Current Plant Biology 9–10 (2017) 29–36

Contents lists available at ScienceDirect

Current Plant Biology

journal homepage: www.elsevier.com/locate/cpb

Comparative landscape of alternative splicing in fruit plants☆

Gaurav Sablok[a,*], Brian Powell[b], Jonathan Braessler[b], Feng Yu[b], Xiang Jia Min[a,*]

[a] Department of Biological Sciences, Youngstown State University, Youngstown, OH 44555, USA
[b] Department of Computer Science and Information Systems, Youngstown State University, Youngstown, OH 44555, USA

ARTICLE INFO

Article history:
Received 12 May 2017
Received in revised form 17 June 2017
Accepted 26 June 2017

Keywords:
Alternative splicing
Fruit plant
Apple
Grape
Orange
Strawberry
Expressed sequence tags

ABSTRACT

Alternative splicing (AS) has played a major role in defining the protein diversity, which could be linked to phenotypic alternations. It is imperative to have a comparative resolution of AS to understand the pre-mRNAs splicing diversity. In the present research, we present a comparative assessment of the AS events in four different fruit plants including apple (*Malus domestica*), grape (*Vitis vinifera*), sweet orange (*Citrus sinensis*), and woodland strawberry (*Fragaria vesca*), using spliced mapping of the expressed sequence tags and mRNA sequences. We identified a total of 2039 AS events in apple, 2454 in grape, 1425 in orange, and 631 in strawberry, respectively. In this study grape displayed the maximum number of genes (1588) associated with the splicing, followed by apple (1580), orange (1133) and strawberry (444). Transcripts mapping analysis shows that grape plant has relatively larger intron sizes than introns in other fruit species. The data provide a basis for further functional characterization of the genes undergoing AS and can be accessed at Plant Alternative Splicing Database (http://proteomics.ysu.edu/altsplice/plant/).

1. Introduction

Plant development depends upon a complex interaction of proteins and as such protein diversity contributes to these interactions. Protein diversity has not only played an intricate role in regulating the transcriptional and post-transcriptional responses but also has played a major role in regulating the stress responses [1,2]. The factor that contributes to increasing protein diversity, termed as alternative splicing (AS) is a key mechanism, which leads to spliceosomal alternations resulting in production of more than one splice transcripts [3–6]. It has been widely elucidated that these splicing events not only affect the developmental patterns, but also play an important role in regulating the stress responses in fruit species [2,7], the fate and divergence of the duplicated genes [8], and also contribute to the mechanistic understanding of the miRNA regulation [9]. Patterns of alternatively spliced transcripts have been widely observed with reports suggesting that 90% of human genes containing multiple exons are alternatively spliced [10], thus, demonstrating exon skipping as a major splice event in humans. In plants, due to the presence of long introns, often intron-retention has been seen as a major splice event, with as many as 60% of multi-exon genes undergoing AS in the model plant *Arabidopsis thaliana* [1,11].

mRNA transcript isoforms are generally generated through four basic events in AS: [1] intron retention (IR) in the mature mRNA; [2] exon skipping (ES) resulting from alternative exon usage (AEU); [3] alternative donor site (AltD) and [4] alternative acceptor site (AltA) that are resulted from the use of cryptic splice sites that may elongate or shorten an exon [4,11–13]. Approximately 60–75% of AS events occur within the protein coding regions of mRNAs, resulting changes in binding properties, intracellular localization, protein stability, enzymatic, and signaling activities [14]. In plants, IR has been shown to be the most dominant form with reports suggesting the proportions of intron containing genes undergoing AS in plants ranged from ~30% to >60% depending the depth of available transcriptome data [1,15]. In addition to the above mentioned basic AS events, various complex types can be formed by combination of basic events [1,13,15]. AS isoforms might encode distinct functional proteins or might be nonfunctional, which harbor a premature termination codon. These non-functional isoforms generated through the process called "regulated unproductive splicing and translation" (RUST) are degraded by a process known as nonsense-mediated decay (NMD) [4,12,16]. Recent study on serine/arginine (SR) genes, which are spatiotemporally regulated and also show a varied amount of splicing diversity, has revealed widespread coupling of AS with NMD in SR gene family, suggesting a strong link between unproductive splicing and the abundance of functional transcripts [17]. Nonetheless, association of AS and RNA-binding proteins has been established, specifically RNA-binding protein AtGRP8 up-regulation has been shown to promote the use of cryptic 5′ splice site thus producing a splice transcript, which

☆ This article is part of a special issue entitled "Plant Development", published in the journal Current Plant Biology 9–10, 2017.

* Corresponding authors.
E-mail addresses: sablokg@gmail.com (G. Sablok), xmin@ysu.edu (X.J. Min).

http://dx.doi.org/10.1016/j.cpb.2017.06.001

in mutants revealed a target direct of NMD in *A. thalina* [18]. Identification of AS events have been widely done in several plant species such as *A. thaliana* [12,19–21], *Oryza sativa* (rice) [12], *Zea mays* (maize) [22,23], *Sorghum bicolor* (sorghum) [23,24], *Nelumbo nucifera* (sacred lotus) [25], *Vitis vinifera* (grape) [7,9], *Brachypodium distachyon* [13,15], and *Ananas comosus* (pineapple) [26]. The SR proteins, which belong to the class of RNA-binding proteins, have been shown to play key role in regulating the splicing machinery [27].

