

Features of evolutionarily conserved alternative splicing events between Brassica and Arabidopsis

Aude Darracq and Keith L. Adams

New Phytologist **199** (2013) 252–263.

Speaker: Joseph Chuang-Chieh Lin

The Comparative & Evolutionary Genomics/Transcriptomics Lab.
Genomics Research Center, Academia Sinica
Taiwan

15 January 2014





Outline

- 1 Introduction
- 2 Materials & methods
- 3 Results
- 4 Discussion



Alternative splicing (AS)

- A cellular process producing multiple types of mRNA isoforms from the same type of pre-mRNA.
 - Differential inclusion/exclusion of intronic/exonic sequences.
- AS plays a role in gene expression regulation (e.g., coupled with the nonsense-mediated decay (NMD) pathway).
- Effect of AS on proteins:
 - Loss/gain of function;
 - altered subcellular localization;
 - protein stability and affinity;
 - enzyme activity;
 - ...



Alternative splicing (AS) (contd.)

- $\geq 50\%$ of genes in *Oryza sativa* and *Arabidopsis thaliana* are affected by AS [Lu *et al. Genome Res.*, 2010; Marquez *et al. Genome Res.*, 2012].
- AS is more prevalent in mammalian genes ($\geq 95\%$) [Pan *et al., Nat. Genet.*, 2008; Nilsen & Graveley *Nature*, 2010].



Evolutionary conservation of AS patterns in plants

- The evolutionary conservation of AS events in plants is largely unexplored.
- Only a small number of AS events have been identified as conserved between divergent species.



Contribution of this paper

- A large-scale analysis of cDNA data from *Brassica* and *Arabidopsis* to identify and characterize conserved AS events.
- 537 conserved AS events in 485 genes are identified.
 - AltD & AltA events are significantly overrepresented;
 - IntronR and ExonS events are underrepresented.
 - Conserved AS events are significantly shorter, less likely to be in the 3'UTR, and enriched for genes whose products function in the chloroplast.



Datasets

- *Arabidopsis thaliana* genes, TAIR (The Arabidopsis Information Resource) gene models and cDNA sequences.
 - TAIR10.
- ESTs
 - dbEST at NCBI (8 Dec. 2010).
 - Extract sequences from *Arabidopsis thaliana* and all Brassica species.
- *Brassica rapa* genome annotation.
 - BRAD [Cheng *et al.*, *BMC Plant Biol.*, 2011]).



Sequence alignments

- *A. thaliana* gene models, cDNAs and ESTs → *Arabidopsis* genes.
 - ★ **Gmap** [Wu & Watanabe, *Bioinformatics*, 2005].
 - Best-hit, identity & coverage $\geq 95\%$.
 - Include only sequences spanning ≥ 1 intron.
- *Brassica* ESTs → *Arabidopsis* genes.
 - ★ **Gmap** and **Sim4 cc** [Zhou *et al.*, *Nucl. Acids Res.*, 2009].
 - Best-hit, identity & coverage $\geq 80\%$.
 - Only ESTs for which the same structure was both found are kept.



Alignment results

	Number of sequences	Mapped to nuclear genes ^a	Mapped to protein coding genes	Mapped to AS genes	AS genes
Same species alignments ^b					
ESTs	1532 829	1075 885	1060 673	491 978	
cDNA	78 096	73 572	72 579	26 032	
TAIR gene models	41 671	41 392	35 176	11 090	
Total	1652 596	1190 849 (72%)	1168 428 (71%)	529 097 (32%)	5686
Cross-species alignments ^c					
ESTs	1026 869	421 932 (41%)	416 993 (41%)	180 060 (18%)	4726

AS, alternative splicing; EST, expressed sequence tag.

^aBest hits with identity and coverage scores ≥ 0.95 for same-species alignments and ≥ 0.80 for cross-species alignments.

^b*Arabidopsis thaliana* sequences aligned on *A. thaliana* genes.

