Quantitative Proteomic and Transcriptomic Study on Autotetraploid Paulownia and Its Diploid Parent Reveal Key Metabolic Processes Associated with Paulownia Autotetraploidization

Yanpeng Dong¹, Minjie Deng¹, Zhenli Zhao¹ and Guoqiang Fan¹, ²*

¹ Department of Forestry, College of Forestry, Henan Agricultural University, Zhengzhou, China, ² Institute of Paulownia, Henan Agricultural University, Zhengzhou, China

Polyploidy plays a very important role in speciation and plant evolution by way of genomic merging and doubling. In the process of polyploidy, rapid genomic, and transcriptomic changes have been observed and researched. However, proteomic divergence caused by the effects of polyploidization is still poorly understood. In the present study, we used iTRAQ coupled with mass spectrometry to quantitatively analyze proteomic changes in the leaves of autotetraploid Paulownia and its diploid parent. A total of 2963 proteins were identified and quantified. Among them, 463 differentially abundant proteins were detected between autotetraploid Paulownia and its diploid parent, and 198 proteins were found to be non-additively abundant in autotetraploid Paulownia, suggesting the presence of non-additive protein regulation during genomic merger and doubling. We also detected 1808 protein-encoding genes in previously published RNA sequencing data. We found that 59 of the genes that showed remarkable changes at mRNA level encoded proteins with consistent changes in their abundance levels, while a further 48 genes that showed noteworthy changes in their expression levels encoded proteins with opposite changes in their abundance levels. Proteins involved in posttranslational modification, protein turnover, and response to stimulus, were significantly enriched among the non-additive proteins, which may provide some of the driving power for variation and adaptation in autopolyploids. Quantitative real-time PCR analysis verified the expression patterns of related protein-coding genes. In addition, we found that the percentage of differentially abundant proteins that matched previously reported differentially expressed genes was relatively low.

Keywords: proteome, transcriptome, autotetraploid, Paulownia, iTRAQ

INTRODUCTION

Polyploidy is a widespread and prominent process that plays an important role in the evolution of all angiosperm plants through genomic merging and doubling. There are three main types of polyploids: autopolyploids, allopolyploids, and segmental allopolyploids (Stebbins, 1947). Autopolyploids such as Solanum tuberosum, Saccharum officinarum, and Medicago sativa, combine
more than two sets of identical genomes in their nucleus (Chen, 2010). They often thrive more vigorously and can survive in severe environments because, through genome duplication (Soltis and Soltis, 2009), they have traits that are superior to their parents. Many hypotheses have been proposed to explain the function of a doubled genome. One such hypothesis is that genome duplication endows distinct superiorities that enable polyploids to survive and thrive in circumstances that are too severe for their parents (Otto, 2007; Weiss-Schneweiss et al., 2013). Despite the widespread occurrence and survival superiorities of polyploids, and some progress in this area, there is still not enough information to explain the cause and direct effect of the assigned biological functions of the increased content of the doubled genome.

High-throughput genomic and transcriptomic sequencing techniques have contributed greatly to gaining a comprehensive perspective on the genomic information of autopolyploids. However, because changes at the transcriptional level do not always correlate with changes at the translational level, it is also necessary to investigate the differences between autotetraploids and their corresponding progenitors at the proteome level. The correlation between the expression of RNA transcripts and protein abundance is usually not direct because processes such as posttranscriptional regulation and posttranslational modifications can make it difficult to predict patterns of protein abundance. Thus, applying proteomic approaches to investigate autopolyploids will greatly increase the understanding of their evolution and adaption. Until recently, proteomic changes in autopolyploids and their diploid parents have been rarely reported, and only a few studies on Arabidopsis (Ng et al., 2012) and Manihot esculenta Crantz (An et al., 2014) were found. In most of these researches, the method of two-dimensional electrophoresis (2-DE) combined with mass spectrometry was used. In general, there are three kinds of methods involving in proteome: 2-DE, label, and shotgun. They are complementary to each other. Coupled with mass spectrometry analysis, iTRAQ has proven to be a powerful method to investigate proteomic changes because many proteins can be identified simultaneously and proteomic changes can be measured with high sensitivity in related species (Ross et al., 2004; O’Brien et al., 2010). Therefore, an iTRAQ-based proteomic analysis in autopolyploids should provide valuable information and new insights into the effects of autopolyploidization at the translational level (Koh et al., 2012).

Paulownia is an economically important genus of trees that are native to China. Paulownia “Yuza I” is an improved species that was bred from P. tomentosa × P. fortunei. The autopolyploid “Yuza I” does not exist in nature; however, it has valuable traits, including drought and weeds’ broom diseases resistance, that were acquired as a result of its autopolyploidy (Liu et al., 2013; Li et al., 2014; Xu et al., 2014). In this study, iTRAQ technology combined with liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) was employed to investigate the effects of polyploidization on the “Yuza I” proteome. Our results provide a survey of protein changes as a result of genome duplication and provide a better understanding of the mechanisms of autopolyploidization in Paulownia. We also cataloged differences in the abundance of proteins and expression levels of mRNAs using previously reported RNA sequencing data (Li et al., 2014). Finally, potential target genes identified in this study could be engineered for acclimation in plants, especially stress.

