Running title: Plastoglobular functions

#for correspondence and proof: Klaas J. van Wijk, Dept. of Plant Biology, Emerson Hall 332, Cornell University, Ithaca, NY 14853. Tel: 1-607-255-3664; Fax: 1-607-255-3664; Email: kv35@cornell.edu

Note for the reviewers:
All mass spectral data (the mgf files reformatted as PRIDE XML files) are available via the Proteomics Identifications database (PRIDE) at http://www.ebi.ac.uk/pride/. Accession numbers: 18969-18988;
Reviewer Account- Username: review54381; Password: D#nejxk#
The functional network of the *Arabidopsis thaliana* plastoglobule proteome based on quantitative proteomics and genome-wide co-expression analysis

Peter K Lundquist, Anton Poliakov, Nazmul H. Bhuiyan, Boris Zybaiov, Qi Sun, Klaas J. van Wijk

*Department of Plant Biology, Cornell University, Ithaca, NY 14853, USA.*

*Computational Biology Service Unit, Cornell University, Ithaca, NY 14853, USA*

#for correspondence: Klaas J. van Wijk, Dept. of Plant Biology, Emerson Hall 332, Cornell University, Ithaca, NY 14853. Tel: 1-607-255-3664; Fax: 1-607-255-5407; Email: kv35@cornell.edu

1 Current address: Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, 4301 W. Markham, Slot #516, Little Rock, AR 72205

**ABSTRACT** Plastoglobules (PGs) in chloroplasts are thylakoid-associated monolayer lipoprotein particles containing prenyl- and neutral lipids, and several dozen proteins mostly with unknown functions. An integrated view of the role of the PG is lacking. Here we better define the PG proteome and provide a conceptual framework for further studies. The PG proteome from Arabidopsis leaf chloroplasts was determined by mass spectrometry of isolated PGs and quantitative comparison with the proteomes of unfractionated leaves, thylakoids and stroma. Scanning electron microscopy showed the purity and size distribution of the isolated PGs. Compared to previous PG proteome analyses, we excluded several proteins and identified six new PG proteins, including an M48 metallopeptidase and two ABC1 atypical kinases, confirmed by immunoblotting. This refined PG proteome consisted of 30 proteins, including 6 ABC1 kinases and 7 fibrillins together comprising more than 70% of the PG protein mass. Other fibrillins located predominantly in the stroma or thylakoid and not PG; we discovered that this partitioning can be predicted by their isoelectric point and hydrophobicity. A genome-wide co-expression network for the PG genes was then constructed from mRNA expression data. This revealed a modular network with four distinct modules that each contained at least one ABC1K and/or fibrillin gene. Each module showed clear enrichment in specific functions, including chlorophyll degradation/senescence, isoprenoid biosynthesis, plastid proteolysis and plastid redox and kinase regulators of electron flow (e.g. thioredoxins, STN7, RAP38). We propose a new testable model for the PGs, in which sets of genes are associated with specific PG functions.
INTRODUCTION

Plastoglobules (PGs) are lipoprotein particles localized in the various types of photosynthetic and non-photosynthetic plastids of photosynthetic organisms. In chloroplasts, PGs are contiguous with the thylakoid membrane – reviewed in (Brehelin et al., 2007; Brehelin and Kessler, 2008; Singh and McNellis, 2011). PGs can be released from the thylakoid membrane by sonication, and purification based on their low density has facilitated analysis of their metabolites (Lohmann et al., 2006; Vidi et al., 2006; Gaude et al., 2007; Zbierzak et al., 2009) and protein composition (Vidi et al., 2006; Ytterberg et al., 2006). This has shown that PGs contain various prenyl lipids and neutral lipids, including plastoquinone, phylloquinone, α-tocopherol, fatty acid phytyl esters, and triacylglycerol. For many years it was believed that PGs solely represented a passive lipid deposition site for the plastid. Thus, the discovery of several dozen PG-localized proteins, mostly with unknown functions, was highly surprising (Vidi et al., 2006; Ytterberg et al., 2006). The PG proteome contains a number of proteins of which only a few, such as tocopherol cyclase (VTE1) and allene oxide synthase (AOS), have an established function. In addition to these various (putative) enzymes, the PG proteome contains a number of proteins of the fibrillin (FBN) family, initially believed to play more structural functions. However, based on the presence of lipocalin domains in some of these FBNs, it has recently been suggested that they may also play a role in metabolite transport (Singh and McNellis, 2011). Finally, the PG proteome also contains a number of ABC1K proteins; these are putative atypical kinases with homology to an ABC1K in yeast and E. coli, where they regulate ubiquinone synthesis (Poon et al., 2000; Tauche et al., 2008; Xie et al., 2011). The presence of several of the ABC1K proteins in PGs is puzzling and we earlier suggested that they play a central role in regulation of PG metabolism (Ytterberg et al., 2006; Brehelin et al., 2007).

PGs appear to play a role in chloroplast development, senescence, and stress defense. Their shrinking and swelling in response to (a)biotic stresses and during developmental transitions, as well as in plastid biogenesis mutants is well documented (Gaude et al., 2007; Simkin et al., 2007; Singh et al., 2010; Zhang et al., 2010). Recent results suggest that PGs are involved in active channeling of hydrophobic metabolites between the thylakoid and PG, permitted by the contiguous association of the two structures (Austin et al., 2006; Gaude et al., 2007). In particular during various abiotic stresses (e.g. N-starvation, drought or light stress), but also during senescence, components of thylakoid degradation, such as fatty acids and phytol tails from chlorophyll, are channeled into the PGs, likely accounting for the massive swelling. Within the PG, several of the observed proteins likely play a role in the recycling of such thylakoid catabolites, in addition to a role in synthesis of isoprenyl lipids such as tocopherol and plastoquinone (Vidi et al., 2006; Ytterberg et al., 2006).

Despite the recent progress on PG analysis, it remains unclear how the PGs fit into plastid metabolism and chloroplast homeostasis, mostly because functions for many PG-localized proteins are
unknown. Key questions about PGs are: i) what determines and controls their size, shape and content, ii) how are proteins recruited to the PG proteome and how does the PG proteome change in response to changes in developmental state or (a)biotic conditions, iii) what are the functions of the PG proteins and how are they related to each other? The current study aims to provide a better framework to help answer these questions by defining a functional network.

We first examined the quantitative protein composition of PGs isolated from leaves subjected to 5 days of increased light intensity (500 umol photons.m\(^2\).s\(^{-1}\)) and compared this quantitatively against proteomics datasets of leaf, thylakoid, and stroma preparations to identify proteins specifically enriched in the PG. Because we used a far more sensitive mass spectrometer than used in previous PG proteome analyses (Vidi et al., 2006; Ytterberg et al., 2006), combined with both in-gel and in-solution digestions, we expected to discover more low abundant members of the PG proteome. Indeed, we identified a number of new PG proteins, including an M48 metallopeptidase (M48), two additional ABC1K proteins and a senescence-associated gene (SAG) protein. The surprising localization of M48, as well as two ABC1K proteins, to PGs was confirmed by immunoblotting.

Transcripts or proteins involved in related biological pathways or complexes often accumulate simultaneously. Therefore, co-expression often implies the presence of functional or physical linkages between genes or proteins, allowing for identification of new components of processes or protein complexes. Indeed, co-expression analysis has been used extensively in plant biology to identify putative protein functions and determine physical or functional connections between proteins (Cartieaux et al., 2003; Rohde et al., 2004; Biehl et al., 2005; Vanderauwera et al., 2005; Dalcorso et al., 2008; Sawada et al., 2009; Takabayashi et al., 2009; Bischoff et al., 2010; Fu and Xue, 2010; Ozaki et al., 2010; Lin et al., 2011).

Here we employed such a transcriptional genome-wide co-expression analysis, using the core PG proteome as input, to provide a better framework for PG functions and to associate PG proteins to functional activities. This identified a co-expression network with 4 modules, each with a specific set of enriched functions, including plastid proteases, redox regulators, cyclic electron flow components and genes encoding for a specific subset of proteins involved in plastid prenyl-lipid metabolism. Specific ABC1K proteins and FBNs were centrally positioned in different modules within the network. This study better defines the core PG proteome and its functions in leaves. We propose a new conceptual model for the PGs, suggest a parallel to lipid rafts, and provide an intellectual and practical framework for further analysis.

**RESULTS**
**Size distribution, extractability, coalescence, and purity of isolated PGs from light-shifted Arabidopsis leaves** As starting material for our study, we grew Arabidopsis plants on soil for 2.5 weeks at 120 µmol photons.m\(^{-2}\).s\(^{-1}\) and transferred them to 520 µmol.m\(^{-2}\).s\(^{-1}\) for 5 days. The higher light intensity increased PG volume and yield and made them more amenable for experimental analysis (Ytterberg, 2006 #10500). The mild light stress treatment accelerated vegetative growth, increased anthocyanin accumulation in the leaves, and resulted in only minor visible damage at the edges of the oldest leaves, but no damage to younger leaves (Supplemental Fig. 1).

Isolation of PGs from the Arabidopsis leaf rosettes was performed by sonication of isolated thylakoid membranes followed by flotation density centrifugation, as in (Ytterberg et al., 2006). The enrichment of PGs was confirmed by immunodetection of the VTE1 protein (Fig. 1A), known to be uniquely localized in the PG as determined by YFP localization and immunogold labeling (Vidi et al., 2006), immunoelectron tomography (Austin et al., 2006) and proteomics (Vidi et al., 2006; Ytterberg et al., 2006). Densitometric analysis of the immunoblots indicated a more than 400-fold enrichment of VTE1 in the PG preparations compared to the original thylakoid membranes (the starting material). We also measured a ~4-fold depletion of VTE1 in thylakoids following sonication, indicating that ~75% of the PG material is extracted from the thylakoids by sonication (Fig. 1B). Hence, our results demonstrate that the isolated PGs were highly enriched for PG particles and that the majority of the PGs were successfully extracted from the thylakoid membrane.

Transmission electron micrographs (TEM) of PGs *in vivo* have demonstrated a broad size distribution, even within the same chloroplast (Austin et al., 2006), however the relationship between PG size and PG proteome is not known. For a correct and meaningful quantitative and qualitative analysis of the PG proteome, it was therefore critical to extract PGs representing the entire *in vivo* population, without bias for size or other (unknown) properties, while keeping contamination from thylakoids and other compartments to a minimum. Therefore we compared the size distribution of the extracted PG particles to the *in vivo* size distribution. TEM of the leaf tissue showed a broad distribution of PG sizes, with diameters ranging from ~175 nm to ~600 nm, and peaking between 250-350 nm (Figs. 1C and D). PG preparations were analyzed by scanning electron microscopy (SEM) and also showed a broad size distribution from ~50 to ~600 nm, peaking between 100-250 nm (Fig. 1C and E). This demonstrated that PGs of all physiologically relevant sizes were extracted efficiently, with a small bias to smaller particles. Interestingly, the micrographs of PG preparations sometimes showed PGs in grape-like clusters, similar to those described in TEM micrographs of leaves (Rey et al., 2000; Austin et al., 2006; Simkin et al., 2007; Zbierzak et al., 2009) (Supplemental Fig. 2A and B). Despite the clustering found in the preparations, each PG clearly maintained its individual structure and they did not coalesce. Apparently, component(s) at the PG-solution interface act to maintain PG structural integrity and are extractable with the PGs, likely
a FBN coat surrounding the PG periphery. Evidence of minor amounts of thylakoid membrane fragments was also found in the micrographs. While the isolated PGs demonstrated remarkably smooth surfaces, the SEM also showed infrequent amorphous structures scattered generally attached to PGs (Supplemental Fig. 2C). The size of these structures, the presence of attached PGs, and their amorphous shape suggested that they are thylakoid membrane fragments. Importantly, these amorphous structures were far less abundant than the PGs, indicating high PG purity, which was further confirmed by the proteomics experiments (see next section). Summarizing, our results demonstrate that more than ~75% of the PGs, from all physiologically relevant sizes, were successfully extracted from the thylakoid membrane into highly enriched PG preparations.

