

[19] 中华人民共和国国家知识产权局



[12] 发明专利说明书

专利号 ZL 01808867.8

[51] Int. Cl.

C12N 15/82 (2006.01)

C12N 15/29 (2006.01)

C12N 5/14 (2006.01)

A01H 5/00 (2006.01)

A61P 43/00 (2006.01)

[45] 授权公告日 2008 年 10 月 1 日

[11] 授权公告号 CN 100422335C

[22] 申请日 2001.3.30 [21] 申请号 01808867.8

[30] 优先权

[32] 2000.3.31 [33] ES [31] P200000814

[86] 国际申请 PCT/ES2001/000127 2001.3.30

[87] 国际公布 WO2001/073088 西 2001.10.4

[85] 进入国家阶段日期 2002.12.4

[73] 专利权人 科学研究高级委员会

地址 西班牙马德里

共同专利权人 新生物技术股份有限公司

[72] 发明人 M·D·冈梅兹吉米尼兹

L·A·卡纳斯克莱门特

F·马杜诺阿尔比

J·P·比尔特兰伯特

[56] 参考文献

WO9626283A1 1996.1.1

WO9325695A1 1993.1.1

EPO578611A1 1994.1.1

审查员 吴希哲

[74] 专利代理机构 中国国际贸易促进委员会专利
商标事务所

代理人 唐伟杰

权利要求书 1 页 说明书 31 页 附图 10 页

[54] 发明名称

调控基因的花药特异性表达的序列及其在生产
雄性不育植物和杂种种子中的应用

[57] 摘要

本发明涉及用于在花药中特异性地表达目的核
苷酸序列的调控序列,例如,其表达受所述调控序
列控制的序列引起花药在其发育早期阶段被完全缺
损,并可以获得用于生产杂种种子的雄性不育植
株。

1. 调控基因在花药中特异性表达的核苷酸序列，选自：
 - a) 由 SEQ ID NO 1 的核苷酸 1-2731 组成的核苷酸序列；和
 - b) 由 SEQ ID NO 4 的核苷酸 1-1526 组成的核苷酸序列。
2. DNA 构建物，其含有根据权利要求 1 的核苷酸序列，和编码肽，多肽，蛋白质，酶或 RNA 的核苷酸序列。
3. 重组载体，其含有根据权利要求 1 的核苷酸序列，或根据权利要求 2 的 DNA 构建物。
4. 转化的细胞，其含有根据权利要求 1 的核苷酸序列，或根据权利要求 2 的 DNA 构建物，或根据权利要求 3 的重组载体。
5. 转基因细胞，其含有插入其基因组中的，根据权利要求 1 的核苷酸序列，或根据权利要求 2 的 DNA 构建物，或根据权利要求 3 的重组载体。
6. 转基因植物在生产杂种种子中的用途，所述的转基因植物包含至少根据权利要求 5 的转基因细胞。
7. 生产雄性不育植物的方法，包括用根据权利要求 2 的 DNA 构建物转化植物，所述的 DNA 构建物包含植物不育促进者酶的编码序列，该序列受根据权利要求 1 的核苷酸序列调控，使得所述编码序列的所述表达产生花药的缺损。
8. 恢复雄性不育植物育性的方法，包括用根据权利要求 2 的 DNA 构建物转化雄性不育植物，所述构建物包含编码具有抑制花药缺损活性而回复或恢复可育性的酶的序列，该序列由根据权利要求 1 的核苷酸序列调控。

调控基因的花药特异性表达的序列及其在 生产雄性不育植物和杂种种子中的应用

发明领域

农业。本发明涉及获得调控 DNA 序列（启动子），和应用所述的序列构建新的能够在转基因植株的花药中特异表达的嵌合基因。同时本发明还涉及雄性不育植株和杂种种子的生产。

发明背景

为了激发转基因植株的花药中特定基因的特异表达，已经分离鉴定并应用了不同的启动子。这些启动子的表达对花药的给定组织是特异的。其中一些应用于细胞缺损（cell ablation）技术来生产对生产杂种种子具有很大利用价值的雄性不育植株。当前，植物细胞的遗传缺损用于研究特异的组织或器官中细胞功能与信号传导，并用于产生雄性不育（雄性-不育）。

遗传缺损基于诱导细胞死亡，通过表达任何一种能够摧毁细胞完整性的酶如蛋白酶，脂肪酶和 RNA 酶（例如芽孢杆菌 RNA 酶和 T1 RNA 酶）而实现。采用将有毒物质引入细胞的办法也可达到同样的效果（Day CD and Irish VF. Trends Plant Sci., 2:106-111, 1997）。后一种方法的事例是应用灭活核糖体的多肽，例如白喉棒杆菌（*Corynebacterium diphtheria*）（DTA）中毒素的 A 链和铜绿假单胞菌（*Pseudomonas aeruginosa*）中的外毒素 A。几个小组提出通过生长调节因子的过量表达产生雄性不育。采用非植物基因如 Ti 质粒的基因 1 和 2，或 Ri 质粒的基因 rolA, B 和 C 合成植株生长素和细胞分裂素，代表这些方法，其中由于表达的规模和不恰当的表达时机这些因子变成了毒素。

通过促使细胞对特异的缺损试剂敏感而不是直接摧毁组织的方法也有所开发。这个方法的实例是应用先前已知基因的“反义”RNA。这

提供了对化学试剂的遗传抗性，如对除草剂的耐药性（Fabijanski SF 等，*In vitro Cell. Dev. Biol.* 28: 46-52, 1992）。反义 RNA 的效果在于特异性的去除化学抗性，如在花粉中，如此以来应用这种除草剂导致花粉的破坏。这种方法将除草剂转化成杀配子剂。

为了得到有效的遗传缺损，关键是得到只在它表达的地方起作用的细胞毒素基因以及调控该细胞毒素基因时空表达的合适的启动子。因此，表征不同类型细胞和/或花药的不同分化阶段中的活性基因启动子，允许以细胞缺损的办法来检测配子形成时期花药不同组织的功能（Mariani C 等，*Nature*, 347: 737-741, 1990; Paul W 等，*Plant Mol. Biol.*, 19: 611-622, 1992; Hird DL 等., *Plant J.*, 4: 1023-1033, 1993），重点特别在于细胞间相互作用（Roberts MR 等., *Sex. Plant Reprod.*, 8: 299-307, 1995）。相似的，大型生产种子公司利用这种办法开发雄性不育，在获得杂种种子的过程中雄性不育是富有价值的特点（Williams ME, *Trends Biotechnol.*, 13: 344-349, 1995）。

在几项工作中分析了生殖结构发育过程中的细胞间相互作用。例如，采用几种不同的启动子来调节花药绒毡层细胞中细胞毒素基因的表达，目的是确定缺损对花粉发育的影响（Mariani C 等 . *Nature*, 347: 737-741, 1990; Roberts MR 等., *Sex. Plant Reprod.*, 8: 299-307, 1995）。不同的研究中缺损绒毡层细胞对花粉发育有不同的影响。采用烟草绒毡层细胞中特异性的启动子（TA 29）调节花粉发育的四分体阶段中芽孢杆菌 RNA 酶（核酸酶）基因的表达，来产生雄性不育，这说明该阶段绒毡层对产生可育花粉是必要的（Mariani C 等. *Nature*, 347: 737-741, 1990）。相反的，特异于花粉小孢子阶段的绒毡层用拟南芥的 APG 启动子代替 TA29 启动子，对花粉没有任何效果（Roberts MR 等., *Sex. Plant Reprod.*, 8: 299-307, 1995）。后者的数据显示绒毡层对四分小孢子分裂形成花粉来说不是关键的。采用 TA29-芽孢杆菌 RNA 酶构建物对芸苔属（*Brassica*）转基因植株花药的形成功能进行的组织化学分析说明绒毡层细胞内 RNA 的降解伴随着来自小孢子的 RNA 的消失（De Block M and Debrouwer D, *Planta* 189: 218-225, 1993）。

这个发现提示孢粉素沉积开始之后以及在小孢子发育的后期阶段小孢子能渗漏小分子，因为 TA29 启动子并不直接调节小孢子中基因的表达。

Beals TP 和 Goldberg RB (Plant Cell , 9: 1527-1545, 1997) 采用新的细胞缺损策略来确定花药中哪些种类型的细胞涉及开裂过程。他们用两个构建物来转化烟草植株：构建物由 TA56 启动子与芽孢杆菌 RNA 酶基因构成，该启动子在隔膜 (*spetum*)，裂缝或药隔 (花药) 组织中有活性，同时伴有以下替换形式的构建物之一：a) 在花药的大多数组织中有活性的 TP12 启动子与 *barstar* 基因 (芽孢杆菌 RNA 酶抑制剂)，b) 在花药的大多数组织中均有活性但分布模式与 TP12 不同的 TA20 启动子与 *barstar* 基因，和 c) 在药隔 (花药) 组织，隔膜和裂缝的细胞中有活性的大豆凝集素基因启动子与 *barstar* 基因。分析转基因植物不同的表型说明分裂过程只依赖于功能性裂缝的存在。

Shull (J. Ind. Abst. Vererb. 12:97-149, 1914) 第一个引入单词杂种优势来描述杂合现象带来的关于有机体在细胞分裂，成长或其它的生理活动等方面的优势。上述优势的结果如大小、活力或产量的增长，早熟，抗病等几十年来一直促使大家努力获得杂交品种 (Tsaftaris SA, Physiol. Plantarum 94:362-370, 1995)。大多数自身具有雄性和雌性繁殖器官的植物自花传粉 (玉米，水稻，大豆，番茄等)，而这为生产杂种种子带来问题 (Kriete G et al., Plant J. 9:809-818, 1996)。为了避免这个问题必须采用某种系统来控制自花传粉。这个系统可以是机械的，化学的或遗传的。

机械的系统包括人工摘除花的花药，这是费力和费钱的工作。

化学的方法基于应用可以特异性破坏花粉而导致短时间雄性不育的化学产品 (杀配子剂)。当遇到花期较长或变化的，不可控制的花期时这种近似的方法便没有显著的效果。另外，通过杀配子剂来商业生产杂种细胞的特点是昂贵，和化学产品相对较高的效率。

大多数生产杂种种子的商业系统都是基于遗传学方法来控制开

花，如此以来应用相互不兼容的植株或雄性不育植株，即不产生花粉的植株，该植株不能释放花粉或使花粉发育从而不能产生自体受精繁殖(Horner HT and Palmer RG, Crop Sci. 35: 1527-1535, 1995)。生产雄性不育植株对获得杂种种子有很大的作用。雄性不育去除了植株自身传粉的可能性，有利于杂种种子的生产，后者在遗传改良计划中有重要的应用。

目前所有的获得杂种种子的方法都有局限性，并且不能应用于许多有农业价值的重要作物。现在，正在发展基于基因工程的生产雄性不育植株的新方法(Gates P, Biotechnol. Genet. Engineering Rev. 13:181-195, 1994)。这种方法的发展基于重组DNA和鉴定涉及花粉发育的基因，已经允许激发核雄性不育(NMS)的分子系统的增殖(Scott RJ et al., Plant Sci. 80:167-191, 1991)。如前所述，细胞缺损方法可以阻碍小孢子或导致小孢子发育的组织(绒毡层和花粉囊壁)的发育，采用这个方法能获得分子系统的雄性不育。另外，也可能产生具有完全功能的花粉，但因为花药结构的缺陷花粉不能释放。为了得到用于生产杂种种子的雄性不育植株，重要的量得到对涉及花粉发育组织有特异性的启动子，用于选择雄性不育系的系统，和用于恢复F1杂交系育性的系统。

特异性启动子

大多数植物基因的转录过程受到空间和时间上的调控。基因活性的调节通过基因启动子区域的反式因子和顺式调控元件的相互作用来完成。因而，启动子就是直接调控结构基因转录的DNA序列，位于基因的5'区域。

那些只在花的生殖器官(雄蕊和心皮)中表达的基因值得引起特殊的注意，因为它们的启动子具有潜在的能力来调控其他基因向所述器官表达，并引发花的雄性或雌性不育。这就是花药中某些特异表达的基因启动子，联合只影响花粉发育的缺损试剂，已经应用于生物技术方法来生产不能自花传粉的雄性不育植株。因此，这些类型的方法极大依赖于能提供适当表达的启动子的存在。过去的一些年中，已经

分离并鉴定了大量在涉及花粉发育的组织或细胞中特异表达的基因，因此这些启动子可用于这个目的 (Scott RJ et al., Plant Sci. 80: 167-191, 1991)。

选择用来表达毒性物质或那些使组织退化的物质的启动子通常是发育调控启动子，对靶细胞有充分的活性和特异性。如果靶细胞是绒毡层组织，启动子应该在小孢子发育的早期起作用，这样才可能在小孢子不依赖绒毡层支持之前阻止过程。相似的，如果致死基因是半合子的，启动子也应该在减数分裂之前起作用来防止部分小孢子缺乏退化的活性。

