

[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 00818930.7

[51] Int. Cl.

C12N 15/82 (2006.01)

C07K 14/415 (2006.01)

C12N 7/06 (2006.01)

[45] 授权公告日 2008 年 10 月 15 日

[11] 授权公告号 CN 100425701C

[22] 申请日 2000.12.12 [21] 申请号 00818930.7

[30] 优先权

[32] 1999.12.16 [33] US [31] 60/171,173

[86] 国际申请 PCT/US2000/033633 2000.12.12

[87] 国际公布 WO2001/044457 英 2001.6.21

[85] 进入国家阶段日期 2002.8.12

[73] 专利权人 孟山都技术有限公司

地址 美国密苏里州

[72] 发明人 K·L·芬彻 S·辅拉辛斯基

J·Q·威尔金森

[56] 参考文献

WO9923232 A 1999.5.14

WO0032800 A 2000.6.8

US5106739 A 1992.4.21

WO0037661 A 2000.6.29

WO9200377 A 1992.1.9

WO9905265 A 1999.2.4

"Constitutive and tissue - specific differential expression of the cryIA (b) gene in transgenic rice plants conferring resistance to rice insect pest." DATTA K ET AL.: THEORETICAL AND APPLIED GENETICS, Vol. 97 No. 1.2, 1998

审查员 周 洋

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 徐雁漪

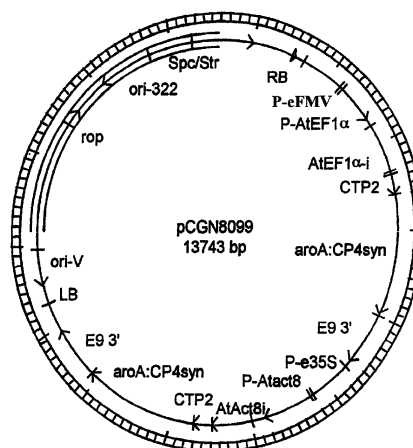
权利要求书 3 页 说明书 83 页 附图 18 页

[54] 发明名称

新型植物表达构建物

[57] 摘要

本发明涉及新型植物表达构建物。更具体地说,本发明提供了包含 5' 调节序列以调节有效连接的基因在植物中表达的 DNA 构建物。



1. 一种嵌合启动子, 所述 DNA 序列包含至少一个花椰菜花叶病毒启动子的顺式增强子元件, 所述顺式增强子元件有效连接至少一个 Efl α 启动子的顺式元件。

2. 权利要求 1 的嵌合启动子, 其中至少一个 Efl α 启动子的顺式元件包含增强子。

3. 权利要求 1 的嵌合启动子, 其中 Efl α 启动子是 SEQ ID NO: 12。

4. 权利要求 1 的嵌合启动子, 其中所述花椰菜花叶病毒启动子是花椰菜花叶病毒(CaMV)或玄参花叶病毒(FMV)启动子。

5. 权利要求 4 的嵌合启动子, 其中所述 CaMV 是 35S CaMV。

6. 权利要求 4 的嵌合启动子, 其中所述 FMV 是 34S FMV。

7. 权利要求 1 的嵌合启动子, 其中所述嵌合启动子是 FMV:Efl α 。

8. 权利要求 7 的嵌合启动子, 其中所述嵌合启动子是 SEQ ID NO:28。

9. 一种 DNA 构建物, 所述 DNA 构建物包含至少一个含有权利要求 1-8 任一项的嵌合启动子的表达盒和结构 DNA 序列; 其中所述结构 DNA 序列有效连接所述嵌合启动子。

10. 权利要求 9 的 DNA 构建物, 其中所述结构 DNA 序列编码一种除草剂抗性蛋白。

11. 权利要求 10 的 DNA 构建物, 其中所述结构 DNA 序列编码一种草甘膦抗性蛋白。

12. 权利要求 11 的 DNA 构建物, 其中所述草甘膦抗性蛋白是 5-烯醇丙酮酸莽草酸-3-磷酸合酶(EPSPS)蛋白或草甘膦氧化还原酶(GOX)蛋白。

13. 权利要求 11 的 DNA 构建物, 其中所述草甘膦抗性蛋白是

aroA:CP4 EPSPS 蛋白。

14. 权利要求 11 的 DNA 构建物, 其中编码草甘膦耐性蛋白的结构 DNA 序列进一步与编码叶绿体转运肽(CTP)的结构 DNA 序列有效连接。

15. 权利要求 14 的 DNA 构建物, 其中所述叶绿体转运肽是拟南芥属 CTP2。

16. 权利要求 11 的 DNA 构建物, 其中编码草甘膦耐性蛋白的结构 DNA 序列进一步与转录终止子有效连接。

17. 权利要求 16 的 DNA 构建物, 其中所述转录终止子是 E9。

18. 权利要求 9 的 DNA 构建物, 其中至少一种表达盒包含与编码 *aroA*:CP4 EPSPS 的结构 DNA 序列有效连接的 FMV:EF1 α , 所述结构 DNA 序列进一步与编码 CTP2 的结构 DNA 序列和 E9 3'终止区有效连接。

19. 权利要求 9 的 DNA 构建物, 其包含第一和第二表达盒, 其中第一表达盒包含与编码 *aroA*:CP4 EPSPS 的结构 DNA 序列有效连接的 FMV:EF1 α , 所述结构 DNA 序列进一步与编码 CTP2 的结构 DNA 序列和 E9 3'终止区有效连接; 第二表达盒包含与编码 *aroA*:CP4 EPSPS 的结构 DNA 序列有效连接的 CaMV:Act8, 所述结构 DNA 序列进一步与编码 CTP2 的结构 DNA 序列和 E9 3'终止区有效连接。

20. 权利要求 9 的 DNA 构建物, 其中所述结构 DNA 序列是苏云金芽孢杆菌(*Bacillus thuringiensis*)昆虫防治基因。

21. 转基因植物细胞, 其包含权利要求 9-20 任一项的 DNA 构建物。

22. 一种制备转基因植物的方法, 其包括由权利要求 21 的转基因植物细胞再生转基因植物, 其中所述转基因植物包含所述 DNA 构建物。

23. 权利要求 22 的方法, 其中转基因植物还具有营养性和生殖性草甘膦耐性。

24. 权利要求 23 的方法, 其中转基因植物能够耐受暴露于最高达每 4000 平方米 7248 克的草甘膦。

25. 权利要求 24 的方法, 其中转基因植物能够耐受暴露于每 4000 平方米 453-1812 克的草甘膦。

26. 权利要求 22-25 任一项的方法, 其中所述转基因植物是单子叶作物。

27. 权利要求 26 的方法, 其中所述单子叶作物是大麦、玉米、水稻、黑麦、高粱或小麦。

28. 权利要求 22-25 任一项的方法, 其中所述转基因植物是双子叶作物。

29. 权利要求 28 的方法, 其中所述双子叶作物是番茄、大豆、棉花、canola、苜蓿或向日葵。

30. 权利要求 29 的方法, 其中所述双子叶作物是棉花。

31. 权利要求 30 的方法, 其中在 4 叶期之后, 所述棉花在暴露于草甘膦后的棉铃图值大于或等于 3, 其中所述棉铃图值是通过以下方法测定的: 用每 4000 平方米 1359 克的草甘膦喷洒 4 叶期之后的棉花, 使棉花生长至成熟, 测定头五个棉铃的最初位置上棉铃的数目。

32. 一种在植物中表达结构 DNA 序列的方法, 该方法包括: (1) 提供权利要求 9-20 任一项的 DNA 构建物; (2) 将所述 DNA 构建物引入植物细胞; 然后(3)使所述植物细胞再生而产生植株, 使得所述结构 DNA 序列在所述植株中可表达。

33. 一种防除杂草的方法, 该方法包括: (1)提供用权利要求 10-19 任一项的 DNA 构建物转化的作物; 和(2)对所述作物施用足量草甘膦以防除杂草而不损害所述作物。

34. 权利要求 1-8 任一项的嵌合启动子在构建转基因植物中的用途, 其中所述转基因植物包含所述嵌合启动子。

新型植物表达构建物

发明领域

本发明涉及分离和应用核酸分子以控制植物中的基因表达，所述核酸分子具体是新型植物启动子。

发明背景

植物遗传工程的一个目标是产生具有重要农学特征或性状的植物。遗传工程的最近发展已经提供了产生包含和表达外源基因的转基因植物的必要工具(Kahl 等, World J. of Microbiol. Biotech. 11: 449-460, 1995)。对于植物遗传工程尤其需要的目的性状或品质包括但不限于：昆虫抗性、真菌病抗性、对其它害虫和病原体的抗性、除草剂耐性、更高的稳定性或更长的保存期、更高产量、环境耐性以及营养强化(nutritional enhancement)。植物转化和再生领域内的技术进步已经使得研究者能够从异源来源或天然来源提取外源 DNA (如一个基因或多个基因)，然后将所述外源 DNA 掺入植物基因组。在一种方法中，通常不在特定植物或特定植物组织中表达的新型基因的表达可能赋予所需的表型效应。在另一种方法中，以反义取向的基因或基因部分的转录可能通过防止或抑制内源基因的表达而产生所需效应。

为产生转基因植物，将包括异源基因序列的构建物引入植物细胞，所述异源基因序列当在植物中表达时赋予所需表型。所述构建物还包括有效连接所述异源基因序列的植物启动子，该植物启动子通常在一般情况下不与所述异源基因连接。将所述构建物引入植物细胞，产生转化植物细胞，然后将所述转化植物细胞再生为转基因植物。所述启动子控制该启动子有效连接的所引入 DNA 序列的表达，并因此影响所述 DNA 序列所赋予的所需特性。

利用多种启动子定制基因表达可能是有利的，以便一个基因或多个基因在植物生长和发育的恰当时间、在所述植物中的最佳位点并且以产生所需效应所必需的量有效转录。例如，基因产物的组成型表达在植物的一个位点可能是有利的，但在该植物的另一部分就不是那么有利。在其它情况下，在植物某个发育阶段或在应答某些环境或化学刺激时产生基因产物可能是有利的。遗传改良种质的商业化发展也已经前进到将多种性状引入作物的阶段，该方法常常被称为基因堆积方法(gene stacking approach)。在该方法中，可以将赋予不同目的性状的多种基因引入植物。当将多种基因引入植物时，重要的是调整或控制每个基因以获得最佳表达，并且使得调节元件多样化，以降低基因沉默的可能性。根据这些和其它的考虑，在植物生物工程技术中基因表达的最佳控制和调节元件多样化显然是重要的。

发明概述

本发明涉及包含以下元件的 DNA 植物表达构建物：拟南芥属肌动蛋白(Act)启动子序列 Act1a、Act1b 和延伸因子 1 α (EF1 α)启动子序列、以及由有效连接于在作物细胞中起作用的异源结构基因序列的这些启动子获得的片段和顺式元件。

因此，根据本发明的一个实施方案，提供以有效连接包含下列元件的重组 DNA 构建物：在作物细胞中起作用的启动子，所述启动子包括：来自 SEQ ID NO:12、SEQ ID NO:22、SEQ ID NO:23 的至少一种顺式元件；对于所述启动子异源的结构 DNA 序列；和在植物中起作用的 3'非翻译区，该 3'非翻译区引起在 RNA 序列的 3'末端添加聚腺苷酸化核苷酸。例如，所述启动子可以主要包含来自下面任何一种的 5'调节区：SEQ ID NO:12、SEQ ID NO:22、SEQ ID NO:23 (包括或不包括它们中间的任何内含子序列)。所述结构基因可以包含任何异源核苷酸序列，其中所述序列的表达导致转基因作物中在农学上有用的性状或产品。

根据本发明的一方面是一种 DNA 构建物, 该 DNA 构建物包含有效连接本发明启动子序列的结构 DNA 序列, 所述结构 DNA 序列编码赋予作物除草剂耐性的蛋白。该除草剂耐性蛋白包括但不限于草甘膦耐性蛋白基因如单独的草甘膦抗性 EPSP 合酶基因, 或者该基因与一种或多种草甘膦降解蛋白基因的组合。

根据本发明的另一实施方案, 提供如上文所述的那些 DNA 构建物, 其中所述启动子是杂种或嵌合启动子, 包含有效连接异源启动子序列的由选自以下的一种或多种序列获得的至少一种顺式元件: SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:25 和 SEQ ID NO:26, 所述异源启动子序列如花椰菜花叶病毒启动子, 例如花椰菜花叶病毒 35S 启动子或玄参花叶病毒启动子。

根据本发明的又一实施方案, 提供串联的例如如上文所述的 DNA 构建物, 其中所述启动子是杂种或嵌合启动子, 包含有效连接在转基因作物细胞中表达的异源基因序列的由选自以下的一种或多种序列获得的至少一种顺式元件: SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:25 和 SEQ ID NO:26。所述嵌合启动子序列更具体地是包含在 SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29 和 SEQ ID NO:30 中定义的序列。

根据本发明的再一实施方案, 提供例如如上文所述的 DNA 构建物, 其中所述结构 DNA 序列是草甘膦耐性基因, 使得当将所述 DNA 构建物引入植物细胞时, 它赋予所述植物细胞对于每升至少包括 50 克酸当量(g a.e./l)草甘膦的草甘膦水溶液制剂的耐性。在其它相关实施方案中, 所述 DNA 构建物赋予所述植物细胞对于草甘膦浓度更高(例如每升至少 300 克酸当量草甘膦)的草甘膦制剂的耐性。根据一个实施方案, 所述 DNA 构建物赋予所述植物细胞对于以下除草剂施用的耐性: 例如以每 4,000 平方米 453 ml 的比率施用 Roundup Ultra®至少一次,

而在其它实施方案中，草甘膦耐性扩展到例如以每 4,000 平方米 453 ml、每 4,000 平方米 906 ml 或每 4,000 平方米 1812 ml 施用一到二次或更多次。

根据本发明的另一实施方案，提供用上文所述 DNA 构建物转化的转基因作物，其中包括单子叶植物物种和双子叶植物物种。我们已经发现：当使用拟南芥属肌动蛋白和拟南芥属 EF1 α 启动子在其它作物物种例如棉花、番茄和向日葵中控制草甘膦耐性基因(如 aroA:CP4)表达时，所述启动子充分有活性，所述植物耐受草甘膦的商业化施用比率，展现良好的营养性耐性以及繁殖组织的损伤低。也可以使用这样的启动子在植物中表达其它目的基因，所述目的基因包括但不限于赋予下列性状的基因：除草剂耐性、昆虫防制、抗病性、稳定性增加或保存期增加、产量提高、营养强化、表达药用多肽产物或其它需要的多肽产物、或植物生理学或形态学的所需改变等等。

根据本发明的另一实施方案，提供用多种 DNA 构建物转化的转基因植物，所述 DNA 构建物包含拟南芥属肌动蛋白和拟南芥属 EF1 α 启动子，所述启动子在其它作物物种例如棉花、番茄和向日葵中控制草甘膦耐性基因(如 aroA:CP4)表达时充分有活性，所述植物耐受草甘膦的商业化施用比率，展现良好的营养性耐性以及繁殖组织的损伤低。也可以使用这样的启动子在植物中表达其它目的基因，所述目的基因包括但不限于赋予下列性状的基因：除草剂耐性、昆虫防制、抗病性、稳定性增加或保存期增加、产量提高、营养强化、表达药用多肽产物或其它需要的多肽产物、或植物生理学或形态学的所需改变等等。

根据本发明的再一实施方案，提供用 DNA 构建物转化的转基因作物，所述 DNA 构建物包含拟南芥属肌动蛋白和拟南芥属 EF1 α 启动子作为与花椰菜花叶病毒 DNA 分子融合的嵌合 DNA 分子，所述嵌合

DNA 分子在植物中具有启动子活性,在其它作物物种例如棉花、番茄和向日葵中充分有活性,例如当用来控制草甘膦耐性基因(如 aroA:CP4)表达时,所述植物耐受草甘膦的商业化施用比率,展现良好的营养体耐性以及对生殖组织的损伤低。也可以使用这样的启动子在植物中表达其它目的基因,所述目的基因包括但不限于赋予下列性状的基因:除草剂耐性、昆虫防制、抗病性、稳定性增加或保存期增加、产量提高、营养强化、表达药用多肽产物或其它需要的多肽产物、或植物生理学或形态学的所需改变等等。

根据本发明的又一实施方案,提供在植物中表达结构 DNA 序列的方法。这样的方法包括:提供如上文所述的 DNA 构建物,将所述 DNA 构建物引入植物细胞,然后再生所述植物细胞产生植株,以便在所述植物中可表达所述结构 DNA。根据一个相关实施方案,提供防除杂草的方法,其中所述 DNA 构建物包含草甘膦耐性基因,对用所述 DNA 构建物转化的作物施用一次其用量足以防除杂草而不会显著损害所述作物的草甘膦。

附图简述

图 1 是 pCGN8086 的质粒图谱。

图 2 是 pMON45325 的质粒图谱。

图 3 是 pMON45331 的质粒图谱。

图 4 是 pMON45332 的质粒图谱。

图 5 是 pCGN9190 的质粒图谱。

图 6 是 pCGN9153 的质粒图谱。

图 7 是 pCGN8099 的质粒图谱。

图 8 是 pCGN8088 的质粒图谱。

图 9 是 pCGN8068 的质粒图谱。

图 10 是 pCGN8096 的质粒图谱。

图 11 是 pCGN9151 的质粒图谱。

图 12 是 pMON10156 的质粒图谱。

图 13 是 pMON52059 的质粒图谱。

图 14 是 pMON54952 的质粒图谱。

图 15 是 pMON54953 的质粒图谱。

图 16 是 pMON54954 的质粒图谱。

图 17 是 pMON54955 的质粒图谱。

图 18 是 pMON54956 的质粒图谱。

序列表简述

SEQ ID NO:1 是用于分离 Act2 启动子的正向 PCR 引物。

SEQ ID NO:2 是用于分离 Act2 启动子的反向 PCR 引物。

SEQ ID NO:3 是用于分离 Act8 启动子的正向 PCR 引物。

SEQ ID NO:4 是用于分离 Act8 启动子的反向 PCR 引物。

SEQ ID NO:5 是用于分离 Act11 启动子的正向 PCR 引物。

SEQ ID NO:6 是用于分离 Act11 启动子的反向 PCR 引物。

SEQ ID NO:7 是用于分离 EF1 启动子的正向 PCR 引物。

SEQ ID NO:8 是用于分离 EF1 启动子的反向 PCR 引物。

SEQ ID NO:9 是 Act2 启动子的序列，其中包括 Act2 基因的内含子序列。碱基位置 1-764 代表启动子序列；碱基位置 765-1215 代表内含子，其后是在 ATG 之前的 5 碱基 5'非翻译区(5' UTR)；转录起始位点位于碱基位置 597。

SEQ ID NO:10 是 Act8 启动子的序列，其中包括 Act8 基因的第一个内含子。碱基位置 1-797 代表启动子序列；碱基位置 798-1259 代表内含子，其后是在 ATG 之前的 10 碱基 5'UTR；转录起始位点位于碱基位置 646。

SEQ ID NO:11 是 Act11 启动子的序列，其中包括 Act11 基因的第一个内含子。碱基位置 1-1218 代表启动子序列；碱基位置 1219-1381 代表内含子，其后是在 ATG 之前的 10 碱基 5'UTR；转录起始位点位于

于碱基位置 1062。

SEQ ID NO:12 是 EF1 启动子的序列，其中包括 EF1 基因的第一个内含子。碱基位置 1 – 536 代表启动子序列；碱基位置 537 – 1137 代表内含子，其后是在 ATG 之前的 22 碱基 5'UTR；转录起始位点位于碱基位置 481。

SEQ ID NO:13 是用于分离 Act1a 启动子的正向 PCR 引物。

SEQ ID NO:14 是用于分离 Act1b 启动子的正向 PCR 引物。

SEQ ID NO:15 是用于分离 Act1a 和 Act1b 启动子的反向 PCR 引物。

SEQ ID NO:16 是用于分离 Act3 启动子的正向 PCR 引物。

SEQ ID NO:17 是用于分离 Act3 启动子的反向 PCR 引物。

SEQ ID NO:18 是用于分离 Act7 启动子的正向 PCR 引物。

SEQ ID NO:19 是用于分离 Act7 启动子的反向 PCR 引物。

SEQ ID NO:20 是用于分离 Act12 启动子的正向 PCR 引物。

SEQ ID NO:21 是用于分离 Act12 启动子的反向 PCR 引物。

SEQ ID NO:22 是 Act1a 启动子的序列，其中包括 Act1a 基因的第一个内含子。碱基位置 1 – 1033 代表启动子序列；碱基位置 1034 – 1578 代表内含子和 5'UTR。

SEQ ID NO:23 是 Act1b 启动子的序列，其中包括 Act1b 基因的第一个内含子。碱基位置 1 – 914 代表启动子序列；碱基位置 915 – 1468 代表内含子和 5'UTR 序列。

SEQ ID NO:24 是 Act3 启动子的序列，其中包括 Act3 基因的第一个内含子。碱基位置 1 – 1023 代表启动子序列；碱基位置 1024 – 1642 代表内含子和 5'UTR 序列。

SEQ ID NO:25 是 Act7 启动子的序列，其中包括 Act7 基因的第一个内含子。碱基位置 1 – 600 代表启动子序列；碱基位置 601 – 1241 代表内含子和 5'UTR 序列。

SEQ ID NO:26 是 Act12 启动子的序列，其中包括 Act12 基因的第

一个内含子。碱基位置 1-1017 代表启动子序列；碱基位置 1018-1313 代表内含子和 5'UTR 序列。

SEQ ID NO:27 是嵌合 FMV-Act11 启动子的序列，其中包括 Act11 基因的第一个内含子。碱基位置 1-536 代表 FMV 启动子序列；碱基位置 553-1946 代表拟南芥属肌动蛋白 11 启动子、内含子和 5'UTR 序列。

SEQ ID NO:28 是嵌合 FMV-EF1 α 启动子的序列，其中包括 EF1 α 基因的第一个内含子。碱基位置 1-536 代表 FMV 启动子序列；碱基位置 553-1695 代表 EF1 α 启动子、内含子和 5'UTR 序列。

SEQ ID NO:29 是 CaMV-Act8 启动子的序列，其中包括 Act8 基因的第一个内含子。碱基位置 1-523 代表 CaMV 启动子序列；碱基位置 534-1800 代表 Act8 启动子、内含子和 5'UTR 序列。

SEQ ID NO:30 是 CaMV-Act2 启动子的序列，其中包括 Act2 基因的第一个内含子。碱基位置 1-523 代表 CaMV 启动子序列；碱基位置 534-1742 代表 Act2 启动子、内含子和 5'UTR 序列。

发明详述

本申请要求于 99 年 12 月 16 日申请的美国临时申请第 60/171,173 号的权益。提供下面的定义和方法以更好地定义本发明，并指导本领域内的一般技术人员实施本发明。除特别指出，应当按照相关领域内一般技术人员的常规使用来理解术语。使用在 37 CFR § 1.822 陈述的 DNA 碱基命名法。使用标准的氨基酸残基单字母命名法和三字母命名法。

“核酸(序列)”或“多核苷酸(序列)”指基因组来源或合成来源的单链或双链 DNA 或 RNA，即分别是从 5'(上游)端读到 3'(下游)端的脱氧核糖核苷酸碱基或核糖核苷酸碱基的聚合物。核酸可以代表有义链或互补(反义)链。

“天然的”指天然出现的(“野生型”)核酸序列。

“异源的”序列指源自外来来源或物种的序列，或者当来自同一来源时指由其原始形式受到修饰的序列。

“分离的”核酸序列是与所述核酸天然出现的生物细胞中其通常结合的其它核酸序列(即其它染色体 DNA 或染色体外 DNA)基本分离或纯化出来的。该术语包括经过生物化学纯化以便基本去除污染的核酸和其它细胞成份的核酸。该术语还包括重组核酸和化学合成的核酸。

本文所用术语“基本纯化的”指与其天然状态通常结合的其它分子分离开来的分子。更优选基本纯化的分子是在制备物中主要存在的种类。基本纯化的分子可以不含有天然混合物中存在的 60%、优选 75%、更优选 90%其它分子(不包括溶剂)。术语“基本纯化的”并不包括以天然状态存在的分子。

假如第一种核酸序列在与参考核酸(或其互补链)进行最佳序列对比(在整个比较窗口上,适当地插入或缺失低于参考序列的 20%的核苷酸)时,在至少 20 个核苷酸位置、优选至少 50 个核苷酸位置、更优选至少 100 个核苷酸位置、最优选所述第一种核酸全长的比较窗口上,存在至少约 75%核苷酸序列同一性、优选至少约 80%同一性、更优选至少约 85%同一性、最优选至少约 90%同一性,那么该种核酸序列就与所述参考序列显示“基本同一性”。可以通过下面方法进行比较窗口的最佳序列对比: Smith 和 Waterman, Adv. Appl. Math. 2:482, 1981 的局部同源性算法; Needleman 和 Wunsch, J. Mol. Biol. 48: 443, 1970 的同源性序列对比算法; Pearson 和 Lipman, Proc. Natl. Acad. Sci. USA 85: 2444, 1988 的相似性搜寻方法; 优选使用在 Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, WI 上这些算法的计算机化实现形式(GAP、BESTFIT、FASTA 和 TFASTA)。所述参考核酸可以是全长分子或较长分子的一部分。或者,假如一种核酸分子在严格条件下与另一种核酸分子杂交,如下文所定义,那么这两种核酸分子就具有基本同一性。

假如两种核酸序列的布置使得第一种核酸序列影响第二种核酸序列的功能,那么所述第一种核酸序列就与所述第二种核酸序列“有效连接”。优选这两种序列是单一连续核酸分子的组成部分,或者更优选是临近的。例如,假如一种启动子调节或介导一种基因在细胞中的转录,那么所述启动子就与所述基因有效连接。

“重组”核酸是通过人工组合两种在其它情况下是分离的序列区段而获得的,例如通过化学合成或通过遗传工程技术操作分离的核酸区段。进行核酸操作的技术是众所周知的(参见例如 Sambrook 等, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, 1989; Mailga 等, *Methods in Plant Molecular Biology*, Cold Spring Harbor Press, 1995; Birren 等, *Genome Analysis*: 第一卷, 分析 DNA, (1997), 第二卷, 检测基因, (1998), 第三卷, 克隆系统, (1999), 第四卷, 基因组作图, (1999), Cold Spring Harbor, New York)。

化学合成核酸的方法在例如 Beaucage 和 Carruthers, *Tetra. Letts.* 22: 1859-1862, 1981 和 Matteucci 等, *J. Am. Chem. Soc.* 103: 3185, 1981 中讨论。可以例如在商业化的自动寡核苷酸合成仪上进行核酸的化学合成。

可以设计并化学合成“合成核酸序列”以增强在特定宿主细胞中的表达并用于克隆进合适的构建物。宿主细胞常常显示优选的密码子选择(Murray 等, 1989)。因此设计用于增强在特定宿主中表达的合成 DNA 应当反映在该宿主细胞中的密码子选择模式。可以获得用于这些目的的计算机程序,包括但不限于 Sequence Analysis Software Package, Genetics Computer Group, Inc., University of Wisconsin Biotechnology Center, Madison, WI 53711 的“BestFit”或“Gap”程序。

核酸“扩增”或“核酸复制”指产生额外的核酸序列拷贝,并使用聚合酶链式反应(PCR)技术进行。多种扩增方法是本领域内已知的并且已经得到描述,尤其是美国专利第 4,683,195 号和第 4,683,202 号以及 PCR Protocols: A Guide to Methods and Applications, Innis 等编辑,

Academic Press, San Diego, 1990。在 PCR 中，引物指退火到 DNA 模板以起始聚合酶链式反应的确定序列的短寡核苷酸。

“转化的”、“转染的”或“转基因”指已经引入外源核酸(如重组构建物)的细胞、组织、器官或生物。所引入的核酸最好整合入受体细胞、组织、器官或生物的基因组 DNA，以便所引入的核酸通过随后的后代遗传。“转基因”或“转化的”细胞或生物还包括所述细胞或生物的后代，以及由使用所述“转基因”植物作为杂交亲本的育种程序产生、并表现由于重组构建物或构建物的存在而引起的表型改变的后代。

术语“基因”指染色体 DNA、质粒 DNA、cDNA、合成 DNA 或其它编码肽、多肽、蛋白或 RNA 分子的 DNA、以及在编码序列两侧参与表达调节的区。一些基因可以转录成为 mRNA 并翻译成为多肽(结构基因)；而其它基因可以转录成为 RNA (如 rRNA、tRNA)；其它类型基因作为表达调节物起作用(调节基因)。

基因“表达”指基因转录产生对应 mRNA 并且该 mRNA 翻译产生对应基因产物，即肽、多肽或蛋白。调节元件控制或调整基因表达，所述调节元件包括 5' 调节元件如启动子。

“遗传成分”指可以构成表达构建物的成分或部分的任何核酸序列或遗传元件。遗传成分的例子包括但不限于启动子区、5' 非翻译前导区、内含子、基因、3' 非翻译区和其它调节序列或影响一个或多个核酸序列转录或翻译的序列。

术语“重组 DNA 构建物”、“重组构建物”、“表达构建物”或“表达盒”指来自任何来源、能够整合入基因组或自主复制的任何因子如质粒、粘粒、病毒、BAC (细菌人工染色体)、自主复制型序列、噬菌体、或者线性或环状单链或双链 DNA 或 RNA 核苷酸序列，包括其中一种或多种 DNA 序列使用众所周知的重组 DNA 技术以功能性可操作方式连接的 DNA 分子。

“互补”指核酸序列通过碱基配对(A-G-T 与互补序列 T-C-A 配对)

的天然结合。假如两个单链分子间只有部分核酸对是互补的，那么它们之间的互补性是部分的；假如所有碱基对都是互补的，那么它们的互补性就是完全的。互补性的程度影响杂交和扩增反应的效率和强度。

“同源性”指核酸序列或氨基酸序列分别按照核苷酸或氨基酸位置同一性百分率而言的相似性水平(即序列相似性或同一性)。同源性也指在不同核酸或蛋白之间相似功能特性的概念。

