Supplemental text of experimental procedures:

**Plant Materials.** For the construction of the libraries (RAFL1, RAFL15, and RAFL19), *Arabidopsis thaliana* plants (ecotype Columbia) were grown in soil for 4 to 8 weeks under continuous light at 22°C. For the construction of the libraries (RAFL2, RAFL3, RAFL4, RAFL5, RAFL7, RAFL8, RAFL12, RAFL13, RAFL14, RAFL17 and RAFL18), *Arabidopsis* plants were grown on germination-medium (GM)-containing agar plates for 3 to 4 weeks with a cycle of 16-h light/8-h dark at 22°C. For the construction of the libraries (RAFL6, RAFL9, RAFL11, and RAFL21), the plants grown in soil under continuous light at 22°C and the plants grown on GM-containing agar plates were used. After treatment with various stresses and hormones, the following plant materials were frozen in liquid nitrogen and stored at -80°C until RNA preparation for the construction of the cDNA libraries. For the construction of RAFL1 library, leaves and stems of the plants grown for 8 weeks were frozen in liquid nitrogen. For the RAFL2 library, the plants grown for 4 weeks were removed from the agar, desiccated for 2 h in plastic dishes under dim light, and frozen. For the RAFL3 library, the plants grown for 4 weeks were removed from the agar and frozen. For the RAFL4 and RAFL7 libraries, the plants grown for 4 weeks were grown in an incubator for 1, 2, 5, 10 and 24 h at 4°C, and frozen. For the RAFL5 and RAFL8 libraries, plants grown for 4 weeks were removed from the agar, desiccated for 1, 2, 5, 10, and 24 h in plastic dishes under dim light, and frozen. For the RAFL6 and RAFL9 libraries, the plants harvested every 5 days after sowing until 11 weeks after sowing, the 4-week-old plants treated with dehydration (1, 2, 5, 10, and 24 h) and cold (1, 2, 5, 10, and 24 h), tissues of mature seeds, siliques and flowers from the plants grown in soil for 6 to 11 weeks and old, yellow leaves induced by natural senescence were used for the construction of the cDNA libraries. For the RAFL11 library, the plants harvested every 5 days after sowing until 11 weeks after sowing, the 4-week-old plants treated with various stresses, such as dehydration (1, 2, 5, 10, and 24 h), cold (1, 2, 5, 10, and 24 h), high-salinity (1, 2, 5, 10, and 24 h), heat and UV, and abscisic acid (ABA) (1, 2, 5, 10 and 24 h) stress treatments, the plants grown for 7 days under dark conditions after sowing, and silique tissues from the plants grown in soil for 6 to 11 weeks were used for the construction of the cDNA library. ABA and high-salinity-stress-treatments were performed as described previously (Yamaguchi-Shinozaki and Shinozaki,
1994). For the heat treatment, the plants grown for 4 weeks were grown in an incubator for 5 h at 37°C. For the UV treatment, the plants grown for 4 weeks were treated with UV-C (100 mJ/cm² and 200 mJ/cm²), and grown on a cycle of 16-h light/8-h dark at 22°C for 24 h. For the RAFL12 library, the 4-week-old plants treated with cold (1, 2, 5, 10 and 24 h) were used. For the RAFL13 library, the 4-week-old plants treated with dehydration (1, 2, 5, 10 and 24 h) were used. For the RAFL14 library, the root tissues from the plants grown for 3 weeks on GM agar medium were used. For the RAFL15 library, the tissues of the flowers and siliques from the plants grown in soil for 6 to 11 weeks were used. For the RAFL16 library, the plants grown for 7 days under dark conditions after sowing were used. For the RAFL17 library, the 4-week-old plants treated with dehydration (1, 2, 5, 10, and 24 h) and rehydration after dehydration were used. For the rehydration treatment, the plants grown for 4 weeks were removed from the agar, desiccated for 10 h in plastic dishes under dim light, transferred to water, and grown under dim light for 1, 2, 5, 10 and 24 h. For the RAFL18 library, the 4-week-old plants treated with cold (1, 2, 5, 10 and 24 h) were used. For the RAFL19 library, the flowers and siliques from the plants grown for 6 to 11 weeks were used. For the RAFL21 library, the 4-week-old-plants treated with various stresses, such as heat and UV, hormones, such as ABA, auxin, cytokinin, ethylene, gibberellin (GA), jasmonic acid (JA) and salicylic acid (SA), and benzo [1,2,3] thiadiazole-7-carbothioic acid S-methyl ester (BTH) were used. For the heat and UV treatments, we carried out the same treatments as that for the RAFL11 library. ABA treatment (2 h) was performed as described previously (Yamaguchi-Shinozaki and Shinozaki, 1994). For the auxin treatment, the plants grown for 4 weeks were removed from the agar, transferred to and grown hydroponically in water of 10 uM indole-3-acetic acid (IAA) under dim light for 1 and 3 h. For the ethylene treatment, the plants grown for 4 weeks were removed from the agar, transferred to and grown hydroponically in water of 2.5 mM ethephon under dim light for 5 h. For the JA treatment, the plants grown for 4 weeks were sprayed with a 200 uM methyl jasmonate (MeJA) solution and grown on a cycle of 16-h light/8-h dark at 22°C for 5 h. SA treatment was performed as described previously (Narusaka et al., 1999). The plants grown for 4 weeks were sprayed with 5 mM SA and grown on a cycle of 16-h light/8-h dark at 22°C for 24 h. For the GA treatment, the plants grown for 4 weeks were removed from the agar, transferred to and grown hydroponically in water of 100
uM gibberellic acid (GA$_3$) solution under dim light for 3 h. For the cytokinin treatment, the plants grown for 4 weeks were removed from the agar, transferred to and grown hydroponically in water of 100 uM benzyladenine (BA) under dim light for 30 min and 6 h. The treatment with BTH was performed as described previously (Narusaka et al., 1999). The plants grown for 4 weeks were sprayed with 0.5 mM BTH (Novartis Crop Protection AG) solution and grown on a cycle of 16-h light/8-h dark at 22°C for 24 h.

