专论与综述

绿色植物分子生物学研究进展

胡友纪　王世杰
(中国科学院上海植物生理研究所)

七十年代以来由于重组 DNA 技术的建立，使真核生物特定基因的分离有了可能．随着基因结构分析技术的飞速发展，分子生物学的研究出现了崭新的局面，同时绿色植物基因结构和基因工程的研究也有了很大的发展．近年来对植物核基因——贮藏蛋白基因、细胞器——叶绿体，线粒体基因，以及有希望成为植物系统载体的 Ti 质粒 DNA 和花椰菜花叶病毒 DNA 的研究，都取得了显著的进展．

一、植物细胞核基因组

植物基因组的大小随不同植物而异，一般为 2×10^8—8×10^{10} 碱基对，相当于大肠杆菌基因组的 10^3—10^4 倍．但是植物核基因组和其他真核生物基因组一样，在庞大的基因组中约有 80% 的顺序位于基因之间的间隙区 (Spacer)，它们是由没有编码能力的重复顺序组成，这种间隙区的重复顺序在 DNA 特定区域内成簇地连续反复，拷贝数高达 10^6 以上．除了间隙区(非编码区)的重复顺序，在编码区还有另一类重复顺序，即重复基因．这类重复顺序拷贝数较少，一般为 50—30,000个[1]．例如 rRNA 基因，组蛋白基因．也有拷贝数只有几个到十几个的重复基因，如贮藏蛋白基因．R. Havel 研究禾本科植物核基组的重复顺序，并提出了它形成的假说，他认为植物核 DNA 非编码区在进化过程中由于增幅或缺失，较早地改变了顺序，进而通过顺序上特定部位的转位，产生新的重复顺序．植物核基因组顺序的改变不仅发生在非编码区，也发生在编码区．B. Burr 证明玉米 DNA 的“皱缩区”是蔗糖合成酶的编码部位，它可由转座子 (Transposable element) 使皱缩部位的结构缺失或插入而发生改变．

植物核基因组结构中目前研究较多的是 rRNA、贮藏蛋白、RuBP 羧化酶小亚基基因．植物细胞质核糖体属于 80S 核糖体，由 60S 大亚基和 40S 小亚基组成，大亚基含 25S、5.8S、5S rRNA、小亚基含 17S 或18S rRNA．其中 17S、25S、5.8S rRNA 基因在核基因组上成簇连续排列，组成一个重复单位，例如甘薯的 rRNA 基因由7800个碱基对组成一个重复单位[3]，这个重复单位也是一个转录单位，在转录时首先转录成一个共同的前体，然后“加工”成各种 rRNA．植物核 5S rRNA 基因和其他真核生物一样，独自形成一个重复区，例如小麦 5S rRNA 基因是以 420 个和 500 个碱基对，二种重复单位反复排列组成一个约 30,000 碱基对的重复群．植物核 5S rRNA 基因转录起始信息不在 5S rRNA 基因的上游区，而是在 5S 基因内．

植物核基因组中研究最充分的蛋白基因是贮藏蛋白基因．植物种子成熟期形成大量的贮藏蛋白 mRNA，利用 Oligo (dT) 纤维素或 Poly(u) 纤维素亲和层析技术很容易分离和纯化种子贮藏蛋白的 mRNA，然后用反转录作与 mRNA 互补的 C-DNA，C-DNA 克隆后就能作为探针分析基因结构．R. B. Goldberg[4] 对大豆胚中四种 mRNA 的 C-DNA 克隆后分析贮藏蛋白基因的结构，发现贮藏蛋白基因都有重复顺序，这些基因有的有插入顺序，有的没有．

他们还证明这些基因的表达在转录阶段就受到调节控制. T. C. Hall 和 F. A. Bliss[5-9] 检查大豆中云扁豆蛋白基因的结构，证明其中有三个部位有插入顺序. K. A. Marcker[9,10] 比较了蚕豆豆血红蛋白 mRNA 的 C-DNA 和核 DNA 的结构，指出蚕豆豆血红蛋白的基因中也有三个插入顺序. 不论云扁豆蛋白基因或豆血红蛋白基因，插入顺序的 5′-端为 GT，3′-端为 AG，这和至今报道的动物系统的插入顺序极其相似.

