

doi:10.3969/j.issn.1672-5565.2013.03.03

基因外显子连接序列与相应内含子序列的相互作用

张强,李宏*,赵小庆 许士华

(内蒙古大学物理科学与技术学院,内蒙古 呼和浩特 010021)

摘要: 剪接后的内含子与相应 mRNA 序列的相互作用在基因表达调控过程中起着非常重要的作用。基于 27 个物种的核糖核蛋白基因序列,采用 Smith-Waterman 局域比对方法得到外显子连接序列与相应内含子序列的最佳匹配片段,分析了外显子连接序列上的匹配频率分布和匹配片段的序列特征。发现一些低等真核生物 EJC 结合区域的匹配频率明显低于其它区域,所有物种 EJC 结合区域的序列构成呈现出相对低的结构序。最佳匹配片段的平均长度和配对率分布与 siRNA 和 miRNA 的结合特征相同。推测 EJC 和内含子在与外显子序列结合的过程中存在相互竞争和相互协作的关系,内含子中部序列在基因表达调控过程中起着重要的作用。

关键词: 核糖核蛋白基因,外显子连接序列,内含子,局域比对,外显子连接复合体

中图分类号: Q61 **文献标识码:** A **文章编号:** 1672-5565(2013)-03-172-09

Interactions between Exon-exon sequence and its corresponding intron sequence of genes

ZHANG Qiang, LI Hong*, ZHAO Xiao-qing, XU Shi-hua

(School of Physical Science and Technology, Inner Mongolia University, Hohhot 010021, China)

Abstract: There is a kind of interaction between mRNA sequences and corresponding introns after splicing, which plays a significant role in gene expression regulation process. Based on the ribosomal protein genes of 27 genomes, Smith-Waterman local alignment method were used to obtain the optimal matching segments between exon-exon sequences and their corresponding intron sequences. The distributions of matching frequency on exon-exon sequences and the sequence characters of optimal matching segments were analyzed. We found that matching frequency on EJC bonding region are lower obviously for some lower eukaryotes. The sequence composition of EJC bonding region shows lower composed order. The average length and the matching rate distribution of the optimal matching elements are the same as siRNA and miRNA. We think that it exist a kind of competitive and cooperative relations between EJC and introns in the process of bonding on exons. The intermediate section of intron sequences may play very important roles in gene regulation and gene expression.

Key words: Ribosomal Protein Gene, Exon-Exon Sequence, Introns, Local Alignment, Exon Junction Complex

近年来,内含子对基因表达的影响越来越受重视。研究表明,内含子中不仅含有涉及基因转录、mRNA 加工(尤其是可变剪接)等众多基因表达调控元件,也含有微 RNA (microRNA)、核仁小 RNA (snoRNA)等多种非编码 RNA^[1-2]。已经发现,剪切后的内含子与相应编码序列之间存在相互作用,

这种相互作用在 mRNA 运输和基因表达调控过程中起到非常重要的作用^[3]。这种相互作用机制从另外一个角度给出了一种探索 mRNA 上蛋白结合因子的思路。

外显子连接复合体(exon-exon junction complex, EJC)是结合在 mRNA 序列上一类非常重要的

收稿日期:2013-03-08;修回日期:2013-04-24。

基金项目:国家自然科学基金资助项目(31260219)。

作者简介:张强,男,内蒙古大学 2010 级硕士研究生;E-mail:zhangqiang829@163.com。

* 通讯作者:李宏,博士生导师,电话:(0471)4996004;E-mail:ndlihong@imu.edu.cn。

结合因子,但序列特征分析很难找到 EJC 结合位点的信息。探索 EJC 结合位点的方法是人们寻找其它蛋白结合位点的重要问题之一。EJC 在 RNA 加工中具有重要作用^[4-6]。1994 年,Cheng 首次报道 mRNA 剪接时遗留下来的外显子-外显子连接点可作为顺式作用元件通过一种蛋白复合物在意义介导的 mRNA 降解(NMD)中发挥作用,降解含有提前出现终止密码子的 mRNA^[7-8]。在哺乳动物细胞中 EJC 与剪接一起形成,在剪接连接点上游 24 核苷酸处与成熟的 mRNA 紧密结合^[9-14]。在哺乳动物中,EJC 参与一系列的生命过程,包括监控 mRNA 的运转、定位、指导 mRNA 的准确翻译以及参与 NMD 过程等^[15-16],EJC 同时在这些不同过程中出现,说明 EJC 在基因表达调控过程中的作用不同寻常。

研究表明,内含子与 EJC 之间存在相互作用。内含子可以通过与 EJC 之间的相互作用来增加 mRNA 与核糖体之间的相互作用,以此来提高翻译速度^[17]。因此,通过分析内含子与相应外显子连接序列的相互作用来探讨 EJC 的结合区域或揭示两者之间的关系具有重要意义。基于此思路我们对 27 个真核生物的核糖核蛋白基因序列,分析外显子连接序列与剪切后的内含子最佳匹配强度分布,探讨外显子连接处的特征,研究匹配频率分布与 EJC 结合区域的关系,同时对可能的其它蛋白结合区域做一个推测。

