Update TRENDS in Genetics Vol.xxx No.x

Full text provided by www.sciencedirect.com ScienceDirect

Genome Analysis

A glimpse of a putative pre-intron phase of eukaryotic evolution

Alexander V. Sverdlov[1], Miklos Csuros[2], Igor B. Rogozin[1] and Eugene V. Koonin[1]

[1] National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

[2] Department of Computer Science and Operations Research, Université de Montréal, Montréal, Québec QC H3C 3J7, Canada

Comparison of the exon–intron structures of ancient eukaryotic paralogs reveals the absence of conserved intron positions in these genes. This is in contrast to the conservation of intron positions in orthologous genes from even the most evolutionarily distant eukaryotes and in more recent paralogs. The lack of conserved intron positions in ancient paralogs probably reflects the origination of these genes during the earliest phase of eukaryotic evolution, which was characterized by concomitant invasion of genes by group II self-splicing elements (which were to become introns in the future) and extensive duplication of genes.

The mystery of the intron invasion

The hallmark of eukaryotic gene structure is the interruption of protein-coding sequences by noncoding sequences known as introns [1–4]. Introns are present in all of the eukaryotic genomes that have been sequenced, including the most compact ones, those of parasitic, unicellular eukaryotes [5–7]. Accordingly, the key components of the spliceosome – the elaborate RNA- and protein-containing molecular machine that mediates the removal of introns and the joining of exons – are conserved throughout the eukaryotic domain of life [8,9]. Moreover, the positions of numerous introns are conserved between orthologous genes from the most evolutionarily distant eukaryotes, in particular between animals and plants [10–12]. Thus, introns seem to have been present in eukaryotic genes since the earliest stages of eukaryogenesis and might have had an important role in the emergence of the nucleus and other features of eukaryotic cell organization [13,14]. By contrast, and counter to the introns-early hypothesis [15,16] (see Glossary), there is no evidence that typical spliceosomal introns or the spliceosome itself were ever present in prokaryotes, although spliceosomal introns are thought to be derived from group II self-splicing elements, which are found in many bacteria, in eukaryotic organelles and in certain archaea. Some of these group II self-splicing elements are bona fide introns, whereas others insert themselves in intergenic regions [2,17–19]. The lack of information on the transitional phase between the almost intronless genomes of prokaryotes (the presence of the occasional self-splicing introns notwithstanding) and the genomes of eukaryotes, which are riddled with introns in protein-coding genes, leaves a large gulf in our understanding of the origins of eukaryotic cell organization.

Lack of conservation of intron positions in ancient eukaryotic paralogs

We sought insight into the events that occurred during the early, formative, phase of eukaryogenesis by comparing the intron positions in paralogous genes that are duplicated in all sequenced eukaryotic genomes but not in any prokaryotic genomes. More than 2000 sets of such ancient eukaryotic paralogs have been identified and inferred to have evolved by duplications during the early phase of eukaryotic evolution (i.e. between the emergence of the eukaryotic cell and the radiation of the extant eukaryotic lineages) [20]. Many of the ancient paralogs show limited sequence conservation, which complicates the accurate alignment of these sequences and the identification of conserved introns. We generated amino acid sequence alignments for the 157 most-conserved sets of ancient paralogs, and we mapped the intron positions onto the alignments, essentially as previously described [11]. We then counted the intron positions that were shared between the ancient paralogs in unambiguously aligned regions (see the supplementary material online). This analysis was carried out either for paralogs within a genome or for multiple alignments of paralogs from a set of genomes (Figure 1). Unexpectedly, we observed very low conservation of intron positions between ancient paralogs: in each of the comparisons, $<3\%$ of the intron positions were shared (Figure 1; Table S2 in the supplementary material online), which is similar to the level of independent intron insertion that is expected in the same positions

Glossary

Group II self-splicing elements: a distinct class of mobile element that combines the ability to undergo ribozyme-catalyzed self-splicing with transposition mediated by the reverse transcriptase that is encoded by the element. These elements insert themselves in protein-coding genes (mainly in the organelles of plants, fungi and protozoa) or in intergenic regions (mainly in bacteria and certain archaea). They are thought to be ancestors of eukaryotic spliceosomal introns.

