Supplementary data

Table S1a. Putative plastid-targeted proteins in *P. wickerhamii* for which the full-length targeting-peptide was identified.

	Protein[a]	**TargetP** [b]	**ChloroP** [c]	**iPSORT** [d]	**Predotar**[e]	**Intracellular localization in plants and algae** [f]
1	>AY616038; 60 kDa chaperonin alpha subunit MMAQPFTGSRIQSRPIRGGRAGRGARQVVRAEAKDILFDNESRRKMQVGIN KIADAVAVTLGPRGRNVVLEQAYGTPQVINDGVSIARAIELADPVENAGAQ LIKEVAGKTNDTAGDGTTTATVLAREMIHLGLQVVTAGANPVSIKRGIDKA CTYLVGELAARARPVTGSDDIRAVAAISAGNDDAIGAMIADALEKV **Note:** The estimate of the length of the N-terminal extension is the result of multiple alignments; the first residues after the sequence in **blue** represent the first stretch of conserved residues among land plants and algae, including *P. wickerhamii*.	**M**	**-**	**P**	**M**	**P**[#]
2	>AY616039; 60 kDa chaperonin beta subunit MNSATLTCGPSTSGLVGGHGAFQKARSGSRAAVRRSVQTRAAKQLHFNTN GEALKRMQAGVDKLASVVGVTLGPKGRNVVLETKYGSPKIVNDGVTVARE VELEDPVENIGAKLVRQAASKTNDLAGDGTTTATILAAGIIQEGMRIVAAGV NPVQVIRGLERTVTHLVAELAALAKPVSDAELKDVATVSAGGSAAIGELIA AAMARVGRQGVVTMEES	**P**	**P**	**P**	**P**	**P**[#]
3	>AY616045; catalytic subunit of ClpP5 protease MQVLGLAQPDASPTFDPTTLTAKAFYVGDAVTESSRQSTRMAQDTRAPGR APPDLPSLLLDGRICYIGMPLVPAVTELVISELLWLNYSSPTKPVYVYINSTG SQREGGESVGFETEATAIMDTMAYIRPDIYTLVIGQACGNAAMLLASGKKG HRFALPHS	**-**	**P**	**-**	**-**	**P**[#]
4	>AY616049; similarity to thylakoid lumenal 29.8 kDa protein from *A. thaliana* MLSSKVFANACDKFVRHGGDCKRAFPSNTPRNVRVAAAGYDAAQSGTSN DSLDGFEVGRRQALAALVTVPAILQANAALAVQGLTAGRIPGVSSTADADG FYRYTRPEGKSGGHGVGWSEMPRYTFKVRPGWEETPVSIADLGGTEIDLRF ACPEEGSLQVVVAPVLRFADVGYNADVRITDLNKPENIISGFAPELFGRPLD DEDVLEQRVLEKEGVPY	**-**	**P**	**-**	**M**	**P**[#]
5	>AY616050; putative plastid ATP/ADP-transporter	**P**	**P**	**P**	**P**	**P**[#]

