

# Characterization and comparison of intron structure and alternative splicing between *Medicago truncatula*, *Populus trichocarpa*, *Arabidopsis* and rice

Jong-Min Baek · Paul Han · Alberto Iandolino · Douglas R. Cook

Received: 3 October 2007 / Accepted: 1 April 2008 / Published online: 27 April 2008  
© Springer Science+Business Media B.V. 2008

**Abstract** Alignment of transcripts and genome sequences yielded a set of alternatively spliced transcripts in four angiosperm genomes: three dicotyledon species *Medicago truncatula* (*Medicago*), *Populus trichocarpa* (poplar) and *Arabidopsis thaliana* (*Arabidopsis*), and the monocotyledon *Oryza sativa* (rice). Intron retention was the predominant mode of alternative splicing (AS) in each species, consistent with previous reports for *Arabidopsis* and rice. We analyzed the structure of 5'-splice junctions and observed commonalities between species. There was dependency of base composition between sites flanking the 5'-splice junction, with the potential to create a subset of splice sites that interact more weakly or strongly than average with U1 snRNA. Such altered nucleotide composition was correlated with splicing fidelity in all four species. For *Medicago*, poplar and *Arabidopsis*, but not in rice, alternative splicing was most prevalent for introns with decreased UA content, consistent with lower UA content for monocot introns and potentially reflecting evolved differences in splicing mechanisms. Similarly, the occurrence of AS between transcript Gene Ontology categories was positively correlated between *Arabidopsis* and

*Medicago*, with no correlation between dicots and rice. Analysis of within-species paralogs and between-species reciprocal best-hit homologs yielded rare cases of potentially conserved AS events. Reverse transcriptase PCR and amplicon sequencing were used to confirm a subset of the in silico-predicted AS events within *Medicago*, as well as to characterize conserved AS events between *Medicago* and *Arabidopsis*.

**Keywords** Alternative splicing · Exon–intron junction · Gene ontology · *Medicago truncatula* · UA content

## Introduction

Alternative splicing (AS) enables a single gene to increase its coding capacity, producing several structurally distinct isoforms (Blencowe 2006; Reddy 2007). Although the functional importance of AS has been demonstrated during various stages of development, differentiation, and diseases in metazoan systems, less is known about the extent of AS function in higher plants (Reddy 2007). Recent genome and expressed sequence tag (EST) sequencing efforts in several plant species enable a comparative study of plant exon/intron structure and AS frequency across important plant families.

Bioinformatics analyses of AS have been reported in several higher eukaryotes. In humans, alignment of genomic and cDNA/EST sequence suggested that 40–70% of human genes experience AS (Kan et al. 2001; Modrek and Lee 2002; Consortium 2004; Nagasaki et al. 2005). Recent studies involving *Arabidopsis* and rice estimate that more than one-fifth of genes are subject to AS during their pre-mRNA processing (Wang and Brendel 2006; Campbell et al. 2006), significantly higher than previous estimates

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s11103-008-9334-4) contains supplementary material, which is available to authorized users.

---

J.-M. Baek (✉) · P. Han · D. R. Cook  
College of Agricultural and Environmental Sciences Genomics  
Facility, University of California, 117 Robbins hall, Davis,  
CA 95616, USA  
e-mail: jbaek@ucdavis.edu

A. Iandolino · D. R. Cook  
Department of Plant Pathology, University of California, Davis,  
CA 95616, USA

for *Arabidopsis thaliana*, where 5–10% of genes were suggested to have more than one transcript isoform (Iida et al. 2004; Nagasaki et al. 2005; Ner-Gaon et al. 2007). AS may take several forms, including exon skipping, intron retention, and alternative 5' (donor) or 3' (acceptor) splice site usage, with considerable variation in the frequency of AS patterns reported between certain species (Brett et al. 2002). Intron retention, for example, accounts for 30–50% of AS in *Arabidopsis* and rice (*Oryza sativa*), while exon skipping appears to be the predominant mechanism of transcript variation in human (Iida et al. 2004; Ner-Gaon et al. 2004; Nagasaki et al. 2005; Wang and Brendel 2006).

Relative to other characterized eukaryotic systems, plant introns display an unusual UA- or U-rich compositional bias, despite the fact that plant and animal mRNA processing systems appear to share a similar general mechanism. The U-rich content of plant introns confers a position-independent effect that distinguishes the function of plant introns from the polypyrimidine tracts of metazoan introns, and the U-rich nature of plant introns is required for correct mRNA processing (Goodall and Filipowicz 1989; Ko et al. 1998; Lorkovic et al. 2000). U-rich mRNA is known to interact with splicing factors and other nuclear proteins, such as plant U-rich RNA binding proteins (UBPs), which have been shown to enhance intron splicing (Lambermon et al. 2000; Hori and Watanabe 2005).

In addition to the possible role of intron content in AS, exon/intron junction sequences can act in *cis* to define splice sites. The first step in mRNA splicing is binding of U1 snRNP to the 5'-splice site via base-pairing with the 5' region of U1 snRNA. The nine nucleotides of the 5' end of U1 snRNA are highly conserved in all eukaryotes, and modification of base-pairing between the 5'-splice site and U1 snRNA has been shown to influence constitutive and alternative splicing of human introns (Ast 2004; Roca et al. 2005). The role of the splice donor site has been dissected further in human and mouse by examining the contribution of different nucleotide positions within the donor site to base-pairing with the U1 snRNA (Carmel et al. 2004). The results revealed dependence between the exonic and the intronic positions surrounding the splice donor site, such that stronger interactions (lower free energy) between U1 snRNA with the intronic portion of the splice donor site compensate for weak interactions in the exonic portion of the donor site, and vice versa. In plants, the function of the 5' splice site in pre-mRNA splicing and its effect on AS is less well characterized.

In vertebrates, alternatively spliced transcripts can contribute in important ways to tissue-specific and development-specific physiology (Stamm et al. 2005; Blencowe 2006). Less data is available in plants, although the functional relevance of certain AS-derived isoforms has been suggested (Reddy 2007). Certain plant disease-

resistance genes, for example, exhibit temporal variation in AS (Ayliffe et al. 1999; Jordan et al. 2002; Ferrier-Cana et al. 2005). In the case of the tobacco (*Nicotiana tabacum*) N gene and the *Arabidopsis* RPS4 gene both full-length and alternatively-spliced transcripts are required for efficient disease resistance (Dinesh-Kumar and Baker 2000; Zhang et al. 2004). In addition to disease-resistance genes, the *Arabidopsis* FCA gene, which encodes an RNA binding protein that promotes flowering, regulates its own expression by AS (Quesada et al. 2003). AS of a chloroplast ascorbate peroxidase pre-mRNA in spinach and tobacco is regulated in a tissue-specific manner, with possible consequences for oxidative stress resistance (Yoshimura et al. 2002), while in rice the Waxy (Wx) gene encodes a granule-bound starch synthase for which alternative splicing is temperature sensitive, potentially contributing to poor grain quality when seed maturation occurs at low temperature (Larkin and Park 1999).

In this study we used alignments between transcripts, represented by EST tentative consensus (TC) sequences, and genomic sequences to deduce gene structure and collect a set of alternatively-spliced transcripts in *Medicago*, poplar, *Arabidopsis* and rice. Prior studies involving transcript splicing in *Arabidopsis* and/or rice have focused largely on the frequency of alternative splicing and correlations with intron size and GC content. Here we extend these studies by analyzing the structure of 5' splice site junctions and correlating splice site structure with the occurrence of alternative splicing. We have also compared AS between species as a function of Gene Ontology (GO) categories. Selected *in silico* predictions of AS were verified experimentally, providing correlations between patterns of AS and plant development, as well as evidence of conserved AS between species and between closely related paralogs within species.

