

# Salinity stress adaptation competence in the extremophile *Thellungiella halophila* in comparison with its relative *Arabidopsis thaliana*

Qingqiu Gong<sup>1</sup>, Pinghua Li<sup>1</sup>, Shisong Ma<sup>1,3</sup>, S. Indu Rupassara<sup>1,3</sup> and Hans J. Bohnert<sup>1,2,\*</sup>

Departments of <sup>1</sup>Plant Biology, and <sup>2</sup>Crop Sciences, University of Illinois at Urbana-Champaign, 1201 W. Gregory Drive, Urbana, IL 61801, USA, and

<sup>3</sup>Physiological and Molecular Plant Biology Graduate Program, University of Illinois, 602 S. Goodwin Avenue, Urbana, IL 61801, USA

Received 16 June 2005; revised 29 August 2005; accepted 1 September 2005.

\*For correspondence (fax +1 217 333 5574; e-mail bohnerth@life.uiuc.edu).

---

## Summary

In stark contrast to *Arabidopsis*, a related species, *Thellungiella halophila* (*Thellungiella salsuginea*; salt cress), displays extreme tolerance to high salinity, low humidity and freezing. High nucleotide sequence identity permits the use of tools developed for *Arabidopsis* for *Thellungiella* transcript profiling, for which a microarray platform with >25 000 DNA elements (70-mer oligonucleotides) was used. Microarray transcript profiling and intensity analysis, quantitative RT-PCR, and metabolite profiles define genes and pathways that showed shared and divergent responses to salinity stress in the two species. Shared responses are exemplified by 40% of the regulated genes functioning in confining ribosomal functions, photosynthesis and cell growth, as well as activating osmolyte production, transport activities and abscisic acid-dependent pathways. An additional 60% of regulated genes distinguished *Thellungiella* from *Arabidopsis*. Analysis of the differences showed that *Arabidopsis* exhibited a global defense strategy that required bulk protein synthesis, while *Thellungiella* induced genes functioning in protein folding, post-translational modification and protein redistribution. At 150 mM NaCl, *Thellungiella* maintained unimpeded growth. Transcript intensity analyses and metabolite profiles supported the microarray results, pointing towards a stress-anticipatory preparedness in *Thellungiella*.

**Keywords:** *Arabidopsis thaliana*, *Thellungiella halophila*, salt stress, comparative genomics, microarray, metabolite analysis.

---

## Introduction

The *Arabidopsis thaliana* (Col-0) genome sequence and resources developed to understand the plant's biology also provide powerful tools that can initiate comparative genomic studies within crucifers and beyond (Borevitz and Ecker, 2004; Irish and Benfey, 2004; Paterson *et al.*, 2004; Taji *et al.*, 2004). Interspecies comparisons of genome structure, gene complement, transcript expression and dynamic behaviour of the proteome provide insights into the evolution of species-specific characters and evolutionary niche adaptations. When comparing *Arabidopsis* with related species, mechanisms that defined speciation may be uncovered and their genetic basis traced. Here, *Arabidopsis* reactions to high salinity were matched with those of a relative, *Thellungiella halophila* (salt cress, synonym: *T. salsuginea*; Al-Shehbaz and O'Kane, 1995; Al-Shehbaz

*et al.*, 1999), a species characterized by a markedly different lifestyle compared with *Arabidopsis*. *Thellungiella* shows 'extremophile' characteristics manifested by extreme tolerance to a variety of abiotic stresses, among them low humidity, freezing and high salinity. The species is a halophyte based on its ability to grow and reproduce in seawater-strength solutions of NaCl (Bressan *et al.*, 2001; Inan *et al.*, 2004; Wang *et al.*, 2003). The genetic or biochemical basis that imparts such high tolerance is unknown, but it may be contrasted with reactions detected previously in other species (Chinnusamy *et al.*, 2004; Hasegawa *et al.*, 2000; Seki *et al.*, 2003; Wang *et al.*, 2003; Wasternack and Hause, 2002; Xiong and Zhu, 2002; Zhu, 2001, 2002, 2003).

Evolutionarily, *T. halophila* occupies an intermediary position between the Brassica and *Arabidopsis* clades,

separated from *Arabidopsis* by 15–30 Myr at a time during which crucifers experienced polyploidization events (Blanc *et al.*, 2003; Koch *et al.*, 2000). Conceivably, protective measures to ionic or osmotic imbalances that distinguish species could have evolved at several levels, including degrees of protein activity and structure; the behaviour of the transcription machinery at different control points; or altered responsiveness to external factors or internal, hormonal signals. Apart from adaptive evolutionary changes in coding regions, an expansion of the gene complement leading to paralogy, for example through gene or (partial) genome duplications, could equally have been at the basis of *Thellungiella*'s adaptation to extreme habitats. The appearance and establishment of paralogs may be at the root of how species in many plant families acquired the ability to thrive in habitats from which most other family members became excluded (Hwang *et al.*, 2002; Lespinet *et al.*, 2002; Madlung and Comai, 2004; Tognolli *et al.*, 2002). Considering the vast variability in the capacity to adapt to stressful conditions shown by individual species in many plant families, or even ecotypes within a species, contrasting the two crucifers should provide clues about mechanisms that reveal their distinctive stress-response capacities.

We describe transcript profiles and metabolite levels in *Arabidopsis* and *Thellungiella* grown under controlled conditions in the absence and presence of salt stress at 150 mM NaCl with *Arabidopsis*, which represents a break point for this species, and at 150 and 250 mM NaCl, conditions that represent a minor irritation or a very moderate stress, respectively, for *Thellungiella*. The use of the *Arabidopsis* long-oligonucleotide array platform is justified by a high nucleotide sequence identity between the species in the range of 92–95% for most genes (Inan *et al.*, 2004; Wang *et al.*, 2004; Wong *et al.*, 2005). We define at least three different *Thellungiella* response categories that are not activated or are not represented in *Arabidopsis*. In addition to the shared salt-stress responses, *Thellungiella* activated additional pathways different from the intense activation of disparate pathways in *Arabidopsis*. The response of *Thellungiella* also included an increase in metabolites that fail to increase in *Arabidopsis*. Second, pre-stress intensities of genes in certain pathways were significantly higher in *Thellungiella*, especially in categories that highlight *Thellungiella*-specific, salt stress-inducible pathways. Third, deduced from intensity profiles and metabolites, *Thellungiella* appears to possess novel stress-relevant genes whose homologs in *Arabidopsis* are not responsive to stress.

## Results

### *Microarrays, normalization, statistics and data clustering*

For the microarray study, 4-week-old *Arabidopsis* (At) and 6-week-old *Thellungiella* (Th), similar in size and before



**Figure 1.** Plants used for the experiments. The plants, *Arabidopsis thaliana* Col-0 (left) and *Thellungiella halophila* Shandong (right), grown in artificial soil, were 4 (*Arabidopsis*) and 6 (*Thellungiella*) weeks old at the start of the experiment.

bolting (Figure 1), were treated with 150 mM NaCl (At-150, Th-150), and 250 mM NaCl (Th-250) for 3 and 24 h, respectively. The gene-expression levels were compared with those of untreated controls (see Experimental procedures). Each experimental condition was represented by six slides from two biological repeats, including dye swaps. In addition to controls, the microarray platform included 25 425 probes spotted as 70-mer oligonucleotides. After background subtraction, normalization and filtering (TIGR-TM4-MIDAS), approximately 80% of the signals remained on each slide (Table 1). Only probes with at least five out of six intensity values were considered further, resulting in the acceptance of  $9708 \pm 865$  probes for each of the six experimental conditions. The accepted transcripts were processed by one-way ANOVA ( $P = 0.01$ ; TIGR-TM4-MEV), then filtered for high degrees of freedom (d.f.  $\geq 27$ ), resulting in the acceptance of 2090 transcripts considered as highly significantly regulated across all experimental conditions.

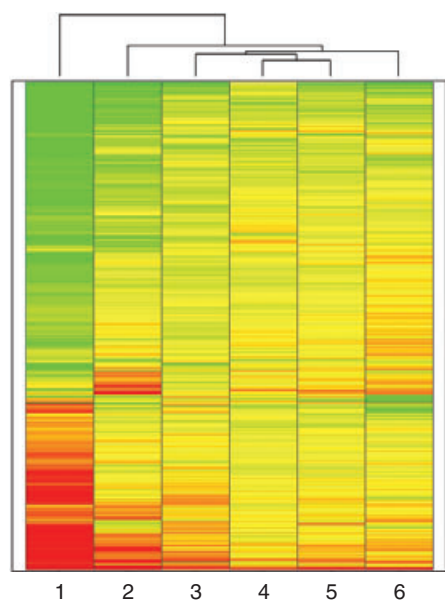
Hierarchical clustering (Figure 2) showed that the *Thellungiella* 3-h experiments grouped closely, followed by the *Arabidopsis* 3-h data set, then *Thellungiella* 24-h (150 mM), then *Thellungiella* 24-h (250 mM), and finally the *Arabidopsis* 24-h (150 mM) data. Furthermore, with respect to the apparently most strongly regulated genes, At-150 mM and Th-250 mM were more similar than Th-150 mM and Th-250 mM. The latter observation was consistent with the one-class *t*-test results generated from each data set (Table 1). Both observations suggested that *Thellungiella* and *Arabidopsis* dealt with salt stress by regulating both similar and different genes and pathways, and that *Thellungiella* might respond in a more determined way.

