Quantitative Analysis of the Mitochondrial and Plastid Proteomes of the Moss *Physcomitrella patens* Reveals Protein Macrocompartmentation and Microcompartmentation\(^1\)\(^{[W]}\)\(^{[OPEN]}\)

Stefanie J. Mueller, Daniel Lang, Sebastian N.W. Hoernstein, Erika G.E. Lang, Christian Schuessele, Anton Schmidt, Melanie Fluck, Desirée Leisibach, Christina Niegl, Andreas D. Zimmer, Andreas Schlosser, and Ralf Reski^*^

Plant Biotechnology, Faculty of Biology (S.J.M., D.L., S.N.W.H., E.G.E.L., C.S., A.Schm., M.F., D.L., C.N., A.D.Z., R.R.), Trinational Institute for Plant Research (R.R.), and Spemann Graduate School of Biology and Medicine (S.J.M., R.R.), University of Freiburg, 79104 Freiburg, Germany; Rudolf-Virchow-Zentrum für Experimentelle Biomedizin, University of Wuerzburg, 97080 Wuerzburg, Germany (A.Schl.); BLOSS Centre for Biological Signalling Studies, 79104 Freiburg, Germany (R.R.); and Freiburg Institute for Advanced Studies, 79104 Freiburg, Germany (R.R.)

ORCID ID: 0000-0002-5496-6711 (R.R.).

Extant eukaryotes are highly compartmentalized and have integrated endosymbionts as organelles, namely mitochondrion and plastids in plants. During evolution, organellar proteomes are modified by gene gain and loss, gene subfunctionalization and neofunctionalization, and by changes in protein targeting. To date, proteomics data for plastids and mitochondria are available for only a few plant model species, and evolutionary analyses of high-throughput data are scarce. We combined quantitative proteomics, cross-species comparative analysis of metabolic pathways, and localizations by fluorescent proteins in the model plant *Physcomitrella patens* in order to assess evolutionary changes in mitochondrial and plastid proteomes. This study implements data-mining methodology to classify and reliably reconstruct subcellular proteomes, to map metabolic pathways, and to study the effects of postendosymbiotic evolution on organelляр pathway partitioning. Our results indicate that, although plant morphologies changed substantially during plant evolution, metabolic integration of organelles is largely conserved, with exceptions in amino acid and carbon metabolism. Retargeting or regulatory subfunctionalization are common in the studied nucleus-encoded gene families of organelle-targeted proteins. Moreover, complementing the proteomic analysis, fluorescent protein fusions revealed novel proteins at organelle interfaces such as plastid stromules (stroma-filled tubules) and highlight microcompartments as well as intercellular and intracellular heterogeneity of mitochondria and plastids. Thus, we establish a comprehensive data set for mitochondrial and plastid proteomes in moss, present a novel multilevel approach to organelle biology in plants, and place our findings into an evolutionary context.

Endosymbiosis has enabled and shaped eukaryotic evolution. The engulfment of an ancestral α-proteobacterium by a presumably archaeabacterial host cell stands at the origin of mitochondrial and eukaryotic evolution over 1.5 billion years ago (Dyall et al., 2004). In plants, the subsequent uptake of a photosynthetic bacterium between 1.5 and 1.2 billion years ago led to the formation of chloroplasts (Dyall et al., 2004). Plants thereby evolved by the integration of three distinct genetic compartments. After the establishment of endosymbiosis, genes were transferred to a great extent, mainly from mitochondria and plastids to the nucleus (Bock and Timmis, 2008), necessitating an orchestrated flux of information in the form of proteins and metabolites between the compartments of eukaryotic cells to ensure homeostasis, growth, and development. This communication between organelles is facilitated by physical interactions (Kornmann et al., 2009), control of protein import (Ling et al., 2012), and retrograde signaling (Nargund et al., 2012). During radiation and diversification, especially of land plants, nuclear genomes substantially changed due to endosymbiotic and horizontal (Yue et al., 2012) gene transfer, genome duplication, and gene gain and loss (Duarte et al., 2006; Lang et al., 2010; Martin, 2010), obtruding the question of whether these phenomena are linked to alterations in metabolic pathway partitioning between organelles. Retained paralogs can either introduce a new function (neofunctionalization) or reconstitute existing functions (subfunctionalization; Duarte et al., 2006), for example

\(^{1}\) This work was supported by the Excellence Initiative of the German Federal and State Governments (grant nos. EXC 294 BIOSS and GSC–4, Spemann Graduate School of Biology and Medicine) and cofunded by the European Union (European Regional Development Fund) in the framework of the program INTERREG IV Upper Rhine (project A17 “ITP-TIP”).

\(^{*}\) Address correspondence to ralf.reski@biologie.uni-freiburg.de.

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 (www.plantphysiol.org) is: Ralf Reski (ralf.reski@biologie.uni-freiburg.de).

\(^{[W]}\) The online version of this article contains Web-only data.

\(^{[OPEN]}\) Articles can be viewed online without a subscription. www.plantphysiol.org/cgi/doi/10.1104/pp.114.235754
by distinct spatiotemporal expression profiles or distinct subcellular localizations, resulting in the modulation or introduction of metabolic functions in the respective cellular compartments. Moreover, proteins can localize to several subcellular compartments, a phenomenon called dual or multiple targeting (Yoge and Pines, 2011; Xu et al., 2013). Consequently, many eukaryotic metabolic pathways, as well as the plastid and mitochondrial proteomes, are constituted of a mosaic of proteins of diverse evolutionary origins (Szklarczyk and Huynen, 2010), and evolution has shaped variable organellar functionalities across taxa. To date, the evolution and variability of postendosymbiotic metabolic partitioning is largely not characterized on a high-throughput level. So far, large-scale mitochondrial proteome data sets are only available for the green alga *Chlamydomonas reinhardtii* (Atteia et al., 2009), rice (*Oryza sativa*; Huang et al., 2009), and the model flowering plant *Arabidopsis thaliana*; Millar et al., 2001; Heazlewood et al., 2004), whereas plastid proteomics in plants is on an advanced level and covers more species (Polyakov et al., 2010; van Wijk and Baginsky, 2011).

While higher plants diversified relatively recently but massively, simple moss plants can be traced back 330 million years (Hubers and Kerp, 2012), identifying them as prime candidates for an evolutionary view of organellar proteomes and organelle biology at a genome-wide scale. In contrast to specialized flowering plants, mosses are generalists with few tissues, high metabolic variability, and ancestral features such as high abiotic stress tolerance (Frank et al., 2007) and few plastid types (Cove, 2005).

By integrating quantitative proteomics, multivariate analysis, metabolic pathway maps, phylogenomics, and localization with fluorescent proteins, we reliably characterize subcellular proteomes and gene family diversification. Key characteristics of postendosymbiotic organellar proteome evolution are identified by cross-species comparative analysis. In support of our high-throughput analyses, we conduct single-protein analyses and identify proteins that mark microcompartments within organelles and localize to dynamic contact sites between organelles. These proteins may facilitate the exchange of proteins and metabolites, while others influence the dynamics of individual chloroplasts and mitochondria. This study characterizes the mitochondrial and plastid proteomes of moss and reveals the heterogeneity of organelles within a single cell.

**RESULTS**

**The Moss Plastid and Mitochondrial Proteomes**

Here, we applied shotgun proteomics to simultaneously isolated organelles (Lang et al., 2011) of the model moss *Physcomitrella patens* (Lang et al., 2008; Rensing et al., 2008) and identified 628 protein sets from chloroplasts and 572 from mitochondria, totaling 1,710 protein isoforms encoded by 1,246 genes (Supplemental Fig. S1; Supplemental Table S1). Of these, proteins from 177 genes were found in both organelles in the initial tandem mass spectrometry (MS/MS) identifications (Fig. 1A; Supplemental Fig. S1).

Quantitative proteomics like isotope labeling techniques can be employed to determine contaminants in organellar fractions. Previous studies used, for example, the relative protein abundance across a density gradient in comparison with known organellar marker proteins (Lilley and Dunkley, 2008; Breckels et al., 2013). In contrast, we established and employed full metabolic labeling with 15N to determine the relative protein abundance between purified subcellular fractions (i.e., labeled chloroplasts and unlabeled mitochondria). Thereby, we separate contaminants, identify dually targeted proteins, and allocate metabolic pathways to organelles (Fig. 1B). While putative chloroplast proteins have a low light/heavy (L/H) ratio, putative mitochondrial proteins have a high L/H ratio (Fig. 1C). In three biological replicates of the metabolic labeling experiment, 819 protein isoforms from 579 genes were quantified with high reproducibility (Supplemental Figs. S1 and S2; Supplemental Table S1). L/H ratios are indicative of relative protein abundance in chloroplasts and mitochondria. Thus, abundant mutual contaminants, previously identified in both organellar proteomes, can be separated (Fig. 1C). However, 17 protein-encoding loci were attributed an intermediate L/H ratio (between 0.5 and 2), thus indicating a similar abundance in plastid and mitochondrial samples (Supplemental Table S1).

**Intermediate L/H Ratios Can Indicate Localization at Organellar Interfaces**

In theory, an intermediate L/H ratio should indicate contaminations from other compartments, like the endoplasmic reticulum (ER; Supplemental Fig. S3), or proteins that are dually targeted to chloroplasts and mitochondria. To validate subcellular localization, we analyzed the in vivo localization of some of the proteins with an L/H ratio between 0.5 and 2 via tagging with fluorescent proteins and unexpectedly found several of them at contact sites between organelles (Table I).

The protein PpLs12_3792V6 has a LysM domain (PFAM PF01476) and a LEA-related domain (PANTHER PTHR23241). While LEA proteins are stress regulators, the LysM domain has the ability to bind oligosaccharides (Zhang et al., 2009). The moss protein shows a colocalization with MitoTracker (Fig. 2A) and a second localization in the ER, as additional punctate and tubular structures and the nuclear envelope were GFP labeled (Fig. 2A, right). Moreover, the transient overexpression of the LEA-like protein caused the aggregation of mitochondria, pointing to a defect in mitochondrial dynamics (Scott and Logan, 2011) reminiscent of defects in ER-mitochondria interactions in yeast (Kornmann et al., 2009). Because of the dual localization in mitochondria and the ER, we named the protein mitochondria-ER-localized LEA-related LysM domain protein1 (MELL1).

Furthermore, five proteins with intermediate L/H ratios localized to plastid stromules (stroma-filled tubules; Table I; Fig. 2, B–F). These dynamic tubules emanate from plastids and may facilitate intracellular communication by increased surface area to boost...
transport processes for proteins and metabolites and/or enable protein degradation by subjecting vesicles of plastid material to autophagy (Ishida et al., 2008; Hanson and Sattarzadeh, 2011). We identified stromule proteins involved in fatty acid biosynthesis, redox homeostasis, and metabolite transport, respectively, as follows. (1) KAR (for 3-ketoacyl-ACP reductase) is part of the plastid fatty acid synthase complex. The moss KAR is homogenously localized in chloroplasts with additional foci in stromules, especially at emanating protrusions (Fig. 2B). This pattern indicates a demand of fatty acid synthesis, putatively linked to ER-stromule membrane contact sites (Schattat et al., 2011), where lipids and fatty acids are exchanged between organelles. (2) Peroxiredoxins (Prx) are involved in redox homeostasis and act as redox sensors (Dietz et al., 2006). The plant-specific subfamily PrxQ is localized in plastids with a function in photosynthesis (Petersson et al., 2006). Of the three moss PrxQ proteins (Pitsch et al., 2010), PrxQA has an intermediate L/H ratio and was evenly distributed in plastids in 50% of the cells, while in the other 50% it accumulated in stromules, often in foci (Fig. 2C). As proteins can migrate through stromules, sometimes in batches (Köhler et al., 2000), this pattern suggests either an unexpected function of PrxQA in stromules or a transport of damaged PrxQA through these tubules to autophagic vesicles. (3) The PRAT proteins are preprotein and amino acid transporters across organelle membranes. They are divided into subfamilies (Supplemental Fig. S4), which localize either to the inner or the outer membrane of plastids and mitochondria. Among them, OEP16 proteins reside in the outer envelope of plastids and are putative amino acid channels (Pudelski et al., 2010). We found three OEP16 proteins with intermediate L/H ratios (OEP16-1.3, OEP16-2.1, and OEP16-2.2). A phylogenetic tree of PRAT family proteins in plant model species is shown in Supplemental Figure S4.

