This article is a *Plant Cell* Advance Online Publication. The date of its first appearance online is the official date of publication. The article has been edited and the authors have corrected proofs, but minor changes could be made before the final version is published. Posting this version online reduces the time to publication by several weeks.

LARGE-SCALE BIOLOGY ARTICLE

Genome-Wide Analysis of Alternative Splicing in *Zea mays*: Landscape and Genetic Regulation[C][W]

Shawn R. Thatcher,[a] Wengang Zhou,[b] April Leonard,[a] Bing-Bing Wang,[b,c] Mary Beatty,[b] Gina Zastrow-Hayes,[b] Xiangyu Zhao,[a,d] Andy Baumgarten,[b] and Bailin Li[a,1]

[a] DuPont Pioneer, Wilmington, Delaware 19880
[b] DuPont Pioneer, Johnston, Iowa 50131
[c] Huazhi Rice Biotech Company, Changsha, Hunan 410125, China
[d] Shandong Agricultural University, Shandong 271000, China

Alternative splicing enhances transcriptome diversity in all eukaryotes and plays a role in plant tissue identity and stress adaptation. To catalog new maize (*Zea mays*) transcripts and identify genomic loci that regulate alternative splicing, we analyzed over 90 RNA-seq libraries from maize inbred lines B73 and Mo17, as well as Syn10 doubled haploid lines (progenies from B73 × Mo17). Transcript discovery was augmented with publicly available data from 14 maize tissues, expanding the maize transcriptome by more than 30,000 and increasing the percentage of intron-containing genes that undergo alternative splicing to 40%. These newly identified transcripts greatly increase the diversity of the maize proteome, sometimes coding for entirely different proteins compared with their most similar annotated isoform. In addition to increasing proteome diversity, many genes encoding novel transcripts gained an additional layer of regulation by microRNAs, often in a tissue-specific manner. We also demonstrate that the majority of genotype-specific alternative splicing can be genetically mapped, with *cis*-acting quantitative trait loci (QTLs) predominating. A large number of *trans*-acting QTLs were also apparent, with nearly half located in regions not shown to contain genes associated with splicing. Taken together, these results highlight the currently underappreciated role that alternative splicing plays in tissue identity and genotypic variation in maize.

INTRODUCTION

Alternative splicing (AS) of pre-mRNA is a crucial regulatory mechanism that significantly increases transcriptome and proteome diversity in all eukaryotic organisms (Stamm et al., 2005). AS events fall into five broad categories: intron retention (IR), exon skipping, alternative donor, alternate acceptor, and alternate position (change in both donor and acceptor positions). The most common form of AS in plants is thought to be intron retention, although its prevalence may be overestimated due to sequencing of pre-mRNA intermediates (Lorković et al., 2000; Marquez et al., 2012). AS is catalyzed by massive ribonucleoprotein complexes known as spliceosomes (Saltzman et al., 2011), and selection of intron removal is regulated most strongly by *cis*-acting elements within each exon known as consensus splice sequences, with nearly all plant introns possessing a 5′ GU and a 3′ AG. Additionally, introns themselves are biased toward AU compared with exons that are more GC rich, with some variation occurring between monocots and dicots (Black, 2003; Reddy et al., 2013). Other *cis*-acting elements, known as exonic and intronic splicing enhancers as well as exonic and intronic splicing silencers, have also been shown to contribute to splice site selection in plants (Pertea et al., 2007). *Trans*-acting factors play a role in AS through the action of serine-arginine (SR) proteins that typically promote intron removal and heterogeneous nuclear ribonucleoproteins (HnRNPs) that typically function to inhibit it (Erkelenz et al., 2013). In addition to generating additional protein variants, AS events can affect transcript stability through alteration of untranslated regions (UTRs) (Kalyna et al., 2012) or localization (Göhring et al., 2014). Transcripts generated by AS also have the potential to gain or lose microRNA (miRNA) binding sites, which may change their expression pattern significantly (Yang et al., 2012).

Many alternatively spliced isoforms are known to have tissue-preferential expression patterns (Emrich et al., 2007), and differential AS has been demonstrated to play important roles in development (Rühl et al., 2012; Staiger and Brown, 2013) as well as stress responses (Li et al., 2013; Staiger and Brown, 2013; Cui et al., 2014). Genes encoding regulatory proteins are much more likely to have alternatively spliced isoforms, highlighting the importance of AS for gene expression networks (Duque, 2011). AS often leads to the generation of transcripts coding for truncated proteins, which have been thought to act primarily by stimulating nonsense-mediated decay through inclusion of premature stop codons (Lewis et al., 2003). This mechanism, termed regulation of unproductive splicing and translation, has been

[1] Address correspondence to bailin.li@cgr.dupont.com.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Bailin Li (bailin.li@cgr.dupont.com).
[C] Some figures in this article are displayed in color online but in black and white in the print edition.
[W] Online version contains Web-only data.
www.plantcell.org/cgi/doi/10.1105/tpc.114.130773

 © 2014 American Society of Plant Biologists. All rights reserved.

shown to play a role in a number of biological processes, including circadian clock control (Filichkin and Mockler, 2012). Recent work on microproteins has challenged the assumption that all proteins encoded by transcripts with premature stop codons are nonfunctional by demonstrating that some truncated proteins can retain domains capable of interacting with substrates or protein complexes in order to directly antagonize the action of full-length proteins (Staudt and Wenkel, 2011).

It is currently estimated that more than 60% of intron-containing genes undergo AS in the model plant *Arabidopsis thaliana*, with additional studies expected to push this percentage significantly higher (Filichkin et al., 2010; Syed et al., 2012). Despite its major role in gene regulation and function, the transcript diversity resulting from AS has yet to be fully explored in many other model plants, including maize (*Zea mays*). Beyond its importance as a major food crop, maize offers many advantages for the study of AS. Many maize genotypes and mutants have been shown to have substantial phenotypic variation, some of which may be the result of changes in alternatively splicing (Jiang et al., 2012; Li et al., 2012; Xing et al., 2014). Additionally, maize has well-established mapping populations that are suitable for determining the quantitative trait loci (QTLs) that regulate genotype-dependent AS (Yu et al., 2008). Mapping populations are indispensable for examining AS regulation, since intron removal is regulated by *cis*-acting consensus sequences as well as *trans*-acting SRs and HnRNPs. The fine mapping of *trans*-acting QTLs found through analyses of these populations also has the potential to uncover novel regulators of AS.

In this study, we used more than 200 Illumina RNA-seq libraries to identify transcripts arising from a variety of tissues as well as two different genotypes of maize. Novel and known transcripts were examined for tissue-specific expression patterns and new isoforms were investigated to determine how their divergence from annotated transcripts affects their protein coding and regulation by miRNAs. Additionally, the IBM Syn10 DH population (Hussain et al., 2007) was used to map QTLs regulating AS events that differ between B73 and Mo17 genotypes. Taken together, our results expand the maize transcriptome by more than 25%, uncover thousands of tissue-specific isoform expression patterns, and demonstrate that the majority of genotype-dependent AS variations map to *cis*-loci.

RESULTS

Prediction of Novel Transcripts

In order to discover and map novel transcripts in maize, 94 paired-end RNA-seq libraries were constructed from 5-week-old leaves, resulting in more than six billion genome-matched reads, with an average length of 50 nucleotides (Table 1). These libraries included three B73, three Mo17, and 88 intermated B73 × Mo17 (IBM) Syn10 doubled haploid (DH) lines. The IBM mapping population was originally created through 10 generations of B73 and Mo17 intermating, followed by doubled haploid generation, which resulted in a population containing highly recombinant fixed alleles (Hussain et al., 2007). Transcript discovery was augmented by the inclusion of 142 publically available B73 RNA-seq libraries originating from 14 different tissue types, totaling over two billion genome-matched reads (Table 1). Full details for individual libraries are available in Supplemental Data Set 1. All libraries were genome matched using TopHat2 (Kim et al., 2013), followed by novel transcript discovery using the Cufflinks pipeline (Trapnell et al., 2010) with the working gene set (WGS) of 137,000 annotated maize transcripts (version 5a; http://www.maizesequence.org) as a reference transcriptome.

Transcript prediction from public data and the IBM population were initially performed as two separate analyses, yet generated a novel transcript set with a high degree of overlap (Figure 1). Merging of the two predicted novel transcript sets resulted in 92,079 potential novel isoforms of annotated genes and 11,524 new transcripts (length > 50) originating from intergenic regions. Novel isoforms of known genes were discovered from genes of various lengths, with the most new isoforms coming from 2- to 6-kb-long genes (Supplemental Figures 1A and 1B). They contained a total of over 119,000 novel introns, which were

Table 1. Summary Statistics for RNA-seq Libraries

Description	Genotype	Libraries	Total Reads	Genome Matched	Percentage Matched
B73 hydroponic	B73	3	252,961,366	249,428,288	99%
Mo17 hydroponic	Mo17	3	405,071,583	393,962,382	97%
IBM hydroponic	IBM	88	5,795,253,356	5,599,170,515	97%
Anther	B73	1	38,074,756	36,554,492	96%
Ear	B73	4	104,293,259	98,987,393	95%
Embryo	B73	7	60,710,425	55,861,189	92%
Endosperm	B73	13	144,540,885	131,347,944	91%
Leaf	B73	42	664,025,044	618,671,115	93%
Ovule	B73	1	36,964,181	35,379,281	96%
Pollen	B73	1	38,623,695	37,342,145	97%
Root	B73	18	296,713,582	272,807,740	92%
Shoot apical meristem	B73	10	148,544,984	135,325,790	91%
Seed	B73	20	346,866,162	320,235,834	92%
Seedling	B73	2	23,661,408	22,675,374	96%
Shoot	B73	14	136,391,616	121,490,509	89%
Silk	B73	1	24,398,322	23,372,552	96%
Tassel	B73	8	175,790,705	166,472,788	95%

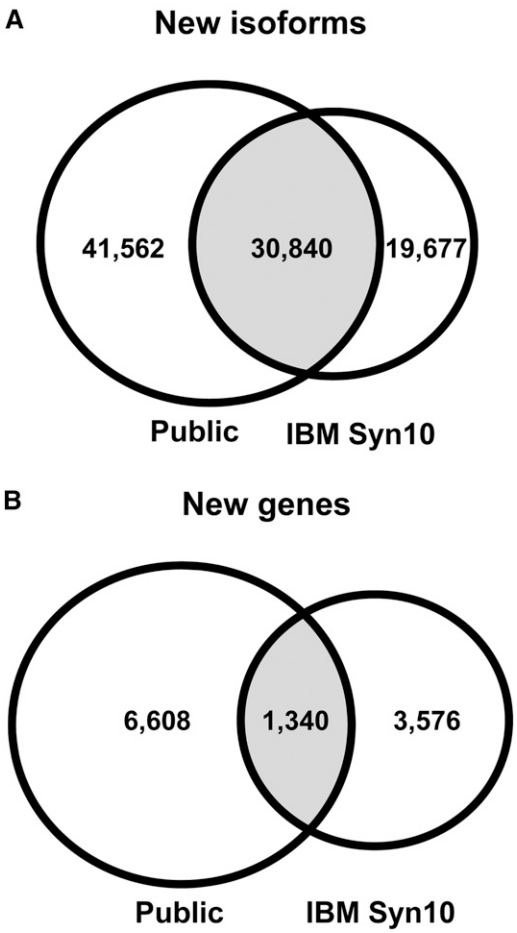

Figure 1. Overlap of Computationally Predicted Transcripts Generated from IBM Syn10 Mapping Population Libraries Compared with Public B73 Tissue Libraries.

