Transcriptomic analysis of cold stress-responsive gene expression and biochemical pathways in *Liriodendron chinense*

D. Hwarari[1], Y. Liming[1,2]
[1]College of Biotechnology and Environment, Nanjing Forestry University, Nanjing, 210037, China
* Email: yangliming@njfu.edu.cn

Background

Liriodendron chinense (Hemsl.) Sarg is a protected tree species belonging to the Magnolia family, *Magnoliaceae,* widely distributed in the southern part of China with important commercial and ecological benefits. Cold stress is a common abiotic stress which limits the growth and productivity of *Liriodendron. chinense* during the winter season in mainland China's seasonal cycle. Seedlings of *Liriodendron. chinense* exposed to extreme low temperatures have stunted growth and structural injuries, thus limiting its planting regions. In order to expand the cultivation of Liriodendron in cold regions, it is of great theoretical and practical significance to study the response of Liriodendron to low temperature stress.

Methods

RNA quantification and qualification

RNA degradation and contamination was monitored on 1% agarose gels. RNA purity was checked using the Nanophotometer® spectrophotometer (IMPLEN, CA, USA). RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

Library preparation for Transcriptome sequencing

A total amount of 1 μg RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext® UltraTM RNA Library Prep Kit for Illumina® (NEB, USA) following manufacturer's recommendations and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer(5X). First strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H-). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H.

Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, NEBNext Adaptor with hairpin loop structure were ligated to prepare for hybridization. In order to select cDNA fragments of preferentially 250~300 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA). Then 3 μl USER Enzyme (NEB, USA) was used with size-selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95 °C before PCR. Then PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. At last, PCR products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system.

Clustering and sequencing

The clustering of the index-coded samples was performed on a cBot Cluster. Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumia) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina Novaseq platform and 150 bp paired-end reads were generated.

Data Analysis

Quality control

Raw data (raw reads) of fastq format were firstly processed through in-house perl scripts. In this step, clean data (clean reads) were obtained by removing reads containing adapter, reads containing ploy-N and low-quality reads from raw data. At the same time, Q20, Q30 and GC content the clean data were calculated. All the downstream analyses were based on the clean data with high quality.

Reads mapping to the reference genome

Reference genome and gene model annotation files were downloaded from genome website directly. Index of the reference genome was built using Hisat2 v2.0.5 and paired-end clean reads were aligned to the reference genome using Hisat2 v2.0.5. We selected Hisat2 as the mapping tool for that Hisat2 can generate a database of splice junctions based on the gene model annotation file and thus a better mapping result than other non-splice mapping tools.

Novel transcripts prediction

The mapped reads of each sample were assembled by StringTie (v1.3.3b) (Mihaela Pertea.et al. 2015) in a reference-based approach. StringTie uses a novel network flow algorithm as well as an optional de novo assembly step to assemble and quantitate full length transcripts representing multiple splice variants for each gene locus.

Quantification of gene expression level

Feature Counts v1.5.0-p3 was used to count the reads numbers mapped to each gene. And then FPKM of each gene was calculated based on the length of the gene and reads count mapped to this gene. FPKM, expected number of Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced, considers the effect of sequencing depth and gene length for the reads count at the same time, and is currently the most commonly used method for estimating gene expression levels.

Differential expression analysis

Differential expression analysis of two conditions/groups (two biological replicates per condition) was performed using the DESeq2 R package (1.16.1). DESeq2 provide statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting P-values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with an adjusted P-value <0.05 found by DESeq2 were assigned as differentially expressed.

Prior to differential gene expression analysis, for each sequenced library, the read counts were adjusted by edgeR program package through one scaling normalized factor. Differential expression analysis of two conditions was performed using the edgeR R package (3.18.1). The P values were adjusted using the Benjamini & Hochberg method.

Corrected P-value of 0.05 and absolute foldchange of 2 were set as the threshold for significantly differential expression.

