Journal of proteome research

Article

pubs.acs.org/jpr

Isotope Labeling-Based Quantitative Proteomics of Developing Seeds of Castor Oil Seed (*Ricinus communis* L.)

Fábio C. S. Nogueira,[†] Giuseppe Palmisano,[‡,§] Veit Schwämmle,[‡] Emanuela L. Soares,[⊥] Arlete A Soares,[∥] Peter Roepstorff,[‡] Gilberto B. Domont,*[,†] and Francisco A. P. Campos*[,⊥]

[†]Proteomic Unit, Institute of Chemistry, Universidade Federal do Rio de Janeiro, Av. Athos da Silveira Ramos, 149 - CT-Bloco A, Lab 543, Rio de Janeiro 21941-909, Brazil

[‡]Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense, Denmark

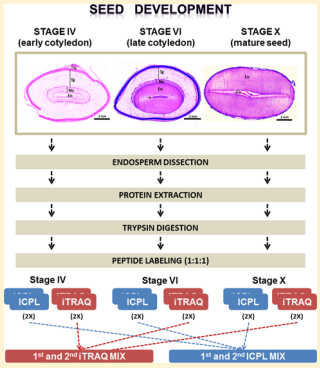
[§]Departamento de Parasitologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, Av. Prof. Lineu Prestes, 1374 - Edifício Biomédicas II, Cidade Universitária "Armando Salles Oliveira", 05508-000 São Paulo, Brazil

[∥]Department of Biology, Universidade Federal do Ceará, Campus do Pici - Bloco 906, 60020-181 Fortaleza, Brazil

[⊥]Department of Biochemistry and Molecular Biology, Universidade Federal do Ceará, Campus do Pici - Bloco 907, 60020-181 Fortaleza, Brazil

Supporting Information

ABSTRACT: In this study, we used a mass spectrometry-based quantification approach employing isotopic (ICPL) and isobaric (iTRAQ) labeling to investigate the pattern of protein deposition during castor oil seed (*Ricinus communis* L.) development, including that of proteins involved in fatty acid metabolism, seed-storage proteins (SSPs), toxins, and allergens. Additionally, we have used off-line hydrophilic interaction chromatography (HILIC) as a step of peptide fractionation preceding the reverse-phase nanoLC coupled to a LTQ Orbitrap. We were able to identify a total of 1875 proteins, and from these 1748 could be mapped to extant castor gene models, considerably expanding the number of proteins so far identified from developing castor seeds. Cluster validation and statistical analysis resulted in 975 protein trend patterns and the relative abundance of 618 proteins. The results presented in this work give important insights into certain aspects of the biology of castor oil seed development such as carbon flow, anabolism, and catabolism of fatty acid and the pattern of deposition of SSPs, toxins, and allergens such as ricin and 2S albumins. We also found, for the first time, some genes of SSP that are differentially expressed during seed development.



KEYWORDS: *ICPL, iTRAQ, quantitative plant proteomics, Ricinus communis, seed development*

INTRODUCTION

Castor (*Ricinus communis L.*) is cultivated as the source of a hydroxylated fatty acid, ricinoleic acid, which has several industrial and pharmaceutical applications.[1] Currently, it is considered to be an as yet untapped source of feedstock for biodiesel production, especially in countries such as Brazil in which several federally funded programs are in place to exploit the production of ethanol and biodiesel from plant sources.[2] The exploitation of castor seeds' potential as a biodiesel feedstock is hampered by a lack of knowledge regarding the molecular and biochemical mechanisms related to the synthesis of fatty acids, whose understanding would allow the tailoring of fatty acid composition in the seeds to specific applications,[3] and by the presence of a highly toxic ribosome-inactivating protein called ricin and allergenic 2S albumins, which together preclude the use of the protein meal obtained after oil extraction as animal feed[1] and represent occupational and environmental hazards.[4]

The publication of the draft genome sequence of castor with a 4.6-fold coverage and more than 31 thousand gene models derived from it[5] created an important database to foster gene expression and protein identification studies so that the questions related to fatty acid biosynthesis and the toxicity/allergenicity issues could be addressed.[6] Following this, two major proteomic analysis of castor seeds were published, focusing on the nucellus of developing seeds[7] and on an enriched fraction of reticulum endoplasmic from developing seeds,[8] which resulted in the identification, respectively, of 542 and 792 proteins. These studies were preceded by the publication of the first systematic quantitative proteomics study based on 2-D gel electrophoresis of seed filling in castor, which resulted in the identification of 304 proteins and in the expression profiles for 660 protein spot groups.[9] More recently, we used LC−MS to compare isotopic (ICPL) and isobaric (iTRAQ) chemical-labeling techniques to quantify proteins in the endosperm of castor seeds in three developmental stages. By comparing labeling performance, it was shown that iTRAQ

Special Issue: Agricultural and Environmental Proteomics

Received: July 4, 2013
Published: October 3, 2013

   dx.doi.org/10.1021/pr400685z | *J. Proteome Res.* 2013, 12, 5012−5024

was able to label 99.8% of all identified unique peptides, while 94.1% were labeled by ICPL, thus showing that both labels were able to quantify successfully proteins present in the endosperm of developing castor seeds.[10] This study concluded that although the use of iTRAQ and ICPL as labeling reagents allows quantification of a comparable number of proteins, the two methods when used in combination result in an increased number of quantified proteins.[10]

In the study presented here, we used ICPL and iTRAQ labeling strategies to undertake an in-depth quantitative analysis of the proteome of developing castor seeds and explored the usefulness of subjecting the peptides labeled with ICPL and iTRAQ to fractionation by hydrophilic interaction liquid chromatography (HILIC) prior to RP-HPLC to increase the number of identified proteins. This led us to identify 1845 proteins, from which 618 were statistically evaluated. Additionally through fuzzy c-means clustering we grouped 975 of the identified proteins in three clusters, each of which reflected the trend in deposition pattern of the proteins during seed development. Besides considerably expanding the repertoire of castor proteins identified and quantified, our results give important insights into certain aspects of the biology of castor seed development such as carbon flow, anabolism, and catabolism of fatty acid and the pattern of deposition of seed-storage proteins (SSPs), toxins, and allergens such as ricin and 2S albumins.

EXPERIMENTAL SECTION

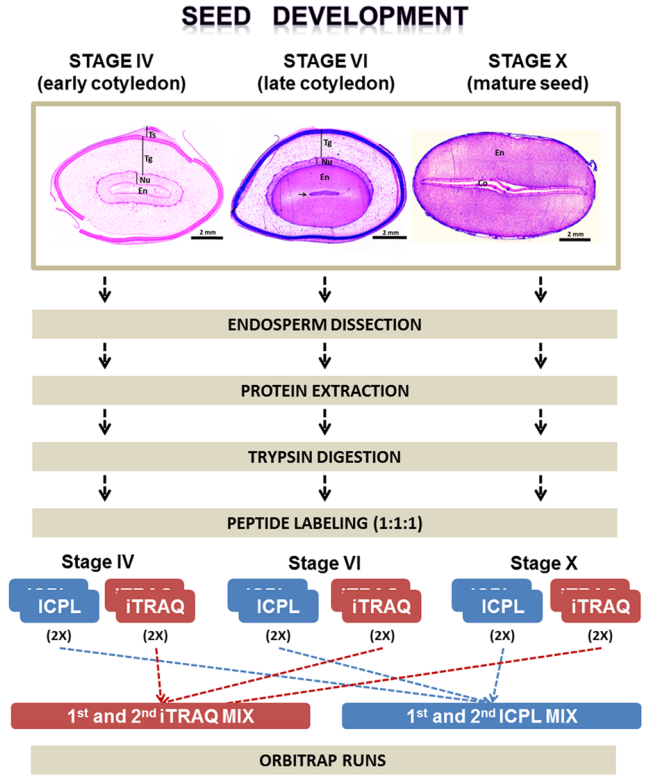
An additional description of the sample preparation, peptide labeling, mass spectrometry, and data analysis can be found in Nogueira et al. (2012).[10]

Sample Preparation

Plant materials were harvested as previously described in ref 10, and the endosperms were dissected from seeds at developmental Stages IV, VI, and X as defined by Greenwood and Bewley.[11] Protein extraction from endosperm was performed accordingly with Nogueira et al.[12] In brief, pyridine buffer (50 mM pyridine, 10 mM thiourea, and 1% SDS, pH 5.0) and polyvinyl-polypyrrolidone were added to endosperm powders in the proportion of 1:40:2 (w/v/w), the mixture was stirred for 2 h at 4 °C, and the proteins were precipitated by acetone/TCA precipitation, followed by washing with cold acetone (three times). Protein pellets were dried under vacuum and dissolved in 7 M urea, 2 M thiourea, and 200 mM TEAB. Protein amounts were measured in duplicate by amino acid analysis using a Biochrom 30+ amino acid analyzer (Biochrom) following the manufacturer's instructions. Proteins were submitted to in-solution trypsin digestion (Promega) (1:50 w/w) after nine times dilution with 200 mM TEAB reduction followed by alkylation with 10 mM DTT (1 h, 25 °C) and 40 mM IAA (40 min, room temperature in the dark), respectively. Digestion was stopped by adding formic acid to a final concentration of 2%, the peptides were cleaned using Oasis cartridges (Millipore), and their concentrations were measured using a fluorometric assay (Qubit, Invitrogen).

