Biochemical and Biophysical Research Communications 459 (2015) 60–65

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Transcriptome sequencing and differential gene expression analysis in *Viola yedoensis* Makino (Fam. Violaceae) responsive to cadmium (Cd) pollution

Jian Gao [b, 1], Mao Luo [c, 1], Ye Zhu [a], Ying He [a], Qin Wang [a], Chun Zhang [a, *, 1]

[a] Department of Pharmacy of Luzhou Medical College, Luzhou, Sichuan, China
[b] Key Laboratory of Biology and Genetic Improvement of Maize in Southwest Region, Ministry of Agriculture, Maize Research Institute of Sichuan Agricultural University, Wenjiang, Sichuan, China
[c] Drug Discovery Research Center of Luzhou Medical College, Luzhou, Sichuan, China

ARTICLE INFO

Article history:
Received 10 February 2015
Available online 26 February 2015

Keywords:
Viola yedoensis
Transcriptome assembly
Functional annotation
Differential expression genes (DEGs)

ABSTRACT

Viola yedoensis Makino is an important Chinese traditional medicine plant adapted to cadmium (Cd) pollution regions. Illumina sequencing technology was used to sequence the transcriptome of *V. yedoensis* Makino. We sequenced Cd-treated (VIYCd) and untreated (VIYCK) samples of *V. yedoensis*, and obtained 100,410,834 and 83,587,676 high quality reads, respectively. After de novo assembly and quantitative assessment, 109,800 unigenes were finally generated with an average length of 661 bp. We then obtained functional annotations by aligning unigenes with public protein databases including NR, NT, SwissProt, KEGG and COG. In addition, 892 differentially expressed genes (DEGs) were investigated between the two libraries of untreated (VIYCK) and Cd-treated (VIYCd) plants. Moreover, 15 randomly selected DEGs were further validated with qRT-PCR and the results were highly accordant with the Solexa analysis. This study firstly generated a successful global analysis of the *V. yedoensis* transcriptome and it will provide for further studies on gene expression, genomics, and functional genomics in Violaceae.

1. Introduction

Heavy metals, such as cadmium (Cd), copper (Cu), lead (Pb), chromium and mercury are major environmental pollutants, has caused a serious metal contamination issues, particularly in soils. Heavy metal accumulation in soils is of concern in agricultural production due to the adverse effects on food safety. Cadmium (Cd), as a non-essential element to plants [32], which has a higher tendency to accumulate in plant tissues and is easily taken up by plant, thereby posing health risks to humans and animals by entering the food chain through ingestion of contaminated plants [12]. In the current time, a large number of Cd-tolerant species and several Cd-hyperaccumulating species have been identified and regarded as potentially ideal models with Cd hyperaccumulate or hypertolerate to investigate the physiological mechanisms of Cd tolerance in plants, especially the Cruciferae plant family [2]. In addition, a number of *Viola* metallophytes have been also found in Central Africa and Europe in the Violaceae family, such as *Viola lutea* ssp. *westfalica* and *V. lutea* ssp. *calaminaria* only occur on contaminated soil [10,29].

Viola yedoensis Makino, a new species of Violaceae in China, distributed in a Cadmium (Cd)/zinc (Zn) mine in Sichuan Province, and it was identified as a newly Cd hyperaccumulator plant. In the present study, we utilized Illumina paired-end sequencing technology to characterize the transcriptome of *V. yedoensis* through the analysis of large-scale transcript sequences. It would offer valuable sequence resource to Violaceae community and explore some unknown Cd-defensive strategies that possibly exist in the Viola species. It would also provide an efficient, less cost and reliable approach that can be readily adopted by researchers studying non-model organisms using transcriptome sequencing.

2. Materials and methods

2.1. Seed sterilization and experiment design

The seeds of *V. yedoensis* Makino (Fam. Violaceae) were sown on filter papers saturated with distilled water and incubated at 26 °C in

* Corresponding author. Department of Pharmacy of Luzhou Medical College, Luzhou, Sichuan 646000, China.
E-mail address: zc83good@126.com (C. Zhang).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.bbrc.2015.02.066
0006-291X/

the dark. About three days later, seedlings selected for almost uniform growth were transplanted into an aerated complete nutrient solution (see Table S1 in File S1) and kept for 3 days in a growth chamber with a photoperiod of 14 h light/10 h dark at 26 °C and a relative humidity of 70%. After about three months later, the almost uniform growth plants were randomly divided into two groups, CK−grown (CK) seedlings, grown only in half-strength Hoagland solution, Cd−grown (Cd) in CK + Cd $(NO_3)_2$ (2.5 mM) causing Cd heavy metal stress. The pH of all solutions was adjusted to 6.0. According to *V. yedoensis* phenotypes under Cd exposure, we found that the leaves turned yellow and the leaf tips of the seedlings began to wilt after 72 h and it is probable that 72 h was the maximum number of threshold for seedlings living (Fig. S1). Both Cd−grown (Treat) and CK−grown (CK) whole plant were separately harvested at 72 h treatments and selected as the samples for sequencing. All samples were harvested from each of the three *V. yedoensis* plants, and three independent replicates were collected for each sample.

