

DOI:10.12113/201911004

mRNA 序列与相应内含子序列匹配的普适性分析

张强¹, 赵小庆³, 薄素玲², 曹艳娟¹, 苏文霞¹

(1. 内蒙古农业大学 理学院, 呼和浩特 010018; 2. 内蒙古医科大学 计算机信息学院, 呼和浩特 010110; 3. 内蒙古自治区农牧业科学院, 呼和浩特 010031)

摘要: 剪切后的内含子在基因表达和调控过程中发挥重要作用, 由此在成熟 mRNA 与相应的内含子之间也存在相互作用, 并且二者协同进化。为了验证这一理论, 以 13 个物种的基因组作为研究样本, 凭借 Smith-Waterman 的局域比对方法, 最终得到在成熟 mRNA 与其内含子序列之间的最佳匹配片段, 同时在 mRNA 序列上显示出最佳匹配强度的分布区。然后对最佳匹配片段的长度、配对率的分布这两项参数进行分析之后, 发现最佳匹配片段的这两个参数的特征, 与 siRNA 和 miRNA 的结合特征是相近的; 在 mRNA 序列上, 得出 UTR 区与内含子相互作用强, GC 含量低的片段偏好结合到 3' UTR 区, 相反 GC 含量高的片段更倾向与 5' UTR 区发生相互作用。因此, 最佳匹配片段的序列特征符合 RNA-RNA 相互作用规律, 可以把内含子看成是一种具有基因表达调控功能的序列。以上研究对于进一步探讨内含子的功能和进化具有重大意义。

关键词: mRNA 序列; 内含子; 局域比对; 最佳匹配片段; GC 含量

中图分类号: Q61 **文献标志码:** A **文章编号:** 1672-5565(2020)03-149-08

Conservative analysis on matching between mRNA and corresponding intron

ZHANG Qiang¹, ZHAO Xiaoqing³, BO Suling², CAO Yanjuan¹, SU Wenxia¹

(1. College of Science, Inner Mongolia Agriculture University, Hohhot 010018, China;
2. College of Computer and Information Inner Mongolia Medical University, Hohhot 010110, China;
3. Inner Mongolia Academy of Agricultural & Animal Husbandry Sciences, Huhhot 010031, China)

Abstract: Introns after splicing play an important role in expression and regulation. Therefore, it is found that there is interaction between mature mRNA and its corresponding intron, and the relation is coevolution. In order to verify this theory, the genome sequences of 13 species were taken as research objects, and the optimal matched segments between mature mRNA and its intron sequences were obtained by using the local matching method named Smith-Waterman. Meanwhile, the distribution of the optimal matched segments on mRNA was obtained. Analysis results show that the average length and matching rate distribution of the optimal matched segments were similar to the combined characteristics of siRNA and miRNA. The UTR regions on mRNA had strong interaction with introns. Segments with low GC content were more likely to interact with 3' UTR region, while those with high GC content were prone to interact with 5' UTR region. Hence, the sequence characteristics of the optimal matched segments are coincident with the general regulation of RNA-RNA interaction, and introns can be regarded as functional segments of regulating gene expression. The conclusion is of constructive significance for further study the function and evolution of introns.

Keywords: mRNA; Intron; Local alignment; Optimal matched segments; GC content

近几年, 诸多研究都开始重视内含子对基因表达的影响^[1]。大量研究表明, 内含子是一类具有生

物学功能的序列。许多基因表达调控元件, 例如基因转录和 mRNA 加工(尤其是可变剪接), 都属于内

收稿日期: 2019-11-21; 修回日期: 2020-01-17.

基金项目: 内蒙古农业大学博士科研启动基金(No. NDYB2016-13); 国家自然科学基金(No. 11747084); 内蒙古自然科学基金(No. 2018BS03001)和内蒙古农业大学基础学科科研启动基金(No. JC2017003).

作者简介: 张强, 男, 讲师, 博士, 研究方向: 生物物理学. E-mail: zhangqiang829@163.com.

含子序列的一部分。各种非编码 RNA, 例如 microRNA 和 snoRNA, 也属于内含子^[2]。对内含子序列而言, 它的丢失和获得对非编码 RNA 变异和基因重组有影响, 这是关系到真核基因进化的主要方面^[3-7]。许多疾病的发生是由于内含子的突变造成的^[8-9]。内含子两端和中间序列的突变, 都是因为激活隐性切割位点进而导致的疾病。虽然部分基因的表达并不需要内含子序列的参与甚至不存在内含子, 然而内含子在许多情况下能够最大限度地增强转基因生物的基因表达^[10-11]。内含子序列已经成为对转基因生物中外源基因表达进行改善不可或缺的一部分^[12-15]。以上所有情况间接表明内含子序列的存在与否对基因表达有显著差别。同时, 在 pre-mRNA 上, 所有剪接过程的发生, 都可以看作内含子和外显子序列相互作用后, 所产生的必然结果。序列匹配是体现上述相互作用最基本的形式^[16-19]。那么必定只能由更加复杂的序列结构才能组成内含子序列并执行多种生物功能。比如, 内含子通过与 mRNA 的序列匹配, 调节结合蛋白因子与 mRNA 的相互作用, 才能进行调节^[20-22]。在 mRNA 序列上存在功能区域, 例如翻译起始与终止位点, 还存在与内含子匹配的特殊形式, 这些形式对基因的表达调控具有至关重要的作用^[23]。因此, 寻找 mRNA 序列与相应内含子的最佳匹配区域是解决问题的思路。