1 数据与方法

1.1 基因序列数据

选取 27 个基因组中核糖核蛋白基因作为研究

样本。核糖核蛋白基因在一个物种内和不同物种间相对其它基因而言具有较好的功能和序列保守性,另外它们的内含子数目不多,内含子长度差异相对较小。故选取核糖核蛋白基因作为研究样本是合适的。核糖核蛋白基因序列取自 Ribosomal Protein Gene Database (RPG) (<http://ribosome.med.miyazaki-u.ac.jp/>),所选基因信息见表 1。分析表明,核糖核蛋白基因外显子的最可几长度约为 100bp(见图 1),而最佳匹配区域的最可几长度在 15bp 左右(见图 2)。我们研究的主要目的是外显子结合处和 EJC 结合区域的特征,实验指出 EJC 多位于外显子上游约 24bp 处。基于上述考虑,选取外显子和外显子交接处向前后各取 50 bp 构建长度为 100bp 的外显子连接序列,它即可以显示可能的 EJC 区域,又不受边界的影响,同时使得样本序列数目最大化,提高统计显著性。最后得到外显子连接序列共计 3 245 条。然后将每个外显子连接序列与该连接处的内含子进行比对。

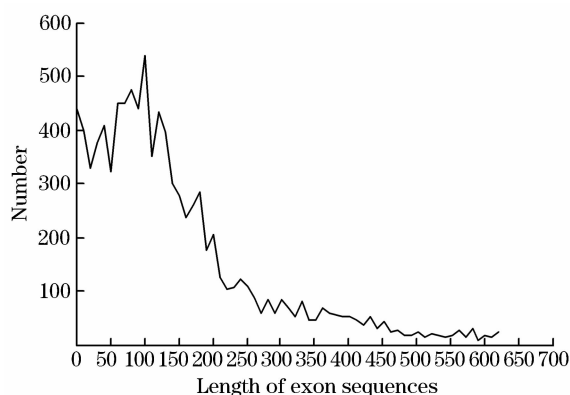


图 1 27 个物种核糖核蛋白基因外显子序列的长度分布

Fig. 1 The length distribution of exons for ribosomal protein genes of 27 genomes

表 1 27 个物种核糖核蛋白基因数据
Table 1 Ribosomal Protein genes of 27 genomes

物种名	分类	基因数	内含子	外显子	外显子连接序列
<i>Caenorhabditis elegans</i> 线虫	A1	84	173	257	143
<i>Coprinus cinereu</i> 鬼伞菌	A2	80	261	341	77
<i>Ustilago maydis</i> 玉蜀黍黑粉菌	A3	79	78	157	22
<i>Arabidopsis thaliana</i> 拟南芥	A4	226	630	856	344
<i>Schizosaccharomyces pombe</i> 粟酒裂殖酵母	A5	141	87	228	19
<i>Rhizopus oryzae</i> 米根霉	A6	304	579	883	128
<i>Dictyostelium discoideum</i> 盘基网柄菌	A7	78	105	183	17
<i>Cryptococcus neoformans</i> 新型隐球菌	A8	78	263	341	80
<i>Anopheles gambiae</i> 蚊子	A9	68	138	206	76

续(表 1)

物种名	分类	基因数	内含子	外显子	外显子连接序列
Neurospora crassa 红面包菌	A10	80	214	294	62
Homo sapiens 人类	B1	79	364	443	172
Rattus norvegicus 褐家鼠	B2	79	351	430	158
Mus musculus 小家鼠	B3	79	350	429	159
Apis mellifera 蜜蜂	B4	79	241	320	137
Oryza sativa 水稻	B5	243	843	1 086	445
Chlamydomonas reinhardtii 绿藻	B6	79	249	328	92
Ciona intestinalis 玻璃海鞘	B7	77	248	325	142
Brugia malayi 马来丝虫	B8	77	199	276	128
Nematostella vectensis 海葵	B9	80	269	349	101
Populus trichocarpa 毛果杨	B10	78	254	332	151
Fugu rubripes 河豚	B11	80	322	402	127
Volvox carteri 团藻	B12	80	269	349	136
Drosophila melanogaster 果蝇	B13	84	171	255	91
Plasmodium falciparum 恶性疟原虫	C1	73	88	161	37
Magnaporthe grisea 稻瘟病菌	C2	79	199	278	63
Fusarium graminearum 禾谷镰刀菌	C3	80	195	275	58
Toxoplasma gondii 刚地弓形虫	C4	79	152	231	80
总计	27	2 723	7 292	10 015	3 245

1.2 比对方

将外显子连接处的内含子序列转化成其互补序列,然后用 Smith-Waterman 局域比对软件 <http://mobyle.pasteur.fr/cgi-bin> 与该外显子连接序列进行局域比对分析,找出一个最佳匹配区域。在局域比对计算中采用 Ednafull 矩阵,选取的参数如下:每个空隙罚分 (Gap penalty) 为 50.0,空隙中每延伸一个碱基位点罚分 (Extend penalty) 为 5.0,得到两者仅有一个可信的最佳局域配对区域,也是两条序列相互作用几率最大的区域。如图 2 所示。

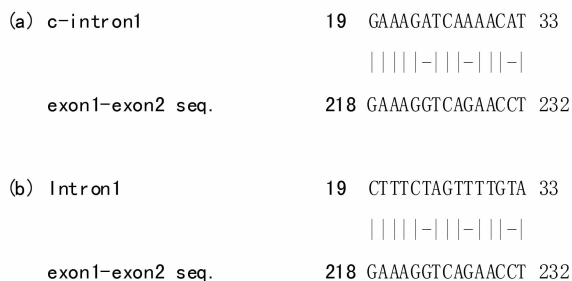


图 2 Smith-Waterman 局域比示例 (a) 将内含子转化成其互补序列后的 Smith-Waterman 局域比对结果, (b) 真实的最佳匹配区域

Fig. 2 Sketch map for Smith-Waterman local alignment (a) Smith-Waterman local alignment between the complementary sequence of intron and exon-exon sequence, (b) Authentic optimal matching region between exon-exon sequence and intron sequence

