Phylogeny of Galactolipid Synthase Homologs Together with their Enzymatic Analyses Revealed a Possible Origin and Divergence Time for Photosynthetic Membrane Biogenesis

Yuichi Yuzawa[1], Hidenori Nishihara[1], Tsuyoshi Haraguchi[1], Shinji Masuda[2], Mie Shimojima[2], Atsushi Shimoyama[1], Hideya Yuasa[1], Norihiro Okada[1], and Hiroyuki Ohta[2,*]

Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 B-65 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan[1] and Center for Biological Resources and Informatics, Tokyo Institute of Technology, 4259 B-65 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan[2]

*To whom correspondence should be addressed. Tel. +81 45-924-5736. Fax. +81 45-924-5823.
Email: ohta.h.ab@m.titech.ac.jp

Abstract

The photosynthetic membranes of cyanobacteria and chloroplasts of higher plants have remarkably similar lipid compositions. In particular, thylakoid membranes of both cyanobacteria and chloroplasts are composed of galactolipids, of which monogalactosyldiacylglycerol (MGDG) is the most abundant, although MGDG biosynthetic pathways are different in these organisms. Comprehensive phylogenetic analysis revealed that MGDG synthase (MGD) homologs of filamentous anoxygenic phototrophs Chloroflexi have a close relationship with MGDs of Viridiplantae (green algae and land plants). Furthermore, analyses for the sugar specificity and anomeric configuration of the sugar head groups revealed that one of the MGD homologs exhibited a true MGDG synthetic activity. We therefore presumed that higher plant MGDs are derived from this ancestral type of MGD genes, and genes involved in membrane biogenesis and photosystems have been already functionally associated at least at the time of Chloroflexi divergence. As MGD gene duplication is an important event during plastid evolution, we also estimated the divergence time of type A and B MGDs. Our analysis indicated that these genes diverged ~323 million years ago, when Spermatophyta (seed plants) were appearing. Galactolipid synthesis is required to produce photosynthetic membranes; based on MGD gene sequences and activities, we have proposed a novel evolutionary model that has increased our understanding of photosynthesis evolution.

Key words: monogalactosyldiacylglycerol; MGDG synthase; galactolipid; *Roseiflexus castenholzii*; plastid evolution

1. Introduction

It is generally accepted that chloroplasts evolved from a photosynthetic prokaryote that entered into an endosymbiotic relationship with a non-photosynthetic host cell.[1] It is also accepted that this photosynthetic prokaryote was an ancestor of the phylum Cyanobacteria.[2] It is plausible, therefore, that most of the genes encoding chloroplast-localized proteins, including components of the photosynthetic machinery, are derived from this ancestral cyanobacterium.[3] Cyanobacteria and chloroplasts of higher plants share common features in many biological contexts, including photosynthesis and cell division.[4,5] In addition, the lipid compositions of biological membranes (photosynthetic membranes in particular) show high degrees of similarity between cyanobacteria and chloroplasts.[6] Thylakoid membranes within a

chloroplast are ~80% galactolipids, whereas the biological membranes of prokaryotic and eukaryotic cells are generally composed of phospholipids. Monogalactosyldiacylglycerol (MGDG) is the most abundant lipid in thylakoid membranes of both cyanobacteria and higher plants. Thus, it is likely that genes encoding enzymes of the galactolipid synthetic pathways were acquired in higher plants through the endosymbiotic event.

MGDG is essential for oxygenic phototrophs. Crystal structure analyses of cyanobacteria have revealed that MGDG binds to photosystem complexes, suggesting that MGDG is an important component of the photosynthetic machinery.[7,8] Furthermore, in MGDG-deficient *Arabidopsis thaliana*, photosynthetic activity is completely absent, and development is arrested at the early seedling stage.[9] Thus, MGDG is involved in both photosynthesis and chloroplast development.

The pathways for MGDG synthesis in cyanobacteria and higher plants, however, are quite different.[10] Higher plants utilize uridine diphosphate galactose (UDP-Gal) and diacylglycerol (DAG) to synthesize MGDG. In contrast, cyanobacteria must first synthesize monoglucosyl-DAG (MGlcDG) using UDP glucose (UDP-Glc) and DAG, which is subsequently epimerized to produce MGDG. In addition, sequence comparison indicates that MGDG synthase (MGD) genes belong to the glycosyltransferase 28 (GT28) family, whereas MGlcDG synthase (MGlcD) genes belong to the family GT2 [according to the Carbohydrate-Active Enzymes database (www.cazy.org)].[11−13] In Viridiplantae (green algae and land plants), therefore, the precise origin of the pathway for MGDG synthesis is not clear.

As MGDG is essential for photosynthesis, understanding the evolution of this pathway could potentially reveal important details concerning phototroph evolution in general. MGDG is found in some anoxygenic phototrophs as well as oxygenic phototrophs.[14] Chloroflexi is an anoxygenic phototrophic group of bacteria whose divergence preceded that of Cyanobacteria.[15−17] The MGD homolog of *Chloroflexus aurantiacus*, a member of the phylum Chloroflexi,[18] belongs to the GT28 family and one of the encoded proteins has an MGDG producing activity.[19] However, the phylogenetic relationship between these prokaryotic MGD homologs and higher plant MGDs has not been well established. Here, we performed comprehensive phylogenetic analyses on MGD homologs from both prokaryotic and eukaryotic organisms and revealed that eukaryotic MGDs including Viridiplantae, Rhodophyta, and Heterokontophyta are monophyly and probably have a single ancestor. Moreover, we found that some of Chloroflexi MGD homologs are highly associated with the eukaryotic MGDs. Analyses of the

sugar specificity and anomeric configuration of the sugar head groups for three MGD homologs in Chloroflexi, *Roseiflexus castenholzii*,[20] revealed that only one of the MGD homologs in *R. castenholzii* exhibits a genuine MGD activity, suggesting a possible origin of higher plant MGDs.

To understand plastid evolution, it is also important to characterize MGD gene duplication. In *A. thaliana*, there are two types of MGD enzymes, type A (MGD1) and type B (MGD2 and MGD3).[21] Type A MGDs reside in membranes of the inner envelope and, under normal circumstances, generate most MGDGs found in the chloroplast. In contrast, type B MGDs are found in membranes of the outer envelope and are involved in membrane remodeling under phosphate-starved conditions.[22] MGDG synthesized in this context is converted into digalactosyl-DAG, which is used to generate cell membranes when phospholipids are not available. As type B MGDs are only found in angiosperms, this type of membrane lipid remodeling is an important characteristic of angiosperms.[23−26] Under normal growth conditions, interestingly, the expression of type B MGD are low in leaves but high in floral organs.[27] Thus, it is possible that the emergence of type B MGDs was an important step toward Angiospermae (angiosperm) or Spermatophyta (seed plant) evolution. Two types of MGDs are essential for plants to survive under low-phosphate conditions, and yet a detailed analysis of MGD gene divergence has not been reported. In this study, we present detailed data concerning MGD gene divergence. Our data provide a new insight into the plastid evolution based on the biogenesis of photosynthetic membranes.

2. Materials and methods

2.1. Phylogenetic analyses

To obtain MGD gene sequences, blastp and tblastn searches were performed against the Protein (nr), GenBank, and EST databases at the National Center for Biotechnology Information website (www.ncbi.nlm.nih.gov/), using *Arabidopsis* MGD amino acid sequences as queries. In addition, a blastp search was conducted against genomic data at the Department of Energy Joint Genome Institute (DOE-JGI) website (www.jgi.doe.gov/). The resulting data are listed in Supplementary Table S1.

MGD amino acid sequences were aligned using MACSIMS[28] and MUSCLE,[29] followed by a partial manual correction. To refine these data for phylogenetic analysis, the 5′- and 3′-terminal regions of MGD gene and regions where >15% of the species had a sequence gap were eliminated from the alignment. The refined data were then subjected to phylogenetic

analyses with Bayesian inference and maximum likelihood (ML). MrBayes ver 3.1.2[30] was used for Bayesian inference under the WAG + Γ_4 model. Two runs with four chains of Markov chain Monte Carlo (MCMC) iterations were performed for 4 000 000 generations. Trees were sampled every 100 generations, and the first 10 000 trees (1 000 000 generations) were discarded as burn-in. Treefinder ver. October 2008[31] was used for ML analyses under the WAG-F + Γ_8 model. Approximate bootstrap supports (LR-ELW; expected likelihood weights by the local rearrangement) were calculated with 10 000 replicates. RAxML ver. 7.0.4[32] was also used for ML analyses under the WAG-F + Γ_4 model with 1000 bootstrap replications.

