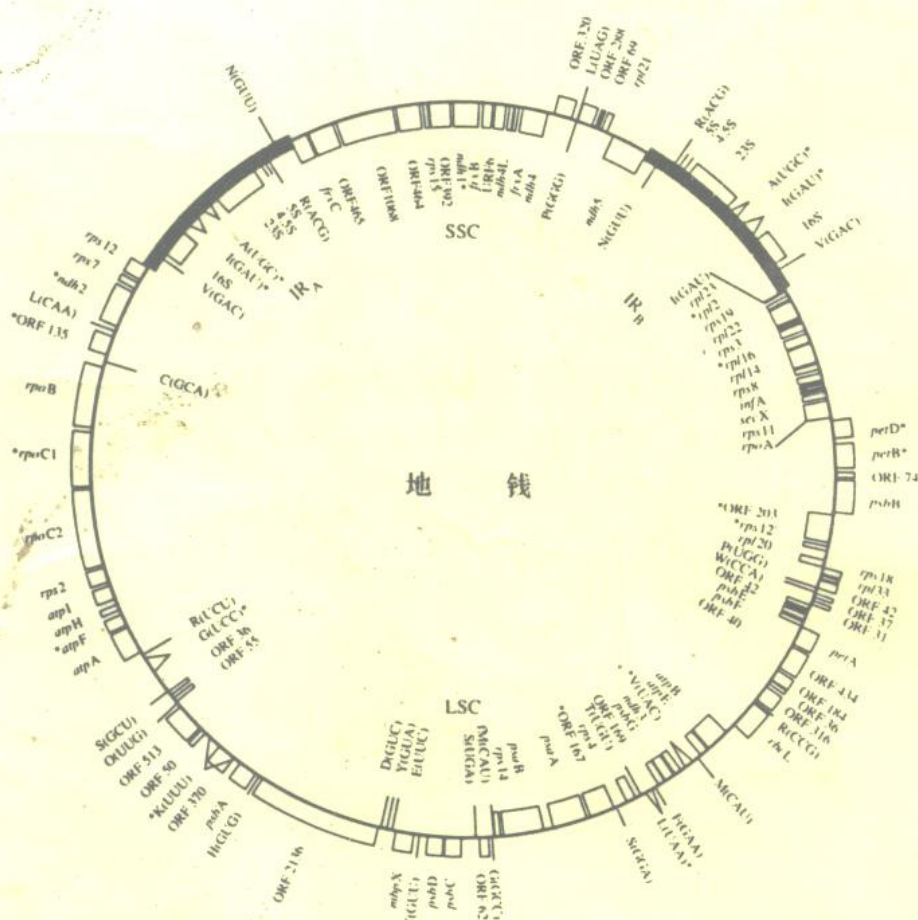


植物基因组:构建、表达和调控

赵微平



首都师范大学 植物生理及分子生物学区研究室

植物基因组：构建、表达和调控

赵微平

首都师范大学 植物生理及分子生物学研究室

京准字 93—228 号

内部交流教材

首都师范大学 植物生理及分子生物学研究室

(北京西三环北路 105 号, 100037)

1993 年 10 月

开本: 787×1092 毫米 1/16

印张: 12.0

字数: 300 千字

印数: 0—500 册

目 录

前 言	1
1. 叶绿体基因组	2
1.1. RNA 基因	3
1.1.1. rRNA 基因	3
1.1.2. tRNA 基因	4
1.2. 蛋白质基因	6
1.2.1. r-蛋白的基因	6
1.2.2. RNA 聚合酶基因(rpoA,B,C)	7
1.2.3. Rubisco 大亚基基因(rbcL)	8
1.2.4. 光系统 II 基因(psbA,B,C,D,E,F,H,I)	11
1.2.5. petA,B,D,E	11
1.2.6. 光系统 I 基因(psaA,B,C)	12
1.2.7. ATP 合成酶基因(atpA,B,E,F,H,I)	14
1.2.8. NADH 去氢酶基因(ndh)	15
2. 线粒体基因组	15
2.1. RNA 基因	17
2.1.1. rRNA 基因(rDNA)	17
2.1.2. tRNA 基因(trn)	18
2.2. 蛋白质基因	18
2.2.1. r-蛋白基因(rpS)	18
2.2.2. 编码其它蛋白的基因	18
3. 核基因组	23
3.1. 核蛋白体基因和组蛋白基因	23
3.1.1. rRNA 基因(rDNA)	23
3.1.2. tRNA 基因	27
3.1.3. 编码蛋白质基因的特征	27
3.1.4. r-蛋白质基因	29
3.1.5. 组蛋白基因	31
3.2. 光合基因	33
3.2.1. Rubisco 小亚基基因(rbcS)	33
3.2.2. 叶绿体 a/b 结合蛋白基因(cab)	36
3.2.3. psd,E,F,G,H	44
3.2.4. pet 基因	46
3.2.5. 编码 PSII 亚基的基因	47
3.2.6. 编码氧释放复合物的基因	47
3.2.7. 叶绿体果糖-1,6-二磷酸酶的基因(Fbp)	50

3.2.8.	磷酸核酮糖激酶基因 (Prk)	52
3.2.9.	磷酸甘油酸激酶基因 (Pgk)	52
3.2.10.	甘油醛磷酸去氢酶基因 (Gap)	52
3.2.11.	磷酸烯醇式丙酮酸羧化酶基因 (Pepc)	55
3.2.12.	丙酮酸正磷酸二激酶基因 (Ppdk)	57
3.2.13.	硫氧还蛋白基因	59
3.2.14.	碳酸酐酶基因	60
3.2.15.	蔗糖运输载体基因	61
3.2.16.	多酚氧化酶基因	62
3.3.	贮存蛋白基因	63
3.3.1.	小麦醇溶蛋白基因 (Gl) 和谷蛋白基因 (Glu)	63
3.3.2.	大麦醇溶蛋白基因	64
3.3.3.	水稻醇溶蛋白基因	65
3.3.4.	玉米醇溶蛋白基因	65
3.3.5.	高粱醇溶蛋白基因	68
3.3.6.	大豆球蛋白基因	68
3.3.7.	豌豆贮存蛋白基因	69
3.3.8.	蚕豆豆球蛋白基因	70
3.3.9.	菜豆球蛋白基因	70
3.3.10.	油菜小蛋白基因和十字花科蛋白基因	71
3.3.11.	油蛋白基因	72
3.3.12.	马铃薯蛋白基因	72
3.3.13.	甘薯蛋白基因	73
3.3.14.	贮铁蛋白基因	74
3.3.15.	凝集素基因 (Lec)	75
3.3.16.	胰蛋白酶/ α -淀粉酶抑制物基因	76
3.4.	糖代谢基因	79
3.4.1.	α -淀粉酶基因 (α -Amy)	79
3.4.2.	β -淀粉酶基因 (β -Amy)	84
3.4.3.	ADP 葡萄糖焦磷酸酶基因	84
3.4.4.	UDP 葡萄糖淀粉糖基转移酶基因 (Waxy)	85
3.4.5.	淀粉分支酶基因	86
3.4.6.	蔗糖合成酶基因	86
3.4.7.	磷酸葡萄糖异构酶基因	87
3.4.8.	醛缩酶基因	88
3.4.9.	NADP-山梨糖-6-磷酸去氢酶基因	88
3.4.10.	酸性转化酶基因	88
3.4.11.	丙酮酸脱羧酶基因	88
3.4.12.	乙醇脱氢酶基因 (Adh)	89
3.4.13.	乳酸脱氢酶基因	90

