Supplementary data

Table S2. Putatively plastid-targeted proteins in *P. wickerhamii* for which the N-terminal methionine could not be assigned unambiguously or the 5' start codon is missing (truncated transit-peptide)

	Protein[a]	**Comments**	**Intracellular localization in plants and algae**[b]
1	>AY616046; similarity to Toc33, putative protein import apparatus, plastid outer membrane DRTYQDACLPWRNAAVTAKTGSFTVLLLGKGGVGKSATVNSLLGERAAVVSSFQATGLGVHMHTRTLPGGFALNLIDTPSLLDQDSVSTSRLEQIGSAIKGVKIDAVLFLDRLDVYSTDTLDEQVVDGVTAYFGEDMWDHAVLGLTRATSSAPPLSTDFGEWVVERSLQLQSLIAKARAQYAPRRRRGEAPPLTPAGCRESQECPRKLRGSGSSPATAPGSPTCMPRWRSAATGAGQAWLHDAARR	TOC 33/TOC 34 proteins from land plants are part of the protein import apparatus located on the plastid outer membrane and probably they do not have plastid-targeting peptides (see the results section).	**P**[#]
2	>AY616047; similarity to Tic22, putative protein import apparatus, plastid inner membrane ARSFNGELQCAAQRCVESWQRNIAKQSAALHATVSEHVEGARRGVQSVTQQLSRLRVVPSAQHASQPSNLLASVTTAARAVTGSFPQPSFDLALAPDEIKSRLANVPIFTVVNAKDEFVLISGEQAEKRQLGLFFFNESDAQALVATMKAADAVVGRQARVLPTSLSSVYDFAVAPRASPGPESVVFRFVPEPSQVAAA	The length of the truncated N-terminal extension was estimated by taking into account the predicted cleavage site for BAD35192.	**P**[#]
3	>AY616051; hexose transporter HRTCGTWRRPVGPWTSAEERAFTRPSRFFMSRRVASPIVNAMQGHAGESLISTPRPAHHLSPQVSLGVGVAGMGALCFGYHLGVVNGPLDAIAADLGFSGNAMLQGLAVSSLLAFAAVGSLGGSGLADRLGTAPRLPAGLRPTPPRPPAQLHGHLPRPAHRWACHL	The length of the truncated N-terminal extension was estimated by taking into account the predicted cleavage site for AAF74565.	**P**[#]
4	>AY616055; phosphoglycerate dehydrogenase KRPGNVGSHFASQVLGTSGTCMGGRAATAPFRTVTVRRASRGMRASTLVRAAIGTNGRPTVLVAEKLGAGGVDMLKEVADVRTVLNMSKEQLLENISSVDAIVIRSATKVTREVIEASKGRLRVVGRAGVGIDNVDLTAASESGCLVVNAPTANTIAAAEHGISLLTALARNVAQADASMKAGGWDRTTYVGVSVVGKTLAIIGFGKVGGEVARRAKGLGMTVIAFDPYAS	Actinobacteria, NZ_AAEB01000043.	**P**

5	>AY616066; acetyl-CoA carboxylase alpha subunit [c] GTYASLTSCCTDDRTPSSHGAAGQALSSVAILGCIRAPLLPCTLSKTPCYARHGRAGPPDTAQGSGGRREAGDWFQMLLARFNPISEKASNTAVLDFEKPLVELDHKIREVRKVAEQNGVDVSSQIQELEARARQLRKETYAKLTPVQRLQVARHPNRATFLDVALNITDKFVELHGDRMGLDDQPWCAASPGMDDTSFMFIGHQKGRSTKENIYRNFGMPQPNGYRKALRFMRMADKFGFPI	Cyanobacteria, AAM82646.	**P, C, ER**
6	>AY616068; enoyl-[acyl-carrier protein] reductase SSRKPHALWAPCGQQGVTRALPVPGARQSRTSQVQMRRTSAPVKHVAASRGMSVRAAAGNGAAAPSVGLPIDLRGKKAFVAGVADDNGFGWAIAKQLAEAGAEITLGVWVPALNIFESNYRRGKFDASRKLSNGSLMEFAHIYPMDAVFDTVDDV	Cyanobacteria, NP_440356	**P**
7	>AY616073; [Acyl-carrier protein] S-malonyltransferase THVEQQERDGGHPAHEVCCWAAKGGRHRRARTVDRPAGPWPPRPTASLTVTSPTRPSWFPGQGAQSVGMAGELVKAVPAAAHLFDQASDILGYDLLKICLEGPKEKLDSTAISQPAIYVASLAAVEKLRAEGGEEALDAVDVAAGLSLGEYTALAFAGAFSFADGLSLVALRGRSMQAAADAAPSGMVSVIGLSAPAVDALCEAANAEVAPEQAVRIANYLCTGNYAVS	Cyanobacteria, ZP_00111994	**P**
8	>AY616076; 3-deoxy-7-phosphoheptulonate synthase (DHAP synthase) [d] AEPAAVPFLSSADHLQEWGPQSWREHEALQQPNYPSKDALDEAVSHIASFPPLVFAGECRTLQERLAAAAAGRAFIVQGGDCAEAFSQFSANRIRDLFRLLLQMAVVLMYGGGVPVVKIGRLAGQFAKPRSADLESIGGESLPSYRGDIINGPEFTAAARVPDPQRLLRGYNQSAATLNLLRGFATGGYAALDRVMQWNLDFTRGSDEGRAYRDLAARIEESISFMRACGLDMHSSIMTETDFYVSHEALLLEY	Proteobacteria, ZP_00267844	**P, C**
9	>AY616078; prephenate dehydratase RKAVATGTSRGLAANPSSSNPAPFPVGMSTILRSTLTAANAEAGPSSMLQMSTGFRPISRVAYQGAPGAYSEMAALKALPNWEPMPCEQFEVAFQALSQWLAERAVLPVENSLGGSIHDVYDLLLHYRLHIVGEVSVVVNHCLLALPGV	Chlorobia, NP_662549	**P**[#]
10	>AY616079; tryptophan synthase, beta chain PPTGSRRAFSAGVTTRARPVRCQAAMASPPTDTKAEDLGIYSTINRPDATGRFGKYGGKYVPETLIVALQELEVAYAEAMRDPAFIAELDELLKVYVGRPSPLYHADRLSEHHRRADGSRPEIYLKREDLNHTGAHKINNSLGQALLCQRMGKKRVIAETGAGQHGVATATVCARAGLQCVVYMGTKDMERQALNVFRMRLLGAEVRPVASGTATLKDATSEAIRDWVTNVETTHYILGSVAG	Cyanobacteria, NZ_AADZ01000001	**P**

