Opinion TRENDS in Genetics Vol.22 No.1 January 2006

Full text provided by www.sciencedirect.com

The biology of intron gain and loss

Daniel C. Jeffares[1,2], Tobias Mourier[1,2] and David Penny[3]

[1]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK, CB10 1SA
[2]Department of Evolutionary Biology, Biological Institute, University of Copenhagen, Denmark
[3]Allan Wilson Center for Molecular Ecology and Evolution, Massey University, PO Box 11-222, Palmerston North, New Zealand

Intron density in eukaryote genomes varies by more than three orders of magnitude, so there must have been extensive intron gain and/or intron loss during evolution. A favored and partial explanation for this range of intron densities has been that introns have accumulated stochastically in large eukaryote genomes during their evolution from an intron-poor ancestor. However, recent studies have shown that some eukaryotes lost many introns, whereas others accumulated and/or gained many introns. In this article, we discuss the growing evidence that these differences are subject to selection acting on introns depending on the biology of the organism and the gene involved.

Intron accumulation and loss: the debate

Spliceosomal introns appear to be a universal feature of eukaryote genomes but are absent from bacterial and archaeal genomes. This discrepancy has fuelled considerable debate about the initial origin of introns and the spliceosome, and the extent to which intron-dense genomes are an ancestral or a derived state (Box 1). The debate continues and currently there is no clear consensus about the initial origin of introns. For detailed reviews, see Refs [1–4].

Because prokaryotes have no spliceosomal introns, our understanding of intron gain or loss (IGL) must come from analysis of eukaryote genomes. A large body of research into eukaryote IGL has shown that although the positions of certain introns (10–40%) are conserved between highly divergent eukaryotes, the number and placement of most introns is dynamic during evolution [5–8]. Comparative eukaryote genomics is showing the dynamic range of intron density (see Glossary) ever more clearly. The intron density of annotated eukaryotic genomes varies by more than three orders of magnitude: from ~140 000 introns in the human genome (intron density 8.4 introns per gene) to only 15 introns in the microsporidian *Encepalitozoon cuniculi* (intron density 0.0075 introns per gene) [9–11].

Early models of intron evolution were based primarily on examining introns in sets of orthologous genes from model eukaryotes, mostly metazoans and higher plants (Refs [1,2,4] and references therein). These studies often assumed that probabilities of IGL were similar among all taxa, and that patterns of IGL observed in any particular group of genes could be extrapolated across the entire genome.

We review recent genomic studies showing that rates of IGL differ significantly between eukaryotic genomes; some have undergone extensive intron loss, whereas others have accumulated and retained many introns. We also describe studies showing that intron dynamics differ between particular types of genes. We present the view that IGL is affected by selective pressures that depend on the biology of the organism and the gene concerned.

Eukaryotes differ in their tendencies to gain and lose introns

The eukaryotic tree displays a range of intron densities. An examination of intron density and eukaryote phylogeny shows that it is not always the case that early branching eukaryotes are intron poor and that late branching eukaryotes are intron rich (Figures 1 and 2). The current view from several recent studies is that differences in intron densities are due to different histories of IGL dynamics – some groups of organisms appear to have gained many introns, whereas others have lost many introns [6,11–15]. For example, a paucity of introns in some specialized eukaryotes is now being interpreted as a result of extensive intron loss (Box 2).

Rates of IGL also differ significantly between lineages. An analysis of orthologs from distantly related eukaryotes (crown group and *Plasmodium*) concluded that rates of intron gain vary tenfold between these species, whereas the rates of intron loss varied by approximately eightfold. The ratio of the rate of intron gain to the rate of intron loss varied by >20-fold [15]. An important conclusion of this study was that all species (including metazoan genomes) had more losses than gains.

Glossary

Nucleomorph: a eukaryote endosymbiont alga whose cellular structure is reduced to the point that only a membrane-bound nucleus and a chloroplast remain. Nucleomorph genomes are compact compared with those of free living eukaryotes.

5′ intron bias: an excess of introns in the 5′ regions of genes. Many unicellular eukaryotes have 5′ intron biases, multicellular genomes have no bias of intron positions.

Intron density: the average number of introns per gene over an entire genome (or set of genes).

***r*-selection:** natural selection on the basis of maximum reproductive and/or growth rate (e.g. *Escherichia coli* or baker's yeast).

K-selection: natural selection that favors organisms with slow growth and long life cycles that live in stable environments.

cDNA recombination: the process of transcription, splicing, reverse transcription by endogenous reverse transcriptases and recombination of the intron-less cDNAs with the genome. It is known to occur in many eukaryote genomes.