“启动子”指位于基因的可读框(或蛋白编码区)翻译起始密码子5'或上游、涉及RNA聚合酶II和其它蛋白(反式作用转录因子)的识别和结合以起始转录的核酸序列。“植物启动子”是在植物细胞中有功能的天然或非天然启动子。组成型启动子在植物的整个发育过程在大多数或所有植物组织中起作用。组织特异性启动子、器官特异性启动子或细胞特异性启动子分别仅在或主要在特定组织、器官或细胞类型中表达。启动子可能在植物的一个部分(如细胞类型、组织或器官)比在植物的其它部分呈现“增强的”表达(即更高的表达水平)，而不是在特定组织、器官或细胞类型中“特异性”表达。时序调节的启动子仅在或主要在植物发育的某些时期或一天内的某些时刻起作用，例如与昼夜节律有关的基因就是这种情况。诱导型启动子应答内源刺激或外源刺激的存在，例如化合物(化学诱导物)，或应答环境信号、激素信号、化学信号和/或发育信号，选择性表达有效连接的DNA序列。诱导型或调节型启动子包括例如由光、热、胁迫、洪涝或干旱、植物激素、创伤或化学药品如乙醇、茉莉酮酸酯、水杨酸或安全剂调节的启动子。

可以用任何植物启动子作为5'调节序列以调节特定一个基因或多个基因的表达。一种优选的启动子是植物RNA聚合酶II启动子。植物RNA聚合酶II启动子象其它高等真核生物的RNA聚合酶II启动子一样，有复杂的结构并由几个独立元件组成。一种这样的元件是真核基因在体外正确表达和在体内准确有效地起始转录所必需的TATA框或

Goldberg-Hogness 框。TATA 框通常位于约-25 到-35, 即在定义为位置 +1 的转录起始位点或加帽位点上游(5') 25 到 35 碱基对(bp) (Breathnach 和 Chambon, *Ann. Rev. Biochem.* 50: 349-383, 1981; Messing 等, *Genetics Engineering of Plants*, Kosuge 等编辑, 第 211-227 页, 1983)。另一个常见元件 CCAAT 框位于-70 和-100 bp 之间。在植物中, CCAAT 框可以具有与哺乳动物启动子的功能类似序列不同的共有序列(所述植物类似物已经命名为“AGGA 框”, 以将其与动物中的类似元件区分开来; Messing 等, *Genetics Engineering of Plants*, Kosuge 等编辑, 第 211-227 页, 1983)。此外, 事实上所有启动子都包括从约-100 bp 延伸到-1,000 bp 或转录起始位点更上游的其它上游激活序列或增强子(Benoist 和 Chambon, *Nature* 290: 304-310, 1981; Gruss 等, *Proc. Nat. Acad. Sci. USA* 78: 943-947, 1981; 和 Khoury 和 Gruss, *Cell* 27: 313-314, 1983)。

当所述启动子与异源 DNA 序列融合时, 所述启动子通常引起所融合的序列以类似于所述启动子正常连接的基因序列的转录方式进行转录。可以加入包括调节序列的启动子片段(例如融合到具有其自身的部分或完全调节序列的活性启动子的 5'端, 或插入其中)(Fluhr 等, *Science* 232: 1106-1112, 1986; Ellis 等, *EMBO J.* 6: 11-16, 1987; Strittmatter 和 Chua, *Proc. Nat. Acad. Sci. USA* 84: 8986-8990, 1987; Poulsen 和 Chua, *Mol. Gen. Genet.* 214: 16-23, 1988; Comai 等, *Plant Mol. Biol.* 15: 373-381, 1991)。或者, 可以在无活性的截短的启动子 5'上游区加上异源调节序列, 所述无活性的截短的启动子例如仅包括核心 TATA、有时还包括 CCAAT 元件的启动子(Fluhr 等, *Science* 232: 1106-1112, 1986; Strittmatter 和 Chua, *Proc. Nat. Acad. Sci. USA* 84: 8986-8990, 1987; Aryan 等, *Mol. Gen. Genet.* 225: 65-71, 1991)。

启动子通常包含多个独立的“顺式作用转录调节元件”或简称“顺式元件”, 每个顺式元件赋予对基因表达总体控制的不同方面(Strittmatter 和 Chua, *Proc. Nat. Acad. Sci. USA* 84: 8986-8990, 1987; Ellis 等, *EMBO J.* 6: 11-16, 1987; Benfey 等, *EMBO J.* 9: 1677-1684, 1990)。

“顺式元件”结合调节转录的反式作用蛋白因子。一些顺式元件结合不止一种因子，而反式作用转录因子可以与不止一个顺式元件以不同亲和性相互作用(Johnson 和 McKnight, *Ann. Rev. Biochem.* 58: 799-839, 1989)。已经讨论过植物转录因子、对应的顺式元件和它们相互作用的分析，例如参见 Martin, *Curr. Opinions Biotech.* 7: 130-138, 1996; Murai, *Methods in Plant Biochemistry and Molecular Biology*, Dashek 编辑, CRC Press, 1997, 第 397-422 页; 和 *Methods in Plant Molecular Biology*, Maliga 等编辑, Cold Spring Harbor Press, 1995, 第 233-300 页。本发明的启动子序列可以包含赋予或调节基因表达的“顺式元件”。

可以通过多种技术鉴定顺式元件，包括：缺失分析，即从启动子的 5' 末端或内部缺失一个或多个核苷酸；使用 DNA 酶 I 足迹法的 DNA 结合蛋白分析、甲基化干扰、电泳迁移率变动分析、通过连接介导的 PCR 的体内基因组足迹分析和其它常规分析；或通过常规序列比较方法分析与已知顺式元件基序的序列相似性。可以通过诱变(或取代)顺式元件中的一个或多个核苷酸，或通过其它常规方法，进一步研究所述元件的精细结构(参见例如，*Methods in Plant Biochemistry and Molecular Biology*, Dashek 编辑, CRC Press, 1997, 第 397-422 页; 和 *Methods in Plant Molecular Biology*, Maliga 等编辑, Cold Spring Harbor Press, 1995, 第 233-300 页)。

可以通过化学合成或从包含顺式元件的启动子中克隆，获得所述顺式元件。也可以合成顺式元件，使其具有添加的包含有用限制酶位点的侧翼序列，以利于亚序列操作。在一个实施方案中，所述启动子由多个独立的“顺式元件”组成。在一个优选的实施方案中，使用计算机程序鉴定包含以下核酸序列的“顺式元件”的序列区：SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:25 和 SEQ ID NO:26，其中所述计算机程序包括但不限于专门设计用于鉴定序列内顺式元件或结构域或基序的 MEME 或 SIGNALSCAN。

本发明包括 SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:25 和 SEQ ID NO:26 的片段或顺式元件,或者与本发明的核酸序列具有同源性的已知影响基因调节的顺式元件的同源物。这样的核酸片段可以包括已公开序列的任何区。如 SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:25 和 SEQ ID NO:26 所示的本发明的启动子区或部分启动子区可以包含至少一种调节元件,所述调节元件包括但不限于能够例如在雄性生殖组织中调节有效连接的 DNA 的表达的“顺式元件”或结构域。

植物启动子也可以包括通过操作已知启动子以获得合成、嵌合或杂种启动子而产生的启动子。这样的启动子也可以结合来自一个或多个启动子的顺式元件,例如通过在具有其自身部分或完全调节序列的活性启动子上添加异源调节序列(Ellis 等, EMBO J. 6: 11-16, 1987; Strittmatter 和 Chua, Proc. Nat. Acad. Sci. USA 84: 8986-8990, 1987; Poulsen 和 Chua, Mol. Gen. Genet. 214: 16-23, 1988; Comai 等, Plant. Mol. Biol. 15: 373-381, 1991)。也已经通过无活性的截短启动子的 5'上游区添加异源调节序列,发展出嵌合启动子,所述无活性的截短启动子即仅包括核心 TATA 并任选包括 CCAAT 元件的启动子(Fluhr 等, Science 232: 1106-1112, 1986; Strittmatter 和 Chua, Proc. Nat. Acad. Sci. USA 84: 8986-8990, 1987; Aryan 等, Mol. Gen. Genet. 225: 65-71, 1991)。

依照本发明的嵌合或杂种启动子可以包括至少一种已知的顺式元件,如由各种环境因子如光、热或胁迫调节的元件;由病原体或化学药品等等调节或诱导的元件。根据所述条件,这类元件可以或者正调节或者负调节基因的表达。顺式元件的例子包括但不限于氧效应元件(Cowen 等, J. Biol. Chem. 268(36): 26904, 1993)、光调节元件(参见例如, Bruce 和 Quail, Plant Cell 2: 1081, 1990 和 Bruce 等, EMBO J. 10:

3015, 1991)、响应茉莉酮酸甲酯处理的顺式元件(Beaudoin 和 Rothstein, Plant Mol. Biol. 33: 835, 1997)、水杨酸效应元件(Strange 等, Plant J. 11: 1315, 1997)、热激反应元件(Pelham 等, Trends Genet. 1: 31, 1985)、响应创伤和非生物性胁迫的元件(Loace 等, Proc. Natl. Acad. Sci. U.S.A. 89: 9230, 1992; Mhiri 等, Plant Mol. Biol. 33: 257, 1997)、冷效应元件(Baker 等, Plant Mol. Biol. 24: 701, 1994; Jiang 等, Plant Mol. Biol. 30: 679, 1996; Nordin 等, Plant Mol. Biol. 21: 641, 1993; Zhou 等, J. Biol. Chem. 267: 23515, 1992)以及干旱效应元件(Yamaguchi 等, Plant Cell 6: 251-264, 1994; Wang 等, Plant Mol. Biol. 28: 605, 1995; Bray E.A. Trends in Plant Science 2: 48, 1997)。

在另一实施方案中, 如 SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29 和 SEQ ID NO:30 所示的核苷酸序列包括任何长度能够调节有效连接的 DNA 序列的所述核苷酸序列。例如, 如 SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29 和 SEQ ID NO:30 所公开的序列可以截短或缺失部分序列, 但仍然能够调节有效连接的 DNA 序列的转录。在一个相关实施方案中, 所公开序列的顺式元件可以赋予特定的特异性, 例如赋予有效连接的 DNA 序列在某些组织中的增强表达。因此, SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29 和 SEQ ID NO:30 所公开序列的任何序列片段、部分或区可以用作调节序列, 其中包括但不限于所公开序列的顺式元件或基序。例如, 可以从启动子序列的 5' 或 3' 末端缺失一个或多个碱基对, 以产生“截短的”启动子。也可以在启动子序列内部插入、缺失或置换一个或多个碱基

对。可以通过例如将启动子片段置于最小启动子的上游而构建启动子，以便所述启动子片段或元件有效连接。最小启动子或基础启动子是能够募集并结合基本转录机器的 DNA 片段。真核细胞中基本转录机器的一个例子是 RNA 聚合酶 II 复合物及其辅助蛋白。所述基本转录机器的酶促成份能够利用最小或基础启动子，起始和延伸特定基因的转录。也就是说，在启动子区没有添加的能够募集和结合调节转录(如增强转录、阻抑转录、使转录成为激素依赖型等等)的转录因子的顺式作用序列。可以结合取代、缺失、插入或它们的任何组合以产生最终构建物。

可以修饰本发明的启动子序列，例如以在其它植物系统内表达。在另一方法中，可以通过多种方法设计或工程化新型杂种启动子。许多启动子包含激活、增强或确定所述启动子强度和/或特异性的上游序列(Atchison, *Ann. Rev. Cell Biol.* 4: 127, 1988)。例如，T-DNA 基因包含确定转录起始位点的“TATA”框以及其它位于转录起始位点上游调整转录水平的上游元件(Gelvin, *Transgenic Plants*, Kung, S.-D. 和 Us, R. 编辑, San Diego: Academic Press, 第 49-87 页, 1988)。另一种嵌合启动子将章鱼氨酸合酶(*ocs*)激活剂的三聚体结合到甘露氨酸合酶(*mas*)激活子加启动子上，据报道该启动子增加报道基因的表达(Min Ni 等, *The Plant Journal* 7: 661, 1995)。本发明的上游调节序列可以用于构建这样的嵌合或杂种启动子。构建本发明变异型启动子的方法包括但不限于组合不同启动子的控制元件或重复启动子的部分或区(参见例如美国专利 5,110,732 和美国专利 5,097,025)。本领域内的技术人员熟悉构建、操作和分离大分子(如 DNA 分子、质粒等等)、产生重组生物、筛选和分离基因的特定条件和程序(参见例如 Sambrook 等, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, 1989; Mailga 等, *Methods in Plant Molecular Biology*, Cold Spring Harbor Press, 1995; Birren 等, *Genome Analysis*: 第 1 卷, 分析 DNA, (1997), 第 2 卷, 检测基因, (1998), 第 3 卷, 克隆系统, (1999), 第 4 卷, 基因组作图, (1999),

Cold Spring Harbor, New York).

本发明还包括设计、构建和应用包含 SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:25 和 SEQ ID NO:26 中一种或多种顺式元件的嵌合或杂种启动子以调整或调节有效连接的核酸序列的表达。

SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29 和 SEQ ID NO:30 的启动子序列、片段、区或其顺式元件能够在多种组织中转录有效连接的 DNA 序列，因此能够在多种组织中选择性调节基因表达。

对于许多农学性状，需要在多种组织中转录一个或多个目的基因以赋予所需的特征。需要获得调节有效连接的基因在选定靶组织中转录的合适启动子，因为可能不希望在所有组织中表达基因，而是仅在某些组织中表达基因。例如，假如人们希望选择性表达靶基因以表达除草剂耐性基因，人们可能希望在营养性组织和生殖组织中表达所述除草剂抗性基因。本发明的启动子序列可以用于在多种组织中调节基因表达，所述组织包括但不限于快速生长的分生组织、雄性生殖组织(雄蕊群)如花粉、花药和花丝、雌性生殖组织(雌蕊群)如柱头、花柱和子房、叶、萼片和花瓣。因此，本发明的启动子可用于表达除草剂耐性基因，例如当在植物发育的多种组织和阶段需要耐性时表达。本发明的启动子序列可用于调节任何靶基因的转录，所述靶基因包括但不限于控制下列性状的基因：育性、产量、耐虫性、真菌耐性、除草剂耐性或任何所需性状。尤其优选的基因包括除草剂耐性基因或耐虫性基因。

在一个实施方案中，当需要在多种组织中表达除草剂耐性基因时，本发明的启动子尤其可用于调节除草剂耐性基因的表达。例如，所述除草剂耐性基因可以赋予对除草剂草甘膦的耐性。合适的草甘膦

耐性基因的例子包括但不限于草甘膦抗性 EPSP 合酶基因或降解草甘膦的基因产物如草甘膦氧化还原酶和膦酸 N-乙酰转移酶。对于任何植物生物工程策略而言，获得 5'调节元件的广泛选择是重要的，以便获得对于所需的表达分布型最为有效的合适调节元件。

在另一实施方案中，本发明的启动子可用于确定基因功能。许多基因的功能是未知的，而本发明的启动子可以用作构建物中的遗传元件，以允许对以有义或反义取向表达的一个或多个基因进行表型评价。当需要在组成组织和生殖组织进行高水平基因表达时，本发明的启动子可以成为开发用于高通量测定的植物表达构建物的成分。

可以选择任何植物来鉴定基因和调节序列。用于分离基因和调节序列的合适植物靶的例子包括但不限于苜蓿、苹果、杏、拟南芥属、朝鲜蓟、芝麻菜、石刁柏、鳄梨、香蕉、大麦、豆类、甜菜、黑莓、越桔、嫩茎花椰菜、抱子甘蓝、卷心菜、canola、网纹甜瓜、胡萝卜、木薯、蓖麻籽、花椰菜、芹菜、樱桃、菊苣、芫荽叶、柑桔、克里曼丁、三叶草、椰子、咖啡、玉米、棉花、酸果蔓、黄瓜、花旗松、茄子、苣荬菜、宽叶苦苣、桉树、茴香、无花果、大蒜、葫芦、葡萄、葡萄柚、蜜瓜、豆薯、猕猴桃、莴苣、韭葱、柠檬、酸橙、火炬松、亚麻籽、芒果、甜瓜、蘑菇、油桃、坚果、燕麦、油棕、油料种子油菜(oil seed rape)、秋葵、洋葱、橙、观赏植物、棕榈、番木瓜、欧芹、欧洲防风、豌豆、桃、花生、梨、胡椒、柿、松、菠萝、车前、李、石榴、杨、马铃薯、南瓜、温柏、radiata pine、radicchio、萝卜、油菜籽、树莓、水稻、黑麦、高粱、南方松、大豆、菠菜、笋瓜、草莓、甜菜、甘蔗、向日葵、甘薯、枫香、红桔、茶树、烟草、番茄、小黑麦、草皮、芫菁、藤、西瓜、小麦、薯蓣和西葫芦。对于鉴定调节序列尤其优选的植物是拟南芥属、玉米、小麦、大豆和棉花。

本发明的启动子序列分离自拟南芥(*Arabidopsis thaliana*)植物 DNA。在一个优选的实施方案中，构建物包括有效连接可转录序列的本发明启动子序列，以及合适的终止子和调节元件。可以将这样的构

建物转化进合适的靶植物。可以使用任何植物作为包含本发明启动子序列的核酸构建物的合适宿主。合适靶植物的例子包括但不限于苜蓿、嫩茎花椰菜、卷心菜、canola、花椰菜、玉米、棉花、酸果蔓、黄瓜、苦苣、豌豆、杨、松、马铃薯、洋葱、水稻、树莓、大豆、甘蔗、甜菜、向日葵、番茄和小麦。

启动子分离和修饰方法

可以使用多种方法分离本文公开的启动子序列的片段。可以采用公众可得到的序列信息，使用基于 PCR 的方法从植物基因组文库扩增侧翼区。本领域内的技术人员已知多种扩增与已知序列核心区邻近的未知 DNA 序列的方法。方法包括但不限于反向 PCR (IPCR)、载体盒 (vectorette) PCR、Y 型(Y-shaped) PCR 和基因组步查方法。对于本发明，通过可得的序列信息通过设计 PCR 引物从拟南芥属分离所述核酸分子。

也可以通过其它技术获得核酸片段，例如通过化学方法直接合成片段，正如通常使用自动寡核苷酸合成仪所进行的实践。也可以通过应用核酸复制技术获得片段，例如使用分子生物学领域技术人员众所周知的重组 DNA 技术的 PCR (聚合酶链式反应)技术。对于使用特定扩增引物对扩增靶核酸序列(如通过 PCR)，“严格 PCR 条件”指这样的条件：允许所述引物对仅与具有对应野生型序列(或其互补物)的引物将会结合的靶核酸序列杂交，并且最好产生单一扩增产物。

本领域内的技术人员知道制备植物基因组 DNA 的方法。在一种方法中，可以用限制酶部分消化，然后根据大小选择处于特定大小范围内的 DNA 片段，可以从选定物种制备基因组 DNA 文库。可以将所述基因组 DNA 克隆进合适的构建物，所述构建物包括但不限于噬菌体，使用从许多销售商(参见例如 Stratagene, La Jolla CA 或 Gibco BRL, Gaithersburg, MD)获得的合适克隆试剂盒用合适的构建物如噬菌体制备。

在另一实施方案中，可以修饰本文公开的启动子的核苷酸序列。本领域内的技术人员可以创造在核苷酸序列上有改变的 DNA 分子。可以修饰或改变如 SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29 和 SEQ ID NO:30 所示的本发明的核苷酸序列，以增强它们的控制特性。例如，可以通过在基于 PCR 的 DNA 修饰方法中插入、缺失或取代模板序列而修饰所述序列。“变异型”DNA 分子是包含如下改变的 DNA 分子：其中天然序列的一个或多个核苷酸发生了缺失、添加和/或置换，但最好同时基本保留启动子功能。在启动子片段的情况下，“变异型”DNA 可以包括影响其有效连接的最小启动子转录的改变。例如通过标准 DNA 诱变技术或通过化学合成所述变异型 DNA 分子或其部分，可以产生变异型 DNA 分子。

本发明的启动子序列除可用于调节基因表达外，还可用作核酸杂交实验中的探针或引物。本发明的核酸探针和引物可以在严格条件下与靶 DNA 序列杂交。术语“严格杂交条件”定义为：在该条件下探针或引物特异性地与靶序列杂交，而不与非靶序列杂交，该条件可以根据经验确定。术语“严格条件”按照通过特定杂交程序(参见例如 Sambrook 等, 1989, 在 9.52-9.55, Sambrook 等, 1989 在 9.47-9.52, 9.56-9.58; Kanehisa, Nucl. Acids Res. 12: 203-213, 1984; 和 Wetmur 和 Davidson, J. Mol. Biol. 31: 349-370, 1968)使核酸探针与靶核酸的杂交(即与特定目的核酸序列的杂交)，在功能上加以定义。促进 DNA 杂交的合适严格杂交条件是本领域内技术人员已知的，例如，在 6.0 x 氯化钠/柠檬酸钠(SSC)中约 45℃，然后用 2.0 x SSC 在 50℃洗涤，或者所述条件可以在实验室手册中找到，所述实验室手册包括但不限于 Current Protocols in Molecular Biology, John Wiley & Sons, N. Y., 1989, 6.3.1-6.3.6。例如，可以在以下范围内选择洗涤步骤中的盐浓度：从约 2.0 x SSC 在 50℃的低严格性到约 0.2 x SSC 在 50℃的高严格性。此外，

洗涤步骤中的温度可以从室温(约 22℃)的低严格性条件增加到约 65℃的高严格性。可以同时改变温度和盐,或者可以使温度或盐浓度保持恒定,而改变另一变量。对于高严格性,例如,使用 DNA 或 RNA 探针或引物的杂交可以在以下条件下进行: 6 x SSC, 0.5% SDS, 5 x Denhardt's, 100 µg/mL 非特异性 DNA (如超声处理过的鲑精 DNA), 65℃, 用 0.5 x SSC, 0.5% SDS 在 65℃下洗涤。

假如一种核酸分子与另一种核酸分子显示完全互补性,那么说前者是后者的“互补物”。如本文所用,当其中一种分子的每一个核苷酸都与另一种分子的核苷酸互补,就说这两种分子显示“完全互补性”。假如两种分子相互杂交,其稳定性足以允许它们至少在常规的“低严格性”条件下保持相互退火,那么这两种分子就是“最小互补的”。相似地,假如两种分子相互杂交,其稳定性足以允许它们在常规的“高严格性”条件下保持相互退火,那么这两种分子就是“互补的”。人们考虑:假如探针或引物与靶序列结合的特异性是保守的,那么就可以用低严格性条件如低杂交和/或洗涤温度来鉴定具有较低程度序列相似性的相关序列。因此,由于本发明核苷酸序列与互补 DNA 片段序列段选择性形成双链体分子的能力,因此可以使用它们。通过杂交检测 DNA 区段是本领域内技术人员众所周知的,因此根据所设想的应用,人们会希望使用不同的杂交条件以获得探针对靶序列不同程度的选择性,而选择的方法将取决于所希望的结果。常规严格性条件描述于 Sambrook 等, *Molecular Cloning, A Laboratory Manual*, 第二版, Cold Spring Harbor Press, Cold Spring Harbor, New York, 1989 和 Haymes 等, *Nucleic Acid Hybridization, A Practical Approach*, IRL Press, Washington, DC, 1985。

在本发明的一个实施方案中,在其它植物组织的杂交测定中使用核酸序列 SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:25 和 SEQ ID NO:26、或这些序列的片段、区、顺式元件或寡聚物,以鉴

定密切相关或同源的基因和相关的调节序列。这些包括但不限于在任何基体上进行的 DNA 印迹分析或 RNA 印迹分析,所述基体包括但不限于合适制备的植物组织、纤维素、尼龙或组合滤膜、芯片或载玻片。这样的方法是本领域内众所周知的,并且可以在由经销商提供的试剂盒或制剂中获得。

本文所用的核酸片段是并非全长的核酸的部分。例如,对于本发明,任何长度短于所公布的 SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:25 和 SEQ ID NO:26 的核苷酸序列的核苷酸序列都被认为是片段。片段也可以包括能够在如上文所定义的严格杂交条件下与天然核酸特异性杂交的至少最小长度。这样的最小片段的长度最好是天然核酸序列的至少 8 个核苷酸、更优选 15 个核苷酸、甚至更优选至少 20 个核苷酸、最优选至少 30 个核苷酸。

“探针”是分离的核酸,其上连接有常规的可检测标记或报道分子,例如放射性同位素、配体、化学发光试剂或酶。“引物”是分离的核酸,其通过核酸杂交与互补靶 DNA 链退火,形成引物和靶 DNA 链之间的杂交体,然后由聚合酶如 DNA 聚合酶沿着靶 DNA 链延伸。可以使用引物对扩增核酸序列,如通过聚合酶链式反应(PCR)或其它常规核酸扩增方法。

探针和引物的长度一般为 11 个核苷酸或更长,最好 18 个核苷酸或更长,更优选 25 个核苷酸、最优选 30 个核苷酸或更长。所述探针和引物在高严格性杂交条件下与靶 DNA 或 RNA 序列特异性杂交,在低严格性条件下与另一物种的靶天然序列特异性杂交。依照本发明的探针和引物最好与天然序列有完全序列相似性,虽然探针与天然序列不同并且可以通过常规方法设计保留与靶天然序列杂交的能力。制备和使用探针和引物的方法描述于,例如, *Molecular Cloning: A Laboratory Manual*, 第二版, 第 1-3 卷, Sambrook 等编辑, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989 (下文称为

“Sambrook 等, 1989”); Current Protocols in Molecular Biology, Ausubel 等编辑, Green Publishing and Wiley-Interscience, New York, 1992 (定期更新)(下文称为“Ausubel 等, 1992”); 和 Innis 等, PCR Protocols: A Guide to Methods and Applications, Academic Press: San Diego, 1990。可以从已知序列衍生出 PCR 引物对, 例如通过使用为该目的的计算机程序如 Primer (Version 0.5, © 1991, Whitehead Institute for Biomedical Research, Cambridge, MA)。可以使用基于本文公开的天然启动子序列的引物和探针确认所公开的序列, 必要时使用所述引物和探针通过常规方法如重克隆和重测序来修饰所公开的序列。

构建物和表达构建物

可以将依照本发明的天然或合成的核酸掺入能够引入宿主细胞并在宿主细胞中复制的重组核酸构建物中, 一般是 DNA 构建物。在一个优选的实施方案中, 将如 SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29 和 SEQ ID NO:30 所示的本发明的核苷酸序列或其片段、变形体或衍生物掺入表达盒, 所述表达盒包括与遗传元件例如可选择、可筛选或可评价的标记基因有效连接的本发明的启动子区。

在另一实施方案中, 如 SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29 和 SEQ ID NO:30 所示的本发明所公开的核酸序列有效连接到遗传元件, 如赋予与下列性状有关的所需特征的核酸: 植物形态学、生理学、生长和发育、产量、营养强化、抗病性或有害生物耐性、或环境或化学药品耐性。这些遗传元件如标记基因或目的农学基因可以在转化植物细胞或转化植物的鉴定中起作用, 或产

生具有农学应用的产物。

在另一实施方案中，一种遗传元件产生作为选择装置的产物，该产物在可再生植物组织中起作用，产生赋予所述植物组织对在其它情况下有毒的化合物的抗性的化合物。用作可选择、可筛选或可评价标记的目的基因包括但不限于 GUS (β -葡糖醛酸糖苷酶的编码序列)、GFP (绿荧光蛋白的编码序列)、LUX (萤光素酶的编码基因)、抗生素抗性标记基因或除草剂耐性基因。转座子和相关抗生素抗性基因的例子包括转座子 Tns (bla)、Tn5 (npt II)、Tn7 (dhfr)、青霉素、卡那霉素(和新霉素、G418、博来霉素); 氨甲蝶呤(和三甲氧苄二氨嘧啶); 氯霉素; 卡那霉素和四环素。

在一份关于微生物应用的报道中已经概述了对于植物选择标记有用的特征(Advisory Committee on Novel Foods and Processes, 1994 年 7 月)。这些包括产生最小数目未转化组织的严格选择、不明显干扰再生的大量独立转化事件、应用于大量物种、以及评价组织中所述标记存在的测定的可用性。

许多选择标记基因是本领域内已知的，几种抗生素抗性标记满足这些标准，包括卡那霉素抗性标记(npt II)、潮霉素 B 抗性标记(aph IV)和庆大霉素抗性标记(aac3 和 aacC4)。有用的显性选择标记基因包括编码抗生素抗性的基因(如潮霉素抗性、卡那霉素抗性、博来霉素抗性、G418 抗性、链霉素抗性或壮观霉素抗性); 和除草剂抗性基因(如膦丝菌素乙酰转移酶)。选择除草剂抗性转化子的有用策略描述于，例如，Vasil, Cell Culture and Somatic Cell Genetics of Plants, 第 I - III 卷, Laboratory Procedures and Their Applications Academic Press, New York, 1984。尤其优选用于本发明中的选择标记基因是赋予化合物抗性的基因，所述化合物如抗生素如卡那霉素，以及除草剂如草甘膦(Della-Cioppa 等, Bio/Technology 5(6), 1987, 美国专利 5,463,175, 美国专利 5,633,435)。也可以实施其它选择装置，这些选择装置也处于本发明的范围内。

为实施本发明, 使用制备和应用 DNA 构建物和宿主细胞的常规组合物和方法, 如特别是在 Sambrook 等, 1989 中讨论的。在一个优选的实施方案中, 所述宿主细胞是植物细胞。许多适于转染植物细胞或适于建立转基因植物的 DNA 构建物已经在如 Pouwels 等, Cloning Vectors: A Laboratory Manual, 1985, 增刊 1987); Weissbach 和 Weissbach, Methods for Plant Molecular Biology, Academic Press, 1989; Gelvin 等, Plant Molecular Biology Manual, Kluwer Academic Publishers, 1990; 和 R.R.D. Croy Plant Molecular Biology LabFax, BIOS Scientific Publishers, 1993 中描述。植物表达构建物可以包括, 例如, 处于 5' 和 3' 调节序列转录控制下的一个或多个克隆植物基因。它们也可以包括如所述选择含有所述表达构建物的宿主细胞的选择标记。这样的植物表达构建物还包含一个启动子调节区(如控制诱导型或组成型表达、环境调节或发育调节表达、细胞特异性或组织特异性表达的调节区)、一个转录起始位点、一个核糖体结合位点、一个 RNA 加工信号、一个转录终止位点和一个聚腺苷酸化信号。也包括其它细菌起源的序列以允许所述构建物在细菌宿主中克隆。所述构建物一般还将包含广宿主范围的原核生物复制起点。在一个尤其优选的实施方案中, 所述宿主细胞是植物细胞, 所述植物表达构建物包含: 一个如 SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29 和 SEQ ID NO:30 所公开的启动子区; 一个有效连接的可转录序列; 和一个转录终止序列。设想作为表达载体中遗传成分的其他调节序列包括但不限于可以与所述启动子偶联的非翻译前导序列。在一个尤其优选的实施方案中, 所述宿主细胞是植物细胞, 所述植物表达构建物包含: 一个如 SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29 和 SEQ ID NO:30 所公开的启