3.4.14.	需 NADH 的羟基丙酮酸还原酶基因	90
3.4.15.	柠檬酸合成酶基因	90
3.4.16.	异柠檬酸去氢酶基因	92
3.4.17.	苹果酸合成酶基因	93
3.5.	氮代谢基因	93
3.5.1.	硝酸还原酶基因(nia)	93
3.5.2.	亚硝酸还原酶基因	95
3.5.3.	谷氨酰胺合成酶基因(gln)	96
3.5.4.	谷氨酸草酰乙酸转氨酶基因(Got)	99
3.5.5.	丙氨酸氨基转移酶基因	99
3.5.6.	谷氨酸去氢酶基因	100
3.5.7.	天冬酰胺酶基因	101
3.5.8.	天冬氨酸转氨酶基因	101
3.5.9.	半胱氨酸合成酶基因	102
3.5.10.	羧肽酶基因	102
3.5.11.	内肽酶基因	102
3.5.12.	氨肽酶基因	103
3.5.13.	硫醇蛋白酶基因	103
3.5.14.	5-烯醇式丙酮酸莽草酸-3-磷酸合成酶(EPSPS)基因	103
3.5.15.	二硫化物异构酶基因	103
3.5.16.	豆血红蛋白基因	104
3.6.	脂肪代谢基因	104
3.6.1.	乙酰载体蛋白基因	104
3.6.2.	烯酰乙酰载体蛋白基因	104
3.6.3.	去饱和酶基因	105
3.6.4.	脂类转运蛋白基因	106
3.7.	生长发育	106
3.7.1.	植物光敏素基因(phy)	106
3.7.2.	SAM 合成酶基因	110
3.7.3.	编码 1-氨基环丙烷-1-羧酸(ACC)合成酶的基因	110
3.7.4.	生长素结合蛋白基因	112
3.7.5.	对生长素反应的基因	113
3.7.6.	对脱落酸反应的基因	113
3.7.7.	茉莉酸诱导蛋白基因	114
3.7.8.	3-羟-3-甲基戊二酸-辅酶 A 还原酶基因	115
3.7.9.	Em 蛋白基因	115
3.7.10.	腺苷酸激酶基因	117
3.7.11.	胚发生蛋白基因	118
3.7.12.	细胞周期调节基因	119
3.7.13.	周期蛋白基因	119

3.7.14.	聚半乳糖醛酸酶基因.....	120
3.7.15.	钙调蛋白基因.....	120
3.7.16.	蛋白质磷酸化酶基因.....	121
3.7.17.	parB 基因	122
3.7.18.	休眠基因.....	122
3.8.	逆境反应.....	123
3.8.1.	热激蛋白基因(hsp)	123
3.8.2.	调渗蛋白基因.....	124
3.8.3.	寒冷诱导基因.....	124
3.8.4.	盐逆境基因.....	124
3.8.5.	过氧化物酶基因.....	126
3.8.6.	抗坏血酸氧化酶基因.....	127
3.8.7.	过氧化氢酶基因.....	127
3.8.8.	超氧歧化酶基因.....	129
3.8.9.	甜菜碱醛脱氢酶基因.....	131
3.8.10.	细胞壁蛋白基因.....	131
3.8.11.	异黄酮还原酶基因.....	133
3.8.12.	查尔酮合成酶基因.....	134
3.8.13.	苯丙氨酸—氨裂解酶基因.....	135
3.8.14.	肉桂醇去氢酶基因.....	136
3.8.15.	木素 O—甲基转移酶基因.....	137
3.8.16.	几丁质酶基因.....	137
3.9.	转录和翻译	138
3.9.1.	转录因子基因.....	138
3.9.2.	RNA 聚合酶 II 最大亚基基因	138
3.9.3.	RNA 螺旋酶基因	138
3.9.4.	翻译起始因子基因.....	139
3.9.5.	延长因子基因.....	139
3.9.6.	泛在蛋白基因.....	140
3.9.7.	蛋白质激酶基因.....	143
3.9.8.	腺嘌呤磷酸核糖基转移酶基因.....	143
3.10.	细胞骨架.....	143
3.10.1.	微管蛋白基因.....	143
3.10.2.	肌动蛋白基因.....	144
3.11.	膜蛋白基因.....	146
3.11.1.	ATP 酶基因	146
3.11.2.	H ⁺ /葡萄糖共运输载体基因	151
3.11.3.	腺嘌呤核苷酸传递体基因.....	152
3.11.4.	磷脂转移蛋白基因.....	152
4.	基因表达的调节.....	152

4.1.	DNA 甲基化	152
4.2.	转录水平上的调节.....	152
4.3.	转录因子.....	154
4.4.	光调节基因.....	160
4.5.	转录后调节.....	161
4.6.	前-mRNA 拼接.....	161
4.7.	翻译控制.....	162
4.8.	果实成熟与基因表达.....	162
4.9.	植物和器官发育与基因表达.....	164
4.10	生物钟控制植物基因的表达.....	166
	参考文献.....	168

