Cytokinin treatment of rice induces minimal effect on differential splicing compared to splicing between tissues

Nowlan Freese[1*], April R. Estrada[1*], Ivory C. Blakley[1], Jinjie Duan[2], Ann Loraine[1**]
[1]Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC, USA
[2]Bioinformatics Research Center, Aarhus University, C.F. Moellers Alle 8, Aarhus C, Denmark.

[*]Equal contribution.
[**]Corresponding author.

ABSTRACT

Alternatively spliced genes produce multiple spliced isoforms, called transcript variants. In differential alternative splicing, the relative abundance of transcript variants differs across sample types. Differential alternative splicing is common in animal systems, but its extent and significance is not as well known in plants. We examined differential alternative splicing in rice seedlings using RNA-Seq data that included approximately 40 million sequence alignments per library, three libraries per sample type, and four sample types: roots and shoots plus and minus treatment with exogenous cytokinin, a versatile plant hormone that controls a myriad of developmental and stress-related processes in plants. Comparing read alignments to gene model annotations found that for 77% of alternative splicing events proposed in the gene models, each splicing choice was supported by at least one RNA-Seq read alignment. Most genes annotated as alternatively spliced favored one dominant isoform. Of splicing choices where there was abundant support for minor forms, most alternative splicing events affected protein-coding sequence. Statistical testing of read count proportions identified 90 genes as differentially spliced between rice roots and shoots. By contrast, only four genes were detected as differentially spliced in response to cytokinin. Ten differential splicing events were selected for validation via capillary gel electrophoresis analysis of reverse transcriptase-PCR products, using newly prepared samples from an independent experiment. In nine of ten cases, differential splicing between tissue types was confirmed. A tool for visualizing protein annotations in the context of genomic sequence (ProtAnnot) together with a genome browser (Integrated Genome Browser) were used to visualize and assess effects of differential splicing on gene function. In general, differentially spliced regions coincided with conserved regions in the encoded proteins, indicating that differential alternative splicing is likely to affect protein function between root and shoot tissue in rice.

INTRODUCTION

Differential splicing of pre-mRNA transcripts, called alternative splicing, enables one gene to produce multiple transcript variants encoding different functions. Alternative splicing is an almost universal phenomenon in higher eukaryotes, occurring to varying degrees in every animal and plant genome examined to date (Kalsotra and Cooper, 2011; Reddy et al., 2013). In animals, differential expression of splice variants has been recruited as a regulatory mechanism in multiple processes, such as sex determination in invertebrates and neuronal differentiation in mammals (Kalsotra and Cooper, 2011; Salz, 2011; Barbosa-Morais et al., 2012).

In plants, less is known about the functional significance and patterns of alternative splicing. However, several trends are apparent. Genes involved in circadian regulation are highly alternatively spliced, often producing multiple splice variants that fluctuate in concert with day/night cycling along with overall transcript abundance (Filichkin et al., 2015). The SR family of RNA-binding, splicing regulatory proteins is greatly expanded compared to mammals and includes many plant-specific forms (Kalyna and Barta, 2004; Barbosa-Morais et al., 2006; Plass et al., 2008; Filichkin et al., 2015). SR transcripts themselves are also highly alternatively spliced in plants, with the relative abundance of these alternative transcripts varying according to environmental stresses and hormones (Palusa et al., 2007; Gulledge et al., 2012; Filichkin et al., 2015; Keller et al., 2016; Mei et al., 2017).

There is a growing body of evidence indicating cell and tissue specific regulation of alternative splicing in plants (Vitulo et al., 2014; Li et al., 2016; Sun et al., 2018). We previously found through analysis of expressed sequence tags (ESTs) that alternative splicing in the model plant *Arabidopsis thaliana* was highly skewed toward expression of a single dominant isoform (English et al., 2010). Analysis of RNA-Seq data from *Arabidopsis* pollen found the relative abundance of splice variants to be similar between leaves and pollen, despite the differences between the two tissues (Loraine et al., 2013). However, this latter analysis was limited by having just one biological replicate for pollen and only two biological replicates for leaves. A more comprehensive analysis of multiple *Arabidopsis* data sets found a high incidence of isoform switching, in which the identity of the most prevalent variant differs between sample types (Vaneechoutte et al., 2017). However, this splicing diversity may have arisen in part from the heterogeneity of the data sets used, which were produced using different technologies at different times.

In this study, we used a well-replicated RNA-Seq data set from rice to re-examine prevalence of alternative splicing between tissues and sample types. This data set was previously generated to investigate cytokinin regulation of gene expression in roots and shoots from 10-day old rice seedlings (Raines et al., 2016). The data set included three biological replicates per sample type and four sample types – roots and shoots treated with exogenous cytokinin or a mock, vehicle-only treatment. The previous study found that cytokinin treatment triggered differential expression of around 5,000 and 2,300 genes in roots and shoots, respectively. Previous work in *Arabidopsis* found that SR transcripts exhibit differential splicing when exposed to cytokinin, however, less is known about the effect of cytokinin on alternative splicing in rice (Palusa et al., 2007).

We found that the relative abundance of splice variants for most alternatively spliced genes was stable, with very few differentially spliced genes between cytokinin treated and control samples. There was nonetheless a small but significant number of genes that were differentially spliced between roots and shoots, with the majority of the differential splicing occurring within the protein-coding sequence. These results provide new evidence that differential alternative splicing likely contributes to gene function diversification between roots and shoots while playing little role in cytokinin response in rice.

MATERIALS & METHODS

RNA-Seq library preparation and sequencing

Samples were prepared and sequenced as described in (Raines et al., 2016). Rice seedlings (Nipponbare) were grown hydroponically for ten days in a growth chamber set to 14 hours light (28°C) and 8 hours of dark (23°C) with light intensity 700 mmol m^{-2} s^{-1}. Around six to ten seedlings were grown in the same pot, in four pots. On the tenth day, culture media was replaced with new media containing 5 mM of the cytokinin benzyladenine (BA) or 0.05 mN NaOH as a control. After 120 minutes, roots and shoots were harvested separately. Roots and shoots from treatment or control pots were pooled to form three replicates per treatment. RNA was extracted and used to synthesize twelve libraries from BA-treated and mock-treated roots and shoots. Libraries were sequenced on an Illumina HiSeq instrument for 100 cycles, yielding 100 base, single end reads. Sequence data are available from the Sequence Read Archive under accession SRP04905. Aligned, processed data are available from the Oct. 2011 rice genome assembly IgbQuickload directories at http://www.igbquickload.org.

