Bioprocess 生物过程, 2011, 1, 9-12
http://dx.doi.org/10.12677/bp.2011.12003 Published Online December 2011 (http://www.hanspub.org/journal/bp/)

Research Progress in Eukaryotic Intron*

Jun Cao
Institute of Life Science, Jiangsu University, Zhenjiang
Email: cjinfor@163.com
Received: Sep. 9th, 2011; revised: Oct. 6th, 2011; accepted: Oct. 28th, 2011.

Abstract: Intron is the genome sequence that is cut out in mature RNA transcripts. Full sequencing of a number of eukaryotic genome gives us some help to understand the intron evolution, exon-intron organization etc. This paper reviewed some progress in intron distrubtion, intron generated hypothesis, spliceosome and major splice sites, intron acquisition and lose and its mechanisms, factors affecting the evolution of intron, and so on.

Keywords: Intron; Evolution; Intron Acquisition and Lose

真核生物内含子研究进展*

曹　军
江苏大学生命科学研究院，镇江
Email: cjinfor@163.com
收稿日期：2011 年 9 月 9 日；修回日期：2011 年 10 月 6 日；录用日期：2011 年 10 月 28 日

摘　要：内含子是成熟 RNA 转录本中被剪切掉的基因组序列。一些真核基因组测序的完成有助于了解内含子进化问题，如真核基因外显子/内含子的结构等等。本文就内含子分布、内含子产生假说、拼接体及主要拼接位点、内含子获得和丢失及其机制、影响因素等方面做简要综述。

关键词：内含子；进化；内含子获得与丢失

1. 引言

内含子是基因内间隔的一段序列，它不出现在成熟的 RNA 分子中，在转录后通常被加工去除。自上世纪七十年代内含子被发现以来，这种独特的结构元件一直备受人们关注。内含子存在于已知所有类型生物中，根据内含子存在宿主的不同，可以把内含子分为：mRNA 内含子、rRNA 内含子、snoRNA 内含子及 snRNA 内含子等。根据内含子剪接机制的不同，又可把内含子以下四类：I 类内含子、II 类内含子、III 类内含子和 IV 类内含子。其中，I 类内含子存在于叶绿体、线粒体、某些低等真核生物的 rRNA 基因及原核生物的噬菌体中；II 类内含子主要存在于细胞器(如线粒体)及细菌中，它的剪接位点类似于 III 类内含子，并同样遵从 GT|AG 规律，但 II 类内含子不同于 III 类内含子主要表现在：II 类内含子具有自我剪接功能，不需要剪接体和 snRNA 的参与，也不需要 ATP 功能；最为常见的是 III 类内含子，又叫拼接体内含子，存在于绝大多数真核细胞 mRNA 前体中；另外一种是 IV 类内含子，存在于核 tRNA 前体中，又称核 tRNA 前体内含子，它是 tRNA 的基因及其转录产物中的内含子。本文主要针对拼接体内含子研究现状做简要综述。

2. 拼接体内含子特点及其基因组分布

拼接体内含子仅存在于真核生物核基因组中，其长度往往与所处物种有关，如线虫、昆虫中内含子长度有的仅几十个碱基，而哺乳动物中有的却高达几百万以上。

根据内含子插入密码子位置的不同又可把拼接体内含子分为三类：内含子插入在两密码子之间的为 0 相内含子；插入在密码子第一个碱基之后的为 1 相内含子；插入在密码子第二个碱基之后的为 2 相内含子。研究表明：在所有已知基因组中，0 相内含子所占比例最大。内含子数目与基因组复杂性有关，一般认为：

*资助信息：江苏省自然科学基金(BK2011467)，江苏大学高级人才启动基金(10JDG027)。

越复杂的生物含有的内含子数目越多。此外，内含子在基因中的分布位置也因生物体不同而表现差异：对单细胞真核生物而言，内含子往往分布于基因的5’端；而在多细胞生物体中，这种分布较为均匀[1]。

3. 内含子早现和内含子晚现假说

目前主要有两种假说来解释真核基因外显子/内含子结构的进化：内含子早现假说和内含子晚现假说。内含子早现假说认为：内含子是远古就有的，现存真核基因中几乎所有的内含子均起源于原核祖先，同源基因结构的差异主要是由内含子丢失所致[2]。相反，内含子晚现假说却认为：真核生物的内含子是后来才出现的，内含子插入促使真核细胞基因结构发生改变[3]。研究表明：0 相内含子在远古基因组中占绝对优势，而现存真核细胞中的 0 相内含子很可能是远古基因留下的遗迹。

