

TECHNICAL ADVANCE

Real-time RT-PCR profiling of over 1400 *Arabidopsis* transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes

Tomasz Czechowski, Rajendra P. Bari, Mark Stitt, Wolf-Rüdiger Scheible* and Michael K. Udvardi*
Max-Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Golm, Germany

Received 18 November 2003; revised 20 January 2004; accepted 27 January 2004.

*For correspondence (fax +49 331 5678 250; e-mail Scheible@mpimp-golm.mpg.de, Udvardi@mpimp-golm.mpg.de).

Summary

To overcome the detection limits inherent to DNA array-based methods of transcriptome analysis, we developed a real-time reverse transcription (RT)-PCR-based resource for quantitative measurement of transcripts for 1465 *Arabidopsis* transcription factors (TFs). Using closely spaced gene-specific primer pairs and SYBR[®] Green to monitor amplification of double-stranded DNA (dsDNA), transcript levels of 83% of all target genes could be measured in roots or shoots of young *Arabidopsis* wild-type plants. Only 4% of reactions produced non-specific PCR products. The amplification efficiency of each PCR was determined from the log slope of SYBR[®] Green fluorescence versus cycle number in the exponential phase, and was used to correct the readout for each primer pair and run. Measurements of transcript abundance were quantitative over six orders of magnitude, with a detection limit equivalent to one transcript molecule in 1000 cells. Transcript levels for different TF genes ranged between 0.001 and 100 copies per cell. Only 13% of TF transcripts were undetectable in these organs. For comparison, 22K *Arabidopsis* Affymetrix chips detected less than 55% of TF transcripts in the same samples, the range of transcript levels was compressed by a factor more than 100, and the data were less accurate especially in the lower part of the response range. Real-time RT-PCR revealed 35 root-specific and 52 shoot-specific TF genes, most of which have not been identified as organ-specific previously. Finally, many of the TF transcripts detected by RT-PCR are not represented in *Arabidopsis* EST (expressed sequence tag) or Massively Parallel Signature Sequencing (MPSS) databases. These genes can now be annotated as expressed.

Keywords: transcript profiling, transcription factor, root, shoot, real-time RT-PCR, *Arabidopsis*.

Introduction

Transcription factors (TFs) are master-control proteins in all living cells. They often exhibit sequence-specific DNA binding and are capable of activating or repressing transcription of multiple target genes. In this way, they control or influence many biological processes, including cell cycle progression, metabolism, growth and development, and responses to the environment. As many TFs are themselves regulated at the level of transcription (Chen *et al.*, 2002), knowing where and when TF genes are transcribed, and how such transcription is affected by internal or external cues can be valuable in elucidating the specific biological roles of the cognate proteins.

With the completion of the *Arabidopsis thaliana* genome sequence, it became possible for the first time to carry out a

census of putative TFs in a higher plant. The *Arabidopsis* genome contains about 30 000 annotated loci (<http://www.arabidopsis.org>), approximately 5–6% of which code for putative TFs (Davuluri *et al.*, 2003; Ratcliffe and Riechmann, 2002; Riechmann *et al.*, 2000). Less than 10% of these have been characterized genetically. Given that a large proportion (approximately 40%) of *Arabidopsis* genes remain to be annotated with regard to function (AGI, 2000), it is likely that the number of TF genes will increase; in fact novel classes of TFs are still being discovered (Riechmann, 2002). TF genes are generally expressed at low levels in plants, frequently in a cell-type or tissue-specific manner, and often only transiently during development (e.g. *LEAFY*

(Weigel *et al.*, 1992); *SHATTERPROOF1* and 2 (Liljegren *et al.*, 2000); *WUSCHEL* (Mayer *et al.*, 1998) or *MONOPTEROS* (Hardtke and Berleth, 1998). Although DNA and oligonucleotide arrays, such as Affymetrix chips, that contain most of the predicted genes of *Arabidopsis* are now available for transcriptome analysis, it is likely that the transcripts of many TF genes will be difficult to detect and quantify with DNA array technologies. Reverse transcription (RT)-PCR is estimated to be at least 100-fold more sensitive than DNA arrays in detecting transcripts (Horak and Snyder, 2002). In yeast, for instance, kinetic or real-time RT-PCR was able to detect transcripts of virtually all TF genes, which varied in abundance by over four orders of magnitude. In contrast, DNA arrays were unable to detect most yeast TF transcripts in a reliable manner (Holland, 2002). The limitations of DNA arrays for TF transcript detection are likely to be even greater in *Arabidopsis*, which contains a large number of different cell types, only a fraction of which will express a particular TF, e.g. *WUS* (Mayer *et al.*, 1998). For this reason, we have developed a library of more than 1400 PCR primer (oligonucleotide) pairs that can be used to quantify transcripts of the majority of TF genes in *Arabidopsis* by real-time RT-PCR. Using these primers, together with SYBR[®] Green and an ABI PRISM[®] 7900HT 384-well-plate PCR system, we are able to measure the abundance of virtually all *Arabidopsis* TF transcripts (via cDNA) in the same sample in a single day.

Here, we present the first results obtained with this new resource. Besides providing the first comprehensive estimate of the range of TF transcript levels in *Arabidopsis*, we identify 36 putative root-specific and 52 putative shoot-specific TF genes in *Arabidopsis*, which may play important roles in the development or function of these distinct organs. In addition, a comparison between real-time RT-PCR and Affymetrix chip technology for measuring gene transcript levels is made, which highlights the value of this new resource with respect to its sensitivity and its ability to provide quantitative data.

Results

PCR primer design and reaction specificity

To ensure maximum specificity and efficiency during PCR amplification of TF cDNA under a standard set of reaction conditions, a stringent set of criteria was used for primer design. This included predicted melting temperatures (T_m) of $60 \pm 2^\circ\text{C}$, primer lengths of 20–24 nucleotides, guanine–cytosine (GC) contents of 45–55% and PCR amplicon lengths of 60–150 base pairs (bp). In addition, when possible, at least one primer of a pair was designed to cover an exon–exon junction (see Supplementary Material, Table S1), according to the gene structure models at MIPS (The

Munich Information Center for Protein Sequences; <http://mips.gsf.de>) and/or TAIR (The Arabidopsis Information Resource; <http://www.arabidopsis.org>). This was the case for approximately 74% of all primer pairs.

The specificity of PCR primers was tested using first-strand cDNA derived from either plate-grown *Arabidopsis* seedling shoots or roots, or whole seedlings grown in axenic cultures. Total ribonucleic acid (RNA) was always treated with DNaseI prior to purification of poly(A)⁺ RNA. Before proceeding with first-strand cDNA synthesis, complete degradation of genomic DNA in RNA preparations was confirmed by PCR analysis. All 1465 TF primer pairs (Supplementary Material, Table S1) were tested for their efficacy in amplifying the specific target cDNA from roots and shoots. For each tissue, a single pool of cDNA was used to seed all real-time RT-PCRs, each of which contained a unique pair of TF primers. Approximately 83% of all primer pairs produced a single DNA product of the expected size, as exemplified in Figure 1(b). Only 4% of reactions yielded more than one PCR product. Thirteen per cent (193) of reactions yielded no PCR product from root or shoot cDNA after 40 PCR cycles, indicating that the target genes were probably not expressed in these organs and growth

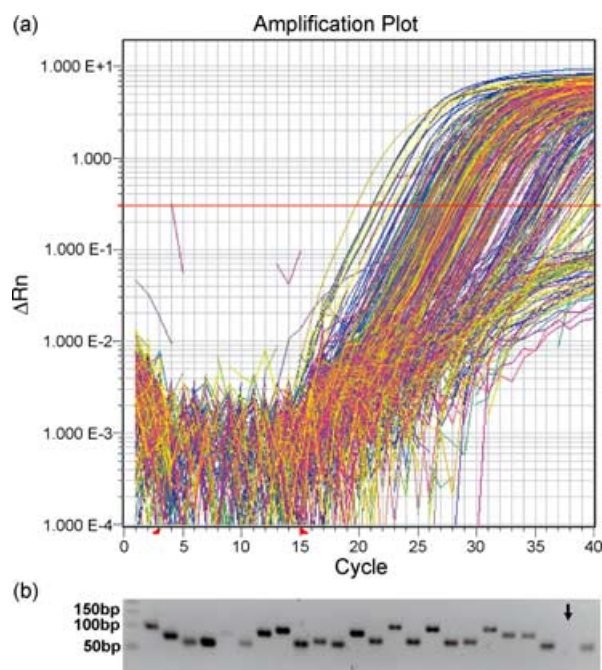


Figure 1. Specificity of RT-PCR.

(a) Typical real-time RT-PCR amplification plots of 384 TF genes showing increase in SYBR[®] Green fluorescence (ΔR_n , log scale) with PCR cycle number. Note the similar slope of most curves as they cross the fluorescence threshold of 0.3, which reflects similar amplification efficiencies. Note also the low proportion of amplification curves that does not cross the fluorescence threshold. Such reactions yield no detectable product when visualized on agarose gels.

(b) Separation of RT-PCR products on 4% (w/v) agarose gels revealed single products of the expected size for most reactions, with few reactions yielding no product (arrow). Size standards in bp are indicated at the left.

conditions (see Supplementary Material, Table S1). Primer pairs for 56 of these genes were complementary to exon sequences only, which enabled us to check the primers on genomic DNA. Forty-four of these primer pairs were tested, and all produced a unique PCR product of the expected size from genomic DNA. This result confirmed not only that the primers were effective, but also that the target genes were not expressed in plants under the conditions studied. The remaining 137 primer pairs contained at least one primer spanning an intron, which prohibited a similar check of primer efficacy using genomic DNA. Nonetheless, the percentage (approximately 71%) of intron-spanning primer pairs amongst those that failed to yield PCR amplicons in our experiments was not higher than that of such primer pairs (approximately 74%) that did yield specific amplicons. Therefore, failure to predict intron-splicing sites correctly probably does not account for failure to detect these transcripts/cDNA in our experiments.