## Methods

### Data source

Genomic sequences of *Arabidopsis* and rice were obtained from public ftp sites of The *Arabidopsis* Information Resource (<ftp://ftp.arabidopsis.org/home/tair/Sequences>) and TIGR ([ftp://ftp.tigr.org/pub/data/Eukaryotic\\_Projects/o\\_sativa/annotation\\_dbs](ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs)), respectively. *Medicago* genomic data was in the form of assembled BAC sequences obtained from the *Medicago* project web site (<http://www.medicago.org>) on January 16, 2006. Assembled sequences of poplar genome were downloaded from public database of the DOE Joint Genome Institute ([http://genome.jgi-psf.org/Poptr1\\_1/Poptr1\\_1.home.html](http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html)). Transcript sequences were obtained from the TIGR Gene Indices specific to each

species, with the following database versions: AtGI v12.1, MtGI v8.0, PppGI v3.0, and OsGI v16.0.

#### Alignment between TC and genomic sequences

TC sequences from each species were aligned against the corresponding genome sequences using the Spidey algorithm (Wheelan et al. 2001) and exon/intron structure information was extracted from the Spidey outputs using PERL scripts of the FELINES tool (<http://www.genome.ou.edu/FELINES.html>) (Drabenstot et al. 2003). Additional PERL scripts and JAVA codes were developed in this study for analyzing data from the Spidey and FELINES outputs. The Linux executables of Spidey and BLASTALL were obtained from the National Center for Biotechnology Information (NCBI) and used to pair TCs with their corresponding genomic sequences. The alignment process was performed in a batch-wise automated manner with an e-value cutoff of  $\leq 1 \times 10^{-20}$  for BLASTN. After initial alignment, sequence pairs were further selected so that genome sequence covered at least 90% of a given TC, with a minimum nucleotide identity of 98% throughout the aligned region. Introns with a size range between 20 and 2,000 bp were considered in subsequent analyses.

#### Characterization of exon/intron structure

Alternative splicing was inferred when multiple TCs aligned to the genome region using the above criteria, but with differing exon/intron structures. Nucleotide sequences of exons, introns, and exon–intron junctions were parsed from the Spidey alignments. The UA content and size of exons and introns were calculated directly. Intron–exon junctions were parsed to contain 3 exonic and 12 intronic nucleotides for each splice donor site and 22 intronic with 3 exonic nucleotides for each splice acceptor site. Pictograms for the junction regions were constructed from <http://genes.mit.edu/pictogram.html> using the collected intron–exon boundaries.

#### Plant materials and RT-PCR

In silico predictions of alternative splicing were validated by means of reverse transcriptase PCR. Oligonucleotide primers were designed to span target introns using the PRIMER3 program ([http://frodo.wi.mit.edu/cgi-bin/primer3/primer3\\_www.cgi](http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)). *Medicago truncatula* cv Jemalong genotype A17 was grown as previously described (Cook et al. 1995). Total RNA was extracted from leaf and root tissues of *Arabidopsis* and *Medicago* using an RNeasy™ Plant Mini Kit (Qiagen) and quantified using the Quant-iT™ RiboGreen® RNA Assay Kit (Invitrogen). Semi-quantitative RT-PCR was

performed using the StrataScript™ RT-PCR kit (Stratagene). Amplified products were excised from agarose gels and sequenced directly to verify amplicon content.

## Results

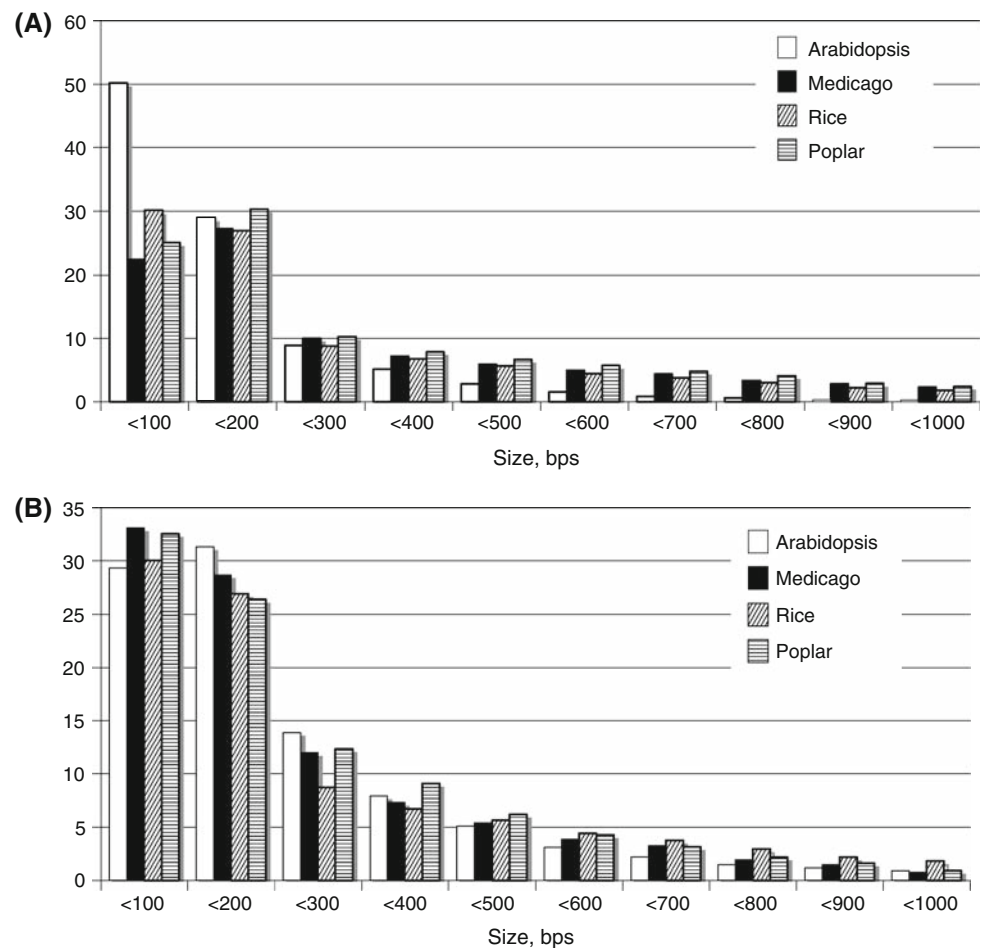
#### Alignment and intron prediction

To provide a common basis for comparison between *Medicago*, poplar, *Arabidopsis* and rice, gene structures were predicted by comparing transcript sequences (tentative consensus sequences or TCs) obtained from The Institute for Genomics Research (TIGR) Gene Indices EST database (AtGI v12.1; MtGI v8.0; OsGI v16.0; PpGI v3.0), with genomic DNA sequences obtained from the public databases of each species. Features of deduced gene structures are given in Table 1. On average, *Arabidopsis* and rice TCs are both more numerous and longer than those predicted in *Medicago* and poplar reflecting efforts to characterize full-length cDNAs in *Arabidopsis* and rice (Iida et al. 2004; Campbell et al. 2006; Yamada et al. 2003; Kikuchi et al. 2003; Liu et al. 2007), but not in *Medicago* and poplar. Most likely as a consequence of this situation, the *Medicago* and poplar gene models derived from alignments between TC and genomic sequence also contained fewer introns per gene. Despite these differences, the percentage of alignments corresponding to genes with introns was generally similar between the four species, with a low of 71% and 72% intron-containing genes in *Medicago* and poplar, respectively, and a high of 79% intron-containing genes in *Arabidopsis*. As shown in Fig. 1b, exon size distribution was also well conserved between species. By contrast, intron size varied significantly between species, increasing in rank order of genome size; the largest differences between species were the frequencies of introns smaller than 100 bp and introns larger

