Heredity (2007) 99, 143–150

www.nature.com/hdy

ORIGINAL ARTICLE

Plastic and adaptive gene expression patterns associated with temperature stress in *Arabidopsis thaliana*

WR Swindell[1], M Huebner[1] and AP Weber[2,3]
[1]*Department of Statistics and Probability, Michigan State University, East Lansing, MI, USA and* [2]*Department of Plant Biology, Michigan State University, East Lansing, MI, USA*

Transcriptional profiling using DNA microarrays has become a widely used approach for identifying genes with important roles in stress-regulatory networks. In previous studies, genes exhibiting a plastic expression pattern with respect to stress and control treatments have been identified as candidates with putative roles in stress-response pathways. This approach, however, often identifies numerous genes, and it is difficult to discern which genes have major effects that impact the fitness of individuals under stress. In this study, we investigated the impacts of temperature stress (cold and heat) on gene expression in the *Arabidopsis thaliana* model system. We identified genes exhibiting plastic patterns of gene expression with respect to temperature stress, but in contrast to previous studies, we also considered the adaptive significance of genes by examining their expression patterns among 10 *Arabidopsis* ecotypes indigenous to a range of latitudes. Our findings support a general association between plasticity of gene expression and adaptive value. In comparison to non-plastic genes, genes exhibiting plastic expression patterns were associated with greater among-ecotype variation in expression levels, and such variation was more strongly correlated with geographical temperature gradients. Surprisingly, while more than 16 000 genes were associated with plastic expression patterns, significant evidence of both expression plasticity and adaptive value was obtained for only 43 genes. These selected genes represent strong candidates for future experimental investigations into the molecular basis of temperature acclimation in the *A. thaliana* model system.
Heredity (2007) **99,** 143–150; doi:10.1038/sj.hdy.6800975; published online 2 May 2007

Keywords: adaptation; gene expression; microarray; plasticity; reaction norm

Introduction

Environmental stress plays a key role in determining the evolutionary history of populations and the geographical distribution of plants and animals (Hoffmann and Parsons, 1991). The impacts of stress are especially critical among plant species, since plants are sessile and unable to evade stressful environments through migration (Huey *et al*., 2002). Plants have therefore evolved complex tolerance mechanisms that allow for continued survival and reproduction under harsh conditions. In model systems, such as *Arabidopsis thaliana*, the biochemical basis of these tolerance mechanisms has been examined with the long-term goal of understanding the regulatory pathways that underlie stress perception, signal-transduction pathways and the molecular mechanisms that confer increased tolerance and survival within stressful environments (Denby and Gehring, 2005; Bohnert *et al*., 2006). Understanding these processes requires the identification and analysis of major genes that underlie stress-regulatory networks. DNA microarrays allow genes exhibiting transcriptional induction under stress to be identified, and have therefore provided considerable insight into processes underlying stress tolerance at the gene expression level (Bohnert *et al*., 2006). While the knowledge generated by DNA microarrays does have limitations (Feder and Walser, 2005), the technology, experimental methods and statistical analysis tools are still undergoing development (Canales *et al*., 2006; Ji and Davis, 2006), such that considerable potential remains for understanding stress-response pathways through the application of microarrays.

Microarrays have been used to identify genes exhibiting large transcriptional induction in response to a wide range of stress conditions in plant species, including temperature extremes, drought, high-intensity light, wounding and pathogen infection (Kreps *et al*., 2002; Hazen *et al*., 2003; Takahashi *et al*., 2004; Seki *et al*., 2004; Liu *et al*., 2005). These previous studies have assigned importance to genes that exhibit large transcriptional induction under a stress treatment in comparison to a benign control treatment. This criterion suggests that genes exhibiting a *plastic* pattern of gene expression play the greatest role in stress-response pathways. The reasoning that underlies this approach is well founded.

Correspondence: Dr WR Swindell, Department of Statistics and Probability, Michigan State University, A-413 Wells Hall, East Lansing, MI 48824, USA.
E-mail: swindel5@msu.edu
[3]*Current address: Institute for Plant Biochemistry, Heinrich-Heine-University, 40225 Düsseldorf, Germany.*
Received 15 August 2006; revised 12 December 2006; accepted 16 February 2007; published online 2 May 2007

Differential expression of a gene between stress and control treatments indicates that transcription is responsive to the stress condition, which implies that to some extent, the gene occupies a role in the relevant stress-regulatory network. A common difficulty with this approach, however, is that an extremely large number of genes may be identified as differentially expressed between a stress and control condition. The study of Kreps *et al*. (2002), for instance, examined the effects of cold stress in *Arabidopsis* using a cDNA microarray representing 8100 *Arabidopsis* genes, and found that 2086 transcripts (25%) exhibited greater than twofold induction under cold stress in comparison to a benign temperature treatment. This large number of gene candidates limits the practical significance of identified genes, and presents a challenge in choosing which genes to investigate further in experimental studies. In addition, it has been argued that the genes identified by this methodology are of little relevance, since plasticity alone provides little indication that a gene's expression is consequential for fitness under stress (Feder and Walser, 2005).