Data from gel-electrophoresis analysis of the amplified PCR products (Figure 1b) were confirmed by melting curve analysis, which was performed by the PCR machine after cycle 40. A more stringent test of the specificity of PCRs was performed by sequencing the products of nine Myb/Myb-like genes (AT3G01140; AT3G02940; AT3G61250; AT4G05100; AT5G02320; AT5G15310; AT5G16770; AT5G54230; and AT5G65230) and eight basic helix-loop-helix (bHLH)-type genes (AT3G19860; AT3G56970; AT3G56980; AT5G08130; AT5G09750; AT5G10570; AT5G37800; and AT5G46830). Genes were chosen from these two families because each family contains many members (>100) with a high degree of sequence similarity in the DNA-binding domains. The chosen genes also exhibited a wide range (>10³) of expression levels. In each case, the sequence of the PCR product matched that of the intended target cDNA, although primers were sometimes placed in conserved regions, confirming the exquisite specificity of the primer pairs.

Dynamic range, sensitivity and robustness of real-time PCR

The threshold cycle C_T is the cycle number (rarely a whole number) at which SYBR[®] Green fluorescence (ΔR_n) in a PCR reaches an arbitrary value during the exponential phase of DNA amplification (set at 0.3 in all of our experiments: see Figure 1a). For an ideal reaction, the number of double-stranded DNA (dsDNA) molecules doubles after each PCR cycle. In this case, a difference in C_T (ΔC_T) of 1.0 indicates a 2-fold difference in the amount of DNA at the start of a reaction, a ΔC_T of 2.0 is equivalent to a fourfold difference, etc. Therefore, C_T is inversely proportional to the logarithm of the amount of target DNA present at the start of a PCR (Figure 2a), or 2^{C_T} is inversely proportional to the amount of target DNA. To make data from real-time RT-PCR easier to understand, we often plot it as 2^{40-C_T} , which is directly

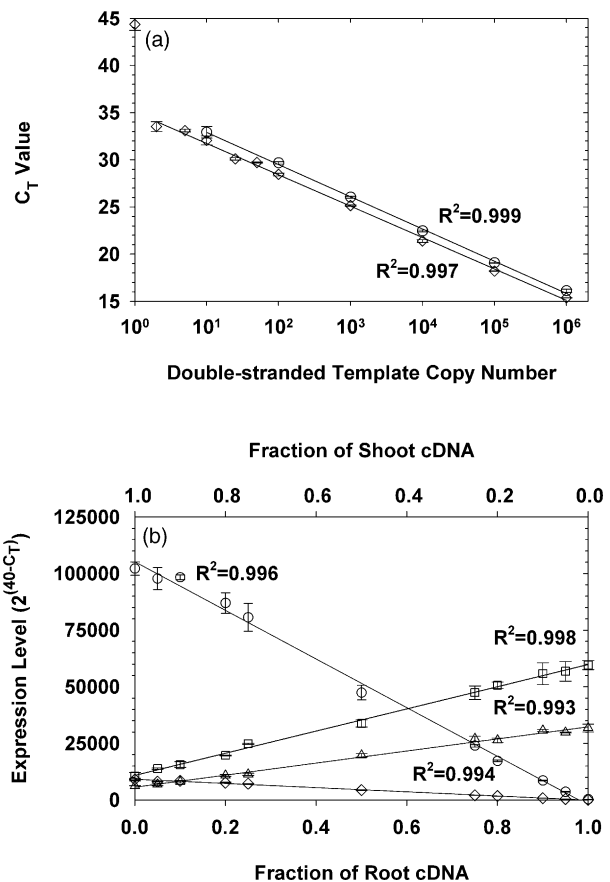


Figure 2. Sensitivity and robustness of RT-PCR.

(a) Relationship between amplification kinetics (C_T) and copy number of a *LUC* gene (O) and an intragenic DNA fragment (◇) in reactions containing a complex pool of 1 ng *Arabidopsis* cDNA.

(b) Relationship between the expression level, 2^{40-C_T} , and the fraction of root or shoot cDNA in a mixture of the two totalling 1 ng, for the four genes AT1G13300 (O); AT1G34670 (◇); AT4G32980 (△) and AT5G44190 (□). Symbols in both panels represent the mean and SD of three replicate measurements.

proportional to target DNA amount. The number 40 above is somewhat arbitrary but was chosen because PCRs are typically stopped at cycle 40.

The sensitivity and robustness of quantification by real-time RT-PCR were investigated in two ways. In the first approach, C_T was measured for a cloned luciferase (*LUC*) gene and an amplified intergenic region of *Arabidopsis*, which were diluted serially from 1 million copies to a single copy and added to a complex matrix of *Arabidopsis* root cDNA (1 ng or approximately 10⁹ cDNA molecules). Amplification of the 60-bp *LUC* gene fragment and the 75-bp intergenic region resulted in C_T values of approximately 16 when 1 million copies of template DNA were introduced into reactions (Figure 2a). An inverse linear relationship between the logarithm of copy number and C_T was observed down to 10 or 2 copies of the *LUC* gene and the intergenic fragment, respectively, reflecting a PCR efficiency of greater

than 98% in both cases (Pfaffl, 2001). With fewer than 10 copies of the *LUC* gene at the start of PCR, a non-specific product was amplified (not shown), which resulted in an effective detection limit of 10 molecules in this case. The effective detection limit for the intergenic region was two copies; the template was undetectable in further dilutions, which can most easily be explained by a complete absence of the template in these reactions (Figure 2a). Thus, we were able to detect as little as two double-stranded copies of a target gene within a complex mixture of 1 ng cDNA. Assuming that the average length of an mRNA (cDNA molecule) is 1.3 kb (AGI, 2000; Haas *et al.*, 2002) and that the average number of transcripts per plant cell is 2×10^5 (Kamalay and Goldberg, 1980; Kiper, 1979; Ruan *et al.*, 1998), we estimate the detection limit of our system to be close to one transcript per 1000 cells, or 0.001 transcripts per cell.

The second approach to assess the sensitivity, robustness and linearity of quantification by real-time RT-PCR involved mixing different amounts of root and shoot cDNA prior to determining C_T values for four root- or shoot-specific genes in each mixture. A linear relationship between 2^{40-C_T} and root/shoot cDNA amount was obtained for each gene over the whole range of mixtures (Figure 2b), which showed that the precision of real-time PCR measurements is not influenced by the complex milieu of molecules present in typical PCRs.

Precision of real-time RT-PCR

The technical precision or reproducibility of real-time RT-PCR measurements was assessed by performing replicate measurements in separate PCR runs, using the same pool of cDNA (intra-assay variation) or two different pools of cDNA obtained independently from the same batch of total RNA (interassay variation; Figure 3). Precision, as reflected by the correlation coefficient, was high in both cases, with the intra-assay variation ($R^2 = 0.9953$), lower than the inter-assay variation ($R^2 = 0.9571$), as expected. Transcript levels varied over five orders of magnitude (for example, see

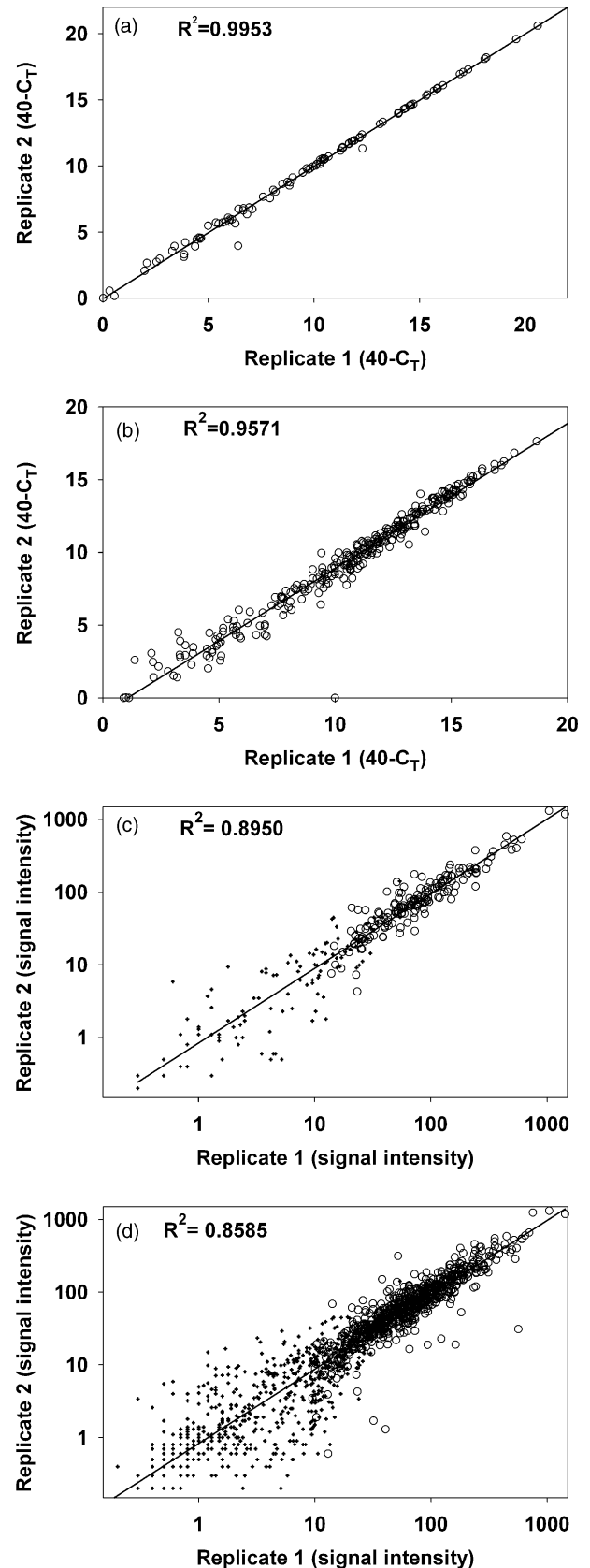


Figure 3. Technical precision of real-time RT-PCR and Affymetrix full genome arrays.

