



# Genome-wide survey of Alternative Splicing in *Sorghum Bicolor*

Bahman Panahi · Bahram Abbaszadeh ·  
Mehdi Taghizadeghan · Esmail Ebrahimie

Received: 16 April 2014 / Accepted: 9 June 2014 / Published online: 29 June 2014  
© Prof. H.S. Srivastava Foundation for Science and Society 2014

**Abstract** *Sorghum bicolor* is a member of grass family which is an attractive model plant for genome study due to interesting genome features like low genome size. In this research, we performed comprehensive investigation of Alternative Splicing and ontology aspects of genes those have undergone these events in *sorghum bicolor*. We used homology based alignments between gene rich transcripts, represented by tentative consensus (TC) transcript sequences, and genomic scaffolds to deduce the structure of genes and identify alternatively spliced transcripts in sorghum. Using homology mapping of assembled expressed sequence tags with genomics data, we identified 2,137 Alternative Splicing events in *S. bicolor*. Our study showed that complex events and intron retention are the main types of Alternative Splicing events in *S. bicolor* and highlights the prevalence of splicing site recognition for definition of introns in this plant. Annotations of the alternatively spliced genes revealed that they represent diverse biological process and molecular functions, suggesting a fundamental role for Alternative Splicing in affecting the development and physiology of *S. bicolor*.

**Keywords** Alternative splicing · Ontology · Complex events · Intron retention

Bahman Panahi and Esmail Ebrahimie contributed equally to this work.

B. Panahi (✉) · M. Taghizadeghan  
Department of Biotechnology and Plant Breeding,  
University of Tabriz, Tabriz, Iran  
e-mail: panahibahman@ymail.com

B. Abbaszadeh  
Meshgin shahr Branch, Islamic Azad University,  
Meshgin Shahr, Iran

E. Ebrahimie  
School of Molecular and Biomedical Science,  
The University of Adelaide, Adelaide, Australia

## Abbreviations

AS Alternative Splicing  
GO Gene Ontology  
EST Expression Sequence Tag  
TC Tentative Consensus

## Introduction

Alternative Splicing is a means that process pre-mRNA into mature isoforms. Alternative RNA splicing permits the production of diverse transcript isoforms and protein products from an individual gene and it is a significant mechanism in transcriptome and proteome plasticity especially in distinct developmental stages and also different environmental conditions. Based on the sites that undergo Alternative Splicing events, this process can affect structural properties, function, and cellular localization (Labadorf et al. 2010). There are several different types of Alternative Splicing (AS) events, which can be classified into four main subgroups. The first type is exon skipping, in which a type of exon known as a cassette exon is spliced out of the transcript together with its flanking introns (Fig. 1a). The second and third types are alternative 3' splice site (3' SS) and 5' SS selection (Fig. 1b, c). These types of AS events occur when two or more splice sites are recognized at one end of an exon. The fourth type is intron retention (Fig. 1d), in which an intron remains in the mature mRNA transcript. This is the rarest AS event in vertebrates and invertebrates, accounting for less than 5 % of known events (Alekseyenko et al. 2007; Kim et al. 2008; Sakabe and de Souza 2007; Sugnet et al. 2004). In contrast, intron retention is the most prevalent type of AS in plants, fungi, and protozoa (Kim et al. 2008; Reddy 2007; Johnson et al. 2003). Less frequent, complex events that give rise to alternative transcript variants include mutually exclusive exons (Fig. 1e), alternative promoter usage (Fig. 1f) and alternative polyadenylation (Fig. 1g) (Ast 2004; Kim et al. 2008;

Black 2003). Another rare form of AS involves reactions between two primary transcripts in trans (Labrador and Corces 2003). The dissimilarities in these splice-variant ratios can be associated to common splice site recognition mechanism, granted that splicing variations are more expected to result in retained introns with the intron definition model (fungi and plants), and in cassette exons with the exon definition model (animals) (Reddy 2007; McGuire et al. 2008) These different proportions of AS types have clear implications for the functional meaning of AS events.