可由诱导型启动子来替代发育调控启动子的作用。然而，具有特异化学诱导性的已述启动子的数量非常小。当需要保持并增加母系时，应用诱导型启动子来阻止花粉发育有很多优点，因为植株可育并能通过自交繁殖。如果雄性不育基于抑制基因在涉及花粉发育的组织中表达，如通过“反义”，遗传活性或诱导型启动子也可以应用。

选择雄性不育系

为了生产工业数量的杂种种子，增加母系是必须的。尽管可以通过体外繁殖的方法为具有农业价值的植物生产许多雄性不育植株，但这样会太昂贵。通常繁殖雄性不育系的策略是给缺损基因连接抗除草剂基因 (Mariani C et al., Nature 357: 384-387, 1992)。用同一基因的雄性不育系杂交可育系，没遗传到不育性的植株通过除草剂去除。经过类推，任何允许区别这两个表型的基因都可以应用，如影响种子色素沉着的基因。

恢复可育

恢复杂种植物的可育性，可以通过将他们与表达特异于用于产生雄性不育的毒性酶的抑制剂的转基因植株杂交，如 barstar 基因，该基因的产物抑制芽孢杆菌 RNA 酶 (Mariani C et al., Nature 357: 384-387, 1992)，或所用致死基因的“反义”RNA (Schmulling T et al., Mol. Gen. Genet. 237: 385-394, 1993)。

获得雄性不育植株的分子系统

用来生产雄性不育的缺损方法由 Mariani et al 第一个提出 (Nature 347: 737-741, 1990)。特异于绒毡层的 TA29 烟草基因的启动子用来调节两种不同的 RNA 酶 (来自米曲霉 (*Aspergillus oryzae*) 的 T1 RNA 酶和来自解淀粉芽孢杆菌 (*Bacillus amyloliquefaciens*) 的芽孢杆菌 RNA 酶) 在烟草和欧洲油菜 (*Brassica napus*) 中的表达。得到的雄性不育转基因植株缺少绒毡层, 并具有无小孢子或无花粉粒的花粉囊。TA29-芽孢杆菌 RNA 酶构建物与 *bar* 基因融合而允许从群体中选择出雄性不育植株, *bar* 基因带来对 ammonium gluphosinate 类除草剂的抗性。应用除草剂除去杂交后代中的雄性可育植株, 而提高分离不育植株的效率。尽管如此, 如果不存在回复雄性不育的可能性, 上述转基因植株不优于自然变异 (Mariani C et al., Nature 357: 384-387, 1992)。采用携带含 TA29 启动子调控的核酸酶芽孢杆菌 RNA 酶抑制基因 (*barstar*) 构建物的转基因植株可以解决问题。TA29-*barstar* 植株作为恢复系, 当它们与 TA29-芽孢杆菌 RNA 酶构建物转化的植株杂交就可以获得雄性可育的植株。

除了这个系统, 文献还包含有其他生产雄性不育系的方法。

在碧冬茄 (*Petunia hybrida*) 中, 通过抑制花药中类黄酮的合成而激发核雄性不育, 类黄酮阻止花粉成熟。上述结果可以通过两条不同的途径完成: 采用查尔酮合成酶基因 (*chalconsynthetase*) RNA 的“反义”效果 (van der Meer IM 等, Plant Cell 4:253-262, 1992), 和采用共抑制查尔酮合成酶基因 (Taylor LP and Jorgensen R, J. Hered. 83:11-17, 1992), 该酶涉及类黄酮合成路线。为了恢复可育性, 可将类黄酮应用到柱头或同花粉混合, 来允许雄性不育系通过自花传粉来繁殖, 这样以来不需要用任何选择标记就可获得雄性不育表型的纯合基因系 (Ylstra B T 等, Plant J. 6:201-212, 1994)。

在烟草中, 表达融合启动子 *CaMV* 35S 的毛根农杆菌 (*Agrobacterium rhizogenes*) 的 *rolC* 基因也可以诱导雄性不育。遗憾的是, 转基因植株中雄性不育表型伴有其他的表型改变 (Schmulling T 等, EMBO J. 7:2621-2629, 1988)。在 F1 杂种中表

达基因 *rol C* 的反义 RNA 可以恢复不育 (Schmulling T et al., Mol. Gen. Genet. 237: 385-394, 1993)。

O'Keefe 等 (Plant Physiol. 105: 473-482, 1994) 描述了以在烟草绒毡层细胞中表达 P450_{SU1} 细胞色素为基础的诱导型雄性不育。该蛋白可以将外源加入的 R7402 杀配子剂的无害衍生物转化成它的活性形式 (毒性加强 500 倍)。然而, 可能因为 R7402 的快速代谢, 在应用该种化合物期间雄性不育局限在处于特定发育阶段的花中。另外, R7402 本身有毒, 当用量是传统生产雄性不育用量的四倍时它就开始抑制生长。

另外一种获得诱导型雄性不育的系统建立在应用绒毡层 TA29 启动子和来自 *E. coli* 的 *argE* 基因的基础上。该基因的产物将 N-乙酰-L-phosphinotrycin 脱乙酰基转化为细胞毒性化合物 gluphosinate。当 N-乙酰-L-phosphinotrycin 应用于烟草植株时, 绒毡层退化而获得雄性不育植株 (Kriete G 等, Plant J. 9: 809-818, 1996)。最后, 我们想指出, 除了是获得杂种种子的工具, 雄性不育也是植株在缺少受精 (单性结实的果实) 条件下果实能够发育的有利特点。在这种类型的果实中缺少种子, 从而消费者对它的消费或接受就会增加 (Rotino G et al., Nat. Biotechnol. 15: 1398-1401, 1997)。一些有农业价值的单性结实的植物有: 梨树, 柠檬, 黄瓜, 葡萄和枣椰子。

发明简述

总之, 本发明涉及的问题关于开发用于在植物中产生雄性不育的系统。

本发明提供的解决方法基于分离并鉴定植物发育早期调节花药特异表达的启动子, 尤其是来源于豌豆 (*Pisum sativum* L.) 的 *END1* 基因启动子。应用上述启动子可以产生表达特异于花药的基因的转基因植物, 如激发花药缺损的基因, 从而产生雄性不育植株, 在为获得杂种种子的遗传改良计划中有极大的用途。在实施例 1 中, 记述了生产含有不同构建物的转基因植株, 这些构建物包含融合报告基因的 *END1*

启动子，观察到它调节花药报告基因表达的特异性。实施例 2 描述了采用芽孢杆菌 RNA 酶基因通过 *END1* 启动子来生产拟南芥雄性不育植株。得到的结果说明构建物激发花药在发育很早阶段的完全缺损，阻碍其中花粉的形成并以 100% 效率得到雄性不育植株。

因此，本发明的目的就是调控基因在花药特异性表达的核苷酸序列，该序列包含 SEQ ID NO 1 中所示的核苷酸序列，或其中一段，或与所述序列基本同源的核苷酸序列。应用所述核苷酸序列生产雄性不育植株和杂种种子成为本发明附加的目的。

本发明另一个附加目的是构建包括整个或部分所述核苷酸序列的 DNA 构建物，以及含所述的序列或 DNA 构建物的载体和上述载体转化的细胞。

本发明另一个附加目的是应用所述的 DNA 序列或所述的 DNA 构建物生产在花药中特异表达目的 DNA 序列的转基因植株，如表达细胞毒性多肽和 RNA 的转基因雄性不育植株，而这种表达会诱导花药的缺损。获得的转基因植株构成本发明另一个附加目的。

本发明另一个附加目的是生产杂种种子的方法，包括向植株引入所述 DNA 构建物。

附图简述

图 1 表示豌豆花不同阶段纵向切片中 *END1* 基因表达的时空分析。紫色的沉积指示表达位置。A. 开花前 12 天的花芽，其中普通的原基尚没有分化出花的花瓣和雄蕊特征，而 *END1* 基因开始表达。B. 开花前 10 天的花。C. 原位杂交的对照：采用“正义”核糖核酸探针（riboprobe）杂交的部分。D. 开花前 8 天的花。E 和 F. 开花前 6 天的花，分别采用较小和较大的倍数摄影。G 和 H. 开花前 5 天，花的连续的切片分别与“正义”核糖核酸探针（对照）和“反义”核糖核酸探针杂交。C, 心皮；E, 雄蕊；P, 花瓣；Co, 药隔组织；M, 小孢子；T, 绒毡层；1, 2, 3 和 4, 花的轮生体 1, 2, 3 和 4。

图 2 表示 *END1* 基因启动子部分的序列。两个可能的 TATA 框（tataaat 和 tatatata），CAAT 框和两个 CCAAT 框以下划线表示。

蓝色，表示三个可能的 CarG 框，绿色表示 GTCAAAA 基元，粉红表示 ACGTCA 元件，红色的表示三个 C(A)_{6/8} 基元。GTCAAAA 框最后五个核苷酸与一个 C(A)_{6/8} 框共有。符号 +1 表示克隆 162 的 cDNA 序列。翻译起始密码子 (ATG) 用黑体字表示。另外，表示出了推导的氨基酸序列。黑体的核苷酸表示终止翻译的密码子 (taa) 和可能的多聚酰基酸化信号 (aataaa)。

图 3 表示与 *END1* 启动子区域推定的对等序列连接 AGAMOUS DNA (CarG 框和邻近的序列) 的共同序列的堆叠。

图 4 表示用于转化拟南芥属 (*Arabidopsis*) 植物，烟草和番茄的构建物 pBI101-F3 和 pBI101-F1.5 的示意图。用来构建构建物的质粒 pBI101 包括：胭脂碱合成酶的遗传启动子 (*nos-pro*) 同带来抗卡那霉素抗性的 *npt II* 基因融合，编码酶 E-葡萄糖苷酸酶 (GUS-内含子) 的 *uidA* 基因，和位于两种基因的 3' 末端的胭脂碱合成酶的多聚酰基酸化信号 (*nos-ter*)。

图 5 表示融合于 *uidA* 基因 (GUS-内含子) 的豌豆 *END1* 启动子在转基因植物拟南芥中的表达结果。A 和 B: 对处于不同发育阶段的花中进行相应的 E-葡萄糖苷酸酶的活性 (GUS) 组织化学检验以后，所述活性只能在花药 (蓝色) 中检测到，而不会在花的其他器官或植株的其余部分中检测到。C: 包被于石蜡中有 GUS 活性的雄蕊纵向切片。在形成上皮，药隔组织，内皮组织的细胞中检测到 GUS 活性 (蓝色)。绒毡层或花粉中没有检测到 GUS 活性。尽管图中一些花粉粒是蓝色的，但那是人为造成的，因为蓝色只显现在花粉粒外壁而非内部。Co: 药隔，En: 内皮，Ep: 上皮。

图六表示融合于 *uidA* 基因 (GUS-内含子) 的豌豆 *END1* 启动子在转基因烟草 (*Nicotiana tabacum*) 植物中的表达结果。A, B 和 C: 组织化学检测在不同发育阶段烟草花中的 GUS 活性显示活性只存在于花药 (蓝色)。D: 展示花粉粒的花药新鲜横切面，并且剩余的绒毡层组织没有蓝色。E: 雄蕊，C: 心皮，P: 花瓣，S: 萼片，Po: 花粉。

图 7 表示融合于 *uidA* 基因 (GUS-内含子) 的豌豆 *END1* 启动子在

转基因番茄 (*Lycopersicon esculentum*) 植株中的表达结果。a: 组织化学检测转基因番茄花的 GUS 活性表明活性特异存在于花药 (St)。b: 上述花的石蜡切片, GUS 活性局限于形成花粉囊的组织 (内皮, 上皮, 药隔等), 但不存在于绒毡层或花粉中。Se: 萼片, Pe: 花瓣, St: 雄蕊, Ca: 心皮, Ep: 上皮, Co: 药隔, En: 内皮, Tp: 绒毡层和 Po: 花粉。

图 8 表示采用 *END1* 启动子调控芽孢杆菌 RNA 酶基因得到转基因拟南芥雄性不育植株的结果。a: 拟南芥属对照植株 (WT) 显示以成熟花粉受精子房的结果形成的长角果 (箭头, 右框)。除此以外, 可以看到 *END1* 启动子调控芽孢杆菌 RNA 酶基因的表达对转基因植株的影响: 衰老的子房 (右框) 因为没有受粉 (箭头) 而留在植株中。b: 以花药分裂进行柱头传粉之后野生型拟南芥植株的正常花 (阶段 14)。花药成熟以后, 花丝伸长, 直到花粉囊位于柱头的顶点从而有助于它们传粉。c: 在同一阶段的转基因植株的花, 显示芽孢杆菌 RNA 酶影响花药发育。花药缺损是完全的, 没有花粉囊形成。花丝没有经历伸长过程, 因为花药的发育不完全。在这种条件下, 传粉是不可能的, 并且植株是 100% 不育的。

