

# Alternative splicing in plants – coming of age

Naeem H. Syed<sup>1</sup>, Maria Kalyna<sup>2</sup>, Yamile Marquez<sup>2</sup>,  
Andrea Barta<sup>2</sup> and John W.S. Brown<sup>1,3</sup>

<sup>1</sup> Division of Plant Sciences, University of Dundee at the James Hutton Institute, Invergowrie, Dundee DD2 5DA, Scotland, UK

<sup>2</sup> Max F. Perutz Laboratories, Medical University of Vienna, Dr Bohr-Gasse 9/3, A-1030 Vienna, Austria

<sup>3</sup> Cell and Molecular Sciences, The James Hutton Institute, Invergowrie, Dundee DD2 5DA, Scotland, UK

Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

**More than 60% of intron-containing genes undergo alternative splicing (AS) in plants. This number will increase when AS in different tissues, developmental stages, and environmental conditions are explored. Although the functional impact of AS on protein complexity is still understudied in plants, recent examples demonstrate its importance in regulating plant processes. AS also regulates transcript levels and the link with nonsense-mediated decay and generation of unproductive mRNAs illustrate the need for both transcriptional and AS data in gene expression analyses. AS has influenced the evolution of the complex networks of regulation of gene expression and variation in AS contributed to adaptation of plants to their environment and therefore will impact strategies for improving plant and crop phenotypes.**

## Frequency and consequences of alternative splicing (AS)

AS (see [Glossary](#)) produces multiple mRNAs from the same gene through variable selection of splice sites during pre-mRNA splicing. It plays a key regulatory role in modulating gene expression during development and in response to environmental signals [1–4]. Regulation of AS in different cell types and under different conditions depends on sequence elements in pre-mRNAs and the interactions of RNA-binding proteins which vary in their concentration and activity. The phenotype of a cell is determined by transcriptional, post-transcriptional, and post-translational networks, which include as a key component the regulated AS of thousands of genes. In humans, where >95% of genes are alternatively spliced, extensive protein diversity is largely a result of AS [5]. Next generation sequencing has revolutionized research into AS and global mapping of human splicing regulatory proteins to their target RNA sequences has led to the development of a splicing code that will allow prediction of tissue-dependent AS [6]. AS not only contributes to proteome diversity but also can generate truncated proteins that are potentially regulatory or detrimental to the cell. It also plays a role in gene expression by regulating transcript levels through production of isoforms which are degraded by the nonsense-mediated decay (NMD) pathway.

In plants, computational analyses of AS events based on expressed sequence tags and more recently high-throughput transcriptome sequencing have examined the frequency of occurrence of AS in different species [42% and 33% of intron-containing genes in *Arabidopsis* (*Arabidopsis thaliana*) and rice (*Oryza sativa*), respectively] and of different types of AS events [7–10]. In *Arabidopsis*, the frequency of occurrence has risen significantly over the past 10 years (Figure 1). Recently, an extensive RNA-seq analysis has significantly increased the observed frequency of AS in *Arabidopsis* to more than 61% of intron-containing genes showing AS. This estimate of 61% of AS is based on analysis of plants grown under normal growth conditions [11] and it is likely that this level will increase further as different tissues at various developmental stages and growth conditions are analyzed [11]. Of the most common types of AS (Figure 2b), intron retention (IR) has been shown to be the most frequent AS event in plants [7–10]. However, some IR events were recently shown to be more likely to represent partially spliced transcripts due to their

## Glossary

**Alternative 5' and 3' splice sites:** use of alternative splice sites at either end of the intron (in the intron or exon sequence) adds or removes sequences.

**Alternative splicing:** precursor mRNAs (pre-mRNAs) are spliced differently to generate different mRNA isoforms.

**Exon skipping/inclusion:** an exon can be removed in a single splicing event or included by two splicing events. Such exons are called alternative exons or cassette exons.

**Heterogeneous nuclear ribonucleoproteins (hnRNP):** RNA-binding proteins which bind RNA in the cell.

**Intron retention (IR):** one or more introns is/are not removed from a pre-mRNA. It is often difficult to determine whether intron retention is due to DNA contamination or partial splicing of pre-mRNAs at the time of RNA extraction or are actively retained.

**Micro-protein (miP):** micro-proteins are usually generated by translation of a transcript containing a premature termination codon, lack one or more functional domains, and have fewer than 100 amino acids [82].

**Nonsense-mediated decay (NMD):** a cellular quality control mechanism that recognizes mRNA transcripts containing PTCs and targets them for degradation.

**Polypyrimidine tract-binding protein (PTB):** an hnRNP protein which binds pyrimidine-rich sequences in RNA to regulate alternative splicing and other mRNA biogenesis processes.

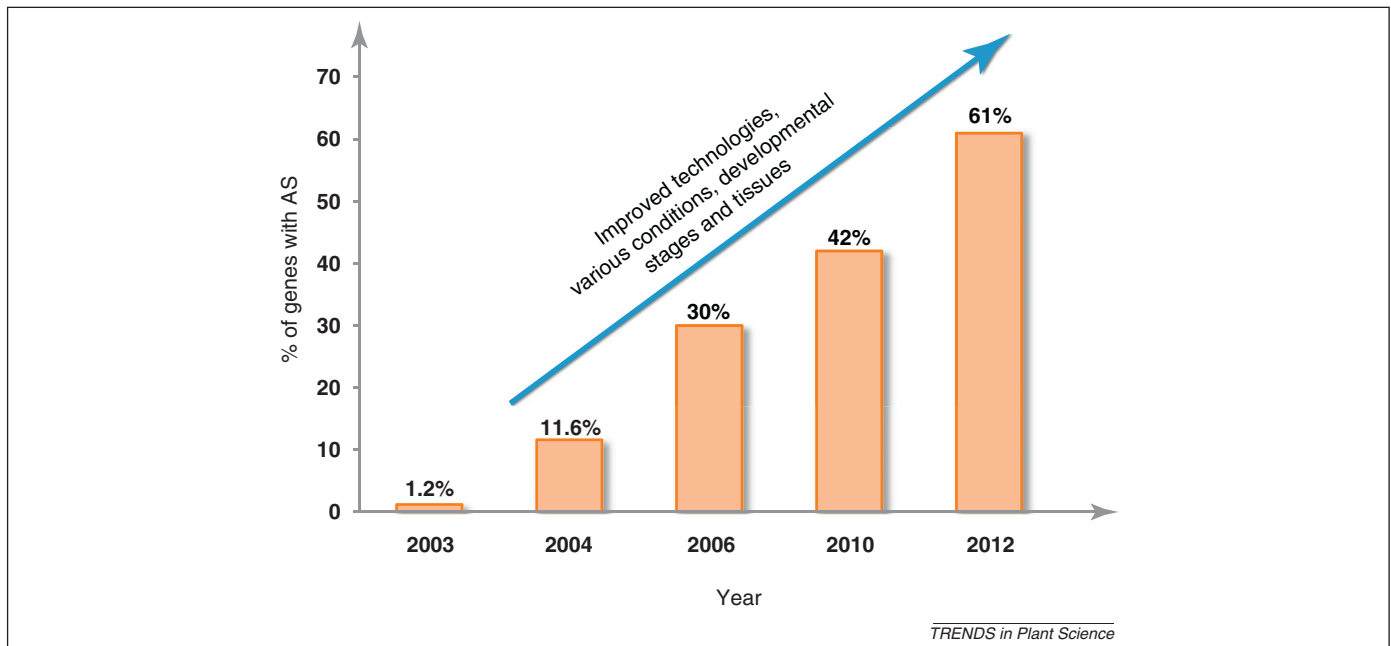
**Precursor messenger RNA (pre-mRNA):** the primary transcript of a gene which is processed to mRNA.