Genome wide investigation of AS is an important approach that can help to disclose the global influence the this process on gene function. Considering this, whole genome sequences and expression sequencing tags get a critical chance to discover knowledge of the distribution and biological impacts of AS events by homology mapping-based approach of ESTs to genome sequences (Gupta et al. 2004).

Computational investigations based on alignment of ESTs/complementary DNAs (cDNAs) against the human genome and experimental analyses using RT-PCR and more recently splicing-sensitive microarrays that can differentiate between AS isoforms have estimated that over 70 % of human genes undergo AS (Modrek and Lee 2002; Johnson et al. 2003; Blencowe 2006).

In plants, such comprehensive investigates are limited mainly because of relatively smaller numbers of ESTs and lack of high-throughput tools such as splicing-sensitive arrays (Blencowe 2006; Reddy 2007). Nevertheless, with available plant ESTs/cDNAs, it is now approximated that at least 30 % of plant genes are alternatively spliced (Wang and Brendel 2006; Chen et al. 2007).

Ner-Gaon et al. (2004) identified 436 alternatively spliced genes in *Arabidopsis* by EST-pair alignment. Three other investigations with a lesser assembled of EST/cDNA data briefly described fewer AS events in *Arabidopsis* (Iida et al. 2004; Zhu et al. 2003; Haas et al. 2003). All these pioneering investigations revealed that reduced parts of genes of 5–10 % are alternatively spliced, with Intron retention the most prevalent AS type in *Arabidopsis* (Iida et al. 2004). However, none of the above studies included comprehensive analyses on the position and conclusion of AS. Two recent studies with full-length cDNAs described alike fractions of alternatively spliced genes but conflicting predominant AS types and places of AS events relative to the coding regions (Nagasaki et al. 2005; Alexandrov et al. 2006).

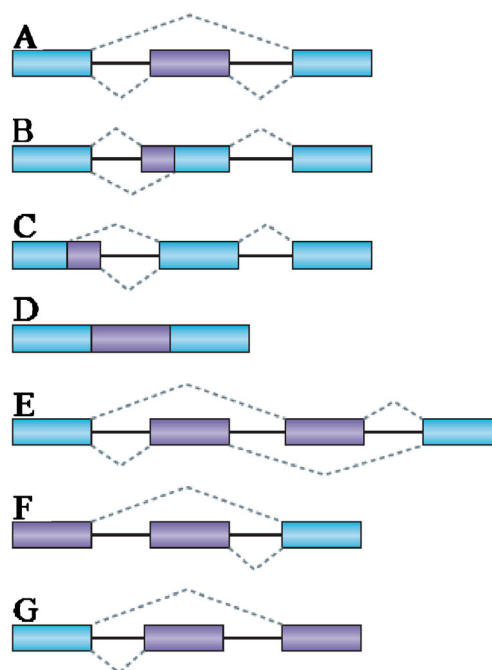
*S. bicolor* is a member of grass family, that for its interesting features of genome such as low genome size (in comparison to crop plants), its low grade of gene duplication and his high number of repetitive elements, is an appealing model plant for functional and structural genomics studding.

In this research, we performed comprehensive investigation of alternative splicing and ontology aspects of genes those have undergone these events in *S. bicolor*. We used homology based alignments between gene rich transcripts, represented by tentative consensus (TC) transcript sequences, and genomic scaffolds to deduce the structure of genes and identify alternatively spliced transcripts in sorghum.

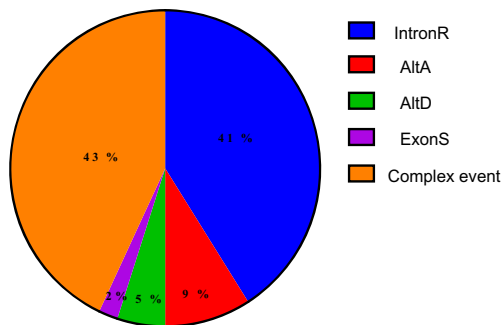
## Materials and methods

### Data sets for AS analysis

First of all, the 209835 EST of *S. bicolor* were searched in NCBI Gen Bank and then those ESTs were analyzed for the contaminating vectors and then homo polymer tails were removed from the EST. After cleaning of the EST, 191300 ESTs were clustered into the tentative 42181 unique genes. Unique genes were mapped onto the Sorghum Chromosome available as scaffold from the Phytozome. Only the 10 chromosome (SBI-01–SBI-10) were used for the further analysis and rest of the short contigs were avoided to prevent spurious mapping.



