



SUBJECT AREAS:
MOLECULAR EVOLUTION
PHYLOGENETICS
PLANT EVOLUTION
PHYLOGENY

Received
26 September 2012

Accepted
27 November 2012

Published
11 December 2012

Correspondence and
requests for materials
should be addressed to
A.R.-P. (areyes@unb.ca)

* Equal contribution
made by these authors.

Plastid-localized amino acid biosynthetic pathways of Plantae are predominantly composed of non-cyanobacterial enzymes

Adrian Reyes-Prieto^{1*} & Ahmed Moustafa^{2*}

¹Canadian Institute for Advanced Research and Department of Biology, University of New Brunswick, Fredericton, Canada, ²Department of Biology and Biotechnology Graduate Program, American University in Cairo, Egypt.

Studies of photosynthetic eukaryotes have revealed that the evolution of plastids from cyanobacteria involved the recruitment of non-cyanobacterial proteins. Our phylogenetic survey of >100 *Arabidopsis* nuclear-encoded plastid enzymes involved in amino acid biosynthesis identified only 21 unambiguous cyanobacterial-derived proteins. Some of the several non-cyanobacterial plastid enzymes have a shared phylogenetic origin in the three Plantae lineages. We hypothesize that during the evolution of plastids some enzymes encoded in the host nuclear genome were mistargeted into the plastid. Then, the activity of those foreign enzymes was sustained by both the plastid metabolites and interactions with the native cyanobacterial enzymes. Some of the novel enzymatic activities were favored by selective compartmentation of additional complementary enzymes. The mosaic phylogenetic composition of the plastid amino acid biosynthetic pathways and the reduced number of plastid-encoded proteins of non-cyanobacterial origin suggest that enzyme recruitment underlies the re-compartmentation of metabolic routes during the evolution of plastids.

Primary plastids of plants and algae are the evolutionary outcome of an endosymbiotic association between eukaryotes and cyanobacteria¹. The establishment of the permanent photosynthetic endosymbionts involved critical evolutionary scenarios, such as cyanobacteria surviving the digestive process², the establishment of mechanisms for metabolite exchange between both partners^{3,4}, the evolution of system for the transport of cytoplasmic-translated proteins into the endosymbiotic cells⁵, and the loss or transfer of genes from the endosymbiont into the host nuclear genome⁶. These latter mechanisms contributed to a significant reduction in the plastid gene-coding capacity. Typical plastid genomes encode only circa 10% of the plastid proteome⁷. Recent surveys of diverse plant and algal nuclear genomes have identified dozens to hundreds of plastid-targeted proteins of non-cyanobacterial origin⁸ and cases of plastid-localized pathways composed of enzymes of diverse phylogenetic origin. These pathways include the Calvin Cycle^{9,10} and the shikimate biosynthetic route¹¹.

In addition to their photosynthetic capabilities, key biochemical pathways such as the *de novo* synthesis of fatty acids¹², isoprenoid synthesis¹³, and critical steps of nitrogen assimilation¹⁴, including several pathways for amino acid biosynthesis (Fig. 1) occur in plastids of angiosperms. Nitrogen assimilation (NA) in plastids begins with nitrite (NO₂⁻) uptake and its subsequent reduction to ammonia (NH₃) by the enzyme nitrite reductase. Ammonia and glutamate are converted into glutamine by the plastid glutamine synthetase (GS) and then the amido group of the glutamine is transferred to a molecule of 2-oxoglutarate by the glutamate synthetase (GOGAT), producing a net gain of one glutamate molecule. The concerted activity of these two plastid enzymes constitutes the “GS/GOGAT cycle”, a pivotal step in the biosynthesis of diverse amino acids and other metabolites (Fig. 1). Numerous enzymes participating in the biosynthesis of chorismate¹¹, histidine¹⁵, aromatic¹⁶, branched-chain¹⁷, and aspartate-derived¹⁸ amino acids are localized in plastids of angiosperms and green algae¹⁹. The biosynthesis of methionine²⁰ and cysteine²¹ apparently occurs, but not exclusively, in plastids as well. Enzymes involved in the biosynthesis of proline²² and arginine²³ have been also identified in plant plastids. Thus, diverse experimental and *in silico* evidences strongly suggest that a number of enzymes catalyzing critical reactions for the biosynthesis of several amino acids are localized in angiosperm plastids¹⁴.

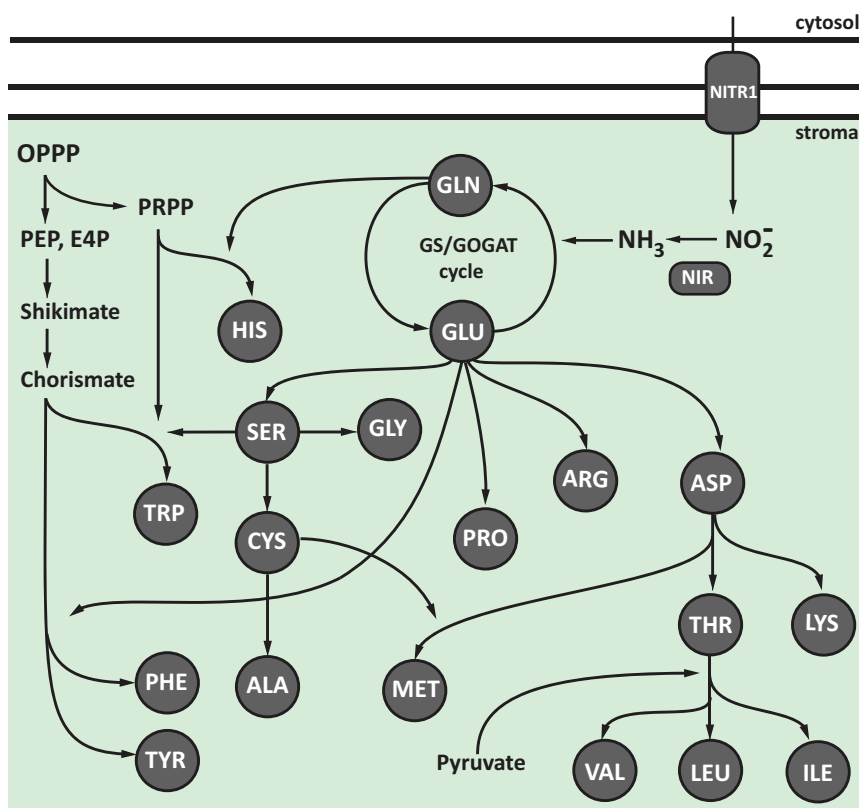


Figure 1 | Schematic metabolic map of the amino acid biosynthetic pathways localized in the *Arabidopsis thaliana* plastids. See Supplementary Table S1 for details of particular enzyme subcellular localization based on experimental and/or sequence evidence. **OPPP**, oxidative pentose-phosphate pathway. **PRPP**, phosphoribosyl-pyrophosphate. **E4P**, Erythrose-4-phosphate. **NITR1**, nitrite transporter. **NIR**, nitrite reductase. Amino acids are indicated with the standard three-letter abbreviation code.

Our central aims are to evaluate if the dissimilar phylogenetic history described for some *Arabidopsis* plastid biochemical routes prevails in the plastid amino acid biosynthesis (AAB) pathways, and, importantly, if this phylogenetic mosaicism is shared among the three different Plantae (*sensu* Cavalier-Smith²⁴) lineages: Viridiplantae (plants and green algae), Rhodophyta (red algae), and Glaucophyta. Considering the ancient (circa 1.5 billions years ago^{25,26}) evolutionary divergence between the three Plantae lineages, we would expect some dissimilarities in the subcellular localization of certain enzymatic reaction or even entire metabolic pathways (i.e., metabolic compartmentation²⁷) as consequence of more than one billion years of independent evolution. However, if the Plantae plastids have a single origin through primary endosymbiosis, we expect as well to identify a number of shared non-cyanobacterial enzymes that represent common recruitments during the early evolution of the plastid proteome. We used the well-characterized and curated biochemical and genomic knowledgebase from *Arabidopsis* as a reference to investigate the evolution of the plastid-localized AAB pathways.

