BMB
Reports

BMB Rep. 2013; 46(9): 439-447
www.bmbreports.org

Invited Mini Review

Splicing and alternative splicing in rice and humans

Zhiguo E[1], Lei Wang[1] & Jianhua Zhou[2,]*
[1]China National Rice Research Institute, Hangzhou, 310006, [2]Nantong University, Nantong 226001, P.R. China

Rice is a monocot gramineous crop, and one of the most important staple foods. Rice is considered a model species for most gramineous crops. Extensive research on rice has provided critical guidance for other crops, such as maize and wheat. In recent years, climate change and exacerbated soil degradation have resulted in a variety of abiotic stresses, such as greenhouse effects, lower temperatures, drought, floods, soil salinization and heavy metal pollution. As such, there is an extremely high demand for additional research, in order to address these negative factors. Studies have shown that the alternative splicing of many genes in rice is affected by stress conditions, suggesting that manipulation of the alternative splicing of specific genes may be an effective approach for rice to adapt to abiotic stress. With the advancement of microarrays, and more recently, next generation sequencing technology, several studies have shown that more than half of the genes in the rice genome undergo alternative splicing. This mini-review summarizes the latest progress in the research of splicing and alternative splicing in rice, compared to splicing in humans. Furthermore, we discuss how additional studies may change the landscape of investigation of rice functional genomics and genetically improved rice. [BMB Reports 2013; 46(9): 439-447]

INTRODUCTION: GENETICS OF RICE, MAIZE AND WHEAT

Rice (*Oryza Sativa L.*), maize and wheat are the three most produced and consumed crops worldwide. Rice is a diploid plant with 12 pairs of chromosomes, and approximately 382 million base pairs (382 Mb) in its genome. Rice has roughly 55,000 genes (1). Similar to rice, maize is also a diploid plant. However, the core genome of maize is 4-5 times larger than rice, with 10 pairs of chromosomes, roughly 2,000 Mb and 110,000 genes (2). In contrast to rice and maize, wheat is an allohexaploid plant with 21 pairs of chromosomes, and 17,000 Mb in its genome. It is predicted that wheat has roughly 94-96 thousand genes (3).

Rice, maize and wheat all originated from a species with only five pairs of chromosomes. During the evolutionary process, genome replication and chromosome translocation and fusion led to the formation of 12 intermediate chromosomes. Rice, wheat, maize and other crops were then gradually differentiated out from this basis. As a result, rice retained the original 12 intermediate states chromosomes, while loss and and/or integration of the original 12 chromosomes led to the current genome of wheat and maize (4). Therefore, although the rice genome is much smaller than that of maize or wheat, it has retained the genetic diversity and fingerprint of its ancestral species. The only major difference is that the repeat sequences in rice are not as large as those in wheat or maize. In light of evolutionary conservation and genome miniaturization, rice has become a model species for the genomic research of gramineous crops.

RNA SPLICING AND ALTERNATIVE SPLICING

RNA splicing is a biological process that removes introns from pre-mRNA, and ligates exons together (5). Since its discovery more than 35 years ago (6, 7), RNA splicing has been extensively studied, particularly in vertebrate animals. In humans, there are approximately 3,200 million base pairs of DNA; however, 98.5% are not transcribed. The remaining 1.5% of DNA contains approximately 25,000-35,000 genes (8-12). The size of genes varies, with an average estimate of 30,000 base pairs (bp) per gene in humans. Many genes are larger than 100,000 bps, with the largest known gene dystrophin being approximately 2.4 million base pairs (8). Genes are transcribed from DNA into pre-mRNAs, which are normally longer than 30,000 nucleotides. However, the average size of mature mRNA that codes a protein is usually shorter than 2,000 nucleotides (8). The size discrepancy between pre-mRNAs and mature mRNA is explained by the so-called RNA splicing process, in which a large portion of a pre-mRNA is trimmed. More interestingly, it has been determined that many individual exons or introns may be included or excluded in some mRNAs, but not in other mRNAs, through an alternative splicing (AS) process, leading to generation of multiple protein isoforms from a single gene. It has been estimated that more than 95% of multiple-exon pre-mRNAs in humans un-

*Corresponding author. Tel: +1-617-953-4160; Fax: +1-617-505-1856; E-mail: jianhua@ntu.edu.cn; jianhua55@msn.com

http://dx.doi.org/10.5483/BMBRep.2013.46.9.161

Received 10 July 2013

Keywords: Alternative splicing, Next generation sequencing (NGS), Rice, Splicing, Splicing elements

ISSN: 1976-670X (electronic edition)

dergo alternative splicing (12), exponentially increasing biological information flow in cellular processes, leading to an estimated 90,000 protein species in humans, despite there being only roughly 25,000 genes (13). Extensive studies have demonstrated that splicing and alternative splicing regulate almost every biological process, including signal transduction and energy transfer in metazoan and plants. A variety of diseases in humans have been found to be caused by defects in pre-mRNA splicing (5, 14). Correction of defective splicing has recently become a target for the treatment of such diseases in humans. However, in contrast to humans and other vertebrates, studies of RNA splicing and alternative splicing in plants, such as in rice, are limited. But the mechanisms of splicing in both metazoans and plants are believed to be similar, although many differences are known to exist in actual splicing between metazoans and plants, as discussed below.

