Introduction

Abiotic stress, such as drought, salinity, and extreme temperature, is one of the primary causes of crop loss worldwide, reducing the average production of major crop plants. Salinity is a serious worldwide problem, with more than 800 million hectares of land affected throughout the world, which accounts for 6% of world’s total land area (Munns and Tester, 2008). In China, more than 90 million hectares are exposed to salinity or secondary salinity. Xinjiang occupies one-sixth of the area of China, but more than one-third of the land is salinized. Therefore, developing salt-tolerant varieties of crops is an important breeding goal in Xinjiang. Owing to the extreme ecological environment, salt-tolerant organisms are abundant in Xinjiang, with abundant gene resources. An important research project is to mine salt-resistance genes from such plants which can be used in future crop improvement.

Olimarabidopsis pumila (Stephan) Al-Shehbaz, O’Kane & R. A. Price [synonym: Arabidopsis pumila (Stephan) N. Busch] habitates semi-arid and semi-salinized areas in Xinjiang region of China. O. pumila is a close relative of the model plant Arabidopsis thaliana but, unlike A. thaliana, it is a salt-tolerant ephemeral plant that is widely distributed in semi-arid and semi-salinized regions of the Xinjiang region of China, thus providing an ideal candidate plant system for salt tolerance gene mining. A good-quality cDNA library was constructed using cap antibody to enrich full-length cDNA with the gateway technology allowing library construction without traditional methods of cloning by use of restriction enzymes. A preliminary analysis of expressed sequence tags (ESTs) was carried out. The titers of the primary and the normalized cDNA library were $1.6 \times 10^6$ cfu/mL and $6.7 \times 10^6$ cfu/mL, respectively. A total of 1093 clones were randomly selected from the normalized library for EST sequencing. By sequence analysis, 894 high-quality ESTs were generated and assembled into 736 unique sequences consisting of 72 contigs and 664 singletons. The resulting unigenes were categorized according to the gene ontology (GO) hierarchy. The potential roles of gene products associated with stress-related ESTs are discussed. The 736 unigenes were similar to A. thaliana, A. lyrata, or Thellungiella salsuginea. This research provides an overview of the mRNA expression profile and first-hand information of gene sequence expressed in young leaves of O. pumila.

Key words: Olimarabidopsis, Comparative Genomics, Gene Expression
of salt tolerance, salt tolerance in plants is still a very complex problem. Salt stress signal transduction still needs to be explored further.

Expressed sequence tags (ESTs) are short (usually about 300–500 bp), single-pass sequence reads from mRNA (cDNA), which are used in the identification of gene transcripts, in gene discovery, and in gene sequence determination (Adams et al., 1991; Boguski et al., 1993; Alba et al., 2004). ESTs have proven to be an efficient, rapid means to identify novel genes involved in tolerance to environmental stress (Jha et al., 2009; Nishiuchi et al., 2010). Large-scale cDNA sequencing and EST analyses have been successfully used to identify stress tolerance genes in a large number of plants, such as *Suaeda salsa* (Zhang et al., 2001), *Avicennia marina* (Mehta et al., 2005), *Thellungiella halophila* (Du et al., 2008; Taji et al., 2008), *Salicornia brachiata* (Jha et al., 2009), and *Suaeda asparagoides* (Ayarpadikannan et al., 2012).

In this study, we made use of the gateway technology to construct a high-quality normalized and full-length cDNA library from young leaves of *O. pumila* plants exposed to 500 mM NaCl stress, and performed a preliminary functional analysis of ESTs in young shoots of *O. pumila*.

**Material and Methods**

**Plant material and salt stress treatment**

*O. pumila* seeds were collected from the natural semi-salinized land surrounding the Marshroom Lake in Shihezi that is located in the north of Xinjiang province of China. Seeds were surface-sterilized and placed in square Petri dishes with 0.5 × Murashige and Skoog (MS) medium (pH 5.7), with 1% (w/v) sucrose and 0.1% (w/v) agar, at 4 °C for 5 d in the dark to synchronize germination. The plates were incubated in a growth chamber under a 14-h light/10-h dark cycle at 22 °C. After 7 d, the seedlings were transplanted into soil and kept in a growth room with a 14-h photoperiod. When plants were 4 weeks old, the soil was watered with 0.5 × MS nutrient solution supplemented with 500 mM NaCl and the plants were grown for 14 h as described previously (Ni et al., 2007) before being sampled for RNA isolation.

