

The MEP-pathway genes are polyphyletic origin and evolutionary conserved

Liping Zeng

University of California Riverside

Katayoon Dehesh (✉ kdehesh@ucr.edu)

UC Riverside: University of California Riverside <https://orcid.org/0000-0002-9312-8032>

Research article

Keywords: isoprenoid, MEP-pathway, plastid-bearing eukaryotes, phylogenetic, polyphyletic

Posted Date: November 24th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-112826/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on February 26th, 2021. See the published version at <https://doi.org/10.1186/s12864-021-07448-x>.

Abstract

Background

Isoprenoids are amongst the most ancient and essential class of metabolites produced in all organisms, either via mevalonate (MVA)-and/or methylerythritol dicyclophosphate (MEP)-pathways. The MEP-pathway is present in all plastid-bearing organisms and most eubacteria. However, no comprehensive study reveals the origination and evolutionary characteristics of MEP-pathway genes in eukaryotes.

Results

Here, detailed bioinformatics analyses of the MEP-pathway provide an in-depth understanding the evolutionary history of this indispensable biochemical route, and offer a basis for the co-existence of the cytosolic MVA- and plastidial MEP-pathway in plants given the established exchange of the end products between the two isoprenoid-biosynthesis pathways. Here, phylogenetic analyses establish the contributions of both cyanobacteria and Chlamydiae sequences to the plant's MEP-pathway genes. Moreover, Phylogenetic and inter-species syntenic block analyses demonstrate that six of the seven MEP-pathway genes have predominantly remained as single-copy in land plants in spite of multiple whole-genome duplication events (WGDs). Substitution rate and domain studies display the evolutionary conservation of these genes, reinforced by their high expression levels. Distinct phenotypic variation among plants with reduced expression levels of individual MEP-pathway genes confirm the indispensable function of each nuclear-encoded plastid-targeted MEP-pathway enzyme in plant growth and development.

Conclusion

Collectively, these findings reveal the polyphyletic origin and restrict conservation of MEP-pathway genes, and reinforce the potential function of the individual enzymes beyond production of the isoprenoids intermediates.

Background

With over 55,000 molecules, isoprenoids are amongst the most ancient group of structurally and functionally diverse metabolites essential for all kingdoms of life (Brocks, et al. 1999). Isoprenoid-derived compounds in free-living organisms range from hormones, lipids, pigments, vitamins, electron transport chain and defense compounds, and as such of industrial interests for drugs, agrochemicals, rubber and fragrances (Hoshino and Gaucher 2018). However, despite their diversity, isoprenoids are derived from two universal five-carbon precursors, isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP) (Sacchettini and Poulter 1997). These precursors are synthesized either by mevalonate- (MVA) (Wright 1961) and/or by the alternative route methyl erythritol phosphate (MEP)-pathway (Rohmer, et al. 1993). Almost all eukaryotes, archaeobacteria and some gram-positive bacteria employ the MVA-pathway, whereas most gram-negative bacteria, cyanobacteria and green algae

exclusively use MEP-pathway (Fig. 1A) (Lombard and Moreira 2010). Plastid-bearing eukaryotes however are unique as they have retained both pathways compartmentalized in the cytosol (MVA) and plastids (MEP) (Hoshino and Gaucher 2018). It is suggested that retention of the two pathways in the two different subcellular compartments of the plastid-bearing eukaryotic cell is to regulate isoprenoid biosynthesis according to the availability of carbon and energy currencies, and as a strategy to balance resource allocation between growth and adaptive responses to unfavorable environmental inputs (Ruiz-Sola, et al. 2016). Given the established metabolic exchanges between MVA- and MEP-pathway in higher plants (Bick and Lange 2003; Hemmerlin, et al. 2003; Laule, et al. 2003), the biological grounds for the indispensable function of each of these pathways in plants has remained an enigma.

One of the most profound outcome of evolution is the emergence of plastids through a single endosymbiotic event accompanied by a complex mix of loss, movement and replacement in the ancestor of eukaryotes (Keeling 2010). The endosymbiotic events that led to the origination of plastids were ensued by the transfer of genetic material from the endosymbiont to the nuclear genome of the host, followed by the evolution of protein import machinery for transferring nuclear-encoded plastid-targeted proteins and by extension the inevitable establishment of plastids-to-nucleus (retrograde signaling) signaling cascades (Archibald 2015; Woodson and Chory 2008). The retrograde signaling cascade is instrumental for coordination of vital activities between the two subcellular genomes in plastid-bearing eukaryotes.

One essential plastidial biochemical route is the MEP-pathway, responsible for catalyzing glyceraldehyde 3-phosphate and pyruvate into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), the central intermediates in the biosynthesis of isoprenoids (Fig. 1B). The MEP-pathway is comprised of seven nuclear genes encoding plastid-localized enzymes. Intriguingly, one of the MEP-pathway intermediates, MEcPP (2-C-methyl-D-erythritol-2, 4-cyclopyrophosphate), is found to be a bi-functional entity severing as a precursor of isoprenoids and as a stress-specific retrograde signaling metabolite coordinating expression of selected stress-response nuclear genes (Walley, et al. 2015; Xiao, et al. 2012).

Given the antiquity and essential function of isoprenoids, the evolutionary history of the MVA-pathway in eukaryotes is extensively examined (Hoshino and Gaucher 2018; Lombard and Moreira 2010) whereas, the characteristics of MEP-pathway genes is thus far restricted to limited species and as such incomplete (Lange, et al. 2000). Understanding the evolutionary history of the MEP-pathway across a wide range of species offers a novel insight into their contribution to the evolution of primary plastids.

Here, extensive and integrated phylogenetic analyses identify alpha-proteobacteria, cyanobacteria and Chlamydiae as the bacterial lineages that have contributed to the evolution of MEP-pathway genes in plastid-bearing eukaryotes. Syntenic analyses establish the predominant presence of MEP-pathway genes as single-copy resulted from the loss of duplicated copies post whole genome duplication (WGD) events in land plants. Inter-species syntenic block and substitution rate analyses reveal the evolutionarily conservation of the MEP-pathway genes. Moreover, genetics analyses establish essential but differential functions of the MEP-pathway enzymes in plant growth and development.

In summary, the finding uncovers the evolutionary history and characteristics of plastidial isoprenoid biosynthesis-pathway genes, and reinforces the uniqueness of the MEP-pathway for unmasking the origins and evolution of plastids.

Results

MEP-pathway genes in plastid-bearing eukaryotes are derived from different bacteria lineages

To gain insight into the evolution of the MEP-pathway genes, we constructed phylogenetic trees for individual genes by using protein sequences of a wide range of species from eukaryotes, cyanobacteria, PVC (Planctomycetes, Verrucomicrobia and Chlamydiae) group eubacteria, and other non-cyanobacteria and non-PVC group bacteria (hereafter named them as 'other-eubacteria'). These analyses reveal the multiple origins of MEP-pathway genes in plastid-bearing eukaryotes (Figs. 2A-F, 3 and S1-S7).

The phylogenetic tree analyses show *DXS* and *MDS* in plastid-bearing eukaryotes and other-eubacteria are sister groups. It is of note that while *DXS* in plastid-bearing eukaryotes is clearly derived from alpha-proteobacteria (Figs. 2A and S1), the specific inheritance source for *MDS* remains unclear (Figs. 2B and S2). The maximum likelihood tree of *MDS* moderately supports *Aquifex Aeolicus* and *Leptospira interrogans* as the closest relatives of plastid-bearing eukaryotes (Figure S2a), whereas, the Bayesian tree clusters *Deinococcus-Thermus* bacteria (*Thermus thermophiles*, *Deinococcus radiodurans*) with plastid-bearing eukaryotes (Fig. S2B).

The phylogenetic trees of *DXR* and *HDR* group eukaryotes sequences with cyanobacteria (Figs. 2C-D, and S3-S4), and eukaryotic CMS and CMK in cluster with Chlamydiae (Figs. 2E-F and S5-S6).

Interestingly, the phylogenetic tree of *HDS* separates the plastid-bearing eukaryotes into two clades; one clade clusters with Chlamydiae and the other is closest to the cyanobacteria homologue (Figs. 3A and S7). Moreover, protein structure analyses show that depending on the organism, HDS enzymes have two different types of gcpE domains. Eubacteria HDS contain type I gcpE domain comprised of two N- and C-terminal parts, whereas the type II domain present in plants contains an additional domain between the N- and C-terminal parts of the protein, thought to enable the enzyme to function as a monomer (Fig. 3B) (Liu, et al. 2012). Domain analyses identify red algae HDS as an eubacteria-like type I-enzyme rather than the expected type II-enzyme in eukaryotes, while Chlamydiae HDS possesses the type II domain instead of the expected eubacterial type I domain (Fig. 3B).

Collectively the results display contributions of different bacterial lineages to the origins of MEP-pathway genes in plastid-bearing eukaryotes.

Duplicated *DXS* copies are not functionally redundant

Among the seven MEP-pathway enzymes, *DXS* catalyzes the first step in isoprenoid biosynthesis (Kuzuyama, et al. 2000) (Fig. 1B). Phylogenetic analyses of *DXS* show the presence of one gene copy in

examined algae and eubacteria, and its expansion into three subfamilies (I to III) in land plants (Figs. 2A and S1).

