**CSIRO** PUBLISHING
*Crop & Pasture Science*, 2019, **70**, 273–282
https://doi.org/10.1071/CP18165

# Comparative transcriptome analysis of five *Medicago* varieties reveals the genetic signals underlying freezing tolerance

*Lijun Xu*[A], *Xuejuan Tang*[A], *Bo Wang*[B], *Xiaoping Xin*[A,G], *Qizhong Sun*[C,G], *Yalu Li*[D], *Jinqiang Chen*[A], *Gele Qing*[E], *and Mingying Guo*[F]

[A]Hulunber Grassland Ecosystem Observation and Research Station/Chinese Academy of Agricultural Sciences Institute of Agricultural Resources and Regional Planning, Beijing 100081, China.
[B]Conservation Program Center, Chinese Academy of Engineering, Beijing 100088, China.
[C]Chinese Academy of Agricultural Sciences Grassland Research Institute, Hohhot 010010, Inner Mongolia, China.
[D]College of Grassland, Resources and Environment/Inner Mongolia Agricultural University, Hohhot 010000, Inner Mongolia, China.
[E]Xilinhot Vocational School, Xinlinhot 026000, Inner Mongolia, China.
[F]Hulunber Grassland Station, Hulunber, Inner Mongolia 021000, China.
[G]Corresponding author. Email: xinxiaoping@caas.cn; sunqz@126.com

**Abstract.** *Medicago* plants such as *M. sativa* (lucerne, alfalfa) are important forage all over the world. Freezing-tolerance capacity is one of the key determinants of the survival and production of *Medicago*. In order to explore the molecular basis underlying freezing tolerance, we sequenced the root transcriptomes of five *Medicago* varieties belonging to two species, *M. sativa* and *M. varia*, and compared their gene expression and molecular evolution. A range of 19.5–23.8 Gb clean bases was obtained, and *de novo* transcriptome assembly generated 205 238–268 520 unigenes. The GO (Gene Ontology) terms of basic biological processes such as binding, cell and metabolism were most represented for the unigenes. In addition, a large number of unigenes related to GO terms and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways of membrane, signalling, transcription and response to stimulus were identified in functional annotation. In total, 12 455 orthologs were identified among the five *Medicago* varieties. Among the orthologs, many unigenes that directly related to freezing tolerance were highly expressed in all five varieties, including genes for WRKY transcription factors, calcium-binding factors, and antioxidant enzymes such as catalase and ascorbate peroxidase. Molecular evolution testing showed that the unigenes involved in membrane shared high $K_a/K_s$ (non-synonymous/synonymous substitution rate) across all the five *Medicago* varieties. Positively selected genes were mainly involved in transcription regulation, metabolism and signal transduction. Our study provides a large transcriptome dataset in the *Medicago* genus and brings new insights into the freezing tolerance for *Medicago* species.

**Additional keywords:** APX, CAT, CBF.

Received 20 April 2018, accepted 13 January 2019, published online 12 March 2019

## Introduction

*Medicago sativa* (lucerne, alfalfa) is an important perennial leguminous forage, and widely cultivated all over the world for its high yield, high protein content, environmental adaptability and good regeneration capacity (Biazzi *et al*. 2017). However, freezing-tolerance capacity is an issue that affects the survival and yield of *M. sativa* in high-latitude and high-altitude regions. Conventional breeding to improve freezing tolerance of *M. sativa* and other *Medicago* species is mainly achieved through a combination of crossing with freezing-tolerant varieties and artificial selection, which has low efficiency and is time-consuming and laborious. Thus, understanding the molecular mechanisms underlying freezing tolerance will provide valuable information for molecular marker-assisted selection (MAS) to select *Medicago* species with better freezing adaptation.

Many plants acquire freezing-tolerance capacity after cold acclimation, a process of exposure to a low, non-freezing temperature (Thomashow 1999). Cold acclimation and acquisition of freezing tolerance in plants involve extensive physiological and biochemical changes (Kaye and Guy 1995), which are mainly mediated by reprograming gene expression (Thomashow 1999; Fowler and Thomashow 2002; Winfield *et al*. 2010). Studies in *Medicago* species have identified many genes

Journal compilation © CSIRO 2019 

that respond to cold acclimation, such as *C-REPEAT BINDING FACTOR* (*CBF*) and *COLD-ACCLIMATION-SPECIFIC* (*CAS*) genes (Pennycooke *et al*. 2008). For example, the *CBF*s lead to cold acclimation and freezing tolerance by activating the expression of cold-responsive (*COR*) genes (Chinnusamy *et al*. 2007). Moreover, overexpression of plasma membrane intrinsic protein 2 (MfPIP2) and a temperature-induced lipocalin gene (*MfTIL1*) were shown to improve freezing tolerance in *M. falcate* (He *et al*. 2015; Zhuo *et al*. 2016). Although many genes that respond to cold have been identified, the transcriptome level of molecular mechanisms underlying freezing tolerance and genome-wide identification of genes in response to cold-stress in *Medicago* species remain limited.

Recent studies based on transcriptome sequencing have indicated that the molecular mechanisms underlying freezing tolerance in *M. sativa* and many other plants share some similarity. For example, many genes encoding membrane proteins, signal transduction components, transcription factors and antioxidants were upregulated under freezing stress (Winfield *et al*. 2010; An *et al*. 2012; Chen *et al*. 2014; Beike *et al*. 2015; Song *et al*. 2016; Chen *et al*. 2017), indicating that transcriptional regulations play vital roles in cold adaption of plants. In this study, we performed transcriptome sequencing of the roots of five *Medicago* varieties under freezing stress. We aimed to investigate the molecular basis underlying cold adaption and identify freezing-tolerance-related genes in *Medicago* by comparative transcriptome analysis. The transcriptome datasets of our study will contribute to functional genomics studies of *Medicago* plants. The simple sequence repeats (SSRs) identified in this study are potential molecular markers for germplasm identification, and can be used as functional markers in *Medicago* molecular breeding processes.

## Materials and methods

### Plant material

Five *Medicago* varieties belonging to two species, *M. sativa* and *M. varia*, were included in this study. Four were *M. sativa* varieties: AC Caribou and BR4010 were obtained from Beijing Kelaowo Seed Co., WL656HQ was obtained from Beijing Zhengdao Seed Co., and Gannong No. 3 (GN3) was obtained from the Gansu Agricultural University. The *M. varia* accession Zhongyu No. 1 (ZY1) was provided by the Institute of Agricultural Resources and Agricultural Regionalisation of the Chinese Academy of Agricultural Sciences. All five *Medicago* varieties are tetraploid and winter-hardy (Supplementary Materials table S1, available at the journal's website). Seeds were sown in the experiment field (15 kg seed $ha^{-1}$, row spacing 20 cm) in Hulun-Beir grassland (49°19′35′N, 119°56′52′; altitude 650 m a.m.s.l.) in 2012. Annual average temperature and annual minimum temperature in Hulun-Beir grassland are –2°C and –48°C, respectively. The *Medicago* plants were grown under natural conditions. Over-wintering rates of these *Medicago* varieties were recorded in four successive years (2014–17) of field trials. The roots used for transcriptome sequencing were collected on 14 November 2016 (soil temperature –14.2°C). To eliminate individual errors, equal amounts of roots from three different plants were pooled as a replicate, and three replicates were used for each accession. In total, 15 transcriptomes were sequenced.