**Table 1** Deduced characters of the transcript data set used

	<i>Arabidopsis</i>	<i>Medicago</i>	Poplar	Rice
Genome size (Mb)	125	520	480	430
No. TC used	28,900	18,612	44,755	36,381
No. TC aligned	23,319	6,584	11,796	23,062
No. of genes aligned	19,842	5,926	9,718	19,972
No. of genes with 2 or more TC	1,680	321	1,710	1,523
% of alternative splicing	8.5	5.4	17.6	7.6
Avg. size of TC (bp)	1,476.2	993.3	840.3	1,493.1
Avg. intron size (bp)	256	423	405	375
Intron density per gene	4.5	3.1	2.5	3.9
Avg exon size (bp)	251	247	296	293

**Fig. 1** Size distribution of introns (a) and exons (b). Sizes of introns and exons were collected and their occurrence was counted in every 100 base interval. For normalized comparison among four plant species, percentage of the counted intron/exon in each size range was calculated based on the total number of introns/exons. Each bar represents the percentage of number of introns/exons in the given range of size



than 400 bp (Fig. 1a). Greater than 99% of introns in all four species conformed to the GT-AG rule for splice donor and splice acceptor sites. Moreover, the context of the splice donor and acceptor sites was nearly identical in the four species (Supplemental Fig. S1) and similar to that of vertebrates (Ast 2004). Consistent with known plant gene structure, the intron-side of predicted splice donor and acceptor sites was rich in U (shown as T; Supplemental Fig. S1).

#### Identification of alternatively spliced transcripts

In *Arabidopsis* and rice, 1,680 and 1,523 genomic loci could be aligned perfectly with two or more TIGR transcript TCs, respectively, consistent with multiple transcript isoforms derived from the same genomic region (Table 1). Thus, 8.5% of the 19,842 *Arabidopsis* genes and 7.6% of the 19,972 rice genes with matching TIGR TCs revealed evidence of AS. In *Medicago* and poplar, 321 and 1,710 genomic regions could be aligned with multiple TIGR TCs, providing evidence for AS in 5.4% and 17.6% of genes in

these species, respectively. Differences in the extent and quality of sequence resources in these species are likely to influence the numbers of recorded AS isoforms. Moreover, the fact that we have used EST assemblies (TIGR's tentative consensus (TC) sequences) rather than individual EST reads precludes an assessment of AS frequency in individual genes. Our approach also misses genomic regions with identity to singleton ESTs, which, although it does not impact predictions of alternative splicing, it does restrict the intron set available for structural analysis. Previous studies that used EST data to predict AS in *Arabidopsis* yielded widely ranging estimates for the fraction of genes with transcript isoforms, with values of ~5% to ~22% of genes. This variance has been attributed to differences in computational approaches and depth of sequence resources (Wang and Brendel 2006; Campbell et al. 2006; Ner-Gaon et al. 2007). Nevertheless, our value of 8.5% *Arabidopsis* genes with alternate transcript isoforms is within the range of previous studies, and for our purposes the total number of putative AS events provides a substantial data set for subsequent analysis of transcript



structure. The most common form of AS in the TC data was intron retention in all four species, corresponding to 39%, 34%, 45%, and 51% of recorded AS events in *Medicago*, poplar, *Arabidopsis*, and rice, respectively, similar to previous reports from computational studies of plant AS (Iida et al. 2004; Ner-Gaon et al. 2004; Nagasaki et al. 2005; Wang and Brendel 2006).

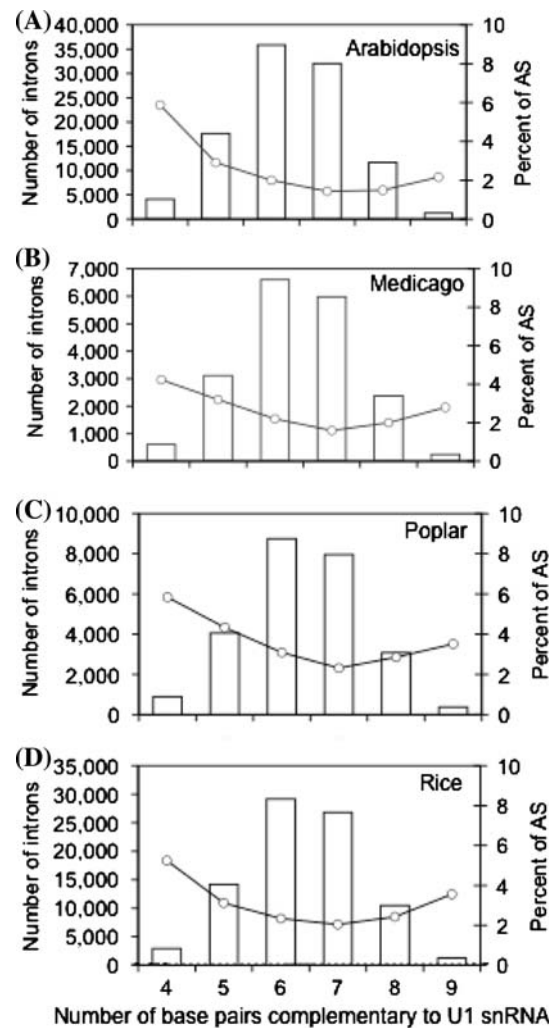
#### Base pairing potential between 5'-splice sites and U1 snRNA

Most eukaryotes possess several U1 snRNA paralogs, e.g., 4 in human and 14 in *Arabidopsis* (Wang and Brendel 2004). The nine nucleotides (5'-ACUUACCUG-3') at the 5'-end of U1 snRNA paralogs are fully conserved across plants and animals, as shown for example for eight selected plant species, representing 4 legumes, rice, wheat, *Arabidopsis* and tomato (Supplemental Fig. S2). As shown in Fig. 2, alignment of 5' exon–intron junctions, predicted above, with the U1 snRNA 5' 9 nucleotide sequence reveals that the average number of complementary base pairs between the U1 snRNA and the splice donor site, counting the non-canonical RNA G:U pair, is 6.31 (*Arabidopsis*), 6.37 (*Medicago*), 6.36 (poplar) and 6.34 (rice). In human, the average number of complementary bases was 7.05 (Carmel et al. 2004).

The base pairing potential of 5'-splice junctions is similarly distributed within each species, from a low of 4 to a high of all 9 bases complementary to the cognate U1 snRNA (Fig. 2). Moreover, the occurrence of AS was well correlated with the base-pairing potential of the splice donor site, with the greatest relative AS incidence associated with splice sites of low base pairing potential (Fig. 2). Thus, the context of the splice donor site is well correlated with the fidelity of mRNA processing. Interestingly, extended complementarity involving all 9 bases was correlated with a slight increase in the incidence of AS in all four species. Extended base pairing has been shown to be inhibitory to splicing at a late step in spliceosome assembly, leading to the suggestion that stable U1 snRNA binding is advantageous for assembly of U1 snRNP, but delays its release and thus impedes entry of subsequent splicing components such as U4/U6 (Lund and Kjems 2002).