11	>AY616082; ketol-acid reductoisomerase (acetohydroxy acid isomeroreductase) SDLRTGSHGNHNLPPSVRPAPAPPAALPSLWLAPGFAAIVRPVSLRAPRSRTSVSVHAAVKYDYDTKVFQKELVKFAGTEEYIYRGGRDKFNKLPEAFKGIKQIGVIGWGSQAPAQARNLRDSFAEAGMDTKVVIGLRKGSPSETEARAVDFTEQDGTLGEVFDVISSSDFVILLISDAAQAKLYPRILAAMKPGGHLGLSHGFLLGVMQNDGAD	Proteobacteria, ZP_00038347	**P**
12	>AY700206; Glutamate dehydrogenase (NADP+) VDSKLQKLVKGIQNDGLRSILEDVYARNAEQDEFLQAVRGMVVSIAPVFEKHPEYLTAFKLLTEPERAIIFRVPWVDDAGQWQMNRGMRVQFSSAIGPYKGGLRFHPSVTLSIIKFLGFEQIFKNALTTLPMGGGKGGSDFDPKGRSEAEIMRFCQSFMTELYRHIGSDVDVPAGDIGVGGREIGYMYGQYKRITRDRGGVLTGKGMGWGGSEI	Firmicutes, ZP_00313337	**P**
13	>AY700209; Histidinol-phosphate transaminase ADGMKSTPLTQCQAGEARLAQRPVVARAVKSGASFVRPHLLALAPYKPIEPLDVLSARLGRPASEIVKLDANENPYGPPPEVRAALADLEYPNIYPDPECRALREAMAQWYGIPKEHFMIGAGADEIIDFLMRCVLEPGDCIIDCPPTFTMYVYDAAVNAARTVTVPRLDGFRLDIKGIKRAVEEFNPKILFLTSPTTQTGASSQTETCWSCWSCRCWWCLTRH	Firmicutes, ZP_00330909	**P**
14	>AY616095; adenylosuccinate lyase ISCSSATAQSAVASNGKTQTLPLNDLTALGPLDGRYGSKVAALRATFSEYGLIRFRVLVECQWLKTLSKLEGVPEVPAFSSEAEALLDDLALNFSVEDAAEVKQEERTTNHDVKAVEYVIKRRLGKNAELSKVLEFTHFACTSEDINNLAHALM	Proteobacteria, ZP_00263521	**P**[#]
15	>AY616103; thioredoxin m[e] AGRTSNWLKVSGRIPRSAVFRHTLQASRSRPSISVLASESVAAQGAGVVNDETWEEVVLKSPVPVLVDFWAPWCGPCRMIAPLVDELSVEYGDKLRTLKLNTDESPTVATDYGIRSIPTVMLFKDGTKLDAVIGAVPKATLKQAIDKHLANQGN	Cyanobacteria, NP_484096	**P, C**
16	>AY616110; porphobilinogen synthase (delta-aminolevulinic acid dehydratase)[g] AGSHVFTHAFQTRRPVSSVRGHLGRAPRVRAIAEVEWTDPPG**PPKA**RV**PEGTPVV**PTMDLPSRPRRNRRSESVRRAFSETVLRPANFILPVFIHDGDASIPIASMPGVSRLGWKTGLVEEVARARSYGVNAIDLFP	Cyanobacteria, NP_892336	**P, C**
17	>AY616112; uroporphyrinogen decarboxylase TATSKLKCGQTACASGPEPLMVRASRGETVERSPCWMMRQAGRYQKAY	Cyanobacteria, ZP_00177900	**P**[#]

