



Profiling Alternative 3' Untranslated Regions in Sorghum using RNA-seq Data

Min Tu and Yin Li*

Waksman Institute of Microbiology, Rutgers, The State University of New Jersey, Piscataway, NJ, United States

Sorghum is an important crop widely used for food, feed, and fuel. Transcriptome-wide studies of 3' untranslated regions (3'UTR) using regular RNA-seq remain scarce in sorghum, while transcriptomes have been characterized extensively using Illumina short-read sequencing platforms for many sorghum varieties under various conditions or developmental contexts. 3'UTR is a critical regulatory component of genes, controlling the translation, transport, and stability of messenger RNAs. In the present study, we profiled the alternative 3'UTRs at the transcriptome level in three genetically related but phenotypically contrasting lines of sorghum: Rio, BTx406, and R9188. A total of 1,197 transcripts with alternative 3'UTRs were detected using RNA-seq data. Their categorization identified 612 high-confidence alternative 3'UTRs. Importantly, the high-confidence alternative 3'UTR genes significantly overlapped with the genesets that are associated with RNA N⁶-methyladenosine (m⁶A) modification, suggesting a clear indication between alternative 3'UTR and m⁶A methylation in sorghum. Moreover, taking advantage of sorghum genetics, we provided evidence of genotype specificity of alternative 3'UTR usage. In summary, our work exemplifies a transcriptome-wide profiling of alternative 3'UTRs using regular RNA-seq data in non-model crops and gains insights into alternative 3'UTRs and their genotype specificity.

Keywords: crop, sorghum, RNA-seq, 3' untranslated regions, alternative 3'UTR, transcriptome analysis, mRNA N⁶-methyladenosine

OPEN ACCESS

Edited by:

Frederic J.J. Chain,
University of Massachusetts Lowell,
United States

Reviewed by:

Guoli Ji,
Xiamen University, China
Wenbin Mei,
Inari Agriculture, United States

*Correspondence:

Yin Li
liyin.plant@hotmail.com

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 28 April 2020

Accepted: 30 September 2020

Published: 26 October 2020

Citation:

Tu M and Li Y (2020) Profiling
Alternative 3' Untranslated Regions
in Sorghum using RNA-seq Data.
Front. Genet. 11:556749.
doi: 10.3389/fgene.2020.556749

INTRODUCTION

Sorghum bicolor is a crop globally used for food, fodder, and fuel. The importance of sorghum in agriculture and bioenergy is due to its advantages in diversity, genetics, and genomics (Calvino and Messing, 2012; Boyles et al., 2019; Xie and Xu, 2019). Sorghum cultivars are divided into groups based on their usage: grain sorghum, sweet sorghum, forage sorghum, and energy sorghum. Grain sorghum ranks 5th in global cereal production (Boyles et al., 2019); grain sorghum and forage sorghum serve as significant sources for animal feed; sweet sorghum and energy sorghum are promising bioenergy crops for sugar- and lignocellulosic-based biofuels (Mullet et al., 2014; Mathur et al., 2017; Li et al., 2018; Yang et al., 2020). Sorghum has a relatively small diploid genome with several reference assemblies (Paterson et al., 2009; Deschamps et al., 2018; McCormick et al., 2018; Cooper et al., 2019).

RNA-seq has become a ubiquitous tool in shaping our understanding of the functions of the genomic components in plants (Stark et al., 2019). RNA-seq data are not only valuable in

hypothesis-driven research but also powerful for data-driven analysis to generate new insights and testable hypotheses, directing future functional studies. Its primary applications are to measure the gene expression levels and to calculate differential gene expression (DGE). Like in model species and major crops, RNA-seq has recently been widely employed in sorghum research and advanced our understanding in many aspects of sorghum including development (Davidson et al., 2012; Kebrom et al., 2017; Turco et al., 2017; Leiboff and Hake, 2019), accumulation of sugar/biomass (McKinley et al., 2016, 2018; Mizuno et al., 2016, 2018; Zhang L. M. et al., 2018; Li et al., 2019a,b; Zhang et al., 2019; Hennem et al., 2020), stress responses and tolerance (Dugas et al., 2011; Gelli et al., 2014; Sui et al., 2015; Fracasso et al., 2016; Yang et al., 2018; Varoquaux et al., 2019), senescence (Johnson et al., 2015; Wu et al., 2016a), and regulation of miRNA and long non-coding RNA (Calvino et al., 2011; Sun et al., 2020).

3'UTR harbors important regulatory elements, *i.e.*, poly(A) signals and *cis*-acting elements. Poly(A) signals determine polyadenylation (PA), while *cis*-acting elements in 3'UTR interact with *trans*-acting factors, such as miRNAs and RNA-binding proteins (Ji et al., 2014). 3'UTR-mediated gene regulation affects mRNA localization, translation, transport, and stability (Srivastava et al., 2018). The usage of 3'UTR is complex and regulated. Multiple transcripts can be produced from the same gene through alternative splicing (AS), which could lead to alternative 3'UTRs (Abdel-Ghany et al., 2016). Additionally, a transcript can produce several isoforms with different 3'UTRs varied in length through alternative polyadenylation (APA) (Wu et al., 2011). Alternative 3'UTRs and poly(A) sites are dynamically regulated by developmental and environmental conditions and are tissue specific (Fu et al., 2016; Lin et al., 2017; Lorenzo et al., 2017; Sun et al., 2017; Téllez-Robledo et al., 2019).

To characterize the dynamics of 3'UTRs and poly(A) sites at the genome-wide level, extensive efforts have been made both in experimental and bioinformatic approaches. Several methods based on oligo(dT)-priming have been established to sequence mRNA poly(A) tails, including polyA capture (Mangone et al., 2010), sequencing APA sites (Fu et al., 2011), PolyA Site Sequencing (Shepard et al., 2011), and PA-seq (Hafez et al., 2013). Besides these, other specialized sequencing methods to capture the 3' ends of mRNAs have been developed, such as poly(A)-position profiling by sequencing (Jan et al., 2010) and 3' region extraction and deep sequencing (Hoque et al., 2013). In particular, PolyA-Tag sequencing has been extensively used for genome-wide APA studies in higher plants, including *Arabidopsis* (Shen et al., 2011; Wu et al., 2011, 2016b; Hong et al., 2018), *Chlamydomonas reinhardtii* (Bell et al., 2016), rice (Fu et al., 2016; Ye et al., 2019; Zhou et al., 2019), and *Medicago* (Wu et al., 2014). In contrast to the experimental approaches targeting poly(A) tails and/or 3'UTRs, many bioinformatic methods and pipelines have been developed to identify poly(A) sites and/or detect differential APA usage between regular RNA-seq samples [reviewed by Chen et al. (2020)]. These bioinformatic tools fall into four major types: type 1 tools include QAPA (Ha et al., 2018) and PAQR (Gruber et al., 2018) and rely on a pre-existed annotation of poly(A) sites, such as PolyA_DB (Lee et al., 2007), Polysite (Gruber et al., 2016), APASdb (You et al., 2014), and

PlantAPA (Wu et al., 2016b); type 2 tools depend on RNA-seq-based transcript reconstruction to infer 3'UTRs, including 3USS (Le Pera et al., 2015) and ExUTR (Huang and Teeling, 2017); type 3 tools, such as Kleat (Birol et al., 2015) and ContextMap 2 (Bonfert and Friedel, 2017), identify poly(A) sites using poly(A)-capped reads in RNA-seq data; and type 4 software include a majority of tools, namely, PHMM (Lu and Bushel, 2013), GETUTR (Kim et al., 2015), Change-Point (Wang W. et al., 2014), EBChangePoint (Zhang and Wei, 2016), IsoSCM (Shenker et al., 2015), DaPars (Xia et al., 2014), APATrap (Ye et al., 2018), and TAPAS (Arefeen et al., 2018). The type 4 methods identify APA dynamics based on modeling read density fluctuation in RNA-seq data. The advantages and the limitations for each type of the bioinformatic tools have been reviewed in detail, and their performance was compared using benchmarking datasets (Chen et al., 2020).

Previous studies focused on the experimental methods for sequencing poly(A) tails have enhanced our understanding of the functions of the core components of poly(A) machinery, cleavage and polyadenylation specificity factor complex (F), including CPSF30, CPSF100, and FIP1. CPSF30 affects poly(A) signal recognition on the near-upstream element and thus the choice of poly(A) sites for many genes in *Arabidopsis* (Thomas et al., 2012; Chakrabarti and Hunt, 2015). CPSF100 affects the poly(A) site choices through the far-upstream element, resulting in transcriptional read-through for many genes (Lin et al., 2017). Another subunit called FIP1 (factor interacting with PAP1), which acts as a bridge between poly(A) polymerase (PAP) and CPSF (Preker et al., 1995), has complex roles in regulating poly(A) site selection and responses to ABA and abiotic stresses in *Arabidopsis* (Téllez-Robledo et al., 2019). While these studies immensely contribute to understanding the dynamics, mechanisms, and variations in plant polyadenylation sites, these methods require specialized library-prep protocols and bioinformatic pipelines. Unlike routine RNA-seq, they are technically challenging and more expensive, preventing from becoming widely available for plant scientists. Many plant scientists work on agriculturally and economically important crops with large and complex genomes [*e.g.*, maize (Jiao et al., 2017), sorghum (Paterson et al., 2009; McCormick et al., 2018), wheat (International Wheat Genome Sequencing Consortium [IWGSC], 2008), and sugarcane (Zhang J. et al., 2018)], which present potential technical difficulties when performing multi-omics experiments. Although sequencing costs are rapidly decreasing, the majority of published RNA-seq data available in plants, especially those in non-model crops, were produced using the Illumina short-read sequencing platforms (Stark et al., 2019). For example, in sorghum, a large amount of RNA-seq studies were conventional short-read sequencing used for DGE analysis, with a few analyzing AS (Abdel-Ghany et al., 2016). These situations suggest that 3'UTR analysis using conventional RNA-seq data should be valuable and significant for non-model plants.

In this study, we sought to gain insights into alternative 3'UTRs in sorghum. Due to the importance of sorghum stem as a carbon reservoir and a conduit for the mobilization of water, nutrients, and signaling molecules, we chose to analyze the RNA-seq data of sorghum stems, which was previously generated by our group in three genetically related

but phenotypically contrasting lines—Rio, BTx406, and R9188—to identify genes and regulatory networks associated with stem sugar accumulation in sweet sorghum (Li et al., 2019a,b). Our present study has the following objectives: (1) to characterize alternative 3'UTRs in sorghum stems, (2) to gain insights into 3'UTR regulation, and (3) with the identified alternative 3'UTRs, to generate testable hypotheses for future in-depth functional studies of 3'UTR-mediated regulation in sorghum.

MATERIALS AND METHODS

RNA-seq for Profiling 3' Untranslated Regions

The RNA-seq dataset (NCBI accession PRJNA413691) used for 3'UTR analysis was generated previously by us to study genes and networks associated with soluble sugar accumulation in internodes (internodes 2, 3, and 4, numbered from top to bottom) at four post-stem elongation stages from three sorghum genotypes, Rio, BTx406, and R9188 (Li et al., 2019a). The three genotypes contrast in the phenotypes of stem sugar accumulation: Rio accumulates high contents of sugar in stems during the post-flowering stages (~20% Brix in the stem-extracted juice), while BTx406 has a low stem sugar content (<10% Brix), with R9188 having an intermediate stem sugar content especially during the post-flowering stages (Li et al., 2019a). R9188 is an introgression line developed from the BTx406/Rio cross followed by one backcross to sweet sorghum Rio and contains the early flowering and dwarf loci introgressed from grain sorghum BTx406 (Ritter et al., 2004). The dataset has an advantage such that the three genetically related genotypes allow us to investigate inter-genotype 3'UTR variations.

