RESEARCH ARTICLE

Enrichment and Analysis of Intact Phosphoproteins in Arabidopsis Seedlings

Uma K. Aryal\textsuperscript{a}, Andrew R. S. Ross\textsuperscript{b\*}, Joan E. Krochko

National Research Council of Canada, Saskatoon, SK, S7N 0W9, Canada

\textsuperscript{a} Current Address: Fisheries and Oceans Canada, Institute of Ocean Sciences, Sidney, British Columbia, V8L 4B2, Canada
\textsuperscript{b} Current Address: Department of Biochemistry, Purdue University, West Lafayette, Indiana, 47907, United States of America

* andrew.ross@dfo-mpo.gc.ca

Abstract

Protein phosphorylation regulates diverse cellular functions and plays a key role in the early development of plants. To complement and expand upon previous investigations of protein phosphorylation in Arabidopsis seedlings we used an alternative approach that combines protein extraction under non-denaturing conditions with immobilized metal-ion affinity chromatography (IMAC) enrichment of intact phosphoproteins in Rubisco-depleted extracts, followed by identification using two-dimensional gel electrophoresis (2-DE) and liquid chromatography-tandem mass spectrometry (LC-MS/MS). In-gel trypsin digestion and analysis of selected gel spots identified 144 phosphorylated peptides and residues, of which only 18 phosphopeptides and 8 phosphosites were found in the PhosPhAt 4.0 and P3DB Arabidopsis thaliana phosphorylation site databases. More than half of the 82 identified phosphoproteins were involved in carbohydrate metabolism, photosynthesis/respiration or oxidative stress response mechanisms. Enrichment of intact phosphoproteins prior to 2-DE and LC-MS/MS appears to enhance detection of phosphorylated threonine and tyrosine residues compared with methods that utilize peptide-level enrichment, suggesting that the two approaches are somewhat complementary in terms of phosphorylation site coverage. Comparing results for young seedlings with those obtained previously for mature Arabidopsis leaves identified five proteins that are differentially phosphorylated in these tissues, demonstrating the potential of this technique for investigating the dynamics of protein phosphorylation during plant development.

Introduction

Seedling establishment is a critical stage in plant development, involving the transition from heterotrophic to autotrophic growth.\cite{1} In Arabidopsis, seed germination is driven largely by the metabolism of storage products other than lipids, whereas seedling establishment involves the mobilization of seed oil reserves.\cite{2} Triacylglycerol (TAG) is the predominant source of carbon in the seeds of Arabidopsis and related species, including Brassica napus (canola).\cite{3}
and mobilization of TAG supplies the energy and molecular building blocks required for seedling establishment.\cite{1,4} Utilization of TAG and other seed reserves is thought to be controlled and regulated by multiple pathways,\cite{5} and although considerable progress has been made in understanding dormancy and seed germination\cite{6–8} the cellular mechanisms involved in seedling establishment are less well understood.

Following germination the glycerol released from TAG through lipase action is converted to glyceraldehyde-3-phosphate (G-3-P) and then by isomerization to dihydroxyacetone phosphate (DHAP), which can either undergo glycolysis to pyruvate or conversion to hexose via gluconeogenesis.\cite{9} The free fatty acids are catabolized by $\beta$-oxidation in the glyoxysome. A more complete understanding of how this metabolic program is regulated in \textit{Arabidopsis} would increase our knowledge of post-embryonic development in plants and assist in the improvement of canola and other oilseed crops.

One way to achieve this is to study protein phosphorylation during early stage of seedling establishment because reversible phosphorylation of proteins regulates a wide variety of cellular processes during plant growth and development.\cite{10} However, the analysis of protein phosphorylation can be challenging due to the low relative abundance of phosphoproteins and the possibility of phosphorylation at multiple sites within a given protein.\cite{11,12} Affinity enrichment of phosphorylated proteins and/or the component phosphopeptides obtained by proteolysis can significantly enhance the identification of such proteins and the mapping of phosphorylation sites. However, phosphoproteome analysis of certain plant tissues is complicated by the presence of D-ribulose bisphosphate carboxylase/oxygenase (Rubisco),\cite{11,13} an abundant phosphoprotein that inhibits the detection and analysis of other, less abundant plant proteins. Rubisco depletion columns (e.g. Seppro IgY spin columns; GenWay Biotech, San Diego CA, USA) are commercially available and have been used successfully to deplete Rubisco in total protein extracts.\cite{14,15} Advances in phosphopeptide enrichment strategies have also enabled large-scale phosphoproteomic studies in \textit{Arabidopsis}, providing new insights regarding the potential involvement of protein phosphorylation in various stages of plant development.\cite{11,16–23} Despite these efforts, our knowledge of protein phosphorylation events during the transition from heterotrophic to photoautotrophic growth in young seedlings remains incomplete.

Much of the information currently available in protein phosphorylation databases has been generated using peptide-level enrichment strategies, and although affinity purification of intact phosphoproteins has been demonstrated\cite{11,24} the use of protein-level enrichment for phosphoproteome analysis in plants remains largely unexplored. To complement and expand upon previous investigations involving phosphopeptide enrichment we carried out a survey of protein phosphorylation in post-embryonic \textit{Arabidopsis} seedlings (hereafter referred to as young seedlings) using Rubisco depletion and enrichment of intact phosphoproteins by immobilized metal-ion affinity chromatography (IMAC) combined with two-dimensional gel electrophoresis (2-DE) and liquid chromatography-tandem mass spectrometry (LC-MS/MS). The results of this study were then compared with those obtained previously using Rubisco depletion and protein-level enrichment of phosphoproteins from \textit{Arabidopsis} mature leaves\cite{11} to evaluate this approach for monitoring the dynamics of protein phosphorylation during plant development.

**Materials and Methods**

**Materials**

Seppro Rubisco IgY Spin Columns (GenWay Biotech, San Diego, CA, USA) were obtained from Sigma-Aldrich (St. Louis, MO, USA; Product No. SEP070). Acrylamide, bisacrylamide
solution, IPG dry strips (pH 3–10, NL, 17 cm), carrier ampholytes, Precision Plus Protein standards, TEMED, TBP, DTT, and IAA were purchased from Bio-Rad (Hercules, CA). Urea was from Merck KGaA (Darmstadt, Germany), Tris base from Roche Diagnostics (Indianapolis, IN), PPS silent surfactant from Protein Discovery Inc. (Knoxville, TN), and trypsin (sequencing grade) from Promega (Madison, WI, USA). PHOS-Select iron affinity gel beads and Sigma-Prep spin columns were purchased from Sigma-Aldrich (St.Louis, MO, USA). All other chemicals were also from Sigma-Aldrich unless otherwise stated, and were of analytical research grade.

Plant growth and protein extraction

*Arabidopsis thaliana* (L) Heynh (Col-0) seeds were treated with 50% bleach in MilliQ water (v/v) containing 5.25% sodium hypochlorite for 2 min and then with 50% (v/v) ethanol for 2 min before washing 4 times with sterilized Milli-Q water and cultivating in Petri dishes containing 0.5× Murashige and Skoog [25] mineral salts with BactoAgar. Seeds were allowed to imbibe at 4°C for 4 days in the dark before transferring them to a growth chamber maintained at 22°C and a 16-h light/8-h dark cycle. Week-old whole seedlings (including roots) with 2 seed leaves were collected for protein extraction. One gram of seedlings was ground to a powder in liquid nitrogen with 0.5% (w/w) PVPP and homogenized in 2 ml of ice-cold extraction buffer (pH 7.4) containing 10 mM Tris-HCl, 150 mM NaCl, the serine protease inhibitor PMSF (1 mM, freshly prepared in DMSO) and a protease inhibitor cocktail developed for plant cell and tissue extracts (Sigma P-9596, 0.2% v/v), together with phosphatase inhibitors 20 mM sodium fluoride, 1 mM sodium molybdate, 1 mM sodium orthovanadate, and 1 mM sodium β-glycerophosphate. The slurry was stirred for 30 min on ice, filtered through two layers of cheese cloth and centrifuged at 10,000 × g for 15 min at 4°C. After discarding the pellet the amount of protein in the supernatant was determined using the Bradford assay (Bio-Rad) with BSA as the standard, and the final concentration of the sample adjusted to 1 mg/ml using the extraction buffer.

Rubisco depletion

Each protein sample was filtered through a 0.45 μm spin filter (Millipore) and 500 μl of the extract, containing about 500 μg of protein, was loaded onto a Seppro IgY column. Rubisco was removed according to the manufacturer’s instructions. The protein flow-through and the bound fraction were collected separately, and each precipitated with 5 volumes of ice-cold methanol and 100 mM ammonium acetate at -20°C overnight. After centrifugation at 10,000 × g for 20 min at 4°C, the resulting pellets were thoroughly washed twice with ice-cold 100% methanol and then with 80% ice-cold methanol. Each pellet was briefly dried using a SpeedVac, re-dissolved in the column incubation buffer (6 M urea, 0.25% CHAPS, 50 mM sodium acetate, pH 4.0) to approximately 1 mg/ml, and used for phosphoprotein enrichment by immobilized metal-ion affinity chromatography (IMAC).