(A) Overlap of new isoforms arising from known genes predicted from separate analyses of IBM Syn10 and public libraries.
(B) Overlap of entirely new genes predicted from separate analyses of IBM Syn10 and public libraries.

compared with the 351,000 known maize introns to determine any deviation from the expected consensus sequences and AU bias. Annotated introns have an average AU content of 58% and more than 96% contain a 5′ GU and 3′ AG, while <1% of randomly generated introns would be expected to contain both consensus splice sequences. Novel introns generated by Cufflinks were highly similar to known introns, with nearly 95% containing both splice sequences and an average AU content of 59%. Analysis of the size distribution of novel introns (Supplemental Figures 1C and 1D) revealed that they were significantly longer than known introns (average length 3038 versus 628). This difference may arise from a bias against discovery of larger introns in earlier annotation efforts, especially those utilizing computational prediction (Roy and Penny, 2007). Although GU-AG introns are by far the most abundant class, a smaller set of U12-type introns that typically have AU-AC donor/acceptor sequence have been shown to have significant regulatory importance (Patel et al., 2002). Our analysis revealed 298 known introns with AU-AC sequences (0.084% of all known introns), while our discovery efforts nearly tripled the number of introns with this important regulatory motif, adding 560 AU-AC introns (0.32% of all new introns).

Comparison of Novel Transcripts with an Artificial Randomly Generated Set

In order to assess the quality and ideal abundance cutoff for novel isoforms, a set of artificial isoforms based on known transcripts was created. One artificial isoform was randomly generated for each annotated transcript by modification of the known transcript based on AS categories: intron retention, exon skipping, alternative donor, alternative acceptor, and alternative position. The 137,000 artificial transcripts each differed from a known transcript by one random splicing modification, making them an ideal set to compare against. While 95% of the introns generated via Cufflinks contained both 5′ GU and 3′ AG, analysis of this artificial set showed that only 0.3% of its randomly generated new introns contained these consensus sequences and the average AU content was only 53%. The bias for consensus sequence and AU content among computationally predicted introns compared with randomly generated ones implies that they are indeed bona fide junctions. However, it is still likely that many of the new transcripts represent low abundance processing intermediates.

To determine an abundance cutoff for novel isoforms, known and randomly generated transcripts were quantified using Cuffdiff (Roberts et al., 2011) in the 14 public libraries, as well as B73, Mo17 parents, and IBM DH lines. Cutoffs ranging from 0.01 to 10 fragments per kilobase per million reads (FPKM) were applied and the fraction of transcripts having expression above each cutoff in at least one tissue was determined. At an extremely low expression cutoff of 0.01 FPKM, 72% of known transcripts were expressed in at least one tissue, while 57% of artificially generated transcripts were similarly expressed. Taking 0.01 as the basal expression level, the loss of known transcripts as the abundance cutoff increased (false negatives) was then plotted against the loss of artificial transcripts (false positives; Supplemental Figure 2). In order to maximize the retention of known transcripts and minimize the retention of artificial ones, 1.3 FPKM was chosen as the expression cutoff. After application of this strict abundance filter to the novel transcripts generated via Cufflinks, 34,545 isoforms of known transcripts and 2630 transcripts from intergenic regions remained. The cDNA sequence, protein translation, and expression data for all known and new transcripts are available in Supplemental Data Set 2. The average expression of the new transcripts was 5.1 FPKM with a SD of 8.7 FPKM, while known transcripts also had an average expression of 5.1 with a SD of 7.0. These new transcripts increase the percentage of intron-containing genes that undergo AS from 28% (WGS) to 40% (WGS plus novel transcripts), bringing maize closer in line with other model plants (Syed et al., 2012).

Frequency of Alternative Splicing Types

Novel isoforms above the 1.3 FPKM cutoff were then examined to determine their effect on the alternative splicing landscape in

maize. AS Pipe (Wang et al., 2008) was used to categorize splicing events of genes in both the WGS as well as the WGS with novel transcripts included, revealing that many genes utilized multiple different splicing mechanisms (Table 2). Genes with alternative acceptor sites had the largest relative increase, overtaking exon skipping as the second most common AS event (Table 2) and bringing maize's splicing landscape closer in line with *Arabidopsis* (Reddy et al., 2013). Still, many genes were found to have novel transcripts resulting from exon skipping, increasing the number of genes with this AS mechanism by 88%. Although intron retention was the most prevalent form of AS in the WGS, it has the lowest relative increase (50%) when novel transcripts were included. This may be the result of the strict abundance filter applied to transcript discovery, which would tend to exclude lower abundance processing intermediates, or may be due to most IR events already having been identified. The average expression of known transcripts with IR was 3.1 FPKM, compared with 2.1 FPKM for novel transcripts and 5.1 for all transcripts, indicating that an abundance of very low level processing intermediates does not explain the prevalence of IR in known transcripts. Regardless of the splicing mechanism, this large increase in new alternatively spliced transcripts has the potential to alter the potential maize proteome dramatically.

Potential Effect of New Transcripts on the Maize Proteome

Potential proteins were computationally predicted for all novel transcripts. It is important to note that not all predicted proteins may be expressed, as some novel transcripts may be subject to nonsense-mediated decay or sequestered away from translation machinery. In order to determine their possible effect on the maize proteome, computationally predicted proteins encoded by novel isoforms of known genes were compared with annotated proteins generated by their most similar known transcript. HMMER3 (Eddy, 2011) was used to identify conserved domains within both sets of proteins, which were then compared in a pairwise manner. A total of 11.7% of the novel proteins lost at least one functional domain, but still retained some conserved regions. *3BETAHSD/D2*, which encodes an endoplasmic reticulum-localized enzyme involved in sterol synthesis, exemplifies such a gene (Figure 2A). The annotated protein contains both the catalytic and endoplasmic reticulum association domains, while a new isoform lacks the endoplasmic reticulum association domain, potentially altering its function significantly. A total of 3.4% of the novel proteins gained additional functional domains compared with their most similar annotated protein. This group included INOSITOL-REQUIRING ENZYME1 (IRE1), a highly conserved gene that is crucial for the unfolded protein response (UPR) under stress (Shamu and Walter, 1996; Bernales et al., 2006). It has been shown in yeast and mammalian systems that IRE1 is activated when it forms homodimers via the interaction its luminal dimerization domains. After activation, IRE1 catalyzes splicing of a transcription factor that activates UPR genes. In maize, the annotated IRE1 homolog lacks the dimerization domain (Figure 2B), but transcript discovery uncovered a 5′ extension that encodes this missing domain (Figure 2B). Transcripts encoding the known truncated protein as well as the newly discovered full-length isoform both have broad tissue expression (Supplemental Data Set 2), suggesting that the alternative start sites of IRE1 may play an additional role in regulating the UPR.

Although the majority of domain gains, losses, and swaps occurred at the N or C terminus, a substantial amount of novel proteins had internal additions or losses without alteration of their reading frame. The annotated WALL ASSOCIATED KINASE3 (WAK3; GRMZM2G050536), for instance, lacks an EF hand calcium binding activation domain in the middle of the protein, which a novel transcript encodes (Figure 2C). Interestingly, a key member of the spliceosome also encodes a novel isoform that has extensively modified internal domains. *PRE-mRNA-PROCESSING-SPLICING FACTOR8* (*PRP8*) codes for a new isoform that retains a nuclear localization signal, core spliceosome interaction domain, and DEAD-box binding domain but lacks RNA and U5/U6 interaction domains (Figure 2D). A small fraction of the computationally predicted proteins (1.8%) contained functional domains that completely differed from those encoded by their most similar annotated transcript. GRMZM2G119248 encodes two annotated transcripts that translate into an asparagine synthase, but two new isoforms were discovered that instead encode a putative bromodomain-containing transcription factor (Figure 2E). Although both transcripts share a large amount of overlap, differences in translation start site cause the 3′ UTR of the annotated transcripts to become the coding region of the novel transcripts, resulting in entirely different proteins (Supplemental Figure 3). In most tissues, the asparagine synthase transcripts predominate, but in pollen and anther, one of the transcription factor-encoding isoforms is almost exclusively expressed (Figure 3A). A total of 64.8% of the novel transcripts encoded proteins containing the same domains as their most similar annotate protein. Although they contained the same domains, the majority did have somewhat altered protein coding that could affect unannotated domains or result in important modifications of known domains. In all, only 14% of the novel transcripts encoded proteins that were

Table 2. Frequency of Alternative Splicing Types in the WGS and WGS with Computationally Predicted Transcripts Included

Class	WGS	Percentage[a]	WGS+Novel	Percentage	Increase
Intron retention	6,123	62%	9,166	58%	50%
Exon skipping	3,303	33%	6,202	39%	88%
Alternate acceptor	2,895	29%	6,227	39%	115%
Alternate donor	2,137	22%	4,126	26%	93%
Alternate position	1,086	11%	1,746	11%	61%
Genes with AS	9,872		15,771		60%

[a]Many genes utilize multiple splicing categories, causing percentages to sum to greater than 100.