Ultraconserved elements: a class of conserved elements identified in metazoan genomes that share 100% identity across at least 200 bp [28].[28]

Corresponding author: Jeffares, D.C. (dcj@sanger.ac.uk).
Available online 14 November 2005

www.sciencedirect.com 0168-9525/$ - see front matter © 2005 Elsevier Ltd. All rights reserved. doi:10.1016/j.tig.2005.10.006

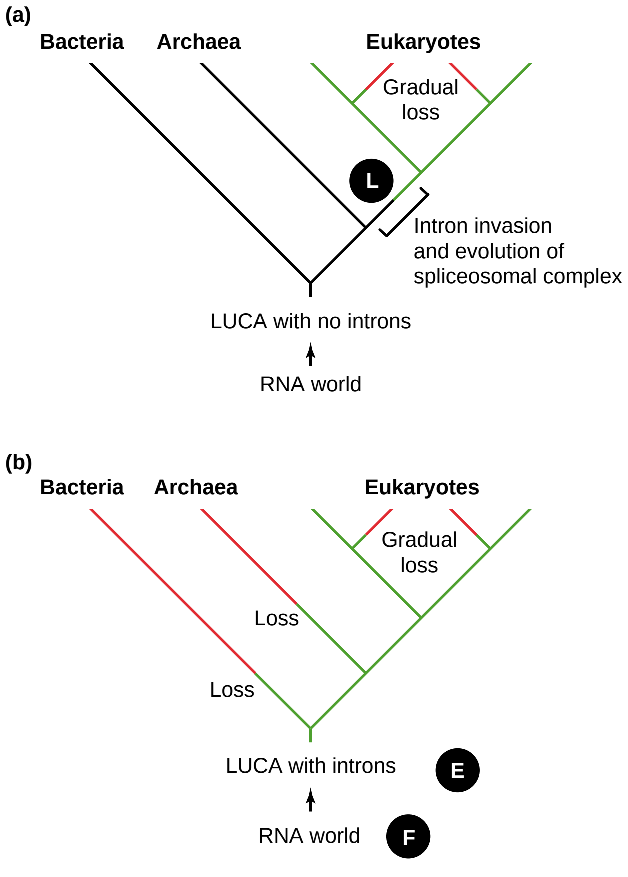
Box 1. The origin of introns

The ubiquity of spliceosomal introns in eukaryotes and their absence in prokaryotes has prompted three different theories regarding the initial origin of introns and the spliceosome.

(i) **Introns late:** according to this theory, introns and the spliceosome arose within the eukaryote lineage (depicted as 'L' in Figure Ia) and then the introns accumulated in eukaryote genomes (see Refs [1,2,4] for more details). Subsequent sources of new introns include 'reverse splicing' and insertion of transposable elements [46–48].

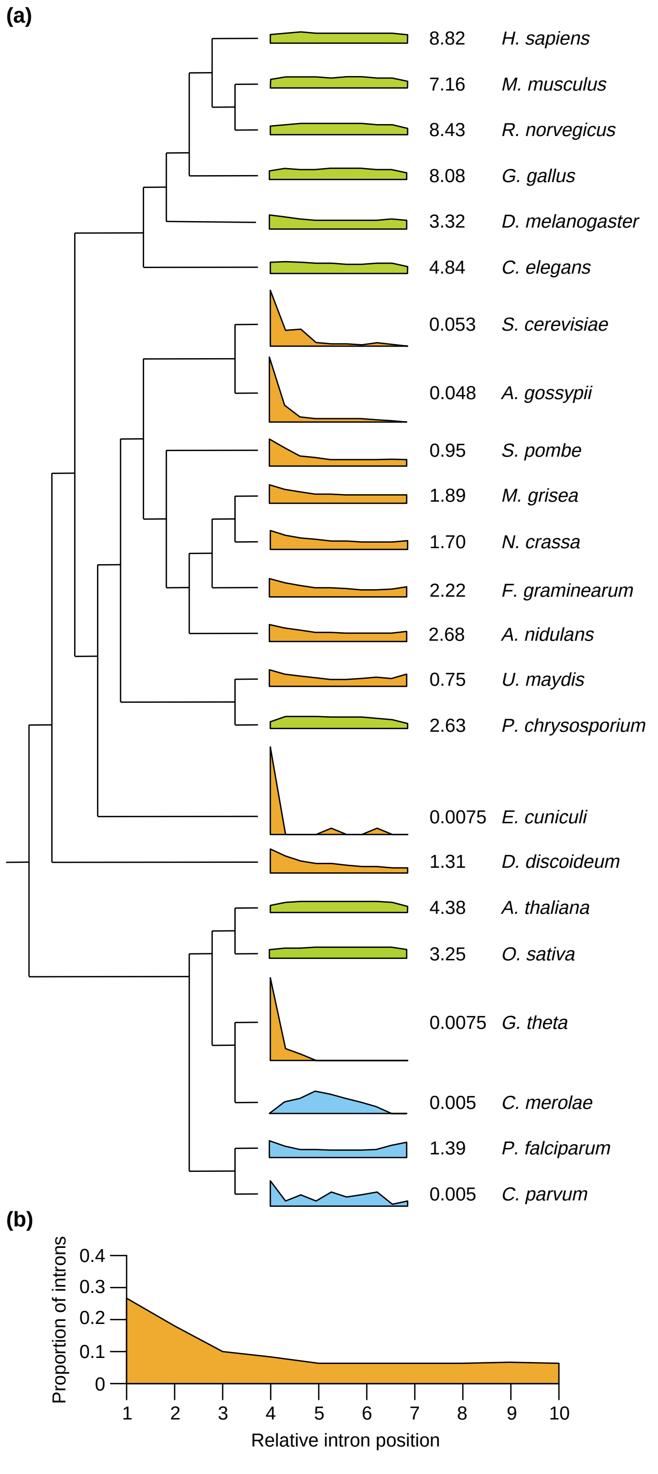
(ii) **Introns early:** this model proposes that the intron-exon structure of genes was present in the last universal common ancestor (LUCA) ('E' in Figure Ib) – and possibly earlier – where primordial protein domains were shuffled to facilitate protein evolution [49]; for a review, see Ref. [3]. Introns were subsequently lost from archaea and bacteria. The molecular mechanism for intron loss via the recombination of spliced cDNAs [50], and the antiquity of the reverse transcriptase required present a clear mechanism of loss. This model depends (as does the introns first model) only on known processes.