现已证明诸如贮藏蛋白之类在细胞中含量丰富的蛋白基因都是重复顺序. K. A. Marcker 证明豆血红蛋白至少有 6 个重复顺序，其中四个是相邻的. A. Bradt 等证明大麦醇溶蛋白基因在第 5 染色体上有二个重复顺序. R. B. Goldberg 报道大豆胚贮藏蛋白重复顺序的位置虽不邻接，但这些基因转录过程是协同进行，在转录阶段接受统一调节.

植物核基因组中被研究的另一种基因是与光合作用二氧化碳同化有关的 RuBP 羧化酶小亚基基因. 该酶是由八个大亚基和八个小亚基组成，两种亚基分别由核基因和叶绿体基因编码，后者是叶绿体基因. S. M. Smith[11,12]等分离了小亚基的 mRNA，并作了它的 C-DNA 克隆. 此 C-DNA 克隆由 766 碱基组成，碱基顺序与 RuBP 羧化酶小亚基的氨基酸顺序比较，小亚基前体蛋白比小亚基蛋白的氨基酸残基多 13 个以上. 用 C-DNA 作探针分析小亚基 mRNA 为 900—1000 个碱基. 小亚基 mRNA 的量在照光时急剧增长，表明基因表达的调节发生在转录阶段.

二、植物细胞器基因组

绿色植物叶绿体基因组的结构和表达是植物基因研究中进展最快的[13-20]. 叶绿体基因组是原核型双链环状 DNA，大小约 135—160 Kbp. 至今已确定叶绿体 rRNA (23S, 16S, 5S, 4.5S)，tRNA (25—30种) 基因，以及 RuBP 羧化酶大亚基和由光诱导产生的 32,000d (P-32) 蛋白的基因(或称光基因)，细胞色素 f，叶绿体偶联因子三种亚基的基因等在叶绿体基因组上的定位，此外还存在 2—5 种核糖体蛋白、细胞色素 B_{559}、延长因子 EFG、EFTu、脂肪酸合成酶、色素蛋白复合体蛋白等近25种蛋白基因，但它们至今尚未在基因组上定位.

各种类型高等植物叶绿体 DNA 内切酶图谱极其相似，多数植物叶绿体 DNA 上存在约 23 Kbp 的二个逆向重复顺序，这是 rRNA 基因区，rRNA 基因集中于此. 在 rRNA 基因间隙区还存在 1—2 个 tRNA 基因. 但在蚕豆和豌豆叶绿体 DNA 上未见此重复顺序，rRNA 基因集中在一个片段中. 眼虫藻叶绿体 DNA 上有三个同向重复顺序，rRNA 基因也集中其上. 由于 rRNA 基因集中排列在重复顺序上，几种 rRNA 基因形成一个转录单位，转录过程先转录成一共同的前体，然后在"转录后"过程再加工、断裂、修饰形成各种 rRNA. J. D. Rochaix 指出衣藻叶绿体 23S rRNA 基因中存在一个约 1000 bp 的插入顺序，并确定了该顺序附近的碱基顺序，指出在编码区存在 CGT 的重复顺序，但此插入顺序与 5′-端是 AA，3′-端是 TG 的 Chambon 规则不一致. 目前在其他植物叶绿体 23S rRNA 基因中尚未见此种插入顺序. 最近篠崎和衫浦测定烟草叶绿体 rRNA 基因起始部位的碱基顺序，指出起始部位中存在原核生物基因中常见的 Pribnow Box 和-35-Base 区的结构. 叶绿体 5S rRNA 基因和兰藻 5S rRNA 基因非常相似，但和细胞质 5S rRNA 基因全然不同.