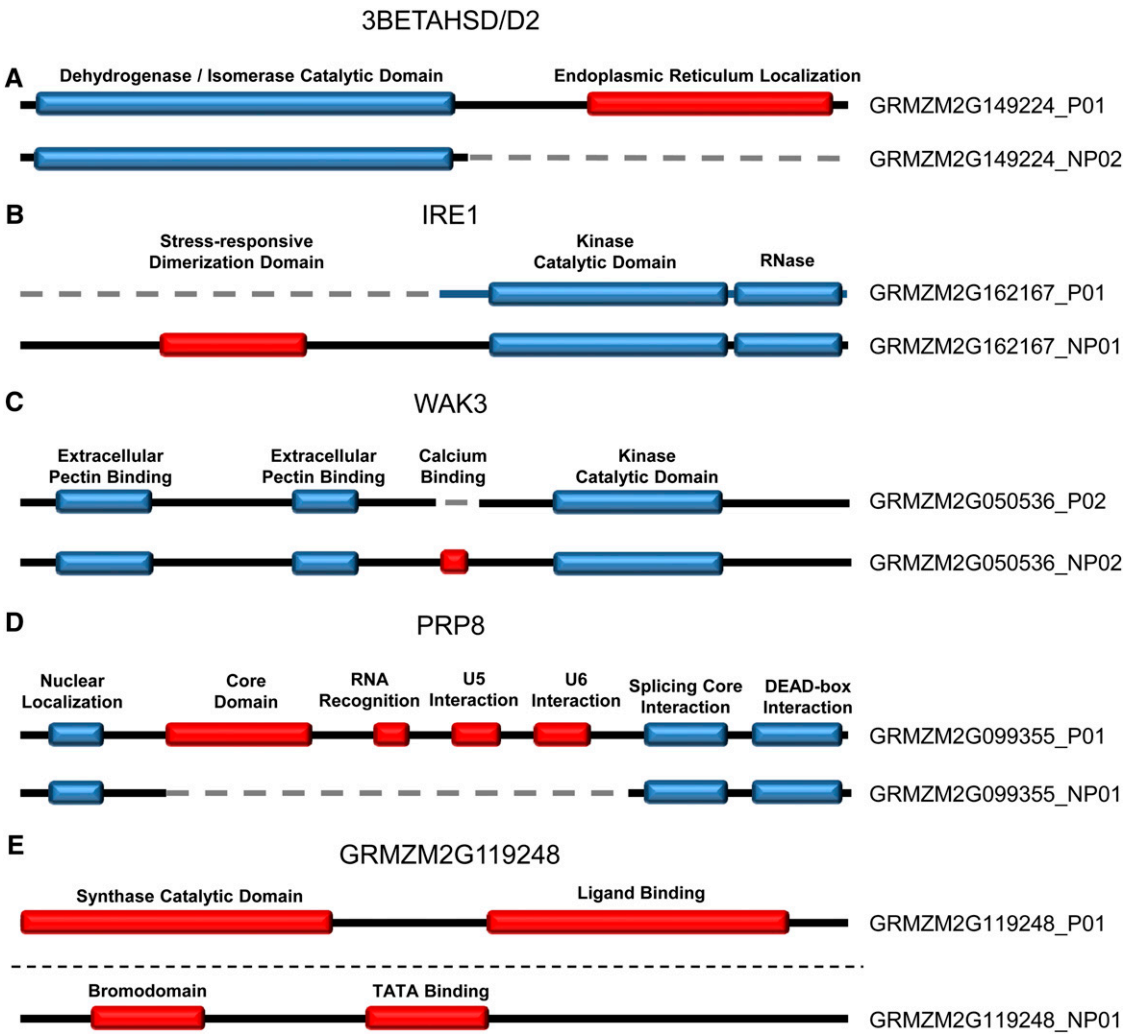

Figure 2. Examples of Proteins Encoded by Known Transcripts (Top) Compared with Those Encoded by Novel Transcripts (Bottom).

(A) 3BETAHSD/D2's loss of reticulon ER localization domain.
(B) IRE1's gain of unfolded protein-responsive luminal dimerization domain.
(C) WAK3's gain of EF hand calcium binding activation domain.
(D) PRP8's loss of RNA and U5/U6 interaction domains.
(E) GRMZM2G119248's switch from asparagine synthase to a putative bromodomain-containing transcription factor.
[See online article for color version of this figure.]

100% identical to those generated by their most similar known transcripts. However, these transcripts still possessed different UTRs, which could affect their stability and regulation by miRNAs.

Effect of New Transcripts on Maize miRNA Targets

miRNAs are a class of small RNAs that typically function by guiding cleavage of target mRNAs through base-pairing. They have been shown to play important roles in a variety of processes, including development and response to environmental stress (Hsieh et al., 2009; Lin et al., 2010; Debernardi et al., 2012). In order to assess how known miRNAs interact with the new isoforms and genes, miRNA targets were first predicted against all annotated transcripts using previously described methods (Fahlgren et al., 2007). Using a miRNA/target pairing score cutoff of 3.0, 393 known genes encoding 1719 known isoforms were predicted targets of at least one annotated maize miRNA. Inclusion of the novel transcripts resulted in 47 new genes and 68 new isoforms that are potential miRNA targets (Supplemental Data 3). Many of these events were the result of novel transcripts with IR, such as GRMZM2G357595, which produces two novel transcripts with a miR159-sensitive retained intron (Figure 4A). Both of these novel isoforms are expressed at lower levels than the annotated transcripts arising from this gene (Supplemental Data Set 2), but other transcripts that gained miRNA target sites were actually the predominant isoform in specific tissues.

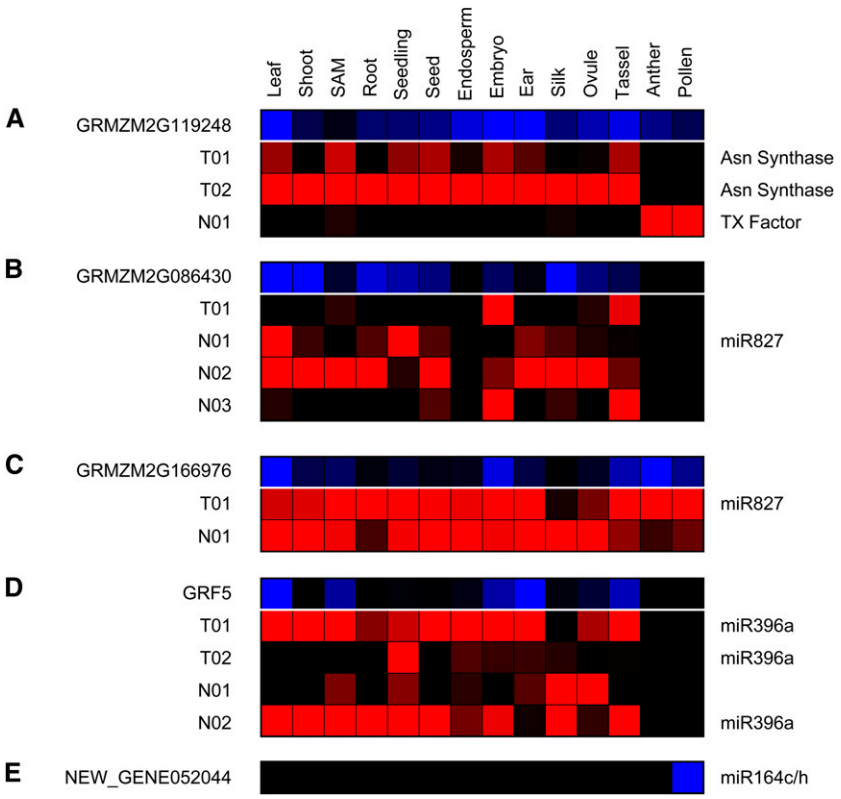

Figure 3. Relative Expression Levels of Isoforms for Genes with Substantial Protein or miRNA Binding Modifications in the Novel Transcript Set.

Relative gene expression differences between tissues are denoted in blue. Relative isoform abundance within each tissue is denoted in red.
(A) One new isoform of GRMZM2G119248 encoding a putative transcription factor.
(B) One new isoform of GRMZM2G086430 is targeted by miR827.
(C) One new isoform of GRMZM2G166976 loses the target binding site for miR827.
(D) One new isoform of GRMZM2G034876 loses the target binding site for miR396a.
(E) Novel pollen-specific gene targeted by miR164c/h.

Interestingly, this set included two SPX domain-containing genes (GRMZM2G086430 and GRMZM2G018018), which typically play a role in phosphate homeostasis. miR827 has a very well established role in nutrient level-based regulation of SPX domain proteins in other systems (Hsieh et al., 2009; Lin et al., 2010), but the annotated transcripts for these genes in maize did not include regions of miR827 complementarity. Novel transcripts were discovered for both genes that include UTRs that are targets of miR827 (Figure 4B; Supplemental Figure 4A). The new isoform for each gene codes for the same protein as the annotated transcripts, with the only variation occurring in the 5′ UTR where miR827 binds. The relative expression levels of GRMZM2G086430's isoforms varied by tissue, with some tissues predominantly expressing the miR827-sensitive isoform and others the miR827-insensitive isoform (Figure 3B).

Novel isoforms that were similar to known transcripts with established miRNA target sites were also examined for the loss of their target site. Forty novel isoforms were found to have lost an miRNA binding site relative to their most similar annotated isoform, potentially altering their expression pattern significantly (Supplemental Data Set 3). Interestingly, this group included another SPX domain-containing gene, GRMZM2G166976, which was found to produce a novel miR827-insensitive transcript with broad tissue expression (Figure 3C; Supplemental Data Set 3). *GENERAL REGULATORY FACTOR5* (*GRF5*), a member of the well-conserved GRF family whose regulation by the miR396 family is crucial for development (Debernardi et al., 2012), was also found to produce a novel insensitive target. Maize encodes three annotated isoforms of *GRF5*, and two additional novel isoforms were predicted for it. The exon targeted by the miR396 family is skipped in one new isoform (Figure 4C), removing the ability of miR396 to regulate it. This miR396-insensitive isoform is expressed at its highest levels in silk and ovule, while miR396-sensitive isoforms predominate in most other tissues (Figure 3D). Several completely new genes were also found to be targets of known miRNAs, one of which was a predicted target of both miR164c and miR164 h (Figure 4D). Examination of the expression level of this gene revealed that it is exclusively expressed in pollen (Figure 3E), where the miR164 family is thought to play a developmental role (Válóczi et al., 2006).