Peptide Labeling

ICPL peptide labeling was performed in duplicate using the ICPL Quadruplex-Kit (SERVA) following the manufacturer's instructions, adjusted to the "post-digest optimized protocol", and for iTRAQ peptide labeling, it was accomplished in duplicate following the manufacturer's recommendations (Figure 1). In brief, for ICPL labeling, 33 μg of peptides of



1ST iTRAQ		2ND iTRAQ	
Orbi Run	Replicate n.	Orbi Run	Replicate n.
RUN 1	1	RUN 1	5
RUN 2	2	RUN 2	6
RUN 3	3	RUN 3	7
RUN 4	4	RUN 4	8

1ST ICPL		2ND ICPL	
Orbi Run	Replicate n.	Orbi Run	Replicate n.
RUN 1	9	RUN 1	13
RUN 2	10	RUN 2	14
RUN 3	11	RUN 3	15
RUN 4	12	RUN 4	16

2nd iTRAQ HILIC FRACTIONATION		
Orbi Run	Fractions	Replicate n.
RUN 1	11	HILIC1

2nd ICPL HILIC FRACTIONATION		
Orbi Run	Fractions	Replicate n.
RUN 1	17	HILIC2

Figure 1. Quantitative proteomic workflow used for the analysis of *Ricinus communis* seed development. Histological analysis of the developing castor seeds is shown at the top. En, endosperm; Nu, nucellus; Tg, tegmen; Ts, testa; and Co, cotyledon. After sample preparation, peptides from each stage were labeled with ICPL or iTRAQ. A total of four labeling replicates were analyzed through 16 Orbitrap runs. The second iTRAQ and ICPL labeling were fractionated off-line by hydrophilic chromatography prior to mass spectrometric analysis.

each developmental stage were dissolved in 30 μL of 20 mM TEAB, pH 8.5 and submitted for labeling using 3 μL of ICPL-labeling reagent for 90 min at room temperature under an argon atmosphere. Another 1.5 μL of aliquot of reagent was then added and incubated for 90 min at room temperature under an argon atmosphere. The reaction was stopped using the stop solution (20 min), and samples were mixed and neutralized with 2 N NaOH (20 min), followed by 2 N HCl (20 min). Labeled peptides mixtures were purified by custom-made chromatographic Poros 50 R2 (PerSeptive Biosystems) and Poros Oligo R3 (Applied Biosystems) reverse-phase microcolumns prior to MS analysis.[13] For iTRAQ labeling, tryptic peptides were dissolved in 30 μL of 20 mM TEAB, pH 8.5. To each iTRAQ 4-plex reagent vial, 70 μL of ethanol was added, and the mixture was combined with the peptide sample

dx.doi.org/10.1021/pr400685z | J. Proteome Res. 2013, 12, 5012−5024

and incubated at room temperature for 1 h. After incubation, samples were acidified using formic acid and dried. Peptide mixtures were purified using a TSK Amide-80 HILIC microHPLC system.[14]

Fractionation of Labeled Samples by HILIC

The second mixture of peptides labeled with ICPL and iTRAQ reagents (~33 μg) was fractionated by TSK amide-80 HILIC microHPLC column (length: 15 cm, diameter: 2 mm, particle size: 3 μm) prior to LC−MS/MS analysis, as previously described.[10,14,15] The peptides were resuspended in 90% ACN and 0.1% TFA, eluted, and fractionated with an increase in aqueous solvent (0.1% TFA) during the gradient at 6 μL/min that led to the obtaining of 17 and 11 fractions for ICPL- and iTRAQ-labeled sample mixture, respectively (Figure 1).

Mass Spectrometry Analysis

Labeled peptide mixtures were dissolved in 0.1% formic acid, and 1 μg of them was loaded onto a C_{18} reverse-phase pulled needle capillary column (18 cm length, 100 μm inner diameters, packed in-house with ReproSil-Pur C_{18}-AQ 3 μm resin). The fractions obtained by HILIC were resuspended in 0.1% formic acid and loaded onto a C_{18} reverse-phase pulled needle capillary column as well. The samples were analyzed by an EASY-nano LC system (Proxeon Biosystems) coupled online to an ESI-LTQ-Orbitrap XL mass spectrometer (Thermo Scientific). Elution was performed using a gradient from 100% phase A (0.1% formic acid) to 35% phase B for 120 min (for HILIC fractions was used 80 min), 35−100% phase B for 5 min, and 100% B for 8 min (flow rate of 250 nL/min). The column was washed with 90% phase B and re-equilibrated with phase A. Spectra were acquired in a positive mode, applying DDA automatic survey MS scan and tandem mass spectra (MS/MS) acquisition. An MS scan (400−1800 m/z) in the Orbitrap mass analyzer set at resolution 60,000 at 400 m/z, 1×10^6 AGC target, and 500 ms maximum ion injection time was followed by MS/MS of the top five multiply charged ions in the LTQ at 30 000 signal threshold, 20 000 gain control target, 300 ms maximum ion injection time, 2.5 m/z isolation width, 30 ms activation time at 35 normalized collision energy, and dynamic exclusion enabled for 30 s with a repeat count of 1 (applied in the ICPL samples). For iTRAQ samples, data-dependent CID MS/MS analysis of the three most intense ions was performed in the LTQ, followed by HCD MS/MS analysis of the corresponding ions with detection in the Orbitrap. MS/MS for acquiring CID was set as described, and MS/MS for acquiring HCD was set at 7500 at 400 m/z resolution, 30 000 signal threshold, 5 ms activation time at 48 normalized collision energy, and dynamic exclusion enabled for 30 s with a repeat count of 1. A 5% ammonia−water solution in a 500 μL Eppendorf tube with holes in the cover was put under the pulled needle to avoid the supercharged effect of the iTRAQ 4plex tag, a phenomenon previously described.[16]

Data Analysis

Raw data were viewed in Xcalibur v.2.1 (Thermo Scientific), and data processing was performed using Proteome Discoverer v.1.2. Database searches were performed using Proteome Discoverer with Mascot v.2.3 algorithm against a concatenated *Ricinus communis* database downloaded from Uniprot database in January 2011. The searches were performed with the following parameters: MS accuracy 10 ppm, MS/MS accuracy 0.6 Da for CID and 0.1 Da for HCD, trypsin digestion with two missed cleavage allowed, fixed carbamidomethyl modification of cysteine and variable modification of oxidized methionine and ICPL (label 0, monoisotopic mass = 105.02), ICPL:13C(6) (label 6, monoisotopic mass = 111.04), and ICPL:13C(6)-2H(4) (label 10, monoisotopic mass = 115.06), and iTRAQ 4plex (monoisotopic mass = 144.102) for N-terminus and Lys was set as variable modification for ICPL and iTRAQ labeling, respectively. The number of proteins and protein groups and the numbers of peptides and the quantitative values for each label were estimated using Proteome Discoverer. False discovery rates of 1% and a peptide rank of 1 were applied as cutoff limits.

The experiment consists of a total of 16 mass spectrometric runs, four replicates forward "Stage IV, Stage VI and Stage X" (1st iTRAQ and ICPL labeling) and four replicates in the reverse "Stage X, Stage IV, Stage VI" (2nd iTRAQ and ICPL labeling) for each labeling method (Figure 1), as described in the previous article.[10] Additionally, HILIC fractionation was performed for each reverse labeling method, generating 11 and 17 fractions for the second iTRAQ and ICPL labeling, respectively (Figure 1). All data were merged and analyzed using Protein Center (Thermo Scientific), from where the number of identified proteins was estimated.

Statistical Analysis

The data consist of four labeling replicates (first and second iTRAQ and first and second ICPL) coming both from different labeling methods and fractionation techniques; see Figure 1. For each replicate and each fractionation, we normalized each of the three conditions IV, VI, and X by subtracting the median from log 2-transformed reporter ion intensities. HILIC fractionation lead to a considerably greater number of peptide hits. Data of the same labeling replicate were merged. Relative protein abundances were obtained by scaling, filtering, and averaging the corresponding peptide intensities using the RRollup function of the DanteR package.[17] We used the option "mean" instead of the default "median" and modified the function to require a minimum of three peptides measurements per protein. The PCA plots (Figure 3) show that the conditions IV, VI, and X separate well even when using different experimental techniques.

Detection of significantly differently expressed proteins was carried out on log 2 values of the ratios between conditions VI versus IV and X versus IV. Because the two replicates coming from HILIC fractionation yielded a much larger number of proteins, we applied a method suitable for data sets with many missing values[18] that combines the usage of the limma package[19] and a modified version of rank products.[20,21] Proteins with q values (p values corrected for multiple testing according to Storey[22]) below 0.01 for at least one of both statistical tests were considered regulated. Furthermore, fuzzy c-means clustering[23] was used to detect common patterns within all proteins. The required parameters were determined according to Schwämmle,[24] yielding 3.78 for the fuzzifier and three clusters. We considered a protein to be member of a cluster if the corresponding membership value was larger than 0.5.

RESULTS AND DISCUSSION

The choice of the developmental stages of castor seeds to be analyzed was based on early work by Greenwood and Bewley,[11] which by following strict morphological descriptors, defined the developmental stages of the seeds. Stages IV, VI, and X (Figure 1) represent stages in which cellularization of the endosperm is

dx.doi.org/10.1021/pr400685z | J. Proteome Res. 2013, 12, 5012−5024

A

STRATEGY	PEPTIDES	Unique PEPTIDES	PROTEINS
Without HILIC	47,683	3,254	1,142
With HILIC	12,164	7,404	1,364
TOTAL	59,847	7,596	1,845

