SCIENTIFIC REP✲RTS

Proteomic analysis of phytase transgenic and non-transgenic maize seeds

Yanhua Tan[1,2], Zheng Tong[1], Qian Yang[1], Yong Sun[1], Xiang Jin[1], Cunzhi Peng[1], Anping Guo[1,2] & Xuchu Wang[1,2]

Proteomics has become a powerful technique for investigating unintended effects in genetically modified crops. In this study, we performed a comparative proteomics of the seeds of phytase-transgenic (PT) and non-transgenic (NT) maize using 2-DE and iTRAQ techniques. A total of 148 differentially expressed proteins (DEPs), including 106 down-regulated and 42 up-regulated proteins in PT, were identified. Of these proteins, 32 were identified through 2-DE and 116 were generated by iTRAQ. It is noteworthy that only three proteins could be detected *via* both iTRAQ and 2-DE, and most of the identified DEPs were not newly produced proteins but proteins with altered abundance. These results indicated that many DEPs could be detected in the proteome of PT maize seeds and the corresponding wild type after overexpression of the target gene, but the changes in these proteins were not substantial. Functional classification revealed many DEPs involved in posttranscriptional modifications and some ribosomal proteins and heat-shock proteins that may generate adaptive effects in response to the insertion of exogenous genes. Protein-protein interaction analysis demonstrated that the detected interacting proteins were mainly ribosomal proteins and heat-shock proteins. Our data provided new information on such unintended effects through a proteomic analysis of maize seeds.

Genetically modified crops (GMCs) have been globally commercialized since 1996, and the cumulative planting area of GMCs has reached an unprecedented 2 billion hectares globally in 2015. A total of 28 countries allow planting of GMCs[1]. Although GMCs have delivered substantial benefits to farmers and our society, it is still debated whether GMCs should be widely accepted in many countries, including China, because of their unintended, unexpected and uncontrolled negative effects[2]. Therefore, determining the potential unintended effects of GMCs is important to understand the benefits of these crops to farmers and consumers.

Proteomics (protein profiling) is a powerful omics-based technique for investigating unintended effects at the protein level[2]. Proteomics is increasingly used because of its accuracy in describing protein diversity[3, 4]. Many proteins cannot be determined based only on the genes encoding their sequences, due to splicing events, enzymatic processing, or posttranslational modifications[5]. Traditionally, the high-resolution two-dimensional polyacrylamide gel electrophoresis (2-DE) technique has been used to detect potential unintended effects in proteomics research[6–10]. Recently, a non-gel-based protein analysis technique referred to as isobaric tags for relative and absolute quantification (iTRAQ) has been proposed for the quantitation of protein species and comparisons among samples. iTRAQ is a comparative proteomic approach that employs two-dimensional liquid chromatography (2D-LC) coupled directly with tandem MS (MS/MS) detection[11–13]. iTRAQ is considered more accurate and reliable than traditional 2-DE gel-based analysis[14–16] and is thus now widely used for comparative proteomic analyses.

Maize is an important crop used for both food and feed, and transgenic maize has attracted much attention worldwide. Researches on commercialization of transgenic maize in China have made important potential contributions to maize production for meeting both national and global food and feed needs. Although maize is imported into China in large quantities (3.3 million tons in 2015), domestic production of transgenic maize has not yet been implemented to date. Many efforts have been devoted to transgenic maize, including Bt maize, herbicide-tolerant maize and phytase maize, but phytase-transgenic (PT) maize, which shows great potential in

[1]Key Laboratory of Biology and Genetic Resources for Tropical Crops, Institute of Tropical Biosciences and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan, 571101, China. [2]College of Agriculture, Hainan University, Haikou, Hainan, 570228, China. Correspondence and requests for materials should be addressed to A.G. (email: guoanping@itbb.org.cn) or X.W. (email: xchwanghainan@163.com)
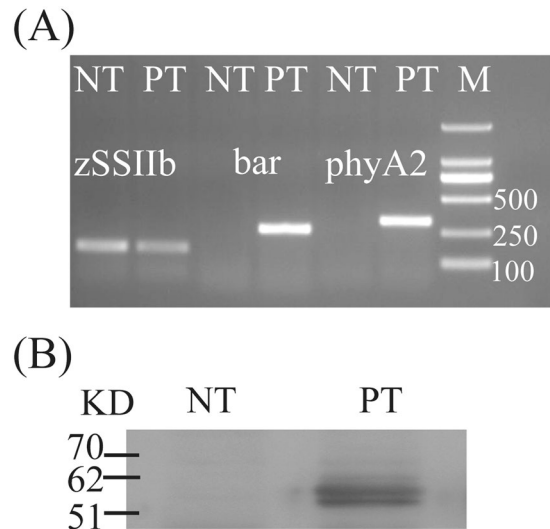
(A)

(B)

Figure 1. Detection of the exogenous gene expression and target proteins in maize seeds. (**A**) RT-PCR results for the exogenous genes in maize seeds; (**B**) Western blot analysis of target proteins in maize seeds.

the animal feed industry, is the only genetically modified maize variety that has been officially issued a biosafety certificate in China[17]. PT maize was produced by achieving seed-specific overexpression of *Aspergillus niger* phytase (*phy*A2), an enzyme catalyzing the release of phosphate from phytase, which improves phosphorus availability and reduces the effect of animal production on the environment[18]. Many studies have been performed to ensure the safety of PT maize, including the investigation of its nutritional value[19], used as livestock feed[20], and the associated arthropod communities in the field[21]. We also performed a comparative proteomic analysis of leaves from PT maize and the non-transgenic (NT) isogenic variety[22]. However, any untargeted effects on the seeds of PT maize have yet to be studied through a proteomic analysis.

In the present study, we used traditional gel-based 2-DE and newly developed gel-free iTRAQ approaches to compare the quantitative differences in the proteomes of maize seeds between PT maize and the NT isogenic variety to investigate any unintended effects in PT maize seeds.

Results

Identification of Exogenous Genes and the Target Protein in PT Maize.
The expression of exogenous genes was detected *via* semi-quantitative RT-PCR. The results presented in Fig. 1A show that the *bar* and *phy*A2 transcripts were detected in PT maize seeds. Neither the *phy*A2 nor *bar* transcript was detected in the NT control seeds of LIYU16. These results suggested that these exogenous genes were successfully expressed in PT maize seeds.

Western blotting analysis revealed a band of approximately 60 kDa in the PT maize seed samples (Fig. 1B), whereas the target band was not observed in the NT samples, indicating that the *phy*A2 protein accumulated only in these PT seeds.

Comparison of Protein Profiles on 2-DE Gels.
To compare the protein profiles of PT maize seeds and the corresponding wild-type variety, 2-DE maps of total proteins were obtained (Fig. 2 and Figure S1). Each sample was produced from three biological replicates, and for each biological replicate of the protein extracts, triplicate 2-DE gels were prepared as technical replicates. An average of 852 ± 94 spots were detected in the nine NT maize seed gels, while 826 ± 102 spots were detected in each of the nine PT maize seed gels, and there were 660 matching spots between the NT and PT maize seed gels (Fig. 2). Only spots showing a change >1.5-fold that were detected in all replicates were considered as differentially expressed proteins (DEPs). Analysis of the 2-DE images revealed 43 DEPs between the PT and NT maize seed samples (18 with higher and 25 with lower abundance compared with NT).

