

Functional genomics using RIKEN *Arabidopsis thaliana* full-length cDNAs

Motoaki Seki · Kazuo Shinozaki

Received: 3 February 2009 / Accepted: 8 April 2009 / Published online: 2 May 2009
© The Botanical Society of Japan and Springer 2009

Abstract Full-length cDNAs are essential for the correct annotation of genomic sequences as well as for the functional analysis of genes and their products. We have isolated about 240,000 RIKEN *Arabidopsis* full-length (RAFL) cDNA clones. These clones were clustered into about 17,000 non-redundant cDNA groups, i.e., about 60% of all *Arabidopsis* predicted genes. The sequence information of the RAFL cDNAs is useful for promoter analysis, and for the correct annotation of predicted transcriptional units and gene products. We prepared cDNA microarrays containing independent full-length cDNA groups and studied the expression profiles of genes under various stress- and hormone-treatment conditions, and in various mutants and transgenic plants. These expression profiling studies have shown the expression levels of many genes as a detailed snapshot describing the

state of a biological system in planta under various conditions. We have applied RAFL cDNAs to the functional analysis of proteins using the full-length cDNA over-expressing (FOX) gene hunting system and the wheat germ cell-free protein synthesis system. The RAFL cDNA collection was also used for determination of the domain structure of proteins by NMR. In this review, we summarize the present state and perspectives of functional genomics using RAFL cDNAs.

Keywords *Arabidopsis thaliana* · DNA microarray · Environmental stress · Functional genomics · RAFL cDNA

Introduction

Arabidopsis thaliana has been adopted as a model organism in the study of plant biology because of its small size, short generation time, and high efficiency of transformation (Meinke et al. 1998). The whole genome sequence has been determined by the *Arabidopsis* Genome Initiative (AGI) (The *Arabidopsis* Genome Initiative 2000).

About 1,500,000 expressed sequence tags (ESTs) from *Arabidopsis* have been deposited in the EST database (dbEST) as of 5 January 2009, including sequences from large-scale EST projects in France (Höfte et al. 1993; Cooke et al. 1996), the United States (Newman et al. 1994; White et al. 2000), and Japan (Asamizu et al. 2000). These projects have produced EST data from different tissues, organs, and developmental stages (Höfte et al. 1993; Newman et al. 1994; Cooke et al. 1996; Asamizu et al. 2000; White et al. 2000). However, as of 1996, only about 50,000 *Arabidopsis* ESTs were registered (Höfte et al. 1993; Newman et al. 1994; Cooke et al. 1996) and most of these EST projects were based on cDNA libraries in which

Motoaki Seki is the recipient of the BSJ Award for Young Scientists, 2005.

M. Seki (✉)
Plant Genomic Network Research Team,
Plant Functional Genomics Research Group,
RIKEN Plant Science Center (PSC),
RIKEN Yokohama Institute,
1-7-22 Suehiro-cho, Tsurumi-ku,
Yokohama 230-0045, Japan
e-mail: mseki@psc.riken.jp

M. Seki
Kihara Institute for Biological Research,
Yokohama City University, 641-12 Maioka-cho,
Totsuka-ku, Yokohama 244-0813, Japan

K. Shinozaki
Gene Discovery Research Team,
RIKEN Plant Science Center (PSC), RIKEN Tsukuba Institute,
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

most of the inserts are not full-length. ESTs are useful for making a catalog of expressed genes, but not for further study of gene function. Consequently, genome-scale collections of the full-length cDNAs of expressed genes are important for the analysis of the structure and function of genes and their products in this era of functional genomics.

Since 1996, we have constructed *Arabidopsis* full-length cDNA libraries from plants grown under different conditions (Seki et al. 1998, 2002a) using the biotinylated CAP trapper method exploiting trehalose-thermoactivated reverse transcriptase (Carninci et al. 1996, 1997, 1998) and about 240,000 RIKEN *Arabidopsis* full-length (RAFL) clones have been isolated (Seki et al. 2002a; Sakurai et al. 2005). At present, there are numerous *Arabidopsis* full-length cDNAs produced and deposited in the GenBank database by other groups, such as Ceres (Haas et al. 2002), Genoscope (Castelli et al. 2004), and others. Information on these full-length cDNAs is available at http://www.arabidopsis.org/portals/masc/ORFeomics_2008Report.pdf. Full-length cDNAs have many advantages for improvement of genome annotation and functional genomics in the post-sequencing era (Fig. 1) (Seki et al. 2001a, 2002b, 2004a).

In this review, we summarize the present state and perspectives of analyses using RAFL cDNAs, including their collection and annotation, their application to expression profiling, and the structural and functional analysis of plant proteins.

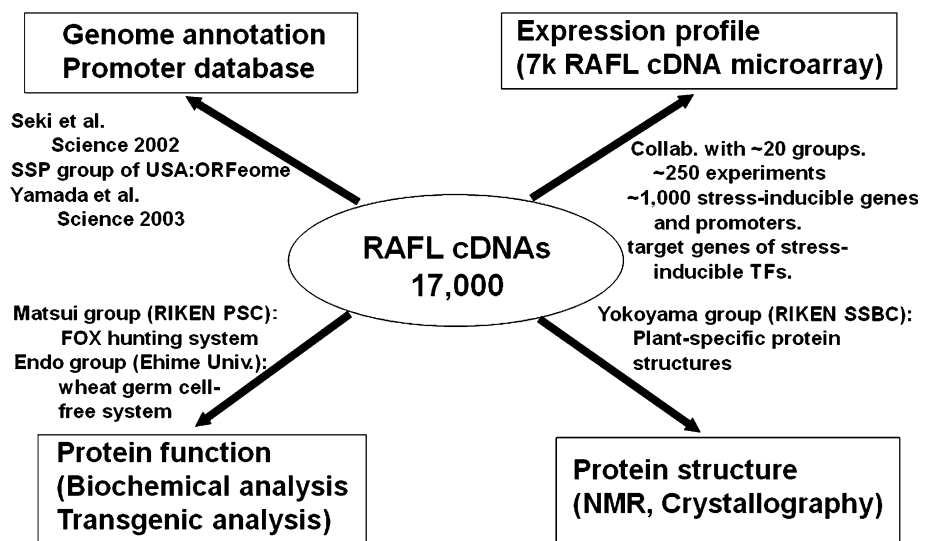
Collection and sequencing of RAFL cDNAs

As reported previously, we have constructed *Arabidopsis* full-length cDNA libraries from plants grown under various stress, hormone and light conditions, from plants at various developmental stages, and from various plant

tissues (Seki et al. 1998, 2002a) using the biotinylated CAP trapper method with trehalose-thermoactivated reverse transcriptase (Carninci et al. 1996, 1997, 1998). The overall strategy for preparing cDNA libraries, including standard, normalized, and subtracted libraries, has been described (Seki et al. 2001b). We have isolated about 240,000 RAFL cDNA clones, clustered into about 17,000 non-redundant cDNA groups, representing about 60% of all *Arabidopsis* predicted genes (Fig. 1; Seki et al. 2002a; Sakurai et al. 2005). Note that all *Arabidopsis* full-length cDNAs including the RAFL cDNAs are mapped on about 19,000 loci in the *Arabidopsis* genome.

Using the 5'-end sequences of mRNAs, promoter sequences can be obtained by comparison with *Arabidopsis* genomic sequences. We obtained 5'-ESTs of the RAFL cDNA clones and constructed a promoter database (Seki et al. 2002a; Sakurai et al. 2005) using the plant cis-acting regulatory DNA elements (PLACE) database (Higo et al. 1999). The *Arabidopsis* promoter database thus constructed contains information on genomic sequences 1,000-bp upstream from the 5'-terminus of each RAFL cDNA clone, and cis-acting elements known from plants, and is available as part of the RIKEN *Arabidopsis* Genome Encyclopedia (RARGE) database (<http://rarge.gsc.riken.go.jp/>; Sakurai et al. 2005). Several established plant promoter databases are also available today, such as the *Arabidopsis* Gene Regulatory Information Server (AGRIS, [http://arabidopsis.med.ohio-state.edu.](http://arabidopsis.med.ohio-state.edu/); Davuluri et al. 2003). Yamamoto et al. (2007) have applied local distribution of short sequences (LDSS) analysis to extract promoter constituents by genome-wide statistical analysis, and have identified 1,000 octamer sequences as LDSS-positive promoter elements. The information on core promoters thus extracted is available at the plant promoter database (PPDB, <http://www.ppdb.gene.nagoya-u.ac.jp>).

