CellPress
REVIEWS

Review

Introns: Good Day Junk Is Bad Day Treasure

Julie Parenteau[1] and Sherif Abou Elela[1,*]

Introns are ubiquitous in eukaryotic transcripts. They are often viewed as junk RNA but the huge energetic burden of transcribing, removing, and degrading them suggests a significant evolutionary advantage. Ostensibly, an intron functions within the host pre-mRNA to regulate its splicing, transport, and degradation. However, recent studies have revealed an entirely new class of *trans*-acting functions where the presence of intronic RNA in the cell impacts the expression of other genes in *trans*. Here, we review possible new mechanisms of intron functions, with a focus on the role of yeast introns in regulating the cell growth response to starvation.

Highlights

Introns are ubiquitous in large genomes and most human genes have an average of eight introns

In small genomes, few genes have introns, mostly in highly expressed genes.

Introns regulate all levels of their host gene expression, including transcription, export, RNA stability, and even translation through the exon junction complex.

Introns can also act beyond their host genes.

In yeast, introns are dispensable when nutrients are abundant but required during starvation.

Introns: a Ubiquitous Eukaryotic Landmark in Search of a Function

Introns (see Glossary) likely arose from the invasion of an early eukaryotic ancestor by hostile genetic elements that had encrypted within them the means for their removal [1–3]. As multiple genes were invaded at the same time, the early cell therefore had no mechanism by which to overcome the intron burden without compromising its own viability [4]. A **population bottleneck** would have led to the cementing of introns in all extant eukaryotes in which introns are now removed after their transcription through the complex process of splicing to generate mature coding mRNAs [4–6]. Since their inception as junk DNA and RNA the genetic space that introns occupy has acquired new functions as regulators of gene expression and sources of protein diversity [7]. Indeed, introns are the playground for the development of new proteins as a function of alternative splicing [8]. Point mutations that affect splicing can inhibit cell growth and development [9]. Indeed, splicing errors cause a large number of human diseases [9–11], and most disease-causing SNPs are in introns [12–14].

Clearly, removing introns correctly from pre-mRNAs is important but cannot explain their ubiquitous preservation in genomic DNA across the course of evolution. Certainly, they are not simply disposable junk [15–17]. In organisms with large genomes, introns are indispensable for the process of alternative splicing, which is essential for regulating gene expression and function. However, introns are also preserved in organisms where alternative splicing and splicing-dependent regulation of gene expression are rare. Why would single cells with compact genomes tolerate the energetic cost of transcribing, splicing, and degrading introns if their only function is associated with the act of their removal? Could introns serve functions other than regulating the expression of host genes? In the past few years, sequencing advances and our ability to delete introns from eukaryotic genomes have started to provide answers to some of these questions [18,19]. Introns are now known to provide a reservoir of small noncoding (nc)RNAs that act in *trans* on other genes [20]. Recent studies have shown that intronic RNAs also function as direct regulators of multiple cell-response genes during nutrient depletion [21,22]. In the next section, we attempt to describe the general features of introns and their various functions. Finally, as an example of the new functions of introns at large, we will focus on yeast introns and their emerging *trans*-acting functions.

Intron Distribution and Preservation in Eukaryotic Genomes

The number and size of introns vary greatly between organisms (Figure 1A). Introns constitute 24% of mammalian genomes, and 95% of all human genes contain an average of eight introns per gene, ranging in size between <60 nucleotides and >1 megabase (Figure 1A,B) [23,24]. By contrast, there are only 295 confirmed introns located in only 4.7% of the genes of the yeast *Saccharomyces cerevisiae* (Figure 1B,C), and about 36 introns in the obligate microsporidian parasite genome of *Encephalitozoon cuniculi* [25–27]. Only a few organisms with compact genomes and reduced functions have lost introns, such as the obligate microsporidian parasite *Encephalitozoon bieneusi* and endosymiont *Hemiselmis andersenii* [28,29]. Intron number also varies between organisms (Figure 1A,B). In most metazoan genomes, most genes have more than one intron, whereas the compact genomes of

[1]Département de microbiologie et d'infectiologie, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Sherbrooke, QC J1E 4K8, Canada

*Correspondence:
sherif.abou.elela@usherbrooke.ca

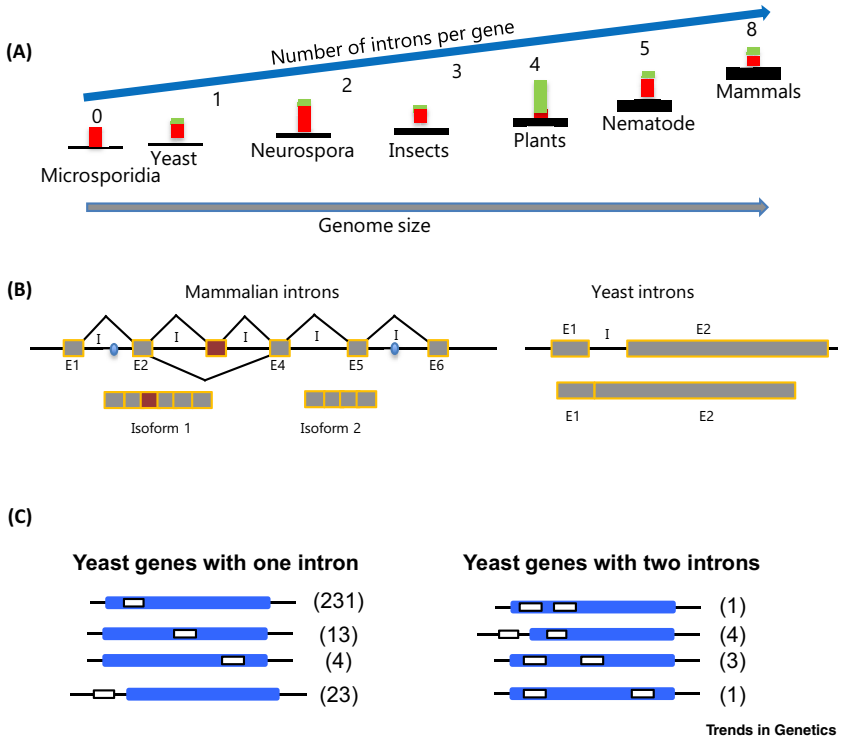

Figure 1. Intron Distribution in Different Organisms.

(A) Introns are abundant in large genomes. The line thickness indicates average intron density (or number of introns per gene) based on a study by Koonin *et al.* [105]. The average number of introns in different organisms taken from Grau-Bové *et al.* is indicated on top [106]. The intron loss and gain relative to the nearest ancestor as estimated by Roy and Gilbert are indicated in red and green bars, respectively [107]. (B) Generalized representation of intron number and position in mammalian and yeast genes. Lines represent introns (I) and boxes represent exons (E). Blue circles indicate noncoding RNAs that are often found in mammalian introns but that are virtually absent from yeast introns. The different isoforms produced by alternative splicing of mammalian genes are indicated below. Most yeast genes (shown on the right) are constitutively spliced and contain just one intron near the 5′ end. (C) Classes of intron-containing genes in budding yeast. Exons, introns, and untranslated regions are shown in blue, white boxes, and lines, respectively. Genes with one intron are shown on the left and those with two introns are on the right. The number of genes with introns in the 5′ untranslated region, 5′ end, middle, and 3′ end of the reading frame is indicated.