**MATERIALS AND METHODS**

**Plant Material**

All biological materials used in this study were obtained from the Institute of Paulownia, Henan Agricultural University, Zhengzhou, Henan Province, China.

Uniformly grown tissue cultured seedlings of the “Yuza I” diploid (Y2, “Yuza I” diploid) and autotetraploid (Y4, “Yuza I” autotetraploid) were cultured in 100-mL triangular flasks on 1/2 MS medium for 30 days. Healthy leaves were harvested, and equal numbers of leaves from each of the three plants in each of the replicate groups [Y2, Y2-2 (replicates for “Yuza I” diploid)]; [Y4,Y4-2 (replicates for “Yuza I” autotetraploid)] were pooled to form the two samples. Terminal buds of 1.5 cm in length of the plantlets in the four treatment groups were sheared, immediately frozen in liquid nitrogen, and stored at −80°C.

**Protein Extraction, Digestion, and iTRAQ Labeling**

The Y2 and Y4 samples were ground into powder in liquid nitrogen, extracted with lysis buffer containing 1 mM PMSF and 2 mM EDTA. After that, protein was extracted according to the method of Tang et al. (2016). Total protein was digested and iTRAQ-labeled according to the method of Meng et al. (2014). The samples were labeled with iTRAQ tags as follows (sample, tag): Y2, 113; Y4, 116; Y2-2, 117; Y4-2, 118. Four other samples were labeled with the same tags as repeats. The peptides were labeled with the isobaric tags and incubated at room temperature for 2 h. The labeled peptide mixtures were then pooled for strong cation exchange and dried by vacuum centrifugation.

**Strong Cation Exchange**

Strong cation exchange chromatography was performed with a LC-20AB HPLC pump system. The iTRAQ-labeled peptide mixtures were reconstituted with 4 mL buffer A and loaded onto a 4.6 × 250 mm Ultremex SCX column containing 5-μm particles. The peptides were eluted at a flow rate of 1 mL min⁻¹ with a gradient of buffer A for 10 min, 5–60% buffer B for 27 min, and 60–100% buffer B for 1 min. The system was then maintained at 100% buffer B for 1 min before equilibrating with buffer A for 10 min prior to the next injection. Elution was monitored by measuring the absorbance at 214 nm, and fractions were collected every 1 min. The eluted peptides were pooled into 20 fractions, desalted with a Strata X C18 column and vacuum-dried.

**LC-MS/MS Analysis**

Mass spectroscopy analysis was performed using an AB SCIEX TripleTOF™ 5600 mass spectrometer, coupled with an online micro-flow HPLC system as described in Section Strong Cation Exchange. The peptides were separated using the method of Qiao et al. (2012).
iTRAQ Protein Identification and Quantification

The acquired raw data files were converted into Mascot generic format files according to the method of Lin et al. (2013). Protein identification was performed using the Mascot search engine against a Paulownia transcriptome database that contained 82,934 sequences. The sequencing data have been submitted to the Short Reads Archive under accession number SRP034738 (Li et al., 2014). Protein identification and quantification was performed according to the method of Guo et al. (2014).

To reduce the probability of false peptide identification, only peptides with significance scores in the 99% confidence interval greater than “identity” in a Mascot probability analysis were counted as identified. Each confident protein identification involved at least one unique peptide.

For protein quantization, each protein was required to contain at least two unique peptides. The quantitative protein ratios were weighted and normalized by the median ratio in Mascot. Only ratios with $p < 0.05$ and fold changes $>1.2$ were considered significant.

Bioinformatics Analysis

Functional analysis of the identified proteins was conducted using Gene Ontology annotations and the proteins were categorized according to their biological processes, molecular functions and cellular localizations. The proteins were further analyzed using the Clusters of Orthologous Groups of proteins database and the Kyoto Encyclopedia of Genes and Genomes database (Fan et al., 2014). Principal Component Analysis (PCA) of the samples were done using the software of SIMCA-P (Wu et al., 2010).

RNA Preparation and Quantitative RT-PCR

The RNA samples from Y2, Y4, Y2-2, and Y4-2 were extracted with Trizol (Sangon, Shanghai, China). The RNA was then precipitated with isopropanol. Purified and concentrated RNA was denatured and first-strand cDNAs for all the samples were synthesized using a PrimeScript RT reagent Kit (Takara, Dalian, China). Primers were designed using Beacon Designer version 7.7 (Premier Biosoft International, Ltd., Palo Alto, CA, USA; Table 1). The genes encoding 10 differentially abundant proteins (DAPs) were selected for qRT-PCR analysis. The qRT-PCR reactions were run in So Fast EvaGreen Supermix starting with 1 $\mu$L of the cDNA template in a standard 20-$\mu$L reaction. The cDNAs were then amplified in a Bio-Rad CFX96TM Real-Time System with SYBR Premix Ex Taq TM II. The PCR cycles were as follows: 95°C for 1 min, followed by 40 cycles of 95°C for 10 s, and 55°C for 15 s. Relative expression levels of the genes were calculated using the $2^{- \Delta \Delta Ct}$ method and normalized with 18S rRNA from “Yuza 1.”