Data processing

Read sequences were aligned onto the *O. sativa japonica* genome assembly Os-Nipponbare-Reference-IRGSP-1.0 (released October 2011) using TopHat version and BowTie2 with default parameter settings except that the maximum intron size was set to 5,000 bases (Kawahara et al., 2013).

A command-line, Java program called "FindJunctions" was used to identify exon-exon junctions from gapped read alignments in the RNA-Seq data. FindJunctions produces BED format files containing junction features, and the score field of the BED file lists the number of read alignments that supported

the junction. Only reads that aligned to a unique location in the genome were considered. Source code and compiled versions of FindJunctions are available from https://bitbucket.org/lorainelab/findjunctions.

Identification of alternative splicing events and differential splicing

To date, there have been two major releases of *O. sativa japonica* gene models: the MSU7 gene set (Kawahara et al., 2013) and the RAP-Db gene set (Sakai et al., 2013). The two gene model sets contain mostly the same data, but the MSU7 gene models appear to be the most heavily used and are well-supported by informatics resources such as Gramene and Gene Ontology Consortium. For simplicity, and to take advantage of available functional annotations, we used the MSU7 annotations here.

Annotated alternative splicing events and the number of reads supporting each alternative were identified using the exon-intron overlap method described in (English et al., 2010). Exons and introns from pairs of gene models from the same locus were compared to identify alternatively spliced regions. Regions where an intron in one model overlapped an exon in another model on the same strand were identified and used to define mutually exclusive splicing choices (Figure 1). Gene models that included an alternatively spliced region were named "L" (for "Long"). Likewise, models that lacked an alternatively spliced region were designated "S" (for "Short"). Alternatively spliced regions were classified according to the type of alternative splicing, as follows. Regions flanked by alternative donor sites were designated "DS" for alternative donor site. Regions flanked by alternative acceptor sites were labeled "AS" for alternative acceptor site. AS and DS events that coincided with exon skipping were labeled "AS/ES" and "DS/ES". Alternatively spliced regions arising from introns that the spliceosome sometimes failed to excise were designated "RI" for retained intron.

For each alternatively spliced region, read alignments that unambiguously supported one or the other splicing choice were counted. For AS and DS events, only gapped reads that aligned across intron junctions were counted as support. For RI events, gapped reads that aligned across the retained intron were counted as support for the intron-removed (S) form, and un-gapped reads that overlapped at least 20 bases within the intron were counted as support for the intron-retained (L) form.

For each alternatively spliced region in each biological replicate, the number of reads supporting L or S, but not both, were used to calculate the percent-spliced-in (PSI) as N/M*100, where N was the number of reads supporting the L form and M was the number of reads that supported S or L but not both. This is the same as the splicing index described in (Katz et al., 2010). A two-sided t-test with equal variance assumption was used to compare PSI between sample types. Because PSI variance was large for events with small M (very few informative reads), only alternatively spliced regions where M was 10 or more in at least three replicate libraries were tested. A false discovery rate (FDR) was calculated for each test using the method of Benjamini and Hochberg, as implemented in the R programming language "p.adjust" method. Alternative splicing events with FDR less than or equal to 0.1 were considered differentially alternatively spliced.

Software used to identify and quantify alternative events is available from https://bitbucket.org/lorainelab/altspliceanalysis. Data analysis code used to analyze RNA-Seq data is available from https://bitbucket.org/lorainelab/ricealtsplice. Data analysis code is implemented as R Markdown files designed to be run in the RStudio development environment. Readers interested in experimenting with different analysis parameters can clone the repository, modify the code, and re-run analyses as desired.

RT-PCR and capillary gel electrophoresis analysis of alternative splicing

Differential alternative splicing detected by analysis of RNA-Seq was re-tested using RT-PCR-based fragment analysis method described in (Stamm et al., 2012). Differentially spliced regions were PCR-amplified and quantified using capillary gel electrophoresis. One benefit of the method is that the results are expressed as relative abundances of splice variants within a sample, thus eliminating the need to normalize using reference genes as in traditional qRT-PCR experiments aimed at measuring overall gene expression.

For splicing validation, new rice seedlings equivalent to the non-BA treated (control) samples used to generate RNA-Seq data were grown and harvested. Seedlings were grown hydroponically in pots containing either liquid media only or calcined clay granules (recommended in (Eddy et al., 2012)) watered with liquid media. After twelve days of cultivation, plants were removed from the pots and roots and shoots were collected separately. Roots and shoots from the same pot were combined to form paired biological replicates. Samples were frozen on liquid nitrogen and stored at -80 degrees C prior to RNA extraction.

RNA was extracted using the RNeasy Plant Mini Kit from Qiagen following the manufacturer's instructions. First strand cDNA was synthesized using oligo dT primers and 1 µg of total RNA per 20 µL reaction. PCR amplification of cDNA was performed using primers flanking differentially spliced regions, including one primer labeled with 6-carboxyfluorescein (6-FAM) to enable amplicon detection via microcapillary electrophoresis. Cycle parameters included denaturation at 94°C for 2 minutes, followed by 24 cycles of 94°C for 15 sec, 50°C for 30 sec and 70°C for 1 min, with a final elongation step of 72°C for 10 minutes. This was essentially the same regime described in (Stamm et al., 2012) but with fewer cycles to ensure reactions were stopped before exiting the logarithmic phase. PCR products were combined with size standards and separated on a 3730 Genetic Analyzer (Life Technologies). Amplicons were quantified using manufacturer-provided software by calculating the area under each amplicon peak. The percentage of the variant containing the alternatively spliced region (%L, also called PSI for "percent-spliced-in") was calculated by dividing the long form area by the total area for both long and short forms. Spreadsheets with data exported from the instrument, along with PSI calculations, are available in the project git repository in a subfolder named "Experimental Testing."