1.3 最佳匹配频数分布

对 m 条外显子连接序列,设 N_{i1} 、 N_{i2} 分别为第 i 条外显子连接序列上最佳匹配区域的起始位点和终止位点 ($i = 1, 2, \dots, m - 1, m$)。将 m 条序列对齐,每个位点上定义赋值函数 f_{ij} ,见公式(1)。

$$f_{ij} = \begin{cases} 1 & N_{i1} \leq j \leq N_{i2} \\ 0 & j < N_{i1} \text{ or } j > N_{i2} \end{cases} \quad (1)$$

f_{ij} 为第 i 条序列中第 j 位点的赋值函数 ($j = 1, 2, \dots, 99, 100$)。表示最佳匹配区域内的位点赋值为 1,最佳匹配区域外的位点赋 0 值。

定义第 j 位点的匹配频率 F 值 (Matching frequency) 如下:

$$F = 1/m \sum_{i=1}^m f_{ij} \quad (2)$$

F 反映了外显子连接序列第 j 个碱基与内含子序列相互作用的强度。 F 越高,说明该位点参与匹配的概率越大。由此得到匹配频率 F 在外显子连接序列上的分布。

2 结果分析

2.1 匹配频率 F 的分布特征

对 27 个物种核糖核蛋白基因的外显子连接序列上匹配频率的分布进行分析,发现匹配频率分布对各个物种是有区别的。在外显子结合处两端匹配频率的分布没有明显的差异,但多数物种具有较高的匹配频率值或作用强度。远离外显子结合点的分

布大体可分为三种情况。第一种有 10 个物种显示出了有 EJC 痕迹,这些物种记为 A1-A10;第二种有 13 个物种未显示出 EJC 位置,记为 B1-B13;第三种有 4 个物种未显示 EJC 位置但下游有保守的低匹配区域,记为 C1-C4。下面对这 3 类情况进行分析。

2.1.1 有 EJC 痕迹的物种

发现有 10 个物种在外显子交接处上游 20bp 左右存在 F 值很低的区域(见图 3),即这个区域内含子与编码序列的作用强度较低,而这个区域正是实

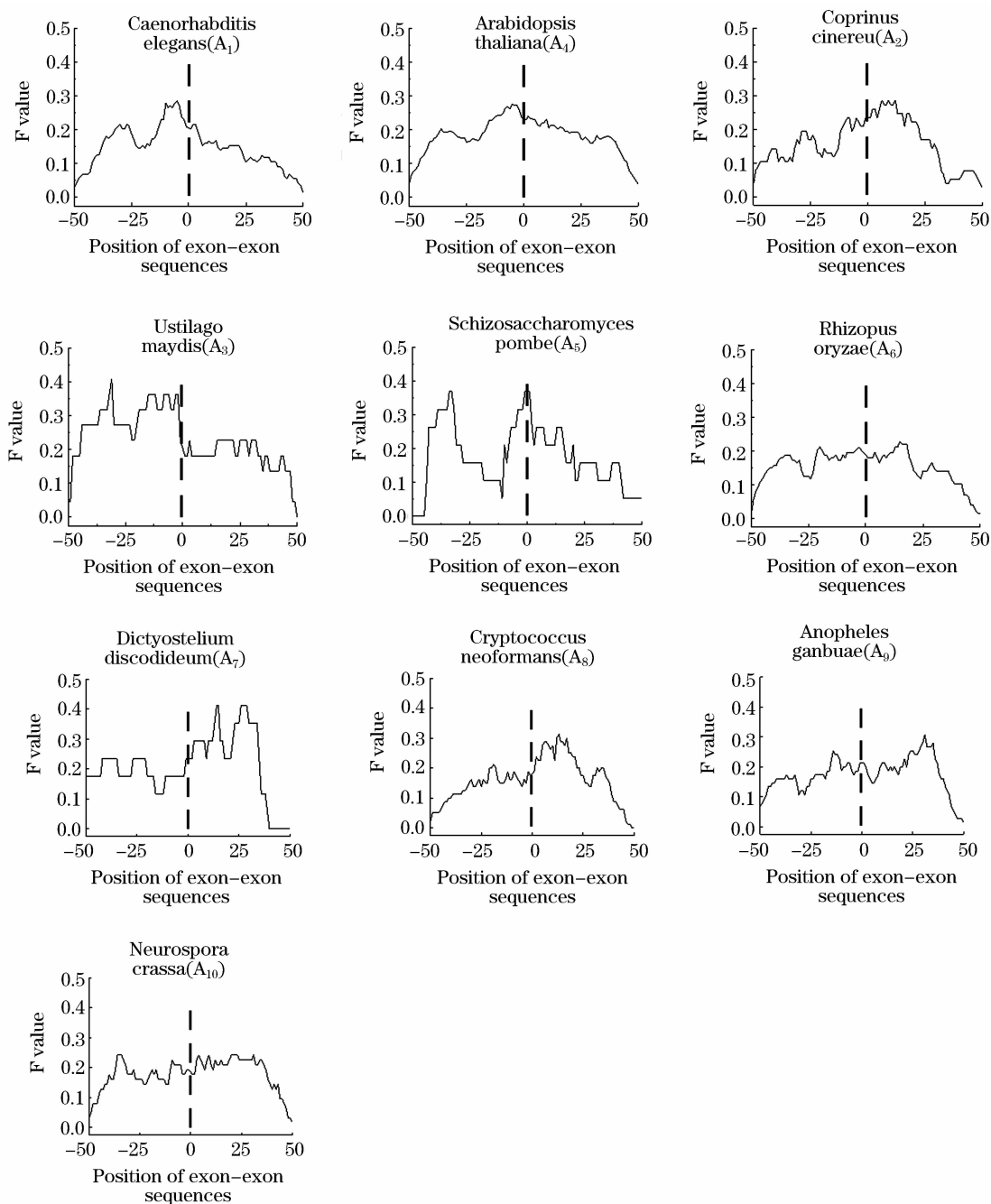


图 3 A 类物种外显子连接序列和它们之间的内含子序列局域比对的相对频率分布

Fig. 3 Distributions of matching frequency between exon-exon sequences and intron sequences which are located in the exon-exon sequence for the genomes of class A

