Genome-wide profiling of stored mRNA in *Arabidopsis thaliana* seed germination: epigenetic and genetic regulation of transcription in seed

Kazumi Nakabayashi¹, Masanori Okamoto^{1,2}, Tomokazu Koshiba², Yuji Kamiya¹ and Eiji Nambara^{1,*}

¹Plant Science Center, RIKEN, Suehiro-cho 1-7-22, Tsurumi, Yokohama, Kanagawa 230-0045, Japan, and ²Department of Biological Sciences, Tokyo Metropolitan University, Minami-Ohsawa 1-1, Hachioji, Tokyo 192-0397, Japan

Received 29 November 2004; accepted 6 December 2004. *For correspondence (fax +81 45 503 9665; e-mail nambara@postman.riken.go.jp).

Summary

To reveal the transcriptomes of Arabidopsis seed, comprehensive expression analysis was performed using ATH1 GeneChips (Affymetrix, Santa Clara, CA, USA). In the dry seed, more than 12 000 stored mRNA species were detected, including all ontological categories. Statistical analysis revealed that promoters of highly expressed genes in wild-type dry seeds overrepresented abscisic acid-responsive elements (ABREs) containing the core motif ACGT. Although the coupling element and seed-specific enhancer RY motif alone were not prominently overrepresented in genes with high expression, the presence of these elements in combination with ABRE was associated with particularly high gene expression. The transcriptome of the imbibed seeds differed from that of the dry seed even at 6 h after seed imbibition. After imbibition many upregulated and downregulated genes were co-regulated in clusters of three to five genes. Genes for which expression was affected by the abi5 mutation tended to be located in clusters, suggesting that transactivation by ABI5 is not restricted to a single gene, but affects other proximal genes. Furthermore, cytosine methylation was observed not only in large silent retrotransposon clusters in centromeric regions, but also in non-centromeric silent gene clusters in the seed. These results suggest that such regions might be transcriptionally silenced by methylation or heterochromatin structures. Our analyses reveal that transcriptomes of Arabidopsis seed are characterized by multiple regulatory mechanisms: epigenetic chromatin structures, chromosomal locations (e.g. co-regulated gene clusters) and cis-acting elements.

Keywords: seed, stored RNA, ABRE, ABI, co-expression, methylation.

Introduction

In the course of embryogenesis and seed maturation, the embryo enters a quiescent phase (Bewley, 1997). The seed remains quiescent, even if watered, until conditions such as temperature, light and nutrients become permissive for germination. Dry mature seed contains a large number of mRNA species. Stored RNA in mature dry seed was first found in cotton (Dure and Waters, 1965) and is universal in plant species (Almoguera and Jordano, 1992; Ishibashi *et al.*, 1990; Kuligowski *et al.*, 1991). The function of stored RNA remains unknown, however it is thought to play a role in protein synthesis during the early stages of germination.

Developmental phases during late embryogenesis and subsequent germination are characterized by spatial and temporal patterns of gene expression (Nambara *et al.*, 2000; Parcy *et al.*, 1997). Expression analyses have shown that the transition from late embryogenesis to germination occurs gradually and therefore characteristics of both developmental phases co-exist during the transition (Harada *et al.*, 1988; Nambara *et al.*, 2000). In accordance with this, sequence analysis suggests that the stored RNA comprises transcripts necessary for both late embryogenesis and seed germination (Comai *et al.*, 1989; Hughes and Galau, 1989, 1991). Subsets of the stored mRNA in dry seed are well described in several plant species, however a genome-wide analysis of stored RNA has not been reported, and the regulation and function of the dry seed transcriptome is poorly understood.

Abscisic acid (ABA) plays a primary regulatory role in seed maturation and dormancy. ABA promotes expression of a

number of genes such as Em and Late embryogenesis abundant (Lea) during seed development and maturation (Busk and Pagès, 1998; Cuming, 1999). Functional analysis of ABA-regulated gene promoters has identified the motifs mediating ABA response. The most widely known ABAresponsive element (ABRE) is a sequence resembling a G-Box with an ACGT core (Marcotte et al., 1989; Shen et al., 1993). The ABRE requires a second element to form a functional ABA response complex (ABRC) (Shen and Ho, 1995). The coupling element (CE) has been identified as one such motif, and it elicits a response to ABA when located adjacent to an ABRE (Shen and Ho, 1995). The RY/Sph motif is involved in seed-specific gene expression (Dickinson et al., 1988; Hattori et al., 1992). Similar to the CE, the RY element also interacts synergistically with ABRE (Ezcurra et al., 1999; Vasil et al., 1995). Several other cis-acting motifs have been also reported (reviewed in Thomas, 1993). These analyses mostly infer a complicated regulation system that involves multiple cis-acting motifs controlling ABA-mediated transcription during seed development and maturation.

Genetic studies have identified components involved in ABA biosynthesis and signaling in the seed. These mutants were defective in accumulation of seed storage reserves and seed dormancy (Finkelstein and Somerville, 1990; Karssen et al., 1983; Koornneef et al., 1984; Nambara et al., 1992). Three ABA-insensitive loci, ABI3, ABI4 and ABI5, have been identified, encoding B3, AP2 and bZIP-type transcription factors, respectively (Finkelstein and Lynch, 2000; Finkelstein et al., 1998; Giraudat et al., 1992; Lopez-Molina and Chua, 2000). Biochemical approaches revealed their target sequences as the RY/Sph motif for ABI3, CE1-like sequences for ABI4, and ACGT-containing ABRE for ABI5 (Carles et al., 2002; Kim et al., 2002; Mönke et al., 2004; Niu et al., 2002). These transcriptional regulators are conserved in both monocots and dicots (Hobo et al., 1999a; Niu et al., 2002; Suzuki et al., 1997). Physical interaction between Arabidopsis ABI3 and ABI5, or their rice orthologs, VP1 and TRAB1, has been demonstrated (Hobo et al., 1999a; Nakamura et al., 2001). This suggests that in response to ABA, ABI3 and ABI5 act in concert, via RY motifs and ABREs, to regulate gene expression in the maturating seed. While the genes expressed during seed development and germination have been extensively analyzed in many species (Dong et al., 2004; Girke et al., 2000; reviewed in McCarty, 1995; Ogawa et al., 2003; White et al., 2000), the in vivo effect of ABI transcription factors on global expression profiles has not yet been fully analyzed.

To explore genome-wide expression patterns as well as regulatory mechanisms in Arabidopsis seeds, we conducted mRNA profiling in dry and imbibed seeds using an oligonucleotide-based microarray. By comparing expression profiles between wild type and *abi* mutants, we found that the ABRE predominantly defines the transcriptome in the dry seed. We also observed a consistent trend of co-regulation of neighboring genes. Cytosine-methylation was detected in centromeric regions and in silent gene clusters. Multiple regulations by epigenetic chromatin structures, chromosomal locations (e.g. co-regulated gene clusters) and *cis*-acting elements within Arabidopsis seed transcriptomes are discussed.

Results

Gene expression profiles of dry and germinating seeds

Stored mRNA in mature dry seeds might reflect the gene expression pattern prior to seed desiccation. To reveal the gene expression profiles in wild-type dry and imbibed seeds in Arabidopsis, we performed microarray analysis using the Affymetrix GeneChip ATH1. Among 22 476 genes represented on the microarray, 12 470 mRNA species were quantitatively detected in dry seeds. Duplicate microarray analysis using the different seed batch gave consistent results. A similar number of transcripts were detected in 24 h-imbibed seeds (14395 genes) (see Supplementary Material S1).