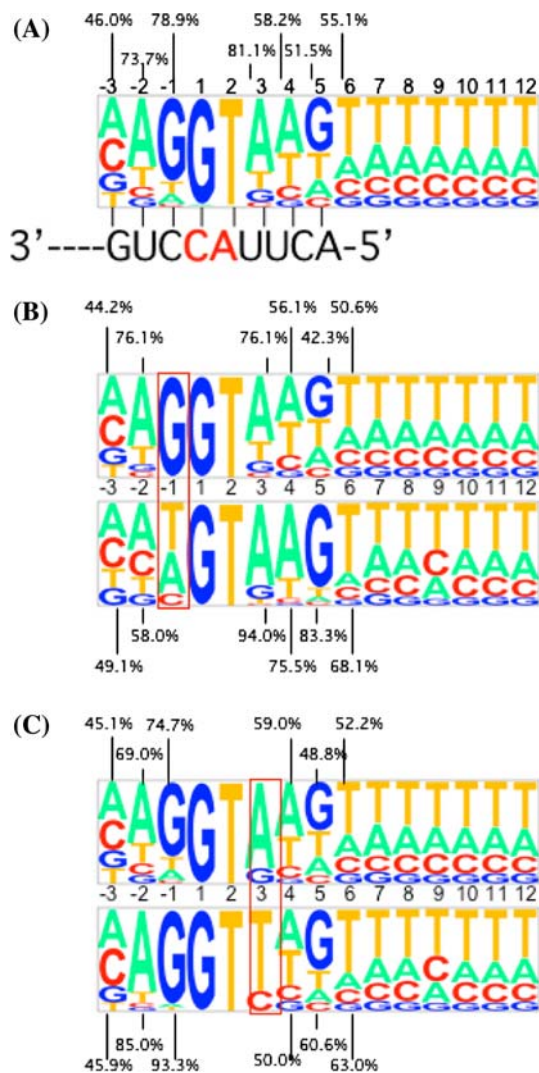
#### Dependency between exonic and intronic sequences at the splice donor site

Based on a comparison of human and mouse exon–intron junctions, Carmel et al. (2004) observed that weak exonic base complementarity with U1 snRNA was typically compensated by strong base pairing potential in intronic sequences, and vice versa. Exon/intron junction sequences



**Fig. 2** Effect of base-pairing at donor sites on AS in *Arabidopsis* (a), *Medicago* (b), poplar (c) and rice (d). Each bar represents the number of introns with the given number of base pair. Percentage of AS introns in each group was plotted with line

from *Medicago*, poplar, *Arabidopsis* and rice were examined to test for dependency between exonic and intronic sequences. Similar results were obtained for all four species and thus only the data from *Medicago* are presented in Fig. 3. A twelve-nucleotide interval, beginning 3 nucleotides on the exon-side of the splice donor site and extending into the intron, was parsed from 18,371 *Medicago* introns and the nucleotide frequency at each position was recorded (Fig. 3a). The splice donor sites were subdivided into two groups based on the presence or absence of base complementarity (i.e., G) at position -1 (Fig. 3b), and based on the presence or absence of base complementarity (A or U) at position +3 (Fig. 3c). For splice donor sites with T, A, or C at the exonic -1 position (no complementary base), base pairing potential at intronic



**Fig. 3** Dependency between *Medicago* exonic and intronic sequences of intron donor sites on base-pairing with U1 snRNA. The 5'-end sequence of U1 snRNA is shown with *Medicago* introns (a). Pictograms were obtained from <http://genes.mit.edu/pictogram.html> using the 5'-splice site data of introns and the frequency of base-pairing nucleotide was shown in percentage for each location of exon/intron junction. The *Medicago* introns were divided into two groups based on the base-pairing on the exonic -1 position (b) and on the intronic +3 position (c). The frequency of base-pairing in each location was indicated

+3, +4, +5, and +6 positions was increased by 17.5 to 41% relative to introns with a complementary “G” at position -1 (Fig. 3b). This inferred compensation was twice as frequent at position +5, which, notably, represents the only intron-side G–C base-pair. The lack of complementarity at position -1 was also correlated with a decrease in base pairing potential frequency at exonic position -2, indicating that a subset of 5' splice junctions may have a substantially weaker interaction with U1 snRNA on the exon side. Similar relationships were obtained when the nucleotide content at the intronic +3 position was analyzed.

In particular, introns without base-pairing potential at position +3 (T or C) possess increased base complementarity at positions -1 and -2, and to a lesser extent at +5 and +6 (Fig. 3c).

UA content is correlated with AS in *Medicago*, poplar, and *Arabidopsis*, but not rice

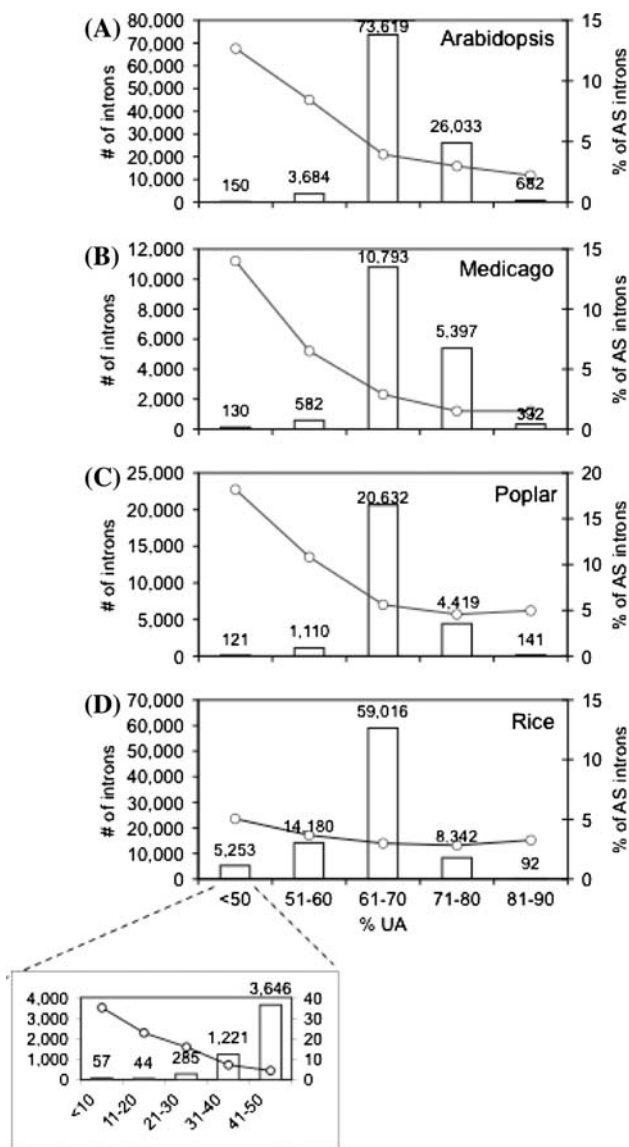
High U(T)A content is a distinguishing feature of plant introns, with functional implications for intron splicing. UA content was examined in both exons and introns of the four plant species, primarily to confirm that our data set was reflective of previously established patterns. The compiled genes of dicot species, *Medicago*, poplar, and *Arabidopsis*, have comparable UA content, with 66.8/69.7/66.7% and 56.3/58.7/57.1% in exons and introns (*Arabidopsis*/*Medicago*/poplar), respectively. The rice gene set exhibited the same relative difference, i.e., 63.1% and 53.5% UA (exons and introns respectively), but with lower average UA content (Supplemental Fig. S3). To determine whether AS introns are unusual with respect to UA content, intron sequences involved in AS were collected and analyzed. In general, AS introns in *Arabidopsis*, *Medicago*, and poplar were less UA-rich than non-AS introns (Table 2). This difference was even more pronounced when AS incidence was plotted over a range of UA content, with a 5-fold higher incidence of AS when UA content was <50%, compared to 60–70% (Fig. 4). Interestingly, rice UA content was not as strongly correlated with AS incidence, despite the fact that numerous rice introns were analyzed throughout the <50–90% UA range. We note that Wang and Brendel (2006) observed a significant association of AS for rice introns of UA <35%. Such introns are relatively rare in rice (Supplemental Fig. S3), comprising less than 1% of the total introns. When we consider only this GC-rich intron class AS incidence is significantly increased, although on a numerical basis accounts for only 3.7% of total recorded AS isoforms (see inset of Fig. 4).

