ATLANTIS PRESS

Advances in Engineering Research, volume 63

5th International Conference on Sustainable Energy and Environment Engineering (ICSEEE 2016)

Deyeuxia angustifolia's trancriptional response to elevated CO_2 concentration in Sanjiang plain, Northeast China

Mingyi Xu[1,a], Hongjie Cao[1,b], Xin Sui[1,c], Yu Zhang[1,d], Liyuan Wang[1,e], Hainan Leng[1,f], Haixiu Zhong[1,g] and Hongwei Ni [1,h*]

[1] National Wetland and Ecological Conservation Engineering Lab, Institute of Natural resources and Ecology, Heilongjiang Academy of Sciences, Harbin, 150040, China

[a]xumingyi1982@163.com, [b]hjcao781228@163.com,[c]xinsui_cool@126.com, [d]lengning1029@126.com, [e]wangliyuan0503@163.com, [f]lenghn1983@163.com, [g]615729478@qq.com,[e]nihongwei2000@163.com

*Corresponding authors

Keywords: *Deyeuxia angustifolia*, Transcriptome, elevated CO_2 concentration, photosynthesis, carbon fixation

Abstract. *Deyeuxia angustifolia* is the dominant species in the natural grassland of Sanjiang plain of Heilongjiang province of China. In this study, two sequencing libraries prepared from control (T1) and samples treaded by elevated CO_2 (T2) were sequenced to investigate the changes of *D. angustifolia* transcriptome responsed to high CO_2 concentration. In total, 80,215 unigenes were found and 49,094 (61.20%) were annotated with gene descriptions, conserved domains, gene ontology terms and metabolic pathways with a cut-off E-value of 10^{-5}. These annotated unigenes included 61 GO terms, 193 KEGG pathways, and 25 COG families. By comparing two transcriptomes from control and elevated CO_2-treated plants, 4,109 differentially expressed genes were found. These genes were significantly enriched in 105 KEGG pathways and 58 GO terms. Meanwhile, 62 gene fragments related to photosynthesis and carbon fixation process were indentified and analyzed. This study will provide the theoretical basis for elucidating the mechanism of wetland plants' adaption to climate change

Introduction

Since the industrial revolution, the mass intensification of human activities has led to an increasing concentration of CO_2 in the atmosphere that has triggered global climate change [1]. The concentration of CO_2 in the atmosphere is projected to reach 700 ppm by the end of the 21st century [2]. The effect of elevated CO_2 concentration on important metabolic processes such as photosynthesis has been well documented in a wide range of plants [3, 4]. Wetlands play an important role in the land and global carbon cycle and have direct close contact with global climate change and display high sensitivity [5, 6]. Therefore, adaptation mechanism research of wetland vegetation responding to elevated CO_2 is very necessary. *D. angustifolia* (*Deyeuxia* genus, family gramineae) is the dominant species in the Sanjiang Plain. The response of elevated CO_2 of *D. angustifolia* plays an important role in cabon cycle of vegetation system in wetland of the Sanjiang. Therefore, the response and adaptation mechanisms of *D. angustifolia* to elevated CO_2 concentration are key links in revealing the carbon balance and carbon cycling of the Sanjiang Plain wetland vegetation ecosystem under global changes, and these mechanisms have important guiding significance in terms of production.

Our previous research suggested that short-term exposure to high CO_2 concentration increased the rate of CO_2 assimilation by *D. angustifolia* photosynthesis [7]. However, the response molecular mechanism of *D. angustifolia* remains unclear. Although some research has been conducted on the ecophysiological aspects of *D. angustifolia* [8, 9, 10], comprehensive analyses of molecular genetic information are lacking. We compared the transcriptomes of CO_2 elevated and control plants to identify genes displaying transcript changes and identified the functions of the transcripts and the KEGG pathways that displayed changes. We speculate that the assembled, annotated transcriptome

sequences and transcript abundance patterns will provide a valuable genetic resource for further investigations of the molecular mechanisms of photosynthetic acclimation in this species and possibly in other plants. This study will provide the theoretical basis for elucidating the mechanism of plant adaption to climate change.

