

叶绿体基因组的结构研究进展

蒋 达 和

(武 汉 大 学)

提 要

本文根据最近由两组日本科学家首次测定出来的两种植物(地钱和烟草)叶绿体基因组的完整序列,通过比较孢子植物(地钱)和被子植物(烟草)的叶绿体基因组,主要介绍叶绿体基因组中的基因表达系统、能量系统、代谢物质运输系统和RNA加工系统的基因及其研究近况。

关键词 叶绿体,光合作用, DNA 序列,基因组, RNA 加工

光合作用是地球上最大规模地贮存太阳能的过程,也是规模最大的积累有机物的过程,为人类提供物质和能量的主要来源。而叶绿体是光合作用的重要细胞器,它能将光能转变成化学能和生物能,用于固定空气中的CO₂合成各种有机物。以前人们对叶绿体的作用机制、遗传控制、进化历史及其与核基因组的关系等都了解不多,关键的问题就是对叶绿体基因组了解甚少。1986年,两组日本科学家分别报道了两种植物叶绿体基因组的完整序列,这无疑是在叶绿体生物学研究的重大进展,所测定的完整DNA序列,为全面地了解上述问题,对于叶绿体分子生物学和遗传学研究以及改造叶绿体、提高光合效率,都具有重要价值。本文主要介绍叶绿体基因组的结构及其研究的新进展。

一、叶绿体基因组遗传图

植物叶绿体基因组为双链环状DNA分子,长约120—160千碱基对(kbp),分子内具有一对倒转重复序列(IR_A和IR_B),之间由一个大单拷贝(LSC)区和一个小单拷贝(SSC)区分隔开(图1)。通过比较地钱(*liverwork*)叶绿体基因组(121024 bp)^[1]和烟草(*tobacco*)叶绿体基因组(155844 bp)^[2]的序列,发现烟草

叶绿体基因组比地钱叶绿体基因组大25%,其基因组大小的差异主要是IR区的大小不同所致(地钱为10058 bp;烟草为25339 bp)。基因组结构的主要区别是在

叶绿体基因组含有4种rRNA基因和31种(地钱)或30种(烟草)tRNA基因以及约90种各类蛋白质基因,在这90种蛋白质基因中有50多个是通过与已知的叶绿体、线粒体和细菌的蛋白质氨基酸序列进行比较而确定的,还有30多种未确定其功能的开放读框(ORF)基因,在地钱和烟草两种叶绿体基因组中显示出几乎相同的氨基酸序列,暗示着它们是有某种功能的基因。虽然地钱(属于苔藓植物)和烟草(被子植物)在进化上相距很远,但其叶绿体基因组结构和基因数目如此相似,意味着所有绿色植物叶绿体基因组都是从同一祖先进化来的。

二、光合作用和电子传递的基因

目前在各种植物的叶绿体中已确定了 20 个编码电子传递和光合固碳的基因：二磷酸核酮糖羧化酶的大亚基 (*rbc L*)、光合系统 I 的两个亚基 (*psa*)、光合系统 II 的 8 个亚基 (*psb*)、 H^+ -ATP 酶的 6 个亚基 (*atp*)、细胞色素 *b/f* 复合物的 3 个亚基 (*pet*)。光合作用的基因在基因组中主要以若干小簇排列。如：*psaA-psaB*、*psbD-psbC*、*psbE-psbF*、*atpB-atpE*、*atpI-atpH-atpF-atpA* 和 *psbB-psbH-petB-petD* 等 (图 1)。这种结构在地钱和烟草的叶绿体基因组中都是很保守的。这些基因簇不仅是表达