**Premature termination codon (PTC):** a translational stop codon found in transcripts upstream of the authentic stop codon; PTCs can be generated by mutations in DNA, errors in transcription or splicing, or alternative splicing.

**Serine/arginine-rich (SR) proteins:** constitutive and alternative splicing factors containing RNA-binding motifs and a RS-rich domain.

**Small-interfering peptides (siPEPs):** short peptides which interfere with cellular processes.

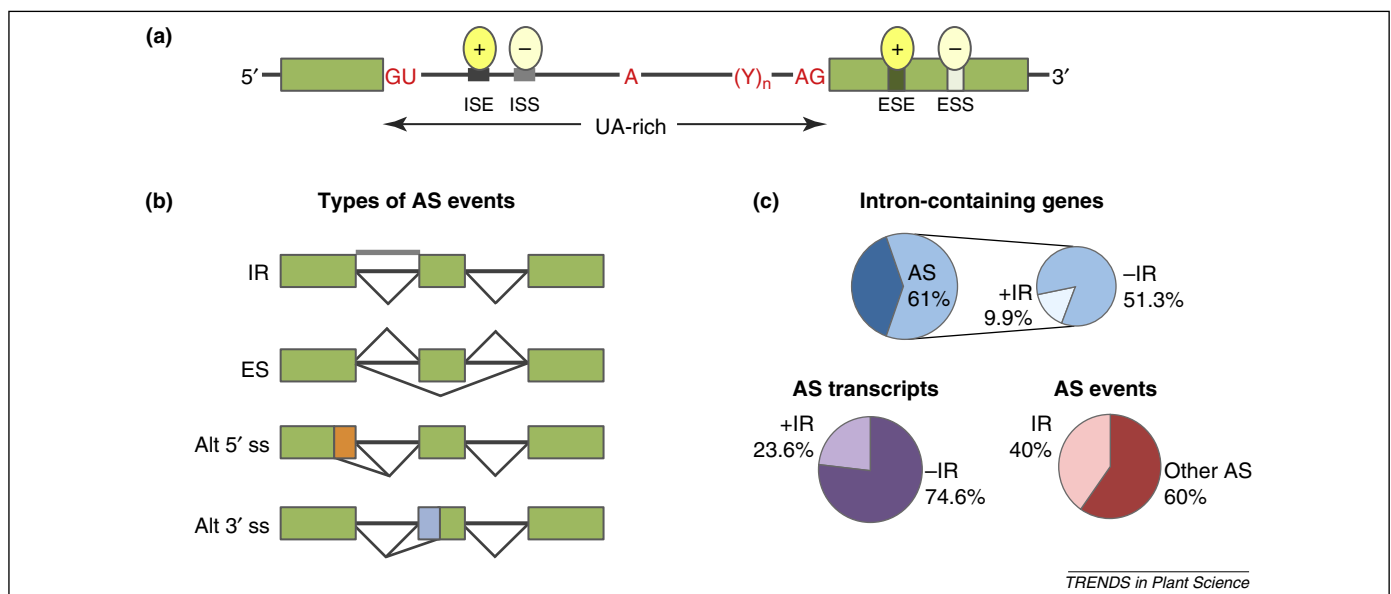
Corresponding author: Brown, J.W.S. ([j.w.s.brown@dundee.ac.uk](mailto:j.w.s.brown@dundee.ac.uk)), ([John.Brown@hutton.ac.uk](mailto:John.Brown@hutton.ac.uk)).



**Figure 1.** Increasing frequency of occurrence of alternative splicing (AS) in *Arabidopsis* with time. In 2003, a study using EST (expressed sequence tag) libraries estimated that only 1.2% of the genes in *Arabidopsis* undergo AS [101]. Subsequently, greater coverage of ESTs and cDNAs libraries allowed the discovery of many more AS events (2004–2006, [7,102–104]). The advent of high-throughput technologies [9,11] has resulted in significant increases in the frequency of AS (almost 60-fold over the past 10 years).

low abundance [11]. In addition, in the genome-wide analysis above, IR was still the most frequent AS event (40%) but it only occurred in assembled AS transcripts of 23% of the genes providing a more reasonable estimate of the impact of IR to AS plants (Figure 2c) [11]. More importantly, 51% of intron-containing genes utilize alternative 5' or 3' splice sites or exon skipping events which can affect the protein

coding sequence or generate unproductive mRNAs to affect transcript levels [11]. The widespread occurrence of AS and the range of functional gene groups which it affects supports an essential role for AS in plant development, physiology, metabolism, and responses to environmental conditions and pathogens, all of which have important consequences on plant/crop phenotypes [7,12–18].

