

## Invited Mini Review

## Splicing and alternative splicing in rice and humans

Zhiguo E<sup>1</sup>, Lei Wang<sup>1</sup> & Jianhua Zhou<sup>2,\*</sup><sup>1</sup>China National Rice Research Institute, Hangzhou, 310006, <sup>2</sup>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

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 alternative 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](http://rice.plantbiology.msu.edu/analyses_facts.shtml)).

Therefore, most genes in humans undergo splicing, to generate mature mRNA.

The human genome contains approximately 25,000-35,000 genes, a number that is much smaller than initially estimated. One explanation for the smaller number of genes in humans, is that many genes undergo alternative splicing (AS), leading to multiple protein isoforms from a single gene, which in turn results in a much greater number of proteins, compared to genes. Alternative splicing is an event that is found in both metazoans and plants. Since its discovery in early 1980, during characterization of the immunoglobulin mu and calcitonin gene (24, 25), alternative splicing has been described in many genes of different species. With the emergence of new technologies, and the rapid growth of expressed sequence tag (EST) and cDNA sequence libraries during the last two decades, it is now estimated that more than 95% of genes with multiple exons in humans undergo alternative splicing (Table 1) (12). Alternative splicing plays an important role in the regulation of gene expression, by affecting mRNA stability, through nonsense-mediated decay (NMD) and translation efficiency (12). Alternative splicing events are heavily regulated at different developmental stages, in different tissues, in different cell types and under different conditions. Abnormal alternative splicing has been implicated in a number of human diseases, such as cancers (breast and lung cancers via the Bcl-x gene) (26), neurodegenerative diseases (spinal muscular atrophy via SMN splicing) (27), frontotemporal dementia with parkinsonism-17 (FTFP-17) via tau splicing) (28), and other diseases (14). Strategies have been developed to target abnormal alternative splicing in these diseases, as potential treatments (5, 28). In fact, several clinical trials are underway, to evaluate the potential to treat diseases by correcting aberrant splicing in humans.

Similar to humans, the majority of genes in rice contain multiple exons and introns, therefore requiring splicing to generate mature mRNAs. Although mechanisms of splicing through the spliceosome have been well characterized in humans, as described above, spliceosomes have yet to be isolated from rice, or other plants. Nonetheless, it is assumed that rice has identical splicing machineries, as in humans. But unlike in humans, studies of alternative splicing in rice have not been as extensive, and most studies have mainly focused on individual genes or gene families. Described in the 1980s (29), ribulose biphosphate carboxylase/oxygenase activase in spinach and Arabidopsis were the first examples of alternative spliced genes in plants. Subsequently, alternative splicing has been found in many other genes in plants, from such species as Arabidopsis and rice. Most of such genes are involved in splicing, transcription, flowering regulation, disease resistance, enzyme activity and other biological processes, in response to conditions such as stress and salinization (30). Specific examples of these genes in rice include homeobox genes (31), waxy gene (32), starch synthase genes (33), OsMET1 genes (34), MYB genes (35), and the HKT transporter gene (36). Although alternative splicing has been described in many of these genes

in different plant species, it was believed that AS in plants was not common. With growing collections of ESTs, and more extensive coverage of plant genomes, researchers have begun to reevaluate splicing and alternative splicing in rice and other plants. In addition, recent advancements of microarray technologies and next generation sequencing have provided extra tools and greater opportunities to examine the depth of splicing and alternative splicing in rice (30, 37-39). As a result, numerous papers published during the last few years have demonstrated extensive alternative splicing events in rice. In one of such papers, Zhang *et al.* (38) described a transcriptome atlas from different organs of cultivated rice, and found that alternative splicing events occur in 33% of rice genes. A higher percentage of genes that undergo alternative splicing in rice has been reported by other groups (39). In addition, these groups found that 58% of AS-related genes have multiple AS events, leading to numerous transcripts from a single gene. Interestingly, they also demonstrated that 59% of the AS events in rice were organ-specific, suggesting that AS plays a role in the functional complexity of rice. Gene structure analyses of AS genes indicated that 76% of AS events in rice occurred in protein-coding region. While 48.5% of AS causes frame shifts, or produces premature termination codons, potentially leading to nonsense-mediated mRNA decay (NMD), 27.5% of AS events retain the same reading-frame. These results suggest that, similar to humans, AS also plays an important role in gene regulation, protein diversity and complexity in rice.

