Full-length transcriptome sequencing reveals a low-temperature-tolerance mechanism in Medicago falcata roots

Guowen Cui
Northeast Agricultural University
Hua Chai
Northeast Agricultural University
Hang Yin
Northeast Agricultural University
Mei Yang
Northeast Agricultural University
Guofu Hu
Northeast Agricultural University
Mingying Guo
Hulunbuir Grassland Station
Rugeletu Yi
Hulunbuir Grassland Station
Pan Zhang (zhangpgrass@126.com)
Northeast Agricultural University https://orcid.org/0000-0002-2628-4852

Research article

Keywords: Medicago falcata, Low-temperature stress, Single-molecular long-read sequencing, RNA-Seq, Weighted gene coexpression network analysis

Posted Date: August 15th, 2019

DOI: https://doi.org/10.21203/rs.2.12899/v1

Version of Record: A version of this preprint was published on December 21st, 2019. See the published version at https://doi.org/10.1186/s12870-019-2192-1.

Abstract

Background Low temperature is one of the main environmental factors that limits crop growth, development and production. Medicago falcata is an economically and ecologically important legume that is closely related to alfalfa and exhibits better tolerance to low temperature than alfalfa. Understanding the low-temperature-tolerance mechanism of M. falcata is important for the genetic improvement of alfalfa. Results In this study, we explored the transcriptomic changes in low-temperature-treated M. falcata roots by combining SMRT and NGS technologies. A total of 115,153 nonredundant sequences were obtained, and 8,849 AS events, 73,149 SSRs and 4,189 LncRNAs were predicted. A total of 111,587 genes from SMRT were annotated, and 11,369 DEGs were identified in this paper that are involved in plant hormone signal transduction, protein processing in endoplasmic reticulum, carbon metabolism, glycolysis/gluconeogenesis, starch and sucrose metabolism, and endocytosis pathways. We characterized 1,538 TF genes into 45 TF gene families, and the most abundant TF family was WRKY, followed by ERF, MYB, bHLH and NAC. A total of 134 genes were differentially coexpressed at all five temperature points, including 101 upregulated genes and 33 downregulated genes. PB40804, PB75011, PB110405 and PB108808 were found to play crucial roles in the tolerance of M. falcata to low temperature. The WGCNA results showed that the MEbrown module was significantly correlated with low-temperature stress in M. falcata. Electrolyte leakage was correlated with most genetic modules and corroborated that electrolyte leakage can be used as direct stress markers to reflect cell membrane damage from low-temperature stress in physiological assays. The consistency between the qRT-PCR results and RNA-Seq analyses confirm the validity of the RNA-Seq data and the analysis of the regulation of low-temperature stress in the transcriptome. Conclusions The full-length transcripts generated in this study provided a full characterization of the gene transcription of M. falcata and are useful for mining new low-temperature stress-related genes specific to M. falcata. These new findings facilitate the understanding of low-temperature-tolerance mechanisms in M. falcata.

Background

Low temperature is one of the main environmental factors that limits plant growth, development and geographical distribution [1]. Low-temperature stress, consisting of chilling stress (<10 °C) and freezing stress (<0 °C), can reduce crop productivity to some extent [2]. The effects of low temperature on plants depend on developmental stage and exposure time. Under low-temperature stress, young tissues and organs are more seriously damaged than old tissues and organs, and the reproductive stage is more sensitive to low temperature than the vegetative stage [3]. Exposure to low temperature causes physiological and molecular changes in plants, such as the inhibition of photosynthetic activity, reduced water uptake, oxidative stress via an increase in reactive oxygen species (ROS) accumulation, an increase in intracellular pH and osmotic pressure, and functional abnormalities in chloroplasts, mitochondria and other organelles, which lead to metabolic disorders in plants [4-6]. Furthermore, low temperature temporarily inhibits sucrose synthesis, and the rearrangement of the membrane causes changes in the

stability and mobility of proteins and a shift in redox homeostasis, which decreases enzyme activities and alters metabolism homeostasis and gene transcription [7, 8]

Plants have developed many mechanisms and pathways that enable them to minimize the negative effect of low temperature and to grow and reproduce successfully [9]. The osmolytes induced by low-temperature stress, including proline, soluble sugar, and cold-induced stress proteins (dehydrins and LEA proteins), can improve the cell osmotic potential, protect the stability of biological membranes, alleviate oxidative damage limitation, and even act as signals to regulate the expression of stress-related genes [10, 11]. The overexpression of *SlNAM1*, a typical NAC gene, improves low-temperature tolerance in transgenic tobacco by improving osmolytes and reducing the H_2O_2 and superoxide anion radical ($O_2^{\cdot-}$) contents under low temperature, which contribute to alleviate the oxidative damage of the cell membrane after low-temperature stress [12]. Low temperature induces the production of Ca^{2+}, which can be sensed by corresponding receptors, among which lipid Ca^{2+} channels may be the primary cryogenic signal receptors, and can then activate calcium response protein kinase (CPKs CIPKs, and CRLK1) and MAPK cascade responses, which regulate cold-responsive (COR) gene expression [13, 14]. The overexpression of COLD1 (jap) significantly enhances chilling tolerance by interacting with the G-protein alpha subunit to activate the Ca^{2+} channel to sense low temperature and to accelerate G-protein GTPase activity [15]. Low-temperature stress rapidly induces the expression of many transcription factors (TFs), including the AP2 domain proteins CBFs, which then activate the expression of numerous downstream COR genes [16-18]. The expression of the CBF gene is controlled by upstream TFs, such as the bHLH TF ICE1. ICE1 is subjected to sumoylation and polyubiquitylation and subsequent proteasomal degradation, mediated by the SUMO E3 ligase SIZ1 and ubiquitin E3 ligase HOS1, respectively [13, 16].

Single-molecule long-read sequencing (SMRT), which was developed by PacBio Biosciences RSII, provides a third-generation sequencing platform. Because of its long reads, it is widely used in genome sequencing and has eliminated many restrictions in sequencing by generating full-length or long sequences [19-22]. Next-generation sequencing (NGS) technology (RNA-Seq) can provide expression profiles of either coding or noncoding RNAs and generate digital data of gene expression that enable rapid and cost-effective genomic and transcriptomic studies for most major crops, including rice [23], wheat [24], and grape [25]. The approach of combining NGS and SMRT sequencing has been applied to generate comprehensive information frequently at the transcriptional level, allowing the identification of key functional and regulatory genes involved in abiotic stress resistance [26-28]. The changes in genes responding to low-temperature stress revealed by transcriptome analyses are mainly separated into three parts. The first part consists of signaling components, including TFs and protein kinases, which regulate downstream gene expression. The second part consists of functional components, including some enzymes in metabolic pathways that directly protect the plant cell from low-temperature damage. The last part consists of some small noncoding RNAs, mainly miRNAs, which regulate target gene expression [29, 30].

Medicago falcata L., an economically and ecologically important legume herbage with an expanse from northern Mediterranean regions to northern Russia, is closely related to alfalfa and exhibits better tolerance to low temperature than alfalfa [31-33]. Understanding the mechanism of low-temperature tolerance in *M. falcata* is important for the genetic improvement of alfalfa, the most important forage legume with high biomass productivity, optimal nutritive profile and adequate persistence [34]. Although the low-temperature tolerance of *M. falcata* is a research hotspot from the morphological level to the physiological biochemistry and molecular biology levels [35-38], few studies have been conducted on the low-temperature tolerance of *M. falcata* at the transcriptome level. In this paper, we combined SMRT and NGS sequencing to generate expression profiles of *M. falcata* roots under low-temperature stress. In total, 115,153 nonredundant sequences were obtained from *M. falcata* roots, and 11,369 differentially expressed genes (DEGs) were identified in this paper, including 134 genes that were differentially coexpressed at all five temperature points. These findings provide a complete characterization of gene transcription and facilitate the understanding of low-temperature-tolerance mechanisms in *M. falcata.*

Results

Physiological responses of *M. falcata* under low-temperature stress

We examined the electrolyte leakage; malondialdehyde (MDA) content; superoxide dismutase (SOD), catalase (CAT) and peroxidase (POD) activities; and superoxide anion radical ($O_2^{\cdot-}$), soluble protein, reduced glutathione (GSH), proline and soluble sugar contents to investigate the physiological changes in *M. falcata* roots exposed to low-temperature stress for 2 h (Fig. 1). Under low-temperature stress, the electrolyte leakage increased gradually with decreasing temperature (Fig. 1A). The highest electrolyte leakage was observed at -15 °C, which was 4.18 times higher than that in the control environment. The MDA content decreased gradually and peaked at -10 °C (relative to the control) and slightly decreased at -15 °C (Fig. 1B). No obvious difference was observed in SOD activity, and this value increased after low-temperature treatment and decreased from 4 °C to -15 °C (Fig. 1C). The changes in CAT and POD activities were completely contrasting under different temperatures. CAT activity first increased, and POD activity first decreased (Fig. 1D - E). The content of $O_2^{\cdot-}$ increased significantly after low-temperature treatment and reached a peak at -10 °C (Fig. 1F). The content of soluble protein decreased first after low-temperature treatment and increased after 4 °C and then decreased significantly after 0 °C (Fig. 1G). The GSH content increased significantly after low-temperature treatment and reached a peak at -5 °C (Fig. 1H). The proline content decreased slightly in the low-temperature environment (Fig. 1I). The highest soluble sugar content was measured at -10 °C (Fig. 1J).

M. falcata transcriptome sequencing

To identify and characterize the transcriptome of *M. falcata* roots under low-temperature stress, we measured the roots under CK (room temperature), 4 °C, 0 °C, -5 °C, -10 °C and -15 °C for 2 h by joining the PacBio SMRT and NGS technologies for whole-transcriptome profiling. In total, 125.48 Gb clean data were obtained by RNA-Seq, yielding 418,495,643 reads with a GC content of 42.61% and a QC 30 of

85.51% (Additional file 1: Table S1). With SMRT, a total of 8 cells were used to obtain 19.27 Gb clean data. A total of 1,202,336 polymerase reads were obtained, and then the polymerase reads with fragments less than 50 bp and sequence accuracy less than 0.75 were filtered. Then, 8,428,385 subreads were obtained by filtering the remaining sequences from the linker, filtering out the linker sequence and filtering out the subreads with fragments less than 50 bp (Table 1). A total of 552,818 reads of inserts (ROIs) were extracted from the original sequence, of which 270,750 were full-length nonchimeric reads (FLNC), and 223,319 were nonfull-length reads (Table 1). The full-length sequences were clustered using the RS_IsoSeq module of SMRT Analysis software. A total of 131,118 consensus isoforms were obtained; 99,490 high-quality isoforms were obtained using the nonfull-length sequence alignment, and 31,628 low-quality isoforms were obtained and corrected using RNA-Seq data [39]. The redundancy in the high-quality and corrected low-quality transcript sequences of each sample were eliminated by CD-HIT software [40], and 115,153 nonredundant sequences were obtained. Based on the nonredundant sequences of each sample, we predicted a total of 8,849 alternative splicing (AS) events with the IsoSeq_AS_de_novo script (Additional file 2: Table S2) [41], 73,149 simple sequence repeats (SSRs) with the MIcroSAtellite identification tool (Additional file 3: Table S3), and 4,189 long noncoding RNAs (LncRNAs) with Coding Potential Calculator (CPC) [42], Coding-Non-Coding Index (CNCI) [43], Coding Potential Assessment Tool (CPAT) [44], and Protein family (Pfam) (Additional file 4: Figure S1).