(iii) **Introns first:** this is similar to the introns early theory, but proposes that introns and the spliceosome are remnants from the RNA world ('F' in Figure Ib) [45,51]. This model was initiated from the observation that putatively ancient snoRNA genes are often encoded by introns. Because RNAs were the only catalysts for the assembly of an all-RNA ribosome before the advent of proteins, snoRNAs must have been used for the assembly of the proto-ribosome as it evolved towards full protein producing capacity [42]. Hence, the introns that contain snoRNAs pre-date the protein-coding exons that surround them. The splicing of snoRNA-encoding introns from transcripts with no protein coding potential, and the processing of pre-rRNA and pre-tRNAs by RNase P are examples of how RNA processing might have occured before proteins evolved [45,51].



TRENDS in Genetics

Figure I. The tree of life. The origin of introns as explained by **(a)** the introns late theory and **(b)** the introns early and introns first theories. The green branches indicate lineages containing introns, the black branches denote pre-intron stages and the red branches indicate secondary loss of introns.



TRENDS in Genetics

Figure 1. Intron density and 5′ bias in eukaryote genomes. **(a)** General eukaryotic phylogeny. The number of introns per gene in each species is shown alongside a schematic illustration of the relative intron position in each species colored according the type of intron bias [green, no bias of intron distribution; orange, 5′ biased introns; blue, other bias; a full explanation of the format is given in (b)]. The general tree topology was adapted from Ref. [55], and the topology of the fungal clade was provided by Jason Stajich (personal communication) (http://www.duke.edu/~jes12/sequenced_fungi_tree.html). The intron data were obtained from the following sources: UCSC Genome Browser (http://genome.ucsc.edu/) (*Drosophila melanogaster*, *Caenorhabditis elegans*, *Mus musculus*, *Gallus gallus*, *Rattus norvegicus* and *Homo sapiens*); Motomichi Matsuzaki provided information about *Cyanidioschyzon merolae*; Ashbya gossypii Genome Browser (http://agd.unibas.ch/) (*A. gossypii*); TIGR Rice Genome Project (http://www.tigr.org/tdb/e2k1/

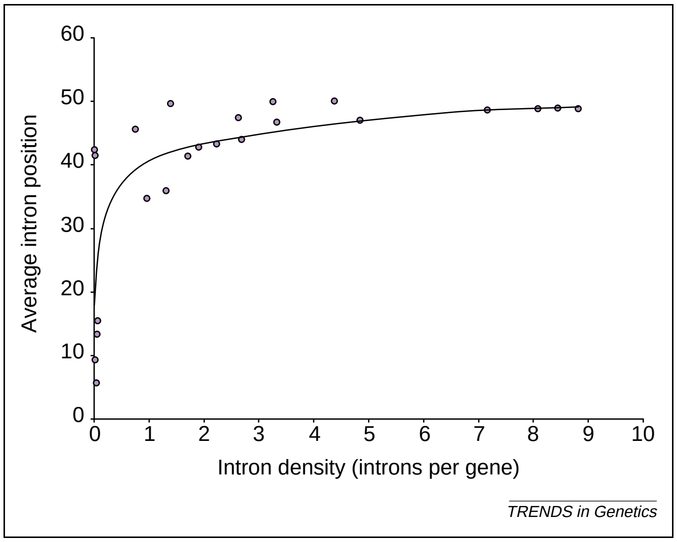


Figure 2. Intron position bias and intron density. For the 23 different genomes (shown in Figure 1a), the average intron position (for all introns in the genome) with respect to the coding sequence was plotted against the intron density. Genomes with no bias of intron position will have an average intron position at 50% of the length of the coding sequence. Those with an increasingly strong 5′ bias of intron positions have average positions approaching zero. Data and calculations are as described in Figure 1.

Box 2. Extensive intron loss

One mechanism for intron loss is by reverse transcription of spliced mRNAs (which do not contain introns), followed by homologous recombination of the cDNA with the genomic copy of the gene. Because reverse transcription proceeds in 5′–3′ direction and often terminates prematurely (see Figure 1b), this process would result in a preferential loss of introns at the 3′ end of transcripts. It has been shown that reverse transcription and cDNA recombination does occur *in vivo* [52], and several analyses of eukaryote genomes have indicated that 3′ intron loss has occurred in unicellular eukaryotes [11,14,48,53].