Introns-early hypothesis: a hypothesis that states that introns in protein-coding genes are an ancestral feature of gene organization that was present in the last universal common ancestor of cellular life but was eliminated in prokaryotes during 'genome streamlining'. The opposing hypothesis – the introns-late hypothesis – states that introns were inserted in protein-coding genes during eukaryotic evolution.

Paralogs: genes that evolved by duplication of an ancestral gene.

Corresponding author: Koonin, E.V. (koonin@ncbi.nlm.nih.gov).

Available online xxxxxx.

  doi:10.1016/j.tig.2007.01.001

Please cite this article in press as: Sverdlov, A.V. et al., A glimpse of a putative pre-intron phase of eukaryotic evolution, Trends Genet. (2007), doi:10.1016/j.tig.2007.01.001

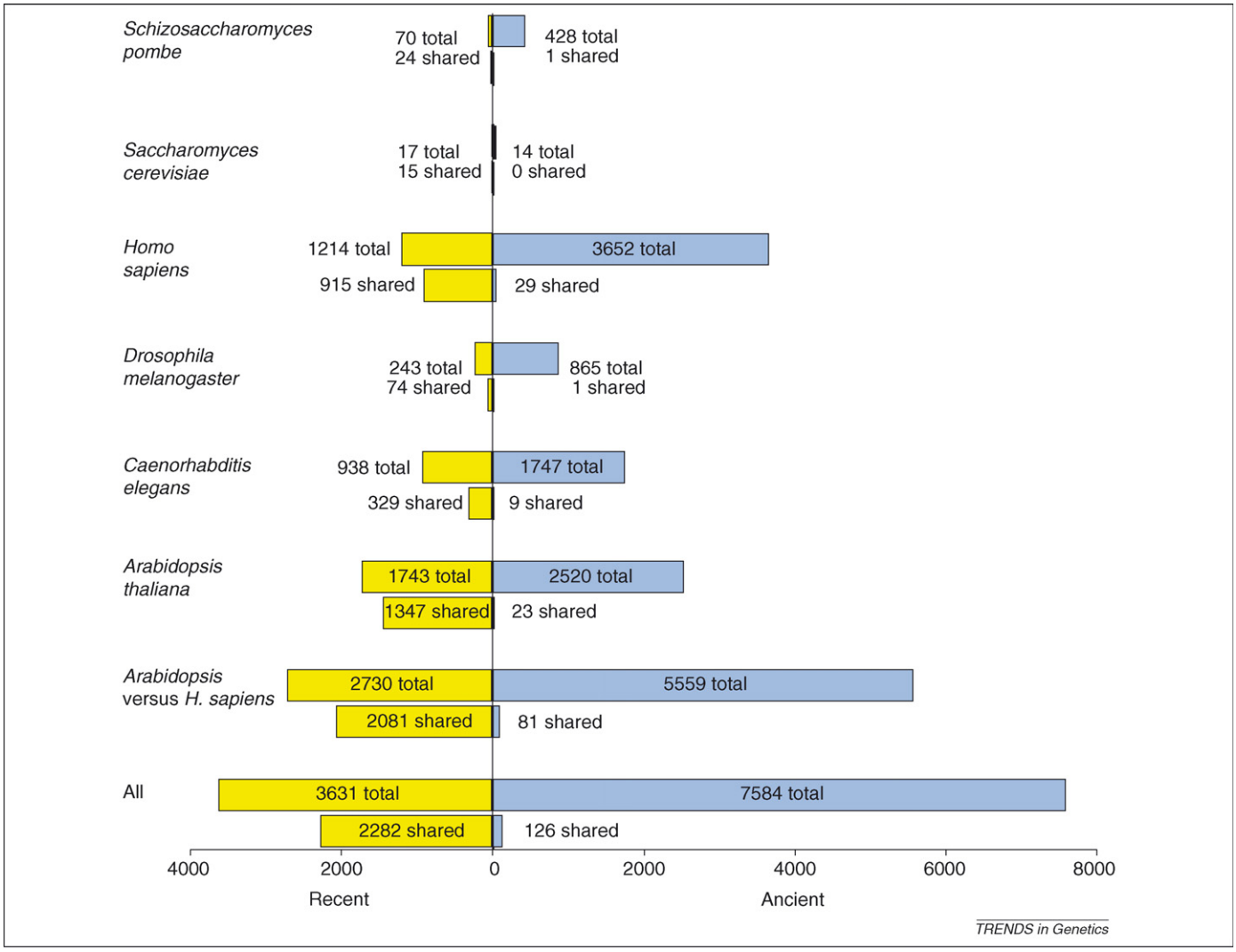

Figure 1. Conservation of intron positions in ancient and recent eukaryotic paralogs. Conservation of introns was assessed by the following: (i) multiple alignments of paralogous sequences from six species (i.e. 12 sequences); (ii) alignments of paralogs from two of these species, *Arabidopsis thaliana* and *Homo sapiens* (two sequences against two sequences); and (iii) alignments of paralogs from each of the six species separately (six separate twofold alignments). In the first two comparisons, an intron position was considered to be conserved if it was shared by any pair of paralogs. The results are given for the alignment stringency ±5 (see the supplementary material online).

of homologous genes [21]. The few shared intron positions seemed to be randomly scattered over the analyzed set of ancient paralogous genes; therefore, there is no indication that the exon–intron structure is conserved in any subset of ancient paralogs (data not shown). As a control, we examined the conservation of intron positions in more-recent, lineage-specific paralogs of the same genes (e.g. those resulting from duplications only in animals or only in plants) and observed a much greater level of conservation (Figure 1), in agreement with previous findings [22].