### Library construction and RNA sequencing

Total RNA was isolated by using an RNAprep Pure Plant Kit (BioTeKe, Beijing) following the manufacturer's instructions. The cDNA libraries were constructed by using a NEBNext Ultra Directional RNA Library Prep Kit (New England BioLabs, Ipswich, MA, USA) according to the manufacturer's instructions. Briefly, the mRNA was first fragmented into 250–450 bp before being used for synthesising the first strand of cDNA. To synthesise the second strand of cDNA, dUTP was added to mark the second strand. The double-stranded cDNA was digested with uracil-DNA glycosylase (UDG) before PCR reaction. In this way, only the first strands of cDNA were retained and sequenced. Transcriptome sequencing was carried out on the Hiseq 4000 (Illumina, San Diego, CA, USA) platform using a paired-end run (2 × 150 bp).

The raw transcriptome data were deposited at the NCBI Short Read Archive database (accession numbers are SRR7969362, SRR7969363, SRR7969364, SRR7969365 and SRR7969366).

### Transcriptome assembly and annotation

Raw reads were first processed to remove adaptor sequences and reads with quality below Q20. All of the clean reads from each *Medicago* accession were pooled and used for *de novo* assembly using Trinity software (version 2.2.0, https://github.com/trinityrnaseq/trinityrnaseq/wiki) with default parameters (fixed *k*-mer value of 25). Trinity assembles the transcriptome by using de Bruijn graph strategy (Grabherr *et al*. 2011). Briefly, the clean reads were first assembled into contigs, then the overlapped contigs overlapped and the reads that astride contigs were assigned into a same group. Finally, the longest transcripts in each group were defined as unigenes. All of the assembled unigenes were used as reference sequences for following analyses.

The CDS (coding sequences) and protein sequences of unigenes were predicted by using TransDecoder (http://transdecoder.github.io/). Functional annotations of unigenes were performed by using the protein sequences of unigenes to BLAST (E-value $<10^{-5}$) against the following public databases: NR (NCBI Non-Redundant protein sequences: https://www.ncbi.nlm.nih.gov/refseq/about/nonredundantproteins/), COG (NCBI Clusters of Orthologous Groups of proteins: http://www.ncbi.nlm.nih.gov/COG/), and Swiss-Prot (a manually annotated and reviewed protein sequence database: http://www.ebi.ac.uk/uniprot). The GO (Gene Ontology: http://geneontology.org/) annotation of unigenes were carried out by using WEGO software based on their NR annotation (Ye *et al*. 2006). KEGG pathway annotations were conducted by KAAS (KEGG Automatic Annotation Server) (Kyoto Encyclopedia of Genes and Genomes: http://www.genome.jp/kegg/).

### Quantitative RT-PCR analysis

For quantitative RT-PCR (qRT-PCR) analysis, 1 μg root total RNA of each sample was first treated with the RNase-Free DNase I Set (Thermo Scientific, Waltham, MA, USA). Reverse transcription was conducted with a ReverAid First Strand cDNA Synthesis Kit (Thermo Scientific, Waltham, MA,

**Table 1. Summary of transcriptome assemblies**

| | *M. sativa_WL656HQ* | *M. sativa_AC Caribou* | *M. sativa_GN3* | *M. sativa_BR4010* | *M. varia_ZY1* |
|---|---|---|---|---|---|
| Raw reads | 82 391 295 × 2 | 94 137 102 × 2 | 82 202 846 × 2 | 80 400 078 × 2 | 77 379 764 × 2 |
| Clean reads | 71 862 917 × 2 | 82 356 665 × 2 | 72 996 288 × 2 | 70 046 937 × 2 | 67 519 629 × 2 |
| Clean base (bp) | 20 867 014 793 | 23 791 297 375 | 21 149 714 551 | 20 312 348 887 | 19 510 644 577 |
| GC content (%) | 43.35 | 43.7 | 43.49 | 44.22 | 43.2 |
| Q20 (%) | 97.78 | 98.28 | 98 | 97.78 | 97.78 |
| No. of unigenes | 240 175 | 242 302 | 241 338 | 205 238 | 268 520 |
| Mean length (bp) | 444.87 | 457 | 450.64 | 470 | 440.32 |
| Median length (bp) | 310 | 313 | 311 | 319 | 309 |
| N50 (bp) | 764 | 781 | 770 | 805 | 755 |
| Total bases (bp) | 106 845 797 | 110 732 169 | 108 756 079 | 96 462 751 | 118 234 304 |

**Table 2. Functional annotations of the *de novo* transcriptomes in five databases**

| | *M. sativa_WL656HQ* | *M. sativa_AC Caribou* | *M. sativa_GN3* | *M. sativa_BR4010* | *M. varia_ZY1* |
|---|---|---|---|---|---|
| COG (Clusters of Orthologous Groups) | 63 439 (26.41%) | 58 365 (24.09%) | 64 350 (26.66%) | 49 429 (24.08%) | 69 444 (25.86%) |
| SwissProt | 77 285 (32.18%) | 71 119 (29.35%) | 78 535 (32.54%) | 61 292 (29.86%) | 83 643 (31.15%) |
| NR (Non-Redundant protein sequences) | 137 560 (57.27%) | 129 803 (53.57%) | 136 180 (56.43%) | 111 299 (54.23%) | 149 723 (55.76%) |
| KEGG (Kyoto Encyclopedia of Genes and Genomes) | 42 797 (17.82%) | 21 153 (8.73%) | 47 145 (19.53%) | 34 989 (17.05%) | 54 867 (20.43%) |
| GO (Gene Ontology) | 78 357 (32.62%) | 73 786 (30.45%) | 78 781 (32.64%) | 62 426 (30.42%) | 86 037 (32.04%) |
| Annotated in all databases | 21 831 (9.09%) | 10 615 (4.38%) | 24 883 (10.31%) | 17 391 (8.47%) | 27 946 (10.41%) |
| Annotated in at least one database | 139 116 (57.92%) | 131 447 (54.25%) | 138 688 (57.47%) | 112 888 (55.00%) | 151 877 (56.56%) |

USA). Real-time qPCR was performed by using a SYBR Green Master Mix (Bio-Rad, Hercules, CA, USA) with a LightCycler 96 (Roche, Basel, Switzerland). Primers of the genes for qRT-PCR analysis are listed in Supplementary Materials table S2.

### *Detection of SSRs*

The SSRs were detected by using MISA software (version 1.0) (Thiel *et al*. 2003). The minimum repeat numbers for unit size of mono-, di-, tri-, tetra-, penta-, and hexa-nucleotides were set as 10, 6, 5, 5, 5 and 5, respectively.

### *Estimation of positive selected genes (PSGs)*

The number of non-synonymous substitutions per non-synonymous site ($K_a$) and the number of synonymous substitutions per synonymous site ($K_s$) of each single-copy orthologue were calculated using KaKs Calculator software (version 2.0) (Zhang *et al*. 2006). The PSGs were estimated using KaKs Calculator software (version 2.0) with MYN model (Zhang *et al*. 2006). Putative PSGs were inferred with $K_a/K_s$ value >1 and adjusted *P*-value <0.05 (Benjamini and Yekutieli 2001).

