Published online September 27, 2004

5096–5103 *Nucleic Acids Research, 2004, Vol. 32, No. 17*
doi:10.1093/nar/gkh845

Genome-wide analysis of alternative pre-mRNA splicing in *Arabidopsis thaliana* based on full-length cDNA sequences

Kei Iida[1,2], Motoaki Seki[1,3], Tetsuya Sakurai[1,2], Masakazu Satou[1,2], Kenji Akiyama[1,2], Tetsuro Toyoda[2], Akihiko Konagaya[2] and Kazuo Shinozaki[1,3,*]

[1]Plant Mutation Exploration Team, Plant Functional Genomics Research Group and [2]Genomic Knowledge Base Research Team, Bioinformatics Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan and [3]Laboratory of Plant Molecular Biology, RIKEN Tsukuba Institute, 3-1-1 Koyadai, Tsukuba 305-0074, Japan

Received July 14, 2004; Revised and Accepted September 6, 2004

ABSTRACT

We mapped RIKEN *Arabidopsis* full-length (RAFL) cDNAs to the *Arabidopsis thaliana* genome to search for alternative splicing events. We used 278 734 full-length and 3′/5′ terminal reads of the sequences of 220 214 RAFL cDNA clones for the analysis. Eighty-nine percent of the cDNA sequences could be mapped to the genome and were clustered in 17 130 transcription units (TUs). Alternative splicing events were found in 1764 out of 15 214 TUs (11.6%) with multiple sequences. We collected full-length cDNA clones from plants grown under various environmental conditions or from various organs. We then analyzed the correlation between alternative splicing events and environmental stress conditions. Alternative splicing profiles changed according to environmental stress conditions and the various developmental stages of plant organs. In particular, cold-stress conditions affected alternative splicing profiles. The change in alternative splicing profiles under cold stress may be mediated by alternative splicing and transcriptional regulation of splicing factors.

INTRODUCTION

Arabidopsis thaliana Heynh. is a model organism used to study various molecular systems in the development, environmental responses and metabolism of higher plants. Its complete genomic sequence has been determined (1), and extensive large-scale, full-length cDNA collections have been made (2–4). From work on the human genome sequence, alternative splicing is now thought to be important to the complexity of gene function (5). Alternative splicing events produce additional transcripts from genes to mediate the complicated functions of the human body. Alternative splicing events are also important in higher plants. Large-scale alternative splicing in *A.thaliana* was first analyzed by Haas *et al*. (3,6). They used ∼180 000 expressed sequence tags (ESTs), including ∼80 000 RIKEN *Arabidopsis* full-length (RAFL) cDNAs, to detect 1188 genes containing alternative splicing variations. In addition, Zhu *et al*. (7) have reported 327 alternative splicing events in an analysis using ∼180 000 ESTs. In our project, we had collected more than 270 000 sequences by the end of 2003 (Figure 1a). We expected that a large number of EST sequences would allow us to detect many alternative splicing events. It was also of help that our EST sequences were obtained from full-length cDNA, because all EST sequences have information on the 5′ and/or 3′ terminal sites.

Our RAFL cDNA collection has an additional advantage for the analysis of alternative splicing events. We have constructed 18 cDNA libraries of expressed genes from *Arabidopsis* plants grown under various environmental conditions or from plant organs at various developmental stages. Therefore, each RAFL cDNA clone has associated information on the conditions or organs in which it is expressed. To use this information, we analyzed the relationship between the expression of alternatively spliced transcripts and plant growth conditions. Previous studies suggest that alternative splicing events occur in response to environmental changes or at particular developmental stages (8–10). However, there have been few reports on changes in alternative splicing profiles according to expressional conditions at the whole transcriptome level. We discuss the molecular mechanism of cold-inducible changes in alternative splicing profiles.

MATERIALS AND METHODS

Data set

We used 278 734 sequences from RAFL cDNA clones. They included 92 654 RAFL 5′ terminal read sequences, 172 653 RAFL 3′ terminal read sequences and 13 427 RAFL full-length read sequences (Figure 2). We analyzed 248 514 mapped cDNA clones. About 190 000 unpublished sequences were also used for the analysis. These sequences can be

*To whom correspondence should be addressed. Tel: +81 29 836 4359; Fax: +81 29 836 9060; Email: sinozaki@rtc.riken.jp
Present address:
Kei Iida, Faculty of Bio-Science, Nagahama Institute of Bio-Science and Technology, 1266 Tamura-cho, Nagahama 526-0829, Japan

Nucleic Acids Research, Vol. 32 No. 17 © Oxford University Press 2004; all rights reserved

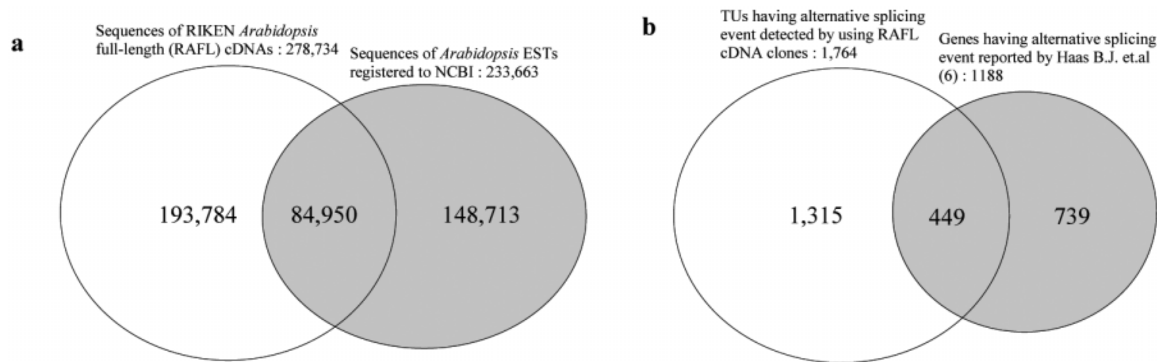

Figure 1. Comparison of sequencing resources and detected alternative splicing events between this work and previous work. (**a**) *A.thaliana* cDNA sequence resources used in our analysis (in December 2003), (**b**) Venne diagram of the genes with alternative splicing events detected in this work and in previous work by Hass *et al*. (6).