验报道的 EJC 结合区域。这表明外显子的这个区域存在与 EJC 结合的序列结构信息。这个区域的 F 值并非为零值,还有一些内含子作用在这个区域上。这表明内含子和 EJC 与编码序列的作用存在相互协作和竞争关系。在这 10 个物种的核糖核蛋白基因中,内含子和 EJC 与编码序列的作用存在较明显的相互协作的因素,编码序列上内含子和 EJC 的作用位点有一定的回避性,编码序列本身就存在 EJC 结合区域的结构域。但从统计上看,还有一些内含子作用在 EJC 的结合区域上,表明了内含子和 EJC 在编码序列的结合上还存在相互竞争的因素,哪一方能否取得优势地位是由基因所处的环境决定的。分析这 10 个物种发现,它们都是低等真核生物。比如线虫和粟酒裂殖酵母,在它们的核糖核蛋白基因中存在 EJC 结合的序列特征。

2.1.2 未显示 EJC 位置的物种

有 13 个物种在两个外显子连接处均有较高的

匹配频率或作用强度(见图 4),在外显子交接点上游 20bp 左右并没有显示出 EJC 结合的信息,也就是说用这种统计方法显示不出 EJC 的结合区域,因此也无法否认 EJC 结合的事实。我们发现没有 EJC 信息分布的生物多数是较高等的生物,如人类和鼠类。

一些生物(如低等生物)显示了 EJC 的位置信息(见 2.1.1),而另一些生物(如高等生物)没显示出 EJC 的位置信息。我们认为没有显示 EJC 信息的生物在进化和环境压力下其结合位点的序列结构信息被弱化了,EJC 和内含子之间的位置竞争增强了。也就是说,随着生物进化有内含子参与的基因调控网络功能不断加强,而外显子序列本身的结构特征被弱化了。但有一些生物比如线虫和粟酒裂殖酵母,由于特定的选择压力,在其核糖核蛋白基因中仍保留了一定的 EJC 结构信息。