前言

植物的生长发育是植物体(主要是细胞)在多种代谢和生理过程基础上所发生的基因在时间上和空间上表达的综合现象,它涉及到遗传信息的贮存和传递、多种多样酶和物质的合成和分解、生理功能的相互协调、激素调控、细胞分化和形态建成等等。各种类型细胞结构上的可见差异性和重要的代谢特征都是由蛋白质和酶存在的独特样式所决定的,而蛋白质的合成又取决于核酸的存在和变化,所以研究基因的分子生物学对最后弄清植物生长发育的机理具有重大的意义。在细胞内,DNA 是染色质的主要成分,而染色质又是细胞核的主要组份。基因则是染色体上的具有一定位座的遗传单位,是DNA 分子一定长度的核苷酸序列。在植物细胞内,一般存在有三个染色体组或基因组,也就是叶绿体基因组、线粒体基因组和核基因组。

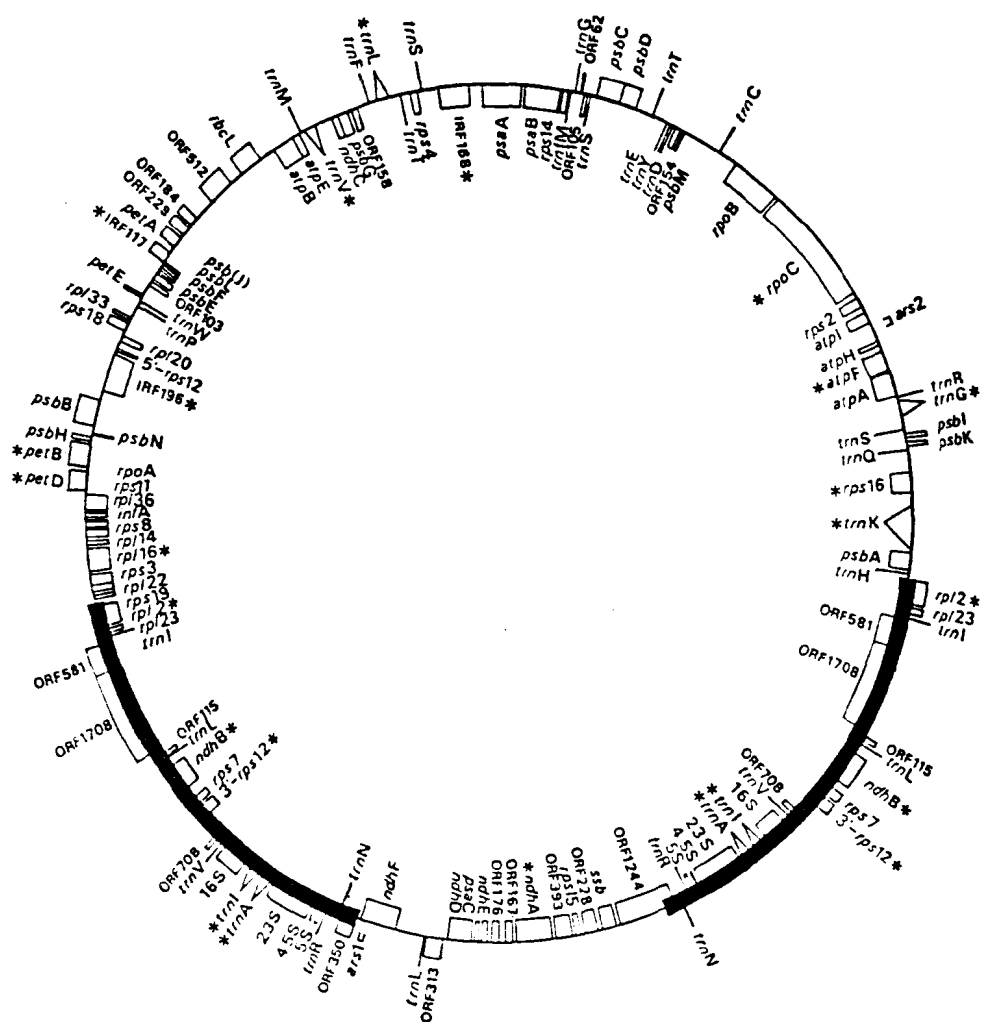


图1 烟草叶绿体基因组的环形基因图。环外面列出的基因是A 股上的,是逆时针转录的,环内列出的基因是存在B 股上的,是顺时针转录的。* 表示断裂的基因。主要的开放读框(ORF)都包括了进来。(依 Shinczaki 等,1986)

由于光合作用重要性和叶绿体 DNA 比较简单,所以对叶绿体基因组的分子生物学研究开始得比较早。

1. 叶绿体基因组

叶绿体基因组是一环状的双螺旋 DNA 分子,因此也叫叶绿体 DNA(cpDNA)。它们的大小一般变动在 120 kbp 至 217 kbp 之间。均具有携带 rRNA 基因(rDNA)的重复序列,重复区的长度在 10 kbp(拟南芥菜)和 25 kbp(烟草)之间。大多数植物的叶绿体基因组是两个含有 rDNA 的倒置重复(IR)被一个大的单拷贝区(LSC,大约长 80kbp)和一个小的单拷贝区(SSC,大约长 20kbp)分开而组成的。一些豆类(如蚕豆和豌豆)则只有一个 rDNA 单位。烟草叶绿体 DNA 研究比较详细,它是一环状的由 155844 个 bp 组成的,含有两个 25339 bp 倒置重复,被一个大的 86684 bp 和一个小的 18482 bp 单拷贝区分开。整个基因组含有四种 rRNA (16S, 23S, 5S, 4.5S)、30 种 tRNA、44 种蛋白质以及 9 种其他蛋白质的基因(图 1)。所有这些基因都是在叶绿体内转录的。最近发现的同大肠杆菌核蛋白体 50S 亚基结合的 L36 r 蛋白也是叶绿体的 r 蛋白,其基因存在于 *infA* 基因和 *rpS11* 基因之间。*rpL36* 以前也叫 *secX*。

cpDNA 是相当保守的,DNA-DNA 分子杂交显示,在种内甚至在种间,限制酶切图和序列都是一样或类似的。在无关的植物间至少有 30% 序列是共同的。这些保守序列分散存在,可能与密码区和非密码区有关。烟草、菠菜、撞羽朝颜(矮牵牛)和黄瓜的 cpDNA 基本上都是线性相关的。

根据对多种植物的分析,质体基因组大约共含有 123 个基因,可分成为两大类:

(1)同光合有关的基因

Rubisco	<i>rbcL</i>
PSI	<i>psaA, B, C, I, J.</i>
PSII	<i>psbA, B, C, D, E, F, G, H, I, J, K, L. M. N.</i>
ATP 合成酶	<i>atpA, B, E, F, H, I.</i>
细胞色素复合物	<i>petA, B, D, E. G.</i>
NADH 去氢酶	<i>ndhA, B, C, D, E, F, G, H, J, K.</i>

! 叶绿体基因表达所需要的基因

转录

RNA 聚合酶	<i>rpoA, B. C.</i>
---------	--------------------

翻译

rRNA	rDNA (16S, 23S, 5S, 4.5S)
tRNA	<i>trn</i> (30 个)
r 蛋白小亚基	<i>rpS</i> (12 个)
r 蛋白大亚基	<i>rpL</i> (8 个)
起始因子	<i>infA</i>
延长因子	<i>tufA</i>
复制	
单股结合蛋白	<i>ssb</i>

未鉴定

>100 个密码(19 个)

<100 个密码(18 个)

```
1 gacaattaaa tctaattttg cccattcagc taagagtgtt ccacggaggc ccgtttaagt
61 taaaaaaaaa aatagtcgcg ttgaatttct cgaccttttg acttaggatt agtcagtcctt
      "-35 box"      "-10 box"
121 ttttctcgat tagggacaag ggaAGGGGTA TAACTCAGCG CTAGAGTGTC ACCTTGACGT
      →tRNAVal (GAC)
181 GGTGGAAGTC ATCAGTTCGA ACCTGATTAT CCCTAaaccc aatatcaatt tttccatttt
241 gaggacgaga tccaatctga gtagataaga ggatagggag ttgacacaag ggggggtaag
      "-35 box"
301 gccatataat atttatggga ggcaactccg ggcgaatagt aagcccatgg atacaagtca
      "-10 box"      ▲
361 agttatgtct tctcagttca gtaactgaaa tcaaatttaa gttcagtaaa tgaaatcaaa
421 ttccgaatca gctttgtcta gaaacaaggga agctataagt aatgcaacta gGAAGCTCAT
      →16S rRNA
481 GGAGAGTTTG ATCCTGGCTC AGGATGAACG CTGGCGGCAT GCTTTACACA TGCAAGTCGG
541 ACGGGAAGTG GTGTTTCCAG TGGCGAACGG GTGAGTAACG CGTAAGAACC TGCCCTTGGG
601 AGGGGGACAA CAGCTGGAAA CGGCTGCTAA TACCCCGTAG GCTGAGGAGC GAAAGGAGGA
661 ATCCGCCCAA GGAGGGGCTC CCGTCTGATT AGCTAGTTGG TGAGGTAATA GCTTACCAAG
721 GCGATGATCA GTAGCTGGTC CGAGAGGATG ATCAGCCACA CTGGGACTGA GACAAGGTC
781 AGACTCCTAC GGGAGGCAGC AGTGGGGAAT TTTCCGCAAT GGGCGAAAGC CTGACGGAGC
841 AATGCCGCGT GGAGGTAGAG GCCCTGGGT CATGAACTTC TTTTCCCGGA GAAGAAAAAA
901 TGACGGTATC CGGGGAATAA GCATCGGCTA ACTCTGTGCC AGCAGCCGCG GTAAGACAGA
961 GGATGCAAGC GTTATCCGGA ATGATTGGGC GTAAAGCGTC TGTAGGTGGC TTTTAAAGTT
1021 CGCTGTCAAA TACCAGGGCT CAACCCTGGA CAGGTGGTGA AAACACTAA GCTAGAGTAC
1081 GGTAGGGGCA GAGGGAATTT CCGTGGGAGC GATGAAATGC GTAGAGATCG GAAGGAACAC
1141 CAACGGCGAA AGCACTCTGC TGGGCCGACA CTGACACTGA GAGACGAAAG CTAGGGGAGC
1201 GAATGGGATT AGAGACCCCA GTAGTCTTAG CCGTAAACGA TGGATACTAA TGGCTGTGC
1261 GTATCGACCC GTGCAACGCT GTAGCTAACG CGTTAAGTAT CCCGCCTGGG GAGTACGTTT
1321 GCAAGAATGA AACTCAAAGG AATTGACGGG GGCCCGACA AGCGGTGGAG CATGTGGTTT
1381 AATTCGATGC AAAGCGAAGA ACCTTACCAG GGCTTGACAT GCGCGAATCC TCTTGAAGA
1441 GAGGAGTGCC TTCGGGAATG CGACACAGGT GGTGCATGGC TGTCGTCAGC TCGTGCCTG
1501 AGGTGTTGGG TTAAGTCCCG CAACGAGCGC AACCCTCGTG TTTAGTTGCC AACGTTTAGT
1561 TTGGAACCTCT GAACGAGCTG CCGGTGATAA GCCGGAGGAA GGTGAGGATG ACGTCAAGTC
1621 ATCATGCCCC TTATGCCCTG GGCTACACAC GTGCTACAAT GGACCGGACA AAGGATCGCG
1681 ACCCGCGGAG GGTGAGCTAA CTTCAAAAAC CTGTCTCAG TTCGGATTGT AGGCTGCAAC
1741 TCGCTACAT GAAGCCGGAA TCGCTAGTAA TCGCCGGTCA GCCATACGGC GG'TGAATTCG
1801 TTCCCGGGCC TTGTACACAC CGCCCGTCAC ACTATGGGAG CTGGCCATGC CCGAAGTCAT
1861 TACCTTAACC GCAAGGAGGG GGATGCCGAA GGCAGGGCTA GTGACTGGAG TGAAGTCGTA
1921 ACAAGGTAGC CGTACTGGAA GGTGCGACTG GATCACCTCC Ttttcaggga gaactaatgc

1981 tttttgtgta attaggtttt acactgcgtc actcacaccc cggttaaagta aagaaaagct
2041 acatcgcact agtaggtaga tttccgttat gtttcgttcc tcgacgggtga agtcacacct
2101 agctcaggag ctagcttatt aacctattca gggagaacag agaacgcttt ttgggtgatt
2161 gggttttaca ctgtttcaca ccgctgaagc gaaggggtta tcgaggcaaa gttagtaggt
2221 ggaagtctcc tttcgtttcc tgtgggtgat gagaatgtca gtgacactaa tctcatgagc
2281 ttattatcct aggtcggaac aacttgatag gagct
```

