Acquisition, conservation and loss of dual-targeted proteins in land plants

Lin Xu\(^1\), Chris Carrie\(^2\), Simon Law\(^1\), Monika W. Murcha\(^1\), and James Whelan\(^1\)*

\(^1\)ARC Centre of Excellence in Plant Energy Biology, Bayliss Building M316 University of Western Australia, 35 Stirling Highway, Crawley 6009, Western Australia, Australia. \(^2\)Department of Biology I, Botany, Ludwig-Maximilians-Universität München, Großhaderner Strasse 2-4, D-82152, Planegg-Martinsried, Germany.

*Corresponding author: James Whelan
Email: jim.whelan@uwa.edu.au

Running title: An evolutionary view of dual-targeted proteins

Expected page number: 15.2

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors is James Whelan (jim.whelan@uwa.edu.au)

High resolution figures available:
URL: http://www.plantenergy.uwa.edu.au/transfer/
Username: pebguest
Password: p@nth3r23
Abstract
The dual targeting ability of a variety of proteins from Physcomitrella patens, Oryza sativa and Arabidopsis thaliana were tested to determine when dual targeting arose and to what extent it is conserved in land plants. Overall the targeting ability of over 80 different proteins from rice and Physcomitrella representing 42 dual-targeted proteins in Arabidopsis were tested. It was found that dual targeting arose early in land plant evolution as it was evident in many cases with Physcomitrella proteins, which were conserved in rice and Arabidopsis. Furthermore, it was evident that the acquisition of dual targeting ability is still occurring, evident in Physcomitrella, as well as rice and Arabidopsis. The loss of dual targeting ability appears rare, but does occur. Ascorbate peroxidase represents such an example, following gene duplication in rice, individual genes encode proteins that are targeted to a single organelle. Whilst dual targeting was generally observed to be conserved, the ability to detect dual-targeted proteins differed depending on the cell types used. Furthermore it appears that small changes in the targeting signal can result in a loss (or gain) of dual targeting ability. Overall, examination of the targeting signals within this study did not reveal any clear patterns that would predict dual targeting ability. The acquisition of dual targeting ability also appears to be coordinated between proteins, Mia40, a protein involved in oxidative folding in mitochondria and peroxisomes, provides an example where acquisition of dual targeting is accompanied by the dual targeting of substrate proteins.
Introduction

Gene transfer to the host nucleus followed the endosymbiotic events that led to the formation of mitochondria and plastids in plant cells (Adams et al., 2000; Dyall et al., 2004; Kleine et al., 2009; Keeling, 2010). This resulted in a reduction of the coding capacity of these organelles to ~5% of the original endosymbionts genome (Pfannschmidt, 2010) and therefore the majority of organellar proteins are encoded in the nucleus, translated in the cytosol and imported into their respective organelles. This process of protein targeting required that new machinery, not present in the original endosymbionts be acquired to specifically recognize and translocate thousands of proteins across their respective organelle membranes (Dolezal et al., 2006). Studies into mitochondrial and plastid protein import revealed that targeting and import is specific for each organelle (Rudhe et al., 2002; Glaser and Whelan, 2007). This specificity is believed to be due to a number of factors, the nature of the targeting signals, the presence of cytosolic “targeting” factors and the presence of protein receptors on the organelle surface, all of which contribute to maintain the specificity of protein import (Chew and Whelan, 2004). The molecular mechanisms of how these features maintain specificity is largely unknown and even further complicated with the growing identification of proteins that can be targeted to multiple organelles.

The initial report that glutathione reductase from *Pisum sativum* was targeted to both mitochondria and plastids revealed that targeting could occur to two organelles and protein targeting was not location specific (Creissen et al., 1995). Since this initial report 107 proteins have now been reported to be dual-targeted to mitochondria and plastids in a variety of plants (Carrie et al., 2009a; Carrie and Small, 2012). The dual targeting of proteins can occur by a variety of mechanisms (Peeters and Small, 2001; Yogev and Pines, 2011), such as ambiguous targeting signals, where a single targeting signal has the ability to target a protein to two distinct locations, or via alternative transcription/translation to produce altered targeting signals for each respective organelle (Peeters and Small, 2001; Yogev and Pines, 2011). Dual targeting by ambiguous signals is of particular interest as it questions how two distinct organelle import machineries can recognize these ambiguous signals...
and yet distinguish between organelle specific signals. In addition to dual targeting to mitochondria and plastids, dual targeting of proteins to mitochondria and peroxisomes has also been reported (Carrie et al., 2008; Carrie et al., 2009b). The mechanism of targeting differs in that proteins dual-targeted to mitochondria and peroxisomes contain two signals, which result in the same protein being imported into two distinct locations.

To date, 72 proteins in *Arabidopsis thaliana* have been shown to be dual-targeted (Carrie et al., 2009a; Carrie and Small, 2012) but as many as 500 are predicted to be dual-targeted by containing ambiguous signals (Mitschke et al., 2009). The necessity for dual targeting remains largely unknown though it has been suggested that it may be related to gene copy number and restriction of genome size (Morgante et al., 2009), or required for the co-ordination of organelle function (Chew et al., 2003; Carrie et al., 2009a). There is only a limited amount of information available regarding the extent of dual targeting of orthologous proteins between species. Dual targeting of four proteins (methionine aminopeptidase, monodehydroascorbate reductase, glutamyl- tRNA synthetase and tyrosyl-tRNA synthetase) have been shown to be conserved in *Oryza sativa* (rice) and Arabidopsis (Morgante et al., 2009). Organellar seryl-tRNA is dual targeted in Arabidopsis and *Zea Mays* (maize) (Rokov-Plavec et al., 2008). Additionally the MutS HOMOLOG 1 was also found to be dual-targeted in a number of dicot plants (Xu et al., 2011).

In order to gain a better insight into dual targeting of proteins to mitochondria and plastids or mitochondria and peroxisomes, in terms of i) when did dual targeting arise in plant evolution? ii) is dual targeting of proteins conserved? iii) is acquisition or loss of dual targeting of proteins still occurring? and, iv) is dual targeting of proteins associated with gain of function in organelles?, a study was undertaken to determine if dual-targeted proteins in Arabidopsis were also dual-targeted in rice and *Physcomitrella patens*. Furthermore, if differences were observed in the dual targeting ability of proteins from within these three species, the targeting ability of *Picea glauca* orthologs was additionally investigated, thus examining the dual targeting ability of proteins spanning 500 million years of land plant evolution.
Overall these questions were posed to gain a better understanding of the purpose of dual targeting of proteins.

Results

In land plants, 107 proteins have been experimentally reported to be dual-targeted to mitochondria and plastids or mitochondria and peroxisomes (Supplemental Table S1 & S2) (Carrie et al., 2009a; Carrie and Small, 2012). In order to determine when dual targeting arose in land plant evolution and if dual targeting ability is conserved, we investigated if the orthologs of some dual-targeted proteins in Arabidopsis were also dual-targeted in Physcomitrella and rice. As Physcomitrella diverged from Arabidopsis ~450 million years ago (Figure 1A), it represents the earliest land plant with a complete genome sequence, allowing identification of all orthologs within gene families of Arabidopsis dual-targeted proteins. It should be noted that for many dual-targeted proteins, location specific orthologs also exist, so identification of all gene family members is required to ensure that all orthologous proteins are being identified to test for targeting ability. The orthologs from Chlamydomonas reichardtii (Merchant et al., 2007) and Chlorella variabilis (Blanc et al., 2010) were also identified and included in the phylogenetic analysis to obtain a more robust characterization of orthologs. Therefore the ortholog that showed the highest level of sequence identity and similarity to the Arabidopsis dual-targeted protein was defined as the closest ortholog and its targeting ability was tested (Supplemental dataset S1). The targeting of Physcomitrella proteins were tested in Arabidopsis cell suspensions and onion epidermal cells, in addition to Physcomitrella protonemal tissues as there have been no previous reports of targeting Physcomitrella proteins in non-homologous plant tissues. The mitochondrial targeting signals from alternative oxidase (AOX) and the alpha subunit of the mitochondrial processing peptidase (MPP alpha), plastid targeting signals of the small subunit of 1, 5 ribulose biphosphate carboxylase oxygenase (SSU Rubisco) and Photosystem I (PS I) subunit 2, and peroxisomal targeting signals from thiolase and malate synthase were tested in Arabidopsis cell suspensions, onion epidermal cells and Physcomitrella tissue to define the fluorescence characteristics of these organelles in the various tissues tested.
Additionally this demonstrated that the mitochondrial, plastid and peroxisomal RFP markers previously used in Arabidopsis to define these organelles can also be used with Physcomitrella (Carrie et al., 2009b).

**Dual targeting arose early and is conserved during land plant evolution**

A number of orthologs to dual-targeted proteins in Arabidopsis were also found to be dual-targeted from rice and Physcomitrella. DNA Topoisomerase represents an example of a protein that is dual-targeted to mitochondria and plastids in Arabidopsis (Carrie et al., 2009b) and was similarly found to be dual-targeted to mitochondria and plastids from rice and Physcomitrella (Figure 2). Analysis of Topoisomerase genes reveals that there are three in Arabidopsis, four in rice and five in Physcomitrella (Figure 2A). While the targeting predictions of the proteins differ using different prediction programs (Figure 2B), it was apparent that Topoisomerases were dual-targeted in all three species when tested by biolistic transformation (Figure 2C). Noticeably while dual targeting of the rice Topoisomerase was readily apparent in Arabidopsis cell suspensions and onion epidermal cells, dual targeting of the Physcomitrella Topoisomerase was only apparent in onion epidermal cells and in Physcomitrella protonemal tissues, where the mitochondria were elongated and cylinder-like in shape (Figure 2C). This elongated shape was similar to that observed with the control mitochondrial markers (Figure 1B). Thus while dual targeting of Topoisomerases was conserved, the dual targeting of the Physcomitrella Topoisomerase was not observed in all cell types tested.

Analysis of a large number of other proteins revealed that while many were observed to be dual-targeted, the ability for dual targeting differed between cell types. In the case of DNA helicase (Figure 3), previously shown to be dual-targeted from Arabidopsis (Carrie et al., 2009b), the Physcomitrella ortholog tested (PpHel1) and the only rice ortholog (OsHel) were also determined to be dual-targeted in all tissues tested (Figure 3). Thus it differed from the Physcomitrella Topoisomerase that was not observed to be dual-targeted in Arabidopsis cell suspensions. However, in the case of DNA Polymerase (Figure 4), whilst the two Arabidopsis orthologs AtPolgamma 1 & 2 and the rice ortholog OsPol1 were observed to be dual-targeted to
mitochondria and plastids (Carrie et al., 2009b), the Physcomitrella orthologs showed a different pattern (Figure 4). Of the four Physcomitrella DNA polymerase orthologs, three, PpPol2, PpPol3 and PpPol1 branched with the Arabidopsis proteins. However PpPol1 was only targeted to plastids, in all tissues tested (Figure 4C). In contrast, whilst PpPol2 was predominantly targeted to mitochondria in Arabidopsis cell suspensions, plastid and mitochondrial targeting was evident in onion epidermal cells (Figure 4C). In Physcomitrella protonemal tissues, as in the Arabidopsis, mitochondrial targeting was dominant with plastid targeting only weakly observed (Figure 4C).

Whilst only some examples have been discussed above, overall in the analysis of 38 Arabidopsis proteins that were previously published to be dual-targeted to mitochondria and plastids, in 28 out of 38 cases rice contains an ortholog that was dual-targeted (Supplemental Figure S1, Supplemental Table S1, shaded in green and yellow¹), and 9 out of 13 cases rice and Physcomitrella were shown to be dual-targeted in the tissues tested (Supplemental Table S1, Supplemental Figure S1, shaded in green²).