For each genotype, the internode samples were collected at four time points, namely, T1, T2, T3, and T4 (flag leaf stage, 100% flowering, 10 days after flowering, and 15 days after flowering, respectively), in the experimental field under a split plot design. RNA was extracted with the TRIzol method, processed to libraries, and sequenced using 150-bp pair-end according to standard protocols. Regular quality control and filtering steps were applied to the raw reads and then mapped to the sorghum reference genome BTx623 (v2.1) with Tophat (Paterson et al., 2009; Trapnell et al., 2012). To calculate gene expression (in reads per kilobase of exon per million mapped sequence reads, RPKM) and differentially expressed genes (DEGs), uniquely mapped reads were used, and DEGs were determined with DEseq and edgeR (Anders and Huber, 2010; McCarthy et al., 2012). Moreover, 18,275, 19,727, and 19,102 genes were expressed in Rio, BTx406, and R9188, respectively (total number of the expressed genes for all timepoints, genic reads per replicate ≥ 10 , average RPKM per sample ≥ 1).

Transcriptome-Wide Analysis of Alternative 3'UTRs

The publicly available bioinformatic method priUTR (Program for RNA-seq-based Identification of Alternative 3'UTR) was used in this study to identify alternative 3'UTRs using regular

RNA-seq data from sorghum (Tu, 2020). The program and its user manual are available at <https://github.com/mint1234/3UTR-priUTR>. priUTR is a type-2 bioinformatic method for identifying alternative 3'UTRs, which relies on RNA-seq-based transcript reconstruction to infer mRNA 3' end. Detailed information about this method is provided in **Supplementary Figure S1**. The alternative 3'UTR results for each genotype and timepoint were obtained with priUTR after mapping the sorghum RNA-seq data to the BTx623 reference genome using a reference-guided mode with Tophat. The reference-guided transcriptome assembly (in GTF format) and gene expression data generated from the Tophat-Cufflinks pipeline were used as input files for the priUTR program. Only those alternative 3'UTRs detected in all three replicates per genotype and time point were kept. Hereafter, a gene or a transcript with an alternative 3'UTR is mentioned as an alternative 3'UTR gene or an alternative 3'UTR transcript.

We integrated the BTx406-introgressed genes/alleles to find out alternative 3'UTR genes associated with the genetic relationship among R9188, Rio, and BTx406 (Li et al., 2019a). A total of 1,805 genes introgressed from BTx406 into R9188 were previously identified based on RNA-seq-derived SNPs with SAMTools, followed by a series of stringent filters, including the mapping quality, read depth, bi-allelism, and homogeneity and homozygosity of the SNPs. The RNA-seq read mapping data were examined for the alternative 3'UTR genes and compared between Rio, BTx406, and R9188 using a genome browser to identify the alternative 3'UTRs that were specifically detected in Rio but not in BTx406/R9188 or *vice versa*.

Expression of Sorghum m⁶A Functional Factors

The functional factors (“reader,” “writer,” and “eraser” proteins) of m⁶A modifications in *Arabidopsis* were reported previously (Duan et al., 2017; Ruzicka et al., 2017; Arribas-Hernández et al., 2018; Scutenaire et al., 2018; Wei et al., 2018), with their maize homologs phylogenetically identified (Miao et al., 2020). The protein sequences of the m⁶A functional factors were used for a BLAST search in the sorghum genome, with a threshold of E value $\leq 10^{-5}$. The best BLAST hits in sorghum were also verified by the gene orthology between maize and sorghum (Zhang et al., 2017).

Functional Annotation and Enrichment Analysis

To comprehensively understand the functions of alternative 3'UTR genes, the functional annotation containing three major sources of annotation, Gene Ontology (GO), MapMan, and KEGG, was employed (Li et al., 2019a). The GO annotation was from Phytozome and AgriGO (Du et al., 2010). For MapMan, the second-level binocodes of the annotation was used (Usadel et al., 2009). Enriched functional terms of the high-confidence alternative 3'UTR geneset (group 4, see “RESULTS”) or the m⁶A associated geneset were

calculated using clusterProfiler (*hypergeometric* < 0.05; Yu et al., 2012).

RESULTS

Transcriptome-Wide Identification of Alternative 3'UTRs in Sorghum Using RNA-seq Data

We performed a transcriptome-wide analysis of alternative 3'UTRs using the sorghum internode RNA-seq datasets (Li et al., 2019a). A total of 1,371 transcripts were identified to have alternative 3'UTRs for all the three genotypes. Among them, 754, 776, and 848 transcripts derived from 750, 771, and 841 genes were identified in Rio, BTx406, and R9188, respectively (Figures 1A,B). We focus on analyzing the features of the alternative 3'UTR genes in terms of their 3'UTR lengths, gene order, and functions rather than the 3'UTR dynamics over the time points. Therefore, the alternative 3'UTR transcripts from the four time points were pooled together for downstream analysis. The read depth within a mRNA tends to drop rapidly when reaching to its 3' end in RNA-seq data, which is known as the 3' bias of RNA-seq (Li et al., 2015). Due to the 3' bias of RNA-seq, we removed any transcripts of which the predicted alternative 3'UTR differed from its corresponding annotation, with no more than 50 bp in length to avoid potential bioinformatic artifacts. After that, 1,197 alternative 3'UTRs were kept. Because priUTR is a type-2 bioinformatic tool for alternative 3'UTR detection and could have inherent limitations of transcript assembly tools, the read mapping results of these 1,197 alternative 3'UTRs were visualized on Integrative Genomics Viewer (IGV; Thorvaldsdóttir et al., 2013) and manually compared with their corresponding annotations in BTx623 reference genome in order to find out potential sources of false prediction and to identify high-confidence alternative 3'UTR (Figures 1A,C). Based on the types of possible mis-prediction, the 1,197 transcripts were categorized into five groups (Figure 1A). Group 1 (namely, the "3' overlapping gene" group) represents the alternative 3'UTRs that could be likely mis-predicted because a gene is located at the 3' downstream of the alternative 3'UTR gene and overlaps with the predicted 3'UTR (Figure 2A). Group 2 (namely, the "3' adjacent gene" group) represents the alternative 3'UTRs that could be mis-predicted because a gene locates at the 3' downstream of the alternative 3'UTR gene and overlaps only with the predicted 3'UTR extension but not with the annotated 3'UTR region (Figure 2B). Group 3 (namely, the "mixed transcript" group) stands for those potentially false predictions, likely due to gene models producing multiple transcripts. In such a situation, the alternative 3'UTR transcript could be embedded within a longer transcript or be mis-predicted due to other transcripts with a similar 3'UTR length (Figure 2C). Group 4 is a collection of high-confidence alternative 3'UTR transcripts of which the results in IGV supported the predicted alternative 3'UTRs (Figures 2D,E). Group 5 are the transcripts of which the alternative 3'UTR seems to be mis-predicted by other miscellaneous reasons. Groups 1,

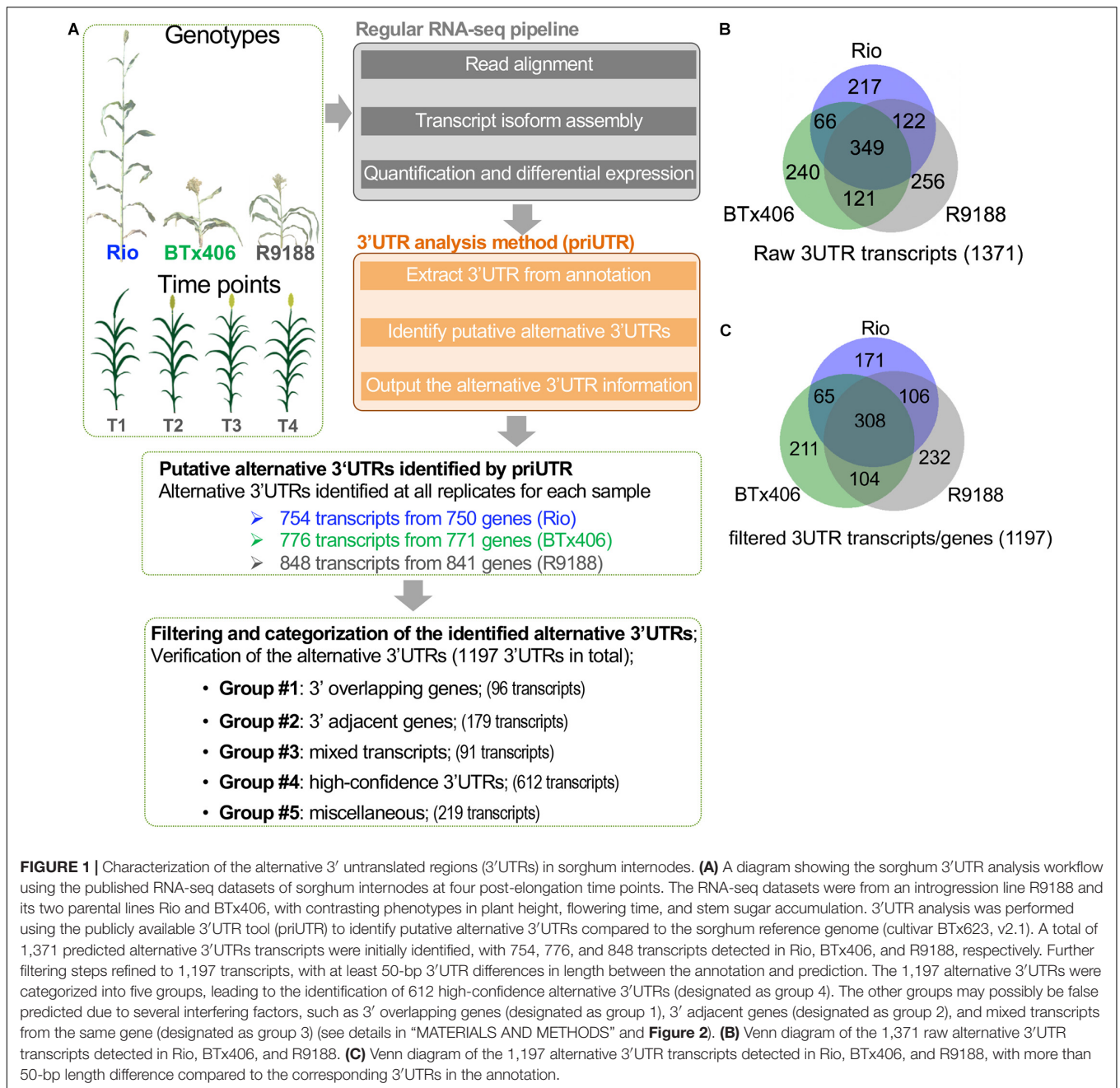
2, 3, 4, and 5 contains 96, 179, 91, 612, and 219 transcripts, respectively (Figure 1A). Examples of the genome browser views for groups 1, 2, 3, and 4 are shown in Figures 2A–D, respectively.