2.2. V. yedoensis whole plant collection and RNA isolation

Two RNA libraries were constructed using RNA extracted and pooled from the CK- grown and Cd-grown *V. yedoensis* whole plant at 72 h. According to the manufacturer's instructions, total RNA samples were prepared as follows in two replicates: equal quantities of RNA isolated from whole plant were pooled for stress stage using Trizol Reagent (Invitrogene, Nottingham, UK). DNase I (Promega) was used to degrade contaminating genomic DNA. The total RNA was quantified by its absorbance at $\lambda = 260$ nm (A260nm) on a Nanodrop spectrophotometer (Thermo Fisher Scientific Inc., MA, USA) and the purity was estimated from its A260nm/A280nm ratio. Furthermore, RNA integrity was assessed on an Agilent Bioanalyzer 2100 system using the Agilent RNA Pico Chip Kit (Agilent Technologies, Inc., Santa Clara, California, USA). The purified RNA was dissolved in RNase-free water and stored in −80 °C freezer until subsequent analysis.

2.3. cDNA preparation and sequencing

Based on the protocol of the Paired-End sample Preparation kit (Illumina, USA), the paired-end library was developed to construct a cDNA library for sequencing of the *V. yedoensis* transcriptome. Two libraries named VIYCK (CK−grown seedlings) and VIYCd (Cd−grown seedlings) were constructed respectively. In brief, we used oligo (dT) magnetic beads (NEB) to get the purified poly-A mRNA, which was then fragmented into fragments 200-500 bp in size. Then we synthesized first-strand cDNA with hexamer primers and reverse transcriptase (Promega) and the second-strand cDNA with DNA polymerase I and RNase H from the fragmented mRNA pieces. After that, the cDNA fragments were ligated to index adapters (Illumina) after purified, end-repaired, A-tailed. The ligated products were subsequently to generate the final cDNA libraries by PCR-amplified. Illumina GA IIX sequencing platform were used to sequence the cDNA library, with the average 75 nt length of sequenced reads.

2.4. Data filtering and de novo assembly

Firstly, we removed reads that do not pass the built-in Illumina's software Failed-Chastity filter, according to the relation "failed-chastity $<= 1$ using a chastity threshold of 0.6 for the first 25 cycles. Secondly, all reads with adaptor contamination were discarded and low-quality reads with more than 5% ambiguous sequences "N" were ruled out. At last, the reads with more than 20% Q < 20 bases were also removed. We used Trinity program with the default settings except K-mer value to conduct the de novo assembly [8]. Additionally, if there are mutli-duplication's reads, only one read copy will be retained for assembly and redundant duplication reads be eliminated. Trinity assembled some reads that had overlapped nucleic acid sequence and generated contigs. To obtain the unigene, the paired-end reads were used for constructing scaffolds with the paired end information by realigning to contigs. Then, these contigs in one transcript were assembled by the Trinity and gained the sequence not being extended on either end defined as unigenes. At last, all usable reads were realigned to the unigenes using SOAPaligner to evaluate the depth of coverage [17] with the default settings.

2.5. Function annotation and classification of assembled transcriptomes

Various bioinformatics approaches were used for further annotation of unigenes. Firstly, sequence-based alignments were performed against four public databases, including the NCBI non-redundant protein (NR) database, RefSeq protein (NT) at NCBI, UniProt (SwissProt and TrEMBL), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) using BLASTX algorithm with a significant E-value threshold of 1e−5. According to the Kyoto Encyclopedia of Genes and Genomes pathway database, pathway assignments were carried out using BLASTx with E value with the threshold of 10-5 [13]. After that, domain/family searches contained Hidden Markov Model (HMM) domain/family searches in Pfam databases and BLASTX alignments against the EuKaryotic Orthologous Groups (KOG) database and Gene ontology (GO) at NCBI. The E-value threshold was also set at ≤1e−5. Blast2GO [4] program was used to get GO annotation according to molecular function, biological process and cellular component ontologies. The unigene sequences were also aligned to the COG database to predict and classify possible functions. In addition, the WEGO software was then used to perform GO functional classification of all unigenes [4]. This analysis mapped all of the annotated unigenes to GO terms in the database and calculated the number of unigenes associated with every GO term [1].