RESULTS

Proteomics Characterization

A total of 366,435 spectra were generated from the iTRAQ-based quantitative proteomics analysis of the proteins extracted from Y2, Y2-2, Y4, and Y4-2 seedling leaves. The analysis using Mascot software identified 21,423 spectra that matched known spectra. Among them, 18,165 unique spectra were matched to 7477 unique peptides and 2963 proteins (Figure 1A). ~54% of which consisted of at least two of the peptides (Figure 1B). The vast majority of these proteins were larger than 10 kDa, although their molecular weights covered a wide range (Figure 1C). Most of the identified proteins had good peptide coverage; 59% had more than 10% sequence coverage, and 33% had 20% sequence coverage (Figure 1D). Further, to avoid identification omissions, we also confined the peptide matching error of the database search strategy to less than 2 ppm. The reproducibility of the proteomic analysis is shown in Figure 2. These results indicate that the proteomics analyses were reliable. The results of PCA of the samples were shown in Figure S1. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al., 2016) partner repository with the dataset identifier PXD004237. Data are available via ProteomeXchange with identifier PXD004237 (Vizcaíno et al., 2014). Reviewer account details were as follows: username was reviewer36112@ebi.ac.uk and password was TH1qVbgd.

To carry out a functional analysis, we assigned GO terms to all the quantitated proteins. Under biological process, metabolic process (1950) and cellular process (1830) were the most represented groups; under cellular component, cell (2290) and cell part (2290) were the two largest groups; and under molecular function, catalytic activity (1518) and binding (1331) were most represented (Figure 3). These results indicate that the identified proteins are involved in almost every aspect of “Yuza 1” metabolism.

To further understand the functions of the 2963 proteins, they were assigned to 23 categories using the COG database. The most highly represented functional categories were “general function prediction only” and “post translational modification, protein turnover, chaperones,” which represented ~16 and 14% of the identified proteins, respectively (Figure 4).

Next, the annotated proteins were mapped onto 121 KEGG pathways (Table S1). Among them, “metabolic pathways” (758) was significantly more highly represented than the other pathways, and was followed by “biosynthesis of secondary metabolites” (455) and “ribosome” (116). Based on the KEGG analysis, we concluded that most of the mapped proteins may affect cellular component biogenesis, post translational modification, protein turnover, and response to stimulus.

Overview and Analysis of Differentially Abundant Proteins Related to Autotetraploidy

DAPs were defined as proteins that showed a greater than 1.2-fold change in relative abundance and a $p < 0.05$. A total of 463 DAPs were identified in the Y4 vs. Y2 comparison; 265 were more abundant and 198 were less abundant (Table S2). Among the top 10 more abundant DAPs, only five could be aligned to proteins with known functions in GenBank. These proteins were annotated as lipid transfer protein 2 (Salvia miltiorrhiza, ABP01769.1gi|144601657),


| Unigene-ID | Potential gene function | Size (bp) | Up/down* | Primer sequence |
|------------|-------------------------|-----------|----------|-----------------|
| CL2443.Contig2_All | Translation initiation factor 1A | 2332 | Up | ATTCGTGAGCATTCGCCATTC |
| | | | | AGCATTTGAGCAGGAAACCC |
| | | | | GAGGGCACTAACGAGGAGAGG |
| | | | | GTTTGAGCATTCGACAGTTC |
| Unigene18898_All | GTP-binding protein SAR1A | 372 | Up | TGCGTCTCCTAAATCTATACAC |
| | | | | CATCTTTGATAGATTGTTGAC |
| CL4141.Contig1_All | Photosystem I subunit III | 1528 | Up | CTGCTTACTGCAACTCTC |
| | | | | GTGATCTGTTATGATGACGC |
| Unigene20285_All | Ubiquinol-cytochrome C reductase | 422 | Up | TTGACTCCACTAATGACAGC |
| | | | | ATCCATCGGCAAGAAAGGC |
| CL2464.Contig1_All | Histone H4 | 2099 | Up | ATCAATCGGCGGAATAGGC |
| | | | | AGATGCGGTCACTCACAAAC |
| Unigene4043_All | Lipid transfer protein 2 | 575 | Up | AGGACTCTGAATAGAACTTATGG |
| | | | | TTGACTCCACTAATGACAGC |
| CL4400.Contig1_All | Photosystem I reaction center subunit XI | 1239 | Up | TGCGATATCGACACAAATGAG |
| | | | | AACATCAACTACACACCTTCC |
| CL9028.Contig1_All | Nitrate excretion transporter 1 | 1067 | Up | GACTTATGCTCTGGTTGTTG |
| | | | | AGATTCGGTGAGGCTAATAC |
| CL5761.Contig1_All | Vacuolar ATP synthase subunit G1 | 778 | Up | TGGTTCCAAGTCATTAGCC |
| | | | | TGTCAAAGAAAGAAGAAATACG |
| Unigene4328_All | Monocopper oxidase-like protein SKS1 | 1138 | Up | TGCGATACTGACACAACTAAG |
| | | | | AAATCTCAACATTACCCTTC |
| CL8969.Contig1_All | Pathogenesis-related protein 10.4 | 955 | Down | ACGACGACATAATGATGGAAAC |
| | | | | CTAATCCAGTTTTCGCAACT |
| CL392.Contig1_All | Brassinosteroid-regulated protein BRU1 | 1588 | Down | TGTCGACATGTATCAGTTAATG |
| | | | | CTGTCGACATGTATCAGTTAATG |
| CL3302.Contig1_All | Acetylajmalan acetyl esterase | 3003 | Down | TGTCGACATGTATCAGTTAATG |
| | | | | GATGGAGTGACAGTGGTTGG |
| CL2061.Contig4_All | Translation initiation factor IF-2 | 1191 | Down | TTACATCCACAGTTTCACAAATC |
| | | | | GGCTTCCACTTCACATTCC |

