**Computational Biology and Bioinformatics**
2023; 11(1): 13-18
http://www.sciencepublishinggroup.com/j/cbb
doi: 10.11648/j.cbb.20231101.12
ISSN: 2330-8265 (Print); ISSN: 2330-8281 (Online)

# The Importance of Stiff Change of U(T) Content Around Splicing Sites in Efficient Plant Intron Splicing -- A Case Study in Rice

**Fangyu Zhang[1], Zhengfeng Zhang[2], Enci Wang[2], Chengqi Wang[1], Benze Xiao[1, *]**

[1]College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, People's Republic of China

[2]Hubei Key Laboratory of Genetic Regulation and Integrative Biology, School of Life Sciences, Central China Normal University, Wuhan, People's Republic of China

## Email address:

benzexiao@mail.hzau.edu.cn (Benze Xiao)

[*]Corresponding author

## To cite this article:

Fangyu Zhang, Zhengfeng Zhang, Enci Wang, Chengqi Wang, Benze Xiao. The Importance of Stiff Change of U(T) Content Around Splicing Sites in Efficient Plant Intron Splicing -- A Case Study in Rice. *Computational Biology and Bioinformatics*. Vol. 11, No. 1, 2023, pp. 13-18. doi: 10.11648/j.cbb.20231101.12

**Received**: April 20, 2023; **Accepted**: May 12, 2023; **Published**: June 5, 2023

**Abstract:** Pre-mRNAs splicing is one of the fundamental process which generates multiple transcripts from a single gene, contributing to transcriptome and proteome diversity. AS is regulated by the cooperation of trans-factors and cis-elements. In plants, extensive alternative splicing occurs not only in tissue-specific manner but also in response to stress conditions. Intron retention is the most predominant splicing type. However, the cis-elements regulating intron retention are still ambiguous in plants, especially under environmental stresses. This study aimed to elucidate the cis-elements underlying intron retention in plants under adverse enrironments. Using RNA-seq data of rice cultivars IRAT109 and ZS97 under drought environments, we compared the sequence characteristics between constitutive and retained introns. The results show that the main AS types include intron retention (IR), alternative acceptor sites (AA), alternative donor sites (AD) and cassette exon (exon skipping, ES). Among of them, IR was the prevelent pattern with frequencies of 30.8-31.2%. Motif analysis of 5' and 3' 200bp intron sequences found rich U(T) in the motifs for both constitutive and retained introns. By further analysis of base composition of sequences flanking splice sites, we detected a notable difference in U(T) content between introns and their neighboring exons in constitutive introns, but not in retained introns. The results in this study suggested that the lack of significant changes in U(T) content between retained introns and neighboring exons might be a potential cis feature of intron retention.

**Keywords:** Alternative Splicing, Intron Retention, Cis-elements, Rice, Drought Stress

## 1. Introduction

Alternative splicing (AS) is a process by which different combinations of exons can be joined together to produce a diverse array of messenger RNA (mRNA) isoforms from a single primary transcript. This process is a common mechanism for controlling gene expression and generating protein diversity in higher eukaryotic cells [1, 4, 36].

Several types of AS events have been explored, including intron retention (IR), exon skipping (ES), alternate 5′ splice sites (alternate donor site, AD) and alternate 3′ splice site (alternative acceptor site, AA) [4]. The accurate regulation of AS needs the cooperation of trans-factors and cis-elements [26, 34]. Increasing evidences show that additional mechanisms help generate plastic AS events with high specificity and fidelity, including the regulators of spliceosome assembly, chromatin remodeling, epigenetic modification, RNA structure and so on [2, 24].

In recent years, cis-acting elements have been identified with a profound influence on splicing outcome [3, 5, 25]. The cis-elements regulating intron recognition and excision include the consensus splice site sequences (mainly GT and AG at the intron termini), branch point sequence YURAC, exon splicing enhancers (ESE), exon splicing silencers (ESS), intron splicing enhancers (ISE) and intron splicing silencers (ISS) [22]. Besides these major cis-features for intron

excision, others like exon length, intron length, RNA modification and base content of introns and exons also have an impact on the correct identification and splicing of introns [19, 32]. Increasing efforts were put into AS behavior and regulation in computational biology field, known as "Splicing Codes" [12, 14, 17, 20].