Protein Identification *via* MALDI TOF/TOF MS.
A total of 43 DEPs were selected for MALDI TOF/TOF MS analysis after in-gel digestion with trypsin[23], and 32 protein spots were successfully identified through MS/MS analysis. Among these positively identified protein spots, 11 were up-regulated, and 21 were down-regulated, as shown in Fig. 2. Protein identification was based on homology to *Zea mays* proteins. Several spots were identified as containing more than one protein *via* MS/MS, in which the protein with the highest score that matched several members of grouped proteins in maize was chosen for further functional analysis (Table S1). The proteins that were indicated as uncharacterized were subjected to BlastP (protein-protein Blast) analysis in NCBI (http://blast.ncbi.nlm.nih.gov/ Blast.cgi) to determine their protein identities.

Identification of Proteins in PT and NT maize seeds *via* iTRAQ.
After merging data from the four biological replicates, a total of 300,833 spectra were generated from two independent iTRAQ experiments. Mascot search results identified a total of 73,765 distinct peptides and 6,313 detectable proteins from maize.
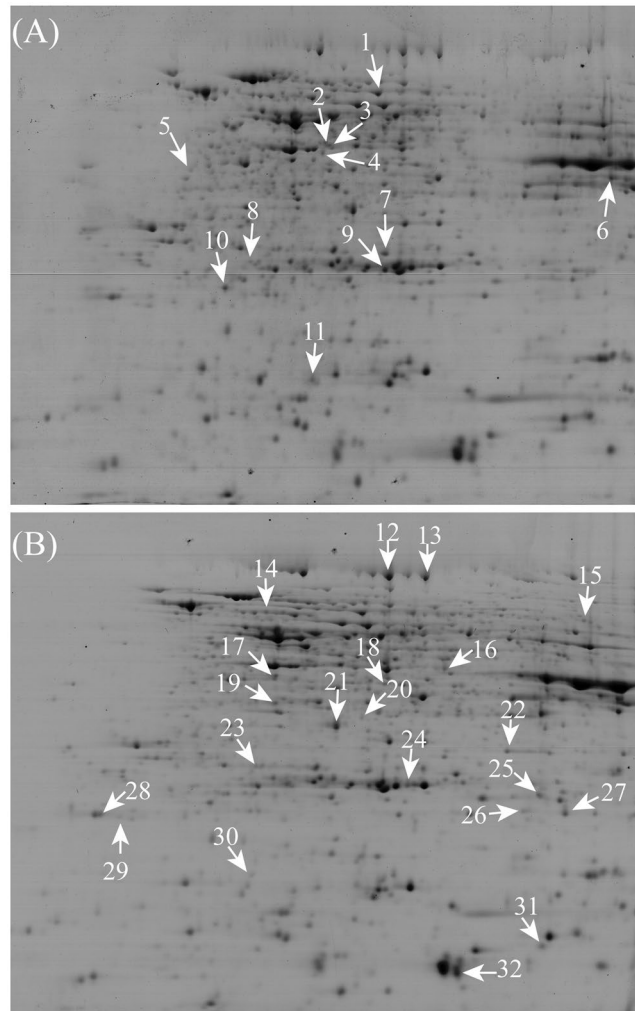
Figure 2. Typical 2-DE gels of total proteins from maize seeds. A total of 32 identified DEPs between PT and NT maize seeds were indicated with arrows in the gel images. The 11 DEPs with increased abundance (**A**) and 21 down-regulated DEPs (**B**) in PT maize seeds were highlighted in the 2-DE gels. The full-length original 2-DE gels were also presented in Supplementary Figure S1.

An analysis of variance of the data from two independent iTRAQ replicates was conducted to assess quantitative variations and reproducibility (Fig. 3A,B). The 114 and 115 tags indicate the two NT maize replicates, and 116 and 117 tags indicate the two PT maize replicates in each iTRAQ experiment. Thus, we compared the $\log_2$ iTRAQ ratios of 114:115 and 116:117 (Fig. 3). Plots of comparable quantification results for the 114:115 and 116:117 ratios were determined *via* linear regression analysis. The results showed a slope of ≈ 0.96 at an R^2 of 0.94 for 114:115 and ≈ 0.97 at an R^2 of 0.90 for 116:117, suggesting that the quantitative iTRAQ data from replicates were reproducible[24, 25].

Determination of DEPs by iTRAQ from the PT and NT maize seeds.
A total of 6313 non-redundant proteins were detected *via* iTRAQ under a false discovery rate (FDR) of less than 1%. Among all identified proteins, 116 DEPs between PT and NT maize seeds were obtained based on the following criteria: confident peptides ≥ 2, error factor < 2, change ratio > 1.5 and P-values < 0.05. Among the 116 DEPs, 31 proteins exhibited an increase of more than 1.5-fold in PT maize seeds, while 81 proteins exhibited a decrease to less than 0.667-fold in the PT maize seeds compared with NT seeds (Fig. 3C and Table S2).

These results, combined with those of 2-DE, showed a total of 148 identified DEPs, 42 of which showed higher expression, while 106 showed lower expression in PT than in NT.

Bioinformatics Analysis of DEPs Identified in iTRAQ and 2D Gels.
Bioinformatics analysis was performed for all 148 DEPs identified through 2-DE and iTRAQ analyses.

To identify the significantly enriched GO functional groups of DEPs between NT and PT maize seeds, GO annotation was performed using WEGO software (http://wego.genomics.org.cn/cgi-bin/wego/index.pl). GO information was obtained with BLAST2GO 3.0. The results showed that 141 DEPs were successfully mapped with GO annotations, which were classified into three ontologies containing 32 functional groups (Fig. 4 and Table S3). At the cellular GO level, 8 GO terms were obtained, corresponding to 35 proteins (23.8%) in the cell
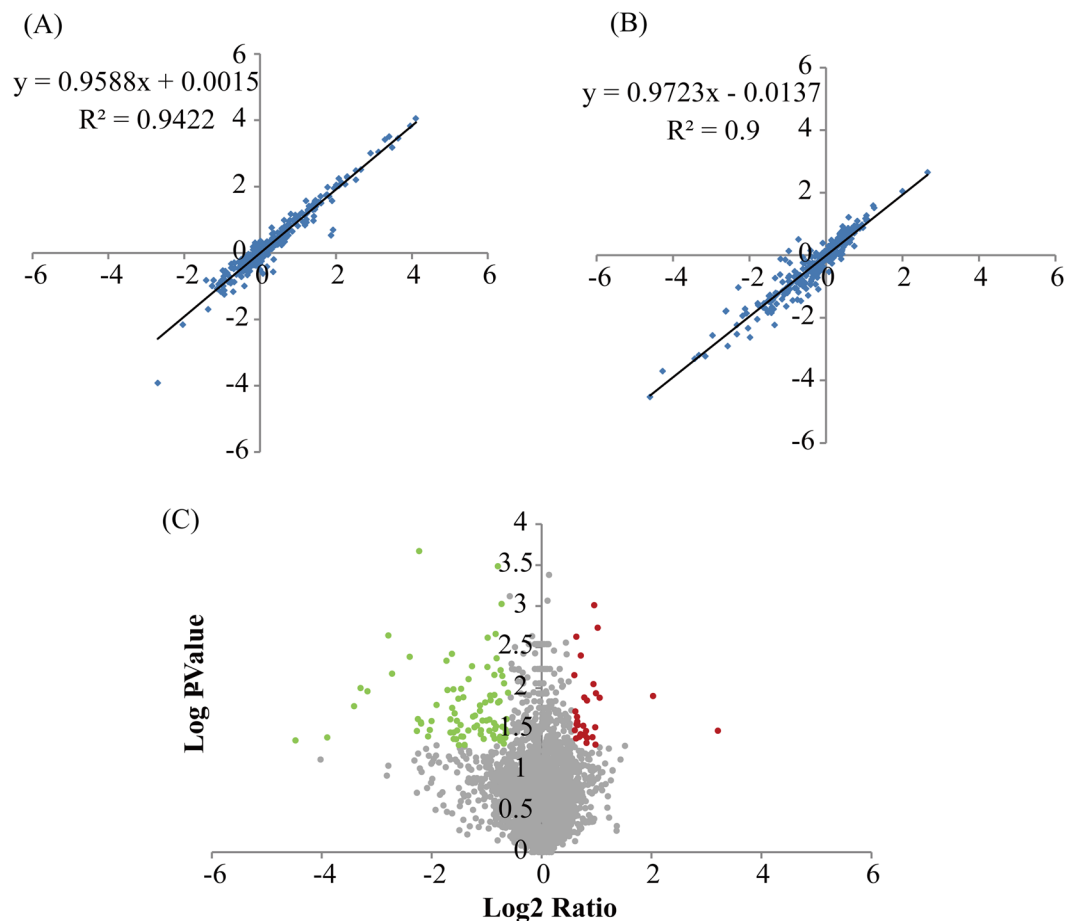
(A)

