Molecular Characterization and Phylogeny of U2AF^{35} Homologs in Plants[1][W][OA]

Bing-Bing Wang and Volker Brendel*

Department of Genetics, Development and Cell Biology (B.B.-W., V.B.) and Department of Statistics (V.B.), Iowa State University, Ames, Iowa 50010

U2AF (U2 small nuclear ribonucleoprotein auxiliary factor) is an essential splicing factor with critical roles in recognition of the 3′-splice site. In animals, the U2AF small subunit (U2AF^{35}) can bind to the 3′-AG intron border and promote U2 small nuclear RNP binding to the branch-point sequences of introns through interaction with the U2AF large subunit. Two copies of U2AF^{35}-encoding genes were identified in Arabidopsis (*Arabidopsis thaliana*; atU2AF^{35}a and atU2AF^{35}b). Both are expressed in all tissues inspected, with atU2AF^{35}a expressed at a higher level than atU2AF^{35}b in most tissues. Differences in the expression patterns of atU2AF^{35}a and atU2AF^{35}b in roots were revealed by a promoter::β-glucuronidase assay, with atU2AF^{35}b expressed strongly in whole young roots and root tips and atU2AF^{35}a limited to root vascular regions. Altered expression levels of atU2AF^{35}a or atU2AF^{35}b cause pleiotropic phenotypes (including flowering time, leaf morphology, and flower and silique shape). Novel slicing isoforms were generated from FCA pre-mRNA by splicing of noncanonical introns in plants with altered expression levels of atU2AF^{35}. U2AF^{35} homologs were also identified from maize (*Zea mays*) and other plants with large-scale expressed sequence tag projects. A C-terminal motif (named SERE) is highly conserved in all seed plant protein homologs, suggesting it may have an important function specific to higher plants.

Splicing is an essential process in eukaryotic gene expression. The precise excision of introns from pre-mRNA requires a dynamically assembled RNA protein complex (spliceosome). Many proteins participate in intron and exon definition prior to the assembly of U1 and U2 small nuclear RNP (snRNP). U2AF is such a splicing factor. Before spliceosome assembly, U2AF binds to the polypyrimidine (Py) tract between the intron branch-point and the 3′-AG dinucleotide intron boundary to recruit U2 snRNP to the branch-point sequence. The U2AF protein is composed of a large subunit (U2AF^{65}) and a small subunit (U2AF^{35}; Zamore and Green, 1989), with U2AF^{65} binding directly to the Py tract (Zamore et al., 1992) and U2AF^{35} binding to the 3′-AG boundary (Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal, 1999). U2AF^{35} can promote binding of U2AF^{65} to the Py tract by interacting simultaneously with the U2AF^{65} and arginine-serine (SR) proteins (Zuo and Maniatis, 1996). It was also shown that the binding of U2AF^{35} can trigger events in spliceosome assembly in addition to stabilizing U2AF^{65} binding (Guth et al., 2001). In vivo studies in Drosophila revealed that U2AF^{38} is an essential factor because mutations in U2AF^{38} caused lethality and development defects (Rudner et al., 1996). In vitro studies, however, suggested that some pre-mRNAs, including human β-globin pre-mRNA and adenovirus major late pre-mRNA, do not require U2AF^{35} in splicing (Guth et al., 1999). These seemingly conflicting results indicate that U2AF^{35} may function in a substrate-specific manner.

Genes encoding U2AF^{35} were designated U2AF1. Single to multiple copies of U2AF1 were identified in fission yeast (*Schizosaccharomyces pombe*; Wentz-Hunter and Potashkin, 1996), worm (Zorio and Blumenthal, 1999), fly (Rudner et al., 1996), fish (Tassone et al., 1999), chicken (Pacheco et al., 2004), and mammals (Zhang et al., 1992; Tupler et al., 2001; Shepard et al., 2002). U2AF^{35} contains a pseudo-RNA recognition motif (ψRRM) flanked by two highly conserved $C_8C_5C_3H$ zinc fingers (Kielkopf et al., 2001, 2004). The ψRRM and two zinc fingers are essential for U2AF^{35} function (Webb and Wise, 2004). U2AF^{35} also contains a C-terminal RS domain (Zhang et al., 1992), which was found to be dispensable in vivo (Rudner et al., 1998). In mammals, a recently duplicated copy of U2AF1 encodes a 26-kD protein (U2AF^{26}; Shepard et al., 2002). U2AF^{26} lacks the C-terminal RS domain, but is still able to functionally substitute for U2AF^{35} in splicing (Shepard et al., 2002). An alternatively spliced isoform of the U2AF1 pre-mRNA is conserved from fish to human (Pacheco et al., 2004). The encoded protein has seven amino acid differences in the ψRRM,

[1] This work was supported in part by the National Science Foundation (grants DBI–0110189 and DBI–0321600) and the United States-Israel Binational Agricultural Research and Development Fund (research grant no. IS–3454–03).

* Corresponding author; e-mail vbrendel@iastate.edu; fax 515–294–6755.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Volker Brendel (vbrendel@iastate.edu).

[W] The online version of this article contains Web-only data.

[OA] Open Access articles can be viewed online without a subscription.

Article, publication date, and citation information can be found at www.plantphysiol.org/cgi/doi/10.1104/pp.105.073858.

but is still able to bind to U2AF^{65} and promote U2AF splicing activity in vitro (Pacheco et al., 2004). Multiple copies and isoforms of U2AF1 may contribute to the fine-tuned control of pre-mRNA splicing in vertebrates (Pacheco et al., 2004). In yeast (*Saccharomyces cerevisiae*), however, no ortholog of U2AF^{35} exists, although a functional ortholog of U2AF^{65} was identified (Mud2p; Abovich et al., 1994). No SR proteins were found either, possibly because the conserved branch consensus sequence eliminates the requirements for these factors in yeast.

Mammalian introns can be classified into AG-dependent and AG-independent types (Reed, 1989). In the AG-dependent introns, the Py tract is short (weak) and the adjacent 3′-AG boundary is essential for splicing (Reed, 1989). The function of U2AF^{35} is important in these introns because it stabilizes the binding of U2AF^{65} with the weak Py tract (Zuo and Maniatis, 1996). AG-independent introns, however, have a long (strong) Py tract and the AG is not required for splicing (Reed, 1989). U2AF^{65} alone is sufficient for recruiting the U2 snRNP to the branch-point sequence (Wu et al., 1999). Plant introns have neither conserved branch-point sequences nor a Py tract. Two U2AF^{65} homologs isolated from wild tobacco (*Nicotiana plumbaginifolia*) can complement the in vitro splicing of adenovirus pre-mRNA in HeLa cell extracts depleted of U2AF factor (Domon et al., 1998). Previous results revealed that Arabidopsis (*Arabidopsis thaliana*) has three copies of genes encoding the U2AF large subunit and one possible pseudogene (Wang and Brendel, 2004). These results suggest that the mechanism of 3′-splice site (ss) recognition is conserved in plants. Because plant introns lack strong branch-point and Py-tract motifs, it is very likely that 3′-ss recognition in plants relies more on U2AF^{35} to stabilize binding of the U2AF large subunit to pre-mRNA and to provide a bridge between the U2AF large subunit and the SR proteins.

Currently, very little is known about the detailed splicing mechanism in plants. Our recent survey revealed that most metazoan splicing factors are conserved and more than half of them are duplicated in plants (Wang and Brendel, 2004).We proposed that plants share the general splicing mechanism with metazoans but have distinct regulatory mechanisms (Wang and Brendel, 2004). Compared with nine SR proteins in humans, a total of 19 SR proteins were identified from Arabidopsis, including four families of novel SR proteins (Lazar et al., 1995; Lopato et al., 1996a, 1996b, 1999a, 1999b, 2002; Golovkin and Reddy, 1998, 1999; Kalyna and Barta, 2004; Wang and Brendel, 2004). Some novel SR proteins were found to be conserved in maize (*Zea mays*; Gupta et al., 2005), suggesting that these genes are possibly conserved in the plant kingdom. Recent studies using fluorescent protein tags revealed that SR proteins are dynamically distributed in nuclear speckles (Ali et al., 2003; Docquier et al., 2004; Fang et al., 2004; Lorkovic et al., 2004; Tillemans et al., 2005). The overall nuclear localization of Arabidopsis SR proteins is similar to each other and to animal homologs, although differences also exist (Tillemans et al., 2005). Overexpression of atSRp30 and atRSZ33 changes the alternative splicing pattern of some endogenous genes and causes morphological and developmental abnormalities (Lopato et al., 1999b; Kalyna et al., 2003). It is likely that atSRp30 and atRSZ33 may have important roles in ss selection. Because SR proteins function in 3′-ss recognition by binding to exonic splicing enhancers (Tian and Maniatis, 1993) and U2AF^{35} bridges SR protein binding to exonic splicing enhancers, with U2AF^{65} binding to the Py tract to stabilize the interactions (Zuo and Maniatis, 1996), it is of great interest to characterize the U2AF^{35} homologs to understand the mechanism of 3′-ss recognition in plants.

No U2AF^{35} homolog has been identified experimentally in plants. Database searches revealed two copies of potential U2AF^{35} genes in Arabidopsis (Domon et al., 1998; Lorkovic and Barta, 2002; Wang and Brendel, 2004). These two genes are highly conserved with their metazoan counterparts on the sequence level, indicating that their functions may also be similar. In this study, we report the experimental characterization of the two genes. Expression pattern differences and functional divergences were found to distinguish the genes. Computational identification of U2AF^{35} homologs in other plants revealed a highly conserved C-terminal motif specific to the plant clade of U2AF^{35} homologs.