Materials and methods

Plant materials and experimental treatment

D.angustifolia were collected from Field Experimental Base of Institute of Natural Resources and Ecology, Heilongjiang Academy of Sciences—Honghe National Nature Reserve (47◦49'N, 133◦40'E) in Northeast China. Two kinds of treatments were designed to test the effects of elevated CO_2 on the photosynthesis of leaves as follows: (1) 370±10 μ mol mol^{-1} (ambient CO_2, T1); (2) 700±10 μ mol mol^{-1} (elevated CO_2, T2). Leaves were collected after the seedlings were treated with a 700 μmol mol^{-1} CO_2 concentration for 24 h.

RNA extraction, cDNA library construction and RNA sequencing

Total RNA were isolated using TRIzol reagent (Invitrogen, Carlsbad, CA, USA), and the RNA samples were treated with DNase I (TaKaRa, Japan) for 4 h. The integrity of the RNA samples was examined with an Agilent 2100 Bioanalyzer. High-quality RNA samples were sent to the Biomarker Technology Company (Beijing, China) for cDNA library construction and sequencing using an *Illumina* HiSeq™ 2000 (Illumina, San Diego, CA, USA). The cDNA library construction method and *Illumina* deep-sequencing processes were the same as described by Xu et al [11].

De novo Assembly of Sequences

Reads from each library were assembled separately. The raw reads were cleaned by removing adapter sequences, low-quality sequences, and reads with more than 10% Q-value <20 bases. The Trinity method was used for *de novo* assembly of *Illumina* reads [12].

Functional annotation

We annotated unigenes based on a set of sequential BLAST searches designed to find the most descriptive annotation for each sequence. The assembled unigenes were compared with sequences in National Center for Biotechnology Information (NCBI) non-redundant (Nr) protein and nucleotide (Nt) databases, the Swiss-Prot protein database, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database, the Cluster of Orthologous Groups (COG) database, the Translated EMBL Nucleotide Sequence Database (TrEMBL) and InterPro database. The Blast2GO program was used to obtain GO annotation of the unigenes. The WEGO software was then used to perform GO functional classification of all unigenes to view the distribution of gene functions.

Results

De novo assembly data and basic bioinformatics analysis

Two sequencing libraries were constructed from T1 and T2 leaf tissue samples to investigate the transcriptomic responses to high CO_2 concentration in *D. Angustifolia* (Table 1). Using the Trinity *de novo* assembly program, these cleaned reads were assembled into 6,458,005 (T1) and 2,643,439 (T2) contigs, and 91,206 (T1) and 77,358 (T2) scaffolds were generated. 54,605 (T1) and 46,354 (T2) unigenes were clustered after removal of redundancy. Using the BLAST-Like Alignment Tool assembler, we combined T1 and T2 clean reads and obtained 80,215 all-unigenes, with an average length of 545 bp and a N50 of 734 bp (Table 2).

Table 1 Summary of *Illumina* transcriptome sequencing for T1 and treated T2 leaf tissue samples

	Total number of reads	Total base pairs [bp]	Average read length [bp]	GC [%]	Q20 [%]
T1	20,197,874	4,077,392,500	202	51.06	93.91
T2	25,803,418	5,211,673,964	202	54.34	85.46

Using the BLAST-Like Alignment Tool assembler, we combined T1 and T2 clean reads and obtained 80,215 all-unigenes with an average length of 545 bp and N50 of 734 bp. The length distribution of the unigenes from both T1 and T2 was similar, unigenes with 200-300 bp range maked

up 39.96% and 36.84% of T1 and T2 unigenes, respectively. More than 25,000 unigenes were with length of > 500 bp in assembled sequences (Fig. 1), and of these all-unigenes, the length of > 1 kb were 10, 433.