RESULTS

Most genes annotated as alternatively spliced favored one dominant isoform

Using the exon-intron overlap method described previously (English et al., 2010), alternative splicing events within each gene were identified and annotated as shown in Figure 1. Following annotation of alternative splicing events, RNA-Seq read alignments from the libraries described in (Raines et al., 2016) were used to assess alternative splicing in the four sample types: roots and shoots from seedlings treated with the cytokinin compound benzyladenine (BA) or with a mock, control treatment. For each alternative splicing event, the number of sequence alignments unambiguously supporting each alternative was counted. These counts were used to calculate percent-spliced-in (PSI), the percentage of read alignments supporting the longer (L) isoform.

In the combined data from all libraries, 77% of AS events had at least one read supporting each of the two splicing choices, and 19.8% had support for just one splicing choice. Only 2.8% of AS events has no reads supporting either form; these corresponded to genes with low or no expression in any of the sample types tested.

Among the events with support for both choices, the distribution of read support skewed toward supporting a single choice (Figure 2). In 54% of annotated alternative splicing events, only 10% or fewer of informative read alignments supported the minor choice. In 70% of annotated alternative splicing events, only 20% or fewer alignments supported the lower frequency choice. Conversely, 30% of alternative splicing choices showed simultaneous expression of multiple choices in non-trivial amounts, e.g., at least 2 in 10 alignments supported the minor form.

Genes with abundant support for both alternative splicing choices exhibited a wide range of functions

We used standard methods for Gene Ontology term enrichment to determine if specific categories of genes were enriched with genes in which alternative isoform expression was high. Specifically, we asked if there some Gene Ontology terms with significantly enriched with genes containing alternative splicing events in which the minor form frequency was at least 20%, i.e., genes occupying the center trough in Figure 2. Interestingly, we found that these genes exhibited a diversity of gene functions, but no

203 significant enrichment of functional categories were identified (Young et al., 2010). Thus, alternative
204 splicing in which minor forms are highly prevalent affects genes with many functions in rice.
205
206 **Many rice genes are differentially spliced between roots and shoots but cytokinin hormone**
207 **application has minimal effect on splicing**
208 In animals, differential splicing between cell or tissue types contributes to cellular differentiation,
209 especially in the nervous system (Naftelberg et al., 2015). Less is known about the role of alternative
210 splicing in regulating cellular differentiation and other processes in plants. Rice shoots and roots are
211 profoundly different tissues, but our previous analysis of this same data set found that many of the same
212 genes were expressed in both (Raines et al., 2016). This raises the question of how these two different
213 tissues are able to carry out their specialized roles, and suggest the hypothesis that differential splicing
214 could enable differential functions in genes expressed in both tissues (Reddy et al., 2013). Our previous
215 study of cytokinin on rice roots and shoots identified significant differences in gene expression in
216 response to cytokinin exposure (Raines et al., 2016). However, little is known about the role alternative
217 splicing during cytokinin response, except for one study in Arabidopsis that reported a shift in splicing of
218 SR protein genes following cytokinin hormone treatment (Palusa et al., 2007). Therefore, we examined
219 differential splicing in the rice RNA-Seq data set comparing root and shoot tissue with or without
220 cytokinin.
221 First, we asked: When an alternatively spliced gene was expressed in two different sample types, was
222 the relative abundance of splice variants the same or different? To address this, we examined correlation
223 of PSI between roots and shoots or between BA-treated versus mock-treated samples (Figure 3). We
224 found that PSI was similar between treated and untreated samples, as revealed by the tighter clustering of
225 scatter plot points in Figures 3A and 3B. This indicated that genes that were alternatively spliced in BA-
226 treated samples were also alternatively spliced in the controls, and that the relative abundance of splice
227 variants was similar. Thus, the cytokinin hormone treatment had minimal effect on splicing. By contrast,
228 there were many genes where the relative abundance of splice variants was different between roots and
229 shoots (Figure 3C). Consistent with Figure 3, statistical testing of PSI differences between sample types
230 identified 90 genes where PSI was significantly different between roots and shoots (FDR ≤ 0.1) but only
231 four and two genes where PSI was different between cytokinin-treated samples and controls in roots and
232 shoots, respectively (See Supplemental Table). Thus, we observed limited but non-trivial levels of
233 differential alternative splicing between roots and shoots but minimal differential alternative splicing
234 between control and BA-treated samples.
235
236 **Alternative splicing remodeled protein-coding sequence more often than disrupting it**
237 Alternative splicing can occur within the UTR or protein-coding regions of genes. Interestingly, 67%
238 of the differential splicing between roots and shoots occurred within protein-coding regions (Table I and
239 Supplemental Table I), suggesting that differential splicing is likely to affect gene function at the level of
240 the protein product. In nearly every instance, major and minor isoforms were both detected, with
241 differential splicing observed as a change in the relative abundance of the two forms.
242 In general, when alternatively spliced regions overlap the coding region of genes and the lengths of
243 these regions are not multiples of three, then inclusion of these differential regions in transcripts is likely
244 to introduce a frame shift, resulting in a premature stop codon and a truncated protein product. As shown
245 in Table I, there was an enrichment of alternatively spliced regions in rice that were evenly divisible by
246 three in coding regions versus non-coding in all subsets of the data. These subsets included all annotated
247 alternatively spliced regions, regions where the minor form was unusually prevalent (the trough region of
248 Figure 2), and differentially spliced regions. Thus, alternative splicing within the coding regions of genes
249 was biased against introducing frame shifts and promoted protein remodeling rather than truncation.
250 To further understand the effects of splicing on protein-coding sequences, we visualized differentially
251 spliced regions together with RNA-Seq alignments, coverage graphs, and inferred junctions using
252 genome browsers. Two genome browsers were used to visualize the data - Integrated Genome Browser
253 (Freese et al., 2016) and ProtAnnot (Mall et al., 2016). Integrated Genome Browser (IGB) was used to