Fig. 1 Application of RIKEN *Arabidopsis thaliana* full-length (RAFL) cDNAs to plant functional genomics



Although many algorithms have been written to predict a transcriptional unit (TU) from genomic sequence data, the accuracy of such predictions is still limited. A more direct and efficient approach to identify coding sequences is to sequence full-length cDNAs (Fig. 1). We have been determining the full-length sequences of the RAFL cDNA clones in collaboration with the *Arabidopsis* SSP group in the United States (Yamada et al. 2003), which comprises the Salk Institute [principal investigator (PI): J. R. Ecker], the Stanford Genome Technology Center (PI: R. W. Davis) and the Plant Gene Expression Center (PI: A. Theologis), and the Japanese group (K. Hanada et al., unpublished results), which comprises the RIKEN BioResource Center (BRC) (PI: M. Kobayashi), the National Institute of Genetics (PI: Y. Kohara) and the Genome Core Technology Facilities of RIKEN Genomic Sciences Center (GSC) (PI: Y. Sakaki). The RAFL cDNA clones are publicly available from the RIKEN BRC (<http://www.brc.riken.go.jp/lab/epd/Eng/>).

Application of full-length cDNAs to genome sequence annotation

Dramatic improvements in *Arabidopsis* genome sequence annotation have been achieved by mapping of RAFL cDNA sequences to the *Arabidopsis* genome (Seki et al. 2002a; Yamada et al. 2003).

Genome-wide analysis of alternative splicing events in *Arabidopsis* found that more than 4,700 transcribed premature mRNAs were alternatively spliced (Iida et al. 2004; Wang and Brendel 2006). Iida et al. (2004) found that the pattern of alternative splicing events was affected by cold stress conditions. Recent full-length sequencing analysis of 1,800 RAFL cDNAs that had 5'- and/or 3'- sequences previously found to have alternative splicing events or alternative transcription start sites revealed the presence of 601 novel alternatively spliced/structure variant transcripts in *Arabidopsis* (Iida et al. 2009).

More than 1,000 overlapping sense-antisense (SAT) pairs have been identified by a genome-wide search of *Arabidopsis* cDNAs (Seki et al. 2004b; Jen et al. 2005; Wang et al. 2005). Antisense RNAs have been believed to control expression of sense transcripts negatively in plants (Borsani et al. 2005). Recently, we identified about 8,000 SAT pairs via *Arabidopsis* tiling array analysis under abiotic stresses. Many non-protein coding transcripts were found to belong to SAT transcripts, and the expression ratios (treated/untreated) of sense transcripts and the ratios of antisense transcripts showed a significant linear correlation (Matsui et al. 2008). Antisense RNAs have been shown to participate in a broad range of types of regulation,

such as gene silencing, RNA stability, RNA editing, RNA masking, and methylation. Our recent tiling array analysis also demonstrated that several non-protein-coding antisense RNAs are suppressed by the nonsense-mediated mRNA decay (NMD) pathway (Kurihara et al. 2009).

One significant class of genes missing from the existing genome annotation is non-protein-coding RNAs. In addition to their role in protein synthesis (ribosomal and transfer RNAs), non-protein-coding RNAs have been implicated in control processes such as chromosomal silencing, transcriptional regulation, developmental control, and responses to stress (MacIntosh et al. 2001). Recently, we identified about 7,000 putative non-protein-coding RNAs in unannotated intergenic regions using an *Arabidopsis* Affymetrix tiling array (Matsui et al. 2008). These include non-protein-coding RNAs present 5'-upstream and 3'-downstream of AGI code genes. Interestingly, 27 promoter-associated short RNA (PASR)-like transcriptional units (TUs) (Kapranov et al. 2007) and 27 terminator-associated short RNA (TASR)-like TUs (Kapranov et al. 2007), which are supported by full-length cDNAs, have been identified in the tiling array analysis (Fig. 2b, d; Matsui et al. 2008). Eight PASR-like TUs and ten TASR-like TUs that are supported by full-length cDNAs show ABA- or stress-responsive gene expression. Martianov et al. (2007) demonstrated that a non-protein-coding transcript upstream of the human *dihydrofolate reductase* (*DHFR*) gene has a critical function in transcriptional repression of the *DHFR* gene. Several novel PASR-like TUs in 5'-upstream regions might act as negative regulators of the downstream main TUs. Our tiling array analysis also showed that the 5'- and 3'-end regions of the 67 and 34 AGI code genes (Fig. 2a, c; Matsui et al. 2008), respectively, are shorter in a previous TAIR6 gene model than the gene model of the AGI code genes detected by the "ARTADE" (*Arabidopsis* tiling array-based detection of exons) program (Toyoda and Shinozaki 2005), as also supported by full-length cDNAs. These results show that the tiling array is also a useful tool for improvement of genome sequence annotation.

RAFL cDNA microarray analysis

cDNA microarrays are a powerful tool for the systematic analysis of expression profiles of large numbers of genes, including stress-inducible gene expression and changes in the expression profiles of mutants or transgenics (Seki et al. 2004a). One interesting type of application of microarray analysis is the identification of novel cis-elements that regulate the expression of genes in response to various experimental treatments (Simpson et al. 2003). By identifying subsets of the genes that have a common expression

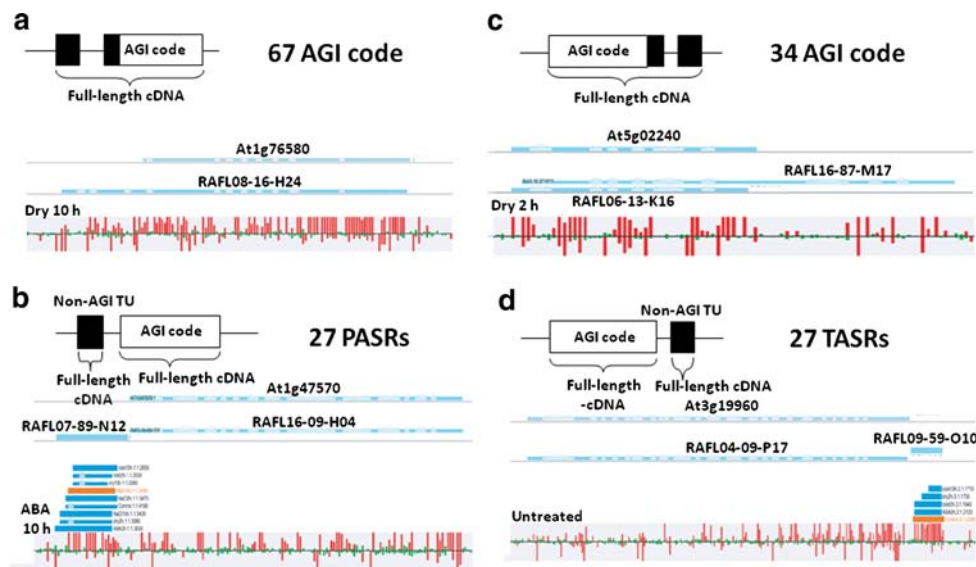


Fig. 2 New gene models supported by full-length cDNAs and tiling array analysis. **a, c** *Arabidopsis* genome initiative (AGI) code genes whose 5'-end- (**a**) or 3'-end- (**c**) regions are short in the TAIR 6 gene model. **b** Promoter-associated short RNA (PASR)-like transcriptional units (TUs). The full-length cDNA sequences support the non-AGI TUs mapped on the promoter region of the AGI code genes.

d Termini-associated short RNA (TASR)-like TUs. The full-length cDNA sequences support the non-AGI TUs mapped around the 3'-termini of the AGI code genes. The *lower panels* represent examples of the tiling array expression data supporting the new gene models (Matsui et al. 2008; <http://omicspace.riken.jp/gps/group/psca1>)

profile, it might be possible to identify conserved motifs in promoter regions. For example, promoter databases have been used for systematic analysis of cis-acting elements in *Arabidopsis* (Fig. 1).