动子区；一个有效连接的可转录序列；和一个转录终止序列。植物表达构建物也可以包含其它序列，其中包括但不限于包含可用于克隆目的的限制酶位点的多接头序列。

植物表达构建物中的遗传元件

植物表达构建物可以包含不止一个各自与不同启动子有效连接的可表达基因序列。许多启动子可用于任何目的基因的植物基因表达，所述目的基因包括但不限于选择标记、评价标记、有害生物耐性基因、抗病性基因、营养强化基因和任何其它有农学意义的基因。可用于植物基因表达的组成型启动子的例子包括但不限于：花椰菜花叶病毒(CaMV) P-35S 启动子，该启动子赋予在大多数植物组织，包括单子叶植物(参见，例如，Dekeyser 等, *Plant Cell* 2: 591, 1990; Terada 和 Shimamoto, *Mol. Gen. Genet.* 220: 389, 1990)中的组成型、高水平表达(参见，例如，Odel 等, *Nature* 313: 810, 1985); CaMV 35S 启动子的串联重复形式、增强的 35S 启动子(P-e35S)、胭脂氨酸合酶启动子(An 等, *Plant Physiol.* 88: 547, 1988)、章鱼氨酸合酶启动子(Fromm 等, *Plant Cell* 1: 977, 1989); 如美国专利第 5,378,619 号所述的玄参花叶病毒(P-FMV)启动子和 FMV 启动子的增强形式(P-eFMV)(其中 P-FMV 的启动子序列串联重复)、花椰菜花叶病毒 19S 启动子、甘蔗杆状病毒启动子、鸭跖草黄斑病毒启动子以及其它已知在植物细胞中表达的植物 DNA 病毒启动子。

多种对环境信号、激素信号、化学信号和/或发育信号应答的植物基因启动子可用于在植物细胞中表达有效连接的基因，所述启动子包括由下列信号调节的启动子：(1)热(Callis 等, *Plant Physiol.* 88: 965, 1988)，(2)光(如豌豆 *rbcS*-3A 启动子，Kuhlemeier 等, *Plant Cell* 1: 471, 1989; 玉米 *rbcS* 启动子，Schaffner 和 Sheen, *Plant Cell* 3: 997, 1991; 或叶绿素 a/b 结合蛋白启动子，Simpson 等, *EMBO J.* 4: 2723, 1985)，(3)激素如脱落酸(Marcotte 等, *Plant Cell* 1:969, 1989)，(4)创伤(如 *wun I* ,

Siebertz 等, Plant Cell 1: 961, 1989); 或(5)化学药品如茉莉酮酸甲酯、水杨酸或安全剂。使用(6)器官特异性启动子(如 Roshal 等, EMBO J. 6: 1155, 1987; Schernthaner 等, EMBO J. 7: 1249, 1988; Bustos 等, Plant Cell 1: 839, 1989)也是有利的。

本发明的启动子是能够在快速生长的分生组织和生殖组织中转录有效连接的 DNA 序列、并且可以在表达构建物中有效连接任何目的基因的植物启动子。

植物表达构建物可以包括位于转基因中多肽编码序列上游或下游的 RNA 加工信号如内含子。此外,所述表达构建物可以包括来自植物基因 3'非翻译区的其它调节序列(Thornburg 等, Proc. Natl. Acad. Sci. USA 84: 744 (1987); An 等, Plant Cell 1: 115 (1989)),例如包括 3'终止子区以增加所述 mRNA 的 mRNA 稳定性,例如马铃薯的 PI-II 终止子区或章鱼氨酸或胭脂氨酸合酶 3'终止子区。mRNA 的 5'非翻译区在翻译起始中起重要作用,也可以成为植物表达构建物中的遗传成分。例如,已经证明来自热激蛋白基因的非翻译 5'前导序列增强植物中的基因表达(参见,例如,美国专利 5,362,865)。这些其它的上游和下游调节序列可以来自对于所述表达构建物中存在的其它元件为天然或异源的来源。

用本发明的启动子序列控制植物细胞中的基因表达。所公开的启动子序列是构成用于植物转化的构建物的一部分的遗传成分。本发明的启动子序列可以与任何合适的植物转化质粒或构建物一起使用,所述质粒或构建物包含如所述的可选择或可筛选标记以及相关调节元件,以及一种或多种以足以赋予特定所需性状的方式表达的核酸。本发明设想的有农学意义的合适结构基因的例子包括但不限于:一种或多种耐虫性基因如苏云金芽孢杆菌(*Bacillus thuringiensis*) (*B.t.*)基因、有害生物耐性基因如防制真菌病的基因、除草剂耐性基因如赋予草甘膦耐性的基因、改善品质的基因,所述品质指例如产量、营养强化、环境或胁迫耐性、或任何在植物生理学、生长、发育、形态学或植物

产品上的所需改变。例如，结构基因将包括任何赋予耐虫性的基因，包括但不限于如下列文献所述的芽孢杆菌属昆虫防制蛋白基因：WO 9931248，特此通过引用完整地结合到本文中，美国专利第 5,689,052 号，特此通过引用完整地结合到本文中，美国专利第 5,500,365 号和第 5,880,275 号，特此通过引用完整地结合到本文中。在另一实施方案中，所述结构基因可以赋予对除草剂草甘膦的耐性，赋予所述耐性的基因包括但不限于农杆菌属(*Agrobacterium*)菌株 CP4 草甘膦抗性 EPSPS 基因(*aroA*:CP4)，如美国专利第 5,633,435 所述，特此通过引用完整地结合到本文中，或草甘膦氧化还原酶基因(GOX)，如美国专利第 5,463,175 所述，特此通过引用完整地结合到本文中。

或者，所述 DNA 编码序列可以通过编码非翻译 RNA 分子，例如通过反义介导的机制或共阻抑介导的机制引起内源基因表达的定向抑制，从而影响这些表型(参见，例如，Bird 等, *Biotech. Gen. Engin. Rev.* 9: 207, 1991)。所述 RNA 也可以是工程化以切割所需内源 mRNA 产物的催化性 RNA 分子(即核酶)(参见例如，Gibson 和 Shillito, *Mol. Biotech.* 7: 125, 1997)。因此，产生表达目的表型改变或形态学改变的蛋白或 mRNA 的任何基因可用于实施本发明。

除位于上游(5')或 DNA 序列内的调节元件或序列外，还有下游(3')序列影响基因表达。因此，本文所用术语调节序列指与细胞的蛋白质合成装置联合作用，控制、介导或影响基因产物表达的位于 DNA 序列上游、中间或下游的任何核苷酸序列。

本领域内的技术人员知道适于植物转化的构建物。最好使用所述的可筛选或可评价标记，将本发明的启动子序列掺入表达构建物中，并在瞬时分析中进行测试以在转化植物中提供基因表达的指标。在瞬时测定中检测基因表达的方法是本领域内技术人员已知的。已经报道了使用各种植物、组织和 DNA 传递系统瞬时表达标记基因。例如，瞬时分析的类型可以包括但不限于在任何瞬时植物测定中使用任何目的植物物种，通过对组织的电穿孔或粒子轰击直接传递基因。这样的瞬

时系统将包括但不限于来自小麦悬浮培养物的原生质体(Zhou 等, Plant Cell Reports 12: 612, 1993)、小麦叶原生质体的电穿孔(Sethi 等, J. Crop Sci. 52: 152, 1983); 由玉米组织制备的原生质体的电穿孔(Sheen, J. The Plant Cell 3:225, 1991)、或对特定目的组织的粒子轰击。本发明包括应用任何瞬时表达系统来评价与选定报道基因、标记基因或目的农学基因有效连接的调节序列。设想在瞬时测定中通过合适传递系统测试的植物组织的例子将包括但不限于叶基组织、愈伤组织、子叶、根、胚乳、胚、花组织、花粉和表皮组织。

在瞬时测定中可以使用任何可评价或可筛选标记。对本发明启动子或 5'调节序列进行瞬时分析优选的标记基因包括 β -葡萄糖醛酸糖苷酶(GUS)基因或绿荧光蛋白(GFP)基因。将包含有效连接标记基因的 5'调节序列的表达构建物传递到组织中, 然后根据所述标记, 通过合适的机制分析所述组织。使用定量或定性分析作为工具, 评价所述 5'调节序列当有效连接目的农学基因时在稳定植物中的可能表达分布型。最后, 将本发明的 5'调节序列直接掺入合适的植物转化表达构建物, 所述构建物包含有效连接可转录目的 DNA 序列的 5'调节序列, 将所述构建物转化进植物, 分析稳定转化植株及其后代的所述 5'调节序列赋予的所需表达分布。

将外源 DNA 引入植物细胞的合适表达构建物包括但不限于用于农杆菌属介导的方法中的无毒性(disarmed) Ti 质粒。这些构建物可以包含一个抗性标记、1-2 个 T-DNA 边界、大肠杆菌(*E. coli*)和农杆菌属的复制起点以及一个或多个目的基因和相关调节区。本领域内的技术人员知道在农杆菌介导的方法中, 可以利用许多菌株和方法。这样的菌株包括但不限于农杆菌属菌株 C58、LBA4404、EHA101 和 EHA105。尤其优选的菌株是根癌农杆菌(*Agrobacterium tumefaciens*)菌株。

可以通过本发明方法引入的核酸的例子包括, 例如, 来自另一物种的 DNA 序列或基因, 或甚至起源于同一物种或在同一物种中存在、但通过并非传统繁殖或育种技术的遗传工程方法掺入受体细胞的基因

或序列。然而，术语“外源的”也用于指不在受转化的细胞中正常存在、或者可能仅仅不以所述转化 DNA 区段或基因中出现的形式、结构存在的基因，或者指正常存在、但人们希望以不同于天然表达模式的方式(如过量表达)表达的基因。因此，术语“外源的”基因或 DNA 用于指任何引入受体细胞的基因或 DNA 区段，而不论所述细胞中是否已经存在相似基因。外源 DNA 中包括的 DNA 类型包括：已经在所述植物细胞中存在的 DNA、来自其它植物的 DNA、来自不同生物的 DNA、或外部产生的 DNA 如包含基因反义信息的 DNA 序列或编码基因的合成或修饰形式的 DNA 序列。

可以通过任何植物转化方法将包含本发明启动子序列的植物转化构建物引入植物。可以利用几种方法将 DNA 序列引入植物细胞，这些方法是本领域内众所周知的。合适的方法包括但不限于细菌感染(如用上文所述的农杆菌属)、二元细菌人工染色体构建物、直接传递 DNA (如通过 PEG 介导的转化、干燥/抑制介导的 DNA 摄入、电穿孔、与碳化硅纤维一起搅拌)以及加速 DNA 包被的粒子(在 Potrykus, Ann. Rev. Plant Physiol. Plant Mol. Biol., 42: 205, 1991 中综述)。

特异性转化双子叶植物的方法主要使用根癌农杆菌。例如，已经报道的转基因植物包括但不限于棉花(美国专利第 5,004,863 号；美国专利第 5,159,135 号；美国专利第 5,518,908 号，WO97/43430)、大豆(美国专利第 5,569,834 号；美国专利第 5,416,011 号；McCabe 等, Bio/Technology, 6: 923, 1988；Christou 等, Plant Physiol., 87: 671, 1988)；芸苔属(*Brassica*)(美国专利第 5,463,174 号)以及花生(Cheng 等, Plant Cell Rep., 15: 653, 1996)和苜蓿(Masoud, S.A.等, Transgen. Res., 5: 313, 1996)。

已经报道了转化单子叶植物的相似方法。已经针对多种作物描述了使用这些方法的转化和植物再生，所述作物包括但不限于石刁柏(*Asparagus officinalis*；Bytebier 等, Proc. Natl. Acad. Sci. U.S.A., 84: 5345, 1987)；大麦(*Hordeum vulgare*；Wan 和 Lemaux, Plant Physiol., 104: 37,

1994); 玉米(*Zea mays*; Rhodes, C.A.等, Science, 240: 204, 1988; Gordon-Kamm 等, Plant Cell, 2: 603, 1990; Fromm 等, Bio/Technology, 8: 833, 1990; Koziel 等, Bio/Technology, 11: 194, 1993); 燕麦(*Avena sativa*; Somers 等, Bio/Technology, 10: 1589, 1992); 鸡脚草(*Dactylis glomerata*; Horn 等, Plant Cell Rep., 7: 469, 1988); 水稻(*Oryza sativa*, 包括 indica 变种和 japonica 变种, Toriyama 等, Bio/Technology, 6: 10, 1988; Zhang 等, Plant Cell Rep., 7: 379, 1988; Luo 和 Wu, Plant Mol. Biol. Rep., 6: 165, 1988; Zhang 和 Wu, Theor. Appl. Genet., 76: 835, 1988; Christou 等, Bio/Technology, 9: 957, 1991); 高粱(*Sorghum bicolor*; Casas, A.M.等, Proc. Natl. Acad. Sci. U.S.A., 90: 11212, 1993); 甘蔗(*Saccharum spp.*; Bower 和 Birch, Plant J., 2: 409, 1992); 苇状羊茅(*Festuca arundinacea*; Wang, Z.Y.等, Bio/Technology, 10: 691, 1992); 草皮草(*Agrostis palustris*; Zhong 等, Plant Cell Rep., 13: 1, 1993); 和小麦(*Triticum aestivum*; Vasil 等, Bio/Technology, 10: 667, 1992; Weeks T.等, Plant Physiol., 102: 1077, 1993; Becker 等, Plant, J. 5: 299, 1994)。对于本领域内技术人员明显的是: 可以利用和改变多种转化方法, 从多种目的作物产生稳定的转基因植物。

植物分析方法

分析转化植物中目的基因的存在以及本发明启动子序列赋予的表达水平和/或分布型。本领域内的技术人员知道可用于分析转化植物的多种方法。使用多种方法评价基因表达, 并确定所引入的基因是否整合、是否正常起作用以及是否如所预期地遗传。对于本发明, 可以通过测定启动子有效连接的基因的表达水平, 评价所述启动子。使用报道基因, 通过瞬时测定方法, 可以确定对启动子功能的初步评价, 但从对稳定植物的分析可以获得更确定性的启动子评价。植物分析方法包括但不限于 DNA 印迹分析或 RNA 印迹分析、基于 PCR 的方法、生物化学分析、表型筛选方法、田间评价和免疫诊断学测定。

本发明的方法包括但不限于 PCR 技术、分离基因组 DNA、构建

表达构建物、瞬时测定和植物转化方法，这些方法是本领域内技术人员众所周知的，并使用标准技术或其修改形式执行。

草甘膦喷洒试验

在一个实施方案中，进行草甘膦耐性的温室或田间评价。本文用术语“草甘膦”统称母体除草剂 N-膦酰甲基甘氨酸(也称为草甘膦酸)、其盐或酯、或在植物组织中转化为 N-膦酰甲基甘氨酸的化合物、或以其它方式提供离子形式 N-膦酰甲基甘氨酸(也称为草甘膦离子)的化合物。作为例证，Franz 的美国专利第 3,799,758 号和第 4,405,531 号公开了本文使用的水溶性草甘膦盐，所述文献的公开内容通过引用结合到本文中。可以按照本发明使用的草甘膦盐包括但不限于碱金属盐，例如钠盐和钾盐；铵盐；C₁₋₁₆ 烷基铵盐，例如二甲基铵盐和异丙基铵盐；C₁₋₁₆ 链烷醇铵盐，例如单乙醇铵盐；C₁₋₁₆ 烷基铊盐，例如三甲基铊盐；它们的混合物等等。草甘膦酸分子有三个酸位点，这三个酸位点有不同的 pKa 值；因此可以使用一代盐、二代盐和三代盐，或它们的混合物，或任何中间中和水平的盐。

草甘膦盐有商业意义，部分因为它们是水溶性的。许多铵盐、烷基铵盐、链烷醇铵盐、烷基铊盐和碱金属盐是高度水溶性的，允许配制成高度浓缩的水溶液，在使用时用水稀释。

这样的浓缩水溶液可以含有按酸当量(g a.e./l)表示的每升约 50 克到约 500 克草甘膦。优选更高的草甘膦浓度，例如约 300 到约 500 g a.e./l。

或者将草甘膦盐配制成水溶性组合物或水分散性组合物，例如采用粉剂、颗粒剂、丸状或片剂的形式。所述组合物常常称为干制剂，但词语“干”在这里不应当理解为意味着完全不含水。通常干制剂含有少于约 5%(重量)的水，例如约 0.5%到约 2%(重量)的水。这样的制剂需要在使用时溶解或分散于水中。

设想的干草甘膦制剂可以含有按酸当量(% a.e.)表示的约 5%到约

80%(重量)的草甘膦。优选在上述范围内的较高草甘膦浓度,例如约50%到约80% a.e.。用于制造干制剂的尤其有用的草甘膦盐是钠盐和铵盐。

本发明的植物处理组合物和液体以及干浓缩组合物可以任选地含有一种或多种所需的赋形剂成份。对于草甘膦组合物尤其有用的赋形剂成份是表面活性剂,表面活性剂协助水性喷洒溶液保留在相对疏水的植物叶表面,也帮助草甘膦穿透叶的蜡质外层(角质层)并因此接触叶内的活组织。表面活性剂也可以执行其它有用功能。

可以用于本发明草甘膦组合物中的表面活性剂在类型和化学种类上没有限制。在特定情形下,非离子型、阴离子型、阳离子型和两性型或者一种以上所述类型的组合都是有用的。然而,一般优选至少一种存在的表面活性剂(如果有的话)应当是非阴离子型的,即至少一种所述表面活性剂应当是非离子型、阳离子型或两性型的。

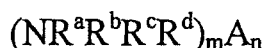
在此有用的许多表面活性剂具有包含一个或多个部分的化学结构,所述部分各由一个 C_{2-4} 烯化氧单位或 C_{2-4} 烯化氧单位聚合链或共聚合链组成。所述表面活性剂称为聚氧化烯表面活性剂,包括非离子型、阴离子型、阳离子型和两性型。在目前设想的组合物中有用的聚氧化烯表面活性剂包含约2到约100个 C_{2-4} 烯化氧单位。在优选的聚氧化烯表面活性剂中,烯化氧单位形成一条或多条环氧乙烷链或环氧乙烷和环氧丙烷共聚物链,烯化氧单位的每条链具有一个氢(hydrido)端基或用 C_{1-4} 烷基或 C_{1-4} 烷酰基封端。

在本发明组合物中有用的表面活性剂的疏水部分可以是基本基于烃的,在这种情况下所述疏水部分一般是 C_{8-24} 、优选 C_{12-18} 的烷基、链烯基、烷芳基、烷酰基或链烯酰基链。这些链可以是线性或分支的。或者,所述疏水部分可以含有硅原子,例如采取硅氧烷基团的形式如七甲基三硅氧烷基团,或者所述疏水部分可以含有氟原子,例如作为部分氟化的烷基链或全氟烃基链。

在非离子型表面活性剂中,尤其优选的种类包括聚氧乙烯烷基、

链烯基或烷芳基醚(例如乙氧基化伯醇或仲醇或烷基酚)、聚氧乙烯烷基或链烯基酯(例如乙氧基化脂肪酸)、聚氧乙烯脱氧山梨糖醇烷基酯、甘油基烷基酯、蔗糖酯、烷基多聚糖苷等等。所述非离子型表面活性剂的代表性具体例子包括聚氧乙烯(9)壬基酚、Shell 的 Neodol™ 25-7 (聚氧乙烯(7)C₁₂₋₁₅ 线性伯醇)、Union Carbide 的 Tergitol™ 15-S-9 (聚氧乙烯(9)C₁₂₋₁₅ 仲醇)、ICI 的 Tween™ 20 (聚氧乙烯(20)脱水山梨糖醇单月桂酸酯)和 Henkel 的 Agrimul™ PG-2069 (C₉₋₁₁ 烷基多聚糖苷)。

在阳离子型表面活性剂中, 尤其优选的种类包括聚氧乙烯叔烷基胺或链烯基胺(如乙氧基化脂肪胺)、季铵表面活性剂、聚氧乙烯烷基醚胺等等。所述阳离子型表面活性剂有代表性的具体例子包括聚氧乙烯(5)椰油胺(cocoamine)、聚氧乙烯(15)牛油脂肪胺、氯化二硬脂酰基二甲基铵、溴化十六烷基三甲基铵、氯化甲基双(2-羟乙基)椰油铵、氯化 N-十二烷基吡啶和聚氧丙烯(8)氯化乙氧基三甲基铵。尤其优选的聚氧乙烯烷基醚胺是 PCT 公开说明书 WO 96/32839 公开的那些化合物。本领域内知道许多阳离子叔铵表面活性剂可以与草甘膦组合使用, 并且可以用于本文所设想的组合物中; 所述叔铵表面活性剂具有式



其中 A 是合适的阴离子如氯离子、溴离子、碘离子、乙酸根、硫酸根或磷酸根, m 和 n 是整数, 使得阳离子(NR^aR^bR^cR^d)的正电荷与阴离子 A 的负电荷相平衡, R^a、R^b、R^c 和 R^d 的选择包括但不限于:

- (i) R^a 是苄基或 C₈₋₂₄, 优选 C₁₂₋₁₈、烷基或链烯基, 而 R^b、R^c 和 R^d 独立地是 C₁₋₄ 烷基, 优选甲基;
- (ii) R^a 和 R^b 独立地是 C₈₋₂₄、优选 C₁₂₋₁₈ 烷基或链烯基, 而 R^c 和 R^d 独立地是 C₁₋₄ 烷基, 优选甲基;
- (iii) R^a 是 C₈₋₂₄、优选 C₁₂₋₁₈ 烷基或链烯基, R^b 是具有约 2 到约 100 个 C₂₋₄ 烯化氧单位(优选环氧乙烷单位)的聚氧化烯链, 而 R^c 和 R^d 独立地是 C₁₋₄ 烷基, 优选甲基;
- (iv) R^a 是 C₈₋₂₄、优选 C₁₂₋₁₈ 烷基或链烯基, R^b 和 R^c 是具有总共

约 2 到约 100 个 C_{2-4} 烯化氧单位(优选环氧乙烷单位)的聚氧化烯链, 而 R^d 是 C_{1-4} 烷基, 优选甲基;

- (v) R^a 是具有约 2 到约 100 个 C_{2-4} 烯化氧单位的聚氧化烯链, 其中主要是 C_{3-4} 烯化氧单位(优选环氧丙烷单位), 而 R^b 、 R^c 和 R^d 独立地是 C_{1-4} 烷基, 优选甲基或乙基。

该类型中尤其优选的季铵表面活性基是那些在 Claude 等的美国专利第 5,464,807 号中公开的化合物。

在一个实施方案中, 与所述季铵表面活性基结合的所述阴离子 A 可以是草甘膦阴离子。

在两性型表面活性剂, 包括本领域内习惯更正确地描述为两性离子型的表面活性剂中, 尤其优选的种类包括聚氧乙烯烷基氧化胺、烷基甜菜碱、烷基取代的氨基酸等等。所述两性型表面活性剂有代表性的例子包括氧化十二烷基二甲基胺、聚氧乙烯(2)氧化椰油胺和硬脂酰基二甲基甜菜碱。

本领域内技术人员可以从标准参考来源中选择不限于上述种类的合适表面活性剂, 所述标准参考来源包括 Gower 出版的 *Handbook of Industrial Surfactants*, 第二版(1997)、MC Publishing Company 出版的 *McCutcheon's Emulsifiers and Detergents*, 北美和国际版(1997)和 Cosmetic, Toiletry and Fragrance Association 出版的 *International Cosmetic Ingredient Dictionary*, 第六版(1995)卷 1 和卷 2。

本发明组合物的其它可选成份包括改变颜色、粘度、胶凝特性、冰点、吸湿性、结块行为、溶解速率、分散性或其它配制特性的试剂。

草甘膦商业制剂的例子包括但不限于: Monsanto Company 销售的 ROUNDUP®、ROUNDUP® ULTRA、ROUNDUP® CT、ROUNDUP® EXTRA、ROUNDUP® BIACTIVE、ROUNDUP® BIOFORCE、RODEO®、POLARIS®、SPARK®和 ACCORD®除草剂, 所有这些除草剂都含有异丙基铵盐形式的草甘膦; Monsanto Company 销售的 ROUNDUP® DRY 和 RIVAL®除草剂, 所述除草剂含有铵盐形式的草

甘膦；Monsanto Company 销售的 ROUNDUP® GEOFORCE 除草剂，所述除草剂含有钠盐形式的草甘膦；Zeneca Limited 销售的 TOUCHDOWN®除草剂，所述除草剂含有三甲基铈盐。

选择有生物活性的草甘膦制剂的施用量在一般农业技术人员知识范围内。本领域内技术人员同样会认识到：在实施本发明的过程中，单株条件、天气条件和生长条件影响所获得的结果。二十多年的草甘膦应用和关于所述应用的已发表的研究已经提供了丰富的信息，杂草防除实施人员可以根据这些信息选择对于特定物种在特定环境条件下在特定生长阶段有效除草的草甘膦施用量。

本发明的一种方法适用于草甘膦作为除草剂或植物生长调节剂对其有生物活性的任何或所有植物物种。这包括世界范围内各种各样的植物物种。同样，本发明的组合物可应用于草甘膦对其有生物活性的任何和所有植物物种。

在一个实施方案中，施用含草甘膦的除草剂于含本发明 DNA 构建物的植株，然后评价所述植株对所述草甘膦除草剂的耐性。可使用任何草甘膦制剂测试含本发明 DNA 构建物的植株。例如，可以使用草甘膦组合物如 Roundup Ultra™。评价植物草甘膦耐性的测试参数将随多种因素而变。因素将包括但不限于：草甘膦制剂的类型、所述制剂中使用的草甘膦的浓度和用量、植物类型、施用期间植物的发育阶段、环境条件、施用方法以及特定制剂的施用次数。例如，可以在温室环境中使用喷洒施用方法测试植物。使用 Roundup Ultra™的测试范围可以包括但不限于 4,000 平方米 226.5 ml 到 4,000 平方米 7248 ml。根据具体作物和植物发育阶段，Roundup Ultra™的优选商业有效范围可以从 4,000 平方米 453 ml 到 4,000 平方米 1812 ml。一种作物可以至少喷洒施用一次草甘膦制剂。为在棉花中测试，在 3 叶期以 4,000 平方米 906 ml 施用一次，然后在后面的发育阶段再施用。对于小麦，可以在 3-5 叶期以 4,000 平方米 906 ml 施用 Roundup Ultra™一次，然后根据所测试的小麦类型，在收获前或收获后施用一次。可以优化每种

作物的测试参数,以便找出赋予所需商业有效草甘膦耐性水平的包含本发明构建物的特定植株。

提供下面的实施例以示范本发明的优选实施方案。本领域内的技术人员应该认识到:在下面实施例中公开的技术代表本发明人发现的在本发明实施中作用良好的技术。然而,本领域内的技术人员根据本公开物应当理解:在不偏离本发明的精神和范围的情况下,对于所公开的具体实施方案可以作出许多改变而仍然获得类似或相似的结果,因此在附图中陈述或显示的所有内容应当被解释为是说明性、而不是限制性的。

实施例

实施例 1

使用的质粒构建物是 pUC 克隆构建物或双边界植物转化构建物,所述构建物包含大肠杆菌复制起点如 *ori322*、广宿主范围复制起点如 *oriV* 或 *oriRi*、和选择标记的编码区如编码赋予壮观霉素或链霉素抗性的 *Tn7* 氨基糖苷腺苷转移酶(*aadA*)的 *Spc/Str* 或庆大霉素(*Gm*, *Gent*)选择标记。为进行植物转化,所述宿主细菌菌株是根癌农杆菌 *ABI* 或 *LBA4404*。

所述遗传元件描述如下: *P-e35S* 是来自 *CaMV* 的 35S RNA,包含重复的-90-300 区,如美国专利第 5,424,200 号所述,该文献特此通过引用完整地结合到本文中; *P-FMV* 是来自玄参花叶病毒的 34S 启动子,如美国专利第 5,378,619 号所述,该文献特此通过引用完整地结合到本文中; *P-eFMV* 是所述 *FMV* 启动子的衍生物,包含重复的 *FMV*; *CTP2* 是拟南芥属 *EPSP* 合酶的转运肽区,如美国专利第 5,633,435 号所述; *aroA:CP4syn* (*aroA:CP4*)是如美国专利第 5,633,435 号所述的 *CP4 EPSP* (合成序列)编码区,或根据特定植物物种中的密码子选择而修饰以在植物中表达; *E9 3'*是豌豆 *RbcS* 基因分离物的作为聚腺苷酸化信号起作用的 3'末端; *nos* 是胭脂氨酸合酶基因的作为聚腺苷酸化信号起作用的 3'末端; *Hsp70* 是来自矮牵牛

(*Petunia hybrida*)的非翻译前导序列,如美国专利第 5,362,865 号所述,该文献特此通过引用完整地结合到本文中;GUS 是大肠杆菌 β -葡萄糖醛酸糖苷酶编码序列(Jefferson, R. A. *Proc. Natl. Acad. Sci. U.S.A.*, 83: 8447-8451, 1987);右边界(RB)和左边界(LB)来自根癌农杆菌章鱼氨酸和胭脂氨酸菌株的 Ti 质粒。P-AtAct2 是来自拟南芥肌动蛋白 2 基因的启动子;AtAct2i 是拟南芥肌动蛋白 2 基因 5'非翻译区(UTR)中的内含子;P-AtAct8 是拟南芥肌动蛋白 8 基因的启动子;AtAct2i 是拟南芥肌动蛋白 8 基因 5' UTR 中的内含子;P-AtAct11 是拟南芥肌动蛋白 11 基因的启动子;AtAct11i 是拟南芥肌动蛋白 11 基因 5'UTR 中的内含子;P-AtAct1a 是拟南芥肌动蛋白 1a 基因的启动子,L-AtAct1a 是所述肌动蛋白 1a 基因的非翻译前导序列,I-AtAct1a 是来自所述肌动蛋白 1a 基因的基因组 DNA 的内含子;P-AtAct1b 是拟南芥肌动蛋白 1b 基因的启动子,L-AtAct1b 是来自所述肌动蛋白 1b 基因的非翻译前导序列,I-AtAct1b 是来自肌动蛋白 1b 基因的基因组 DNA 的内含子;P-AtAct3 是拟南芥肌动蛋白 3 基因的启动子,L-AtAct3 是来自所述肌动蛋白 3 基因的非翻译前导序列,I-AtAct3 是来自所述肌动蛋白 3 基因的基因组 DNA 的内含子;P-AtAct7 是拟南芥肌动蛋白 7 基因的启动子,L-AtAct7 是来自所述肌动蛋白 7 基因的非翻译前导序列,I-AtAct7 是来自所述肌动蛋白 7 基因的基因组 DNA 的内含子;P-AtAct12 是拟南芥肌动蛋白 12 基因的启动子,L-AtAct12 是来自所述肌动蛋白 12 基因的非翻译前导序列,I-AtAct12 是来自所述肌动蛋白 12 基因的基因组 DNA 的内含子;P-AtEF1 α (P-AtEF1 或 EF1 α)是拟南芥延伸因子基因 1 α 的启动子,AtEF1 α -i (AtEF1-i)是拟南芥延伸因子基因 1 α 的 5'UTR 中的内含子。