基于此思路, 以 13 个生物基因的编码序列作为研究样本, 目的在于找到 mRNA 序列上, 能够与内含子序列存在相互作用关系的最佳匹配片段, 进而分析这些片段的序列特征, 讨论在 mRNA 序列上匹配频率的分布特点, 研究该分布在不同物种中的进化规律。

1 数据与方法

1.1 基因序列数据

13 个物种 1 号染色体编码基因序列取自 The Exon-Intron Database (EID) 数据库。13 个物种中包括 4 个植物, 2 个无脊椎动物和 7 个有脊椎动物。在选取样本 (见表 1) 中去掉了含有重复元件和 ncRNA 等已知非剪接功能的基因。

1.2 比对方法

mRNA 序列与相应内含子序列存在碱基互补匹配片段, 该片段能够反映它们之间的相互特点。因此它们之间的相互作用可由最佳匹配片段表征出来。首先把内含子进行转化, 得到它的互补序列, 其次通过软件 Smith-Waterman, 对其进行局部相似性比对, 再次比对互补序列与其所对应的 mRNA 序列, 进而找出最佳相似片段, 最后通过转化得到两者

最佳匹配片段。

1.3 最佳匹配频数分布

定义 1: 匹配打分函数

对给定序列的每个碱基位点赋予一个分值, 在最佳匹配区域内, 则碱基位点赋值为 1; 如果不在最佳匹配区域内, 则赋碱基位点值为 0。定义匹配打分函数如下:

$$f_{ij} = \begin{cases} 1 & N_{is} \leq j \leq N_{ie} \\ 0 & j < N_{is} \text{ 或 } j > N_{ie} \end{cases} \quad (1)$$

公式中, i 代表第 i 条序列, j 是第 i 条序列中, 第 j 个碱基位点 ($j = 1, 2, \dots, L_i$); 那么 L_i 是第 i 条序列的长度, 则 N_{is} 和 N_{ie} 分别是第 i 条序列上最佳匹配片段的起始和终止碱基位点。这样一条序列就转换成由 0 和 1 组成的数字串, 1 代表最佳匹配片段位置。

表 1 13 种真核生物基因

Table 1 Protein coding genes of 13 eukaryotes

物种名	基因数	内含子数
人类	2 068	20 750
小鼠	1 138	11 346
大鼠	1 315	11 974
狗	769	8 307
鸡	788	7 017
牛	381	3 443
斑马鱼	530	4 558
线虫	233	1 850
果蝇	2 711	13 785
拟南芥	4 083	24 673
葡萄	485	2 485
水稻	2 026	11 709
杨树	692	3 381

定义 2: 匹配频率函数

在所分析的 m 条序列上, 统计第 j 碱基位点上出现 1 的次数, 除以 m 就得到该位点的匹配频率。匹配频率函数 $F(j)$ 定义如下:

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

函数 F 是匹配频率, 它代表的是比对序列的第 j 个碱基位点, 与被比对序列两者的匹配强度, 或者是它们相互作用的强度。由此可以推出, 根据 F 值的大小, 可以对此位点参与最佳匹配的概率进行评价。如果 $F(j)$ 等于 1, 那么就说明, 所有 m 条序列的第 j 个碱基位点, 应该全都位于最佳匹配区域内。由此得到匹配频率 F 值在序列上的分布。

定义 3: 序列长度标准化

实际序列可以是编码序列, 或者是内含子序列, 针对它们长度各异这一情况, 序列的最佳匹配片段要进行相对位置分布比较, 为使其比较起来更方便,

则需要把序列长度标准化,得到长度为 L 的目标序列。长度标准化采用如下方法:

$$n_{ij} = \begin{cases} [L * N_{ij}/L_i] & L * N_{ij}/L_i \text{ is integer} \\ [L * N_{ij}/L_i] + 1 & L * N_{ij}/L_i \text{ is non-integer} \end{cases} \quad (3)$$

分析以上公式,将其中第 i 条比对序列长度设置为 L_i ;那么位于第 i 条比对序列上,第 j 个碱基位点是 N_{ij} ;在长度标准化之后,与第 i 条比对序列相对应,第 j 个相对位点是 N_{ij} 。高斯取整函数用方括号代表,要求是,取一个实数的整数部分。如此,就在 m 条长度各不相同的序列经过转化之后,得到 L 长度的目标序列。实际上,序列长度标准化的过程就是,按照一定的比例将长度各异的多条序列缩放成等长的序列,便于更好地对最佳匹配区域的相对位置分布进行分析。

2 结果

分析最佳匹配片段的序列构成,有助于深入研究内含子与相应 mRNA 序列之间相互作用。通过对 mRNA 序列和相应内含子序列之间的比对,得到各物种 mRNA 序列上全部最佳匹配片段集合。要对所有最佳匹配片断进行序列特征分析,那么就要用到最佳匹配片段中的 3 个特征参数,它们分别为最佳匹配片段长度、最佳配对率及其中的 GC 含量。