H. Kössel 研究 rRNA 基因区间隙部位的 tRNA 基因的碱基顺序，指示 16S rRNA 基因上方的 $tRNA^{Val}$ 基因和大肠杆菌 tRNA 基因有 60% 以上的碱基顺序相同. L. Bogorad[19] 指出玉米叶绿体 $tRNA^{Leu}$，$tRNA^{His}$ 基因的顺序和大肠杆菌的碱基结构也有 60—70% 相同. 但在玉米叶绿体 16S 和 23S rRNA 基因之间的 $tRNA^{Ile}$、$tRNA^{Ala}$ 基因中发现有 800—900bp 的插入顺序，此结构在其他原核 tRNA 基因中未见到，因此叶绿体 tRNA 基因结构与原核基因可谓是大同小异.

叶绿体 RuBP 羧化酶大亚基和 P-32 蛋白基因已在叶绿体基因组上定位，并已对它们的结构进行分析. L. Bogorod[20] 把含有 RuBP 羧化酶大亚基基因的 Bam 片断克隆后分析碱基顺序，证明叶绿体蛋白合成是从 AUG 密码子开始，这和线粒体基因不同. 叶绿体基因 AUG 上方 5—9 个碱基的部位和 16S rRNA 的 3′-端相补，转录起始点在距 AUG 顺序上方第 60 个碱基部位. 叶绿体基因也是以 UAA 密码子终止转录，这和线粒体基因情况相同. 最近大亚基因的全碱基顺序测定亦已完成，并由此推算出大亚基 475 个氨基酸残基顺序. 叶绿体为蛋白编码的基因中至今未发现插入顺序.

线粒体是另一类具自身基因组的细胞器. 酵母和动物系统的线粒体基因组精细结构已有充分研究，但植物线粒体基因组和动物线粒体迥然不同，而且不同植物线粒体基因组也各不相同. 最近发现同一植物线粒体中也存在大小各异的环状 DNA[21-23]，例如玉米线粒体中存在周长分别为 30, 22, 16μm 三种环状 DNA 和长度各为 0.8, 0.6, 0.5, 0.47μm 的四种线状分子. 内切酶片段分析也证实了这种异质性. 这种异质性在不同植物中各不相同，但周长 30μm 的环状分子在各种植物线粒体中似乎是普遍存在的.

植物线粒体基因组的编码能力比动物和酵母线粒体大得多. 已确定的植物线粒体基因组编码的基因有线粒体 2S, 18S, 5S, 4S rRNA 和 tRNA 基因[24,25]. 此外线粒体体外蛋白能合成 20 多种蛋白，它们可能也是由线粒体基因组编码. 植物线粒体 rRNA 基因并不集中形成操纵子，例如小麦线粒体 26S, 18S, 5S rRNA 基因都分散在不同的部位. 线粒体 tRNA 基因有一些是成簇排列，另一些则也分散在基因组不同的部位. 线粒体蛋白基因中研究较多的是辅酶 I 和辅酶 II 基因，小麦线粒体辅酶 II 基因已克隆化，碱基顺序也已测定[26]，其顺序与酵母辅酶 II 基因有 47% 相同的顺序，与牛心线粒体辅酶 II 基因有 40% 相同. 植物线粒体辅酶 II 基因中有一794 bp 的插入顺序，显然和叶绿体蛋白基因的特点不同.