We prepared the following two types of cDNA microarray: (1) a 1.3 K RAFL cDNA microarray (Seki et al. 2001a) containing about 1,300 RAFL cDNA clones, and (2) a 7K RAFL cDNA microarray (Seki et al. 2002b) containing about 7,000 RAFL cDNA clones. Using these cDNA microarrays, we have studied the expression profiles of *Arabidopsis* genes under various stress conditions (Fig. 1), such as drought, cold, and high-salinity-stresses (Seki et al. 2001a, 2002b), and high light stress (Kimura et al. 2003), as well as various treatment conditions, such as abscisic acid (ABA) (Seki et al. 2002c), rehydration treatment after dehydration (Oono et al. 2003), ethylene (Narusaka et al. 2003), jasmonic acid (JA) (Narusaka et al. 2003), salicylic acid (SA) (Narusaka et al. 2003), reactive oxygen species (ROS)-inducing compounds such as paraquat and rose bengal (Narusaka et al. 2003), UV-C (Narusaka et al. 2003), proline (Pro) (Satoh et al. 2002), and inoculation with pathogen (Narusaka et al. 2003). We have also studied expression profiles in various mutants and transgenic plants (Fig. 1; Seki et al. 2001a; Osakabe et al. 2002; Abe et al. 2003; Dubouzet et al. 2003; Nanjo et al. 2003; Chini et al. 2004; Kamei et al. 2005; Noutoshi et al. 2005; Osakabe et al. 2005). Note that various types of oligonucleotide DNA microarrays that are available from Affymetrix (http://www.affymetrix.com/products_services/

[index.affx#1_1](http://www.chem.agilent.com/en-US/products/instruments/dnamicromicroarrays/Pages/default.aspx)) and Agilent Technologies (<http://www.chem.agilent.com/en-US/products/instruments/dnamicromicroarrays/Pages/default.aspx>) etc. have been widely used recently instead of cDNA microarrays. This might be due to the fact that oligonucleotide DNA microarrays are superior to cDNA microarrays in terms of the number of genes that are contained on each microarray, as well as easier management of the microarray system. The oligonucleotide microarrays have been prepared using sequence information from the updated gene models of the *Arabidopsis* genome. Expression profiling studies using these microarrays have shown the expression levels of many genes as a detailed snapshot describing the state of a biological system in plants under certain conditions.

Identification of genes regulated by drought, cold, high-salinity-stress or abscisic acid

Plant growth is affected greatly by environmental abiotic stresses, such as drought, high salinity, and low temperature. Plants respond and adapt to these stresses in order to survive. These stresses induce various biochemical and physiological responses in plants. Several thousand genes have been identified that respond to drought, high-salinity or cold stress at the transcriptional level (Thomashow 1999; Hasegawa et al. 2000; Seki et al. 2002b; Zhu 2002; Matsui et al. 2008). It is important to study the function of stress-inducible genes not only to understand the molecular

mechanisms of stress tolerance and responses in plants but also to improve stress tolerance by genetic engineering. Stress-inducible genes have been used to improve the stress tolerance of plants by gene transfer (Thomashow 1999; Hasegawa et al. 2000; Shinozaki and Yamaguchi-Shinozaki 2000).

Several years ago, we prepared a full-length cDNA microarray (7K RAFL cDNA microarray) containing ca. 7,000 independent *Arabidopsis* full-length cDNA groups (Seki et al. 2002b), and applied the 7K RAFL cDNA microarray to identify new drought-, cold-, high-salinity- or abscisic acid (ABA)-inducible genes. We identified 299 drought-inducible genes, 54 cold-inducible genes, 213 high-salinity-stress-inducible genes and 245 ABA-inducible genes (Seki et al. 2002b, c). Venn diagram analysis indicated the existence of significant crosstalk between drought and high-salinity stress signaling processes (Seki et al. 2002b). Many ABA-inducible genes are induced after drought- and high-salinity-stress treatments, which indicates the existence of significant crosstalk between drought and ABA responses (Seki et al. 2002c). These results indicate the presence of strong overlaps of gene expression in response to drought, high-salinity, and ABA (Shinozaki and Yamaguchi-Shinozaki 2000), and partial overlap of gene expression in response to cold and osmotic stress.

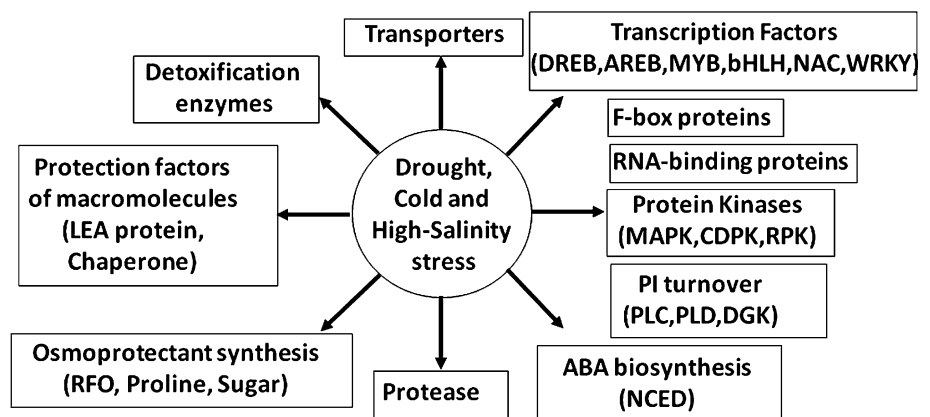
The products of the drought-, high-salinity- or cold-stress-inducible gene products can be classified into two groups (Fig. 3; Shinozaki and Yamaguchi-Shinozaki 2000; Seki et al. 2002b). The first group includes functional proteins, or proteins that probably function in stress tolerance. They are late-embryogenesis abundant (LEA) proteins, heat shock proteins, KIN (cold-inducible) proteins, osmoprotectant-biosynthesis-related proteins, carbohydrate-metabolism-related proteins, water channel proteins, sugar transporters, potassium transporters, detoxification enzymes, proteases, senescence-related proteins, protease inhibitors, ferritin, and lipid transfer proteins (Seki et al. 2002b).

The second group contains regulatory proteins, that is, protein factors involved in further regulation of signal transduction and gene expression that probably function in the response to stress (Shinozaki and Yamaguchi-Shinozaki 2000; Seki et al. 2002b, c). These include various transcription factors, protein kinases, protein phosphatases, enzymes involved in phospholipid metabolism, and other signaling molecules such as calmodulin-binding protein (Seki et al. 2002b, c). We identified many stress-inducible transcription factor (TF) genes, such as dehydration-responsive element (DRE)-binding protein (DREB), ethylene-responsive element binding factor (ERF), zinc finger, WRKY, MYB, basic helix-loop-helix (bHLH), bZIP, NAC and homeodomain-leucine zipper (HD-ZIP) TF genes, suggesting that various transcriptional regulatory mechanisms function in the drought-, cold- or high-salinity-stress signal transduction pathways (Seki et al. 2002b, c). These transcription factors probably regulate various stress-inducible genes cooperatively or separately.

Identification of candidate genes regulated by stress-inducible transcription factors

Transcriptional activation of some stress-responsive genes, such as the *RD29A/COR78/LTI78* gene (responsive to dehydration/cold-regulated/low-temperature-induced) is well understood. The promoter of this gene contains both an ABRE (abscisic acid-responsive element) and a DRE/CRT (dehydration responsive element/C-repeat) (Yamaguchi-Shinozaki and Shinozaki 2005, 2006). ABRE and DRE/CRT are cis-acting elements that function in ABA-dependent and ABA-independent gene expression in response to stress, respectively. Transcription factors belonging to the ERF/AP2 (ethylene-responsive element binding factor/apetala 2) family that bind to DRE/CRT were isolated and termed DREB1/CBF (DRE-binding protein 1/C-repeat-binding factor) and DREB2 (Yamaguchi-Shinozaki and Shinozaki

Fig. 3 Drought-, cold- and high-salinity-stress-inducible genes and their possible functions in stress tolerance and response



2005, 2006). Their conserved DNA-binding motif is A/GCCGAC. The DREB1/CBF genes are rapidly and transiently induced in response to cold stress, and these transcription factors in turn activate the expression of target genes.

