Computational Biology and Chemistry 86 (2020) 107251

Contents lists available at ScienceDirect

Computational Biology and Chemistry

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

Research Article

5' and 3' splicing signals evolution in vertebrates: Analysis in a conserved gene family

Maria A. Panaro, Rosa Calvello*, Vincenzo Mitolo, Antonia Cianciulli

Department of Biosciences, Biotechnologies and Biopharmaceutics, University of Bari, via Orabona, 4, I-70126 Bari, Italy

ARTICLE INFO

Keywords:
Splicing signals
Mitochondrial solute carrier genes
Zebrafish
Chicken
Mouse
Human

ABSTRACT

The mitochondrial solute carrier genes (SLC25) are highly conserved during vertebrate evolution. In most SLC25 genes of zebrafish, chicken, mouse, and human, the introns are located at exactly superimposable positions. In these topographically corresponding introns we studied the composition of the initial and terminal hexanucleotides (5'ss and 3'ss) which are instrumental in splicing signaling, focusing on the evolutionary conservation/mutation dynamics of these genetically related sequences. At each position, the per cent conservation of zebrafish individual nucleotides in chicken, mouse and human is proportional to their percent frequency in zebrafish; furthermore, nucleotide mutations are biased in favor of the more represented nucleotides, thus compensating for those highly represented zebrafish nucleotides which have not been conserved. As a result of these evolutionary dynamics, the general nucleotide composition at each position has remained relatively conserved throughout vertebrates. At 5'ss, following the canonical GT, A and G are largely prevailing at position +3, A at +4 and G at +5 (GT[A/G]AGx). At 3'ss, T and C are largely prevailing at positions −6, −5 and −3, preceding the canonical intron terminal AG ([C/T] [C/T]x[C/T]AG). However, the actual composition of the tetranucleotides at 5' and 3' often does not conform to the above scheme. At 5'ss the more canonical sequence is completely expressed in 63% of cases and partially (2 or 1 matches) in 37 % of cases. At 3'ss the more canonical sequence is completely expressed in 71 % of cases and partially (2 or 1 matches) in 29 % of cases. The nucleotide conservation loss (nucleotide mutation) is higher in the evolution from fish to the last common ancestor of birds and mammals (58 %), then diminishes in the successive evolution steps up to the mammalian common ancestor (10 %), and becomes still lower at the divergence of rodents and primates (5 %).

1. Introduction

During the splicing process of the pre-mRNAs the non-coding introns are removed, eventually resulting in mature mRNAs, formed by the coding exons only. The splicing is a complex process involving specific pre-mRNA signaling sequences, a host of proteins forming the spliceosome and some small nuclear ribonucleoproteins (snRNPs). The exact locations of the exon/intron cutting points at each mRNA precursor is responsible for generating homogeneous canonical protein products ("normal" proteins). During evolution, alternative splicing events can generate modified proteins, which might be selected positively and activate novel biochemical pathways. On the contrary, occasional splicing "mistakes" occurring in individuals may cause severe diseases (Burset et al., 2000; Nilsen, 2003; Chasin, 2007; Schwartz et al., 2008; Ke and Chasin, 2011; Arias et al., 2015; Kadowaki, 2015; Wan et al., 2019; Wilkinson et al., 2019; Ule and Blencowe, 2019). Among the hypothesized "intrinsic" splicing signals residing in the pre-mRNA itself, the nucleotides at both intron ends, making up the 5' splicing signal (5'ss) and the 3' splicing signal (3'ss), are believed to play major roles. Further evidence indicates that the critical nucleotides at 5'ss and 3'ss are approximately the first six, and the last six nucleotides, respectively (Abril et al., 2005; Schwartz et al., 2008; Calvello et al., 2013). The first two and last two intron nucleotides are almost invariably GT and AG, respectively, which are mutated by less than 1 % (Burset et al., 2001; Abril et al., 2005), while the other 5'ss nucleotides (+3 to +6) and 3'ss nucleotides (−6 to −3) are widely variable.

The structure of the splicing sites, including the 5' and 3' ss sections, has been studied in different species (Abril et al., 2005; Schwartz et al., 2008; Calvello et al., 2013; Baralle and Baralle, 2018). However, the comparative/evolutionary studies were usually carried on large mixed populations of genes in order to discover the 5'ss and 3'ss specific general traits in different animal families or genera. By contrast, we decided to study the splicing signal evolution in a family of genes which are highly conserved in vertebrates. The model we selected for this

* Corresponding author.
E-mail address: rosa.calvello@uniba.it (R. Calvello).

https://doi.org/10.1016/j.compbiolchem.2020.107251
Received 12 July 2019; Received in revised form 8 March 2020; Accepted 13 March 2020
Available online 18 March 2020
1476-9271/

study is the mitochondrial solute carrier family of genes (SLC25 genes, A1 to A54; Palmieri, 2013; Palmieri and Monnè, 2016). These genes are well conserved during evolution (Palmieri and Pierri, 2010) and in vertebrates most of the SLC25 pre-mRNAs are spliced according to a topographic scheme which is strictly conserved in each carrier from fish to primates. The corresponding 5'ss and 3'ss are, therefore, genetically related and are well suited to study the evolution of these splicing signals. In particular, we studied the 5'ss and 3'ss sections in the SLC25 genes of zebrafish, chicken, mouse and human. Rather than describing the structural differences between species, we focused on the evolutionary conservation/mutation dynamics studied at the level of the individual nucleotides, from fish to the common bird/mammal ancestor and the mammalian common ancestor, down to the extant species.

The SLC25 genes are thought to be expressed in all tissues, though to different extents, possibly according to the metabolic activity. However, some of the genes are expressed at especially high levels in specific tissues: A7 (adipose tissue), A3 (isoform A) and A12 (heart and skeletal muscle), A14 and A27 (brain), A30 (kidney), A31 (testis), and A 38 (erythroid cells) (https://www.uniprot.org/uniprot/).

The conservation of the 5'ss and 3'ss sections was studied separately in the group of tissue-specific genes (see above) and in all other SLC25 genes.

2. Material and methods

The NCBI bank of homologous genes (https://www.ncbi.nlm.nih.gov/homologene) was used to select the homologous mitochondrial solute carrier genes (SLC25 genes) in different species. Genomic and all mRNA *canonical* sequences of the SLC25 genes, (A1 to A54) of human (*Homo sapiens*), mouse (*Mus musculus*), chicken (*Gallus gallus*), and zebrafish (*Danio rerio*) were derived from the NCBI GenBank (http://www.ncbi.nlm.nih.gov/).