Experimental validation of AS in *Medicago truncatula*

To validate predictions from computational analysis, twenty candidate genes were selected from the *Medicago*

**Table 2** UA content in exons and introns of *Arabidopsis*, *Medicago*, poplar, and rice

	Exons	Introns	AS Introns
<i>Arabidopsis</i>	56.8%	66.8%	64.5%
<i>Medicago</i>	58.7%	69.7%	65.9%
Poplar	57.1%	66.7%	63.2%
Rice	53.5%	63.1%	62.7%



**Fig. 4** UA content of introns and its effect on AS in *Arabidopsis* (a), *Medicago* (b), poplar (c) and rice (d). Each bar represents number of introns in the given range of %UA and the percentage of introns with AS was shown as a line

dataset and semi-quantitative RT-PCR was performed to evaluate the prevalence of transcript isoforms (Table 3). Gene specific PCR primers (Supplemental Table S1) were designed to amplify across variably spliced introns. RNA samples were prepared from leaves and roots, as well as roots of increasing age. The number and size of RT-PCR products were used for comparison to in silico data, with direct sequencing of gel purified amplicons used for final confirmation of the structure of AS transcripts. Four types of AS were observed in the test gene set, including intron retention (RI), exon skipping (ES), alternative donor site (AD) and alternative acceptor site (AA). For 17 of the 20 test genes, the number of RT-PCR products and their

observed sizes were consistent with in silico predictions from analysis of TIGR *Medicago* GeneIndex. Only single amplicons were observed for transcripts derived from a stress-related kinase gene and an ubiquitin protease gene, which may reflect the absence of AS in specific tissues or perhaps genomic DNA contamination in cDNA libraries used for EST sequencing. In the case of a MYB transcription factor gene, three splice variant were detected by means of RT-PCR, whereas two isoforms were predicted from in silico analysis.

#### Structure and differential abundance of AS products

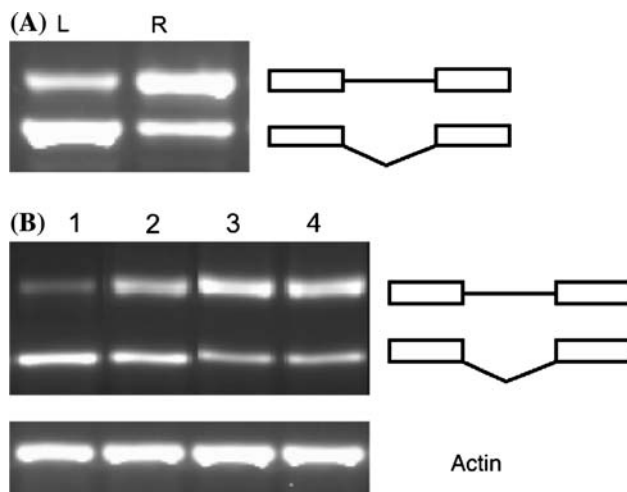
The impact of AS on transcript structure and open reading frame content was examined for the 20 test genes listed in Table 3. The most frequent consequence of AS among the 20 test genes was introduction of an early stop codon, observed for 11 genes. In addition to yielding truncated protein isoforms, termination of translation >50 nt 5' of the 3'-most splice junction can trigger transcript degradation by a mechanism known as non-sense mediated mRNA decay (NMD), providing a means to remove aberrant transcripts and/or regulate transcript abundance. Certain AS events were predicted to yield modified polypeptide sequences as a consequence of N- or C-terminal extensions, the insertion/deletion of amino acids within proteins, or variable sequence content resulting from frame shift events. Three isoforms from a *Medicago* MYB transcription factor gene (MtMYB1) differed in the location of their start codon, resulting in altered amino acid sequence and size of the N-terminal peptide, where a putative DNA binding motif is located (Supplemental Fig. S4). AS in genes coding for an auxin-repressed protein and a PR protein resulted in variable C-termini (Supplemental Fig. S5a, b), and an AS isoform of the peroxidase2 gene is predicted to yield insertion of 6 amino acids within the deduced polypeptide sequence (Supplemental Fig. S5c). We also observed three examples of alternatively spliced 5' or 3' UTR sequences, where altered structure can potentially impact transcript stability, subcellular localization of the mRNA, or translation efficiency. In particular, alteration of the 3' UTR was observed for the branch point protein and malate dehydrogenase transcripts, while the bZIP transcription factor transcripts were derived from differential retention of a region internal to the 5'-UTR.

The RT-PCR experiments conducted to validate transcript structure (Table 3) revealed examples of tissue- or development-specific patterns of AS. Figure 5 shows two examples of genes where AS involves variable retention of the first intron, with relative transcript abundance assessed by densitometry of ethidium bromide stained agarose gels. The peroxidase gene shown in Fig. 5a contains four exons

**Table 3** *Medicago* genes used for RT-PCR experiments

Gene name	TC pair	AS type	No. of amplicons	Outcome of AS
Branch point protein	TC101317/TC101318	AA	2	AS at 3'-UTR
bZIP Transcription factor	TC101061/TC101062	ES/AA/AD	2	AS at 5'-UTR
Auxin response factor	TC96747/TC96748	AA/D	2	Early STOP
Auxin-repressed protein, Drm4	TC107811/TC107812	AD	2	Different C-terminal
Embryonic protein	TC107158/TC107159	RI	2	Early STOP
Extensin	TC100212/TC100227	RI	2	Early STOP
Glycine-rich RNA-binding protein	TC93939/TC93940/TC94153	RI, AD	3	Early STOP
Malate dehydrogenase	TC94390/TC94392	RI	2	AS at 3'-UTR
Splicing factor	TC99613/TC108924	ES	2	Early STOP
MYB transcription factor	TC101529/TC101530	RI, AD	3	Different N-terminal
Nodulin	TC101084/TC101085	RI	2	Early STOP
Oxysterol-binding	TC107755/TC107756	RI	2	Early STOP
Peroxidase1	TC95153/TC95154	RI	2	Early STOP
Peroxidase2	TC94677/TC94679	AA	2	6 AA gap in peptide
phospholipase C	TC106978/TC106980	RI	2	Early STOP
PR protein	TC94562/TC98367	RI	2	different C-terminal
R protein	TC94428/TC94430	RI	2	Early STOP
Leguminosins	TC94716/TC94717	ES	2	Early STOP
Stress-related kinase	TC105431/TC107574	–	1	–
Ubiquitin protease	TC100891/TC100892	–	1	–

AS types were detected in the test genes by means of direct sequencing of gel-purified PCR products, yielding examples of intron retention (RI), exon skipping (ES), alternative donor site (AD) and alternative acceptor site (AA). The number of RT-PCR products was indicated in each gene with structural outcome of the AS



**Fig. 5** Tissue- and development-specific AS in *Medicago*. **(a)** Total RNA was extracted from leaf (L) and root (R) tissues of 2-week old plants and RT-PCR was performed with primer set of a peroxidase gene. **(b)** Development-specific AS of an R protein gene in *Medicago* root. Root tissues of 1, 2, 3, and 4-week-old plants were used for RNA extraction and RT-PCR

and is represented by two AS transcript isoforms (TC95153 and TC95154). Correct splicing of the first intron accounted for approximately 90% of transcript in leaves, while

transcripts that retained the first intron were the predominant isoform in roots. Figure 5b documents changes in the relative abundance of transcript isoforms for a nucleotide binding site-leucine rich repeat (NBS-LRR) resistance gene homolog. This putative disease resistance gene is composed of three exons and two introns, with two isoforms (TC94428 and TC94430). The fully-spliced transcript was dominant in young (1 week old) roots, while the isoform derived from retention of intron 1 increased over a 4 week time course, during which significant changes to root architecture occur.