Intron retention has been revealed to be an important component to finely tune transcriptomes in mammals [7, 13]. Weak 3' splice site recognition and differential GC content between exons and introns have been shown to establish splicing strategies of introns in human [27]. A recent study in Drosophila has suggested that the introns with increased retention have exon-like characteristics, including the length, GC content, and RNA polymerase II occupancy [6]. In plants, GC-rich exon sequences in pre-mRNAs have been reported to act as recognition targets [16]. The unusual UA- or U-rich compositional bias of plant introns was thought to be important for correct mRNA processing and splice-site selection [11, 16].

Substantial evidence suggests AS is an essential post-transcriptional regulatory mechanism in plant development and environmental response [15, 19, 33]. Although IR has been extensively demonstrated to be the predominant type of AS in plants [8, 18, 29, 35, 37], the mechanism influencing efficient splicing of retained introns is yet to be determined, especially in plant responsive to abiotic stresses. In the present study, we aim to decode the sequences characteristics regarding to intron retention events in plant under environmental stress through analyzing RNA-seq data of rice under drought stress. We found the potential role of the stiff UA-content differences between intron and exon in promoting intron splicing.

# 2. Data and Methods

### *2.1. Plant Materials, cDNA Libraries Construction and Illumine Sequencing*

Rice accessions IRAT109 and ZS97, drought-resistant and drought-sensitive variety, respectively were grown in separated pots in greenhouse, for 4 weeks in well-water conditions and followed by one-week with or without watering as control or drought stress treatment. Total RNA was extracted from the leaves of two independent pots separately as two biological replicates for each variety using plant RNA kit (OMEGA Bio-Tek, USA). cDNA libraries and Illumina sequencing was performed at Novogene Bioinformatics Technology Co., Ltd., Beijing, China (www.novogene.cn) individually for each using Illumina sequencing platform (HiSeqTM 2500). The detailed RNA preparation, libraries construction and sequencing were described in a previous report [37].

### *2.2. Transcriptome Assembly and Alternative Splicing Analysis*

FastQC analysis was performed to estimate the quality of the raw reads. After trimming adapter sequences and filtering low-quality reads with >5% ambiguous bases and low-quality reads, the clean reads were used to assembly the transcriptome based on the reference genome of Os-Nipponbare-Reference-IRGSP-1.0. TopHat 2.0.11 program was used for read mapping and cufflinks were used for transriptome assembly [30]. Cuffmerge were employed to generate comprehensive transcripts [10]. Splice junction information was analyzed using ASTALAVISTA software [9]. The detailed analyzing method were previously described [37].

### *2.3. Investigation of AS-Associated Cis Features*

The datasets for motif analysis were obtained using in-house Perl scripts with genome sequences and exon/intron boundary position information (Figure 1). In brief, to compare potential motif in retained and constitutive introns, the 5' and 3' 200bp sequences were retrieved from the constitutive and retained introns, respectively. To compare the base content of sequences relevant to intron splicing, the 200bp sequences flanking donor and acceptor sites were retrieved for retained and constitutive introns, respectively. MEME suite web version was employed for motif searching with 6- and 15-mers as the minimum and maximum width of each motif, respectively [21]. Sequences retrievement and calculation of base content in the sequences was performed with in-house Perl scripts.

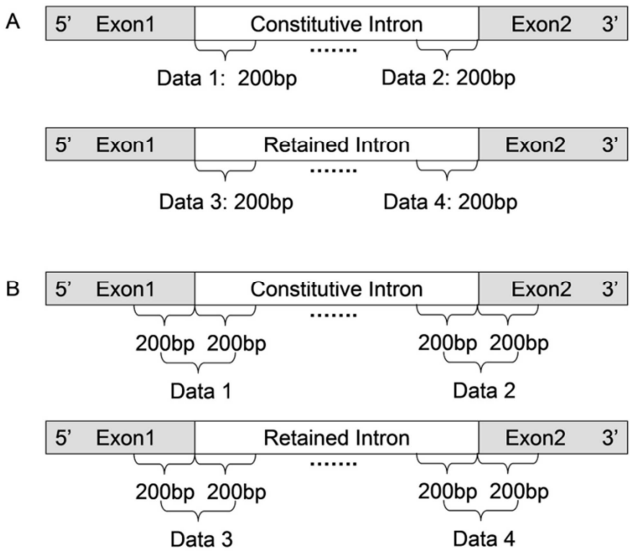


***Figure 1.*** *Sequence retrievement for obtaining datasets for cis-feature analysis of intron retention. (A). The parameters used to retrieve the four datasets used to compare potential cis-elements in retained and constitutive introns. (B) The parameters used to retrieve the four datasets used to compare the base content of sequences flanking donor and acceptor sites of retained and constitutive introns. The gray boxes represent exons and the white boxes indicate introns.*