	MQLAPVCTATLPVASCRPRLTACRPIARAACLSTSKSAMAAQPLMAKPMGMASLPSQQRVKSASVRVEASNSAATPAPKEGSFLGIAAFTWQKIIPLGLMFFCILFNYTILRDTKDVLVVTAPGSGAEIIPFLKTWVNLPMAIGFTILYTKLANVLNNEQLFYVCIIPFLAFFSAFAFLLYPCATSCTPPSGPPRRSPSMAP					
6	>AY616056; pyruvate kinase, similar to isozyme G MASPCSSLRPTRLLGPSGSQLAYWPDHTPDCMHRDPPPSRSNGKTDSSRPRASTSTPALPGYAKGVSLSSSNAKMMAEDPASEASSLDIPTRKTKTVCTIGPTSCDREAFFRLADAGMNVVRLNMSHGDHASHQQVIDLVREYNALGRRNLAIMLDTKGPEVRSGDLTQPLDLEKGDLITFTIVAGADGTNNRIGVNYDGFIDDVEVGDMLLVDGGIMTMLVKSKSDTEVLVEVLGLGHEISRHLTFGEVR	**P**	**-**	**P**	**-**	**P, C**
7	>AY616059; dihydrolipoamide S-acetyltransferase MQASSAVMMQCTSVLAGRPTRVSLPRGPGGRRILRPLSAVKDVFMPALSSTMTEGKIVSWLKSPGDKVAKGESIVVVESDKADMDVEAFAEGILGCITVPEGGVAGVGSAIAYIAETEADLEAAKAKGDSSAGTTAAAPAPVAAAV	**P**	**P**	**P**	**P**	**P, M**
8	>AY616060; pyruvate dehydrogenase E1 beta subunit (acetyl-transferring) MAALVGSRISRTCPGASKSQTSTLSPACIFIRSNVAPKGRLSAGRMMIKTHAKKEMMMWESLREGLDEEMERDPNVCLMGEDVGHYGGSYKVSYGLHKKYGDMRLLDTPICENGFMGMGVGAAMTGLRPVVEGMNMGFLLLAFNQISNNCGMLHYTSGGQFKVPMVIRGPGGVGRQLGAEHSQRLESYFQSIPGVQLVAVSTARNAKGLLK	**P**	**P**	**P**	**P**	**P, M**
9	>AY616061; small subunit of glucose-1-phosphate adenylyltransferase MSELPLSPLRTAPSARCGVHSLSQDAPDWTGRLARPAYACRSLPKSHAFTSPPRQSKAVQPRATAQLDRDVASILNLTNGSPDPRPADISRTVLGIILGGGAGSRLYPLTKQRAKPAVPLGANYRLIDIPVSNCINSGVNKIYCLTQFNSASLNRHLSQRYNSYNSNVGGYNARGFVEVLAASQTTTQKNWFQGTADAVRQYLWLFEEATRGGVEDYLILSGDHLYRMDYSEFVTAHRRAGAEITVAALPCTEKEAGAFGLMKIDDSGRILDFAEKPKGEELRGMQVDTTLLGLDAESAARKPYIASMGIYLCKASAMHHLLTDRFATQNDFGSDII	**P**	**P**	**-**	**-**	**P, C**
10	>AY616063; 1,4-alpha-glucan branching enzyme MQLASAVRPRGVFGGTRHSARVHGIVTAGAAPDGLGSPIELARRENAALKRSIDQAETSVSKLESGLASAGVPIPPLSTTSSPDSSARKNGSNGDAVEAAEPASDAGPLEPEDYWSPVQHVPDGVALKETLGPITPVPDHDGTDCLKWDASLWSHADHFKYRWHIFKSIRAAIDANEGGLEQFSQGYKYYGINRGEHEGK	**M**	**-**	**M**	**M**	**P**

	Note: The estimate of the length of the N-terminal extension is the result of multiple alignments with sequences from land plants; the first residue after the sequence in **blue** corresponds to the N-terminal methionine in proteins that lack the N-terminal extension.					
11	>AY616065; biotin carboxylase, subunit of acetyl-CoA carboxylase MAYVAPPGQLRASAGWVPSATQPSMRVAAVRPARRSVARMTVRAEGAGNGVAAAARTPITKVLIANRGEIAVRVIRACREMGIKSVAVYSVADRDCLHVQLADEAVCIGEASSAESYLNVPTIIAAAMSRGADAIHPGYGFLSENASFVDICKDHGIEF	**P**	**P**	**M**	**P**	**P, C, ER**
12	>AY616069 3-oxoacyl-[acyl-carrier protein] synthase MATQSLAFPRVGSSTHCFGRREPRNRRHVTQHVSCEQSSSAGGGKPKRRVVVTGQGVVTSLGQKTKQFYEKLLAGTSGIAPIEGFDTTGYSTTFAGEVKSVDAGPYVARKWVKRMDEVMKFMFVAGKQALEDAGLPFEGPDLDQLDRALC	**P**	**P**	**M**	**-**	**P**
13	>AY616070; 3-oxoacyl-[acyl-carrier protein] reductase MLRHLQVPRALLRQHGSCARSTFAARPSTARRSTVNVQVRADSAVAEAPAKSVQQGLEEGVTLVTGSSRGIGRAIALRLGAAGGRIVVNYAASSKAAEEVAHEIEALGGSAIVVKGDLAKVSDAQALVKAAVDKWGTLDALVNNAGITRDTLMLRMKPEQWQQVIDVNLSGVFYATQAAVKVMAKKRKGRIVNISSVVGLTGNAGAGQLLSGSKA	**M**	**P**	**M**	**M**	**P**[#]
14	>AY616071; acyl-[acyl-carrier protein] desaturase MALTSVRPAFHAMSADLRRGGPSASAPRPSRASLRVRAAIASEAPVAAMSKASPGPTVYSKLDRAHTLTPERMELINGMADFAEERILPVLQPVEKLWQPQDMLPDPESEDFLDQVAELRARSEGVPDAYYVVLVGDMITEEALPTYMAMLNTLDGVRDETGAADHPWGRWTRQWVAEENRHGDLLNKYCWLTGRVNMKAIEVTIQNLI	**P**	**P**	**P**	**P, M**	**P, C**
15	>AY616072; beta-hydroxyacyl-[acyl-carrier protein] dehydratase (similar to FabZ from *E.coli*) MASVCVANPRAAHSLPRFQVSSSVSRARCVPARSRRSNLAAQANGVASSAATESLAVVKSGPNMKALKDIQEIMEILPHRYPFLLVDRVLEWETGKYAVGYKCITVNDNFFPGHFPTRAIMPGVLQVEAMAQLAGIAMIDPADRGQQNNFFFGGVDGVRWKKPVVPGDVLMMRVDVIKFNKRHGICKVDAKAYVGTDLVCQGELTLVMLK	**P**	**P**	**M**	**P**	**P**[#]
16	>AY616074; 1,2-diacylglycerol 3-beta-galactosyltransferase (MGDG synthase type A)	**-**	**-**	**M**	**-**	**P**