^c*Brassica* sequences aligned on *A. thaliana* genes.



Detection of AS

AS detection:

- Pairwise comparisons of Gmap gene structures for each locus.
- Only TAIR gene models where the AS was supported by ≥ 1 cDNA/EST sequence were used.



Detection of conserved AS

Conserved AS detection:

- Conserved **position** events (**CPs**);
 - 1 The same type of events;
 - 2 The same position in the orthologous genes between *Brassica* and *Arabidopsis*;
 - 3 Affect the **homologous intron** at the same exon-intron junctions.
- Conserved **junction** events (**CJs**).
 - 1 The same type of events;
 - 2 Affect the same exon-intron junctions yet the AS event is not in exactly the same position.
- ★ Events found only in *A. thaliana*: **NEC** (having **no** evidence for conservation).



Analysis of AS events

- Length (not for ME and AltP).
- Locations (5'UTR, CDS, 3'UTR).
- Effect on protein-coding:
 - Only pairs of isoforms showing only one AS event.
 - Consider the longest ORF.
 - PTC (premature termination codon): a stop-codon position in the AS isoform resulting in a **shorter** ORF compared with the constitutive isoform.
 - “Possible-PTC”?



Possible-PTC

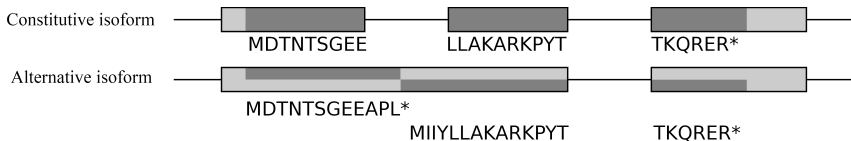


Figure S2 - Example of AS leading to two possible isoform translations.

One translation has the same start as the constitutive isoform, whereas the second leads to a longer ORF compared to the first translation. Boxes represent exons and lines are introns.

Light gray shading: UTR regions, dark gray: coding regions. Asterisks indicate stop codons.



Gene ontology and domain analysis

Gene ontology (GO) enrichment analysis:

- R topGO package (version 2.2.0).
- Compare GO categories for genes having conserved AS with all genes having an AS.

Protein domain analysis:

- Identified with hmmscan from the HMMER3 program suite [Durbin *et al.*, 1998].
 - Using PfamA HMM downloaded from Pfam25.0 [Finn *et al.*, *Nucl. Acids Res.*, 2010].



Distribution of AS categories among CP compared with NEC

Table 2 Alternative splicing (AS) distribution

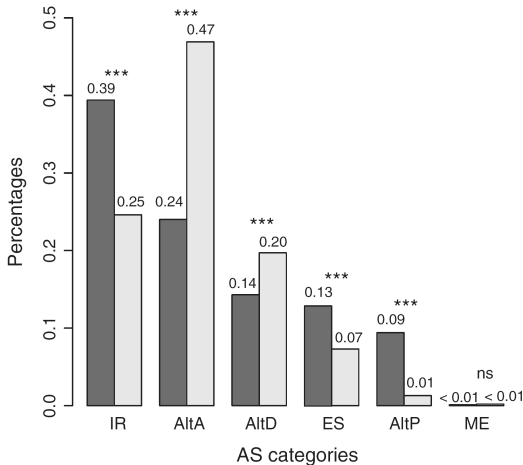
	IR	AltA	AltD	ES	AltP	ME	Total	Number of genes
Conserved position AS (CP)								
Frequencies	132	252	106	39	7	1	537	485
Percentages	24.6	46.9	19.7	7.3	1.3	0.2		
No evidence for conservation (NEC) AS								
Frequencies	3887	2375	1410	1273	925	8	9878	5426
Percentages	39.4	24.0	14.3	12.9	9.4	0.1		
G-test (on CP vs NEC)								
P-value	***	***	***	***	***	ns		

IR, intron retention/exclusion; AltA, alternative acceptor; AltD, alternative donor; ES, exon skipping/retention; AltP, alternative positions; ME, mutually exclusive exons. ns, nonsignificant; ***, $P < 0.001$.