*Up or down regulation indicated the expression in the Y4 vs. Y2 comparison.

40S ribosomal protein S20-2 isofrom 1 (Vitis vinifera, XP_002265347.1), phenylcoumaran benzylic ether reductase homolog F11 (Forsythia × intermedia, AAF64174.1), glucose-6-phosphate dehydrogenase (Capsicum annuum, AAT75322.1), and chlorophyllase (Sesbania rostrata, BAG55223.1). Among the top 10 least abundant proteins, eight could be aligned to proteins with known functions in GenBank. These proteins were annotated as acetylajmalan acetyl esterase (Striga asiatica, ABD98038.1), glyoxalase I (Avicennia marina, AAK06838.1), acetylajmalan acetyl esterase (Striga asiatica, ABD98038.1), brassinosteroid-regulated protein BRU1 (V. vinifera, XP_002270375.1), methyltransferase DDB_G0268948 (Solanum lycopersicum, XP_004237334.1), protein VITISV_028080 (V. vinifera, CAN70727.1), miraculin-like protein (V. vinifera, XP_002266302.2), and PRUPE_ppa005409mg (Prunus persica, EMJ12790.1). Interestingly, proteins playing important roles in paulownia stress tolerance were found in the comparison of Y4 vs. Y2. These proteins include XP_002266352.1, XP_002329905.1, XP_004234931.1, XP_002513962.1, AFA35119.1, XP_004229610.1, XP_004234945.1, XP_002512963.1, XP_002263538.1, XP_004146530.1, CAA54303.1, XP_004236869.1, XP_004246134.1, AFB74186.1, AEO19903.1, XP_002276841.1, AFU95415.1, ABB16972.1, AHF08831.1, NP_178050.1, XP_002333019.1, AAK06838.1, XP_004234985.1, P32980.1, AEA86324.1, CBI20688.3, ABK32073.1, ABK55669.1, XP_002320236.1, AFC01205.1, CAA06961.1, EMJ12650.1, ABO97414.1, AEO19903.1, AFA35119.1, CBI20065.3, O49996.1, CAJ00339.1, ADK70385.1, NP_001117239.1, ABK23883.1, CAJ19270.1, CAJ43709.1, XP_003626827.1, BAF62340.1, CAA82994.1, AEO45784.1,

Frontiers in Plant Science | www.frontiersin.org 4 June 2016 | Volume 7 | Article 892
Comparison of the Abundance Profiles of DAPs between the Y2 and Y4 Samples

Based on the Y4 vs. Y2 comparison, DAPs were retrieved and analyzed using the GO, COG, and KEGG databases. The DAPs were assigned to 51 functional groups under the three main GO categories (Figure 5 and Table S3). Under biological process, metabolic process and cellular process were the most represented groups; under cellular component, cell and cell part were the two largest groups; and under molecular function, catalytic activity and binding were the most represented. Interestingly, the GO classifications for the DAPs were similar to the classifications for all the identified proteins. The most enriched GO terms under biological process, cellular component, and molecular function were the same for all the proteins and for the DAPs subset (Table 2). We suggest that these enriched GO terms may be related to autotetraploidy.

To predict and classify the possible functions of the DAPs, we assigned them to COG categories. Based on their sequence homology with known proteins in GenBank, 355 DAPs, which comprised 11.98% of all the proteins, were divided into 20 categories (Table 3). The “posttranslational modification, protein turnover, chaperones” category contained 48 proteins and was the largest, followed by “energy production and conversion” (47), “translation, ribosomal structure and biogenesis” (46), “general function prediction only” (41), “carbohydrate transport and metabolism” (38), “amino acid transport and metabolism” (27), and “coenzyme transport and metabolism” (17). The smallest group was “intracellular trafficking, secretion, and vesicular transport” with only one protein (Figure S2).