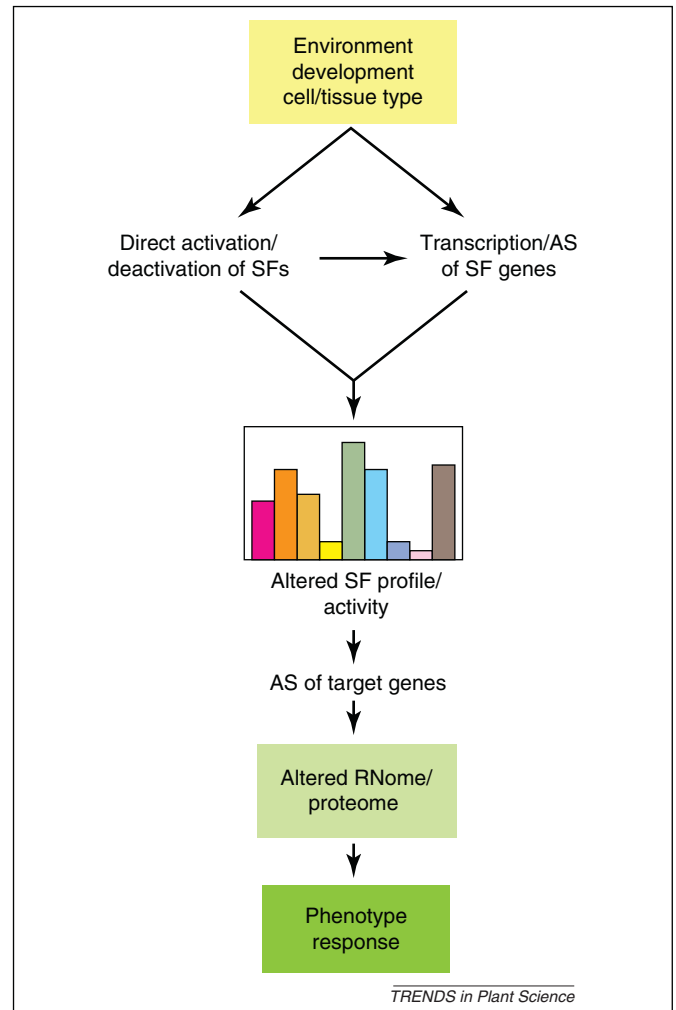


**Figure 2.** Main types of alternative splicing (AS) events and frequency in *Arabidopsis*. (a) Splicing of pre-mRNA is directed by *cis* elements which include splice sites, branch point, and polypyrimidine tract sequences. Selection of alternative splice sites is affected by *trans*-acting factors binding to auxiliary exonic and intronic *cis* elements, termed splicing enhancers and silencers. (b) Types of AS events. (c) Frequency of occurrence of intron retention in *Arabidopsis*. Intron retention is the most frequent AS event in *Arabidopsis* (40%) but its contribution to transcript diversity is much lower [11]. Of the 61% of *Arabidopsis* intron-containing genes with AS, 51% produce AS transcripts which do not involve intron retention (–IR). Among alternatively spliced transcripts, 23.6% contain one or more retained introns (+IR), whereas the rest (74.6%) are produced by other AS events. Colored boxes, exons; lines, introns; GU, 5' splice site which includes highly conserved GU dinucleotide; AG, 3' splice site which includes highly conserved AG dinucleotide; A, branch point adenosine; (Y)<sub>n</sub>, polypyrimidine tract; ovals, positive and negative splicing regulators; carets, splicing events; thick gray line, unspliced (retained) intron. Abbreviations: ESE, exonic splicing enhancers; ESS, exonic splicing silencers; ISE, intronic splicing enhancers; ISS, intronic splicing silencers; Alt 3' ss, alternative 3' splice sites; Alt 5' ss, alternative 5' splice sites; ES, exon skipping; IR, intron retention.

### Regulation of alternative splicing

Assembly of the spliceosome during intron removal and ligation of exons is directed by sequence features of the pre-mRNA. The *cis* sequences in the pre-mRNA include splice sites, branch point, polypyrimidine tract, enhancer, and suppressor sequences (Figure 2a). In addition, it is well documented in plants that UA-richness of introns contributes to their recognition and is essential for efficient splicing. The splice sites situated at the transition between UA-rich introns and GC-rich exons are preferentially selected for splicing (for review see [19]). The compositional bias for UA-richness has been always considered as a distinguishing feature of plant introns; however, recently it has been shown also for animals that exons have higher GC content than introns [20,21]. Regulation of AS depends on *cis* signals and their recognition by *trans*-acting splicing factors. Serine/arginine-rich (SR) and heterogeneous nuclear RNP (hnRNP) proteins function as constitutive and AS splicing factors controlling splice site choice in a concentration-dependent manner [22]. SR proteins are highly conserved in metazoa and plants and, typically, contain one or two RNA recognition motifs (RRMs) and a C-terminal domain (CTD) rich in serine and arginine residues (RS domain). Interestingly, plants possess plant-specific SR proteins and in general they have almost a double number of SR proteins of that in nonphotosynthetic organisms [23]. Many of them have different spatiotemporal expression patterns, implicating diverse target specificities and biological functions [24]. hnRNP proteins, by contrast, constitute a structurally diverse group of RNA-binding proteins associated with nascent pre-mRNA molecules and, in addition to their role in splicing, are involved in a variety of molecular processes [25]. SR and hnRNP proteins bind splicing signals and intronic and exonic enhancer/silencer sequences and through multicomponent interactions with other splicing factors (including cell- and tissue-specific factors) determine splice site selection and where the spliceosome assembles. In general, SR proteins promote splicing and the hnRNP proteins inhibit splice site selection. However, both SRs and hnRNPs can also have opposite functions with, for example, SRSF10 (also known as SRp38) [26] being a negative regulator and polypyrimidine tract-binding protein (PTB; also known as hnRNPI) being a positive regulator [27].

Variation in abundance and activity of splicing factors determines the AS profiles of target genes and therefore their differential expression under different growth conditions and during development [28–31]. Importantly, different growth conditions modulate AS of SR and hnRNP genes causing dynamic changes in the splicing factor profile, further impacting expression of target genes (Figure 3). For example, AS of many *Arabidopsis* SR genes is affected by temperature, light, salt, hormones, etc. [30–32]. In addition, the activity or localization of SR proteins can be affected by phosphorylation [4,33,34] and protein kinases which phosphorylate plant SR proteins have been identified [35–37]. Not surprisingly therefore, SR protein overexpression and knockout lines show a variety of developmental and growth phenotypes, demonstrating the importance of these proteins to normal growth and development and broad effects on gene expression [28]. The best-studied plant



**Figure 3.** Dynamic regulation of expression by alternative splicing. Developmental or environmental cues activate signaling pathways which can directly modulate splicing factor activity by post-translational modifications, relocalization, etc. Signaling also directs changes in transcription of splicing factor genes and alternative splicing of these genes changes the abundance, composition, and activity of the splicing factor population. Expression of other target genes including transcription factors is also modulated by alternative splicing responding to the dynamic changes in splicing factor profiles. Changes in the proteome feedback to transcription and alternative splicing (and other post-transcriptional mechanisms) ultimately generating the cellular and organismal phenotype and response. Boxes of colored bars, abundance and/or composition of SFs. Abbreviations: AS, alternative splicing; SF, splicing factor.

hnRNP proteins to date are the *Arabidopsis* orthologs of the animal negative splicing regulator, PTB [38], and the glycine-rich RNA-binding proteins, GRP7 and GRP8, components of a slave oscillator coupled to the circadian clock [39–41]. The *Arabidopsis* PTBs auto- and crossregulate the AS within the family [38]. GRP7 and GRP8 also autoregulate their own AS and crossregulate each other's AS to generate unproductive mRNAs which are targeted by NMD to reduce mRNA and protein levels [39,41]. Finally, the nuclear cap-binding complex (the CBC) consists of two subunits, AtCBP20 and AtCBP80, and AtCBP20 contains a canonical RNA-binding domain (RBD). Mutation of these subunits showed the CBC to be involved in AS of at least some *Arabidopsis* genes and to preferentially influence AS of the first intron, particularly at the 5' splice site [42].