A comparison of the alternative 3'UTR lengths from three sorghum genotypes with those of the annotation showed significantly longer 3'UTRs in our results (Figure 2F). Moreover, a comparison of the alternative 3'UTR lengths between the five groups indicated that groups 1 and 2 have longer alternative 3'UTRs than the other groups, matching with our findings that many of them are likely false prediction due to the inclusion of transcribed neighboring genes (Figure 2G). Besides that, the "mixed transcript" group (group 3) has significantly higher numbers of transcripts per gene (based on the genome annotation) when compared with the remaining four groups, in agreement with the possible source of false prediction as "mixed transcripts" (Figure 2H). Additionally, a comparison of the expression levels between the five groups showed that the expression levels of the "3' overlapping" genes (group 1), but not other groups, were lower than those of the high-confidence alternative 3'UTRs (group 4), suggesting that high expression levels may not be a major reason for the high-confidence prediction of alternative 3'UTRs (Supplementary Figure S3). This manual verification step highlights that the high-confidence alternative 3'UTRs consist of a large fraction of all predictions (group 4, 612 out of 1,197) and identifies several major sources of false prediction.

Association Between Alternative 3'UTRs and RNA m⁶A Modification

Since APA and RNA m⁶A modification are two molecular phenomena that could be associated with changes in 3'UTR in plants, we therefore sought to test whether the alternative 3'UTRs identified here are associated with APA or RNA m⁶A modification.

APA is one important approach currently known to produce length variations in 3'UTR in plants. Alternative polyadenylation and/or 3'UTRs are regulated by genetic factors, environmental conditions, and developmental context. Polyadenylation (PA) is directly controlled by the CPSF complex that contains six subunits, CPSF30, CPSF73, CPSF100, CPSF160, Wdr33, and FIP1, and recognizes the polyadenylation sites on 3'UTR (Mandel et al., 2008; Shi et al., 2009). Some subunits of the PA machinery (CPSF30, CPSF100, and FIP1) have been functionally characterized in plants (Thomas et al., 2012; Lin et al., 2017; Téllez-Robledo et al., 2019). FPA is another *trans*-acting factor regulating the 3' end formation of mRNA in *Arabidopsis* (Duc et al., 2013). Besides that, abiotic stresses, such as dehydration and salt, can also induce alternative polyadenylation (Sun et al., 2017; Téllez-Robledo et al., 2019). Recently, a comprehensive profiling of APAs in rice across several developmental stages and tissues marks tissue specificity in APA patterns (Fu et al., 2016). Based on the abovementioned observation, it can be hypothesized that, if the 612 high-confidence alternative 3'UTR transcripts (group 4) use their alternative 3'UTRs are due, at least partly, to the previously studied PA machinery proteins or environmental



conditions, a significant fraction of the 612 sorghum transcripts would be the orthologs of the APA genes in *Arabidopsis* or rice which are associated with PA proteins or stresses. To test the APA hypothesis, we retrieved the APA-associated genesets (genesets A1, A2, B, D1, D2, D3, E, and G) from the above-mentioned studies in *Arabidopsis* and rice. The details of these genesets are provided in **Figure 3A** and **Supplementary Tables S2, S3**. The gene IDs in *Arabidopsis* and rice were converted to sorghum gene IDs using the established gene orthologous relationships between *Arabidopsis* and sorghum (Van Bel et al., 2018) and those between rice and sorghum (Sakai et al., 2013), respectively. Hypergeometric tests failed to reveal any significant

enrichment of the APA-associated genesets within the five groups of sorghum alternative 3'UTR genes ($P < 0.05$; **Figure 3A** and **Supplementary Figure S4**), not supporting the APA hypothesis that the alternative 3'UTR genes observed here are largely contributed by the known polyadenylation machinery proteins or the stresses.

We then analyzed whether the alternative 3'UTRs identified in sorghum could be related to RNA m⁶A modification. RNA m⁶A modification is another factor known to be related to 3'UTR, as a large portion of m⁶A sites in plants are located within mRNA 3'UTR regions (Luo et al., 2014, 2020). m⁶A is installed by m⁶A methyltransferase [known as "m⁶A writer,"

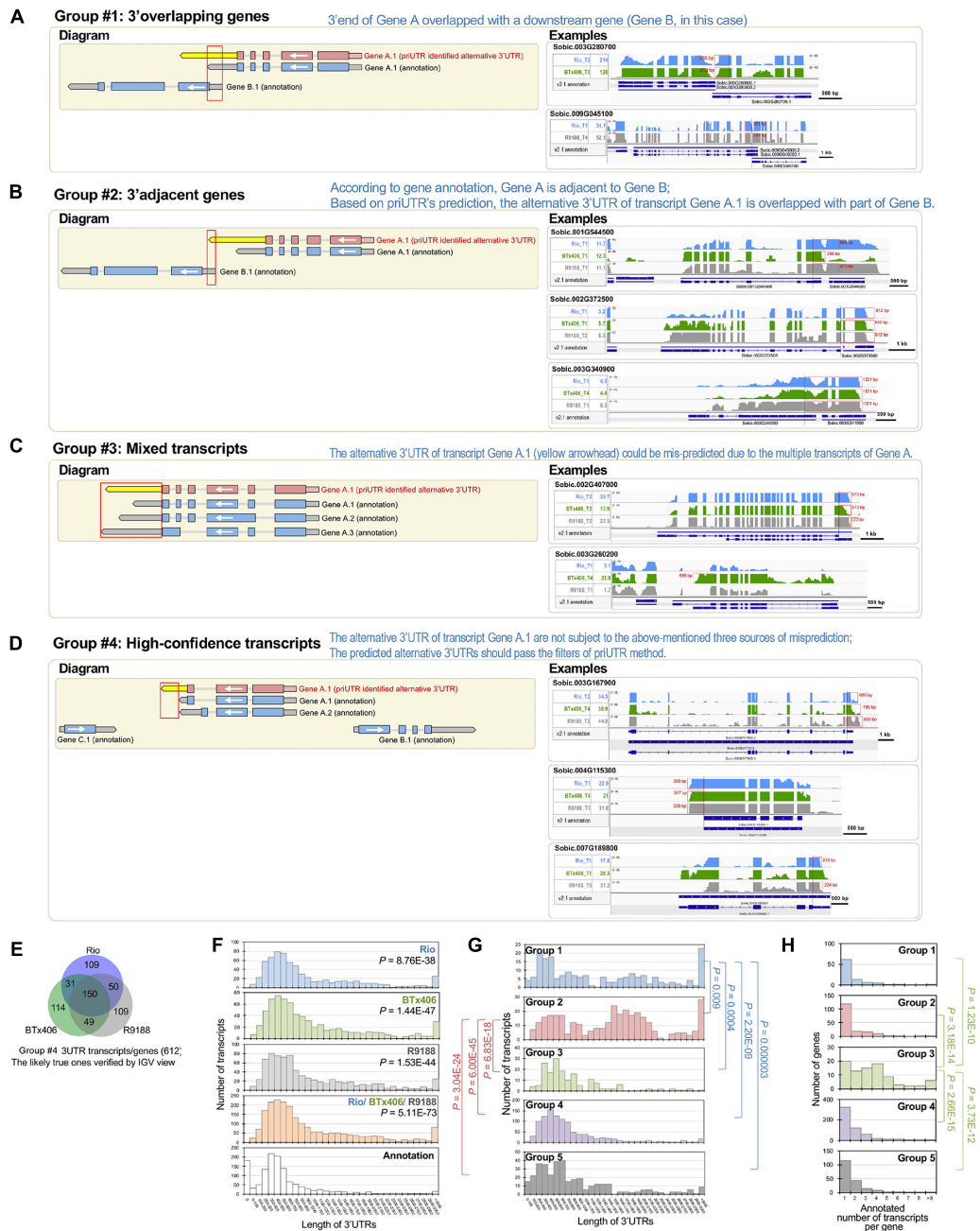


FIGURE 2 | Categorization of the predicted 3'UTRs highlights the high-confidence group of alternative 3'UTRs in sorghum. **(A–D)** Schematic diagrams showing that the predicted 3'UTRs in sorghum were categorized into groups, with groups 1, 2, 3, and 4, respectively. In the left panel of the diagrams, the white arrow indicates gene orientation. The red gene model denotes the reconstructed transcript, while the blue gene model denotes the annotated transcript, with the yellow arrow denoting the identified alternative 3'UTR. In the right panel, representative Integrative Genomics Viewer (IGV) views of the predicted alternative 3'UTRs in groups 1, 2, 3, and 4 **(A–D)**, respectively) are shown. The select IGV tracks of the genotype and time point are shown, in which the alternative 3'UTRs were detected. The gene expression levels at those samples provided (in reads per kilobase of exon per million mapped sequence reads). The alternative 3'UTRs are highlighted in red boxes, with their length labeled. **(E)** Venn diagram of the 612 high-confidence alternative 3'UTR transcripts. **(F)** Histograms showing the length of the high-confidence alternative 3'UTRs detected in Rio, BTx406, and R9188, all of the three genotypes and their corresponding annotations. The results indicated that the identified alternative 3'UTRs tend to be longer than the annotations (Wilcoxon rank sum test, $P < 0.05$). **(G)** The histograms of the alternative 3'UTR length of the five groups demonstrated that groups 1 and 2 differed significantly in 3'UTR length from the other groups (Wilcoxon rank sum test, $P < 0.05$), while the length distribution of 3'UTRs was not different among groups 3, 4, and 5. Longer 3'UTRs in groups 1 and 2 are consistent with the classification that the alternative 3'UTRs in groups 1 and 2 could be mostly mis-detected due to the 3' downstream genes overlapping with or adjacent to the 3' ends of the alternative 3'UTR genes, respectively. **(H)** Histograms of the number of transcripts per gene for the five groups showed that group 3 has significantly more transcripts per gene than the other groups (Wilcoxon rank sum test, $P < 0.05$), matching the likely source of false positive for group 3 that the alternative 3'UTRs could be mis-predicted by mixing multiple 3'UTRs of the transcripts from the same gene.