**Fig. 1** Different types of alternative splicing. **a** Exon skipping. **b** Alternative 3' SS selection. **c** Alternative 5' SS selection. **d** Intron retention. **e** Mutually exclusive exons. **f** Alternative promoters. **g** Alternative poly(A). (Form **e**, **f**, **g** and other trans forms (not shown) categorized as a complex events). In the figure, constitutive exons are shown in blue and alternatively spliced regions in purple. Introns are represented by solid lines and dashed lines indicate splicing options



**Fig. 2** Landscape of AS events in *S. bicolor*: *Intron R* intron retention, *AltD* alternative donor splice site, *AltA* alternative acceptor splice site, *ExonS* exon skipping

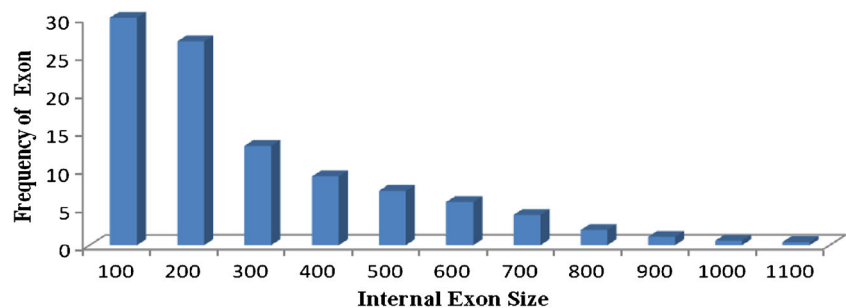
### Homology mapping

Mapping of the transcripts on the chromosome was done using the sim4 algorithm, and then, exon files were parsed for the generation of the GTF file. For genome mapping using tentatively unique gene sequences, we have utilized a threshold of 95 % identity, a smallest alignment of 40 bp, and a smallest criterion that the aligned length of a tentatively unique gene sequence should be 75 % of the total tentatively unique gene sequence length. The GTF file was further investigated using the Astalavista web server (<http://genome.crg.es/astalavista/>), and landscape files were produced for further analysis.

### Functional impact of Alternative Splicing

GO analysis was conducted by using data obtained from Ensemble using the BioMart tool. The text file consisting of Gene Ontology for AS-affected genes was analyzed through the AgriGO tool (<http://bioinfo.cau.edu.cn/agriGO/>) (Du et al. 2010) to obtain the GO plot and the corresponding values for the molecular function, biological process, and cellular component) tool to obtain the GO plot and the corresponding values for the molecular function, biological process, and cellular component.

**Fig. 3** Distribution of internal exon size: Bin sizes are right inclusive (e.g., bin 100 comprises sequences of lengths 1–100 bp)



## Result and discussion

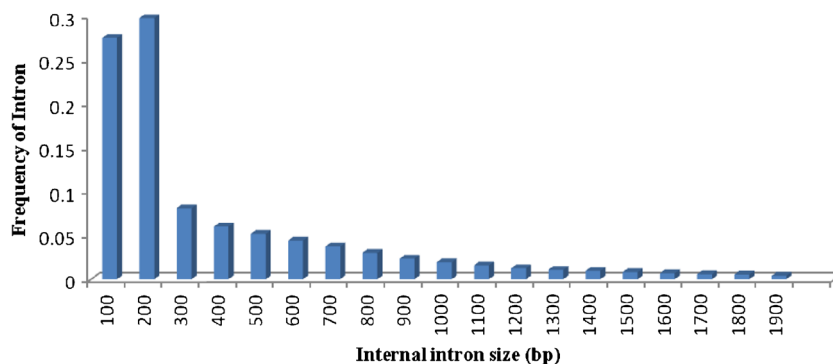
### Detection of Alternative Splicing events