Although the abundance and activity of splicing factors determines the AS profiles of downstream genes, very few

examples of the biological relevance of AS-derived protein isoforms of splicing factors have been published. Functional diversity of AS isoforms was elegantly demonstrated when two isoforms of SR45 (which differ by only eight amino acids and include putative phosphorylation sites), differentially complemented petal or root developmental phenotypes in the *sr45* mutant [43]. Hence, isoforms with very similar sequences can have substantially different morphological outcomes which will reflect a host of gene expression changes in different organs of the plant.

### Epigenetic control of alternative splicing

An extensive body of evidence from human and yeast shows that splicing is often coupled to transcription [44–46]. The CTD of RNA polymerase II serves as a landing pad for recruitment of proteins involved in capping, splicing, polyadenylation, and export [47–52]. The rate of transcription elongation by RNA polymerase II may affect splice site choice and thereby AS outcomes [45,46]. A slower rate of transcription tends to favor inclusion of weak upstream exons before the splicing complex is committed to splicing of a stronger downstream exon [45,46]. Importantly, efficiency of the splicing process may also influence the rate of transcription elongation, where, for example, transient pausing of RNAP II at the 3' splice site of an intron coincides with the appearance of spliced product [53]. The rate of transcript elongation may depend on the chromatin state [50]. For example, nucleosome occupancy varies along a gene with GC-rich exons being relatively nucleosome-rich compared with GC poor introns [20,21,52,54] and transcription through nucleosome-rich regions with compact chromatin tends to be slower [50]. Furthermore, nucleosome occupancy is also lower in alternatively spliced exons compared with constitutively spliced exons [20]. The relations between chromatin state, nucleosome occupancy, RNAP II elongation rates, splicing efficiency, and AS outcomes are key to understanding AS regulation. Indeed, a direct link between histone modifications and AS was demonstrated recently [55]. The splicing of two PTB-dependent mutually exclusive exons in the human fibroblast growth factor receptor 2 (FGFR2) gene depended on histone modifications H3K36me3 and H3K4me1 acting in one direction, and H3K27me3, H3K4me3, and H3K9me1 acting in the opposite direction. The changes in chromatin state are read by the chromatin-binding adapter protein MRG15 which then recruits the splicing factor PTB to the pre-mRNA and affects splicing outcomes [45,46,55]. In plants, the direct link between AS and either chromatin state or RNAP II elongation rates (transcription) has not yet been demonstrated. However, clearly the impact of environmental cues and signaling pathways on chromatin and transcription to generate different AS variants has the potential to further our understanding how plants respond to their ever-changing environment.

### Alternative splicing affects protein complexity and transcript levels and stability

The different forms of AS (Figure 2b) and, in particular, alternative 5' and 3' splice site selection and exon skipping

often lead to changes in protein sequences from the inclusion or removal of a few amino acids (see above for SR45 [43]) to large regions of proteins affecting protein domains, or generating changes in N-terminal or C-terminal regions [56,57]. Different protein isoforms have the potential for differential functions as highlighted by several recent studies. For example, cold-induced sweetening is a serious problem in potatoes where starch is converted to glucose and fructose by vacuolar acid invertase: lines showing resistance to cold-induced sweetening have higher expression of two splice variants (INH2 $\alpha$  and INH2 $\beta$ ) of the invertase inhibitor gene (INH2) [58]. The flavin-dependent monooxygenase gene, *YUCCA4*, involved in auxin biosynthesis, undergoes tissue-specific AS to generate isoforms with different intracellular localization. One isoform is expressed in all tissues and is distributed throughout the cytosol, whereas a second is restricted to flowers and is attached to the endoplasmic reticulum [59]. *In silico* analysis of MADS-box MIKC-type (MADS, Intervening, Keratin-like and C-terminal domain) transcription factors in *Arabidopsis* predicted protein isoforms which affect dimerization properties or higher order protein complex formation [57]. The potential for AS to influence function was shown by the differential effects on flowering time and floral development of overexpression of isoforms of *SHORT VEGETATIVE PHASE*, consistent with their different protein–protein interactions [57]. This systematic analysis of a large gene family illustrates the potential of AS to affect key protein domains and function as well as the impact of AS in the evolution of gene families and protein interaction networks.

AS can regulate mRNA levels through the production of AS isoforms containing premature termination codons (PTCs) which are targeted for degradation by NMD [60,61]. Plants possess orthologs of the key eukaryotic NMD proteins, UPF1, UPF2, UPF3, and SMG-7 (but not SMG-1, SMG-5, or SMG-6) and these have been shown to be involved in degrading mRNAs with PTCs [62–68]. Rules for NMD in plants have been established mainly by studying mutations in a small number of model transcripts. The mechanisms of recognition of NMD substrates in plants appear to be fundamentally similar to those in other eukaryotes relying on the distance from a PTC to the 3' end of the transcript (long 3'UTR) or downstream splice junctions (splicing-dependent) [18,64,69–71]. Recently, by analyzing a large population of endogenous *Arabidopsis* transcripts, coupled AS, and NMD has been shown to be a widespread mechanism for regulating gene expression with 11–18% of alternatively spliced transcripts being turned over by NMD [18]. This study also showed that transcripts containing PTCs which are NMD substrates are often readily detectable and can contribute significantly to steady-state transcript levels of genes. In addition, some PTC-containing transcripts were not turned over by NMD. For example, some transcripts containing retained introns or parts of introns were unaffected in NMD mutants, suggesting that not all NMD triggering signals or transcript arrangements are understood [18]. Thus, the generation of unproductive AS transcripts can influence the levels of functional mRNAs (full-length protein coding), as has been observed in the regulation of human SR and



hnRNP proteins through AS and ultraconserved elements [72–74]. Similarly in plants, SR and PTB genes are regulated by AS [28,38,75] which gives rise to PTC-containing transcripts, suggesting a regulatory function via unproductive mRNAs.

### Alternative splicing in the plant circadian clock

Regulation of expression by AS generating unproductive transcripts has recently been demonstrated for core circadian clock genes in *Arabidopsis*. Circadian clocks have approximately 24-h rhythms and allow organisms to anticipate the day–night cycle and coordinate their genetic, biochemical, and physiological responses [76–78]. Core clock gene expression is regulated at multiple different levels: transcription, protein degradation, and modification with AS being an emerging theme in regulation of the clock [77,78]. Until recently, examples of AS in clock genes and their functional significance were rare. For example, an IR event in CCA1 was conserved in at least four plant species and levels of IR-containing transcripts increased in high light and decreased in the cold [9]. In addition, a mutant in PROTEIN ARGININE METHYL TRANSFERASE 5 (PRMT5) showed a longer circadian period and dramatic changes in the levels of unproductive AS transcripts of PRR9 [79,80]. Recently, extensive AS in the majority of the core clock genes in *Arabidopsis* was identified and dynamic changes in AS profiles for many AS events were observed in response to changes in temperature and particularly to lower temperatures [81]. AS events were either induced by low temperatures or increased in abundance to 10–50% of the transcripts; the majority of these events were nonproductive resulting in a reduction of functional mRNAs potentially impacting protein levels [81]. Furthermore, the partially redundant gene pairs, LHY and CCA1, and PRR7 and PRR9 behaved differently with respect to AS, implying functional differences between these related genes. Thus, temperature-associated AS modulating the balance between productive and nonproductive mRNA isoforms is an additional mechanism involved in the operation and control of the plant circadian clock.