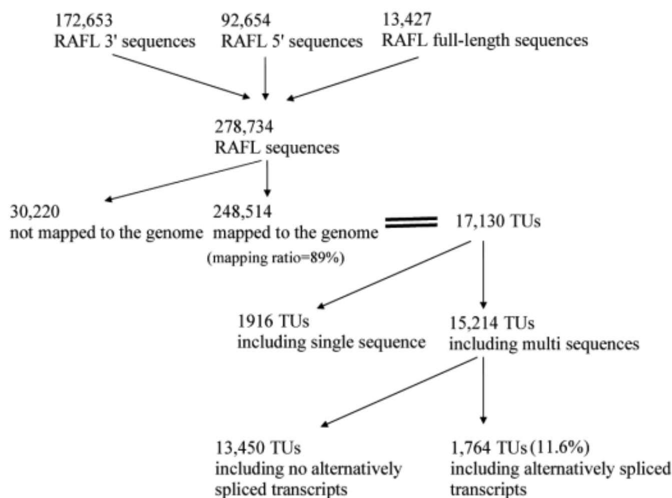

Figure 2. Data flow of clustering for the analysis of alternative splicing events in RAFL sequences.

downloaded from RARGE (http://rarge.gsc.riken.jp/) and have been deposited in the DNA database of Japan (DDBJ).

Mapping the RAFL cDNA clone sequences to the *Arabidopsis* genome

We mapped the RAFL cDNA sequences to the *Arabidopsis* genome using BLAST (11). We clustered the results in two steps. In the first step, to detect long and identical exons, we chose sequences with ⩾95% identity and a length of ⩾50 bp as exons. In the second step, to detect micro-exons (3) or other small exons, we chose sequences with ⩾85% identity and a length of ⩾15 bp where each HSP (high-scoring segment pair) was consistent with exons detected in the first step. Although a micro-exon is defined as an exon with a length of 3–25 bp (3), we did not treat HSPs with a length of <15 bp as exons. It is difficult to detect such micro-exons using BLAST. In some cases, this problem causes the incorrect detection of exon skip-type (ES-type) alternative splicing events. In addition, ~10 bp sequences on exon–intron boundaries usually belong to both of the two neighboring exons. To avoid the incorrect detection of exon–intron structure as a result, we used 15 bp sequences as a spacer to check the consistency of the exon–intron structure.

After mapping the RAFL cDNA sequences to the genome, we clustered mapped sequences into transcription units (TUs) according to the method of Okazaki *et al*. (12). Sequences with the same direction and overlapping nucleotides were clustered into single TUs.

Detection of alternative splicing

Before searching for alternative splicing events, we surveyed the genomic exon–intron structure of each TU. We aligned sequences clustered into single TUs on the genome sequence and considered contiguous nucleotides as genomic exons if each nucleotide was on an exon in at least one RAFL cDNA sequence. Following detection of the genomic exon–intron structure, we searched for alternative splicing events. For ES-type alternative splicing, we simply searched for sequences without genomic exons. If an exon loss occurred on the 3′ terminal site of a TU, the alternative splicing event was categorized as an alternative terminal exon type (AT-type) one. The genomic exon reflected the longest exon. For alternative donor type (AD-type) and alternative acceptor type (AA-type) alternative splicing, we searched for exons shorter than genomic exons. We did not check the 5′ sites of initial exons or 3′ sites of terminal exons because these are not splice sites. We used an additional rule in the search for AD/AA-type alternative splicing events to avoid misdetection. The BLAST algorithm was sometimes not good at finding identical sequences at an exon–intron boundary if the similarities of the boundary sequences were not high because of sequence read errors. To detect AD/AA-type events, we used the lengths of sequences that could not be aligned on the genome as spacers. In considering spacers, we considered a sequence as an alternatively spliced clone of the AD/AA type if $L_d - L_s > L_{t\ (AD/AA)}$ [where L_d is the length of the sequence difference between the genomic and sequence exon–intron boundaries, L_s is the spacer length and $L_{t\ (AD/AA)}$ is a threshold length for the detection of an AD/AA-type event; we used $L_{t\ (AD/AA)} = 10$]. For a retained intron type (RI-type) alternative splicing event, we searched for introns on genomic exons. As in the detection of other types of alternative splicing, we used nucleotides that could not be aligned to the genome as spacers. We considered a sequence to be an RI-type alternatively spliced clone if $L_i - L_s > L_{t\ (RI)}$ (where L_i is the length of an intron found on a genomic exon, L_s is the spacer length and

$L_{t\ (RI)}$ is a threshold length for the detection of RI-type alternative splicing; we used $L_{t\ (RI)} = 20$).

