Comparative transcriptome analysis of two Ipomoea aquatica Forsk. cultivars targeted to explore possible mechanism of genotype dependent accumulation of cadmium

Ying-Ying Huang, Chuang Shen, Jing-Xin Chen, Chun-Tao He, Qian Zhou, Xiao Tan, Jiangang Yuan, and Zhongyi Yang

J. Agric. Food Chem., Just Accepted Manuscript • DOI: 10.1021/acs.jafc.6b01267 • Publication Date (Web): 06 Jun 2016

Downloaded from http://pubs.acs.org on June 8, 2016
Comparative transcriptome analysis of two *Ipomoea aquatica* Forsk. cultivars targeted to explore possible mechanism of genotype dependent accumulation of cadmium

Ying-Ying Huang, Chuang Shen, Jing-Xin Chen, Chun-Tao He, Qian Zhou, Xiao Tan, Jian-Gang Yuan and Zhong-Yi Yang*

State Key Laboratory for Biocontrol, School of Life Sciences, Sun Yat-Sen University, Xingang Xi Road 135, Guangzhou, 510275, China

*Corresponding Author
Mail address: Xingang Xi Road 135, Guangzhou, 510275, China.
Email: adszy@sysu.edu.cn
Tel: +86 2084113220
ABSTRACT

A low-shoot-Cd (QLQ) and a high-shoot-Cd cultivar (T308) of water spinach (Ipomoea aquatica Forsk.) were used to investigate molecular mechanism of the genotype difference in cadmium (Cd) accumulation. RNA-Seq under 9 h and 72 h cadmium exposures (5 mg L\(^{-1}\)) were undertaken to explore Cd induced genotype differences in molecular processes. In total, 253,747,540 clean reads were assembled into 57,524 unigenes. Among them, 6,136 and 10,064 unigenes were differentially expressed in QLQ and T308, respectively. Cell wall biosynthesis genes such as GAUT and laccase and three Cd efflux genes (Nramp5, MATE9 and YSL7) had higher expression levels in QLQ, while the genes in sulfur and glutathione metabolism pathway e.g. sulfate transporter and cysteine synthase showed higher expression levels in T308. These findings would be useful for further understanding of the mechanisms related to genotype-dependent Cd accumulation and developing the molecular assisted screening and breeding of low-shoot-Cd cultivars for water spinach.

KEYWORDS: Low-shoot-Cd accumulation, Water spinach (Ipomoea aquatica Forsk.), Transcriptome, DEGs, Molecular mechanisms
INTRODUCTION

Cadmium (Cd) is a non-essential element with high toxicity and can be accumulated in the human body over time from the digestion of food containing Cd, being a threat to human health with excessive intake\(^1,2\). Cd contamination has been a worldwide environmental problem\(^3\). Therefore, take efficient technologies to reduce Cd content in crops is crucially important. Wide variations among cultivars in the concentration of Cd have been reported in many species, including rice (\textit{Oryza sativa})\(^4\), asparagus bean (\textit{Dolichos sinensis})\(^5\), hot pepper (\textit{Capsicum annuum})\(^6,7\) and water spinach (\textit{Ipomoea aquatic} Forsk.)\(^8\). It has been indicated that low-Cd cultivars can be used to reduce the movement of Cd into the human diet\(^1\).

Water spinach is an important leafy vegetable in southern China, it can accumulate Cd easily when growing in Cd-contaminated soils\(^8,9\). A high-shoot-Cd cultivar (T308) and a low-shoot-Cd cultivar (QLQ) were identified in our previous studies\(^8,10\). When growing in Cd contaminated soil (0.593 mg kg\(^{-1}\) dry weight, DW), the differences between shoot concentrations of the two cultivars were 3.1~3.4 fold\(^9\).

A reciprocal grafting experiment of QLQ and T308 showed that the plants of T308 scion with QLQ rootstock accumulated less Cd than the shoot of non-grafted T308 and the plants of QLQ scion with T308 rootstock accumulated higher Cd than the shoots of non-grafted QLQ. These observations indicated that the genotype-dependent Cd accumulation of water spinach largely depends on bio-processes occurring in roots\(^11\). Cd-induced gene expression differences of the two water spinach cultivars (QLQ and T308) had been observed by suppression subtractive hybridization (SSH) in our
previous study\textsuperscript{12}. Although there are some studies focus on the molecular mechanisms of Cd hyperaccumulation in several species\textsuperscript{13, 14}, the molecular mechanisms of cultivar dependent difference in Cd accumulation of crops are poorly understood\textsuperscript{1, 15}. The objective of this study is to uncover the possible molecular mechanisms of low-shoot-Cd accumulation of water spinach at transcriptional level.

For water spinach, which has no reference genome, \textit{de novo} assembly using high-throughput deep sequencing technology is an effective approach for transcriptome analysis. RNA-Seq has been successfully used in studies of Cd stressed \textit{Noccaea caerulescens} and \textit{Sedum alfredii} which has also no reference genome\textsuperscript{16, 17}.

Accordingly, RNA-Seq is employed to explore the differences in root transcriptional responses to Cd stress between the two cultivars after short- (9 h) and long-term (72 h) Cd treatment. To our knowledge, this is the first attempt to use RNA-Seq to detect Cd-induced transcriptional differences between two water spinach cultivars with different Cd accumulation abilities. It is expected that the results will help to clarify molecular mechanisms that are responsible for low-shoot-Cd accumulation in water spinach and will be useful to build up breeding methodology of low Cd accumulation cultivars based on molecular assistant breeding techniques.

**METHODS AND MATERIALS**

**Plant Materials and Treatments**

Two water spinach cultivars, QLQ (low-shoot-Cd cultivar) and T308 (high-shoot-Cd cultivar) were used in this study. The two cultivars were selected from 30 cultivars in our previous study, and the differences between shoot Cd
concentrations of the two cultivars were 3.1~3.4 fold when grew in Cd contaminated soil (0.593 mg kg\(^{-1}\) dry weight, DW) \(^9\). When grew in Hoagland solution containing 6 mg L\(^{-1}\) CdCl\(_2\) for 24 h, the differences between shoot Cd concentrations was 1.6 fold \(^12\).