(a, b) Real-time RT-PCR was used to obtain duplicate measurements of TF cDNA levels (a) from the same RT reaction or (b) from two separate RT reactions. The same sample of total RNA from shoots was used throughout to preclude biological variation. Thus, (a,b) illustrate intra- and interassay technical variabilities, respectively. Two separate measurements of (a) 101 genes and (b) 298 genes are compared.

(c, d) The Affymetrix full genome array (ATH1) was used to measure TF transcript levels via cRNA derived from two separate RT reaction products starting with the same RNA sample as for (a,b). (c) Interassay technical variability is illustrated for the 277 TF genes on the Affymetrix array that correspond to the 298 genes shown in panel (b). The 169 genes that were categorized 'present' in both replicates by the Affymetrix software are depicted as circles and those called 'absent' as cross-hairs. A regression line and the corresponding correlation coefficient (R^2) is shown for the entire set of 277 genes. Interassay technical variability of all 1275 TF genes represented on the Affymetrix array is depicted in (d).

Figure 3b), with the most highly expressed TF genes apparently as active as the house-keeping gene *Ubiquitin10* (*UBQ10*, AT4G05320).

Efficiency of PCRs

The number of cycles needed to reach a given fluorescence intensity depends on not only the amount of cDNA in the extract but also the amplification efficiency (E). In the ideal case, when the amount of cDNA is doubled in each reaction cycle, $E = 1$. As mentioned above, PCR primers were designed to produce short amplicons, typically between 60 and 150 bp (see Supplementary Material, Table S1), to maximize E . While preliminary measurements (see Figure 2, for example) showed that efficiencies of virtually 100% were achieved in some reactions, we expected that a significant fraction of the 1465 TF-specific PCRs would have lower efficiency.

Different methods are available for estimating PCR efficiency (for a compilation, see <http://www.weihenstephan.de/gene-quantification/>). The classical method uses C_T values obtained from a series of template dilutions, as illustrated in Figure 2(a) (e.g. Pfaffl, 2001). An alternative method utilizes absolute fluorescence data captured during the exponential phase of amplification of each real-time PCR (Ramakers *et al.*, 2003). Comparison of the two methods yielded very similar amplification efficiencies for a subset of 46 TF primer pairs (Supplementary Material, Table S2). Hence, we used the latter method to establish amplification efficiencies for all 1465 primer pairs, as it does not require standard curves for every primer pair, and because it allows estimation of the efficiency for each individual PCR.

The E -value is derived from the log slope of the fluorescence versus cycle number curve for a particular primer pair, using the equation $(1 + E) = 10^{\text{slope}}$ (Ramakers *et al.*, 2003). Inspection of Figure 1(a) reveals that each PCR shows a lag and then enters an exponential phase, which appears in the logarithmic plot as a linear increase. The positions of the lines are offset, reflecting the different amount of cDNA for each transcription factor. The slopes of the lines are, in most cases, very similar showing that E is similar for most of the primer pairs. However, a small subgroup with a lower slope can be distinguished. The E -values for all of the primer pairs are summarized in Supplementary Material (Table S1). Of the 1465 primer pairs, 71 had E -values >0.90 , 402 between 0.90 and 0.81, 495 between 0.80 and 0.71, 244 between 0.70 and 0.61, 86 between 0.60 and 0.51 and 51 between 0.50 and 0.41. One hundred and sixteen primer pairs had E -values ≤ 0.40 , but *nota bene* they usually belonged to TF genes of categories 3 and 4 (see Supplementary Material, Table S1), which were barely or not at all detected in shoots or roots. Efficiency values were taken into account in all subsequent calculations, including calculations of the ratios of transcript levels in the shoot and root.

Comparison of technologies: real-time RT-PCR versus Affymetrix chips

As Affymetrix chips have become a 'gold-standard' for *Arabidopsis* transcriptome analysis, we were interested in comparing the results of real-time RT-PCR measurements of TF transcript levels with corresponding data from 'whole-genome' chips. Using the same preparations of RNA that had been used for RT-PCR analysis, Affymetrix chips detected (called 'present' twice in at least one organ by Affymetrix software) less than 55% of the putative transcription factors listed in Supplementary Material (Table S1). Interassay variation between replicate Affymetrix chips was greater than that of real-time RT-PCR, which indicated a lower precision of the Affymetrix technology, especially for low-abundance transcripts (see Figure 3c,d).

We did not necessarily expect a good correlation between signals obtained for the levels of the individual transcripts by real-time RT-PCR and Affymetrix chips. Unlike quantitative RT-PCR, hybridization-based technologies like Affymetrix chips are qualitative, and there is no strict linear relationship between signal strength and transcript amount for different genes (Holland, 2002). Nonetheless, genes determined to be highly expressed by real-time RT-PCR typically yielded high signals on Affymetrix chips. A large majority (90%) of the 503 genes that were categorized as 'absent' by Affymetrix software were detected by real-time PCR (see above) albeit at lower levels, as expected. Overall, there was little quantitative agreement between the two data sets for 1083 TF genes that were analysed from shoots (Figure 4) or roots (data not presented).

Identification of root- and shoot-specific TF genes by real-time RT-PCR

The real-time RT-PCR resource for TF transcript profiling was used to identify root- and shoot-specific TF genes, to test its efficacy in identifying known organ-specific TFs, and to identify novel root- or shoot-specific TFs for future study. From amongst the 1214 TF gene transcripts that were detected by real-time RT-PCR in roots and shoots, 438 (36%) were differentially expressed (shoot/root (S/R) ratio >4 or <4 ; Figure 5). Approximately 10.5% (127/1214) of the TF genes exhibited a greater than 20-fold difference in expression level in shoots compared to roots (indicated by the dashed lines in Figure 5). We considered these as putative shoot- or root-specific genes. Many of these genes were not previously reported to be organ-specific, and several of the genes are not represented on the Affymetrix ATH1 array (Table 1, Table 2; Supplementary Material, Table S1). Organ-specific expression was confirmed for the 87 TF genes shown in Table 1 and Table 2 by repeating the real-time RT-PCR with a biological replicate (Supplementary Material, Figure S1 and Table S1). Biological replication

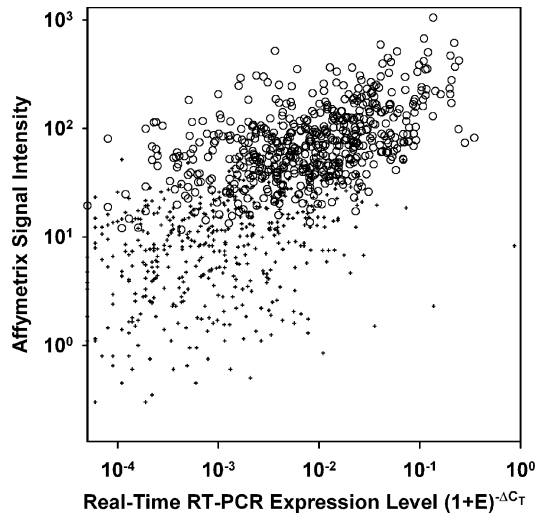


Figure 4. Comparison of shoot TF transcript levels measured by real-time RT-PCR and Affymetrix whole genome arrays. Normalized raw data from RT-PCR ($(1 + E)^{-\Delta C_T}$) were compared to normalized raw data from Affymetrix chips (\log_{10} of fluorescence signal) for the 1083 TF genes, which were detectable in shoots on the real-time RT-PCR platform and also present on the Affymetrix ATH1 gene array. Genes categorized as 'present' or 'absent' by Affymetrix software are depicted as circles or cross-hairs, respectively.

was also performed using Affymetrix analysis. The correlation coefficients calculated from the real-time RT-PCR data, and the Affymetrix data were comparable and higher than 0.70, when gene expression levels were plotted (shown for the root data in Figure S1, panels a and b). Plotting the replicated S/R ratios yielded a R^2 value of 0.87 for real-time RT-PCR (Figure S1, panel c) and a R^2 value of 0.78 for the Affymetrix approach (Figure S1, panel d). The mean S/R