**Improved coverage and quantification of the PG proteome** The first comprehensive PG proteome analyses were carried out using a Q-TOF mass spectrometer (Ytterberg et al., 2006) or a LCQ Deca XP ion trap mass spectrometer (Vidi et al., 2006). Recent improvements in sensitivity, mass accuracy, and speed of mass spectrometers have enabled detection of lower abundant proteins in complex mixtures and also facilitate mass spectrometry-based label-free proteome quantifications using spectral counting (Bantscheff et al., 2007; Mann and Kelleher, 2008; Domon and Aebersold, 2010). Thus a much more sensitive and quantitative analysis of the PG proteome should now be possible. The spectral counting technique is based on the observation that the number of successful MS/MS acquisitions of peptides coming from a protein shows a positive and linear correlation to the relative concentration of this protein in the studied sample (Liu et al., 2004; Old et al., 2005; Zybailov et al., 2005; Sandhu et al., 2008). Spectral counting is particularly effective to detect large quantitative differences, as expected in our study where we compare (sub)cellular or suborganellar fractions that are very different in function and composition. We previously optimized the SPC workflow and tested it for *Arabidopsis* and maize organelles, cell-types and complexes, e.g. (Zybailov et al., 2008; Friso et al., 2010; Majeran et al., 2010; Olinares et al., 2011). The relative normalized abundance (relative mass contribution) of each protein within each sample, NadjSPC, was calculated from the number of adjusted matched MS/MS spectra (adjSPC), normalized to the total adjSPC per sample, as defined previously (Friso et al., 2010). Thus a protein with a NadjSPC= 0.01 contributes ~1% of the protein mass of the analyzed sample. As a general rule, the accuracy of quantification improves with the number of adjSPC per protein.

Here we employed a LTQ-Orbitrap mass spectrometer (Hu et al., 2005) coupled to a nanoLC system, to search for additional, more low-abundant proteins located in the PG. Moreover, we re-evaluated previous assignments of proteins to the PG (Vidi et al., 2006; Ytterberg et al., 2006) based on quantitative comparative proteome information. Using three independent PG preparations, the PG proteome was analyzed in two different ways: i) PG proteins were separated by SDS-PAGE, each lane cut
in 5 slices, and in-gel digested with trypsin, and ii) unfractionated PGs were delipidized and digested in-solution with trypsin. These protein digests were then analyzed by MS/MS in the LTQ-Orbitrap (Fig. 2A). The rationale for using these two different protein extraction/separation methods was to: i) maximize detection of low abundant proteins, ii) increase the robustness of protein quantification, and iii) improve protein sequence coverage. Proteins identified by only a single peptide sequence, irrespective of post-translational modification or charge state, were discarded to increase the robustness of the analysis and avoid any false-positive protein identification; these proteins represented <1% of the protein mass in PGs.

**Defining the core PG proteome** The combined proteome analysis identified 234 proteins, with 129 identified by both in-solution and in-gel workflows and respectively 6 and 99 proteins identified in only the in-solution or in-gel digestions (Fig. 2A) (see Supplemental Table 1 for details). The 129 proteins identified by both methods represented ~99% of the PG protein mass, showing that only the least abundant proteins were not identified by both methods. The in-solution and in-gel methods showed a good correlation for the relative protein abundance for proteins with abundance >0.001, i.e. proteins that each represent >0.1% of the protein mass of the PGs (Fig. 2B - grey area). Protein sequence coverage was on average 26% for the in-gel method and 16% for the in-solution method; this increased to 37% and 27% respectively if we only considered the 129 proteins identified by both methods. The correlation of the average NadjSPC values (combining both in-gel and in-solution data) between the three biological replicates was excellent, with pair-wise correlation coefficients between 0.902 and 0.960 (Fig. 2C).

We then determined those proteins highly-enriched in the PGs, hereafter named the ‘core’ PG proteome, using the workflow as depicted in figure 3A. The core PG proteins were distinguished from non-plastid contaminants or proteins localized primarily elsewhere in the chloroplast by comparing the abundance in PGs to their average abundance in total leaf extracts (5 biological replicates with 2 replicates from Zybailov et al. (2009) and 3 from this study), and isolated thylakoid and stromal fractions (Zybailov et al., 2008). Supplemental Table 2 provides the quantitative and qualitative data about these proteomes. Furthermore, core PG proteins were required to have a minimal abundance (NadjSPC >0.001) and be observed in the PGs by both in-gel and in-solution methods (Fig. 3A).

For selection of core PG proteins, we first discarded those proteins with a PG/leaf abundance ratio below 10; only 52 proteins out of the 234 proteins passed this first filter. We emphasize that this was a relatively ‘relaxed’ minimal threshold considering that the PG proteome represents less than 10% of the leaf proteome – however, this was already very effective to remove non-plastid contaminants, as well as the abundant proteins of the photosynthetic apparatus, and other proteins not truly enriched in PGs. Importantly, it also removed several proteins that were earlier assigned to the PG, including fructose bisphosphate aldolase 1 and 2 (FBPA-1 and -2) (Vidi et al., 2006; Ytterberg et al., 2006). Furthermore,
FBN3a and FBN10 were also eliminated, because they showed PG/leaf abundance ratios of only 0.9 and 4.7 respectively (Table I; Supplemental Table 1). The relative abundance and distribution of the remaining 52 proteins between PGs, thylakoids and stroma is displayed in Figures 3B,C.

As a next step, we removed proteins that failed to show at least a 5-fold enrichment in the isolated PGs compared to the thylakoid (Fig. 3A). This resulted in removal of 4 proteins – a DnaJ domain protein, a glutaredoxin (GRX), a protein with an unknown function (At5g62140) and allene oxide synthase (AOS) (Fig. 3B). Finally, 4 proteins with a PG/stroma abundance ratio below 20 were discarded – these were thioredoxin m4 (TRXm4), UV-B/Ozone similarly-regulated protein (UOS1), an unknown protein with DUF1350 domain and fructose-bisphosphate aldolase-3 (FBPA-3) (Fig. 3B).

The remaining 44 proteins were then evaluated for abundance and frequency of identification in the PG proteome analysis (Fig. 3C). 12 proteins with a relative abundance below NadjSPC of 0.001 (corresponding to ≤0.1% of the protein mass) (Fig. 3C), or only identified by one of the methods, were discarded (We note that none of these proteins were co-expressors of the PG core genes – see further below). Finally, we manually evaluated the remaining 32 proteins for known subcellular localization and/or function. Two proteins were discarded from the core proteome based on literature evidence. The extra-plastidic calcoxin protein RD20 (AT2G33380) has been shown to be localized in cytosolic lipoprotein particles (Aubert et al., 2010), while a PLAT/LH2 domain protein (AT2G22170) is likely ER-localized based on GFP-tagging (http://gfp.stanford.edu/index.html). Thus, these proteins were removed from the final list.

Because the PGs were isolated from plants shifted for 5 days to higher light intensities (500 μmol photons.s⁻¹.s⁻¹), we also determined and quantified the total leaf proteome of these plants (three independent replicates; Supplemental Table 2). However, using these quantitative total leaf proteome data in the workflow (of Figure 3A) did not affect the final selection of core PG proteins.

The core PG proteome Table II summarizes the core PG proteome with their relative abundance (including the coefficient of variation - CV) and enrichment as compared to other plastid compartments. The CV of protein abundance across the 3 biological replicates was 24%, indicating an excellent reproducibility. Twenty-three of the 30 core proteins were previously assigned to the PG in (Vidi et al., 2006) or (Ytterberg et al., 2006), with 18 identified in both studies (Table II). VTE1 showed a PG/thylakoid ratio of 131, consistent with the high ratio determined by the immunoblot analysis (Fig. 1A), and was not detected in chloroplast stroma.

Another seven proteins were newly identified as plastoglobular, namely two ABC1 kinases (ABC1K6 and K7), a PLAT/LH2-domain protein (PLAT/LH2-1), an esterase-domain protein (Esterase 1), two proteins of unknown function (Unknown-2 with DUF1350 and Unknown Senescence-Associated
Gene - SAG), and a metallopeptidase M48-domain protein. Six of these seven proteins are the lowest abundant proteins of the core PG proteome (Table II), explaining their previous lack of detection. Thirteen proteins previously assigned to the PG did not pass our filters (Table II); these were AOS, FBPA-1,2,3, FBN3a, two RNA-associated proteins (Rap38&41), an ATPase, WAVE3, peroxiredoxin Q, and three proteins of unknown function (see DISCUSSION).

The four most abundant proteins were all FBN proteins (FBN1a, 1b, 2, and 4) and were also found previously to be the homologues of the FBNs in red pepper chromoplast plastoglobules (Ytterberg et al., 2006), suggesting they may hold a general function in maintenance of plastid lipid body structure. The six FBN core proteins constituted 53% of the PG proteome mass (Fig. 3D). The second most abundant class of core proteins consisted of six ABC1 kinases, together constituting 19% of the core PG proteome mass. The original ABC1K proteins, identified in *Saccharomyces cerevisiae* (Abc1p/Coq8p) and *Escherichia coli* (UbiB), are implicated in regulation of ubiquinone metabolism (Poon et al., 2000; Do et al., 2001). In particular, phosphorylation of several members of the ubiquinone biosynthetic complex is dependent on Abc1p/Coq8p (Xie et al., 2011). However, the role and possible targets of the PG-localized ABC1K proteins are unknown. Carotenoid cleavage dioxygenase 4 (CCD4), with specificity for 8’-apo-β-caroten-8’-al in *Arabidopsis* (Huang et al., 2009), was 3.3% of the proteome mass. The VTE1 protein, involved in tocopherol biosynthesis (Porfirova et al., 2002), was 2.6% of the proteome mass and the NAD(P)H dehydrogenase C1 (NDC1), which reduces plastoquinone to plastoquinol and is necessary for phylloquinone synthesis (Eugeni-Piller et al., 2011) was 2.5% of the proteome mass. The M48 protein was only 0.3% of the proteome mass (Table II).

**Confirmation of PG localization of ABC1K1, ABC1K3 and Peptidase M48 by immunoblotting** To further validate our quantitative proteomics analysis, we generated specific antisera against three PG core proteins, ABC1K1, ABC1K3 and the low abundant M48 protease because of its novelty as a potential PG-localized protease. Specific, polyclonal antisera were raised against affinity purified *E. coli* overexpressed domains of each of the three proteins. After confirming the specificity of the sera, we compared isolated thylakoid fractions and isolated PGs for protein abundance of M48, ABC1K1 and ABC1K3 using immunoblots. Figure 4 shows that ABC1K1, ABC1K3 and M48 were ~10, 20 and >50-fold enriched in isolated PGs compared to (untreated) thylakoids, in agreement with the PG/thylakoid abundance ratios of respectively 11, 16 and 17 measured by mass spectrometry (Table II). This provides independent evidence that metallopeptidase M48 and the two ABC1K proteins are highly enriched in the PG, and indicate that our mass spectrometry–based quantitative analysis does provide reliable information about the core PG proteome.
Partitioning of the FBN proteins between PGs and the thylakoid or stroma

We identified all 12 known FBNs, as well as a FBN-like protein (AT1G18060), in our collective proteome data sets of leaves, chloroplast stroma, thylakoids and PGs (Table I; Supplemental Tables 1 and 2). However, we assigned only 7 FBNs to the PG core proteome, based on our quantitative analysis (Fig. 3A), because the other FBNs did not preferentially locate to PGs. Therefore, we searched for physical-chemical properties of the FBN protein family that correlated with subplastid localization. We also included two truncation products of FBN7a, FBN7a (1-133) and FBN7a (1-290), which localizations were determined by YFP-tagging as localized to stroma and PG, respectively (Vidi et al., 2007). The combination of isoelectric point (pI) and hydrophobicity (calculated as the GRAVY index) for each of the FBN correlated surprisingly well with their relative distribution between stroma, thylakoids and PG (Fig. 5; Table I). The FBN proteins could be placed in one of four groups: i) strongly enriched in the PG (at least 10-fold), ii) equal enrichment between the PG and thylakoid (PG/Thy ≈ 1), iii) strongly enriched in the thylakoid as compared to PG (>10-fold), and iv) stroma-localized, not identified in PG or Thy. All 7 PG-localized FBNs, as well as the truncated FBN7a (1-290), were found to display low pIs and (on average) higher hydrophobicity indices. Conversely, all 6 FBNs strongly enriched in the thylakoid membrane fraction displayed higher pIs and lower hydrophobicity indices. Importantly, FBN10, the only FBN with approximately equal ratio between PG and thylakoid (PG/thylakoid = 1.8) showed an intermediate pI and hydrophobicity index. Finally, the stroma-localized FBN5 and FBN7a (1-133) demonstrated a low pI and the lowest hydrophobicity index of the 16 proteins products. The pI and GRAVY index however did not predict subplastid localization of other members of the core PG proteome, likely because they have very diverse secondary structures.