254 examine RNA-Seq read alignments and compare alignments to the annotated gene structures. ProtAnnot,
255 an IGB App, was used to search the InterPro database of conserved protein motifs to find out how (or if)
256 splicing inferred from RNA-Seq data was likely to affect gene function through remodeling of protein
257 motifs as detected by the InterProScan Web service (Finn et al., 2017).
258 Of the 105 differentially spliced regions, 71 overlapped protein-coding sequence regions, suggesting
259 that in these cases, alternative splicing affected protein function. All but one (70/71) of the differentially
260 spliced regions embedded in coding regions overlapped a predicted functional motif (e.g., a predicted
261 transmembrane helix) or a region found by protein classification systems (e.g., Pfam (Finn et al., 2016) or
262 PANTHER (Thomas et al., 2003)) to be conserved among members of the same protein family
263 (Supplemental Table I and Figure 4).
264
265 **RT-PCR with capillary gel electrophoresis confirmed differential splicing for nine of ten genes**
266 **tested**
267 We used a method based on capillary gel electrophoresis of fluorescently tagged PCR products to
268 assay alternative splicing of ten genes detected as differentially spliced between rice roots and shoots
269 (Stamm et al., 2012). New rice seedlings were grown under a close-to-identical replication of the RNA-
270 Seq experiment. Primers were designed to amplify differentially spliced regions, including one primer
271 that was conjugated to a fluorescent tag. Following PCR amplification of cDNA prepared from the new
272 rice samples, products were resolved on a capillary-based sequencer and PSI calculated (Table II). In nine
273 out of ten genes, differential alternative splicing was confirmed. In the one case where differential
274 alternative splicing was not confirmed, there were very few RNA-Seq read alignments covering the
275 differentially spliced region, suggesting this was likely a false positive result. The FDR cutoff used to
276 detect differential splicing in the RNA-Seq data was 0.1, corresponding to 1 in 10 false discoveries, in
277 line with results from the microcapillary-based analysis.
278
279 **DISCUSSION**
280 In this study we profiled the prevalence of alternative splicing using a well-replicated RNA-Seq data
281 from 10-day old rice roots and shoots, treated or untreated with exogenous benzyl adenine, a synthetic
282 cytokinin hormone. We found there was at least one RNA-Seq read supporting each annotated alternative
283 splicing choice for most of the annotated alternative splicing events, with a distribution skewed toward
284 supporting a single splicing choice. We identified very few differences in splicing between cytokinin
285 treated samples and mock, whereas we observed limited but significant differential splicing between roots
286 and shoots, with the majority of those differences falling within the protein coding region of transcripts.
287 The observation of very little differential splicing between cytokinin-treated and mock-treated
288 samples was surprising due to previous evidence in *Arabidopsis* (Palusa et al., 2007). Cytokinin exposure
289 of *Arabidopsis* seedlings by Palusa et al. identified several SR genes that were differentially spliced when
290 exposed to cytokinin (Palusa et al., 2007). SR proteins regulate splicing choices in other species as well as
291 being alternatively spliced themselves (Filichkin et al., 2015). Therefore, we had expected to observe
292 cytokinin-induced differential splicing of SR genes in rice that would also affect other genes. However,
293 we detected no rice SR protein differential splicing of SR proteins between cytokinin- and mock-treated
294 rice samples and only four differentially spliced genes between roots and two in shoots.
295 The relative lack of differential splicing between cytokinin-treated and mock-treated samples suggests
296 that cytokinin signaling does not employ alternative splicing as a regulatory mechanism in rice. Cytokinin
297 signaling involves transfer of phosphate groups between successive elements of a phosphorelay signaling
298 pathway culminating in phosophorylation-dependent activation of Myb-type transcription factor proteins
299 called type B ARRs (Raines et al., 2016). Thus, regulation of type B ARR transcription factor activity
300 occurs at the level of protein, and cytokinin treatment has no or little effect on transcription of type B
301 ARRs. In addition, most type B transcriptional regulators and other members of the cytokinin signaling
302 pathway are not highly alternatively spliced. By contrast, a closely related family of similar genes
303 encoding so-called "pseudo-response regulators" have similar sequence to type B ARRs and are highly
304 alternatively spliced. Further study is necessary to discern why cytokinin exposure in rice has so few

differences in splicing, whether this is true for all monocots, and how the response to cytokinin compares to dicots such as *Arabidopsis* and if that indicates differences in cytokinin response in general.

In comparison to the number of differentially spliced genes due to cytokinin exposure, there was a comparatively large number of differentially spliced genes between shoots and roots. There is a growing body of evidence that alternative splicing is cell, tissue, and stage specific (Vitulo et al., 2014; Gupta et al., 2015; Li et al., 2016; Sun et al., 2018). Although the alternatively spliced genes with high minor form frequency were not enriched for particular Gene Ontology terms, this lack of concentrated function suggests that alternative splicing has been recruited as a mechanism for diversification of gene function. It is of note that the majority of differential splicing between roots and shoots fell within protein-coding sequence, and of those events, nearly all splicing events affected a predicted motif.

In our previous comparison of rice root and shoot tissue, we found that the overall diversity of gene expression in roots and shoots was similar, but with the most highly expressed genes reflecting tissue-specific differences (Raines et al., 2016). Sun et al. found that clustering of cucumber tissues by alternative splicing profiles placed leaf/stem far from root tissue (Sun et al., 2018). It is likely that alternative splicing plays a role in further delineating differences in expression between these tissues.

CONCLUSION

By analyzing the number of reads that supported different splice variants, we identified examples of differential splicing with confirmation by RT-PCR with capillary gel electrophoresis. There were 90 genes differentially spliced between root and shoot tissues, but only four between cytokinin-treated and non-treated samples. For most differential splicing events, the protein-coding regions were affected, strongly suggesting that differential splicing is playing a role in modulating gene function between roots and shoots.

FUNDING AND SUPPORT

This material is based upon work supported by the National Science Foundation through the Postdoctoral Research Fellowship in Biology under Grant No.1523814 awarded to Nowlan Freese and the Plant Genome Research Program award number IOS-1238051. NIH award 2R01GM103463-06A1 to AL supports development of the IGB software.