图 9 表示细胞类型的扫描电镜 (SEM) 研究结果, 该细胞类型存在于带有 p*END1*-芽孢杆菌 RNA 酶构建物转基因植株的花药位置形成的结构中。a: 对照拟南芥花的雄蕊。黑色的箭头代表存在于花药 (锯齿状边缘) 上皮的细胞类型, 黑的代表花丝 (伸长的)。b: 从相反的位置看同一雄蕊。在花丝到达花药 (箭头) 的地方观察到细胞类型的变化 (更多圆的)。c: 转基因植物的雄蕊, 观察花丝中存在的细胞类型, 圆的位于插入带, 但并非花粉囊上皮的那些。d: 花丝末端带的结构细节, 在那里花丝伸长的细胞类型和插入带的圆形类型共存。

图 10 表示石蜡切片的组织化学研究结果, 关于带有 p*END1*-芽孢杆菌 RNA 酶构建物转基因植物中花药位置形成的结构。a: 花药发育早期阶段拟南芥花对照的纵向切片。红的是经过连续分裂将会形成成熟花粉的小孢子母细胞团。b: 在较高级发育阶段的对照花药 (12-13,

分裂之前), 可以观察到绒毡层几乎全部消失, 且花粉正在成熟(红色)。花丝开始伸长。c 和 d: 转基因花分别在阶段 13 和 12 的纵向切片, 可观察到在雄蕊位置形成的结构。在末端带产生的加宽部分内可观察到小团花粉母细胞, 可能是某些停止发育(没有分裂)的绒毡层(红色)外部细胞。花丝没有开始加长。e: 图 d 中所示结构之一的细节, 其中展示圆的花粉母细胞团和可能的绒毡层细胞细节。

发明详述

本发明提供调控基因在花药中特异表达的核苷酸序列, 下文中称为本发明的核苷酸序列, 其选自:

- a) 包含 SEQ ID NO 1 中所示的核苷酸序列的核苷酸序列;
- b) 所述 SEQ ID NO 1 中保持有在花药中调控特异表达能力的片段;

和

- c) 与 a) 或 b) 中定义的核苷酸序列基本类似的核苷酸序列。

在本说明书所用的意义上, 单词“类似”意在包括任何至少具有在花药中调控特异表达能力的核苷酸序列。典型的类似 DNA 序列:

以 SEQ ID NO1 中所示的核苷酸序列为基础可从任何产该类似序列的生物中分离得到, 或

以 SEQ ID NO1 中所示的核苷酸序列为基础通过替换一或多个碱基, 在序列中插入一或多个碱基, 在序列的任一末端添加一或多碱基或, 在任一末端或序列内部缺失一或多个碱基构建而得。如, 类似的 DNA 序列可能是 SEQ ID NO1 中所示的核苷酸序列的子序列。

一般说来, 类似的 DNA 序列与 SEQ ID NO1 中定义的核苷酸序列基本同源。从本说明书的意义来看, “基本同源”意即在核苷酸水平所讨论的核苷酸序列至少有 60% 的同一性, 优选至少 85% 或更优选至少 95%。

本发明的核苷酸序列可源于包含它的任一生物, 如来自豌豆 (*Pisum sativum* L.) 或来自该 DNA 序列转化的宿主生物。本发明的核苷酸序列可用传统技术分离, 应用本发明核苷酸序列的信息制备寡聚核苷酸探针从任何其他物种的 DNA 中起始。

在具体的实施方案中, 本发明的核苷酸序列是 SEQ ID NO1 中所示

的序列，与豌豆 (*Pisum sativum* L.) *END1* 基因启动子的核苷酸序列相对应。

在我们实验室以前进行的工作中，我们分离并表征仅在花药的特定组织中表达的豌豆 (*Pisum sativum*) 基因。它采用先分离并鉴定它编码的蛋白的方法来进行对它的分离和表征。为了做到这一点，用花的其他器官（萼片，花瓣和心皮）的抽提液来免疫兔子得到多克隆血清，用该血清进行雄蕊抽提液的免疫减法过程。这个过程的结果就是产生了富含特异于雄蕊的蛋白抽提液，去除了那些在花其他器官中共有的蛋白。用所述的富集抽提液免疫小鼠，获得产生特异于特定雄蕊蛋白的单克隆抗体的杂交瘤系。其中之一 (MabA1) 识别在雄蕊中大量存在的一种 26kD 蛋白 (END1)，它的原位免疫说明它仅在花药中组成花粉囊结构的组织中积累。采用亲和层析，我们纯化了一定量的所述蛋白，测定了它 N-末端 20 个氨基酸序列。得到的序列与先前所述的豌豆子叶中作用未知的另外一种蛋白 (PA2) 有高的同源性 (Higgins 等, *Plant Mol Biol.* 8:37-45, 1987)。采用 PA2 的 cDNA 作探针筛选豌豆花的基因文库，我们分离到带有完全编码序列 (910bp, 我们称之为 END1, SEQ ID NO 2) 的克隆，其表现和 PA2 有 72% 的同源性。Northern 印迹分析说明这种新蛋白表达的特异性，由于它局限于花药的抽提液，而在花的其他器官，子叶和植株组织中没有检测到。另一方面，原位杂交检测证实了 *END1* 基因在豌豆花药组织中在发育不同阶段的表达特异性。(图 1) 早期发育阶段 (在花瓣和雄蕊中普通原基细胞分化) 它开始表达，并持续到花药裂开，只在上皮，药隔，中层和内皮表达。该基因的表达在营养组织 (绒毡层) 和生殖组织 (花粉) 中检测不到。同时检测花的其他器官，该植株的其他部分 (茎，叶，根等) 和种子 (子叶)，结果均呈阴性。

END1 基因在花药的形成花粉囊的组织中特异性的表达提示我们分离并分析所述基因的启动子。研究这个启动子具有不同原因的价值。首先，我们能够确定启动子的哪一个基元决定基因在花药中的特异表达。相似的，分析何种转录元件调控 *END1* 基因的表达，及确定它们是

否与调控花器官身份的 MADS-框同源异型基因相关, 尤其是是否与涉及雄蕊发育的基因组 B 和 C 相关是很有趣的。带着这些目标, 我们采用 *END1* cDNA (SEQ ID NO 2) 完整片段作为探针筛选豌豆基因组 DNA 文库。初筛了 500,000 噬菌斑, 其中我们发现了阳性克隆(克隆 162)。再三次筛选后进行分离与纯化。采用培养物和阳性噬菌体裂解我们纯化得到它的 DNA。采用 Hind II 消化 DNA, 因为该酶可在起始第 210 个碱基处裂解 cDNA 片段, 从噬菌体消化物中我们获得了包含 cDNA 前 210 个碱基的 DNA 片段以及其上游更多的启动子序列。采用克隆 162 的 cDNA 片段 5'末端的 210bp 作探针与消化的 DNA 杂交做 southern 杂交分析。该片段用作探针来确定限制酶切得到的哪一个片段是由 210 个碱基和部分或全部的启动子序列组成的。southern 印迹分析表明来自噬菌体 DNA Hind II 消化的大约 3Kb 的片段与探针杂交。采用 QUIARX II 的方法纯化 3Kb 片段, 在 Bluescript-KS(-)载体的 Hind II 克隆位点进行克隆然后测序。图 2 和 SEQ ID NO 1(图 2 中的核苷酸从 -2736 和 1 与 SEQ ID NO 1 中的核苷酸 1 和 2766 分别相关)表示了 *END1* 启动子的全长核苷酸序列。

测定克隆片段的序列之后, 我们发现我们已经分离得到 2946bp 的片段。其中, 3'末端的 210bp 与 cDNA 162 克隆前 210bp 的片段相对应。在 -263 位置 (SEQ ID NO1, 2474) 和 -56 (SEQ ID NO1, 2681), 将分离得到的 cDNA162 的第一个作为核苷酸 +1 (在 SEQ ID NO1 中核苷酸 +1 对应于核苷酸 2377), 发现有两个可能的 TATA 框, 在位置 -347 (SEQ ID NO1, 2390) 和 -66 (SEQ ID NO1, 2671), 两个推定的 CCAAT 框, 和在位置 -401 CAAT 框 (SEQ ID NO1, 2336) (图 2)。除了这些真核生物启动子中普通的框, 典型的植物启动子框也被发现。在表 1 中从 *END1* 基因启动区域的核苷酸 -400 (SEQ ID NO1, 1977) 和 +1 之间发现的植物启动子的所有基元, 采用 National Institute for Agro-biological Resource of Japan 的数据库定义。我们以启动子中的功能框相对转录起点的距离通常不超过 400 个核苷酸为基础选择这个区域, 这是根据参考研究不同框的科学文献得到的。

表 I. 位于 *endl* 启动子区域核苷酸序列 -400 到 -1 之间片段中的顺式元件的描述

基元	位置 (链)	共有序列	参考文献
ELEMENT -300	-157 (+)	TGHAAARK	Thomas 和 Flavell, 1990
AP3	-209 (+)	TGTGGWWW	Mercurio 和 Karin, 1989
ASF1	-122 (-)	TGACG	Terzaghi 和 Cashmore, 1995
G-BOX	-84 (+)	CACGTG	Foster 等., 1994
E-BOX	-333, -295, -30 (+)	CANNTG	Kawagoe 等., 1994
GATA	-265, -205 (+)	GATA	Gilmartin 等., 1990
GT1	-278, -156 (+)	GRWAAW	Lawton 等., 1991
HEXMOTIF1	-127 (+)	ACGTCA	Mikami 等., 1989
I-BOX	-356 (-), -194 (-)	GATAA	Giuliano 等., 1988
MYB-PH3	-234 (+)	CNGTTA	Solano 等., 1995
MYB-ST1	-206 (+)	GGATA	Baranowskij 等., 1994
ROOTMOTIF1	-202(+), -96 (+)	ATATT	Elmayan 和 Tepfer, 1995
RY	-34 (+)	CATGCAY	Fujiwara 和 Beachy, 1994
SIF	-215 (+)	ATGGTA	Zhou 等., 1992
SEF1	-392 (+)	ATATTTAWW	Lessard 等., 1991
SEF4	-276 (+)	RTTTTTR	Lessard 等., 1991
MYB-P	-349 (+)	CCAACC	Grotewold 等., 1994
SBF1	-150 (+)	TTAA	Lawton 等., 1991
CarG	-103 (-)	CC(A/T) ₆ GG	Shiraishi 等., 1993

对启动子区域序列的进一步分析可以检测到其他调控基元的存在。其中两个是 CarG 框 ($-^{329}\text{TGAAAATACC}^{-320}$ (SEQ ID NO1²⁴⁰⁸TGAAAATACC²⁴¹⁷) 和 $-^{300}\text{GGTTTCAACT}^{-291}$ (SEQ ID NO1²⁴³⁷GGTTTCAACT²⁴⁴⁶))，尽管不如第一个与共同序列相似，但也起相同作用。起作用的 CarG 框的例子，尽管与共同序列不同，由拟南芥属中 AP3 基因启动子的三个 CarG 框中的两个 (CCTTTCATGG 和 CCATTTT TAG) 代表 (Tilly JJ 等, Development 125:1647-1657, 1998)。我们鉴定的其他的基元是 $-^{290}\text{GTCAAAA}^{-284}$ (SEQ ID NO1²⁴⁴⁷GTCAAAA²⁴⁵³)，在基因 *Zm13* 和 *LAT52* 中存在 (Zou JT 等, Am. J. Bot. 81: 552-561, 1994)，基元 $-^{127}\text{ACGTCA}^{-122}$ (SEQ ID NO1²⁶¹⁰ACGTCA²⁶¹⁵) 位于基因 *Bp19* (Zou JT 等, Am. J. Bot. 81: 552-561, 1994)，和重复三次的元件 C(A)_{6/8} (位于 -507 (SEQ ID NO 2230 - 2236)，-288 (SEQ ID NO 2449-2457) 和 -247 (SEQ ID NO 2490-2496)) 在基因 *OlnB4* 和 *OlnB19* 的启动子上 (Hong HP 等, Plant Mol. Biol. 34:549-555, 1997)。

END1 启动子区域发现的不同的基元可以按照它们所在基因类型的功能分组。基元 SEF1, SEF4, E-框, RY, 和元件 -300 以它们存在于编码种子储备蛋白的基因启动子中而表征。

基元 GATA, GT1, I-框和 ASF-1 存在于受光调控其表达的基因上。有些框识别与 Myb 动物蛋白同源的因素: MYB-P, MYB-PH3 和 MYB-ST1, 存在于涉及 *phenylpropanoids* 生物合成路线的基因中。G-框基元存在于多数由不同环境和生理因子调控的植物基因中。菜豆 (*phaseolus vulgaris*) 的防御基因 *chs15* 中已经描述过 SBF1. Rootmotif1 已被鉴定存在于玉米根的两个基因 *rold* 和 *poxI* 的启动子序列中。元件 GTCAAAA, ACGTCA 和 C(A)_{6/8} 存在于欧洲油菜的花药特异基因如 *Bp19* 中 (Zou JT 等, Am. J. Bot. 81: 552-561, 1994; Hong HP 等, Plant Mol. Biol. 34:549-555, 1997)。上述所有基元均列于 SEQ ID NO1。

最后，基元 CARG 是一个由 MADS 结构域识别的 DNA 共有序列，该结构域表征花器官发育涉及的植物同源异型基因 (Huang H 等, Nuc.

