Proteomic Analysis of Embryo Isolated From Mature *Jatropha curcas* L. Seeds

Ayesha Ramzan¹, Mohibullah Shah*¹, Najeeb Ullah¹, Sheheryar², José R. S. Nascimento², Francisco A. P. Campos³, Gilberto B. Domont³, Fábio C. S. Nogueira³,⁴ and Magda H. Abdellattif*⁵

¹ Department of Biochemistry, Bahauddin Zakariya University, Multan, Pakistan, ² Department of Biochemistry and Molecular Biology, Federal University of Ceará, Fortaleza, Brazil, ³ Department of Biochemistry, Institute of Chemistry, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, ⁴ Laboratory of Proteomics/LADETEC, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, ⁵ Department of Chemistry, College of Science, Taif University, Taif, Saudi Arabia

*Correspondence: Mohibullah Shah
mohibush@gmail.com; mohib@bzu.edu.pk
Magda H. Abdellattif
m.hasan@tu.edu.sa

Jatropha curcas L. is a non-edible oilseed containing almost 40% of seed oil and is famous as the best source of raw material for biofuel production. *J. curcas* seeds contain three main tissues, such as inner integument, endosperm, and embryo. To best understand the physiological events related to specific tissues, it is important to perform the proteome analysis of these tissues. Previously we have explored the pattern of reserves deposition and tissue-specific biological pathways by analyzing the proteome of the inner integument and endosperm and organelles, such as plastids and gerontoplasts isolated from these tissues. The focus of the present study was to perform the proteomic analysis of embryo isolated from the mature seeds of *J. curcas*.

This analysis resulted in the identification of 564 proteins of which 206 are not identified previously from any other tissue of this plant. The identified proteins were functionally classified using the MapMan classification system revealing various proteins involved in different functionalities. The proteins involved in transport functions and those with proteolytic activity were determined through the Transporter Classification Database (TCDB) and MEROPS database, respectively. In addition to identify a large number of proteins participating in various metabolic processes, we found several proteins involved in defense functions, such as the members of chaperones and the ubiquitin-proteasome system. Similarly, members of the legumin and vicilin family of seed storage proteins (SSPs) were identified which in addition to their storage function, are involved in defense. In addition, we have reported that proteases belonging to different mechanistic classes and are involved in diverse physiological functions. Last but not the least, several classes of transport-related proteins were identified that are discussed concerning their function in the transportation of different nutrients across the embryo. To the best of our knowledge, this study reported the highest number of proteins identified from the embryo of mature *J. curcas* seeds, most of which are essential for seed germination, reflecting the fact that many proteins required for germination are already present in the mature embryo.

**Keywords:** *Jatropha curcas*, subproteome, embryo, biodiesel, oilseed
INTRODUCTION

The seeds of embryonic plants are comprised of three basic components, an embryo, embryo proper tissue, and a seed coat to insulate seed from environmental stress (De Smet et al., 2010). Nutrients can be stored in various tissues, such as cotyledons, endosperm, or megagametophyte. In a few seeds, an embryo is trapped in the endosperm, and in other cases, the endosperm is absorbed by the embryo. In the latter case, the embryo develops inside the growing seeds, and cotyledons become enriched with the storage products (Miernyk et al., 2011). Jatropha curcas L. is a non-edible oil rich shrub and is best known for its potential to be used as a feedstock for biodiesel production. The key reasons are its oil richness, rapid growth, drought resistance ability, and necessary adaptation to widespread environmental conditions. Mature J. curcas seeds contain a tiny embryo enclosed in a dense endosperm (Liu et al., 2009). Both embryo and endosperm have distinct ploidies and interact in a coordinated manner to control the seed growth.

The availability of the draft genome sequence of J. curcas enabled the scientific community to perform sub-proteome analysis associated with the seed development. In this aspect, our first study was focused on plastids (Pinheiro et al., 2013) isolated from the endosperm of mature seeds resulting in the identification of 923 unique proteins associated with the biosynthesis of amino acids and fatty acids. This study was followed by the comparative proteome analysis of the two distinct regions of the inner integument in which a total of 1,770 proteins were identified (Soares et al., 2014). This study highlighted the role of different enzymes, i.e., proteases, nucleases, lipases, and carbohydrate-acting enzymes in programmed cell death (PCD), which are associated with the provision of nutrients for growing embryo and endosperm. Though these two studies provided insight into the diverse biological aspects of the seeds, the changes associated with the proteome during seeds development remained unanswered. To fill this gap, a comprehensive proteomic analysis of the endosperm at five different developmental stages was performed (Shah et al., 2015). The major classes of 1,760 identified proteins were those, related to seed storage, lipids and carbohydrates metabolism, xenobiotics metabolism, and proteolysis. Moreover, this study revealed the deposition of various isoforms of these proteins during different developmental stages of the seed. To determine the biological aspects of the plastids transition to gerontoplast, we had performed their histological and proteomic investigations. For this purpose, the plastids were isolated from the inner integument of developing J. curcas seeds and identified 1,923 proteins after proteomic analysis (Shah et al., 2016). Here, we reported and discussed proteins involved in a myriad of functions related to the dismantling of this tissue, providing nutrients to other tissues under development. Despite the importance of the embryo in the seeds, its proteome is yet unanalyzed to reveal the biological pathways associated with this tissue in J. curcas seeds. Previously two different studies utilized the two-dimensional gel electrophoresis (2DE) based approach for analyzing the embryo proteome of J. curcas seeds (Liu et al., 2009, 2011). However, these studies collectively identified < 30 proteins, most of which were related to reserve mobilization during germination.

In hitherto studies, we have targeted various tissues and organelles of J. curcas which collectively resulted in the identification of 6,188 proteins of unique physiological characteristics. In this regard, the current study is aimed to analyze the proteome of an embryo isolated from mature J. curcas seeds to disseminate biological pathways related to this important tissue. Here, we obtained 364 proteins with 206 newly identified and were never reported in our previous results from other tissues. The identified proteins were discussed with emphasis on their involvement in seed development and reserves deposition and mobilization during germination.

MATERIALS AND METHODS

Embryo Isolation and Protein Extraction

Mature J. curcas seeds were collected from the Punjab Province of Pakistan. Seeds were manually dehulled and the embryos were separated from seeds (Figure 1) with a scalpel to prevent its contact with endosperm and inner integument. To remove the lipids, the embryos were cut into small fragments and placed under gentle stirring in acetone for almost 30 h. The acetone was changed every 5 h. This material was then dried at room temperature and ground into powder with mortar and pestle by using liquid nitrogen. The powdered material was then stored at −80°C until further use. Powdered embryos were subjected to protein extraction following a previously established method (Vasconcelos et al., 2005). For the extraction of soluble proteins, 3 replicates of 0.1 g powdered embryo were weighed and homogenized in 5 ml of pyridine buffer (50 mM pyridine, 10 mM thiourea, and 1% SDS, pH 5.0) with polyvinyl-polypyrrolidone in a ratio of 1:40:2 (w/v/w). The mixture was kept at stirring for 3 h at 4°C. The centrifugation was performed for 30 min at 10,000 rpm. Proteins were precipitated from the supernatant using trichloroacetic acid (10%) in acetone. Cold acetone was used to wash the pellets three times, centrifuged, and dried at room temperature. The dried pellets were solubilized in sample buffer containing 7 M urea/2 M thiourea and 100 mM triethylammonium bicarbonate (TEAB) buffer. The same process was repeated for all three biological replicates. Bradford assay (Bradford, 1976) was used for the identification of 923 unique proteins associated with the important tissue. Here, we obtained 564 proteins with 206 newly identified and were never reported in our previous results from other tissues. The identified proteins were discussed with emphasis on their involvement in seed development and reserves deposition and mobilization during germination.

![FIGURE 1](https://example.com/figure1.png) | The mature seed of Jatropha curcas. (A) Mature seed with seed coat. (B) Mature seed without a seed coat. (C,D) Endosperm. (E) Embryo.
to measure the concentration of proteins using bovine serum albumin (BSA) as standard.

**Samples Preparation for Liquid Chromatography With Tandem Mass Spectrometry (LC-MS/MS)**

In-solution trypsin digestion was carried out with 50 μg of proteins following our previously used protocol (Pinheiro et al., 2013). Before trypsinization, the proteins were reduced using 10 mM tris (2-carboxyethyl) phosphine (TCEP) for 1 h at 30°C and alkylated with 40 mM iodoacetamide at room temperature in the dark for 30 min. To reduce the urea concentration to less than 1 M, the samples were diluted with 100 mM TEAB (1:9 v/v). Finally, samples were subjected to trypsin digestion for 18 h at 37°C. Proteins digestion was quenched with TFA. After digestion, the resulting peptides were cleaned through a spin column, dried in speed-vacuum (SpeedVac), and stored at −80°C for further use.