GO and KEGG enrichment analysis of differentially expressed genes

Gene Ontology (GO) enrichment analysis of differentially expressed genes was implemented by the clusterProfiler R package, in which gene length bias was corrected. GO terms with corrected Pvalue less than 0.05 were considered significantly enriched by differential expressed genes. KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-through put experimental technologies (http://www.genome.jp/kegg/). We used clusterProfiler R package to test the statistical enrichment of differential expression genes in KEGG pathways.

Results

This research investigated the expression profiles of genes in the leaves of *Liriodendron. chinense* seedling under 4 ˚C by transcriptome sequencing. More than 42.7 million (Table. 1) clean reads were obtained from seven time points (0h, *1h, 3h, 6h, 12h,1d* and *3d*) of cold stress and de novo transcriptome assembly was performed, RNA was extracted from the leaves.

Table 1. Summary of transcriptome sequencing assembly results

Treatment	raw reads	clean reads	error rate	Q20	Q30	GC pct
Col	46999785	45175039	0.03	97.92	93.83	48.55
1hr	45898918	43658286	0.02	98.43	95.05	48.53
3hr	45022251	42709432	0.02	98.37	94.93	48.54
6hr	45227409	42513602	0.02	98.5	95.24	48.54
12hr	46685207	44208161	0.02	98.42	95.03	48.77
1d	46666157	44264520	0.02	98.41	95.05	48.24
3d	47103033	44598493	0.03	98.09	94.3	47.11

A total of 35 126 DEGs were significantly differentially expressed ($p<0.05$), using the DESeq2 R package (1.16.1), a model based on the negative binomial distribution. P-values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate ([FDR] < 0.05). 52.1% of the total DEGs were up-regulated and 47.8% were down regulated. Differential Expression analysis showed that, most of DEGs were expressed during long time periods, for instance 4062, 6951, 8401were expressed at *12h, 1d* and *3d* respectively. Analysis of the expression patterns of 20 differentially expressed genes (DEGs) by quantitative real-time RT-PCR (qRT-PCR) confirmed the accuracy of the RNA-Seq results.

Gene Ontology and KEGG pathway functional enrichment analyses allowed us to better understand these DEGs. Functional classification of DEGs was achieved using Gene Ontology enrichment analysis. These DEGs were divided into three categories of GO: biological process, molecular function, and cellular component. The enrichment pathways of the DEGs in different treatments reflected the preferential biological functions of the different temperature treatments. For instance, long period exposed plants had more genes upregulated in the BP category, with respect to

transcription DNA template processes, regulation of RNA metabolic process, regulation of macromolecules biosynthesis and in regulation of biosynthesis process. The most significant transcriptomic changes observed under cold stress were related to plant-pathogen interaction, starch and sucrose metabolism, plant-hormone signal transduction and protein processing in Endoplasmic Reticulum.