The application of microarrays to populations that have undergone selection for increased stress resistance provides an alternative method of identifying genes involved in stress-response pathways. Since gene expression levels have a heritable component (Gibson and Weir, 2005), genes the expression of which influences the survival of individuals under stress should exhibit differential expression between stress-selected and non-selected populations. Microarray studies of populations selected for stress resistance may therefore add a second dimension to the data obtained from plasticity of gene expression responses under stress. The combination of these two approaches, moreover, may generate shorter lists of candidate genes with more highly supported roles in stress-response pathways. For rapidly reproducing species, artificial selection experiments performed in the laboratory may be used to identify transcripts responsive to selection for increased stress resistance (e.g., Riehle *et al*., 2003; Fong *et al*., 2005). For species with lengthy generation times, artificial selection experiments are less feasible, but considerable insight can still be obtained from species that exhibit extensive natural variation (Oleksiak *et al*., 2002, 2005; Whitehead and Crawford, 2006). Many types of stress such as temperature exhibit strong geographic gradients, such that the influence of selection on expression patterns can be detected by examining the covariance between gene expression and environmental factors associated with natural populations (e.g., Lempe *et al*., 2005; Whitehead and Crawford, 2006). This approach allows genes the expression of which has likely been influenced by natural selection to be distinguished from those whose expression levels have diverged due to random genetic drift (Khaitovich *et al*., 2004).

In this study, we investigated both plastic and adaptive gene expression patterns associated with temperature stress in the *A. thaliana* model system. We used large-scale Affymetrix microarray data sets that have been made publicly available by the AtGenExpress consortium, which include expression measurements from approximately 80% of all known *Arabidopsis* genes (Schmid *et al*., 2005). These data include transcript levels of *Arabidopsis* genes under three different temperature treatments (4°C (cold stress), 24°C (benign) and 38°C (heat stress)), in addition to gene expression measurements from 10 different *Arabidopsis* ecotypes. The available data therefore allowed the identification of genes exhibiting plastic expression responses under temperature stress, as well as putatively adaptive genes with expression patterns among ecotypes that correlated with geographical temperature gradients. Our analysis had two primary objectives: since previous *Arabidopsis* studies have identified genes underlying temperature-stress response based upon plasticity of expression patterns alone (e.g., Kreps *et al*., 2002), it was of interest to evaluate whether, in general, genes exhibiting significant plasticity under temperature stress were more likely than non-plastic genes to be of adaptive value (as determined from gene expression patterns among ecotypes). Our second objective was to combine data related to plasticity of gene expression with inferences regarding adaptive significance to generate a short list of gene candidates with highly supported roles in temperature-stress-regulatory pathways.

Materials and methods

The preprocessed and normalized gene expression data examined in this study were downloaded from AtGenExpress at http://www.weigelworld.org/resources/microarray/AtGenExpress/. The abiotic stress series data consists of gene expression measurements performed on *A. thaliana* (col-0) shoot tissue samples with duplicate biological replications. From this data set, we considered gene expression measurements from the cold (4°C), high temperature (38°C) and control environmental treatments (24°C). For each treatment, gene expression measurements were obtained from 16- to 18-day-old plants at six different time points (0.5, 1, 3, 6, 12 and 24 h). In the cold-stress treatment, plants were subjected to low temperature for the entire 24-h period, while in the heat-stress treatment, plants were exposed to high temperature for only the first 3 h. An overview of the experimental procedures followed in temperature and control treatments is available at http://www.uni-tuebingen.de/plantphys/AFGN/atgenextable2.htm, while more detailed descriptions can be obtained from TAIR (http://www.arabidopsis.org/) (submission numbers: ME00325 and ME00340). The ecotype data set consists of gene expression measurements from 10 *Arabidopsis* accessions (TAIR submission number ME00374). These ecotypes are listed in Table 1 along with the average yearly temperatures associated with their geographic points of origin. A world map displaying the geographical distribution of ecotypes is provided in Supplementary Data File 1. For each ecotype, plants were grown under continuous light and RNA was isolated from aerial parts of 4-day-old seedlings. Triplicate gene expression measurements originated from plants that were grown at the same time (under non-stressful conditions) and processed individually with independent RNA preparation and probe synthesis (Markus Schmid, personal communication). All gene expression measurements in both data sets were generated using the ATH1 Affymetrix microarray platform (Hennig *et al*., 2003; Redman *et al*., 2004), and expression estimates were generated using GeneChip RMA normalization (Wu *et al*., 2004). A total of 22 810 probes were included on the ATH1

Table 1 *Arabidopsis* ecotypes and geographic locations from which accessions originated

Ecotype	*Location*	*Latitude*	*Longitude*	*Temperature (range)*
Bay-0	Bayreuth, Germany	49 N	11 E	7.6 (−1.9–17.1)
C24	Coimbra, Portugal	40 N	8 W	14.8 (9.3–20.9)
Col-0	Columbia, United States	39 N	93 W	12.1 (−3.1–25.6)
Cvi	Cape Verdian Islands	16 N	24 W	24.5 (22.2–26.9)[a]
Est	Estonia, Russia	59 N	26 E	4.5 (−6.5–16.8)[b]
Kin-0	Kinneville, United States	43 N	85 W	7.2 (−6.8–20.7)[c]
Ler	Landsberg, Poland	53 N	16 E	7.2 (−1.6–16.6)[d]
Nd-1	Niederzenz, Germany	50 N	8 E	9.8 (0.2–19.0)[e]
Shakdara	Pamiro–Alay, Tadjikistan	37 N	71 E	13.1 (−2.3–27.2)[f]
Van-0	Vancouver, British Columbia	50 N	123 W	8.0 (3.3–12.9)

All temperature data were obtained from the WorldClimate online database (www.worldclimate.com). The temperatures listed for each accession represent yearly averages over a span of at least 30 years. Ranges reflect the mean temperatures associated with the coldest and warmest months respectively.
[a]Praia, Cape Verde (14 N, 23 W).
[b]Helsinki/Seutula (60 N, 25 E).
[c]Bad Axe, Huron County (44 N, 83 W).
[d]Koszalin, Poland (54 N, 16 E).
[e]Frankfurt, Germany (50 N, 8 E).
[f]Fergana, Tadjikistan (40 N, 71 E).

platform, including 64 control probes not corresponding to *Arabidopsis* genes. Our analysis was therefore based on a total of $N = 22\,746$ genes, representing approximately 80% of all known *Arabidopsis* genes (Schmid *et al*., 2005).