Table I. L/H ratios and subcellular localization of selected proteins

<table>
<thead>
<tr>
<th>Gene Model</th>
<th>Description</th>
<th>L/H Ratio 1</th>
<th>L/H Ratio 2</th>
<th>L/H Ratio 3</th>
<th>Median L/H Ratio</th>
<th>Subcellular Localization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pp1s317_70V6.1</td>
<td>PRAT3.1</td>
<td>12.13</td>
<td>11.94</td>
<td>na</td>
<td>12.04</td>
<td>M + P</td>
</tr>
<tr>
<td>Pp1s97_59V6.1</td>
<td>Germin-like protein5</td>
<td>1.99</td>
<td>2.84</td>
<td>na</td>
<td>2.42</td>
<td>Other (ER; Supplemental Fig. S3)</td>
</tr>
<tr>
<td>Pp1s157_44V6.1</td>
<td>OEP16-2.2</td>
<td>1.37</td>
<td>2.49</td>
<td>1.91</td>
<td>1.91</td>
<td>P (stromules)</td>
</tr>
<tr>
<td>Pp1s68_202V6.1</td>
<td>OEP16-1.3</td>
<td>0.60</td>
<td>1.28</td>
<td>1.24</td>
<td>1.24</td>
<td>P (stromules)</td>
</tr>
<tr>
<td>Pp1s233_104V6.1</td>
<td>PrxQA</td>
<td>1.05</td>
<td>1.10</td>
<td>na</td>
<td>1.08</td>
<td>P (stromules and plastid main body)</td>
</tr>
<tr>
<td>Pp1s12_379V6.1</td>
<td>MELL1</td>
<td>0.60</td>
<td>0.68</td>
<td>na</td>
<td>0.64</td>
<td>M + ER</td>
</tr>
<tr>
<td>Pp1s181_107V6.1</td>
<td>OEP16-2.1</td>
<td>0.55</td>
<td>0.74</td>
<td>0.63</td>
<td>0.63</td>
<td>P (stromules)</td>
</tr>
<tr>
<td>Pp1s33_206V6.1</td>
<td>KAR</td>
<td>0.53</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>P (stromules and plastid main body)</td>
</tr>
<tr>
<td>Pp1s72_282V6.1</td>
<td>PreP</td>
<td>0.90</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>M + P</td>
</tr>
</tbody>
</table>
PpOEP16-1.3, PpOEP16-2.1, and PpOEP16-2.2 GFP fusions localized to partly large tubular stromules and smaller protrusions on chloroplasts, whereas only low fluorescence was visible in the plastid bodies (Fig. 2, D–F). No colocalization but often close association with mitochondria was visible (Fig. 2, E and F, right). For OEP16-2.1: GFP-containing structures, the localization in the membrane of a stromule was evident (Fig. 2E) in a cross section, whereas for OEP16-1.3 and OEP16-2.2, the tubular structures were expanded and deformed, suggesting an effect of the overexpression on stromule morphology. By changing the protein-to-lipid ratio, overexpression of outer membrane proteins can cause membrane protrusions in plastids (Machet之上 et al., 2011). In addition, stromule morphology is dependent on interactions with the ER (Schattat et al., 2011) and putatively on the lipid reservoir in membrane protrusions, also termed the “mobile jacket” of plastids (Hanson and Sattarzadeh, 2008). Here, we found that OEP16 proteins are novel abundant constituents of plastid envelope protrusions and stromules, pointing to their importance in interorganellar transport processes. Thus, we infer that stromules are microcompartments of plastids that accumulate specific proteins to serve specialized functions.

**Dually Targeted to a Different Extent: PreP and PRAT3.1**

We also confirmed two proteins as being dually targeted to mitochondria and plastids. (1) Presequence peptidase (PreP), a metalloprotease of the pitrilysin family, degrades the targeting peptides that are cleaved off from preproteins subsequent to their import into mitochondria and plastids. Both Arabidopsis PreP proteins are dually targeted (Bhushan et al., 2005), as is the single moss PreP (Fig. 2G). The L/H ratio of 0.9 confirms the similar relative abundance of this protein in both endosymbiont-derived organelles. (2) The transporter PRAT3.1 (Supplemental Fig. S4) was found in the membranes of both organelles, albeit to higher amounts in mitochondrial membranes, in accordance with its higher L/H ratio (12.04). Thus, dual targeting can favor one organelle and does not necessarily result in an intermediate L/H ratio. Overexpression of PRAT3.1 led to clustered mitochondria with altered morphology (Fig. 2H), implying a role of this protein in the homeostasis of mitochondria. So far, plant PRAT3 proteins were exclusively localized in mitochondria (Murcha et al., 2007). As we also found PRAT3.1 in moss plastids, we conclude that evolutionary changes between moss and higher plants occurred in the subcellular targeting of these proteins, revealing changes in interorganellar transport during land plant evolution.

**Assessing Metabolic Partitioning by Multivariate Clustering and Classification of Subcellular Proteomes and Metabolic Pathways**

To scrutinize the evolution of metabolic partitioning between organelles at a genome-wide scale, we combined the protein sets identified here with our quantitative data obtained from metabolic labeling, Gene...
Ontology (GO) annotations, gene family definitions (Zimmer et al., 2013), and plant metabolic networks (MossCyc and AraCyc; Caspi et al., 2008). In order to reliably reconstruct and separate the quantified subcellular proteomes in silico, principal component analysis (PCA) and subsequent model-based clustering were employed, resulting in seven protein clusters (Fig. 3; Supp. Fig. S5). These represent organelle and abundance-level clusters, with predominantly mitochondrial proteins in clusters 1 and 2 and plastid proteins in clusters 4 to 6, while cluster 3 contains putative contaminants from peroxisomes (Supplemental Table S2, worksheet 2). Cluster 7 comprises very low-abundance plastid or mitochondrial proteins, dually targeted proteins (PreP; Fig. 2G), and contaminants from cytosol or ER (Supplemental Fig. S3). Interestingly, the five stromule-localized proteins identified here are at the edges of the plastid clusters 4 and 5, facing both mitochondrial clusters (Fig. 3, green squares). The inferred classifications are strongly supported by the results of GO-based Gene Set Enrichment Analysis of the resulting clusters (TopGO; Alexa et al., 2006; Supplemental Table S2, worksheets 4 and 5). The complete classification can be depicted as an overlay of protein localization and annotation of metabolic pathways (Supplemental Table S2, worksheet 2; www.plantcyc.org Omics viewer tool).

In order to take the analysis to the level of metabolic pathway partitioning among organelles, we employed multiple correspondence analysis (MCA) with subsequent hierarchical clustering of MossCyc pathway annotations for the quantified proteins (Supplemental Fig. S6; Supplemental Table S2, worksheets 6–9). The complete classification can be depicted as an overlay of protein localization and annotation of metabolic pathways (Supplemental Table S2, worksheet 2; www.plantcyc.org Omics viewer tool).

Figure 3. Organelle and abundance clustering and pathway analysis. PCA of three biological metabolic labeling replicates is shown. Eigenvectors are the L/H ratios from three biological replicates, the corresponding SPKM values, and the SPKM values of organellar proteomes without metabolic labeling (“mitochondrion” and “plastid”). In the cluster analysis of PCA, proteins are contained in seven clusters representing organelle and abundance clusters. Squares depict the positions of stromule-localized proteins with intermediate L/H ratios (Fig. 2). Full resolution is shown in Supplemental Figure S5.
resulting three clusters (Supplemental Fig. S6) reflect subcellular localizations of 272 pathway components, with 57 pathways associated to the mitochondrial cluster and 36 to the plastid cluster. Enzymes of photospiration are in a third, separate cluster. Eleven pathways containing enzymes with isoforms for both organelles (e.g. ascorbate-glutathione cycle), enzymes involved in several different pathways (e.g. enzymes of the glyoxylate cycle), or requiring the metabolite exchange between plastids and mitochondria (Gln biosynthesis) show an ambiguous clustering reflecting the relation to both organelles (Supplemental Fig. S6; Supplemental Table S2). The biosynthetic pathways of other amino acids clustered to a single compartment and mirror the situation in higher plants (Rolland et al., 2012), with the exception of Met and Cys (Supplemental Text S1). In contrast to higher plants, Cys biosynthesis does not occur in moss mitochondria (Birke et al., 2012), revealing changes in the partitioning of metabolic pathways and thus metabolites during land plant evolution.

Functional and Regulatory Diversification of Organelle-Targeted Proteins

To analyze the functional diversification of gene families, we utilized our subcellular and abundance protein classification, inferred phylogenetic relationships for candidate gene families, and predicted that, from a total of the 324 families analyzed here, only 13 (4%) gene families encode proteins that are targeted to different organelles (Supplemental Table S2, worksheet 10) and only 45 (14%) gene families encode proteins with distinctly different abundance levels (Supplemental Table S2, worksheet 11; phylogenetic trees, Supplemental Figs. S7–S9). These values reveal very low functional diversification rates of paralogous gene functions in the moss compared with higher plants, where more than 55% of the families show patterns of regulatory or functional diversification (Wang et al., 2011). This result is consistent with the reported high metabolic redundancy in moss (Lang et al., 2005, 2008; Rensing et al., 2007) and the differential evolutionary rates observed for bryophytes and flowering plants (Stenøien, 2008). Functional diversification of the identified gene families was analyzed in more detail, utilizing the inferred phylogenetic trees, our subcellular classification, the Arabidopsis SUBA database (Tanz et al., 2013), and the publication record. HSP70 proteins (Supplemental Fig. S7) are involved in intracellular communication by facilitating protein import into organelles. In moss, as in higher plants, the mitochondrial and plastid isoforms of HSP70 cluster into two clades. In contrast to their two plastid-localized homologs in Arabidopsis, we found a mitochondrial and a plastid isoform of prolyl-cis-trans-isomerases (Supplemental Fig. S8) in moss. β-Carboxyl anhydrases (Supplemental Fig. S9), which are putatively involved in carbon-concentrating mechanisms in different compartments (Fabre et al., 2007), exhibit a high variability of subcellular localization. Thus, we infer that altered subcellular targeting of organelar proteins contributed to functional diversification during evolution but occurred at different, family-dependent frequencies.

In addition, the relative protein abundances inferred by metabolic labeling can be used to deduce regulatory or expression diversification of paralogous genes. For example, the gene family of Rubisco small subunits (rbcs; Supplemental Fig. S10) is substantially expanded in P. patens and shows evidence of regulatory subfunctionalization, as we identified members of one clade to be less abundant than members of a second clade. Thus, the repeated duplication of rbcs genes resulted in dissimilar protein abundances of the paralogs in moss, indicating either subfunctionalization resulting in a differential spatiotemporal expression pattern or to balance gene dosage effects, as reported for rbcs homologs in cotton (Gossypium hirsutum; Gong et al., 2012). Similarly, our data suggest regulatory diversification of members of the mitochondrial isocitrate dehydrogenase and plastid ATPase β-chain families (Supplemental Table S2, worksheet 11).