2.1 最佳匹配长度分布

通过 mRNA 序列与相应内含子序列之间的比

对分析,得到了 13 个物种 mRNA 序列上的最佳匹配长度分布(见图 1)。考虑这 13 类生物的最佳匹配片段长度分布形式是一样的,与 $\alpha > 1$ 的伽马分布相接近。长度大多在 10~80 bp 之内分布。无脊椎动物、植物,它们的长度平均为 20 bp 左右;而有脊椎动物,长度大约平均为 30 bp。我们发现,对于高等真核生物,其最佳匹配片段的长度,平均起来要长于低等真核生物。这就是说生物进化的同时,最佳匹配片段长度,亦随之增加。那么也就是说,物种愈复杂,内含子与其相应 mRNA 序列二者之间的相互作用模式,也会愈复杂。

2.2 最佳匹配片段的配对率

对配对率给出如下定义:把 L 设定为最佳匹配片段长度,如果 k 个碱基可以进行完全配对,也就是存在 C-G 和 U-A,那么, k/L 即是该最佳匹配片段的配对率。最佳匹配片段的配对率分布图(见图 2)。不同的物种的配对率分布基本无差别,大多分布在 60%~85%,甚至极少部分最佳匹配片段的配对率能达到 100%,这些片段的长度都很短,低于 15 bp。配对率分布曲线上有一些小的峰值出现,比较明显的峰值出现在配对率为 0.75, 0.8, 0.85 和 0.95 处,这些峰值分布对各个物种都是一样的,说明了最佳匹配区域序列匹配是有规律的。最佳匹配片段配对率分布在各个物种中具有普适性,反映了配对率的保守性。这表明不同进化水平的物种其 mRNA 与相应内含子的相互作用或序列匹配方式遵循同一个匹配机制。

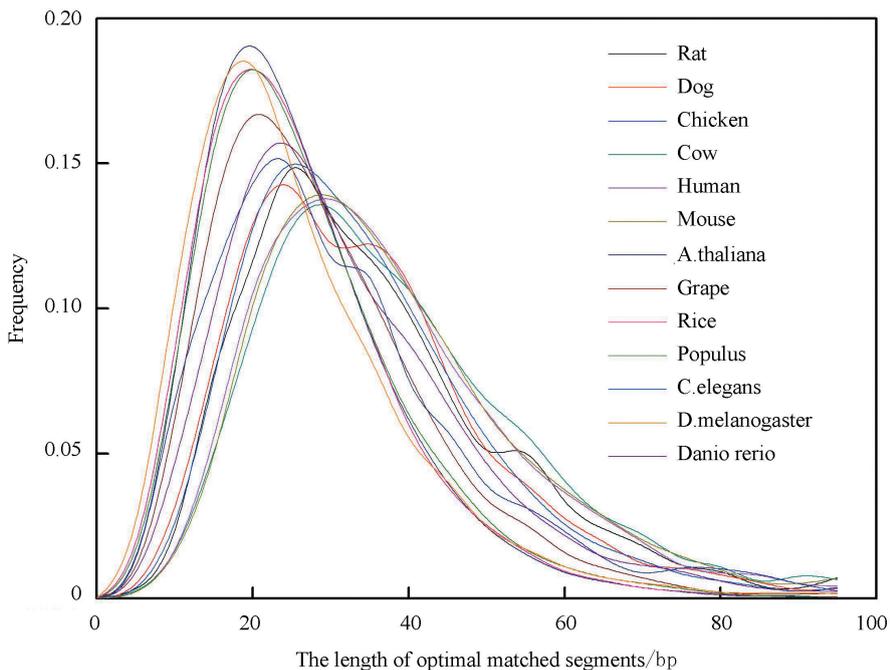


图 1 最佳匹配片段长度分布

Fig.1 Length distributions of optimal matched segments

2.3 最佳匹配片段 GC 含量分布

对 mRNA 的序列特征分析时,发现,GC 含量是非常重要的参数。13 类生物基因的 mRNA 序列上的最佳匹配片段的 GC 含量分布情况(见图 3)。四种植物最佳匹配片段的 GC 含量分布在一个比较窄的范围(0.2~0.6)之内,最概然分布是 0.35,两个脊椎动物鸡和斑马鱼的分布范围比植物要宽,在 0.1~0.7,最概然 GC 含量是 0.45。线虫的分布与植物相似,果蝇的分布与脊椎动物的分布相似。哺乳动物

最佳匹配区域的 GC 含量分布各不相同,大鼠和小鼠的最概然 GC 含量更大一些(约为0.55),牛的最概然 GC 含量很小为 0.3。但它们的分布范围很宽,主要在 0.1~0.8。人类和狗的分布与其它物种的分布不同,不仅 GC 含量分布更宽,在0.1~0.9,而且最佳匹配区域的 GC 含量分布是双峰分布。第一个峰的最概然 GC 含量值是 0.3,第二个峰的最概然 GC 含量值是 0.7。