A PG co-expression network shows strong, specific enrichment for genes of four plastid functions

Because the functions of most PG proteins are unknown and hard to predict, we employed a genome-wide transcript co-expression analysis to identify putative functions for the PG core proteins, identify potential targets for the ABC1K proteins, generate testable hypotheses and provide a better framework for further studies. Several co-expression analysis tools have been developed and employed in plant co-expression analysis, each offering their own suite of functions and set of normalized microarray experiments, e.g. see (Steinhauser et al., 2004; Mutwil et al., 2008; Usadel et al., 2009). We tested and compared three different publicly available co-expression tools – MetaOmGraph (Wurtele et al., 2007), BAR (Toufighi et al., 2005) and ACT (Manfield et al., 2006) - for their ability to identify co-expression relationships among functionally and physically associated gene products. Using the well-studied gene family encoding for the ClpPR protease complex in plastids and mitochondria (Olinares et al., 2011) and 10 genes encoding for enzymes involved in tetrapyrrole biosynthesis, we first demonstrated that although the three software programs, MetaOmGraph, BAR and ACT, show quantitative differences in co-expression rankings, true
co-expressers were consistently found (see Supplemental text and Supplemental Fig. 3). We also tested to see if PG core genes preferentially expressed with other PG core genes, rather than genes encoding for plastid proteins in general or with genes encoding for extra-plastidic proteins. This showed that PG core genes generally preferentially co-express with other PG genes at higher Pearson correlation coefficients (PCC) (Supplemental text and Supplemental Fig. 4). Importantly, these tests also showed that genes encoding for plastid proteins are clearly not co-expressed as a single group, and thus that we should be able to find specific co-expression patterns for PG core proteins. The selection of test sets, procedures and results are described in more detail in the Supplemental Text.

We chose to employ the MetaOmGraph program to investigate the PG co-expression network because of its user-friendly nature and validated the final results with the other two programs. We discarded from the analysis those probes measuring multiple genes to ensure that we were testing specific gene-gene co-expression relationships. The resulting set contained 21158 Affymetrix microarray probes, including 25 of the 30 PG core genes. PG core genes FBN1a and 1b, as well as DGAT4 had to be excluded because they were not represented by unique probes (see further in DISCUSSION), whereas the SOUL and esterase 1 genes were not represented on the microarrays.

A PG network was constructed from a genome-wide search for each of the 25 PG core genes on the Affymetrix microarray. Some of the PG genes (e.g. FBN2, 4, Aldo/keto reductase, unknown-2) had several hundred co-expressing genes above a PCC threshold of 0.7 (or in some cases even above 0.8), whereas other genes (ABC1K7, UbiE-2, M48, DGAT3, unknown SAG) had none above that threshold. Therefore, we used the twenty strongest co-expressing genes for each PG core gene to construct a PG network, rather than applying a minimal PCC threshold. All such co-expression relationships had a PCC above 0.65 with the exception of the PLAT/LH2 domain protein and the SAG protein with unknown function. Strong negative correlations between genes can be relevant, however negative PCC values never exceeded an absolute value of 0.67, and only positive correlations became therefore part of the PG network.

The resulting network contained 374 nodes (genes) and 500 edges (co-expression interactions) (Fig. 6). Of the 374 nodes, 201 (54%) were assigned to the plastid based on experimental information (Supplemental Table 3). Interestingly, the PG core proteins differed strongly in the subcellular localization of their co-expressers. For instance, in the case of the 5 FBNs, NDC1, ABC1K3 and several others, 17-20 out of 20 of the co-expressers were plastid-localized, however core proteins VTE1, UbiE1, DGAT3, and PLAT/LH2 domain protein each had three or less plastid-localized co-expressers. This immediately suggests that the latter proteins are primarily post-transcriptionally regulated, or that their transcriptional regulation is integrated with extra-plastidic functions and needs.
To better assign functions to the PG core genes and the PG as a whole, co-expressing genes were categorized by their assigned functional category (using the MapMan bin system as the basis to organize the functions), and edges connecting to each bin were counted. Because some bins were much larger than others and thus had a much greater opportunity to be represented in the PG network, we normalized the representation of each bin by its size. As indicated in Table III, a strong enrichment was found for plastid-localized proteases (17 in total; LON, Prep1, EGY2, FtsH1, 2, 5, 8, 9, ClpR2, R3, P4, P5, P6, S, C1, D and DegP1, 8), proteins involved in cyclic/alternative electron flow (five NDH subunits, PGR5, PGRL1A, B, PTOX, PIFI) and regulators of the light reaction state transition kinase (STN7 and phosphatase TAP38), plastid-localized isoprenoid metabolism (in particular carotenoid metabolism PDS, ZDS, LYC, ZEP), chlorophyll degradation (PaO/ACD1, pheophytinase, ACD2), and plastid redox regulation (Trx-M1, 2, 4, Trx-F1, Fd-Trx-Reductase subunits, NADPH-reductase, PrxQ and others) (Table III, Supplemental Table 3). To further substantiate the findings from the MetaOmGraph co-expression analysis, we analyzed functional enrichment of the top 20 co-expressers using the two other software programs, BAR and ACT (Supplemental Tables 3 and 4). Clearly, the distribution of functional groups was consistent between all three programs, strengthening the significance of the MetaOmGraph analysis.

The PG co-expression network shows four modules  The co-expression network showed that most core genes had associations with other PG genes, producing a gene expression network with 4 clusters of nodes, which we refer to as ‘modules’. Modules are parts of biological networks in which nodes are densely connected with each other, but between which there are only sparse connections. Thus within each of these modules, genes co-express more tightly to each other than with genes outside the module (Figs. 6; Table III). The modular nature is an important property of biological networks. As will be detailed below, the four modules each showed enrichment for specific functions. The remainder of the PG core genes (FRed-1, UbiE-1, PLAT/LH2-1 and VTE1) had no or weaker associations with other PG genes (Fig. 6); moreover, they had in common that most of their co-expressors encoded for extra-plastidic proteins.

Module-1, with four PG core genes (DGAT3, ABC1K7, SAG and M48 metalloprotease), was enriched for senescence functions, in particular chlorophyll degradation (pheophorbide a oxygenase (PaO) and pheophytinase (PPH)) and a variety of proteases outside the plastid (including a senescence-associated cystein protease), as well as the senescence-induced Clp protease chaperone ClpD1. The two chlorophyll degradation enzymes co-expressed with ABC1K7 and SAG (Fig. 6 – nodes 1 and 2) and we note that a 3rd, more down-stream enzyme (red chlorophyll catabolite reductase; RCCR), was found in module-3 co-expressing with FBN4. Strikingly, only 35% of the edges in module-1 were plastid-localized, compared to 71-95% for the other modules, consistent with the observation that senescence
leads to controlled breakdown of the whole cell, and is not limited to plastids. We also point out an interesting plastid-localized putative tyrosine kinase that co-expressed with both DGAT3 and ABC1K7 (Fig. 6 – node #18). The role for M48 protease is completely unknown and its co-expressors included 5 plastid proteins with unknown function, and 4 plastid-localized proteins: PPH, ClpR3, thylakoid alternative oxidase (PTOX) and a glutaredoxin-thioredoxin.

The most extensive module (module-2) was centrally located in the PG network and comprised eight PG genes (ABC1K-1,3,6, aldo-keto reductase, NDC1, flavin-reductase 2, CCD4, UbiE-2). It was particularly enriched for carotenoid metabolism enzymes (see Supplemental Fig. 5 for the complete pathway and the connections to co-expressors) and plastid proteases (22 edges); 71% of the nodes encoded for plastid proteins indicating tight integration with plastid functions (Table III; Fig. 6). In addition to the plastid carotenoid enzymes, also upstream cytosolic solanesyl diphosphate synthase (SPS1) and its plastid isoform (SPS2-responsible for synthesizing the hydrophobic tail of plastoquinone – PQ9) were part of this module (we note that MEP pathway enzymes are only found in module-4). Interestingly, GOLDEN2-LIKE 1 (GPRI1) transcription factor, known to co-regulate expression of a suite of nuclear photosynthetic genes (Fitter et al., 2002), was also part of this module as a co-expressor of CCD4. Within module-2, ABC1K3, AKred and NDC1 were particularly tightly connected, mostly through co-expressing plastid proteases. The top 20 co-expressors of ABC1K3 were almost exclusively involved in carotenoid metabolism (ZEP, PDS, ZDS, CCD1), and protein degradation (FtsH1,2,5,8,9, ClpR3 and ClpC, PREP1, DegP1), but also included thylakoid kinase STN7, involved in phosphorylation of LHC proteins facilitating state-transitions to optimize electron flow (Rochaix, 2011).

The third module involved five PG core genes (ABC1K9, FBN2, 4, 7a and unknown-1) and was particularly enriched in redox regulators and ‘photoacclimation’. Eighty-nine percent of the edges encoded for assigned plastid proteins. The module was highly enriched for plastid redox regulators (including thioredoxins M-1, -2, -4, and F-1, two Fd-TRX reductase subunits, glutaredoxins) and non-linear electron flow components (NDH, PGRL1A, PIFI), as well as several plastid proteases (ClpR2,P5 and S; DegP1, FtsH2), (Fig. 6). Also the thylakoid phosphatase TAP38 and the gene coding for the ‘acclimation of photosynthesis to environment 1’ (APE1) were part of this module.

The fourth and smallest module contained PG core proteins FBN7b and unknown-2, and was strongly enriched for proteins involved in various aspects of plastid biogenesis, including proteases, and the Calvin cycle. Remarkably, PG core protein unknown-2, co-expressed with six different Calvin cycle genes (FBPase, FBPA, sedoheptulose-bisphosphatase, phosphoribulokinase, G3P-DH, phosphoglycerate kinase) as well as the catalytic subunit of glycine decarboxylase, critical for photorespiration. FBN7b appears to be involved in plastid/thylakoid biogenesis; among its top co-expressors are genes coding for ‘thylakoid formation 1’ (TF1), ‘vesicle-inducing protein in plastids’ (VIPP1), plastid division protein
Giant Chloroplasts 1 (GC1), and several genes of protein synthesis, assembly, folding and targeting. MEP enzymes IspF and HDS were also found in this module-4, as co-expressers of FBN-2, 4 and ABC1K9.

FBN8 was positioned between modules 2 and 4 and its co-expressors (all 20 were plastid-localized) were enriched in plastid biogenesis, photosynthesis and several proteins without known function. ABC1K5 connected to both modules 2 and 4 and its co-expressors (16 were plastid-localized) were enriched in NDH subunits, transporters and various unknowns.

It is important to note that only 1 of the genes (CF1-γ) encoding for known structural proteins of the linear electron transport chain and ATP-synthase (e.g. Photosystems I or II, the cytb6f complex or ATPsynthase) co-expressed with the PG core genes. However two of the three known genes that control state transitions (both STN7 and TAP38) and several structural components of cyclic (NDH and PGR components) or alternative electron flow (PTOX) were part of the co-expression network. This suggests that the PG function is tightly integrated with cyclic electron flow or the balance between PSI and PSII activity. We note that 4 lumenal OEC-23-like proteins with unknown function, as well as two unusual low abundant LHCI-5 and LHCII-7 proteins (At1g45474 and At1g76570) were found as co-expressors, suggesting that they have functions related to optimization of the light reactions under stress conditions. Indeed, LHCI-5 is a component of the PSI-NDH supercomplex and necessary for its formation and stability, particularly under times of stress (Peng and Shikanai, 2011). The co-expression profile was found to be very similar to NPQ4 (PsbS) and LIL3 involved in chlorophyll or tocopherol biosynthesis (Klimmek et al., 2006). LHCII-7 was found to be up-regulated in response to light stress (Alboresi et al., 2011) and blue or far-red light treatment (Sawchuk et al., 2008).

Three of the PG core proteins were not placed in the co-expression network because they were on the same probe (on the microarrays) as a close homologue. Indeed evaluation of DGAT4 on the same probe as a closely related non-plastid homologue (At3g26820), showed that the top 20 co-expressors were mostly involved with senescence but not in the plastid, and they did not connect well to the PG network. However, homologues FBN1a and b, both PG core proteins and together on a single probe spot, connected tightly in the network, with co-expression with core protein ABC1K3, and its co-expressors RbcX and FtsH8, and also co-expressing with RD20 and POT, both co-expressors of ABC1K7. Thus FBN1a/b is located in the network between module-1 and module-2.