- 220 CP AS events were not referenced in TAIR gene models; 139 *Arabidopsis* genes were not described as AS in TAIR.



Distribution of AS categories among CP compared with NEC



Comparison of length of AS between CP and NEC

Table 3 Length of alternative splicing (AS) (bp)

	IR	AltA	AltD	ES
Conserved position AS (CP)				
Mean	107.90	15.21	16.87	119.7
Median	90.50	6.00	5.00	70.00
No evidence for conservation (NEC) AS				
Mean	156.2	39.06	72.1	133.3
Median	99.00	11.00	17.00	96.00
Wilcoxon test				
P-value	***	***	***	*

IR, intron retention/exclusion; AltA, alternative acceptor; AltD, alternative donor; ES, exon skipping/retention. *, $P < 0.05$; ***, $P < 0.001$.



Comparison of AS positions between CP and NEC

Table 4 Alternative splicing (AS) positions

	5'UTR	Coding	3'UTR	Number in coding region ^a	Number in UTR	Total number
Conserved position AS (CP)						
No. of AS events	45	255	3	255	48	303
%	14.9	84.2	1.0	84.4	14.7	
No evidence for conservation (NEC) AS						
No. of AS events	755	3165	189	3165	944	4118
%	18.3	77.1	4.6	77.1	22.9	
G-test/Fisher's exact test (CP vs NEC AS)						
<i>P</i> -value	ns/ns	**/**	***/**	***/**		

^aThe AS position is assigned as 'in coding region' if the AS is in the coding region of one isoform.

ns, nonsignificant; **, $P < 0.01$; ***, $P < 0.001$. Only AS events contained in TAIR gene models and cDNAs for isoforms containing only one AS event were taken into account.



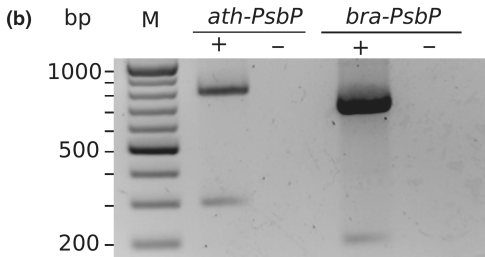
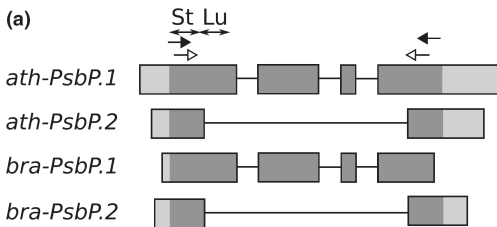
Types of genes with conserved AS

Enrichment in genes for **chloroplast proteins**.

- GO analysis reveals: it is possible that
 - AS plays an important role in the function of chloroplast proteins;
 - AS is a major part of regulation of nuclear genes for chloroplast.



A photosynthesis gene having conserved AS



Domain analysis of genes with conserved AS

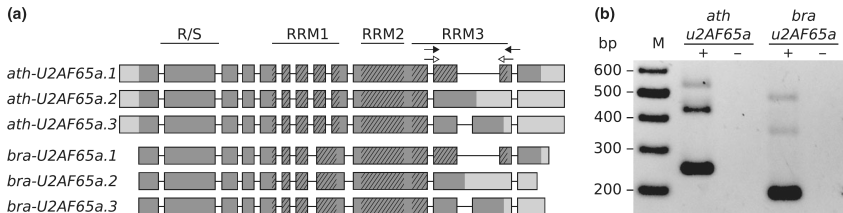
Table 5 Examples of conserved position (CP) alternative splicing (AS) modifying a known domain

