Tree Physiology 31, 452–461
doi:10.1093/treephys/tpr015

Research paper

Genome-scale transcriptome analysis of the desert poplar, *Populus euphratica*

Qiang Qiu[1], Tao Ma[1], Quanjun Hu, Bingbing Liu, Yuxia Wu, Haihong Zhou, Qian Wang, Juan Wang and Jianquan Liu[2]

Molecular Ecology Group, Key Laboratory of Arid and Grassland Ecology, Lanzhou University, Lanzhou 730000, China; [1]These authors contributed equally to this work; [2]Corresponding author (liujq@nwipb.ac.cn)

Received October 6, 2010; accepted February 10, 2011; published online March 22, 2011; handling Editor Ron Sederoff

Populus euphratica is well-adapted to extreme desert environments and is an important model species for studying the effects of abiotic stresses on trees. Here we present the first deep transcriptomic analysis of this species. To maximize representation of conditional transcripts, mRNA was obtained from living tissues of desert-grown trees and two types of callus (salt-stressed and unstressed). *De novo* assembly generated 86,777 Unigenes using Solexa sequence data. These sequences covered 92% of previously reported *P. euphratica* expressed sequence tags (ESTs) and 90% of the TIGR poplar ESTs, and a total of 58,499 high-quality unique sequences were annotated by BLAST similarity searches against public databases. We found that 27% of the total Unigenes were differentially expressed (up- or down-regulated) in response to salt stress in *P. euphratica* callus. These differentially expressed genes are mainly involved in transport, transcription, cellular communication and metabolism. In addition, we found that numerous putative genes involved in ABA regulation and biosynthesis were also differentially regulated. This study represents the deepest transcriptomic and gene-annotation analysis of *P. euphratica* to date. The genetic knowledge acquired should be very useful for future studies of the molecular adaptation of this tree species to abiotic stress and facilitate genetic manipulation of other poplar species.

Keywords: abiotic stress, Illumina/Solexa sequencing, *Populus euphratica*, transcriptome.

Introduction

Salinity and drought are two major environmental stressors that inhibit tree growth in semiarid and arid areas (Boyer 1982). To survive in such areas, plants require complex responses to these abiotic stressors (Bohnert et al. 1995), involving signal transduction, regulation of gene expression, ionic homeostasis, scavenging of reactive oxygen species, accumulation of compatible solutes and growth regulation (e.g. Hasegawa et al. 2000, Seki et al. 2001, Bartels and Sunkar 2005, Yamaguchi-Shinozaki and Shinozaki 2006, Munns and Tester 2008). These responses have been partly elucidated in both trees and herbs through the identification and characterization of genes that are activated in response to abiotic stress (Ingram and Bartels 1996, Seki et al. 2002, Xiong et al. 2002, Zhu 2002, Shinozaki et al. 2003).

Populus euphratica Oliv. (Salicaceae) is naturally distributed in western China and adjacent Middle-Eastern countries (Browicz 1977). It plays an important role in maintaining local arid ecosystems (Ma et al. 1997) and can grow in deserts with extremely hot and dry summers, while other congeners are known to be drought sensitive (Wang et al. 2007, 2008, Ding et al. 2010). In addition, *in vitro* experiments have indicated that this species can tolerate salt concentrations up to 450 mM and mannitol concentrations up to 400 mM (Watanabe et al. 2000, Gu et al. 2004*b*). Thus, *P. euphratica* has been widely considered as a model species for elucidating abiotic resistance mechanisms of trees, e.g., responses to salinity or drought stress (Chen et al. 1997, 2009, Gu et al. 2004*a*, Ottow et al. 2005, Wu et al. 2007, Zhang et al. 2007, Sun et al. 2009, 2010, Ye et al. 2009, Zeng et al. 2009). In a recent

transcriptomal effort, Brosché et al. (2005) recovered 14,000 expressed sequence tags (ESTs) from normalized and subtracted cDNA libraries of normal and stress-exposed samples of *P. euphratica* trees using Sanger EST sequencing. This study yielded valuable information, but the known scale of the *P. trichocarpa* genome (Tuskan et al. 2006), and findings that EST sequences generated from Sanger sequencing do not cover the entire transcriptional profile (Hale et al. 2009) indicate that a minority of the *P. euphratica* genes were recovered.

Genome-scale transcript analysis aims to capture an unbiased view of the complete RNA transcript profile of a species (Wilhelm et al. 2008), allowing the transcriptional level of each gene in a given tissue at a given point in its life cycle to be monitored. The development of next-generation, massively parallel sequencing technologies (e.g., Illumina/Solexa-based RNA-Seq technology) has significantly improved the possible depth of transcriptome sequencing, with significantly reduced costs (Wang et al. 2009). Using this high-throughput RNA sequencing, it is possible to recover transcripts that are expressed at extremely low levels, and to isolate substantial numbers of novel transcripts that previous large-scale Sanger EST sequencing procedures could not identify (Morozova and Marra 2008, Wang et al. 2009). In addition, such sequencing-based methods detect absolute expression, rather than relative gene expression changes, and thus overcome many of the inherent limitations of microarray-based systems (Hoen et al. 2008, Sultan et al. 2008, Wilhelm and Landry 2009). This method is now being widely applied to non-model species for which genomic information is not available, and many ESTs and numerous novel transcripts are being recovered (Trick et al. 2009, Libault et al. 2010*a*, 2010*b*, Wu et al. 2010).

Here, we present a *de novo* assembly of the *P. euphratica* transcriptome using Solexa data. Our data were collected by sequencing cDNA libraries of living tissues from mature trees growing in the Talim Basin desert, which had a long period to adapt to the local conditions before salt-stressed callus and unstressed callus were sampled. We specially examined the gene expression dynamics of this species in response to salt stress and identified a core set of stress-related transcripts. The acquired information should facilitate attempts to elucidate response mechanisms of this species to abiotic stress and to develop stress-tolerant poplar trees through genetic manipulation of wild genetic resources.

Materials and methods

Plant material

Three sets of samples representing desert-grown trees, control-callus samples and salt-stressed callus samples were examined in this study. We collected three replicate samples (all of the same fresh weight) of roots, leaves, flower buds, flowers, xylem and phloem from two mature male *P. euphratica* trees and one mature female tree in the Talim Basin desert in Xinjiang. In addition, we cultivated calli using the method described by Zhang et al. (2004*a*), replaced the growth medium of one set with fresh, unamended medium and that of another set with fresh medium supplemented with 100 mM NaCl (to impose salt stress), and then harvested both sets 24 h later. All samples were rapidly stored at −80 °C until required for RNA extraction.