To calculate the divergence time of type A and type B MGDs, the nucleotide sequences of land plant MGDs (the first and second codon positions) were subjected to Bayesian and ML phylogenetic analyses. Sequence data are listed in Supplementary Table S2. MrBayes ver 3.1.2 was used for Bayesian-tree inference with 1 000 000 generations under the GTR + Γ_8 model. Trees were sampled every 100 generations, and the first 2500 trees (250 000 generations) were discarded as burn-in. The ML tree was estimated using RAxML ver. 7.0.4 under the GTR + Γ_4 model with 1000 bootstrap replications. In both phylogenetic analyses, different model parameters were estimated between the first and the second codon position data. Divergence times were estimated based on the first and second codon sequences under a relaxed clock model using the MCMCTREE program in the PAML4 package.[33] The model parameters were separately estimated between the first and the second codon sequences under the HKY + Γ_5 model. The number of MCMC samples and burn-in were set to 10 000 and 2500, respectively. The following constraints were used for time calibrations (similar to Yoon *et al.*[34]): (i) divergence of vascular plants from other land plants <432 million years ago (MYA) (Kenrick and Crane[35] provided the origin of land plants), (ii) divergence of seed plants from other vascular plants >355 MYA (Gillespie *et al.*[36] provided the oldest known seeds), (iii) 290–320 MYA for the Angiospermae–Gymnospermae split, and (iv) 90–130 MYA for the Monocotyledoneae–Dicotyledoneae split.[37]

2.2. Expression of MGD homologs

The genes of three MGD homologs from *R. castenholzii* were amplified by polymerase chain reaction and cloned into the pET28a vector, respectively. Vectors were transformed into BL21 (DE3) competent *Escherichia coli*. Transformed cells were grown in Luria–Bertani medium at 37°C until they reached an OD_{600} reading of 0.7. Protein expression was induced using 1 mM isopropyl β-D-thiogalactopyranoside for 3 h [we designated three genes as MGD of *R. castenholzii* (RcMGD), MGlcD of *R. castenholzii* (RcMGlcD), and diglucosyl-DAG (DGlcDG) synthase of *R. castenholzii* (RcDGlcD) according to their functions determined].

2.3. Assay for glycolipid synthesis

Glycolipid synthetic activity was assayed as described by Awai *et al.*[13] Radiolabeled lipids were separated by one-dimensional thin-layer chromatography (TLC) in chloroform–hexane–isopropanol–tetrahydrofuran–water (50:100:80:1:2, by volume) or hexane–tetrahydrofuran–isopropanol–water (40:0.4:50:10, by volume) and detected by Image Analyzer (STORM860, Molecular Dynamics, Sunnyvale, CA, USA, or BAS2000, FUJIX).

2.4. Isolation of glycolipids

Cultures of *E. coli* expressing RcMGD, RcMGlcD, or RcDGlcD were centrifuged at 3500*g* for 10 min at 4°C. Pellets were homogenized in 5 ml of 50 mM TES-KOH buffer (pH 7.0) and 125 µl of 1 M $MgCl_2$. The enzyme solutions were sonicated and mixed with 115 µl of 30 mM UDP-Gal or UDP-Glc. These mixtures were incubated at 37°C for 3 h and total lipids were extracted as described previously.[38] Total lipids were then purified by column chromatography (InterSep SI, GL Sciences). Elution solvents included chloroform, acetone–isopropanol (9:1, by volume), and methanol. Glycolipids were collected in the acetone–toluene fraction and then isolated by one-dimensional TLC as above or using hexane–tetrahydrofuran–isopropanol–water (40:0.4:50:10, by volume). Finally, glycolipids were eluted using chloroform–methanol (2:1, by volume).

Roseiflexus castenholzii provided from Prof. Keizo Shimada was grown as described by Yamada *et al.*[39] Roseiflexus total lipid was also extracted as described above.[38] This lipid was developed by two-dimensional TLC using chloroform–methanol–7 N ammonium hydroxide (120:80:8, by volume) in the first dimension and chloroform–methanol–acetic acid–water (170:20:15:3, by volume) in the second.

2.5. Acetylation of glycolipids

The glycolipids isolated by one-dimensional TLC were evaporated and then eluted using 0.5 ml of pyridine and 0.25 ml of acetic acid anhydride. These mixtures were incubated overnight in the dark. Subsequently, 5 ml of toluene was added to these mixtures and evaporated. This step was repeated three times. Finally, acetylated lipids were dissolved

in chloroform and used for nuclear magnetic resonance (NMR).

2.6. Structural analysis of glycolipids by NMR

Acetylated lipids were dissolved in $CDCl_3$. These structures were analyzed by ^{1}H-NMR at 600 MHz (Bruker AV-600), 400 MHz (Varian 400-MR), or 270 MHz (JEOL JNM-EX-270) for protons. Internal tetramethylsilane [σ 0 parts per million (ppm)] in $CDCl_3$ was used as the standard. Chemical shifts were expressed in ppm with reference to the standard. The multiplicity of signals was abbreviated as follows: s, singlet; d, doublet; dd, doublet of doublets; t, triplet; and m, multiplet. The protons of the sugar head groups were denoted as H-1, H-1′, H-2, H-2′, H-3, H-3′, H-4, H-4′, H-5, H-5′, H-6a, H-6′a, H-6b, and H-6′b. The protons of the glycerol backbone were denoted as H-1″a, H-1″b, H-2″, H-3″a, and H-3″b. In MGDG synthesized by RcMGD, ^{1}H-NMR (600 MHz, $CDCl_3$) δ 5.39−5.38 (d, 1H, $J_{3',4'} = 3.3$ Hz, H-4), 5.36−5.34 (m, 1.7H, fatty acid), 5.20−5.17 (m, 2H, H-2 and H-2″), 5.01 (dd, 1H, $J_{2',3'} = 10.3$ Hz, $J_{3',4'} = 3.3$ Hz, H-3), 4.48 (d, 1H, $J_{1',2'} = 7.9$ Hz, H-1), 4.31 (dd, 1H, $J_{1a,1b} = 12.0$ Hz, $J_{1a,2} = 3.4$ Hz, H-1″a), 4.18−4.08 (m, 3.5H, H-6a, H-6b, and H-1″b), 3.96 (dd, 1H, $J_{3a,3b} = 11.0$ Hz, $J_{2,3a} = 5.0$ Hz, H-3″a), 3.90 (t, 1H, $J_{5',6'a} = J_{5',6'b} = 6.7$ Hz, H-5), and 3.77−3.66 (m, 1.7H, $J_{3a,3b} = 11.0$ Hz, $J_{2,3b} = 5.7$ Hz, H-3″b). In MGlcDG synthesized by RcMGlcD, ^{1}H-NMR (600 MHz, $CDCl_3$) δ 5.35−5.34 (m, 1H, fatty acid), 5.20 (t, 1H, $J_{2',3'} = J_{3',4'} = 9.5$ Hz, H-3), 5.21−5.18 (m, 1H, H-2″), 5.08 (t, 1H, $J_{3',4'} = J_{4',5'} = 9.5$ Hz, H-5), 4.97 (dd, 1H, $J_{1',2'} = 7.9$ Hz, $J_{2',3'} = 9.5$ Hz, H-2), 4.52 (d, 1H, $J_{1',2'} = 7.9$ Hz, H-1), 4.30 (dd, 1H, $J_{6'a,6'b} = 12.0$ Hz, $J_{5',6'} = 3.4$ Hz, H-6a), 4.26 (dd, 1H, $J_{1a,1b} = 12.4$ Hz, $J_{1a,2} = 4.7$ Hz, H-1″a), 4.14−4.11 (m, 2H, H-6b and H-1″b), 3.95 (dd, 1H, $J_{3a,3b} = 10.9$ Hz, $J_{2,3a} = 4.9$ Hz, H-3″a), and 3.69−3.66 (m, 3H, H-5, H-3″b, and fatty acid). In DGlcDG synthesized by RcDGlcD, ^{1}H-NMR (400 MHz, $CDCl_3$) δ 5.36−5.33 (m, 4H, fatty acid), 5.21−5.16 (m, 3H, H-2″, H-4, and H-4′), 5.07 (t, 1H, $J_{2'',3''} = J_{3'',4''} = 10.0$ Hz, H-3′), 4.98 (dd, 1H, $J_{1',2'} = 7.9$ Hz, $J_{2',3'} = 9.5$ Hz, H-2), 4.92 (dd, 1H, $J_{1'',2''} = 9.7$ Hz, $J_{2'',3''} = 7.9$ Hz, H-2′), 4.89 (t, 1H, $J_{2',3'} = J_{3',4'} = 9.5$ Hz, H-3), 4.57 (d, 1H, $J_{1',2'} = 7.9$ Hz, H-1), 4.47 (d, 1H, $J_{1'',2''} = 7.9$ Hz, H-1′), 4.31−4.26 (m, 2H, H-1″a and H-6′a), 4.14−4.09 (m, 2H, H-1″b and H-6′b), 3.97 (dd, 1H, $J_{6'a,6'b} = 10.8$ Hz, $J_{5',6'} = 5.1$ Hz, H-6a), 3.87 (dd, 1H, $J_{3a,3b} = 10.8$ Hz, $J_{2,3a} = 2.0$ Hz, H-3″a), and 3.73−3.65 (m, 4H, H-5, H-5′, H-3″b, and H-6b). In MGDG isolated from *R. castenholzii*, ^{1}H-NMR (270 MHz, $CDCl_3$) δ 5.38 (d, 1H, $J_{3',4'} = 3.3$ Hz, H-4), 5.20 (dd, 1H, $J_{1',2'} = 7.9$ Hz, $J_{2',3'} = 10.3$ Hz, H-2), 5.01 (dd, 1H, $J_{2',3'} = 10.5$ Hz, $J_{3',4'} = 3.3$ Hz, H-3), 5.00−4.93 (m, 1H, H-2″), 4.47 (d, 1H, $J_{1',2'} = 7.9$ Hz, H-1), 4.20−4.05 (m, 2H, H-6a and H-6b), 3.88 (dd, 2H, $J = 11.9$, 6.0 Hz, H-5 and H-1″a), 3.60 (dd, 1H, $J_{1a,1b} = 11.1$ Hz, $J_{1b,2} = 4.1$ Hz, H-1″b), and 3.15−3.10 (m, 1.7H, H-3″a and H-3″b).