Concomitant intron invasion and extensive gene duplication in the early phase of eukaryotic evolution

Our analysis shows the absence of appreciable conservation of intron positions between ancient paralogs, in contrast to more-recent paralogs of the same genes. This could be explained either by the rapid loss of introns after gene duplication during the early stages of eukaryotic evolution or by the absence of shared intron positions in ancient paralogs immediately after gene duplication. However, rapid intron loss is unlikely to be the main factor underlying the absence of conserved intron positions in ancient paralogs because, within orthologous sets, the same genes have many intron positions that are conserved between the most evolutionarily distant eukaryotic genomes available [11]. Specifically, in the genes comprising the ancient paralogous sets analyzed in *Homo sapiens* and *Arabidopsis thaliana*, the positions of 193 of the 6624 introns present (2.9%) are conserved between these genomes, compared with 1230 introns (18.6%) between orthologs in the same gene set.

Thus, it seems most probable that the ancient paralogs lacked introns at the time of duplication and/or duplicated genes by a mechanism, such as reverse transcription, that involved loss of introns in one of the copies. Conceivably, both of these factors made important contributions to the observed pattern (Figure 2). If the duplications that generated ancient paralogs occurred concomitantly with the initial process of intron insertion in the genes of the emerging eukaryote [13,14], then many of these duplications would involve intronless or intron-poor genes. Presumably, this early phase of eukaryogenesis was also characterized by a high level of reverse transcription

Please cite this article in press as: Sverdlov, A.V. et al., A glimpse of a putative pre-intron phase of eukaryotic evolution, Trends Genet. (2007), doi:10.1016/j.tig.2007.01.001