$y = 0.9588x + 0.0015$
$R^2 = 0.9422$

(B)

$y = 0.9723x - 0.0137$
$R^2 = 0.9$

(C)

Figure 3. Comparison of $\log_2$ ratios and the volcano plot for the identified proteins from iTRAQ. $\log_2$ values for the ratios of 114:115 in NT (**A**) and 116:117 ratios in PT (**B**) were used to show good replication in the iTRAQ experiments. The volcano plot shows the collected p-values (**C**) for the changes in the patterns of all identified proteins. Red spots represent up-regulated DEPs; gray spots are unchanged proteins; and green spots are down-regulated DEPs.
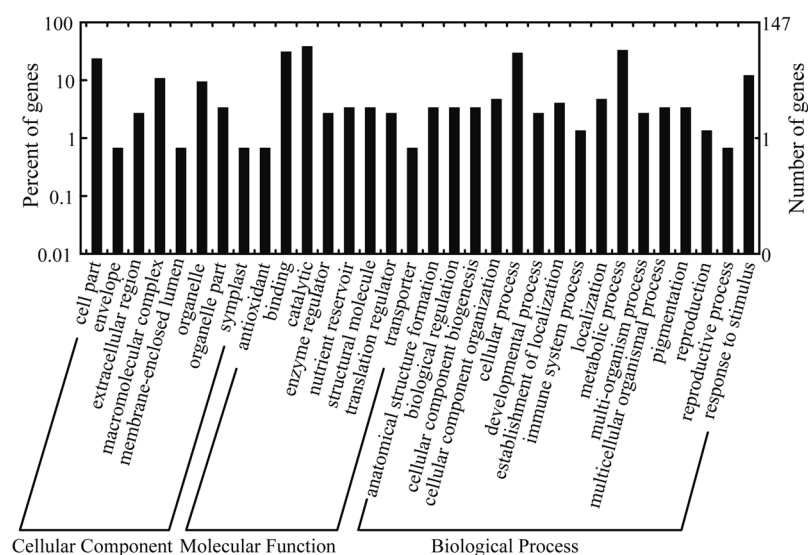
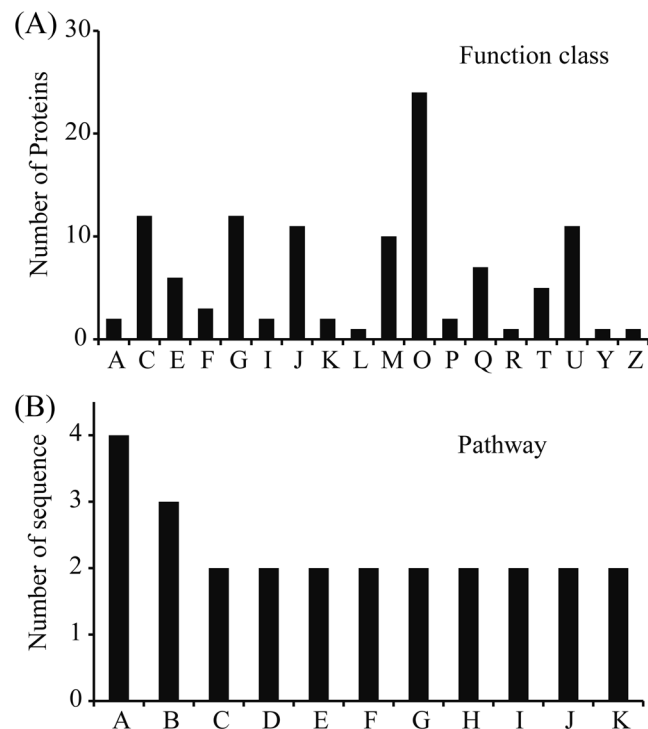**Figure 4.** GO annotation of DEPs in PT and NT maize seeds.

Figure 5. COG classification and KEGG pathway analysis of DEPs between PT and NT maize seeds. Functional class was produced by COG (**A**), and the abbreviations in the figure (**A**) are as follows: A, RNA processing and modification; C, Energy production and conversion; E, Amino acid transport and metabolism, Coenzyme transport; F, Nucleotide transport and metabolism; G, Carbohydrate transport and metabolism; I, Lipid transport and metabolism; J, Translation, ribosomal structure and biogenesis; K, Transcription; L, Transcription, Replication, recombination and repair; M, Cell wall/membrane/envelope biogenesis; O, Post-translational modification, protein turnover, and chaperones; P, Inorganic ion transport and metabolism; Q, Secondary metabolites biosynthesis, transport, and catabolism; R, General function prediction only; T, Signal transduction mechanisms; U, Intracellular trafficking, secretion, and vesicular transport; Y, Nuclear structure; Z, Cytoskeleton. The related KEGG pathways were classified into 11 main categories (**B**), and the abbreviations in the figure (**B**) are as follows: A, Amino sugar and nucleotide sugar metabolism; B, Pyruvate metabolism; C, Starch and sucrose metabolism; D, Galactose metabolism; E, Metabolism of xenobiotics by cytochrome P450; F, Histidine metabolism; G, Carbon fixation in photosynthetic organisms; H, Glutathione metabolism; I, Drug metabolism - cytochrome P450; J, Aminobenzoate degradation; K, Pentose and glucuronate interconversions.

part category (GO:00044464), and 16 proteins (10.9%) in the macromolecular complex category (GO:0032991). For the molecular function ontology, 8 GO terms were found; the major functional groups were catalytic activity (GO:0003824), with 57 proteins (38.8%), and binding functional groups (GO:0005488), with 46 proteins (31.3%). In the biological process category, 16 GO terms were assigned. The major functional groups of the DEPs were involved in metabolic processes (GO:0008152), including 49 proteins (33.3%), followed by cellular processes (GO:0009987), including 44 proteins (29.9%).