Results

Even though we inferred ML phylogenies of the 158 *Arabidopsis thaliana* proteins (summary of all phylogenetic results and ML trees are presented in the Supplementary Table S1 and Figs. S1-S156) involved in AAB, our central analysis was focused on a subset of 103 proteins that we identified as plastid-localized products (see Methods and Supplementary Table S1). As expected, not all ML trees inferred from single-locus alignments are easily interpretable. To alleviate irregular taxa distribution and stochastic errors inherent to molecular phylogenetic estimations, such as short sequences, compositional biases, use of oversimplified substitution models and incomplete lineage sorting, we focused our analysis on 92 ML trees,

including 62 plastid-localized proteins (Table 1), where at least two Plantae lineages branch together in the same clade, regardless of whether they form a monophyletic group or are intermingled with algal lineages harboring plastids of secondary origin. In order to contrast each ML tree against a *null hypothesis*, we defined our reference “type tree” as the topological pattern where the *Arabidopsis* (Plantae in general) plastid proteins branch with cyanobacterial homologs (e.g., imidazoleglycerol-phosphate dehydratase, Fig. 2; glutamate synthase, Supplementary Figs. S9–10), reflecting an ancestral endosymbiotic origin of the encoding gene. After visual inspection of the 62 trees of plastid-localized, we identified only 21 proteins (corresponding to 13 different enzymatic activities) of unambiguous cyanobacterial origin. Additionally, we identified 8 trees where the cyanobacterial provenance of the plastid-enzyme cannot be entirely discerned (Supplementary Figs. S41-42, S50, S76, S95-98). In contrast, we identified 33 trees presenting topologies of apparent incongruence with the cyanobacterial ancestry of the corresponding enzyme (Table 1). The non-cyanobacterial origin of these 33 proteins was additionally supported in 15 phylogenetic estimations that recovered cyanobacterial homologs in the same tree but branching in different clades and distant from *Arabidopsis* and other Plantae (details in Supplementary Table S1), rejecting the possibility of cyanobacterial origin of those enzymes in Plantae. We identified as well four cases of cyanobacterial-derived enzymes likely localized in the cytoplasm of the plant cells.

The enzymes of the GS/GOGAT cycle. The plastid glutamine synthetase (GS) in *Arabidopsis*, viridiplants in general, and red algae has a putative host origin²⁸. The *Cyanophora* homolog was not recovered in our plastid GS tree, but it was retrieved in the several ML trees estimated when *Arabidopsis* non-plastid GS isoenzymes were used as queries (Supplementary Table S1). The



Table 1 | Phylogenetic origin of 62 plastid-localized enzymes involved in amino acid biosynthesis

Amino acid biosynthesis pathway	Plastid-localized isoenzymes	Cyanobacterial	Non-cyanobacterial	ML bootstrap (%)
Glutamine				
Glutamine synthetase	1		1	100
Glutamate				
Ferredoxin-dependent Glutamate synthase (GOGAT)	2	2		>63
Tryptophan				
Anthranilate synthase alpha subunit	2		2	>70
Anthranilate synthase beta subunit	2	2		<50
Anthranilate phosphoribosyl transferase	1		1	<50
Phosphoribosyl anthranilate isomerase	3		3	95
Tryptophan synthase alpha subunit	2	2		>75
Tryptophan synthase beta subunit	2	2		100
Phenylalanine/Tyrosine				
Arogenate dehydratase/prephenate dehydratase	6		6	<50
Arogenate dehydrogenase	2		2	<50
Histidine				
ATP phosphoribosyltransferase	2		2	<50
BBMII isomerase	1		1	<50
Imidazoleglycerol-phosphate dehydratase	2	2		>94
Histidinol-phosphate aminotransferase	2		2	>95
Histidinol phosphate phosphatase	1		1	<50
Aspartate-derived				
Aspartate kinase	1		1	<50
Lysine				
L-diaminopimelate aminotransferase	1		1	<50
Diaminopimelate epimerase	1	1		58
Diaminopimelate decarboxylase	2		2	>94
Threonine				
Threonine synthase	1		1	<50
Methionine				
Cystathionine gamma-synthase	1		1	100
Isoleucine and Valine				
Acetolactate synthase large subunit	1		1	99
Acetolactate synthase small subunit	1	1		95
Ketolacid reductoisomerase	1		1	<50
Leucine				
Isopropylmalate isomerase	4		4	<50
Isopropylmalate dehydrogenase	4	4		>75
Arginine				
N-Acetylglutamate synthase	1		1	100
N-Acetylglutamate kinase	1	1		69
N-Acetylglutamate-5-P reductase	1		1	91
N2-Acetylornithine:glutamate acetyltransferase	1	1		99
Carbamoyl-P synthetase Large subunit	1	1		100
Ornithine carbamoyltransferase	1		1	<50
P II nitrogen sensing protein GLB I	1	1		<50
Serine				
3-phosphoglycerate dehydrogenase	2		2	<50
Glycine				
Serine (glycine) hydroxymethyltransferase	2		2	65
Serine (alanine):glyoxylate aminotransferase	1		1	94
Alanine				
Cysteine desulfurase	1	1		<50
Total	62	21	33+8	

only GS encoded in the genome of the red alga *Cyanidioschyzon merolae* is probably cytosolic-localized²⁹; however, the GS subcellular localization in glaucophyte algae is unknown. In contrast to the convoluted evolution of the plastidic GS, the two *Arabidopsis* plastid Fd-GOGAT are of cyanobacterial provenance (Table 1). The Fd-GOGAT encoding gene is still present in plastid genomes of red algae. The plastid NADH-dependent GOGAT, more abundant in non-photosynthetic tissues, is of a putative host origin. Given that GS is likely cytosolic in *Cyanidioschyzon* and the Fd-GOGAT is plastid-localized, the existence of a plastid GS/GOGAT cycle in plastids of red algae is unlikely³⁰. These results and previous

evidence indicate that the plastid-localization of the GS/GOGAT cycle is a particularity of streptophytes and that those two enzymatic players have distinct ancestral origins.

Aromatic amino acids and histidine biosynthesis. Our analyses indicate that the plastid routes for the biosynthesis of the aromatic amino acids tryptophan, tyrosine, and phenylalanine (Fig. 1) comprise only 4 out of 11 (i.e., < 40%) enzymatic components of cyanobacterial origin (see Table 1). An interesting case is the plastid-localized alpha subunits of the anthranilate synthase (ASA), which have non-cyanobacterial origin in Plantae lineages, branching in

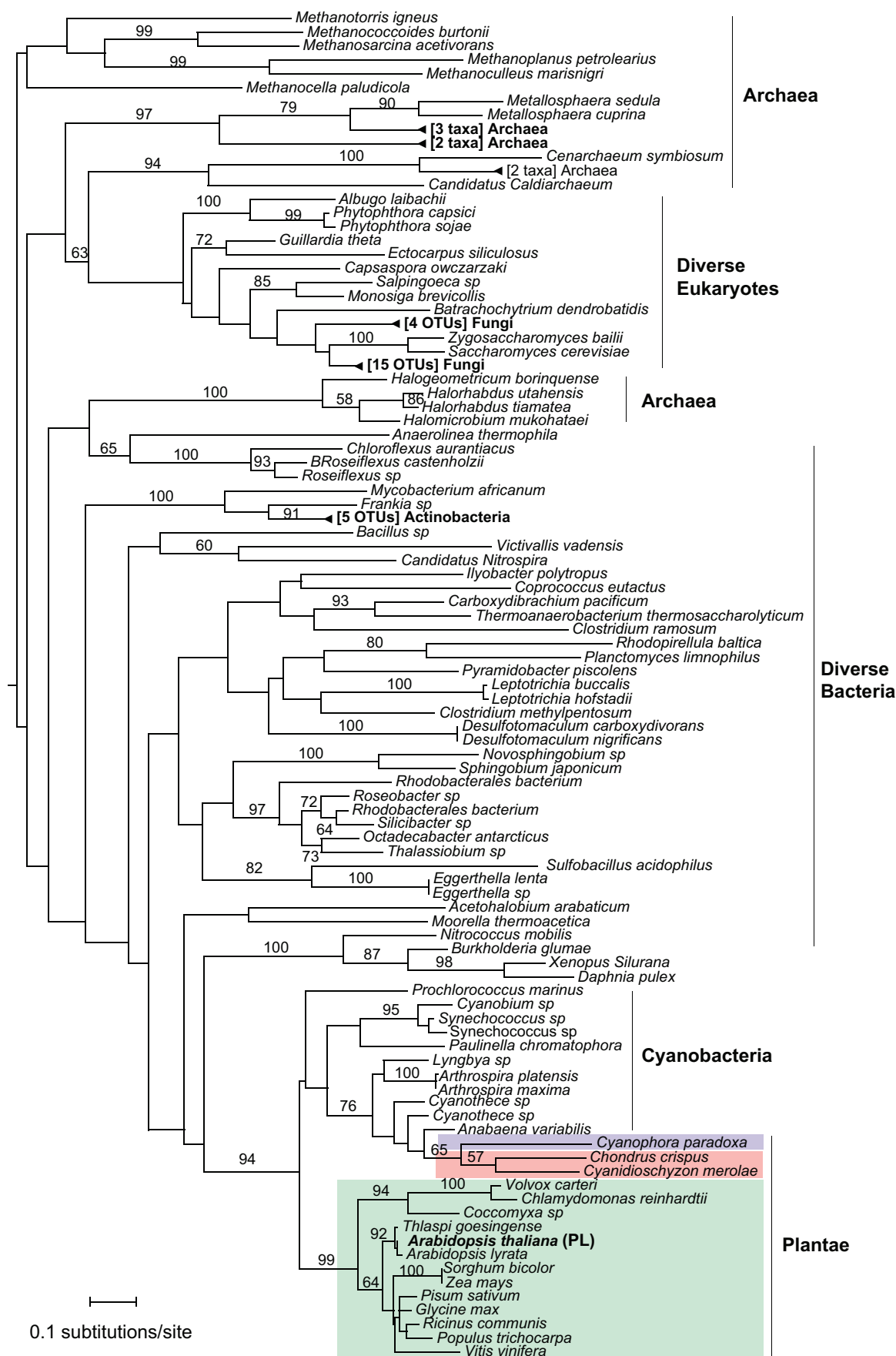


Figure 2 | Maximum Likelihood phylogenetic tree of the imidazoleglycerol-phosphate dehydratase (At4g14910.1). The numbers at the nodes indicate RaxML bootstrap values (only values >50% are shown). Branch lengths are proportional to the number of substitutions per site (see the scale bars). Colored boxes highlight Plantae lineages (glaucoophytes, red algae and viridiplants). **PL**, plastid localized protein.