Tissue-Specific Transcript Expression

Many known and novel transcripts were found to be expressed above 1.3 FPKM in only one tissue type, with expression of <0.1

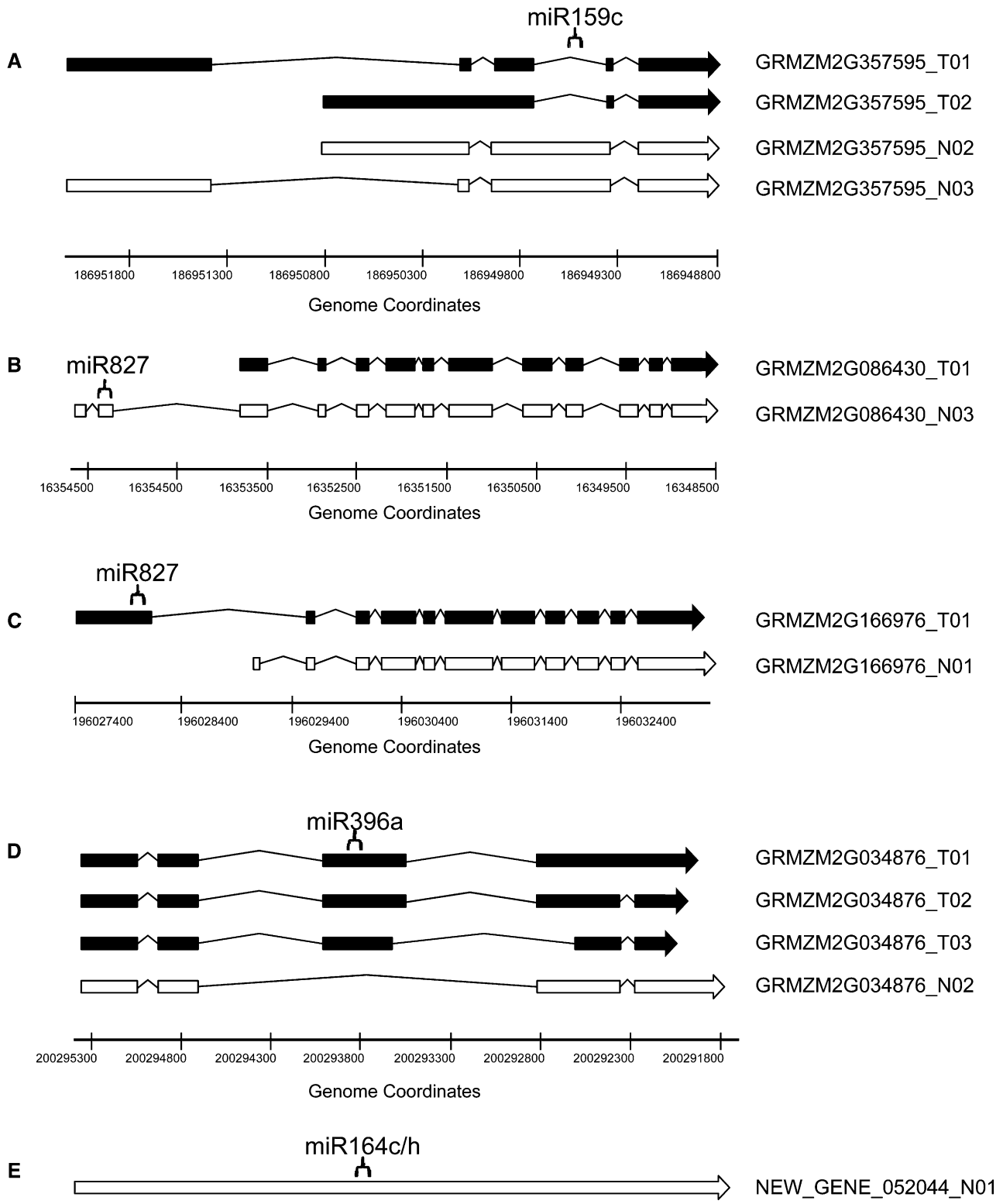

Figure 4. Gain and Loss of miRNA Target Sites in the Novel Transcript Set.

Novel transcripts (gray) that gained or lost miRNA target sites relative to their most similar known isoforms (black) are shown.
(A) Two new isoforms of GRMZM2G357595 are targeted by miR159c.
(B) One new isoform of GRMZM2G086430 is targeted by miR827.

FPKM in all others. The number of unique transcripts found varied significantly by tissue, with shoot apical meristem possessing the most unique isoforms (1039) and embryo possessing the least (49; Figure 5). A more detailed examination of these tissue-specific transcripts revealed that 63% of the genes that encode them were also only expressed above the cutoff in a single tissue, indicating that these transcripts' tissue specificity occurs at the transcriptional level. The remaining 37% of tissue-specific transcripts had other isoforms that were expressed in different tissues, implying that their tissue specificity likely results from alternative splicing (Figure 5).

Additionally, more than 3600 transcripts were expressed above 1.3 FPKM in only two tissue types, with expression of <0.1 FPKM in all others. Transcripts that were expressed in only two tissues were examined to determine that tissues overlapped most frequently. These transcript overlaps served as a quality control, revealing tissue-specific isoform expression patterns that fit well with previous gene expression work. Pollen, which is known to have a unique tissue expression pattern (Becker et al., 2003; Loraine et al., 2013), overlapped well with only anther (Supplemental Figure 5A). On the other hand, shoot apical meristem, which contains many undifferentiated cells, had a very even distribution of tissue overlap (Supplemental Figure 5B). In addition to demonstrating expected tissue overlap, some interesting patterns were revealed from this analysis. Leaf and root have very similar tissue overlap patterns, pairing strongly with seedling, seed, and shoot tissues (Supplemental Figures 5C and 5D). Despite a similar pattern of overlap with other tissues, only 29 transcripts were expressed solely in root and leaf tissue themselves (Supplemental Figures 5C and 5D).

Tissue-Specific Alternative Splicing

Since absolute isoform expression is the result of both transcriptional activity and splice site selection, an analysis of isoform abundance relative to total gene expression was also performed. First, the relative percentage that each isoform represents of total gene expression was determined for each gene in each tissue. The isoform percentages of each gene were then compared against those of the other 13 tissues in a pairwise manner, and the average difference was determined. Across all tissues the average difference was only 17%, with pollen's relative isoform expression differing the most from other tissues (25%) and seed differing the least (14%) (Supplemental Figure 6A). Cuffdiff was then used to test for statistically significant AS events among different tissue data sets. Anther, ovule, pollen, seedling, and silk did not have enough biological replicates to generate statistically significant AS calls and were excluded from this analysis (Table 1). A pairwise analysis of the remaining tissues revealed 1918 different statistically significant ($q \leq 0.05$) events from 803 different genes. Embryo had the highest number of AS events when compared with other tissues, followed by shoot apical meristem (Supplemental Figure 6B). Of the 803 genes with any significant tissue-specific AS, 322 were present in the differential splicing test between shoot apical meristem and embryo. This result is in keeping with recent work in soybean (*Glycine max*) (Shen et al., 2014) and further confirms the importance of AS in rapidly growing and differentiating tissues.

In order to assess the possible functional importance of tissue-specific AS events, a comparison of the conserved protein domains between tissues was performed. For each gene with significant tissue-specific AS, the two isoforms that had the largest change (relative to total gene expression) was determined. These isoform pairs represent a switch from one isoform to the other when comparing tissues, with one gaining the most relative abundance and the other losing the most. HMMER3 was used to find conserved domains for each of the two proteins, which were then compared against each other. Thirty percent of the proteins pairs gained, lost, or exchanged domains relative to each other, while a small subset (1%) coded for proteins with completely different sets of domains. Although the majority of pairs had the same set of domains (69%), only 26% coded for identical proteins.

Isoform pairs were also examined for the gain or loss of miRNA target sites. Seven pairs had one isoform that was capable of being targeted by a miRNA, while the other was not, although five of these had target scores of 3.5 to 4.0 (Fahlgren et al., 2007) and are likely to have miRNAs that only affect translation (Supplemental Data Set 3). Interestingly, all seven pairs had tissue-specific AS involving embryo. The two clearest cases involved a loss of a miR393 target site in the auxin signaling gene GRMZM5G848945 in embryo relative to ears and the gain of a miR169 target site in the translation elongation gene GRMZM2G343543 in embryo relative to shoot apical meristem.

Expression of Splicing-Related Genes

The relative isoform expression of alternatively spliced genes is thought to be regulated by a combination of *cis*-acting sequence elements as well as *trans*-acting splicing factors (Erkelenz et al., 2013). In animal system, SR and HnRNP proteins have been demonstrated to have a high degree of tissue specificity, which is thought to be a major regulatory factor in determining differential splicing between tissues (Grosso et al., 2008). In order investigate the expression pattern of these genes in maize, 76 genes that have Gene Ontology categories associated with splicing were quantified using Cuffdiff. Nearly all of these genes had expression differences across the 14 tissues included in this analysis, with an average SD of over 14 FPKM (Supplemental Data Set 4). Interestingly, the total expression levels of all splicing-related genes varied across tissue type. Embryo and shoot apical meristem, which were shown to possess the largest amount of differentially spliced genes (Supplemental Figure 6B), also had the highest average expression of splicing-related genes (Supplemental Data Set 4). In contrast to the significant fluctuation of splicing factors

Figure 4. (continued).

(C) One new isoform of GRMZM2G166976 loses the target binding site for miR827.

(D) One new isoform of GRMZM2G034876 loses the target binding site for miR396a.

(E) Novel pollen-specific gene targeted by miR164c/h.

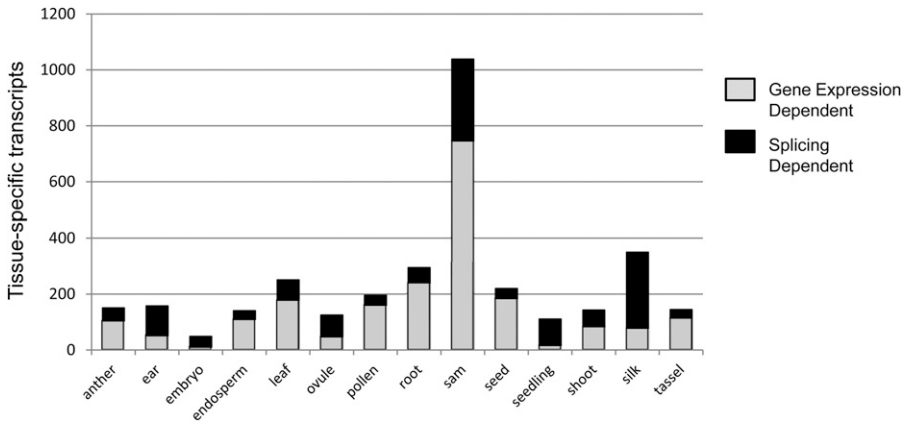

Figure 5. Single Tissue-Specific Transcripts.

Number of known and novel transcripts that are expressed at or above 1.3 FPKM in only one tissue and <0.1 FPKM in all others. Gray areas represent the portion of transcripts that are only expressed in one tissue whose genes are also only expressed in that tissue (gene expression dependent). Black areas represent the portion of transcripts that are only expressed in one tissue, with other isoforms present in other tissues (alternative splicing dependent). Transcript abundances were normalized by average transcript expression per library to account for variable tissue transcriptome complexity.

across tissues, very few of these genes were differentially expressed when comparing B73 to Mo17 genotypes (average SD of 3 FPKM; Supplemental Data Set 4).

