

小米 *psbA* 基因 5' 末端非编码区的结构特征

黄 瑶* 吴乃虎**

(中国科学院发育生物学研究所 北京 100080)

摘 要 以小米 (*Setaria italica*) 为材料, 克隆了含有叶绿体 *psbA* 基因的 2.2kb EcoR I 片段, 测定了该基因 5' 末端非编码区的核苷酸序列。序列分析显示 *psbA* 基因 5' 末端非编码区存在着与原核类似的启动子结构, 其“-10”区的序列为 TATACT, 与原核生物“-10”共有基序 (Consensus motif) 仅相差一个核苷酸; 其“-35”区的序列为 TTGACA, 与原核生物“-35”共有基序完全相同。另外, 在“-10”区和“-35”区之间还存在着一个类似真核启动子结构的“TATATA”保守序列。这些结果表明小米 *psbA* 基因的启动子既具有原核的特征又具有真核的特征。小米 *psbA* 基因的 mRNA 前导序列长 87bp, 与高粱完全一致, 而比水稻多出了“CTATTTT”7个核苷酸, 比小麦、大麦和黑麦多出了“TTTT”4个核苷酸。因此推测在禾本科的 C_3 和 C_4 植物之间, *psbA* 基因 mRNA 前导序列区的差异可能具有普遍性。计算机分析结果显示, 以上 6 种植物的 *psbA* 基因 mRNA 前导序列区内均能形成小的茎环结构, 而且这段“CTATTTT”额外序列恰好位于茎环结构中, 造成了 6 种植物间茎环大小的差异。这一小的二级结构可能对 *psbA* 基因的表达调控有一定的影响。

关键词 小米, *psbA* 基因, 启动子, 茎环结构

已经知道大多数叶绿体基因的启动子都含有类似原核启动子的“-10”和“-35”元件, 而在高等植物中, 编码光系统反应中心蛋白 Q_B 的叶绿体 *psbA* 基因在“-10”和“-35”元件之间还含有类似真核启动子的“TATA”元件⁽¹⁾。研究表明, *psbA* 基因是受光调控的, 其光效表达的调控区正是位于 5' 上游的启动子区域内⁽²⁾。

psbA 基因的表达调控发生在转录、转录后、翻译和翻译后等多层次上。对 *psbA* 基因转录本的“run-on”分析表明, 在黑暗中生长了 8 d 的大麦幼苗经光照后, 其转录水平发生显著变化, 前 4 h 转录水平提高 13 倍⁽³⁾。对高粱质体 RNA 水平进行“run-on”分析显示, 光诱导的 RNA 水平的提高是由于转录水平的提高而不是 RNA 稳定性的改变⁽⁴⁾。而菠菜叶绿体的“run-on”转录实验则表明这种依赖光的 mRNA 的积累主要发生在转录后水平, 与转录本的稳定性密切相关⁽⁵⁾, 而且 RNA 的稳定性可能是通过 3' 末端的茎环结构及结合蛋白决定的⁽⁶⁾。虽然在叶绿体和黑暗条件下发育的质体中均存在着 *psbA* 的 mRNA, 但 mRNA 的转译却是直接或间接地受光调控的⁽⁷⁾。已从衣藻中分离到了与 *psbA* 及 *psbC* mRNA 的 5' 末端结合的蛋白因子, 它似乎可以与 *psbA* 基因上游的茎环结构发生专一性结合, 可能通过调整 RNA 二级结构以促进转译的起始⁽⁸⁾。由此可见, *psbA* 基因由于其在光合作用中的重要作用及其多层次的表达调控, 已被广泛地用于光合作用的分子生物学研究, 为我们深入探明质体基因表达调控机制提供了一个好材料。

鉴于我们在先前的工作中已经发现 C_4 植物高粱和 C_3 植物水稻相比, 在其 *psbA* 基因 5' 末端非编码区的转录与转译起始点之间, 前者比后者多出了一段由 7 个核苷酸组成