三、植物 RNA 病毒基因组

植物病毒也有 RNA 和 DNA 型两类. RNA 病毒中 RNA 既是基因组又是 mRNA. 近年来烟草花叶病毒 (TMV) RNA 基因组的研究取得了很大进展[27-30]. TMV 基因组 RNA 的分子量为 2×10^6，相当于 6400 个碱基，它的 5′ 端有一"帽子"结构. TMV 基因组至少包括四种蛋白基因，然而它的基因表达相当于一个单顺反子，也即基因组中仅靠近 5′-端的蛋白基因能得到解读，其他基因均不能表达，例如 3′-端的外壳蛋白基因. 近年对 3′-端附近的基因的表达有了一些了解. H. Hirth 等从感染植物中分离出一种具有外壳蛋白 mRNA 活性的低分子量 RNA，称 CP-mRNA，并测定了它的一级结构，指出 CP-mRNA 相当于 TMV RNA 3′-端一段约 700 个碱基组成的片段. CP-mRNA 的 5′-端也有一"帽子"结构，这似乎说明外壳蛋白基因是在它的 5′-端从 TMV 基因组 RNA 上脱落，并接上一"帽子"结构后才开始解读. 在外壳蛋白基因 5′-端上方有一分子量 3,000d 的蛋白基因，它的基因表达和 CP-mRNA 极其相似，该蛋白是以称为 I_2-RNA 的一段 RNA 为模板合成的. I_2-RNA 的结构也是该基因的 5′-端从 TMV 基因组上脱落后，接上"帽子"结构组成的 mRNA. 不论 CP-mRNA 或是 I_2-RNA 都是 TMV 基因组 RNA 上的一段接上"帽子"结构组成，因此称这类 mRNA 为亚基因组 mRNA. 推测这种亚基因组 mRNA 就是 3′-端附近的基因进行表达的形式，这种 mRNA 同样也是单顺反子的. 目前对亚基因组形成和表达的机制尚不清楚. 在植物 RNA 病毒中以亚基因组 mRNA 形式表达可能是一种普遍的方式.

四、Ti 质粒和花椰菜花叶病毒 (CaMV) DNA

有希望成为植物系统基因载体的 Ti 质粒，以及花椰菜花叶病毒 DNA 的结构和基因表

达的研究，近年来发展非常迅速[31-33]。Ti 质粒是存在于格兰氏阴性细菌——根癌农杆菌中的一种分子量为 1×10^8d 的质粒。这种菌感染双子叶植物时，Ti 质粒的一个片段 T-DNA 插入宿主细胞核基因组，引起植物发生肿瘤——冠瘿瘤。Ti 质粒 DNA 除有诱导肿瘤的基因外，还有鳞肉碱（octopine）基因。T-DNA 分子量约 1.4×10^6d。它插入宿主基因组是通过 T-DNA 两端具有重复顺序的区域整合到植物核 DNA 上。T-DNA 在宿主 DNA 上能形成重复顺序。这种整合的机制和细菌 DNA 转导非常相似。

如果使 T-DNA 发生突变，使植物不发生肿瘤，同时把我们希望的基因插入 Ti 质粒，就有可能以 Ti 质粒作为载体把外来基因导入植物细胞，因此 T-DNA 诱发肿瘤的机制是一个极有兴趣的问题。国际上有很多实验室正在作 T-DNA 克隆，研究它们的结构，探索应用的可能性。

花椰菜花叶病毒也是 DNA 型病毒，基因组是环状双链 DNA，其全碱基顺序已经确定，是由 8024 个碱基对组成，它的 α-链上有一"缺刻"（gap）称 gap1。它的 β-链有二个"缺刻"，称 gap 2 和 gap 3。α-链的"缺刻"部位缺少 1—2 个碱基，但是 β-链的 gap 2 和 3 部位并不缺失碱基，只是在"缺刻"区出现碱基的重叠。最近研究指出 CaMV 双链 DNA 中只有 α-链能转录，β-链没有转录能力。α-链转录产物有六种，它们的基因都已在 DNA 上定位。