图2 豌豆叶绿体含有 tRNA^{Val} 和 16S rRNA 基因的核苷酸序列。大写字母是编码区,启动子成分下画有横线,16S rRNA 基因的转录起始位点由箭头表示。依 Cerutti 和 Jagendorf, 1991)

1.1. RNA 基因

1.1.1. rRNA 基因(rDNA) 为四种叶绿体核蛋白体 RNA(16S, 23S, 4.5S 和 5SrRNA) 编码的基因成串地存在于倒置重复区,其排列的顺序是 16SrDNA—区间—23SrDNA—区间—5SrDNA,与大肠杆菌和兰藻类似。在高等植物,叶绿体尚有一种小的 4.5SrRNA,它的基因接近于 23SrDNA 的 3' 端。最近也发现,两种低分子量的 3S 和 7SrRNA 存在于衣藻 50S 核蛋白体亚

基上,由 23SrDNA 5' 端前面的序列为它们编码。在眼虫的 16S-23SrDNA 区间含有 tRNA^{Pro}和 tRNA^{Asn}基因。许多植物的 16S-23SrDNA 区间皆大于 2000 个碱对,这主要都是由于存在有 tRNA 基因含有内含子的缘故。叶绿体 16SrDNA 的大小估计为 1491 个 bp,而大肠杆菌的为 1541 个 bp,其中有 1144 个是相同的(74%同源)。烟草和眼虫的序列有 96%和 80%与玉米的序列是相同的。Cerutti 和 Jagendorf(1991)从豌豆获得叶绿体 DNA 的 2.3kb 片段并克隆进 pBluescript 载体中进行了全序列分析。他们证明,这个克隆含有 tRNA^{Val}基因(72 个核苷酸)和 16S rRNA 的基因(1490 个核苷酸)(图 2)。

对玉米和烟草叶绿体 23S 和 4.5SrDNA 克隆的 cpDNA 也进行过序列分析。烟草 23SrDNA 的 3' 端后面是 101bp 的区间,然后则是 103bp 的 4.5SrDNA。某些植物四种 rRNA 是共转录的,而另一些植物 5SrRNA 可能来自于分开的转录单位。rRNA 的丰度同转录的高速度相关。在衣藻叶绿体的 23SrDNA 中最先发现有内含子。凯木(*Alnus*)叶绿体 23SrRNA 基因的核苷酸序列见图 3。

1.1.2. tRNA 基因 在 cpDNA 分子上有三十余种 tRNA 基因,其中有 7 种存在于倒置重复中,其他则分散在整个基因组中。一般认为,叶绿体的蛋白质合成不需要由细胞质输入 tRNA。叶绿体 tRNA 基因见表 1。

表 1 植物叶绿体 tRNA 基因

基因	tRNA	基因	RNA
trnA-UGC	Ala-tRNA(UGC)	trnL-UAG	Leu-tRNA(UAG)
trnR-ACG	Arg-tRNA(ACG)	trnK-UUU	Lys-tRNA(UUU)
trnR-UCU	Arg-tRNA(UCU)	trnM-CAU	fMet-tRNA(CAU)
trnR-CCG	Arg-tRNA(CCG)	trnM-CAU	Met-tRNA(CAU)
trnN-GUU	Asn-tRNA(GUU)	trnF-GAA	Phe-tRNA(GAA)
trnD-GUC	Asp-tRNA(GUC)	trnP-UGG	Pro-tRNA(GAA)
trnC-GCA	Cys-tRNA(GCA)	trnS-GGA	Ser-tRNA(GGA)
trnQ-UUG	Gln-tRNA(UUG)	trnS-UGA	Ser-tRNA(UGA)
trnE-UUC	Glu-tRNA(UUC)	trnS-GC	Ser-tRNA(GCU)
trnG-GCC	Gly-tRNA(GCC)	trnT-GGU	Thr-tRNA(GGU)
trnG-UCC	Gly-tRNA(UCC)	trnT-UGU	Thr-tRNA(UGU)
trnH-GUG	His-tRNA(GUG)	trnW-CCA	Trp-tRNA(CCA)
trnI-GAU	Ile-tRNA(GAU)	trnY-GUA	Tyr-tRNA(GUA)
trnI-CAU	Ile-tRNA(CAU)	trnV-GAC	Val-tRNA(GAC)
trnL-UAA	Leu-tRNA(UAA)	trnV-UAC	Val-tRNA(UAC)
trnL-CAA	Leu-tRNA(CAA)		

(依 Sugura, 1992)

对眼虫 cpDNA 的 EcoRI 片断 G1.6kb 部分的序列分析证明,在 367 个 bp 内存在有 4 个 tRNA 基因。排列的情况是:tRNA^{Val}-16bp 区间-tRNA^{Asn}-3bp 区间-tRNA^{Asp}-45bp 区间-tRNA^{Pro}。烟草叶绿体 tRNA^{Asn}基因存在于倒置重复区,离 5S rDNA 远端大约 900bp 处和小的单