	MGGSTQYERHLCASSITMTSLPLASSAGRVCEPTLVASAPSWDWKFWKTGR DGLKPFLGPDASGRGTGKVALPFKPGQGKCSSGPGIRLSNSGLRLPRLAMG GGSGHGAGAYSFFMIQAPRLWPVSWRIPPKKGSAGEDKQRILILMSDTGGG HRASAQALHSGFQELYGDKYKIDVLDIWTNYTPWP **Note:** Multiple alignments with sequences from land plants revealed a weak conservation in the first 120-130 amino acids; the N-terminal methionine in proteins of prokaryotic origin (Firmicutes; YP_080396) is located downstream of residue 130 from *P. wickerhamii.* The sequence in **blue** represents a rough estimate of the length of the N-terminal extension based on the average of the cleavage sites predicted by TargetP for four land plant sequences.					
17	>AY616075; glycerol-3-phosphate O-acyltransferase MAAAMGCVAFPVGPISTHRVRHHSRIQRISQNAMSRRSMRPRELKTLASAA PATAYVSPFSKVEHEEQLFAILKAGTASGKVPKRLVDGITELYGNYKKAVV GSNKPGATADFVAKVMASVCERVMIQITSPFVFQSKHERMLEPYNYYEFGQ RYIRGLVDFDTSRLGHADRFATIQEQLARGENVVLLANHQTEADPAIFALL	**P**	**P**	**P**	**P**	**P, M, C, ER**
18	>AY616077; 3-dehydroquinate synthase MKGVCPNSGLGPVTCGSRALPRPFVPTPSRHSPFLSGPSPHAGVTQPLRRQT RRGHSLKAKAASAPPAPPPAMETVEVGLGTRTYPIYIGRGLLRQGELLRRHI VGSSALIVTNQTVSKLYLDATLAALRGGNSGLRVESVVLPDGEQNKTMEVL QQVWTAALEARLDRNTTFVALGGGVVGDMAGFAAAAYQRGVAFVQIPTT VMAQVDSSVGGKTGVNHPLGKNMIGA	**P**	**P**	**P**	**P**	**P**
19	>AY616080; acetolactate synthase, small subunit MQLTGAICQGAVVARPSSRLAFHAAHAPRCLTTPPRVHGTRVPARAASAVA RASSETTDSPATSASEGEGYDVYVYEAEPLEEGVSKHTLNVFVGDEAGMIN RVAGVFARRGANIESLAVGLTQDKALFTIVATGTDATVANLCKQLAKLVN VRYVEDITSSEHIQRELFIAKVAAPPGVTRTEIQ	**P**	**P**	**P**	**-**	**P**
20	>AY616081; acetolactate synthase, large subunit MLGLCQLSTCERKPRNPGAVKSQTGLFGAAPSTSRAARHALARRVVRGTPP TAALARDQAGPDSPALAKAREAAHASLASPPPKEWVDRFGTAPRKGSDILV QALEREGVDVLFAYPGGASMEIHQALTRSHTVRNILCRHEQGEVFAAEGYA KVSGRVGVCIATSGPGATNLVTGLADALMDSVAMVAITGQVPRSLIGSDGF PGDAHRGGD	**P**	**P**	**P**	**P, M**	**P**
21	>AY616083; 3-isopropylmalate dehydrogenase	**P**	**P**	**P**	**M**	**P**$^{\#}$