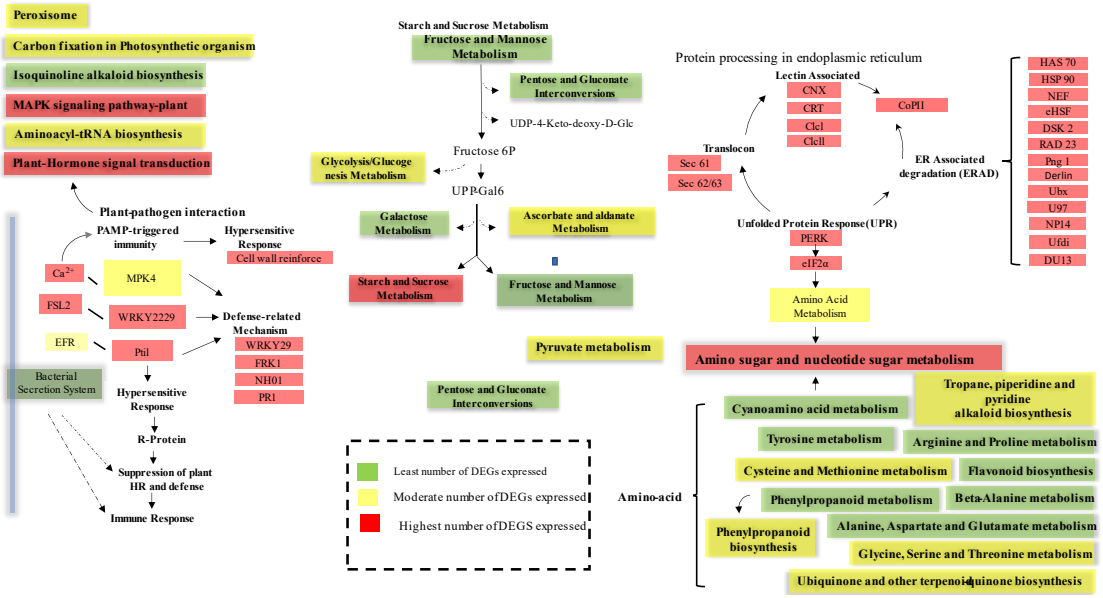

Figure 1. Depicting significant pathways with differing enrichment levels and corresponding number of DEGs, green (low enrichment and least no. of DEGs), yellow (moderate enrichment and no. of DEGs) and red (highly enriched with the highest number of DEGs.

We used KEGG database resource to better understand high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information. Many KEGG pathways were significantly enriched, among which the 'plant hormone signal transduction', 'plant-pathogen interaction', 'starch and sucrose metabolism', 'mitogen-activated protein kinase (MAPK) signaling pathway', 'oxidative phosphorylation' pathways and 'ribosome' pathways were the most highly represented. The 'plant hormone signal transduction' pathway (ath04075) exhibited the most DEGs, suggesting that plant hormones play significant roles in resistance to cold stress in P. chinense. The second largest number of DEGs were in the 'plant-pathogen interaction' category (ath04626), indicating that plants are vulnerable to pathogenic bacteria during exposure to cold stress. The 'MAPK signaling pathway' (ath04016) exhibited the fourth largest number of DEGs, indicating that during cold stimulation, the expression of internal genes in plants is regulated by various signaling substances for adaptation to the cold environment. 'Glycerol metabolism, starch and sucrose metabolism and ribosome metabolism' (ath00500) were the most highly represented.