# 3. Results

### *3.1. Transcripts Assembly and Alternative Splicing Analysis*

About 30.6 million 100-bp paired-end clean reads pairs for

each library were generated after adapter trimming and low-quality reads filtered. On average, about 53.7 and 51.3 million reads were uniquely mapped to reference in Zhenshan97 and IRAT109 respectively. AS events were identified using ASTALAVISTA [9]. 19120 events of 9342 genes and 20091 AS events of 9644 genes were detected in Zhenshan97 and IRAT109 respectively. Totally 5301 genes show AS in both samples, whereas 4041 and 4343 genes show AS either in Zhenshan97 or in IRAT109. The main AS types include intron retention (IR), alternative acceptor sites (AA), alternative donor sites (AD), cassette exon (exon skipping, ES). IR was the prevelent pattern with frequencies of 30.8% and 31.2% in Zhenshan97 and IRAT109, respectively [37].

### *3.2. The Cis-features of Retained and Constitutive Introns*

While numerous studies have attempted to investigate the mechanisms underlying AS, many questions still remain, particularly the cis features underlying intron retention in plants. Regulatory cis elements in AS regions can serve as binding sites for trans-splicing factors to recruit the splicing machinery [23]. The prevalence of IR events in plant encouraged us to explore the different cis-elements between retained introns (RIs) and constitutive introns (CIs). The base content and ratio between RIs and CIs in the two genotypes were analyzed using the genome sequences and transcriptome data. To acquire the constitutive introns data sets, the merged gene model file was used to identify introns larger than 200 bp in a given genes and spliced out in all transcripts (Figure 1A). For the retained introns data sets, genes containing introns larger than 200 bp and from IR events common to the two rice genotypes were identified (Figure 1A). The 5′ and 3′ 200 bp sequences from 1596 constitutive introns and 1498 retained introns were used to screen 6- to 15-mer conserved motifs using the MEME suite [21]. The constitutive introns generated a highly conserved 15-nucleotide motif with enriched C and T nucleosides in 5′ sequences and a 12-nucleotide motif with enriched T nucleosides in 3′ sequences, especially T-quadruplex at first 4 sites (Figure 2, CI 5′ and CI 3′). Slightly different motifs were identified in the 5′ and 3' sequences of retained introns. These motifs were similar with a high proportion of the nucleoside ‘T’, while a T-quadruplex present at last 4 sites of RI 5’ sequences (Figure 2, RI 5′ and RI 3′). This was anticipated as it is generally believed that high U(T) levels in introns are important for splice site recognition [11, 16].