本研究系国家“863-101-07-05”基金资助。

* 现在通讯地址: 中国科学院动物研究所 北京 100080。

** 论文联系人。

本文于 1994 年 12 月 14 日收到。

的额外序列⁽⁹⁾，为了进一步探明这种差异是否具有普遍性，我们选用了另一种 C₃ 植物小米为材料，克隆了含 *psbA* 基因的 2.2kb *EcoR* I 片段，对 *psbA* 基因 5' 末端非编码区进行了核苷酸序列分析，并对小米、高粱⁽⁹⁾和 水稻⁽¹⁰⁾的 *psbA* 基因的启动区序列和结构进行了比较分析，发现前两者在转录和转译起始点之间都比水稻多出了“CTATTTT”7 个核苷酸额外序列，计算机分析显示这些植物的 *psbA* 基因 mRNA 前导序列区均能形成小的茎环结构，而这段额外序列正好位于茎环结构中。

1 材料与方 法

1.1 化学试剂及酶制剂

核酸限制酶购自 New England Biolabs 公司和 Bethesda Research Laboratories 公司。DNA 序列分析试剂盒、蛋白酶 K、碱性磷酸酶的 T4DNA 连接酶购自 Boehringer Mannheim 公司。低熔点琼脂糖 (LMP) 和缺口翻译试剂盒购自 BRL 公司。放射性同位素 α -³²PdNTPs (比活性为 3000Ci/mmol) 购自 Amersham 公司。DTT 等其它化学试剂均购自 Sigma 公司或 Bio-Rad 公司。

1.2 DNA 的制备、酶切、电泳和序列分析

均参照 Wu 等人^(10,11)的方法进行。

1.3 计算机分析

采用 DNASIS 程序，版本 5.00，由 HITACHI, SOFTWARE ENGINEERING CO., LTD 出品。

2 结果与讨论

2.1 小米 *psbA* 基因的克隆及其 5' 末端非编码区结构特征分析

小米叶绿体 DNA (cpDNA) 分别用 *Bam*H I, *Eco*R I, *Hind* III 和 *Pvu* I 消化，并经过 1% 琼脂糖凝胶电泳分离之后，转移到硝酸纤维滤膜上，用 ³²P 标记的水稻 *psbA* 基因⁽¹⁰⁾作为探针进行 Southern 杂交，结果表明 *psbA* 基因定位在 2.2kb *Eco*R I 片段上，与水稻⁽¹⁰⁾和高粱⁽¹¹⁾的情况完全相同。这也进一步证实了在高等植物的叶绿体基因组中，基因的排列是相当保守的。

为了克隆小米的 *psbA* 基因，我们构建了小米叶绿体基因组的专一性文库。小米的 cpDNA 经 *Eco*R I 酶切后，LMP 回收 1.5~3.0kb 之间的全部片段，与经过 *Eco*R I 酶解的 pBR322 载体 DNA 连接，获得了大约由 1000 个菌落组成的小米 cpDNA 之 *Eco*R I 限制片段富集的基因文库。使用 ³²P 标记的水稻 *psbA* 基因作为探针进行原位菌落杂交，共获得了 22 个阳性克隆。从这些阳性菌落中提取重组质粒 DNA，发现所有的重组质粒都含有 2.2kb *Eco*R I 插入片段，进一步的 Southern 杂交分析证明，所用的探针只能同 2.2kb *Eco*R I 插入片段杂交，而不能同载体 DNA 杂交，说明这条 2.2kb *Eco*R I 插入片段确实含有小米 *psbA* 基因。我们将其中一个阳性克隆命名为 pFMpsbA1。

本实验采用不同的单酶或双酶组合来消解 pFMpsbA1 DNA，分别用来自水稻的 1.0kb *Bgl* II -*Eco*R I (含 *psbA* 基因 5' 末端非编码区) 和 1.2kb *Bgl* II -*Eco*R I (包括完整的 *psbA* 基因编码区)⁽¹⁰⁾作为探针，构建了小米 *psbA* 基因及其 5' 末端非编码区的限制图。从所用的 *Ava* I, *Bam*H I, *Bgl* II, *Hind* I, *Pvu* I 和 *Xba* I 等 6 种限制酶来看，其