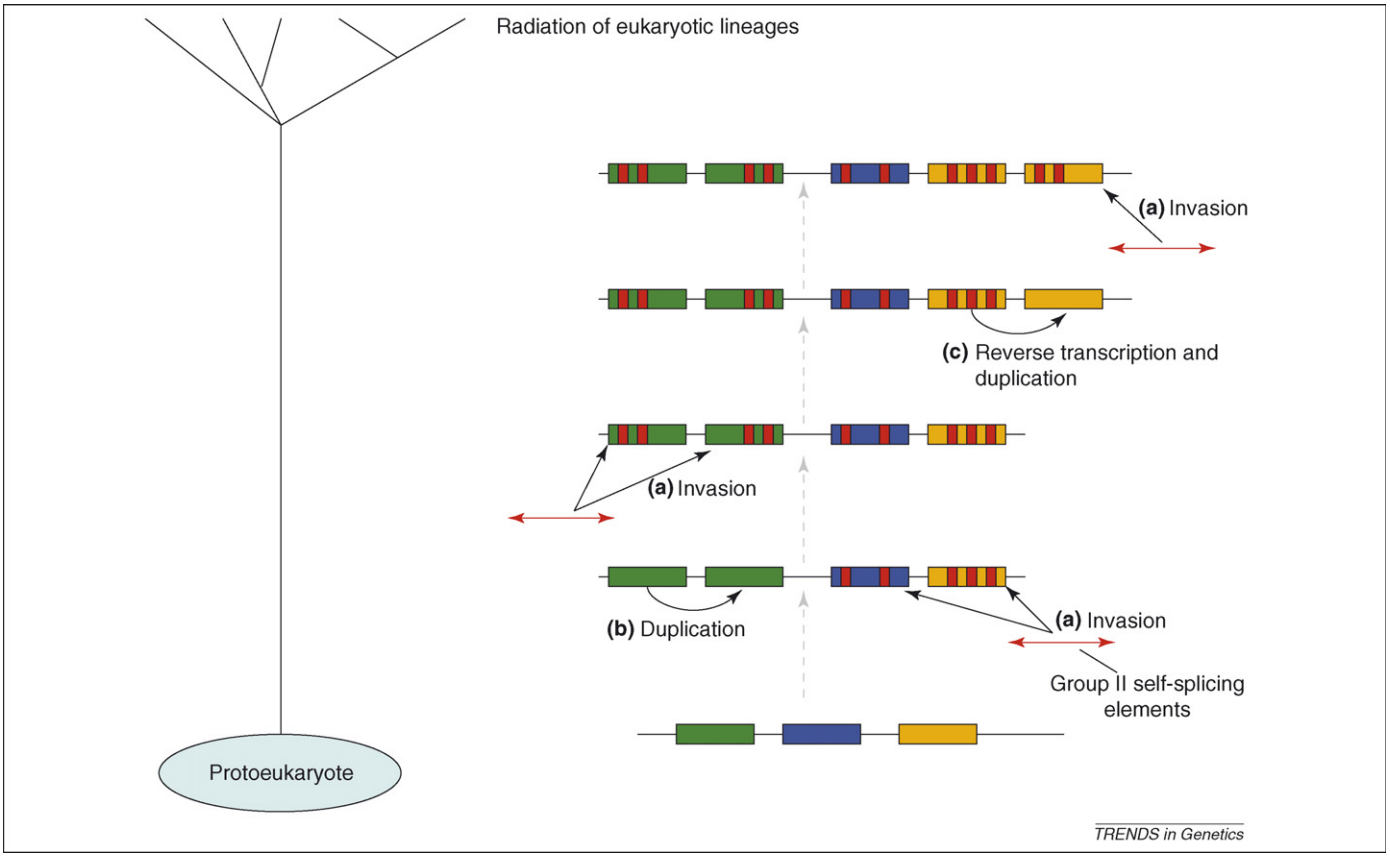

Figure 2. The processes that probably account for the lack of conservation of intron positions between ancient eukaryotic paralogs. **(a)** Ongoing invasion of group II self-splicing elements into eukaryotic genes, giving rise to spliceosomal introns. **(b)** Duplication of an intronless gene followed by differential insertion of introns into the paralogs. **(c)** Reverse-transcription-mediated duplication of an intron-containing gene, yielding an intronless paralog that, subsequently, accumulates introns in different positions. A schematic tree of eukaryotic evolution is shown, emphasizing that all of these processes are attributed to the time between the emergence of the eukaryotes and the radiation of the known eukaryotic lineages.

conferred by proliferating group II self-splicing elements, the progenitors of introns. This would drive extensive reverse-transcription-mediated gene duplication, with any previously inserted introns lost from the new copies of the duplicated genes. Therefore, the lack of conservation of intron position between ancient paralogs is likely to reflect their origination by gene duplication during the earliest, formative, phase of eukaryotic evolution, which pre-dates the relatively intron-rich state that has been inferred [3,11,12] for the genes of the last common ancestor of modern eukaryotes.