well-supported clades with Planctomycetes bacteria homologs (Fig. 3a and Table 1; >70% bootstrap support, BS). In contrast, the two different plastid-localized ASA beta subunits seem to be cyanobacterial-derived in Viridiplantae and Glaucophyta (Fig. 3b; no BS); however, the branching position of the red algal sequences is not well resolved (Supplementary Fig. S14). The ASA beta subunit is still plastid-encoded in both glaucophytes and red algae. These results suggest that the plastidic ASA is, at least in viridiplants and glaucophytes, an oligomeric complex constituted by enzyme subunits of disparate phylogenetic origins. The overall result illustrates the mosaic composition of the Plantae plastid tryptophan biosynthesis pathway, similarly as described before in the secondary plastids of diatoms and other stramenopiles³¹. Several of the enzymes (e.g., chorismate mutase, prephenate aminotransferase, arogenate dehydratase/prephenate dehydratases, and arogenate dehydrogenases) of non-cyanobacterial origin involved in tyrosine and phenylalanine biosynthesis have no apparent homologs in cyanobacterial genomes (Table S1).

The plastidic route for histidine biosynthesis includes only one enzyme (out of 11 enzymes) of unambiguous cyanobacterial provenance in all Plantae: the imidazoleglycerol-phosphate dehydratase (two isoenzymes; Supplementary Figs. S46–47)¹⁵. The ML trees of the ATP phosphoribosyltransferase and histidinol phosphate phosphatase are inconclusive and a cyanobacterial origin for these enzymes cannot be conclusively disproved (Table S1). Some enzymes in the histidine biosynthetic pathway were likely recruited for plastid roles from bacterial sources other than the cyanobacterial ancestor of the plastid (Supplementary Table S1). Our results demonstrate that the plastid pathway for the synthesis of histidine has a mosaic phylogen-

etic constitution as well and depicts a complex evolutionary history that involves independent losses and gains of non-cyanobacterial enzymes in the different Plantae lineages. For example, the histidinol-phosphate aminotransferase (Fig. 4) has a shared non-cyanobacterial origin in the tree Plantae lineages, indicating this enzyme was recruited for plastid functions by the Plantae ancestor from Chloroflexi bacteria ($\geq 95\%$ BS). Other cases indicate that some non-cyanobacterial plastid enzymes are unique evolutionary innovations in viridiplants, such as the host-derived histidinol dehydrogenase (Supplementary Table S1).

Aspartate-derived and branched-chain amino acid biosynthesis.

The phylogenetic scrutiny of the plastid-localized enzymes involved in the biosynthesis of aspartate, lysine, threonine, methionine, and isoleucine (22 enzymatic activities represented by 41 nuclear-encoded proteins) revealed only three proteins of unambiguous cyanobacterial origin involved in the biosynthesis of these amino acids (Supplementary Table S1). The three cyanobacterial-derived enzymes are the dihydrodipicolinate reductase (only present in streptophytes), the diaminopimelate epimerase and the small subunit of the acetolactate synthase (Fig. 5b; 95% BS). The lysine biosynthesis sub-route encompasses three enzymes of non-cyanobacterial origin in viridiplants (dihydrodipicolinate synthase) and Plantae lineages (diaminopimelate aminotransferase and two diaminopimelate decarboxylases), respectively (Table 1). Both the biosynthesis of threonine from aspartate-semialdehyde and methionine comprises enzymes of non-cyanobacterial provenances (Supplementary Table S1). Remarkably, the phylogenetic tree of the large subunit of acetolactate synthase (ALS; Fig. 5a) illustrates the

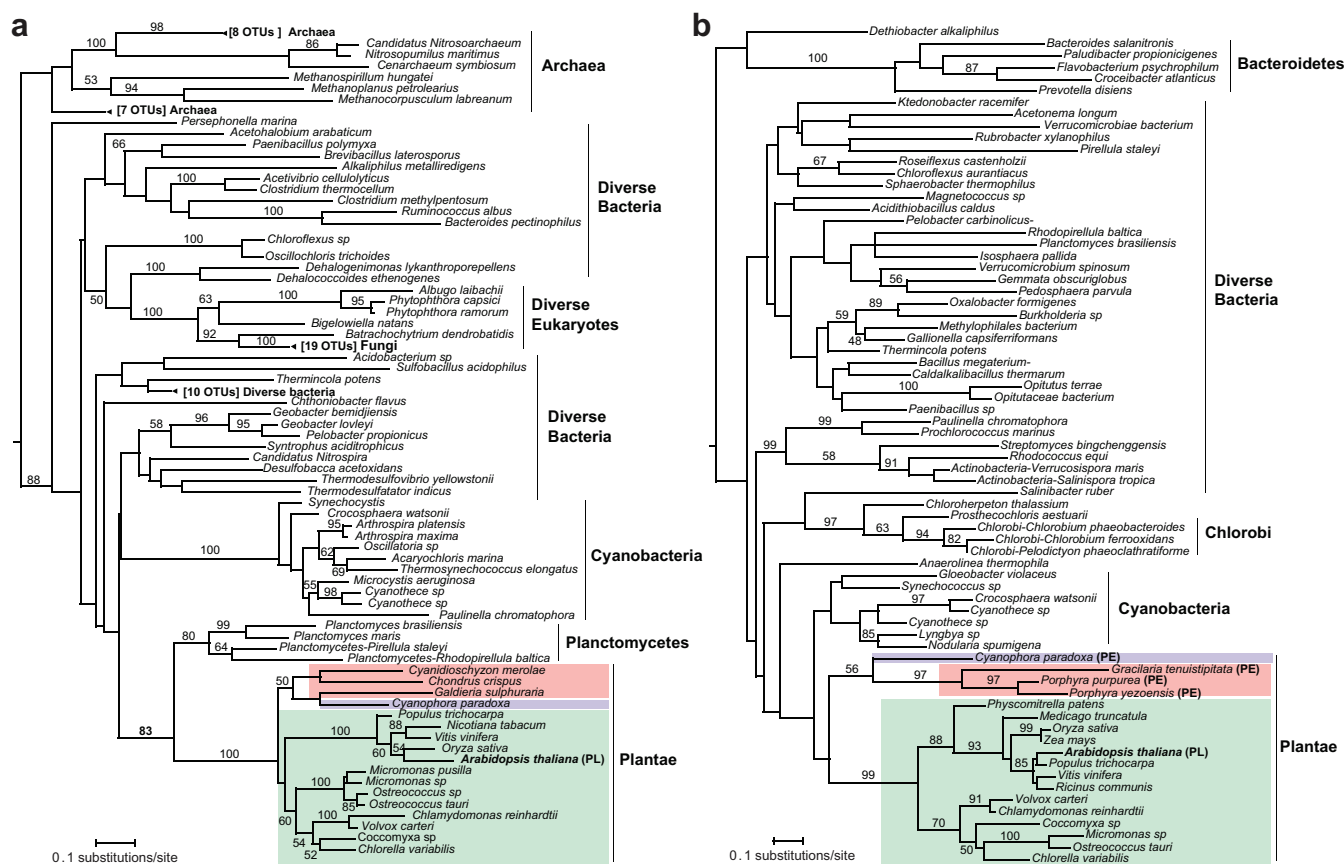


Figure 3 | Phylogenetic analysis of the plastid anthranilate synthase complex. Maximum Likelihood phylogenetic tree of the (a) alpha subunit, or component I (At5g05730.1) and (b) beta subunit, or component II (At1g25220.1) of the anthranilate synthase. The numbers at the nodes indicate RaxML bootstrap values (only values >50% are shown). Branch lengths are proportional to the number of substitutions per site (see the scale bars). Colored boxes highlight Plantae lineages **PL**, plastid localized protein; **PE**, plastid encoded protein.