**RNA preparation and construction of Arabidopsis full-length cDNA libraries.** RNA preparation was done essentially as reported previously (Seki et al., 1998). Full-length cDNA libraries were constructed essentially as reported previously (Carninci et al., 1996, 1997; Seki et al., 1998) by biotinylated CAP trapper using trehalose-thermoactivated reverse transcriptase (Carninci et al., 1998). We used the lambda-ZAP (Carninci et al., 1996; Seki et al., 1998) and lambda-FLC (Carninci et al., 2001) vectors for the construction of the cDNA libraries. The lambda-FLC vectors accommodate cDNAs in a broad range of sizes and are useful for the high-efficiency cloning of long cDNA fragments (Carninci et al., 2001). The lambda-FLC vectors can also be bulk-excised by a Cre-<i>lox</i>-based system free of size bias to produce the plasmid libraries. In the construction of the some full-length cDNA libraries (RAFL12, 13, 14, 15, 16, 17, 18, 19 and 21; Table 1), a new single-strand linker ligation method (SSLLM) (Shibata et al., 2001), which uses DNA ligase to add a double-stranded (ds) DNA linker to single-stranded (ss) full-length cDNA, was also introduced. This method is advantageous because the sequencing of the clones and the translation of proteins from full-length cDNA are easier and more efficient due to the elimination of GC tail. Normalization and subtraction procedures (Carninci et al., 2000) were also introduced in the construction of the some full-length cDNA libraries (RAFL11, 12, 13, 17, 18, 19 and 21; Table 1) to reduce the frequency of highly expressed mRNAs in the library, and to remove cDNAs already categorized by using one-pass sequencing, respectively.