**Nanoscale Liquid Chromatography Coupled to Tandem Mass Spectrometry (NanoLC-MS/MS) and Data Analysis**

Before the introduction to nanoscale liquid chromatography coupled to tandem mass spectrometry (NanoLC-MS/MS), peptides were solubilized in 20 μl of 0.1% formic acid and diluted to 5x. These peptides were quantified with a Qubit protein assay kit. Diluted samples (4 μl) were subjected to NanoLC-MS/MS system interfaced online to ESI-LTQ Orbitrap Velos MS. Peptides were loaded onto a 150 μm × 2 cm trap column packed with C18 ReproSil 3 μm resin and then eluted onto an analytical column of 100 μm × 15 cm packed with the similar resin. The separated peptides were collected through a gradient from 100% of A (0.1% formic acid) to 35% of B (0.1% formic acid and 95% acetonitrile) for 150 min, followed by 35–90% of solution B for 15 min and 90% for 5 min. For MS1 spectra, each data-dependent acquisition mode comprised of a survey scan covering a range of m/z 300–90% for 5 min. For MS1 spectra, each data-dependent acquisition mode comprised of a survey scan covering a range of m/z 300–1800 and 60,000 resolution with a targeted value of 1 × 10−6 ions. Tandem mass spectrometry (MS/MS) fragmentation of the ten major intense ions was acquired using a normalized collision-induced dissociation of formerly fragmented ions. The m/z of fragmented precursor ions were excluded for 60 s. Each biological replicate was injected three times which resulted in nine technical replicates for three biological replicates of the sample.

An Xcalibur v.2.1 (Thermo Fisher Scientific) was used to view the raw files while database search was performed using the SequestTM algorithm embedded in Proteome Discoverer 2.1 (Thermo Fisher Scientific) against the combined database of *J. curcas* nuclear (Chan et al., 2010) and plastid (Asif et al., 2010) genomes. The search parameters were: MS accuracy 10 ppm, MS/MS accuracy 0.1 Da, trypsin digestion with two missed cleavages, carbamidomethylation of cysteine as fixed oxidized methionine as variable modification. A false discovery rate of 1% was used at the protein and peptide level. The identified proteins of *J. curcas* embryo that appeared in almost 2 biological replicates were used for the downstream analysis. For annotation of the identified proteins, we performed a local BLAST of the identified proteins against *Arabidopsis thaliana* and *Ricinus communis* protein databases that were downloaded from TAIR1 and UniProt,2 with e-value of 1 × 10−5 for determining the respective orthologous proteins. Protein functional classification was performed according to the bincodes of MapMan.3 For the classification of proteases, identified proteins were scanned against MEROPS database (Rawlings et al., 2018)4 using blastp with an e-value of 1 × 10−15. For the classification of transporters, TCDB (Saier et al., 2016)5 was used with an e-value of 1 × 10−15.

**RESULTS AND DISCUSSION**

**Proteins Identification and Functional Classification**

Proteomic studies on *J. curcas* resulted in a wealth of information regarding biological events associated with various tissues of the seeds. However, limited attention was given to the embryo. In this study, we have identified 564 proteins (Supplementary Table 1) from the *J. curcas* embryo of which 208 were previously not identified from other tissues of this plant. The identified proteins were functionally classified using a MapMan classification system (Schwacke et al., 2019). The major functional classes include protein metabolism, lipid metabolism, carbohydrate metabolism, defense-related proteins, and proteases among the others (Figure 2).

**Lipid Metabolism Related Proteins**

*Jatropha curcas* is an industrially significant oilseed crop mainly due to the high quantity of oil in their seeds. Although in *J. curcas* seeds, endosperm is the main site for the oil accumulation, however, lipids are also present in the embryo (Chen et al., 2011; de Lopes et al., 2013). In the present study, 27 lipid metabolism-related proteins were identified (Table 1) corresponding to almost 5% of the total identified proteins and are involved in the numerous metabolic pathways of lipids. The initial phase of this pathway relies on acetyl-CoA for the biosynthesis of FA in plastids. Acetyl-CoA is provided by the plastidial pyruvate dehydrogenase complex, of which several subunits are identified here. Similarly, acetyl-CoA carboxylase, a biotin-bound one carbon carrier protein (Jcr4s01232.50) catalyzes the abridgment of acetyl-CoA to malonyl-CoA, is also identified. In addition, we identified ketoacyl-ACP synthase I (Jcr4s02541.50) catalyzing the synthesis of 18 carbon FA via condensation reactions. A series of desaturation reactions introduce desaturation and play a significant role in lipids quality (Gu et al., 2012). Previously, transcripts for multiple desaturase proteins, e.g., stearate, oleate, and linoleate desaturase were identified from *J. curcas* seeds (Costa et al., 2010). Significantly, four isoforms of stearate desaturase (Arabidopsis homologous) were identified from the endosperm proteome of *J. curcas* seeds4 (Shah et al., 2013).

---

1http://www.arabidopsis.org/
2http://www.uniprot.org/
3https://mapman.gabipd.org/
4https://www.ebi.ac.uk/merops/
5http://www.tcdb.org/
2015). However, in the present study, we did not identify any desaturase from the embryo of *J. curcas* seeds. The possible reason might be due to lower identifications as compared with our previous studies.

Fatty acids are condensed with glycerol skeleton to form triacylglycerols (TAGs) and finally aggregate in the form of oil bodies (OBs) inside the endoplasmic reticulum (ER). OBs consist of a core of TAGs accompanied by various proteins,
such as oleosins and caleosins (calcium-binding peroxigenases) incorporated in a phospholipid monolayer. In this study, we have identified two isoforms of oleosins (Jcr4S05992.20 and Jcr4S01276.90) and one caleosin (Jcr4S01398.30). In addition to these two oleosins, three other isoforms of oleosins were reported from our previous study on J. curcas endosperm (Shah et al., 2015). This may indicate the fact that different isoforms of the same proteins are active in different tissues. Oleosins are responsible for altering the OBs size and lipid accumulation in different seed tissues of Arabidopsis. Such storage deposits are typically large insoluble compounds that can remain stable for long spells in desiccated seeds (Siloto et al., 2006). Our results suggest that in J. curcas seeds embryo, oleosins might be involved in stabilizing OBs during the seed desiccation phase. This might prevent their degradation until they are needed for germination.

We have identified phospholipase D1 (Jcr4S00086.110) responsible for the large-scale breakdown of lipids related to cell death (Shah et al., 2015) and most often contributes to caspase dependent cellular death signaling mechanism. Previously we found them in the inner integument (Soares et al., 2014) and endosperm (Shah et al., 2015) of J. curcas seeds along with other hydrolases. They were believed to be involved in PCD in these different tissues. A previous study showed that the enhanced activity of proteins involved in β-oxidation and the glyoxyly pathway resulted in the deposition of particular enhanced activity of proteins involved in these different tissues. A previous study showed that the enhanced activity of proteins involved in β-oxidation and the glyoxyly pathway resulted in the deposition of particular fatty acids in growing rapeseeds (Chia et al., 2005). Here, we identified two glyoxylate cycle enzymes including two isoforms of aconitate (Jcr4S09697.10 Jcr4S00736.30) and one of malate synthase (Jcr4S00100.200). Acetyl CoA (end product of β-oxidation) serves as a substrate for the glyoxylate cycle (a modified form of TCA cycle) that skips the decarboxylation phase and permits the net production of carbon skeletons (succinate) without carbon loss as CO2 (Cornah and Smith, 2002). These findings suggest that during seed germination, the stored lipids will be degraded through the successive operation of β-oxidation, lipoxygenase, and glyoxylate cycle to provide carbon and energy for the seedling growth.