The dual targeting of proteins to mitochondria and peroxisomes or plastids and peroxisomes has also been reported with 11 cases identified to date (Supplemental Table S2). The dual targeting of orthologs from rice and Physcomitrella of 6-Phosphogluconolactonase (dual-targeted to plastids and peroxisomes) (Reumann et al., 2007), and Alanine:glyoxylate aminotransferase (dual-targeted to mitochondria and peroxisomes) (Carrie et al., 2009b) were tested to determine if dual targeting was also conserved when it involved another organelle, the peroxisomes. The orthologs tested were found to be targeted to both plastids and peroxisomes or mitochondria and peroxisomes (Supplemental Figure S2.1 & S2.2) in all tissues.

Overall, in the analysis of four Arabidopsis proteins that were dual-targeted to mitochondria and peroxisomes or plastids and peroxisomes, three

¹ The following proteins were not dual-targeted to mitochondria and plastids in rice compared to Arabidopsis. Ascorbate peroxidase, Glutamine synthetase, Co-chaperone GrpE protein, Methylene-tetrahydrofolate reductase, Organellar single-stranded DNA binding protein, Rhodanase-like protein, SAL1, Sulfiredoxin, Thiazole biosynthetic enzyme and AAA-type ATPase family protein.

² The following proteins were not dual-targeted to mitochondria and plastids in rice and Physcomitrella compared to Arabidopsis. Ascorbate peroxidase, Co-chaperone GrpE protein, Toc33, Toc159, DNA ligase, SAL1.
out of four (Supplemental Table S2) orthologs from rice tested in Arabidopsis suspension and onion were dual-targeted and two out of three proteins from rice and Physcomitrella were dual-targeted to mitochondria and peroxisomes or plastids and peroxisomes (Supplemental Table S2).

**Dual targeting is gained and lost during land plant evolution**

While a variety of proteins examined were observed to be dual-targeted, Physcomitrella DNA polymerase hinted that targeting to one organelle may be stronger or more efficient than to another (Figure 4), and that this may be associated with gene duplication. Therefore a detailed study was carried out investigating the dual targeting ability of a number of proteins that are known to be encoded by small gene families. Enzymes of the ascorbate-gluthione reductase cycle were chosen as, glutathione reductase (GR), ascorbate peroxidase (APX) and monodehydroascorbate reductase (MDHAR) contain orthologs that are dual-targeted in Arabidopsis and are part of multi-gene families (Figure 5A) (Chew et al., 2003). GR orthologs from rice (OsGR1) and Physcomitrella (PpGR1) were similarly observed to be dual-targeted in Arabidopsis, onion and Physcomitrella tissue (Figure 5D). In the case of APX, four genes encode APX in Physcomitrella, five in Picea, seven in Arabidopsis and eight in rice (Figure 6A). The Arabidopsis APX designated as a plastid stromal isoform (AtSAPX) has been previously reported to be dual-targeted (Chew et al., 2003), and this was confirmed in this study (Figure 6). However, examination of Physcomitrella APX1, showed it was only targeted to plastids in all three tissue types tested (Figure 6B). Examination of the rice APX proteins, OsAPX5, OsAPX6, OsAPX7 and OsAPX8 which grouped with and displayed the highest sequence identity to AtSAPX (Figure 6A & C) showed that OsAPX5 and OsAPX6 were targeted to mitochondria, while OsAPX7 and OsAPX8 were targeted to plastids (Figure 6B). To gain insight into when dual targeting of APX was gained (or lost), the targeting of the most orthologous APX from Picea, PgAPX1, was examined and found to be dual-targeted (Figure 6). Thus it is proposed that the dual targeting ability

---

3 The following protein was not dual-targeted to mitochondria and peroxisomes in rice compared to Arabidopsis Malonyl-CoA decarboxylase.

---

Downloaded on December 30, 2020. - Published by https://plantphysiol.org
Copyright (c) 2020 American Society of Plant Biologists. All rights reserved.
of APX arose following the split between Physcomitrella and Picea and subsequently lost in rice following monocot divergence. In rice, gene duplication resulted in two genes with each encoding organelle specific proteins (Figure 6A). The gene family of MDHAR has multiple members identified with four genes in Physcomitrella, three in Picea, five in Arabidopsis and six in rice (Figure 7A). AtMDHAR6, previously shown to be dual-targeted (Chew et al., 2003) has two orthologs in Physcomitrella, one of which showed targeting to both mitochondria and plastids. As MDHAR dual-targeted orthologs could be detected in all four species, including Picea (Figure 7), it suggests that as with GR, dual targeting ability arose early in land plant evolution and has been conserved. However PpMDHAR1 and OsMDHAR1.2 are not dual-targeted, it suggests that dual targeting of these isoforms may have been lost following gene duplication in these organisms, so as only a single dual-targeted isoform is conserved in each organism.

Hexokinase proteins represent another example of the acquisition of dual targeting ability throughout land plant evolution. Whilst the targeting of Arabidopsis orthologs have not previously been studied, a recent study (Nilsson et al., 2011) in Physcomitrella had identified isoforms that were dual-targeted to mitochondria and plastids. In this study, of the 11 genes encoding hexokinase (PpHXK1-11) (Figure 8A), at least six isoforms are dual-targeted (Figure 9). Compared to the previous study some differences have been observed (Nilsson et al., 2011), in that PpHXK5 and PpHXK9 were found to be dual-targeted whilst PpHXK8 was not (Figure 8B). However, none of the four hexokinase orthologs, AtHXK1-4 (or one hexokinase-like AtHKL1) from Arabidopsis displayed dual targeting ability, with four orthologs targeted to mitochondria and one plastids (Figure 8C). Phylogenetic analysis of the dual-targeted hexokinase isoforms from Physcomitrella reveals that five out of the six isoforms branch together (Figure 8A), suggesting that a single gene in Physcomitrella encoding a dual-targeted hexokinase underwent gene duplication as all proteins in this group display dual targeting ability (Figure 8D). On the other hand, Hexokinase 5 (PpHHK5) also displays dual targeting ability (Figure 8D), most likely acquired throughout evolution as its closest orthologs (PpHXK1 & 6) (Figure 8A) are not dual-targeted (Figure 8D). As Physcomitrella hexokinases are more similar to each other than to
hexokinases in other land plants, and have “undergone concerted evolution” (Nilsson et al., 2011), this is consistent with the proposal that dual targeting arose after Physcomitrella had diverged from the lineage that gave rise to other land plants.

**Acquisition of dual targeting ability may allow organelles to gain additional functions**

In the case of peroxisomes, Mia40 represents an example of a protein that has acquired dual targeting ability during land plant evolution. Mia40 was first identified in Arabidopsis as an ortholog to the essential yeast protein, Mitochondrial intermembrane space import and assembly protein 40 (Mia40) (Carrie et al., 2010b). However subsequent work on the Arabidopsis Mia40 showed that it differed to the yeast protein. Firstly, the Arabidopsis Mia40 was shown to be a non-essential protein and was dual-targeted to both the mitochondrial intermembrane space and the peroxisomal matrix (Carrie et al., 2010b). It was also demonstrated that the rice Mia40 was similarly dual-targeted to mitochondria and peroxisomes (Carrie et al., 2010b). Further analyses of the protein sequences from other plant Mia40 proteins showed that in higher plants Mia40 contains a known peroxisomal targeting sequence (PTS1), whilst Mia40 from lower plant species does not contain this PTS1 sequence (Supplemental Figure S3), thus the Physcomitrella Mia40 (PpMia40) was chosen to determine its subcellular location. GFP analysis of PpMia40 found that it did not target to either mitochondria or peroxisomes and instead appeared to reside in the cytosol as evidenced by a diffuse GFP signal in all tissue types tested (Figure 9C). Whilst the lack of peroxisomal targeting is expected due to the lack of a PTS1 sequence at the C-terminus, the lack of mitochondrial targeting is surprising considering all other known Mia40 proteins have been shown to be located within mitochondria. In addition, the PpMia40 sequence was observed to contain the conserved cysteine residues and is highly similar to the Arabidopsis Mia40 in the enzymatically active domain. PpMia40 is missing 28 amino acids from the N-terminal end when compared to Arabidopsis Mia40 (Supplemental Figure S3). In an attempt to deduce the targeting ability of Mia40 over a wider range of plants, Mia40 sequences from eleven plant species, ranging from
Chlamydomonas reinhardtii to Arabidopsis were examined for predicted targeting ability (Supplementary Figure S3, Figure 10), and also the proposed substrates for Mia40, Ccs1 (copper chaperone for superoxide dismutase 1), CSD1 (copper zinc superoxide dismutase), 2 and 3 (Figure 10). It was observed that mitochondrial targeting of Mia40, and dual targeting of Mia40, arose relatively late in plant evolution (Figure 10). It was notable that Physcomitrella, Volvox (Volvox Carteri), Chlorella and to a lesser extent Chlamydomonas lacked the first 30 to 40 amino acids compared to the predicted mitochondrial targeted Mia40 from other plants (Supplemental Figure S3). However the peroxisomal acquisition of Mia40 is also accompanied by peroxisomal targeting ability of Ccs1 and CSD3. Note that as previously demonstrated that Erv1 can carry out the oxidative folding of proteins in mitochondria alone (Carrie et al., 2010b) so it is likely that Ccs1 and CSD like proteins can be assembled in mitochondria without a mitochondrial Mia40, in Physcomitrella and Picea (Figure 10).

Discussion

This study used GFP tagging of proteins from several land plants to determine when dual targeting of proteins arose in land plant evolution, and if it was conserved. While it is desirable to use a variety of approaches to determine the location of a protein (Millar et al., 2009), for the analysis of 96 proteins from Arabidopsis, rice, Physcomitrella and three from Picea, GFP tagging is the only realistic approach to determine targeting ability. The use of various other approaches to determine the presence of a protein, either by immunodetection or mass spectrometry was not feasible due to large gene families and thus the requirement of isoform specific antibodies. Mass spectrometric approaches are only feasible for highly purified organelles and may not identify proteins that are relatively low in abundance. Even in Arabidopsis, the most intensively studied plant in subcellular proteomics, only 13 of the 72 dual-targeted proteins that have been determined by GFP tagging, were also detected in two organelles via proteomic studies (Heazlewood et al., 2007). The in vivo targeting assay using GFP tagging offers a realistic approach to assess targeting as if targeting to one organelle,
mitochondria, plastids or peroxisomes, is observed, it indicates that the protein is in an import competent state.

It was observed that in many cases, dual targeting was conserved. If dual targeting ability arose early in evolution, it remained conserved from Physcomitrella to Arabidopsis and rice, as 11 out of 16 tested proteins were confirmed to be dual-targeted in all three species (Supplemental Table S1 and S2, shaded in green). Similarly if dual targeting arose later in plant evolution, dual targeting remained conserved with two out of five tested proteins confirmed to be dual-targeted from rice and Arabidopsis. Loss of dual targeting could be concluded with confidence as was observed with APX, where the loss of dual targeting in rice was accompanied by gene duplication followed by neo-functionalization, in that the duplicated genes encoded proteins that were targeted to single locations. A similar scenario also appears to have occurred with MDHAR isoforms in rice and Physcomitrella. It could also be observed that dual targeting ability is being acquired as a number of proteins (seven) are dual-targeted from Arabidopsis alone. The dual targeting ability of Hexokinases from Physcomitrella, is likely to be a derived feature, rather than it being lost from Arabidopsis.