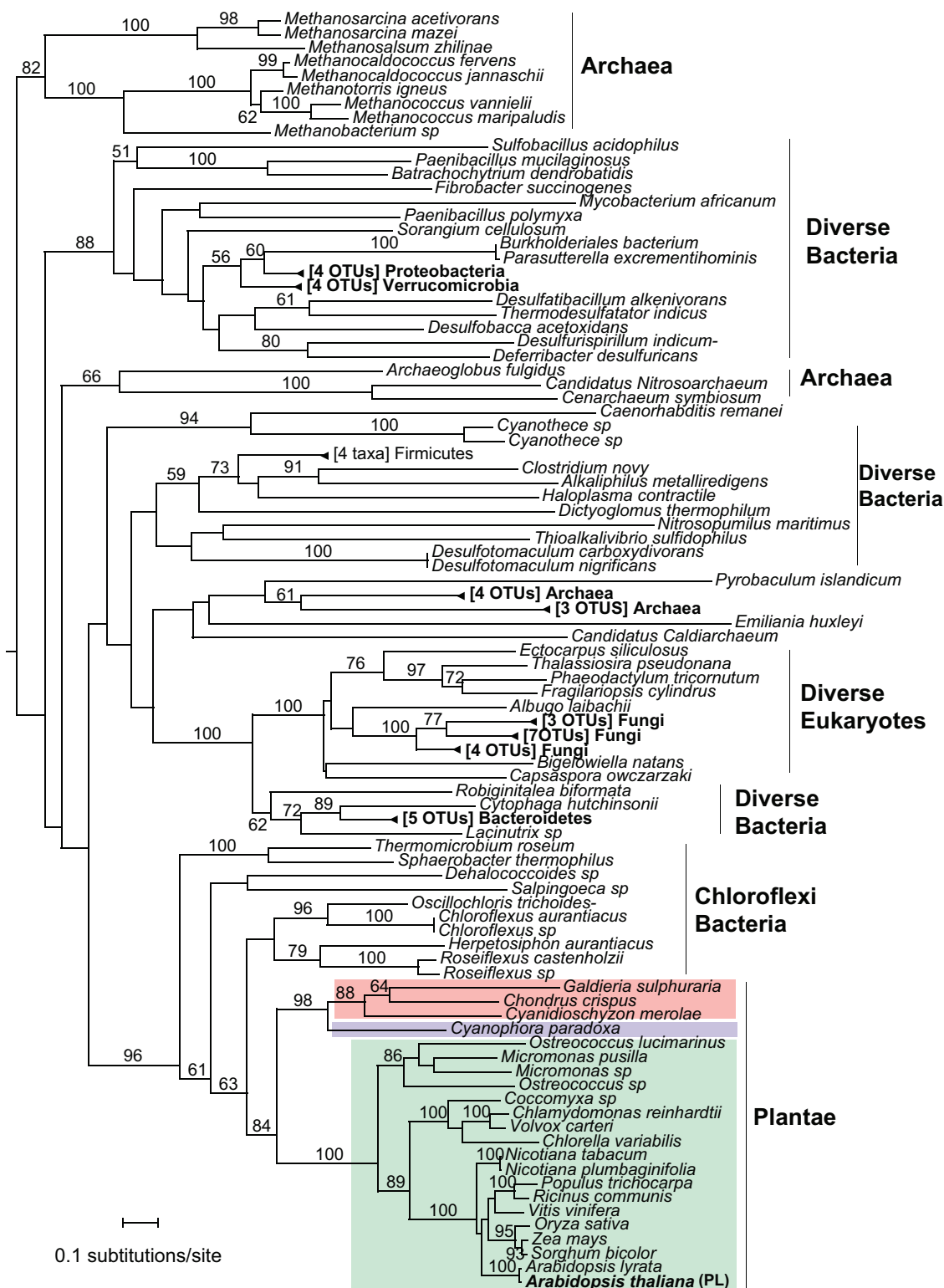


Figure 4 | Phylogenetic analysis of the plastid histidinol-phosphate aminotransferase (At1g71920.1). The numbers at the nodes indicate RaxML bootstrap values (only values >50% are shown). Branch lengths are proportional to the number of substitutions per site (see the scale bars). Colored boxes highlight Plantae lineages. **PL**, plastid localized protein.

non-cyanobacterial origin of this enzyme in viridiplants and *Cyanophora*, but its cyanobacterial derivation in the case of the plastid-encoded red algal homolog. Considering that the ALS small subunit branches in the same clade with cyanobacterial homologs (Fig. 5b), the overall ALS results (Fig. 5) suggest that the ALS in viridiplants and *Cyanophora* is another example of a plastid

oligomeric complex constituted by proteins of disparate ancestral origins. The other five enzymes (three reactions) participating in isoleucine biosynthesis have non-cyanobacterial origins (Table 1). The plastid biosynthesis of the branched-chain (BCH) amino acids leucine and valine from pyruvate involves the activity as of ALS as well and the non-cyanobacterial enzymes ketol-acid reductoiso-

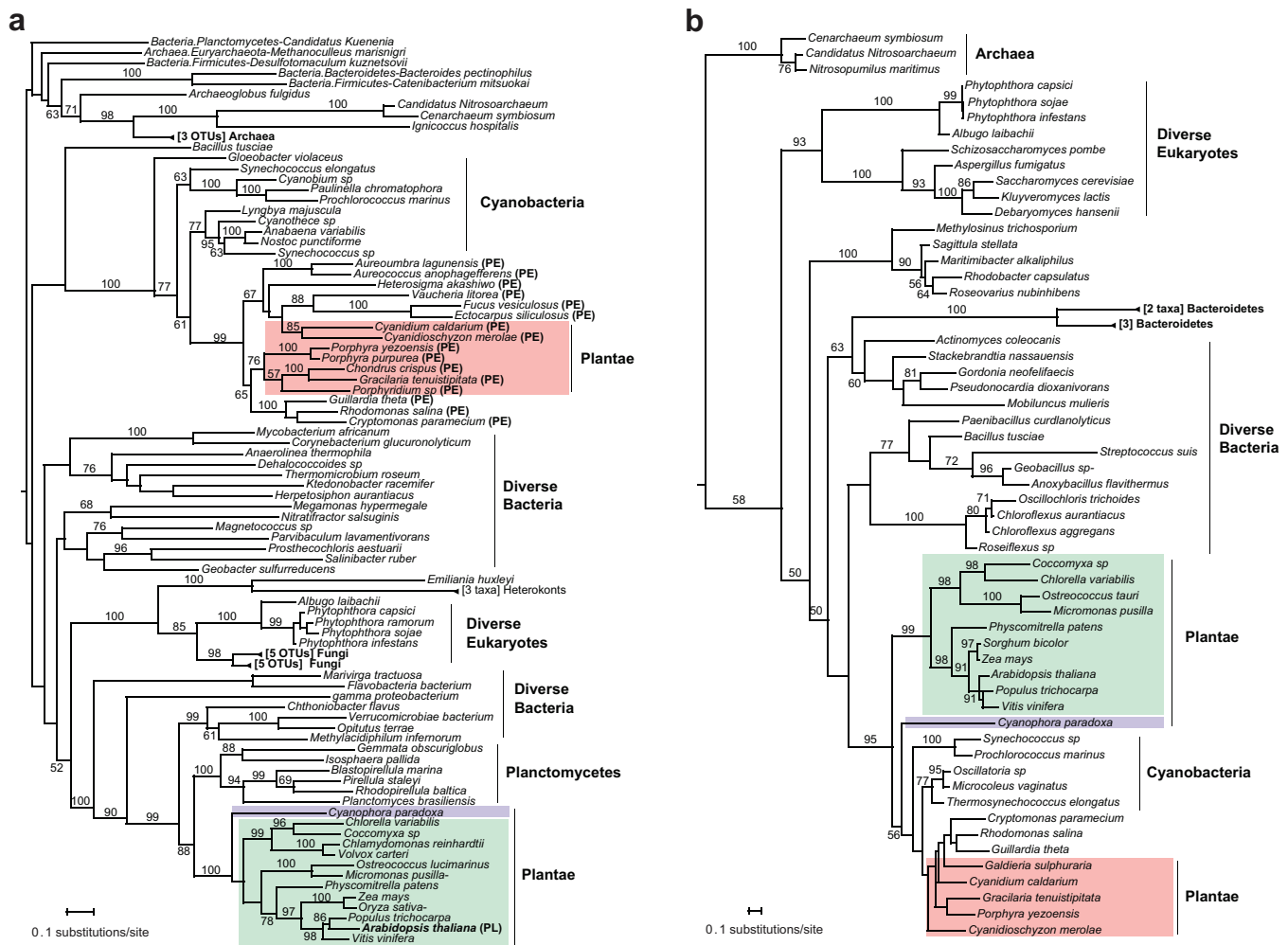


Figure 5 | Phylogenetic trees of the acetolactate synthase. (a) large (At3g48560.1) and (b) small (At2g31810) subunits. The numbers at the nodes indicate RaxML bootstrap values (only values >50% are shown). Branch lengths are proportional to the number of substitutions per site (see the scale bars). **PL**, plastid localized protein; **PE**, plastid encoded protein.

merase, dihydroxy acid dehydratase (only present in viridiplants) and several BCH aminotransferases (Supplementary Table S1). Finally, several isopropylmalate dehydrogenases (Figs. S99–102) involved in leucine biosynthesis represent the only cases of cyanobacterial-derived enzymes in this pathway (Table 1).