Differential expression analysis

Genes exhibiting differential expression (plasticity) under cold or heat stress were identified using the limma linear modeling package available in the R Bioconductor software suite (Smyth, 2004). In this approach, a linear model was fit for each gene (separately for cold and heat stress), which allowed individual genes to be tested for differential expression at each of the six time points of measurement. Each time point of measurement corresponded to one of six effects included in the linear model associated with each stress, where each effect yielded a coefficient estimate, which was converted into moderated *T*-statistics used to evaluate differential expression at each of the six time points (Smyth, 2004). The *P*-values generated by this analysis were adjusted for multiple comparisons using the Benjamini–Hochberg method (Benjamini and Hochberg, 1995). In contrast to conservative *P*-value adjustments aimed at controlling the familywise error rate (e.g., Holm, 1979), the Benjamini–Hochberg method controls the false discovery rate, which leads to greater power when testing thousands of hypotheses such as in microarray experiments (Allison *et al*., 2006). Individual genes were tested for differential expression at a total of 12 stress–time combinations. At each stress–time combination, the analysis yielded a set of n differentially expressed genes, whose expression levels exhibited significant plasticity with respect to either cold or heat stress.

Expression patterns among ecotypes

Patterns of gene expression among ecotypes were characterized by two different measures (F and β). Both measures were computed after gene expression values among ecotypes were centered, such that the mean expression intensity associated with each individual gene was equal to zero. For each individual gene, a total of 30 gene expression measurements were available among the 10 ecotypes (10 ecotypes, 3 replicates per ecotype). These expression measurements were treated as the response variable in a one-way analysis of variance in which ecotype was treated as a class variable with 10 levels. The F ratio generated from the ANOVA reflects the variance among ecotypes relative to the average variance of triplicate gene expression measurements associated with individual ecotypes. For the ith gene, therefore, the ratio F_i provides a summary measure for the overall level of differentiation among ecotypes with respect to gene expression (Nuzhdin *et al*., 2004). The value of F reflects the divergence among ecotypes resulting from the combined effects of many different environmental factors (not temperature alone). A large F ratio suggests a gene's expression has been subjected to divergent selection among ecotypes, while small values of F are evidence for stabilizing selection among ecotypes (Falconer and Mackay, 1996; Whitehead and Crawford, 2006). These inferences are suggestive, but not conclusive evidence of selection, since the expectation of F under random genetic drift was not known (Khaitovich *et al*., 2004).

The second measure (β) reflected the covariance between ecotype expression levels and the average yearly temperature associated with the geographic locations from which ecotypes originated (see Table 1). This measure was calculated by least-squares regression analysis in which gene expression values for a given gene were treated as the response variable, and average yearly temperatures listed in Table 1 were treated as the predictor variable. The estimate of β obtained for the ith gene was equal to the least-squares slope obtained from this regression analysis. The significance of individual β estimates was assessed by a two-tailed *T*-test, with Benjamini–Hochberg adjustments for multiple testing. Significant estimates of β indicate that variation in gene expression among ecotypes strongly covaries with geographical temperature gradients. Since such a relationship between gene expression levels and geographical temperature gradients is not expected under a random genetic drift model (Khaitovich *et al*., 2004),

significant estimates of β provide evidence to suggest that a gene's expression pattern has been influenced by divergent selection related to temperature.

Association between plastic and adaptive expression patterns

Differential expression analyses identified several sets of n genes exhibiting plastic expression responses to either cold or heat stress at particular time points of exposure. To examine the relationship between plasticity of expression and adaptive significance, it was of interest to determine whether each set of n plastic genes exhibited expression patterns among ecotypes that were distinguishable from all other *Arabidopsis* genes. For each set of n plastic genes, therefore, we compared patterns of gene expression among ecotypes to all other genes represented on the ATH1 array. In particular, we determined whether expression levels of the n genes collectively exhibited greater among-ecotype differentiation (F) or covariance with geographical temperature (β) than all $N = 22\,746$ *Arabidopsis* genes examined.

Two summary statistics were developed to characterize the values of F and β associated with each set of n genes differentially expressed at each stress–time combination. These statistics are designated as $T(F)$ and $T(\beta)$, respectively, and represent the median values of F_i and $|\beta_i|$ associated with each set of genes $i = 1 \ldots n$.

$$T(F) = \underset{i=1\ldots n}{\text{median}}\,(F_i) \quad (1)$$

$$T(\beta) = \underset{i=1\ldots n}{\text{median}}\,(|\beta_i|) \quad (2)$$

The significance of each statistic was evaluated under the null hypothesis that the set of n genes is a random sample of the N genes represented on the ATH1 array, versus the alternative that the set is a non-random sample yielding a T-statistic larger than expected within a random sample. To evaluate this hypothesis, 10 000 random samples of n genes were chosen at random from among all N genes, and the appropriate T-statistic was calculated from each of the 10 000 random samples. The resampling procedure was carried out for each T-statistic with respect to each of the 12 stress–time combinations. This yielded null distributions of each T-statistic corresponding to each stress–time combination, which were used to evaluate the significance of observed T-statistics calculated from each set of n differentially expressed genes. An observed T-statistic was significant with respect to a given stress–time combination, when the proportion of random samples yielding a larger or equal T-statistic was less than $\alpha = 0.05$.

