: a challenge en route to a tree of genomes.** *Bioessays* 1999, **21**:99-104.
6. Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D: **A molecular timeline for the origin of photosynthetic eukaryotes.** *Mol Biol Evol* 2004, **21**:809-818.

7. Hedges SB, Chen H, Kumar S, Wang DY, Thompson AS, Watanabe H: **A genomic timescale for the origin of eukaryotes.** *BMC Evol Biol* 2001, **1**:4.
8. Reyes-Prieto A, Weber AP, Bhattacharya D: **The origin and establishment of the plastid in algae and plants.** *Annu Rev Genet* 2007, **41**:147-168.
9. Rice DW, Palmer JD: **An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters.** *BMC Biol* 2006, **4**:31.
10. NCBI Organelle Genome Resources [[http://www.ncbi.nlm.nih.gov/genomes/ORGANELLES/plastids\\_tax.html](http://www.ncbi.nlm.nih.gov/genomes/ORGANELLES/plastids_tax.html)]
11. Morden CW, Delwiche CF, Kuhse M, Palmer JD: **Gene phylogenies and the endosymbiotic origin of plastids.** *Biosystems* 1992, **28**:75-90.
12. Chu KH, Qi J, Yu ZG, Anh V: **Origin and phylogeny of chloroplasts revealed by a simple correlation analysis of complete genomes.** *Mol Biol Evol* 2004, **21**:200-206.
13. Sato N, Ishikawa M, Fujiwara M, Sonoike K: **Mass identification of chloroplast proteins of endosymbiont origin by phylogenetic profiling based on organism-optimized homologous protein groups.** *Genome Inform* 2005, **16**:56-68.
14. Reyes-Prieto A, Hackett JD, Soares MB, Bonaldo MF, Bhattacharya D: **Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions.** *Curr Biol* 2006, **16**:2320-2325.
15. Moustafa A, Bhattacharya D: **PhyloSort: a user-friendly phylogenetic sorting tool and its application to estimating the cyanobacterial contribution to the nuclear genome of *Chlamydomonas*.** *BMC Evol Biol* 2008, **8**:6.
16. Delwiche CF, Palmer JD: **Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids.** *Mol Biol Evol* 1996, **13**:873-882.
17. Khan H, Parks N, Kozera C, Curtis BA, Parsons BJ, Bowman S, Archibald JM: **Plastid genome sequence of the cryptophyte alga *Rhodomonas salina* CCMP1319: lateral transfer of putative DNA replication machinery and a test of chromist plastid phylogeny.** *Mol Biol Evol* 2007, **24**:1832-1842.
18. Gross J, Cho WK, Lezhneva L, Falk J, Krupinska K, Shinozaki K, Seki M, Herrmann RG, Meurer J: **A plant locus essential for phyloquinone (vitamin K1) biosynthesis originated from a fusion of four eubacterial genes.** *J Biol Chem* 2006, **281**:17189-17196.
19. Meganathan R: **Biosynthesis of menaquinone (vitamin K2) and ubiquinone (coenzyme Q): a perspective on enzymatic mechanisms.** *Vitam Horm* 2001, **61**:173-218.
20. Wade Johnson T, Naithani S, Stewart C Jr, Zybailev B, Daniel Jones A, Golbeck JH, Chitnis PR: **The men D and menE homologs code for 2-succinyl-6-hydroxyl-2,4-cyclohexadiene-1-carboxylate synthase and O-succinylbenzoic acid-CoA synthase in the phyloquinone biosynthetic pathway of *Synechocystis* sp. PCC 6803.** *Biochim Biophys Acta* 2003, **1557**:67-76.
21. Shimada H, Ohno R, Shibata M, Ikegami I, Onai K, Ohto MA, Takamiya K: **Inactivation and deficiency of core proteins of photosystems I and II caused by genetical phyloquinone and plastoquinone deficiency but retained lamellar structure in a T-DNA mutant of *Arabidopsis*.** *Plant J* 2005, **41**:627-637.
22. Lohmann A, Schöttler MA, Bréhélin C, Kessler F, Bock R, Cahoon EB, Dörmann P: **Deficiency in phyloquinone (vitamin K<sub>1</sub>) methylation affects prenyl quinone distribution, Photosystem I abundance, and anthocyanin accumulation in the *Arabidopsis AtmenG* mutant.** *J Biol Chem* 2006, **281**:40461-40472.
23. Kim HU, Oostende CV, Basset GJ, Browse J: **The AAE14 gene encodes the *Arabidopsis* osuccinylbenzoyl-CoA ligase that is essential for phyloquinone synthesis and photosystem I function.** *Plant J* 2008. Epub ahead of print.
24. Hackett JD, Yoon HS, Li S, Reyes-Prieto A, Rümmele SE, Bhattacharya D: **Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates.** *Mol Biol Evol* 2007, **24**:1702-1713.
25. Reyes-Prieto A, Bhattacharya D: **Phylogeny of Calvin cycle enzymes supports Plantae monophyly.** *Mol Phylogenet Evol* 2007, **45**:384-391.
26. Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, DeLong EF, Chisholm SW: **Genomic islands and the ecology and evolution of *Prochlorococcus*.** *Science* 2006, **311**:1768-1770.
27. Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chisholm SW: **Transfer of photosynthesis genes to and from *Prochlorococcus* viruses.** *Proc Natl Acad Sci USA* 2004, **101**:11013-11018.
28. Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, Hattori M: **The 160-kilobase genome of the bacterial endosymbiont *Carsonella*.** *Science* 2006, **314**:259-260.
29. Zhaxybayeva O, Gogarten JP, Charlebois RL, Doolittle WF, Papke RT: **Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events.** *Genome Res* 2006, **16**:1099-1108.
30. Marin B, Nowack EC, Glockner G, Melkonian M: **The ancestor of the *Paulinella chromatophora* obtained a carboxysomal operon by horizontal gene transfer from a *Nitrococcus*-like gamma-proteobacterium.** *BMC Evol Biol* 2007, **7**:85.
31. Race HL, Herrmann RG, Martin W: **Why have organelles retained genomes?** *Trends Genet* 1999, **15**:364-370.
32. Lange BM, Rujan T, Martin W, Croteau R: **Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes.** *Proc Natl Acad Sci USA* 2000, **97**:13172-13177.
33. Nara T, Hshimoto T, Aoki T: **Evolutionary implications of the mosaic pyrimidine-biosynthetic pathway in eukaryotes.** *Gene* 2000, **257**:209-222.
34. Illingworth C, Mayer MJ, Elliott K, Hanfrey C, Walton NJ, Michael AJ: **The diverse bacterial origins of the *Arabidopsis* polyamine biosynthetic pathway.** *FEBS Lett* 2003, **549**:26-30.
35. Richards TA, Dacks JB, Campbell SA, Blanchard JL, Foster PG, McLeod R, Roberts CW: **Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements.** *Eukaryot Cell* 2006, **5**:1517-1531.
36. Fast NM, Kissinger JC, Roos DS, Keeling PJ: **Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids.** *Mol Biol Evol* 2001, **18**:418-26.
37. Nosenko T, Lidie KL, Van Dolah FM, Lindquist E, Cheng JF, Bhattacharya D: **Chimeric plastid proteome in the Florida "red tide" dinoflagellate *Karenia brevis*.** *Mol Biol Evol* 2006, **23**:2026-2038.
38. Tyra HM, Linka M, Weber AP, Bhattacharya D: **Host origin of plastid solute transporters in the first photosynthetic eukaryotes.** *Genome Biol* 2007, **8**:R212.
39. Deusch O, Landan G, Roettger M, Gruenheit N, Kowalik KV, Allen JF, Martin W, Dagan T: **Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor.** *Mol Biol Evol* 2008, **25**:748-761.
40. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
41. Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Marchler GH, Mullokandov M, Shoemaker BA, Simonyan V, Song JS, Thiesens PA, Yamashita RA, Yin JJ, Zhang D, Bryant SH: **CDD: a Conserved Domain Database for protein classification.** *Nucleic Acids Res* 2005, **33**:D192-196.
42. NCBI [<http://www.ncbi.nlm.nih.gov/>]
43. DOE Joint Genome Institute [<http://www.jgi.doe.gov/>]
44. Softberry [<http://linux1.softberry.com/berry.phtml>]
45. EBI [<http://www.ebi.ac.uk/WISE2/>]
46. Thompson JD, Higgins DG, Gibson TJ: **Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**:4673-4680.
47. Debashish Bhattacharya lab [<http://www.biology.uiowa.edu/debweb/>]
48. Ronquist F, Huelsenbeck JP: **MRBAYES 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**:1572-1574.
49. Guindon S, Gascuel O: **PhyML – A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003, **52**:696-704.
50. Kumar S, Tamura K, Nei M: **MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment.** *Brief Bioinform* 2004, **5**:150-163.
51. Abascal F, Zardoya R, Posada D: **ProtTest: Selection of bestfit models of protein evolution.** *Bioinformatics* 2005, **21**:2104-2105.

52. Maddison WP, Maddison DR: **Interactive analysis of phylogeny and character evolution using the computer program MacClade.** *Folia Primatol* 1989, **53**:190-202.
53. Schmidt HA, Strimmer K, Vingron M, Haeseler A: **TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing.** *Bioinformatics* 2002, **18**:502-504.
54. Shimodaira H, Hasegawa M: **CONSEL: For assessing the confidence of phylogenetic tree selection.** *Bioinformatics* 2001, **17**:1246-1247.

Publish with **BioMed Central** and every scientist can read your work free of charge

*"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."*

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