## Results and discussion

### *Summary and assembly of transcriptomes*

Transcriptome sequencing obtained a range of 77 379 764 × 2 to 94 137 102 × 2 raw reads for the five *Medicago* varieties (Table 1), which generated a range of 19.5–23.8 Gb clean bases after filter and quality control. The clean GC content of the five varieties ranged from 43.2% to 44.22% (Table 1). *De novo* transcriptome assembly generated 240 175, 242 302, 241 338, 205 238 and 268 520 unigenes for *M. sativa_WL656HQ*, *M. sativa_AC Caribou*, *M. sativa_GN3*, *M. sativa_BR4010* and *M. varia_ZY1* (Table 1), respectively. The N50 values of the five transcriptome were 764, 781, 770, 805 and 755 bp (Table 1), respectively. The length distributions of unigenes of the five *Medicago* varieties are presented in Supplementary Materials figure S1. These transcriptome datasets add to the sequence to the whole *Medicago* genus and will provide valuable information for the study of molecular biology in this genus.

### *Functional annotation of unigenes*

In total, 139 116 (57.92%), 131 447 (54.25%), 138 688 (57.47%), 112 888 (55.00%) and 151 877 (56.56%) unigenes in *M. sativa_WL656HQ*, *M. sativa_AC Caribou*, *M. sativa_GN3*, *M. sativa_BR4010*, and *M. varia_ZY1*, respectively, were annotated in at least one of the five public databases that were aligned in this study (Table 2). Of the annotated unigenes, almost all of them were matched in NR database. GO classification assigned these unigenes into 46 level-two GO terms of the three categories (Fig. 1). With regard to molecular function category, the 'catalytic activity' (GO:0003824) and 'binding' (GO:0005488) terms contained the greatest number of unigenes (Fig. 1). For the cellular-component category, the top three terms represented were 'cell' (GO:0005623), 'cell part' (GO:0044464) and 'organelle' (GO:0043226) (Fig. 1). In the biological-process category, the 'metabolic process' (GO:0008152), 'cellular process' (GO:0009987) and 'single-organism process' (GO:0044699) terms were the most represented GO terms in all five *Medicago* varieties (Fig. 1). The GO-term distributions of unigenes in the five varieties were quite similar; this indicates that the sampling and transcriptome sequencing in this study had no bias, and the cold response in the different *Medicago* species was similar. In addition, the terms 'membrane' (GO:0016020), 'membrane part' (GO:0044425), 'signaling' (GO:0023052) and 'response to stimulus'

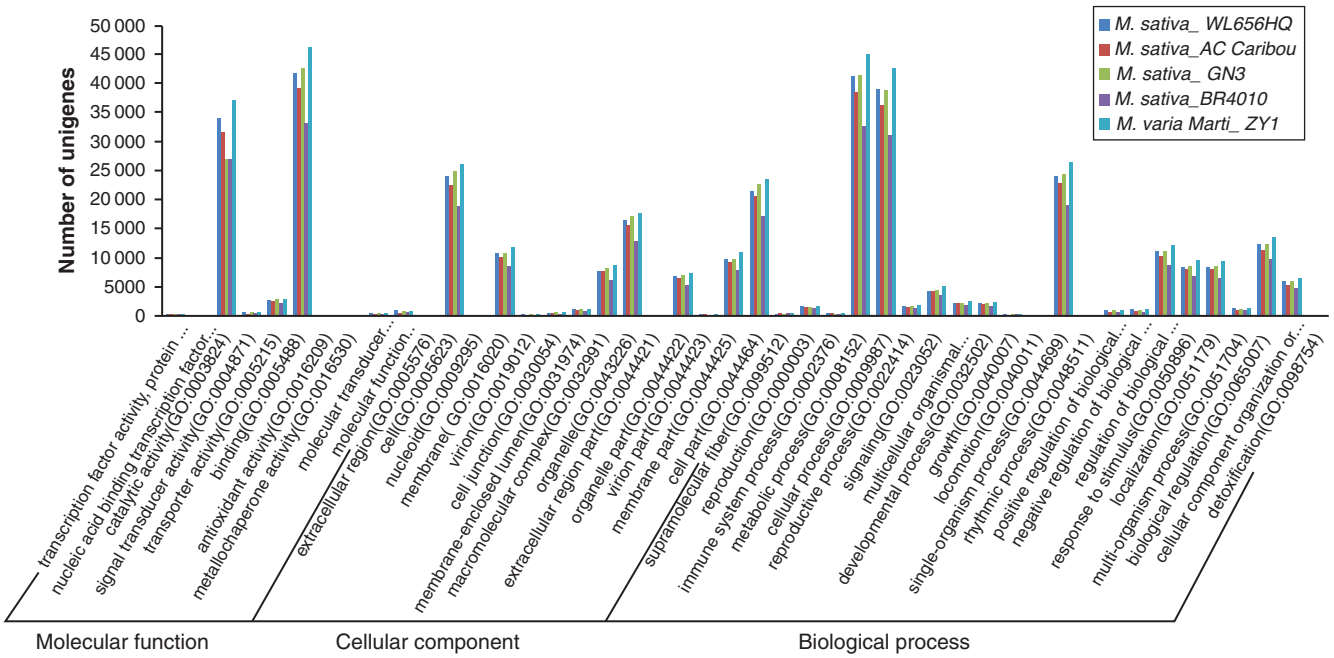


**Fig. 1.** GO (Gene Ontology) term distributions of unigenes in the five *Medicago* varieties; *x*-axis indicates the level-two GO terms and *y*-axis indicates the number of unigenes in each GO term.