#### Conservation of AS between *Medicago* and *Arabidopsis*

With the intent of comparing AS patterns between *Medicago*, *Arabidopsis* and rice, we analyzed clusters of highly similar genes present in the TIGR Eukaryotic Gene Ortholog (EGO) database (Lee et al. 2002). The EGO data set is derived from a reciprocal BLASTN best hit approach; importantly, while the paired transcripts are highly similar at the nucleotide level, they are not necessarily orthologous genes. 2,974 *Medicago* and *Arabidopsis* homolog pairs were identified in the EGO clusters, including 195 and 321 cases of AS in *Medicago* and *Arabidopsis*, respectively.



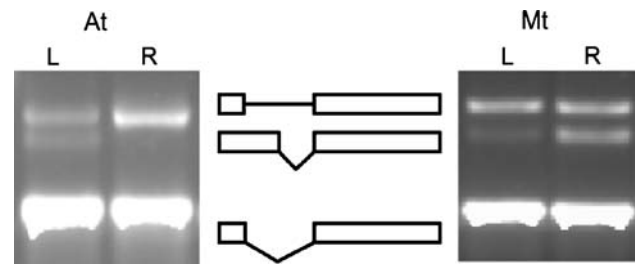
**Table 4** Cross-species conservation of AS

Species	Best hit pair	AS in each species	AS in both species
Mt/At	2974	195/321	31
Mt/Os	1644	113/136	13
Os/At	5106	392/682	45
Mt/Os/At	1360	97/125/176	6

Reciprocal best homologs between *Medicago*, *Arabidopsis* and Rice were obtained from the TIGR Eukaryotic Gene Orthology (EGO) database. AS incidence in each species, as well as the incidence of AS in conserved genes between species, is shown

Thirty of the homologous pairs, representing 1% of the total 2,974 transcript pairs and 10–15% of AS transcript pairs, were subject to AS in both species. Similar frequencies of AS in homologous genes were observed between *Medicago* and rice (0.8%), and between rice and *Arabidopsis* (0.9%) (Table 4). Among the 111 pairwise AS events only 6 are structurally similar in AS (Supplemental Tables S2–S4). Taken together, these results suggest that few of these AS events are likely to represent conserved AS. When *Medicago*, *Arabidopsis* and Rice were considered together, 1,360 EGO clusters were identified that contained transcripts from each species (Table 4), including 6 homologs with AS events in all three species, two of which were structurally similar between species (Supplemental Table S5). We speculate that such rare conserved AS events are conserved and may be functionally important.

To validate those rare AS events that are similar between species, we analyzed splicing of a Glycine-rich RNA-binding protein (GRP) gene. The regulation of *Arabidopsis* AtGRP7 is known to involve AS, and the corresponding protein is implicated in circadian rhythm and response to cold stress (Staiger et al. 2003). As shown in Supplemental Table S2, EGO cluster ID 894090 includes both AtGRP7 and a *Medicago* homolog. The *Medicago* GRP gene (MtGRP1) is represented by 3 transcript isoforms in the TIGR database (TC93939, TC93940, TC94153). TC93939 results from complete splicing of the 315bp intron; TC93940 results from use of an alternative donor (AD) site that incorporates 128bp of the 5' intron region into the transcript; and TC94153 retains the full-length intron. *Arabidopsis* GRP7 (At2g21660) shares 63% identity at the nucleotide level with MtGRP1, and 75% identity at the protein level (Supplemental Fig. S7a). MtGRP1 and AtGRP7 also have well-conserved exon/intron structure and possess similar patterns of AS as revealed by RT-PCR (Fig. 6). A rice GRP homolog (LOC\_Os12g43600) was also found with two AS isoforms (TC261824 and TC248372), where the latter sequence retains the full-length intron. No AD variant of rice GRP was detected in the current rice TC database.



**Fig. 6** Conserved AS pattern of a GRP gene in *Arabidopsis* and *Medicago*. Total RNA was extracted from leaf (L) and root (R) tissues of *Arabidopsis* and *Medicago* and RT-PCR was performed with gene-specific primer sets. Three different amplicons were detected in each GRP gene from agarose gel electrophoresis. Individual band was extracted from agarose gel and verified with DNA sequencing

## Discussion

Alternative transcript splicing is increasingly recognized as an important factor in the complexity of eukaryotic proteomes. In this study, we have analyzed and compared transcript structures and alternative splicing patterns in the model plants *Arabidopsis*, *Medicago*, poplar, and rice. We observed splicing-related characters that are common to all four angiosperms, and aspects of splicing that distinguish between species, with divergence most commonly observed between dicot and monocot species.

Common to *Arabidopsis*, *Medicago*, poplar, and rice was a dependency between exonic and intronic sequences flanking the intron donor site in pre-mRNA, similar to relationships previously established for mammalian transcripts (Carmel et al. 2004). Such altered splice site context may mis-stabilize pre-mRNA-snRNA interactions and thereby alter AS frequency; consistent with this suggestions, AS incidence was increased when splice site flanking sequences were under- or over-represented for base-pairing potential with the U1 snRNA. An additional contextual feature that influences intron recognition and splicing of plant introns is the occurrence of extended UA-polymer tracts. We observed that intron UA content was inversely proportional to AS frequency. This correlation was restricted to the dicots *Arabidopsis*, *Medicago*, and poplar, but absent in rice, supporting the observation of Goodall et al. (1991) that monocot introns lack an absolute requirement for UA-rich sequences (although their presence facilitates intron splicing).

The vast majority of plant introns contain the GT dinucleotide at their splice donor sites. According to Campbell et al. (2006), substitution of GT by GC at splice donor sites is correlated with increased incidence of AS, especially in the case of alternative donor (AD) sites. We observe the same correlation in our data sets. In particular, Chi-square analysis of rice donor sites at the AD class showed a significantly increased frequency of the GC

dinucleotide ( $P = 1.06 \times 10^{-17}$ ) compared to non-AS donor sites. GC enrichment was also observed in *Arabidopsis* AD donor sites ( $P = 0.001$ ). *Medicago* and poplar data were not shown here because the number of observed GC substitution was not enough for statistical analysis.

AS is proposed to provide adaptive benefit to eukaryotic organisms by increasing the plasticity of the proteome in response to developmental and external cues (Kazan 2003; Palusa et al. 2007). In cases where AS confers a selective advantage one might expect to observe conserved AS in orthologous genes, especially for related species (e.g., plants) where the system context for protein function is likely to be similar. Using the eukaryotic gene ortholog (EGO) data set (Lee et al. 2002) as a basis for comparing between species, we observed two likely cases of conserved splicing patterns between *Medicago* and *Arabidopsis*; although the inferred rice ortholog did possess an AS isoform, it was structurally-distinct from the dicot AS isoforms and thus splicing of the rice ortholog is unlikely to be a homologous character.