Annotation and expression of transcripts under low-temperature stress

To acquire the most comprehensive annotation, all full-length transcripts from SMRT were aligned with NCBI nonredundant protein database (NR), SwissProt, Gene Ontology (GO), Clusters of Orthologous Groups (COG), euKaryotic Ortholog Groups (KOG), Pfam, and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases using BLAST software (version 2.2.26) [45], and a total of 111,587 genes from SMRT were annotated, of which 7,155 genes' length were >= 300 bp and <1,000 bp, and 104,432 genes' length were >= 1,000 bp (Additional file 5: Table S4). Among these annotated sequences, 111,384 sequences had significant matches in the Nr database, 89,117 sequences had significant matches in the Pfam database, 82,433 sequences had matches in the SwissProt database, and 3,385 had effective matches in the GO database. Based on the homology with sequences of different species, 97,831 (87.87%) sequences were found against *M. truncatula*, and 3,463 (3.11%) sequences had clandestine hits with *Cicer arietinum*, followed by *M. sativa* (1,662, 1.49%), *Glycine max* (1,011, 0.91%), *Rhizoctonia solani* (528, 0.47%), *Fusarium oxysporum* (388, 0.35%), *G. soja* (369, 0.33%), *Phaseolus vulgaris* (349, 0.31%), *Vitis vinifera* (266,0.24%), and *M. falcata* (181, 0.16%). Only 5,283 (4.75%) annotated sequences had similarity with other plant species (Additional file 6: Figure S2).

To evaluate gene expression levels in response to low-temperature stress, we mapped all clean data back onto the assembled transcriptome, and the readcount for each gene was obtained from the mapping results by using RNA-Seq by Expectation Maximization (RSEM) software (Additional file 7: Table S5) [46]. The mapped readcount for each gene was then converted into the expected number of fragments per kilobase of transcript per million mapped reads (FPKM) (Additional file 8: Table S6). FPKM values can eliminate the impact of transcript length and sequencing differences on computational expression. The

boxplot diagram of FPKM values indicated that gene expression levels were not evenly distributed in the different experimental environments (Fig. 2A). The Pearson correlation coefficient was used to evaluate the correlations of each biological sample, with r^2 values close to 1 indicating a strong correlation between two replicate samples (Fig. 2B). Then, all sequences were used for further DEG analysis after excluding the abnormal samples.

Analysis of DEGs in response to low-temperature stress

In total, 11,369 DEGs displaying up- or downregulation between samples (fold change≥2 and false discovery rate (FDR)<0.01) collected at any pair of temperature points were identified by comparing gene expression levels under low-temperature stress (Additional file 9: Table S7). Clustering patterns of DEGs under low-temperature stress were determined by hierarchical cluster analysis of all DEGs (Fig. 3A). All DEGs with the same or similar expression levels were clustered, and a set of genes was quickly activated during the early stage of low-temperature stress (4 °C), and other genes were activated under a freezing temperature (-10 °C). The 11,369 DEGs identified were grouped into six subclusters by K-means coexpression cluster analysis (Fig. 3B). The expression level of genes in subcluster 1 (1,271 genes) began to increase after the temperatures dropped to -5 °C and reached a maximum at -10 °C; then, the expression levels decreased rapidly. KEGG analysis of genes in subcluster 1 revealed that most were involved in starch and sucrose metabolism, plant-pathogen interaction, and galactose metabolism. The expression level of genes in subcluster 2 (2,226 genes) increased significantly after low-temperature treatment, then the expression level first decreased slowly and then rapidly at each temperature point below 4 °C. Genes in this subcluster functioned mostly in circadian rhythm, plant-pathogen interaction, and plant hormone signal transduction. The genes in subcluster 3 (1,460 genes) and subcluster 4 (2,695 genes) were both upregulated after low-temperature treatment and downregulated after 4 °C treatment. The difference was that the decrease in the expression level of genes in subcluster 3 fluctuated while that in subcluster 4 was continuous. Genes involved in starch and sucrose metabolism, protein processing in endoplasmic reticulum, fatty acid degradation, and tyrosine metabolism were enriched in subcluster 3. Genes in subcluster 4 were enriched in plant-pathogen interaction, starch and sucrose metabolism and plant hormone signal transduction. Genes in subcluster 5 (1,613 genes) were weakly downregulated after low-temperature treatment, then strongly upregulated from 4 °C to 0 °C, and then upregulated again under -15 °C treatment. Genes in this subcluster mostly functioned in plant-pathogen interaction, phenylalanine metabolism, and the plant hormone signal transduction pathway. Genes in subcluster 6 (2,104 genes) were downregulated at all times, and most of them functioned in phenylpropanoid biosynthesis, circadian rhythm, and ubiquitin-mediated proteolysis.

Identification of putative TFs

TFs play an important role in cell function and development and directly regulate gene expression through interactions with themselves and other proteins to participate in plant stress regulations, including low-temperature-related processes. In this study, 1,538 TF genes were differentially expressed between different temperature points and were classified into 45 TF gene families according to

PlantTFDB (http://planttfdb.cbi.pku.edu.cn/) (Additional file 10: Table S8). The most abundant TF family was the WRKY (186 genes) family, followed by the ERF (165 genes), MYB (143 genes), bHLH (131 genes) and NAC (111 genes) families.

Comparison of coexpressed genes between the low-temperature-treated and control samples

Under low-temperature stress, a total of 8,683 DEGs were identified by a comparison of each temperature point to the control environment. Interestingly, the number of upregulated DEGs (5,876 genes) was much higher than the number of downregulated DEGs (2,807 genes), and 134 genes were differentially coexpressed at all five temperature points (Additional file 11: Table S9). As shown in Fig. 4A, there were 101 upregulated genes and 33 downregulated genes in all five comparisons. These 134 coexpressed genes were assigned into the GO categories of biological process, cellular component and molecular function (Fig. 4B). In the biological process category, “metabolic process”, “cellular process”, “single-organism process” and “biological regulation” were the most enriched terms. In the cellular component category, “cell”, “cellular component” and “organelle” were the most enriched. In the molecular function category, “binding” was the most enriched term, followed by “catalytic activity”. COG functional classification of the 134 coexpressed genes showed that most of the genes were enriched in “General function prediction only”, “Transcription”, “Replication, recombination and repair”, “Signal transduction mechanisms” and “Carbohydrate transport and metabolism” (Fig. 4C). KEGG enrichment showed that most of the coexpressed genes were enriched in the “Circadian rhythm - plant”, “Cysteine and methionine metabolism”, “Arginine and proline metabolism”, “Phenylpropanoid biosynthesis”, “Galactose metabolism”, “Plant-pathogen interaction” and “Biosynthesis of amino acids” pathways (Fig. 4D).

Weighted gene coexpression network analysis (WGCNA) of DEGs in response to low-temperature stress

WGCNA was performed to obtain a better understanding of which genes within these complex signaling networks were the most connected hubs. The number of genes in the module was clustered according to the expression levels, and genes with a high clustering degree were allocated to the same models. The 8,683 DEGs identified by comparing low-temperature-treated and control samples were clustered based on topological overlap, and then the gene modules were obtained from the dynamic tree cut. Finally, 12 gene modules were identified after merging modules with similar expression patterns (Fig. 5A). The magenta modules contained the most genes, 1,092 genes, and the violet nodules contained the fewest genes, 70 genes (Additional file 14: Table S10). The gray module was not a true module but a place to categorize the remaining genes that were not well correlated with any one of the significant colored modules. The kME (module eigengene-based connectivity) was calculated for each gene to every module, and 449 genes were found to act as a hub in more than one module.

All 12 genetic modules with a module characteristic value $P < 0.05$ were used to find the modules that were highly correlated with samples and physiological indicators (Fig. 5B). The control samples were highly correlated with the MEblue module. The samples under 4 °C treatment were highly correlated with the MEmagenta module. The samples under 0 °C treatment were highly correlated with the MElightcyan module. The samples under -5 °C treatment were highly correlated with the MElightyellow module. The

samples under -10 °C treatment were highly correlated with the MEbrown module. The samples under -15 °C treatment were highly correlated with the MEgray60 module. The MEbrown module was found to be significantly correlated with physiological indicators and may play a key role in low-temperature tolerance in *M. falcata*. COG classification results showed that the MEbrown module genes were involved in 23 major categories, including “General function prediction only”, “Signal transduction mechanisms”, “Transcription” and “Replication, recombination and repair” (Additional file 13: Figure S3A). GO analysis results showed that the MEbrown module genes were mainly involved in 13 biological processes, such as “protein phosphorylation”, “regulation of transcription, DNA-templated”, and “oxidation-reduction process”, were distributed to eight cellular component terms, such as “integral component of membrane”, “plasmodesma”, “chloroplast stroma” and “nucleus”, and consisted of 14 molecular functions, which included ATP binding, protein serine/threonine kinase activity, and cation binding (Additional file 13: Figure S3B). KEGG enrichment showed that most of the genes in the MEbrown module were enriched in “Starch and sucrose metabolism”, “Plant-pathogen interaction”, “Circadian rhythm - plant”, “Protein processing in endoplasmic reticulum”, “Galactose metabolism” and “Plant hormone signal transduction” (Additional file 13: Figure S3C).

Confirmation of RNA-Seq sequencing data by qRT-PCR analysis

The DEGs associated with low-temperature stress were selected for qRT-PCR assays to confirm the SMRT sequencing data. Eighteen genes were selected randomly from 134 DEGs coexpressed at all five temperature points. We found that the fold-changes in the expression calculated by sequencing data did not exactly match the expression values detected by qRT-PCR analysis, but the expression profiles were basically consistent for all 18 genes (Additional file 14: Figure S4). These analyses confirmed the reliability of the gene expression values generated from SMRT sequencing data.