Table 2 Summary of Illumina transcriptome assembly for T1 and treated T2 leaf tissue samples

	T1 [number]	Mean length [bp]	N50 [bp]	T2 [number]	Mean length [bp]	N50 [bp]
Contigs	6,458,005	40	30	2,643,439	66	78
Scaffolds	91,206	687	1,038	77,358	610	825
Unigenes	54,605	573	817	46,354	535	712
All-unigene	Number:80,215		Mean length [bp]:546		N50 [bp]:734	

Functional annotation of assembled unigenes

The unigenes were annotated in Nr, UniProt/Swiss-Prot, KEGG, COG and UniProt/TrEMBL. Of the 80,215 all-unigenes, 48,856 (60.9%) had significant matches in the Nr databases, and 32,199 (40.1%) unigenes had similarity to proteins in the Swiss-Prot database. Based on sequence homology, 39,747 unigenes were classified into 61 GO terms and distributed into three categories, with47.10% in biological processes, 38.35% in cellular components, and 14.55% in molecular functions (Fig. 2). Interestingly, 12 and 9 unigenes were classified into the categories ‘carbon fixation’ and ‘carbon utilization’, respectively. These results indicated that most of the sequenced genes were responsible for fundamental biological regulation and metabolism.

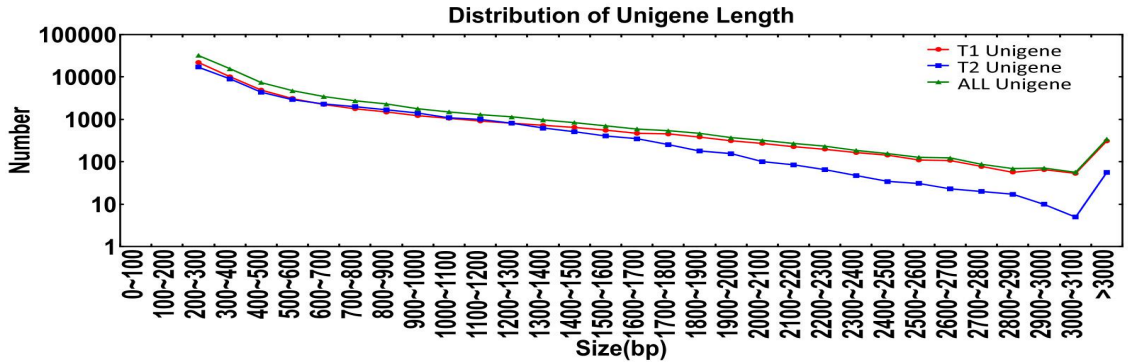

Fig. 1 Length distributions of T1-, T2- and all-unigenes

Based on the Nr annotation, only 12,484 of 80,215 (15.6%) all-unigenes could be aligned to the COG database for predicting and classifying possible functions. COG-annotated putative proteins, involved in cellular structure, biochemistry metabolism, molecular processing, signal transduction and so on, were functionally classified into at least 25 protein families. (Fig. 3). The unigenes involved in carbon utilization and carbon fixation were found in the categories of ‘inorganic ion transport and metabolism’ and ‘energy production and conversion’, respectively.