RNA extraction and quality determination

Total RNA was extracted three times from each of the sample sets, using a CTAB procedure (see Chang et al. 1993). A_{260}/A_{280} ratios of the RNA samples dissolved in 10 mm Tris (pH 7.6) ranged from 1.9 to 2.1. The integrity of the RNA samples was examined with an Agilent 2100 Bioanalyzer and their RIN (RNA integrity number) values ranged from 8.6 to 10.0, with no sign of degradation. RNA from each replicate was pooled (in equal volumes) to obtain a single RNA sample for cDNA preparation and RNA-Seq, and equal amounts of mRNA from different tissues of the desert-grown trees were pooled to make single samples.

Illumina cDNA library preparation and sequencing

For cDNA synthesis and Solexa sequencing, 20 μg of total RNA was used, at a concentration of ≥400 ng/μl. Poly(A) mRNA was first purified using beads with oligo(dT). Then, the mRNA was fragmented into small pieces using divalent cations at an elevated temperature. Based on these cleaved RNA fragments, we used random hexamer-primer and reverse transcriptase (Invitrogen) to synthesize first-strand cDNA. Second-strand cDNA was synthesized using RNase H (Invitrogen) and DNA polymerase I (New England BioLabs). We constructed three paired-end cDNA libraries with insert sizes of 200 bp, and then sequenced the cDNA using an Illumina (San Diego, CA, USA) Genome Analyzer according to the manufacturer's protocols with a read length of 75 bp.

De novo assembly and assessment

Reads from each library were assembled separately. Adapter sequences and reads containing too many (>8) unknown bases or low-quality bases (>50% of the bases with a quality score ≤5) were filtered using in-house Perl scripts. The average proportion of clean reads for each library was ~96.5%. The transcriptome sequence was assembled into distinct contigs with short reads by SOAPdenovo software (Li et al. 2010) (http://soap.genomics.org.cn), which applies the de Bruijn graph data structure to construct contigs. The reads were then realigned to the contig sequences, and the paired-end relationships between the reads were used to construct scaffolds between contigs. To fill the intra-scaffold gaps, we then used the paired-end information to retrieve read pairs that had one read well-aligned on the contigs and another read located in

the gap region, and then locally assembled the collected reads. After gap closure, we constructed a non-redundant Unigene set from all three assembled datasets using the EST assembly program TGICL (Pertea et al. 2003).

To obtain high-quality sequences for further annotation and analysis, we excluded Unigene sequences that might represent non-coding RNAs, identified by comparing our sequences with known non-coding RNAs deposited in the Rfam database (http://www.sanger.ac.uk/Software/Rfam/, release 10.0). Unigene sequences assigned to microbial (MBGD: http://mbgd.genome.ad.jp/), fungal and virus (based on data downloaded from the NCBI database) sources were also filtered out. In addition, sequences for which >50% of the bases aligned with sequences in UTRdb (http://utrdb.ba.itb.cnr.it/) and/or contained <200 non-UTR bases were excluded.

In order to assess the sequence coverage of the transcriptome assemblies, we searched our Unigene sequences by BLAST (Altschul et al. 1997) against (i) 13,845 unassembled *P. euphratica* ESTs from GenBank; (ii) 172,068 assembled ESTs from 14 poplar species in the TIGR Plant Transcript Assemblies database (Childs et al. 2007) (accessed August 2009; http://plantta.tigr.org/index.shtml); (iii) *P. trichocarpa* protein-coding transcript and genome sequences (JGI release 2.0; http://www.phytozome.net/poplar.php); and (iv) all available expressed sequences identified from 254 plant species in the TIGR Plant Transcript Assemblies database.

Functional annotation

We annotated sequences based on a set of sequential BLAST searches (Altschul et al. 1997) designed to find the most descriptive annotation for each sequence. The assembled unique transcripts were compared with sequences in GenBank's non-redundant database using the BLASTN algorithm, the GI accessions of best hits were retrieved, and the GO accessions were mapped to GO terms according to molecular function, biological process and cellular component ontologies (http://www.geneontology.org/). The remaining sequences that putatively encoded proteins were searched against the Swiss-Prot protein database (http://www.expasy.ch/sprot), the KEGG pathways database (Kanehisa et al. 2008) and the COG database (http://www.ncbi.nlm.nih.gov/COG/), applying a typical *E*-value cutoff level of <1E – 5.

Analysis of differential EST expression

Gene expression levels were measured in the RNA-Seq analyses as numbers of reads per kilobase of exon region in a given gene per million mapped reads (RPKM) (Mortazavi et al. 2008). To identify genes regulated by salt stress, we determined the number of reads for each coding region in the control and salt-stress callus libraries (after normalization to account for the difference in number of total reads), and then calculated the ratio of reads in the two libraries. The statistical significance of the differential expression value for each gene was determined using the method described by Audic and Claverie (1997), and the results of all statistical tests were corrected for multiple testing with the Benjamini–Hochberg false discovery rate (FDR). Sequences were deemed to be significantly differentially expressed if the adjusted *P* value obtained by this method was <0.001 and there was at least a twofold change (>1 or <– 1 in log 2 ratio value) in sequence count between two libraries.

To ascertain the biological significance of the detected differences in gene expression profiles, ESTs with significant expression differences under salt stress were assigned to functional classifications according to corresponding *Arabidopsis* locus identifiers by BLASTX to *Arabidopsis* gene models (TAIR ver. 9, http://www.arabidopsis.org/), applying an *E*-value cutoff level of E – 10. The data were then entered into a functional catalogue according to the Munich Information Center for Protein Sequence (MIPS) classification scheme (http://www.helmholtz-muenchen.de/en/ibis; see Table S4 available as Supplementary Data at Tree Physiology Online for a complete list).