The sequences used contained both 5′ and 3′ terminal read sequences of the RAFL clones. We deselected uninformative sites of each sequence to check alternative splicing. (For example, the 3′ terminal sites of the 5′ terminal read sequences were not used to check AT-type alternative splicing.)

We used the BLAST algorithm for mapping the RAFL cDNA clones to the genome. The BLAST algorithm is good for searching for highly similar sequences, but sometimes it will not detect sequences with low similarity, especially exon–intron boundary sequences. Other programs are better tuned for mapping sequences to a genome, e.g. GeneSeqer (13) and Sim4 (14). We tried these programs, but they gave worse results than BLAST because our sequence set included some poor-quality terminal read sequences. As we did not use sequences with <95% identity, good sequences remained that could be mapped to the genome. Considering these issues, we chose BLAST.

Analysis of alternative splicing profiles in each library

We analyzed the relationships between alternative splicing events and environmental conditions or the various developmental stages of plant organs. In this analysis, we compared the measured and expected numbers of clones that have specific alternative splicing profiles in each library. For this comparison, we counted clones without alternative exons in ES-type alternative splicing events, those with short terminal sites in AT-type alternative splicing, those with short exons in AD or AA sites and those with unspliced introns in RI-type alternative splicing. Expected numbers were calculated as the product of the number of library members and the probability of each alternative splicing event. We used the chi-square test to check whether the differences between measured and expected values were statistically significant. We used the statistics software 'R' (http://cran.r-project.org/) to compute P-values in the chi-square test.

RESULTS

Mapping the RAFL cDNA clone sequences to the *Arabidopsis* genome

We mapped 248 514 (89%) of 278 734 RAFL cDNA clone sequences to the *A.thaliana* genome (1) using the BLAST search (11). We used a mapping rule in which each exon has ⩾95% identity to the genome in a ⩾50 bp region. Haas *et al*. (3) reported micro-exons in some *Arabidopsis* genes. To detect micro-exons or other small exons, we used an additional rule in which exons ⩾15 bp are considered to be micro-exons only if they occur between mapped exons. cDNA clones with mapping coverage of <90% of the corresponding full-length exons were not used. After mapping these sequences to the genome, we constructed TUs (12). Sequences that are encoded on the same strands of the same chromosome and overlap by at least 1 nt were clustered into single TUs. Using this rule, we analyzed the whole *Arabidopsis* genome to identify 17 130 TUs out of 248 514 sequences (Figure 2). Each TU was estimated to contain 14.5 sequences on average. The TU with the most sequences contained 1335 sequences, encoding the *dna*k-type molecular chaperone hsc 70.1. TUs with ~30 sequences accounted for almost 90% of all TUs.

Detection of alternative splicing

We searched for alternative splicing events in 15 214 TUs with two or more sequences. We divided events into five types: ES-type, AT-type, AD-type, alternative AA-type and RI-type events, following Zhou *et al*. (15) and Haas *et al*. (3) (Figure 3). Although AT is a subtype of ES, we treat AT-type events as an independent category. These two alternative splicing events may belong to different types, because selection of an alternative terminal exon can occur as a result of alternative polyadenylation.

We detected 1764 TUs (11.6% of multisequence TUs) with splicing variants (Table 1; http://rarge.gsc.riken.jp/a_splicing/) including 1315 TUs that were newly detected as alternatively spliced TUs (Figure 1b). Among the 15 214 TUs, we identified 273 (1.8%) ES-type and 458 (3.0%) AT-type events. In both ES-type and AT-type alternative splicing events, there was at least one alternative exon. If the alternative exon was on the terminal site of the TU, this alternative splicing event was classified as an AT-type one. These results indicate that the alternative exon tended to be on the terminal sites of TUs rather than on their initial or middle sites. We identified 250 (1.6%) AD-type and 321 (2.1%) AA-type events. When TUs contained a shorter exon than the genomic exon, they were considered to be AD-type or AA-type events (details in Materials and Methods). We could not find much difference among occurrences of these two alternative splicing events. Of the 15 214 TUs, 790 (5.2%) had RI-type alternative splicing events. RI-type TUs were slightly more common than the other types. Some RI-type alternative splicing events may

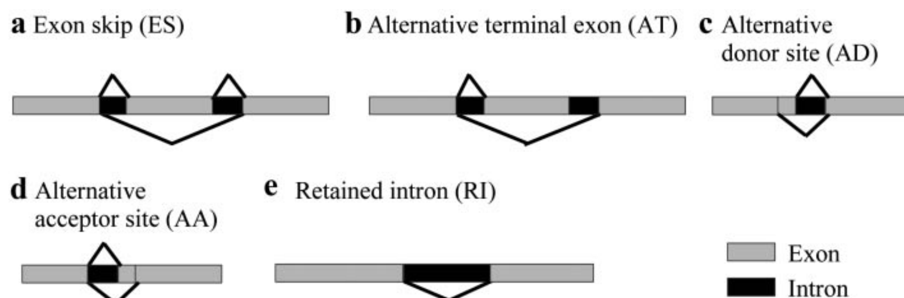

Figure 3. Types of alternative splicing: (**a**) exon skip, (**b**) alternative terminal exon (subtype of exon skip), (**c**) alternative donor site, (**d**) alternative acceptor site, (**e**) retained intron. [based on Zhou *et al*. (15) and Haas *et al*. (3).