An alternative source of functional diversity for the transcriptome and proteome is gene duplication, with new paralogs potentially facilitating the evolution of both novel patterns of transcription and neo- or sub-functionalization of protein isoforms. (Kopelman et al. 2005) observed that AS incidence was inversely correlated with gene duplication, consistent with the possibility that functional diversity generated upon gene duplication may supercede the role of AS. A corollary to this prediction is that during later stages of paralog evolution, newly acquired AS would yield paralog-specific AS isoforms (Su et al. 2006). During our PCR and sequence analysis of GRP gene transcripts in *Medicago* a total of three MtGRP paralogs were identified, with ~80% of nucleotide sequence identity in coding regions and highly conserved AS isoforms (Supplemental Fig. S7). These three MtGRP genes (MtGRP1, 2, and 3) are apparently the product of recent tandem duplication, as they reside with a 50 kb interval on the same *Medicago* BAC clone (Supplemental Fig. S6; GenBank accession# AC134242). In *Arabidopsis*, AtGRP7 is a close paralog (i.e., ~80% nucleotide identity) of AtGRP8, with the two paralogs located on different chromosomes. Phylogenetic analysis suggests that the duplications in *Arabidopsis* and *Medicago* postdate their speciation (Supplemental Fig. S7). As was observed in *Medicago*, the structure of AS isoforms is conserved between these *Arabidopsis* paralogs. No comparable duplication with conserved exon/intron structure was found for the rice GRP ortholog (LOC\_Os12g43600). Certain members of the MYB transcription family in *Arabidopsis* and rice have also been reported to possess conserved patterns of AS (Li et al. 2006). As shown in Supplemental Fig. S8 we identified two closely related MYB paralogs in *Medicago*, with MtMYB1

exhibiting AS isoforms similar to those reported for *Arabidopsis* and rice (Supplemental Fig. S4). The two MtMYBs are 57% identical in their inferred polypeptide sequences and are approximately 50% conserved with the respective MYB proteins of *Arabidopsis* and rice (Supplemental Fig. S8). Interestingly, sequence similarity is almost exclusively located in the N-terminal half of the protein, where a DNA-binding domain resides. The fact that similar AS isoforms in *Medicago*, *Arabidopsis* and rice are predicted to yield MYB proteins with common modifications to their N-terminal domain, suggests a conserved role in MYB gene regulation. We conclude that, although rare, certain AS events are well conserved following both speciation and within genome duplications, with the GRP and MYB proteins providing examples of both conservation histories.

Despite the example of AtGRP7/8 and MtGRP1/2/3, the current analysis suggests that most forms of AS are not conserved between species, consistent with the possibility that many AS isoforms lack functional significance and therefore functional constraint, or conversely that AS diversity evolves quickly and underlies some of the functional differences that are so evident between angiosperms. Although the AS mechanisms in plants are largely unknown, recent studies suggest that plant AS is an important posttranscriptional regulatory system in modulating gene expression (reviewed in Reddy 2007). When AS transcripts were classified into Gene Ontology (GO) categories, we observed uneven distribution of AS frequency between GO categories. This observation is consistent with the possibility that AS incidence is under various levels of selectivity depending on the function of the encoded protein (Supplemental Fig. S9). Interestingly, the relationship of AS frequency with GO category is similar between dicot species and distinct from rice, indicating that considerable difference in AS incidence may have evolved subsequent to the dicot–monocot divide (Supplemental Fig. S10). We emphasize, however, that these observations are based on relatively small data sets, and will require further and more detailed analyses before firm conclusions are warranted.

A trivial source of variation in observed AS incidence could be the dissimilar origins of cDNA libraries that comprise each species' EST database. For example, over 60% of cDNA libraries in the *Medicago* EST database originate from root tissues while *Arabidopsis* and rice EST collections contain only 13% and 8% of root ESTs, respectively (<http://www.tigr.org/tdb/tgi/plant.shtml>). This biased distribution of EST sources could explain the underrepresentation of AS isoforms for genes of the 'development' GO category in *Medicago* (Supplemental Fig. S9). Similarly, more than 75% of *Medicago* cDNAs derive from tissues challenged by abiotic or biotic stress, while only 3%

of *Arabidopsis* and 15% of rice cDNAs derive from stressed tissues. Consistent with this fact, AS bias in the ‘response to stimulus’ category was highest for *Medicago* and lowest for *Arabidopsis*.

Computational alignment of ESTs with genomic data provided a valuable resource to identify genes with AS in four model plant genomes, *Medicago*, poplar, *Arabidopsis* and rice, and to correlate the sequence context of splice donor sites with AS incidence. The inclusion of three dicot and one monocot species allowed inference of homologous characters that are either common to angiosperms or shared by dicot species but diverged from the monocot lineage. However, EST data lack both sampling depth and developmental breadth, and thus may be insensitive to rare transcript isoforms or isoforms that occur only under specific conditions. New sequencing technologies can likely overcome these limitations and thus offer opportunities to extend these studies, for example by examining genome-scale AS in response to developmental and/or environmental cues. Similarly, the presence of reference genome sequences for several dicot species (e.g., rice, *Arabidopsis*, *Medicago* and poplar) offers an important opportunity to explore both conserved and species-specific AS events within a documented phylogenetic context. Based on the assumption that structural conservation belies functional importance, it will be particularly interesting to catalog and eventually test by reverse genetic and biochemical assay the role of AS isoforms and their encoded proteins.