The evidence obtained for plasticity of expression patterns was combined with inferences of adaptive value in order to identify candidate genes with highly supported roles in temperature stress-response pathways. For both cold and heat stress separately, sets of n genes identified at each of the six time points of stress exposure were pooled. This yielded two large sets of genes, one exhibiting plastic expression with respect to at least one time point under cold stress, and one exhibiting plastic expression with respect to at least one time point under heat stress. These two sets were then intersected with the set of genes associated with significant estimates of β (indicative of adaptive value) to identify those genes associated with both plastic and adaptive expression patterns. The web-based GOstat tool was then used to determine whether any gene ontology terms were overrepresented among the candidate genes we identified by this method (Beissbarth and Speed, 2004).

Results

A total of 12 843 genes exhibited differential expression with respect to at least 1 time point under the cold-stress treatment, while 13 367 genes exhibited differential expression with respect to at least 1 time point under the heat-stress treatment. Among all genes represented on the ATH1 array, 16 637 exhibited differential expression under either heat or cold stress with respect to at least 1 time point, which represents 70% of the 22 746 genes examined in this study. The great majority of genes we considered, therefore, exhibited some plasticity under either cold- or heat-stress treatments. In the cold-stress treatment, 1697 genes were differentially expressed after 0.5 h of stress exposure, with increasing numbers of genes differentially expressed over time, such that more than 9000 genes were differentially expressed at the 24 h time point of stress exposure. In contrast, for the heat-stress treatment, the number of differentially expressed genes peaked at the 3 h time point (9113 genes), with fewer genes differentially expressed during the early and later stages (812–6667 genes).

Table 2 lists the statistics $T(F)$ and $T(\beta)$ associated with the n genes differentially expressed at each time point of

Table 2 *T*-statistics associated with sets of n genes exhibiting significant plasticity with respect to each of six time points under cold and heat stress

Treatment	N	T(F)	T(β) × 10^3
Cold			
0.5 h	1697	0.883[a]	10.19[a]
1 h	2508	0.893[a]	8.70[a]
3 h	3057	0.902[a]	10.02[a]
6 h	5298	0.938[a]	9.71[a]
12 h	8094	0.935[a]	8.98[a]
24 h	9407	0.886[a]	8.73[a]
0.5–24 h	116	0.838[b]	12.44[a]
Heat			
0.5 h	1696	0.827[a]	10.52[a]
1 h	6667	0.775[a]	7.81[a]
3 h	9113	0.825[a]	7.86[a]
6 h	3760	0.889[a]	7.34[a]
12 h	869	0.894[a]	10.30[a]
24 h	812	0.859[a]	7.84[a]
0.5–24 h	11	2.071[c]	9.42[d]

The row labeled 0.5–24 h lists statistics associated with the n genes exhibiting significant plasticity with respect to all six time points in either heat or cold stress. The value of $T(F)$ indicates the median level of differentiation among ecotypes with respect to gene expression within a given set of n genes. The value of $T(\beta)$ represents the median level of covariance between gene expression and geographical temperature with respect to a given set of n genes (see Equation 2 in Materials and methods).
The footnotes associated with each statistic indicate P-values, which represent the probability of observing a larger statistic based upon 10 000 random samples of size n from the $N = 22\,746$ *Arabidopsis* genes represented on the ATH1 array (see Materials and methods).
[a]P-value is <0.001.
[b]P-value is 0.086.
[c]P-value is 0.028.
[d]P-value is 0.116.

measurement. In both stress treatments, and with respect to every time point, the statistics $T(F)$ and $T(\beta)$ were associated with significant P-values. This indicated that the median values of F and β associated with each set of n genes were larger than that expected from a sample of n genes taken at random from all genes represented on the ATH1 array. Among all 22 746 genes on the ATH1 array, the median values of F and $|\beta| \times 10^3$ were equal to 0.61 and 4.55, respectively. However, the median values of these parameters were considerably larger among each set of n differentially expressed genes ($0.77 < T(F) < 2.07$; $7.34 < T(\beta) \times 10^3 < 12.44$). Genes whose expression levels exhibited significant plasticity under temperature stress, therefore, were generally associated with greater among-ecotype variance in gene expression (F) as well as stronger covariance between expression levels and geographic temperature gradients (β).

A total of 666 individual genes were associated with significant β estimates ($P < 0.05$). This small number of genes, exhibiting an adaptive expression pattern, is considerably less than the more than 16 000 genes exhibiting plasticity under cold or heat stress. The relationship among adaptive and plastic gene sets is illustrated by the Venn Diagrams shown in Figure 1. At a significance level of $\alpha = 0.05$, only 602 ($101 + 423 + 78$) genes exhibiting plastic expression were also associated with significant β estimates (see Figure 1a). Consideration of β estimates, therefore, reduced the number of candidate genes to 3.6% of those identified on the basis of plasticity of gene expression patterns alone. When a significance level of $\alpha = 0.001$ was considered, the overlap between significantly plastic genes and genes with significant β estimates was reduced to 128 genes, representing just 0.77% of genes exhibiting differential expression.