图 1-18 提供包含一到三个植物表达盒的植物转化构建物的例子。植物分子生物学领域内的技术人员能够不经过分的试验,就可构建包含本发明启动子和遗传元件的植物表达盒的多种组合并在作物中测试。不应当认为在附图中显示的构建物是能够装配的仅有的构建

物, 所述构建物仅仅是作为例子提供给本领域内的技术人员。图 1 (pCGN8086)提供了包含一个表达盒的植物转化构建物的例子, 该表达盒包含与目的基因(CTP2-*aroA*:CP4syn)有效连接的一个本发明的启动子(P-AtAct8)。图 2 (pMON45325)提供了包含两个表达盒的植物转化构建物的例子, 所述表达盒包含与至少一个目的基因(CTP2-*aroA*:CP4syn)有效连接的至少一个本发明的启动子(P-AtAct11)。图 3 (pMON45331)提供了包含一个表达盒的植物转化构建物的例子, 所述表达盒包含与至少一个目的基因(CTP2-*aroA*:CP4syn)有效连接的一个本发明的启动子(P-AtEF1 加内含子)。图 4 (pMON45332)提供了包含两个表达盒的植物转化构建物的例子, 所述表达盒包含与至少一个目的基因(CTP2-*aroA*:CP4syn)有效连接的至少一个本发明的启动子(P-AtEF1 加内含子)。图 5 (pMON9190)提供了包含三个表达盒的植物转化构建物的例子, 在所述表达盒中, 至少两个本发明的启动子(P-AtEF1 加内含子, AtEF1 α -i; P-AtAct2 加内含子, AtAct2i)与至少一个目的基因(CTP2-*aroA*:CP4syn)有效连接, 而 P-eFMV 启动子与 CTP2-*aroA*:CP4syn 有效连接。图 6 (pMON9153)植物表达盒与图 4 (pMON45332)显示的植物表达盒相同, 显示该质粒图谱的目的是鉴定在说明书中显示的数据表中关于植物表型所示数据的表达盒。图 7 (pCGN8099)提供了包含两个表达盒的植物转化构建物的例子, 所述表达盒包含驱动目的基因(*aroA*:CP4syn)转录的本发明杂种启动子 P-FMV-AtEF1 α 和 P-e35S-AtAct8。图 8 (pCGN8088)提供了包含两个表达盒的植物转化构建物的例子, 所述表达盒包含本发明的一个启动子 P-AtAct8 加内含子、AtAct8i 和驱动目的基因(*aroA*:CP4syn)表达的 P-eFMV 启动子。图 9 (pCGN8068)提供了包含两个表达盒的植物转化构建物的例子, 所述表达盒包含本发明的一个启动子 P-AtAct2 加内含子、AtAct2i 和驱动目的基因(*aroA*:CP4syn)表达的 P-eFMV 启动子。图 10 (pCGN8096)提供了包含两个表达盒的植物转化构建物的例子, 所述表达盒包含驱动目的基因(*aroA*:CP4syn)转录的本发明杂种启动子 P-FMV/-AtAct11 和 P-

e35S-AtAct2。图 11 (pCGN9151)提供了包含两个表达盒的植物转化构建物的例子, 所述表达盒包含驱动目的基因(*aroA*:CP4syn)转录的本发明杂种启动子 P-FMV-AtEF1 α 和 P-e35S-AtAct2。图 12 (pMON10156)提供了包含一个表达盒的植物转化构建物的例子, 所述表达盒包含驱动目的基因 *aroA*:CP4syn 表达的 P-eFMV 启动子, 该载体用于与本发明启动子序列进行比较的目的。图 13 (pMON52059)提供了包含一个表达盒的植物转化构建物的例子, 所述表达盒包含驱动目的基因 (*aroA*:CP4syn)表达的杂种启动子(P-FMV-AtEF1 α)。图 14 (pMON54952)提供了包含一个表达盒的植物转化构建物的例子, 所述表达盒包含与至少一个目的基因(CTP2-*aroA*:CP4syn)有效连接的一个本发明启动子(P-AtAct1a 加 AtAct1a 内含子)。图 15 (pMON54953)提供了包含一个表达盒的植物转化构建物的例子, 所述表达盒包含与至少一个目的基因 (CTP2-*aroA*:CP4syn)有效连接的一个本发明启动子(P-AtAct1b 加 AtAct1b 内含子)。图 16 (pMON54954)提供了包含一个表达盒的植物转化构建物的例子, 所述表达盒包含与至少一个目的基因 (CTP2-*aroA*:CP4syn)有效连接的一个本发明启动子(P-AtAct3 加 AtAct3 内含子)。图 17 (pMON54955)提供了包含一个表达盒的植物转化构建物的例子, 所述表达盒包含与至少一个目的基因(CTP2-*aroA*:CP4syn)有效连接的一个本发明启动子(P-AtAct7 加 AtAct7 内含子)。图 18 (pMON54956)提供了包含一个表达盒的植物转化构建物的例子, 所述表达盒包含与至少一个目的基因(CTP2-*aroA*:CP4syn)有效连接的一个本发明启动子(P-AtAct12 加 AtAct12 内含子)。

实施例 2

克隆构建物和 GUS 构建物列于表 1。使用拟南芥 *Landsberg erecta* DNA 作为模板(Rogers 和 Bendich, *Plant Mol. Biol.* 5: 69, 1998), 用 SEQ ID NO:1 (正向引物)和 SEQ ID NO:2 (反向引物)在下面反应中分离拟南芥属肌动蛋白 2 启动子和内含子(Genbank 登记号 U41998, 如 An 等,

Plant J. 10: 107-121, 1996 所述): 0.5 µg 模板 DNA、25 pmole 每种引物、taq 聚合酶(BMB, Indianapolis, IN), 用蜡珠进行“热起始”PCR。PCR 热循环仪条件如下: 94℃1 分钟; 30 个循环的 92℃40 秒, 55℃1 分钟, 72℃1 分钟 30 秒; 5 分钟 72℃延伸。所述 PCR 用 GeneClean II (Bio101 Inc., Vista, CA)纯化, 用 HindIII 和 Nco I 消化, 然后连接进用 HindIII 和 Nco I 消化的构建物 pMON26149 (表 1)。证实启动子克隆的序列, 并将得到的构建物命名为 pMON26170 (表 1)。

表 1. 包含拟南芥属肌动蛋白和 EF1 启动子序列的克隆构建物和 GUS 构建物

<u>构建物</u>	<u>描述</u>	<u>启动子*/基因/3'</u>
pMON26149	克隆构建物	
pMON26170	植物表达构建物	Act2/GUS/nos
pMON26171	植物表达构建物	Act8/GUS/nos
pMON8677	克隆构建物	
pMON48407	植物表达构建物	Act11/GUS/nos
pMON26152	克隆构建物	
pMON26177	植物表达构建物	EF1/GUS/nos
pMON11750	植物表达构建物	e35S/GUS/nos
pMON15737	植物表达构建物	FMV/GUS/nos

* 所述肌动蛋白和延伸因子启动子序列也包含来自对应基因的 5' UTR 的内含子序列。

实施例 3

使用拟南芥 *Landsberg erecta* DNA 作为模板, 应用实施例 2 所述的 PCR 条件和纯化方法, 用引物 SEQ ID NO:3 (正向引物)和 SEQ ID NO:4 (反向引物)分离拟南芥属肌动蛋白 8 启动子和内含子(Genbank 登记号 U42007, 如 An 等, *Plant J.* 10: 107-121, 1996 所述)。如实施例 2

所述用限制酶克隆所述启动子，证实所述启动子的序列，并将获得的构建物命名为 pMON26171 (表 1)。

实施例 4

使用拟南芥 *Landsberg erecta* DNA 作为模板，应用实施例 2 所述的 PCR 条件和纯化方法，用引物 SEQ ID NO:5 (正向引物)和 SEQ ID NO:6 (反向引物)分离拟南芥属肌动蛋白 11 启动子和内含子(Genbank 登记号 U27981, 如 Huang 等, *Plant Mol. Biol.*, 33: 125-139, 1997 所述)。用限制酶 EcoRV 和 Nco I 克隆所述启动子并将其连接进 pMON8677 (表 1), 证实所述启动子的序列，并将获得的构建物命名为 pMON48407 (表 1)。

实施例 5

使用拟南芥 *Landsberg erecta* DNA 作为模板，应用实施例 2 所述的 PCR 条件和纯化方法，用引物 SEQ ID NO:7 (正向引物)和 SEQ ID NO:8 (反向引物)分离拟南芥属延伸因子 1 α (AtEF1 α)启动子和内含子(Genbank 登记号 X16430, 如 Axelos 等, *Mol. Gen. Genet.* 219: 106-112, 1989; Curie 等, *NAR* 19: 1305-1310; Curie 等, *Plant Mol. Biol.* 18: 1083-1089, 1992; Curie 等, *Mol. Gen. Genet.* 238: 428-436, 1993 所述)。如实施例 2 所述用限制酶 HindIII 和 Nco I 克隆所述启动子并将其连接进 pMON26152 (表 1), 证实所述启动子的序列，并将获得的构建物命名为 pMON26177 (表 1)。

实施例 6

将所述植物转化构建物通过交配转移进农杆菌属。基本如 WO/0036911 所述进行棉花转化，该文献特此通过引用完整地结合到本文中。如 Ye 等, *Plant Journal* 19: 249-257, 1999 所述进行拟南芥属转化。如美国专利第 5,565,347 号所述进行番茄转化，该文献特此通过引

用完整地结合到本文中。

实施例 7

通过合适的传递系统例如农杆菌介导的转化方法将 DNA 构建物转化进靶作物(参见例如美国专利第 5,569,834 号,该文献特此通过引用完整地结合到本文中,美国专利第 5,416,011 号,该文献特此通过引用完整地结合到本文中,美国专利第 5,631,152 号,该文献特此通过引用完整地结合到本文中,美国专利第 5,159,135 号,该文献特此通过引用完整地结合到本文中,美国专利第 5,004,863 号,该文献特此通过引用完整地结合到本文中,以及美国临时申请第 60/111795 号,该文献特此通过引用完整地结合到本文中)。或者,可以使用粒子轰击方法(参见例如专利申请 WO 92/15675、WO 97/48814 和欧洲专利申请 586,355 和美国专利第 5,120,657 号、美国专利第 5,503,998 号、美国专利第 5,830,728 号、美国专利第 5,015,580 号,所有这些文献都通过引用完整地结合到本文中)。

可以利用大量转化和再生系统,并且这些转化和再生系统是本领域内技术人员众所周知的。然后通过多种本领域内技术人员已知的分子、免疫诊断学、生物化学和/或田间评价方法,分析稳定转化植株和后代中所述基因在目标组织中的表达,所述评价方法包括但不限于在生长室或田间环境中以商业化有效的浓度使用草甘膦制剂进行喷洒测试。

实施例 8

通过本领域内技术人员已知的常规方法进行 GUS 测定(参见,例如,Jefferson 等, *EMBO J.* 6: 3901, 1987)。对于棉花,测试 R0 植株。在发育不同阶段根据大小选择组织,取样,汇集样品以进行分析。收获棉花花序芽,取雄性生殖组织样品(花药和花丝)、雌性生殖组织样品(整个柱头、花柱和子房)和花冠(萼片和花瓣)。为进行大小选择,选

择每个阶段的三个花序芽, 所述花序芽包括小(短于 0.5 cm)、中(从 0.5-0.7 cm)和大(candle stage 或开花)的几种大小。将棉花植株放置在温室中约 1-2 周后收集叶样品, 并在约 1-2 月后收集其它样品。不收集初花(头五个结果位置保持完整)。

对于拟南芥属, 分析 V1 植株, 并仅测试纯合分离体和杂合分离体。每种构建物分析八到十个事件(每事件 5 个植株)。拟南芥属的 GUS 结果代表 8-10 个事件的汇集样品。所公开的表(表 2 和表 3)中的值代表对于所指定组织的平均 GUS 表达(pmol/MU/分钟/mg)。

实施例 9

分析植株在叶组织和生殖组织中的 GUS 表达, 所述生殖组织包括未成熟的花序芽和花。结果显示于表 2。测试的构建物包括 pMON48407 (P-AtAct11 + 内含子/GUS/nos)、pMON26170 (P-AtAct2 + 内含子/GUS/nos)、pMON26171 (P-AtAct8 + 内含子/GUS/nos)、pMON11750 (e35S/GUS/nos)、pMON26177 (P-EF1 α + 内含子/GUS/nos) 和 pMON15737 (P-FMV/GUS/nos)。所述肌动蛋白和延伸因子启动子在多种组织包括生殖组织中赋予高水平 GUS 表达。

表 2. 平均拟南芥属 V1 GUS 表达

构建物	叶	未成熟的花序芽	花	雌蕊群	雄蕊群
pMON48407	6944	7394	8359	ND	ND
pMON26170	45238	74099	54502	73623	217292
pMON26171	29343	35884	37125	76311	207100
pMON11750	60844	14032	16263	35882	115049
pMON26177	47598	72871	96420	191066	507370
pMON15737	28314	57903	84457	44696	87876

实施例 10

测试 R0 棉花植株选定组织不同发育阶段的 GUS 报道基因表达。

根据大小对花序芽划分阶段(小、中和大; 大=candle 和开花)。雄蕊群代表雄性生殖组织。雌蕊群代表包括整个花托(柱头、花柱和子房)的雌性生殖组织。花冠样品包括萼片和花瓣。如实施例 8 所述制备所述组织并进行 GUS 测试。结果概述于表 3。测试的构建物包括 pMON48407 (P-EF1 α + 内含子/gus/nos)、pMON26170 (P-AtAct2 + 内含子/gus/nos)和 pMON48407 (P-AtAct11 + 内含子/gus/nos)。

对于 pMON26177 测试六株植株并获得平均 GUS 值。对于 pMON26170 测试二十株植株并获得平均 GUS 值。对于 pMON48407 测试八株植株并获得平均 GUS 值。结果证明: 所述肌动蛋白和延伸因子启动子可用于尤其在生殖组织中有效表达有效连接的基因。

表 3. 棉花植株的 GUS 测定结果

构建物	启动子/内含子	测试组织	GUS 结果
pMON26177	EF1 α	叶	11600
pMON26177	EF1 α	小花冠	396
pMON26177	EF1 α	小雌蕊	8670
pMON26177	EF1 α	小雄蕊	13771
pMON26177	EF1 α	中花冠	362
pMON26177	EF1 α	中雌蕊	3318
pMON26177	EF1 α	中雄蕊	8006
pMON26177	EF1 α	大花冠	351
pMON26177	EF1 α	大雌蕊	500
pMON26177	EF1 α	大雄蕊	15512
pMON26170	Act2	叶	12718
pMON26170	Act2	小花冠	1296
pMON26170	Act2	小雌蕊	16684
pMON26170	Act2	小雄蕊	7570
pMON26170	Act2	中花冠	742

pMON26170	Act2	中雌蕊	10041
pMON26170	Act2	中雄蕊	7893
pMON26170	Act2	大花冠	289
pMON26170	Act2	大雌蕊	3218
pMON26170	Act2	大雄蕊	42737
pMON48407	Act11	叶	28289
pMON48407	Act11	小花冠	10
pMON48407	Act11	小雌蕊	40755
pMON48407	Act11	小雄蕊	47834
pMON48407	Act11	中花冠	742
pMON48407	Act11	中雌蕊	52495
pMON48407	Act11	中雄蕊	35573
pMON48407	Act11	大花冠	1072
pMON48407	Act11	大雌蕊	4869
pMON48407	Act11	大雄蕊	42737

实施例 11

在一个温室喷洒测试中,使用 Roundup Ultra™草甘膦制剂和 Track Sprayer 装置(Roundup Ultra 是 Monsanto Company 的注册商标),测试转化植株。植株处于“二”真叶期或更晚的生长阶段,在应用 Roundup®喷洒前使叶干燥。使用的制剂是3磅/加仑 a.e. (酸当量) Roundup Ultra™制剂。使用如下的校准:

对于 20 加仑/英亩喷洒体积:

喷嘴速度: 9501 恒流(evenflow)

喷洒压力: 40 psi

喷洒高度: 天棚顶和喷嘴尖之间 18 英寸

轨道(Track)速度: 1.1 英尺/秒, 对应于 1950 – 1.0 伏的读数

制剂: Roundup Ultra™ (3 磅 A.e./加仑)

根据所需测试范围，喷洒浓度将有所变化。例如，如所需喷洒率为 8 盎司/英亩，则使用 3.1 ml/L 的工作溶液，如所需喷洒率为 64 盎司/英亩，则使用 24.8 ml/L 的工作溶液。

根据作物、植物发育阶段和所需的耐性水平，评价周期将有所变化。

实施例 12

用于番茄转化的植物表达构建物列于表 4。在使用草甘膦(Roundup Ultra™)制剂的温室草甘膦喷洒试验中，筛选包含构建物的番茄植株(T0)赋予转基因番茄植株草甘膦耐性的效率，其中所述构建物包含有效连接一个 *aroA*:CP4 草甘膦耐性基因的至少一个肌动蛋白或延伸因子启动子(包含内含子)。任选地通过喷洒应用草甘膦(Roundup Ultra™)，筛选转化进番茄植株的有效连接一个 *aroA*:CP4 基因的至少一个肌动蛋白或延伸因子启动子序列以及有效连接一个 *aroA*:CP4 的一个 eFMV 花椰菜花叶病毒启动子。以 48 盎司/英亩喷洒番茄植株，然后在应用后两周进行评价，分析营养性耐性，并在应用后直到 60 天进行评价，分析生殖性耐性。结果显示于表 4，并根据选择生殖耐受系的效率进行分级。营养性耐性百分率是筛选出的显示对草甘膦损害有足够营养性耐性的系的百分率，考虑进一步研究所述系的农学性状以制备商业化候选物。生殖性耐性百分率是也显示有足够繁殖耐性的营养耐受系的百分率，考虑对所述系进行进一步的农学评价。所有构建物都证明能够为转基因番茄植株有效提供营养性耐性和生殖性耐性。各种启动子组合能够增加在本试验中通过筛选选择营养性耐性系和生殖性耐性系的效率。包含拟南芥属 EF1 α 启动子的构建物更加特异性地与营养性耐受系的高百分率相关。P-Act2 启动子与 P-eFMV 和 P-AtEF1 α (pCGN9190)的组合提高了通过本方法筛选的繁殖性耐受系的百分率。

表 4. 温室轨道喷洒试验, 其中喷洒率为 48 盎司/英亩*

构建物	描述	#测试的系	营养耐受性% ¹	繁殖耐受性% ²
pCGN9190	eFMV/CP4+EF1 α /CP4+ Act2/CP4	930	83.2	52.4
pCGN9153	EF1 α /CP4+eFMV/CP4	391	73.9	38.9
pCGN8086	Act8/CP4	21	47.6	38.1
pCGN8099	FMV- EF1 α /CP4+Act8/CP4	71	84.5	36.6
pCGN8088	eFMV/CP4+Act8/CP4	144	79.9	34.7
pMON45325	eFMV/CP4+Act11/CP4	90	70.0	34.4
pCGN8096	FMV-Act11/CP4+Act2/CP4	201	62.7	10.4
pCGN8067	Act2/CP4	205	67.3	8.8

*施用一次

¹从 25 次筛选汇集的结果。施用后 14 天计数²从 25 次筛选汇集的结果。施用后直到 60 天计数

使用番茄种子产量作为本发明使用的赋予转基因植物草甘膦耐受性的各种启动子序列和表达盒组合的效率的衡量。在表 5 中, 显示了转基因番茄植株三个田间试验的结果, 所述转基因番茄植株包含用本发明的启动子驱动草甘膦耐受性的 *aroA*:CP4 编码序列表达的构建物。试验 1 是用从构建物产生的植株的测试, 所述构建物包含天然形式的玄参花叶病毒启动子(P-FMV)或其重复形式(P-eFMV)以及也在用于测试本发明启动子序列的构建物中出现的所述构建物中的其它遗传元件。也测试其它遗传元件例如 5'非翻译序列和 叶绿体转运肽的来源。构建物 pMON20998 包含连接矮牵牛 Hsp70 5' UTR 的 P-eFMV、连接拟南芥属 EPSPS 叶绿体转运肽(CTP-2)的前导序列, 并连接到 E9 3'终止区。构建物 pMON20999 与 pMON20998 的不同之处仅在于其启动子是 P-FMV。构建物 pMON10156 与 pMON20998 的不同之处仅在于其 CTP 来自矮牵牛 EPSPS 叶绿体转运肽(CTP-4)。构建物 pMON45312 与 pMON20998 的不同之处仅在于其前导序列是天然 FMV 前导序列。

将番茄植株移植到大田, 排列成行。在田间用 Roundup 除草剂以

48 盎司/英亩的比率喷洒处理所述植株。从果实收集番茄种子并称重。一个未喷洒的番茄植株系作为对照以进行比较，每种构建物的效力以对照的百分率表示。试验 1 的结果(表 5 的栏 1)是：FMV 启动子和 P-eFMV 仅提供未喷洒对照种子产量的 5-11%。试验 2 和 3 在不同地点测试本发明的构建物(表 5 的栏 2 和 3)。试验 2 在试验 1 的同一地点进行，构建物 pCGN8099 (图 7)、pCGN9151 (图 11)和 pCGN9190 (图 5)表现很好，提供了相对于未喷洒对照的 25-46%种子。在一个生长季节较冷的不同地点，试验 3 证明：pCGN8068 (图 9)、pCGN8088 (图 8)、pCGN8099、pCGN9151、pCGN9153 (图 6)和 pMON45325 (图 2)能够赋予番茄足够的草甘膦耐性，产生相对于未喷洒对照约 34-77%的正常种子。

表 5. 番茄种子产量试验

	试验 1		试验 2		试验 3	
	种子相对于对照 克数的重量百分比		种子相对于对照 克数的重量百分比		种子相对于对照 克数的重量百分比	
pMON20998	0.52	5.3				
pMON20999	0.84	8.6				
pMON10156	0.50	5.1				
pMON45312	1.07	11.0				
pCGN8068			0.48	8.4	7.06	77.8
pCGN8088			0.43	7.6	3.09	34.1
pCGN8096			0.40	7.0		
pCGN8099			1.85	32.5	6.93	76.4
pCGN9151			1.46	25.7	6.11	67.4
pCGN9153			0.68	12.0	4.03	44.4
pCGN9190			2.64	46.4		
pMON45325			0.31	5.4	3.37	37.2
pCGN8067						
对照	9.73	100.0	5.69	100.0	9.07	100.0

实施例 13

SEQ ID NO:1-8 和 SEQ ID NO:13-21 是由拟南芥 Act1、Act2 (Genbank #U41998)、Act3、Act7、Act8 (Genbank #ATU42007)、Act11 (Genbank #ATU27981)、Act12 和 E1f1 α (Genbank #X16430)基因公共可得的信息设计的 PCR 引物。这些序列用于使用聚合酶链式反应(PCR)扩增技术延伸所述核酸序列(参见例如 Mullis 等, Cold Spring Harbor Symp. Quant. Biol. 51: 263, 1986; Erlich 等, 欧洲专利申请 50,424; 欧洲专利申请 84,796、欧洲专利申请 258,017、欧洲专利申请 237,362; Mullis, 欧洲专利申请 201,184; Mullis 等, 美国专利第 4,683,202 号; Erlich, 美国专利 4,582,788; 和 Saiki 等, 美国专利第 4,683,194 号)。许多 PCR 扩增方法是本领域内技术人员已知的, 并用于鉴定与已知序列邻近的核酸序列。例如, 已经描述了反向 PCR (IPCR)方法, 该方法用于扩增与已知序列核心区邻近的未知 DNA 序列。也可以利用其它方法, 例如捕获 PCR (Lagerstrom M.等, PCR Methods Applic. 1: 111, 1991)和步查 PCR (Parker 等, Nucleic Acids Res 19: 3055, 1991)。许多生产商也已经开发了基于这些方法的改进形式的试剂盒以鉴定目的序列。技术进步包括引物和连接物设计的改进、聚合酶的改进、以及热循环仪的能力已经促进了更快的分离目的序列的有效方法。

表 5a. 分离拟南芥属肌动蛋白和 EF1 α 启动子序列的引物序列

At.肌动蛋白 2 正向: TTTTTTTTGATATCAAGCTTCAACTATTTTTATGTATGC
 At.肌动蛋白 2 反向:
 GCCTCAGCCATGGTGAGTCTGCTGCAAACACACAAAAAGAGTTCAAT
 At.肌动蛋白 8 正向:
 TTTTTTTTGATATCAAGCTTCCATTTTTCTTTGCATAATTC
 At.肌动蛋白 8 反向:
 GCATCGGCCATGGTGAGTCTTCTGCAATCAAAAACATAAAGATCTGA
 At.肌动蛋白 11 正向:
 TTTTTTTTAAAGCTTGATATCACAACCAAATGTCAAATGG
 At.肌动蛋白 11 反向: CCATCTGCCATGGTCTATATCCTGTC
 At. EF1 α 正向: TTTTTTTTAAAGCTTGATATCGGAAGTTTCTCTCTTG
 At. EF1 α 反向: CTTTCCCATGGTAGATCTCTGGTCAACAAATC

At.肌动蛋白 1a 正向: CCAAGCTTAAATGACATCAGATACACGC
At.肌动蛋白 1b 正向: CATAAGCTTAGAGGTCCAAATTCA
At.肌动蛋白 1 反向: CCATCAGCCATGGTCTTCTACCTTTATGCAAA
At.肌动蛋白 3 正向: CCAAGCTTACCACACTCAGATGCATAAACAAACACA
At.肌动蛋白 3 反向: CATCAGCCATGGTCTACTCTCTGCAAAAACA
At.肌动蛋白 7 正向: GCAAAGCTTACTAGTCAACAATTGGCC
At. 肌动蛋白 7 反向: GATCGGCCATGGTTCATAAAAAAAG
At. 肌动蛋白 12 正向: GGAAGCTTGCGGCCGCTTTCTACTCTACATGTTTCT

将拟南芥小植株的叶(1 g)在 9 ml CTAB 缓冲液(Saghai-Maroo等, 1984, PNAS 81: 8014-8018)中匀浆。所述 CTAB 缓冲液含 100 mM TrisHCl, pH 7.8, 700 mM NaCl, 50 mM EDTA, 1% CTAB (溴化烷基三甲基铵)和 140 mM 2-巯基乙醇。在 65℃温育 90 分钟后,加入 4.5 ml 氯仿:异戊醇(24:1),混合样品 10 分钟。通过在 1500 g 离心 10 分钟,分离水层,并用氯仿:异戊醇再萃取。第二次离心后,将水层转移到含 50 ml 10 mg/ml RNA 酶 A (无 DNA 酶)的试管中,在室温下温育 30 分钟以除去 RNA。用 6 ml 异丙醇沉淀 DNA,并重悬浮于 1 ml 10 mM TrisHCl 缓冲液 pH 8.5 中。DNA 溶液用等体积苯酚萃取一次,然后用等体积氯仿:异戊醇萃取一次。离心后,在水层加入 1/10 体积乙酸钠(3 M, pH 5.2),然后加入 2.5 倍体积乙醇。钩起 DNA,在 70%乙醇中洗,然后风干并重悬浮于 0.2 ml 10 mM TrisHCl 缓冲液。

在 50 ml PCR 反应中使用拟南芥属基因组 DNA (100 ng)。包含显示于表 5a 的所述引物的反应包含 0.2 mM 反向和正向引物溶液、200 nM dNTP 和含镁的 PCR 缓冲液、来自 Expand™ High Fidelity PCR System (Roche Molecular Biochemicals)的 DNA 聚合酶混合物。在开始于 94℃变性 2 分钟后,反应物以 94℃0.5 分钟、55℃0.5 分钟和 72℃1.5 分钟循环 35 次。在 1%琼脂糖凝胶上通过电泳分析 PCR 产物。用 T4 DNA 激酶将凝胶分离的代表肌动蛋白 1a、肌动蛋白 1b、肌动蛋白 7 和肌动蛋白 12 序列的 DNA 片段磷酸化,连接进去磷酸化并用 Sma I 切割的 pUC19 克隆构建物中。筛选白色菌落中合适插入片段的存在,用 M13 测序以确认肌动蛋白启动子的存在。选定的克隆命名为 pMON54941

(P-AtAct1a)、pMON54942 (P-AtAct1b)、pMON54943 (P-AtAct7) 和 pMON54944 (P-AtAct12)。随后, 用 HindIII 和 Nco I 消化包含所述插入片段序列的 pUC19 构建物, 释放肌动蛋白启动子 DNA 序列, 凝胶分离所述 DNA 片段并连接进用同样限制酶消化的 pMON26165。用 HindIII 和 Nco I 消化肌动蛋白 3 启动子(P-AtAct3)的 PCR 产物, 然后直接克隆进 pMON26165, 形成 pMON54951。pMON26165 包含 GUS/nos 终止子基因区段。与该启动子区段的连接使得能够通过植物细胞中 β -葡萄糖醛酸糖苷酶的表达测定每种启动子的功能活性。所述植物细胞可以是分离自例如烟草叶原生质体, 或者所述植物细胞可以包含在一种植物组织或器官中, 如叶、根、子叶、下胚轴、胚、花或贮藏器官中。

在大豆下胚轴中测定与 P-e35S 启动子驱动的 GUS 相比, 由这些启动子驱动的 GUS 表达水平(表 6)。将质粒 DNA/金粒子轰击进大豆下胚轴, 48 小时后用组织化学方法测定 GUS 活性。该测定中测试的所有肌动蛋白启动子都在下胚轴组织中显示功能活性, 证明它们在异源作物物种中表达转基因的实用性。为了在作物中稳定表达草甘膦抗性 EPSPS, 在农杆菌双元植物转化构建物中制备包含由本发明拟南芥属肌动蛋白 1a (pMON54952)、肌动蛋白 1b (pMON54953)、肌动蛋白 3 (pMON54954)、肌动蛋白 7 (pMON54955)和肌动蛋白 12 (pMON54956)启动子驱动的 *aroA*:CP4 EPSPS 的构建物。将这些构建物转化进大豆和棉花细胞, 在含草甘膦的组织培养培养基上选择所述细胞并再生成植株, 然后分析 *aroA*:CP4 蛋白的表达以及对施用草甘膦的耐性。进一步通过常规育种方法将草甘膦耐性性状转移进适于栽培的种质中, 产生显示商业上可接受的草甘膦耐性的植株。