Seeds of the cultivars were surface disinfected with 1‰ carbendazim (m/v) for 0.5 h, then thoroughly washed with deionized water, and sown into pots filled with sand (3 plants per pot) and then were irrigated with a half strength Hoagland nutrient solution for 4 weeks. The plants were cultured at constant temperature condition (25±1°C) with a photoperiod of 14 h (12000 lx). Three Cd exposure lengths were then carried out as treatment by irrigating with the half strength Hoagland solution containing 5 mg L\(^{-1}\) CdCl\(_2\) for 0 h, 9 h and 72 h, respectively. The three treatment groups were denoted as E0, E9 and E72, respectively, and each treatment repeated 3 times.

The shoots and roots of the three plants for each treatment were harvested separately and then washed with deionized water. Thereafter one plant is randomly selected from each pot for further study.

**Determination of Cd Concentration**

The roots were washed thorough with tap water and soaked in 25 mM CaCl\(_2\) twice (5 min each time) to remove heavy metal ions from the root surface. All samples were thoroughly washed with deionized water, dried at 105°C for 30 min, and then dried at 70°C to constant weight. The dried samples were digested (HNO\(_3\): H\(_2\)O\(_2\), 5:1) and then analyzed by atomic absorption spectroscopy (Hitachi Z-5300,
Japan. A certified reference material (CRM) of plant (GBW-07603, provided by the National Research Center for CRM, China) with a Cd concentration of 0.38 mg kg$^{-1}$ was used to control the precision of the analytical procedures. Data statistical analyses including independent samples t-test and one-way ANOVA with the least significant difference (LSD) test were performed using excel 2013 and SPSS 18.0.

**RNA Extraction**

Total RNA were extracted separately from the root samples of the three plants as replicates for each treatment with the aid of the RNA Easyspin Isolation System (Aidlab, China) according to the manufacturer’s instruction, and the genomic DNA residues were removed during RNA extraction. RNA concentration and quality were tested in an Agilent 2100 Bioanalyzer (Agilent, USA). RNA quality was also confirmed by RNase free agarose gel electrophoresis. Equal amounts of total RNA extracted from the three replicate plants of each treatment were pooled to construct the cDNA library, so 6 cDNA libraries (2 cultivars × 3 treatments) were constructed in the present study. Pooling is a cost-effective strategy to identify gene expression profiles and has been well-justified based on statistical and practical considerations.

**cDNA Library Construction and Sequencing**

Poly (A) mRNA was isolated by using oligo-dT beads (Qiagen, USA) and broken down into short fragments (200 nt) by adding fragmentation buffer. First-strand cDNA was generated using random hexamer-primed reverse transcription and second-strand cDNA was synthesized using RNase H and DNA polymerase I.
The cDNA fragments were purified with a QIAquick PCR purification kit (Qiagen, USA) and resolved with EB buffer for end reparation, poly (A) addition and ligated to sequencing adapters. After agarose gel electrophoresis, suitable fragments (about 200 nt) were used to construct the final cDNA library and was sequenced using Illumina HiSeq™ 2000 (pair-end sequencing, length 125 bp).

**De Novo Transcriptome Assembly and Annotation**

Illumina adapter sequences were cut, and low quality reads that did not meet the quality threshold were removed by a Perl program (version 5.18.4). Raw reads containing adapter, reads with more than 10% unknown nucleotides, and low quality reads with over 50% low Q-value (≤5) base were removed as well. The clean reads were *de novo* assembled using Trinity 2.0. The resulting transcripts were annotated with BLASTX (version 2.2.29), a tool widely used to search sequence similarities, by comparing against the database of NCBI non-redundant protein database (Nr, http://www.ncbi.nlm.nih.gov/), Cluster of Orthologous Groups of proteins database (COG, http://www.ncbi.nlm.nih.gov/COG/), Kyoto Encyclopedia of genes and genomes (KEGG, http://www.genome.jp/kegg/) and Swiss-Prot (http://www.expasy.ch/sprot), with an e-value cut off of 10⁻⁵.

**Identification of Differentially Expressed Genes**

Clean reads were used to map reference sequence by SOAPaligner/soap2 (version 2.2.1), a tool designed for short sequence alignment. The expression levels of all genes were determined using the RPKM (reads per kb per million reads) method, according to the formula of RPKM (A)=10⁶C/(NL 10⁻³). In which, RPKM stands for
the expression of gene A; C stands for the number of reads uniquely mapped to gene A; N stands for the number of reads uniquely mapped to all genes; L stands for the total length of gene A. For genes with more than one alternative transcript, the longest transcript was used to calculate the RPKM \(^{20}\). The RPKM method can eliminate the influence of different gene lengths and sequencing discrepancies on the gene expression calculation. The false discovery rate (FDR) was calculated to adjust the threshold of \(p\) value using the method described by Benjamini et al \(^{21}\). Unigenes between each two samples were considered differentially expressed when the false discovery (FDR) \(\leq 0.001\) and the absolute value of \(\log_2\) Ratio\(\geq 1\).

**GO and Pathway Enrichment Analyses of DEGs**

To determine the main biological functions of the differentially expressed genes, Gene Ontology (GO, http://www.geneontology.org/) terms were assigned by using Blast2GO (version 2.8) to search the Nr database \(^{22}\). GO classification was performed in WEGO (http://wego.genomics.org.cn/cgi-bin/wego/index.pl) \(^{23}\), and the GO categorization results were expressed as 3 independent hierarchies for molecular function, biological process and cellular component. The KEGG pathway annotation was performed using BLASTX software against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database according to the method described by Kanehisa et al \(^{24}\). For GO enrichment analysis, all \(p\)-values were adjusted with the Bonferroni correction, and a corrected \(p\)-value \(\leq 0.05\) was chosen as the threshold value to determine significantly enriched GO terms. For KEGG enrichment analysis, the pathway with a FDR value \(\leq 0.05\) was considered as an enriched pathway. Pearson
correlation coefficient analysis and principal component analysis (PCA) on all samples were performed by R package (version 3.1.3, http://www.r-project.org/) using the log transformed RPKM values of all genes.