The categories of the DEPs between NT and PT maize seeds were further grouped according to their primary biological activities, as defined by the functional catalogue of Clusters of Orthologous Groups (COG) of proteins. The results showed that 131 DEPs (76.4%) were classified into 18 COG categories, among which, "posttranslational modification, protein turnover, chaperones" represented the largest group (group O, 24 DEPs), followed by "energy production and conversion" (group C, 12 DEPs) and "carbohydrate transport and metabolism" (group G, 12 DEPs). Approximately 10% of the DEPs were related to "translation, ribosomal structure and biogenesis" (group J, 11 DEPs) and "intracellular trafficking, secretion, and vesicular transport" (group U, 11 DEPs) (Fig. 5A and Table S4).

To further investigate the biological functions of these DEPs, KEGG pathway analysis was performed using the BLAST2GO 3.0 program. A total of 77 DEPs (52%) were mapped to 37 pathways in the KEGG database. The most represented pathway was "amino sugar and nucleotide sugar metabolism," which contained four sequences (Spots 52, 71, 96, and 124). The other major pathway was "pyruvate metabolism" which contained three sequences (Spots 52, 64, and 113). Two DEPs were involved in each of the following pathways: "starch and sucrose metabolism," "galactose metabolism," "metabolism of xenobiotics by cytochrome P450," "histidine metabolism," "carbon fixation in photosynthetic organisms," "glutathione metabolism," "drug metabolism - cytochrome P450," "aminobenzoate degradation," and "pentose and glucuronate interconversions". The remaining pathways contained only one sequence each (Fig. 5B and Table S5).
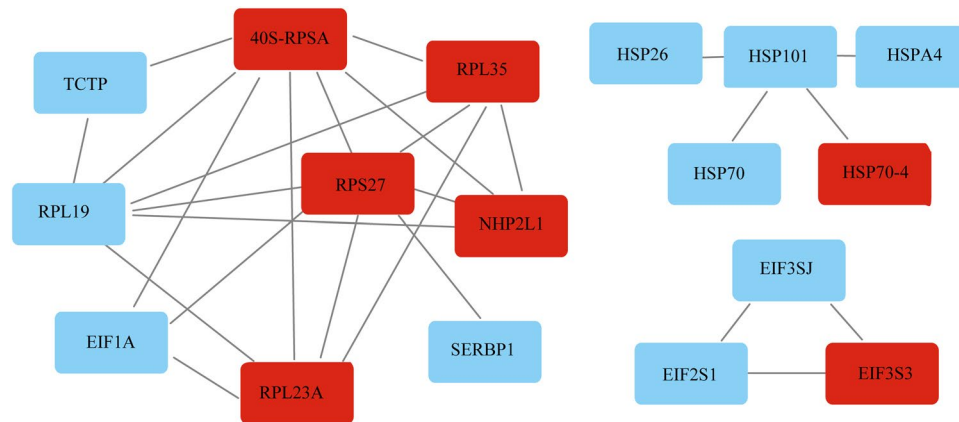
Figure 6. Protein-protein interaction analysis of DEPs. A total of 17 proteins were involved in protein-protein interactions that contained 3 main clusters with 7 up-regulated (marked with red blocks) and 10 down-regulated proteins (blue blocks).

To identify protein-protein interaction networks, the identified 148 DEPs were subjected to STRING (v10) analysis and visualized by Cytoscape 3.40 software with high confidence. In the 148 proteins, 17 were involved in protein-protein interactions, which were included in three main clusters, with seven up-regulated and 10 down-regulated proteins (Fig. 6). Among these proteins, eight interacting proteins were mainly related to "translation, ribosomal structure and biogenesis," while five interacting proteins were related to "post-translational modification, protein turnover, and chaperones" among the COG categories. The 40S ribosomal proteins SA (40S-RPSA) and S27 (RPS27) were found to be the central core proteins in these interacting networks, associating with seven other proteins. The other core proteins were the ribosomal protein L19 and the 60S ribosomal protein L35, which interacted with six other proteins, and a 60S ribosomal protein knows as L23a, which interacted with five other proteins.

Comparison of Protein and Transcript Expression Patterns. To explore the correspondence between the level of mRNA transcripts and the abundance of accumulated proteins, transcriptional analysis of 58 selected DEPs, including 42 up-regulated and 16 down-regulated protein species was performed *via* qRT-PCR. Most of the genes exhibited a similar change trend at the transcriptional and translational levels (Fig. 7). Among the 42 up-regulated DEPs, the transcript levels of 25 genes displayed the same change trend as the abundance of the corresponding protein species. Additionally, 12 genes showed no difference at the transcript level and five genes exhibited the opposite trend from the observed translation levels. The five inconsistent proteins were pro-resilin, MIR-interacting saposin-like protein, xyloglucan endotransglucosylase/hydrolase, activator of 90 kDa heat-shock protein ATPase and an uncharacterized protein. Among the 16 down-regulated DEPs, 14 genes exhibited a similar change trend and two genes showed no difference at the transcript level. The discrepancy between the transcription and translational levels probably resulted from the presence of various posttranslational modifications[26] or the higher sensitivity of iTRAQ compared with that of qRT-PCR[27].

Discussion

The 2-DE method has been a commonly applied separation technique for decades in proteomic studies[28]. However, as a separation method for proteomic analyses, 2-DE presents several limitations[29]. The routine 2-DE approach allows the detection of a limited number of protein spots compared with iTRAQ and other high-throughput quantitative proteomic methods[30]. It is considered that 2D PAGE can resolve up to only 5000 protein spots simultaneously, depending on the gel size and pH gradient used[31]. Moreover, hydrophobic proteins, proteins with a very high or very low molecular weight, and highly acidic or highly basic proteins are often lost during 2-DE gel separation[29]. However, 2-DE is especially suitable for visual detection of high abundant proteins and their isoforms[32].

The iTRAQ method is a gel-free method for shotgun quantitative analysis. It is a second-generation proteomic technique with high sensitivity, excellent accuracy and high throughput. Proteins with low quantity or that are difficult to separate *via* 2-DE can be analyzed by iTRAQ. However, iTRAQ cannot detect many high abundant proteins that can be detected by 2-DE; this inconsistency is clear in the comparison of proteomic studies on maize (*Zea mays*) under salt and drought stress performed using 2-DE and iTRAQ approaches[32–36]. The partial overlap between the outputs of 2-DE and iTRAQ approaches is limited, due to their different characteristics[32, 37].