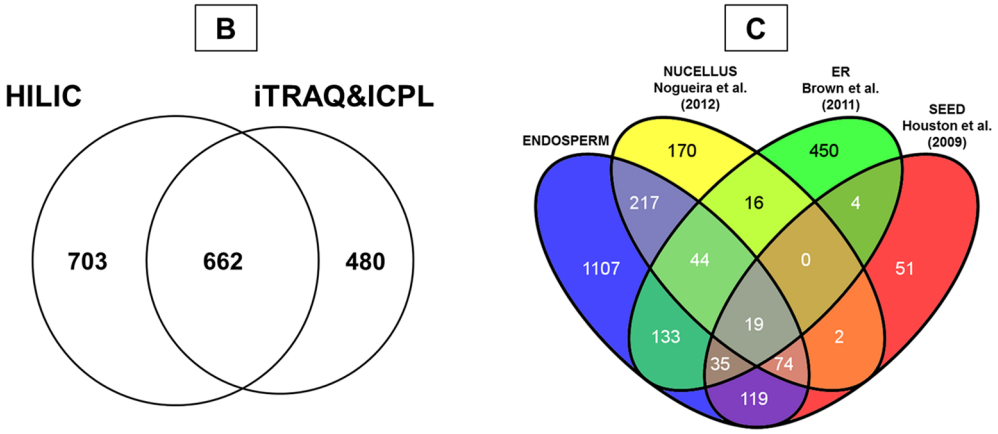


Figure 2. (A) Number of peptides, unique peptides and proteins identified using different strategies. (B) Venn diagram shows the identified proteins using off-line hydrophilic chromatography fractionation (HILIC) or not (iTRAQ and ICPL). (C) Comparative analysis of the proteins mapped to castor gene models[5] identified in this work (ENDOSPERM) compared with three previous works – NUCELLUS, Endoplasmic Reticulum (ER), and Seed Development (SEED).

proceeding in a wave-like manner to the chalazal end of the seed (Stage IV), cellularization is complete and the endosperm occupies most of the seed space (Stage VI), and endosperm is fully mature and the seed is in a desiccated state (Stage X). The endosperm tissues isolated from seeds in these developmental stages are therefore in the initial, intermediary, and final stages of deposition of reserves, as evidenced by previous studies.[9,25,26] The labeling with ICPL and iTRAQ of the peptides obtained by in-solution digestion with trypsin of the protein extracts led to the identification of 1142 proteins (corresponding to 47 683 peptides and 3254 unique peptides) (Figure 2A,B, Table S1 in the Supporting Information), but the number of identifications rose to 1365 (corresponding to 12 164 peptides and 7404 unique peptides) when the samples labeled with iTRAQ and ICPL were fractionated by HILIC prior to separation by RP-HPLC (Figure 2A,B, Table S1 in the Supporting Information). Even though 662 (36%) identified proteins were common to both identification strategies, a sizable number were identified only by ICPL and iTRAQ (480 or 26%) or by HILIC (703 or 38%), showing that the additional step of fractionation yielded considerable improvement in the number of proteins and unique peptides identified (Figure 2A,B). Altogether, a total of 1845 proteins (corresponding to 59 847 peptides and 7596 unique peptides) were identified (Figure 2A,B, Table S1 in the Supporting Information), and from these 1748 could be mapped to extant castor gene models,[5] thus considerably expanding the repertoire of castor proteins so far identified.[7−9,27]

According to the rules we established for normalization and measurement of relative protein abundances between conditions (Stages VI, IV, and X) (see Figure 3 and Experimental Section for details), we determined the relative levels of 1663, 1184, and 980 proteins present in two or more labeling replicates and with at least one, two, or three unique peptides (Table S2 in the Supporting Information). For the following analysis, we have decided to use a stringent criterion by: filtering for proteins with at least three unique peptides. The conditions were well-distinguishable on a PCA plot compared with variance coming from different labeling methods (Figure 3A). By combining the usage of two statistical tests (see refs 19−21), better suitable than a standard t test for data sets with few replicates and missing values, we were able to statistically evaluate the 618 proteins and identified 459 differentially regulated proteins using a threshold of 0.05 for the q values (Table S3 in the Supporting Information). Additionally, through fuzzy c-means clustering (see Experimental Section for details), 975 proteins with three unique peptides and present in at least one replicate could be grouped into three clusters, each of which reflecting the trend in deposition pattern of the proteins during seed development (Figure 4, Table S4 in the Supporting Information). Proteins grouped in Clusters I and III have deposition levels at Stage X that are lower than in Stage IV including all of the less-abundant proteins shown in Table S3 in the Supporting Information. Most of the proteins involved in the primary metabolism are grouped into these clusters, as, for example, the ribosomal proteins, elongation factors, aminoacyl-tRNA-synthetases, proteasome components, chaperones, and proteins involved in amino acid metabolism. These two clusters, besides the proteins involved in secondary metabolism, also contain the enzymes involved in the pathways related to the biosynthesis and catabolism of carbohydrates and lipids. Proteins grouped in Cluster II show increasing deposition levels throughout the developmental stages and within this cluster all of the up-regulated proteins in Stage X compared with Stage IV are found shown in Table S3 in the Supporting Information. Cluster II proteins have diverse roles

dx.doi.org/10.1021/pr400685z | J. Proteome Res. 2013, 12, 5012−5024

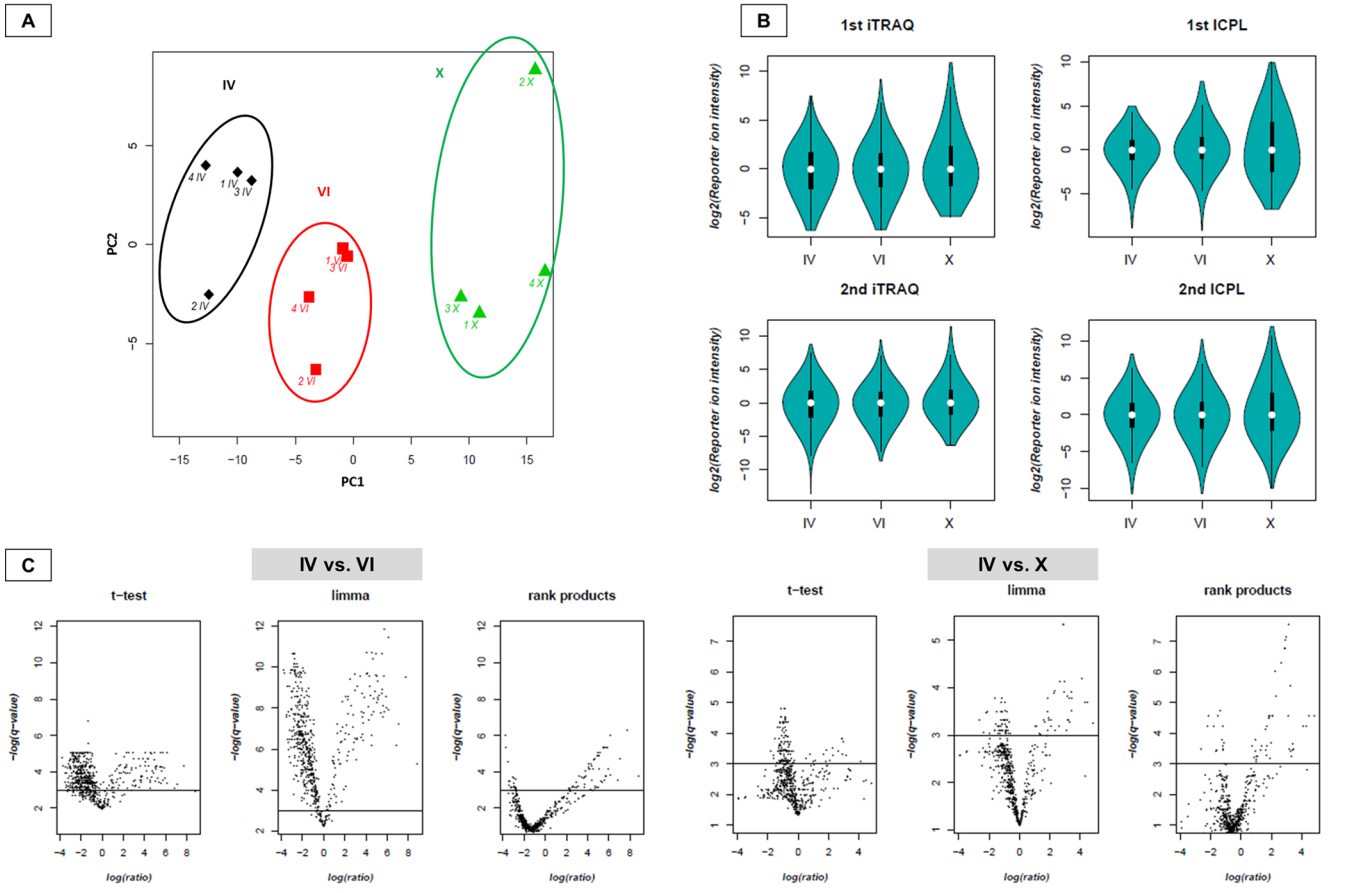


Figure 3. PCA plot showing a well-distinguishable separation of the stages in all labeling replicates (A). Violin plots of merged and normalized technical fractions for each labeling sample (B). Volcano plots showing the distribution for the ratios Stage VI/Stage IV and Stage X/Stage IV against the *q* value of the statistical tests used (*t* test, limma, and rank products) (C).