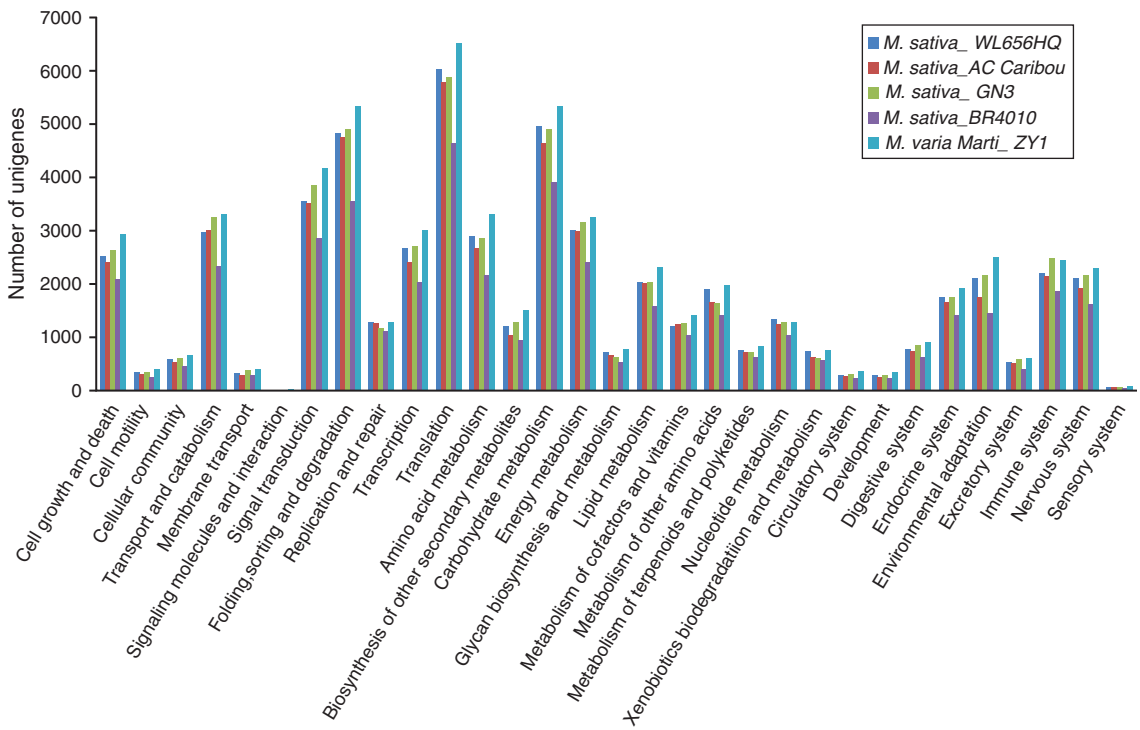


**Fig. 2.** KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway distributions of unigenes in the five *Medicago* varieties; *x*-axis indicates the hierarchy two pathways and *y*-axis indicates the number of unigenes in each pathway.

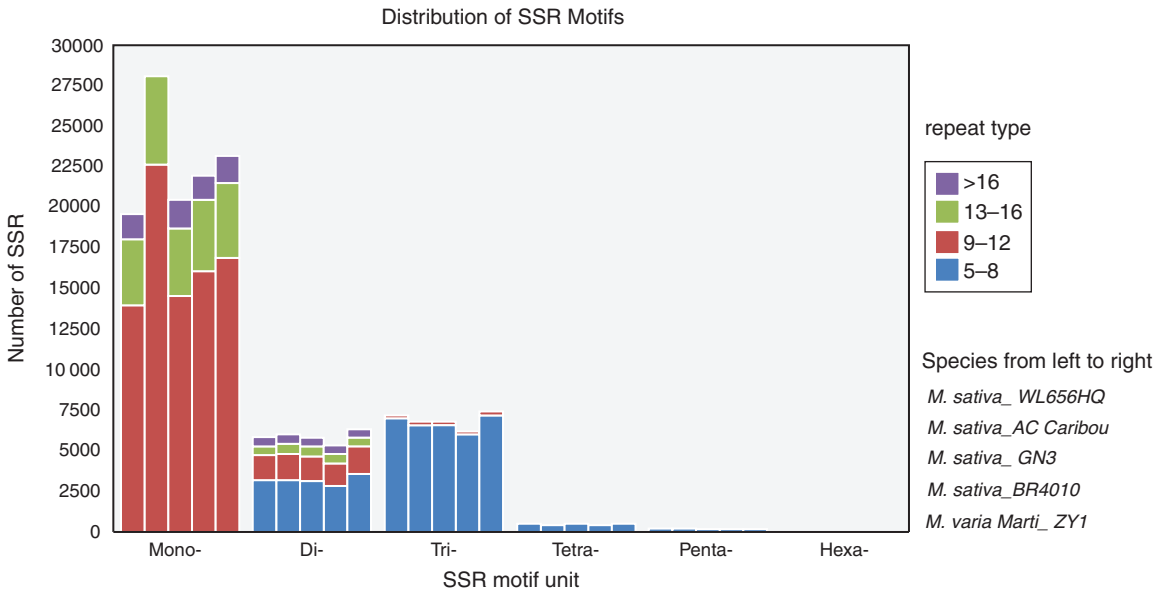


**Fig. 3.** Distribution of SSR motifs in the five *Medicago* varieties. Different colour bars represent different repeat types (repeat number ranges of SSR motif unit).

(GO:0050896) contained a high number of unigenes (Fig. 1). Further KEGG pathway classification showed that metabolism pathways such as amino acid, carbohydrate, energy, glycan, lipid, and cofactors and vitamins contained the largest number of unigenes (Fig. 2). In addition, the pathways 'cell growth and death', 'transport and catabolism', 'signal transduction', 'folding, sorting and degradation', 'translation', 'transcription' and 'environmental adaption' were also highly represented in all five *Medicago* varieties (Fig. 2).

Previous studies in different plant species showed that freezing tolerance is a complex trait involving many physiological and biochemical changes (Winfield *et al*. 2010). The membrane was predicted to perceive temperature stimulus by altering membrane fluidity (Los and Murata 2004). Genes involved in the metabolism of sugar, lipid and amino acid were shown to be upregulated under cold stress (Beike *et al*. 2015; Chen *et al*. 2017). These molecules are important effectors that work to alleviate stress in cells. Moreover, many transcription factors such as WRKYs (Wang *et al*. 2014; Zhang *et al*. 2016) and calcium-binding factors (CBFs) (Pennycooke *et al*. 2008; Winfield *et al*. 2010), and genes that participate in pathways of signal transduction (Beike *et al*. 2015; Hu *et al*. 2016) and trans-membrane transport (Chen *et al*. 2017), were also upregulated in response to cold-stress. Taken together, our functional annotation of the expressed unigenes under cold stress were consistent with the previous studies, which confirmed that metabolic processes, expression of transcription factors and signal transductions were largely activated in plants under cold environments.

### *Detection of SSRs*

In total, 33 637, 43 813, 33 942, 34 318 and 38 051 SSRs were detected in *M. sativa_WL656HQ*, *M. sativa_AC Caribou*, *M. sativa_GN3*, *M. sativa_BR4010*, and *M. varia_ZY1*, respectively (Fig. 3). The numbers of SSR-containing sequences for the five varieties were 29 042, 37 303, 29 110, 29 195 and 32 734, respectively (Supplementary Materials table S3). Most of the SSRs were mononucleotide repeats (range 58.49–79.17%), and mainly $(A/T)_n$ (Supplementary Materials table S3). The second-largest SSR group for each *Medicago* accession was trinucleotide repeats, which were slightly more common (1.3–4.2%) than dinucleotide repeats. These RNA-Seq SSRs markers can be used in MAS, and will facilitate hybridisation breeding between different *Medicago* species.

### *Freezing-tolerance-related genes*

In order to explore the molecular basis of freezing tolerance in *Medicago*, single-copy orthologs among the five *Medicago* varieties were identified and the orthologs that related to freezing tolerance were selected for further expression comparison. The expression levels of the unigenes were measured by using fragments per kb per million reads (FPKM) methods (Li and Dewey 2011). In total, 12 455 orthologs were identified (Supplementary Materials table S4). The expression levels of these orthologs are displayed in Supplementary Materials figure S2, which shows that the expression level of most of the orthologs was similar across the five *Medicago* varieties, indicating that gene regulation networks in the five varieties were similar under frozen conditions. Thus, identifying cold-response orthologs with similar expression level may help to find a common molecular basis for freezing tolerance of *Medicago*.