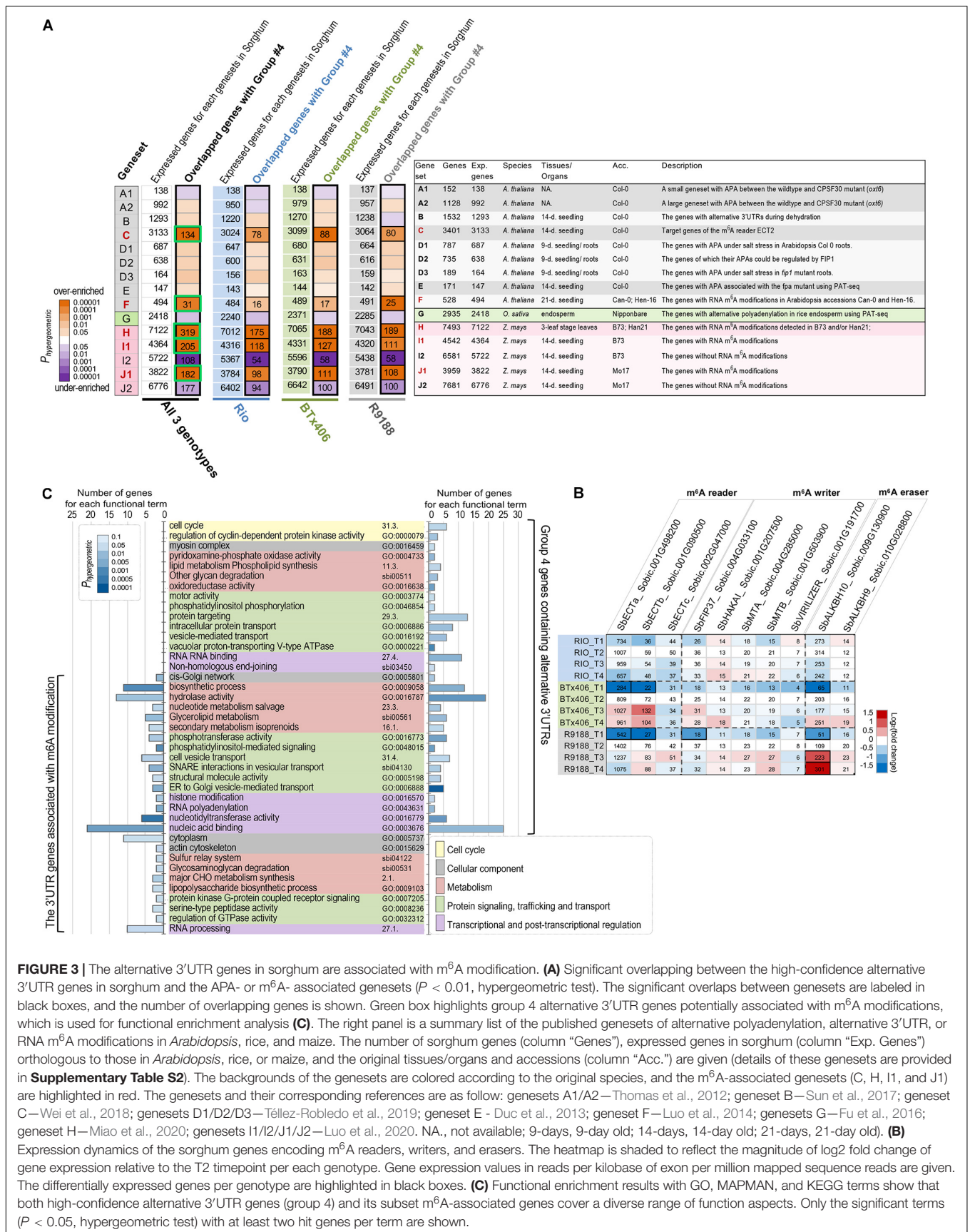


FIGURE 3 | The alternative 3'UTR genes in sorghum are associated with m⁶A modification. **(A)** Significant overlapping between the high-confidence alternative 3'UTR genes in sorghum and the APA- or m⁶A- associated genesets ($P < 0.01$, hypergeometric test). The significant overlaps between genesets are labeled in black boxes, and the number of overlapping genes is shown. Green box highlights group 4 alternative 3'UTR genes potentially associated with m⁶A modifications, which is used for functional enrichment analysis **(C)**. The right panel is a summary list of the published genesets of alternative polyadenylation, alternative 3'UTR, or RNA m⁶A modifications in *Arabidopsis*, rice, and maize. The number of sorghum genes (column "Genes"), expressed genes in sorghum (column "Exp. Genes") orthologous to those in *Arabidopsis*, rice, or maize, and the original tissues/organs and accessions (column "Acc.") are given (details of these genesets are provided in **Supplementary Table S2**). The backgrounds of the genesets are colored according to the original species, and the m⁶A-associated genesets (C, H, I1, and J1) are highlighted in red. The genesets and their corresponding references are as follow: genesets A1/A2—Thomas et al., 2012; geneset B—Sun et al., 2017; geneset C—Wei et al., 2018; genesets D1/D2/D3—Télez-Robledo et al., 2019; geneset E—Duc et al., 2013; geneset F—Luo et al., 2014; genesets G—Fu et al., 2016; geneset H—Miao et al., 2020; genesets I1/I2/J1/J2—Luo et al., 2020. NA., not available; 9-days, 9-day old; 14-days, 14-day old; 21-days, 21-day old. **(B)** Expression dynamics of the sorghum genes encoding m⁶A readers, writers, and erasers. The heatmap is shaded to reflect the magnitude of log₂ fold change of gene expression relative to the T2 timepoint per each genotype. Gene expression values in reads per kilobase of exon per million mapped sequence reads are given. The differentially expressed genes per genotype are highlighted in black boxes. **(C)** Functional enrichment results with GO, MAPMAN, and KEGG terms show that both high-confidence alternative 3'UTR genes (group 4) and its subset m⁶A-associated genes cover a diverse range of function aspects. Only the significant terms ($P < 0.05$, hypergeometric test) with at least two hit genes per term are shown.

e.g., methyltransferase-like 3 (METTL3), METTL14, Wilms tumor 1-associated protein (WTAP); Liu et al., 2014; Wang Y. et al., 2014; Ping et al., 2014], recognized by “m⁶A reader” (EVOLUTIONARILY CONSERVED C-TERMINAL REGION proteins, ECT2, 3, 4 in plants; Wei et al., 2018) and removed by “eraser” m⁶A RNA demethylase [e.g., alkylated DNA repair protein AlkB homolog 5 (ALKBH5); Zheng et al., 2013]. Previously, mRNA m⁶A methylome studies in *Arabidopsis* (Luo et al., 2014), rice (Zhang et al., 2019), and maize (Luo et al., 2020; Miao et al., 2020) reveal a functional association of m⁶A with chloroplast, sporogenesis, stress response, and translational status, highlighting new functions of m⁶A functional factors, the effects of m⁶A modification, and m⁶A inter-variety variation. In the present study, if the 612 alternative 3'UTRs in sorghum are associated with or regulated by m⁶A modification, a significant fraction of the 612 transcripts would be orthologs of the genes with m⁶A modifications in *Arabidopsis* or maize (designated as the m⁶A hypothesis). To investigate the hypothesis, several m⁶A-associated genesets were obtained from the m⁶A-modified genes reported in *Arabidopsis* (genesets C and F) and maize (geneset H, I1, and J1; **Figure 3A**; Luo et al., 2014, 2020; Miao et al., 2020). The genesets I2 and J2 exclusively contain the expressed genes without m⁶A modification in maize seedlings, serving as the controls for genesets I1 and J1, respectively (Luo et al., 2020). Our analysis clearly showed that the high-confidence alternative 3'UTR genes (group 4) are significantly over-enriched with the orthologs of m⁶A-associated genesets in both *Arabidopsis* and maize but are under-enriched with the m⁶A control genesets (I2 and J2) ($P_{\text{hypergeometric}} < 0.01$; **Figure 3A**). For example, many genes which are orthologous targets of m⁶A reader ECT2 in *Arabidopsis* were significantly over-enriched in group 4. The hypergeometric tests were first performed using all of the 612 alternative 3'UTR genes pooled from three sorghum genotypes. We further repeated the analysis in each of the three sorghum lines to avoid potential statistical artifacts. The results for each sorghum genotype confirmed the over-enrichment of m⁶A-associated genes (**Figure 3A** and **Supplementary Tables S4, S5**). The statistical significance of the enrichment was further assessed using 400 permutation tests for each of the three sorghum genotypes. In group 4, there are 340, 344, and 358 genes from Rio, BTx406, and R9188, respectively. To avoid potential influences of expression levels in the permutation tests, we examined the distribution of expression levels of the 612 alternative 3'UTR genes within each sorghum genotype (**Supplementary Table S5**). The deciles of the expression levels of group 4 genes were determined in Rio, BTx406, and R9188, respectively, and used to randomly select 36 expressed genes within each expression decile per genotype from all of the expressed genes, yielding a set of 360 random genes with a distribution of expression levels similar to that of the group 4 genes. The results of permutation tests showed that both the number of overlapped genes and the hypergeometric *p*-values were significantly correlated with m⁶A genesets C, F, H, I1, and J1 (**Supplementary Figures S5, S6**). In contrast, the number of overlapped genes and the hypergeometric *p*-values were significantly smaller than the random distributions for m⁶A control genesets I2 and J2 (**Supplementary Figures S5, S6**), supporting a significant under-enrichment of non-m⁶A

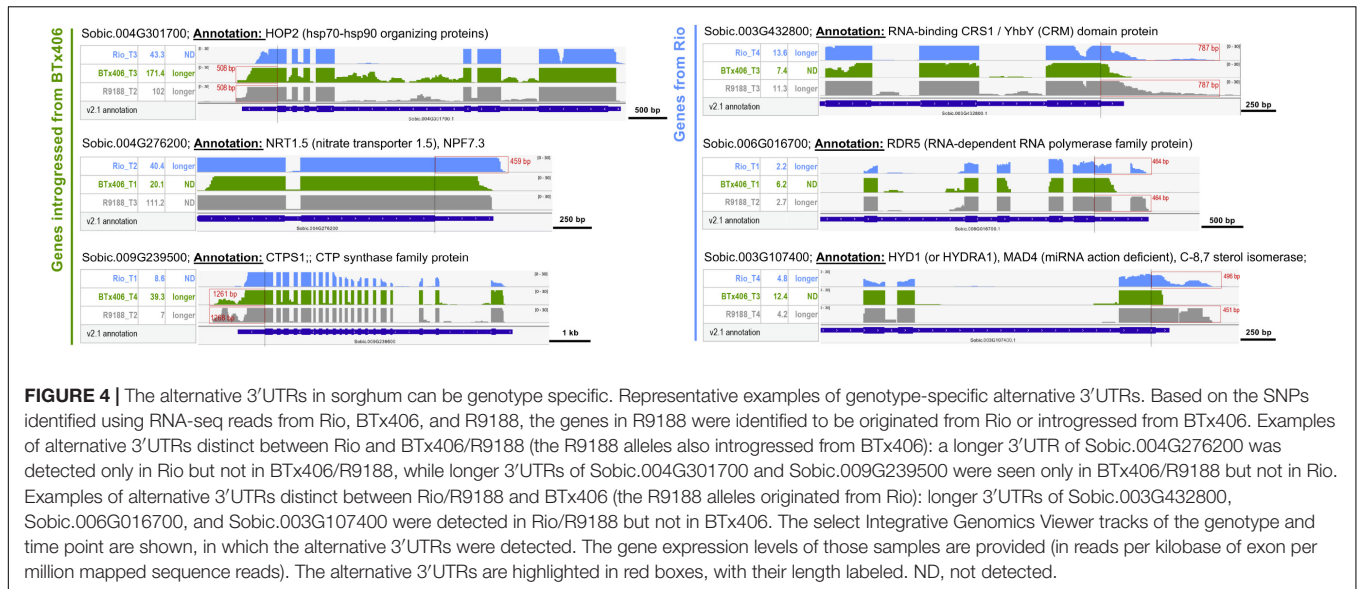
modified genes in the group 4 genes. Overall, the analysis between genesets is suggestive of the m⁶A hypothesis.

To ascertain m⁶A occurrence in sorghum internodes, we found that multiple genes putatively encoding m⁶A reader, writer, and eraser, respectively, were expressed in the RNA-seq samples, with many of them expressed at high levels. The genes encoding m⁶A readers (Sobic.001G498200 and Sobic.001G090500) and eraser (Sobic.009G130900) were upregulated during post-anthesis stages in BTx406 and R9188 (**Figure 3B**).