Analyses in this paper are based on comparisons of homologous hexanucleiotides at the 5' and 3' ends of introns of SLC25 genes between two or more species. For some carriers, however, reliable DNA sequences were not available for some species (especially the chicken), or the introns were not strictly homologous. For these reasons, the zebrafish/chicken comparisons were based on 183 couples of hexanucleotides for both 5'ss and 3'ss; the zebrafish/mouse on 259 couples; the zebrafish/human on 252 couples; and the comparisons among all species cumulatively on 176 intron sequences. Due to the different raw data used in different comparisons, the percentages of conservation or mutation of specific nucleotides in each species may be slightly different in the different analyses.

All statistical analyses were performed using the VassarStats suite, Website for Statistical Computation (http://vassarstats.net/).

3. Results

3.1. Conservation of the individual 5'ss zebrafish nucleotides in chicken, mouse and human (SUP-Table 1)

The conservation of each of the four variable 5'ss nucleotides has been determined in couples of *homologous* hexanucleotides of zebrafish/chicken, zebrafish/mouse and zebrafish/human.

At *nucleotide +3* of the 5'ss the purines A and G are highly conserved, whereas the pyrimidines C and T are only rarely conserved.

At *nucleotide +4,* the zebrafish As are highly conserved (more than 75 %) in chicken, mouse, and human, while Gs, Ts, and Cs are poorly conserved.

At *nucleotide +5* of the 5'ss the zebrafish Gs are highly conserved (more than 80 %) in chicken, mouse, and human, while As, Ts, and Cs are poorly conserved.

At *nucleotide +6* of the 5'ss the Ts of the zebrafish sequence are more conserved (50–60 %) than all other nucleotides in chicken, mouse and human.

No statistically significant difference in the nucleotide conservation was found between the tissue-specific SLC25 genes (A3, A7, A12, A14, A27, A30, A31, and A 38) and the other SLC25 genes.

3.2. Conservation of the individual 3'ss zebrafish nucleotides in chicken, mouse and human (SUP-Table 1)

The conservation of each of the four variable 3'ss nucleotides has been determined in couples of *homologous* hexanucleotides of zebrafish/chicken, zebrafish/mouse and zebrafish/human.

At *nucleotides −6 and −5* of the 3'ss the pyrimidines T and C are the most conserved (about 50 % and 30 %, respectively) in chicken, mouse and human.

At *nucleotide −4* of the 3'ss the conservation is relatively low (20 % - 30 %) and tends to be equal in all nucleotides.

At *nucleotide −3* of the 3'ss the zebrafish Cs are conserved by about 70 % and the Ts by about 30 % in chicken, mouse and human sequences.

As at the 5'ss terminal, no statistically significant difference in the nucleotide conservation was found between the tissue-specific SLC25 genes and the other SLC25 genes.

3.3. Global conservation at the different 5'ss and 3'ss positions of the zebrafish nucleotides in birds and mammals (SUP-Table 2)

At each 5'ss and 3'ss position the conservation is remarkably similar in chicken, mouse, and human.

At 5' the global conservation in nucleotides 3, 4, and 5 is similar (50 %–60 %), whereas at nucleotide 6 the conservation is significantly lower (35 %–40 %).

At 3', conservation averages 40 %–45 % at nucleotides −6 and −5; the nucleotide −4 is the least conserved (25 %–30 %); the nucleotide -3 is the most conserved (60 %).

3.4. Global conservation of the different zebrafish nucleotides in birds and mammals, at 5'ss and 3'ss (SUP-Table 3)

The global conservation of the four zebrafish nucleotides is very similar in chicken, mouse and human, at both 5'ss and 3'ss.

At 5' the purines A and G conservation averages 40 %, while the pyrimidines C and T are much less conserved (on average, 10 % and 20 %, respectively).

Conversely, at 3' both pyrimidines are more conserved (about 40 %), while A averages 20 % and G averages only 10 %.

3.5. Frequency of individual nucleotides in zebrafish and the conservation in chicken, mouse and human (SUP-Table 4)

The Table plots on the *abscissa* the per cent frequency of each zebrafish nucleotide at each position in the 5'ss or 3'ss and on the *ordinate* the per cent conservation in chicken, mouse or human. The correlation coefficients and their 95 % confidence limits are also shown.

In all instances the two percentages proved highly positively correlated.

3.6. Mutation of the individual 5'ss and 3'ss zebrafish nucleotides in chicken, mouse and human (SUP-Table 5)

In the joint Table, each possible transition/transversion of a zebrafish nucleotide during evolution to birds and mammals is presented. Nucleotide changes are expressed as the *percentage* of a given zebrafish nucleotide transforming into a different nucleotide in chicken, mouse or human. For example, A→C stands for A transformed into C. Reciprocal changes (e.g., A→C and C→A) are tabulated side by side. In the Table, for each transition/transversion the per cent frequency (± SE) is reported; the two last columns indicate whether the difference between

the two reciprocal changes is statistically not significant ('ns') or significant ($p < 0.05$): in the latter instance the prevailing direction of change is indicated.

At *nucleotide 3* of the 5'ss there is a significant shift of zebrafish pyrimidine nucleotides towards purine nucleotides in chicken and mammals; furthermore, a significant G→A shift is observed in some instances.

At *nucleotide 4* of the 5'ss the G→A transition and the C→A and T→A transversions are highly significant in the evolution to birds or mammals.

At *nucleotide 5* of the 5'ss the A→G transition and the C→G and T→G transversions are highly significant in the evolution to both birds and mammals.

At *nucleotide 6* of the 5'ss the shifts A→T and C→T are clearly the dominant transformations in chicken, mouse and human.

At *nucleotides -6 and -5* of the 3'ss in chicken, mouse and human there is a significant shift from both purines to T and C, but also a shift from C to T, especially remarkable at -5ss.

At *nucleotide −4* of the 3'ss the probability of all possible transitions and transversions is roughly equal.

At *nucleotide −3* of the 3'ss the zebrafish As and Ts tend to change into C; nucleotide G is virtually non-expressed in all species studied.

3.7. Nucleotide frequencies in 5'ss and 3'ss sequences of zebrafish, chicken, mouse and human (SUP-Table 6)

The actual nucleotide composition of 5'ss and 3'ss of the extant chicken, mouse and human is determined by the concurrent effects of the nucleotide conservation and mutations. In SUP-Table 6 are presented, for comparison, the nucleotide frequencies at the 5'ss and 3'ss of zebrafish, chicken, mouse and human. The graphs show that, on the whole, the frequencies do not differ significantly in the species studied. There are, however, a few significant differences: (i) at 5'ss the sequence GTxxGx is significantly less expressed in zebrafish than in mammals (68.25 % in zebrafish; 76.47 % in mouse; 83.64 % in human); (ii) at position +6 of the 5'ss the C is significantly lower and T significantly higher in zebrafish than in all other species; (iii) at 3'ss the sequence xxCxAG is significantly lower and the sequence xxTxAG significantly higher in zebrafish than in all other species.