Overexpression of the *DREB1A/CBF3* cDNA under the control of the cauliflower mosaic virus (CaMV) 35S promoter or the stress-inducible *rd29A* promoter in transgenic plants gave rise to strong constitutive expression of stress-inducible DREB1A target genes, and increased tolerance to freezing and drought stresses (Jaglo-Ottosen et al. 1998; Liu et al. 1998; Kasuga et al. 1999). Kasuga et al. (1999) identified six DREB1A target genes. However, it remains poorly understood how overexpression of the *DREB1A* cDNA in transgenic plants increases stress tolerance to freezing and drought stresses. We applied the RAFL cDNA microarrays to identify new target genes of *DREB1A* (Seki et al. 2001a; Maruyama et al. 2004) and identified more than 40 DREB1A target genes. The downstream target genes include C2H2 zinc-finger-type- and ERF/AP2-type-TFs, RNA-binding proteins, sugar transport proteins, LEA proteins, KIN proteins, RFO (raffinose family oligosaccharides)-biosynthesis-related proteins, and protease inhibitors. Conserved sequences in the promoter regions of the DREB1A/CBF3 target genes were searched, and A/GCCGACNT was found in their promoter regions between –51 and –450 as a consensus DRE (Maruyama et al. 2004). These results showed that the DNA microarray is a useful system with which to identify target genes of stress-related transcription factors and potential cis-acting DNA elements by combining expression data with genomic sequence data.

We have also applied the RAFL cDNA microarray to identify the target genes of the following stress-related transcription factors: ERF/AP2 TF family, e.g., DREB2A (Sakuma et al. 2006); bZIP TF family, e.g., AREB1 (Fujita et al. 2005); MYB TF family, e.g., AtMYB2 (Abe et al. 2003); bHLH TF family, e.g., AtMYC2 (Abe et al. 2003); NAC TF family, e.g., RD26/ANAC072 (Fujita et al. 2004; Tran et al. 2004), ANAC019 (Tran et al. 2004) and ANAC055 (Tran et al. 2004). The roles of TFs in the abiotic stress signaling and the expression profiling results are summarized in recent reviews (Bartels and Sunkar 2005; Seki et al. 2005; Yamaguchi-Shinozaki and Shinozaki 2005, 2006). Information on the target genes is useful for understanding the transcriptional regulatory networks involved in cellular responses to abiotic stresses.

Transcriptome analysis in the recovery process following stress

Analysis of genes involved in the recovery from stress as well as stress-inducible genes is also important, not only

for the understanding of the molecular responses to abiotic stresses but also for improving the stress tolerance of crops by gene manipulation. Oono et al. (2003) applied the analysis of the 7K RAFL cDNA microarray to the identification of genes that are induced during the rehydration process after dehydration stress treatment, and identified 152 rehydration-inducible genes. These genes can be classified into the following three major groups: (1) regulatory proteins involved in further regulation of signal transduction and gene expression, (2) functional proteins involved in the recovery process after dehydration-induced damage, and (3) functional proteins involved in plant growth (Oono et al. 2003). Venn diagram analysis also showed that among the rehydration-inducible genes, at least two gene groups existed, i.e., genes functioning in adjustment of cellular osmotic conditions and those functioning in the repair of drought-stress-induced damage, and that most of the rehydration-downregulated genes are dehydration-inducible (Oono et al. 2003).

Oono et al. (2006) analyzed the gene expression profiles in the process of cold acclimation and deacclimation (recovery from cold stress) using two microarray systems: the 7K RAFL cDNA microarray and the Agilent 22K oligonucleotide array. Both microarray analyses identified 292 genes up-regulated and 320 genes down-regulated during deacclimation, and 445 cold up-regulated genes and 341 cold down-regulated genes during cold acclimation. Many genes up-regulated during deacclimation were found to be down-regulated during cold acclimation, and vice versa.

Application of RAFL cDNAs to functional analysis of proteins

Endo's group at Ehime University has established an efficient wheat germ cell-free protein synthesis system to produce milligram quantities of proteins (Madin et al. 2000; Sawasaki et al. 2002). We have applied the wheat germ cell-free protein synthesis system using the RAFL cDNAs to study the functional characteristics of *Arabidopsis* proteins (Fig. 1). Sawasaki et al. (2004) used about 400 RAFL cDNAs encoding protein kinases for protein synthesis using the wheat germ system. The assay revealed about 200 products with autophosphorylation activity. Seven proteins out of 26 calcium-dependent protein kinases phosphorylated a synthetic peptide substrate in the presence of calcium ions, demonstrating that the translation products retained their substrate specificity. Recently, we also demonstrated that the wheat germ system is useful for studying the biochemical characteristics of TFs (T. Sawasaki et al. unpublished results) and proteins involved in ubiquitination (Takahashi et al. 2009).

We also applied the RAFL cDNAs to a gain-of-function gene hunting technique—full-length cDNA over-expressing (FOX) gene hunting system (Ichikawa et al. 2006; Fig. 1)—that involves the random overexpression of a normalized full-length cDNA library. Ichikawa et al. (2006) introduced about 10,000 independent RAFL cDNAs under the control of the CaMV 35S promoter into *Arabidopsis*, and found about 1,500 possible morphological mutants with various phenotypes, e.g., pale green, dwarf, and bushy phenotypes, from about 15,000 transformants. Fujita et al. (2007) focused on stress-inducible TFs; the full-length cDNAs of 43 stress-inducible TFs were mixed to create the FOX lines. After screening for salt-stress-resistant lines, a number of salt-tolerant lines were found to harbor a bZIP-type transcription factor (*AtbZIP60*) involved in the endoplasmic reticulum stress response. Full-length rice cDNAs have been introduced into *Arabidopsis* (Kondou et al. 2009) and rice (Nakamura et al. 2007) plants using the FOX system, and many FOX lines showing altered growth or morphological characteristics, such as super-dwarf mutants, have been obtained. These results demonstrate that the FOX system is a useful method to screen for genes with valuable functions. Ogawa et al. (2008) introduced 96 metabolism-related RAFL cDNAs into *Arabidopsis* suspension-cultured T87 cells by *Agrobacterium*-mediated transformation to study the plant metabolome.

ORFeome clones, that is, cDNA clones containing full-length open reading frames (ORFs) are a valuable research tool for functional proteomics. In collaboration with the *Arabidopsis* SSP group in the United States (Yamada et al. 2003), we used the RAFL cDNAs to construct the U (pUNI) clone, an *Arabidopsis* ORFeome clone. We have constructed about 10,500 U clones and determined full-length sequences of the intact ORF regions for confirmation of error-free ORFeome clones. The U clones are publicly available from the *Arabidopsis* Biological Resource Center (ABRC; <http://www.biosci.ohio-state.edu/~plantbio/Facilities/abrc/abrchome.htm>). Several groups, such as the Yale group (Gong et al. 2004), also have constructed various *Arabidopsis* ORFeome clones. Information on ORFeome clones is available at http://www.arabidopsis.org/portals/masc/ORFeomics_2008Report.pdf.

Application of RAFL cDNAs to structural analysis of proteins

Full-length cDNAs are useful resources for determining the three-dimensional structures of proteins by X-ray crystallography and NMR spectroscopy (Seki et al. 2001b) (Fig. 1). We have determined the three-dimensional structures of plant proteins using *Arabidopsis* full-length

cDNAs by NMR spectroscopy in the RIKEN Structural Genomics Initiative (Yokoyama et al. 2000), using cell-free protein synthesis systems for protein expression. Cell-free in vitro systems have three advantages over conventional in vivo expression systems: (1) cell-free systems are suitable for automated, high-throughput expression, as proteins can be produced without the need for cloning genes into expression vectors; (2) milligram quantities of proteins can be obtained in several hours; and (3) proteins that are difficult to express in vivo can be produced in vitro.

We have applied this system to plant protein expression and determined the domain structure of 29 proteins containing plant-specific-type TFs, such as the DNA-binding domain of squamosa promoter-binding protein (Yamasaki et al. 2004b, 2006), the B3 DNA-binding domain of the cold-responsive transcription factor RAV1 (RAV for related to ABI3/VP1) (Yamasaki et al. 2004a), the C-terminal WRKY domain of the WRKY4 protein (Yamasaki et al. 2005a), and the DNA-binding domain of an ethylene-insensitive 3 (EIN3) protein, EIL3 (Yamasaki et al. 2005b). Determination of the three-dimensional structure of the DNA-binding domains of stress-inducible TFs might be applied to alter the target genes for improvement of stress tolerance.

Conclusions and future perspectives

We have demonstrated that full-length cDNAs are an important resource for improvement of genome sequence annotation, expression profiling studies, and functional and structural analysis of plant proteins in the post-sequencing era. The RAFL cDNA clones are publicly available from the RIKEN BRC, and the BRC has distributed about 24,000 RAFL cDNAs to more than 500 labs as of 22 October 2008. The RAFL cDNAs are a standard resource in the *Arabidopsis* research community.