3. Results and discussion

3.1. Evolutionary relationship of MGD homologs

To comprehensively understand the relationships among MGD homologs, we compared amino acid sequences of 56 MGD homologs from prokaryotes and phototrophic eukaryotes. The homologs similarly located were excluded from the phylogenetic tree. A complete list of these proteins is shown in Supplementary Table S1. Our phylogenetic analyses identified monophyletic clades of Viridiplantae, Rhodophyta (red algae), and Heterokontophyta (heterokonts), which were clearly supported by both Bayesian inference (MrBayes) and ML analyses (Treefinder and RAxML; node A in Fig. 1). Our data provide compelling evidence that a single origin of MGD in a common ancestor of these groups and the pathway for MGDG synthesis is monophyletic among Viridiplantae, Rhodophyta, and Heterokontophyta. As indicated above, MGDs from Heterokontophyta formed a single group, suggesting that the common ancestor of Heterokontophyta acquired MGDs (node B in Fig. 1). This grouping was strongly supported by MrBayes, Treefinder, and RAxML. Notably, Rhodophyta and Heterokontophyta formed a single group (node C in Fig. 1). Recent phylogenetic analyses of plastid proteins have revealed an extremely close relationship between heterokontophytes and rhodophytes. These data suggest that heterokont plastids originated from endosymbiosis with Rhodophyta.[40,41] MGDs of heterokonts were likely acquired from Rhodophyta through this endosymbiotic event.

Gloeobacter violaceus is one of the most ancient cyanobacterial lineages, and the *Gloeobacter* MGD homolog is a likely candidate for the origin of plastid MGDs. Our phylogenetic analysis of MGDs, however, indicated that this gene formed a clade with different MGD homologs in green non-sulfur bacteria (*C. aurantiacus* and *Chloroflexus aggregans*[42]) (node D in Fig. 1). Since chloroplasts may be derived from ancient types of cyanobacteria, it was expected that the eukaryote MGDs may be directly evolved from a cyanobacterial MGD origin that might be acquired by the primary endosymbiotic event, although it might be lost in most cyanobacterial species. However, here, the grouping between the *Gloeobacter* and Chloroflexi MGD homologs was robustly supported by MrBayes, Treefinder, and RAxML (node D in Fig. 1). This result strongly suggests that
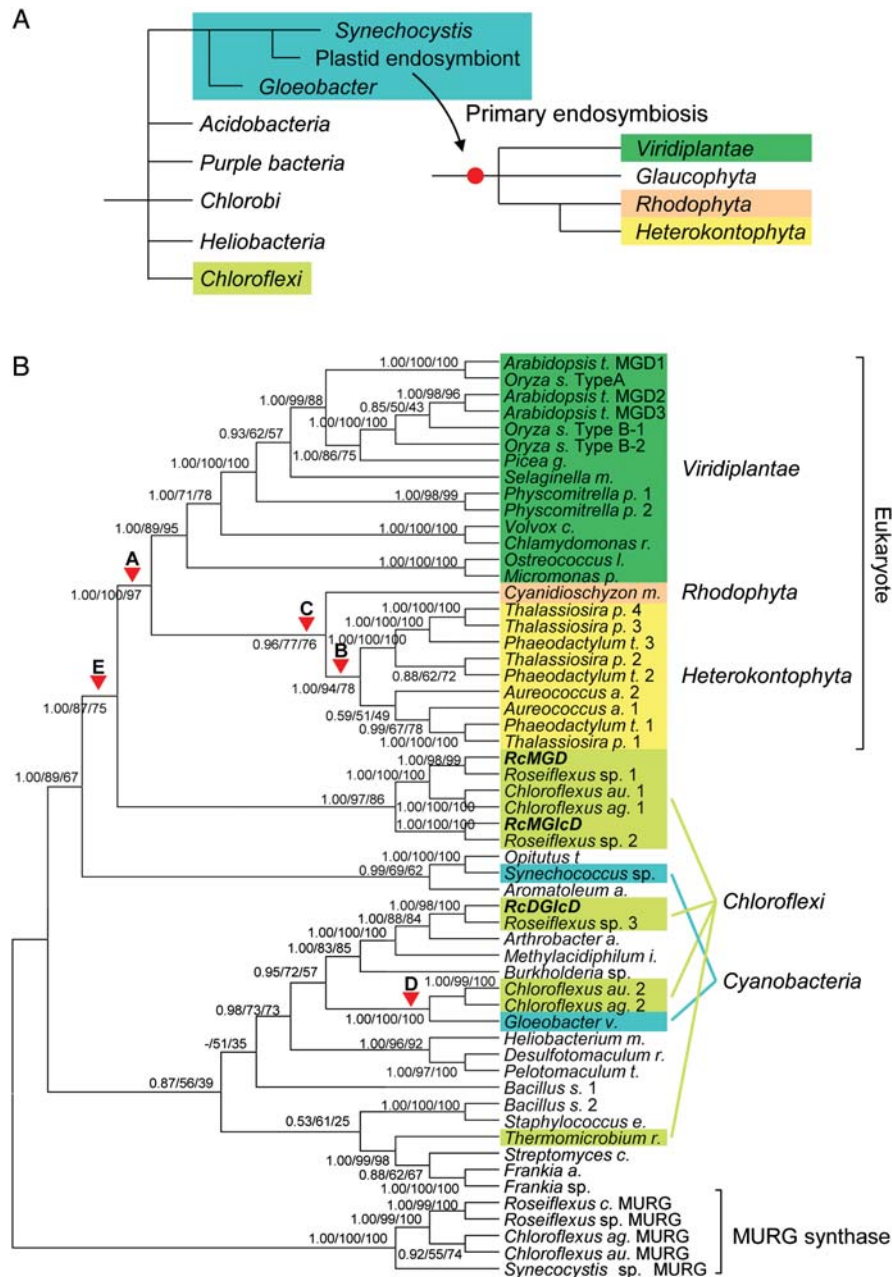
Figure 1. Phylogenetic relationship of MGD homologs. (**A**) An accepted model of chloroplast evolution.[2, 22, 53] Each red point indicates an endosymbiotic event. (**B**) A phylogenetic tree of all MGD homologs from plastids and bacteria (based on amino acid sequences). To obtain MGD sequences, blastp and tblastn searches were performed using Arabidopsis MGD amino acid sequences as queries. Above each branch is shown: 1) the posterior probability estimated with Bayesian inference, 2) LR-ELW edge supports from Treefinder analysis, and 3) the bootstrap probability from RAxML analysis (in that order). Viridiplantae, Rhodophyta, Heterokontophyta, Chloroflexi, and Cyanobacteria are indicated by green, light magenta, yellow, light green, and light blue, respectively. MURG synthase genes were used as the outgroup. The original phylogenetic trees from the 3 separate analyses (Bayesian, Treefinder, and RAxML) are shown in supplementary figure 1−3. Arrowheads indicate (**A**) the acquisition of MGD in plastids, (**B**) the acquisition of MGD homologs in Heterokontophyta, (**C**) the grouping of red algal plastids, (**D**) the horizontal gene transfer between Gloeobacter and Chloroflexus, and (**E**) the close relationship between plastids and Chloroflexi.