Stored mRNA species in dry seeds are thought to be required for late embryogenesis and seed germination. In order to differentiate these two classes of genes, we analyzed expression profiles obtained from seeds at 6, 12 and 24 h after imbibition. Of the 12 470 stored mRNA species, more than 10 000 mRNA species were also quantitatively detected during the course of seed imbibition. K-means clustering analysis classified the genes detected in dry seeds into seven kinetic patterns (Figure 1a). This temporal profiling of genes represented as stored mRNA species reflected the overlapping nature of both late embryogenic and germinative states. The largest gene set was group 3, which showed relatively constant expression levels for 24 h after imbibition. Other sets of genes were dramatically or moderately downregulated (groups 1 and 2, respectively), and more than a 1000 genes were upregulated gradually or dramatically after 12 h (groups 6 and 7, respectively). Genes in groups 4 and 5 showed transient increases in transcript level at 6 and 12 h, respectively. Expression patterns of several genes from each group were validated by semiquantitative RT-PCR and representative results are shown in Figure 1(a).

Figure 1(b) shows the ontological classification of the genes in each group. A number of highly expressed genes for late embryogenesis are included in group 1. Genes related to metabolism were overrepresented in groups 1, 6 and 7. This indicated that embryogenic genes were largely downregulated within 6 h of imbibition, and that induction of metabolic genes involved in germination started around 12 h after imbibition. During this shift, expression of genes involved in cell cycle, DNA processing, transcription and protein synthesis were activated, and these categories were significantly overrepresented in groups 5 and 6.



Figure 1. Classification of mRNA species detected in dry seeds.

(a) Kinetic patterns of accumulation of mRNA species detected in dry seeds during imbibition by K-means clustering. The number of genes categorized in each group is indicated in parenthesis. The color of individual genes in graphs reflects expression level in dry seed. Semiquantitative RT-PCR for representative genes in each group is shown in insets. The sample points are dry seed, 6, 12 and 24 h-imbibed seed from left to right. Selected genes are: group 1, At3g51810 (top) and At5g56100 (bottom); group 2, At4g27150 and At4g19960; group 3, At5g03280 and At5g47210; group 6, At1g05190 and At1g15500; and group 7, At3g16430 and At1g47128. The result of 18S rRNA is shown as control.

(b) Ontological classification of mRNA species in each group in Figure 1(a). The nomenclature of 13 groups and the ratio of those 13 groups in all 22 476 genes are indicated at the bottom. Genes related to protein synthesis were not found in group 7.

Upstream regions of highly expressed genes in wild-type dry seeds enriched ABREs

Approximately 500 of the most highly expressed genes in wild-type dry seeds were searched for any common motifs located within a 1 kb region upstream of the translational start site. A statistical search identified multiple sequences containing an ACGT core, and the most prominent were CACGTG-related sequences (Table 1 and Tables S1 and S2). These were consistent with one of the most typical ABREs

Molecular profiling in Arabidopsis seed 699

previously identified by promoter and binding assays (Guiltinan *et al.*, 1990; Shen *et al.*, 1993; Skriver *et al.*, 1991). The frequency of CGTGTC-related sequences was also significantly high, with a false-positive probability of 1.744e-6 (Table 1 and Tables S1 and S2). These sequences share a common motif with CE3, and function in concert with ACGT-containing ABREs to regulate ABA-inducible gene expression (Hobo *et al.*, 1999b; Shen *et al.*, 1996). In addition, the frequency of TATCCA-related sequences was also high with a false-positive probability of 3.4e-4 (Table 1 and Tables S1 and S2). These motifs are reportedly involved in sugar repression (Lu *et al.*, 1998) and also show a resemblance to a light-responsive element (Argüello-Astorga and Herrera-Estrella, 1996).

ACGT-containing ABREs are less operative in the abi5 seed

ABI5 binds to ACGT-containing ABREs. We performed expression analysis on the *abi5* seed and compared with wild type. The *abi5* dry seed had 527 nuclear genes with at least twofold less expression than in wild type. A statistical search of promoter regions of these genes revealed a significant excess of CACGTG-containing sequences (Table 1 and Tables S3 and S4). This clearly showed that ABRE-mediated gene expression was largely debilitated in *abi5* seeds.

Intensity and kinetics of ABRE-mediated gene expression during germination

We further analyzed the effect of the selected CACGTG motif on gene expression in the seed. In our analysis, absolute values of expression level of each gene ranged from 0 to approximately 35 000. Approximately 95% of the genes had expression levels of less than 5000 in the dry seed. Of those lower than 5000 in the wild-type dry seed, only 14% had at least one CACGTG sequence within 1 kb upstream region of each gene. The percentage increased in the gene sets which showed higher expression levels. More than half the genes with an expression level higher than 30 000 contained an ABRE (Figure 2a, dry seed). These values are significantly higher than the overall percentage (15.5%) of genes containing this ABRE. The positive correlation between high level expression and the presence of ABRE diminished in the 24 h-imbibed seed (Figure 2a, imbibed seed).

We performed time-course analysis on the frequency of ABRE-containing genes among highly expressed genes during wild-type seed imbibition. The frequency of ABRE-containing genes was positively correlated with expression level in the dry seed, as described earlier (Figure 2b), but this correlation diminished by 6 h after imbibition (Figure 2b). Consistent with this, the endogenous ABA level in the seed decreased dramatically within 6 h after imbibition (Figure 2c).

700 Kazumi Nakabayashi et al.

Table 1	Sequences	enriched in the	1-kb upstream	region of	selected	gene sets
						•

Selected gene sets	Sequence	Observed (%)	Observed in other genes (%)	Random rate (%)	<i>P</i> -value
Expression level above 10 000 in wild-type dry seed	MCACGTGK	136/473 (28.8)	7.522	5.823	3.132e-38
	ACGTGKCA	78/473 (16.5)	4.16	2.955	3.857e-20
	GMCACGT	126/473 (26.6)	7.897	11.319	5.786e-16
	CGTGTCA	67/473 (14.2)	3.588	5.829	1.744e-6
	TATCCA	172/473 (36.4)	24.91	21.377	3.40e-4
Downregulated in <i>abi5</i>	ACACGTGK	80/527 (15.2)	4.485	2.955	2.098e-16
-	ACGTGTC	62/527 (11.8)	4.972	5.829	1.11e-2

Sequences enriched in the 1-kb upstream region of 473 nuclear genes that showed expression level above 10 000 in wild-type dry seed and of 527 downregulated genes in *abi5* dry seed are shown.

Combined effects of ABRE and other motifs

We next tested whether the transcriptome of dry seed reflects the characters of known *cis*-elements. Previous promoter analyses have demonstrated that a single ABRE provides relatively weak transcriptional activation, while multiple copies of ABRE confer a stronger ABA responsiveness (Shen and Ho, 1995). To examine whether seed transcriptome contains this character, we surveyed the effect of copy number of the selected ABRE on expression level in the dry seed. The percentage of genes with a single, double or triple or more repeat(s) of the ABRE among genes with various expression levels was compared with the respective overall rates (see Experimental procedures, Table S5). Genes with high expression were more likely to contain multiple ABREs (Figure 3a, Col). The degree of enrichment of ABRE in the abi5 seed was generally lower than in wild type, however Figure 3(a) clearly shows that the multiple copies of ABREs were functional even in abi5.