Postendosymbiotic Evolution of Mitochondrial Pathways

As P. patens is the first nonvascular land plant with high-throughput mitochondrial proteome data, we screened for proteins with conserved mitochondrial localization but unknown function as well as for proteins previously not identified in plant mitochondria (Supplemental Tables S3 and S4; Supplemental Text S1). The results of this analysis include the presence of (1) a moss acetate kinase, revealing an anaerobic metabolism known for algae (Atteia et al., 2009) that is lost in higher plants, (2) several aldolactone oxidoreductases, implying alternative routes of ascorbate biosynthesis in moss (Supplemental Fig. S11), and (3) a formate dehydrogenase known from higher plant and fungal mitochondria (Olson et al., 2000).

Furthermore, the in vivo subcellular localization of six proteins representative for pathways with putative evolutionary changes of compartmentation (Supplemental Table S5; Supplemental Text S1) was verified using internal tagging with a fluorescent reporter at the endogenous locus via homologous recombination (Supplemental Fig. S12). Neither N- nor C-terminal targeting signals were masked by this approach, and the fusion protein was expressed from the native genomic context, which is so far unavailable for other plants, where most such studies rely on ectopic expression.

Hydroxypyruvate reductase (HPR) and the “multifunctional protein” are in moss peroxisomes (Fig. 4, A and B), suggesting that these steps of photospiration and β-oxidation were already transferred from mitochondria to peroxisomes in the ancestor of land plants. SAMS is in mitochondria of algae but in the cytosol of moss (Fig. 4C), fungi, animals, and higher plants. The moss homolog of “HIV-tat-interactive protein,” a cell death-related protein so far only known from fungi and animals, is dually targeted to cytosol and mitochondria in P. patens (Fig. 4D). mitochondria and peroxisomes are...
often very closely associated (Fig. 4E), putatively causing the copurification of abundant peroxisomal proteins with mitochondrial fractions. Our comparative analysis of postendosymbiotic organelle evolution is summarized in Figure 4F.

One isoform of fumarase (Fum) accumulated to low levels in mitochondria of protonema cells but was abundant in the cytosol of meristematic bud cells (Fig. 5, A and B). It is known from fungi, animals, and flowering plants that Fum is dually targeted to the...
cytosol and to mitochondria, although by different mechanisms (Pracharoenwattana et al., 2010; Yogev et al., 2011; Fig. 4F). Mitochondrial Fum participates in the tricarboxylic acid cycle, while cytosolic Fum is involved in DNA damage repair in yeast (Yogev et al., 2011) and influences nitrogen assimilation in Arabidopsis (Pracharoenwattana et al., 2010). ESTs from the P. patens gene (www.cosmoss.org) indicate that tissue-specific targeting in moss is regulated by alternative splicing. The evolutionary conservation of dual targeting, although by profoundly different mechanisms, suggests an important conserved moonlighting function of cytosolic Fum in eukaryotes. In addition to this heterogeneity between mitochondria of different tissues, we observed heterogeneity of mitochondria within the same cell by tracing a caseinolytic protease proteolytic subunit (ClpP). These ATP-dependent endopeptidases hydrolyze proteins in organelles. The ClpP analyzed here is the ortholog of Arabidopsis mitochondrial ClpP (Heazlewood et al., 2004) and is present to different levels in the matrix of all mitochondria but additionally accumulates in foci (Fig. 5C). The formation of these foci may be related to the fusion of a fluorescent protein to ClpP, as demonstrated for bacterial ClpP (Landgraf et al., 2012). However, the level of homogenously distributed ClpP differed between mitochondria of the same cell and thus indicates regulation in a heterogenous population of mitochondria inside a single cell.

**DISCUSSION**

**Large-Scale Multilevel Analysis of Organellar Proteomes**

During evolution, most of the DNA originally contained in the endosymbionts was transferred to the nucleus, and in return, most organellar proteins are synthesized on cytoplasmic ribosomes and imported into mitochondria and plastids. Once in the nuclear genome, many protein-coding genes acquired targeting signals that would enable the proteins to be retargeted to their original subcellular compartment or to explore other subcellular localizations, modulating or introducing metabolic functions in a different cellular compartment (Martin, 2010). Combined with the ubiquity of gene and genome duplications in the evolution of land plants (Duarte et al., 2006; Lang et al., 2010; Martin, 2010), these processes have the potential to generate part of the necessary variability and adaptations to explain the morphological, physiological, and ecological diversity observed in extant plant lineages.

Currently, we are a long way from the experimental validation of functions and localizations of every protein encoded by the growing lists of sequenced genomes. In consequence, for most proteins, we have to rely on computational predictions and high-throughput data. Although the quality of ab initio predictions of subcellular localization by machine learning or other data-mining approaches is improving continuously, the sensitivity and accuracy of such bioinformatics predictions vary greatly and are influenced by the choice of algorithms and model species (Heazlewood et al., 2004; Fuss et al., 2013). This raises the need for sufficiently large species-specific training sets with a low rate of false-positive identifications, especially where evolutionary analyses and targeting peptide predictions are concerned (Zybailov et al., 2008).

Using shotgun proteomics identifications and full metabolic labeling in moss in combination with multivariate clustering and classification (PCA and MCA), we provide a comprehensive study of plastid and mitochondrial proteomes and pathway allocation between organelles in the model plant P. patens. The metabolic labeling approach yields relative protein abundances that we employ in combination with MS/MS identifications to (1) remove frequent mutual contaminants, (2) identify organellar isoforms of enzymes and reconstruct gene family diversification, (3) assess dual targeting and its extent, and (4) detect novel proteins that localize to organelar contact interfaces. In contrast, by relying on subtractive proteomics, the proteins in the overlap of two characterized proteomes would be excluded from the analysis. Moreover, the comparative analysis of two purified organellar fractions leads to an increased number of protein identifications in the organelles of interest, compared with quantitative methods using gradient profiles (Dunkley et al., 2004). Thus, our method is particularly suitable for biological questions.
focusing on two organelles as well as for organisms or tissues with problematic purification of subcellular compartments. To date, for ease of purification, studies on mitochondrial proteomes in plants often relied on nonphotosynthetic or young tissue, such as suspension cultures, tubers, or seedlings (Millar et al., 2001; Heazlewood et al., 2004; Huang et al., 2009). These approaches exclude the study of many physiologically and developmentally important conditions, which may be analyzed with the methodology presented here.

While our study focuses on plastids and mitochondria, additional data sets employing a direct comparison of proteomes between other organelles such as mitochondria and peroxisomes would be useful to exclude contaminants and to further develop subcellular proteomics data in P. patens. As the sensitivity of tandem mass spectrometers and modern proteomics pipelines is ever increasing (van Wijk and Baginsky, 2011), as is identifying low-abundance proteins in complex mixtures, false-positive identifications are and will be a frequently addressed issue for subcellular proteomes. The availability of high-quality data sets is limited and requires the generation, integration, and curation of experimental evidence from many sources, ideally including quantitative proteomics data and single protein validations by other techniques such as the use of fluorescent tags.

Our study demonstrates that P. patens is an attractive system highly amenable to high-throughput proteomics studies, as it has few tissues and enables full metabolic labeling because it can be easily grown in liquid culture (Frank et al., 2005). Due to its intermediate evolutionary position between algae and vascular plants (Lang et al., 2008), it is additionally an excellent evolutionary model system.

“Evolutionary Proteomics”

The compartmentation of metabolism introduces diverse options for a eukaryote to adapt functionally and morphologically to the terrestrial environment, with P. patens representing a basal model system for alternative metabolic organization in nonvascular plants (Rensing et al., 2007, 2008; Lang et al., 2010).

In order to characterize postendosymbiotic evolution, several levels of data analysis needed to be combined (protein localization, protein abundance, putative protein function, and phylogenetic relationships) and subsequently integrated with corresponding data from homologs of other species. This procedure was already successfully realized by ortholog assignment between model systems to create reference plastid proteome sets for Arabidopsis and maize (Zea mays; Huang et al., 2013). Recently, we inferred gene families for 28 sequenced model species of the Plantae using phylogenomics (Zimmer et al., 2013). These clusters of orthologs provide the phylogenomic framework for an in-depth and genome-wide comparative analysis of the subcellular proteomes identified in this study. Thus, moss data are now available for further orthoproteome (Huang et al., 2013) analyses.

We used our subcellular and abundance classification in combination with pathway maps (Caspi et al., 2008) to overlay high-throughput protein localization and abundance data with predicted pathway functions. Furthermore, applying this classification to phylogenomic analysis of the identified proteins, we found evidence for distinct patterns of gene family diversification between P. patens and vascular plants. Thus, our data set constitutes a comprehensive resource for the comparative study of organelle biology and protein targeting in plants. By comparison with the literature and available proteomic data sets from animals, fungi, algae, and flowering plants, we retraced the evolutionary conservation and plasticity of metabolic partitioning as well as novel or unique organellar protein constituents. The future task is to improve the coverage of subcellular proteomes of model species (e.g. by sampling of suborganellar proteomes and different developmental stages) to validate putative protein functions in single-protein studies and to add new species to the data pool. The challenge in cross-species comparisons is still the integration of information from various localization, pathway, and orthology databases (Chen et al., 2006; Heazlewood et al., 2007; Caspi et al., 2008; Proost et al., 2009; Huang et al., 2013).

In addition to high-throughput data, fluorescent protein fusions are often necessary to support and clarify subcellular localizations (Pagliarini et al., 2008). We used the amenability of P. patens to gene targeting (Reski, 1998; Prigge and Bezanilla, 2010) to tag organelle proteins internally via knockin of yellow fluorescent protein (Citrine) at the genomic locus. This approach is particularly useful to obtain reliable data for proteins where N and/or C termini may contain targeting signals. In addition, the fusion protein is expressed in the native genomic context, allowing the tissue-specific analysis of abundance level and targeting, exemplified by moss Fum. Thus, single-protein validation can additionally give important insights into the evolutionary variability of protein macrocompartmentation as well as intracellular and intercellular heterogenous organelle populations.

Interorganellar Contact Sites and Organellar Heterogeneity

The results of our cell biological verifications showed the need to link protein content to organelle heterogeneity, organelle dynamics, and cross talk. Recently, the biochemical continuity of ER and chloroplasts was demonstrated for the biosynthesis of several nonpolar compounds using an approach termed transorganellar complementation (Mehrshahi et al., 2013). The existence of putative membrane contact sites between different organelle types is of high interest, as they may provide bidirectional access to substrates in organelle-spanning pathways (Mehrshahi et al., 2013). Moss lines with fluorescence-labeled organelles (Furt et al., 2012) can
constitute powerful tools to answer cell biological questions. We show that the similar abundance of several proteins detected by quantitative proteomics in our study is connected to specific and informative subcellular localizations, as observed by fluorescence tagging.

Based on the confocal data on MELL1 localization, it is tempting to speculate that MELL1 localizes to ER-mitochondria interfaces and either positively influences mitochondrial fusion or inhibits mitochondrial division directly or indirectly. Further investigation will reveal which role mitochondria-ER contact sites play in mitochondrial dynamics in plants. So far, components of an ER-mitochondria encounter structure are only known from fungi and animals (Kornmann et al., 2009; Schwarzländer et al., 2012). The detected intermediate L/H ratio of MELL1 could be due to a copurification of closely associated ER with plastids.

The intermediate L/H ratio of stromule-localized proteins is reflected by a marginal position in the plastid cluster in our PCA (Fig. 3), facing the mitochondrial cluster. Single-protein validation via fluorescence tagging confirmed the subcellular localization as plastidic but highlighted the accumulation of these proteins in stromules. So far, plastid stromules remain enigmatic structures, as no mutants without stromules have been identified and stromule frequency is dependent on various features, such as chloroplast number, growth conditions, and stress (Hanson and Sattarzadeh, 2011; Gray et al., 2012). Stromules are discussed as interaction hubs between subcellular compartments, including the ER (Schattat et al., 2011), and as structures supporting the autophagy of plastid material (Ishida et al., 2008). The detected intermediate L/H ratio could be due to a copurification of mitochondria and disrupted stromules in the density gradient during organelle purification, caused either by a similar density or a close association of stromules and mitochondria. Physical attachment could be caused by links to actin filaments (Hanson and Sattarzadeh, 2008). Alternatively, vesicles originating from stromules and serving the disposal of plastid material via autophagy (Ishida et al., 2008) could exhibit a similar density than mitochondria and copurify in the density gradient.