Arginine and proline biosynthesis. Several enzymes that participate in arginine biosynthesis from glutamate and ornithine, have been localized in plastids^{32,33} (Supplementary Table S1). The overall phylogenetic analysis of the 12 enzymes involved in arginine biosynthesis revealed that the N-acetylglutamate kinase, N2-acetylornithine-glutamate acetyltransferase, and the large subunit of the carbamoyl phosphate synthetase (CPS; Fig. S110) are the only three proteins of unambiguous cyanobacterial origin (Table 1 and Supplementary Table S1). The phylogenetic tree of the small subunit of the CPS (Supplementary Fig. S111) indicates the non-cyanobacterial origin of this protein in Viridiplantae but cyanobacterial-derived and plastid-encoded in red algae. These results suggest CPS is another possible case of a plastid oligomer constituted by subunits of different origins, similar to the ASA (Fig. 3) and ALS cases (Fig. 5). Even though the plastid occurrence of the arginine pathway awaits further experimental verification, our results show a mosaic composition for this pathway with only a minor contribution of cyanobacterial-derived components. Plastid proline biosynthesis from glutamate involves the activity of the non-

cyanobacterial enzymes pyrroline-5-carboxylate synthase (only present in viridiplants) and the pyrroline-5-carboxylate reductase (Supplementary Table S1).

Serine, glycine, cysteine, and alanine biosynthesis. The three plastidic enzymes that participate in serine biosynthesis from 3-phosphoglycerate have non-cyanobacterial origin (Supplementary Table S1). The phylogeny of the plastid enzyme serine:glyoxylate aminotransferase (SGAT; Fig. 6), which is part of the route for glycine biosynthesis from glyoxylate, suggests a unique non-cyanobacterial origin in the three Plantae lineages (94% BS). This result suggests that the Plantae common ancestor recruited SGAT for plastid functions from other bacterial sources. The plastid-localized glutamate:glyoxylate aminotransferase, involved in glycine biosynthesis from glyoxylate, is of a non-cyanobacterial origin as well. Thus, the two plastid-localized alternative routes for glycine biosynthesis evolved from enzymes of different origins (Supplementary Table S1). The plastid enzymes serine O-acetyltransferase and cysteine synthase/O-acetylserine lyase, which catalyze the conversion of serine to cysteine, are other cases of non-cyanobacterial proteins recruited for plastid functions. Finally, the plastid cysteine desulfurase, which catalyzes the alanine synthesis from cysteine, is a protein of cyanobacterial origin (albeit with no significant bootstrap support), which is present only in viridiplants and *Cyanophora* (Table 1).

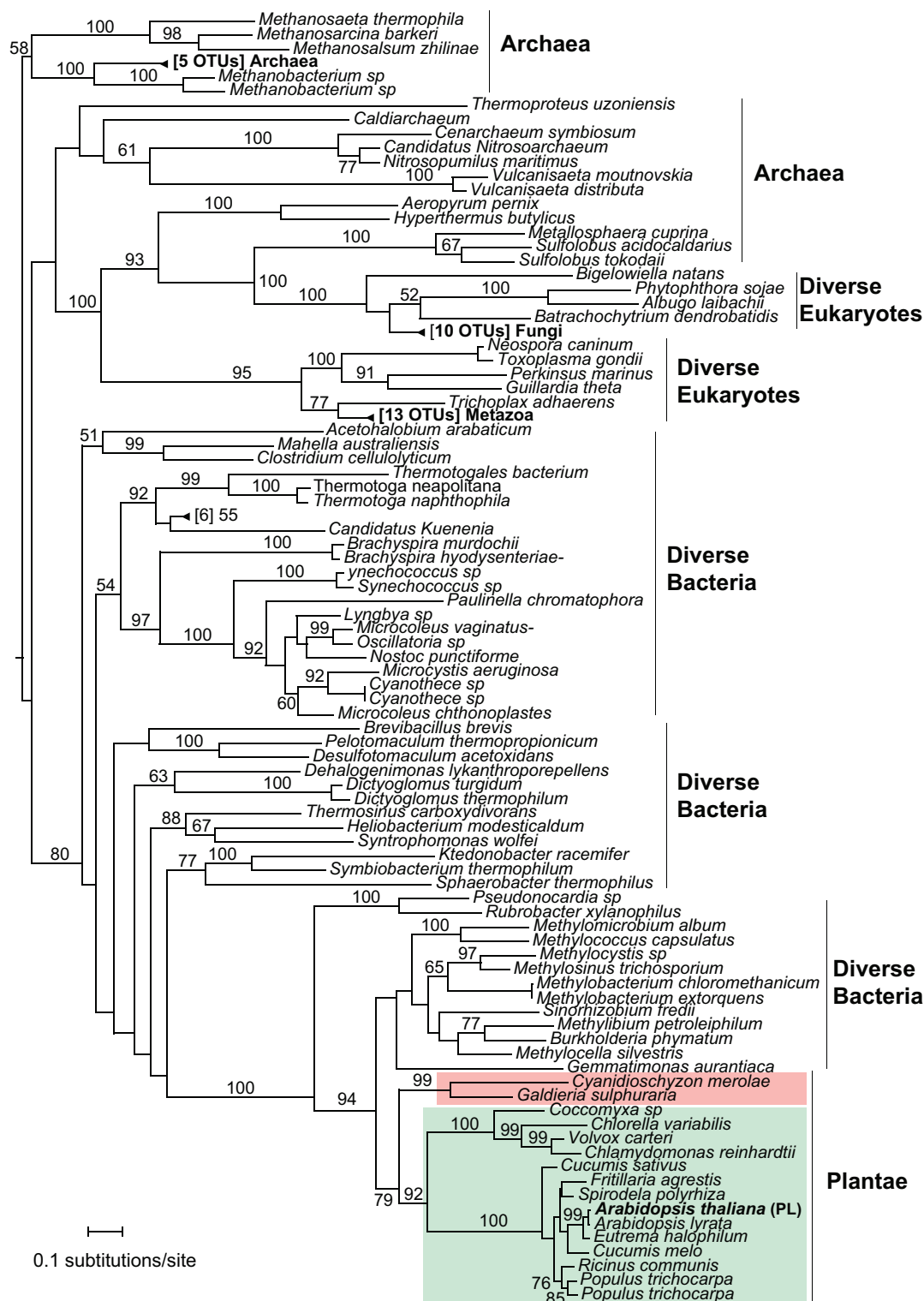


Figure 6 | Phylogenetic trees of the plastid-localized serine: glyoxylate aminotransferase (At2g13360.2). The numbers at the nodes indicate RaxML bootstrap values (only values >50% are shown). Branch lengths are proportional to the number of substitutions per site (see the scale bars). PL, plastid localized protein.

Discussion

Our phylogenetic survey shows that circa two-thirds (41/62 proteins and 26/39 enzymatic activities) of the *Arabidopsis* (and streptophytes in general) plastid enzymes involved in AAB evolved from non-cyanobacterial sources. We suggest that the phylogenetic mosaicism of the plastid AAB pathways is, in part, an outcome of ancient recruitment of non-cyanobacterial enzymes for plastid functions

throughout the establishment of the endosymbiotic relationships that gave rise to the photosynthetic organelle. A key question is whether the phylogenetic mosaic composition of the AAB plastid pathways in *Arabidopsis*, and streptophytes in general, is an ancestral trait shared with other Plantae lineages. Our results reveal that several of the streptophytes plastid non-cyanobacterial enzymes branch in moderate to well-supported clades (>80% BS) together with green



algae, red algae, or glaucophytes (e.g., Figs. 3a, 4, 5a, 6 and Supplementary Figs. S26, S48, S74–75, S78, S106). Other ML trees with no significant support are consistent with this branching pattern as well (Supplementary Figs. S25, S28, S33–38, S39, S40, S59, S72, S112, S120, S122–123). Therefore, the most parsimonious scenario to explain the shared non-cyanobacterial ancestry of the Plantae plastid enzymes is that several alien enzymes were “re-compartmentalized” by the host protein sorting system into the novel photosynthetic organelle early during the evolution of the Plantae ancestor^{34,35}. The common non-cyanobacterial ancestry of several enzymes involved in AAB illustrates that, in addition to their photosynthetic role, plastids became critical compartments for the “assembly” of the high energy consuming routes of nitrogen assimilation.