To portrait the representative molecular functions, biological processes, and metabolic pathways, functional enrichment was performed for the alternative 3'UTR genes and the subset of 3'UTR genes associated with m⁶A modifications. The subset of m⁶A-associated 3'UTR genes was generated by combining the overlapping genes across several m⁶A genesets (green boxes in **Figure 3A**, a total of 398 genes). **Figure 3C** reveals that both the alternative 3'UTR genes and the m⁶A-associated 3'UTR genes cover a diverse set of functions, which can be grouped into five major categories, including cell cycle, cellular components, metabolism, protein signaling, trafficking and transport, and transcriptional and post-transcriptional regulation. The enrichment results of these two genesets highlight functions related to metabolism (*i.e.*, biosynthetic process, hydrolase activity) and nucleic acid regulation (*i.e.*, nucleic acid binding and RNA processing). Moreover, we found that the sorghum m⁶A-associated 3'UTR genes are highly enriched in RNA processing, nucleic acid binding, as well as functions associated with protein localization and transportation (for example, cell vesicle transport, SNARE in vesicle transport, *cis*-Golgi network) (**Figure 3C**), consistent with the representative functions of m⁶A-modified genes previously seen in *Arabidopsis* and maize (Luo et al., 2014, 2020).

Genotype-Specific Alternative 3'UTRs

Utilizing the genetic relationship among Rio, BTx406, and R9188, we sought to address whether alternative 3'UTR can be genotype-specific. The RNA-seq read mapping data of the 612 alternative 3'UTRs were compared between the three genotypes. Twenty-three genes were identified to have distinct alternative 3'UTRs either between Rio and BTx406/R9188 (seven genes, **Supplementary Figure S7A**) or between Rio/R9188 and BTx406 (16 genes, **Supplementary Figures S7B,C**). Many of them (19 out of 23) are homologous to the genes with known annotations or proven important functions in *Arabidopsis*. The genotype-specific alternative 3'UTRs comprise only a small fraction of the identified alternative 3'UTRs (~3.7%, 23 out of 612). To further validate the parental origins of these 23 alternative 3'UTR genes in R9188, we utilized the 1,805 genes that are introgressed from BTx406 into R9188 (Li et al., 2019a). All of the seven genes in R9188 with identical 3'UTR length to BTx406 are indeed introgressed from BTx406, while among the 16 R9188 genes with identical 3'UTR length to Rio, eight are validated to be originated from Rio, with the remaining eight genes lacking SNPs for validation (**Supplementary Figure S7**). The representative genes in R9188 that have genotype-specific alternative 3'UTRs are shown in **Figure 4**. Many homologs of the genes identified here are critical for plant development.



For example, Sobic.004G276200 is a homolog of AtNRT1.5 (AT1G32450) that regulates root architecture and leaf senescence through nitrate response and root-to-shoot nitrate transport and potassium translocation (Meng et al., 2015; Zheng et al., 2016). Sobic.006G016700 is homologous to AtHYD1 (AT1G20050), a sterol biosynthetic gene that affects cell wall synthesis, phytohormone signaling, and miRNA activity, crucial for embryo and root development (Schrack et al., 2004; Souter et al., 2004; Brodersen et al., 2012).

DISCUSSION

Our work exemplifies a transcriptome-wide 3'UTR analysis using conventional RNA-seq data in non-model plants, identifying several hundred genes with high-confidence alternative 3'UTRs in sorghum. While many bioinformatic tools for identifying mRNA 3' ends have been reported, the priUTR program was used here to analyze sorghum 3'UTRs for several reasons (Tu, 2020). First, the genome assembly and annotation files of non-model crop species, such as maize and sorghum, can be readily used as input files of priUTR (e.g., gff3 annotation), while some other programs appear to lack clear guidelines for preparing input files using non-model plant species. Second, results from the priUTR tool should be considered largely reliable, since priUTR resembles a previous online tool for detecting alternative 3'UTRs, 3USS, in the principles of the algorithms (Le Pera et al., 2015). priUTR adopts the idea of intron-chain matching from 3USS, which required the transcript reconstruction from Cufflinks or Scripture as one of the input files and predicted alternative 3'UTRs once all of the intron chain could be matched between the assembled and annotated transcript. However, 3USS provided alternative 3'UTR prediction for model animals and *Arabidopsis* and is no longer operational. Unlike 3USS, priUTR utilizes partial intron-chain matching algorithm, in which three consecutive introns before the stop codon are required to be matched and holds the flexibility that

the assembled transcript could have alternative splicing in its 5' part or have additional unannotated introns in the predicted 3'UTR regions (Supplementary Figure S1). Third, the priUTR tool could be used for analyzing multiple RNA-seq samples by program loop or parallelism. Fourth, the priUTR program fits our computer systems (Scientific Linux release 6.10 Carbon), while other programs might not be readily compatible with our operational system.

Extensive bioinformatics efforts have been made to identify poly(A) sites or mRNA 3' tails using RNA-seq data. These previous bioinformatic tools have been classified into four groups based on their principles of the algorithms (Chen et al., 2020). Some of the programs have been compared using benchmarking datasets from model species (i.e., human, mouse, and *Arabidopsis*). The advantages and the limitations for each type of poly(A)-site prediction tools have been well discussed. Type 1 tools depend on a prior database of poly(A) sites in model animals or *Arabidopsis* and can effectively identify numerous poly(A) sites. Type 1 tools are limited to detect previously known poly(A) sites. Type 2 programs apply relatively mature software for transcriptome assembly, such as Scripture (Guttman et al., 2010), Cufflinks (Trapnell et al., 2012), and StringTie (Pertea et al., 2016). While type 2 programs could detect both 3'UTR extending and shortening events, they inherit the limitations of transcript assembly tools that are: (1) failure to distinguish the transcripts which overlap or reside closely between each other in the genome and (2) inability to distinguish a short transcript isoform embedded in longer ones from the same gene. Type 3 tools heavily depend on poly(A)-capped reads, which are scarce in regular RNA-seq data (Bayerlova et al., 2015; Kim et al., 2015), leading to low sensitivity of these methods. Type 4 programs detect sudden fluctuations of read density at the 3' end of mRNAs in order to model 3'UTR usage. Type 4 programs tend to identify 3'UTR shortening events and have more inaccurate predictions caused by the heterogeneity of read coverage and non-biological variations (Szkop and Nobeli, 2017). Based on the information above, the type 2 methods which rely on transcript reconstruction

and infer 3'UTR from assembled transcripts may be a potentially better choice for alternative 3'UTR study for non-model plant species with reference genome available.

As a type 2 bioinformatic tool for alternative 3'UTR prediction, priUTR has some limitations. It heavily relies on the accuracy of transcript annotation, and the accuracy of RNA-seq-based transcript reconstruction depends on the genic context. When multiple transcripts are produced from a gene model, of which the transcript with a shorter 3'UTR sequence embedded within a longer transcript, transcript reconstructions are prone to be inaccurate. Such cases have been identified in our study as the group 3 alternative 3'UTRs (**Figure 2**). In other situations, when several gene models are located closely in the same genomic locus, the prediction of 3'UTR extension tends to be false because the transcript could be mis-assembled to include neighboring genes. In this study, these likely mis-predicted alternative 3'UTRs were categorized as groups 1 and 2 (**Figure 2**). Another limitation of priUTR is that it discards those transcripts with less than 3 introns (including the intronless ones) since the transcript do not meet the requirement for partial intron-chain matching. This would limit the number of predicted alternative 3'UTRs. Besides that, priUTR is not applicable to non-model organisms without reference genomes. When priUTR is applied to an organism with a reference genome and poor annotations, caution should be taken as predicted alternative 3'UTRs could be previously unannotated 3'UTR regions.

Thanks to the identification of potential sources of mis-predicted alternative 3'UTRs, several improvements could be made to the priUTR tool in the near future. First, filters could be made to discard genes with multiple overlapped transcripts. Second, filters could be applied to discard those genes with neighbor genes located at 3' downstream to avoid mis-assembly of transcript. Third, methods could be established to identify alternative 3'UTR transcripts with less than three introns.

Besides the application of priUTR method for characterizing alternative 3'UTRs in sorghum, we attempted to gain insights into the factors associated with the alternative 3'UTRs. Our results strongly indicate an association between m⁶A modification and alternative 3'UTR (**Figure 3**). Recently, a maize study of RNA m⁶A profiling reveals a clear indication between the m⁶A modification and APA (Luo et al., 2020). There are both similarities and differences between the results of the maize m⁶A study and the results of the present alternative 3'UTR study. In the maize study, genes were split into two sets, one set of genes with single poly(A) site and the other set of genes that contain at least two poly(A) sites and could have APA events. With such an approach for dividing the maize expressed genes, significant associations between potential APA genes and m⁶A modified genes were observed (Luo et al., 2020). The maize study was a genome-wide, large-scale profiling of m⁶A methylated mRNAs without actually determining the mRNAs with alternative 3'UTRs or poly(A) sites. In our sorghum study, the association between APA and m⁶A methylation was discovered the other way around. We first identified the genes using alternative 3'UTRs and then discovered that many of the alternative 3'UTR genes could carry m⁶A methylation. Another feature of our results is that a few hundreds of alternative

3'UTR genes were identified in sorghum, with the potential to carry m⁶A modification. Together with enriched functions, our results could help to narrow down candidate genes for further functional studies with a focus on the cause-and-effect relationship between m⁶A methylation and APA for specific functionally important genes. In addition, our results provide pieces of evidence that alternative 3'UTR usage for some genes (many functionally important) are genotype specific and could be inherent genetically.

For the m⁶A hypothesis, public data from maize studies have been used (Luo et al., 2020; Miao et al., 2020). Maize and sorghum are close relatives, split just 11.9 million years ago (Mya) from a common progenitor (Swigonova et al., 2004). After the divergence, maize experienced allotetraploidization followed by diploidization (Gaut and Doebley, 1997; Xu and Messing, 2008). Therefore, the orthologous relationship between maize and sorghum genes is well established, and many gene regulations remain conserved (Zhang et al., 2017). Such conservation could help to establish an association between RNA m⁶A modification and alternative 3'UTR. For the APA hypothesis, it remains possible that PA machinery proteins, stress responses, or developmental context might be involved in alternative 3'UTRs in sorghum. *Arabidopsis* and rice are plant model species divergent from sorghum. The authentic results of APA or alternative 3'UTRs related to stresses have only been reported by a handful of studies in plants (Sun et al., 2017; Téllez-Robledo et al., 2019). As far as our knowledge is concerned, the target genes of PA machinery proteins have only been studied in *Arabidopsis* (Thomas et al., 2012; Duc et al., 2013; Lin et al., 2017; Téllez-Robledo et al., 2019). Whether or to what extent these APA genes are evolutionarily conserved between *Arabidopsis* and sorghum remains to be determined. Thus, one of the possible explanations for the non-significant results between APA-associated genesets and the sorghum alternative 3'UTR genes could be that the APA genes that respond to stresses or are induced by mutations of PA proteins could be largely unconserved between sorghum and *Arabidopsis*. Overall, our results clearly lead to important testable hypotheses that: (1) m⁶A modifications are functionally related to alternative 3'UTR/APA, yet the cause-and-effect relationship between them needs clarification, and (2) whether the genes with inter-genotype variations in alternative 3'UTR/APA could have functional consequences in sorghum; if yes, what are the consequences in molecular and phenotypic levels.