Acids Res. 21:4769-4776, 1993; Shiraisi H 等, Plant J. 4:385-398, 1993), 基因 *APETALA-1*(*AP1*; Mandel MA 等, Nature 360:273-277, 1992), *APETALA-3*(*AP3*, Jack T 等, Cell 68:683-697, 1992), *PISTILLATA*(*PI*, Goto K 等, Genes Dev. 8:1548-1560, 1994) 和 *AGAMOUS*(*AG*, Ranofsky MF 等, Nature 346:35-39, 1990) 是拟南芥属中调控花器官身份的同源异型基因。这些基因属于其序列中包含 MADS 结构域的蛋白家族。MAD 结构域由 56 个氨基酸组成, 涉及到 MADS 基因自身或与其他的 MADS 基因的二聚体化, 以及这些二聚体与 DNA 的结合 (Shore P and Sharrocks AD, E. J. Biochem. 229:1-13, 1995)。

B 型同源异型基因 (拟南芥中的 *AP3* 和 *PI*) 控制花瓣和雄蕊的发育, 而 C 型 (拟南芥中的 *AG*) 控制雄蕊和心皮的生成。*END1* 基因在雄蕊中表达, 因而可以成为 B 型与 C 型 MADS 基因的靶基因。*END1* 启动子区域存在上述推定的 CarG 框支持这个假说。Huang H 等 (Nuc. Acids Res. 21:4769-4776, 1993) 与 Shiraisi H 等 (Plant J. 4:385-398, 1993) 确定 *AG* 基因应该需要邻接 CarG 框的共有序列来结合 DNA。*AG* 中 MADS 结构域结合 DNA 的共有序列鉴定为: 根据 Shiraisi H 等 (Plant J. 4:385-398, 1993) 的 5'-TT(A/T/G)CC(A/T)GG(A/T/C)AA-3' 和根据 Huang H 等 (Nuc. Acids Res. 21:4769-4776, 1993) 的 TT(A/T)CC(A/T)(A/t)₂(T/A)NNGG(-G)(A/t)₂。在 *END1* 启动子位置 -103 处邻接 CarG 框的序列与由先前作者阐述的 *AGAMOUS* 基因 MADS 结构域识别的序列相吻合, 除两个核苷酸例外。这个事实支持 *END1* 基因能够成为花同源异型基因的直接作用靶点的假设, 尤其是它会是 *AGAMOUS* 基因的作用靶点。

为了确定从 *END1* 克隆的启动子序列是否调控基因在花药中特异性的表达, 以及它能否在豌豆 (*P. sativum*) 以外的植物中起作用, 我们把启动子融合到基因 E-葡萄糖醛酸糖苷酶 (*uidA*, GUS-内含子) 的编码序列来转化烟草植株 (*Nicotiana tabacum*), 拟南芥 和 番茄 (*Lycopersicon esculentum*) (实施例 1)。采用质粒 pBI101:pBI101-F3

和 pBI101-F1.5 完成了两种不同的构建物（图 4）。第一个构建物包含从核苷酸 - 2736 至 - 6 (SEQ ID NO1, 1 到 2731) 的核苷酸序列，第二个包含从 - 1531 至 - 6 (SEQ ID NO1, 1206 到 2731) 的残基的序列。在最初的尝试中检测后一个构建物标记克隆片段中组成启动子区域的序列。采用根癌土壤杆菌 (*Agrobacterium tumefaciens*) C58 菌株将两种构建物引入到拟南芥，采用根癌土壤杆菌 LBA4404 菌株将构建物 pBI101-F3 引入到烟草和番茄。上述所有的植物又用对照的 pBI101 来转化，然后在不同的组织中分析报告基因的表达。

上述实施例的转化结果说明从 *END1* 基因中分离得到的启动子序列起到充分的作用，因为它在豌豆以外植物的花药中调控基因以特异性的方式异源表达。而且，该启动子位于 - 1531 和 - 6 之间核苷酸序列的这种能力已被描述过。另一方面，基本保持有在花药中调控特异表达的能力的最小核苷酸序列可用相似的实验定义。如此以来，它们组成了本发明所有核苷酸序列的一部分，无论是具有 *END1* 启动子区域的全部序列或是基本保持有在花药中调控特异表达的能力的核苷酸序列的任何片段。另一方面，从豌豆 *END1* 基因启动子序列开始，基于分子生物学技术普通知识，可以从其他植物种类中获得与 *END1* 同源的两个基因的启动子区域。所有与本发明中所述的一个序列同源，并基本保持有在花药中调控特异表达的能力的核苷酸序列构成本发明的一部分。

这些调控序列可用于开发还含有目的基因编码序列 DNA 构建物，该 DNA 构建物又可整合入重组表达载体。为了做到这一点，本领域已知的不同技术都可以应用 (Sambrook et al, "Molecular cloning, a Laboratory Manual", 2nd ed., Cold Spring Harbor Laboratory Press, N. AND., 1989 Vol 1-3), 其中一些在本发明中展示。上述重组载体允许表达不同的肽，多肽，蛋白质或 RNA 编码序列，如本发明发明背景所述，它们表达后会对所在的组织有毒从而诱导该组织缺损并由此导致植株雄性不育。这些基因的实例公布于欧洲专利申请 EP412006, 其可以和本发明发明背景所述的那些一起引作参考。另一方面，这些

构建物可以包含其他在花药中调控特异表达的元件，这取决于所用的重组载体，要转化的植物等。所有可能的含作为共同支配者 (common dominator) 的本发明核苷酸序列的上述 DNA 构建物，和它们在生产雄性不育植物中的应用组成本发明的一部分。

所以，本发明提供 DNA 构建物，它包括本发明的核苷酸序列和目的肽，多肽，蛋白质，活性或 RNA 的编码序列。在具体的实施方案中，所述 DNA 构建物包括在本发明核苷酸序列调控下促进植株不育的活性编码序列。在另一个具体的实施方案中，所述 DNA 构建物包括具有回复或恢复植物育性的活性的编码序列，如在本发明核苷酸序列的调控下具有抑制促植株不育活性的活性编码序列。所述 DNA 构建物也可以包含操作性连接于例如转录终止序列的某些调控元件而实现花药特异表达，这取决于所用的重组载体，转化的植株等。

本发明的核苷酸序列，或本发明提供的 DNA 构建物，可以插入合适的载体。因此，本发明也涉及载体如表达载体，含本发明所述的核苷酸序列和所述 DNA 构建物。选择载体取决于后来所要引入的宿主细胞。以实施例中的方法，所述 DNA 序列所引入的载体可以是质粒或当引入宿主细胞后会整合到所述细胞的基因组中并随同所整合的染色体一起复制的载体。为了获得所述载体，可以应用本领域周知的常规方法 [Sambrook 等，1989，前面已引用]。

本发明也提供细胞，其含有本发明的核苷酸序列，或含所述核苷酸序列的 DNA 构建物或上文提及的载体。可以转化本发明核苷酸序列的宿主细胞可以是原核细胞，优选真核细胞如植物组织细胞。转化植物组织细胞同样可用常规方法实现。关于基因转移到植物的综述包括载体，DNA 转移的方法等，可见 Marta Izquierdo, Ed. Piramide (1999) 的书 "Ingenieria genetica and transferencia genica"，具体到第九章，第 283-316 页，标题为 "Transferencia genica a plantas"。

本发明的核苷酸序列可用于转化植物并获得转化的植物。在现有技术中广泛描述植物转化。众所周知，可采用多种系统例如质粒载体，脂质体，电穿孔，微量注射，扩散，基因枪，磷酸钙共沉淀，使用病

毒载体等。在本发明中，采用质粒载体转化植物作为许多技术可能性的实施例被描述，其所有都组成本发明的部分。从这个意义上说，在实施例 1 中转化烟草植株 (*N. tabacum*)，番茄 (*L. sculentum*) 和拟南芥作为植物转化的例子被描述。

本发明的核苷酸序列可用来在植物中调控目的肽，多肽，蛋白质，活性或 RNA 序列在花药中特异表达。在本发明具体的实施方案中，本发明的核苷酸序列可用来生产雄性不育的植物（雄性不育），也就是说雄性不育植物，这包括采用本发明提供的 DNA 构建物转化植物，所述的 DNA 构建物包含促进植株不育活性的编码序列例如在本发明核苷酸序列控制下的细胞毒性基因序列，如编码核糖核酸酶活性的基因，从而使得包含于所述构建物中的编码序列的表达从花药发育的最初阶段引发其完全缺损，阻止其中花粉的生成并以 100%效率生产雄性不育植株。所得雄性不育植物可通过包括在允许其发育和成熟的条件下培育的方法用于产生杂种种子。

本发明也提供了一种恢复雄性不育植物育性的方法，包括用所述 DNA 构建物转化所述雄性不育植物从而获得可育植物，所述的 DNA 构建物包括在本发明核苷酸序列的调控下具有回复或恢复植株育性活性的编码序列。在本发明具体的实施方案中，雄性不育是因为表达芽孢杆菌 RNA 酶而恢复育性是因为表达具有回复或恢复植株育性活性的编码序列如表达 barstar，这种活性可以抑制由芽孢杆菌 RNA 酶引起的花药缺损活性。

对本发明核苷酸序列的转化敏感的植株可以是单子叶植物或双子叶植物，如具有农业价值的单子叶植物或双子叶植物如谷类，园艺植物如烟草，番茄，瓜，西瓜，黄瓜等。

在另外一个具体的实施方案中，本发明提供含有整合到其基因组中的本发明核苷酸序列的转基因细胞，也包括含至少一种所述的转基因细胞的转基因植物。所述转基因植物组成本发明的附加目的，可以通过常规技术获得，如应用常规反义 mRNA 技术和/或过量表达（正义沉默）或其他，如采用二元载体或其他当前不同的转基因技术可用的

载体。构成本发明部分的转基因植株实例包括单子叶植物和双子叶植物。

本发明的核苷酸序列用于获得雄性不育植物和生产杂种种子方面，因为如前所述，它调控形成花粉囊的组织中编码产生细胞缺损的特定的酶基因的表达。这个系统，联合其他相似系统，产生那些激发细胞缺损的酶的抑制物而恢复育性，将对基于生产杂种种子（杂种优势）的基因改良计划有很大的用途。带着同样的目的，以前试过不同的启动子，但是通过它们没有获得具有100%的不育植物群体（Zhan X et al., Sex. Plant Reprod. 9: 35-43, 1996; Roberts MR 等, Sex. Plant Reprod. 8: 299-307, 1995）。这种现象是因为在花粉实际上已经成熟的发育后期阶段激活上述启动子，所以进行细胞缺损的组织（绒毡层）不再是花粉达到完全成熟所必需的。这种方法学的严重问题是最后作为结果得到杂交与非杂交种子混和的群体。本发明的核苷酸序列替代所述的启动子，因为当雄蕊原基细胞开始分化的时候它被激活。在这个早期阶段，蛋白酶或核酸酶将通过破坏早期阶段生发形成花粉囊组织的细胞系来阻碍花药正常的发育。*END1* 表达的组织是对花药提供支持和结构（结构）的组织，从而很难想像没有上述组织花药可以产生花粉。从这种意义上说，实施例2描述以100%效率生产拟南芥雄性不育植株。可以合理的期望本发明的核苷酸序列（具体为 *END1* 启动子）能够在不同的双子叶和单子叶植物中保持活性，恰如在上述的几种植物中已鉴定C类型的MADS基因（与 *AGAMOUS* 同源）可激活如双子叶植物中的启动子，所以可以期待本发明核苷酸序列将对生产具有农业价值的双子叶和单子叶植物雄性不育有用。

实施例1

转基因植物拟南芥，烟草（*Nicotiana tabacum*）和番茄（*Lycopersicon esculentum*）中 *END1* 启动子的功能研究

1.1- 设计 pBI101-F3 和 pBI101-F1.5 构建物

从 pBluescript-KS(+) 质粒克隆的 DNA 基因组片段中以 TB1, TB2 和 TB3 寡聚物采用 PCR 技术扩增得到 *END1* 基因启动子区域（表2）。