The microsporidian *Encephalitozoon cuniculi* (an intracellular parasite of mammals) and the *Guillardia theta* nucleomorph (an endosymbiont alga) provide clear examples of extensive intron loss at the 3′ end of transcripts. Both organisms contain <20 introns each per genome [33,35], and all introns are located at the extreme 5′ ends of genes. Because both organisms are related to relatively intron-rich groups of eukaryotes (see Figure 1a), they do not appear to have been ancestrally intron poor.

The genomes of these ‘mini-eukaryotes’ have been derived by reduction from more complex, free-living organisms; *E. cuniculi* is an amitochondrial fungus, and *G. theta* is derived from a red alga that is now reduced to a organelle-like nucleomorph [33,35]. Many features of their genomes are highly reduced or compact; they have overlapping genes, shorter coding regions than typical eukaryote homologs and their genomes contain few genes and lack some of the core eukaryote protein set [33,35,54].

The derived nature of the *E. cuniculi* and *G. theta* nucleomorph genomes, their phylogenetic placement within relatively intron-rich eukaryote groups and the association of 5′ biased intron density with intron loss provide strong evidence that these two organisms have undergone extensive intron loss.

The intestinal parasite *Giardia lamblia* also appears to be derived from a more intron-rich ancestor. Although only one intron has been found in the *Giardia* genome so far, conservative criteria have identified 27 spliceosomal proteins in this species, representing a complex spliceosome [43]. The scenario that one or a few introns have necessitated the evolution of a complex spliceosome with dozens of proteins and five small nuclear RNAs appears unlikely. It seems more likely that the complex spliceosome of *Giardia* is a molecular fossil of its intron-rich ancestry.

Metazoan genomes provide the best examples of differing IGL rates over similar evolutionary timescales. Mammalian genomes appear almost static in terms of IGL because a loss of only 0.003 introns per gene and no clear intron gains could be identified between human and mouse genomes (separated by ~90 million years) [16]. By contrast, a comparison of *Caenorhabditis elegans* and *Caenorhabditis briggsae* genomes (separated by ~100 million years) identified ~0.5 IGL events per gene [17], and a comparison of the *Drosophila* and *Anopheles* genomes (separated by 125 million years) identified approximately one IGL event per gene [17].

In the remainder of this article we attempt to explain such differences in intron dynamics. In particular, we argue that life cycle parameters are important for understanding intron evolution.

The selective advantages of intron loss

Using population genetics theory, Lynch has shown that if intron-containing alleles are slightly deleterious, they will be less tolerated in small organisms that have large populations [18]. There is an inverse relationship between the size of organism and its effective population size; therefore, single-cell organisms will usually have much larger populations than multicellular organisms, and thus will experience selection against even slightly deleterious introns.

We found that intron density correlates with the logarithm of generation time (Figure 3). That is, organisms that reproduce rapidly tend to have fewer introns than organisms that have longer life cycles. This correlation could be due to selection for smaller genomes, selection for genes that can produce proteins quickly in response to external stimuli (e.g. heat shock proteins), or both.

We expect that small, *r*-selected organisms will experience selection pressure to reduce processing times for mRNAs. RNA polymerase II proceeds at 1–1.5 kb per minute and intron excision takes approximately three minutes [19]; therefore, intron removal is a significant part of mRNA-processing time. The consideration of an extreme example illustrates how these RNA-processing times might be crucial in organisms with short generation times, and unimportant in organisms with much longer generation times. The human Usher syndrome 2A gene (*USH2A*) encodes a 790-kb transcript (including ~70

osa1/) (*Oryza sativa*); National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) [*Arabidopsis thaliana*, *Encephalitozoon cuniculi*, *Guillardia theta*, *Saccharomyces cerevisiae*, *Neurospora crassa*, *Schizosaccharomyces pombe*, *Dictyostelium discoideum* (Chromosome 2) and *Cryptosporidium parvum*]; GenBank (*Plasmodium falciparum*); The US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/) (*P. chrysosporium*); Broad Institute of MIT and Harvard (http://www.broad.mit.edu) (*Magnaporthe grisea*, *Ustilago maydis*, *Aspergillus nidulans* and *Fusarium graminearum*); Fungal Genomics Laboratory at North Carolina State University (http://www.fungalgenomics.ncsu.edu) and Broad Institute of MIT and Harvard (http://www.broad.mit.edu/) (*Magnaporthe grisea*). **(b)** A detailed relative intron-position plot for *S. pombe*. For each intron, its relative position is calculated as: the number of base pairs upstream of the intron in the coding sequence (i.e. the intronic sequence was removed) ÷ the total number of base pairs of coding sequence. Values are then pooled into ten classes. The proportion of introns (*y*-axis) that fall into each of the ten classes (*x*-axis) is shown.