While conventional RNA-seq studies are prevalent in sorghum, studies on alternative 3'UTR/APA are scarce in non-model plant species. Previously, the full-length transcriptome of sorghum seedling was characterized by PacBio Iso-Seq and identified a few thousand transcripts with APA, demonstrating APA as a common phenomenon (Abdel-Ghany et al., 2016). We acknowledge that the priUTR-based 3'UTR analysis has limitations in the number of alternative 3'UTRs identified and the mis-predictions compared to those specialized pipelines based on the more advanced long-reads sequencing technologies (*i.e.*, PacBio and Oxford Nanopore) (Abdel-Ghany et al., 2016; Parker et al., 2020). These limitations are partly due to the short-read

derived transcriptome assembly and to the limited ability in detecting transcripts with large variations using reference-guided mapping (Stark et al., 2019). Recently, new aspects of mRNA 3' ends have been revealed to affect post-transcriptional regulation, including poly(A) length and sequence composition (Zhao et al., 2019). This suggests that a long-read-based method to integrate profiles of alternative 3'UTR/APA, poly(A) length and sequence, and m⁶A modifications will be a powerful tool to understand 3'UTR-mediated regulation in the near future.

In summary, our work presents a transcriptome-wide profiling of alternative 3'UTRs in sorghum and identified hundreds of genes as the candidate genes to study the functional effects of alternative 3'UTR usage. The new insights reported here suggest future research directions for 3'UTR genotype specificity and the link between 3'UTR and epi-transcriptome modifications. Additionally, this study exemplifies alternative 3'UTR analysis using conventional RNA-seq, signifying 3'UTR analysis as a valuable addition to routine RNA-seq analysis in plants.

DATA AVAILABILITY STATEMENT

The RNA-seq data of sorghum used in this study can be found at NCBI Sequence Read Archive (SRA) under accession PRJNA413691. The priUTR program for 3'UTR analysis is available at github (<https://github.com/mint1234/3UTR->). All data generated or analyzed during this study are included in this published article and its **Supplementary Material**.

AUTHOR CONTRIBUTIONS

MT and YL designed the study, performed the data analysis, and revised and finalized the manuscript. MT generated the data. YL wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Waksman Genomics Facility and Waksman Institute of Microbiology.

REFERENCES

- Abdel-Ghany, S. E., Hamilton, M., Jacobi, J., Ngam, P., Devitt, N., Schilkey, F., et al. (2016). A survey of the sorghum transcriptome using single-molecule long reads. *Nat. Commun.* 7:11706.
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11:R106.
- Areifeen, A., Liu, J., Xiao, X., and Jiang, T. (2018). TAPAS: tool for alternative polyadenylation site analysis. *Bioinformatics* 34, 2521–2529. doi: 10.1093/bioinformatics/bty110
- Arribas-Hernández, L., Bressendorff, S., Hansen, M. H., Poulsen, C., Erdmann, S., and Brodersen, P. (2018). An m(6)A-YTH module controls developmental timing and morphogenesis in Arabidopsis. *Plant Cell* 30, 952–967. doi: 10.1105/tpc.17.00833

ACKNOWLEDGMENTS

We thank Dr. Yaping Feng and Dr. Dibyendu Kumar at Waksman Institute of Microbiology, Rutgers University, for their contributions in the project of developing 3' untranslated region analysis program, priUTR.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.556749/full#supplementary-material>

Supplementary Figure S1 | Schematic diagram of the priUTR algorithm.

Supplementary Figure S2 | Distribution of the expression levels for the five groups of predicted alternative 3'UTR genes.

Supplementary Figure S3 | Comparison of groups 1, 2, 3, and 5 of sorghum alternative 3'UTR genes with the APA- or m⁶A- associated genesets in *Arabidopsis*, rice, and maize.

Supplementary Figure S4 | The distribution of the number of genes overlapped between the public genesets and the randomly selected genes in sorghum by using a permutation test.

Supplementary Figure S5 | The distribution of hypergeometric *p*-values between the public genesets and the randomly selected genes in sorghum by using a permutation test.

Supplementary Figure S6 | Genome browser views of the 23 genes with genotype-specific alternative 3'UTRs.

Supplementary Table S1 | Information about the 612 high-confidence alternative 3'UTR genes.

Supplementary Table S2 | Summary information about the public available genesets associated with m⁶A and APA.

Supplementary Table S3 | Genesets for m⁶A and APA.

Supplementary Table S4 | Hypergeometric tests for the overlap between sorghum alternative 3'UTR genes and the APA- or m⁶A- associated genesets.

Supplementary Table S5 | Distribution of the expression levels of the high-confidence alternative 3'UTR genes in Rio, BTx406, and R9188, respectively.

Supplementary Table S6 | Functional enrichment results of group 4 genes and the m⁶A-associated 3'UTR subset.

Supplementary Table S7 | Information about the 23 genes with genotype-specific alternative 3'UTR.

- Bayerlova, M., Klemm, F., Kramer, F., Pukrop, T., Beibarth, T., Bleckmann, A., et al. (2015). Newly constructed network models of different WNT signaling cascades applied to breast cancer expression data. *PLoS One* 10:e0144014. doi: 10.1371/journal.pone.0144014
- Bell, S. A., Shen, C., Brown, A., and Hunt, A. G. (2016). Experimental genome-wide determination of RNA polyadenylation in *Chlamydomonas reinhardtii*. *PLoS One* 11:e0146107. doi: 10.1371/journal.pone.0146107.g007
- Birol, I., Raymond, A., Chiu, R., Nip, K. M., Jackman, S. D., Kreitzman, M., et al. (2015). Kleat: cleavage site analysis of transcriptomes. *Pac. Symp. Biocomput.* 2015, 347–358.
- Bonfert, T., and Friedel, C. C. (2017). Prediction of poly(A) sites by poly(A) read mapping. *PLoS One* 12:e0170914. doi: 10.1371/journal.pone.0170914
- Boyles, R. E., Brenton, Z. W., and Kresovich, S. (2019). Genetic and genomic resources of sorghum to connect genotype with phenotype

- in contrasting environments. *Plant J.* 97, 19–39. doi: 10.1111/tj.14113
- Brodersen, P., Sakvarelidze-Achard, L., Schaller, H., Khafif, M., Schott, G., Bendahmane, A., et al. (2012). Isoprenoid biosynthesis is required for miRNA function and affects membrane association of ARGONAUTE 1 in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1778–1783. doi: 10.1073/pnas.1112501109
- Calvino, M., and Messing, J. (2012). Sweet sorghum as a model system for bioenergy crops. *Curr. Opin. Biotech.* 23, 323–329. doi: 10.1016/j.copbio.2011.12.002
- Calvino, M., Bruggmann, R., and Messing, J. (2011). Characterization of the small RNA component of the transcriptome from grain and sweet sorghum stems. *BMC Genomics* 12:356. doi: 10.1186/1471-2164-12-356
- Chakrabarti, M., and Hunt, A. G. (2015). CPSF30 at the interface of alternative polyadenylation and cellular signaling in plants. *Biomolecules* 5, 1151–1168. doi: 10.3390/biom5021151
- Chen, M., Ji, G., Fu, H., Lin, Q., Ye, C., Ye, W., et al. (2020). A survey on identification and quantification of alternative polyadenylation sites from RNA-seq data. *Brief. Bioinform.* 21, 1261–1276. doi: 10.1093/bib/bbz068
- Cooper, E. A., Brenton, Z. W., Flinn, B. S., Jenkins, J., Shu, S., Flowers, D., et al. (2019). A new reference genome for Sorghum bicolor reveals high levels of sequence similarity between sweet and grain genotypes: implications for the genetics of sugar metabolism. *BMC Genomics* 20:420. doi: 10.1186/s12864-019-5734-x
- Davidson, R. M., Gowda, M., Moghe, G., Lin, H., Vaillancourt, B., Shiu, S. H., et al. (2012). Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. *Plant J.* 71, 492–502.
- Deschamps, S., Zhang, Y., Llaca, V., Ye, L., Sanyal, A., King, M., et al. (2018). A chromosome-scale assembly of the sorghum genome using nanopore sequencing and optical mapping. *Nat. Commun.* 9:4844.
- Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. (2010). AgriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* 38, W64–W70.
- Duan, H. C., Wei, L. H., Zhang, C., Wang, Y., Chen, L., Lu, Z., et al. (2017). ALKBH10B is an RNA N6-methyladenosine demethylase affecting Arabidopsis floral transition. *Plant Cell* 29, 2995–3011. doi: 10.1105/tpc.16.00912
- Duc, C., Sherstnev, A., Cole, C., Barton, G. J., and Simpson, G. G. (2013). Transcription termination and chimeric RNA formation controlled by Arabidopsis thaliana FPA. *PLoS Genet.* 9:e1003867. doi: 10.1186/1471-2164-12-514
- Dugas, D. V., Monaco, M. K., Olson, A., Klein, R. R., Kumar, S., Ware, D., et al. (2011). Functional annotation of the transcriptome of Sorghum bicolor in response to osmotic stress and abscisic acid. *BMC Genomics* 12:514.
- Fracasso, A., Trindade, L. M., and Amaducci, S. (2016). Drought stress tolerance strategies revealed by RNA-Seq in two sorghum genotypes with contrasting WUE. *BMC Plant Biol.* 16:115. doi: 10.1186/s12870-016-0800-x
- Fu, H., Yang, D., Su, W., Ma, L., Shen, Y., Ji, G., et al. (2016). Genome-wide dynamics of alternative polyadenylation in rice. *Genome Res.* 26, 1753–1760. doi: 10.1101/gr.210757.116
- Fu, Y., Sun, Y., Li, Y., Li, J., Rao, X., Chen, C., et al. (2011). Differential genome-wide profiling of tandem 3' UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Res.* 21, 741–747. doi: 10.1101/gr.115295.110
- Gaut, B. S., and Doebley, J. F. (1997). DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc. Natl. Acad. Sci. U.S.A.* 94, 6809–6814. doi: 10.1073/pnas.94.13.6809
- Gelli, M., Duo, Y., Konda, A. R., Zhang, C., Holding, D., and Dweikat, I. (2014). Identification of differentially expressed genes between sorghum genotypes with contrasting nitrogen stress tolerance by genome-wide transcriptional profiling. *BMC Genomics* 15:179. doi: 10.1186/1471-2164-15-179
- Gruber, A. J., Schmidt, R., Ghosh, S., Martin, G., Gruber, A. R., van Nimwegen, E., et al. (2018). Discovery of physiological and cancer-related regulators of 3' UTR processing with KAPAC. *Genome Biol.* 19, 44.
- Gruber, A. J., Schmidt, R., Gruber, A. R., Martin, G., Ghosh, S., Belmadani, M., et al. (2016). A comprehensive analysis of 3' end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. *Genome Res.* 26, 1145–1159. doi: 10.1101/gr.202432.115
- Guttman, M., Garber, M., Levin, J. Z., Donaghey, J., Robinson, J., Adiconis, X., et al. (2010). Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* 28, 503–510. doi: 10.1038/nbt.1633
- Ha, K. C. H., Blencowe, B. J., and Morris, Q. (2018). QAPA: a new method for the systematic analysis of alternative polyadenylation from RNA-seq data. *Genome Biol.* 19:45.
- Hafez, D., Ni, T., Mukherjee, S., Zhu, J., and Ohler, W. (2013). Genome-wide identification and predictive modeling of tissue-specific alternative polyadenylation. *Bioinformatics* 29, i108–i116.
- Hennet, L., Berger, A., Trabanco, N., Ricciuti, E., Dufayard, J., Bocs, S., et al. (2020). Transcriptional regulation of sorghum stem composition: key players identified through co-expression gene network and comparative genomics analyses. *Front. Plant Sci.* 11:224. doi: 10.3389/fpls.2020.00224
- Hong, L., Ye, C., Lin, J., Fu, H., Wu, X., and Li, Q. Q. (2018). Alternative polyadenylation is involved in auxin-based plant growth and development. *Plant J.* 93, 246–258. doi: 10.1111/tj.13771
- Hoque, M., Ji, Z., Zheng, D. H., Luo, W., Li, W., You, B., et al. (2013). Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat. Methods* 10, 133–139. doi: 10.1038/nmeth.2288
- Huang, Z., and Teeling, E. C. (2017). ExUTR: a novel pipeline for large-scale prediction of 3'-UTR sequences from NGS data. *BMC Genomics* 18:847.
- International Wheat Genome Sequencing Consortium [IWGSC] (2008). (Shifting)the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:661.
- Jan, C. H., Friedman, R. C., Ruby, J. G., and Bartel, D. P. (2010). Formation, regulation and evolution of Caenorhabditis elegans 3' UTRs. *Nature* 469, 97–101. doi: 10.1038/nature09616
- Ji, G., Guan, J., Zeng, Y., Li, Q. Q., and Wu, X. (2014). Genome-wide identification and predictive modeling of polyadenylation sites in eukaryotes. *Brief. Bioinform.* 16, 304–313. doi: 10.1093/bib/bbu011
- Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., et al. (2017). Improved maize reference genome with single-molecule technologies. *Nature* 546, 524–527.
- Johnson, S. M., Cummins, I., Lim, F. L., Slabas, A. R., and Knight, M. R. (2015). Transcriptomic analysis comparing stay-green and senescent Sorghum bicolor lines identifies a role for proline biosynthesis in the stay-green trait. *J. Exp. Bot.* 66, 7061–7073. doi: 10.1093/jxb/erv405
- Kebrom, T. H., McKinley, B., and Mullet, J. E. (2017). Dynamics of gene expression during development and expansion of vegetative stem internodes of bioenergy sorghum. *Biotechnol. Biofuels* 10:159.
- Kim, M., You, B. H., and Nam, J. W. (2015). Global estimation of the 3' untranslated region landscape using RNA sequencing. *Methods* 83, 111–117. doi: 10.1016/j.ymeth.2015.04.011
- Le Pera, L., Mazzapioda, M., and Tramontano, A. (2015). 3USS: a web server for detecting alternative 3'UTRs from RNA-seq experiments. *Bioinformatics* 31, 1845–1847. doi: 10.1093/bioinformatics/btv035
- Lee, J. Y., Yeh, I., Park, J. Y., and Tian, B. (2007). PolyA_DB 2: mRNA polyadenylation sites in vertebrate genes. *Nucl. Acids Res.* 35, D165–D168.
- Leiboff, S., and Hake, S. (2019). Reconstructing the transcriptional ontogeny of maize and sorghum supports an inverse hourglass model of inflorescence development. *Curr. Biol.* 20, 3410–3419. doi: 10.1016/j.cub.2019.08.044
- Li, X., Nair, A., Wang, S., and Wang, L. (2015). “Quality Control of RNA-Seq Experiments,” in *RNA Bioinformatics in Methods in Molecular Biology*, Vol. 1269, ed. E. Picardi (New York, NY: Humana Press), 137–146. doi: 10.1007/978-1-4939-2291-8_8
- Li, Y., Mehta, R., and Messing, J. (2018). A new high-throughput assay for determining soluble sugar in sorghum internode-extracted juice. *Planta* 248, 785–793. doi: 10.1007/s00425-018-2932-8
- Li, Y., Tu, M., Feng, Y., Wang, W., and Messing, J. (2019b). Common metabolic networks contribute to carbon sink strength of sorghum internodes: implications for bioenergy improvement. *Biotechnol. Biofuels* 12:274.
- Li, Y., Wang, W., Feng, Y., Tu, M., Wittich, P. E., Bate, N. J., et al. (2019a). Transcriptome and metabolome reveal distinct carbon allocation patterns during internode sugar accumulation in different sorghum genotypes. *Plant Biotech. J.* 17, 472–487. doi: 10.1111/pbi.12991