Phosphoprotein enrichment

Enrichment of intact phosphoproteins from Rubisco-depleted samples was carried out as previously described.[11] Briefly, a 500 μl slurry of PHOS-Select iron affinity gel beads (Sigma) was washed 3 times with 0.1% TFA in 30% acetonitrile and equilibrated 3 times with 500 μl of incubation buffer (6 M urea, 0.25% CHAPS, 50 mM sodium acetate, pH 4.0) with centrifugation at 1,000 × g for 1 min between each step before being loaded onto the column. Two ml of Rubisco-depleted protein sample in incubation buffer were loaded onto each spin column (about 2 mg total protein per 500 μl of bead slurry) and incubated for 1 h at room temperature with gentle shaking. Phosphoproteins bound to the IMAC columns were eluted three times.
with 200 μl of elution buffer (6 M urea, 50 mM Tris-acetate pH 7.5, 0.1 M EDTA, 0.1 M EGTA, 0.25% CHAPS), each time incubating at room temperature for 10 min with gentle shaking, and then centrifuged at 1,000 × g for 1 min. The 3 eluates were pooled and precipitated with methanol as previously described before re-suspension in lysis buffer (7 M urea, 2 M thiourea, 2% CHAPS) to obtain a total protein concentration of 1 μg/μl prior to 2-D gel electrophoresis (2-DE).

Gel electrophoresis and in-gel digestion

One-dimensional gel electrophoresis was used to resolve proteins from the bound and flow-through fractions obtained during Rubisco depletion on the Seppro column. Ten μl of each fraction containing approximately 10 μg of protein was mixed with 10 μl of gel sample buffer (0.2 M Tris-HCl, pH 6.8, 2% SDS, 10% glycerol, 0.02% bromophenol blue) and separated on a 1.0 mm, 12.5% Criterion Tris/HCl gel in a Criterion Cell (Bio-Rad) (13.3 cm × 8.7 cm) at a constant voltage of 150 V. The separated proteins were visualized using Bio-Safe Coomassie Blue stain (Bio-Rad).

For phosphoproteome analysis, 200 μl (200 μg) of IMAC-enriched phosphoprotein sample was mixed with 200 μl of rehydration buffer (7 M urea, 2 M thiourea, 2% CHAPS, 10 mM DTT, 0.5% IPG buffer, pH 3–10), resolved by 2-DE and visualized by silver staining. Gel images were recorded using an ImageScanner (GE Healthcare) and Phoretix 2D software (v2004) was used to measure the total number of protein spots visualized in each 2-DE gel image. Proteins of interest were excised manually from each gel and digested with trypsin using a MassPREP protein digestion station, according to the protocol (digestion 5.0) recommended by the manufacturer (Micromass, Manchester, UK). Preparation of tryptic peptide samples for LC-MS/MS analysis was carried out as previously described.

Mass spectrometry and protein identification

Six μl of each 2-D gel protein digest was analyzed using a nanoAQUITY UPLC system (Waters, Milford, MA, USA) interfaced to a quadrupole time-of-flight (Q-TOF) Ultima Global hybrid tandem mass spectrometer (Waters, Mississauga, ON, Canada). Separations were performed using a Waters BEH130 C18 nanoAQUITY UPLC analytical column (75 μm, 1.75 mm × 100 mm) at an initial flow rate of 400 nl/min. Mobile phase solvent A was 0.2% formic acid in water and solvent B was 0.2% formic acid in 100% acetonitrile. Separations were performed using the following 55-min solvent program: 99:1 (%A:%B) for 1 min, changing to 90:10 at 16 min, 55:45 at 45 min, and 20:80 at 46 min, at which point the flow rate was changed to 800 nl/min and the gradient held until 52 min before reverting to 99:1 at 53 min. A 5 min seal wash with 10% acetonitrile in water was carried out after the completion of each run.

The Q-TOF MS was operated in the positive ion mode and TOF MS spectra were acquired over the m/z range 400–1900 at the rate of one scan/s. Of the multiply charged (2+, 3+, or 4+) peptide ion peaks rising over a threshold, the three most abundant were automatically selected for CID, and product-ion spectra were acquired over the m/z range 50–1900 in TOF MS/MS mode. The CID collision energy was selected automatically according to the m/z ratio and charge state of the precursor ion. A real-time exclusion window was used to prevent precursor ions with the same m/z from being selected for CID and TOF MS/MS within 2 min of their initial acquisition. Data were also acquired using pre-programmed exclusion lists for keratin and trypsin.

Data were processed using MassLynx 4.1 (Waters, Milford, MA) and searched against NCBI protein sequence database for Arabidopsis thaliana (thale cress) using an in-house Mascot server (Version 2.2, Matrix Sciences, UK) and the following parameters: carbamidomethylation of
cysteine as the fixed modification; oxidation of methionine and phosphorylation of serine, threonine and tyrosine as variable modifications; mass tolerances of 0.2 Da for MS and 0.5 Da for MS/MS data; and one missed cleavage for tryptic peptides. Peptide MS/MS spectra used for protein identification had to be of sufficient quality, with a signal-to-noise ratio of 3 or greater for annotated fragment ions, including neutral loss peaks associated with de-phosphorylation during CID. Only peptides matched with significant ion scores ($P < 0.05$) and low expectation values ($e$-value < 0.01) were selected. For unambiguous identification, each peptide MS/MS spectrum had to contain at least three sequential y- or b-type ions. Protein identification was regarded as positive if the Mascot score exceeded the 95% confidence threshold, the matched protein contained at least four top-ranking unique peptides, and protein sequence coverage by the matching peptides was >15%. If the same set of peptides matched multiple members of a protein family, or a protein appeared under different names and accession numbers in the database, the entry with the highest score and/or most descriptive name was reported. When protein isoforms were observed, the data were inspected manually. If several isoforms shared the same set of identified peptides the protein with the most matching peptides was accepted as the correct result. The presence of protein isoforms was confirmed and reported based on the identification of at least two unique peptides.

Since the error tolerance of the MS method used (200 mDa) was greater than the mass difference between phosphorylation and sulfation (9.5 mDa), a second error-tolerant search reporting masses to 0.1 mDa was performed to allow sulfation and phosphorylation to be distinguished. Raw MS/MS spectra matched to phosphorylated peptides in the Mascot search were manually inspected and validated using MassLynx 4.1. The spectra were processed to give singly charged, monoisotopic, centroided peaks and compared with the in silico fragmentation masses for the matched peptide to confirm neutral loss of phosphoric acid for serine and threonine phosphorylation, or the mass increment of 80 Da associated with phosphorylation of tyrosine.

Results

Phosphoproteome analysis of young seedlings

A schematic representation of our analytical approach is shown in Fig 1. The molecular weight distributions of proteins in the bound and flow-through samples following Rubisco depletion were investigated by 1-DE (Fig 2). Results show that the Seppro IgY Rubisco-depletion columns are efficient at removing Rubisco from the protein extracts of young seedlings. The Rubisco protein concentrated in the bound fraction is predominantly the small subunit (SSU), whereas both small and large subunits of Rubisco were evident in a previous study of mature Arabidopsis leaves.[11] That study also found that Rubisco depletion significantly increased the number of identified phosphoproteins, even without IMAC enrichment, and that only Rubisco and other relatively abundant phosphoproteins were recovered from non-depleted extracts using IMAC, whereas IMAC enrichment more than doubled the number of phosphoproteins identified in depleted extracts. It has recently been demonstrated that the Rubisco SSU up-regulates expression of the Rubisco large subunit (LSU) at the transcriptional level. This coordinated expression of subunits may explain the relatively small amount of Rubisco LSU observed during early growth in young seedlings.[26]

IMAC-purified phosphoproteins from the Rubisco-depleted flow-through fraction were subsequently resolved by 2-DE (Fig 3). The reproducibility of both 1- and 2-DE experiments was confirmed by analyzing and comparing three biological replicates (not shown). An average of 175 protein spots were detected in replicate 2-DE gels following IMAC enrichment of Rubisco-depleted extracts. These were excised, trypsinized and analyzed by LC-MS/MS, which
identified 156 of the spots based on our acceptance criteria for protein identification (see above). Of these, 105 spots (i.e. 60% of the 175 detected following IMAC) were found to contain a total of 82 different phosphoproteins based on the detection of 144 tryptic phosphopeptides, not counting methionine-oxidized and non-oxidized forms of the same peptide (Table 1, S3 Fig). The spot in which each phosphoprotein had been identified with highest confidence was subsequently labeled on a representative 2-DE gel image (Fig 3). Although significantly depleted in the flow-through fraction (Fig 2) Rubisco SSU was still detectable in 2-DE gels (Fig 3, spot 65).

The 144 detected phosphopeptides contained a total of 144 unique sites of protein phosphorylation, of which 48% (69) were serine, 48% (69) were threonine, and 4% (6) were tyrosine residues (Table 2; Fig 4A). To assess any differences in phosphorylation occupancy among the S, T and Y residues, we compared our results with those from previous studies that utilized different enrichment methods and plant tissues (Table 2). The distribution observed in this study for Arabidopsis seedlings is similar to that obtained using Rubisco depletion and IMAC enrichment of intact phosphoproteins from mature Arabidopsis leaves,[11] which contained 52% phosphoserine (pS), 40% phosphothreonine (pT), and 8% phosphotyrosine (pY) residues (Table 2). However, these results differ from those obtained using IMAC to enrich phosphopeptides generated by trypsin digestion of plant phosphoproteins. For example, previous results reported 88% pS, 11% pT and 1% pY in 22-day-old Arabidopsis seedlings [22]; 85% pS, 13% pT and 2% pY in 9-day-old Arabidopsis seedlings [23]; 85% pS, 11% pT and 4% pY in cultured Arabidopsis cells [21]; 86% pS, 13% pT and 1% pY in Medicago truncatula roots [26]; and 81% pS, 17% pT, and 2% pY in dormant poplar (Populus simonii × P. nigra) buds [27] when IMAC enrichment was performed at the phosphopeptide level. Rao and Moller [28] reported the occurrence of 77% pS, 17.5% pT and 5.5% pY in eukaryotic phosphoproteins based on a combined Uniprot, Phospho.ELM and Phosida database analysis, which also differs
from the present study. By way of comparison, the average pS:pT:pY ratio observed for cellular phosphoproteins in mammals is approximately 1800:200:1,[29] corresponding to 89.95% pS, 10.00% pT and 0.05% pY. These results suggest that peptide- and protein-level enrichment strategies complement each other to some extent and that the latter provides access to a greater proportion of phosphorylated threonine and tyrosine residues, at least in plant phosphoproteins.