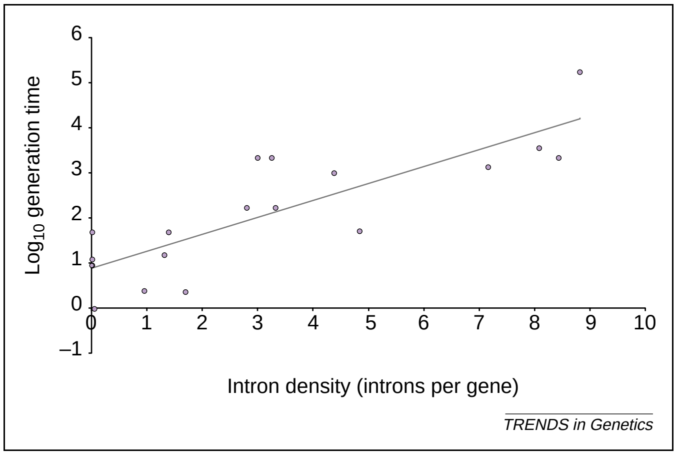


Figure 3. Intron density is correlated with generation time. Intron density (introns per gene) correlates with the log$_{10}$ of generation time across a broad range of eukaryotes including metazoans, plants, fungi and unicellular eukaryotes (Spearman Rank correlation coefficient $r_s=0.82$, $P=1.9\times10^{-5}$). This might be a result of selection for rapid cell division, rapid gene expression or a combination of both. The organisms that were included in this comparison are: *Homo sapiens* (20 y), *Rattus norvegicus* (90 d), *Gallus gallus* (150 d), *Mus musculus* (56 d), *Caenorhabditis elegans* (52 h), *Arabidopsis thaliana* (42 d), *Drosophila melanogaster* (7 d), *Neurospora crassa* (2.5 h), *Plasmodium falciparum* (48 h), *Dictyostelium discoideum* (15 h), *Schizosaccharomyces pombe* (2.5 h), *Cryptosporidium parvum* (13 h), *Encephalitozoon cuniculi* (48 h), *Cyanidioschyzon merolae* (9 h), *Oryza sativa* (90 d), *Anopheles gambiae* (ten days), *Danio rerio* (90 d), *Saccharomyces cerevisiae* (1 h), *L. major* (9 h), *C. albicans* (1.5 h), *C. reinhardtii* (7 d). The generation times in years (y), days (d) or hours (h), as appropriate, are given in parenthesis. The intron densities are the same as those in Figures 1 and 2.

introns that contribute 770 kb of its length) [20]. Although introns can enhance gene expression [21], it would take at least eight hours to transcribe this transcript. An intronless *USH2A* (which encodes an 18.8-kb transcript) could be transcribed in just 12 minutes. This difference might not be important in human biology, but would be important in an organism such as *Saccharomyces cerevisiae*, which has a generation time of approximately one hour.

The unicellular eukaryote genomes that are fully sequenced are biased towards ultra-small, highly reduced (or compact) genomes, parasites and hemiascomycetous yeasts. Analysis of these genomes shows that extreme intron paucity can be associated with compact, reduced genomes. But this is not always the case; for example, the parasites *Cryptosporidium parvum* and *Leishmania major* and the alga *Cyanidioschyzon merolae* are also intron poor (Figure 1a), but their genomes are not otherwise reduced compared with those of the more typical unicellular eukaryotes (Table 1). More sequence data from additional unicellular eukaryotes are required to assess the generality of intron loss in unicellular eukaryotes. There is at least one free-living unicellular eukaryote, a Jakobid flagellate, that appears to have an intron density comparable to that of vertebrates (approximately eight introns per gene) [22].

Intron gain or retention in complex genomes

Many unicellular eukaryotes (particularly parasites) appear to be under pressure to lose introns, whereas multicellular eukaryotes have more intron-dense genomes. This appears to be due to intron accumulation [5,6] and/or intron retention [6,13,23].

New introns can accumulate or be retained in multicellular eukaryotes because organisms that have small populations will have less stringent selection against mildly deleterious elements, such as functionless introns, in their genomes compared with organisms that have large populations [18]. Although this factor might account for the maintenance of any recently gained functionless introns, there are many indications that a proportion of introns are essential functional components of eukaryote genomes, rather than being mildly deleterious.

Introns are required for alternative splicing. At least 40% of genes in both multicellular plants and animals are alternatively spliced [24], and possibly as many as 70% in the human genome [25]. Up to 60% of alternative splice variants are conserved between mouse and human [26], suggesting that many of these splice forms are necessary components of the genomic toolkit.

Introns also encode a variety of untranslated RNAs including microRNAs, small nucleolar RNAs (snoRNAs) and guide RNAs for RNA editing [27]. The exact number of such transcripts is unknown, but they might be as abundant as mRNAs in some genomes. For example, a recent study identified >5000 novel intronic transcripts on human chromosomes 21 and 22 (which contain 770 annotated genes) [25]. The functions of most of these transcripts have not been investigated, and a proportion of these might result from non-functional aberrant transcription. Even if only 10% of these transcripts were functional, there would be 500 intronic transcripts produced from just 770 genes.