的转录单位,而且也是调控单位,其中有的是重叠基因,如在 *AtpB* 操纵子中,编码 H^+ -ATPase 的 β 亚基的 *atpB* 基因其终止密码子就与编码 e 亚基的 *atpE* 基因的起始密码子重叠。最近 Kohchi 等^[3]首次在地钱叶绿体基因组中发现位于 *psbB* 操纵子互补股 DNA 链上的 ORF₄₃ 基因,其 mRNA 的 5' 前导序列是差向重叠的,由于这种重叠是位于互补链上,因此把 *psbB* 操纵子和 ORF₄₃ 基因的转录称为差向重叠转录 (divergent overlapping transcription)。操纵子中存在差向重叠转录现象对操纵子的基因表达是具有一定的调控作用的。

通过分析这两种叶绿体完整 DNA 序列,

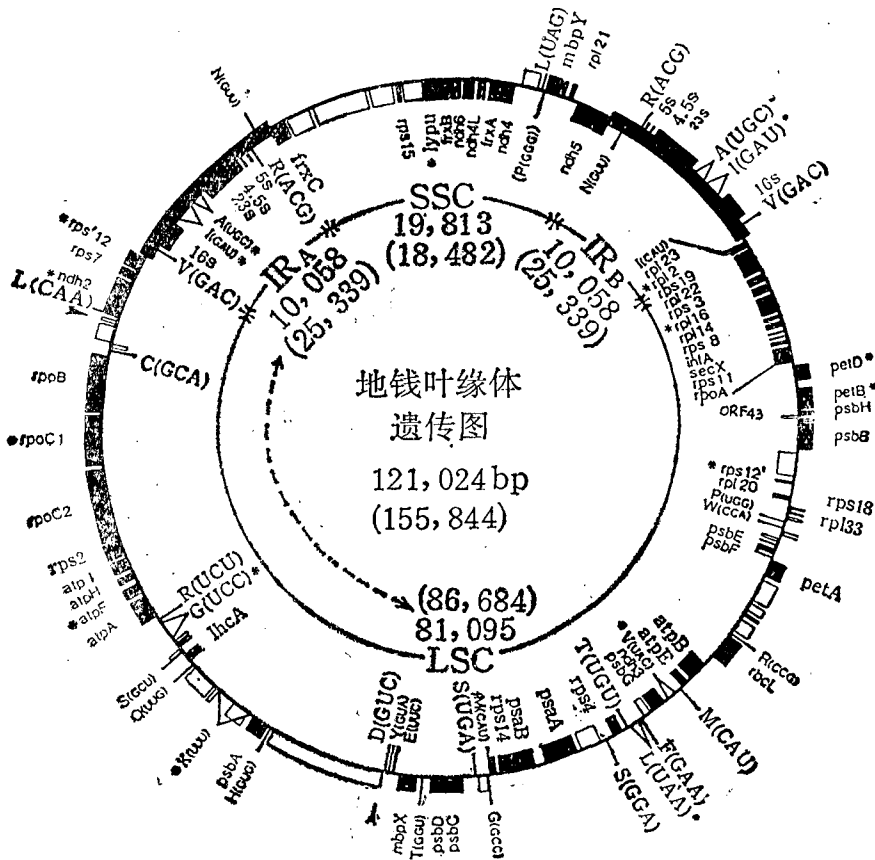


图 1 地钱叶绿体基因组的遗传图

图外圈所示基因以反时针方向转录,内圈基因以顺时针方向转录,基因缩写符号解释见正文。图内的圆括号内之数目字为烟草叶绿体 DNA 的核苷酸数,虚线和箭头表示烟草叶绿体基因组的倒转排列区。空白框架表示未确定功能的 ORF 基因,星号表示在编码区内含有内含子的基因。(此图引自 Ohyama K *et al.* *TIBS*, 1988; 13; 20)

又发现两组新基因: *ndh* 和 *frx* 基因。两种叶绿体基因组中都有 7 个 *ndh* 基因 (*ndh*₁, *ndh*₂,

*ndh*₃, *ndh*₄, *ndh*_{4L}, *ndh*₅ 和 *ndh*₆) 相当于人线粒体 NADH 脱氢酶基因 (*ND*₁, *ND*₂, *ND*₃,