4. 拼接体与主要拼接位点

大多数真核蛋白编码基因被许多内含子分开，这些内含子在拼接供体和受体位点处被切断，从而使邻近的外显子连接在一起，该过程主要由五种小分子核蛋白组成的拼接体参与完成。拼接体与内含子及其邻近外显子的特定部分相互作用，以确保精确有效的剪接。在该过程中，内含子及外显子邻近处的核苷参与了该反应，共同序列为(A/C)AG|GU(A/G)AGU。拼接供体序列与U1 snRNA的 5’端序列互补，它们相互作用而引起拼接反应。核苷CAG|G是典型的拼接受体序列，它能被U5 snRNA识别。除了拼接供体和受体序列外，还有一个短的分支点信号，它一般位于拼接受体位点上游的内含子序列内，并且该序列中含有参与套索结构形成的腺苷。首先U1 snRNP与mRNA前体的供体位点结合，然后U2 snRNP以碱基互补配对方式与分支点结合，接着U4/U6 和U5 snRNP复合体组装进拼接体，通过整理snRNA分子而完成拼接反应[4]。内含子附近(A/C)AG|G序列是内含子插入的识别信号，即主要拼接位点，它是内含子进化的主要组成成分。绝大多数情况下内含子插入都会集中在主要拼接位点内，即供体位点为|GT，受体位点为AG|。然而在一些脊椎动物、昆虫和植物中，一些内含子供体拼接位点为|AT，而受体位点为AC|。拼接位点是如何被选择的？人们对此知之甚少。当把肌动蛋白基因中正常拼接位点去除时，往往会引起其它拼接位点的活化[5]。内含子及拼接体均出现在早期的真核细胞中，在进化关系较远的物种间，内含子位置、拼接体及其相关蛋白高度保守，这说明真核细胞的祖先至少含有部分内含子及最初的拼接体，所以真核细胞可能起源于具有复杂拼接体和内含子的祖先[6]。

5. 内含子获得与丢失

真核生物中约 30%的内含子位置高度保守，其它内含子在进化过程中呈动态分布，如人类基因组内含子密度为8.4，而小孢子虫(*E. cuniculi*)的仅为0.0075[1,7]。可见不同的基因组，内含子获得或丢失的频率存在很大差异：一些基因组中内含子丢失占优势；而另一些却以获得为主。哺乳动物中，平均每个基因约有 0.003 个内含子丢失，且并没有发现明显的内含子获得现象[8]。相反，在线虫和果蝇基因组中，该值约为 0.5~1。分析发现：酵母几乎丢失了所有的远古内含子；而哺乳动物不仅获得了大量的内含子，其远古内含子丢失也很少；而其它脊椎动物中仅发现内含子丢失现象[8,9]。由此可见，在长期的进化过程中，真核生物内含子丢失占绝对优势，而内含子插入(获得)却随着不同的进化分枝而改变。内含子密度差异主要是由不同的内含子获得和丢失所致。一般认为，基因组中内含子的数量通常与有效群体大小及各物种特有的突变率有关。此外，各种选择性压力也可能影响内含子的获得和丢失。Carmel 等[10]研究发现：在真核生物进化过程中，内含子获得和丢失率在不同的进化分枝中各不相同，如真菌、啮齿动物、双翅目昆虫等物种中内含子丢失占优势；鞭毛生物、后生动物等物种中内含子获得占优势；而木兰、水稻、盘菌亚门等物种中内含子获得和丢失基本持平。这说明可能会有很多机制参与内含子获得和丢失。

真核物种内含子数目的变化并不像系统发生树中物种分离那么简单，在进化过程中一定发生了大量的内含子丢失或获得现象。目前主要有两个模型来解释内含子丢失现象：一个是基因转换(mRNA 转录本的反转录及两次重组)，另一个是基因组删除。两者具有明显差异：第一，mRNA 反转录可以精确地去除内含子，而基因组删除却不能，通常基因组删除也会删掉邻近的编码序列；第二，mRNA 反转录介导的内含子丢失需要 mRNA 作为中介，而基因组删除则不需要；第三，