Some introns are known to enhance or be necessary for normal levels of mRNA transcription, processing and transport [21], and most of these are found in metazoans

Table 1. Intron paucity is not always associated with compact genomes

Species[a]	Intron density (introns per gene)[b]	Number of genes	Genome size (Mb)	Genes per Mb	Biology
Leishmania major	0.0032	8022	33.6	239	Euglenozoan parasite
Cryptosporidium parvum	0.013	3396	10.4	327	Alveolate parasite
Plasmodium falciparum	1.39	5319	22.8	233	Alveolate parasite
Guillardia theta (nucleomorph)	0.035	486	0.55	884	Algal nucleomorph
Encephalitozoon cuniculi	0.0075	1996	2.5	798	Fungi, intracellular parasite
Saccharomyces cerevisiae	0.047	5770	12.4	465	Fungi, budding yeast
Schizosaccharomyces pombe	0.96	4929	12.4	398	Fungi, fission yeast
Cyanidioschyzon merolae	0.005	5331	16.5	323	Alga, acidophilic and thermophilic

[a]Species were grouped and color-coded according to broad phylogenetic groups, and then sorted by intron density.

[b]All data are from complete genomes: *C. parvum* data are from Ref. [56]; *L. major* genome annotation is from The Wellcome Trust Sanger Institute (http://www.sanger.ac.uk/Projects/L_major/); our intron density calculation assumed the minimum intron size was 20 bp. All other genome data sources are the same as shown in Figure 1.

or higher plants. Introns also contain many highly conserved elements that can function at the DNA level. More than a fifth of the 481 ultraconserved elements (100/481) shared by human, mouse and rat that were identified recently [28] are intronic. There are >5000 shorter ultraconserved elements (of at least 100 bp), suggesting that there might be 1000 ultraconserved intronic elements in mammalian genomes.

The indications from alternative splicing, intronic non-coding transcripts and ultraconserved elements are that a certain proportion of introns have important functions in multicellular eukaryotes, but just how many introns are functional has yet to be determined [29]. Nevertheless, some intronic functions appear to have become essential for the production of complex multicellular organisms [30], and, therefore, once they accumulate they cannot be lost.

Some genes use introns more than others

Many studies of intron evolution implicitly assume that patterns of intron evolution displayed by one gene are representative of the entire genome. However, there is evidence that different functional classes of genes have different propensities for intron gain and loss. Even in metazoans, which have smaller populations than unicells, some genes are apparently under selective pressure to minimize introns. Genes that are highly expressed in humans and *C. elegans*, and human genes that are expressed in all tissues (probably a similar set of genes) tend to have short introns [31,32], arguably reflecting selection against the expense and time of transcription and processing.

By contrast, introns are more abundant in the ribosomal genes of *E. cuniculi*, *Guillardia theta* and *S. cerevisiae* and its related yeasts than in the rest of their genomes [14,33–35]. These introns might have been retained because ribosomal proteins are more likely to contain intron-encoded snoRNAs [36], or because their expression is coupled to the transcription-splicing-mRNA transport process [34,37]. Intron length also appears to be positively correlated with expression in unicellular eukaryotes and negatively correlated in multicellular eukaryotes [38].

Several studies have shown that certain groups of genes have different rates of IGL, illustrating how different intron densities are established. Rates of IGL vary in genes that have been duplicated; paralogs differ more than orthologs in *Plasmodium* species [39], and genes that have undergone lineage-specific expansions have increased rates of intron gain [23].

The ancient evolutionary history of introns in the tree of life

The factors determining the evolutionary fate of any given intron will depend on the intron itself (e.g. introns that contain untranslated RNA genes), the gene in which it resides (which might be highly expressed or an alternatively spliced gene) and the host organism (depending on population size and generation time) (Figure 4). Different intron densities might be selected for because of 'incidental' facets of the biology of genes; for example, human genes involved in cell communication and enzyme regulation are more likely to be alternatively spliced, which will constrain loss of introns [40]. It follows that different evolutionary 'rules' apply to different introns.

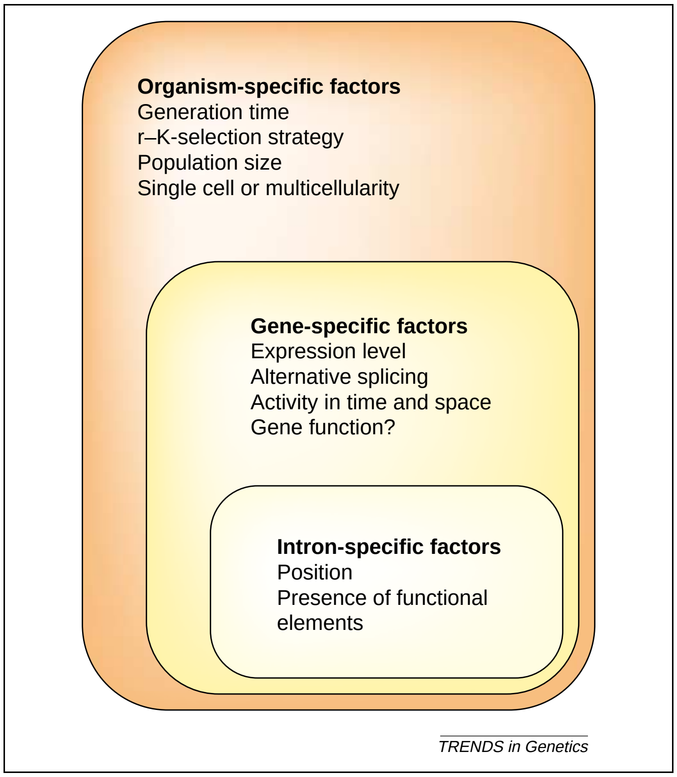


Figure 4. Factors that can affect the gain and loss of introns. The fate of any given intron is dependent on three facets of the genes biology. Organism-specific factors will determine the overall rates of gain or loss. Gene-specific factors differentiate between particular groups of genes, such as those with extensive alternative splicing. Finally, certain introns within a gene are more likely to be gained or lost, such as the bias for loss in 5′ introns (Box 2). Examples of possible factors at each level are shown.