334 **Table I**. Alternative splicing choices producing difference regions evenly divisible by three or with
335 remainder of 1 or 2. P-value obtained from a binomial test of the null hypothesis that the true probability
336 of a differentially spliced region having a length divisible by three is 1 in 3 and an alternative hypothesis
337 that the probability is greater than 1 in 3.
338

Alternative Splicing		Divisible by 3	Remainder of 1	Remainder of 2	P-value
Location	Event				
Coding region	Annotated as alternatively spliced	3,248	2,466	2,411	3e-36
UTR	Annotated as alternatively spliced	1,152	1,127	1,113	1
Coding region	Minor form is expressed	173	149	153	8e-6
UTR	Minor form is expressed	34	20	13	0.03
Coding region	Differentially spliced	33	24	18	0.03
UTR	Differentially spliced	6	5	11	0.79

339

340 **Table II.** Differential splicing detected using RNA-Seq and re-tested using capillary gel electrophoresis
341 (CGE). P is the p-value obtained from comparing roots and shoots PSI from CGE.
342

Gene ID	Description	AS type	Avg. RPKM Expression		RNA-Seq PSI (%L)		CGE PSI (%L)		P-value
			Root	Shoot	Root	Shoot	Root	Shoot	
LOC_Os01g25484	ferrodoxin nitrite reductase	RI	300	131	74.2	31.7	29.9	13	5e-05
LOC_Os01g35580	unknown	AS	55.7	34.9	44.0	66.3	49.9	66.9	2.47e-04
LOC_Os01g45274	carbonic anhydrase	ES	171	1,380	96.8	24.3	97.4	15.4	3e-09
LOC_Os01g51290	protein kinase	RI	49.6	49.9	88.4	95.1	13.3	17.2	0.03466
LOC_Os03g05390	citrate transporter	RI	219	174	86.0	96.3	86	95.5	1.2e-04
LOC_Os12g08260	dehydrogenase E1	RI	8.03	30.3	55.7	2.9	4.1	0.87	3e-05
LOC_Os01g61670	ureidoglycolate hydrolase	DS	59	37.8	78.0	31.0	59	37.8	1.6e-09
LOC_Os05g48040	MATE efflux family protein	DS	7.05	6.72	88.1	100.0	89.4	88.27	0.632
LOC_Os02g05830	ribulose bisphosphate carboxylase	RI	4.79	12.3	88.2	10.1	23.9	3.06	8e-04
LOC_Os06g05110	superoxide dismutase	RI	12.5	39.7	38.5	13.6	22.7	6.5	1.3e-07

343
344

Supplemental Table I. Excel spreadsheet file (SuppTable_PSI.xlsx) is provided as a supplement. The spreadsheet contains output of R script SplicingAnalysis.Rmd run with "coverage_threshold" parameter set to 10, where coverage_threshold is the number of reads overlapping a differentially spliced region which supports one or the other splicing choice when comparing mutually exclusive splicing choices. The R script resides in the project source code repository (https://bitbucket.org/lorainelab/ricealtsplice) in the "AltSplice" subfolder. Data files produced by the script reside in subfolder "AltSplice/results". The spreadsheet SuppTable_PSI.xlsx contains worksheets copied from script output files rootsVshoots_MSU7_10.xlsx, roots_MSU7_10.xlsx, and shoots_MSU7_10.xlsx. The worksheet comparing roots and shoots alternative splicing (from rootsVshoots_MSU7_10.xlsx) was edited by hand to include information obtained from manual examination of differentially spliced genes and regions using Integrated Genome Browser and ProtAnnot. A key describing each column and its meaning is also included. The first column of each data worksheet contains hyperlinked alternatively spliced region identifiers. To use the hyperlinks, users should first download and launch Integrated Genome Browser from http://bioviz.org. Clicking the links sends a message to Integrated Genome Browser instructing it to zoom and scroll to the alternatively spliced region. Users can then open and view RNA-Seq data sets from the study by selecting the RNA-Seq folder in the **Available Data Sets** section of the IGB **Data Access** tab.