Results and discussion

De novo assembly and quantitative assessment of the Illumina ESTs

After removing low-quality sequences and trimming adapter sequences, totals of 28 million, 55 million and 57 million clean reads of 75 bp were generated from the desert-grown tree, control-callus and salt-stressed callus cDNA libraries, respectively, in the Illumina GA runs (Table 1). The total length of the reads was >10.6 gigabases (Gb), equivalent to ~25-fold coverage of a genome of *P. trichocarpa* size. *De novo* assembly was carried out by SOAPdenovo, a genome assembly program developed specifically for next-generation short-read sequences (Li et al. 2010). The average contig size exceeded 170 bp in all three libraries. After using paired-end information to join the contigs into scaffolds and local assembly, we generated 44,593 scaffolds for desert-grown trees, 71,876 scaffolds for control-callus samples and 63,655 scaffolds for salt-stressed callus samples, with average lengths of 452, 591 and 589 nt, respectively. These scaffolds were assembled into 94,196 Unigenes by TGICL clustering tools. The sequencing throughout ranged from 1 to 24,406 × (average 96×), and the average length of the Unigenes was 671 bp (Table 1). The size distribution and gap ratio of the assembly is shown in Figures S1 and S2 available as Supplementary Data at Tree Physiology Online. We excluded possible non-coding RNA, untranslated region, microbial, fungal and virus sequences identified by comparing our Unigene sequences against entries in databases listed in the Materials and methods section. Finally, we identified a total of 86,777 high-quality Unigene sequences when 7419 possibly polluted sequences were excluded. All Unigene sequences

Table 1. Overview of the sequencing and assembly.

Sequences	Control callus	Salt-stressed callus	Desert-grown trees
Total number of clean sequences	55,112,298	57,974,944	28,272,830
Total clean nucleotides (nt)	4,133,422,350	4,348,120,800	2,092,189,420
Contig			
Number of contigs	227,855	202,350	151,930
Length of all contigs (nt)	51,630,167	45,123,631	25,976,461
Average contig size	227	223	171
Range of contig length	75–5234	75–6857	75–2123
Scaffolds >200 bp			
Number of scaffolds	71,876	63,655	44,593
Length of all scaffolds (nt)	42,468,437	37,474,605	20,170,369
Average scaffold sizes	591	589	452
Range of scaffold lengths	200–7218	200–8351	200–3685
Gap% (*n*/size)	0.9	1.2	3.3
Unigene sequences >200 bp			
Number of Unigenes	94,196		
Length of all Unigenes (nt)	63,208,033		
Average Unigene size	671		
Range of Unigene length	200–8514		
Mean throughout[1]	96		

[1]Base number of mapped reads/Unigene sequence length.

obtained in this study can be accessed through File 1 available as Supplementary Data at Tree Physiology Online.

A BLAST comparison revealed that the assembly contains 92% of the NCBI *P. euphratica* ESTs and 90% of the TIGR poplar ESTs. In addition, we compared the recovered *P. euphratica* transcripts against *P. trichocarpa* transcripts from the genome assembly (Tuskan et al. 2006). Homologues of 40,249 sequences (~88% of the *P. trichocarpa* transcripts) were represented in our assembled transcriptome (Table 2). These results suggest that our assembly covered >90% of all *P. euphratica* transcripts.

Functional annotation and characterization of transcripts

The entire Unigene sets were then annotated on the basis of similarities to known or putative sequences in the public databases. Using the best hits found by BLAST, an inferred putative function was assigned to the sequences and they were sorted into major functional categories (Figure 1). Among the 86,777 high-quality unique sequences, 58,499 (67.4%) had at least one significant match to an existing gene model in BLASTX searches (see Table S1 available as Supplementary Data at Tree Physiology Online). The remaining 28,278 (32.6%) unannotated Unigene sequences were further compared against the *P. trichocarpa* gene and genome sequences, and all available ESTs identified from 254 plant species. Using an *E*-value cutoff of <E – 10, at least 65 sequences did not show any significant match to known genes longer than 400 bp (see Table S2 available as Supplementary Data at Tree Physiology Online) and may be *P. euphratica*-specific novel ESTs.

The 100 most abundant transcripts distinctly differed among the three samples. In control callus, genes involved in auxin signaling, cell division and biogenesis were highly expressed, while stress-responsive genes (such as several antiporter, stress-induced transcription factor and diverse protease encoding genes) were strongly expressed in the salt-stressed callus. By contrast, the desert-growing trees strongly expressed photosynthesis-related genes and stress-related proteins, for example, aquaporin and glutathione transferase (see Table S3 available as Supplementary Data at Tree Physiology Online). Together, the recovered expression profiles largely support the phenotypic and physiological characterization of our three sample types.

Transcript differences between control and salt-stressed callus

According to the applied criteria (twofold or more change and $P < 0.001$), 23,512 ESTs were identified as differentially expressed between the 24-h salt-stressed callus and control-callus samples: 7109 up-regulated and 16,403 down-regulated (Figure 2). This 24-h salt-stress treatment had apparently modified the expression of almost 27% of the total Unigenes, in accordance with previous observations in *Arabidopsis* (Kreps et al. 2002). Therefore, it is obvious that *P. euphratica* responds to abiotic stress by moderate reprogramming of its

Table 2. Assessment of sequence coverage by the Unigene assembly based on EST and *P. trichocarpa* gene model.

Sequence collection	Number	Total length (bp)	Average length (bp)	Base coverage (%)	Total aligned sequences
P. euphratica EST	13,845	6,624,164	478	81	12,718 (92%)
Populus EST	172,068	100,558,182	584	79	155,210 (90%)
P. trichocarpa CDS	45,778	51,551,039	1,126	85	40,249(88%)

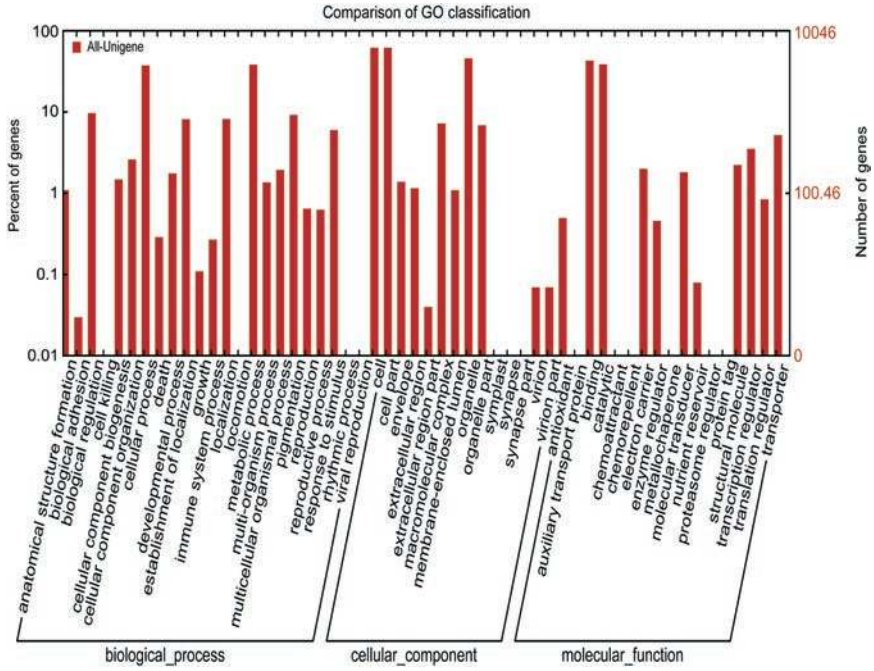

Figure 1. GO categories of the *P. euphratica* Unigenes. The percentage and total number of Unigenes in each category is shown.