While dual targeting of proteins was conserved in many cases, differences in targeting ability of proteins to mitochondria and plastids were observed with the different tissues/cells used, e.g. Physcomitrella DNA Polymerase 2. This does not appear to be co-ordinated with differences in the mitochondrial protein import apparatus of Physcomitrella, rice and Arabidopsis. The outer mitochondrial membrane protein import receptor 64 (OM64), which is derived from an ancestral gene encoding a plastid outer envelope protein is only present in rice and Arabidopsis (Chew et al., 2004; Carrie et al., 2010a), yet dual targeting occurs in Physcomitrella. The differences observed may be due to a variety of reasons; firstly there are isoforms of the protein import components present in plastids and mitochondria (Soll and Schleiff, 2004; Lister et al., 2007; Jarvis, 2008) and for plastids it has been proposed that these different isoforms may import different sets of proteins. Thus the difference in import between systems may reflect the different abundance of various isoforms in various cells, and/or the fact that there is co-evolution or specialization of precursor proteins to bind to
specific isoforms of protein import components. An analysis of the Tom20 import family of proteins in Arabidopsis suggested that different Tom20 isoforms may display some preference or differences for binding different precursor proteins (Lister et al., 2007; Duncan et al., 2012). Furthermore, although targeting signals are generally considered to be well conserved across wide phylogenetic gaps, an analysis of mitochondrial targeting signals from rice and Arabidopsis revealed differences in length and amino acid composition could be detected, suggesting that subtle differences between species are likely that may affect the efficiency of targeting in different cell types (Huang et al., 2009). Another reason for the differences between various cells tested is that even within a single species the extent of dual targeting varies in cells from different tissues (Carrie et al., 2009b). Finally it has previously been reported that while some proteins are dual-targeted, that targeting to a single organelle is only observed in a single cell (Beardslee et al., 2002), although the reason(s) for this are unknown. Thus the reasons that some variation in dual targeting may be observed between various systems with some proteins may be due to a variety of reasons.

Analysis of the targeting signals of dual-targeted proteins does not reveal any specific motifs or residues that are associated with dual targeting. In this study protein isoforms that display very high levels of sequence similarity were observed to differ in dual targeting ability. In the case of rice MDHAR, one splicing isoform was dual-targeted, OsMDAHR1.1 (Figure 7), yet another with only four amino acids different, OsMDAHR1.2 was not dual-targeted (Figure 7) (Supplemental Figure S4). With Physcomitrella MDHAR1 or 2, four amino acid differences in the predicted targeting region result in PpMDAHR1 not being targeted to mitochondria or plastids, and PpMDAHR2 being dual-targeted (Figure 7) (Supplemental Figure S5). Thus overall it appears that small changes in protein sequences do apparently result in large differences in targeting, it is likely not an all or nothing situation. The threshold of GFP detection may result in a more dramatic difference that occurs in vivo, and as observed with DNA helicases and DNA Polymerase where the ability to detect dual targeting differed between cell types (Figure 3 and 4).
One of the incentives for this study was to gain a better understanding of the purpose of dual targeting. As outlined above, dual targeting appears to be conserved once it arises, and thus under positive selection to be maintained. However, the functional characterization of several dual-targeted proteins has largely concluded that the effects of inactivation of genes encoding dual-targeted proteins are only observed in a single organelle. For example, inactivation of a dual-targeted RNA polymerase, targeted to mitochondria and plastids, only resulted in changes to mitochondrial transcript abundance (Kuhn et al., 2009). The MUTS homolog 1, a dual-targeted protein that maintains genome stability in both plastids and mitochondria (Xu et al., 2011), suppression of protein abundance resulted in responses that were plastid in origin (Xu et al., 2012). In the case of a dual-targeted mitochondrial carrier protein AtBT1, shown to be additionally targeted to plastids (Bahaji et al., 2011b), complementation of a T-DNA mutant using a mitochondrial specific form of AtBT1 resulted in the restoration of a normal growth phenotype (Bahaji et al., 2011a). These studies suggest that dual-targeted proteins appear to have a predominantly central role in a single organelle, yet this study indicates that dual targeting is conserved which suggests that dual-targeted proteins may well function in both organelles and be under some kind of positive selection. These apparently conflicting conclusions may be reconciled by the fact that it is likely that while other location specific isoforms can restore the normal phenotype under a given set of conditions, it does not indicate that the underlying complex molecular interactions (transcriptome, proteome, metabolome) are restored to normal. Consequently, these hemi-complemented plants provide a valuable resource to explore more cell or condition specific functions of dual-targeted proteins, which drives the conservation of dual targeting.

The evolutionary history of the targeting of Mia40 makes an interesting case. Mia40 is believed to be found in most eukaryotic species but is best studied in yeast model systems (Chacinska et al., 2004). In yeast, Mia40 is located within the mitochondrial intermembrane space where it plays an essential role in the oxidation and folding of intermembrane space proteins (Chacinska et al., 2004). It was originally thought that this was the role of Mia40 in all eukaryotic species. However, in Arabidopsis it was demonstrated
that a mia40 knockout is not lethal and that it was also targeted to peroxisomes (Carrie et al., 2010b). The role for Arabidopsis Mia40 is thought to involve the oxidation and folding of both the mitochondrial and peroxisomal located Copper chaperone for superoxide dismutase 1 (Ccs1) and copper zinc superoxide dismutase (CSD1 and CSD3) (Carrie et al., 2010b). It is interesting to note that not all plant Mia40 proteins are targeted to peroxisomes. It appears that the dual targeting of Mia40 has arisen later in plant evolution. The interesting point is that a dual-targeted Mia40 is accompanied by Ccs1 and CSD proteins with a clear PTS1 sequence in Arabidopsis (Supplemental Figure S3) (Figure 10). So in lower plant species such as Physcomitrella and Picea Mia40 is not dual-targeted, their peroxisomes are not predicted to contain Ccs1 or CSD proteins (Supplemental Figure S3) (Figure 10). In higher plants such as Arabidopsis and Populus trichocarpa, not only does Mia40 contain a PTS1 sequence, so do their Ccs1 and CSD proteins. Rice and other monocots appear to be an intermediate of the above case where rice has a dual-targeted Mia40, but has no predicted peroxisomal targeted Ccs1 or CSD proteins (Figure 10). This hints that over time biochemical pathways involving several proteins may become dual-targeted.

It appears that as observed in Arabidopsis dual targeting of proteins is widespread in various land plant lineages. Overall it appears that dual targeting is well conserved, but with gene duplication and amplification, dual targeting may be lost with the resulting individual genes encoding proteins that are location specific. The fact that dual targeting of proteins appears well conserved, along with the fact that acquisition of dual targeting may result in additional biochemical pathways within organelles, as observed with Mia40. This suggests that at least one role for dual targeting is that it allows the development of metabolic complexity and functional diversification, in that the same protein targeted to two locations can result in two different functions without the introduction of any change in the genome per sec. Once this occurs and these pathways are established gene duplication followed by neo-functionalization may result in proteins that are targeted to only one organelle. Phylogenetic analysis of the plastid and mitochondrial proteome suggest that they are composed of proteins that arose from a variety of evolutionary
sources (Gray et al., 2001; Martin, 2010; Suzuki and Miyagishima, 2010; Szklarczyk and Huynen, 2010), and in fact proteins from the endosymbionts that gave rise to mitochondria and plastids do not in fact represent the majority of proteins in these organelles. Thus dual targeting may represent an ancient means to develop metabolic/proteomic complexity in plant cells, which is still ongoing.
Material and methods

Bioinformatic Analyses

The tree showing the evolutionary relationship between the six plant species used in this study was determined according to their time of divergence (in millions of years) according to previous studies (Bowman et al., 2007; Rensing et al., 2008; Carrie et al., 2010b).

Protein sequences of all published Arabidopsis dual-targeted proteins (Supplemental Table S1 & S2) were obtained from TAIR (http://www.arabidopsis.org/) and used to BlastP (Altschul et al., 1990) against Oryza sativa, Physcomitrella patens and Chlamydomonas reihardtii protein databases using Phytozome (http://www.phytozome.net/). Orthologs from Picea glauca were identified using tBLASTn (Altschul et al., 1990) against EST sequences from the NCBI database (http://blast.ncbi.nlm.nih.gov/blast.cgi/). BlastP searches for Chlorella variabilis NC64 genome (Blanc et al., 2010) was done on http://genome.jgi-psf.org/ChlNC64A_1/ChlNC64A_1.home.html. Candidate proteins with a similarity percentage above 50%, and containing the same functional domains were included in this study. The ortholog with the highest percentage similarity and identity to the known Arabidopsis dual-targeted protein was selected for GFP targeting (Supplemental Dataset, indicated in red). When the ortholog did not exhibit dual targeting ability the next highest percentage similarity and identity ortholog was chosen. Subcellular localizations were predicted using Predotar (Small et al., 2004), TargetP (Emanuelsson et al., 2007) and WoLFPSORT (Horton et al., 2007). The Ambiguous Targeting Predictor (ATP) predictor was also used to predict the probability of a protein being dual-targeted (Mitschke et al., 2009). Multiple sequence alignments of protein family members were conducted using MAFFT (Katoh et al., 2005). Multiple align show (http://www.bioinformatics.org/sms/multi_align.html) was used to visualize multiple sequence alignment. MEGA 5 (Tamura et al., 2011) was used to construct phylogenetic trees using statistical method of neighboring-joining with the number of bootstrap replications of 1000. MatGAT 2 (http://bitincka.com/ledion/matgat/) was used to determine the percentage identity and similarity matrices.
Construction of GFP fusion vectors

RNA extraction from Arabidopsis thaliana, Oryza sativa and Physcomitrella patens was carried out using the RNeasy kit (Qiagen, Melbourne) according to manufacturer's instructions. Picea glauca RNA was obtained from Dr Olivier Keech (Umea Plant Science Center). Reverse transcription was carried out using SuperScript™ III First-strand synthesis system (Invitrogen, Sydney). The translational start sites for all Picea glauca genes were confirmed by 5′–RACE using CapFishing™ full-length cDNA premix kit (Seegene, South Korea). Full-length cDNA was amplified using gene specific primers flanked by Gateway recombination cassettes (see Supplemental Table S3) according to manufacturer's instructions. A number of genes (see Supplemental Table S3) were cloned directly from cDNA clones obtained from the Knowledge-based Oryza Molecular biological Encyclopedia (KOME) (Kikuchi et al., 2003) or the RIKEN Bioresource Center (Nishiyama et al., 2003). The targeting signals of Glycine max alternative oxidase (AOX), Pisum sativum small subunit of 1,5 – ribulose bisphosphate carboxylase/oxygenase (SSU) and Cucubita sp. malate synthase were fused to RFP and used as a mitochondrial, plastid and peroxisomal marker respectively (Carrie et al., 2007; Carrie et al., 2009b). COXIV-mCherry (Nelson et al., 2007) was used as the mitochondrial marker in onion epidermal cells. PpAOX-mCherry was generated by replacing COXIV targeting signal (1-29 aa) with full length Physcomitrella alternative oxidase (PpAOX) (1-365 aa) in front of mCherry fluorescence protein, and used as the mitochondrial marker in Physcomitrella protonemal tissues.