The coupling element is not regarded as a mediator of ABA signals, but is rather an enhancer of ABA-inducible expression via ABRE in the seed (Shen et al., 1996). The combined effects of ABRE and CE motifs on high-level expression were also examined in the same manner. We selected the 6-bp sequence CACCG(A/C) as representative of CE, which has been reported as the target sequences of ZmABI4 (Niu et al., 2002). The effect of CE(a) (CACCGA) alone on high-level expression was relatively small in the seed [Figure 3b, CE(a)]. Genes containing both ABRE and CE(a) showed up to sevenfold higher frequency relative to the random percentage [Figure 3b, CE(a)+ABRE]. A similar pattern was also observed when the CE(c) (CACCGC) sequence was present (data not shown). These analyses are also consistent with the previous promoter analysis (Shen et al., 1996).

The RY/Sph motif has been reported to regulate seedspecific gene expression (Bäumlein *et al.*, 1992; Hattori *et al.*, 1992). We surveyed the frequency of typical 7-bp RY sequences, CATGCAT [RY(t)] and CATGCAA [RY(a)], in wildtype dry seed. Analysis of RY(a) showed that the motif was moderately overrepresented in highly expressed genes in the dry seed (Figure 3b). However, analysis of upstream regions of highly expressed genes did not show significant increase in the percentage of RY(t) motifs (data not shown). The combined effects of RY(a) and ABRE showed up to 12-fold enrichment in highly expressed gene sets [Figure 3b, RY(a)+ABRE]. A similar pattern, but to a less extent than in RY(a), was observed with RY(t) (data not shown). These results suggest that the association of RY motif with ABRE facilitates strong gene induction in the seed.

We performed the same analysis on the abi5 and abi4 mutants, comparing them with wild type. As shown in Figure 3(b), there was only weak overrepresentation of ABRE in abi5. The frequency of CE alone in highly expressed genes in abi5 was similar to that in wild type, however frequency of the genes containing both CE and ABRE was apparently low. Analyses of RY(a) in the abi5 seed showed a slight less frequency in the highly expressed groups, although ABI5 does not bind directly to the RY motif. Finally, the combination of RY(a) and ABRE was reduced in genes with a high-level expression in abi5. Slight reduction of the effect of ABRE on high-level gene expression was also observed in the abi4 seed (Figure 3b). Although the effect of CE alone on expression levels in abi4 seed was similar to that in wild type and abi5, the effect of the abi4 mutation on the percentage of highly expressed genes containing both CE and ABRE was visible. The RY motif also occurred less frequently in highly expressed genes, and the frequency of RY and ABRE was largely reduced in the *abi4* mutant. These results showed fluctuations in the effect of cis-acting motifs, with ABRE having the most noticeable influence in the seed.

Co-regulation of chromosomal gene clusters in the seed

We also found co-regulation of adjacent genes occurred in the seed. In many cases, three to five adjoining genes showed related expression patterns after seed imbibition (Figure 4a). For example, four genes from At1g01080 to At1g01110 underwent upregulation regardless of their expression level, while other sets of genes





Figure 2. Changes of selected ABRE activity and ABA content in wild-type seeds during imbibition.

(a) Distribution pattern of genes in expression levels were compared in dry and 24 h-imbibed seeds, and numbers of the genes containing at least one CACGTG in the 1 kb promoter were pointed up in shaded box in the bar at the top. The number shown above the bar represents a percentage of the genes that have ABRE in each group. Insets are close ups of the groups where the expression level above 10 000.

(b) Frequency of ABRE-containing genes in each group. The percentage of genes containing at least one CACGTG in (a) was compared with the percentage when they appear evenly (15.5%). The number one in 'ABRE frequency' means that no enrichment of the ABRE is observed.

(c) A change in endogenous ABA content in the seed during imbibition.

(At1g01160-At1g01180, and from At1g01230 to At1g01250) were downregulated during this period. Co-regulation of some neighboring genes was verified by semiquantitative RT-PCR (Figure 4b). The tendency was observed overall on chromosomes. The frequency of three-gene-clusters



Figure 3. Copy number and combinatorial effect of the selected ABRE on high-level expression in wild-type and *abi* mutant seed.

(a) Effect of copy number of the selected ABRE on high-level expression. The percentage of genes containing CACGTG in each copy number was calculated in the expression group, and was compared with the percentage when they appear evenly.

(b) Combinatorial effects of CE or RY with ABRE on high-level expression in wild type, *abi4* and *abi5* mutant. Enrichment of genes containing at least one indicated motif in each expression group was calculated in wild type, *abi4* and *abi5* dry seed. Selected motifs are as follows: ABRE, CACGTG; CE(a), CACCGA; and RY(a), CATGCAA.

showing co-downregulation and co-upregulation after 24 h imbibition (149 and 136 clusters, respectively) in 3701 genes was significantly higher than that in the same



Figure 4. Regional expression profile in chromosome I during imbibition.

(a) Changes in expression level during imbibition. The data obtained from GeneChip analysis for genes in a representative region in chromosome I are shown. (b) Expression profile from distinct regions and semiquantitative RT-PCR. The data obtained from GeneChip analysis for selected regions are shown. Red asterisks indicate no probe sets on GeneChip. Semiquantitative RT-PCR was performed to examine the expression pattern of genes from At1g55070 to At1g55120 and from At1g55325 to At1g55365. The result of 18S rRNA is shown as control. DS, dry seeds; 6 h, 6 h-imbibed seeds; 12 h, 12 h-imbibed seeds; 24 h, 24 h-imbibed seeds; n.d., not detected.

number of genes chosen randomly [84.8 \pm 9.8 (t = 100), see Experimental procedures].

Clustered expression patterns were also observed in dry seed of the *abi4* and *abi5* mutants. These mutants displayed some variations in expression pattern compared with wild type, and especially in the *abi5* mutant. The overall pattern of expression clusters was, however, mostly similar in wild type and in both mutants as shown in Figure 5(a,b). This suggests that changes in gene expression observed in these mutants were not due to the progression of developmental stages, but rather the direct effect of *abi* mutations.

We then surveyed the chromosomal location of the genes whose expression was altered by *abi* mutations. Co-regulated gene clusters were commonly composed of

three to five genes, therefore, we assessed the frequency of downregulated gene pairs located within a distance of five genes. In *abi5* dry seed, 527 genes exhibited at least twofold decrease in expression level compared with wild type. Among them, 77 downregulated genes (14.6%) had another downregulated gene nearby. In contrast, in *abi4* dry seed only 18 of 309 downregulated genes (5.8%) were co-located. This trend was also found among upregulated genes. Ninety-five of 716 upregulated genes (13.3%) in *abi5* dry seed were co-located, compared with 25 of the 438 genes (5.7%) in *abi4* dry seed. Genes whose expression was affected by the *abi5* mutation co-localized more frequently than those affected by the *abi4* mutation. This trend was more apparent when the distance of co-regulated genes

Figure 5. Clustered expression profiles in *abi* mutants.

(a, b) Expression levels obtained from GeneChip analysis are shown for comparison between wild type and abi mutants. (a) A region in chromosome III and (b) a region in chromosome V. (c, d) Co-location frequencies of ABI4- and ABI5regulated genes. The number of co-downregulated (c) or co-upregulated (d) genes, which are located within five-gene distance in abi4 and abi5 are shown, (c) Seventy-seven and 18 genes that were twofold downregulated in abi5 and abi4 dry seeds were co-located within five genes, respectively. Distances between downregulated genes were counted. (d) Ninety-five and 25 genes that were twofold upregulated in abi5 and abi4 dry seeds were co-located within five genes, respectively. Distances between upregulated genes were counted.



was plotted. Co-location of *abi5* downregulated genes was distance-dependent (Figure 5c), while this was not the case in *abi4* downregulated genes. A similar trend was also observed for upregulated genes (Figure 5d). These results suggest that the effect of ABI5 in trans-activation of gene expression is not merely restricted to a single gene, but affects expression of other proximal genes.