### Alternative splicing and regulation by small interfering peptides/micro-proteins

The fate of alternatively spliced transcripts containing PTCs is expected to be degradation by the NMD pathway but some PTC-containing transcripts are stable and appear to avoid the NMD machinery [18]. PTC-containing transcripts also have the potential to be translated into truncated proteins or peptides. In plants, intron-containing mRNA transcripts were found associated with polyosomes and recently ribosome profiling in mouse has identified novel upstream open reading frames (ORFs) and ORFs in long noncoding RNAs [82,83]. In animal and plant systems, small interfering peptides (siPEPs) or micro-proteins (miPs) named after their analogy with siRNAs and miRNAs have been described [84,85]. Such peptides can have altered functionality by only containing particular domains (e.g., DNA binding, transcriptional activators) and can act as both positive and negative regulators and affect regulatory feedback loops [86]. For

example, in animals, an AS isoform generates a miP of the ETS1 transcription factor which regulates growth and development responses, lacks the transactivation domains, and interacts physically with ETS1 blocking ETS1-mediated expression of target genes in a dominant negative manner [86–88]. Similarly, miP AS protein isoforms of the animal transcription factor MEIS2 also interact in a dominant negative manner [89]. Interestingly, one of the splice variants (MEIS2E) is structurally similar to a plant protein ‘KNATM’, which is a member of the TALE homeodomain transcription regulators (controlling meristem formation, organ position and morphogenesis, and some aspects of reproductive phase) [86]. KNATM, however, lacks a homeodomain and by forming nonfunctional heterodimers with the BELL TALE protein regulates leaf pattern [90,91]. It is intriguing that a protein generated via AS in animals appears to exist as a miP equivalent in plants.

Genome-wide analysis of AS in *Arabidopsis* suggested that 78% of alternative transcripts introduced in-frame PTCs [9], providing a huge potential for production of miPs. Examples of miPs in plants with functional consequences are rare but a recent study showed that the *Arabidopsis* transcription factor gene, *IDD14*, produces a splice variant (IDD14 $\beta$ ) which lacks the DNA-binding domain but interacts with the functional IDD14 $\alpha$  isoform to produce heterodimers. The IDD14 $\alpha/\beta$  heterodimer has reduced binding affinity for the promoter of the Qua-Quine Starch (QQS) gene which regulates starch accumulation by initiating starch degradation [84]. Starch accumulation is one response of plants to cold and as the IDD14 $\beta$  splice variant is only expressed under cold conditions, starch degradation is reduced providing an AS/miP-dependent strategy for maintaining starch reserves at low temperatures. Interestingly, the core clock proteins CCA1 and LHY can both homo- and heterodimerize [92,93], and different combinations have different binding affinity to their target sequences [94]. It is interesting to speculate whether the extensive production of PTC-containing transcripts in core clock genes by AS could add a further level of regulation by miPs. Furthermore, siPEPs/miPs offer another mechanism to modify or knock-down expression of endogenous genes. Artificial siPEPs/miPs encoding dimerization domains could be transformed into plants to reduce the activity of a target gene. As they function at the protein level and depend on homo- or heterodimerization, this should improve specificity and reduce off-target silencing [85].

### Alternative splicing diversity in ecotypes and polyploids

Extensive AS occurs under altered growth or stress conditions in plants. Similarly, extensive variation in AS is expected in diverse ecotypes adapted to very different climates thereby achieving environmental and phenotypic plasticity. To address such diversity, genomic and transcriptomic sequencing is being performed on geographically and phenotypically diverse accessions of *Arabidopsis*. Sequencing of the genomes and transcriptomes of 18 *Arabidopsis* accessions identified extensive single nucleotide polymorphism and indel variation among the genotypes [95]. When compared with Col-0 (TAIR10) one-third of protein coding genes were disrupted/altered in at least

one accession although re-annotation restored coding potential in most cases. Sequence variations affected translation start and stop sites, introduced PTCs, or changed the frame of the coding sequence, or potentially generated protein isoforms in different accessions. Two-thirds of 2572 genes with disrupted splice sites when compared to TAIR10 had new splice sites and a quarter of these sites were close to the splice sites in Col-0 [95]. Clearly, naturally occurring sequence variation can disrupt splice sites and RNA-binding motifs for splicing factors. Such mutations impact protein expression and activity and provide a basis for selection for adaptation of different ecotypes to their environments [7,8,10,57].

Extensive duplication or polyploidization has occurred in the evolutionary history of many plant species [96]. Theoretically, such events could generate immediate and drastic changes in the abundance, composition, and activity of splicing factors which in turn could affect splice site choice among variable analogous splice site sequences. Further mutation, gene loss, or changes in expression or protein functionality provide the basis for selection and continued evolution of the species. Recently, AS patterns have been studied among natural and synthetic polyploids of *Brassica napus*. Interestingly, two independently synthesized lines showed parallel loss of AS events after polyploidy [97]. This is intriguing because it shows that even in two independent events of polyploidy, using the same species, results in identical pattern of AS loss, pointing towards a non-random response of the so-called 'genomic shock' after two genomes physically interact with each other. The same study also showed that 26–30% of the duplicated genes show changes in AS, compared with the parents, with some showing organ specificity or response to abiotic stress [97]. It is likely that AS has played an important role in the evolution and adaptation of cultivated crops to different environmental conditions and niches, because many crop species are polyploids and this whole area will be one of great interest in the future, particularly with the power of high-throughput sequencing.

### Evolutionary aspects of alternative splicing factor diversity

The *Arabidopsis* genome encodes 18 SR proteins which is nearly double the number found in humans. At least 12 of the 18 SR genes are located in duplicated genomic regions and the current data indicate that most of them have different spatiotemporal expression patterns, suggesting functional diversification [24]. In different plant lineages, the number of paralogous SR genes is highly variable. For example, the plant-specific RS subfamily of SR proteins is encoded by four genes in *Arabidopsis*, two in rice, at least five in *Pinus taeda* and by one gene both in *Physcomitrella patens* and *Chlamydomonas reinhardtii* [98]. Differential or common, redundant functions of SR paralogs in different species remain to be determined. However, it is interesting that several SR paralogs and orthologs are regulated by AS events conserved from unicellular green algae to land plants. These events occur in the analogous long introns situated in the RRM coding regions [98,99] and, moreover, they involve unusually highly conserved sequences around alternative splice sites [98], suggesting an important biological function of such

regulation. A recent systematic survey of SR genes in 27 eukaryotic genomes showed that flowering plants on average possess nearly double the number of SR genes than nonphotosynthetic species [23]. Moreover, most of the plant SR genes are under purifying selection, ensuring that paralogous genes which originated due to the duplication events maintain their structure and function, whereas redundancy is reduced via diversification of gene expression [23].