- Lin, J., Xu, R., Wu, X., Shen, Y., and Li, Q. Q. (2017). Role of cleavage and polyadenylation specificity factor 100: anchoring poly(A) sites and modulating transcription termination. *Plant J.* 91, 829–839. doi: 10.1111/tbj.13611
- Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* 10, 93–95. doi: 10.1038/nchembio.1432
- Lorenzo, L., Sorenson, R., Bailey-Serres, J., and Hunt, A. G. (2017). Noncanonical alternative polyadenylation contributes to gene regulation in response to hypoxia. *Plant Cell* 29, 1262–1277. doi: 10.1105/tpc.16.00746
- Lu, J., and Bushel, P. R. (2013). Dynamic expression of 3' UTRs revealed by Poisson hidden Markov modeling of RNA-Seq: implications in gene expression profiling. *Gene* 527, 616–623. doi: 10.1016/j.gene.2013.06.052
- Luo, G. Z., MacQueen, A., Zheng, G., Duan, H., Dore, L. C., Lu, Z., et al. (2014). Unique features of the m6A methylome in Arabidopsis thaliana. *Nat. Commun.* 5, 5630.
- Luo, J. H., Wang, Y., Wang, M., Zhang, L. Y., Peng, H. R., Zhou, Y. Y., et al. (2020). Natural variation in RNA m6A methylation and its relationship with translational status. *Plant Physiol.* 182, 332–344. doi: 10.1104/pp.19.00987
- Mandel, C. R., Bai, Y., and Tong, L. (2008). Protein factors in pre-mRNA 3'-end processing. *Cell Mol. Life Sci.* 65, 1099–1122. doi: 10.1007/s00018-007-7474-3
- Mangone, M., Manoharan, A. P., Thierry-Mieg, D., Thierry-Mieg, J., Han, T., Mackowiak, M., et al. (2010). The landscape of *C. elegans* 3' UTRs. *Science* 329, 432–435.
- Mathur, S., Umakanth, A. V., Tonapi, V. A., Sharma, R., and Sharma, M. (2017). Sweet sorghum as biofuel feedstock: recent advances and available resources. *Biotechnol. Biofuels.* 10:146.
- McCarthy, D. J., Chen, Y., and Smyth, K. G. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucl. Acids Res.* 40, 4288–4297. doi: 10.1093/nar/gks042
- McCormick, R. F., Truong, F. K., Sreedasyam, A., Jenkins, J., Shu, S., Sims, D., et al. (2018). The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.* 93, 338–354. doi: 10.1111/tbj.13781
- McKinley, B., Casto, A., Rooney, W., and Mullet, J. (2018). Developmental dynamics of stem starch accumulation in Sorghum bicolor. *Plant Dir.* 2, 1–15.
- McKinley, B., Rooney, W., Wilkerson, C., and Mullet, J. (2016). Dynamics of biomass partitioning, stem gene expression, cell wall biosynthesis, and sucrose accumulation during development of Sorghum bicolor. *Plant J.* 88, 662–680. doi: 10.1111/tbj.13269
- Meng, S., Peng, J. S., He, Y. N., Zhang, G. B., Yi, H. Y., Fu, Y. L., et al. (2015). Arabidopsis NRT1.5 mediates the suppression of nitrate starvation-induced leaf senescence by modulating foliar potassium level. *Mol. Plant.* 9, 461–470. doi: 10.1016/j.molp.2015.12.015
- Miao, Z., Zhang, T., Qi, Y., Song, J., and Han, Z. (2020). Evolution of the RNA N6-methyladenosine methylome mediated by genomic duplication. *Plant Physiol.* 182, 345–360. doi: 10.1104/pp.19.00323
- Mizuno, H., Kasuga, S., and Kawahigashi, H. (2016). The sorghum SWEET gene family: stem sucrose accumulation as revealed through transcriptome profiling. *Biotechnol. Biofuels* 9: 127.
- Mizuno, H., Kasuga, S., and Kawahigashi, H. (2018). Root lodging is a physical stress that changes gene expression from sucrose accumulation to degradation in sorghum. *BMC Plant Biol.* 18:2. doi: 10.1186/s12870-017-1218-9
- Mullet, J., Morishige, D., McCormick, R., Truong, S., Hillel, J., McKinley, B., et al. (2014). Energy Sorghum—a genetic model for the design of C4 grass bioenergy crops. *J. Exp. Bot.* 65, 3479–3489. doi: 10.1093/jxb/eru229
- Parker, M. T., Knop, K., Sherwood, A. V., Schurch, N. J., Mackinnon, K., Gould, P. D., et al. (2020). Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA processing and m6A modification. *eLife.* 2020:e49658. doi: 10.7554/eLife.49658
- Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, H., Haber, G., et al. (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature* 457, 551–556.
- Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT. StringTie and Ballgown. *Nat. Prot.* 11, 1650–1667. doi: 10.1038/nprot.2016.095
- Ping, X. L., Sun, B., Wang, L., Xiao, W., Yang, X., Wang, W., et al. (2014). Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res.* 24, 177–189. doi: 10.1038/cr.2014.3
- Preker, R. J., Lingner, J., Miville-Sebastia, L., and Keller, W. (1995). The FIP1 gene encodes a component of a yeast pre-mRNA polyadenylation factor that directly interacts with poly(A) polymerase. *Cell* 81, 379–389. doi: 10.1016/0092-8674(95)90391-7
- Ritter, K., Chapman, S., Jordan, D., Godwin, I., and McIntyre, L. (2004). “Investigating the use of sweet sorghum as a model for sugar accumulation in sugarcane,” in *Proceedings of the 4th International Crop Science Congress*, Brisbane.
- Ruzicka, K., Zhang, M., Campilho, A., Bodi, Z., Kashif, M., Saleh, M., et al. (2017). Identification of factors required for m6A mRNA methylation in Arabidopsis reveals a role for the conserved E3 ubiquitin ligase HAKAI. *New Phytol.* 215, 157–172. doi: 10.1111/nph.14586
- Sakai, H., Lee, S. S., Tanaka, T., Numa, H., Kim, J., Kawahara, Y., et al. (2013). Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol.* 54:e6. doi: 10.1093/pcp/pcs183
- Schrick, K., Fujioka, S., Takatsuto, S., Stierhof, Y. D., Stransky, H., Yoshida, S., et al. (2004). A link between sterol biosynthesis, the cell wall, and cellulose in Arabidopsis. *Plant J.* 38, 227–243. doi: 10.1111/j.1365-313x.2004.02039.x
- Scutenaire, J., Deragon, J. M., Jean, V., Benhamed, M., Raynaud, C., Favory, J. J., et al. (2018). The YTH domain protein ECT2 is an m6A reader required for normal trichome branching in Arabidopsis. *Plant Cell* 30, 986–1005. doi: 10.1105/tpc.17.00854
- Shen, Y., Venu, R. C., Nobuta, K., Wu, X., Notibala, V., Demirci, C., et al. (2011). Transcriptome dynamics through alternative polyadenylation in developmental and environmental responses in plants revealed by deep sequencing. *Genome Res.* 21, 1478–1486. doi: 10.1101/gr.114744.110
- Shenker, S., Miura, P., Sanfilippo, P., and Lai, E. C. (2015). IsoSCM: improved and alternative 3' UTR annotation using multiple change-point inference. *RNA* 21, 14–27. doi: 10.1261/rna.046037.114
- Shepard, P. J., Choi, E. A., Lu, J., Flanagan, L. A., Hertel, K. J., and Shi, Y. (2011). Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA* 17, 761–772. doi: 10.1261/rna.2581711
- Shi, Y., Di Giammartino, D. C., Taylor, D., Sarkeshik, A., Rice, W. J., Yates, J. R. III, et al. (2009). Molecular architecture of the human pre-mRNA 3' processing complex. *Mol. Cell.* 33, 365–376. doi: 10.1016/j.molcel.2008.12.028
- Souter, M. A., Pullen, M. L., Topping, J. F., Zhang, X., and Lindsey, K. (2004). Rescue of defective auxin-mediated gene expression and root meristem function by inhibition of ethylene signalling in sterol biosynthesis mutants of Arabidopsis. *Planta* 219, 773–783.
- Srivastava, A. K., Lu, Y., Zinta, G., Lang, Z., and Zhu, J. K. (2018). UTR-dependent control of gene expression in plants. *Trends Plant Sci.* 23, 248–259. doi: 10.1016/j.tplants.2017.11.003
- Stark, R., Grzelak, M., and Hadfield, J. (2019). RNA sequencing: the teenage years. *Nat. Rev. Genet.* 20, 631–656. doi: 10.1038/s41576-019-0150-2
- Sui, N., Yang, Z., Liu, M., and Wang, B. (2015). Identification and transcriptomic profiling of genes involved in increasing sugar content during salt stress in sweet sorghum leaves. *BMC Genomics* 16:534.
- Sun, H., Li, Y., Niu, Q., and Chua, N. H. (2017). Dehydration stress extends mRNA 3' untranslated regions with noncoding RNA functions in Arabidopsis. *Genome Res.* 27, 1427–1436. doi: 10.1101/gr.218669.116
- Sun, X., Zheng, H., Li, J., Liu, J., Zhang, X., and Sui, N. (2020). Comparative transcriptome analysis reveals new lncRNAs responding to salt stress in sweet sorghum. *Front. Bioeng. Biotech.* 8:331. doi: 10.3389/fbioe.2020.00331
- Swigonova, Z., Lai, J., Ma, J., Ramakrishna, W., Llaça, V., Bennetzen, J. L., et al. (2004). Close split of sorghum and maize genome progenitors. *Genome Res.* 14, 1916–1923. doi: 10.1101/gr.2332504
- Szkop, K. J., and Nobeli, I. (2017). Untranslated parts of genes interpreted: making heads or tails of high-throughput transcriptomic data via computational methods: computational methods to discover and quantify isoforms with alternative untranslated regions. *Bioessays* 39:1700090. doi: 10.1002/bies.201700090
- Téllez-Robledo, B., Manzano, C., Saez, A., Navarro-Neila, S., Silva-Navas, J., Lorenzo, L., et al. (2019). The polyadenylation factor FIP1 is important for plant development and root responses to abiotic stresses. *Plant J.* 99, 1203–1219. doi: 10.1111/tbj.14416