Besides the aforementioned relatively minor differences, the main features, which are common to all the species, are the following. At position +3 of 5'ss the purines A and, to a lesser extent, G are highly represented, while both pyrimidines are scarcely represented. A at position +4 and G at position +5 account for the great majority of nucleotides at these positions. At position +6 no nucleotide is clearly prevailing.

At the −6 and −5 positions of the 3'ss the pyrimidines T and, to a lesser extent, C are more represented, while both purines are scarcely represented. At -4 no nucleotide is prevailing. At -3 the pyrimidines C and, to a lesser extent, T are more represented.

We studied the actual occurrence of each of these nucleotides at the appropriate positions, in both the 5'ss and 3'ss. In particular, at 5' we searched for the occurrence of either A or G at +3, A at +4, and G at +5. No analysis was carried for the "neural" nucleotide +6. Results are shown in Table 1, listing all these 3-nucleotide sequences found in the material studied, their frequencies and the number of nucleotides matching the prevailing scheme. In 63 % of the sequences there was full matching of this scheme with 3 identities found. However, 30 % of sequences exhibited a matching score of two, the unmatched nucleotide being either the first, the second or the third. Furthermore, in an additional 7 % only one of the nucleotides matched the reference sequence.

A similar analysis was carried at the 3' end, assuming as a reference sequence [C/T] [C/T]x[C/T]AG, i.e., either C or T at positions −6, −5 and −3, the x at -4 representing indifferently one of the four nucleotides. Here a variety of sequences, comprising *all* the combinations of Cs and Ts, matches the reference sequence for a total of 71 %. Almost 26 % exhibit 2 identities over 3, and 2.8 % exhibit only 1 identity.

It is noteworthy that, both at 5'ss and 3'ss, all of the expressed sequences share at least one identity with the reference sequences.

Table 1
The 5'ss reference sequence is assumed as GT[A/G]AGx, x being indifferently one of the four nucleotides. * denotes the unmatched nucleotide.

	GT???x			
Identities	+3 A/G +4 A +5 G	**+3/+4/+5**	**%**	**Total %**
3	AAG	AAG	26.14	
	GAG	GAG	36.93	
				63.07
2	AA*	AAA	3.41	
		AAC	3.98	
		AAT	2.84	
	A*G	ACG	0.57	
		AGG	6.82	
		ATG	6.82	
	G*G	GTG	0.57	
		GGG	2.27	
		GCG	1.14	
	*AG	CAG	1.70	
				30.12
1	A**	ACA	0.57	
		ACC	0.57	
		ACT	0.57	
		AGT	0.57	
		ATT	2.27	
	G**	GCA	0.57	
		GGT	0.57	
	**G	TGG	0.57	
		TTG	0.57	
				6.83

3.8. Estimation of the per cent conservation of zebrafish nucleotides in the birds and mammals common ancestor and in the mammals common ancestor

Separately for each nucleotide type at each position in the 5' and 3' zebrafish hexanucleotides, we determined the number of chicken *and/or* mouse *and/or* human sequences which had kept an identical nucleotide at the same position. The nucleotide under consideration was regarded as having been conserved throughout evolution up to the birds/mammals common ancestor. Otherwise, this nucleotide was regarded as having changed in the evolution from fish and the birds/mammals common ancestor. Similarly, when the zebrafish nucleotide under consideration was present in the mouse *and/or* the human the latter was assumed to have been conserved throughout evolution up to the mammals common ancestor.

For instance, at position +3 of the 5'ss the nucleotide A was present in zebrafish in 104 cases (over the 176 zebrafish-chicken-mouse-human sequences); nucleotide A was present in that position in 85 cases (82 %) in at least one of the chicken-mouse-human sequences and in 74 cases (71 %) in at least one of the mouse-human sequences.

Although these estimates of the common ancestors under consideration are somewhat biased due to overlooking possible revertant mutations, this approach seemed to be indicative for general evolutionary patterns.

Tables 3 and 4 summarize the conservation percentages of zebrafish nucleotides, at 5'ss and 3'ss, in the birds/mammals ancestor, in mammals ancestor and in the extant chicken, mouse and human.

At 5'ss the percentage of conservation in chicken appears to be similar to the conservation in the mammals ancestor; as for mouse and

human, a further loss of conserved nucleotides takes place in the evolution starting from their common ancestor. The conservation loss appears to be higher in the evolution from zebrafish to the common birds/mammals ancestor (about 58 %) than in the further evolution to the common mammalian ancestor (about 10 %) and in the final evolution to the extant mammals (about a further 5 %).

At 3'ss, on the contrary, the percentage of conservation in the extant chicken appears to be similar to the conservation in the extant mouse and human. Once again, the conservation loss is higher in the evolution from zebrafish to the common birds/mammals ancestor (about 55 %) than in the further evolution to the common mammalian ancestor (about 10 %) and in the final evolution to the extant mammals (about a further 8–9 %).

Another significant index to evaluate the 5'ss and 3'ss conservation during evolution from zebrafish to birds and mammals is the percentage of unmodified complete (6 nucleotides) sequences. The zebrafish/chicken conservation is 7.7 % at 5'ss and 3.3 % at 3'ss; the zebrafish/mouse conservation is 8.8 % at 5'ss and 3.8 % at 3'ss; the zebrafish/human conservation is 7.1 % at 5'ss and 3.2 % at 3'ss.

The significance of the different conservation indexes reported in this section will be considered in the *Discussion* section.

3.9. Frequency of individual nucleotides in zebrafish and the conservation in the last common ancestor of birds and mammals (SUP-Table 7)

The Table plots on the *abscissa* the per cent frequency of each zebrafish nucleotide at each position in the 5'ss or 3'ss and on the *ordinate* the estimated per cent conservation in the last common ancestor of birds and mammals. The data plotted refer to those nucleotides whose absolute frequency in zebrafish is higher than 5.

The two parameters are strongly positively correlated: the r coefficient is 0.62, with 95 % confidence limits 0.23 and 0.84.