	MSALAPSTRIGVGAPARRAFAARSRCPRARPALATCAAHRVTVLPGDGIGPEITAVTLSVLEAAGKAEGESFTFTEALIGGAAYDATGDPYPDATYRACADSDAVLLAAIGGYKWDALPSVSKPETGLLRLRSSLNAFANLRPATVIPELADASSLKREVLEGVDLLIVRELVGGIYFGQPRGFEIRNGEKVGFSTDIYSEFEVERIAHVA					
22	>AY616084; glycine hydroxymethyltransferase MQAIASSQPRFAGIRAETCGPGRQVQRLASRPNQARLVTPAQSRVVCLAATLERQATGSMFFEDGSLDEVDPEIASIIRKEKVRQVTGLELIASENFTSRAVMTAVGSCMTNKYSEGLPGARYYGGNEFIDQAESLCQRRALEAFGLDPAEWGVNVQPHSGSPANFAVYTALLSPHDRIMGLDLPHGGHLTHGFQTPKRRVSATSVYFESMP	**P**	**P**	**P**	**-**	**P, M, C**
23	>AY616085; 5,10-methylene-tetrahydrofolate dehydrogenase and 5,10-methenyltetrahydrofolate cyclohydrolase, bi-functional protein MSALCFGLQNRIPLRSPTERRSPRTVVAHPPQRRVSLTLRGSVVSAKRAVATASAMPAALIDGKKIAETIRGELKVEVDELRAQYNTVPGLAVVLVGERKDSQTYVRSKKKACDEVGINSFGADLPETATEEEVLAVVAKFNDDPNVHGILVQLPLPEHINEKRVLDAISLEKDVDGFHPNNIGELAMRGRSPTYVSCTPKGCLELLKRAGVTLAGKRAAVVGRSNIVGLPAALLLQQEDATVT	**P**	**P**	**P**	**M**	**P, C**
24	>AY616086; cysteine synthase (= O-acetylserine (thiol)-lyase) MGGSHSKKSGATTDPLVTVLENKPSRSQATPPADTAPPPAKKNIVSSNITQVIGNTPLVRLNHLAETEGAKATILLKMESLEPNSSVKDRIGLAMIEDAEQSGKITPGKTTLVEPTSGNTGIALAFVAAARGYKLVLTMPATMSLERRIVLRAFGAQLVLTDPAKGMKGALLKAEEIAAATPDSYILQQFENPANPAIHYTTTGPEIWNDAGGDVDVFIAG	**P**	**P**	**M**	**-**	**P, M, C**
25	>AY616088; aspartate-semialdehyde dehydrogenase MPLSLSSSRCPAVALGPRRASVTRASRSGPSKSRCTRSKDGVRAYIAIAGVTGAVGQEFLRVLTERNFPYSNIKMLASARSAGRQYEFEGETYTVEELTDKSFEGVDIALFSAGGSISKKYAAVASAAGATVVDNSSAFRMTEGVPLVIPEVNPEAMAGMKAGKGGIIANPNCSTIIALMAVAPLHRLAGVKRMVVSTYQAASGAGQAAMEELEL	**P**	**P**	**P**	**P**	**P**
26	>AY616089; threonine synthase MWGASRAAVERCAPAHRHLPPRGTSGVTLSCVPARRAPSGPSTCPSSPWLPAPRAPRRGRPTPWDDVVYRAKDGGLLDVQHDFEALAKYDAKYWKTLFDGRVGTARWPHGSGVWSKKEWVLPQISDDDIVSMFEGNSNLFWAERFGRQIGMSDLWVKQCGNSHTGSFKDLGMTVLVSQVNPAEQDQA	**P**	**P**	**M**	**M**	**P**