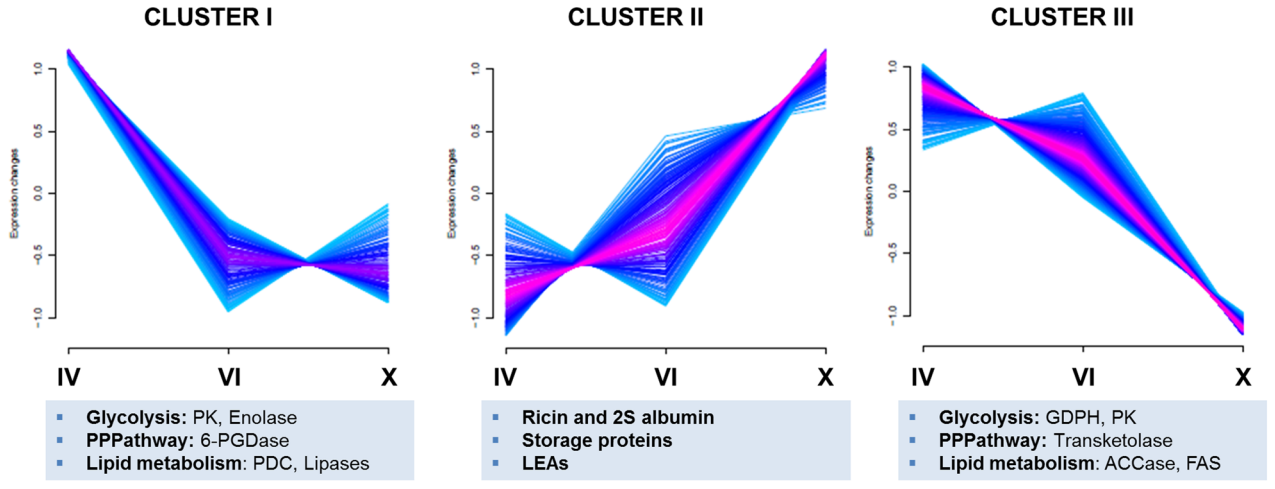


Figure 4. Plots showing the trend patterns of proteins identified in this work that were grouped in three clusters. Classes of the proteins present in each cluster are shown at the bottom. PK − pyruvate kinase, 6-PGDase −6-phosphogluconate dehydrogenase, PDC − pyruvate dehydrogenase, LEA − late embryogenesis abundant protein, GDPH − Glyceraldehyde 3-phosphate dehydrogenase, ACCase − acetyl-CoA carboxylase and FAS − fatty acid synthase.

in the desiccation phase of seed development and in the wide range of reactions that are directed to support the mobilization of seed reserves to the growing embryo. In this cluster, we found the SSP (11S globulins, 2S albumins, etc.), the oleosins that help to stabilize the structure of the oil bodies in which the lipid reserves are accumulated,[28] and an array of peptidases, lipases, and carbohydrases that will be involved in the mobilization of the protein and lipid reserves during germination. Thioredoxin, a protein that due to its ability to reduce the SSP enhances susceptibility to peptidases,[29] is also present, together with many proteins (glutathione peroxidase, glutathione reductase, glutaredoxin, peroxiredoxin, etc.) that

dx.doi.org/10.1021/pr400685z | J. Proteome Res. 2013, 12, 5012−5024

Table 1. Correlation between Ricin, 2S Albumin, and Storage Protein Identifications Using Uniprot Database and the Gene Models Retrieved from Chan et al.[5] [a]

gene model	Uniprot AC	protein description	Cluster	ratio VI/IV	ratio X/IV
Ricin Family					
60637.m00004	B9T7K1	rRNA *N*-glycosidase	II	6.23	25.21
60638.m00023	B9T1S0	rRNA *N*-glycosidase	II	3.98	30.99
60638.m00019	B9T1S2	rRNA *N*-glycosidase	II	1.97	13.11
29791.m000533	B9SSU3	rRNA *N*-glycosidase	II	3.28	44.73
60627.m00002	B9T8T6	ricin homologue	II	8.61	216.96
2S Albumin					
28166.m001081	P01089	2S albumin	II	1.53	5.72
28166.m001074	B9SA29	2S albumin	II	1.94	14.97
28166.m001073	B9SA28	2S albumin	II	4.96	31.28
28166.m001083	B9SA37	2S sulfur-rich seed-storage protein large subunit	II	3.56	10.92
30156.m001719	B9RPN9	2S sulfur-rich seed-storage protein large subunit	II	3.58	3.07
Storage Protein					
30005.m001289	B9SF36	legumin A	II	4.21	117.81
30005.m001288	B9SF35	legumin A	II	7.49	297.17
29611.m000223	B9T1B8	legumin A	II	3.21	26.28
30005.m001290	B9SF37	legumin A	II	1.72	34.36
29200.m000167	B9T5E6	legumin B	II	5.09	19.48
29600.m000564	Q9M4Q8	legumin B	II	7.48	23.89
29600.m000561	B9SDX6	legumin B	II	7.82	17.37
29716.m000305	B9SW16	11S globulin subunit β	II	1.32	23.08
29629.m001355	B9S9Q7	11S globulin subunit β	II	3.40	12.85
29200.m000169	B9T5E7	glutelin type-A 3	II	2.74	23.05
29709.m001187	B9SKF4	nutrient reservoir	II	9.42	12.89
30111.m000716	B9SK34	nutrient reservoir	II	3.28	6.86
29844.m003250	B9RTM9	vicilin GC72-A	II	3.01	7.72

[a]All proteins were identified with a least one unique peptide and clustered. The relative abundance was measured using the Stage IV as reference.

will act to prevent the damage resulting from the oxidative stress associated with the intense catabolic activity during germination. Furthermore, Cluster II contained all enzymes of the glyoxylate cycle that are responsible for the formation of oxaloacetate via acetyl-CoA in the gluconeogenesis pathway.[30]

Ricin Isoforms

The draft of the castor genome revealed 28 putative ricin genes, including potential pseudogenes or gene fragments.[5] Here, using the Uniprot database, we identified 82 ricin entries, corresponding to 15 gene models, including two (29988.m000128 and 29988.m000129) that were thought to be pseudogenes.[5] Five of the proteins coded by these genes were identified with unique peptides, and these isoforms were found to be more abundant throughout seed development and are grouped in Cluster II (Table 1). Previously, we have identified in the nucellus of castor three ricin isoforms,[7] thus raising the number of isoforms that could be mapped to extant gene models to eight. However it should be noted that ricin is known to undergo extensive proteolytic processing, and this may give rise to many peptides not encoded by the genome. Transcriptomic studies[3] revealed the presence of ricin transcripts in leaf and male flowers, thus giving weight to our previous suggestion that ricin action is not circumscribed to the embryo proper but extends to maternal tissues. We find it significant that none of the ricin isoforms we identified in the endosperm were also identified in the nucellus and vice versa, thus indicating that this tissue-specificity may be related to the role/action of a particular isoform within these tissues.

Seed Storage Proteins

Our studies led to the identification of 13 proteins that belong to the two major classes of SSPs from castor: the 11S globulins with 11 identifications and the 7S globulins with 2 identifications (Table 1). The 11S globulin gene family in *R. communis* was previously characterized as having a total of 11 nonallelic genes, including two putative pseudogenes.[31] The high level of sulfur-containing amino acids is a feature distinguishing the 7S globulins from the 11S globulins, and for this reason the 7S globulins are considered to be a specialized reserve for sulfur-containing amino acids to be mobilized to the growing plantlet during germination.[32] Our data show that all of these 13 proteins are expressed in the developing seed in a coordinated manner, and we find it relevant to point out that each of them was identified with unique peptides and that each one could be assigned to a unique annotated castor gene model.[5] Consistent with the role of these proteins as nutrient reserves, the quantitative analysis of the MS data led us to group these proteins in Cluster II (Figure 4). Notwithstanding sequence similarities and coordinated expression, sharp quantitative differences in the deposition pattern of the 11S globulins could be observed, thus suggesting the occurrence of a tight genetic control in the expression of each individual gene. For example, the relative concentration of the legumin A (gene model 30005.m001288) at Stage X is 297 times higher than that at Stage IV, while for another legumin A (gene model 29611.m000223, the concentration ratio is 26 (Table 1). These differences raise the question as to whether there exists a differential mobilization of each individual protein during germination,

dx.doi.org/10.1021/pr400685z | J. Proteome Res. 2013, 12, 5012−5024

Table 2. Correlation between LEAs and Oleosins Protein Identifications Using Uniprot Database and the Gene Models Retrieved from Chan et al.[5a]

gene model	Uniprot AC	protein description	cluster	ratio VI/IV	ratio X/IV
Late Embryogenesis Abundant Protein (LEA)					
29601.m000448	B9SWY7	late seed maturation protein P8B6	II	2.86	44.05
29814.m000729	B9SQP0	embryonic abundant protein	II	3.14	49.43
29889.m003401	B9S010	late embryogenesis abundant	II	1.86	8.05
29647.m002058	B9S3Z7	late embryogenesis abundant protein D-34	II	2.04	8.50
29634.m002070	B9S696	dehydrin xero	II	1.95	32.79
29815.m000495	B9SRL2	late embryogenesis abundant	II	1.25	1.79
30190.m011335	B9RBC1	late embryogenesis abundant	II	1.11	2.08
29841.m002806	B9RV15	late embryogenesis abundant protein D-7	II	1.22	9.12
29844.m003281	B9RTR0	late embryogenesis abundant protein D-34	II	1.99	3.47
30131.m007082	B9RH90	late embryogenesis abundant protein	II	1.71	2.58
30174.m008670	B9RDY8	late embryogenesis abundant protein Lea14-A			
30128.m008646	B9REU2	late embryogenesis abundant protein D-34			
29647.m002057	B9S3Z6	late embryogenesis abundant protein D-34			
29634.m002127	B9S6F3	late embryogenesis abundant			
28266.m000192	B9T526	late embryogenesis abundant protein Lea14-A			
Oleosin					
29794.m003372	B9RRX4	oleosin	II	0.97	4.90
29917.m001992	Q5VKJ9	oleosin (oleosin1)	II	1.66	2.44
30147.m014333	Q5VKJ8	oleosin	II	1.90	2.65
30147.m013891	B9RAW7	oleosin 18.2 kDa	II	0.49	4.67
30174.m008728	B9RE45	oleosin	II	0.61	1.62

[a]All proteins listed below were identified with a least one unique peptide, and some of them were clustered. The relative abundance was measured using the Stage IV as reference.

but we considered the answer to this question to be beyond the scope of this work.

The 2S albumins comprise another family of SSPs from which we have identified five members (Table 1). These proteins are relevant not only for their role as a nutrient reservoir during seed germination but also because they are known to be powerful allergens, representing both a potent occupational sensitizer and an environmental allergen.[33,34] As demonstrated by our quantitative studies, the deposition of these five proteins is highly coordinated, and consistent with their role as storage proteins they were grouped in Cluster II (Figure 4, Table 1). As seen with the 11S and 7S globulins, our quantitative studies also pointed out significant differences in the deposition pattern of the 2S albumins (Table 1).