3.10. Evolution of 5'ss and 3'ss nucleotides from the common birds/mammals ancestor to chicken, mouse and human (SUP-Table 8 and SUP-Table 9)

We also evaluated the rate of nucleotide conservation during the evolution from the last common birds/mammals ancestor to the extant chicken and, through a common mammalian ancestor, to mouse and human. In the SUP-Table 8 the following data are plotted: the per cent conservation of zebrafish nucleotides in the *birds/mammals common ancestor* against (i) the ratio: conservation in the mammals common ancestor divided by conservation in the birds/mammals common ancestor, (ii) the ratio: conservation in the extant chicken divided by conservation in the birds/mammals common ancestor, (iii) the ratio: conservation in the extant mouse divided by conservation in the birds/mammals common ancestor, and (iv) the ratio: conservation in the extant human divided by conservation in the birds/mammals common ancestor.

These ratios measure the level of nucleotide conservation in the further evolution from the last common ancestor of birds and mammals. The theoretical minimum value of the ratio is zero, when no nucleotide is conserved; the maximum value is 1, denoting that all nucleotides have been conserved.

In SUP-Table 8 the following correlations are also shown: the per cent conservation of zebrafish nucleotides in the *mammals common ancestor* against (i) the ratio: conservation in the extant mouse divided by conservation in the mammals common ancestor, and (ii) the ratio: conservation in the extant human divided by conservation in the mammals common ancestor. These ratios measure the level of nucleotide conservation in the further evolution from the last common ancestor of mammals to mouse or human.

The referred data refer to nucleotides whose absolute frequency is higher than 5 and the data from 5'ss and 3'ss are pooled.

All interpolating straight lines have a positive slope, indicating that the higher the nucleotide conservation in the ancestors the higher the conservation in the further evolution. Note, however, that the angular coefficient is significantly higher when comparing the extant mouse or human with the birds/mammals common ancestor than in the comparison with the more recent mammals common ancestor (angular coefficients about 0.008 and 0.002, respectively).

In the SUP-Table 9 the whole set of the above correlations is summarized. The estimated correlation coefficients are reported together with the 95 % confidence limits; the significant (statistically > 0) correlation coefficients are marked with an asterisk. The last column of the Table lists the angular coefficients.

The correlations between the birds/mammals common ancestor and the mammals common ancestor are positive both at 5' and 3', but may be not-significant at the 3'. However, the correlation with pooled 5' and 3' is significant.

The correlations between the chicken, mouse and human sequences with the birds/mammals common ancestor are positive and significant at both 5'ss and 3'ss, except in the case of the chicken 5'ss, in which, however, the correlation is significant for 3'ss and for pooled 5' and 3'.

The correlations between the mouse and human sequences with the mammals common ancestor are positive at both 5'ss and 3'ss, but possibly not-significant, except in the case of the mouse 5'ss, where the correlation is significant.

The slope figure of the interpolating straight line is always higher in the correlations with the birds/mammals ancestor than in the correlations with the mammals ancestor.

3.11. Frequency of couples of 5'ss and 3'ss nucleotides in zebrafish and the conservation in chicken, mouse and human (SUP-Table 10)

The conservation of *couples* of 5'ss and 3'ss zebrafish nucleotides in chicken, mouse and human was investigated. The couples considered consisted of both adjacent and non-adjacent nucleotides, but couples with less than six nucleotides were excluded.

By comparison with the actual experimental data, we calculated a theoretical conservation percentage by multiplying the conservation percentages of the individual constitutive nucleotides (*Sections 1 and 2*).

As an example, the actual per cent conservation of the $+4$ and $+5$ dinucleotide AG (i.e., GTxAGx) of chicken (the most highly expressed dinucleotide) is 67.02 %, while the theoretical conservation is 61.45 (76.52×80.31).

SUP-Table 10 is a representative example of this analysis at 5'ss and 3'ss. The frequency of a given nucleotide couple in the zebrafish (abscissa) is plotted against the actual percentage of *conservation* of this couple in chicken (ordinate); the interpolating straight line is also represented. The correlation coefficient is 0.844 (95 % confidence limits 0.73 and 0.91) at 5' and 0.735 (95 % confidence limits 0.55 and 0.85) at 3'. The statistical parameters of the theoretical distribution are very similar and the corresponding interpolating straight line is practically superimposable to the experimental straight line.

The analysis of mouse and human data yielded similar results.

To conclude, the conservation of nucleotide couples at 5'ss and 3'ss in birds and mammals is proportional to the frequency in zebrafish, and the rate of conservation is determined, at least prevailingly, by the combined conservations of the two elements of the couple.

4. Discussion

The majority of mitochondrial solute carrier gene pre-mRNAs transcripts of zebrafish, chicken, mouse, and human share a complete homology in the alternation of exonic and intronic sections, so that corresponding introns are topographically homologous, although usually differing in composition. By studying couples of these strictly homologous zebrafish/chicken, zebrafish/mouse, and zebrafish/human introns we analyzed in detail the evolution of each nucleotide of the initial and terminal hexanucleotides, which are thought to be major

agents in splicing signaling.

The evolutionary period covered spans from the appearance of the bony fishes (Euteleostomi) about 420–450 million years ago (MYA) to the emergence of a last common (Sarcopterygian) ancestor that gave rise to birds and mammals approximately 300–310 MYA and eventually the rodents/primates divergence some 65–100 MYA (Foote et al., 1999; Lee, 1999; Nei et al., 2001; Nobrega and Pennacchio, 2004; Broughton et al., 2013; Betancur et al., 2013).

At 5' the first two intronic nucleotides are, in accordance with the general rule, G and T, but the other members of the hexanucleotide are highly variable: e.g., in 252 couples of zebrafish and human introns we found 69 different configurations in zebrafish and 49 different configurations in human (out of the $4^4 = 256$ possible configurations), several of the configurations being present only once.

At 3' the last two intronic nucleotides are, as usual, always A and G, but the other members of the hexanucleotide are even more variable than at 5': e.g., in the same couples of zebrafish and human introns we found 68 different configurations in zebrafish and 71 different configurations in human.

It should be appreciated that, despite the structural changes in the initial and terminal hexanucleotides during evolution, the location of the intronic inserts has remained unaltered, whereas the length and the structure of the corresponding introns has often changed dramatically (*see* further on).

We studied the evolutionary dynamics of each of the 5'ss and 3'ss nucleotides, from fish (zebrafish) to birds (chicken) or mammals (mouse and human). The two parameters considered are the conservation of zebrafish nucleotides in chicken, mouse and human (dealt with in *Sections 1 and 2*) and the changes of nucleotides from a given type to a different type (dealt with in *Section 6*).