## REGULATION OF SPLICING AND ALTERNATIVE SPLICING IN HUMANS AND RICE

A majority of the 25,000 genes in humans and the 50,000 genes in rice contain multiple exons and introns. Therefore, to produce mature mRNAs, splicing is required in both humans and rice. Humans and rice share a number of similarities with regard to the regulation of splicing and alternative splicing, which are controlled by splicing signals, cis-elements and trans-factors. One of the most important similarities is that both humans and rice have similar consensus 5' and 3' splicing sites, with C(A)AG/GTAA and TGCAG/G as the donor and acceptor, respectively. These two sites guarantee precise cleavage between introns and exons. In addition, similar to humans, rice contains many homologues of Uridylate-Rich Small Nuclear RNAs (UsnRNAs) and protein components of spliceosomes, indicating that splicing in rice likely occurs via a spliceosome complex.

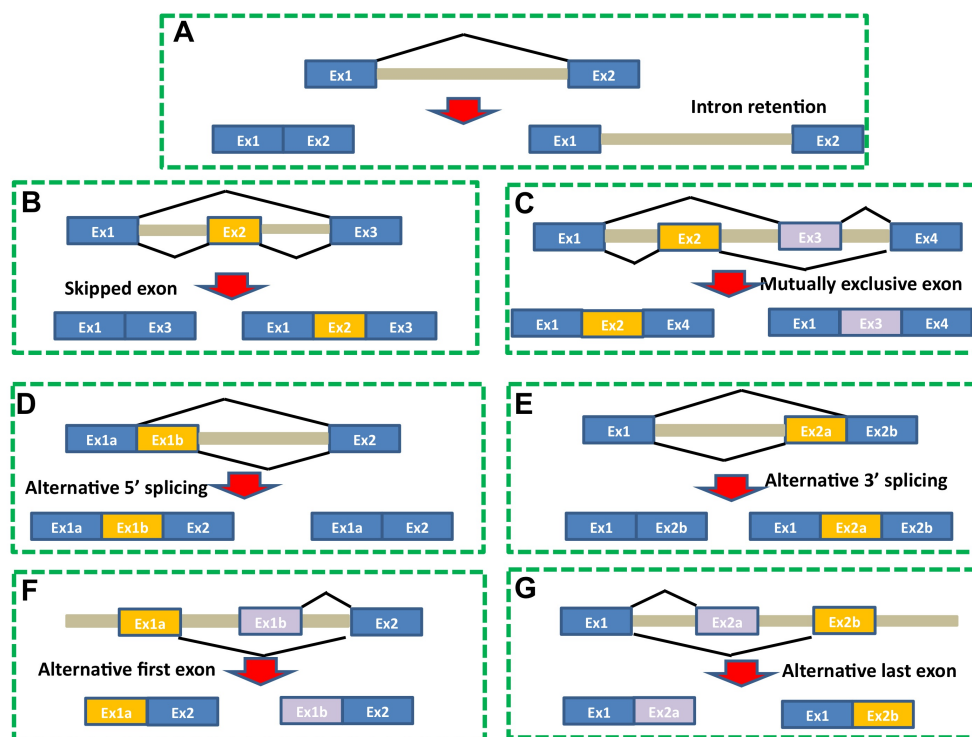
However, several major differences with regard to the regulation of splicing also exist between humans and rice. First, human genes typically have massive size variations in introns and exons, with an average intron of 3500-5500 bps, and small internal exons (140-180 bps) (8, 23). In contrast, introns in rice genes are typically much smaller, and exons can be as large as those in humans (38, 39) (Table 1) (MSU Rice

Genome Annotation, Release 7: [http://rice.plantbiology.msu.edu/analyses\\_facts.shtml](http://rice.plantbiology.msu.edu/analyses_facts.shtml)). Second, a branch point is usually required for the splicing of pre-mRNA in humans. Although there seems to be a branch point upstream of the 3' splicing site in rice, the site is not conserved. Third, a polypyrimidine tract between the branch point and the 3' splicing site is a major feature in regulating the splicing of pre-mRNAs in humans. Instead of a polypyrimidine tract, a U-rich sequence in rice pre-mRNA was discovered, in place of a polypyrimidine tract (38–40).