表 6. 在瞬时测定中, 不同拟南芥属肌动蛋白启动子与 P-e35S 相比的活性

构建物	GUS 活性
Pe35S/GUS	+++
P-AtAct1a/GUS	++
P-AtAct1b/GUS	++
P-AtAct3/GUS	++
P-AtAct7/GUS	++
P-AtAct12/GUS	+

实施例 14

棉花产量与结蕾头 4 到 5 周内所结棉蕾的数目相关。这些棉蕾保持到成熟的棉铃以及它们对皮棉收成的贡献是产量的关键因素。当测定转基因构建物赋予棉花除草剂耐性的效力时, 棉铃保留的量是效力的量度, 也是所需的性状。在温室条件下测定包含本发明启动子的转基因棉花植株(表 7)的棉铃保留。所述启动子指导赋予草甘膦耐性表型的 *aroA*:CP4 编码序列的表达。通过农杆菌介导的方法或粒子枪方法转化植物。所述粒子枪构建物包含一个额外的含有 GUS 的表达盒, 所述表达盒可用于组织化学定位来自本发明启动子的 β -葡萄糖醛酸糖苷酶活性。在含有草甘膦的培养基上再生转基因植物, 然后在生根培养基使植物生根。生根的小植株转移到盆栽土壤, 并转移到生长室达到种子硬化(hardening off)期。收集这些植株系的种子并种植。来自每个系的十五个植株在 4 叶期以 48 盎司/英亩喷洒草甘膦。来自每个系的至少 8 个存活的植株在 8 叶期再次以 48 盎司/英亩喷洒草甘膦。成熟时, 计数产生头五个棉铃的最初位置上棉铃的数目。喷洒草甘膦后保留的最初位置棉铃数为 3 或更多的那些系(植株图 ≥ 3)用于进一步的研究。表 7 显示了该研究获得的数据。作图系的数目指出第一次施用草甘膦喷洒后存活的系的数目。商业标准是系 1445 (pMON17136), 该系含有驱

动 CTP2-*aroA*:CP4 基因/E9 3'表达的 P-FMV 启动子, 该系在头 5 个棉铃中保留了不到 1 个棉铃。构建物 pCGN8099、pCGN9153、pCGN8088、pCGN8068 在棉花中提供了足够的生殖性草甘膦耐性, 以至于来自所述构建物 14-35%检测的系进一步进行后面的农学试验。

表 7. 温室棉铃保留研究

构建物	启动子	#作图的系	植株图 > 3	% > 3
PCGN8099	FMV:EF1a+e35S: Act8	104	36	34.6%
pCGN9153	EF1α+FMV	36	12	33.3%
pCGN9165	EF1α+35S/GUS	3	1	33.3%
pCGN9152	EF1a	7	0	0.0%
pCGN8088	Act8+FMV	43	6	14.0%
pCGN8086	Act8	7	0	0.0%
pCGN8068	Act2+FMV	37	7	18.9%
pCGN8067	Act2	37	0	0.0%
pCGN8084	Act2+FMV+35S/ GUS	5	0	0.0%
pCGN8085	Act2+FMV/GUS	1	0	0.0%
pCGN9164	Act11+35S/GUS	21	1	4.8%
pMON45325	Act11+FMV	43	0	0.0%
pCGN8096	FMV:Act11+e35S :Act2	14	0	0.0%
pCGN9154	FMV:Act11+e35S :Act2	16	1	6.3%
系 1445	FMV		<1.0	

实施例 15

棉花产量与结蕾头 4 到 5 周内所结棉蕾的数目相关。这些棉蕾保持到成熟的棉铃以及它们对皮棉收成的贡献是产量的关键因素。当测

定转基因构建物赋予棉花除草剂耐性的效力时，棉铃保留的量是效力的量度，也是所需的性状。在两个地点，在田间条件下测定包含本发明启动子的转基因棉花植株的棉铃保留。将转基因棉花系 502-254-2 (pCGN8068)、701-178-2 (pCGN8068)、53-2 (pCGN8088)、178-1 (pCGN9153)和 60-1 (pCGN9153)与 1445 (草甘膦耐受系)和包含构建物 pMON17136 (P-FMV/CTP2-aroA:CP4/E93')相比较，另外还包括一个野生型非转基因系 Coker 130。田间设计是包括 2 行 x 20-30 英尺 x 3 个重复的随机完全区组设计。在棉花植株发育的 8 叶期以 1.12 磅 ai/A = 48 盎司产物和 1.5 磅 ai/A = 64 盎司产物的比率施用作为 Roundup Ultra™ 制剂的草甘膦。积极管理所有棉花小区的杂草防除和害虫防制以及其它农学因素，例如种植时间、施肥、灌溉、PGR 应用和脱叶。棉铃保留百分率如下确定：随机选择每个小区中间两行中间的十个植株(每行五株)进行作图，绘制每个保留的棉铃的位置。应该在第一次开花后 4 周进行第一次作图(生长中期图)，收获时进行第二次作图。收集的数据包括底部五个花节最初位置棉铃的数目，该数目指示转基因棉花系对草甘膦的生殖性耐性。表 8 说明了本发明启动子的好处：它们使得转基因棉花植株能够保留棉铃。该增强的生殖性耐性导致皮棉产量增加(表 9)和种子产量增加(表 10)。

表 8. 底部 5 个最初位置棉铃在生长中期植株作图时的棉铃保留

	位置 1			位置 2		
	未处理	48 oz/A	64 oz/A	未处理	48 oz/A	64 oz/A
(17136)1445	68	67	53	81	63	62
(8068) 502-254-2	87	72	64	77	80	69
(8068) 701-178-2	85	77	60	84	86	76
(8088) 53-2	89	81	80	79	76	73
(9153) 178-1	77	83	73	85	71	79
(9153) 60-1	80	89	81	77	82	87
PM1218BR	92	56	63			

表 9. 皮棉产量(磅/英亩)和产量百分率(位置 1)

栽培品种	未处理	48 oz/A	48 oz/A %	64 oz/A	64 oz/A %
8068-502-254-2-4	1103	960	87.0%	858	77.8%
8068-701-178-2-2	1326	1219	91.9%	1177	88.8%
9153-60-1-1	1177	1206	102.5%	1171	99.5%
9153-178-1-1	1112	769	69.2%	750	67.4%
8088-53-2-11	1283	1071	83.5%	1097	85.5%
1445	1018	563	55.3%	490	48.1%
C130	1200	0	0.0%	0	0.0%
PM 1218 BR	1092	826	75.6%	713	65.3%

表 10. 种子产量(磅/英亩)和产量百分率(位置 1)

栽培品种	未处理	48 oz/A	48 oz/A %	64 oz/A	64 oz/A %
8068-502-254-2-4	3357	2923	87.1%	2646	78.8%
8068-701-178-2-2	3720	3521	94.7%	3328	89.5%
9153-60-1-1	3294	3413	103.6%	3316	100.7%
9153-178-1-1	3468	2355	67.9%	2218	64.0%
8088-53-2-11	3404	2950	86.7%	2968	87.2%
1445	2835	1624	57.3%	1372	48.4%
C130	3272	0	0.0%	0	0.0%
PM 1218 BR	3036	2192	72.2%	1885	62.1%

实施例 16

在转基因拟南芥中比较驱动 CTP2-*aroA*:CP4 编码序列表达的杂种启动子 P-FMV-AtEF1 α (图 13, pMON52059) 和 P-FMV/CTP2-*aroA*:CP4/E93'(pMON15737)的效力。通过真空渗入产生转基因拟南芥植株(Bechtold 等, C R Acad Paris Life Sci 316: 1194-1199), 种子盆栽在土壤中, 置于调整到 24℃、16 小时光(120 $\mu\text{Em}^{-2}\text{s}^{-1}$)周期的生长室内以允许植株的正常生长和发育。通过以 24 盎司/英亩的比率喷洒施用草甘膦除草剂, 选择 pMON52059 V1 事件草甘膦耐性转基因拟南芥属植株, 将存活的植株移植到单个的盆中。在约 16 天后, 对应于抽苔的观察时, 以 24 盎司/英亩的比率第二次喷洒 8 个 pMON52059 V1 植株和

8 个 pMON15737 纯合植株。所述第二次喷洒将确定所述两种构建物赋予生殖性耐性的效力。观察植株对施用草甘膦的营养性效应。所有植株都有完全的营养性耐性，没有观察到不正常的花。然而，出现长角果败育指示在败育的长角果中没有结种子。计数每个植株产生的长角果总数以及包含种子的长角果(能育长角果)的总数，并制成表格。结果显示于表 9，该结果表明：杂种启动子构建物 pMON52059 获得的能育长角果有 89%，比 pMON15737 的 8%改善 10 倍以上。能育结果结构的数目与能够产生的种子数量相关，这对于产量与种子数量相关的作物尤其重要。对于如棉花、大豆、canola、小麦和玉米等作物，对草甘膦的生殖性耐性对于好的产量是绝对必要的。

表 11. 比较杂种启动子 P-FMV-EF1 α (pMON52059) 和 P-FMV (pMON15737)赋予拟南芥属植株对草甘膦的生殖性耐性

pMON52059				pMON15737			
植株 编号	能育 长角果	总 长角果	育性 百分率	植株 编号	能育 长角果	总 长角果	育性 百分率
8819	39	50	78.0%	1	74	540	13.7%
8820	626	691	90.6%	2	23	600	3.8%
8821	507	561	90.4%	3	1	470	0.2%
8822	0	69	0.0%	4	20	646	3.1%
8823	512	534	95.9%	5	43	717	6.0%
8827	326	354	92.1%	6	22	651	3.4%
8833	432	461	93.7%	7	178	868	20.5%
8838	323	374	86.4%	8	40	520	7.7%
总计	2765	3094	89.4%	总计	401	5012	8.0%

实施例 17

向日葵(*Helianthus annuus L.*)是用于油和食物的重要农业作物。将本发明构建物 pMON45325 (图 2)、pMON45332 (图 4)和 pMON45331 (图 3)转化进向日葵。已经报道了农杆菌介导的向日葵转化 (Schrammeijer 等, *Plant Cell Reports*, 9: 55-60, 1990; EP 0 486 234)。本领域内技术人员已知的用转基因表达构建物转化植物的方法包括下胚轴、顶端分生组织、原生质和其它向日葵组织。转基因向日葵系

SFB250-27 包含 pMON20999 (P-FMV/CTP2-*aroA*:CP4/E93')表达盒;
SFB288-01、SFB295-09 包含 pMON45325 (P-FMV-*aroA*:CP4/E93':P-
AtAct11+内含子/CTP2-*aroA*:CP4/E93'); SFB289-01 包含 pMON45332
(P-AtEF1 α + 内 含 子 /CTP2-*aroA*:CP4/E93':P-eFMV/CTP2-
aroA:CP4/E93'); SFB303-08、SFB303-09、SFB303-11 和 HA300B 包含
pMON45331 (P-AtEF1 α +内含子/CTP2-*aroA*:CP4/E9)。测试这些系的草
甘膦耐性, 结果显示于表 12。

向日葵对草甘膦的生殖性耐性可以作为正常头状花序百分率、正
常头状花序大小百分率和花粉产生的函数来测量。在 V-4 叶期和 V-8
叶期, 以 0 盎司/英亩、32 盎司/英亩或 64 盎司/英亩的比率用草甘膦喷
洒这些植株。评价所述向日葵植株对草甘膦的营养性耐性。在植株发
育的 V4 期和 V8 期、在 32 盎司/英亩和 64 盎司/英亩的草甘膦喷洒水
平达到了营养性耐性。

评价草甘膦营养性耐性转基因向日葵系的头状花序数目、正常头
状花序百分率、正常头状花序大小百分率和正常脱落花粉百分率。在
一个地点的一个田间试验中评价这些性状。头状花序计数和花粉产生
的制表显示于表 12。从本发明构建物选出的系显示更高的正常头状花
序百分率、一般更高的正常头状花序大小百分率和更好的花粉脱落。

表 12. 向日葵草甘膦抗性评价

系编号	头状花序数	%正常头状花序	%正常头状花 序大小	%脱落的花粉
SFB250-27	28	29	75	36
SFB288-01	11	36	73	73
SFB295-09	28	57	64	68
SFB289-01	13	38	92	38
SFB303-08	25	68	92	64
SFB303-09	43	81	88	88
SFB305-11	45	71	84	100
HA300B	30	100	97	97
非转基因分离子	0	0	0	0

实施例 18

可以通过多种方法鉴定适当启动子调节所必需的顺式作用调节元件。在一种方法中,进行缺失分析以去除所述启动子的多个区,然后分析获得的启动子片段的启动子活性。假如截短的启动子的活性与原始启动子片段相比发生改变,就认为该 DNA 片段对于启动子调节是必需的。通过这种缺失分析,可以鉴定对于转录正调节或负调节必需的小 DNA 区。也可以鉴定启动子序列基序,工程化新型启动子使其包含这些顺式元件,以调节有效连接的可转录序列的表达。见例如美国专利第 5,223,419 号,特此通过引用完整地结合到本文中,美国专利第 4,990,607 号,特此通过引用完整地结合到本文中,美国专利第 5,097,025 号,特此通过引用完整地结合到本文中。

另一种可采用的方法是寻找启动子间具有相似表达分布型的相似序列。具有重叠活性模式的启动子可能具有共同的调节机制。可以利用几种计算机程序鉴定启动子间保守的序列基序,所述计算机程序包括但不限于 MEME、SIGNAL SCAN 或 GENE SCAN。这些基序可能代表调节启动子的转录因子的结合位点。一旦鉴定了所述序列基序,就可以测定它们的功能。例如,可以从启动子中缺失所述基序序列以确定所述基序对于正常启动子功能是否是必需的。或者,可以将所述基序添加到最小启动子上,以测试它是否足以激活转录。可以如下测试怀疑的负调节元件的充分性:将其添加到活性启动子中,检测启动子活性的降低。一些顺式作用调节元件可能需要其它元件以发挥功能。因此,可以通过本领域内技术人员已知的任何方法,以不同组合测试多种元件。

一旦已经鉴定了有功能的启动子元件,就可以在核苷酸水平上修饰启动子元件以影响蛋白结合。所述修饰可能导致更高或更低的亲和性结合,这将影响从该启动子开始的转录水平。

启动子元件可以加性作用或协同作用影响启动子活性。在这点上,可以串联排列来自不同 5'调节区的启动子元件,获得具有不同活

性范围或不同表达分布型的启动子。因此，来自异源来源的启动子元件组合或者相似元件或相同元件的重复可能赋予有效连接的可转录序列的更高水平表达。例如，可以形成一种启动子元件的多聚体，以增加由该启动子元件所实现模式的特异性表达的水平。

构建包含新型工程化5'调节元件的表达构建物所需的技术方法是本领域内技术人员已知的。在表达构建物中测试所述工程化启动子，并通过有效连接所述新型启动子到合适的报道基因如 GUS 然后在瞬时植物测定中测试，瞬时测试所述启动子。将所述新型启动子有效连接到一个或多个目的基因，并与一个或多个其它调节元件掺入植物转化构建物，然后通过合适的 DNA 传递系统转化进目标植物。通过适于评价所需特征的多种分子方法、免疫诊断学方法、生物化学方法、表型方法或田间方法，评价稳定转化的植株及其后代。

在举例说明并描述本发明的原理后，本领域内的技术人员应当清楚可以在安排和细节上改变本发明而不偏离这些原则。我们要求保护在所附权利要求的精神和范围内的改变。

本说明书中引用的所有出版物和公开专利文件都特此通过引用结合到本文中，其程度等同于具体并分别指出各个出版物或专利申请通过引用结合到本文中。

序 列 表

<110> Fincher, Karen L.
Flasinski, Stanislaw
Wilkinson, Jack Q.

<120> 新型植物表达构建物

<130> 11898.0018.00PC00 MOBS018

<140> US 60/171,173

<141> 1999-12-16

<160> 30

<170> PatentIn version 3.0

<210> 1

<211> 39

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 1

ttttttttga tatcaagctt caactatattt tatgtatgc

39

<210> 2

<211> 47

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 2

gcctcagcca tggtagtct gctgcaaaca cacaaaaaga gttcaat

47

<210> 3

<211> 42

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 3
ttttttttga tatcaagctt ccatttttct ttgcataat tc 42

<210> 4
<211> 47
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 4
gcatcgcca tggtagtct tctgcaatca aaaacataaa gatctga 47

<210> 5
<211> 40
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 5
ttttttttta agcttgatat cacaacaaa tgtcaaatgg 40

<210> 6
<211> 26
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 6
ccatctgcca tggctatat cctgtc 26

<210> 7
<211> 37
<212> DNA
<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 7

ttttttttta agcttgatat cggaagtttc tctcttg	37
--	----

<210> 8

<211> 33

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 8

cttttcccat ggtagatctc tggtaacaaa atc	33
--------------------------------------	----

<210> 9

<211> 1219

<212> DNA

<213> Arabidopsis Thaliana

<400> 9

caactatddd tatgtatgca agagtcagca tatgtataat tgattcagaa tggttttgac	60
gagttcggat gtagtagtag ccattattda atgtacatac taatcgtgaa tagtgatatg	120
atgaaacatt gtatcttatt gtataaatat ccataaacac atcatgaaag acactttctt	180
tcacggctctg aattaattat gatacaattc taatagaaaa cgaattaaat tacgttgaat	240
tgtatgaaat ctaattgaac aagccaacca cgacgacgac taaogttgcc tggattgact	300
cggtttaagt taaccactaa aaaaacggag ctgtcatgta acacgcggat cgagcaggtc	360
acagtcatga agccatcaaa gcaaaagaac taatccaagg gctgagatga ttaattagtt	420
taaaaattag ttaacacgag ggaaaaggct gtctgacagc caggtcacgt tatctttacc	480
tgtggtcgaa atgattcgtg tctgtcgatt ttaattattd ttttgaaagg ccgaaaataa	540
agttgtaaga gataaaccgc cctatataaa ttcatatatt ttcctctccg ctttgaattg	600
tctcgttgtc ctctcactt tcatcagccg ttttgaatct ccggcgactt gacagagaag	660
aacaaggaag aagactaaga gagaaagtaa gagataatcc aggagattca ttctccgttd	720

tgaatcttcc tcaatctcat ctctctccgc tctttcttcc caaggtaata ggaactttct 780
 ggatctactt tatttgctgg atctogactt tgttttctca atttccttga gatctggaat 840
 tcgtttaatt tggatctgtg aacctccact aaatcttttg gttttactag aatcgatcta 900
 agttgaccga tcagtttagct cgattatagc taccagaatt tggcttgacc ttgatggaga 960
 gatccatggt catgttacct gggaaatgat ttgtatatgt gaattgaaat ctgaactggt 1020
 gaagtttagat tgaatctgaa cactgtcaat gtttagattga atctgaacac tgtttaagtt 1080
 agatgaagtt tgtgtataga ttcttcgaaa ctttaggatt tgtagtgcg tacgttgaac 1140
 agaaagctat ttctgattca atcagggttt atttgactgt attgaactct tttgtgtgt 1200
 ttgcagcaga ctcacccatg 1219

<210> 10
 <211> 1271
 <212> DNA
 <213> Arabidopsis Thaliana

<220>
 <221> misc_feature
 <222> (535)..(536)
 <223> y = c 或 t/u

<220>
 <221> misc_feature
 <222> (561)..(561)
 <223> y = c 或 t/u

<400> 10
 ccatttttct ttgcataat tcatgtttat ttttttattt ttttcatctt gcataattca 60
 tgtttaaaag gatataatac tgggtctact acattctcct gacattacgt tttatgtgtt 120
 tgtctttctga aaataatcat caaaatattt caggacttgt ttacgttttc aggagaaaaa 180
 aaataactgt acccttttca atatagaaat aacatttgta gaaatcgtgg attttcctta 240
 ataaacaatc caaaacaaga ccaccgttgt ctctctgact cggtaacacc cgatcgccga 300
 ottgaaaatt agaagaaaaa tgaaaagaat aataaaaaaa aaaaaggaaat gatgattgaa 360
 gctgtcatat atgtcgacc cttcacagtc aatccaatag cctatattcg ccaactgata 420

tatccaacgg ctcacaaatt ttcacaaact tttcaaaaa gtataataaa agaggctgtc 480
 tgacagccat gtcacgttat actttttccg tatgatcgaa atgattcgtc tttggygaat 540
 ttaattatatt ccaaaattga ygactctaaa gaaaaaaaa tagtttttca gataaaccog 600
 cctatataaa tagttcaaca ctcggtttat ttctttctcc ctcaaagaat tgcctcgtcg 660
 tcttcagctt catcgccgt tgcatttccc ggcgataaga gagagaaaga ggagaaagag 720
 tgagccagtt cttcatcgtc gtggttcttg tttcttctc gatctctcga tcttctgctt 780
 ttgcttttcc gattaaggta attaaaacct ccgatctact tgttcttggt ttggatctcg 840
 attacgattt ctaagttacc ttcaaaagtt gtttcgatt tgattttgat tggaatttag 900
 atcgggtcaaa ctattggaaa tttttgatcc tggcacccgat tagctcaacg attcatgttt 960
 gacttgatct tgogttgtat ttgaaatoga tccggatcct ttcgcttctt ctgtcaatag 1020
 gaatctgaaa tttgaaatgt tagttgaagt ttgacttcag attctgttga tttattgact 1080
 gtaacatttt gtcttcgat gagtatggat tcgttgaaat ctgctttcat tatgattcta 1140
 ttgatagata catcatacat tgaattgaat ctactcatga atgaaaagcc tggtttgatt 1200
 aagaaagtgt tttcggtttt ctcgatcaag attcagatct ttatgttttt gattgcagat 1260
 cgtagaccat g 1271

<210> 11
 <211> 1393
 <212> DNA
 <213> Arabidopsis Thaliana

<400> 11
 acaaccaaatt gtcaaatgga atgcatcaga gaccaaacct gtaagagtcc acaaaacaat 60
 tcaaagaaag aatatcaaca attcagagat tcaatcctaa aacaaaaaga gaactgaaac 120
 caaatcgtac ctacacgacc agtgaagata ccaatagaga gctctgttgt agaatacaac 180
 acattaagcg caattagcag aaacagtctc ttcattctgcc gatttccact tgtcactact 240
 ccaaaaacct cccaaaccat ttccaaaaca gacacttttg ccatgtctac atctttccct 300

```

tccccgaaaa acacatcatt tccatcaacg gagtaaataat ccggcggcat atcgatgctc 360
gagaccgtcc tatcgagaaa aggettagcc gcttccgtga ccgccggcgt tcgtggaccg 420
tgagattgct gaaacgagcg agaataagca agcctccgat cattagcagc atatccgaca 480
tcgtgctcc gatcatcagg gagctcgta tcgcctcgag gattaaagga aatggatctc 540
tccattttct tctttgatct taaagttcca acttcggcaa atactaaaat caacagtcag 600
tcgtacaaag aaactctgct tatacagtaa agtcaatggg ccactgttct aagcccatat 660
ataattttag aagcccatag aatacaaaag agtcaagaag cattgaccgc acaagaaaaa 720
aacaattgtt aaaaagggtt ggtagtgtg tatgtatata tgaaatgcaa caaacattat 780
acagcccat taaatagggt gttatagta gatgtcccca ttaaggaact ttatccagcc 840
cattaaatta ctttacagag taaaagagag agagaagatt tacagttacg ttaccaaatt 900
ttcgaaatga ttaattagt aataaataaa taattaaatg tcagttactc tctttagaaa 960
gctaaataag acagctgttt ccaccaacaa cgtgactggc cgtgggggtc tccttcgttc 1020
aaagtgatat tcagaaatca acggtgaga tcttctocat caatatttat tacgggccta 1080
ttccttcctt ttttaaactt caattctcgg gtcacattc tcttcttcat tcgctcgtt 1140
tctctctcaa aaactacaca ccgtaccac accaccacc tcctogtttc ctcagagatc 1200
ccctctctaa cttctaaggt aatcacattt ccataacgtt ccatcgatcat tgattcttca 1260
ttagtatgcg tttatgaagc tttttcaatt taattctctt tggtagatct taagattcct 1320
ctgtttcttg caaaataaag ggttcaatta tgctaataatt ttttatatca attttgacag 1380
gatatagacc atg 1393

```

<210> 12
 <211> 1160
 <212> DNA
 <213> Arabidopsis Thaliana

<220>
 <221> misc_feature
 <222> (565).. (565)
 <223> y = c 或 t/u

<220>

<221> misc_feature

<222> (571).. (571)

<223> y = c 或 t/u

<220>

<221> misc_feature

<222> (638).. (638)

<223> n = a 或 g 或 c 或 t/u

<220>

<221> misc_feature

<222> (656).. (656)

<223> n = a 或 g 或 c 或 t/u

<400> 12

```

ggaagtttct ctcttgaggg aggttgctcg tggaatggga cacatatggt tgttataata      60
aaccatttcc attgtcatga gatcttgagg ttaatatata ctttacttgt tcattatttt      120
atttggtggt tgaataaatg atataaatgg ctcttgataa tctgcattca ttgagatata      180
aaatatattac tctagagaag agtgtcatat agattgatgg tccacaatca atgaaatttt      240
tgggagacga acatgtataa ccatttgctt gaataacctt aattaaaagg tgtgattaaa      300
tgatgtttgt aacatgtagt actaaacatt cataaaacac aaccaacca agaggtattg      360
agtattcacg gctaaacagg ggcataatgg taatttaaag aatgatatta ttttatgtta      420
aaccctaaca ttggtttcgg attcaacgct ataaataaaa ccactctcgt tgcgtattcc      480
atttatcggt cttattgacc ctagccgcta cacacttttc tgcgatattc ctgaggtaaag      540
cgtaaacgta cccttaratc gtctyttttc yttttcgtct gctgatcggt gctcatatta      600
tttcgatgat tgttggaatc gatgctcttt gttgattnat cgttctgaaa attctnatct      660
gttggtttaga ttttatcgat tgtaaatatc aacgtttcac tgcttctaaa cgataattta      720
ttcatgaaac tattttccca ttctgatcga tcttggtttg agattttaat ttgttcgatt      780
gattgttggt tgggtgatct atatacgagt gaacttggtg atttgcgat ttaagatgta      840
tgctgatttg aattgtgatt gggtaattct ggagtagcat aacaaatcca gtgttccttt      900

```

tttctaaggg taattctcgg attgtttgct ttatatctct tgaaattgcc gatttgattg 960
 aatttagctc gcttagctca gatgatagag caccacaatt tttgtggtag aaatcggttt 1020
 gactccgata goggcgttttt actatgattg ttttgtgtta aagatgattt tcataatggt 1080
 tatatatgtc tactgttttt attgattcaa tatttgattg ttcttttttt tgcagatttg 1140
 ttgaccagag atctaccatg 1160

<210> 13
 <211> 29
 <212> DNA
 <213> 人工序列

<220>
 <223> 合成寡核苷酸

<400> 13
 cccaagctta aatgacatca gatacacgc 29

<210> 14
 <211> 24
 <212> DNA
 <213> 人工序列

<220>
 <223> 合成寡核苷酸

<400> 14
 cataagctta gaggtccaaa ttca 24

<210> 15
 <211> 32
 <212> DNA
 <213> 人工序列

<220>
 <223> 合成寡核苷酸

<400> 15
 ccatcagcca tggctctteta cctttatgca aa 32

<210> 16
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 16
ccaagcttac cacaactcaga tgcataaaca aacaca 36

<210> 17
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 17
catcagccat ggtctactct ctgcaaaaac a 31

<210> 18
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 18
gcaaagctta ctagtcaaca attggcc 27

<210> 19
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 19

gatcggccat ggttcactaa aaaaaaag 28

<210> 20

<211> 36

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 20

ggaagcttgc ggccgctttc tactctacat gtttct 36

<210> 21

<211> 31

<212> DNA

<213> 合成寡核苷酸

<400> 21

gactagccgc catggttcaa tctctagctg a 31

<210> 22

<211> 1578

<212> DNA

<213> Arabidopsis Thaliana

<400> 22

taaatgacat cagatacacg cttgtgaacc atctttaag tattgatgga ctcttacta 60

tgaaagctct ctttaaaatt aattttcttt gtacatgtct ctaagcaatg tcaaattaat 120

tagaggtcca aattcaaaaa aatgtcgtat tgaatcatc cattactaaa ttggttcaat 180

gtcagattta aacagcctag ggataattta gtgagatatg agattctact ttcaacatat 240

actaatccta aatctctagc aactttttat ataagctata aatatcatga aaatgtattt 300

taatcgtttc ataatttatg cagtcacact aatggaaaa aggccaatta ttattatttt 360

cttcagacta taaatgaaaa cataaattaa aatgcagatt agttttaaatt ttttaataagt 420

aagtaaaatg cttatagcct tatacaaaat catatttgga agtttctaac attgttgcaa 480

tttgttatca caaatcacag taatatttgt atactaatta gtaattacaa ctatacacia 540

atttaaattgg gtaatcatat atttgtgtcc agtggattga acaaatatgc tcggcccatg 600
 cggaagtaat gccaattttg ggtgagtaaa gcccatgcga aattttcaca taagaaatgc 660
 atgctttttg ttttcaacga catgagttgc atgcttttta tcattgctta tatagttgca 720
 agtttgcaac tccttgatat tttttttatg tagacactac taccaccaa aacttttggt 780
 ctgcttattc ttgtttacta tgtaaaaaaa ataatgaat tgtttattta ctccgatttg 840
 atggagtctg gtttatgagg ttttatagcc tttacagaaa attgatagtt acaaaaatat 900
 tttcaaaaa taaaagggtta aaaccgtcat ttcaagttgt tattgttttg ggggactgga 960
 tttgaaatga aatatagaac oggaaaacaa ggtgagccga agtcgaagcc tttggaccog 1020
 tttttatatt tactcctccc attcctttct ccttcaatcc ttccttcctc ctccctcctt 1080
 cttcttcttc cctcttttca ttttccagcc actacaaact tttctatctc tacttttttt 1140
 cctctcgatt tcaggtactt tttgagaccc tttgttgtga ttttgaaca cacaccccaa 1200
 ttaagtttga tttttgatcc cgcacgatt tcaattcacc cgtttctgag tttcttttgg 1260
 atctgggtgt cttgagctaa tcttttccgat ctgttggtta tcgattttac tcatgcgtat 1320
 gttcattaca ccatttggtta tttgtttaat caacaaaag actcatgttt ttcaaatgtc 1380
 tttaatataa tttttctgat tgaattttat aatatttaca tgattctgga tccagaatat 1440
 ccttcttctt ctccatttt gtccgttatt gatttgtctt tgaaaaagga ttgttctttg 1500
 tatctgtatt ggtgaaaaag gattgttatt tgttgataaa aatttgatct ttaaacaatg 1560
 tttggttttg cataaagg 1578