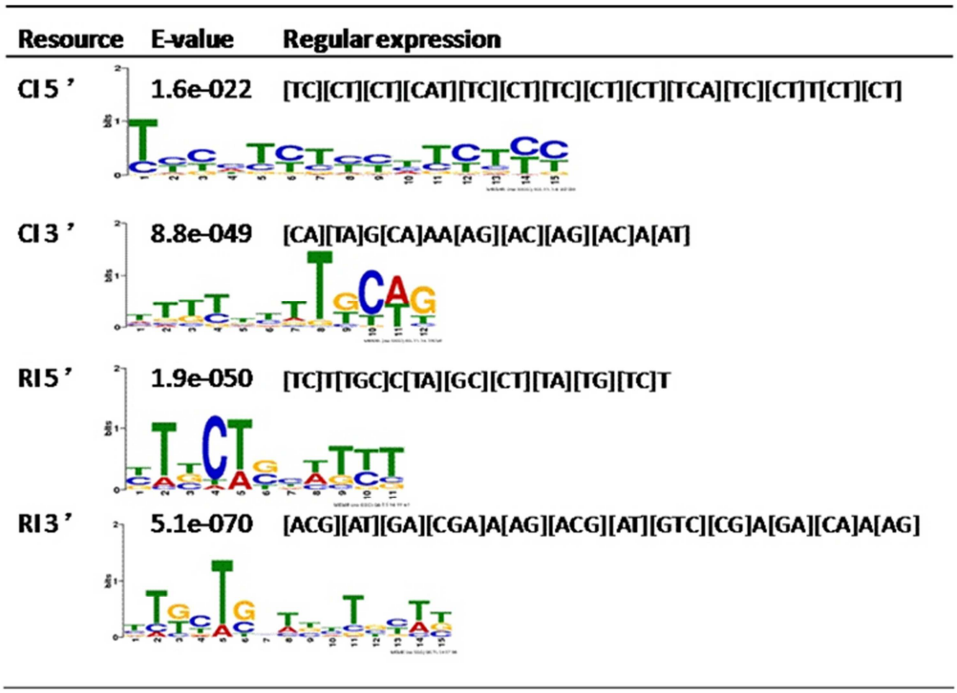


***Figure 2.*** *Motifs information generated using MEME analysis of constitutive and retained introns. Constitutive intron (CI) 5' and CI 3' refer to the 5' and 3' terminal sequences of constitutively spliced introns, respectively. Retained intron (RI) 5' and RI 3' refer to the 5' and 3' terminal sequences of retained introns, respectively.*

### *3.3. The Cis-features of Splicing Sites Relevant to Retained and Constitutive Introns*

To decipher the possible role of intron and exon nucleotide content in intron retention, we further compared the flanking sequences of splicing donor and acceptor sites from the retained and constitutive introns. Four data sets were obtained including: (1) 200-bp-long sequences upstream and downstream of 1596 constitutive intron donor sites; (2) 200-bp-long sequences upstream and downstream of 1498 retained intron donor sites; (3) 200-bp-long sequences upstream and downstream of 1596 constitutive intron acceptor sites and (4) 200-bp-long sequences upstream and downstream of 1498 retained intron acceptor

sites (Figure 1B). The frequencies of T, A, C, and G for each 10-bp sliding window were calculated for all four data sets (Figure 3). For the constitutive intron donor sites, the upstream (exon) region is composed of about 40% AT(U) and 20% T(U), whereas in the downstream (intron) region the percentage of AT(U) and T(U) increased sharply to an average of 55% and 31%, respectively. A similar phenomenon was observed for the upstream (intron) and downstream (exon) regions of constitutive intron acceptor sites (Figure 3). The data from the constitutive introns indicated that the sharp increase in AT content in the intronic compared with the exonic region might be a key feature for efficient intron splicing. In contrast, no profound differences in the nucleotide fractions were observed between the exon and intron regions around splicing sites for the retained introns (Figure 3). This suggests that the lack of stiff changing of AU(T) content around the splicing sites might hinder the precise recognition by spliceosome and ultimately cause intron retention.