RDLALKHPSFRERSEDVGLIVEISLQPWESFKPDGVIIFSDIMTPLPAMGIPF
EIDDARGPRIADPIRSREQVGQLHALDLDQLSFVGDALRALRATVAGQAA
VLGFVGSPWTLATYIVEGGSSLTYKTIKSLCYTARGCWRPC

---

[a] The sequences in **red** or **blue** represent the estimated length of truncated N-terminal extensions. Sequences in **red** represent an estimate based on the cleavage site predicted by TargetP in the best-BLAST hit proteins; the accession number for these proteins can be found in the third column of this table. Sequences in **blue** represent N-terminal extensions with respect to the N-terminal methionine from eukaryotic cytosolic isozymes and/or from their prokaryotic/cyanobacterial homologues; the accession number for these sequences can be found in the third column of this table.

[b] The intracellular localization in land plants and algae of the proteins which are similar to those identified in *P. wickerhamii* was compiled from the Brenda (http://www.brenda.uni-koeln.de), TAIR (http://www.arabidopsis.org/index.jsp) and Gramene (http://www.gramene.org) databases, and from journal articles related to the best-BLAST proteins. The Brenda database provides comprehensive information about enzymes for which the intracellular localization was determined by experiments such as immunolocalization and subcellular fractionation. For enzymes for which this kind of data is not present in the Brenda database and for proteins that are not enzymes, we used the other aforementioned sources. In this situation the abbreviation for intracellular localization is marked with #. Abbreviations: P, plastid targeted; C, cytosolic.

[c] AY616066. Two structurally distinct types of acetyl-CoA carboxylases (ACC) are found in nature. Bacteria and most plant plastids contain a multi-subunit ACC enzyme (MS-ACC) while the second type of ACC, a single large multifunctional polypeptide can

be found in the cytosol of plants, fungi and mammals. We clearly identified the alpha subunit of MS-ACC, the plastid-localized enzyme. Acetyl-CoA carboxylase alpha subunit from *P. wickerhamii* shows the highest similarity with acetyl-CoA carboxylase alpha subunit from green alga *Chlamydomonas reinhardtii* (cluster 20021010.1560.3 retrived from http://www.biology.duke.edu/chlamy) (BLASTP e value $6e^{-68}$), which has a plastid-targeting peptide predicted by ChloroP and TargetP.

[d]AY616076. We identified only one form of 3-deoxy-7-phosphoheptulonate synthase (DHAP synthase) in the green alga *C. reinhardtii* (BLASTP e value $1e^{-86}$) but in land plants there are several. We found two isozymes in *Arabidopsis thaliana* (accession numbers AAL91255 and AAK32811) while in *Oryza sativa* we identified four different forms of the DHAP synthase (XP_483024, XP_506482, AAR06362 and BAD14927). However, all the isozymes from *A. thaliana* and *O. sativa* as well as DHAP synthases from other land plants have N-terminal extensions relative to prokaryotic/cyanobacterial sequences; TargetP and ChloroP predict that all these isozymes are plastid targeted. DHAP synthase from *C. reinhardtii* (cluster 20021010.959.1 retrieved from http://www.biology.duke.edu/chlamy) has an N-terminal extension and the prediction is plastid (ChloroP) and mitochondrial (TargetP).

[e]AY616103. Thioredoxin m represents the plastid–located isoform while thioredoxin h represents the cytosolic one. Thioredoxin m from *P. wickerhamii* shows the highest similarity with thioredoxin m, CAA56851, from *C. reinhardtii* (BLASTP e value $8e^{-39}$). The BLASTP score for thioredoxin h, CAA55399, from *C. reinhardtii* is significantly lower (BLASTP e value $8e^{-15}$). Due to the high difference on BLAST e values between the two forms of thioredoxin, the output of the BLASTX and TBLASTN search in GenBank with thioredoxin m from *P. wickerhamii* will not list the scores for any thioredoxin h.

[f] AY616110. The first methionine in *P. wickerhamii* porphobilinogen synthase (delta-aminolevulinic acid dehydratase) corresponds to the N-teminal methionine in NP_892336 and to other sequences from cyanobacteria (highlighted in red). However, the PPKA and P(X)GTPVV sequences that are located upstream of the aforementioned methionine are conserved in *P. wickerhamii* and the best BLAST-hit CAA52955 (the moss *Selaginella martensii*). Therefore, it is unlikely that this methionine represent the N-teminal methionine in *P. wickerhamii*. Also, in CAA52955 the N-terminal extension is 30 amino acids longer than that of *P. wickerhamii*; for AA52955 all four programs used to determine the plastid targeting sequences predicted the presence of a plastid transit-peptide.