Discussion

The approach of combining NGS and SMRT has become increasingly popular in researching plant responses to adverse environments for providing high-quality and increasingly complete assemblies at the transcriptome level [26-28]. SMRT can generate full-length or long sequences, and the high error rate can be overcome using short and high-accuracy NGS reads [19, 22]. In this study, we combined SMRT and RNA-Seq methods to analyze the transcriptome assembly for *M. falcata* roots under low-temperature stress and identify key functional and regulatory genes involved in low-temperature tolerance. Ultimately, we obtained 115,153 nonredundant sequences, and the average ROI was long enough to represent the full-length transcripts (Table 1). We also predicted a total of 8,849 AS events, 73,149 SSRs and 4,189 LncRNAs.

Changes in physiological indicators

Our results indicated the complexity of physiological changes in *M. falcata* in response to low-temperature stress. Under low-temperature stress, the roots displayed relatively more extensive changes in membrane, antioxidants and osmolytes. An increase in the contents of MDA, proline, soluble sugar and

electrolyte leakage has been demonstrated in cold-treated wheat and wild tomato [25, 47]. During low-temperature domestication, *M. falcata* accumulates more sucrose and proline and high sucrose phosphate synthase (SPs) and sucrose synthase activity [48]. In this paper, the increase in MDA, electrolyte leakage, proline and soluble sugar contents was observed in low-temperature-treated *M. falcata* roots, suggesting that osmolytes might protect plant cell membranes, increase membrane stabilization, and balance osmotic pressure during low-temperature-induced dehydration in *M. falcata*; additionally, electrolyte leakage can be used as a direct stress marker to reflect cell membrane damage by low-temperature stress (Fig. 1). Low-temperature stress induces the activities of cell apoptosis factors, and it has been proven that the injury in *M. falcata* under low temperature is related to proteins in cells [49]. The enzymatic antioxidant system is a protective mechanism employed to eliminate or reduce ROS and increase a plant's capacity for stress tolerance under low-temperature stress [50, 51]. The changes in CAT, POD, and GSH may play a key role in the detoxification of ROS induced by low temperature in *M. falcata* roots.

Gene expression of *M. falcata* roots in response to low-temperature stress

In this study, a total of 11,369 DEGs were identified as responsive to low-temperature stress at all temperature points, of which 68.8% were induced and 31.2% were repressed under low-temperature stress. All DEGs were grouped into six subclusters (Fig. 3B), and an enrichment analysis was conducted with the KEGG pathway. In the organismal systems category, the most enriched pathway was "Plant-pathogen interaction", indicating a basic plant immunological response in *M. falcata* roots under low-temperature stress [52]. In the category of environmental information processing, most DEGs were involved in the plant hormone signal transduction pathway. Our results demonstrated the critical role of phytohormones in plants in the response to external and internal cues to regulate growth and development [37]. In the genetic information processing category, the most enriched pathway was "Protein processing in endoplasmic reticulum". Under low-temperature stress, the membrane protein synthesis rate and membrane protein number increased in cold-adapted alfalfa [53]. Our data showed the facilitation and monitoring of proper folding by chaperone interactions and the formation of assemblies into multimeric proteins in the endoplasmic reticulum. In the category of metabolism, most DEGs were involved in "Carbon metabolism", followed by "Glycolysis/Gluconeogenesis" and "Starch and sucrose metabolism", suggesting that carbon and energy supply was very important for the adaptation of *M. falcata* to low temperatures. In the category of cellular processes, most DEGs functioned in as endocytosis. Endocytosis regulates the entry of membrane proteins, lipids, and extracellular molecules into the cell under adverse environmental conditions and plays a key role in alleviating ROS [54, 55].

Genes encoding TFs in response to low-temperature stress

TFs play an important role in cell function and development and directly regulate gene expression through interactions with themselves and proteins to participate in plant stress regulatory processes, including low-temperature-related processes. Many TFs, including bHLH, bZIP, MYB, C2H2, ERF, NAC and WRKY, that confer tolerance to low temperature to plants have been identified using transcriptomic

approaches [56]. In this study, 1,538 TF genes were differentially expressed between different temperature points and classified into 45 TF gene families. The most abundant TF family was WRKY, followed by ERF, MYB, bHLH and NAC. Our results are consistent with previous reports on TFs in plant low-temperature stress and suggest that the WRKY family plays a critical role in *M. falcata* low-temperature tolerance.

Identification of genes responsible for the response to low temperature

Plants have developed many mechanisms and pathways that enable them to minimize the negative effect of low temperature. Global analysis of stress-responsive genes facilitates the understanding of the plant response to low-temperature stress. In this paper, a total of 8,683 DEGs were identified by a comparison of each temperature point to the control environment, and 134 genes were differentially coexpressed at all five temperature points, including 101 upregulated genes and 33 downregulated genes (Fig. 4). However, only 7 genes were successfully annotated with GO, and two genes (PB40804 and PB75011) were enriched in all three terms. These two genes were both upregulated under low-temperature stress. PB40804 functions as a phenylalanine ammonia-lyase. Phenylalanine ammonia-lyase functions as a smart switch directly controlling the accumulation of calycosin and calycosin-7-O-beta-D-glucoside in *Astragalus membranaceus* plants during low-temperature treatment [57]. PB75011 functions as decarboxylase and is involved in cell wall/membrane/envelope biogenesis. Ornithine decarboxylase and arginine decarboxylase control the synthesis of polyamines in plants. The response of *Arabidopsis thaliana* to low-temperature stress emphasizes the involvement of transcriptional regulation in arginine decarboxylase gene expression [58]. In the COG annotation, we identified 18 enriched genes, including 15 upregulated genes and 3 downregulated genes. PB40804 was also annotated as “Amino acid transport and metabolism” in COG. The abundance of amino acid transporters is correlated with multitude fundamental roles in plant growth and development, and low-temperature stress could decrease the amino acid concentrations and alter their composition [59, 60]. “Circadian rhythm - plant” was the most enriched pathway in KEGG, and we found that PB110405, the gene with the most GO terms in the biological process category, was enriched in this pathway. PB110405 was annotated with 9 GO terms in the biological process category, including “regulation of transcription, DNA-templated”, “temperature compensation of the circadian clock”, “response to hydrogen peroxide”, “starch metabolic process” and “response to cold”. The “Circadian rhythm-plant” pathway suggested that ambient temperatures in *M. falcata* were substantially influenced by low temperature. PB108808, a putative ortholog of MYB-related TF LHY in *Arabidopsis*, is a gene that is induced by low temperature and that indicates the presence of interplay between circadian rhythm and the response to low temperature in *M. falcata.* The following pathway was “Arginine and proline metabolism”. Arginine and proline metabolism is one of the central pathways for the biosynthesis of the amino acids arginine and proline [27]. Proline accumulation is a well-known measure for alleviating abiotic stress in plants [61]. Combined with the changes in proline contents under low-temperature stress, our results indicated that osmotic regulatory substances, protective protein molecules in *M. falcata*, play important roles in the response to low-temperature stress, and the composition of aromatic compounds may change under low-temperature stress.

Identification of genetic modules corresponding to low-temperature stress

WGCNA, known as gene coexpression network analysis, is a systematic biological method that can be applied to the study of biological processes with multiple sources [62]. It has been proven that WGCNA can be an efficient data mining method, specifically for screening genes related to traits and for conducting modular classification to obtain coexpression modules with high biological significance [63]. In this paper, 8,683 DEGs were identified by comparing low-temperature-treated and control samples and were clustered into 12 gene modules after merging modules with similar expression patterns (Fig. 5). It was found that the MEbrown module was significantly correlated with low-temperature stress in *M. falcata*. GO enrichment analysis of MEbrown showed that regulatory pathways with biological significance could be obtained in this module. For example, we enriched “cold-response pathways”, “response to stress” and “intracellular signal transduction” terms in GO annotation. COG classification illustrated that the MEbrown module was enriched in many DEGs in general function prediction only and the signal transduction mechanism. KEGG pathway enrichments revealed that there were many DEGs involved in starch and sucrose metabolism. We also found that electrolyte leakage was correlated with more genetic modules than other physiological indicators, which corroborated our finding in physiological assays that electrolyte leakage can be used as a direct stress marker to reflect cell membrane damage from low-temperature stress.

Conclusions

Taken together, we explored the transcriptomic changes in low-temperature-treated *M. falcata* roots by combining SMRT and NGS technologies. A total of 115,153 nonredundant sequences were obtained, and 8,849 AS events, 73,149 SSRs and 4,189 LncRNAs were predicted. A total of 111,587 genes from SMRT were annotated, and 11,369 DEGs were identified in this paper that are involved in plant hormone signal transduction, protein processing in the endoplasmic reticulum, carbon metabolism, glycolysis/gluconeogenesis, starch and sucrose metabolism, and endocytosis pathways. We characterized 1,538 TF genes into 45 TF gene families, and the most abundant TF family was WRKY, followed by ERF, MYB, bHLH and NAC. A total of 134 genes were differentially coexpressed at all five temperature points, including 101 upregulated genes and 33 downregulated genes. PB40804, PB75011, PB110405 and PB108808 were found to play crucial roles in the tolerance of *M. falcata* to low temperature. The WGCNA results showed that the MEbrown module was significantly correlated with low-temperature stress in *M. falcata*. Electrolyte leakage was correlated with most genetic modules and corroborated that electrolyte leakage can be used as a direct stress marker to reflect cell membrane damage from low-temperature stress in physiological assays. These findings provide a complete characterization of gene transcription and facilitate the understanding of the mechanisms of tolerance to low temperature in *M. falcata.*

Methods

Plant cultivation and low-temperature treatment

Seeds of *M. falcata* L. cv. Hulunbuir, a registered grass variety approved by National Grass Variety Approval Committee of China, were obtained from Hulunbuir Grassland Station in Inner Mongolia province of China, and disinfected with 5% sodium hypochlorite solution for 5 minutes and then washed in distilled water. Then, the seeds were germinated on wet filter paper in culture dishes in a dark growth chamber at 25 °C. Five-day-old seedlings were transplanted to plastic pots filled with a mixture of vermiculite, pearlite and peat soil (1:1:1) in the greenhouse with an average temperature of 25 °C and 20 °C and a relative humidity of 55% and 70% during the day and night, respectively. All seedlings were watered with 1/2 strength Hoagland [64] nutrient solution every two days. Ninety days after transplantation, the uniform seedlings were transported to the growth chamber for the low-temperature treatment. The treatment temperatures were 4 °C, 0 °C, -5 °C, -10 °C and -15 °C. The normal environment temperature was used as a control. For temperatures below 0 °C (freezing damage), *M. falcata* seedlings were acclimated for 2 days at 4 °C and exposed to low-temperature stress. The hypothermia schedule was a decrease of 1 °C every 1 h from 4 °C, and the low-temperature stress treatment was conducted at each studied temperature for 2 h. The roots were harvested, immediately frozen in liquid nitrogen and stored at -80 °C for laboratory analysis.