ND₁, ND_{4L} 和 ND₆)^[4]。DNA-RNA 分子杂交分析表明,叶绿体中的 *ndh* 基因都具转录活性,在莱茵衣藻 (*Chlamydomonas reinhardtii*) 中已证实 *ndh* 基因是编码 NADH-质体醌 (PQ) 氧化还原酶,当用 PQ 氧化抑制剂阻断光合系统 I 和 II 之间的电子传递通道时,叶绿体呼吸的电子传递则不受影响^[5],说明叶绿体含有不同于光合作用的质体醌为氧化还原载体氧化 NADH 的电子传递链。

地钱叶绿体中有三个 *frx* 基因 (*frxA*、*B* 和 *C*)。其中的 *frxA* 和 *frxB* 相当于细菌的 4Fe-4S 型铁氧还蛋白 (ferredoxin), 具有特征性的氨基酸序列: -Cys-N-N-Cys-N-N-Cys-N-N-N-Cys-Pro-。烟草的 ORF₁₆₇ 相当于地钱的 *frxB*, 而两者的 *frxA* 基因仅相差一个核苷酸,表明 *frxA* 和 *B* 基因在大多数叶绿体中都存在。最近已从菠菜 (spinach) 叶绿体中分离出 *frxA* 基因产物, 分子量为 8 kDa, 是光系统 I 的成分^[6]。表明在叶绿体电子传递系统中除了由核基因编码的 2Fe-2S 型铁氧还蛋白外, 还

有叶绿体本身编码的 4Fe-4S 型铁氧还蛋白。*frxB* 基因产物还未鉴定。地钱 *frxC* 基因产物与固氮菌 *nifH* 基因编码的一种铁蛋白即固氮酶的还原酶具有高度同源性^[7], 但 *frxC* 基因产物能否固定分子氮, 是否还有别的功能, 都有待于进一步研究。烟草叶绿体中未检测到 *frxC* 基因。

三、运输系统的基因

由于叶绿体是一种由内外两层膜构成的细胞器, 在膜上肯定存在输送各种代谢物质的蛋白质。地钱叶绿体的 *mbpX* 和 *mbpY* 基因编码的氨基酸序列与伤寒沙门氏菌 (*Salmonella typhimurium*) 组氨酸输送系统 *hisP* 和 *hisQ* 基因产物的氨基酸序列具有高度的同源性^[8], 也类似于 *E. coli* 麦芽糖输送系统的 *malK* 和 *malF* 基因产物, 因而推测 *mbp* 基因产物可能是一种透性酶 (permease)。*mbpX* 基因产物的 N 末端具有类似于动物和细菌的核苷酸驱动输送系统 (nucleotide-driven transport system)。