Quantitative Real-Time Reverse Transcription PCR

The expression levels of 20 genes were determined by quantitative real-time reverse transcription PCR (qRT-PCR). First strand cDNA was generated by purified total RNA using a PrimeScript™ RT regent kit (Takara, Japan). The water spinach actin (unigene0033718) was selected as a reference gene. All primers for qRT-PCR were shown in Table S1 (supporting information). qRT-PCR was performed on a LightCycler® 480 Real-Time PCR System (Roche, Germany) with SYBR GreenII PCR Master Mix (Takara, Japan). The qPCR cycling program was run as follows: 95°C for 3 min, followed by 45 cycles of 95°C for 15 sec and 60°C for 1 min. Each qRT-PCR analysis was performed in triplicate. Results were analyzed with the integrated LightCycler® 480 service software. Expression levels of the tested genes were determined by CT values and calculated by the ∆∆Ct method.

RESULTS

Cd Concentration of Water Spinach

Root dry weight in both two cultivars at E9 and E72 did not show significant differences compared with those at E0 (Figure S1). Shoot and root Cd concentrations of water spinach cultivars QLQ and T308 are shown in Figure 1. Concentrations in both cultivars increased with duration of Cd treatment. T308 always accumulated higher Cd in shoots than QLQ, but only the difference at E72 was significant.
The shoot Cd concentration in QLQ and T308 at E72 were 4.22 mg kg\(^{-1}\) and 8.78 mg kg\(^{-1}\) (dry weight basis), respectively, indicating that the genotype dependent Cd accumulation is reproducible for water spinach. After Cd treatment, root Cd concentrations in QLQ were always higher than that of T308, but the differences had no significance (\(p>0.05\)).

**De Novo Sequence Assembly and Functional Annotation**

RNA-Seq analyses of the two cultivars were performed on the Illumina system to explore the differences that may contribute to their metal-related traits at transcriptional level in roots. As mentioned before that there is no appropriate reference genome sequence available for water spinach, therefore all the 253,747,540 clean reads were *de novo* assembled using Trinity to construct unique consensus sequence. The assembly yielded a total of 57,524 unigenes with an N50 unigene size of 1,627 bp, an average length of 936.74 bp, a minimum length of 201 bp and a maximum length of 17,439 bp. A summary of the unigenes size distribution is given in Figure S2. The raw reads of RNA sequencing were deposited in the Sequence Read Archive (SRA) database of NCBI with the accession number of SRP071620.

Only 55.81\% (32104) unigenes of all the assembled unigenes (57,524) could be annotated to the 4 databases (Nr, Swiss-Prot, COG and KEGG). GO analysis was performed to classify the function of annotated unigenes. GO analysis categorized 16,394 unigenes belonging to cellular component, biological process and molecular function in which cell (8,049), metabolic process (8,734), and catalytic activity (7,981) was the dominant groups, respectively (Figure S3). In addition, 9,166 unigenes were
mapped to 125 KEGG pathways. The dominant pathway with most unigenes was metabolic pathways, followed by biosynthesis of secondary metabolite, ribosome and plant hormone signal transduction (Table S2).

The Pearson correlation between QLQ-72 h and QLQ-0 h was lower than that between QLQ-9 h and QLQ-0 h, while it was reversed in T308. The correlations between two cultivars decreased with duration of Cd treatment (Table S3). Principal component analysis (PCA) showed that 68.8% of the transcriptional changes associated with Cd stress were explained by the first two PCA components. The first and second component did not cluster the samples by Cd treatment time in both two cultivars, but the transcriptional response to Cd stress of QLQ and T308 could be separated by PCA analyses. For the first principal component, assignment of QLQ-72 h was similar to that of T308-9 h (Figure S4).

**Differential Expressed Genes (DEGs)**

Unigenes were commonly considered differentially expressed when the FDR \( \leq 0.001 \) and the absolute value of \( \log_2 \text{Ratio} \geq 1 \). RNA-Seq data showed that the gene expression profiles of the both cultivars changed significantly after exposing to 5 mg L\(^{-1}\) CdCl\(_2\) at E9 and E72 as compared to E0. The number of DEGs between two cultivars increased over time, and this result was comparable with the increment of the differences in shoot Cd concentrations between two cultivars (Table S4). DEGs between QLQ and T308 at E9 and E72 obtained from Venn analysis are shown in Figure 2. A total of 79 up-regulated and 98 down-regulated unigenes were shared by QLQ and T308 at different treatment time (Table S5).
Trend Analysis of DEGs

To examine the expression profile of the DEGs in QLQ and T308, trend analysis were performed by Short Time-series Expression Miner (STEM) Software. A total of 6,136 DEGs in QLQ were clustered into 8 profiles, and only 2 profiles recorded with a significance \( p < 0.05 \) (Q3 with 1,245 DEGs and Q4 with 881 DEGs) (Figure 3A). Of all the 10,064 DEGs in T308, 7,936 DEGs were clustered into 4 profiles and all of them were with significance \( p < 0.05 \) (Figure 3B). Among them, Profile T2, T3, T4, and T5 contained 2,047, 1,849, 1,790 and 2,250 DEGs, respectively. The GO functions within each main profile were then analyzed, and the enriched GO terms of QLQ and T308 were shown in Figure 3. In QLQ, Q4 profile contained most GO terms. In T308, T3 and T5 profile contained the most cellular component GO terms, while T4 profile contained most molecular function GO terms, and T2 and T3 profile contained the most biological process GO terms. Compared with QLQ, gene regulation at transcriptional level in T308 was more active coping with short-term Cd stress (E9). The major transcriptional regulation of QLQ occurred at E72.