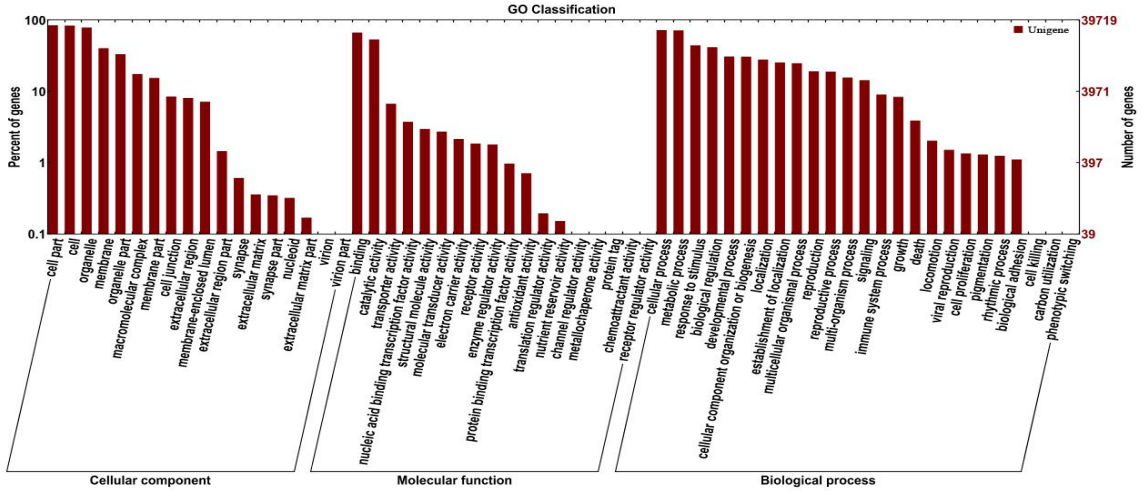

Fig. 2 GO functional annotation

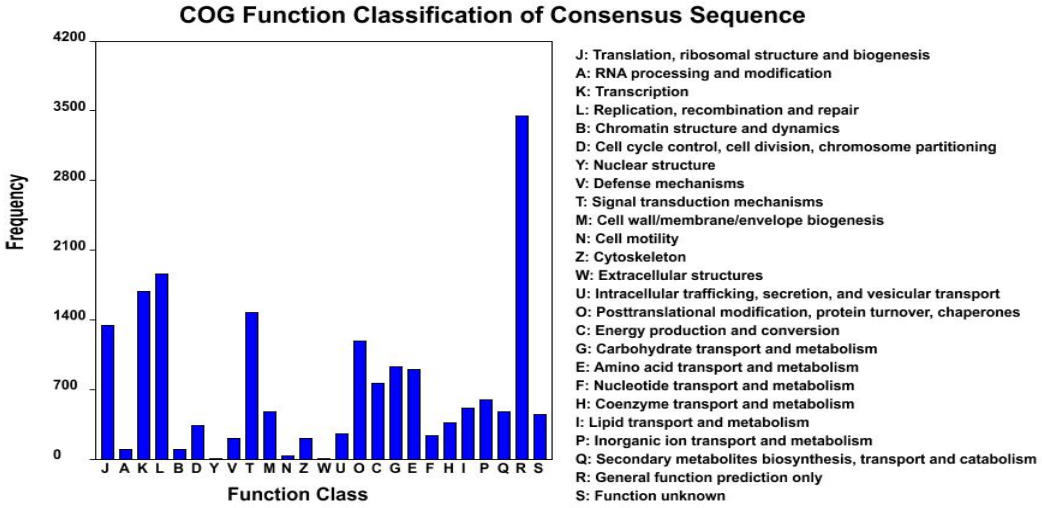

Fig. 3 COG classification of putative proteins

The function gene was categorized according to the KEGG database which is categorize by biochemical pathways. In totally, 9,732 unigenes, accounting for 12.13%, were assigned to 193 KEGG pathways. In our sequence dataset, 245 unigenes related to carbon fixation in photosynthetic organisms, photosynthesis and photosynthesis-antenna proteins pathways were found to be potentially related to high CO_2 concentration.

Identification and annotation of potential differentially expressed genes (DEGs)

This analysis identifies genes that display different expression levels in the T1 and T2 samples from *D. angustifolia*. A total of 4,569 differentially expressed genes were detected between T1 and T2 libraries, with 2,778 genes up-regulated and 1,791 genes down-regulated in T2 sample contrast toto T1 sample (Fig. 4). However, among these DEGs, only 4,109 DEGs were annotated. As shown in Fig. 5, 2,644 and 1,465 genes were up- and down-regulated relative to the T1 sample, respectively (P-value <0.001)

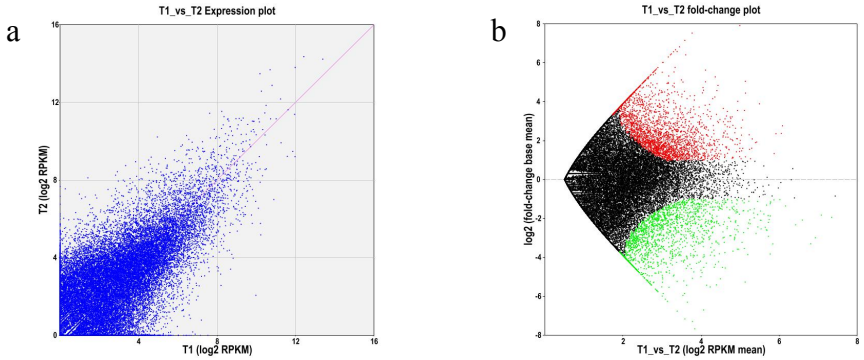

Fig. 4 Identification of DEGs between T1 and T2.