**RNA extraction and construction of the primary cDNA library**

Total RNA was isolated from the leaves using the plant RNA Mini-Prep kit (Qiagen, Hilden, Germany). The mRNA was isolated from total RNA with the FastTrack MAG mRNA isolation kit (Invitrogen, Carlsbad, CA, USA). With the materials, a full-length cDNA library was prepared using the Superscript full-length library construction kit (Invitrogen) according to the manufacturer’s protocol. First-strand cDNA was synthesized using Superscript III reverse transcriptase with the 5′ primer attB2-(d)T22 VN: 5′-biotinyl-GGGGAGAACAATTTGTACAAAGAAGGTGGG (T)22VN-3′, where N = A, C, G, or T; V = A, G, or C. The underlined sequences are attB2 sequences. The 3′ primer is biotinylated to block blunt-end ligation of the 5′ primer adapter to the 5′ end of the first-strand cDNA during the adapter ligation step. Truncated cDNA/RNA was digested by RNase I treatment of incomplete RNA: DNA hybrids. Full-length first-strand cDNA was magnetically captured by Cap-antibody beads and then eluted by wash buffer (Invitrogen). The 5′ prime adapter with the 5′ attB1 primer sequence 5′-ACCACTTTGTACAAAAAGTTGGG-3′ was ligated to the 3′ end of the first-strand cDNA.

The second-strand cDNA was synthesized using the 5′ primer extension method. The 100-μL reaction mixture containing 10 μL 10× high-fidelity PCR buffer, 4 μL dNTP mix (10 mM of each dNTP), 79 μL first-strand cDNA (about 1 μg) with the 5′ primer adaptor, 1 μL 5′ primer (100 ng/μL), and 1 μL Platinum Taq DNA polymerase high fidelity (Invitrogen) was incubated at 68 °C for 20 min, then at 72 °C for 20 min. Double-strand cDNA was size-fractionated by Sephacyrl® S-500 HR column chromatography (Invitrogen) to generate a cDNA library with an average cDNA insert size of approximately 1,000 bp.

To construct a gateway entry cDNA library, the size-fractionated cDNA was ligated into the plasmid vector pDONR222 (Invitrogen) using BP Clonase (Invitrogen). The recombination reaction was allowed to proceed at 25 °C for 16 h, and terminated by mixing with 2 μL proteinase K (Invitrogen) and incubating at 37 °C for 10 min. The resultant mixture was used to transform the ElectroMAX™ DH10B™ T1 phage-resistant cells (Invitrogen). Stocks of electro-transformed...
Construction of the normalized cDNA library

Genomic DNA of *O. pumila* was isolated using the cetyltrimethyl ammonium bromide (CTAB) method (Doyle and Doyle, 1987). Genomic DNA (120 μg) was digested with *Eco*RI (TaKaRa, Dalian, China). After purification and recovery, the DNA fragments were filled in using the Klenow fragment (Invitrogen). The treated genomic DNA was mixed with Dynal magnetic beads (Dynabeads® M-280 Streptavidin) (Invitrogen) to form the affinity system.

The isolation of the plasmids containing the primary cDNA library was performed as previously described by Xiang et al. (1994). The mixture of the plasmids was added to the above affinity system for saturation hybridization (Reqan et al., 2000), and then eluted with wash buffer (Invitrogen). The eluted plasmids were transformed into the ElectroMAX™ DH10B™ T1 phage-resistant cells. After recovery, cells were plated on Luria-Bertani (LB) agar medium containing 50 μg/mL kanamycin. A fraction of the bacterial colonies was used for plasmid preparation and library checking, whereas the remaining bacterial colonies were washed off the agar plates with liquid LB medium containing 50 μg/mL kanamycin and completely resuspended. The suspension was mixed with an equal volume of 80% glycerol, and aliquots were stored as normalized cDNA library stocks at –80 °C.

**Titration of the cDNA library**

Ten μL electro-transformed *E. coli* cells carrying the primary and normalized library, respectively, were diluted 1000-fold, and 50 μL were spread on an LB plate containing 50 μg/mL kanamycin. The plates were incubated at 37 °C overnight to determine the colony forming units (cfu). The number of clones was counted to calculate the library titer according to the formula: 

\[
\text{cfu/mL} = \frac{\text{number of plaques} \times \text{dilution factor} \times 10^6}{\mu L}
\]

The size of the inserted fragments was determined by polymerase chain reaction (PCR) in 40 to 60 randomly selected clones as described by Ni et al. (2007).