Once all the *Arabidopsis* full-length cDNAs are functionally annotated, the database will be the Rosetta Stone for understanding the network of gene functions in higher plants (Appels et al. 2003; Clarke et al. 2003). Information gained from these full-length cDNAs can be applied to other crops, such as rice (Kikuchi et al. 2003), wheat (Ogihara et al. 2004), barley (Sato et al. 2009), soybean (Umezawa et al. 2008), and cassava (Sakurai et al. 2007), to trees, such as poplar (Nanjo et al. 2007), *Cryptomeria japonica* (Futamura et al. 2008), and *Sitka spruce* (Ralph et al. 2008), and to model plants, such as *Physcomitrella patens* (Nishiyama et al. 2003) and *Thellungiella halophila* (Taji et al. 2008). Full-length cDNAs will be used for improvement of plants in the future.

Recently, the SABRE (systematic consolidation of *Arabidopsis* and other botanical resource) database

(<http://saber.epd.brc.riken.jp/sabre/SABRE0101.cgi>) has been developed to provide organized information on plant full-length cDNA resources that are available from RIKEN BRC. The SABRE database will help researchers access counterpart full-length cDNA resources in other plant species for basic and applied science.

Complete genome sequences of various plant species, such as rice (International Rice Genome Sequencing Project 2005), poplar (Tuskan et al. 2006), grapevine (The French-Italian Public Consortium for Grapevine Genome Characterization 2007) and *Physcomitrella* (Rensing et al. 2008) have been determined. Recently, next generation sequencers that are available from 454 Life Sciences (<http://www.454.com/>; Margulies et al. 2005), Illumina (<http://www.illumina.com/>) and Applied Biosystems (<http://www.appliedbiosystems.com/>) have been applied to whole-genome sequencing in various plant species, and to the identification of whole-genome sequence variation in 1,001 natural strains of *Arabidopsis* (<http://1001genomes.org>; Ossowski et al. 2008). Paired-end diTag (PET; Ruan et al. 2007) analysis, a useful method to characterize both ends of DNA fragments, using such next generation sequencers, will be applied to the analysis of the full-length cDNAs of many plant species and strains in the future, and will contribute to the discovery of useful genes and our understanding of natural variation and evolution in plants.

Acknowledgments We would like to thank our many collaborators for their valuable participation in this work. This work was supported by a grant for Genome Research from RIKEN, the Program for Promotion of Basic Research Activities for Innovative Biosciences, the Special Coordination Fund of the Science and Technology Agency, and a Grant-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan to K.S. It was also supported by a Grant-in-Aid for Scientific Research on Priority Areas (C) “Genome Science” from the Ministry of Education, Culture, Sports, Science and Technology of Japan to M.S.

References

- Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell* 15:63–78
- Appels R, Francki M, Chibbar R (2003) Advances in cereal functional genomics. *Funct Integr Genomics* 3:1–24
- Asamizu E, Nakamura Y, Sato S, Tabata S (2000) A large scale analysis of cDNA in *Arabidopsis thaliana*: generation of 12, 028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries. *DNA Res* 7:175–180
- Bartels D, Sunkar R (2005) Drought and salt tolerance in plants. *Crit Rev Plant Sci* 24:23–58
- Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* 123:1279–1291
- Carninci P, Kvam C, Kitamura A, Ohsumi T, Okazaki Y, Itoh M, Kamiya M, Shibata K, Sasaki N, Izawa M, Muramatsu M, Hayashizaki Y, Schneider C (1996) High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics* 37:327–336
- Carninci P, Westover A, Nishiyama Y, Ohsumi T, Itoh M, Nagaoka S, Sasaki N, Okazaki Y, Muramatsu M, Schneider C, Hayashizaki Y (1997) High efficiency selection of full-length cDNA by improved biotinylated CAP trapper. *DNA Res* 4:61–66
- Carninci P, Nishiyama Y, Westover A, Itoh M, Nagaoka S, Sasaki N, Okazaki Y, Muramatsu M, Hayashizaki Y (1998) Thermoactivation and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. *Proc Natl Acad Sci USA* 95:520–524
- Castelli V, Aury JM, Jaillon O, Wincker P, Clepet C, Menard M, Cruaud C, Quetier F, Scarpelli C, Schachter V, Temple G, Caboche M, Weissenbach J, Salanoubat M (2004) Whole genome sequence comparisons and “full-length” cDNA sequences: a combined approach to evaluate and improve *Arabidopsis* genome annotation. *Genome Res* 14:406–413
- Chini A, Grant J, Seki M, Shinozaki K, Loake G (2004) Drought tolerance established by enhanced expression of the CC-NBS-LRR gene, ADR1, requires salicylic acid, EDS1 and ABI1. *Plant J* 38:810–822
- Clarke B, Lambrecht M, Rhee SY (2003) *Arabidopsis* genomic information for interpreting wheat EST sequences. *Funct Integr Genomics* 3:33–38
- Cooke R, Raynal M, Laudie M, Grellet F, Delseny M, Morris PC, Guerrier D, Giraudat J, Quigley F, Clabault G, Li YF, Mache R, Krivitzky M, Gy IJ, Kreis M, Lecharny A, Parmentier Y, Marbach J, Fleck J, Clément B, Philipps G, Hervé C, Bardet C, Tremoussaygue D, Lescure B, Lacomme C, Roby D, Jourjon MF, Chabrier P, Charpentreau JL, Desprez T, Amselem J, Chiapello H, Höfte H (1996) Further progress towards a catalogue of all *Arabidopsis* genes: analysis of a set of 5, 000 non-redundant ESTs. *Plant J* 9:101–124
- Davuluri RV, Sun H, Palaniswamy SK, Matthews N, Molina C, Kurtz M, Grotewold E (2003) AGRIS: *Arabidopsis* gene regulatory information server, an information resource of *Arabidopsis* cis-regulatory elements and transcription factors. *BMC Bioinformatics* 4:25
- Dubouzet JG, Sakuma Y, Ito Y, Kasuga M, Dubouzet EG, Miura S, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) OsDREB genes in rice, *Oryza sativa* L, encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. *Plant J* 33:751–763
- Fujita M, Fujita Y, Maruyama K, Seki M, Hiratsu K, Ohme-Takagi M, Tran LSP, Yamaguchi-Shinozaki K, Shinozaki K (2004) A dehydration-induced NAC protein, RD26 is involved in ABA-dependent stress signaling pathway. *Plant J* 39:863–876
- Fujita Y, Fujita M, Satoh R, Maruyama K, Parvez MM, Seki M, Hiratsu K, Ohme-Takagi M, Shinozaki K, Yamaguchi-Shinozaki K (2005) AREB1 is a transcription activator of novel ABRE-dependent ABA-signaling that enhances drought stress tolerance in *Arabidopsis*. *Plant Cell* 17:3470–3488
- Fujita M, Mizukado S, Fujita Y, Ichikawa T, Nakazawa M, Seki M, Matsui M, Yamaguchi-Shinozaki K, Shinozaki K (2007) Identification of stress-tolerance-related transcription-factor genes via mini-scale full-length cDNA over-expressor (FOX) gene hunting system. *Biochem Biophys Res Commun* 364:250–257
- Futamura N, Totoki Y, Toyoda A, Igasakai T, Nanjo T, Seki M, Sakaki Y, Mari A, Shinozaki K, Shinohara K (2008) Characterization of expressed sequence tags from a full-length enriched cDNA library of *Cryptomeria japonica* male strobili. *BMC Genomics* 9:383
- Gong W, Shen YP, Ma LG, Pan Y, Du YL, Wang DH, Yang JY, Hu LD, Liu XF, Dong CX, Ma L, Chen YH, Yang XY, Gao Y, Zhu D, Tan X, Mu JY, Zhang DB, Liu YL, Dinesh-Kumar SP, Li Y,

