

Supplementary data

Table S3. Putatively plastid-targeted proteins in *P. wickerhamii* for which the N-half of the protein was not identified (N-half truncated clones). The sequences presented in the table cover the middle or the C-half of the mature protein.

	Protein	Comments	Intracellular localization in plants and algae ^a
1	>AY616041; heat shock 70 protein AVVSLVERLAKKKPNVTVPNPDEVVALGAAVQGGVLAGEVSDIVLLDVTPLSLGLETLGGVATKVIPR NTTLPTSKTETTFSTAADSQSSVEVNVLQGEREFARDNKSLGNFRLDGIPPAPRGVVPQIEVRFDIDANGIL SVTASDKGTGKSQDIKITGASTLPSDEVERMVQDAERFS		P[#]
2	>AY616043; protein disulfide isomerase HPARGHLLRRQMFDES DVVVLTVKNFEVELRKQKYALVEFYAPWCGHCKSLKPAYAAAATEVKKL GLPVVLAKVDATVEADLGSQFEVRYPTLKWVNGKVASDFGGGRTKDDIVRWVSKKSGPATTDLE SVAALET	See ^b	P[#]
3	>AY616044; poly(A) binding protein RB47 homologue LDRDVTEAQLYEIFAQIGPVASIRVCRDAVTRRSLGYAYVNYNSGVDAGAAERALEQLNYVPLGRTP HAADVEPPRPSYRKSGVGNIFIKNLGRLPLTRPWFDTFSTFGNILSC		P[#]
4	>AY616048; SecA-type chloroplast protein transport factor – related SQLRCAFLAIKAEFKA VTDVERTRVVALGGLHVIGTERHESRRIDNQLRGRSGRQGDPTTQFFLSLED NIFRVFGGDRIQTLMSALHIGDLPIESKMLTGALDGAQRKVEAYYFDARKQLFDYDQVLDIQRGRVYE ERRRALLAPDLTPVMLDYARKTVDDILEANVDPKTSVDSWPIEALASKMVQYCPHIDFGPECLWQA ARTKGYAGLRALLWRRGEAA		P[#]
5	>AY616053; phosphoglucomutase EAVQIILKLAAGNGLQKVVVGRNGITATPAMSAIIRRRGLYGLIMSASHNPAGPNEDWGIKFNYSSG EPAPEKITDKIYGFTQSIKELQVADIPDVLDSVLGTTTTFGDFEVEVIDYTVDYFNQLKEVDFEALRTLL ASPDFSFVYDALHAVTGAYAEPFLVDELGASPDLSKNCVPLEDFGGGHP	See ^c	P, C

6	>AY616062; starch synthase isoform SS III VKLEPLSEAVVVKRNGLSPGSWVGAKVDIDKLAMLVDFVFSNVARTTWDNNDSDADYHSLVSGALKDK ELGLVLASDAMARDAGEAAKQEDLAVKRAVEKAEIKYTSAKQRRAVQAAFLYTRPLTPRAGQAVEV FYNPDLTPLRGRPEVYVRGGFDRWTRNTFSPQAMRATGTGGFVTTTIQVPRTAHVMDLVFMDSGDA HGGFIDNRGLDYHMPVAGAPGKL	P
7	>AY616087; cystathionine gamma-synthase (O-succinylhomoserine (thiol)-lyase) PKMGISATVIDPADLAGLEAALEAHPVALFFSESPTNPYLRCVDIPRIKALCEPRGAAVVIDSTFATPIN QRALALGADLVLHSATKYLSGHNDVLAGAIAGRSDLVASVRAMHNVLGGVMDPHAAAYLLLRGMKT LDLRVGRANATAQLLAERLSEHPRVARVHYPLPSHPDHIALRQMDGYGGVVSFEVAGDLW	P
8	>AY616090; L aspartate oxidase AHPLPGRGHHAGHGGVVRGSCTPLTTNPGVATGDGMAMAARAGARLGGLEFVQFHPTVLVAPAAAG AGGHADADRAFLISEAVRGEGRLFNVDGERFMNRYDPARRELAPRDVVARAIHAEMASTGRDHVL LDVSHLAREEVLSHFPTIAARCAALGLDLTRQALPVAPGQHYMCGGVATGLHGETSLSGLFAAGEVA SSG	P [#]
9	>AY616092; ferredoxin-dependent glutamate synthase LIRADGGMRTGRDVLVAAALGADEYGFQTVAMIATGCIMARVCHTNNCPVGVASQREELRKRFPGT PEDLVNYFMFIGEEVRASLASLGLRSLDDLIGRGNLYRQRSDVTLAKTASLDLSILTRYAGDCARSSTR RTASPHDNGSDWDDVILADPEVQAAIAGQGTVARSYTIVNTDRAALGRLGGAIARLHGDDRFEGR IDDLRGSAGQS	P
10	>AY616098; ATP phosphoribosyltransferase. VAREFFASRGFEHVKLLSADGALEAAPAMGFADIILDLVSSGVTLRENNLREIEGGRVMASQGVLVAS RAALLERPGLLELVHELL	P [#]
11	>AY700210; Histidinol dehydrogenase KEVRAAFDIARANIAAFHAAQASPDLVHETMPGVSCSRVTRAIGAVGLYVPGGTAVLPSSALMLGVP AGLAGCRTIVLATPPGPDGSIPAEVLYCARAAGVTHVLAAGGAQAVAALAWGTASCPKVDKVLGPG NQYVTAAKVALAAGEAMVAIDMPAGPSEVLVADAGADLVHVALDLLSQAHEGPDQSAILVALPGL DVAALRAEVARQCDALP	P
12	>AY616102; ferredoxin-NADP reductase HEQLESIGLPLNTFNRRKPFARIVSVDRIVGPKATGETYHVVLETRGEIPFAEGQSYGVIPPGSKVNSK KKEVPHGTRLYSIASTRYGDFFDGRTASLCVRRATYWDPETGAEDPAKKGICSNFLCDAKPGQEV	P