2.6. Transcriptome quantification

Transcript quantification of the de novo assembly was carried out with RSEM to assess the transcript abundances based on the mapping of RNA-seq reads to the assembled transcriptome in this study [14]. Briefly, a newly modified Burrows-Wheeler transform (BWT) aligner, RSEM (Bowtie2 v 2.0.1), was applied for aligning the reads back to a reference genome or transcriptome and alignment results were converted to BAM format using SAMtools [15]. Normalizing and quantifying gene expression levels from ambiguous alignment results are statistical challenges when performing high-throughput RNA sequencing. The developed software RSEM was used to accurately quantify the abundance of transcript-level sequences and to calculate the FPKM (fragments per kilobase of transcript per million mapped reads). Only transcripts with an FPKM ≥1 were considered to be significant expressed [30].

2.7. Differential gene expression analysis

Differentially expressed genes were calculated by edgeR package v3.0.8 [25]. In addition, FDR ≤0.001 and the absolute value of |log2Ratio| ≥ 1 were set as the cutoff threshold to determine their significant expression. GO enrichment analysis of differentially expressed genes was performed using Blast2GO software, with a P-value cutoff value of 0.05 against the annotated unigenes during the Fisher's exact test. Significantly regulated genes were also clustered

using Cluster v3.0 software [7] by performing Hierarchical clustering. Gene expression values were extracted from the edgeR-normalized FPKM data sets. Moreover, Euclidean distance and complete-linkage methods calculate Matrix distance were used for expression heatmap with original FPKM values log-transformed and centered, and then constructed using TreeView v1.1.6 [27] and MeV v4.8.1 [26].

2.8. Real-time quantitative (qRT-PCR) validation

The isolated RNA sequencing samples mentioned above were also used to perform real-time quantitative (qRT-PCR) analysis. First-strand cDNA was synthesized using M-MLV Reverse Transcriptase (Promega, U.S.). After that, we used Primer Premier 5.0 (Premier Biosoft International, Palo Alto, CA, U.S.) to design the primers, which is showed in (Table S2 in File S1). 18sRNA (FJ669717) was used as an internal control. According to the standard protocol, the qRT-PCR was carried out using the ABI7500 system. The amplification for PCR were performed as follows: 40 cycles including 95 °C for 30s, 60 °C for 30s and 72 °C for 30 s, and then generated the melt curves for verification of amplification specificity by a thermal denaturing step. We set all reactions in triplicate, as well as non-template controls. The relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method [19].

3. Results and discussion

3.1. Illumina sequencing and de novo assembly

To examine the coverage of *V. yedoensis* mRNA by RNA sequencing, RNA libraries were constructed by pooling RNA isolated from sample of whole plants collected from the two groups named VIYCK (CK—grown seedlings) and VIYCd (Cd—grown seedlings) respectively, with two biology replicates after 72 h respectively. After filtered the only adaptor sequences, containing N sequences, as well as low quality sequences. VIYCK and VIVIYCd RNA-Seq libraries still generated over 50 million and 41 million clean reads for each replicate in each library respectively.

Due to the absence of reference genomic sequences, we used de novo assembly to construct transcripts from these RNA-seq reads performed using Trinity [8]. The quality trimmed reads ($Q \geq 20$) were then de novo assembled into transcripts using Trinity, with a fixed k-mer of 25. We applied the “Reduce” option within the modified version of the Trinity software package to reduce redundancy in assembled transcriptomes. In total, 244,906 contigs were constructed from the raw sequence reads. The sequence length of these assembled contigs were at an average length of 1,112 bp (Fig. 1A). Using paired-end reads, contigs from the same transcript were connected and extended in both ends to get the longest unigenes possible. A fairly large number (10,0027 of 24,4906) of assembled unigenes were between 200 bp and 500 bp in length, indicating the presence of assembled fragment. After that, 109,800 unigenes were generated, with an average unigene length of 660 bp (Fig. 1B). The result is longer than those assembled in previous studies for example, *Eucalyptus grandis* (247 bp) [23], *Coral larval* (440 bp) [21], *Lodegpole pine* (500 bp) [24]. As we all known, Illumina transcriptome or whole genome de novo sequencing and assembly have been successfully used for model [9,18,31,33], and non-model organisms [33], especially with the great advantage of paired-end sequencing [20]. Consistent with these publications, our results also indicated that relatively short reads from Illumina paired-end sequencing can be effectively assembled and used for unique gene discovery in non-model organism.