In the present study, we used 2-DE- and iTRAQ-based quantitative proteomic methods to compare the proteomic characteristics of the seed between PT maize and its NT isogenic counterpart. By taking advantages of these two quantitative techniques, we performed an in-depth study and obtained much new information. In our study, only 43 DEPs were observed in 2D gel, among which 32 proteins were successfully identified, whereas 6313 proteins were successfully identified *via* iTRAQ, 116 of which (1.83%) were considered as DEPs. However, only a few of the proteins identified *via* 2-DE overlapped with the iTRAQ DEPs. This discrepancy might be related to many factors, such as protein physicochemical properties, peptide preparation, quantification way, and proteins with post translational modifications. We consider that these two methods are complementary and multiple
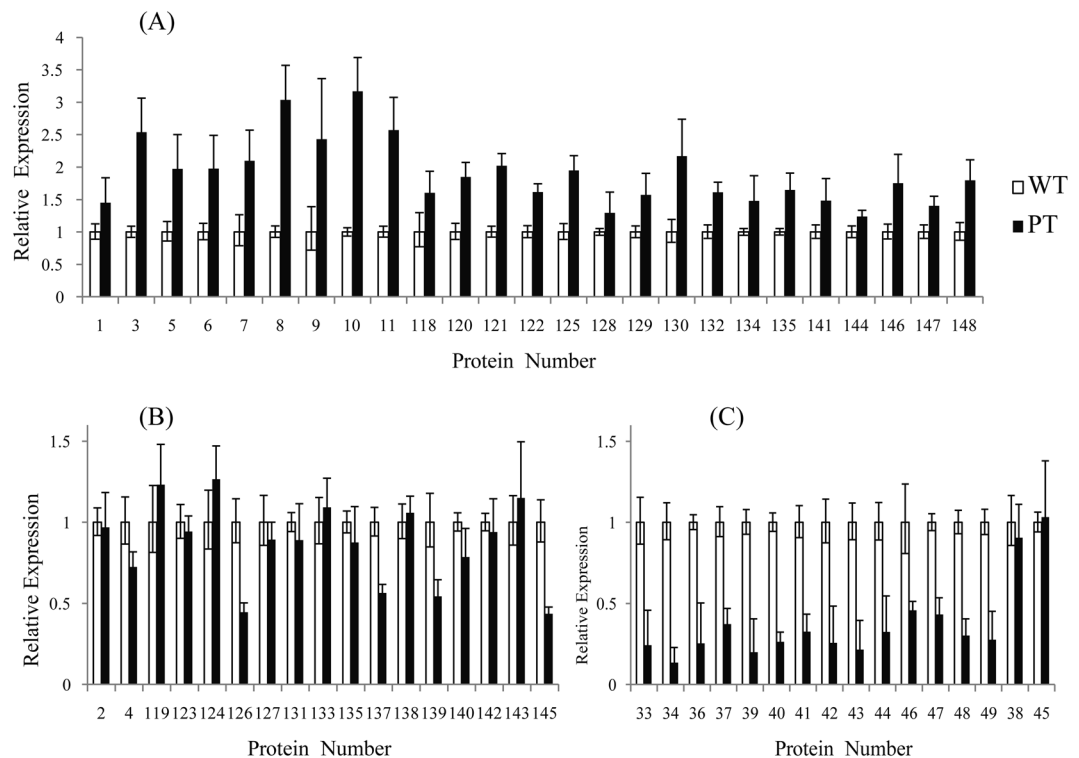
Figure 7. qRT-PCR analysis of the gene expression patterns of the 58 identified DEPs. Among the 42 up-regulated DEPs, the transcript levels of 25 genes displayed the same trend, with increased abundance of the corresponding protein species (**A**), while the results for 17 genes were inconsistent with observed translation levels (**B**); among the 16 down-regulated DEPs, 14 genes exhibited similar change trends, and two genes showed no significant difference at the transcript level (**C**).

strategies are required to obtain high proteomics data. Therefore, the outputs of the two methods should be merged for the subsequent analysis.

A new GMC may exhibit potential unintended effects due to the random insertion of exogenous genes[2]. To assess the effects caused by *phy*A2 gene insertion, a comparative proteomic analysis of the seeds between PT maize and its NT isogenic counterpart was conducted. Using 2-DE- and iTRAQ-based quantitative proteomics techniques, approximately 148 identified DEPs were detected. Compared with the total number of proteins identified using iTRAQ (6313), the number of DEPs identified was significantly smaller, and a limited number of proteins were found to be affected by gene insertion[38]. Moreover, the DEPs identified in the transgenic plants were not identified as new proteins but as proteins showing changes in abundance. These results agreed with our findings in maize leaves[22] and the findings of many other studies[39, 40], indicating that DEPs were detectable but showed no substantial differences in the seeds of PT versus NT maize after overexpression of target genes[8, 22, 41].

Unfortunately, we did not detect the target phytase protein, encoded by the *phy*A2 gene, as a DEP. The DEPs identified using the 2-DE approach may have been due to low expression of the target gene and the limited accumulation of the target protein in PT seeds, as described previously[22]. Regarding the DEPs obtained by iTRAQ, which is a comparative quantitative proteomic analysis method, it must be considered that iTRAQ shows considerable deficits in differential protein detection[30].

COG functional classification showed that approximately 20% of DEPs were associated with "posttranslational modification, protein turnover, chaperones" and 11% with "translation, ribosomal structure and biogenesis". Therefore, approximately 31% of the DEPs were involved in posttranscriptional regulation and posttranscriptional modifications in PT maize seeds. Regarding the function of posttranscriptional regulation, most of the related DEPs were ribosomal proteins. In the nucleolus, ribosomal proteins form the mature 40S and 60S subunits of the mature ribosome (the site of protein synthesis), which is a process of fundamental importance in cell biology[42]. In the present study, the abundance of most ribosomal proteins was found to increase, and changes in mRNA expression levels were up-regulated in PT maize seeds. These up-regulated effects might be a response to the insertion of exogenous genes.

In terms of posttranscriptional modifications, most of the related DEPs were heat shock-proteins. HSPs promote the degradation of abnormal proteins, functioning as "molecular chaperones" to prevent the aggregation of denatured proteins and promote the proper refolding of these proteins[43, 44]. Our results showed that most HSP70s were up-regulated in PT maize seeds. Conversely, the abundance of HSP26 and other chaperonin family proteins was decreased in PT maize seeds (Table S4); this finding suggested that different members of HSPs exhibit different functions in response to the insertion of exogenous genes.

Protein-protein interaction analysis also showed that most interacting proteins were associated with "post-translational modification, protein turnover, chaperones" and "translation, ribosomal structure and biogenesis," and those interacting proteins were mainly ribosomal proteins and heat-shock proteins.

KEGG analysis revealed that only 52% of the DEPs could be mapped to the 37 pathways, and only 3 up-regulated DEPs could be mapped to 5 pathways. In these pathways, most of the enzymes were dehydrogenases, which were involved in 20 pathways, followed by phosphatases involved in 5 pathways. GO annotation also showed that the major functional groups were catalytic activity and binding functional groups in terms of molecular function ontology. Phytases are classified as phosphatases that catalyze the production of less-phosphorylated myo-inositol derivatives and inorganic phosphate from phytic acid[45]. These phosphatase signals may have occurred in response to the insertion of the *phy*A2 gene.

In conclusion, we conducted a comparative proteomic analysis to evaluate the unintended effects of PT maize seeds by combining 2-DE and iTRAQ approaches. A total of 148 DEPs were successfully identified between PT and NT maize seeds, among which 42 were up-regulated and 106 were down-regulated in PT maize seeds compared with NT maize seeds. Bioinformatics analysis showed that most of the DEPs were involved in post-transcriptional regulation and posttranscriptional modification functions in PT maize seeds. Ribosomal proteins and heat shock proteins might present adaptive effects in response to the insertion of exogenous genes. By taking advantages of the 2-DE and iTRAQ approaches, the data generated in the present study provided more information on such unintended effects than the data from target-oriented analyses. However, the number of proteins that could be analyzed was still limited. Unintended effects are widely observed during conventional plant breeding and are not unique to GM plants. Biosafety assessment of GM plants should be performed in a case-by-case manner. Our proteomics data for PT maize seeds may be beneficial for the commercialization of PT maize and will provide much more information for the biosafety assessment of GMCs in the future.