Even though enzyme re-compartmentation is a likely explanation for the phylogenetic mosaicism of the plastid AAB routes, in principle, we cannot discard that the plastid ancestor acquired the genes encoding non-cyanobacterial enzymes via horizontal gene transfer (HGT) prior to engulfment by the eukaryotic host. HGT has been largely recognized as a major force in prokaryote genome evolution and we assume this pervasive force had some impact on the evolution of the genome of the plastid ancestor. It is possible to speculate that a fraction of genes recruited via HGT by the plastid ancestor was subsequently transferred into the host nuclear genome via endosymbiotic gene transfer (EGT⁶). However, it is indispensable to consider, based upon the magnitude of ancestral HGT observed in plastid genomes, whether this scenario is sufficient to explain the extensive phylogenetic mosaicism evident in plastid AAB pathways. The genes of the RuBisCO operon (*rbcL*, *rbcS*)³⁶, the gene *cbhX*^{37,38} from red alga, and the seven genes involved in menaquinone/phyloquinone biosynthesis (*menF*, *menD*, *menC*, *menB*, *menE*, *menH* and *menA*)³⁹ of cyanidiales are the few well-documented cases of Plantae plastid genes of likely ancient (i.e., before plastid origins) non-cyanobacterial origin. However, our blastp search of the proteins encoded in seven Plantae plastid genomes (Supplementary Table S2; see methods) versus all bacterial sequences in GenBank suggests that most plastid protein-encoding genes are of cyanobacterial origin (> 90% in *Cyanophora paradoxa*, *Pyropia yezoensis* and 3 diverse viridiplants, and > 70% in extremophilic cyanidiales). Thus, the low number of plastid genes of possible non-cyanobacterial origin (< 10%) in mesophilic Plantae makes it unlikely that the pre-endosymbiosis HGT scenario best explains the extensive phylogenetic mosaicism of the plastid AAB pathways (i.e., 66% of non-cyanobacterial enzymes). There is no reason to assume that HGT has more extensively affected genes encoding enzymes involved in the diverse pathways for AAB than other gene sets before the plastid establishment. As indirect comparative reference, it is important to note that the genome of the cyanobacterial-derived organelle in the filose amoeba *Paulinella chromatophora* FK01 contains only 33 genes acquired via HGT before the endosymbiosis that gave rise to the chromatophore (i.e., 4% of the chromatophore genes are non-cyanobacterial)⁴⁰. This evidence provides an independent estimation of the relatively low number of alien genes present in other organelle genome of cyanobacterial origin. In summary, these results indicate that the non-cyanobacterial phylogenetic (66% of the proteins) signal observed in the composition of the plastid-localized AAB pathways is higher than that estimated for the gene repertoires in cyanobacterial-derived organelles (4–10%) of independent origins. This comparison suggests that post-endosymbiosis enzyme recruitment and re-compartmentation are the most likely explanations for our findings.

A key aspect of the plastid proteome evolution is the elucidation of the evolutionary forces that might have driven the re-compartmentation of the several enzymes involved in AAB into the organelle. Intracellular compartmentation by the presence of diverse membranous organelles generates a non-homogenous distribution of soluble compounds and enzymes critical for the organization and regulation of the eukaryotic cell metabolism⁴¹. Thus, differential

metabolite concentrations permit efficient enzymatic activities and effective regulation of metabolic routes. Most noticeable cases of metabolic intracellular compartmentation involve the localization of entire biochemical pathways in particular organelles, such as the fatty acid beta-oxidation pathway, the tricarboxylic acid cycle and oxidative phosphorylation in the mitochondrion, the glycolysis in the cytosol, and the Calvin cycle, and terpenoid biosynthesis in the plastid. There are cases of re-compartmentation of complete pathways to different organelles such as the glycolysis relocation inside peroxisomes in trypanosomatids⁴². As envisaged by Ginger *et al.*, physical retargeting of entire metabolic pathways to new cellular compartments implies establishment and retuning of the regulatory mechanisms⁴². We hypothesize that re-compartmentation, or even *de novo* assembly, of AAB pathways into the photosynthetic organelle comprised both adaptive and non-selective changes derived from the stochastic traffic (i.e., mistargeting) of cytosolic and mitochondrial enzymes into the plastids. In principle, it is plausible to suppose that the physical relocation of non-cyanobacterial enzymes into the photosynthetic organelle occurred by incidental mistargeting via the TIC/TOC protein import machinery (ChloroP 1.1 analysis showed 87/103 of the analyzed plastid enzymes have predicted plastid transit peptides; Supplementary Table S1). Subsequently, we hypothesize that the catalytic activities of some of the mistargeted enzymes were transiently and randomly coupled to the plastid metabolite pools, biochemical intermediaries, effectors and the activity of native enzymes of the photosynthetic organelle. However, how were entire non-cyanobacterial pathways compartmentalized and assembled in the new photosynthetic organelle? If we assume that several proteins were occasionally mistargeted to subcellular locations different from the “correct” compartment, then it is plausible to predict that some enzymes catalyzing reactions associated with anabolic pathways were stochastically active in “wrong” organelles (“minor mistargeting model”²⁷). Moreover, a certain degree of intermingling between the mistargeted proteins and the endosymbiont enzymes catalyzing similar, preceding or subsequent (e.g., fortuitous substrate channeling; use of common or non-specific substrates) reactions possibly sustained a steady state of novel catalyzed reactions inside the photosynthetic organelle. It is reasonable to expect that transitory enzymatic activities, sustained by recurrent mistargeting of non-cyanobacterial enzymes, produced slightly diverse plastid phenotypes²⁷. These new plastid-localized enzymatic activities were initially neutral. However, the evolution of the obligate host-organelle interdependence in the Plantae ancestor opened a unique window for selective forces acting over the alien enzymatic activities and their incidental interactions in plastids. Thus, the advantageous re-compartmentation of some foreign enzymes into the organelle and the inexorable tendency of endosymbiont genomes to be reduced overtime established an ideal scenario for both replacement of some endosymbiont enzymes and subsequent selection of other “complementary” mistargeted enzymes. Overall, the phylogenetic results reflect a process of protein recruitment underlying the re-compartmentation of the AAB routes during the evolution of the plastid proteome.

Nevertheless, we still need to explain the ulterior evolution of novel regulatory mechanism and metabolic control exerted by the light quality and intensity and the redox intermediaries of the photosynthesis (e.g., phytochromes, ferredoxin, and NADPH)¹⁴ over the plastid amino acid production. Plastid nitrogen assimilation is tightly coupled with the flow of carbon compounds (e.g., products of photorespiration, glycolysis, and the tricarboxylic acid cycle) into the organelle, which are an essential source of carbon skeletons for amino acid biosynthesis¹⁴. Under this scenario, we predict that the active transport (i.e., availability) and the concentration of carbon compounds (e.g., organic acids) and ammonia were also critical conditions for the successful re-compartmentation of several AAB pathways in the plastid. Consistent with this scenario, it is well known that nitrogen assimilation in plastids depends mostly on



photosynthetic energy: in plant photosynthetic cells, 80% of the redox intermediaries required for nitrogen assimilation are regenerated by photochemically-reduced ferredoxin¹⁴. The activities of GS and Fd-GOGAT are up-regulated by light- and sugar-signaling in plant plastids⁴³. Additionally, GS activity decreases when the photosynthetic rate is low⁴⁴ and the GOGAT activity is mediated by light via phytochrome⁴⁵. Overall, different plastid routes such as the GS/GOGAT cycle, the chorismate biosynthesis, key steps of histidine biosynthesis, and synthesis of the aspartate-family require ATP and NADPH¹⁴, which are actively produced in the plastid during the photosynthesis light-dependent reactions. The plastid pathways for the synthesis of branched-chain amino acids from pyruvate are light-regulated as well⁴⁶. The ATP/AMP and NADPH/ NADP⁺ ratios inside the photosynthetic organelle provided a favorable microenvironment to randomly sustain significant catalytic activities of mistargeted enzymes. At particular plastid internal substrates concentrations, some transitory enzymatic activities were sufficient to generate slightly different biochemical phenotypes. Thus, we consider that the rate of production of ATP and NADPH during light-dependent reactions of photosynthesis was an important selective element that favored the assembly of energy-demanding and redox-regulated AAB routes during the evolution of the plastid. In summary, diverse metabolic data suggest that the plastid redox-energy balance favored the re-compartmentation of these enzymes into the photosynthetic organelle.

Our results demonstrate that in addition to the quintessential role of plastids for the emergence of the eukaryotic photoautotrophic lifestyle (which relies mostly on an ancestral protein core of cyanobacterial origin), these photosynthetic organelles resulted in target compartments for the gradual assembly and re-compartmentation of entire biosynthetic pathways by recruiting non-cyanobacterial enzymes. Our results reveal that several of these non-cyanobacterial plastid enzymes are encoded by genes likely acquired over time by the host via HGT from diverse prokaryotic sources^{47,48}. Well-known examples of this evolutionary process include essential plastid proteins such as the ATP/ADP translocator⁴⁹ and enzymes involved in starch metabolism⁵⁰, that likely originated from ancestral parasitic Chlamydiae-like bacteria. These scenarios suggest that transitory energy parasites were important genetic donors for the establishment of the plastid as well^{50,51}. The fact that many of these non-cyanobacterial plastid enzymes are shared between green, red, and glaucophyte algae suggests that at least some part of this foreign enzymatic collection was anciently assembled in their common ancestor. Although, our comparative search (Supplementary Table S2) and phylogenetic evidence⁵² do not support the possibility that the Plantae non-cyanobacterial plastid enzymes of the AAB pathways were originally present in the genome of the plastid ancestor, and later transferred to the nucleus of the host through EGT, it is still possible that extant cyanobacterial genomes are radically different from those ancient cyanobacteria that gave rise to the Plantae plastid more than a billion years in the past.