Due to these similarities and differences in splicing signals, regulatory elements and gene structures, one would expect similarities and differences in splicing patterns between humans and rice. While constitutive splicing in both humans and rice complies equally well with splicing rules via spliceosome machineries, differences in the types of alternative splicing and extent of alternative splicing are also noticeable. More than 95% of human genes with multiple exons undergo alternative splicing, while only 50% of rice genes are estimated to have alternative splicing. One explanation for this difference is that the coverage of rice transcriptome and collections of ESTs in rice are not yet as extensive as in humans. However, more exons and introns in an average human gene (Table 1) may also contribute to the difference, that humans have a higher per-

centage of genes with alternative splicing.

Seven types of alternative splicing have been observed in both humans and rice (Fig. 1), including exon skipping, intron retention, 5' alternative splicing; 3' alternative splicing, mutually exclusive exons, alternative first exons, and alternative last exons. The most common type of alternative splicing in humans is exon skipping, which comprises more than 42% of total AS events (41). Exon skipping occurs when a middle exon is skipped, leading to two separate exons joining together. Conversely, making up only 9% of total AS events in humans, intron retention (IR) occurs when an intron is not spliced out (41). In contrast to humans, intron retention is a common type of alternative splicing in rice (>47%), whereas less than 25% of alternative splicing in rice involves exon skipping. Interestingly, it has been previously reported that the average size of retained intron is only 183 bp, which is much smaller than the average size of introns (450 bp) in rice, suggesting that intron-retention alternative splicing is related to the size of introns. Such results are consistent with the hypothesis that organisms such as rice with average small introns, use intron-definition splicing mechanisms, by which introns are initially recognized by spliceosome; whereas organisms like humans with large introns, use exon-definition splicing mechanisms, by which exons are first recognized by spliceosomes.



**Fig. 1.** Seven types of alternative splicing. (A) Intron retention (IR), (B) Exon skipping, (C) Mutually exclusive exons, (D) Alternative 5' splicing, (E) Alternative 3' splicing, (F) Alternative first exons, (G) Alternative last exons. Intron retention is the major alternative splicing event in rice, whereas exon skipping is the most frequent alternative splicing in humans.

One other aspect of alternative splicing between humans and rice that needs to be discussed is cis regulatory elements. Numerous exon splicing enhancers (ESE), exon splicing silencers (ESS), intron splicing enhancers (ISE) and intron splicing silencers (ISS) have been identified and characterized in many genes in humans (5). These elements are primarily targeted by splicing trans-factors for positive (SR proteins) or negative regulation (hnRNPs: heterogeneous nuclear ribonucleoproteins) of alternative splicing of specific genes. While potential ESEs have been described in Arabidopsis (42), ESEs, ESSs, ISEs and ISSs have not been well studied in rice.

Finally, it should be noted that Zhang *et al.* (38) reported that similar to humans, rare events such as trans-splicing can still occur in rice. Trans-splicing is coined as a splicing event in which chimeric RNAs comprise exons from two or more different genes (38, 43). Although this is thought to be a rare occurrence, the number of trans-splicings may be higher than expected. Identification and characterization of trans-splicing RNAs may provide insight on how genes are regulated in both humans and rice.

## TRANS SPlicing ELEMENTS IN HUMANS AND RICE

Regulation of splicing and alternative splicing is an exquisite biological process that requires not only splicing signals, cis-elements and gene structures, but also many protein and RNA components, to form spliceosomes. In addition, trans-elements such as SR proteins and hnRNPs are essential for the regulation of alternative splicing in different organs, at different developmental stages, and under different conditions. Among trans splicing elements, SR proteins are the most well studied group of splicing factors (16, 18). SR proteins are a family of splicing regulators that contain long repeats of serine (S) and arginine (R) amino acid residues (RS domain) and 1-2 RNA recognition motifs (RRM) at the N-terminus. SR proteins play key roles in regulating both the constitutive and alternative