- Thomas, P. E., Wu, X., Liu, M., Gaffney, B., Ji, G., Li, Q., et al. (2012). Genome-wide control of polyadenylation site choice by CPSF30 in Arabidopsis. *Plant Cell* 24, 4376–4388. doi: 10.1105/tpc.112.096107
- Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 14, 178–192. doi: 10.1093/bib/bbs017
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Prot.* 7, 562–579. doi: 10.1038/nprot.2012.016
- Tu, M. (2020). *priUTR* [Source Code]. Available online at: <https://github.com/mint1234/3UTR-> (accessed April 15, 2020).
- Turco, G. M., Kajala, K., Kunde-Ramamoorthy, G., Ngan, C. Y., Olson, A., Deshpande, S., et al. (2017). DNA methylation and gene expression regulation associated with vascularization in Sorghum bicolor. *New Phytol.* 214, 1213–1229. doi: 10.1111/nph.14448
- Usadel, B., Poree, F., Nagel, A., Lohse, M., Czedik-Eysenberg, A., and Stitt, M. (2009). A guide to using MapMan to visualize and compare omics data in plants: a case study in the crop species. *Maize. Plant Cell Environ.* 32, 1211–1229. doi: 10.1111/j.1365-3040.2009.01978.x
- Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van de Peer, Y., et al. (2018). PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucl. Acid Res.* 46, D1190–D1196.
- Varoquaux, N., Cole, B., Gao, C., Pierroz, G., Baker, C. R., Patel, D., et al. (2019). Transcriptomic analysis of field-droughted sorghum from seedling to maturity reveals biotic and metabolic responses. *Proc. Natl. Acad. Sci. U.S.A.* 116, 27124–27132. doi: 10.1073/pnas.1907500116
- Wang, W., Wei, Z., and Li, H. (2014). A change-point model for identifying 3'UTR switching by next-generation RNA sequencing. *Bioinformatics.* 30, 2162–2170. doi: 10.1093/bioinformatics/btu189
- Wang, Y., Li, Y., Toth, J. I., Petroski, M. D., Zhang, Z., and Zhao, J. C. (2014). N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat. Cell. Biol.* 16, 191–198. doi: 10.1038/ncb2902
- Wei, L., Song, P., Wang, Y., Lu, Z., Tang, Q., Yu, Q., et al. (2018). The m6A reader ECT2 controls trichome morphology by affecting mRNA stability in Arabidopsis. *Plant Cell* 30, 968–985. doi: 10.1105/tpc.17.00934
- Wu, X., Gaffney, B., Hunt, A. G., and Li, Q. Q. (2014). Genome-wide determination of poly(A) sites in *Medicago truncatula*: evolutionary conservation of alternative poly(A) site choice. *BMC Genomics* 15:615. doi: 10.1186/1471-2164-15-615
- Wu, X., Hu, W., Luo, H., Xia, Y., Zhao, Y., Wang, L., et al. (2016a). Transcriptome profiling of developmental leaf senescence in sorghum (*Sorghum bicolor*). *Plant Mol. Biol.* 92, 555–580. doi: 10.1007/s11103-016-0532-1
- Wu, X., Liu, M., Downie, B., Liang, C., Ji, G., Li, Q. Q., et al. (2011). Genome-wide landscape of polyadenylation in Arabidopsis provides evidence for extensive alternative polyadenylation. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12533–12538. doi: 10.1073/pnas.1019732108
- Wu, X., Zhang, Y., and Li, Q. Q. (2016b). PlantAPA: a portal for visualization and analysis of alternative polyadenylation in plants. *Front. Plant Sci.* 7:889. doi: 10.3389/fpls.2016.00889
- Xia, Z., Donehower, L. A., Cooper, T. A., Neilson, J. R., Wheeler, D. A., Wagner, E. J., et al. (2014). Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat. Commun.* 5:5274.
- Xie, Q., and Xu, Z. (2019). Sustainable agriculture: from sweet sorghum planting and ensiling to ruminant feeding. *Mol. Plant.* 12, 603–606. doi: 10.1016/j.molp.2019.04.001
- Xu, J., and Messing, J. (2008). Organization of the prolamin gene family provides insight into the evolution of the maize genome and gene duplications in grass species. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14330–14335. doi: 10.1073/pnas.0807026105
- Yang, Z., Li, J., Liu, L. U., Xie, Q., and Sui, N. (2020). Photosynthetic regulation under salt stress and salt-tolerance mechanism of sweet sorghum. *Front. Plant Sci.* 10:1722. doi: 10.3389/fpls.2019.01722
- Yang, Z., Zheng, H., Wei, X., Song, J., Wang, B., and Sui, N. (2018). Transcriptome analysis of sweet Sorghum inbred lines differing in salt tolerance provides novel insights into salt exclusion by roots. *Plant Soil.* 430, 423–439. doi: 10.1007/s11104-018-3736-0
- Ye, C., Long, Y., Ji, G., Li, Q. Q., and Wu, X. (2018). APAtrap: identification and quantification of alternative polyadenylation sites from RNA-seq data. *Bioinformatics* 34, 1841–1849. doi: 10.1093/bioinformatics/bty029
- Ye, C., Zhou, Q., Wu, X., Ji, G., and Li, Q. Q. (2019). Genome-wide alternative polyadenylation dynamics in response to biotic and abiotic stresses in rice. *Ecotoxicol. Environ.* 183:109485. doi: 10.1016/j.ecoenv.2019.109485
- You, L., Wu, J., Feng, Y., Fu, Y., Guo, Y., and Long, L. (2014). APASdb: a database describing alternative poly(A) sites and selection of heterogeneous cleavage sites downstream of poly(A) signals. *Nucl. Acids Res.* 43, D59–D67.
- Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118
- Zhang, F., Zhang, Y., Liao, J., Yu, Y., Zhou, Y., Feng, Y., et al. (2019). The subunit of RNA N6-methyladenosine methyltransferase OsFIP regulates early degeneration of microspores in rice. *PLoS Genet.* 15:e1008120. doi: 10.1371/journal.pgen.1008120
- Zhang, J., and Wei, Z. (2016). An empirical Bayes change-point model for identifying 3' and 5' alternative splicing by next-generation RNA sequencing. *Bioinformatics.* 32, 1823–1831.
- Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., et al. (2018). Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat. Genet.* 50, 1565–1573.
- Zhang, Y., Ngu, D. W., Carvalho, D., Liang, Z., Qiu, Y., Roston, R. L., et al. (2017). Differentially regulated orthologs in sorghum and the subgenomes of maize. *Plant Cell* 29, 1938–1951. doi: 10.1105/tpc.17.00354
- Zhang, L. M., Leng, C. Y., Luo, H., Wu, X. Y., Liu, Z. Q., Zhang, Y. M., et al. (2018). Sweet sorghum originated through selection of Dry, a plant-specific NAC transcription factor gene. *Plant Cell* 30, 2286–2307. doi: 10.1105/tpc.18.00313
- Zhao, T., Huan, Q., Sun, J., Liu, C., Hou, X., Yu, X., et al. (2019). Impact of poly(A)-tail G-content on Arabidopsis PAB binding and their role in enhancing translational efficiency. *Genome Biol.* 20:189. doi: 10.1186/s13059-019-1799-8
- Zheng, G., Dahl, J. A., Niu, Y., Fedorcsak, P., Huang, C. M., Li, C. J., et al. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell.* 49, 18–29. doi: 10.1016/j.molcel.2012.10.015
- Zheng, Y., Drechsler, N., Rausch, C., and Kunze, R. (2016). The Arabidopsis nitrate transporter NPF7.3/NRT1.5 is involved in lateral root development under potassium deprivation. *Plant Signal. Behav.* 11:e1176819. doi: 10.1080/15592324.2016.1176819
- Zhou, Q., Fu, H., Yang, D., Ye, C., Zhu, S., Lin, J., et al. (2019). Differential alternative polyadenylation contributes to the developmental divergence between two rice subspecies, japonica and indica. *Plant J.* 98, 260–276. doi: 10.1111/tpj.14209

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Tu and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.