<210> 23

<211> 1468

<212> DNA

<213> Arabidopsis Thaliana

<400> 23

ttagaggtcc aaattcaaaa aatgtcgta ttgaatcatt ccattactaa attggttcaa 60

tgtcagattt aaacagccta gggataattt agtgagatat gagattctac tttcaacata 120

tactaatcct aaatctctag caacttttta tataagctat aaatatcatg aaaatgtatt	180
ttaatcgttt cataatttat gcagtcacac taatggaaaa aaggccaatt attattattt	240
tcttcagact ataaatgaaa acataaatta aaatgcagat tagtttaaaa ttttaataag	300
taagtaaaat gcttatagcc ttatacaaaa tcatatttgg aagtttctaa cattgttgca	360
atttgttato acaaatcaca gtaatatttg tataactaatt agtaattaca actatacaca	420
aatttaaatg ggtaatcata tatttgtgtc cagtggattg aacaaatag ctogggccat	480
gcggaagtaa tgccaatttt gggtagtaa agcccatgag aaattttcac ataagaaatg	540
catgcttttt gtttcaacg acatgagttg catgcttttt atcattgctt atatagttgc	600
aagtttgcaa ctcttgata ttttttttat gtagacacta ctaccaccaa aaacttttgg	660
tctgcttatt cttgtttact atgtaaaaa aataaatgaa ttgtttattt actcagattt	720
gatggagtct ggtttatgag gttttatagc ctttacagaa aattgatagt tacaaaaata	780
ttttcaaaa ataaaaggtt aaaacgtca tttcaagttg ttattgtttt gggggactgg	840
atttgaaatg aaatatagaa ccggaaaaca aggtgagccg aagtcgaagc ctttgaccc	900
gtttttatat ttactctcc cattcccttc tcttcaatc ctctcttct cctcctccct	960
tcttcttctt cccctcttcc attttccagc cactacaaac ttttctatct ctactttttt	1020
tcctctcgat ttcaggtact ttttgagacc ctttgttggtg attttcgaa acacacccca	1080
attacgtttg atttttgatc ccgcacgat ttcaattcat ccgtttctga gtttcttttg	1140
gatctgggtg tcttgagcta atcttttcca tctgttggtt atcgatttta ctcatgcgta	1200
tggtcattac accatttggt atttgtttaa tcaacaaaa gactcatgtt tttcaaatgt	1260
ctttaatata atttttctga ttgaatttta taatatttac atgattctgg atccagaata	1320
tccttcttct tcttccattt tgtcctgtat tgatttgtct ttgaaaaagg attgttcttt	1380
gtatctgtat tggtgaaaaa ggattgttat ttgttgataa aaatttgatc tttaaacaat	1440
gtttggtttt gcataaaggt agaagacc	1468

<210> 24

<211> 1642

<212> DNA

<213> Arabidopsis Thaliana

<400> 24

tcaagcttac cacactcaga tgcataaaca aacacagcaa gaagattgcc acaaaaaatca	60
taacgaaata atcaagagat agctatcaaa tcgccacogg cgaatcatgt catactcagt	120
atcagaaaca gatatgatag ctcaaaatat ggattaataa tgttactaaa cacatggaca	180
ataatgcac aatattgaaa gaaagaaaat ggttttagcag aagcaaatg gtttagaaag	240
taatgaacta cacattcaca aaggtgaaga attcgtcaag cctacaataa caaatgtcta	300
tactttatga gccacaaaag agatacatca cactatctga acgaaactaa agcaacctaa	360
catagtctag aaactactaa aatgaatgtt tcaaaacaat tttacagaa ggcaaaagtg	420
aaacaacata ctcccttgcg agaacgagga cgaggagcta attcacgtct ggtaacaaca	480
tgcccttgt tcaacccaac gaacaaaccg gtcttcactt gtggagttgt catcttctgt	540
aaatttcata gacaacaaac aaacaaaact ttctattcaa tacaaaatca aattttacaa	600
gagacggatt cagagataat aaagagatga agagagttaa atcaaaaggg attgatagaa	660
gatacctaata caatggatcg agctcctccg gtggttcaga caaagaagg acgccgactg	720
aaaattacat ttttgtatat ataccagaga gactcaagaa aaaaccctag tccagtttgg	780
gcttttattg ggccttataa attttgggtc agttttgaca aagtaaatac aaggctatag	840
ctgctttgct aacgtgatta attatttacc atttaccaaa agccttaacc gaggccgagc	900
gagaaaaaaa aacaaaaaaa aggtagaggg caagaacgtc atttcacaa ggaattgaat	960
cggaaaacga ggtgtgccga attogaagcc tttggaccgg tttttatact tttttacttg	1020
ccattcgttt tttttgttc attggcctca tttgattact tgtttctttg atttctcctt	1080
ccatagaacc gaattgtttt cagtctgaga tttctcctgc cgagagaacg attttaatct	1140
attttcctcg gtaatgttat agcctaattt gtgttttttt ctttttcctg atccggatat	1200
cgttattctg attgacaatt gtcagtttca tcttctatct tgtgaaattt tgattttttt	1260
ccgatctgtg atttcgtcat tgtatcagcg tgcttatatg cgtttgaggc gtaaattgagt	1320

```

gtgtacctca tttatcattt gctatgtttt tttttttaac agagatcttc agctgtaata 1380
ttataataga ttgaattgat aacgtgattc tggatctgga atatatatat gtcacattct 1440
tcttaggatt tgattttgtc tctcttttga tattaatatt cttcactccc ttgaaaatga 1500
atctgtttat tataatgttt agatatattc cttaccggca tttgttttag cataaatatg 1560
aaacatagca ttgactgatt tgtcttttta ttattcttgt ttttttgcca aattggtctc 1620
atgtttttgc agagagtaga cc 1642

```

<210> 25
 <211> 1241
 <212> DNA
 <213> Arabidopsis Thaliana

<220>
 <221> misc_feature
 <222> (26)..(26)
 <223> n = a 或 g 或 c 或 t/u

```

<400> 25
actagtcaac aattggccaa tctttngttc taaattgcta ataaacgacc atttccgtca 60
attctccttg gttgcaacag tctaccogtc aaatgtttac taatttataa gtgtgaagtt 120
tgaattatga aaaacgaaat cgtattaaaa attcacaaga ataaacaact ocatagattt 180
tcaaaaaaac agtcacgaga aaaaaaccac agaccgtttg tctgctcttc tagtttttat 240
tatttttcta ttaatagttt tttgttattt cgagaataaa atttgaacga tgtccgaacc 300
acaaaagccg agccgataaa tcctaagccg agcctaactt tagccgtaac catcagtcac 360
ggtccccggg ctaattcatt tgaaccgaat cataatcaac ggtttagatc aaactcaaaa 420
caatctaog gcaacataga cgcgtcggtg agctaaaaag agtgtgaaag ccaggtcacc 480
atagcattgt ctctcccaga ttttttattt gggaaataat agaagaaata gaaaaaaata 540
aaagagtgag aaaaatcgta gagctatata ttgcacatg tactcgtttc gctttcetta 600
gtgttagctg ctgccgctgt tgtttctect ccatttctct atctttctct ctcgctgctt 660
ctcgaatctt ctgtatcacc ttcttcttct tcaaggtgag tctctagatc cgttcgcttg 720

```

attttgcgc tcgttagtcg ttattgttga ttctctatgc cgatttcgct agatctgttt 780
 agcatgcgtt gtggttttat gagaaaatct ttgttttggg ggttgcttgt tatgtgattc 840
 gatccgtgct tgttggatcg atctgagcta attcttaagg tttatgtgtt agatctatgg 900
 agtttgagga ttcttctcgc ttctgtcgat ctctcgctgt tatttttgtt tttttcagtg 960
 aagtgaagtt gtttagttcg aaatgacttc gtgtatgctc gattgatctg gttttaatct 1020
 tcgatctgtt aggtgttgat gtttacaagt gaattctagt gttttctctt tgagatctgt 1080
 gaagtttgaa cctagttttc tcaataatca acatatgaag cgatgtttga gtttcaataa 1140
 acgtcgttaa tcttcgaaac taagttgtga tctgattcgt gtttacttca tgagcttacc 1200
 caattcattt cggtttcatt ttactttttt ttagtgaac c 1241

<210> 26

<211> 1313

<212> DNA

<213> Arabidopsis Thaliana

<400> 26

ttctactct acatgtttct tgttattagg taaagtatta ggtctttttt ttaaaaaaaa 60
 tgcttaatcc tctgggtacc tcgaaaagg aataatactc tagttagata agtcagcga 120
 tcaacatgac aaaatgaatg aatgtttgct ttaattgggtg gctaaaagct aaatacacag 180
 aaaagtcaaa attcaatctc aaaatcaacc cctctgtctc caatgtccct aatctatacc 240
 aaaatgtcaa tttattttct tgatcatata ttccactaat taaaaataaa tcttctctc 300
 atgaaatttg tcaaggcctt ggaagcctag ttttaaatat taaatggaaa ctatttctc 360
 aacaatcaca ctgttattta gtattgttgt atgttgttca ctactttctt catttgttt 420
 gtaagaaact ataataagca aaaacacata ataaagtctc atgtcaaata atgaatctta 480
 tgcacatgct tgattatttt acttgcacat atccctatca tcattatcac atttgatcaat 540
 taccgttacc atcattactc tcattcttcc cagaactttt tcagcaattt ccatacctca 600
 cccactaaga tcttttacc ttttcttaa ttatagtttg gatagcactc ttttcatag 660

cactgaaatt tcggttgaac acataaatta ctagaaacta gaaggaaatg ttactgaaat 720
 ttcactgatt gtctaaaatt gaataatcta aagaaaatgg ccttttaacc tttttcttag 780
 gcccaaattg gctcattacc actcatgctt gttcggtgac cggattcttc cggtaaaaca 840
 gagcctaaac cgtattttca ggtaggctg gtgttttctt aattctccaa cctaaaaata 900
 gatggacacg tgtctataga ggctgagata ttggtctcaa tgaagaaaac taacggctca 960
 gaccctgta tgaacgatat taagggccaa agttgcttct gttttccaga aatttttgaa 1020
 acccaatttc agggcacgat tccacaacct cttcttttct ttctagatct acgtaaattc 1080
 atcaggatca tggtattttt ttgttttatt tgatgtcaaa attttgatca caaggaggca 1140
 aaaccaatat aaatgtaacg ctaatgcgtt tgattatggt atacgtaacg aattagattt 1200
 aatgggtaca ttttattggt ttagatttag ttatgagatt ggcattaatt attggtgttt 1260
 cctttgaatt tgctatgttt cttatgttga tgtaatcagc tagagattga acc 1313

<210> 27

<211> 1946

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 27

aattctcagt ccaaagcctc aacaaggctc ggttacagag tctccaaacc attagccaaa 60
 agctacagga gatcaatgaa gaatcttcaa tcaaagtaaa ctactgttcc agcacatgca 120
 tcatggctcag taagtttcag aaaaagacat ccaccgaaga cttaaagtta gtgggcatct 180
 ttgaaagtaa tcttgtcaac atcgagcagc tggttgtgg ggaccagaca aaaaaggaat 240
 ggtgcagaat tgtaggcgc acctaccaa agcatctttg cctttattgc aaagataaag 300
 cagattcctc tagtacaagt ggggaacaaa ataacgtgga aaagagctgt cctgacagcc 360
 cactcactaa tgcgtatgac gaacgcagtg acgaccacaa aagaattagc ttgagctcag 420
 gatttagcag cattccagat tgggttcaat caacaaggta cgagccatat cactttattc 480

aaattggtat cgccaaaacc aagaaggaac tcccatcctc aaaggtttgt aaggaagaat	540
tcgatatccc cgcggcgcgc ttatcacaac caaatgtcaa atggaatgca tcagagacca	600
aacctgtaag agtccacaaa acaattcaaa gaaagaatat caacaattca gagattcaat	660
cctaaaacaa aaagagaact gaaaccaaat cgtacctaca cgaccagtga agataccaat	720
agagagctct gttgtagaat acaacacatt aagcgcaatt agcagaaaca gtctcttcat	780
ctgccgattt ccacttgtca ctactccaaa aacctcccaa accatttcca aaacagacac	840
ttttgccatg tctacatctt tcccttcccc gaaaaacaca tcatttccat caacggagta	900
aatatccggc ggcatatcga tgctcgagac cgtcctatcg agaaaaggct tagccgcttc	960
cgtgaccgcc ggcgttcgtg gaccgtgaga ttgctgaaac gagcgagaat aagcaagcct	1020
ccgatcatta gcagcatatc cgacatcgct gctccgatca tcaggagct cgttatcgcc	1080
tcgaggatta aaggaaatgg atctctccat tttcttcttt gatcttaaag ttocaacttc	1140
ggcaaatact aaaatcaaca gtcagtcgta caaagaaact ctgcttatac agtaaagtca	1200
atgggccact gttctaagcc catatataat tttagaagcc catagaatac aaaagagtca	1260
agaagcattg accgcacaag aaaaaacaa ttgttaaaaa gggttggtta gtgtgtatgt	1320
atatatgaaa tgcaacaaac attatacagc ccattaaata tggttgttat aggtagatgt	1380
ccccattaag gaactttatc cagccatta aattacttta cagagtaaaa gagagagaga	1440
agatttacag ttaagttacc aaattttcga aatgatttaa ttagtaataa ataaataatt	1500
aatgtcagt tactctcttt agaaagctaa ataagacagc tgtttcacc aacaacgtga	1560
ctggctgtgg ggtcctcctt cgttcaaagt gatattcaga aatcaacggc tgagatcttc	1620
tccatcaata ttattacgg gcctattcct tcttttttta aacttcaatt ctccggtca	1680
cattctcttc ttcattcgct ccgtttctct ctcaaaaact acacaccgt accacaccac	1740
cacctctctc gtttctcag agatccccct tctaacttct aaggtaatca catttcata	1800
acgttccatc gtcattgatt cttcattagt atgcgtttat gaagcttttt caatttaatt	1860
ctctttggta gatcttaaga ttctctgtt tcttgcaaaa taaagggttc aattatgcta	1920

atatttttta tatcaatttt gacagg

1946

<210> 28

<211> 1695

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<220>

<221> misc_feature

<222> (1198)..(1198)

<223> n = a 或 g 或 c 或 t/u

<220>

<221> misc_feature

<222> (1216)..(1216)

<223> n = a 或 g 或 c 或 t/u

<220>

<221> misc_feature

<222> (1125)..(1125)

<223> y = c 或 t/u

<220>

<221> misc_feature

<222> (1131)..(1131)

<223> y = c 或 t/u

<400> 28

aattctcagt ccaaagcctc aacaaggta gggtaacagag tctccaaacc attagccaaa 60

agctacagga gatcaatgaa gaatcttcaa tcaaagttaa ctactgttcc agcacatgca 120

tcatggtcag taagtttcag aaaaagacat ccaccgaaga cttaaagtta gtgggcatct 180

ttgaaagtaa tcttgtaac atcgagcagc tggcttgtgg ggaccagaca aaaaaggaat 240

gggtgcagaat tgtaggcgc acctaccaa agcatctttg cctttattgc aaagataaag 300

cagattcctc tagtacaagt ggggaacaaa ataacgtgga aaagagctgt cctgacagcc 360

cactcactaa tgcgtatgac gaacgcagt acgaccacaa aagaattagc ttgagctcag 420

gatttagcag cattccagat tgggttcaat caacaaggta cgagccatat cactttattc 480

```

aaattggtat cgccaaaacc aagaaggaac tcccatcctc aaaggtttgt aaggaagaat    540
tcgatatcaa gcttgatata ggaagtttct ctcttgaggg aggttgctcg tggaatggga    600
cacatatggt tgttataata aaccatttcc attgtcatga gattttgagg ttaatataata    660
ctttacttgt tcattatfff atttggtgtt tgaataaatg atataaatgg ctcttgataa    720
tctgcattca ttgagatata aaatatttac tctagagaag agtgcataat agattgatgg    780
tccacaatca atgaaatfff tgggagacga acatgtataa ccatttgctt gaataacctt    840
aattaaaagg tgtgattaaa tgatgtttgt aacatgtagt actaaacatt cataaaacac    900
aaccaacca agaggtattg agtattcacg gctaaacagg ggcataatgg taatttaaag    960
aatgatatta ttttatgtta aaccctaaca ttggtttogg attcaacgct ataaataaaa   1020
ccactctcgt tgetgattcc atttatcgtt cttattgacc ctagccgcta cacacttttc   1080
tgcgatatct ctgaggtaa ggttaacgta cccttarata gtccyttttc yttttcgtct   1140
gctgatcggt gctcatatta tttcgatgat tgttggtatc gatgctcttt gttgattnat   1200
cgttctgaaa attcctnatct gttgtttaga ttttatcgat tgtaatatc aacgtttcac   1260
tgcttctaaa cgataattta ttcatgaaac tattttccca ttctgatcga tcttgttttg   1320
agattttaat ttgttcgatt gattgttggt tgggtgatct atatacgagt gaacttgttg   1380
atttgogtat ttaagatgta tgtcgatttg aattgtgatt gggtaattct ggagtagcat   1440
aacaatatca gtgttccttt tttctaaggg taattctogg attgtttgct ttatatctct   1500
tgaaattgcc gatttgattg aatttagctc gcttagctca gatgatagag caccacaatt   1560
tttgtggtag aaatcggttt gactccgata gcggcttttt actatgattg ttttgtgtta   1620
aagatgattt tcataatggt tatatatgtc tactgttttt attgattcaa tatttgattg   1680
ttcttttttt tgcag                                           1695

```

<210> 29

<211> 1800

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<220>

<221> misc_feature

<222> (1068)..(1069)

<223> y = c 或 t/u

<220>

<221> misc_feature

<222> (1094)..(1094)

<223> y = c 或 t/u

<400> 29

```

ggtccgatgt gagacttttc aacaaagggg aatatccgga aacctcctcg gattccattg      60
cccagctatc tgtcacttta ttgtgaagat agtggaaaag gaagggtggct cctacaaatg      120
ccatcattgc gataaaggaa aggccatcgt tgaagatgcc tctgccgaca gtggtcccaa      180
agatggaccc ccaccacga ggagcatcgt ggaaaaagaa gacgttccaa ccaogtcttc      240
aaagcaagtg gattgatgtg atgggtccgat gtgagacttt tcaacaaagg gtaatatccg      300
gaaacctect cggattccat tgcccagcta tctgtcactt tattgtgaag atagtggaaa      360
aggaaggtgg ctactacaaa tgccatcatt gcgataaagg aaaggccatc gttgaagatg      420
cctctgccga cagtgggtccc aaagatggac cccacccac gaggagcatc gtggaaaaag      480
aagacgttcc aaccaogtct tcaaagcaag tggattgatg tgatatcaag cttccatttt      540
tcttttgcac aattcatggt tattttttta ttttttcat cttgcataat tcatgtttta      600
aaggatatat acatgggtct actacattct cctgacatta cgttttatgt gtttgtcttc      660
tgaaaataat catcaaaata tttcaggact tgtttacgtt ttcaggagaa aaaaaataac      720
tgtacccttt tcaatataga aataacattt gtagaaatcg tggattttcc ttaataaaca      780
atccaaaaca cgaccaccgt tgtctcctcg actcggtaac acccgatcgc cgacttgaaa      840
attagaagaa aatgaaaag aataataaaa aaaaaaagg aatgatgatt gaagctgtca      900
tatatgtoga ccctatcaca gtcaatccaa tagcctatat tcgccaactg atatatccaa      960
cggtcacaa attttcaca acttttcaa aaagtataat aaaagaggct gtctgacagc     1020

```

catgtcacgt tatacttttt cegtatgac gaaatgattc gtctttgyyg aatttaatta 1080
 tttccaaaat tgaagactct aaagaaaaaa aaatagtttt tcagataaac cgcctatat 1140
 aaatagttca acactcggtt tattttctct cccctcaaag aattgcctcg tegtcttcag 1200
 cttcatcggc cgttgcatth cccggcgata agagagagaa agaggagaaa gagttagcca 1260
 gatcttcacg gtctgtggtc ttgtttcttc ctcatctctc cgatcttctg cttttgcttt 1320
 tccgattaag gtaattaaaa cctccgatct acttggtctt gtgttgatc tcgattaaga 1380
 tttctaagtt accttcaaaa gttgtttccg atttgatttt gattggaatt tagatcggtc 1440
 aaactattgg aaatttttga tcttggcacc gattagctca acgattcatg tttgacttga 1500
 tcttgcggtg tatttgaaat cgatccggat ctttcgctt cttctgtcaa taggaatctg 1560
 aaatttgaat tgtagttga agtttgactt cagattctgt tgatttattg actgtaacat 1620
 tttgtcttcc gatgagtatg gattcgttga aatctgcttt cattatgatt ctattgatag 1680
 atacatcata cattgaattg aatctactca tgaatgaaaa gcctggtttg attaagaaa 1740
 tgttttcggt tttctcgatc aagattcaga tctttatggt tttgattgca gatcgtagac 1800

<210> 30
 <211> 1742
 <212> DNA
 <213> 人工序列

<220>
 <223> 合成的

<400> 30
 gtccgatgtg agacttttca acaaagggtg atatccggaa acctcctcgg attccattgc 60
 ccagctatct gtcactttat tgtgaagata gtggaaaagg aagggtggctc ctacaaatgc 120
 catcattgcg ataaaggaaa ggccatcggt gaagatgcct ctgccgacag tgggtccaaa 180
 gatggacccc caccacagag gagcatcgtg gaaaaagaag acgttccaac cactcttca 240
 aagcaagtgg attgatgtga tggtcgatg tgagactttt caacaaaggg taatatccgg 300
 aaacctcctc ggattocatt gccagctat ctgtcacttt attgtgaaga tagtggaana 360

ggaaggtggc tctacaaat gocatcattg cgataaagga aaggccatcg ttgaagatgc	420
ctctgccgac agtgggtccca aagatggacc cccacccacg aggagcatcg tggaaaaaga	480
agacgttcca accacgtctt caaagcaagt ggattgatgt gatatcaagc ttcaactatt	540
tttatgtatg caagagtcag catatgtata attgattcag aatcgttttg acgagttcgg	600
atgtagtagt agccattatt taatgtacat actaatcgtg aatagtata tgatgaaaca	660
ttgtatctta ttgtataaat atccataaac acatcatgaa agacactttc tttcacggtc	720
tgaattaatt atgatacaat tctaatagaa aacgaattaa attacgttga attgtatgaa	780
atctaattga acaagccaac cagcagcagc actaacgttg cctggattga ctcggtttaa	840
gttaaccact aaaaaaacgg agctgtcatg taacacgcgg atcgagcagg tcacagtcac	900
gaagccatca aagcaaaaga actaatccaa gggctgagat gattaattag tttaaaaatt	960
agttaacacg agggaaaagg ctgtctgaca gccaggtcac gttatcttta cctgtggtcg	1020
aaatgattcg tgtctgtcga ttttaattat tttttgaaa ggccgaaaat aaagttgtaa	1080
gagataaacc cgccatata aattcatata ttttcctctc cgctttgaat tgtctcgttg	1140
tcctcctcac tttcatcagc cgttttgaat ctccggcgac ttgacagaga agaacaagga	1200
agaagactaa gagagaaagt aagagataat ccaggagatt cattctccgt tttgaatctt	1260
cctcaatctc atcttcttcc gctctttctt tccaaggtaa taggaacttt ctggatctac	1320
tttatgtgct ggatctcgat cttgtttctt caatttctt gagatctgga attcgtttaa	1380
tttgatctg tgaacctcca ctaaactctt tggttttact agaatcgatc taagttgacc	1440
gatcagttag ctcgattata gctaccagaa tttggcttga ccttgatgga gagatccatg	1500
ttcatgttac ctgggaaatg atttgtatat gtgaattgaa atctgaactg ttgaagttag	1560
attgaatctg aacactgtca atgttagatt gaatctgaac actgtttaag ttagatgaag	1620
tttgtgtata gattcttoga aacttttagga tttgtagtgt cgtacgttga acagaaagct	1680
atttctgatt caatcagggt ttatttgact gtattgaact ctttttgtgt gtttcagca	1740
ga	1742