These findings are of particular significance given the emerging importance of tyrosine phosphorylation in plant processes such as germination, growth, development, and abiotic stress responses.[30] In particular, our discovery of 6 new tyrosine phosphorylation sites (Table 1) in proteins involved in the mobilization of seed reserves (NAD+ MDH), cell defence (MLP), cellular signaling (cyclase family protein), oxidative stress response (GST9), protein degradation (20S proteasome alpha subunit B) and protein folding (chaperonin 20) represents a significant contribution to the list of potential substrates for known and predicted protein tyrosine kinases in plants.[30] It also helps to address the apparent discrepancy between the
predicted frequency of pY residues in the Arabidopsis proteome [13] and that observed using peptide-level affinity enrichment strategies, during which the phosphorylated residues in each protein are distributed between tryptic peptides containing only one or two such residues, of which those carrying the more abundant pS modification are likely to predominate in terms of recovery and analysis.

Of the 144 phosphopeptides and 144 phosphosites reported in the present study, only 10 peptides and 1 phosphorylation site matched those identified during a recent survey of the phosphoproteome in hydroponically-grown Arabidopsis seedlings, which utilized Ti$^{4+}$-IMAC enrichment of tryptic phosphopeptides from whole protein digests (Table 1).[23] Of those 144 phosphopeptides, 10 phosphopeptides and 5 phosphorylation sites were found in both the
### Table 1. List of identified phosphoproteins.

| Spot No. | Gene locus | Protein name | MW/ pI | Functional group | Phosphopeptide (a) | psite |
|----------|------------|--------------|--------|-----------------|--------------------|-------|
| 1 | gi| 15233613 | O-acetylserine (thiol) lyase (OASA1) | 33.9/5.9 | Amino acid biosynthesis | DVPTELIGNTPLVYLNNVAEGCVGR | T10 |
| 2 | gi| 18404496 | Catalytic co-enzyme binding protein (b) | 35.8/8.4 | Amino acid biosynthesis | ALDLaSKEGTGTPTK (c) | S302 |
| 3 | gi| 15218373 | Cystidine/deoxyxystidylate deaminase (b,d) | 20.4/5.6 | Amino acid biosynthesis | YpTDPTAHAETVIR | T75 |
| 4 | gi| 15235213 | Caffeoyl-CoA 3–0 methyltransferase | 29.3/5.1 | Amino acid biosynthesis | TSPSTNGEDQKQSQNLR pTSTNGEDQKQSQNLR | S13 T11 |
| 5 | gi| 30691732 | Aminoacylase, putative | 48.0/5.9 | Amino acid biosynthesis | pTKGEAGTNVVEAVR | T165 |
| 6 | gi| 15224470 | Pyridoxin biosynthesis protein PDX1.1 (b) | 33.1/5.8 | Amino acid biosynthesis | TpSKPEIFPASTDAR | T387 |
| 7 | gi| 15233161 | Peroxiglycinamidine cycloligase | 41.6/5.3 | Amino acid biosynthesis | GLAhpTGggFTDNIpR | T296 |
| 8 | gi| 42573371 | Carbonic Anhydrase 2 (CA2) (b,d) | 28.7/5.4 | Cellular metabolism | GNEpSYEDAEIALKK (b) KlpTAELQAAASSSKD (c,f) | S5 T35 S98 S191 |
| 9 | gi| 7769871 | NAD malate dehydrogenase, mitochondrial (b) | 37.2/8.5 | Cellular metabolism | VCPpSHVLDFHPDGFVVR | T175 T251 T114 S146 |
| 10 | gi| 15219721 | Malate dehydrogenase cytoplasmic 1 (b,d) | 35.9/6.1 | Cellular metabolism | VQpTSSGKEPpR (c,f) | S116 T90 T90 T4 |
| 11 | gi| 15226185 | Fructose bisphosphate aldolase (b) | 42.5/8.2 | Cellular metabolism | YpSAEGENGAdaKK | S372 |
| 12 | gi| 30678347 | Carbonic anhydrase 1 chloroplast (CA1) (b) | 29.8/5.5 | Cellular metabolism | VCPpSHVLDFQPpDAVFVR | S98 S189 |
| 13 | gi| 15232468 | Malate dehydrogenase (NAD), mitochondrial | 36.0/8.3 | Cellular metabolism | VVILGAGGIGQpLpSLLMK | S46 |
| 14 | gi| 16398 | Nucleotide diphosphate kinase (b) | 16.3/7.9 | Cellular metabolism | NVIHqppSDSVESAp (c,f) IIGApNp accumulation MEGqTFM* pKpPGVQR | S116 T90 T90 T4 |
| 15 | gi| 15230595 | Phosphoglycerate kinase 1 (PGK1) (b) | 50.1/6.1 | Cellular metabolism | VLPqVIALDEApVT | T480 |
| 16 | gi| 15220167 | 3-Isopropyl malate dehydrogenase 2 cytoplasmic | 43.5/5.3 | Cellular metabolism | VpSPEVIAEHTVpTvPAAPVpAVFLGDSGpSEEEATp | S239 T254 S84 |
| 17 | gi| 15231715 | Fructose bisphosphate aldolase (FBA), putative (b,d) | 38.8/6.1 | Cellular metabolism | IGEpSEHISINDAYGpLpLqGDggLpSLLMK | S155 S350 T333 T33 S42 T230 S233 |
| 18 | gi| 15222848 | Glyceraldehyde-3-phosphate dehydrogenase C-2 (GAPC-2) (b) | 37.0/6.7 | Cellular metabolism | pTLFPKEpTVpFGIR | T70 S152 |
| 19 | gi| 15229231 | Glyceraldehyde-3-phosphate dehydrogenase C subunit (GAPC), cytosolic (b,d) | 37.0/6.6 | Cellular metabolism | pTLFPKEpTVpFGIR | T70 |
| 20 | gi| 15218869 | Isocitrate dehydrogenase (IDH) (b) | 46.1/6.1 | Cellular metabolism | pTIEAEAAAHGTVTR | T302 |

(Continued)
Table 1. (Continued)

| Spot No.* | Gene locus | Protein name | MW/ pI | Functional group | Phosphopeptide (a) | psite |
|-----------|------------|--------------|--------|------------------|--------------------|-------|
| 21 gil    | gi| 15238559    | Glutamine synthatase 2 (GS2), mitochondrial *(b) | 47.8/ 6.4 | Cellular metabolism | pTIEKPVEDPSELPK *(c) | T302, T154, T162 |
| 22 gil    | gi| 15222551    | Phosphoribulose kinase (PRK), chloroplastic | 44.7/ 5.7 | Cellular metabolism | HADFPgSNNGTGLFQTIVGKL | S360 |
| 23 gil    | gi| 4539316     | Fructose bisphosphate aldolase, putative | 43.1/ 6.8 | Cellular metabolism | RLDSIGLENPTEANR | T91 |
| 24 gil    | gi| 15226185    | Fructose bisphosphate aldolase | 42.5/ 8.2 | Cellular metabolism | YpSAEGENEDAKK | S372 |
| 25 gil    | gi| 15227752    | Peroxisomal malate dehydrogenase (PMDH1) | 37.8/ 8.1 | Cellular metabolism | KLMVpTMLDVVR KPGM* pTRDLFNINAGIR | T138, T127 |
| 26 gil    | gi| 15228198    | PYK 10 binding protein 1 (PBP1) *(b,d) | 32.1/ 5.5 | Cell defense | pSPEEVGTGEEHK | S196, T299 |
| 27 gil    | gi| 18405982    | Avirulence-responsive protein | 19.6/ 5.0 | Cell defense | LHACpSPSINGLGKTVEVLPDTSEKK | S56, T97 |
| 28 gil    | gi| 15236568    | Major latex protein related/ MLP-related | 17.5/ 5.9 | Cell defense | EIDDEpTKTLTLVRpVpDYYQFIPK | T79, Y98, S139 |
| 29 gil    | gi| 15223957    | Major latex protein related/ MLP related | 18.0/ 6.4 | Cell defense | FVpTSLAADMDDHILK | T138 |
| 30 gil    | gi| 15236566    | Major latex-related/ MLP-related | 17.8/ 5.9 | Cell defense | RNDFPpSRYM | S131 |
| 31 gil    | gi| 1755154     | Germin-like protein | 22.0/ 6.8 | Cell defense | AAVpTPAFAPAYINGGLVSLAR | T72 |
| 32 gil    | gi| 15228199    | Jacalin lectin family protein *(b) | 32.2/ 5.9 | Cell signaling | pSPEEVGTGEEHK | S195, T299 |
| 33 gil    | gi| 15228216    | Jacalin lectin family protein | 32.0/ 5.1 | Cell signaling | pTSPPYGLETQKKKVHVGQGQDV | T104, S37 |
| 34 gil    | gi| 15226403    | Cupin family protein *(b) | 55.9/ 5.8 | Cell signaling | NRQPFLVgpSNSLIRpTGPEFVGFTTSAHKGpSGSSECEDSYNNYDKK *(e) | S456, T433, S320 |
| 35 gil    | gi| 18418598    | Cyclase family protein | 30.0/ 5.6 | Cell signaling | AGLpYSVHCLPR | Y249 |
| 36 gil    | gi| 15241018    | A. thaliana Ferretin 1 (ATFER1), chloroplastic *(a) | 28.1/ 5.7 | Electron transport | ADLAIPpTSHASLAR *(c) | T80 |
| 37 gil    | gi| 9843639     | Rieske FeS protein | 24.6/ 8.8 | Electron transport | GAPpSLALAHIDEAGK | S196 |
| 38 gil    | gi| 15231176    | ATP synthase D chain, mitochondrial (ATPQ) | 19.6/ 5.1 | Membrane transport | VpTPEYKPK | T96 |
| 39 gil    | gi| 7525040     | ATP synthase CF1 beta subunit | 53.9/ 5.4 | Membrane transport | IVGEEHYEpTAAQVK | T387 |
| 40 gil    | gi| 15231008    | Translocase of outer mitochondrial membrane 40 | 34.2/ 6.3 | Membrane transport | GKiDpSNGVASALLEER | S269 |
| 41 gil    | gi| 1143394     | V-type proton ATPase | 26.2/ 6.0 | Membrane transport | IDYpSMQLNASR pSNPGLHCSGGVVLASR | S71, S178 |
| 42 gil    | gi| 15236722    | ATP synthase family *(b) | 23.9/ 5.8 | Membrane transport | ALDpSQAIALSDEVKK *(a) | S203 |
| 43 gil    | gi| 7708276     | ATP synthase beta subunit | 52.5/ 5.2 | Membrane transport | INPPpTpGSGVMTLEK | T5, T6, S7 |
| 44 gil    | gi| 15227104    | putative ATP synthase subunit | 27.6/ 6.3 | Membrane transport | EKiIpTLDPEPDAVK | T71 |