3.2. Functional annotation and classification

All assembled sequences were first aligned to protein sequences from a custom-made NCBI non-redundant database (NR) filtered, which returned 52,266 significant BLAST hits (47.6%, see Table 1). The mapping rates of the *V. yedoensis* contigs against the NP RefSeq and UniProt databases were 24,612 (22.41%) and 42,714 (38.9%), respectively (Table 1). To further exploit the potential function of conserved domains about the transcriptome sequences in *V. yedoensis*, their conserved domains were identified against the Pfam, GO and KOG databases. Considering of proteins has little similarity at sequence level with conserved structural domains, we identified their similarity at domain level against these databases. Pfam analysis showed 43,098 contigs were annotated (Table 1) contained different kinds of Pfam domains, among which Zinc finger C2H2 type, WD domain G-beta repeat and Cytochrome P450 were highly represented (629, 346 and 214 times, respectively).

The annotated unigenes were then assigned to Gene Othology (GO) terms for functional classification. Three main categories of GO classification were analyzed separately to investigate their functional distribution, such as biological process, molecular function and cellular component. We assigned the annotated sequences to GO- slim terms to simplify the functional distribution of plants to obtain a “thin” version of classification [3]. Cellular process (GO:0009987) and metabolic process (GO:0008152) within biological process, binding activity (GO:0005488) and catalytic

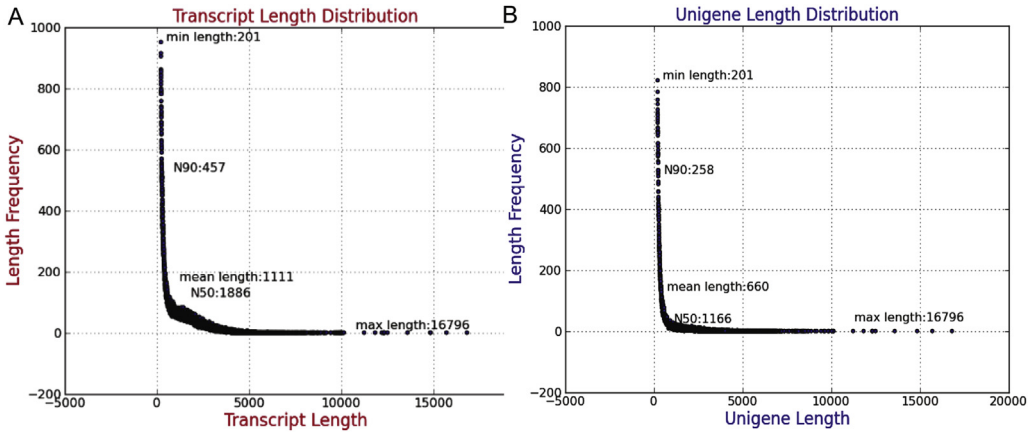

Fig. 1. Sequence length distributions of assembled transcripts (A) and unigenes (B). from *V. yedoensis* transcriptome. Transcripts and unigenes were assembled from raw sequence data after reads with adapters, reads with >5% unknown nucleotides, and low quality reads were removed.

Table 1
BLASTx and BLASTn annotation against various databases.

	Number of unigenes	Percentage (%)
Annotated in NR	52266	47.6
Annotated in NT	24612	22.41
Annotated in KEGG	24199	22.03
Annotated in SwissProt	42714	38.9
Annotated in PFAM	43098	39.25
Annotated in GO	54479	49.61
Annotated in KOG	27907	25.41
Annotated in all databases	9301	8.47
Annotated in at least one database	61591	56.09
Total unigenes	109800	100

activity (GO:0003824) within molecular function and cells (GO:0005623) and organelles (GO:0043226) within cellular component were the most representative level 2 GO terms in all three data sets (Table S3 in File S1).

To further elucidate the functionality of the *V. yedoensis* transcriptome, the annotated unigenes were categorized into different functional groups based on the COG database (Cluster of Orthologus Groups) (Fig. 2). Out of 52,226 unigenes, 31,285 could be classified into 26 COG categories. Out of 31,285 unigenes (Table S4 in File S1), 4021 (12.85%) were assigned into the COG category of general function prediction, which represented the largest functional group of the 26 COG categories, followed by post-translational modification, protein turnover, chaperon (3,520, 11.25%), translation (3,142, 10.04%), signal transduction (3,043, 9.73%), intracellular trafficing and secretion (1,709, 5.46%), and transcription (1,698, 5.17%).