The fuzzy *k*-means clustering method with principle components analysis (PCA) was introduced as a novel way of clustering the microarray results (Gasch and Eisen, 2002; Figure 3). By this process, centroids were generated according to the assigned parameter, *k*, combined with the PCA result. Hence the method captures and assembles the most typical gene-expression patterns. Each gene is linked to each

**Table 1** Regulated genes identified by one-class *t*-test

Experiment	Total genes (d.f. $\geq 4$ )	Significant genes	Significant genes (d.f. $\geq 4$ )	Upregulated > twofold	Downregulated > twofold	Up (1.5-fold)	Down (1.5-fold)
At-150-24	9365	3930	2500	277	332	553	866
At-150-3	10 571	2637	604	30	4	124	30
Th-150-24	9150	1406	620	33	12	137	58
Th-150-3	9989	1312	549	6	13	55	54
Th-250-24	8479	4200	2212	122	66	336	328
Th-250-3	10 695	1595	634	12	11	47	65

*t*-test parameters:  $P = 0.01$ ; permutation = 64 (all possible combinations from six slides); adjusted Bonferroni correction for alpha. Experimental conditions identify species, NaCl concentration (mM) and time of stress (h).

**Figure 2.** Hierarchical clustering over all experimental conditions.

The 2090 significantly differently regulated genes were clustered using hierarchical clustering. The clustering method was the unweighted pair-group method using arithmetic averages (UPGMA); similarity measure was based on Euclidean distance. Ordering function was based on average value. The visualization dendrogram is based on similarity between profiles. Columns 1–6 (left to right): At-150 mM-24-h; Th-250-24-h; At-150-3-h, Th-150-3-h, Th-250-3-h, Th-150-24-h. The data show the At-3-h profile as a clear precursor of the At-24-h profile, while the Th patterns remain unique, particularly at the 24-h time points. Red indicates upregulation; green, downregulation; yellow, no change.

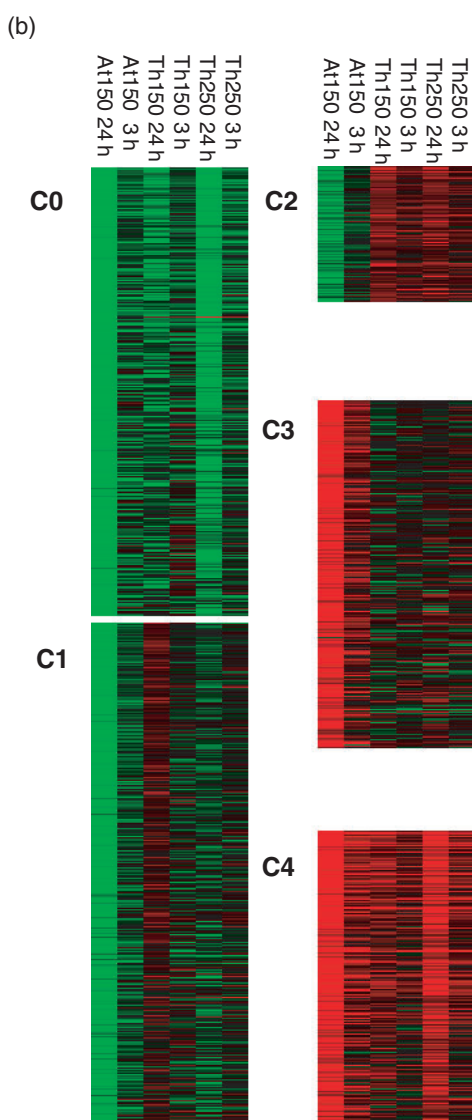
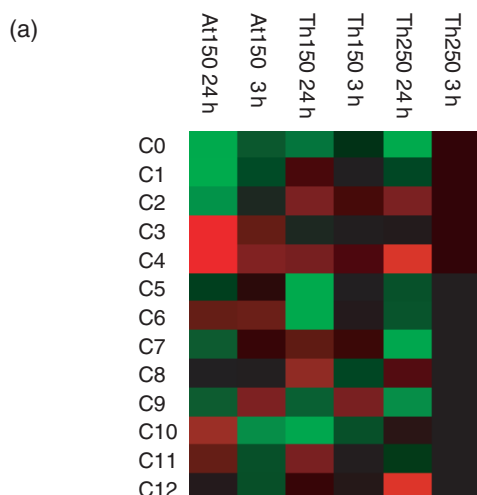
centroid with different membership values, with a total membership value of 1.0 apportioned to all centroids. A higher membership value for a centroid corresponds to an expression pattern more similar to genes associated with this centroid. Finally, clusters based on centroids were generated (see Experimental procedures). Using this method, we obtained 12 clusters that properly described the behaviour of 80% of the 2090 transcripts. These included five large clusters (C0–C4) that included 75% of all data, and seven small clusters (C5–C11) with 5% of the data (Table S4). The remaining 20% of the transcripts could not be placed

with a certainty higher than 0.50 and were therefore not analysed (Figure 4).

An overview of these clusters showed that both species had a balanced number of up- and downregulated genes. Two of the 12 clusters (C0 and C4), containing 40% of all regulated genes, showed consistent upregulation (C4) or downregulation (C0) among the three sets of data (At-150, Th-250, Th-150), and 3 and 24-h time points, respectively. In each species, the 24-h time points represented an amplification of the 3-h response. One small cluster (C9) showed upregulation at the 3-h time point and downregulation at the 24-h time point in all three sets of data. The remaining nine clusters, which included >50% of the data, revealed different time-specific and, in particular, species-specific regulation. Cluster C1 showed moderate downregulation in At-150 at both 3 and 24 h; slight downregulation at 24 h in Th-250; and slight to moderate upregulation in Th-150 at the 24-h time point. C2 showed upregulation in Th-150 and Th-250, and downregulation in At-150. C3 showed upregulation in At-150 but practically no change in Th-150 and Th-250.

#### *Similarity between the two species: downregulated genes (C0)*

The 411 genes in C0 were consistently downregulated in all six experimental conditions, with the decline in At-150 more dramatic than in Th-250 or Th-150. They divided into four groups. The first and largest group contained cytosolic and chloroplast ribosomal subunit genes. Consistent with this, four FKBP genes, including FKBP13 and TWD1/UCU2, appeared as downregulated. Histone genes and genes functioning in DNA synthesis, replication, DNA damage repair, amino acid synthesis (except for proline synthesis), protein synthesis (including two tRNA synthetase genes, elongation factor genes), and cell division were also included. The second group consisted of genes that function in photosynthesis, including genes encoding photosystem I and II subunits, ferredoxins and ferredoxin-NADP reductases, magnesium-chelatase subunits GUN5, and CHL12; protchlorophyllide reductases PORA, PORB and PORC; coproporphyrinogen III oxidase LIN2, oxygen-evolving



enhancer *PsbQs*, RuBisCo activase, Calvin cycle genes, and other chloroplast-located gene products for chloroplast protein import, chloroplast growth and division. As the third group, a considerable number of genes specify their functions in cell-wall degradation and modification, including cell elongation (*EXGT-A1* and *XTH9*), expansions (*EXP6* and *EXP8*), and cell structure (tubulins, *TUA6* and *TUB1*). Similarly downregulated were genes encoding proteins in fatty acid synthesis and elongation genes, including *ACP1*, *ACP4*, *AAE14*, *MOD1*, *LTA2/PLE2*, lipid-transfer protein genes, and *GDSL*-motif lipase/hydrolase genes. The latter group combined auxin-responsive *SAURs* and gibberellin-responsive genes.