RESULTS

Identification of Arabidopsis U2AF Small Subunit Homologs

We recently performed a genome-scale survey of splicing-related genes in Arabidopsis and identified two U2AF^{35} homologs (At1g27650 and At5g42820; Wang and Brendel, 2004). At1g27650 maps to the short arm of chromosome 1 and encodes a predicted polypeptide of 296 amino acids. At5g42820 maps to the long arm of chromosome 5 and encodes a predicted polypeptide of 283 amino acids. A third gene (At1g10320) shows significant similarity to human U2AF^{35} but is presumably the ortholog of a mammalian U2AF^{35}-related sequence, U2AF1-RS1 (Kitagawa et al., 1995). The gene nomenclature conventions of previous genome-scale studies were followed and atU2AF^{35}a and atU2AF^{35}b were used to represent At1g27650 and At5g42820, respectively. At the time of this study, 16 expressed sequence tag (EST) sequences and eight cDNA sequences could be aligned to the atU2AF^{35}a region (see displays at http://www.plantgdb.org/AtGDB). Spliced alignment of the full-length cDNAs reveals a 402-nucleotide (nt) intron in the 5′-untranslated region (UTR) of atU2AF^{35}a. Two ESTs (gi2757034 and gi5839839) indicate that the 5′-UTR intron may be retained in some tissues. For atU2AF^{35}b, only three ESTs were available,

and all matched to the 3′-end of the gene. One of the EST clones (accession no. AI997531) was ordered (Genome Systems) and sequenced from both directions. The full-length sequence identified the 5′-end of the atU2AF^{35}b gene, which includes a 277-nt intron in the 5′-UTR region. It also revealed a poly(A) tail at the 3′-end (which was missing in the original EST sequence). The sequence was deposited to GenBank (accession no. AF409140). As shown in Supplemental Figure 1, the atU2AF^{35}a and atU2AF^{35}b proteins contain most of the conserved domains of hsU2AF^{35}, including the ψRRM, one RS domain, two zinc fingers, and the two regions for interacting with the U2AF large subunit (Kielkopf et al., 2001). Both proteins lack the stretch of Glys present in hsU2AF^{35}. The sequences are overall 83% identical and each Arabidopsis homolog shows approximately 70% similarity to hsU2AF^{35} (BLAST2 alignments using default parameters).

Expression Patterns and Alternative Splicing of atU2AF^{35}

To verify the gene structure and check the expression patterns of the atU2AF^{35}a and atU2AF^{35}b genes, specific primers were designed from the 5′- and 3′-UTRs of both genes. Reverse transcription (RT)-PCR was conducted using RNAs extracted from Arabidopsis 7-d seedlings, leaf before flowering (LeafBF), leaf after flowering (LeafAF), meristem after flowering (MeriAF), root after flowering (RootAF), stem, flower, and silique tissues. As shown in Figure 1A, both atU2AF^{35}a and atU2AF^{35}b genes express in all these tissues. No clear intron retention product can be identified for either atU2AF^{35}a or atU2AF^{35}b, indicating that the 5′-UTR intron is spliced efficiently. Intriguingly, atU2AF^{35}a seems to have noncanonical introns in addition to the 5′-UTR intron. Several smaller bands in addition to the main product were observed in the RT-PCR for atU2AF^{35}a. Sequencing results revealed that two additional segments in the 3′-end could be removed from the main transcript. These additional segments are possible introns and named alternative intron (AltIntron) 1 and AltIntron2. AltIntron1 (287 nt) and AltIntron2 (345 nt) overlap with each other. The position of these introns is shown in Figure 1B. Neither of the AltIntrons is canonical; both have a repeat region flanking the intron-exon junction (AltIntron1, AGGAGCA; AltIntron2, AAAAC). Thus their real borders are difficult to determine. Independent RT-PCRs using different transcriptase and RNA preparations confirmed the existence of additional products. Splicing of AltIntron1 and AltIntron2 removes the coding sequences for the C terminus of atU2AF^{35}a protein. The truncated Arabidopsis U2AF^{35}a proteins retain the conserved N-terminal domains and a shortened SR domain. This domain structure is similar to U2AF^{26}, a duplicated copy of U2AF^{35} in human and mouse (Shepard et al., 2002). Whether the truncated atU2AF^{35}a proteins have physiological roles is unknown.

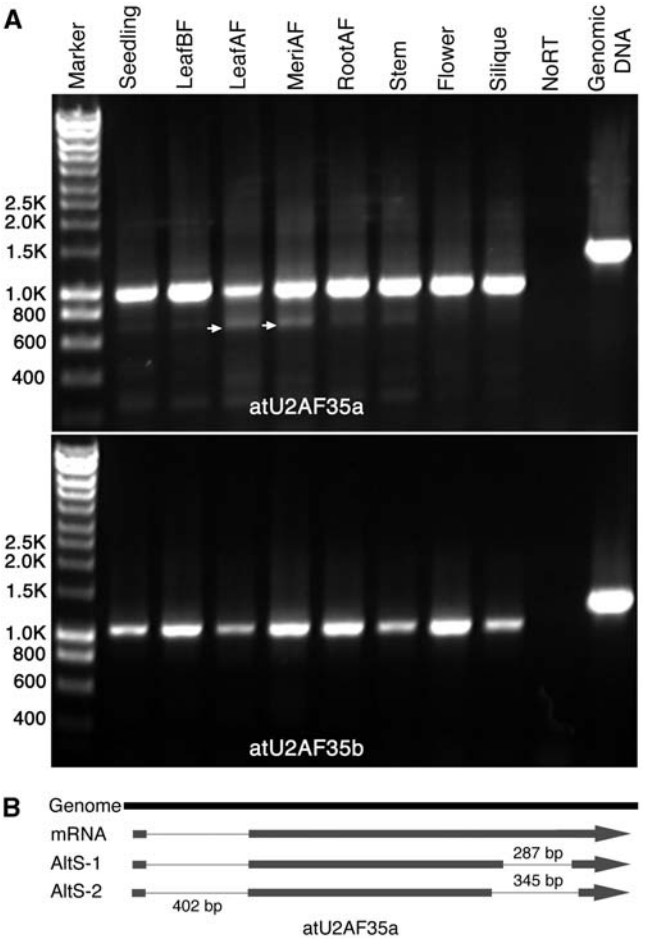

Figure 1. RT-PCR analysis of atU2AF^{35}a and atU2AF^{35}b transcripts. A, Gel pictures of RT-PCR results. RNA tissues include seedling, LeafBF, LeafAF, MeriAF, RootAF, stem, flower, and silique. NoRT is the negative control, where equal amounts of different tissue RNAs were mixed and used in the RT-PCR reaction without adding reverse transcriptase. Genomic DNA was used as positive control. The arrows point to noncanonical alternative splicing isoforms. B, Noncanonical alternative splicing pattern. Solid boxes represent exons and lines represent introns. Numbers indicate the sizes of the corresponding introns.

atU2AF^{35}a Has Higher Expression Levels Than atU2AF^{35}b in Most Tissues

To compare the expression level of atU2AF^{35}a and atU2AF^{35}b in different tissues, real-time RT-PCR was employed using primers designed from the 5′-UTR region for both genes. Because the two genes are very similar to each other at the nt level, it is not feasible to design gene-specific probes for northern analysis. To eliminate possible DNA contamination, the reverse primers were designed from the exon junctions of the 5′-UTR intron for both genes. A validation experiment using different dilutions of cDNAs confirmed that the primers for the atU2AF^{35}a and atU2AF^{35}b genes have similar amplification efficiency (data not shown). The expression of atU2AF^{35}a in LeafBF was arbitrarily selected as a calibrator and the other samples were compared with it to get relative expression levels.

As shown in Figure 2, real-time RT-PCR revealed that the atU2AF^{35}a transcript is more abundant than atU2AF^{35}b in all tissues tested. In flower, stem, silique, RootAF, and MeriAF, the atU2AF^{35}a transcript level is significantly higher than the atU2AF^{35}b level (t test, $P < 0.05$). atU2AF^{35}a expresses at a level over 2-fold higher than atU2AF^{35}b in these tissues. In seedlings and leaf tissues (LeafAF and LeafBF), the difference between atU2AF^{35}a and atU2AF^{35}b is not statistically significant. The atU2AF^{35}a level is less than twice that of atU2AF^{35}b in leaves. It seems that both atU2AF^{35}a and atU2AF^{35}b express at a relatively stable level in different tissues. Compared with the expression level in LeafBF, no significant difference was observed for atU2AF^{35}b in other tissues. For atU2AF^{35}a, only flower and MeriAF have significantly higher expression. From these results, we concluded that atU2AF^{35}a expresses in a level similar to or a little higher than atU2AF^{35}b in most tissues before flowering. After flowering, both genes have an increased level in leaves. The expression of atU2AF^{35}a increases significantly in meristem and flower, while expression of atU2AF^{35}b seems to be decreased in roots.