For instance, the absolute incidence of GTAxxx in zebrafish is 109 nucleotides (out of a total 183, i.e., 59.56 %); of these, 78 (71.56 %) are conserved in chicken (*Supplementary Material*, SUP-Table 1); the conserved nucleotides account for 42.62 % only of the total chicken nucleotides; however, a portion of G, C, and T zebrafish nucleotides (i.e., 29 Gs, 7 Cs, and 4 Ts) are transformed into As; as a result, the GTAxxx eventually accounts in chicken for a total of 118 nucleotides (64.48 %), a number similar to that in zebrafish, although a little higher.

The percentages of conservation of each element of boundary hexanucleotides in the evolution along different paths, at 5' and 3', are shown in *Supplementary Material*, SUP-Table 1. In summary, in all species the more conserved nucleotide types are A at positions 3 and 4, G at position 5 and T at position 6 of the 5'ss.

The relative uniformity in the conservation parameters in chicken, mouse and human is evidenced in *Supplementary Material*, SUP-Table 3, also showing a prevalent purination at 5'ss and a prevalent pyrimidation at 3'ss.

In addition, *Supplementary Material*, SUP-Table 2 shows that at each position of the 5'ss and 3'ss sequences the *total* conservation (all nucleotides) is almost the same in chicken, mouse and human (*Section 3*).

The analysis of the conservation in chicken, mouse and human of the individual 5'ss and 3'ss nucleotides as a function of the actual frequency in zebrafish reveals a significant positive correlation between the two parameters (*Section 5* and *Supplementary Material*, SUP-Table 4).

The detailed analysis of the frequencies of all possible transitions and transversions (*Supplementary Material*, SUP-Table 5) reveals the existence of preferred transformations in the evolutionary paths from zebrafish to chicken or mouse and human.

At 5'ss the more significant shifts are, in all species, towards A and C at +3, towards A at +4, towards G at +5 and towards T at +6.

At 3'ss the more significant shifts are, in all species, towards C and T at -6 and -5, towards C at -3. Remarkably, at -4 all couples of reciprocal changes are virtually balanced.

Thus, aside from minor variations, the evolutionary changes favor at each location a shift towards the more conserved nucleotides, which in turn correspond to the nucleotides with a higher frequency in zebrafish. While the conservation is never total even in the more represented nucleotides, the evolutionary changes of type of the other nucleotides tend to compensate for the loss of non-conserved nucleotides.

As a result of these mutually compensating evolutionary dynamics the net effect is virtually a long-lasting preservation, over about 400 million years from fish to birds and mammals, of the nucleotide composition of the 5' and 3' signals (*Section 7* and *Supplementary Material*, SUP-Table 6).

However, in the *Section 7* a few relatively small, albeit significant, differences between zebrafish and chicken/mammals at specific 5'ss and 3'ss positions are recorded. These still poorly understood differences appear to deserve further investigation.

In conclusion, the *general* composition of the 5'ss and 3'ss sequences is, on the whole, similar in the four species studied (*Section 7* and SUP-Table 6) and in broad terms matches the results of previous investigations of vertebrate introns (e.g., Schwartz et al., 2008; Calvello et al., 2016). This is not surprising considering the "damping" effect of the mutually compensating evolutionary dynamics described above, operating at least throughout the whole vertebrate evolution.

A clear pyrimidine-rich signal near the 3′ end of introns has been described in metazoans (Abril et al., 2005; Burset et al., 2000; Schwartz et al., 2008). In our material, at 3'ss the pyrimidines C and T are the more represented at positions -6 and -5 (with the exception of 18.6 % of cases only; *see* Table 2).

According to previous reports (e.g., Schwartz et al., 2008) at position -3 (immediately before the terminal AG) the frequency of bases is C, T, and A in a decreasing order. In our material, C, T, and A represent 66 %, 30 %, and 4%, respectively (*see* Table 2), while G is only rarely

Table 2
The 3'ss reference sequence is assumed as [C/T] [C/T]x[C/T]AG, x being indifferently one of the four nucleotides. * denotes the unmatched nucleotide.

	??x?AG			
Identities	−6 C/T −5 C/T −3 C/T	−6/-5/-3	%	**Total %**
3	CCxC	CCxC	11.30	
	CCxT	CCxT	1.69	
	CTxC	CTxC	9.60	
	CTxT	CTxT	3.39	
	TCxC	TCxC	5.65	
	TCxT	TCxT	2.82	
	TTxC	TTxC	22.60	
	TTxT	TTxT	14.12	
				71.17
2	*CxC	ACxC	3.39	
	*CxT	ACxT	0.56	
	*TxC	ATxC	3.39	
	*TxT	ATxT	2.26	
	C*xC	CAxC	1.69	
	CCx*	CCxA	0.56	
	CTx*	CTxA	0.56	
	*CxT	GCxT	0.56	
	*TxC	GTxC	2.82	
	*TxT	GTxT	0.56	
	T*xC	TAxC	3.39	
	T*xT	TAxT	1.13	
	TCx*	TCxA	0.56	
	T*xC	TGxC	1.13	
	T*xT	TGxT	2.26	
	TTx*	TTxA	1.13	
				25.95
1	**xC	AAxC	0.56	
	**xT	AAxT	0.56	
	Tx	ATxA	0.56	
	C*x*	CAxA	0.56	
	**xC	GGxC	0.56	
				2.80

Table 3
Conservation of 5'ss zebrafish nucleotides.

5'ss	Conservation (%)		
Birds/Mammals Ancestor		42.55	
Mammals Ancestor		32.30	
Extant	Chicken	Mouse	Human
	31.16	27.13	27.61

Table 4
Conservation of 3'ss zebrafish nucleotides.

3'ss	Conservation (%)		
Birds/Mammals Ancestor		45.32	
Mammals Ancestor		35.10	
Extant	Chicken	Mouse	Human
	26.69	26.19	27.43

present at this position (*see also* Akerman and Mandel-Gutfreund, 2006).

For the differential splicing of the 3' tandem sequence NAGNAG, *see* below.

In order to trace the main evolutionary steps of the 5' and 3' splicing signals, we attempted to estimate the conservation of the zebrafish nucleotides to the last common ancestor of birds and mammals and also the conservation of the zebrafish nucleotides to the mammals common ancestor. *Section 8* details the criteria which were followed in order to estimate the conservation of the zebrafish nucleotides up to the birds/mammals common ancestor or the mammals common ancestor. In this section the conceptual limits and possible drawbacks of this approach are also discussed, although it seemed to us to be able to offer some indicative clues.

An interesting conclusion of this analysis is that at both 5'ss and 3'ss the loss of nucleotide conservation was clearly higher in the evolution from fish to a birds/mammals common ancestor, while the evolution was more conservative thereafter (*Section 8* and Tables 3 and 4). A closer analysis revealed that the conservation of zebrafish nucleotides in the last common ancestor of birds and mammals is positively correlated with the actual frequency in zebrafish (*Section 9* and *Supplementary Material* – SUP-Table 7). Furthermore, the nucleotide conservation from the last common ancestor of birds and mammals to the extant chicken is positively correlated with the frequency in the birds/mammals ancestor (*Supplementary Material* – SUP-Table 8, B).