The DAPs were mapped to 77 KEGG metabolic pathways (Table S4): “metabolic pathways,” “biosynthesis of secondary
metabolites,” “ribosome,” and “photosynthesis” were highly enriched, followed by “glyoxylate and dicarboxylate metabolism,” “amino sugar and nucleotide sugar metabolism,” “pyruvate metabolism,” “carbon fixation in photosynthetic organisms,” “other glycan degradation,” “glycolysis/gluconeogenesis,” “starch and sucrose metabolism,” and “protein processing in endoplasmic reticulum.”

**Correlation between Proteins and Transcripts**

To compare changes in protein abundance with transcript levels, we compared the more and less abundant DAPs identified by iTRAQ with the up- and down-regulated differentially expressed genes identified from a previous transcriptome analysis. Differentially expressed unigenes were screened based on an absolute fold change value of log2 ratio \( >1 \) with \( p < 0.001 \) and FDR \(< 0.001 \). A total of 1808 corresponding protein-encoding genes were detected by RNA sequencing. We found that 59 of the genes that showed noteworthy changes in their expression levels encoded proteins with corresponding changes in their abundance levels; 37 of these genes were up-regulated and 22 were down-regulated, while the corresponding proteins were more and less abundant, respectively (Table S5). A further 48 genes that showed noteworthy changes in their expression levels encoded proteins with opposite changes in their abundance levels (Table S5). We found 252 genes that showed noteworthy changes in transcript levels, while the corresponding proteins showed no changes in their abundances (Table S6). Conversely, 356 proteins showed noteworthy changes in their abundance levels, while the corresponding encoding genes showed no changes in their expression levels (Table S7).

**Confirmation of Differentially Expressed Proteins by qRT-PCR**

To confirm the differential expression of proteins identified by iTRAQ analysis, qRT-PCR was performed to detect the transcript expression levels of the corresponding genes. In Y4 vs. Y2, the qRT-PCR results shown that 15 DAPs were consistent with the iTRAQ LC-MS/MS analysis results, 2 DAPs displayed opposite protein and mRNA expression patterns (Figure 6), 1 transcript of the DAP failed to be amplified in the qRT-PCR experiment. This discrepancy may be attributed to posttranscriptional and posttranslational regulatory processes.
FIGURE 3 | Gene Ontology classification of distinct proteins that were detected in “Yuza 1” leaves 2683 proteins (90.55% of total) were categorized into 51 function groups.

FIGURE 4 | COG function classification of distinct proteins that were detected in “Yuza 1” leaves protein or domains were annotated and divided into 23 specific categories.
FIGURE 5 | Gene Ontology (GO) classification of DAPs that were detected in “Yuza 1” leaves 321 proteins (69.33% of total) were categorized into 41 function groups. (A) Biological process of GO; (B) cellular component of GO; (C) molecular function of GO.
TABLE 2 | Statistical analysis of the biological process, cell component and molecular function of the differentially abundant proteins (DAPs) in the Y4 vs. Y2 comparison.

| Ontology Class                          | Class                                      | DAP-Number |
|----------------------------------------|--------------------------------------------|------------|
| Biological process                     | Biological regulation                      | 28         |
| Cell component organization or biogenesis | Cellular component organization or biogenesis | 22         |
| Cellular process                       | Cellular process                           | 185        |
| Developmental process                  | Developmental process                      | 14         |
| Establishment of localization          | Establishment of localization              | 19         |
| Growth                                 | Growth                                     | 9          |
| Immune system process                  | Immune system process                      | 3          |
| Localization                          | Localization                               | 19         |
| Metabolic process                      | Metabolic process                          | 239        |
| Multi-organism process                 | Multi-organism process                     | 9          |
| Multicellular organism process         | Multicellular organism process              | 9          |
| Negative regulation of biological process | Negative regulation of biological process | 3          |
| Positive regulation of biological process | Positive regulation of biological process | 1          |
| Regulation of biological process       | Regulation of biological process           | 26         |
| Reproduction                           | Reproduction                               | 2          |
| Reproductive process                   | Reproductive process                       | 2          |
| Response to stimulus                   | Response to stimulus                       | 69         |
| Signaling                              | Signaling                                  | 4          |
| Single-organism process                | Single-organism process                    | 57         |
| Cell component                         | Cell                                       | 200        |
| Cell junction                          | Cell junction                              | 2          |
| Cell part                              | Cell part                                  | 200        |
| Extracellular region                   | Extracellular region                       | 30         |
| Macromolecular complex                 | Macromolecular complex                     | 72         |
| Membrane                               | Membrane                                  | 105        |
| Membrane part                          | Membrane part                              | 32         |
| Membrane-enclosed lumen                | Membrane-enclosed lumen                    | 24         |
| Nucleoid                               | Nucleoid                                   | 2          |
| Organelle                              | Organelle                                  | 149        |
| Organelle part                         | Organelle part                             | 92         |
| Symplast                               | Symplast                                   | 1          |
| Molecular function                     | Antioxidant activity                       | 14         |
| Binding                                | Binding                                    | 159        |
| Catalytic activity                     | Catalytic activity                         | 188        |
| Electron carrier activity              | Electron carrier activity                  | 15         |
| Enzyme regulator activity              | Enzyme regulator activity                  | 6          |
| Molecular transducer activity          | Molecular transducer activity              | 1          |
| Nutrient reservoir activity             | Nutrient reservoir activity                 | 1          |
| Receptor activity                      | Receptor activity                          | 1          |
| Structural molecule activity           | Structural molecule activity               | 30         |
| Transporter activity                   | Transporter activity                       | 8          |