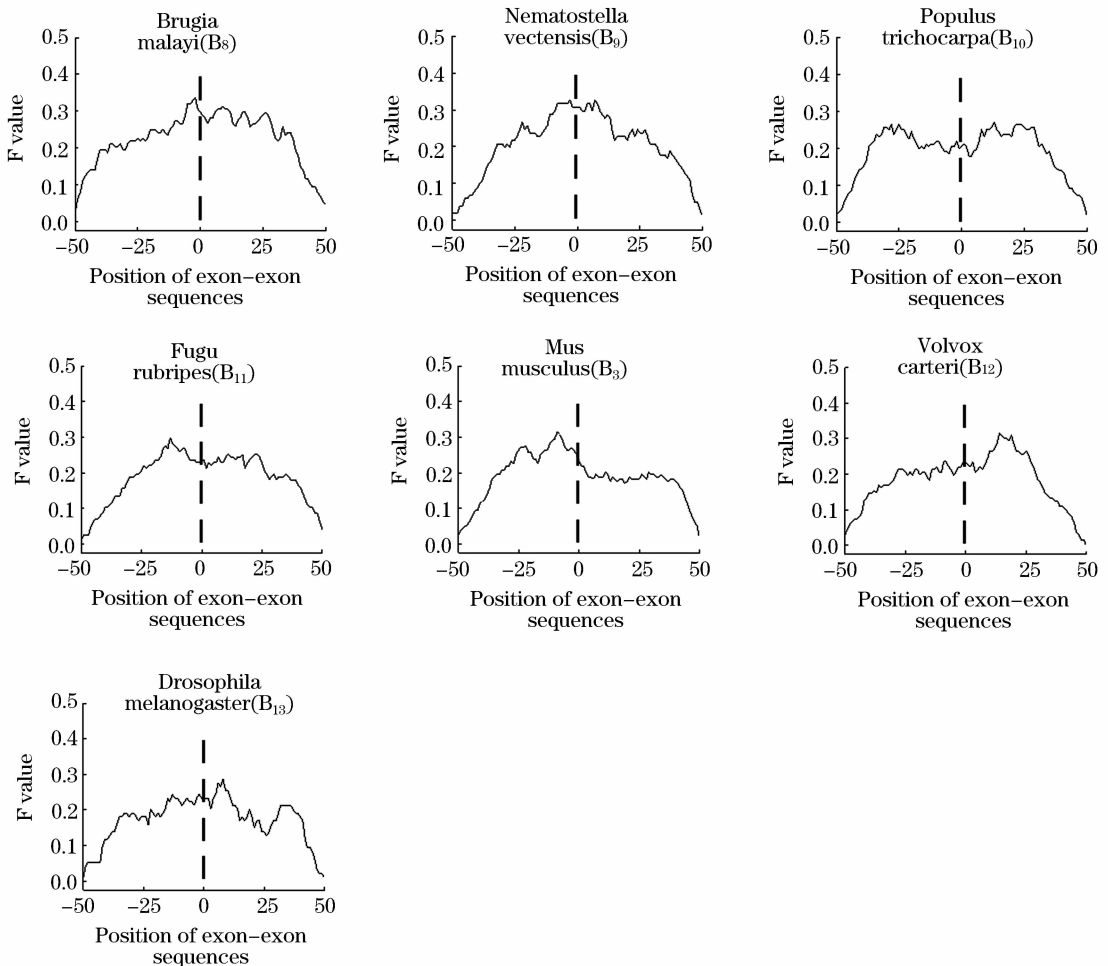


图 4 B 类物种外显子连接序列和它们之间的内含子序列局域比对的相对频率分布

Fig. 4 Distributions of matching frequency between exon-exon sequences and intron sequences which are located in the exon-exon sequence for the genomes of class B

2.1.3 有其它特征的物种

在所分析的物种中有4个物种在外显子交接位点下游0~20bp内存在较低的匹配频率分布(见图5)。这4个物种均是较低等的真核生物。按照2.1.1节和2.1.2节的分析,我们推测这一低匹配区

域可能是某个结合蛋白或蛋白复合体的结合区域。在这些物种的序列上显现了蛋白结合的序列特征。或者说,恰好在这些物种的基因序列上这个结合蛋白与内含子之间的协作关系明显,这些迹象给我们提供了一个探测结合蛋白的窗口。

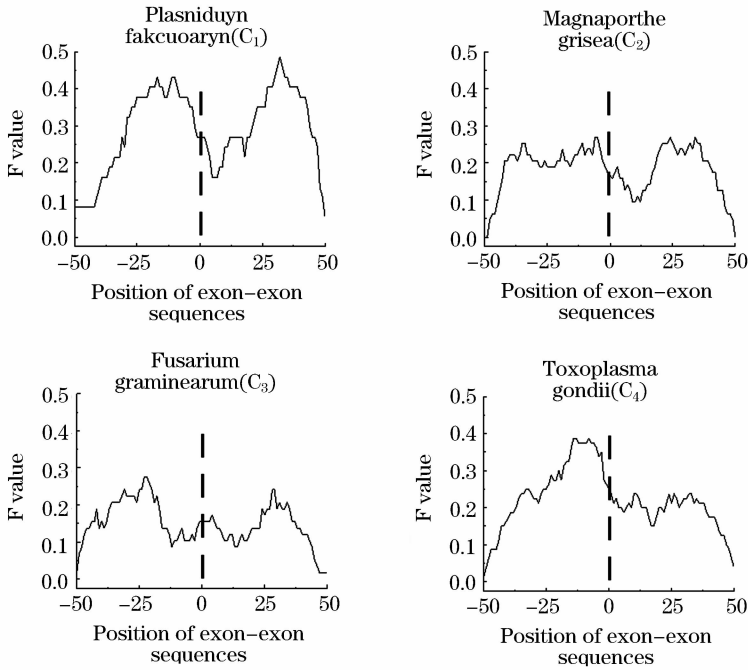


图5 C类物种外显子连接序列和它们之间的内含子序列局域比对的相对频率分布

Fig. 5 Distributions of matching frequency between exon-exon sequences and intron sequences which are located in the exon-exon sequence for the genomes of class C

2.2 内含子长度对 EJC 位置的影响

在图3中线虫、鬼伞菌和粟酒裂殖酵母的外显子连接序列中有较明显的 EJC 结合位置信息,将这3个物种合在一起作为一个样本,对其 EJC 结合区域进行分析。由于内含子长度差异很大,根据 Haligan^[18]。研究结论,以80bp为界将内含子分为长内含子和短内含子两类(见图6(a))。以长、短两类内含子分别与相应外显子连接序列比对,发现长内

含子与外显子连接序列的匹配频率分布中在 EJC 区域的匹配频率更低,或 EJC 位置更加明显(见图6(b)),而短内含子的分布中 EJC 的位置不明显(见图6(c))。这表明长内含子与外显子连接序列结合时有意避开 EJC 区域,短内含子的这种作用趋势较弱。可见内含子的长度对 EJC 的结合是有影响的。我们认为,长内含子是与 EJC 位置协作的主体,短内含子是与 EJC 位置竞争的主体。

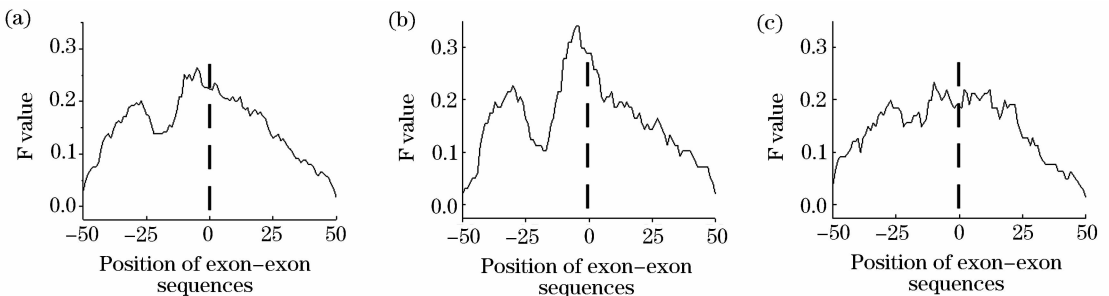


图6 A1、A2 和 A3 物种的匹配频率分布与长短内含子的关系 (a) 所有内含子, (b) 长内含子, (c) 短内含子

Fig. 6 Distributions of matching frequency for long introns and short introns in genome A1, A2 and A3