**EST sequencing, editing, and assembly**

Clones of the normalized cDNA library were randomly selected from LB agar plates supplemented with 50 μg/mL kanamycin. After manual picking, clones were grown overnight in standard LB/kanamycin medium, and plasmids were isolated by the alkaline lysis method (Birnboim and Doly, 1979). In addition, selected clones were stored at –80 °C as glycerol stocks. Sequencing was carried out from the 5’ end of the cDNA inserts with the M13 forward primer (5’-GTAAAACGACGGCCAG-3’) using an ABI PRISM 3730xl automated DNA sequencer (Applied Biosystems, Grand Island, NY, USA) at the Sequencing Center of the Beijing Genomics Institute, Beijing, China.

The trimming process of all sequences, which included the base calling, the removal of the low-quality sequences, sequences from the vectors, and poly (A) tails, was conducted essentially as described by Lai et al. (2011). After cleaning, sequences shorter than 100 nucleotides were discarded. High-quality ESTs were aligned and assembled into contigs using Codon Code Aligner software (http://www.codoncode.com), when the criterion of a minimum identity of 95% over 40 bp was met. When an EST could not be assembled with others in a contig, it remained as a “singleton”. The contigs and singletons should thus correspond to sequences of unique genes (unigenes).

**Functional annotation of unigenes**

All unique sequences were searched for putative open reading frames (ORF) with the program gorf (http://www.ncbi.nlm.nih.gov/gorf/gorf.html), and the largest ORF sequences were used for functional analysis. Sequence similarity searches were performed using the BLAST program (http://ncbi.nlm.nih.gov/blast/) and the sequences then compared to those in a variety of databases including NCBI nt (non-redundant nucleotide database), NCBI nr (non-redundant protein database) (http://www.ncbi.nlm.nih.gov), and Swiss-Prot (http://www.expasy.org/sprot/) which contains all nucleotide or protein sequences submitted to the public databases, with a cutoff value of *E* < 1e-5. To assign gene ontology (GO) terms, the unique sequences were functionally categorized using the GO annotation tool in TAIR (http://www.arabidopsis.org/tools/bulk/go/).
The distribution of GO terms in the main ontology categories of cellular components, molecular functions, and biological processes was examined.

**Results**

*Total RNA isolation and double-strand cDNA synthesis*

To construct the cDNA library, total RNA was isolated from *O. pumila* leaves stressed by 500 mM NaCl for 14 h. The $A_{260}/A_{280}$ ratio of isolated RNA was found to be 2.0 confirming good quality of the isolated RNA. Agarose gel electrophoresis of the total RNA revealed distinct bands of 28S and 18S rRNA, respectively, the concentration of 28S rRNA being about twice that of 18S rRNA. Double-strand cDNA was analysed on an 1.0% agarose gel. The ds cDNA appeared as a 0.5- to 5-kb smear on the gel, which confirmed the successful synthesis of ds cDNA.

*Construction and analysis of the full-length cDNA library*

A primary cDNA library was constructed with a titer of $1.6 \cdot 10^6$ cfu/mL. The total library capacity was $4.8 \cdot 10^6$ cfu. The quality of the library was assayed by PCR amplification of 96 colonies randomly selected. The recombination rate was more than 95%. The size of the inserts ranged from 500 to 3000 bp, with an average size of about 1000 bp.

The titer of the normalized cDNA library was $6.0 \cdot 10^6$ cfu/mL, and the capacity of the total library was $6.7 \cdot 10^6$ cfu. PCR amplification of randomly picked clones confirmed that the average insert size was again about 1000 bp (Fig. 1). The recombination rate was more than 95%. Thus the normalized and full-length cDNA library, respectively, had high titers, high recombination rates, and large inserts.

*Generation, assembly, and analysis of *O. pumila* ESTs*

From the normalized cDNA library, a total of 1093 cDNA clones were randomly picked and sequenced from the 5' terminal using the primer M13-F. All raw EST sequences were trimmed of vector sequences and the poly (A) tails. The

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total ESTs</td>
<td>1093</td>
</tr>
<tr>
<td>High-quality ESTs</td>
<td>894</td>
</tr>
<tr>
<td>Contigs</td>
<td>72</td>
</tr>
<tr>
<td>ESTs in contigs</td>
<td>230</td>
</tr>
<tr>
<td>Singletons</td>
<td>664</td>
</tr>
<tr>
<td>Unique sequences</td>
<td>736</td>
</tr>
<tr>
<td>Redundancy (%)</td>
<td>17.86</td>
</tr>
<tr>
<td>Average length of unigene sequences (bp)</td>
<td>831.5</td>
</tr>
</tbody>
</table>
low-quality sequences below the minimum length (100 bp) were discarded. This resulted in 894 high-quality ESTs with an average length of 820 bp. These EST resources are suitable for salt tolerance gene discovery and as molecular markers.