We suggest that stromules have the ability to accumulate a specific protein content that is important for the transport processes of plastid-synthesized fatty acids and amino acids into other subcellular compartments. This hypothesis might agree with the disposal of plastid material by autophagic vesicles, with the common principle of stromule outgrowth to bridge diffusion distances. Coupling to other subcellular compartments like the ER, mitochondria, and vacuole may require introrganellar redistribution of metabolic functions and damaged proteins, resulting in the intr.organelle heterogeneity observed here.

Metabolic integration of an endosymbiont-derived organelle implies nuclear control of biogenesis and autophagy but also a certain degree of organelar self-organization and differences in stress level within an organelar population (Wikstrom et al., 2009; Schwarzländer et al., 2012).

These concepts result in features such as tissue-specific protein content and heterogenous populations of organelles in a single cell such as those exemplified by the intermitochondrial differences in ClpP abundance. Mitochondria containing increased levels of ClpP (Fig. 5) might exhibit distinct properties in terms of protein content, respiratory activity, or density compared with mitochondria showing only low levels of ClpP. Further experiments are needed to unravel the effects and the regulation of ClpP abundance in plant mitochondria and to map the heterogeneity on a (sub)mitochondrial level.

We show that the study of subcellular proteomes benefits from microscopic localizations via fluorescent reporters in order to move to the level of suborganellar proteomes and microcompartments. By tagging specific organelle proteins, we revealed that moss mitochondria display distinct protein contents in a single cell and highly variable shapes, whereas moss chloroplasts exhibit a large number of protrusions and stromules. ER-linked dynamics in mitochondria as well as heterogeneity on the suborganellar level are present in basal land plants and will give new insights into organelle biology. Moreover, knockin of fluorescent proteins at the endogenous locus using techniques such as internal tagging in organisms capable of high rates of homologous recombination such as moss (Reski, 1998) can open new avenues of research by enabling the tissue-specific and dynamic study of macrocompartments and microcompartments.

CONCLUSION

Eukaryotes arose by the engulfment of prokaryotes and are thus genetic mosaics with two (animals and fungi) or three (plants) DNA-containing organelles. The integration of the third genetic compartment, the plastids, has led to photoautotrophic eukaryotes that are the nutritional basis for most life on earth. Plants had to evolve alternative means of metabolic coupling and organelar interaction hubs. We present a data set promoting the moss P. patens as a model organism for organelle biology, based on its evolutionary position and amenability to proteomics as well as microscopic studies. Comparative quantitative proteomics integrating validation on the single-protein level and metabolic pathway databases provides evidence that protein compartmentation and metabolic partitioning are highly flexible but well regulated in different kingdoms of life, different lineages within a kingdom, different tissues of a given species, between individual organelles of a single cell, and even at the suborganellar level by the formation of dynamic microcompartments in plastids and mitochondria.

MATERIALS AND METHODS

Cell Culture

Protocenoma of Physcomitrella patens was cultured in modified liquid Knop medium according to Reski and Abel (1985) containing 250 mg L\(^{-1}\) KH\(_2\)PO\(_4\), 250 mg L\(^{-1}\) KCl, 250 mg L\(^{-1}\) MgSO\(_4\) \(\cdot\) 7H\(_2\)O, 1 g L\(^{-1}\) Ca(NO\(_3\)\(_2\)) \(\cdot\) 4H\(_2\)O, and...
12.5 mg L⁻¹ FeSO₄·7H₂O (pH 5.8). In order to prevent the formation of g/emophores, moss cultures were disrupted weekly with an ULTRA-TURRAX (IKA) at 18,000 rpm for 90 s. Cultures were incubated at 23°C in a 16-h-light/8-h-dark cycle at 70 μmol m⁻² s⁻¹ (Osrarn TLD 36 W/25). For bubble flask culture, protonema filaments were disrupted with an ULTRA-TURRAX (IKA) at 18,000 rpm for 90 s before inoculation. Round-bottom flasks containing 5 L of Knop medium were inoculated with 0.3 g dry weight and aerated with 0.3 volume of air per volume of medium and minute (vvm) at 25°C under long-day conditions (see above).

Organelle Isolation

Chloroplasts and mitochondria were isolated from 7-d-old bubble flask cultures according to Lang et al. (2011).

Internal Tagging with Citrine

The DNA sequence encoding Citrine with linker sequences at the N and C termini was obtained from the Jackson Laboratory (Tian et al., 2004). Citrine with linker sequences was amplified using the forward primer 5'-GGACGGTGAGAGCTTGAAGAGGACGAG-3' (P1), 5'-GGCCCCCTGCTCACAACCTTGCTCCACCTGTTAGACTGTA-3' (P2), 5'-TTTCTA-3' (P3), and the reverse primer 5'-CAAATCCCAGGTCTTCACTG-3' (P4); Pp1s75_175, 5'-ACATT-TACGCATGTTTTTCGTCAGA-3' (P5), and the forward primer 5'-GCTGTTGATGCATTTGTCTG-3' (P6). Flanking homologous sequences were amplified from genomic DNA using P1/P2 and P3/P4, with P2 and P3 containing overlapping regions to the Citrine-linker construct: Pp1s414_8, 5'-GCTCACAACCTTGCTCCACCTGTTAGACTGTA-3' (P1), 5'-GGCCCCCTGCTCACAACCTTGCTCCACCTGTTAGACTGTA-3' (P2), 5'-TTTCTA-3' (P3), and the reverse primer 5'-CAAATCCCAGGTCTTCACTG-3' (P4). Overexpression Constructs

Coding sequences of Pp1s12_379 (MELL1), Pp1s33_206 (KAR), Pp1s233_104 (ProXQA), Pp1s181_107 (PpOE1P6-2), Pp1s157_44 (PpOE1P6-2.2), and Pp1s72_282 (PreP) were PCR amplified from wild-type complementary DNA. The constitutive promoter of the ACTINS gene (Weise et al., 2006) was used to drive the expression of a C-terminal fusion to a fluorescent protein using the Xhol and Sall restriction sites in a modified pmGFP vector (Kircher et al., 1999) containing mGFP4 (Haseloff et al., 1997; MELL1, KAR, PreXQA, and PreP) or linker Citrine (Griesbeck et al., 2003; Tian et al., 2004; OPE1-6-3, OPE1-6-2, OPE1-6-2.2, and PRAT3-1). Protoplasts were transiently transfected (Strepp et al., 1998) with 20 to 50 μg of uncut plasmid and kept for 24 to 72 h in the dark prior to confocal microscopy.

Mass Spectrometry

Sample Preparation

Proteins were extracted from washed mitochondrial pellets either using urea buffer 1 (7.5 M urea, 2.5 M thiourea, 12.5% [v/v] glycerol, 62.5 mM Tris-HCl, 2.5% [w/v] n-octylglycopropylcyanoside, and 0.1% protease inhibitor) or urea membrane buffer (5 M urea, 2.5 M thiourea, 10% [v/v] glycerol, 50 mM Tris-HCl, 2% [w/v] n-octylglycopropylcyanoside, 2.5% [w/v] SDS-10, and 0.1% [v/v] protease inhibitor) and subsequently precipitated with methanol/chloroform (modified after Wessel and Flugge, 1984). The protein pellet was resuspended in urea buffer 2 (50 mM Tris and 8 M urea), and protein concentration was determined using protein quantification according to Bradford (1976). Sequential chloroplast protein extraction was performed as described previously (Lang et al., 2011). Prior to SDS-PAGE, protein samples were mixed with Laemmli buffer (Bio-Rad), resuspended with 30 mM diisothreitol and denatured (60°C, 15 min), followed by alkalization of Cys residues by adding iodoadamantane to a concentration of 100 mM and shaking in the dark for 20 min. Residual iodoadamantane was reduced by adding diisothreitol to a final concentration of 100 mM. Finally, 10 to 50 μg of reduced and alkylated proteins was separated using SDS-PAGE on ready-made SDS gels (Ready gels, 12%; Bio-Rad) and stained with Coomassie blue (Invitrogen; SimplyBlue). Gel lanes were cut in six to 23 bands, and in-gel digestion was performed with trypsin. The excised gel bands were destained with 30% (v/v) acetonitrile and 70% (v/v) 100 mM NH₄HCO₃ for 10 min, equilibrated with 100 mM NH₄HCO₃ for 10 min, shrunk with 100% acetonitrile, and dried in a vacuum concentrator (concentrator 5301; Eppendorf). Digestion with trypsin was performed overnight at 37°C in 0.05 M NH₄HCO₃ (pH 8). About 0.1 μg of protease was used for one gel band. The supplement of the digest was combined with the supernatant after one round of extraction from the residual gel with 5% (v/v) formic acid for 30 min.

Mitochondrial and Plastid Proteomes

Liquid chromatography-MS/MS analyses were performed on an ion-trap mass spectrometer (Agilent 6340; Agilent Technologies) coupled to a 1200 Agilent nanoflow system via an HPLC-Chip cubic electrospray ionization interface. Peptides were eluted with 0.1% formic acid (H₂O) and 0.1% formic acid (acetonitrile) were 95% Peptide Prophet probability, 90% Protein Prophet probability, and at least one identified peptide (protein decay false positive). After an MS survey scan, the three peptides were selected for MS/MS (collision-induced dissociation, standard enhanced mode). The automated gain control was set to 350,000. The maximum accumulation time was set to 300 ms. Mascot Distiller 2.1 was used for raw data processing and for generating peak lists, essentially with standard settings for the Agilent ion trap. Mascot Server 2.2 was used for database searching with the following parameters: peptide mass tolerance, 1.1 Da; MS/MS mass tolerance, 0.3 Da. 1C; 1 enzyme, trypsin with maximum of two missed cleavages; variable modifications, Gln→pyroGlu (N-terminal Q), oxidation (M), and carbamidomethyl (C). The Scaffold V3 software (Proteome Software) was used to generate a decoy database, by reversing all protein hits with the database including the reversed sequences from Scaffold. Protein hits were validated using the Scaffold V3 software, by reversing all protein hits with the database including the reversed sequences from Scaffold. Protein hits were validated using the Scaffold V3 software, by reversing all protein hits with the database including the reversed sequences from Scaffold. Protein hits were validated using the Scaffold V3 software, by reversing all protein hits with the database including the reversed sequences from Scaffold. Protein hits were validated using the Scaffold V3 software, by reversing all protein hits with the database including the reversed sequences from Scaffold. Protein hits were validated using the Scaffold V3 software, by reversing all protein hits with the database including the reversed sequences from Scaffold.
Metabolic Labeling and Quantitative Proteomics

Protonema of P. patens was cultured under standard conditions (see above) with the exception that the Ca(NO₃)₂ in the medium was replaced by Ca(¹⁵NO₃)₂ (Cambridge Isotope Laboratories). For each replicate of the metabolic labeling experiments on organelles, moss was freshly labeled by growing protonema for 5 weeks in Knop medium containing Ca(¹⁵NO₃)₂. In parallel, two unlabeled cultures were inoculated in normal Knop medium containing Ca(NO₃)₂. Heavy and light cultures were processed in parallel under the same conditions and derived from the same moss inoculum. After 5 weeks, each culture was transferred to round-bottom flasks containing 5 L of medium and aerated with 0.3 volumes of air per volume of medium per minute for further growth for 1 week. For each replicate, two bubble flasks containing light culture and one bubble flask containing heavy culture were harvested, and both mitochondria and chloroplasts were isolated (Lang et al., 2011). Proteins were extracted from washed organelle pellets using urea buffer 1 (7.5 M urea, 2.5 M thiourea, 12.5% [v/v] glycerol, 62.5 mM Tris-HCl, 2.5% [w/v] n-octylglycropyranosid, and 0.1% [v/v] protease inhibitor) and subsequently precipitated with methanol/ chloroform (modified after Wessel and Flügge [1984]). The protein pellet was resuspended in urea buffer 2 (50 mM Tris and 8 M urea), and protein concentration was determined using protein quantification according to Bradford (1976). A total of 35 μg of protein from ¹³C-labeled chloroplast was mixed one by one with 35 μg of protein from ¹⁵N-containing mitochondria and mixed with SDS-Sample buffer (Bio-Rad) for SDS-PAGE. For each replicate, 22 gel slices were cut. Destaining of the gel and preparation for liquid chromatography–MS/MS were performed as described for the organelar proteomes.