transcriptome. Based on the MIPS functional catalogue, the 5775 differentially regulated ESTs with well-functional annotations and classifications were grouped into 18 main functional categories (Figure 3 and Table S5 available as Supplementary Data at Tree Physiology Online), the most heavily represented being cellular transport (10%), transcription (8%), cellular communication (10%) and metabolism (23%). Four important expression characteristics were further identified from the following detailed analyses.

Firstly, a total of 579 differentially regulated ESTs were categorized as transporters, many of which (246) were strongly up-regulated in response to salt stress (see Table S5 available as Supplementary Data at Tree Physiology Online). For example, the expressions of two EST sequences (Unigene45498_All and Unigene24445_All), which are homologous to *Arabidopsis* potassium ion transmembrane transporter (HAK5) and Na^+/H^+ antiporter (NHX1) genes, were up-regulated 48-fold and 66-fold, respectively. The 100 most abundant transcripts under salt stress encompassed sequences encoding 17 transport-related proteins, including sodium and potassium ion transmembrane transporters, and chloride channel and ABC transporters (see Table S3 available as Supplementary Data at Tree Physiology Online), corroborating previous findings (Wang et al. 2008, Sun et al. 2009, Ye et al. 2009). These transporters, which are important for maintaining and re-establishing homeostasis of the cytoplasm, are induced and highly expressed in response to salinity stress (Hasegawa et al. 2000).

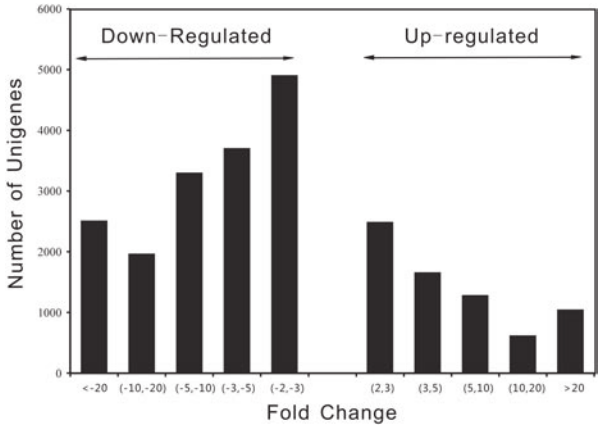

Figure 2. Distribution of transcript changes in salt-stressed callus compared with control callus. The distribution includes those genes showing a two fold or more increase (up-regulated) or decrease (down-regulated) in RPKM change, with an adjusted *P* value of 0.001 or less. A complete list of these ESTs is provided in Table S4 available as Supplementary Data at Tree Physiology Online.

Secondly, there were also considerable changes in the expression of genes involved in transcription and cellular communication. We found 454 differentially regulated ESTs involved in transcription: 132 up-regulated and 322 down-regulated. Five hundred and ninety-six of the differentially expressed ESTs were involved in cellular communication: 166 up-regulated and 430 down-regulated (see Table S5 available as Supplementary Data at Tree Physiology Online). Several of the transcription factors have homologues, e.g. *AP2-EREBP* and *bZIP*, which are known to be stress-induced in model herb species (*Arabidopsis* and rice) (Novillo et al. 2004, Oh et al. 2005). In addition, two transcription factors (*PeDREB2* and *PeSCL7*) have been isolated from *P. euphratica*, and their over-expression has been recently shown to increase drought- and salt-stress tolerance in transgenic

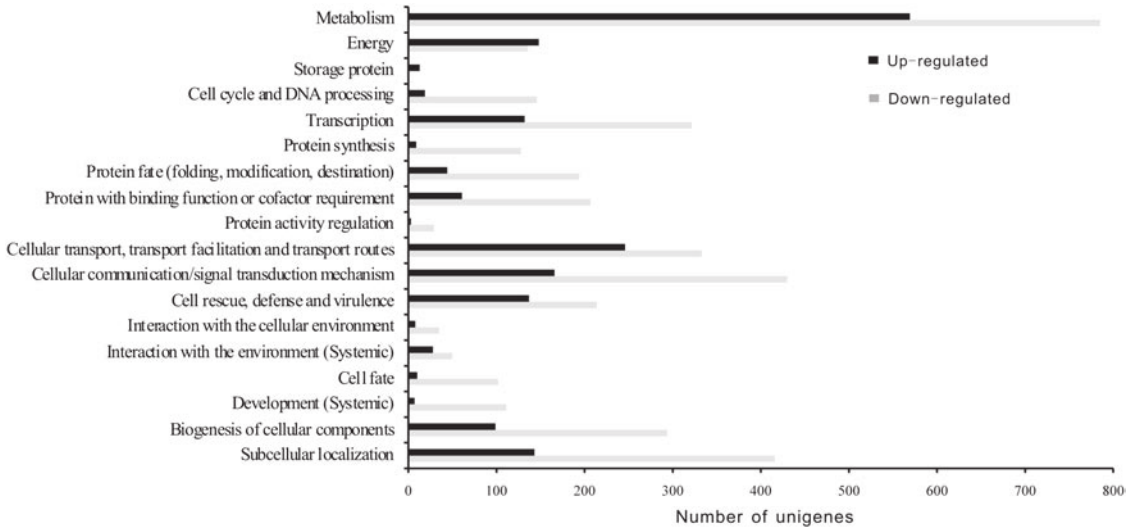

Figure 3. Functional categories of stress-responsive ESTs. The stress-responsive ESTs were assigned to main functional categories based on the *Arabidopsis* MIPS classification scheme. A complete functional classification is presented in Table S5 available as Supplementary Data at Tree Physiology Online.

plants (Chen et al. 2009, Ma et al. 2010). However, there have been no functional analyses of homologues for a greater number of the transcription factors (e.g. *AP2-EREBP11* and *bZIP53*, homologues of Unigene45134_All, Unigene3935_All, respectively) that were up-regulated >60-fold in our salt-stressed callus samples. Functional analyses of these transcription factors (Figure 4 and see Table S6 available as Supplementary Data at Tree Physiology Online) should provide more information on the complex regulatory networks involved in responses of trees to salt stress (Yamaguchi-Shinozaki and Shinozaki 2006).