TB1 引入限制性位点 BamHI, 而 TB2 和 TB3 引入 SalI 限制性位点。扩增两个片段: 其中 2731bp 的一个包含几乎全部分离得到的启动子区域 (-2736/-6), 1526bp 的另外一个邻近基因的编码区域(图 1 中 -1531/-6) (SEQ ID NO 4)。上述两个片段均不含有编码序列。两个片段均在质粒载体 PCR_{2.1} 中克隆 (Clar, J. M. (1998) Nuc. Acids. Res. 16:9677-9686; Mead, D., et al (1991) Bio Technology 9:657-663)。然后用限制性内切酶 BamHI 和 SalI 将克隆的插入片段释放出来, 再克隆到质粒 pBI101 中调控 E-葡萄糖苷酶基因 uidA (GUS-内含子) 的表达 (Vancanneyt G 等, Mol. Gen. Genet. 220:245-250, 1990)。在步骤的最后, 得到两个不同的构建物 pBI101-F3 和 pBI101-F1.5。第一个包含 2731bp 的片段, 第二个包含 1526bp 的片段(图 4)。

引物	序列 (5' 3')	DNA 模板
TB1	GAGAGCCTAGGAAGGTTATGTTGTGAGC	F3 克隆
TB2	GACTCGAGGTCGACTTCAACCTTATTAGT G	F3 克隆
TB3	GACTCGAGGTCGACAACCAGTGTGCATAT ATC	F3 克隆

表 2. PCR 扩增采用的引物

1.2. - 拟南芥, 烟草, 番茄植物的转化

为了获得烟草的转基因植株, 应用了两个烟草栽培种: Samsun 和 Pettit Habane SRI。采用 Horsch et al 所述 (Science, 223:496-498, 1984) 经 Fisher 和 Guiltinan 修改 (Plant Mol. Biol. Reporter 13:2780289, 1995) 的方法, 用叶盘与根癌土壤杆菌 (LBA4404 菌株) 在 MSS 培养基 (Marashige and Skoog 培养基 4.4g/L, 蔗糖 2%, Mes 100mg/L, phytigel 3.5g/L, pH 5.9) 中 24℃ 黑暗共培养三天来实施转化。培养以后, 叶盘转移到有再生和筛选培养基 (MSS 培养基加入 IAA 0.2mg/L, 6-BAP 2.2 mg/L, 羧苄青霉素 400mg/L 和卡那霉素 100mg/L) 的盘中, 出现的叶芽转移到生根培养基 (MSS 培养

基加入 IAA 0.2mg/L, 羧苄青霉素 200mg/L 和卡那霉素 100mg/L)。再生植株转移到盛有(1:1)泥炭:蛭石的罐子中,保存直到它们产出种子。在整个过程中培养条件是光照 12 小时,温度 24℃。

用 Columbia 栽培种生产拟南芥转基因植株。转化方法按照 Bechtold et al 所述(C R Acad. Sci. Paris, Life Sci., 316: 1194-1199, 1993)通过真空渗透和筛选抗卡那霉素的植株实现。采用根癌土壤杆菌 C58 菌株。抗卡那霉素的植株转移到泥炭:珍珠岩:蛭石(1: 1: 1)的罐子于培养箱在 22℃长日照条件下生长直到最后收集到种子。

为了获得转基因番茄植株,采用已知为 VC82b 的生长变种按照 Ellul 等,的方法(Teor. Appl. Genet. In press, 2001)用从萌发种子(12 天)起的双子叶植物作为起始材料,在卡那霉素培养基中以 npt II 基因作标记筛选转化体。

1. 3. - 组织化学分析(GUS)转基因植株

转基因植株的第一代采用组织化学分析 E-葡萄糖苷酶的活性。采用两次 5 分钟真空脉冲在 0.5nM 铁氰化物, 0.5nM 亚铁氰化物, 0.1% 的 Triton X-100 和 2mM X-葡萄糖酸的 0.1M pH7.0 的磷酸缓冲液中渗透被研究的组织,在该溶液中 37℃培养 16 小时。然后,采用 50°, 70°, 90°乙醇连续冲洗脱色。蓝色的鉴定为 GUS 阳性带。固定显现蓝色的组织,在脱水之前包埋于石蜡中(Canas LA et al., Plant J. 6:597-604, 1994),以此通过切片鉴定具体何种类型的细胞表现 E-葡萄糖苷酶活性。采用 MZ8 镜头(Leica)拍摄未经处理的组织,石蜡中的组织切片采用光学显微镜 Eclipse600(Nikon)拍摄。

1. 4. 分析拟南芥转基因植株

采用组织化学检测抗卡那霉素第一代转化植株组织的 E-葡萄糖苷酶活性来研究基因 uidA(GUS-内含子)的表达。检测的器官有:花,叶,茎,根和发芽的苗。

我们分析了 26 株转化了 pBI101-F3 构建物的植株,其中 24 株表现出特异的 E-葡萄糖苷酶活性,剩下的两株在任何被检测组织中都没

有表现活性(图5)。对于 pBI101-F1.5 构建物转化的 19 株抗卡那霉素植株,只有两株缺少 GUS 活性。剩下的 17 株特异地在花药中观察到蓝色。表现 GUS 活性的花药组织同表达豌豆 END1 基因的一样:上皮,药隔,中层和内皮。雄蕊花丝,中央维管柱,营养组织(绒毡层),花粉母细胞和成熟的花粉在任何情况下均无 GUS 活性。pBI101 质粒转化的植株(阴性对照)在任何分析的组织中没有表现 GUS 活性。

用一种或另一种构建物转化的植株间,GUS 活性没有观察到不同。然而,我们观察到无论采用何种构建物,同一植株的花的着色浓度有变化,因为我们已经区别了在特定阶段不表现 GUS 活性的花与在同一阶段确实表现 GUS 活性的花。可能,这个现象是因为组织间渗透的检测剂量的不同。两种构建物取得相似的结果显示所有活性必须的调控片段包含在残基 -1531 和 -6 之间,尽管需要更多的研究来确定保持有启动子活性的最小片段。

1.5. - 分析转基因烟草植物 (*Nicotana tabacum*)

进行关于 pBI101-F3 构建物转化的第一代植物中 *uidA* 基因表达的研究。检测的器官有:花,叶,茎和根。

我们分析了 12 株抗卡那霉素的烟草,其中 10 株在它们的花药中表现出 GUS 活性,而两株在任何被检测组织中都没有表现出活性(图6)。从相同植株中我们发现和在拟南芥植株中观察到的同样的植株间蓝色色带的变化。如同预期,pBI101 质粒转化的植株(阴性对照)在任何组织中没有表现 E-葡萄糖苷酶活性。

1.6. - 分析转基因番茄 (*Lycopersicon esculentum*)

分析了 10 株不同的抗卡那霉素植株,其中一半在分析花药中 *uidA* 基因的表达时显示阳性。研究 pBI101-F3 构建物转化的第一代植株中所述基因的表达。检测的组织有:花,叶,茎和根。恰如前面的情况,结果实际上是一致的,观察到 GUS 基因只在花药形成花粉囊的组织中表达(图7)。

实施例 2

采用 END1 启动子调控的芽孢杆菌 RNA 酶基因来生产拟南芥的雄性不育

植株

2.1. - 设计 pBI101 - pEND1 - 芽孢杆菌 RNA 酶/*barstar* 构建物 pBI101-F3 构建物包括 END1 启动子的 2731bp 片段和 GUS 基因, 被 BamHI 和 SacI 限制性内切酶消化得到 GUS 片段。采用 Quiaex II 系统 (Quiagen) 从琼脂糖凝胶上分离对应于带 BamHI 和 SacI 末端 (termini) 的 pBI101 - pEND1 质粒的片段。芽孢杆菌 RNA 酶/*barstar* 片段 (Mariani et al. Nature, 347: 737-741, 1990; Mariani et al., Nature 357: 384-387, 1992), 克隆在质粒 pBluescriptKS(+) 的 BamHI 位点, 采用以下寡聚物扩增:

T7: 5'TAATACGACTCACTATAGGG3', e

Inhi II: 5'GCGAGCTCTTAAGAAAGTATGATGGTGATG3'

用第一个在芽孢杆菌 RNA 酶基因的 ATG 水平保持住 BamHI 的拼接位点, 而用第二个在 *barstar* 基因的终止密码子水平产生 SacI 拼接位点。PCR 反应的片段产物连接到 pGEM-T 载体 (Promega) 上。最后, 用 BamHI 和 SacI 释放出克隆的芽孢杆菌 RNA 酶/*barstar* 插入片段 (918bp), 并连接到带有 BamHI 和 SacI 的末端 pBI101 - pEND1 质粒上。

2.2. - 结果和评价

得到的结果表示在图 8 中。以 pEND1 - 芽孢杆菌 RNA 构建物转化的植株显示它们花的心皮是如何不育的, 如何不能如同对照植株发生的那样形成相应的长角果。另一方面, 转基因植株显示花丝没有伸长, 其末端出现一些非常初级的结构而不是花药。得到的所有抗卡那霉素的 17 株转基因植株显示其花药完全被缺损从而具有 100% 的雄性不育。

采用扫描电子镜 (SEM) 进行的研究表明 (图 9), 所述的结构中存在的细胞类型与花药上皮中存在的并不对应, 只检测形成花丝的类型或存在于花丝连接花药的地带的类型。以石蜡结构切片并以 Alcian blue 和 Safranin 染色来检测涉及花粉发育的细胞系来完成这些研究。在图 10 中, 我们发现花药母细胞是多么小的一组, 在花丝末端的结构

内部中观察到的可能是外部绒毡层（红色）。与正常产生未转化花药的过程相比，这些细胞看起来已经停止了它们的发育。

如前面所解释的，应用芽孢杆菌 RNA 酶/*barstar* 系统成功的生产了雄性不育植株，因为通过采用特异性的启动子（一般是绒毡层的）它被导向植株的花药。这个系统主要的缺点在于使用所述的启动子，它在花药发育和涉及为形成花粉的细胞进行营养的组织（绒毡层）中起作用太晚。这种作用导致植株的一些花中生产一定量的花粉，引起了该系统的逃逸，相应于效率损失 (<59%)。

采用 *END1* 基因的启动子，这些问题就不存在了，因为三个原因：

- 在非常早期的发育阶段进行芽孢杆菌 RNA 酶 (RNase) 的表达，导致将要形成花粉囊的组织：上皮，内皮，药隔，中层等的细胞系完全的缺损。

- 只在花丝的底部形成花丝，但从不会伸长到达心皮的柱头。

- 尽管将要形成外部绒毡层和花粉母细胞的细胞系没有经历芽孢杆菌 RNA 酶的影响，因为它们不是 *END1* 启动子的目标，然而它们在早期发育阶段停止因为它们没有花药剩余组织的支持，而这也是成熟花粉释放（裂开）过程所需的基本支持。

序列表

<110> CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS

<120> 调控基因花药特异性的表达的序列及其在生产雄性不育植物和杂种种子中的应用

<130> 花药中表达的启动子区域

<140>

<141>

<160> 4

<170> PatentIn Ver. 2.1

<210> 1

<211> 2766

<212> ADN

<213> 豌豆

<220>

<221> 启动子

<222> (1)..(2736)

<220>

<221> TATA_信号

<222> (2474)..(2479)

<220>

<221> TATA_信号

<222> (2681)..(2688)

<220>

<221> CAAT_信号

<222> (2336)..(2339)

<220>

<221> CAAT_信号

<222> ()..(2394)

<220>

<221> CAAT_信号

<222> ()..(2675)

<220>

<221> C_区域

<222> (2408)..(2417)

<220>

<221> C_区域

<222> (2437)..(2446)

<220>

<221> C_区域
<222> 补足 ((2625)..(2634))

<220>
<221> GC_信号
<222> (2447)..(2453)

<220>
<221> 等位基因
<222> (2610)..(2615)

<220>
<221> STS
<222> (2230)..(2236)

<220>
<221> STS
<222> (2449)..(2457)

<220>
<221> STS
<222> (2490)..(2496)

<220>
<221> 基因
<222> (2737)..(2766)