(Continued)
| Spot No. | Gene locus | Protein name                                           | MW/ pl   | Functional group | Phosphopeptide (a) | psite |
|---------|------------|--------------------------------------------------------|----------|------------------|-------------------|------|
| 45      | gi|15236678| Ascorbate peroxidase 4 (APX4), chloroplastic           | 38.1/   | Oxidative stress | AENEGLpSDGLSLIEEVKK | S155 |
| 46      | gi|15223576| Dehydroascorbate reductase 1 (DHR1)                   | 23.4/   | Oxidative stress | pTPAESVGSNIFGTFLK  | T91  |
| 47      | gi|15224582| Glutathione S-transferase 10 (ATGSTF10)               | 24.2/   | Oxidative stress | VLpTIVAPLFASSK     | T4   |
| 48      | gi|15224581| Glutathione S-transferase 9 (ATGSTF9) (b,d)           | 24.1/   | Oxidative stress | QPAVLQOPFGpTVAPVGGDK Y146 |
| 49      | gi|15218640| Glutathione S-transferase 6 (ATGSTF6)                 | 23.5/   | Oxidative stress | VFGHPASTpApTR       | T15  |
| 50      | gi|15226610| ATPDIL2-1/MEE30/UNE5 (PD1) (b)                        | 39.8/   | Oxidative stress | GpSYASKETER ELVAApSEDEKK | S321 |
| 51      | gi|15231718| Peroxiredoxin type 2, chloroplastic                    | 24.7/   | Oxidative stress | QM*GLpSDFK1VALS2H   | S152 |
| 52      | gi|3093971 | Universal stress protein family protein               | 17.9/   | Oxidative stress | DLKLPsIVMSGR       | S125 |
| 53      | gi|15232567| A. thaliana thioredoxin M-type 4 (ATHM4), chloroplastic | 21.3/   | Oxidative stress | INpTDESPPANTAR OpSIIAGVPRETK | T144 |
| 54      | gi|6539610 | Thioredoxin M2, chloroplastic (a)                      | 20.6/   | Oxidative stress | TTLpTSSLKFLP LNpTDESPPGQYVR | T178 |
| 55      | gi|3121825 | 2-Cys peroxiredoxin, chloroplast precursor            | 29.0/   | Oxidative stress | pSGGLGDNLPSDVTK     | S161 |
| 56      | gi|15223049| Ascorbate peroxidase 1 (APX1), cytosolic              | 27.8/   | Oxidative stress | QM*GlpsSDK1VALS2H   | S152 |
| 57      | gi|15219086| Protein disulfide isomerase (PDI)-like protein        | 55.8/   | Oxidative stress | pSADDASEVVS1DKK (c) | S149 |
| 58      | gi|2097312 | Glutathione S-transferase 6 (GST6) (b)                | 24.1/   | Oxidative stress | AlpTYLAEEYSEKGEK    | T72  |
| 59      | gi|18415155| 2-Cys peroxiredoxin, chloroplastic                    | 29.9/   | Oxidative stress | pSFGVLIPDQIALR pS1GLGDNLPSDITK | S189 |
| 60      | gi|15228407| Superoxide dismutase 1 (MSD1), mitochondrial (b)      | 25.5/   | Oxidative stress | YApSEVYEKKN        | S223 |
| 61      | gi|7653843 | Peroxiredoxin IIF (b)                                 | 21.3/   | Oxidative stress | LAEgpTD1SAAPVSLQK pSGLGDNLPSDITK | T35  |
| 62      | gi|15228194| Sedoheptulose-1,7-bisphosphatase, chloroplastic (b,d) | 42.7/   | Photosynthesis   | GFgpT1HEELL1DEGK   | T235 |
| 63      | gi|15229349| Ribose 5-phosphate isomerase-related (b,d)            | 29.4/   | Photosynthesis   | LLpSGGELDIVGIFTSK pS1GLGDNLPSDITK | S86  |
| 64      | gi|414550  | Cytosolic triosephosphate isomerase (TPI) (b)          | 27.4/   | Photosynthesis   | AILNepsSSEFVGDK1AYAQLK | S106 |
| 65      | gi|13926229| Rubisco small chain 1A (RBCS1A) (b)                   | 14.9/   | Photosynthesis   | EH1NgpT5PGYGD (e) KFe1pT3L1P0L0CLSE1AK (f) | S58  |
| 66      | gi|16194   | Rubisco small subunit (RbcS)                          | 20.6/   | Photosynthesis   | EH1NgpT5PGYGD KFe1pT3L1P0L0CLSE1AK (f) | T113 |
| 67      | gi|15223217| Glycine cleavage system H protein, mitochondrial (b)  | 18.0/   | Photosynthesis   | VKP1SSPAELEALMPGK V1KpSSPAELEALMGK | S141 |
| 68      | gi|84468442| Putative Rubisco subunit binding protein (b)           | 48.6/   | Photosynthesis   | HEAAGDPpTTTASILAR  | T8   |

(Continued)
Table 1. (Continued)

|Spot No.*| Gene locus | Gene name | Protein name | MW/ pI | Functional group | Phosphopeptide (a) | psite |
|----------|------------|-----------|--------------|--------|-----------------|-------------------|------|
|69 gi| 13926291| PS II oxygen-evolving complex 1 (PSBO1) | PS II oxygen-evolving complex 1 (PSBO1) | 35.3/ 5.6 | Photosynthesis and respiration | QLDApSGKPDSFTGK | S221 |
|70 gi| 20260472| Glycolate reductase | Glycolate reductase | 36.7/ 8.5 | Photosynthesis and respiration | pSKCDPLVGGLGAK pSYGLSDEDFSAVIEALK | S85 S320 |
|71 gi| 18416540| CIP amino terminal domain containing protein | CIP amino terminal domain containing protein | 26.1/ 9.2 | Protein degradation | pSMNEDVLSFKK | S224 |
|72 gi| 18390982| ATP dependent Clp protease proteolytic subunit (CLPP), chloroplast | ATP dependent Clp protease proteolytic subunit (CLPP), chloroplast | 36.4/ 8.6 | Protein degradation | pSVAYNEHRPR VPSpGLML*PASDVLIR | S107 S241 |
|73 gi| 15224993| 20S proteasome subunit PAA2 | 20S proteasome subunit PAA2 | 27.4/ 5.8 | Protein degradation | LLDQpSSVSHLFPVTK A pTSAGMKQEAVNFLEK | S64 T166 |
|74 gi| 2511588| Multicatalytic endopeptidase complex | Multicatalytic endopeptidase complex | 27.2/ 5.6 | Protein degradation | LLDQpSSVTHLFPITK | S63 |
|75 gi| 15219317| 20S Proteasome alpha subunit B, putative | 20S Proteasome alpha subunit B, putative | 25.7/ 5.8 | Protein degradation | KLPpSILVDEASVOK LpYKEPIPVQLVR | S54 |
|76 gi| 15242045| Chaperonin 20 (CPN20), chloroplastic | Chaperonin 20 (CPN20), chloroplastic | 26.8/ 8.9 | Protein folding | YpTSDMELDAAHTAILTK | T179 T194 |
|77 gi| 15226314| Chaperonin HSP 60A (CPN60A) | Chaperonin HSP 60A (CPN60A) | 62.2/ 5.1 | Protein folding | pTAKHDIPILEK | T272 |
|78 gi| 16221| Chaperonin HSP60 | Chaperonin HSP60 | 61.6/ 5.7 | Protein folding | VpTKDGVTVAK | T80 |
|79 gi| 62321455| Putative cruciferin 12S seed storage protein | Putative cruciferin 12S seed storage protein | 19.9/ 7.9 | Seed storage | GLPLEVIpTNGYQISPEEAK VFDQEpSSGQLLVPGFSSVM*K | T143 |
|80 gi| 166678| 12S storage protein | 12S storage protein | 50.9/ 6.8 | Seed storage | VFDQEpSSGQLLVPGFSSVM*K GLPLEVIpTNGYQISPEEAKR | T420 |
|81 gi| 9758672| Unnamed protein product | Unnamed protein product | 29.0/ 5.9 | Unclassified | pTNENQENVTLNAGR | T395 |
|82 gi| 18391006| Unknown protein | Unknown protein | 20.0/ 5.4 | Unclassified | EIpSMPNGLLPLK | S33 |