In the subfamily I, gene duplications in each common ancestor of Brassicaceae (Cruciferae, e.g. cabbage and turnip) and Fabaceae (legume, e.g. soybean) resulted in the presence of two genes (*DXS1* and *DXS2*). In the subfamily II, there is only one *DXS* copy, designated *DXS3* in *Arabidopsis thaliana*, Brassicaceae and the ancestor of Fabaceae. Moreover, the subfamily II is absent in gymnosperms, but duplicated copies are present in several moss and lycophyte species. Strikingly, the subfamily III branch is lost in Brassicaceae family, whereas Fabaceae and grape display species-specific duplication(s), and gymnosperms maintain two copies of subfamily III in their ancestor.

Interestingly, despite of the presence of three *DXS* subfamilies in land plants, only one is reported to function as a housekeeping MEP-pathway gene, such as *DXS1* in *A. thaliana* that encodes the functional MEP-pathway enzyme (Carretero-Paulet, et al. 2013). This is supported by plastidial localization of *DXS1* in Fabaceae species *Medicago* and soybean, in line with the function of the enzyme catalyzing the first step of the MEP-pathway (Walter, et al. 2002; Zhang, et al. 2009). The *DXS2*, which has no *DXS* activity, is assumed to synthesize specific isoprenoids related to mycorrhiza in *Medicago* (Floß, et al. 2008). *DXS3*, as the most divergent member of the family, has the expected *DXS* enzyme activity, but is expressed at very low levels, provoking the idea of its involvement in the synthesis of phytohormones in maize (Cordoba, et al. 2011).

In summary, despite the preservation of duplicated *DXS* copies in land plants, only one copy has retained the ancestral function of catalyzing the first step of the MEP-pathway, alluding to a possible loss or neo-functionalization of additional copies.

MEP-pathway genes are predominantly single copies

WGD events are most prevalent during angiosperm's diversity and are found in the common ancestors of seed plants (Ren, et al. 2018). However in spite of gene duplication events there are ~3124 nuclear-encoded single-copy genes, comprising ~11% of *Arabidopsis* genome, shared by other angiosperms (De Smet, et al. 2013). Excluding *DXS*, and with exception of few species that experienced a recent WGD, such as soybean and *Brassica oleracea* (Figs. S2-S7), the remainder six MEP-Pathway genes are among the single-copy genes in all algae and most land plants. Even in exceptional cases of soybean and *Brassica oleracea*, the MEP-pathway gene such as *CMS* remained as single copy (Fig. S5).

Despite multiple WGD events, the predominant presence of MEP-pathway genes as single-copy in most land plants leads to the question of when the duplicated copies were lost. To address this, we constructed intra-species conserved syntenic blocks of MEP-pathway genes for *A. thaliana* and *Oryza sativa*, the model eudicot and monocot species respectively. Both species are current diploids even though their most recent common ancestors experienced two WGDs. Except for the *DXS* in *A. thaliana*, the MEP-pathway genes in both species are single-copy, separately positioned in a syntenic block surrounded by

pairs of paralogs retained after WGD(s) (Fig. 4 and Data S1). This data supports the notion of loss of MEP-pathway genes post WGD.

To gain insight into the fate of the 'lost' copy of MEP-pathway genes, we searched for remnants of duplication events, but found no evidence such as the presence of a pseudogene for any of the MEP-pathway genes in *A. thaliana* genome.

MEP-pathway genes are evolutionarily conserved

To investigate the evolutionary characteristics of MEP-pathway genes, we examined the evolutionary rate, and domain architectures of the encoded proteins.

The evolutionary divergence of DNA can be estimated by the ratio of substitution rates at non-synonymous (dN; amino acid altering) and synonymous (dS; silent) sites, a measure of the dynamics of molecular evolution (Yang and Nielsen 1998). That is, a significantly low ratio of dN/dS marks slow evolution and as such the conserved nature of the protein. To investigate the MEP-pathway genes' evolutionary rates we examined their respective dN/dS ratios in selected species from represented lineages of plastid-bearing eukaryotes. The markedly low dN/dS median values ranging 0.04-0.14 suggests a strong purification selection for all the seven MEP-pathway genes, thereby supporting their evolutionary conservation (Fig. 5A and Table S1).

Moreover the analyses of the protein domain(s) structure of MEP-pathway enzymes establishes that, with the exception of HDS, an enzyme with two different types of gcpE domains (Fig. 3B), the protein structures of the remainder of MEP-pathway enzymes are universally conserved (Jin, et al. 2009).

MEP-pathway genes are highly expressed

There are two theories regarding gene conservation as the result of evolutionary rate of proteins; i) an inverse relationship between the expression levels and the evolutionary rate (Zhang and Yang 2015); and ii) a slow evolution of functionally critical genes as opposed to less critical ones (Jordan, et al. 2002). To test the potential contribution of these two scenarios to the high conservation of the MEP-pathway genes, we obtained and ranked expression levels of all MEP-pathway genes by analyzing the publicly available genome-wide transcriptomic datasets of representative land species, such as eudicots (*A. thaliana* and soybean), monocots (*O. sativa* and *Zea mays*), gymnosperm (*Picea abies*), moss (*P. patens*) and lycophyte (*S. moellendorffii*). The data illustrate high expression levels for most MEP-pathway genes with the exception of the three duplicated copies of *DXS* and two duplicates of *HDS* in *P. patens*, and one duplicated copy of *CMK* in soybean (Fig. 5B and Table S2). Notably, in most species, expression-ranking data places the first two genes (*DXS* and *DXR*) and the last three genes (*MDS*, *HDS* and *HDR*) amongst the top 5-10% most abundant transcripts.

To compensate for the absence of transcriptomic datasets for several lineages, we recruited a widely used quantitative method, Codon Adaptation Index (CAI), to predict the expression level of a gene based on its codon sequence. The rationale of CAI is based on codon degeneracy, and that the highly expressed

genes are biased towards the codon decoded by the most abundant tRNA species (Jansen, et al. 2003). We therefore calculated CAIs of all MEP-pathway genes from represented species with and without transcriptomic datasets in all life lineages. In most analyzed species, the MEP-pathway genes have a CAI value higher than 0.7 (Table S3). The median CAI values for the MEP-pathway genes (0.76-0.80), denote their high expression levels in all life lineages analyzed (Fig. 5C).

In summary, the high expression levels of the MEP-pathway genes support their evolutionary conservation.

MEP-pathway genes are indispensable for plant growth

Except green algae, plants possess both the cytosolic MVA- and the plastidial MEP-pathways, despite the established exchanges of the end products between the two isoprenoid producing routes (Laule, et al. 2003). Given the indispensable function of MEP-pathway genes in eubacteria (Goodall, et al. 2018; Rubin, et al. 2015), we employed genetic approaches to test the likelihood of the essentiality of the MEP-pathway genes in plants.

Unavailability of T-DNA insertion lines for the MEP-pathway genes, led us to employ the previously generated RNAi lines that were maintained as segregating population for individual MEP-pathway genes in *A. thaliana* (Xiao, et al. 2012). Homozygous RNAi lines, each with 92-95% reduced expression levels of the corresponding MEP-pathway genes (Xiao, et al. 2012), displayed seedling size and variegation leaf phenotypes distinct from each other and from those of the wild type plants transformed with an empty vector (EV). These visibly altered phenotypes include dwarfed stature of *asDXR*, *asMDS* and *asHDS* lines; in concert with pale-yellow leaves phenotype of the *asMDS* seedlings, and an albino phenotype of true leaves in all the other six RNAi lines (Fig. 6).

In summary, the phenotypes of RNAi lines confirm the indispensable function of MEP-pathway enzymes in plant growth and development, and that the markedly different size and phenotypic characteristics of each line suggest the involvement of these enzymes in distinct functions in addition of their role as intermediates in isoprenoids biosynthesis pathway.

Discussion

The MEP-pathway is comprised of seven nuclear-encoded plastid-localized enzymes, essential for plant growth and key to stress-specific retrograde signaling as evidenced by the function of the MEP-pathway intermediate, methylerythritol cyclodiphosphate (MEcPP) as a retrograde signaling metabolite (Xiao, et al. 2012). The retrograde signaling function of MEcPP offers an exciting justification regarding the necessity of the MEP-pathway existence, not only for the production of the isoprenoids but also for retrograde signaling function of each of intermediates essential for coordinated action of the two organelles. This possibility could also explain the coexistence of MVA- and MEP-, the two isoprenoid producing pathways in plants.

MEP-pathway genes are resistant to duplication

In land plants, all the MEP-pathway genes with the exception of *DXS*, are present as single-copy in all the analyzed diploid plants in spite of ancient WGD events. In fact, although *DXS* experienced duplications, only one copy maintained the MEP-pathway-based enzyme activity (Carretero-Paulet, et al. 2013; Walter, et al. 2002; Zhang, et al. 2009). The critical nature of gene duplication as a source of evolutionary innovation and adaptation (Rensing 2014), raises the question of why the MEP-pathway genes have remained single-copies. One explanation might be that under the relaxation of selective pressure, the duplicated copy is inclined to accumulate deleterious mutations (Zhang 2003), which in turn could result in a dominant negative inhibition of the other functional copy. Indeed this is in stark contrast with the existence of multiple copies of the cytosolic MVA-pathway genes, such as functionally redundant *AACT1* and *AACT2* (anthocyanin-5-aromatic acyl transferase-like) both of which encode the initial enzyme of the MVA-pathway (Jin, et al. 2012), or *HMG1* and *HMG2* encoding the HMGR (3-hydroxy-3-methylglutaryl CoA reductase) (Suzuki, et al. 2004), and *MVD1* and *MVD2* encoding the MVD (mevalonate diphosphate decarboxylase) (Pulido, et al. 2012). Plants lacking *AACT1* or *HMG2* are viable with no apparent phenotypes, in contrast to indispensability of MEP-pathway genes.