Note: a. The values of Log_2 RPKM in T1 and T2 genes of two samples; b The values of Log_2 T1-RPKM/T2-RPKM in T1 and T2 samples. Red spots represent up-regulated DEGs and green spots indicate down-regulated DEGs. Those shown in black are unigenes that did not show obvious changes in elevated CO_2-treated *D.angustifolia*

Enrichment analysis was conducted to clarify the biological functions of the identified DEGs. The results indicated that 3,741 DEGs were enriched in 58 GO terms ($P \leq 0.05$, Fig. 6). They were mainly involved in processes relative to cell structure, catalytic activity, organelle binding, molecular binding and metabolic processes. The KEGG metabolic pathway analysis also indicated that genes were expressed differentially between T1 and T2 samples. Totally, 1,213 DEGs were enriched in 105 metabolic pathways. The most enriched metabolic pathways included 'carbohydrate metabolism' (229 DEGs), 'amino acid metabolism' (144 DEGs) and 'energy metabolism' (111 DEGs). In addition, the pathways of photosynthesis (25 DEGs), photosynthesis-antenna proteins (13 DEGs), and carbon fixation in photosynthetic organisms (24 DEGs) were detected in two samples. These analyses

indicated that the expressions of some genes were radically altered between the T1 and T2 samples after treatment with a high concentration of CO_2.