Number of DEGs involved in the selected GO terms of the two cultivars at E0, E9 and E72 are shown in Table 1. For the cellular component GO terms, QLQ had more DEGs than T308 in the photosystem, nucleus, ribosomal submit and ribonucleoprotein complex. In QLQ, most of these DEGs were clustered to profile 4 where genes did not show significant difference at E9 but up-regulated at E72. While in T308, these DEGs showed different expression patterns, and a substantial part of them were clustered into profile 5 where genes up-regulated at E9 but down-regulated
at E72. However, GO terms and pathways related to DNA repair did not show
significant time dependent change as well as significant difference between the two
cultivars at both E9 and E72. Genes involved in DNA repair, such as *MSH, MLH, XRCC, RAD* and *REV1*, showed the same expression patterns (Table S6). In the GO
terms related to ion activities, such as response to metal ion, ion transport, ion
transmembrane transporter activity and ion binding, more unigenes were found to be
differentially expressed in T308 than in QLQ. T308 had more DEGs involved in the
cell wall related GO terms, such as cell wall, cell wall biogenesis and lignin
metabolism process, but QLQ had higher percentages of up-regulated genes (profile 4)
involved in these GO terms than that of T308.

**KEGG Pathway Classification of DEGs**

KEGG pathway enrichment analysis was also performed to further understand
the biological meanings of all the DEGs. In QLQ, 1391 unigenes were clustered to
112 KEGG pathways, in which, only 19 pathways were significantly enriched (*p*<0.05)
(Table S6). In T308, 1978 unigenes were clustered to 121 pathways, but only 23
pathways were significantly enriched (Table S7). The pathways that the DEGs
enriched were similar in the two cultivars. Interestingly, glutathione metabolism
pathway was only significantly enriched in T308.

**Genes in the Sulfur and Glutathione Metabolisms Pathway in Response to Cd**

KEGG pathway analysis showed that glutathione metabolism pathway was only
significantly enriched in T308. The DEGs involved in sulfur and glutathione (GSH)
metabolisms pathway were compared between the two cultivars (Figure 4)
Most of the genes involved in sulfur metabolism pathway, including sulfur transporter and ATP sulfurylase were induced by Cd treatment in both cultivars, and most of them had higher expression in T308 than in QLQ. It was also observed that the expression profile of cysteine synthase was similar to sulfur transporter. Glutathione-ascorbic acid (GSH-AsA) recycle related enzymes, such as glutathione reductase, glutathione peroxidase, ascorbate peroxidase and dehydroascorbate reductase were also expressed higher in T308. In general, glutathione S-transferases expressed higher in T308 at E72. Interestingly, the expression of GSH1 (encoding one of the limit enzymes for the GSH biosynthesis) and PCS1 did not show significant time dependent change as well as significant difference between the two cultivars at both E9 and E72.

**Genes Involved in Heavy Metal Transport and Detoxification**

Metal transporters and chelating compounds play an important role in coping with Cd stress by Cd exclusion, chelation and compartmentalization. In the present study, most heavy metal transporters showed higher expression levels in T308 than in QLQ, such as HMA2 (unigene0025087), IRT1 (unigene0006857), MRP-like ABC transporter (unigene0033818), PDR-like ABC transporter (unigene0026448), copper transporter 6-like (unigene0008597) and Nramp3 (unigene0025280). However, transporters Nramp5 (unigene0029032), YSL7 (unigene0029192) and MATE efflux family protein (unigene0037925), which are genes functioned to metal efflux, expressed at higher levels in QLQ than in T308 under Cd stress (Table 2).

Most of the genes associated with metal detoxification expressed at higher levels
in T308 than in QLQ at E9, while reversed at E72. Among them, most MTs were more highly expressed in QLQ after long-term Cd treatment (E72). MT1 (unigene0008188) and MT2 (unigene0022813) had high expression in both cultivars but they did not performed as typical DEGs between neither different time stages nor different cultivars. MT3 was slightly down-regulated in QLQ, and was greatly up-regulated at E9 and rapidly down-regulated at E72 in T308. The expression of MT3 at E0 and E72 were significantly higher in QLQ than in T308, but it reversed at E9. In addition, nicotianamine synthase was up-regulated about 3-fold and 1-fold in QLQ and T308 at E72, respectively.

Validation of RNA-Seq Data by qRT-PCR Analysis

Comparative analysis of the expression data obtained by RNA-Seq with qRT-PCR is a widely adopted approach for data validation. To confirm the reliability of the RNA-Seq data in the present study, 20 genes were selected and their expression were detected by qRT-PCR. A good consistency was verified between RNA-Seq and qRT-PCR results for the 20 candidate unigenes ($r^2=0.3936$, $p<0.0001$), indicating that the expression data obtained by RNA-Seq experiment of the present study was reliable (Figure 5). For each gene, the expression level of RNA-Seq data exhibited similar expression profile comparing with the results of qRT-PCR (Figure S5).

DISCUSSION

About 56% of unigenes are annotated in water spinach

Exploring the regulatory mechanisms underlying Cd accumulation relevant
processes including uptake, detoxification and translocation are of great importance to reduce Cd content in crops via molecular assisted breeding. For water spinach, a SSH study was conducted and 400 Cd induced genes differently expressed between QLQ and T308 were detected\textsuperscript{12}. In the present study, much more Cd-induced DEGs (6,136 in QLQ and 10,064 in T308) were identified using RNA-Seq technology, which may thus be helpful to further explain the mechanisms involving the genotype difference in Cd accumulation at transcriptional level.

Some analyses of transcriptome have been conducted to identify genes and regulatory mechanisms, which are responsible for heavy metal stress\textsuperscript{15-17}. In the present study, 32,104 unigenes were annotated to the 4 databases, which rated 55.81% of all the assembled unigenes (57,524). This proportion was comparable with other studies using next-generation sequencing techniques to characterize the transcriptome of non-model species, such as *Elodea nuttallii*, *Camellia sinensis*, and *Sedum alfredii*\textsuperscript{17,25,26}.