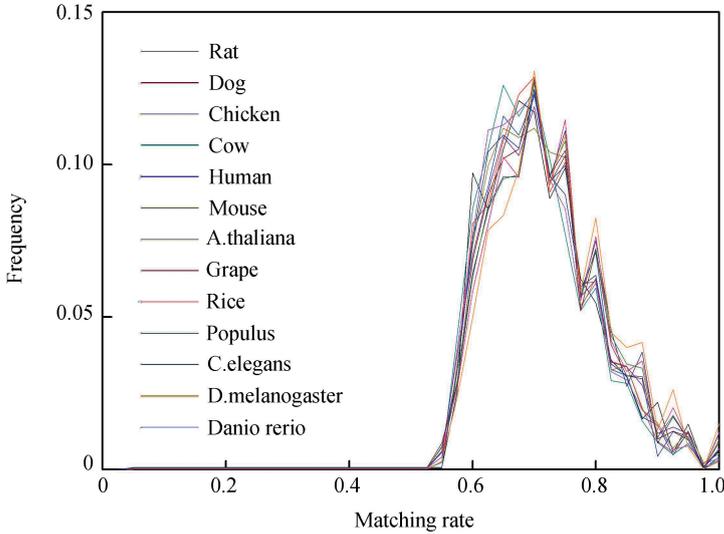


图 2 最佳匹配片段配对率分布

Fig.2 Matching rate distributions of optimal matched segments

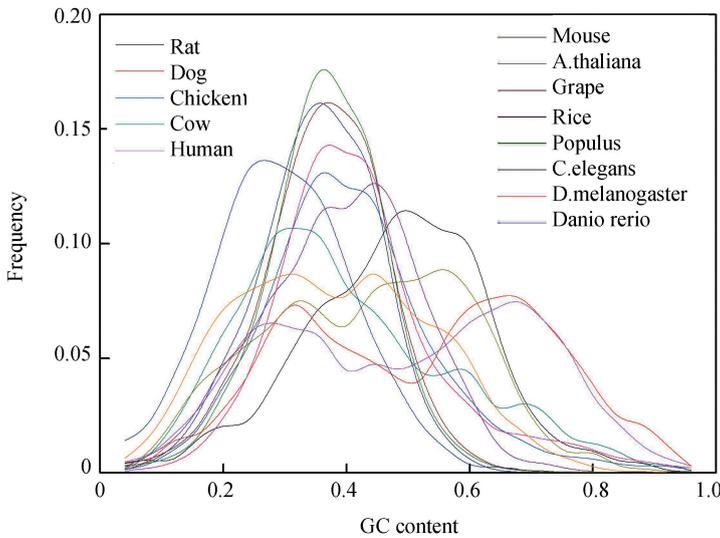


图 3 最佳匹配片段的 GC 含量分布

Fig.3 GC content distribution of optimal matched segments

可见最佳匹配片段的最概然 GC 含量随着物种进化在逐渐增加,这与基因组序列 GC 含量随物种进化而增加的现象是一致的。值得注意的是最佳匹配片段的 GC 含量分布范围很广,在 0.1~0.9 广泛

分布。由此说明最佳匹配片段的发生不受 GC 含量水平的影响,有的 GC 含量高,有的 GC 含量低。

对于大多数低等生物,最佳匹配片段中 GC 含量均偏低,接近内含子的 GC 含量。这表明基因中

mRNA 序列上,最佳匹配多发生在低 GC 区域。而且内含子与 mRNA 之间存在一种弱相互作用,从而形成弱双链结构。然而对高等生物而言,最佳匹配片段中 GC 含量多数偏高,与其外显子中 GC 含量接近。在人类和狗中,最佳匹配片段中 GC 含量分布特征与线虫和小鼠的分布一致。尽管人类与小鼠基因在进化上很接近,但最佳匹配片段中 GC 含量分布却有差别。这说明内含子与 mRNA 序列之间的最佳匹配序列组成在一定程度上反映了表观遗传的进化差异。

2.4 最佳匹配片段在 mRNA 序列上分布特征

针对 13 个物种基因序列,把它们的内含子和对应的成熟 mRNA 放在一起,进行局域比对。在成熟 mRNA 上,出现最佳匹配区域位置。因成熟 mRNA 序列长度不一致,为研究 mRNA 序列上各个相对位置的匹配强度,所以把成熟 mRNA 序列进行长度标

准化,长度标准化到 100 bp。发现成熟 mRNA 上出现相对匹配频率的分布情况(见图 4)。结果如下:(1)所有物种的匹配强度分布趋势一致,在序列 UTR 区(两端)的匹配频率明显高于编码序列(中部),特别是在 3' UTR 区出现极大值分布。(2)无脊椎动物在 3' UTR 区的峰值最大,比如果蝇和线虫的极大值大约是它们极小值的 7 倍(见图 4b),植物的匹配频率的极大值略高于有脊椎动物的峰值。(3)在 mRNA 序列的 5' UTR 区,有 3 个有脊椎动物(人类,狗和老鼠)的匹配 F 值高于它们的 CDS 区(见图 4a),而其它有脊椎物种则低于 CDS 区。单子叶植物(水稻)的 F 值低于双子叶植物的 F 值,也低于 CDS 区的 F 值(见图 4c)。(4)所有物种在 CDS 区匹配强度大小较一致,并且 F 值都较低。从整体上来说,植物物种在 CDS 区的 F 值相对最高,无脊椎动物的 F 值相对最低,而有脊椎动物 F 值介于它们之间。