The polyphyletic origin of MEP-pathway genes

Among seven MEP-pathway genes, *DXS* and *MDS* have originated from 'other-eubacteria'. The closest sister clade of plastid-bearing eukaryotes *DXS* is alpha-proteobacteria, also the known ancestor of mitochondrion (Martijn, et al. 2018). This suggests that plastid-bearing eukaryotes *DXS* might have originated directly from alpha-proteobacteria via horizontal gene transfer (HGT), or indirectly via endosymbiotic gene transfer (EGT) from the mitochondrion genome.

We were unable to place the origin of eukaryotic *MDS*, but through expanded phylogenetic analyses we determined notable homology between Chlamydiae and eukaryotes sequences of three (*CMS*, *CMK*, and *HDS*) of the seven MEP-pathway genes. Specifically, the phylogenetic trees of *CMS* and *CMK*, show that eukaryotes lineage form a sister cluster with the corresponding Chlamydiae gene, suggestive of HGTs between Chlamydiae and the common ancestor of eukaryotes. In addition, phylogenetic analyses of *HDS* depict clustering of red algae with cyanobacteria as opposed to other plastid-bearing eukaryotes that form a sister group with Chlamydiae. One potential explanation for this bifurcated clustering is that the ancestral plastid-bearing eukaryotes acquired *HDS* from Chlamydiae, but in red algae ancestor, the chlamydial *HDS* was lost as the result of two major phases of genome reduction (Qiu, et al. 2015), but later it was replaced by the second HGT event from cyanobacteria.

The necessity of Chlamydiae like *HDS* enzyme in plastid bearing organisms potentially could be justified as a response to the changing environmental conditions over time. Based on the oldest eukaryotic algae fossils findings in conjunction with the molecular clock data, plastids are predicted to have originated in Mesoproterozoic era ~0.9-1.7 billion years ago (McFadden 2014). During the Proterozoic era, oxygen began to rise and built up to above 10% of the levels existed in the atmosphere at Mesoproterozoic era (Harada, et al. 2015; Zhang, et al. 2018). Simultaneously, the earth entered a warm period ending

glaciations and raising the tropical mean sea surface temperatures from ~19.4-28.7°C (Fiorella and Sheldon 2017).

It is well established that HDS, a [4Fe-4S]-protein reactive to oxygen species, is hypersensitive to high radiation and supra-optimal temperatures. Under these unfavorable conditions inhibition of HDS results in accumulation of its substrate, MEcPP, that in turn protects MEP-pathway activity by restricting oxidative stress (Ostrovsky, et al. 1998; Ostrovsky, et al. 2003; Rivasseau, et al. 2009). Plastid-bearing eukaryotes are frequently and simultaneously exposed to reactive oxygen species, high light irradiance and hot temperatures, and one could consider HDS enzyme as the gatekeeper maintaining the MEP-pathway's functionality.

Accordingly, we propose that the evolutionary pressures resulted from high oxygen and higher temperatures at the era of plastid establishment may have led to acquisition of a monomeric Chlamydiae like HDS in plastid-bearing organisms. The presence of a middle domain in the monomeric enzyme would have provided a higher ratio of protein/ labile [Fe-S]-iron cluster, thereby a functionally more efficient enzyme than the dimeric form in a high oxygen and high temperature environment. As such, plastid-bearing eukaryotes, acquired the more efficient monomeric HDS donated by Chlamydiae.

Our overall finding poses the question of how multiple donors could have contributed to the MEP. The simplified schematic (Fig. 7) depicts the three potential scenarios addressing the question. Scenario I proposes inherited chimerism by EGT, and that the occurrence of prokaryotic HGT to the cyanobacteria genome happened prior to endosymbiosis event leading to plastid formation (Ku, et al. 2015a). If so, cyanobacteria must have acquired chlamydial MEP-pathway genes through HGT before EGT in plants (Fig. 7). In such a case, one would expect the presence of chlamydial type *CMS*, *CMK* and *HDS* sequences in *Gloeomargarita lithophora* genome, the prime candidate for extant relative of the cyanobacterial plastid progenitor (Ponce-Toledo, et al. 2017). However, clustering of the three genes in *G. lithophora* with cyanobacteria and not with Chlamydiae (Figures S3-S4, and S6), diminishes the probability of scenario I.

Scenario II suggests that *CMS*, *CMK* and *HDS* in eukaryotes are the result of HGT from Chlamydiae after the endosymbiosis (Fig. 7). But, the inability of Chlamydiae to infect current photosynthetic eukaryotes or plastid-containing organisms renders this scenario less plausible.

Scenario III supports co-contribution of cyanobacteria and Chlamydiae to the origin of the primary plastid (Fig. 7), once proposed as the `tripartite (*ménage-à-trois*- 'household of three') symbiotic relationship between the extant order Chlamydiales, a cyanobacterium, and an eukaryotic host for the establishment of the eukaryotes lineages (Ball, et al. 2016; Ball, et al. 2013; Cenci, et al. 2016; Huang and Gogarten 2007). The tripartite endosymbiosis supported by phylogenomic analyses of a considerable number of nuclear genes in eukaryotes related to chlamydial homologues, proposes that the chlamydial partner injected effector proteins into the ancestral eukaryotes as a strategy to manipulate host cell carbohydrate metabolism to the parasite's advantage (Ball, et al. 2013; Cenci, et al. 2017; Huang and Gogarten 2007). However, counter arguments question the correct evolutionary models of phylogenomic analyses, the high frequency of HGTs among prokaryotes and among prokaryote-to-eukaryote (Domman, et al. 2015;

Gould 2016; Ku, et al. 2015b; Soucy, et al. 2015). Our analyses based on the best-fitting evolutionary models for constructing phylogenetic trees of individual MEP-pathway genes support the chlamydial origination for three of seven MEP-pathway genes in plastid-bearing eukaryotes, even though the evolutionary pressure(s) that led to plastid-bearing eukaryotes harboring a chimera MEP-pathway remains an enigma.

Our data clearly presents contribution of both cyanobacteria and Chlamydiae to plastid-bearing eukaryotes MEP-pathway and by extension to the origin of the primary plastid.

Conclusion

The MEP-pathway genes are highly conserved and are essential for the survival of plastid-bearing eukaryotes. The plastid-bearing eukaryotes MEP-pathway genes originated from both cyanobacteria and Chlamydiae indicating their co-contributions to the evolution of primary plastids. The nuclear-encoded plastid-destined MEP-pathway enzymes enable the host eukaryotes to control plastids in a stable endosymbiosis system, while in return MEcPP, the plastid-produced intermediate of the MEP-pathway, coordinates expression of selected nuclear stress-response genes and the corresponding physiological ramifications. These bilateral controls mediated by MEP-pathway may also shed light on the basis of the co-existence of cytosolic and plastidial isopreneoid biosynthesis pathways in eukaryotes.

In summary, these findings uncover the evolutionary history and characteristics of the plastidial isoprenoid biosynthesis-pathway genes and its implications in origin and evolution of primary plastid.

Methods And Materials

Identification the homologues of MEP-pathway

Plant genome sequences were downloaded from the Phytozome v12 (<https://phytozome.jgi.doe.gov/pz/portal.html>), *Amborella* Genome Database (<http://amborella.huck.psu.edu/data>), Spruce Genome Project (<http://congenie.org/start>) and JGI Genome portal (<https://genome.jgi.doe.gov/>). Algae genomes were downloaded from Phytozome and Greenhouse (<https://greenhouse.lanl.gov/greenhouse/organisms/>). Annotated genome sequences of *Chara braunii* (Nishiyama, et al. 2018), *Klebsormidium nites* (Hori, et al. 2014) are downloaded. Genome sequences of selected eubacteria were downloaded from Ensembl (release 90) (<ftp://ftp.ensembl.org/pub/>).

The names and IDs of MEP-pathway genes in *A. thaliana* are *DXS* (1-deoxy-D-xylulose-5-phosphate synthase): *AT4G15560*; *DXR* (1-deoxy-D-xylulose 5-phosphate reductoisomerase): *AT5G62790*; *CMS* (4-Diphosphocytidyl-2C-methyl-D-erythritol synthase): *AT2G02500*; *CMK* (4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase activity): *AT2G26930*; *MDS* (2C-methyl-d-erythritol 2,4-cyclodiphosphate synthase): *AT1G63970*; *HDS* (4-hydroxy-3-methylbut-2-enyl diphosphate synthase): *AT5G60600*; and *HDR* (4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase): *AT4G34350*. The protein domain information

for each MEP-pathway gene in *A. thaliana* was obtained from Phytozome v12, which are 1-deoxy-D-xylulose-5-phosphate synthase as PF13292 for DXS, 1-deoxy-D-xylulose-5-phosphate reductoisomerase as PF02670 and 1-deoxy-D-xylulose-5-phosphate reductoisomerase C-terminal as PF08436 for DXR, MobA-like NTP transferase as PF12804 for CMS, GHMP kinases N terminal as PF00288 and GHMP kinases C terminal as PF08544 for CMK, YgbB as PF02542 for MDS, GcpE as PF04551 for HDS, and LytB protein as PF02401 for HDR, respectively. Hidden Markov Models (HMM) (Eddy 1996) matrix presenting each domain of MEP-pathway enzymes was downloaded from Pfam (<https://pfam.xfam.org/>). Then, hmmsearch and fastacmd were used to obtain protein sequences in selected whole-genome sequenced species. And protein sequences of MEP-pathway genes in PVC bacteria were retrieved from PVCbase (<http://pvcbacteria.org/pvcbase/>) using BLASTP. All the identified proteins were examined on Pfam website to confirm the presence of the corresponding protein domain.