Clusters of gene families were often observed on chromosomes, and some of these are targets of co-regulation by ABI5 (Table S6). The downregulated clusters of genes include Lea proteins, 2S seed storage proteins and dehydrins, while the upregulated gene clusters include two sets of NAM transcription factors and genes related to metabolism.

We also found a number of large gene clusters that were hypo-expressed in dry seeds, each composed of more than 10 genes, and these clusters are located on many different chromosomes (data not shown). In most cases, these regions contain transposon/retro-transposon-related genes, pseudogenes or tandem/inverted repeats of the gene family(-ies), such as serine/threonine protein kinases or germins. In particular, we found five silent regions which

cover more than 1 Mb. Transposons and retrotransposons in plants are normally methylated, leading to the formation of heterochromatin structures and transcriptional silencing (Hirochika et al., 2000; Kato et al., 2003). To examine whether these silent regions are methylated, Southern blot analysis was performed on genomic DNA isolated from dry seed, following digestion by either one of two isoscizomers Hpall or Mspl. Transposon/retrotransposon-related genes as well as pseudogenes on different chromosomes all gave larger hybridization signals in Hpall digestion compared with those in *Mspl* digestion, indicating that these genes were methylated (Figure 6a). Larger hybridization signals than expected were often detected, even after Mspl digestion, which suggests dense methylation of these regions including non-CG methylation. Figure 6(b) shows methylation status in tandem repeats of gene families. Inhibition of Hpall digestion was observed for these genes. In contrast, hybridization patterns following digestion by both enzymes were unchanged when ABI5 and LEC1 were used as probes (Figure 6c). This indicates that the non-expressed LEC1 as well as the highly expressed ABI5 were free from methylation in the seed.



Figure 6. Southern hybridization analysis for methylation status in dry seeds.

Genomic DNA prepared from dry seeds was cleaved with *Msp*I or *Hpa*II and hybridized with either genomic DNA or cDNA fragment of each gene as a probe. H, *Hpa*II digest and M, *Msp*I digest.

(a) Heterochromatin-related transposon genes or pseudogenes. 1, At2g07760; 2, At2g10256; 3, At3g30813; 4, At3g30834; 5, At3g30839; and 6, At4g07452.

(b) Genes from silent clusters that comprise tandem/inverted repeat of gene family. 7, At1g51805 (LRR-protein kinase); and 8, At4g23210 (Ser/Thr protein kinase).

(c) Genes from a region that shows three to five gene expression cluster. 9, At2g36270 (*ABI5*); and 10, At1g21970 (*LEC1*). List for each gene identifier, product name, and expected band sizes for completely digestion at CCGG sites are described in Supplementary Material S1.

Discussion

Stored mRNA species and phase transition in the seed

Mature dry seeds contain large amounts of mRNAs that are not only reservoirs from embryogenesis and seed maturation but also a provision for germination. Previous reports have identified only a subset of highly expressed mRNA species, such as the genes for seed reserve synthesis, Lea proteins, protein synthesis and degradation (Dal Degan *et al.*, 1994; Dietrich *et al.*, 1989). Our microarray analysis describes mRNA species representing more than half of all genes. This number was also equivalent to the transcriptome reported in other organs (http://www.arabidopsis.org/info/expression/ATGenExpress.jsp). Transcripts from all the ontological categories were included. However, among the 2-3% genes (approximately 500 genes) with the highest expression in the dry seed, genes for metabolism as well as protein synthesis and degradation were overrepresented (Figure S1, Table S7). Ribosomal proteins, translation initiation and elongation factors were especially conspicuous. This may reflect the indispensability of protein synthesis in early seed germination, and is consistent with the observation that germination process is completely halted by cycloheximide application (Rajjou et al., 2004). Interestingly, abundant mRNA species include the ones not only for the functionally described proteases/ peptidases but also for proteins related to the ubiquitin/ proteasome system. This suggests involvement of selective proteolysis during early seed germination, in addition to the massive breakdown in protein bodies.

Early studies using cotton seeds have suggested that long-lived mRNA can contribute to protein synthesis during early stages in germination (Dure and Waters, 1965). In addition to the role of long-lived mRNA, our expression profiling revealed that transcripts of a number of genes (Figure 1a, group 4) exhibited a transient accumulation within 6 h after imbibition (Figure 1a). Indeed, the association of ABRE with high gene expression that was observed in the dry seed was not seen in the 6 h-imbibed seed, showing that the mRNA profile is largely altered within this short period.

Cis-acting elements determining transcriptome of stored mRNA

Synthesis of subsets of mRNAs that are stored in the seed is known to be positively regulated by ABA. ABA-mediated gene expression is in turn regulated through several distinct ABREs, however these ABREs are not necessarily involved in all ABA-mediated processes. Our expression profiling identified a G-box-like CACGTG sequence as the most prominent ABRE in dry seed transcriptome. The enrichment of this element is well correlated with the endogenous ABA level that dramatically decreases within 6 h, almost reaching the basal level by 12 h (Figure 2b,c). In addition, similar sequences were more frequent in downregulated genes in abi5 seeds (Table 1 and Tables S1-S4). Furthermore these sequences are overrepresented in upregulated genes in seeds imbibed in 30 µM ABA (data not shown). These sequences play an important role in ABA-induced synthesis of embryonic mRNA in the seed.

Downregulated genes in *abi5* dry seed displayed overrepresented numbers of ABRE in the upstream region. However, upregulated genes in this mutant seed exhibited no significant overrepresentation of ABRE. This suggests that ABI5 contributes mostly to activation of ABRE-mediated transcription rather than repression.

Our analyses also showed that multiple copies of ABRE and combining CE or RY with ABRE results in increased expression in seed (Figure 3a,b), which is consistent with previous promoter analyses (Shen et al., 1996; Vasil et al., 1995). By monitoring the expression of 22 000 genes in the seed, these effects were confirmed to be functional in vivo. Although a contribution of CE and RY to regulation of gene expression was observed, these elements showed only weak correlation with high-level expression in wild-type seed, when analyzed individually (Figure 3b). A statistical cis-element search in downregulated genes in abi4 seed did not identify CE (data not shown). More research will be needed to clarify the role and function of the ABI4 and CEs. However, recent comprehensive expression analysis using Arabidopsis overexpressing maize VP1 showed that VP1regulated genes were enriched in ABRE (Suzuki et al., 2003). Similarly, our expression analysis of the abi3-8 mutant seed (Nambara et al., 2002) showed an increased occurrence of ABRE, but not RY, in downregulated genes in abi3 relative to the wild type (K. Nakabayashi and E. Nambara, unpublished data). Collectively, our analyses indicate that CE and RY play a role in vivo in controlling ABRE-mediated gene expression in seed.