Thirdly, to obtain a better understanding of the functional significance of the differentially regulated ESTs, we focused on

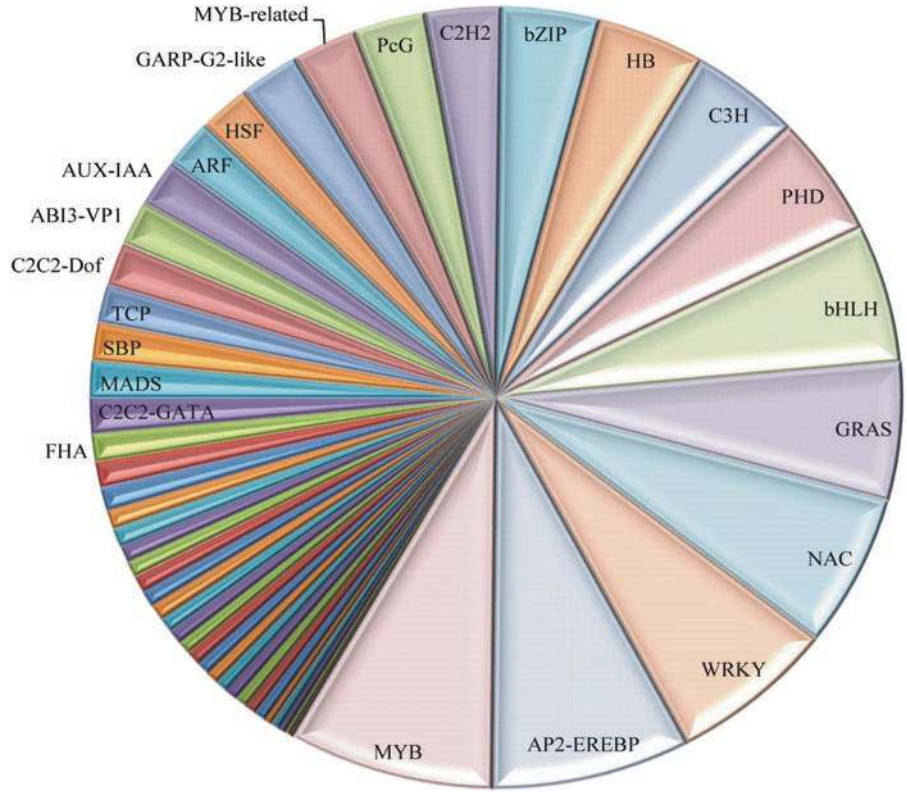

Figure 4. Distribution of identified *P. euphratica* transcription factors in gene families.

those that play important roles in well-characterized metabolic pathways. Key aspects of the salt-stress responses of *P. euphratica* callus cultured *in vitro* include the accumulation of proline and sugars, which provide osmotic balance (Watanabe et al. 2000). We found that the expression of genes related to 40 metabolic pathways was significantly changed under salt stress ($P < 0.05$), including genes involved in carbohydrate, amino acid, energy, lipid, secondary metabolite, cofactor and vitamin, terpenoid and polyketide metabolism (Table 3). Therefore, *P. euphratica* may establish a new energetic and

Table 3. Statistical enrichment analysis for KEGG metabolic pathways.

Metabolism pathway	Pathway ID	Number			FDR *P* value
		Background	Up	Down	
Carbohydrate metabolism					
Starch and sucrose metabolism	ko:00500	724	92	275	5.35E–14
Pentose and glucuronate interconversions	ko:00040	204	30	98	5.35E–14
Ascorbate and aldarate metabolism	ko:00053	177	11	83	9.75E–08
Amino sugar and nucleotide sugar metabolism	ko:00520	295	33	88	5.49E–03
Amino acid metabolism					
Phenylalanine metabolism	ko:00360	267	98	56	5.35E–14
Tyrosine metabolism	ko:00350	165	51	31	2.35E–05
Cysteine and methionine metabolism	ko:00270	330	69	77	2.79E–05
Alanine, aspartate and glutamate metabolism	ko:00250	185	57	31	7.97E–05
Metabolism of other amino acids					
Cyanoamino acid metabolism	ko:00460	215	39	62	4.09E–05
Glutathione metabolism	ko:00480	199	52	31	1.66E–02
Biosynthesis of other secondary metabolites					
Phenylpropanoid biosynthesis	ko:00940	567	157	132	5.35E–14
Flavonoid biosynthesis	ko:00941	309	79	70	5.19E–08
Stilbenoid, diarylheptanoid and gingerol biosynthesis	ko:00945	282	83	55	5.53E–08
Isoquinoline alkaloid biosynthesis	ko:00950	61	27	8	3.53E–04
Flavone and flavonol biosynthesis	ko:00944	95	25	21	4.94E–03
Glucosinolate biosynthesis	ko:00966	63	14	19	5.06E–03
Tropane, piperidine and pyridine alkaloid biosynthesis	ko:00960	82	25	14	1.56E–02
Indole alkaloid biosynthesis	ko:00901	30	9	8	2.37E–02
Energy metabolism					
Nitrogen metabolism	ko:00910	207	66	61	5.35E–14
Methane metabolism	ko:00680	181	51	45	7.90E–08
Sulphur metabolism	ko:00920	93	34	8	2.78E–02
Lipid metabolism					
Steroid biosynthesis	ko:00100	96	14	51	2.86E–11
Linoleic acid metabolism	ko:00591	94	20	23	2.11E–02
Synthesis and degradation of ketone bodies	ko:00072	13	9	0	3.16E–02
Metabolism of cofactors and vitamins					
Retinol metabolism	ko:00830	81	23	25	5.58E–06
Ubiquinone and other terpenoid-quinone biosynthesis	ko:00130	117	42	13	4.23E–03
Metabolism of terpenoids and polyketides					
Zeatin biosynthesis	ko:00908	65	20	23	2.93E–07
Brassinosteroid biosynthesis	ko:00905	55	21	13	5.64E–05
Limonene and pinene degradation	ko:00903	286	65	58	5.74E–04
Carotenoid biosynthesis	ko:00906	165	37	32	2.72E–02
Diterpenoid biosynthesis	ko:00904	80	20	16	4.72E–02
Xenobiotics biodegradation and metabolism					
Metabolism of xenobiotics by cytochrome P450	ko:00980	129	50	29	2.72E–10
Drug metabolism—cytochrome P450	ko:00982	139	53	30	4.88E–10
Gamma-hexachlorocyclohexane degradation	ko:00361	245	68	50	1.54E–06
Naphthalene and anthracene degradation	ko:00626	226	62	43	4.75E–05
3-Chloroacrylic acid degradation	ko:00641	79	19	20	7.05E–03
Fluorobenzoate degradation	ko:00364	25	9	6	2.10E–02
Fluorene degradation	ko:00628	23	9	5	2.21E–02
1,1,1-Trichloro-2,2-bis(4-chlorophenyl)ethane (DDT) degradation	ko:00351	23	9	5	2.21E–02
1- and 2-Methylnaphthalene degradation	ko:00624	66	17	14	3.86E–02