**A More Complicated and Faster Mechanism Responding to Cd stress was Developed in T308**

Comparing with QLQ, more Cd induced DEGs were identified in T308, and the DEGs of T308 could be clustered to more profiles, indicating that T308 had developed a more complicated mechanism to cope with Cd stress. PCA analysis showed that QLQ-72 h and T308-9 h was similar for the first principal component, suggesting that the response to Cd stress at transcriptional level in T308 at E9 was similar to that of QLQ at E72, indicating a delay in response to Cd stress comparing
with T308. Therefore, it is considered that the response to Cd stress in T308 is faster than in QLQ. Pearson correlation coefficient and expression profile analyses also suggested this observation.

Genes related to DNA repair, such as \( MSH, MLH, XRCC, RAD \) and \( REV1 \), did not differentially expressed in both two cultivars, suggesting that the two cultivars were not suffered from DNA damage until 72 h. This result could be connected to the fact that root biomasses in both tested cultivars were not significantly influenced by Cd treatment. QLQ had more Cd induced DEGs related to the photosystem, nucleus, ribosomal submit and ribonucleoprotein complex GO terms, which were considered to be involved in repairing cellular damages induced by Cd stress. The different expression of these genes displayed the possibility that the Cd tolerance of QLQ may be depending on the mechanisms to repair cellular damages induced by Cd treatment.

These results suggested that the two cultivars might have similar Cd tolerance, however, T308 had clearly developed a more complicated and efficient mechanism responding to Cd stress, which should be related to its high Cd accumulation ability in shoots.

**Cell Wall Biosynthesis Should Be One of the Key Processes Deciding the Genotype Dependent Shoot Cd Accumulation**

Cell wall is the first barrier against Cd stress in plant, Nishizono et al. (1989) indicated that about 70\%–90\% of the root cadmium was located in the cell wall. The pentose and glucuronate interconversions pathway and starch and sucrose metabolism pathway are related to cell wall biosynthesis, and most of the DEGs
involved in these two pathways expressed at higher levels in QLQ than in T308. Pectin, cellulose and lignin are important components of plant cell wall. Exogenous nitric oxide can enhance rice Cd tolerance through increasing pectin and hemicellulose contents in root cell wall. Galacturonosyltransferase (GAUT) and cellulose synthase is the key enzyme of plant cell wall pectin and cellulose synthesis, respectively. Laccase is involved in lignin synthesis. Most GAUT, cellulose synthase related unigenes and two laccase related genes (lac14 and lac17) were expressed at higher levels in QLQ. All these results suggested that the cell wall biosynthesis was more active in QLQ than in T308, and thus might be one of the key processes associating to the genotype difference in Cd accumulation.

In our previous studies, root Cd concentration was higher in QLQ than in T308, which may be attributed to that the phellem in the main root of QLQ was much thicker than that in T308 under Cd exposure, and thus reducing Cd translocation from roots to shoots. The results observed in the present study indicated that the high root but low shoot Cd concentration of QLQ may rely on an enhanced cell wall biosynthesis of roots.

**Regulation of Sulfur and Glutathione Metabolisms was More Active in T308**

Sulfur and glutathione metabolisms is one of the best studied mechanisms of heavy metal detoxification and antioxidant process, in which, glutathione (GSH), phytochelatins (PCs) and other low molecular weight sulfur-containing compounds play important roles. GSH is a low molecular weight tripeptide widely distributed in plant cells. It has been reported that the genes involved in sulfur assimilation and
GSH metabolism were induced in roots of Arabidopsis and rice when exposed to Cd. Sulfate uptake by roots is the first step of sulfur assimilation. It was found in the present study that sulfate transporter related unigenes were induced in the two tested cultivars at E72, suggesting that Cd might induce sulfate uptake by improving the expression of sulfate transporters. The synthesis of cysteine was more efficient in T308 than in QLQ. Such a response may represent an adaptive approach required to ensure a sufficient supply of sulfur to meet the needs of cysteine for GSH and PCs biosynthesis. This result is similar to the finding in Arabidopsis where an increased cysteine biosynthesis is essential for Cd tolerance and accumulation. However, the insignificant expression of GSH1 at E9 and E72 implied that the difference of GSH synthesis between the two cultivars might be mainly caused by the previous processes, i.e. cysteine biosynthesis.

AsA-GSH cycle played an important role in removing reactive oxygen species (ROS) produced during metabolic process. GSH is an important antioxidant of this cycle. Most of the expression of AsA-Cycles and synthesis of GSH related genes had higher expression in T308, indicating that the higher shoot Cd accumulation in T308 may be related to its higher antioxidant ability.

GSH was also involved in metal homeostasis. Glutathione S-transferase (GST) catalyzed the formation of GSH-Cd, and could help to compartmentalize heavy metal ions to the vacuole. PCs are synthesized from GSH catalyzed by phytochelatin synthase, and PCs can chelate metal ions to form metal-chelate compounds and sequester them in the vacuole. PCS1 was not significantly induced by Cd treatment.
in both two cultivars in the present study, suggesting that water spinach may tend to maintain a high GSH level to cope with the risk of oxidative stress caused by Cd. This result is consistent to the observation that Cd hypertolerance was not rely on PCs, and the synthesis of GSH was enhanced in *T. caerulescens*<sup>35,36</sup>. Nishikawa et al. (2006) also indicated that PCs content did not correlate closely with the level of Cd accumulation, and maintaining a high GSH level was more important for Cd accumulation in *Chlamydomonas* spp<sup>37</sup>. These results explained why the expression level of *PCS1* did not correlate with Cd accumulation in water spinach. It is considered that to maintain a high expression level of GSH metabolism related genes seemed to be a crucial way for Cd accumulation in water spinach.

**Heavy Metal Transport and Detoxification Related genes Responded Differently between QLQ and T308**

Heavy metal transporter related genes were usually enhanced in hyperaccumulators, and this was considered as a typical property to bring about heavy metal hyperaccumulation<sup>38</sup>. The GO analysis of DEGs showed that T308 had more DEGs in ion transport related GO terms, suggesting that the higher shoot Cd concentration in T308 may depend on the high ion transport activity.