Digested protein samples from the metabolic labeling experiments were measured with an electrospray ionization quadrupole time-of-flight device (Agilent 6520) coupled to an Agilent 1200 nanoflow system via an HPLC-Chip cube interface. Because of the high sample complexity, a large-capacity chip (Agilent 500nm 300A 185/αLC chip II) with a 160-nL trap column packed with Zorbax 300SB C-18 (5 μm particle size) was used. Peptides were eluted with a linear acetonitrile gradient. MS/MS analysis was performed using the data-dependent mode. After an MS scan (standard enhanced mode), four precursors were selected for MS/MS (collision-induced dissociation, standard enhanced mode). Mascot Server 2.2 was used for database searching with the following parameters: peptide mass tolerance, 50 ppm; MS/MS mass tolerance, 0.05 Da; ¹³C; 1; enzyme, trypsin with maximum of two missed cleavages; variable modifications, Gln→pyroGlu (N-terminal Q), oxidation (M), and carbamidomethyl (C). Mascot search was performed with the P. patens version 1.6 database (Zimmer et al., 2013) including the decoy database containing reversed sequences generated by Scaffold (see above). Protein hits were validated using Scaffold V3 software. Settings for Scaffold analysis were as followed: standard sample, each file will be analyzed separately; condensing off, keep all unmatched spectra for further export. Settings for the protein identification were 80% Protein/Prophet probability, 96% Protein/Prophet probability, and at least two identified peptides. Mascot Distiller 2.4 was used for quantification assuming 98% ¹³C and 2% ¹⁵N [Constrain Search, no; Protein Ratio Type, median; Protein Score Type, mudpit; Report Detail, yes; Show subsets, 0.0; Require bold red, yes; Minimum Peptides, 1; Significance Threshold, 0.05; Allow mass time match, yes; Integration Source, survey; Simple Ratio, no; Allow Elution Shift, no; Elution Profile Correlation Threshold, 0.1; All Charge States, yes; All Charge States Threshold, 0.2; Matched Rho, 0.85; XIC Threshold, 0.1; XIC Max Width, 250; XIC Smoothing, 3; Minimum precursor charge, 2; Isolated Precursor, No; Minimum a(1), 0.0; Peptide Threshold Type, minimum score; Unique Peptide seq, no; Isolated Precursor Threshold, 0.5]. Ratios of peptides from distinct protein bands from one gel lane were combined using an in-house Perl script, and median L/H ratios were calculated for every protein from the L/H ratios of assigned peptides. The script also included the Protein Peptide and the Peptide Prophet values to validate the quantified proteins and peptides. Finally, proteins were taken into account that were at least quantified with two peptides and had a Protein Prophet probability greater than or equal to 99%.

Classification of Organellar Fractions and Protein Abundance Levels

PCA of the data matrix was performed using R (http://www.r-project.org/). Loadings of the resulting principal components were analyzed and used to select the first two components representing the variances separating the mitochondrial and plastid fractions. Individual PCA scores are available in worksheet 1 in Supplemental Table S2. Subsequently, the two components were used to cluster the quantified proteins using the R mclust package (Fraley and Raftery, 2003). The best model and number of clusters were selected based on the Bayesian Information Criterion (VVI model; seven clusters).

Protein abundance levels from shotgun spectra were derived by k-means clustering each of the five scaled SPKM columns (one median mitochondrial, one median plastid, and three pooled SPKM) into five clusters representing the following linear abundance categories: 1, not detectable, very low abundance; 2, very low, low abundance; 3, low, medium abundance; 4, medium, high abundance; 5, high, very high abundance.

The clustered data matrix, resulting organellar clusters, cluster uncertainties and probabilities, protein abundance levels, along with known protein names and descriptions (www.cosmoss.org) are provided in worksheet 2 of Supplemental Table S2.

GO Enrichment Analysis

Differences in terms of biological process, molecular function, and subcellular components of the protein clusters were assessed using GO enrichment analysis using the R Bioconductor TopGO package and the latest cosmoss.org GO annotation (February 4, 2013; https://www.cosmoss.org/phytosome_project/wiki/Downloads#Master_branch_-_Nightly-build). Enrichment results at the 95% confidence level for the comparison against the entire genome and within the clustered sets are available in worksheets 4 and 5 in Supplemental Table S2.

Multiple Correspondence Analysis and Partitioning of Metabolic Pathways

Clustered proteins were mapped to metabolic pathways using the MosMyc 1.0 annotations kindly provided by the Plant Metabolic Network database 7.0 (http://www.plantcyc.org; Caspi et al., 2008), resulting in 616 protein-in-pathway mappings. Classifications for the seven organellar clusters and 107 pathways were used as qualitative variables together with the SPKM, L/H, and cluster uncertainty and cluster probability values as supplementary quantitative variables for multiple correspondence analysis (method MCA) using the R FactoMineR package (Lé et al., 2008). Resulting dimensions were clustered into three pathway partitions (orange1, green2, and red) using the HCPC (for Hierarchical Clustering on Principal Components) method implemented in FactoMineR. Pathways present in both the mitochondrial and plastid cluster fractions were assigned a fourth (blue) partition (Supplemental Fig. S6). Summary statistics of MCA dimensions and pathway partitioning and full classification of proteins into pathways at 95% confidence are provided in worksheets 6 to 9 in Supplemental Table S2.

Gene Family Functional Diversification Analysis

Gene family definitions were based on our recent phylogenomics analysis based on homology clustering of protein sequences from 28 sequenced Plantae genomes (Zimmer et al., 2013). The resulting protein clusters in most cases represent gene families or subfamilies that can be traced back to a single gene

Data Preprocessing for Multivariate Analysis

The number of total spectra observed (spectral counts) for each protein isoform was normalized analogously to RNASeq count data to SPKM (spectral count per kilo amino acids per million spectra). SPKM values from 18 plastid and 10 mitochondrial shotgun proteomics samples were condensed using two median values.

Reproducibility of the biological replicates was assessed visually by plotting scatter and Lowes smooth-fitted curves of log-transformed L/H ratios and by testing for significant correlation using Kendall’s τ. Up to two missing L/H values per isoform and replicate were imputed using the R Amelia package (Honaker et al., 2011). Imputation was performed 100 times, and the resulting matrices were condensed to mean values. SPKM and L/H values were log transformed and scaled for further analysis. The resulting data matrix had eight columns (three L/H, three pooled L/H SPKM, one median mitochondrial SPKM, and one median plastid SPKM).
in the last common ancestor of Plantae, Viridiplantae, and Embryophyta. Proteins in organellar clusters were annotated according to gene families to define the following subsets: gene families with proteins in at least two subcellular compartments (clusters 1–3 versus clusters 4–6 versus cluster 7) as possible evidence for subcellular diversification and gene families with evidence for regulatory diversification based on the distribution of members among different intrageneric clusters representing divergent protein abundance levels (e.g. in plastids split across clusters 4–6). The resulting lists of proteins are provided in tables 10 and 11 of Supplemental Table S2. Phylogenetic relationships of some of the gene families were reconstructed as described below, in order to distinguish between ancestral or more recent species-specific duplications and subsequent subfunctionalizations or neofunctionalizations.

PhylogeneticAnalyses

Homologs of putative aldonolactone oxidoreductases have been retrieved from PLAZA (Proost et al., 2009) and our phylogenomic clustering (Zimmer et al., 2013) and aligned using Jalview (Waterhouse et al., 2009) using the MAFFT algorithm (default settings). The multiple sequence alignment was subsequently trimmed to the D-arabinono-1,4-lactone oxidase domain and FAD-binding domains, as identified by Plant (Finn et al., 2010). Correct alignment of domains was checked manually. A phylogenetic tree was constructed using MrBayes (version 3.1.2) software (Ronquist and Huelsenbeck, 2003) with the following settings: aamodel = mixed, ngen = 500 000, samplefreq = 100, burnins = 1250. The deviation of split frequencies was well below 0.01 after 500000 generations.

Phylogenetic relationships of the PRAT superfamily were reconstructed based on a manually curated MAFFT (Katoh et al., 2002) multiple sequence alignment of representative sequences from the PFAM domain PF02466 and from the OrthoMCL clusters (see above), representing the PRAT gene families from 16 representative plant species covering the entire range of Plantae. The multiple sequence alignments of the families studied for gene functionalization in more detail were aligned using the PRANK (Löytynoja and Goldman, 2010) algorithm. An appropriate evolutionary model was selected using ProtTest (Abascal et al., 2005). The chosen LG model (http://mbe.oxfordjournals.org/content/25/7/1307.full) was used to infer phylogenetic relationships with posterior probabilities using the aayes method implemented in PHYML (Guindon et al., 2010) with four rate categories, empirical amino acid frequencies, maximum likelihood estimates of transition/transversion ratio, invariable sites and g parameters, and optimization of tree topology and branch lengths.

The FigTree software version 1.2.3 (A. Rambaut; http://tree.bio.ed.ac.uk/software/gtree/) was used to generate figures.

Microscopy

Confocal images were taken on a Zeiss LSM 510 with an Axiovolt 200 inverted microscope. Water-immersion objectives were used: the LD LCI Plan-Apochromat 25×/0.8 DIC Imm Korr (UV) VIS-IR and the C-Apochromat 63×/1.2W Korr. Citrine and chlorophyll were excited with an argon laser (488 nm). Chlorophyll auto-fluorescence was detected above 650 nm, and fluorescence was detected above 650 nm, and fluorescence was detected between 560 and 615 nm. Fluorescence signals are false colored in Magenta (chlorophyll), green (Citrine), and orange (MTO). Three-dimensional reconstructions of z-stacks were performed with the Imaris software (Bitplane). Screenshots were taken from three-dimensional reconstructions using IrfanView (www.irfanview.de). Confocal planes were exported from the ZEN2010 software (Zeiss) and further processed with Gimp (http://www.gimp.org/).

SupplementalData

The following materials are available in the online version of this article.

Supplemental Figure S1. Overview of identified and quantified proteins.

Supplemental Figure S2. Metabolic labeling; variability of biological replicates.

Supplemental Figure S3. Subcellular localization of Germin-like protein 5.

Supplemental Figure S4. Phylogeny of PRAT proteins.

Supplemental Figure S5. PCA (full resolution).

Supplemental Figure S6. MCA and factor map.

Supplemental Figure S7. Phylogeny of HSP70.

Supplemental Figure S8. Phylogeny of peptideyl-prolyl cis-trans-isomerasers.

Supplemental Figure S9. Phylogeny of 3-carbonic anhydrases.

Supplemental Figure S10. Phylogeny of rcs.

Supplemental Figure S11. Phylogeny of aldonolactone oxidoreductases.

Supplemental Figure S12. Internal tagging at the endogenous locus.

Supplemental Table S1. Protein identification and quantification.

Supplemental Table S2. Analyses of metabolic labeling and MS identifications: PCA, MCA, and MossCyc pathways.