Table 1. Numbers of TUs including alternatively spliced transcripts

Type of alternative splicing	Number of alternative splicing events	Percentage[a]
Exon skip type (ES-type)	273	1.8
Alternative terminal exon type (AT-type)	458	3.0
Alternative donor type (AD-type)	250	1.6
Alternative acceptor type (AA-type)	321	2.1
Retained intron type (RI-type)	790	5.2
Total	2092	13.8
Total (non-redundant)	1764	11.6

[a]Percentage of 15 214 multisequence TUs.

be due to immature mRNAs. In our data, there were 267 TUs (34%) in which RI-type alternative splicing events were supported by more than two transcripts. When using other EST data registered with the NCBI, the proportion of TUs including more than two transcripts with an unspliced intron was 65% (data not shown).

Change of alternative splicing profiles

We analyzed whether alternative splicing profiles change with environmental conditions or the developmental stages of plants. We collected RAFL cDNAs from 18 cDNA libraries prepared from *Arabidopsis* plants grown under various abiotic stress conditions (2). Thus, each RAFL cDNA clone was associated with a library (Table 2). Using this information, we could analyze the relationship between alternative splicing profiles and growth conditions. Significant changes of alternative splicing profiles were observed in several libraries.

From the analysis of ES-type alternative splicing (Figure 3a), we compared the measured and expected numbers of clones without alternatively spliced exons in each library. Libraries RAFL07 (cold-treated plants), RAFL15 (siliques and flowers), RAFL17 (rehydration) and RAFL21 (various stresses) had statistically significant differences ($P < 1.0 \times 10^{-7}$) between measured and expected values (Table 2 and Figure 4a). In those except RAFL21, the numbers of alternatively spliced clones without alternative exons were fewer than expected.

For AT-type alternative splicing, we compared the measured and expected numbers of clones with short termination events. Libraries RAFL07, RAFL21, RAFL06 (various developmental stages) and RAFL19 (siliques and flowers) had statistically significant differences between measured and expected values (Figure 4b). In RAFL06 and RAFL07, the observed numbers of clones with short termination sites were greater than expected. In libraries RAFL19 and RAFL21, the numbers were smaller than expected.

For AD-type and AA-type alternative splicing, we analyzed the number of clones with short exons. These two types of alternative splicing showed similar results. RAFL07 and RAFL21 had statistically significant differences (Figure 4c and d). In RAFL07 (cold-treated plants), the number of clones with short exons as a result of AD/AA-type alternative splicing was lower than expected. In contrast, in RAFL21 (various stresses), the number of clones with short exons was greater than expected.

For RI-type alternative splicing, we compared the measured and expected number of clones with at least one unspliced intron. RAFL04 (cold-treated plants), RAFL08 (dehydration-treated plants), RAFL09 (various developmental stages) and RAFL21 had statistically significant differences between the two values (Figure 4e). In all those except RAFL21, the numbers of alternatively spliced clones with unspliced introns were greater than expected.

These results suggest that alternative splicing profiles are significantly affected by environmental stress conditions. In particular, RAFL07 (cold-treated plants) and RAFL21 (various stresses) had very different alternative splicing profiles from those of other libraries (Figure 4).

DISCUSSION

This study presents a large-scale analysis of alternative splicing profiles in the *A.thaliana* transcriptome. We used 278 734 cDNA sequences obtained from our collection of *Arabidopsis* full-length cDNAs [(2,4) and Figure 1a]. We detected 1764 TUs with alternatively splicing transcripts out of 15 214 TUs with two or more sequences (Figure 2). We were able to detect many more TUs with alternative splicing events than reported previously [(6) and Figure 1b] because of the number of sequences we analyzed. Moreover, all the sequences used in this study had associated information on at least one terminal site, which means that cDNA sequences have more information than EST sequences. For example, we could determine both the 5′ initiation and 3′ termination sites of 13 580 out of 17 130 TUs mapped on the genome. The percentage of TUs with alternative splicing events was 11.6%. The percentages of alternatively spliced genes out of the detected transcripts are 42% in human (16), 41% in mouse (12) and ~10% in rice (17). Johnson *et al*. (18) reported that the percentage of alternative splicing events was 74% in human transcripts on the basis of a search of alternative splicing events using microarray analysis. The occurrence of alternative splicing events in *Arabidopsis* is similar to that in rice. The large difference in alternative splicing frequencies between higher animals and higher plants may cause differences in variations of transcripts.

Alternative splicing profiles were affected by environmental stress conditions at the transcriptome level (Figure 4 and Table 2). In particular, libraries RAFL07 (cold-treated plants) and RAFL21 (plants treated with various stresses) had statistically significant differences between the measured and expected numbers of alternatively spliced clones in many types of alternative splicing. Because RAFL21 included cDNA clones isolated from plants grown under various stress and hormone-treatment conditions, the relationship between alternative splicing and the exact environmental stress was not clear. However, on the basis of the data on alternative splicing events in RAFL07, we can discuss the effect of cold stress on alternative splicing.