the *G. violaceus* MGD homolog is not the origin of MGDs in Viridiplantae and that horizontal gene transfer may have occurred much more recently between *Gloeobacter* and Chloroflexi.

However, notably, other MGD homologs of Chloroflexi strains, *Chloroflexus* and *Roseiflexus*, were located closest to the clade of chloroplasts (node E in Fig. 1). *Chloroflexus* au. 1, an MGD of *C. aurantiacus*, belongs to the GT28 family and the encoded protein has an MGDG synthetic activity.[19] In addition, *R. castenholzii* has three MGD homologs that we have named RcMGD, RcMGlcD, and RcDGlcD according

to their functions determined (see below). These genes also belong to the GT28 family, and RcMGD and RcMGlcD form a clade with *Chloroflexus* au. 1. Although MGD homologs from Chloroflexi and chloroplasts were closely related phylogenetically, information regarding substrate specificity was only available for *Chloroflexus* au. 1. We have therefore examined in more detail the enzymatic properties of Chloroflexi MGD homologs.

3.2. Glycolipid synthetic activities of R. castenholzii MGD homologs

To characterize the glycolipid synthetic activities of the three *R. castenholzii* MGD homologs, we expressed these recombinant proteins in *E. coli*. When radiolabeled UDP-Gal was used as a substrate, RcMGD was capable of synthesizing a radiolabeled lipid (Fig. 2). We used UDP-Gal as a substrate because the proposed pathway for MGDG synthesis of higher plants utilizes UDP-Gal. Cucumber (*Cucumis sativus*) MGD1 (CsMGD) was used as a positive control for MGDG synthesis in this experiment. In contrast, RcMGD extract did not contain a radiolabeled product, when UDP-Glc was provided as a substrate, suggesting that lipid synthesis via RcMGD was UDP-Gal-dependent. On the other hand, the enzyme encoded by RcMGlcD synthesized MGlcDG when UDP-Glc was provided as a substrate. The MGlcD of *Synechocystis*

sp. PCC 6803 (Sll1377) was used as a positive control for MGlcDG synthesis in this experiment.

Using TLC, lipids from two samples (RcMGD/UDP-Gal and RcMGlcD/UDP-Glc) were purified and subjected to ^{1}H-NMR analysis to reveal the sugar head groups of these lipids (Fig. 3). In Figs 3A and B, the doublet peak indicated by the black arrow at 4.5−4.6 ppm indicates H1 of the hexose moiety of the lipid. These peaks demonstrated that the sugar head group was bound to a glycerol backbone via a β-anomeric configuration at the C3 position, an MGDG structure that is characteristic of land plants.[14] Consequently, the clade in Fig. 1, which includes *Chloroflexus* and *Roseiflexus*, represents green non-sulfur bacterial MGDs. The present results indicate that MGD genes occur at least at the time of Chloroflexi divergence (an ancient bacterial lineage). Botté et al.[43] has independently performed phylogenetic analysis of MGDs. Although they indicated similar results for phylogeny of MGDs, they have not included prokaryotic MGD homologs except for Chloroflexi in
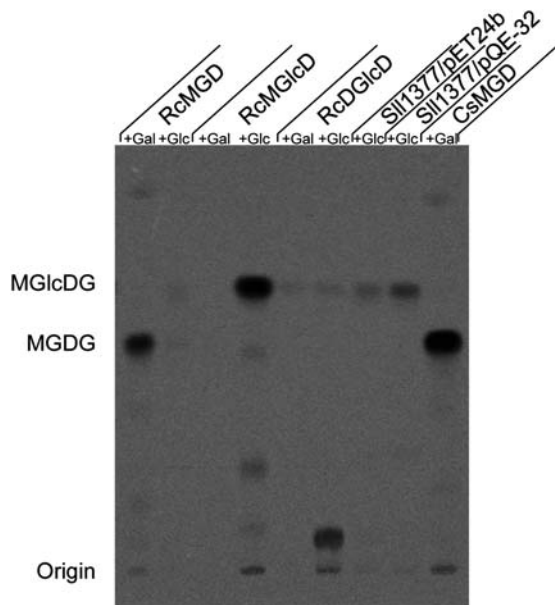

Figure 2. Analysis of sugar transferase activity. Radiolabeled UDP-Gal (+Gal) or UDP-Glc (+Glc) were used as substrates. The resulting lipids were separated by TLC (chloroform-hexane-tetrahydrofuran-isopropanol-water, 50:100:1:80:2 by volume) and visualized by autoradiography. CsMGD and Sll1377 (pET24b and pQE-32) were used as controls for MGDG and MGlcDG, respectively.
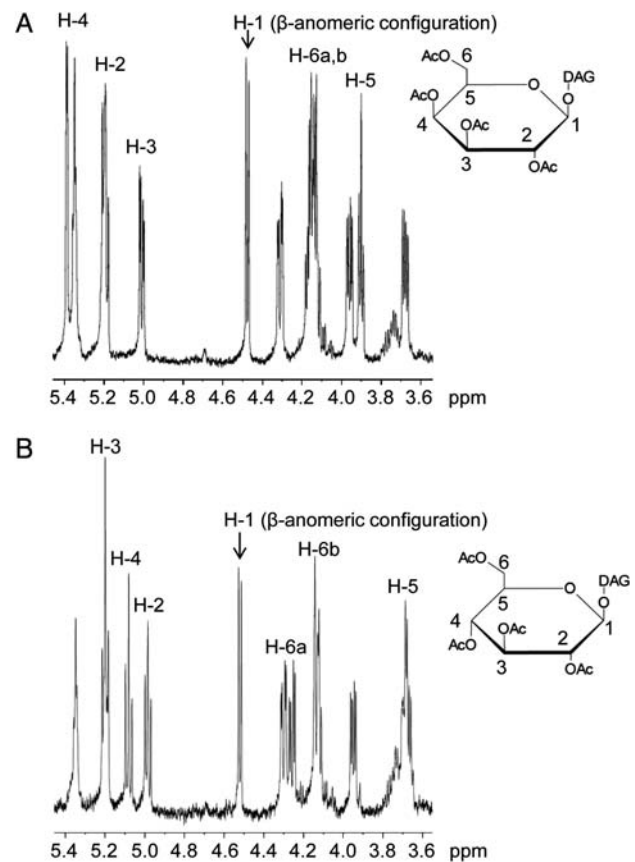