27	>AY616091; glutamate-cysteine ligase (gamma-glutamylcysteine synthetase) MSCMLEKHWDAGSAAPSVSTTRPHRSAVPEVETGPVPPINRVLFYIHIMCTDKVATPVGAAPHLTLEDLVENLRKGCKPRSSWRIGTEAREVGVRLRLTPCPATHQQISDLLTVIHDSNGWEYIKEGEAKIGLAHANETVTLEPGGQTELSGAPLKDLHAVAQETKEHLIQVRAACDKVGIDYMNIGFDPKWGFEDVPKMPKTRYRYMREYMPKVGTLGHDMMFRSCTIQVNLDFESEADMVEKFRVSLALQNVAGALFANSPFRDGKPTGYKSWRLNVWTDVDNARCGRLPFVFDADFSFARYAEWALDVPMYFLYRDGEYHDVGAQPFRDFFHGKLAAFPGLFPTMDDWDVHLTTVFPDVRLKRFLEMRGADGGGWEFITALPALWVGLLYDAGAQAEAARLVADWTPEELDALQADVPRQALQATFRS	**P**	**-**	**-**	**-**	**P, C**
28	>AY700207; acetylglutamate kinase MSALLPGTHLVSRPTKGWTPVRPARNVRPAGMGVSRAPMRPTICQSSTVAADTGTKLERFSPLDRVSVLSEALPYLQRFRGKTVVIKYGGAAMKDPTLKAGVVADLVLLATVGIRPVLVHGGGPEINIWLTKLGIEPNFKNGLRVTDEATMDVVEMVLGGRVNKSLVTLIQQSGGRAVGLCGKDSDIILARQMVEKDIGFVGEVSGVR	**P**	**P**	**P**	**P**	**P**$^{\#}$
29	>AY700208; argininosuccinate lyase MAAMACSTHIGAQLPRQTSGSRVGSRASTISAPSSRTRTLRCTAAAAPADAPAAKKLWGGRFTGATDPLMEKFNESLPFDRRMWAEDIRGSQAYAKSLARVGVLTEEEASTLVDGLAAVAEEWAAGTFKVVPGDEDIHTANERRLSELVGPVGGKLHTGRSRNDQCVTDTRLWLTGAVRELRASLQALIG	**P**	**P**	**P**	**P**	**P**$^{\#}$
30	>AY616099; Phosphoribosyl-AMP cyclohydrolase. MATAVGIASPAGRLRTVQAGAFCRATGVGHHTSVGRRAGAVGRVERAATLAPCSAGSTPTPLASLPSLKYDRQGLIPAIVQHVDTGEVLMQAYSDQAAILETLETGLATFYSRSRKGRWCKGETSGHFIKVLSVHMDCDRDSLIYLGSPIGPACHTNAPTCYFTALVPEAGGVVAEKGDSTSRAQAPLPSLFALCPAAHHRRAQGPRPRCQVRGARPWRVGGPGCWVAGQTGWGSGPGLG	**M**	**P**	**P**	**P**	**P**$^{\#}$
31	>AY616093; phosphoribosylformylglycinamidine cyclo-ligase MSAASLLRQAACQRTAGEPSCSYAAPLHRVFCTPRRLMGHRARAAHLIPRASGLSYKDAGVDIDAGNELVRRIQKMNPNIGGFSGMVPFGDSFLVAGTDGVGTKLKLAFDVDKHDTVGIDLVAMSVNDIITSGAQPMFFLDYFATGKLEVDTAEQVIKGIVEGCRQSDCTLMGGETAEMPGFYSPGEYDLGWLCGGVCEAEPRD	**M**	**-**	**M**	**-**	**P**$^{\#}$**, M**$^{\#}$
	Note: The estimate of the length of the N-terminal extension is the result of multiple alignments with sequences from land plants that have N-terminal extensions and					

	from prokaryotic (including cyanobacterial) sequences; the first residue after the sequence in **blue** corresponds to the N-terminal methionine in a proteobacterial protein (NP_274275).					
32	>AY616094; phosphoribosylaminoimidazole-succinocarboxamide synthase MAVQCSSARALRQTGPIPSAGAVRRLSRAHVRLCAVPQQQPRIRRCIMHAVPSAHVEAVAGLDASLTPIIQAALDNCLTETSLELPGEHYVGKVRDTYDLGDKLLIVTTDRQSAFDRHLASIPFKGQILNQTSAWWMEATRSIVPNALLGLPDPNASLMTKCTVFPVEFVCRGYMTGSTDTSLWTHYAAGERRYCGNDFPDGMVKNARLPVNVITPTTKSAEHDVPISP	**P**	**P**	**P**	**-**	**$P^{\#}$**
33	>AY616096; phosphoribosylaminoimidazolecarboxamide formyltransferase MTGVAAQVVASPSRCLFASSQRTSKTASTSRCPSSRRWASLTTPSAMSQTSPAAAAPTKVPKGDKGQVLVSVSDKAELIPFVQGLVELGYDIISTGGSATSLDAAGVPVERVENVTGFPEMLDGRVKTLHPAVHGGILARRELPSHMAALAEHDIDLINIVVVNLYPFRQTVTAAVKPAYEVAVENIDIGGPSMIRAAAKNHAHVLVVVDPADYAMVLERLRAGTADAALRRQLAWKAFQHTATYDATVAEWLWGEVGTGPAPERCVPMQLCSTLRYGENPHQAAAFYTDLSLGEAGRGGVATAVQHHGKELSYNNYLDADAAYGAVCDIAAPACAIVKHTNPCGVAARDDLREAYRLAVRADPTSAFGGIVAFNREVDADLARELREFRS	**P**	**P**	**P**	**-**	**$P^{\#}$**
34	>AY616097; adenylosuccinate synthetase MACLSVSPAASMAVCSSRGHAACRLSSRLVQLPRQCARKSSRPGAVAARGFGSVSPAQRLRAGASAAPASVKTEEFDPQVVVVLGMQWGDEGKGKLVDILAQKYEIVARAQGGANAGHTIYDTEGNKWKLHLLPSGILNPKATCVVGNGVVVHLPSLLDEIKGMEARGVSVKGRLHISSRAHLLFDLHKEIDARREEELAGT	**P**	**P**	**P**	**P**	**$P^{\#}$**
35	>AY616100; ferredoxin MALAMTPASLVWPPGRAPAQPSRSSPLVPARLAVIRSPVALLKHSVVCQSYKVTFQLENGEEETIEVPEDQYILDAADDAGLDLPYSCRSGTCSTCVGKVVSGELDQSDQSFLDDAQMEKGYALLCVAYPTTDCVIQTHKEEELY	**P**	**P**	**P**	**P**	**$P^{\#}$**
36	>AY616101; similar to thioredoxin y MGSLALASVGCRTSLPLQSTTRFGARASFHESRRQAPRTARRVSPCQRQHGIVARAGTSDLIQEVSARELEEAVQARAKPLVIDFYATWCGPCVLQAKELVKVAEQLGDGVQILKIDTDDNPDISNQLQIMGLPTVIIVPTDPSKPALRTEGLLSAEKIVGLINSSGSPATVAPSDVGRPACFEGRTCFTYERVLTTVIPALHGCSMHV	**P**	**P**	**M**	**P**	**P, C**
37	>AY616105; adenylyl-sulfate reductase (glutathione) MLTQALPLAHSVPVRCTGRFHGSKFLATPCQGTAACPPSPANPGSAHLRRG	**P**	**P**	**P**	**-**	**P**