Glossary

Intron: noncoding sequence that needs to be removed from the pre-mRNA through the process of splicing to create the mature mRNA coding used for protein synthesis.

Long noncoding RNA (lncRNAs): loose term that is often used to refer to untranslated RNA longer than 200 nucleotides.

Piwi-interacting RNAs (piRNAs): large class of regulatory ncRNAs forming a complex with the piwi-subfamily of Argonaute proteins to regulate gene expression.

Population bottleneck: term used to refer to a sharp reduction in a population, which can lead to reduced genetic variation and changes in, for example, the retention of introns over time.

smaller organisms rarely have more than one intron in the same gene, and multiple introns are virtually absent in *S. cerevisiae* genes [30]. In short, intron number varies among species, and although many eukaryotes have lost most of their introns, they have invariably held on to a few introns in selected group of genes like ribosomal protein genes (Figure 1A,B). The size of introns also varies among species and within the same organism. In general, it is estimated that the average intron in human cells is about 6 kb, with introns containing ncRNAs being larger [31]. By contrast, in compact genomes, introns are generally small, being under 500 nucleotides in fungi and as small as 25 nucleotides in small unicellular parasitic species [32–36].

In most intron-poor genomes, the introns are primarily at the 5′ end of genes [37,38]. One proposed explanation for this bias is that, during evolution, 3′ introns were lost more readily than 5′ introns due to homologous recombination with incompletely reverse-transcribed cDNAs [39]. Indeed, reverse-transcription-dependent homologous recombination, along with genomic deletion is considered to be the main mechanism of intron loss [34,40,41]. However, the retention of 5′ end introns due to links with transcription or translation cannot be ruled out.

Intron Variation and Conservation

Introns evolve more rapidly than coding sequences, which makes the comparison of introns among species difficult [4]. In fact, most studies addressing the issue of intron conservation refer to the maintenance of introns within the gene or at a specific position, and do not delve as far as sequence identity [42]. Comparison of intron position in ancient lineages, the inferred universal common ancestor, and modern organisms, clearly shows that intron position is not conserved [43]. However, there is some intron conservation between certain animal, plant, and fungal genomes. Comparison of eight different genomes identified intron conservation across an estimated 1.5 billion years of evolution [35,44]. However, it remains unclear if this conservation reflects (i) conservation of intron role in the regulation of the host genes; (ii) the presence of required noncoding elements within the intron; or (iii) shared limitations of the intron loss process (e.g., population bottlenecks that prevent intron removal from certain genes). Conservation of intron role in the regulation of the host genes is suggested by the specific conservation of introns near tissue-regulated exons in vertebrates [45]. Introns harboring ncRNA are highly conserved and more examples are being uncovered [19,46]. For example, introns containing small nucleolar (sno)RNA are generally more conserved than those with no apparent ncRNA element [47]. Shared limitations of the intron loss process may include small population size that do not permit weak mutation pressure against newborn introns to become significant or problems associated with reverse transcription which is less efficient when introns are found in the 5′ end of genes [4,40].

Introns as Stable ncRNAs or Precursors of ncRNAs

A large number of ncRNA genes reside in introns, including snoRNAs, **piwi-interacting RNAs** (piRNAs), and **long noncoding RNAs** (lncRNAs) [46,48,49]. In many cases, these intron-derived RNAs are produced as part of the host gene primary transcript, but some piRNAs and lncRNAs are produced from their own intron-embedded promoters [20]. In addition, introns in their entirety, or intron-derived fragments, accumulate in stable linear or circular form [50]. Sequencing of mouse introns indicated that thousands of stable spliced intronic (sis)RNAs, whose overall abundance changes in response to inflammation, have structural and sequence features of functional RNA [46]. In human and other metazoan cells, thousands of spliced-out intronic RNAs were detected and their expression varied in different tissues and cell lines [50,51]. In certain cases, intronic lncRNAs were linked with different aspects of cancer biology [52]. As many intronic ncRNAs are being discovered, we are still awaiting a final inventory. It is even conceivable that all introns produce stable ncRNA in one form or another in at least one specific tissue or conditions. With blurring of functions of true introns and intron-borne *trans*-acting genes, it is difficult to study the innate functions of introns. Intronic mutations or deletion phenotypes could be caused by defects in hidden *trans*-acting ncRNAs or the loss of the intronic sequence itself. Accordingly, a clear definition of intronic ncRNA is needed for any systematic studies of intronic sequence.

Burden of Maintaining Introns

Introns are a major burden on eukaryotic cells given the energy needed to transcribe, excise and degrade them in every mRNA [7,53]. The presence of only one 5-kb intron doubles the average energetic cost of transcribing an average eukaryotic gene [54]. In many cases, especially in human genomes, introns greatly exceed the length of their adjacent exons and the cell needs to transcribe millions of nucleotides to generate a protein encoded by only 2000 nucleotides [53,55]. Aside from the extra investment in nucleotides for transcription, the removal of introns is also very costly and complex, requiring the coordinated expression and assembly of hundreds of genes. Indeed, the spliceosome, or the machine enacting splicing, is among the largest ribonucleoprotein complexes in the cell, comprising five small nuclear (sn)RNAs and >150 proteins [56,57]. This number is even higher in eukaryotes, which also have a second class of spliceosomal introns, called U12 introns, that are removed by the minor spliceosome [58]. In addition, targeting the spliceosome to the exon junction requires a host of *cis*- and *trans*-regulatory factors and elements. The dependence of splicing on the recognition of the splice junction by the spliceosome and requirement of *cis* regulatory elements make the cell vulnerable to synonymous and noncoding mutations that would otherwise be ineffective [59]. Mutations in the spliceosome components also increase the potential of splicing malfunction, which would otherwise be irrelevant in the absence of introns [60]. In complex multicellular organisms with large genome and stable cellular environment the burden of intron could be tolerated for longer terms

advantage. However, it is less clear how short-lived unicellular organisms with rapid replication cycle tolerate the high cost of intron preservation.