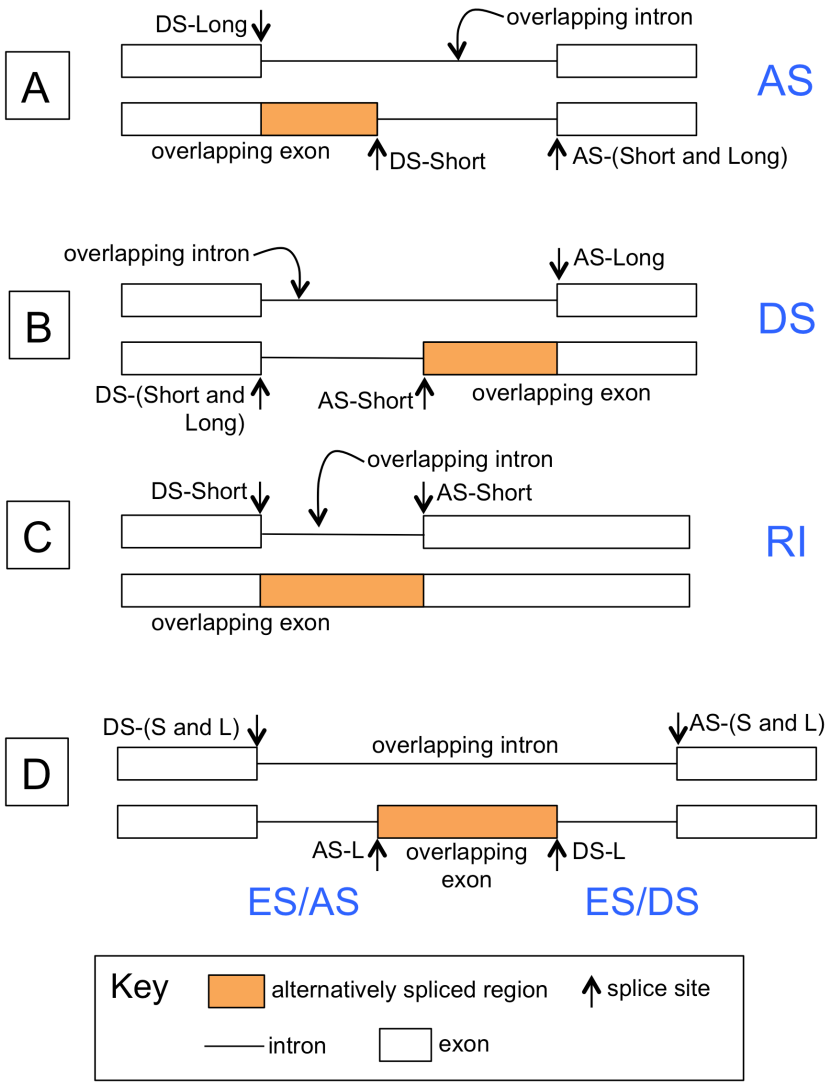
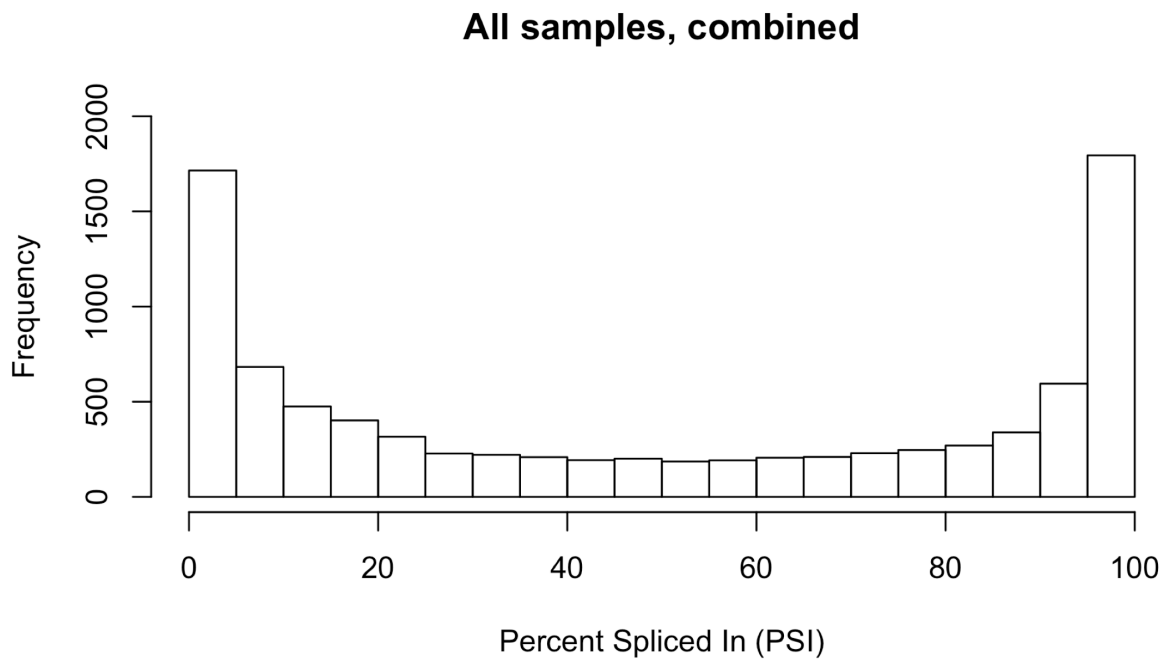


Figure 1. Alternative splicing annotation. The overlap between an intron in one gene model and an exon in another gene model defines an alternatively spliced region. Arrows indicate splice sites, named AS for acceptor site and DS for donor site. Use of sites named AS-L or DS-L causes inclusion of the differentially spliced region, generating the longer (L) isoform. Similarly, DS-S and AS-S refer to sites that exclude the differentially spliced region and generate the shorter (S) isoform. (**A**) Alternative donor sites, in which the U2 snRNP complex forms at alternative locations on the 5' end of introns. (**B**) Alternative acceptor sites, in which the U1 snRNP complex forms at alternative sites near the 3' end of alternatively spliced introns. (**C**) Alternatively spliced intron, in which a donor/acceptor site pairing can either be used or not used, forming a retained intron (RI). (**D**) Alternatively spliced, skipped exon. In exon skipping, alternative splicing involves four sites, indicated by DS-S/L, AS-L, DS-L, and SD-S/L. Exon inclusion requires assembly of two spliceosome complexes linking DS-S/L with AS-L and DS-L with AS-S/L, while exon skipping requires linking DS-S/L and AS-S/L only.



376
377 **Figure 2. Distribution of percent-spliced-in (PSI) for annotated splicing events where each choice**
378 **was supported by at least one RNA-Seq alignment.** PSI was calculated as 100*L/(S+L), where L and
379 S were the number of reads that supported the splicing choice that included (L) or excluded (S) the
380 differentially spliced region. Read alignment counts from all twelve libraries were combined to obtain a
381 global view of alternative splicing occurrence in rice seedlings. The U-shaped character of the distribution
382 persisted whether lower or higher thresholds of informative reads were used.
383

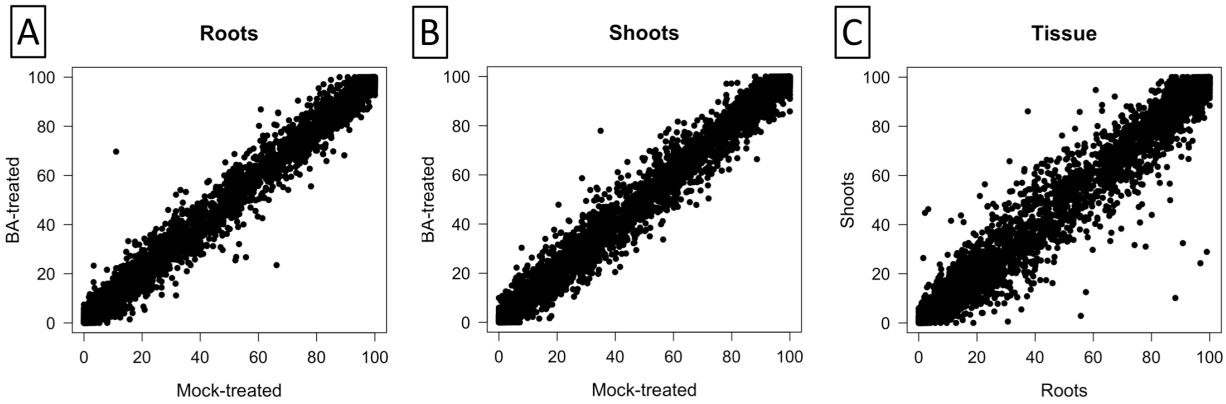


Figure 3. Scatter plots comparing percent- spliced-in (PSI) between sample types for annotated splicing events. PSI was calculated from RNA-Seq reads obtained from sequencing rice seedling shoots and roots grown hydroponically and subjected to a two-hour treatment with BA, a cytokinin analog, or a mock-treatment (control). PSI is the average of three biological replicates. Only events with at least 15 informative read alignments in all six samples being compared were included. (A) BA-treated rice roots (y axis) compared to mock roots (x axis). (B) BA-treated rice shoots (y axis) compared to mock shoots (x axis). (C) Mock shoots (y axis) compared to mock roots (x axis)