Figure 3. Analysis of glycolipids by ^{1}H-NMR. Each glycolipid synthesized by RcMGD and RcMGlcD was analyzed by ^{1}H-NMR at 600 MHz. (**A**) MGDG synthesized by RcMGD expression and (**B**) MGlcDG synthesized by RcMGlcD expression were isolated from *E. coli*. H-1 doublet peaks (arrows) indicate that these glycolipids formed β-anomeric configurations to the glycerol backbone.
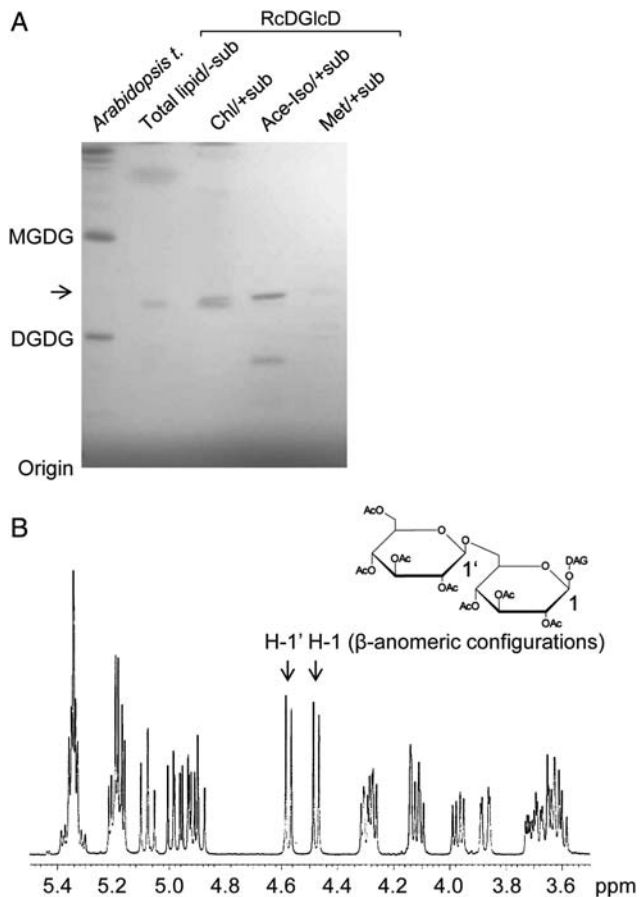
Figure 4. Analysis of glycolipids synthesized by RcDGlcD. Glycolipids synthesized by RcDGlcD were analyzed by TLC (**A**) and ^{1}H-NMR at 400 MHz (**B**). (**A**) RcDGlcD was incubated with UDP-Glc and DAG (+sub), or without substrate (−sub). Synthesized lipids were purified by column chromatography. Fractions eluted either with chloroform (Chl), acetone-isopropanol (9:1, by volume, Ace-Iso), or methanol (Met) were developed using hexane-tetrahydrofuran-isopropanol-water (40:0.4:50:10, by volume). Glycolipids were detected by the anthrone reagent (MGDG, DGDG, and the lipid indicated by an arrow). (**B**) ^{1}H-NMR lipid analysis resulted in doublet H-1 peaks (arrows), indicating β-anomeric configurations.

their phylogenetic analysis, and thus, the phylogenetic relationship between eukaryotic and prokaryotic MGD homologs remains uncertain. Our detailed analyses suggest that this ancestral type of MGD genes is the origin of higher plant MGDs.

RcDGlcD synthesized a glycolipid that consisted of a number of sugars, using UDP-Glc as a substrate (Fig. 2). The crude lipid extract from *E. coli* overexpressing RcDGlcD contained much phospholipids derived from *E. coli*. To remove phospholipids and concentrate glycolipids, the extract was purified by silica column chromatography. The separated fractions were further developed by TLC and stained using the anthrone reagent (Fig. 4A). The lipid mobility was associated with DGlcDG.[19] ^{1}H-NMR analysis identified this lipid as DGlcDG. The analysis showed two doublet peaks

H1 and H′1 (4.4−4.6 ppm) indicating two hexose moieties in the lipid (Fig. 4B). The spectrum including these peaks revealed that this lipid contained two glucoses and these glucoses were bound to each other via a β-anomeric configuration. Furthermore, one of these glucoses was also bound to a glycerol backbone via a β-anomeric configuration. MGD gene homologs share a high degree of sequence similarity, yet sugar-donor specificity and the number of molecules involved in the reaction are clearly diverse. Detailed comparisons of MGD homolog sequences could potentially reveal novel protein domains that determine sugar-donor specificity.

3.3. Lipid analysis of R. castenholzii

We determined the activity of the RcMGD. Subsequently, to define whether *R. castenholzii* possesses MGDG, MGlcDG, and DGlcDG or not, we analyzed the lipid composition of *R. castenholzii*. Total lipid was extracted and developed by two-dimensional TLC, and three glycolipids were detected by the anthrone reagent (Fig. 5A). We compared the mobility on two-dimensional TLC between *A. thaliana* and *R. castenholzii* (Supplementary Fig. S4). The mobility of a *Roseiflexus* glycolipid was very similar to *Arabidopsis* MGDG. Using ^{1}H-NMR, we confirmed that this glycolipid was a galactolipid and formed a β-anomeric configuration to the glycerol backbone (Fig. 5B). From these results, we concluded that this glycolipid is MGDG. Although the peak pattern derived from the glycerol backbone was slightly different from MGDG synthesized by RcMGD in *E. coli*, it is probably due to the difference in fatty acid composition of these MGDGs (Supplementary Table S3). This result strongly supports that RcMGD actually functions *in vivo* to synthesize MGDG.

As indicated above, we also detected two other glycolipids on two-dimensional TLC. One of them was assumed to be sulfoquinovosyl-DAG (SQDG) from the mobility on TLC. In fact, there is a homolog of *Arabidopsis* SQDG synthase 1 in *Roseiflexus* and the presence of SQDG in *C. aurantiacus* was reported in a previous paper.[44] On the other hand, in another one (UL4, unknown lipid 4, in Fig. 5A), the corresponding lipid was not existent in *A. thaliana*. This glycolipid was likely DGlcDG. However, ^{1}H-NMR analysis revealed that this lipid was not DGlcDG, although it consisted of at least two sugar head groups (data not shown). We could not detect MGlcDG in total lipid fraction from *Roseiflexus*. In Cyanobacteria, MGlcDG is a minor lipid, less than 1% of total lipid.[10] Similarly, MGlcDG may not be abundant in *Roseiflexus*. Therefore, these results suggest that RcDGlcD mainly incorporates glucose to a major monoglycolipid MGDG, not MGlcDG *in vivo*.

To confirm the glucose transfer activity of RcDGlcD, we used MGDG and MGlcDG as sugar acceptors and compared the diglycolipid producing activity (Fig. 6). Obviously, RcDGlcD could transfer glucose to MGDG, although DGlcDG is still produced by it's processive activity. When MGlcDG was used as a substrate, the band corresponding to DGlcDG was increased, confirming that RcDGlcD could also transfer glucose to MGlcDG. However, UL4 was not DGlcDG and MGDG is one of the most abundant lipids in *Roseiflexus*

(Supplementary Table S3). Consequently, we presume that the main sugar acceptor of RcDGlcD *in vivo* is MGDG and UL4 is glycosylgalactosyl-DAG. However, since we only utilized UDP-Gal and UDP-Glc as *in vitro* substrates (Fig. 2), its true substrate is still uncertain, and further experiments are needed to determine the structure of UL4.