- Wang XP, Gu HY, Qu LJ, Bai SN, Lu YT, Li JY, Zhao JD, Zuo J, Huang H, Deng XW, Zhu YX (2004) Genome-wide ORFeome cloning and analysis of *Arabidopsis* transcription factor genes. *Plant Physiol* 135:773–782
- Haas BJ, Volfovsky N, Town CD, Troukhan M, Alexandrov N, Feldmann KA, Flavell RB, White O, Salzberg SL (2002) Full-length messenger RNA sequences greatly improve genome annotation. *Genome Biol* 3:1–12
- Hasegawa PM, Bressan RA, Zhu JK, Bohnert HJ (2000) Plant cellular and molecular responses to high salinity. *Annu Rev Plant Physiol Plant Mol Biol* 51:463–499
- Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database. *Nucleic Acids Res* 27:297–300
- Höfte H, Desprez T, Amselem J, Chiapello H, Caboche M, Moisan A, Jourjon MF, Charpentreau JL, Berthomieu P, Guerrier D, Giraudat J, Quigley F, Thomas F, Yu DY, Mache R, Raynal M, Cooke M, Grellet F, Delseny M, Parmentier Y, Marcillac GD, Gigot C, Fleck J, Philipps G, Axelos M, Bardet C, Tremousaygue D, Lescure B (1993) An inventory of 1,152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. *Plant J* 4:1051–1061
- Ichikawa T, Nakazawa M, Kawashima M, Iizumi H, Kuroda H, Kondou Y, Tshuhara Y, Suzuki K, Ishikawa A, Seki M, Fujita M, Motohashi R, Nagata N, Takagi T, Shinozaki K, Matsui M (2006) The FOX hunting system: an alternative gain-of-function gene hunting technique. *Plant J* 45:974–985
- Iida K, Seki M, Sakurai T, Satou M, Akiyama K, Toyoda T, Konagaya A, Shinozaki K (2004) Genome-wide analysis of alternative pre-mRNA splicing in *Arabidopsis thaliana* based on full-length cDNA sequences. *Nucleic Acids Res* 32:5096–5103
- Iida K, Fukami-Kobayashi K, Toyoda A, Sakaki Y, Kobayashi M, Seki M, Shinozaki K (2009) Analysis of multiple occurrences of alternative splicing events in *Arabidopsis thaliana* using novel sequenced full-length cDNAs. *DNA Res* (in press)
- International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* 436:793–800
- Jaglo-Ottosen KR, Gilmour SJ, Zarka DG, Schabender O, Thomashow MF (1998) *Arabidopsis* CBF1 overexpression induces *cor* genes and enhances freezing tolerance. *Science* 280:104–106
- Jen CH, Michalopoulos I, Westhead DR, Meyer P (2005) Natural antisense transcripts with coding capacity in *Arabidopsis* may have a regulatory role that is not linked to double-stranded RNA degradation. *Genome Biol* 6:R51
- Kamei A, Seki M, Umezawa T, Ishida J, Satou M, Akiyama K, Zhu JK, Shinozaki K (2005) Analysis of gene expression profiles in *Arabidopsis* salt overly sensitive mutants, *sos2* and *sos3* mutants. *Plant Cell Environ* 28:1267–1275
- Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H, Gingeras TR (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 316:1484–1488
- Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K, Shinozaki K (1999) Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. *Nat Biotechnol* 17:287–291
- Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H, Hotta I, Kojima K, Namiki T, Ohneda E, Yahagi W, Suzuki K, Li CJ, Ohtsuki K, Shishiki T, Otomo Y, Murakami K, Iida Y, Sugano S, Fujimura T, Suzuki Y, Tsunoda U, Kurosaki T, Kodama T, Masuda H, Kobayashi M, Xie Q, Lu M, Narikawa R, Sugiyama A, Mizuno K, Yokomizo S, Niikura J, Ikeda R, Ishibiki J, Kawamata M, Yoshimura A, Miura J, Kusumegi T, Oka M, Ryu R, Ueda M, Matsubara K, Kawai J, Carninci P, Adachi J, Aizawa K, Arakawa T, Fukuda S, Hara A, Hashidume W, Hayatsu N, Imotani K, Ishii Y, Itoh M, Kagawa I, Kondo S, Konno H, Miyazaki A, Osato N, Ota Y, Saito R, Sasaki D, Sato K, Shibata K, Shinagawa A, Shiraki T, Yoshino M, Hayashizaki Y (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. *Science* 301:376–379
- Kimura M, Yamamoto YY, Seki M, Sakurai T, Satou M, Abe T, Yoshida S, Manabe K, Shinozaki K, Matsui M (2003) Identification of *Arabidopsis* genes regulated by high light stress using cDNA microarray. *Photochem Photobiol* 77:226–233
- Kondou Y, Higuchi M, Takahashi S, Sakurai T, Ichikawa T, Kuroda H, Yoshizumi T, Tsumoto Y, Horii Y, Kawashima M, Hasegawa Y, Kuriyama T, Matsui K, Kusano M, Albinsky D, Takahashi H, Nakamura Y, Suzuki M, Sakakibara H, Kojima M, Akiyama K, Kurotani A, Seki M, Fujita M, Enju A, Yokotani N, Saitou T, Ashidate K, Fujimoto N, Ishikawa Y, Mori Y, Nanba R, Takata K, Uno K, Sugano S, Natsuki J, Dubouzet JG, Maeda S, Ohtake M, Mori M, Oda K, Takatsuki H, Hirochika H, Matsui M (2009) Systematic approaches to using the FOX hunting system to identify useful rice genes. *Plant J* 57:883–894
- Kurihara Y, Matsui A, Hanada K, Kawashima M, Ishida J, Morosawa T, Tanaka M, Kaminuma E, Mochizuki Y, Matsushima A, Toyoda T, Shinozaki K, Seki M (2009) Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in *Arabidopsis*. *Proc Natl Acad Sci USA* 106:2453–2458
- Liu Q, Kasuga M, Sakuma Y, Abe H, Miura S, Yamaguchi-Shinozaki K, Shinozaki K (1998) The transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. *Plant Cell* 10:1391–1406
- MacIntosh GC, Wilkerson C, Green PJ (2001) Identification and analysis of *Arabidopsis* expressed sequence tags characteristics of non-coding RNAs. *Plant Physiol* 127:765–776
- Madin K, Sawasaki T, Ogasawara T, Endo Y (2000) A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. *Proc Natl Acad Sci USA* 97:559–564
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP,irage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380
- Martianov I, Ramadass A, Barros AS, Chow N, Akoulitchev A (2007) Repression of the human dihydrofolate reductase gene by a noncoding interfering transcript. *Nature* 445:666–670
- Maruyama K, Sakuma Y, Kasuga M, Ito Y, Seki M, Goda H, Shimada Y, Yoshida S, Shinozaki K, Yamaguchi-Shinozaki K (2004) Identification of cold-inducible downstream genes of the *Arabidopsis* DREB1A/CBF3 transcriptional factor using two microarray systems. *Plant J* 38:982–993
- Matsui A, Ishida J, Morosawa T, Mochizuki Y, Kaminuma E, Endo TA, Okamoto M, Nambara E, Nakajima M, Kawashima M, Satou M, Kim JM, Kobayashi N, Toyoda T, Shinozaki K, Seki M (2008) *Arabidopsis* transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using a tiling array. *Plant Cell Physiol* 49:1135–1149