In addition to the SSP, we identified a plethora of proteins that are known to be related to the desiccation phase of seed development (Table 2) and that are thought to be useful markers of seed maturation.[35] As seen with the storage proteins, these proteins could be grouped in Cluster II, and a closer analysis of their deposition patterns indicates substantial differences in the kinetics of deposition among the different proteins. For example, while for the late maturation protein (29601.m000448), dehydrin (29634.m002070), and the embryonic abundant protein (29814.m000729) the relative concentration at Stage X is, respectively, 44, 49, and 32 times higher, for three other late embryogenesis abundant proteins (29815.m000495, 30190.m011335, and 30131.m007082), these values range from 1.8 to 2.5, thus suggesting that these proteins may also play a role in earlier stages of seed development.

The oleosins are structural proteins that play a crucial role in oil seeds by preventing the coalescence of oleosomes during the desiccation phase of seed development, also playing a role as a binding site for lipases during germination.[28] Here we could identify five oleosin isoforms (Table 2) and could assign to each a unique gene model in the castor genome.[5] As expected by their involvement in the package of triacylglycerols into oleosomes, our quantitative analysis indicated a steady increase in its deposition during seed development (Table 2), and consistent with this they were grouped in Cluster II (Figure 4). Recently, Hyun et al.[36] performed a computational identification and phylogenetic analysis of oil-body structural proteins and have found a total of six putative oleosin genes; here we have identified the products of four of these genes. These authors also identified two putative caleosin genes, one of which was shown to be expressed only in seeds (30008.m000820), while the other (29673.m000932) was expressed in seeds and young leaves. Here we identified only the protein product corresponding to the seed specific gene.

Carbon Flow

Carbon assimilation during seed development begins with the transport of sucrose into seeds.[37] In line with the crucial role of sucrose synthase (SUS) in this process, we identified four SUS isoforms, which are predicted to be cytosolic (Table S1 in the Supporting Information). Two of these isoforms (B9RT94 and B9T3H2) were identified with only one peptide and therefore could not be quantified. Isoforms B9SAU6 and B9RR41were grouped in Clusters I and III, respectively (Figure 4, Table 3), indicating that these two isoforms may have distinct roles both in embryogenesis proper and in seed filling, as previously observed in developing flax seeds.[38] We were unable to identify any invertase isoform, even though 14 invertase gene models in castor were proposed by Chan et al.,[5] thus giving weight to the notion that SUS plays a predominant role vis-à-vis invertase in the entry of carbon into metabolism in nonphotosynthetic cells.[39]

Our data highlight that in developing castor seeds the availability of carbon compounds to be used for the synthesis of fatty acids in plastids is guaranteed by the operation of the

Table 3. Correlation between Identified Proteins Related to Carbon Flow (Glycolysis and Pentose Phosphate Pathway) Using Uniprot Database and the Gene Models Retrieved from Chan et al.[5] [a]

path no.	gene model	Uniprot AC	protein description	cluster	ratio VI/IV	ratio X/IV	TargetP
1	29739.m003693	B9RR41	sucrose synthase	III	0.59	0.31	
1	29726.m004086	B9RT94	sucrose synthase				
1	29986.m001646	B9SAU6	sucrose synthase	I	0.37	0.41	
1	29951.m000143	B9T3H2	sucrose synthase				
2	29630.m000793	B9SJB6	phosphorylase				chloro
3	30174.m008935	B9RDH7	fructokinase				chloro
3	30204.m001759	B9S525	fructokinase				secret
3	28609.m000204	B9T2J6	fructokinase				
3	29569.m000151	B9T544	fructokinase	I	0.43	0.38	secret
4	30147.m014138	B9R9J6	phosphoglucomutase				chloro
4	29692.m000529	B9SP64	phosphoglucomutase	I	0.63	0.66	
4	27516.m000169	B9T3D2	phosphoglucomutase	II	0.85	1.38	mito
5	30170.m014025	B9R7S8	glucose-6-phosphate isomerase	I	0.91	0.35	
5	29908.m006236	B9RJU9	glucose-6-phosphate isomerase	III	0.49	0.90	chloro
6	30170.m013807	B9R8U5	phosphofructokinase	I	0.43	0.46	
6	29842.m003527	B9RXE7	phosphofructokinase	III	0.72	0.27	chloro
7	30131.m007128	B9RHD4	fructose-bisphosphate aldolase				chloro
7	30154.m001149	B9S0W4	fructose-bisphosphate aldolase				
7	29660.m000779	B9SJY9	fructose-bisphosphate aldolase	III	0.63	0.20	chloro
7	29092.m000460	B9SRH4	fructose-bisphosphate aldolase	III	0.82	0.29	
7	28623.m000391	B9SZT9	fructose-bisphosphate aldolase				chloro
7	29988.m000122	B9T5T6	fructose-bisphosphate aldolase	III	0.67	0.15	
8	29711.m000323	B9STC9	triosephosphate isomerase	I	0.71	0.91	chloro
8	30028.m000254	B9SZM2	triosephosphate isomerase	III	0.79	0.60	
8	27383.m000156	B9T4H8	triosephosphate isomerase	III	0.93	0.53	
8	32915.m000014	B9TBD2	triosephosphate isomerase	II	0.87	1.13	
9	30147.m013783	B9RAL0	glyceraldehyde 3-phosphate dehydrogenase	III	0.75	0.38	mito
9	30147.m013787	B9RAL3	glyceraldehyde 3-phosphate dehydrogenase				mito
9	30190.m010986	B9RBN8	glyceraldehyde 3-phosphate dehydrogenase	III	0.73	0.18	
9	30169.m006270	B9RHV9	glyceraldehyde 3-phosphate dehydrogenase	I	0.02	0.14	chloro
9	47543.m000013	B9TBB7	glyceraldehyde 3-phosphate dehydrogenase				
10	30169.m006296	B9RHY3	phosphoglycerate kinase	I	0.25	0.32	
10	30169.m006297	B9RHY4	phosphoglycerate kinase	I	0.26	0.43	chloro
10	29950.m001117	B9SHE4	phosphoglycerate kinase				secret
11	30073.m002196	B9S1V6	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	III	0.82	0.38	
11	30073.m002196	P35493	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	III	0.59	0.26	
12	30147.m014180	B9R9N6	enolase	III	0.56	0.14	
12	30174.m009039	B9RE72	enolase	III	0.61	0.30	chloro
12	29751.m001807	B9S376	enolase				
12	29092.m000447	B9SRG1	enolase				
12	29092.m000447	P42896	enolase	III	1.02	0.43	
13	30131.m07299	B9RGK5	pyruvate kinase	I	0.48	0.44	
13	30169.m006558	B9RIP4	pyruvate kinase	I	0.47	0.34	chloro
13	29827.m002563	B9S1I3	pyruvate kinase	III	0.41	0.21	chloro
13	30099.m001688	B9S7Y5	pyruvate kinase				chloro
13	28842.m000927	B9SBM7	pyruvate kinase				
13	29950.m001161	B9SHI6	pyruvate kinase				chloro
13	29815.m000503	B9SRM0	pyruvate kinase	III	0.43	0.20	chloro
13	28448.m000357	B9ST42	pyruvate kinase				
13	27428.m000106	B9T6R6	pyruvate kinase				
13	29827.m002563	P55964	pyruvate kinase isozyme G, chloroplastic	I	0.34	0.17	
13	29844.m003195	Q43117	pyruvate kinase isozyme A, chloroplastic				chloro
13	29844.m003195	Q43117−2	pyruvate kinase isozyme A, chloroplastic	III	0.62	0.25	chloro
14	29929.m004613	B9RMA8	glucose-6-phosphate 1-dehydrogenase				chloro
14	27721.m000026	B9T8C1	glucose-6-phosphate 1-dehydrogenase				secret
15	30213.m000666	B9RWU5	6-phosphogluconolactonase				
16	30190.m010775	B9RCL8	6-phosphogluconate dehydrogenase, decarboxylating	I	0.45	0.42	secret
16	29841.m002898	B9RVA7	6-phosphogluconate dehydrogenase, decarboxylating				secret
16	29832.m000313	B9SXT4	6-phosphogluconate dehydrogenase, decarboxylating	I	0.70	0.59	
17			phoshopentose isomerase				

dx.doi.org/10.1021/pr400685z | J. Proteome Res. 2013, 12, 5012−5024

Table 3. continued

path no.	gene model	Uniprot AC	protein description	cluster	ratio VI/IV	ratio X/IV	TargetP
18			ribose 5P epimerase				
19	30174.m008857	B9RDA1	transketolase	III	0.58	0.16	chloro
19	28507.m000160	B9T0B2	transketolase				chloro
20	30128.m008566	B9RG09	transaldolase	III	0.62	0.15	chloro
20	28200.m000195	B9T2V8	transaldolase total2	III	0.37	0.18	chloro
21	30170.m014025	B9R7S8	glucose-6-phosphate isomerase	I	0.91	0.35	
21	29908.m006236	B9RJU9	glucose-6-phosphate isomerase	III	0.49	0.90	chloro

[a]All proteins were identified with at least one unique peptide, and some of them were clustered. The relative abundance was measured using the Stage IV as reference. The pathway number (path no.) is correlated with the metabolic pathway. Protein localization was performed using protein-targeting prediction program − TargetP.

glycolytic and pentose phosphate pathways both in the plastid and in the cytoplasm (Table 3). With the exception of phosphoglyceromutase (PGM, EC 5.4.2.1) from which only the cytosolic representative was found, we identified all of the cytosolic and plastidic isoforms of the glycolytic enzymes. PGM catalyzes the interconversion of the phosphate group between the C-3 carbon of 3-phosphoglycerate (3-PGA) and the C-2 carbon of 2-phosphoglycerate (2-PGA),[40] and two classes of these enzymes are known to exist: a cofactor dependent requiring 2,3-bisphosphoglycerate for activity (dPGM), which is present in vertebrates, certain fungi, and bacteria, and a cofactor independent (iPGM) that occurs in plants, some invertebrates, fungi, and bacteria.[40] PGM has been purified from castor endosperm[41] and is shown to be an iPGM. These authors confirmed that this isoform is present in the cytosol, being absent in plastids, and they predicted that if a plastid form of PGM is present in castor endosperm it must be structurally unrelated to cytosolic PGM (Huang et al., 1993). As previously mentioned, we identified one iPGM (corresponding to gene model 30073.m002196), which is predicted to be cytoplasm located and has a deposition pattern (Table 3 and Tables S2 and S3 in the Supporting Information) similar to the one determined by Huang et al.[41] The annotation of the castor genome established 2 gene models for iPGM and 10 gene models for dPGM.[5] Here, although we identified one of the two iPGM isoforms, we were unable to identify any of the 10 putative dPGM, but a preliminary proteome analysis of the plastids from developing castor seeds lead us to identify one dPGM isoform (unpublished results). Recently, iPGM isoforms were identified in the proteomes of chloroplast stroma of *Arabidopsis*,[42] in chloroplasts of maize,[43] and in tomato chromoplast,[44] thus suggesting the possibility that in plants iPGM is cytoplasmic while dPGM is plastidial.