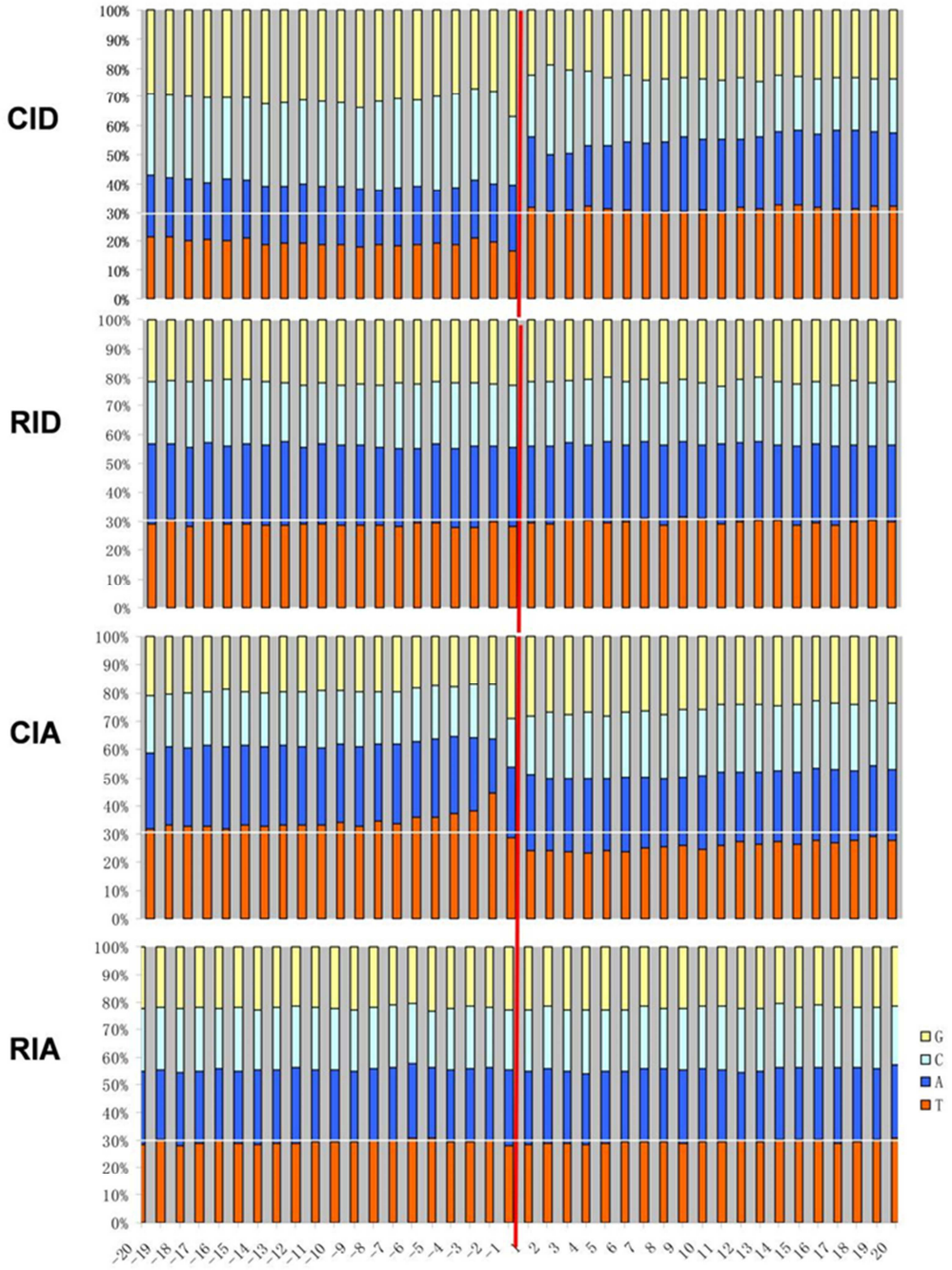


***Figure 3.*** *The frequencies of nucleotides around splicing donor and acceptor sites for constitutive and retained introns. Frequencies of nucleotides were generated by counting the numbers of each type of nucleotide in sliding windows of 10 nucleotides, fromsequences flanking splice sites of constitutive and retained introns, grouped as constitutive intron donor (CID), retained intron donor (RID), constitutive intron acceptor (CIA), and retained intron acceptor (RIA) sites. The X-axis scale is 10 bp and the Y-axis shows the % frequency of T, A, C and G in each 10-bp region. The vertical red line represents the boundary of introns and exons; the white horizontal lines show nucleotide frequency of 30%.*

# 4. Discussion

*The Constant Base Content Around Splicing Sites Might Be Responsible for Intron Retention*

The studies of cis elements underlying AS in animals generally focused on intronic sequences flanking constitutive and alternatively spliced exons, because exon skipping is the predominant AS type in animals [31]. Comprehensive intron recognition and excision are determined by sequence elements in pre-mRNA, including 5′ and 3′ splice sites, branch point and poly-pyrimidine tract, and the binding of trans-splicing factors [4]. It has been revealed that high U(T)A, and especially high U(T), levels in plant introns are crucial for the correct recognition and splicing of introns [11, 28]. In the present study, we found rich U(T) content was found in the motifs of both constitutive and retained introns in the present study (Figure 2). As there was no clear difference in the U(T) content between the constitutive and retained introns, we looked further for any possible cis-differences between them. By examining the base composition of sequences flanking splice sites for constitutive and retained introns, a notable difference in base content between introns and their neighboring exons was identified for constitutive introns (Figure 3). In contrast, a consistent U(T) level around splice sites was detected for retained introns. Therefore, We propose that simple U(T)-rich introns is not enough for intron recognition and excision, while a relatively higher U(T) content and lower GC content in introns than neighboring exons is essential for splicing of introns. The absence of a U(T) content difference between introns and their neighboring exons might lead to a low intron recognition efficiency and therefore, a high occurrence of IR events.

By using RNA-Seq data, the present results not only support the role of U(T)-rich regions in splice efficiency but also expand the idea that only rich U(T) in intron sequences is not sufficient for high splicing efficiency. We infer that the sharp changing of U(T) content between introns and their neighboring exons plays an essential role in intron recognition and excision. Conversely, the lack of this contrast will decrease the recognition, splicing efficiency and as a consequence, intron retention.