Phosphorylated proteins, peptides and residues (phosphosites) identified by mass spectrometry in protein extracts from young Arabidopsis seedlings after rubisco depletion, IMAC enrichment and 2-DE separation of phosphoproteins. (a) pS, pT and pY = phosphorylated serine, threonine and tyrosine residues; M* = oxidized methionine. (b) Protein reported in the PhosPhAt 4.0 database. (c) Peptide reported in the PhoPhAt 4.0 database with a different protein phosphorylation site. (d) Protein previously reported in Arabidopsis seedlings. (e) Peptide previously reported in Arabidopsis seedlings with the same protein phosphorylation site. (f) Peptide previously reported in Arabidopsis seedlings with a different protein phosphorylation site. (g) Peptide previously reported in Arabidopsis seedlings with the same protein phosphorylation site.

doi:10.1371/journal.pone.0130763.t001

Table 2. Distribution of phosphorylated residues identified in plant proteins using immobilized metal-ion affinity chromatography of phosphorylated proteins or peptides.

| Plant species | Tissue | IMAC target | % pS | % pT | % pY | References |
|---------------|--------|-------------|------|------|------|------------|
| Arabidopsis thaliana | seedlings | pProteins | 48 | 48 | 4 | This study |
| Arabidopsis thaliana | leaves | pProteins | 52 | 40 | 8 | 11 |
| Arabidopsis thaliana | cultured cells | pPeptides | 85 | 11 | 4 | 21 |
| Arabidopsis thaliana | seedlings | pPeptides | 88 | 11 | 1 | 22 |
| Arabidopsis thaliana | seedlings | pPeptides | 85 | 13 | 2 | 23 |
| Medicago truncatula | roots | pPeptides | 86 | 13 | 1 | 25 |
| Poplar simonii × P. nigra | dormant buds | pPeptides | 81 | 17 | 1 | 26 |

doi:10.1371/journal.pone.0130763.t002
P3DB (http://www.p3db.org/) and PhosPhAt 4.0 (http://phosphat.uni-hohenheim.de/) databases, with an additional 8 peptides and 3 phosphorylation sites found only in the PhosPhAt 4.0 database.

Of the identified phosphoproteins previously reported in *Arabidopsis thaliana* seedlings [23] (Table 1) two are isoforms of the same protein, GAPC (spots 18 and 19). Detection of the novel phosphopeptide pTLLFGEKPVTVGIR in both isoforms indicates that both are phosphorylated at T70. However, a second phosphopeptide SDLDIVpSNASCTTNCLAPLAK, which had previously been detected in *Arabidopsis* seedlings [23] (though with a different site of phosphorylation), was also identified in one of the isoforms (spot 18) indicating phosphorylation at S152 (Table 1). The concomitant reduction in pI relative to the other isoform (spot 19) is consistent with horizontal separation of these two proteins on the 2-DE gel (Fig 3), demonstrating the utility of our gel-based approach for resolving differentially phosphorylated phosphoproteins.

![Distribution and functional classification of identified phosphoproteins](http://www.genome.jp/kegg/pathway.html). Proteins involved in carbohydrate/energy metabolism, oxidative stress/redox regulation and photosynthesis/respiration account for over 50% of the identified phosphoproteome.

doi:10.1371/journal.pone.0130763.g004
forms of a given protein. Similarly, vertical separation of two Rubisco polypeptides (Fig 3, spots 65 and 66) reflects the difference in molecular weight between the matched proteins, each of which contained the same number (3) of identified phosphorylation sites (Table 1).

IMAC purification, 2-DE separation, and digestion of intact phosphoproteins to produce a mixture of phosphorylated and non-phosphorylated peptides may have contributed to the relatively small number of multiply-phosphorylated peptides identified during this study, compared with studies in which only phosphorylated peptides were enriched and analyzed by mass spectrometry. However, the average number of phosphopeptides identified per plant protein (1.8 in young Arabidopsis seedlings and 1.9 in mature leaves) compares well with studies that utilize peptide-level enrichment. Furthermore, the phosphoproteins we identified in young seedlings using protein-level enrichment include basic proteins (e.g. APX1, APX4, nucleotide diphosphate kinase) and proteins previously identified as plasma membrane proteins (e.g. CA2, PGK, DHAR1) in Arabidopsis seedlings, suggesting minimal bias towards proteins of a particular polarity, pl or molecular weight. By enabling protein identification using both phosphorylated and non-phosphorylated peptides our approach also provides high confidence in the identification of phosphoproteins and hence, their selection as candidates for further investigation of the role of protein phosphorylation during plant development (which lies beyond the scope of the present study).

Functional classification of phosphoproteins

The identified phosphoproteins were sorted into functional groups using the KEGG Pathway database (http://www.genome.jp/kegg/pathway.html). The two largest groups were those involved in carbohydrate/energy metabolism (22%) and oxidative stress/redox regulation (20%), which together with photosynthesis and respiration (11%) accounted for more than half of the identified phosphoproteins (Fig 4). Many of these, including glyceraldehyde-3-phosphate dehydrogenase (GAPC-2), triosephosphate isomerase (TPI), phosphoglycerate kinase (PGK1), fructose bisphosphate aldolase (FBA), and malate dehydrogenase (MDH), play important role in processes such as glycolysis, gluconeogenesis and the Calvin cycle during seed germination and the early stages of seedling establishment. Identification of phosphorylated 20S proteasome subunits, proteases, chaperonins, thioredoxins, glutathione transferases (GSTs), dehydroascorbate reductase (DHAR1) and manganese superoxide dismutase (MSD1) is also consistent with the role of proteolytic events in mobilizing TAG and other seed reserves. Comparison of our experimental results with the PhosPhAt 4.0 Arabidopsis thaliana phosphorylation site database and P3DB database with supplementary information from a recently published survey of the Arabidopsis seedling phosphoproteome showed that 43 of the 82 phosphoproteins identified in our study have not been reported before (Table 1), and that we were able to identify new phosphorylation sites in previously characterized phosphoproteins such as FBA, GAPC-2, TPI and PMDH1, GSTs, PRK and IDH. New and known phosphorylation sites were also identified in 12S seed storage proteins (Table 1), further demonstrating the utility of our approach for identifying novel sites of protein phosphorylation in plant tissues.

Discussion

Phosphorylation of enzymes involved in post-embryonic development

Many of the enzymes known to be important during the early stages of plant growth were found to be phosphorylated in Arabidopsis young seedlings. The glycolytic enzyme triosephosphate isomerase (TPI), for example, plays a central role in chloroplast development and other biochemical pathways by equilibrating the cytosolic pool of DHAP and G-3-P. The latter
is required for 1,5-bisphosphate production in the Calvin cycle, whereas DHAP suppresses the production of chlorophyll and 1,5-bisphosphate. Phosphorylation of human TPI has been shown to reduce its activity in converting G-3-P to DHAP, and although it has been suggested that TPI can be phosphorylated at S21 there is evidence that other sites may be subject to phosphorylation. Our discovery of phosphorylated S106 and S178 residues in Arabidopsis TPI (Table 1, spot 64) provides new information with which to investigate the role of protein phosphorylation in controlling the activity of this enzyme and thus regulating chloroplast development in young seedlings.

NAD+ MDH, a key enzyme in carbohydrate metabolism, is responsible for regenerating NAD+ and is involved in the mobilization of seed oil reserves and the photosynthetic assimilation of carbon in developing leaves. We identified several sites of phosphorylation in mitochondrial NAD+ MDH (Table 1, spot 9), as well as single site of phosphorylation in cytosolic MDH (Table 1, spot 10). We also observed phosphorylation of 3-isopropyl malate dehydrogenase (spot 16), which is primarily involved in leucine biosynthesis.

Carbonic anhydrase (CA), a major chloroplast protein, is involved in photosynthesis and the mobilization of seed reserves during the early stages of post-embryonic growth. CA1 is also known to form part of a Rubisco-containing Calvin cycle enzyme complex. Identification of phosphorylation sites in CA (spots 8 and 12), ribose-5-phosphate isomerase (spot 63), Rubisco SSU (spot 66) and PRK (spot 22) may help to elucidate the role of protein phosphorylation in controlling the assimilation and utilization of carbon reserves during the early stages of seedling establishment. Other identified phosphoproteins include members of the jacalin-lectin (Fig 5A), cupin, and cyclase families (spots 32 to 34), all of which are involved in cell signaling. A cupin domain protein (AtPirin1) has also been found to interact with G protein α-subunit GPA1 in Arabidopsis to regulate seed germination and seedling development.