Carbohydrate Metabolism Related Proteins

We identified 51 carbohydrate metabolism-related proteins (Table 2) corresponding to 9% of the total identifications. In oilseeds, reserves are primarily synthesized during the seed-filling phase characterized by alterations in the morphological, cellular, and metabolic processes of embryo and endosperm (Hill et al., 2003). De novo synthesis of lipids requires carbon, energy, and reducing equivalents and is provided directly or indirectly via glycolysis in the form of acetyl-CoA, ATP, and nicotinamide adenine dinucleotide phosphate (NADPH) (Plaxton and Podestá, 2006). In the present study, several cytosolic and plastidial isoforms of glycolytic proteins were identified (Table 2). We identified fructose-1,6-bisphosphate aldolase (FBA) (Jcr4S14120.10) that catalyzes the aldol cleavage of Fru-1,6-bisP to glyceraldehyde-3-P (GAP) and dihydroxyacetone phosphate (DHAP). A previous research indicated that in heterotrophic embryos of sunflower, triose-phosphates were the primary source of carbon for the biosynthesis of fatty acids, but did not find the source of triose-phosphate production (Alonso et al., 2007).

We identified four cytosolic isoforms of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (Jcr4S07385.10, Jcr4S00953.40, Jcr4S00273.150, and Jcr4S00205.140) that catalyze the formation of 1,3-bisphosphoglycerate (BPGA) from Gly-3-P. In rapeseeds, the specific expression of GAPDH resulted in a 3–4-fold increase in G3P which ensured 40% improved oil deposition (Vigeolas et al., 2007). Such a role of GAPDH suggests that it might be the essential protein responsible for the significant amounts of oil deposition in an embryo. Additionally, we identified cytosolic isoform of phosphoglycerate mutase (Jcr4S28329.10) that was not identified in our previous studies from other seed tissues. It catalyzes the phosphate group interconversion between the C-3 carbon of 3-phosphoglycerate and C-2 carbon of 2-phosphoglycerate. Pyruvate, the outcome of glycolysis, serves as a precursor for acetyl-CoA. The production of acetyl-CoA primarily occurs through the pyruvate dehydrogenase complex (PDHC) inside the plastids (Joshee, 2006) and cannot penetrate the membrane. We have identified all the three subunits of plastidial PDHC (Jcr4S01924.60, Jcr4S00112.160, and Jcr4S04485.30) that support the biosynthesis of fatty acids within the plastids. All such proteins of the glycolytic pathway have already been reported from the endosperm and non-photosynthetic plastids of J. curcas seeds (Pinheiro et al., 2013; Shah et al., 2015). These consistent results support the notion that glycolytic pathway provides most of the carbon for FAS in growing embryos (Schwender et al., 2003).

Oxidative pentose phosphate pathway (OPPP) is another preferred route responsible for the generation of reducing power for fatty acids biosynthesis. We identified one oxidative phase protein PGL (Jcr4S16847.20) that catalyzes the hydrolysis of 6-phosphogluconolactone hydrolysis and one non-oxidative phase protein transketolase (Jcr4S00057.90) as well. It was observed that in sunflower, OPPP produces the bulk of reducing equivalents for the biosynthesis of fatty acids (Alonso et al., 2007). This study identified many functional glycolytic and OPPP proteins suggesting an active metabolism in the embryo.

Phosphoglucomutase (PGM) (Jcr4S00445.90) was identified here, catalyzing the reversible interconversion of Glc-1-P and Glc-6-P. The plastidial PGM is predominantly involved in the biosynthesis and degradation of starch molecules. Plastidial PGM mutant Arabidopsis seeds result in 40% less oil accretion in comparison with its wild-type seeds (Periappuram et al., 2000). Its cytosolic and plastidial isoforms have been reported in the endosperm (Shah et al., 2015), inner integument (Soares et al., 2014), and plastid of J. curcas seeds (Pinheiro et al., 2013). Identification of PGM establishes its significant impact on the accumulation of storage products in the seeds of J. curcas. We have identified cytosolic isoform of phosphoenolpyruvate carboxykinase (PEPCK) (Jcr4S08285.10), which catalyzes the synthesis of phosphoenolpyruvate from oxaloacetate. It has a vital role in the gluconeogenic production of sugar from stored oil during
TABLE 2 | Proteins related to carbohydrate metabolism identified from the embryo of mature J. curcas seeds.

| Protein accession | Description                                      | Best hit *Ricinus* | Best hit *Arabidopsis* |
|-------------------|-------------------------------------------------|--------------------|------------------------|
| Jcr4S00433.130    | Phosphofructokinase                              | B9RXE7             | AT1G12000.1            |
| Jcr4S14120.10     | Fructose-bisphosphate aldolase                  | B9SRH4             | AT2G36460.1            |
| Jcr4S02610.20     | Fructose-bisphosphate aldolase                  | B9T5T6             | AT2G36460.1            |
| Jcr4S01786.40     | Fructose-bisphosphate aldolase                  | B9SJJ9             | AT2G01140.1            |
| Jcr4S00484.40     | Fructose-bisphosphate aldolase                  | B9SWW4             | AT4G26530.2            |
| Jcr4S07385.10     | Glyceraldehyde-3-phosphate dehydrogenase         | B9RNW8             | AT2G24270.3            |
| Jcr4S00953.40     | Glyceraldehyde-3-phosphate dehydrogenase         | B9RNW8             | AT2G24270.3            |
| Jcr4S02273.150    | Glyceraldehyde-3-phosphate dehydrogenase         | B9RAL0             | AT3G04120.1            |
| Jcr4S00205.140    | Glyceraldehyde-3-phosphate dehydrogenase         | B9RAL0             | AT1G13440.1            |
| Jcr4S000043.160   | Phosphoglycerate kinase                          | B9RHY3             | AT1G79550.2            |
| Jcr4S28329.10     | Phosphoglycerate mutase                          | B9RF09             | AT3G05170.1            |
| Jcr4S00171.20     | Phosphophorylpyruvate hydratase (enolase)        | B9RIN6             | AT2G36530.1            |
| Jcr4S00445.90     | Phosphoglucomutase                               | B9SP64             | AT1G70730.3            |
| Jcr4S00575.60     | Pyruvate dehydrogenase E1 component alpha subunit| B9S2H9             | AT1G59900.1            |
| Jcr4S01924.60     | Dihydrolipoamide acetyltransferase              | B9SLI7             | AT3G52200.1            |
| Jcr4S00306.100    | Dihydrolipoamide acetyltransferase              | B9SSV2             | AT1G54220.2            |
| Jcr4S04485.30     | Dihydrolipoyline-residue succinyltransferase     | B9SVA1             | AT4G26910.1            |
| Jcr4S00112.10     | Dihydrolipoyline dehydrogenase                  | B9RZ2N             | AT3G16960.1            |
| Jcr4S08285.10     | Phosphoenolpyruvate carboxykinase                | B9R6O4             | AT4G37870.1            |
| Jcr4S18847.20     | 6-Phosphogluconolactonase                       | B9RUV5             | AT5G24400.1            |
| Jcr4S00057.90     | Transketolase                                    | B9RDA1             | AT2G45290.1            |
| Jcr4S00380.50     | Aldose 1-epimerase                               | B9SYV8             | AT3G17940.1            |
| Jcr4S00100.200    | Malate synthase                                  | B9RAK0             | AT5G03860.2            |
| Jcr4S03576.70     | Malic enzyme                                     | B9RS07             | AT5G25880.1            |
| Jcr4S03356.50     | Malic enzyme                                     | B9RK6              | AT5G25880.1            |
| Jcr4S03281.10     | Malic enzyme                                     | B9RQE8             | AT1G79750.1            |
| Jcr4U29909.10     | Alcohol dehydrogenase                            | B9RBJ1             | AT1G22440.1            |
| Jcr4S00609.30     | Alcohol dehydrogenase                            | B9RBJ1             | AT1G22440.1            |
| Jcr4S00529.80     | Alcohol dehydrogenase                            | B9SJ18             | AT1G77120.1            |
| Jcr4S02965.20     | Alcohol dehydrogenase                            | B9SJ18             | AT1G77120.1            |
| Jcr4S03546.10     | Aldehyde dehydrogenase (NAD +)                  | B9RB9H             | AT3G48000.1            |
| Jcr4S03703.10     | Aminobutyrate aminotransferase, putative         | B9SAY5             | AT3G22200.1            |
| Jcr4S02325.10     | (α–β)-beta-glucan endohydrolase                  | B9SP5S             | AT2G01650.1            |
| Jcr4S02004.20     | Phosphoprotein phosphatase, putative             | B9RJC0             | AT3G19420.1            |
| Jcr4S01159.50     | Catalase                                         | B9RBN4             | AT4G35090.1            |
| Jcr4S00684.10     | UTP-glucose-1-phosphate uridylyltransferase      | B9SKS5             | AT5G17310.2            |
| Jcr4S00313.50     | Glucan endo-1,3-beta-glucosidase                 | B9T103             | AT1G66250.1            |
| Jcr4S09697.10     | Aconitate hydratase (Aconitase)                  | B9SX8              | AT4G35830.1            |
| Jcr4S00736.30     | Aconitate hydratase (Aconitase)                  | B9T2U5             | AT2G05710.1            |
| Jcr4S00125.80     | Succinyl-CoA synthetase                          | B9RL91             | AT5G23250.1            |
| Jcr4S06387.30     | Succinyl-CoA synthetase                          | B9RL91             | AT5G23250.1            |
| Jcr4S03295.10     | Malate dehydrogenase                             | B9SE47             | AT1G53240.1            |
| Jcr4S00279.20     | Malate dehydrogenase                             | B9TSE4             | AT4G34330.1            |
| Jcr4S00055.90     | Cytochrome C oxidase, putative                   | B9YV06             | AT1G80230.1            |
| Jcr4S01269.80     | ATP synthase subunit beta                        | B9TV8              | AT5G08680.1            |
| Jcr4S00914.20     | ATP synthase subunit beta                        | B9TV8              | AT5G08680.1            |
| Jcr4S00868.20     | ATP synthase subunit gamma                       | B9SZ3              | AT2G33040.1            |
| Jcr4S01672.20     | Ferredoxin                                       | B9SCU0             | AT2G27510.1            |
| Jcr4S01794.30     | Chlorophyll a-b binding protein                  | B9SID9             | AT5G54270.1            |
| Jcr4S01456.10     | Phosphoglycerate dehydrogenase                   | B9RYA3             | AT4G34200.1            |