# 5. Conclusion

By comparing base content around the splicing sites between the constitutive and retained introns we propose that the lack of significant differences in U(T) content between introns and neighboring exons might be an important cis-feature of intron retention. This discovery extends previous research findings and will encourage future studies on gene splicing in plants. However, one limitation of this study is that we only explore the sequence characteristics for intron retention with RNA-seq data from rice under drought stress. Data from more diverse plant species, different growth conditions and variant tissues is necessary to decode the precise cis-features underlying different AS types. Furthermore, the rapid increasing RNA-seq data of populations will promote the analysis of splicing QTL for alternative splicing, which will provide causing cis-elements underlying splicing efficiency. Hereafter, we will employ more comprehensive data to decode more general theory regarding to cis-elements underlying AS in the next project.

# Author Contributions

BX designed the project, conceived the study and wrote the manuscript. FZ, ZZ, EW and CW performed data analysis. All authors read and approved the final manuscript.

# Conflict of Interest Statement

The authors declare that they have no competing interests.

# Acknowledgements

This work was supported financially by the National Natural Science Foundation of China (Codes: 31000116 and 30800680), the self-determined research fund of Central China Normal University (CCNU18QN027) and the project of Hubei Key Laboratory of Genetic Regulation and Integrative Biology (GRIB201701).

# References

[1] Amit M, Donyo M, Hollander D, Goren A, Kim E, Gelfman S, Lev-Maor G, Burstein D, Schwartz S, Postolsky B, Pupko T, Ast G (2012) Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. Cell Rep 1: 543-556.

[2] Apostolidi M, Stamatopoulou V (2023) Aberrant splicing in human cancer: An RNA structural code point of view. Front Pharmacol 14: 1137154.

[3] Bao S, Moakley DF, Zhang C (2019) The Splicing Code Goes Deep. Cell 176: 414-416.

[4] Baralle F, Giudice J (2017) Alternative splicing as a regulator of development and tissue identity. Nat Rev Mol Cell Biol 18: 437-451.

[5] Baralle M, Baralle FE (2018) The splicing code. Biosystems 164: 39-48.

[6] Bou Sleiman M, Frochaux MV, Andreani T, Osman D, Guigo R, Deplancke B (2020) Enteric infection induces Lark-mediated intron retention at the 5' end of Drosophila genes. Genome Biol 21: 4.

[7] Braunschweig U, Barbosa-Morais NL, Pan Q, Nachman EN, Alipanahi B, Gonatopoulos-Pournatzis T, Frey B, Irimia M, Blencowe BJ (2014) Widespread intron retention in mammals functionally tunes transcriptomes. Genome Res 24: 1774-1786.

[8] Chang CY, Lin WD, Tu SL (2014) Genome-Wide Analysis of Heat-Sensitive Alternative Splicing in Physcomitrella patens. Plant Physiol 165: 826-840.

[9] Foissac S, Sammeth M (2007) ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. Nucleic Acids Res 35: W297-299.

[10] Ghosh S, Chan CK (2016) Analysis of RNA-Seq Data Using TopHat and Cufflinks. Methods Mol Biol 1374: 339-361.

[11] Goodall GJ, Filipowicz W (1989) The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. Cell 58: 473-483.

[12] Hia F, Yang SF, Shichino Y, Yoshinaga M, Murakawa Y, Vandenbon A, Fukao A, Fujiwara T, Landthaler M, Natsume T, Adachi S, Iwasaki S, Takeuchi O (2019) Codon bias confers stability to human mRNAs. EMBO Rep 20: e48220.

[13] Jacob AG, Smith CWJ (2017) Intron retention as a component of regulated gene expression programs. Hum Genet 136: 1043-1057.

[14] Lemaire S, Fontrodona N, Aube F, Claude JB, Polveche H, Modolo L, Bourgeois CF, Mortreux F, Auboeuf D (2019) Characterizing the interplay between gene nucleotide composition bias and splicing. Genome Biol 20: 259.

[15] Liu J, Zhang Y, Zheng Y, Zhu Y, Shi Y, Guan Z, Lang K, Shen D, Huang W, Dou D (2023) PlantExp: a platform for exploration of gene expression and alternative splicing based on public plant RNA-seq samples. Nucleic Acids Res 51: D1483-D1491.