Promoter::β-Glucuronidase Assays Reveal Similarities and Differences between the Expression Patterns of atU2AF^{35}a and atU2AF^{35}b

RT-PCR results only revealed the expression patterns of atU2AF^{35}a and atU2AF^{35}b in major tissues. A promoter::β-glucuronidase (GUS) assay was employed to identify the expression patterns in more detail. For both atU2AF^{35}a and atU2AF^{35}b, three to five independent lines transformed with two promoter constructs (the long and short ones, atU2AF^{35}a [876 and 1,358 bp]; atU2AF^{35}b [555 and 982 bp]; see ''Materials and Methods'' and Supplemental Fig. 2) all revealed similar GUS staining patterns, indicating that the short promoter region is a functional unit. The strong promoter control (*Cauliflower mosaic virus* [CaMV] 35S) showed strong GUS expression throughout the plants. No GUS activities were detected in the negative control transgenic plants (data not shown).

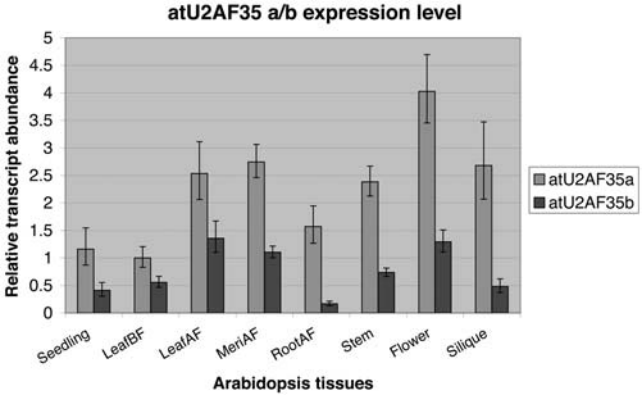

Figure 2. Expression levels of atU2AF^{35}a and atU2AF^{35}b. Real-time RT-PCR was performed to compare the expression levels of atU2AF^{35}a and atU2AF^{35}b in different tissues. The RNA tissues used are described in the legend to Figure 1. The level of atU2AF^{35}a in leafBF was arbitrarily chosen as the calibrator and other expression levels were compared with it. Relative expression level is indicated by bar heights. The thin lines above the bars represent the SE among three experimental repeats.

As shown in Figure 3, the GUS staining patterns are consistent with RT-PCR results and demonstrate that both atU2AF^{35}a and atU2AF^{35}b genes express in most tissues. The spatial and temporal expression of the two genes is similar in most tissues. Strong GUS activities are detected in 2- to 5-d seedlings (Fig. 3, A, F, and G). Shoot meristem, leaf primordia, and young leaves, including trichomes, show intense GUS activity (Fig. 3, B, C, and I–K). In large leaves, GUS expression is decreased and localized around vascular tissues of leaves (Fig. 3I). Adult leaf blades show weak GUS activity, while the vasculature and petioles have stronger GUS expression (Fig. 3, B and I). The most intense GUS activity is found in flowers. Flower buds, sepals, stamens, anthers, pollen, stigma, and the basal region of flowers all show strong GUS activity (Fig. 3, D and L–N). Petals show weak, but detectable, GUS expression. In siliques, the placenta and funiculus have strong GUS expression (Fig. 3E).

Differences between the expression of atU2AF^{35}a and atU2AF^{35}b were also discovered in flowers and young roots. As shown in Figure 3, M to Q, strong GUS activities were detected on the tops of pistils in control plants (CaMV 35S). The atU2AF^{35}a transformants also have detectable GUS expression. atU2AF^{35}b transformants, however, have no clear GUS activities. A distinct expression pattern was discovered in young roots. For atU2AF^{35}a, expression in young roots is limited to vascular regions (Fig. 3P). Root tips and hairs do not show clear GUS expression (Fig. 3S). The atU2AF^{35}b and CaMV 35S promoter drive strong GUS activities on the whole root (Fig. 3, Q and R), with the strongest expression in root tips (Fig. 3, T and U).

Both atU2AF^{35}a and atU2AF^{35}b Proteins Localize to the Cell Nucleus

As pre-mRNA splicing takes place in the nucleus, the atU2AF^{35}a and atU2AF^{35}b gene products should have nuclear localization if they are indeed splicing factors. atU2AF^{35}a and atU2AF^{35}b open reading frame (ORF) sequences were fused in frame downstream of the green fluorescent protein (GFP) coding sequence driven by the CaMV 35S promoter in gateway vector pMDC43 (Curtis and Grossniklaus, 2003). The unmodified pMDC43 was used as a control vector, with fluorescence detected in both the cell nucleus and the cytoplasm (Wang, 2005). Both atU2AF^{35}a and atU2AF^{35}b proteins are clearly enriched in nuclei and no differences were detected between them. As shown in Figure 4, strong green fluorescence was detected in the nuclei of root cells (Fig. 4C), leaf cells, guard cells (Fig. 4, D and E), and trichomes (Fig. 4F). Detailed study using confocal microscopy revealed that the distribution of atU2AF^{35}a and atU2AF^{35}b protein in

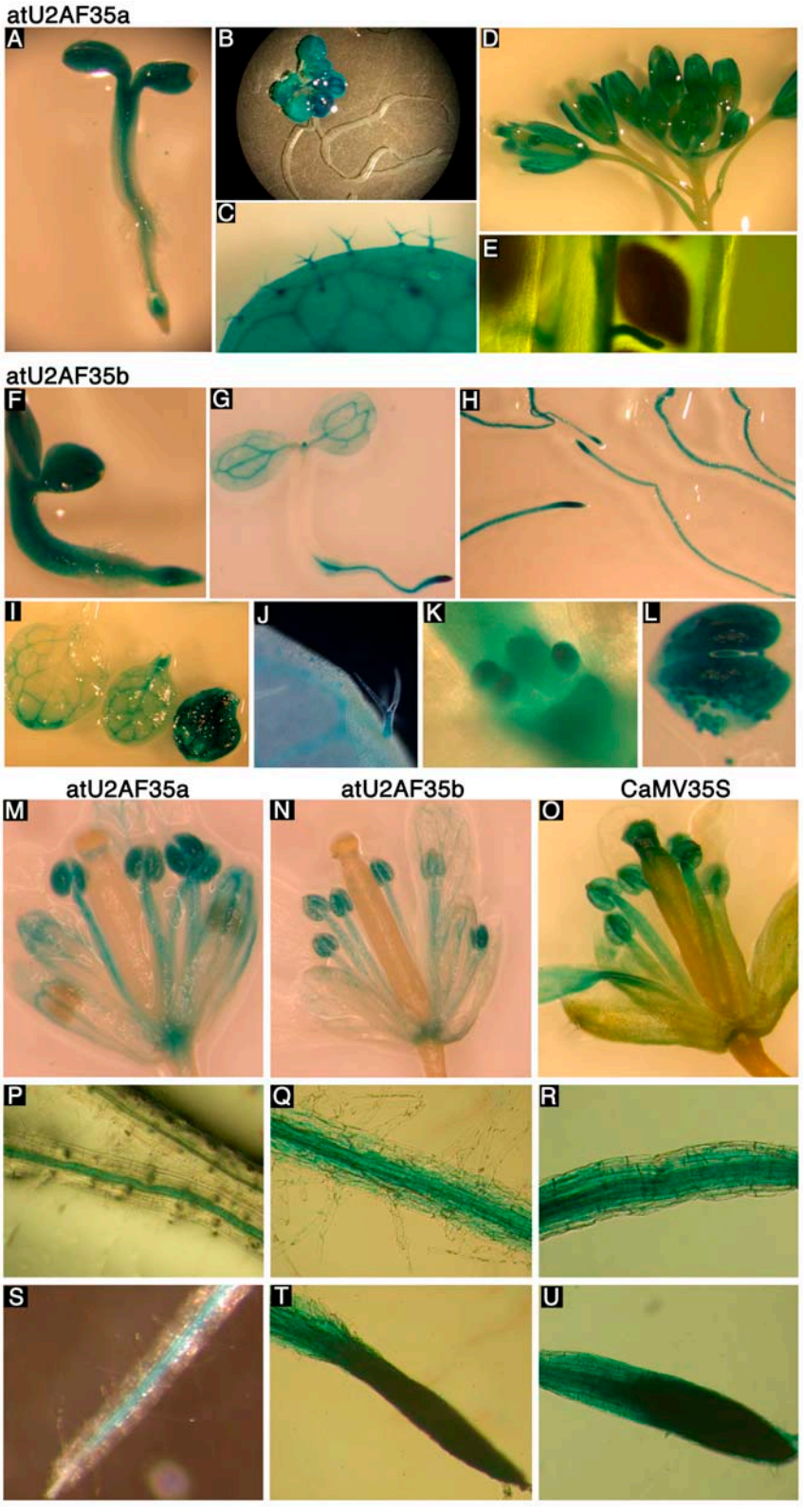

Figure 3. GUS staining patterns for atU2AF^{35}a and atU2AF^{35}b promoters. The tentative atU2AF^{35}a or atU2AF^{35}b promoters were linked to the GUS gene. A to E, M, P, and S, atU2AF^{35}a. F to L, N, O, and T, atU2AF^{35}b. O, R, and U, CaMV 35S promoter control. A, B, F, and G, Whole seedlings in different stages. C, I, and J, Leaves and trichomes. D, M to O, Flowers and inflorescences. E, Silique. K, Leaf primordia and meristem. L, Anther and pollen. H, P to U, Primary roots.