Multiple sequence alignment and phylogenetic tree construction

The MEP-pathway protein sequences were aligned using MUSCLE (Edgar 2004) v3.8.31 with default parameters. Prottest (Darriba, et al. 2011) was used to select out the best-fitting evolutionary model for each aligned protein matrix of MEP-pathway gene. Then the evolutionary model of WAG + G was specified for DXS, LG + I +G was specified for DXR, HDS and HDR, and VT + I +G was specified for CMS, CMK and MDS. Phylogenetic trees were constructed by RAxML (Silvestro and Michalak 2012) v7.1.0 and MrBayes (Ronquist, et al. 2012) v3.2.7. As an exception to the MEP-pathway genes, the *CMK* belongs to the GHMP gene family with 13 copies in *A. thaliana*. All protein sequences of this family in each species were firstly retrieved and preliminary ML tree using aligned sequences was constructed. Lastly, members of CMK and MVK (mevalonate kinase in the MVA-pathway) were selected out for constructing the final phylogenetic tree. The MVK branch was set as outgroup.

Synteny analysis

The Locus search function in PGDD (Lee, et al. 2012) (<http://chibba.agtec.uga.edu/duplication/>), a public database for cluster identification of plant genes based on intra- or cross-genome syntenic relationships, was implemented for identifying the intra-species duplication blocks around 500 kb region of each MEP-pathway genes in *A. thaliana* and *O. sativa*.

dN/dS analyses

Nucleotide sequences for represented species in each lineage, namely *A. thaliana* a eudicot, *O. sativa* a monocot, *A. trichopoda* an early-diverging angiosperm, *P. abies* a gymnosperm, *P. patens* a moss, *S. moellendorffii* a lycophyte, *Volvox carteri* a green algae and *Cyanidioschyzon merolae* a red algae, were retrieved to calculate the nonsynonymous to synonymous rate ratio ($\omega = dN/dS$) between *A. thaliana* and all other species. The ω was calculated by yn00 contained in the software PAML v4.5 (Yang 2007) using the Yang & Nielsen method, where $0 < \omega < 1$ indicates purifying selection, $\omega = 1$ corresponds to neutral selection, and $\omega > 1$ implicates positive selection. The distributions of all dN/dS values for each MEP-pathway gene were drawn by the boxplot function in the R (Team 2013) program.

Expression levels, Codon Adaption index (CAI)

The sources of the RPKM values of all expressed genes in wild type plant in represented species were listed as following: *A. thaliana* (Bjornson, et al. 2017), *Z. mays* (Kebede, et al. 2018) (2 replicates of mock treated wild type), *G. max* (SoyBase Soybean Genome Annotation Page: https://soybase.org/soyseq/tables_lists/index.php), *O. sativa* (Rice Genome Annotation Project: <http://rice.plantbiology.msu.edu/expression.shtml>, use four libraries of four-leaf stage seedling), *P. abies* (Spruce Genome Project: ftp://plantgenie.org/Data/ConGenIE/Picea_abies/v1.0/Expression/), *P. patens* (Khraiwesh, et al. 2015) and *S. moellendorffii* (You, et al. 2017). Genes with RPKM ≥ 1 were retained for further analyses. The relative expression ranking of each gene was calculated using the formula: $1 - (\text{order of the gene}) / (\text{total number of all expressed genes})$. The relative expression ranking of all represented species for each gene were presented as scatter-boxplot in the R program.

Nucleotide sequences for each MEP-pathway gene in selected species were retrieved in corresponding datasets. Codon usage table for each selected species was obtained from Condon Usage Database (<http://www.kazusa.or.jp/codon/>). Lastly, fasta format of each nucleotide sequence and codon usage table of each species were inputted to calculate the CAI on the CAIcal SERVER (Puigbò, et al. 2008) (<http://ppuigbo.me/programs/CAIcal/>).

Plant material and growth conditions

We employed the RNAi lines for all MEP-pathway genes in *A. thaliana* that were previously generated (Xiao, et al. 2012). Sterilized seeds sowed on 1/2 MS medium were maintained for 48 hours at 4 °C. Two-week-old seedlings were grown at 22 °C under 16/8h light/dark period.

Declarations

Ethics Approval and Consent to Participate

Not applicable

Consent for Publication

Not applicable

Availability of Data and Materials

Not applicable

Competing Interest

The authors declare that they are no competing interests.

Funding

This work was supported by the U.S. National Science Foundation (NSF) grant IOS-1352478 to KD, and the U.S. National Institutes of Health (NIH) grant R01GM107311 to KD. The funders had no role in design of the study, or in the collection, analysis, and interpretation of data or in the writing of the manuscript.

Author Contributions

L. Z. and K. D. designed this study. L. Z. performed bioinformatics analyses. L. Z. drafted the manuscript and K. D. revised the manuscript. All authors have read and approved the manuscript.

Corresponding Author

Correspondence to Katayoon Dehesh with kdehesh@ucr.edu

Acknowledgements

Not applicable

References

- Archibald JM. 2015. Genomic perspectives on the birth and spread of plastids. *Proceedings of the National Academy of Sciences* 112(33):10147-10153.
- Ball SG, Bhattacharya D, Weber AP. 2016. Infection and the first eukaryotes—Response. *Science* 352(6289):1065-1066.
- Ball SG, Subtil A, Bhattacharya D, Moustafa A, Weber AP, Gehre L, Colleoni C, Arias M-C, Cenci U, Dauvillée D. 2013. Metabolic effectors secreted by bacterial pathogens: essential facilitators of plastid endosymbiosis? *The Plant Cell* 25(1):7-21.
- Bick JA, Lange BM. 2003. Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane. *Archives of biochemistry and biophysics* 415(2):146-154.
- Bjornson M, Balcke GU, Xiao Y, Souza A, Wang JZ, Zhabinskaya D, Tagkopoulos I, Tissier A, Dehesh K. 2017. Integrated omics analyses of retrograde signaling mutant delineate interrelated stress-response strata. *The Plant Journal* 91(1):70-84.
- Brocks JJ, Logan GA, Buick R, Summons RE. 1999. Archean molecular fossils and the early rise of eukaryotes. *Science* 285(5430):1033-6.
- Carretero-Paulet L, Cairo A, Talavera D, Saura A, Imperial S, Rodriguez-Concepcion M, Campos N, Boronat A. 2013. Functional and evolutionary analysis of DXL1, a non-essential gene encoding a 1-deoxy-D-xylulose 5-phosphate synthase like protein in *Arabidopsis thaliana*. *Gene* 524(1):40-53.

- Cenci U, Bhattacharya D, Weber AP, Colleoni C, Subtil A, Ball SG. 2017. Biotic host–pathogen interactions as major drivers of plastid endosymbiosis. *Trends in plant science* 22(4):316-328.
- Cenci U, Ducatez M, Kadouche D, Colleoni C, Ball SG. 2016. Was the Chlamydial Adaptive Strategy to Tryptophan Starvation an Early Determinant of Plastid Endosymbiosis? *Frontiers in Cellular and Infection Microbiology* 6.
- Cordoba E, Porta H, Arroyo A, San Román C, Medina L, Rodríguez-Concepción M, León P. 2011. Functional characterization of the three genes encoding 1-deoxy-D-xylulose 5-phosphate synthase in maize. *Journal of experimental botany* 62(6):2023-2038.
- Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27(8):1164-1165.
- De Smet R, Adams KL, Vandepoele K, Van Montagu MC, Maere S, Van de Peer Y. 2013. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proceedings of the National Academy of Sciences* 110(8):2898-2903.
- Domman D, Horn M, Embley TM, Williams TA. 2015. Plastid establishment did not require a chlamydial partner. *Nature communications* 6:6421.
- Eddy SR. 1996. Hidden markov models. *Current opinion in structural biology* 6(3):361-365.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* 32(5):1792-1797.
- Fiorella RP, Sheldon ND. 2017. Equable end Mesoproterozoic climate in the absence of high CO₂. *Geology* 45(3):231-234.
- Floß DS, Hause B, Lange PR, Küster H, Strack D, Walter MH. 2008. Knock-down of the MEP pathway isogene 1-deoxy-d-xylulose 5-phosphate synthase 2 inhibits formation of arbuscular mycorrhiza-induced apocarotenoids, and abolishes normal expression of mycorrhiza-specific plant marker genes. *The Plant Journal* 56(1):86-100.
- Goodall EC, Robinson A, Johnston IG, Jabbari S, Turner KA, Cunningham AF, Lund PA, Cole JA, Henderson IR. 2018. The essential genome of *Escherichia coli* K-12. *MBio* 9(1):e02096-17.
- Gould SB. 2016. Infection and the first eukaryotes. *Science* 352(6289):1065-1065.
- Harada M, Tajika E, Sekine Y. 2015. Transition to an oxygen-rich atmosphere with an extensive overshoot triggered by the Paleoproterozoic snowball Earth. *Earth and Planetary Science Letters* 419:178-186.
- Hemmerlin A, Hoeffler J-F, Meyer O, Tritsch D, Kagan IA, Grosdemange-Billiard C, Rohmer M, Bach TJ. 2003. Cross-talk between the cytosolic mevalonate and the plastidial methylerythritol phosphate

pathways in tobacco bright yellow-2 cells. *Journal of Biological Chemistry* 278(29):26666-26676.