The nucleotide conservation in the evolution from the last common ancestor of birds and mammals to the last mammalian ancestor and eventually to the extant mouse or human is illustrated in the graphs in *Supplementary Material* – SUP-Table 8 (C/D and E/F, respectively). In all these evolutionary steps the conservation in each succeeding step is proportional to the conservation in the preceding step. The level of the proportionality in conservation is expressed by the correlation coefficient and the slope of the interpolating straight line in the graphs (*Supplementary Material* – SUP-Table 10). It is noteworthy that the proportionality is more prominent in the evolution from the last common ancestor of birds/mammals to the mammalian ancestor than in the evolution from the last common mammalian ancestor to the extant mouse and human.

In summary, throughout all the evolutionary steps from fish to birds and mammals the nucleotide conservation at the individual 5'ss and 3'ss nucleotides is proportional to the actual frequency in zebrafish. Furthermore, at these sites the nucleotide mutations do not happen randomly but are biased in favor of the nucleotides which are more represented in zebrafish. These dynamics result in an overall virtual evolutionary stability of the 5' and 3' splicing signals. Within this reference scenario, some nucleotides at specific positions appear to be endowed with a more significant information content, due to their particularly high incidence and conservation.

As described at *Section 7*, at 5', A or G are more represented at position +3, A is more represented at +4 and G at +5. The high conservation rate of AG at positions +4 and +5 (together with a significant rate of gain of novel As and Gs by mutation of other nucleotides) is in keeping with the finding that, comparatively, the dinucleotide AG has the highest probability of base-pairing with the snRNA (short nuclear RNA) U1, an event which initiates the splicing process (O'Reilly et al., 2013; Guiro and O'Reilly, 2015). At position +6 no nucleotide is clearly prevailing and thus it is likely that this nucleotide is scarcely informative for splicing.

At 3' T or C are prevailing at positions -6, -5 and -3 and thus should be regarded as endowed with a high information content. On the contrary, at position -4 no nucleotide is clearly prevailing and thus it is likely that this nucleotide is scarcely informative for splicing. At -3 A accounts for less than 4 % and is possibly poorly informative; at this position G is only exceptionally present and indeed is prone to disappear by mutation (Lev-Maor et al., 2003; Calvello et al., 2013). In the present material, the terminal GAG sequence appears only once in a human intron, but the corresponding mouse sequence is AAG. In another intron the terminal GAG is present in chicken, but corresponds to TAG in zebrafish, mouse and human. Thus, possibly G at this position carries a strong signal disqualifying the sequence, in most cases, as an intron terminal.

As remarked in *Section 7*, in the majority of the expressed 5'ss and 3'ss sequences the "informative" nucleotides which are conserved at their canonical positions are two or, more often, three; in a minority of cases (about 7 % at 5'ss and 3 % at 3'ss) only a single match is conserved; however, no expressed sequence is completely devoid of any such match. These data might suggest that at least a single nucleotide at its canonical position in the 5'ss and 3'ss sequences is sufficient to support a correct splice. However, this conclusion must be considered with caution because several extrinsic factors play major roles in the splicing process.

The strong pyrimidation at -6 and -5 suggests that these nucleotides might represent the downstream end of the polypyrimidine tract, possibly extending down to -5 in vertebrates (Schwartz et al., 2008). Thus, considering the "neutral" role of the -4 nucleotide, the effective 3' signal could include the last three intron nucleotides only, i.e., CAG, TAG, or, much less frequently, AAG.

Since the individual 5'ss and 3'ss nucleotides were shown to have been differentially conserved, we addressed the question whether selected *couples* of nucleotides were differentially conserved by virtue of the specific nucleotide association, or else the expression of each nucleotide couple depended exclusively on the conservation levels of the nucleotides composing the couple. To this end we determined, for couples of adjacent and non-adjacent nucleotides, the per cent frequency in zebrafish and the *actual* percentage of conservation in chicken, mouse and human. Then we calculated a corresponding *theoretical* percentage of conservation by multiplying the conservation percentages of the two nucleotides of the couple. The statistical analysis demonstrated no significant difference between the actual and calculated percentages of conservation (*Section 11*). It may thus be concluded that the biological control of conservation and mutation during evolution is exerted prevailingly, if not exclusively, at the level of individual nucleotides.

Other factors possibly contribute to the overall relative conservation of 5'ss and 3'ss. Primarily, the rarity of "physiological" alternatively spliced transcripts. The only well documented alternative splicing was described in the SLC25A3 gene with the expression of two similar isoforms (Dolce et al., 1994, 1996; Fiermonte et al., 1998), but the generation of the two isoforms was demonstrated to result from a duplication of a section of the gene (Calvello et al., 2018). Splicing variants of other SLC25 genes have also been described, but these apparently generated short-lived protein products only (Del Arco, 2005; Bassi et al., 2005).

Another proof of an inherent bias towards the stability of the splicing patterns is given by the splicing events at the 3' tandem sequence NAGNAG, which could be alternatively spliced, contributing to the structural and functional protein diversity (Hiller et al., 2006, 2008; Yan et al., 2015; Hujová et al., 2019). In mammals the majority of such sites are spliced after the proximal AG (the remaining NAG being the 5' end of the next exon), while in about one tenth of instances the splicing occurs after the second AG (Akerman and Mandel-Gutfreund, 2006). In our series of human SLC25 genes 3' NAGNAG sequences occur 15 times and the splicing always occurs after the proximal AG, indicating a strong evolutionary conditioning against an alternate splicing which would alter the coding sequence.

Transposons are mobile segments of genetic material which may settle in introns modifying their structure (Chalopin et al., 2015), but it has been demonstrated that the 5' and 3' ends of introns are relatively refractory to the transposon invasion (Cianciulli et al., 2017).

In addition, the global conservation of the intron sequences (as evaluated from the length of sections that can be significantly aligned between two species) averages 0.23-0.27 % from zebrafish to chicken, mouse and human (Calvello et al., 2019).

To summarize, in the SLC25 genes the evolutionary conservation between vertebrate species is estimated to vary from about 70–80 % in exons, from 26 % to 31 % at the intron 5'ss and 3'ss (*Section 8*) to become dramatically reduced to about 0.25 % in the whole introns.