Introns Are Enriched in Yeast Ribosomal Protein Genes

Over a third of all yeast introns (104/295) are found in ribosomal protein genes and most of these are found in duplicated ribosomal protein genes, which are not strictly essential for growth [30,61]. Human ribosomal protein genes have shorter than average introns and, as in yeast, the AUG start codon is often close to the first intron boundary [62,63]. Indeed, while size of ribosomal protein gene pre-mRNAs and the number of exons therein varies between eukaryotes, the position of the first 5′ intron is highly conserved [64]. Recent studies in yeast indicate that introns in ribosomal protein genes may have been preserved throughout evolution because they regulate responses to changes in growth conditions [65,66]. However, whether ribosomal protein gene introns achieve similar function between closely related yeast species and other eukaryotes including the human species remains unclear. Plocik and Guthrie proposed that *trans*-regulation of gene expression is universally conserved in some ribosomal protein gene introns. For example, it was shown that an intron in the gene coding for the small subunit protein S9 regulates RNA degradation in yeast, flies, and humans [67]. However, while most examined genomes contain at least one ribosomal protein gene with introns, not all ribosomal protein genes contain introns [68]. The rest of yeast introns are located in meiosis and RNA metabolism genes. Most of these intron encoding genes are heavily regulated at the RNA level, consistent with the presumed role of introns as regulator of gene expression. [21,65].

Yeast Helps Us Understand Intron Function

The budding yeast *S. cerevisiae* is an excellent model for understanding the global impact of introns on transcriptome architecture and cellular homeostasis, as it has only 295 introns located in 280 genes (out of >6275) and only nine genes have more than one intron [26,30,69] (J. Parenteau *et al.*, unpublished). *S. cerevisiae* introns are well annotated, and their boundaries and splicing efficiency have been described both *in silico* and *in vivo*. Furthermore, their conservation across yeast species is also well studied, which facilitates their study and functional assessment [26,30,69]. Intron-containing yeast genes are mostly nonessential for growth, and they have no functions in common, nor are they restricted to a given pathway aside from ribosomal protein genes (J. Parenteau *et al.*, unpublished).

Most yeast introns are shorter than 150 nucleotides and they lack obvious short RNA transcripts or other functional features. Unlike, metazoan introns, which are replete with snoRNA and miRNA, only eight of 295 yeast introns contain a snoRNA. Another, 19 yeast introns contain hypothetical *in silico* predicted ncRNAs, and 20 overlap with an open reading frame on the opposing strand (Figure 1B,C) [21,70]. Therefore, in most cases, a clean deletion of the intron from the yeast genome is unlikely to kill the cell by disrupting the expression of essential genes or sequences [26]. Most importantly, it is much easier to create clean intron deletions in *S. cerevisiae* than in most other organisms. Accordingly, intron deletions were created through a modified two-step (pop-in/pop-out) method that uses homologous recombination and leaves no trace sequence or markers behind [26]. More recently, yeast introns were efficiently deleted using CRISPR technology. However, this method requires a particular strain background and could cause extra nontarget passenger mutations [22]. The creation of an intronless eukaryote will be a powerful tool for studying and understanding how the eukaryotic gene circuitry evolved and how splicing or introns became integrated at different levels of gene expression. Other organisms with fewer introns than *S. cerevisiae*, like certain obligate algae and obligate parasites, are also attractive options for the study of basic intron function and the creation of an intronless eukaryote. However, the lack of comprehensive molecular and phenotypic characterization of these organisms makes them less attractive than other more established model systems with more introns, like budding yeast [6,27,71].

Impact of Introns on Yeast Gene Expression and Growth

The systematic deletion of yeast introns indicates that half of the introns repress the expression of the host gene, whereas a quarter of them induce the expression of the host mRNA [21,65,66] (J. Parenteau

et al., unpublished). Thus, while introns may enhance the expression of certain genes, they mostly function as repressors of tightly regulated genes, including those encoding proteins implicated in RNA metabolism and meiosis [21,26,65] (J. Parenteau *et al.*, unpublished). This is consistent with introns being a burden that repress gene expression. Consistently, deletion of most introns does not affect or enhance cell proliferation in the exponential phase of growth [26]. Analyzing the growth of intron deleted strains in different growth conditions indicated that 31% of introns enhance growth under stress, while 33% repress growth, and 36% have no effect on growth (J. Parenteau *et al.*, unpublished). Strikingly, in fitness tests, 96% of the intron-deleted strains outcompeted the wild-type strain, suggesting that in rich media introns are a burden (J. Parenteau *et al.*, unpublished). Only five introns are essential for growth, and four of these are in *MTR2,* which is one of the few alternatively spliced pre-mRNAs in yeast. The introns of this gene cooperate with its promoter to produce the desired amount of proteins [21,26]. The fifth essential intron is in *ERV46,* and its deletion disrupts the expression of an essential gene that resides in the opposite strand of *ERV46*. These unintended effects of intron deletion highlight the care needed for interpreting the phenotype of intron deletion strains since certain deletion may affect neighboring or overlapping genes. Therefore, at least in yeast, the majority of introns compromise both gene expression and cell proliferation.

Effect of Introns on Gene Expression in Mammalian Cells

In mammalian cells, genes are regulated by intronic splicing regulatory sequences and intron-binding splicing factors. In addition, it was shown that 5′-proximal introns may regulate transcription by harboring transcription enhancers or repressors [72,73]. Some introns may also contain 3′-end processing or termination signals, as is the case of the second intron of the human β-globin gene [74]. Introns can also up- and downregulate gene expression through the link between non-sense-mediated decay (NMD) and alternative splicing (Figure 2) [75]. NMD is a surveillance mechanism that degrades mRNAs with premature termination codons. Mammalian introns often contain NMD signals that are revealed if their splicing is inhibited, when intron retention leads to stop codons that are recognized as premature because of one or several downstream exon junction complexes [76,77]. Introns affect all the processes downstream of splicing including RNA export, as splicing enhances export to the cytoplasm [55]. However, the paradigm of intronic effects on gene expression is related to aberrant splicing. Aberrant splicing may repress gene expression by directing mRNA to degradation through NMD or by changing translation efficiency by repressing protein function through the introduction of missense mutations [75,78]. However, the most prominent function of introns is in alternative splicing where the shuffling of exons, or portions thereof, generates many proteins with distinct functions [78–83].

Reasons for Intron Preservation

Why do most eukaryotic genomes contain introns? There are two main schools of thought concerning the general preservation of introns across evolution. The first posits that the intronic burden remains due to severe population bottlenecks, while the second says that introns are retained due to a positive evolutionary pressure linked to genome complexity [7,34,41,42]. In neither scheme do introns have any innate host independent function, and they are preserved either because the cell cannot remove them or because they regulate their host genes. In human and mouse cells, it is also suggested that *de novo* generation of introns from repeated sequences that become alternatively spliced may serve to increase protein diversity [84]. While these hypotheses could be valid for certain organisms or certain conditions none of them explain why the most preserved introns are found in nonessential genes and in particular duplicated ribosomal protein genes as the case in yeast. If the only criterion for intron attrition is population size and recombination capacity, as the population bottleneck theory posits, we should not observe any enrichment based on host gene function as is the case of ribosomal protein genes. If however, the key is the contribution of introns to host gene expression we should find more introns in essential genes than in nonessential genes, or at least we should observe conservation of introns in the same gene across different species. The evidence from intron-poor genomes, like yeast, suggest yet another model where the burden of introns as repressor of gene expression and hence decelerator of exponential cell growth may provide an advantage for cell resistance to starvation [30,65]. As such, this model, which is discussed in detail below, provides an explanation