Materials and Methods

Plant Materials. The transgenic maize variety 10TPY006 is a phytase gene-overexpressing variety (PT maize), and the corresponding non-genetically modified control is the conventional hybrid LIYU16 (NT maize). LIYU16 planted in China is commonly used as feed for animals. The genetic background of the two plant materials was described in our previous study[22].

Seeds of PT and NT maize were provided by the Beijing Origin Seed Technology Inc. The PT and NT seeds were planted in a greenhouse, and each line was planted in three micro-plots as three replicates. After sowing, the plants were treated with the following standard agricultural practices. The ears were harvested after physiological maturity. Ears of three plants from each micro-plot were harvested at the same time on the same day, and immediately stored at −80 °C for further study. For the following analysis, seeds derived from the central position of each ear were ground into fine powders.

RT-PCR and Weston Blot Detection of the Target Protein. Semi-quantitative RT-PCR was performed to examine the expression of exogenous genes in transgenic maize. Total RNA was isolated from seeds by the TRIzol method using the TIRpure reagent (Bioteke, Beijing, China). The *zSSIIb* gene was amplified using the *zSSIIb*-F (5′-CGGTGGATGCTAAGGCTGATG-3′) and *zSSIIb*-R (5′-AAAGGGCCAGGTTCATTATCCTC-3′) primers as an internal control. The *bar*-F (5′-GAAGGCACGCAACGCCTACGA-3′) and *bar*-R (5′-CCAGAAACCCACGTCATGCCA-3′) primers were employed for the *bar* gene, and the *phy*A2 event-specific primers used for RT-PCR were P-F (5′-TCAAACCCTTCACGAAGCTATCCC-3′) and P-R (5′-TACTTTCCCGCTCAACTCCACTCT-3′) as described previously[22]. Reference materials were provided by Beijing Origin Seed Technology Inc. and were used as positive and negative controls.

Weston blot detection was performed as described previously[18]. Proteins were extracted from seeds using GLY-HCl buffer (0.05 mol/L, pH 3.0)[46], and approximately 20 μg of the total protein in the supernatant was employed for SDS-PAGE. After separation on an SDS-PAGE gel, the proteins were transferred onto a polyvinylidene difluoride (PVDF) membrane (GE Healthcare) for Western blot analysis. The blot was probed with a polyclonal antibody for the phyA2 protein provided by the Biotechnology Institute, Chinese Academy of Agricultural Sciences (CAAS) (1:2000 dilution), and goat anti-rabbit IgG-labeled with horseradish peroxidase (HRP) was used as the secondary antibody.

Protein Extraction from Seeds. Each sample contained three biological replicates. Total seed proteins were extracted using a modified Borax/PVPP/Phenol (BPP) protein extraction method as described[47]. Approximately 3 g of the frozen maize seeds were ground into fine powders in liquid nitrogen and resuspended in 10 mL of extraction buffer. After 5 min, an equal volume of Tris-saturated phenol (pH 8.0) was added into the suspension, and the mixtures were vortexed for 10 min. The mixtures were then centrifuged (16,000 g, 15 min, 4 °C), and the upper phase was transferred to a new centrifuge tube. Then protein precipitates were obtained by adding five volumes of ammonium sulfate saturated-methanol, followed by incubation at 22 °C for at least 6 h. The precipitated proteins were centrifuged as indicated above, and the protein pellet was resuspended and rinsed twice with ice-cold methanol and ice-cold acetone. Finally, the protein pellets were air-dried and recovered with lysis buffer to determine the protein concentration according to the Bradford method, using bovine serum albumin as a protein standard, or were stored at −20 °C.

2-DE Electrophoresis. Approximately 1,300 μg of the protein samples were diluted to a volume of 450 μL with lysis buffer (7 M urea, 2 M thiourea, 2% CHAPS, 13 mM DTT), followed by loading onto a 24-cm IPG strip (immobilized pH gradient) with a linear pH gradient of 4–7 (GE Healthcare, Uppsala, Sweden). The strips were hydrated for 18 h at room temperature, and then subjected to IEF on an Ettan IPGphor isoelectric focusing system, following the instructions of the manufacturer (2-DE Manual, GE Healthcare), with some modifications as

described previously[22]. After IEF, the IPG strips were equilibrated first in an equilibration solution containing 1% DTT, followed by an equilibration solution with 4% iodoacetamide. The strips were then transferred to an Ettan Dalt system (GE Healthcare) to perform SDS-PAGE using 12.5% SDS polyacrylamide gels[47].

The gels were visualized *via* the GAP staining method[48] and scanned with ImageMaster Labscan V3.0 (GE Healthcare, Uppsala, Sweden), and image analysis was performed with the ImageMaster 2D Platinum software package (GE Healthcare, Uppsala, Sweden). Spots that were present in all replicate gels, showing Student's *t* test p-values <0.05 and a relative fold change of at least a 1.5 in their quantity, were further analyzed.

Protein Identification in 2-DE Gels *via* MALDI-TOF MS/MS.

The target protein spots of DEPs were manually excised from the 2-DE gels and digested in-gel with bovine trypsin (Trypsin, Roche, Cat. 11418025001) as described previously[23]. Digested proteins were then mixed with alpha-cyano-4-hydroxycinnamic acid (CHCA) matrix for determination of mass spectra and analyzed using an AB SCIEX MALDI TOF-TOF 5800 system (AB SCIEX, Shanghai, China) equipped with a neodymium with laser wavelength 349 nm. Mass spectra were obtained as described previously[49].

The raw MS and MS/MS spectra were combined and searched against the *Zea mays* amino acid sequence database (including 87,603 sequences), which was downloaded from Uniprot (http://www.uniprot.org), using in-house MASCOT software for protein identification. The searched parameters were set as previously described[22]. If peptides matched multiple proteins, the protein with the highest score was recorded in this study for bioinformatics analysis. For the unnamed proteins, BLAST search using NCBI (http: //www.ncbi.nlm) was performed to identify homologous proteins.

iTRAQ Labeling, Fractionation and Nano-LC-MS/MS Analysis.

Proteins from two biological replicate samples were used for quantitative proteomic analysis. The iTRAQ experiment was conducted on a Triple TOF 5600 system (AB Sciex) coupled with a Nanoflex microchip system (Eksigent, Dublin, USA) as described[50]. A total of 100 μg of proteins was collected from each sample solution and digested with trypsin (AB Sciex, No. 4352157). The digested peptides were labeled using iTRAQ 4-plex kits (Applied Biosystems, USA), following the instructions of the manufacturer. The NT maize replicates were labeled with the iTRAQ 114 and 115 tags, and the PT maize replicates were labeled with the 116 and 117 tags. The labeled peptide mixtures were pooled and then dried by vacuum centrifugation.

The labeled peptide mixtures were re-dissolved with buffer A (20 mM ammonium formate, pH 10.0) and fractioned using a C18 column (4.6 mm × 250 mm, 5 μm 100 Å, Agela). The peptides were eluted at a flow rate of 0.8 mL/min, and the gradient of fraction was set as follows: 5% buffer B (20 mM ammonium formate, pH 10.0, 80% CAN) for 0–5 min, 5–25% buffer B for 5–30 min, 25–40% buffer B for 30–45 min, 40–90% buffer B for 45–55 min, 90% buffer B, 55–64 min, and 90–5% buffer B for 64–65 min. The chromatogram was monitored through a UV detector, and UV wavelengths were set at 280 and 214 nm. Fractions were collected at 1 min intervals, with a total of 65 fractions being collected and lyophilized. The eluted peptides were pooled into 10 fractions for subsequent nano-LC-MS/MS analysis.