**Experimental verification of the co-expression network**

The co-expression network suggested that a subset of the PG localized proteins (module 1) is involved in senescence responses. Therefore, we tested therefore for 5 genes (ACD1, PPH, DGAT3, ABC1K7, MCS) from module-1 if transcript accumulation is indeed up-regulated during natural senescence. As a control, we also tested two genes from module-4 (ABC1K9 and FBN4) which have no obvious senescence...
association in the network. To that end, Arabidopsis rosette leaves were harvested during bolting and flowering, during which leaves show increased visual signs of natural leaf senescence. Because PGs are found to accumulate fatty acid phytol esters, with the phytol generated by breakdown of chlorophyll (ref), the uncharacterized esterase identified in our PG core proteome is an excellent candidate enzyme for the esterification of free phytol at the PG. The flux into phytol esterification would be expected to be highest during senescence-induced chlorophyll degradation and we thus tested whether expression of the esterase, which is not represented on the 22K microarrays, is also senescence-induced. RT-PCR experiments were then carried out on three biological replicates (Fig. 7). Indeed, expression of the five genes from the senescence-associated module-1, and the esterase, but not the two genes from module-4, are induced by senescence, thus providing support for our co-expression network and our hypothesized functions for the esterase and MCS gene products.

**DISCUSSION**

*The core PG proteome* We identified and quantified proteins highly enriched in the thylakoid-associated PG as compared to other subplastid locations (the core PG proteome), and we associated key functions to the PG using genome-wide co-expression network analyses. PG localization for core proteins ABC1K1, 3 and M48 metalloprotease was confirmed by immunodetection. We determined that 13 proteins, previously assigned to the PG, were not particularly enriched in the PGs; instead they appeared primarily localized in the stroma or thylakoid membranes. Indeed, 8 of these were not found in the co-expression network, whereas the others had only a single connection to a PG core gene (RAP38 and FBPA-2 to Unknown2; RAP41 to ABCK1; FBPA-1 to FRed2; PrxQ to FBN7b). Importantly, we extended the known PG proteome with six new proteins of low abundance, including M48 protease. These new PG proteins were well-integrated in the co-expression network, providing further support for their PG localization and function.

*Chloroplast protein distribution and recruitment of proteins to PGs* It is not known how proteins are recruited to the PGs. This could occur by *de novo* synthesis and direct targeting to PGs. Alternatively, proteins could be recruited from other locations (e.g. stroma or thylakoid) to the PG through (ir)reversible protein modifications or through changes in the lipid/metabolite composition of the PGs. We can draw a parallel with recruitment of proteins to lipid rafts, which are membrane microdomains with a distinct lipid and protein composition (Simon-Plas et al., 2011). Such lipid rafts in plant plasma membranes have emerged as a regulatory mechanism governing physiological responses, in particular with a role as signal transduction platforms during stress. In the lipid rafts, proteins (typically low abundant) are brought...
physically together such that they form functional modules to carry out specific functions. Because the PG can rapidly change in size and number in response to (a)biotic stresses, it seems likely that both de novo synthesis, as well as recruitment of existing proteins could occur. In the latter case, proteins should show dual localization between PGs and other plastid compartments (as is the case for most PG core proteins), whereas the first group should be exclusively localized to PGs with changing cellular concentrations dependent on PG size and abundance.

Indeed, we determined that some of the PG core proteins showed a far stronger enrichment to the PGs (e.g. ABC1K9 and VTE1) than others (Table II). Dynamic changes in localization have been reported for some FBN proteins, but mechanism are unknown. The FBN1a homolog in tobacco and pepper distributed primarily in the stroma under optimal conditions, but redistributed to the thylakoid (including PGs) in response to light or drought stress (Rey et al., 2000; Simkin et al., 2007). Based on our experimental data and empirical relationship between physico-chemical parameters of FBN proteins and their distribution between PG and thylakoid preparations (Fig. 5), we suggest that FBN10 is a good example of a protein dual-localized between PGs and the stromal-exposed thylakoid surface. The direct membrane continuity between the thylakoid and PG, demonstrated elegantly by (Austin et al., 2006), could permit movement of proteins between these two membrane systems. How unique protein compositions are maintained between them has not been demonstrated conclusively, but it is likely that protein modifications such as (de)phosphorylation, prenylation or redox regulation may alter the distribution.

FBPA1,2 and 3 were previously identified in isolated PGs (Vidi et al., 2006; Ytterberg et al., 2006), and transient expression of GFP-tagged FBPA1 and 2 (At2g21330, At4g38970) in isolated protoplasts demonstrated association to PGs (Vidi et al., 2006). Our current quantitative, comparative analysis clearly demonstrated that these abundant FBPA mostly localized to the stroma, with only a small portion found in isolated PGs. We suggest that small amounts of these FBPA could be recruited to the PG (but their function is not understood) and that the concentration effect at the PG surface, compared to the diffuse signal from the much larger stroma volume, explains the apparent, more exclusive PG localization observed by GFP-tagging.

Our evidence that a significant number of genes involved in plastid isoprenoid/carotenoid accumulation are transcriptionally coordinated with genes encoding for PG proteins suggests that at least a subset of PG proteins are synthesized de novo concurrent with isoprenoid metabolism. Consistent with this notion, expression of Erwinia uredovora phytoene desaturase in potato tuber enhanced carotenoid metabolism, while simultaneously increasing transcript levels of the FBN homolog CDSP34 (Ducreux et al., 2005)
The PG co-expression network suggests several PG functions - An integrated model

PGs are believed to function in chloroplast development, stress responses, lipid metabolism and senescence including chlorophyll degradation - reviewed in (Brehelin et al., 2007; Brehelin and Kessler, 2008; Singh and McNellis, 2011). Because PGs function in so many diverse processes, and because most PG proteins have no known function, it has been difficult to obtain an integrated view of the role of the PG and assign PG core proteins to specific tasks or processes. To provide a framework for PG function, and to associate putative functions or processes to PG proteins with unknown functions, we determined the co-expression network based on the twenty most tightly associated co-expressors for each PG core gene. This resulted in four modules, each with a clear enrichment for specific functions, indicating that subsets of the PG core proteins work together to carry out specific roles.

Based on the core PG proteome information and modular structure of the co-expression network, as well as extensive published information, we created a summarizing model that integrates PG functions with chloroplast photosynthesis and metabolism, chloroplast responses to abiotic stress, and senescence (Fig. 8). A detailed description explaining the various pathways and processes is provided in the legend of the figure. In the remainder of this DISCUSSION, we will briefly summarize suggested PG functions and summarize our conclusions for FBNs and ABC1K proteins.

Function 1. The role of PG during leaf senescence

During senescence, the thylakoid membrane is dismantled, resulting in the free monogalactosyl- and digalactosyl glycerols and free fatty acids. These can be used as substrate by AOS for production of jasmonic acid (JA), or stored as TAG by DGAT3/4 (Fig. 8). The PG likely serves as a transient storage space for these glycerols and fatty acids. Concomitant with breakdown of the thylakoid bilayer, thylakoid protein complexes and associated pigments such as chlorophylls are degraded. The first steps in chlorophyll degradation are the removal of Mg$^{2+}$ from the porphyrin ring by an unknown protein, tentatively named ‘metal chelating substance’ (MCS), and cleavage of the phytol from the porphyrin ring by pheophytinase (PPH), one of the co-expressors in module-1 (Fig. 8) (Hortensteiner and Krautler, 2011). The toxic free phytol has been shown to become esterified to fatty acid and deposited in the PG (Gaude et al., 2007). We speculate that the PG localized esterase could be responsible for this esterification and its transcript levels did increase during natural senescence; experiments are now underway to test this hypothesis. AT5G17450, a co-expressor in senescence module-1 is a candidate for MCS because it has a metal binding domain (HMA). This protein must have a very low abundance as it has not been identified by proteomics.

Function 2. PG function in isoprenoid metabolism

The largest set of PG core proteins and their co-expressors were involved with plastid isoprenoid metabolism, in particular carotenoid metabolism,
including two PDS isoforms, ZDS, LYC-β, β-OHase, ZEP, and CCD1 (see Supplemental Fig. 5 for the complete isoprenoid pathway and the projected co-expressors). Particularly interesting was the finding that PG genes co-expressed with PDS and ZDS, as these enzymes transfer electrons from their carotenoid substrate to the plastoquinone pool, a major component of the PG metabolome (Bailey and Whyborn, 1963; Greenwood et al., 1963; Tevini and Steinmuller, 1985). The other isoprenoid genes, upstream of the carotenoid biosynthetic pathway, were both isoforms of solanesyl diphosphate synthase (SDS1,2), and MDS and HDS of the MEP pathway. Plastid-localized SDS2 is responsible for synthesizing the hydrophobic tail of plastoquinone (PQ-9). In particular ABC1K3, and to a lesser degree ABC1K1, was part of the network of these isoprenoid genes. The surprising linkage within the PG network of isoprenoid metabolism to plastid proteolysis can be easiest explained by the observation that these plastid proteases, in particular the thylakoid FtsH complex, are ‘household’ proteases, thus removing proteins that are unwanted or damaged, followed by release of chlorophylls and carotenoids. Indeed, it has been demonstrated that carotenoids and Chl a are continuously synthesized and degraded in photosynthesizing leaves and indicate distinct acclimatory responses of their turnover to changing irradiance (Beisel et al., 2010).

The abundant PG compounds α-tocopherol, PQ-9, and plastochromanol-8 (PC-8) are effective antioxidants in vivo (Havaux et al., 2005; Szymanska and Kruk, 2008, 2010). All three compounds are known to accumulate in response to light stress, most of which is likely accumulating in the PG (Vidi et al., 2006; Zbierzak et al., 2009). Within the PG, these antioxidants can re-reduce the sequestered oxidized lipids. As part of module-2, four enzymes with (putative) oxidoreductase activity are present in the PG (NDC1, aldo/keto reductase, and flavin reductase 1 and 2) which (may) act in regeneration of spent antioxidants in the PG by reducing carbonyl groups. Consistent with this possibility, vitamin K epoxide/naphthoquinone reductase was found to specifically reduce phylloquinone and menaquinone to their quinol forms in vitro (Furt et al., 2010). NDC1 has recently been demonstrated to display NAD(P)H reductase activity towards a plastoquinone (PQ-9) analog decyl-PQ (Eugeni-Piller et al., 2011). We speculate that the four PG oxidoreductases are active in re-reducing oxidized lipophilic compounds sequestered in the PG, thereby affecting the thylakoid redox state (see function 3 below).

Function 3. Contribution of PGs in optimization of photosynthesis, light acclimation and repair

Among the predominant genes co-expressing with members of the PG are the state transition kinase STN7 involved in balancing PSI and PSII activity, structural components of cyclic electron flow (NDH and PGR) and alternative oxidase (PTOX), nearly the complete plastidic thioredoxin regulatory system, as well as six enzymes of the Calvin cycle. This strongly suggests that the PG is intimately involved with optimization of the dark and light reactions, in particular via the redox state. Interestingly, recent work
suggested that the chloroplast redox status (or ROS) regulate cyclic electron flow, which in turn helps to achieve the correct ratio of ATP and redox energy required for the Calvin cycle and chloroplast metabolism in general (Livingston et al., 2010; Livingston et al., 2010). Interestingly, the PG co-expression network also included the chloroplast sensor kinase CSK (AT1G67840), co-expressing in module-3 with both ABC1K9 and FBN2, and many components of the chloroplast redox network. CSK was recently shown to be involved in redox-coupled transcriptional regulation of chloroplast genes (Puthiyaveetil et al., 2008). Furthermore, zeaxanthin epoxidase (ZEP) involved in the reversible conversion of zeaxanthin to violaxanthin (via antheroxanthin) within the xanthophyll cycle was centrally located in the gene expression network with connections to ABC1K3, ABC1K6 and CCD4. We speculate that ZEP activity may be regulated by one of these ABC1K proteins. Consistently, it was suggested that ZEP activity is controlled by a direct, as yet unidentified, modification, which does not involve the state transition kinases (Reinhold et al., 2008). There is some indirect evidence that phosphorylation of ZEP significantly impedes its in vivo activity (Xu et al., 1999). Collectively, it appears that the PG plays a key role in short term regulation and balancing of photosynthetic activities. Perhaps surprising, none of the well-known enzymes involved in detoxification of soluble reactive oxygen species (superoxide and hydrogen peroxide) such as superoxide dismutases, or thylakoid and stromal APX, were found in the PG co-expression network.