With the advent of the next generation sequencing approaches, several classification approaches have been used for the identification of AS types, which includes differential splicing (Diffseq), differential exon-usage (DEXseq) and application of Bayesian (rMATS), splicing graph based detections and count based approaches (Spladder) [28–31]. Although the application of these approaches have revealed the AS landscape variations in different plant species with the splicing information based on the method, application of these approaches are limited in elucidating the AS landscape in polyploid species, mainly due to the heterozygosity and the large genome size of these polyploid species with ancient genome duplication events. Previously, EST/mRNA based approaches have been widely used to understand the AS events in polyploid species and have provided robust estimates of the splice detection in the polyploid species. Taking into account these above mentioned considerations, in this work, we carried out a survey of AS landscape in four fruit plants, which include apple (*Malus domestica*), grape (*Vitis vinifera*), sweet orange (*Citrus sinensis*), and woodland strawberry (*Fragaria vesca*), using mRNA/ESTs spliced alignment. Accurate spliced alignment of the transcripts and identification of these AS events allows for further functional characterization to reveal the role of these identified spliced transcripts played in the regulation process, which can pave the way for understanding the physiological events in fruit plants.

2. Materials and methods

2.1. Genome sequences and transcripts

For the prediction of the comparative AS events, respective genome sequences including gene models for four fruit plant species were downloaded from different data sources respectively. Briefly, the genome sequence for sweet orange data were downloaded from Citrus genome database (http://citrus.hzau.edu.cn/orange/download/data.php) [32], woodland strawberry and apple plant data were downloaded from GDR database (the Genome Database for Rosaceae) (http://www.rosaceae.org/species/fragaria/fragaria_vesca/genome_v1.0; and http://www.rosaceae.org/species/malus/malus_x_domestica/genome_v1.0) [33,34], grape genome was downloaded from Phytozome database (ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Vvinifera/annotation/) [36]. The assembled ESTs and mRNA transcripts of orange and grape plants were downloaded from the PlantGDB database (http://www.plantgdb.org/prj/ESTCluster/) [36]. The strawberry and apple plant mRNA and ESTs were downloaded from the ESTs and nucleotide database at National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/) and assembled in-house using the CAP3 program with the following parameters: −p 95–o 50–g 3–y 50–t 1000 [37].

2.2. Putative unique transcripts (PUTs) to genome mapping, identification and functional annotation of AS isoforms

The PUTs were mapped to their corresponding genomes using ASFinder (http://proteomics.ysu.edu/tools/ASFinder.html/) [38]. ASFinder uses SIM4 program [39] to align PUTs to the genome and then parse the SIM4 output file to generate a file with gene transfer format (gtf) and also extract those PUTs that are mapped to the same genomic location but have variable exon-intron boundaries. The output file (AS.gtf) of ASFinder was then subsequently submitted to AStalavista server (http://genome.crg.es/astalavista/) for AS event analysis [40]. To avoid the call of the spurious alternative splicing events, we applied a threshold of minimum of 95% identity of aligned PUT with a genomic sequence, a minimum of 80 bp aligned length, and >75% of a PUT sequence aligned to the genome [13]. Application of the above identity percentage and the aligned length minimizes the chance of the false positive AS events calling as a result of gene and genome duplication events. The percentage of alternative spliced genes was estimated using the genome predicted gene models with the spliced genes having at least one PUT spliced alignment.

The assembled PUTs were further annotated for their coding regions using the ORFPredictor [41] and the full–length transcript coverage was assessed using TargetIdentifier [42]. Functional classification was assigned to the PUTs by performing BLASTX searches against UniProtKB/Swiss-Prot with a cutoff E-value of 1E-5. The predicted protein sequences from ORFPredictor were further functionally annotated using rpsBLAST against the PFAM database (http://pfam.xfam.org/). Following the mapping, the exonic and the intronic boundaries were extracted from the AS mapping files and the sequence logo for the intronic and the exonic boundaries were made using the Web Logo 3 available from http://weblogo.threeplusone.com [43]. For the phylogenetic representation, *A. thaliana* SR proteins was used as a query across the sequenced genomes from each clade available from Phytozome [35] and protein alignment was done using MSAProbs [44], followed by RAxML ancestral phylogenetic analysis using RAxML version 8 [45] with PROTCATWAG model. Phylogenetic tree was rooted using *Ambroella trichopoda* as a basal angiosperm.

2.3. Data access and visualization of AS

AS events identified in this study along with the integrated genomic tracks are available from Plant Alternative Splicing Database (http://proteomics.ysu.edu/altsplice/) [13,23]. The user interface allows choosing a species and then searching the database using a PUT ID, gene ID, keywords in functional annotation; PFAM; or AS event types. Additionally; the identified AS events can be visualized and compared with predicted gene models using GBrowse for comparative assessment. BLASTN search for the PUTs and AS isoforms is also supported. The assembled sequence data with annotation information and other related intermediate data files are publicly available for downloading at: http://proteomics.ysu.edu/publication/data/FruitAS/

3. Results and discussion

3.1. Analysis and annotation of PUTs

Genome-wide analyses of alternative splicing have established its nearly ubiquitous role in gene regulation in many organisms [46]. Identification of spliced transcripts plays an important role in understanding the ecotypic responses and also has played fundamental role in understanding the regulome of plant species [1–4]. It is noteworthy to highlight that the previous estimates using ESTs/mRNAs mapping based approaches provided relatively accurate results of the AS events, where high resolution based mRNA-seq is lacking or in sequenced species where the genome fragmentation is largely present [47]. In the present research, we used putative unique transcripts (PUTs) for genome mapping to unravel the splicing diversity in four fruit species and presented a