Physiological assays of low-temperature-treated *M. falcata* roots

Leaf cell membrane damage was determined as electrolyte leakage [65]. MDA was determined using a modified thiobarbituric acid (TBA) method [66]. The activity of SOD was measured by the nitroblue tetrazolium (NBT) method [67]. The activity of CAT was measured following the method of Maehly and Chance [68]. The activity of POD was measured according to the method of Zaharieva et al. [69]. The superoxide anion radical ($O_2^{\cdot-}$) contents were determined following Elstner's method [70]. The soluble protein content was determined according to the Bradford method [71]. The content of reduced glutathione (GSH) was fluorometrically estimated [72]. The proline content was determined by the ninhydrin method [73]. The soluble sugar content was determined following the method of Dreywood [74].

All assays described above were repeated four times on four biological replicates. The data shown as the mean ± SD were subjected to ANOVA to determine significant differences. The least significant differences (LSD) of means were determined using Duncan's test at the level of significance defined as $\alpha=0.05$.

RNA isolation, library preparation and sequencing

Total RNA from each sample was isolated using the TRIzol Reagent (Invitrogen, USA) according to the manufacturer's protocol, and genomic DNA was removed via digestion with DNase I (TaKaRa, Japan). Then, the purity, concentration and nucleic acid absorption peak were measured with a Nanodrop ND-1000 spectrophotometer (NanoDrop, USA). RNA integrity was detected by an Agilent 2100 Bioanalyzer (Agilent, USA). Genomic DNA contamination was detected by electrophoresis.

The library was prepared after the samples passed the quality tests. For Illumina cDNA library preparation, 20 μg of total RNA from each pool was enriched with magnetic beads with oligo (dT) and randomly interrupted by adding fragmentation buffer. Then, first-strand cDNA was synthesized with random hexamers by using mRNA as a template. Second-strand cDNA was synthesized after adding buffer, dNTPs, RNase H and DNA polymerase I. The cDNA was purified with AMPure XP beads. The purified double-stranded cDNA was subjected to end repair, the addition of a poly-A tail and ligation with the sequencing linker, and the fragment size was selected using AMPure XP beads. Finally, the cDNA library was prepared by PCR enrichment.

For PacBio Iso-Seq library preparation, the cDNA was synthesized by using the SMARTer™ PCR cDNA Synthesis Kit (TaKaRa, Japan). cDNA libraries of different sizes were generated using BluePippin. Then, the screened cDNA was amplified by PCR, end-repaired, connected to the SMRT dumbbell type connector, and exonuclease digested. Finally, the library was prepared after a secondary screening using BluePippin. A total of eight SMRT cells were used for the three libraries at three size ranges: 1-2 kb, 2-3 kb, and 3-6 kb.

After the accurate quantification of libraries using Qubit 2.0 and the library sizes were qualified using Agilent 2100, the libraries were sequenced using PacBio RS II with 8 SMRT cells and an Illumina HiSeq 2500 platform in the Biomarker Institute (Biomarker, China). The 1-2 kb, 2-3 kb and 3-6 kb libraries were sequenced with 3, 3 and 2 SMRT cells, respectively.

Quality filtering and transcriptome assembly

Raw reads were processed into error-corrected ROIs using the ToFu pipeline with full passes>=0, and the accuracy of the sequence was greater than 0.75 (https://github.com/PacificBiosciences/cDNA_primer/wiki/Understanding-PacBio-transcriptome-data#readexplained). High-quality clean data were obtained by removing reads containing connectors, low-quality reads (including those with N removal ratio greater than 10%, and reads where the number of bases with mass value Q≤10 accounted for more than 50% of the reads). Next, FLNC transcripts were determined by searching for the poly-A tail signal and the 5’ and 3’ cDNA primers in ROIs. Iterative Clustering for Error Correction (ICE) was used to obtain consensus isoforms, and full-length consensus sequences from ICE were polished using SMRT Analysis (v2.3.0). Full-length transcripts with postcorrection accuracy above 99% were generated for further analysis. Redundant reads were removed from the Iso-Seq high-quality full-length transcripts using CD-HIT (identity > 0.99). The resulting transcript sequence was directly used for subsequent analysis of AS events, SSRs and LncRNAs. The second-generation data were used to quantify and differentially analyze the new coding sequence (CDS).

Identification of AS events, SSRs, LncRNAs and CDSs

We used Iso-SeqTM data directly to run all-vs-all BLAST with high-identity settings; BLAST alignments that met all criteria were considered products of candidate AS events: There were two high-scoring segment pairs (HSPs) in the alignment; the two HSPs had the same forward/reverse direction, within the same alignment, one sequence was continuous, or with a small "Overlap" size (smaller than 5 bp); the

other one was distinct to show an "AS Gap", and the continuous sequence mostly completely aligned to the distinct sequence. The AS gap was larger than 100 bp and at least 100 bp away from the 3'/5' end. SSRs of the transcriptome were identified using MISA (http://pgrc.ipk-gatersleben.de/misa/). Transcripts with lengths greater than 200 nt and with more than two exons were selected as LncRNA candidates and further screened using CPC/CNCI/CPAT/Pfam, which have the power to distinguish the protein-coding genes from the noncoding genes. TransDecoder (https://github.com/TransDecoder/TransDecoder/releases) was used to identify CDS regions within transcript sequences.

Gene function annotation

To acquire the most comprehensive annotation, all full-length transcripts from SMRT were aligned with the NCBI NR, SwissProt, GO, COG, KOG, Pfam, and KEGG databases using BLAST software (version 2.2.26) [45]. GO enrichment analysis was implemented by the GOseq R package-based on Wallenius noncentral hypergeometric distribution [75]. KOBAS software was used to test the statistical enrichment of DEGs in KEGG pathways [76].

Quantification of gene expression levels

The gene expression level of each sample was identified by RSEM [46]. Clean data were mapped back onto the assembled transcriptome, and readcount for each gene was obtained from the mapping results. Bowtie2 software was used to compare the clean data from Illumina sequencing to the SMRT sequencing data. Quantification of gene expression levels was estimated by FPKM considering the effect of the sequence depth and gene length on the fragments.

Identification of DEGs

Differential expression analysis was performed using the DESeq R package (1.10.1) to identify DEGs between the low-temperature-treated and control samples and samples collected at different temperature points [77]. DESeq provides statistical analyses for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. In the detection of DEGs, fold change≥2 and FDR<0.01 were taken as the screening criteria. The fold change represents the ratio of expression amount between two samples (groups). The FDR was obtained by correcting the p-value of difference significance. Because the differences in the transcriptome sequencing expression analysis is the transcribed expression values of a large number of independent statistical hypothesis tests, false positives are a concern; thus, in the process of analyzing DEGs, the recognized Benjamini-Hochberg correction method of hypothesis testing with the original significance *p* values for correction and, eventually, the FDR were used as key indicators of screening DEGs.

Identification of putative TFs

All DEGs were BLAST searched to a plant TF database (PlantTFDB4.0, http://planttfdb.cbi.pku.edu.cn) to identify putative TFs. TF information was annotated based on the comparison results.

Coexpression network analysis with WGCNA

Coexpression networks were constructed using the WGCNA package in R from all DEGs [62]. The modules were obtained using the automatic network construction function blockwise modules with default settings. The eigengene value was calculated for each module and used to test the association with each physiological index. The total connectivity and intramodular connectivity (function soft connectivity), kME (for modular membership, also known as eigengene-based connectivity), and kME-P values were calculated for the DEGs.

Validation of RNA-Seq data by qRT-PCR

RNA samples isolated above were used as templates and reverse transcribed with the HiScript II Q Select RT SuperMix for qPCR (gDNA eraser) kit (Vazyme, China). Primers used in this study were designed using Primer 5 with RefSeq and listed in Additional file 15: Table S11. The expression of the *β-actin* gene was used as the internal control. qRT-PCR was performed using ChamQ^{TM} Universal SYBR qPCR Master Mix (Vazyme, China) on a LightCycler480 II58 (Roche, Switzerland) following the manufacturer's protocol. Relative gene expression levels were evaluated using the $2^{-\Delta\Delta CT}$ method [78].

List Of Abbreviations

DEG: Differentially expressed genes; FPKM: Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; COG: Clusters of Orthologous Groups; KOG: EuKaryotic Ortholog Groups; NR: NCBI non-redundant protein; Pfam: Protein family; qRT-PCR: Quantitative real-time PCR; ROS: Reactive oxygen species; RSEM: RNA-Seq by Expectation Maximization; NGS: Next-generation sequencing; SMRT: Single-molecule real-time; MDA: Malondialdehyde; SOD: Superoxide dismutase; POD: Peroxidase; CAT: catalase; GSH: reduced glutathione; O2.-: superoxide anion radical; SPs: sucrose phosohate synthase; TF: Transcription factor; NGS: Next-generation sequencing; AS: alternative splicings; SSRs: simple sequence; CPC: Coding Potential Calculator; CNCI: Coding-Non-Coding Index; CPAT: Coding Potential Assessment Tool; FDR: False Discovery Rate; ROI: Reads of Insert; ICE: Iterative Clustering for Error Correction; FLNC: full length non-chimeric read; WGCNA: coexpression network analysis.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

All relevant supplementary data are provided within this manuscript as Additional files. The PacBio SMRT reads and the Illumina SGS reads generated in this study were submitted to the NCBI Sequence Read Archive under accession numbers BioProject PRJNA 549099 and 520970. Address is as follows: http://www.ncbi.nlm.nih.gov/sra.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was financially supported by the National Science Foundation of China (31502007), the Young Talents Project of Northeast Agricultural University (16QC23), the Special Funds for Science and Technology Innovation Talent Research in Harbin (2017RAQXJ008), and the Project of Nature Scientific Foundation of Heilongjiang Province, China (QC2015019). Each of the funding bodies granted the funds based on a research proposal. They had no influence over the experiment design, data analysis or interpretation, or writing the manuscript.