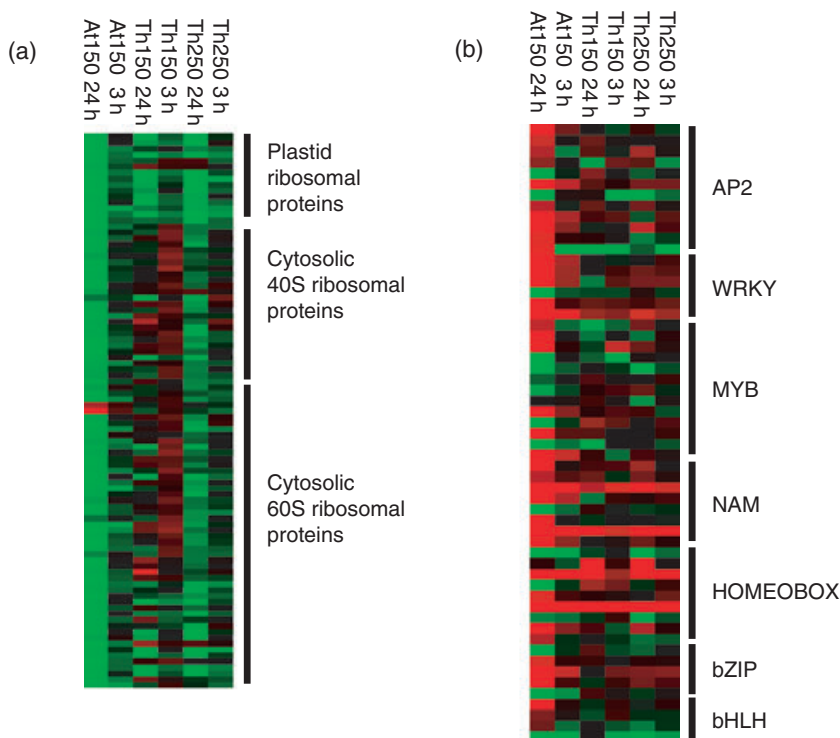
#### Similarity between the two species: upregulated genes (C4)

The 272 genes in C4 were consistently upregulated in all six experimental conditions, with regulation in At-150 slightly more dramatic than in Th-250 and Th-150. Although functionally diverse, many of these genes had been reported in studies on individual genes as strongly abiotic stress-regulated. These included the abscisic acid (ABA)-responsive PP2Cs *ABI1*, *ABI2* and *HAB2*; the ABA-responsive transcription factors *ATHB7*, *ATHB12* and growth regulators *ATAF1*, *NAP*, *BLH1*, *SPL*; proline biosynthesis genes *P5CS2*, *ALDH2*, *P5CDH* and several alcohol dehydrogenases; the CBL-interacting protein kinases *CIPK11/PKS5* and *CIPK21*; and other abiotic stress-responsive genes such as *RAB18*, *RD21A*, *RD29A*, *RD29B*, *KIN2*, *ERD7*, *ERD10*, *ATTPS11* and *MT2A*. Finally, several G-proteins also appeared in C4.

The cluster was further characterized by the presence of genes for transporters, several of the ABC type, and for amino acids, proton-dependent oligopeptides, and sulphate (Table S1), the calcium-transporting ATPase *ACA8*, two auxin polar transporters including *APP1*, and several mitochondria and chloroplast carrier protein genes involving in energy transfer. Identified were genes for several chloroplast-located proteins, previously suggested as stress-responsive (*ATFER1*, *ELIP1*), several leaf senescence markers (*ASP3/YLS4*, *ACD1*, *DIN2*), and genes involved in lipid degradation and  $\beta$ -oxidation pathways were greatly upregulated.

As a smaller group, aspartyl proteases, cysteine proteases, genes involved in the protein ubiquitination pathway, autophagy gene *APG8h*, and vacuolar processing enzyme

**Figure 3.** Fuzzy *k*-means clustering over all experimental conditions. The 2090 significantly differently regulated genes were clustered using Fuzzy *k*-means clustering combined with a principle components analysis. The six experimental conditions are shown. Red indicates upregulation; green, downregulation; black, no change.  
(a) Overview of the 13 original centroids C0–C12.  
(b) The five large clusters, C0–C4. A cut-off value of 0.50 was applied to each membership. Numbers of genes in each cluster: 411 (C0); 464 (C1); 128 (C2); 318 (C3); 272 (C4).



**Figure 4.** Transcript profiling of ribosomal subunit genes and transcription factors. Red indicates upregulation; green, indicates downregulation; black, no change.

(a) Fifteen plastid ribosomal subunit genes; 27 40S ribosomal subunit genes; 51 60S ribosomal subunit genes.

(b) Twelve AP2 domain-containing transcription factors; six WRKY family transcription factors; 13 MYB family transcription factors; eight NAM family proteins; nine homeobox transcription factors; six bZIP transcription factors; three bHLH family proteins.

$\gamma$ -VPE were also included in C4. Not surprising, many redox homeostasis genes were greatly upregulated, including glutaredoxins, glutathione peroxidases (ATGPX2 and ATGPX6), and glutathione *S*-transferases (ATGSTZ1 and ATGSTU7). Many lignin biosynthesis genes, the wax synthesis gene WAX2, and a few cell-wall modification genes formed a group that indicated the process of secondary growth. Finally, a number of NAM transcription factors, zinc-finger family protein genes, and two WRKY transcription factors were included in this cluster.

#### *Differences between the two species: genes upregulated in Arabidopsis only (C3)*

The 318 genes in C3 were greatly upregulated in Arabidopsis, yet were mostly unaffected in *Thellungiella*. In contrast to C4, very few genes in this cluster have been characterized previously as osmotic/salt/ABA-responsive. At first glance, many of these have defense-related functions. These consist of genes for a number of MATE efflux family proteins (presumably involved in toxin resistance), a metallothionein (MT1A), the universal stress protein RD2, senescence-associated SAG13, other short-chain dehydrogenase/reductases, hydrolases, jacalin lectin family proteins, disease-resistance proteins, a terpene synthase, and trypsin inhibitors. The associated WRKYs 17, 25 and 33 might be grouped as defense-related transcription factors.

Interestingly, a group of genes specifying calcium-binding proteins, and calcium-dependent (calcium-bind-

ing) and independent (lipid-mediated) vesicle trafficking functions appeared almost exclusively in C3. They included ECA1, CPK9, BON1, several SEC14 genes, an SEC1 gene, and two SNAREs (VAMP7B and VAMP7C). The appearance of some genes in fatty acid synthesis in C3 could be consistent with the increased intracellular trafficking.

Many genes encoding Myb transcription factors, AP2s, bZIPs, two NAMs, and zinc fingers appeared in C3 as specifically upregulated in Arabidopsis. Many protein kinases and PP2Cs, similar to C4, also appeared in this cluster. Additionally, we found a few MAP kinases, including ATMKK9, ATMPK11, ATMPK5 and WNK4, in this group. The presence of a variety of translation initiation factors, ribosomal subunits RPL10B and RPL10C (both also have an RPL16 domain that binds directly to 23S rRNA) could be consistent with the appearance of the MAPKs.

#### *Differences between the two species: genes upregulated in Thellungiella only (C2)*

The 128 genes in C2 were upregulated steadily at a comparable level in Th-150 and Th-250, and downregulated in Arabidopsis. As expected, as presumably species-specific functions they were relatively difficult to categorize. As many as half the genes are identified as functionally unknown by both The Arabidopsis Information Resource (TAIR) and The Institute for Genomic Research (TIGR). The remainder fell into many divergent and often unique

categories. MSBP1, ATTRX4, ATGSTU20 and other redox homeostasis genes were included. Differently from cluster C3, we observed very few transcription factors in C2. Instead, *Thellungiella* induced small nuclear ribonucleoproteins and a Methyl-CpG binding protein (known to function in gene silencing). A few peptidyl-prolyl *cis-trans* isomerases, including PIN1AT (known to facilitate protein folding), were specific to this cluster. Further C2-specific genes were MMT (a methionine *S*-methyltransferase), a UbiE/COQ5 methyltransferase involved in coenzyme synthesis, a CoA-synthase homolog, two cytosolic aminopeptidases, and several genes specifying protein-targeting functions. Evaluating the collection of functions in C2, it may be suggested that, rather than activating multiple transcription and translation processes, *Thellungiella* diverged from the Arabidopsis model by enhancing functions in protein interaction and modification, and in protein redistribution processes.

#### *Thellungiella prefers sodium (C1)*

C1, the largest cluster with 464 genes, was composed of distinctly differently regulated genes. Arabidopsis repressed the expression of these genes at both time points, most strongly at 24 h. *Thellungiella* showed no regulation at 3 h at both concentrations of NaCl. At 24 h, however, Th-250 showed slight downregulation, and Th-150 showed upregulation. One-third of the genes in C1 were annotated by TAIR and TIGR as functionally unknown/no ontology. The rest could be grouped clearly into different categories. Similarly to C2, C1 showed a large number of genes for protein interactions, modification and translocation, including protein kinases and phosphatases, chaperonins, DNAJ proteins, and protein targeting. Genes for cytosolic ribosomal subunits abounded, suggesting that protein synthesis pathways were induced in Th-150, conceivably to sustain growth, or to restructure the proteome, or to make up for initial losses. The upregulation of FKBP15-2, Sec61 gamma, and cell-cycle genes (CDC2A, ROC2) appeared to support this view. In addition, increased transcription involved other anabolic functions: DNA repair (e.g. RecA); nucleotide synthesis and salvage pathways (e.g. UMPS, PUR5, ADK2, UPT1); fatty acid synthesis (e.g. CAC2, SMT3, LACS2, KAS I, KAS III); and amino acid synthesis (e.g. NMT1, MAM1/IMS3, CARA).

No plastid ribosomal subunits were found in C1, which may suggest that plastid growth was restrained. However, many genes encoding proteins involved in glycolysis, the TCA and Calvin cycles, mitochondrial electron transport/ATP synthesis, and even tetrapyrrole biosynthesis appeared in C1. The non-photochemical quenching gene NPQ4/PSBS was severely downregulated in Arabidopsis, yet stayed constant in both Th-150 and Th-250.

