

INTRON ORIGINS AND TRANSLATIONAL STREAMLINING

David Elder

1/42 High St

Grange SA 5022 Australia

Email: david_elder28@hotmail.com.au

08 January 2008

Abstract

There is evidence that introns are basically ancient, though intron insertion events can also occur. Introns in the various classes of split gene may have been derived by 'translational streamlining' from ancestral elements formerly performing functional roles in early cells, roles which later became redundant. For example, the exons of structural genes could have originally been small 'minigene' mRNAs cotranscribed polycistronically, and the spliceosomal introns could formerly have been excised spacers or cleavage sites between the minigene cistrons. Similar ideas can be advanced for introns of tRNA and perhaps of rRNA.

The intron problem

The discovery of introns in the late 1970s was highly unexpected (Crick, 1979). Introns of various forms have been found in a variety of gene classes, not only in nuclear genomes, but also in non-nucleated genomes, those of organelles, mitochondrion and chloroplast, and to a degree even in prokaryote genomes. The origin of these strange intron sequences is a major area of debate. Here it is argued that introns are basically early, though with some later additions also. It is further argued that an introns-early model called 'translational streamlining' can explain the presence of introns in each kind of split gene. This model was stated previously (Elder, 1987). It can now be restated in an improved form, and with greater experimental support for it. Before discussing this model, it is appropriate to first review the various types of introns, and also the empirical evidence bearing on their origins.

Intron types

The great majority of introns can be classified into three groups:

- (1) group I
- (2) lariat, with two subclasses, (a) group II, and (b) spliceosomal
- (3) tRNA-type.

To briefly consider each in turn:

- (1) *Group I introns* (reviewed by Haugen et al., 2005). These are large, complexly folded introns excising by 5' addition of an incoming guanosine and cyclisation. They are

capable of ribozymal self-splicing in vitro, though protein maturases assist their excision in vivo. Group I introns are the commonest intron type in ribosomal RNA genes; they can be found in a variety of positions in the rRNA cistrons, and can be present in this gene class in many genomes, of eukaryotic nuclei and organelles, and sometimes of bacteria (Jackson et al., 2002). The bacterial examples may represent lateral transfers from eukaryotic organelles (Nesbo and Doolittle, 2003). Group I introns are not restricted to rRNA genes. For example, this intron type is sometimes seen in tRNA genes of eubacteria (Vepritskiy et al., 2002). In one of these cases, a split leucyl-tRNA gene, the group I intron appears to have been vertically inherited by the eubacteria's endosymbiotic descendent the chloroplast; if so, this is a very ancient intron (Besendahl et al., 2000). Group I introns can also sporadically be found in structural genes of organelles (Michel and Dujon, 1983), of eubacteria (Braun et al., 2000; Ko et al., 2002), and of viruses both prokaryotic (Landthaler and Shub, 1999) and eukaryotic (Yamada et al., 1994).

(2) *Lariat introns*, splicing out by a 2',5' linkage to an internal adenosine. As noted above there are two subclasses:

(a) *Group II introns* of organellar/prokaryotic genomes. Again this is a large and structurally complex intron form, capable of self-splicing in vitro, but using maturase proteins to help in excision in vivo. These introns can be found in organellar genes for mRNA, tRNA and rRNA (reviewed by Bonen and Vogel, 2001). Group II introns can also be present in prokaryotes, both eubacterial and archaea, where they commonly behave like mobile elements of retron character. Such group II introns appear to have been present in prokaryotes from very early times (Toro, 2003).

(b) *Spliceosomal introns* of nuclear genomes, requiring the large snRNP complex of the

spliceosome to excise. These introns are virtually synonymous with those of nuclear structural genes; they are the majority intron class. They share with the organellar/prokaryotic group II class both the 2',5'-lariat excision pathway and similar consensus motifs at the intron ends, typically GU-AG in spliceosomal introns and GU-APy in group II (Michel and Dujon, 1983). (There is also a minority spliceosomal variant class of introns commonly ending in AT-AC: see e.g. Russell et al., 2006 and references therein.) The elaborate secondary structure of the group II intron shows some notable resemblances to the complex of a spliceosomal intron with its spliceosomal snRNAs. It has therefore been speculated that the group II introns gave rise to the spliceosomal ones by 'deconstruction' into core intron plus separate spliceosomal snRNAs (reviewed by Michel and Ferat, 1995, and Bonen and Vogel, 2001). It is important to note however that there is an alternative possibility: spliceosomal introns came first and gave rise to group II ones by *accretion* of elements like spliceosomal snRNA motifs into the intron. In this way, group II introns could have survived in prokaryotes even while spliceosomal introns and the spliceosome itself were being lost there. No spliceosomal introns are currently known in prokaryotes; the latter do have sequence relatives of the Sm proteins, key components of the spliceosome, but it seems that these prokaryotic relatives have roles other than splicing (Sun et al., 2002). Also, sequence homologues of spliceosomal snRNA U6 are known in mycoplasmas, but their function is evidently unknown, and they may have arisen through lateral transfer from eukaryotes (Ushida and Muto, 1993; Ushida et al., 2003).

(3) *tRNA-type*, a family of small introns which are commonly found in tRNA genes of archaeobacteria and of eukaryotic nuclei. (The tRNA introns found in eubacteria and

organelles are of other groups, as mentioned above: group I in split eubacterial tRNAs, Vepritskiy et al., 2002, and group II in split chloroplast tRNAs, Bonen and Vogel, 2001.) The tRNA-type introns often lie in a common position in both archaeobacteria and eukaryotes, after the base following the anticodon, with some exceptions (Marck and Grosjean, 2002). Also the splicing machinery for these introns shows notable similarities between archaea and eukaryotes (Lykke-Andersen et al., 1997; Tocchini-Valentini et al., 2005). In archaeobacteria the 'tRNA-type' introns can also sometimes be found in other gene classes, in rRNA loci (Jackson et al., 2002) and in structural genes (Watanabe et al., 2002).

One notes that tRNA genes are unusual among the loci for smaller RNA in commonly being split; genes for other smaller RNAs like 5S RNA, 7SL RNA, snRNAs and snoRNAs are typically unsplit, with rare exceptions (see for example Takahashi et al., 1996). One also notes that some introns apparently do not fit readily into any of the three major groups – group I, the spliceosomal and group II lariats, tRNA-type - discussed above (Henze et al., 1995; Kawahara et al., 1998; Tan et al., 2005). But such cases are a minority.