表 1 示线粒体 (mt.)、叶绿体 (chl.) 和大肠杆菌 (*E. c.*) 含有的 tRNA 分子

密码子	氨基酸	tRNA 的反密码子			密码子	氨基酸	tRNA 的反密码子			密码子	氨基酸	tRNA 的反密码子			密码子	氨基酸	tRNA 的反密码子		
		mt.	chl.	E. c.			mt.	chl.	E. c.			mt.	chl.	E. c.			mt.	chl.	E. c.
UUU	Phe				UCU				UAU				UGU	Cys					
UUC		GAA	GAA	GAA	UCC	Ser		GGA	GGA	UAC	Tyr	GUA	GUA	GUA	UGC		GCA	GCA	GCA
UUA	Leu	UAA	UAA	(NAA)	UCA		UGA	UGA	UGA	UAA	-				UGA	-	UCA		
UUG			CAA	CAA	UCG			CGA	CGA	UAG	-				UGG	Trp		CCA	CCA
CUU					CCU				CAU						CGU		ACG	ACG	ACG
CUC	Leu			GAG	CCC	Pro	(ggg)		CAC	His	GUG	GUG	GUG	CGC	Arg				
CUA		UAG	UAG	UAG	CCA		UGG	UGG	CAA	Gln	UUG	UUG	UUG	CGA					
CUG				CAG	CCG			CGG	CAG				CUG	CGG		(CCG)	CCG		
AUU	Ile				ACU				AAU	Asn				AGU	Ser				
AUC		GAU	GAU	GAU	ACC	Thr	GGU	GGU	AAC		GUU	GUU	GUU	AGC		GCU	GCU	GCU	
AUA	Met _f		CAU	CAU	ACA		UGU	UGU	AAA	Lys	UUU	UUU	UUU	AGA	Arg	UCU	UCU	UCU	
AUG		CAU	CAU	CAU	ACG			CGU	AGA					AGG				CCU	
GUU					GCU				GAU	Asp				GGU					
GUC	Val		GAC	GAC	GCC	Ala		GCC	GAC		GUC	GUC	GUC	GGC		GCC	GCC	GCC	
GUA		UAC	UAC	UAC	CCA		UGC	UGC	GAA	Glu	UUC	UUC	UUC	GGA	Gly	UCC	UCC	UCC	
GUG					GCG				GAG					GGG				CCC	

注: 酵母线粒体中 CUN 密码代表苏氨酸而不是亮氨酸, UGA 是色氨酸密码子。大肠杆菌的 NAA 意义不明, 因为 N 似乎是 A 的衍生物。叶绿体的 (ggg) 在地钱中是一个假基因, 而 (CCG) 至今仅在地钱叶绿体中发现。

组分中的氨基酸序列^[9], 这表明在地钱叶绿体中可能存在一种活性的依赖于蛋白质的核苷酸驱动的输送系统。mbp 基因也未在烟草叶绿体中检测到, 可见, mbp 基因与 frx C 一样, 在进化上都是不保守的。

四、蛋白质合成的基因

叶绿体蛋白质合成机构由核基因产物和叶绿体基因产物两部分组成。约 1/2 的叶绿体基因是编码该机构成分的基因, 其中包括 rRNA、tRNA、RNA 多聚酶和核糖体蛋白质等的基因。

叶绿体核糖体与原核细胞核糖体在大小、抗菌素敏感性和交叉免疫反应等方面有许多相同之处。活性的 70S 核糖体也是由 30S 和 50S 两个亚基构成, 含有 4 种 rRNA (16S, 23S, 5S 和 4.5S)。4.5S rRNA 为植物叶绿体所特有, 它与大肠杆菌 23S rRNA 的 3' 端部分相同^[10]。叶绿体的所有 rRNA 基因都位于 IR 区, 以两个相同基因群的形式排列, 每个基因群中间由间隔子 trn K (GAU) 和 trn A (UGC) 基因分隔开(见图 1)。

分子杂交实验证实, 全部的叶绿体 tRNA 都是由叶绿体基因组编码的。地钱和烟草叶绿体都具有 30 种 tRNA 基因, 地钱还另外有一种活性的 tRNA^{Asp} (CCG) 基因和 tRNA^{Phe} (ggg) 假基因(见表 1)。显而易见, 叶绿体不需要核基因编码的 tRNA 也能完成转运各种氨基酸去合成蛋白质的任务。此外, 有实验证实, 叶绿体内叶绿素的生物合成需要一种叫 RNA^{DALA} 的成分, 该成分与叶绿体编码的 tRNA^{Glu} (UUC) 相同^[11]。由于叶绿体基因组中编码的 tRNA^{Glu} (UUC) 只有单一的一个基因, 因而有理由推测, 该基因产物可能具有叶绿素和蛋白质生物合成的双重功能。