...
provision of sugars during radical protrusion from embryo and seedling growth.

During the early phase of seed germination (oxygen deficient condition), the output of the TCA cycle is not sufficient and energy requirements are mostly met mainly by glycolysis and anaerobic respiration. In embryo, alcohol dehydrogenase is believed to be the part of glycol-metabolism where it catalyzes the synthesis of alcohol from pyruvic acid reduction (Yang et al., 2007). Here, we have identified multiple isoforms of alcohol dehydrogenase including one unique isoform (Table 1) suggesting a major energy producer in hypoxia to facilitate the seed germination. As the energy from alcohol fermentation cannot fulfill the entire needs of germinating seeds. Therefore, at this point, TCA cycle produces the maximum amount of energy after the cellular environment is rich in oxygen (Miro et al., 2017). We identified three proteins of the TCA cycle, such as aconitate hydratase (Jcr4S09697.10), succinyl-CoA synthetase (Jcr4S00125.80), and malate dehydrogenase (Jcr4S03295.10). Although the TCA cycle is a major producer of energy during germination, the above three proteins identification hints that these proteins might be involved in reserves mobilization and provide energy for seedling growth. The identification of a significant number of the proteins related to carbohydrate metabolism indicates the importance of this pathway for energy production and providing precursors for lipid metabolism.

**Proteostasis and Defense Related Proteins**

Many seeds can resist the environmental threats and develop into the new plant under favorable conditions. The emergence of adaptable strategies enables the seeds to defend themselves against stress. The failure can lead to the death of a new plant. Chaperones protect the proteins in their functionally active form and are concerned with assembly, folding and sustainability, and proteolysis (Trivedi et al., 2016). In this study, we identified 53 defense-related proteins (Table 3) corresponding to 9% of the total identified proteins. Eight isoforms of the chaperone (HSP70), an important class of chaperones associated with protein folding, and translocation to various cellular organelles, were identified. They inhibit protein accumulation and enable refolding of native proteins in normal and stress conditions as well. Additionally, we identified four isoforms of HSP 90 which are exclusively involved in the signaling pathways. Steroid hormone receptors and protein kinases are regarded as their substrates (Young et al., 2001).

Further, four isoforms of HSP60 (Jcr4S04936.50, Jcr4S00653.20, Jcr4S03954.10, and Jcr4S02144.50) were identified which are ATP dependent mitochondrial chaperons. They are mainly involved in the import, refolding, and assembling of misfolded or unfolded proteins inside the mitochondrial matrix during stress conditions. In addition, these proteins facilitate the optimal growth and development of chloroplasts, embryos, and seedlings (Wang et al., 2004). Thus, the identification of various isoforms of heat shock proteins (HSPs) in the embryo of J. curcas seeds is indicative of the presence of the phenomenon of protein folding, translocation, and degradation in normal and stressed conditions.

The ubiquitin associated deterioration pathway plays a significant role in multiple aspects of plant growth. In the present study, the components of the ubiquitin proteasome system (UPS) were identified, such as isoforms of ubiquitin protein, 205 proteasome subunit alpha (Jcr4S02802.40), and regulatory component RPT5 of 265 proteasome (Jcr4S00385.110) (Table 3). E3 ubiquitin-protein ligase (Jcr4S02462.10) which acts as a modulator of plant responses to abiotic stresses, such as cold, heat, radiation, desiccation, salt, and nutrient deficiency were identified. The UPS promotes adaption to abiotic stress by managing the functioning of stress hormones like abscisic acid. It is accomplished through the activity of several ubiquitin ligases that control the signaling of various stress hormones (Komander and Rape, 2012). The identification of UPS in our data indicates the importance of this regulatory mechanism. It protects the embryo against environmental stress by degrading potentially harmful proteins, which monitor the concentration of key enzymes and regulatory proteins to maintain cellular homeostasis in the embryo of mature J. curcas seeds.

We identified eight hSHPs isoforms not identified previously from other tissues of J. curcas seeds (Table 3). These ubiquitous proteins are produced in response to a high temperature but are equally found during the particular phases of plant growth (Wehmeyer and Vierling, 2000). Here, the identification of multiple unique isoforms of hSHPs indicates that these proteins are embryo specific. They may have distinct regulatory controls and probably diverse functions during seed maturation. Additionally, they may be involved in protecting embryos during seeds desiccation tolerance, dormancy, and high temperature stress.

**Seed Storage Related Proteins**

In J. curcas seeds, the nutrients are primarily deposited inside the endosperm and relatively less in the embryo itself (de Lopes et al., 2013). We identified 21 SSPs (Table 4) belonging to different classes. Five isoforms of legumins were identified which consist of six subunit pairs that interact non-covalently. Each subunit consists of an acidic (α-subunit 30–40 kDa) and a basic (β-subunit 20 kDa) unit joined covalently through a single disulfide bond. These chains are assembled within the protein bodies, yielding the mature forms and deposited in a particular temporal order with 7S (trimeric) and 11–12S globulins (hexameric) (Gruis et al., 2002). Six isoforms of 2S albumins including four embryo specific isoforms were identified here (Table 4). These are cysteine-containing water-soluble proteins found in a wide variety of dicotyledonous seeds having a protective role against fungus. The amino acids profile of these proteins from multiple plant species has shown a higher level of sulfur-containing amino acids (Moreno and Clemente, 2008) and identified as potential food allergens. Albumins were also reported from the endosperm (Shah et al., 2015) as well as the inner integument of J. curcas seeds (Soares et al., 2014).