Studies in *Arabidopsis* and other plants have shown that genes encoding transcription factors (Fowler and Thomashow 2002; Lee *et al*. 2005), effector molecules (Winfield *et al*. 2010) and antioxidants (Guo *et al*. 2014) play essential roles in freezing tolerance in plants. Transcription factors play important roles in signal transduction and control the target genes by binding to the *cis*-acting elements in the promoter regions of the target genes (Winfield *et al*. 2010). Effector molecules act to regulate osmotic

stress in cells and can prevent cold-induced dehydration of plants. In this study, many such cold-response genes were identified, including those encoding WRKY transcription factors, CBFs, antioxidant enzymes and effector proteins such as osmotin, thaumatin and chaperone. The heat map of the expression levels of these cold-response genes, and four housekeeping genes encoding histone H2A (*H2A*), tubulin α-6 chain (*Tub*), elongation factor 1-α (*EF1*) and glyceraldehyde-3-phosphate dehydrogenase (*Gapdh*), is displayed in Fig. 4, and the FPKM values of these genes are shown in Supplementary Materials table S5. One of the genes for a WRKY transcription factor, *WRKY6*, displayed high expression levels in all five *Medicago* varieties (Fig. 4). The WRKY family is one of the largest transcription-factor families in plants, and plays important roles in stress tolerance (Agarwal *et al*. 2011). Expression of genes for WRKY transcription factors confers cold resistance in many plants (Wang *et al*. 2014; Zhang *et al*. 2016). Our results suggest that *WRKY6* may be one of the candidate genes that improve freezing tolerance of *Medicago* plants.

It is generally accepted that the perception of temperature changes in plants is mediated by $Ca^{2+}$ influx into the cytosol (Winfield *et al*. 2010). The $Ca^{2+}$-binding proteins in cytosol, CBFs and annexin (An), are important signal transporters that directly interact with downstream target proteins and initiate large-scale reprogramming of gene expression to preventing cold damage (Chinnusamy *et al*. 2007). There are three major CBFs in plants: calcineurin B-like proteins (CBLs), calcium-dependent protein kinases (CDPKs) and calmodulin-like proteins (CMLs) (Fowler and Thomashow 2002). Transgenic plants overexpressing *CBL* genes were shown to display increased cold tolerance (Deng *et al*. 2013; Zhou *et al*. 2016). Additionally, overexpression of the rice *OsCDPK7* gene and maize *CDPK* gene in *Arabidopsis* also showed enhanced cold tolerance (Saijo *et al*. 2000) (Wang and Shao 2013). In this study, two *CBF* genes, *CBL3* and *CDPK4*, were highly expressed in the five *Medicago* varieties (Fig. 4), which indicates that CBL3 and CDPK4 may play vital roles in cold adaption.

Cold-stress was reported to increase reactive oxygen species (ROS) in plant cells (Suzuki *et al*. 2012; Baxter *et al*. 2014), which are harmful metabolites in the cell and can cause damage to DNA and proteins and lead to plant death. Cold stress can also increase antioxidant levels and antioxidant enzyme activities in plants (Streb *et al*. 2003; Król *et al*. 2015). In this study, we found that the unigenes encoding SOD (superoxide dismutase), CAT (catalase), APX (ascorbate peroxidase), PrxR (peroxiredoxin), GPX (glutathione peroxidase), GR (glutathione reductase) and GST (glutathione S-transferase) were highly expressed in all five *Medicago* varieties (Fig. 4). Moreover, a gene in the flavonoid biosynthesis pathway, encoding FGT (flavonoid glucosyltransferase) (Fig. 4), was also high expressed. Flavonoids are important antioxidants in plants and play significant roles in cold tolerance (Schulz *et al*. 2016). Antioxidant enzymes play important roles in scavenging ROS and alleviating oxidant stress in cells (Choudhary *et al*. 2012; Bela *et al*. 2015). The high expression level of many of the antioxidant enzymes suggests that the antioxidant biological processes may play major roles in freezing tolerance in *Medicago*.

Effector molecules such as osmotin, sugar, thaumatin, chaperone, LTPs (lipid transfer proteins) and LEAs (late embryogenesis abundant proteins) also play vital roles in freezing tolerance of plants by modulating membrane fluidity and osmotic stress in cells (Winfield *et al*. 2010). Transcriptome profiling of different plants under cold stress revealed that many genes encoding these effector molecules were upregulated (Svensson *et al*. 2006; Winfield *et al*. 2010). In this study, we detected one unigene coding for osmotin, three unigenes for sugar transporter, one for thaumatin, five for chaperone, one for LTP and one for LEA that were highly expressed in the five *Medicago* varieties. Among these unigenes, those coding for thaumatin, chaperone20, LTP and LEA3 were highly expressed in all five *Medicago* varieties (Fig. 4).

For expression validation, we selected eight genes for qRT-PCR analysis, including four antioxidant-related genes, *SOD*, *CAT*, *APX* and *GST*, one calcium signal gene, *CDPK4*, one transcription factor, *WRKY4*, and two effector-molecule genes, *LEA3* and *chaperone20* (Fig. 5). In accordance with the transcriptome sequencing results, almost all of the cold-response genes in each variety had higher expression level than the housekeeping gene *ACTIN*. This result indicated that our transcriptome sequencing is reliable.

Our transcriptome profiling of the roots of different *Medicago* varieties under freezing temperature (–14.2°C) indicates that the molecular mechanism underlying freezing tolerance in *Medicago* is similar to that of other winter-hardy plants such as wheat (*Triticum aestivum* L.) and barley (*Hordeum vulgare* L.).

### Detection of PSGs

In contrast to the model species *M. truncatula*, which is sensitive to low temperature (–4°C) (Pennycooke *et al*. 2008), the five *Medicago* varieties used for transcriptome sequencing in this study are all winter-hardy and can survive in temperature as low as –40°C. To identify genes that had experienced or are undergoing positive selection for cold adaption, we performed molecular evolution evaluation of the single-copy orthologs across the five *Medicago* varieties by comparing with the cold-sensitive *M. truncatula*. In total, 494 single-copy orthologs were identified among the six *Medicago* varieties (Supplementary Materials table S6). Average $K_a/K_s$ ratios of the orthologue clusters according to GO categories showed that orthologs in the 'membrane' (GO:0016020) term shared high $K_a/K_s$ ratios in all five *Medicago* varieties (Supplementary Materials figure S3). This result is consistent with recent transcriptome analysis of *M. sativa* showing that freezing adaptation involved in many membrane proteins (Song *et al*. 2016), suggesting the membrane plays vital roles in cold adaption.

Further PSG estimation identified 24, 3, 11, 3 and 8 PSGs ($K_a/K_s$ >1, $P < 0.05$) in *M. sativa_WL656HQ*, *M. sativa_AC Caribou*, *M. sativa_GN3*, *M. sativa_BR4010*, and *M. varia_ZY1*, respectively (Supplementary Materials table S7). Functional annotation showed that these PGSs were mainly involved in transcription regulation, signalling and metabolism (Fig. 6, Supplementary Materials table S7). Among the PSGs, the *LEA* gene, which displayed positive selection in *M. sativa_WL656HQ* and *M. varia_ZY1*, was an important cold-response gene. The *LEA* genes confer freezing tolerance by encoding hydrophilic protein to prevent plants from suffering cold-