In the recent years, it has been demonstrated that the incorporation of bacterial genes has been an important factor in the evolution of primary plastids^{8,51}. It has been estimated that circa 40% of the plastid proteins shared between red algae and Viridiplantae originated from the host repertoire and/or diverse bacteria⁸. The detailed evolutionary history of the plastid NA network, which we present here, emphasizes the role of the metabolic re-compartmentation and evolutionary innovation for the assembly of several amino acid biosynthetic pathways and their relevance in the integration of the plastid protein repertoire. The overall implication is that these processes demonstrate the ancestral and critical participation of the cyanobacterial-derived compartment in the evolution of the whole plant cell metabolism beyond its primary distinctive contribution as photoautotrophic partner.

Methods

Identification of enzymes involved in plastid amino acid biosynthesis. We used information of the Plant Metabolic Network database (PMN; www.plantcyc.org, May–August, 2010) to identify 158 reported, or predicted, *Arabidopsis thaliana* proteins involved in AAB. The corresponding amino acid sequences were retrieved from the *Arabidopsis* Information Resource database (TAIR; http://www.arabidopsis.org) for further phylogenetic analyses. The final protein set included several cases of paralogous sequences encoded in the *Arabidopsis thaliana* genome (see details of analyzed sequences in Supplementary Table S1).

Protein subcellular localization. We analyzed information from previous experimental investigations that report the subcellular localization of most of the 158 nuclear-encoded *Arabidopsis* proteins involved in AAB (see Supplementary Table S1). Additionally, we used three different computational protocols (Predotar, TargetP, and WoLF PSORT)^{53–55} to predict the subcellular localization of the 158 *Arabidopsis* proteins. We also used ChloroP 1.1⁵⁶ to investigate the presence of transit peptides in the predicted plastid-targeted proteins.

Phylogenetic analysis of *Arabidopsis* enzymes involved in amino acid biosynthesis. Each of the 158 retrieved *Arabidopsis* proteins was used as individual query to identify homolog sequences from our local protein database comprising the RefSeq GenBank data and conceptual translations of transcriptomic datasets from diverse eukaryotes, which do not have genome sequence available yet⁵⁷. We established a blastp cutoff E-value $\leq 1e-10$ to identify homolog proteins. In order to partially reduce sampling biases given the taxonomic composition of the database and to maximize the taxon coverage, we constrained the incorporation of the BLAST hits into the multiple alignment with an increasing maximum number on entries for each major taxonomic rank, starting from one sequence per species up to a maximum of 200 sequences per “Kingdom” according to the NCBI Taxonomy definitions (http://www.ncbi.nlm.nih.gov/taxonomy)⁵⁷. The homologues protein sequences were aligned with MUSCLE⁵⁸. Resulting multiple protein alignments were manually verified and edited using Se-AL 2.0 (http://tree.bio.ed.ac.uk/software/seal/). Partial sequences with more than 50% of missing amino acid residues were discarded from further analyses. We estimated maximum likelihood (ML) trees from each multiple alignment using the protein substitution model LG+FF+ Γ implemented in RAxML HPC-MPI 7.2.8 with 100 bootstrap replicates⁵⁹. Visual inspection of the estimated unrooted ML trees was carried out with Archaeopteryx 0.997 beta version (http://www.phylosoft.org/archaeopteryx/). During the visual inspection of the unrooted ML trees, we used the Archaeopteryx *Root/Reroot* option to manually select at least three alternative rooting nodes (i.e., outgroup selection) in each tree to reduce potential misidentification of taxa branching patterns (i.e., monophyletic groups) as result of possible rooting artifacts. Our tree visual inspection aimed to distinguish between plastid enzymes of unambiguous cyanobacterial origin and those plastid proteins branching with homologs from taxonomic groups different from cyanobacteria. We distinguished plastid proteins of cyanobacteria provenance when the *Arabidopsis* (viridiplants in general) sequences branch in the same clade with cyanobacterial homologs and either any member of the Plantae group and/or algae with plastids of secondary origin (e.g., “chromists”, chlorarachniophytes, and euglenids). In contrast, we defined plastid proteins of non-cyanobacterial origin in those cases as the *Arabidopsis* protein branches with other Plantae, algal homologs, and lineages other than cyanobacteria. ML trees in Newick format and refined multiple protein alignments are available upon request.

Searching for non-cyanobacterial genes in plastids genomes. In order to identify plastid genes acquired via horizontal gene transference (HGT), we carried out a blastp reciprocal search (cutoff E-value $\leq 1e-15$) of all proteins encoded in the plastid genomes of seven diverse Plantae (*Arabidopsis thaliana* [85 protein models], *Mesostigma viride* [105], *Chlamydomonas reinhardtii* [69], *Cyanophora paradoxa* [149], *Pyropia yezoensis* [209], *Cyanidioschyzon merolae* [203], and *Cyanidium caldarium* [209]) against all prokaryotic sequences present in GenBank (as of June 2012). We used the taxonomic profile of the top-ten BLASTP hits as proxy to identify the phylogenetic origin of each plastid-encoded query. We arbitrarily defined a plastid-encoded protein of cyanobacterial origin when at least six out of ten top hits (i.e., $\geq 60\%$) were cyanobacterial (different species) homologs. In contrast, those plastid queries with less than five cyanobacteria within the top-ten hits (i.e., $< 50\%$) were considered non-cyanobacterial. Considering that horizontal incorporation of protein-coding genes into plastids genomes is highly constrained⁶², this blastp similarity approach provides a conservative estimation of the number of alien genes present in the genome of the plastid ancestor.

1. Reyes-Prieto, A., Weber, A. P. & Bhattacharya, D. The origin and establishment of the plastid in algae and plants. *Annu Rev Genet* **41**, 147–168 (2007).
2. Sagan, L. On the origin of mitosing cells. *J. Theor. Biol.* **14**, 225–IN6 (1967).
3. Bhattacharya, D., Archibald, J. M., Weber, A. P. & Reyes-Prieto, A. How do endosymbionts become organelles? Understanding early events in plastid evolution. *Bioessays* **29**, 1239–1246 (2007).
4. Weber, A. P. M., Linka, M. & Bhattacharya, D. Single, ancient origin of a plastid metabolite translocator family in Plantae from an endomembrane-derived ancestor. *Eukaryot. Cell* **5**, 609–12 (2006).
5. Cavalier-Smith, T. The origins of plastids. *Biol. J. Linn. Soc.* **17**, 289–306 (1982).