Genetic Regulation of Genotype-Dependent Variations in AS

In order to identify genes that were differentially spliced in a genotype-dependent manner, Cuffdiff was run using B73 and Mo17 parent libraries. This initial analysis revealed 328 statistically significant ($q \leq 0.05$) alternatively spliced genes (Supplemental Data Set 5). The majority of these genes had isoform expression patterns that segregated cleanly in the IBM population (Figures 6A to 6C). Both parents and 88 IBM DH lines were then genotyped using the Illumina MaizeSNP50 DNA analysis kit (Illumina). The relative percentage of total gene expression that each isoform represents was mapped using Spotfire (Kaushal and Naeve, 2004), with a minimum of 5% of total gene expression in at least 10 IBM lines required to attempt mapping. In total, 704 isoforms' relative expression was mapped successfully using this method, covering 235/328 (72%) of the genes that Cuffdiff determined to be differentially spliced between B73 and Mo17 genotypes. These isoforms included 270 computationally predicted novel transcripts, further increasing confidence in their validity.

Variations in splicing for majority of genes (87%) mapped near their annotated locations, with a median P value of 1.1×10^{-12}. Some of the genes with *cis*-acting QTLs were putative splicing factors, including GRMZM2G154278, whose human homolog *Spliceosome-Associated Protein* (*CWC15*) has been implicated as a component of the spliceosome that plays a role in cell cycle control (Hegele et al., 2012). GRMZM2G154278 has two annotated isoforms, and two additional novel isoforms were predicted for it. Out of these four isoforms, one annotated and one predicted isoform were abundant enough in the IBM population to attempt mapping. While B73 predominantly expresses the annotated isoform, Mo17 preferentially expresses a novel isoform that translates into a truncated and likely nonfunctional splicing protein, potentially contributing to the differential splicing in these genotypes (Supplemental Data Set 2). This expression pattern segregated very cleanly in the IBM population (Figure 6A) and both isoforms mapped to the region surrounding GRMZM2G154278 (Figure 7A), indicating that splicing of this gene is regulated by strong *cis*-acting elements. In order to determine a possible cause for this genotypic variability, Mo17 whole-genome shotgun reads from the Joint Genome Institute were used to assemble the region surrounding GRMZM2G154278. This analysis revealed that the B73 and Mo17 loci are highly similar, but Mo17 has a 36-bp deletion that removes a consensus AG acceptor site at the border between exons four and five (Figure 8). The loss of the consensus sequence relative to B73 appears to result in the use of an alternate acceptor site that causes in premature termination in the Mo17 transcript.

AS variations in a smaller number of isoforms (13%) mapped to locations at least 20 Mb away from the genes that encode them, with a median P value of 9.7×10^{-8}. These putative *trans*-acting QTLs were distributed evenly across the genome, and 53% of them were within 10 MB of at least one of the 79 genes falling into Gene Ontology categories associated with splicing (Supplemental Figure 7). In *Arabidopsis*, more than 200 proteins are thought to be involved in spliceosome assembly and regulation (Reddy et al., 2013; Staiger and Brown, 2013). The other 47% of *trans*-acting QTLs that are not near known splicing factors likely represent unidentified maize splicing genes, which could be further examined in the future by fine mapping and cloning. Genes with *trans*-acting QTLs are exemplified by the conserved myosin tail binding protein GRMZM2G136455, which is thought to be involved in Golgi vesicle-mediated transport (Hashimoto et al., 2008). Maize encodes two annotated isoforms that have identical protein translations but different 3′ UTRs. Both isoforms are expressed in all B73 tissues surveyed (Supplemental Data Set 2), but GRMZM2G136455 isoform two has no detectible

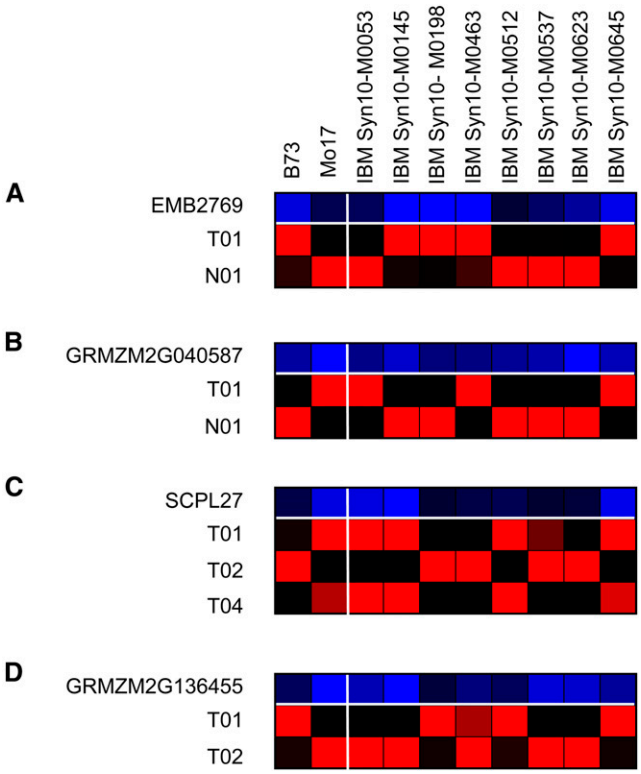

Figure 6. Relative Expression Level of Isoforms for Genes with Differential Genotype-Dependent Alternative Splicing Variations in B73, Mo17, and IBM Syn10 Lines.

Relative gene expression between genotypes is denoted in blue. Relative isoform abundance within each genotype is denoted in red.

(A) GRMZM2G154278, which has a strong *cis*-acting QTL with clean segregation.

(B) GRMZM2G040587, which has a strong *cis*-acting QTL with clean segregation.

(C) GRMZM2G155232, which has a strong *cis*-acting QTL.

(D) GRMZM2G136455, which has a strong *trans*-acting QTL.

expression in Mo17. GRMZM2G136455's isoform expression segregates nearly as cleanly as genes with *cis*-acting QTLs (Figure 6D), and both isoform expression patterns map to the same location in the genome (Chr1: 208941992; Figure 7B). Manual examination of the recombination breakpoints surrounding this region reduced the interval to a 2.2 million base pair region centered on Chr1: 209301649. There are 115 annotated genes within this region, many of which have unknown functions and could represent novel splicing factor candidates for future fine mapping efforts.

Gene Expression Variations Compared with Alternative Splicing Variations

The relative importance and prevalence of AS variations compared with gene expression variations in defining phenotypic variations and tissue identity remains an open question. Although our pairwise Cuffdiff analysis revealed more than 1900 genes with statistically significant ($q \leq 0.05$) tissue-specific splicing between nine tissues (Supplemental Figure 6B), it also uncovered nearly 34,000 significant ($q \leq 0.05$) gene expression differences (Supplemental Data Set 6). Similarly, using the same significance cutoffs, 328 genes with genotype-specific splicing were found, while nearly 1700 genes with differential expression were revealed in the same analysis (Supplemental Data Set 6). The predominance of variations in gene expression compared with AS may represent a real biological importance for variations at the expression level but could also arise from computational difficulties inherent in assigning RNA-seq reads to specific isoforms.

To assess the effect of expression level on the statistical power of identifying differentially spliced and expressed genes, the median expression levels for genes with significant splicing or expression variations were determined. Significantly differentially expressed genes in tissue and genotype manners had medians of only 4.5 and 3.0 FPKM, respectively, while differentially spliced genes had a median of 14.0 FPKM in both tests (Supplemental Data Set 6). This disparity is further compounded by the fact that differential splicing calls require gene expression above a threshold in at least two tissues, compared with one tissue being required for differential gene expression. In tissue-specific tests, 14,241 genes have expression above the differential splicing median in at least two tissues, while 30,997 have expression above the differential expression median in at least one. This result implies that while the increased expression requirement for splicing tests may cause some underestimation of the prevalence of differential splicing relative to gene expression, it cannot explain the 18-fold higher level of significant tissue-specific gene expression changes. Analysis of genotype-specific tests, on the other hand, revealed that 7074 genes have expression above the differential splicing median, while 25,399 have expression above the differential expression median. Therefore, while 5 times as many genes showed differential genotypic gene expression compared with splicing, a large part of the difference may be accounted for by the higher expression requirement for determining statistically significant AS between genotypes. Additional RNA-seq data with deeper sequencing coverage should result in the identification of more novel transcripts, as well more differentially expressed AS isoforms.

PCR Validation of Novel Transcripts

In order to assess the reliability of computationally predicted transcripts, a subset of 20 genes with novel isoforms was chosen for RT-PCR validation. This set included eight genes noted previously for their impact on the proteome and miRNA pathways (Figures 2 to 4; Supplemental Figure 4), which had expression in B73 or Mo17 leaf and generated isoforms that could be distinguished via RT-PCR. The remaining 12 genes were randomly selected from the pool of genes for which genotypic splicing variation was successfully mapped (Supplemental Data Set 5). As a positive control, primers were first designed to amplify regions shared by all known and novel isoforms whenever possible (Supplemental Data Set 7), with an 88% success rate (Supplemental Figure 8). Amplification was next attempted for abundant annotated isoforms from the same set of genes. In order to achieve specificity, primers were designed to include exons or span splice junctions that were unique to each isoform. In total, 83% of known isoforms were able to be amplified in at least one genotype (Supplemental Figure 9). Finally, computationally predicted novel isoforms were

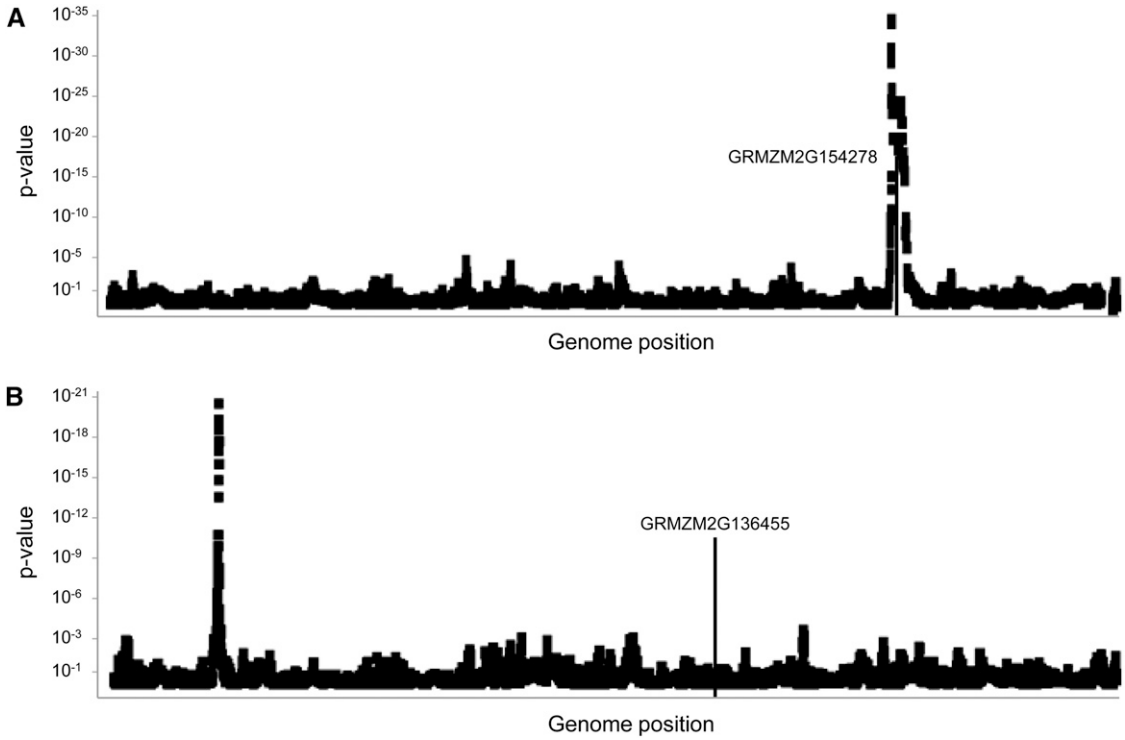

Figure 7. Mapping Results Reveal QTLs Regulating Genotype-Dependent Alternative Splicing Variations.