It was noteworthy that, although C1 contained a comparably large number of transcriptional regulators like cluster

C3, these fell into diverse categories including small nuclear ribonucleoproteins, histone deacetylases, chromatin-assembly factors, various RNA- or DNA-binding proteins, or transcriptional coactivators. Only a few have been classified as transcription factors (CUC2, SVP, MS1, RAP2.7, WRKY54). Their presence, however, was consistent with the conclusions drawn for cluster C2.

#### *The seven small clusters*

Clusters C5–C12 included 118 genes altogether. There were from seven to 26 genes associated with these clusters, indicating that their regulation patterns were relatively insignificant. A description of these clusters is given in Appendix S1.

#### *Pre-stress expression differences between Thellungiella and Arabidopsis*

*Thellungiella* transcripts generally hybridized to the printed Arabidopsis probes with intensities that were equal to those seen in homologous hybridizations. After intensity normalization, standard deviation filtering and *t*-test ( $P = 0.05$ ), 2620 genes remained as having statistically significant intensity differences between *Thellungiella* and Arabidopsis. Among these, 1207 had higher intensities in Arabidopsis and 1313 had higher intensities in *Thellungiella*, but in both species very few transcripts showed large differences. When a threefold difference was used as a cut-off, there were 42 transcripts with higher intensities in Arabidopsis, and 51 higher in *Thellungiella* (Table 2).

Among the 51 transcripts with higher intensity in *Thellungiella*, several groups emerged. The first group included four GDSL lipases and a lipid transfer protein, together with FAD3, and a few cell wall-synthesis genes. The second group contained five genes involved in redox control, including ATGPX4, ATGSTU26, APR2 and APR3; and three abiotic stress-inducible genes including RAB18 and TCH2. The third group included eight genes with a function in transcriptional regulation and growth modulation, including MFT, ATK5 and CYP78A5. Finally, we identified four genes functioning in protein-ubiquitination pathways. Transcripts for genes encoding transcription factors were not present. When all 2620 genes were considered in terms of differences in gene families, a few families stood out (Table S3). More highly expressed in *Thellungiella* were genes with functions in ABA synthesis and ABA responsiveness, lipid transfer proteins and GDSL lipases, heat-shock proteins, peptidyl-prolyl *cis-trans* isomerases, histone genes and small nuclear ribonucleoproteins, redox control proteins, and C2H2 zinc fingers. In contrast, genes involved in ethylene response, cyclins, transposons and bZIP transcription factors showed higher intensities in Arabidopsis.

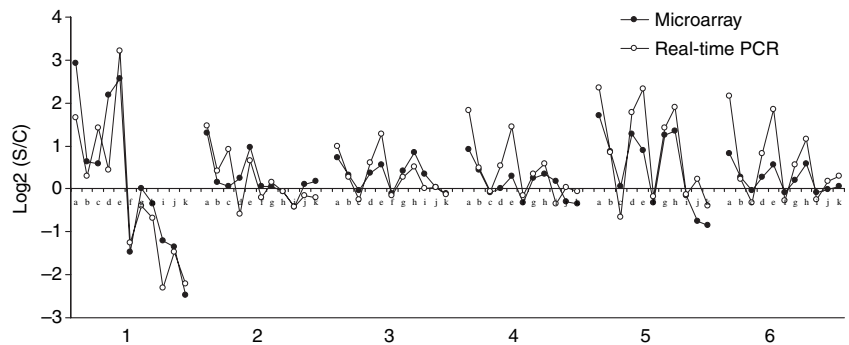
**Table 2** Significant intensity differences between *Thellungiella* and Arabidopsis transcripts

AGI	$I_{At}$	$I_{Th}$	$P$	$I_{Th}/I_{At}$	Annotation
At2g30540	154.23	844.56	7.21E-05	5.48	Glutaredoxin family protein
At5g60060	126.24	642.80	5.17E-05	5.09	F-box family protein, similar to SKP1 interacting partner 2 (SKIP2)
At3g12850	9582.07	47081.29	9.85E-06	4.91	COP9 signalosome complex-related/CSN complex-related
At3g12143	242.98	1087.19	1.87E-05	4.47	Pseudogene, similar to polygalacturonase-inhibitor protein
At1g14020	674.36	2908.86	1.52E-05	4.31	Growth regulator protein similar to <i>Nicotiana tabacum</i> auxin-independent growth promoter axi 1
At1g62180	2176.35	9259.54	3.26E-05	4.25	$\beta$ 5'-adenylylsulphate reductase 2, chloroplast (APR2) (APSR) with glutaredoxin activity
At2g20210	132.56	558.21	0.00016	4.21	Leucine-rich repeat family protein
At5g37770	1376.25	5778.89	1.71E-05	4.20	TCH2, a calmodulin-related protein in response to touch, darkness and temperature shock
At5g42040	156.22	636.22	0.000197	4.07	26S proteasome non-ATPase regulatory subunit, putative
At1g12100	553.23	2165.80	9.86E-05	3.91	Protease inhibitor/seed storage/lipid transfer protein family protein
At5g41010	935.86	3596.24	9.57E-05	3.84	DNA-directed RNA polymerases I, II and III, 7-kDa subunit, putative
At1g33265	374.20	1432.76	5.54E-05	3.83	Expressed protein
At3g14225	162.30	614.08	3.66E-06	3.78	GDSL-motif lipase/hydrolase family protein
At1g27020	1153.25	4362.60	6.77E-05	3.78	Expressed protein
At3g16370	4142.32	15438.33	0.000306	3.73	GDSL-motif lipase/hydrolase family protein
At5g45910	1025.12	3768.27	0.000307	3.68	GDSL-motif lipase/hydrolase family protein
At1g56040	576.94	2119.39	0.000316	3.67	U-box domain-containing protein, protein ubiquitination
At5g26220	167.35	609.71	0.000122	3.64	ChaC-like family protein, similar to <i>E. coli</i> pH-independent $Ca^{2+}/H^{+}$ antiporter
At3g29780	350.23	1270.87	0.000118	3.63	RALFL27, similar to tobacco rapid alkalization factor (RALF)
At4g23500	129.07	459.50	1.27E-05	3.56	Glycoside hydrolase family 28 protein
At2g29980	954.13	3374.16	7.59E-05	3.54	FAD3, ER omega-3 fatty acid desaturase, biosynthesis of 18:3 fatty acids from phospholipids
At5g38530	258.78	892.58	7.2E-05	3.45	Tryptophan synthase-related protein
At1g18100	188.66	650.67	2.98E-05	3.45	Mother of FT and TF1 protein (MFT), a flowering-time modulator
At5g66400	2120.21	7310.74	4.96E-05	3.45	RAB18, a dehydrin
At5g18180	261.49	895.83	0.000131	3.43	Gar1 RNA-binding region family protein
At5g61290	526.74	1770.73	2.34E-05	3.36	Flavin-containing monooxygenase family protein
At5g47960	147.67	495.83	0.000401	3.36	Ras-related GTP-binding family protein
At4g21990	1226.72	4083.71	0.000473	3.33	b 5'-adenylylsulphate reductase (APR3)/PAPS reductase homolog (PRH26)
At1g09180	668.68	2220.06	9.72E-05	3.32	Putative GTP-binding protein, similar to SAR1
At2g48150	157.90	517.22	8.72E-05	3.28	ATGPX4, a glutathione peroxidase
At5g27290	140.99	461.49	0.000202	3.27	Expressed protein with a DNA ligase domain
At3g58100	134.54	439.73	1.11E-05	3.27	Glycosyl hydrolase family 17 protein
At1g24440	341.59	1108.94	1.09E-05	3.25	Expressed protein, weakly similar to MTD2
At1g17190	1529.60	4935.25	0.000133	3.23	ATGSTU26, a glutathione S-transferase
At1g18650	953.11	3058.88	0.000511	3.21	Glycosyl hydrolase family 17 protein
At2g25500	323.42	1030.31	0.000352	3.19	Expressed protein
At1g13710	215.54	684.63	0.000218	3.18	CYP78A5, a $P_{450}$ that marks the shoot apical meristem boundary in Arabidopsis
At2g01890	444.68	1403.31	0.000521	3.16	PAP8, a purple acid phosphatase
At1g18140	838.15	2630.01	0.000289	3.14	Laccase, putative similar to high-pI laccase (LAC2-1)
At2g24580	224.80	703.50	0.000494	3.13	Sarcosine oxidase family protein
At4g37470	1792.23	5547.08	2.45E-05	3.10	Hydrolase, alpha/beta fold family protein
At1g29670	768.70	2376.24	0.000146	3.09	GDSL-motif lipase/hydrolase family protein
At1g07070	879.22	2708.43	0.000295	3.08	60S ribosomal protein L35a (RPL35aA)
At4g05190	226.96	698.00	0.000261	3.08	ATK5, a kinesin protein involved in microtubule spindle morphogenesis
At1g21510	112.14	343.59	0.000272	3.06	Hypothetical protein
At4g30580	588.75	1800.14	0.000349	3.06	Phospholipid/glycerol acyltransferase family protein
At2g22270	244.11	745.77	9.78E-05	3.06	Expressed protein
At5g55400	205.20	624.41	0.000657	3.04	Fimbrin-like protein, putative; similar to fimbrin-like protein ATFIM1
At2g37510	177.00	535.46	0.000326	3.03	RNA-binding protein, putative
At1g17100	1454.11	4368.53	8.22E-05	3.00	SOUL, heme-binding family protein
At4g38410	140.31	420.35	0.000542	3.00	Dehydrin, similar to dehydrin ERD10

$P$ -values calculated using two-tailed Student's  $t$ -test with two-sample unequal variances.  $I_{Th}$  and  $I_{At}$  values calculated by averaging raw intensities, after background subtraction, extracted from the 12 *Thellungiella* and six Arabidopsis 24-h control slides. AGI, Arabidopsis Genome Initiative.

