

文章编号: 0529-6579 (1999) 01-0049-53

基因组的进化与内含子中的基因的进化^{*}

张尚宏, 屈良鹄

(中山大学生物工程研究中心, 广州 510275)

摘 要: 论述了基因组整体进化过程中内含子所含的基因的进化. 分析了内含子的起源方式与内含子中的基因种类的关系, 并结合基因组在大小、组成等方面进化过程中内含子的演化趋势, 探讨了内含子中的基因, 特别是核仁小分子 RNA 基因的进化规律. 同时, 对于内含子中的基因的进化, 提出了一些可能的途径.

关键词: 基因组进化, 内含子, 内含子中的基因

分类号: Q 754 **文献标识码:** A

基因组的进化包括在大小、组成、序列、三维结构等方面的进化, 它是一种具有整体效应的过程. 从而, 基因组进化的各个方面都是相互联系, 并与基因组的整体进化密切相关的. 其中, 基因组的大小和组成的进化就很能体现出基因组整体发展的趋势, 而且基因组组成成分的特点还能反映出基因组进化的大体途径^[1]. 因此, 基因组的进化可以通过其个别方面来进行深入研究.

内含子是基因组组成中的一种重要而又令人迷惑的成分. 继在基因中发现非编码的内含子后, 又发现某些内含子可含有编码与它们活动有关的蛋白质的基因^[2], 近年更在一些核 mRNA 内含子中发现编码核仁小分子 RNA (snoRNA) 的基因^[3]. 虽然对内含子中含有基因这种特殊结构形式已有不少研究, 但至于这种结构是如何起源和发展变化, 却是有待进一步探讨的问题. 无疑, 这一方面对基因组的进化也是起着十分重要的作用. 因此, 对其进行深入研究, 必将能加深对基因组进化的认识.

1 内含子的起源与内含子中的基因

虽然内含子的起源仍然是一个没有定论的问题, 但内含子中的基因或多或少是会与内含子的起源方式有关的. 因此, 分析内含子各种可能的起源与内含子中的基因的状况, 就可以为后者的起源与进化问题的研究提供一些新的启示. 对于内含子的起源, 目前有“后起源”和“先起源”两类观点. “后起源”观点认为, 内含子都是作为间隔序列插入到连续编码的基因序列中而形成, 内含子在较高级的功能基因或在真核生物出现之后才产

^{*} 基金项目: 国家自然科学基金重点项目 (39730300)

收稿日期: 1998-06-15 第一作者简介: 张尚宏, 男, 36 岁, 副教授

生^[4]。在发现了内含子也存于原核生物的基因组中之后,支持该观点的依据主要还有内含子本身可具移动性这一点。“后起源”观点必须面对的一个难题是:内含子最初如何能插入到连续编码的基因中而对其功能丝毫无损?此外,该观点对内含子本身的序列是怎么产生也难以说明清楚,往往只能认为它们是基因组进化的副产物,是一些“自私的 DNA”^[4]。

“先起源”观点则认为:早期的内含子具有自催化、自我复制等能力,因此它们是原始基因和基因组的组织与复制必不可少的^[5]。内含子在原始基因组中就已存在,现代的内含子则是一类进化遗迹,它们能继续存在,是因为具有重新组合基因组中的外显子以形成新的基因的能力,即内含子能赋予其携带者更大的进化潜力。支持“先起源”观点的依据很多^[6],不过并不能排除内含子“后起源”这种产生方式。原则上,“先起源”观点也并不能完全否认一些内含子是“后起源”的。

在内含子“先起源”的前提下,对于内含子本身的序列的来源,则有观点认为是早期编码序列的上下游非编码部分成为了原始的内含子^[7];也有观点认为外显子是源于原始的微型基因,而内含子则是这些外显子的衍生物^[8];此外,还有观点认为最原始的基因组是由可自复制的生物大分子作为结构单元而组成的重复序列,外显子和内含子最初都是直接起源于这样的重复序列^[9]。在后两种观点的模式中,内含子与外显子实际上是同源的。

至于内含子中的基因,目前已知主要有两大类。第一大类是一些 I, II 类内含子和一些原细菌 rRNA 内含子中编码蛋白质的基因^[9, 10]。这些基因编码的蛋白质一般都是与内含子本身的活动有关:如起到类似逆转录酶、内切酶的作用,参与催化内含子在基因组内^[9]或基因组之间^[10]的移位或扩增;另外,起到所谓成熟酶的作用,参与催化有关内含子的剪接过程^[9]。另一大类是真核生物核 mRNA 内含子中的 snoRNA 基因^[3]。这些基因编码的 snoRNA 在真核生物 rRNA 的加工与修饰中起着十分重要的作用^[3, 11]。