Co-expression is frequently observed in seed

In addition to regulation of individual genes, our analysis showed that genes for which mRNA is stored in the seed are more often clustered on a chromosome rather than dispersed randomly. Stored mRNA from gene clusters of three to five genes (15-20 kb) was observed in the dry seed of wild type, abi4 and abi5 mutants (Figure 5a,b). Interestingly, our kinetic expression profiling revealed that proximal genes tend to be co-regulated during seed germination (Figure 4). Gene clustering was frequently observed throughout all chromosomal regions. Collectively, this indicates that regulation of both individual genes and regions containing multiple genes determines the transcriptome of dry and imbibed seed. A number of the clustered gene families include genes for 2S albumins and Lea proteins (Table S6). The expression of these genes is coordinated during seed maturation and late embryogenesis. Therefore, gene clustering may contribute to such developmentally controlled expression.

Our analysis reveals that in *abi5* dry seed, both upregulated and downregulated genes tend to be located in clusters (Figure 5c,d). This is not the case in *abi4* dry seed. This suggests that ABI5-dependent transcriptional regulation may contribute at least in part, to clustered gene expression. The upregulated genes in *abi5* do not enrich the ABRE, but they tend to be located proximally. Accordingly, the effect of ABI5 on regulation of gene clusters does not necessarily require ABRE binding. Some transcription factors are known to interact with nucleosome remodeling factors, therefore it is possible that ABI5 or its complex may influence chromosomal structure. On the contrary, ABI4-dependent regulation does not appear to affect expression of neighboring genes in dry seed (Figure 5c,d).

Large silent regions display characteristics of heterochromatin

Aside from observing cis-acting regulatory elements in gene expression, we also observed a number of silent regions throughout chromosomes. Although the size of such regions varies considerably, clusters became apparent where hypo-expressed genes (evaluated as 'absent' expression) were located in groups of more than 10 genes (approximately 50 kb). Most of these silent regions contain either pseudogene(s), retrotransposon(s) or clustered homologous genes. These traits are reminiscent of heterochromatin regions where a large number of proximally located genes are repressed. Methylation is often linked with heterochromatin that is associated with repetitive sequences including the centromere, transposable elements and rRNA repeats (Bender, 2004). Our results clearly illustrate that the large silent genomic regions are heavily methylated (Figure 6a). Most methylated regions contain more than 100 transposon-related genes, and in some cases almost 300 genes are sequentially linked, and map to the centromeric regions. The sizes of these large clusters are well matched to the estimated sizes of Arabidopsis centromeres (Haupt et al., 2001). Thus, these genes may be silenced transcriptionally by heterochromatin structure.

The genes in repeated non-transposon arrays were also targets of methylation (Figure 6b). In contrast to the centromeric retrotransposon elements, in which only a small number of fairly large bands were detected in Southern analysis of Hpall-digested DNA, CG methylation in the nontransposon genes or in the clusters was restricted. This restricted methylation pattern is similar to that observed in SUP and FWA genes and it is reported that hypomethylation of these genes leads to an increase in their expression (Jacobsen et al., 2000; Soppe et al., 2000). Interestingly, the silent cluster of serine/threonine protein kinases that we selected was found to be expressed in axillary buds (K. Tatematsu and E. Nambara, unpublished data). This suggests that this gene (or gene clusters) may be transcriptionally silenced in the seed as documented elsewhere (reviewed in Finnegan et al., 2000; Lohe and Chaudhury, 2002).

Our analyses suggest that expression profiles in Arabidopsis seeds during different phases of development are defined by multiple regulatory mechanisms: epigenetic heterochromatin structures and conditional silencing by methylation, unknown factor(s) controlling expression of clustered genes, and individual *cis*-acting elements.

In silico analysis: a powerful tool for dissecting expression profiles

The accumulation of microarray data has provided us with a large amount of information on gene expression. In many cases, however, microarray data have been used to show a change in the expression level of specific genes with a particular treatment or a difference between wild-type and mutant plants. One of our approaches to fully utilize expression data from GeneChip analyses was to display the distribution of expression levels of 22 000 Arabidopsis genes without comparison to other GeneChip data as shown in Figures 2 and 3. Based on the distribution of these expression levels, particular ontological categories harboring selected single or multiple *cis*-acting elements were shown to be disproportionately represented in subsets with particular expression levels.

In addition, microarray analysis of 22 000 Arabidopsis genes enabled us to examine gene expression as influenced by chromosomal location. It is generally assumed that specific cell types will have specific patterns of gene clustering as in Arabidopsis root cells (Birnbaum et al., 2003). However, in this study, we could see both genetic and epigenetic regulation within the transcriptome of stored RNA in dry and imbibed seeds by analyzing the whole seed (without any separation of particular cell types). This may be due to the simplicity of the seed structure. In addition, there may be few differences in the composition of cell types in dry and 24 h-imbibed seeds, as radicle protrusion was initiated only after 30-36 h of imbibition. Our expression profiling demonstrates that seed is an excellent system in which to study epigenetic regulation using microarraybased expression profiles.

A number of detailed studies of ABA-inducible promoters have identified ABRCs that are necessary and sufficient for ABA-induced transcription. These previous *in vitro* and transient expression studies provided the basis that enabled us to use the ABI5/ABRE system as a model to evaluate our expression profiling. With validated data sets and analytical conditions, our results describe the transcriptional profile of the seed, and verify previous results on *in vitro* expression patterns by showing that these also occur in *in vivo* situations. The next challenge is to identify new *cis*-acting elements in novel mutants using our profiling system.

Experimental procedures

Plants and growth conditions

Plants used in this study were *Arabidopsis thaliana* (L.) Heynh of ecotype Columbia, and mutants *abi4-11* and *abi5-7* in a Columbia background (Nambara *et al.*, 2002). All plants were grown or treated as previously described (Kushiro *et al.*, 2004). Seeds were imbibed under continuous light without stratification, after

washing in 0.04% Triton X-100 and rinsing several times with water.

DNA microarray analysis

Extraction of RNA was performed using RNAqueous columns (Ambion, Austin, TX, USA), and microarray analysis was performed using the GeneChip Arabidopsis ATH1 Genome Array as described previously (Kushiro *et al.*, 2004). Basic data analyses to obtain values for signal intensity and detection call were conducted using MicroArray Suite 5.0 (Affymetrix, Santa Clara, CA, USA). Sample quality was assessed by calculating 3' to 5' intensity ratios of certain genes before subsequent analysis. For all microarray experiments, RNA from two independent biological replicates was used for hybridization and subsequent analyses. Data for dry and 24 himbibed seeds in Col-0, *abi5* and *abi4* are shown in Tables S7–S12.

Data analysis

Advanced data analyses were performed using Microsoft Excel and GeneSpring 4.2 software (Silicongenetics, Redwood, CA, USA). First, a per-chip normalization was performed using the 50th percentile of all measurements in order to adjust total signal intensity in each chip. For the clustering analysis, a per-gene normalization (using the median for each gene) as well as a per-chip normalization (the 50th percentile) was applied. For the selection of upregulated or downregulated genes, we only used the set of genes of which detection call was 'present' at least in either condition in a basic analysis by MicroArray Suite software. To search for the candidate cis-regulatory sequences, the 'Find Potential Regulatory Sequences' tool in GeneSpring was used based on the following criteria: from 6- to 8-nucleotide length without any point discrepancies within 10-1000 bases upstream region from the translational start site of each gene. In all statistical analysis relative to upstream region of other (unselected) genes, the cutoff P-value was set at 0.05. For calculation of 'fold increase' in distribution of the cis-elements, we first searched for the motif of interest in the nuclear genes represented on the ATH1 GeneChip, identified the number of genes containing the motif, and calculated the overall distribution rate. Subsequently, the frequency of the motif in each expression level group was calculated and compared with the overall rate. A onefold increase indicates that the frequency was the same as the overall rate. The search for genes with multiple ABRE within 1 kb region upstream of the coding region was carried out using a developing program generated by Mr J. Koch. The overall rates for one, two, and more than three copies of ABRE were calculated as 12.6, 2.60, and 0.54%, respectively.