splicing of many genes, including their own genes, in all species (29, 44-46). However, until as recently as 2010, it was ambiguous as to which proteins actually constituted member of the SR protein family, because proteins with RS domains but no RRP motifs exist. To simplify classification, Manley and Krainer (46) redefined SR proteins, based on studies in mammals, as proteins with one or two N-terminal RRMs, followed with a downstream c-terminal arginine serine-rich domain (RS domain) (Table 2). A RS domain contains at least 50 amino acids, with >40% RS or SR repeats. However, this definition would have excluded a number of proteins in plants from SR families. Therefore it was later clarified by Barta *et al.* (47), and endorsed by many scientists, including Manley and Krainer themselves, that Manley and Krainer's SR classification fits well to SR proteins in metazoans (humans), but SR proteins in plants are more diversified, with several additional plant (Rice) specific subgroups (Table 2) (47). One of these rice specific subgroups has RRM motif(s) and RS domains, but with two distinguished zinc knuckles and/or an acidic N-terminal extension rich in Ser and Pro residues, whereas the SC35-like protein subgroup in rice contains only one RRM and RS domain, but possesses an N-terminal extension rich in Arg, Pro, ser, Gly and Tyr. In addition, it was found that another rice specific subgroup has two RRMs. However, the conserved SWQDLKD motif is not present in the 2nd RRM of this group of SR proteins. Similar to this classification, in a more recent study Richardson *et al.* (48) identified and described more than 272 SR proteins in 27 different species, from fungi to humans. They analyzed these SR proteins from different species, with the RRM domain as a phylogenetic marker. They classified SR proteins into five major groups, which can be further divided into 11 sub-groups. Five of these sub-groups are mainly composed of plant SR proteins, while the six remaining sub-groups are mainly comprised of mammal SR proteins. Interestingly, one group of SR proteins, SC35, is conserved among many species, possibly due to its roles in both splicing,

**Table 2.** SR proteins in humans and rice (47)

	Rice	Humans
Number of SR proteins	22	12
RS domains	> 50 aas long; > 20% RS content or RS dipeptide	> 50 aas long; > 40% RS content of RS or SR repeat
RRM domains	1-2 Conserved	1-2 Conserved
Additional domains	1-2 zinc knuckle; acidic -terminal extension rich in Ser and Pro	1 zinc knuckle
Not rice specific sub-groups		
SR subfamily	2 RRMs with SWQDLKD motif	SF2/ASF: 2 RRMs with SWQDLKD motif
RSZ subfamily	1 RRM, 1 zinc knuckle	SF7/9G8: 1 RRM, 1 zin knuckle
SC subfamily	1 RRM+SR domain	SF2/SC35: 1 RRM+SR domain
Rice specific sub-groups		
SCL subfamily	1 RRM+ 1 N-terminal charged extension	N/A
SC2Z subfamily	1 RRM+ 2 zinc Knuckles + 1 acidic-terminal extension rich in Ser and Pro	
RS subfamily	2 RRMs but without SWQDLKD motif	

and transcriptional elongation (49).

While SR proteins in humans have been extensively studied by both *in vitro* assays and *in vivo* examination, studies of SR proteins and other trans-factors in rice are primarily based on *in vitro* analysis, and on sequence predictions (50). According to currently available information (50), there seems to be several differences between SR proteins of humans and rice. Due to genome duplication throughout evolution, there is a much higher number of SR proteins in rice, than in humans. In fact, the number of SR proteins in rice ( $n = 22$ ) is almost twice as many as in humans ( $n = 12$ ). However, duplication of the genome in rice has also led to redundancy of the SR proteins. Interestingly, plants, including rice, may have undergone a so-called purifying process, leading to the compromise of biological efficacy, and protein-protein interactions of orthologues. In addition, it is worthy of note that redundant SR genes are not expressed equally, but rather that one SR gene is usually preferably expressed, than its paralogue in rice. Lastly, although the significance of this observation is still unclear, SR genes in rice have a higher percentage of alternative splicing (95%), than those from humans (40-50%) (50).