We analyzed the alternative splicing variants of transcripts for splicing factors because alternative splicing profiles could be affected by splicing factors. In the *Arabidopsis* genome, there are 33 genes annotated as splicing factors (ftp://ftpmips.gsf.de/cress/). We obtained 26 TUs corresponding to these genes annotated as splicing factors; 13 contained splicing variation (Table 3). Among these, 11 were identified

Table 2. Alternative splicing events in each library

Library name	Plant materials	Number of DNA clones[a]	ES-type[b]		AT-type[b]		AD-type[b]		AA-type[b]		RI-type[b]	
			Measured	Expected	Measured	Expected	Measured	Expected	Measured	Expected	Measured	Expected
RAFL02	Rosette plants	331	4	1.8	0	3.4	8	1.6	4	1.2	17	5.8
RAFL03	Dehydration-treated plants	321	2	1.7	3	3.3	2	1.6	1	1.2	14	5.6
RAFL04	Cold-treated plants	1 138	18	6.2	21	11.8	7	5.6	7	4.2	46	19.8
RAFL05	Dehydration-treated plants	2 402	25	13.1	51	24.8	22	11.9	24	8.9	64	41.9
RAFL06	Plants at various developmental stages and those treated with dehydration and cold	6 005	27	32.7	106	62.1	18	29.7	31	22.4	151	104.7
RAFL07	Cold-treated plants	21 026	22	114.4	369	217.4	42	104.0	22	78.3	412	366.5
RAFL08	Dehydration-treated plants	2 608	17	14.2	46	27.0	8	12.9	11	9.7	88	45.5
RAFL09	Plants at various developmental stages and those treated with dehydration and cold	21 290	92	115.8	289	220.1	61	105.3	68	79.3	558	371.1
RAFL11	Plants at various developmental stages and those subjected to various stress [(dry, cold, NaCl, heat and ultraviolet (UV)] and ABA treatments. Plants grown under dark conditions. Silique tissues	1 900	12	10.3	39	19.6	6	9.4	3	7.1	29	33.1
RAFL12	Cold-treated plants	20	0	0.1	1	0.2	0	0.1	0	0.1	2	0.3
RAFL13	Dehydration-treated plants	68	0	0.4	1	0.7	0	0.3	0	0.3	3	1.2
RAFL14[c]	Roots	27 908	162	151.8	325	288.6	172	138.0	129	103.9	461	486.4
RAFL15	Siliques and flowers	12 534	17	68.2	167	129.6	31	62.0	25	46.7	242	218.4
RAFL16[c]	Dark-grown plants	29 892	154	162.6	300	309.1	128	147.8	112	111.3	487	521.0
RAFL17	Dehydration-treated plants—rehydration (after dry 10 h) treated plants	12 886	22	70.1	161	133.2	33	63.7	41	48.0	177	224.6
RAFL18	Cold-treated plants	1 138	3	6.2	25	11.8	3	5.6	3	4.2	25	19.8
RAFL19[c]	Siliques and flowers	45 418	267	247.1	241	469.6	233	224.6	142	169.1	631	791.6
RAFL21[c]	Plants treated with various stresses (heat and UV), hormones (ABA, auxin, ethylene, JA, SA, GA and cytokinin) and BTH treatments	33 329	354	181.3	132	344.6	315	164.8	197	124.1	431	580.9
	Total	220 214	1198		2277		1089		820		3838	
	Percentage of alternatively spliced clones[a]		0.54%		1.03%		0.49%		0.37%		1.74%	

Numbers with gray shading show a significant difference between the measured and expected values of alternative splicing events ($P < 1.0 \times 10^{-7}$).
[a]Expected values are calculated as (number of cDNA clones) × (percentage of alternatively spliced clones).
[b]Numbers of specific types of alternative splicing variants. Details are given in the text.
[c]The RAFL14, RAFL16, RAFL19 and RAFL21 libraries included the RAFL23, RAFL24, RAFL22 and RAFL25 libraries, respectively.