**Figure 5.** Verification of microarray results by real-time quantitative RT-PCR.

$\text{Log}_2(\text{stress/control})$  values for genes (a–k) in the six experimental conditions. Genes: (a) At1g01720; (b) At1g11910; (c) At3g20410; (d) At3g44880; (e) At4g26080; (f) At5g66190; (g) At1g74920; (h) At1g80530; (i) At2g16360; (j) At2g39730; (k) At4g05180. Experimental conditions 1–6 (left to right): At-150 mm-24-h, At-150-3-h, Th-150-24-h, Th-150-3-h, Th-250-24-h, Th-250-3-h. Primers are listed in Table S2.



#### Verification by real-time RT-PCR

Quantitative RT-PCR analyses were conducted for all six experimental conditions using 11 transcripts chosen from different fuzzy *k*-means clusters, apparently up- or down-regulated or unaffected (Figure 5; Table S2). The results, while the absolute fold-regulation varied between microarrays and quantitative amplifications for these transcripts, showed a correlation of 0.83.

#### Metabolite analysis

With a focus on the polar fraction of metabolites (Table 3), drastic differences distinguished the two species. *Thellungiella* maintained higher metabolite levels in both the absence and presence of salt stress compared with *Arabidopsis*. Most of the differences, and higher concentrations in *Thellungiella*, could be attributed to increases under stress in the amount of hexoses, disaccharides and additional complex sugars, the identity of which has not been determined conclusively; and in the amount of sugar alcohols, especially inositol and galactinol. The increase in the category amino acids in both species was nearly exclusively based on an increase in proline, a known stress response in both species (Bressan *et al.*, 2001; Taji *et al.*, 2004), which was more pronounced in *Thellungiella*. Among the entry metabolites for nitrate into the amino acid pool, *Thellungiella* contained much higher amounts for glutamic and aspartic acids, the corresponding amines. In addition, *Arabidopsis* and *Thellungiella* were distinguished by different amounts of organic acids, especially malate and citrate which decreased slightly during the stress. Again, *Thellungiella* showed much higher amounts than *Arabidopsis* under pre-stress conditions, which was also the case for inorganic ions, particularly phosphate.

#### Discussion

Transcript profiles have been reported, using a cDNA-based platform of 7000 *Arabidopsis* full-length cDNAs, in salt treatments that compared *Thellungiella* plants with *Arabid-*

*opsis* plants exposed to a lethal (250 mM) concentration of NaCl (Taji *et al.*, 2004). The results indicated less pronounced stress responses by *Thellungiella* compared with *Arabidopsis*. The authors interpreted their results by further identifying constitutively higher expression of genes in *Thellungiella* as the cause for its superior performance. We used a 70-mer oligonucleotide-based platform of >25 000 *Arabidopsis* genes, with more reasonable and physiologically comparable stress conditions for both species. Rigorous statistical analyses, careful selection of clustering methods, and a comparison with Affymetrix data ensured high confidence in data interpretation. Combined with pre-stress intensity analysis and metabolite profiling, the results revealed genes highlighting a clear distinction between the two species. In addition, we used plants of comparable developmental age, taking into account *Thellungiella*'s slower growth, and performed experiments when complete rosettes had formed and before flowering was initiated.

#### Selection of suitable tools for analysis

To depict the expression pattern for 2090 genes that passed one-way ANOVA with high statistical confidence, we introduced the fuzzy *k*-means clustering method (Gasch and Eisen, 2002) which, to our knowledge, has not been applied before to plant transcriptome studies. In conventional *k*-means clustering, each gene is forced into one and only one cluster, and therefore co-expression and co-regulation patterns will be masked by only marginally relevant genes, resulting in misinterpretations. In fuzzy *k*-means clustering the boundary of each cluster is set by a selected membership cut-off value. By optimizing this cut-off value, irrelevant genes can be removed. Furthermore, PCA guaranteed capture of the most informative centroids; automated convergence of very similar centroids during computation ensures uniqueness of the final clusters. Compared with conventional *k*-means clustering, the requirement of accurately estimating *k* is greatly reduced, leading to increased precision. This approach placed 75% of the regulated genes into five major clusters with distinct expression patterns. In contrast, a test using conventional *k*-means clustering gave

Table 3 Metabolite analysis

Metabolite	Arabidopsis		Thellungiella		
	Control	150 mM	Control	150 mM	250 mM
Aspartic acid	1.05	0.80	ND	2.13	1.49
Citric acid	42.99	45.76	71.31	57.54*	60.48*
Fructose/sorbose	7.31	9.22	51.94	54.70	53.92
Fumaric acid	39.50	34.07	6.03	3.64**	2.33*
Galactinol	2.38	3.65	2.43	5.42*	6.56*
Glucose	28.03	20.43	67.70	95.00*	96.88*
Glutamic acid	15.12	19.45	11.51	30.84*	30.95*
Glycerol	5.23	4.35	2.33	3.52	4.24
Glycine	0.90	4.91	0.94	1.53	2.52
Inositol	9.32	9.50	12.03	17.56**	19.93**
Unknown (putative complex sugar)	1.70	4.40*	ND	4.69*	3.91*
Malic acid	42.87	51.36	155.84	135.67	124.4**
Maltose	1.50	1.14	2.29	1.08	0.97
Mannitol	2.18	1.56	ND	ND	ND
Palmitic acid	2.06	2.00	4.01	1.85	2.00
Phosphoric acid	39.10	50.52	95.79	78.7*	66.64*
Proline	4.10	14.68*	6.55	25.45*	55.35*
Raffinose	ND	2.40	ND	1.36	2.80
Serine	5.04	10.47	3.61	8.87	10.60
Stearic acid	3.28	3.43	6.39	3.36	3.58
Succinic acid	4.92	3.36	9.31	10.42	9.09
Sucrose	96.52	118.72**	150.26	115.7*	102.92*
Threonine	8.38	9.83	5.43	11.78	14.47
Trehalose	2.62	2.58	1.11	2.42**	2.45**
Trehalose-like	1.84	1.80	12.64	35.86*	31.39*

Mean values are expressed as percentage of total signal (internal standard, ribitol, was taken as 100%). *Thellungiella* contains many of these metabolites at higher amounts under control conditions, especially fructose, sucrose, complex sugars, malate and proline. The nature of all complex sugars has not been identified unequivocally. Samples were taken from control plants and plants treated at 150 and 250 mM NaCl, all harvested at the 24-h time point. ND, not detected; difference control versus stress: \* $P = 0.01$ ; \*\* $P = 0.05$ .

results that were difficult to analyse, due to the ambiguous placement of a subset of the genes and similarities between different clusters (results not shown).

#### *Arabidopsis salt-stress responses*

The importance of maintaining ion homeostasis during salt stress has long been recognized. The Salt-Overly-Sensitive (SOS) pathway that connects a calcium signal through a calcium-binding protein (SOS3) via a kinase transmitter (SOS2) to a sodium exclusion (SOS1) function (Zhu, 2002, 2003) has been established as a major ionic ( $\text{Na}^+$ ) stress-regulatory system. Also shown has been a salt-protection function provided by transgenic overexpression of  $\text{Na}^+/\text{H}^+$  (or alkali cation) antiporters of the NHX/NHE family which confine sodium to vacuoles, or by maintaining potassium homeostasis (Apse *et al.*, 1999; Blumwald, 2003; Sottosanto *et al.*, 2004).

However, transcripts for these functions are not readily observed in microarray studies (Kreps *et al.*, 2002; Seki *et al.*, 2002). These genes could have low transcript intensities as one reason, or their induction might be post-

transcriptional. Possibly most relevant is that many of these genes have been characterized in *Arabidopsis* mutant screens at low  $\text{Na}^+$  concentrations, while they exert an effect that is insufficient to save *Arabidopsis* at moderate or high  $\text{Na}^+$ . Accordingly, studies done with *Arabidopsis* at higher  $\text{Na}^+$  concentrations point to the importance of osmotic adjustments, the pivotal role of ABA, and genes and pathways shared with other abiotic, and even biotic, stress responses (Fowler and Thomashow, 2002; Kreps *et al.*, 2002; Seki *et al.*, 2002, 2003). A salt stress that challenges growth is not merely ionic in nature; it requires elicitation of a complex response network (Zhu, 2001).