Intron functions

Spliceosomal intron sequences, aside from the conserved motifs needed for splicing, typically display rapid divergence compared with the exons. This might suggest that these intron sequences are merely 'junk'. However, some functional roles have been

associated with them. A familiar example is the role of spliceosomal introns in facilitating new gene creation by exon shuffling; however, this is presumably a long-term consequence rather than an immediate cause of their presence (Doolittle, 1978). Other well-known examples of intron-encoded functions are the presence in some spliceosomal introns of transcriptional enhancers, and the involvement of certain of these introns in alternative splicing. Spliceosomal introns can also encode small RNAs like microRNAs involved in post-transcriptional control (Ying & Lin, 2004). Such introns can be involved in nonsense-mediated decay of aberrant transcripts (Zhang et al., 1998). Spliceosomal introns appear to be quite important in assisting 3' maturation of hnRNA to mRNA (Nesic and Maquat, 1994; Antoniou et al., 1998). Such roles make some sense of the presence of these introns. However, some of these roles, alternative splicing or transcriptional enhancement for example, are only sporadically associated with the introns. And even with an apparently more general role like 3' hnRNA processing, one would still like to know why nature settled upon such an odd arrangement, the interspersions of often-numerous introns throughout a coding region, to achieve such ends.

As for functional roles for non-spliceosomal introns: the tRNA-type introns can be involved in base-modification of the cloverleaf (Jian et al., 1997; Clouet d'Orval et al., 2001). Introns of various groups can contain ORFs for homing endonucleases which maintain the intron in the population, and for maturases which aid excision of the introns. These homing and maturase functions mostly seem to benefit the intron itself rather than the host cell. Such ORFs are largely confined to organelles and prokaryotes, where they can be found in introns of groups I and II, and also in some archaeal 'tRNA-type' introns

located atypically in the rRNA genes. In nuclear genomes such intronic ORFs are rare, known only in certain nuclear group I rRNA introns. The majority spliceosomal intron class appears to lack such ORFs (reviewed by Chevalier and Stoddard, 2001).

In conclusion: introns can have functional roles. But it is not clear that these roles provide a fully satisfying explanation for the existence of the split-gene state.

Intron origins

There are two schools of thought concerning intron origins. They could be late insertions (Crick, 1979; Stoltzfus et al., 1994; Logsdon, 1998, 2004; Cousineau et al., 2000; Qiu et al., 2004). Or they could be relics of an ancient split-gene state, all but lost in prokaryotes due to genomic streamlining, intron shedding to facilitate genome compaction for rapid replication (Darnell, 1978; Doolittle, 1978; Elder, 1987, 2000; Senapathy, 1988; Cerff, 1995; de Souza et al., 1996). The two models are not mutually exclusive (Roy and Gilbert, 2006). The present treatment will accept something of both.

Intron insertion by homing endonucleases. Insertion events are well documented for those introns which encode the aforementioned homing endonucleases, capable of inserting introns at specific sites in DNA. As their name implies, these endonucleases primarily home the intron back to intron-minus alleles of the locus; but they can sometimes place an intron copy at new positions. Could such events be central to intron origins in general? This seems questionable since as noted previously, endonuclease

ORFs are absent to date from the majority intron class, the nuclear spliceosomal group (see review by Chevalier and Stoddard, 2001).

Origins of spliceosomal introns. There is evidence for spliceosomal intron insertion into nuclear structural genes by various methods (reviewed by Roy and Gilbert, 2006). For example, in the pufferfish some introns appear from close phylogenetic analysis to be insertions; and there are hints that these may have arisen by duplication of exonic sequences containing a motif resembling the consensus sequence for splice junctions, canonically AG|GU or MAG|R where M = A,C (Venkatesh et al., 1999; see also Rogers, 1989). Another possible explanation for new spliceosomal introns is grossly ectopic back-splicing of an intron, followed by retrotranscription via some helper function, then chromosomal incorporation of the DNA copy. Fungal genes for spliceosomal snRNAs appear to have acquired spliceosomal introns in this manner (Takahashi et al., 1996); and structural genes could presumably acquire new ones in a similar way. Whatever the mechanism(s) of intron gain, the vital factor for the intron-origin debate is of course the rate of gain – is it enough to justify introns-late? Evidence has been claimed for significant levels of spliceosomal intron gain in nematodes (Coghlan and Wolf, 2004), in drosophilids (Tarrío et al., 2003), and in sundry other taxa, as for example with the triose-phosphate isomerase (TPI) locus (Logsdon et al., 1995). Such introns are putatively inserted because they are phylogenetically isolated. But some caveats apply here. In the nematode many of the putatively new, apparently phylogenetically isolated introns proved on further study to be not so phylogenetically isolated after all; they do not suggest common intron gain but a marked preponderance of loss over gain in this

organism (Roy and Penny, 2006; for other similar cases see Schmitt and Brower, 2001, and Krzywinski and Besansky, 2002). Several of the phylogenetically isolated introns in the TPI gene could easily be explained by small slippages; these could occur by compensatory mutations at each end of an intron (Brenner and Corrochano, 1996) or by slightly ectopic back-splicing followed by retronic events (Martinez et al., 1989). Finally, one of the best candidates for a spliceosomal intron insertion, in the xanthine dehydrogenase gene of drosophilids, may actually be a reinsertion into a site which may initially have possessed an intron, may subsequently have lost it, but retained flanking sequences ideal for splicing and hence also for intron reacquisition - a compromise between introns-early and introns-late (Tarrío et al., 2003). However, one need not doubt that some spliceosomal intron addition occurs.