叶绿体核糖体蛋白则只有一部分是由叶绿体基因组编码, 已证实地钱中有 19 种核糖体蛋白基因, 主要是根据其基因所推测出的氨基酸序列与 *E. coli* 核糖体蛋白进行比较而确定的。如叶绿体的 10 个 rps 基因 (30S 亚基) 编码 *E. coli* S₂, S₃, S₄, S₇, S₈, S₁₁, S₁₂, S₁₄, S₁₅, S₁₈ 和 S₁₉

的类似物; 8 个 rpl 基因 (50S 亚基) 编码 *E. coli* L₂, L₄, L₁₆, L₂₀, L₂₁, L₂₂, L₂₃ 和 L₃₃ 的类似物。烟草叶绿体基因组缺 rpl₂₁, 而另外编码 rps₁₆。这些结果与菠菜叶绿体的生化分析较一致, 菠菜叶绿体核糖体蛋白 1/3 是叶绿体基因组编码, 其余都是由核基因编码^[12]。

以 *E. coli* 翻译起始因子 I 和 X 蛋白质基因(最近已证实该基因是 *E. coli* 50S 核糖体亚基蛋白的一个新基因^[13])的同源性为基础, 确定了叶绿体的 inf A 和 sec X 基因。莱茵衣藻和纤细裸藻 (*Euglena gracilis*) 等绿藻叶绿体具有能编码延伸因子 Tu 的 tuf A 基因^[10], 但在其他植物叶绿体中还未发现该基因的存在。

目前, 叶绿体的转录基因还不太清楚。虽然地钱叶绿体中已检测到四种(烟草为三种)相当于编码 *E. coli* RNA 多聚酶核心酶亚基的 rpo 基因: rpo A 编码 α 亚基, rpo B 编码 β 亚基, rpo C₁ 和 C₂ 分别编码 β' 亚基的 N-和 C-末端部分, 但未发现有类似 *E. coli* RNA 多聚酶 σ 亚基的叶绿体基因。而 Lerbs 等^[14]从菠菜叶绿体中提纯了 RNA 多聚酶, 并证实所有亚基都是从带有多聚 (A)⁺ 的 mRNA 翻译来的, 说明这是核基因的产物。如何将这两种结果统一起来, 有人在叶绿体中检测到两种不同的 RNA 多聚酶活性, 认为核基因编码的多聚酶可选择性地转录 rRNA 基因。而叶绿体编码的多聚酶则可能是转录其他的基因。

五、RNA 加工及其基因

叶绿体基因及其表达方式与原核系统有许多共同的特征, 但叶绿体基因含有内含子(这在原核中极少), 因而在表达过程中需要对转录后的 RNA 分子进行拼接加工。地钱和烟草叶绿体基因组中, 有 6 种 tRNA 和 9 种蛋白质基因都具有内含子, 另外地钱还有三个 ORF (烟草为二个 ORF) 也含内含子。

对于 mRNA 前体中内含子的加工, 以前的许多报道都是在同一条 RNA 链上进行分子内的剪接即以顺式拼接 (cis-splicing) 方式加工, 绝大部分具内含子的叶绿体基因转录产物都是

以这种方式进行加工^[15]。最近在叶绿体基因组中发现编码 30 S 核糖体蛋白 S₁₂ 的 mRNA 则是在分子间以反式拼接 (trans-splicing) 机制进行加工形成的。该基因由三个外显子组成, 第一个外显子 (图中的 rps 12') 与第二和第三个外显子 (rps' 12) 之间由数万 bp 的内含子 (内含子中又有多个基因) 隔开^[16]。这种结构也在许多植物叶绿体基因组中存在。被隔断的基因分别转录, 两个 RNA 分子通过反式拼接机制加工就可以形成完整的 mRNA 分子^[16]。最近报道, 衣藻的 psa 基因也是以反式拼接机制进行加工的^[17]。

已经证实, 基因中的 I 类和 II 类内含子都可以进行体外自我拼接, 这表明催化反应的活性存在于内含子 RNA 本身。但在叶绿体中还未检测到内含子的体外自我拼接现象, 说明叶绿体内含子的拼接反应可能需要蛋白质因子参与。