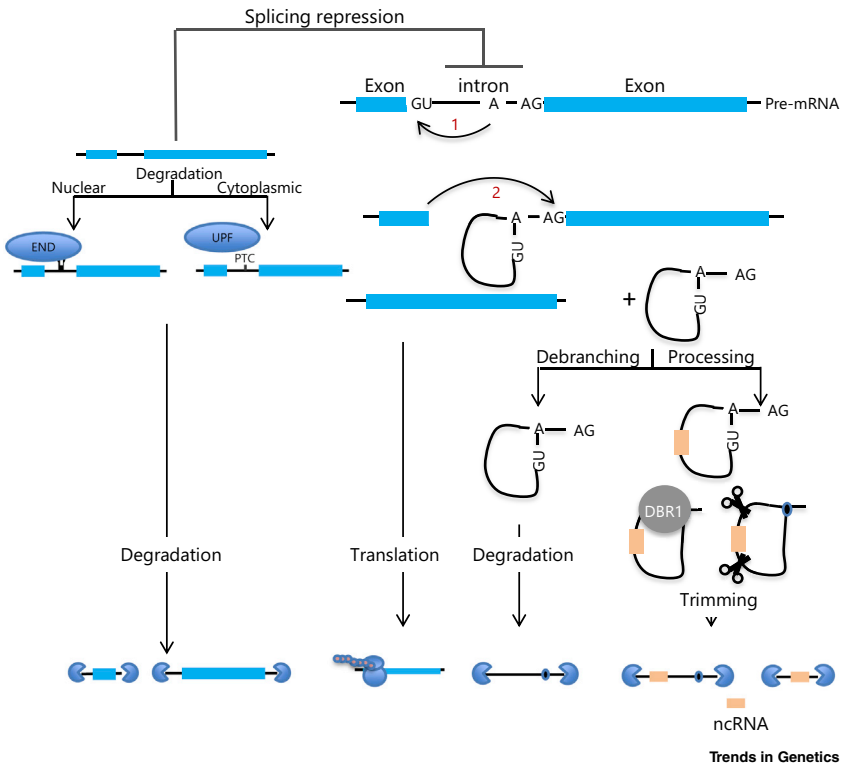

Figure 2. Different Fates of Introns after Splicing.

The different steps of splicing and intron release are shown on the right. Introns containing noncoding (nc)RNAs are processed, either after debranching, as is the case for the most part in mammals, or through endonuclease cleavage (shown as scissors) of the lariat. Introns lacking ncRNA are often degraded after debranching by the debranching enzyme (DBR1). The degradation of unspliced RNA containing premature stop codons (PTC) by non-sense-mediated decay, by the UPF complex, which is prevalent in mammalian cells, is shown on the left. END indicates the degradation of unspliced RNA by endoribonucleases targeting intronic sequence, which occurs in yeast RNA.

for how the burden of introns stipulated by the bottleneck model may translate into a long-term advantage for yeast and perhaps other unicellular organisms, as suggested by the positive selection model.

Alternative Model for Intron Preservation in Yeast

The recent demonstration of an intron effect on cell growth and starvation, by acting as ncRNAs beyond their host gene, puts a new twist on the reason for intron preservation. Indeed, as diffusible RNAs, excised introns enhance the repression of ribosomal protein genes in *trans* during starvation through the sequestration of the spliceosome, which explains why introns might be preserved in all genomes in at least some genes (Figure 3, Key Figure) [21,22]. Only a few genes need to be spliced for this starvation-dependent repression of ribosome biogenesis mechanism to work and it is not important from which genes the introns come. What is needed in this case is at least one intron that accumulates in sufficient quantities during starvation to sequester the spliceosome and repress the splicing of a few ribosomal protein genes leading to repression of translation (Figure 3) [21,22]. Additionally, functions of introns in regulating the expression of duplicated ribosomal protein genes may also add to the reasons for their conservation in this class of genes [65]. Based on these observations, it is possible that, while a population bottleneck or recombination potential have slowed intron loss, it is the capacity of introns to provide global regulation of ribosome biogenesis that directed their preservation in ribosomal protein genes. A possible example of this comes from the red alga

Key Figure

Autonomous Functions of Introns as *trans* Regulators of Splicing and Starvation Response in Yeast

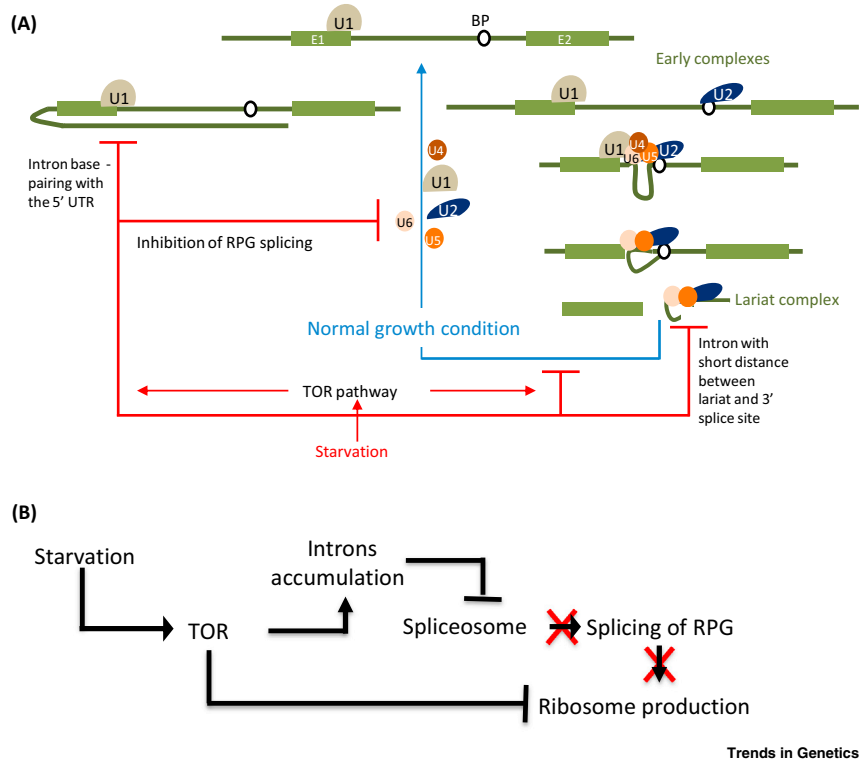

Figure 3. (A) Different mechanisms by which introns sequester the spliceosome upon nutrient depletion. The different steps of splicing are shown and the recycling of splicing factors under normal growth conditions is shown in blue while the sequestration of splicing factors upon nutrient depletion (starvation) is indicated by the red lines. Introns may regulate splicing under starvation as part of the pre-mRNA (shown on the left) or as free spliced introns (shown on the right). In this model starvation would lead to the accumulation of introns either as part of the pre-mRNA or as spliced linear introns. The accumulation of introns leads to repression of ribosomal protein genes (RPG) splicing by sequestering early (in case of unspliced introns) or late (in the case of spliced introns) components of the spliceosome. (B) Intron-mediated repression of ribosomal protein genes as part of the TOR-mediated response to starvation. Upon starvation introns accumulate in the form of pre-mRNA or as spliced introns and associate with the spliceosome. The association of the introns with the spliceosome represses ribosome biogenesis through the repression of ribosomal protein gene splicing. Abbreviations: BP, branch point; E1 and E2, two exons; U1, 2, 4–6, small nuclear RNAs.