Supplemental Table S3. Evolutionarily conserved proteins with unknown function.

Supplemental Table S4. Mitochondrial proteins not conserved in all land plant lineages.

Supplemental Table S5. Proteins chosen for internal tagging at the endogenous loci with Citrine and subcellular localization in distinct lineages.

Supplemental Text S1. Further comparative analyses of organellar proteomes.

ACKNOWLEDGMENTS

We thank Robert Gawlik for help with data processing for the metabolic labeling experiments and Christine Glockner and Richard Haas for technical assistance. We thank the Jackson Laboratory for the gift of the Citrine-linker construct, the team of the life imaging center in Freiburg for support with confocal microscopy, and Dr. Peter Beyer, Chris Meisinger, Markus Schwarzländer, and Marco Vervliet-Scheebaum for helpful discussions.

Received January 14, 2014; accepted February 7, 2014; published February 10, 2014.

LITERATURE CITED


Macrocompartmentation and Microcompartmentation in Moss
Macrocompartmentation and Microcompartmentation in Moss


Supplemental Data:

Supplemental Figures:

Supplemental Figure S1. Overview of identified and quantified proteins

(A) 572 protein sets were identified in the mitochondrial fractions, whereas 628 protein sets were identified in the plastid fractions (based on V1.6 protein models, www.cosmoss.org). Due to the high redundancy of genes, many peptides match to several genomic loci or splice variants and are assembled in protein sets. (B-C) In concordance with the high metabolic redundancy and large gene families of the moss genome (Rensing et al., 2008), a high percentage of the genomic loci in the plastid and the mitochondrial datasets are identified with peptides matching several encoded isoforms and thus grouped into protein sets: 33.2% of the proteins identified in the plastid fractions are assembled in sets, whereas 66.8% distinct proteins could be identified with unique peptides. In the mitochondrial fraction 49.6% of the proteins could be identified with unique peptides. (D) 177 protein-encoding loci were identified in both mitochondrial as well as plastid fractions, of which 154 were quantified in the metabolic labeling experiment, corresponding to 87% of the overlap of the organellar proteomes (Venn diagram (Hulsen et al., 2008)).
Supplemental Figure S2. Metabolic labeling: variability of biological replicates

L/H ratios of proteins quantified in the biological replicates of the metabolic labeling experiments show high reproducibility. L/H ratios of proteins quantified in all three biological replicates are plotted against each other (logarithmic scale). The correlation for each pairwise comparison is given.

Supplemental Figure S3. Subcellular localization of Germin-like protein 5 (GLP5)

Transient overexpression of GLP5:GFP (see also Table I) in moss protoplasts. GFP fluorescence is depicted in green, MitoTracker CMTMRos in orange and chlorophyll autofluorescence in magenta. GLP5:GFP does not co-localize with MitoTracker or plastids but is localized in the ER, as shown by the network structure of the signal and the fluorescence of the nuclear envelope (arrowhead). Scale bars 5µm.
Supplemental Figure S4. Phylogeny of PRAT proteins

The PRAT superfamily comprises the mitochondrial TIM complex components TIM17, TIM23, TIM22, the mitochondrial PRAT3 and PRAT4 proteins, the chloroplast OEP16 (outer envelope protein of 16kDa), PRAT1 proteins and the putatively dually localized PRAT2 (SAM (sterile alpha motif) domain containing) proteins (Murcha et al., 2007; Pudelski et al., 2010). Subfamilies are colored according to putative subcellular localization: orange=mitochondria, green=plastids. Identifiers are from UniProt, Phytozome (Goodstein et al., 2011) and cosmoss (www.cosmoss.org). Identifiers of proteins with organellar MS/MS evidence (according to SUBA (Heazlewood et al., 2007) and this study) are colored in green for plastid, orange for mitochondria and blue for plastids and mitochondria. Empty circles correspond to GFP evidence from (Murcha et al., 2007; Pudelski et al., 2010) and filled shapes to GFP evidence from this study: green for plastid, orange for mitochondria, blue for plastids and mitochondria and green
square for stromule localization (this study). The OEP16-1 and OEP16-2 family proteins are likely amino acid transporters (Pudelski et al., 2010) accumulating in stromules. In addition to putatively dual targeted PRAT2 proteins (Murcha et al., 2007; Pudelski et al., 2010), PpPRAT3.1 is dual targeted to mitochondria and plastids.
Supplemental Figure S5: PCA (full resolution)

Principal component analysis of three metabolic labeling replicates with “heavy” chloroplasts and “light” mitochondria. SPKM (spectral counts per kilo amino acids per million spectra) from the metabolic labeling experiments as well as from the identifications in the mitochondrial and plastid proteomes (eigenvector “mitochondrion” and “plastid”) were included. Clustering divides the protein isoforms into seven clusters of which cluster 1-2 contain predominantly mitochondrial proteins, cluster 3 putative peroxisomal contaminants, cluster 4, 5 and 6 plastid proteins and cluster 7 predominantly putative contaminants in both fractions and mitochondrial and plastid proteins of low abundance.
Supplementary variables on the MCA factor map
Supplemental Figure S6. MCA and factor map: (A) Factor map: dimension 1 of multiple correspondence analysis (MCA) including pathway data from the MossCyc database (Caspi et al., 2007) separates mitochondrial pathways, plastidic pathways and pathways with higher fit uncertainty in clusters. Cluster 1 and 2 contain predominantly mitochondrial proteins, cluster 3 putative peroxisomal contaminants, and cluster 4, 5 and 6 plastidic proteins, whereas cluster 7 contains putative contaminants in both organellar fractions and low abundant proteins. (B) Clustered MCA: metabolic pathways form three distinct clusters. The resulting three clusters reflect subcellular localizations of 272 pathway components with 57 pathways associated to the mitochondrial cluster (orange) and 36 to the plastid cluster (green). Enzymes of photorespiration are in a third, separate cluster (red). Eleven pathways (blue) either containing enzymes with isoforms for both organelles (e.g. ascorbate-glutathione cycle), enzymes involved in several distinctively localized pathways (e.g. enzymes of glyoxylate cycle), or require the metabolite exchange between plastids and mitochondria (glutamine biosynthesis), show an ambiguous clustering.
HSP70 proteins are present in plastids as well as in mitochondria. The functional diversification in the superfamily is known and dates back to the algal ancestors of land plants. Protein isoforms identified and quantified in this study are colored according to their assigned PCA cluster (light blue=cluster 5, orange=cluster 2, brown=cluster 1). Posterior probabilities are depicted at internal nodes.
Supplemental Figure S8. Phylogeny of peptidyl-prolyl cis-trans-isomerases

Whereas the homologs in *A. thaliana* both localize to chloroplasts and are important for redox-signaling (Dominguez-Solís et al., 2008) and NAD(P)H dehydrogenase complex assembly (Sirpiö et al., 2009), one isoform in *P. patens* clusters with mitochondrial proteins and the other with plastid proteins. Protein isoforms identified and quantified in this study are colored...
according to their assigned PCA cluster (dark blue=cluster 4, orange=cluster 2). Posterior probabilities are depicted at internal nodes.
Supplemental Figure S9. Phylogeny of beta carbonic anhydrases

Beta carbonic anhydrases are present in plastids as well as in mitochondria. The mitochondrial isoform identified in *P. patens* in this study is in the same clade as cytosolic, plastid and plasma membrane localized isoforms of *A. thaliana*. Moss protein isoforms identified and quantified in this study are colored according to their assigned PCA cluster (dark green=cluster 6, orange=cluster 2). Proteins localized in *A. thaliana* and *C. reinhardtii* are colored and annotated according to their identified subcellular localization (Fabre et al., 2007; Moroney and Ynalvez, 2007). Posterior probabilities are depicted at internal nodes.
Figure S10
Supplemental Figure S10. Phylogeny of *rbcs*

Protein isoforms of the small subunit of ribulose-1,5-bisphosphate-carboxylase/oxygenase (*rbcs*) identified and quantified in this study are colored according to their assigned PCA cluster (dark blue=cluster 4 (less abundant), dark green=cluster 6 (high abundance)). The gene family of RubisCO-SUs is expanded in *P. patens* and subdivided in two clades whose members show distinct abundance in plastids and thus regulatory diversification and subfunctionalization. Whether the subfunctionalization is ancestral or specific to moss is not clear, as gene conversion could mask ancestral diversification. Posterior probabilities are depicted at internal nodes.
Supplemental Figure S11: Phylogeny of aldonolactone oxidoreductases

Ascorbate and its analogue dehydro-D-arabinono-1,4-lactone can be synthesized from different precursors by enzymes named aldonolactone oxidoreductases. This tree depicts a phylogenetic analysis of these enzymes showing well-separated clades of which some are not yet functionally characterized. Node labels correspond to posterior probabilities indicating the reliability of a node (0 - 1.0). Identifiers: At = A. thaliana, Pp = P. patens V1.6 (locus number alone is shown when V1.2 and V1.6 models clustered together in the analysis), Chlre = Chlamydomonas reinhardtii, Selmo = Selaginella moellendorfii, Vc = Volvox carteri, Rn = Rattus norvegicus, Mm = Mus musculus, Sc = Saccharomyces cerevisiae, Ca = Candida albicans, Ss = Streptomyces spec.. Enzymes identified in this study in moss mitochondria are boxed in orange.*MS/MS evidence mitochondrial (SUBA) (Heazlewood et al., 2007).
Supplemental Figure S12. Internal tagging at the endogenous locus

The sequence encoding Citrine and linkers is flanked by sequences homologous to the target locus in the genome. After homologous recombination, Citrine is integrated into the coding sequence of the gene and a fusion protein is produced. Neither the N- nor the C-terminus of the fusion protein is modified leaving putative targeting sequences functional. Interference between Citrine and target protein folding are minimized by the flexible linker peptides (Tian et al., 2004).
Supplemental Tables

Supplemental Table S1. Protein identification and quantification (supplemental file)

Supplemental Table S2. Analyses of metabolic labeling and MS identifications: PCA, MCA, MossCyc pathways (supplemental file)

Supplemental Table S3. Evolutionary conserved proteins with unknown function

Evidence for mitochondrial localization is from SUBA (Heazlewood et al., 2007) (only the best available mitochondrial evidence is shown in table). Ortholog cluster information is from PLAZA (Proost et al., 2009) and OrthoMCL (Chen, 2006).