DISCUSSION

Insights into the Proteomic and Transcriptomic Analyses of Autopolyploids

An important aspect of polyploid genomics research is the observation of genome expansion and contraction. The C-value, which is a measure of the amount of DNA in the unreplicated gametic nucleus, is often used to estimate the ploidy level in plants (Udall et al., 2006). It is assumed that polyploids have large C-values that increase in direct proportion with ploidy (Poggio et al., 2014). However, all three possible fates of genome size evolution after polyploidization, namely expansion, contraction, and no change, have been reported. Genome expansion has been reported in only a few studies. In a recent study of five Nicotiana polyploids, the genomes of the polyploids were found to have increased in size (Leitch and Leitch, 2008). Conversely, a large-scale analysis of 3008 angiosperms that combined available genome size data, suggested that genome contraction was a widespread biological response after polyploid formation (Leitch and Bennett, 2004; Ammiraju et al., 2010). It is also been reported that in some polyploid series, there is no change in genome size after polyploidization, and such phenomenon especially takes place in newly-formed ones (Yang et al., 2011).

In our work, the genome size of the autotetraploid Paulownia was found to be less than twice the diploid genome size, suggesting genome contraction had occurred in the process of autopolyploidization. We predicted that the contraction may be due to an insertion/deletion bias that could result in frequent and larger deletions and result in a reduction in genome size because of an increase in illegitimate recombinations (Grover et al., 2007).

Our proteomic and transcriptomic data gave quite different abundance/expression results. This may be because gene
expression differences were not translated into protein abundance differences because of posttranscriptional regulation and/or posttranslational modification, or because differences between the autotetraploid and its diploid parent were underestimated at the proteomic level because the iTRAQ-based analysis determines the number of DAPs based on an estimate of the fold change of different samples. It has been reported that iTRAQ-based methods may have problems with accuracy because they may systematically underestimate ratios, especially when the ratio changes are large (Bantscheff et al., 2008; Ow et al., 2009; Karp et al., 2010). In this work, as in most iTRAQ studies, fold changes ≥1.2 or ≤0.8 were employed because, in iTRAQ quantitation, small changes may reflect larger differences.

**Correlation between Protein Abundance and Transcript Expression Is Limited**

To interpret biological processes, such as gene expression, protein interactions, and the structures and functions of cellular systems, it is essential to investigate the correlation of protein abundance and mRNA transcript levels. However, the coefficient of correlation between them has been found to be different among different studies and different plant tissues. In yeast, good correlation has been reported between transcript levels and protein abundances (Lackner et al., 2012), while in most plant tissues, such as *Arabidopsis thaliana* leaves and roots (Lan et al., 2012; Ng et al., 2012), *Brassica napus* (Marmagne et al., 2010), *Tragopogon mirus* (Koh et al., 2012), and *Triticum aestivum* (Song et al., 2007), only limited correlation has been found. In general, low correlation between protein abundance and mRNA expression levels is more often observed and is more widespread. In a study conducted by Lan et al. RNA sequencing and iTRAQ-based proteomics were both employed to generate genomic expression and proteomic abundance data of *Arabidopsis* roots, and discordant changes were reported. However, interestingly, they found that for highly up-regulated genes, the concordance between expression levels and protein abundance was generally high (Lan et al., 2012). In our study, we used RNA sequencing data and iTRAQ-based quantitative proteomics to compare the transcript levels and protein abundances in “Yuza 1” autotetraploid and diploid Paulownia leaves. Among the 2963 proteins detected by iTRAQ, 1808 genes that encoded these proteins were detected in the RNA sequencing data. Among them, 37 up-regulated genes corresponded to the more abundant proteins, and 22 down-regulated genes corresponded to the less abundant proteins in the Y4 vs. Y2 comparison. The expressions of a further 48 genes showed an opposite trend to the abundances of the corresponding proteins. We found that most of the genes and corresponding proteins that exhibited the same trend at the transcriptional and translational levels, respectively, were related to response to stimulus or oxidation-reduction. In an earlier study, we also found that up-regulated genes in Paulownia autotetraploids were significantly related with response to stimulus (Li et al., 2014). These data imply that changes in transcript levels may be one of the main determinants for changes in protein abundances.