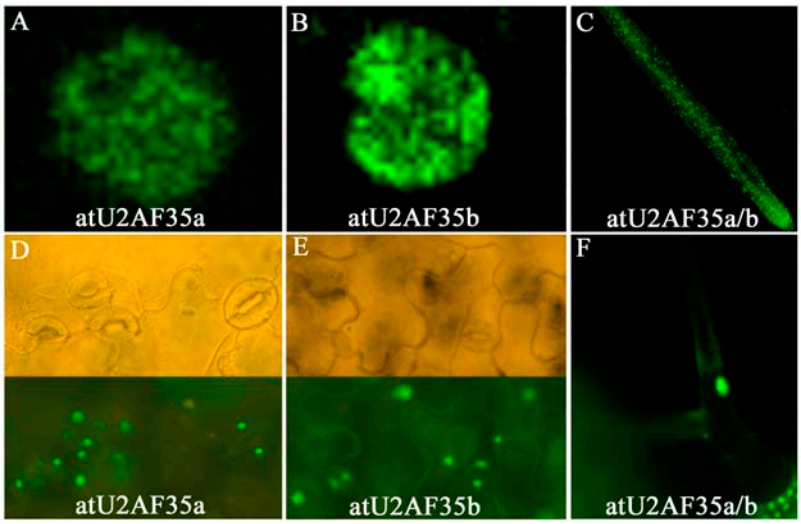

Figure 4. Cellular localization of atU2AF^{35}a and atU2AF^{35}b proteins. The ORF of atU2AF^{35}a or atU2AF^{35}b was fused with GFP in frame. Nuclear localization was observed for both constructs using a fluorescence microscopy unless denoted otherwise. A and D, GFP-atU2AF^{35}a. B and E, GFP-atU2AF^{35}b. C and F, Both GFP-atU2AF^{35}a and GFP-atU2AF^{35}b. A to C, Young roots. D and E, Leaf epidermal and guard cells, with top image taken without UV excitation and bottom image taken with excitation. F, Trichome. A and B are images taken by confocal microscopy and show the uneven distribution of atU2AF^{35}a and atU2AF^{35}b in the nucleus. The pattern is similar to the speckle distribution of SR proteins. A GFP control showed targeting to both nucleus and cytoplasm (data not shown).

the nucleus is not even (Fig. 4, A and B). They both are likely organized in nuclear speckles, a pattern similar to known SR proteins (Ali et al., 2003; Docquier et al., 2004; Lorkovic and Barta, 2004; Lorkovic et al., 2004; Tillemans et al., 2005).

Plants with Altered Expression Levels of atU2AF^{35} Genes Show Pleiotropic Phenotypes

One line with T-DNA inserted into the 5′-UTR intron (SALK_050678; Alonso et al., 2003) was identified for atU2AF^{35}a. No insertion line was found for the atU2AF^{35}b gene. To knock down the atU2AF^{35}b gene, antisense and RNA interference (RNAi) vectors were constructed and transformed into Arabidopsis (vector constructions shown in Supplemental Fig. 3). Real-time RT-PCR revealed that the expression level of atU2AF^{35}a is down-regulated by 2.5-fold in atU2AF^{35}a T-DNA insertion plants (Fig. 5F). The atU2AF^{35}b RNAi plants have the atU2AF^{35}b gene knocked down to 7-fold lower expression. The atU2AF^{35}b antisense plants surprisingly have a normal atU2AF^{35}b expression level, but the atU2AF^{35}a gene is up-regulated to 2-fold. The atU2AF^{35}b gene is down-regulated relative to atU2AF^{35}a in antisense plants. Pleiotropic phenotypes were observed in the mutants and transgenic lines. The homozygous atU2AF^{35}a T-DNA insertion plants and atU2AF^{35}b transgenic plants are all late flowering under both long- and short-day conditions. As shown in Figure 5A, wild-type Arabidopsis (Col-CS60000) flowers at 20 to 25 d with 11 to 12 leaves under continuous light conditions. The atU2AF^{35}a T-DNA plants flower at 25 to 30 d with 12 to 15 leaves. The atU2AF^{35}b RNAi and antisense plants flower at 26 to 32 d with 13 to 15 leaves. In addition to the late-flowering phenotype, all three mutants have shorter flowers with enlarged bottoms compared with wild type (Fig. 5B). A few atU2AF^{35}a and atU2AF^{35}b plants show an arrested primary stem at a certain stage (Fig. 5C, indicated by a red arrow), with branches below the terminus growing normally. Distinct silique shape was observed in atU2AF^{35}a T-DNA plants. As shown in Figure 5, D and E, normal siliques are cylindrical, tapering at the ends. In contrast, the siliques in atU2AF^{35}a T-DNA plants are flattened and widened at the distal end. Additionally, silique number is increased at the terminus of the stem in atU2AF^{35}a mutants. Leaf morphology changes were also observed in atU2AF^{35}a/b plants, with atU2AF^{35}a mutants showing larger, yellowish, and flatter leaves and atU2AF^{35}b transgenic plants showing smaller, dark green, and serrated leaves (Wang, 2005).

Noncanonical Splicing Patterns in Transgenic Plants

Because atU2AF^{35}a and atU2AF^{35}b presumably function in splicing, RT-PCR was carried out using total RNA from wild type, atU2AF^{35}a mutant, and atU2AF^{35}b transgenic lines to probe expression of 16 genes, including 12 genes predicted to be alternatively spliced, three genes (FLOWERING LOCUS C [FLC], FCA, and FPA) involved in the flowering pathway (Macknight et al., 1997; Sheldon et al., 1999; Schomburg et al., 2001), and AtDBR1 (At4g31770) as a constitutively spliced multiple-intron single-copy gene (Wang et al., 2004). No significant differences were observed for the 12 alternatively spliced genes and FPA among these plants. FLC expression levels are much higher in atU2AF^{35}b antisense plants and several minor PCR products were observed in AtDBR1 mutant and transgenic plants (Wang, 2005). The increased levels of FLC transcripts are consistent with the late-flowering phenotype in atU2AF^{35}b antisense plants. For FCA, four transcript isoforms (α, β, γ, and δ) are known (see Fig. 6B), of which FCA-γ is the only isoform producing full-length protein (Macknight et al., 1997). The γ and δ isoforms were detected in all lines using primers designed from the UTRs, al-

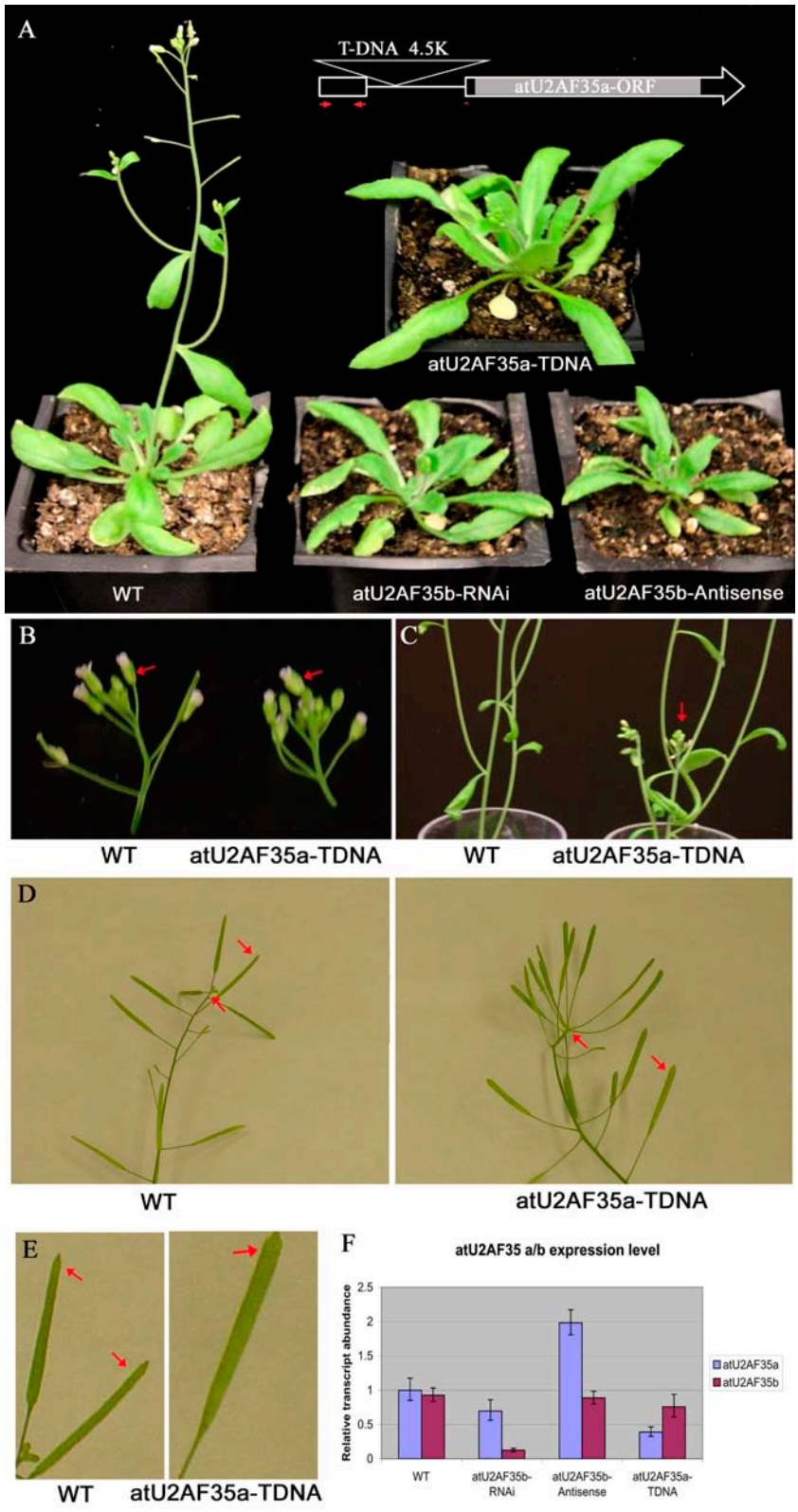

Figure 5. Phenotypes of atU2AF^{35}a mutant and atU2AF^{35}b transgenic lines. Pleiotropic phenotypes are observed in atU2AF^{35}a/b mutants. A, All atU2AF^{35}a and atU2AF^{35}b plants are late flowering. The gene structure and T-DNA insertion of atU2AF^{35}a are shown above the atU2AF^{35}a mutant. Boxes represent the exons and the thick line represents the intron. The gray rectangle denotes the protein coding region. The T-DNA insertion is indicated by a triangle. The red arrows indicate the positions of the real-time RT-PCR primers. Note that the reverse primer is flanking the exon junction. B, Flower shape changes in atU2AF^{35}a mutants as well as in atU2AF^{35}b plants. C, Main stem stops growing in some atU2AF^{35}a mutants. Some atU2AF^{35}b plants also show the same phenotype. D and E, Shape of siliques is changed in atU2AF^{35}a mutants. F, Real-time RT-PCR results on atU2AF^{35}a/b plants and wild type. The level of atU2AF^{35}a in wild type was chosen as a calibrator. Bars indicate relative expression levels; thin lines above the bars represent SEs.