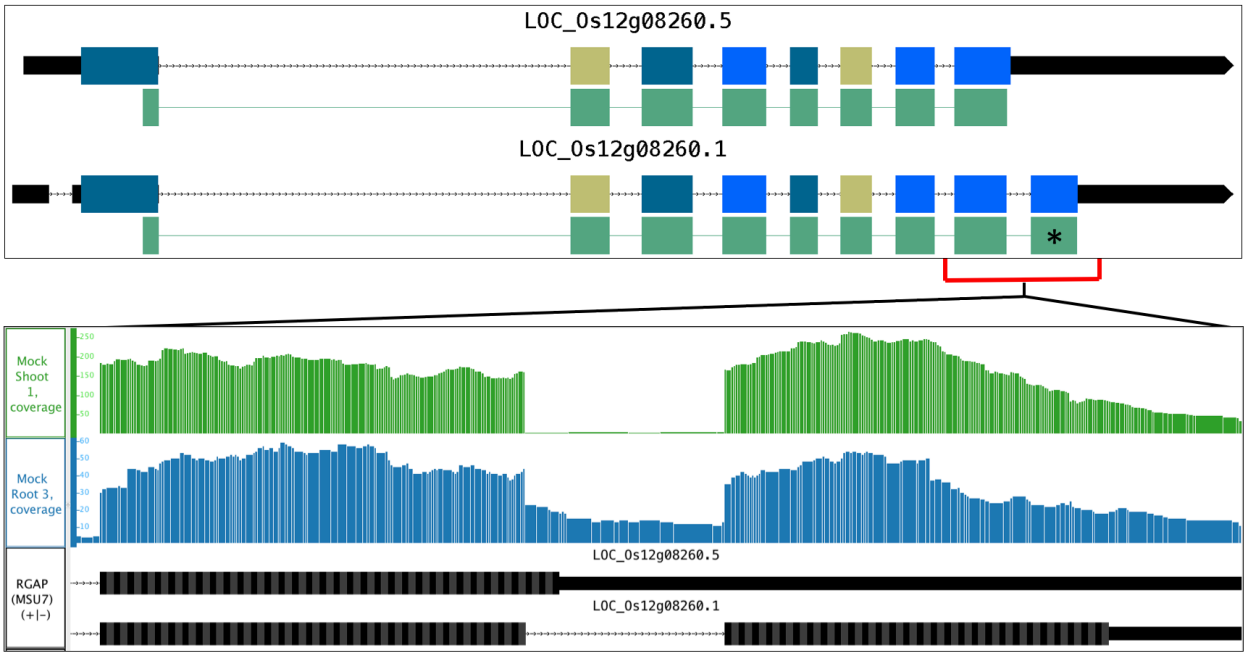


Figure 4. ProtAnnot and IGB images showing difference in splicing between shoot and root. ProtAnnot (upper panel) shows coding region exons color-coded by frame, with regions matching InterPro profiles indicated by green, linked rectangles. Asterisk highlights difference in the PANTHER InterPro profile PTHR11516 between isoforms 1 and 5 of the LOC_Os12g08260 gene. Integrated Genome Browser (lower panels) shows a zoomed-in view of RNA-Seq coverage graphs from rice root (blue) and shoot (green). Y-axis is the number of RNA-Seq aligned sequences with MSU7 gene models in black below.

403 **REFERENCES CITED**

404

405 **Barbosa-Morais NL, Carmo-Fonseca M, Aparício S** (2006) Systematic genome-wide annotation of
406 spliceosomal proteins reveals differential gene family expansion. Genome Research **16:** 66-77
407 **Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C,**
408 **Watt S, Colak R, Kim T, Misquitta-Ali CM, Wilson MD, Kim PM, Odom DT, Frey BJ,**
409 **Blencowe BJ** (2012) The evolutionary landscape of alternative splicing in vertebrate species. Science
410 **338:** 1587-1593
411 **Eddy R, Hahn DT, Moulton T** (2012) Optimizing Greenhouse Rice Production: What Is the Best Root
412 Medium? Purdue Methods for Rice Growth
413 **English AC, Patel KS, Loraine AE** (2010) Prevalence of alternative splicing choices in Arabidopsis
414 thaliana. BMC Plant Biol **10:** 102
415 **Filichkin S, Priest HD, Megraw M, Mockler TC** (2015) Alternative splicing in plants: directing traffic
416 at the crossroads of adaptation and environmental stress. Current Opinion in Plant Biology **24:** 125-
417 135
418 **Filichkin Sergei A, Cumbie Jason S, Dharmawardhana P, Jaiswal P, Chang Jeff H, Palusa**
419 **Saiprasad G, Reddy ASN, Megraw M, Mockler Todd C** (2015) Environmental Stresses Modulate
420 Abundance and Timing of Alternatively Spliced Circadian Transcripts in Arabidopsis. Molecular
421 Plant **8:** 207-227
422 **Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang H-Y, Dosztányi Z, El-**
423 **Gebali S, Fraser M, Gough J, Haft D, Holliday GL, Huang H, Huang X, Letunic I, Lopez R, Lu**
424 **S, Marchler-Bauer A, Mi H, Mistry J, Natale DA, Necci M, Nuka G, Orengo CA, Park Y,**
425 **Pesseat S, Piovesan D, Potter SC, Rawlings ND, Redaschi N, Richardson L, Rivoire C,**
426 **Sangrador-Vegas A, Sigrist C, Sillitoe I, Smithers B, Squizzato S, Sutton G, Thanki N, Thomas**
427 **PD, Tosatto Silvio C E, Wu CH, Xenarios I, Yeh L-S, Young S-Y, Mitchell AL** (2017) InterPro in
428 2017—beyond protein family and domain annotations. Nucleic Acids Research **45:** D190-D199
429 **Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi**
430 **M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A** (2016) The Pfam protein families
431 database: towards a more sustainable future. Nucleic Acids Research **44:** D279-D285
432 **Freese NH, Norris DC, Loraine AE** (2016) Integrated genome browser: visual analytics platform for
433 genomics. Bioinformatics **32:** 2089-2095
434 **Gulledge AA, Roberts AD, Vora H, Patel K, Loraine AE** (2012) Mining Arabidopsis thaliana RNA-
435 seq data with Integrated Genome Browser reveals stress-induced alternative splicing of the putative
436 splicing regulator SR45a. American journal of botany **99:** 219-231
437 **Gupta V, Estrada AD, Blakley I, Reid R, Patel K, Meyer MD, Andersen SU, Brown AF, Lila MA,**
438 **Loraine AE** (2015) RNA-Seq analysis and annotation of a draft blueberry genome assembly
439 identifies candidate genes involved in fruit ripening, biosynthesis of bioactive compounds, and stage-
440 specific alternative splicing. Gigascience **4:** 5
441 **Kalsotra A, Cooper TA** (2011) Functional consequences of developmentally regulated alternative
442 splicing. Nature Reviews. Genetics **12:** 715-729
443 **Kalyna M, Barta A** (2004) A plethora of plant serine/arginine-rich proteins: redundancy or evolution of
444 novel gene functions? Biochemical Society transactions **32:** 561-564
445 **Katz Y, Wang ET, Airoldi EM, Burge CB** (2010) Analysis and design of RNA sequencing experiments
446 for identifying isoform regulation. Nat Methods **7:** 1009-1015
447 **Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz**
448 **DC, Tanaka T, Wu J, Zhou S, Childs KL, Davidson RM, Lin H, Quesada-Ocampo L,**
449 **Vaillancourt B, Sakai H, Lee SS, Kim J, Numa H, Itoh T, Buell CR, Matsumoto T** (2013)
450 Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and
451 optical map data. Rice **6:** 1-10