Authors' contributions

GC, HC and PZ designed the research. PZ and HC wrote the paper with contributions and discussion from all of the coauthors. HC, HY, MY, GH, MG and RY conducted the research. All authors have read and approved the manuscript.

Acknowledgements

We thank the Biomarker Corporation (Beijing, China) for SMRT and RNA-Seq sequencing and raw data analysis.

References

1. Jin J, Kim J: **Cold stress signaling networks in Arabidopsis**. *Journal of Plant Biology* 2013, **56**(2):69-76.
2. Zhou MQ, Shen C, Wu LH, Tang KX, Lin J: **CBF-dependent signaling pathway: A key responder to low temperature stress in plants**. *Crit Rev Biotechnol* 2011, **31**(2):186-192.
3. Wong CE, Li Y, Labbe A, Guevara D, Nuin P, Whitty B, Diaz C, Golding GB, Gray GR, Weretilnyk EA *et al*: **Transcriptional profiling implicates novel interactions between abiotic stress and hormonal responses in Thellungiella, a close relative of Arabidopsis**. *Plant Physiol* 2006, **140**(4):1437-1450.
4. Bg DLR, Yun SJ, Herath V, Xu F, Park MR, Lee JI, Kim KY: **Phenotypic, physiological, and molecular evaluation of rice chilling stress response at the vegetative stage**. *Methods Mol Biol* 2013, **956**:227-241.

5. Gehan MA, Greenham K, Mockler TC, Mcclung CR: **Transcriptional networks — crops, clocks, and abiotic stress**. *Curr Opin Plant Biol* 2015, **24**:39-46.
6. Janska A, Marsik P, Zelenkova S, Ovesna J: **Cold stress and acclimation - what is important for metabolic adjustment?** *Plant Biol* 2010, **12**(3):395-405.
7. Majlath I, Darko E, Palla B, Nagy Z, Janda T, Szalai G: **Reduced light and moderate water deficiency sustain nitrogen assimilation and sucrose degradation at low temperature in durum wheat**. *J Plant Physiol* 2016, **191**:149-158.
8. Hu Z, Fan J, Xie Y, Amombo E, Liu A, Gitau MM, Khaldun ABM, Chen L, Fu J: **Comparative photosynthetic and metabolic analyses reveal mechanism of improved cold stress tolerance in bermudagrass by exogenous melatonin**. *Plant Physiol Biochem* 2016, **100**:94-104.
9. Zheng YL, Li WQ, Sun WB: **Effects of acclimation and pretreatment with abscisic acid or salicylic acid on tolerance of Trigonobalanus doichangensis to extreme temperatures**. *Biol Plant* 2015, **59**(2):382-388.
10. Theocharis A, Clement C, Barka EA: **Physiological and molecular changes in plants grown at low temperatures**. *Planta* 2012, **235**(6):1091-1105.
11. Trischuk RG, Schilling BS, Low NH, Gray GR, Gusta LV: **Cold acclimation, de-acclimation and re-acclimation of spring canola, winter canola and winter wheat: The role of carbohydrates, cold-induced stress proteins and vernalization**. *Environmental & Experimental Botany* 2014, **106**(1):156-163.
12. Li XD, Zhuang KY, Liu ZM, Yang DY, Ma NN, Meng QW: **Overexpression of a novel NAC-type tomato transcription factor, SlNAM1, enhances the chilling stress tolerance of transgenic tobacco**. *J Plant Physiol* 2016, **204**:54-65.
13. Zhu JK: **Abiotic Stress Signaling and Responses in Plants**. *Cell* 2016, **167**(2):313-324.
14. Shi YT, Ding YL, Yang SH: **Cold Signal Transduction and its Interplay with Phytohormones During Cold Acclimation**. *Plant and Cell Physiology* 2015, **56**(1):7-15.
15. Ma Y, Dai XY, Xu YY, Luo W, Zheng XM, Zeng DL, Pan YJ, Lin XL, Liu HH, Zhang DJ *et al*: **COLD1 Confers Chilling Tolerance in Rice**. *Cell* 2015, **160**(6):1209-1221.
16. Chinnusamy V, Zhu J, Zhu J-K: **Cold stress regulation of gene expression in plants**. *Trends Plant Sci* 2007, **12**(10):444-451.
17. Jagloottosen KR, Gilmour SJ, Zarka DG, Schabenberger O, Thomashow MF: **Arabidopsis CBF1 overexpression induces COR genes and enhances freezing tolerance**. *Science* 1998, **280**(5360):104-106.
18. Jeknić Z, Pillman KA, Dhillon T, Skinner JS, Veisz O, Cuesta-Marcos A, Hayes PM, Jacobs AK, Chen TH, Stockinger EJ: **Hv-CBF2A overexpression in barley accelerates COR gene transcript accumulation and acquisition of freezing tolerance during cold acclimation**. *Plant Mol Biol* 2014, **84**(1-2):67-82.
19. Vanburen R, Bryant D, Edger PP, Tang H, Burgess D, Challabathula D, Spittle K, Hall R, Gu J, Lyons E: **Single-molecule sequencing of the desiccation-tolerant grass Oropetium thomaeum**. *Nature* 2015, **527**(7579):508.

20. Lan T, Renner T, Ibarra-Laclette E, Farr KM, Chang TH, Cervantes-Pérez SA, Zheng C, Sankoff D, Tang H, Purbojati RW: **Long-read sequencing uncovers the adaptive topography of a carnivorous plant genome**. *Proceedings of the National Academy of Sciences of the United States of America* 2017(3):201702072.
21. Chaisson MJ, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, Antonacci F, Surti U, Sandstrom R, Boitano M: **Resolving the complexity of the human genome using single-molecule sequencing**. *Nature* 2014, **517**(7536):608-611.
22. Chen X, Bracht JR, Goldman AD, Dolzhenko E, Clay DM, Swart EC, Perlman DH, Doak TG, Stuart A, Amemiya CT: **The architecture of a scrambled genome reveals massive levels of genomic rearrangement during development**. *Cell* 2014, **158**(5):1187-1198.
23. Zhang T, Zhao X, Wang W, Pan Y, Huang L, Liu X, Zong Y, Zhu L, Yang D, Fu B: **Comparative transcriptome profiling of chilling stress responsiveness in two contrasting rice genotypes**. *Plos One* 2012, **7**(8):e43274.
24. Karki A, Horvath DP, Sutton F: **Induction of DREB2A pathway with repression of E2F, jasmonic acid biosynthetic and photosynthesis pathways in cold acclimation-specific freeze-resistant wheat crown**. *Functional & Integrative Genomics* 2013, **13**(1):57-65.
25. Xu W, Li R, Zhang N, Ma F, Jiao Y, Wang Z: **Transcriptome profiling of Vitis amurensis, an extremely cold-tolerant Chinese wild Vitis species, reveals candidate genes and events that potentially connected to cold stress**. *Plant Molecular Biology* 2014, **86**(4-5):527-541.
26. Xu ZC, Peters RJ, Weirather J, Luo HM, Liao BS, Zhang X, Zhu YJ, Ji AJ, Zhang B, Hu SN *et al*: **Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of Salvia miltiorrhiza and tanshinone biosynthesis**. *Plant J* 2015, **82**(6):951-961.
27. Yang L, Jin Y, Huang W, Sun Q, Liu F, Huang X: **Full-length transcriptome sequences of ephemeral plant Arabidopsis pumila provides insight into gene expression dynamics during continuous salt stress**. *BMC Genomics* 2018.
28. Chao Y, Yuan J, Li S, Jia S, Han L, Xu L: **Analysis of transcripts and splice isoforms in red clover (Trifolium pratense L.) by single-molecule long-read sequencing**. *BMC Plant Biol* 2018.
29. Yang Y, Yu Q, Yang Y, Su Y, Ahmad W, Guo J, Gao S, Xu L, Que Y: **Identification of cold-related miRNAs in sugarcane by small RNA sequencing and functional analysis of a cold inducible ScmiR393 to cold stress**. *Environ Exp Bot* 2018, **155**:464-476.
30. Shen C, Li D, He R, Fang Z, Xia Y, Gao J, Shen H, Cao M: **Comparative transcriptome analysis of RNA-seq data for cold-tolerant and cold-sensitive rice genotypes under cold stress**. *Journal of Plant Biology* 2014, **57**(6):337-348.
31. Zhang LL, Zhao MG, Tian QY, Zhang WH: **Comparative studies on tolerance of Medicago truncatula and Medicago falcata to freezing**. *Planta* 2011, **234**(3):445-457.
32. Riday H, Brummer EC: **Forage Yield Heterosis in Alfalfa**. *Crop Sci* 2002, **42**(3):716-723.

33. Gréard C, Barre P, Flajoulot S, Santoni S, Julier B: **Sequence diversity of five Medicago sativa genes involved in agronomic traits to set up allele mining in breeding**. *Molecular Breeding* 2018, **38**(12).
34. Tan J, Zhuo C, Guo Z: **Nitric oxide mediates cold- and dehydration-induced expression of a novel *MfHyPRP* that confers tolerance to abiotic stress**. *Physiol Plant* 2013, **149**(3):310-320.
35. Zhang LL, Zhao MG, Tian QY, Zhang WH: **Comparative studies on tolerance of Medicago truncatula and Medicago falcata to freezing**. *Planta* 2011, **234**(3):445-457.
36. Qu YT, Duan M, Zhang ZQ, Dong JL, Wang T: **Overexpression of the Medicago falcata NAC transcription factor MfNAC3 enhances cold tolerance in Medicago truncatula**. *Environ Exp Bot* 2016, **129**:67-76.
37. Miao ZY, Xu W, Li DF, Hu XN, Liu JX, Zhang RX, Tong ZY, Dong JL, Su Z, Zhang LW *et al*: **De novo transcriptome analysis of Medicago falcata reveals novel insights about the mechanisms underlying abiotic stress-responsive pathway**. *BMC Genomics* 2015, **16**.
38. Liu M, Wang TZ, Zhang WH: **Sodium extrusion associated with enhanced expression of SOS1 underlies different salt tolerance between Medicago falcata and Medicago truncatula seedlings**. *Environ Exp Bot* 2015, **110**:46-55.
39. Hackl T, Hedrich R, Schultz J, Förster F: **proovread: large-scale high-accuracy PacBio correction through iterative short read consensus**. *Bioinformatics* 2014, **30**(21):3004.
40. Li W, Godzik A: **Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences**. *Bioinformatics* 2006, **22**(13):1658.
41. Liu X, Mei W, Soltis PS, Soltis DE, Barbazuk WB: **Detecting alternatively spliced transcript isoforms from single-molecule long-read sequences without a reference genome**. *Molecular ecology resources* 2017, **17**(6):1243-1256.
42. Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L, Gao G: **CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine**. *Nucleic acids research* 2007, **35**(Web Server issue):W345-W349.
43. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, Liu Y, Chen R, Zhao Y: **Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts**. *Nucleic acids research* 2013, **41**(17):e166-e166.
44. Wang L, Park HJ, Dasari S, Wang S, Kocher J-P, Li W: **CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model**. *Nucleic acids research* 2013, **41**(6):e74-e74.
45. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic acids research* 1997, **25**(17):3389-3402.
46. Li B, Dewey CN: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome**. *BMC Bioinformatics* 2011, **12**:323.
47. Chen H, Chen X, Chen D, Li J, Zhang Y, Wang A: **A comparison of the low temperature transcriptomes of two tomato genotypes that differ in freezing tolerance: *Solanum lycopersicum* and *Solanum habrochaites***. *BMC Plant Biol* 2015, **15**(1):132.