Cyanidioschyzon merolae. This acidothermophile was found to have only 27 introns in its genome, of which only four are in genes clearly related to translation: two in ribosomal proteins, one in eIF1a, and one in a predicted valine tRNA ligase [85]. Recent analysis of the splicing machinery of *C. merolae* revealed a highly reduced spliceosome with few of the auxiliary factors that regulate splicing in other organisms [6]. If the introns themselves are involved in starvation tolerance, through their impact on the translation machinery, it must be only a small number of introns that are responsible. Given the highly interdependent nature of translation, perhaps preservation of just one intron in a single ribosomal protein gene regulates the entire ribosome synthesis pathway. Indeed, most introns in yeast

required for resistance to starvation are not found in ribosomal protein genes. Instead, they indirectly affect ribosome biogenesis by interfering with splicing and expression of other genes involved in this process [21]. Intron deletion in *C. morolae* or other organisms with ultra-compact genomes will provide a strong test of the idea that a key role of introns is to regulate translation.

ncRNA *trans* Functions of Introns

The effect of introns on the expression and function of their host genes is well documented, but their innate effects on other genes are just starting to be uncovered. One of the most accepted non-host functions of introns is as precursors of ncRNAs, like snoRNAs, miRNAs, and lncRNAs [46]. ncRNAs are generated from introns, often after the splicing and debranching of the lariats. However, ncRNAs can also be excised directly from the unspliced introns by specialized endoribonucleases (Figure 2) [86–90]. It is likely that these intronic ncRNAs were added to introns late in evolution to take advantage of their long size and weak selection constraints [91,92]. Stable intronic sequences devoid of obvious ncRNA were also documented, especially those produced from viral RNA, but the functional significance of these introns remains unclear [50].

Profiling of splicing in yeast under different growth conditions suggested that splicing is repressed upon starvation [93]. More recently, it was demonstrated that, in *S. cerevisiae,* the RNA of certain introns accumulates when nutrients are depleted. Deletion of the introns reduces resistance to starvation, as they function to sensitize the cells to the TOR nutrient sensing pathway that represses ribosome synthesis [21,22]. Surprisingly, 90% of budding yeast introns affect growth under starvation. This general effect of introns in *S. cerevisiae* does not act through the host gene or any ncRNA sequence within it. In most cases, introns promote cell survival under starvation as part of the unspliced pre-mRNA but a small subset of introns with short distances between the lariat branch-point and 3′ splice site were also shown to promote cell survival as linear post-splicing introns (Figure 3) [21,22]. It is currently unclear if the unspliced and spliced introns promote cell resistance to starvation using the same mechanism; however, it is clear that at least the TOR pathway works via both spliced and unspliced introns [21,22]. To summarize, the discovery of introns that function beyond their host genes refutes the previous view of introns as entities that only function to modify their host gene's expression or a reservoir of ncRNA genes.

Impact of Splicing Equilibrium on Nutrient Sensing

Eukaryotic cells have many introns, and even in cells with intron-poor genomes, like *S. cerevisiae,* there are hundreds of introns being spliced at any one time, raising questions of how introns are selected for splicing by the spliceosome, and whether splicing is rate-limiting for gene expression. An obvious mechanism of splicing regulation involves targeted modulation of splicing rates by RNA-binding proteins. Classical studies focused on auxiliary splicing factors and their binding sites as the main criteria for selecting messages and which exons are spliced and when [94]. Most of these studies characterized the regulatory networks by depleting individual factors and studying the impact on splicing. This revealed that individual factors caused certain introns to be more spliced than others to affect the pattern of alternative splicing [94–96]. Other studies identified motifs or binding sites near splice sites to explain the splicing pattern. However, most of these studies did not consider the impact of the availability, or concentration, of introns on splicing equilibrium. According to this theory the rate constant determines which introns are spliced and at what levels. This theory gives eminence to the availability of spliceosomes and the 'spliceability' of introns, which depends on splice site consensus sequences [73,97–99].

Six years ago, a pioneering study in yeast, demonstrated that splicing can be regulated by competition between different RNAs for access to the splicing apparatus. Thus, depletion of highly spliced mRNAs induced the splicing of poorly spliced mRNAs [100]. However, the sensitivity of this equilibrium and the functional consequences of splicing competition remained unclear. Recently this question was answered through the demonstration that deletion of a single intron alters the splicing of as much as 15% of the intron-containing pre-mRNA in *S. cerevisiae*. Strikingly, deletion of different introns increased the splicing of a similar subset of highly spliced genes and especially ribosomal

protein genes [21]. This indicates that the splicing equilibrium is very sensitive to the differential affinities of RNAs for the spliceosome.

Under the equilibrium model, the deletion of any 'sticky' introns that are not readily released from the spliceosome would induce the splicing of the same set of efficiently spliced introns. Linear debranched introns that accumulate during the stationary phase form stable complexes with spliceosomal proteins, suggesting these stable introns actively sequester the spliceosome (Figure 3) [22]. We demonstrated that nonexcised introns affect splicing by folding into secondary structures with the 5′ untranslated region. We propose that these pre-mRNA structures, which are required for growth under starvation, also sequester the spliceosome (Figure 3) [21]. It remains to be seen if the spliceosome machinery can count the number of genes being spliced or if there is an overall intron concentration threshold below which splicing is diminished. In general, deleting more than one intron in one cell did not progressively decrease cell viability or resistance to starvation, suggesting that this effect of introns is not additive [21]. By contrast, multiple deletions of the introns that accumulate in linear forms increased resistance to drugs that repress ribosome synthesis, like rapamycin, suggesting that this effect of linear introns might be cumulative [22]. However, it is not clear if progressive intron deletion would lead to a proportional concomitant shift in splicing equilibrium. A systematic examination of the effect of deleting increasing numbers of introns on splicing is needed to settle this question. In any case, it is now imperative to consider the quantity and type of mRNA accumulating in the cell when studying the impact of mutations and factors on splicing.

Outstanding Questions

How does the deletion of single introns among hundreds changes the splicing equilibrium?