6. Martin, W. & Herrmann, R. Gene transfer from organelles to the nucleus: how much, what happens, and Why? *Plant. Physiol.* **118**, 9–17 (1998).
7. Leister, D. Chloroplast research in the genomic age. *Trends Genet* **19**, 47–56 (2003).
8. Suzuki, K. & Miyagishima, S. Y. Eukaryotic and eubacterial contributions to the establishment of plastid proteome estimated by large-scale phylogenetic analyses. *Mol Biol Evol* **27**, 581–590 (2010).
9. Martin, W. & Schnarrenberger, C. The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. *Curr. Gen.* **32**, 1–18 (1997).
10. Reyes-Prieto, A. & Bhattacharya, D. Phylogeny of Calvin cycle enzymes supports Plantae monophyly. *Mol Phylogenet Evol* **45**, 384–391 (2007).
11. Richards, T. A. *et al.* Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements. *Eukaryot. Cell* **5**, 1517–31 (2006).
12. Ohlrogge, J. B., Kuhn, D. N. & Stumpf, P. K. Subcellular localization of acyl carrier protein in leaf protoplasts of *Spinacia oleracea*. *Proc Natl Acad Sci U S A* **76**, 1194–1198 (1979).
13. Lichtenthaler, H. K., Schwender, J., Disch, A. & Rohmer, M. Biosynthesis of isoprenoids in higher plant chloroplasts proceeds via a mevalonate-independent pathway. *FEBS Lett* **400**, 271–274 (1997).
14. Lancien, M., Lea, P. J. & Azevedo, R. A. Amino Acid Synthesis in Plastids in Wise, R. R. & Hooper, J. K. *The Structure and Function of Plastids*, Springer. 355–385 (2006).
15. Stepansky, A. & Leustek, T. Histidine biosynthesis in plants. *Amino Acids* **30**, 127–142 (2006).
16. Rippert, P., Puyaubert, J., Grisolle, D., Derrier, L. & Matringe, M. Tyrosine and phenylalanine are synthesized within the plastids in *Arabidopsis*. *Plant Physiol* **149**, 1251–1260 (2009).
17. Diebold, R., Schuster, J., Daschner, K. & Binder, S. The branched-chain amino acid transaminase gene family in *Arabidopsis* encodes plastid and mitochondrial proteins. *Plant Physiol* **129**, 540–550 (2002).
18. Jander, G. & Joshi, V. Aspartate-Derived Amino Acid Biosynthesis in *Arabidopsis thaliana*. *The Arabidopsis Book* **7**, 1–15 (2008).
19. Vallon, O. & Spalding, M. H. Amino acid metabolism in Stern, D. & Harris, E. *The Chlamydomonas Sourcebook 2nd Edition, Volume 2: Organellar and Metabolic Processes*, Elsevier. 115–158 (2009).
20. Ravel, S. *et al.* Methionine metabolism in plants: chloroplasts are autonomous for de novo methionine synthesis and can import S-adenosylmethionine from the cytosol. *J Biol Chem* **279**, 22548–22557 (2004).
21. Schmutz, D. & Brunold, C. Intercellular Localization of Assimilatory Sulfate Reduction in Leaves of *Zea mays* and *Triticum aestivum*. *Plant Physiol* **74**, 866–870 (1984).
22. Rayapati, P. J., Stewart, C. R. & Hack, E. Pyrroline-5-Carboxylate Reductase Is in Pea (*Pisum sativum* L.) Leaf Chloroplasts. *Plant Physiol* **91**, 581–586 (1989).
23. Peltier, J. B. *et al.* The oligomeric stromal proteome of *Arabidopsis thaliana* chloroplasts. *Mol Cell Proteomics* **5**, 114–133 (2006).
24. Cavalier-Smith, T. Eukaryote kingdoms: seven or nine? *Biosystems* **14**, 461–481 (1981).
25. Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G. & Bhattacharya, D. A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* **21**, 809–18 (2004).
26. Hedger, S. B., Blair, J. E., Venturi, M. L. & Shoe, J. L. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol. Biol.* **4**, 2 (2004).
27. Martin, W. Evolutionary origins of metabolic compartmentalization in eukaryotes. *Philos Trans R Soc Lond B Biol Sci* **365**, 847–855 (2010).
28. Ghoshroy, S., Binder, M., Tartar, A. & Robertson, D. L. Molecular evolution of glutamine synthetase II: Phylogenetic evidence of a non-endosymbiotic gene transfer event early in plant evolution. *BMC Evol Biol* **10**, 198 (2010).
29. Terashita, M., Maruyama, S. & Tanaka, K. Cytoplasmic localization of the single glutamine synthetase in a unicellular red alga, *Cyanidioschyzon merolae* 10D. *Biosci. Biotech. Bioch.* **70**, 2313–5 (2006).
30. Imamura, S. *et al.* Nitrate assimilatory genes and their transcriptional regulation in a unicellular red alga *Cyanidioschyzon merolae*: genetic evidence for nitrite reduction by a sulfite reductase-like enzyme. *Plant Cell Physiol* **51**, 707–717 (2010).
31. Jiroutova, K., Horak, A., Bowler, C. & Obornik, M. Tryptophan biosynthesis in stramenopiles: eukaryotic winners in the diatom complex chloroplast. *J Mol Evol* **65**, 496–511 (2007).
32. Ferrario-Méry, S., Besin, E., Pichon, O., Meyer, C. & Hodges, M. The regulatory PII protein controls arginine biosynthesis in *Arabidopsis*. *FEBS letters* **580**, 2015–20 (2006).
33. Slocum, R. D. Genes, enzymes and regulation of arginine biosynthesis in plants. *Plant Physiol Biochem* **43**, 729–745 (2005).
34. Rodríguez-Ezpeleta, N. *et al.* Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr. Biol.* **15**, 1325–30 (2005).
35. Price, D. C. *et al.* *Cyanophora paradoxa* Genome Elucidates Origin of Photosynthesis in Algae and Plants. *Science* **335**, 843–847 (2012).
36. Delwiche, C. F. & Palmer, J. D. Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. *Mol. Biol. Evol.* **13**, 873–882 (1996).
37. Maier, U. G., Fraunholz, M., Zauner, S., Penny, S. & Douglas, S. A nucleomorph-encoded CbbX and the phylogeny of RuBisCo regulators. *Mol. Biol. Evol.* **17**, 576–83 (2000).
38. Fujita, K., Ehira, S., Tanaka, K., Asai, K. & Ohta, N. Molecular phylogeny and evolution of the plastid and nuclear encoded cbbX genes in the unicellular red alga *Cyanidioschyzon merolae*. *Genes Genet. Syst.* **83**, 127–33 (2008).
39. Gross, J., Meurer, J. & Bhattacharya, D. Evidence of a chimeric genome in the cyanobacterial ancestor of plastids. *BMC Evol. Biol.* **8**, 117 (2008).
40. Reyes-Prieto, A. *et al.* Differential gene retention in plastids of common recent origin. *Mol. Biol. Evol.* **27**, 1530–7 (2010).
41. Bowsher, C. G. & Tobin, A. K. Compartmentation of metabolism within mitochondria and plastids. *J Exp Bot* **52**, 513–527 (2001).
42. Ginger, M. L., McFadden, G. I. & Michels, P. A. Rewiring and regulation of cross-compartmentalized metabolism in protists. *Philos Trans R Soc Lond B Biol Sci* **365**, 831–845 (2010).
43. Thum, K. E., Shasha, D. E., Lejay, L. V. & Coruzzi, G. M. Light- and carbon-signaling pathways. Modeling circuits of interactions. *Plant Physiol* **132**, 440–452 (2003).
44. Woodall, J., Boxall, J. G., Forde, B. G. & Pearson, J. Changing perspectives in plant nitrogen metabolism: the central role of glutamine synthetase. *Science Progress* **1–26** (1996).
45. Becker, T. W., Nef-Campa, C., Zehnacker, C. & Hirel, B. Implication of the phytochrome light regulation of the tomato gene(s) encoding ferredoxin dependent glutamate synthase. *Plant. Physiol. Bioch.* 725–729 (1993).
46. Schulze-Siebert, D., Heineke, D., Scharf, H. & Schultz, G. Pyruvate-Derived Amino Acids in Spinach Chloroplasts: Synthesis and Regulation during Photosynthetic Carbon Metabolism. *Plant Physiol* **76**, 465–471 (1984).
47. Doolittle, W. F. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet* **14**, 307–311 (1998).
48. Watkins, R. F. & Gray, M. W. The frequency of eubacterium-to-eukaryote lateral gene transfers shows significant cross-taxa variation within amoebzoa. *J Mol Evol* **63**, 801–814 (2006).
49. Linka, N. *et al.* Phylogenetic relationships of non-mitochondrial nucleotide transport proteins in bacteria and eukaryotes. *Gene* **306**, 27–35 (2003).
50. Ball, S., Colleoni, C., Cenci, U., Raj, J. N. & Tirtiaux, C. The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. *J Exp Bot* **62**, 1775–1801 (2011).
51. Moustafa, A., Reyes-Prieto, A. & Bhattacharya, D. Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. *PLoS One* **3**, e2205 (2008).
52. Rice, D. W. & Palmer, J. D. An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. *BMC Biol.* **4**, 31 (2006).
53. Small, I., Peeters, N., Legeai, F. & Lurin, C. Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* **4**, 1581–1590 (2004).
54. Emanuelsson, O., Brunak, S., Von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2**, 953–971 (2007).
55. Horton, P. *et al.* WoLF PSORT: protein localization predictor. *Nucleic Acids Res* **35**, W585–7 (2007).
56. Emanuelsson, O., Nielsen, H. & von Heijne, G. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* **8**, 978–84 (1999).
57. Moustafa, A., Bhattacharya, D. & Allen, A. E. iTREE: A high-throughput phylogenomic pipeline. *5th Cairo International Biomedical Engineering Conference* 103–107 (2010).
58. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).
59. Stamatakis, A. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).

Acknowledgments

The authors acknowledge Andreas P.M. Weber and Shawn R. MacLellan for their comments on the manuscript. ARP further acknowledges the Integrated Microbiology Program of the Canadian Institute for Advanced Research. The present work was supported by the Natural Sciences and Engineering Research Council of Canada (project 402421-2011), the Canada Foundation for Innovation (project 28276), and the New Brunswick Innovation Foundation (project RIF2012-006) awarded to ARP.

Author contributions

A.R.P. and A.M. conceived, designed and performed the experiments. A.R.P. and A.M. participated equally in discussions, analysis of the results and towards writing of the manuscript.

Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>



Competing financial interests: The authors declare no competing financial interests.

License: This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

How to cite this article: Reyes-Prieto, A. & Moustafa, A. Plastid-localized amino acid biosynthetic pathways of Plantae are predominantly composed of non-cyanobacterial enzymes. *Sci. Rep.* 2, 955; DOI:10.1038/srep00955 (2012).