- Meinke DW, Cherry JM, Dean CD, Rounsley S, Koornneef M (1998) *Arabidopsis thaliana*: a model plant for genome analysis. *Science* 282:662–682
- Nakamura H, Hakata M, Amano K, Miyao A, Toki N, Kajikawa M, Pang J, Higashi N, Ando S, Toki S, Fujita M, Enju A, Seki M, Nakazawa M, Ichikawa T, Shinozaki K, Matsui M, Nagamura Y, Hirochika H, Ichikawa H (2007) A genome-wide gain-of-function analysis of rice genes using the FOX-hunting system. *Plant Mol Biol* 65:357–371
- Nanjo T, Fujita M, Seki M, Kato T, Tabata S, Shinozaki K (2003) Toxicity of free proline revealed in an *Arabidopsis* T-DNA-tagged mutant deficient in proline dehydrogenase. *Plant Cell Physiol* 44:541–548
- Nanjo T, Sakurai T, Totoki Y, Toyoda A, Nishiguchi M, Kado T, Igasa N, Futamura N, Seki M, Sakaki Y, Shinozaki K, Shinohara K (2007) Functional annotation of 19,841 *Populus nigra* full-length enriched cDNA clones. *BMC Genomics* 8:448
- Narusaka Y, Narusaka M, Seki M, Ishida J, Nakashima M, Kamiya A, Enju A, Sakurai T, Satou M, Kobayashi M, Tosa Y, Park P, Shinozaki K (2003) The cDNA microarray analysis using an *Arabidopsis pad3* mutant reveals the expression profiles and classification of genes induced by *Alternaria brassicicola* attack. *Plant Cell Physiol* 44:377–387
- Newman T, DeBruijn FJ, Green P, Keegstra K, Kende H, McIntosh L, Ohlrogge J, Raikhel N, Somerville S, Thomashow M, Retzel E, Somerville C (1994) Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. *Plant Physiol* 106:1241–1255
- Nishiyama T, Fujita T, Shin T, Seki M, Nishide H, Uchiyama I, Kamiya A, Carninci P, Hayashizaki Y, Shinozaki K, Kohara Y, Hasebe M (2003) Comparative genomics of the *Physcomitrella* gametophytic transcriptome and *Arabidopsis* genome: implication for the land plant evolution. *Proc Natl Acad Sci USA* 100:8007–8012
- Noutoshi Y, Ito T, Seki M, Nakashita H, Yoshida S, Marco Y, Shirasu K, Shinozaki K (2005) A single amino acid insertion in the WRKY domain of the *Arabidopsis* TIR-NBS-LRR-WRKY type disease resistance protein SLH1 (SENSITIVE TO LOW HUMIDITY 1) causes activation of defense responses and hypersensitive cell death. *Plant J* 43:873–888
- Ogawa Y, Dansako T, Yano K, Sakurai N, Suzuki H, Aoki K, Noji M, Saito K, Shibata D (2008) Efficient and high-throughput vector construction and *Agrobacterium*-mediated transformation of *Arabidopsis thaliana* suspension-cultured cells for functional genomics. *Plant Cell Physiol* 49:242–250
- Ogihara Y, Mochida K, Kawaura K, Murai K, Seki M, Kamiya A, Shinozaki K, Carninci P, Hayashizaki Y, Shin-I T, Kohara Y, Yamazaki Y (2004) Construction of a full-length cDNA library from young spikelets of hexaploid wheat and its characterization by large-scale sequencing of expressed sequence tags. *Genes Genet Syst* 79:227–232
- Oono Y, Seki M, Nanjo T, Narusaka M, Fujita M, Satoh R, Satou M, Sakurai T, Ishida J, Akiyama K, Iida K, Maruyama K, Sato S, Yamaguchi-Shinozaki K, Shinozaki K (2003) Monitoring expression profiles of *Arabidopsis* gene expression during rehydration process after dehydration using ca. 7000 full-length cDNA microarray. *Plant J* 34:868–887
- Oono Y, Seki M, Satou M, Iida K, Akiyama K, Sakurai T, Fujita M, Yamaguchi-Shinozaki K, Shinozaki K (2006) Monitoring expression profiles of *Arabidopsis* genes during cold acclimation and deacclimation using DNA microarrays. *Funct Integr Genomics* 6:212–234
- Osakabe Y, Miyata S, Urao T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2002) Overexpression of *Arabidopsis* response regulators, ARR4/ATRR1/IBC7 and ARR8/ATRR3, alters cytokinin responses differentially in the shoot and in callus formation. *Biochem Biophys Res Commun* 293:806–815
- Osakabe Y, Maruyama K, Seki M, Satou M, Shinozaki K, Yamaguchi-Shinozaki K (2005) An LRR receptor kinase, RPK1, is a key membrane-bound regulator of abscisic acid early signaling in *Arabidopsis*. *Plant Cell* 17:1105–1119
- Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D (2008) Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res* 18:2024–2033
- Ralph SG, Chun HJE, Kolosova N, Cooper D, Oddy C, Ritland CE, Kirkpatrick R, Moore R, Barber S, Holt RA, Jones SJM, Marra MA, Douglas CJ, Ritland K, Bohlmann J (2008) A conifer genomics resource of 200,000 spruce (*Picea* spp.) ESTs and 6,464 high-quality, sequence-finished full-length cDNAs for Sitka spruce (*Picea sitchensis*). *BMC Genomics* 9:484
- Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perraud PF, Lindquist EA, Kamisugi Y, Tanahashi T, Sakakibara K, Fujita T, Oishi K, Shin-I T, Kuroki Y, Toyoda A, Suzuki Y, Hashimoto S, Yamaguchi K, Sugano S, Kohara Y, Fujiyama A, Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho SH, Dutcher SK, Estelle M, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hicks KA, Hughes J, Lohr M, Mayer K, Melkozernov A, Murata T, Nelson DR, Pils B, Prigge M, Reiss B, Renner T, Rombauts S, Rushton PJ, Sanderfoot A, Schween G, Shiu SH, Stueber K, Theodoulou FL, Tu H, Van de Peer Y, Verrier PJ, Waters E, Wood A, Yang L, Cove D, Cuming AC, Hasebe M, Lucas S, Mishler BD, Reski R, Grigoriev IV, Quatrano RS, Boore JL (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319:64–69
- Ruan Y, Ooi HS, Choo SW, Chiu KP, Zhao XD, Srinivasan KG, Yao F, Choo CY, Liu J, Ariyaratne P, Bin WG, Kuznetsov VA, Shahab A, Sung WK, Bourque G, Palanisamy N, Wei CL (2007) Fusion transcripts and transcribed retrotransposed loci discovered through comprehensive transcriptome analysis using Paired-End diTags (PETs). *Genome Res* 17:828–838
- Sakuma Y, Maruyama K, Osakabe Y, Feng Q, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2006) Functional analysis of an *Arabidopsis* transcription factor, DREB2A, involved in drought-responsive gene expression. *Plant Cell* 18:1292–1309
- Sakurai T, Satou M, Akiyama K, Iida K, Seki M, Kuromori T, Ito T, Konagaya A, Toyoda T, Shinozaki K (2005) RARGE: a large-scale database of RIKEN *Arabidopsis* resources ranging from transcriptome to phenome. *Nucleic Acids Res* 33:D647–D650
- Sakurai T, Plata G, Rodriguez-Zapata F, Seki M, Salcedo A, Toyoda A, Ishiwata A, Tohme J, Sakaki Y, Shinozaki K, Ishitani M (2007) Sequencing analysis of 20,000 full-length cDNA clones from cassava reveals lineage specific expansions in gene families related to stress response. *BMC Plant Biol* 7:66
- Sato K, Shin-I T, Seki M, Shinozaki K, Yoshida H, Takeda K, Yamazaki Y, Conte M, Kohara Y (2009) Development of 5,006 full-length cDNAs in barley: a tool for accessing cereal genomics resources. *DNA Res* 16:81–89
- Satoh R, Nakashima K, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2002) ACTCAT, a novel *cis*-acting element for proline- and hypoosmolarity-responsive expression of the *ProDH* gene encoding proline dehydrogenase in *Arabidopsis*. *Plant Physiol* 130:709–719
- Sawasaki T, Ogasawara T, Morishita R, Endo Y (2002) A cell-free protein synthesis system for high-throughput proteomics. *Proc Natl Acad Sci USA* 99:14652–14657
- Sawasaki T, Hasegawa Y, Morishita R, Seki M, Shinozaki K, Endo Y (2004) Genome-scale, biochemical annotation method based on the wheat germ cell-free protein synthesis system. *Phytochemistry* 65:1549–1555
- Seki M, Carninci P, Nishiyama Y, Hayashizaki Y, Shinozaki K (1998) High-efficiency cloning of *Arabidopsis* full-length cDNA by biotinylated CAP trapper. *Plant J* 15:707–720