We have identified two isoforms of vicilin-like SSPs (Jcr4S01353.60, Jcr4S15278.20) that belong to the cupin superfamily (nutrient reservoirs). They are highly diverse in
| Protein accession | Description | Best hit Ricinus | Best hit Arabidopsis |
|-------------------|-------------|-----------------|----------------------|
| Jcr4s00582.50     | Calreticulin (CRT) | B9RF12         | AT1G56340.2         |
| Jcr4s03149.70     | Calnexin (CNX)     | B9RA41         | AT5G61790.1         |
| Jcr4s08787.20     | Heat shock cognate 70 kDa protein | B9SR13         | AT3G12580.1         |
| Jcr4s00002.460    | Heat shock 70 kDa protein | B9RGN3         | AT1G16030.1         |
| Jcr4s02598.30     | Chaperone (mtHsc70) | B9RX55         | AT5G00950.1         |
| Jcr4s11847.20     | Chaperone (Hsp70)  | B9T228         | AT5G02500.1         |
| Jcr4s07413.10     | Chaperone (Hsp70)  | B9SP17         | AT5G02500.1         |
| Jcr4s03314.10     | Chaperone (Hsp70)  | B9T228         | AT5G02500.1         |
| Jcr4s01616.40     | Chaperone (Hsp70)  | B9RGN3         | AT1G16030.1         |
| Jcr4s01616.30     | Chaperone (Hsp70)  | B9RGN3         | AT1G16030.1         |
| Jcr4s00047.50     | HSP70-chaperone (BiP) | B9SQC9         | AT1G79930.1         |
| Jcr4s03592.10     | Chaperone (Hsp90)  | B9T1Q8         | AT5G56000.1         |
| Jcr4s02234.80     | Chaperone (Hsp90)  | B9R1T0         | AT5G52640.1         |
| Jcr4s00535.40     | Chaperone (Hsp90)  | B9RCG8         | AT5G52640.1         |
| Jcr4s00006.110    | Endoplasmic, putative (Hsp90) | B9BA7         | AT4G24190.1         |
| Jcr4s04936.50     | Chaperonin-60 kD, ch60, putative | B9SS82         | AT3G23990.1         |
| Jcr4s00065.20     | Chaperonin-60 kD  | B9SBN5         | AT3G13470.1         |
| Jcr4s03964.10     | Chaperonin-60 alpha | B9T7 × 8      | AT2G28000.1         |
| Jcr4s02414.50     | Chaperonin CPN60-2, mitochondrial | B9RW2         | AT3G23990.1         |
| Jcr4s00940.30     | Groes chaperonin, putative | B9SJS50      | AT1G23100.1         |
| Jcr4s00187.50     | Chaperone protein CipB1 | B9RLP7        | AT1G74310.1         |
| Jcr4s00077.20     | Nascent polyepptide associated complex alpha subunit, putative | B9SHV0        | AT3G12390.1         |
| Jcr4s01075.10     | Heat shock 70 kDa protein, putative | B9S3M9         | AT4G16660.1         |
| Jcr4s02348.10     | Nascent polyepptide associated complex alpha subunit, putative | B9SSM4         | AT3G49470.1         |
| Jcr4s07320.10     | Jun activation domain binding protein, putative | B9SP01         | AT1G71230.1         |
| Jcr4s00045.200    | Ubiquitin, putative | B9R0I09        | AT5G20620.1         |
| Jcr4s00168.60     | Ubiquitin, putative | B9R0I09        | AT5G03240.3         |
| Jcr4s00518.10     | Ubiquitin, putative | B9R0I09        | AT4G05320.6         |
| Jcr4s00540.60     | Ubiquitin, putative | B9R0I09        | AT5G03240.3         |
| Jcr4s01667.40     | Ubiquitin, putative | B9R0I09        | AT1G31340.1         |
| Jcr4s03519.90     | Ubiquitin, putative | B9R0I09        | AT4G05050.3         |
| Jcr4s04671.10     | Ubiquitin, putative | B9R0I09        | AT1G31340.1         |
| Jcr4s25965.10     | Ubiquitin, putative | B9R0I09        | AT4G05050.3         |
| Jcr4s02001.10     | Ubiquitin         | B9SVD7         | AT3G52590.1         |
| Jcr4s02833.30     | Ubiquitin         | B9RN74         | AT2G47110.2         |
| Jcr4s08473.50     | Ubiquitin         | B9SHC5         | AT2G47110.2         |
| Jcr4s030993.20    | Small ubiquitin-related modifier (SUMO) | B9S1S0         | AT4G26840.1         |
| Jcr4s01038.60     | Small ubiquitin-related modifier (SUMO) | B9S1S0         | AT5G55160.1         |
| Jcr4s00385.110    | Regulatory component RPT5 of 26S proteasome | B9RFBS        | AT3G05530.1         |
| Jcr4s02802.40     | 20S Proteasome subunit alpha 2 | B9R0I05        | AT1G16470.2         |
| Jcr4s02462.10     | E3 Ubiquitin-protein ligase | B9SM23         | AT1G70320.1         |
| Jcr4s00009.170    | Transitional endoplasmic reticulum ATPase | B9R2Y1         | AT5G03340.1         |
| Jcr4s01757.40     | Transitional endoplasmic reticulum ATPase | B9S01         | AT5G03340.1         |
| Jcr4s02560.50     | Transitional endoplasmic reticulum ATPase | B9S03         | AT5G03340.1         |
| Jcr4s00024.10     | Class-C-I small heat-shock-responsive protein | B9S395         | AT1G07400.1         |
| Jcr4s00071.90     | Class-M-I small heat-shock-responsive protein | B9SGS1         | AT4G25200.1         |
| Jcr4s01251.90     | Class-C-I small heat-shock-responsive protein | B9SKB2         | AT1G07400.1         |
| Jcr4s04545.30     | Class-C-I small heat-shock-responsive protein | B9SS390        | AT1G53540.1         |
| Jcr4s10013.10     | Class-C-I small heat-shock-responsive protein | B9SS395        | AT1G07400.1         |
| Jcr4s11638.20     | Class-C-I small heat-shock-responsive protein | B9SSW04        | AT1G53540.1         |
| Jcr4s18690.10     | Class-C-I small heat-shock-responsive protein | B9SS395        | AT1G07400.1         |
| Jcr4s02800.10     | Class-C-I small heat-shock-responsive protein | B9SS395        | AT1G07400.1         |
| Jcr4s03109.50     | Class-M-II small heat-shock-responsive protein | B9RV59        | AT1G52560.1         |
the terms of polypeptide composition and involved in the plant defense response. These proteins have been reported from the embryo of *A. angustifolia* mature seeds, in which their accumulation was related to cotyledon differentiation and a defensive function against insect predation (de Sales et al., 2001).

Furthermore, we revealed the isoforms of antimicrobial proteins, such as vicilin-like (Jcr4s17767.20) and non-specific lipid-transfer proteins (nsLTPs) (Jcr4s20386.10 and Jcr4s00353.50). They exhibit antimicrobial activities because of their ability to permeabilize the cell membrane of phytopathogens (Scheuer and Schülke, 2018). This class of proteins was used as a biomarker to explore culture conditions during *Elaeis guineensis* somatic embryos maturation (Morcillo et al., 2001). The nsLTP was reported from the tomato endosperm and found to be involved in the transition of lipids from endosperm to embryo. They are also involved in the synthesis of a protective coating of cutin and suberin over the plant surface, and the defense against pathogens during seed germination (Scheuer and Schülke, 2018). Identifying these proteins in our analysis suggest their role in embryogenesis, seed maturation, lipid mobilization, signaling, and direct defense against pathogens during germination.

Besides these SSPs, we identified isoforms of the late embryogenesis abundant (LEA) (Table 4) proteins. These LEA proteins are synthesized during various phases of late embryogenesis in the seed embryo and under varying stress conditions, such as desiccation. The expression and deposition of LEA proteins in embryos indicate seed maturation and correlate water deficiency in various plant tissues and seed dehydration (Olvera-Carrillo et al., 2011). Since *J. curcas* seeds are orthodox type (desiccation resistant), LEA protein might be one of the major proteins contributing to the induction of this capacity to preserve seeds viability.

### Proteases

In the present study, 50 proteases and their inhibitors belonging to different mechanistic classes were identified (Table 5). Aspartic proteases (APs) are the most abundant class of peptidases, followed by Metallo (MPs), serine (SPs), and cysteine proteases (CPs) (Figure 3). APs were involved in the proteolysis and mobilization of reserve proteins (gliadin and globulin) in germinating wheat and rice seeds. In castor seeds, the involvement of APs in conjunction with vacuolar processing enzyme was proposed for the proteolytic processing and maturation of pro2S albumin pro-peptide (Hiraiwa et al., 1997). Six isoforms of 2S albumin protein were identified here (Table 4) which indicate that in the embryo of *J. curcas* these APs might be involved in their maturation. It was also hypothesized that during PCD, APs could be the part of nucellar cells deterioration and be involved in the synthesis of new proteins from nucellar cell death proteins for the growth of embryo and endosperm in barley (Chen and Foolad, 1997). Previously, we identified
six homologs of *Arabidopsis* phytepsins from the endosperm of *J. curcas* seeds (Shah et al., 2015), however, in the present study only one isoform of phytepsin (Jcr4S00547.20) was identified, suggesting that similar isomers of the enzyme may be active in different tissues of this plant. The identification of APs in our data suggests their multiple roles, such as maturation of reserves, 