Hori K, Maruyama F, Fujisawa T, Togashi T, Yamamoto N, Seo M, Sato S, Yamada T, Mori H, Tajima N. 2014. *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nature communications* 5:3978.

Hoshino Y, Gaucher EA. 2018. On the origin of isoprenoid biosynthesis. *Molecular biology and evolution* 35(9):2185-2197.

Huang J, Gogarten JP. 2007. Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? *Genome biology* 8(6):R99.

Jansen R, Bussemaker HJ, Gerstein M. 2003. Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models. *Nucleic Acids Research* 31(8):2242-2251.

Jin H, Song Z, Nikolau BJ. 2012. Reverse genetic characterization of two paralogous acetoacetyl CoA thiolase genes in *Arabidopsis* reveals their importance in plant growth and development. *The Plant Journal* 70(6):1015-1032.

Jin J, Xie XY, Chen C, Park JG, Stark C, James DA, Olhovsky M, Linding R, Mao YY, Pawson T. 2009. Eukaryotic Protein Domains as Functional Units of Cellular Evolution. *Science Signaling* 2(98).

Jordan IK, Rogozin IB, Wolf YI, Koonin EV. 2002. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome research* 12(6):962-968.

Kebede AZ, Johnston A, Schneiderman D, Bosnich W, Harris LJ. 2018. Transcriptome profiling of two maize inbreds with distinct responses to *Gibberella* ear rot disease to identify candidate resistance genes. *BMC genomics* 19(1):131.

Keeling PJ. 2010. The endosymbiotic origin, diversification and fate of plastids. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365(1541):729-748.

Khraiwesh B, Qudeimat E, Thimma M, Chaiboonchoe A, Jijakli K, Alzahmi A, Arnoux M, Salehi-Ashtiani K. 2015. Genome-wide expression analysis offers new insights into the origin and evolution of *Physcomitrella patens* stress response. *Scientific reports* 5:17434.

Ku C, Nelson-Sathi S, Roettger M, Garg S, Hazkani-Covo E, Martin WF. 2015a. Endosymbiotic gene transfer from prokaryotic pangenomes: Inherited chimerism in eukaryotes. *Proceedings of the National Academy of Sciences* 112(33):10139-10146.

Ku C, Nelson-Sathi S, Roettger M, Sousa FL, Lockhart PJ, Bryant D, Hazkani-Covo E, McInerney JO, Landan G, Martin WF. 2015b. Endosymbiotic origin and differential loss of eukaryotic genes. *Nature* 524(7566):427.

- Kuzuyama T, Takagi M, Takahashi S, Seto H. 2000. Cloning and characterization of 1-deoxy-D-xylulose 5-phosphate synthase from *Streptomyces* sp. Strain CL190, which uses both the mevalonate and nonmevalonate pathways for isopentenyl diphosphate biosynthesis. *Journal of Bacteriology* 182(4):891-7.
- Lange BM, Rujan T, Martin W, Croteau R. 2000. Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proceedings of the National Academy of Sciences* 97(24):13172-13177.
- Laule O, Fürholz A, Chang H-S, Zhu T, Wang X, Heifetz PB, Grussem W, Lange M. 2003. Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* 100(11):6866-6871.
- Lee T-H, Tang H, Wang X, Paterson AH. 2012. PGDD: a database of gene and genome duplication in plants. *Nucleic acids research* 41(D1):D1152-D1158.
- Liu YL, Guerra F, Wang K, Wang WX, Li JK, Huang C, Zhu W, Houlihan K, Li Z, Zhang Y and others. 2012. Structure, function and inhibition of the two- and three-domain 4Fe-4S IspG proteins (vol 109, pg 8558, 2012). *Proceedings of the National Academy of Sciences of the United States of America* 109(26):10605-10605.
- Lombard J, Moreira D. 2010. Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Molecular biology and evolution* 28(1):87-99.
- Martijn J, Vosseberg J, Guy L, Offre P, Ettema TJ. 2018. Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature* 557(7703):101-105.
- McFadden GI. 2014. Origin and evolution of plastids and photosynthesis in eukaryotes. *Cold Spring Harbor perspectives in biology* 6(4):a016105.
- Nishiyama T, Sakayama H, de Vries J, Buschmann H, Saint-Marcoux D, Ullrich KK, Haas FB, Vanderstraeten L, Becker D, Lang D. 2018. The Chara genome: secondary complexity and implications for plant terrestrialization. *Cell* 174(2):448-464. e24.
- Ostrovsky D, Diomina G, Lysak E, Matveeva E, Ogel O, Trutko S. 1998. Effect of oxidative stress on the biosynthesis of 2-C-methyl-D-erythritol-2, 4-cyclopyrophosphate and isoprenoids by several bacterial strains. *Archives of microbiology* 171(1):69-72.
- Ostrovsky DN, Dyomina GR, Deryabina YI, Goncharenko AV, Eberl M, Shumaev KB, Shashkov AS. 2003. Properties of 2-C-methyl-D-erythritol 2,4-cyclopyrophosphate, an intermediate in nonmevalonate isoprenoid biosynthesis. *Applied Biochemistry and Microbiology* 39(5):497-502.
- Ponce-Toledo RI, Deschamps P, López-García P, Zivanovic Y, Benzerara K, Moreira D. 2017. An early-branching freshwater cyanobacterium at the origin of plastids. *Current Biology* 27(3):386-391.

- Puigbò P, Bravo IG, Garcia-Vallve S. 2008. CAIcal: a combined set of tools to assess codon usage adaptation. *Biology direct* 3(1):38.
- Pulido P, Perello C, Rodriguez-Concepcion M. 2012. New insights into plant isoprenoid metabolism. *Molecular plant* 5(5):964-967.
- Qiu H, Price DC, Yang EC, Yoon HS, Bhattacharya D. 2015. Evidence of ancient genome reduction in red algae (Rhodophyta). *Journal of phycology* 51(4):624-636.
- Ren R, Wang H, Guo C, Zhang N, Zeng L, Chen Y, Ma H, Qi J. 2018. Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Molecular plant* 11(3):414-428.
- Rensing SA. 2014. Gene duplication as a driver of plant morphogenetic evolution. *Current Opinion in Plant Biology* 17:43-48.
- Rivasseau C, Seemann M, BOISSON A, Streb P, Gout E, Douce R, Rohmer M, Bligny R. 2009. Accumulation of 2-C-methyl-D-erythritol 2, 4-cyclopyrophosphate in illuminated plant leaves at supraoptimal temperatures reveals a bottleneck of the prokaryotic methylerythritol 4-phosphate pathway of isoprenoid biosynthesis. *Plant, cell & environment* 32(1):82-92.
- Rohmer M, Knani M, Simonin P, Sutter B, Sahm H. 1993. Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. *Biochemical Journal* 295(Pt 2):517.
- Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology* 61(3):539-542.
- Rubin BE, Wetmore KM, Price MN, Diamond S, Shultzaberger RK, Lowe LC, Curtin G, Arkin AP, Deutschbauer A, Golden SS. 2015. The essential gene set of a photosynthetic organism. *Proceedings of the National Academy of Sciences* 112(48):E6634-E6643.
- Ruiz-Sola MÁ, Barja MV, Manzano D, Llorente B, Schipper B, Beekwilder J, Rodriguez-Concepcion M. 2016. A single *Arabidopsis* gene encodes two differentially targeted geranylgeranyl diphosphate synthase isoforms. *Plant physiology* 172(3):1393-1402.
- Sacchettini JC, Poulter CD. 1997. Creating isoprenoid diversity. *Science* 277(5333):1788-1789.
- Silvestro D, Michalak I. 2012. raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity & Evolution* 12(4):335-337.
- Soucy SM, Huang J, Gogarten JP. 2015. Horizontal gene transfer: building the web of life. *Nature Reviews Genetics* 16(8):472.