We mapped tentative consensus sequences to ten genomic scaffolds. In our study, for increase of correctness of outcomes and decrease of rate of false positive AS isoforms that possibly identified, we utilized tentative consensus sequences instead of individual ESTs. Using of the putative map, we observed that complex events and intron retention are the common type of AS in sorghum as a previous studies in *Oryza* and *Arabidopsis* (Iida et al. 2004; Ner-Gaon et al 2004; Wang and Brendel 2006; Barbazuk et al. 2008; Filichkin et al. 2010). Although the estimated rates of complex events in sorghum were higher than of other plants.

In general, about 41 % of the events in the assembled transcripts are intron retention. Our result showed that ratio of intron retention events in sorghum is less than of this basic form in *Populus*, *Oryza*, and *Chlamydomonas* (Labadorf et al. 2010; Baek et al. 2008).

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 (Marquez et al. 2012). This estimate is based on analysis of *Arabidopsis* grown under normal growth conditions, and it is likely that this level will increase further as different tissues at various developmental stages and growth conditions are analyzed (Marquez et al. 2012). Of the most common types of AS, intron retention (IR) has been shown to be the most frequent AS event in plants (Filichkin et al. 2010). However, some IR events were recently shown to be more likely to represent partially spliced transcripts due to their low abundance (Marquez et al. 2012). 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 (Marquez et al. 2012). 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 (Marquez et al. 2012).

**Fig. 4** Distribution of internal Intron size: Bin sizes are right inclusive (e.g., bin 100 comprises sequences of lengths 1–100 bp)



Also, the high rate of intron retention (41 %) in sorghum and in other plants supports the intron definition model. Based on this model, introns are recognized by the splicing machinery spliceosomes during pre-mRNA processing in plants. Whereas, the exon definition model seen in animals where exon skipping events occur much more frequently than intron retention events (McGuire et al. 2008). Most intron retention events result in the insertion of an in-frame premature termination codon inside the transcript, which can lead an mRNA to different fates. If this codon is located more than 50 nucleotides upstream of an exon–exon junction, the transcript will be targeted for degradation by nonsense-mediated decay, an mRNA surveillance means that is believed to prevent accumulation of truncated, and potentially harmful, proteins (Lewis et al. 2003).

Alternative donor sites and alternative acceptor sites each accounted for ~5–9 %, and exon skipping was the least prevalent event type, consistent with AS event distributions reported in other plant species (Wang and Brendel 2006; Walters et al. 2013). Also, we find that the use of alternative 3' splice sites is higher than the use of alternative 5' splicing site (Fig. 2) consistent with former finding (Wang and Brendel 2006).

#### Features of exons and introns

Intron and exon lengths reflect the constraints imposed by splicing recognition and frequency of AS. Following the mapping of the ESTs to the genomic sequences, we calculated the lengths of 98,340 internal exons and 61,515 introns with computer scripts. We observed that most of exons had a size below 500 bp (86 %) and 57 % lied between 10 and 200 bp (Fig. 3). The observed exon length is agreed with the observed mean internal exon length in *Arabidopsis thaliana* (172 bp) and *Oryza sativa* (193 bp) (Wang and Brendel 2006). However, compared with exon size, the distributions of intron size in sorghum ranged from 2 to 1,900 bp, and 77 % of introns had a size range of 100 to 500 bp (Fig. 4). In *Drosophila melanogaster*, most of the introns flanking alternatively spliced exons are long, whereas constitutively spliced exons are flanked by short introns. The length of the upstream

intron was found to have a greater influence on exon selection than that of the downstream intron (Fox-Walsh et al. 2005). This shows a large contribution of the exon–intron size in *D. melanogaster* to the frequency of AS. In humans, exons flanked by long introns are subject to exon skipping more often than those flanked by short introns (Fox-Walsh et al. 2005). An additional illustration of the effect of exon–intron architecture on AS is revealed when the size of mammalian exons is examined. Usually, enlarged exons lead to exon skipping, but if the flanking introns are short, the enlarged exon is included (Sterner et al. 1996). The average intron size in *S. bicolor* was close to rice and *Brachypodium* intron size (433 bp) and longer than the average intron size (173 bp) of *Arabidopsis* (Wang and Brendel 2006).