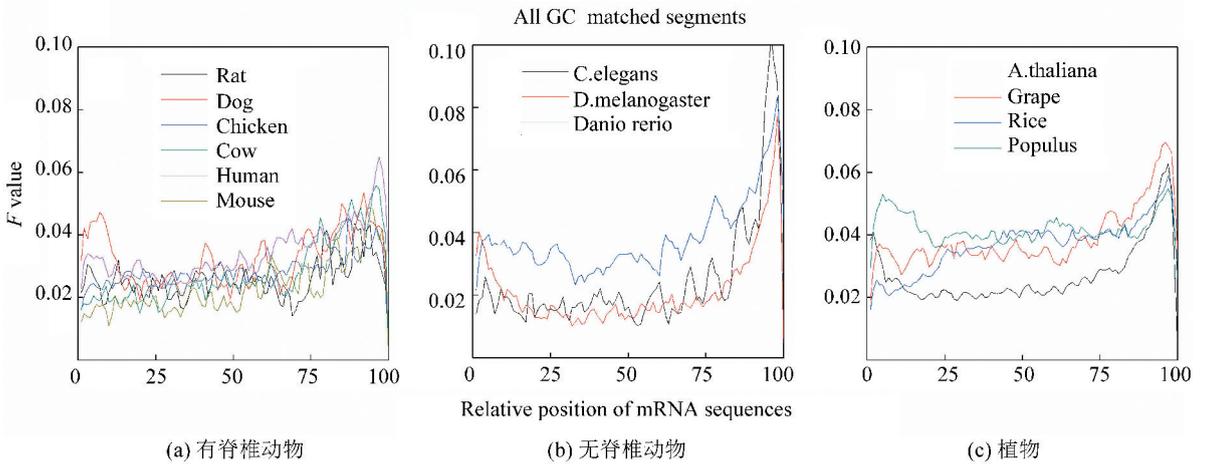


图 4 mRNA 序列和相应的内含子序列局域比对的相对频率分布

Fig.4 Distributions of matching frequency between mRNA sequences and coresponding intron sequences

图 3 中,为找到最佳匹配片段的 GC 含量分布情况,把最佳匹配片段按照 GC 含量的大小分为 2 组:GC 含量高于 0.5 的高 GC 片段和 GC 含量小于 0.3 的低 GC 片段。得到它们在 mRNA 序列的上分布(见图 5、图 6)。分析发现,在 mRNA 序列的 UTR 区,低 GC 片段在其匹配强度更高,而在编码区对应的 F 值则更低。无脊椎动物和植物在 3' UTR 区有显著的峰值(见图 5b, 5c),对有脊椎动物来说,3' UTR 区出现了多个极值的匹配区域(见图 5a)。4 个植物物种和果蝇仍在 mRNA 序列的 5' UTR 区出现极大值分布,且约为极小值的 3 倍(见图 5c),但是其他物种在该区域的匹配频率极低。

对于高 GC 片段在 mRNA 序列上的匹配强度分

布,我们发现,无脊椎动物在 mRNA 序列上的各个位点的 F 值大小相近(见图 6b),然而植物和有脊椎动物在 5' UTR 区出现极大值,其中有脊椎动物出现显著的极大值约为极小值的 5.5 倍(见图 6a, 6c)。另外在 CDS 区和 3' UTR 区的匹配频率一致,且 F 值均较低。

mRNA 序列上存在许多与内含子序列的匹配区域,低 GC 片段偏好与 UTR 区作用,特别是在 mRNA 序列的 3' UTR 区。高 GC 片段偏好与 mRNA 序列的 5' UTR 区匹配。高 GC 片段在 mRNA 序列的 3' UTR 区和编码区分布没有显著的差别,说明在内含子上还存在一些高 GC 含量区域,它们与整个 mRNA 序列都有作用。