- Suzuki M, Kamide Y, Nagata N, Seki H, Ohyama K, Kato H, Masuda K, Sato S, Kato T, Tabata S. 2004. Loss of function of 3-hydroxy-3-methylglutaryl coenzyme A reductase 1 (HMG1) in Arabidopsis leads to dwarfing, early senescence and male sterility, and reduced sterol levels. *The Plant Journal* 37(5):750-761.
- Team RC. 2013. R: A language and environment for statistical computing.
- Walley J, Xiao Y, Wang JZ, Baidoo EE, Keasling JD, Shen Z, Briggs SP, Dehesh K. 2015. Plastid-produced interorganelle stress signal MEcPP potentiates induction of the unfolded protein response in endoplasmic reticulum. *Proc Natl Acad Sci USA* 112(19):6212-7.
- Walter MH, Hans J, Strack D. 2002. Two distantly related genes encoding 1-deoxy-d-xylulose 5-phosphate synthases: differential regulation in shoots and apocarotenoid-accumulating mycorrhizal roots. *The Plant Journal* 31(3):243-54.
- Woodson JD, Chory J. 2008. Coordination of gene expression between organellar and nuclear genomes. *Nat Rev Genet* 9(5):383-95.
- Wright LD. 1961. Biosynthesis of isoprenoid compounds. *Annual Review of Biochemistry* 30(1):525-548.
- Xiao YM, Savchenko T, Baidoo EEK, Chehab WE, Hayden DM, Tolstikov V, Corwin JA, Kliebenstein DJ, Keasling JD, Dehesh K. 2012. Retrograde Signaling by the Plastidial Metabolite MEcPP Regulates Expression of Nuclear Stress-Response Genes. *Cell* 149(7):1525-1535.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* 24(8):1586-1591.
- Yang Z, Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *Journal of Molecular Evolution* 46(4):409-18.
- You C, Cui J, Wang H, Qi X, Kuo L-Y, Ma H, Gao L, Mo B, Chen X. 2017. Conservation and divergence of small RNA pathways and microRNAs in land plants. *Genome biology* 18(1):158.
- Zhang J. 2003. Evolution by gene duplication: an update. *Trends in ecology & evolution* 18(6):292-298.
- Zhang J, Yang J-R. 2015. Determinants of the rate of protein sequence evolution. *Nature Reviews Genetics* 16(7):409.
- Zhang K, Zhu X, Wood R, Shi Y, Gao Z, Poulton S. 2018. Oxygenation of the Mesoproterozoic ocean and the evolution of complex eukaryotes. *Nature Geoscience* 11:345-350.
- Zhang M, Li K, Zhang C, Gai J, Yu D. 2009. Identification and characterization of class 1 DXS gene encoding 1-deoxy-D-xylulose-5-phosphate synthase, the first committed enzyme of the MEP pathway from soybean. *Molecular Biology Reports* 36(5):879-87.

Supporting Information

Figure S1 Phylogenetic tree of DXS.

Figure S2 Phylogenetic tree of MDS.

Figure S3 Phylogenetic tree of DXR.

Figure S4 Phylogenetic tree of HDR.

Figure S5 Phylogenetic tree of CMS.

Figure S6 Phylogenetic tree of CMK.

Figure S7 Phylogenetic tree of HDS.

Table S1. Synonymous (dN) and nonsynonymous (dS) substitution rates estimated by PAML.

Table S2. The relative expression ratio of the MEP-pathway genes in represented species.

Table S3. The Codon Adaption Index of the MEP-pathway genes.

Data S1. Paralogous syntenic blocks display loss of duplicated copies of MEP-pathway genes

Figures

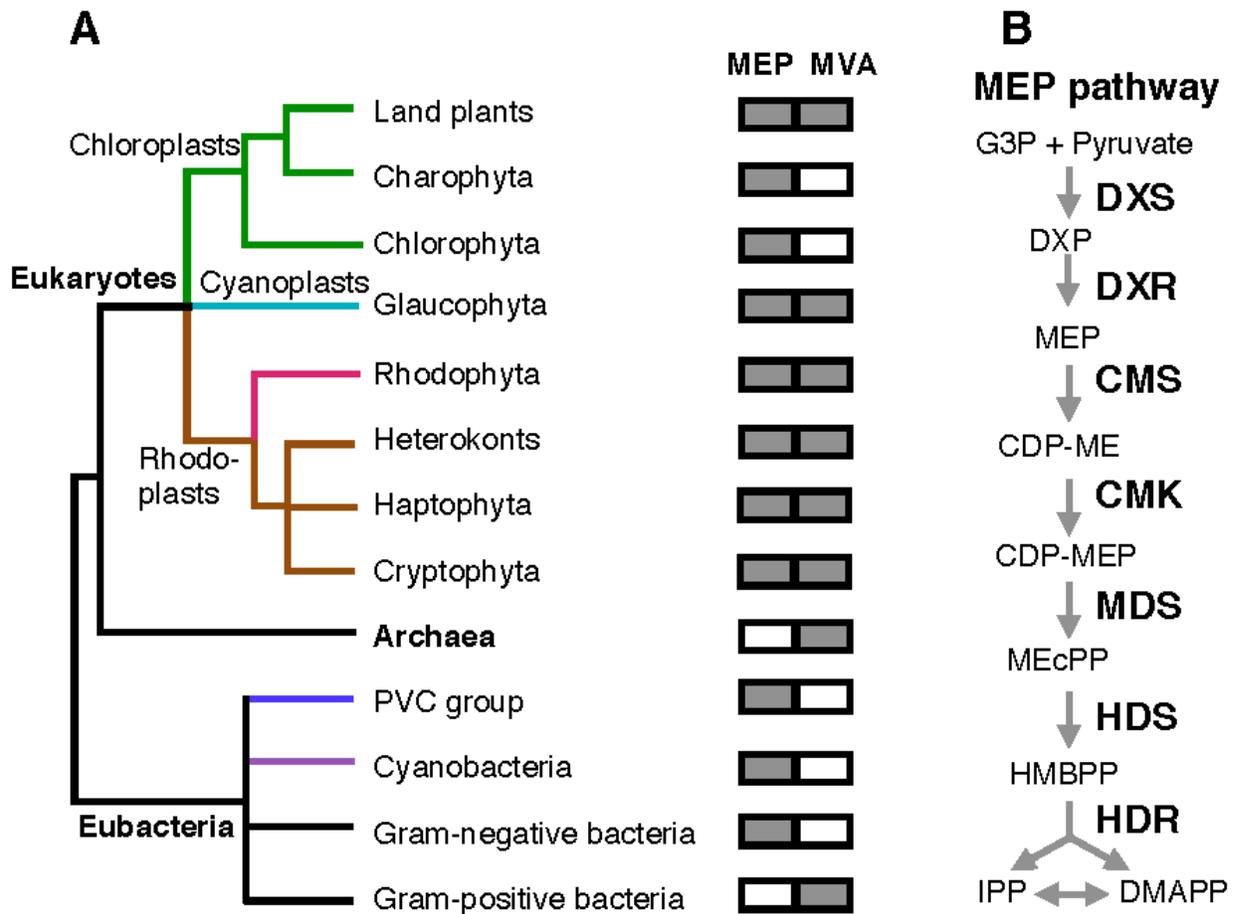


Figure 1

Distribution of isoprenoid biosynthesis-pathways across lineages. (a) Distribution of mevalonate (MVA) and/or methylerythritol dicyclopentyl phosphate (MEP) pathways, the two isoprenoid biosynthesis-pathways across lineages of eukaryotes, Archaea, and eubacteria. The rectangular boxes display the presence (filled) or absence of (blank) MVA- or MEP-pathway in each of the lineages. (b) Schematic presentation of the seven enzymes of the MEP-pathway producing the two universal five-carbon precursors, isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP). DXS: 1-deoxy-D-xylulose-5-phosphate synthase. DXR: 1-deoxy-D-xylulose 5-phosphate reductoisomerase. CMS: 4-Diphosphocytidyl-2C-methyl-D-erythritol synthase. CMK: 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase activity. MDS: 2C-methyl-d-erythritol 2, 4-cyclodiphosphate synthase. HDS: 4-hydroxy-3-methylbut-2-enyl diphosphate synthase. HDR: and 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase.