#### *Comparisons with Affymetrix gene-chip results*

In an additional evaluation, our results were compared with publicly available Affymetrix gene-chip data from *Arabidopsis* 3-h and 24-h 150 mM NaCl stress experiments (AtGen-Express, TAIR gene chip database: <http://www.arabidopsis.org/info/expression/ATGenExpress.jsp>). The two data sets, although based on different platforms including different numbers of genes, and processed through different

statistical methods, shared common trends of regulation. For the 3-h time point, 81% of upregulated genes and 82% of downregulated genes from our study (*t*-test,  $P = 0.05$ ) were regulated in the same direction as suggested by the Affymetrix data set. For the 24-h time point the numbers were 79% for upregulation and 88% for downregulation (S.M. and Q.G., unpublished data). This comparison provided additional confidence with respect to the microarray platform used, as well as suggesting that the Arabidopsis salt stress-response mechanisms are now relatively well defined at the level of transcripts.

#### *Similarities between the Arabidopsis and Thellungiella salt-stress transcriptomes*

The similarities are clearly depicted in the fuzzy *k*-means clusters C0 (downregulation, 411 genes) and C4 (upregulation, 272 genes). Cluster C0 proved simplest with respect to its composition: ribosomal subunit genes and genes in photosynthesis functions represented 12 and 20%, respectively, of the downregulated, functionally known genes in C0, suggesting repression of protein synthesis and plastid activities. Additionally, cell-wall synthesis and modification were equally reduced, together with other growth and expansion processes. Interestingly, in other large-scale analyses of the Arabidopsis abiotic stress transcriptome, transcripts for ribosomal proteins rarely changed (Fowler and Thomashow, 2002; Kreps *et al.*, 2002; Seki *et al.*, 2002, 2003). However, an analysis of Affymetrix gene-chip data (AtGenExpress, TAIR gene chip database: <http://www.arabidopsis.org/info/expression/ATGenExpress.jsp>) revealed 70% of the ribosomal subunit genes as significantly downregulated only in osmotic and salt-stress experiments (S.M. and Q.G., unpublished data). This suggests that the osmotic or salt-stress treatments typically used are of such severity that Arabidopsis ceases to grow, and that the reactions measured are largely defense responses.

Cluster C4 included a large number of abiotic stress-responsive (especially ABA-responsive) genes, many identifying functions in osmolyte production. Also observed were glutaredoxins, glutathione peroxidases and glutathione *S*-transferases, suggesting a necessity for enhanced redox homeostasis. Further upregulated were transcription factors and growth regulators, suggesting remodelling of growth; a variety of transporters, probably involved in nutrient redistribution, signalling molecules and energy; and many proteases and protein ubiquitination genes, the latter suggesting protein-recycling processes. Consistent with cluster C0, leaf senescence markers were upregulated, showing decreased chloroplast activity. The upregulation of secondary cell-wall and wax biosynthesis suggest attempts at preventing water loss.

Similarities in regulation between Arabidopsis and *Thellungiella* showed that both salt-sensitive and salt-tolerant

plants, encountering increased salinity, transcriptionally rapidly curb growth-related functions, including (principally) chloroplast activities. Abscisic acid was clearly the crucial hormone used by both species. At 3 h these adjustments were already well under way in both species. Presence and upregulation of protein kinases (including CDPKs) and phosphatases (especially PP2Cs) suggested responses involving protein modification.

#### *Differences between salinity-sensitive Arabidopsis and tolerant Thellungiella*

Clusters C1, C2 and C3 (910 genes) outline differences between Arabidopsis and *Thellungiella*. C3, clustering genes upregulated in Arabidopsis but downregulated in *Thellungiella*, identified mainly general defense-related genes and transcription factors, together with protein translation-initiation functions. Even at 150 mM NaCl, Arabidopsis initiates what may be termed a global defense response that apparently requires the presence of newly synthesized proteins. In contrast, cluster C2, upregulated in *Thellungiella* at both concentrations of NaCl (and especially at 250 mM) and downregulated in Arabidopsis, was enriched in genes with a function in protein folding, post-translational modification and protein distribution. Here the choices of pathways diverged dramatically between both species. It is tempting to suggest that the '*Thellungiella* way' conserved resources and energy, hence making it a more superior strategy under stress conditions.

The composition of cluster C1 is also illuminating. Apart from genes suggesting modifications at the protein level, as in C2, we observed stress recovery already at the 3-h time point. These processes included enhanced nucleic acid salvage, DNA repair, carbon assimilation, and amino acid and lipid synthesis. In addition, transcripts in glycolysis, TCA cycle, Calvin cycle and ATP synthesis had reverted to the pre-stress status at the 24-h time point. Although some chloroplast ribosomal protein genes remained downregulated, transcripts for cytosolic ribosomal subunit proteins were upregulated. It appeared that *Thellungiella* was able to evaluate stress intensity quickly, and took distinct, different strategies facing 150- and 250-mM NaCl stress.

A recent study, reporting 6578 unigene expressed sequence tags (ESTs) from cold-, drought- and salt-stressed cDNA libraries of *Thellungiella* (ecotype Yukon) (Wong *et al.*, 2005) concluded that salt-induced responses diverged significantly between *Thellungiella* and Arabidopsis. Our results agree with this conclusion. In this study, 569 out of the 2090 genes analysed here had at least one matching EST, evenly distributed over the 12 clusters that show overall sequence identity of 92–95% with their Arabidopsis homologs, further supporting the suggestion that the Arabidopsis-based microarray platform is suitable for *Thellungiella* transcript profiling.

### Metabolites and stress tolerance

Fundamental differences, reflecting what the transcript profiles appear to predict, are prominently reflected when comparing metabolite profiles (Table 3). *Thellungiella* exhibits higher pre-stress concentrations of several compounds that have been shown to have protective functions in the osmotic imbalance that is a component of salt stress (Hasegawa *et al.*, 2000). The concentrations for all major metabolite groups were often several-fold higher in *Thellungiella* and, in some groups, stress (even short-term stress) caused even more disparity. The high amounts of hexoses (approximately fivefold), sugar alcohols (twofold) and organic acids (two- to threefold, malate, citrate, but not fumarate) in *Thellungiella* were unexpected. Some of these increased during the 24-h stress period. Proline, the amino acid that increased most dramatically in both species, as previously noted (Bressan *et al.*, 2001; Taji *et al.*, 2004); inositols; and increased amounts of hexoses and complex sugars are substantiated by increased expression levels of transcripts involved in their metabolism.

### Transcription factors with species specific patterns

The majority of transcription factor transcripts identified at significantly regulated levels showed co-ordinated behaviour. As a group, the plant-specific transcriptional regulators NAM (no apical meristem), including ATAF1 and CUC1, were all upregulated significantly in at least one condition or species. WRKY transcription factors, known to have regulatory functions in pathogen defense and other stress responses (Eulgem *et al.*, 2000; Ulker and Somssich, 2004), were also upregulated, especially in Arabidopsis. However, Myb transcription factors showed diverse responses, and none appeared in C4 as upregulated in both species. AP2 transcription factors mainly appeared in C3 as upregulated only in Arabidopsis. bZIP transcription factors, having higher intensities in Arabidopsis under control condition, were further induced mainly in Arabidopsis.

The largest group of responders, however, included transcripts for 40 zinc-finger proteins. Although originally identified as nucleic acid-binding proteins, zinc-finger domains are versatile in structure and involved in diverse functions, particularly in protein-protein interactions (Taka-tsuiji, 1999). In our experiments, the regulated zinc-finger genes appeared as one of the largest functional categories, found in every cluster. For instance, C2H2-type zinc fingers have been suggested as stress regulators (Sakamoto *et al.*, 2004). They showed generally higher transcript intensities in *Thellungiella* under control condition, and many were induced especially in Arabidopsis during salt stress. The precise functions of all observed zinc-finger proteins remain to be elucidated.

### Differences in transcript intensity

Recent studies have probed the possibilities of acquiring both expression ratios and absolute or relative gene-expression intensities from long oligonucleotide arrays (t'Hoen *et al.*, 2004). We analysed all 24-h control intensities. Even before normalization, we observed comparable hybridization intensities, and the Pearson correlation coefficient between *Thellungiella* and Arabidopsis slides, based on the averaged intensity for each gene, was 0.90. After stringent statistical analysis, 2620 genes showed significant intensity differences between the two species. Consistent with the observation by Taji *et al.* (2004), with high confidence, we observed ABA biosynthesis and ABA-responsive genes with higher intensities in *Thellungiella* (Table S3). We further identified genes with a function in redox control, protein folding, histone genes, and cell-wall synthesis genes as more highly expressed in *Thellungiella*.