Peptides were eluted from the HPLC column *via* the application of a gradient from 5% buffer B (98% ACN v/v, 0.1% FA v/v) to 55% buffer B for 79 min, running at 300 nL/min, followed by a linear gradient to 80% buffer B in 1 min. The obtained peptides were subjected to analysis using a Triple TOF 5600 mass spectrometer, which was operated in an information-dependent data acquisition mode. The parameters were set as follows: ion spray voltage: 2.3 kV; GS1: 4; curtain gas: 30; DP: 100; m/z: 350–1250; accumulation time: 0.25 s; product ion scan: IDA number: 30; m/z: 100–1500; accumulation time: 0.1 s; dynamic exclusion time: 25 s; rolling CE: enabled; adjust CE when using iTRAQ reagent: enabled; CES: 5.

iTRAQ Protein Identification and Quantification.

Protein identification was performed using ProteinPilot™ software (AB Sciex Inc., USA) against the *Zea mays* database (87,603 sequences) from the UniprotKB database. To evaluate the consistency among replicates, quantitative ratios of protein species abundance between the four biological replicates were compared with a value of 1. For relative protein quantification, one protein was required to contain at least two unique peptides. The quantitative protein ratios were weighed and normalized according to the median ratio in Mascot. Proteins showing a fold change of at least 1.5 and a P-value <0.05 were further analyzed in four biological replicates.

DEPs Functional Classification and Pathway Analysis.

Functional annotations of the DEPs identified through combined 2-DE and iTRAQ analyses were performed. COG analysis of the proteins was performed for functional classification *via* searches in a database (http://eggnogdb.embl.de/). GO classification according to GO terms was further performed online using WEGO software (http://wego.genomics.org.cn)[51]. In addition, KEGG pathway analysis was conducted to determine the reaction networks of all identified DEPs using Blast2GO 4.0 against the non-redundant protein database (NR; NCBI) (http://www.geneontology.org/). Finally, protein-protein interactions were analyzed using the STRING v10 database (http://string-db.org) for the identified proteins to determine the functions and pathways that were most strongly associated with the protein list.

RNA Extraction and Quantitative Real-time PCR.

Total RNA was extracted with the TRIzol reagent, and 1 μg of RNA was used for reverse transcription into cDNA with a reverse transcriptase kit (TaKaRa, Tokyo, Japan). The mixed solution of the qRT-PCR reaction with a 20 μL volume was prepared in triplicate by adding 1 μL of each cDNA dilution to SYBR Green PCR Master Mix (TaKaRa) and run on an Mx3005 P sequence detection system, following the instructions of the manufacturer. Primers were designed using Primer 5.0 software (Premier Biosoft International, Palo Alto, CA, USA), and the maize *zSSIIb* gene was employed as the internal control. All primer pairs used for qRT-PCR are listed in Table S6. Data were analyzed using MxPro software.

References

1. James, C. 20th anniversary (1996 to 2015) of the global commercialization of biotech crops and biotech crop highlights in 2015. *ISAAA Brief.* 51 (ISAAA: Ithaca, NY, 2016).
2. Gong, C. Y. & Wang, T. Proteomic evaluation of genetically modified crops: current status and challenges. *Front Plant Sci.* **4**, 41, doi:10.3389/fpls.2013.00041 (2013).
3. Smith, L. M. & Kelleher, N. L. Proteoform: a single term describing protein complexity. *Nat Methods.* **10**, 186–187, doi:10.1038/nmeth.2369 (2013).
4. Schluter, H., Apweiler, R., Holzhutter, H. G. & Jungblut, P. R. Finding one's way in proteomics: a protein species nomenclature. *Chem Cent J.* **3**, 11, doi:10.1186/1752-153X-3-11 (2009).
5. Rappsilber, J. & Mann, M. What does it mean to identify a protein in proteomics? *Trends Biochem. Sci.* **27**, 74–78 (2002).
6. Balsamo, G. M., Cangahuala-Inocente, G. C., Bertoldo, J. B., Terenzi, H. & Arisi, A. C. Proteomic analysis of four Brazilian MON810 maize varieties and their four non-genetically-modified isogenic varieties. *J Agric Food Chem.* **59**, 11553–11559, doi:10.1021/jf202635r (2011).
7. Zolla, L., Rinalducci, S., Antonioli, P. & Righetti, P. G. Proteomics as a complementary tool for identifying unintended side effects occurring in transgenic maize seeds as a result of genetic modifications. *J Proteome Res.* **7**, 1850–1861, doi:10.1021/pr0705082 (2008).
8. Vidal, N., Barbosa, H., Jacob, S. & Arruda, M. Comparative study of transgenic and non-transgenic maize (Zea mays) flours commercialized in Brazil, focussing on proteomic analyses. *Food Chem.* **180**, 288–294, doi:10.1016/j.foodchem.2015.02.051 (2015).
9. Coll, A., Nadal, A., Rossignol, M., Puigdomenech, P. & Pla, M. Proteomic analysis of MON810 and comparable non-GM maize varieties grown in agricultural fields. *Transgenic Res.* **20**, 939–949, doi:10.1007/s11248-010-9453-y (2011).
10. Albo, A. G. *et al.* Proteomic analysis of a genetically modified maize flour carrying Cry1Ab gene and comparison to the corresponding wild-type. *Maydica.* **52**, 443–455 (2007).
11. Ross, P. L. *et al.* Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics.* **3**, 1154–1169, doi:10.1074/mcp.M400129-MCP200 (2004).
12. Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J. & Gygi, S. P. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res.* **2**, 43–50, doi:10.1021/pr025556 (2003).
13. Washburn, M. P., Dirk, W. & Yates, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol.* **19**, 242–247, doi:10.1038/85686 (2001).
14. Zieske, L. R. A perspective on the use of iTRAQ TM reagent technology for protein complex and profiling studies. *J Exp Bot.* **57**, 1501–1508, doi:10.1093/jxb/erj168 (2006).
15. Ghochani, Bb. F. N. M. & Gilany, K. Proteomics a key tool for a better understanding of endometriosis: a mini- review. *J Paramedical Sci.* **2**, 51–58 (2011).
16. Fukao, Y. *et al.* iTRAQ analysis reveals mechanisms of growth defects due to excess Zinc in arabidopsis. *Plant Physiol.* **155**, 1893–1907, doi:10.1104/pp.110.169730 (2011).
17. Chen, R. *et al.* Corn seeds as bioreactors for the production of phytase in the feed industry. *J. Biotechnol.* **165**, 120–126, doi:10.1016/j.jbiotec.2013.01.030 (2013).
18. Chen, R. *et al.* Transgenic maize plants expressing a fungal phytase gene. *Transgenic Res.* **17**, 633–643, doi:10.1007/s11248-007-9138-3 (2008).
19. Gao, C. Q. *et al.* Evaluation of the compositional and nutritional values of phytase transgenic corn to conventional corn in roosters. *Poult Sci.* **91**, 1142–1148, doi:10.3382/ps.2011-01915 (2012).
20. Li, S. F. *et al.* Energy, amino acid, and phosphorus digestibility of phytase transgenic corn for growing pigs. *J Anim Sci.* **91**, 298–308, doi:10.2527/jas.2012-5211 (2013).
21. Zhang, Y., Liu, C., Li, Y. & Wu, K. Phytase transgenic maize does not affect the development and nutrition utilization of Ostrinia furnacalis and Helicoverpa armigera. *Environ Entomol.* **39**, 1051–1057, doi:10.1603/EN09380 (2010).
22. Tan, Y. *et al.* Comparative proteomics of leaves from phytase-transgenic maize and its non-transgenic isogenic variety. *Front Plant Sci.* **7**, 1211, doi:10.3389/fpls.2016.01211 (2016).
23. Wang, X. *et al.* Comparative proteomic analysis of differentially expressed proteins in shoots of Salicornia europaea under different salinity. *J Proteome Res.* **8**, 3331–3345, doi:10.1021/pr801083a (2009).
24. Yang, Q. S. *et al.* Quantitative proteomic analysis reveals that antioxidation mechanisms contribute to cold tolerance in plantain (Musa paradisiaca L.; ABB Group) seedlings. *Mol Cell Proteomics.* **11**, 1853–1869, doi:10.1074/mcp.M112.022079 (2012).
25. Gan, C. S., Chong, P. K., Pham, T. K. & Wright, P. C. Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). *J Proteome Res.* **6**, 821–827, doi:10.1021/pr060474i (2007).
26. Wang, X. *et al.* iTRAQ-based quantitative proteomic analysis reveals new metabolic pathways responding to chilling stress in maize seedlings. *J Proteomics.* **146**, 14–24, doi:10.1016/j.jprot.2016.06.007 (2016).
27. Li, J. *et al.* Differential proteomics analysis to identify proteins and pathways associated with male sterility of soybean using iTRAQ-based strategy. *J Proteomics.* **138**, 72–82, doi:10.1016/j.jprot.2016.02.017 (2016).
28. Rogowska-Wrzesinska, A., Le Bihan, M. C., Thaysen-Andersen, M. & Roepstorff, P. 2D gels still have a niche in proteomics. *J Proteomics.* **88**, 4–13, doi:10.1016/j.jprot.2013.01.010 (2013).
29. Lenz, C. & Dihazi, H. Introduction to Proteomics Technologies. *Methods Mol Biol.* **1362**, 3–27, doi:10.1007/978-1-4939-3106-4_1 (2016).
30. Wu, X. & Wang, W. Increasing confidence of proteomics data regarding the identification of stress-responsive proteins in crop plants. *Front Plant Sci.* **7**, 702, doi:10.3389/fpls.2016.00702 (2016).
31. Lilley, K. S., Razzaq, A. & Dupree, P. Two-dimensional gel electrophoresis recent advances in sample preparation, detection and quantitation. *Curr Opin Chem Biol.* **6**, 46–50 (2001).
32. Benešová, M. *et al.* The physiology and proteomics of drought tolerance in maize: early stomatal closure as a cause of lower tolerance to short-term dehydration? *PLoS One.* **7**, e38017, doi:10.1371/journal.pone.0038017 (2012).
33. Cheng, Y. *et al.* Salt-induced root protein profile changes in seedlings of maize inbred lines with differing salt tolerances. *Chil J Agr Res.* **74**, 468–476, doi:10.4067/s0718-58392014000400014 (2014).
34. Cui, D. *et al.* Proteomic analysis of seedling roots of two maize inbred lines that differ significantly in the salt stress response. *PLoS One.* **10**, e0116697, doi:10.1371/journal.pone.0116697 (2015).
35. Sanggon, K. *et al.* Physiological and proteomic analysis of the response to drought stress in an inbred Korean maize line. *Plant Omics.* **8**, 159–168, doi:10.1007/s12892-014-0110-5 (2015).
36. Zörb, C., Schmitt, S. & Mühling, K. H. Proteomic changes in maize roots after short-term adjustment to saline growth conditions. *Proteomics.* **10**, 4441–4449, doi:10.1002/pmic.201000231 (2010).
37. Alvarez, S. *et al.* Comprehensive analysis of the Brassica juncea root proteome in response to cadmium exposure by complementary proteomic approaches. *Proteomics.* **9**, 2419–2431, doi:10.1002/pmic.200800478 (2009).
38. Arruda, S. C. C., Barbosa, H. S., Azevedo, R. A. & Arruda, M. A. Z. Comparative studies focusing on transgenic through cp4EPSPS gene and non-transgenic soybean plants: An analysis of protein species and enzymes. *J Proteomics.* **93**, 107–116, doi:10.1016/j.jprot.2013.05.039 (2013).