在内含子中的基因种类与内含子的起源方式的关系方面,首先考虑“先起源”的情况。如果内含子与外显子同源,那么在内含子中含有基因,特别是一些十分原始的基因,是不足为奇的,这类基因的性质应该是由原始基因组的序列所决定。在起源上,编码功能 RNA 的基因似乎要比编码蛋白质的基因更早产生,因此,“先起源”的内含子若含有基因,最有可能就是一些编码小分子 RNA 的基因。从 snoRNA 的功能和结构看, snoRNA 基因应是一类相当原始的基因。在这种意义上,核 mRNA 内含子中编码 snoRNA 的基因很可能是与内含子的“先起源”联系在一起的。

在内含子“后起源”的情况下,它们则不大可能含有十分原始的基因,但却有可能含有一些次生性的基因;或者,含有一些有利于内含子本身移动、扩增的基因。一些 I, II 类内含子似乎就是属于这种情况。当然,这只能说明部分移动性较强的 I, II 类内含子是“后起源”的。而且,这些内含子所编码的蛋白质种类较多,编码序列的位置也多变(如 I 类内含子^[9]),因此它们所含的基因也比较有可能是在内含子的进化过程中才产生。

综合起来,有理由推测:内含子这种结构在原始基因中就已存在,而且当时内含子的序列与外显子序列有一定的同源性,也可以含有一些原始的基因;在进化中,仍有一部分内含子继续保留并发展其内的基因,还有一些内含子获得或形成了次生性的基因,可移动性增强,在基因组中移位或扩增,产生了“后起源”的内含子。因此,在现代基因组中,“先起源”和“后起源”的内含子是同时存在的。对于含有基因的内含子来说,其内基因

的种类就可以在一定程度上反映出其起源。

2 基因组进化中内含子的进化

内含子在基因组进化中无疑是起着相当重要的作用。另一方面, 它们又是基因组组成中的一种独特成分, 或者说是一种“非必需成分”(原始基因组中的情况除外)。因此, 随着基因组的进化, 内含子的进化除了其本身的进化(包括各类内含子的产生及其进化关系), 还有内含子在基因和基因组中的地位的进化。

根据各类内含子的结构与剪接方式的相似程度, 可以把它们归纳为 3 大类: ① I 类内含子; ② II, III 类内含子与核 mRNA 内含子; ③ 核 tRNA 内含子与原细菌的内含子(主要是 tRNA 和 rRNA 基因中的内含子)^[10]。其中, II, III 类内含子以及核 mRNA 内含子之间的进化关系已有比较多的事实依据支持^[9], 而 I 类内含子与它们的差别主要是在剪接过程的第一个步骤, 在其他方面还是有不少相似之处^[10]; 核 tRNA 内含子与原细菌的内含子则很可能是有共同起源的^[10], 但它们与其他内含子却似乎没有什么进化关系。因此, 目前在内含子本身的进化问题上, 基本上公认的是核 mRNA 内含子源于原始的 II 类内含子, 对 I 类内含子与 II 类及核 mRNA 内含子是否有进化关系则有不同的意见^[9, 12]。

作为一种“非必需成分”, 内含子在基因和基因组中地位的进化可以从其含量的变化趋势体现出来, 而后者又是与基因组大小的进化有密切联系^[1]。在核(类核)基因组、线粒体基因组和叶绿体基因组中, 都有“小基因组”型与“大基因组”型两种结构形式。这 3 大类基因组的“小基因组”型总的来说都以基因组相对较小(对于同一类基因组而言), 基因排列紧密, 不含或只含少量内含子与重复序列为特点; 而“大基因组”型则以基因组相对较大或很大, 含有大量的非编码序列(基因间隔、内含子和重复序列等)为特点^[1, 6]。类核基因组、动物型线粒体基因组、大多数藻类和高等植物的叶绿体基因组就分别是 3 大类基因组中所对应那一类的“小基因组”型; 而核基因组、植物型线粒体基因组、以伞藻属为代表的叶绿体基因组则分别属于所对应基因组类别中的“大基因组”型^[1, 6]。“小基因组”型与“大基因组”型的存在说明 3 大类基因组在大小和组成方面的进化都有两种与其结构特点相应的途径, 即“小基因组”途径和“大基因组”途径^[1]。在原始基因组中就有内含子和重复序列的情况下, 那么, “小基因组”进化途径就是以丢失、残留内含子和重复序列为特点, 包括属于“小基因组”型的各种基因组的进化; “大基因组”进化途径则以保留、发展内含子和重复序列为特点, 包括属于“大基因组”型的各种基因组的进化^[1]。因此, 内含子在基因组进化过程中, 既可以被淘汰掉, 又可以得到扩增和发展, 至于是哪一种情况则取决于基因组在大小方面的进化趋势。