In T308, the higher expression levels of most heavy metal transporters related to Cd translocation were observed. These genes including *IRT1, HMA2, MRP-like ABC transporter, PDR-like ABC transporter, copper transporter 6-like* and *Nramp3* have been proved to enhance Cd translocation from roots to shoots or to be associated with Cd detoxification in plants. The zinc/Fe transporter *IRT1* is the major entry of Cd into
plants and mediate the accumulation of additional metal ions \(^{39}\). IRT1 was also involved in the high-affinity Cd uptake in roots of \(T. caerulescens\) \(^{40}\). In addition, \(HMA\) family were suggested to be involved in heavy metal transport \(^{41, 42}\). Many studies have also exhibited that \(ABCCs\) involve in heavy metal tolerance and detoxification \(^{43, 44}\).

On the other hand, three genes related to Cd efflux, including \(Nramp5\), \(YSL7\) and \(MATE\) efflux family protein showed higher expression in QLQ than in T308 (Table 2). The processes regulating the Cd efflux should be the crucial mechanism leading to low Cd accumulation in QLQ, which has also been discovered in other plants. For example, a RNAi experiment indicated that \(Nramp5\) is a major Fe, Mn and Cd transporter in rice, and the \(Nramp5\) knockdown rice plants accumulated more Cd shoots \(^{45, 46}\). The yellow stripe-like (YSL) transporter family can participate in transporting of heavy metals, and it has been reported that \(TcYSL7\) was involved in the efflux of NA-complexes \(^{47}\). Furthermore, multidrug and toxic compound extrusion (MATE), as a member of ABC transporter super family, has been reported to regulate Cd detoxification via exporting Cd to outside of cytoplasm \(^{48, 49}\). Therefore, the metal efflux mechanism obviously participated the genotype difference in shoot Cd accumulation of water spinach.

\(MTs\) are ubiquitous low-molecular weight, cysteine-rich proteins, which can bind metal ions. In the present study, it was observed that the \(MTs\) (except \(MT1\)) expressed at higher levels in QLQ than in T308 at E72. Grispen et al. (2011) indicated that overexpression of \(AtMT2b\) enhanced Cd tolerance and reduce Cd translocation from
root to shoot in tobacco under 1.6, 6.4 and 25.6 μM Cd stress. Therefore, the higher expression of MTs may restrict Cd translocation from roots to shoots and thus result in higher root and lower shoot Cd accumulation in water spinach.

In summary, two main findings were concluded in the present study. Firstly, strategies to block Cd in roots, such as enhancing Cd detoxification system and cell wall biosynthesis, may be developed in QLQ. In T308, however, heavy metal transporters were more activated to further transport Cd to shoot. Secondly, GSH metabolisms seemed to play a more prominent role in Cd detoxification rather than PCS1 in water spinach. Furthermore, the Cd efflux genes including Nramp5, MATE and YSL7 may be responsible for the low Cd accumulation in QLQ. These results provided the directions of further studies on the molecular mechanisms of low-shoot-Cd accumulation. On the complicate Cd induced transcriptomic changes observed in the present study, further bioinformatic analysis and function verification of the explored genes and pathway are need. We believe that the new observations will help to clarify molecular mechanisms associating to Cd accumulation in water spinach and will be useful for speeding up the breeding process of low-shoot-Cd cultivars based on molecular assistant breeding technique.

**Funding Sources**

This research was supported by National Natural Science Foundation of China (Grant No 21277178).

**Notes:**

The authors declare no competing financial interest.
Supporting Information Available

Table S1. Primers used for qRT-PCR. Table S2. The KEGG pathway analysis of assembled unigenes. Table S3. The Pearson correlation coefficients between cultivars and treatment duration. Table S4. Number of DEGs between cultivars and treatment duration. Table S5. Unigenes up-regulated in two cultivars after short- and long-term Cd stress. Table S6. Genes involved in DNA repair. Table S7. The significantly enriched pathways of all the DEGs in QLQ. Table S8. The significantly enriched pathways of all the DEGs in T308. Figure S1. Root dry weight of QLQ and T308. Asterisks indicate significantly differences between the two cultivars (p<0.05). Error bars represent standard deviation (n=3). Figure S2. Length distribution of assembled unigenes. Figure S3. Gene ontology classification of assembled unigenes. Figure S4. Principal component analysis (PCA) for all RNA-Seq samples. Figure S5. Expression of the selected 20 genes inferred by RNA-Seq and qRT-PCR. Q0, Q9, Q72, T0, T9 and T72 represents QLQ-0 h, QLQ-9 h, QLQ-72 h, T308-0 h, T308-9 h and T308-72 h, respectively. Error bars represent standard deviation (n=3).

This material is available free of charge via the Internet at http://pubs.acs.org.
Reference


46. Gendre, D.; Czemic, P.; Conéjéro, G.; Pianelli, K.; Briat, J. F.; Lebrun, M.; Mari, S., TcYSL3, a member of the YSL gene family from the hyper-accumulator Thlaspi caerulescens, encodes a


FIGURE CAPTIONS

Figure 1. Cd concentration of shoots (left) and roots (right) in QLQ (a low-shoot-Cd cultivar) and T308 (a high-shoot-Cd cultivar). Cd concentration in shoots and roots of QLQ and T308 were determined by AAS after exposure to 5 mg L$^{-1}$ CdCl$_2$. **indicate significantly differences between the two cultivars ($p<0.01$). Error bars represent standard deviation (n=3).

Figure 2. Venn diagrams of differential expressed genes: up-regulated (left) and down-regulated (right) at E9 and E72 in two cultivars.