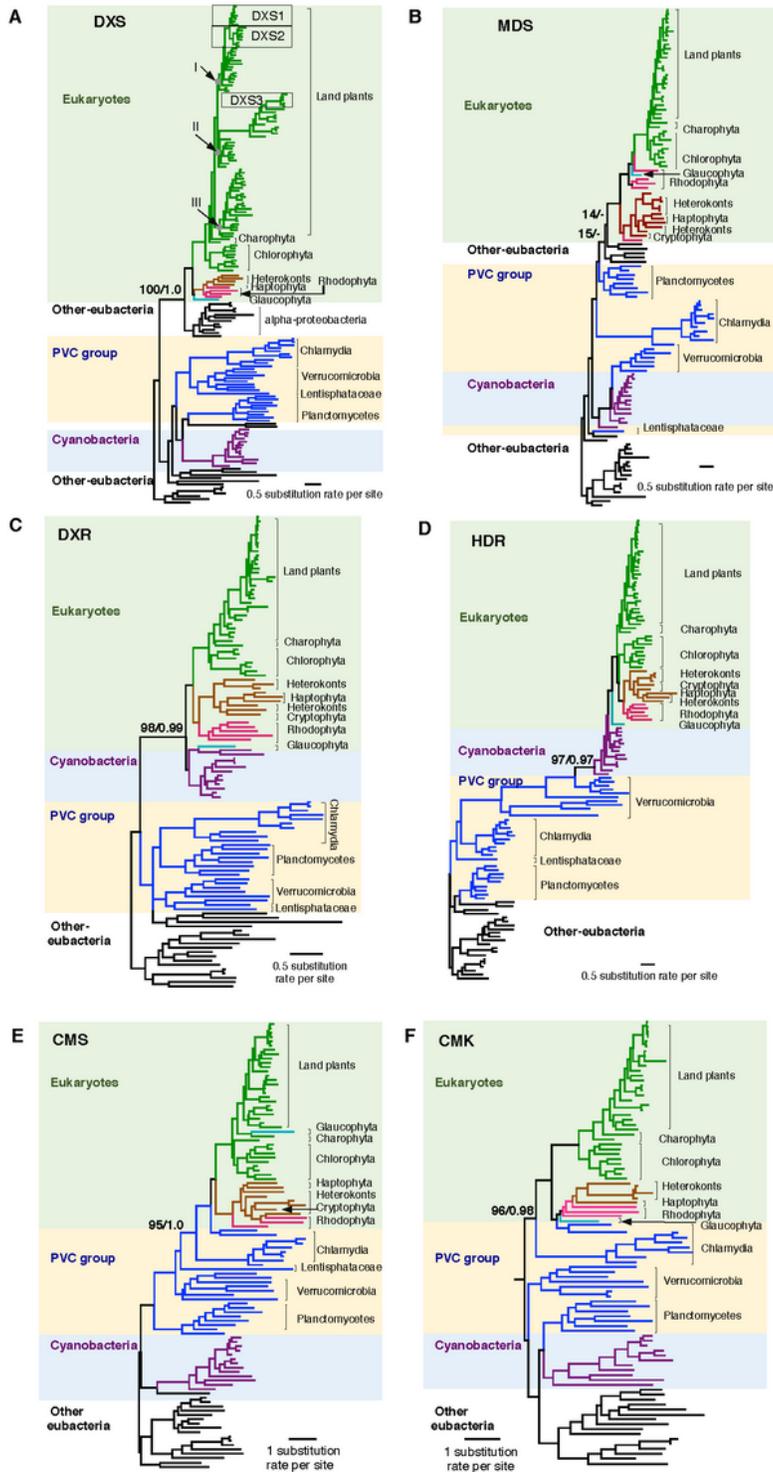
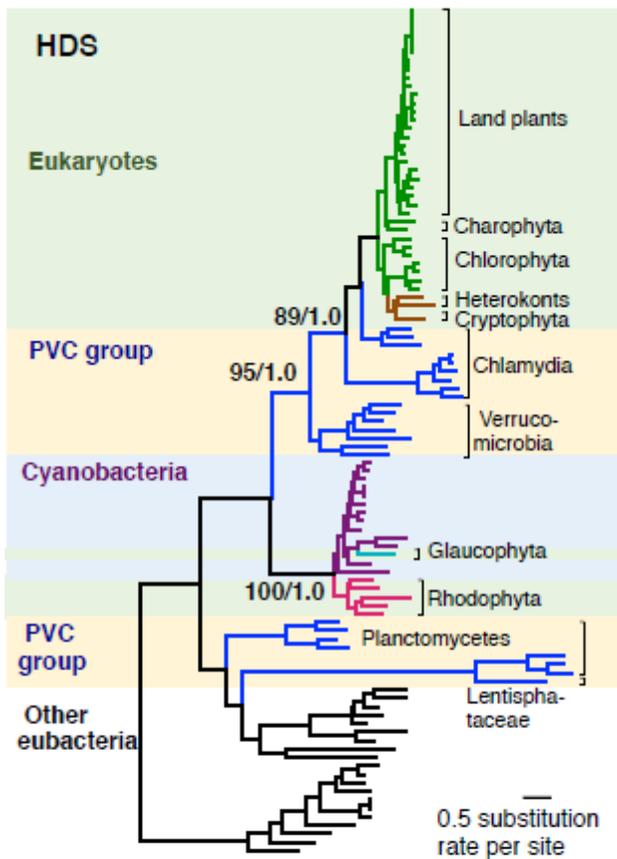


Figure 2

Origin of plastid-bearing eukaryotes MEP-pathway genes. The cladograms display clustering of plastid-bearing eukaryotes DXS (a), MDS (b), DXR (c), HDR (d), CMS (e) and CMK (f) with other-eubacteria, obtained from RAxML using amino acid sequences. Taxa in various major groups are shown in different colors. Species in eukaryotes, PVC groups and Cyanobacteria are shaded with light green, light yellow and light blue, respectively. Numbers associated with branches are bootstrap (BS) values obtained from

RAXML and posterior probability obtained from MrBayes. The dash associated with the branches of eukaryotes and the other-eubacteria shown as 14/- and 15/- indicate that the relationships are not supported by MrBayes.

A



B

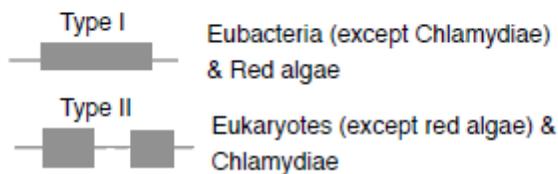


Figure 3

Origin and domain structure of HDS. (a) Cladograms cluster plastid-bearing eukaryotes HDS with chlamydia amino acid sequences. The cladograms are obtained from RAXML using amino acid sequences. Taxa in different major groups are shown in different colors. Species in plastid-bearing eukaryotes, PVC groups and Cyanobacteria are shaded with light green, light yellow and light blue, respectively. Numbers associated with branches are bootstrap (BS) values obtained from RAXML and posterior probability are obtained from MrBayes. (b) Schematics of the types (type I and type II) of gcpE domain in HDS enzyme and their presence in their corresponding species.

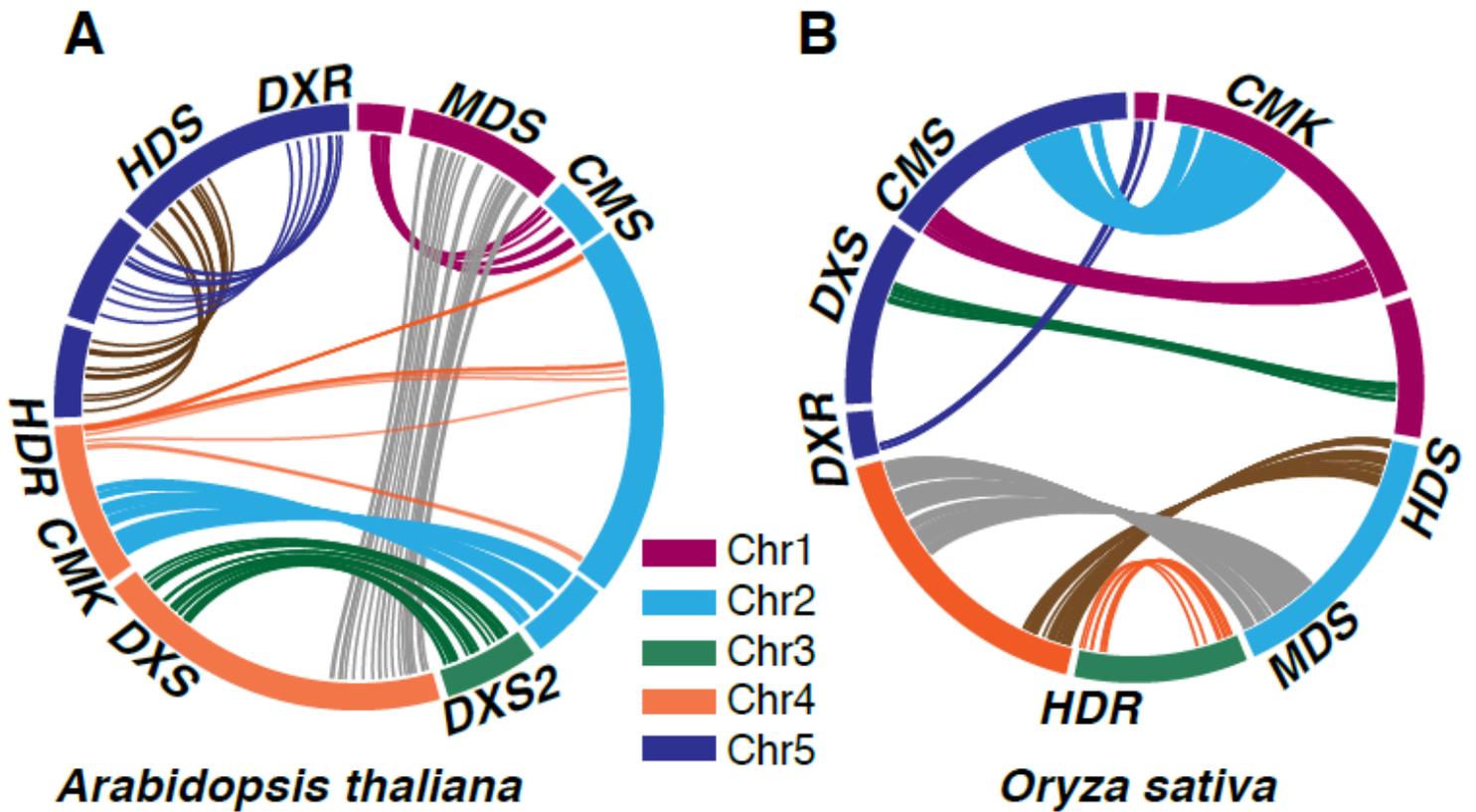


Figure 4

Paralogous syntenic blocks display loss of MEP-pathway duplicated copies. Arabidopsis (a) and rice (b) genes presented in the same colored bands are pairs of paralogues. The band weight correlates to gene numbers on chromosome fragments depicted in different colors.