(a) All intron. (b) Long intron. (c) Short intron

2.3 外显子连接点上游序列特征分析

在外显子连接位点上游的匹配频率分布中,有些生物显示出 EJC 结合信号,有些生物则没有。我们认为所有生物在 EJC 结合区域仍存在相应的序列特征,随着进化一些生物(特别是高等真核生物)的这个序列特征被弱化了,而内含子与 EJC 的竞争关系逐渐占主导地位,使得 EJC 结合区域的序列信息被掩盖了。我们用二阶信息冗余 D_2 值的分布来检验 EJC 结合区域的序列特征。

二阶(紧邻碱基关联)信息冗余定义为:

$$D_2 = \sum_{i,j} p_{ij} \log_2(p_{ij}/p_i p_j) \quad (3)$$

其中 p_i 或 p_j 为外显子连接序列中碱基 i 或 j 出现的概率($i, j = A, C, G, U, \dots$), p_{ij} 为二核苷 ij 在序列中出现的联合概率^[19]。一般 $D_2 \geq 0$, 对于长度为 N 的随机序列在 99% 置信度下, $D_2 = 8.15/N$ 。如果实际序列的 $D_2 \geq 8.15/N$, 则表明序列是非随机的, 碱基之间存在关联。 D_2 值实际上反映了序列的结

构序的强弱。

具体做法是:将各个物种的核糖核蛋白基因外显子连接位点上游序列分成 3 个区间。靠近外显子连接点的 15bp 长的序列为一区,一区上游 10bp 的序列记为二区,这个区域正是实验给出的 EJC 结合区域,二区上游的序列记为三区,有 25bp 长。分别将 A 类和 B 类物种各个区的序列连接起来组成 3 个序列,然后计算每个序列的二阶信息冗余 D_2 值,结果见表 2。

对所分析的 A 类和 B 类物种而言,三个区域的序列结构具有共同特征。一区序列的 D_2 值最高,二区序列的 D_2 值基本上是最底的,三区序列的 D_2 值介于两者之间。表明外显子 3' 端区域的序列具有更强的结构序,相比而言 EJC 结合区域的结构序最低或序列的随机性更强些。这一结果显示出外显子序列具有可区分的内部结构。

表 2 外显子连接序列的序列特征

Table 2 Sequence characters of different exon regions

外显子分区	D_2 值($A_1 - A_3$)	D_2 值($A_1 - A_{10}$)	D_2 值($B_1 - B_3$)	D_2 值($B_1 - B_{13}$)
-26 bp ~ -50 bp	0.036 (0.001 0)	0.028 (0.000 3)	0.030 (0.000 7)	0.019 (0.000 20)
-16 bp ~ -25 bp	0.030 (0.003 0)	0.025 (0.000 8)	0.027 (0.002 0)	0.020 (0.000 40)
-1 bp ~ -15 bp	0.044 (0.002 0)	0.035 (0.000 6)	0.045 (0.001 0)	0.033 (0.000 30)
-1 bp ~ -50 bp	0.038 (0.000 7)	0.034 (0.000 2)	0.035 (0.000 5)	0.026 (0.000 08)

注:括号中的数据指相同长度的随机序列在 99% 置信度下的 D_2 值。

Notes: The datas in parantheses refer to D_2 value of the same random sequence of length 99% confidence level.

A 类物种在 EJC 结合区域表现出低匹配频率,而对应的外显子序列具有低的结构序,特别是 $A_1 - A_3$ 物种,这一特征更加明显。但在 B 类物种中,尽管在 EJC 结合区域没有明显的低匹配频率出现,但在此区域里外显子序列仍显示了较低的结构序。这说明了两方面的问题:一是多数生物在外显子上游 20bp 附近存在 EJC 结合的序列特征。相对于外显子其他区域,EJC 结合区域的序列组成更随机一些。这一序列特征是造成基于序列预测 EJC 结合位点困难的根本原因之一。二是结论支持了我们提出的内含子和 EJC 之间在与外显子结合的过程中存在协调和竞争的关系。在一些高等真核生物中,竞争的因素更加明显,EJC 结合的信息被内含子竞争作用所掩盖。

2.4 最佳匹配长度和配对率分析

我们用最佳匹配片段来刻划内含子与外显子序列的相互作用情况。这些最佳匹配片段是否会发生是值得研究的,在实验还无法验证的情况下,理论分析是必要的。一个最佳匹配片段中有两个参数特

征:一是最佳匹配长度,二是片段的配对率。配对率定义为:设最佳匹配长度为 m bp,若有 k 个碱基完全配对,即 C-G 和 U-A 出现,则这个最佳匹配片段的配对率为 k/m 。对 27 个物种所有核糖核蛋白基因外显子连接序列上与相应内含子序列的最佳匹配长度和配对率分布进行分析,结果见图 7 和图 8。

在最佳匹配长度分布中大于 50bp 和小于 5bp 的匹配很少出现,最佳匹配长度的平均值是 19.3bp。我们另一个工作给出了拟南芥和果蝇全部基因的平均最佳匹配长度分别为 20.2bp 和 19.6bp。它们均在 20bp 左右。最佳匹配片段的配对率主要分布在 55% ~ 95%,多数最佳匹配片段的配对率集中在 65% ~ 80%,另外还有少部分片段的配对率为 100%。内含子与编码序列之间存在的最佳匹配片段也是 RNA-RNA 相互作用,目前人们对这类相互作用还缺少认识。我们知道目前有两类主要的 RNA-RNA 相互作用,它们是 RNA 干涉和 RNA 抑制。承担这两个功能的非编码序列分别是 siRNA 和 miRNA。siRNA 的长度在 21 ~ 25bp,由 Dicer 加