Figure 3. DEGs expression profiles and their GO terms functional enrichment analyses (A and B). Expression profiles of genes in the main clusters of QLQ (Q3 and Q4) and T308 (T2, T3, T4 and T5) obtained by STEM. Each compartment represented an expression profile, the grey lines represented the unigenes contained in this profile, and the thick black line represented the expression tendency of all these unigenes. The number of unigenes of each profile was shown above each compartment. The GO terms enrichment of 2 main profiles in QLQ (A) and 4 main profiles in T308 (B). The significances of the GO terms in each main profile were indicated using log-transformed P-value (red). The cellular component, molecular function and biological process GO terms were indicated with red, green and black font, respectively.
Figure 4. Expression levels of unigenes involved in sulfur and GSH metabolisms in roots of QLQ and T308. The expression heatmaps from left to right were arranged in the following order: QLQ-0h, QLQ-9h, QLQ-72h, T308-0h, T308-9h and T308-72h. Data for unigene expression level were normalized to z-score.

Figure 5. Correlation between qRT-PCR and RNA sequencing for the 20 selected genes. Data was log10 transformed and plotted against RNA-Seq data (n=120).
Table 1. The number of DEGS involved in selected GO terms of the two cultivars during 9 h to 72 h Cd treatment. QA means all profiles of QLQ, Q3 and Q4 means profile 3 and profile 4 of QLQ, TA means all profiles of T308, T2, T3, T4 and T5 means profile 2, profile 3, profile 4 and profile 5 of T308.

<table>
<thead>
<tr>
<th>GO terms</th>
<th>QA</th>
<th>Q3</th>
<th>Q4</th>
<th>TA</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cellular component</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0005618 cell wall</td>
<td>39</td>
<td>4</td>
<td>12</td>
<td>93</td>
<td>22</td>
<td>15</td>
<td>15</td>
<td>23</td>
</tr>
<tr>
<td>GO:0009521 photosystem</td>
<td>26</td>
<td>1</td>
<td>15</td>
<td>21</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>GO:0005634 nucleus</td>
<td>116</td>
<td>12</td>
<td>41</td>
<td>102</td>
<td>34</td>
<td>15</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>GO:0009536 plastid</td>
<td>372</td>
<td>47</td>
<td>72</td>
<td>507</td>
<td>84</td>
<td>112</td>
<td>51</td>
<td>185</td>
</tr>
<tr>
<td>GO:0005840 ribosome</td>
<td>114</td>
<td>9</td>
<td>53</td>
<td>177</td>
<td>70</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>GO:0033279 ribosomal subunit</td>
<td>117</td>
<td>6</td>
<td>45</td>
<td>62</td>
<td>40</td>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>GO:0005737 cytoplasm</td>
<td>690</td>
<td>89</td>
<td>128</td>
<td>915</td>
<td>211</td>
<td>159</td>
<td>139</td>
<td>268</td>
</tr>
<tr>
<td>GO:0030529 ribonucleoprotein complex</td>
<td>196</td>
<td>13</td>
<td>54</td>
<td>150</td>
<td>74</td>
<td>15</td>
<td>10</td>
<td>21</td>
</tr>
<tr>
<td><strong>Biological process</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0010038 response to metal ion</td>
<td>87</td>
<td>9</td>
<td>15</td>
<td>139</td>
<td>39</td>
<td>14</td>
<td>36</td>
<td>19</td>
</tr>
<tr>
<td>GO:0006790 sulfur metabolic process</td>
<td>12</td>
<td>2</td>
<td>4</td>
<td>24</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>GO:0042546 cell wall biogenesis</td>
<td>10</td>
<td>0</td>
<td>7</td>
<td>20</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>GO:0006811 ion transport</td>
<td>84</td>
<td>14</td>
<td>12</td>
<td>146</td>
<td>25</td>
<td>22</td>
<td>44</td>
<td>32</td>
</tr>
<tr>
<td>GO:0009808 lignin metabolic process</td>
<td>12</td>
<td>0</td>
<td>6</td>
<td>15</td>
<td>3</td>
<td>0</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>GO:0042445 hormone metabolic process</td>
<td>12</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>GO:0006796 phosphate metabolic process</td>
<td>160</td>
<td>44</td>
<td>20</td>
<td>348</td>
<td>64</td>
<td>65</td>
<td>90</td>
<td>65</td>
</tr>
<tr>
<td>GO:0051716 cellular response to stimulus</td>
<td>91</td>
<td>19</td>
<td>15</td>
<td>137</td>
<td>27</td>
<td>24</td>
<td>24</td>
<td>32</td>
</tr>
<tr>
<td>GO:0050896 response to stress</td>
<td>371</td>
<td>52</td>
<td>55</td>
<td>545</td>
<td>150</td>
<td>63</td>
<td>113</td>
<td>96</td>
</tr>
<tr>
<td><strong>Molecular function</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0016740 transferase activity</td>
<td>444</td>
<td>90</td>
<td>67</td>
<td>876</td>
<td>158</td>
<td>138</td>
<td>180</td>
<td>175</td>
</tr>
<tr>
<td>GO:0005198 structural molecule activity</td>
<td>142</td>
<td>9</td>
<td>48</td>
<td>108</td>
<td>51</td>
<td>14</td>
<td>2</td>
<td>22</td>
</tr>
<tr>
<td>GO:0005057 receptor signaling protein activity</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>28</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>GO:0015075 ion transmembrane transporter activity</td>
<td>73</td>
<td>15</td>
<td>12</td>
<td>145</td>
<td>28</td>
<td>21</td>
<td>47</td>
<td>28</td>
</tr>
<tr>
<td>GO:0016209 antioxidant activity</td>
<td>45</td>
<td>1</td>
<td>7</td>
<td>76</td>
<td>21</td>
<td>5</td>
<td>24</td>
<td>8</td>
</tr>
<tr>
<td>GO:0043167 ion binding</td>
<td>439</td>
<td>54</td>
<td>68</td>
<td>667</td>
<td>149</td>
<td>96</td>
<td>138</td>
<td>150</td>
</tr>
</tbody>
</table>
Table 2. Selected genes associated with Cd accumulation and tolerance.