<table>
<thead>
<tr>
<th>P. patens V1.2</th>
<th>P. patens V1.6 locus</th>
<th>Description</th>
<th>A. thaliana homolog</th>
<th>Method</th>
<th>Evidence for mitochondrial localization</th>
<th>TAIR description</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phypa_16 2532</td>
<td>Pp1s44_182</td>
<td>expressed protein</td>
<td>AT4G26410 AT2G45060</td>
<td>PLAZA</td>
<td>prediction (1)</td>
<td>uncharacterized conserved protein UCP022280</td>
<td>InterPro:IPR016803 Plant-specific</td>
</tr>
<tr>
<td>Phypa_14 8013</td>
<td>Pp1s270_16</td>
<td>expressed protein</td>
<td>AT2G20820</td>
<td>PLAZA</td>
<td>prediction (8)</td>
<td>unknown protein</td>
<td>Plant-specific?</td>
</tr>
<tr>
<td>Phypa_16 2786</td>
<td>Pp1s47_175</td>
<td>expressed protein</td>
<td>AT1G65270</td>
<td>PLAZA, OrthoMCL</td>
<td>prediction (1)</td>
<td>unknown protein</td>
<td></td>
</tr>
<tr>
<td>Phypa_23 3279</td>
<td>Pp1s75_233</td>
<td>expressed protein</td>
<td>None; but conserved in other flowering plants</td>
<td></td>
<td></td>
<td>InterPro:IPR021467 Protein of unknown function DUF3119</td>
<td></td>
</tr>
<tr>
<td>Phypa_12 4578</td>
<td>Pp1s50_23</td>
<td>expressed protein</td>
<td>AT3G27930</td>
<td>OrthoMCL</td>
<td>prediction (2)</td>
<td>unknown protein</td>
<td>Plant-specific?</td>
</tr>
<tr>
<td>Phypa_13 0678/ Phypa_20 6609/ Phypa_47 294</td>
<td>Pp1s84_251/Pp1s34_228/Pp1s20_93</td>
<td>expressed protein</td>
<td>AT2G16460</td>
<td>OrthoMCL</td>
<td>MS/MS (4)</td>
<td>Protein of unknown function (DUF1640)</td>
<td>InterPro:IPR012439 Protein of unknown function DUF1640 InterPro: only metazoan</td>
</tr>
<tr>
<td>Phypa_22 7472</td>
<td>Pp1s421_8</td>
<td>HVA22-like protein</td>
<td>AT4G24960 AT5G50720</td>
<td>PLAZA, PLAZA</td>
<td>prediction (2) prediction (3)</td>
<td>HVA22D HVA22E</td>
<td>InterPro: IPR004345 TB2/DP1/HVA22</td>
</tr>
<tr>
<td>Phypa_16 9186</td>
<td>Pp1s200_12</td>
<td>Lectin-like</td>
<td>AT1G19715</td>
<td>PLAza, OrthoMCL</td>
<td>none</td>
<td>Mannose-binding lectin superfamily protein</td>
<td>InterPro: IPR001229 Mannose-binding lectin</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
<td>-------------</td>
<td>-----------</td>
<td>----------------</td>
<td>------</td>
<td>-----------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Phypa_11 0085</td>
<td>Pp1s118_113</td>
<td>MAM33-like protein</td>
<td>AT3G55605</td>
<td>OrthoMCL</td>
<td>MS/MS (1)</td>
<td>Mitochondrial glycoprotein family protein</td>
<td>InterPro: IPR003428 Mitochondrial glycoprotein Eukaryote-specific</td>
</tr>
<tr>
<td>Phypa_21 5382</td>
<td>Pp1s114_140</td>
<td>midA-like protein</td>
<td>AT3G28700</td>
<td>PLAza, OrthoMCL</td>
<td>prediction (5)</td>
<td>Protein of unknown function (DUF185)</td>
<td>InterPro: IPR003788 Protein of unknown function DUF185</td>
</tr>
<tr>
<td>Phypa_14 4349/ Phypa_55 615</td>
<td>Pp1s214_37/ Pp1s15_225</td>
<td>MA3</td>
<td>AT4G24800</td>
<td>OrthoMCL</td>
<td>none</td>
<td>MA3 domain-containing protein</td>
<td>4x InterPro: IPR003891 Initiation factor eIF-4 gamma, MA3 Eukaryote-specific</td>
</tr>
<tr>
<td>Phypa_12 8768</td>
<td>Pp1s172_71</td>
<td>MA3</td>
<td>AT4G24800</td>
<td>OrthoMCL</td>
<td>none</td>
<td>MA3 domain-containing protein</td>
<td>4x InterPro: IPR003891 Initiation factor eIF-4 gamma, MA3 Eukaryote-specific</td>
</tr>
<tr>
<td>Phypa_11 4781</td>
<td>Pp1s11_325</td>
<td>CBS/PB1 domain-containing protein</td>
<td>AT5G50530</td>
<td>PLAza</td>
<td>prediction (3)</td>
<td>CBS domain-containing protein / octicosapetide/Phox/Bemp1 (PB1) domain-containing protein</td>
<td>InterPro: IPR000644 Cystathionine beta-synthase, core InterPro: IPR000270 Octicosapetide/Phox/Bem1p</td>
</tr>
<tr>
<td>Phypa_25 822</td>
<td>Pp1s12_415</td>
<td>GRAM-domain containing protein</td>
<td>AT5G13200</td>
<td>OrthoMCL</td>
<td>prediction (1)</td>
<td>GRAM domain family protein</td>
<td>InterPro: IPR004182 GRAM</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
<td>--------------------------------</td>
<td>-----------</td>
<td>-----------</td>
<td>---------------</td>
<td>-----------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Phypa_17 2747</td>
<td>Pp1s376_53</td>
<td>mitofilin domain-containing protein</td>
<td>AT4G39690</td>
<td>PLAza, OrthoMCL</td>
<td>MS/MS (2)</td>
<td>Unknown mitochondrial protein</td>
<td>InterPro: IPR019133 Mitofilin</td>
</tr>
<tr>
<td>Phypa_15 5619</td>
<td>Pp1s475_2</td>
<td>putative sucrase-related protein</td>
<td>AT4G26620</td>
<td>OrthoMCL</td>
<td>none</td>
<td>Sucrase/ferredoxin-like family protein</td>
<td>InterPro: IPR009737 Sucraseferredoxin-like IPR012336 Thioredoxin-like fold</td>
</tr>
<tr>
<td>Phypa_16 8317</td>
<td>Pp1s172_80</td>
<td>tetratricopeptide repeat (TPR)-containing protein, putative kinesin light chain</td>
<td>AT2G31240</td>
<td>PLAza</td>
<td>prediction (4)</td>
<td>Tetratricopeptide repeat (TPR)-like superfamily protein</td>
<td>InterPro: IPR011990 Tetratricopeptide-like helical IPR019734 Tetratricopeptide repeat IPR013026 Tetratricopeptide repeat-containing</td>
</tr>
<tr>
<td>Phypa_17 6822</td>
<td>Pp1s19_316</td>
<td>tetratricopeptide repeat (TPR)-containing protein, putative kinesin light chain</td>
<td>AT2G31240</td>
<td>OrthoMCL</td>
<td>prediction (4)</td>
<td>Tetratricopeptide repeat (TPR)-like superfamily protein</td>
<td>InterPro: IPR011990 Tetratricopeptide-like helical IPR019734 Tetratricopeptide repeat</td>
</tr>
</tbody>
</table>
Supplemental Table S4. Mitochondrial proteins not conserved in all land plant lineages

For several proteins identified in moss mitochondrial samples no homologs were identified in seed plants thus suggesting that a loss of these proteins occurred during land plant evolution.

Chlre: *Chlamydomonas reinhardtti*; Chlsp: *Chlorella* spec.; Phypa: *Physcomitrella patens*; Selmo: *Selaginella moellendorfii*; Volca: *Volvox carteri*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Homologs in / cluster size</th>
<th>Putative function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phypa_110540</td>
<td>Pp1s317_20</td>
<td>Chlre / 2, Chlsp / 1, Phypa / 2, Selmo / 1, Volca / 2</td>
<td>putative acetate kinase (anaerobic metabolism)</td>
</tr>
<tr>
<td>Phypa_234235</td>
<td>Pp1s128_98</td>
<td>Phypa / 2, Selmo / 1</td>
<td>putative phytanoyl-CoA dioxygenase (alpha oxidation)</td>
</tr>
<tr>
<td>Phypa_165031</td>
<td>Pp1s81_153</td>
<td>Moss-specific</td>
<td>unknown</td>
</tr>
<tr>
<td>Phypa_162755</td>
<td>Pp1s47_69</td>
<td>Moss-specific</td>
<td>unknown</td>
</tr>
<tr>
<td>Phypa_163182</td>
<td>Pp1s53_22</td>
<td>Similarity to metazoan sequences</td>
<td>putative C1orf93 homolog (unknown)</td>
</tr>
<tr>
<td>Phypa_163217</td>
<td>Pp1s53_159</td>
<td>Phypa / 4, Chlre / 1</td>
<td>putative D-arabinono-1,4-lactone oxidase (biosynthesis of dehydro-D-arabinono-1,4-lactone)</td>
</tr>
<tr>
<td>Phypa_93324</td>
<td>Pp1s240_100</td>
<td>Phypa / 2, Selmo / 1</td>
<td>nucleic acid binding protein (unknown)</td>
</tr>
<tr>
<td>Phypa_165160</td>
<td>Pp1s83_192</td>
<td>No cluster of homologs but BLAST hits seed plants</td>
<td>putative glutathione S-transferase (detoxification)</td>
</tr>
<tr>
<td>Phypa_177778</td>
<td>Pp1s26_72</td>
<td>No cluster of homologs but BLAST hits in seed plants</td>
<td>putative glutathione S-transferase (detoxification)</td>
</tr>
</tbody>
</table>
Supplemental Table S5. Proteins chosen for internal tagging at the endogenous locus with Citrine and subcellular localization in distinct lineages. Evolutionary hypotheses derived from proteomics data of moss (this study) and other mitochondrial/plastid model organisms were tested by internal tagging with a Citrine-linker construct at the endogenous genomic locus in moss.

<table>
<thead>
<tr>
<th>Description/Locus ID</th>
<th>Biological Process</th>
<th>Ophistokonts</th>
<th>Green algae</th>
<th>P. patens</th>
<th>A. thaliana</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hydroxypyruvate reductase (HPR)/Pp1s414_8V6</td>
<td>Photorespiration</td>
<td>Mitochondria, peroxisomes (dependent on species)</td>
<td>Peroxisomes (Fig. 4)</td>
<td>Peroxisomes</td>
<td></td>
</tr>
<tr>
<td>Multifunctional protein (MFP)/Pp1s75_175V6</td>
<td>Beta-oxidation</td>
<td>Mitochondria, peroxisomes</td>
<td>Mitochondria, peroxisomes</td>
<td>Peroxisomes (Fig. 4)</td>
<td>Peroxisomes</td>
</tr>
<tr>
<td>SAM Synthase (SAMS)/Pp1s109_127V6</td>
<td>S-adenosyl-methionine synthesis</td>
<td>Cytosol</td>
<td>Mitochondria</td>
<td>Cytosol (Fig. 4)</td>
<td>Cytosol</td>
</tr>
<tr>
<td>HIV-tat-interactive protein 2-homolog (HTI)/Pp1s79_8V6</td>
<td>Stress response, cell death (?)</td>
<td>Mitochondria, cytosol, nucleus</td>
<td>Unknown</td>
<td>Cytosol, mitochondria (Fig. 4)</td>
<td>Not conserved</td>
</tr>
<tr>
<td>Fumarase/Pp1s49_154V6</td>
<td>TCA</td>
<td>Mitochondria, cytosol</td>
<td>Mitochondria</td>
<td>Mitochondria, cytosol (Fig. 5)</td>
<td>Mitochondria, cytosol</td>
</tr>
<tr>
<td>Caseinolytic protease proteolytic subunit (ClpP)/Pp1s102_15V6</td>
<td>Mitochondrial quality control</td>
<td>Mitochondria</td>
<td>Mitochondria</td>
<td>Mitochondria (Fig. 5)</td>
<td>Mitochondria</td>
</tr>
</tbody>
</table>
Supplemental Text

So far, large scale mitochondrial proteome data sets are available from the following model organisms: human (Taylor et al., 2003), mouse (Mootha et al., 2003; Pagliarini et al., 2008), rat (Forner et al., 2006), the yeast Saccharomyces cerevisiae (Reinders et al., 2006; Pagliarini et al., 2008), the protist Tetrahymena thermophila (Smith et al., 2007), the green alga Chlamydomonas rheinhardtii (Atteia et al., 2009), rice (Huang et al., 2009) and the model flowering plant Arabidopsis thaliana (Millar et al., 2001; Heazlewood et al., 2004). These were used for our comparative analyses.