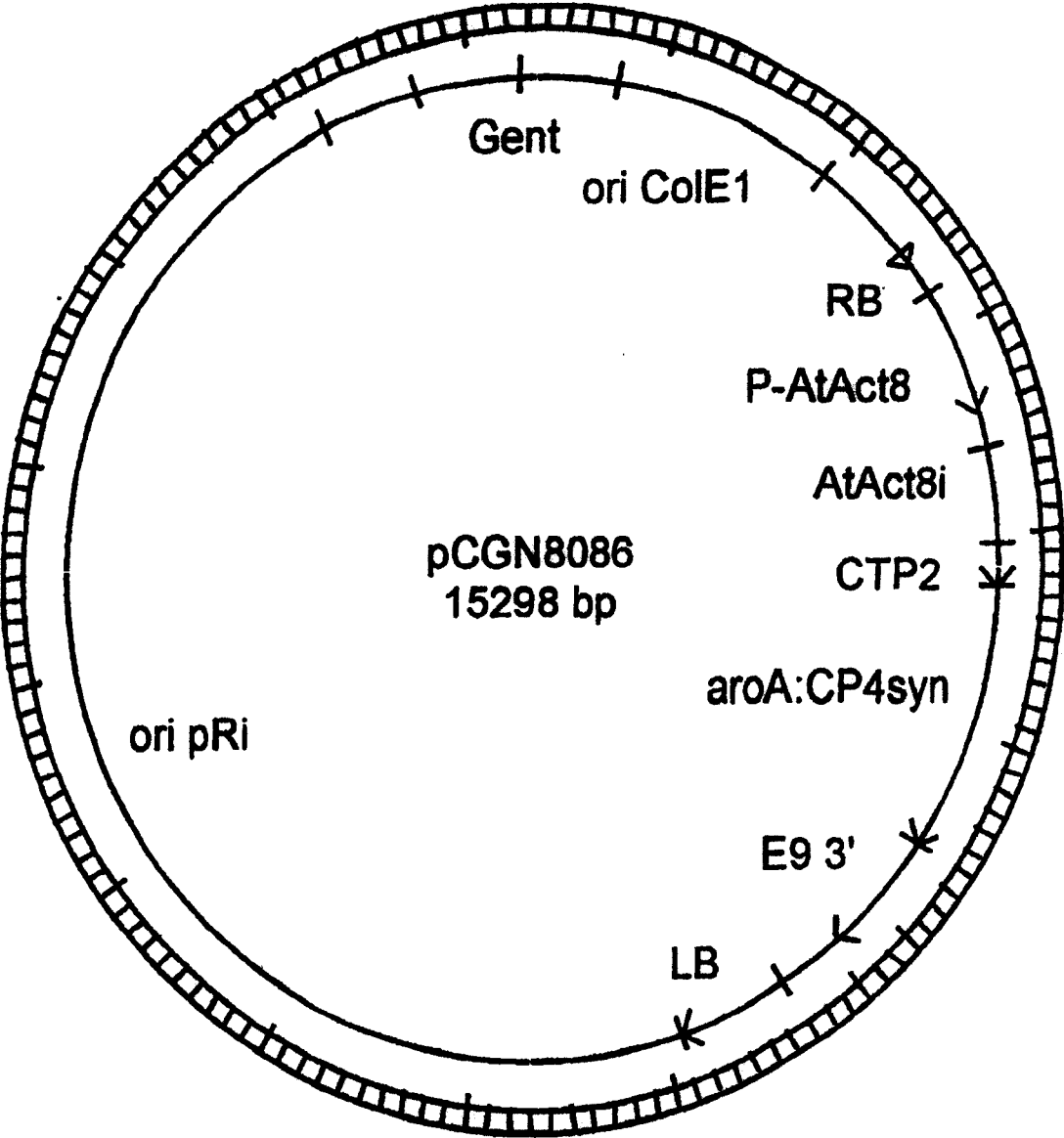


图 1

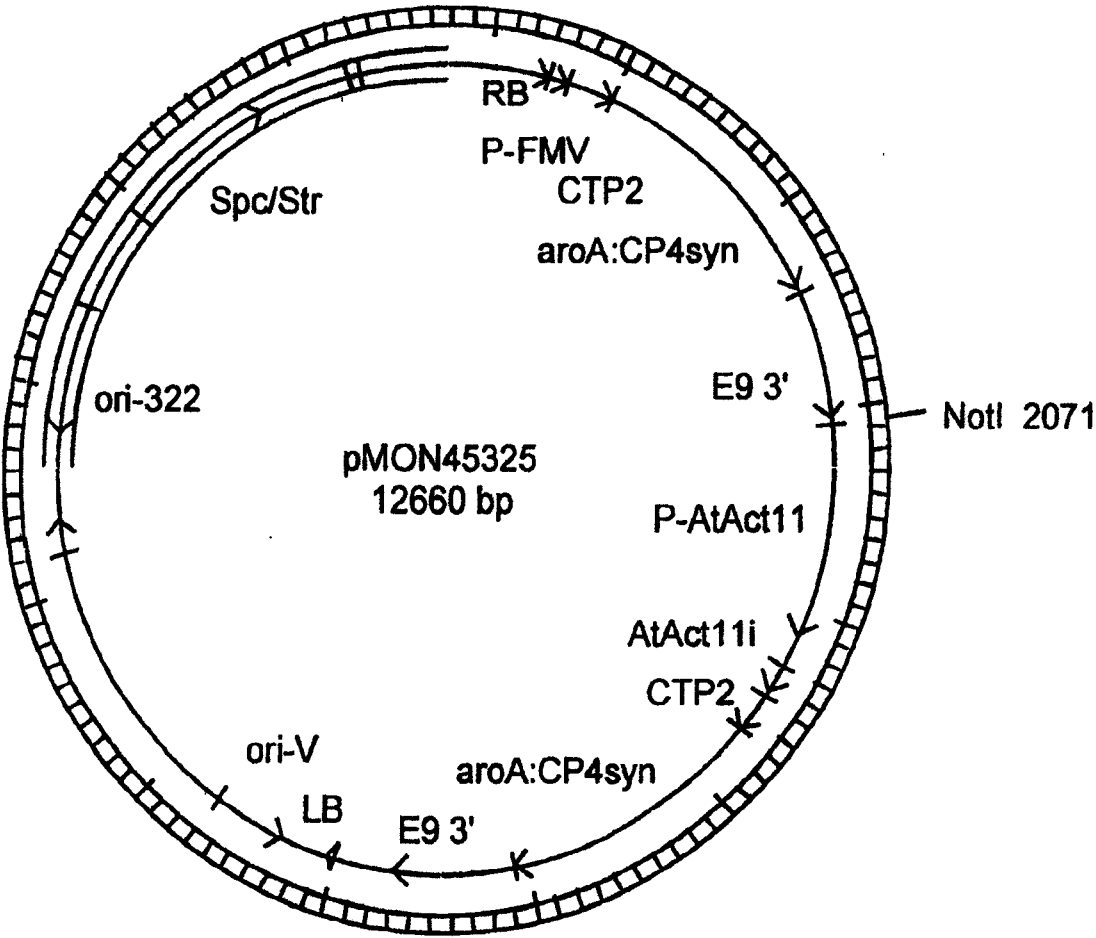


图 2

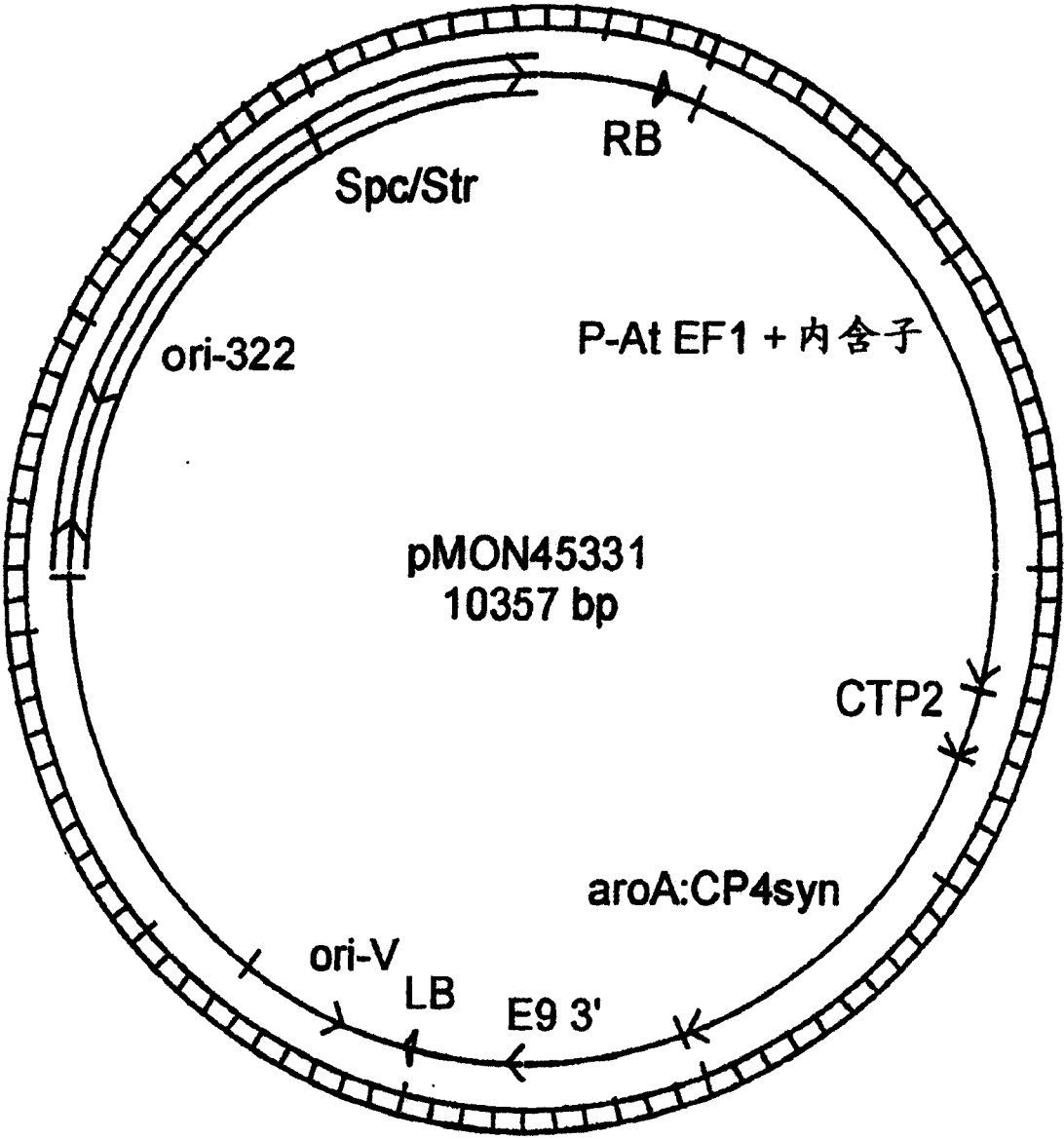


图 3

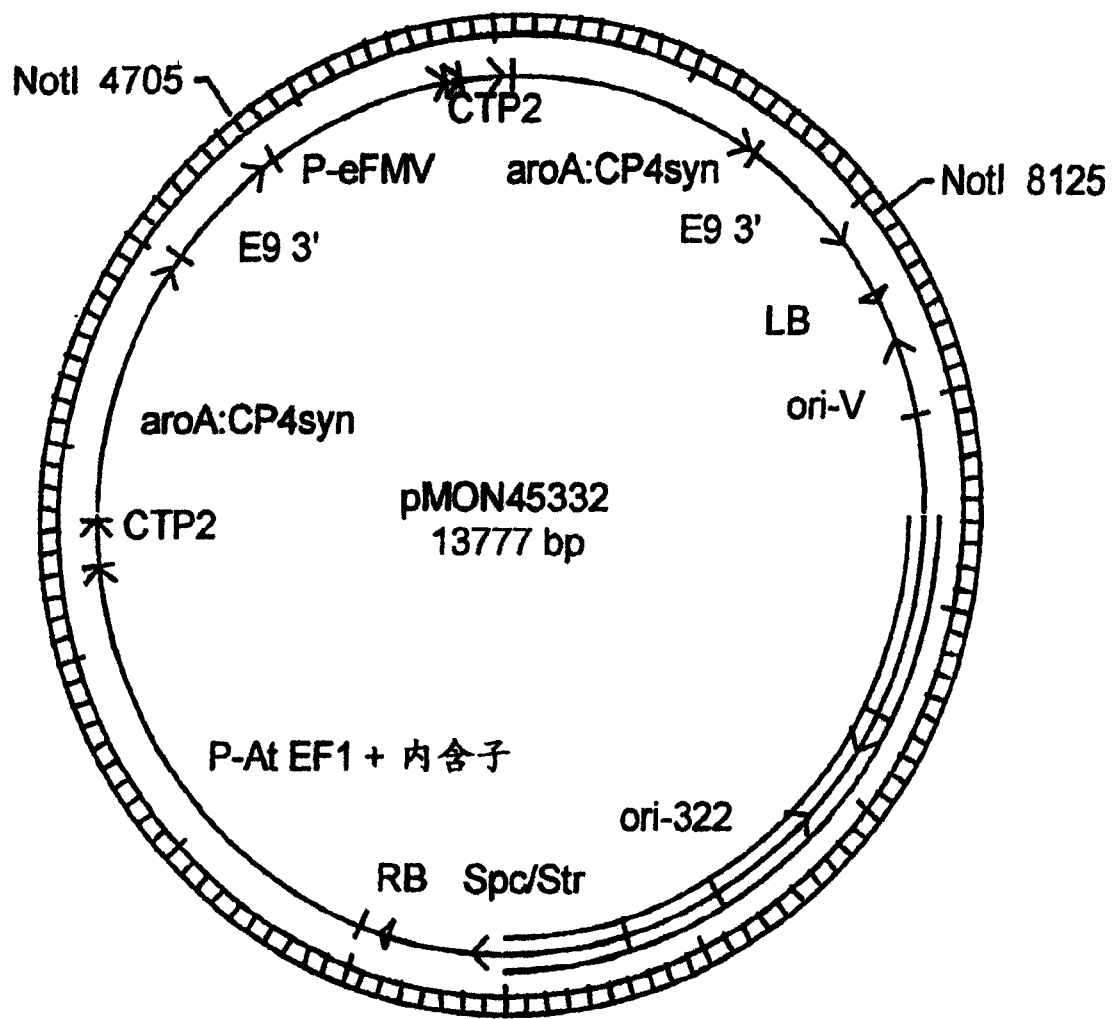


图 4

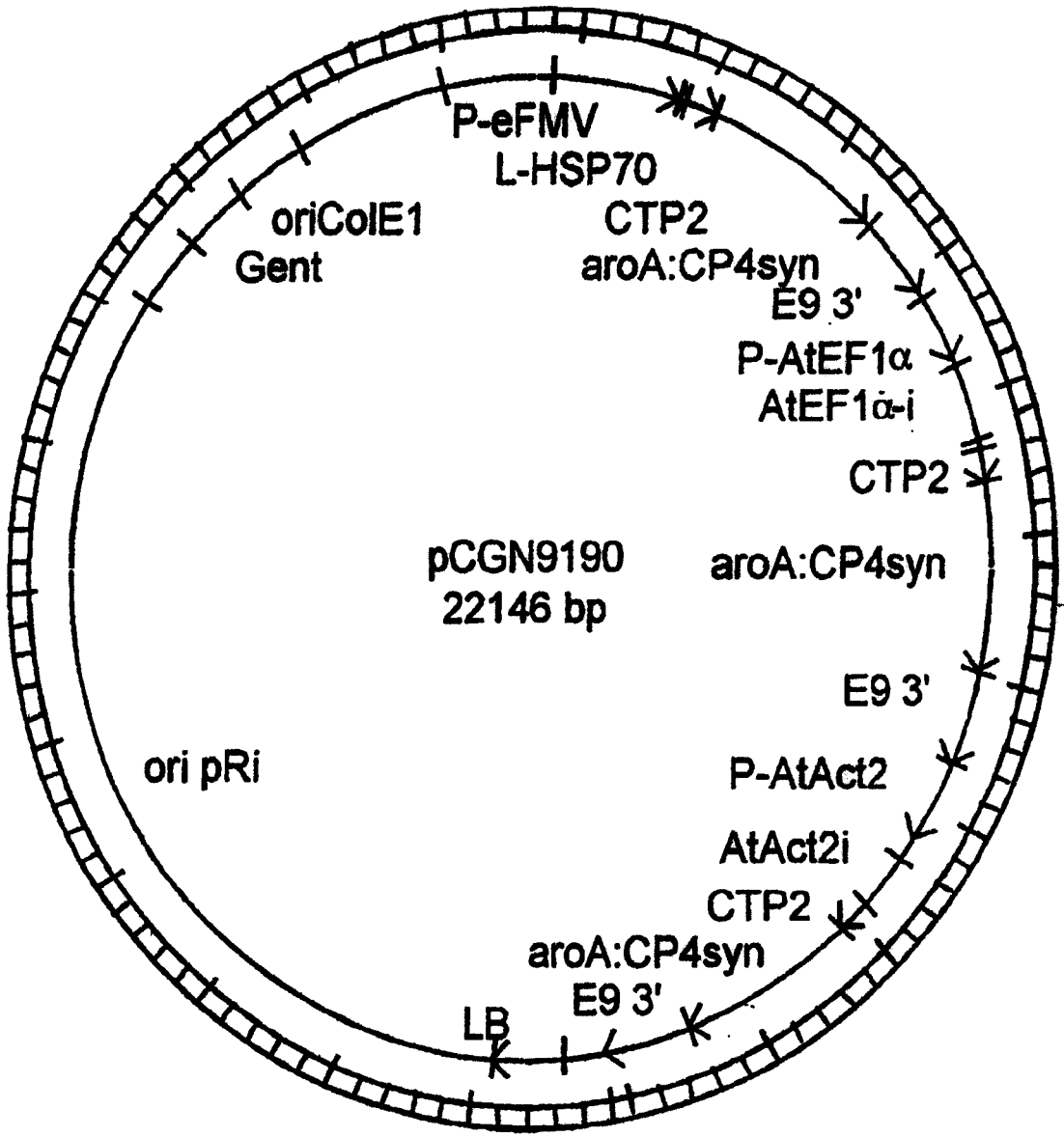


图 5

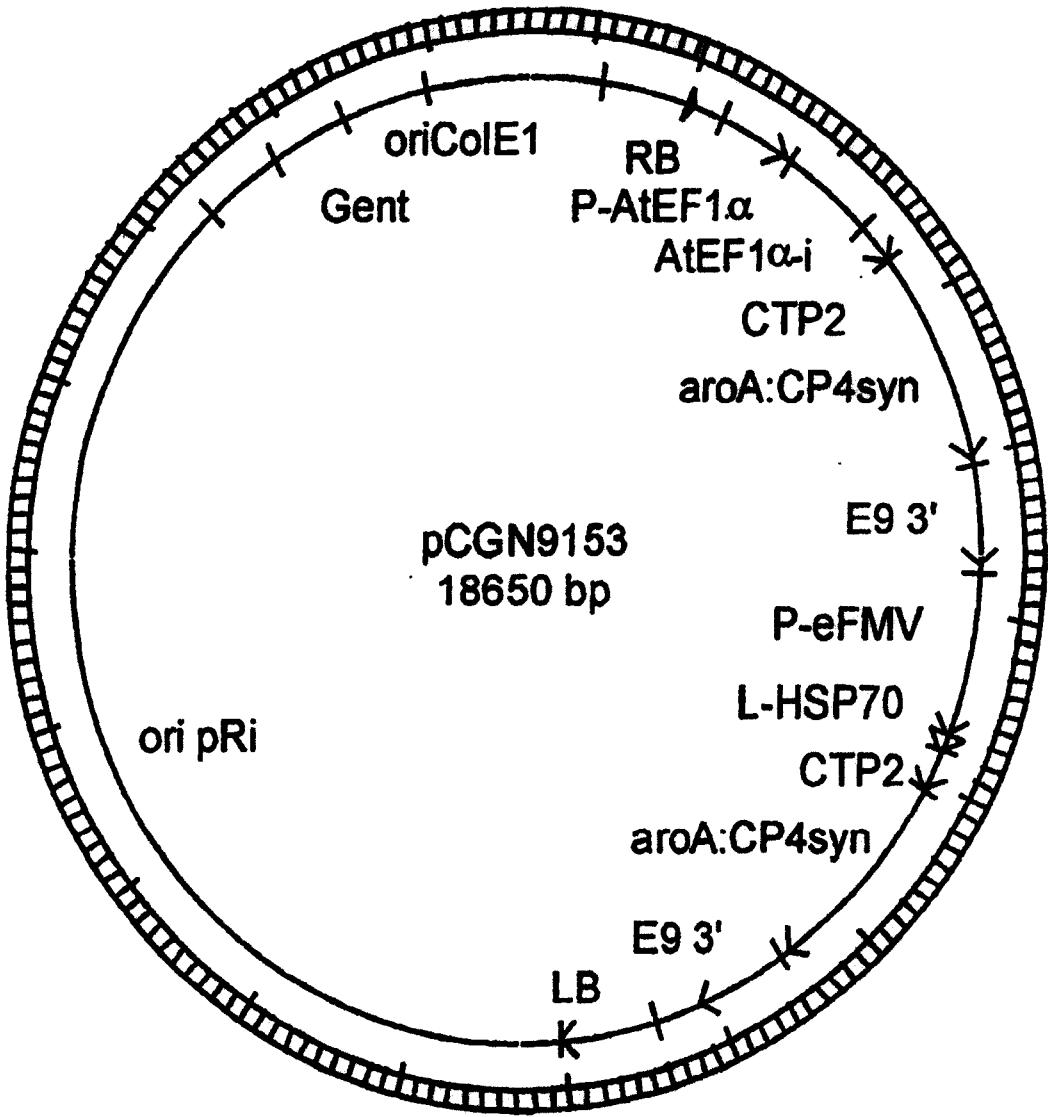


图 6

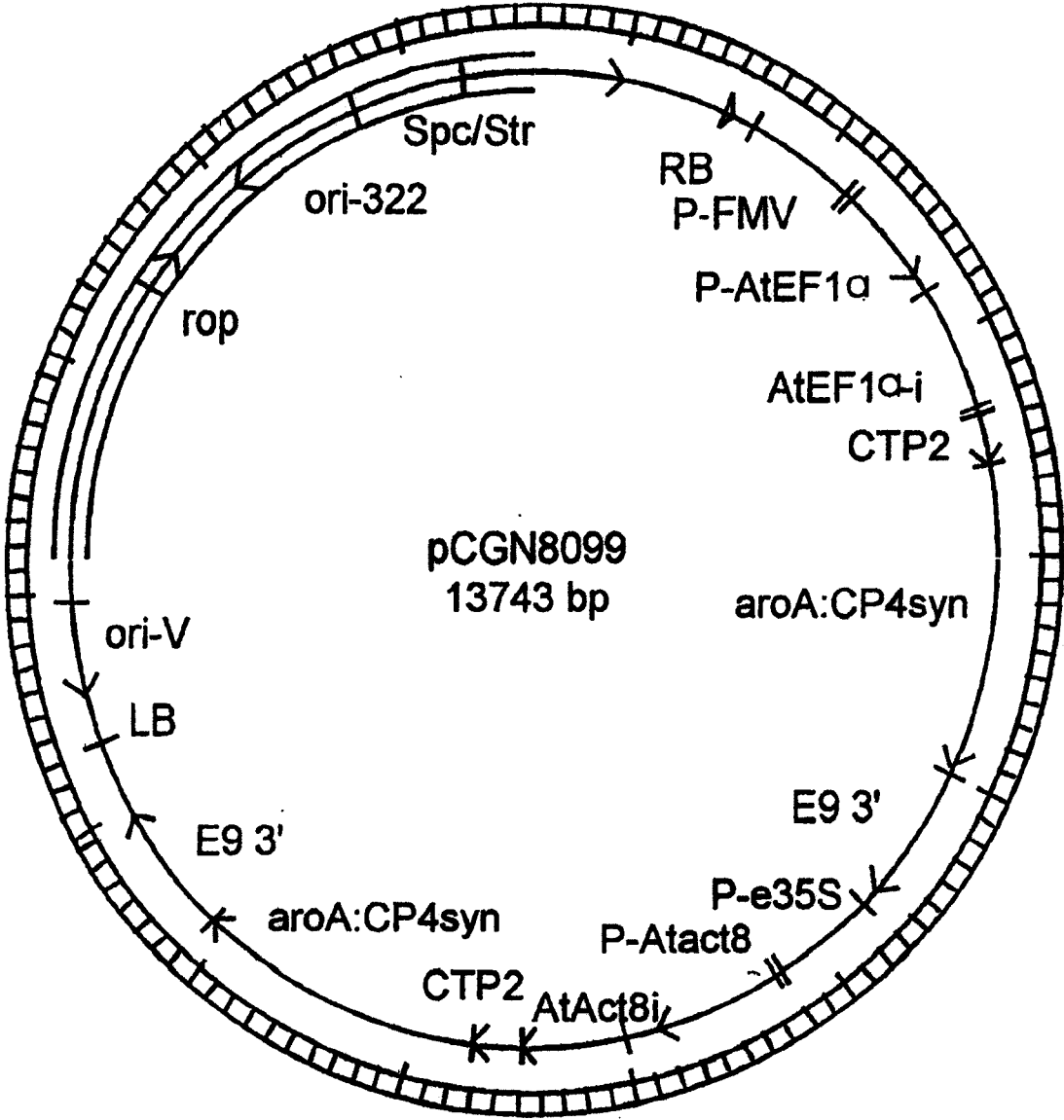


图 7

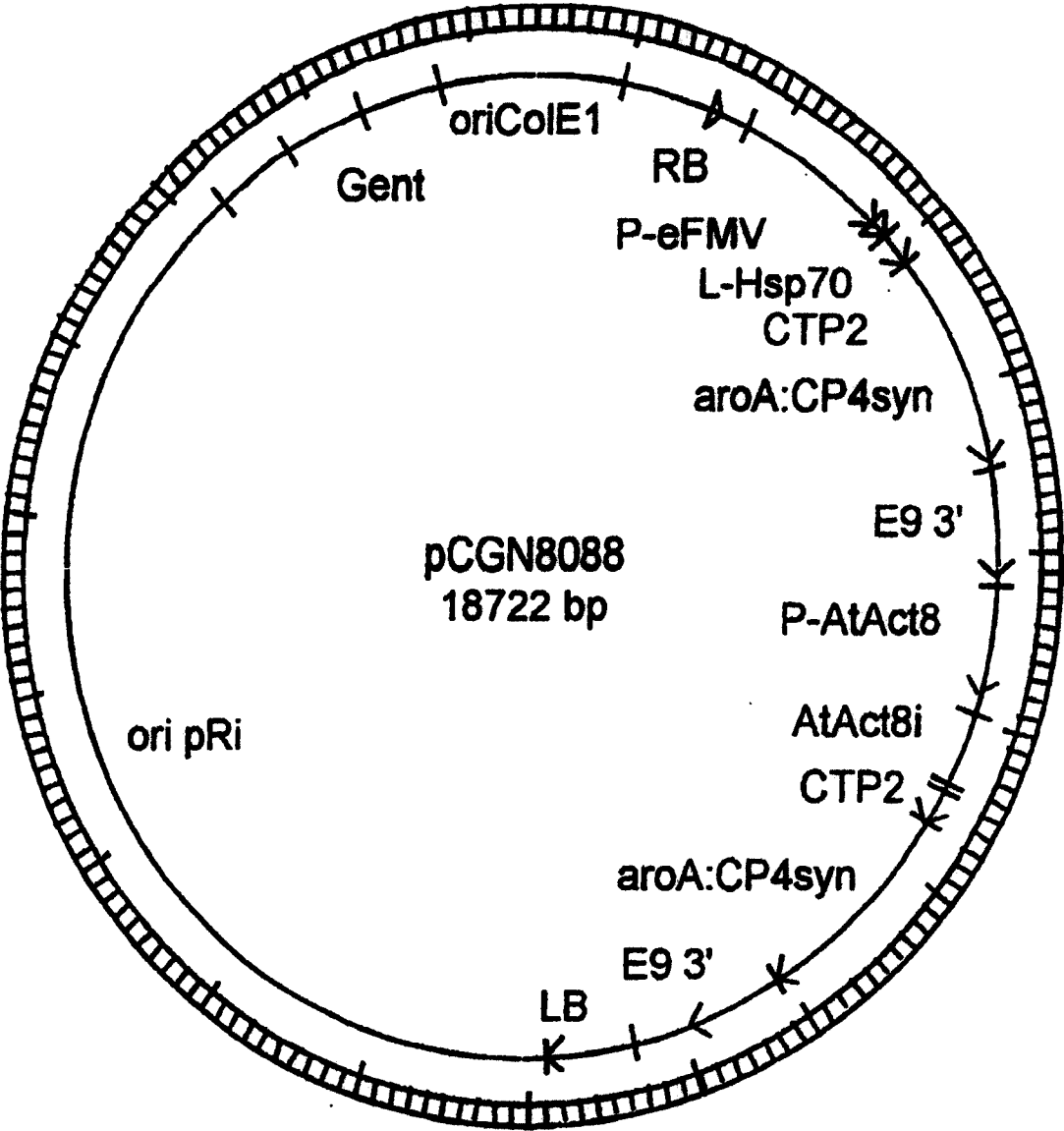


图 8

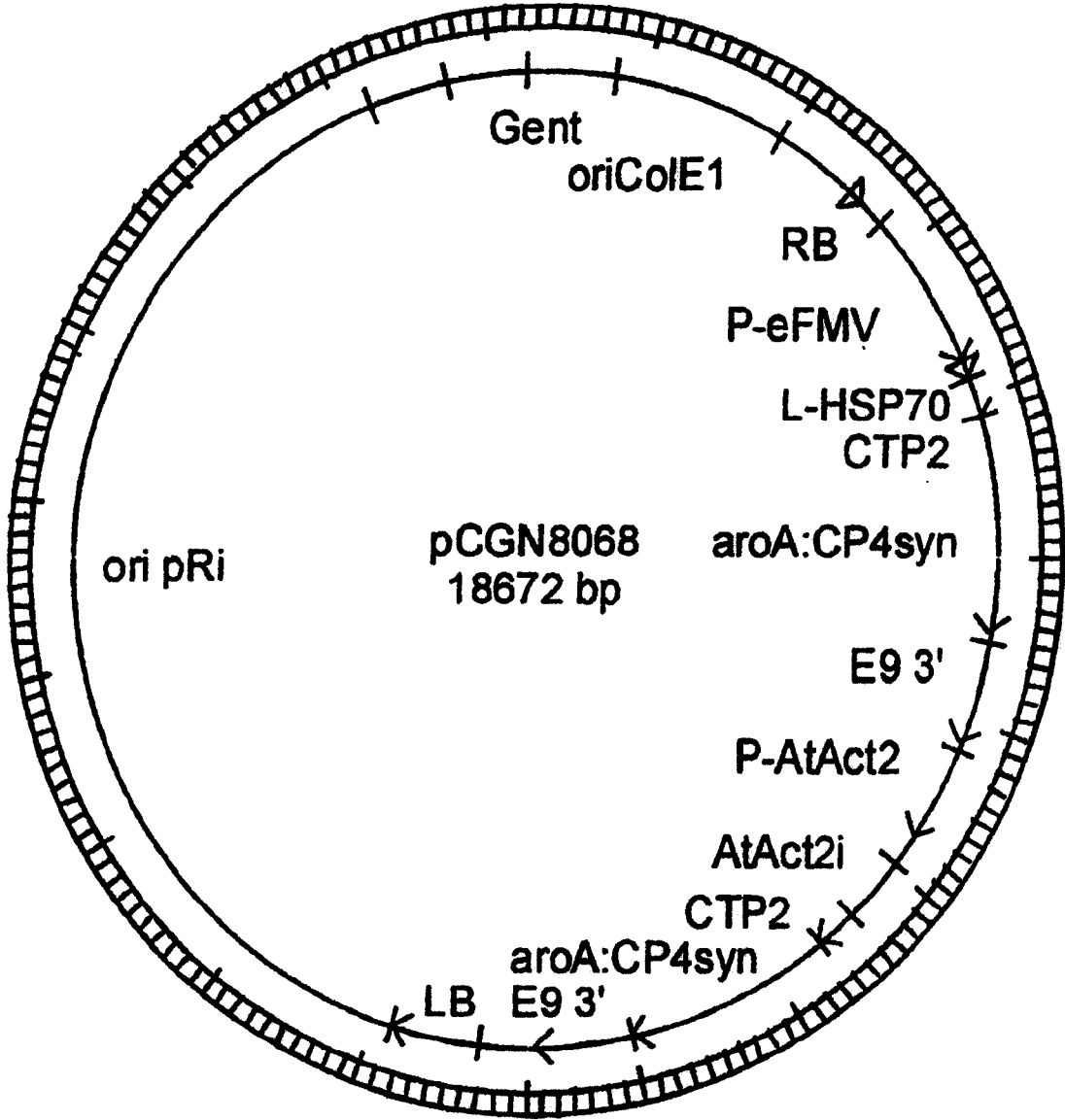


图 9