Thylakoid membranes or membranous structures of anoxygenic photosynthetic bacteria involve complexes composed of proteins, pigments, and lipids. In particular, these types of membranes are characterized by a unique and conserved lipid composition. Genes related to the photosynthetic membrane biogenesis, therefore, have also played an important role in phototroph evolution. Recently, we reported that a distinct type of MGD belonging to the GT1 family was found in the green sulfur bacterium *Chlorobacterium tepidum* and proved to be involved in chlorosome biogenesis.[45] This work also demonstrated that MGDG is commonly important in both anoxygenic and oxygenic photosynthetic organisms,
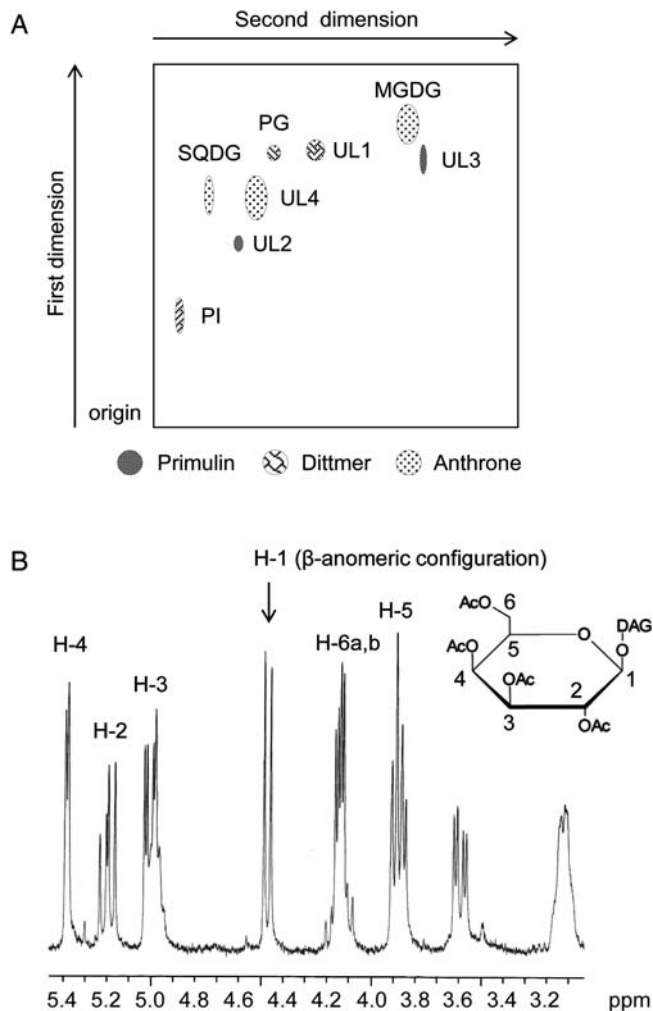

Figure 5. Analysis of MGDG isolated from *R. castenholzii*. (**A**) Description of 2-dimensional TLC in *R. castenholzii*. The TLC analysis was performed with chloroform-methanol- 7 N ammonium hydroxide (120:80:8, by volume) in the first dimension and chloroform-methanol-acetic acid-water (170:20:15:3, by volume) in the second. The lipids were visualized by spraying with primulin reagent and viewing under ultraviolet light. The glycolipids were visualized by anthrone reagent and The phospholipids were visualized by Dittmer's reagent. PG, phosphatidylglycerol; SQDG, sulfoquinovosyldiacylglycerol; PI, phosphatidylinositol; UL, unknown lipid. (**B**) MGDG isolated from *R. castenholzii* was analyzed by ^{1}H-NMR at 270 MHz. The MGDG formed a β-anomeric configuration to the glycerol backbone.
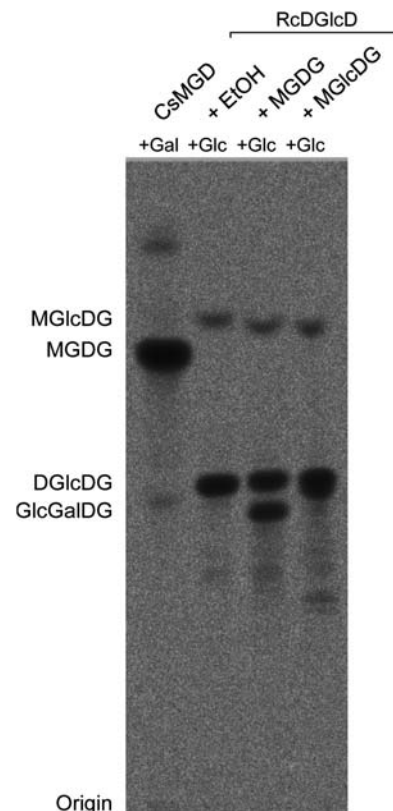

Figure 6. Analysis of sugar transferase activity of RcDGlcD. Radiolabeled UDP-Gal (+Gal) or UDP-Glc (+Glc) were used as substrates. Additionally, MGDG or MGlcDG eluted in EtOH were added as acceptors. The resulting lipids were separated by TLC (hexane-tetrahydrofuran-isopropanol-water, 40:0.4:50:10, by volume) and visualized by autoradiography. CsMGD was used as controls for MGDG. GlcGalDG, glucosylgalactosyldiacylglycerol.

although green sulfur and non-sulfur bacteria have independently acquired different types of MGDG synthetic machineries. Given the conservation of MGD gene families and the presence of MGDG in Chloroflexi, a functional association between the photosystem and the membrane galactolipid MGDG was established before endosymbiosis event, leading to the establishment of highly organized photosynthetic membranes.

Our results suggested that the photosynthetic eukaryote has acquired MGD gene from ancestral Chloroflexi in an early event, even though Cyanobacteria has a different MGDG synthetic pathway. These pathways were subsequently conserved during the course of evolution and have played important roles in chloroplast biology. A recent large-scale phylogenetic analysis also revealed that a considerable amount of non-cyanobacterial genomic material has contributed to the establishment of the plastid before the split of red and green algae.[46] Eukaryotic MGD genes may be acquired through this kind of large-scale gene transfer from ancestral Chloroflexi.

3.4.   The divergence point of type A and type B MGDs

Determining the time at which the two types of MGDs (type A and type B) diverged is important to understand the evolution of photosynthetic membranes in land plants. Using both Bayesian inference and ML methods, we performed phylogenetic analyses on MGD gene sequences of land plants, excluding third codon positions (Supplementary Figs S5 and S6). Results from these analyses generally agreed with accepted phylogenetic relationships of land plants. For example, Angiospermae were clearly divided into Monocotyledoneae (monocots) and Dicotyledoneae (dicots) in both types of MGDs. The groupings received strong support from both Bayesian inference and ML methods (Supplementary Figs S5 and S6). These results indicate that MGDs have been conserved throughout the course of land plant evolution. Importantly, one full-length MGD gene sequence of *Picea glauca* was available among Gymnospermae, and both Bayesian and ML analyses placed the *Picea* sequence close to the type B MGD clade. Although a Gymnospermae type A MGD has not yet been identified, this result suggests that the type A/B divergence may have preceded that of Angiospermae and Gymnospermae.

Based on the phylogenetic tree, we estimated an MGD type A/B divergence time using four calibration points (similar to Yoon *et al.*[34]). Our analysis indicated that MGDs likely diverged ~323 MYA, with 298−357 MYA representing the 95% confidence interval (Fig. 7). This divergence time corresponds to the

Carboniferous period, which came after Spermatophyta appeared (355−370 MYA). We propose, therefore, that the acquisition of two types of MGDs occurred in the common ancestor of Spermatophyta.

Our data indicate that MGD gene duplication and subsequent functional differentiation occurred in Spermatophyta (Fig. 8). Jiao *et al.*[47] showed that a whole-genome duplication (WGD) event occurred in the common ancestor of all seed plants. This WGD allowed for major adaptive changes in these organisms and contributed to the dominance of seed plants. Interestingly, type B *MGD* expression is elevated during the reproduction stage of seed plants.[27] During the Carboniferous period, spore-producing plants were dominant,[48] whereas this dominance shifted to seed plants shortly thereafter. The Carboniferous and Permian periods were also characterized by dramatic changes to the global climate, primarily drying. Expression of type B MGDs is elevated during phosphate deprivation, a response that is regulated by both auxin and cytokinin.[49] These regulatory mechanisms may have worked in response to these climate changes. More extensive research concerning the MGDs of Gymnospermae is required for a thorough understanding of MGD gene evolution. It is
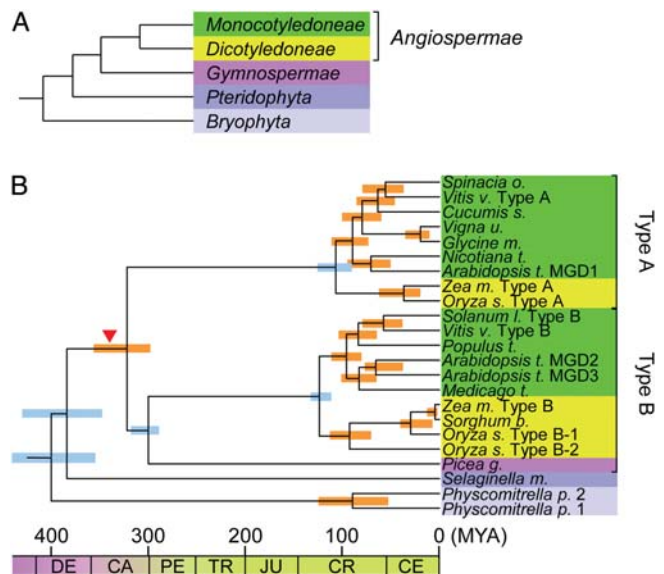