	TLEAPVADIPARTLDWPSIAQNLDAKSPLEIMDHALATFGDTIGIAWSGAEDVALIEYAHLTGRPFRVFSLDTGRLNPETYRVFAAVEKHYGIKIEYTFPDAQETMDLVREKGQFSFYEDGHGECCRVRKVRPLRRQLSGLRAWITGQRKDQSPGTRQAVPV					
38	>AY616106; geranylgeranyl diphosphate synthase (farnesyltranstransferase) MGFSTMPAVSAPCLRAFTHGPRYMPGRPPRSSRNHLQSPRGTAEVSSPPSSANRVHPLSVFDFRVYMSQRAALVDAALDVAAPLQYPERLHSAMRYSLLGGGKRVRPALCLAACELVGGCLEVAMPAACAVEMIHTMSLIHDDLPAMDDDDVRRGQPANHRAFGEDIAILAGDALLALAFEHVA	**P**	**P**	**P**	**M**	**P**
39	>AY616113; Mg chelatase, subunit MLLAKPFMRLVMGPMRCPRNAMAAAPLSRKHLRVQAVAAVAENGAAVQAAEGQARPIFPFTAVIGQEEMKLALILNVIDPKIGGVMIMGDRGTGKTTTIRALADLLPQVPVVAKDPFNSDPFDAELMSEEVRQRLKAGEELPVGSMKIPMVDLPLGATEDRVCGTIDIEKALTEGVKAFEPGLLA	**M**	**P**	**M**	**M**	**P**

[a] The amino acid sequence in **red** corresponds to the sequence located between the N-terminal methionine and the cleavage site predicted by TargetP, or for AY616045, AY616049, AY616070 and AY616113, the cleavage site predicted by ChloroP. When the aforementioned two programs failed to predict a plastid-targeting peptide, the length of the N-terminal extensions was estimated by multiple alignments which included *P. wickerhamii* sequences, best-BLAST proteins with plastid targeting-peptides, prokaryotic/cyanobacterial homologues and, when applicable, eukaryotic proteins cytosol-localized. The first amino acid after the sequence in **blue** corresponds to the N-terminal methionine in proteins that do not have such extensions or for AY616038, to the first stretch of conserved residues. The note at the end of the amino acid sequence of these clusters provides detailed information for each case; the accession numbers are for GenBank. In all situations, the sequences that were identified as targeting-peptides represent extensions compared with their prokaryotic/cyanobacterial homologues or/and with eukaryotic proteins that lack N-terminal extensions, i.e., cytosol-localized.

[b]TargetP (http://www.cbs.dtu.dk/services/TargetP) predicts N-terminal sorting peptides, including those for plastid and mitochondrial localization; the dash (-) implies that the program was not able to identify any N-terminal sorting peptides. The detailed results of the prediction can be found in Table S1b.

[c]ChloroP (http://www.cbs.dtu.dk/services/ChloroP) predicts only the presence of plastid-targeting peptides, the dash (-) indicates that the program was not able to identify this peptide. The detailed results of the prediction can be found in Table S1c.

[d]IPSORT (http://biocaml.org/ipsort/iPSORT) predicts the presence of a signal peptide, as well as the presence of a plastid or mitochondrial targeting peptide. The dash (-) implies that the program was not able to identify any N-terminal sorting signals.