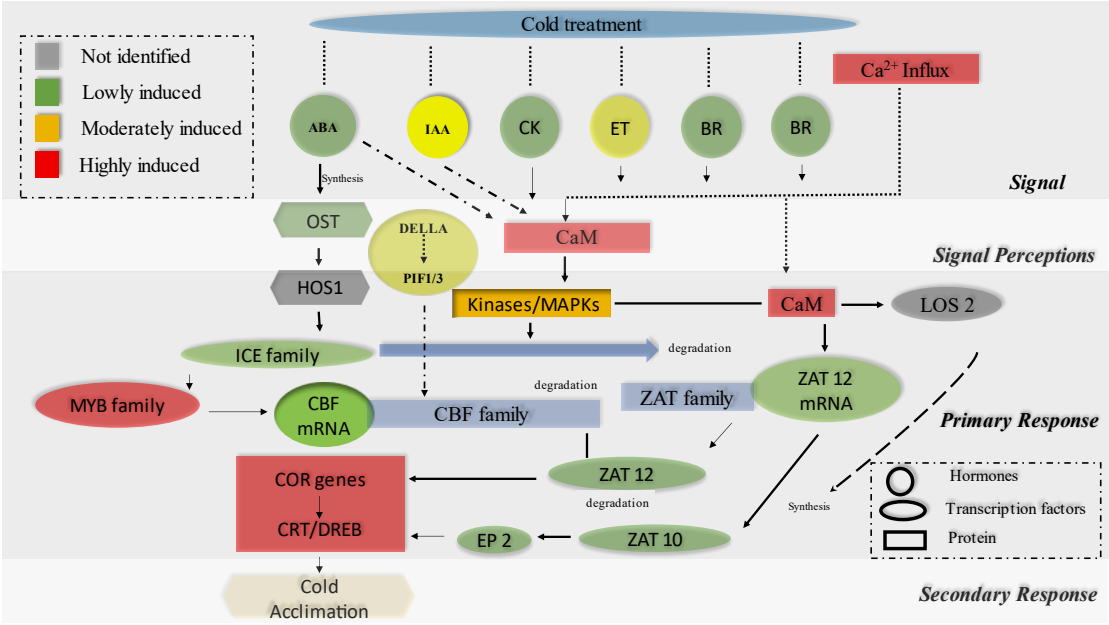

Figure 2. Shows gene pathways and the level of gene induction, shown by different colors.

In addition, 99 transcription factors, including members of the CBF, bHLH, MYB, NAC, ZAT, bZIP families, other Kinases and COR genes, were identified as cold responsive. Results showed that, the COR genes were highly induced at long period of cold stress (Figure. 1), 12h, 1d and 3d. Hormonal pathway was also observed to be lowly induced.

Conclusion

In this study, we provide the transcriptome-level patterns of gene expressions in cold-stressed *L. chinense*, and also provide important clues not only for understanding the molecular mechanisms of cold stress in *Liriodendron chinense*.

Project Funding: The work was supported by the National Natural Science Foundation of China (No.31971682)

References

Banerjee, A. and A. Roychoudhury (2018). "Abiotic stress, generation of reactive oxygen species, and their consequences: an overview." Revisiting the role of reactive oxygen species (ROS) in plants: ROS Boon or bane for plants: 23-50.

Bhatla, S. C. (2018). Abiotic Stress. Plant Physiology, Development and Metabolism, Springer**:** 969-1028.

Gong, Z., et al. (2020). "Plant abiotic stress response and nutrient use efficiency." SCIENCE CHINA Life Sciences **63**(5): 635-674.

Miki, Y., et al. (2019). "Temporal proteomics of Arabidopsis plasma membrane during cold-and de-acclimation." Journal of proteomics **197**: 71-81.

Palta, J. P. and L. S. Weiss (2018). "ICE FORMATION AND FREEZING INJURY: ANOVERVIEW ON THE SURVIVAL MECHANISMS AND MOLECULAR ASPECTS OF INJURY AND COLD ACCLIMATION IN HERBACEOUS PLANTS." Advances in plant cold hardiness: 11.

Rosa, S. (2011). "Chromatin dynamics in Arbidopsis development: a live cell imaging approach."

Wang, W., et al. (2019). "The late embryogenesis abundant gene family in tea plant (Camellia sinensis): Genome-wide characterization and expression analysis in response to cold and dehydration stress." Plant Physiology and Biochemistry **135**: 277-286.

Wu, Z., et al. (2019). "Cold stress activates disease resistance in Arabidopsis thaliana through a salicylic acid dependent pathway." Plant, cell & environment **42**(9): 2645-2663.

Xu, J., et al. (2020). "combined transcriptomic and metabolomic analyses uncover rearranged gene expression and metabolite metabolism in tobacco during cold acclimation." Scientific reports **10**(1): 1-13.

Yang, X., et al. (2019). "Transcriptome profiling of Populus tomentosa under cold stress." Industrial Crops and Products **135**: 283-293.

Yuan, P., et al. (2018). "Calcium signaling in plant autoimmunity: A guard model for ATSR1/CAMTA3-mediated immune response." Molecular Plant **11**(5): 637-639.

Zhao, M., et al. (2020). "Sesquiterpene glucosylation mediated by glucosyltransferase UGT91Q2 is involved in the modulation of cold stress tolerance in tea plants." New Phytologist **226**(2): 362-372.