To further identify the active biochemical pathways of *V. yedoensis*, we mapped the *V. yedoensis* unigenes to the reference canonical pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG). KEGG is thought to provide a basic platform for systematic analysis of gene function in terms of the network of gene products. A total of 24,199 unigenes were annotated based on a BLASTX search of the KEGG database (Table S5 in File S1): 263 biosynthesis pathways were predicted and classified into five categories. Of which, the metabolic pathway was the largest, containing 9,884 members (Unigene products) (Fig. 3).

3.3. Differentially expressed genes involved in the response to Cd stress in V. yedoensis Makino

To better investigate the critical genes with alteration in *V. yedoensis* Makino responding to heavy metal Cd stress, we used

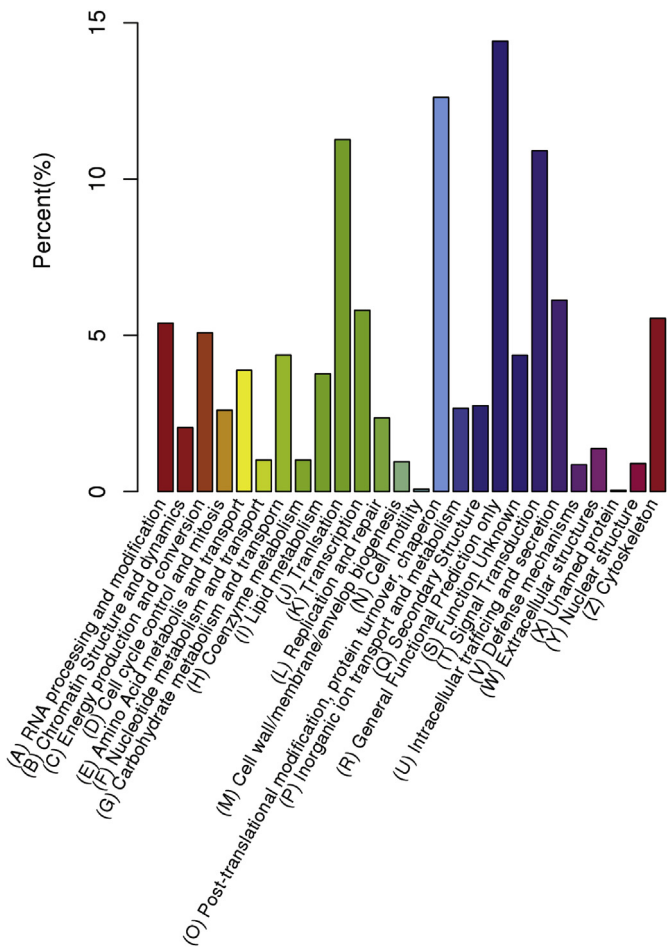

Fig. 2. Histogram presentation of COG classification. In total, 52,226 unigenes showing significant homology to the COGs database at NCBI (E-value 1.0e-5) have a COG classification among the 26 categories. The capital letters in x-axis indicate the COG categories as listed on the right and the y-axis indicates the number of unigenes in each category.