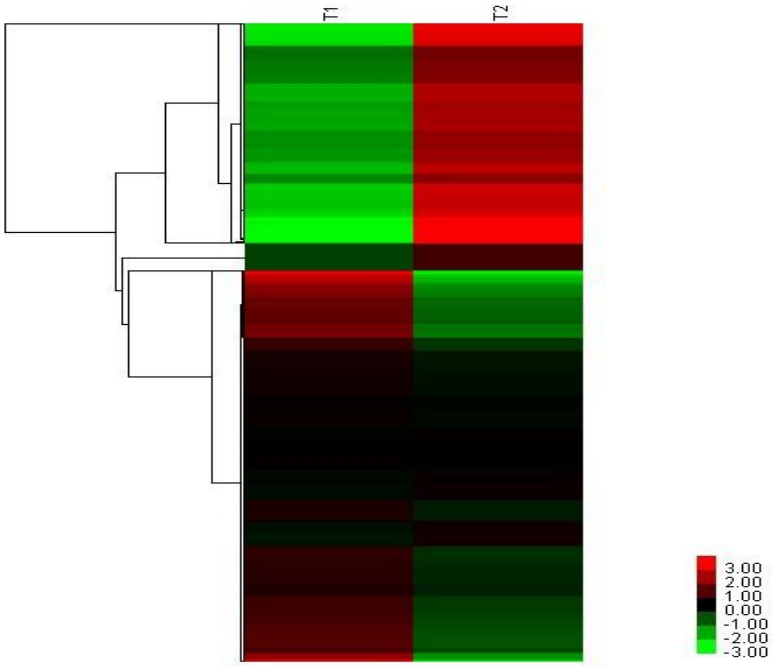

Fig. 5 The clustering map of differently expressed genes

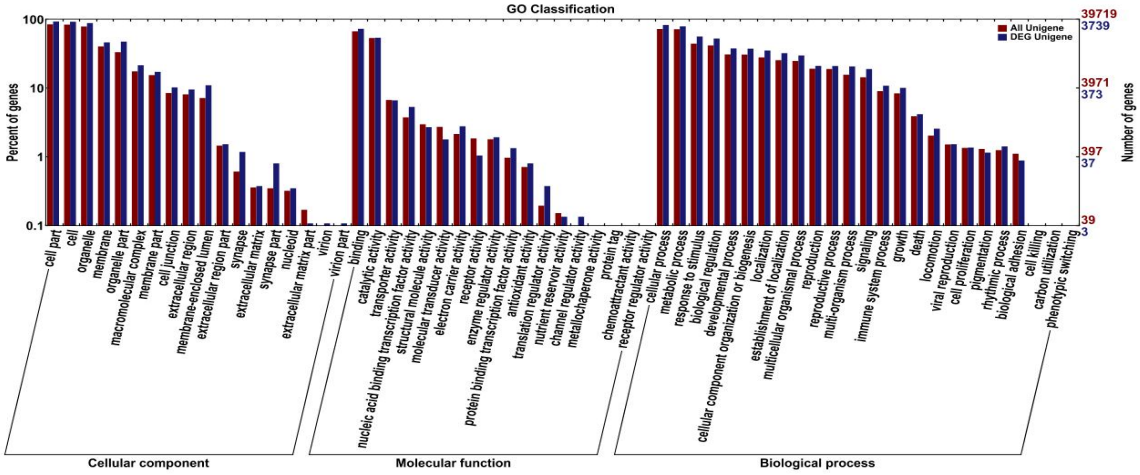

Fig.6 GO functional annotation.of DEGs

Discussion

Construction of an informative transcriptome dataset for *D. angustifolia*

The genomic information of many non-model species were too few to understand the genetic mechanisms and structures underlying their unique features. Therefore,in our present research, the transcriptomes of *D. angustifolia*, a member of the gramineae, were accomplished by the High-throughput *Illumina* sequencing technology, which is the new sequencing technique widely used to find out the important physiological questions in genomic levels.

A total of 9.29 G bp of data with 46,001,292 high-quality reads were obtained from the platform and 80,215 unigenes were identified by *de novo* assembly. Among them, 61.20% were annotated by BLAST analysis and several diferent functional bioinformatics analyses. There were 24,859 unigenes displaying strong homology with the sequences that hits in the database (E-value < $1.0E^{-50}$).

Out of all assembled unigenes, 27.82% had lengths between 300 and 500 bp, among which 81.06% had significant BLAST hits in the public databases. 74.52% of 33,976 unigenes with sequence lengths > 500 bp had significant BLAST scores. The amount of hits was positively correlated with the number of long sequences. Out of all unannotated unigenes (31,121), 58.57%

unigenes that were shorter than 300 bp had no BLAST matches, whereas only 1.04% uigenes that were longer than 1000 bp had no matchest. We also observed the similar correlation in other studies [13, 14, 15]. This demonstrated that the lack of annotation for these unigenes could be ascribed to a genuine lack of hits to sequences in the database rather than a shorter sequence length . The non-annotated unigenes may represent poorly conserved regions [16], such as untranslated regions (UTRs), short sequences, noncoding RNA that do not contain a protein domain or that are mistakely assembled.

Totally, the top five species with BLAST hits, which can annotate the largely genomes, were belong to *Brachypodium distachyon*, *Hordeum vulgare*, *Oryza sativa*, *Sorghum bicolor*, and *Zea mays* species (Fig. 7). The results illustrate that the sequences of the *D. angustifolia* unigenes generated were assembled and annotated correctly.

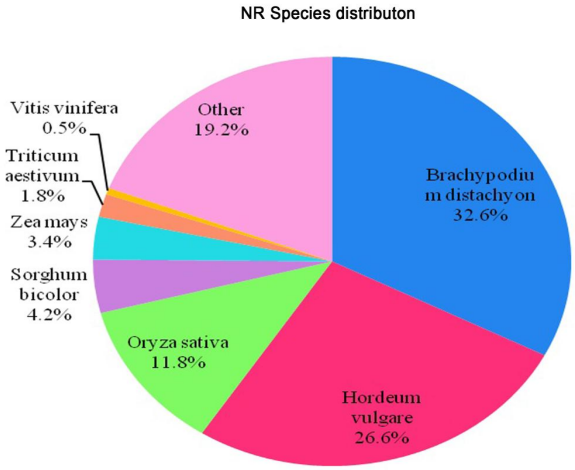

Fig. 7 Species distributions of the top seven BLASTX matches to the nr protein database

We, furthermore, estimated the genes or proteins names, or ORFs of the assembled unigenes, and predicted the conserved domains, gene ontology terms and potential metabolic pathways. The average length of *D. angustifolia* unigenes was shorter compared with previous transcriptomic studies using the same platform in other plants, such as *Litsea cubeba* [17], *Dendrocalamus latiflorus* [18], *Chorispora bungeana* [19], and *Ipomoea batatas* [15].A large numbers of unigenes were allocated to many different GO categories and COG classifications. Using the KEGG database, most representative transcripts were distributed to specific pathways, such as cellular processes, metabolism, environmental information processing and genetic information processing. On the other hand, we found unigenes involved in the process of carbon fixation and carbon utilization. The metabolism of chlorophyll was also detected. These transcripts may control specific pathways and will be useful for studying gene function.

