November 2016. Horticultural Plant Journal, 2 (6): 338–350.

Horticultural Plant Journal

Available online at www.sciencedirect.com
The journal's homepage: http://www.journals.elsevier.com/horticultural-plant-journal

Landscape and Fruit Developmental Regulation of Alternative Splicing in Tomato by Genome-Wide Analysis

WANG Ketao [a,1], JIAO Zhicheng [a,1], XU Meng [a,1], WANG Yongsheng [b], LI Ren [a], CUI Xia [a], GU Lianfeng [b,*], and ZHANG Shuaibin [a,*]

[a] *Key Laboratory of Biology and Genetic Improvement of Horticultural Crops of the Ministry of Agriculture, Sino-Dutch Joint Laboratory of Horticultural Genomics, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China*
[b] *Haixia Institute of Science and Technology (HIST), Fujian Agriculture and Forestry University, Fuzhou 350002, China*

Received 15 August 2016; Received in revised form 19 September 2016; Accepted 7 November 2016

Available online 9 January 2017

Abstract

In eukaryotes, alternative splicing (AS) is one of the posttranscriptional regulatory mechanisms that play important roles by generating transcriptome diversity. To obtain a global view of AS and its dynamics during tomato fruit development, we analyzed the AS events using a large amount of transcriptome datasets. Same with other plant species, about half of the expressed multiexonic genes were alternatively spliced in tomato. Besides that, our further analyzation of RNA-seq datasets of ovule and pericarp at early fruit developmental stages reveals that the dynamic alteration of AS events occurred in specific tissues and AS was regulated spatially and temporally during early fruit development in tomato. By investigating the sequence variations at splice sites causing differential AS events between tomato cultivar 'Moneymaker' and wild species *Solanum pimpinellifolium* PI365967, we uncover that AS may play the regulatory roles during domestication of tomato. Taken together, our results provided the global AS pattern in tomato and highlighted the importance of AS during tomato fruit development and domestication.

Keywords: tomato; alternative splicing (AS); fruit development; domestication

1. Introduction

Alternative splicing (AS) is a crucial posttranscriptional regulatory mechanism that modulates gene expression and increases the diversity of proteome in all eukaryotic organisms (Lareau et al., 2004). Most of the AS events affect the coding sequence to produce proteins that differ in their sequence and domain arrangement and then lead to diversities of localization, stability and biological function (Stamm et al., 2005). In addition, AS also regulates transcription through producing premature termination codons (PTCs) and committing the transcripts to degradation via nonsense-mediated decay (NMD) (Lareau et al., 2007). There are four main types of AS events including exon skipping (ES), intron retention (IR), alternative donor sites (AD) and alternative acceptor sites (AA). The most common form of AS in human is the ES (Wang et al., 2008), whereas the IR is the main type in plant species, such as *Arabidopsis*, rice, soybean and maize (Wang and Brendel, 2006; Filichkin et al., 2010; Zhang et al., 2010; Marquez et al., 2012; Shen et al., 2014; Thatcher et al., 2014). AS is catalyzed by spliceosomes constituted by massive ribonucleic protein (Saltzman et al., 2011), and regulated by trans-acting proteins (Deng et al., 2010; Erkelenz et al., 2013) or corresponding cis-acting regulatory sites (silencers and enhancers) on the pre-mRNA (Pertea et al., 2007). Moreover, other factors may also influence AS, such as histone modification (Luco et al., 2010), chromatin environment (Kornblihtt et al., 2004) and transcription factors recruited to the promoters (Auboeuf et al., 2004).

In humans, an estimate of more than 95% transcripts from multiexonic genes are alternatively spliced by genome-wide analysis of expressed sequence (Pan et al., 2008). Disorder of AS is

* Corresponding authors. Tel.: +86 10 82109541; +86 591 86392267
E-mail addresses: lfgu@fafu.edu.cn; zhangshuaibin@caas.cn
[1] These authors contribute equally to this work.
Peer review under responsibility of Chinese Society for Horticultural Science (CSHS) and Institute of Vegetables and Flowers (IVF), Chinese Academy of Agricultural Sciences (CAAS)

http://dx.doi.org/10.1016/j.hpj.2017.01.007

well known to correlate with numerous diseases (Soleymanlou et al., 2005; Tan et al., 2005; Ule et al., 2005; Agrawal and Eng, 2006; Zhong et al., 2006). For example, more than 50% of AS events in ovarian and breast tissue are altered in tumors (Tazi et al., 2009; Venables et al., 2009). Unlike in animals, AS is underestimated in plants for a long period until next-generation sequencing (NGS) technologies and comprehensively tissue-specific transcriptome were applied to identify the alternative spliced genes. By now, 40%–60% intron-containing transcripts are alternatively spliced in *Arabidopsis*, rice, maize and soybean (Zhang et al., 2010; Marquez et al., 2012; Shen et al., 2014; Thatcher et al., 2014).