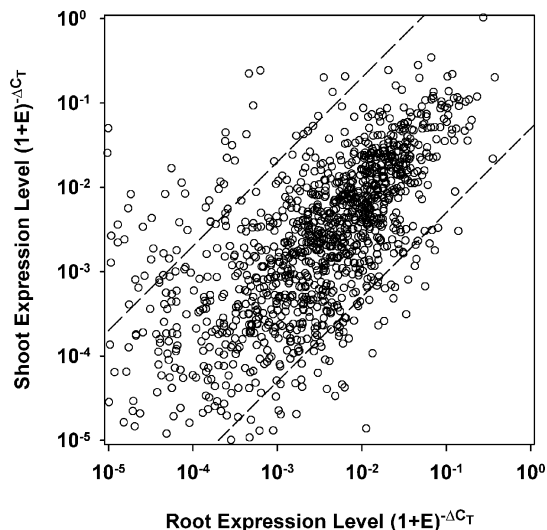


Figure 5. Comparison of TF transcript levels in shoots and roots. Normalized expression values ($(1 + E)^{-\Delta C_T}$) from real-time RT-PCR amplification of cDNA from shoots and roots are compared for 1214 genes (category 1 genes; see Supplementary Material, Table S1). Dashed lines indicate 20-fold differences in the shoot-to-root transcript levels.

ratio obtained for the confirmed organ-specific genes was compared to publicly available data from Massively Parallel Signature Sequencing (MPSS; <http://mpss.udel.edu/at/java.html>) of *Arabidopsis* (Table 1 and Table 2). The MPSS database contained signatures for 73 of the 87 genes that were found by real-time RT-PCR to show strong (>20-fold) differences in expression levels between the shoot and root. For this subset of 73, there was remarkably good qualitative agreement between the two technologies. In all but four cases, genes with a high S/R transcript ratio measured by RT-PCR also had a high ratio as determined by MPSS. In most of these cases, signature sequences were completely absent for roots. For genes with a very low S/R ratio, there was even better qualitative agreement between real-time RT-PCR and MPSS data. In general, data from Affymetrix arrays were also in qualitative agreement with real-time RT-PCR and MPSS data. Only in very few cases were data from the three different technologies at odds with one another.

To further investigate the reasons for discrepancies between real-time RT-PCR and Affymetrix chip data, the S/R ratios were calculated for both complete data sets, and plotted against each other (Figure 6a). At first glance, there was only weak agreement between the ratios obtained with the two technologies ($R^2 = 0.472$ for the entire set of 975 considered genes). A different picture emerged when the data set was split into groups of genes according to their Affymetrix shoot expression level (Figure 6b). For example, when the 50 TF genes with the highest Affymetrix shoot expression levels were analysed, there was quite good agreement with the S/R ratios estimated from real-time RT-PCR data ($R^2 = 0.727$). When genes with lower expression level were introduced (see Figure 6b), the correlation coefficient dropped continuously. In general, there was a clear correlation between the 'discrepancy' in the S/R ratios determined by the two technologies and the frequency of genes that were flagged 'absent' by Affymetrix software (Figure 6c). For example, about 7% of the genes showed a >10-fold discrepancy in the S/R ratio obtained from real-time RT-PCR and Affymetrix chips, and of these, about 80% were called 'absent' by the Affymetrix software. In contrast, 75% of the genes had similar S/R expression ratios (less than threefold discrepancy) in both data sets, of which only 46% were called 'absent' by the Affymetrix software (Figure 6c).

Discussion

We have developed a unique public resource for studying the expression of transcription factor genes in *Arabidopsis*. This resource, which is based upon highly multiplexed real-time RT-PCR with gene-specific primers, enabled us to measure transcript levels in roots or shoots of *Arabidopsis* seedlings for 1247 TF genes with high specificity and

Table 1 Shoot-specific TF genes identified by real-time RT-PCR

Gene identifier	Gene name	Real-time RT-PCR S/R expression ratio	Affymetrix S/R expression ratio	MPSS ^c	
				Shoot ^d signatures	Root ^d signatures
AT1G73870	<i>COL7</i>	8678 ^a	101 ^b	12	0
AT4G00180	<i>YAB3</i>	4284 ^a	6.8 ^b	67	0
AT2G30420		3855 ^a	6.9 ^b	11	0
AT2G45190	<i>YAB1</i>	1578 ^a	147 ^b	56	11
AT2G41940	<i>ZFP8</i>	1351	9.7 ^b	89	2
AT2G43010	<i>PIF4</i>	458	5.9	172	0
AT1G68520	<i>COL6</i>	458	29	993	0
AT5G25390		451 ^a	2.4 ^b	0	0
AT5G54630		423	10 ^b	91	0
AT3G15270	<i>SPL5</i>	408	8.3 ^b	0	0
AT1G73830	<i>BEE3</i>	373 ^a	3.6	31	0
AT5G46690		333	10.0 ^b	6	0
AT3G59060	<i>PIL6</i>	272	130 ^b	138	0
AT3G06120		264 ^a	0.81 ^b	N	N
AT1G18710		223	5.1 ^b	35	0
AT5G49330		218	5.3 ^b	0	0
AT5G15310		189	12 ^b	90	8
AT5G46880		184 ^a	2.4 ^b	2	0
AT3G58070		177 ^a	8.7 ^b	5	0
AT1G25440	<i>COL16</i>	158	N	250	0
AT1G62360		157	11	21	0
AT3G18010		152 ^a	N	19	0
AT3G57600		136	2.4	9	0
AT5G10570		131	N	2	7
AT4G01460		128	11 ^b	29	0
AT1G71030		126	9.3	223	2
AT2G36610		126	2.0 ^b	18	29
AT1G68190		122	11 ^b	71	6
AT5G56860		110	53 ^b	28	0
AT5G11190		97	N	N	N
AT3G24140		73	5.4	27	0
AT5G57150		73	0.32	32	269
AT4G14540		70	138 ^b	100	10
AT1G76110		64	33 ^b	25	0
AT2G39250		63	5.0 ^b	11	0
AT5G15800	<i>AGL2</i>	61	0.25 ^b	0	0
AT2G02450		51	29 ^b	44	0
AT3G61950		48 ^a	7.5	9	0
AT4G25490	<i>CBF1</i>	45	3.3 ^b	3	0
AT1G53160	<i>SPL4</i>	43	2.5	48	0
AT2G33810	<i>SPL3</i>	43	43 ^b	34	3
AT5G15850	<i>COL1</i>	37	15 ^b	297	0
AT5G44190	<i>GLK2</i>	35	8.6	193	7
AT4G32980	<i>ATH1</i>	33	13	47	0
AT1G75490		30	2.0 ^b	1	0
AT1G08810		27	6.8 ^b	21	0
AT4G25470	<i>CBF2</i>	27	1.7 ^b	33	0
AT1G33760		27	6.1 ^b	N	N
AT2G17950	<i>WUS</i>	26 ^a	0.78 ^b	3	0
AT3G02380	<i>COL2</i>	25	28 ^b	77	0
AT5G47220	<i>ERF2</i>	25	3.8 ^b	112	11
AT2G46870		20	9.6 ^b	19	2

Data for genes exhibiting a more than 20-fold ratio (mean value of two biological replicas) in transcript abundance between roots and shoots are presented. Data from Affymetrix chips and Massively Parallel Signature Sequencing (<http://mpss.udel.edu/at/java.html>) are included for comparison.

^aNominal value (transcripts undetectable in one kind of organ: C_T value = 40, in both biological replicas).

^bTranscripts called absent by Affymetrix software in at least one organ and in both biological replicas).

^cUnspecific MPSS signatures were not considered. N denotes gene not represented on Affymetrix chip or MPSS database.

^dp.p.m.

Table 2 Root-specific TF genes identified by real-time RT-PCR

Gene identifier	Gene name	Real-time RT-PCR R/S expression ratio	Affymetrix R/S expression ratio	MPSS ^c	
				Shoot ^d signatures	Root ^d signatures
AT4G37940	<i>AGL21</i>	136	174 ^b	0	19
AT5G06840		106	5.1 ^b	0	29
AT4G13620		89	N	0	47
AT5G15130		86	45 ^b	5	118
AT1G13300		81	12	12	171
AT2G22630	<i>AGL17</i>	68	8.0 ^b	N	N
AT4G10350		65	2.6 ^b	N	N
AT5G56960		53 ^a	N	N	N
AT4G33880		52	N	0	42
AT5G18560		43	0.9	0	4
AT3G09290		41	16 ^b	0	6
AT5G19790		40	18 ^b	18	115
AT5G02350		38	23 ^b	N	N
AT5G14340		38	4.5 ^b	0	44
AT2G42660		31	5.6 ^b	1	45
AT2G33720		31 ^a	0.7 ^b	N	N
AT4G01350		30	1.5	0	4
AT1G68150		29	9.7 ^b	0	43
AT5G52170		29	2.0 ^b	2	13
AT5G19520		28	13 ^b	4	463
AT1G29280		26	7.4	18	346
AT1G68880		26	47 ^b	3	72
AT3G24310		26	11 ^b	0	4
AT3G20840		26 ^a	2.7 ^b	0	0
AT3G25790		25	14 ^b	0	20
AT5G65790		25	27 ^b	1	76
AT1G64000		25	4.8 ^b	0	4
AT1G74500		24	38 ^b	0	176
AT3G45170		24 ^a	N	N	N
AT1G28160		22 ^a	1.3 ^b	0	13
AT3G12720		22	19 ^b	0	13
AT1G66470		21	7.3 ^b	0	44
AT1G17950		21	31 ^b	0	18
AT1G69810		20	26 ^b	0	94
AT1G79580		20	2.1	0	4

For explanation see legend to Table 1.

precision. Single PCR products of the expected size were obtained following RT-PCR for all of these genes, and sequencing of a subset of them confirmed the specificity of each PCR. Four per cent of the 1465 different TF RT-PCRs yielded more than the single expected product. Synthesis of new primer pairs should enable specific measurements to be made on the transcripts of these genes in the future.