<table>
<thead>
<tr>
<th>ID</th>
<th>Gene description</th>
<th>QLQ 0 h</th>
<th>QLQ 9 h</th>
<th>QLQ 72 h</th>
<th>T308 0 h</th>
<th>T308 9 h</th>
<th>T308 72 h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unigene0037925</td>
<td>PREDICTED: MATE efflux family protein 9-like</td>
<td>69.82</td>
<td>139.36</td>
<td>190.59</td>
<td>77.52</td>
<td>71.54</td>
<td>131.61</td>
</tr>
<tr>
<td>Unigene0025087</td>
<td>PREDICTED: cadmium/zinc-transporting ATPase HMA2-like</td>
<td>19.07</td>
<td>17.51</td>
<td>10.05</td>
<td>25.93</td>
<td>29.38</td>
<td>15.60</td>
</tr>
<tr>
<td>Unigene0009244</td>
<td>heavy metal ATPase</td>
<td>5.78</td>
<td>5.33</td>
<td>5.05</td>
<td>4.82</td>
<td>6.23</td>
<td>10.50</td>
</tr>
<tr>
<td>Unigene006857</td>
<td>iron transporter protein IRT1</td>
<td>32.69</td>
<td>158.32</td>
<td>480.60</td>
<td>231.52</td>
<td>218.35</td>
<td>597.53</td>
</tr>
<tr>
<td>Unigene0023769</td>
<td>PREDICTED: ABC transporter B family member 25-like</td>
<td>12.83</td>
<td>6.82</td>
<td>15.40</td>
<td>17.88</td>
<td>10.22</td>
<td>40.64</td>
</tr>
<tr>
<td>Unigene0039216</td>
<td>PREDICTED: ABC transporter C family member 3-like</td>
<td>2.56</td>
<td>1.10</td>
<td>2.07</td>
<td>1.661</td>
<td>1.33</td>
<td>10.82</td>
</tr>
<tr>
<td>Unigene0033818</td>
<td>MRP-like ABC transporter</td>
<td>19.67</td>
<td>17.25</td>
<td>15.51</td>
<td>22.51</td>
<td>29.71</td>
<td>52.00</td>
</tr>
<tr>
<td>Unigene0026448</td>
<td>PDR-like ABC transporter</td>
<td>5.59</td>
<td>6.70</td>
<td>9.44</td>
<td>11.71</td>
<td>2.63</td>
<td>22.45</td>
</tr>
<tr>
<td>Unigene0030973</td>
<td>PREDICTED: ABC transporter G family member 36-like</td>
<td>207.64</td>
<td>166.74</td>
<td>195.78</td>
<td>160.99</td>
<td>74.23</td>
<td>283.41</td>
</tr>
<tr>
<td>Unigene0008597</td>
<td>PREDICTED: copper transporter 6-like</td>
<td>2.25</td>
<td>1.16</td>
<td>6.25</td>
<td>2.83</td>
<td>3.47</td>
<td>21.17</td>
</tr>
<tr>
<td>Unigene0053235</td>
<td>metal transporter</td>
<td>8.46</td>
<td>8.39</td>
<td>18.58</td>
<td>7.50</td>
<td>3.71</td>
<td>23.91</td>
</tr>
<tr>
<td>Unigene0025280</td>
<td>Nramp3</td>
<td>2.50</td>
<td>13.37</td>
<td>6.24</td>
<td>6.31</td>
<td>40.48</td>
<td>31.78</td>
</tr>
<tr>
<td>Unigene0035060</td>
<td>PREDICTED: metal tolerance protein 10</td>
<td>13.42</td>
<td>32.68</td>
<td>27.21</td>
<td>48.63</td>
<td>37.72</td>
<td>107.22</td>
</tr>
<tr>
<td>Unigene0029032</td>
<td>Nramp5</td>
<td>1.49</td>
<td>0.90</td>
<td>1.68</td>
<td>0.57</td>
<td>0.68</td>
<td>0.94</td>
</tr>
<tr>
<td>Unigene0029192</td>
<td>probable metal-nicotiamine transporter YSL7-like isoform 2</td>
<td>24.86</td>
<td>28.88</td>
<td>58.70</td>
<td>11.22</td>
<td>4.35</td>
<td>28.51</td>
</tr>
<tr>
<td>Unigene0054660</td>
<td>metallothionein-like protein</td>
<td>30.92</td>
<td>56.53</td>
<td>17.46</td>
<td>17.18</td>
<td>23.34</td>
<td>7.70</td>
</tr>
<tr>
<td>Unigene0008188</td>
<td>metallothionein-like type 1 protein</td>
<td>3776.56</td>
<td>3498.57</td>
<td>2852.20</td>
<td>4720.24</td>
<td>6843.60</td>
<td>4437.47</td>
</tr>
<tr>
<td>Unigene0022813</td>
<td>metallothionein-like type 2 protein</td>
<td>2227.63</td>
<td>2035.24</td>
<td>2958.26</td>
<td>1689.33</td>
<td>2252.59</td>
<td>2052.91</td>
</tr>
<tr>
<td>Unigene0021970</td>
<td>metallothionein type 3-like protein</td>
<td>2843.30</td>
<td>2629.41</td>
<td>1809.33</td>
<td>877.93</td>
<td>3459.76</td>
<td>818.88</td>
</tr>
<tr>
<td>Unigene0003113</td>
<td>phytochelatin synthase</td>
<td>12.41</td>
<td>12.61</td>
<td>13.95</td>
<td>11.28</td>
<td>11.86</td>
<td>18.44</td>
</tr>
<tr>
<td>Unigene0032198</td>
<td>PREDICTED: nicotiamine synthase-like</td>
<td>34.74</td>
<td>98.20</td>
<td>126.26</td>
<td>70.26</td>
<td>142.74</td>
<td>159.69</td>
</tr>
</tbody>
</table>
Figure 1
Figure 2
Figure 3
Figure 4
Figure 5

$q^2 = 0.3936$
$p < 0.0001$
Comparative transcriptome analysis of two *Ipomoea aquatica* Forsk. cultivars targeted to explore possible mechanism of genotype dependent accumulation of cadmium

Ying-Ying Huang, Chuang Shen, Jing-Xin Chen, Chun-Tao He, Qian Zhou, Xiao Tan, Jian-Gang Yuan and Zhong-Yi Yang*