工而成双链 RNA,通过与靶 mRNA 严格互补来指导 mRNA 沉默^[20]。miRNA 的长度在 18 ~ 25bp,由 Dicer 加工而成单链 RNA,通过与目标 mRNA 不同程度的互补来抑制靶 mRNA 转录和表达^[21],统计表明 miRNA 与靶 mRNA 之间的匹配率在 65% ~ 95%^[22],miRNA 被认为在调控发育过程中有重要作用。

对比内含子与外显子的最佳匹配长度和配对率的范围,我们发现它们与 miRNA 和 siRNA 的特性惊人的一致,这是值得深思的。我们认为剪切后的内含子与相应 mRNA 的相互作用是存在的,而这一现象是普遍存在的。内含子内的一些区域正是通过与 mRNA 序列不同程度的互补来参与基因调控和表达过程的。在配对率分布中,绝大多数的片段是非严格配对的,只有少数片段的配对率为 100%。表明内含子序列中即包含了与 miRNA 特性一致的序列也包含了与 siRNA 特性一致的序列。从这个角度来看,真核基因本身就包含了应有的基因调控原件,作为一类非编码序列内含子不仅完成剪接和可变剪接的功能,还是基因表达调控的重要执行者之一。作为一类 RNA - RNA 相互作用的具有功能的非编码序列集合,内含子序列的这一特征应该受到关注。

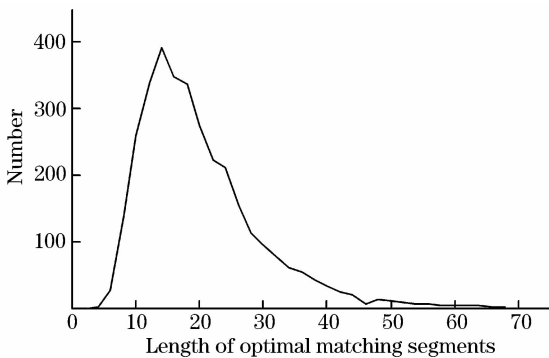


图7 最佳匹配片段的长度分布

Fig. 7 The length distribution of optimal matching segments

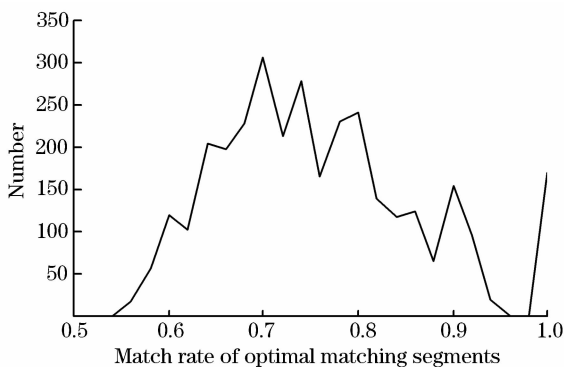


图8 最佳匹配片段的配对率分布

Fig. 8 Distribution of matching rates for all optimal matching segments

3 结论与讨论

对匹配频率分布的分析表明,在外显子连接序列上确实存在许多与内含子结合的最佳匹配区域,也存在一些低匹配区域。一些生物的核糖核蛋白基因外显子连接处上游 20bp 左右存在一个低匹配区,这与 EJC 结合区域一致,有 EJC 痕迹的生物多数是低等生物,有一些生物特别是高等真核生物没有显示出 EJC 位置的信息。还有一些生物在外显子结合点下游 0 ~ 20bp 范围内存在低匹配区域,猜测这些区域很可能是某些蛋白因子的结合区域。在有较明显 EJC 结合区域的生物中发现长内含子与外显子连接序列的相互作用中有避开 EJC 区域的现象,而短内含子的相互作用没有避开 EJC 区域的趋势。对外显子的序列结构进行分析,发现无论有没有低匹配频率分布出现,在 EJC 结合区域,它们的序列均具有低的结构序或序列构成更加随机。这些结论说明蛋白结合因子和内含子这两类调控元件在与 mRNA 序列结合的过程中存在相互竞争和相互协作的关系。EJC 能否结合到相应的位置不仅取决于外显子上存在的序列结构信息,还取决于它们与内含子的竞争是否占有优势。能否取得优势地位是由基因所处的环境决定的,这个过程恰好反映了基因的调控机制。两类调控元件协作和竞争的信息可以通过研究内含子与相应编码序列之间的相互作用和外显子的序列特征展现出来。

通过分析所有最佳匹配区域的长度和配对率分布,发现最佳匹配区域的平均长度均在 20bp 左右,配对率主要分布在 65% ~ 80%,少数片段的配对率达到 100%。它们与 siRNA 和 miRNA 这两类非编码序列的结合特征是一样的。猜测它们功能应该相似,作为一类非编码序列,剪切后的内含子通过与 mRNA 序列不同程度的互补来参与基因调控和表达过程的,内含子应该成为第三类 RNA-RNA 相互作用的具有功能的非编码序列集合。

对真核生物内含子与相应的 mRNA 序列的相互作用尚需进一步的研究,这里尽管揭示了许多有趣的问题,因所取样本的规模有限,一些结论的可靠性尚需通过扩大样本规模来做进一步的检验。今后将改进局域比对过程能够进行集约化比对减少耗时,使之对全基因组的基因分析成为可能。

参考文献 (References)

- [1] John S. Mattick, Michael J. Gagen. The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms [J]. *Molecular Biology and Evolution*, 2001, 18(9):1611-1630.