Genes with statistically significant ($q < 0.05$) splicing differences between B73 and Mo17 were selected for mapping using Spotfire. The relative percentage of total gene expression that each isoform represents was used as the phenotype and compared with marker data from the Illumina MaizeSNP50 DNA analysis kit (Illumina).

(A) GRMZM2G154278, exemplifying the case of a typical *cis*-acting QTL, with one clear peak surrounding the gene's annotated position on chromosome 8.

(B) GRMZM2G136455, exemplifying a *trans*-acting QTL. The QTL is located on chromosome 5, while the gene is located on chromosome 1.

amplified using isoform-specific primers, with 19/23 (83%) detected, validating their prediction (Supplemental Figure 10). In many cases, known and novel isoform-specific amplifications also generated additional larger bands that likely represent intron retention events that fell below the abundance cutoff for isoform discovery. The isoforms that were unable to be amplified by primers targeting their novel splice junctions or exons may be the result of false computational predictions, low transcript expression in B73 and Mo17 leaf tissue, or nonideal design of primers, which were required to achieve isoform specificity in some genes. In addition to confirming the majority of the attempted known and novel transcripts, this analysis also validated genotype-specific expression of several known (Supplemental Figure 11A) and novel (Supplemental Figure 11B) transcripts. Overall, the success in amplifying this subset of novel isoforms, coupled with the similar successful rate of mapping genotypic variation of novel and known isoforms, serves as a strong indicator of the validity of the larger set of new transcripts.

DISCUSSION

Deep sequencing of maize B73 and Mo17 inbreds, their IBM doubled haploid progeny, as well as inclusion of published B73 data from 14 different tissues enabled the identification of more than 30,000 new maize transcripts, bringing the percentage of intron-containing maize genes with alternative splicing closer in line with other model plants. The frequency of different splicing categories in maize was mostly similar to other plants, with the exception of exon skipping, which was significantly more prevalent in maize than it is in any other model plant (Campbell et al., 2006; Marquez et al., 2012; Shen et al., 2014). It has been proposed that the prevalence of intron retention in plants may actually be the result of sequencing low abundance splicing intermediates (Lorković et al., 2000; Marquez et al., 2012). However, our analysis of both known and novel transcripts arising from IR indicated that they have comparable expression to the entire pool of maize transcripts.

The majority of novel splicing events lead to transcripts encoding new proteins and our analysis revealed that they often contained domain losses, gains, or substitutions relative to their most similar known isoform (Table 3). Intriguingly, the core splicing component, *PRP8*, was found to encode a novel transcript that retains its nuclear localization and spliceosome interaction domains but loses mRNA, U5, and U6 binding domains, which have been shown to be crucial for splice site selection and catalysis (Umen and Guthrie, 1995). By binding to nuclear-localized

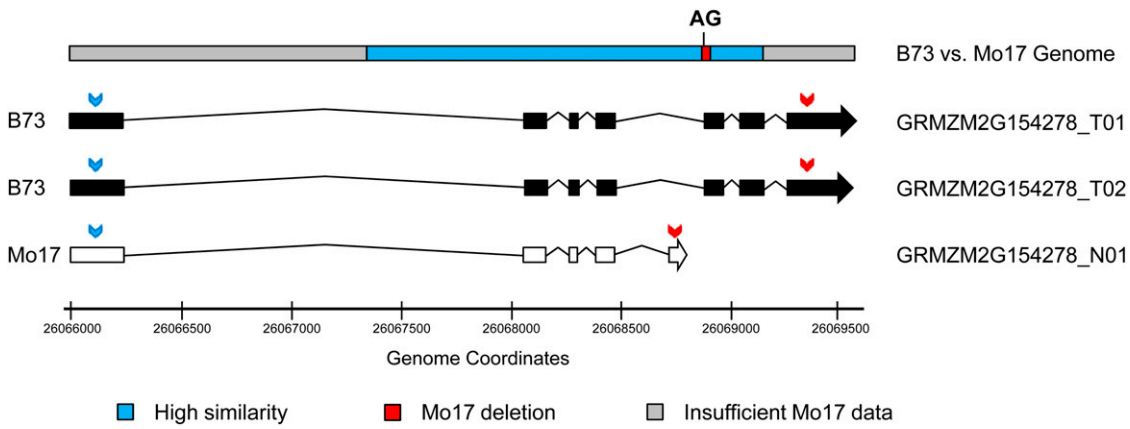

Figure 8. Splicing Factor EMB2769 Is Differentially Spliced in B73 Compared with Mo17, Which Has a 36-bp Deletion That Removes the AG 3′ Consensus Acceptor Sequence.

Translation start site for the longest possible protein is denoted with blue arrows while the stop codon for that protein is denoted in red.

spliceosomes without the ability to interact with RNA, this novel protein may inhibit splicing. The expression of the transcript encoding truncated PRP8 varied substantially across different tissues, ranging from zero (embryo and tassel) to over 40% (ovule) of the total *PRP8* gene expression (Supplemental Data Set 2). Such large differences in the expression of the full-length versus truncated protein for a core component of the spliceosome could have dramatic effects on splicing and may therefore represent an important regulator of tissue-specific AS.

Many new alternatively spliced transcripts had the potential to code for entirely different proteins compared with their most similar annotated isoform (Figure 2). One such case has been previously reported in mangrove (*Bruguiera gymnorrhiza*), in that a chloroplast-encoded ribosomal protein gene (RPL32) was inserted into the exon of a Cu-Zn superoxide dismutase, resulting in a chimeric gene that produces both proteins through AS (Cusack and Wolfe, 2007). Our transcript discovery revealed that an annotated maize asparagine synthase (GRMZM2G119248) produces additional novel isoforms that code for a putative transcription factor (Figure 2E) that is almost exclusively expressed in anther and pollen (Figure 3A). The production of these disparate proteins is achieved through alternative transcription and translation start sites that turn the 3′ UTR of the asparagine synthase into the coding region of the transcription factor (Supplemental Figure 3). BLASTP analysis revealed that the predicted protein generated by this novel isoform is conserved at greater than 75% identity in wheat (*Triticum aestivum*), millet (*Pennisetum glaucum*), *Brachypodium*, and rice (*Oryza sativa*), further supporting a possible functional role. Interestingly, this novel protein has 36% identity to an *Arabidopsis* transcription factor (*TAF8*) that is also expressed most strongly in pollen. The strong and apparently conserved tissue specificity of the isoform encoding the transcription factor implies that it may play a specialized role regulating gene expression in anther and pollen.

Table 3. Comparison of Proteins Generated by Novel Transcripts to Proteins Generated by Their Most Similar Known Isoform

Class	Proteins	Percentage
Same domains	23,079	64.8%
Lost all functional domains	4,176	11.7%
Lost some functional domains	3,876	10.9%
Exchanged some functional domains	1,590	4.5%
Gained new functional domains	1,197	3.4%
Only novel has functional domains	1,051	3.0%
Exchanged all functional domains	657	1.8%

Our analysis of the interaction between splicing and miRNA regulation uncovered an additional layer of regulation for many genes, with the nutrient-responsive miR827's targets showing a particularly interesting pattern. Two new and one known miR827 target belong to the SPX family, while a fourth is a homolog of the *Arabidopsis* developmental regulator *PASTICCINO1* (Supplemental Figure 4B) (Vittorioso et al., 1998). All three SPX genes' miRNA sensitivity is regulated via alternative start sites that result in the inclusion of miR827-sensitive exons in their 5′ UTRs (Figure 4; Supplemental Figure 4A), while the *PASTICCINO1* homolog AC216247.3_FG005 gains miR827 sensitivity through extension of its 3′ UTR (Supplemental Figure 4B). Interestingly, all four miR827 target genes produce abundant miR827-sensitive and -insensitive isoforms, often in a tissue-dependent manner (Figure 3). GRMZM2G166976 has an interesting expression pattern in reproductive tissue, producing more abundant miR827-sensitive isoforms in tassel, anther, and pollen, while miR827-insensitive forms predominate in silk and ovule (Figure 3C). Genotype-specific expression adds another layer onto miR827 regulation, with Mo17 expressing significant levels of both miR827-sensitive and -insensitive GRMZM2G018018, whereas B73 has almost no detectible expression level for this gene (Supplemental Figure 7 and Supplemental Data Set 2). These findings present an interesting example of transcription start site, splicing, and miRNA regulation all potentially affecting tissue and genotype-specific responses to nutrient levels.

Our use of multiple different maize tissues enabled the identifications of nearly 2000 cases of tissue-specific alternative splicing. The vast majority of these cases resulted in changes at the protein level, although only 30% of these differences were significant enough to result in gains, losses, or substitutions of whole domains.

Protein differences that did not appear to alter overall domain architecture may be representative of tissue-specific fine tuning, such as minor alterations in terminal domains affecting localization, but likely also include changes in unannotated domains. Only a few cases of tissue-specific AS affecting miRNA's interaction with their target genes were uncovered, although, interestingly, all of them involved embryo. The auxin signaling gene GRMZM5G848945's loss of its auxin-regulated miR393 target site in embryo had the most obvious possible functional implications, as auxin is known to play a role in embryo development (Möller and Weijers, 2009).