限制图与高粱的⁽¹¹⁾完全相同，而与水稻的⁽¹⁰⁾不同，例如 Ava I 和 Hind I 在小米和高粱的 psbA 基因的 5' 末端上游非编码区的 0~80bp 的范围内，都有 1 个识别位点，而在水稻中则没有相应的位点。这种限制性位点的多态性说明，两种 C₄ 植物小米和高粱在 psbA 基因的 5' 末端非编码区的核苷酸序列结构上非常相似，而与 C₃ 植物水稻有明显的差异。

序列分析结果表明，和大多数叶绿体基因一样，在小米 psbA 基因的 5' 末端非编码区也存在着与原核生物类似的启动子结构。其“-10”区序列为 TATACT，原核生物“-10”区共有基序为 TATAAT，两者仅相差一个核苷酸；“-35”区序列为 TTGACA，原核生物“-35”区共有基序为 TTGACA，两者完全相同（图 1）。除了以上两个序列之外，小米 psbA 基因的启动区在“-10”区和“-35”区之间还存在着一个类似真核生物启动子的“TATATA”保守序列。这些结果表明，小米 psbA 基因的启动区既具有原核的特征又具有真核的特征。从表 1 可以看出，不同植物来源的 psbA 基因的启动子，在这三种元件的核苷酸序列结构上都是相当保守的。

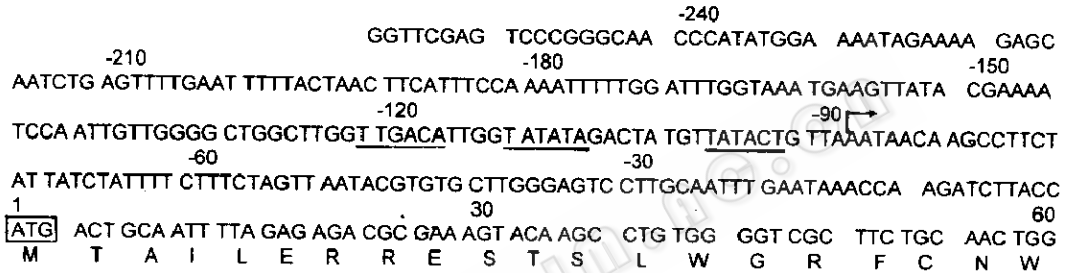


图 1 小米 psbA 基因的 5' 末端非编码区的核苷酸序列
Fig. 1 Sequence of the 5'-noncoding region of millet psbA gene.

- Indicates the putative transcription initiation site deduced from the published psbA sequences.
- Indicates start codon which is 87 bp distant from the transcription initiation site. This 87bp fragment is the mRNA leader region.
- Indicates “-10” box, “-35” box and “TATA” box.

2.2 小米 psbA 基因 mRNA 前导序列区的结构分析

在大多数叶绿体基因的 5' 末端上游非编码区都存在着一段 mRNA 前导序列，它的长度可以从几个核苷酸到几百个核苷酸，但对于已研究过的大多数植物的 psbA 基因而言，其 mRNA 前导序列的长度相当稳定，一般为 80±4 个核苷酸，例如 C₃ 植物水稻 psbA 基因的 mRNA 前导序列为 80bp，另外 3 种 C₃ 植物小麦、大麦和黑麦的为 83bp，而 C₄ 植物高粱和小米的却为 87bp，比水稻多出了“CTATTTT”7 个核苷酸，比小麦、大麦和黑麦多出了“TTTT”4 个核苷酸（表 2）。如果去掉“CTATTTT”这段额外序列，小米和高粱 psbA 基因的 mRNA 前导序列的其余部分与水稻的完全相同。水稻、小麦、大麦和黑麦是分属于单子叶植物禾本科植物 (*Gramineae*) 四个属 (genus) 的 C₃ 植物，而小米和高粱则是分属于禾本科黍亚科 (*Panicoideae*) 的两种 C₄ 植物，因此在禾本科的 C₃ 和 C₄ 植物之间，这种 psbA 基因 mRNA 前导序列的差异可能具有普遍性。计算机分析结果显示，以上 6 种禾本科 psbA 基因 mRNA 前导序列区均能形成小的茎环结构（图 2），而这段 7bp 的额外序列恰好位于茎环结构中，造成了茎环大小的差异。小米和高粱的茎 (stem) 区为 9 对碱基，其中 8 对严格配对，环 (loop) 区为 9 个碱基；水稻的茎区为 8