13	>AY616108; hydroxymethylbutenyl 4-diphosphate synthase, similar to <i>E. coli</i> gcpE gene product LHRDGSVLSALSLEDLHKPDATYRSLGCKMAVGMPEFKDIATSDSVYLTEVPAVDDAVARRALRRLQE VGVHVLAEADALVASPLPDSIAVVSLAEAVAAAREGRSLLPPGAVRLALAIIDGTESEAEAAATGGLDA TLALLRTAPGLSRVHASRRVFEALARAHCTLPVIHALAFQAGTGREALVLAAGALVGALMVDGR	P [#]
14	>AY616109; glutamate 1-semialdehyde aminotransferase RHQNPKILDRPGVYEQDRVTNVLVAGLLQAGRETGHDMCGGSINAMFGFFFAKGPVTNFEDAKTA DGAKFAKFHRGMLEHG	P
15	>AY616111; uroporphyrinogen III synthase SESCYEHTNNASVPRAGRLVRRKCDSEFARWSSRRLVVLTRHEGKNEDESLLCARGIESLKLPMIET VDGPDRLLLGSFLCDEDFEWICLSSPEAAAVFMRAWQEVGQPPPIAAVGSQTARVLRHLGGPECIPGF MPSVATAEHFAGELPWGKGRNLNRVLYPTSARAXQSLQWILEGRGFAVTRLNTYDTIPVKHVDADFLE QARNAAMVAISSTSAGRSWISTIGQ	P [#]

^aThe 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; M, mitochondrial targeted; C, cytosolic; ER, endoplasmic reticulum (i.e., microsomal fraction).

^bAY616043. The best blast hits are for the protein disulfide isomerases from the algal species *C. reinhardtii* (accession number AAD02069), *Volvox carteri* f. *nagariensis* (accession number AAD55566), and *Helicosporidium* sp. (AAU07697). All these proteins

have N-terminal extensions predicted to represent a signal-peptide (ER-localization) by both TargetP and iPSORT. ChloroP and Predotar, which can discriminate only between mitochondrial and plastid transit-peptides also failed to identify these extensions as plastid transit-peptides. Moreover, the protein disulfide isomerases from the algal species *C. reinhardtii* and *V. carteri* and from land plants possess the C-terminal ER-retention signal KDEL (for *Helicosporidium* sp. protein disulfide isomerase the C-half of the protein is not available in databases). Evidence for plastid localization comes from *C. reinhardtii* protein disulfide isomerase, for which several studies demonstrated that it is a regulator of the plastid translational activation (27, 54). However, based on the abovementioned structural features, a dual localization, i.e., in the plastid and ER, was also suggested for this protein (54).

^cAY616053. For the enzyme phosphoglucomutase, the best BLASTX hits are for the plastid isozyme (bit score 263, e value $2e^{-69}$) followed by the cytosolic isozyme (bit score 232, e value $3e^{-60}$), both from land plants. The alignment of the 187 amino acids from *P. wickerhamii* with plastid and cytosolic land plant isozymes, *C. reinhardtii* and cyanobacterial enzymes revealed that the cytosolic isozymes from land plants have three insertion of 6, VG(A/V)DGS, 1 and 4 residues (S/T)GPE, respectively, that are not present in *P. wickerhamii*.