TTCAAACGAG GAAAGGCTTA CGGTGGATAC CTAGGCACCC AGAGACGAGG AAGGGCGTAG TAAGCGACGA 70
 AATGCTTCGG GGAGTTGAAA ATAAGCGTAG ATCCGGAGAT TCCCGAATAG GTCAACCTTT CAAACTGCTG 140
 CTGAATCCAT GGGCAGGCAA GAGACAACCT AACGAATGA AACATCTTAG TAGCCAGAGG AAAAGAAAGC 210
 AAAAGCGATT CCCGTAGTAG CAGGCGAGCG AAATGGGAGC AGCCTAAACC GTGAAAACGG GGTGTGGGA 280
 GAGCAATACA AGCGTCGTGC TGCTAGGCGA AGCGGTGAAG TGCCGAACCC TAGATGGCTA GAGTCCAGTA 350
 GCCGAAGCA TCACTAGCTT ACGCTCTGAC CCGAGTAGCA TGGGGCACGT GGAATCCCGT GTGAATCAGC 420
 AAGGACCACC TTGCAAGGCT AAATACTCCT GGGTGACCGA TAGTGAAGTA GTACCGTGAG GGAAGGGTGA 490
 AAAGAACCCC CACCGGGGAG TGAATAGAA CATGAACCG TAAGCTCCCA AGCAGTGGGA GGAGCCTGGG 560
 GCTCTGACCG CGTGCTGTT GAAGAATGAG CCGGCGACTC ATAGGCAGTG GCTTGTTAA GGAACCCAC 630
 CGGAGCCGTA GCGAAAGCGA GTCTTCATAG GGCAATTGTC ACTGCTTATG GACCCGAACC TGGGTGATCT 700
 ATCCATGACC AGGATGAAGC TTGGGTGAAA CTAAGTGGAG GTCCGAACCC ACTGATGTTG AAGAATCAGC 770
 GGATGAGTTG TGGTTAGGGG TGAATGCCA CTCGAACCCA GAGCTAGCTG GTTCTCCCGG AAATGCGTTG 840
 AGGCGCAGCA GTTGAAGTGA CATCTAGGGG TAAAGCACTG TTTCCGTGCG GGGCGCGAGA GCGGTACCAA 910
 ATCAGGCGAA ACTCTGAATA CTAGATATGA CCTCAAAATA ACAGGGGTCA AGGTCCGCCA GTGAGACGAT 980
 GGGGGATAAG CTTCATCGTC GAGAGGGAAA CAGCCCGGAT CACCAGCTAA GGGCCCTAAA TGACCGCTCA 1050
 GTGATAAAGG AGGTAGGGGT GCAGAGACAG CCAGGAGGTT TGCCTAGAAG CAGCCACCCT TGAAAGAGTG 1120
 CGTAATAGCT CACTGATCGA GCGCTCTTGC GCCGAAGATG AACGGGGCTA AGCGATCTGC CGAAGCTGTG 1190
 GGATGTAAAA AGGCATCGGT AGGGGAGCGT TCCGCCCTAG GGGGAAGCAC CTGCGCGAGC AGCGGTAGAC 1260
 GAAGCGAAG CGAGAATGTC GCGTTGAGTA ACGCAAACAT TGGTGAGAAT CCAATGCCCC GAAAACCTAA 1330
 GGGTCTCTCC GCAAGGTTG TCCACGGAGG GTGAGTCAGG GCCTAAGATC AGGCCGAAAG GCGTAGTCGA 1400
 TGGACAACAG GTGAATATTC CTGTACTGCC CTTGTGTTGT CCCGAGGGAC GGAGGAGGCT AGGTTAGCCG 1470
 AAAGATGGTT ATCGGTTCAA GAGACGAAGG TGCCCTTGCT TTTTCAGGTT AAGAAGGGGT AGAGAAAATA 1540
 CCTCGAGCCA ATGTTCCAGT ACCAGGCGCT ACGGCGCTGA AGTAACCCAT GCCATACTCC CAGGAAAAGC 1610
 TCGAACGACC TTCAACAAAA GGGTACCTGT ACCCGAACC GACACAGGTA GGTAGGTAGA GAATACCTAG 1680
 GGGCGCGAGA CAACTCTCTC TAAGGAACTC GGCAAAATAG CCCCCTAACT TCGGGAGAAG GGGTGCCTCC 1750
 TCACAAAGGA GGTGCGAGTG ACCAGGCCCG GCGGACTGTT TACCAAAAC ACAGGTCTCC GCAAAGTCGT 1820
 AAGACCATGT ATGGGGCTG ACGCCTGCCC AGTGCCGGAA GGTCAAGGAA GTTGGTGACC TGATGACAGG 1890
 GGAGCCAGCG ACCGAAGCCC CGGTGAACGG CGGCCGTAACT TATAACGGTC CTAAGGTAGC GAAATTCCTT 1960
 GTCGGGTAAG TTCCGACCG CACGAAAGGC GTAACGATCT GGGCACTGTC TCGGAGAGAG ACTCGGTGAA 2030
 ATAGACATGT CTGTGAAGAT CGCGACTACC TGCACCTGGA CAGAAAGACC CTATGAAGCT TCACTGTTCC 2100
 CTGGGATTGG CTTTGGGCTT TTCTGCGCA GCTTAGGTGG AGGGCGAAGA AGGCCTCCTT CCGGGGGGGC 2170
 CCGAGCCGTC AGTGAGATAC CACTCTGGAA GAGCTAGAAT TCTAACCTTG TGTGAGGACC TACGGGCCAA 2240
 GGGACAGTCT CAGGTAGACA GTTCTATGG GCGCTAGGCC TCCCAAAAGG TAACGGAGGC GTGCAAAGGT 2310
 TTCCTCGGGC CAGACCGAGA TTGGCCCTCG AGTGCAAAGG CAGAAGGGAG CTTGACTGCA AGACCCACCC 2380
 GTCGAGCAGG GACGAAAGTC GGCCTTAGTG ATCCGACGGT CCCGAGTGA AGGGCCGTCG CTCACCGGAT 2450
 AAAAGTTACT CTAGGGATAA CAGGTGATC TTCCCAAGA GCTCACATCG ACGGGAAGGT TTGGCACCTC 2520
 GATGTGGCT CTTCGCGACC TGGGGCTGTA GTATGTTCCA AGGGTTGGGC TGTTCGCCCA TTAAGCGGT 2590
 ACGTGAGCTG GGTTCAGAAC GTCGTGAGAC AGTTCGGTCC ATATCCGGTG TGGGCGTTAG AGCATTGAGA 2660
 GGACCTTCC CTAGTACGAG AGGACCGGGA AGGACGCACC TCTGGGTGAC CAGTTATCGT GCCACCGTA 2730
 AACGCTGGGT AGCCAAGTGC GGAGCGGATA ACTGCTGAAA GCATCTAAGT ACTAAGCCCA CCCCAAGATG 2800
 AGTGCTCTCC T 2811

图3 凯木(*Alnus incana*)叶绿体 23S rRNA 基因的核苷酸序列。(依 Levesque 和 Johnson, 1992)

拷贝区接合处。玉米 $tRNA^{Val}$ 基因差不多是在从 16S rDNA 起点上行 300bp 处。

Yu 和 Spreitzer (1991) 从衣藻 (*Chlamydomonas reinhardtii*) 克隆出叶绿体的编码 $tRNA^{Arg}$ 的 $trnR-ACG$ 基因并进行了序列分析。利用能鉴别反密码子区的合成的寡聚核苷酸探针他们证明此基因存在于叶绿体 *EcoRI* 片段 1 的 1056 个 bp 的 *Hind III*—*Hind III* 亚片段上。 $trnR$ DNA 序列也形成二级结构。