Are there one or more mechanisms by which linear spliced and unspliced introns affect splicing equilibrium and/or cell resistance to starvation?

Is intron function restricted to regulating nutrient-sensing and resistance to starvation?

The Balancing Act between Selfish and Selfless Introns

The realization that introns can function beyond their host genes creates a new area of research about how cells control the interplay between the *cis*- and *trans*-acting functions of introns. For introns to function in *trans*, they need to accumulate in the cell, either within the unspliced pre-mRNA or as excised introns (Figure 3) [21,22]. When the introns accumulate as unspliced pre-mRNA, this represses host gene expression. If the *trans*-acting introns accumulate as free spliced-out introns, the host gene must use an additional means of regulation downstream of the splicing step, like targeted degradation of the spliced mRNA. Recent studies indicate that both mechanisms of intron function exist (Figure 3) [21,22]. Examples of introns that function within the context of premature or unspliced pre-mRNA can be divided into three classes: (i) genes that are repressed constitutively through repression of splicing (e.g., meiotic genes that are repressed in the vegetative state); (ii) genes that are conditionally repressed through splicing in response to starvation (e.g., metabolic genes); and (iii) autoregulatory feedback loops between duplicated genes with a certain degree of functional redundancy (e.g., ribosomal protein genes) [21]. In the first two cases the inhibition of splicing has no side effects since the genes are repressed in the condition requiring intron accumulation. In the third group, duplication of function (e.g., duplicated ribosomal protein genes) may permit the repression of one paralog with no serious impact on the cell.

Introns, or parts of introns, that have an independent life after splicing, need to be regulated in ways that do not strictly depend on the expression of the host gene especially if the host gene protein and introns have different or perhaps opposing function. For many genes in yeast, this is achieved by specific RNA processing or degradation signals in the intron (Figure 2) [90]. The presence of a stem-loop structure in introns permits cleavage of either the pre-mRNA or the spliced RNA by the double-stranded-RNA-specific ribonuclease III [90,101]. This releases the intron-embedded snoRNAs even when splicing is inhibited [90]. A recent study demonstrated that the accumulation of spliced introns increases as the distance between the 3′ splice site and the branch point is shortened [22]. These introns with branch points close to the 3′ splice site accumulated in saturated culture. This suggests that intron accumulate is a response to growth conditions. However, the mechanism by which the cell regulates the abundance of these linear introns without perturbing the host gene products' abundance or function remains unclear. What is clear however is that, in all cases, the genes carrying functional introns must be transcribed, at least at low level, regardless of the cell need for the mature mRNA to allow introns to function and be regulated post-transcriptionally.

Concluding Remarks

Modern introns are not junk DNA, but rather they are noncoding sequences that have gained functions through evolution to merit their maintenance. It is possible that a population bottleneck, or inability to eliminate introns, provided the initial reason for their preservation and allowed introns to develop these functions. However, it is becoming increasingly clear that these introns have rapidly gained inherent functions beyond regulating their host gene that further supported their maintenance through evolution. The recent discovery of a function for introns as regulators of splicing efficiency that in turn determines cell survival under starvation indicates that the lack of an effect on the host gene and the absence of identifiable sequence or structural features do not necessarily indicate a lack of function. Most importantly, the effects of intron mutations and deletions can no longer be assumed to solely impact host gene function. It is not clear at this moment if introns contribution to cell resistance to starvation is restricted to yeast or preserved in other organisms. Several lines of evidence suggest that senses, dietary restriction, and aging may influence splicing and may lead to intron retention [102–104]. However, it remains unclear if they may function independent of their host genes in complex genomes like that of human cells. Indeed, we must now benefit from this fresh opportunity to re-examine whether deep intronic mutations (mutations that are a long way from splice sites) can alter the function of splicing regulators and thus splicing in *trans* (see Outstanding Questions).

Acknowledgments

We thank Benoit Chabot and Stephen Rader for reading the manuscript. The work was supported by a research grant from the Canadian Institute of Health Research (CIHR) and Research Chair in RNA Biology and Cancer Genomics to S.A.

Disclaimer Statement

The authors declare no competing interests.

References

1. de Souza, S.J. *et al.* (1996) Introns and gene evolution. *Genes Cells* 1, 493–505
2. Hartung, F. *et al.* (2002) Intron gain and loss in the evolution of the conserved eukaryotic recombination machinery. *Nucleic Acids Res.* 30, 5175–5181
3. Mattick, J.S. (1994) Introns: evolution and function. *Curr. Opin. Genet. Dev.* 4, 823–831
4. Lynch, M. (2002) Intron evolution as a population-genetic process. *Proc. Natl. Acad. Sci. U S A* 99, 6118–6123
5. Irimia, M. and Roy, S.W. (2014) Origin of spliceosomal introns and alternative splicing. *Cold Spring Harb. Perspect. Biol.* 6, a016071
6. Stark, M.R. *et al.* (2015) Dramatically reduced spliceosome in *Cyanidioschyzon merolae*. *Proc. Natl. Acad. Sci. U S A* 112, E1191–E1200
7. Jo, B.S. and Choi, S.S. (2015) Introns: the functional benefits of introns in genomes. *Genomics Inform* 13, 112–118
8. McAlinden, A. *et al.* (2004) Regulation of protein diversity by alternative pre-mRNA splicing with specific focus on chondrogenesis. *Birth Defects Res. C Embryo Today* 72, 51–68
9. Kuyumcu-Martinez, N.M. and Cooper, T.A. (2006) Misregulation of alternative splicing causes pathogenesis in myotonic dystrophy. *Prog. Mol. Subcell Biol.* 44, 133–159
10. Solis, A.S. *et al.* (2008) Splicing fidelity, enhancers, and disease. *Front. Biosci.* 13, 1926–1942
11. Baralle, D. and Baralle, M. (2005) Splicing in action: assessing disease causing sequence changes. *J. Med. Genet.* 42, 737–748
12. Kwan, T. *et al.* (2008) Genome-wide analysis of transcript isoform variation in humans. *Nat. Genet.* 40, 225–231
13. Kwan, T. *et al.* (2007) Heritability of alternative splicing in the human genome. *Genome Res.* 17, 1210–1218
14. Tessier, M.C. *et al.* (2006) Type 1 diabetes and the OAS gene cluster: association with splicing polymorphism or haplotype? *J. Med. Genet.* 43, 129–132
15. Bulman, S. *et al.* (2007) Intron-rich gene structure in the intracellular plant parasite *Plasmodiophora brassicae*. *Protist* 158, 423–433
16. Roy, S.W. and Penny, D. (2006) Large-scale intron conservation and order-of-magnitude variation in intron loss/gain rates in apicomplexan evolution. *Genome Res.* 16, 1270–1275
17. Russell, A.G. *et al.* (2005) An ancient spliceosomal intron in the ribosomal protein L7a gene (Rpl7a) of *Giardia lamblia*. *BMC Evol. Biol.* 5, 45
18. Hesselberth, J.R. (2013) Lives that introns lead after splicing. *Wiley Interdiscip. Rev. RNA* 4, 677–691
19. Chorev, M. and Carmel, L. (2012) The function of introns. *Front. Genet.* 3, 55
20. Boivin, V. *et al.* (2018) Protein coding genes as hosts for noncoding RNA expression. *Semin. Cell Dev. Biol.* 75, 3–12
21. Parenteau, J. *et al.* (2019) Introns are mediators of cell response to starvation. *Nature* 565, 612–617
22. Morgan, J.T. *et al.* (2019) Excised linear introns regulate growth in yeast. *Nature* 565, 606–611
23. Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921