**FBNs and ABC1K proteins – distribution, functions and targets** The seven PG-localized FBNs (1a, 1b, 2, 4, 7a, 7b, 8) and the six ABC1K proteins constituted more than 70% of the PG protein mass. These six ABC1K proteins are expected to act as enzyme regulators, possibly via phosphorylation (Do et al., 2001), and the notion of a regulatory function is strengthened by their position as hubs in the PG network (Fig. 8). Their PG localization suggests that they are regulating enzymes that locate, at least transiently, to the PG; the co-expression network provides potential target genes that should now be experimentally tested. We note that ABC1K homologue AT5G64940, that we annotated as ABC1K8, was never found in the PG nor in the co-expression network. ABC1K8 was identified as a chloroplast inner-envelope protein and reduced expression resulted in increased sensitivity toward oxidative stress and high light (Jasinski et al., 2008). The FBN proteins are suggested to primarily function as structural proteins, likely determining PG size, some involved in adaptation to environmental stress and others possibly influencing metabolite and protein content. Information about their possible functions is summarized in a recent review (Singh and McNellis, 2011). The seven PG-localized FBNs were distributed across the co-expression network thus providing further suggestions for functions.

**MATERIALS AND METHODS**
Preparation of PG and thylakoid material – The PG isolation method was adapted from Ytterberg, et al (Ytterberg et al., 2006). For each PG preparation, two flats (~150 individuals) of *A. thaliana* (Col-0) were grown on soil for 2.5 weeks, under 120 µmol photons m\(^{-2}\) s\(^{-1}\) with a 16 hour photoperiod. Plants were then transferred to 520 µmol photons.m\(^{-2}\).s\(^{-1}\) conditions during the dark period. In the morning of the 6th day leaf tissue was harvested and homogenized in grinding buffer (50 mM Hepes-KOH pH 8.0, 5 mM MgCl\(_2\), 100 mM sorbitol, 5 mM ascorbic acid, 5 mM reduced cysteine, 0.05 % (w/v) BSA). Homogenate was filtered through four layers of 20-µm miracloth and thylakoid membranes were pelleted by centrifugation for 6 minutes at 1800xg. Thylakoid pellets were washed once in 4 volumes of grinding buffer and resuspended in Medium R (50 mM Hepes-KOH pH 8.0, 5 mM MgCl\(_2\), cocktail of protease inhibitors) containing 0.2 M sucrose. An aliquot of resuspended thylakoid material was stored at -80° C to be used as the pre-sonicated thylakoid fraction. The remainder was sonicated 4x 5s at output power 23 Watts (Fischer Scientific, sonic dismembrator Model 100), returning the samples to ice between each sonication event. Sonicated samples were centrifuged for 30 minutes at 150,000 xg and PGs released from the thylakoid floated to the surface of the solution. PGs were removed and combined with Medium R with 0.7M sucrose to achieve a sucrose concentration of 0.5M, which was then overlayed with Medium R with 0.2M sucrose and Medium R with no sucrose. The gradient was centrifuged 90 minutes, 150,000 xg. The resulting floating pad of PGs was removed, flash frozen in liquid N\(_2\), and stored at -80° C.

Antiserum generation - Nucleotide sequences encoding the soluble part of the M48 protein (amino acids 72-325) and the C-termini of ABC1K3 (556-711) and ABC1K1 (578-582) were amplified by PCR. The resulting DNA fragments were ligated into restriction sites of the pET21a expression vector, coding for a C-terminal His affinity tag. The vector was transformed into BL21 *E. coli* cells, and over-expressed protein was harvested from liquid culture after incubation in 1mM IPTG for 3 h at 37° C. Proteins were solubilized in 200mM NaCl, 50mM Tris, and 8M Urea at pH 8 and purified on a nickel-nitrotriacetic acid agarose resin matrix, and polyclonal antibodies were raised in rabbits by injecting purified antigen.

Immunoblotting – Protein concentrations were estimated by the BCA method (Smith et al., 1985) using a BCA kit (Pierce). Protein samples were solubilized in 1x Laemmli buffer (125 mM Tris-HCl pH 6.8, 2% SDS, 5% β-mercaptoethanol, 10% glycerol), heated for 10 mins at 75° C, and separated in a SDS-PAGE gel (6% acrylamide stacking, 12% separation). Proteins were blotted to nitrocellulose, probed with purified anti-PeptidaseM48, anti-ABC1K3, anti-ABC1K1 or anti-VTE1 serum (a gift of Dr. Dean DellaPenna), and visualized by the horseradish peroxidase-based enhanced chemiluminescence system. Densitometric analysis of relevant spots was performed using the ImageJ software program (http://rsbweb.nih.gov/ij/).
TEM – Leaf tissue from 3 individuals of each genotype at each time point was harvested 1 hour after beginning of the photoperiod. Leaf margins and midribs were excluded and the remaining leaf tissue was divided into 1x2 mm sections with a fresh razor blade. Sections were fixed in 2% glutaraldehyde, 2% paraformaldehyde, 0.1% tannic acid, 70 mM PIPES buffer pH 6.8 for 2 hours and then washed 3 times in 70 mM PIPES buffer pH 6.8. Tissues were fixed in 1% osmium tetroxide (OsO₄), 70 mM PIPES pH 6.8 for 2 hours and washed 3 times in 70 mM PIPES pH 6.8. Tissues were then stained in 2% uranyl acetate for 1 hour and washed twice in ultrapure H₂O. Fixed and stained tissues were carried through an acetone series of increasing concentrations. Dehydrated tissue was then embedded with Spurr’s resin (Electron Microscopy Sciences; Hatfield, PA) in increasing concentrations of resin in acetone, according to manufacturer’s instructions. Fully embedded tissue was cured in resin blocks at 60°C overnight. Cured resin blocks were sectioned and imaged at Electron Microscopy Services; Colorado Springs, Colorado.

SEM – 2-3 µl of purified PG sample were spotted onto a silica wafer. A 3 µl drop of 2% OsO₄ in 70 mM potassium phosphate pH 7.2 was added to the 3 µl drop of the PG sample on the silica wafer. The buffered OsO₄ was allowed to remain in contact with the PGs for one hour at 4°C. After one hour the wafers were floated on a droplet of 70 mM potassium phosphate buffer at pH 7.2 for 10 minutes. This was done three times at 4°C. The wafers were then floated on drops of 2% gluteraldehyde in 70 mM potassium phosphate pH 7.2 for one hour at 4°C. After one hour the wafers were floated on drops of 70 mM potassium phosphate pH 7.2 for 10 minutes at 4°C. The wafers were then floated on drops of distilled water for 10 minutes at 4°C. The wafers were dehydrated by floating on first 25%, then 50%, then 75%, then 95% and finally drops of 100% ethanol for approximately 10 minutes each at 4°C. The wafers were then critical point dried in 100% ethanol (Bal Tec - Leica Microsystems, Inc., Bannockburn, IL). The wafers were then mounted on specimen supports and sputter coated with gold/palladium (Denton Vacuum, LLC, Moorestown, NJ). The wafers were viewed at 3 kV in a Hitachi S4500 scanning electron microscope (Hitachi High-Technologies Corp., Tokyo, Japan).

In-solution and in-gel digestion of isolated PGs - For in-solution digestion, isolated PGs were precipitated in 10% tri-chloro acetic acid (TCA) overnight at 4°C. Precipitated proteins were pelleted by centrifugation and washed once with 100% acetone and once with 80% acetone, 10% methanol, 0.1% acetic acid, by incubating at -20°C for 1.5 hours each. Washed pellets were resuspended in dimethyl sulfoxide and quantified by the BCA method (Smith et al., 1985) using a BCA kit (Pierce). 5 µg of protein was digested with modified trypsin (Promega), 40:1 (protein: trypsin). Salts and detergents were removed by C18 Ziptip (Millipore) and dried down in a speed-vac. Digested and washed samples were resuspended in 15 µl 2% formic acid immediately prior to loading on the LC-MS/MS instrument. For gel-based separation and in-gel digestion, PG samples were lyophilized and solubilized in a modified Laemmli solubilization buffer (125 mM Tris-HCl pH 6.8, 6% SDS, 10% β-mercaptoethanol, 20%
glycerol). Samples were shaken gently at 30°C for 15 minutes to ensure complete solubilization and subsequently heated at 80°C for 10 min. Samples were centrifuged to remove insoluble material and proteins were separated by SDS PAGE (6% acrylamide stacking, 12% separation). Each gel lane was cut in 5 slices and proteins were digested with trypsin, as described in (Friso et al., 2011).

Proteome analysis of total leaf extracts. Wild-type plants (Col-0) were grown on soil for 30 days under a short-day cycle (10h/14h of light/dark) at 120 μmol photons.m⁻².s⁻¹. The complete leaf rosettes were then harvested and proteins were immediately quantitatively extracted in presence of SDS (in triplicate) as described in detail in (Friso et al., 2011). Alternatively, plants were transferred were grown on soil for 2.5 weeks under similar conditions as above, but transferred to 520 μmol photons m⁻² s⁻¹ conditions. In the morning of the 6th day leaf tissue was harvested and extracted as above (in triplicate).

Proteome analysis by nanoLC-LTQ-Orbitrap and data processing – Peptides prepared from in-gel digestion and in-solution digestion were analyzed by data-dependent tandem mass spectrometry (MS/MS) using on-line LC-LTQ-Orbitrap (Thermo Electron) with dynamic exclusion, similar as described in (Zybailov et al., 2008). Peak lists (.mgf format) were generated using DTA supercharge (v1.19) software (http://msquant.sourceforge.net/) and searched with Mascot v2.2 (Matrix Science) against a combined database containing the Arabidopsis genome with protein-coding gene models and 187 sequences for known contaminants (e.g. keratin, trypsin) (total 33,013 entries) and concatenated with a decoy database where all the sequences were randomized; in total this database contained 66,026 protein sequences. Off-line calibration for all precursors ions was done as described in (Olinares et al., 2010). Each of the peak lists were searched using Mascot v2.2 (maximum p-value of 0.01) for full tryptic peptides using a precursor ion tolerance set at ±6 ppm, fixed cysteine carbamido-methylation and variable methionine oxidation, protein N-terminal acetylation, asparagine/glutamine (N/Q) deamidation and maximally one missed cleavage allowed. The maximum fragment ion tolerance (MS/MS) was 0.8 Da. For semi-tryptic peptides the search was performed with a precursor ion tolerance set at ±3 ppm, fixed cysteine carbamido-methylation and variable methionine oxidation, N-terminal acetylation, glutamine deamidation and maximally one missed cleavage allowed. Minimal ion score threshold was chosen such that a peptide false discovery rate (FDR) below 1% was achieved. Using an in-house written filter, the search results were further filtered as follows: For identification with two or more peptides, the minimum ion score threshold was set to 30. For protein identification based on a single peptide, the minimum ion score threshold was set to 33, and the mass accuracy of the precursor ion was required to be within ±3 ppm. The peptide false discovery rate (FDR) was calculated as: 2 × (decoy hits) / (target + decoy hits) and was below 1%. The FDR of proteins identified with two or more peptides was zero. Peptides with less than seven amino acids were discarded. All mass spectral data (the mgf files reformatted as PRIDE XML
files) are available via the Proteomics Identifications database (PRIDE) at http://www.ebi.ac.uk/pride/.
Accession numbers: 18969-18988; Reviewer Account- Username: review54381; Password: D#nejxk#

Several Arabidopsis genes have more than one gene model, and in such cases the protein form with the highest number of matched spectra was selected; if two gene models had the same number of matched spectra, the model with the lower digit was selected. For quantification, each protein accession was scored for total spectral counts (SPC), unique SPC (uniquely matching to an accession) and adjusted SPC [Friso, 2011 #13929]. The latter assigns shared peptides to accessions in proportion to their relative abundance using unique spectral counts for each accession as a basis. The normalized adjSPC (NadjSPC) for each protein was calculated through division of adjSPC by the sum of all adjSPC values for the proteins from that gel lane. NadjSPC provides a relative protein abundance measure by mass, whereas NSAF estimates relative protein concentration within a particular sample.

**Genome-wide co-expression calculations and network visualization** – The Pearson correlation coefficients of all pairwise combinations between PG (bait) genes and all single-gene probes of the *A.thaliana* 22K Affymetrix microarray (Affymetrix, Inc.) were calculated using three different software programs: i) the MetaOmGraph software program (http://metnetdb.org) (Wurtele et al., 2007), ii) the Botany Array Resource (BAR) expression angler (http://142.150.214.117/welcome.htm) (Toufighi et al., 2005), and the Arabidopsis Co-expression data mining Tool (ACT) website (http://www.arabidopsis.leeds.ac.uk/act/index.php) (Manfield et al., 2006). MetaOmGraph analysis used the publicly available “Affy.ath1.data1 project” containing normalized, averaged Arabidopsis experimental datasets obtained from NASCArrays (http://affymetrix.arabidopsis.info/) and PlexDB (http://plexdb.org) from 71 experiments and 424 microarray chips from diverse environmental and genotypic conditions and tissue types and developmental stages. Correlations were calculated using the Pearson correlation algorithm. Visualization of the MetaOmGraph-derived network was performed in Cytoscape v2.8.0 (http://cytoscape.org/) (Shannon et al., 2003), applying the force-directed layout algorithm. Co-expression analysis using the BAR expression angler was performed for each PG gene by searching in the ‘NASCArrays 392’ dataset available at the website. Analysis at the ACT website was performed for each PG gene by using the “Co-expression analysis over available array experiments” option.