Now it has been argued that such insertions constitute the general origin of spliceosomal introns, because they seemed to be elusive in the oldest forms of life, the prokaryotes and the most ancient eukaryotic protists, and to become increasingly common as one passes through the protists to invertebrate metazoans and on to vertebrates (Logsdon, 1998). However, it is now known that the earliest eukaryote lineages do have some spliceosomal introns (Archibald et al., 2002; Nixon et al., 2002; Simpson et al., 2002; Vanacova et al., 2005; Russell et al., 2005; Slamovits and Keeling, 2006). And some loci from older eukaryotic protistan lineages are well split, and appear to have long been so (Jean et al., 2001; Archibald et al., 2002; Slamovits and Keeling, 2006). The absence of spliceosomal introns in prokaryotes, and their rarity in early-arising protists, may reflect genomic streamlining, the heavy purging of introns, which is expected to be most in evidence in

these typically small and fast-replicating forms. Such genomic streamlining is thought to have occurred by retronic processes in a number of intron-poor protists; their few introns are typically concentrated towards the 5' ends of genes, a canonical footprint of retronic intron removal (Mourier and Jeffares, 2003; see also Vanacova et al., 2005). Similarly for the apparent rise in intron numbers from fly or worm to vertebrate (Logsdon, 1998); this again could reflect genomic streamlining, in the two compact-genomed invertebrates. Both have much higher rates of intron loss than vertebrates (Banyai and Patthy, 2004; Cho et al., 2004). Also supporting this interpretation is the presence of high levels of intron-exon conservation with respect to vertebrates in sponges (Muller et al., 2002; see also Exposito et al., 1993), coelenterates (Sullivan et al., 2006) and annelids (Raible et al., 2005). And although the pufferfish as noted above shows some intron insertions (Venkatesh et al., 1999), overall its intron complement is very similar to that of humans despite some 430 million years of phylogenetic separation (Elgar et al., 1996). So in at least some lineages, insertion is evidently rare. And even in other lineages with higher estimated insertion rates, these rates still do not seem high enough to convincingly sustain an introns-late scenario (Roy and Gilbert, 2005b).

Thus although spliceosomal intron addition can occur, there does not appear to be any compelling evidence that insertion is the dominant factor in the evolutionary dynamics of these introns. And there is evidence for the contrary view that they are substantially ancestral. Firstly, spliceosomal introns commonly show positional conservation, or near-conservation plausibly attributable to slippage, between widely separated lineages such as plant vs. metazoan or early protist vs. higher eukaryote. On this basis Roy and Gilbert

(2005a) estimate that (at least) some 40% of introns of plants, animals and fungi were already present in early eukaryotic protists. Such intron conservation patterns might be attributed by introns-late advocates to parallel selective insertion at sites providing the ideal intron-flanking splice consensus. However, substantial intron positional conservation can be observed in cases where there is little conservation of flanking sequences, as with the aspartyl protease gene family (Jean et al., 2001), or the split genes of those early protists the parabasalids (Vanacova et al., 2005). These observations suggest that early eukaryotic genomes were well split. And there are indications that spliceosomal introns may be even older than this. There is a pair of GAPDH loci deriving from a duplication which occurred back in prokaryotic times; these paralogues show a number of intron matches and near-matches. Again this is not easy to explain by parallel selective insertion at sites with the ideal flanking sequences for splicing; for the intron-flanking sequences in question show only modest matching to any ideal consensus (Cerff, 1995). This suggests that spliceosomal introns may be of prokaryotic vintage.

Similar conclusions arise from consideration of genes with internal sequence repetitions. Introns often tend to map in phase, or nearly so (putative slippage), with respect to such repeats, as expected on the introns-early model (Traut, 1988). In a class of collagen genes found from sponges to vertebrates, the introns are periodic at a sixfold multiplet of the basic Gly-X-Pro repeat, or related spacings easily derivable from this – a pattern particularly awkward for insertion models as there is nothing visible in the DNA sequence of the locus to direct intron insertions to every sixth repeat (Wozney et al., 1981; Exposito et al., 1993). For some genes with internal repeats and introns commonly

in or nearly in phase with them, the internal repetitions are of prokaryotic antiquity. This is so for loci encoding several glycolytic enzymes (Elder, 2000) and the transcription initiation factors TATA-binding protein and TFIIB/TFIIIB (unpublished observations). It has also been suggested for the carbamoyl-phosphate synthetase gene family (van den Hoff et al., 1995). Again this suggests that spliceosomal introns could go back to the prokaryotic era.

This conclusion enjoys further supports. Introns in translational phase 0 are unique in being considerably over-represented in structural genes as a class, about 50% vs. one-third expected on a random basis. This could perhaps be a relic of primitive exon-sized minigenes necessarily in phase zero for an open reading frame (Traut, 1988). Introns in phase 0 are also distinctive in showing a correlation with the boundaries of compact modules in ancient genes, possible counterparts of early exon-sized minigenes and miniproteins (de Souza et al., 1998). Such compact modules in ancient proteins most commonly fall in the size range 15-30 residues (de Souza et al., 1996); these values are of the same order of magnitude as the average observed exon size for intron-rich higher eukaryotes at 30-40 codons (Deutsch and Long, 1999). Both figures are also of the same order as the ancestral exon sizes inferred on introns-early premises from the known intron complements of some well-studied genes, commonly around 20 residues (e.g. van den Hoff et al., 1995; Shpakovski and Lebedenko, 1999). Such figures are also of the order expected for the open reading frames of the earliest cells. Assume the earliest nucleotide sequences were generated by more or less random processes. With a genetic code comprising 64 codons of which three are terminators, the expected mean length of an

open reading frame is then of the order of $64/3$ or about 20 codons. This is an order of magnitude shorter than the typical modern protein containing several hundred amino acids; but it compares well with the sort of dimensions inferred for the first minigenes on introns-early premises (Elder, 1991; see also Senapathy, 1988). It is also easier to understand the evolution of a very complex entity like a typical protein if it was assembled in a modular fashion from ancestral miniproteins; and the introns-early idea gives a ready basis for this. In sum, one must evidently take seriously the prospect that at least a significant proportion of the spliceosomal introns are ancient. As for the related group II lariat introns of prokaryotes and organelles: they could, as noted in the introductory discussion, have derived from early spliceosomal introns by ‘accretion’ of spliceosomal snRNA motifs.

Origins of tRNA introns. The tRNA-type introns could also be substantially ancestral, on the strength of their aforementioned frequent presence at or near a common position just after the anticodon in both archaea and eukaryotes (Marck and Grosjean, 2002), and of their use of related splicing enzymes in both kingdoms (Lykke-Andersen et al., 1997; Tocchini-Valentini et al., 2005). Such an intron may thus have been ancestrally present in the anticodon loop of the common ancestor of many or even all tRNAs. Some mobility of these tRNA-type introns is also inferred, since they sometimes lie in highly non-canonical positions in the cloverleaf (Marck and Grosjean, 2002; Kawach et al., 2005). And in archaea ‘tRNA-type’ introns can as noted earlier appear sometimes in non-tRNA gene classes, for rRNA (Jackson et al., 2002) and structural genes (Watanabe et al., 2002); one may well suspect that these very unusually located cases represent insertions.