The 894 high-quality ESTs were assembled into 736 unigenes, including 72 (9.8%) contigs and 664 (90.2%) singletons (Table I). The average length of the unigene sequences was 831.5 bp, ranging from 216 bp to 1424 bp (Table I). Fig. 2 compares the distribution of the sequence lengths before and after sequence assembly. The distribution of EST frequencies after clustering is shown in Fig. 3. Of the 72 contigs, 46 (63.9%) contained 2 ESTs, 10 (13.9%) contained 3 ESTs, 7 (9.7%) contained 4 ESTs, and 6 (8.3%) contained 5 ESTs. There were few sequences (4.2%) representing more than six ESTs (Fig. 3), and there were only 16 contigs representing more than 4 ESTs (Table II), suggesting that the redundance in the normalized library was relatively low. Each of these clusters contained ≥ 5 ESTs, representing 12% of the total number of ESTs obtained. On average, each contig was assembled from 3.2 sequences due to highly redundant ESTs, and the unigene average size was only 1.2 sequences. These data indicate the good quality of the normalized *O. pumila* cDNA library.

**Functional annotation and categorization of unigenes**

GO analysis has been widely used to classify gene functions (Ashburner et al., 2000). Three structured controlled vocabularies (ontologies) have been defined that describe gene products in terms of their associated biological processes,
Y.-X. Zhao et al. · cDNA Library from Olimarabidopsis pumila

... cellular components, and molecular functions in a species-independent manner (Berardini et al., 2004). All 736 unigenes were annotated, indicating that all of them had significant matches against genes of A. thaliana, A. lyrata, and other plants, and were divided among the three GO categories (Fig. 4). On the whole, 652 genes were categorized under the “cellular component” category, 654 under “molecular function”, and 669 under “biological process”. The total number of GO mapping in each of the three ontologies may exceed the number of unigenes, because a given gene product may be assigned to more than one GO term.

Within the “cellular component” category, chloroplast (13.4%) and plastid (8.3%) genes are most enriched (Fig. 4A). Genes encoding proteins of membranes, including the plasma membrane (4.2%) and other membranes (9.6%), the nucleus (6.3%) and the cytosol (3.6%) are also significantly enriched (Fig. 4A). Transcription factors are crucial for regulating plant responses to biotic and abiotic stress (Singh et al., 2002). In the "molecular function" category, most of the GO terms (40%) were included in "binding" proteins, which often act as transcription factors, such as nucleotide binding (8%), protein binding (6.7%), DNA and RNA binding (5.6%), nucleic acid binding (1.6%), and other binding (17.8%) (Fig. 4B). Proteins having hydrolase activity (11.3%) and transferase activity (10%) are dramatically enriched compared to other terms (Fig. 4B). The unknown molecular function term (7.7%) is rather enriched (Fig. 4B).

The GO “biological process” is helpful for the functional classification of analysed genes. In the "biological process" category (Fig. 4C), a considerable number of ESTs were involved in response to abiotic and biotic stimuli (10%), response to stress (9.5%), and transport (5.5%), which is consistent with previous studies showing that plant adaptation to environmental stress involves the expression of stress-responsive genes and activation of various physiological and metabolic responses (Thomashow, 1999; Shinozaki et al., 2003). These three categories form the basis for mining stress-regulated genes.

**Discussion**

**Construction and composition of a cDNA library of O. pumila**

Construction and analysis of a cDNA library is still an essential technology in modern biology and functional genomics, as it provides much more detailed information on genomic mecha-
Fig. 4. Representation of GO mapping results for *O. pumila* unigenes: (A) GO term "cellular component"; (B) GO term "molecular function"; (C) GO term "biological process".
nisms underlying diverse processes of the organism. The key to the success of a cDNA library construction are the quality of cDNA and the ability to obtain high-quality cDNAs from limited amounts of mRNA. Numerous improvements carried out in the past few years have been made in the construction of full-length cDNA libraries (Seki et al., 1998; Zhu et al., 2001; Kato et al., 2005; Ni et al., 2007). Nevertheless, the construction of high-quality cDNA libraries is still a great challenge, and there is no single method that can resolve all the common problems associated with ligation-assisted conventional cDNA library construction. In the present study, we made use of cap antibody beads to capture and enrich the full-length cDNA, and constructed a normalized and full-length cDNA library by using site-specific recombination, avoiding that genes were cut by restriction enzymes. Using this approach, we constructed an entry cDNA library of young leaves of *O. pumila* plants treated with 500 mM NaCl. The titer of the cDNA library, the recombination rate, and the average insert size in the primary and normalized cDNA library are suitable for the needs of a standard cDNA library (Fig. 1).