24. Venter, J.C. *et al.* (2001) The sequence of the human genome. *Science* 291, 1304–1351
25. Juneau, K. *et al.* (2007) High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. *Proc. Natl. Acad. Sci. U S A* 104, 1522–1527
26. Parenteau, J. *et al.* (2008) Deletion of many yeast introns reveals a minority of genes that require splicing for function. *Mol. Biol. Cell* 19, 1932–1941
27. Whelan, T.A. *et al.* (2019) Microsporidian introns retained against a background of genome reduction: characterization of an unusual set of introns. *Genome Biol. Evol.* 11, 263–269
28. Lane, C.E. *et al.* (2007) Nucleomorph genome of *Hemiselmis andersenii* reveals complete intron loss and compaction as a driver of protein structure and function. *Proc. Natl. Acad. Sci. U S A* 104, 19908–19913
29. Keeling, P.J. *et al.* (2010) The reduced genome of the parasitic microsporidian *Enterocytozoon bieneusi* lacks genes for core carbon metabolism. *Genome Biol. Evol.* 2, 304–309
30. Spingola, M. *et al.* (1999) Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA* 5, 221–234
31. Hube, F. and Francastel, C. (2015) Mammalian introns: when the junk generates molecular diversity. *Int. J. Mol. Sci.* 16, 4429–4452
32. Angrist, M. (1998) Less is more: compact genomes pay dividends. *Genome Res.* 8, 683–685
33. Hawkins, J.D. (1988) A survey on intron and exon lengths. *Nucleic Acids Res.* 16, 9893–9908
34. Ma, M.Y. *et al.* (2015) Evaluation of the mechanisms of intron loss and gain in the social amoebae *Dictyostelium*. *BMC Evol. Biol.* 15, 286
35. Rogozin, I.B. *et al.* (2003) Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. *Curr. Biol.* 13, 1512–1517
36. Neuveglise, C. *et al.* (2011) The intronome of budding yeasts. *C R Biol.* 334, 662–670
37. Sakurai, A. *et al.* (2002) On biased distribution of introns in various eukaryotes. *Gene* 300, 89–95
38. Mourier, T. and Jeffares, D.C. (2003) Eukaryotic intron loss. *Science* 300, 1393
39. Fink, G.R. (1987) Pseudogenes in yeast? *Cell* 49, 5–6
40. Cohen, N.E. *et al.* (2012) The role of reverse transcriptase in intron gain and loss mechanisms. *Mol. Biol. Evol.* 29, 179–186
41. Ma, M.Y. *et al.* (2015) Imprecise intron losses are less frequent than precise intron losses but are not rare in plants. *Biol. Direct.* 10, 24
42. Lynch, M. (2006) The origins of eukaryotic gene structure. *Mol. Biol. Evol.* 23, 450–468
43. Cho, G. and Doolittle, R.F. (1997) Intron distribution in ancient paralogs supports random insertion and not random loss. *J. Mol. Evol.* 44, 573–584
44. Kondrashov, F.A. and Koonin, E.V. (2003) Evolution of alternative splicing: deletions, insertions and origin of functional parts of proteins from intron sequences. *Trends Genet.* 19, 115–119
45. Sugnet, C.W. *et al.* (2006) Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLoS Comput. Biol.* 2, e4
46. St Laurent, G. *et al.* (2012) Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells. *BMC Genomics* 13, 504
47. Chorev, M. *et al.* (2017) Identification of introns harboring functional sequence elements through positional conservation. *Sci. Rep.* 7, 4201
48. Singh, G. *et al.* (2018) Genome-wide profiling of the PIWI-interacting RNA-mRNA regulatory networks in epithelial ovarian cancers. *PLoS One* 13, e0190485
49. Boivin, V. *et al.* (2019) The cellular landscape of mid-size noncoding RNA. *Wiley Interdiscip. Rev. RNA* 10, e1530
50. Chan, S.N. and Pek, J.W. (2019) Stable intronic sequence RNAs (sisRNAs): an expanding universe. *Trends Biochem. Sci.* 44, 258–272
51. Osman, I. *et al.* (2016) Stable intronic sequence RNAs (sisRNAs): a new layer of gene regulation. *Cell Mol. Life Sci.* 73, 3507–3519
52. Wang, M. *et al.* (2017) Prognostic role of the long non-coding RNA, SPRY4 Intronic Transcript 1, in patients with cancer: a meta-analysis. *Oncotarget* 8, 33713–33724
53. Gorlova, O. *et al.* (2014) Genes with a large intronic burden show greater evolutionary conservation on the protein level. *BMC Evol. Biol.* 14, 50
54. Lynch, M. and Marinov, G.K. (2015) The bioenergetic costs of a gene. *Proc. Natl. Acad. Sci. U S A* 112, 15690–15695
55. Shaul, O. (2017) How introns enhance gene expression. *Int. J. Biochem. Cell Biol.* 91, 145–155
56. Zhang, X. *et al.* (2017) An atomic structure of the human spliceosome. *Cell* 169, 918–929 e14
57. Yan, C. *et al.* (2015) Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* 349, 1182–1191
58. Turunen, J.J. *et al.* (2013) The significant other: splicing by the minor spliceosome. *Wiley Interdiscip. Rev. RNA* 4, 61–76
59. Lopez-Bigas, N. *et al.* (2005) Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett.* 579, 1900–1903
60. Wang, G.S. and Cooper, T.A. (2007) Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat. Rev. Genet.* 8, 749–761
61. Ares, M., Jr. *et al.* (1999) A handful of intron-containing genes produces the lion's share of yeast mRNA. *RNA* 5, 1138–1139
62. Ishii, K. *et al.* (2006) Characteristics and clustering of human ribosomal protein genes. *BMC Genomics* 7, 37
63. Zhang, Q. *et al.* (2016) The evolution mechanism of intron length. *Genomics* 108, 47–55
64. Yoshihama, M. *et al.* (2002) The human ribosomal protein genes: sequencing and comparative analysis of 73 genes. *Genome Res.* 12, 379–390
65. Parenteau, J. *et al.* (2011) Introns within ribosomal protein genes regulate the production and function of yeast ribosomes. *Cell* 147, 320–331
66. Petibon, C. *et al.* (2016) Introns regulate the production of ribosomal proteins by modulating splicing of duplicated ribosomal protein genes. *Nucleic Acids Res.* 44, 3878–3891
67. Plocik, A.M. and Guthrie, C. (2012) Diverse forms of RPS9 splicing are part of an evolving autoregulatory circuit. *PLoS Genet.* 8, e1002620
68. Yoshihama, M. *et al.* (2007) Intron dynamics in ribosomal protein genes. *PLoS One* 2, e141
69. Grate, L. and Ares, M., Jr. (2002) Searching yeast intron data at Ares lab Web site. *Methods Enzymol.* 350, 380–392
70. Hooks, K.B. *et al.* (2016) Novel intronic RNA structures contribute to maintenance of phenotype in *Saccharomyces cerevisiae*. *Genetics* 203, 1469–1481
71. Hudson, A.J. *et al.* (2015) Splicing diversity revealed by reduced spliceosomes in *C. merolae* and other organisms. *RNA Biol.* 12, 1–8
72. Gaunitz, F. *et al.* (2005) An intronic silencer element is responsible for specific zonal expression of glutamine synthetase in the rat liver. *Hepatology* 41, 1225–1232
73. Sun, S. *et al.* (2012) Mechanisms of activation and repression by the alternative splicing factors RBFOX1/2. *RNA* 18, 274–283