CaMV DNA 作为植物系统基因载体有二个缺点，首先是 CaMV 病毒繁殖能力差，要得到一定量 CaMV DNA 有困难，其次宿主范围窄。目前正致力于研究解决这两个困难，正在用重组 DNA 技术使病毒 DNA 在大肠菌内繁殖，可望解决第一个问题。宿主范围窄主要取决于侵入宿主细胞和脱外壳蛋白，目前正在用突变的办法使 CaMV 病毒不形成外壳蛋白。此外正在研究把裸露的 DNA 包埋在脂质体内导入植物原生质体。

本文承殷宏章先生审阅，特此致谢。

参 考 文 献

[1] Kaina, B. et al: *FEBS Lett*, **96**, 19, 1978.
[2] Burr, B. et al.: Stdler Genet. Symp., 13, **115**, 1981.
[3] Varsanyi-Breiner, A., et al.: *Gene*, **7**, 317, 1979.
[4] Fischer, R. L., et al.: *Cell*, **29**, 615, 1982.
[5] Slightom, J. L., et al.: *Cell*, **29**, 651, 1983.
[6] Sun, S. M., et al.: *Nature* **289**, 37, 1981.
[7] Frank, G., et al.: Gene. **13**, 227, 1981.
[8] Karn, F., et al.: *Proc. Natl. Acad. Sci. U. S. A.*, **77**, 5172, 1980.
[9] Wiborg, O., *EMBO J.*, **2**, 449, 1983.
[10] Bohsen, K., et al.: *EMBO J.*, **2**, 1165, 1983,
[11] Bedbrook, J., et al.: Nature, 287, 692, 1980.
[12] Smith, S. M., et al.: *J. Mol. Appl. Genet.*, **1**, 127, 1981.
[13] Orozco, E. M. Jr. et al.: *J. Biol. Chem.*, **255**, 109, 1980.
[14] Bogorad, L., et al.: *Photosyn. Prokaryotes Cell Differ. Funct. Proc. Sped. FEBS Meet* 1982 (pub. 1983) p. 119, 1983.
[15] Palmer, J. D.: *Nucleic Acids. Res.*, **10**, 177i,
[16] Miyata, T., et al.: *Nucleic Acids. Res.*, **10**, 1658, 1982.
[17] De Heig, H. H.: *Cur. Genet.*, **7**, 1, 1983.
[18] Karabin, C. D.: *J. Biol. Chem.*, **258**, 5512, 1983.
[19] Bogorad, D., et al.: *J. Biol. Chem.*, **258**, 5503, 1983.
[20] Bogorad, L.: in J. K. Setlow (ed.) Genetic Engineering Principes and Methods Vol. I. p. 205, 1979.
[21] Quetier, J.: in C. J. Leaver (ed.) *Genome organization and expression in plants, NATO Adavanced Study* Institute Series, Vol. **29**, Plenum Press, New York, p. 401, 1980.
[22] Ward, B. L. et al.: *Cell*, **25**, 793, 1981.
[23] Wallace, D. C.: *Microbiol. Rev.*, **46**, 208, 1982.
[24] Bonen, L.: et al.: *Nucleic Acids. Res.*, **8**, 319. 1980.
[25] Forde, B. G. et al.: *Proc. Natl. Acad. Sci. U. S. A.*, **75**, 3841, 1980.
[26] Hall, T. et al.: *Nucleic Acids, in Plants* Vol. II, CRC press, p. 378.
[27] Hirth, L., et al.: *Adv. Virus Res.*, **26**, 145, 1981.
[28] Haseloff, J., et al.: *Nucleic Acids, Res.*, 9, 2741, 1981.
[29] Fukuda, M., et al.: Proc. Natl. Acad. Sci. U. S. A., 78, 4231. 1981.
[30] Rackwitz, H-R. et al.: Nature 291, 297, 1981.
[31] Wilson, T. D. et al.: *Cell* 17, 77, 1979.
[32] Fukunaga, Y., et al.: Virology, 113, 752, 1981.
[33] Nagata, T., et al.: 1982, Mol. Gen. Genet., 183, 182, 1982.

[本文于1984年4月3日收到]