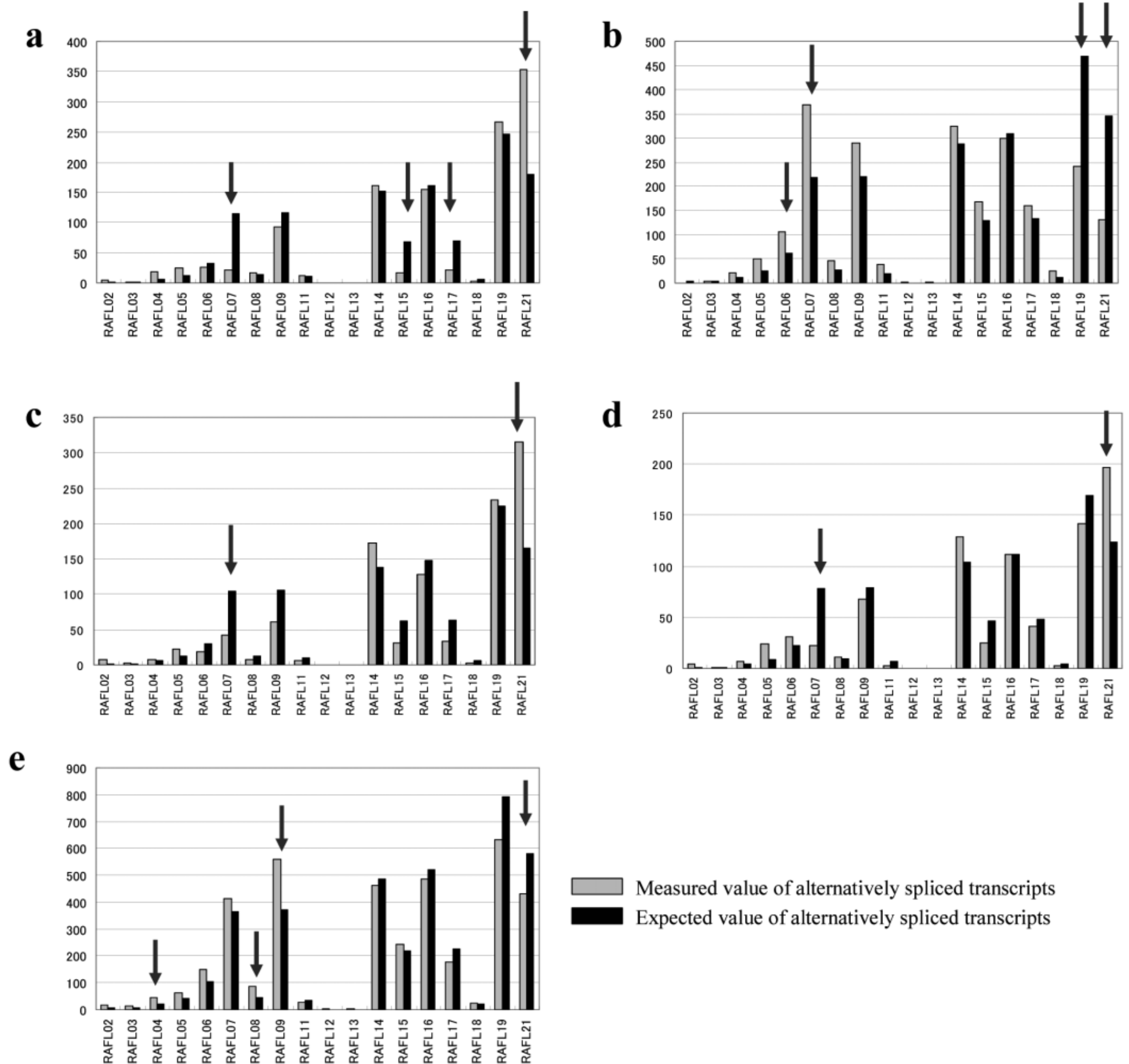

Figure 4. Comparison of the measured values and expected values of alternatively spliced transcripts in each library: (**a**) exon skip, (**b**) alternative terminal exon, (**c**) alternative donor site, (**d**) alternative acceptor site, (**e**) retained intron. Arrows indicate the libraries with statistically significant differences ($P < 1.0 \times 10^{-7}$) between measured values and expected values.

from RAFL libraries prepared from cold-treated plants (RAFL04, 07, 12, 18). Transcripts for splicing factors expressed under cold stress showed different splicing from those identified under other growth conditions. The transcript for splicing factor SR1 has splicing variants under cold stress. The SR1 transcript is alternatively spliced at an acceptor site of exon 11 (Figure 5). Splicing variants of SR1 expressed under cold stress (RAFL04-19-E10) have a long exon 11 as a result of the selection of the acceptor site (Figure 5). Lazar and Goodman (19) reported that the SR1 transcript has splicing variants, and that its alternative splicing is regulated by temperature increase. They suggested that alternative splicing profiles may be changed at the genomic level as a result of alternative splicing of SR1 at high temperature. SR1 is a homolog of human general/alternative splicing factor SF2/ASF (20), which can affect alternative splicing of several genes (21,22). These results suggest that the change in the splicing profile of SR1 by temperature change affects the splicing profiles of various transcripts.

In addition to SR1, transcripts of other splicing factors showed splicing variations under cold conditions (Table 3). Our microarray analyses [(23); data available at http://rarge.gsc.riken.jp/microarray/] revealed that some genes for splicing factors were induced (>2.5 times) under cold stress (gray cells in Table 3). Fourteen splicing factor genes on the 7k full-length cDNA microarray can be divided into two groups: a