### TABLE 5 | A list of peptidases and their inhibitors identified from the embryo of mature *J. curcas* seeds.

| Protein accession | MEROPS ID | Description | Catalytic type |
|-------------------|-----------|-------------|---------------|
| Jcr4S00063.130    | MER412334 | Aspartic proteinase, putative | Aspartic |
| Jcr4S00107.20     | MER178448 | Uncharacterized protein | Aspartic |
| Jcr4S00547.20     | MER496613 | Aspartic proteinase (phytepsin) putative | Aspartic |
| Jcr4S00739.60     | MER170824 | Aspartic proteinase nepenthein-1, putative | Aspartic |
| Jcr4S01204.30     | MER200104 | Uncharacterized protein | Aspartic |
| Jcr4S01708.30     | MER200104 | Uncharacterized protein | Aspartic |
| Jcr4S02365.20     | MER177024 | Poly [ADP-ribose] polymerase (PARP) | Aspartic |
| Jcr4S02367.30     | MER168322 | CCHC-type domain-containing protein | Aspartic |
| Jcr4S02531.20     | MER200104 | Uncharacterized protein | Aspartic |
| Jcr4S03044.70     | MER200104 | Uncharacterized protein | Aspartic |
| Jcr4S06099.10     | MER200104 | Uncharacterized protein | Aspartic |
| Jcr4S06992.10     | MER165268 | Chaperone binding protein, putative | Aspartic |
| Jcr4S07071.50     | MER161139 | CCHC-type domain-containing protein | Aspartic |
| Jcr4S07203.10     | MER165958 | Uncharacterized protein | Aspartic |
| Jcr4S08659.10     | MER200073 | Uncharacterized protein | Aspartic |
| Jcr4S09276.10     | MER199445 | Uncharacterized protein | Aspartic |
| Jcr4S09731.30     | MER201206 | Uncharacterized protein | Aspartic |
| Jcr4S10915.10     | MER200446 | Retrotrans gag domain-containing protein | Aspartic |
| Jcr4S11449.10     | MER176399 | Uncharacterized protein | Aspartic |
| Jcr4S16426.20     | MER199779 | Uncharacterized protein | Aspartic |
| Jcr4S16499.10     | MER164887 | Uncharacterized protein | Aspartic |
| Jcr4S19565.30     | MER200104 | Uncharacterized protein | Aspartic |
| Jcr4S20272.10     | MER199779 | Uncharacterized protein | Aspartic |
| Jcr4S27524.10     | MER199659 | Glyceraldehyde-3-phosphate dehydrogenase | Aspartic |
| Jcr4S28308.20     | MER199779 | Uncharacterized protein | Aspartic |
| Jcr4S00009.170    | MER269263 | Transitional endoplasmic reticulum ATPase, putative | Metallo |
| Jcr4S00309.10     | MER413614 | Transitional endoplasmic reticulum ATPase, putative | Metallo |
| Jcr4S00343.100    | MER172798 | Leucine aminopeptidase, putative | Metallo |
| Jcr4S00385.110    | MER278187 | 26S Protease regulatory subunit 6a, putative | Metallo |
| Jcr4S01168.90     | MER172955 | Oligopeptidase A (TOP1 peptidase), putative | Metallo |
| Jcr4S01757.40     | MER172955 | Oligopeptidase A (TOP1 peptidase), putative | Metallo |
| Jcr4S02560.50     | MER278187 | Transitional endoplasmic reticulum ATPase, putative | Metallo |
| Jcr4S03720.10     | MER143437 | Jun activation domain binding protein, putative | Metallo |
| Jcr4S00788.120    | MER157367 | WD-repeat protein, putative | Serine |
| Jcr4S00899.60     | MER178907 | Transporter, putative | Serine |
| Jcr4S00913.60     | MER148853 | Nucleotide binding protein, putative | Serine |
| Jcr4S01323.70     | MER135263 | WD-repeat protein, putative | Serine |
| Jcr4S01500.30     | MER170587 | Tripeptidyl-peptidase II | Serine |
| Jcr4S02039.30     | MER171392 | Xylem serine proteinase 1, putative | Serine |
| Jcr4S00072.60     | MER038019 | Stem-specific protein TSJT1, putative | Cysteine |
| Jcr4S01609.40     | MER158866 | Cysteine protease (RD21), putative | Cysteine |
| Jcr4S02110.50     | MER178955 | Caspase (metacaspase), putative | Cysteine |
| Jcr4S04275.10     | MER126050 | Uncharacterized protein | Cysteine |
| Jcr4S16229.10     | MER292732 | Cysteine protease, putative | Cysteine |
| Jcr4S00079.140    | MER180116 | Protein Z (AtSerpin1), putative | Unknown |
| Jcr4S02989.80     | MER202315 | Cysteine proteinase inhibitor (phytocystatin) | Unknown |
| Jcr4S17412.10     | MER169281 | Cysteine proteinase inhibitor B (cystatin), putative | Unknown |
| Jcr4S28633.30     | MER200074 | Uncharacterized protein | Unknown |
| Jcr4S02820.40     | MER170530 | Proteasome subunit alpha type | Non-peptidase homolog |
| Jcr4S02821.10     | MER176444 | Proliferation-associated 2g4, putative | Non-peptidase homolog |
mobilization, and proteolytic hydrolysis of stored reserves during seed germination and in PCD.

We identified eight MPs including Arabidopsis homolog of thimet oligopeptidase known as TOP1 peptidase (Jcr4S01168.90), constituting a class of salicylic acid (SA) binding proteins (Moreau et al., 2013). As SA plays important role in plant immune response, its identification in our data suggests its interaction with SA. It may play important role in defense mechanisms to protect the embryo from various types of biotic and abiotic stresses. FtsH endopeptidases were identified that are ATP-dependent zinc MPs and are embryo-specific associated with a broad range of cellular functions. The identification of MPs indicates that these proteins might be responsible for the rapid turnover and stability of large protein complexes to maintain cellular homeostasis inside the embryo of J. curcas seeds.

In addition, five CPs were identified here (Table 5) including 2 papain like proteases (Jcr4S01609.40 and Jcr4S16229.10). As seeds are the major sites for structural, metabolic, defensive, and reserve proteins, PLCPs are primarily involved in the mobilization and degradation of these proteins during seed germination. In the growing seedlings of wheat and maize, CPs account for 90% of the overall protease activity of prolamsins (Grudkowska and Zagdańska, 2004). During germination, seedling requires nutrients and energy, provided by the seed reserves. In this context, the identification of these proteases indicates their involvement in the mobilization and degradation of reserves during seed germination.
Plants lack caspases homologs but contain a phylogenetically distinct class of CPs known as metacaspases. These are also identified in our data (Table 5). Although the exact function of metacaspases in the PCD mechanism is still unclear, their differential expression was found to affect the PCD in plant embryos (Suarez et al., 2004). Thus, the identification of metacaspase indicates its significance in PCD during embryonic pattern formation and also presents a link between PCD and plant embryogenesis. Furthermore, a previous study reported that KDEL-tailed cysteine peptidases along with vacuolar processing enzymes (VPEs) were responsible for the occurrence of PCD in the endosperm (Shah et al., 2015) and inner integument (Soares et al., 2014) of J. curcas seeds. These results collectively reveal the involvement of a variety of peptidases in PCD in the different tissues of J. curcas seeds.

Additionally, we identified proteinase inhibitors (PIs), such as two cysteine protease inhibitors, namely, cystatins/phytocystatins (Jcr4S02989.80 and Jcr4S17142.10) and one serine protease inhibitor, namely, serpin (Jcr4S00079.140) (Table 5). Cystatin is actively produced in growing seeds and vegetative storage tissues to outnumber CPs and support reserve proteins. A high cystatin/CPs balance retained in quiescent tissues enables the pool of stored proteins to be preserved over dormancy and made available upon seed germination. After seed imbibition, the upregulation of CPs encoding genes while the downregulation of cystatin encoding genes causes a sharp decrease in cysteine/cystatin protease balance which favors the mobilization and hydrolysis of stored proteins in amaranth and Arabidopsis (Hwang et al., 2009). The identification of cystatin in this study suggests its role in regulating CPs activity to prevent an unscheduled hydrolysis of SSFs as well as protecting the embryo from the exogenous CPs of phytopathogens. Serpins inhibit cysteine proteases specially RD21 cysteine protease (Jcr4S01609.40) which is also identified here. Such proteins are concerned with desiccation response and pathogenic defense (Lampl et al., 2010). Serpin and RD21 cysteine protease were identified from the endosperm of J. curcas seeds (Shah et al., 2015). The identification of serpin in our data reveals its role to retain RD21 inactive until required during germination.