Approximately 13% of TF gene transcripts were not detected in samples of roots or shoots of vegetative plants grown under the conditions used in these experiments. Of these, about a quarter of the genes have primers that do not span exon-exon junctions. All primer pairs tested from this subset yielded unique PCR products of the expected size from genomic DNA as template, showing that the primers have been correctly designed and do function. This indicates that these genes are expressed at extremely low

levels or not at all under these conditions. Transcripts of another third of these genes have meanwhile been detected in *Arabidopsis* siliques or in seedlings exposed to various nutrient stresses (A. Blacha, T. Czechowski, W.-R. Scheible and M. Udvardi; unpublished results).

The sensitivity and robustness of TF transcript quantification by real-time RT-PCR were outstanding. As few as two copies of a target DNA could be detected in a complex mixture of 10^9 cDNA molecules (Figure 2a). This corresponds to a detection limit of about one transcript per 1000 cells, or 0.001 transcripts per cell which is similar to values obtained for yeast (Holland, 2002). In contrast, detection limits of DNA arrays are three orders of magnitude higher, at one transcript per cell (Holland, 2002; Horak and Snyder, 2002). Robustness of cDNA quantification was demonstrated in a second way: a linear relationship between

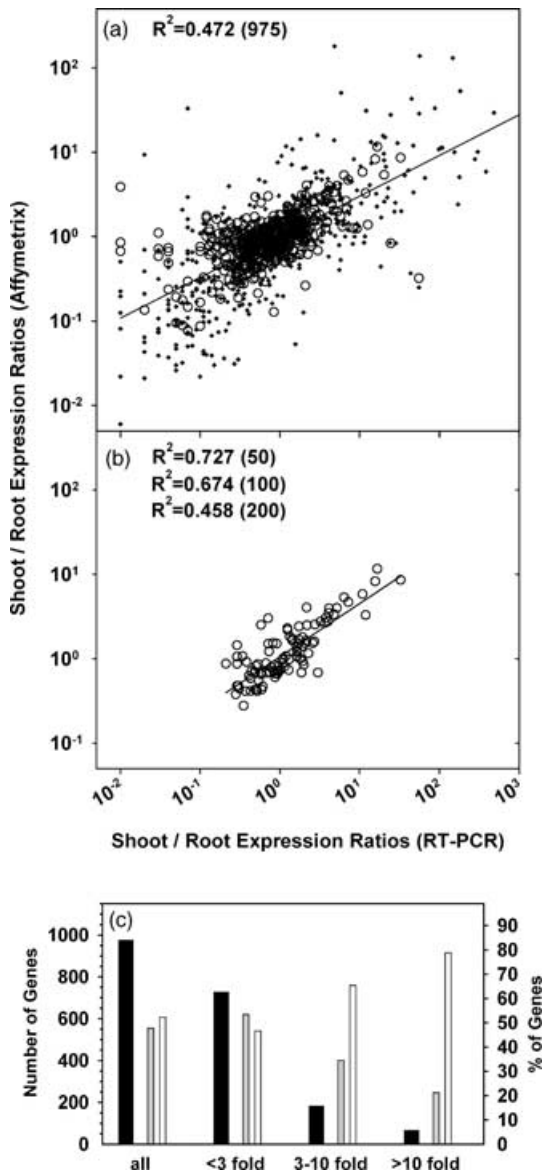


Figure 6. Comparison of S/R expression ratios obtained from real-time RT-PCR and Affymetrix data.

(a) Expression ratios are shown for the subset of 975 genes, which were clearly detectable by real-time RT-PCR (C_T value < 40) and were also present on the Affymetrix ATH1 gene chip. Circles represent genes that were called 'present' in the shoot as well as the root by Affymetrix software. Cross-hairs stand for genes that received 'absent' calls in either shoot or root or in both organs.

(b) A subset of 100 genes (called 'present', see (a)), with highest shoot expression levels, according to Affymetrix technology, is depicted. Correlation coefficients (R^2) of S/R expression ratios for this subset of 100 genes, a subset of 50 and one of 200 genes, obtained by the two technologies are given.

(c) Displays a distribution of the 975 genes, according to the ratio of $(S/R)_{RT-PCR}$ to $(S/R)_{Affymetrix}$. Genes are categorized in four subsets (black bars): all genes, the 727 genes for which the expression ratios obtained with the two technologies varies less than threefold, the 182 genes for which the expression ratios differ from 3- to 10-fold, and the 66 genes with more than 10-fold difference. The percentage of genes called 'present' (grey bar) or 'absent' (white bar) by Affymetrix software is depicted for each subset.

output signal (2^{40-C_T}) and target cDNA amount was maintained over a wide range of mixtures of root and shoot cDNA (Figure 2b). Such robustness has never been shown for DNA arrays, to our knowledge. Precise quantification of transcripts by real-time RT-PCR depends upon having uniformly high amplification efficiency, or having a method to determine the amplification efficiency for each individual PCR. The latter was achieved using the method described by Ramakers *et al.* (2003). This allows the amplification efficiency to be determined for each technical and biological replicate, and the relative transcript abundance to be calculated accordingly. The technical precision of real-time RT-PCR measurements of TF transcript levels was high. Very low intra-assay variation was observed in duplicate measurements of the same pool of cDNA, made in separate runs on the PCR machine (Figure 3a). Interassay variation was estimated by measuring cDNA produced from two separate RT reactions that began from the same sample of RNA. As expected, interassay technical variation was slightly higher than intra-assay variation (Figure 3b). Interassay variability of Affymetrix chips was greater than that of real-time RT-PCR (Figure 3c,d), especially for genes expressed at low levels. The signal to noise ratio for hybridization-based methods of transcript detection is known to decrease exponentially with decreasing amounts of transcript (Holland, 2002; Figure 3c,d). This was not the case for real-time RT-PCR measurements, although variability in duplicate measurements increased slightly as TF transcript levels decreased in our experiments (Figure 3a,b).

Real-time RT-PCR indicated that TF transcript levels in *Arabidopsis* range over five orders of magnitude (for example, see Figure 5). Such a range in TF gene expression levels has never before been reported for plants. Presumably, this great range reflects not only differences in the expression level of different TF genes within any one cell-type, but also differences between cells of different tissues and organs. Given their role(s) as regulators of gene expression, it is to be expected that many TF genes will be expressed in a precise spatial and temporal manner in response to developmental and/or environmental cues. TF genes that orchestrate developmental transitions are known to be amongst the lowest expressed of all genes, and transcripts of these genes are often only detectable by RT-PCR or RNA *in situ* hybridization (Long *et al.*, 1996; Mayer *et al.*, 1998; Putterill *et al.*, 1995; Siegfried *et al.*, 1999). The most-highly expressed TF genes are presumably transcribed constitutively throughout the plant. Some of these may bind non-specifically to DNA. We are aware that not all of the genes that we have targeted are necessarily TF genes. These genes were selected because they encode DNA-binding and other domains that are shared by TF proteins, which does not necessarily mean that they are transcription factor genes. Nonetheless, it is interesting to compare the range of transcript levels that we measured for TFs in *Arabidopsis*

with that measured using the same technique in yeast. Levels of TF transcripts in the single-celled yeast *Saccharomyces cerevisiae* varied over four orders of magnitude (Holland, 2002), which is one order of magnitude less than that observed by us in the more complex, multicellular plant.

It is also interesting to compare the data on TF transcript abundance obtained by real-time RT-PCR with those obtained for the same RNA samples using Affymetrix chips (Figures 3 and 4). The range of values obtained with real-time RT-PCR was two orders of magnitude greater than that obtained with Affymetrix chips (10^5 versus 10^3). As shown above, real-time RT-PCR yields a constant ΔC_T for each X-fold change in initial DNA concentration over the whole range of detectable DNA concentrations (Figure 2a). This is not true for DNA array-based methods, which suffer from an exponential decrease in signal intensity as transcript levels fall, because of second order kinetics of hybridization (Holland, 2002). This could account for the narrower range of values obtained with Affymetrix chips compared to real-time RT-PCR (Figure 4).

Although real-time RT-PCR exhibited greater precision in replicate measurements than Affymetrix chips, this does not necessarily imply greater accuracy. To address the issue of accuracy directly, we used both methods to identify TF genes with extreme shoot to root expression ratios and compared these data with that available in an *Arabidopsis* MPSS database. MPSS represents an alternative means by which to estimate the relative abundance of gene transcripts in a particular organ. Like serial analysis of gene expression (SAGE; Velculescu *et al.*, 1995), MPSS (Brenner *et al.*, 2000a,b) generates short sequence tags produced from a defined position within an mRNA, and the relative abundance of these tags in a given library represents a quantitative estimate of expression of that gene. The *Arabidopsis* MPSS data set contained 3 645 414 tags from a root cDNA library and 2 885 229 tags from shoots. As described above, there was good qualitative agreement between real-time RT-PCR and the MPSS data (Table 1).

We also compared the quantitative accuracy of real-time RT-PCR and Affymetrix chips. A plot of the absolute signals given by the two methods revealed a rather weak correlation in the range corresponding to highly expressed genes and no correlation for genes expressed at lower levels (Figure 4). Unlike quantitative RT-PCR, hybridization-based technologies like Affymetrix chips are qualitative, and there is no strict linear relationship between signal strength and transcript amount for different genes. Thus, it is not possible to conclude with confidence that transcripts of one gene are more abundant than transcripts of another gene, simply based on greater signal strength in the former case on an Affymetrix chip. It is generally assumed that it will not affect the reliability of conclusions drawn from the changes in the Affymetrix signal for a given gene across different

chips, i.e. the Affymetrix chips do provide reliable information about the relative levels of a transcript in different tissues or conditions. To check this, we compared the S/R ratios for all of the TFs that we measured, calculated from real-time RT-PCR data and from Affymetrix arrays (Figure 6a). Indeed, the agreement was good, provided abundantly expressed transcripts were compared (Figure 6b,c). This confirms the accuracy and reliability of both methods. For about half of the TFs, however, the signal obtained by the Affymetrix technology was in a range where accurate results could not be obtained (Figure 6b,c). As already indicated, these discrepancies were most widespread for genes that show a low signal on the arrays.