另一方面, 由于核基因组本身大小的变化范围很大, 在基因组组成上也有类似“小基因组”型与“大基因组”型的分化。例如, 酵母的核基因组是真核生物中最小的, 只有约 1 000 万碱基对, 以至与个别原核生物的基因组大小差不多^[13]。在结构组成上, 它们也具有类似“小基因组”型的特点^[14]。至于绝大部分真核生物, 它们的核基因组都很大, 在结构组成上自然就属于“大基因组”型了。因此, 在绝大部分真核生物的核基因组沿“大基因组”途径进化的同时, 也有小部分真核生物的核基因组是基本上按照“小基因组”进化途径的模式发展的, 这就使得真核基因组中的内含子也可以有两种进化趋势。

3 内含子的进化与内含子中的基因的进化

随着内含子在基因组中的种类和含量的进化, 内含子中的基因也会相应地进化. 一方面, 内含子本身的进化有可能为内含子中的基因提供进一步发展的机会, 或者使内含子中的基因特异化. 核 mRNA 内含子出现后, 其中一些就很适合 snoRNA 基因的存在; 现代的 I, II 类内含子则比较适合含有一些编码与其活动有关的蛋白质的基因. 另一方面, 内含子含量在基因组进化中的变化使得内含子中的基因必须适应或服从这种变化.

在内含子“先起源”的模式中, 若基因组沿“大基因组”途径进化, 则内含子中的基因原则上可以继续保留并进一步发展, 也有可能由于序列的分化而失去编码能力或由于移位而成为独立的基因; 但是, 若基因组沿“小基因组”途径进化, 则内含子中的基因就要面临被淘汰或被移位的命运. 在后一种途径中, 如果原来内含子中的基因除了与内含子的活动有关外, 并没有什么其他功能, 那么它们就会随着内含子的消失而消失, 从而一般只能在残留的内含子中继续存在. I, II 类内含子中的某些基因在“小基因组”进化途径中就很有可能属于这种情况. 相反, 如果原来内含子中的基因是生命活动所必需或是具有一些与内含子本身无关的功能, 那么它们就必须在内含子消失之前被移位到基因组的其他区域, 成为独立的基因或以其他方式继续存在. 对于 snoRNA 基因, 情况很可能就是这样.

在高等真核生物中, snoRNA 基因只有少数种类是独立基因或成基因簇, 而大部分都在内含子中^[3, 11, 15, 16]. 酵母的 snoRNA 基因却大部分是独立基因^[3]或成基因簇(待发表), 只有小部分在内含子中^[3]. 很明显, 这些状况是与核基因组的相对大小密切相关的. 有理由设想: snoRNA 基因最初是起源于原始内含子中的片段并作为内含子中的基因而存在. 其后, snoRNA 基因在“大基因组”进化途径的生物中仍主要以内含子中的基因的形式存在, 并可通过扩增、在内含子之间的移位等来进一步发展. 而在酵母这些相当于“小基因组”进化途径的生物中, 大部分 snoRNA 基因则随着内含子的消失而移位, 成为独立的基因. snoRNA 基因这种进化模式基本上能说明它们在现代真核生物中的各种状况.