Concluding remarks

We found that, in contrast to orthologs from even the most distant eukaryotic species or to relatively recent paralogs, ancient paralogs that evolved by duplication before the divergence of the main lineages of eukaryotes have almost no shared intron positions. It seems most probable that the duplications that generated the ancient eukaryotic paralogs occurred concomitantly with the massive invasion of group II self-splicing elements (the ancestors of spliceosomal introns) into protein-coding genes. Therefore, the ancient duplications might often have involved intronless or intron-poor genes. In addition, many of these duplications might have occurred by the reverse-transcription pathway, mediated by the reverse transcriptase encoded by group II self-splicing elements. This pathway also would have yielded intronless gene duplicates, which, subsequently, would have accumulated introns independently of their paralogs.

Acknowledgements

We thank Kira Makarova, Sergei Mekhedov and Yuri Wolf for help and advice during the early stages of this project. This research was supported by the Intramural Research Program of the National Institutes of Health (National Center for Biotechnology Information, National Library of Medicine).

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2007.01.001.

References

1 Gilbert, W. (1978) Why genes in pieces? *Nature* 271, 501
2 Lynch, M. and Richardson, A.O. (2002) The evolution of spliceosomal introns. *Curr. Opin. Genet. Dev.* 12, 701–710
3 Roy, S.W. and Gilbert, W. (2006) The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat. Rev. Genet.* 7, 211–221
4 Rodriguez-Trelles, F. *et al.* (2006) Origin and evolution of spliceosomal introns. *Annu. Rev. Genet.* 40, 47–76
5 Embley, T.M. and Martin, W. (2006) Eukaryotic evolution, changes and challenges. *Nature* 440, 623–630
6 Nixon, J.E. *et al.* (2002) A spliceosomal intron in *Giardia lamblia*. *Proc. Natl. Acad. Sci. U. S. A.* 99, 3701–3705
7 Simpson, A.G. *et al.* (2002) Eukaryotic evolution: early origin of canonical introns. *Nature* 419, 270
8 Jurica, M.S. and Moore, M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* 12, 5–14
9 Collins, L. and Penny, D. (2005) Complex spliceosomal organization ancestral to extant eukaryotes. *Mol. Biol. Evol.* 22, 1053–1066

Please cite this article in press as: Sverdlov, A.V. et al., A glimpse of a putative pre-intron phase of eukaryotic evolution, Trends Genet. (2007), doi:10.1016/j.tig.2007.01.001

10 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
11 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
12 Roy, S.W. and Gilbert, W. (2005) Complex early genes. *Proc. Natl. Acad. Sci. U. S. A.* 102, 1986–1991
13 Koonin, E.V. (2006) The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biol. Direct*,1, 22 DOI: 10.1186/1745-6150-1-22 (www.biology-direct.com)
14 Martin, W. and Koonin, E.V. (2006) Introns and the origin of nucleus–cytosol compartmentalization. *Nature* 440, 41–45
15 Doolittle, W.F. (1978) Genes in pieces: were they ever together? *Nature* 272, 581–582
16 Gilbert, W. and Glynias, M. (1993) On the ancient nature of introns. *Gene* 135, 137–144
17 Rogers, J.H. (1990) The role of introns in evolution. *FEBS Lett.* 268, 339–343
18 Lambowitz, A.M. and Zimmerly, S. (2004) Mobile group II introns. *Annu. Rev. Genet.* 38, 1–35
19 Robart, A.R. and Zimmerly, S. (2005) Group II intron retroelements: function and diversity. *Cytogenet. Genome Res.* 110, 589–597
20 Makarova, K.S. *et al.* (2005) Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res.* 33, 4626–4638
21 Sverdlov, A.V. *et al.* (2005) Conservation versus parallel gains in intron evolution. *Nucleic Acids Res.* 33, 1741–1748
22 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

Please cite this article in press as: Sverdlov, A.V. et al., A glimpse of a putative pre-intron phase of eukaryotic evolution, Trends Genet. (2007), doi:10.1016/j.tig.2007.01.001