In addition to the transport to the plastid of phosphoenolpyruvate (PEP) generated by cytosolic glycolysis, another major route of carbon flow in developing seeds is the conversion of PEP to oxaloacetic acid, followed by its conversion to malate, the import of malate in the plastid and its decarboxylation to pyruvate that is then converted to acetyl-CoA.[38] Here we found evidence of the functioning of this route, as we have identified all of the relevant enzymes as well as an oxoglutarate/malate carrier protein (Table SI in the Supporting Information).

The presence of a third route of carbon flow in developing castor seeds is indicated by the identification of the small and large subunits of Rubisco, even though we did not identify any of the enzymes directly involved in photosynthesis (Table SI in the Supporting Information). When first observed in green oilseeds, the action of Rubisco without the Calvin cycle was considered to be an adaptive feature counterbalancing for the loss of one-third of the carbon as CO_2 as a result of the conversion of carbohydrate to oil through glycolysis:[45] In developing seeds of *Brassica napus*, Rubisco action provides 20% more acetyl-CoA for fatty acid synthesis and results in 40% less loss of carbon as CO_2.[45,46] At first the action of Rubisco apart of the Calvin cycle was thought to occur only in green oilseeds, but subsequent studies with nongreen oilseeds provided evidence of the widespread occurrence of this action.[9,38,47,48] Our results thus support that this route is fully functional in the endosperm of developing castor seeds.

Lipid Metabolism

In oil seeds like castor, fatty acid synthesis occurs in the plastids and requires acetyl-CoA to synthesize mainly oleic and palmitic acid. The first step of this pathway depends on the acetyl-CoA that can be supplied by the plastidial pyruvate dehydrogenase complex (PDC) or acetyl-CoA synthetase, although the PDC provides acetyl-CoA and NADH for de novo fatty acid biosynthesis.[49,50] Here we found plastidial isoforms of all subunits of the PDC (E1 α and E1 β − pyruvate dehydrogenase, E2 − dihydrolipoamide acetyltransferase, and E3 − dihydrolipoyl dehydrogenase) (Table 4). Further steps for fatty acid synthesis require the acetyl-CoA carboxylase (ACCase) and fatty acid synthase (FAS) complexes, and isoforms from both of them were identified in this study (Table 4). Different enzymes are required to assemble the fatty acids in triacylglycerols; the majority of them were identified in a previous study where the endoplasmic reticulum proteome was investigated.[8] In our analysis, we confirmed the identification of some of these enzymes as, for example, glycerol-phosphate acyltransferases (B9T011 and B9T753) and phospholipases (B9RV56) (Table 4).

Besides the identification of a full gamut of proteins related to lipid biosynthesis (Table S1 in the Supporting Information), a wide range of proteins related to lipid catabolism were also identified, including the four enzymes of the β-oxidation pathway, enzymes of the glyoxylate cycle such as isocitrate lyase and malate synthase, several mono- and triacylglycerol lipases, and two phospholipases (Table 4). The mobilization of fatty acids during seed development has been previously studied, especially in the case of developing embryos of *Brassica napus*, where it was shown that at least 10% of the triacylglycerol is lost during the desiccation phase of seed development.[51] Transcriptional studies in developing castor seeds have shown the presence of significant levels of transcripts related to lipid breakdown,[3] and this was followed by the identification of enzymes from the β-oxidation and glyoxylate pathways in the nucellus of developing castor seeds.[7] In this study, the role of several peptidases in programmed cell death (PCD) in the

dx.doi.org/10.1021/pr400685z | J. Proteome Res. 2013, 12, 5012−5024

Table 4. Correlation between Identified Proteins Related to Triacylglycerol Metabolism Using Uniprot Database and the Gene Models Retrieved from Chan et al.[5] [a]

enzyme	gene model	Uniprot AC	protein description	cluster	ratio VI/IV	ratio X/IV	TargetP
PDC	30128.m009036	B9RFW4	pyruvate dehydrogenase	I	0.46	1.05	mito
PDC	30076.m004445	B9RNK3	pyruvate dehydrogenase	I	0.52	0.38	chloro
PDC	29693.m001991	B9S0Z5	pyruvate dehydrogenase	I	0.44	0.36	chloro
PDC	29736.m002097	B9S2H9	pyruvate dehydrogenase	I	0.43	0.64	chloro
PDC	29706.m001326	B9S5V2	dihydrolipoamide acetyltransferase	II	1.17	1.08	
PDC	29728.m000838	B9SLH2	dihydrolipoamide acetyltransferase	III	0.79	0.53	chloro
PDC	28883.m000717	B9ST02	dihydrolipoyllysine-residue acetyltransferase	III	0.59	0.37	chloro
PDC	29889.m003358	B9RZW7	dihydrolipoyl dehydrogenase	II	1.03	1.19	mito
PDC	29889.m003272	B9RZN2	dihydrolipoamide dehydrogenase	III	0.69	0.32	chloro
ACCase	29729.m002301	B9S484	biotin carboxyl carrier protein	III	0.55	0.24	chloro
ACCase	29630.m000809	B9SJD0	biotin carboxyl carrier protein	I	0.21	0.35	chloro
ACCase	30185.m000954	B9S1E2	biotin carboxylase subunit	III	0.42	0.21	chloro
ACCase	30185.m000954	Q41133	biotin carboxylase				mito
ACCase	27798.m000585	B9SPE5	carboxyl-transferase, subunit				chloro
ACCase	28890.m000006	B9TAH3	carboxyl-transferase, subunit				
FAS	28455.m000368	A6N6J4	ketoacyl-ACP synthase III (KAS III)				chloro
FAS	30113.m001448	B9RRJ0	malonyl-CoA: ACP acyltransferase	III	0.50	0.20	chloro
FAS	29693.m002034	Q41135	50 kDa ketoacyl-ACP synthase	II	0.63	1.44	chloro
FAS	29739.m003711	Q41134	chloroplast β-ketoacyl-ACP synthase	III	0.69	0.50	chloro
FAS	29929.m004732	B9RMM7	ketoacyl-ACP reductase (KAR)	III	1.11	0.22	chloro
FAS	29950.m001148	B9SHH3	short-chain type dehydrogenase	III	0.93	0.26	chloro
FAS	30200.m000354	B9SDU7	hydroxyacyl-ACP dehydrase	I	0.71	0.59	chloro
FAS	27843.m000160	B9T5D7	enoyl-ACP reductase	III	0.61	0.41	chloro
FAS	29650.m000277	B9SVD7	short-chain dehydrogenase				chloro
FAS	29929.m004514	B5AY35	acyl-[acyl-carrier-protein] desaturase	III	0.54	0.17	mito
FAS	27985.m000877	B5AY36	acyl-[acyl-carrier-protein] desaturase	III	0.51	0.18	chloro
FAS	29929.m004514	B9RM11	acyl-[acyl-carrier-protein] desaturase	III	0.47	0.23	mito
FAS	29929.m004515	B9RM12	stearoyl-ACP desaturase				mito
FAS	29929.m004515	C0KWD3	stearoyl-acyl-carrier protein desaturase				
FAS	30020.m000203	P22337	acyl-[acyl-carrier-protein] desaturase, chloroplastic	III	0.58	0.25	mito
FAS	30217.m000262	A9XK92	acyl-ACP thioesterase FatA				chloro
FAS	30217.m000262	A9XK92	acyl-ACP thioesterase FatA				chloro
FAS	29841.m002798	B9RV07	acyl-protein thioesterase				mito
FAS	28320.m001083	B9SEM5	acyl-protein thioesterase				
FAS	30066.m000711	B9SLW9	acyl-CoA thioesterase				chloro
ACS	29908.m006186	B9RJ88	long-chain-fatty-acid CoA ligase				secret
ACS	29732.m000322	B9T1U3	long-chain-fatty-acid CoA ligase				
GPAT	27568.m000266	B9T011	ER glycerol-phosphate acyltransferase				
GPAT	28350.m000105	B9T753	ER glycerol-phosphate acyltransferase				mito
LIPASE	29841.m002847	B9RV56	phospholipase D	III	0.69	0.18	
LIPASE	29841.m002847	Q41142	phospholipase D α 1 (PLD 1)	I	0.53	0.40	
LIPASE	30183.m001305	A0T1K1	acidic triacylglycerol lipase 2	II	1.82	14.37	
LIPASE	27660.m000084	B9T6U9	monoglyceride lipase				
LIPASE	29884.m000182	B9T502	triacylglycerol lipase				
LIPASE	29884.m000182	Q5S8F1	lipase				
LIPASE	29935.m000048	Q5VKJ7	lipase	II	1.24	4.93	
ACS	29908.m006186	B9RJ88	long-chain-fatty-acid CoA ligase				secret
ACS	29732.m000322	B9T1U3	long-chain-fatty-acid CoA ligase				
ACOX	29646.m001117	B9SGN6	acyl-coenzyme A oxidase				
AD	29726.m004084	B9RT92	putative uncharacterized protein (acyl-CoA dehydrogenase)				
HAD	29912.m005496	B9RKN5	3-hydroxyacyl-CoA dehyrogenase				
HAD	29726.m004067	B9RT76	3-hydroxyacyl-CoA dehyrogenase	II	0.86	1.55	
HAD	29848.m004538	B9RNV7	enoyl-CoA hydratase				mito
Th	29904.m002984	B9RWL7	3-ketoacyl-CoA thiolase B	I	0.72	0.90	
ACAT	29732.m000314	B9T1T5	acetyl-CoA acetyltransferase	III	0.82	0.47	
ACAT	34989.m000014	B9TF55	acetyl-CoA acetyltransferase				
LIPID BINDING	30128.m008670	B9REW6	acyl carrier protein	III	0.59	0.27	chloro
LIPID BINDING	29726.m003980	B9RSZ1	acyl carrier protein				chloro
LIPID BINDING	30128.m008808	B9RF93	acyl-CoA binding protein				
LIPID BINDING	29827.m002594	B9S1L3	acyl-CoA-binding protein	III	0.72	0.46	