## OPPORTUNITIES FOR AND CHALLENGES TO SPLICING STUDIES IN RICE

Splicing is a universal biological event that occurs in almost all organisms. It is presumed that mechanisms to regulate splicing in both humans and rice are identical, and primarily occur through the spliceosome machineries. However, despite the rapid progress of studies in alternative splicing, the importance of alternative splicing in rice is still poorly understood. It was initially thought that alternative splicing in rice regulates gene expression, by generating premature termination codon (PTC), leading to non-sense mediated mRNA decay (NMD) (51, 52). However, more recent studies suggest that the majority of alternative splicing events in rice occur in coding regions (76.5%), leading to the generation of a diverse array of proteins (38, 39). In addition, it has been shown that alternative splicing in rice may be organ-specific, and induced by specific conditions, such as stress (35, 53). These studies suggest that alternative splicing in rice is not only required for the regulation of gene expression, but perhaps also for general diversification of the proteome (38, 39). Advancements in next generation sequencing and analytical tools have led to significant strides in our understanding of the mechanisms by which alternative splicing occurs in rice and other plants, ultimately leading to better understanding of the relationship between alternative splicing and the regulation of biological processes.

Due to climate change and exacerbated soil degradation over recent years, many challenges remain in the study of rice, in order to cope with a variety of abiotic stresses, and many other hostile conditions. Traditional genetic studies to improve the yield and quality of rice have reached the level at which additional improvements have become increasingly difficult.

On the other hand, genomic studies, including gene expression and splicing, may provide novel insights. Although continued investigation of gene expression, signal pathways and alternative splicing may revolutionize how rice is cultivated, much more work is needed to understand how gene expression, splicing and alternative splicing are related to the yield, quality and nutrition of rice, and to stress conditions.

## Acknowledgements

This work was supported by the Special Fund for Agro-scientific Research in the Public Interest (201203029, 201203031), and the Special Foundation for Basic Research and Development of Central Level Scientific Research Institutes in China (2012RG006) (ZGE, LW), and by the Priority Academic Program Development of Jiangsu Higher Education Institution (PAPD) (JZ).

## REFERENCES

1. Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., Schwartz, D. C., Tanaka, T., Wu, J. Z., Zhou, S. G., Childs, K. L., Davidson, R. M., Lin, H. N., Quesada-Ocampo, L., Vaillancourt, B., Sakai, H., Lee, S. S., Kim, J., Numa, H., Itoh, T., Buell, C. R. and Matsumoto, T. (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* **6**, 4.
2. [http://www.maizesequence.org/Zea\\_mays/Info/StatsTable?db=core](http://www.maizesequence.org/Zea_mays/Info/StatsTable?db=core), date cited: 2013-09-03.
3. Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G. L., D'Amore, R., Allen, A. M., McKenzie, N., Kramer, M., Kerhornou, A., Bolser, D., Kay, S., Waite, D., Trick, M., Bancroft, I., Gu, Y., Huo, N., Luo, M. C., Sehgal, S., Gill, B., Kianian, S., Anderson, O., Kersey, P., Dvorak, J., McCombie, W. R., Hall, A., Mayer, K. F., Edwards, K. J., Bevan, M. W. and Hall, N. (2013) Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* **491**, 705-710.
4. Eckardt, N. A. (2008) Grass genome evolution. *Plant Cell* **20**, 3-4.
5. Zhou, J., Zheng, X. and Shen, H. (2012) Targeting RNA-splicing for SMA treatment. *Mol. Cells* **33**, 223-228.
6. Chow, L. T., Gelinis, R. E., Broker, T. R. and Roberts, R. J. (1977) An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**, 1-8.
7. Berget, S. M., Moore, C. and Sharp, P. A. (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 3171-3175.
8. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrum, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman,

- D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Showkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chisoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramsay, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordliek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kasprzyk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S. and Chen, Y. J. (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.
9. Eichler, E. E., Clark, R. A. and She, X. (2004) An assessment of the sequence gaps: unfinished business in a finished human genome. *Nat. Rev. Genet.* **5**, 345-354.
10. Stein, L. D. (2004) Human genome: end of the beginning. *Nature* **431**, 915-916.
11. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferreira, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hosten, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A. and Zhu, X. (2001) The sequence of the human genome. *Science* **291**, 1304-1351.

12. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. and Blencowe, B. J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413-1415.
13. Kim, E., Magen, A. and Ast, G. (2007) Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res.* **35**, 125-131.
14. Treisman, R., Orkin, S. H. and Maniatis, T. (1983) Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes. *Nature* **302**, 591-596.
15. Chiara, M. D. and Reed, R. (1995) A two-step mechanism for 5' and 3' splice-site pairing. *Nature* **375**, 510-513.
16. Black, D. L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72**, 291-336.
17. Maniatis, T. and Tasic, B. (2002) Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **418**, 236-243.
18. Matlin, A. J., Clark, F. and Smith, C. W. (2005) Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol.* **6**, 386-398.
19. Dreyfuss, G., Matunis, M. J., Pinol-Roma, S. and Burd, C. G. (1993) hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.* **62**, 289-321.
20. Jurica, M. S. and Moore, M. J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* **12**, 5-14.
21. Reed, R. (1996) Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr. Opin. Genet. Dev.* **6**, 215-220.
22. Reinhold-Hurek, B. and Shub, D. A. (1992) Self-splicing introns in tRNA genes of widely divergent bacteria. *Nature* **357**, 173-176.
23. Sakharkar, M. K., Chow, V. T. and Kanguane, P. (2004) Distributions of exons and introns in the human genome. *In Silico Biol.* **4**, 387-393.
24. Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. and Hood, L. (1980) Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways. *Cell* **20**, 313-319.
25. Rosenfeld, M. G., Lin, C. R., Amara, S. G., Stolarsky, L., Roos, B. A., Ong, E. S. and Evans, R. M. (1982) Calcitonin mRNA polymorphism: peptide switching associated with alternative RNA splicing events. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 1717-1721.
26. Lebedeva, I., Rando, R., Ojwang, J., Cossum, P. and Stein, C. A. (2000) Bcl-xL in prostate cancer cells: effects of over-expression and down-regulation on chemosensitivity. *Cancer Res.* **60**, 6052-6060.
27. Lorson, C. L., Hahnen, E., Androphy, E. J. and Wirth, B. (1999) A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 6307-6311.
28. Zhou, J., Yu, Q. and Zou, T. (2008) Alternative splicing of exon 10 in the tau gene as a target for treatment of tauopathies. *BMC Neurosci.* **9** (Suppl 2), S10.
29. Long, J. C. and Caceres, J. F. (2009) The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.* **417**, 15-27.
30. Wang, B. B. and Brendel, V. (2006) Genomewide comparative analysis of alternative splicing in plants. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 7175-7180.
31. Tamaoki, M., Tsugawa, H., Minami, E., Kayano, T., Yamamoto, N., Kano-Murakami, Y. and Matsuoka, M. (1995) Alternative RNA products from a rice homeobox gene. *Plant J.* **7**, 927-938.
32. Cai, X. L., Wang, Z. Y., Xing, Y. Y., Zhang, J. L. and Hong, M. M. (1998) Aberrant splicing of intron 1 leads to the heterogeneous 5' UTR and decreased expression of waxy gene in rice cultivars of intermediate amylose content. *Plant J.* **14**, 459-465.
33. Frances, H., Bligh, J., Larkin, P. D., Roach, P. S., Jones, C. A., Fu, H. and Park, W. D. (1998) Use of alternate splice sites in granule-bound starch synthase mRNA from low-amylose rice varieties. *Plant Mol. Biol.* **38**, 407-415.
34. Yamauchi, T., Moritoh, S., Johzuka-Hisatomi, Y., Ono, A., Terada, R., Nakamura, I. and Iida, S. (2008) Alternative splicing of the rice OsMET1 genes encoding maintenance DNA methyltransferase. *J. Plant Physiol.* **165**, 1774-1782.
35. Li, J., Li, X., Guo, L., Lu, F., Feng, X., He, K., Wei, L., Chen, Z., Qu, L. J. and Gu, H. (2006) A subgroup of MYB transcription factor genes undergoes highly conserved alternative splicing in Arabidopsis and rice. *J. Exp. Bot.* **57**, 1263-1273.
36. Cotsaftis, O., Plett, D., Shirley, N., Tester, M. and Hrmova, M. (2012) A two-staged model of Na<sup>+</sup> exclusion in rice explained by 3D modeling of HKT transporters and alternative splicing. *PLoS One* **7**, e39865.
37. Campbell, M. A., Haas, B. J., Hamilton, J. P., Mount, S. M. and Buell, C. R. (2006) Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. *BMC Genomics.* **7**, 327.
38. 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. and Wang, J. (2010) Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res.* **20**, 646-654.
39. Lu, T., Lu, G., Fan, D., Zhu, C., Li, W., Zhao, Q., Feng, Q., Zhao, Y., Guo, Y., Li, W., Huang, X. and Han, B. (2010) Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. *Genome Res.* **20**, 1238-1249.
40. Jurica, M. S. and Moore, M. J. (2002) Capturing splicing complexes to study structure and mechanism. *Methods* **28**, 336-345.
41. Barbazuk, W. B., Fu, Y. and McGinnis, K. M. (2008) Genome-wide analyses of alternative splicing in plants: opportunities and challenges. *Genome Res.* **18**, 1381-1392.
42. Pertea, M., Mount, S. M. and Salzberg, S. L. (2007) A computational survey of candidate exonic splicing enhancer motifs in the model plant Arabidopsis thaliana. *BMC Bioinformatics.* **8**, 159.
43. Zhou, J., Liao, J., Zheng, X. and Shen, H. (2012) Chimeric RNAs as potential biomarkers for tumor diagnosis. *BMB Rep.* **45**, 133-140.
44. Blencowe, B. J., Bowman, J. A., McCracken, S. and Rosonina, E. (1999) SR-related proteins and the processing of messenger RNA precursors. *Biochem. Cell Biol.* **77**, 277-291.
45. Fu, X. D. (1995) The superfamily of arginine/serine-rich splicing factors. *RNA* **1**, 663-680.