## References

- Ast G (2004) How did alternative splicing evolve? *Nat Rev Genet* 5:773–782
- Ayliffe MA, Frost DV, Finnegan EJ, Lawrence GJ, Anderson PA, Ellis JG (1999) Analysis of alternative transcripts of the flax L6 rust resistance gene. *Plant J* 17:287–292
- Blencowe BJ (2006) Alternative splicing: new insights from global analyses. *Cell* 126:37–47
- Brett D, Pospisil H, Valcarcel J, Reich J, Bork P (2002) Alternative splicing and genome complexity. *Nat Genet* 30:29–30
- Campbell MA, Haas BJ, Hamilton JP, Mount SM, Buell CR (2006) Comprehensive analysis of alternative splicing in rice and comparative analyses with *Arabidopsis*. *BMC Genomics* 7:327
- Carmel I, Tal S, Vig I, Ast G (2004) Comparative analysis detects dependencies among the 5′ splice-site positions. *RNA* 10:828–840
- Consortium IHGS (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431:931–945
- Cook D, Dreyer D, Bonnet D, Howell M, Nony E, VandenBosch K (1995) Transient induction of a peroxidase gene in *Medicago truncatula* precedes infection by *Rhizobium meliloti*. *Plant Cell* 7:43–55
- Dinesh-Kumar SP, Baker BJ (2000) Alternatively spliced N resistance gene transcripts: their possible role in tobacco mosaic virus resistance. *Proc Natl Acad Sci USA* 97:1908–1913
- Drabenstot SD, Kupfer DM, White JD, Dyer DW, Roe BA, Buchanan KL, Murphy JW (2003) FELINES: a utility for extracting and examining EST-defined introns and exons. *Nucleic Acids Res* 31:e141
- Ferrier-Cana E, Macadre C, Seignac M, David P, Langin T, Geffroy V (2005) Distinct post-transcriptional modifications result into seven alternative transcripts of the CC-NBS-LRR gene JA1tr of *Phaseolus vulgaris*. *Theor Appl Genet* 110:895–905
- Goodall GJ, Filipowicz W (1989) The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell* 58:473–483
- Goodall GJ, Filipowicz W, Goodall GJ, Filipowicz W, Ko CH, Brendel V, Taylor RD, Walbot V (1991) Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants. *EMBO J* 10:2635–2644
- Hori K, Watanabe Y (2005) UPF3 suppresses aberrant spliced mRNA in *Arabidopsis*. *Plant J* 43:530–540
- Iida K, Seki M, Sakurai T, Satou M, Akiyama K, Toyoda T, Konagaya A, Shinozaki K (2004) Genome-wide analysis of alternative pre-mRNA splicing in *Arabidopsis thaliana* based on full-length cDNA sequences. *Nucleic Acids Res* 32:5096–5103
- Jordan T, Schornack S, Lahaye T (2002) Alternative splicing of transcripts encoding Toll-like plant resistance proteins—what’s the functional relevance to innate immunity? *Trends Plant Sci* 7:392–398
- Kan Z, Rouchka EC, Gish WR, States DJ (2001) Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. *Genome Res* 11:889–900
- Kazan K (2003) Alternative splicing and proteome diversity in plants: the tip of the iceberg has just emerged. *Trends Plant Sci* 8:468–471
- Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H, Hotta I, Kojima K, Namiki T, Ohneda E, Yahagi W, Suzuki K, Li CJ, Ohtsuki K, Shishiki T, Otomo Y, Murakami K, Iida Y, Sugano S, Fujimura T, Suzuki Y, Tsunoda Y, Kurosaki T, Kodama T, Masuda H, Kobayashi M, Xie Q, Lu M, Narikawa R, Sugiyama A, Mizuno K, Yokomizo S, Niikura J, Ikeda R, Ishibiki J, Kawamata M, Yoshimura A, Miura J, Kusumegi T, Oka M, Ryu R, Ueda M, Matsubara K, Kawai J, Carminci P, Adachi J, Aizawa K, Arakawa T, Fukuda S, Hara A, Hashizume W, Hayatsu N, Imotani K, Ishii Y, Itoh M, Kagawa I, Kondo S, Konno H, Miyazaki A, Osato N, Ota Y, Saito R, Sasaki D, Sato K, Shibata K, Shinagawa A, Shiraki T, Yoshino M, Hayashizaki Y, Yasunishi A (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. *Science* 301:376–379
- Ko CH, Brendel V, Taylor RD, Walbot V (1998) U-richness is a defining feature of plant introns and may function as an intron recognition signal in maize. *Plant Mol Biol* 36:573–583
- Kopelman NM, Lancet D, Yanai I (2005) Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. *Nat Genet* 37:588–589
- Lambermon MH, Simpson GG, Wicczorek Kirk DA, Hemmings-Mieszczak M, Klahre U, Filipowicz W (2000) UBP1, a novel hnRNP-like protein that functions at multiple steps of higher plant nuclear pre-mRNA maturation. *EMBO J* 19:1638–1649
- Larkin PD, Park WD (1999) Transcript accumulation and utilization of alternate and non-consensus splice sites in rice granule-bound starch synthase are temperature-sensitive and controlled by a single-nucleotide polymorphism. *Plant Mol Biol* 40:719–727
- Lee Y, Sultana R, Pertea G, Cho J, Karamycheva S, Tsai J, Parvizi B, Cheung F, Antonescu V, White J, Holt I, Liang F, Quackenbush J (2002) Cross-referencing eukaryotic genomes: TIGR Orthologous Gene Alignments (TOGA). *Genome Res* 12:493–502
- Li J, Li X, Guo L, Lu F, Feng X, He K, Wei L, Chen Z, Qu LJ, Gu H (2006) A subgroup of MYB transcription factor genes undergoes

- highly conserved alternative splicing in *Arabidopsis* and rice. *J Exp Bot* 57:1263–1273
- Liu X, Lu T, Yu S, Li Y, Huang Y, Huang T, Zhang L, Zhu J, Zhao Q, Fan D, Mu J, Shanguan Y, Feng Q, Guan J, Ying K, Zhang Y, Lin Z, Sun Z, Qian Q, Lu Y, Han B (2007) A collection of 10,096 indica rice full-length cDNAs reveals highly expressed sequence divergence between *Oryza sativa indica* and *japonica* subspecies. *Plant Mol Biol* 65:403–415
- Lorkovic ZJ, Wiczyorek Kirk DA, Lambermon MH, Filipowicz W (2000) Pre-mRNA splicing in higher plants. *Trends Plant Sci* 5:160–167
- Lund M, Kjems J (2002) Defining a 5' splice site by functional selection in the presence and absence of U1 snRNA 5' end. *RNA* 8:166–179
- Modrek B, Lee C (2002) A genomic view of alternative splicing. *Nat Genet* 30:13–19
- Nagasaki H, Arita M, Nishizawa T, Suwa M, Gotoh O (2005) Species-specific variation of alternative splicing and transcriptional initiation in six eukaryotes. *Gene* 364:53–62
- Ner-Gaon H, Halachmi R, Savaldi-Goldstein S, Rubin E, Ophir R, Fluhr R (2004) Intron retention is a major phenomenon in alternative splicing in *Arabidopsis*. *Plant J* 39:877–885
- Ner-Gaon H, Leviatan N, Rubin E, Fluhr R (2007) Comparative cross-species alternative splicing in plants. *Plant Physiol* 144:1632–1641
- Palusa SG, Ali GS, Reddy AS (2007) Alternative splicing of pre-mRNAs of *Arabidopsis* serine/arginine-rich proteins: regulation by hormones and stresses. *Plant J* 49:1091–1107
- Quesada V, Macknight R, Dean C, Simpson GG (2003) Autoregulation of FCA pre-mRNA processing controls *Arabidopsis* flowering time. *Embo J* 22:3142–3152
- Reddy AS (2007) Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu Rev Plant Biol* 58:267–294
- Roca X, Sachidanandam R, Krainer AR (2005) Determinants of the inherent strength of human 5' splice sites. *RNA* 11:683–698
- Staiger D, Zecca L, Wiczyorek Kirk DA, Apel K, Eckstein L (2003) The circadian clock regulated RNA-binding protein AtGRP7 autoregulates its expression by influencing alternative splicing of its own pre-mRNA. *Plant J* 33:361–371
- Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, Thanaraj TA, Soreq H (2005) Function of alternative splicing. *Gene* 344:1–20
- Su Z, Wang J, Yu J, Huang X, Gu X (2006) Evolution of alternative splicing after gene duplication. *Genome Res* 19:182–189
- Wang BB, Brendel V (2004) The ASRG database: identification and survey of *Arabidopsis thaliana* genes involved in pre-mRNA splicing. *Genome Biol* 5:R102
- Wang BB, Brendel V (2006) Genomewide comparative analysis of alternative splicing in plants. *Proc Natl Acad Sci USA* 103:7175–7180
- Wheelan SJ, Church DM, Ostell JM (2001) Spidey: a tool for mRNA-to-genomic alignments. *Genome Res* 11:1952–1957
- Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, Pham P, Cheuk R, Karlin-Newmann G, Liu SX, Lam B, Sakano H, Wu T, Yu G, Miranda M, Quach HL, Tripp M, Chang CH, Lee JM, Toriumi M, Chan MM, Tang CC, Onodera CS, Deng JM, Akiyama K, Ansari Y, Arakawa T, Banh J, Banno F, Bowser L, Brooks S, Carninci P, Chao Q, Choy N, Enju A, Goldsmith AD, Gurjal M, Hansen NF, Hayashizaki Y, Johnson-Hopson C, Hsuan VW, Iida K, Karnes M, Khan S, Koesema E, Ishida J, Jiang PX, Jones T, Kawai J, Kamiya A, Meyers C, Nakajima M, Narusaka M, Seki M, Sakurai T, Satou M, Tamse R, Vaysberg M, Wallender EK, Wong C, Yamamura Y, Yuan S, Shinozaki K, Davis RW, Theologis A, Ecker JR (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* 302:842–846
- Yoshimura K, Yabuta Y, Ishikawa T, 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
- Zhang Y, Dorey S, Swiderski M, Jones JD (2004) Expression of RPS4 in tobacco induces an AvrRps4-independent HR that requires EDS1, SGT1 and HSP90. *Plant J* 40:213–224