A conservation analysis of both the individual nucleotides and the whole sequences of the 5' and 3' splicing sites of all SLC25 genes between zebrafish and chicken, mouse and human is reported in *Section 8*. The per cent conservation values concerning the individual nucleotides shown in Tables 3 and 4 vary between 27 and 31 % at 5'ss and 26 and 27 % at 3'ss. From these data it can be calculated the theoretical concurrent conservation of all four variable nucleotides (the fourth power of the data of the individual nucleotides), which should vary between 0.5 % and 0.9 %. By contrast, the percentage of the complete zebrafish sequences conserved in chicken, mouse and human varies between 7 % and 9 % at 5'ss and 3 % and 4 % at 3'ss, which would correspond to a calculated conservation average of the individual nucleotides of 53 % at 5'ss and 43 % at 3'ss. These differences depend on which of the two nucleotide populations is considered: in the first case, the whole population of 5'ss and 3'ss nucleotides is considered, irrespective of their percent representation in the splicing signals; on the contrary, the second statistics is based on the nucleotides represented in the more conserved sequences and thus preferentially represented. This stresses the existence of a bias of the splicing signals composition towards specific nucleotide types, in spite of the wide structural variability.

In conclusion, the family of the SLC25 homologous genes presents aspects of an overall long-term evolutionary exon/intron architecture stability; this stability seems to be supported by different evolutionary mechanisms ensuring a certain, albeit low, degree of conservation of the splicing signals at the intron ends. However, more generally, all available evidence (Shi, 2017; Baralle and Baralle, 2018) suggest that the correct physiological splicing depends upon a large number of different factors, which, in addition, may be modulated to a certain extent by the milieu in which they are read, including not only the genomic contest but also the tissue concerned and the developmental stage. For these reasons, despite a substantial amount of available information, the achievement of a robust general algorithm of the splicing process remains elusive.

Author contributions

Data analysis: V. Mitolo; wrote the paper: R. Calvello and A. Cianciulli; review of the paper: M.A. Panaro

Declaration of Competing Interest

The authors declare no conflict of interest

Acknowledgments

Thanks are due to Mrs Mary V. C. Pragnell for linguistic text revision.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.compbiolchem.2020.107251.

References

Abril, J.F., Castelo, R., Guigó, R., 2005. Comparison of splice sites in mammals and chicken. Genome Res. 15 (1), 111–119. https://genome.cshlp.org/content/15/1/111.

Akerman, M., Mandel-Gutfreund, Y., 2006. Alternative splicing regulation at tandem 3' splice sites. Nucleic Acids Res. 34 (1), 23–31. https://doi.org/10.1093/nar/gkj408.

Arias, M.A., Lubkin, A., Chasin, L.A., 2015. Splicing of designer exons informs a biophysical model for exon definition. RNA 21 (2), 213–229. https://doi.org/10.1261/rna.048009.114.

Baralle, M., Baralle, F.E., 2018. The splicing code. Biosystems 164, 39–48. https://doi.org/10.1016/j.biosystems.2017.11.002.

Bassi, M.T., Manzoni, M., Bresciani, R., Pizzo, M.T., Della Monica, A., Barlati, S., Monti, E., Borsani, G., 2005. Cellular expression and alternative splicing of SLC25A23, a member of the mitochondrial Ca^{2+}-dependent solute carrier gene family. Gene 345 (2), 173–182. https://doi.org/10.1016/j.gene.2004.11.028.

Betancur-R, R., Broughton, R.E., Wiley, E.O., Carpenter, K., López, J.A., Li, C., Holcroft, N.I., Arcila, D., Sanciangco, M., Cureton Ii, J.C., Zhang, F., Buser, T., Campbell, M.A., Ballesteros, J.A., Roa-Varon, A., Willis, S., Borden, W.C., Rowley, T., Reneau, P.C., Hough, D.J., Lu, G., Grande, T., Arratia, G., Ortí, G., 2013. The tree of life and a new classification of bony fishes. PLoS Curr. http://currents.plos.org/treeoflife/index.html%3Fp=4341.html.

Broughton, R.E., Betancur-R, R., Li, C., Arratia, G., Ortí, G., 2013. Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. PLoS Curr. http://currents.plos.org/treeoflife/index.html%3Fp=2607.html.

Burset, M., Seledtsov, I.A., Solovyev, V.V., 2000. Analysis of canonical and non-canonical splice sites in mammalian genomes. Nucleic Acids Res. 28 (21), 4364–4375. https://doi.org/10.1093/nar/28.21.4364.

Burset, M., Seledtsov, I.A., Solovyev, V.V., 2001. SpliceDB: database of canonical and non-canonical mammalian splice sites. Nucleic Acids Res. 29 (1), 255–259. https://doi.org/10.1093/nar/29.1.255.

Calvello, R., Cianciulli, A., Panaro, M.A., 2013. Conservation/Mutation in the splice sites of cytokine receptor genes of mouse and human. Int. J. Evol. Biol. 2013, 818954. https://doi.org/10.1155/2013/818954. 2013.

Calvello, R., Panaro, M.A., Salvatore, R., Mitolo, V., Cianciulli, A., 2016. Conservation/Mutation in the splice sites of mitochondrial solute carrier genes of vertebrates. J. Mol. Evol. 83 (3-4), 147–155. https://doi.org/10.1007/s00239-016-9762-8.

Calvello, R., Cianciulli, A., Panaro, M.A., 2018. Unusual structure and splicing pattern of the vertebrate mitochondrial solute carrier SLC25A3 gene. J. Genet. 97 (1), 225–233. https://doi.org/10.1007/s12041-018-0906-z.

Calvello, R., Cianciulli, A., Mitolo, V., Porro, A., Panaro, M.A., 2019. Conservation of intronic sequences in vertebrate mitochondrial solute carrier genes (zebrafish, chicken, mouse and human). Noncoding RNA 5 (1). https://doi.org/10.3390/ncrna5010004. pii: E4.

Chalopin, D., Naville, M., Plard, F., Galiana, D., Volff, J.N., 2015. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. Genome Biol. Evol. 7 (2), 567–580. https://doi.org/10.1093/gbe/evv005.

Chasin, L.A., 2007. Searching for splicing motifs. Adv. Exp. Med. Biol. 623, 85–106. https://www.ncbi.nlm.nih.gov/pubmed/18380342.

Cianciulli, A., Calvello, R., Panaro, M.A., 2017. Transposable elements: a comparative study in the introns and UTRs of the homologous mitochondrial solute carrier genes of Human, Mouse and Zebrafish. Int. J. Struct. Comput. Biol. 1 (1), 1–24. https://symbiosisonlinepublishing.com/quantitative-computationalbiology/structural-computational-biology02.php.