表 1 不同植物的 psbA 基因启动区核苷酸的同源性比较

Table 1 Homology comparison of nucleotide sequences among different plant psbA promoters

<i>E. coli</i>	"-35 Box" tgTTGACA			"Pribnow Box" attgtTATAATg			
	-51	-41	-31	-21	-11	-1	
MILLET	CAATTGTTGGGGCTGGCTTGGTTGACATTGGTATATAGACTATGTTATACTGTTA						
SORGHUM	-----						
RICE	C	C			T	A	
WHEAT	T	ATC	T	A		T	
BARLEY	C	A	C	T	N	AT	C
MUSTARD	ATC	A	CCATTT	ACA		G	C
ALFALFA					CC	AGTC	G
PETUNIA					CGA	AGTCG	G
PEA	T	CGGTGGATATTG	TA	G	CCC	AGTC	T
SOYBEAN	T	C	ATGGATATTG	TA	C	AGTC	G
<i>N. plumbaginifolia</i>	AT	GATC	ACATACAC		CGA	AGTC	G
<i>N. debneyi</i>	AT	GATC	ACATACAC		CGA	AGTC	G
SPINACH	AT	GATC	CACTAGATA		CG	C	AGGC
COTTON	T	T	CTATTTGATA		CG		AGTC
SOLANUM	AT	GATCCA	ATACA		CGA		AGTC

a. "N" indicates the deleted nucleotide. b. "-" indicates the same nucleotide.

表 2 小米与禾本科其它 5 种植物之间 psbA 基因 mRNA 前导序列区的同源性比较

Table 2 Homology comparison of nucleotide sequences of mRNA

leader regions between millet and other five gramineae species.

	-60	-50	-40
MILLET	AATAACAAGCCTTCTATTATCTATTTCTTTCTAGTTAATACGTGTGCTTGGGAG		
SORGHUM	-----		
RICE	NNNNNNN		
WHEAT	NNNNNTA		A
BARLEY	NNNNNTA		
RYE	NNNNNTA		

a. "N" indicates the deleted nucleotide.
b. "-" indicates the same nucleotide.

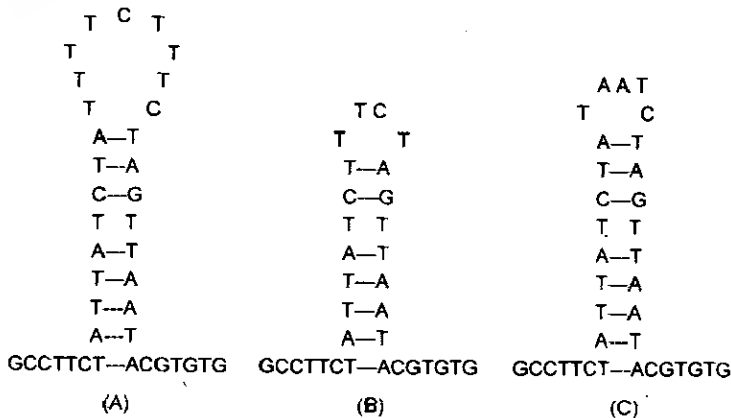


图 2 推导的 6 种禾本科植物 psbA 基因前导序列区的二级结构

Fig. 2 Predicted secondary structures for the mRNA leader regions of psbA genes of six Gramineae species.