though the δ isoform was clearly visible only in the atU2AF^{35}b antisense (column BA in Fig. 6A) plants. As shown in Figure 6A, extra bands different from the known isoforms were produced from all mutant and transgenic lines. Three of these were cloned and sequenced, including two bands (FCA-m1 [1.4 kb] and FCA-m2 [1.2 kb]) from atU2AF^{35}b antisense and one band (FCA-m3 [547 bp]) from atU2AF^{35}b RNAi plants. Sequence alignment of the three isoforms against the genome revealed four novel introns (SI1–4) shown in Figure 6, B and C. All these introns have noncanonical intron borders, with repeated segments on both intron-exon junctions. The sequences of the repeated junctions are distinct for the different introns. Multiple exons were skipped in all novel sequenced FCA isoforms by splicing of a rather long (>4.5 kb) noncanonical intron. FCA-m1 also retained the sixteenth intron. These novel FCA isoforms demonstrate that the splicing pattern of some genes can be changed in plants with altered atU2AF^{35}a/b expression levels.

Motifs and Molecular Phylogeny of Plant U2AF^{35} Homologs

Two U2AF^{35} homologs are known in rice (*Oryza sativa*; Domon et al., 1998). We identified maize orthologs of the two rice genes by RT-PCR (Wang, 2005). All these monocot homologs have introns in the 5′-UTR, which can be alternatively spliced (Wang, 2005). By searching the EST sequences in other plant species, 15 more full-length homologs were identified, including three homologs in wheat (*Triticum aestivum*), two homologs in barley (*Hordeum vulgare*), soybean (*Glycine max*), tree cotton (*Gossypium arboretum*), and potato (*Solanum tuberosum*), and one homolog in Medicago (*Medicago truncatula*), tomato (*Lycopersicon esculentum*), pine (*Pinus taeda*), and unicellular green algae (*Chlamydomonas reinhardtii*). Sequences of these plant homologs are available in Supplemental File 1. As shown in Supplemental Figure 1, all these U2AF^{35} homologs contain a degenerate RNA-binding domain, two CCCH-type zinc fingers, and one RS domain of variable length. A plant-specific conserved region was observed in the C-terminal regions. As shown in Figure 7A, U2AF^{35} homologs from monocot plants, dicot plants, and loblolly pine contain a conserved stretch of 16 amino acids in their C terminus. The consensus sequence is SEERRA(K/R) IEQWNRERE, where underlined amino acids are completely conserved in higher plants. For the (K/R) site, all monocot homologs, except osU2AF^{35}a, have K, while most dicot homologs have R. We refer to this region as the SERE motif. Four homologs in wheat, barley, potato, and tomato (taU2AF^{35}a, hvU2AF^{35}p, stU2AF^{35}b, and leU2AF^{35}b) have two SERE motifs in their C terminus (Wang, 2005). The U2AF^{35} homolog in unicellular green algae (crU2AF^{35}a) does not contain the SERE motif, indicating a possible function of the motif in higher plant-specific splicing mechanisms.

A phylogenetic tree was constructed based on the multiple sequence alignment of the N-terminal regions upstream of the RS domain shown in Supplemental Figure 1. As shown in Figure 7B, all plant homologs cluster into one group and all animal homologs cluster into another group. In the plant group, the seed plant homologs separate into four clades, including two monocot and two dicot clades. The green algae homolog appears as an outgroup. zmU2AF^{35}b (maize), osU2AF^{35}b (rice), hvU2AF^{35}a (barley), taU2AF^{35}b and taU2AF^{35}c (wheat) group into the monocot I clade, and zmU2AF^{35}a, osU2AF^{35}a, hvU2AF^{35}p, and taU2AF^{35}a group into the monocot II clade. For dicot plants, atU2AF^{35}a and atU2AF^{35}b (Arabidopsis),

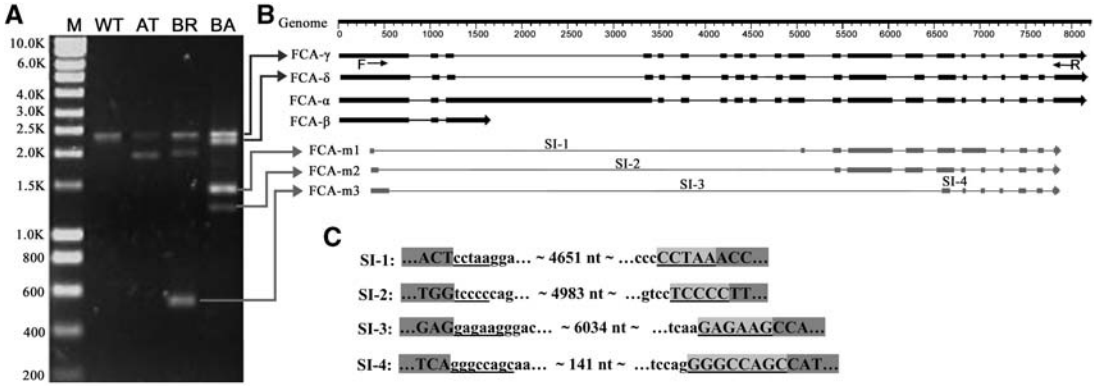

Figure 6. Novel splicing isoforms generated for the FCA gene in plants with altered levels of atU2AF^{35}. A, RT-PCR results. M, Marker; WT, wild type; AT, atU2AF^{35}a-TDNA plants; BR, atU2AF^{35}b RNAi plants; BA, atU2AF^{35}b antisense plants. Three bands indicated by the gray connectors were cloned and sequenced. B, Spliced alignment of FCA transcripts against the genome sequence. Boxes indicate exons and lines indicate introns. The four known FCA isoforms are represented in black and novel splicing isoforms are represented in gray. Vertical bars on the intron-exon junctions indicate ss different from the ones used in the FCA-γ isoform. Black arrows indicate the forward (F) and reverse (R) primer positions. C, Junctions of noncanonical introns. Capitalized letters indicate exon sequences and lower case letters indicate intron sequences. The unambiguous exon sequences are highlighted in gray and uncertain exon sequences are highlighted in light gray. Repeated segments are underlined.