由于反转录酶是从 RNA 分子的 3’端向 5’端起作用，并且经常产生不完整的转录本，那么反转录酶产物趋向于 3’端序列，这样就产生了偏重 3’端的内含子丢失现象[11]。内含子获得也有很多途径，包括内含子转座插入、串联基因组复制、内含子转换等。一方面，祖先生物若含有较少的内含子，而现有内含子丰富的物种一定经历了很多的内含子插入，物种间内含子位置相关性将不明显；另一方面，若几乎所有的内含子都来自祖先的话，现存内含子较少的物种在进化过程一定丢失了大量的内含子，这样，物种间内含子位置的相关性将非常明显。因而，可以说没有一个模型可以清晰地解释内含子的起源，内含子的进化可能是多种因素共同作用的结果[11]。

6. 影响内含子进化的因素

内含子可以参与基因调控和选择性剪接等生物学功能，因而在进化中它会受到选择压力的影响。通常，影响内含子进化的因素包括如下几种。

6.1. 选择性剪切

选择性剪切在高等生物基因组进化中扮演重要角色，研究表明：高等生物中约有 40%~70%基因要经历选择性剪切，通过选择性剪切不同的内含子而产生不同的基因产物[12]。所以，内含子是选择性剪切的必需成分，选择性剪切影响内含子进化。

6.2. 重组

内含子的存在可以增加基因间的重组率，研究表明：果蝇基因的重组率与内含子及其长度有关[13]。可见，内含子在基因组进化中扮演重要角色。

6.3. 基因保守性

Carmel 等[14]比较了 19 个真核物种中的 391 个保守基因，发现内含子的获得速率与编码序列进化率呈高度负相关，而内含子的丢失率与编码序列进化率呈高度正相关，即进化保守的基因更容易积累内含子。

6.4. 基因表达水平

高度表达的基因通常进化较慢，基因的表达水平可能是序列进化的主要决定因素。

6.5. 功能性需求

内含子中含有很多保守的功能性元件，它们可以增强 mRNA 的转录、加工和运输水平等。一些内含子的功能可能与多细胞生物体的产生有关[15]，故功能性需求在内含子进化过程中也起一定作用。

6.6. 基因功能

不同功能的基因往往具有不同的内含子获得和丢失率，如核糖体基因在内原虫(*E. cuniculi*)、酿酒酵母(*S. cerevisiae*)中具有较多的内含子，这可能与核糖体蛋白包含有内含子编码 snoRNA，在进化过程中使这些内含子保留下来[16]。

6.7. 细胞数目、世代时间和群体大小

世代时间较长和群体小的多细胞物种趋向于拥有高密度的内含子，而世代时间较短且群体大的单细胞物种趋向于拥有较少的内含子。内含子密度与生物个体世代时间也有一定关系，即繁殖迅速的生物体一般含有更少的内含子。个体较小的生物体通过选择压力来降低 mRNA 的加工时间，一般情况下 RNA 聚合酶 II 平均每分钟移动 1~1.5 千碱基对，但内含子切除却要花费大约三分钟，因此，内含子切除是 mRNA 处理的主要部分[17]。mRNA 处理时间对短世代生物体来说至关重要，如人类 *USH2A* 基因编码 790 kb 的转录本，其中包括约 770 kb 的内含子部分，这样转录这个转录本至少要花费 8 个小时。但是对不含内含子的 *USH2A* 基因(约 18.8 kb)来说，只需 12 分钟就能把它转录完。这种差异对人类来说没有什么区别，但对世代时间仅有一小时的酵母等生物体来说却至关重要[18]。

7. 内含子应用

7.1. 系统发生标记

一般认为物种间共享内含子位置片段随着进化距离的延长而降低，因此可以利用内含子的保守性作为一种系统发生标记。Verkatesh 等比较鱼类视网膜色素基因内含子发现：远古脊索动物视网膜色素基因的四个内含子在条鳍鱼中同时丢失[19]。所以在进化过程中，可以利用同源内含子的有无来定义进化分枝。

7.2. 进化距离测定

与基因的编码区相比，大多数内含子并不含有功能元件，因而受到较小的选择压力，通过计算内含子碱基替换速率测定进化距离。

7.3. 参与基因表达调控

研究表明，在转基因系统中引入内含子，可以提高目的基因的表达效率等[20]。

当前，一些研究主要集中在拼接体内含子进化方面，如内含子起源的时间和增殖、内含子丢失、获得的机制以及推动内含子进化的作用力等。尽管已取得一些进展，但很多问题仍没有解决，如内含子的产生，内含子的插入机制，不同物种间内含子数目的差异等等。比较分析物种间内含子位置及特异性内含子丢失和获得等方面的研究将有助于更好地解释内含子进化的分子机制等。