[e]Predotar (http://www.inra.fr/predotar) can identify putative mitochondrial and plastid targeting sequences. If the score is higher than 0.5 for both the mitochondrion and the plastid, the prediction will be for targeting to both cellular compartments. The dash (-) indicates that the program was not able to identify any N-terminal mitochondrial or plastid sorting peptides. The detailed results of the prediction can be found in Table S1d.

[f]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; M, mitochondrial targeted; C, cytosolic; ER, endoplasmic reticulum (i.e., microsomal fraction).

Table S1b. TargetP prediction results for plastid-targeted proteins from *P. wickerhamii*

Name	Length	cTP	mTP	SP	other	Loc.	RC	TPlen
AY616038	199	0.276	0.564	0.019	0.209	M	4	31
AY616039	220	0.920	0.206	0.005	0.071	C	2	30
AY616045	162	0.044	0.226	0.015	0.873	_	2	-
AY616049	220	0.121	0.260	0.034	0.566	_	4	-
AY616050	202	0.884	0.103	0.014	0.019	C	2	66
AY616056	251	0.433	0.251	0.003	0.125	C	5	76
AY616059	146	0.906	0.214	0.020	0.023	C	2	21
AY616060	211	0.895	0.358	0.009	0.017	C	3	32
AY616061	337	0.529	0.325	0.002	0.181	C	4	54
AY616063	200	0.063	0.811	0.001	0.103	M	2	60
AY616065	159	0.773	0.527	0.004	0.017	C	4	43
AY616069	150	0.793	0.288	0.022	0.062	C	3	34
AY616070	215	0.170	0.925	0.005	0.017	M	2	41
AY616071	209	0.826	0.479	0.009	0.017	C	4	37
AY616072	210	0.905	0.470	0.009	0.012	C	3	47
AY616074	188	0.108	0.232	0.054	0.516	_	4	-
AY616075	204	0.978	0.211	0.005	0.013	C	2	47
AY616077	230	0.930	0.049	0.007	0.077	C	1	59
AY616080	186	0.701	0.627	0.011	0.008	C	5	45
AY616081	213	0.692	0.354	0.001	0.118	C	4	52
AY616083	211	0.764	0.612	0.028	0.013	C	5	36
AY616084	212	0.756	0.248	0.005	0.116	C	3	48
AY616085	244	0.901	0.285	0.003	0.034	C	2	71
AY616086	221	0.613	0.057	0.025	0.430	C	5	60
AY616088	215	0.939	0.406	0.012	0.013	C	3	24
AY616089	187	0.758	0.590	0.007	0.015	C	5	33
AY616091	431	0.598	0.086	0.028	0.389	C	4	17
AY700207	208	0.943	0.281	0.005	0.040	C	2	50
AY700208	190	0.924	0.504	0.007	0.018	C	3	41
AY616099	240	0.437	0.500	0.010	0.038	M	5	25
AY616093	204	0.466	0.768	0.007	0.026	M	4	38
AY616094	229	0.870	0.487	0.003	0.010	C	4	52
AY616096	391	0.948	0.227	0.004	0.005	C	2	71
AY616097	202	0.958	0.127	0.019	0.005	C	1	61
AY616100	145	0.832	0.298	0.013	0.026	C	3	47
AY616101	209	0.899	0.545	0.003	0.007	C	4	53
AY616105	213	0.625	0.446	0.005	0.047	C	5	57
AY616106	184	0.859	0.072	0.002	0.082	C	2	40
AY616113	185	0.065	0.903	0.018	0.091	M	1	112

The following description of the TargetP output was adapted from: http://www.cbs.dtu.dk/services/TargetP-1.0/pages/output-expl.html).

Name is the GenBank accession number for the putatively plastid targeted proteins from *P. wickerhamii*; **Length** is the length of the submitted sequence; **cTP/mTP/SP/other**: The neural network output score for each of the possible categories. The location with the highest score is the most likely one according to TargetP, and the relation between the scores may be an indication of how certain the prediction is (see column RC). **Loc.** The subcellular localization predicted by TargetP: **C**: Chloroplast, *i.e. the sequence contains a chloroplast transit peptide, cTP;* **M**: Mitochondrion, *i.e. the sequence contains a mitochondrial targeting peptide, mTP;* **S**: Secretory pathway, *i.e. the sequence contains a signal peptide, SP;* _ : any other location; **RC**: Reliability Class: a measure of the size of the difference (diff) between the highest (winning) and the second highest output scores. **TPlen**: For sequences predicted to contain a cTP/mTP/SP, this is the predicted *length* of the presequence. For SPs, SignalP is used in this prediction, and for cTPs, ChloroP is used.