Also, if the ancestral tRNA intron was of the 'tRNA-type', one necessarily infers that those tRNA introns which are of other types, that is, of group I in eubacterial tRNAs (Vepritskiy et al., 2002) or group II in chloroplast tRNAs (Bonen and Vogel, 2001), are insertions. Indeed, the group I introns of eubacterial tRNA genes mostly have limited phylogenetic distributions consistent with late insertion (Vepritskiy et al., 2002). There is one apparent exception, the leucyl-tRNA group I intron found more widely, in both cyanobacteria and the chloroplasts derived from them (Besendahl et al., 2000). If this was an insertion, it presumably happened very early. But that is not impossible. For example: this group I leucyl-tRNA intron may have originated from the rRNA genes, if they contained group I introns at an early date, as they not infrequently do in extant organisms (see further discussion below). The group II tRNA introns of chloroplasts also appear from their phylogenetic distribution to be relatively late insertions (Manhart and Palmer, 1990). A notable sidelight here is that tRNA introns of all classes frequently reside in or near the same standard position just after the anticodon. This is largely true not just for the archaeal/eukaryotic 'tRNA-type' introns but also for the eubacterial group I tRNA introns (Vepritskiy et al., 2002) and the chloroplast group II tRNA introns (Vogel and Hess, 2001). How should this be interpreted? Perhaps the anticodon loop ancestrally had a tRNA-type intron which was lost in many tRNAs, but which left a conformation in this region of the cloverleaf which rendered it prone to intron reacquisition – for example, by ectopic back-splicing and retronic events - even with introns of other classes like groups I and II. Here is a case where parallel site-specific insertion is indeed credible. But a plausible case can still be made that the tRNA-type introns of archaeal and eukaryotic tRNAs are substantially ancestral.

Origins of rRNA introns. The rRNA introns include representatives of all classes; but by far the commonest is the group I type. These group I introns can occupy many positions in the rRNA; but a good number map to certain specific sites in diverse organisms. On the other hand, however, such positional conservation does not necessarily imply ancestral status; it may reflect site-specific insertion mechanisms, such as ectopic retrosplicing into preferred positions in the rRNA followed by retronomic genomic incorporation (reviewed by Jackson et al., 2002, and Haugen et al., 2005). Likewise the discovery of group I introns in a few eubacterial rRNA genes is not clinching evidence for their ancestral presence; as noted earlier, these introns may be lateral transfers from eukaryotic organelles (Nesbo and Doolittle, 2003). The most an introns-early advocate can say here is that the common association of group I introns with rRNA genes might perhaps hint at some ancestral connection between the two.

Intron origins: summary. Introns of each type can be mobile to greater or lesser extents. But a case can be made that structural genes were ancestrally well split by spliceosomal introns, and that a tRNA-type intron ancestrally resided in the anticodon loop of this gene class. The ancestral intron status of rRNA genes is debatable; but if they were originally split, the group I intron family currently seems the likeliest candidate. Note incidentally that this scenario would ancestrally allocate the three great intron classes to the three great split-gene classes on a temptingly neat one-to-one basis – group I introns to rRNA genes; spliceosomal lariat introns to structural genes (with the group II lariat introns presumably arising from the spliceosomal ones by accretion); and tRNA-type introns of

course to tRNA genes.

Translational streamlining: structural genes

Translational streamlining is an introns-early model which aims to provide a unified explanation for the introns in each kind of split gene (Elder, 1987). This model holds that the introns are remnants of ancient motifs which initially had functional roles, but eventually became redundant, and were then removed by splicing. For example, the introns could have derived from specific RNA-processing sites in ancient transcriptional units. Such primitive transcriptional units could have been polycistronic precursors containing short segments of RNA for small primitive proteins, or for small sections of ancient tRNA or rRNA; these short segments could have ancestrally been cut free from their precursors. Such short, cotranscribed but cleaved ancestral genes could then be amalgamated into modern genes which are larger, but split, through conversion of the intercistronic cleavage sites into introns. This is one concrete example of how translational streamlining could work.

On such a scenario, the evolution of splicing would be expedited on two grounds. Firstly, the original RNA-processing motifs would already have part of the splicing process, in the form of cleavage and/or excision steps. Secondly, the motifs involved in RNA processing commonly take the form of hairpins or related foldback structures; this brings the flanking sequences into proximity favouring their ligation as exons. Examples of such foldback structure at RNA-processing sites are seen in the self-shearing hammerhead

ribozyme of plant viroids and virusoids (Symons, 1997), and in the hairpin target sites for cleavage of the polycistronic early mRNA of phage T7 by the enzyme RNase III (Rosenberg and Kramer, 1977; Robertson et al., 1977).

One could also imagine other versions of translational streamlining; for it to occur, an obsolete motif may not need to have been an excised spacer or a cleavage site. If any sort of obsolete motif was in the form of a foldback structural module, like those abounding in tRNA and rRNA for example, then the foldback topology alone might be sufficient to encourage conversion of the motif into an intron, simply by bringing the potential flanking exon sequences into the proximities required by an evolving splicing apparatus.

In such ways, then, ancient motifs in the form of excised spacers or cleavage sites, or modular foldback elements in general, on becoming obsolete, could be marked out at the RNA level in ways that could facilitate their removal by splicing. In contrast, such obsolete sequences would not be marked out in the DNA double helix in any such way; and it could be a long wait for a deletion removing just the unwanted sequence. On this basis one could understand why so odd a process as splicing should have come about. It is likely to be easier for evolution to get rid of obsolete motifs by RNA splicing than by DNA deletion.

One can first consider how translational streamlining could work for spliceosomal introns of structural genes. On the introns-early approach, a typical transcriptional unit in a very early cell would have contained polycistronic 'operons' of small genes of about 20

codons; we could call them minigenes or protogenes. The intercistronic spacer regions between the minigenes would have become introns, while the minigenes themselves became exons (Darnell, 1978). Now internal minigene cistrons in such polycistronic transcripts would have needed some means of achieving translational initiation at their internal AUG codons. Modern prokaryotes can readily perform such internal initiation via the short Shine-Dalgarno box. But in eukaryotes, usually an mRNA is simply scanned by the ribosome for the first AUG from the 5' end, and internal AUG codons cannot be accessed. It also appears that a number of archaeal mRNAs use this 5' scanning mechanism for translation initiation (Londei, 2005). A few eukaryotic mRNAs can atypically undergo internal translational initiation, via a motif called the IRES; it is much larger and more complex than the bacterial Shine-Dalgarno box (Le and Maizel, 1997). In any case, if early cells had polycistronic minigene operons, the intercistronic regions between the minigenes are expected to have contained motifs promoting internal translational initiation in some way; and such motifs might then have become translationally streamlined into introns.