Phosphorylation of 20S proteasome subunit PtpPBA1, and increased expression of 20S proteasome α-subunit B and regulatory subunit RPN10, have been observed in poplar dormant terminal buds. We observed phosphorylation of 20S proteasome α-subunit B (spot 75) at S54, T179, T194 and Y101 and of the 20S proteasome subunit PAA2 (spot 73) at S64 and T166 in Arabidopsis young seedlings (Table 1). ATP dependent Clp protease proteolytic subunit (CLPP) is a highly conserved, multimeric serine protease that degrades large globular proteins in the presence of an AAA ATPase complex. CLPP (spot 72) was found to be phosphorylated at S107 and S241, and a Clp amino terminal domain-containing protein (spot 71) at S224.

Although there is growing evidence of crosstalk between redox signaling and hormonal response pathways during seed germination, the molecular components involved in this process during post-embryonic development remain elusive. We identified phosphorylated forms of several proteins known to be key regulators of stress response, including APX1 (spot 56), APX4 (spot 45), GST6 (spot 58), ATGSTF9, -10 and -6 (spots 47 to 49 and S2A Fig and S2B Fig), dehydroascorbate reductase 1 (spot 46), thioredoxins M2 and M4 (spots 53 and 54), peroxiredoxins (spots 51 and 59), and manganese superoxide dismutase (spot 60). Phosphorylation of APX1, APX4, peroxiredoxin type-2, GST6, and MSD1 was also observed in mature leaves but at sites other than those observed in young seedlings (Table 3). Thioredoxins and other H2O2-scavenging enzymes help to protect plants from damage caused by the production of reactive oxygen species (ROS) during seed germination and seedling development. Germin-like protein, which generates H2O2 from the oxidative breakdown of oxalate, was also found to be phosphorylated in our study (Fig 5B).

Heat shock proteins (HSPs) are involved in bud dormancy and phosphorylation of HSPs and chaperonin has been reported in Arabidopsis and poplar. Our results
confirm phosphorylation of these proteins in *Arabidopsis* seedlings and identify sites of phosphorylation in HSP60 (T80) and chaperonin 20 (adjacent residues Y59, T60 and S61) that, to the best of our knowledge, have not been reported before (S2C Fig and S2D Fig).

Comparing protein phosphorylation at different stages of development

In a previous study we used IMAC to recover and identify 132 phosphoproteins with 252 component phosphopeptides in mature *Arabidopsis* leaf extracts (Fig 4A), following polyethylene glycol (PEG) fractionation to deplete Rubisco.[11] Having now used IMAC to recover and identify intact phosphoproteins in Rubisco-depleted extracts from young seedlings we decided...
Table 3. Changes in protein phosphorylation between post-embryonic seedlings and mature leaves. Phosphorylated proteins, peptides and residues (S = serine, T = threonine, Y = tyrosine) identified in post-embryonic seedlings and mature leaves of Arabidopsis thaliana. Common phosphopeptides with conserved phosphosites are highlighted in bold and common phosphopeptides with different phosphosites are highlighted in bold and italics.

| Gene locus | Protein name | Young seedlings | Mature leaves | Phospho-site (s) |
|------------|--------------|-----------------|---------------|-----------------|
| gi|18404496  | Catalytic co-enzyme binding | ALDLApSKPEGTGTPTK | ALDLApSKPEGTGTPTK | S302 |
| gi|15218869  | Isocitrate dehydrogenase | pTIEAEEAHGTVTR | - | T302 |
| gi|15219721  | Malate dehydrogenase (MDH) | VOpTSSGEEKPVR | - | T203 |
| gi|15222848  | G3P cytosolic-2 (GAPC-2) | pTLLFGEKPVTVFGR | pTLLFGEKPVTVFGR | T70 |
| gi|15231715  | Fructose bisphosphate aldolase | pTVPAAVPAIVFSLGSQSEEAEATR | pTVPAAVPAIVFSLGSQSEEAEATR | T254 |
| gi|15227752  | Malate dehydrogenase (PMDH1) | KLMGVpTMLDVVR | - | T138 |
| gi|15229231  | G3P cytosolic (GAPC) | pTLLFGEKPVTVFGR | pTLLFGEKPVTVFGR | T70 |
| gi|16398     | Nucleotide diphosphate kinase | NVIHOpSDESESAR | NVIHOpSDESESAR | S116 |
| gi|414550    | Cytosolic triose phosphate isomerase | AILNEpSSEFVGDKVAYALAQQGLK | AILNEpSSEFVGDKVAYALAQQGLK | S106 |
| gi|42573371  | Carbonic Anhydrase 2 (CA2) | VApSPAQAQEVHDELRK | VApSPAQAQEVHDELRK | S178 |
| gi|7769871   | NAD-malate dehydrogenase | RTQDGpgTEVEAK | pTQDGGEVEAK | T251, T246 |

(Continued)
| Gene locus | Protein name | Young seedlings | Mature leaves | Phospho-site(s) |
|------------|--------------|-----------------|---------------|----------------|
| gil15228198 | PYK 10 binding protein 1 (PBP1) | pSPEEVVTGEEHGK | pSPEEVVTGEEHGK | S196 |
| | | QpTSPPPFLEAGTFELKEEGHK | QTPSPPFLEAGTFELKE | T254, S255 |
| | | TpSPPFLEAGTFELKE | TpSPPFLEAGTFELKE | S20 |
| | | GpSPPFLEAGTFELKE | GpSPPFLEAGTFELKE | S325 |
| | | VpVYQADGK | VpVYQADGK | T175 |
| gil1755154 | Germin-like protein | AAVpTAPAFAPAYAGINGGLGVSLAR | - | T72 |
| | | GpSVMFGQQLHGFQNLNGK | GpSVMFGQQLHGFQNLNGK | S140 |
| gil18405982 | Avirulence-responsive protein | LHACIpSPENGLINGK | - | S56 |
| | | TVEVLPDTSEKK | TVEVLPDTSEKK | T97, T91 |
| gil9843639 | Rieske FeS protein | FLCPCGHspSQYNAGQR | FLCPCGHspSQYNAGQR | S180 |
| | | GpAPLspSLAHADIDEAGK | GpAPLspSLAHADIDEAGK | S196 |
| | | GpSAPLspSLAHADIDEAGK | GpSAPLspSLAHADIDEAGK | T138 |
| gil1143394 | V-type proton ATPase | IDYpSMQLNAR | IDYpSMQLNAR | S71 |
| | | GpSNDPGHLHCQGGVVLAR | GpSNDPGHLHCQGGVVLAR | S178 |
| gil752040 | ATP synthase CF1 beta subunit | IVEEHYEpTAQVK | IVEEHYEpTAQVK | T387 |
| | | TNPPpTTSNPEVSIR | TNPPpTTSNPEVSIR | T3 |
| | | VGLpTALTMAEYFR | VGLpTALTMAEYFR | T252 |
| gil15223049 | L-ascorbate peroxidase (APX1) | QM*GLpSDKDIVALSGAHTLGR | - | S152 |
| | | ELLpSGLKEGLIQLQVSDK | ELLpSGLKEGLIQLQVSDK | S196 |
| gil15226610 | ATPDIL2-1/MEE30/UNE5 | AGHDYDGRDLDFVpSFINEK | DLDDFVpSFINEK | S243 |
| | | ELVApSEDEKK | - | S280 |
| | | GpSDYASKETER | - | S321 |
| gil15228407 | Mn-superoxide dismutase (MSD1) | YApSEVVEKDN | - | S223 |
| | | GpSLGSIDAHFSGLELGK | GpSLGSIDAHFSGLELGK | S124 |
| | | HHOAYVTNpYNNAELDQAVK | HHOAYVTNpYNNAELDQAVK | Y67 |
| gil15231718 | Peroxiredoxin type 2 | pTILFAVPGAFTTCSQS | pTILFAVPGAFTTCSQS | T108 |
| | | VLNLEEGGAFpTNSSAEMDLK | VLNLEEGGAFpTNSSAEMDLK | T223 |
| | | LDPpSTSylDLPSTGDKV | LDPpSTSylDLPSTGDKV | S82 |
| gil15236678 | Ascorbate peroxidase 4 (APX4) | AENGLpSDGLSIEEVKK | - | S155 |
| | | GGPpSYADIIQLAQQAVK | GGPpSYADIIQLAQQAVK | S178 |
| gil20197312 | Glutathione S-transferase (GST6) | AlpTYLAEYSEKGEK | AlpTYLAEYSEK | T72 |
| | | GMFGMtpTTPDAVQLEGK | GMFGMtpTTPDAVQLEGK | T129 |
| | | QEAHLALNPFGQIPALEDDGpTFLFSR | QEAHLALNPFGQIPALEDDGpTFLFSR | T64 |

(Continued)
to compare the results of the two studies. Of the 82 phosphoproteins identified in post-embryonic seedlings 28 were also identified in mature leaves, with 26 component phosphopeptides showing the same sites of phosphorylation in both tissues (Fig 4A, Table 3). For example, phosphorylation of the Rubisco small chain 1A at T58, T14 and T78 was observed in both seedlings and leaves, confirming phosphorylation of the protein at those sites. However, some of the phosphopeptides spanning the same amino acid sequence in both tissues showed a difference in protein phosphorylation state between young seedlings and mature leaves. For example, the CA2 peptide VLAESESSAFEDQCGR was identified in both tissues but was phosphorylated at S191 in seedlings and at S193 in leaves. Tryptic peptides showing differential phosphorylation of four other proteins (NAD+ MDH, PBP1, avirulence responsive protein, and ribose 5-phosphate isomerase) were also observed (Table 3), suggesting that these proteins may play a significant role in Arabidopsis development.