Yu 等 (1992) 又研究了衣藻叶绿体编码 $tRNA^{Ser}$ 和 r -蛋白 L20 的基因的序列。在他们所分离出来的 *Eco*—29 片段上含有 $trnW-CCA$, $trnS-UGA$ 和 $rpl20$ 三个基因。

在 *cpDNA* 的 16S 和 23S rDNA 之间也含有 $tRNA$ 基因。眼虫的 16S—23S rDNA 区间长 258 个 bp 左右, 基因是由 5'→3' 组建的, 也就是 16S rDNA—87bp 区间— $tRNA^{Asp}$ —9bp 区间— $tRNA^{Ala}$ —15bp 区间—23S rDNA。

在叶绿体 $tRNA$ 基因中有内含子最先是在玉米的 $trn1$ 和 $trnA$ 中证明的。以后在烟草的少数叶绿体基因 (6 个 trn , 4 个 r -蛋白基因, $atpF$, $petB$, $petD$, $ndhA$, B) 中发现有内含子。

而拟南芥菜分别有 6 个 $tRNA$ 基因和 12 个蛋白质基因有内含子。有些内含子非常长 (几百个碱对), 虽然其所编码的 $tRNA$ 才不过 70—80 个核苷酸。应当指出的是, 玉米内含子 I 和 II 序列分别含有 123 和 45 个密码子的开放读框 (ORF)。烟草内含子 I 和 II 分别含有 71 和 36 个密码子的 ORF。

表 2 叶绿体核蛋白体蛋白质 (r -蛋白) 基因

基因	30S 小亚基蛋白	基因	50S 大亚基蛋白
$rpS2$	S2	$rpL2$	L2
$rpS3$	S3	$rpL14$	L14
$rpS4$	S4	$rpL16$	L16
$rpS7$	S7	$rpL20$	L20
$rpS8$	S8	$rpL22$	L22
$rpS11$	S11	$rpL23$	L23
$rpS12$	S12	$rpL33$	L33
$rpS14$	S14	$rpL36$	L36
$rpS15$	S15		
$rpS16$	S16		
$rpS18$	S18		
$rpS19$	S19		

(依 Sugiura, 1989)

1. 2. 蛋白质基因

1. 2. 1. r -蛋白的基因 质体核蛋白体的生物发生比较复杂, 由核基因组和质体基因组共同编码。叶绿体 70S 核蛋白体有 58 至 62 个核蛋白体蛋白质 (r -蛋白)。其中有三分之一是叶绿体编码的。在这些基因中, $rpS7$, $rpL2$ 和 $rpL23$ 存在于倒置重复区。 $rpS12$, $rpS16$, $rpL2$ 和 $rpL6$ 含有 536 至 1020bp 长的内含子。眼虫的 $rpS12$ 由 125 个密码子组成, 而且没有内含子。烟草的 $rpS12$ 基因是分开的, 它的外显子存在于倒置重复 (IRA 和 IRB) 中。

烟草的叶绿体 *r*-蛋白基因列于表 2 内。*rpL23*, *rpL2*, *rpS19*, *rpL22*, *rpS3*, *rpL16*, *rpL14*, *rpS8*, *rpL36* 和 *rpS11* 是按这样的顺序成串排列的,这与大肠杆菌的情形非常类似。

编码 *S15* 的基因在质体 DNA 上的位座,烟草和地钱是一个,黑麦、水稻和玉米是两个。*Fitzky* 和 *Subramanian* (1990) 对玉米叶绿体 *rpS 15* 基因 (*IR-II* 位座) 进行了核苷酸序列分析。基因是利用黑麦 *rpS 15* 基因作为探针从玉米叶绿体 DNA 的 *EcoRI* 降解物经 *Southern* 印迹分析鉴定的。杂交区亚克隆成为 *EcoRI/EcoRV* 片段 (600 bp), 然后两股的序列都作了分析 (图 4)。假定的 *rpS 15* 长 270 bp, 从 96 处的 AUG 密码子开始。第二个 AUG 密码子出现在 132 处, 与黑麦和水稻的情形一样。推导的 *S15* 的氨基酸序列同黑麦和水稻, 烟草, 地钱和大肠杆菌分别有 97—98%, 63%, 44% 和 32% 相同。在玉米叶绿体基因组的物理图中 *S15* 基因存在于 6 060—6 550 处 (*IR-II*) 和 132 370—132 860 (*IR-I*) 处。

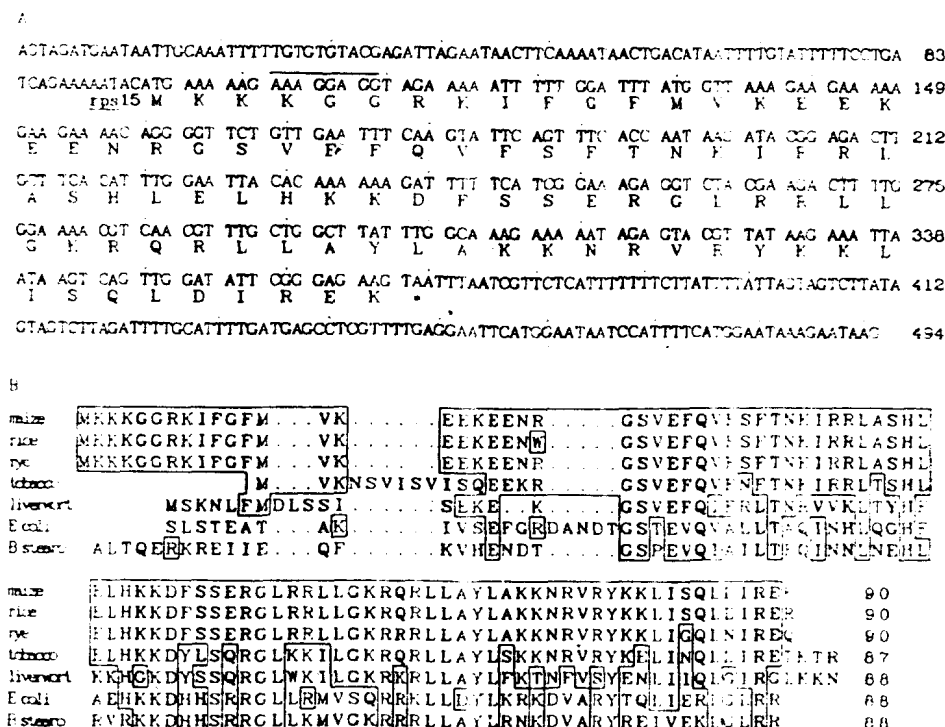


图 4 玉米叶绿体 DNA 编码核蛋白体蛋白质 *S15* 的核苷酸序列 (a) 以及与其他植物 *S15* (b) 的比较。(依 *Fitzky* 和 *Subramanian*, 1990)