Because intron density can vary so much under the influence of selection, we cannot yet make a realistic estimation of the intron density or genome size of the earliest eukaryotes. Our understanding of the processes occurring in extant eukaryotes, however, relates to theories about the absence of introns in prokaryotes (Box 1). The observation that introns have been almost entirely purged from highly reduced genomes [11,33,35] indicates that even prokaryote genomes could be derived from an intron-containing ancestor [41,42].

If this were so, we would expect the spliceosome to be an ancient component of cells, and the biology of the spliceosome appears to be consistent with this idea. First, many spliceosomal proteins are conserved across all eukaryotes including basal eukaryotes [43]. The discovery that modern spliceosomes might be the largest RNA-protein complex in the cell, exceeding even the size of ribosomes [44], and the coupling between splicing and essential RNA-processing steps [37] is consistent with a model of the spliceosome evolving gradually over a long period of time. This could have occurred in the (possibly long) branch of the tree of life between the divergence of prokaryotes and eukaryotes and earliest the divergence of the first eukaryote taxa (Box 1). But the involvement of

RNA at the core of the spliceosome and the catalytic potential of the RNA components of the spliceosome is consistent with the evolution of the spliceosome much earlier, in the RNA world [45].

Whatever the initial origins of intron and the spliceosome were, at present it is apparent that intron evolution is a dynamic process in eukaryotes. Introns are gained and lost to varying extents in different genomes in response to strong selective pressures, so that intron use is an important component of genome adaptation. Only a few of the factors that influence intron density have been identified to date, we can expect many more to become apparent in the future.

Acknowledgements

D.C.J. was supported by The Carlsberg Foundation and by the Wellcome Trust. T.M. was supported by The Carlsberg Foundation. D.P. was supported by The Marsden Fund. We thank A. Ivens for permission to use the *L. major* genome annotation, Jason Stajich for use of the fungal phylogeny, Motomichi Matsuzaki for the *C. merolae* annotation and all those who assisted with critical reading of this article.