**Analysis of transcript accumulation during natural senescence.** Wild-type Arabidopsis col-0 was grown on soil. Leaf tissue was selected from five time points during the course of natural leaf senescence: 1 = leaf rosette from plants beginning to bolt; 2 = leaf rosette from plant beginning to flower; 3 = senescing leaf ~10% chlorotic, 4 = senescing leaf ~50% chlorotic; 5 = senescing leaf ~50% chlorotic, one week later in senescence. Total RNA was extracted from leaf tissue using the RNeasy plant miniprep kit (Qiagen) according the manufacturer’s instructions. 700 ng of total RNA was used for synthesis of cDNA.
using oligo dT(20) primer and the Superscript III cDNA synthesis kit (Invitrogen) according to the manufacturer’s instructions. cDNA samples were diluted to equal concentration by normalizing according to amplification of the actin2 gene using 20 cycles. Each gene was then amplified for 25 cycles using an equal volume of template and an appropriate gene specific primer pair. Signal intensity was quantified using the Alpha Imager 2200 v5.5 software package. The forward and reverse primers are: for PAO/ACD1 5'- GAT GCG AAA CTA CCA ATC GTC G -3' and 5'- CAT CAG AAG GAA CAC CAG CCG -3'; for PPH 5'- CAA TCA TGC TTG CTC CTG GTG -3' and 5'- CTA CCA ATC CTG GAC TCC TCC -3'; for DGAT3 5'- GCC AGA GGA GCT TCA TTT TAC T -3' and 5'- GGG TAT GCC CAT TGT CCT T -3'; for ABC1K7 5'- ATC CGC ACC CAG GAA ACC TT -3' and 5'- ACA GAT CCT GCC ATA GAA AGG AGG -3'; for MCS 5'- GAA ATC GGT GGA GGT GAA CC -3' and 5'- GGT TGG TTG GCT CAC ATG AT -3'; for ESTERASE 5'- GCT AAC TGC TGT TAC GTG CTC CTG GTG GCC -3' and 5'- CAA ACT CCG AAT GTT CTG GCC -3'; for ABC1K9 5'- GCA GCT TGG TCT ACT GTC TC -3' and 5'- CAC ATT AAG CGC GGT AAT AAC G -3'; for FBN4 5'- TTC TTT CCG ACC ACC GTT CTG CAC ATG AT -3' and 5'- CAA ACT TGT GTG CCA ATG TCG C -3'; for ACTIN2 5'- GCC AGA AAG AGG GCT GGA ACA AGA CT -3' and 5'- GCA ACT GGG ATG ATA TGG AAA AGA -3'

Calculation of protein physicochemical parameters – Parameters were calculated by the ProtParam tool (Gasteiger et al., 2005) available through the ExPasy website (http://expasy.org/tools/).

ACKNOWLEDGEMENTS

We thank Carole Daugherty for her critical assistance with the SEM analysis of PG preparations, Richard Medville of Electron Microscopy Sciences (Colorado Springs, Colorado) for collection of TEM micrographs and Dr. Dean DellaPenna for his generous donation of anti-VTE1 serum. PKL was in part funded by an NIH Chemical-Biological Interface (CBI) training grant NIH (# 5T32GM008500). Part of this work was carried out using the resources of the Computational Biology Service Unit of Cornell University which is partially funded by the Microsoft Corporation.

Accession Numbers with protein names

AT4G04020 (FBN1a), AT3G23400 (FBN4), AT4G22240 (FBN1b), AT2G35490 (FBN2), AT5G05200 (ABC1K9), AT1G79600 (ABC1K3), AT4G31390 (ABC1K1), AT3G58010 (FBN7a), AT4G19170 (CCD4), AT4G32770 (VTE1), AT5G08740 (NDC1), AT1G54570 (DGAT3), AT2G42130 (FBN7b), AT1G32220 (FR-like1), AT4G13200 (Unknown1), AT2G46910 (FBN8), AT3G10130 (SOUL-like), AT2G41040 (UbiE-like1), AT1G71810 (ABC1K5), AT1G06690 (AKR-like), AT2G34460 (FR-like2), AT1G78140 (UbiE-like2), AT3G26840 (DGAT4), AT3G24190 (ABC1K6), AT4G39730 (PLAT/LH2-1), AT3G43540 (Unk2), AT2G22170 (PLAT/LH2-2),
AT3G07700 (ABC1K7), AT1G73750 (UnkSAG), AT3G27110 (M48 metalloprotease), AT5G41120 (esterase1).
### Table I. Subplastid localization of fibrilllin proteins and their variants

<table>
<thead>
<tr>
<th>accession #</th>
<th>name</th>
<th>PG/stroma (a)</th>
<th>PG/thylakoid (a)</th>
<th>PG/leaf (a)</th>
<th>PG core</th>
<th>thylakoid / stroma (a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT4G04020.1</td>
<td>FBN1a</td>
<td>176</td>
<td>34</td>
<td>170</td>
<td>yes</td>
<td>5</td>
</tr>
<tr>
<td>AT4G22240.1</td>
<td>FBN1b</td>
<td>713</td>
<td>40</td>
<td>570</td>
<td>yes</td>
<td>18</td>
</tr>
<tr>
<td>AT2G35490.1</td>
<td>FBN2</td>
<td>1188</td>
<td>59</td>
<td>151</td>
<td>yes</td>
<td>20</td>
</tr>
<tr>
<td>AT3G26070.1</td>
<td>FBN3a</td>
<td>not in stroma</td>
<td>0</td>
<td>1</td>
<td>no</td>
<td>only thylakoid</td>
</tr>
<tr>
<td>AT3G26080.1</td>
<td>FBN3b</td>
<td>not in stroma</td>
<td>not in PG</td>
<td>only thylakoid</td>
<td>no</td>
<td>only thylakoid</td>
</tr>
<tr>
<td>AT3G23400.1</td>
<td>FBN4</td>
<td>121</td>
<td>32</td>
<td>79</td>
<td>yes</td>
<td>4</td>
</tr>
<tr>
<td>AT5G09820.1</td>
<td>FBN5</td>
<td>not in PG</td>
<td>—</td>
<td>only stroma</td>
<td>no</td>
<td>only stroma</td>
</tr>
<tr>
<td>AT5G19940.1</td>
<td>FBN6</td>
<td>not in PG</td>
<td>not in PG</td>
<td>not in PG</td>
<td>no</td>
<td>15</td>
</tr>
<tr>
<td>AT3G58010.1</td>
<td>FBN7a</td>
<td>146</td>
<td>15</td>
<td>342</td>
<td>yes</td>
<td>10</td>
</tr>
<tr>
<td>AT2G42130.4</td>
<td>FBN7b</td>
<td>23</td>
<td>11</td>
<td>55</td>
<td>yes</td>
<td>2</td>
</tr>
<tr>
<td>AT2G46910.1</td>
<td>FBN8</td>
<td>434</td>
<td>29</td>
<td>388</td>
<td>yes</td>
<td>15</td>
</tr>
<tr>
<td>AT4G00030.1</td>
<td>FBN9</td>
<td>—</td>
<td>—</td>
<td>only leaf</td>
<td>no</td>
<td>—</td>
</tr>
<tr>
<td>AT1G51110.1</td>
<td>FBN10</td>
<td>not in stroma</td>
<td>2</td>
<td>5</td>
<td>no</td>
<td>only thylakoid</td>
</tr>
<tr>
<td>AT1G18060.1</td>
<td>FBN-like</td>
<td>not in PG</td>
<td>not in PG</td>
<td>not in PG</td>
<td>no</td>
<td>23</td>
</tr>
<tr>
<td>variant (b)</td>
<td>FBN7a (1-290)</td>
<td>found in stroma based on GFP visualization</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>variant (b)</td>
<td>FBN7a (1-133)</td>
<td>found in thylakoid based on GFP visualization</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) Abundance ratio based on NadjSPC in PG and other chloroplast compartments
(b) from Vidi et al., 2007
Table II. The plastoglobule core proteome determined by quantitative comparative proteomics

<table>
<thead>
<tr>
<th>Accession</th>
<th>Protein name</th>
<th>NadjSPC (a)</th>
<th>CV (%) (b)</th>
<th>% mass PG core (%) (c)</th>
<th>PG/thylakoid (d)</th>
<th>PG/stroma (d)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT4G04020</td>
<td>Fibrillin 1a (FBN1a)</td>
<td>0.100</td>
<td>15</td>
<td>16.1</td>
<td>34</td>
<td>176</td>
<td>x x x</td>
</tr>
<tr>
<td>AT3G23400</td>
<td>Fibrillin 4 (FBN4)</td>
<td>0.074</td>
<td>30</td>
<td>11.9</td>
<td>32</td>
<td>121</td>
<td>x x x</td>
</tr>
<tr>
<td>AT4G22240</td>
<td>Fibrillin 1b (FBN1b)</td>
<td>0.059</td>
<td>12</td>
<td>9.6</td>
<td>40</td>
<td>713</td>
<td>x x x</td>
</tr>
<tr>
<td>AT2G35490</td>
<td>Fibrillin 2 (FBN2)</td>
<td>0.044</td>
<td>4</td>
<td>7.1</td>
<td>59</td>
<td>1188</td>
<td>x x x</td>
</tr>
<tr>
<td>AT5G05200</td>
<td>ABC1K9</td>
<td>0.032</td>
<td>8</td>
<td>5.2</td>
<td>440</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT4G31390</td>
<td>ABC1K1</td>
<td>0.028</td>
<td>8</td>
<td>4.5</td>
<td>16</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT1G79600</td>
<td>ABC1K3</td>
<td>0.027</td>
<td>23</td>
<td>4.3</td>
<td>11</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT3G58010</td>
<td>ABC1K7a (FBN7a)</td>
<td>0.022</td>
<td>44</td>
<td>3.5</td>
<td>15</td>
<td>146</td>
<td>x x x</td>
</tr>
<tr>
<td>AT4G19170</td>
<td>Carotenoid dioxygenase 4 (CCD4)</td>
<td>0.021</td>
<td>25</td>
<td>3.3</td>
<td>18</td>
<td>42</td>
<td>x x x</td>
</tr>
<tr>
<td>AT4G32770</td>
<td>Tocopherol cyclase (VTE1)</td>
<td>0.016</td>
<td>5</td>
<td>2.6</td>
<td>131</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT1G54570</td>
<td>Diacylglycerol Acyltransferase 3 (DGAT-3)</td>
<td>0.016</td>
<td>12</td>
<td>2.6</td>
<td>31</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT5G08740</td>
<td>NAD(P)H dehydrogenase C1 (NDC1)</td>
<td>0.015</td>
<td>8</td>
<td>2.5</td>
<td>15</td>
<td>322</td>
<td>x x x</td>
</tr>
<tr>
<td>AT2G42130</td>
<td>Fibrillin 7b (FBN7b)</td>
<td>0.013</td>
<td>45</td>
<td>2.1</td>
<td>19</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT1G32220</td>
<td>Flavin-reductase-related 1</td>
<td>0.013</td>
<td>23</td>
<td>2.1</td>
<td>11</td>
<td>23</td>
<td>x x x</td>
</tr>
<tr>
<td>AT4G13200</td>
<td>Unknown 1</td>
<td>0.012</td>
<td>17</td>
<td>1.9</td>
<td>22</td>
<td>102</td>
<td>x x x</td>
</tr>
<tr>
<td>AT3G10100</td>
<td>SOUL domain protein</td>
<td>0.011</td>
<td>10</td>
<td>1.8</td>
<td>11</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT2G46910</td>
<td>Fibrillin 8 (FBN8)</td>
<td>0.011</td>
<td>12</td>
<td>1.8</td>
<td>61</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT1G71610</td>
<td>ABC1K5</td>
<td>0.011</td>
<td>12</td>
<td>1.7</td>
<td>29</td>
<td>434</td>
<td>x x x</td>
</tr>
<tr>
<td>AT1G78140</td>
<td>UbE methyltransferase-related 1</td>
<td>0.009</td>
<td>43</td>
<td>1.5</td>
<td>-</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT1G05690</td>
<td>Aldo/keto reductase</td>
<td>0.009</td>
<td>35</td>
<td>1.5</td>
<td>48</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT2G34460</td>
<td>Flavin-reductase related 2</td>
<td>0.009</td>
<td>26</td>
<td>1.5</td>
<td>13</td>
<td>765</td>
<td>x x x</td>
</tr>
<tr>
<td>AT2G41040</td>
<td>UbE methyltransferase-related 2</td>
<td>0.009</td>
<td>17</td>
<td>1.5</td>
<td>6</td>
<td>75</td>
<td>x x x</td>
</tr>
<tr>
<td>AT3G26840</td>
<td>Diacylglycerol Acyltransferase 3 (DGAT 3)</td>
<td>0.009</td>
<td>14</td>
<td>1.4</td>
<td>72</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT3G24190</td>
<td>ABC1K6</td>
<td>0.016</td>
<td>14</td>
<td>2.6</td>
<td>14</td>
<td>80</td>
<td>-</td>
</tr>
<tr>
<td>AT4G39730</td>
<td>PLAT/LH2-1</td>
<td>0.010</td>
<td>29</td>
<td>1.6</td>
<td>17</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT3G43540</td>
<td>Unknown 2 (DUF1350)</td>
<td>0.008</td>
<td>42</td>
<td>1.3</td>
<td>-</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT3G07700</td>
<td>ABC1K7</td>
<td>0.005</td>
<td>27</td>
<td>0.8</td>
<td>37</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT1G73750</td>
<td>Unknown SAG</td>
<td>0.002</td>
<td>80</td>
<td>0.4</td>
<td>-</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT3G27110</td>
<td>M48 protease</td>
<td>0.002</td>
<td>42</td>
<td>0.3</td>
<td>17</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT5G41120</td>
<td>Esterase 1</td>
<td>0.002</td>
<td>30</td>
<td>0.3</td>
<td>-</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT5G42650</td>
<td>Allene oxide synthase (AOS)</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT1G09340</td>
<td>Rap38/CSP41B</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT1G26090</td>
<td>Anion-transporting ATPase</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT1G28150</td>
<td>unknown</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT4G01150</td>
<td>unknown</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT3G63140</td>
<td>Rap41/CSP41A</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT5G01730</td>
<td>WAVE3</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT2G21330</td>
<td>FBPA-1</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT4G38970</td>
<td>FBPA-2</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT2G01140</td>
<td>FBPA-3</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT3G28060</td>
<td>Peroxiredoxin Q (Prx Q)</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT1G52590</td>
<td>unknown (DUF39)</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x x x</td>
</tr>
<tr>
<td>AT3G26070</td>
<td>Fibrillin (FBN3a)</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x x x</td>
</tr>
</tbody>
</table>