48. Tondelli A, Francia E, Barabaschi D, Pasquariello M, Pecchioni N: **Inside the CBF locus in Poaceae**. *Plant Sci* 2011, **180**(1):39-45.
49. Dubé M-P, Castonguay Y, Cloutier J, Michaud J, Bertrand A: **Characterization of two novel cold-inducible K3 dehydrin genes from alfalfa (*Medicago sativa* spp. sativa L.)**. *Theor Appl Genet* 2013, **126**(3):823-835.
50. Jithesh MN, Prashanth SR, Sivaprakash KR, Parida AK: **Antioxidative response mechanisms in halophytes: Their role in stress defence**. *Journal of Genetics* 2006, **85**(3):237-254.
51. Mittler R, Vanderauwera S, Gollery M, Van BF: **Reactive oxygen gene network of plants**. *Trends in Plant Science* 2004, **9**(10):490-498.
52. Huot B, Castroverde CDM, Velásquez AC, Hubbard E, Pulman JA, Yao J, Childs KL, Tsuda K, Montgomery BL, He SY: **Dual impact of elevated temperature on plant defence and bacterial virulence in Arabidopsis**. *Nature Communications* 2017, **8**(1):1808.
53. Mohapatra SS, Wolfraim L, Poole RJ, Dhindsa RS: **Molecular cloning and relationship to freezing tolerance of cold-acclimation-specific genes of alfalfa**. *Plant Physiol* 1989, **89**(1):375-380.
54. Fan LS, Li RL, Pan JW, Ding ZJ, Lin JX: **Endocytosis and its regulation in plants**. *Trends Plant Sci* 2015, **20**(6):388-397.
55. Leborgne-Castel N, Luu DT: **Regulation of endocytosis by external stimuli in plant cells**. *Plant Biosystems* 2009, **143**(3):630-635.
56. Rihan HZ, Al-Issawi M, Fuller MP: **Advances in physiological and molecular aspects of plant cold tolerance**. *Journal of Plant Interactions* 2017, **12**(1):143-157.
57. Pan HY, Wang YG, Zhang YF, Zhou TS, Fang CM, Nan P, Wang XQ, Li XB, Wei YL, Chen JK: **Phenylalanine ammonia lyase functions as a switch directly controlling the accumulation of calycosin and calycosin-7-O-beta-D-glucoside in Astragalus membranaceus var. mongholicus plants**. *J Exp Bot* 2008, **59**(11):3027-3037.
58. Hummel I, Bourdais G, Gouesbet G, Couee I, Malmberg RL, El Amrani A: **Differential gene expression of ARGININE DECARBOXYLASE ADC1 and ADC2 in Arabidopsis thaliana: characterization of transcriptional regulation during seed germination and seedling development**. *New Phytol* 2004, **163**(3):519-531.
59. Liu X, Bush DR: **Expression and transcriptional regulation of amino acid transporters in plants**. *Amino Acids* 2006, **30**(2):113-120.
60. Zhu XC, Song FB, Liu FL: **Altered amino acid profile of arbuscular mycorrhizal maize plants under low temperature stress**. *J Plant Nutr Soil Sci* 2016, **179**(2):186-189.
61. Saxena SC, Kaur H, Verma P, Petla BP, Majee M: **Osmoprotectants: Potential for Crop Improvement Under Adverse Conditions**; 2013.
62. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis**. *BMC Bioinformatics* 2008, **9**(1):559.

63. Zhao W, Langfelder P, Fuller T, Dong J, Hovarth S: **Weighted Gene Coexpression Network Analysis: State of the Art**. *J Biopharm Stat* 2010, **20**(2):281-300.
64. Hoagland D, Arnon D: **The water culture method for growing plants without soil**. *California Agr Expt S* 1950, **347**(1):32.
65. Peever TL, Higgins VJ: **Electrolyte Leakage, Lipoxygenase, and Lipid Peroxidation Induced in Tomato Leaf Tissue by Specific and Nonspecific Elicitors from Cladosporium fulvum**. *Plant Physiol* 1989, **90**(3):867-875.
66. Wang WB, Kim YH, Lee HS, Kim KY, Deng XP, Kwak SS: **Analysis of antioxidant enzyme activity during germination of alfalfa under salt and drought stresses**. *Plant Physiol Biochem* 2009, **47**(7):570-577.
67. Giannopolitis CN, Ries SK: **Superoxide dismutases: I. Occurrence in higher plants**. *Plant Physiol* 1977, **59**(2):309-314.
68. Maehly AC: **The assay of catalases and peroxidases**. In: *Methods Biochem Anal.* John Wiley & Sons, Inc.; 2006: 357-424.
69. Zaharieva T, Yamashita K, Matsumoto H: **Iron Deficiency Induced Changes in Ascorbate Content and Enzyme Activities Related to Ascorbate Metabolism in Cucumber Roots**. *Plant Cell Physiol* 1999, **40**(3):273-280.
70. Elstner E, Heupel A: **Inhibition of nitrite formation from hydroxylammoniumchloride: a simple assay for superoxide dismutase**. *Anal Biochem* 1976, **70**(2):616-620.
71. Bradford MM: **A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding**. *Anal Biochem* 1976, **72**(1-2):248-254.
72. Hissin PJ, Hilf R: **A fluorometric method for determination of oxidized and reduced glutathione in tissues**. *Anal Biochem* 1976, **74**(1):214-226.
73. Bates LS, Waldren RP, Teare ID: **Rapid determination of free proline for water-stress studies**. *Plant Soil* 1973, **39**(1):205-207.
74. Dreywood R: **Qualitative test for carbohydrate material**. *Ind Eng Chem, Anal Ed* 1946, **18**(8):499-499.
75. Young MD, Wakefield MJ, Smyth GK, Oshlack A: **Gene ontology analysis for RNA-seq: accounting for selection bias**. *Genome Biol* 2010, **11**(2):R14-R14.
76. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li C-Y, Wei L: **KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases**. *Nucleic Acids Research* 2011, **39**(Web Server issue):316-322.
77. Anders S, Huber W: **Differential expression analysis for sequence count data**. *Genome Biol* 2010, **11**(10):R106-R106.
78. Livak KJ, Schmittgen TD: **Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2 ΔΔC T Method**. *METHODS* 2001, **25**:402-408.

Table 1

Table 1 Statistics of SMRT sequencing data

cDNA size	1-2 kb	2-3 kb	> 3 kb	All
SMRT cells	3	3	2	8
Polymerase reads	450,876	450,876	300,584	1,202,336
Postfilter number of subreads	4,761,465	2,481,575	1,185,345	8,428,385
Reads of insert	240,441	189,779	122,598	552,818
Number of five prime reads	122,516	120,763	84,001	327,280
Number of three prime reads	139,768	126,512	87,086	353,366
Number of poly-A reads	133,709	124,866	86,112	344,687
Number of filtered short reads	42,363	11,932	2,853	57,148
Number of nonfull-length reads	100,142	75,634	47,543	223,319
Number of full-length reads	97,936	102,213	72,202	272,351
Number of full-length nonchimeric reads	97,156	101,959	71,635	270,750
Average full-length nonchimeric read length	1,309	2,282	3,536	2,264
Full-length percentage (FL%)	40.73%	53.86%	58.89%	49.27%
Artificial concatemers (%)	0.80%	0.25%	0.79%	0.59%

Additional Files

Additional file 1: Table S1. Overview of the quality of the sequence data obtained by SGS sequencing.

Additional file 2: Table S2. Statistics of alternative splicing. The QueryName and SubjectName were the ID for the identified alternative splicing events. The QhspStart1, QhspEnd1, QhspStart2, QhspEnd2, ShspStart, ShspEnd1, ShspStart2 and ShspEnd2 were the start and end positions of the 2 HSPs for these two alternative splicing events, respectively.

Additional file 3: Table S3 Statistics of SSRs.

Additional file 4: Figure S1. Venn diagram of LncRNAs by CPC, CNCI, CPAT and Pfam.

Additional file 5: Table S4. Statistics of annotated transcripts.

Additional file 6: Figure S2. Nr homologous species distribution.

Additional file 7: Table S5. Summary of reads from the RNA-Seq data and their matches with the full-length transcription sequences.

Additional file 8: Table S6. The FPKMs and function annotations of all transcripts.

Additional file 9: Table S7. Information on the 11,369 DEGs under low-temperature stress.

Additional file 10: Table S8. Summary of the 1,538 transcription factors in response to low-temperature stress.

Additional file 11: Table S9. Summary of coexpressed genes between low-temperature-treated and control samples.

Additional file 12: Table S10. The genetic modules and kME values of genes in the WGCNA results.

Additional file 13: Figure S3. Functional analysis of genes in the MEbrown module. A, COG function classification. B, GO classification. C, KEGG enrichment.

Additional file 14: Figure S4. qRT-PCR assays of genes selected for RNA-Seq sequencing data confirmation. The blue line represents the normalized expression (log_{10}(FPK+1)) of RNA-Seq shown on the Y-axis to the left. The red line represents the relative qRT-PCR expression level shown on the Y-axis to the right.

Additional file 15: Table S11. List of primers used in this paper.