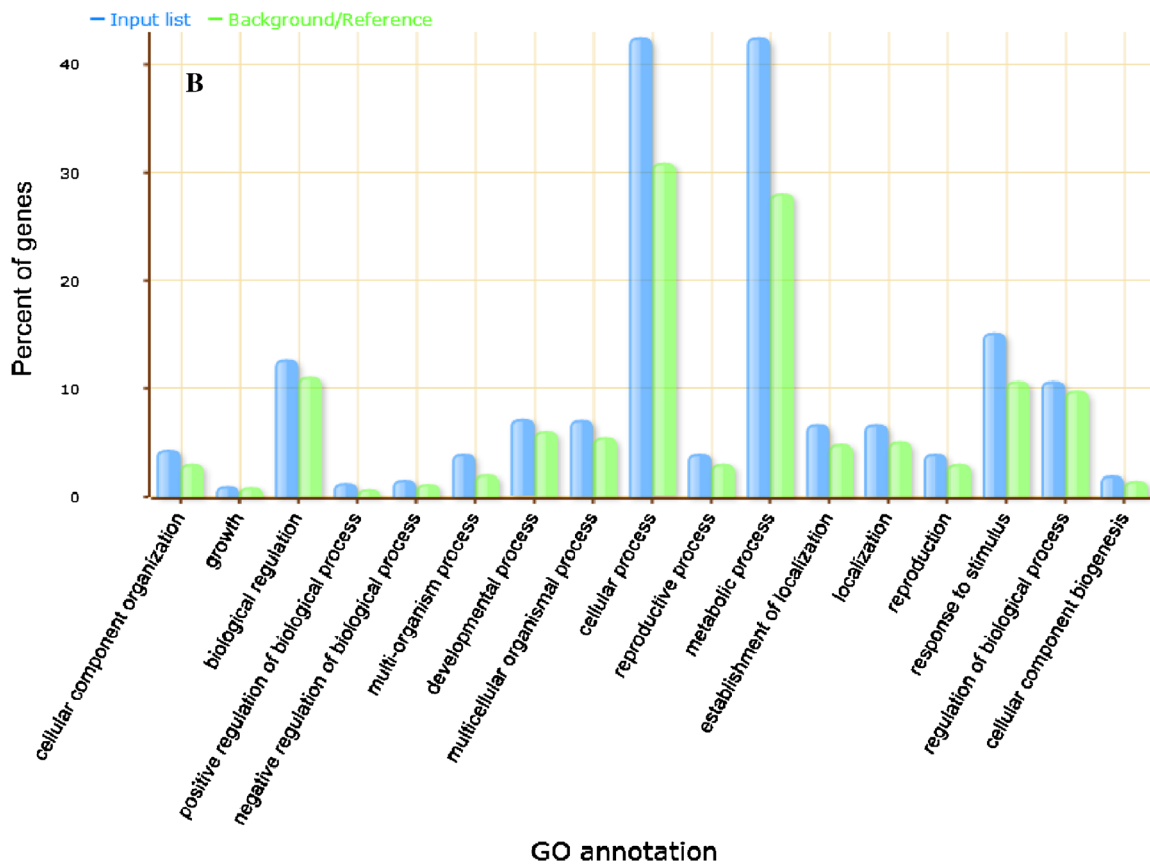