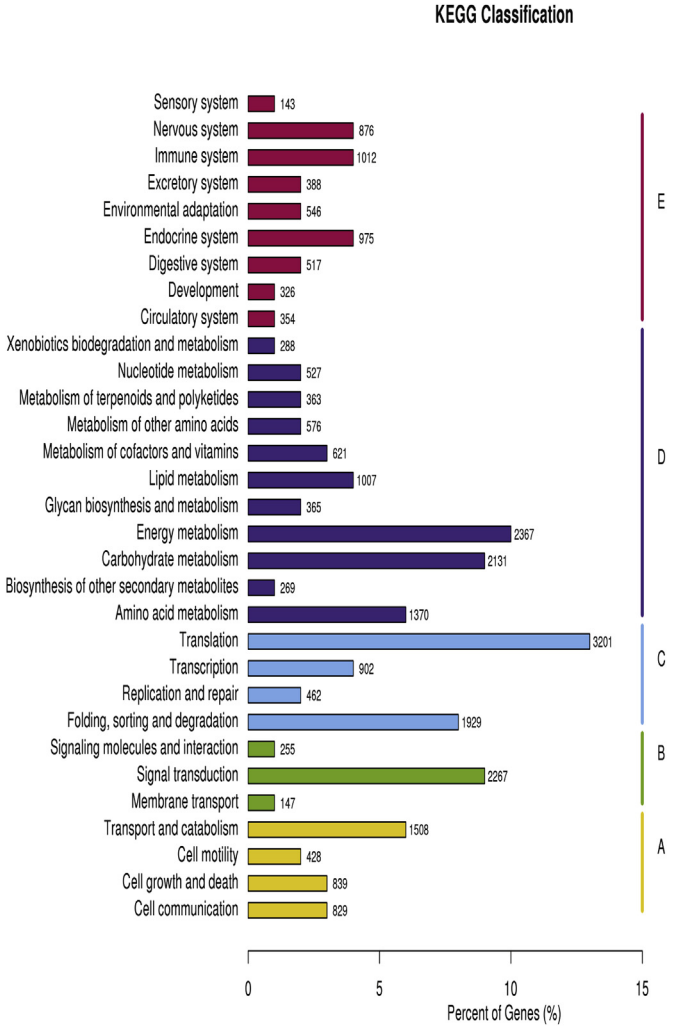

Fig. 3. Functional classification and pathway assignment of unigenes by KEGG. Non-redundant unigenes for *V. yedoensis* transcriptome were assigned X pathways within X clades under five major categories: I, Metabolism; II, Genetic information processing; III, Environmental information processing; IV, Cellular processes; V, Organismal systems, including 31 sub-categories. The y-axis indicates the name of the KEGG metabolic pathways. The x-axis indicates the percentage of the number of genes annotation under the pathway in the total number of genes annotation.

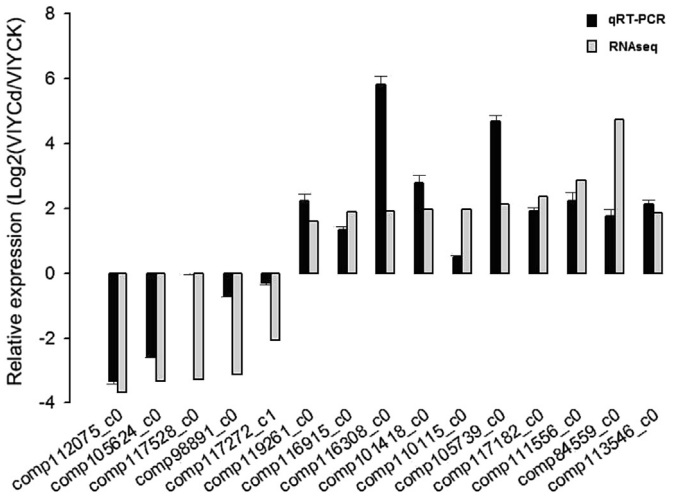

Fig. 4. The differential expression genes levels randomly selected from differential expressed genes with high abundance enriched in VIYCd and VIYCK libraries respectively and validation with qRT-PCR. The relative gene expression levels as expressed by $2^{-\Delta\Delta Ct}$ were determined separately for each treatment as the mean ± SE. The validation of selected differential expressed known genes enriched in three libraries indicated that the results from DGE sequencing were generally agreed well with the qRT-PCR results.

Bowtie2 software to map the clean reads of the VIYCd and the VIYCK libraries to the de novo assembly transcriptome reference sequences respectively and qualify transcriptome by assigning to unigenes and isoforms with the RSEM (RNA-Seq by Expectation Maximization) software [14]. We used the FPKM (fragments mapped per kilobase of exon per million reads mapped) cutoff of 1 to calculate the assigned unigene and isoform expression levels. A log-fold expression change (log2 ratio) greater than 2 or less than −2 was used to determine the differentially expressed genes (DEGs) (including unigene and isoform), accompanying with a greater statistically significant value ($P \leq 0.005$), as well as false discovery rates (FDR ≤ 0.001). In total, 892 DEGs were detected between the two libraries, including both up-regulated (305 transcripts) and down-regulated genes (587 transcripts) under the Cd treatment (Fig. S2, Table S6 in File S1).

3.4. qRT-PCR validation of the candidate DEGs responsive to Cd stress

To validate the reliability of the Solexa analysis, RNA samples isolated for RNA sequencing were used to perform qRT-PCR analysis. In this study, 15 candidate DEGs at different level of expression were randomly selected and detected by qRT- PCR. It is showed that the expression changes of these candidate DEGs in VIYCK and VIYCd libraries are similar to the results of Solexa sequencing (Fig. 4). The discrepancies with respect to ratio should be attributed to the essentially different algorithms and sensitivity between the two techniques [16,28].