Locus	Gene name	AS	AS length (bp)	Domain predicted	Modification predicted
AT1G24260	AGL9	AltD	3	K-box region SRF-type transcription factor (DNA-binding and dimerization domain)	Modified Same
AT2G44750	TPK2	AltA	6	Thiamin pyrophosphokinase, vitamin B1 binding domain	Modified
AT2G45770	CPFTSY	AltA	21	Thiamin pyrophosphokinase, catalytic domain SRP54-type protein, GTPase domain	Same Modified
AT3G57630	Unknown	AltA	6	SRP54-type protein, helical bundle domain Exostosin family	Same Same
AT4G21580	Unknown	AltA	18	EGF-like domain (2 occurrences)	Modified and same
AT4G31550	WRKY11	AltA	3	Zinc-binding dehydrogenase Alcohol dehydrogenase GroES-like domain	Modified Same
AT5G23260	ABS	AltA	15	Plant zinc cluster domain WRKY DNA -binding domain	Modified Same
AT5G60980	Unknown	AltA	3	SRF-type transcription factor (DNA-binding and dimerization domain)	Same
				K-box region RNA recognition motif. (aka RRM, RBD, or RNP domain) Nuclear transport factor 2 (NTF2) domain	Modified Modified Same

- 141 cases where AS modified \geq domains in one isoform compared with the other.



An example of a conserved AS event affecting protein domain



- U2AF⁶⁵a: playing a key role in the assembly of splicing complex.
 - Three RNA recognition motif domains & one N-terminal RS domain.
- On *Arabidopsis* and *Brassica*, both AS events cause PTCs that would lead to the loss of one RRM domain.
- In regulating the spliceosome assembly?
 - Though their exact functions remain uncharacterized.



Features of conserved AS events

- The largest number of conserved AltA events: **3 bp**.
 - Unlikely disrupting the reading frame.
 - Shorter AS length → smaller effects on the coding region.
- Conserved AS events are enriched for genes whose products function in the chloroplast.
 - AS may be important for the regulation of nuclear genes for chloroplast proteins;
 - AS may be involved in producing alternative isoforms of many chloroplast proteins.



Evolutionary conservation of AS in angiosperms

- One possibility to explain previous results on the legume species that the conserved AS events are not so many.
 - The number of ESTs available from the legume species was relatively small.
- Previous results and this paper almost certainly underestimate the number of conserved AS events between plant species.
 - \therefore the ESTs are not comprehensive w.r.t. AS.



Potential functional and regulation importance of conserved AS events

- Not all conserved events are likely to be functional (e.g., secondary structure?).
- Events that are not evolutionary conserved are not necessarily nonfunctional (species-specific function?).



Potential functional and regulation importance of conserved AS events

- ★ AS can regulate mRNA levels by creating transcripts targeted to the NMD pathway for degradation.
- ★ Transcripts that contain PTCs, especially those having a stop codon:
 - > 50 bp upstream of the last exon–exon junction or
 - > 350 bp upstream of the transcript end,

⇒ likely targets for degradation by the NMD machinery.
- ♣ Yet, a recent study [Kalyna *et al.*, *Nucl. Acids Res.*, 2012] revealed that many predicted NMD targets might NOT actually be degraded.



Potential functional and regulation importance of conserved AS events

microProteins (miPs):

- Dominant-negative regulatory proteins.
- Disturb the formation of functional proteins dimers by forming nonfunctional, homotypic protein complexes with their targets.
- ★ MiPs could result from an AS event that deletes the catalytic, activation or other functional domain, but **retains the protein binding domain**.
 - Identified in this work: Jas1 (or Jaz10; AT5G13220).
 - It is conserved among several plants.
 - ▷ **Hypothesis:** AS transcripts containing PTCs that are not degraded by NMD might be translated into miPs.



Future work

- Genome-wide studies of NMD.
- Examine AS event conservation between species at different phylogenetic depths.
- Using RNA-seq approach.
- Tissue-specific conserved AS.



Thank you.