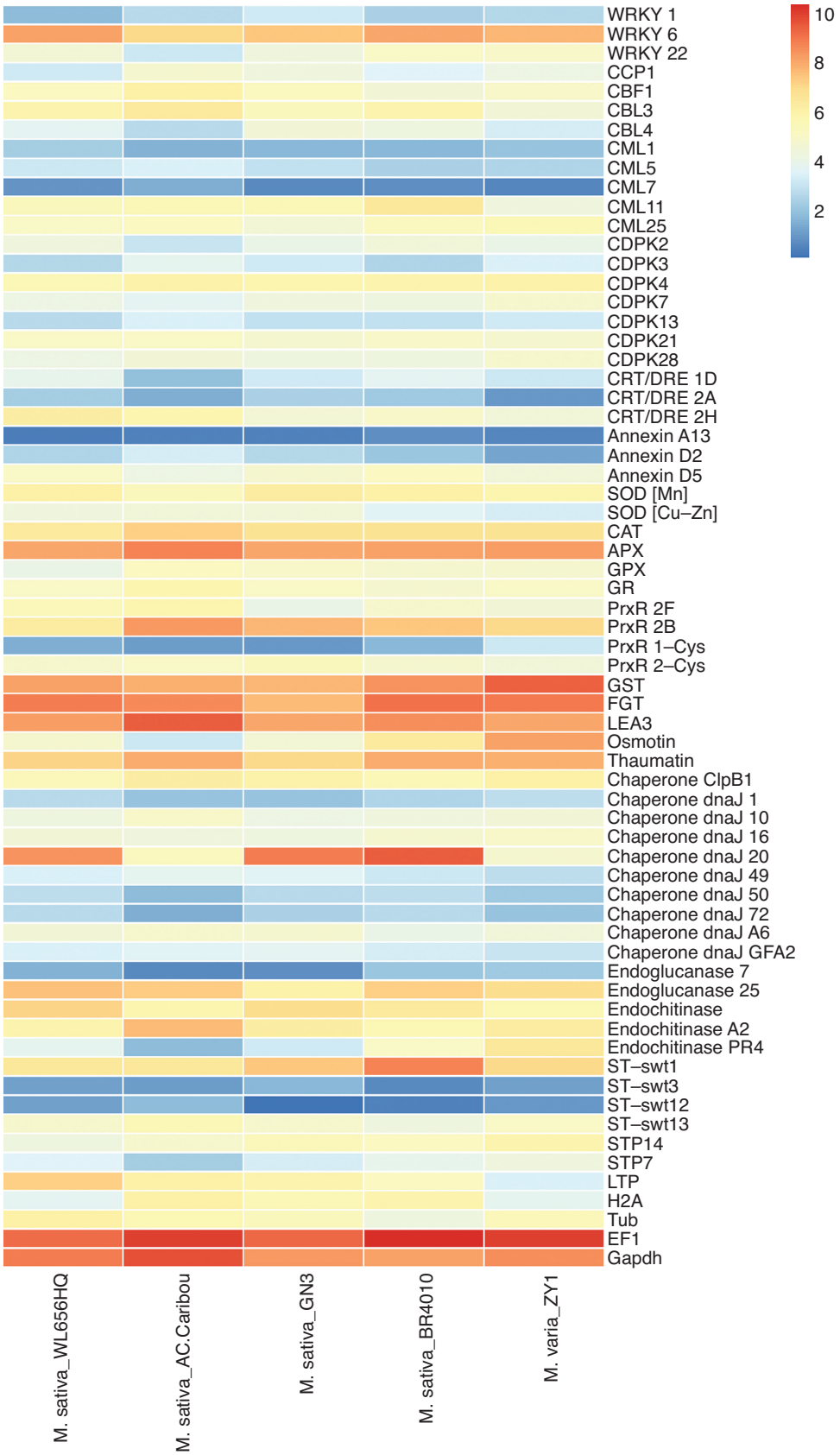


**Fig. 4.** Expression heat map of cold-related unigenes in the five *Medicago* varieties. Horizontal lines show the expression of a unigene in different varieties. Gene expression levels are represented by $\log_2$(FPKM + 1) value. Colours from blue to red indicate expression level from low to high.

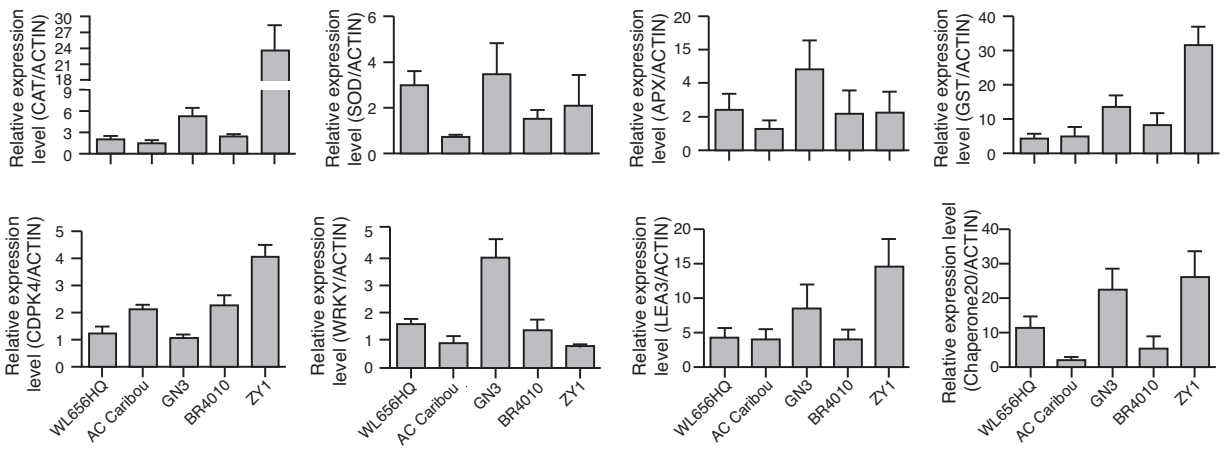


**Fig. 5.** Expression levels of cold-response genes by qRT-PCR analysis.

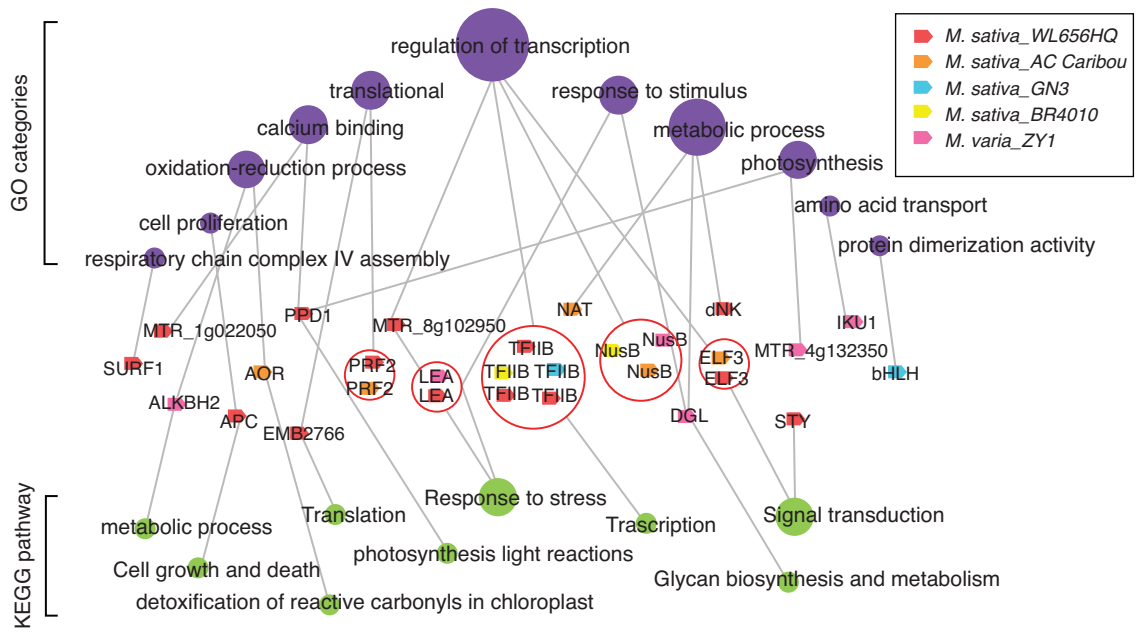


**Fig. 6.** Genetic network of putative positively selected genes (PSGs) in the five *Medicago* varieties. Functions of the PSGs are defined by using GO and KEGG annotations.