原核系统中内含子的存在, 提出了一个使人困惑的问题, 即在相当于核膜的任何类似结构存在下, 细胞器内的核糖体是如何将 mRNA 与其前体区别开来, 有人认为, 叶绿体含有两种核糖体即基质核糖体和结合类囊体膜核糖体, 分别具有翻译不同产物的功能^[18], 这可能意味着它们可以选择 mRNA, 而拼接反应就是选择过程的一个部分。

叶绿体的成熟 tRNA 和 rRNA 也不是直接转录形成的, 也有一个加工过程。用叶绿体提取物证实, RNA 加工酶 RNase P 具有精确加工 tRNA 分子 5' 末端的活性^[19]。该酶由蛋白质和 RNA 两种组分构成, Shinozaki 等^[20]观察到叶绿体 RNase P 的活性对绿豆 (mung bean) 核酸酶敏感, 表明酶中的 RNA 组分也与 RNase P 活性有关。但最近 Wang 氏等^[21]用小球菌核酸酶 (micrococcal nuclease) 处理菠菜叶绿体酶提取物, 发现仍具有 RNase P 活性, CsCl 密度梯度离心表明叶绿体 RNase P

分子量亦小于大肠杆菌的 RNase P, 因而认为叶绿体 RNase P 可能不同于细菌的 RNase P, 也不含有 RNA 成分。目前由于还未测出 RNase P 的完整序列, 所以不清楚叶绿体是否具有编码该酶基因。

六、展 望

地钱和烟草两种完整 DNA 序列的测定, 是叶绿体分子生物学和分子遗传学的良好开端, 它有可能为人工改造叶绿体及叶绿体遗传工程提供基础理论依据。当然, 目前还存在许多问题, 如叶绿体的基因组中还有许多 ORF 基因未确定其功能; 叶绿体 DNA 的复制及其复制基因仍然是一个谜; 其基因的表达和调控及叶绿体呼吸电子传递系统等都了解甚少。这些都有待于进一步研究才能解决。

本文承蒙卢文筠教授指导和审阅, 特此致谢。

参 考 文 献

- 1 Ohyama K *et al.* *Nature*, 1986; 322: 572
- 2 Shinozaki K *et al.* *EMBO J*, 1986; 5: 2043
- 3 Kobchi T *et al.* *EMBO J*, 1988; 7: 885
- 4 Chomyn A *et al.* *Science*, 1986; 234: 614
- 5 Maione T E *Plant Physiol*, 1986; 80: 364
- 6 Oh-Oka H *FEBS Lett*, 1987; 218: 52
- 7 Haselkora R *Annu Rev Microbiol*, 1986; 40: 525
- 8 Higgins C F *et al.* *EMBO J* 1985; 4: 1033
- 9 Higgins C F *et al.* *Nature* 1986; 323: 448
- 10 Palmer J D *Annu Rev Genet*, 1985; 19: 325
- 11 Schön A *et al.* *Nature*, 1986; 322: 281
- 12 Dorne A M *et al.* *Plant Mol Biol*, 1984; 3: 83
- 13 Wade A *et al.* *J Biochem.* 1987; 101: 817
- 14 Lerbs S *et al.*: *EMBO J* 1985; 4: 1661
- 15 Tanaka M *et al.* *MGG*, 1987; 209: 427
- 16 Koller B *et al.* *Cell*, 1987; 48: 111
- 17 Kück U *et al.* *EMBO J*, 1987; 6: 2185
- 18 Minami E *et al.* *Arch Biochem Biophys*, 1984; 235: 562
- 19 Greenberg B M *et al.* *Plant Mol Biol*, 1984; 3: 97
- 20 Yamaguchi-Shinozaki K *et al.* *FEBS Lett*, 1987; 215: 132
- 21 Wang M J *et al.* *EMBO J*, 1988; 7: 1567

【本文于1988年9月20日收到】