Several factors may explain the reported lack of concordance between protein abundance and mRNA expression levels.
First, it is widely assumed that posttranscriptional regulation and posttranslational modifications can control translational efficiency, and that these processes could lead to discordance between mRNA expression levels and protein abundance. For example, small RNAs, including microRNAs, which regulate the expression of their target genes, play important roles in many biological processes (Ha et al., 2009). Second, the limitations of the different experimental techniques in detecting the expression of mRNAs and abundance of proteins could contribute in a major way to this problem. Third, the use of different plant tissues might lead to different results because the transcriptional and translational processes are both temporal and spatial. Proteomic studies on whole tissues may draw a more comprehensive picture, but studies on specific organelles will likely show differences more exactly. In summary, despite the limited correlation, integrated analyses of the transcriptome and the proteome data are essential for understanding the effects of polyploidization.

AUTHOR CONTRIBUTIONS

GF conceived the experiments. YD wrote the paper. MD performed the experiments. ZZ analyzed the data. YD contributed reagents/materials/analysis tools.

FUNDING

This work was supported by the Fund of the Transformation Project of the National Agricultural Scientific and Technological Achievement of China (2012GB2D000271), the Central Financial Forestry Science Promotion Project (GTH [2012]01), the Fund of the Technology Innovation Team Project of Zhengzhou (121PCXTD515), the Joint Funds of the National Natural Science Foundation of China (NSFC) (Grant No. U1204309), the Fund of the Research Cooperation Project of Henan Province (Grant No. 152107000097) and the Fund of Zhongyuan Scholarship Foundation of Henan Province (122101110700).

ACKNOWLEDGMENTS

We thank WZ for their interest and invaluable assistance during this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2016.00892
Table S1 | KEGG pathway analysis of all identified proteins for “Yuza 1.”
Table S2 | Differentially abundant proteins identified in the comparison of Y4 vs. Y2 Up-regulated proteins were marked in red while down-regulated ones in green.
Table S3 | Gene Ontology analysis of differentially abundant proteins in the comparison of Y4 vs. Y2 321 proteins (69.33% of total) were categorized into 41 function groups.
Table S4 | KEGG analysis of the identified differentially abundant proteins in the comparison of Y4 vs. Y2.
Table S5 | Proteins changed noteworthy in expression levels with corresponding in transcript level in the comparison of Y4 vs. Y2. Same trend were marked in red while opposite ones in green.

Table S6 | Proteins showed no changes in their translational levels while noteworthy changes in transcript level in the comparison of Y4 vs. Y2.

REFERENCES

Accamando, A. K., and Cronin, J. T. (2012). Costs and benefits of jasmonic acid induced responses in soybean. Environ. Entomol. 41, 551–561. doi: 10.1603/EN11277

Amirraju, J. S., Fan, C., Yu, Y., Song, X., Cranston, K. A., Pontaroli, A. C., and Kuster, B. (2008). Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer. Mol. Cell. Proteomics 7, 1702–1713. doi: 10.1074/mcp.M800029-MCP200

Casneuf, T., De Bodt, S., Raes, J., Maere, S., and Van de Peer, Y. (2006). Nonrandom flowering plant genome downsizing in polyploid plants: GENOME DOWNSIZING IN POLYPLOIDS. Biol. J. Linnean Soc. 82, 651–663. doi: 10.1111/j.1095-8312.2004.00349.x

Dong, Y., Fan, G., Deng, M., Xu, E., and Zhao, Z. (2014a). Genome-wide expression profiling of the transcripts of four Paulownia tomentosa accessions in response to drought. Genomics 104, 295–305. doi: 10.1016/j.ygeno.2014.08.008

Dong, Y., Fan, G., Zhao, Z., and Deng, M. (2014b). Complementary proteome and metabolome study of paulownia and its related species. Genomics 55, 73–83. doi: 10.1016/j.ygeno.2014.09.057

Lin, X. Q., Liang, S. L., Han, S. Y., Zheng, S. P., Ye, Y. R., and Lin, Y. (2013). Quantitative iTRAQ LC-MS/MS proteomics reveals the cellular response to heterologous protein overexpression and the regulation of HAC1 in Pichia pastoris. J. Proteomics 91, 58–72. doi: 10.1016/j.jprot.2013.06.031

Marmagne, A., Brabant, P., Thiellement, H., and Alix, K. (2010). Analysis of genes related to witches' broom disease in Paulownia fortunei by a de novo assembled transcriptome. PLoS ONE 6:e28038. doi: 10.1371/journal.pone.0080283

Manzaneda, A. J., Rey, P. J., Bastida, J. M., Weiss-Lehman, C., Raskin, E., and Mittell-Olds, T. (2012). Environmental aridity is associated with cytotype segregation and polyploidy occurrence in Brachypodium distachyon (Poaceae). New Phytol. 193, 797–805. doi: 10.1111/j.1469-8137.2011.03988.x

Ng, D. W., Zhang, C., Miller, M., Shen, Z., Briggs, S. P., and Chen, Z. J. (2012). Discovering of microRNAs and transcript targets related to witches' broom disease in Paulownia fortunei by high-throughput sequencing and degradome approach. Mol. Genet. Genomics 291, 181–191. doi: 10.1007/s00438-015-1102-y