<400> SEQ ID NO 1
gacttcaacc ttattagtga atggacaata aaggttataa gctcctttac tgtgaaagcc 60
caccagtaac atcaccttgc ttatatcatt cagcttcttt ctagtaacat ttggaacgtg 120
tttataacag aaaaaaaccc aaaaactctg aaaagactca cacttttctt atctccagtc 180
cacctctcaa aaggaacaat ttccttcagc ttcttggttg gacacctgtt gagcacatat 240
gctgcagtg caacagtttc tccccacaaa gtgtaggaa gcttcttctc cttcagcatg 300
ttccttgtca tatcaagcaa agttcggttt tcaacaagac cattatggtg aggagtatat 360
ggatcagtca cttcatgctc aattccattc tctttacaga acttcttgaa ctctgtagag 420
ttatactcac ctccaccac agttctgaga atcttcagaa gtctgaccac tttatttctc 480
agccttgatt atgaatttct taaattcagc aaacacctcg tgtttgaatt ttataagga 540
tacctatgct atccttgtga actcatccat aaatgacata aagtattatt ccctcctagt 600
gaaaggtttg taatgggcca cacacataag aatgcactac tcctaaagca tgttttgctc 660
tttgagctac ttttgatgaa aatggcagtc ttggttgctt ccctttcatg cacacattac 720
atgacttttt tggtttctta attgtaggaa ttccacgtac cagtttcttt gaattcaaat 780
tccctaagct cctaaagttc aatgaccaa atcttttggtt ccacaactca ctttccttca 840
caacacttgt tgcgctaagg cattcagagt ctgcagtttt aacattcgcc ttgaatggtt 900
tactccttcc atgttctgac tccataatca acttctgata acagtcatac agcttcaaaa 960
gaatgtcatt catggtaact ggaaatccct tttcaattaa ttgacctaca ctcatcagat 1020
tgctcttcat gccagaagcg taccaagacg ttctgaatta atgcagattt tctattattc 1080
ataatcactc taacattccc cattccttta gcatttagtt acttatcatc agcacatcta 1140
atcttggttt tcttctaga gtcaaaatca accagccatt tcttatttcc agtatgatgg 1200
tttgaacaac cagtgtccat atatcaccag tcttctatag acgcaactac ataactagaa 1260
gccattaata gcacatggtc atcatggtgc tcagatcctt agaatgttca attgctacaa 1320
cgatgtaate aaactgatga gtaagagatc taagtacctt ctcaatgata ctttcctcat 1380
aaagagtttc tccatgagac ttcatctcat ttgtgatcag aatcactcta gagatgtagt 1440
cagataactt ctcatgttcc ttcatgctta gatttctata ctgctcacgt agagactgaa 1500
gtttcacctt ctacactgat gcactactat cgtagcacca caccagtctg tctcacacaa 1560
ccttttccgt cattgaatca acgatttctt taaacacggt cacatccaca cactgatgga 1620

```

tgtagaacaa cgcattctga tccttcttcc tcatatcaca ctgagcattt ctttgcgcat 1680
ccgttgcatt ttctagaagt gaagcataaa cttcgttgat gagatcaaga acatcttgag 1740
caccaaataa cacacacatc tgaatcatcc aacgattcca gttggtgtcg tcgaacaatg 1800
gnagcntggg gcacagattc acaacgatat attataantt ttgttttatg aaatttaaga 1860
acaaatttcc attattctta aaatgtttac aactgatgt agactgcaa aggaataaag 1920
atacaatttg ttcacaccac tcaactgctg aaatataagt gagagttaat gagaaatact 1980
aaaataccct ctaaaattat gaattaattc taacaatctc taatgttagt ataatccatt 2040
aaacactttg atggcaggta taacaagggt gtaagttagt gtatacatat taggctctta 2100
ttatttttat attatctctg cttttcttct tcatgttctc actaatatga tattatctcc 2160
cttccctaaa ttatttatat ttattagaaa aagagtttca ttttttaaaa atatattacc 2220
gtaatttttc aaaaaataaa atttaaatat attttataaa aaaattattt aataatttat 2280
ttacattaat gcataaatat aaataaatac tgtcatttaa tatttaacct tttaacaata 2340
aattatattt atttaattca actaatataa gctaagttat ctcatccaac caattaaaaa 2400
gatcatttga aaataccttt ttatttagtt tgtggcgggt tcaactgtca aaaaaaagga 2460
atttttacga cgatataaat ttaaaccagc aaaaaattga agcagttaag cgaaccaact 2520
catggtagtg ggatataatt atctttgtcg tttatatcgg attcgaatct ctataatgat 2580
gaaaaattaa tatcaaaactt taataaagaa cgtcatttat agagccattt tgggaaacac 2640
atatttcatg tacacgtgat tcgcaaattt ccaataactc tatatatagc cctcctcagt 2700
ttcatgcatt tgctcacaac ataaccttcc ttgaattcga tatctaccta agatgacaaa 2760
accagg
2766

```

<210> 2
<211> 910
<212> ADN
<213> 豌豆

<220>
<221> CDS
<222> (17)..(706)

<400> SEQ ID NO 2

```

tcgatatcta cctaag atg aca aaa cca ggt tac att aat gct gct ttt cgt 52
Met Thr Lys Pro Gly Tyr Ile Asn Ala Ala Phe Arg
      1             5             10

tca tct ttc aac ggc gaa cgt tac tta ttc atc gat gat aag tat gtg 100
Ser Ser Phe Asn Gly Glu Arg Tyr Leu Phe Ile Asp Asp Lys Tyr Val
      15             20             25

ttg gta gat tat gca ccg gga acc cgc gac gat aag ctc tta aac ggg 148
Leu Val Asp Tyr Ala Pro Gly Thr Arg Asp Asp Lys Leu Leu Asn Gly
      30             35             40

cct ctt cct ctt cct gct ggg ttt aaa tca ctt gat ggt aca gta ttt 196
Pro Leu Pro Leu Pro Ala Gly Phe Lys Ser Leu Asp Gly Thr Val Phe
      45             50             55             60

gga acc tac gga gtt gac tgt gcc ttt gac acc gat aac gac gaa gca 244
Gly Thr Tyr Gly Val Asp Cys Ala Phe Asp Thr Asp Asn Asp Glu Ala
      65             70             75

ttc atc ttt tat gag aac ttt act gct ctc ata aac tat gct cca cat 292
Phe Ile Phe Tyr Glu Asn Phe Thr Ala Leu Ile Asn Tyr Ala Pro His
      80             85             90

```

```

act tac aat gac aaa atc atc tcg ggt ccg aag aaa atc tcg gac atg 340
Thr Tyr Asn Asp Lys Ile Ile Ser Gly Pro Lys Lys Ile Ser Asp Met
          95                      100                      105

ttt cct ttt ttc aaa gga acc gtg ttt gaa aac ggg att gac gct gca 388
Phe Pro Phe Phe Lys Gly Thr Val Phe Glu Asn Gly Ile Asp Ala Ala
          110                      115                      120

ttc agg tca act aag gag aaa gaa gtt tat tta ttc aaa gga gac ttg 436
Phe Arg Ser Thr Lys Glu Lys Glu Val Tyr Leu Phe Lys Gly Asp Leu
          125                      130                      135                      140

tat gct cgt ata gac tat gga aaa aac tat ctg gtt caa agt atc aag 484
Tyr Ala Arg Ile Asp Tyr Gly Lys Asn Tyr Leu Val Gln Ser Ile Lys
          145                      150                      155

aac att agc act ggg ttc cct tgt ttc act gga acc gtc ttt gaa aat 532
Asn Ile Ser Thr Gly Phe Pro Cys Phe Thr Gly Thr Val Phe Glu Asn
          160                      165                      170

gga gtg gat gct gct ttt gct tct cac agg acc aat gaa gca tac ttt 580
Gly Val Asp Ala Ala Phe Ala Ser His Arg Thr Asn Glu Ala Tyr Phe
          175                      180                      185

ttc aaa gga gat tac tat gca ctt gtc aag att agc ccg ggc gga ata 628
Phe Lys Gly Asp Tyr Tyr Ala Leu Val Lys Ile Ser Pro Gly Gly Ile
          190                      195                      200

gat gac tat att atc ggt ggt gtg aag ccc att ctt gag aat tgg cct 676
Asp Asp Tyr Ile Ile Gly Gly Val Lys Pro Ile Leu Glu Asn Trp Pro
          205                      210                      215                      220

tct ctt cgt ggt ata ata cct cag aaa agt taaatgtggc tctctgtgtg 726
Ser Leu Arg Gly Ile Ile Pro Gln Lys Ser
          225                      230

tgtgtgatat catcagtcaa gtaggtatt aagaataaag actattgttg tcgttgttgt 786

gtgtttcttt ttcattgtgt ttctagttct taatgtttgc ttatgttgtt catgtgaact 846

atgtaatgac atgcactgtg tacgcgcaga gtgaaaataa tatattactg tgtatgttga 906

ttac 910

<210> 3
<211> 230
<212> PRT
<213> 豌豆

<400> SEQ ID NO 3
Met Thr Lys Pro Gly Tyr Ile Asn Ala Ala Phe Arg Ser Ser Phe Asn
  1          5          10          15

```

Gly Glu Arg Tyr Leu Phe Ile Asp Asp Lys Tyr Val Leu Val Asp Tyr
 20 25 30
 Ala Pro Gly Thr Arg Asp Asp Lys Leu Leu Asn Gly Pro Leu Pro Leu
 35 40 45
 Pro Ala Gly Phe Lys Ser Leu Asp Gly Thr Val Phe Gly Thr Tyr Gly
 50 55 60
 Val Asp Cys Ala Phe Asp Thr Asp Asn Asp Glu Ala Phe Ile Phe Tyr
 65 70 75 80
 Glu Asn Phe Thr Ala Leu Ile Asn Tyr Ala Pro His Thr Tyr Asn Asp
 85 90 95
 Lys Ile Ile Ser Gly Pro Lys Lys Ile Ser Asp Met Phe Pro Phe Phe
 100 105 110
 Lys Gly Thr Val Phe Glu Asn Gly Ile Asp Ala Ala Phe Arg Ser Thr
 115 120 125
 Lys Glu Lys Glu Val Tyr Leu Phe Lys Gly Asp Leu Tyr Ala Arg Ile
 130 135 140
 Asp Tyr Gly Lys Asn Tyr Leu Val Gln Ser Ile Lys Asn Ile Ser Thr
 145 150 155 160
 Gly Phe Pro Cys Phe Thr Gly Thr Val Phe Glu Asn Gly Val Asp Ala
 165 170 175
 Ala Phe Ala Ser His Arg Thr Asn Glu Ala Tyr Phe Phe Lys Gly Asp
 180 185 190
 Tyr Tyr Ala Leu Val Lys Ile Ser Pro Gly Gly Ile Asp Asp Tyr Ile
 195 200 205
 Ile Gly Gly Val Lys Pro Ile Leu Glu Asn Trp Pro Ser Leu Arg Gly
 210 215 220
 Ile Ile Pro Gln Lys Ser
 225 230

<210> 4

<211> 1561

<212> ADN

<213> 豌豆

<220>

<221> 启动子

<222> (1)..(1561)

<400> SEQ ID NO 4

acaaccagtgtg tccatatatc accagttcttc tatagacgca ctatcataac tagaagccat 60
 taatagcaca gtgtcatcat ggtgctcaga tccttagaat gttcaattgc tacaacgatg 120

```

taatcaaact gatgagtaag agatctaagt acctttctcaa tgatactttc ctcataaaga 180
gtttctccat gcgacttcat ctcatttgtg atcagaatca ctctagagat gtagtcagat 240
aaccttctcat tgttcttcat gcttagattc tcatactgct cacgtagaga ctgaagtttc 300
accttctaca ctgatgcatc actatcgtag caccacacca gtctgtctca cacaacctt 360
tccgtcattg aatcaacgat tttcttaaac acgttcacat ccacacactg atggatgtag 420
aacaacgcat tctgatcctt cttcctcata tcacactgag catttctttg cgcatccgtt 480
gcattttcta gaagtgaagc ataaacttgc ttgatgagat caagaacatc ttgagcacca 540
aataacacac acatctgaat catccaacga ttccagttgt tgtcgtcga caatggnagc 600
ntggtgcaca gattcacaac gatataattt aanttttggt ttatgaaatt taagaacaaa 660
tttccattat tcttaaaatg ttacacact gatgtagact gcaaaaggaa taaagataca 720
atgtgtcac accactcact tgcgtaaata taagtgagag ttaatgagaa atactaaaat 780
accctctaaa attatgaatt aattctaaca atctctaatt ttagtataat ccattaaaca 840
ctttgatggc aggtataaca aggggtgaag ttagtgtata catattaggc tcttattatt 900
tttatattat ctctgctttt cttcttcatg ttctcactaa tatgatatta tctcccttcc 960
ctaaattatt tatatttatt agaaaaagag tttcattttt taaaaatata ttaccgtaat 1020
ttttcaaaaa ataaaattta aatatatttt ataaaaaat tatttaataa tttatttaca 1080
ttaatgcata aatataaata aatactgtca tttaatattt aaccttttaa caataaatta 1140
tatttattta attcaactaa tataagctaa gttatctcat ccaaccaatt aaaaagatca 1200
tttgaanaata cttttttatt tagtttggg cggtttcaac tgtcaaaaaa aaggaatttt 1260
tacgacgata taaattttaa ccagcaaaaa attgaagcag ttaagcgaac caactcatgg 1320
tatgtggata ttttatctt tgcgttttat atcggattcg aatctctata atgatgaaaa 1380
attaatatca aactttaaat aagaacgtca tttatagagc cattttggga aacacatatt 1440
tcatgtacac gtgattcgca aatttccaat aactctatat atagccctcc tcagtttcat 1500
gcatttgctc acaacataac cttccttgaa ttcgatattc acctaagatg acaaaaccag 1560
g 1561