Biolistic transformation and microscopy

Biolistic co-transformation of the GFP and RFP fusion vectors was performed on Arabidopsis cell suspensions and onion epidermal cells as previously described (Carrie et al., 2009b). Briefly, 5 µg of GFP and RFP/mCherry plasmids were co-precipitated onto gold particles and bombarded onto 4-day-old Arabidopsis cell suspensions and freshly peeled
onion epidermal cells using the PDS-1000/He biolistic transformation system (Bio-Rad, Sydney). For putative Physcomitrella proteins, transformation was also performed on 7-day-old protonemal tissues. Following incubation for 12-24 h at 22°C (25 °C for Physcomitrella) in the dark, GFP and RFP/mCherry expression was visualized at 100X magnification using a BX61 Olympus microscope (Olympus, Melbourne) with the excitation wavelengths of 460/480 nm (GFP) and 535/555 nm (RFP/mCherry), and emission wavelengths of 495–540 nm (GFP) and 570–625 nm (RFP/mCherry). Images were captured using CellR imaging software (Olympus, Melbourne) as previously described (Carrie et al., 2009b).

Acknowledgments

This work was supported by a WA state government Postgraduate Scholarship to L.X., and the Australian Research Council Discovery Grant DP0664692 to JW. CC is supported by a Humboldt Research Fellowship for Postdoctoral Researchers from the Alexander von Humboldt Foundation. We would like to thank Dr Olivier Keech for his kind donation of Picea RNA.

Author Contributions

JW and CC designed the experiments, LX, SL and CC performed the experimental procedures, LX, CC, MWM and carried out the data analysis, and all authors contributed to the writing of the manuscript.
Figure Legends

Figure 1. Experimental design to investigate dual targeting of proteins in land plants. A, Tree diagram showing the approximate time (in millions of years ago) that the major group of plants diverged. The four species of land plants used in this study are shown, Arabidopsis thaliana (Arabidopsis), Oryza sativa (rice), Picea glauca (Picea) and Physcomitrella patens (Physcomitrella). Additionally, Chlamydomonas reinhardtii (Cr) and Chlorella variabilis (Cv), were included in the phylogenetic analysis. B, Verifying the Physcomitrella derived organelle specific markers for mitochondria (alpha subunit of the mitochondrial processing peptidase = MPP alpha, Alternative oxidase = AOX), plastids (Small subunit of ribulose 1, 5 bisphosphate carboxylase/oxygenase = RBCS1, Subunit 2 of photosystem I = PSI subunit 2) and peroxisomes (Thiolase and Malate synthase) fused to GFP. Accession numbers are indicated. The biolistic transformation of each Physcomitrella derived organelle marker was carried out with previously published organelle marker sets, the mitochondrial (Alternative oxidase = AOX), plastid (Small Subunit of RBCS = SSU) and peroxisomal (Cucubita sp. malate synthase = SRL) targeting signal fused to RFP (Carrie et al., 2009b). In addition, COXIV-mCherry (Nelson et al., 2007) and PpAOX-mCherry were used as mitochondrial marker in onion epidermal cells and Physcomitrella protonemal tissues, respectively (Cytochrome c oxidase IV = COXIV). Targeting was tested in Arabidopsis cell suspensions, onion epidermal cells and Physcomitrella protonemal tissues. Scale bar indicates 20 μM. Mitochondria (M), plastids (Pl) and peroxisomes (Px) are indicated respectively.

Figure 2. Dual targeting of DNA Topoisomerase to mitochondria and plastids. A, Phylogenetic analysis of genes encoding DNA Topoisomerases (Top) from Arabidopsis (At) rice (Os) Physcomitrella (Pp), Picea (Pg), Chlamydomonas (Cr) and Chlorella (Cv) using MEGA 5 (see methods). B, Table of Top proteins from Arabidopsis, rice, Physcomitrella, Picea, Chlamydomonas and Chlorella with genomic loci numbers, targeting predictions (see methods) and experimental localization based on GFP targeting. C, GFP images of the targeting ability of Top proteins tested, dual targeting of AtTopIa1 and OsTopI
was evident in Arabidopsis cell suspensions and onion epidermal cells. In contrast, dual targeting of PpTop was only evident in onion epidermal cells and Physcomitrella protonemal tissues. Mitochondria (M) and plastids (Pl) are indicated respectively. Scale bar indicates 20 μM. M, Mitochondria; Pl, plastids; Cyto, cytosol; N, nuclear; S, secretory pathway.

Figure 3. Dual targeting of DNA Helicase to mitochondria and plastids. A, Phylogenetic analysis of genes encoding DNA Helicase (Hel) from Arabidopsis (At) Rice (Os) Physcomitrella (Pp), Chlamydomonas (Cr) and Chlorella (Cv) using MEGA 5 (see methods). There was no ortholog found in Picea. B, Table of Hel proteins from Arabidopsis, rice, Physcomitrella, Chlamydomonas and Chlorella with genomic loci numbers, targeting predictions and experimental localization based on GFP tagging. C, GFP images of the targeting ability of Hel proteins tested. Dual targeting of OsHel and PpHel1 was evident in Arabidopsis cell suspensions, onion epidermal cells and Physcomitrella protonemal tissues. Mitochondria (M) and plastids (Pl) are indicated respectively. Scale bar indicates 20 μM. M, Mitochondria; Pl, plastids; Cyto, cytosol; ER, endoplasmic reticulum; N, nuclear.

Figure 4. Dual targeting of DNA Polymerase to mitochondria and plastids. A, Phylogenetic analysis of genes encoding DNA Polymerases (Pol) from Arabidopsis (At) Rice (Os) Physcomitrella (Pp), Picea (Pg) and Chlamydomonas (Cr) using MEGA 5 (see methods). There was no ortholog found in Chlorella. B, Table of the Pol proteins from Arabidopsis, rice, Physcomitrella, Picea and Chlamydomonas with genomic loci numbers, predictions using a variety of prediction programs, experimental localization based on GFP tagging. C, GFP images of targeting ability of Pol proteins tested. Dual targeting of OsPol1 was evident in all three cells types tested. Only plastid targeting could be detected for PpPol1 in all three tissues, while plastid and mitochondrial targeting was detected for PpPol2 in onion epidermal cells and Physcomitrella protonemal tissues. Mitochondria (M) and plastids (Pl) are indicated respectively. Scale bar indicates 20 μM. M, Mitochondria; Pl, plastids; Cyto, cytosol; N, nuclear.
**Figure 5.** Dual targeting of the enzymes of the ascorbate glutathione cycle. A, Overview of the ascorbate glutathione cycle. B, Phylogenetic analysis of the genes encoding glutathione reductase from Arabidopsis (At), Rice (Os), Picea (Pg), Physcomitrella (Pp), Chlamydomonas (Cr) and Chlorella (Cv) using MEGA 5. C, Table of GR proteins from Arabidopsis, rice, Physcomitrella, Picea, Chlamydomonas and Chlorella with genomic loci numbers, targeting predictions and experimental localization based on GFP tagging. D) GFP images of targeting ability of GR proteins tested. Dual targeting of GR from Arabidopsis, rice and Physcomitrella was detected in all tissues tested. Mitochondria (M) and plastids (Pl) are indicated respectively. Scale bar indicates 20 µM. GR, Glutathione reductase; DHAR, dehydroascorbate reductase; MDHAR, Monodehydroascorbate reductase; APX, Ascorbate reductase; ASC, ascorbate; DHA, Dehydroascorbate; GSH, Glutathione; GSSG, Glutathione disulfide; M, Mitochondria; Pl, plastids; Cyto, cytosol.

**Figure 6.** Dual targeting of Ascorbate Peroxidase to mitochondria and plastids. A, Phylogenetic analysis of genes encoding Ascorbate Peroxidase (APX) from Arabidopsis (At), Rice (Os), Picea (Pg), Physcomitrella (Pp) and Chlamydomonas (Cr) using MEGA 5 (see methods). There was no ortholog found in Chlorella. B, GFP images of targeting ability of APX proteins tested. Dual targeting of AtSAPX was evident in Arabidopsis cell suspensions and onion epidermal cells. Four rice APX proteins were tested, OsAPX5 and OsAPX6 showed targeting to mitochondria, while OsAPX7 and OsAPX8, showed targeting to plastids. PgAPX1 showed targeting to both mitochondria and plastids in onion epidermal cells, while PpAPX1 showed targeting to plastids only in all tissues tested. C, Table of the APX proteins from Arabidopsis, rice and Physcomitrella with genomic loci numbers, targeting predictions and experimental localization based on GFP targeting. Mitochondria (M) and plastids (Pl) are indicated respectively. Scale bar indicates 20 µM. M, Mitochondria; Pl, plastids; Cyto, cytosol; ER, endoplasmic reticulum.

**Figure 7.** Dual targeting of Monodehydroascorbate Reductase to mitochondria and plastids. A, Phylogenetic analysis of genes encoding
Monodehydroascorbate Reductase (MDHAR) from Arabidopsis (At), Rice (Os), Picea (Pg), Physcomitrella (Pp) and Chlamydomonas (Cr) using MEGA 5 (see methods). There is no ortholog found in Chlorella. B, GFP images of targeting ability of the MDHAR proteins tested. Dual targeting of AtMDHAR6 was evident in Arabidopsis cell suspensions and onion epidermal cells. While OsMDHAR1.1 displayed dual targeting to mitochondria and plastids, OsMDAHR1.2 only showed mitochondrial targeting ability. Likewise for both Physcomitrella and Picea, of the two proteins tested for each species only one (PgMDHAR2 and PpMDHAR2 for Picea and Physcomitrella respectively) displayed dual targeting ability. C, Table of the MDHAR proteins from Arabidopsis, rice, Picea, Physcomitrella and Chlamydomonas with genomic loci numbers, targeting predictions and experimental localization based on GFP tagging. Mitochondria (M) and plastids (Pl) are indicated respectively. Scale bar indicates 20 µM. M, Mitochondria; Pl, plastids; Cyto, cytosol; ER, endoplasmic reticulum; EX, extracellular; S, secretory pathway.

**Figure 8.** Dual targeting of Hexokinase to mitochondria and plastids. A, Phylogenetic analysis of genes encoding Hexokinase (HXK) from Arabidopsis (At), Rice (Os), Picea (Pg), Physcomitrella (Pp), Chlamydomonas (Cr) and Chlorella (Cv) using MEGA 5 (see methods). B, Table of the HXK proteins from Arabidopsis, rice, Picea, Physcomitrella, Chlamydomonas and Chlorella with genomic loci numbers, targeting predictions and experimental localization based on GFP tagging. C, GFP tagging of several HXK proteins from Arabidopsis showed that no dual targeting was evident for any HXK protein in Arabidopsis. D and E, GFP tagging of several HXK proteins from Physcomitrella revealed that PpHXK 2,3,5,7,9 and 11 displayed dual targeting ability. Mitochondria (M) and plastids (Pl) are indicated respectively. Scale bar indicates 20 µM. M, Mitochondria; Pl, plastids; Cyto, cytosol; ER, endoplasmic reticulum; EX, extracellular; S, secretory pathway.

**Figure 9.** Targeting ability of Mia40 in Physcomitrella. A, Phylogenetic analysis of genes encoding Mia40 in Arabidopsis (At), Rice (Os), Picea (Pg), Physcomitrella (Pp), Chlamydomonas (Cr) and Chlorella (Cv) using MEGA 5 (see methods). B, Table of the Mia40 proteins from Arabidopsis, rice, Picea, Physcomitrella, Chlamydomonas and Chlorella with genomic loci numbers, targeting predictions and experimental localization based on GFP tagging. C, GFP tagging of several Mia40 proteins from Arabidopsis revealed that no dual targeting was evident for any Mia40 protein in Arabidopsis. D and E, GFP tagging of several Mia40 proteins from Physcomitrella revealed that PpMia40 2, 3, 7, 9 and 11 displayed dual targeting ability. Mitochondria (M) and plastids (Pl) are indicated respectively. Scale bar indicates 20 µM. M, Mitochondria; Pl, plastids; Cyto, cytosol; ER, endoplasmic reticulum; EX, extracellular; S, secretory pathway.
Physcomitrella, Picea, Chlamydomonas and Chlorella with genomic loci numbers, targeting prediction and experimental localization based on GFP tagging. C, GFP tagging of Physcomitrella Mia40, displaying cytosolic targeting in all tissues tested. Scale bar indicates 20 µM. M, Mitochondria; Px, peroxisomes; Pl, plastids; Cyto, cytosol; EX, extracellular; N, nuclear; Cyto, cytosol.