One obvious example: if the ancient minigene operons had IRES elements or some equivalent for internal translation initiation, these elements could have been translationally streamlined into spliceosomal introns. However, I know of no further evidence for this particular scheme. But there is another attractive version of this approach. The primitive minigene polycistronic mRNAs may initially have been cut down to monocistronic ones, each of which could be freely translated by the ribosome 5' scanning mechanism. That is, the intercistronic regions between the minigenes were

originally excised spacers or cleavage sites. These would then have become translationally streamlined into spliceosomal introns (Elder, 1987). Some empirical precedent for this proposal can now be advanced. There are certain eukaryotes which unusually have polycistronic operons. This is well documented in trypanosomes and nematodes. These eukaryote operons are superficially similar to classical prokaryotic ones; however, there is a fundamental difference. The eukaryotic operonic polycistron mRNAs are converted to monocistrons by cleavage of intercistronic spacer sequences. The 5' ends of the freed monocistron mRNAs undergo addition of leader RNAs, encoded elsewhere in the genome. One notes in passing that monocistronically transcribed mRNAs can also acquire such leader RNAs, in the nematode and elsewhere. Returning to the polycistronic cases: the polycistron-cleaving 5' leader addition occurs by a process of trans-splicing; and this trans-splicing process uses components clearly related to the spliceosomal ones used in normal (cis) splicing (reviewed by Blumenthal, 1995).

Of course, the internal cistrons in these eukaryotic operonic mRNAs encode full-sized modern-type proteins, not the minigene ORFs postulated for the earliest cells. Nevertheless, this sort of trans-splicing system is notably like what was postulated for ancient cells. The strongest version of translational streamlining here would be to speculate that the trans-splicing process actually existed for intercistronic cleavage in minigene operons in primitive cells, and then went over into conventional intronic cis-splicing. The common presence of trans-splicing of operonic mRNAs in organisms like trypanosomes and nematodes could then represent a partial reversion towards an ancestral state. Of course, a devil's advocate could argue instead that trans-splicing in

such organisms is purely a de novo derivation from normal cis-splicing. Nonetheless, the existence of these trans-splicing operon systems provides a significant empirical precedent for the idea that spliceosomal introns could have originally been sites for processive cleavage between polycistronic minigenes (Elder, 1987).

Further to this: if the spliceosomal introns were indeed originally excised spacers or cleavage sites between minigene mRNAs, these excised-spacer/cleavage sites could have formerly assisted in polyadenylation of the 3' ends of each upstream minigene mRNA as it was processed free as a monocistron. Poly A addition is a very ancient process, common to both eukaryotes and prokaryotes (Sarkar, 1997). It might well have also been needed by minigene mRNAs freed from primitive polycistronic transcription units; and as the polyadenylation sites would have been in proximity to the excised-spacer/cleavage sites, some involvement of the latter in the poly A addition processes would not be a surprise. Then if the excised-spacer/cleavage sites were translationally streamlined into introns, it would also not be surprising if the introns continued to be utilised to aid polyadenylation down at the 3' end of contemporary exon-spliced mRNA. This is a kind of role which could be carried over from intercistronic excised-spacer/cleavage sites to introns derived from them. Indeed, as noted earlier, the spliceosomal introns have been implicated in 3' maturation of hnRNA to mRNA (Nesic and Maquat, 1994; Antoniou et al., 1998).

Translational streamlining: transfer RNA genes

Introns in tRNA could often also be dealt with by translational streamlining. Evidence was presented to suggest that a tRNA-type intron was present in the anticodon loop in tRNA genes from very early times (see earlier discussion). Again the intron could have once been an excised spacer or cleavage site. That is, the two exons in a typical split tRNA gene could originally have been two half-cloverleaves, held together only by base-pairing in the mature functioning tRNA. The 5' half-cloverleaf would have had the anticodon near its distal end, conveniently free to move during the dynamic process of protein synthesis. Note also that the two half-cloverleaves could easily have arisen as inverted duplications of one another; one can confirm that this could generate the basic form of the cloverleaf structure at a single stroke. And the junction between the inverted repeats would be in the anticodon loop, where the tRNA-type intron typically lies. (Di Giulio (1992) has discussed various earlier ideas along similar lines, including a version of his own with direct rather than inverted repeats.) In any case: assuming a primitive tRNA gene with co-transcribed half-cloverleaves flanking an excised spacer or cleavage site; it should not be unduly difficult to translationally streamline the excised-spacer/cleavage motif into a tRNA-type intron. An excision/cleavage step would already be operative. And the two half-cloverleaves would have been base-paired together in a conformation favouring their ligation as exons by the appearance of splicing.

The excised-spacer or cleavage site scenarios are not the only possible prior roles for a motif ancestral to the tRNA-type intron; diverse other possibilities could be imagined. One example is base-modification of the tRNA, a role which can be associated with tRNA-type introns as previously noted (Jian et al., 1997; Clouet d'Orval et al., 2001).

Other possibilities for an archaic role of the pre-intron motif could be binding to the ribosome; or binding to the aminoacyl-tRNA synthetases; or being itself a ribozymal aminoacyl-tRNA synthetase (as speculated by Ho, 1988, and Szathmary, 1993, though they assumed that tRNA genes were originally split by a group I intron rather than a tRNA-type one). It is also conceivable that a motif ancestral to the tRNA-type intron initially had more than one role, serving for example as an excised spacer or a cleavage site, but also having a further role such as tRNA base-modification. Such a second role might be retained through the translational streamlining of an excised-spacer/cleavage site into an intron.

Translational streamlining: ribosomal RNA genes

As for the rRNA genes, we saw above that it is debatable whether they originally had introns, though they do have an intron class, group I, that is often associated with them today at least. If early rRNA genes did contain introns of some sort, these could for example have been translationally streamlined from excised spacers between primitive shorter segments of rRNA in a polycistronic transcription unit. Relevant precedents are the freeing of the 5S RNA from the polycistronic rRNA precursor in eubacteria, and the additional fragmentation of the larger rRNAs in some organisms (Clark, 1987). Again, alternative or additional roles to excised-spacers/cleavage sites could be entertained for motifs ancestral to the introns; for example, involvement in base-modification of the mature rRNAs.