```

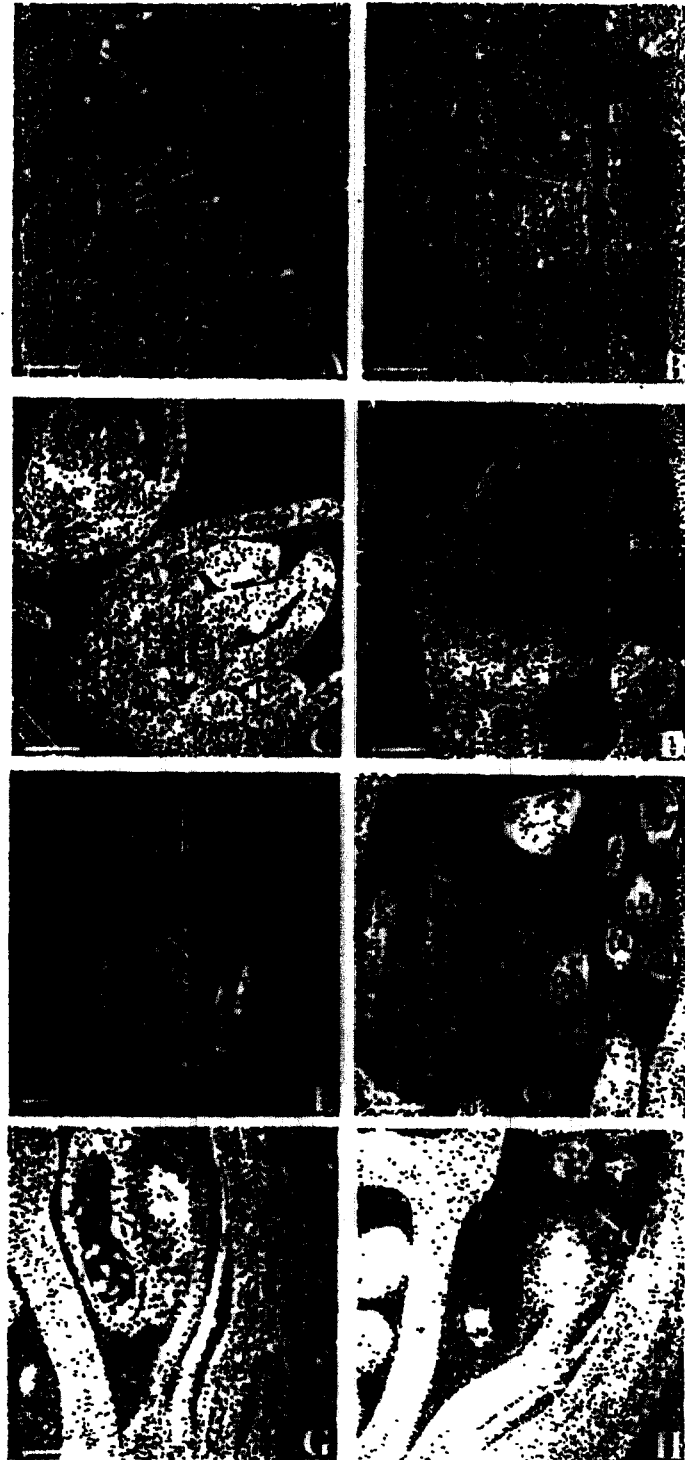


图 1

```

-2736 gacttcaaac ttattagtga atggacaata aaggttataa gctcctttac
-2686 tgtgaagacc caccagtaac atcaccttgc ttataatoatt oagcttottt
-2636 ctagttaacat ttggaacgtg ttataadag aaaaaaaccc aaaaactotg
-2586 aaaagactca cacttttctt atctccagtc caoctctdaa aaggaacaat
-2536 ttocctcagc ttcttggttg gacacctggt gagcacatat gctgcagtg
-2486 caacagtttc tocccacaaa gtgttaggaa gottctcttc cttcagcatg
-2436 ttoccttgta tatcaagcaa agttcgggtt tcaacaagac cattaagtgt
-2386 aggagtatat ggtcagtcct cttoatgctc aattccattc tctttacaga
-2336 acttcttgaa ctctgtagag ttatactcac ctccaccatc agttctgaga
-2286 atcttcagaa gtotgaccac tttatttoto agocttgatt atgaatttct
-2236 taatttcagc aaacaacctg tgtttgaatt ttataagggg tacoccatgtc
-2186 actcttgta actatccat aatgaacata aagtattatt coctootagt
-2136 gaaaggtttg taatgggcca cacacataag aatgcactac toctaacgca
-2086 tgttttgctc tttgagctac ttttgatgaa aatggcagtc ttggttgott
-2036 coctttcatg cacacattac atgacttttt gaattcaaat tooctaagct octaaagttc
-1986 ttocacgtac cagtttcttt caacaaactc ctttccctca caaaccttgt
-1936 aatgaccaa atcttttgtt ctgcagtttt aacattcgcc ttgaatgttt
-1886 tgcgctaagg ctttcagagt atgtctgac tocataatca acttctgata acagtcatac
-1836 taotccttcc gaatgtcatt catggttaot ggaatccct tttcaattaa
-1786 agttcctaca ctcatcagat tgctcttcat gccaaagacg taccacagc
-1736 ttotgaatta atgcagattt tctattatcc ataactactc taacattccc
-1686 catctottta gcaattagtt acttatoatc agaacatota atcttggttt
-1636 ttctcctaga gtcaaaatca accagccatt tottatttcc agtatgtgg
-1586 tttgaacaac cagtgtccat atatcaccag tctctatag acgcactatc
-1536 ataactagaa gccattaata gcaactgttc atcatggtgc tcagatcctt
-1486 agaattgtca attgctacaa cgatgtaatc aaactgatga gtaagagatc
-1436 taagtacctt ctcaatgata ctttccctat aaagagtttc tcoatggac
-1386 ttcatctcat ttgtgatcag aatcactcta gagatgtagt cagataactt
-1336 ctcatctgtc ttcatgctta gattctcata ctgctcaagt agagactgaa
-1286 gtttcacctt ctacactgat gcatoactat cgtagcaccg caccagtctg
-1236 tocoacaca cttttccgtt ctttgaatca acgattttct taacaagttg
-1186 caotocaca caotgatgga tgtagaacaa cgoattotga tecttcttcc
-1136 toatatoaca ctgagcaatt otttgogoat cogttgcaat ttctagaagt
-1086 gagcataaaa cttoyttgat gagatcaaga acatottgag caccaaaata
-986 caacacato tgaatcatoc aacgattcca gttgtgtctg togaacaatg
-936 gnagcctggt gcacagattc acacagatat attataantt ttgttttatg
-886 aattttaaga acaaatcttc attattotta aatgtttac acactgatgt
-836 agactgaaa agaatcaag atacaatttg ttccacccc tcaactgogt
-786 aatataagt gagagttaat gagaataact aaataccct ctaaaattat
-736 gaattaattc taacaatctc taatgttagt ataactcatt aaacactttg
-686 atgpcaggtc taacaaggtt gtaagttagt gtatacatat taggctotta
-636 ttatttttat attatctctg cttttctctc tcatgttctc actaatatga
-586 tattatctoc cttooctaaa ttatttatat ttattagaaa aagagtttca
-536 ttttttaaaa atatattacc gtaatttttc aaaaaataa atttaaatat
-486 attttataaa aaaattattt aataatttat ttacattant gataaatat
-436 aataaantac tgtcatttaa tatttaacct ttttaagaaa aattatattt
-386 atttaattca actaatataa gctaagttat ctoatccaac gaattaaaa
-336 gatcatttga aataaccttt ttatttagtt tgtggcgytt tcaactgtga
-286 aaaaaagga atttttaaga cgtatataat ttaaacccagc aaaaaattga
-236 agcagtttaag cgaaccaact catggtatgt ggatatattt atotttctg
-186 tttatatogy attcgaatct ctataatgat gaaaattaa tatcaactt
-136 taantaagaa cgtcatttat agagocattt tgggaaacac atatttcatg
-86 tacacgtgat togcaaattt ccaataactc tataatagc cctcctcagt
-36 ttcatgcatt tgctcacaac ataaccttcc ttgaattcca TATCTACCTA
AGATGACAAA ACCAGG +1

```

图 2

END1

```

togatatctacctaagatgacaaaaccaggttacattaatgctgcttttcgttcatcttt 60
      M F K P G Y I N A A F R S S Y 15
caacggcgaaacggttacttattcatcgatgataagatggtggttagattatgcaacggg 120
      N G E R Y L F I D D K Y V L V D Y A P G 35
aaccggcgcgataagctcttaaacgggcctctctctctctctgctgggtttaaactact 180
      F R D D K L L N G P L P L P A G F K S L 55
tgatggtacagtatttggaaacctacggagttagctgtgctttgacacccgataacgacga 240
      D G T V F G T Y G V D C A F D T D N D E 75
agcattcatcttttatgagaacttactgctctcataaaactatgctccacatacttacia 360
      A F I F Y H N F T A L I M Y A P H T Y H 95
tgacaaaatcatctcgggtccgaagaaaatctcggacatggttccctttttcaagggaac 360
      D K I I S G P K K I S D M F P F Y K G T 115
cgtgtttgaaaacgggattgacgctgcattcaggtcaactaaggagaagaagtatttttt 440
      V P H N G I D A A F R S T K H K H V Y L 135
attcaaggagacttgtagctcgtatagactatggaaaaactatctgggttcaaagtat 480
      F K G D L Y A R I D Y G K H Y L V Q S I 155
caagaacattagcactgggttcccttgtttcactggaacogtctttgaaaatggagtgga 540
      K H I S T G F P C F T G T V F E N G V D 175
tgctgcttttgcttctcacagaccatgaagcatactttttcaaaggagattactatgc 600
      A A F A S H R T H E A Y F Y K G D Y Y A 195
acttgtcaagattagccggggcgaatagatgactatattatcgggtggtggaagcccat 660
      L V K I S F G G I D D Y I I G G V X P I 215
tcttgagaattggccttctctctcgtggtataatacctcagaaaagttaaattggtgctctc 720
      L H N W F S L R G I I P Q K S * 230
tgtgtgtgtgatatcatcagtcagttatggtattaagaataaagactattgtgtgct 780
      tgttgtgtgttcttttcatggttctctagttctttaaattgtttgcttatggtgtcatg 840
      tgaactatgtaatgacatgcactgtgtacgcagagtgaaaataatatactactgtgta 900
      tgttgattac 910

```

图 2 (续)