dx.doi.org/10.1021/pr400685z | J. Proteome Res. 2013, 12, 5012−5024

Table 4. continued

enzyme	gene model	Uniprot AC	protein description	cluster	ratio VI/IV	ratio X/IV	TargetP
LIPID BINDING	30190.m011003	B9RBQ4	lipid binding protein				secret
LIPID BINDING	30174.m008919	B9RDG1	calcium lipid binding protein				secret
LIPID BINDING	30169.m006486	B9RIH2	lipid binding protein	III	2.39	0.42	secret
LIPID BINDING	29929.m004644	B9RMD9	lipid binding protein	II	0.87	4.15	secret
LIPID BINDING	28515.m000321	B9SWA9	lipid binding protein	III	0.87	0.66	secret
LIPID BINDING	29653.m000295	B9SXR9	lipid binding protein	III	0.98	0.60	secret
LIPID BINDING	30133.m000237	B9SY04	lipid binding protein				secret

[a]All proteins were identified with a least one unique peptide, and some of them were clustered. Protein localization was performed using protein-targeting prediction program − TargetP.

nucellus of developing castor seeds was discussed as well as the role of PCD in providing reducing power and carbon and nitrogen sources to the developing endosperm. In addition to the breakdown of proteins, the breakdown of fatty acids in cells undergoing PCD could very well fulfill this role not only in the nucellus itself but also in the developing endosperm, as PCD in cell layers surrounding the developing embryo is known to occur. Put together, these observations indicate that lipid turnover may be a normal feature of the development of castor and other oil seeds, and its occurrence has an obvious implication for oil yield and also puts into question the idea that controlling degradation of oil and fatty acids during embryo development could be a useful strategy to improve oil yield in existing crops or in those engineered to produce other industrial fatty acids.[52]

CONCLUSIONS

In conclusion, we have successfully applied iTRAQ and ICPL labeling to quantify proteins present in the castor seed development. Indeed we have identified with high confidence a total of 1875 proteins and described their modulation during seed development. Additionally, this work give important insights into certain aspects of the biology of castor seed development such as carbon flow, anabolism, and catabolism of fatty acid and the pattern of deposition of SSP, toxins, and allergens such as ricin and 2S albumins. We also found, for the first time, some genes of SSP that are differentially expressed during seed development. Taken together these data represent a step forward in understanding the molecular mechanisms that regulate the development of Castor oil seed (*Ricinus communis* L.), a plant of industrial importance.

ASSOCIATED CONTENT

Supporting Information

Supplementary Table 1 − Identified protein groups from endosperm of castor seeds generated by Protein Center analysis (Thermo Scientific). In this table are reported the strategy used, replicate origin, Uniprot, NCBI and gene number and gene model originated from the genome draft (Chan et al., 2010), the presence of alternative spliced, transmembrane domain (TM), signal peptide (SIGNAL), PFAM and gene ontology slim (GO_Slim), and the number of unique peptides, protein coverage, and peptide sequence. Supplementary Table 2 − Normalized values of proteins present in two or more labeling replicates and with at least one, two, or three unique peptides. Tables are shown in different spread sheets. NA, not applicable. Supplementary Table 3 − List of the proteins submitted to statistic evaluation. These proteins were identified with at least three unique peptides with quantitative values, presence in two or three stages, and presence in a minimum of two labeling replicates. Supplementary Table 4 − Fuzzy c-means clustering of 975 protein with three unique peptides and present in at least one replicate grouped into three clusters, each of which reflecting the trend in deposition pattern of the proteins during seed development. To be member of a cluster, the corresponding membership value has to be larger than 0.5. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: gilberto@iq.ufrj.br. Phone: +55-21-25627353. Fax: +55-21-24955765 (G.B.D.).
*E-mail: bioplant@ufc.br. Phone: +55-85-33669824. Fax: +55-85-33669829 (F.A.P.C.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

F.C.S.N., G.B.D., and F.A.P.C. were supported by PETROBRAS and the Brazilian National Research Council (CNPq); G.P. is supported by the Danish Medical Science Research Council (GP Grant No. 11-107551); and V.S. is supported by the Danish Council for Independent Research, Natural Sciences (FNU).

REFERENCES

(1) Severino, L. S.; Auld, D. L.; Baldanzi, M.; Cândido, M. J. D.; Chen, G.; Crosby, W.; Tan, D.; He, X.; Lakshmamma, P.; Lavanya, C.; Machado, O. L. T.; Mielke, T.; Milani, M.; Miller, T. D.; Morris, J. B.; Morse, S. A.; Navas, A. A.; Soares, D. J.; Sofiatti, V.; Wang, M. L.; Zanotto, M. D.; Zieler, H. A Review on the Challenges for Increased Production of Castor. *Agron. J.* **2012**, *104* (4), 853−880.

(2) da Silva César, A.; Otávio Batalha, M. Biodiesel production from castor oil in Brazil: A difficult reality. *Energy Policy* **2010**, *38* (8), 4031−4039.

(3) Brown, A. P.; Kroon, J. T.; Swarbreck, D.; Febrer, M.; Larson, T. R.; Graham, I. A.; Caccamo, M.; Slabas, A. R. Tissue-specific whole transcriptome sequencing in castor, directed at understanding triacylglycerol lipid biosynthetic pathways. *PLoS One* **2012**, *7* (2), e30100.

(4) Gressel, J. Transgenics are imperative for biofuel crops. *Plant Sci.* **2008**, *174* (3), 246−263.

(5) Chan, A. P.; Crabtree, J.; Zhao, Q.; Lorenzi, H.; Orvis, J.; Puiu, D.; Melake-Berhan, A.; Jones, K. M.; Redman, J.; Chen, G.; Cahoon, E. B.; Gedil, M.; Stanke, M.; Haas, B. J.; Wortman, J. R.; Fraser-Liggett, C. M.; Ravel, J.; Rabinowicz, P. D. Draft genome sequence of the oilseed species Ricinus communis. *Nat. Biotechnol.* **2010**, *28* (9), 951−6.

(6) Feuillet, C.; Leach, J. E.; Rogers, J.; Schnable, P. S.; Eversole, K. Crop genome sequencing: lessons and rationales. *Trends Plant Sci.* **2011**, *16* (2), 77−88.