In any event, the rRNA genes are of great interest in relation to translational streamlining in another way. In archaeobacteria the process of liberation of the rRNAs from their precursor exhibits some striking similarities to intron splicing. The excised leader and spacer sequences flanking the pre-16S rRNA become ligated together; the resultant molecule is thought to function in base-modification operations (Tang et al., 2002). These ligated leader and spacer sequences are rather akin to exons. In fact, this is almost translational streamlining; but not quite. Translational streamlining is expected to convert a spacer into an intron, rather than a quasi-exon as seen here. The discrepancy is explained by the particular geometry of the cleavage sites in the rRNA precursor – foldback pairing between the ends of each pre-rRNA, not foldback between the ends of a leader/spacer. This geometry favours this transformation of the leader/spacer into an ‘exon’ rather than an intron. Still, here a system for processive cleaving of a precursor RNA at a foldback motif has evolved into something very close to classical splicing, by a pathway very similar to that predicted by translational streamlining.

This paper is dedicated to the memory of my close colleague and friend the late Dr R. J. Harris for his invaluable assistance and support.

REFERENCES

Antoniou, M., Geraghty, F., Hurst, J., Grosveld, F., 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.

Archibald, J. M., O'Kelly, C. J., Doolittle, W. F., 2002. The chaperonin genes of jakobid and jakobid-like flagellates: implications for eukaryotic evolution. *Mol. Biol. Evol.* **19**, 422-431.

Banyai, L., Patthy, L., 2004. Evidence that human genes of modular proteins have retained significantly more ancestral introns than their fly or worm orthologues. *FEBS Lett.* **565**, 127-132.

Besendahl, A., Qiu, Y. L., Lee, J., Palmer, J. D., Bhattacharya, D., 2000. The cyanobacterial origin and vertical transmission of the plastid tRNA(Leu) group-I intron. *Curr. Genet.* **37**, 12-23.

Blumenthal, T., 1995. Trans-splicing and polycistronic transcription in *Caenorhabditis elegans*. *Trends Genet.* **11**, 132-136.

Bonen, L., Vogel, J., 2001. The ins and outs of group II introns. *Trends Genet.* **17**, 322-331.

Braun, V., Mehlig, M., Moos, M., Rupnik, M., Kalt, B., Mahony, D. E., von Eichel-Streiber, C., 2000. A chimeric ribozyme in *Clostridium difficile* combines features of group I introns and insertion elements. *Mol Microbiol.* **36**, 1447-1459.

Brenner, S., Corrochano, L. M., 1996. Translocation events in the evolution of aminoacyl-tRNA synthetases. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8485-8489.

Cerff, R., 1995. The chimeric nature of nuclear genomes and the antiquity of introns as demonstrated by the GAPDH gene system, in: Go, M., Schimmel, P. (Eds.), *Tracing Biological Evolution In Protein And Gene Structures*, Elsevier Science, pp. 205-227.

Chevalier, B. S., Stoddard, B. L., 2001. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res.* **29**, 3757-3774.

Cho, S., Jin, S. W., Cohen, A., Ellis, R. E., 2004. A phylogeny of *Caenorhabditis* reveals frequent loss of introns during nematode evolution. *Genome Res.* **14**, 1207-1220.

Clark, C. G., 1987. On the evolution of ribosomal RNA. *J. Mol. Evol.* **25**, 343-350.

Clouet d'Orval, B., Bortolin, M. L., Gaspin, C., Bachellerie, J. P., 2001. Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNA^{Trp} intron guides the formation of two ribose-methylated nucleosides in the mature tRNA^{Trp}. *Nucleic Acids*

Res. **29**, 4518-4529.

Coghlan, A, Wolfe, K. H., 2004. Origins of recently gained introns in *Caenorhabditis*. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 11362-11367.

Cousineau, B., Lawrence, S., Smith, D., Belfort, M., 2000. Retrotransposition of a bacterial group II intron. *Nature* **404**, 1018-1021.

Crick, F., 1979. Split genes and RNA splicing. *Science* **204**, 264-271.

Darnell, J. E. Jr., 1978. Implications of RNA-RNA splicing in evolution of eukaryotic cells. *Science* **202**, 1257-1260.

De Souza, S. J., Long, M., Schoenbach, L., Roy, S. W., Gilbert, W., 1996. Intron positions correlate with module boundaries in ancient proteins. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 14632-14636.

De Souza, S. J., Long, M., Klein, R. J., Roy, S., Lin, S., Gilbert, W., 1998. Toward a resolution of the introns early/late debate: only phase zero introns are correlated with the structure of ancient proteins. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5094-5099.

Deutsch, M., Long, M., 1999. Intron-exon structures of eukaryotic model organisms. *Nucleic Acids Res.* **27**, 3219-3228.

Di Giulio, M., 1992. On the origin of the transfer RNA molecule. *J. Theor. Biol.* **159**, 199-214.

Doolittle, W. F., 1978. Genes in pieces - were they ever together? *Nature* **272**, 581-582.

Elder, D., 1987. Split genes, introns and translational streamlining. *Rivista di Biologia-Biology Forum* **80**, 499-511.

Elder, D., 1991. Evolution of split genes. *J. theor. Biol.* **152**, 427-428.

Elder, D., 2000. Split gene origin and periodic introns. *J. theor. Biol.* **207**, 455-472.

Elgar, G., Sandford, R., Aparicio, S., Macrae, A., Venkatesh, B., Brenner, S., 1996. Small is beautiful: comparative genomics with the pufferfish (*Fugu rubripes*). *Trends Genet.* **12**, 145-150.

Exposito, J. Y., van der Rest, M., Garrone, R., 1993. The complete intron/exon structure of *Ephydatia mulleri* fibrillar collagen gene suggests a mechanism for the evolution of an ancestral gene module. *J. Mol. Evol.* **37**, 254-259.

Haugen, P., Simon, D. M., Bhattacharya, D., 2005. The natural history of group I introns. *Trends Genet.* **21**, 111-119.

Henze, K., Badr, A., Wettern, M., Cerff, R., Martin, W., 1995. A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 9122-9126.

Ho, C. K., 1988. Primitive ancestry of transfer RNA. *Nature* **333**, 24.