Transport Related Proteins

Besides other important proteins, we identified 109 different transport-related proteins corresponding to almost 19% of the total identifications (Supplementary Table 2). Based on TCDB database information, these proteins belong to seven different classes. Channel/pores are the most representative class of transporters identified, followed by primary active transporters, accessory factors involved in transport, electrochemical potential-driven transporters, incompletely characterized transport systems, transport electron carriers, and group translocators (Figure 4). We identified 45 members of the channels/pores family of transporters (Supplementary Table 2). Among them, 10 transporters belong to the family of HSPs (HSF70) which are abundantly present in multiple living species. They are responsible for the transport of proteins in different cellular organelles and in translocating misfolded proteins toward the UPS for deterioration. They are also involved in the transport of many transmembrane proteins, aids in their folding, and protect them from stressful conditions (Trivedi et al., 2016). The identification of HSP70 in our study indicates their involvement in the translocation of misfolded and other functional proteins toward their target site to maintain cellular homeostasis in normal and stressful conditions. We identified Aquaporin PIP2 (Jcr4S02148.40), a channel protein that belongs to the membrane intrinsic protein family. They promote the passive movement of water and other solute molecules through the membrane, facilitated by osmotic or solute gradients. Genes coding for PIP along with other Aquaporin, i.e., PIP2, TIP1 were found to express in cotyledons and seed coats of growing pea seeds (Schuurmans et al., 2003). Though information regarding the functioning of PIPs during seed maturation is not yet available, however, 11 Aquaporins have been reported from the dry seeds of Arabidopsis (Vander Willigen et al., 2006). They are supposed to be pre-formed during the growth and maturation of seeds. Since mature seeds are dried and require water for germination, the identification of Aquaporin in our study suggests that they might be responsible for the transport of water and probably other solute molecules in the embryo to facilitate efficient seed germination.

Five members of electrochemical potential-driven transporters were identified here (Supplementary Table 2). They include amino acid transporters isoforms (Jcr4S90615.20 and Jcr4S07701.10), previously not identified from this species. These transporters are important for the transport of amino acids to cells and organelles, such as mitochondria, peroxisomes, chloroplast, and vacuole to ensure an equal supply of organic nitrogen (Dinkeloo et al., 2018). In addition, studies indicated the important functions of various transporters, i.e., AAP1 and AAP8 in Arabidopsis during the embryonic distribution of amino acids to promote the seed growth and reserves deposition. Among them, particularly AAP1 was found to be involved in the transfer of amino acids toward the embryo. AAP1 deficient embryos revealed the lower import of amino acids leading to the lower accumulation of reserve proteins (Sanders et al., 2009). Similarly, it was hypothesized that the upregulation of AAP1 encoding gene ensued increased deposition of reserves in pea seeds broad beans (Miranda et al., 2001). The presence of amino acid transporter indicates that these may be embryo-specific and present a positive correlation with the accumulation of SSFs identified here. The appearance of these proteins indicates their role in regulating the synthesis of SSFs and reveals their importance for the completion of the nitrogen cycle in seeds.

CONCLUSION

In this study, we presented a proteome analysis of the embryo isolated from the mature J. curcas seeds. This analysis resulted in the identification of 564 proteins of which 206 proteins were not identified from other tissue of this plant so far. We were able to identify the proteins responsible for the provision of carbon source and energy, including but not limited to carbohydrate metabolism, lipids metabolism, transport-related proteins, and proteases of different mechanistic classes. In the case of J. curcas...
seeds, nutrients are primarily deposited inside the endosperm and relatively less in the embryo. However, the identification of several classes of the SSPs, including the members of legumins and vicilins as well as the transporters, such as amino acids transporters indicates the potential of this tissue for supporting the germination process in the terms of the provision of nutrients. Our analysis furnishes insight into the important pathways and certain unique features of the embryo from the mature seeds of J. curcas. The findings presented here are a step forward in our efforts to create a proteome catalog of J. curcas seeds, which will serve as a significant resource for studies on the developmental biology of oilseeds.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: Mass spectrometry raw data files are available on the Chorus repository web site (chorusproject.org; project ID: 1747, project name: Jatropha curcas Embryo Proteome).

AUTHOR CONTRIBUTIONS

MS, MA, and FN designed the experiment and wrote the manuscript. AR, NU, Sheheryar, and JN performed the experiments and analyzed the data. FC and GD proofread the manuscript. All authors read and approved of its content, read and agreed to the published version of the manuscript.

REFERENCES

Alonso, A. P., Goffman, F. D., Ohlrogge, J. B., and Shachar-Hill, Y. (2007). Carbon conversion efficiency and central metabolic fluxes in developing sunflower (Helianthus annuus L.) embryos. Plant J. 52, 296–308. doi: 10.1111/j.1365-313X.2007.02325.x

Asif, M. H., Mantri, S. S., Sharma, A., Srivastava, A., Trivedi, I., Gupta, P., et al. (2010). Complete sequence and organisation of the Jatropha curcas (Euphorbiaceae) chloroplast genome. Tree Genet. Genomes 6, 941–952. doi: 10.1007/s11295-010-0363-300

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248–254. doi: 10.1016/0003-2697(76)90527-3

Chan, A. P., Crabtree, J., Zhao, Q., Lorenzi, H., Orvis, J., Puiu, D., et al. (2010). Draft genome sequence of the oilseed species Ricinus communis. Nat. Biotechnol. 28, 951–956. doi: 10.1038/nbt.1674

Chen, F., and Foolad, M. R. (1997). Molecular organization of a gene in barley which encodes a protein similar to aspartic protease and its specific expression in nucellar cells during degeneration. Plant Mol. Biol. 35, 821–831. doi: 10.1023/a:1005833207007

Chen, M. S., Wang, G. J., Wang, R. L., Wang, J., Song, S. Q., and Xu, Z. F. (2011). Analysis of expressed sequence tags from biodiesel plant Jatropha curcas embryos at different developmental stages. Plant Sci. 181, 696–700. doi: 10.1016/j.plantsci.2011.08.003

Chia, T. Y. P., Pike, M. J., and Rawsthorne, S. (2005). Storage oil breakdown during embryo development of Brassica napus (L.). J. Exp. Bot. 56, 1285–1296. doi: 10.1093/jxb/eri129

Cornah, J. E., and Smith, S. M. (2002). “Synthesis and function of glycolylyte cycle enzymes,” in Plant Peroxisomes, eds A. Baker and I. A. Graham (Berlin: Springer), 57–101.

Costa, G. G. L., Cardoso, K. C., Del Bem, L. E. V., Lima, A. C., Cunha, M. A. S., de Campos-Leite, L., et al. (2010). Transcriptome analysis of the oil-rich seed of the bioenergy crop Jatropha curcas L. BMC Genomics 11:462. doi: 10.1186/1471-2164-11-462

de Lopes, L. S., Gallão, M. I., and de Bertini, C. H. C. M. (2013). Mobilisation of reserves during germination of Jatropha seeds. Rev. Cien. Agron. 44, 371–378. doi: 10.1590/S1806-66902013000200021

de Sales, M. P., Pimenta, P. P., Paes, N. S., Grossi-de-Sá, M. F., and Xavier-Filho, J. (2001). Vicilins (7S storage globulins) of cowpea (Vigna unguiculata) seeds bind to chitinous structures of the midgut of Callosobruchus maculatus (Coleoptera: Bruchidae) larvae. Brazilian J. Med. Biol. Res. 34, 27–34. doi: 10.1590/S0100-879x2001000100003

De Smet, L., Lau, S., Mayer, U., and Jürgens, G. (2010). Embryogenesis—the humble beginnings of plant life. Plant J. 61, 959–970.

Dinkeloo, K., Boyd, S., and Pilot, G. (2018). Update on amino acid transporter functions and on possible amino acid sensing mechanisms in plants. Sem. Cell Dev. Biol. 74, 105–113. doi: 10.1016/j.semcdb.2017.07.010

Grudkowska, M., and Zagdańska, B. (2004). Multifunctional role of plant cysteine proteinases. Acta Biochim. Pol. 51, 609–624.