**Figure 10.** The evolution of targeting of Mia40 and putative substrate proteins. In plants such as *Arabidopsis thaliana*, *Glycine max* and *Poplar trichocarpa* Mia40 is dual targeted to mitochondria and peroxisomes. This dual targeting is also accompanied by the targeting of Mia40 substrates to the same organelles. However in plants such as *Chlamydomonas reinhardtii*, *Volvox carteri*, *Chlorella variabilis*, *Physcomitrella patens* and *Picea glauca* Mia40 is not predicted to be targeted to any organelle, and this was confirmed for Physcomitrella in this study (Figure 9). This is accompanied by either the absence of substrate proteins or the absence of targeting of substrate proteins to peroxisomes. Monocot species such as *Oryza sativa*, *Brachypodium distachyon* and *Zea mays* appear to be an intermediate as they contain a dual-targeted Mia40 but lack substrate proteins in the peroxisome. This suggests how dual targeting of one protein may facilitate the acquisition of whole metabolic pathways between different organelles. Ccs, Copper chaperone for superoxide dismutase; CSD, copper zinc superoxide dismutase. See Supplemental Figure S3 for sequence alignment of Mia40 from various plants and the predicted targeting signals highlighted.
**Supplemental Table S1.** List of all known dual-targeted proteins to mitochondria and plastids from Arabidopsis and their orthologs in other plants. Proteins are listed by gene family. The localization of each protein is indicated as mitochondrial (M), plastid (Pl), cytosolic (Cyto) or nuclear (N) as determined by previous studies and/or GFP tagging in this study. Green shading indicates proteins dual-targeted in Arabidopsis, rice and Physcomitrella. Yellow shading indicates proteins dual-targeted in Arabidopsis and rice.

**Supplemental Table S2.** List of all known dual-targeted proteins to mitochondria and peroxisomes or plastids and peroxisomes from Arabidopsis and their orthologs in other plants. Proteins are listed by gene family. The localization of each protein is indicated as mitochondria (M), plastid (Pl), peroxisomes (Px), endoplasmic reticulum (ER) or cytosol (Cyto) as determined by previous studies and/or GFP tagging in this study. Green shading indicates proteins dual-targeted in Arabidopsis, rice and Physcomitrella.

**Supplemental Table S3.** List of primers and cDNA clones used in this study.

**Supplemental dataset.** Percentage of similarity and identity scores generated by MatGAT for protein sequences in gene families.
Supplemental Figures S1.1-S1.31. Targeting ability of orthologs of known Arabidopsis dual-targeted proteins to mitochondria and plastids.

Supplemental Figures S2.1-S2.3. Targeting ability of orthologs to known Arabidopsis dual-targeted proteins to mitochondria and peroxisomes or plastids and peroxisomes.

Supplemental Figure S3. Sequence alignment of orthologs of Mia40 proteins from different plant species. Species are Arabidopsis thaliana (At), Glycine max (Gm), Populus trichocarpa (Pt), Zea mays (Zm), Brachyodium distachyon (Bd), Oryza sativa (Os), Picea glauca (Pg), Physcomitrella patens (Pp), Chlorella variabilis (Cv), Chlamydomonas reinhardtii (Cr) and Volvox carteri (Vc). Arrows indicate conserved cysteine residues. The 28 amino acids present in Arabidopsis but absent in Physcomitrella are highlighted by the red box. Putative PTS-1 sequences are underlined in blue.

Supplemental Figure S4. Sequence alignment of the two isoforms of rice MDHAR1 (LOC_Os08g05570.1 and LOC_Os08g05570.2). OsMDHAR1.1 was dual-targeted while OsMDHAR1.2 was only targeted to mitochondria (Figure 7). OsMDHAR1.1 contains four extra amino acids, highlighted in red.

Supplemental Figure S5. Sequence alignment of Physcomitrella MDHAR1 and 2. PpMDHAR2 was dual-targeted and PpMDHAR1 was cytosolic (Figure 7). Black arrow indicates the putative cleavage site of targeting signal for PpMDHAR2 predicted by TargetP (see methods). No targeting signal was predicted in PpMDHAR1. Four amino acids differences between PpMDHAR1 and 2 are indicated in red.
References

Adams, K.L., Daley, D.O., Qiu, Y.L., Whelan, J., and Palmer, J.D. (2000). Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. Nature 408, 354-357.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of molecular biology 215, 403-410.

Bahaji, A., Munoz, F.J., Ovecka, M., Baroja-Fernandez, E., Montero, M., Li, J., Hidalgo, M., Almagro, G., Sesma, M.T., Ezquer, I., and Pozueta-Romero, J. (2011a). Specific delivery of AtBT1 to mitochondria complements the aberrant growth and sterility phenotype of homozygous Atbt1 Arabidopsis mutants. Plant J 68, 1115-1121.

Bahaji, A., Ovecka, M., Barany, I., Risueno, M.C., Munoz, F.J., Baroja-Fernandez, E., Montero, M., Li, J., Hidalgo, M., Sesma, M.T., Ezquer, I., Testillano, P.S., and Pozueta-Romero, J. (2011b). Dual targeting to mitochondria and plastids of AtBT1 and ZmBT1, two members of the mitochondrial carrier family. Plant Cell Physiol 52, 597-609.

Beardslee, T.A., Roy-Chowdhury, S., Jaiswal, P., Buhot, L., Lerbs-Mache, S., Stern, D.B., and Allison, L.A. (2002). A nucleus-encoded maize protein with sigma factor activity accumulates in mitochondria and chloroplasts. Plant J 31, 199-209.

Blanc, G., Duncan, G., Agarkova, I., Borodovsky, M., Gurnon, J., Kuo, A., Lindquist, E., Lucas, S., Pangilinan, J., Polle, J., Salamov, A., Terry, A., Yamada, T., Dunigan, D.D., Grigoriev, I.V., Claverie, J.M., and Van Etten, J.L. (2010). The Chlorella variabilis NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. Plant Cell 22, 2943-2955.

Bowman, J.L., Floyd, S.K., and Sakakibara, K. (2007). Green genomics comparative genomics of the green branch of life. Cell 129, 229-234.

Carrie, C., and Small, I. (2012). A reevaluation of dual-targeting of proteins to mitochondria and chloroplasts. Biochim Biophys Acta PMID: 22683762

Carrie, C., Giraud, E., and Whelan, J. (2009a). Protein transport in organelles: Dual targeting of proteins to mitochondria and chloroplasts. FEBS J 276, 1187-1195.

Carrie, C., Murcha, M.W., and Whelan, J. (2010a). An in silico analysis of the mitochondrial protein import apparatus of plants. BMC Plant Biol 10, 249.

Carrie, C., Murcha, M.W., Millar, A.H., Smith, S.M., and Whelan, J. (2007). Nine 3-ketoacyl-CoA thiolases (KATs) and acetoacetyl-CoA thiolases (ACATs) encoded by five genes in Arabidopsis thaliana are targeted either to peroxisomes or cytosol but not to mitochondria. Plant Mol Biol 63, 97-109.
Carrie, C., Kuhn, K., Murcha, M.W., Duncan, O., Small, I.D., O'Toole, N., and Whelan, J. (2009b). Approaches to defining dual-targeted proteins in Arabidopsis. Plant J 57, 1128-1139.

Carrie, C., Murcha, M.W., Kuehn, K., Duncan, O., Barthe, M., Smith, P.M., Eubel, H., Meyer, E., Day, D.A., Millar, A.H., and Whelan, J. (2008). Type II NAD(P)H dehydrogenases are targeted to mitochondria and chloroplasts or peroxisomes in Arabidopsis thaliana. FEBS Lett 582, 3073-3079.

Carrie, C., Giraud, E., Duncan, O., Xu, L., Wang, Y., Huang, S., Clifton, R., Murcha, M., Filipovska, A., Rackham, O., Vrielink, A., and Whelan, J. (2010b). Conserved and novel functions for Arabidopsis thaliana MIA40 in assembly of proteins in mitochondria and peroxisomes. J Biol Chem 285, 36138-36148.

Chacinska, A., Pfannschmidt, S., Wiedemann, N., Kozjak, V., Sanjuan Szklarz, L.K., Schulze-Specking, A., Truscott, K.N., Guiard, B., Meisinger, C., and Pfanner, N. (2004). Essential role of Mia40 in import and assembly of mitochondrial intermembrane space proteins. The EMBO journal 23, 3735-3746.

Chew, O., and Whelan, J. (2004). Just read the message: a model for sorting of proteins between mitochondria and chloroplasts. Trend Plant Sci 9, 318-319.

Chew, O., Whelan, J., and Millar, A.H. (2003). Molecular definition of the ascorbate-glutathione cycle in Arabidopsis mitochondria reveals dual targeting of antioxidant defenses in plants. J Biol Chem 278, 46869-46877.

Chew, O., Lister, R., Qbadou, S., Heazlewood, J.L., Soll, J., Schleiff, E., Millar, A.H., and Whelan, J. (2004). A plant outer mitochondrial membrane protein with high amino acid sequence identity to a chloroplast protein import receptor. FEBS Lett 557, 109-114.

Creissen, G., Reynolds, H., Xue, Y., and Mullineaux, P. (1995). Simultaneous targeting of pea glutathione reductase and of a bacterial fusion protein to chloroplasts and mitochondria in transgenic tobacco. Plant J 8, 167-175.

Dolezal, P., Likic, V., Tachezy, J., and Lithgow, T. (2006). Evolution of the molecular machines for protein import into mitochondria. Science 313, 314-318.

Duncan, O., Murcha, M.W., and Whelan, J. (2012). Unique components of the plant mitochondrial protein import apparatus. Biochim Biophys Acta PMID: 22406071

Dyall, S.D., Brown, M.T., and Johnson, P.J. (2004). Ancient invasions: from endosymbionts to organelles. Science 304, 253-257.

Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. Nature protocols 2, 953-971.
Glaser, E., and Whelan, J. (2007). Import of nuclear-encoded mitochondrial proteins. In Plant Mitochondria D.C. Logan, ed (Oxford: Blackwell), pp. 97-128.

Gray, M.W., Burger, G., and Lang, B.F. (2001). The origin and early evolution of mitochondria. Genome biology 2, REVIEWS1018.

Heazlewood, J.L., Verboom, R.E., Tonti-Filippini, J., Small, I., and Millar, A.H. (2007). SUBA: the Arabidopsis Subcellular Database. Nucleic acids research 35, D213-218.

Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., and Nakai, K. (2007). WoLF PSORT: protein localization predictor. Nucleic acids research 35, W585-587.

Huang, S., Taylor, N.L., Whelan, J., and Millar, A.H. (2009). Refining the definition of plant mitochondrial presequences through analysis of sorting signals, N-terminal modifications, and cleavage motifs. Plant Physiol 150, 1272-1285.