References

1 Logsdon, J.M., Jr. (1998) The recent origins of spliceosomal introns revisited. *Curr. Opin. Genet. Dev.* 8, 637–648
2 de Souza, S.J. (2003) The emergence of a synthetic theory of intron evolution. *Genetica* 118, 117–121
3 Roy, S.W. (2003) Recent evidence for the exon theory of genes. *Genetica* 118, 251–266
4 Lynch, M. and Richardson, A.O. (2002) The evolution of spliceosomal introns. *Curr. Opin. Genet. Dev.* 12, 701–710
5 de Souza, S.J. *et al*. (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
6 Rogozin, I.B. *et al*. (2003) Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. *Curr. Biol.* 13, 1512–1517
7 Fedorov, A. *et al*. (2002) Large-scale comparison of intron positions among animal, plant, and fungal genes. *Proc. Natl. Acad. Sci. U. S. A.* 99, 16128–16133
8 Sverdlov, A.V. *et al*. (2005) Conservation versus parallel gains in intron evolution. *Nucleic Acids Res.* 33, 1741–1748
9 Gilson, P.R. and McFadden, G.I. (1996) The miniaturized nuclear genome of eukaryotic endosymbiont contains genes that overlap, genes that are cotranscribed, and the smallest known spliceosomal introns. *Proc. Natl. Acad. Sci. U. S. A.* 93, 7737–7742
10 Lander, E.S. *et al*. (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
11 Mourier, T. and Jeffares, D.C. (2003) Eukaryotic intron loss. *Science* 300, 1393
12 Nielsen, C.B. *et al*. (2004) Patterns of intron gain and loss in fungi. *PLoS Biol.* 2, e422
13 Banyai, L. and 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
14 Bon, E. *et al*. (2003) Molecular evolution of eukaryotic genomes: hemiascomycetous yeast spliceosomal introns. *Nucleic Acids Res.* 31, 1121–1135
15 Roy, S.W. and Gilbert, W. (2005) Rates of intron loss and gain: implications for early eukaryotic evolution. *Proc. Natl. Acad. Sci. U. S. A.* 102, 5773–5778
16 Roy, S.W. *et al*. (2003) Large-scale comparison of intron positions in mammalian genes shows intron loss but no gain. *Proc. Natl. Acad. Sci. U. S. A.* 100, 7158–7162
17 Zdobnov, E.M. *et al*. (2002) Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* 298, 149–159
18 Lynch, M. (2002) Intron evolution as a population-genetic process. *Proc. Natl. Acad. Sci. U. S. A.* 99, 6118–6123
19 Neugebauer, K.M. (2002) On the importance of being co-transcriptional. *J. Cell Sci.* 115, 3865–3871
20 van Wijk, E. *et al*. (2004) Identification of 51 novel exons of the Usher syndrome type 2A (*USH2A*) gene that encode multiple conserved functional domains and that are mutated in patients with Usher syndrome type II. *Am. J. Hum. Genet.* 74, 738–744
21 Le Hir, H. *et al*. (2003) How introns influence and enhance eukaryotic gene expression. *Trends Biochem. Sci.* 28, 215–220
22 Archibald, J.M. *et al*. (2002) The chaperonin genes of jakobid and jakobid-like flagellates: implications for eukaryotic evolution. *Mol. Biol. Evol.* 19, 422–431
23 Babenko, V.N. *et al*. (2004) Prevalence of intron gain over intron loss in the evolution of paralogous gene families. *Nucleic Acids Res.* 32, 3724–3733
24 Brett, D. *et al*. (2002) Alternative splicing and genome complexity. *Nat. Genet.* 30, 29–30
25 Kampa, D. *et al*. (2004) Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res.* 14, 331–342
26 Thanaraj, T.A. *et al*. (2003) Conservation of human alternative splice events in mouse. *Nucleic Acids Res.* 31, 2544–2552
27 Mattick, J.S. and Gagen, M.J. (2001) The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol. Biol. Evol.* 18, 1611–1630
28 Bejerano, G. *et al*. (2004) Ultraconserved elements in the human genome. *Science* 304, 1321–1325
29 Johnson, J.M. *et al*. (2005) Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. *Trends Genet.* 21, 93–102
30 Graveley, B.R. (2001) Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* 17, 100–107
31 Castillo-Davis, C.I. *et al*. (2002) Selection for short introns in highly expressed genes. *Nat. Genet.* 31, 415–418
32 Eisenberg, E. and Levanon, E.Y. (2003) Human housekeeping genes are compact. *Trends Genet.* 19, 362–365
33 Katinka, M.D. *et al*. (2001) Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* 414, 450–453
34 Ares, M., Jr. *et al*. (1999) A handful of intron-containing genes produces the lion's share of yeast mRNA. *RNA* 5, 1138–1139
35 Douglas, S. *et al*. (2001) The highly reduced genome of an enslaved algal nucleus. *Nature* 410, 1091–1096
36 Maxwell, E.S. and Fournier, M.J. (1995) The small nucleolar RNAs. *Annu. Rev. Biochem.* 64, 897–934
37 Maniatis, T. and Reed, R. (2002) An extensive network of coupling among gene expression machines. *Nature* 416, 499–506
38 Vinogradov, A.E. (2001) Intron length and codon usage. *J. Mol. Evol.* 52, 2–5
39 Castillo-Davis, C.I. *et al*. (2004) Accelerated rates of intron gain/loss and protein evolution in duplicate genes in human and mouse malaria parasites. *Mol. Biol. Evol.* 21, 1422–1427
40 Johnson, J.M. *et al*. (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 302, 2141–2144
41 Forterre, P. and Philippe, H. (1999) Where is the root of the universal tree of life? *BioEssays* 21, 871–879
42 Poole, A. *et al*. (1999) Early evolution: prokaryotes, the new kids on the block. *BioEssays* 21, 880–889
43 Collins, L. and Penny, D. (2005) Complex spliceosomal organization ancestral to extant eukaryotes. *Mol. Biol. Evol.* 22, 1053–1066
44 Jurica, M.S. and Moore, M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* 12, 5–14
45 Jeffares, D.C. *et al*. (1998) Relics from the RNA world. *J. Mol. Evol.* 46, 18–36
46 Giroux, M.J. *et al*. (1994) *De novo* synthesis of an intron by the maize transposable element dissociation. *Proc. Natl. Acad. Sci. U. S. A.* 91, 12150–12154
47 Coghlan, A. and Wolfe, K.H. (2004) Origins of recently gained introns in *Caenorhabditis*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11362–11367
48 Sverdlov, A.V. *et al*. (2004) Preferential loss and gain of introns in 3′ portions of genes suggests a reverse-transcription mechanism of intron insertion. *Gene* 338, 85–91
49 Gilbert, W. (1978) Why genes in pieces? *Nature* 271, 501

50 Fink, G.R. (1987) Pseudogenes in yeast? *Cell* 49, 5–6

51 Poole, A.M. *et al*. (1998) The path from the RNA world. *J. Mol. Evol.* 46, 1–17

52 Derr, L.K. and Strathern, J.N. (1993) A role for reverse transcripts in gene conversion. *Nature* 361, 170–173

53 Roy, S.W. and Gilbert, W. (2005) The pattern of intron loss. *Proc. Natl. Acad. Sci. U. S. A.* 102, 713–718

54 Keeling, P.J. and Fast, N.M. (2002) Microsporidia: biology and evolution of highly reduced intracellular parasites. *Annu. Rev. Microbiol.* 56, 93–116

55 Baldauf, S.L. *et al*. (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290, 972–977

56 Abrahamsen, M.S. *et al*. (2004) Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* 304, 441–445