Figures

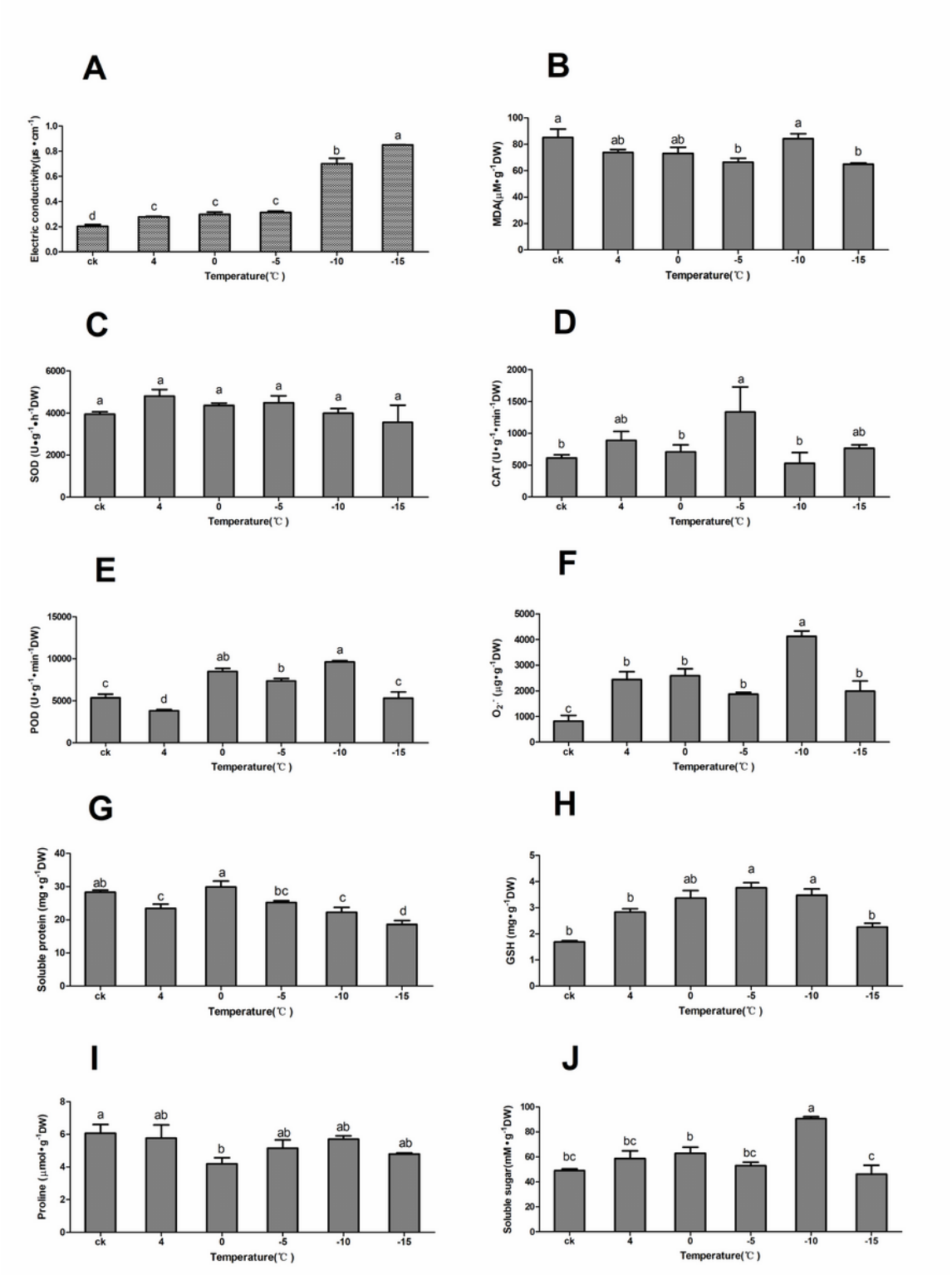

Figure 1

Determination of physiological indices of M. falcata roots under low-temperature stress. A, Electrolyte leakage. B, MDA content. C, SOD activity. D, CAT activity. E, POD activity. F, Superoxide anion radical (O2.-) content. G, Soluble protein content. H, GSH content. I, Proline content. J, Soluble sugar content. Data are shown as the means ± SD of four independent experiments. Different letters represent statistically significant differences as determined by one-way ANOVA ($p < 0.05$, Duncan's multiple range test).

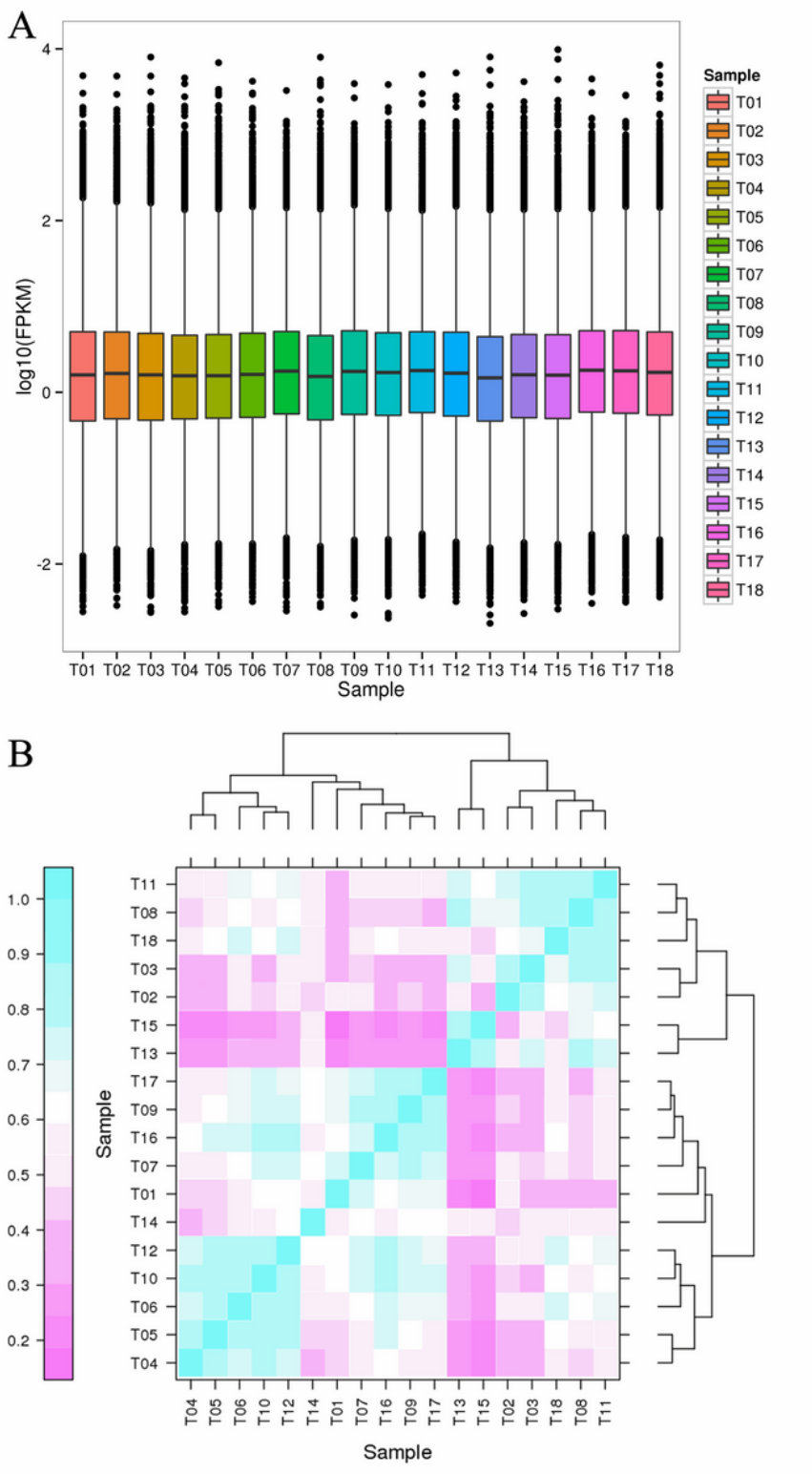

Figure 2

Comparison of gene expression levels under low-temperature stress. (A) Boxplot showing the distribution of FPKM values of each sample under low-temperature stress. The X-axis in the boxplot was the ID of each sample. The Y-axis was the log10 (FPKM). (B) The heat map of the Pearson correlation coefficient of each sample under low-temperature stress.

