Morphology, metabolomic and transcriptomic analysis revealed the mechanism of foliar application of triacontanol enhances Cd enrichment in *Tagetes patula* L.

Luqi Mi, Yetong Liu, Qingqing Huang, Lijie Zhao, Xu Qin*, Yuebing Sun*, Boyan Li

Supplementary Material (Text)

Captions

Text S1 Chemical forms of Cd in *T. patula* leaves.

Text S2 Metabolomic analysis steps *of T. patula* roots.

Text S3 Transcriptomic analysis steps *of T. patula* roots.

Text S1 Chemical forms of Cd in *T. patula* leaves

Frozen leaves were grounded to powder with pre-chilled mortar, and placed in 50-mL plastic centrifuge tubes. The leaves were mixed with extraction solutions at a ratio of 1:10 (w/v), and the mixture was incubated at 25 °C and 100 rpm/min for 22 h, and then the homogenate was centrifuged at 4000 g for 20 min. The supernatants were separated and then evaporated on an electric plate at 70 °C until constant weight. The sediment underwent the same procedure with the next extraction solutions in the indicated order. Each of the evaporated supernatants was digested with 8 mL HNO_3. The concentrations of different chemical forms of Cd were measured by ICP-MS.

Text S2 Metabolomic analysis steps *of T. patula* roots

1. Metabolite Extraction

50 mg roots sample were accurately weighed, and the metabolites were extracted using a 400 μL methanol:water (4:1, v/v) solution with 0.02 mg/mL L-2-chlorophenylalanin as internal standard. The mixture was allowed to settle at -10 ℃ and treated by High throughput tissue crusher Wonbio-96c (Shanghai Wanbo Biotechnology Co., LTD) at 50 Hz for 6 min, then followed by ultrasound at 40 kHz for 30 min at 5 ℃. The samples were placed at -20 ℃ for 30 min to precipitate proteins. After centrifugation at 13000 g at 4 ℃ for 15 min, the supernatant were carefully transferred to sample vials for UHPLC-MS/MS analysis.

2. Quality control sample

As a part of the system conditioning and quality control process, a pooled quality control sample (QC) was prepared by mixing equal volumes of all samples. The QC samples were disposed and tested in the same manner as the analytic samples. It helped to represent the whole sample set, which would be injected at regular intervals (every 10 samples) in order to monitor the stability of the analysis.

3. UHPLC-MS/MS analysis.

The instrument platform for LC-MS analysis was UHPLC-Q Exactive system of Thermo Fisher Scientific.

Chromatographic conditions:

2 μL of sample was separated by HSS T3 column (100 mm × 2.1 mm i.d., 1.8 μm) and then entered into mass spectrometry detection. The mobile phases consisted of 0.1% formic acid in water:acetonitrile (95:5, v/v) (solvent A) and 0.1% formic acid in acetonitrile:isopropanol:water

(47.5:47.5:5, v/v)(solvent B). The solvent gradient changed according to the following conditions: from 0 to 0.1 min, 0% B to 5% B; from 0.1 to 2 min, 5% B to 25% B; from 2 to 9 min, 25% B to 100% B ; from 9 to 13 min, 100% B to 100% B; from 13 to 13.1 min, 100% B to 0% B; from 13.1 to 16 min, 0% B to 0% B for equilibrating the systems. The sample injection volume was 2 μL and the flow rate was set to 0.4 mL/min. The column temperature was maintained at 40 ℃. During the period of analysis, all these samples were stored at 4 ℃.

MS conditions:

The mass spectrometric data was collected using a Thermo UHPLC-Q Exactive Mass Spectrometer equipped with an electrospray ionization (ESI) source operating in either positive or negative ion mode. The optimal conditions were set as followed: heater temperature, 400 ℃; Capillary temperature, 320 ℃; sheath gas flow rate, 40 arb; Aux gas flow rate, 10 arb; ion-spray voltage floating (ISVF), -2800V in negative mode and 3500V in positive mode, respectively; Normalized collision energy, 20-40-60V rolling for MS/MS. Full MS resolution was 70000, and MS/MS resolution was 17500. Data acquisition was performed with the Data Dependent Acquisition (DDA) mode. The detection was carried out over a mass range of 70-1050 m/z.

4. Data preprocessing and annotation

After the mass spectrometry detection was completed, the raw data of LC/MS was preprocessed by Progenesis QI (Waters Corporation, Milford, USA) software, and a three-dimensional data matrix in CSV format was exported. The information in this three-dimensional matrix included: sample information, metabolite name and mass spectral response intensity. Internal standard peaks, as well as any known false positive peaks (including noise, column bleed, and derivatized reagent peaks), were removed from the data matrix, deredundant and peak pooled.

At the same time, the metabolites were searched and identified, and the main database was the HMDB (http://www.hmdb.ca/), Metlin (https://metlin.scripps.edu/) and Majorbio Database .

The data after the database search was uploaded to the Majorbio cloud platform (https://cloud.majorbio.com) for data analysis. Metabolic features detected at least 80% in any set of samples were retained. After filtering, minimum metabolite values were imputed for specific samples in which the metabolite levels fell below the lower limit of quantitation and each Metabolic features were normalized by sum. In order to reduce the errors caused by sample preparation and instrument instability, the response intensity of the sample mass spectrum peaks was normalized by the sum normalization method, and the normalized data matrix was obtained. At the same time, variables with relative standard deviation (RSD) > 30% of QC samples were removed, and log10 logarithmization was performed to obtain the final data matrix for subsequent analysis.