(a) Abundance of each PG core protein
(b) Coefficient of Variation (CV) of the average NadjSPC across the three biological replicates
(c) Contribution of each protein to protein mass of the PG core proteome as percentage of total core proteome
(d) Abundance ratio based on NadjSPC in PG and other chloroplast compartments
(e) Vidi, et al.(2006)
(g) current analysis
Table III. Functional group enrichment of the PG network produced by MetaOmGraph.

<table>
<thead>
<tr>
<th>Functional group</th>
<th>Bin size</th>
<th>whole network</th>
<th>module 1</th>
<th>module 2</th>
<th>module 3</th>
<th>module 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein degradation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein degradation - plastid</td>
<td>42</td>
<td><strong>3.3</strong></td>
<td><strong>1.2</strong></td>
<td><strong>6.5</strong></td>
<td><strong>3.3</strong></td>
<td><strong>3.6</strong></td>
</tr>
<tr>
<td>Protein degradation - not plastid</td>
<td>1313</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Core plastoglobule (PG)</td>
<td>25</td>
<td><strong>3.4</strong></td>
<td><strong>4.0</strong></td>
<td><strong>4.5</strong></td>
<td><strong>4.8</strong></td>
<td><strong>2.0</strong></td>
</tr>
<tr>
<td>Light reaction</td>
<td>139</td>
<td>0.8</td>
<td>0.4</td>
<td>0.5</td>
<td>1.4</td>
<td>1.1</td>
</tr>
<tr>
<td>light stress-Lil-Sep/Ohp</td>
<td>8</td>
<td>0.5</td>
<td>0.0</td>
<td><strong>1.6</strong></td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>NDH dependent &amp; independent, Immutans, PIFI</td>
<td>29</td>
<td><strong>1.5</strong></td>
<td><strong>1.7</strong></td>
<td>0.4</td>
<td><strong>2.8</strong></td>
<td>0.0</td>
</tr>
<tr>
<td>PSI, PSII, ATPsynt, cytb6f, FNR, electron carriers (PC, Fd)</td>
<td>94</td>
<td>0.5</td>
<td>0.0</td>
<td>0.3</td>
<td>0.9</td>
<td><strong>1.6</strong></td>
</tr>
<tr>
<td>thylakoid-bound regulators, including kinases and phosphatases</td>
<td>5</td>
<td><strong>3.2</strong></td>
<td>0.0</td>
<td><strong>5.0</strong></td>
<td><strong>8.0</strong></td>
<td>0.0</td>
</tr>
<tr>
<td>Isoprenoid metabolism</td>
<td>124</td>
<td>0.7</td>
<td>0.2</td>
<td><strong>1.6</strong></td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>Isoprenoid metabolism - plastid</td>
<td>59</td>
<td><strong>1.1</strong></td>
<td>0.0</td>
<td><strong>2.5</strong></td>
<td><strong>1.0</strong></td>
<td><strong>0.8</strong></td>
</tr>
<tr>
<td>Isoprenoid metabolism - not plastid</td>
<td>65</td>
<td>0.3</td>
<td>0.4</td>
<td>0.8</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Redox</td>
<td>187</td>
<td>0.7</td>
<td>0.4</td>
<td>0.3</td>
<td><strong>2.5</strong></td>
<td>0.3</td>
</tr>
<tr>
<td>Tetrapyrrole metabolism</td>
<td>50</td>
<td>0.5</td>
<td></td>
<td><strong>2.0</strong></td>
<td>0.0</td>
<td>0.4</td>
</tr>
<tr>
<td>Stress</td>
<td>690</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>1274</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Protein - other</td>
<td>1535</td>
<td>0.1</td>
<td>0.0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.4</td>
</tr>
<tr>
<td>Protein - other - plastid</td>
<td>211</td>
<td>0.7</td>
<td>0.0</td>
<td>0.8</td>
<td>0.4</td>
<td><strong>2.8</strong></td>
</tr>
<tr>
<td>Protein - other - not plastid</td>
<td>1326</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Not assigned</td>
<td>7707</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>Development</td>
<td>521</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>CHO metabolism (c)</td>
<td>441</td>
<td>0.3</td>
<td>0.0</td>
<td>0.2</td>
<td>0.2</td>
<td>0.9</td>
</tr>
<tr>
<td>Transport</td>
<td>944</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>Other (d)</td>
<td>2128</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Signalling</td>
<td>1048</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Lipid metabolism</td>
<td>331</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>DNA/RNA</td>
<td>2659</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

(a) number of genes (represented by a single probe spot on the 22K Affymetrix microarray chip) in each bin

(b) Number of edges per bin, normalized for bin size, and normalized for number of PG core genes per module. Values in bold are enriched functions

(c) includes major and minor carbohydrate metabolism, gluconeogenesis, glycolysis, TCA cycle, C1 metabolism, fermentation, oxidative pentose phosphate pathway, Calvin cycle, and all other dark reactions

(d) includes cofactor and vitamin metabolism, metal handling, xenobiotics, amino acid metabolism, nucleotide metabolism, cytoskeleton, mitochondrial electron transport, cell wall, cell, cell division, cell cycle, nitrogen metabolism, photosynthesis, polyamine metabolism, sulfur assimilation, secondary metabolism (excluding isoprenoids/tetrapyrrole), and hormone metabolism
CITED REFERENCES


Mann M, Kelleher NL (2008) Special Feature: Precision proteomics: The case for high resolution and high mass accuracy. Proc Natl Acad Sci U S A


FIGURE LEGENDS

Figure 1. PG extraction and purification efficiency. A, Immunoblot of a thylakoid sample prior to sonication and purified PG fractions. Proteins were separated by 1D SDS-PAGE and probed with antibody against PG marker protein, VTE1. 1x corresponds to 2.2 μg protein. B, Immunoblot of thylakoid samples prior to sonication and after sonication for removal of PGs. A Ponceau stain of protein is included as a loading control. 1x corresponds to 10 μg protein. C, Histogram illustrating the distribution of PG diameters from transmission electron micrographs (TEM) of chloroplasts in mature leaf tissue and scanning electron micrographs (SEM) of purified PG preparations, n ≥ 263 measurements. D, Transmission electron micrograph (TEM) of leaf chloroplast of A. thaliana (Col-0), representative of the time point PG preparations were made. An image of a typical light stress Arabidopsis rosette plant is shown. PGs in the TEM, marked by white arrows, appear as gray circles with black periphery or, less frequently as solid black circles. E, SEM of A. thaliana (Col-0) purified PGs, demonstrating efficient isolation of PGs with varying diameters. Scale bars in (D) and (E) = 500 nm.

Figure 2. Identification of the PG proteome by in-gel and in-solution methods. A, Three independent preparations of PGs were made from leaf rosettes of A. thaliana plants grown at 120 μmol photons.m⁻².s⁻¹ and 16 light/8 dark cycle, that were shifted for 5 days to moderate light intensities (500 μmol photons.m⁻².s⁻¹). Thylakoids were isolated from total leaf tissue and sonicated to release PGs. Aliquots of each PG preparation were then separated by SDS-PAGE and in-gel digested or delipidized, in-solution digested, and zip-tipped. In-solution and in-gel digested peptides were analyzed with a nanoLC-LTQ-Orbitrap mass
spectrometer. 234 unique proteins were identified in total, 129 of which were identified by both experimental methods. B, Comparison of NadjSPC between each of the 234 proteins between the in-gel and in-solution methods demonstrates consistent quantification of proteins above 0.001 NadjSPCs, marked in grey. C, Comparison of protein abundance (measured as NadjSPC) between each of the three biological replicates (repl). \( R^2 = \) Correlation Coefficient. \( N = \) number of proteins. Proteins only present in one of each pair were included in the correlation analysis and were represented by a zero value when absent.

**Figure 3.** Determination of the core PG proteome. A, The PG proteome flow diagram. 234 unique proteins were identified and quantified from the in-solution and in-gel experiments of *A. thaliana* (Col-0) PGs. These 234 proteins were passed through 4 sequential filters, first by comparing the protein abundance between the PG and representative preparations of total leaf, thylakoid and stroma (measured as average NadjSPC) and then discarding proteins with abundances less than 0.001 NadjSPC in PG preparations and not found in both methods. 32 proteins passed all 4 filters. Manual curation then found the callosin, RD20, and PLAT/LH2-2, to be ER-localized (Aubert et al., 2010) (http://gfp.stanford.edu/index.html) and were manually removed, resulting in a core PG proteome comprised of 30 proteins. This core proteome then served as the input for a genome-wide co-expression analysis. B and C, Sub-organelar distribution (B) and abundance (C) of the 52 proteins passing the total leaf filter (enriched > 10-fold in PGs) were plotted. Proteins with PG/stroma < 20 or PG/thylakoid < 5 are gray and were eliminated from the core proteome by the filter series. D, Relative mass contributions of the 30 PG core proteins to the total core PG proteome.

**Figure 4.** M48 metalloprotease, ABC1K3 and ABC1K1 are enriched in the PG preparations. Immunoblots of a thylakoid sample (prior to sonication) and the PGs (subsequently extracted by sonication), illustrate enrichment levels comparable to those determined by mass spectrometry. The Ponceau stain is included for each blot as a loading control. 1x = 10\( \mu \)g.