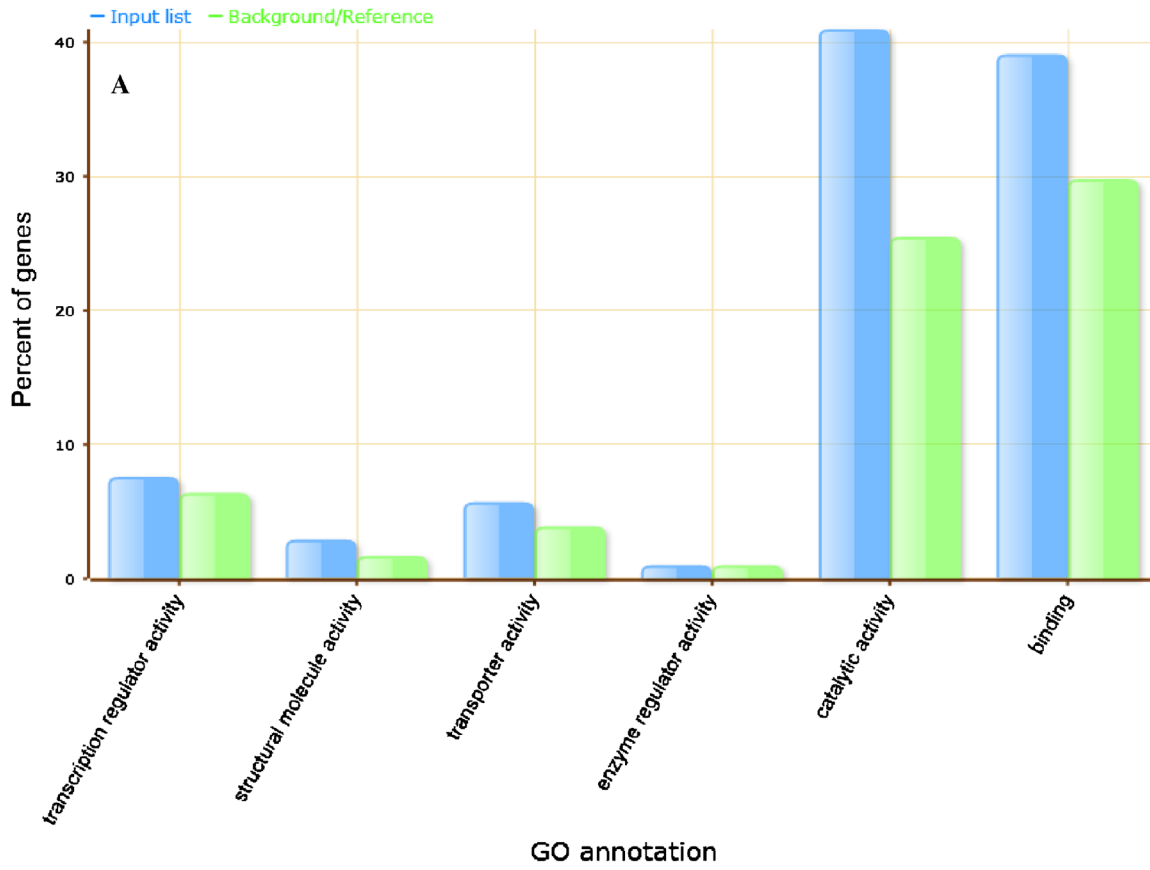
#### Functional ontology of AS genes