developmental equilibrium under salt stress (Bartels and Sunkar 2005, Ottow et al. 2005).

Finally, we found that numerous genes involved in abscisic acid (ABA) signalling regulation and biosynthesis were significantly up- or down-regulated in response to salt stress. This was consistent with expectations, since ABA plays crucial roles in drought and high-salinity stress responses, as well as diverse aspects of plant growth and development (Zhu 2002, Galvez-Valdivieso et al. 2009, Raghavendra et al. 2010). Furthermore, its concentration is known to increase during salt stress in *P. euphratica* (Chen et al. 2001, 2002). Three key ABA-binding receptors have been identified for pathways regulated by ABA: the flowering-time control protein (FCA), the Mg-chelatase H subunit and the G-protein-coupled receptor (GCR2) (Hirayama and Shinozaki 2007). Our results showed that homologues of genes encoding two of these receptors—*FCA* (Unigene484_All) and Mg-chelatase subunit H (Unigene2835_All)—were highly expressed in salt-stressed callus. Phosphatidic acid (PA), an important second messenger, is produced by phospholipase D (PLD) upon ABA treatment (Mishra et al. 2006), and two ESTs (Unigene7349_All and Unigene16355_All) that putatively encode homologues of *Arabidopsis* PLDα1 and PLDβ1 were also significantly up-regulated in our salt-stressed callus sample. Furthermore, two protein-phosphatase 2C (PP2C)-like proteins encoded by ABA-insensitive loci ABI1 and ABI2 negatively regulate ABA responses in *Arabidopsis* (Zhang et al. 2004*b*), and we found that 46 ESTs homologous to *Arabidopsis PP2C* genes were differentially regulated: 20 were significantly up-regulated and 26 significantly down-regulated. Recent studies have also revealed connections between PP2C and SNF1-related kinases (SnRKs), particularly that SnRK2- and SnRK3-type kinases are important ABA regulators (Kim et al. 2003, Fujii et al. 2007). Accordingly, we found that one (Unigene38424_All) and two (Unigene14960_All, Unigene7553_All) ESTs that encode putative SnRK2.3- and SnRK3-type kinases, respectively, were up-regulated by salt stress.

The first step of the ABA-specific synthetic pathway is the conversion of zeaxanthin to violaxanthin, which is catalysed by zeaxanthin epoxidase (ZEP)—the first enzyme to be identified as an ABA biosynthetic enzyme (Seo and Koshiba 2002), and lack of which leads to ABA deficiency in mutants such as *Arabidopsis thaliana aba1* (Duckham et al. 1991). Our results showed that a possible *AtZEP* homologue (Unigene4188_All) was highly induced under salt stress. Another enzyme that plays a key role in ABA biosynthesis, 9-*cis*-epoxycarotenoid dioxygenase (NCED) (Nambara and Marion-Poll 2005), has been found to improve drought and salt tolerance by increasing endogenous ABA levels (Iuchi et al. 2001). Our results clearly indicated that one EST (Unigene21682_All), putatively homologous to *AtNCED3*, was also up-regulated under salt stress. Hence, the differential expression profile indicates that key components of the ABA signalling and synthesis pathways are induced during salt stress, confirming the close relationship between ABA responses and salt stress in *P. euphratica*. Such a correlation is also consistent with the observed statistically significant enrichment of genes involved in KEGG metabolic pathways (Table 3), because carotenoid metabolism provides precursors for ABA biosynthesis (Seo and Koshiba 2002). However, the detailed mechanisms whereby components of ABA pathways participate in salt-stress responses in *P. euphratica* remain to be further elucidated.

Conclusions and perspectives

The development of new sequencing technologies in the past decade has provided opportunities for genome-wide transcriptomic analysis of non-model plants (Trick et al. 2009, Libault et al. 2010*a*, 2010*b*, Wu et al. 2010). Our results provide an extensive catalogue of the genes expressed in *P. euphratica*, and confirm that the applied sequence and assembly approach provides coverage of sufficient breadth and depth for potent transcriptome characterization and gene annotation, especially in non-model tree species. We have now increased the recorded number of EST sequences (13,845) sixfold for this species (Brosché et al. 2005). In addition, we have identified numerous potential stress-induced transporters, transcription factors and various components of ABA biosynthesis and signalling pathways that may be involved in the adaptation of this species to extreme environments (Chen et al. 2002, 2009, Umezawa et al. 2006, Wu et al. 2007). These genetic findings should be very useful for future studies of the molecular adaptation of this tree species to abiotic stress and genetic manipulation of other poplar species.

Funding

Financial support was provided by the National High-Tech Research and Development Program (863 Program) of China (grant no. 2009AA10Z101) and the Key Innovation Project of Ministry of Education of China (grant to L.J.Q.).

Supplementary data

Supplementary data for this article are available at *Tree Physiology* Online.

References

Altschul, S.F., T.L. Madden, A.A. Schaffer et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Audic, S. and J.M. Claverie. 1997. The significance of digital gene expression profiles. Genome Res. 7:986–995.

Bartels, D. and R. Sunkar. 2005. Drought and salt tolerance in plants. Crit. Rev. Plant Sci. 24:23–58.

Bohnert, H.J., D.E. Nelson and R.G. Jensen. 1995. Adaptations to environmental stresses. Plant Cell 7:1099–1111.