This reserach is the first time to use the *de novo* sequencing and assemble the *D. angustifolia* transcriptome without a reference genome by using *Illumina* paired-end sequencing technology. The results indicate that our final assembly quality is satisfactory. This work will provide a sequence source and facilitate further studies about gene cloning and functional analysis.

Transcriptome comparison identified genes related to high CO_2-stress in *D. angustifolia*

Gene transcription and/or expression is often compared, between different plant organs, different developmental stages or different plants growth conditions [20, 21, 22]. In our study,gene transcriptional changes of seedlings under high CO_2 stress were identified by comparing with untreated controls. There were 4,569 unigenes displaying differences in transcript abundance and 1,213 unigenes were defined as DEGs using the thresholds of FDR ≤ 0.001 and $|\log_2\text{Ratio}| \geq 1$. The potential biological functions of these DEGs were analyzed by the GO clusterin. The results showed that many DEGs were enriched in GO terms, such as ‘carbohydrate metabolism’, ‘amino acid metabolism’ and ‘energy metabolism.’ This information will be useful to elucidate mechanisms of CO_2 absorption and utilization in plants and to find high CO_2 treatment-related genes specific to *D.*

angustifolia. The KEGG enrichment analysis showed that the metabolic pathways involving the DEGs were significant enriched . Fortunately, the genes relating to photosynthesis, photosynthesis-antenna proteins and carbon fixation pathways were transcribed at much higher levels in the *D. angustifolia* under high CO_2 treatment, which was consistent with the results of previous studies that the functional genes retalted with the three pathways may play a vital role in improving photosynthesis and absorbing CO_2.

Conclusions

This study confirmed that *D. angustifolia* leaves respond to a high CO_2 concentration treatment using second-generation sequencing technology. The substantially assembled sequences represented the transcriptome of this plant with a considerable portion. Transcriptome comparison obviousely proved that the transcripts and metabolic pathways played significant roles of *D. angustifolia* for elevated CO_2. The results of this study are as follow:

1. Using *de novo* assembly, 54,605 unigenes were indentified from T1 transcriptome and 46,354 unigenes were indentified from T2. 80,215 All-Unigenes were obtained by combining T1- and T2-clean reads after removal of redundancy.

2. 49,094 of 80,215 Unigenes were annotation by BLAST hits in the public databases. Among them, 39,747 Unigenes were classified into 61 GO terms, 12,484 Unigenes were categorized into 25 Clusters of COG families, and 9,732 Unigenes were assigned to 193 biochemical pathways in KEGG.

3. A total of 73,301 DEGs were detected between T1 and T2 libraries though gene expression annotation and expression pattern of the differentially expressed genes. And 40,527 were up-regulated and 32,334 were down-regulated relative to the *D. angustifolia* under atmosphere CO_2. 4,569 DEGs were identified, including 2,778 up-regulated and 1,791 down-regulated genes relative to T1. Of the total, 4,109 DEGs were annotated function, may be specific gene resource of *D. angustifolia*.

4. 3,741 DEGs were enriched in 58 GO terms. The GO term for ‘translation regulator activity’, ‘membrane-enclosed lumen’, ‘organelle part’, ‘nucleic acid binding transcription factor activity’, ‘channel regulator activity’ and ‘protein binding transcription factor activity’ were enriched most, and close correlated with elevated CO_2. 1,213 DEGs were enriched in 105 metabolic pathways. Among the pathways, ‘photosynthesis’ , ‘photosynthesis- antenna proteins’ and ‘photosynthetic carbon fixation’ were activation, showed an important effect on responding to elevated CO_2 in *D. angustifolia*.

5. Respectively, 86, 34 and 125 genes among the Unigenes were identified relating to regulation of ‘photosynthesis’ , ‘photosynthesis- antenna proteins’ and ‘photosynthetic carbon fixation’. Among them, 62 genes directly participated light reaction and dark reaction.