Table S1c. ChloroP prediction results for plastid-targeted proteins from *P. wickerhamii*

Name	Length	Score	cTP	CS-score	cTP-length
AY616038	199	0.477	-	6.870	30
AY616039	220	0.566	Y	6.442	30
AY616045	162	0.526	Y	4.663	23
AY616049	220	0.516	Y	8.808	34
AY616050	202	0.563	Y	5.312	66
AY616056	251	0.498	-	1.105	76
AY616059	146	0.553	Y	2.369	21
AY616060	211	0.552	Y	3.426	32
AY616061	337	0.503	Y	5.246	54
AY616063	200	0.465	-	5.207	27
AY616065	159	0.573	Y	9.314	43
AY616069	150	0.537	Y	6.341	34
AY616070	215	0.558	Y	9.844	40
AY616071	209	0.567	Y	11.055	37
AY616072	210	0.577	Y	4.368	47
AY616074	188	0.463	-	3.642	23
AY616075	204	0.547	Y	5.458	47
AY616077	230	0.575	Y	5.718	59
AY616080	186	0.545	Y	8.454	45
AY616081	213	0.538	Y	4.033	52
AY616083	211	0.541	Y	7.047	36
AY616084	212	0.547	Y	3.284	48
AY616085	244	0.545	Y	5.801	71
AY616086	221	0.536	Y	5.488	60
AY616088	215	0.557	Y	5.250	24
AY616089	187	0.541	Y	1.609	33
AY616091	431	0.467	-	0.635	17
AY700207	208	0.573	Y	4.985	50
AY700208	190	0.574	Y	7.890	41
AY616099	240	0.542	Y	4.968	52
AY616093	204	0.485	-	6.354	20
AY616094	229	0.522	Y	5.818	52
AY616096	391	0.587	Y	3.876	71
AY616097	202	0.589	Y	6.208	61
AY616100	145	0.560	Y	6.572	47
AY616101	209	0.561	Y	5.961	53
AY616105	213	0.526	Y	0.551	57
AY616106	184	0.578	Y	4.978	40
AY616113	185	0.507	Y	8.993	48

The following description of the ChloroP output was adapted from: http://www.cbs.dtu.dk/services/ChloroP-1.1/pages/output-expl.php).

Name is the GenBank accession number for the putatively plastid targeted proteins from *P. wickerhamii*; **Length** is the length of the submitted sequence; **Score** is the output score from the second step network. The prediction cTP/no cTP is based solely on this score; **cTP** tells whether or not this is predicted as a cTP-containing sequence; "Y" means that the sequence *is* predicted to contain a cTP; "-" means that is predicted *not* to contain a TP. **CS-score** is the MEME scoring matrix score for the suggested cleavage site. **cTP-length** is the predicted length of the presequence (**Please note** that the prediction of the transit peptide length is carried out and presented *even if* its presence is not predicted).

Table S1d. Predotar prediction results for plastid-targeted proteins from *P. wickerhamii*

Sequence	Mitochondrial score	Plastid score	Prediction
AY616038	0.986	0.001	mitochondrial
AY616039	0.005	0.633	plastid
AY616045	0.000	0.165	neither
AY616049	0.628	0.218	mitochondrial
AY616050	0.170	0.915	plastid
AY616056	0.039	0.180	neither
AY616059	0.420	0.850	plastid
AY616060	0.034	0.995	plastid
AY616061	0.028	0.009	neither
AY616063	0.604	0.262	mitochondrial
AY616065	0.011	0.989	plastid
AY616069	0.016	0.380	neither
AY616070	0.906	0.008	mitochondrial
AY616071	0.816	0.523	both
AY616072	0.426	0.799	plastid
AY616074	0.000	0.000	neither
AY616075	0.011	0.998	plastid
AY616077	0.008	0.998	plastid
AY616080	0.001	0.431	neither
AY616081	0.512	0.747	both
AY616083	0.998	0.000	mitochondrial
AY616084	0.249	0.243	neither
AY616085	0.986	0.043	mitochondrial
AY616086	0.000	0.001	neither
AY616088	0.337	0.776	plastid
AY616089	0.906	0.246	mitochondrial
AY616091	0.000	0.022	neither
AY700207	0.314	0.946	plastid
AY700208	0.012	0.988	plastid
AY616099	0.198	0.895	plastid
AY616093	0.111	0.041	neither
AY616094	0.009	0.141	neither
AY616096	0.010	0.403	neither
AY616097	0.015	0.997	plastid
AY616100	0.030	0.944	plastid
AY616101	0.162	0.601	plastid
AY616105	0.442	0.189	neither
AY616106	0.801	0.398	mitochondrial
AY616113	0.998	0.000	mitochondrial