452 **Keller M, Hu Y, Mesihovic A, Fragkostefanakis S, Schleiff E, Simm S** (2016) Alternative splicing in
453 tomato pollen in response to heat stress. DNA Research**:** dsw051
454 **Li S, Yamada M, Han X, Ohler U, Benfey PN** (2016) High resolution expression map of the
455 Arabidopsis root reveals alternative splicing and lincRNA regulation. Developmental cell **39:** 508-
456 522
457 **Loraine AE, McCormick S, Estrada A, Patel K, Qin P** (2013) RNA-seq of Arabidopsis pollen
458 uncovers novel transcription and alternative splicing. Plant Physiol **162:** 1092-1109
459 **Mall T, Eckstein J, Norris D, Vora H, Freese NH, Loraine AE** (2016) ProtAnnot: an App for
460 Integrated Genome Browser to display how alternative splicing and transcription affect proteins.
461 Bioinformatics **32:** 2499-2501
462 **Mei W, Boatwright L, Feng G, Schnable JC, Barbazuk WB** (2017) Evolutionarily Conserved
463 Alternative Splicing Across Monocots. Genetics **207:** 465-480
464 **Naftelberg S, Schor IE, Ast G, Kornblihtt AR** (2015) Regulation of alternative splicing through
465 coupling with transcription and chromatin structure. Annual review of biochemistry **84:** 165-198
466 **Palusa SG, Ali GS, Reddy ASN** (2007) Alternative splicing of pre-mRNAs of Arabidopsis
467 serine/arginine-rich proteins: regulation by hormones and stresses. The Plant Journal **49:** 1091-1107
468 **Plass M, Agirre E, Reyes D, Camara F, Eyras E** (2008) Co-evolution of the branch site and SR
469 proteins in eukaryotes. Trends in Genetics **24:** 590-594
470 **Raines T, Blakley IC, Tsai Y-C, Worthen JM, Franco-Zorrilla JM, Solano R, Schaller GE, Loraine**
471 **AE, Kieber JJ** (2016) Characterization of the cytokinin-responsive transcriptome in rice. BMC Plant
472 Biology **16:** 260
473 **Reddy ASN, Marquez Y, Kalyna M, Barta A** (2013) Complexity of the Alternative Splicing Landscape
474 in Plants. The Plant Cell **25:** 3657-3683
475 **Sakai H, Lee SS, Tanaka T, Numa H, Kim J, Kawahara Y, Wakimoto H, Yang CC, Iwamoto M,**
476 **Abe T, Yamada Y, Muto A, Inokuchi H, Ikemura T, Matsumoto T, Sasaki T, Itoh T** (2013) Rice
477 Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics.
478 Plant Cell Physiol **54:** e6
479 **Salz HK** (2011) Sex Determination in Insects: a binary decision based on alternative splicing. Current
480 opinion in genetics & development **21:** 395-400
481 **Stamm S, Smith C, Lührmann R** (2012) Alternative pre-mRNA Splicing: Theory and Protocols. Wiley-
482 Blackwell
483 **Sun Y, Hou H, Song H, Lin K, Zhang Z, Hu J, Pang E** (2018) The comparison of alternative splicing
484 among the multiple tissues in cucumber. BMC Plant Biology **18:** 5
485 **Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan**
486 **A, Narechania A** (2003) PANTHER: A Library of Protein Families and Subfamilies Indexed by
487 Function. Genome Research **13:** 2129-2141
488 **Vaneechoutte D, Estrada AR, Lin YC, Loraine AE, Vandepoele K** (2017) Genome-wide
489 characterization of differential transcript usage in Arabidopsis thaliana. Plant J **92:** 1218-1231
490 **Vitulo N, Forcato C, Carpinelli EC, Telatin A, Campagna D, D'Angelo M, Zimbello R, Corso M,**
491 **Vannozzi A, Bonghi C, Lucchin M, Valle G** (2014) A deep survey of alternative splicing in grape
492 reveals changes in the splicing machinery related to tissue, stress condition and genotype. BMC Plant
493 Biology **14:** 99-99
494 **Young MD, Wakefield MJ, Smyth GK, Oshlack A** (2010) Gene ontology analysis for RNA-seq:
495 accounting for selection bias. Genome Biol **11:** R14
496