3.5. Clustering and functional enrichment of DEGs

To further illustrate the relationships between DEGs with various expression patterns, the Euclidean distance method was used to perform the hierarchical clustering of the DEGs identified in our study (Fig. S3A). According to the cluster results, 19 clusters were identified and plotted with expression patterns in SOM (Self Organizing Map), resulted in five main clusters (Fig. S3B). Of those, the sub-cluster1 possessed the most genes (328), followed by sub-cluster5 (206), sub-cluster4 (150), sub-cluster2 (124) and sub-cluster3 (84). GO-enrichment was performed against all of the annotated unigenes of the combined assembly. The over-represented GO-slim terms of DEGs are shown in Fig. S4. Many of the DEGs were involved in response to stress (GO:0006950) and stimulus (GO:0050896), as well as response to abiotic stress (GO:0009628), especially for cell wall (GO:0005618) and apoplast (GO:0048046). In addition, some of DEGs acted as oxidoreductase activity (GO:0016491) and transferase activity (GO:0016757, GO:0016758) (Table S7 in File S1). The plant glutathione transferases, as an important transferases identified mostly with the great alteration in our study, which acted in herbicide detoxification, hormone homeostasis, vacuolar sequestration of anthocyanin, tyrosine metabolism, hydroxyperoxide detoxification, regulation of apoptosis, especially in plant responses to biotic and abiotic stresses [6]. Previous study showed that GST activity was found increased in leaves and roots of Cd-exposed *Pisum sativum* plants [5], in roots of *Phragmites australis* [11] and *Oryza sativa* [22]. This information could gain more direct biological insights to elucidate the molecular mechanism of responsive to Cd stress in *V. yedoensis*. The results were highly consistent with the molecular mechanism of heavy metal stresses in plants [34].

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

This work was supported by the grants from the National Natural Science Foundation of China (Grant No. 81001700, No. 21072088), and Foundation of Luzhou Municipal Science and Technology Bureau of Sichuan Province (Grant No. 2013-S-47 (10/20), 2013 LZLY-K64 and 14JC0137).

Transparency document

Transparency document related to this article can be found online at http://dx.doi.org/10.1016/j.bbrc.2015.02.066.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.bbrc.2015.02.066.

References

[1] E. Altermann, T.R. Klaenhammer, PathwayVoyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, BMC Genomics 6 (2005) 60.
[2] A.G. Assunção, H. Schat, M.G. Aarts, Thlaspi caerulescens, an attractive model species to study heavy metal hyperaccumulation in plants, New. Phytol. 159 (2003) 351–360.
[3] E. Camon, M. Magrane, D. Barrell, V. Lee, E. Dimmer, J. Maslen, D. Binns, N. Harte, R. Lopez, R. Apweiler, The Gene ontology annotation (GOA) database: sharing knowledge in Uniprot with Gene Ontology, Nucleic Acids Res. 32 (2004) D262–D266.
[4] A. Conesa, S. Götz, J.M. García-Gómez, J. Terol, M. Talón, M. Robles, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, Bioinformatics 21 (2005) 3674–3676.
[5] V. Dixit, V. Pandey, R. Shyam, Differential antioxidative responses to cadmium in roots and leaves of pea (Pisum sativum L. cv. Azad), J. Exp. Bot. 52 (2001) 1101–1109.
[6] D.P. Dixon, M. Skipsey, R. Edwards, Roles for glutathione transferases in plant secondary metabolism, Phytochemistry 71 (2010) 338–350.
[7] M.B. Eisen, P.T. Spellman, P.O. Brown, D. Botstein, Cluster analysis and display of genome-wide expression patterns, Proc. Natl. Acad. Sci. 95 (1998) 14863–14868.
[8] B.J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P.D. Blood, J. Bowden, M.B. Couger, D. Eccles, B. Li, M. Lieber, De novo transcript sequence