Del Arco, A., 2005. Novel variants of human SCaMC-3, an isoform of the ATP-Mg/Pi mitochondrial carrier, generated by alternative splicing from 3'-flanking transposable elements. Biochem. J. 389 (Pt3), 647–655. https://doi.org/10.1042/BJ20050283.

Dolce, V., Iacobazzi, V., Palmieri, F., Walker, J.E., 1994. The sequences of human and bovine genes of the phosphate carrier from mitochondria contain evidence of alternatively spliced forms. J. Biol. Chem. 269 (14), 10451–10460. https://www.jbc.org/content/269/14/10451.long.

Dolce, V., Fiermonte, G., Palmieri, F., 1996. Tissue-specific expression of the two isoforms of the mitochondrial phosphate carrier in bovine tissues. FEBS Lett. 399, 95–98. https://doi.org/10.1016/S0014-5793(96)01294-X.

Fiermonte, G., Palmieri, L., Dolce, V., Lasorsa, F.M., Palmieri, F., Runswick, M.J., Walker,

J.E., 1998. The sequence, bacterial expression, and functional reconstitution of the rat mitochondrial dicarboxylate transporter cloned via distant homologs in yeast and Caenorhabditis elegans. J. Biol. Chem. 273 (38), 24754–24759. https://www.jbc.org/content/273/38/24754.

Foote, M., Hunter, J.P., Janis, C.M., Sepkoski Jr., J.J., 1999. Evolutionary and preservational constraints on origins of biologic groups: divergence times of eutherian mammals. Science 283 (5406), 1310–1314. https://science.sciencemag.org/content/283/5406/1310.full.

Guiro, J., O'Reilly, D., 2015. Insights into the U1 small nuclear ribonucleoprotein complex superfamily. Wiley Interdiscip. Rev. RNA 6 (1), 79–92. https://doi.org/10.1002/wrna.1257.

Hiller, M., Huse, K., Szafranski, K., Jahn, N., Hampe, J., Schreiber, S., Backofen, R., Platzer, M., 2006. Single-nucleotide polymorphisms in NAGNAG acceptors are highly predictive for variations of alternative splicing. Am. J. Hum. Genet. 78 (2), 291–302. https://doi.org/10.1086/500151.

Hiller, M., Szafranski, K., Huse, K., Backofen, R., Platzer, M., 2008. Selection against tandem splice sites affecting structured protein regions. BMC Evol. Biol. 8, 89. https://bmcevolbiol.biomedcentral.com/track/pdf/10.1186/1471-2148-8-89.

Hujová, P., Grodecká, L., Souček, P., Freiberger, T., 2019. Impact of acceptor splice site NAGTAG motif on exon recognition. Mol. Biol. Rep. 46 (3), 2877–2884. https://doi.org/10.1007/s11033-019-04734-6.

Kadowaki, T., 2015. Evolutionary dynamics of metazoan TRP channels. Pflugers Arch. 467 (10), 2043–2053. https://doi.org/10.1007/s00424-015-1705-5.

Ke, S., Chasin, L.A., 2011. Context-dependent splicing regulation: exon definition, co-occurring motif pairs and tissue specificity. RNA Biol. 8 (3), 384–388. https://doi.org/10.4161/rna.8.3.14458.

Lee, M.S., 1999. Molecular clock calibrations and metazoan divergence dates. J. Mol. Evol. 49 (3), 385–391. https://doi.org/10.1007/PL00006562.

Lev-Maor, G., Sorek, R., Shomron, N., Ast, G., 2003. The birth of an alternatively spliced exon: 30 splice-site selection in Alu exons. Science 300 (5623), 1288–1291. www.sciencemag.org/cgi/content/full/300/5623/1288/DC1.

Nei, M., Xu, P., Glazko, G., 2001. Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. Proc. Natl. Acad. Sci. U. S. A. 98 (5), 2497–2502. https://doi.org/10.1073/pnas.051611498.

Nilsen, T.W., 2003. The spliceosome: the most complex macromolecular machine in the cell? BioEssays 25 (12), 1147–1149. https://doi.org/10.1002/bies.10394.

Nobrega, M.A., Pennacchio, L.A., 2004. Comparative genomic analysis as a tool for biological discovery. J. Physiol. 554 (Pt1), 31–39. https://doi.org/10.1113/jphysiol.2003.050948.

O'Reilly, D., Dienstbier, M., Cowley, S.A., Vazquez, P., Drozdz, M., Taylor, S., James, W.S., Murphy, S., 2013. Differentially expressed, variant U1 snRNAs regulate gene expression in human cells. Genome Res. 23 (2), 281–291. https://doi.org/10.1101/gr.142968.112.

Palmieri, F., 2013. The mitochondrial transporter family SLC25: identification, properties and physiopathology. Mol. Aspects Med. 34 (2-3), 465–484. https://doi.org/10.1016/j.mam.2012.05.005.

Palmieri, F., Monné, M., 2016. Discoveries, metabolic roles and diseases of mitochondrial carriers: a review. Biochim. Biophys. Acta 1863, 2362–2378. https://doi.org/10.1016/j.bbamcr.2016.03.007.

Palmieri, F., Pierri, C.L., 2010. Mitochondrial metabolite transport. Essays Biochem. 47, 37–52. https://doi.org/10.1042/bse0470037.

Schwartz, S.H., Silva, J., Burstein, D., Pupko, T., Eyras, E., Ast, G., 2008. Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. Genome Res. 18, 88–103. https://doi.org/10.1101/gr.6818908.

Shi, Y., 2017. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. Nat. Rev. Mol. Cell Biol. 18 (11), 655–670. https://doi.org/10.1038/nrm.2017.86.

Ule, J., Blencowe, B.J., 2019. Alternative splicing regulatory networks: functions, mechanisms, and evolution. Mol. Cell 76 (2), 329–345. https://doi.org/10.1016/j.molcel.2019.09.017.

Wan, R., Bai, R., Zhan, X., Shi, Y., 2019. How is precursor messenger RNA spliced by the spliceosome? Annu. Rev. Biochem. https://doi.org/10.1146/annurev-biochem-013118-111024.

Wilkinson, M.E., Charenton, C., Nagai, K., 2019. RNA splicing by the spliceosome. Annu. Rev. Biochem. https://doi.org/10.1146/annurev-biochem-091719-064225.

Yan, X., Sablok, G., Feng, G., Ma, J., Zhao, H., Sun, X., 2015. NAGNAG: identification and quantification of NAGNAG alternative splicing using RNA-Seq data. FEBS Lett. 589 (15), 1766–1770. https://doi.org/10.1016/j.febslet.2015.05.029.