In order to see whether differences in expression intensity were important for salt-stress tolerance, the expression-intensity ratio between *Thellungiella* and Arabidopsis was calculated for each gene in the computed fuzzy *k*-means clusters (data not shown). The only significant difference identified cluster C2. Surprisingly, the averaged ratio in C2 was 1.5, demonstrating that *Thellungiella* not only had higher intensity of genes in this cluster with diametrically opposite regulatory behaviour, but elevated their expression when stressed. Therefore we consider these genes, 129 in total, and the way *Thellungiella* enhances their expression, as prime representatives in salt-stress acclimation and major contributors to the tolerance lifestyle of this species. One-third of these genes are labelled as functionally unknown in the TAIR database; and only 18 out of 129 genes have been studied in Arabidopsis. None of these 18 genes has been indicated before as salt stress-relevant, and it seems possible that they represent stress-inducible paralogs of genes that are not regulated in the same way in Arabidopsis. The high-intensity signal on the Arabidopsis platform and upregulation of these genes in *Thellungiella* could indicate their involvement in *Thellungiella*'s superior stress performance. Also, the high signal intensity may point to the existence of novel functions carried out by paralogs of Arabidopsis genes in *Thellungiella*. Only the sequencing of the *Thellungiella* genome will resolve this question.

With a platform that nearly completely encompassed the genome, our study identified salt-stress response as a defined network in both Arabidopsis and *Thellungiella*, with common responses. Many of the Arabidopsis-specific reactions to high salinity appear to be injury-based, representing extreme defense reactions. Advanced bioinformatics tools then led to the discovery of response pathways that distinguished *Thellungiella* from Arabidopsis. *Thellungiella* appears superior in managing metabolite composition and

maintenance of energy, while *Arabidopsis* expended energy at low stress levels on a multiplicity of pathways. The response precision shown by *Thellungiella*, we suggest, identifies genes that will very probably be novel and important for our understanding of plant salinity stress tolerance.

## Experimental procedures

### Plant growth and stress treatments

*Arabidopsis thaliana* (Col-0) and *Thellungiella halophila* (Shandong) were used (Inan *et al.*, 2004). Surface-sterilized seeds suspended in Murashige and Skoog (MS) medium (Gibco, Carlsbad, CA, USA) containing 2% sucrose were germinated at 24°C with gentle shaking (16/8 h light/dark; approximately 150  $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ ). Two-week-old seedlings were transferred to pots filled with isolate artificial soil (Sundine Enterprises, Arvada, CO, USA) for hydroponic culture. Plants were supplied with 0.5 $\times$  Hoagland's nutrient solution with 4 $\times$  increased Fe.

Four-week-old *Arabidopsis* and 6-week-old *Thellungiella*, similar in size and before bolting, were irrigated with 150 mM NaCl (At-150, Th-150), and one set of *Thellungiella* plants with 250 mM NaCl (Th-250). Plants remained in the presence of the chosen NaCl solution, or nutrient solution for controls. After treatment for 3 and 24 h, respectively, plants, at least 10 per sample, were frozen in liquid N<sub>2</sub>. Two biological repeats, grown separately at different times, were used. We chose these time points because a previous manuscript (Taji *et al.*, 2004) had shown that, within the initial 12 h salt stress, both species showed an increase in NaCl uptake; after 24 h *Arabidopsis* showed a marginally higher concentration of sodium.

### Microarray hybridizations

Total RNA was isolated (RNeasy, Qiagen, Carlsbad, CA, USA). Glass microarray slides consisting of 70-mer oligonucleotide probes (<http://ag.arizona.edu/microarray>) were used in hybridizations. RNA samples (70  $\mu\text{g}$  each) for control and treatment conditions were reverse-transcribed (SuperScript III, Invitrogen, Carlsbad, CA, USA), and hybridization was carried out according to <http://atarays.tigr.org/arabprotocols.shtml> (for detailed protocol see Appendix S2). To avoid bias in microarrays as a consequence of dye-related differences in labelling efficiency, dye labelling for each paired sample (stress/control) was swapped in one of three independent hybridizations. In total, 36 microarray hybridizations were carried out.

### Microarray evaluation and statistical analysis

Signal intensities for each array element were collected (GenePix 4000B, Axon Instruments, Union City, CA, USA) and images analysed (GENEPIX PRO 4.0). Spots with intensities lower than background or with an aberrant spot shape were flagged by the GENEPIX software and checked manually. The resulting GPR files were analysed by TIGR-TM4 (<http://www.tm4.org>; Saeed *et al.*, 2003) and converted by EXPRESSCONVERTER ver. 1.5 to generate co-ordinated MEV and ANN files. MEV files were processed through TIGR-MIDAS ver. 2.18. Total intensity normalization, Lowess (Locfit) normalization, standard deviation regulation, and intensity filtering were done within each slide. After normalization, columns IA and IB in dye-swap slides were swapped so that each column IA represented the control data and each column IB the stress data.

Statistical analyses were carried out using MEV ver. 3.0.3, FUZZYK (<http://rana.lbl.gov/FuzzyK/software.html>) and SPOTFIRE (Spotfire, Somerville, MA, USA). In MEV, a one-class *t*-test for each experimental condition, and a one-way ANOVA for all six experimental conditions, were carried out separately to reveal patterns of regulation. In both cases permutation with a very high stringency (ANOVA:  $P = 0.01$ , permutation = 1000; *t*-test:  $P = 0.01$ , permutation = 64) was used to find significantly differently regulated genes. Adjusted Bonferroni *P*-value correction was applied at the same time in order to reduce false discovery rate. The one-way ANOVA output, log<sub>2</sub>-ratio of stressed to control, was then analysed with SPOTFIRE for hierarchical clustering and with FUZZYK (Gasch and Eisen, 2002) for fuzzy *k*-means clustering, including a PCA. Throughout we use the names of *Arabidopsis* homologs to identify *Thellungiella* genes.

Fuzzy *k*-means analysis was done with the FUZZYK program (<http://rana.lbl.gov/FuzzyK/software.html>) with the parameter  $k = 20$ , and 13 centroids were generated. Genes were grouped into a cluster if they had a 0.50 or higher membership value with a certain centroid. All together, 12 clusters were obtained corresponding to 12 centroids, and the last centroid (C12) was removed because no gene had a membership value higher than 0.50. The results were visualized via the software MAPLE TREE (<http://maple.tree.sourceforge.net>).

Transcript intensity analysis was carried out as follows: the raw intensities after background subtraction for each gene were extracted from the control column of all 24-h slides, 18 in total. After removal of all control spots, total intensity normalization was done with all 18 slides. Then the mean intensity value and standard deviation were calculated separately for each gene in the 12 *Thellungiella* and six *Arabidopsis* slides. Genes with a ratio of standard deviation to mean >0.50 were then filtered out. Finally, a *t*-test ( $P = 0.05$ ) was performed with the remaining genes to find significant differences in transcript intensities.

In order to understand better the salt-stress response at the transcript level, the fuzzy *k*-means clustering results and transcript intensities under control conditions were combined and analysed. For each gene appearing in the fuzzy *k*-means clustering results, the ratio  $r = I_{Th}/I_{At}$  was calculated, with  $I_{Th}$  and  $I_{At}$  representing averaged transcript intensity calculated from all 24-h *Thellungiella* control slides and 24-h *Arabidopsis* control slides. The individual ratios within each cluster were averaged, resulting in 12 averaged ratios.

### Quantitative RT-PCR

For verification of the hybridization results, real-time PCR was conducted. RNA (2  $\mu\text{g}$ ) from each experimental condition was used for first-strand cDNA synthesis exactly as described. Primer sequences are listed in Table S2. *Thellungiella* actin (CX129618) cDNA primers and *Arabidopsis* Ubiquitin-10 cDNA primers were used as internal controls. Detection of RT-PCR products used the fluorescent dye SYBR-green (Applied Biosystems, Foster City, CA, USA) and the ABI PRISM/Taqman 7700 Sequence Detection System (Applied Biosystems). Three repeats per gene were done, and averaged threshold-cycle numbers were used to calculate changes (log<sub>2</sub>).

### Metabolite profiling

Aliquots of the material used for microarrays in liquid N<sub>2</sub> were used and metabolites were extracted (Fiehn *et al.*, 2000a,b). Tissue (50 mg) was extracted and separated into polar (methanol/water) and apolar/lipid (chloroform) phases. Internal standards (acetic

acid, proline, ribitol, trehalose) were added to a separate set of samples before extractions. Vacuum-dried polar fractions were derivatized (Fiehn *et al.*, 2000b). Samples (0.5 µl; split ratio 40:1) were injected into a GC-HP5890 (Spectron-HP, Kirkland, WA, USA) connected to a Micromass 70-VSE magnetic sector mass spectrometer (injection temperature 250°C; interface 315°C; ion source kept at 200°C). Mass spectral *m/z* range was 50–800. Masses were compared with data of the National Institute of Standards and Technology (NIST) and its replicates collection (NISTREP), our own libraries, and reference compounds for peak identification.