- [2] Ajit Nott, Shlomo H. Muslin, Melissa J. Moore. A quantitative analysis of intron effects on mammalian gene expression [J]. *RNA*, 2003, 9(5):607-617.
- [3] Xiaoqing Zhao, Hong Li, Tonglaga Bao. Analysis on the interaction between post-spliced introns and corresponding protein coding sequences in ribosomal protein genes [J]. *Journal of Theoretical Biology*, 2013, 328:33-42.
- [4] Ideue Takashi, Sasaki Yasunori T. F., Hagiwara Masatoshi, Hirose Tetsuro. Introns play an essential role in splicing-dependent formation of the exon junction complex [J]. *Genes & Development*, 2007, 21(16):1993-1998.
- [5] Heather L. Wiegand, Shihua Lu, Bryan R. Cullen. Exon junction complexes mediate the enhancing effect of splicing on mRNA expression [J]. *Proceedings of The National Academy of Sciences of The United States of America*, 2003, 100(20):11327-11332.
- [6] Ajit Nott, Herve Le Hir, and Melissa J. Moore. Splicing enhances translation in mammalian cells: an additional function of the exon junction complex [J]. *Genes & Development*, 2004, 18(2):210-222.
- [7] Michael D Diem, Chia C Chan, Ihab Younis, Gideon Dreyfuss. PYM binds the cytoplasmic exon-junction complex and ribosomes to enhance translation of spliced mRNAs [J]. *Nature Structural & Molecular Biology*, 2007, 14(12):173-179.
- [8] Giorgi Corinna, Yeo Gene W, Stone Martha E, Kata Donaid B, Burge Christopher, Turrigiano Gina, Moore Melissa J. The EJC factor eIF4AIII modulates synaptic strength and neuronal protein expression [J]. *Cell*, 2007, 130(1):179-191.
- [9] Soule Jonathan, Alme Maria, Myrum Craig, Schubert Manja, Kanhema Tambudzai, Bramham Clive R. Balancing arc synthesis, mRNA decay, and proteasomal degradation: maximal protein expression triggered by rapid eye movement sleep-like bursts of muscarinic cholinergic receptor stimulation [J]. *Journal of Biological Chemistry*, 2012, 287(26):22354-22366.
- [10] Herve Le Hir, Elisa Izaurrealde, Lynne E Maquat, Melissa J. Moorel. The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions [J]. *Embo Journal*, 2000, 19(24):6860-6869.
- [11] Lynne E. Maquat. Nonsense-mediated mRNA decay: Splicing, translation and mRNP dynamics [J]. *Nature Reviews Molecular Cell Biology*, 2004, 5(2):89-99.
- [12] Thomas O Tange, Ajit Nott, Melissa J Moore. The ever-increasing complexities of the exon junction complex [J]. *Current Opinion in Cell Biology*, 2004, 16(3):279-284.
- [13] Herve Le Hir, Ajit Nott, Melissa J. Moore. How introns influence and enhance eukaryotic gene expression [J]. *Trends in Biochemical Sciences*, 2003, 28(4):215-220.
- [14] Vienna L. Reichert, Herve Le Hir, Melissa S. Jurica, Melissa J. Moore. 5' exon interactions within the human spliceosome establish a framework for exon junction complex structure and assembly [J]. *Genes & Development*, 2002, 16:2778-2791.
- [15] Lionel Ballut, Brice Marchadier, Aurelie Baguet, Catherine Tomasetto, Bertrand Seraphin, Herve Le Hir. The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity [J]. *Nature Structural & Molecular Biology*, 2005, 12(10):861-869.
- [16] Xiaoju Max Ma, Sang-oh Yoon, Richardson, Kristina Julich, John Blenis. SKAR links pre-mRNA splicing to mTOR/S6K1-mediated enhanced translation efficiency of spliced mRNAs [J]. *Cell*, 2008, 133(2):303-313.
- [17] Zuo Zhang and Adrian R. Krainer. Splicing remodels messenger ribonucleoprotein architecture via eIF4A3 dependent and independent recruitment of exon junction complex components [J]. *Proceedings of The National Academy of Sciences of The United States of America*, 2007, 104(28):11574-11579.
- [18] Daniel L. Halligan, Peter D. Keightley. Ubiquitous selective constraints in the Drosophila genome revealed by a genome-wide interspecies comparison [J]. *Genome Research*, 2006, 16(7):875-884.
- [19] Liaofu Luo. *Theoretic-Physical Approach to Molecular Biology* [M]. Shanghai: Shanghai Scientific & Technical Publishers, 2004. 69-77.
- [20] Thomas A. Volpe, Catheriner Kidner, Ira M. Hall, Grace Teng, Shivl. S. Grewal, Robert A. Martienssen. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi [J]. *Science*, 2002, 297(5588):1833-1837.
- [21] Lee P. Lim, Nelson C. Lau, Philip Garrett-Engele, Andrew Grimson, Janell M. Schelter, John Castle, David P. Bartel, Peter S. Linsley, Jason M. Johnson. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs [J]. *Nature*, 2005, 433(7027):769-773.
- [22] Jingang Cui, Yang Zhao, Sethi Prahlad K, Yanyan Li, Mahta Ali, Culicchia Frank, Lukiw Walter J. Micro-RNA-128 (miRNA-128) down-regulation in glioblastoma targets ARPS (ANGPTL6), Bmi-1 and E2F-3a, key regulators of brain cell proliferation [J]. *Journal of Neurooncology*, 2010, 98(3):279-304.