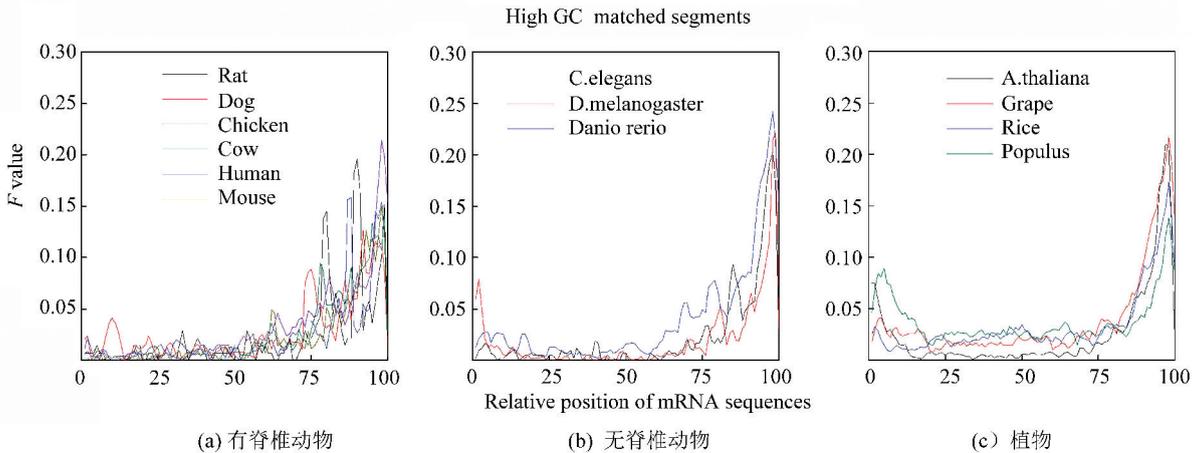


图 5 GC 含量在 0~0.3 内 mRNA 序列匹配频率

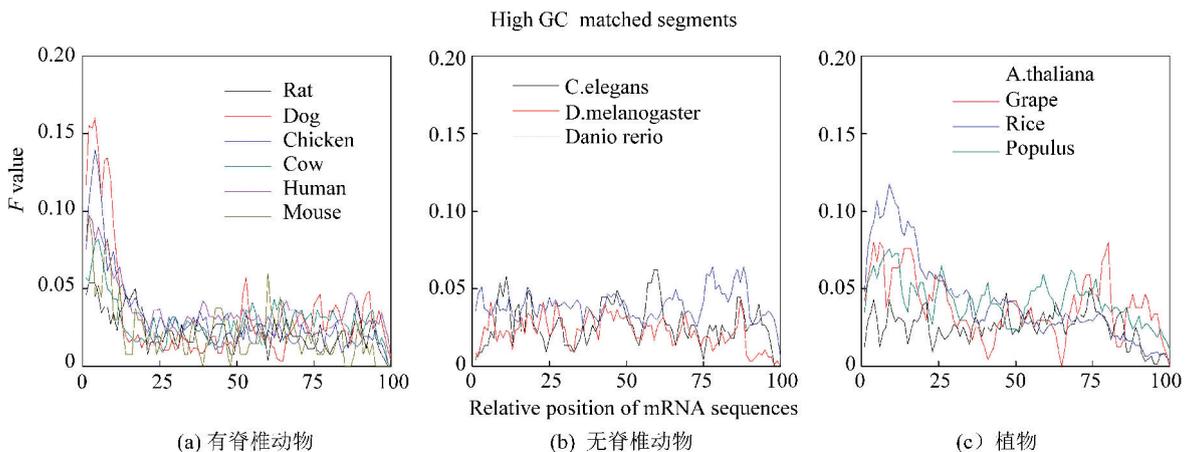
Fig.5 Distribution of F value with GC content between 0.0 and 0.3

图 6 GC 含量大于 0.50 的 mRNA 序列匹配频率

Fig.6 Distribution of F value with GC content greater than 0.5

3 结果与讨论

通过序列比对获得了 13 个物种基因中内含子与 mRNA 二者的最佳匹配片段。通过这些最佳匹配片段序列特征及匹配频率的分布规律,我们发现, siRNA 和 miRNA 的结合特征与得到的最佳匹配片段的平均长度和配对率分布一致; mRNA 上的 UTR 区偏好与内含子相互作用,而 CDS 区域与内含子的匹配程度较低。结论表明内含子与成熟 mRNA 序列存在相互作用。

目前,人们对内含子与 mRNA 序列二者的最佳匹配片段还缺乏一定的了解。小干扰 RNA (siRNA)、微小 RNA (miRNA) 以及 piRNA (Piwi-interacting RNA) 这三个非编码序列承担了 RNA 干扰和 RNA 抑制的功能。一方面, Dicer 可以把长度在 21~25 bp 之间的 siRNA 加工形成双链 RNA, 然后双链 RNA 与靶 mRNA 进行严格地互补再指导 mRNA 沉

默^[24]。另一方面,要形成单链 RNA, Dicer 还会在 18~25 bp 之间,选出 miRNA 进行加工,单链 RNA 与目标 mRNA 这二者之间,在不同程度地进行完成互补,之后对靶 mRNA 转录和表达进行干涉和抑制^[25]。通过统计数据,可以看到 miRNA 与靶 mRNA 的匹配率在 65%~95% 的范围内,那么我们认为 miRNA 在调控发育过程中发挥着关键作用。piRNA 是一类长度为 26~31 bp 单链的小 RNA,大部分集中在 29~30 bp 之间,而且只有通过小 RNA 与属于 PIWI 蛋白的家族成员进行结合, piRNA 才能发挥调控作用。piRNA 的发现对非编码小分子 RNA 的研究开拓出新的领域,因此 Science 把该研究称为 2006 年十大科技进展之一^[26]。

把内含子与 mRNA 之间相互作用的最佳匹配片段的长度以及配对率的数据区间范围与 miRNA 和 siRNA 进行比较,结果显示双方的长度和配对率范围的特性,竟然相当的一致。内含子在经过剪切之后,和对应的 mRNA 之间,相互作用的程度高,无论是 mRNA 与内含

子作用的最佳匹配片段还是 siRNA、miRNA 和 piRNA,从生物选择的功能片段长度到配对率,它们所遵循的生物学机制是一样的。由此推论,在生物基因组序列上,还存在大量的各种形式的类似 miRNA 的功能片段,它们在基因表达调控和实现表观遗传多样性方面起着决定性的作用。在不同程度上,通过和 mRNA 序列之间的互补,一些存在于内含子序列上的区域,才能够对基因调控和表达产生深刻影响。对配对率的分布进行研究之后,我们发现,绝大部分的片段所进行的配对并不严格,然而有极少的片段,严格进行配对,但这些片段的长度很短,远小于 20 bp。按照 RNA 干涉理论,完全匹配对基因表达是致死的,因此在 mRNA 与内含子之间存在的相互作用片段,显然避开了这类致死的匹配片段。从这个角度来看,mRNA-内含子的相互作用理论是合理的。分析认为,对于真核基因这一基因种群,他们自己就具备基因调控所需的原件。因为内含子是非编码序列,所以它可以完成剪接与可变剪接这两个任务,而且能够实践基因表达调控这一重要功能。内含子序列作为具有功能的一类 RNA-RNA 相互作用的非编码序列集合,研究人员需要对此重要特征予以关注。