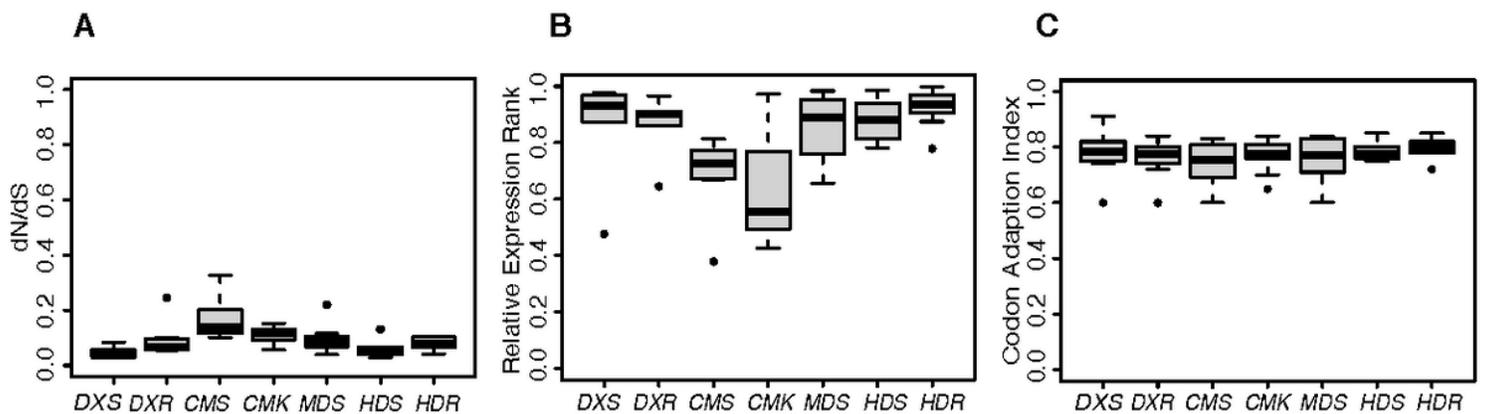


Figure 5

MEP-pathway genes evolved slowly and are highly expressed. (a) The dN/dS ratios of MEP-pathway genes. (b) The relative expression ranking of MEP-pathway genes. (c) The Codon Adaption Index (CAI) of

MEP-pathway genes. All data are displayed as scatter-boxplots with the maximum value of 1.0.

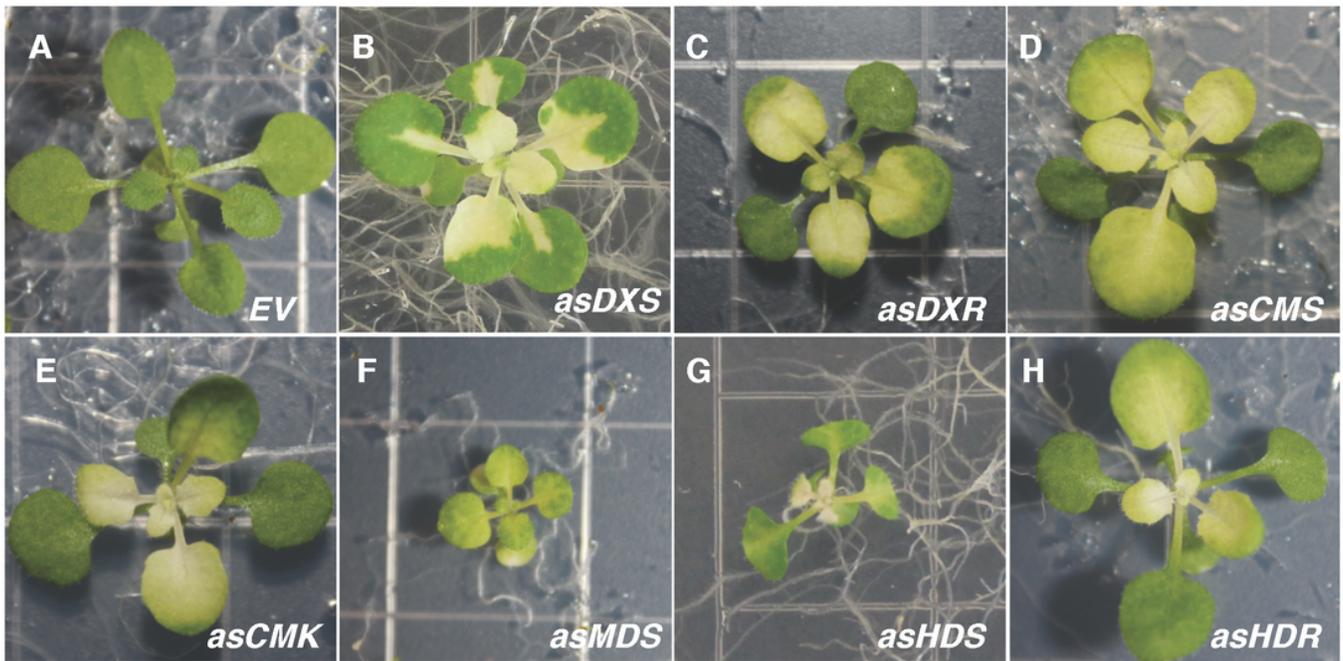


Figure 6

The MEP-pathway genes are indispensable for plant growth. Representative images of 2-week-old seedlings of RNAi lines and wild type transformed with empty vector (EV).

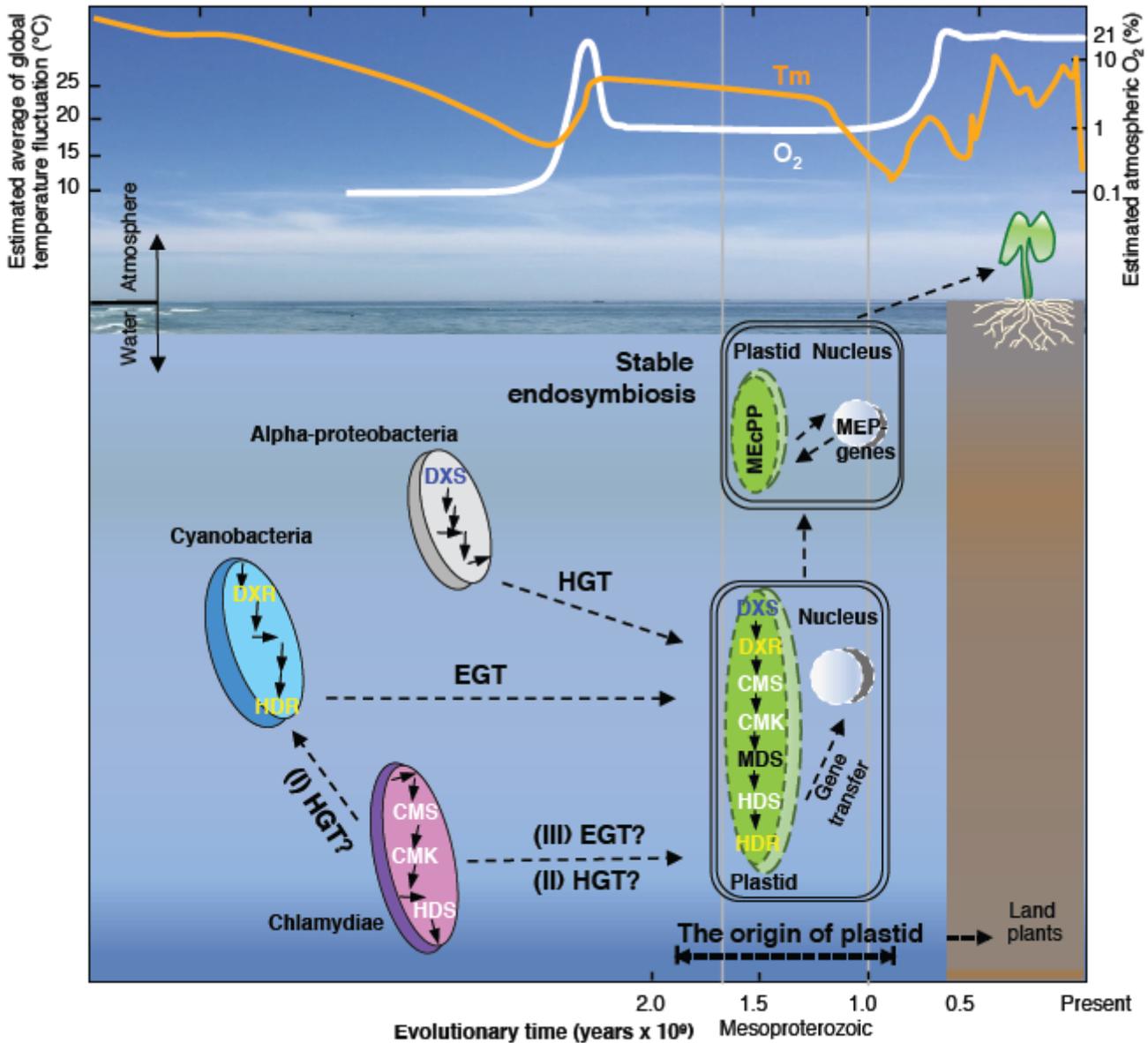


Figure 7

Schematic representation of the probable basis of polyphyletic origin of the MEP-pathway genes. The three proposed scenarios supporting polyphyletic origin of the MEP-pathway genes are: (I) Horizontal Gene Transfer (HGT) of the MEP-pathway genes from chlamydial (magenta) to cyanobacteria (cyan) pangenome followed by the cyanobacterial Endosymbiotic Gene Transfer (EGT) to the eukaryotic host; (II) HGT of genes from Chlamydiae-to-ancestral eukaryotes; (III) Co-contribution of cyanobacteria and Chlamydiae to the MEP-pathway genes. In all the probable cases, the events were ensued by the transfer of genetic material from the endosymbiont(s) to the host nuclear genome (white circle).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [DataS1.Paralogoussyntenicblocks.xlsx](#)
- [SupplementaryTables.docx](#)
- [SupplementaryFigures17.pptx](#)