5. Differential metabolites analysis

Performed variance analysis on the matrix file after data preprocessing. The R package ropls (Version 1.6.2) performed principal component analysis (PCA) and orthogonal least partial squares discriminant analysis (OPLS-DA), and used 7-cycle interactive validation to evaluate the stability of the model. In addition, student's t-test and fold difference analysis were performed. The selection of significantly different metabolites was determined based on the Variable importance in the projeciton (VIP) obtained by the OPLS-DA model and the p-value of student’s T test, and the metabolites with VIP>1, $p<0.05$ were significantly different metabolites. A total of XXX differential metabolites were screened.

Differential metabolites among two groups were summarized , and mapped into their

biochemical pathways through metabolic enrichment and pathway analysis based on database search (KEGG, http://www. genome.jp/kegg/). These metabolites can be classified according to the pathways they involved or the functions they performed. Enrichment analysis was usually to analyze a group of metabolites in a function node whether appears or not. The principle was that the annotation analysis of a single metabolite develops into an annotation analysis of a group of metabolites. scipy.stats (Python packages) (https://docs.scipy.org/doc/scipy/) was exploited to identify statistically significantly enriched pathway using Fisher's exact test.

Text S3 Transcriptomic analysis steps *of T. patula* roots

1. RNA extraction

Total RNA was extracted from the tissue using TRIzol® Reagent. Then RNA quality was determined by 5300 Bioanalyser (Agilent) and quantified using the ND-2000 (NanoDrop Technologies). Only high-quality RNA sample (OD260/280=1.8~2.2, OD260/230≥2.0, RQN≥6.5, 28S:18S≥1.0, >1μg) was used to construct sequencing library.

2. Library preparation and Sequencing

RNA purification, reverse transcription, library construction and sequencing were performed at Shanghai Majorbio Bio-pharm Biotechnology Co., Ltd. (Shanghai, China) according to the manufacturer's instructions. The XX RNA-seq transcriptome librariy was prepared following Illumina® Stranded mRNA Prep, Ligation (San Diego,CA) using 1 μg of total RNA. Shortly, messenger RNA was isolated according to polyA selection method byoligo(dT) beads and then fragmented by fragmentation buffer firstly. Secondly double-stranded cDNA wassynthesized using a SuperScript double-stranded cDNA synthesis kit (Invitrogen, CA) with random

hexamerprimers. Then the synthesized cDNA was subjected to end-repair, phosphorylation and adapter addition accordingto library construction protocol. Libraries were size selected for cDNA target fragments of 300 bp on 2% LowRange Ultra Agarose followed by PCR amplified using Phusion DNA polymerase (NEB) for 15 PCR cycles. Afterquantified by Qubit 4.0, the sequencing library was performed on NovaSeq X Plus platform (PE150) using NovaSeqReagent Kit. (NovaSeq Sequencing platform) OR the sequencing library was performed on DNBSEQ-T7 platform (PE150) using DNBSEQ-T7RS Reagent Kit (FCL PE150) version 3.0. (T7 Sequencing platform)

3. Quality control and De novo Assembly

The raw paired end reads were trimmed and quality controlled by fastp [1] with default parameters. Then cleandata from the samples were used to do de-novo assembly with Trinity[2]. To increase the assembly quality, all theassembled sequences were filtered by CD-HIT[3] and TransRate[4] and assessed with BUSCO (BenchmarkingUniversal Single-Copy Orthologs)[5]. The assembled transcripts were searched against the NCBI protein nonredundant (NR), Clusters of Orthologous Groups of proteins (COG), and Kyoto Encyclopedia of Genes andGenomes (KEGG)[6] databases using Diamond to identify the proteins that had the highest sequence similarity withthe given transcripts to retrieve their function annotations and a typical cut-off E-values less than 1.0×10-5 was set. BLAST2GO[7] program was used to get GO annotations of unique assembled transcripts for describing biologicalprocesses, molecular functions and cellular components.

4. Differential expression analysis and Functional enrichment

To identify DEGs (differential expression genes) between two different samples/groups, the

expression level of each transcript was calculated according to the transcripts per million reads (TPM) method. RSEM[8] was used toquantify gene abundances. Essentially, differential expression analysis was performed using the DESeq2[9] or DEGseq[10]. DEGs with |log2FC|≧1 and FDR<0.05 (DESeq2) or FDR<0.001 (DEGseq) were considered to besignificantly different expressed genes. In addition, functional-enrichment analysis including GO and KEGG wereperformed to identify which DEGs were significantly enriched in GO terms and metabolic pathways at Bonferroni-corrected P-value<0.05 compared with the whole-transcriptome background. GO functionalenrichment and KEGG pathway analysis were carried out by Goatools and Python scipy software, respectively.

Reference

[1] Chen S, Zhou Y, Chen Y, et al. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics, 2018;34:884-890.

[2] Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 2011;29:644-652.

[3] Fu LM, Niu BF, Zhu ZW, et al. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics, 2012;28:3150-3152.

[4] Smith-Unna R, Boursnell C, Patro R, et al. TransRate: reference-free quality assessment of de novo transcriptome assemblies. Genome Research, 2016;26:1134-1144.

[5] Manni M, Berkeley MR, Seppey M, et al. BUSCO Update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Molecular Biology and Evolution, 2021, 38: 4647-4654.

[6] Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids

Research 2000;28:27-30.

[7] Conesa A, Gotz S, Garcia-Gomez JM, et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 2005;21:3674-3676.

[8] Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 2011;12:323.

[9] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 2014;15:550.

[10] Wang L, Feng Z, Wang X, et al. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics, 2009;26:136-138.