74. Antoniou, M. *et al.* (1998) Efficient 3′-end formation of human beta-globin mRNA in vivo requires sequences within the last intron but occurs independently of the splicing reaction. *Nucleic Acids Res.* 26, 721–729
75. Ge, Y. and Porse, B.T. (2014) The functional consequences of intron retention: alternative splicing coupled to NMD as a regulator of gene expression. *Bioessays* 36, 236–243
76. Gerbracht, J.V. and Gehring, N.H. (2018) The exon junction complex: structural insights into a faithful companion of mammalian mRNPs. *Biochem. Soc. Trans.* 46, 153–161
77. Le Hir, H. *et al.* (2016) The exon junction complex as a node of post-transcriptional networks. *Nat. Rev. Mol. Cell Biol.* 17, 41–54
78. Lareau, L.F. *et al.* (2007) The coupling of alternative splicing and nonsense-mediated mRNA decay. *Adv. Exp. Med. Biol.* 623, 190–211
79. Black, D.L. (2000) Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. *Cell* 103, 367–370
80. Blencowe, B.J. (2006) Alternative splicing: new insights from global analyses. *Cell* 126, 37–47
81. Cooper, T.A. (2005) Alternative splicing regulation impacts heart development. *Cell* 120, 1–2
82. Kim, E. *et al.* (2008) Alternative splicing: current perspectives. *Bioessays* 30, 38–47
83. Lopez, A.J. (1998) Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu. Rev. Genet.* 32, 279–305
84. Zhuo, D. *et al.* (2007) Modern origin of numerous alternatively spliced human introns from tandem arrays. *Proc. Natl. Acad. Sci. U S A* 104, 882–886
85. Matsuzaki, M. *et al.* (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* 428, 653–657
86. Brown, J.W. *et al.* (2008) Intronic noncoding RNAs and splicing. *Trends Plant Sci.* 13, 335–342
87. Ooi, S.L. *et al.* (1998) Intronic snoRNA biosynthesis in *Saccharomyces cerevisiae* depends on the lariat-debranching enzyme: intron length effects and activity of a precursor snoRNA. *RNA* 4, 1096–1110
88. Hirose, T. and Steitz, J.A. (2001) Position within the host intron is critical for efficient processing of box C/D snoRNAs in mammalian cells. *Proc. Natl. Acad. Sci. U S A* 98, 12914–12919
89. Petfalski, E. *et al.* (1998) Processing of the precursors to small nucleolar RNAs and rRNAs requires common components. *Mol. Cell Biol.* 18, 1181–1189
90. Ghazal, G. *et al.* (2005) Genome-wide prediction and analysis of yeast RNase III-dependent snoRNA processing signals. *Mol. Cell Biol.* 25, 2981–2994
91. Hoeppner, M.P. *et al.* (2009) Evolutionarily stable association of intronic snoRNAs and microRNAs with their host genes. *Genome Biol. Evol.* 1, 420–428
92. Hoeppner, M.P. and Poole, A.M. (2012) Comparative genomics of eukaryotic small nucleolar RNAs reveals deep evolutionary ancestry amidst ongoing intragenomic mobility. *BMC Evol. Biol.* 12, 183
93. Bergkessel, M. *et al.* (2011) Diverse environmental stresses elicit distinct responses at the level of pre-mRNA processing in yeast. *RNA* 17, 1461–1478
94. Lee, Y. and Rio, D.C. (2015) Mechanisms and regulation of alternative pre-mRNA splicing. *Annu. Rev. Biochem.* 84, 291–323
95. Simard, M.J. and Chabot, B. (2000) Control of hnRNP A1 alternative splicing: an intron element represses use of the common 3′ splice site. *Mol. Cell Biol.* 20, 7353–7362
96. Li, Q. *et al.* (2007) Neuronal regulation of alternative pre-mRNA splicing. *Nat. Rev. Neurosci.* 8, 819–831
97. Holste, D. and Ohler, U. (2008) Strategies for identifying RNA splicing regulatory motifs and predicting alternative splicing events. *PLoS Comput. Biol.* 4, e21
98. Rosenberg, A.B. *et al.* (2015) Learning the sequence determinants of alternative splicing from millions of random sequences. *Cell* 163, 698–711
99. Suyama, M. *et al.* (2010) A network of conserved co-occurring motifs for the regulation of alternative splicing. *Nucleic Acids Res.* 38, 7916–7926
100. Munding, E.M. *et al.* (2013) Competition between pre-mRNAs for the splicing machinery drives global regulation of splicing. *Mol. Cell* 51, 338–348
101. Danin-Kreiselman, M. *et al.* (2003) RNAse III-mediated degradation of unspliced pre-mRNAs and lariat introns. *Mol. Cell* 11, 1279–1289
102. Naro, C. *et al.* (2017) An orchestrated intron retention program in meiosis controls timely usage of transcripts during germ cell differentiation. *Dev. Cell* 41, 82–93 e4
103. Naro, C. and Sette, C. (2017) Timely-regulated intron retention as device to fine-tune protein expression. *Cell Cycle* 16, 1321–1322
104. Heintz, C. *et al.* (2017) Splicing factor 1 modulates dietary restriction and TORC1 pathway longevity in *C. elegans*. *Nature* 541, 102–106
105. Koonin, E.V. (2006) The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biol. Direct* 1, 22
106. Grau-Bové, X. *et al.* (2017) Dynamics of genomic innovation in the unicellular ancestry of animals. *eLife* 6, e26036
107. Roy, S.W. and Gilbert, W. (2005) Rates of intron loss and gain: implications for early eukaryotic evolution. *Proc. Natl. Acad. Sci. U S A* 102, 5773–5778