Jackson, S., Cannone, J., Lee, J., Gutell, R., Woodson S., 2002. Distribution of rRNA introns in the three-dimensional structure of the ribosome. *J. Mol. Biol.* **323**, 35-52.

Jean, L., Long, M., Young, J., Pery, P., Tomley, F., 2001. Aspartyl proteinase genes from apicomplexan parasites: evidence for evolution of the gene structure. *Trends Parasitol.* **17**, 491-498.

Jiang, H. Q., Motorin, Y., Jin, Y. X., Grosjean, H., 1997. Pleiotropic effects of intron removal on base modification pattern of yeast tRNA^{Phe}: an in vitro study. *Nucleic Acids Res.* **25**, 2694-2701.

Kawach, O., Voss, C., Wolff, J., Hadfi, K., Maier, U. G., Zauner, S. 2005. Unique tRNA introns of an enslaved algal cell. *Mol. Biol. Evol.* **22**, 1694-1701.

Kawahara, T., Yanagi, H., Yura, T., Mori, K., 1998. Unconventional splicing of HAC1/ERN4 mRNA required for the unfolded protein response. Sequence-specific and

non-sequential cleavage of the splice sites. *J. Biol. Chem.* **273**, 1802-1807.

Ko, M., Choi, H., Park, C., 2002. Group I self-splicing intron in the *recA* gene of *Bacillus anthracis*. *J. Bacteriol.* **184**, 3917-3922.

Krzywinski, J., Besansky N. J., 2002. Frequent intron loss in the *white* gene: a cautionary tale for phylogeneticists. *Mol. Biol. Evol.* **19**, 362-366.

Landthaler, M., Shub, D. A., 1999. Unexpected abundance of self-splicing introns in the genome of bacteriophage Twort: introns in multiple genes, a single gene with three introns, and exon skipping by group I ribozymes. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 7005-7010.

Le, S. Y., Maizel, J. V. Jr., 1997. A common RNA structural motif involved in the internal initiation of translation of cellular mRNAs. *Nucleic Acids Res.* **25**, 362-369.

Logsdon, J. M. Jr., 1998. The recent origins of spliceosomal introns revisited. *Curr. Opin. Genet. Dev.* **8**, 637-648.

Logsdon, J. M. Jr., 2004. Worm genomes hold the smoking guns of intron gain. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 11195-11196.

Logsdon, J. M. Jr, Tyshenko, M. G., Dixon, C., D-Jafari, J., Walker, V. K., Palmer, J. D.,

1995. Seven newly discovered intron positions in the triose-phosphate isomerase gene: evidence for the introns-late theory. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 8507-8511.

Londei, P., 2005. Evolution of translational initiation: new insights from the archaea. *FEMS Microbiol. Rev.* **29**, 185-200.

Lykke-Andersen, J., Aagaard, C., Semionenkov, M., Garrett, R. A., 1997. Archaeal introns: splicing, intercellular mobility and evolution. *Trends Biochem. Sci.* **22**, 326-331.

Manhart, J. R., Palmer, J. D., 1990. The gain of two chloroplast tRNA introns marks the green algal ancestors of land plants. *Nature* **345**, 268-270.

Marck, C., Grosjean, H., 2002. tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA* **8**, 1189-1232.

Martinez, P., Martin, W., Cerff, R., 1989. Structure, evolution and anaerobic regulation of a nuclear gene encoding cytosolic glyceraldehyde-3-phosphate dehydrogenase from maize. *J. Mol. Biol.* **208**, 551-565.

Michel, F., Dujon, B., 1983. Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast- and nuclear-encoded members. *EMBO J.* **2**, 33-38.

Michel, F., Ferat, J. L., 1995. Structure and activities of group II introns. *Annu. Rev. Biochem.* **64**, 435-461.

Mourier, T, Jeffares, D. C., 2003. Eukaryotic intron loss. *Science* **300**, 1393.

Muller, W. E., Bohm, M., Grebenjuk, V. A., Skorokhod, A., Muller, I. M., Gamulin, V., 2002. Conservation of the positions of metazoan introns from sponges to humans. *Gene* **295**, 299-309.

Nesbo, C. L., Doolittle, W. F., 2003. Active self-splicing group I introns in 23S rRNA genes of hyperthermophilic bacteria, derived from introns in eukaryotic organelles. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 10806-10811.

Nesic, D., Maquat, L. E., 1994. Upstream introns influence the efficiency of final intron removal and RNA 3'-end formation. *Genes Dev.* **8**, 363-375.

Nixon, J. E., Wang, A., Morrison, H. G., McArthur, A. G., Sogin, M. L., Loftus, B. J., Samuelson, J., 2002. A spliceosomal intron in *Giardia lamblia*. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 3701-3705.

Qiu, W. G., Schisler, N., Stoltzfus, A., 2004. The evolutionary gain of spliceosomal introns: sequence and phase preferences. *Mol. Biol. Evol.* **21**, 1252-1263.

Raible, F., Tessmar-Raible, K., Osoegawa, K., Wincker, P., Jubin, C., Balavoine, G., Ferrier, D., Benes, V., de Jong, P., Weissenbach, J., Bork, P., Arendt, D., 2005. Vertebrate-like intron-rich genes in the marine annelid *Platynereis dumerilii*. *Science* **310**, 1325-1326,

Robertson, H. D., Dickson, E., Dunn, J. J., 1977. A nucleotide sequence from a ribonuclease III processing site in bacteriophage T7 RNA. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 822-826.

Rogers, J.H., 1989. How were introns inserted into nuclear genes? *Trends Genet.* **5**, 213-216.

Rosenberg, M., Kramer, R. A., 1977. Nucleotide sequence surrounding a ribonuclease III processing site in bacteriophage T7 RNA. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 984-988.

Roy, S. W., Gilbert, W., 2005a. Complex early genes. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 1986-1991.

Roy, S. W., Gilbert, W., 2005b. Rates of intron loss and gain: implications for early eukaryotic evolution. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 5773-5778.

Roy, S. W., Gilbert, W., 2006. The evolution of spliceosomal introns: patterns, puzzles

and progress. *Nat. Rev. Genet.* **7**, 211-221,

Roy, S. W., Penny, 2006.

Russell, A. G., Shutt, T. E., Watkins, R. F., Gray, M. W., 2005. An ancient spliceosomal intron in the ribosomal protein L7a gene (Rpl7a) of *Giardia lamblia*. *BMC Evol. Biol.* **5**, 45.