This problem may not be unique to TFs. In fact, for any given sample, the fraction of *Arabidopsis* genes that are labelled absent by Affymetrix software typically is 30–40%, and 40–45% of the genes typically display normalized expression signals <32 at a target normalization value of 100 (Figure 7). As observed for the TF genes, additional families of genes may contain a considerably higher than normal fraction of low-expressed members, or members with cell-type specific expression patterns. For example, inspection of our Affymetrix data sets indicated that 56% of the approximately 600 annotated receptor kinases (Shiu and Bleecker, 2001) yielded Affymetrix signals <32 (see Figure 7) when probed with cDNA from shoots and roots. As observed for TFs, the receptor kinases are under-represented amongst the highly expressed genes and over-represented amongst the more lowly expressed genes (Figure 7). A similar picture emerged for the large family of cytochrome P450 genes (not shown). Therefore, dedicated analyses of these and other gene families may benefit

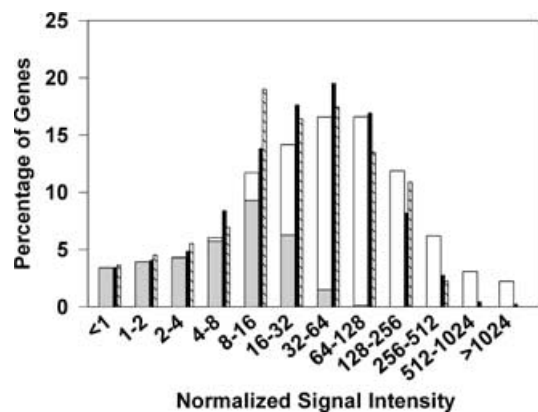


Figure 7. Distribution of normalized signal intensities for an Affymetrix array probed with labelled cRNA derived from shoot material. Normalized fluorescence signal intensities (target normalization value = 100) of the >22 700 sets of gene-specific oligonucleotides were grouped in 12 classes based on twofold difference. The percentage of genes present in each class (white bars) and the percentage of genes that were labelled 'absent' (grey bars overlaid on white bars) were then plotted. Black and cross-hatched bars show the percentage of transcription factor and receptor-kinase genes present in the 12 classes, respectively.

from a real-time RT-PCR approach similar to the one that we have taken for TFs.

TFs control many aspects of plant growth and development by regulating the expression of sets of target genes. Many TF genes are also regulated, in time and space, by internal and/or external cues. Thus, it should be possible to identify TF genes involved in important plant processes through 'Guilt by Association'. To identify TF genes that may play roles in root- or shoot-specific processes, we compared transcript levels of 1214 TF genes in these organs (Figure 5). Approximately 7% (87) of the TF genes repeatedly exhibited greater than 20-fold differences in expression in shoots compared to roots (Table 1). Seventy-three of these were represented in the *Arabidopsis* MPSS data, and as mentioned above, almost all of these were confirmed as essentially root or shoot specific.

There is no published information on the majority of the 87 shoot- or root-specific genes that we identified by real-time RT-PCR (Table 1 and Table 2). Only 14 of the 52 shoot-specific genes have been characterized to some extent in the past. Eight of these were found to be expressed predominantly or exclusively in shoots. These include *AGL2/SEP1* (AT5G15800), *YAB3* (AT4G00180), *YAB1/FIL* (AT2G45190), *ATH1* (AT4G32980), *WUS* (AT2G17950), *SPL3* (AT2G33810), *SPL4* (AT1G53160) and *SPL5* (AT3G15270). Most of these genes have been implicated in plant development. *AGL2/SEP1* is expressed in floral meristems, floral primordia and ovules, and plays a central role in controlling organ identity, such as the development of petals, stamens and carpels (Pelaz *et al.*, 2000). *YAB3* is expressed in all above-ground organs but not in roots, and specifies abaxial tissue development in lateral organs (Siegfried *et al.*, 1999). *YAB1/FIL* is expressed in above-ground vegetative and reproductive meristems and is required for the growth and maintenance of inflorescence and floral meristems (Sawa *et al.*, 1999). *SPL3*, *SPL4* and *SPL5* are expressed in aerial organs, especially in the inflorescence, and control flowering and other aspects of plant development (Cardon *et al.*, 1997, 1999). Other shoot-specific genes from Table 1 that have been described in the literature are: *ATH1*, which is involved in photomorphogenesis (Quaedvlieg *et al.*, 1995); and two genes involved in phytochrome B signalling, *PIF4* (Huq and Quail, 2002) and *PIL6* (Yamashino *et al.*, 2003). Shoot-specific expression of the latter two genes has not been reported previously.

Three genes that we identified as shoot specific encode well-known stress-response regulators: *CBF1/DREB1B* (AT4G25490), *CBF2/DREB1C* (AT4G25470) and *ERF2* (AT5G47220). Expression of the two *CBF* genes, which regulate adaptive responses to cold stress, is induced dramatically by chilling (Medina *et al.*, 1999; Shinwari *et al.*, 1998). However, under non-stress conditions, *CBF1* and *CBF2* transcripts were barely detectable in shoots or roots (Medina *et al.*, 1999; Shinwari *et al.*, 1998). Our results

indicate that the basal or non-induced level of expression of these genes is significantly greater in shoots than in roots, which makes biological sense because the shoot is exposed to more rapid changes in temperature than the root is. *ERF2* is involved in signal reception of ethylene-mediated signalling pathways and also shows modest induction by cold stress (Fujimoto *et al.*, 2000). The *WUS* homeodomain TF gene is expressed in very few cells of the shoot apical meristem during embryogenesis, vegetative growth and flower development, and determines the fate of meristem stem cells (Mayer *et al.*, 1998).

Of the 35 root-specific genes we identified (Table 2), only two have been characterized in the past, namely *AGL21* (AT4G37940) and *AGL17* (AT2G22630). Based on their root-specific expression patterns, roles in root development have been proposed for those two *AGL* genes (Burgeff *et al.*, 2002; Rounsley *et al.*, 1995). Other *AGL* genes have also been characterized as root specific (Alvarez-Buylla *et al.*, 2000; Burgeff *et al.*, 2002; Rounsley *et al.*, 1995), including *AGL14* (AT4G11880) and *AGL19* (AT4G22950). We found transcript levels of both to be approximately 10 times higher in roots than that in shoots (Supplementary Material, Table S1).

Many of the reported genes that we identified as shoot specific appear to be involved in developmental processes. This may simply reflect the way in which most TF genes have been isolated to date, namely via genetic screens for aberrant growth and development. Defects in TF genes involved other plant processes, such as metabolism, may produce more subtle phenotypes, which are difficult to identify. Thus, many of the novel root- and shoot-specific genes that we have identified may eventually be implicated in processes other than development. Obviously, reverse genetics will play a central role in identifying functions for these genes.

Recently, an expression profile matrix for 400 *Arabidopsis* TF genes, derived from a series of Affymetrix chip experiments, was used to identify TF genes that may play roles in responses to different environmental stresses (Chen *et al.*, 2002). Transcripts of about 10% of the genes were not detected under any of the conditions used in that study. Importantly, we detected expression of several of these genes in roots and/or shoots, using real-time RT-PCR (AT4G13480; AT1G73410; AT4G01500 and AT3G12820), which highlights the greater sensitivity of this technique. An interesting anomaly discussed in the paper by Chen *et al.* (2002) was the expression pattern of the *TINY* gene (AT5G25810), which was found by Affymetrix chip analysis to be expressed at high levels in roots but not at all in other organs. *TINY* is required for both vegetative and floral organogenesis (Wilson *et al.*, 1996), which indicates that it is expressed in aerial parts of the plant. We were able to detect transcripts of this gene in both roots and shoots (sevenfold higher in roots than in shoots) using RT-PCR.

To summarize, we have created a resource for real-time RT-PCR profiling of almost 1500 *Arabidopsis* TF genes that, compared to existing technologies such as Affymetrix chips, increases significantly the sensitivity, precision and accuracy with which transcripts of these genes can be measured. This resource is also more flexible than other systems: we can add, remove or replace primer pairs at any time. For instance, we will re-design and replace primer pairs for those PCRs that yielded efficiencies lower than 0.5. On the other hand, we are also aware of a significant number of additional *Arabidopsis* genes that have been attributed a potential role as transcriptional regulators (<http://www.arabidopsis.org>; <http://arabidopsis.med.ohio-state.edu>; Davuluri *et al.*, 2003; <http://genetics.mgh.harvard.edu/sheenweb/AraTRs.html>), and we plan to add primers to these genes to the existing set of TF primers in the near future.

We used real-time RT-PCR in this study to identify a considerable number of novel root- and shoot-specific TF genes, which may play important roles in development or other organ-specific processes. This information will be a valuable starting point for further research on these genes. In addition, we provide the first experimental evidence that the vast majority of the putative TF genes annotated in *Arabidopsis* are indeed expressed. This new resource will help to identify TF genes involved in numerous plant processes, including abiotic stress responses, an area that we are particularly interested in.