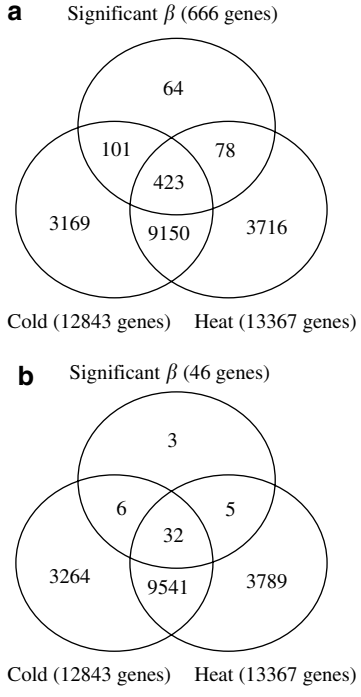

Figure 1 Gene sets exhibiting plastic and/or adaptive expression patterns. Genes included in the cold and heat sets, respectively, exhibited significant differential expression (plasticity) with respect to at least one of six time points at which measurements were obtained. The significant β set corresponds to genes exhibiting expression patterns among ecotypes that correlated significantly with geographical temperature gradients. (**a**) Set membership for all significant genes. (**b**) The significant β set includes only genes for which β estimates were significant with or without the inclusion of the Cape Verdi Island ecotype.

The Cape Verdi Island ecotype was associated with a temperature considerably larger than other ecotypes (see Table 1), and thus had substantial leverage in regression analyses. To account for this, we identified genes whose β estimates were significant regardless of whether the cvi ecotype was included in the anlaysis (see Figure 1b). Applying this criterion yielded 46 genes that show significant evidence of an adaptive gene expression pattern among ecotypes. Of these 46 genes, only 3 were not differentially expressed under either heat or cold stress (see Figure 1b) (At5g10140 (FLC transcription factor), At3g59380 (FTA), At4g26320 (AGP13)). The remaining 43 genes exhibited both plastic and adaptive gene expression patterns, with and without inclusion of the cvi ecotype, and were therefore the most well-supported candidate genes identified in this analysis. These genes are listed in Supplementary Data File 1 along with annotations, functional overviews, hierarchical clustering analyses and absolute expression intensities under control and temperature-stress treatments.

We examined the sensitivity of our set of 43 identified genes to slight methodological modifications. One alternative approach, for example, is to use minimum or maximum geographic temperatures associated with ecotypes (see Table 1) as the predictor variable for calculating β estimates (rather than average temperature). For all 43 genes we identified, however, β estimates remained significant when either minimum or maximum temperature was used as a predictor variable. Our identified gene set was therefore stable to alternative predictor variables associated with average temperature. Another consideration is that the three North American ecotypes included in our analysis (Col-0, Kin-0 and Van-0) were more recently introduced to their associated geographic location (Shimizu, 2002), and have therefore had less opportunity to adapt to the thermal climate. We thus examined the sensitivity of our identified set of 43 genes to the exclusion of these ecotypes. Of the 43 identified genes, 32 remain identified if North American ecotypes are excluded from the analysis (see Supplementary Data File 1). This indicates that, for most identified genes, adaptive expression patterns are robust to exclusion of North American ecotypes.

Figure 2 displays the plastic and adaptive expression patterns associated with MYB29 (At5g07690), which was one of the 3 transcription factors included among the 43 genes we identified. In Figure 2a, the $\log_2$ fold-change exhibited by MYB29 across each time point of stress exposure under cold and heat stress is shown, while expression levels of ecotypes versus geographic temperature are plotted in Figure 2b. Plots similar to Figure 2a and b are shown for the remaining 42 genes within Supplementary Data File 2. Among all 43 genes we identified, no gene ontology terms were significantly overrepresented. However, among the five genes associated with significant β estimates and differential expression under heat (see Figure 1b), the gene ontology term corresponding to aminoacylase activity (GO:0004046) was significantly overrepresented ($P = 0.011$).

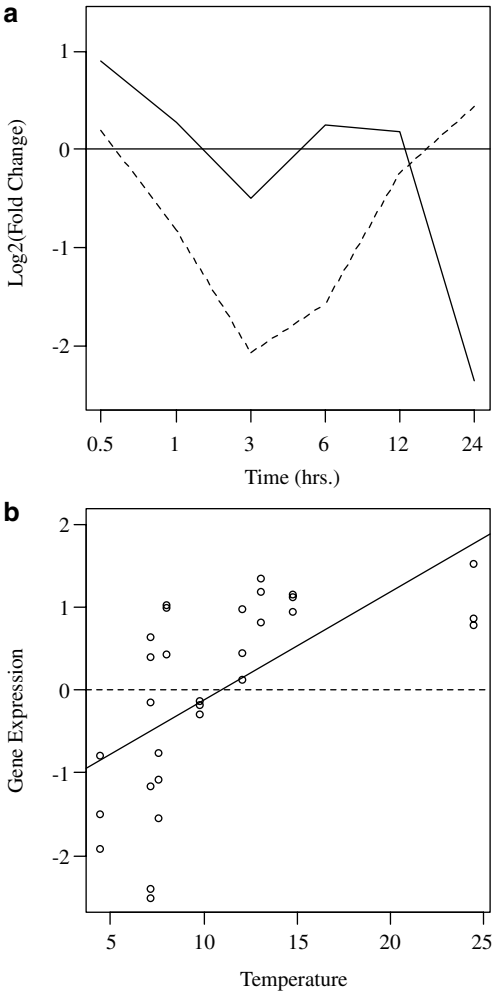

Figure 2 Plastic and adaptive expression patterns associated with the MYB29 transcription factor (At5g07690). MYB29 exhibited differential expression (plasticity) with respect to cold and heat stress ($P<0.001$), and also exhibited expression patterns among ecotypes that correlated with geographic temperature ($P<0.016$). (**a**) The $\log_2$ fold-change associated with MYB29 over six time points at which measurements were obtained. The horizontal line denotes a $\log_2$ fold-change of zero, while solid and dashed lines represent fold changes associated with cold and heat stress, respectively. (**b**) Plots of expression values associated with each ecotype with respect to the average temperature of geographic locations from which ecotypes originated. The three replicate gene expression measurements are shown for each ecotype, and the slanted line represents the least-squares regression estimate (β). The dotted horizontal line represents the mean expression intensity among all ecotypes.