In our investigation of splicing in B73, Mo17, and the Syn10 IBM doubled haploid population, we attempted to map QTLs that regulate genotype-specific AS variations in plants. Over 70% of statistically significant genotype-specific AS events were able to be mapped successfully, indicating a strong level of genetic regulation of this process (Supplemental Data Set 5). Genotype-specific AS events that were unable to be mapped may be the result of variation determined by multiple QTLs with minor effects, which could be missed in this type of analysis. The successfully mapped AS events came from a variety of different genes, and nearly 90% resulted in *cis*-acting QTLs. The *cis* AS QTL tend to be more significant and have larger effects than *trans* AS QTL, which is consistent with the established importance of consensuses sequences at exon/intron boundaries and implies their importance in determining genotypic variations. Still, *trans*-acting QTLs did regulate the expression levels of more than 80 isoforms, many of that could represent unannotated splicing factors that are excellent prospects for future efforts to fine map and clone genes involved in AS regulation in maize. It is also noteworthy that the cutoff used for genotype-dependent AS variations ($q \leq 0.05$) is strict and may result in many real genotype-dependent splicing differences being missed. To determine whether real differential splicing was being missed, the significance cutoff was relaxed to $q \leq 0.15$, resulting in an additional 67 genes, with a 55% mapping success rate (Supplemental Data Set 5). Future efforts with deeper sequencing coverage and utilizing less stringent cutoffs for significant differential splicing, combined with larger mapping populations, additional genotypes, and tissue types, are likely to increase the number of genes with demonstrable genotype-dependent AS.

The mechanisms regulating AS are still being fully elucidated in plants but are thought to involve a combination of *cis*-acting sequence and methylation differences, as well as differential expression, splicing, and sequence variation of *trans*-acting SRs and HnRNPs (Black, 2003; Pertea et al., 2007; Erkelenz et al., 2013). The vast majority of genes involved in splicing had relatively constant expression between Mo17 and B73 (Supplemental Data Set 4), suggesting that the expression level of known splicing factors is not the major driving force behind genotypic AS variation. This finding is further supported by our mapping results, which demonstrate a strong bias for *cis*-acting elements as the regulatory force behind genotypic splicing differences (Supplemental Data Set 5). Tissue-specific splicing differences, on the other hand, are likely to be strongly influenced by large differences in the expression of *trans*-acting splicing factors (Supplemental Data Set 4), although *cis*-acting tissue-specific methylation differences also play an important role (Regulski et al., 2013).

In summary, our analysis greatly expanded the maize transcriptome to include more than 30,000 newly annotated transcripts. This expansion increased proteome diversity and resulted in novel proteins that gained and lost important functional domains, often in tissue-specific manners. Many genes were also found to interact with the miRNA pathway in various ways through novel miRNA-sensitive or -insensitive isoforms. We also demonstrate that the majority of genotype-specific AS maps to *cis*-acting regulatory elements, implying their importance in determining genotypic differences in splicing. The identification of a number of *trans*-acting regulatory loci by genetic mapping will also enable future efforts to identify genes responsible for AS regulation by positional cloning. Taken together, these findings highlight the currently underappreciated role that AS plays in tissue identity and genotypic variation in maize.

METHODS

Plant Growth and Harvesting

B73, Mo17, and 88 Syn10 IBM DH maize (*Zea mays*) lines (Hussain et al., 2007) were grown hydroponically, as described previously (Holloway et al., 2011), under long-day conditions (16/8-h day/night cycle at 26°C/22°C). Two 0.6-cm leaf sections from 5-week-old seedlings were harvested from three plants of each genotype, pooled, and flash-frozen in liquid nitrogen. Tissue was then used for mRNA-seq library preparation and RT-PCR validation.

mRNA-seq Library Preparation and Sequencing

Total RNA was isolated from frozen maize leaf tissues via Qiagen RNeasy kit for total RNA isolation. Libraries from the resulting total RNA were prepared using the TruSeq paired-end mRNA-Seq kit and protocol from Illumina and sequenced on the Illumina HiSequation 2500 system with Illumina TruSeq SBS v3 reagents. On average, 70 million 50-bp paired-end reads were generated for each sample (Supplemental Data Set 1). The resulting sequences were trimmed based on quality scores and mapped to the maize B73 reference genome sequence V2 and maize working gene set V5a with Tophat2 (Kim et al., 2013) using the following modifications from default parameters: maximum intron size, 100,000; minimum intron size, 20; up to two mismatches allowed. Reads aligning to multiple locations were assigned heuristically based on the abundance of surrounding regions (Kim et al., 2013).

Computational Prediction of Novel Isoforms

Genome-matched reads were assembled with Cufflinks (Trapnell et al., 2010) using the following modifications from default parameters: maximum intron size, 100,000; minimum intron size, 20. Annotation of novel junctions required at least 10 reads to span those junctions, and any new transcripts were required to represent at least 10% of the total gene abundance in at least one library. Known and novel transcripts were then quantified in each tissue and genotype with Cuffdiff (Roberts et al., 2011) using default parameters. Novel transcripts with expression less than 1.3 FPKM in all tissues were filtered out.

Generation and Application of Artificial Isoforms

One new transcript was randomly generated from each annotated maize transcript to obtain an artificial isoform set. New isoforms were generated based on known alternative splicing categories: intron retention, exon skipping, alternative donor, alternative acceptor, and alternative position. Artificial isoforms were combined with known transcripts and quantified using Cuffdiff (Roberts et al., 2011). Quantification data were then used to

determine an ideal abundance cutoff for computationally predicted transcripts, as described in the text.

Comparison of Novel and Known Protein Domains

HMMER3 (Eddy, 2011) was used to determine conserved protein domains for all novel isoforms and their most similar known transcripts, as determined by Cufflinks (Trapnell et al., 2010). Resulting domain hits for novel and similar known transcripts were then compared in a pairwise manner and classified relative to the known transcript as domain losses, gains, or substitutions.

PCR Validation of Known and Novel Isoforms

Total RNA was isolated from frozen maize leaf tissues using an E.Z.N.A. RNA extraction kit with the optional DNase digestion step (Omega Bio-Tek). RNA was then reverse transcribed using a Qiagen QuantiTect RT kit. Primers were designed to specifically amplify only one isoform via placement in unique exons or by spanning unique splice junctions with at least eight nucleotides covered on each side of the junction. Thirty-five rounds of PCR amplification using Phusion polymerase (Thermo Scientific) were performed, followed by gel visualization.

Accession Numbers

Sequence data from this article can be found in National Center for Biotechnology Information Gene Expression Omnibus under accession number GSE57337 as well as in the GenBank/EMBL data libraries under the following accession numbers: *3BETAHSD/D2*, *GRMZM2G149224*; *IRE1*, *GRMZM2G162167*; *PRP8*, *GRMZM2G099355*; *GRF5*, *GRMZM2G034876*; *EMB2769*, *GRMZM2G154278*; *SCPL27*, *GRMZM2G155232*.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Size Distribution of Known and Novel Isoforms and Transcripts.

Supplemental Figure 2. Selection of FPKM Cutoff for Novel Transcripts.

Supplemental Figure 3. Transcripts Produced by GRMZM2G119248.

Supplemental Figure 4. Gain and Loss of miR827 Target Sites in the Novel Transcript Set.

Supplemental Figure 5. Overlap of Tissue-Specific Transcript Expression.

Supplemental Figure 6. Tissue-Specific Variation in Transcript Expression and Alternative Splicing.

Supplemental Figure 7. Mapping Pattern of *Trans*-Acting QTLs Relative to Splicing Factors.

Supplemental Figure 8. Amplification of Regions Shared among All Known and Novel Isoforms for Each Gene.

Supplemental Figure 9. RT-PCR Amplification of Known Isoforms of Genes for Which Novel Transcripts Were Discovered.

Supplemental Figure 10. RT-PCR Amplification of Novel Isoforms.

Supplemental Figure 11. PCR Validation of Known and Novel Isoforms with Validated Genotype-Dependent Expression.

The following materials have been deposited in the DRYAD repository under accession number http://dx.doi.org/10.5061/dryad.g3v15.

Supplemental Data Set 1. RNA-seq Libraries Used for Transcript Discovery and Alternative Splicing Analysis.

Supplemental Data Set 2. Known and Novel Transcripts' cDNA, Protein, and Expression Information.

Supplemental Data Set 3. Novel Transcripts Which Gained or Lost miRNA Target Sites Relative to Their Most Similar Known Isoform.

Supplemental Data Set 4. Expression of Genes with Gene Ontology Categories Associated with Alternative Splicing.

Supplemental Data Set 5. Expression and Mapping Results for Genes with Statistically Significant ($q \leq 0.05$) Genotype-Dependent Alternative Splicing Variation.

Supplemental Data Set 6. Expression of Genes with Statistically Significant Expression Differences in Tissue and Genotypic Tests.

Supplemental Data Set 7. Primers Used to Amplify Known and Novel Isoforms.

ACKNOWLEDGMENTS

We thank Antoni Rafalski for insightful discussions during this work. We also thank Nathan Uhlmann and Wang Nan Hu for their technical assistance. This work was supported by a Discovery Grant from DuPont Pioneer.

AUTHOR CONTRIBUTIONS

B.L., S.T., W.Z., B.W., and A.B. conceived and designed the experiments. S.T., W.Z., A.L., M.B., G.Z., and X.Z. performed the experiments and contributed to data analysis. S.T. and B.L. wrote the article.

Received August 7, 2014; revised August 7, 2014; accepted September 10, 2014; published September 23, 2014.

REFERENCES

Becker, J.D., Boavida, L.C., Carneiro, J., Haury, M., and Feijó, J.A. (2003). Transcriptional profiling of Arabidopsis tissues reveals the unique characteristics of the pollen transcriptome. Plant Physiol. **133:** 713–725.

Bernales, S., Papa, F.R., and Walter, P. (2006). Intracellular signaling by the unfolded protein response. Annu. Rev. Cell Dev. Biol. **22:** 487–508.

Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. Annu. Rev. Biochem. **72:** 291–336.

Campbell, M.A., Haas, B.J., Hamilton, J.P., Mount, S.M., and Buell, C.R. (2006). Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. BMC Genomics **7:** 327.

Cui, P., Zhang, S., Ding, F., Ali, S., and Xiong, L. (2014). Dynamic regulation of genome-wide pre-mRNA splicing and stress tolerance by the Sm-like protein LSm5 in Arabidopsis. Genome Biol. **15:** R1.

Cusack, B.P., and Wolfe, K.H. (2007). When gene marriages don't work out: divorce by subfunctionalization. Trends Genet. **23:** 270–272.