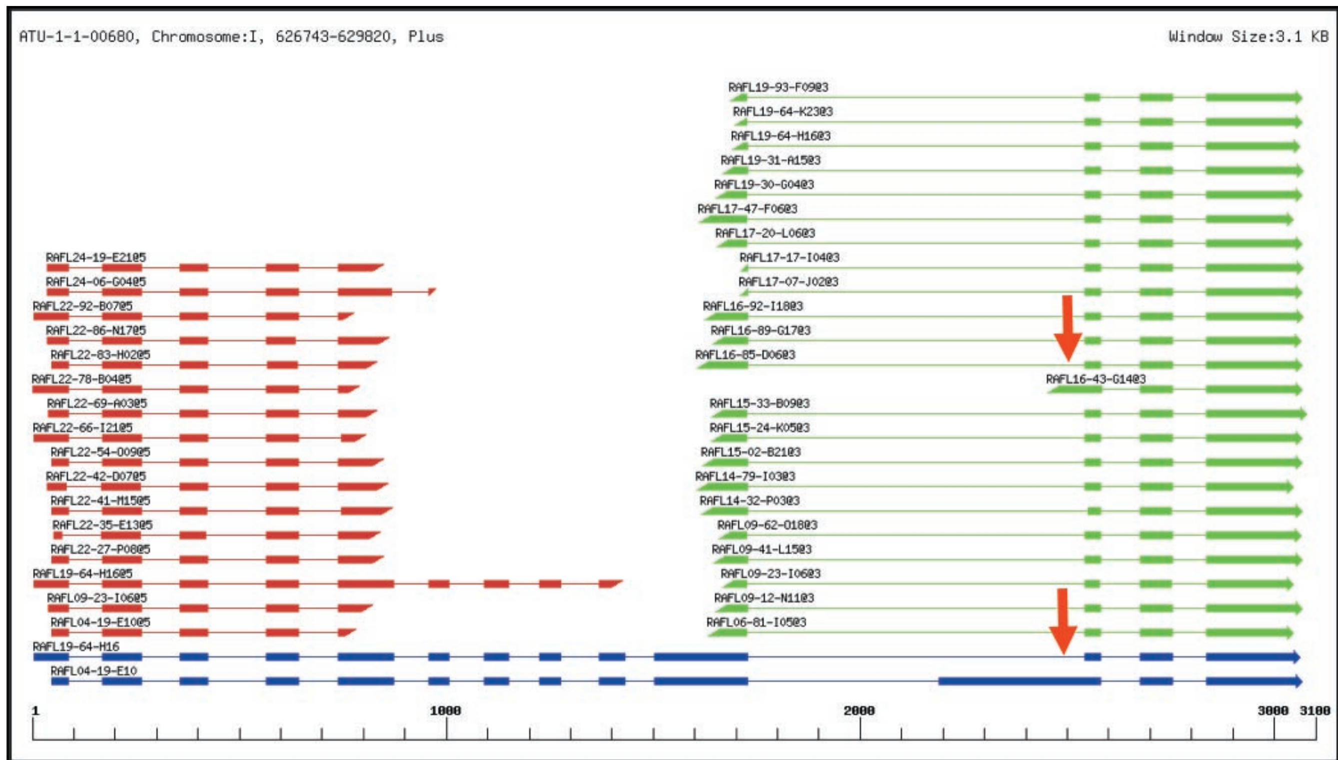

Figure 5. Splicing variations of the splicing factor SR1 transcripts. Full-length read sequences are colored in blue, 5′ terminal read sequences in red and 3′ terminal read sequences in green. AD-type alternative splicing events on the junction site between intron 10 and exon 11 are indicated by red arrows.

Table 3. Alternative splicing events in various splicing factor transcripts

TU name	RAFL clone name on microarray	Alternative splicing type	Clones from cold libraries	AGI code	Gene name
ATU-1-1-00680		AA	o	At1g02840	Splicing factor SR1
ATU-1-0-03240	RAFL08-12-I05		o	At1g09140	Putative SF2/ASF splicing modulator, SRP30
ATU-1-0-03900		AT	o	At1g10580	Putative pre-mRNA splicing factor
ATU-1-0-08320	RAFL05-12-O10			At1g23860	9G8-like splicing factor/SRZ-21
ATU-1-0-17180				At1g65660	Step II splicing factor, putative
ATU-1-1-01210	RAFL04-09-P15		o	At1g04510	Putative pre-mRNA splicing factor PRP19
ATU-1-1-04740	RAFL11-07-B22	AT,AD,RI		At1g14650	Splicing factor, putative
ATU-1-1-11690	RAFL11-09-L02	AT	o	At1g44910	Splicing factor, putative
ATU-2-0-02310		AT	o	At2g16940	Splicing factor-like protein
ATU-2-0-11510			o	At2g43770	Putative splicing factor
ATU-2-0-12560			o	At2g46610	Putative arginine/serine-rich splicing factor
ATU-2-0-12820	RAFL09-09-N06			At2g47250	Putative pre-mRNA splicing factor RNA helicase
ATU-2-1-00840			o	At2g03870	Putative snRNP splicing factor
ATU-2-1-04690			o	At2g24590	RSZp22 splicing factor like-protein
ATU-3-0-13490	RAFL05-17-P20	AA	o	At3g53500	Splicing factor-like protein
ATU-3-0-16270		ES	o	At3g61860	Arginine/serine-rich splicing factor RSP31
ATU-4-0-01200	RAFL05-18-E11		o	At4g03430	Putative pre-mRNA splicing factor
ATU-4-0-11380	RAFL08-12-N11	RI	o	At4g36690	Splicing factor-like protein
ATU-4-1-00880		AD,RI		At4g02430	Putative alternative splicing factor ASF protein (AT4g02430)
ATU-4-1-07370	RAFL04-10-E08	ES,AT	o	At4g25500	Splicing factor At-SRP40
ATU-4-1-09580				At4g31580	Splicing factor 9G8-like SR protein/SRZ-22
ATU-4-1-11640				At4g37120	Step II splicing factor-like protein
ATU-5-0-01650	RAFL04-12-N14	ES	o	At5g06160	Splicing factor 3a
ATU-5-0-02970	RAFL04-19-H09	ES,RI	o	At5g09880	Splicing factor-like protein
ATU-5-1-04250	RAFL08-17-F04			At5g13010	Pre-mRNA splicing factor ATP-dependent RNA helicase-like protein
ATU-5-1-14590	RAFL05-05-P06	ES	o	At5g52040	Arginine/serine-rich splicing factor RSP41 homolog

Expression of the RAFL members shaded with gray was induced (>2.5-fold) in response to cold stress as reported by Seki *et al*. (23).

group of eight genes that have alternative splicing events, and a group of six genes that have no alternative splicing events. Out of the eight genes with splicing variations, five were induced >2.5 times under cold stress. In contrast, out of the six genes without splicing variations, only one was induced under cold stress. These observations suggest a correlation between alternative splicing events and the cold-inducibility of splicing factor genes, and that both transcriptional and splicing regulation of splicing factor genes affects genome-wide alternative splicing profiles under cold stress. According to us, splicing factors change alternative splicing profiles at the transcriptome level by binding to exon–intron junctions for splicing. Further study of the relationship between alternative splicing and expressional and alternative splicing regulation of splicing factor genes is necessary to explain how splicing factors affect alternative splicing.