Gruis, D. F., Selinger, D. A., Curran, J. M., and Jung, R. (2002). Redundant proteolytic mechanisms process seed storage proteins in the absence of seed-type members of the vacuolar processing enzyme family of cysteine proteases. Plant Cell 14, 2863–2882. doi: 10.1105/tpc.005009

Gu, K., Yi, C., Tian, D., Sangha, J., Hong, Y., and Yin, Z. (2012). Expression of fatty acid and lipid biosynthetic genes in developing endosperm of Jatropha curcas. Biotechnol. Biofuels 5:47. doi: 10.1186/1754-6834-5-47

Hill, L. M., Morley-Smith, E. R., and Rawsthorne, S. (2003). Metabolism of sugars in the endosperm of developing seeds of oilseed rape. Plant Physiol. 131, 228–236. doi: 10.1104/pp.010868

Hiraiwa, N., Kondo, M., Nishimura, M., and Hara-Nishimura, I. (1997). An aspartic endopeptidase is involved in the breakdown of propeptides of storage proteins in protein-storage vacuoles of plants. Eur. J. Biochem. 246, 133–141. doi: 10.1111/j.1432-1033.1997.00133.x

Hwang, J. E., Hong, J. K., Je, J. H., Lee, K. O., Kim, D. Y., Lee, S. Y., et al. (2009). Regulation of seed germination and seedling growth by an Arabidopsis phytocystatin isoform, AtCYS6. Plant Cell Rep. 28, 1623–1632. doi: 10.1007/s00299-009-0762-7

Joshee, N. (2006). Book review. Plant Sci. 170:415. doi: 10.1016/j.plantsci.2005.11.003

FUNDING

The authors acknowledge Taif University Researchers Supporting Project Number TURSP-2020/91, Taif University, Taif, Saudi Arabia.

ACKNOWLEDGMENTS

We are highly thankful to the Laboratory of Proteomics/LADETEC, Federal University of Rio de Janeiro, for providing mass spectrometry facility.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.843764/full#supplementary-material

Supplementary Table 1 | A list of proteins and peptides identified from the proteome analysis of the embryo isolated from mature Jatropha curcas seeds.
Komaner, D., and Rape, M. (2012). The ubiquitin code. Annu. Rev. Biochem. 81, 203–229.

Lampe, N., Budai-Hadrian, O., Davydov, O., Joss, T. V., Harrop, S. J., Curmi, P. M. G., et al. (2010). Arabidopsis AtSerpin1, crystal structure and in vivo interaction with its target protease RESPONSIVE TO DESICCATION-21 (RDS21). J. Biol. Chem. 285, 33540–33545. doi: 10.1074/jbc.M109.907579

Liu, H., Liu, Y., Yang, M. F., and Shen, S. H. (2009). A comparative analysis of embryo and endosperm proteome from seeds of Jatropha curcas. J. Integr. Plant Biol. 51, 850–857. doi: 10.1111/j.1744-7909.2008.00839.x

Liu, H., Yang, Z., Yang, M., and Shen, S. (2011). The differential proteome of endosperm and embryo from mature seed of Jatropha curcas. Plant Sci. 181, 660–666. doi: 10.1016/j.plantsci.2011.03.012

Miernyk, J. A., Prešová, A., Olmedilla, A., Klubíková, K., Obert, B., and Hajduch, M. (2011). Using proteomics to study sexual reproduction in angiosperms. Sex. Plant Reprod. 24, 9–22. doi: 10.1007/s00497-010-0149-5

Miranda, M., Borisjuk, L., Tews, A., Heim, U., Sauer, N., Wobus, U., et al. (2001). Amino acid permeases in developing seeds of Vicia faba L.: expression precedes storage protein synthesis and is regulated by amino acid supply. Plant J. 28, 61–71. doi: 10.1046/j.1365-313x.2001.01129.x

Miro, B., Longkumer, T., Entila, F. D., Kohli, A., and Ismail, A. M. (2017). Rice seed germination underwater: morpho-physiological responses and the bases of differential expression of alcoholic fermentation enzymes. Front. Plant Sci. 8:1857. doi: 10.3389/fpls.2017.01857

Morcillo, F., Hartmann, C., Duval, Y., and Tregear, J. W. (2001). Regulation of 7S globulin gene expression in zygotic and somatic embryos of oil palm. Physiol. Plant. 112, 233–243. doi: 10.1046/j.1399-3054.2001.112012.x

Moreau, M., Westlake, T., Zamponga, G., Popescu, G., Tian, M., Noutsos, C., et al. (2013). The Arabidopsis oligopeptidases TOP 1 and TOP 2 are salicylic acid targets that modulate SA-mediated signaling and the immune response. Plant J. 76, 603–614. doi: 10.1111/tpj.12320

Moréno, F. J., and Clemente, A. (2008). 2S albumin storage proteins: a reversible enzyme reaction with an important role in metabolic control. Plant Physiol. 142, 1193–1200. doi: 10.1104/pp.108.11193

Pinheiro, C. B., Shah, M., Soares, E. L., Nogueira, F. C. S., Carvalho, P. C., Soares, A. A., Domont, G. B., Nogueira, F. C. S., et al. (2015). Proteomic analysis of the endosperm ontogeny of Jatropha curcas L. seeds. J. Proteome Res. 14, 2557–2568. doi: 10.1021/acs.jproteome.5b00106

Shah, M., Soares, E. L., Lima, M. L. B., Pinheiro, C. B., Soares, A. A., Domont, G. B., et al. (2016). Deep proteome analysis of gerontoplasts from the inner integument of developing seeds of Jatropha curcas L. J. Proteom. 143, 346–352. doi: 10.1016/j.jprot.2016.02.025

Siloto, R. M. P., Findlay, K., Lopez-Villalobos, A., Yeung, E. C., Nykiforuk, C. L., and Moloney, M. M. (2006). The accumulation of oleosins determines the size of seed oilbodies in Arabidopsis. Plant Cell 18, 1961–1974. doi: 10.1105/tpc.106.041269

Soares, E. L., Shah, M., Soares, A. A., Costa, J. H., Carvalho, P., Domont, G. B., et al. (2014). Proteome analysis of the inner integument from developing Jatropha curcas L. seeds. J. Proteome Res. 13, 3562–3570. doi: 10.1021/pr5004505

Santos, M. E., Milonova, I. H., Smertenko, A., Savenkov, E. I., Clapham, D. H., von Arnold, S., et al. (2004). Metacaspase-dependent programmed cell death is essential for plant embryogenesis. Curr. Biol. 14, R339–R340. doi: 10.1016/j.cub.2004.04.019

Trivedi, D. K., Huda, K. M. K., Gill, S. S., and Tuteja, N. (2016). “Molecular chaperone structure, function, and role in plant abiotic stress tolerance,” in Abiotic Stress Response Plants, eds N. Tuteja and S. S. Gill (Weinheim: Wiley-VCH Verlag GmbH & Co.).

Vander Willigen, C., Postaire, O., Tournaire-Roux, C., Bourisya, Y., and Maurel, C. (2006). Expression and inhibition of aquaporins in germinating Arabidopsis seeds. Plant Cell Physiol. 47, 1241–1250. doi: 10.1093/pcp/pcp094

Vasconcelos, É. A., Nogueira, F. C. S., Abreu, E. F. M., Gonçalves, E. F., Souza, P. A. S., and Campos, F. A. P. (2005). Protein extraction from cowpea tissues for 2-D gel electrophoresis and MS analysis. Chromatographia 62, 447–450. doi: 10.1365/s10337-005-0637-6

Vigolos, H., Waldeck, P., Zank, T., and Geigenberger, P. (2007). Increasing seed oil content in oil-seed rape (Brassica napus L.) by over-expression of a yeast glycerol-3-phosphate dehydrogenase under the control of a seed-specific promoter. Plant Biotechnol. J. 5, 431–441.

Wang, W., Vinocur, B., Shoseyov, O., and Altman, A. (2004). Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends Plant Sci. 9, 244–252. doi: 10.1016/j.tplants.2004.03.006

Wehmeier, N., and Vierling, E. (2000). The expression of small heat shock proteins in seeds responds to discrete developmental signals and suggests a general protective role in desiccation tolerance. Plant Physiol. 122, 1099–1108. doi: 10.1104/pp.122.4.1099

Yang, P., Li, X., Wang, X., Chen, H., Chen, F., and Shen, S. (2007). Proteomic analysis of rice (Oryza sativa) seeds during germination. Proteomics 7, 3358–3368.

Young, J. C., Moarefi, I., and Hartl, F. U. (2001). Hsp90: a specialized but essential protein-folding tool. J. Cell Biol. 154, 267–274. doi: 10.1083/jcb.200104079

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ramzan, Shah, Ullah, Sheheryar, Nascimento, Campos, Domont, and Abdel-Aal. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.