Russell, A. G., Charette, J. M., Spencer, D. F., Gray, M. W., 2006. An early evolutionary origin for the minor spliceosome. *Nature* **443**, 863-866.

Sarkar, N. 1997. Polyadenylation of mRNA in prokaryotes. *Annu. Rev. Biochem.* **66**, 173-197.

Schmitt, D. M., Brower, D. L., 2001. Intron dynamics and the evolution of integrin beta-subunit genes: maintenance of an ancestral gene structure in the coral, *Acropora millepora*. *J. Mol. Evol.* **53**, 703-710.

Senapathy, P., 1988. Possible evolution of splice-junction signals in eukaryotic genes from stop codons. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 1129-1133

Shpakovski, G. V., Lebedenko, E. N., 1999. Molecular evolution and structure of eukaryotic nuclear RNA polymerase subunits in light of the exon-intron organization of

their genes. *Russian J. Bioorganic Chem.* **25**, 828-837.

Simpson, A. G., MacQuarrie, E. K., Roger, A. J., 2002. Eukaryotic evolution: early origin of canonical introns. *Nature* **419**, 270.

Slamovits, C. H., Keeling, P. J., 2006. A high density of ancient spliceosomal introns in oxymonad excavates. *BMC Evol. Biol.* **6**, 34.

Stoltzfus, A., Spencer, D. F., Zuker, M., Logsdon, J. M. Jr., Doolittle, W. F., 1994. Testing the exon theory of genes: the evidence from protein structure. *Science* **265**, 202-207.

Sullivan, J. C., Reitzel, A. M., Finnerty, J. R., 2006. A high percentage of introns in human genes were present early in animal evolution: evidence from the basal metazoan *Nematostella vectensis*. *Genome Inform.* **17**, 219-229.

Sun, X., Zhulin, I., Wartell, R. M., 2002. Predicted structure and phyletic distribution of the RNA-binding protein Hfq. *Nucleic Acids Res.* **30**, 3662-3671.

Symons, R. H., 1997. Plant pathogenic RNAs and RNA catalysis. *Nucleic Acids Res.* **25**, 2683-2689.

Szathmary, E., 1993. Coding coenzyme handles: a hypothesis for the origin of the genetic

code. *Proc. Natl. Acad. Sci. U. S. A.* **190**, 9916-9920.

Takahashi, Y., Tani, T., Ohshima, Y., 1996. Spliceosomal introns in conserved sequences of U1 and U5 small nuclear RNA genes in yeast *Rhodotorula hasegawae*. *J. Biochem.* **120**, 677-683.

Tan, K. S., Ong, G., Song, K. P., 2005. Introns in the cytolethal distending toxin gene of *Actinobacillus actinomycetemcomitans*. *J. Bacteriol.* **187**, 567-575.

Tang, T. H., Rozhdestvensky, T. S., d'Orval, B. C., Bortolin, M. L., Huber, H., Charpentier, B., Branlant, C., Bachellerie, J. P., Brosius, J., Huttenhofer, A., 2002. RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nucleic Acids Res.* **30**, 921-930.

Tarrio, R., Rodriguez-Trelles, F., Ayala, F. J., 2003. A new *Drosophila* spliceosomal intron position is common in plants. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 6580-6583.

Tocchini-Valentini, G. D., Fruscoloni, P., Tocchini-Valentini, G. P., 2005. Coevolution of tRNA intron motifs and tRNA endonuclease architecture in Archaea. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15418-15422.

Toro, N., 2003. Bacteria and Archaea Group II introns: additional mobile genetic elements in the environment. *Environ. Microbiol.* **5**, 143-151.

Traut, T. W., 1988. Do exons code for structural or functional units in proteins? *Proc. Natl. Acad. Sci. U. S. A.* **85**, 2944-2948.

Ushida, C., Muto, A., 1993. A small RNA of *Mycoplasma capricolum* that resembles eukaryotic U6 small nuclear RNA. *Nucleic Acids Res.* **21**, 2649-2653.

Ushida, C., Yoshida, A., Miyakawa, Y., Ara, Y., Muto, A., 2003. Distribution of the MCS4 RNA genes in mycoplasmas belonging to the *Mycoplasma mycoides* cluster. *Gene* **314**, 149-155.

Vanacova, S., Yan, W., Carlton, J. M., Johnson, P. J., 2005. Spliceosomal introns in the deep-branching eukaryote *Trichomonas vaginalis*. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 4430-4435.

Van den Hoff, M. J., Jonker, A., Beintema, J. J., Lamers, W. H., 1995. Evolutionary relationships of the carbamoylphosphate synthetase genes. *J. Mol. Evol.* **41**, 813-832.

Venkatesh, B., Ning, Y., Brenner, S., 1999. Late changes in spliceosomal introns define clades in vertebrate evolution. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10267-10271.

Vepritskiy, A. A., Vitol, I. A., Nierzwicki-Bauer, S. A., 2002. Novel group I intron in the tRNA(Leu)(UAA) gene of a gamma-proteobacterium isolated from a deep subsurface

environment. *J. Bacteriol.* **184**, 1481-1487.

Vogel, J., Hess, W. R., 2001. Complete 5' and 3' end maturation of group II intron-containing tRNA precursors. *RNA* **7**, 285-292.

Watanabe, Y., Yokobori, S., Inaba, T., Yamagishi, A., Oshima, T., Kawarabayasi, Y., Kikuchi, H., Kita, K., 2002. Introns in protein-coding genes in Archaea. *FEBS Lett.* **510**, 27-30.

Wozney, J., Hanahan, D., Tate, V., Boedtker, H., Doty, P., 1981. Structure of the pro alpha 2 (I) collagen gene. *Nature* **294**, 129-135.

Yamada, T., Tamura, K., Aimi, T., Songsri, P., 1994. Self-splicing group I introns in eukaryotic viruses. *Nucleic Acids Res.* **22**, 2532-2537.

Ying, S. Y., Lin, S. L., 2004. Intron-derived microRNAs - fine tuning of gene functions. *Gene* **342**, 25-28.

Zhang, J., Sun, X., Qian, Y., LaDuca, J. P., Maquat, L. E., 1998. At least one intron is required for the nonsense-mediated decay of triosephosphate isomerase mRNA: a possible link between nuclear splicing and cytoplasmic translation. *Mol. Cell. Biol.* **18**, 5272-5283.