(A) Millet and sorghum; (B) Rice; (C) Wheat, barley and rye

对碱基, 其中 7 对严格配对, 环区为 4 个碱基; 小麦、大麦和黑麦的茎区为 9 对碱基, 其中 8 对严格配对, 环区为 5 个碱基。已经有实验表明在衣藻中似乎存在着专一的核基因产物调控 mRNA 的转译^[12,13]。这些核基因产物的作用机制尚不十分清楚, 它们可能通过调整 RNA 的二级结构来促进转译的起始^[8]。但高等植物与衣藻不同, 其多顺反子叶绿体 mRNA 的 5' 末端非编码区相对较短, 多数情况下难以形成二级结构。因此, 这段额外序列对于 psbA 基因的 mRNA 分子的稳定性和转录水平的影响及它同 C₃ 植物的高光合效率之间的内在联系等问题都有待于进一步研究。

参 考 文 献

- [1] Eisermann A, Tiller K, Link G. EMBO J, 1990, 9: 3981~3987.
- [2] Gilmartin P M, Sarokin L, Memelink J *et al.* The Plant Cell, 1990, 2: 369~378.
- [3] Klein R R, Mullet J E. J Biol Chem., 1990, 265: 1895~1902.
- [4] Schrubar H, Wanner G, Westhoff P. Planta, 1990, 183: 101~111.
- [5] Deng X W, Gruissem W. Cell, 1987, 49: 379~387.
- [6] Plant A L, Gray J C. Photosynth Res, 1988, 16: 23~39.
- [7] Klein R R, Mason H S, Mullét J. E. J Cell Biol., 1987, 106: 289~301.
- [8] Danon A, Mayfield S P Y. EMBO J, 1991, 10: 3993~4001.
- [9] Shi X M, Wu N H *et al.* Chinese Sci Bull., 1994, 39: 1224~1227.
- [10] Wu N H, Cote J C, Wu Ray. Developmental Genetics, 1987, 8: 339~330.
- [11] Wu N H, Shi X M. Chinese J Bot., 1991, 3: 81~88.
- [12] Kuchka M R, Mayfield S P Y, Rochaix J-D. EMBO J, 1988, 7: 319~324.
- [13] Rochaix J D, Kuchka M, Mayfield S *et al.* EMBO J., 1989, 8: 1013~1021.

Characterization of the 5'-noncoding Region of Millet Chloroplast psbA Gene

Huang Yao Wu Naihu

(Institute of Developmental Biology, Academia Sinica, Beijing 100080)

Abstract The 2.2kb EcoR I fragment containing the chloroplast psbA gene from millet (*Setaria italica*) has been cloned and the nucleotide sequence of the 5'-noncoding region has been determined. The sequence shows that both prokaryote-like and prokaryote-like promoter structures exist in the 5'-flanking region of psbA. Compared with the consensus sequence of the prokaryotic promoter, only one nucleotide difference is found in the "-10" box TATACT and no difference in the "-35" box TTGACA. Between these two boxes, there is a consensus sequence "TATATA". All these results indicate that millet psbA promoter has both prokaryotic and eukaryotic characteristics. The mRNA leader region of millet psbA gene is 87 bp long which is the same as that of the sorghum gene. There is an additional CTATTTT sequence compared with rice and TTTT with wheat, barley and rye genes. So it is presumably universal that this difference exists between C₃ and C₄ plants in the family of Gramineae. A small stem-loop structure could be found in psbA mRNA leader regions of the above-mentioned six plants by computer analysis. The additional CTATTTT sequence is located in the stem-loop structure, which leads to the different size of the stem-loops of these six species. This small secondary structure may have an effect on the regulation of expression of psbA gene.

Key words Millet, psbA gene, promoter, stem-loop structure