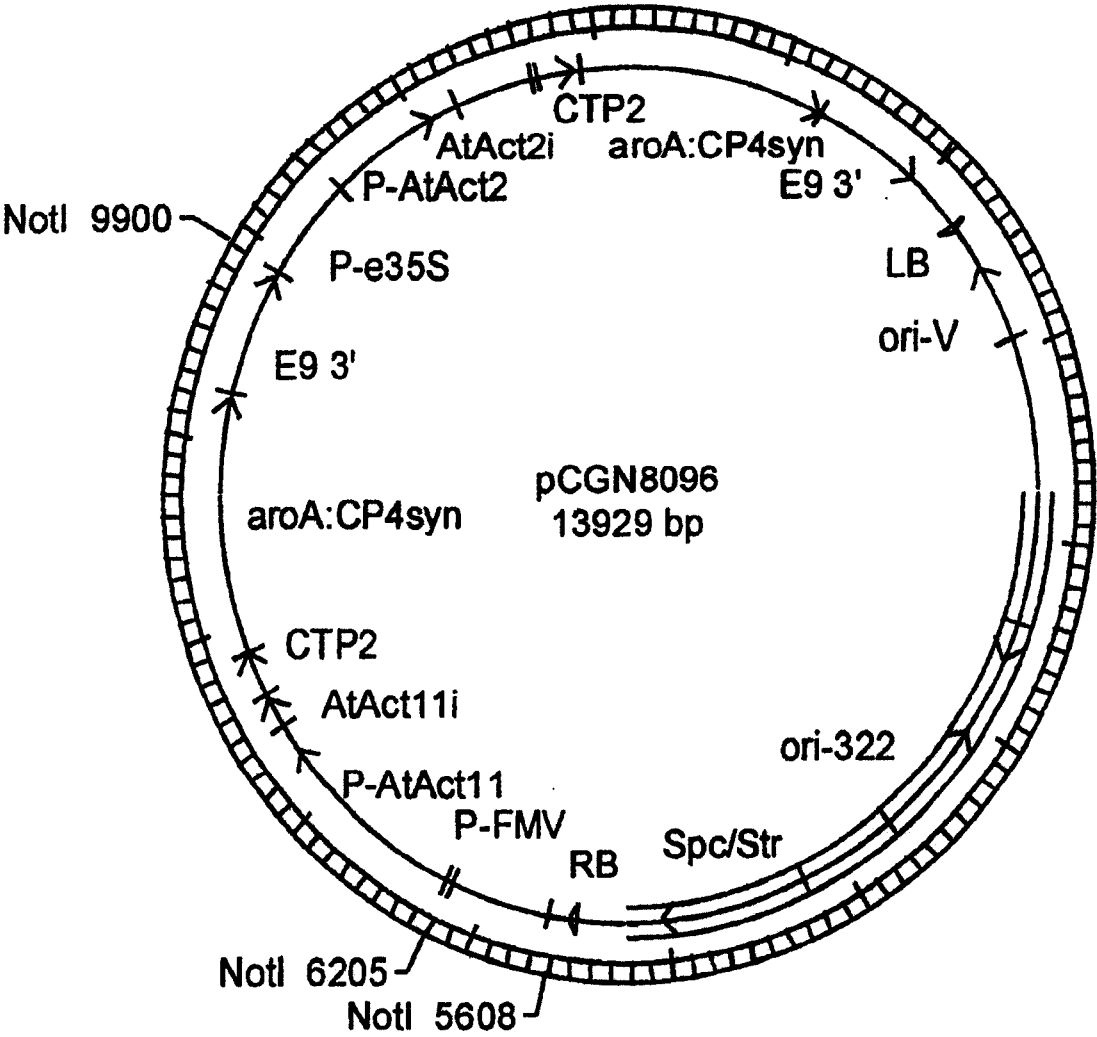


图 10

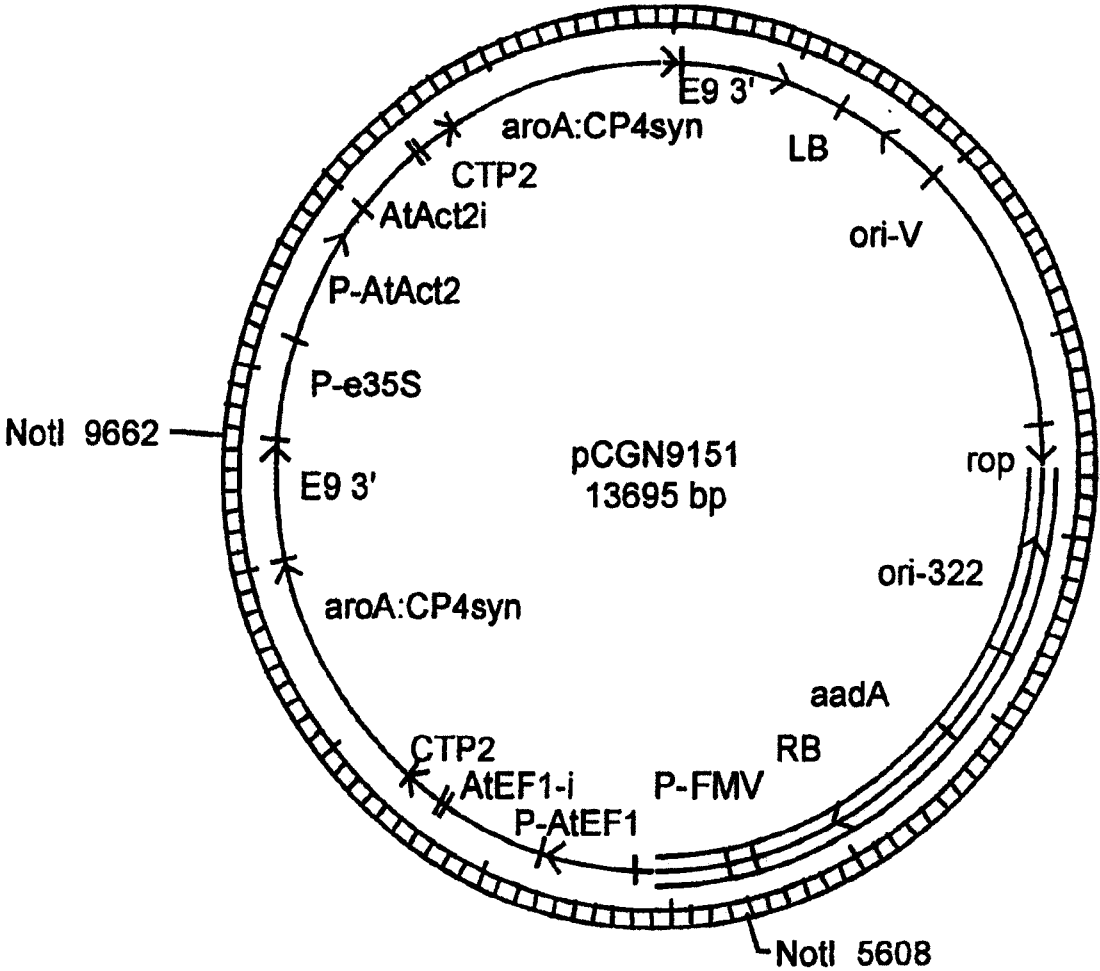


图 11

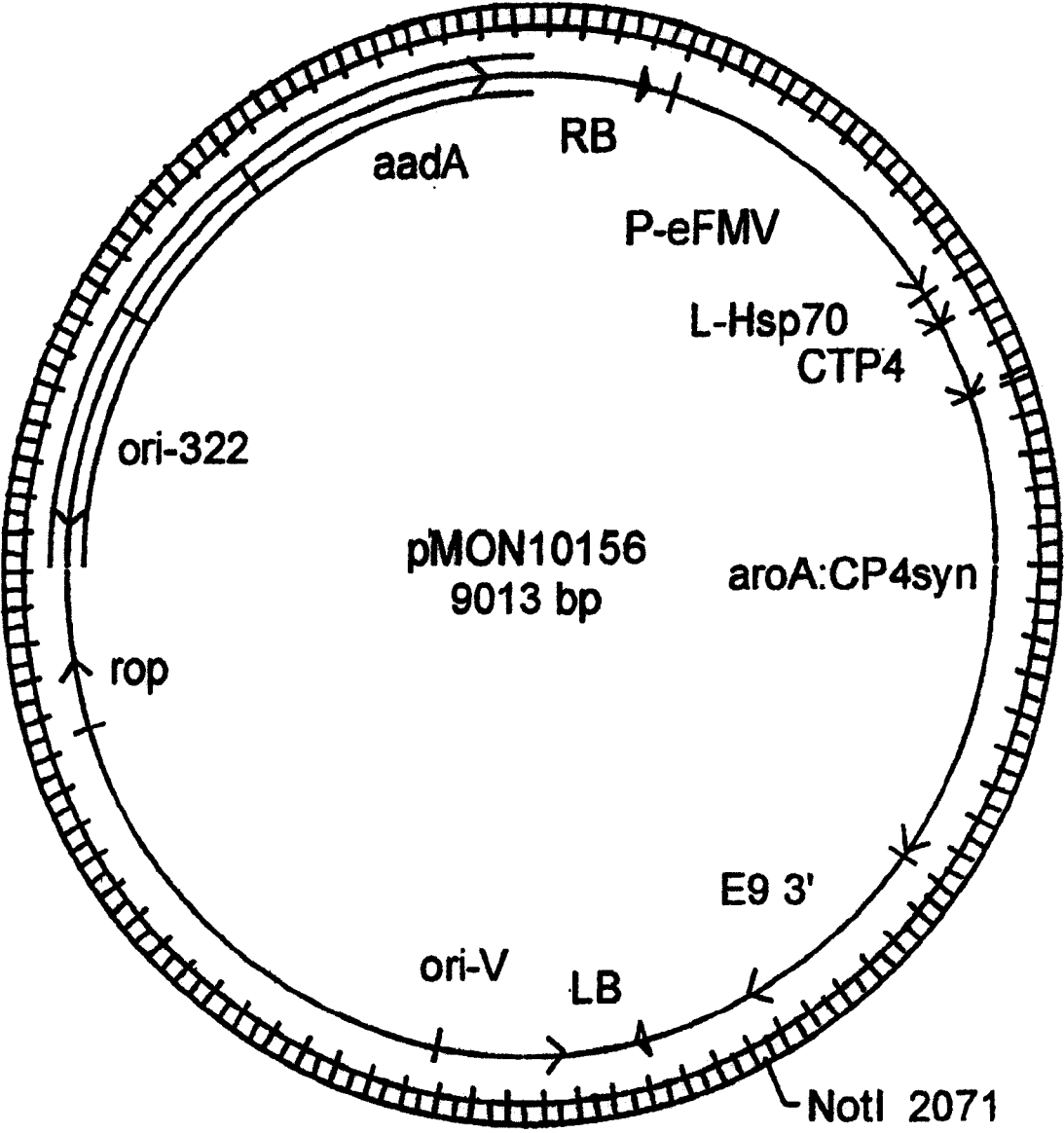


图 12

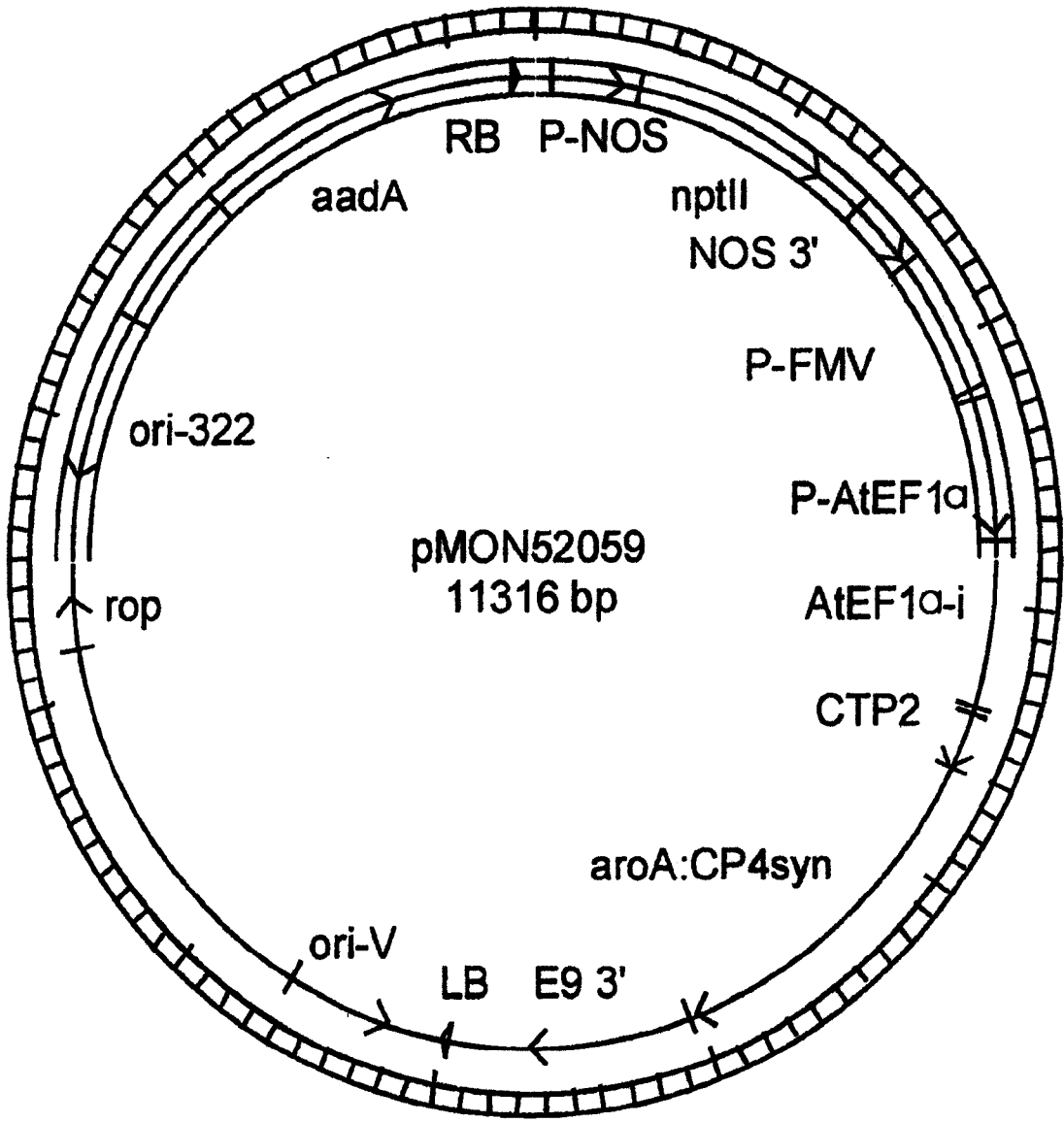


图 13

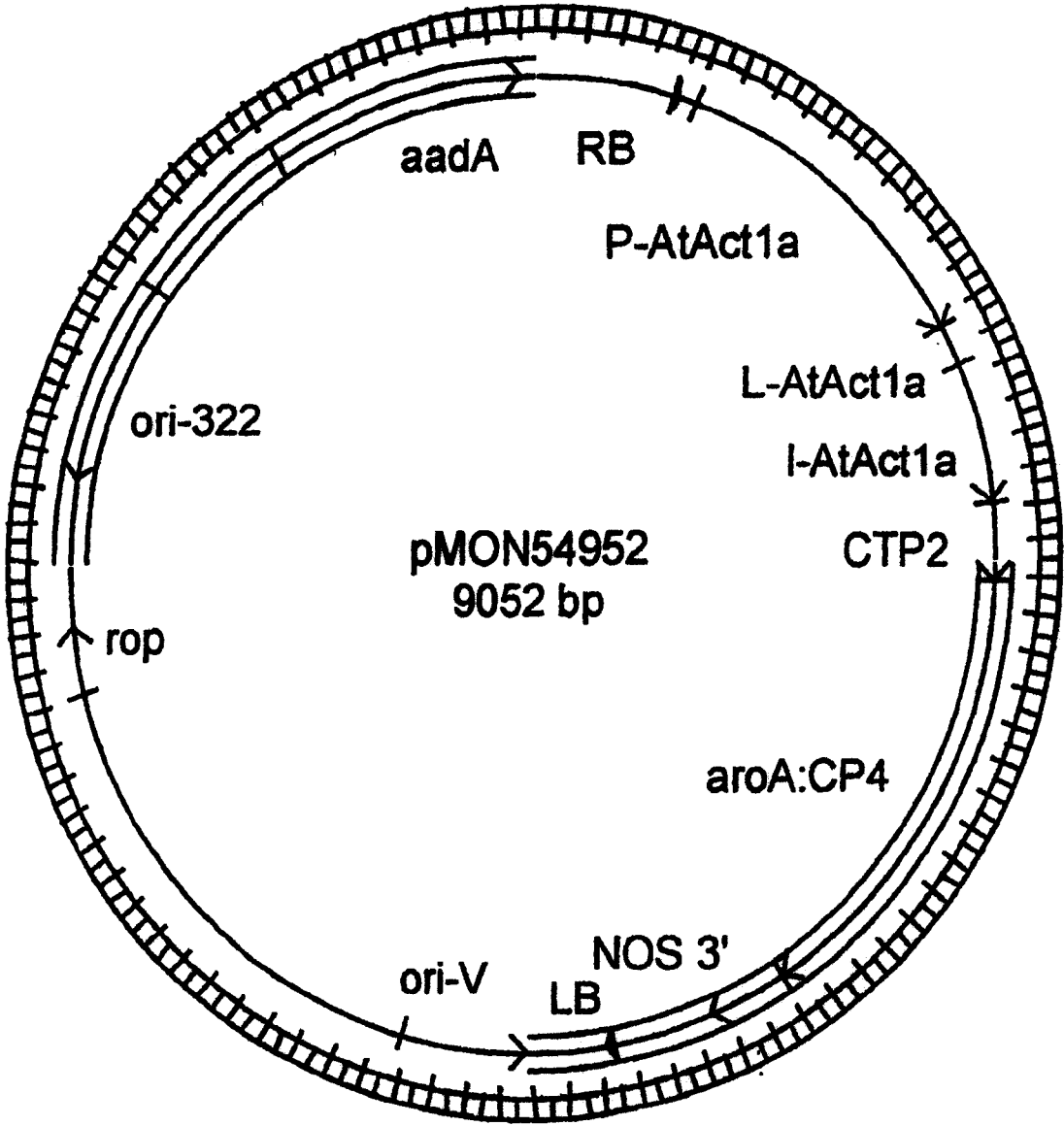


图 14

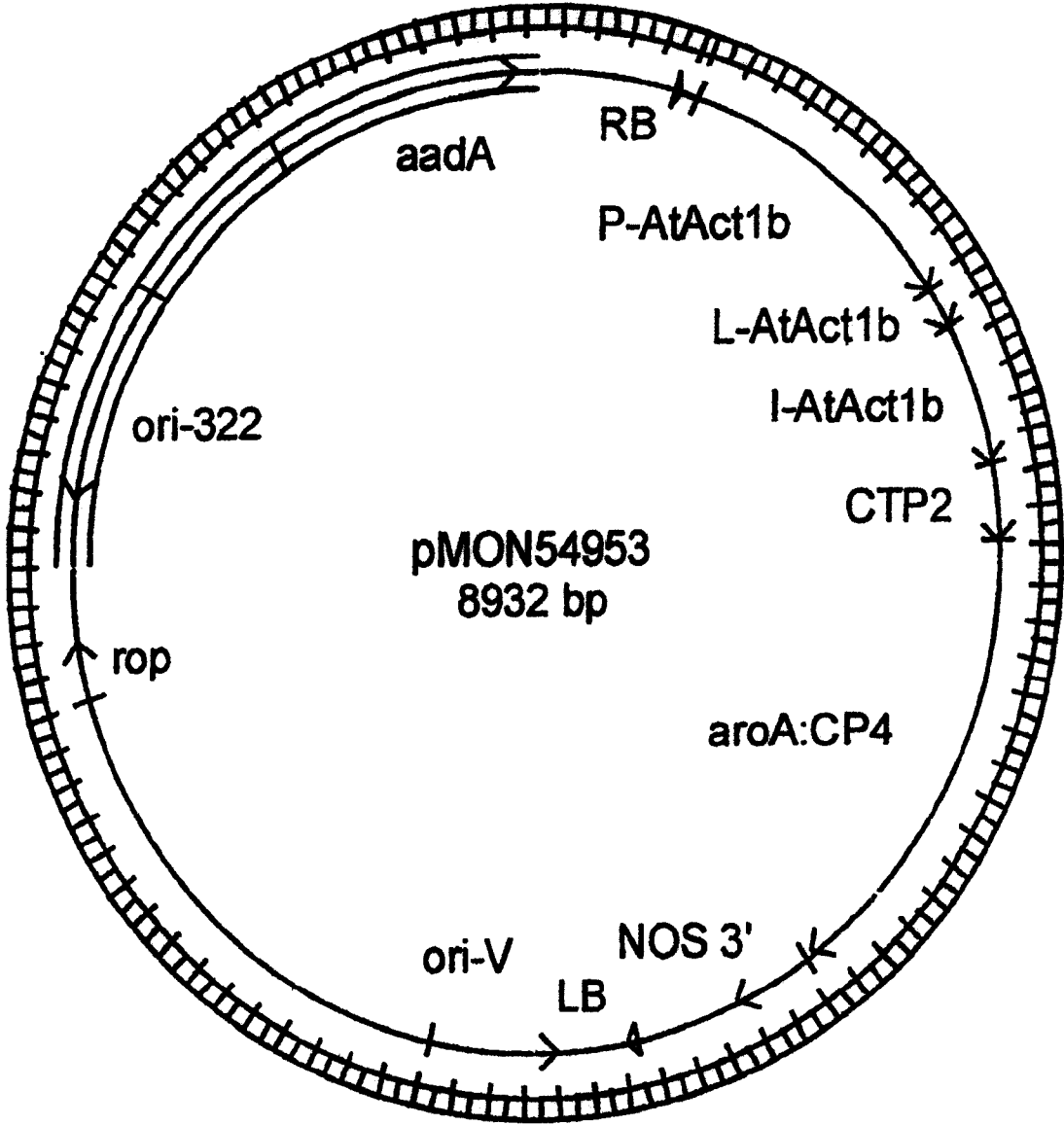


图 15

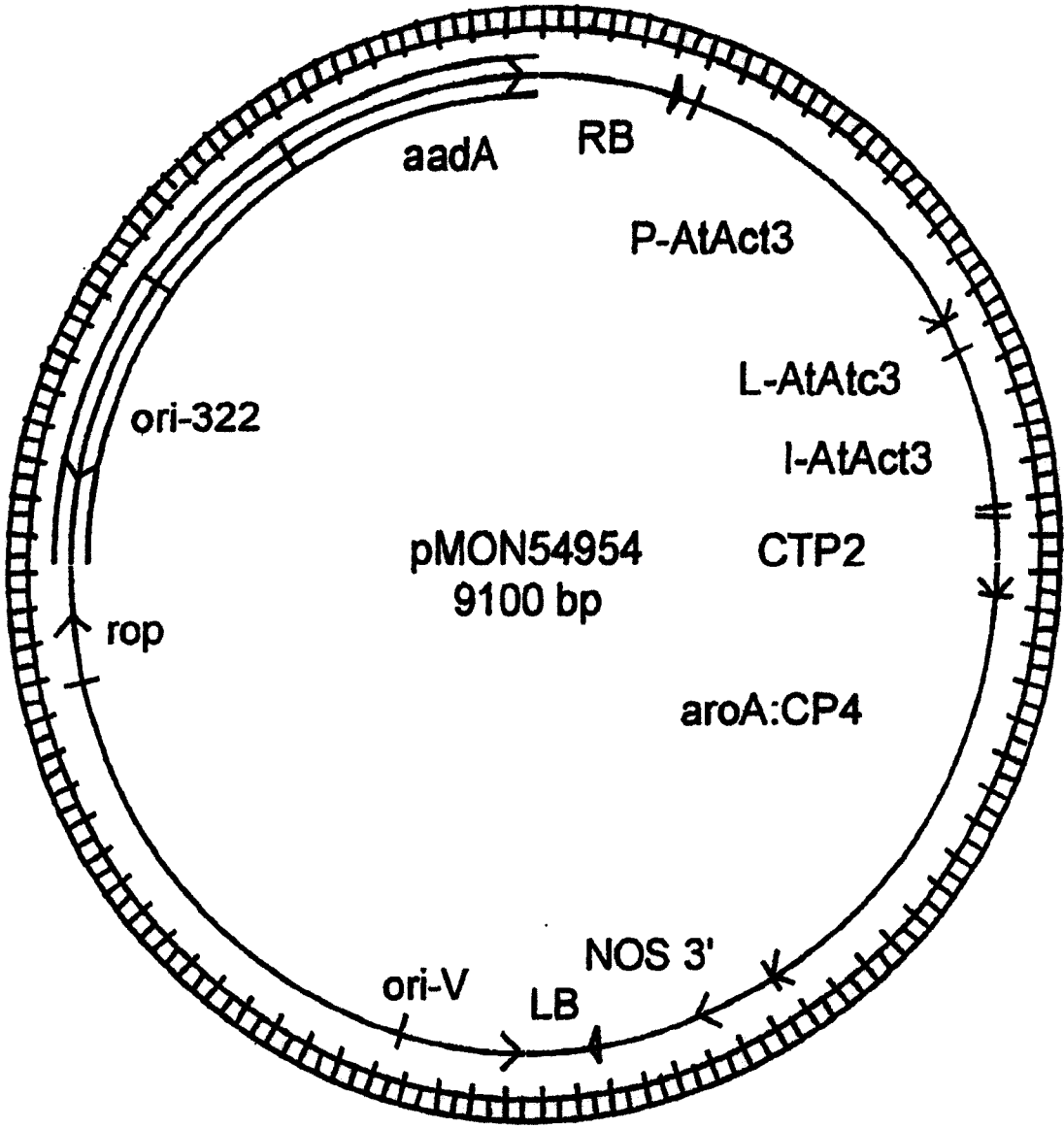


图 16

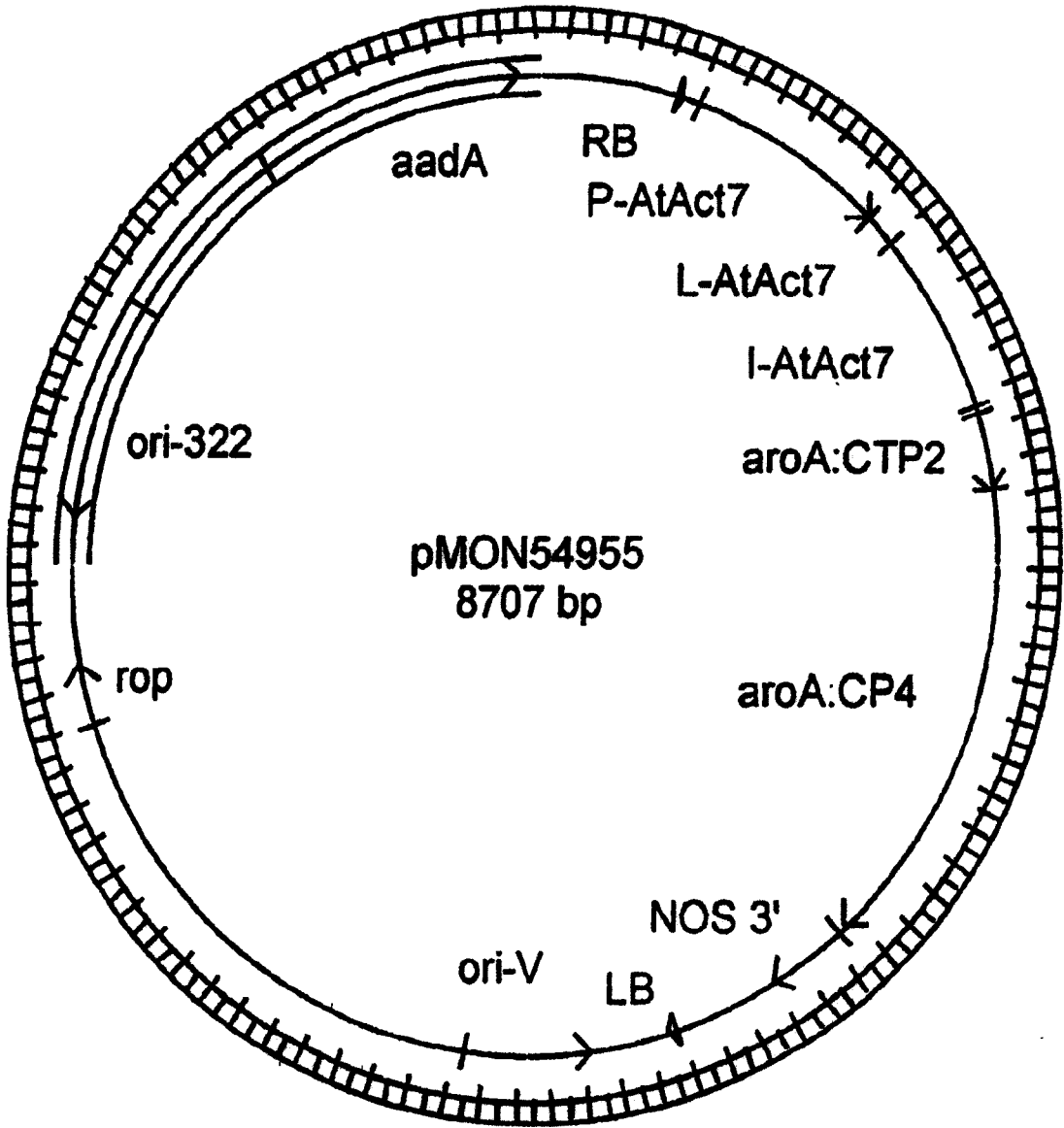


图 17

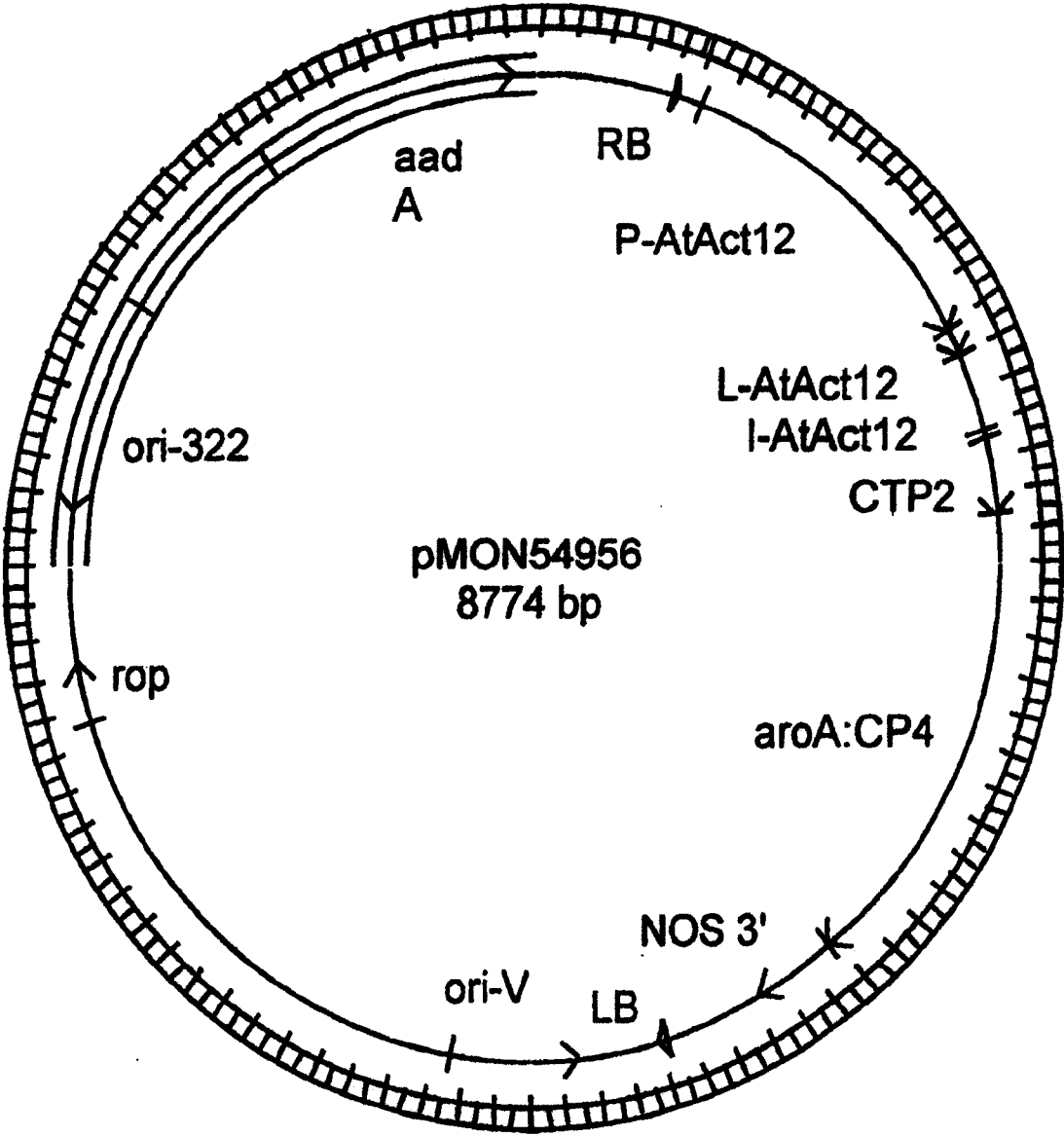


图 18