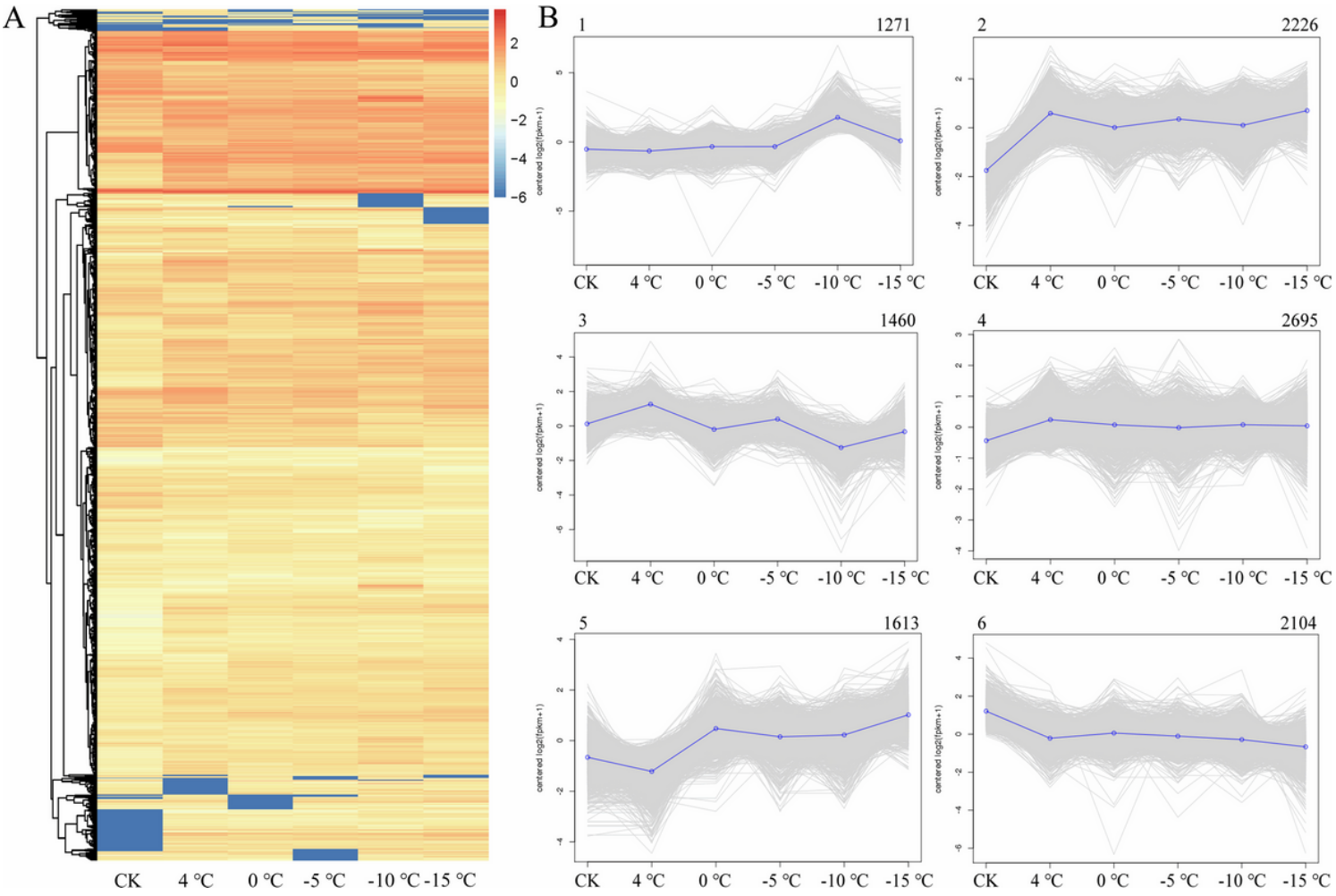

Figure 3

Clustering analysis of the DEGs. A, Hierarchical clustering of the 11,369 DEGs based on the averaged log2 (FPKM+1) values of all genes in each cluster. B, The six subclusters of 11,369 DEGs were clustered. The number of genes in each subcluster is shown at the top of the subcluster. The blue line shows the average values for relative expression levels in each subcluster, and the gray lines represent the relative gene expression levels of each gene in each subcluster.

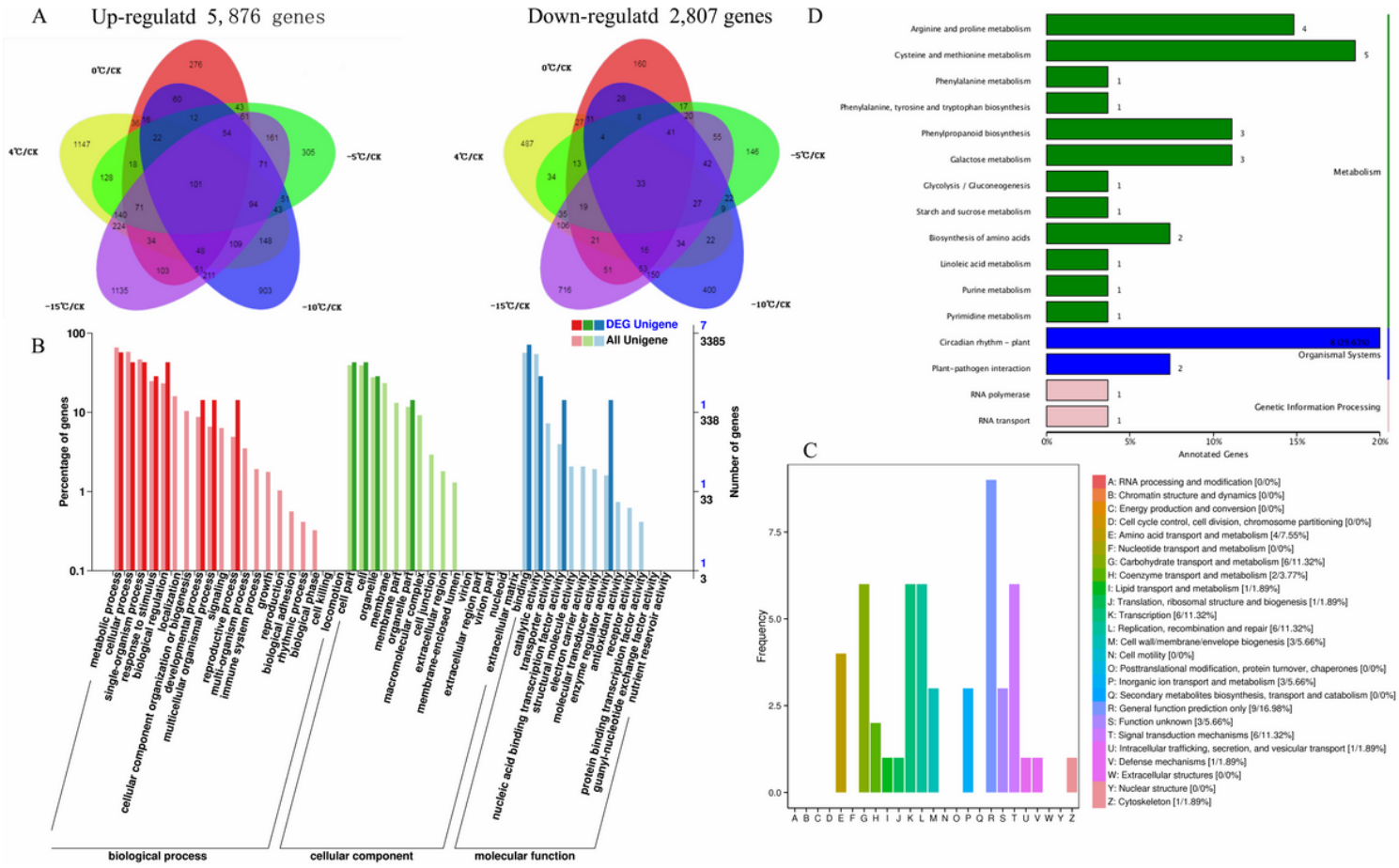

Figure 4

Summary of coexpressed genes between low-temperature-treated and control samples. A, Venn diagram of DEGs identified by a comparison of each temperature point to the control environment. B, GO classification of 134 coexpressed genes. C, COG function classification of 134 coexpressed genes. D, KEGG enrichment of 134 coexpressed genes.

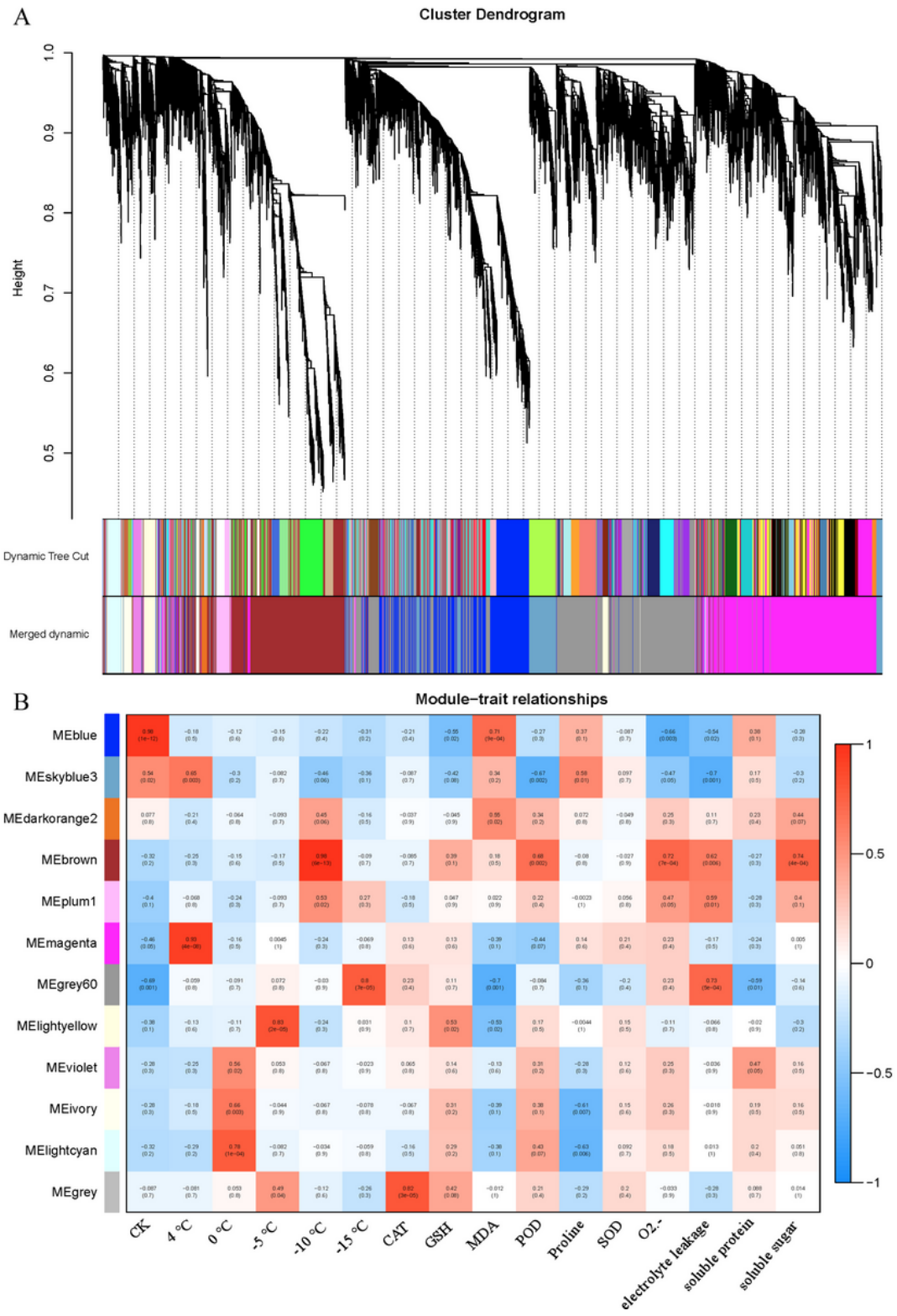

Figure 5

WGCNA of the genetic modules related to each sample and physiological indicators. A, Cluster dendrogram. B, Module-trait relationships.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement1.docx
- supplement1.docx
- supplement3.docx
- supplement4.xlsx
- supplement5.docx
- supplement6.tif
- supplement7.tif
- supplement8.tif
- supplement9.xlsx
- supplement10.xlsx
- supplement11.tif
- supplement12.xlsx
- supplement13.xlsx
- supplement14.xlsx
- supplement15.xlsx