To examine significance of co-regulation, twofold downregulated (3701 genes) and 2.5-fold upregulated (3701 genes) nuclear genes were selected by comparison between dry and 24 h-imbibed seed transcriptomes, and the number of three-gene clusters was counted in those gene sets. Random frequency was obtained by counting the occurrence of three-gene clusters in randomly chosen 3701 elements out of 22 591 nuclear genes represented on the GeneChip (t = 100).

Throughout the data sets, genes were identified by the AGI gene code from Munich Information Center for Protein Sequence (MIPS, http://mips.gsf.de/). Correspondence between the AGI codes and the Affymetrix gene numbers was determined using the microarray element file at The Arabidopsis Information Resource (TAIR, http://www.arabidopsis.org/). Genes were classified into 13 groups described in Figure 1 based on the gene ontology obtained from the Kyoto Encyclopedia of genes and Genomes (KEGG, http://www.genome.ad.jp/kegg/) (Kanehisa *et al.*, 2002) and MIPS *Arabidopsis thaliana* DataBase (MatDB, http://mips.gsf.de/proj/thal/ db/index.html) (Schoof *et al.*, 2002). This classification was only applied to nuclear genes, thus genes encoded in the chloroplast and mitochondria genome were categorized as 'unclassified.'

Measurement of ABA level

Endogenous ABA content was quantified by GC-MS as described in Kushiro *et al.* (2004). ABA quantification was performed three times using independent plant material.

RT-PCR analyses

First strand cDNA was synthesized from total RNA as described (Kushiro *et al.*, 2004). For validation of expression pattern, semiquantitative PCR analysis was performed. Selected genes and primer sequences are shown in Supplementary Material. Duplicate experiments were performed using independent seed batch.

Extraction of genomic DNA and Southern blot analysis

Genomic DNA was prepared from dry seeds according to the method of Wagner et al. (1987) with some modifications. Dry seeds (160 mg) were ground in liquid nitrogen and homogenized in 6.4 ml extraction buffer [0.1 M Tris-HCl, pH 8.0, 0.35 M sorbitol, 10% (w/v) polyethylene glycol (Mr 6000), 0.5% spermine, 0.5% spermidine, 0.5% β-mercaptoethanol]. A pellet was collected by centrifugation (20 000 g, 4°C, 10 min) and repeatedly washed several times with the same extraction buffer. The pellet was resuspended in 2 ml lysis buffer (0.1 м Tris-HCl, pH 8.0, 0.35 м sorbitol, 0.5% spermine, 0.5% spermidine, 0.5% β-mercaptoethanol), and N-lauroylsarcosine was added to a concentration of 1% (w/v). After 10 min at room temperature, equal volume of CTAB extraction buffer [2% (w/v) cetyltrimethylammonium bromide (СТАВ), 0.1 м Tris-HCl, pH 8.0, 20 mм EDTA, 1.4 м NaCl] was added and incubated at 65°C for 15 min. After chloroform/isoamylalcohol (24:1) extraction, equal volume of CTAB precipitation buffer [1% (w/v) CTAB, 50 mm Tris-HCI, pH 8.0, 10 mm EDTA] was added to aqueous phase. A pellet was collected by centrifugation (20 000 g, 24°C, 10 min) and dissolved in TE. After ethanol precipitation, genomic DNA was treated with RNaseA, extracted with phenol/chloroform and finally dissolved in TE buffer.

DNA was cleaved with either *Msp*l or *Hpa*ll and separated on 0.7% agarose. DNA was transferred to Hybond-N+ (Amersham Biosciences, Piscataway, NJ, USA) according to the manufacturer's instruction. Labeling of probes, hybridization and signal detection were performed using *Gene Images* random prime labeling module and *Gene Images* CDP-*Star* detection module (Amersham Biosciences) according to the manufacturer's instruction. Chemiluminescence was detected using the luminescent image analyzer LAS-3000 (Fuji Film, Tokyo, Japan). Duplicate experiments were performed using independent plant material.

Acknowledgements

We thank Mr J. Koch for sharing his developing program for obtaining a gene set containing different numbers of the selected *cis*-element. We also thank Dr Toru Fujiwara for his critical reading of the manuscript, and Ms Kaori Kuwata for her general assistance.

© Blackwell Publishing Ltd, The Plant Journal, (2005), 41, 697–709

Supplementary Material

The following material is available from http://www. blackwellpublishing.com/products/journals/suppmat/TPJ/TPJ2337/ TPJ2337sm.htm

Figure S1. Ontological classification of the gene set showing an expression level >10 000 in dry seed. Ratio of each classification in all genes is indicated at the bottom. Arrows indicate the overrepresented categories in the selected gene set.

Material S1. An excel file containing following content in each worksheet. (1) Overlap of gene sets in duplicate GeneChip analyses, (2) gene identifiers and primer sequences used for RT-PCR analysis in K-means clustering groups, (3) primer sequences for RT-PCR analysis in chromosomal clustering, and (4) gene identifiers and expected signal sizes in Southern hybridization methylation assay.

 Table S1 Individual sequences with increased frequency in the 1 kb

 upstream region of highly expressed genes in wild-type dry seeds

 (first batch)

 Table S2 Individual sequences with increased frequency in the 1 kb

 upstream region of highly expressed genes in wild-type dry seeds

 (second batch)

Table S3 Individual sequences with increased frequency in the 1 kb upstream region of downregulated genes in *abi5* dry seeds (first batch)

Table S4 Individual sequences with increased frequency in the 1 kb upstream region of downregulated genes in *abi5* dry seeds (second batch)

 Table S5
 A list of genes from Affymetrix GeneChip ATH1 that contain ABRE in 1 kb upstream region

 Table S6 Proximally located gene families that are co-regulated by
 ABI5

Table S7 Expression level of genes in wild-type dry seed (first batch). Genes are listed beginning with those which had the highest expression level. Col_ds_Signal, absolute signal value; Col_ds_detection, detection flag based on the statistical algorithm (P, present; A, absent; and M, marginal); Col_ds_detection_P-value, false-positive probabilities in statistical calculation when hypothesized that the gene is expressed

Table S8Expression level of genes in wild-type 24 h-imbibedseed (first batch).Abbreviations in the table are as described inTable S7

Table S9 Expression level of genes in dry seed in *abi5* (first batch).

 Abbreviations in the table are as described in Table S7

Table S10 Expression level of genes in 24 h-imbibed seed in *abi5*(first batch). Abbreviations in the table are as described in Table S7**Table S11** Expression level of genes in dry seed in *abi4* (first batch).Abbreviations in the table are as described in Table S7

 Table S12 Expression level of genes in 24 h-imbibed seed in abi4

 (first batch). Abbreviations in the table are as described in Table S7

 All GeneChip data will be deposited to AtGenExpress.

References

- Almoguera, C. and Jordano, J. (1992) Developmental and environmental concurrent expression of sunflower dry-seed-stored lowmolecular-weight heat-shock protein and Lea mRNAs. *Plant Mol. Biol.* 19, 781–792.
- Argüello-Astorga, G.R. and Herrera-Estrella, L.R. (1996) Ancestral multipartite units in light-responsive plant promoters have structural features correlating with specific phototransduction pathways. *Plant Physiol.* **112**, 1151–1166.
- Bäumlein, H., Nagy, I., Villarroel, R., Inzé, D. and Wobus, U. (1992) Cis-analysis of a seed protein gene promoter: the conservative RY

repeat CATGCATG within the legumin box is essential for tissuespecific expression of a legumin gene. *Plant J.* **2**, 233–239.