Figure 7. The divergence time of type A and type B MGDs. (**A**) A generally accepted evolution model of land.[54,55] (**B**) Divergence time estimation among MGDs of land plants based on the nucleotide dataset of the 1st and 2nd codon positions. Horizontal bars (blue and orange) indicate 95% credible intervals of the divergence time estimates. The 5 nodes used as time constraints are indicated with blue bars. The original ML and Bayesian trees used for the divergence time estimation are shown in supplementary figure 4 and 5. Arrowhead indicates the divergence of type A and type B MGDs. DE, Devonian; CA, Carboniferous; PE, Permian; TR, Triassic; JU, Jurassic; CR, Cretaceous; CE, Cenozoic.
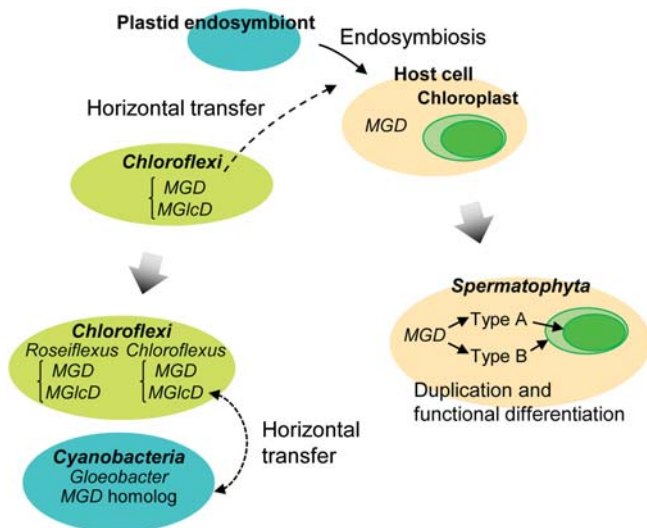
Figure 8. Proposed model for MGD evolution. The requirement of MGDG in photosynthetic processes was established in ancient times. For MGDs, duplication and functional divergence occurred in *Spermatophyta*. Black dashed arrows indicate horizontal gene transfer. Large, shaded black arrows indicate evolutionary progression.

important to determine whether there is a species of Gymnospermae that has two types of MGDs. The evolution of Gymnospermae, however, has not been adequately mapped, in part because a complete genome sequence is not available. Determining the divergence time of type A and B MGDs in Gymnospermae may prove critical if we are to understand the divergence of land plants.

4. Conclusion

Based on phylogenetic and enzymatic analyses, we proposed the evolutionary model of MGDs (Fig. 8). Although no cyanobacteria analyzed possessed the ancestral type of higher plant MGDs, we identified the possible MGD ancestor from Chloroflexi. The results suggest that higher plant MGD originated from this type of Chloroflexi MGD. Eukaryotic MGD may be acquired from ancient Chloroflexi via horizontal gene transfer in parallel with the major endosymbiotic event and contributed to the early plastid evolution. After the acquirement, MGD duplication and functional differentiation of two types of MGD, types A and B, occurred along with land plant evolution. This functional differentiation probably had important meanings in further prosperity of seed plants. It is universally accepted that an ancestral cyanobacterium is the origin of the plastid organelle; however, plastid evolution is still heavily debated. These disagreements can be attributed to a very

complex process that involved both horizontal and intracellular gene transfer during the course of prokaryotic and eukaryotic evolution.[50,51] To generate a more accurate picture of photosynthesis evolution, we believe that focused comprehensive analyses of specific physiological processes are just as important as more global gene sequence comparisons. In particular, galactolipids are highly enriched in photosynthetic membranes, and deciphering the evolution of galactolipid synthetic pathways provides critical information toward understanding the early evolution of photosynthesis.

Supplementary Data: Supplementary Data are available at www.dnaresearch.oxfordjournals.org.

Funding

References

1. McFadden, G.I. 2001, Primary and secondary endosymbiosis and the origin of plastids, *J. Phycol.*, **37**, 951–9.
2. Nelissen, B., Van de Peer, Y., Wilmotte, A. and De Wachter, R. 1995, An early origin of plastids within the cyanobacterial divergence is suggested by evolutionary trees based on complete 16S rRNA sequences, *Mol. Biol. Evol.*, **12**, 1166–73.
3. Martin, W., Rujan, T., Richly, E., et al. 2002, Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus, *Proc. Natl Acad. Sci. USA*, **99**, 12246–51.
4. Wolfe, G.R., Cunningham, F.X., Durnford, D., Green, B.R. and Gantt, E. 1994, Evidence for a common origin of chloroplasts with light-harvesting complexes of different pigmentation, *Nature*, **367**, 566–8.
5. Osteryoung, K.W. and Vierling, E. 1995, Conserved cell and organelle division, *Nature*, **376**, 473–4.
6. Joyard, J., Marechal, E., Miege, C., Block, M.A., Dorne, A.J. and Douce, R. 1998, Structure, distribution and biosynthesis of glycerolipids from higher plant chloroplasts. In: Siegenthaler, P.A. and Murata, N., eds. *Lipid in Photosynthesis: Structure, Function and Genetics*, Kluwer Academic Publishers: Dordrecht, pp. 21–52.

7. Jordan, P., Fromme, P., Witt, H.T., Klukas, O., Saenger, W. and Krauss, N. 2001, Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution, *Nature*, **411**, 909−17.

8. Loll, B., Kern, J., Saenger, W., Zouni, A. and Biesiadka, J. 2005, Towards complete cofactor arrangement in the 3.0 Å resolution structure of photosystem II, *Nature*, **438**, 1040−4.

9. Kobayashi, K., Kondo, M., Fukuda, H., Nishimura, M. and Ohta, H. 2007, Galactolipid synthesis in chloroplast inner envelope is essential for proper thylakoid biogenesis, photosynthesis, and embryogenesis, *Proc. Natl Acad. Sci. USA*, **104**, 17216−21.

10. Sato, N. and Murata, N. 1982, Lipid biosynthesis in the blue-green alga, *Anabaena variabilis* I. Lipid classes, *Biochim. Biophys. Acta*, **710**, 271−8.

11. Shimojima, M., Ohta, H., Iwamatsu, A., Masuda, T., Shioi, Y. and Takamiya, K. 1997, Cloning of the gene for monogalactosyldiacylglycerol synthase and its evolutionary origin, *Proc. Natl Acad. Sci. USA*, **94**, 333−7.

12. Dörmann, P. and Benning, C. 2002, Galactolipids rule in seed plants, *Trends Plant. Sci.*, **7**, 112−8.

13. Awai, K., Kakimoto, T., Awai, C., et al. 2006, Comparative genomic analysis revealed a gene for monoglucosyldiacylglycerol synthase, an enzyme for photosynthetic membrane lipid synthesis in cyanobacteria, *Plant Physiol.*, **141**, 1120−7.

14. Hölzl, G. and Dörmann, P. 2007, Structure and function of glycoglycerolipids in plants and bacteria, *Prog. Lipid Res.*, **46**, 225−43.

15. Woese, C.R. 1987, Bacterial evolution, *Microbiol. Rev.*, **51**, 221−71.

16. Gupta, R.S., Mukhtar, T. and Singh, B. 1999, Evolutionary relationships among photosynthetic prokaryotes (*Heliobacterium chlorum*, *Chloroflexus aurantiacus*, cyanobacteria, *Chlorobium tepidum* and proteobacteria): implications regarding the origin of photosynthesis, *Mol. Microbiol.*, **32**, 893−906.

17. Xiong, J., Fischer, W.M., Inoue, K., Nakahara, M. and Bauer, C.E. 2000, Molecular evidence for the early evolution of photosynthesis, *Science*, **289**, 1724−30.

18. Pierson, B.K. and Castenholz, R.W. 1974, A phototrophic gliding filamentous bacterium of hot springs, *Chloroflexus aurantiacus*, gen. and sp. nov, *Arch. Microbiol.*, **100**, 5−24.