39. Ren, Y. *et al*. A comparative proteomics approach to detect unintended effects in transgenic Arabidopsis. *J Genetics and Genomics.* **36**, 629–639, doi:10.1016/s1673-8527(08)60155-1 (2009).
40. Ruebelt, M. C. *et al*. Application of two-dimensional gel electrophoresis to interrogate alterations in the proteome of gentically modified crops. 3. Assessing unintended effects. *J Agric Food Chem.* **54**, 2169–2177, doi:10.1021/jf052358q (2006).
41. Gong, C. Y., Li, Q., Yu, H. T., Wang, Z. & Wang, T. Proteomics insight into the biological safety of transgenic modification of rice as compared with conventional genetic breeding and spontaneous genotypic variation. *J Proteome Res.* **11**, 3019–3029, doi:10.1021/pr300148w (2012).
42. Robledo, S. *et al*. The role of human ribosomal proteins in the maturation of rRNA and ribosome production. *Rna.* **14**, 1918–1929, doi:10.1261/rna.1132008 (2008).
43. Wang, W., Vinocur, B., Shoseyov, O. & Altman, A. 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 (2004).
44. Parsell, D. A. & Lindquist, S. The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. *Genetics.* **27**, 437–496, doi:10.1146/ annurev.ge. 27.120193. 002253 (1993).
45. Pasamontes, L. & Wyss, M. Biophysical characterization of fungal phytases (myo-inositol hexakisphosphate phosphohydrolases). *Appl Environ Microbiol.* **65**, 359–366 (1999).
46. Zhao, Q. *et al*. Purification of phytase from transgenic maize by immunoaffinity chromatography. *China Biotechnology.* **33**, 125–130 (2013).
47. Wang, X. *et al*. A protein extraction method compatible with proteomic analysis for the euhalophyte Salicornia europaea. *Electrophoresis.* **28**, 3976–3987, doi:10.1002/elps.200600805 (2007).
48. Wang, X. *et al*. Systematic comparison of technical details in CBB methods and development of a sensitive GAP stain for comparative proteomic analysis. *Electrophoresis.* **33**, 296–306, doi:10.1002/elps.201100300 (2012).
49. Yi, X. *et al*. Quantitative proteomics of Sesuvium portulacastrum leaves revealed that ion transportation by V-ATPase and sugar accumulation in chloroplast played crucial roles in halophyte salt tolerance. *J Proteomics.* **99**, 84–100, doi:10.1016/j.jprot.2014.01.017 (2014).
50. Wang, X. C. *et al*. Quantitative proteomics and transcriptomics reveal key metabolic processes ssociated with cotton fiber initiation. *J Proteomics.* **114**, 16–27, doi:10.1016/j.jprot.2014.10.022 (2015).
51. Ye, J. *et al*. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* **34**, W293–297, doi:10.1093/nar/gkl031 (2006).

Acknowledgements

Author Contributions

A.G. and X.W. conceived the project. Y.T. conducted research work and wrote the manuscript. Z.T., Q.Y., Y.S., X.J. and C.P. performed statistical analysis.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-09557-8

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.