The majority of known plant AS events has not been functionally characterized, but several lines of evidence suggest that AS has a biological role. As suggested by Reddy (2007), the majority of intron-containing genes should produce splice variants if most isoforms resulted from random splicing errors. However, AS is predominant in some gene families, while absent in others.

GO classification of molecular functions of the alternative splicing affected genes reveals that about 80 % of genes have catalytic activity and binding properties (Fig. 5a). More dissection of each mentioned groups showed that genes with catalytic activity ontology including of transferase activity (35 %), hydrolase activity (27.6 %), oxidoreductase activity (18.6 %), RNA splicing factor activity (7.8 %) and lyase activity, ligase activity, isomerase and activity genes are in these groups.

Several studies link the occurrence of AS to tissue-specific and/or developmental cues, and alternatively spliced isoforms have been associated with ribosomes. Results of the few functional analyses that have been conducted indicate roles for AS in plant processes such as some metabolic pathways

**Fig. 5** Ontology diversity, **a** and **b**, molecular function and biological process for AS-affected genes, respectively



(Gorlach et al. 1995), catabolic pathways (Kopriva et al. 1995), and mRNA processing (Kalyna et al. 2006), and AS impacts many important plant process such as photosynthesis, defense response, flowering, and cereal grain quality (Reddy 2007).

RNA binding proteins are hub proteins in regulation of mRNA process and also defense responses. results showed that nucleic acid binding proteins is the prevalent proteins that affect by alternative splicing (76 %). other type of binding proteins that affected by alternative splicing including protein binding (31.4 %), chromatin binding (6.2 %), calcium ion binding (6.2 %), and calcium-dependent phospholipids binding (0.7 %).

Cellular component analyses showed about 27 % membrane proteins, 17 % intracellular, and variable distributions in other sub cellular components. These proteins are involved in various biological processes including metabolic, biosynthesis, transport, and responses to stress.

Gene Ontology investigation of the alternatively spliced gene list reported by Wang and Brendel (2006) and Chen et al. (2007) showed that a majority of stress-related genes undergo alternative splicing. Transcription factors are known to affect the expression of genes under specific stress conditions. Interestingly, some transcription factors themselves are subjected to splicing regulation, which in turn may regulate downstream genes and coordinate interactions between major networks at major nodes. Iida et al. (2005) have reported that of the 1,968 total *Arabidopsis* transcription factors, 110 undergo alternative splicing. Computational analyses have identified several exonic and intronic splicing enhancer sequences in *Arabidopsis* (Sheth et al. 2006). It is suggested that these sequences together with SR proteins likely play significant roles in regulating alternative splicing in response to various stresses. This suggestion is supported by the observation that the tissue-specific alternative splicing of ascorbate peroxidase is conferred by an intronic splicing regulatory element (Yoshimura et al. 2002).

Gene Ontology analysis for biological process showed that most of these related to metabolic and cellular process (Fig. 5b). When we dissect metabolic process, we find that about 94 % of these genes belong to primary metabolic process and variable distributions in other process including biosynthetic process, coenzyme metabolic process, ferredoxin metabolic process, and sulfur metabolic process.

The widespread occurrence of AS and the range of functional gene groups which it affects supports an essential role for AS in sorghum development, physiology, metabolism, and responses to environmental conditions and pathogens, all of which have important consequences on sorghum phenotypes.

## Conclusion

Detection of alternative splicing events is one of the first steps essential to link transcriptome to proteome and realize the

connection between their diversity. Our investigations supply a potential survey and will help realize the function of alternative splicing in the sorghum bicolor. Our genome-wide prediction of AS events was based on mapping tentative consensus transcript sequences on genomic scaffolds. Therefore, additional studies, including studies taking tissue or developmental stage specificity into account, will be required to have a more complete view.