Acknowledgements

This work was financially supported by Fund of Heilongjiang academy of sciences (ZR201307).

References

[1] IPCC, in: Clinate change 2001:The Seientific Basis, edited by J. T. Houghton, Y. Ding, D. J. Griggs. Cambridge University Press (2001).

[2] IPCC, in: Summary for Policy Makers. Climate Change, edited by S. Solomon, Q. S., M. Manning. Cambridge University Press (2007).

[3] E. A. Ainsworth, A. Rogers: Plant Cell Env. 30 (2007), p. 258

[4] A. Leakey, F. Xu, K. Gillespie, J. M. McGrath, E. A. Ainsworth and D. R.Ort: Proc. Natl. Acad. Sci. U.S.A. 106 (2009b), p. 3597

[5] W. J. Mitsch, X. Wu, in: Soil Management and Gerenhouse Effect, edited by R. Lal, J. E. Kimble, B. A. Levine, CRC Press (1995).

[6] S. E.Trumbore: Proc Natl Acad Sci USA 94 (1997), p. 8284

[7] M.Y. Xu, A. X. Wang and H. W. Ni: Advanced Materials Research Vols. 726-731 (2013), p. 305

[8] K. P. Ma: Chinese Bulletin of Botany S2 (1995), p.1. In Chinese

[9] H. W. Ni, J. H. Chen, Y. J. Wang, H. T. Wang, Y. L. Song, Z. J. Ma and L. Cheng: Journal of Northeast Forestry University 25 (2) (1997), p. 13. In Chinese

[10] H. W. Ni, X. Zhang, L. Jia, Y. H. Gao and H. Y. Wu: Bulletin of Botanical Research 18 (3) (1998), p. 328. In Chinese

[11] D. L. Xu, H. Long, J. J. Liang, J. Zhang, X. Chen, J. L. Li, Z. F. Pan, G. B. Deng and M. Q. Yu: BMC Genomics 13 (2012), p. 133

[12] M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin and D. A. Thompson: Nat Biotechnol 29 (2011), p. 644

[13] T. L. Parchman, K. S. Geist, J. A. Grahnen, C. W. Benkman and C. A. Buerkle: BMC Genomics 11 (2012), p. 180

[14] W. Wei, X. Qi, L. Wang, Y. Zhang and W. Hua: BMC Genomics 12 (2011), p. 451

[15] Z. Y. Wang, B. P. Fang, J. Y. Chen, X. J. Zhang and Z. X. Luo: BMC Genomics 11(2010), p. 726

[16] R. Hou, Z. Bao, S. Wang, H. Su and Y. Li: PLOS ONE 6 (2011), e21560. doi:10.1371/journal.pone.0021560.

[17] X. J. Han, Y. D. Wang, Y. C. Chen, L. Y. Lin and Q. K. Wu: PLOS ONE 8 (10):e76890. doi:10.1371/journal.pone.0076 890.

[18] M. Liu, G. Qiao, J. Jiang, H. Yang and L. Xie: PLoS ONE 7(10) (2012): e46766. doi:10.1371/journal.pone.0046766.

[19] Z. Zhao, L. Tan, C. Dang, H. Zhang and Q. Wu: BMC Plant Biol 12 (2012): 222. doi:10.1186/1471-2229-12-222. PubMed: 23171377.

[20] C. Y. Shi, H. Yang, C. L. Wei, O. Yu, Z. Z. Zhang, C. J. Jiang, J. Sun, Y. Y. Li, Q. Chen and T. Xia: BMC Genomics 12 (2011), p. 131

[21] S. Zenoni, A. Ferrarini, E. Giacomelli, L. Xumerle, M. Fasoli, G. Malerba, D. Bellin, M. Pezzotti, M. Delledonne: Plant Physiol 152 (2010), p. 1787

[22] P. H. Li, L. Ponnala, N. Gandotra, L. Wang, Y. Q. Si, S. L. Tausta, T. H. Kebrom, N. Provart, R. Patel, C. R. Myers, E. J. Reidel, R. Turgeon, P. Liu, Q. Sun, T. Nelson and T. P. Brutnell: Nat Genet 42 (2010), p. 1060