Experimental procedures

Plant material and growth conditions

Arabidopsis (Col-0) wild-type plants were grown vertically on half-strength Murashige and Skoog medium (Murashige and Skoog, 1962), supplemented with 0.5% (w/v) sucrose and solidified with 0.7% agar, at 22°C under a 16-h day (140 $\mu\text{mol m}^{-2} \text{sec}^{-1}$) and 8-h night regime. Shoots and roots were harvested separately 14 days after germination, and frozen in liquid nitrogen before storage at -80°C.

RNA isolation and cDNA synthesis

Total RNA was isolated from shoots or roots using TRIZOL reagent (Invitrogen GmbH, Karlsruhe, Germany), as described (http://www.arabidopsis.org/info/2010_projects/comp_proj/AFGC/RevisedAFGC/site2RnaL.htm#isolation). RNA concentration was measured in an Eppendorf Biophotometer, and 150 μg of total RNA was digested with RNase-free DNaseI (product number D5307, Sigma-Aldrich, Taufkirchen, Germany), according to the manufacturer's instructions. Absence of genomic DNA contamination was subsequently confirmed by PCR, using primers designed on intron sequence of a control gene (At5g65080). RNA integrity was checked on a 1.5% (w/v) agarose gel prior to, and after DNaseI digestion. Poly(A)+ RNA was purified with an Oligotex mRNA Mini Kit (Qiagen GmbH, Hilden, Germany), using the supplier's batch protocol. RT reactions were performed with SuperScriptTM III reverse transcriptase (Invitrogen GmbH), according to the manufacturer's instruc-

tions. The efficiency of cDNA synthesis was assessed by real-time PCR amplification of control genes encoding *actin2* (primers: AT3G18780F, 5'-TCCCTCAGCACATTCCAGCAGAT-3'; AT3G18780R, 5'-AACGATTCCTGGACCTGCCTCATC-3'), *UBQ10* (AT4G05320F, 5'-CACACTCCACTTGGTCTTGCGT-3'; AT4G05320R, 5'-TGGTCTTCCGGTGAGAGTCTTCA-3'), β -6-*tubulin* (AT5G12250F, 5'-ACCACTCTAGCTTTGGTGATCTG-3'; AT5G12250R, 5'-AGGTTCACTGCGAGCTTCCTCA-3'), *elongation factor 1 α* (AT5G60390F, 5'-TGAGCACGCTCTTCTTGCTTCA-3'; AT5G60390R, 5'-GGTGGTGCCATCTTGTTACA-3'), and *adenosyl-phosphoribosyltransferase* (AT1G27450F, 5'-GTTGCAGGTGTTGAAGCTAGAGGT-3'; AT1G27450R, 5'-TGGCCAATAGCCAACGCAATAG-3'). Only cDNA preparations that yielded similar C_T values (e.g. 20 ± 1) for the control genes were used for comparing TF transcript levels.

PCR primer design

Putative TF genes were identified in the *Arabidopsis* genome by taking advantage of gene annotations and INTERPRO domain number searches (Riechmann *et al.*, 2000) at the MIPS database (<http://mips.gsf.de/cgi-bin/proj/thal/>). The resulting set of sequences was supplemented by performing BLASTP and TBLASTN searches (<http://www.ncbi.nlm.nih.gov/blast/>), to uncover further possible TF genes in the *Arabidopsis* genome. The set of 1465 putative TF genes that we compiled is listed in Supplementary Material (Table S1).

To facilitate RT-PCR measurement of transcripts of all putative TF genes under a standard set of reaction conditions, oligonucleotide primers were required to meet a stringent set of criteria as outlined in the beginning of the section under Results. Primers were designed according to these criteria by Dr Jacqueline Weber-Lehmann at MWG Biotech AG (Ebersberg, Germany) using the PRIME program of GCG[®] Wisconsin PackageTM, version 10.2 (Madison, WI, USA). Global alignments of the suggested primer sequences with genomic and transcript sequences of all *Arabidopsis* genes were performed using the Smith-Waterman nucleotide (SWN) search algorithm in BIOVIEW TOOLKIT (BTK) Software, version 5.0 (Paracel, Pasadena CA, USA). Assessment and choice of primer pairs was realized with PERL scripts specifically designed for our purposes at MWG Biotech AG. The sequences of each primer pair are given in Supplementary Material (Table S1).

Real-time PCR conditions and analysis

Polymerase chain reactions were performed in an optical 384-well plate with an ABI PRISM[®] 7900 HT Sequence Detection System (Applied Biosystems, Foster City, CA, USA), using SYBR[®] Green to monitor dsDNA synthesis. Reactions contained 5 μl 2 \times SYBR[®] Green Master Mix reagent (Applied Biosystems), 1.0 ng cDNA and 200 nM of each gene-specific primer in a final volume of 10 μl . A master mix of sufficient cDNA and 2 \times SYBR[®] Green reagent was prepared prior to dispensing into individual wells, reduce pipetting errors and ensure that each reaction contained an equal amount of cDNA. An electronic Eppendorf multipipette was used to pipette the cDNA-containing master mix, while primers were aliquoted with an Eppendorf 12-channel pipette. The following standard thermal profile was used for all PCRs: 50°C for 2 min; 95°C for 10 min; 40 cycles of 95°C for 15 sec and 60°C for 1 min. Data were analysed using the SDS 2.0 software (Applied Biosystems). To generate a baseline-subtracted plot of the logarithmic increase in fluorescence signal (ΔR_n) versus cycle number, baseline data were collected between cycles 3 and 15. All amplification plots were analysed with an R_n threshold of 0.3 to obtain C_T (threshold cycle)

values. In order to compare data from different PCR runs or cDNA samples, C_T values for all TF genes were normalized to the C_T value of *UBQ10*, which was the most constant of the five house-keeping genes included in each PCR run. The average C_T value for *UBQ10* was 20.04 (± 0.89) for all plates/templates measured in this series of experiments. PCR efficiency (E) was estimated in two ways. The first method of calculating efficiency utilized template dilutions and the equation $(1 + E) = 10^{(-1/\text{slope})}$, as described previously by Pfaffl (2001). The second method made use of data obtained from the exponential phase of each individual amplification plot and the equation $(1 + E) = 10^{\text{slope}}$ (Ramakers et al., 2003). TF gene expression was normalized to that of *UBQ10* by subtracting the C_T value of *UBQ10* from the C_T value of the TF gene of interest. S/R expression ratios were then obtained from the equation $(1 + E)^{\Delta\Delta C_T}$, where $\Delta\Delta C_T$ represents ΔC_{TS} minus ΔC_{TR} , and E is the PCR efficiency. RT-PCR products were resolved on 4% (w/v) agarose gels (3 : 1 HR agarose, Amresco, Solon, OH, USA) run at 4 V cm^{-1} in TBE Tris-Borate-EDTA buffer, along with a 50-bp DNA-standard ladder (Invitrogen GmbH).

Mixing and dilution experiments

Mixtures of root and shoot cDNA were made to give the following amounts (ng) of root cDNA in a total of 1 ng cDNA: 1.0, 0.95, 0.90, 0.80, 0.75, 0.50, 0.25, 0.20, 0.10, 0.05 and 0. Real-time PCR using 1 ng cDNA was performed as described above, with primers for two shoot-specific genes (AT1G13300 (circle); AT1G34670 (diamond)) and two root-specific genes (AT4G32980 (triangle); AT5G44190 (square)).

Plasmid pZPXOmegaL+ (kindly provided by Dr Steve Kay, TSRI, La Jolla, CA, USA) containing the *LUC* gene or a 75-bp intergenic DNA fragment (genetic marker ATC4H; <http://www.arabidopsis.org>), amplified from *Arabidopsis* Columbia-0) genomic DNA, were serially diluted to yield solutions containing from 1 million copies μl^{-1} to 1 copy μl^{-1} . One microlitre of each plasmid or DNA fragment dilution was mixed with 1 ng of cDNA from shoots or roots, *ATC4H* primers or *LUC*-specific primers (*LUC-F*, 5'-ATTGTTCCAG-GAACCAGGGC-3'; *LUC-R*, 5'-GAACCGCTGGAGAGCAACTG-3') and subjected to real-time PCR analysis as described above, with the exception that 50 instead of 40 PCR cycles were performed and recorded.

Hybridization of Affymetrix genome arrays

Twenty micrograms of total RNA, isolated as above, were used for double-stranded cDNA synthesis (SuperScript Choice system, Invitrogen GmbH). Biotin-labelled cRNA was synthesized using the BioArray High Yield RNA Transcript Labelling Kit (Enzo Life Sciences Inc., Farmingdale, NY, USA). All cRNA samples were checked for degradation by gel analysis according to the Affymetrix gene chip expression analysis technical manual (<http://www.affymetrix.com/support/index.affx>). Both samples were checked by hybridization of Test 3 arrays (part number 900341; Affymetrix, Santa Clara, CA, USA). Only satisfactory probes were hybridized with the Affymetrix *Arabidopsis* Full Genome Array (ATH1; part number 900386; Affymetrix). Hybridization, washing, staining and scanning procedures were performed as described in the Affymetrix technical manual. Expression analysis was performed using Affymetrix MICROARRAY SUITE software (version 5.0) and each array was globally normalized to a target value of 100. Basic principles of Affymetrix oligonucleotide arrays were reviewed by Lipshutz et al. (1999) and Lockhart et al. (1996).