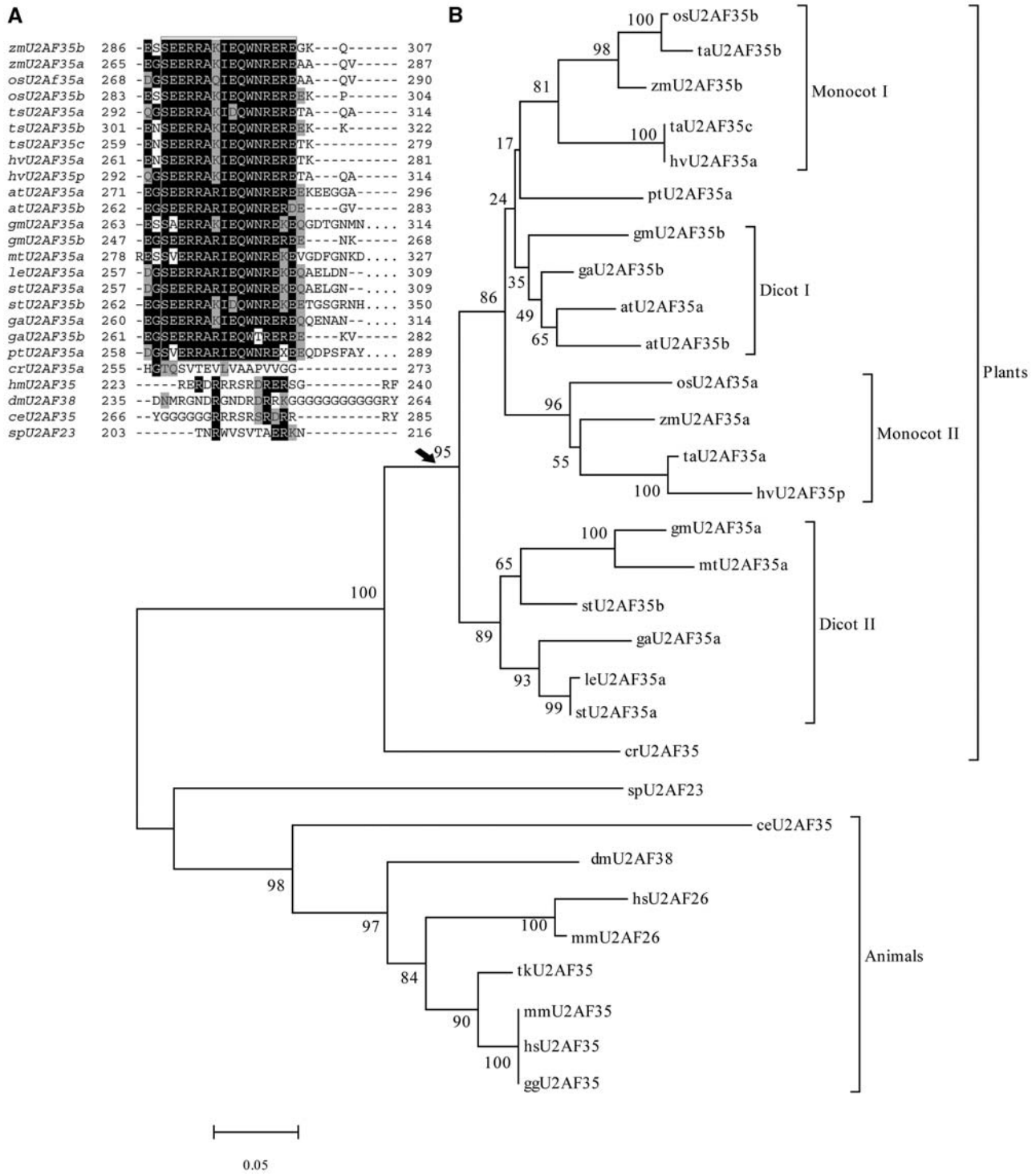

Figure 7. Phylogeny of U2AF^{35} homologs. A, C-terminal alignment of plant U2AF^{35} homologs. The SERE motif is indicated by a line above the alignment. Conserved amino acid residues are highlighted in black. Residues not identical, but similar to the conserved one, are highlighted in gray. The number at the end of each sequence denotes the protein length. B, Phylogenetic tree of U2AF^{35} homologs. Branch lengths indicate distances. Numbers on the branch are bootstrap values of confidence in the displayed branches. zm, Maize; at, Arabidopsis; os, rice; ta, wheat; hv, barley; gm, soybean; mt, *Medicago truncatula*; le, tomato; st, potato; ga, cotton; pt, pine; cr, unicellular algae; hs, human; dm, *Drosophila melanogaster*; ce, nematode; sp, fission yeast; mm, mouse; tk, bony fish; gg, chicken.

gmU2AF^{35}b (soybean), and gaU2AF^{35}b (tree cotton) group into the dicot I clade, and the remaining dicot homologs cluster into the dicot II clade. The dicot I clade is clustered in a big group with the monocot clades and ptU2AF^{35}a (from lobolly pine), indicating the ancient form of U2AF^{35} in the ancestor of seed plants. It seems that the ancient U2AF^{35} gene likely had introns in the 5′-UTR, as suggested by the gene structure of known homologs in Arabidopsis, rice, and maize. atU2AF^{35}a and atU2AF^{35}b both fall into the dicot I clade, indicating a recent duplication event in Arabidopsis.

DISCUSSION

Expression and Function of U2AF^{35} Homologs in Arabidopsis

Two homologs of U2AF^{35} were characterized in Arabidopsis in this study. The protein sequences and domain structures of atU2AF^{35}a and atU2AF^{35}b are very similar to human U2AF^{35}, indicating that the two proteins have exchangeable functions similar to their mammalian homologs. Plants with altered levels of atU2AF^{35}a or atU2AF^{35}b showed similar phenotypes, including late flowering and flower and leaf morphology changes, suggesting that the two genes have redundant functions. Both atU2AF^{35} genes express in all major tissues, indicating that the two homologs may function simultaneously in many cases. atU2AF^{35}a expresses at a higher level than atU2AF^{35}b in most tissues. In some tissue (e.g. root tip), however, atU2AF^{35}b expresses strongly, while atU2AF^{35}a is barely expressed, as suggested by our promoter::GUS assay. Alteration of the levels of atU2AF^{35}a and atU2AF^{35}b caused different novel splicing isoforms in FCA in the atU2AF^{35}a mutant and atU2AF^{35}b transgenic lines, while most other investigated genes were not affected. This suggests that, although most pre-mRNA splicing may not depend on the precise balance of atU2AF^{35}a and atU2AF^{35}b expression levels, some genes may require a finely tuned balance for correct splicing.

Noncanonical Alternative Splicing of U2AF^{35} and Other Splicing Factors

RT-PCR on atU2AF^{35}a genes revealed two extra bands in addition to the constitutively spliced product. Noncanonical alternative splicing events were identified by sequencing the extra products, which in each case removes a segment with repeated borders from the second exon of the atU2AF^{35}a gene. We also found that three extra FCA isoforms were produced by excision of noncanonical introns with repeated junctions in our atU2AF^{35} mutant and transgenic lines. Very likely the change of atU2AF^{35} levels will lead to the usage of noncanonical sites in some pre-mRNAs. It is interesting to note that all these noncanonical introns have repeated junctions. It may be possible that some protein homodimer is involved in symmetric ss recognition when U2AF^{35} levels are abnormal. As the expression level of atU2AF^{35}a and atU2AF^{35}b changes dynamically during growth and development, it is likely that atU2AF^{35}a can autoregulate the level of full functional protein by the noncanonical alternative splicing events. In vertebrates, the U2AF1 gene can also be alternatively spliced by inclusion of an additional exon, producing an isoform with seven amino acid differences in the ψRRM (Pacheco et al., 2004). Two maize SR proteins (ZmSRp31A and ZmSRp31B) also show noncanonical alternatively spliced introns (Gupta et al., 2005), strongly suggesting that many splicing factors can posttranscriptionally regulate their expression by noncanonical alternative splicing.

Evolution of U2AF^{35} Protein Domains and U2AF^{35} Gene Structures

Sequence alignment and phylogenetic analysis demonstrated that the U2AF^{35} gene exists in the ancestor of eukaryotic organisms. The ancient U2AF^{35} contained at least a ψRRM and two CCCH zinc fingers. It may also contain a run of Glys because the animal homologs and some monocot plant homologs have this motif. But this motif was lost in some homologs. In the plant kingdom, a C-terminal motif (SERE) was acquired after the divergence of green algae and probably before the divergence of seed plants. The SERE motif may have plant-specific functions, such as recognizing plant-specific splicing signals or interacting with plant-specific SR proteins. There might be another ancient U2AF^{35} gene in plants, as suggested by the separation of the dicot II clade from the dicot I-monocot-ptU2AF^{35}a group, which was probably lost in monocot and some dicot (Arabidopsis) lineages. After the divergence of monocot and dicot plants, individual duplications of U2AF^{35} genes happened in the ancestor of monocot plants, nightshade family (potato and tomato), Triticeae (wheat and barley), and Arabidopsis. The ancient plant U2AF^{35} gene may have had introns in the 5′-UTR, as suggested by the fact that the U2AF^{35} homologous genes in Arabidopsis, maize, and rice all have introns in the 5′-UTR. None of them have introns in either coding region or the 3′-UTR. Interestingly, the 5′-UTR intron is alternatively spliced in maize and rice homologs and the alternative splicing patterns and some intron segments are well conserved between the two monocot species (Wang, 2005), suggesting important biological functions of the 5′-UTR intron and the alternative splicing events in gene regulation.

Roles of Multiple Copies of U2AF^{35}

Plant introns have neither conserved branch sequences nor a Py tract. The 3′-ss recognition in plants will probably rely more on U2AF than in mammals. What the exact roles of plant U2AF^{35} homologs are and how they achieve their functions are challenging questions. Similar to mammalian introns, some plant

introns may be AG independent and may not require U2AF^{35} for correct splicing. Multiple copies of U2AF^{35} exist in nearly all higher plant genomes. Different U2AF^{35} homologs may preferably interact with different U2AF large subunits to form different U2AF heterodimers, as suggested by variations in the interacting regions of U2AF^{35} from the same species as well as different species (Wang, 2005). Also the RS domain and its surrounding regions of U2AF^{35} contain many variations, indicating the flexibility of interaction between U2AF^{35} and variable SR proteins. It is likely that some SR proteins may preferably interact with one of the U2AF^{35} proteins, which in turn interacts with specific U2AF large subunits. The balance among different U2AF-SR protein complexes on pre-mRNAs may be important to the splicing of particular introns. Disruption of this balance will produce abnormal splicing isoforms and reduce the abundance of normal splicing isoforms, as demonstrated in FCA genes in our mutant and transgenic lines. Thus we would expect to see similar phenotypes independent of which U2AF^{35} gene was disrupted, as observed in our mutant and transgenic lines.