Jarvis, P. (2008). Targeting of nucleus-encoded proteins to chloroplasts in plants. New Phytol 179, 257-285.

Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic acids research 33, 511-518.

Keeling, P.J. (2010). The endosymbiotic origin, diversification and fate of plastids. Philos Trans R Soc Lond B Biol Sci 365, 729-748.

Kleine, T., Maier, U.G., and Leister, D. (2009). DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. Annu Rev Plant Biol 60, 115-138.

Kuhn, K., Richter, U., Meyer, E.H., Delannoy, E., de Longevialle, A.F., O'Toole, N., Borner, T., Millar, A.H., Small, I.D., and Whelan, J. (2009). Phage-type RNA polymerase RPOTmp performs gene-specific transcription in mitochondria of Arabidopsis thaliana. Plant Cell 21, 2762-2779.

Lister, R., Carrie, C., Duncan, O., Ho, L.H., Howell, K.A., Murcha, M.W., and Whelan, J. (2007). Functional definition of outer membrane proteins involved in preprotein import into mitochondria. Plant Cell 19, 3739-3759.

Martin, W. (2010). Evolutionary origins of metabolic compartmentalization in eukaryotes. Philos Trans R Soc Lond B Biol Sci 365, 847-855.

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Marechal-Drouard, L., Marshall, W.F., Qu, L.H., Nelson, D.R., Sanderfoot, A.A., Spalding, M.H., Kapitonov, V.V., Ren, Q., Ferris, P., Lindquist, E., Shapiro, H., Lucas, S.M., Grimwood, J., Schmutz, J., Cardol, P., Cerutti, H., Chanfreau, G., Chen, C.L., Cognat, V., Croft, M.T., Dent, R., Dutcher, S., Fernandez, E., Fukuzawa, H., Gonzalez-Ballesta, D., Gonzalez-Halphen, D., Hallmann, A., Hanikenne, M., Hippler, M., Inwood, W., Jabbari, K., Kalanon, M., Kuras, R., Lefebvre, P.A., Lemaitre, S.T., Lobanov, O., Lohn, M., Manne, A., Meier, I., Mets,
L., Mittag, M., Mittelmeier, T., Moroney, J.V., Moseley, J., Napoli, C., Nedelcu, A.M., Niyogi, K., Novoselov, S.V., Paulsen, I.T., Pazour, G., Purton, S., Ral, J.P., Riano-Pachon, D.M., Riekhof, W., Rymarquis, L., Schroda, M., Stern, D., Umen, J., Willows, R., Wilson, N., Zimmer, S.L., Allmer, J., Balk, J., Bisova, K., Chen, C.J., Elias, M., Gendler, K., Hauser, C., Lamb, M.R., Ledford, H., Long, J.C., Minagawa, J., Page, M.D., Pan, J., Pootakham, W., Roje, S., Rose, A., Stahlberg, E., Terauchi, A.M., Yang, P., Ball, S., Bowler, C., Dieckmann, C.L., Gladyshev, V.N., Green, P., Jorgensen, R., Mayfield, S., Mueller-Roeber, B., Rajamani, S., Sayre, R.T., Brokstein, P., Dubchak, I., Goodstein, D., Hornick, L., Huang, Y.W., Jhaveri, J., Luo, Y., Martinez, D., Ngau, W.C., Otillar, B., Poliakov, A., Porter, A., Szajkowski, L., Werner, G., Zhou, K., Grigoriev, I.V., Rokhsar, D.S., and Grossman, A.R. (2007). The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science 318, 245-250.

Millar, A.H., Carrie, C., Pogson, B., and Whelan, J. (2009). Exploring the function-location nexus: using multiple lines of evidence in defining the subcellular location of plant proteins. Plant Cell 21, 1625-1631.

Mitschke, J., Fuss, J., Blum, T., Hoglund, A., Reski, R., Kohlbacher, O., and Rensing, S.A. (2009). Prediction of dual protein targeting to plant organelles. New Phytol 183, 224-235.

Morgante, C.V., Rodrigues, R.A., Marbach, P.A., Borgonovi, C.M., Moura, D.S., and Silva-Filho, M.C. (2009). Conservation of dual-targeted proteins in Arabidopsis and rice points to a similar pattern of gene-family evolution. Mol Genet Genom 281, 525-538.

Nelson, B.K., Cai, X., and Nebenfuhr, A. (2007). A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. Plant J 51, 1126-1136.

Nilsson, A., Olsson, T., Ulfstedt, M., Thelander, M., and Ronne, H. (2011). Two novel types of hexokinases in the moss Physcomitrella patens. BMC Plant Biol 11, 32.

Peeters, N., and Small, I. (2001). Dual targeting to mitochondria and chloroplasts. Biochim Biophys Acta 1541, 54-63.

Pfannschmidt, T. (2010). Plastidial retrograde signalling—a true "plastid factor" or just metabolite signatures? Trend Plant Sci 15, 427-435.

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.F., Lindquist, E.A., Kamisugi, Y., Tanahashi, T., Sakakibara, K., Fujita, T., Oishi, K., Shin, I.T., Kuroki, Y., Toyoda, A., Suzuki, Y., Hashimoto, S., Yamaguchi, K., Sugano, S., Kohara, Y., Fujiyama, A., Anterola, A., Aoki, S., Ashton, N., Barbazuk, W.B., Barker, E., Bennetzen, J.L., Blankenship, R., Cho, S.H., Dutcher, S.K., Estelle, M., Fawcett, J.A., Gundlach, H., Hanada, K., Heyl, A., Hicks, K.A., Hughes, J., Lohr, M., Mayer, K., Melkozernov, A., Murata, T., Nelson, D.R., Pils, B., Prigge, M., Reiss, B., Renner, T., Rombauts, S., Rushton, P.J., Sanderfoot, A., Schween, G., Shin, S.H., Stueber, K., Theodoulou,
F.L., Tu, H., Van de Peer, Y., Verrier, P.J., Waters, E., Wood, A., Yang, L., Cove, D., Cuming, A.C., Hasebe, M., Lucas, S., Mishler, B.D., Reski, R., Grigoriev, I.V., Quatrano, R.S., and Boore, J.L. (2008). The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science 319, 64-69.

Rudhe, C., Chew, O., Whelan, J., and Glaser, E. (2002). A novel in vitro system for simultaneous import of precursor proteins into mitochondria and chloroplasts. Plant J 30, 213-220.

Small, I., Peeters, N., Legeai, F., and Lurin, C. (2004). Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics 4, 1581-1590.

Soll, J., and Schleife, E. (2004). Protein import into chloroplasts. Nat Rev Mol Cell Biol 5, 198-208.

Suzuki, K., and Miyagishima, S.Y. (2010). Eukaryotic and eubacterial contributions to the establishment of plastid proteome estimated by large-scale phylogenetic analyses. Mol Biol Evol 27, 581-590.

Szklarczyk, R., and Huynen, M.A. (2010). Mosaic origin of the mitochondrial proteome. Proteomics 10, 4012-4024.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28, 2731-2739.

Xu, Y.Z., Arrieta-Montiel, M.P., Virdi, K.S., de Paula, W.B., Widhalm, J.R., Basset, G.J., Davila, J.L., Elthon, T.E., Elowsky, C.G., Sato, S.J., Clemente, T.E., and Mackenzie, S.A. (2011). MutS HOMOLOG1 is a nucleoid protein that alters mitochondrial and plastid properties and plant response to high light. Plant Cell 23, 3428-3441.

Xu, Y.Z., De la Rosa Santamaria, R., Virdi, K.S., Arrieta-Montiel, M.P., Razvi, F., Li, S., Ren, G., Yu, B., Alexander, D., Guo, L., Feng, X., Dweikat, I.M., Clemente, T.E., and Mackenzie, S.A. (2012). The chloroplast triggers developmental reprogramming when MUTS HOMOLOG1 is suppressed in plants. Plant Physiol 159, 710-720

Yogev, O., and Pines, O. (2011). Dual targeting of mitochondrial proteins: mechanism, regulation and function. Biochim Biophys Acta 1808, 1012-1020.
Figure 1. Experimental design to investigate dual targeting of proteins in land plants. A, Tree diagram showing the approximate time (in millions of years ago) that the major group of plants diverged. The four species of land plants used in this study are shown, Arabidopsis thaliana (Arabidopsis), Oryza sativa (rice), Picea glauca (Picea) and Physcomitrella patens (Physcomitrella). Additionally, Chlamydomonas reinhardtii (Cr) and Chlorella variabilis (Cv), were included in the phylogenetic analysis. B, Verifying the Physcomitrella derived organelle specific markers for mitochondria (alpha subunit of the mitochondrial processing peptidase = MPP alpha, Alternative oxidase = AOX), plastids (Small subunit of ribulose 1, 5 bisphosphate carboxylase/oxygenase = RBCS1, Subunit 2 of photosystem I = PSI subunit 2) and peroxisomes (Thiolase and Malate synthase) fused to GFP. Accession numbers are indicated. The biolistic transformation of each Physcomitrella derived organelle marker was carried out with previously published organelle marker sets, the mitochondrial (Alternative oxidase = AOX), plastid (Small Subunit of RBCS = SSU) and peroxisomal (Cucubita sp. malate synthase = SRL) targeting signal fused to RFP (Carrie et al., 2009b). In addition, COXIV-mCherry (Nelson et al., 2007) and PpAOX-mCherry were used as mitochondrial marker in onion epidermal cells and Physcomitrella protonemal tissues, respectively (Cytochrome c oxidase IV = COXIV). Targeting was tested in Arabidopsis cell suspensions, onion epidermal cells and Physcomitrella protonemal tissues. Scale bar indicates 20 μM. Mitochondria (M), plastids (Pl) and peroxisomes (Px) are indicated respectively.
Figure 2. Dual targeting of DNA Topoisomerase to mitochondria and plastids. A, Phylogenetic analysis of genes encoding DNA Topoisomerases (Top) from Arabidopsis (At) rice (Os) Physcomitrella (Pp), Picea (Pg), Chlamydomonas (Cr) and Chlorella (Cv) using MEGA 5 (see methods). B, Table of Top proteins from Arabidopsis, rice, Physcomitrella, Picea, Chlamydomonas and Chlorella with genomic loci numbers, targeting predictions (see methods) and experimental localization based on GFP targeting. C, GFP images of the targeting ability of Top proteins tested. Dual targeting of AtTopIA1 and OsTopI was evident in Arabidopsis cell suspensions and onion epidermal cells. In contrast, dual targeting of PpTop was only evident in onion epidermal cells and Physcomitrella protonemal tissues. Mitochondria (M) and plastids (Pl) are indicated respectively. Scale bar indicates 20 μM. M, Mitochondria; Pl, plastids; Cyto, cytosol; N, nuclear; S, secretory pathway.
Figure 3. Dual targeting of DNA Helicase to mitochondria and plastids. A, Phylogenetic analysis of genes encoding DNA Helicase (Hel) from Arabidopsis (At) Rice (Os) Physcomitrella (Pp), Chlamydomonas (Cr) and Chlorella (Cv) using MEGA 5 (see methods). There was no ortholog found in Picea. B, Table of Hel proteins from Arabidopsis, rice, Physcomitrella, Chlamydomonas and Chlorella with genomic loci numbers, targeting predictions and experimental localization based on GFP tagging. C, GFP images of the targeting ability of Hel proteins tested. Dual targeting of OsHel and PpHel1 was evident in Arabidopsis cell suspensions, onion epidermal cells and Physcomitrella protonemal tissues. Mitochondria (M) and plastids (Pl) are indicated respectively. Scale bar indicates 20 μM. M, Mitochondria; Pl, plastids; Cyto, cytosol; ER, endoplasmic reticulum; N, nuclear.
**Figure 4.** Dual targeting of DNA Polymerase to mitochondria and plastids. A, Phylogenetic analysis of genes encoding DNA Polymases (Pol) from Arabidopsis (At) Rice (Os) Physcomitrella (Pp), Picea (Pg) and Chlamydomonas (Cr) using MEGA 5 (see methods). There was no ortholog found in Chlorella. B, Table of the Pol proteins from Arabidopsis, rice, Physcomitrella, Picea and Chlamydomonas with genomic loci numbers, predictions using a variety of prediction programs, experimental localization based on GFP tagging. C, GFP images of targeting ability of Pol proteins tested. Dual targeting of OsPol1 was evident in all three cells types tested. Only plastid targeting could be detected for PpPol1 in all three tissues, while plastid and mitochondrial targeting was detected for PpPol2 in onion epidermal cells and Physcomitrella protonemal tissues. Mitochondria (M) and plastids (Pl) are indicated respectively. Scale bar indicates 20 μM. M, Mitochondria; Pl, plastids; Cyto, cytosol; N, nuclear.
**Figure 5.** Dual targeting of the enzymes of the ascorbate glutathione cycle. A, Overview of the ascorbate glutathione cycle. B, Phylogenetic analysis of the genes encoding glutathione reductase from Arabidopsis (At), Rice (Os), Picea (Pg), Physcomitrella (Pp), Chlamydomonas (Cr) and Chlorella (Cv) using MEGA 5. C, Table of GR proteins from Arabidopsis, rice, Physcomitrella, Picea, Chlamydomonas and Chlorella with genomic loci numbers, targeting predictions and experimental localization based on GFP tagging. D) GFP images of targeting ability of GR proteins tested. Dual targeting of GR from Arabidopsis, rice and Physcomitrella was detected in all tissues tested. Mitochondria (M) and plastids (Pl) are indicated respectively. Scale bar indicates 20 μM. GR, Glutathione reductase; DHAR, dehydroascorbate reductase; MDHAR, Monodehydroascorbate reductase; APX, Ascorbate reductase; ASC, ascorbate; DHA, Dehydroascorbate; GSH, Glutathione; GSSG, Glutathione disulphide; M, Mitochondria; Pl, plastids; Cyto, cytosol.