- Bender, J. (2004) DNA methylation and epigenetics. Annu. Rev. Plant Biol. 55, 41–68.
- Bewley, J.D. (1997) Seed germination and dormancy. *Plant Cell*, 9, 1055–1066.
- Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W. and Benphy, P.N. (2003) A gene expression map of the *Arabidopsis* root. *Science*, **302**, 1956–1960.
- Busk, P.K. and Pagès, M. (1998) Regulation of abscisic acid-induced transcription. *Plant Mol. Biol.* **37**, 425–435.
- Carles, C., Bies-Etheve, N., Aspert, L., Léon-Kloosterziel, K.M., Koornneef, M., Echeverria, M. and Delseny, M. (2002) Regulation of Arabidopsis thaliana Em genes: role of ABI5. Plant J. 30, 373– 383.
- Comai, L., Dietrich, R.A., Maslyar, D.J., Baden, C.S. and Harada, J.J. (1989) Coordinate expression of transcriptionally regulated isocitrate lyase and malate synthase genes in *Brassica napus* L. *Plant Cell*, 1, 293–300.
- Cuming, A.C. (1999) LEA proteins. In Seed Proteins (Shewry, P.R. and Casey, R., eds). Dordrecht: Kluwer Academic, pp. 753–780.
- Dal Degan, F., Rocher, A., Cameron-Mills, V. and von Wettstein, D. (1994) The expression of serine carboxypeptidases during maturation and germination of the barley grain. *Proc. Natl Acad. Sci.* USA, 91, 8209–8213.
- Dickinson, C.D., Evans, R.P. and Nielsen, N.C. (1988) RY repeats are conserved in the 5'-flanking regions of legume seed-protein genes. *Nucleic Acids Res.* 16, 371.
- Dietrich, R.A., Maslyar, D.J., Heupel, R.C. and Harada, J.J. (1989) Spatial patterns of gene expression in *Brassica napus* seedlings: identification of a cortex-specific gene and localization of mRNAs encoding isocitrate lyase and a polypeptide homologous to proteinases. *Plant Cell*, **1**, 73–80.
- Dong, J., Keller, W.A., Yan, W. and Georges, F. (2004) Gene expression at early stages of *Brassica napus* seed development as revealed by transcript profiling of seed-abundant cDNAs. *Planta*, 218, 483–491.
- Dure, L. and Waters, L. (1965) Long-lived messenger RNA: evidence from cotton seed germination. *Science*, **147**, 410–412.
- Ezcurra, I., Ellerström, M., Wycliffe, P., Ståberg, K. and Rask, L. (1999) Interaction between composite elements in the *napA* promoter: both the B-Box ABA-responsive complex and the RY/G complex are necessary for seed-specific expression. *Plant Mol. Biol.* **40**, 699–709.
- Finkelstein, R.R. and Lynch, T.J. (2000) The Arabidopsis abscisic acid response gene ABI5 encodes a basic leucine zipper transcription factor. Plant Cell, 12, 599–609.
- Finkelstein, R.R. and Somerville, C.R. (1990) Three classes of abscisic acid (ABA)-insensitive mutations of *Arabidopsis* define genes that control overlapping subsets of ABA responses. *Plant Physiol.* 94, 1172–1179.
- Finkelstein, R.R., Wang, M.L., Lynch, T.J., Rao, S. and Goodman, H. (1998) The Arabidopsis abscisic acid response locus ABI4 encodes an APETALA2 domain protein. *Plant Cell*, **10**, 1043– 1054.
- Finnegan, E.J., Peacock, W.J. and Dennis, E.S. (2000) DNA methylation, a key regulator of plant development and other processes. *Curr. Opin. Genet. Dev.* **10**, 217–223.
- Giraudat, J., Hauge, B.M., Valon, C., Smalle, J., Parcy, F. and Goodman, H.M. (1992) Isolation of the Arabidopsis *ABI3* gene by positional cloning. *Plant Cell*, **4**, 1251–1261.
- Girke, T., Todd, J., Ruuska, S., White, J., Benning, C. and Ohlrogge, J. (2000) Microarray analysis of developing Arabidopsis seeds. *Plant Physiol.* **124**, 1570–1681.

- Guiltinan, M.J., Marcotte, Jr, W.R. and Quatrano, R.S. (1990) A plant leucine zipper protein that recognizes an abscisic acid response element. *Science*, 250, 267–271.
- Harada, J.J., Baden, C.S. and Comai, L. (1988) Spatially regulated genes expressed during seed germination and postgerminative development are activated during embryogeny. *Mol. Gen. Genet.* 212, 466–473.
- Hattori, T., Vasil, V., Rosenkrans, L., Hannah, L.C., McCarty, D.R. and Vasil, I.K. (1992) The *Viviparous-1* gene and abscisic acid activate the *C1* regulatory gene for anthocyanin biosynthesis during seed maturation in maize. *Genes Dev.* 6, 609–618.
- Haupt, W., Fischer, T.C., Winderl, S., Fransz, P. and Torres-Ruiz, R.A. (2001) The CENTROMERE1 (CEN1) region of *Arabidopsis thaliana*: architecture and functional impact of chromatin. *Plant J.* 27, 285–296.
- Hirochika, H., Okamoto, H. and Kakutani, T. (2000) Silencing of retrotransposons in Arabidopsis and reactivation by the *ddm1* mutation. *Plant Cell*, **12**, 357–369.
- Hobo, T., Kowyama, Y. and Hattori, T. (1999a) A bZIP factor, TRAB1, interacts with VP1 and mediates abscisic acid-induced transcription. *Proc. Natl Acad. Sci. USA*, **96**, 15348–15353.
- Hobo, T., Asada, M., Kowyama, Y. and Hattori, T. (1999b) ACGTcontaining abscisic acid response element (ABRE) and coupling element 3 (CE3) are functionally equivalent. *Plant J.* **19**, 679– 689.
- Hughes, D.W. and Galau, G.A. (1989) Temporally modular gene expression during cotyledon development. *Genes Dev.* 3, 358– 369.
- Hughes, D.W. and Galau, G.A. (1991) Developmental and environmental induction of *Lea* and *LeaA* mRNAs and the postabscission program during embryo culture. *Plant Cell*, 3, 605–618.
- Ishibashi, N., Yamauchi, D. and Minamikawa, T. (1990) Stored mRNA in cotyledons of *Vigna unguiculata* seeds: nucleotide sequence of cloned cDNA for a stored mRNA and induction of its synthesis by precocious germination. *Plant Mol. Biol.* **15**, 59–64.
- Jacobsen, S.E., Sakai, H., Finnegan, E.J., Cao, X. and Meyerowitz, E.M. (2000) Ectopic hypermethylation of flower-specific genes in *Arabidopsis. Curr. Biol.* 10, 179–186.
- Kanehisa, M., Goto, S., Kawashima, S. and Nakaya, A. (2002) The KEGG database at GenomeNet. *Nucleic Acids Res.* 30, 42–46.
- Karssen, C.M., Brinkhorst-van der Swan, D.L.C., Breekland, A.E. and Koornneef, M. (1983) Induction of dormancy during seed development by endogenous abscisic acid: studies on abscisic acid deficient genotypes of *Arabidopsis thaliana* (L.) Heynh. *Planta*, **157**, 158–165.
- Kato, M., Miura, A., Bender, J., Jacobsen, S.E. and Kakutani, T. (2003) Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis. *Curr. Biol.* 13, 421–426.
- Kim, S.Y., Ma, J., Perret, P., Li, Z. and Thomas, T.L. (2002) Arabidopsis ABI5 subfamily members have distinct DNA-binding and transcriptional activities. *Plant Physiol.* **130**, 688–697.
- Koornneef, M., Reuling, G. and Karssen, C.M. (1984) The isolation and characterization of abscisic acid-insensitive mutants of *Arabidopsis thaliana*. *Physiol. Plant.* 61, 377–383.
- Kuligowski, J., Ferrand, M. and Chenou, E. (1991) Stored mRNA in early embryos of a fern *Marsilea vestita*: a paternal and maternal origin. *Mol. Reprod. Dev.* 30, 27–33.
- Kushiro, T., Okamoto, M., Nakabayashi, K., Yamagishi, K., Kitamura, S., Asami, T., Hirai, N., Koshiba, T., Kamiya, Y. and Nambara,
 E. (2004) The *Arabidopsis* cytochrome P450 CYP707A encodes ABA 8'-hydroxylases: key enzymes in ABA catabolism. *EMBO J.* 23, 1647–1656.
- Lohe, A.R. and Chaudhury, A. (2002) Genetic and epigenetic processes in seed development. Curr. Opin. Plant Biol. 5, 18–25.