Nagano 等 (1991) 对豌豆叶绿体编码核蛋白体蛋白 *L2* 基因 *rpL2* 附近的核苷酸序列进行了分析, 结果发现, 在插入于 *pBR322* 的豌豆叶绿体 DNA 的 9.2kb *PstI* 片段上存在有 *rpL23*、*rpL2* 和 *rpS19* 三个基因。*rpL2* 基因含有 702 个 bp 的内含子, 并且上游序列与 *rpL23* 编码序列类似。他们也研究了 *rpL2* 的表达。在转绿 (去黄化) 过程中, *rpL2* 转录物的积累不需要在叶绿体核蛋白体上新合成蛋白质, 这与大多数叶绿体编码的基因不同。

质体核蛋白体蛋白质的其余三分之二是核编码的, 因为在核蛋白体组装时必需有相同数量的 *r* 蛋白, 所以 *r* 蛋白基因的表达应该受共同机制协调控制。

1.2.2. RNA 聚合酶基因 (*rpoA, B, C*)。从叶绿体也分离出一个大的多亚基 RNA 聚合酶。这个酶含有三个叶绿体编码的亚基, 其基因分别为 *rpoA, B, C*, 与大肠杆菌 RNA 聚合酶

亚基同源。其他亚基是由核编码的。烟草叶绿体 RNA 聚合酶基因与大肠杆菌的 *rpoC* 序列非常类似。Hu 和 Bogorad (1990) 证明, 玉米叶绿体的 *rpoC2*、*rpoB* 和 *rpoA* 基因的产物是 RNA 聚合酶的 180、120 和 38kDa 多肽, 这几种多肽是酶的有功能的组份。此外, 玉米叶绿体 RNA 聚合酶也富有 140、100、95 和 85kDa 多肽。某些种类的叶绿体染色体也被证明有 26—50% 编码 RNA 聚合酶蛋白的基因。



图5 白薯 *rbcL* 的核苷酸序列和推导的氨基酸序列。(依 Haberhausen 和 Zetsche, 1992)

1. 2. 3. *rbcL* (核酮糖 1,5-双磷酸羧化酶/加氧酶或 Rubisco 大亚基基因)。Rubisco 是叶绿体内光合 CO₂ 固定的关键酶。高等植物 Rubisco 是由八个 (Mr 52000—55000) 大亚基和八个 (

Mr12000—15000) 小亚基组成的。大亚基由叶绿体编码,小亚基由核编码,前者在叶绿体内合成,小亚基在细胞质内合成,然后进入叶绿体,加工并同大亚基结合,形成全酶。每个大亚基的分子量约为 55kDa,每个小亚基的分子量约为 14kDa,大亚基含有催化位。*rbcL* 基因存在于大的单拷贝区,同 ATP 酶的 β 亚基和 ϵ 亚基的基因邻近。菠菜、撞羽朝颜、苜蓿、大麦、玉米、豌豆、烟草以及番茄等的 *rbcL* 基因的序列都被进行了分析。高等植物的 *rbcL* 是连续的,长度约为 1.4kb,所编码的多肽差不多有 475 个氨基酸。在不同种的 *rbcL* 之间编码序列是高度保守的,只在基因的 3' 端核苷酸有变化。许多碱基的变化是沉寂的,因此在氨基酸序列之间是高度保守的,例如,在玉米和菠菜的 *rbcL* 核苷酸序列之间的同源性是 84% 而相应的氨基酸序列之间的同源性是 90%。亲缘关系近的撞羽朝颜和烟草之间氨基酸序列有 97% 是相同的,而且 5' 和 3' 旁区的核苷酸序列也是高度相同的。

眼虫叶绿体 DNA 为 Rubisco 大亚基编码的基因含有 9 个插入序列。这是借在眼虫 *rbcL* 和菠菜不含内含子的 *rbcL* 之间的异源双股(heteroduplex)分析、眼虫 *rbcL* DNA—mRNA 杂交体的电泳术以及克隆、限制酶分析和部分 DNA 序列分析等确定的。10 个外显子中有 6 个,其定位和克隆特性都已被鉴定。在 6.5kb *rbcL* 中 9 个内含子的每个长度均接近于 0.5kb 内含子和外显子交界处的 DNA 序列都是高度保守的。

表 3 光系统 I 多肽的叶绿体基因

基因名称	基因产物	密码子数
psbA	32 kDa Q_B 蛋白, D1	353
psbB	47 kDa 叶绿素 a 蛋白	508
psbC	44 kDa 叶绿素 a 蛋白	461
psbD	34 kDa 蛋白, D2	353
psbE	9 kDa 细胞色素 b—559	83
psbF	4 kDa 细胞色素 b—5	39
psbG	24 kDa 多肽	248
psbH	9—10 kDa 磷蛋白	73
psbI	4.5 kDa 反应中心多肽	36
psbJ	假定的蛋白	53
psbK	3.9 kDa 多肽	37
psbL	4.8 kDa 多肽	38
PsbM	M—蛋白	
PsbN	N—蛋白	

(依 Gray 等, 1990)

Haberhausen 和 Zetsche(1992)对白薯(*Ipomoeea purpurea* L.)的 *rbcL* 基因进行了序列分析(图 5)。并且将编码区同烟草、豌豆、玉米和鱼腥草的相应序列进行了比较。同烟草已报导的氨基酸序列有 95% 相同。并且—35 和—10 区均非常保守。

袁卫明等(1992)对籼稻叶绿体 *rbcL* 基因进行了分子克隆并测定了全序列,结构基因全长为 1434bp,编码 477 个氨基酸。基因 5' 端启动子区和 *atpB* 基因 5' 端启动子相隔 159 个核苷酸