| Protein name | Accession/Locus | Prediction | Experimental localization | References |
|--------------|----------------|------------|--------------------------|------------|
| AtGR2       | At3g54660      | PI         | Pl PI                    | M + PI     | Chew et al. 2003 |
| AtGR2       | At3g24170      | -          | Pl                       | 0.29       |            |
| OsGR1       | LOC_Os03g06740 | PI         | Pl PI                    | 0.73       | M + PI     | this study |
| OsGR2       | LOC_Os10g28000 | -          | Cyto                     | 0.47       |            |
| OsGR3       | LOC_Os02g56850 | -          | Cyto                     | 0.25       |            |
| PgGR1       | BT109833       | -          | -                        | 0.31       |            |
| PgGR2       | BT113073       | -          | PI                       | 0.26       |            |
| PgGR3       | BT107715       | M          | M Pl                     | M          |            |
| PgGR3       | Pp1s13_127V6   | M          | PI                       | 0.60       | M + PI     | this study |
| PgGR3       | Pp1s116_56V6   | -          | Cyto                     | 0.41       |            |
| CrGR1       | Cre02.g132850  | -          | M Pl                     | 0.39       |            |
| CrGR2       | Cre06.g262100  | -          | Cyto                     | 0.62       |            |
| CvGR1       | 34977          | M          | M Pl                     | 0.51       |            |
| CvGR2       | 137376         | -          | Cyto                     | 0.62       |            |

**Table C:**

| Protein name | Accession/Locus | Prediction | Experimental localization | References |
|--------------|----------------|------------|--------------------------|------------|
| AtGR2       | At3g54660      | PI         | Pl PI                    | M + PI     | Chew et al. 2003 |
| AtGR2       | At3g24170      | -          | Pl                       | 0.29       |            |
| OsGR1       | LOC_Os03g06740 | PI         | Pl PI                    | 0.73       | M + PI     | this study |
| OsGR2       | LOC_Os10g28000 | -          | Cyto                     | 0.47       |            |
| OsGR3       | LOC_Os02g56850 | -          | Cyto                     | 0.25       |            |
| PgGR1       | BT109833       | -          | -                        | 0.31       |            |
| PgGR2       | BT113073       | -          | PI                       | 0.26       |            |
| PgGR3       | BT107715       | M          | M Pl                     | M          |            |
| PgGR3       | Pp1s13_127V6   | M          | PI                       | 0.60       | M + PI     | this study |
| PgGR3       | Pp1s116_56V6   | -          | Cyto                     | 0.41       |            |
| CrGR1       | Cre02.g132850  | -          | M Pl                     | 0.39       |            |
| CrGR2       | Cre06.g262100  | -          | Cyto                     | 0.62       |            |
| CvGR1       | 34977          | M          | M Pl                     | 0.51       |            |
| CvGR2       | 137376         | -          | Cyto                     | 0.62       |            |
| Protein name | Accession/Locus | Prediction | Experimental | ATP | References |
|--------------|----------------|------------|--------------|-----|------------|
| AtSAPX      |                |            |              |     |            |
| AtTAPX      |                |            |              |     |            |
| AtAPX3      |                |            |              |     |            |
| AtAPX5      |                |            |              |     |            |
| AtAPX2      |                |            |              |     |            |
| AtAPX1      |                |            |              |     |            |
| AtAPX6      |                |            |              |     |            |
| OsAPX7      | LOC_04g344900  |            |              |     |            |
| OsAPX6      | LOC_05g484900  |            |              |     |            |
| OsAPX8      | LOC_06g429000  |            |              |     |            |
| OsAPX3      | LOC_04g349000  |            |              |     |            |
| OsAPX4      | LOC_04g349000  |            |              |     |            |
| OsAPX1      | LOC_04g349000  |            |              |     |            |
| OsAPX2      | LOC_04g349000  |            |              |     |            |
| PgAPX1      | LOC_04g349000  |            |              |     |            |
| PgAPX2      | LOC_04g349000  |            |              |     |            |
| PgAPX3      | LOC_04g349000  |            |              |     |            |
| PgAPX4      | LOC_04g349000  |            |              |     |            |
| PgAPX5      | LOC_04g349000  |            |              |     |            |
| PpAPX1      | LOC_04g349000  |            |              |     |            |
| PpAPX2      | LOC_04g349000  |            |              |     |            |
| PpAPX3      | LOC_04g349000  |            |              |     |            |
| PpAPX4      | LOC_04g349000  |            |              |     |            |
| CreAPX1     | LOC_04g349000  |            |              |     |            |
| CrAPX2      | LOC_04g349000  |            |              |     |            |
| CrAPX3      | LOC_04g349000  |            |              |     |            |

**Figure 6**

https://plantphysiol.org
Arabidopsis cell suspensions

| Gene       | GFP | AOX-RFP | merged | GFP | SSU-RFP | merged |
|------------|-----|---------|--------|-----|---------|--------|
| AtSAPX     |     |         |        |     |         |        |
| OsAPX7     |     |         |        |     |         |        |
| OsAPX5     |     |         |        |     |         |        |
| OsAPX8     |     |         |        |     |         |        |
| PgAPX1     |     |         |        |     |         |        |

Onion epidermal cells

| Gene       | GFP | COXIV-mCherry | merged | GFP | SSU-RFP | merged |
|------------|-----|---------------|--------|-----|---------|--------|
| AtSAPX     |     |               |        |     |         |        |
| OsAPX7     |     |               |        |     |         |        |
| OsAPX5     |     |               |        |     |         |        |
| OsAPX8     |     |               |        |     |         |        |
| PgAPX1     |     |               |        |     |         |        |

Physcomitrella protonemal tissues

| Gene       | GFP | PpAOX-mCherry | merged | GFP | SSU-RFP | merged |
|------------|-----|---------------|--------|-----|---------|--------|
| AtSAPX     |     |               |        |     |         |        |
| OsAPX7     |     |               |        |     |         |        |
| OsAPX5     |     |               |        |     |         |        |
| OsAPX8     |     |               |        |     |         |        |
| PgAPX1     |     |               |        |     |         |        |

**Figure 6.** Dual targeting of Ascorbate Peroxidase to mitochondria and plastids. A, Phylogenetic analysis of genes encoding Ascorbate Peroxidase (APX) from Arabidopsis (At), Rice (Os), Picea (Pg), Physcomitrella (Pp) and Chlamydomonas (Cr) using MEGA 5 (see methods). There was no ortholog found in Chlorella. B, GFP images of targeting ability of APX proteins tested. Dual targeting of AtSAPX was evident in Arabidopsis cell suspensions and onion epidermal cells. Four rice APX proteins were tested, OsAPX5 and OsAPX6 showed targeting to mitochondria, while OsAPX7 and OsAPX8, showed targeting to plastids. PgAPX1 showed targeting to both mitochondria and plastids in onion epidermal cells, while PpAPX1 showed targeting to plastids only in all tissues tested. C, Table of the APX proteins from Arabidopsis, rice and Physcomitrella with genomic loci numbers, targeting predictions and experimental localization based on GFP targeting. Mitochondria (M) and plastids (Pl) are indicated respectively. Scale bar indicates 20 μM. M, Mitochondria; Pl, plastids; Cyto, cytosol; ER, endoplasmic reticulum.
| Protein Name | Accession/Locus | Predicted Target | GFP Localization | References |
|--------------|-----------------|------------------|------------------|------------|
| AtMDHAR1     | At1g63940       | M                | Pl               | Pl Chew et al. 2003 |
| AtMDHAR2     | At3g52880       | ER               | S                | Pl this study |
| AtMDHAR3     | At3g27820       | ER               | M                | Pl this study |
| AtMDHAR4     | At3g09940       | S                | -                | Pl this study |
| OsMDHAR1.1   | LOC_Os08g05570 | M                | Pl               | Pl this study |
| OsMDHAR1.2   | LOC_Os08g05570 | S                | Cyto             | Pl this study |
| OsMDHAR2     | LOC_Os08g05570 | S                | Cyto             | Pl this study |
| OsMDHAR3     | LOC_Os08g05570 | S                | Cyto             | Pl this study |
| OsMDHAR4     | LOC_Os08g05570 | S                | Cyto             | Pl this study |
| OsMDHAR5     | LOC_Os08g05570 | S                | Cyto             | Pl this study |
| PgMDHAR1     | BT117693        | Pl               | Pl               | Pl this study |
| PgMDHAR2     | BT131291        | S                | Cyto             | Pl this study |
| PgMDHAR3     | BT102815        | S                | Cyto             | Pl this study |
| PgMDHAR4     | BT117693        | S                | Cyto             | Pl this study |
| CrMDHAR      | Cre17471100     | Pl               | Pl               | Pl this study |

**Prediction**

| Protein Name | Prediction |
|--------------|------------|
| AtMDHAR1     | M          |
| AtMDHAR2     | M          |
| AtMDHAR3     | M          |
| AtMDHAR4     | M          |
| OsMDHAR1.1   | M          |
| OsMDHAR1.2   | M          |
| OsMDHAR2     | M          |
| OsMDHAR3     | M          |
| OsMDHAR4     | M          |
| OsMDHAR5     | M          |
| PgMDHAR1     | M          |
| PgMDHAR2     | M          |
| PgMDHAR3     | M          |
| PgMDHAR4     | M          |
| CrMDHAR      | M          |