Discussion

Microarray studies in plant species have identified genes underlying stress-response pathways by focusing on genes exhibiting plastic patterns of gene expression (Kreps *et al*., 2002; Hazen *et al*., 2003; Takahashi *et al*., 2004; Seki *et al*., 2004; Liu *et al*., 2005). Patterns of gene expression among naturally occurring variants, however, have not been considered in previous investigations, despite the insight this approach may offer into the adaptive significance of genes that underlie stress-regulatory networks. This study examined both plastic and adaptive gene expression patterns associated with temperature stress in the *A. thaliana* model system. Our results show that genes with plastic expression patterns generally exhibit greater differentiation in expression levels among ecotypes, and moreover, that the covariance of this variation with geographic temperature gradients was larger than among non-plastic genes. To our knowledge, these findings provide the first evidence to suggest that plasticity of gene expression is associated with adaptive value. Our findings also demonstrate that investigations into the plasticity of gene expression patterns, in combination with inferences regarding adaptive value, can drastically reduce the number of candidate genes identified by microarray analyses. While more than 16 000 genes exhibited plastic expression patterns with respect to cold or heat stress, only 1–4% were associated with significant evidence of adaptive value, and we ultimately identified a small set of only 43 gene candidates. These genes provide insight into the molecular basis of temperature stress response in *Arabidopsis* and represent promising candidates for future experimental investigation.

The 43 genes we identified are associated with a diverse range of processes, including transcription factor activity, membrane metabolism and signal transduction. Transcription factors play a widespread role in response to many forms of stress in *Arabidopsis*, including cold- and heat-stress treatments, and have often proven useful in bioengineering applications (Chen and Zhu, 2004). The genes we identified included three transcription factors (At5g07690 (MYB29), At1g04240 (SHY2), At3g56400 (WRKY70)), each of which had previously been associated with stress-signaling hormones (e.g., jasmonic acid, salicyclic acid, auxin) and stress response. Other genes we identified were involved in hydrolase activity (At1g14250, At5g28050), lipid metabolism and glycerol biosynthesis (At2g27360), as well as lipid binding and transport (At3g18280). These genes may play a role in altering the saturation, lipid asymmetry and molecular composition of membranes, which are processes that modulate membrane fluidity under heat and cold temperature stress (Sung *et al*., 2003). One important goal for understanding the physiological response to temperature stress in plants is to elucidate the molecular basis of signal-transduction pathways (Knight and Knight, 2001), many elements of which are thought to overlap between cold and heat stress (Sung *et al*., 2003). Several genes we identified may play a role in signal transduction, including two proteins associated with signal transducer activity gene ontology terms (At5g47800, At4g14010 (RALFL32)), along with additional kinases and proteins involved in phosphorylation (At3g02020, At1g70250, At5g35170). The two glutathione transferases we identified (At1g27130 (ATGSTU13), At1g17190 (ATGSTU26)) likely have an indirect role in signal transduction, since these enzymes aide in detoxification of reactive oxygen species, which are involved in signal transduction and other processes under many different stress treatments (Pastori and Foyer, 2002). A cyclophilin with peptidyl-prolyl *cis-trans* isomerase activity (At2g21130) was among the most interesting genes we identified, since cyclophilins have been associated with a wide range of processes (including signal transduction), and have been found to catalyze the folding of certain proteins while serving as molecular chaperones (Godoy *et al*., 2000).

The identification of temperature-related genes through microarray analysis represents only a first step towards understanding their role in cold- and heat-stress-regulatory pathways. This is certainly the case for not only the genes identified on the basis of plastic expression patterns alone, but is also true of genes associated with both plastic and adaptive patterns of gene expression. Candidate genes with a well-supported role in stress-response pathways provide good prospects for subsequent experimental study, which will generally include the generation of T-DNA insertion mutants along with analysis of post-transcriptional processes. Genes exhibiting both plastic and adaptive gene expression patterns are a highly selective class, and on the basis of microarray analysis, appear to represent the strongest available candidates for subsequent experimental study. It should be emphasized, however, that genes exhibiting plastic expression patterns, without evidence of adaptive value, may still occupy important roles in stress-response pathways. It is possible, for example, that some genes involved in stress-response pathways cross-talk with critical physiological processes that are highly conserved among natural variants. Such genes may exhibit plastic expression responses to stress, but like quantitative traits closely linked with fitness, may be associated with little heritable variation at the population level such that expression levels are not responsive to selection pressure (Mousseau and Roff, 1987). It should not be concluded, therefore, that genes yielding no evidence of adaptive value do not play a role in stress-response pathways. At the same time, however, this reasoning suggests that genes involved in stress-response pathways without exhibiting heritable variation in gene expression would offer poor prospects for stress-resistance engineering, since it may be more likely that modifying the expression of such genes will entail negative side-consequences influencing plant fitness and survival.