We showed a correlation between alternative splicing events and cold-stress conditions. This correlation may be due to the cold-inducible expression of splicing factor genes. This means that alternative splicing events in certain splicing factors affect alternative splicing profiles at the whole transcriptome level.

ACKNOWLEDGEMENTS

We thank Ms Asako Kamiya, Maiko Nakajima, Junko Ishida, Akiko Enjyu and Mari Narusaka for their excellent technical assistance. We also thank Dr Yoshihide Hayashizaki, Prof. Joseph R. Ecker, Athanasios Theologis and Ronald W. Davis for their collaboration, and Prof. Michiko Go for critical reading of the manuscript. This work is supported by the Genome Research Program of RIKEN.

REFERENCES

1. Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.
2. Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T., Nakajima,M., Enju,A., Akiyama,K., Oono,Y. *et al*. (2002) Functional annotation of a full-length *Arabidopsis* cDNA collection. *Science*, **296**, 141–145.
3. Haas,B.J., Volfovsky,N., Town,C.D., Troukhan,M., Alexandrov,N., Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L. (2002) Full-length messenger RNA sequences greatly improve genome annotation. *Genome Biol*., **3**, RESEARCH0029.
4. Yamada,K., Lim,J., Dale,J.M., Chen,H., Shinn,P., Palm,C.J., Southwick,A.M., Wu,H.C., Kim,C., Nguyen,M. *et al*. (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science*, **302**, 842–846.
5. Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W. *et al*. (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
6. Haas,B.J., Delcher,A.L., Mount,S.M., Wortman,J.R., Smith,R.K.,Jr, Hannick,L.I., Maiti,R., Ronning,C.M., Rusch,D.B., Town,C.D. *et al*. (2003) Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res*., **31**, 5654–5666.
7. Zhu,W., Schlueter,S.D. and Brendel,V. (2003) Refined annotation of the *Arabidopsis* genome by complete expressed sequence tag mapping. *Plant Physiol*., **132**, 469–484.
8. Shi,H., Xiong,L., Stevenson,B., Lu,T. and Zhu,J.K. (2002) The *Arabidopsis salt overly sensitive 4* mutants uncover a critical role for vitamin B6 in plant salt tolerance. *Plant Cell*, **14**, 575–588.
9. Macknight,R., Duroux,M., Laurie,R., Dijkwel,P., Simpson,G. and Dean,C. (2002) Functional significance of the alternative transcript processing of the *Arabidopsis* floral promoter FCA. *Plant Cell*, **14**, 877–888.
10. Yoshimura,K., Yabuta,Y., Ishikawa,T. and Shigeoka,S. (2002) Identification of a *cis* element for tissue-specific alternative splicing of chloroplast ascorbate peroxidase pre-mRNA in higher plants. *J. Biol. Chem*., **277**, 40623–40632.
11. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*., **25**, 3389–3402.
12. Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H. *et al*; FANTOM Consortium; RIKEN Genome Exploration Research Group Phase I & II Team. (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*, **420**, 563–573.
13. Usuka,J., Zhu,W. and Brendel,V. (2000) Optimal spliced alignment of homologous cDNA to a genomic DNA template. *Bioinformatics*, **16**, 203–211.
14. Florea,L., Hartzell,G., Zhang,Z., Rubin,G.M. and Miller,W. (1998) A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res*., **8**, 967–974.
15. Zhou,Y., Zhou,C., Ye,L., Dong,J., Xu,H., Cai,L., Zhang,L. and Wei,L. (2003) Database and analyses of known alternatively spliced genes in plants. *Genomics*, **82**, 584–595.
16. Modrek,B., Resch,A., Grasso,C. and Lee,C. (2001) Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res*., **29**, 2850–2859.
17. Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H. *et al*; Rice Full-Length cDNA Consortium; National Institute of Agrobiological Sciences Rice Full-Length cDNA Project Team; Foundation of Advancement of International Science Genome Sequencing & Analysis Group; RIKEN. (2003) Collection, mapping, and annotation of over 28 000 cDNA clones from *japonica* rice. *Science* **301**, 376–379.
18. Johnson,J.M., Castle,J., Garrett-Engele,P., Kan,Z., Loerch,P.M., Armour,C.D., Santos,R., Schadt,E.E., Stoughton,R. and Shoemaker,D.D. (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science*, **302**, 2141–2144.
19. Lazar,G. and Goodman,H.M. (2000) The *Arabidopsis* splicing factor SR1 is regulated by alternative splicing. *Plant Mol. Biol*., **42**, 571–581.
20. Lazar,G., Schaal,T., Maniatis,T. and Goodman,H.M. (1995) Identification of a plant serine-arginine-rich protein similar to the mammalian splicing factor SF2/ASF. *Proc. Natl Acad. Sci. USA*, **92**, 7672–7676.
21. Caceres,J.F., Stamm,S., Helfman,D.M. and Krainer,A.R. (1994) Regulation of alternative splicing *in vivo* by overexpression of antagonistic splicing factors. *Science*, **265**, 1706–1709.
22. Molin,M. and Akusjarvi,G. (2000) Overexpression of essential splicing factor ASF/SF2 blocks the temporal shift in adenovirus pre-mRNA splicing and reduces virus progeny formation. *J. Virol*., **74**, 9002–9009.
23. Seki,M., Narusaka,M., Ishida,J., Nanjo,T., Fujita,M., Oono,Y., Kamiya,A., Nakajima,M., Enju,A., Sakurai,T. *et al*. (2002) 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.