Comparing the phosphorylation status of 12S seed storage protein (cruciferin) in young seedlings (Table 1, spot 80) and dormant Arabidopsis seeds [50] shows that certain phosphorylation sites (T395 and T420) are common to both tissues, thereby validating the current method with reference to results obtained during a previous in-depth study of cruciferin phosphorylation. However, an apparent shift in phosphorylation site from S367 in dormant seeds to S366 in post-embryonic seedlings again demonstrates the ability to detect subtle changes in phosphorylation status that may have implications for seed storage protein mobilization and other processes during plant development,[50] although further investigations are required to confirm the significance of these findings.

Conclusions

Seedling establishment involves the efficient utilization of endogenous protein reserves and external resources, requiring that developmental and metabolic programs adapt to the

Table 3. (Continued)

| Gene locus | Protein name | Young seedlings | Mature leaves | Phospho-site (s) |
|------------|--------------|-----------------|---------------|------------------|
| gi|13926229 | Rubisco small chain 1A (RBCS1A) | EHGNpTPGYDGR | EHGNpTPGYDGR | T58 |
| | | KFEpTLSYLDPDLTDSELAK | KFEpTLSYLDPDLSDVELAK | T14 |
| | | - | FEpTLPpSYPDLDSDVELAK | T14 |
| | | - | KFEpTLPsSYPDLDSDLpELAK | T14, S16 |
| | | LPLFGCPpTDpAQVLK | LPLFGCPpTDpAQVLK | T78 |
| gi|15229349 | Ribose 5-phosphate isomerase | LLpSGpPELYDVpGpITSK | LLpSGpPELYDVpGpITSK | S86, S87 |
| | | pSLGIPpLVpGLpDLpHPR | pSLGIPpLVpGLpDLpHPR | S108 |
| | | LQDFKEFGCEpSK | - | S206 |
| | gil1526314 | Chaperonin 60 alpha (CPN60A) | IpTAIKDIpILEK | - | T272 |
| | | - | HGLlpSVTSGAPVSLK | S150 |
| | gil15242045 | Chaperonin 20 (CPN20) | YpTSIKpLPGLpDR | YpTSIKpLPGLpDR | T60 |
| | | - | YpTSIKpLPGLpDR | Y59 |
| | | pTLGGILLpSTApQSKpQGEVpAVGEGpR | - | T80 |
| | gil16221 | Chaperonin HSP60 | VpTKDGVpTVAK | - | S151 |
| | | - | GIpSMApDApVVpTNLpK | S151 |

doi:10.1371/journal.pone.0130763.t003

Analysis of Intact Seedling Phosphoproteins

PLOS ONE | DOI:10.1371/journal.pone.0130763 July 9, 2015 19 / 23
prevailing environmental conditions.[51] Using a combination of Rubisco depletion and IMAC enrichment of intact phosphoproteins we identified and characterized the phosphorylated forms of 82 proteins expressed in Arabidopsis young seedlings. These included enzymes involved in chloroplast development, mobilization of TAG, and other processes known to be important during the early stages of plant development. Comparison of our results for young seedlings with those obtained previously for Arabidopsis seeds [50] and mature leaves [11] shows that some of these proteins undergo differential phosphorylation during plant growth, and that protein level enrichment appears to enhance detection of pT and pY residues. Our study complements previous investigations by identifying an additional 43 proteins and 136 residues that undergo phosphorylation in Arabidopsis young seedlings. By purifying and enriching phosphorylated proteins under non-denaturing conditions our approach also lends itself to the study of phosphorylation in endogenous protein complexes and during protein-protein interactions.

Supporting Information

S1 Fig. Identification of novel phosphorylation sites using tandem mass spectrometry (MS/MS). The MS/MS spectra correspond to phosphopeptides with the following mass-to-charge (m/z) ratios, as obtained by trypsin digestion of proteins selected from the 2-DE gel shown in Fig 3. (A) m/z 701.377, showing phosphorylation of FBA (spot 17) at S84; (B) m/z 878.968, showing phosphorylation of GAPC-2 (spot 18) at T70; (C) m/z 619.981, showing phosphorylation of cytosolic TPI (spot 64) at S178; and (D) m/z 721.369, showing phosphorylation of PMDH1 (spot 25) at T138.

S2 Fig. Identification of novel phosphorylation sites using tandem mass spectrometry. The MS/MS spectra correspond to phosphopeptides with the following mass-to-charge (m/z) ratios, as obtained by trypsin digestion of proteins selected from the 2-DE gel shown in Fig 3. (A) m/z 745.425, showing phosphorylation of ATGST10 (spot 47) at T4; (B) m/z 612.269, showing phosphorylation of ATGST6 (spot 49) at T15; (C) m/z 549.250, showing phosphorylation of HSP60 (spot 78) at T80; and (D) m/z 615.308, showing phosphorylation of chaperonin 20 (spot 76) at S61.

S3 Fig. Phosphopeptide MS/MS spectra and MASCOT search results for selected phosphoproteins.

Acknowledgments

We thank Doug Olson and Steve Ambrose of the Mass Spectrometry and Protein Research Group at the National Research Council (NRC) in Saskatoon for technical support. UKA also acknowledges the generous support from the National Research Council (NRC) for a postdoctoral Visiting Fellowship. Funding for protein mass spectrometry equipment was provided by the Saskatchewan Provincial Government and NRC. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. This article is contribution number 54680 from the National Research Council of Canada.
Author Contributions
Conceived and designed the experiments: UKA ARSR. Performed the experiments: UKA. Analyzed the data: UKA. Contributed reagents/materials/analysis tools: JEK ARSR. Wrote the paper: UKA ARSR JEK.

References
1. Chen M, Thelen JJ (2010) The plastid isoform of triose phosphate isomerase is required for the postgerminative transition from heterotrophic to autotrophic growth in Arabidopsis. The Plant cell 22: 77–90. doi: 10.1105/tpc.109.071837 PMID: 20097871
2. Cernac A, Andre C, Hoffmann-Benning S, Benning C (2006) WRI1 is required for seed germination and seedling establishment. Plant physiology 141: 745–757. PMID: 16568074
3. O’Neill CM, Gill S, Hobbs D, Morgan C, Bancroft I (2003) Natural variation for seed oil composition in Arabidopsis thaliana. Phytochemistry 64: 1077–1090. PMID: 14568074
4. Graham IA (2008) Seed storage oil mobilization. Annual review of plant biology 59: 115–142. doi: 10.1146/annurev.arplant.59.032607.092938 PMID: 18444898
5. Horvath DP, Anderson JV, Chao WS, Foley ME (2003) Knowing when to grow: signals regulating bud dormancy. Trends in plant science 8: 534–540. PMID: 14607098
6. Martin T, Oswald O, Graham IA (2002) Arabidopsis seedling growth, storage lipid mobilization, and photosynthetic gene expression are regulated by carbon:nitrogen availability. Plant physiology 128: 472–481. PMID: 11842151
7. Rajjou L, Belghazi M, Huguet R, Robin C, Moreau A, Job C, et al. (2006) Proteomic investigation of the effect of salicylic acid on Arabidopsis seed germination and establishment of early defense mechanisms. Plant physiology 141: 910–923. PMID: 16679420
8. Waterworth WM, Ashley MK, West CE, Sunderland PA, Bray CM (2005) A role for phosphorylation in the regulation of the barley scutellar peptide transporter HvPTR1 by amino acids. Journal of experimental botany 56: 1545–1552. PMID: 15824072
9. Beevers H (1956) Utilization of Glycerol in the Tissues of the Castor Bean Seedling. Plant physiology 31: 440–445. PMID: 16654920
10. Bodenmiller B, Mueller LN, Mueller M, Domon B, Aebersold R (2007) Reproducible isolation of distinct, overlapping segments of the phosphoproteome. Nature methods 4: 231–237. PMID: 17293869
11. Aryan UK, Krochko JE, Ross AR (2012) Identification of Phosphoproteins in Arabidopsis thaliana Leaves Using Polyethylene Glycol Fractionation, Immobilized Metal-ion Affinity Chromatography, Two-Dimensional Gel Electrophoresis and Mass Spectrometry. Journal of proteome research 11: 425–437. doi: 10.1021/pr200917t PMID: 22092075
12. Tang W, Deng Z, Oses-Prieto JA, Suzuki N, Zhu S, Zhang X, et al. (2008) Proteomics studies of brassinosteroid signal transduction using prefractionation and two-dimensional DIGE. Molecular & cellular proteomics: MCP 7: 728–738.
13. Lohrig K, Muller B, Davydova J, Leister D, Wolters DA (2009) Phosphorylation site mapping of soluble proteins: bioinformatical filtering reveals potential plastidic phosphoproteins in Arabidopsis thaliana. Planta 229: 1123–1134. doi: 10.1007/s00425-009-0901-y PMID: 19238429
14. Cellar NA, Karnoup AS, Albers DR, Langhorst ML, Young SA (2009) Immunodepletion of high abundance proteins coupled on-line with reversed-phase liquid chromatography: a two-dimensional LC sample enrichment and fractionation technique for mammalian proteomics. Journal of chromatography B, Analytical technologies in the biomedical and life sciences 877: 79–85. doi: 10.1016/j.jchromb.2008.11.020 PMID: 19038586
15. Cellar NA, Kuppannan K, Langhorst ML, Ni W, Xu P, Young SA (2008) Cross species applicability of abundant protein depletion columns for ribulose-1,5-bisphosphate carboxylase/oxygenase. Journal of chromatography B, Analytical technologies in the biomedical and life sciences 861: 29–39. PMID: 18083427
16. Nuhse TS, Bottrill AR, Jones AM, Peck SC (2007) Quantitative phosphoproteomic analysis of plasma membrane proteins reveals regulatory mechanisms of plant innate immune responses. The Plant journal: for cell and molecular biology 51: 931–940.
17. Nuhse TS, Stensballe A, Jensen ON, Peck SC (2004) Phosphoproteomics of the Arabidopsis plasma membrane and a new phosphorylation site database. The Plant cell 16: 2394–2405. PMID: 15308754
18. Nuhse TS, Stensballe A, Jensen ON, Peck SC (2003) Large-scale analysis of in vivo phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry. Molecular & cellular proteomics: MCP 2: 1234–1243.
19. de la Fuente van Bentem S, Anrather D, Roitinger E, Djamei A, Hufnagl T, Barta A, et al. (2006) Phosphoproteomics reveals extensive in vivo phosphorylation of Arabidopsis proteins involved in RNA metabolism. Nucleic acids research 34: 3267–3278. PMID: 16807317