**Accession/Locus**

| Protein Name | Accession/Locus |
|--------------|-----------------|
| AtMDHAR1     | At1g63940       |
| AtMDHAR2     | At3g52880       |
| AtMDHAR3     | At3g27820       |
| AtMDHAR4     | At3g09940       |
| OsMDHAR1.1   | LOC_Os08g05570 |
| OsMDHAR1.2   | LOC_Os08g05570 |
| OsMDHAR2     | LOC_Os08g05570 |
| OsMDHAR3     | LOC_Os08g05570 |
| OsMDHAR4     | LOC_Os08g05570 |
| OsMDHAR5     | LOC_Os08g05570 |
| PgMDHAR1     | BT117693        |
| PgMDHAR2     | BT131291        |
| PgMDHAR3     | BT102815        |
| PgMDHAR4     | BT117693        |
| CrMDHAR      | Cre17471100     |

**GFP Localization**

- M: Mitochondrion
- ER: Endoplasmic Reticulum
- S: Chloroplast
- Pl: Plasma membrane

**References**

- Pl Chew et al. 2003
- Pl this study

**Figure 7**

https://plantphysiol.org
Figure 7. Dual targeting of Monodehydroascorbate Reductase to mitochondria and plastids. A, Phylogenetic analysis of genes encoding Monodehydroascorbate Reductase (MDHAR) from Arabidopsis (At), Rice (Os), Picea (Pg), Physcomitrella (Pp) and Chlamydomonas (Cr) using MEGA 5 (see methods). There is no ortholog found in Chlorella. B, GFP images of targeting ability of the MDHAR proteins tested. Dual targeting of AtMDHAR6 was evident in Arabidopsis cell suspensions and onion epidermal cells. While OsMDHAR1.1 displayed dual targeting to mitochondria and plastids, OsMDHAR1.2 only showed mitochondrial targeting ability. Likewise for both Physcomitrella and Picea, of the two proteins tested for each species only one (PgMDHAR2 and PpMDHAR2 for Picea and Physcomitrella respectively) displayed dual targeting ability. C, Table of the MDHAR proteins from Arabidopsis, rice, Picea, Physcomitrella and Chlamydomonas with genomic loci numbers, targeting predictions and experimental localization based on GFP tagging. Mitochondria (M) and plastids (Pl) are indicated respectively. Scale bar indicates 20 μM. M, Mitochondria; Pl, plastids; Cyto, cytosol; ER, endoplasmic reticulum; EX, extracellular; S, secretory pathway.
| Accession/Locus | Predotar Prediction | TargetP Prediction | WoLF Prediction | Protein Name | GTP | Actin | Predic | References |
|----------------|--------------------|--------------------|-----------------|--------------|-----|------|--------|------------|
| At4g29130      | Pl                 | Pl                 | Pl              | S            | 0.29| M    | this study |
| At2g19860      | Pl                 | Pl                 | Pl              | S            | 0.09| M    | this study |
| At1g47840      | Pl                 | Pl                 | M              | Pl           | 0.63| M    | this study |
| At3g20040      | Pl                 | Pl                 | Pl              | M            | 0.18| M    | this study |
| At1g50460      | Pl                 | Pl                 | Pl              | S            | 0.12| M    | this study |
| At1g50460      | Pl                 | Pl                 | Pl              | M            | 0.12| M    | this study |
| LOC_Os07g26540 | Pl                 | Pl                 | Cyto            | M            | 0.18| M    | this study |
| LOC_Os05g45590 | ER                 | Pl                 | Pl              | M            | 0.20| M    | Nilsson et al., 2011 |
| LOC_Os01g71320 | M                  | Pl                 | Pl              | S            | 0.09| M    | Nilsson et al., 2011 |
| LOC_Os07g09890 | Pl                 | Pl                 | Pl              | Pl           | 0.61| M    | Nilsson et al., 2011 |
| LOC_Os05g44760 | ER                 | Pl                 | Pl              | S            | 0.12| M    | Nilsson et al., 2011 |
| LOC_Os01g53930 | Pl                 | Pl                 | Pl              | M            | 0.27| M    | Nilsson et al., 2011 |
| LOC_Os05g09500 | S                  | Cyto               | Pl              | M            | 0.06| M    | Nilsson et al., 2011 |
| LOC_Os01g09460 | -                  | Cyto               | Pl              | S            | 0.36| M    | Nilsson et al., 2011 |
| LOC_Os05g31110 | M                  | Pl                 | Pl              | S            | 0.27| M    | Nilsson et al., 2011 |
| BT108835       | S                  | Pl                 | Pl              | M            | 0.31| M    | Nilsson et al., 2011 |
| BT110179       | ER                 | S                  | Pl              | M            | 0.21| M    | Nilsson et al., 2011 |
| BT104062       | -                  | Pl                 | Pl              | M            | 0.70| M    | Nilsson et al., 2011 |
| Pp1s150_124V6  | M                  | Pl                 | Cyto            | M            | 0.25| M    | Nilsson et al., 2011 |
| Pp1s414_10V6   | M                  | S                  | Pl              | M + Pl       | 0.12| M    | Nilsson et al., 2011 |
| Pp1s401_23V6   | ER                 | S                  | Pl              | M + Pl       | 0.12| M    | Nilsson et al., 2011 |
| Pp1s45_156V6   | -                  | -                  | Cyto            | Cyto         | 0.29| Cyto | Nilsson et al., 2011 |
| Pp1s84_294V6   | M                  | Pl                 | Pl              | M + Pl       | 0.58| M    | Nilsson et al., 2011 |
| Pp1s58_165V6   | Pl                 | Pl                 | Pl              | M            | 0.43| M    | Nilsson et al., 2011 |
| Pp1s12_19V6    | M                  | M                  | M              | M + Pl       | 0.51| M    | Nilsson et al., 2011 |
| Pp1s3_571V6    | S                  | Pl                 | Pl              | M + Pl       | 0.18| M    | Nilsson et al., 2011 |
| Pp1s122_136V6  | S                  | Pl                 | Pl              | M + Pl       | 0.27| M    | Nilsson et al., 2011 |
| Pp1s210_75V6   | S                  | Cyto               | Pl              | Cyto         | 0.23| Cyto | Nilsson et al., 2011 |
| Pp1s258_84V6   | S                  | Pl                 | Pl              | EX           | 0.00| Pl   | Nilsson et al., 2011 |
| Cre02.g117500  | ER                 | M                  | Pl              | Pl           | 0.15| Pl   | Nilsson et al., 2011 |

Figure 8

https://plantphysiol.org
Downloaded on December 30, 2020. - Published by
Copyright (c) 2020 American Society of Plant Biologists. All rights reserved.
### Figure 8

**Arabidopsis cell suspensions**

| Protein | GFP | AOX-RFP | merged | GFP | SSU-RFP | merged |
|---------|-----|---------|--------|-----|---------|--------|
| AtHXK1  | ![Image](image1) | ![Image](image2) | ![Image](image3) | ![Image](image4) | ![Image](image5) | ![Image](image6) |
| AtHXK2  | ![Image](image7) | ![Image](image8) | ![Image](image9) | ![Image](image10) | ![Image](image11) | ![Image](image12) |
| AtHXK3  | ![Image](image13) | ![Image](image14) | ![Image](image15) | ![Image](image16) | ![Image](image17) | ![Image](image18) |
| AtHXK4  | ![Image](image19) | ![Image](image20) | ![Image](image21) | ![Image](image22) | ![Image](image23) | ![Image](image24) |

**Onion epidermal cells**

| Protein | GFP | COXIV-mCherry | merged | GFP | SSU-RFP | merged |
|---------|-----|---------------|--------|-----|---------|--------|
| AtHXK1  | ![Image](image25) | ![Image](image26) | ![Image](image27) | ![Image](image28) | ![Image](image29) | ![Image](image30) |
| AtHXK2  | ![Image](image31) | ![Image](image32) | ![Image](image33) | ![Image](image34) | ![Image](image35) | ![Image](image36) |
| AtHXK3  | ![Image](image37) | ![Image](image38) | ![Image](image39) | ![Image](image40) | ![Image](image41) | ![Image](image42) |
| AtHXK4  | ![Image](image43) | ![Image](image44) | ![Image](image45) | ![Image](image46) | ![Image](image47) | ![Image](image48) |

*Plants were transformed with CDS 1-496aa, CDS 1-502aa, CDS 1-498aa, CDS 1-493aa, or CDS 1-502aa of AtHXK1, AtHXK2, AtHXK3, AtHXK4, and AtHKL1 respectively.*
Figure 8. Dual targeting of Hexokinase to mitochondria and plastids. A, Phylogenetic analysis of genes encoding Hexokinase (HXK) from Arabidopsis (At), Rice (Os), Picea (Pg), Physcomitrella (Pp), Chlamydomonas (Cr), and Chlorella (Cv) using MEGA 5 (see methods). B, Table of the HXK proteins from Arabidopsis, rice, Picea, Physcomitrella, Chlamydomonas (Cr), and Chlorella (Cv) using MEGA 5 (see methods). C, GFP tagging of several HXK proteins from Arabidopsis showed no dual targeting was evident for any HXK present in Arabidopsis. D and E, GFP tagging of several HXK proteins from Physcomitrella revealed that PpHXK 2, 3, 5, 7, 9, and 11 displayed dual targeting ability. Mitochondria (M) and plastids (Pl) are indicated respectively. Scale bars indicate 20 μM. M, Mitochondria; Pl, plastids; Cyto, cytosol; ER, endoplasmic reticulum; EX, extracellular; S, secretory pathway.
Figure 9. Targeting ability of Mia40 in Physcomitrella. A, Phylogenetic analysis of genes encoding Mia40 in Arabidopsis (At), Rice (Os), Picea (Pg), Physcomitrella (Pp), Chlamydomonas (Cr) and Chlorella (Cv) using MEGA 5 (see methods). B, Table of the Mia40 proteins from Arabidopsis, rice, Physcomitrella, Picea, Chlamydomonas and Chlorella with genomic loci numbers, targeting prediction and experimental localization based on GFP tagging. C, GFP tagging of Physcomitrella Mia40, displaying cytosolic targeting in all tissues tested. Scale bar indicates 20 μM. M, Mitochondria; Px, peroxisomes; Pl, plastids; Cyto, cytosol; EX, extracellular; N, nuclear; Cyto, cytosol.
Figure 10. The evolution of targeting of Mia40 and putative substrate proteins. In plants such as *Arabidopsis thaliana*, *Glycine max* and *Poplar trichocarpa* Mia40 is dual targeted to mitochondria and peroxisomes. This dual targeting is also accompanied by the targeting of Mia40 substrates to the same organelles. However in plants such as *Chlamydomonas reinhardtii*, *Volvox carteri*, *Chlorella variabilis*, *Physcomitrella patens* and *Picea glauca* Mia40 is not predicted to be targeted to any organelle, and this was confirmed for Physcomitrella in this study (Figure 9). This is accompanied by either the absence of substrate proteins or the absence of targeting of substrate proteins to peroxisomes. Monocot species such as *Oryza sativa*, *Brachypodium distachyon* and *Zea mays* appear to be an intermediate as they contain a dual-targeted Mia40 but lack substrate proteins in the peroxisome. This suggests how dual targeting of one protein may facilitate the acquisition of whole metabolic pathways between different organelles. Ccs, Copper chaperone for superoxide dismutase; CSD, copper zinc superoxide dismutase. See Supplemental Figure S3 for sequence alignment of Mia40 from various plants and the predicted targeting signals highlighted.