### Future challenges in alternative splicing research in plants

AS research in plants has made substantial progress in the past 4–5 years. The ever-increasing number of plant genes with AS and the processes in which they are involved point to the importance of understanding the mechanisms and regulation of AS and the functions of AS. The functional impact of AS is one of the most important questions – this is largely due to the relatively small number of examples of AS for which differential functions have been demonstrated for different protein isoforms. However, we draw a parallel with research to discover the extent of AS in plants. Around 10 years ago the first estimate was only 1.2% of plant genes showing AS! With massively improved technologies this number has now grown to more than 61% (Figure 1). Similarly, the increasing number of plant genome sequences and the generation of vast transcriptome data will allow computational analyses to identify conservation of AS events across species and tissue-, developmental-stage and environment-specific regulation of AS providing evidence of functionality. The number of functional examples of AS, whether at the mRNA transcript stability level or protein function, continues to grow and in turn is stimulating wider interest in AS in the plant community. High-throughput sequencing will also address dynamic changes in AS in development and under different environmental conditions and stresses, and how variation in AS patterns in different ecotypes and polyploids contributes to plasticity and adaptation of plant species. Furthermore, we need to understand how signaling pathways affect splicing factor activity directly or via chromatin modification and how transcriptional and AS networks interact [100]. AS is a major mechanism by which plants modulate and fine-tune expression of their genes. The next 5 years will see an explosion of knowledge of the functional significance of AS and understanding its contribution to the complexity of gene expression will offer new opportunities in approaches to modifying plant function for improved phenotypes.

### Acknowledgments

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) [BB/G024979/1 (ERA-NET Plant Genomics (PASAS)); the Scottish Government Rural and Environment Science and Analytical Services Division (RESAS); the Austrian Science Fund (FWF) [SFB 1710, 1711; DK W1207; ERA-NET Plant Genomics (PASAS) I254; SFB RNAREg F43-P10]; the Austria Genomic Program (GENAU III) [ncRNAs]; and the EU FP6 Program Network of Excellence on Alternative Splicing (EURASNET) [LSHG-CT-2005-518238].

### References

- Black, D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* 72, 291–336
- Graveley, B.R. (2001) Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* 17, 100–107