## References

- Alekseyenko AV, Kim N, Lee CJ (2007) Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes. *RNA* 13:661–670
- Alexandrov NN, Troukhan ME, Brover VV, Tatarinova T, Flavell RB, Feldmann KA (2006) Features of Arabidopsis genes and genome discovered using full-length cDNAs. *Plant Mol Biol* 60:69–85
- Ast G (2004) How did alternative splicing evolve? *Nat Rev Genet* 5:773–782
- Baek JM, Han P, Iandolino A, Cook DR (2008) Characterization and comparison of intron structure and alternative splicing between *Medicago truncatula*, *Populus trichocarpa*, *Arabidopsis* and rice. *Plant Mol Biol* 67:499–510
- Barbazuk WB, Fu Y, McGinnis KM (2008) Genome-wide analyses of alternative splicing in plants: opportunities and challenges. *Genome Res* 18:1381–1392
- Black DL (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 72:291–336
- Blencowe BJ (2006) Alternative splicing: new insights from global analyses. *Cell* 126:37–47
- Chen FC, Wang SS, Chaw SM, Huang YT, Chuang TJ (2007) Plant gene and alternatively spliced variant annotator. A plant genome annotation pipeline for rice gene and alternatively spliced variant identification with cross-species EST conservation from seven plant species. *Plant Physiol* 143:1086–1095
- Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) AgriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res* 38:W64–W70
- Filichkin SA, Priest HD, Givan SA, Shen R, Bryant DW, Fox SE, Wong WK, Mockler TC (2010) Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res* 20:45–58
- Fox-Walsh K, Yimeng D, Bianca JL, Hung SP, Baldi PF, Hertel KJ (2005) The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc Natl Acad Sci U S A* 102:16176–16181
- Gorlach J, Raesecke HR, Abel G, Wehrli R, Amrhein N, Schmid J (1995) Organ-specific differences in the ratio of alternatively spliced chorismate synthase (LeCS2) transcripts in tomato. *Plant J* 8:451–456
- Gupta S, Zink D, Korn B, Vingron M, Haas SA (2004) Genome wide identification and classification of alternative splicing based on EST data. *Bioinformatics* 20:2579–2585
- Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK (2003) Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* 31:5654–5666
- Iida K, Seki M, Sakurai T, Satou M, Akiyama 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
- Iida K, Seki M, Sakurai T, Satou M, Akiyama K et al (2005) RARTF: Database and tools for complete sets of arabidopsis transcription factors. *DNA Res* 12:247–256

- Johnson JM, Castle J, Garrett-Engel P, Kan Z, Loerch PM (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 302:2141–2144
- Kalyana M, Lopato S, Voronin V, Barta A (2006) Evolutionary conservation and regulation of particular alternative splicing events in plant SR proteins. *Nucleic Acids Res* 34:4395–4405
- Kim E, Goren A, Ast G (2008) Alternative splicing: current perspectives. *Bioessays* 30:38–47
- Kopriva S, Cossu R, Bauwe H (1995) Alternative splicing results in two different transcripts for H-protein of the glycine cleavage system in the C4 species *Flaveria trinervia*. *Plant J* 8:435–441
- Labadorf A, Link A, Rogers MF, Thomas J, Reddy ASN, Ben Hur A (2010) Genome-wide analysis of alternative splicing in *Chlamydomonas reinhardtii*. *BMC Genomics* 11:14–19
- Labrador M, Corces VG (2003) Extensive exon reshuffling over evolutionary time coupled to trans-splicing in *Drosophila*. *Genome Res* 13:2220–2228
- Lewis BP, Green RE, Brenner SE (2003) Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci* 100:189–192
- Marquez Y et al (2012) Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Res* 22:1184–1195
- McGuire AM, Pearson MD, Neafsey DE, Galagan JE (2008) Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biol* 9:50–57
- 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
- Reddy AS (2007) Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu Rev Plant Biol* 58:267–294
- Sakabe NJ, de Souza SJ (2007) Sequence features responsible for intron retention in human. *BMC Genomics* 8:59
- Sheth N, Roca X, Hastings ML, Roeder T, Krainer AR (2006) Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res* 34:3955–3967
- Stern DA, Carlo T, Berget SM (1996) Architectural limits on split genes. *Proc Natl Acad Sci U S A* 93:15081–15085
- Sugnet CW, Kent WJ, Ares M Jr, Haussler D (2004) Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac Symp Biocomput* 9:66–77
- Walters B, Lum G, Sablok G, Min XJ (2013) Genome-wide landscape of alternative splicing events in *Brachypodium distachyon*. *DNA Res* 20:163–171
- Wang B, Brendel V (2006) Genome wide comparative analysis of alternative splicing in plants. *Proc Natl Acad Sci U S A* 103:7175–7180
- 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
- Zhu W, Schlueter SD, Brendel V (2003) Refined annotation of the *Arabidopsis* genome by complete expressed sequence tag mapping. *Plant Physiol* 132:469–484