**Generation and analysis of ESTs**

High throughout single-pass sequencing, generation and analysis of ESTs have been proven to be a rapid and efficient way of obtaining information on gene expression patterns (Adams et al., 1991; Wu et al., 2002; Phukon et al., 2012). In the present study, a total of 1093 cDNA clones randomly selected from the normalized and full-length cDNA library of *O. pumila* were subjected to sequencing for generation of ESTs. A total of 894 high-quality ESTs were obtained from the cDNA library, putatively representing 736 unigenes, including 72 contigs and 664 singletons.

We found that all ESTs had significant homology with genes from *A. thaliana*, *A. lyrata*, or *Thellungiella salsuginea*. As shown in Fig. 4, the unigenes were classified into cellular components, molecular function, and biological process developed by the Gene Ontology Consortium (Berardini et al., 2004). These genes covered a broad range of the GO functional categories.

A large proportion of genes were found to participate in metabolic processes (30.5%) including protein metabolism, DNA or RNA metabolism, and cell organization and biogenesis (Fig. 4C). Our results showed that the terms of response to abiotic or biotic stimuli, responds to stress, and transport (Fig. 4C) were significantly expressed. At high salinity stress, plant can enhance salinity tolerance by Na\(^+\) exclusion and the control of Na\(^+\) transporters within the plant. Many studies concentrated on the Na\(^+\)/H\(^+\) antiporter proteins in the plasma membrane and tonoplast which play essential roles in Na\(^+\) exclusion and compartmentalization, for example SOS1 (Qiu et al., 2003), NHX1 (Pardo et al., 2006), and HKT1 (Haro et al., 2005). Previous studies showed that the Na\(^+\)/H\(^+\) antiporter (Apse et al., 1999; Gaxiola et al., 2001), H\(^+\)-adenosine triphosphatase (H\(^+\) ATPase), and H\(^+\)-inorganic pyrophosphatase (H\(^+\)-PPase) (Maeshima, 2001) coordinate regulate Na\(^+\) concentration. In a high salt stress environment, plants can survive by exclusion of excess Na\(^+\) from the cytoplasm and sequestration of Na\(^+\) from the cytosol to the vacuole towards the maintenance of ion homeostasis inside the cell. Detailed analysis of genes involved in responses to abiotic and biotic stimuli showed that genes encoding sodium/hydrogen exchanger (NHX1) (JZ151532), vacuole H\(^+\)-ATPase (JZ151828), plasma membrane H\(^+\)-ATPase (JZ151872), late embryogenesis abundant (LEA) protein (JZ151694), NAC domain transcription factor (JZ151841), aquaporin PIP2–3 (JZ152319), and other stress-induced proteins were found in these categories (Fig. 4C).

Of the 72 contigs, there are only nine contigs with more than 5 ESTs (Table II). The gene for which most ESTs (29) were obtained, was the gene encoding the chlorophyll a-b binding protein (CAB1). The genes encoding ribulose bisphosphate carboxylase small subunit 3B, ribulose bisphosphate carboxylase/oxygenase activase, and photosystem II protein psbY-2 (PSBY) were represented by 13, 8, and 5 ESTs, respectively (Table II). The representation of a high number of transcripts encoding the mentioned proteins involved in photosynthesis may indicate that the plants remained healthy even in the presence of NaCl (Jha et al., 2009). Genomic sequencing and comparative genomics indicated that many genes related to cation transport, abscisic acid signaling, and wax production possibly contributed to the success of *Thellungiella salsuginea* in stressful environments (Wu et al., 2012). There are many other genes putatively involved in stress responses of the *O. pumila* genome the functions of which have not yet been analysed.
This is the first study of the transcriptome of O. pumila, a close relative of A. thaliana and A. lyrata, based on a normalized cDNA library derived from young leaves. Through analysis of the ESTs and putative functional gene annotation, a large number of putative stress-regulated genes were identified. Next, we will determine the expression profiles of these genes in response to salt stress. The full length of these genes can be obtained easily, and their specific function in salt tolerance can be investigated by the transgene technology in model organisms, which will eventually provide new gene targets for improving crop resistance to abiotic stress by the genetic engineering technology. In addition, the resulting gateway entry cDNA library obtained can be transferred into various gateway destination vectors for gene expression and functional analysis. The set of ESTs obtained will facilitate the comparable genomic based on unigenes within the different species of the Brassicaceae in the future.

**Accession Numbers**

All ESTs data are available on the internet (GenBank_Accn: JZ151519–JZ152412).

**Acknowledgements**

This work was supported by the National Natural Science Foundation of China (31060149), the Program for Excellent Talents in Shihezi University (RCZX200902), and the Natural Science Foundation of Shihezi University (2011ZRKXTD-06).