46. Manley, J. L. and Krainer, A. R. (2010) A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes. Dev.* **24**, 1073-1074.
47. Barta, A., Kalyna, M. and Reddy, A. S. (2010) Implementing a rational and consistent nomenclature for serine/arginine-rich protein splicing factors (SR proteins) in plants. *Plant Cell* **22**, 2926-2929.
48. Richardson, D. N., Rogers, M. F., Labadorf, A., Ben-Hur, A., Guo, H., Paterson, A. H. and Reddy, A. S. (2011) Comparative analysis of serine/arginine-rich proteins across 27 eukaryotes: insights into sub-family classification and extent of alternative splicing. *PLoS One* **6**, e24542.
49. Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S. and Fu, X. D. (2008) The splicing factor SC35 has an active role in transcriptional elongation. *Nat. Struct. Mol. Biol.* **15**, 819-826.
50. Isshiki, M., Tsumoto, A. and Shimamoto, K. (2006) The serine/arginine-rich protein family in rice plays important roles in constitutive and alternative splicing of pre-mRNA. *Plant Cell* **18**, 146-158.
51. Isshiki, M., Yamamoto, Y., Satoh, H. and Shimamoto, K. (2001) Nonsense-mediated decay of mutant waxy mRNA in rice. *Plant Physiol.* **125**, 1388-1395.
52. Severing, E. I., van Dijk, A. D., Stiekema, W. J. and van Ham, R. C. (2009) Comparative analysis indicates that alternative splicing in plants has a limited role in functional expansion of the proteome. *BMC Genomics* **10**, 154.
53. Kong, J., Gong, J. M., Zhang, Z. G., Zhang, J. S. and Chen, S. Y. (2003) A new AOX homologous gene OslM1 from rice (*Oryza sativa* L.) with an alternative splicing mechanism under salt stress. *Theor. Appl. Genet.* **107**, 326-331.
54. Spingola, M., Grate, L., Haussler, D. and Ares, M. Jr. (1999) Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA* **5**, 221-234.
55. Jain, M., Khurana, P., Tyagi, A. K. and Khurana, J. P. (2008) Genome-wide analysis of intronless genes in rice and *Arabidopsis*. *Funct. Integr. Genomics* **8**, 69-78.
56. Zou, M., Guo, B. and He, S. (2011) The roles and evolutionary patterns of intronless genes in deuterostomes. *Comp. Funct. Genomics* **2011**, 680673.
57. Parenteau, J., Durand, M., Veronneau, S., Lacombe, A. A., Morin, G., Guerin, V., Cecez, B., Gervais-Bird, J., Koh, C. S., Brunelle, D., Wellinger, R. J., Chabot, B. and Abou Elela, S. (2008) Deletion of many yeast introns reveals a minority of genes that require splicing for function. *Mol. Biol. Cell.* **19**, 1932-1941.