- 3 Lareau, L.F. *et al.* (2004) The evolving roles of alternative splicing. *Curr. Opin. Struct. Biol.* 14, 273–282
- 4 Stamm, S. *et al.* (2005) Function of alternative splicing. *Gene* 344, 1–20
- 5 Pan, Q. *et al.* (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–1415
- 6 Barash, Y. *et al.* (2010) Deciphering the splicing code. *Nature* 465, 53–59
- 7 Wang, B.B. and Brendel, V. (2006) Genome-wide comparative analysis of alternative splicing in plants. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7175–7180
- 8 Wang, B.B. *et al.* (2008) Cross-species EST alignments reveal novel and conserved alternative splicing events in legumes. *BMC Plant Biol.* 8, 17
- 9 Filichkin, S.A. *et al.* (2010) Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res.* 20, 45–58
- 10 Barbazuk, W.B. *et al.* (2008) Genome-wide analyses of alternative splicing in plants: opportunities and challenges. *Genome Res.* 18, 1381–1392
- 11 Marquez, Y. *et al.* (2012) Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Res.* 22, 1184–1195
- 12 Ali, G.S. and Reddy, A.S. (2008) Regulation of alternative splicing of pre-mRNAs by stresses. *Curr. Top. Microbiol. Immunol.* 326, 257–275
- 13 Chen, F.C. *et al.* (2007) Plant gene and alternatively spliced variant annotator. A plant genome annotation pipeline for rice gene and alternatively spliced variant identification with cross-species expressed sequence tag conservation from seven plant species. *Plant Physiol.* 143, 1086–1095
- 14 Halterman, D.A. *et al.* (2003) Powdery mildew-induced *Mla* mRNAs are alternatively spliced and contain multiple upstream open reading frames. *Plant Physiol.* 131, 558–567
- 15 Lorkovic, Z.J. (2009) Role of plant RNA-binding proteins in development, stress response and genome organization. *Trends Plant Sci.* 14, 229–236
- 16 Quesada, V. *et al.* (2003) Autoregulation of *FCA* pre-mRNA processing controls *Arabidopsis* flowering time. *EMBO J.* 22, 3142–3152
- 17 Zhu, J. *et al.* (2007) Interplay between cold-responsive gene regulation, metabolism and RNA processing during plant cold acclimation. *Curr. Opin. Plant Biol.* 10, 290–295
- 18 Kalyna, M. *et al.* (2011) Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in *Arabidopsis*. *Nucleic Acids Res.* 40, 2454–2469
- 19 Lorkovic, Z.J. *et al.* (2000) Pre-mRNA splicing in higher plants. *Trends Plant Sci.* 5, 160–167
- 20 Schwartz, S. *et al.* (2009) Chromatin organization marks exon–intron structure. *Nat. Struct. Mol. Biol.* 16, 990–995
- 21 Tilgner, H. *et al.* (2009) Nucleosome positioning as a determinant of exon recognition. *Nat. Struct. Mol. Biol.* 16, 996–1001
- 22 Matlin, A.J. *et al.* (2005) Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol.* 6, 386–398
- 23 Richardson, D.N. *et al.* (2011) Comparative analysis of serine/arginine-rich proteins across 27 eukaryotes: insights into sub-family classification and extent of alternative splicing. *PLoS ONE* 6, e24542
- 24 Kalyna, M. and Barta, A. (2004) A plethora of plant serine/arginine-rich proteins: redundancy or evolution of novel gene functions? *Biochem. Soc. Trans.* 32, 561–564
- 25 Martinez-Contreras, R. *et al.* (2007) hnRNP proteins and splicing control. *Adv. Exp. Med. Biol.* 623, 123–147
- 26 Feng, Y. *et al.* (2008) Phosphorylation switches the general splicing repressor SRp38 to a sequence-specific activator. *Nat. Struct. Mol. Biol.* 15, 1040–1048
- 27 Xue, Y. *et al.* (2009) Genome-wide analysis of PTB–RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol. Cell* 36, 996–1006
- 28 Barta, A. *et al.* (2008) Plant SR proteins and their functions. *Curr. Top. Microbiol. Immunol.* 326, 83–102
- 29 Lazar, G. and Goodman, H.M. (2000) The *Arabidopsis* splicing factor SR1 is regulated by alternative splicing. *Plant Mol. Biol.* 42, 571–581
- 30 Palusa, S.G. *et al.* (2007) Alternative splicing of pre-mRNAs of *Arabidopsis* serine/arginine-rich proteins: regulation by hormones and stresses. *Plant J.* 49, 1091–1107
- 31 Tanabe, N. *et al.* (2007) Differential expression of alternatively spliced mRNAs of *Arabidopsis* SR protein homologs, atSR30 and atSR45a, in response to environmental stress. *Plant Cell Physiol.* 48, 1036–1049
- 32 Duque, P. (2011) A role for SR proteins in plant stress responses. *Plant Signal. Behav.* 6, 49–54
- 33 Stamm, S. (2002) Signals and their transduction pathways regulating alternative splicing: a new dimension of the human genome. *Hum. Mol. Genet.* 11, 2409–2416
- 34 Stamm, S. (2008) Regulation of alternative splicing by reversible protein phosphorylation. *J. Biol. Chem.* 283, 1223–1227
- 35 de la Fuente van Bentem, S. *et al.* (2006) Phosphoproteomics reveals extensive in vivo phosphorylation of *Arabidopsis* proteins involved in RNA metabolism. *Nucleic Acids Res.* 34, 3267–3278
- 36 de la Fuente van Bentem, S. *et al.* (2003) The subcellular localization of plant protein phosphatase 5 isoforms is determined by alternative splicing. *Plant Physiol.* 133, 702–712
- 37 Savaldi-Goldstein, S. *et al.* (2003) Alternative splicing modulation by a LAMMER kinase impinges on developmental and transcriptome expression. *Plant Cell* 15, 926–938
- 38 Stauffer, E. *et al.* (2010) Polypyrimidine tract-binding protein homologues from *Arabidopsis* underlie regulatory circuits based on alternative splicing and downstream control. *Plant J.* 64, 243–255
- 39 Schoning, J.C. *et al.* (2007) Auto-regulation of the circadian slave oscillator component AtGRP7 and regulation of its targets is impaired by a single RNA recognition motif point mutation. *Plant J.* 52, 1119–1130
- 40 Schuttpelz, M. *et al.* (2008) Changes in conformational dynamics of mRNA upon AtGRP7 binding studied by fluorescence correlation spectroscopy. *J. Am. Chem. Soc.* 130, 9507–9513
- 41 Schoning, J.C. *et al.* (2008) Reciprocal regulation of glycine-rich RNA-binding proteins via an interlocked feedback loop coupling alternative splicing to nonsense-mediated decay in *Arabidopsis*. *Nucleic Acids Res.* 36, 6977–6987
- 42 Raczynska, K.D. *et al.* (2010) Involvement of the nuclear cap-binding protein complex in alternative splicing in *Arabidopsis thaliana*. *Nucleic Acids Res.* 38, 265–278
- 43 Zhang, X.N. and Mount, S.M. (2009) Two alternatively spliced isoforms of the *Arabidopsis* SR45 protein have distinct roles during normal plant development. *Plant Physiol.* 150, 1450–1458
- 44 Carrillo Oesterreich, F. *et al.* (2010) Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol. Cell* 40, 571–581
- 45 Luco, R.F. *et al.* (2011) Epigenetics in alternative pre-mRNA splicing. *Cell* 144, 16–26
- 46 Luco, R.F. and Misteli, T. (2011) More than a splicing code: integrating the role of RNA, chromatin and non-coding RNA in alternative splicing regulation. *Curr. Opin. Genet. Dev.* 21, 366–372
- 47 de Almeida, S.F. and Carmo-Fonseca, M. (2008) The CTD role in cotranscriptional RNA processing and surveillance. *FEBS Lett.* 582, 1971–1976
- 48 Fong, N. *et al.* (2003) A 10 residue motif at the C-terminus of the RNA pol II CTD is required for transcription, splicing and 3' end processing. *EMBO J.* 22, 4274–4282
- 49 Goldstrohm, A.C. *et al.* (2001) The transcription elongation factor CA150 interacts with RNA polymerase II and the pre-mRNA splicing factor SF1. *Mol. Cell. Biol.* 21, 7617–7628
- 50 Howe, K.J. (2002) RNA polymerase II conducts a symphony of pre-mRNA processing activities. *Biochim. Biophys. Acta* 1577, 308–324
- 51 McCracken, S. *et al.* (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* 385, 357–361
- 52 Schwartz, S. and Ast, G. (2010) Chromatin density and splicing destiny: on the cross-talk between chromatin structure and splicing. *EMBO J.* 29, 1629–1636
- 53 Alexander, R.D. *et al.* (2010) Splicing-dependent RNA polymerase pausing in yeast. *Mol. Cell* 40, 582–593
- 54 Spies, N. *et al.* (2009) Biased chromatin signatures around polyadenylation sites and exons. *Mol. Cell* 36, 245–254
- 55 Luco, R.F. *et al.* (2010) Regulation of alternative splicing by histone modifications. *Science* 327, 996–1000
- 56 Airoidi, C.A. *et al.* (2010) Single amino acid change alters the ability to specify male or female organ identity. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18898–18902