Biosynthesis of ascorbate and ascorbate-glutathione cycle

As mitochondria contain an electron transport chain that generates ROS, the scavenging of free radicals is crucial to the mitochondrial response to oxidative stress. The final step in the biosynthesis of the important antioxidant L-ascorbic acid is catalyzed by the mitochondrial L-galactono-1,4-lactone dehydrogenase (Wheeler et al., 1998). A putative L-galactono-1,4-lactone dehydrogenase was also identified in moss mitochondria (Pp1s34_241V6). Animals synthesize ascorbate via a different pathway where L-gulono-1,4-lactone is oxidized to yield ascorbate (Valpuesta and Botella, 2004). Some fungi as e.g. S. cerevisiae do not produce ascorbate but its analogue dehydro-D-arabinono-1,4-lactone which is synthesized via D-arabinono-1,4-lactone. Interestingly, in addition to the ortholog to plant L-galactono-1,4-lactone dehydrogenase, two additional putative aldonolactone oxidoreductases (Pp1s53_159V6, Pp1s96_200V6), which show similarities to animal L-gulono-1,4-lactone oxidases and fungal D-arabinono-1,4-lactone oxidases, are present in moss mitochondrial fractions. A phylogenetic tree of putative aldonolactone oxidoreductases is depicted in Supplemental Figure S11. Pp1s96_200V6 and Pp1s53_159V6 are present in moss mitochondria, the first experimental evidence for a mitochondrial localization in their subfamilies. Pp1s96_200V6 is most closely related to AT5G11540 (AtGulLO3), which has been implicated in the synthesis of L-ascorbic acid using L-gulono-1,4-lactone (Maruta et al., 2010) as a substrate. An alternative pathway for de novo biosynthesis of L-ascorbate using GDP-L-gulose has already been discussed for plants (Valpuesta and Botella, 2004), and an important enzyme for the allocation of precursors to the distinct pathways, a putative GDP-D-mannose 3’,5’-epimerase (Pp1s171_56V6), was also identified in moss mitochondrial fractions. Phylogenetic analysis of the protein family (Supplemental Figure S11) has shown that two clades of putative aldonolactone oxidoreductases are not yet characterized and might contain enzymes catalyzing alternative metabolic pathways leading to ascorbate. Moreover one uncharacterized subfamily of putative aldonolactone oxidoreductases contains algal and four distinct moss sequences but no vascular plant sequence. The proteome data suggest the existence of ascorbate biosynthesis via a putative non-vascular plant specific pathway, a plant-like and an animal-like pathway in moss mitochondria.
Cysteine biosynthesis

OAS-TL (O-acetylseryine(thiol)lyase) catalyzes the final step of cysteine synthesis, which is the incorporation of sulfide into the amino acid backbone of O-acetylseryine (OAS). In flowering plants as well as in mosses, sulfide is produced by the assimilatory sulfate reduction pathway (ASRP) in plastids (Wiedemann et al., 2010; Birke et al., 2012). OAS is synthesized by serine acetyltransferase (SAT) of which cytosolic, plastidic and mitochondrial isoforms exist in Arabidopsis (Kawashima et al., 2005). In Arabidopsis, OAS-TL isoforms equally localize to cytosol, plastids and mitochondria (Hell, 2011) whereas most cysteine is synthesized in the cytosol and plastids. Mitochondrial AtOAS-TLC does not contribute much to the synthesis of cysteine but regulates cysteine synthesis indirectly by binding to AtSAT3 which enables the production of the cysteine precursor OAS. We identified two genes encoding OAS-TL isoforms in P. patens (Pp1s17_59V6 and Pp1s71_187V6) which were attributed L/H ratios of 0.10 (median) and 0.06 (one replicate only), respectively, indicating a predominantly plastidic localization. Transient overexpression of GFP fusion construct beginning at different methionines showed that Pp1s17_59V6 is dual targeted to the cytosol and plastids whereas Pp1s71_187V6 was imported into plastids (Birke et al., 2012). The low L/H ratio of the moss OAS-TL isoforms confirms the absence of a mitochondrial isoform in P. patens and thus a distinct compartmentation and less complex regulation of cysteine synthesis compared to flowering plants.

Photorespiration

Hydroxybutyrate reductase (HPR) catalyzes the conversion of hydroxybutyrate to glycerate, a step in photorespiration in plants. Photorespiration comprises a series of enzymatic reactions partitioned among three separate organelles in flowering plants: plastids, mitochondria and peroxisomes. It recycles phospho-glycolate which is formed by the oxygenase activity of RuBisCO (ribulose-1,5-bisphosphate-carboxylase/oxygenase) to be reused in the assimilatory Calvin cycle (Peterhansel, 2010). In chlorophytes as Chlamydomonas reinhardtii, mitochondria catalyze nearly all steps of photorespiration in a modified pathway: they possess a mitochondrial glycolate dehydrogenase instead of the flowering plant peroxisomal glycolate oxidase. Consequently enzymes of photorespiration including a hydroxybutyrate reductase (Chlre3_128310) were identified in the mitochondrial proteome of Chlamydomonas (Atteia et al., 2009). In contrast, the phylogenetic sister group to land plants, the charophytes, contain already a glycolate oxidase and leaf-type peroxisomes (Stabenau and Winkler, 2005). HPR-Citrine (Pp1s414_8V6) does not co-localize with mitochondria in moss, at least under the culture conditions used for isolation of mitochondria (Fig. 4). In consequence, the conversion of hydroxybutyrate to glycerate occurs in peroxisomes in P. patens. HPR-Citrine fluorescence has shown that HPR is an abundant protein in the peroxisomal matrix in all cell types.

Beta-oxidation

Pp1s75_175V6 encodes a putative enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase, also termed multifunctional protein (MFP) which is involved in several enzymatic steps of the
The core reaction of beta-oxidation. Beta-oxidation is necessary for the breakdown of fatty acids to gain metabolic energy and is important for the mobilization of storage triacylglycerols during seed germination in oilseed plant peroxisomes (Rylott et al., 2006). In addition, beta-oxidation plays a role in lipid turnover and hormone synthesis (jasmoneate and auxins). MFP2 (AT3G06860) is a well-established peroxisomal marker in A. thaliana (Tian et al., 2004). In animals and fungi, beta-oxidation takes place partly in mitochondria (for review, see (Graham and Eastmond, 2002)) and enzymes of beta-oxidation have been identified in C. reinhardtii mitochondria, including a multifunctional protein (Chlre3_132158) (Atteia et al., 2009). In algae, enzymes of beta-oxidation are present in both peroxisomes and mitochondria, exclusively in mitochondria or exclusively in peroxisomes (Stabenau et al., 1984; Gross, 1989). In Mougeotia, a charophyte, most activity localized to peroxisomes (Stabenau et al., 1984). The confocal images (Fig. 4) showed a clear localization in the peroxisomal matrix and not in mitochondria. This suggests that proteins of beta-oxidation are abundant peroxisomal proteins in moss. In conclusion, peroxisomal beta-oxidation is either an ancestral feature lost in some lineages of green algae which do not possess leaf-type peroxisomes or it developed before the ancestors of charophytes split off from the lineage leading to land plants (Fig. 4F).

SAM synthesis

S-adenosylmethionine synthase (SAMS) catalyzes the formation of the methyl donor S-adenosylmethionine (SAM) which is needed for transmethylation reactions involving nucleic acids, lipids, proteins, and metabolites (Hanson and Roje, 2001). Following the current model of methionine and SAM synthesis in higher plants, SAM is exclusively synthesized in the cytosol and subsequently imported into organelles to be used by methyltransferases. The resulting S-adenosylhomocysteine is exported back to the cytosol and regenerated in the activated methyl cycle involving S-adenosylhomocysteine hydrolase, methionine synthase, and SAMS. Methionine synthase is additionally present in flowering plant chloroplasts (Ravanel et al., 2004). A putative additional pathway may exist in non-vascular plants, as a SAMS (Chlre3_182408) as well as a cobalamin-independent methionine synthase (Chlre3_154307) were present in the mitochondrial proteome of C. reinhardtii (Atteia et al., 2009), and at least two distinct SAMSs and a cobalamin-independent methionine synthase were identified in the moss mitochondrial proteome in this study. The putative SAMS Pp1s109_127V6 was internally tagged with Citrine and localizes to the cytosol as well as the nucleus with the fluorescence being brightest in meristematic tip cells of moss protonema (Fig. 4). Thus Pp1s109_127V6 was found in mitochondrial fractions without being imported into mitochondria. It is possible that the identifications in the study of C. reinhardtii mitochondria equally have their origin in a contamination by cytosolic proteins. However, the identification of SAMS and methionine synthases in mitochondria of non-vascular plants seems not at random as other cytosolic contaminants, except ribosomal proteins, are largely lacking. It is thinkable that these enzymes are cytosolic but associate dynamically with mitochondrial outer membranes, putatively in order to facilitate substrate transport into the organelles as it was shown for glycolytic enzymes (Graham et al., 2007). Notably, the protein abundance of the investigated SAMS is increased in meristematic cells, where substantial epigenetic modifications take place (Mosquina et al., 2009). Epigenetic regulation by DNA or histone methylation requires S-adenosylmethionine as methyl donor which may explain the increased abundance of SAMS.
Cell-death related protein HTI-like (HIV TAT-interactive protein–like)

Pp1s79_8V6 (HTI-like) is a homolog of a mammalian oxidoreductase which interacts with the TAT (transactivating regulatory protein) protein of HIV and is called HIV-1 TAT-interactive protein 2 (HTATIP2) or 30kDa TAT-interactive protein (TIP30). Depending on the splice variant, the protein can have pro- or antiapoptotic effects and was linked to the regulation of nuclear import as well as to respiratory activity in mitochondria (Whitman et al., 2000; King and Shtivelman, 2004; Chen and Shtivelman, 2010). Putative homologs in the plant kingdom are present in the green algae Chlamydomonas reinhardtii and Volvox carteri, the moss Physcomitrella patens and in the flowering plants Ricinus communis and Oryza sativa. HTI-Citrine fluorescence in the stable lines was very weak and only perceivable in the tip cells of young cultures. Confocal microscopy revealed that HTI-Citrine is cytosolic and co-localizes with part of the mitochondria in some cells (Fig. 4). The yeast ortholog YER004W has been found by MS/MS in the outer membrane of mitochondria (Zahedi et al., 2006) whereas the expression is increased after treatment with a DNA-damaging agent (Dardalhon et al., 2007). Thus HTI-like proteins are candidates for ancient stress and cell-death related processes in animals, fungi and plants.

Conservation and unknown function of mitochondrial proteins

In the moss mitochondrial data set, we identified 18 proteins which were not assigned any function, but which are evolutionary conserved as shown in Supplemental Table S3 and thus represent yet uncharacterized mitochondrial protein families. The protein domains present suggest that their function is likely related to diverse processes important for mitochondria as stress and autophagy (HVA-like proteins), RNA editing (MAM33), protein-protein interactions (MA-3, TPR), regulation and signaling (CBS, GRAM) and morphology (mitofilin).

In contrast, nine proteins identified in this study do not have any homologs in vascular plants and are therefore candidates for mitochondrial functions that have been lost during land plant evolution (Supplemental Table S4). Pp1s317_20V6 encodes a putative acetate kinase whose homolog was identified in C. reinhardtii mitochondria but which has no homologs in land plants except in the lycophyte Selaginella moellendorfii. In C. reinhardtii, the acetate kinase is thought to work together with a phosphotransacetylase in a typically eubacterial pathway for acetate to acetyl-CoA conversion (Ingram-Smith et al., 2006) important for acetate assimilation or metabolism under anaerobic conditions (Atteia et al., 2006). After the divergence of lycophytes this pathway was putatively lost in monilophytes and seed plants, which in general possess a taller plant body less prone to submersion than mosses and lycophytes. Thus changes in anaerobic metabolism may be linked to the evolution of plant body plans.

Further, we identified a putative formate dehydrogenase (Pp1s106_16V6), an enzyme regulating the concentration of one-carbon metabolites that is present in flowering plant mitochondria but is thought to be absent in algal mitochondria (Atteia et al., 2006).
Supplemental References