[16] Lorkovic ZJ, Wieczorek Kirk DA, Lambermon MH, Filipowicz W (2000) Pre-mRNA splicing in higher plants. Trends Plant Sci 5: 160-167.

[17] Louadi Z, Oubounyt M, Tayara H, Chong KT (2019) Deep Splicing Code: Classifying Alternative Splicing Events Using Deep Learning. Genes (Basel) 10.

[18] Marquez Y, Brown JW, Simpson C, Barta A, Kalyna M (2012) Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. Genome Res 22: 1184-1195.

[19] Muhammad S, Xu X, Zhou W, Wu L (2022) Alternative splicing: An efficient regulatory approach towards plant developmental plasticity. Wiley Interdiscip Rev RNA: e1758.

[20] Neitz M, Neitz J (2021) Intermixing the OPN1LW and OPN1MW Genes Disrupts the Exonic Splicing Code Causing an Array of Vision Disorders. Genes (Basel) 12.

[21] Nystrom SL, McKay DJ (2021) Memes: A motif analysis environment in R using tools from the MEME Suite. PLoS Comput Biol 17: e1008991.

[22] Reddy A, Maria K, Barta A (2013) Complexity of the alternative splicing landscape in plants. Plant Cell 25: 3657-3683.

[23] Reddy AS, Rogers MF, Richardson DN, Hamilton M, Ben-Hur A (2012) Deciphering the plant splicing code: experimental and computational approaches for predicting alternative splicing and splicing regulatory elements. Front Plant Sci 3: 18.

[24] Rogalska ME, Vivori C, Valcarcel J (2023) Regulation of pre-mRNA splicing: roles in physiology and disease, and therapeutic prospects. Nat Rev Genet 24: 251-269.

[25] Rong S, Buerer L, Rhine CL, Wang J, Cygan KJ, Fairbrother WG (2020) Mutational bias and the protein code shape the evolution of splicing enhancers. Nat Commun 11: 2845.

[26] Sanchez SE, Petrillo E, Kornblihtt AR, Yanovsky MJ (2011) Alternative splicing at the right time. RNA Biol 8: 954-959.

[27] Smith LD, Lucas CM, Eperon IC (2013) Intron retention in the alternatively spliced region of RON results from weak 3' splice site recognition. PLoS One 8: e77208.

[28] Syed NH, Kalyna M, Marquez Y, Barta A, Brown JW (2012) Alternative splicing in plants--coming of age. Trends Plant Sci 17: 616-623.

[29] Thatcher SR, Zhou W, Leonard A, Wang BB, Beatty M, Zastrow-Hayes G, Zhao X, Baumgarten A, Li B (2014) Genome-wide analysis of alternative splicing in Zea mays: landscape and genetic regulation. Plant Cell 26: 3472-3487.

[30] Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25: 1105-1111.

[31] Xing Y, Wang Q, Lee C (2006) Evolutionary divergence of exon flanks: a dissection of mutability and selection. Genetics 173: 1787-1791.

[32] Xiong HY, Alipanahi B, Lee LJ, Bretschneider H, Merico D, Yuen RK, Hua Y, Gueroussov S, Najafabadi HS, Hughes TR, Morris Q, Barash Y, Krainer AR, Jojic N, Scherer SW, Blencowe BJ, Frey BJ (2015) RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. Science 347: 1254806.

[33] Xue R, Mo R, Cui D, Cheng W, Wang H, Qin J, Liu Z (2023) Alternative Splicing in the Regulatory Circuit of Plant Temperature Response. Int J Mol Sci 24.

[34] Yeap WC, Namasivayam P, Ho CL (2014) HnRNP-like proteins as post-transcriptional regulators. Plant Sci 227: 90-100.

[35] Zhang C, Gschwend AR, Ouyang Y, Long M (2014) Evolution of gene structural complexity: an alternative-splicing-based model accounts for intron-containing retrogenes. Plant Physiol 165: 412-423.

[36] Zhang G, Guo G, Hu X, Zhang Y, Li Q, Li R, Zhuang R, Lu Z, He Z, Fang X, Chen L, Tian W, Tao Y, Kristiansen K, Zhang X, Li S, Yang H, Wang J (2010) Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. Genome Res 20: 646-654.

[37] Zhang Z, Xiao B (2018) Comparative alternative splicing analysis of two contrasting rice cultivars under drought stress and association of differential splicing genes with drought response QTLs. Euphytica 214: 73.