Connection between Splicing and Physiological Responses in Plants

Both atU2AF^{35}a and atU2AF^{35}b mutant and transgenic lines are late flowering, with flower and leaf morphology changes. RT-PCR revealed that the FLC gene is up-regulated in atU2AF^{35}b antisense mutants (Wang, 2005). How exactly does the altered expression level of a splicing factor (atU2AF^{35}a/b) change the expression of flowering pathway genes? A possible explanation is that the production of novel FCA isoforms either reduced the level of the normal FCA isoform or possibly intervened with normal FCA isoform function, and, therefore, the FLC expression is increased. Recently, mutations disrupting the large subunit of cap-binding protein (CBP80/ABH1) were demonstrated to suppress the FRIGIDA-mediated increased FLC mRNA levels (Bezerra et al., 2004), providing another link between a splicing factor and flowering time. Knockout of the small subunit of cap-binding protein (CBP20) causes delayed development and serrated leaves (Papp et al., 2004), a phenotype very similar to our atU2AF^{35}b mutants (Wang, 2005). Both cbp20 and cbp80/abh1 mutants also confer hypersensitivity to abscisic acid. In addition, overexpression of an SR protein kinase in Arabidopsis caused delayed flowering and leaf morphology changes (Savaldi-Goldstein et al., 2003). Altered expression level of some other splicing-related genes will affect the response to drought (Xiong et al., 2001), cold (Gong et al., 2005), and salt resistance (Forment et al., 2002) in plants. It seems likely that many more connections between pre-mRNA splicing and plant responses to stress and environment are yet to be discovered.

MATERIALS AND METHODS

Identification of U2AF^{35} Homologs from Plants with Large Collections of EST Sequences

Plant U2AF^{35} homologs were identified by matching Arabidopsis (*Arabidopsis thaliana*) homologs against plant EST sequences by BLAST. All plant ESTs were downloaded from the National Center for Biotechnology Information (NCBI) Plant Genomes Central (http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/PlantList.html). Hits with E-value less than 10^{-04} were regarded as significant. All hits were then used to match the EST sequences again. The same criterion was used to retrieve all related ESTs. CAP3 (Huang and Madan, 1999) was then used to cluster these related ESTs and make contigs. The contigs were matched against all known U2AF^{35} sequences and against Arabidopsis proteins using BLASTx. Contigs were regarded as U2AF^{35} homologs if (1) they had an E-value of less than e^{-15} when matched against U2AF^{35} homologs; and (2) their best hit in Arabidopsis is a U2AF^{35} homolog. Putative proteins were translated from the contigs by the NCBI-ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html).

Sequence Alignment and Phylogenetic Analysis

Multiple sequence alignments of the U2AF^{35} proteins were generated with ClustalW using default parameters (Thompson et al., 1994). The alignments were visualized using the BioEdit program (version 5.0.9; http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Phylogenetic analysis of the sequences was conducted using MEGA software (version 2.1; http://www.megasoftware.net; Kumar et al., 2001). The phylogenetic trees were constructed using the neighbor-joining method with a bootstrap test. The distance model used was a Kimura 2-parameter. All other parameters used were default.

Arabidopsis Growth Conditions and RNA Extraction

Arabidopsis seeds were sown in soil and grown at 4°C for 4 d, and then the plants were moved to a growth room and grown at 22°C with continuous light. Total plant RNA was isolated using either TRIzol reagent (Invitrogen) or the Plant RNeasy mini kit (Qiagen) from 0.1 to 0.2 g of different tissues. The manufacturer's protocol was followed. For Arabidopsis, root, leaf, meristem, stem, and flower tissues from wild-type ecotype Columbia were used. Total RNA was dissolved in 30 μL diethyl pyrocarbonate-treated water and saved at −20°C.

RT-PCR and Real-Time RT-PCR

The primer sequences are described in Supplemental Table I. Total RNA was treated by RQ1 RNase free DNase according to the manufacturer's protocol (Promega). Two micrograms of treated RNA were then used for first-strand synthesis and PCR according to the manufacturer's protocol (Invitrogen). A mixture of treated RNAs was used as a no-RT control. For real-time RT-PCR, PRIMER EXPRESS version 2.0 software (Applied Biosystems) was used to design oligonucleotide primers. cDNAs were prepared as described above and diluted 600-fold for amplification of the 18S ribosome RNA gene and 3-fold for other genes. One microliter of diluted cDNA was used in a 25-μL reaction with SYBR Green master mix (Applied Biosystems). All reactions were performed in triplicate by using a Prism 5700 sequence detection system (Applied Biosystems). The experiments were replicated twice using different RNA samples. Primer efficiency was checked for each primer pair by constructing a standard curve using an equal mixture of all cDNAs (Applied Biosystems). The expression level of each gene was calculated based on $2^{-\Delta\Delta ct}$ method described in user bulletin number 2 (Applied Biosystems). The relative amount of calculated message was normalized to the level of the 18S rRNA gene.

Promoter::GUS Assay

Two potential promoter regions together with the 5′-UTR region were checked for both atU2AF^{35}a and atU2AF^{35}b. For atU2AF^{35}a, promoter 1 (PGaa1) is the genomic region from 876 nt before the ATG start codon, and promoter 2 (PGaa2) is from 1,358 nt before the start codon. For atU2AF^{35}b, promoter 1 (PGab1) is from 555 nt before the start codon, and promoter 2

(PGab2) is from 982 nt before the start codon. As shown in Supplemental Figure 2, both longer promoters for atU2AF^{35}a and atU2AF^{35}b include part of the first exon of the neighboring gene. These tentative promoters were amplified from Arabidopsis genome DNA. Primers are shown in Supplemental Table I. PCR products were purified and ligated to vector pCAMBIA1381z. The vectors were subjected to sequencing from both ends to make sure the insertions were correct. In addition to the two promoters for each gene, the CaMV 35S promoter was linked with the GUS gene and used as a strong promoter control (PGxx). The empty pCAMBIA1381z (no promoter) was used as a no-promoter control (negative control, PG_{00}). These constructs are all shown in Supplemental Figure 2. The right vectors were used for Arabidopsis transformation by methods described below. Three to five individual transgenic plants from each transformation were subjected to histochemical GUS assays, following the protocol described in Weigel and Glazebrook (2002).

GFP, Antisense, and RNAi Vector Construction

GFP vectors were constructed based on vector pMDC43 (Curtis and Grossniklaus, 2003). The ORFs of atU2AF^{35}a and atU2AF^{35}b genes were amplified and cloned into the pENTR vector, then excised and transferred to the pMDC43 vector using Gateway technology (Invitrogen). Antisense vectors were constructed based on binary vector pCAMBIA1301. The vector diagrams are shown in Supplemental Figure 3. The ORF of atU2AF^{35}b was amplified by PCR, using primers described in Supplemental Table I. The PCR product was isolated, digested, and inserted into the downstream of the CaMV 35S promoter in the reverse direction. To construct an RNAi vector, the atU2AF^{35}b ORF was amplified using the primer set shown in Supplemental Table I. The PCR product and antisense vector were digested by *Mlu*I and *Bsr*GI (New England Biolabs), then ligated by T4-DNA ligase (Promega). The RNAi construct uses the atU2AF^{35}b ORF to replace part of the antisense vector sequence in the sense direction. The resulting transcript forms a hairpin structure that triggers silencing of the endogene.

Arabidopsis Transformation

Different vectors were transformed into Agrobacterium by electroporation methods. Arabidopsis ecotype Columbia was transformed by Agrobacterium using the floral-dip method (Weigel and Glazebrook, 2002). Seeds were screened at 0.8% Arabidopsis selective medium containing 50 μg/mL hygromycin for 7 d, then transformed to 1.5% plates for another 7 d. Resistant plants were transferred to soil and analyzed.

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers AF409139 and AF409140.

ACKNOWLEDGMENTS

We would like to thank Wei Huang, Tiffanie Kuhn, and Zhen Ni Li for help with experiments. We are also grateful to Wei Huang, Dr. Philip Becraft, Dr. Robert Fluhr, and Dr. Thomas Peterson for critical reading of the manuscript. Microscopy pictures were taken in the laboratories of Dr. Shuizhang Fei and Dr. Jo Anne Powell-Coffman and in the Iowa State University Confocal Microscopy Facility.

Received November 7, 2005; revised November 7, 2005; accepted November 26, 2005; published January 11, 2006.

LITERATURE CITED

Abovich N, Liao XC, Rosbash M (1994) The yeast MUD2 protein: an interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. Genes Dev **8:** 843–854

Ali GS, Golovkin M, Reddy AS (2003) Nuclear localization and in vivo dynamics of a plant-specific serine/arginine-rich protein. Plant J **36:** 883–893

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, et al (2003) Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science **301:** 653–657

Bezerra IC, Michaels SD, Schomburg FM, Amasino RM (2004) Lesions in the mRNA cap-binding gene ABA HYPERSENSITIVE 1 suppress FRIGIDA-mediated delayed flowering in Arabidopsis. Plant J **40:** 112–119

Curtis MD, Grossniklaus U (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol **133:** 462–469

Docquier S, Tillemans V, Deltour R, Motte P (2004) Nuclear bodies and compartmentalization of pre-mRNA splicing factors in higher plants. Chromosoma **112:** 255–266

Domon C, Lorkovic ZJ, Valcarcel J, Filipowicz W (1998) Multiple forms of the U2 small nuclear ribonucleoprotein auxiliary factor U2AF subunits expressed in higher plants. J Biol Chem **273:** 34603–34610

Fang Y, Hearn S, Spector DL (2004) Tissue-specific expression and dynamic organization of SR splicing factors in Arabidopsis. Mol Biol Cell **15:** 2664–2673

Forment J, Naranjo MA, Roldan M, Serrano R, Vicente O (2002) Expression of Arabidopsis SR-like splicing proteins confers salt tolerance to yeast and transgenic plants. Plant J **30:** 511–519