(7) Nogueira, F. C.; Palmisano, G.; Soares, E. L.; Shah, M.; Soares, A. A.; Roepstorff, P.; Campos, F. A.; Domont, G. B. Proteomic profile of the nucellus of castor bean (Ricinus communis L.) seeds during development. *J. Proteomics* **2012**, *75* (6), 1933−9.
(8) Brown, A. P.; Kroon, J. T.; Topping, J. F.; Robson, J. L.; Simon, W. J.; Slabas, A. R. Components of complex lipid biosynthetic pathways in developing castor (Ricinus communis) seeds identified by MudPIT analysis of enriched endoplasmic reticulum. *J. Proteome Res.* **2011**, *10* (8), 3565−77.
(9) Houston, N. L.; Hajduch, M.; Thelen, J. J. Quantitative proteomics of seed filling in castor: comparison with soybean and rapeseed reveals differences between photosynthetic and non-photosynthetic seed metabolism. *Plant Physiol.* **2009**, *151* (2), 857−68.
(10) Nogueira, F. C.; Palmisano, G.; Schwammle, V.; Campos, F. A.; Larsen, M. R.; Domont, G. B.; Roepstorff, P. Performance of isobaric and isotopic labeling in quantitative plant proteomics. *J. Proteome Res.* **2012**, *11* (5), 3046−52.
(11) Greenwood, J. S.; Bewley, J. D. Seed development in Ricinus communis (castor bean). I. Descriptive morphology. *Can. J. Bot.* **1982**, *60* (9), 1751−1760.
(12) Nogueira, F. C.; Goncalves, E. F.; Jereissati, E. S.; Santos, M.; Costa, J. H.; Oliveira-Neto, O. B.; Soares, A. A.; Domont, G. B.; Campos, F. A. Proteome analysis of embryogenic cell suspensions of cowpea (Vigna unguiculata). *Plant Cell Rep.* **2007**, *26* (8), 1333−43.
(13) Gobom, J.; Nordhoff, E.; Mirgorodskaya, E.; Ekman, R.; Roepstorff, P. Sample purification and preparation technique based on nano-scale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry. *J. Mass Spectrom.* **1999**, *34* (2), 105−16.
(14) Palmisano, G.; Lendal, S. E.; Engholm-Keller, K.; Leth-Larsen, R.; Parker, B. L.; Larsen, M. R. Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. *Nat. Protoc.* **2010**, *5* (12), 1974−82.
(15) Melo-Braga, M. N.; Verano-Braga, T.; Leon, I. R.; Antonacci, D.; Nogueira, F. C.; Thelen, J. J.; Larsen, M. R.; Palmisano, G. Modulation of protein phosphorylation, N-glycosylation and Lys-acetylation in grape (Vitis vinifera) mesocarp and exocarp owing to Lobesia botrana infection. *Mol. Cell. Proteomics* **2012**, *11* (10), 945−56.
(16) Thingholm, T. E.; Palmisano, G.; Kjeldsen, F.; Larsen, M. R. Undesirable charge-enhancement of isobaric tagged phosphopeptides leads to reduced identification efficiency. *J. Proteome Res.* **2010**, *9* (8), 4045−52.
(17) Taverner, T.; Karpievitch, Y. V.; Polpitiya, A. D.; Brown, J. N.; Dabney, A. R.; Anderson, G. A.; Smith, R. D. DanteR: an extensible R-based tool for quantitative analysis of -omics data. *Bioinformatics* **2012**, *28* (18), 2404−6.
(18) Schwämmle, V.; Leon, I. R.; Jensen, O. N. Assessment and improvement of statistical tools for comparative proteomics analysis of sparse data sets with few experimental replicates. *J. Proteome Res.* **2013**, *12*, 3874−3883.
(19) Smyth, G. K. limma: Linear Models for Microarray Data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*; Gentleman, R., Carey, V., Huber, W., Irizarry, R., Dudoit, S., Eds.; Springer: New York, 2005; pp 397−420.
(20) Breitling, R.; Armengaud, P.; Amtmann, A.; Herzyk, P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett.* **2004**, *573* (1−3), 83−92.
(21) Koziol, J. A. Comments on the rank product method for analyzing replicated experiments. *FEBS Lett.* **2010**, *584* (5), 941−4.
(22) Storey, J. D. A direct approach to false discovery rates. *J. R. Stat. Soc., Ser. B* **2002**, *64* (3), 479−498.
(23) Kumar, L.; Futschik, E. M. Mfuzz: a software package for soft clustering of microarray data. *Bioinformation* **2007**, *2* (1), 5−7.
(24) Schwammle, V.; Jensen, O. N. A simple and fast method to determine the parameters for fuzzy c-means cluster analysis. *Bioinformatics* **2010**, *26* (22), 2841−8.
(25) Greenwood, J. S.; Gifford, D. J.; Bewley, J. D. Seed development in Ricinus communis cv. Hale (castor bean). II. Accumulation of phytic acid in the developing endosperm and embryo in relation to the deposition of lipid, protein, and phosphorus. *Can. J. Bot.* **1984**, *62* (2), 255−261.
(26) Greenwood, J. S.; Bewley, J. D. Seed development in Ricinus communis cv. Hale (castor bean). III. Pattern of storage protein and phytin accumulation in the endosperm. *Can. J. Bot.* **1985**, *63* (12), 2121−2128.
(27) Campos, F. A. P.; Nogueira, F. C. S.; Cardoso, K. C.; Costa, G. C. L.; Del Bem, L. E. V.; Domont, G. B.; Da Silva, M. J.; Moreira, R. C.; Soares, A. A.; Jucá, T. L. Proteome analysis of castor bean seeds. *Pure Appl. Chem.* **2010**, *82* (1), 259−267.
(28) Guelette, B. S.; Benning, U. F.; Hoffmann-Benning, S. Identification of lipids and lipid-binding proteins in phloem exudates from Arabidopsis thaliana. *J. Exp. Bot.* **2012**, *63* (10), 3603−16.
(29) Bjornberg, O.; Maeda, K.; Svensson, B.; Hagglund, P. Dissecting molecular interactions involved in recognition of target disulfides by the barley thioredoxin system. *Biochemistry* **2012**, *51* (49), 9930−9.
(30) Runquist, M.; Kruger, N. J. Control of gluconeogenesis by isocitrate lyase in endosperm of germinating castor bean seedlings. *Plant J.* **1999**, *19* (4), 423−31.
(31) Chileh, T.; Esteban-Garcia, B.; Alonso, D. L.; Garcia-Maroto, F. Characterization of the 11S globulin gene family in the castor plant Ricinus communis L. *J. Agric. Food Chem.* **2010**, *58* (1), 272−81.
(32) Tan-Wilson, A. L.; Wilson, K. A. Mobilization of seed protein reserves. *Physiol. Plant.* **2012**, *145* (1), 140−53.
(33) Sirvent, S.; Palomares, O.; Cuesta-Herranz, J.; Villalba, M.; Rodriguez, R. Analysis of the Structural and Immunological Stability of 2S Albumin, Nonspecific Lipid Transfer Protein, and Profilin Allergens from Mustard Seeds. *J. Agric. Food Chem.* **2012**, *60*, 6011−6018.
(34) Thorpe, S. C.; Kemeny, D. M.; Panzani, R. C.; McGurl, B.; Lord, M. Allergy to castor bean. II. Identification of the major allergens in castor bean seeds. *J. Allergy Clin. Immunol.* **1988**, *82* (1), 67−72.
(35) Amara, I.; Odena, A.; Oliveira, E.; Moreno, A.; Masmoudi, K.; Pagès, M.; Goday, A. Insights into Maize LEA Proteins: From Proteomics to Functional Approaches. *Plant Cell Physiol.* **2012**, *53* (2), 312−329.
(36) Hyun, T. K.; Kumar, D.; Cho, Y. Y.; Hyun, H. N.; Kim, J. S. Computational identification and phylogenetic analysis of the oil-body structural proteins, oleosin and caleosin, in castor bean and flax. *Gene* **2013**, *515* (2), 454−60.
(37) Dennis, D. T.; Miernyk, J. A. Compartmentation of Non-photosynthetic Carbohydrate Metabolism. *Ann. Rev. Plant Physiol.* **1982**, *33* (1), 27−50.
(38) Barvkar, V. T.; Pardeshi, V. C.; Kale, S. M.; Kadoo, N. Y.; Giri, A. P.; Gupta, V. S. Proteome profiling of flax (Linum usitatissimum) seed: characterization of functional metabolic pathways operating during seed development. *J. Proteome Res.* **2012**, *11* (12), 6264−76.
(39) Barratt, D. H.; Derbyshire, P.; Findlay, K.; Pike, M.; Wellner, N.; Lunn, J.; Feil, R.; Simpson, C.; Maule, A. J.; Smith, A. M. Normal growth of Arabidopsis requires cytosolic invertase but not sucrose synthase. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (31), 13124−9.
(40) Zhao, Z.; Assmann, S. M. The glycolytic enzyme, phosphoglycerate mutase, has critical roles in stomatal movement, vegetative growth, and pollen production in Arabidopsis thaliana. *J. Exp. Bot.* **2011**, *62* (14), 5179−89.
(41) Huang, Y.; Blakeley, S. D.; McAleese, S. M.; Fothergill-Gilmore, L. A.; Dennis, D. T. Higher-plant cofactor-independent phosphoglyceromutase: purification, molecular characterization and expression. *Plant Mol. Biol.* **1993**, *23* (5), 1039−53.
(42) Olinares, P. D.; Ponnala, L.; van Wijk, K. J. Megadalton complexes in the chloroplast stroma of Arabidopsis thaliana characterized by size exclusion chromatography, mass spectrometry, and hierarchical clustering. *Mol. Cell. Proteomics* **2010**, *9* (7), 1594−615.
(43) Friso, G.; Majeran, W.; Huang, M.; Sun, Q.; van Wijk, K. J. Reconstruction of metabolic pathways, protein expression, and homeostasis machineries across maize bundle sheath and mesophyll

chloroplasts: large-scale quantitative proteomics using the first maize genome assembly. *Plant Physiol.* **2010**, *152* (3), 1219−50.
(44) Barsan, C.; Sanchez-Bel, P.; Rombaldi, C.; Egea, I.; Rossignol, M.; Kuntz, M.; Zouine, M.; Latche, A.; Bouzayen, M.; Pech, J. C. Characteristics of the tomato chromoplast revealed by proteomic analysis. *J. Exp. Bot.* **2010**, *61* (9), 2413−31.
(45) Schwender, J.; Goffman, F.; Ohlrogge, J. B.; Shachar-Hill, Y. Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature* **2004**, *432* (7018), 779−782.
(46) Goffman, F. D.; Alonso, A. P.; Schwender, J.; Shachar-Hill, Y.; Ohlrogge, J. B. Light Enables a Very High Efficiency of Carbon Storage in Developing Embryos of Rapeseed. *Plant Physiol.* **2005**, *138* (4), 2269−2279.
(47) Troncoso-Ponce, M. A.; Kilaru, A.; Cao, X.; Durrett, T. P.; Fan, J.; Jensen, J. K.; Thrower, N. A.; Pauly, M.; Wilkerson, C.; Ohlrogge, J. B. Comparative deep transcriptional profiling of four developing oilseeds. *Plant J.* **2011**, *68* (6), 1014−27.
(48) Walker, R. P.; Battistelli, A.; Moscatello, S.; Chen, Z. H.; Leegood, R. C.; Famiani, F. Metabolism of the seed and endocarp of cherry (Prunus avium L.) during development. *Plant Physiol. Biochem.* **2011**, *49* (8), 923−30.
(49) Lin, M.; Oliver, D. J. The role of acetyl-coenzyme a synthetase in Arabidopsis. *Plant Physiol.* **2008**, *147* (4), 1822−9.
(50) Tovar-Mendez, A.; Miernyk, J. A.; Randall, D. D. Regulation of pyruvate dehydrogenase complex activity in plant cells. *Eur. J. Biochem.* **2003**, *270* (6), 1043−9.
(51) Chia, T. Y.; Pike, M. J.; Rawsthorne, S. Storage oil breakdown during embryo development of Brassica napus (L.). *J. Exp. Bot.* **2005**, *56* (415), 1285−96.
(52) Graham, I. A. Seed storage oil mobilization. *Annu. Rev. Plant Biol.* **2008**, *59*, 115−42.