19. Hölzl, G., Zähringer, U., Warnecke, D. and Heinz, E. 2005, Glycoengineering of cyanobacterial thylakoid membranes for future studies on the role of glycolipids in photosynthesis, *Plant Cell Physiol.*, **46**, 1766−78.

20. Hanada, S., Takaichi, S., Matsuura, K. and Nakamura, K. 2002, *Roseiflexus castenholzii* gen. nov., sp. nov., a thermophilic, filamentous, photosynthetic bacterium that lacks chlorosomes, *Int. J. Syst. Evol. Microbiol.*, **52**, 187−93.

21. Awai, K., Maréchal, E., Block, M.A., et al. 2001, Two types of MGDG synthase genes, found widely in both 16:3 and 18:3 plants, differentially mediate galactolipid syntheses in photosynthetic and nonphotosynthetic tissues in *Arabidopsis thaliana*, *Proc. Natl Acad. Sci. USA*, **98**, 10960−5.

22. Kobayashi, K., Awai, K., Nakamura, M., Nagatani, A., Masuda, T. and Ohta, H. 2009, Type-B monogalactosyldiacylglycerol synthases are involved in phosphate starvation-induced lipid remodeling, and are crucial for low-phosphate adaptation, *Plant J.*, **57**, 322−31.

23. Andersson, M.X., Stridh, M.H., Larsson, K.E., Liljenberg, C. and Sandelius, A.S. 2003, Phosphate-deficient oat replaces a major portion of the plasma membrane phospholipids with the galactolipid digalactosyldiacylglycerol, *FEBS Lett.*, **537**, 128−32.

24. Gaude, N., Tippmann, H., Flemetakis, E., Katinakis, P., Udvardi, M. and Dörmann, P. 2004, The galactolipid digalactosyldiacylglycerol accumulates in the peribacteroid membrane of nitrogen-fixing nodules of soybean and *Lotus*, *J. Biol. Chem.*, **279**, 34624−30.

25. Jouhet, J., Maréchal, E., Baldan, B., Bligny, R., Joyard, J. and Block, M.A. 2004, Phosphate deprivation induces transfer of DGDG galactolipid from chloroplast to mitochondria, *J. Cell Biol.*, **167**, 863−74.

26. Russo, M.A., Quartacci, M.F., Izzo, R., Belligno, A. and Navari-Izzo, F. 2007, Long- and short-term phosphate deprivation in bean roots: plasma membrane lipid alterations and transient stimulation of phospholipases, *Phytochemistry*, **68**, 1564−71.

27. Kobayashi, K., Awai, K., Takamiya, K. and Ohta, H. 2004, *Arabidopsis* type B monogalactosyldiacylglycerol synthase genes are expressed during pollen tube growth and induced by phosphate starvation, *Plant Physiol.*, **134**, 640−8.

28. Thompson, J.D., Muller, A., Waterhouse, A., et al. 2006, MACSIMS: multiple alignment of complete sequences information management system, *BMC Bioinformatics*, **7**, 318.

29. Edgar, R.C. 2004, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic. Acids Res.*, **32**, 1792−7.

30. Ronquist, F. and Huelsenbeck, J.P. 2003, MrBayes 3: Bayesian phylogenetic inference under mixed models, *Bioinformatics*, **19**, 1572−4.

31. Jobb, G., von Haeseler, A. and Strimmer, K. 2004, TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics, *BMC Evol. Biol.*, **4**, 18.

32. Stamatakis, A. 2006, RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models, *Bioinformatics*, **22**, 2688−90.

33. Yang, Z. 2007, PAML 4: phylogenetic analysis by maximum likelihood, *Mol. Biol. Evol.*, **24**, 1586−91.

34. Yoon, H.S., Hackett, J.D., Ciniglia, C., Pinto, G. and Bhattacharya, D. 2004, A molecular timeline for the origin of photosynthetic eukaryotes, *Mol. Biol. Evol.*, **21**, 809−18.

35. Kenrick, P. and Crane, P.R. 1997, The origin and early evolution of plants on land, *Nature*, **389**, 33−9.

36. Gillespie, W.H., Rothwell, G.W. and Scheckler, S.E. 1981, The earliest seeds, *Nature*, **293**, 462−4.

37. Crane, P.R., Friis, E.M. and Pedersen, K.R. 1995, The origin and early diversification of angiosperms, *Nature*, **374**, 27−33.

38. Bligh, E.G. and Dyer, W.J. 1959, A rapid method of total lipid extraction and purification, *Can. J. Biochem. Physiol.*, **37**, 911−7.

39. Yamada, M., Zhang, H., Hanada, S., Nagashima, K.V., Shimada, K. and Matsuura, K. 2005, Structural and spectroscopic properties of a reaction center complex from the chlorosome-lacking filamentous anoxygenic-phototrophic bacterium *Roseiflexus castenholzii*, *J. Bacteriol.*, **187**, 1702−9.

40. Yoon, H.S., Hackett, J.D., Pinto, G. and Bhattacharya, D. 2002, The single, ancient origin of chromist plastids, *Proc. Natl Acad. Sci. USA*, **99**, 15507−12.

41. Harper, J.T. and Keeling, P.J. 2003, Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids, *Mol. Biol. Evol.*, **20**, 1730−5.

42. Hanada, S., Hiraishi, A., Shimada, K. and Matsuura, K. 1995, *Chloroflexus aggregans* sp. nov., a filamentous phototrophic bacterium which forms dense cell aggregates by active gliding movement, *Int. J. Syst. Bacteriol.*, **45**, 676−81.

43. Botté, C.Y., Yamaryo-Botté, Y., Janouskovec, J., et al. 2011, Identification of plant-like galactolipids in *Chromera velia*, a photosynthetic relative of malaria parasites, *J. Biol. Chem.*, **286**, 29893−903.

44. Knudsen, E., Jantzen, E., Bryn, K., Ormerod, J. and Sirevåg, R. 1982, Quantitative and structural characterization of lipids in *Chlorobium* and *Chloroflexus*, *Arch. Microbiol.*, **132**, 149−54.

45. Masuda, S., Harada, J., Yokono, M., et al. 2011, A monogalactosyldiacylglycerol synthase found in the green sulfur bacterium *Chlorobaculum tepidum* reveals important roles for galactolipids in photosynthesis, *Plant Cell*, **23**, 2644−58.

46. Suzuki, K. and Miyagishima, S.Y. 2010, Eukaryotic and eubacterial contributions to the establishment of plastid proteome estimated by large-scale phylogenetic analyses, *Mol. Biol. Evol.*, **27**, 581−90.

47. Jiao, Y., Wickett, N.J., Ayyampalayam, S., et al. 2011, Ancestral polyploidy in seed plants and angiosperms, *Nature*, **473**, 97−100.

48. DiMichele, W.A., Pfefferkorn, H.W. and Gastaldom, R.A. 2001, Response of late carboniferous and early Permian plant communities to climate change, *Annu. Rev. Earth Planet Sci.*, **29**, 461−87.

49. Kobayashi, K., Masuda, T., Takamiya, K. and Ohta, H. 2006, Membrane lipid alteration during phosphate starvation is regulated by phosphate signaling and auxin/cytokinin cross-talk, *Plant J.*, **47**, 238−48.

50. Raymond, J., Zhaxybayeva, O., Gogarten, J.P., Gerdes, S.Y. and Blankenship, R.E. 2002, Whole-genome analysis of photosynthetic prokaryotes, *Science*, **298**, 1616−20.

51. Keeling, P.J. and Palmer, J.D. 2008, Horizontal gene transfer in eukaryotic evolution, *Nat. Rev. Genet.*, **9**, 605−18.

52. Turner, S., Pryer, K.M., Miao, V.P. and Palmer, J.D. 1999, Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis, *J. Eukaryot. Microbiol.*, **46**, 327−38.

53. Blankenship, R.E. 2010, Early evolution of photosynthesis, *Plant Physiol.*, **154**, 434−8.

54. Pryer, K.M., Schneider, H., Smith, A.R., et al. 2001, Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants, *Nature*, **409**, 618−22.

55. Palmer, J.D., Soltis, D.E. and Chase, M.W. 2004, The plant tree of life: an overview and some points of view, *Am. J. Bot.*, **91**, 1437−45.