Acknowledgements

We acknowledge Jacqueline Weber-Lehmann at MWG Biotech AG (Ebersberg, Germany) for expert design of oligonucleotides, and Ulrike Simon-Rosin (MPI-MPP, Golm, Germany) for experimental assistance during the initial stages of the study, Matthias Scholz and Peter Krüger for bioinformatics assistance, and the MPI-MPP (Golm) and the Max-Planck Society for funding of the real-time PCR system. We also thank our colleagues Ute Krämer and Dirk Hincha for useful discussion, and the anonymous reviewers for their thoughtful comments, which improved the paper significantly. We would be grateful to colleagues with expert knowledge of new TF genes/families if they share this knowledge with us to make the TF real-time PCR resource as comprehensive as possible.

Supplementary Material

The following material is available from <http://www.blackwellpublishing.com/products/journals/suppmat/TPJ/TPJ2051/TPJ2051sm.htm>

Figure S1. Comparison of expression levels and S/R expression ratios obtained from biological replicates using real-time RT-PCR and Affymetrix technology.

Table S1 Complete list of TF genes, primer sequences and experimental data

Table S2 Comparison of RT-PCR efficiencies determined from calibration dilution curves and from increase in absolute fluorescence

References

- AGI (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.
- Alvarez-Buylla, E.R., Liljegren, S.J., Pelaz, S., Gold, S.E., Burgeff, C., Ditta, G.S., Vergara-Silva, F. and Yanofsky, M.F. (2000) MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. *Plant J.* **24**, 457–466.
- Brenner, S., Williams, S.R., Vermaas, E.H. et al. (2000a) *In vitro* cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. *Proc. Natl. Acad. Sci. USA*, **97**, 1665–1670.
- Brenner, S., Johnson, M., Bridgham, J. et al. (2000b) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* **18**, 630–634.
- Burgeff, C., Liljegren, S.J., Tapia-Lopez, R., Yanofsky, M.F. and Alvarez-Buylla, E.R. (2002) MADS-box gene expression in lateral primordia, meristems and differentiated tissues of *Arabidopsis thaliana* roots. *Planta*, **214**, 365–372.
- Cardon, G.H., Hohmann, S., Nettesheim, K., Saedler, H. and Huijser, P. (1997) Functional analysis of the *Arabidopsis thaliana* SBP-box gene *SPL3*: a novel gene involved in the floral transition. *Plant J.* **12**, 367–377.
- Cardon, G., Hohmann, S., Klein, J., Nettesheim, K., Saedler, H. and Huijser, P. (1999) Molecular characterisation of the *Arabidopsis* SBP-box genes. *Gene*, **237**, 91–104.
- Chen, W., Provart, N.J., Glazebrook, J. et al. (2002) Expression profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses. *Plant Cell*, **14**, 559–574.
- Davuluri, R.V., Sun, H., Palaniswamy, S.K., Matthews, N., Molina, C., Kurtz, M. and Grotewold, E. (2003) AGRIS: *Arabidopsis* Gene Regulatory Information Server, an information resource of

- Arabidopsis* cis-regulatory elements and transcription factors. *BMC Bioinformatics*, **4**, 25.
- Fujimoto, S.Y., Ohta, M., Usui, A., Shinshi, H. and Ohme-Takagi, M. (2000) *Arabidopsis* ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. *Plant Cell*, **12**, 393–404.
- Haas, B.J., Volfovsky, N., Town, C.D., Troukhan, M., Alexandrov, N., Feldmann, K.A., Flavell, R.B., White, O. and Salzberg, S.L. (2002) Full-length messenger RNA sequences greatly improve genome annotation. *Genome Biol.* **3**, 0029.1–0029.12 (Research).
- Hardtke, C.S. and Berleth, T. (1998) The *Arabidopsis* gene *MONOPTEROS* encodes a transcription factor mediating embryo axis formation and vascular development. *EMBO J.* **17**, 1405–1411.
- Holland, M.J. (2002) Transcript abundance in yeast varies over six orders of magnitude. *J. Biol. Chem.* **277**, 14363–14366.
- Horak, C.E. and Snyder, M. (2002) Global analysis of gene expression in yeast. *Funct. Integr. Genomics*, **2**, 171–180.
- Huq, E. and Quail, P. (2002) PIF4, a phytochrome-interacting bHLH factor, functions as a negative regulator of phytochrome B signaling in *Arabidopsis*. *EMBO J.* **15**, 2441–2450.
- Kamalay, J.C. and Goldberg, R.B. (1980) Regulation of structural gene expression in tobacco. *Cell*, **19**, 935–946.
- Kiper, M. (1979) Gene numbers as measured by single-copy DNA saturation with mRNA are routinely overestimates. *Nature*, **278**, 279–280.
- Liljegren, S.J., Ditta, G.S., Eshed, Y., Savidge, B., Bowman, J.L. and Yanofsky, M.F. (2000) SHATTERPROOF *MADS*-box genes control seed dispersal in *Arabidopsis*. *Nature*, **404**, 766–770.
- Lipshutz, R.J., Fodor, S.P., Gingeras, T.R. and Lockhart, D.J. (1999) High density synthetic oligonucleotide arrays. *Nat. Genet.* **21**, 20–24.
- Lockhart, D.J., Dong, H., Byrne, M.C. *et al.* (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14**, 1675–1680.
- Long, J.A., Moan, E.I., Medford, J.I. and Barton, M.K. (1996) A member of the KNOTTED class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*. *Nature*, **379**, 66–69.
- Mayer, K.F., Schoof, H., Haecker, A., Lenhard, M., Jurgens, G. and Laux, T. (1998) Role of *WUSCHEL* in regulating stem cell fate in the *Arabidopsis* shoot meristem. *Cell*, **95**, 805–815.
- Medina, J., Bargas, M., Terol, J., Perez-Alonso, M. and Salinas, J. (1999) The *Arabidopsis* CBF gene family is composed of three genes encoding AP2 domain-containing proteins whose expression is regulated by low temperature but not by abscisic acid or dehydration. *Plant Physiol.* **119**, 463–470.
- Murashige, T. and F. Skoog (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant.* **15**, 473–479.
- Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E. and Yanofsky, M.F. (2000) B and C floral organ identity functions require *SEPAL-LATA* *MADS*-box genes. *Nature*, **405**, 200–203.
- Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucl. Acids Res.* **29**, e45.
- Putterill, J., Robson, F., Lee, K., Simon, R. and Coupland, G. (1995) The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell*, **80**, 847–857.
- Quaedvlieg, N., Dockx, J., Rook, F., Weisbeek, P. and Smeekens, S. (1995) The homeobox gene *ATH1* of *Arabidopsis* is derepressed in the photomorphogenic mutants *cop1* and *det1*. *Plant Cell*, **7**, 117–129.
- Ramakers, C., Ruijter, J.M., Deprez, R.H. and Moorman, A.F. (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci. Lett.* **13**, 62–66.
- Ratcliffe, O.J. and Riechmann, J.L. (2002) *Arabidopsis* transcription factors and the regulation of flowering time: a genomic perspective. *Curr. Issues Mol. Biol.* **4**, 77–91.
- Riechmann, J.L. (2002) Transcriptional regulation: a genomic overview (September, 30, 2002). In *The Arabidopsis Book* (Somerville, C.R., and Meyerowitz E.M., eds), Rockville, MD: American Society of Plant Biologists (doi/10.1199/tab.0085, <http://www.aspb.org/publications/Arabidopsis/>).
- Riechmann, J.L., Heard, J., Martin, G. *et al.* (2000) *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science*, **290**, 2105–2110.
- Rounsley, S.D., Ditta, G.S. and Yanofsky, M.F. (1995) Diverse roles for *MADS* box genes in *Arabidopsis* development. *Plant Cell*, **7**, 1259–1269.
- Ruan, Y., Gilmore, J. and Conner, T. (1998) Towards *Arabidopsis* genome analysis: monitoring expression profiles of 1400 genes using cDNA microarrays. *Plant J.* **15**, 821–833.
- Sawa, S., Ito, T., Shimura, Y. and Okada, K. (1999) *FILAMENTOUS FLOWER* controls the formation and development of *Arabidopsis* inflorescences and floral meristems. *Plant Cell*, **11**, 69–86.
- Shinwari, Z.K., Nakashima, K., Miura, S., Kasuga, M., Seki, M., Yamaguchi-Shinozaki, K. and Shinozaki, K. (1998) An *Arabidopsis* gene family encoding *DRE/CRT* binding proteins involved in low-temperature-responsive gene expression. *Biochem. Biophys. Res. Commun.* **250**, 161–170.
- Shiu, S.H. and Bleeker, A.B. (2001) Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc. Natl. Acad. Sci. USA*, **98**, 10763–10768.
- Siegfried, K.R., Eshed, Y., Baum, S.F., Otsuga, D., Drews, G.N. and Bowman, J.L. (1999) Members of the *YABBY* gene family specify abaxial cell fate in *Arabidopsis*. *Development*, **126**, 4117–4128.
- Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. (1995) Serial analysis of gene expression. *Science*, **270**, 484–487.
- Weigel, D., Alvarez, J., Smyth, D.R., Yanofsky, M.F. and Meyerowitz, E.M. (1992) *LEAFY* controls floral meristem identity in *Arabidopsis*. *Cell*, **69**, 843–859.
- Wilson, K., Long, D., Swinburne, J. and Coupland, G. (1996) A dissociation insertion causes a semidominant mutation that increases expression of *TINY*, an *Arabidopsis* gene related to *APETALA2*. *Plant Cell*, **8**, 659–671.
- Yamashino, T., Matsushika, A., Fujimori, T., Sato, S., Kato, T., Tabata, S. and Mizuno, T. (2003) A link between circadian-controlled bHLH factors and the *APRR1/TOC1* quintet in *Arabidopsis thaliana*. *Plant Cell Physiol.* **44**, 619–629.