8. 致谢

本章节为作者提供“致谢”的示例。本研究得到江苏省自然科学基金(BK2011467)和江苏大学高级人才启动基金(10JDG027)的资助，在此深表谢意。

参考文献 (References)

[1] T. Mourier, D. C. Jeffares. Eukaryotic intron loss. Science, 2003, 300: 1393.
[2] W. Gilbert. The exon theory of genes. Cold Spring Harbor Symposia Quantitative Biology, 1987, 52: 901-905.
[3] G. Cho, R. F. Doolittle. Intron distribution in ancient paralogs sup-ports random insertion and not random loss. Journal of Molecular Evolution, 1997, 44(6): 573-584.
[4] M. Rosbash, B. Séraphin. Who’s on first? The U1 snRNP-59 splice site interaction and splicing. Trends in Biochemical Sciences, 1991, 16: 187-190.
[5] T. Sadusky, A. J. Newman and N. J. Dibb. Exon junction sequences as cryptic splice sites: Implications for intron origin. Current Biology, 2004, 14(6): 505-509.
[6] S. Vanacova, W. Yan, J. M. Carlton, et al. Spliceosomal introns in the deep-branching eukaryote Trichomonas vaginalis. P The Proceedings of the National Academy of Sciences USA, 2005, 102: 4430-4435.
[7] A. V. Sverdlov, I. B. Rogozin, V. N. Babenko, et al. Conservation versus parallel gains in intron evolution. Nucleic Acids Research, 2005, 33(6): 1741-1748.
[8] S. W. Roy, A. Fedorov and W. Gilbert. Large-scale comparison of intron positions in mammalian genes shows intron loss but no gain. The Proceedings of the National Academy of Sciences USA, 2003, 100(12): 7158-7162.
[9] V. N. Babenko, I. B. Rogozin, S. L. Mekhedov, et al. Prevalence of intron gain over intron loss in the evolution of paralogous gene families. Nucleic Acids Research, 2004, 32(12): 3724-3733.
[10] L. Carmel, I. B. Rogozin, Y. I. Wolf, et al. Evolutionarily conserved genes preferentially accumulate introns. Genome Research, 2007, 17: 1045-1050.
[11] S. W. Roy, et al. The evolution of spliceosomal introns: Patterns, puzzles and progress. Nature Reviews Genetics, 2006, 7: 211-221.
[12] D. Brett, H. Pospisi, J. Valcárcel, et al. Alternative splicing and genome complexity. Nature Genetics, 2002, 30(1): 29-30.
[13] J. M. Comeron, M. Kreitman. The correlation between intron length and recombination in Drosophila. Dynamic equilibrium between mutational and selective forces. Genetics, 2000, 156(3): 1175-1190.
[14] L. Carmel, Y. I. Wolf, I. B. Rogozin, et al. Three distinct modes of intron dynamics in the evolution of eukaryotes. Genome Research, 2007, 17: 1034-1044.
[15] B. R. Graveley. Alternative splicing: Increasing diversity in the proteomic world. Trends Genet, 2001, 17(2): 100-107.
[16] E. S. Maxwell, M. J. Fournier. The small nucleolar RNAs. Annual Review of Biochemistry, 1995, 64: 897-934.
[17] K. M. Neugebauer. On the importance of being co-transcriptional. Journal of Cell Science, 2002, 115: 3865-3871.
[18] E. van Wijk, R. J. Pennings, H. te Brinke, et al. Identification of 51 novel exons of the Usher syndrome type 2A (USH2A) gene that en-code multiple conserved functional domains and that are mutated in patients with Usher syndrome type II. The American Journal of Human Genetics, 2004, 74(4): 738-744.
[19] B. Venkatesh, Y. Ning and S. Brenner. Late changes in spliceosomal introns define clades in vertebrate evolution. The Proceedings of the National Academy of Sciences USA, 1999, 96(18): 10267-10271.
[20] A. B. Rose. Intron-mediated regulation of gene expression. Current Topics in Microbiology and Immunology, 2008, 326: 277-290.