在内含子“先起源”和 snoRNA 基因源于内含子模式中, snoRNA 基因似乎应该在所有生物中都存在. 而实际上, 在原核生物中至今还没有发现 snoRNA 或类似的基因. 原核生物是沿“小基因组”途径进化的, 其内含子中的基因在进化中或者随内含子的消失而消失, 或者移位成为独立的基因. 原核生物的内含子如果曾经含有 snoRNA 或类似的基因, 那这些基因为什么不像酵母那样进化成为独立的基因呢? 在原核生物中没有发现 snoRNA 基因的事实是否与上述关于 snoRNA 基因起源和进化的模式相矛盾? 这样的问题必须从现代原核生物中 rRNA 的情况加以分析. 现代原核生物中 rRNA 的加工和修饰机制与真核生物的并不一样^[11, 17], 有关的过程完全不需要 snoRNA 的参与. 因此, 不管原核生物中 rRNA 的加工及修饰机制是原始或是次生的, 原始基因组中 snoRNA (基因) 的祖先在原核生物的进化中都会成为一类“非必需成分”, 从而在“小基因组”进化途径中就会被逐渐淘汰^[11]. 由此看来, 在原核生物基因组中是有可能找到类似 snoRNA 基因的进化遗迹的.

本文论述的只是内含子中的基因进化的一些较有可能的情况, 必须进一步研究各种与内含子和内含子中的基因有关的结构及其相互关系, 才能得出更明确的结论.

参考文献:

[1] 张尚宏. 论基因组的起源和进化. 中山大学学报(自然科学版), 1996, 35(5): 96~101

- [2] MOUNT S, HENIKOFF S. Nested genes take flight. *Curr Biol*, 1993, 3: 372~374
- [3] MAXWELL E S, FOURNIER M J. The small nucleolar RNA. *Annu Rev Biochem*, 1995, 35: 897~934
- [4] CAVALIER-SMITH T. Selfish DNA and the origin of introns. *Nature*, 1985, 315: 283~284
- [5] DOOLITTLE W F. The origin and function of intervening sequences in DNA: a review. *Amer Nat*, 1987, 130: 915~928
- [6] ZHANG S H. The origin and evolution of repeated sequences and introns. *Speculat Sci Technol*, 1998, 21: 7~16
- [7] BLAKE C C F. Exons and the evolution of proteins. *Int Rev Cytol*, 1985, 93: 149~185
- [8] SEIDEL H M, POPMILIANO D L, KNOWLES J R. Exons as microgenes? *Science*, 1992, 257: 1489~1490
- [9] MICHEL F, FERAT J L. Structure and activities of group II introns. *Annu Rev Biochem*, 1995, 64: 435~461
- [10] LYKKE-ANDERSEN J, AAGAARD C, SEMIOENKOV M, et al. Archaeal introns: splicing, intercellular mobility and evolution. *Trends Biochem Sci*, 1997, 22: 326~331
- [11] BACHELLERIE J P, CAVAILLE J. Guiding ribose methylation of rRNA. *Trends Biochem Sci*, 1997, 22: 257~261
- [12] CECH T R. The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell*, 1986, 44: 207~210
- [13] CAVALIER-SMITH T. Eukaryote gene number, non-coding DNA and genome size. In: CAVALIER-SMITH T, ed. *The Evolution of Genome Size*. Chichester: John Wiley and Sons Ltd, 1985. 69~103
- [14] GOFFEAU A, AERT R, AGOSTINI-CATBONE M L, et al. The yeast genome directory. *Nature*, 1997, 387 (Suppl.): 1~105
- [15] 屈良鹄, 钟翎, 施苏华, 等. 水稻 Hsp70 基因第一个内含子编码 2 种核仁小分子 RNA. *自然科学进展*, 1997, 7 (2): 234~241
- [16] LEADER D J, CLARK G P, WATTERS J, et al. Clusters of multiple different small nucleolar RNA genes in plants are expressed as and processed from polycistronic pre-snoRNAs. *EMBO J*, 1997, 16: 5742~5751
- [17] MORRISSEY J P, TOLLERVEY D. Birth of the snoRNPs: the evolution of RNase MRP and the eukaryotic pre-rRNA-processing system. *Trends Biochem Sci*, 1995, 20: 78~82

Genome Evolution and the Evolution of Genes in Introns

*
ZHANG Shanghong, QU Lianghu

Abstract: This paper is mainly on the evolution of genes in introns during overall genome evolution. First, the relationship between the ways by which introns originated and the kinds of genes in introns was discussed. And the evolutionary characteristics of genes in introns, especially snoRNA genes, were studied in terms of the tendency for the variation of the amounts of introns during the evolution of genomes in size and in composition. With this consideration, some possible pathways for the evolution of genes in introns are proposed.

Keywords: genome evolution, intron, genes in introns