In plants, AS also plays an important role in many aspects of physiological processes. The variation of splice sites is one of the mechanisms for crop evolution and domestication. A well-known example is the *Waxy* (*Wx*) gene in rice encoding a granule bound starch synthase. The Wx^b allele contains a G to T mutation at 5′ splice site of the first intron which reduces splicing efficiency and generates less amylose compared with the corresponding Wx^a allele (Cai et al., 1998; Isshiki et al., 1998; Larkin and Park, 1999). Moreover, alternatively spliced isoforms are differentially expressed in a tissue- or developmental stage-specific manner to participate in many developmental processes such as flowering, circadian clock and stress response as an endogenous cue (Palusa et al., 2007; Reddy, 2007; Schindler et al., 2008; Filichkin et al., 2010; Wang et al., 2014), indicating that AS is crucial for regulating plant development. However, functions of AS events in other developmental processes of plants are still unclear.

Tomato is a model plant for fleshy fruit development in which many signals and factors are involved. *AUXIN RESPONSE FACTOR* (*ARF*) genes which function in auxin signaling pathway play crucial roles during tomato fruit set and formation (Kumar et al., 2011). About one third of tomato *ARF* family members were found to be alternatively spliced during fruit set (Zouine et al., 2014), suggesting that AS may play important regulatory roles in tomato fruit development. A recent study and our work indicate that about half of the expressed multiexonic genes in tomato were alternatively spliced although different RNA-seq datasets were used (Sun and Xiao, 2015). Moreover, we revealed that AS regulates early fruit development in a spatial and temporal manner by identification of the ovule and pericarp specific AS genes and exploration of the dynamics of AS during early fruit development of tomato. We further study sequence variations at splice sites which induce differential AS events during fruit development between cultivar ‘Moneymaker’ and its wild relative PI365967 and uncovered that some AS might contribute to tomato natural variations during domestication. In brief, our results reveal a landscape of AS in tomato fruit development, which will help us to further explore functions of AS in this process.

2. Materials and methods

2.1. Plant growth and material collection

Tomato plants (*Solanum lycopersicum* ‘Moneymaker’, *S. lycopersicum* var. *cerasiforme* ‘LA1310’ and *S. pimpinellifolium* ‘PI365967’) were grown on soil in greenhouse during autumn of 2015 in Beijing. For the samples of ovule and pericarp at early developing stages, the opening flowers were marked and the ovaries and little fruits were collected at −2 days post anthesis (DPA), 1 DPA, 2 DPA and 5 DPA. Samples of 0 DPAe (‘e’ stands for ‘emasculated’) and 2 DPAe were emasculated at −2 DPA first and collected at 0 DPA and 2 DPA respectively. The ovule and the pericarp were dissected under a microscope and frozen in liquid nitrogen for further use. For fruits of Moneymaker, LA1310 and PI365967 at different stages, the opening flowers were marked and the fruits at 5 DPA, 20 DPA, mature green stage, break stage and red ripening stage as well as the flowers at −2 DPA were harvested and frozen in liquid nitrogen immediately.

2.2. RNA extraction, library construction and sequencing

The RNA was extracted using TRIzol reagent (Invitrogen, 15596-026). The contaminant DNA was removed using TURBO DNA-free™ Kit (Ambion, AM1907). For fruits of Moneymaker, LA1310 and PI365967, RNA from different stages of each species was mixed in the same amount for library construction. Transcriptome libraries were constructed using dUTP method (Parkhomchuk et al., 2009). By this principle, the transcript direction and strand-specific information were reserved. Then the RNA-seq libraries were sequenced separately on the Illumina HiSeq 2500 platform with a pair-end 125 bp (for the ovule and pericarp samples) or a single-end 50 bp (for fruit samples of Moneymaker, LA1310 and PI365967) sequencing strategy.

2.3. RNA-seq data analysis and AS event identification

Tomato genome sequences and gene annotation (ITAG 2.4) were downloaded from ftp://ftp.solgenomics.net/tomato_genome/. RNA-seq reads were aligned to the genome using Tophat-2.0.11 (Kim et al., 2013) with default parameters. The segment mismatch was set to no more than 2. The expression levels for gene models from ITAG 2.4 were measured and normalized as reads per kilobase of exon model per million mapped reads (RPKM) (Mortazavi et al., 2008). Cufflinks (2.1.1) were used to assemble above alignment reads into the transcripts (Trapnell et al., 2010). The minimum isoform fraction for transcripts was set to 0.05. The percentage of reads with small anchor fraction was set to 0.01. The minimum and maximum intron lengths were set to 30 and 15 000 bp respectively. Different types of AS events in each tissue were identified from assembled transcripts using ASTALAVISTA algorithm (Foissac and Sammeth, 2007). Four main types of AS events including IR, ES, AA, and AD were analyzed as previously defined (Sammeth et al., 2008). GO analysis of AS genes was performed using their best homologous genes in *Arabidopsis* with DAVID (The Database for Annotation, Visualization and Integrated Discovery, https://david.ncifcrf.gov/) (Huang et al., 2009). Visualization of reads densities from RNA-seq was performed using the Integrated Genome Browser (Nicol et al., 2009).

2.4. Validation of AS events

A total of 95 AS events in four different AS types distributed across 12 chromosomes were chosen to validate the accuracy