reconstruction from RNA-seq using the Trinity platform for reference generation and analysis, Nat. Protoc. 8 (2013) 1494–1512.
[9] Z. Hegedűs, A. Zakrzewska, V.C. Ágoston, A. Ordas, P. Rácz, M. Mink, H.P. Spaink, A.H. Meijer, Deep sequencing of the zebrafish transcriptome response to mycobacterium infection, Mol. Immunol. 46 (2009) 2918–2930.
[10] U. Hildebrandt, K. Hoef-Emden, S. Backhausen, H. Bothe, M. Bożek, A. Siuta, E. Kuta, The rare, endemic zinc violets of Central Europe originate from Viola lutea Huds, Plant Syst. Evol. 257 (2006) 205–222.
[11] M.A. Iannelli, F. Pietrini, L. Fiore, L. Petrilli, A. Massacci, Antioxidant response to cadmium in *Phragmites australis* plants, Plant Physiology Biochem. 40 (2002) 977–982.
[12] A. Kabata-Pendias, Trace Elements in Soils and Plants, CRC press, 2000.
[13] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (2000) 27–30.
[14] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, BMC Bioinforma. 12 (2011) 323.
[15] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The sequence alignment/map format and SAMtools, Bioinformatics 25 (2009) 2078–2079.
[16] P. Li, L. Ponnala, N. Gandotra, L. Wang, Y. Si, S.L. Tausta, T.H. Kebrom, N. Provart, R. Patel, C.R. Myers, The developmental dynamics of the maize leaf transcriptome, Nat. Genet. 42 (2010a) 1060–1067.
[17] R. Li, Y. Li, K. Kristiansen, J. Wang, SOAP: short oligonucleotide alignment program, Bioinformatics 24 (2008) 713–714.
[18] R. Li, H. Zhu, J. Ruan, W. Qian, X. Fang, Z. Shi, Y. Li, S. Li, G. Shan, K. Kristiansen, De novo assembly of human genomes with massively parallel short read sequencing, Genome Res. 20 (2010b) 265–272.
[19] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method, Methods 25 (2001) 402–408.
[20] C.A. Maher, N. Palanisamy, J.C. Brenner, X. Cao, S. Kalyana-Sundaram, S. Luo, I. Khrebtukova, T.R. Barrette, C. Grasso, J. Yu, Chimeric transcript discovery by paired-end transcriptome sequencing, Proc. Natl. Acad. Sci. 106 (2009) 12353–12358.
[21] E. Meyer, G. Aglyamova, S. Wang, J. Buchanan-Carter, D. Abrego, J. Colbourne, B. Willis, M. Matz, Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx, BMC Genomics 10 (2009) 219.
[22] A. Moons, *Osgstu3* and *osgtu4*, encoding tau class glutathione *S*-transferases, are heavy metal-and hypoxic stress-induced and differentially salt stress-responsive in rice roots, Febs Lett. 553 (2003) 427–432.
[23] E. Novaes, D.R. Drost, W.G. Farmerie, G.J. Pappas, D. Grattapaglia, R.R. Sederoff, M. Kirst, High-throughput gene and SNP discovery in Eucalyptus grandis, an uncharacterized genome, BMC Genomics 9 (2008) 312.
[24] T. Parchman, K. Geist, J. Grahnen, C. Benkman, C.A. Buerkle, Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery, BMC Genomics 11 (2010) 180.
[25] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (2010) 139–140.
[26] A. Saeed, V. Sharov, J. White, J. Li, W. Liang, N. Bhagabati, J. Braisted, M. Klapa, T. Currier, M. Thiagarajan, TM4: a free, open-source system for microarray data management and analysis, Biotechniques 34 (2003) 374.
[27] A.J. Saldanha, Java Treeview—extensible visualization of microarray data, Bioinformatics 20 (2004) 3246–3248.
[28] T. Shi, Z. Gao, L. Wang, Z. Zhang, W. Zhuang, H. Sun, W. Zhong, Identification of differentially-expressed genes associated with pistil abortion in Japanese apricot by genome-wide transcriptional analysis, PloS One 7 (2012) e47810.
[29] A. Słomka, M. Libik-Konieczny, E. Kuta, Z. Miszalski, Metalliferous and non-metalliferous populations of *Viola tricolor* represent similar mode of anti-oxidative response, J. Plant Physiology 165 (2008) 1610–1619.
[30] C. Trapnell, B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. van Baren, S.L. Salzberg, B.J. Wold, L. Pachter, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, Nat. Biotechnol. 28 (2010) 511–515.
[31] M. Trick, Y. Long, J. Meng, I. Bancroft, Single nucleotide polymorphism (SNP) discovery in the polyploid Brassica napus using Solexa transcriptome sequencing, Plant Biotechnol. J. 7 (2009) 334–346.
[32] M. Van der Perk, Soil and Water Contamination: from Molecular to Catchment Scale, Taylor & Francis, 2007.
[33] B. Wang, G. Guo, C. Wang, Y. Lin, X. Wang, M. Zhao, Y. Guo, M. He, Y. Zhang, L. Pan, Survey of the transcriptome of *Aspergillus oryzae* via massively parallel mRNA sequencing, Nucleic Acids Res. 38 (2010) 5075–5087.
[34] Q.-Y. Zhao, Y. Wang, Y.-M. Kong, D. Luo, X. Li, P. Hao, Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study, BMC Bioinforma. 12 (2011) S2.