Pre-mRNA splicing is a two-step process that involves the formation of phosphodiester bonds (15). The first step involves attack of the phosphate group at the 5′ splice site on the hydroxyl group of the adenine at branch point, so that a 2′-5′ phosphodiester bond is formed. The second step engages the phosphate group at the 3′ splice site that attacks the hydroxyl group of the 5′ splice site, to form 3′-5′ phosphodiester bond. The splicing process is not complete until the cleaved exons are ligated together, to form a mature RNA. RNA splicing is a heavily regulated biological process that is dependent on sequence elements in pre-mRNAs. These sequences, termed splicing signals, include such elements as the 5′ splice site, 3′ splice site, branch point and polypyrimidine tract (16). In contrast to metazoans, the branch point and polypyrimidine tract are less conserved in plants. In addition to these essential elements, the pre-mRNA trimming of some introns and inclusion/skipping of some exons are regulated by enhancers or inhibitors in the pre-mRNA sequence (16, 17). The overall coordination among splicing signals, enhancers, and inhibitors, as well as other components that are discussed below, leads to the precise and orchestrated event of pre-mRNA splicing and alternative splicing.

Pre-mRNA splicing occurs in the spliceosome, a large RNA protein complex (18). The spliceosome is a dynamic structure that undergoes multiple complex transitions during the splicing process. While slicing signals in pre-mRNA are required for the splicing of specific exons, U1, U2, U4, U5 and U6 snRNPs (small nuclear RNA protein complexes), as well as U2AF65 and serine-arginine rich (SR) proteins, are essential components that make up the spliceosome (19-21). Assembly of the spliceosome begins with the formation of the first complex, complex E as the early spliceosome, whereby U1 snRNP is recruited to the pre-mRNA and U1 snRNA in U1 snRNP base pairs, with the 5′ splice site of pre-mRNA. Subsequently, a pre-spliceosome complex is formed. In the pre-spliceosome, U2 snRNP is recruited to the pre-mRNA and the U2 snRNA in U2 snRNP base pairs, with the branch-point in the pre-mRNA. Recruitment of U4/U5/U6 snRNPs to the pre-spliceosome leads to maturation of the spliceosome complex, which then proceeds to trim introns and ligate exons, respectively.

Splicing and alterative splicing in humans and in rice

Pre-mRNA splicing was initially described in adenovirus 2 late mRNA (6, 7). Subsequently, it was determined that RNA splicing is a universal event that occurs in all organisms. However, the type and mechanism of splicing varies among species. In prokaryotes, splicing is a rare event that occurs in non-coding RNAs, such as tRNAs (22). On the other hand, in eukaryotes, splicing is mostly referred to as trimming introns and the ligation of exons in protein-coding RNAs. Another major difference in splicing between prokaryotes and eukaryotes is that splicing in prokaryotes does not involve a spliceosome. The frequency of RNA splicing depends on the complexity of gene structures of genomes in particular species. Generally, many more RNA splicing events occur in higher species, such as mammals, compared to lower species, such as single cell organisms like *Saccharomyces cerevisiae* (yeast) (Table 1). Approximately 95% of genes in yeast have a single exon without introns. As such, splicing is not necessary in these genes. The remaining 5% of genes in yeast have either one intron or two introns, suggesting that pre-mRNA splicing in yeast is not as complicated, as it is in other species. On the other hand, >80% of genes in humans and rice have multiple introns. In fact, genes in humans have an average of 8-10 exons (23).

Table 1. Genes, exons and introns in yeast, rice and humans

	Yeast	Rice	Humans
Total genes	6,000	50,000-60,000	25,000-40,000
Intronless genes	>5,700	11,109	6,227
Percentage of intronless genes	95%	~20%	~20%
Average internal exon size	N/A	300 bp-500 bp	140 bp-180 bp
Average numbers of exons	<2	4-5	8-10
Alternative splicing of multiple-exon genes	Rare	>50%	>95%
Splicing mechanism	Spliceosome mediated	Intron definition	Exon definition
Average intron size	<200 bp	300 bp-550 bp	3,500 bp-5,500 bp

Based on previous publications (8, 23, 38, 54-57) and MSU Rice Genome Annotation, Release (http://rice.plantbiology.msu.edu/analyses_facts.shtml).