Debernardi, J.M., Rodriguez, R.E., Mecchia, M.A., and Palatnik, J.F. (2012). Functional specialization of the plant miR396 regulatory network through distinct microRNA-target interactions. PLoS Genet. **8:** e1002419.

Duque, P. (2011). A role for SR proteins in plant stress responses. Plant Signal. Behav. **6:** 49–54.

Eddy, S.R. (2011). Accelerated Profile HMM Searches. PLOS Comput. Biol. **7:** e1002195.

Emrich, S.J., Barbazuk, W.B., Li, L., and Schnable, P.S. (2007). Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Res. **17:** 69–73.

Erkelenz, S., Mueller, W.F., Evans, M.S., Busch, A., Schöneweis, K., Hertel, K.J., and Schaal, H. (2013). Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. RNA **19:** 96–102.

Fahlgren, N., Howell, M.D., Kasschau, K.D., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., Law, T.F., Grant, S.R., Dangl, J.L., and Carrington, J.C. (2007). High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS ONE **2:** e219.

Filichkin, S.A., and Mockler, T.C. (2012). Unproductive alternative splicing and nonsense mRNAs: a widespread phenomenon among plant circadian clock genes. Biol. Direct **7:** 20.

Filichkin, S.A., Priest, H.D., Givan, S.A., Shen, R., Bryant, D.W., Fox, S.E., Wong, W.K., and Mockler, T.C. (2010). Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. Genome Res. **20:** 45–58.

Göhring, J., Jacak, J., and Barta, A. (2014). Imaging of endogenous messenger RNA splice variants in living cells reveals nuclear retention of transcripts inaccessible to nonsense-mediated decay in Arabidopsis. Plant Cell **26:** 754–764.

Grosso, A.R., Gomes, A.Q., Barbosa-Morais, N.L., Caldeira, S., Thorne, N.P., Grech, G., von Lindern, M., and Carmo-Fonseca, M. (2008). Tissue-specific splicing factor gene expression signatures. Nucleic Acids Res. **36:** 4823–4832.

Hashimoto, K., Igarashi, H., Mano, S., Takenaka, C., Shiina, T., Yamaguchi, M., Demura, T., Nishimura, M., Shimmen, T., and Yokota, E. (2008). An isoform of Arabidopsis myosin XI interacts with small GTPases in its C-terminal tail region. J. Exp. Bot. **59:** 3523–3531.

Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C.L., Pena, V., Lührmann, R., and Stelzl, U. (2012). Dynamic protein-protein interaction wiring of the human spliceosome. Mol. Cell **45:** 567–580.

Holloway, B., Luck, S., Beatty, M., Rafalski, J.A., and Li, B. (2011). Genome-wide expression quantitative trait loci (eQTL) analysis in maize. BMC Genomics **12:** 336.

Hsieh, L.C., Lin, S.I., Shih, A.C., Chen, J.W., Lin, W.Y., Tseng, C.Y., Li, W.H., and Chiou, T.J. (2009). Uncovering small RNA-mediated responses to phosphate deficiency in Arabidopsis by deep sequencing. Plant Physiol. **151:** 2120–2132.

Hussain, T., Tausand, P., Graham, G., and Ho, J. (2007). Registration of IBM2 SYN10 doubled haploid mapping population of maize. J. Plant Reg. **1:** 81.

Jiang, F., Guo, M., Yang, F., Duncan, K., Jackson, D., Rafalski, A., Wang, S., and Li, B. (2012). Mutations in an AP2 transcription factor-like gene affect internode length and leaf shape in maize. PLoS ONE **7:** e37040.

Kalyna, M., et al. (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. Nucleic Acids Res. **40:** 2454–2469.

Kaushal, D., and Naeve, C.W. (2004). Analyzing and visualizing expression data with Spotfire. In Current Protocols in Bioinformatics, A. Bateman, W.R. Pearson, L.D. Stein, G.D. Stormo, and J.R. Yates III, eds (Hoboken, NJ: John Wiley & Sons).

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. **14:** R36.

Lewis, B.P., Green, R.E., and Brenner, S.E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc. Natl. Acad. Sci. USA **100:** 189–192.

Li, W., Lin, W.D., Ray, P., Lan, P., and Schmidt, W. (2013). Genome-wide detection of condition-sensitive alternative splicing in Arabidopsis roots. Plant Physiol. **162:** 1750–1763.

Li, X., et al. (2012). Genic and nongenic contributions to natural variation of quantitative traits in maize. Genome Res. **22:** 2436–2444.

Lin, S.I., et al. (2010). Complex regulation of two target genes encoding SPX-MFS proteins by rice miR827 in response to phosphate starvation. Plant Cell Physiol. **51:** 2119–2131.

Loraine, A.E., McCormick, S., Estrada, A., Patel, K., and Qin, P. (2013). RNA-seq of Arabidopsis pollen uncovers novel transcription and alternative splicing. Plant Physiol. **162:** 1092–1109.

Lorković, Z.J., Wieczorek Kirk, D.A., Lambermon, M.H., and Filipowicz, W. (2000). Pre-mRNA splicing in higher plants. Trends Plant Sci. **5:** 160–167.

Marquez, Y., Brown, J.W., Simpson, C., Barta, A., and Kalyna, M. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. Genome Res. **22:** 1184–1195.

Möller, B., and Weijers, D. (2009). Auxin control of embryo patterning. Cold Spring Harb. Perspect. Biol. **1:** a001545.

Patel, A.A., McCarthy, M., and Steitz, J.A. (2002). The splicing of U12-type introns can be a rate-limiting step in gene expression. EMBO J. **21:** 3804–3815.

Pertea, M., Mount, S.M., and Salzberg, S.L. (2007). A computational survey of candidate exonic splicing enhancer motifs in the model plant *Arabidopsis thaliana*. BMC Bioinformatics **8:** 159.

Reddy, A.S., Marquez, Y., Kalyna, M., and Barta, A. (2013). Complexity of the alternative splicing landscape in plants. Plant Cell **25:** 3657–3683.

Regulski, M., et al. (2013). The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. Genome Res. **23:** 1651–1662.

Roberts, A., Pimentel, H., Trapnell, C., and Pachter, L. (2011). Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics **27:** 2325–2329.

Roy, S.W., and Penny, D. (2007). Intron length distributions and gene prediction. Nucleic Acids Res. **35:** 4737–4742.

Rühl, C., Stauffer, E., Kahles, A., Wagner, G., Drechsel, G., Rätsch, G., and Wachter, A. (2012). Polypyrimidine tract binding protein homologs from Arabidopsis are key regulators of alternative splicing with implications in fundamental developmental processes. Plant Cell **24:** 4360–4375.

Saltzman, A.L., Pan, Q., and Blencowe, B.J. (2011). Regulation of alternative splicing by the core spliceosomal machinery. Genes Dev. **25:** 373–384.

Shamu, C.E., and Walter, P. (1996). Oligomerization and phosphorylation of the Ire1p kinase during intracellular signaling from the endoplasmic reticulum to the nucleus. EMBO J. **15:** 3028–3039.

Shen, Y., Zhou, Z., Wang, Z., Li, W., Fang, C., Wu, M., Ma, Y., Liu, T., Kong, L.A., Peng, D.L., and Tian, Z. (2014). Global dissection of alternative splicing in paleopolyploid soybean. Plant Cell **26:** 996–1008.

Staiger, D., and Brown, J.W. (2013). Alternative splicing at the intersection of biological timing, development, and stress responses. Plant Cell **25:** 3640–3656.

Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T.A., and Soreq, H. (2005). Function of alternative splicing. Gene **344:** 1–20.

Staudt, A.C., and Wenkel, S. (2011). Regulation of protein function by 'microProteins'. EMBO Rep. **12:** 35–42.

Syed, N.H., Kalyna, M., Marquez, Y., Barta, A., and Brown, J.W. (2012). Alternative splicing in plants—coming of age. Trends Plant Sci. **17:** 616–623.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. **28:** 511–515.

Umen, J.G., and Guthrie, C. (1995). A novel role for a U5 snRNP protein in 3′ splice site selection. Genes Dev. **9:** 855–868.

Válóczi, A., Várallyay, E., Kauppinen, S., Burgyán, J., and Havelda, Z. (2006). Spatio-temporal accumulation of microRNAs is highly coordinated in developing plant tissues. Plant J. **47:** 140–151.

Vittorioso, P., Cowling, R., Faure, J.D., Caboche, M., and Bellini, C. (1998). Mutation in the Arabidopsis PASTICCINO1 gene, which encodes a new FK506-binding protein-like protein, has a dramatic effect on plant development. Mol. Cell. Biol. **18:** 3034–3043.

Wang, B.B., O'Toole, M., Brendel, V., and Young, N.D. (2008). Cross-species EST alignments reveal novel and conserved alternative splicing events in legumes. BMC Plant Biol. **8:** 17.

Xing, A., Williams, M.E., Bourett, T.M., Hu, W., Hou, Z., Meeley, R.B., Jaqueth, J., Dam, T., and Li, B. (2014). A pair of homoeolog ClpP5 genes underlies a virescent yellow-like mutant and its modifier in maize. Plant J. **79:** 192–205.

Yang, X., Zhang, H., and Li, L. (2012). Alternative mRNA processing increases the complexity of microRNA-based gene regulation in Arabidopsis. Plant J. **70:** 421–431.

Yu, J., Holland, J.B., McMullen, M.D., and Buckler, E.S. (2008). Genetic design and statistical power of nested association mapping in maize. Genetics **178:** 539–551.

Genome-Wide Analysis of Alternative Splicing in *Zea mays*: Landscape and Genetic Regulation

Shawn R. Thatcher, Wengang Zhou, April Leonard, Bing-Bing Wang, Mary Beatty, Gina Zastrow-Hayes, Xiangyu Zhao, Andy Baumgarten and Bailin Li

Plant Cell; originally published online September 23, 2014;
DOI 10.1105/tpc.114.130773

This information is current as of January 2, 2015

Supplemental Data	**http://www.plantcell.org/content/suppl/2014/09/18/tpc.114.130773.DC1.html**
Permissions	https://www.copyright.com/ccc/openurl.do?sid=pd_hw1532298X&issn=1532298X&WT.mc_id=pd_hw1532298X
eTOCs	Sign up for eTOCs at: http://www.plantcell.org/cgi/alerts/ctmain
CiteTrack Alerts	Sign up for CiteTrack Alerts at: http://www.plantcell.org/cgi/alerts/ctmain
Subscription Information	Subscription Information for *The Plant Cell* and *Plant Physiology* is available at: http://www.aspb.org/publications/subscriptions.cfm

© American Society of Plant Biologists
ADVANCING THE SCIENCE OF PLANT BIOLOGY