发现内含子与 mRNA 序列的匹配片段在 mRNA 序列上有分布,其匹配频率,发生在两端非编码序列区域的值较大,在中间编码序列区域,却有较低值。同时,内含子又出现偏好和 mRNA 序列上 3'UTR 区相互作用。在 mRNA 序列的 3'UTR 区,GC 值比较小的片段,拥有相对高的匹配强度,匹配强度却在编码区较低。那么就说明内含子与相应 mRNA 之间存在的相互作用主要是弱键,也就是所谓的 AU 匹配,同时还包括 GC 值大的匹配。但是,大家有可能提出疑问的是,基因序列进化时,UTR 类似于内含子的 GC 值,很清楚的显示出不同于编码序列的 GC 值,这一现象的原因是什么呢?想要出现内含子和 UTR 序列发生较强的相互作用,然后通过该作用对基因进行调控,从而使以上两类序列的进化趋于一致。在 UTR 和编码区上 GC 含量较高的最佳匹配片段分布几乎不存在差别,这也表示一些高 GC 区域存在于内含子序列上。可以看出,要想进一步探究内含子对基因的表达与调控,还必须继续挖掘 mRNA 序列上匹配频率分布的内涵。

综上所述,所有的研究结论,有力地证明了,内含子和 mRNA 之间发生的相互作用,是真实存在的。在维持基因组正常运转的过程中,内含子可能起到了比较关键的调控作用^[27]。然而尚需进一步对真核生物内含子与相应的 mRNA 序列的相互作用进行深入探讨,本文虽然揭示了一些有意思的论

点,然而一些结论尚需通过实验进一步验证。

参考文献(References)

- [1] NOTT A, Meislin S H, Moore M J. A quantitative analysis of intron effects on mammalian gene expression[J]. RNA, 2003, 9 (5): 607 - 617. DOI: 10.1371/journal.ppat.1000799.
- [2] MORELLO L, GIANI S, TROINA F, et al. Testing the IM-Eter on rice introns and other aspects of intron-mediated enhancement of gene expression[J]. Journal of Experimental Botany, 2011, 62(2): 533-544. DOI:10.1093/jxb/erq273.
- [3] YOSHIHAMA M, NGUYEN H D, KENMOCHI N. Intron dynamics in ribosomal protein genes[J]. PLoS one, 2007, 2 (1): e141. DOI:10.1371/journal.pone.0000141.
- [4] ODA T, OHNIWA R L, SUZUKI Y, et al. Evolutionary dynamics of spliceosomal intron revealed by in silico analyses of the P-Type ATPase superfamily genes[J]. Molecular Biology Reports, 2010, 38 (4): 2285 - 2293. DOI: 10.1007/s11033-010-0360-3.
- [5] PARRA G, BRADNAM K, ROSE A B, et al. Comparative and functional analysis of intron-mediated enhancement signals reveals conserved features among plants[J]. Nucleic Acids Research., 2011, 39(13): 5328-5337. DOI:10.1093/nar/gkr043.
- [6] DURET L. Why do genes have introns Recombination might add a new piece to the puzzle[J]. Trends in Genetics, 2001, 17(4): 172 - 175. DOI: 10.1016/s0168-9525(01)02236-3.
- [7] BROCKMOLLER J, CASCORBI L, KERB R, et al. Polymorphisms in xenobiotic conjugation and disease predisposition[J]. Toxicology Letters, 1998, 95(1001): 9-10. DOI: 10.1016/S0378-4274(98)00304-X.
- [8] NORDIN A, LARSSON E, HOLMBERG M. The defective splicing caused by the ISCU intron mutation in patients with myopathy with lactic acidosis is repressed by PTBP1 but can be derepressed by IGF2BP1[J]. Human Mutation, 2012, 33(3): 467-470. DOI:10.1002/humu.22002.
- [9] STOVER D A, VERRELLI B C. Comparative vertebrate evolutionary analyses of type I collagen: Potential of COL1a1 gene structure and intron variation for common bone-related diseases[J]. Molecular Biology and Evolution, 2010, 28 (1): 533-542. DOI:10.1093/molbev/msq221.
- [10] DUNCKER B, DAVIES P, WALKER V. Introns boost transgene expression in Drosophila melanogaster[J]. Molecular and General Genetics MGG, 1997, 254(3): 291-296. DOI:10.1007/s004380050418.
- [11] KO C H, BRENDEL V, TAYLOR R D, et al. U-richness is a defining feature of plant introns and may function as an intron recognition signal in maize[J]. Plant Molecular Biology, 1998, 36 (4): 573 - 583. DOI: 10.1023/A:1005932620374.
- [12] BARTLETT J G, SNAPE J W, HARWOOD W A. Intron-