### Acknowledgements

The authors thank Dr John Quackenbush and the TM4 group, TIGR, for help with statistical analyses. Also, we thank Drs Valeriy Poroyko, Alexander Ulanov and Dong-ha Oh, University of Illinois at Urbana-Champaign (UIUC), for discussions and advice on manuscript preparation. Space constraints made it impossible to list the publications characterizing the 341 functionally characterized salt stress-regulated genes included in Table S1. The work has been supported by NSF DBI-0223905/DBI-0500617, the UIUC Plant Biology and Physiological and Molecular Plant Biology graduate programs, and UIUC institutional funds.

### Supplementary Material

The following supplementary material is available for this article online:

**Table S1** Functionally known genes in fuzzy *k*-means clusters C0–C12 and all genes in fuzzy *k*-means clusters

**Table S2** List of primer sequences used in RT-PCR analysis and the RT-PCR data

**Table S3** List of all transcript intensities with statistically significant differences between *Arabidopsis* and *Thellungiella*

**Table S4** Complete list of fuzzy *k*-means clusters

**Appendix S1.** The seven small clusters

**Appendix S2.** Description of the microarray experiments to document Minimum Information About a Microarray Experiment compliance

This material is available as part of the online article from <http://www.blackwell-synergy.com>

### References

- Al-Shehbaz, I.A. and O’Kane, S.L. (1995) Placement of *Arabidopsis parvula* in *Thellungiella* (Brassicaceae). *Novon*, **5**, 309–310.
- Al-Shehbaz, I.A., O’Kane, S.L. and Price, R.A. (1999) Generic placement of species excluded from *Arabidopsis* (Brassicaceae). *Novon*, **9**, 296–307.
- Apse, M.P., Aharon, G.S., Snedden, W.A. and Blumwald, E. (1999) Salt tolerance conferred by overexpression of a vacuolar Na<sup>+</sup>/H<sup>+</sup> antiporter in *Arabidopsis*. *Science*, **285**, 1256–1258.
- Blanc, G., Hokamp, K. and Wolfe, K.H. (2003) A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Res.* **13**, 137–144.
- Blumwald, E. (2003) Engineering salt tolerance in plants. *Biotechnol. Eng. Rev.* **20**, 261–275.
- Borevitz, J.O. and Ecker, J.R. (2004) Plant genomics: the third wave. *Annu. Rev. Genomics Hum. Genet.* **5**, 443–477.
- Bressan, R.A., Zhang, C., Zhang, H., Hasegawa, P.M., Bohnert, H.J. and Zhu, J.K. (2001) Learning from the *Arabidopsis* experience. The next gene search paradigm. *Plant Physiol.* **127**, 1354–1360.
- Chinnusamy, V., Schumaker, K. and Zhu, J.K. (2004) Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants. *J. Exp. Bot.* **55**, 225–236.
- Eulgem, T., Rushton, P.J., Robatzek, S. and Somssich, I.E. (2000) The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* **5**, 199–206.
- Fiehn, O., Kopka, J., Trethewey, R.N. and Willmitzer, L. (2000a) Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal. Chem.* **72**, 3573–3580.
- Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N. and Willmitzer, L. (2000b) Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **18**, 1157–1161.
- Fowler, S. and Thomashow, M.F. (2002) *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. *Plant Cell*, **14**, 1675–1690.
- Gasch, A.P. and Eisen, M.B. (2002) Exploring the conditional coregulation of yeast gene expression through fuzzy *k*-means clustering. *Genome Biol.* **3**, 1–22.
- Hasegawa, P.M., Bressan, R.A., Zhu, J.K. and Bohnert, H.J. (2000) Plant cellular and molecular responses to high salinity. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **51**, 463–499.
- t’Hoen, P.A., Turk, R., Boer, J.M., Sterrenburg, E., de Menezes, R.X., van Ommen, G.J. and den Dunnen, J.T. (2004) Intensity-based analysis of two-colour microarrays enables efficient and flexible hybridization designs. *Nucl. Acids Res.* **32**, e41.
- Hwang, I., Chen, H.C. and Sheen, J. (2002) Two-component signal transduction pathways in *Arabidopsis*. *Plant Physiol.* **129**, 500–515.
- Inan, G., Zhang, Q., Li, P. *et al.* (2004) Salt cress. A halophyte and cryophyte *Arabidopsis* relative model system and its applicability to molecular genetic analyses of growth and development of extremophiles. *Plant Physiol.* **135**, 1718–1737.
- Irish, V.F. and Benfey, P.N. (2004) Beyond *Arabidopsis*. Translational biology meets evolutionary developmental biology. *Plant Physiol.* **135**, 611–614.
- Koch, M.A., Haubold, B. and Mitchell-Olds, T. (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol. Biol. Evol.* **17**, 1483–1498.
- Kreps, J.A., Wu, Y., Chang, H.S., Zhu, T., Wang, X. and Harper, J.F. (2002) Transcriptome changes for *Arabidopsis* in response to salt, osmotic, and cold stress. *Plant Physiol.* **130**, 2129–2141.
- Lespinet, O., Wolf, Y.I., Koonin, E.V. and Aravind, L. (2002) The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Res.* **12**, 1048–1059.
- Madlung, A. and Comai, L. (2004) The effect of stress on genome regulation and structure. *Ann Bot. Lond.*, **94**, 481–495.
- Paterson, A.H., Bowers, J.E., Chapman, B.A., Peterson, D.G., Rong, J. and Wicker, T.M. (2004) Comparative genome analysis of monocots and dicots, toward characterization of angiosperm diversity. *Curr. Opin. Biotechnol.* **15**, 120–125.
- Saeed, A.I., Sharov, V., White, J. *et al.* (2003) TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*, **34**, 374–378.
- Sakamoto, H., Maruyama, K., Sakuma, Y., Meshi, T., Iwabuchi, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2004) *Arabidopsis* Cys2/His2-type zinc-finger proteins function as transcription repressors under drought, cold, and high-salinity stress conditions. *Plant Physiol.* **136**, 2734–2746.
- Seki, M., Narusaka, M., Ishida, J. *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.
- Seki, M., Kamei, A., Yamaguchi-Shinozaki, K. and Shinozaki, K.** (2003) Molecular responses to drought, salinity and frost: common and different paths for plant protection. *Curr. Opin. Biotechnol.* **14**, 194–199.
- Sottosanto, J.B., Gelli, A. and Blumwald, E.** (2004) DNA array analyses of *Arabidopsis thaliana* lacking a vacuolar Na<sup>+</sup>/H<sup>+</sup> antiporter: impact of AtNHX1 on gene expression. *Plant J.* **40**, 752–771.
- Taji, T., Seki, M., Satou, M., Sakurai, T., Kobayashi, M., Ishiyama, K., Narusaka, Y., Narusaka, M., Zhu, J.K. and Shinozaki, K.** (2004) Comparative genomics in salt tolerance between *Arabidopsis* and *Arabidopsis*-related halophyte salt cress using *Arabidopsis* microarray. *Plant Physiol.* **135**, 1697–1709.
- Takatsuji, H.** (1999) Zinc-finger proteins: the classical zinc finger emerges in contemporary plant science. *Plant Mol. Biol.* **39**, 1073–1078.
- Tognolli, M., Penel, C., Greppin, H. and Simon, P.** (2002) Analysis and expression of the class III peroxidase large gene family in *Arabidopsis thaliana*. *Gene*, **288**, 129–138.
- Ulker, B. and Somssich, I.E.** (2004) WRKY transcription factors: from DNA binding towards biological function. *Curr. Opin. Plant Biol.* **7**, 491–498.
- Wang, W., Vinocur, B. and Altman, A.** (2003) Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta*, **218**, 1–14.
- Wang, Z.I., Li, P.H., Fredricksen, M. et al.** (2004) Expressed sequence tags from *Thellungiella halophila*, a new model to study plant salt-tolerance. *Plant Sci.* **166**, 609–616.
- Wasternack, C. and Hause, B.** (2002) Jasmonates and octadecanoids: signals in plant stress responses and development. *Prog. Nucl. Acid Res. Mol. Biol.* **72**, 165–221.
- Wong, C.E., Li, Y., Whitty, B.R., Diaz-Camino, C., Akhter, S.R., Brandle, J.E., Golding, G.B., Weretilnyk, E.A., Moffatt, B.A. and Griffith, M.** (2005) Expressed sequence tags from the Yukon ecotype of *Thellungiella* reveal that gene expression in response to cold, drought and salinity shows little overlap. *Plant Mol. Biol.* **58**, 561–574.
- Xiong, L. and Zhu, J.K.** (2002) Molecular and genetic aspects of plant responses to osmotic stress. *Plant Cell Environ.* **25**, 131–139.
- Zhu, J.K.** (2001) Plant salt tolerance. *Trends Plant Sci.* **6**, 66–71.
- Zhu, J.K.** (2002) Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* **53**, 247–273.
- Zhu, J.K.** (2003) Regulation of ion homeostasis under salt stress. *Curr. Opin. Plant Biol.* **6**, 441–445.