20. Benschop JJ, Mohammed S, O’Flaherty M, Heck AJ, Slijper M, Menke FL (2007) Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis. Molecular & cellular proteomics: MCP 6: 1198–1214.

21. Sugiyama N, Nakagami H, Mochida K, Daudi A, Tomita M, Shirasu K, et al. (2008) Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in Arabidopsis. Molecular systems biology 4: 193. doi: 10.1038/msb.2008.32 PMID: 18463617

22. Reiland S, Messeri G, Baerenfaller K, Gerrits B, Endler A, Grossmann J, et al. (2009) Large-scale Arabidopsis phosphoproteome profiling reveals novel chloroplast kinase substrates and phosphorylation networks. Plant physiology 150: 889–903. doi: 10.1104/pp.109.138677 PMID: 19376835

23. Wang X, Bian Y, Cheng K, Gu LF, Ye M, Zou H, et al. (2013) A large-scale protein phosphorylation analysis reveals novel phosphorylation motifs and phosphoregulatory networks in Arabidopsis. Journal of proteomics 78: 486–498. doi: 10.1016/j.jprot.2012.10.018 PMID: 23111157

24. Machida M, Kosako H, Shirakabe K, Kobayashi M, Ushiya M, Inagawa J, et al. (2007) Purification of phosphoproteins by immobilized metal affinity chromatography and its application to phosphoproteome analysis. FEBS J 274: 1576–1587. PMID: 17480206

25. Murasige T, Skoog F (1962) A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. Physiologia Plantarum 15: 473–497.

26. Grimsrud PA, den Os D, Wenger CD, Swaney DL, Schwartz D, Sussman MR, et al. (2010) Large-scale phosphoprotein analysis in Medicago truncatula roots provides insight into in vivo kinase activity in legumes. Plant physiology 152: 19–28. doi: 10.1104/pp.109.149625 PMID: 19923235

27. Liu CC, Liu CF, Wang HX, Shen ZY, Yang CP, Wei ZG (2011) Identification and analysis of phosphorylation status of proteins in dormant terminal buds of poplar. BMC plant biology 11: 158. doi: 10.1186/1471-2229-11-158 PMID: 22074553

28. Rao RSP, Moller IM (2012) Large-scale analysis of phosphorylation site occupancy in eukaryotic proteins. Biochimica Et Biophysica Acta-Proteins and Proteomics 1824: 405–412.

29. Mann M, Ong SE, Gronborg M, Steen H, Jensen ON, Pandey A (2002) Analysis of protein phosphorylation by mass spectrometry: deciphering the phosphoproteome. Trends Biotechnol 20: 261–268. PMID: 12007495

30. Ghelis T (2011) Signal processing by protein tyrosine phosphorylation in plants. Plant Signal Behav. 6: 942–951. PMID: 21628997

31. Nakagami H, Sugiyama N, Mochida K, Daudi A, Yoshida Y, Toyoda T, et al. (2010) Large-scale comparative phosphoproteomics identifies conserved phosphorylation sites in plants. Plant physiology 153: 1161–1174. doi: 10.1104/pp.110.157347 PMID: 20466843

32. Alexandersson E, Saalbach G, Larsson C, Kjellborn P (2004) Arabidopsis plasma membrane proteome identifies components of transport, signal transduction and membrane trafficking. Plant & cell physiology 45: 1543–1556.

33. Heazlewood JL, Durek P, Hummel J, Selbig J, Weckwerth W, Walther D, et al. (2008) PhosPhAt: a database of phosphorylation sites in Arabidopsis thaliana and a plant-specific phosphorylation site predictor. Nucleic acids research 36: D1015–1021. PMID: 17984086

34. Durek P, Schmidt R, Heazlewood JL, Jones A, MacLean D, Nagel A, et al. (2010) PhosPHat: the Arabidopsis thaliana phosphorylation site database. An update. Nucleic acids research 38: D828–834. doi: 10.1093/nar/gkp810 PMID: 19880383

35. Yao QM, Ge HY, Wu SQ, Zhang N, Chen W, Xu CH, et al. (2014) (PDB)-D-3 3.0: From plant phosphorylation sites to protein networks. Nucleic Acids Research 42: D1213–1219. doi: 10.1093/nar/gkt1135 PMID: 24243849

36. Lichtenthaler HK (1999) The 1-Deoxy-D-Xylulose-5-Phosphate Pathway of Isoprenoid Biosynthesis in Plants. Annual review of plant physiology and plant molecular biology 50: 47–65. PMID: 15012203

37. Tang GL, Wang YF, Bao JS, Chen HB (1999) Overexpression in Escherichia coli and characterization of the chloroplast triosephosphate isomerase from spinach. Protein expression and purification 16: 432–439. PMID: 10425165

38. Aoyagi K, Nakamoto H (1985) Pyruvate, pi dikinase in bundle sheath strands as well as in mesophyll cells in maize leaves. Plant physiology 78: 661–664. PMID: 16664304

39. He Y, Chen L, Zhou Y, Mawhinney TP, Chen B, Kang BH, et al. (2011) Functional characterization of Arabidopsis thaliana isopropylmalate dehydrogenases reveals their important roles in gametophyte development. The New phytologist 189: 160–175. doi: 10.1111/j.1469-8137.2010.03460.x PMID: 20840499
40. Jebanathirajah JA, Coleman JR (1998) Association of carbonic anhydrase with a Calvin cycle enzyme complex in Nicotiana tabacum. Planta 204: 177–182. PMID: 9487724

41. Teng N, Wang J, Chen T, Wu X, Wang Y, Lin J (2006) Elevated CO2 induces physiological, biochemical and structural changes in leaves of Arabidopsis thaliana. The New phytologist 172: 92–103. PMID: 16945092

42. Lapik YR, Kaufman LS (2003) The Arabidopsis cupin domain protein AtPirin1 interacts with the G protein alpha-subunit GPA1 and regulates seed germination and early seedling development. The Plant cell 15: 1578–1590. PMID: 12837948

43. Porankiewicz J, Wang J, Clarke AK (1999) New insights into the ATP-dependent Clp protease: Escherichia coli and beyond. Molecular microbiology 32: 449–458. PMID: 10320569

44. Kress W, Maglica Z, Weber-Ban E (2009) Clp chaperone-proteases: structure and function. Research in microbiology 160: 618–628. doi: 10.1016/j.resmic.2009.08.006 PMID: 19732826

45. Foreman J, Demidchik V, Bothwell JH, Mylona P, Miedema H, Torres MA, et al. (2003) Reactive oxygen species produced by NADPH oxidase regulate plant cell growth. Nature 422: 442–446. PMID: 12660786

46. Marx C, Wong-JH, Buchanan BB (2003) Thioredoxin and germinating barley: targets and protein redox changes. Planta 216: 454–460. PMID: 12520337

47. Patnaik D, Khurana P (2001) Germins and germin like proteins: an overview. Indian journal of experimental biology 39: 191–200. PMID: 11495276

48. Mazzitelli L, Hancock RD, Haupt S, Walker PG, Pont SD, McNicol J, et al. (2007) Co-ordinated gene expression during phases of dormancy release in raspberry (Rubus idaeus L.) buds. Journal of experimental botany 58: 1035–1045. PMID: 17244630

49. Jones AM, MacLean D, Studholme DJ, Serna-Sanz A, Andreasson E, Rathjen JP, et al. (2009) Phosphoproteomic analysis of nuclei-enriched fractions from Arabidopsis thaliana. Journal of proteomics 72: 439–451. doi: 10.1016/j.jprot.2009.02.004 PMID: 19245862

50. Wan L, Ross AR, Yang J, Hegedus DD, Kermode AR (2007) Phosphorylation of the 12 S globulin cruciferin in wild-type and abi1-1 mutant Arabidopsis thaliana (thale cress) seeds. The Biochemical journal 404: 247–256. PMID: 17313365

51. Eastmond PJ, Graham IA (2001) Re-examining the role of the glyoxylate cycle in oilseeds. Trends in plant science 6: 72–78. PMID: 11173291