Golovkin M, Reddy AS (1998) The plant U1 small nuclear ribonucleoprotein particle 70K protein interacts with two novel serine/arginine-rich proteins. Plant Cell **10:** 1637–1648

Golovkin M, Reddy AS (1999) An SC35-like protein and a novel serine/arginine-rich protein interact with Arabidopsis U1-70K protein. J Biol Chem **274:** 36428–36438

Gong Z, Dong CH, Lee H, Zhu J, Xiong L, Gong D, Stevenson B, Zhu JK (2005) A DEAD box RNA helicase is essential for mRNA export and important for development and stress responses in Arabidopsis. Plant Cell **17:** 256–267

Gupta S, Wang BB, Stryker GA, Zanetti ME, Lal SK (2005) Two novel arginine/serine (SR) proteins in maize are differentially spliced and utilize non-canonical splice sites. Biochim Biophys Acta **1728:** 105–114

Guth S, Martinez C, Gaur RK, Valcarcel J (1999) Evidence for substrate-specific requirement of the splicing factor U2AF(35) and for its function after polypyrimidine tract recognition by U2AF(65). Mol Cell Biol **19:** 8263–8271

Guth S, Tange TO, Kellenberger E, Valcarcel J (2001) Dual function for U2AF(35) in AG-dependent pre-mRNA splicing. Mol Cell Biol **21:** 7673–7681

Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. Genome Res **9:** 868–877

Kalyna M, Barta A (2004) A plethora of plant serine/arginine-rich proteins: redundancy or evolution of novel gene functions? Biochem Soc Trans **32:** 561–564

Kalyna M, Lopato S, Barta A (2003) Ectopic expression of atRSZ33 reveals its function in splicing and causes pleiotropic changes in development. Mol Biol Cell **14:** 3565–3577

Kielkopf CL, Lucke S, Green MR (2004) U2AF homology motifs: protein recognition in the RRM world. Genes Dev **18:** 1513–1526

Kielkopf CL, Rodionova NA, Green MR, Burley SK (2001) A novel peptide recognition mode revealed by the x-ray structure of a core U2AF35/U2AF65 heterodimer. Cell **106:** 595–605

Kitagawa K, Wang X, Hatada I, Yamaoka T, Nojima H, Inazawa J, Abe T, Mitsuya K, Oshimura M, Murata A, et al (1995) Isolation and mapping of human homologues of an imprinted mouse gene U2af1-rs1. Genomics **30:** 257–263

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17:** 1244–1245

Lazar G, Schaal T, Maniatis T, Goodman HM (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

Lopato S, Forstner C, Kalyna M, Hilscher J, Langhammer U, Indrapichate K, Lorkovic ZJ, Barta A (2002) Network of interactions of a novel plant-specific Arg/Ser-rich protein, atRSZ33, with atSC35-like splicing factors. J Biol Chem **277:** 39989–39998

Lopato S, Gattoni R, Fabini G, Stevenin J, Barta A (1999a) A novel family of plant splicing factors with a Zn knuckle motif: examination of RNA binding and splicing activities. Plant Mol Biol **39:** 761–773

Lopato S, Kalyna M, Dorner S, Kobayashi R, Krainer AR, Barta A (1999b) atSRp30, one of two SF2/ASF-like proteins from Arabidopsis thaliana, regulates splicing of specific plant genes. Genes Dev **13:** 987–1001

Lopato S, Mayeda A, Krainer AR, Barta A (1996a) Pre-mRNA splicing in

plants: characterization of Ser/Arg splicing factors. Proc Natl Acad Sci USA **93:** 3074–3079

Lopato S, Waigmann E, Barta A (1996b) Characterization of a novel arginine/serine-rich splicing factor in Arabidopsis. Plant Cell **8:** 2255–2264

Lorkovic ZJ, Barta A (2002) Genome analysis: RNA recognition motif (RRM) and K homology (KH) domain RNA-binding proteins from the flowering plant Arabidopsis thaliana. Nucleic Acids Res **30:** 623–635

Lorkovic ZJ, Barta A (2004) Compartmentalization of the splicing machinery in plant cell nuclei. Trends Plant Sci **9:** 565–568

Lorkovic ZJ, Hilscher J, Barta A (2004) Use of fluorescent protein tags to study nuclear organization of the spliceosomal machinery in transiently transformed living plant cells. Mol Biol Cell **15:** 3233–3243

Macknight R, Bancroft I, Page T, Lister C, Schmidt R, Love K, Westphal L, Murphy G, Sherson S, Cobbett C, et al (1997) FCA, a gene controlling flowering time in Arabidopsis, encodes a protein containing RNA-binding domains. Cell **89:** 737–745

Merendino L, Guth S, Bilbao D, Martinez C, Valcarcel J (1999) Inhibition of msl-2 splicing by sex-lethal reveals interaction between U2AF35 and the 3′ splice site AG. Nature **402:** 838–841

Pacheco TR, Gomes AQ, Barbosa-Morais NL, Benes V, Ansorge W, Wollerton M, Smith CW, Valcarcel J, Carmo-Fonseca M (2004) Diversity of vertebrate splicing factor U2AF35: identification of alternatively spliced U2AF1 mRNAS. J Biol Chem **279:** 27039–27049

Papp I, Mur LA, Dalmadi A, Dulai S, Koncz C (2004) A mutation in the cap binding protein 20 gene confers drought tolerance to Arabidopsis. Plant Mol Biol **55:** 679–686

Reed R (1989) The organization of 3′ splice-site sequences in mammalian introns. Genes Dev **3:** 2113–2123

Rudner DZ, Breger KS, Rio DC (1998) Molecular genetic analysis of the heterodimeric splicing factor U2AF: the RS domain on either the large or small Drosophila subunit is dispensable in vivo. Genes Dev **12:** 1010–1021

Rudner DZ, Kanaar R, Breger KS, Rio DC (1996) Mutations in the small subunit of the Drosophila U2AF splicing factor cause lethality and developmental defects. Proc Natl Acad Sci USA **93:** 10333–10337

Savaldi-Goldstein S, Aviv D, Davydov O, Fluhr R (2003) Alternative splicing modulation by a LAMMER kinase impinges on developmental and transcriptome expression. Plant Cell **15:** 926–938

Schomburg FM, Patton DA, Meinke DW, Amasino RM (2001) FPA, a gene involved in floral induction in Arabidopsis, encodes a protein containing RNA-recognition motifs. Plant Cell **13:** 1427–1436

Sheldon CC, Burn JE, Perez PP, Metzger J, Edwards JA, Peacock WJ, Dennis ES (1999) The FLF MADS box gene: a repressor of flowering in Arabidopsis regulated by vernalization and methylation. Plant Cell **11:** 445–458

Shepard J, Reick M, Olson S, Graveley BR (2002) Characterization of U2AF(6), a splicing factor related to U2AF(35). Mol Cell Biol **22:** 221–230

Tassone F, Villard L, Clancy K, Gardiner K (1999) Structures, sequence characteristics, and synteny relationships of the transcription factor E4TF1, the splicing factor U2AF35 and the cystathionine beta synthetase genes from Fugu rubripes. Gene **226:** 211–223

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22:** 4673–4680

Tian M, Maniatis T (1993) A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. Cell **74:** 105–114

Tillemans V, Dispa L, Remacle C, Collinge M, Motte P (2005) Functional distribution and dynamics of Arabidopsis SR splicing factors in living plant cells. Plant J **41:** 567–582

Tupler R, Perini G, Green MR (2001) Expressing the human genome. Nature **409:** 832–833

Wang B-B (2005) Genome-wide analysis of splicing related genes and alternative splicing in plants. PhD dissertation. Iowa State University, Ames, Iowa

Wang B-B, Brendel V (2004) The ASRG database: identification and survey of Arabidopsis thaliana genes involved in pre-mRNA splicing. Genome Biol **5:** R102

Wang H, Hill K, Perry SE (2004) An Arabidopsis RNA lariat debranching enzyme is essential for embryogenesis. J Biol Chem **279:** 1468–1473

Webb CJ, Wise JA (2004) The splicing factor U2AF small subunit is functionally conserved between fission yeast and humans. Mol Cell Biol **24:** 4229–4240

Weigel D, Glazebrook J (2002) Arabidopsis: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Wentz-Hunter K, Potashkin J (1996) The small subunit of the splicing factor U2AF is conserved in fission yeast. Nucleic Acids Res **24:** 1849–1854

Wu S, Romfo CM, Nilsen TW, Green MR (1999) Functional recognition of the 3′ splice site AG by the splicing factor U2AF35. Nature **402:** 832–835

Xiong L, Gong Z, Rock CD, Subramanian S, Guo Y, Xu W, Galbraith D, Zhu JK (2001) Modulation of abscisic acid signal transduction and biosynthesis by an Sm-like protein in Arabidopsis. Dev Cell **1:** 771–781

Zamore PD, Green MR (1989) Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. Proc Natl Acad Sci USA **86:** 9243–9247

Zamore PD, Patton JG, Green MR (1992) Cloning and domain structure of the mammalian splicing factor U2AF. Nature **355:** 609–614

Zhang M, Zamore PD, Carmo-Fonseca M, Lamond AI, Green MR (1992) Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. Proc Natl Acad Sci USA **89:** 8769–8773

Zorio DA, Blumenthal T (1999) U2AF35 is encoded by an essential gene clustered in an operon with RRM/cyclophilin in Caenorhabditis elegans. RNA **5:** 487–494

Zuo P, Maniatis T (1996) The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. Genes Dev **10:** 1356–1368