- mediated enhancement as a method for increasing transgene expression levels in barley [J]. *Plant Biotechnology Journal*, 2009, 7 (9): 856–866. DOI: 10.1111/j.1467-7652.2009.00448.x.
- [13] 薄素玲, 李宏, 张强, 等. 果蝇成熟 mRNA 序列与其相应内含子序列的匹配特征分析[J]. *内蒙古大学学报*, 2018, 49 (6): 612–619. DOI: 10.13484/j.nmgdxzbk.20180608.
- BO Suling, LI Hong, ZHANG Qiang. et al. Analysis of the matching characteristics on the mature mRNA sequences and their introns in *D. melanogaster* [J]. *Journal of Inner Mongolia University*, 2018, 49 (6): 612–619. DOI: 10.13484/j.nmgdxzbk.20180608.
- [14] BLANCO F J, BERNABEU C. Alternative splicing factor or splicing factor-2 plays a key role in intron retention of the endoglin gene during endothelial senescence [J]. *Cell*, 2011, 10 (5): 896–907. DOI: 10.1111/j.1474-9726.2011.00727.x.
- [15] MATTICK J S, GAGEN M J. The evolution of controlled multitasked gene networks: The role of introns and other noncoding RNAs in the development of complex organisms [J]. *Molecular Biology Evolution*, 2001, 18 (9): 1611–1630. DOI: 10.1093/oxfordjournals.molbev.a003951.
- [16] 张强. 成熟 mRNA 与其内含子序列的相互作用机制 [D]. 呼和浩特: 内蒙古大学, 2016, 9–35.
- ZHANG Qiang. The mechanism of interaction between mature mRNA and its intron sequences [D]. Hohhot: Inner Mongolia University, 2016, 9–25.
- [17] 张强, 李宏, 赵小庆, 等. 基因外显子连接序列与相应内含子序列的相互作用 [J]. *生物信息学*, 2013, 11 (3): 172–180. DOI: 10.3969/j.issn.1672-5565.2013.03.03.
- ZHANG Qiang, LI Hong, ZHAO Xiaoqing, et al. Interactions between Exon-exon sequence and its corresponding intron sequence of genes [J]. *Chinese Journal of Bioinformatics*, 2013, 11 (3): 172–180. DOI: 10.3969/j.issn.1672-5565.2013.03.03.
- [18] ZHANG Qiang, LI Hong, ZHAO Xiaoqing, et al. Analysis on the preference for sequence matching between mRNA sequences and the corresponding introns in ribosomal protein genes [J]. *Journal of Theoretical Biology*, 2016, 392: 113–121. DOI: 10.1016/j.jtbi.2015.12.003.
- [19] ZHANG Qing, LI Hong, ZHAO Xiaoqing, et al. The evolution mechanism of intron length [J]. *Genomics*, 2016, 108: 47–55. DOI: 10.1016/j.ygeno.2016.07.004.
- [20] ZHAO Xiaoqing, LI Hong, BAO Tonglaga. 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. DOI: 10.1016/j.jtbi.2013.03.002.
- [21] 赵小庆, 李宏, 包通拉嘎. 线虫核糖核蛋白基因内含子与相应编码序列的相互作用 [J]. *生物化学与生物物理进展*, 2010, 37: 1006–1015. DOI: 10.3724/SP.J.1206.2010.00186.
- ZHAO Xiaoqing, LI Hong, BAO Tonglaga. Interactions between introns and corresponding protein coding sequences of ribosomal protein genes in *C. elegans* [J]. *Progress in Biochemistry and Biophysics*, 2010, 37: 1006–1015. DOI: 10.3724/SP.J.1206.2010.00186.
- [22] ZHANG Qiang, LI Hong, ZHAO Xiaoqing, et al. Distribution bias of the sequence matching between exons and introns in exon joint and EJC binding region in *C. elegans* [J]. *Journal of Theoretical Biology*, 2015, 364: 296–304. DOI: 10.1016/j.jtbi.2014.09.009.
- [23] BO Suling, LI Hong, ZHANG Qiang, et al. Potential relations between post-spliced introns and mature mRNAs in the *Caenorhabditis elegans* genome [J]. *Journal of Theoretical Biology*, 2019, 467: 7–14. DOI: 10.1016/j.jtbi.2019.01.031.
- [24] SOULE J, ALME M, MYRUM C. 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 Biology Chemistry*, 2012, 287: 22354–22366. DOI: 10.1074/jbc.M112.376491.
- [25] WU L, FAN J, BELASCO J G. MicroRNAs direct rapid deadenylation of mRNA [J]. *PNAS*, 2006, 103 (11): 4034–4039. DOI: 10.1073/pnas.0510928103.
- [26] SAI LAKSHMI S, AGRAWAL S. piRNABank: A web resource on classified and clustered Piwi interacting RNAs [J]. *Nucleic Acids Research*, 2008, 36: 173–177. DOI: 10.1093/nar/gkm696.
- [27] 张强. 外显子连接序列与连接处内含子序列的相互作用 [D]. 呼和浩特: 内蒙古大学, 2013, 1–8.
- ZHANG Qiang. Interaction between exon-exon sequences and intron sequences [D]. Hohhot: Inner Mongolia University, 2013, 1–8.

[责任编辑: 吴永英]