**Figure 5 (in color).** Fibrillin localization is determined by isoelectric point and hydrophobicity. The isoelectric point and hydrophobicity (GRAVY index) were measured for 16 FBN protein products (\(-cTP\)) using the ProtParam tool at the ExPasy website (http://expasy.org/). FBN7a(1-290) and FBN7a(1-133) indicate the two truncation products of FBN7a, analyzed for localization in (Vidi et al., 2006). The 16 proteins were grouped by sub-cellular localization and plotted by hydrophobicity (GRAVY index) and isoelectric point (pI). *FBN9 was only observed with low MOWSE scores in total leaf extracts.
Figure 6 (in color). PG network visualization and functional enrichment. For each PG gene, the 20 strongest co-expressing genes from a genome-wide analysis by MetaOmGraph were compiled into a PG co-expression network and visualized with the Cytoscape software program, using the force-directed layout algorithm. Each gene is represented by a single node. Edges, representing co-expression interactions between PG genes and co-expressed genes, are colored according to the functional annotation of the co-expressed gene. Co-expression relationships between two PG genes are indicated with red. Visualization reveals 4 functional modules in which co-expressed genes are enriched in specific cellular/plastidic processes. Each module is shaded in grey and the enriched cellular processes are indicated. Six PG genes are not included in a functional module. For each, the number of plastid-targeted genes (out of 20) and potential relevant co-expressers are listed. Twenty seven co-expressors that are located at important positions in the network, and/or that have particularly interesting functions, are marked with numbers and are: 1 – Pheophytinase; 2 - Pheophorbide a oxygenase (PaO or ACD1); 3 - FtsH8; 4 - CCD1; 5 - ZDS; 6 - PDS1; 7 - FtsH2; 8 - DegP1; 9 - EF-TU-Lep; 10 - Fd1-like; 11 - Trx M1; 12 - Trx M2; 13, FTR beta; 14 - CcdA cytf assembly; 15 - AKRed-like; 16 - Haloacid dehalogenase domain protein; 17 - methyltransferase domain protein; 18 - Tyr kinase; 19, Beta-glucosidase 9; 20 - ZEP; 21 - NAD kinase 2; 22 - STN7; 23 - TAP38; 24 – PTOX/Immutans; 25 - NDH-N; 26 - NDF1; 27 - NDF2; 28 – MCS; 29 CSK.

Figure 7. Gene expression of eight selected genes in Arabidopsis leaves during natural leaf senescence determined by RT-PCR. Transcript accumulation of five genes from module 1 (ACD1, PPH, DGAT3, ABC1K7, MCS), two genes from module 4 (ABC1K9 and FBN4) and the uncharacterized esterase (AT5G41120), which was not on the microarray experiments, and therefore could not be incorporated in the co-expression network. ACTIN2 was used as internal loading control. Leaf tissue was selected from five time points during the course of natural leaf senescence: 1 = leaf rosette from plants beginning to bolt; 2 = leaf rosette from plant beginning to flower; 3 = senescing leaf ~10% chlorotic, 4 = senescing leaf ~50% chlorotic; 5 = senescing leaf ~50% chlorotic, one week later in senescence. The experiment was carried out in three independent replicates, with similar results; data for one of the replicates is shown.

Figure 8. A model for plastoglobule function in plastid metabolism and short and long term photo response and adaptation. Physical connectivity of the PG and thylakoid permit extensive exchange of metabolites between the two sub-compartments and possibly also facilitates recruitment of proteins from the thylakoid-stromal exposed surface to the PG. During times of high lipid or protein turnover (such as senescence, stress or plastid biogenesis) the role of the PG becomes especially pertinent. We illustrate
here some of the proposed functions of the PG in these processes. Turnover of galactolipids by DAD1-like acylhydrolases will release free fatty acids transported to the PG where they can either i) be incorporated into TAG by DGAT3/4, ii) enter the jasmonic acid synthesis pathway, in the case of linolenic acid (18:3), or iii) be esterified to free phytol into fatty acid phytyl esters (FAPEs) during concurrent chlorophyll degradation. Alternatively, the free phytol can be recycled for incorporation into tocopherols by two subsequent phytol kinases, the first of which has been identified (VTE5) (Valentin et al., 2006). During chlorophyll degradation, the tetrapyrrole head group is captured by the PG-localized SOUL/heme-binding protein (SOUL/HBP) and delivered for further degradation to the stroma. We predict that the four PG oxidoreductases (NDC1, AKRed and FRed1 and 2) are active in re-reducing oxidized lipophilic compounds sequestered in the PG. Supporting this, NDC1 has recently been demonstrated to display NAD(P)H reductase activity towards a plastoquinone (PQ-9) analog (decyl-PQ) (Eugeni-Piller et al., 2011). We expect that NDC1 and the other PG oxidoreductases are responsible for regeneration of oxidized PG quinones following ROS scavenging. PQ-9 is expected to be exchanged between the PG and thylakoid. Selective uptake of reduced (or oxidized) PQ-9 would permit a powerful control over the redox state of PQ-9 in the thylakoid and thus over a number of processes regulated by PQ-9 redox state, including photosynthetic electron flow, retrograde signaling, carotenoid desaturation, and LHCII state transition. The presence of the carotenoid cleavage dioxygenase 4 (CCD4), suggests the presence of carotenoid catabolism at the PG. Carotenoids released from the photosynthetic apparatus (photosystems and light harvesting complexes) can be directed to the PG by FBN4 (or other FBNs) for degradation by CCD4. ABC1K9, positioned as a hub in module-4 (Fig. 6), is regulating the localization or function of fibrillins (FBNs) and thioredoxins (TRXs) as well as components of the cyclic electron flow apparatus (NDH- and/or PGR5-dependent). FBNs will be controlling the size of PGs (dashed arrow), while TRXs will control Calvin Cycle activity to match supply of reducing power produced from photosynthesis. Increased Calvin cycle activity will create additional demand for ATP that can be met by upregulated cyclic electron flow (CEF), either NDH- or PGR5-dependent (Livingston et al., 2010; Livingston et al., 2010). Metabolites are enclosed in grey boxes, PG-localized proteins are marked in red and co-expressers are marked in blue. Abbreviations used in the figure are listed: Thioredoxin (TRX), cyclic electron flow (CEF), fibrillin (FBN), light-harvesting complex (LHC), zeaxanthin epoxidase (ZEP), state transition kinase (STN7), phytoene desaturase (PDS), zeta-carotene desaturase (ZDS), plastoquinone-9 (PQ-9), reactive oxygen species (ROS), 9-cis-epoxycarotenoid dioxygenase (NCED), plastochromanol-8 (PC-8), triacylglycerol (TAG), lipoxygenase (LOX), fatty acid phytyl ester (FAPE), 9,13-hydroperoxy-octadecatetraenoic acid (9,13-HPOT), allene oxide synthase (AOS), phytol kinase (VTE5), SOUL/heme-binding protein domain (SOUL/HBP), pheophytinase (PPH), metal chelating substance (MCS*) possible represented by AT5G17450, pheophorbide a (pheide a), pheophorbide a oxygenase
(PaO), 12-oxo-phytodienoic acid (OPDA), poly-unsaturated fatty acids (PUFA), abscisic acid (ABA),
diacyl glycerol acyl transferase 3 and 4 (DGAT3/4), red chlorophyll catabolite (RCC), red chlorophyll
catabolite reductase (RCCR), primary fluorescent chlorophyll catabolite (pFCC).

SUPPLEMENTAL DATA

Supplemental Table 1. Experimental data of in-gel and in-solution PG proteome analysis
Supplemental Table 2. Comparison of protein abundances in PG, total leaf, thylakoid and stroma
Supplemental Table 3. MetaOmGraph co-expression results – Top 20
Supplemental Table 4. Functional group enrichment of PG co-expressers using different software
programs
Supplemental Figure 1. Five-day light-shifted wild-type A. thaliana plant, representative of those used
in this work.
Supplemental Figure 2. Scanning electron micrographs of PG preparations
Supplemental Figure 3. Co-expression within gene sets of chlorophyllide biosynthesis and the ClpPR
protease complex.
Supplemental Figure 4. PG genes preferentially maintain co-expression with other PG genes at higher
PCCs.
Supplemental Figure 5. Co-expression relationships between PG genes and isoprenoid metabolism
genes found in MetaOmGraph have been projected onto the isoprenoid pathway.

Supplemental Text
Figure 1. PG extraction and purification efficiency. A, Immunoblot of a thylakoid sample prior to sonication and purified PG fractions. Proteins were separated by 1D SDS-PAGE and probed with antibody against PG marker protein, VTE1. 1x corresponds to 2.2 μg protein. B, Immunoblot of thylakoid samples prior to sonication and after sonication for removal of PGs. A Ponceau stain of protein is included as a loading control. 1x corresponds to 10 μg protein. C, Histogram illustrating the distribution of PG diameters from transmission electron micrographs (TEM) of chloroplasts in mature light stressed leaf tissue and scanning electron micrographs (SEM) of purified PG preparations, n ≥ 263 measurements. D, Transmission electron micrograph (TEM) of 5-day light stressed leaf chloroplast of A. thaliana (Col-0), representative of the time point PG preparations were made. An image of a typical light stress Arabidopsis rosette plant is shown. PGs in the TEM, marked by white arrows, appear as gray circles with black periphery or, less frequently as solid black circles. E, SEM of A. thaliana (Col-0) purified PGs, demonstrating efficient isolation of PGs with varying diameters. Scale bars in (D) and (E) = 500 nm.
Figure 2. Identification of the PG proteome by in-gel and in-solution methods. A, Three independent preparations of PGs were made from 5-day light stressed *A. thaliana*. Thylakoids were isolated from total leaf tissue and sonicated to release PGs. Aliquots of each PG preparation were then separated by SDS-PAGE and in-gel digested or delipidized, in-solution digested, and zip-tipped. In-solution and in-gel digested peptides were analyzed with a nanoLC-LTQ-Orbitrap mass spectrometer. 234 unique proteins were identified in total, 129 of which were identified by both experimental methods. B, Comparison of NadjSPC between each of the 234 proteins between the in-gel and in-solution methods demonstrates consistent quantification of proteins above 0.001 NadjSPCs, marked in grey. C, Comparison of protein abundance (measured as NadjSPC) between each of the three biological replicates (repl). $R^2$, Correlation Coefficient. N, number of proteins. Proteins only present in one of each pair were included in the correlation analysis and were represented by a zero value when absent.
Figure 3. Determination of the PG core proteome. A, The PG proteome flow diagram. 234 unique proteins were identified and quantified from the in-solution and in-gel experiments of 5-day light-stressed A. thaliana (Col-0) PGs. These 234 proteins were passed through 4 sequential filters, first by comparing the protein abundance between the PG and representative preparations of total leaf, thylakoid and stroma (measured as average NadjSPC) and then discarding proteins with abundances less than 0.001 NadjSPC in PG preparations and not found in both methods. 32 proteins passed all 4 filters. Manual curation then found the caleosin, RD20, and PLAT/LH2-2, to be ER-localized (Aubert et al., 2010) (http://gfp.stanford.edu/index.html) and were manually removed, resulting in a core PG proteome comprised of 30 proteins. This core proteome then served as the input for a genome-wide co-expression analysis. B and C, Sub-organellar distribution (B) and abundance (C) of the 52 proteins passing the total leaf filter (enriched > 10-fold in PGs) were plotted. Proteins with PG/stroma < 20 or PG/thylakoid < 5 are gray and were eliminated from the core proteome by the filter series. D, Relative mass contributions of the 30 PG core proteins to the total PG core proteome.
Figure 4. M48 protease, ABC1K1 and ABC1K3 are enriched in the PG preparations. Immunoblots of a thylakoid sample (prior to sonication) and the PGs (subsequently extracted by sonication), illustrate enrichment levels comparable to those determined by mass spectrometry. The Ponceau stain is included for each blot as a loading control. 1x = 10µg.
Figure 5 (in color). Fibrillin localization is determined by isoelectric point and hydrophobicity. The isoelectric point and hydrophobicity (GRAVY index) were measured for 16 FBN protein products (-cTP) using the ProtParam tool at the ExPasy website (http://expasy.org/). FBN7a(1-290) and FBN7a(1-133) indicate the two truncation products of FBN7a, analyzed for localization in (Vidi et al., 2006). The 16 proteins were grouped by sub-cellular localization and plotted by hydrophobicity (GRAVY index) and isoelectric point (pl). *FBN9 was only observed with low MOWSE scores in total leaf extracts.
Fig. 6 (in color)
- Legend on next page
Figure 7. Gene expression of eight selected genes in *Arabidopsis* leaves during natural leaf senescence determined by RT-PCR. Transcript accumulation of five genes from module 1 (*ACD1, PPH, DGAT3, ABC1K7, MCS*), two genes from module 4 (*ABC1K9* and *FBN4*) and the uncharacterized esterase (*AT5G41120*), which was not on the microarray experiments, and therefore could not be incorporated in the co-expression network. *ACTIN2* was used as internal loading control. Leaf tissue was selected from five time points during the course of natural leaf senescence: 1 = leaf rosette from plants beginning to bolt; 2 = leaf rosette from plant beginning to flower; 3 = senescing leaves ~10% chlorotic, 4 = senescing leaves ~50% chlorotic; 5 = senescing leaves ~50% chlorotic, one week later in senescence. The experiment was carried out in three independent replicates, with similar results; data for one of the replicates is shown.
Fig. 8 (in color) – legend on next page