O’Brien, R. N., Shen, Z., Tachikawa, K., Lee, P. A., and Briggs, S. P. (2010). Quantitative proteome analysis of pluripotent cells by iTRAQ mass tagging.
reveals post-transcriptional regulation of proteins required for ES cell self-renewal. *Mol. Cell. Proteomics* 9, 2238–2251. doi: 10.1074/mcp.M110.000281
Otto, S. P. (2007). The evolutionary consequences of polyploidy. *Cell* 131, 452–462. doi: 10.1016/j.cell.2007.10.022
Ow, S. Y., Salim, M., Noirel, J., Evans, C., Rehman, I., and Wright, P. C. (2009). iTRAQ underestimation in simple and complex mixtures: “the good, the bad and the ugly.” *J. Proteome Res.* 8, 5347–5355. doi: 10.1021/pr900634c
Poggio, L., Realini, M. F., Fourastié, M. F., Garcia, A. M., and González, G. E. (2014). Genome downsizing and karyotype constancy in diploid and polyploid congers: a model of genome size variation. *AoB Plants* 6:plu029. doi: 10.1093/aobpla/plu029
Qiao, J., Wang, J., Chen, L., Tian, X., Huang, S., Ren, X., et al. (2012). Quantitative iTRAQ LC-MS/MS proteomics reveals metabolic responses to biofuel ethanol in cyanobacterial *Synechocystis* sp. *PCC* 6803. *J. Proteome Res.* 11, 5286–300. doi: 10.1021/pr300504w
Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., et al. (2004). Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* 3, 1154–1169. doi: 10.1074/mcp.M400129-MCP200
Saleh, B., Allario, T., Dambier, D., Ollitrault, P., and Morillon, R. (2008). Tetraploid citrus rootstocks are more tolerant to salt stress than diploid. *C. R. Biol.* 331, 703–710. doi: 10.1016/j.crvi.2008.06.007
Shen, Y., Zhang, Y., Zou, J., Meng, J., and Wang, J. (2015). Comparative proteomic study on *Brassica* hexaploid and its parents provides new insights into the effects of polyploidization. *J. Proteomics* 112, 274–284. doi: 10.1016/j.jprot.2014.10.006
Soltis, P. S., and Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* 60, 561–588. doi: 10.1146/annurev.arplant.043008.092039
Song, X., Ni, Z., Yao, Y., Xie, C., Li, Z., Wu, H., et al. (2007). Wheat (*Triticum aestivum* L.) root proteome and differentially expressed root proteins between hybrid and parents. *Proteomics* 7, 3538–3557. doi: 10.1002/pmic.200700147
Stebbins, G. L. Jr. (1947). Types of polyploids; their classification and significance. *Adv. Genet.* 1, 403–429.
Tang, Z., Du, W., Du, X. L., Ban, Y. Y., and Cheng, J. L. (2016). iTRAQ protein profiling of adventitious root formation in mulberry hardwood cuttings. *J. Plant Growth Regul.* doi: 10.1007/s00344-015-9567-9. [Epub ahead of print].
Udall, J. A., Swanson, J. M., Nettleton, D., Percifield, R. J., and Wendel, J. F. (2006). A novel approach for characterizing expression levels of genes duplicated by polyploidy. *Genetics* 173, 1823–1827. doi: 10.1534/genetics.106.058271
Vizcaíno, J. A., Coords, A., del Toro, N., Dianes, J. A., Griss, J., Lavidas, I., et al. (2016). 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* 44, D447–D456. doi: 10.1093/nar/gkv1145
Vizcaíno, J. A., Deutsch, E. W., Wang, R., Coords, A., Reisinger, F., Rios, D., et al. (2014). ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* 32, 223–226. doi: 10.1038/nbt.2839
Weiss-Schneeweiss, H., Emadzade, K., Jang, T. S., and Schneeweiss, G. M. (2013). Evolutionary consequences, constraints and potential of polyploidy in plants. *Cyto genet. Genome Res.* 140, 137–150. doi: 10.1159/000351727
Wu, Z., Li, D., Meng, J., and Wang, H. (2010). “Introduction to SIMCA-P and its application,” in *Handbook of Partial Least Squares: Concepts, Methods and Applications*, eds V. Esposito Vinzi, W. W. Chin, J. Henseler, and H. Wang (Berlin; Heidelberg: Springer Berlin Heidelberg), 757–774.
Xu, E., Fan, G., Niu, S., Zhao, Z., Deng, M., and Dong, Y. (2014). Transcriptome-wide profiling and expression analysis of diploid and autotetraploid *Paulownia tomentosa* x *Paulownia fortunei* under drought stress. *PLoS ONE* 9:e113313. doi: 10.1371/journal.pone.0113313
Yang, X., Ye, C. Y., Cheng, Z. M., Tschaplinski, T. J., Wullschleger, S. D., Yin, W., et al. (2011). Genomic aspects of research involving polyploid plants. *Plant Cell Tissue Organ Cult.* 104, 387–397. doi: 10.1007/s11240-010-9826-1

**Conflict of Interest Statement**: 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.

**Copyright © 2016 Dong, Deng, Zhao and Fan. 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) or licensor 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.**