Boyer, J.S. 1982. Plant productivity and environment. Science 218:443–448.

Brosché, M., B. Vinocur, E.R. Alatalo et al. 2005. Gene expression and metabolite profiling of *Populus euphratica* growing in the Negev desert. Genome Biol. 6:R101–R117.

Browicz, K. 1977. Chorology of *Populus euphratica* Olivier. Arboretum Kornickie 22:5–27.

Chang, S., J. Puryear and J. Cairney. 1993. A simple and efficient method for isolating RNA from pine trees. Plant Mol. Biol. Rep. 11:113–116.

Chen, J., X. Xia and W. Yin. 2009. Expression profiling and functional characterization of a DREB2-type gene from *Populus euphratica*. Biochem. Biophys. Res. Commun. 378:483–487.

Chen, S., S. Wang, A. Altman and A. Hüttermann. 1997. Genotypic variation in drought tolerance of poplar in relation to abscisic acid. Tree Physiol. 17:797–803.

Chen, S., J. Li, S. Wang, A. Hüttermann and A. Altman. 2001. Salt, nutrient uptake and transport, and ABA of Populus euphratica; a hybrid in response to increasing soil NaCl. Trees 15:186–194.

Chen, S., J. Li, T. Wang, S. Wang, A. Polle and A. Hüttermann. 2002. Osmotic stress and ion-specific effects on xylem abscisic acid and the relevance to salinity tolerance in poplar. J. Plant Growth Regul. 21:224–233.

Childs, K.L., J.P. Hamilton, W. Zhu et al. 2007. The TIGR Plant Transcript Assemblies database. Nucleic Acids Res. 35:D846–D851.

Ding, M., P. Hou, X. Shen et al. 2010. Salt-induced expression of genes related to Na(+)/K(+) and ROS homeostasis in leaves of salt-resistant and salt-sensitive poplar species. Plant Mol. Biol. 73:251–269.

Duckham, S.C., R.S.T. Linforth and I.B. Taylor. 1991. Abscisic-acid deficient mutants at the aba gene locus of *Arabidopsis thaliana* are impaired in the epoxidation of zeaxanthin. Plant, Cell Environ. 14:601–606.

Fujii, H., P.E. Verslues and J.K. Zhu. 2007. Identification of two protein kinases required for abscisic acid regulation of seed germination, root growth, and gene expression in *Arabidopsis*. Plant Cell 19:485–494.

Galvez-Valdivieso, G., M.J. Fryer, T. Lawson et al. 2009. The high light response in *Arabidopsis* involves ABA signaling between vascular and bundle sheath cells. Plant Cell 21:2143–2162.

Gu, R.S., S. Fonseca, L.G. Puskás, L. Hackler Jr, A. Zvara, D. Dudits and M.S. Pais. 2004*a*. Transcript identification and profiling during salt stress and recovery of *Populus euphratica*. Tree Physiol. 124:265–276.

Gu, R.S., Q.L. Liu, D. Pei and X.N. Jiang. 2004*b*. Understanding saline and osmotic tolerance of *Populus euphratica* suspended cells. Plant Cell Tissue Organ Cult. 78:261–265.

Hale, M.C., C.R. McCormick, J.R. Jackson and J.A. Dewoody. 2009. Next-generation pyrosequencing of gonad transcriptomes in the polyploid lake sturgeon (*Acipenser fulvescens*): the relative merits of normalization and rarefaction in gene discovery. BMC Genomics 10:203.

Hasegawa, P.M., R.A. Bressan, J.K. Zhu and H.J. Bohnert. 2000. Plant cellular and molecular responses to high salinity. Annu. Rev. Plant Physiol. Plant Mol. Biol. 51:463–499.

Hirayama, T. and K. Shinozaki. 2007. Perception and transduction of abscisic acid signals: keys to the function of the versatile plant hormone ABA. Trends Plant Sci. 12:343–351.

Hoen, P.A., Y. Ariyurek, H.H. Thygesen et al. 2008. Deep sequencing-based expression analysis shows major advances in robustness, resolution and interlab portability over five microarray platforms. Nucleic Acids Res. 36:141–151.

Ingram, J. and D. Bartels. 1996. The molecular basis of dehydration tolerance in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 47:377–403.

Iuchi, S., M. Kobayashi, T. Taji et al. 2001. Regulation of drought tolerance by gene manipulation of 9-cis-epoxycarotenoid dioxygenase, a key enzyme in abscisic acid biosynthesis in *Arabidopsis*. Plant J. 27:325–333.

Kanehisa, M., M. Araki, S. Goto et al. 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 36:D480–D484.

Kim, K.N., Y.H. Cheong, J.J. Grant, G.K. Pandey and S. Luan. 2003. CIPK3, a calcium sensor-associated protein kinase that regulates abscisic acid and cold signal transduction in *Arabidopsis*. Plant Cell 15:411–423.

Kreps, J.A., Y.J. Wu, H.S. Chang, T. Zhu, X. Wang and J.F. Harper. 2002. Transcriptome changes for *Arabidopsis* in response to salt, osmotic, and cold stress. Plant Physiol. 130:2129–2141.

Li, R., H. Zhu, J. Ruan et al. 2010. *De novo* assembly of human genomes with massively parallel short read sequencing. Genome Res. 20:265–272.

Libault, M., A. Farmer, L. Brechenmacher et al. 2010*a*. Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection. Plant Physiol. 152:541–552.

Libault, M., A. Farmer, T. Joshi, K. Takahashi, R.J. Langley and L.D. Franklin. 2010*b*. An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. Plant J. 63:86–99.

Ma, C., L. Fung, S. Wang and A. Huttermann. 1997. Photosynthetic response of *Populus euphratica* to salt stress. For. Ecol. Manage. 93:56–61.

Ma, H.S., D. Liang, P. Shuai, X.L. Xia and W.L. Yin. 2010. The salt- and drought-inducible poplar GRAS protein SCL7 confers salt and drought tolerance in *Arabidopsis thaliana*. J. Exp. Bot. 61:4011–4019.

Mishra, G., W. Zhang, F. Deng, J. Zhao and X. Wang. 2006. A bifurcating pathway directs abscisic acid effects on stomatal closure and opening in *Arabidopsis*. Science 312:264–266.

Morozova, O. and M.A. Marra. 2008. Applications of next-generation sequencing technologies in functional genomics. Genomics 92:255–264.

Mortazavi, A., B.A. Williams, K. McCue, L. Schaeffer and B. Wold. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5:621–628.