- Seki M, Narusaka M, Abe H, Kasuga M, Yamaguchi-Shinozaki K, Carninci P, Hayashizaki Y, Shinozaki K (2001a) Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses using full-length cDNA microarray. *Plant Cell* 13:61–72
- Seki M, Narusaka M, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y, Shinozaki K (2001b) *Arabidopsis* encyclopedia using full-length cDNAs and its application. *Plant Physiol Biochem* 39:211–220
- Seki M, Ishida J, Narusaka M, Fujita M, Nanjo T, Umezawa T, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Akiyama Y, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y, Shinozaki K (2002a) Monitoring the expression pattern of ca. 7000 *Arabidopsis* genes under ABA treatments using a full-length cDNA microarray. *Funct Integr Genomics* 2:282–291
- Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Akiyama K, Taji T, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y, Shinozaki K (2002b) Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold, and high-salinity stresses using a full-length cDNA microarray. *Plant J* 31:279–292
- Seki M, Narusaka M, Kamiya A, Ishida J, Satou M, Sakurai T, Nakajima M, Enju A, Akiyama K, Oono Y, Muramatsu M, Hayashizaki Y, Kawai J, Carninci P, Itoh M, Ishii Y, Arakawa T, Shibata K, Shinagawa A, Shinozaki K (2002c) Functional annotation of a full-length *Arabidopsis* cDNA collection. *Science* 296:141–145
- Seki M, Satou M, Sakurai T, Akiyama K, Iida K, Ishida J, Nakajima M, Enju A, Narusaka M, Fujita M, Oono Y, Kamei A, Yamaguchi-Shinozaki K, Shinozaki K (2004a) Expression profiling under abiotic stress conditions using RIKEN *Arabidopsis* full-length (RAFL) cDNA microarray. *J Exp Bot* 55:213–223
- Seki M, Satou M, Sakurai T, Akiyama K, Iida K, Ishida J, Nakajima M, Enju A, Narusaka M, Fujita M, Oono Y, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y, Shinozaki K (2004b) Full-length cDNAs for the discovery and annotation of genes in *A. thaliana*. In: Leister D (ed) *Plant Functional Genomics*. Haworth's Food Products Press, New York, pp 3–22
- Seki M, Ishida J, Nakajima M, Enju A, Iida K, Satou M, Fujita M, Narusaka Y, Narusaka M, Sakurai T, Akiyama K, Oono Y, Kamei A, Umezawa T, Mizukado S, Maruyama K, Yamaguchi-Shinozaki K, Shinozaki K (2005) Genomic analysis of stress response. In: Jenks MA, Hasegawa PM (eds) *Plant abiotic stress*. Blackwell, Sheffield, pp 248–265
- Shinozaki K, Yamaguchi-Shinozaki K (2000) Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. *Curr Opin Plant Biol* 3:217–223
- Simpson SD, Nakashima K, Narusaka Y, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Two different novel cis-acting elements of *erd1*, a *clpA* homologous *Arabidopsis* gene function in induction by dehydration stress and dark-induced senescence. *Plant J* 33:259–270
- Taji T, Sakurai T, Mochida K, Ishikawa A, Kurotani A, Totoki Y, Toyoda A, Sakaki Y, Seki M, Ono H, Sakata Y, Tanaka S, Shinozaki K (2008) Large scale collection and annotation of full-length cDNAs from a model halophyte, *Thellungiella halophila*. *BMC Plant Biol* 8:115
- Takahashi H, Nozawa A, Seki M, Shinozaki K, Endo Y, Sawasaki T (2009) A simple and high-sensitivity method for analysis of ubiquitination and polyubiquitination based on wheat cell-free protein synthesis. *BMC Plant Biol* 9:39
- The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815
- The French-Italian Public Consortium for Grapevine Genome Characterization (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449:463–468
- Thomashow MF (1999) Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. *Annu Rev Plant Physiol Plant Mol Biol* 50:571–599
- Toyoda T, Shinozaki K (2005) Tiling array-driven elucidation of transcriptional structures based on maximum-likelihood and Markov models. *Plant J* 43:611–621
- Tran LSP, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, Fujita M, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2004) Functional analysis of *Arabidopsis* NAC transcription factors controlling expression of *erd1* gene under drought stress. *Plant Cell* 16:2481–2498
- Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroev S, Dejardin A, dePamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryabov D, Schmutz J, Schrader J, Segerman B, Shih H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr & Gray). *Science* 313:1596–1604
- Umezawa T, Sakurai T, Totoki Y, Toyoda A, Seki M, Ishikawa A, Akiyama K, Kurotani A, Yoshida T, Mochida K, Kasuga M, Todaka D, Maruyama K, Nakashima K, Enju A, Mizukado S, Ahmed S, Yoshihara K, Harada K, Tsubokura Y, Hayashi M, Sato S, Anai T, Ishimoto M, Funatsuki H, Teraishi M, Osaki M, Shinano T, Akashi R, Sasaki Y, Yamaguchi-Shinozaki K, Shinozaki K (2008) Sequencing and analysis of approximately 40,000 soybean cDNA clones from a full-length enriched cDNA library. *DNA Res* 15:333–346
- Wang BB, Brendel V (2006) Genomewide comparative analysis of alternative splicing in plants. *Proc Natl Acad Sci USA* 103:7175–7180
- Wang XJ, Gaasterland T, Chua NH (2005) Genome-wide prediction and identification of cis-natural antisense transcripts in *Arabidopsis thaliana*. *Genome Biol* 6:R30
- White JA, Todd J, Newman T, Focks N, Girke T, Ilárduya OMD, Jaworski JG, Ohlrogge JB, 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
- Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, Pham P, Cheuk R, Karlín-Neumann G, Liu SX, Lam B, Sakano H, Wu T, Yu G, Miranda M, Quach HL, Tripp M, Chang CH, Lee JM, Toriumi M, Chan MMH, Tang CC, Onodera CS, Deng JM, Akiyama K, Ansari Y, Arakawa T, Banh J, Banno F, Bowser L, Brooks S, Carninci P, Chao Q, Choy N, Enju A, Goldsmith AD, Gurjal M, Hansen NF, Hayashizaki Y, Johnson-Hopson C, Hsuan VW, Iida K, Karnes

- M, Khan S, Koesema E, Ishida J, Jiang PX, Jones T, Kawai J, Kamiya A, Meyers C, Nakajima M, Narusaka M, Seki M, Sakurai T, Satou M, Tamse R, Vaysberg M, Wallender EK, Wong C, Yamamura Y, Yuan S, Shinozaki K, Davis RW, Theologis A, Ecker JR (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* 302:842–846
- Yamaguchi-Shinozaki K, Shinozaki K (2005) Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. *Trends Plant Sci* 10:88–94
- Yamaguchi-Shinozaki K, Shinozaki K (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annu Rev Plant Biol* 57:781–803
- Yamamoto YY, Ichida H, Matsui M, Obokata J, Sakurai T, Satou M, Seki M, Shinozaki K, Abe T (2007) Identification of plant promoter constituents by analysis of local distribution of short sequences. *BMC Genomics* 8:67
- Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Hayami N, Terada T, Shirouzu M, Osanai T, Tanaka A, Seki M, Shinozaki K, Yokoyama S (2004a) Solution structure of the B3 DNA domain of the *Arabidopsis* cold-responsive transcription factor RAV1. *Plant Cell* 16:3448–3459
- Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Nunokawa E, Ishizuka Y, Terada T, Shirouzu M, Osanai T, Tanaka A, Seki M, Shinozaki K, Yokoyama S (2004b) A novel zinc-binding motif revealed by solution structures of DNA-binding domains of *Arabidopsis* SBP-family transcription factors. *J Mol Biol* 337:49–63
- Yamasaki K, Kigawa T, Inoue M, Yamasaki T, Tateno M, Yabuki T, Aoki M, Seki E, Masuda T, Tomo Y, Hayami N, Terada T, Shirouzu M, Osanai T, Tanaka A, Seki M, Shinozaki K, Yokoyama S (2005a) Solution structure of the WRKY DNA-binding domain. *Plant Cell* 17:944–956
- Yamasaki K, Kigawa T, Inoue M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Tomo Y, Terada T, Shirouzu M, Tanaka A, Seki M, Shinozaki K, Yokoyama S (2005b) Solution structure of the major DNA-binding domain of *Arabidopsis* ETHYLENE-INSENSITIVE3-LIKE3. *J Mol Biol* 348:253–264
- Yamasaki K, Kigawa T, Inoue M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Tomo Y, Terada T, Shirouzu M, Tanaka A, Seki M, Shinozaki K, Yokoyama S (2006) Disruption of the second zinc-binding site of an *Arabidopsis* SBP-family transcription factor retains its tertiary structure. *FEBS Lett* 580:2109–2116
- Yokoyama S, Hirota H, Kigawa T, Yabuki T, Shirouzu M, Terada T, Ito Y, Matsuo Y, Kuroda Y, Nishimura Y, Kyogoku Y, Miki K, Masui R, Kuramitsu S (2000) Structural genomics projects in Japan. *Nature Struct Biol* 7:943–945
- Zhu JK (2002) Salt and drought stress signal transduction in plants. *Annu Rev Plant Biol* 53:247–273