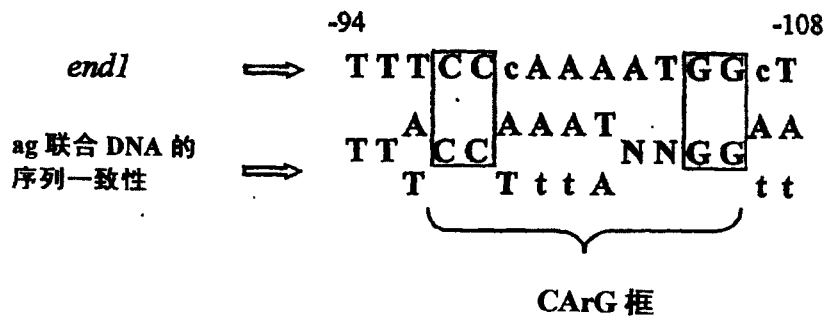


图 3

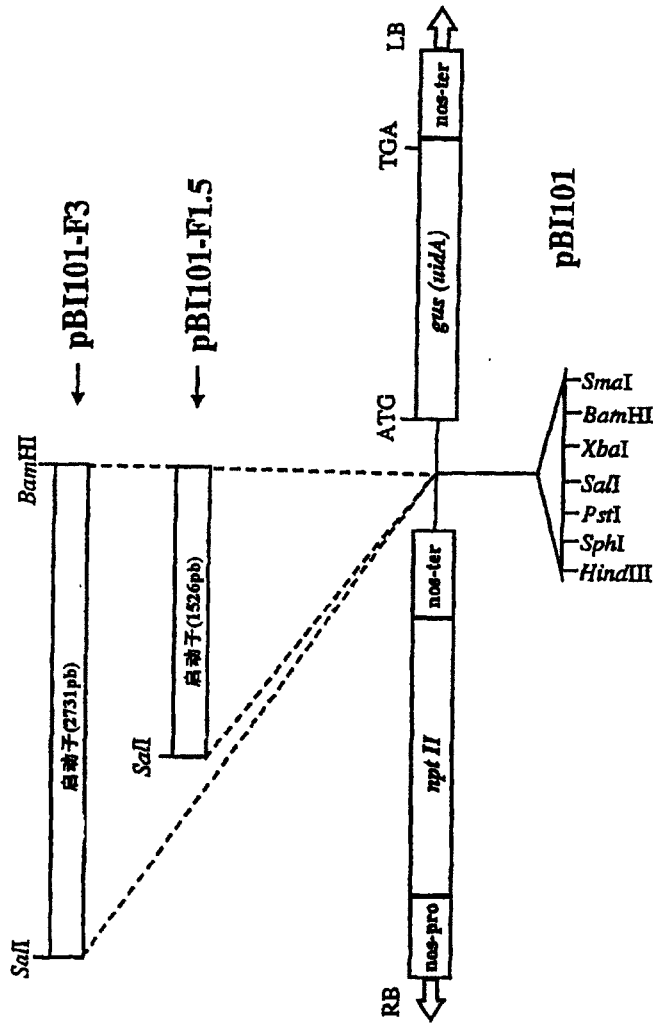


图 4

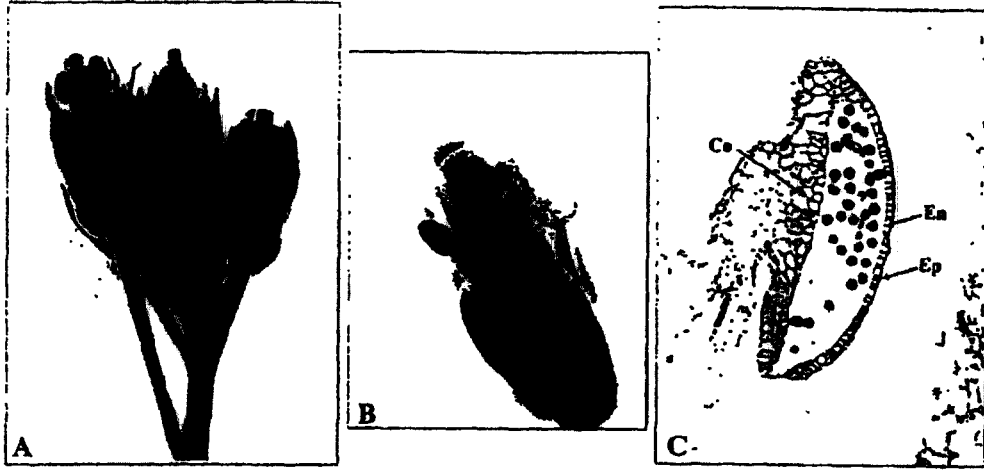


图 5

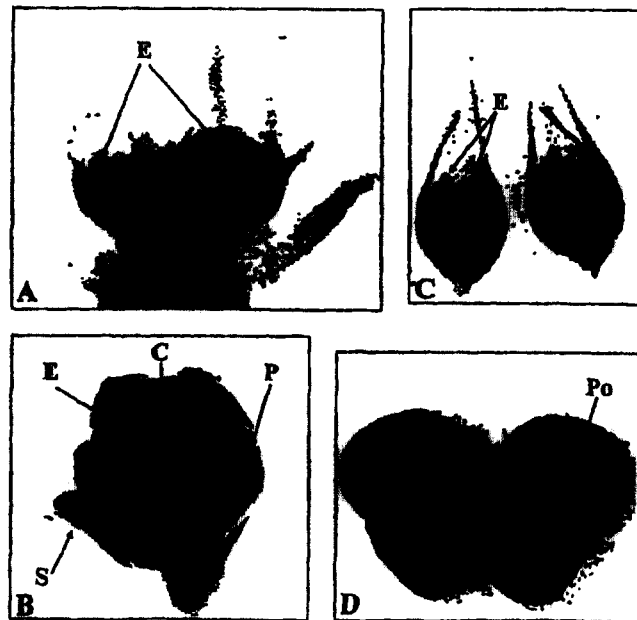


图 6



图 7

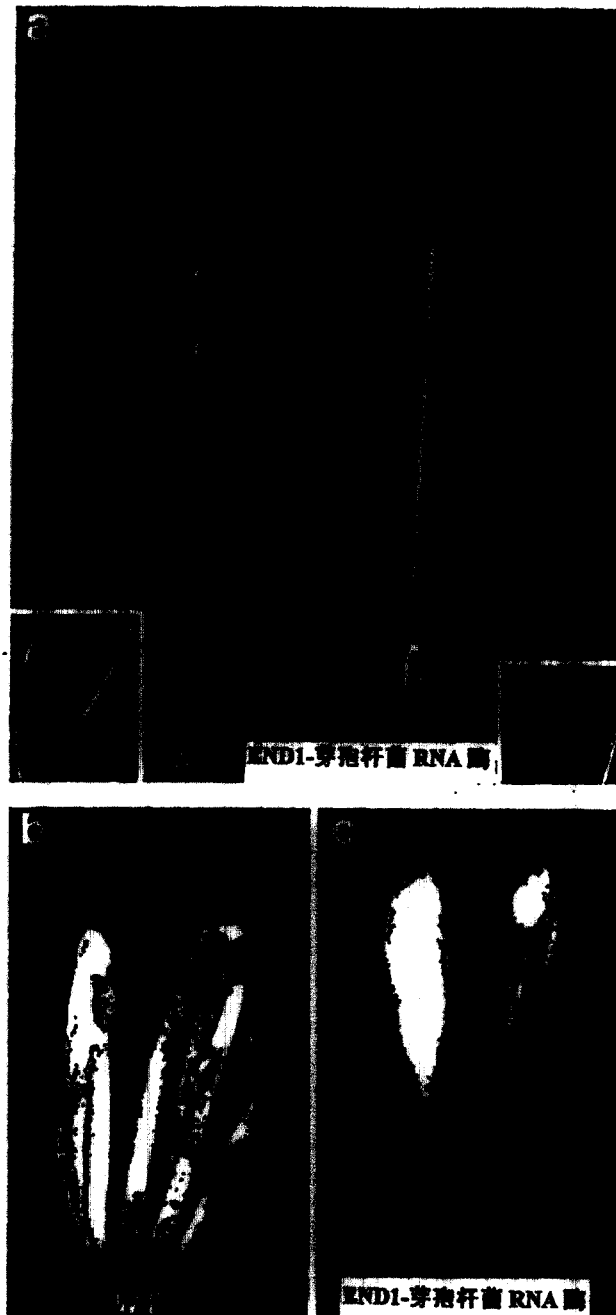


图 8

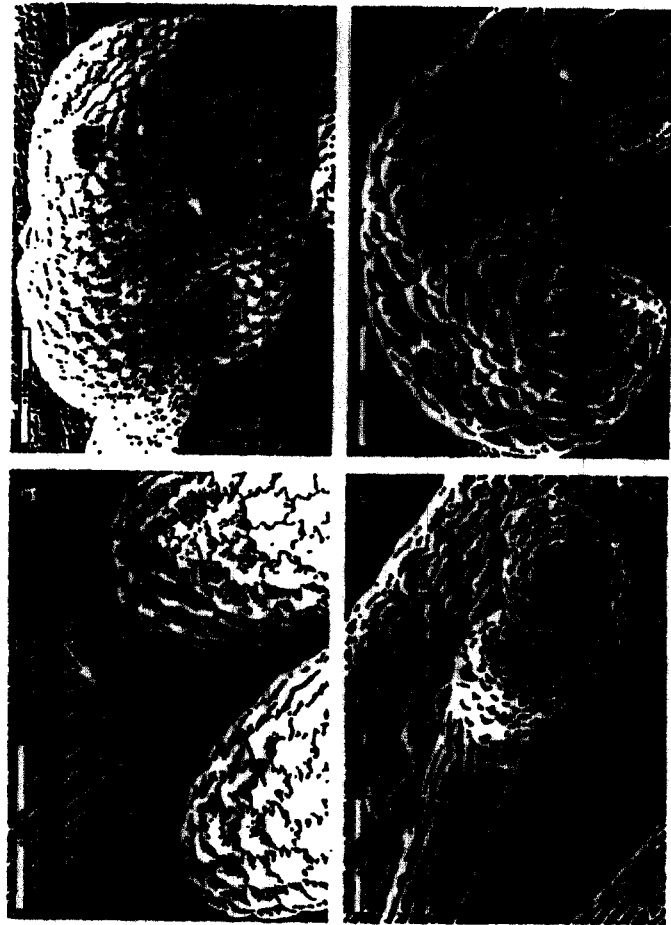


图 9

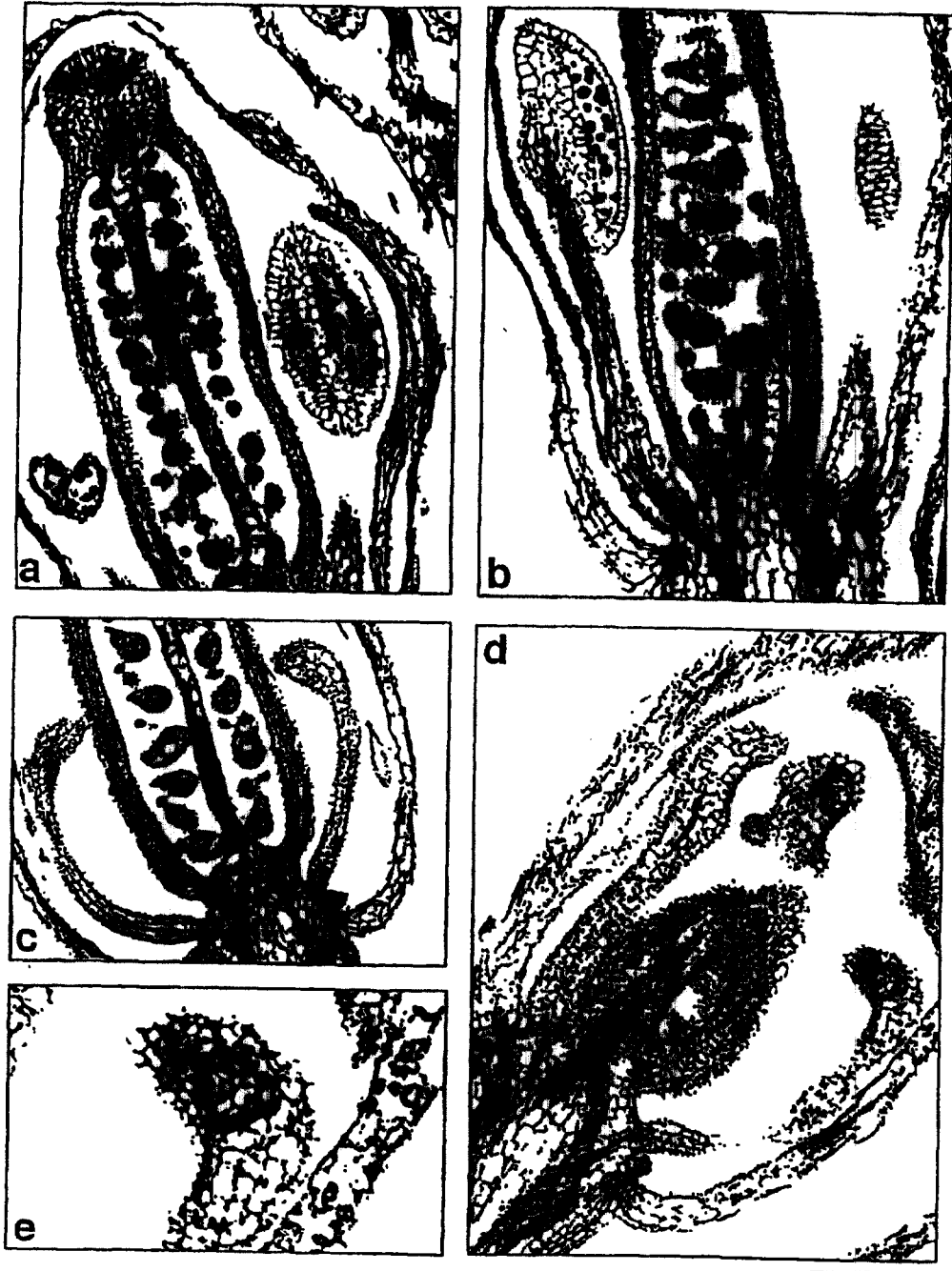


图 10