The adaptive significance of genes in this study was inferred based upon the covariance of expression levels among ecotypes with the temperature associated with geographic locations from which ecotypes originated. This approach provides evidence regarding the adaptive significance of gene expression levels, and is similar to methodologies that have been used to evaluate the adaptive significance of quantitative traits in *Arabidopsis* and other species (e.g., Maloof *et al*., 2001; Stinchcombe *et al*., 2004; Heibo *et al*., 2005; Lempe *et al*., 2005; Umina *et al*., 2005). However, we cannot rule out the possibility that for some genes, other factors covary with temperature and thus serve as lurking variables contributing to the association between expression levels and temperature. One such lurking variable, for instance, could be the genealogical relationships among *Arabidopsis* ecotypes. For cases in which genetic distance data is available, a conservative method of controlling for such relationships has been implemented in a recent study involving four populations of the fish *Fundulus heteroclitus* (Whitehead and Crawford, 2006). Analysis of expression levels within recombinant progeny formed by crossing of divergent strains may also aid detection of adaptively significant variation in gene expression (Filatov *et al*., 2006). Ultimately, however the ideal method of identifying adaptive gene expression patterns is an experimental approach in which replicate lineages are subjected to temperature stress selection along with non-selected control lineages. This methodology has been used to study stress effects in several bacterial species (Riehle *et al*., 2003; Fong *et al*., 2005), but to our knowledge, has not yet been implemented in *Arabidopsis* or other plant species. Microarray analyses of lineages artificially selected for increased stress resistance in combination with already available data on the plasticity of gene expression will provide a powerful combination for identifying genes with critical roles in stress-regulatory networks.

Temperature extremes and other forms of stress have a large impact on the evolution of plant populations in the wild, and are key factors limiting the agronomic yield of commercially valuable crop species (Boyer, 1982). The identification and functional analysis of major genes underlying stress-response pathways will be an important step towards the successful development of stress-resistant plant species (Denby and Gehring, 2005; Vinocur and Altman, 2005). The effectiveness of transcriptional profiling for identifying such genes, however, may be limited for investigations in which plasticity of expression patterns alone is the criterion by which genes are evaluated (Feder and Walser, 2005). The results of this study demonstrate that natural variation can be exploited to provide a new dimension to the data considered in previous transcriptional profiling analyses. This insight into the adaptive significance of gene expression levels has not been widely incorporated into our existing knowledge of stress-regulatory networks. The overall approach we have implemented, moreover, can be applied to identify key genes underlying genomic responses to nearly any type of abiotic or biotic stress factor. We therefore anticipate that the continued availability of large-scale gene expression data sets generated by the AtGenExpress consortium, in combination with new data sets generated by natural or artificial selection experiments, will provide a valuable toolkit for dissecting the molecular mechanisms that coordinate stress response in both plant and animal species.

Acknowledgements

We thank Hsueh-Han Yeh for assistance in preparing Supplementary data, as well as three anonymous reviewers for providing helpful comments on this manuscript. This work was supported by National Science Foundation award MCB-0348074 and a research grant from the Michigan State University Quantitative Biology and Modeling Initiative. Additional support was provided by the Department of Statistics and Probability and Department of Plant Biology at Michigan State University. The authors gratefully acknowledge members of the AtGenExpress consortium (Thomas Altmann, Pascal von Koskull-Döring, Jörg Kudla, Lutz Nover, Detlef Weigel) and the *Arabidopsis* Functional Genomics Network for providing the gene expression data analyzed in this study.

References

Allison DB, Xiangqui C, Page GP, Sabripour M (2006). Microarray data analysis: from disarray to consolidation and consensus. *Nat Rev Genet* **7**: 55–65.

Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: a powerful and practical approach to multiple testing. *J Roy Stat Soc B* **57**: 289–300.
Beissbarth T, Speed TP (2004). Gostat: find statistically over-represented gene ontologies within a group of genes. *Bioinformatics* **20**: 1464–1465.
Bohnert HJ, Qingqiu G, Pinghua L, Ma S (2006). Unraveling abiotic stress tolerance mechanisms – getting genomics going. *Curr Opin Plant Biol* **9**: 180–188.
Boyer JS (1982). Plant productivity and environment. *Science* **218**: 443–448.
Canales RD, Luo Y, Willey JC, Austermiller B, Barbacioru CC, Boysen C *et al.* (2006). Evaluation of DNA microarray results with quantitative gene expression platforms. *Nat Biotech* **24**: 1115–1122.
Chen WJ, Zhu T (2004). Networks of transcription factors with roles in environmental stress response. *Trends Plant Sci* **9**: 591–596.
Denby K, Gehring C (2005). Engineering drought and salinity tolerance in plants: lessons from genome-wide expression profiling in *Arabidopsis*. *Trends Biotechnol* **23**: 547–552.
Falconer DS, Mackay TFC (1996). *Introduction to Quantitative Genetics*. Longman, Harlow, UK.
Feder ME, Walser JC (2005). The biological limitations of transcriptomics in elucidating stress and stress responses. *J Evol Biol* **18**: 901–910.
Filatov V, Dowdle J, Smirnoff N, Ford-Lloyd B, Newbury HJ, Macnair MR (2006). Comparison of gene expression in segregating families identifies genes and genomic regions involved in a novel adaptation, zinc hyperaccumulation. *Mol Ecol* **15**: 3045–3059.
Fong SS, Joyce AR, Palsson B (2005). Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Res* **15**: 1365–1372.
Gibson G, Weir B (2005). The quantitative genetics of transcription. *Trend Genet* **21**: 616–623.
Godoy AV, Lazzaro AS, Casalongue CA, San Segundo B (2000). Expression of a *Solanum tuberosum* cyclophilin gene is regulated by fungal infection and abiotic stress conditions. *Plant Sci* **152**: 123–134.
Hazen SP, Wu Y, Kreps JA (2003). Gene expression profiling of plant responses to abiotic stress. *Funct Integr Genomics* **3**: 105–111.
Heibo E, Magnhagen C, Vøllestad LA (2005). Latitudinal variation in life history traits in Eurasian Perch. *Ecology* **86**: 3377–3386.
Hennig L, Menges M, Murray JA, Gruissen W (2003). *Arabidopsis* transcript profiling on Affymetrix GeneChip arrays. *Plant Mol Biol* **53**: 457–465.
Holm S (1979). A simple sequentially rejective multiple test procedure. *Scand J Statist* **6**: 65–70.
Hoffmann AA, Parsons PA (1991). *Evolutionary Genetics and Environmental Stress*. Oxford University Press, Oxford.
Huey RB, Carlson M, Crozier L, Frazier M, Hamilton H, Harley C *et al.* (2002). Plants versus animals: do they deal with stress in different ways? *Integr Comp Biol* **42**: 415–423.
Ji H, Davis RW (2006). Data quality in genomics and microarrays. *Nat Biotech* **24**: 1112–1113.
Khaitovich P, Weiss G, Lachmann M, Hellmann I, Enard W, Muetzel B *et al.* (2004). A neutral model of transcriptome evolution. *Plos Biol* **2**: 682–689.
Knight H, Knight MR (2001). Abiotic stress signaling pathways: specificity and cross-talk. *Trends Plant Sci* **6**: 262–267.
Kreps JA, Wu Y, Hur-Song C, Zhu T, Wang X, Harper JF (2002). Transcriptomic changes for *Arabidopsis* in response to salt, osmotic, and cold stress. *Plant Phys* **130**: 2129–2141.
Lempe J, Balasubramanian S, Sureshkumar S, Singh A, Schmid M, Weigel D (2005). Diversity of flowering responses in wild *Arabidopsis thaliana* strains. *Plos Genetics* **1**: 109–118.
Liu F, VanToai T, Moy LP, Bock G, Linford LD, Quackenbush J (2005). Global transcription profiling reveals comprehensive insight into hypoxic response in *Arabidopsis*. *Plant Phys* **137**: 1115–1129.
Maloof JN, Borevitz JO, Dabi T, Lutes J, Nehring RB, Redfern JL *et al.* (2001). Natural variation in light sensitivity of *Arabidopsis*. *Nat Genet* **29**: 441–446.
Mousseau TA, Roff DA (1987). Natural-selection and the heritability of fitness components. *Heredity* **59**: 181–197.
Nuzhdin SV, Wayne ML, Harmon KL, McIntyre LM (2004). Common pattern of evolution of gene expression level and protein sequence in *Drosophila*. *Mol Biol Evol* **21**: 1308–1317.
Oleksiak MF, Churchill GA, Crawford DL (2002). Variation in gene expression within and among natural populations. *Nat Genet* **32**: 261–266.
Oleksiak MF, Roach JL, Crawford DL (2005). Natural variation in cardiac metabolism and gene expression in *Fundulus heteroclitus*. *Nat Genet* **37**: 67–72.
Pastori GM, Foyer H (2002). Common components, networks, and pathways of cross-tolerance to stress. The central role of 'redox' and abscisic acid-mediated controls. *Plant Phys* **129**: 460–468.
Redman JC, Haas BJ, Tanimoto G, Town CD (2004). Development and evaluation of an *Arabidopsis* whole genome Affymetrix probe array. *Plant J* **38**: 545–561.
Riehle MM, Bennett AF, Lenski RE, Long AD (2003). Evolutionary changes in heat-inducible gene expression in lines of *Escherichia coli* adapted to high temperature. *Physiol Genomics* **14**: 47–58.
Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M *et al.* (2005). A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* **37**: 501–506.
Seki M, Satou M, Sakurai T, Akiyama K, Lida K, Ishida J *et al.* (2004). Ripken *Arabidopsis* full-length (RAFL) cDNA and its applications for expression profiling under abiotic stress conditions. *J Exp Biol* **55**: 213–223.
Shimizu KK (2002). Ecology meets molecular genetics in *Arabidopsis*. *Popul Ecol* **44**: 221–233.
Smyth GK (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article 3.
Stinchcombe JR, Weinig C, Ungerer M, Olsen KM, Mays C, Halldorsdottir SS *et al.* (2004). A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. *Proc Natl Acad Sci* **101**: 4712–4717.
Sung D-Y, Kaplan F, Lee K-J, Guy CL (2003). Acquired tolerance to temperature extremes. *Trends Plant Sci* **8**: 179–187.
Takahashi S, Seki M, Ishida J, Satou M, Sakurai T, Narusaka M *et al.* (2004). Monitoring the expression profiles of genes induced by hyperosmotic, high salinity, and oxidative stress and abscisic acid treatment in *Arabidopsis* cell culture using a full-length cDNA microarray. *Plant Mol Biol* **56**: 29–55.
Umina PA, Weeks AR, Kearney MR, McKechnie SW, Hoffmann AA (2005). A rapid shift in a classic clinal pattern in *Drosophila* reflecting climate change. *Science* **308**: 691–693.
Vinocur B, Altman A (2005). Recent advances in engineering plant tolerance to abiotic stress: achievements and limitations. *Curr Opin Biotech* **16**: 123–132.
Whitehead A, Crawford DL (2006). Neutral and adaptive variation in gene expression. *Proc Nat Acad Sci* **103**: 5425–5430.
Wu Z, Irizarry R, Gentleman R, Martinez Murillo F, Spencer F (2004). A model based background adjustment for oligonucleotide expression arrays. *J Amer Stat* **99**: 909–917.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)