- 57 Severing, E.I. *et al.* (2012) Predicting the impact of alternative splicing on plant MADS domain protein function. *PLoS ONE* 7, e30524
- 58 Brummell, D.A. *et al.* (2011) Induction of vacuolar invertase inhibitor mRNA in potato tubers contributes to cold-induced sweetening resistance and includes spliced hybrid mRNA variants. *J. Exp. Bot.* 62, 3519–3534
- 59 Kriechbaumer, V. *et al.* (2012) Alternative splicing of the auxin biosynthesis gene *YUCCA4* determines its subcellular compartmentation. *Plant J.* 70, 292–302
- 60 Rebbapragada, I. and Lykke-Andersen, J. (2009) Execution of nonsense-mediated mRNA decay: what defines a substrate? *Curr. Opin. Cell Biol.* 21, 394–402
- 61 Nicholson, P. *et al.* (2010) Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors. *Cell. Mol. Life Sci.* 67, 677–700
- 62 Arciga-Reyes, L. *et al.* (2006) UPF1 is required for nonsense-mediated mRNA decay (NMD) and RNAi in *Arabidopsis*. *Plant J.* 47, 480–489
- 63 Hori, K. and Watanabe, Y. (2005) UPF3 suppresses aberrant spliced mRNA in *Arabidopsis*. *Plant J.* 43, 530–540
- 64 Kerenyi, Z. *et al.* (2008) Inter-kingdom conservation of mechanism of nonsense-mediated mRNA decay. *EMBO J.* 27, 1585–1595
- 65 Riehs, N. *et al.* (2008) *Arabidopsis* SMG7 protein is required for exit from meiosis. *J. Cell Sci.* 121, 2208–2216
- 66 Wu, J. *et al.* (2007) Nonsense-mediated mRNA decay (NMD) silences the accumulation of aberrant trypsin proteinase inhibitor mRNA in *Nicotiana attenuata*. *Plant J.* 51, 693–706
- 67 Yoine, M. *et al.* (2006) *Arabidopsis* UPF1 RNA helicase for nonsense-mediated mRNA decay is involved in seed size control and is essential for growth. *Plant Cell Physiol.* 47, 572–580
- 68 Yoine, M. *et al.* (2006) The *lba1* mutation of *UPF1* RNA helicase involved in nonsense-mediated mRNA decay causes pleiotropic phenotypic changes and altered sugar signalling in *Arabidopsis*. *Plant J.* 47, 49–62
- 69 Kertesz, S. *et al.* (2006) Both introns and long 3'-UTRs operate as cis-acting elements to trigger nonsense-mediated decay in plants. *Nucleic Acids Res.* 34, 6147–6157
- 70 Nyiko, T. *et al.* (2009) Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. *Plant Mol. Biol.* 71, 367–378
- 71 Schwartz, A.M. *et al.* (2006) Stability of plant mRNAs depends on the length of the 3'-untranslated region. *Biochemistry (Mosc)* 71, 1377–1384
- 72 Lareau, L.F. *et al.* (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* 446, 926–929
- 73 Ni, J.Z. *et al.* (2007) Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* 21, 708–718
- 74 Palusa, S.G. and Reddy, A.S. (2010) Extensive coupling of alternative splicing of pre-mRNAs of serine/arginine (SR) genes with nonsense-mediated decay. *New Phytol.* 185, 83–89
- 75 Reddy, A.S. (2007) Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu. Rev. Plant Biol.* 58, 267–294
- 76 Doherty, C.J. and Kay, S.A. (2010) Circadian control of global gene expression patterns. *Annu. Rev. Genet.* 44, 419–444
- 77 Petrillo, E. *et al.* (2011) Alternative splicing adds a new loop to the circadian clock. *Commun. Integr. Biol.* 4, 284–286
- 78 Staiger, D. and Green, R. (2011) RNA-based regulation in the plant circadian clock. *Trends Plant Sci.* 16, 517–523
- 79 Hong, S. *et al.* (2010) Type II protein arginine methyltransferase 5 (PRMT5) is required for circadian period determination in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21211–21216
- 80 Sanchez, S.E. *et al.* (2010) A methyl transferase links the circadian clock to the regulation of alternative splicing. *Nature* 468, 112–116
- 81 James, A.B. *et al.* (2012) Alternative splicing mediates responses of the *Arabidopsis* circadian clock to temperature changes. *Plant Cell* 24, 961–981
- 82 Ingolia, N.T. *et al.* (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802
- 83 Ner-Gaon, H. *et al.* (2004) Intron retention is a major phenomenon in alternative splicing in *Arabidopsis*. *Plant J.* 39, 877–885
- 84 Seo, P.J. *et al.* (2011) Two splice variants of the IDD14 transcription factor competitively form nonfunctional heterodimers which may regulate starch metabolism. *Nat. Commun.* 2, 303
- 85 Seo, P.J. *et al.* (2011) Competitive inhibition of transcription factors by small interfering peptides. *Trends Plant Sci.* 16, 541–549
- 86 Staudt, A.C. and Wenkel, S. (2011) Regulation of protein function by 'microProteins'. *EMBO Rep.* 12, 35–42
- 87 Laitem, C. *et al.* (2009) Ets-1 p27: a novel Ets-1 isoform with dominant-negative effects on the transcriptional properties and the subcellular localization of Ets-1 p51. *Oncogene* 28, 2087–2099
- 88 Lee, G.M. *et al.* (2005) The structural and dynamic basis of Ets-1 DNA binding autoinhibition. *J. Biol. Chem.* 280, 7088–7099
- 89 Yang, Y. *et al.* (2000) Three-amino acid extension loop homeodomain proteins Meis2 and TGIF differentially regulate transcription. *J. Biol. Chem.* 275, 20734–20741
- 90 Hamant, O. and Pautot, V. (2010) Plant development: a TALE story. *C. R. Biol.* 333, 371–381
- 91 Magnani, E. and Hake, S. (2008) KNOX lost the OX: the *Arabidopsis* KNATM gene defines a novel class of KNOX transcriptional regulators missing the homeodomain. *Plant Cell* 20, 875–887
- 92 Lu, S.X. *et al.* (2009) CIRCADIAN CLOCK ASSOCIATED1 and LATE ELONGATED HYPOCOTYL function synergistically in the circadian clock of *Arabidopsis*. *Plant Physiol.* 150, 834–843
- 93 Yakir, E. *et al.* (2009) Posttranslational regulation of CIRCADIAN CLOCK ASSOCIATED1 in the circadian oscillator of *Arabidopsis*. *Plant Physiol.* 150, 844–857
- 94 O'Neill, J.S. *et al.* (2011) Circadian clock parameter measurement: characterization of clock transcription factors using surface plasmon resonance. *J. Biol. Rhythms* 26, 91–98
- 95 Gan, X. *et al.* (2011) Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* 477, 419–423
- 96 Adams, K.L. and Wendel, J.F. (2005) Allele-specific, bidirectional silencing of an alcohol dehydrogenase gene in different organs of interspecific diploid cotton hybrids. *Genetics* 171, 2139–2142
- 97 Zhou, R. *et al.* (2011) Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16122–16127
- 98 Kalyna, M. *et al.* (2006) Evolutionary conservation and regulation of particular alternative splicing events in plant SR proteins. *Nucleic Acids Res.* 34, 4395–4405
- 99 Iida, K. and Go, M. (2006) Survey of conserved alternative splicing events of mRNAs encoding SR proteins in land plants. *Mol. Biol. Evol.* 23, 1085–1094
- 100 Blencowe, B.J. (2006) Alternative splicing: new insights from global analyses. *Cell* 126, 37–47
- 101 Zhu, W. *et al.* (2003) Refined annotation of the *Arabidopsis* genome by complete expressed sequence tag mapping. *Plant Physiol.* 132, 469–484
- 102 Iida, K. *et al.* (2004) Genome-wide analysis of alternative pre-mRNA splicing in *Arabidopsis thaliana* based on full-length cDNA sequences. *Nucleic Acids Res.* 32, 5096–5103
- 103 Xiao, Y.L. *et al.* (2005) Analysis of the cDNAs of hypothetical genes on *Arabidopsis* chromosome 2 reveals numerous transcript variants. *Plant Physiol.* 139, 1323–1337
- 104 Campbell, M.A. *et al.* (2006) Comprehensive analysis of alternative splicing in rice and comparative analyses with *Arabidopsis*. *BMC Genomics* 7, 327