induced dehydration (Kosová *et al*. 2014). Overexpression of a wheat LEA protein in *Arabidopsis* was proved to increase freezing tolerance (NDong *et al*. 2002). In addition, the gene encoding bHLH (basic helix loop helix protein) transcription factor, which was positively selected in *M. sativa_GN3*, was also proved to relate to cold tolerance. Transgenic *Arabidopsis* plants overexpressing the grapevine *bHLH* gene improved cold tolerance by activating the CBF pathway (Xu *et al*. 2014). Despite most of the PSGs annotated in the GO terms and pathways being related to cold response, their roles in cold tolerance are still not clear and need further investigation. It is noteworthy that the gene encoding TFIIB (transcription factor IIB) was positively selected in all five *Medicago* (Fig. 6). Studies in *Arabidopsis* showed that TFIIB was required for pollen tube growth and endosperm development (Zhou *et al*. 2013). Thus, whether TFIIB is related to cold adaption of *Medicago* needs further investigation.

## Conclusion

The transcriptomes of *Medicago* roots under freezing stress provide a large transcriptome dataset for the *Medicago* genus. The dataset provides valuable genomic information for molecular function research of *Medicago* plants. The SSRs based on transcriptome sequencing are potential functional markers for *Medicago* breeding and germplasm identification. The highly expressed unigenes shared by the five *Medicago* varieties, and the PSGs, are candidate genes conferring freezing

tolerance in *Medicago* plants. Genes involved in binding, transcription regulation and membrane composition may play vital roles in cold adaption. In addition, antioxidant enzymes were highly activated and may play major roles in freezing tolerance in *Medicago* plants. Our study therefore brings new insights into the molecular basis underlying freezing tolerance and will facilitate molecular breeding of *Medicago* species.

## Conflicts of interest

The authors declare no conflict of interest.

## Contributions

Xu Lijun, Sun Qizhong, Xin Xiaoping and Chen Jinqiang conceived, designed, and performed the experiments, and wrote the paper. Li Yalu, Wang Bo, Tang Xuejuan, Qing Gele and Guo Mingying prepared the materials and performed data analysis. Xin Xiaoping and Sun Qizhong revised this paper. This article was read and approved by all of the authors.

## Acknowledgements

This work was supported by grants from the '973' Program (2015CB150800), National Natural Science Foundation of China (No. 41703081), the National Key Research and Development Program of China (No. 2016YFC0500603), and the China Agriculture Research System (No. CARS-34).

## References

Agarwal P, Reddy MP, Chikara J (2011) WRKY: its structure, evolutionary relationship, DNA-binding selectivity, role in stress tolerance and development of plants. *Molecular Biology Reports* **38**, 3883–3896. doi:10.1007/s11033-010-0504-5

An D, Yang J, Zhang P (2012) Transcriptome profiling of low temperature-treated cassava apical shoots showed dynamic responses of tropical plant to cold stress. *BMC Genomics* **13**, 64. doi:10.1186/1471-2164-13-64

Baxter A, Mittler R, Suzuki N (2014) ROS as key players in plant stress signalling. *Journal of Experimental Botany* **65**, 1229–1240. doi:10.1093/jxb/ert375

Beike AK, Lang D, Zimmer AD, Wust F, Trautmann D, Wiedemann G, Beyer P, Decker EL, Reski R (2015) Insights from the cold transcriptome of *Physcomitrella patens*: global specialization pattern of conserved transcriptional regulators and identification of orphan genes involved in cold acclimation. *New Phytologist* **205**, 869–881. doi:10.1111/nph.13004

Bela K, Horvath E, Galle A, Szabados L, Tari I, Csiszar J (2015) Plant glutathione peroxidases: emerging role of the antioxidant enzymes in plant development and stress responses. *Journal of Plant Physiology* **176**, 192–201. doi:10.1016/j.jplph.2014.12.014

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* **29**, 1165–1188.

Biazzi E, Nazzicari N, Pecetti L, Brummer EC, Palmonari A, Tava A, Annicchiarico P (2017) Genome-wide association mapping and genomic selection for alfalfa (*Medicago sativa*) forage quality traits. *PLoS One* **12**, e0169234. doi:10.1371/journal.pone.0169234

Chen X, Zhang J, Liu Q, Guo W, Zhao T, Ma Q, Wang G (2014) Transcriptome sequencing and identification of cold tolerance genes in hardy *Corylus* species (*C. heterophylla* Fisch) floral buds. *PLoS One* **9**, e108604. doi:10.1371/journal.pone.0108604

Chen J, Yang X, Huang X, Duan S, Long C, Chen J, Rong J (2017) Leaf transcriptome analysis of a subtropical evergreen broadleaf plant, wild oil-tea camellia (*Camellia oleifera*), revealing candidate genes for cold acclimation. *BMC Genomics* **18**, 211. doi:10.1186/s12864-017-3570-4

Chinnusamy V, Zhu J, Zhu JK (2007) Cold stress regulation of gene expression in plants. *Trends in Plant Science* **12**, 444–451. doi:10.1016/j.tplants.2007.07.002

Choudhary R, Saroha AE, Swarnkar PL (2012) Effect of abscisic acid and hydrogen peroxide on antioxidant enzymes in *Syzygium cumini* plant. *Journal of Food Science and Technology* **49**, 649–652. doi:10.1007/s13197-011-0464-3

Deng X, Zhou S, Hu W, Feng J, Zhang F, Chen L, Huang C, Luo Q, He Y, Yang G, He G (2013) Ectopic expression of wheat *TaCIPK14*, encoding a calcineurin B-like protein-interacting protein kinase, confers salinity and cold tolerance in tobacco. *Physiologia Plantarum* **149**, 367–377.

Fowler S, Thomashow MF (2002) Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. *The Plant Cell* **14**, 1675–1690. doi:10.1105/tpc.003483

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644–652. doi:10.1038/nbt.1883

Guo Z, Tan J, Zhuo C, Wang C, Xiang B, Wang Z (2014) Abscisic acid, $H_2O_2$ and nitric oxide interactions mediated cold-induced S-adenosylmethionine synthetase in *Medicago sativa* subsp. *falcata* that confers cold tolerance through up-regulating polyamine oxidation. *Plant Biotechnology Journal* **12**, 601–612. doi:10.1111/pbi.12166

He X, Sambe MA, Zhuo C, Tu Q, Guo Z (2015) A temperature induced lipocalin gene from *Medicago falcata* (*MfTIL1*) confers tolerance to cold and oxidative stress. *Plant Molecular Biology* **87**, 645–654. doi:10.1007/s11103-015-0304-3

Hu R, Zhu X, Xiang S, Zhan Y, Zhu M, Yin H, Zhou Q, Zhu L, Zhang X, Liu Z (2016) Comparative transcriptome analysis revealed the genotype specific cold response mechanism in tobacco. *Biochemical and Biophysical Research Communications* **469**, 535–541. doi:10.1016/j.bbrc.2015.12.040

Kaye C, Guy CL (1995) Perspectives of plant cold tolerance: physiology and molecular responses. *Science Progress* **78**, 271–299.