Munns, R. and M. Tester. 2008. Mechanisms of salinity tolerance. Annu. Rev. Plant Biol. 59:651–681.

Nambara, E. and A. Marion-Poll. 2005. Abscisic acid biosynthesis and catabolism. Annu. Rev. Plant Biol. 56:165–185.

Novillo, F., J.M. Alonso, J.R. Ecker and J. Salinas. 2004. CBF2/DREB1C is a negative regulator of CBF1/DREB1B and CBF3/DREB1A and plays a central role in stress tolerance in *Arabidopsis*. Proc. Natl Acad. Sci. USA 101:3985–3990.

Oh, S.J., S.I. Song, Y.S. Kim et al. 2005. *Arabidopsis* CBF3/DREB1A and ABF3 in transgenic rice increased tolerance to abiotic stress without stunting growth. Plant Physiol. 138:341–351.

Ottow, E.A., M. Brinker, T. Teichmann et al. 2005. *Populus euphratica* displays apoplastic sodium accumulation, osmotic adjustment by decreases in calcium and soluble carbohydrates, and develops leaf succulence under salt stress. Plant Physiol. 139:1762–1772.

Pertea, G., X. Huang, F. Liang et al. 2003. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19:651–652.

Raghavendra, A.S., V.K. Gonugunta, A. Christmann and E. Grill. 2010. ABA perception and signaling. Trends Plant Sci. 15:395–401.

Seki, M., M. Narusaka, H. Abe, M. Kasuga, K. Yamaguchi-Shinozaki, P. Carninci, Y. Hayashizaki and K. Shinozaki. 2001. Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses by using a full-length cDNA microarray. Plant Cell 13:61–72.

Seki, M., M. Narusaka, J. Ishida et al. 2002. Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. Plant J. 31:279–292.

Seo, M. and T. Koshiba. 2002. Complex regulation of ABA biosynthesis in plants. Trends Plant Sci. 7:41–48.

Shinozaki, K., K. Yamaguchi-Shinozaki and M. Seki. 2003. Regulatory network of gene expression in the drought and cold stress responses. Curr. Opin. Plant Biol. 6:410–417.

Sultan, M., M.H. Schulz, H. Richard et al. 2008. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science 321:956–960.

Sun, J., S. Chen and S. Dai. 2009. NaCl-included alternations of cellular and tissue ion fluxes in roots of salt-resistant and salt-sensitive poplar species. Plant Physiol. 149:1141–1153.

Sun, J., M.J. Wang, M.Q. Ding et al. 2010. H(2)O(2) and cytosolic Ca(2+) signals triggered by the PM H(+)-coupled transport system mediate K(+)/Na(+) homeostasis in NaCl-stressed *Populus euphratica cells*. Plant Cell Environ. 33:943–958.

Trick, M., Y. Long, J. Meng and Bancroft I. 2009. Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing. Plant Biotechnol. J. 7:334–346.

Tuskan, G.A., S. DiFazio, S. Jansson et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313:1596–1604.

Umezawa, T., M. Fujita, Y. Fujita, K. Yamaguchi-Shinozaki and K. Shinozaki. 2006. Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future. Curr. Opin. Biotechnol. 17:113–122.

Wang, R.G., S.L. Chen, L. Deng, E. Fritz, A. Hüttermann and A. Polle. 2007. Leaf photosynthesis, fluorescence response to salinity and the relevance to chloroplast salt compartmentation and anti-oxidative stress in two poplars. Trees 21:581–591.

Wang, R.G., S.L. Chen, X.Y. Zhou et al. 2008. Ionic homeostasis and reactive oxygen species control in leaves and xylem sap of two poplars subjected to NaCl stress. Tree Physiol. 28:947–957.

Wang, Z., M. Gerstein and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10:57–63.

Watanabe, S., K. Kojima, Y. Ide and S. Sasaki. 2000. Effects of saline and osmotic stress on proline and sugar accumulation in *Populus euphratica in vitro*. Plant Cell Tissue Organ Cult. 63:199–206.

Wilhelm, B.T. and J.R. Landry. 2009. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. Methods 48:249–257.

Wilhelm, B.T., S. Marguerat, S. Watt et al. 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature 453:1239–1243.

Wu, Y., N. Ding, X. Zhao, M. Zhao, Z. Chang, J. Liu and L. Zhang. 2007. Molecular characterization of PeSOS1: the putative Na + /H+ antiporter of *Populus euphratica*. Plant Mol. Biol. 65:1–11.

Wu, T., Z. Qin, X. Zhou, Z. Feng and Y. Du. 2010. Transcriptome profile analysis of floral sex determination in cucumber. J. Plant Physiol. 167:905–913.

Xiong, L., K.S. Schumaker and J.K. Zhu. 2002. Cell signaling during cold, drought, and salt stress. Plant Cell 14:(Suppl 1):S165–S183.

Yamaguchi-Shinozaki, K. and K. Shinozaki. 2006. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annu. Rev. Plant Biol. 57:781–803.

Ye, C.Y., H.C. Zhang, J.H. Chen, X.L. Xia and W.L. Yin. 2009. Molecular characterization of putative vacuolar NHX-type Na(+)/H(+) exchanger genes from the salt-resistant tree *Populus euphratica*. Physiol. Plant. 137:166–174.

Zeng, F., H. Yan and S.K. Arndt. 2009. Leaf and whole tree adaptations to mild salinity in desert grown *Populus euphratica*. Tree Physiol. 29:1237–1246.

Zhang, F., Y.L. Yang, W.L. He, X. Zhao and L.X. Zhang. 2004*a*. Effects of salinity on growth, compatible solutes of calluses induced from *Populus euphratica*. In Vitro Cell. Dev. Biol. Plant 40:491–494.

Zhang, W., C. Qin, J. Zhao and X. Wang. 2004*b*. Phospholipase D alpha1-derived phosphatidic acid interacts with ABI1 phosphatase 2C and regulates abscisic acid signaling. Proc. Natl Acad. Sci. USA 101:9508–9513.

Zhang, F., Y. Wang, Y.L. Yang, H. Wu, D. Wang and J.Q. Liu. 2007. Involvement of hydrogen peroxide and nitric oxide in salt resistance in the calluses from *Populus euphratica*. Plant, Cell Environ. 30:775–785.

Zhu, J.K. 2002. Salt and drought stress signal transduction in plants. Annu. Rev. Plant Biol. 53:247–273.