- Lopez-Molina, L. and Chua, N.-H. (2000) A null mutation in a bZIP factor confers ABA-insensitivity in *Arabidopsis thaliana*. *Plant Cell Physiol.* 41, 541–547.
- Lu, C.-A., Lim, E.-K. and Yu, S.-M. (1998) Sugar response sequence in the promoter of rice α-amylase gene serves as a transcriptional enhancer. J. Biol. Chem. 273, 10120–10131.
- Marcotte, Jr, W.R., Russell, S.H. and Quatrano, R.S. (1989) Abscisic acid-responsive sequences from the Em gene of wheat. *Plant Cell*, 1, 969–976.
- McCarty, D.R. (1995) Genetic control and integration of maturation and germination pathways in seed development. Annu. Rev. Plant Physiol. Plant Mol. Biol. 46, 71–93.
- Mönke, G., Altschmied, L., Tewes, A., Reidt, W., Mock, H.-P., Bäumlein, H. and Conrad, U. (2004) Seed-specific transcription factors ABI3 and FUS3: molecular interaction with DNA. *Planta*, 219, 158–166.
- Nakamura, S., Lynch, T.J. and Finkelstein, R.R. (2001) Physical interactions between ABA response loci of *Arabidopsis*. *Plant J.* 26, 627–635.
- Nambara, E., Naito, S. and McCourt, P. (1992) A mutant of *Arabidopsis* which is defective in seed development and storage protein accumulation is a new *abi3* allele. *Plant J.* **2**, 435–441.
- Nambara, E., Hayama, R., Tsuchiya, Y., Nishimura, M., Kawaide, H., Kamiya, Y. and Naito, S. (2000) The role of *ABI3* and *FUS3* loci in *Arabidopsis thaliana* on phase transition from late embryo development to germination. *Dev. Biol.* 220, 412–423.
- Nambara, E., Suzuki, M., Abrams, S., McCarty, D.R., Kamiya, Y. and McCourt, P. (2002) A screen for genes that function in abscisic acid signaling in *Arabidopsis thaliana*. *Genetics*, 161, 1247–1255.
- Niu, X., Helentjaris, T. and Bate, N.J. (2002) Maize ABI4 binds coupling element1 in abscisic acid and sugar response genes. *Plant Cell*, 14, 2565–2575.
- Ogawa, M., Hanada, A., Yamauchi, Y., Kuwahara, A., Kamiya, Y. and Yamaguchi, S. (2003) Gibberellin biosynthesis and response during Arabidopsis seed germination. *Plant Cell*, **15**, 1591–1604.
- Parcy, F., Valon, C., Kohara, A., Miséra, S. and Giraudat, J. (1997) The ABSCISIC ACID-INSENSITIVE3, FUSCA3, and LEAFY COTY-LEDON1 loci act in concert to control multiple aspects of Arabidopsis seed development. Plant Cell, 9, 1265–1277.
- Rajjou, L., Gallardo, K., Debeaujon, I., Vandekerckhove, J., Job, C. and Job, D. (2004) The effect of α-amanitin on the Arabidopsis seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. *Plant Physiol.* **134**, 1598–1613.
- Schoof, H., Zaccaria, P., Gundlach, H., Lemcke, K., Rudd, S., Kolesov, G., Arnold, R., Mewes, H.W. and Mayer, K.F.X. (2002) MIPS

Arabidopsis thaliana Database (MatDB): an integrated biological knowledge resource based on the first complete plant genome. *Nucleic Acids Res.* **30**, 91–93.

- Shen, Q. and Ho, T.-H.D. (1995) Functional dissection of an abscisic acid (ABA)-inducible gene reveals two independent ABAresponsive complexes each containing a G-box and a novel *cis*acting element. *Plant Cell*, 7, 295–307.
- Shen, Q., Uknes, S.J. and Ho, T.-H.D. (1993) Hormonal response complex in a novel abscisic acid and cycloheximide-inducible barley gene. J. Biol. Chem. 268, 23652–23660.
- Shen, Q., Zhang, P. and Ho, T.-H.D. (1996) Modular nature of abscisic acid (ABA) response complexes: composite promoter units that are necessary and sufficient for ABA induction of gene expression in barley. *Plant Cell*, 8, 1107–1119.
- Skriver, K., Olsen, F.L., Rogers, J.C. and Mundy, J. (1991) Cis-acting DNA elements responsive to gibberellin and its antagonist abscisic acid. *Proc. Natl Acad. Sci. USA*, 88, 7266–7270.
- Soppe, W.J.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M. and Peeters, A.J.M. (2000) The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of homeodomain gene. *Mol. Cell*, 6, 791– 802.
- Suzuki, M., Kao, C.Y. and McCarty, D.R. (1997) The conserved B3 domain of VIVIPAROUS1 has cooperative DNA binding activity. *Plant Cell*, 9, 799–807.
- Suzuki, M., Ketterling, M.G., Li, Q.-B. and McCarty, D.R. (2003) Viviparous1 alters global gene expression patterns through regulation of abscisic acid signaling. *Plant Physiol.* **132**, 1664– 1677.
- Thomas, T.L. (1993) Gene expression during plant embryogenesis and germination: an overview. *Plant Cell*, 5, 1401–1410.
- Vasil, V., Marcotte, Jr, W.R., Rosenkrans, L., Cocciolone, S.M., Vasil, I.K., Quatrano, R.S. and McCarty, D.R. (1995) Overlap of Viviparous 1 (VP1) and abscisic acid response elements in the *Em* promoter: G-box elements are sufficient but not necessary for VP1 transactivation. *Plant Cell*, 7, 1511–1518.
- Wagner, D.B., Furnier, G.R., Saghai-Maroof, M.A., Williams, S.M., Dancik, B.P. and Allard, R.W. (1987) Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. *Proc. Natl Acad. Sci. USA*, 84, 2097–2100.
- White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martínez de Ilárduya, O., Jaworski, J.G., Ohlrogge, J.B. and Benning, C. (2000) A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil. *Plant Physiol.* **124**, 1582–1594.