Kosová K, Vítámvás P, Prášil IT (2014) Wheat and barley dehydrins under cold, drought, and salinity—what can LEA-II proteins tell us about plant stress response? *Frontiers of Plant Science* **5**, 343. doi:10.3389/fpls.2014.00343

Król A, Amarowicz R, Weidner S (2015) The effects of cold stress on the phenolic compounds and antioxidant capacity of grapevine (*Vitis vinifera* L.) leaves. *Journal of Plant Physiology* **189**, 97–104. doi:10.1016/j.jplph.2015.10.002

Lee BH, Henderson DA, Zhu JK (2005) The *Arabidopsis* cold-responsive transcriptome and its regulation by ICE1. *The Plant Cell* **17**, 3155–3175. doi:10.1105/tpc.105.035568

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323. doi:10.1186/1471-2105-12-323

Los DA, Murata N (2004) Membrane fluidity and its roles in the perception of environmental signals. *Biochimica et Biophysica Acta* **1666**, 142–157. doi:10.1016/j.bbamem.2004.08.002

NDong C, Danyluk J, Wilson KE, Pocock T, Huner NP, Sarhan F (2002) Cold-regulated cereal chloroplast late embryogenesis abundant-like proteins. Molecular characterization and functional analyses. *Plant Physiology* **129**, 1368–1381. doi:10.1104/pp.001925

Pennycooke JC, Cheng H, Stockinger EJ (2008) Comparative genomic sequence and expression analyses of *Medicago truncatula* and alfalfa subspecies *falcata COLD-ACCLIMATION-SPECIFIC* genes. *Plant Physiology* **146**, 1242–1254. doi:10.1104/pp.107.108779

Saijo Y, Hata S, Kyozuka J, Shimamoto K, Izui K (2000) Over-expression of a single $Ca^{2+}$-dependent protein kinase confers both cold and salt/drought

tolerance on rice plants. *The Plant Journal* **23**, 319–327. doi:10.1046/j.1365-313x.2000.00787.x

Schulz E, Tohge T, Zuther E, Fernie AR, Hincha DK (2016) Flavonoids are determinants of freezing tolerance and cold acclimation in *Arabidopsis thaliana*. *Scientific Reports* **6**, 34027. doi:10.1038/srep34027

Song L, Jiang L, Chen Y, Shu Y, Bai Y, Guo C (2016) Deep-sequencing transcriptome analysis of field-grown *Medicago sativa* L. crown buds acclimated to freezing stress. *Functional & Integrative Genomics* **16**, 495–511. doi:10.1007/s10142-016-0500-5

Streb P, Aubert S, Gout E, Bligny R (2003) Cold- and light-induced changes of metabolite and antioxidant levels in two high mountain plant species *Soldanella alpina* and *Ranunculus glacialis* and a lowland species *Pisum sativum*. *Physiologia Plantarum* **118**, 96–104. doi:10.1034/j.1399-3054.2003.00099.x

Suzuki N, Koussevitzky S, Mittler R, Miller G (2012) ROS and redox signalling in the response of plants to abiotic stress. *Plant, Cell & Environment* **35**, 259–270. doi:10.1111/j.1365-3040.2011.02336.x

Svensson JT, Crosatti C, Campoli C, Bassi R, Stanca AM, Close TJ, Cattivelli L (2006) Transcriptome analysis of cold acclimation in barley *albina* and *xantha* mutants. *Plant Physiology* **141**, 257–270. doi:10.1104/pp.105.072645

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* **106**, 411–422. doi:10.1007/s00122-002-1031-0

Thomashow MF (1999) Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. *Annual Review of Plant Physiology and Plant Molecular Biology* **50**, 571–599. doi:10.1146/annurev.arplant.50.1.571

Wang CT, Shao JM (2013) Characterization of the *ZmCK1* gene encoding a calcium-dependent protein kinase responsive to multiple abiotic stresses in maize. *Plant Molecular Biology Reporter* **31**, 222–230. doi:10.1007/s11105-012-0496-5

Wang L, Zhu W, Fang L, Sun X, Su L, Liang Z, Wang N, Londo JP, Li S, Xin H (2014) Genome-wide identification of *WRKY* family genes and their response to cold stress in *Vitis vinifera*. *BMC Plant Biology* **14**, 103. doi:10.1186/1471-2229-14-103

Winfield MO, Lu C, Wilson ID, Coghill JA, Edwards KJ (2010) Plant responses to cold: Transcriptome analysis of wheat. *Plant Biotechnology Journal* **8**, 749–771. doi:10.1111/j.1467-7652.2010.00536.x

Xu W, Zhang N, Jiao Y, Li R, Xiao D, Wang Z (2014) The grapevine basic helix-loop-helix (*bHLH*) transcription factor positively modulates *CBF*-pathway and confers tolerance to cold-stress in *Arabidopsis*. *Molecular Biology Reports* **41**, 5329–5342. doi:10.1007/s11033-014-3404-2

Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, Wang J (2006) WEGO: a web tool for plotting GO annotations. *Nucleic Acids Research* **34**, W293. doi:10.1093/nar/gkl031

Zhang Z, Li J, Zhao X, Wang J, Wong GK, Yu J (2006) KaKs_Calculator: Calculating Ka and Ks through model selection and model averaging. *Genomics, Proteomics & Bioinformatics* **4**, 259–263. doi:10.1016/S1672-0229(07)60007-2

Zhang Y, Yu H, Yang X, Li Q, Ling J, Wang H, Gu X, Huang S, Jiang W (2016) *CsWRKY46*, a WRKY transcription factor from cucumber, confers cold resistance in transgenic-plant by regulating a set of cold-stress responsive genes in an ABA-dependent manner. *Plant Physiology and Biochemistry* **108**, 478–487. doi:10.1016/j.plaphy.2016.08.013

Zhou J, Liang Y, Niu Q, Chen L, Zhang X, Ye D (2013) The *Arabidopsis* general transcription factor TFIIB1 (AtTFIIB1) is required for pollen tube growth and endosperm development. *Journal of Experimental Botany* **64**, 2205–2218. doi:10.1093/jxb/ert078

Zhou Y, Cheng Y, Yang Y, Li X, Supriyo B, Sun X, Yang Y (2016) Overexpression of *SpCBL6*, a calcineurin B-like protein of *Stipa purpurea*, enhanced cold tolerance and reduced drought tolerance in transgenic *Arabidopsis*. *Molecular Biology Reports* **43**, 957–966. doi:10.1007/s11033-016-4036-5

Zhuo C, Wang T, Guo Z, Lu S (2016) Overexpression of *MfPIP2-7* from *Medicago falcata* promotes cold tolerance and growth under $NO_3^-$ deficiency in transgenic tobacco plants. *BMC Plant Biology* **16**, 138. doi:10.1186/s12870-016-0814-4