**DNA extraction, sequencing and clustering of the cDNA clones.** Bacteria were picked with commercially available picking machines (Q-bot; Genetics, UK) and transferred to 384-microwell
plates. Duplicate plates were used to prepare plasmid DNA. For plasmid DNA, 384-well plates were divided and grown in 4 X 96 deep-well plates. After overnight incubation, plasmids were extracted either manually or automatically (Itoh et al., 1999). DNA sequences were determined typically using the RISA sequencing system (Shibata et al., 2000); and partly using the DNA sequencer (model 3700; Perkin Elmer-Applied Biosystems, CA, USA). DNA clones were subjected to single-pass sequencing from the 3’-ends of the cDNA. The primer was 5’-CAGGAAACAGCTATGAC. The 3’-end single-pass sequencing data was used in the following 2 steps for clustering. The poly-T (for 3’-end), C-stretch (for 5’-end) regions, and the vector sequences were trimmed from the 3’-end single-pass sequencing data. From the trimmed sequences, we selected 100-bp sequences as tag sequences for the first clustering. We used the BLAST program to search for homology between the new tag and the database of nonredundant 100-bp tag sequences; sequences having BLAST parameters of E=1.0e−20 or lower were clustered together. When the tag sequence was not in the database, the tag was added to the database. When the database contained the tag, it was added as a member of the identical tag group. If the tag was found in the database and at the same time when the shift was < 10 bases, the overlap was >80 bases, with >90% identity in the overlap, the sequences were grouped together. After the first clustering, the best quality sequence was chosen as the representative of the group. Then we selected the each 3’-end representative sequence that were longer than 300 bp and that had <1% ambiguity within 300-bp sequences as the tag sequences for the second clustering. We used the BLAST program to search for homology between the new tag and the database of nonredundant >300-bp tag sequences. When the tag sequence was not in the database, the tag was added to the database. When the database contained the tag, it was added as a member of the identical tag group. If the tag was found in the database and at the same time when the overlap was >100 bases, with >98% identity in the overlap, the sequences were grouped together. After the second clustering, the best quality sequence was chosen as the representative of the group. The 3’-EST of each representative clone was then mapped onto the Arabidopsis genome

(ftp://ftpmpis.gsf.de/cress/chr1/chromo1_anno_v040701.dat;
ftp://ftpmpis.gsf.de/cress/chr2/chromo2_anno_v040701.dat;
using the BLASTN program. At the Munich information center for protein sequences (MIPS) Arabidopsis thaliana database mentioned above, the presence of 26,285 genes including 25,556 protein-encoding genes, tRNAs, small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs) and pseudogenes was predicted by the Arabidopsis Genome Initiative (AGI) as of July 5, 2001. The criterion for mapping was identity >95% within >100 bp-overlap. As a result, 13,292 non-redundant representative sequences whose corresponding MIPS protein entry code was identified were mapped on the Arabidopsis genome. For the 2,359 representative sequences mapped on the Arabidopsis genome, the sequence of its corresponding MIPS protein entry code did not exist within 1,000 bp of the mapped position. The 2,359 representative sequences were then mapped on the basis of the mapped position. If the mapped position of the representative sequences is within 500 bp, the sequences were grouped together. As a result, the 2,359 representative sequences were grouped into 1,586 non-redundant representative sequences and 14,878 non-redundant representative sequences were mapped on the Arabidopsis genome.

Next, the 14,878 cDNA clones were subjected to single-pass sequencing from the 5’-end of the cDNA. The primer was 5’-GTTTTCCCAGTCACGAC. The 5’-end sequencing data of the 14,034 clones sequenced was mapped onto the Arabidopsis genome using the BlastN program as mentioned above. As a result, 13,146 non-redundant representative sequences whose corresponding MIPS protein entry code was identified were mapped on the Arabidopsis genome. For the remaining 888 sequences mapped on the Arabidopsis genome, the sequences of its corresponding MIPS protein entry code did not exist within 1000 bp of the mapped position. Furthermore, the 3’- and 5’-end sequencing data of the clones whose protein entry code is different from that for the 3’-EST and that for the 5’-EST were subjected to homology search using the BlastN and Blast X programs. Finally, by combining the results of BlastN and Blast X mentioned above, the corresponding MIPS protein entry code of the 13,831 RAFL cDNA clones were identified and the corresponding MIPS protein entry code of the 837 RAFL cDNA clones were not identified.
References for supplemental material:


