

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200780043165.5

[51] Int. Cl.

C12N 15/82 (2006.01)

A01H 5/00 (2006.01)

[43] 公开日 2009年11月25日

[11] 公开号 CN 101589147A

[22] 申请日 2007.9.24

[21] 申请号 200780043165.5

[30] 优先权

[32] 2006.9.25 [33] US [31] 60/847,304

[86] 国际申请 PCT/US2007/079268 2007.9.24

[87] 国际公布 WO2008/039709 英 2008.4.3

[85] 进入国家阶段日期 2009.5.25

[71] 申请人 先锋高级育种国际公司

地址 美国依阿华州

[72] 发明人 郭梅 M·鲁普 C·西蒙斯

S·西瓦桑卡

[74] 专利代理机构 中国专利代理(香港)有限公司  
代理人 林毅斌 李连涛

权利要求书3页 说明书77页 序列表53页  
附图9页

[54] 发明名称

改善农作物的植物生长、蒸腾效率和耐旱性的玉米 ERECTA 基因

[57] 摘要

本发明提供 Zm ERECTA 基因家族的多核苷酸和相关多肽。本发明提供 Zm ERECTA 基因的核酸序列。Zm ERECTA 负责控制农作物的植物生长、器官大小和产量。

1. 一种分离的多核苷酸，选自：
  - a. 与选自 SEQ ID NO: 5、7、9、11、13、15、17、19、21、23 和 25 的多核苷酸全长序列有至少 70%序列同一性的多核苷酸，序列同一性用 GAP 算法默认参数计算得出；其中所述多核苷酸编码起器官大小调节物作用的多肽；
  - b. 编码选自 SEQ ID NO: 6、8、10、12、14、16、18、20、22、24 和 26 多肽的多核苷酸，和
  - c. 选自 SEQ ID NO: 5、7、9、11、13、15、17、19、21、23 和 25 的多核苷酸；和
  - d. 与(a)、(b)或(c)的多核苷酸互补的多核苷酸。
2. 一种重组表达盒，所述重组表达盒包含权利要求 1 的多核苷酸，其中所述多核苷酸按有义或反义方向与启动子有效连接。
3. 一种宿主细胞，所述宿主细胞包含权利要求 2 的表达盒。
4. 一种转基因植物，所述植物包含权利要求 2 的重组表达盒。
5. 权利要求 4 的转基因植物，其中所述植物是单子叶植物。
6. 权利要求 4 的转基因植物，其中所述植物是双子叶植物。
7. 权利要求 4 的转基因植物，其中所述植物选自：玉米、大豆、向日葵、高粱、油菜、小麦、苜蓿、棉花、水稻、大麦、粟、花生和可可。
8. 一种转基因种子，所述种子得自权利要求 4 的转基因植物。
9. 一种调节植物器官大小的方法，该方法包括：
  - a. 将包含与启动子有效连接的权利要求 1 的多核苷酸的重组表达盒导入植物细胞；和
  - b. 在植物细胞生长条件下培育植物；其中所述植物细胞中的器官大小受到调节。
10. 权利要求 9 的方法，其中所述植物细胞选自以下植物：玉米、

大豆、向日葵、高粱、油菜、小麦、苜蓿、棉花、水稻、大麦、粟、花生和可可。

11. 一种调节植物的整株植物或器官大小的方法，该方法包括：

a. 将包含与启动子有效连接的权利要求1的多核苷酸的重组表达盒导入植物细胞；

b. 在植物细胞生长条件下培养植物细胞；和

c. 由所述植物细胞再生出植物；其中所述植物的器官大小受到调节。

12. 权利要求11的方法，其中所述植物选自：玉米、大豆、高粱、油菜、小麦、苜蓿、棉花、水稻、大麦、粟、花生和可可。

13. 一种由转基因植物组织加工方法得到的产物，所述植物组织表达编码 ERECTA 基因的分离多核苷酸，该方法包括：

a. 用重组表达盒转化植物细胞，所述表达盒包含与选自 SEQ ID NO: 5、7、9、11、13、15、17、19、21、23 和 25 的多核苷酸全长序列有至少 90% 序列同一性的多核苷酸，所述多核苷酸与启动子有效连接；和

b. 在植物细胞生长条件下培养转化的植物细胞；其中所述转化的植物细胞的生长受到调节；

c. 使植物细胞在植物形成条件下生长从而在植物组织中表达所述多核苷酸；和

d. 加工所述植物组织以获得产物。

14. 权利要求13的产物，其中所述多核苷酸还编码选自 SEQ ID NO: 6、8、10、12、14、16、18、20、22、24 和 26 的多肽。

15. 权利要求13的转基因植物，其中所述植物是单子叶植物。

16. 权利要求13的转基因植物，其中所述植物选自：玉米、大豆、向日葵、高粱、油菜、小麦、苜蓿、棉花、水稻、大麦和粟。

17. 权利要求13的产物，其通过所述多核苷酸的过量表达而改善植物的柄强度。

- 
18. 权利要求 13 的产物，其通过增加生物量而提高产量。
  19. 权利要求 13 的产物，其为乙醇(ethanol)组分。
  20. 一种在干旱胁迫期间调节植物整株植物或器官大小的方法，该方法包括：
    - a. 将包含与启动子有效连接的权利要求 1 的多核苷酸的重组表达盒导入植物细胞；
    - b. 在植物细胞生长条件下培养植物细胞；和
    - c. 在干旱胁迫条件下由所述植物细胞再生出植物；其中所述植物的植物生长受到调节。

## 改善农作物的植物生长、蒸腾效率和耐旱性的玉米 ERECTA 基因

### 发明领域

本发明总的来讲涉及分子生物学领域。

### 发明背景

许多植物的驯化与产量大幅提高有关。发生在自然种群中的大多数表型变异是连续的，是通过多种基因影响实现的。在驯化植物中鉴定出引起产量巨大差异的特定基因，已成为农业研究中的重要焦点。

研究表明，在拟南芥中，ERECTA 基因通过促进细胞增殖来控制器官生长和花的发育(Shpak 等(2003) *Plant Cell* 15:1095-1110; *Development* (2004) 131:1491-501)。拟南芥 ERECTA 基因通过促进细胞增殖从而影响花序发育，控制器官生长。异位过量表达 ERECTA 基因的转基因拟南芥植物通过影响气孔密度、表皮细胞膨胀、叶肉细胞增殖和细胞间接触，从而改善植物蒸腾效率和耐旱性。ERECTA 基因编码富含亮氨酸重复序列受体样激酶(leucine-rich repeat receptor-like kinase, LRR-RLK)，可以控制植物生长/器官大小和生物量累积。另外，Masle Gilmore 和 Farquhar 指出除已知对植物构造的不同作用以外，拟南芥 ERECTA 基因还决定了植物的转化效率(Masle Gilmore 和 Farquhar, *Nature* (2005) 436:866)。这在农业，尤其在耐旱性和农艺性能方面具有意义。

ERECTA 与生长加快有关。ERECTA 基因可用于控制玉米或其它农作物整株植物或特定器官的大小。该基因的潜在用途是在转基因植物中使之过量表达以增加生物量累积，使用组织特异性启动子使该基因的表达靶定在特定组织中从而促进根生长、加快幼苗生长以实现快速的林冠

郁闭、叶片变大、穗大小增加、胚加大、胚乳生长适于较大谷粒和操控整个谷粒中的油、蛋白质或淀粉含量等等。通过改变玉米穗丝生长率，可以操控同步化或 ASI (雌雄穗开花间隔(anthesis and silking interval)), 这可改善胁迫耐受性。另一个潜在的应用是改善得自体外组织培养的作物的转化和再生。可以控制此基因的表达以提高细胞增殖率和培养组织生长率。ERECTA 还可用于通过在特定组织中使表达下调，从而操控基因以减小器官大小，例如雄花穗大小。ERECTA 基因可用于通过改善玉米和其它作物的蒸腾效率，从而提高耐旱性。探索此基因的天然等位基因变异可用于通过鉴定与近交种胁迫抗性表型有关的等位基因单元型，从而改善育种或转基因植物。该基因作图位于染色体的干旱 QTL 附近的位置，这表明可能是抗性等位基因变体。

本发明包括鉴定与拟南芥 ERECTA 基因(SEQ ID NO: 1 和 SEQ ID NO: 3)有关的推定的玉米 ERECTA 基因，即 ZmERECTA A 和 ZmERECTA B (SEQ ID NO: 5 和 SEQ ID NO: 7)。具有与拟南芥 ERECTA (SEQ ID NO: 1)最大相似性的直向同源物(ortholog)是 ZmERECTA 1 (SEQ ID NO: 5)。该基因的表达与未成熟的生殖组织有关，主要见于花序分生组织和茎端分生组织，在其它分生组织相关的组织中水平较低。

预期表达 ZmERECTA A (SEQ ID NO: 5)的转基因植物对生物量累积和玉米植物的生长率以及器官大小增大都有着积极的效果。同样预期表达 ZmERECTA 的转基因植物耐旱性得到改善。这些玉米基因可用于改善玉米(和其它作物)的农艺性状。

本发明还包括鉴定其它植物品种的 ERECTA 基因。水稻基因家族由 2 个家族成员表示。大豆(*Glycine max*)中还发现了 4 个基因序列，两色蜀黍(*Sorghum bicolor*)中发现了 3 个基因。本文公开了 ERECTA 拟南芥基因家族的两个成员。

## 发明概述

本发明提供用于控制植物生长和器官大小以提高植物产量的组合物和方法。所述组合物包括得自玉米、大豆、拟南芥、水稻和高粱的 ERECTA 序列。本发明的组合物包含选自 SEQ ID NO: 5-8 的氨基酸序列和核苷酸序列及其变体和片段。

本发明在 DNA 构建体中提供编码 ERECTA 序列的多核苷酸用于在目标植物中进行表达。本发明还提供包含本发明序列的表达盒、植物、植物细胞、植物组成部分和种子。在具体的实施方案中，所述多核苷酸与组成型启动子有效连接。

本发明提供用于调节植物或植物组成部分的 ERECTA 序列水平的方法。该方法包括将包含本发明 ERECTA 序列的异源多核苷酸导入植物或植物组成部分。可以提高或降低 ERECTA 多肽的水平。这类方法可用来提高植物的产量；在一个实施方案中，所述方法用来提高谷类的谷粒产量。

## 附图简述

图 1: B73、Mo17、B73xMo17 F1 杂种和 Mo17xB73 F1 杂种(MG 6006.42, 44)的穗状花序分生组织(EIM)和茎端分生组织(SAM)在三个阶段的 ZmERECTA-A 的 MPSS 表达。所述基因在穗状花序分生组织中以高水平进行表达，而在茎端分生组织中则以略微较低的水平进行表达。

图 2: 显示共有序列和保守区的得自玉米、拟南芥、水稻、高粱和大豆的 ERECTA 序列的比对。

图 3: ERECTA A 基因(SEQ ID NO: 3)优选在分生组织和未成熟穗中表达，在多种组织中的表达详见图中所示。

图 4: 树形图表示得自拟南芥、玉米、大豆、高粱和水稻的 ERECTA 序列的亲缘关系。

## 发明详述

除非另有说明，否则本文所用的所有科技术语具有本发明所属领域普通技术人员通常所理解的含义。除非另有说明，否则本发明所采用或包括的技术是本领域普通技术人员所熟知的标准方法。材料、方法和实施例只是说明性的，而不是限制性的。下面通过举例进行说明，但无意限制本发明的范围。

下文将参照附图对本发明进行更详尽的说明，其中仅显示本发明的一些实施方案，而非全部实施方案。事实上，这些发明可用许多种不同的方式进行表达，因此不得解释为限于本文所提出的实施方案；然而，这些实施方案应按这样的方式提供，即令本公开内容符合适用的法律要求。在通篇说明书中，近似的数量是指相近的元素。

在获得先前描述和有关附图中所提出的教导内容的益处之后，本发明所属领域技术人员将会想到本文所公开的本发明的许多修改和其它实施方案。因此，要了解的是，本发明不限于所公开的具体实施方案，有关修改和其它实施方案均包括在所附权利要求书的范围内。虽然本文采用特定的术语，但是这些术语只按普通描述性意义使用，而不是用于限制目的。

除非另有说明，否则为了实施本发明将应用植物学、微生物学、组织培养、分子生物学、化学、生物化学和重组 DNA 技术的常规技术，这些技术为本领域技术人员所掌握。这类技术详细记载于文献中。参见例如 Langenheim 和 Thimann, (1982) *Botany: Plant Biology and Its Relation To Human Affairs*, John Wiley; *Cell Culture and Somatic Cell Genetics of Plants* (1984), 第 1 卷, Vasil 主编; Stanier 等(1986) *The Microbial World*, 第 5 版, Prentice-Hall; Dhringra 和 Sinclair, (1985) *Basic Plant Pathology Methods*, CRC Press; Maniatis 等(1982) *Molecular Cloning: A Laboratory*



*Manual; DNA Cloning*, 第 I 卷和第 II 卷, (1985) Glover 主编; *Oligonucleotide Synthesis*, (1985) Gait, 主编; *Nucleic Acid Hybridization*, (1984) Hames 和 Higgins 主编; 以及 *Methods In Enzymology* 系列丛书, Colowick 和 Kaplan 主编, Academic Press, Inc., San Diego, CA.

单位、前缀和符号可以按其 SI 可接受的形式表示。除非另有说明, 否则核酸按 5'到 3'的方向自左向右书写; 氨基酸序列按氨基到羧基的方向自左向右书写。限定范围的数字也包括在数字范围内。本文可按周知的氨基酸三字母符号或由 IUPAC-IUB 生物化学命名委员会(IUPAC-IUB Biochemical Nomenclature Commission)推荐的一字母符号使用氨基酸。同样, 可按公认的核苷酸单字母代码使用核苷酸。下面定义的术语将通过参照整个说明书总体上得到更详细的定义。

为了描述本发明, 将采用下列术语, 并按以下说明来定义。

术语“微生物”是指任何微生物(包括真核和原核微生物), 例如真菌、酵母、细菌、放线菌、藻类和原生动物以及其它单细胞结构。

术语“扩增”是指用至少一个核酸序列作为模板构建多拷贝的核酸序列或与该核酸序列互补的多个拷贝。扩增系统包括聚合酶链式反应(PCR)系统、连接酶链式反应(LCR)系统、基于核酸序列的扩增(NASBA、Cangene、Mississauga、Ontario)、Q $\beta$  复制酶系统、基于转录的扩增系统(TAS)和链置换扩增(SDA)。参见例如 *Diagnostic Molecular Microbiology: Principles and Applications*, (1993) Persing 等主编, American Society for Microbiology, Washington, DC。扩增产物称为扩增子。

术语“保守修饰的变体(conservatively modified variant)”适用于氨基酸序列和核酸序列两者。对于特定核酸序列, 保守修饰的变体是指编码相同氨基酸序列或氨基酸序列保守修饰的变体的核酸。由于遗传密码的简并性, 因此许多功能上相同的核酸都编码任一指定蛋白质。例如, 密码子 GCA、GCC、GCG 和 GCU 全都编码氨基酸丙氨酸。因此, 在丙氨

酸出现的每个位置都用一个密码子标明，该密码子可变成所述相应的任何密码子而又不改变所编码的多肽。这类核酸变异是“沉默变异(silent variation)”，代表一种保守修饰的变异。本文编码多肽的每个核酸序列还记载了核酸的每一个可能的沉默变异。普通技术人员应当了解的是，可以对核酸中的每个密码子(AUG除外，它通常是甲硫氨酸的唯一密码子；一个例外是红色微球菌(*Micrococcus rubens*)，GTG是其甲硫氨酸的密码子(Ishizuka等(1993), *J. Gen. Microbiol.* 139:425-32)进行修饰以产生功能相同的分子。因此，编码本发明多肽的核酸的每种沉默变异在各个所述多肽序列中都是隐含的，并且通过引用结合到本文中。

至于氨基酸序列，技术人员应当了解的是，当改变导致氨基酸被化学上相似的氨基酸取代时，在被编码序列上改变、添加或缺失单个氨基酸或小百分比的氨基酸的核酸、肽、多肽或蛋白质序列的各个取代、缺失或添加，即为“保守修饰的变体”。因此，选自1-15任何整数中的任何数目的氨基酸残基可被如此改变。因此，可产生例如1、2、3、4、5、7或10种变化。保守修饰的变体通常提供与未修饰多肽序列(保守修饰的变体从中衍生)类似的生物活性。例如，对于天然底物，底物特异性、酶活性或配体/受体结合一般是天然蛋白质的至少30%、40%、50%、60%、70%、80%或90%，优选60-90%。提供功能相似的氨基酸的保守取代表是本领域众所周知的。

下面6组分别含有彼此保守取代的氨基酸：

- 1) 丙氨酸(A)、丝氨酸(S)、苏氨酸(T)；
- 2) 天冬氨酸(D)、谷氨酸(E)；
- 3) 天冬酰胺(N)、谷氨酰胺(Q)；
- 4) 精氨酸(R)、赖氨酸(K)；
- 5) 异亮氨酸(I)、亮氨酸(L)、甲硫氨酸(M)、缬氨酸(V)；和
- 6) 苯丙氨酸(F)、酪氨酸(Y)、色氨酸(W)。

另参见 Creighton, *Proteins*, (1984) W.H. Freeman and Co..

本文所用“基本由.....组成”是指包含对于目标多核苷酸而言的附加序列(additional sequence)在内,其中所述附加序列在严格性条件下,不与与所述多核苷酸相同的 cDNA 选择性杂交,其中所述杂交条件包括在 65°C 下在 0.1X SSC 和 0.1%十二烷基硫酸钠中的洗涤步骤。

术语“编码”或“编码的”对于特定核酸而言,是指包含用于翻译成特定蛋白质的信息。蛋白质编码核酸可包含核酸翻译区内的非翻译序列(例如内含子),或者可缺乏这类间插非翻译序列(例如 cDNA 中的间插非翻译序列)。编码蛋白质的信息通过密码子使用而被具体化。通常由使用“通用”遗传密码的核酸编码氨基酸序列。然而,当采用以下这些生物表达核酸时,可以使用通用密码的变体,例如存在于某些植物、动物和真菌线粒体、细菌山羊枝原体(*Mycoplasma capricolum*) (Yamao 等, (1985) *Proc. Natl. Acad. Sci. USA* 82:2306-9)或纤毛虫大核(*Macronucleus*)中的通用密码的变体。

当通过合成来制备或改变核酸时,可利用将在其中表达核酸的目标宿主的已知的密码子偏爱性。例如,虽然本发明的核酸序列可同时在单子叶植物品种和双子叶植物品种两者中表达,但是可根据单子叶植物或双子叶植物对特定密码子的偏爱和 GC 含量的偏爱,来对序列进行修饰,研究表明因为这些偏爱有所不同(Murray 等(1989), *Nucleic Acids Res.* 17:477-98, 该文献通过引用结合到本文中)。因此,可从玉米的已知基因序列,得到特定氨基酸的玉米偏爱的密码子。得自玉米植物的 28 个基因的玉米密码子选择见 Murray 等人文献中的表 4, 出处同上。

本文所用“异源”当指核酸时是起源于外源品种的核酸,或者如果得自同一品种,则是由其天然组成形式和/或基因组基因座通过人为刻意干预的基本修饰形式。例如,与异源结构基因有效连接的启动子得自不同于得到结构基因的品种,或者如果都得自同一品种,则一个或两个是其

原始形式的基本修饰形式。异源蛋白可源自外源品种，或者如果得自同一品种，则它是由其原始形式通过刻意人为干预得到的基本修饰形式。

术语“宿主细胞”是指含有载体并支持表达载体复制和/或表达的细胞。宿主细胞可以是原核细胞(例如大肠杆菌(*E. coli*)), 或者是真核细胞, 例如酵母、昆虫、植物、两栖动物或哺乳动物细胞。优选宿主细胞是单子叶植物或双子叶植物细胞, 包括但不限于玉米、高粱、向日葵、大豆、小麦、苜蓿、水稻、棉花、油菜、大麦、粟和番茄。特别优选的单子叶植物宿主细胞是玉米宿主细胞。

术语“杂交复合物(hybridization complex)”包括有关由 2 条单链核酸序列彼此选择性杂交所形成的双链体核酸结构。

有关将核酸插入细胞方面的术语“导入”, 是指“转染”或“转化”或“转导”, 并且包括有关将核酸掺入真核细胞或原核细胞, 其中核酸可掺到细胞的基因组(例如染色体、质粒、质体或线粒体 DNA)、转化成自主复制子或者进行瞬时表达(例如转染的 mRNA)。

术语“分离的”是指这样的材料, 例如核酸或蛋白质, 它基本或根本不含正常时在其天然存在的环境下伴随存在或者与之相互作用的组分。分离材料任选包含在其天然环境中不与该材料一起存在的材料。作为如本文定义的“分离的”核酸, 亦称“异源”核酸。除非另有说明, 否则术语“ERECT 核酸”是指包含编码 ERECTA 多肽的多核苷酸(“ERECTA 多核苷酸”)的核酸。

本文所用“核酸”是指单链或双链形式的脱氧核糖核苷酸或核糖核苷酸多聚体, 除非特别限定, 否则包括已知的具有天然核苷酸基本性质的类似物, 因为它们按与天然存在的核苷酸(例如肽核酸)同样的方式与单链核酸杂交。

术语“核酸文库”是指一批分离的 DNA 或 RNA 分子, 它包含并基本上代表了特定生物基因组的整个转录部分。示例性核酸文库(例如基因组

文库和 cDNA 文库)的构建可参见标准分子生物学文献, 例如 Berger 和 Kimmel, *Guide To Molecular Cloning Techniques, Methods In Enzymology* 系列丛书, 第 152 卷, Academic Press, Inc., San Diego, CA (1987); Sambrook 等(1989), *Molecular Cloning: A Laboratory Manual*, 第二版, 第 1-3 卷; 以及 *Current Protocols In Molecular Biology*, Ausubel 等主编, Current Protocols, Greene Publishing Associates, Inc.与 John Wiley & Sons, Inc.合资企业(1994 增刊)。

本文所用“有效连接”是指第一序列(例如启动子)和第二序列之间的功能性连接, 其中启动子序列启动和介导相当于第二序列的 DNA 序列的转录。一般而言, 有效连接是指待连接的核酸序列是邻接的, 必要时将 2 个蛋白质编码区相邻连接并在同一读框内。

本文所用术语“植物”是指完整植物、植物器官(例如叶、茎、根等)、种子及植物细胞和它们的子代。本文所用植物细胞包括但不限于种子悬浮培养物、胚、分生组织区、愈伤组织、叶、根、茎干(shoot)、配子体、孢子体、花粉和小孢子。可用于本发明方法的植物类别, 通常与适于转化技术的高等植物的类别一样广泛, 包括单子叶植物和双子叶植物, 包括下列属的品种: 南瓜属(*Cucurbita*)、蔷薇属(*Rosa*)、葡萄属(*Vitis*)、核桃属(*Juglans*)、草莓属(*Fragaria*)、百脉根属(*Lotus*)、苜蓿属(*Medicago*)、驴食豆属(*Onobrychis*)、三叶草属(*Trifolium*)、胡卢巴属(*Trigonella*)、豇豆属(*Vigna*)、柑橘属(*Citrus*)、亚麻属(*Linum*)、老鹳草属(*Geranium*)、木薯属(*Manihot*)、胡萝卜属(*Daucus*)、拟南芥属(*Arabidopsis*)、芸苔属(*Brassica*)、萝卜属(*Raphanus*)、白芥属(*Sinapis*)、颠茄属(*Atropa*)、辣椒属(*Capsicum*)、曼陀罗属(*Datura*)、天仙子属(*Hyoscyamus*)、番茄属(*Lycopersicon*)、烟草属(*Nicotiana*)、茄属(*Solanum*)、碧冬茄属(*Petunia*)、毛地黄属(*Digitalis*)、*Majorana*、菊苣属(*Cichorium*)、向日葵属(*Helianthus*)、莴苣属(*Lactuca*)、雀麦属(*Bromus*)、天门冬属(*Asparagus*)、

金鱼草属(*Antirrhinum*)、萱草属(*Hemerocallis*)、*Nemesis*、天竺葵属(*Pelargonium*)、稷属(*Panicum*)、狼尾草属(*Pennisetum*)、毛茛属(*Ranunculus*)、千里光属(*Senecio*)、*Salpiglossis*、黄瓜属(*Cucumis*)、蓝英花属(*Browallia*)、大豆属(*Glycine*)、豌豆属(*Pisum*)、菜豆属(*Phaseolus*)、黑麦草属(*Lolium*)、稻属(*Oryza*)、燕麦属(*Avena*)、大麦属(*Hordeum*)、黑麦属(*Secale*)、葱属(*Allium*)和小麦属(*Triticum*)。特别优选的植物是玉米(*Zea mays*)。

本文所用的“产量”是指收获时经调整谷类水分(通常为 15%)后每英亩谷类农作物蒲式耳。谷类水分在收获的谷物中测量。收获时调整谷类水分水平后测定经调整的谷类测试重量(test weight),以磅/蒲式耳重量表示。

本文所用“多核苷酸”是指具有天然核糖核苷酸基本性质的脱氧核糖多核苷酸(deoxyribopolynucleotide)、核糖多核苷酸(ribopolynucleotide)或其类似物,因为在严格性杂交条件下,它们与天然存在的核苷酸基本相同的核苷酸序列杂交和/或允许翻译成与天然存在的核苷酸所翻译的氨基酸相同的氨基酸。多核苷酸可以是天然基因、异源结构基因或调节基因的全长序列或亚序列。除非另有说明,否则该术语是指特定序列及其互补序列。因此,出于稳定性或其它原因,骨架经修饰的 DNA 或 RNA 是本文所用术语“多核苷酸”。此外,包含稀有碱基例如肌苷或经修饰的碱基例如三苯基甲基化碱基(仅举两个例子)的 DNA 或 RNA 也是本文所用术语多核苷酸。应当了解的是,可以对 DNA 和 RNA 进行各种修饰,以达到本领域技术人员已知的多种有用目的。本文所采用的术语多核苷酸包括这类化学修饰形式、酶促修饰形式或代谢修饰形式的多核苷酸,以及病毒和细胞(还包括简单细胞和复杂细胞)特有的 DNA 和 RNA 的化学形式。

术语“多肽”、“肽”和“蛋白质”本文中可互换使用,是指氨基酸残基的多聚体。该术语适用于其中一个或多个氨基酸残基是相应天然存在的

氨基酸的人工化学类似物的氨基酸多聚体，并适用于天然存在的氨基酸多聚体。

本文所用的“启动子”是指自转录开始并参与识别和结合 RNA 聚合酶和其它蛋白质以启动转录的 DNA 上游区域。“植物启动子”是能够在植物细胞中启动转录的启动子。示例性植物启动子包括但不限于从植物、植物病毒和包含在植物细胞中表达基因的细菌(如土壤杆菌属(*Agrobacterium*)或根瘤菌属(*Rhizobium*))得到的启动子。实例有在某些组织例如叶、根、种子、须根、木质部导管、管胞或厚壁组织中优先启动转录的启动子。这类启动子被称为“组织优选的”启动子。“细胞型”特异性启动子主要驱动在一种或多种器官中的某些细胞类型(例如根或叶的维管细胞)中进行表达。“诱导型”或“可调节”启动子是受环境控制的启动子。可通过诱导型启动子影响转录的环境条件的实例包括厌氧条件或光线的存在。另一类型的启动子是发育调节启动子，例如在花粉发育期间驱动表达的启动子。组织优选的启动子、细胞型特异性启动子、发育调节启动子和诱导型启动子构成“非组成型”启动子类别。“组成型”启动子是在大多数环境条件下都有活性的启动子。

术语“ERECTA 多肽”是指一个或多个氨基酸序列。该术语还包括氨基酸序列的片段、变体、同源物、等位基因或前体(例如前蛋白质或蛋白质原)。“ERECTA 蛋白”包含 ERECTA 多肽。除非另有说明，否则术语“ERECT 核酸”是指包含编码 ERECTA 多肽的多核苷酸(“ERECTA 多核苷酸”)的核酸。

本文所用“重组”是指通过导入异源核酸而被修饰细胞或载体，或者从如此修饰的细胞衍生的细胞。因此，例如重组细胞表达在天然(非重组)形式的细胞中不会以相同形式存在的基因，或者由于人为刻意干预从而表达否则会异常表达、表达不足或完全不表达的天然基因。本文所用术语“重组”不包括由天然存在的事件(例如自发突变、天然转化/转导/转座)

所引起的细胞或载体变化，例如由无人为刻意干预而发生的细胞或载体变化。

本文所用“重组表达盒”是经重组或合成产生的具有一系列特定核酸元件的核酸构建体，该构建体允许特定核酸在靶细胞中进行转录。可将重组表达盒掺到质粒、染色体、线粒体 DNA、质体 DNA、病毒或核酸片段中。表达载体的重组表达盒部分除其它序列以外，通常还包括待转录的核酸和启动子。

术语“残基”或“氨基酸残基”或“氨基酸”在本文中可互换使用，是指掺到蛋白质、多肽或肽(总称“蛋白质”)的氨基酸。氨基酸可以是天然存在的氨基酸，除非另有限定，否则可包括已知的可按与天然存在的氨基酸同样的方式起作用的天然氨基酸的类似物。

术语“选择性杂交”是指在严格性杂交条件下，比起其与非靶核酸序列杂交的程度，核酸序列以可检测的较高程度(例如超过本底的至少 2 倍)与特定核酸靶序列杂交，并且基本不包括非靶核酸。选择性杂交序列彼此间通常具有约至少 40%序列同一性，优选 60-90%序列同一性，最优选 100%序列同一性(即互补)。

术语“严格性条件”或“严格性杂交条件”是指比起其它序列，探针可与其靶序列以可检测的较高程度(例如超出本底的至少 2 倍)杂交的条件。严格性条件是序列依赖性的，在不同情况下将有所不同。通过控制杂交和/或洗涤条件的严格性，可鉴定出与探针互补可高达 100%的靶序列(同源探测)。或者，可调整严格性条件，使得序列中存在一些错配，从而检测较低程度的相似性(异源探测)。探针长度最理想约为 500 个核苷酸，但是也可在长度 500 个核苷酸以下到等于靶序列完整长度之间大幅变化。

严格性条件通常是这样的条件，其中在 pH 7.0-8.3 下盐浓度小于约 1.5 M 钠离子，通常约 0.01-1.0 M 钠离子浓度(或其它盐)，对于短探针(例



如 10-50 个核苷酸), 温度至少约为 30°C, 对于长探针(例如大于 50 个核苷酸)至少约为 60°C。还可加入去稳定剂例如甲酰胺或 Denhardt 试剂来达到严格性条件。示例性的低严格性条件包括在 37°C 下在 30-35%甲酰胺、1 M NaCl、1% SDS (十二烷基硫酸钠)的缓冲溶液中进行杂交, 并在 50-55°C 下用 1X-2X SSC (20X SSC = 3.0 M NaCl/0.3 M 柠檬酸三钠)洗涤。示例性的中等严格性条件包括在 37°C 下在 40-45%甲酰胺、1 M NaCl、1% SDS 中进行杂交, 并在 55-60°C 下用 0.5X-1X SSC 洗涤。示例性的高严格性条件包括在 37°C 下在 50%甲酰胺、1 M NaCl、1% SDS 中进行杂交, 并在 60-65°C 下用 0.1X SSC 洗涤。特异性通常随杂交后洗涤而变化, 关键因素是最终洗涤溶液的离子强度和温度。对于 DNA-DNA 杂交, 可由下列公式得到  $T_m$  的近似值:  $T_m = 81.5^\circ\text{C} + 16.6 (\log M) + 0.41 (\%GC) - 0.61 (\% \text{ form}) - 500/L$  (Meinkoth 和 Wahl, *Anal. Biochem.*, (1984) 138:267-84); 其中 M 是一价阳离子的体积摩尔浓度, % GC 是 DNA 中鸟嘌呤核苷酸和胞嘧啶核苷酸的百分比, % form 是杂交溶液中甲酰胺的百分比, L 是以碱基对为单位的杂交序列长度。 $T_m$  是互补靶序列与完全匹配的探针杂交达 50%时的温度(在规定的离子强度和 pH 下)。对于每 1%的错配,  $T_m$  便降低大约 1°C; 因此, 可以调整  $T_m$ 、杂交和/或洗涤条件从而与所需同一性的序列杂交。例如, 如果要寻获同一性  $\geq 90\%$  的序列, 则可将  $T_m$  降低 10°C。一般选择在规定离子强度和 pH 下低于特定序列及其互补序列热熔点( $T_m$ )约 5°C 的严格性条件。然而, 极严格性条件可采用比热熔点( $T_m$ )低 1°C、2°C、3°C 或 4°C 的杂交和/或洗涤; 中等严格性条件可采用比热熔点( $T_m$ )低 6°C、7°C、8°C、9°C 或 10°C 的杂交和/或洗涤; 低严格性条件可采用比热熔点( $T_m$ )低 11°C、12°C、13°C、14°C、15°C 或 20°C 的杂交和/或洗涤。在使用所述公式、杂交和洗涤组合物及所需的  $T_m$  时, 普通技术人员应清楚杂交和/或洗涤溶液的变化是固有发生的。如果所需的错配程度引起  $T_m$  低于 45°C (含水溶液)或 32°C (甲酰胺溶液), 则优选

增加 SSC 浓度以便可使用较高的温度。核酸杂交的详尽指南可参见 Tijssen, *Biology - Hybridization with Nucleic Acid Probes*, 第 2 章, 第 I 部分, “Overview of principles of hybridization and the strategy of nucleic acid probe assays (杂交原理与核酸探针实验策略综述)”, Elsevier, New York (1993); 以及 *Current Protocols In Molecular Biology*, 第 2 章, Ausubel 等主编, Greene Publishing and Wiley-Interscience, New York (1995)。除非另有说明, 否则本申请中高严格性定义为在 65°C 下在 4X SSC、5X Denhardt 试剂(含 5 g 菲可(Ficoll)、5 g 聚乙烯吡咯烷酮、5 g 胎牛血清白蛋白的 500ml 水)、0.1 mg/ml 煮沸的鲑精 DNA 和 25 mM 磷酸钠中进行杂交, 并在 65°C 下用 0.1X SSC、0.1% SDS 洗涤。

本文所用“转基因植物”是指在其基因组中包含异源多核苷酸的植物。异源多核苷酸一般稳定地整合到基因组中, 使得多核苷酸传递给连续世代。异源多核苷酸可单独或者作为重组表达盒的组成部分整合到基因组中。本文所用“转基因的”包括任何细胞、细胞系、愈伤组织、组织、植物组成部分或植物, 其基因型由于异源核酸的存在而发生了改变, 所述异源核酸包括最初如此改变的转基因植物的异源核酸, 以及由最初的转基因植物经有性杂交或无性繁殖所产生的异源核酸。本文所用术语“转基因的”不包括通过常规植物育种方法或者由天然存在的事件(例如随机异花受精、非重组病毒感染、非重组细菌转化、非重组转座或自发突变)引起的基因组(染色体或染色体外)改变。

本文所用“载体”是指用于转染宿主细胞的核酸, 并且在载体中可插入多核苷酸。载体常常是复制子。表达载体允许插入其中的核酸进行转录。

使用下列术语来描述两个或更多个核酸或多核苷酸或者多肽之间的序列关系: (a) “参比序列”, (b) “比较窗(comparison window)”, (c) “序列同一性”, (d) “序列同一性百分比”, 和(e) “基本相同”。

本文所用“参比序列”是用作序列比较基础的规定序列。参比序列可以是特定序列的亚序列或全序列；例如作为全长 cDNA 或基因序列的区段，或者是完整的 cDNA 或基因序列。

本文所用“比较窗”是指包括多核苷酸序列连续的特定区段，其中可将多核苷酸序列与参比序列进行比较，且其中与用于两个序列的最佳比对的参比序列(其不包含添加或缺失)相比，比较窗的多核苷酸序列部分可包含添加或缺失(即空位)。比较窗长度一般至少为 20 个连续核苷酸，任选可以是 30、40、50、100 个或更长。本领域技术人员清楚，为了避免由于多核苷酸序列含有空位所致的与参比序列有较高的相似性，通常会引入空位罚分并且从匹配数中扣减。

用于比较的核苷酸序列和氨基酸序列的比对方法是本领域众所周知的。局部同源算法(BESTFIT) (见 Smith 和 Waterman, (1981) *Adv. Appl. Math* 2:482), 可进行最佳序列比对以便比较; 还可通过以下方法进行: 同源比对算法(GAP) (Needleman 和 Wunsch, (1970) *J. Mol. Biol.* 48:443-53); 检索相似性方法(Tfasta 和 Fasta) (Pearson 和 Lipman, (1988) *Proc. Natl. Acad. Sci. USA* 85:2444); 这些算法的计算机运行, 包括但不限于 PC/Gene 程序中的 CLUSTAL (Intelligenetics, Mountain View, California)、GAP、BESTFIT、BLAST、FASTA 和 Wisconsin 遗传软件包第 8 版(Wisconsin Genetics Software Package)的 TFASTA (可得自 Genetics Computer Group (GCG<sup>®</sup> 程序 (Accelrys, Inc., San Diego, CA))。有关 CLUSTAL 程序的详情参见 Higgins 和 Sharp, (1988) *Gene* 73:237-44; Higgins 和 Sharp, (1989) *CABIOS* 5:151-3; Corpet 等(1988) *Nucleic Acids Res.* 16:10881-90; Huang 等(1992) *Computer Applications in the Biosciences* 8:155-65; 以及 Pearson 等(1994) *Meth. Mol. Biol.* 24:307-31。用于多个序列的最佳全局比对的优选程序是 PileUp (Feng 和 Doolittle, (1987) *J. Mol. Evol.*, 25:351-60, 该方法与 Higgins 和 Sharp 所述方法(Higgins 和 Sharp

(1989) *CABIOS* 5:151-53)类似, 并通过引用结合到本文中)。可用于数据库相似性检索的一类 BLAST 程序包括: 对照核苷酸数据库序列用于核苷酸查询序列的 BLASTN; 对照蛋白质数据库序列用于核苷酸查询序列的 BLASTX; 对照蛋白质数据库序列用于蛋白质查询序列的 BLASTP; 对照核苷酸数据库序列用于蛋白质查询序列的 TBLASTN; 以及对照核苷酸数据库序列用于核苷酸查询序列的 TBLASTX。参见 *CURRENT PROTOCOLS IN MOLECULAR BIOLOGY*, 第 19 章, Ausubel 等主编, Greene Publishing and Wiley-Interscience, New York (1995)。

GAP 使用 Needleman 和 Wunsch 的算法(出处同上), 来查找使匹配数最大且空位数最小的 2 个全序列的比对。GAP 将所有可能的比对和空位都考虑在内, 创建具有碱基匹配数最大和空位最少的比对。这允许提供以匹配碱基为单位的空位建立罚分(gap creation penalty)和空位延伸罚分(gap extension penalty)。对于其插入的每个空位, GAP 都必须得到(make a profit of)匹配的空位建立罚分数。如果选择空位延伸罚分大于零, 则 GAP 另还必须因每个插入的空位而得到空位长度乘以空位延伸罚分的值。Wisconsin 遗传软件包(第 10 版)中默认空位建立罚分值和空位延伸罚分值分别为 8 和 2。空位建立罚分和空位延伸罚分可用选自 0-100 之间的整数表示。因此, 例如空位建立罚分和空位延伸罚分可以为 0、1、2、3、4、5、6、7、8、9、10、15、20、30、40、50 或以上。

GAP 是最佳比对类别的一员。该类别中可有许多成员, 但其它成员并不具有较好的质量。GAP 显示出比对的 4 项品质因素: 质量(quality)、比率、同一性和相似性。质量是为了对序列进行比对的量度(metric)最大化。比率是质量除以较短区段的碱基数。百分比同一性是实际匹配符号的百分比。百分比相似性是相似符号的百分比。空位跨过的符号可忽略不计。当一对符号的评分矩阵值(scoring matrix value)  $\geq 0.50$  (即相似性阈值)时, 被评为有相似性。Wisconsin 遗传软件包(第 10 版)中所用的评分

矩阵是 BLOSUM62 (参见 Henikoff 和 Henikoff, (1989) *Proc. Natl. Acad. Sci. USA* 89:10915)。

除非另有说明, 否则本文所提供的序列同一性/相似性数值是指应用 BLAST 2.0 程序组采用默认参数所得到的分值(Altschul 等(1997) *Nucleic Acids Res.* 25:3389-402)。

正如本领域普通技术人员所了解的一样, BLAST 假定蛋白质可作为随机序列模型来进行检索。然而, 许多真实蛋白质包含非随机序列区域, 它可以是同聚束(homopolymeric tract)、短周期重复序列(short-period repeat)或富含一种或多种氨基酸的区域。可对无关蛋白质之间的这类低复杂性区域进行比对, 即使蛋白质的其它区域完全不同。可采用多种低复杂性过滤程序(filter program)以减少这类低复杂性比对。可单独或联合使用例如 SEG (Wooten 和 Federhen, (1993) *Comput. Chem.* 17:149-63)和 XNU (Claverie 和 States, (1993) *Comput. Chem.* 17:191-201)低复杂性过滤程序。

本文所用“序列同一性”或“同一性”在两个核酸序列或两个多肽序列的情况下, 是指两个序列中的残基, 当在特定比较窗内对最大对应性进行比对时是相同的。当序列同一性百分比用来提及蛋白质时, 一般认为不相同的残基位置常常因保守氨基酸取代而不同, 其中氨基酸残基被具有相似化学性质(例如电荷或疏水性)的其它氨基酸残基取代, 因此不会改变分子的功能性质。如果序列在保守取代方面不同, 则百分比序列同一性可以向上调整以纠正保守性质的取代。由于这类保守取代而不同的序列, 被称为具有“序列相似性”或“相似性”。进行这类调整的方法为本领域技术人员所熟知。这通常包括将保守取代作为部分错配而不是完全错配进行评分, 从而提高百分比序列同一性。因此, 例如当相同的氨基酸得分为 1, 且非保守取代得分为 0 时, 保守取代的得分介于 0 和 1 之间。例如, 按照 Meyers 和 Miller 算法(Meyers 和 Miller, (1988) *Computer*

*Applic. Biol. Sci.* 4:11-17), 例如执行程序 PC/GENE (Intelligenetics, Mountain View, California, USA), 计算出保守取代的得分。

本文所用“序列同一性百分比”是指在比较窗内对两个最佳比对序列进行比较所测定的值, 其中与用于两个序列最佳比对的参比序列(不包含添加或缺失)相比, 比较窗内的多核苷酸序列部分可包含添加或缺失(即空位)。如下计算百分比: 确定同时在两个序列中出现相同核酸碱基或氨基酸残基的位置数, 得到匹配位置数, 用比较窗的位置总数除匹配位置数, 结果乘以 100 便得到序列同一性的百分比。

术语多核苷酸序列的“基本相同”是指包含与参比序列相比, 序列同一性介于 50-100 %之间的序列的多核苷酸, 优选至少 50%序列同一性, 优选至少 60%序列同一性, 优选至少 70%, 更优选至少 80%, 更优选至少 90%, 最优选至少 95%, 所述序列同一性通过应用所述比对程序之一用标准参数计算得出。技术人员应当了解的是, 可以将密码子简并性、氨基酸相似性、读框定位等考虑在内, 对这些数值进行适当调整, 以确定由两个核苷酸序列编码的相应蛋白质的同一性。出于这些目的的氨基酸序列的基本相同通常是指介于 55-100%之间的序列同一性, 优选至少 55%, 优选至少 60%, 更优选至少 70%、80%、90%, 最优选至少 95%。

核苷酸序列是基本相同的另一个标志是如果两个分子在严格性条件下能彼此杂交。遗传密码的简并性为许多氨基酸取代创造了条件, 该取代导致了编码相同氨基酸的核苷酸序列的多样性, 因此, 有可能出现的是, 可编码相同多肽的 DNA 序列在严格性条件下彼此间不会杂交。例如, 当用遗传密码所允许的最大密码子简并产生核酸拷贝时, 这种情况便可能出现。2 个核酸序列是基本相同的一个标志是这样的多肽, 即第一核酸编码的多肽与第二核酸的多肽有免疫交叉反应性。

术语“基本相同”在肽的情况下是指肽包含在特定比较窗内与参比序列相比序列同一性介于 55-100%的序列, 优选至少 55%序列同一性, 优

选 60%、优选 70%、更优选 80%、最优选至少 90%或 95%序列同一性。优选采用 Needleman 和 Wunsch 的同源比对算法(出处同上)进行最佳比对。两个肽序列是基本相同的标志是一种肽与针对第二种肽所产生的抗体有免疫反应性。因此,例如当两种肽仅因保守取代而不同时,第一肽与第二肽是基本相同的。另外,当两种肽因非保守变化而不同时,如果抗体识别的表位是基本相同的,则第一肽与第二肽是基本相同的。与如上所述序列有共同序列只是残基位置有所不同的“基本相似”的肽,可因保守氨基酸改变而有所不同。

本发明公开了 ERECTA 多核苷酸和 ERECTA 多肽。本发明的新的核苷酸和蛋白质具有的表达模式表明它们在植物发育中起着重要作用。所述多核苷酸在多种植物组织中进行表达。因此,所述多核苷酸和多肽为操控植物发育以改变种子和营养组织发育、时间控制或组成提供了机会。这可用于产生不育植物、无籽植物或胚乳组成改变的植物。

表1 序列鉴定

名称	品种	SEQ ID NO: 多核苷酸	SEQ ID NO: 多肽	SEQ ID NO: 多核苷酸可读框
AtERECTA	拟南芥 ( <i>Arabidopsis thaliana</i> )	1	2	27
AtERECTA 样	拟南芥	3	4	28
ZmERECTA A	玉米	5	6	29
ZmERECTA B	玉米	7	8	30
OsERECTA A	水稻( <i>Oryza sativa</i> )	9	10	31
OsERECTA B	水稻	11	12	32
SbERECTA A	两色蜀黍 ( <i>Sorghum bicolor</i> )	13	14	33
SbERECTA B	两色蜀黍	15	16	34
SbERECTA C	两色蜀黍	17	18	35
GmERECTA A	大豆 ( <i>Glycine max</i> )	19	20	36
GmERECTA B	大豆	21	22	37
GmERECTA C	大豆	23	24	38
GmERECTA D	大豆	25	26	39

### 核酸

本发明此外还提供包含 ERECTA 多核苷酸的 RNA、DNA 及其类似物和/或嵌合体的分离核酸。

本发明还包括使在不同生物中表达最优化的多核苷酸。例如，对于



玉米植物中多核苷酸的表达,可按照 Murray 等的方法(出处同上),使序列改变以符合特定的密码子偏爱,并改变 GC 含量。对于得自玉米植物 28 个基因的玉米密码子选择可参见 Murray 等人(出处同上)的表 4。

本发明的 ERECT 核酸包含分离的 ERECTA 多核苷酸,其包括:

- (a) 编码 ERECTA 多肽的多核苷酸及其保守性修饰的变体和多态性变体;
- (b) 与(a)或(b)的多核苷酸具有至少 70%序列同一性的多核苷酸;
- (c) (a)或(b)的多核苷酸的互补序列。

### 核酸的构建

可采用(a)标准重组方法,(b)合成技术,或其组合,来制备本发明的分离核酸。在一些实施方案中,本发明的多核苷酸可被克隆、扩增,或者另由真菌或细菌构建。

核酸可适宜包含除本发明多核苷酸以外的序列。例如,可将包含一个或多个内切核酸酶限制位点的多克隆位点插入核酸中有助于分离多核苷酸。同样,可以插入可翻译序列有助于分离翻译的本发明多核苷酸。例如,六聚组氨酸标记序列为纯化本发明的蛋白质提供方便的方法。本发明的上述核酸-多核苷酸序列除外-任选为用于本发明多核苷酸克隆和/或表达的载体、衔接子或接头。可将附加序列添加到这类克隆和/或表达序列上以使其在克隆和/或表达时最优化,从而有助于分离多核苷酸,或者促进多核苷酸导入细胞。本发明核酸的长度减去其本发明多核苷酸的长度通常小于 20 千碱基对,常常小于 15 kb,时常小于 10 kb。克隆载体、表达载体、衔接子和接头的使用是本领域众所周知的。示例性的核酸包括诸如这样的载体: M13、λZAP Express、λZAP II、λgt10、λgt11、pBK-CMV、pBK-RSV、pBluescript II、λDASH II、λEMBL 3、λEMBL 4、pWE15、SuperCos 1、SurfZap、Uni-ZAP、pBC、pBS+/-、pSG5、pBK、

pCR-Script、pET、pSPUTK、p3'SS、pGEM、pSK+/-、pGEX、pSPORTI 和 II、pOPRSVI CAT、pOPI3 CAT、pXT1、pSG5、pPbac、pMbac、pMC1neo、pOG44、pOG45、pFRT $\beta$ GAL、pNEO $\beta$ GAL、pRS403、pRS404、pRS405、pRS406、pRS413、pRS414、pRS415、pRS416、 $\lambda$ MOSSlox 和  $\lambda$ MOSElox。用于本发明的任选的载体包括但不限于  $\lambda$ ZAP II 和 pGEX。有关各种核酸的描述参见例如 Stratagene Cloning Systems, 1995、1996、1997 目录(La Jolla, CA); 以及 Amersham Life Sciences, Inc, '97 目录(Arlington Heights, IL)。

### 构建核酸的合成方法

可用以下方法通过直接化学合成来制备本发明的分离核酸: 例如磷酸三酯方法(Narang 等(1979) *Meth. Enzymol.* 68:90-9); 磷酸二酯方法(Brown 等, (1979) *Meth. Enzymol.* 68:109-51); 二乙基亚磷酰胺(diethylphosphoramidite) 方法(Beaucage 等(1981) *Tetra. Letts.* 22(20):1859-62); 固相亚磷酰胺三酯方法(Beaucage 等(出处同上), 该方法例如使用自动合成仪, 参见例如 Needham-VanDevanter 等(1984) *Nucleic Acids Res.* 12:6159-68); 以及固相支持体方法(美国专利第 4,458,066 号)。化学合成法一般产生单链寡核苷酸。这可通过与互补序列杂交或者通过用该单链作为模板用 DNA 聚合酶聚合来转化成双链 DNA。技术人员应当了解的是, 虽然 DNA 的化学合成限于约 100 个碱基的序列, 但是可通过较短序列的连接得到较长的序列。

### UTR 和密码子偏爱

研究发现翻译效率一般受 RNA 的 5'非编码区或非翻译区(5'UTR) 特定序列元件的调节。正向序列基序包括翻译起始共有序列(Kozak, (1987) *Nucleic Acids Res.* 15:8125)和 5'G<sup>+</sup>7 甲基 GpppG RNA 帽子结构

(Drummond 等(1985) *Nucleic Acids Res.* 13:7375)。负向元件包括稳定的分子内 5'UTR 茎 - 环结构(Muesing 等(1987) *Cell* 48:691)和 AUG 序列或位于 5'UTR 中合适 AUG 之后的短可读框(Kozak, 出处同上, Rao 等(1988) *Mol. and Cell. Biol.* 8:284)。因此, 本发明提供 5'和/或 3'UTR 区用于调节异源编码序列的翻译。

此外, 可以对本发明多核苷酸的多肽编码区段进行修饰以改变密码子选择。可利用已改变的密码子选择来改变翻译效率和/或使编码序列在所需宿主中的表达最优化, 或者使在玉米中表达的异源序列中的密码子选择最优化。可以应用获自 University of Wisconsin Genetics Computer Group 的市售软件包例如“Codon Preference”, 对本发明多核苷酸编码区中的密码子选择进行统计分析。参见 Devereaux 等(1984) *Nucleic Acids Res.* 12:387-395; 或 MacVector 4.1 (Eastman Kodak Co., New Haven, Conn.)。因此, 本发明提供至少一个本发明多核苷酸的编码区密码子选择频率特征。可以用来确定密码子选择频率的多核苷酸数(3 个核苷酸/氨基酸), 可以是 3 个至本文所提供的本发明多核苷酸数之间的任何整数。任选多核苷酸可以是全长序列。用于统计分析的示例性序列数可为至少 1、5、10、20、50 或 100。

### 序列改组

本发明提供使用本发明的多核苷酸进行序列改组的方法以及从其中得到的组合物。序列改组可参见 WO 公布号 97/20078。另参见 Zhang 等(1997), *Proc. Natl. Acad. Sci. USA* 94:4504-9; 以及 Zhao 等(1998) *Nature Biotech* 16:258-61。序列改组一般提供用于产生具有所需特征的多核苷酸文库的方法, 所述特征是可被选择或筛选的。重组多核苷酸文库由一群序列相关的多核苷酸产生, 所述序列相关的多核苷酸包含具有基本序列同一性并可在体外或体内进行同源重组的序列区。所述序列重组的多核

核苷酸群包含多核苷酸亚群，该多核苷酸亚群具有所需要的特征或优势特征，并且可通过合适的选择或筛选方法被选出。所述特征可以是能够在筛选系统中被选出或被检测的性质或属性，可包括以下蛋白、元件或序列的性质：编码的蛋白质；转录元件；控制转录、RNA 加工、RNA 稳定性、染色质构象、翻译或基因或转基因的其它表达性质的序列；复制元件；蛋白质结合元件等，例如赋予可筛选或可检测性质的任何特征。在一些实施方案中，选择特征可以是与如本文所提供的野生型蛋白质相比  $K_m$  和/或  $K_{cat}$  改变。在其它实施方案中，由序列改组产生的蛋白质或多核苷酸的配体结合亲和力可大于未改组野生型多核苷酸的配体结合亲和力。在又一些实施方案中，与未改组野生型多核苷酸的相比，由序列改组产生的蛋白质或多核苷酸最适宜 pH 被改变。这类性质的提高可以是野生型数值的至少 110%、120%、130%、140%或 150%以上。

### 重组表达盒

本发明还提供包含本发明核酸的重组表达盒。编码所需的本发明多核苷酸的核酸序列(例如长度足以编码本发明活性蛋白的 cDNA 或编码多肽的基因组序列)，可用来构建可导入所需宿主细胞的重组表达盒。重组表达盒通常可包含与转录起始调节序列有效连接的本发明多核苷酸，该转录起始调节序列可指导多核苷酸在既定宿主细胞(例如转化植物的组织)中转录。

例如，植物表达载体可包括：(1)受 5'和 3'调节序列转录控制的克隆的植物基因，(2)显性选择标记。如有需要，这类植物表达载体还可包含启动子调节区(例如赋予诱导型或组成型表达、环境或发育调节性表达或者细胞或组织特异性/选择性表达的启动子调节区)、转录起始位点、核糖体结合位点、RNA 加工信号、转录终止位点和/或聚腺苷酸化信号。

可以采用植物启动子片段，该片段可指导本发明多核苷酸在再生植

物的所有组织中表达。这类启动子本文称为“组成型”启动子，并且在大多数环境条件和发育或细胞分化状态下都具有活性。组成型启动子的实例包括 1'启动子或 2'启动子(衍生自根癌土壤杆菌(*Agrobacterium tumefaciens*)的 T-DNA)、Smas 启动子、肉桂醇脱氢酶启动子(美国专利第 5,683,439 号)、Nos 启动子、核酮糖二磷酸羧化酶-加氧酶启动子、GRP1-8 启动子、得自花椰菜花叶病毒(CaMV)的 35S 启动子(参见 Odell 等(1985) *Nature* 313:810-2)、水稻肌动蛋白启动子(McElroy 等(1990) *Plant Cell* 163-171)、泛素启动子(Christensen 等(1992) *Plant Mol. Biol.* 12:619-632 和 Christensen 等(1992) *Plant Mol. Biol.* 18:675-89)、pEMU (Last 等(1991) *Theor. Appl. Genet.* 81:581-8)、MAS (Velten 等(1984) *EMBO J.* 3:2723-30)、玉米 H3 组蛋白启动子(Lepetit 等(1992) *Mol. Gen. Genet.* 231:276-85)和 Atanassova 等(1992) *Plant Journal* 2(3):291-300)、ALS 启动子(参见 PCT 专利申请号 WO 96/30530、GOS2 (美国专利第 6,504,083 号)以及得自本领域技术人员已知的各种植物基因的其他转录起始区。对于本发明而言，泛素启动子是优选的用于在单子叶植物中表达的启动子。

或者，植物启动子可指导本发明的多核苷酸在特定组织中表达，或者另可处于更精细的环境或发育控制之下。这类启动子在此称为“诱导型”启动子(Rab17、RAD29)。可影响通过诱导型启动子转录的环境条件包括病原体攻击、厌氧条件或光线的存在。诱导型启动子的实例有由缺氧或冷胁迫诱导的 Adh1 启动子、由热胁迫诱导的 Hsp70 启动子和由光诱导的 PPDK 启动子。

受发育控制的启动子的实例包括仅仅或优先在某些组织(例如叶、根、果实、种子或花)中启动转录的启动子。启动子的操作还可随启动子在基因组中的位置而变化。因此，诱导型启动子在某些位置上可能变成完全组成型或部分组成型。

如果需要多肽表达，则一般需要的是在多核苷酸编码区的 3'端包括

聚腺苷酸化区。可从多种植物基因或者从 T-DNA 中得到聚腺苷酸化区。待添加的 3'端序列可得自例如胭脂碱合酶或章鱼碱合酶基因, 或者得自其它植物基因, 或者次优选得自任何其它真核基因。这类调节元件的实例包括但不限于 3'终止区和/或聚腺苷酸化区, 例如根癌土壤杆菌胭脂碱合酶(nos)基因的 3'终止区和/或聚腺苷酸化区(Bevan 等(1983) *Nucleic Acids Res.* 12:369-85); 马铃薯蛋白酶抑制剂 II (PINII)基因(Keil 等(1986) *Nucleic Acids Res.* 14:5641-50; An 等(1989) *Plant Cell* 1:115-22); 以及 CaMV 19S 基因(Mogen 等(1990) *Plant Cell* 2:1261-72)。

可将内含子序列添加到 5'非翻译区或部分编码序列的编码序列中以增加在细胞溶胶积蓄的成熟信息的量。在植物和动物表达构建体两者的转录单元中包括可剪接内含子显示出在 mRNA 和蛋白质水平上的基因表达提高多达 1000 倍(Buchman 和 Berg, (1988) *Mol. Cell Biol.* 8:4395-4405; Callis 等(1987) *Genes Dev.* 1:1183-200)。这类内含子增强的基因表达通常在置于接近转录单元 5'端时最高。玉米内含子 Adh1-S 内含子 1、2 和 6 及 Bronze-1 内含子的使用是本领域已知的。一般参见 *The Maize Handbook*, 第 116 章, Freeling 和 Walbot 主编, Springer, New York (1994)。

植物信号序列可用于本发明, 所述植物信号序列包括但不限于使蛋白质靶向植物细胞胞外基质的信号肽编码 DNA/RNA 序列(Drathewka-Kos 等(1989) *J. Biol. Chem.* 264:4896-900), 例如白花丹叶烟草(*Nicotiana glauca*)扩延基因(DeLoose 等(1991)基因 99:95-100); 使蛋白质靶向液泡的信号肽, 例如甘薯块根贮存蛋白(sporamin)基因(Matsuka 等(1991) *Proc. Natl. Acad. Sci. USA* 88:834)和大麦凝集素基因(Wilkins 等(1990) *Plant Cell*, 2:301-13); 引起蛋白质分泌的信号肽, 例如 PR1b 信号肽(Lind 等(1992) *Plant Mol. Biol.* 18:47-53)或大麦  $\alpha$  淀粉酶(BAA) (Rahmatullah 等(1989) *Plant Mol. Biol.* 12:119, 该文献通过引用结合到本文中); 或者使蛋白质靶向质体的信号肽, 例如油菜籽烯酰 Acp 还原酶的信号肽

(Verwaert 等(1994) *Plant Mol. Biol.* 26:189-202)。对于本发明,与 ERECTA 多核苷酸融合的大麦  $\alpha$  淀粉酶信号序列是用于在玉米中表达的优选的构建体。

包含得自本发明多核苷酸的序列的载体通常包含标记基因,它赋予植物细胞可选择的表型。选择标记基因一般可通过合适的基因编码抗生素抗性,合适的基因包括抗生素大观霉素抗性编码基因(例如 *aada* 基因)、编码链霉素抗性的链霉素磷酸转移酶(SPT)基因、编码卡那霉素或遗传霉素抗性的新霉素磷酸转移酶(NPTII)基因、编码潮霉素抗性的潮霉素磷酸转移酶(HPT)基因、起抑制乙酰乳酸合酶(ALS)、特别是抑制磺酰脲型除草剂作用的除草剂抗性的编码基因(例如因含有导致这种抗性的突变、特别是 S4 和/或 Hra 突变的乙酰乳酸合酶(ALS)基因)、起抑制谷氨酰胺合酶作用的除草剂(例如膦丝菌素或 *basta*)抗性的编码基因(例如 *bar* 基因)或者本领域已知的其它这类基因。*bar* 基因编码除草剂 *basta* 抗性, ALS 基因编码除草剂氯磺隆抗性。

用于在高等植物中进行基因表达的典型载体是本领域众所周知的,包括得自根癌土壤杆菌诱导根癌的(Ti)质粒的载体,参见 Rogers 等(1987) *Meth. Enzymol.* 153:253-77。这些载体是植物整合性载体,因为在转化时,载体将部分载体 DNA 整合到宿主植物的基因组中。本文可用的示例性根癌土壤杆菌载体是质粒 pKYLX6 和 pKYLX7 (Schardl 等(1987) *Gene* 61:1-11 和 Berger 等(1989) *Proc. Natl. Acad. Sci. USA*, 86:8402-6)。本文另一个可用的载体是可获自 CLONTECH Laboratories, Inc. (Palo Alto, CA) 的质粒 pBI101.2。

### 宿主细胞中蛋白质的表达

采用本发明的核酸,可以在重组改造的细胞例如细菌细胞、酵母细胞、昆虫细胞、哺乳动物细胞或者优选植物细胞中表达本发明的蛋白质。

细胞以非天然状况(例如数量、组成、位置和/或时间)产生出蛋白质,因为它们已通过人为干预如按计划进行了遗传改造。

期望本领域技术人员已掌握可用于本发明蛋白质编码核酸表达的多个表达系统。因此不再试图详述用于在原核生物或真核生物中表达蛋白质的各种已知方法。

简而言之,编码本发明蛋白质的分离核酸的表达通常可通过将例如DNA或cDNA与启动子(它是组成型或是诱导型的)有效连接,随后掺入表达载体中来实现。载体可适于在原核生物或真核生物中复制并整合。表达载体通常含有转录和翻译终止子、起始序列和用于调节编码本发明蛋白质的DNA表达的启动子。为了获得高水平表达的克隆基因,需要构建表达载体至少含有指导转录的强启动子(例如泛素启动子)、用于翻译起始的核糖体结合位点和转录/翻译终止子。可将组成型启动子归类以提供各种组成型表达。因此,有些是弱组成型启动子,有些是强组成型启动子。所谓“弱启动子”一般是指驱动编码序列按低水平表达的启动子。所谓“低水平”是指约1/10,000转录物至约1/100,000转录物至约1/500,000转录物的水平。相反地,“强启动子”驱动编码序列按“高水平”或约1/10转录物至约1/100转录物至约1/1,000转录物进行表达。

技术人员应当了解的是,可对本发明的蛋白质进行修饰但不降低它的生物活性。可以进行某些修饰以促进克隆、表达或寻靶分子掺入融合蛋白。这类修饰为本领域技术人员所熟知,包括例如将甲硫氨酸添加到N端以提供起始位点,或者将额外的氨基酸(例如poly His)置于任一端以形成方便定位的限制位点或终止密码子或纯化序列。

### 原核生物中的表达

原核细胞可用作表达宿主。原核生物最常用的代表是不同的大肠杆菌菌株;然而,也可使用其它的微生物菌株。本文定义的包括用于转录



起始的启动子、任选的操纵基因、以及核糖体结合位点序列一起的常用的原核控制序列，包括诸如以下常用的启动子： $\beta$  内酰胺酶(青霉素酶)和乳糖(lac)启动子系统(Chang 等(1977) *Nature* 198:1056)、色氨酸(trp)启动子系统(Goeddel 等(1980) *Nucleic Acids Res.* 8:4057)和 $\lambda$ 衍生 PL 启动子和 N-基因核糖体结合位点(Shimatake 等(1981) *Nature* 292:128)。在已转进大肠杆菌内的 DNA 载体中包含选择标记也是有用的。这类标记的实例包括赋予氨苄西林、四环素或氯霉素抗性的基因。

选择载体以使目标基因能够导入合适的宿主细胞。细菌载体通常是质粒或噬菌体源的。合适的细菌细胞用噬菌体载体颗粒感染，或者用裸露的噬菌体载体 DNA 转染。如果使用质粒载体，则细菌细胞用质粒载体 DNA 转染。用芽孢杆菌(*Bacillus sp.*)和沙门氏菌(*Salmonella*)可获得用于表达本发明蛋白质的表达系统(Palva 等(1983) *Gene* 22:229-35; Mosbach 等(1983) *Nature* 302:543-5)。得自 Pharmacia 的 pGEX-4T-1 质粒载体是本发明优选的大肠杆菌表达载体。

### 真核生物中的表达

多种真核表达系统例如酵母、昆虫细胞系、植物和哺乳动物细胞为本领域技术人员所知。正如下面简要说明的一样，本发明可在这些真核系统中进行表达。在一些实施方案中，使用如下文中论述的转化/转染的植物细胞，作为生产本发明蛋白质的表达系统。

酵母中异源蛋白质的合成是众所周知的。Sherman 等(1982) *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory)是公认的巨著，它描述了可在酵母中生产蛋白质的各种方法。用于生产真核蛋白质的两种普遍使用的酵母是酿酒酵母(*Saccharomyces cerevisiae*)和巴斯德毕赤酵母(*Pichia pastoris*)。用于在酿酒酵母和毕赤酵母中表达的载体、菌株和方案是本领域已知的，可由供应商(例如 Invitrogen)提供。合适的载体通常根据需要

具有表达控制序列(例如启动子包括 3-磷酸甘油酸激酶或醇氧化酶启动子和复制起点)和终止序列等。

本发明的蛋白质一经表达,便可通过使细胞裂解并将标准蛋白质分离技术用于裂解物或沉淀,从而从酵母中分离出来。可通过应用蛋白质印迹技术或放射免疫测定法或其它标准免疫测定法技术,实现对纯化过程的监测。

本发明蛋白质的编码序列还可与用于转染例如哺乳动物、昆虫或植物源的细胞培养物的各种表达载体连接。哺乳动物细胞系统常常是单层细胞的形式,尽管也可使用哺乳动物细胞悬液。本领域已开发出能够表达完整蛋白质的许多合适的宿主细胞系,包括 HEK293、BHK21 和 CHO 细胞系。用于这些细胞的表达载体可包括表达控制序列,例如复制起点、启动子(例如 CMV 启动子、HSV *tk* 启动子或 *pgk* (磷酸甘油酸激酶)启动子)、增强子(Queen 等(1986) *Immunol. Rev.* 89:49)和必需的加工信息位点(例如核糖体结合位点、RNA 剪接位点、聚腺苷酸化位点(例如 SV40 大 T Ag 加 poly A 位点))和转录终止子序列。可从例如美国典型培养物保藏中心(American Type Culture Collection)细胞系和杂交瘤目录(第 7 版,1992)获得用于生产本发明蛋白质的其它动物细胞。

用于在昆虫细胞中表达本发明蛋白质的合适载体一般得自 SF9 杆状病毒。合适的昆虫细胞系包括蚊幼虫、蚕、粘虫、蛾和果蝇(*Drosophila*)细胞系,例如 Schneider 细胞系(参见例如 Schneider, (1987) *J. Embryol. Exp. Morphol.* 27:353-65)。

如使用酵母的情况一样,当使用高等动物或植物宿主细胞时,聚腺苷酸化序列或转录终止子序列通常被掺入载体中。终止子序列的一个实例是得自牛生长激素基因的聚腺苷酸化序列。还包括用于精确剪接转录物的序列。剪接序列的一个实例是得自 SV40 的 VP1 内含子(Sprague 等(1983) *J. Virol.* 45:773-81)。另外,可将在宿主细胞中控制复制的基因序

列掺到载体中，例如存在于牛乳头瘤型载体的载体(Saveria-Campo, “Bovine Papilloma Virus DNA a Eukaryotic Cloning Vector (一种真核克隆载体牛乳头瘤DNA)”, *DNA Cloning: A Practical Approach*, 第II卷, Glover 主编, IRL Press, Arlington, VA, 第213-38页(1985))。

另外，置于合适植物表达载体上的 ERECTA 基因可用来转化植物细胞。然后从植物愈伤组织分离出多肽，或者转化细胞可用来再生出转基因植物。可收获这类转基因植物，并可对合适的组织(例如种子或叶)进行大规模的蛋白质提取和纯化。

### 植物转化方法

将外源基因导入植物的多种方法是已知的，可用来将 ERECTA 多核苷酸插入植物宿主中，包括生物和物理的植物转化方案。参见例如 Miki 等，“Procedure for Introducing Foreign DNA into Plants (外源 DNA 导入植物的方法)”，*Plant Molecular Biology and Biotechnology*, Glick 和 Thompson 主编, CRC Press, Inc., Boca Raton, 第67-88页(1993)。所选择的方法因宿主植物而异，包括化学转染方法(例如磷酸钙法)、微生物介导的基因转移(例如土壤杆菌)(Horsch 等(1985) *Science* 227:1229-31)、电穿孔、显微注射和生物射弹轰击。

用于植物细胞或组织转化和植物再生的表达盒和载体以及体外培养方法是已知并可获得的。参见例如 Gruber 等，“Vectors for Plant Transformation (植物转化的载体)”，载于 *Plant Molecular Biology and Biotechnology*, 出处同上, 第89-119页。

可以通过通常用于直接递送给细胞的一种或多种技术，将分离多核苷酸或分离多肽导入植物中。这类方案可随基因修饰所靶定的生物、细胞、植物或植物细胞的类型(即单子叶植物或双子叶植物)而变化。转化植物细胞合适的方法包括显微注射(Crossway 等(1986) *Biotechniques*

4:320-334 和美国专利第 6,300,543 号)、电穿孔(Riggs 等(1986) *Proc. Natl. Acad. Sci. USA* 83:5602-5606)、直接基因转移(Paszkowski 等(1984) *EMBO J.* 3:2717-2722)和射弹粒子加速法(ballistic particle acceleration) (参见例如 Sanford 等, 美国专利第 4,945,050 号; WO 91/10725; 以及 McCabe 等(1988) *Biotechnology* 6:923-926)。另参见 Tomes 等, Direct DNA Transfer into Intact Plant Cells Via Microprojectile Bombardment (通过微粒轰击使 DNA 直接转移到完整植物细胞中), 第 197-213 页, 载于 *Plant Cell, Tissue and Organ Culture, Fundamental Methods*, O. L. Gamborg 和 G. C. Phillips 主编, Springer-Verlag Berlin Heidelberg New York, 1995; 美国专利第 5,736,369 号(分生组织); Weissinger 等(1988) *Ann. Rev. Genet.* 22:421-477; Sanford 等(1987) *Particulate Science and Technology* 5:27-37 (洋葱); Christou 等(1988) *Plant Physiol.* 87:671-674 (大豆); Datta 等(1990) *Biotechnology* 8:736-740 (水稻); Klein 等(1988) *Proc. Natl. Acad. Sci. USA* 85:4305-4309 (玉米); Klein 等(1988) *Biotechnology* 6:559-563 (玉米); WO 91/10725 (玉米); Klein 等(1988) *Plant Physiol.* 91:440-444 (玉米); Fromm 等(1990) *Biotechnology* 8:833-839; 以及 Gordon-Kamm 等(1990) *Plant Cell* 2:603-618 (玉米); Hooydaas-Van Slogteren & Hooykaas (1984) *Nature* (London) 311:763-764; Bytebier 等(1987) *Proc. Natl. Acad. Sci. USA* 84:5345-5349 (百合科(Liliaceae)); De Wet 等(1985), 载于 *The Experimental Manipulation of Ovule Tissues*, G.P. Chapman 等, 第 197-209 页, Longman, NY (花粉); Kaeppler 等(1990) *Plant Cell Reports* 9:415-418; 以及 Kaeppler 等(1992) *Theor. Appl. Genet.* 84:560-566 (颈须介导的转化); 美国专利第 5,693,512 号(超声处理); D'Halluin 等(1992) *Plant Cell* 4:1495-1505 (电穿孔); Li 等(1993) *Plant Cell Reports* 12:250-255; 以及 Christou 和 Ford (1995) *Annals of Botany* 75:407-413 (水稻); Osjoda 等(1996) *Nature Biotech.* 14:745-750; 土壤杆菌介导的玉米转化(美国专利第 5,981,840 号); 碳化

硅晶须(silicon carbide whisker)方法(Frame 等(1994) *Plant J.* 6:941-948); 激光方法(Guo 等(1995) *Physiologia Plantarum* 93:19-24); 超声处理方法(Bao 等(1997) *Ultrasound in Medicine & Biology* 23:953-959; Finer 和 Finer (2000) *Lett Appl Microbiol.* 30:406-10; Amoah 等(2001) *J Exp Bot* 52:1135-42); 聚乙二醇方法(Krens 等(1982) *Nature* 296:72-77); 单子叶植物和双子叶植物细胞的原生质体可用电穿孔(Fromm 等(1985) *Proc. Natl. Acad. Sci. USA* 82:5824-5828)和显微注射(Crossway 等(1986) *Mol. Gen. Genet.* 202:179-185 进行转化; 所述所有文献均通过引用结合到本文中。

### 土壤杆菌介导的转化

将表达载体导入植物的最普遍使用的方法以土壤杆菌的天然转化系统为基础。根癌土壤杆菌和发根土壤杆菌(*A. rhizogenes*)是植物致病性土壤细菌, 它从遗传上转化植物细胞。根癌土壤杆菌的 Ti 质粒和发根土壤杆菌的 Ri 质粒分别携带负责遗传转化植物的基因。参见例如 Kado, (1991) *Crit. Rev. Plant Sci.* 10:1。有关用于土壤杆菌介导的基因转移的土壤杆菌载体系统和方法的描述参见 Gruber 等(出处同上)、Miki 等(出处同上)和 Moloney 等(1989) *Plant Cell Reports* 8:238。

同样, 可将基因插入分别得自根癌土壤杆菌的 Ti 质粒或发根土壤杆菌 Ri 质粒的 T-DNA 区内。因此, 可以用这些质粒根据以上所述来构建表达盒。多种控制序列是已知的, 当其与异源编码序列偶联并转化到宿主生物内时, 相对于原始编码序列的组织/器官特异性, 在基因表达时具有保真度。参见例如 Benfey 和 Chua, (1989) *Science* 244:174-81。用于这些质粒的特别合适的控制序列是用于在各种靶植物中的基因组成型叶特异性表达的启动子。其它有用的控制序列包括得自胭脂碱合酶基因(NOS)的启动子和终止子。NOS 启动子和终止子存在于质粒 pARC2 中, 可获自美国典型培养物保藏中心, 并被命名为 ATCC 67238。如果使用这类

系统,则还需存在来自 Ti 或 Ri 质粒的毒力(*vir*)基因,所述毒力基因或者连同 T-DNA 部分一起存在,或者借助其中 *vir* 基因存在于单独载体上的二元系统。这类系统、用于其中的载体,以及转化植物细胞的方法参见美国专利第 4,658,082 号;1986 年 10 月 1 日提交的美国专利申请号 913,914,正如 1993 年 11 月 16 日授予的美国专利第 5,262,306 号中所引用的一样;以及 Simpson 等(1986) *Plant Mol. Biol.* 6:403-15 (同样在'306 专利中引用);所述所有文献均通过引用全部结合到本文中。

这些质粒一经构建,便可置于发根土壤杆菌或根癌土壤杆菌以及用于转化植物品种的细胞的载体中,所述植物品种通常易受镰孢(*Fusarium*)或链格孢(*Alternaria*)感染。本发明还包括若干其它的转基因植物,包括但不限于大豆、玉米、高粱、苜蓿、水稻、三叶草、甘蓝、香蕉、咖啡树、芹菜、烟草、豇豆、棉花、甜瓜和辣椒。根癌土壤杆菌或发根土壤杆菌任一种的选择将取决于其转化的植物。根癌土壤杆菌一般是用于转化的优选生物。大多数双子叶植物、一些裸子植物和少数单子叶植物(例如某些百合目(*Liliales*)和天南星目(*Arales*)的某些成员)易受根癌土壤杆菌感染。发根土壤杆菌的宿主范围也十分广,包括大多数双子叶植物和一些裸子植物,它们包括豆科(*Leguminosae*)、菊科(*Compositae*)和藜科(*Chenopodiaceae*)成员。目前单子叶植物的转化也取得了一定成效。欧洲专利申请号 604 662 A1 公开了利用土壤杆菌转化单子叶植物的方法。欧洲专利申请号 672 752 A1 公开了使用未成熟胚盾盖用土壤杆菌转化单子叶植物的方法。Ishida 等人对通过将未成熟胚暴露于根癌土壤杆菌以转化玉米的方法进行了论述(*Nature Biotechnology* 14:745-50 (1996))。

这些细胞一经转化,则可用来再生转基因植物。例如可通过割伤植物然后将载体引入伤口部位,从而用这些载体感染整株植物。可以割伤植物的任何组成部分,包括叶、茎和根。或者,可将这些载体接种到外植体形式的植物组织中,例如子叶组织或叶盘,并在促进植物再生的条

件下进行培养。用发根土壤杆菌或根癌土壤杆菌接种植物组织转化得到的根或茎干(含有串珠镰孢黄菌素降解酶编码基因),可用作植物组织源,通过体细胞胚胎发生或器官发生,再生出串珠镰孢黄菌素抗性转基因植物。再生植物组织的这类方法的实例公开于 Shahin, (1985) *Theor. Appl. Genet.* 69:235-40; 美国专利第 4,658,082 号; Simpson 等(出处同上); 以及 1986 年 10 月 1 日同时提交的美国专利申请号 913,913 和 913,914, 如在 1993 年 11 月 16 日授予的美国专利第 5,262,306 号中引用一样, 该处的全部公开内容通过引用结合到本文中。

### 直接基因转移

尽管土壤杆菌介导的转化的宿主范围较广, 但是一些主要谷类农作物品种和裸子植物通常抗拒这种模式的基因转移, 即使最近在水稻中取得一定成果(Hiei 等(1994) *Plant Journal* 6:271-82)。已经开发出替代土壤杆菌介导的转化的若干种植物转化方法, 统称为直接基因转移。

普遍适用的植物转化方法是微弹介导的转化, 其中尺寸约为 1-4  $\mu\text{m}$  的微弹表面上携带 DNA。用将微弹加速至 300-600 m/秒速度的生物射弹装置, 将表达载体导入植物组织中, 该速度足以穿透植物细胞壁和细胞膜(Sanford 等(1987) *Part. Sci. Technol.* 5:27; Sanford, (1988) *Trends Biotech* 6:299; Sanford, (1990) *Physiol. Plant* 79:206; 以及 Klein 等(1992) *Biotechnology* 10:268)。

将 DNA 物理性传递到植物中的另一种方法是对靶细胞进行超声处理, 参见 Zang 等(1991) *BioTechnology* 9:996。或者, 使用脂质体或原生质球融合将表达载体导入植物中。参见例如 Deshayes 等(1985) *EMBO J.* 4:2731; 以及 Christou 等(1987) *Proc. Natl. Acad. Sci. USA* 84:3962。有研究报道了采用  $\text{CaCl}_2$  沉淀、聚乙烯醇或聚-L-鸟氨酸使 DNA 直接被摄取进入原生质体。参见例如 Hain 等(1985) *Mol. Gen. Genet.* 199:161; 以及

Draper 等(1982) *Plant Cell Physiol.* 23:451。

还有研究披露了原生质体和整个细胞和组织的电穿孔。参见例如 Donn 等, 载于 *Abstracts of the VIIth Int'l. Congress on Plant Cell and Tissue Culture IAPTC* (IAPTC 第七届国际植物细胞和组织培养会议文摘), A2-38, 第 53 页(1990); D'Halluin 等(1992) *Plant Cell* 4:1495-505; 以及 Spencer 等(1994) *Plant Mol. Biol.* 24:51-61。

### 提高 ERECTA 多肽的活性和/或水平

本发明提供提高本发明 ERECTA 多肽的活性和/或水平的方法。可通过向植物提供 ERECTA 多肽来实现本发明 ERECTA 多肽的水平和/或活性的提高。可通过将编码 ERECTA 多肽的氨基酸序列导入植物、将编码 ERECTA 多肽的核苷酸序列导入植物或者通过对编码本发明 ERECTA 多肽的基因组基因座进行修饰, 来提供 ERECTA 多肽。

如本文其它部分所述, 本领域已知向植物提供多肽的多种方法, 包括但不限于将多肽直接导入植物、将编码具有植物生长调节活性的多肽的多核苷酸构建体(瞬时或稳定地)导入植物。还要了解的是, 本发明的方法可采用无法在转化植物中指导蛋白质或 RNA 表达的多核苷酸。因此, 可通过改变编码 ERECTA 多肽的基因或其启动子来提高 ERECTA 多肽的水平 and/或活性。参见例如 Kmiec, 美国专利第 5,565,350 号; Zarling 等, PCT/US93/03868。因此, 本发明提供携带 ERECTA 基因突变的经诱变的植物, 其中所述突变增加 ERECTA 基因的表达或提高已编码的 ERECTA 多肽的植物生长和/或器官发育的活性。

### 降低 ERECTA 多肽的活性和/或水平

本发明提供了通过用表达抑制 ERECTA 多肽表达的多核苷酸的表达盒转化植物细胞, 来降低或消除本发明 ERECTA 多肽活性的方法。所



述多核苷酸可通过防止 ERECTA 信使 RNA 翻译, 直接抑制 ERECTA 多肽的表达; 或者通过编码抑制编码 ERECTA 多肽的 ERECTA 基因转录或翻译的多肽, 间接抑制 ERECTA 多肽表达。用于在植物中抑制或消除基因在植物中表达的方法是本领域众所周知的, 而且任何的这类方法都可用于本发明以抑制 ERECTA 多肽的表达。

根据本发明, 以下情况说明 ERECTA 多肽的表达受到抑制: ERECTA 多肽的蛋白质水平低于未经遗传修饰或诱变以抑制同一 ERECTA 表达的植物中该 ERECTA 多肽的蛋白质水平的 70%。在本发明的具体实施方案中, 在本发明的改性植物中, ERECTA 多肽的蛋白质水平低于不是突变型或未经遗传修饰以抑制 ERECTA 多肽表达的植物中的同一 ERECTA 多肽蛋白质水平的 60%、低于 50%、低于 40%、低于 30%、低于 20%、低于 10%、低于 5%或低于 2%。可以通过例如测定 ERECTA 多肽在植物细胞或植物中的表达水平, 来直接测定 ERECTA 多肽的表达水平; 或者可以通过例如测定 ERECTA 多肽在植物细胞或植物中的植物生长和/或器官发育活性, 或测定植物的生物量, 来间接测定 ERECTA 多肽的表达水平。用于进行这类测定的方法见本文其它部分的描述。

在本发明的其它实施方案中, 通过用包含编码抑制 ERECTA 多肽活性的多肽的多核苷酸的表达盒转化植物细胞, 降低或消除 ERECTA 多肽的活性。根据本发明, 以下情况说明 ERECTA 多肽的植物生长和/或器官发育活性受到抑制: ERECTA 多肽的植物生长和/或器官发育活性小于未经改性以抑制该 ERECTA 多肽的植物生长和/或器官发育活性的植物的同一 ERECTA 多肽的植物生长和/或器官发育活性的 70%。在本发明的具体实施方案中, 本发明改性植物的 ERECTA 多肽的植物生长和/或器官发育活性小于未经改性以抑制 ERECTA 多肽表达的植物的同一 ERECTA 多肽的植物生长和/或器官发育活性的 60%、小于 50%、小于 40%、小于 30%、小于 20%、小于 10%或小于 5%。当使用本文其它部

分所述测定方法未检出 ERECTA 多肽的植物生长和/或器官发育活性时,则表示本发明 ERECTA 多肽的植物生长和/或器官发育活性被“消除”。ERECTA 多肽的植物生长和/或器官发育活性的测定方法见本文其它部分的描述。

在其它实施方案中,可通过破坏 ERECTA 多肽的编码基因,来降低或消除 ERECTA 多肽的活性。本发明包括在 ERECTA 基因中携带突变的诱变植物,其中突变降低 ERECTA 基因的表达或者抑制编码的 ERECTA 多肽的植物生长和/或器官发育活性。

因此,可以采用多种方法以降低或消除 ERECTA 多肽的活性。另外,可使用不止一种方法来降低单一 ERECTA 多肽的活性。下文中给出了降低或消除 ERECTA 多肽表达的方法的非限制性实例。

### **1. 基于多核苷酸的方法:**

在本发明的一些实施方案中,用能够表达抑制本发明 ERECTA 多肽表达的多核苷酸的表达盒来转化植物。本文所用术语“表达”是指基因产物的生物合成,包括所述基因产物的转录和/或翻译。例如,对于本发明的目的,能够表达抑制至少一种 ERECTA 多肽表达的多核苷酸的表达盒是能够产生抑制至少一种本发明 ERECTA 多肽转录和/或翻译的 RNA 分子的表达盒。由 DNA 分子“表达”或“产生”蛋白质或多肽是指产生蛋白质或多肽的编码序列进行转录和翻译,而由 RNA 分子“表达”或“产生”蛋白质或多肽则是指产生蛋白质或多肽的 RNA 编码序列进行翻译。

抑制 ERECTA 多肽表达的多核苷酸的实例见下文。

#### **i. 有义抑制/共抑制**

在本发明的一些实施方案中,可通过有义抑制(sense suppression)或共抑制(cosuppression)来达到 ERECTA 多肽表达的抑制。对于共抑制,

表达盒被设计成表达 RNA 分子, 该 RNA 分子对当于按“有义”方向编码 ERECTA 多肽的信使 RNA 的全部或部分。该 RNA 分子的过量表达可导致天然基因表达的降低。因此, 对用共抑制表达盒转化的得到多个植物系进行筛选, 以鉴定 ERECTA 多肽表达受最大抑制的植物系。

用于共抑制的多核苷酸可相当于 ERECTA 多肽编码序列的全部或部分、ERECTA 多肽转录物 5'和/或 3'非翻译区的全部或部分、或者编码 ERECTA 多肽的转录物的编码序列以及非翻译区两者的全部或部分。在其中多核苷酸包含 ERECTA 多肽编码区全部或部分的一些实施方案中, 表达盒被设计来消除多核苷酸的起始密码子, 从而不翻译任何蛋白质产物。

可利用共抑制来抑制植物基因的表达, 从而产生这样植物, 其中测不出由这些基因编码的蛋白质的蛋白质水平。参见例如 Broin 等(2002) *Plant Cell* 14:1417-1432。还可利用共抑制来抑制同一植物中多种蛋白质的表达。参见例如美国专利第 5,942,657 号。利用共抑制来抑制植物内源基因表达的方法参见 Flavell 等(1994) *Proc. Natl. Acad. Sci. USA* 91:3490-3496; Jorgensen 等(1996) *Plant Mol. Biol.* 31:957-973; Johansen 和 Carrington (2001) *Plant Physiol.* 126:930-938; Broin 等(2002) *Plant Cell* 14:1417-1432; Stoutjesdijk 等(2002) *Plant Physiol.* 129:1723-1731; Yu 等(2003) *Phytochemistry* 63:753-763; 以及美国专利号 5,034,323、5,283,184 和 5,942,657; 所述每个文献均通过引用结合到本文中。通过在表达盒有义序列的 3'端位置包括 poly-dT 区以及 5'包括聚腺苷酸化信号, 来提高共抑制的效率。参见美国专利公布号 20020048814, 该专利通过引用结合到本文中。这类核苷酸序列通常与内源基因转录物的序列具有基本序列同一性, 宜大于约 65%序列同一性, 更宜大于约 85%序列同一性, 最宜大于约 95%序列同一性。参见美国专利号 5,283,184 和 5,034,323; 所述专利通过引用结合到本文中。

## ii. 反义抑制

在本发明的一些实施方案中,可以通过反义抑制实现对 ERECTA 多肽表达的抑制。对于反义抑制,表达盒被设计来表达与编码 ERECTA 多肽的信使 RNA 的全部或部分互补的 RNA 分子。反义 RNA 分子的过量表达可导致天然基因的表达降低。因此,对用反义抑制表达盒转化得到的多个植物系进行筛选,以鉴定 ERECTA 多肽表达受最大抑制的植物系。

用于反义抑制的多核苷酸可相当于 ERECTA 多肽编码序列的互补序列的全部或部分、ERECTA 转录物 5'和/或 3'非翻译区的互补序列的全部或部分或者编码 ERECTA 多肽的转录物的编码序列和非翻译区两者的互补序列的全部或部分。另外,反义多核苷酸可与靶序列完全互补(即与靶序列的互补序列同一性为 100%)或部分互补(即与靶序列的互补序列的同一性小于 100%)。反义抑制可用来抑制同一植物多种蛋白质的表达。参见例如美国专利第 5,942,657 号。此外,部分反义核苷酸可用来破坏靶基因的表达。一般可采用至少 50 个核苷酸、100 个核苷酸、200 个核苷酸、300、400、450、500、550 个以上核苷酸的序列。用反义抑制以抑制植物内源基因表达的方法参见例如 Liu 等(2002) *Plant Physiol.* 129:1732-1743 和美国专利号 5,759,829 和 5,942,657, 所述文献均通过引用结合到本文中。可通过在表达盒反义序列 3'端位置包括 poly-dT 区以及包括 5'聚腺苷酸化信号,来提高反义抑制的效率。参见美国专利公布号 20020048814, 该专利通过引用结合到本文中。

## iii. 双链 RNA 干扰

在本发明的一些实施方案中,可以通过双链 RNA (dsRNA)干扰实现对 ERECTA 多肽表达的抑制。对于 dsRNA 干扰,在同一细胞中表达如上文所述用于共抑制的有义 RNA 分子以及与有义 RNA 分子完全或部分

互补的反义 RNA 分子，导致对相应内源信使 RNA 表达的抑制。

通过设计同时包含有义序列和反义序列的表达盒，可实现有义和反义分子的表达。或者，可以使用有义序列和反义序列分开的表达盒。用 dsRNA 干扰表达盒转化多个植物系，或者之后对表达盒进行筛选，以鉴定 ERECTA 多肽表达受最大抑制的植物系。用 dsRNA 干扰以抑制内源植物基因表达的方法参见 Waterhouse 等(1998) *Proc. Natl. Acad. Sci. USA* 95:13959-13964; Liu 等(2002) *Plant Physiol.* 129:1732-1743; 以及 WO 99/49029、WO 99/53050、WO 99/61631 和 WO 00/49035; 所述每个文献均通过引用结合到本文中。

#### **iv. 发夹 RNA 干扰和含有内含子的发夹 RNA 干扰**

在本发明的一些实施方案中，可以通过发夹 RNA (hpRNA) 干扰或含有内含子的发夹 RNA (ihpRNA) 干扰，来获得对 ERECTA 多肽之一的表达的抑制。这些方法在抑制内源基因表达中非常有效。参见 Waterhouse 和 Helliwell (2003) *Nat. Rev. Genet.* 4:29-38 及其中引用的参考文献。

对于 hpRNA 干扰，表达盒被设计成表达与自身杂交形成发夹结构的 RNA 分子，所述发夹结构包含单链环区和碱基配对茎。碱基配对茎区包含相当于编码其表达将被抑制的基因的内源信使 RNA 全部或部分的有义序列，以及与该有义序列完全或部分互补的反义序列。因此，该分子的碱基配对茎区一般决定 RNA 干扰的特异性。hpRNA 分子在抑制内源基因表达时十分有效，它们诱导的 RNA 干扰被遗传给植物后代。参见例如 Chuang 和 Meyerowitz (2000) *Proc. Natl. Acad. Sci. USA* 97:4985-4990; Stoutjesdijk 等(2002) *Plant Physiol.* 129:1723-1731; 以及 Waterhouse 和 Helliwell (2003) *Nat. Rev. Genet.* 4:29-38。用 hpRNA 干扰抑制基因表达或使基因表达沉默的方法参见例如 Chuang 和 Meyerowitz (2000) *Proc. Natl. Acad. Sci. USA* 97:4985-4990; Stoutjesdijk 等(2002) *Plant*

*Physiol.* 129:1723-1731; Waterhouse 和 Helliwell (2003) *Nat. Rev. Genet.* 4:29-38; Pandolfini 等, *BMC Biotechnology* 3:7; 以及美国专利公布号 20030175965; 所述每个文献均通过引用结合到本文中。hpRNA 构建体内使基因表达沉默的效率和的瞬时测定法参见 Panstruga 等(2003) *Mol. Biol. Rep.* 30:135-140, 该文献通过引用结合到本文中。

对于 ihpRNA, 该干扰分子具有与 hpRNA 相同的通用结构, 但是该 RNA 分子另包含在 ihpRNA 进行表达的细胞中能够被剪接的内含子。内含子的使用使发夹 RNA 分子环的大小在剪接后减到最小, 这就提高了干扰的效率。参见例如 Smith 等(2000) *Nature* 407:319-320。实际上, Smith 等人指出, 用 ihpRNA 介导的干扰 100%抑制内源基因的表达。用 ihpRNA 干扰以抑制内源植物基因表达的方法参见例如 Smith 等(2000) *Nature* 407:319-320; Wesley 等(2001) *Plant J.* 27:581-590; Wang 和 Waterhouse (2001) *Curr. Opin. Plant Biol.* 5:146-150; Waterhouse 和 Helliwell (2003) *Nat. Rev. Genet.* 4:29-38; Helliwell 和 Waterhouse (2003) *Methods* 30:289-295 以及美国专利公布号 20030180945, 所述每个文献均通过引用结合到本文中。

还可设计这样的 hpRNA 干扰表达盒, 使得有义序列和反义序列不与内源 RNA 相对应。在本实施方案中, 有义序列和反义序列在包含与靶基因内源信使 RNA 全部或部分相应的核苷酸序列的环序列的两侧。因此, 正是该环区决定了 RNA 干扰的特异性。参见例如 WO 02/00904, 该专利通过引用结合到本文中。

#### v. 扩增子介导的干扰

扩增子表达盒包含植物病毒衍生的序列, 该序列含有靶基因的全部或部分, 但是一般不含天然病毒基因的全部。表达盒转录产物中存在的病毒序列使转录产物能够指导自我复制。由扩增子产生的转录物相对于

靶序列(即 ERECTA 多肽的信使 RNA)可以是有义或反义的。用扩增子抑制内源植物基因表达的方法参见例如 Angell 和 Baulcombe (1997) *EMBO J.* 16:3675-3684; Angell 和 Baulcombe (1999) *Plant J.* 20:357-362; 以及美国专利第 6,646,805 号, 所述每个文献均通过引用结合到本文中。

#### **vi. 核酶**

在一些实施方案中, 通过本发明表达盒表达的多核苷酸是催化性 RNA 或具有对 ERECTA 多肽信使 RNA 有特异性的核酶活性。因此, 该多核苷酸引起内源信使 RNA 降解, 导致 ERECTA 多肽表达降低。该方法参见例如美国专利第 4,987,071 号, 该专利通过引用结合到本文中。

#### **vii. 小干扰 RNA 或微小 RNA**

在本发明的一些实施方案中, 可以经微小 RNA (miRNA) 编码基因的表达通过 RNA 干扰来实现对 ERECTA 多肽表达的抑制。miRNA 是由约 22 个核糖核苷酸组成的调节剂。miRNA 在抑制内源基因表达中非常有效。参见例如 Javier 等(2003) *Nature* 425: 257-263, 该文献通过引用结合到本文中。

对于 miRNA 干扰, 表达盒被设计成表达模仿内源 miRNA 基因的 RNA 分子。所述 miRNA 基因编码形成发夹结构的 RNA, 所述发夹结构含有与另一内源基因(靶序列)互补的 22 个核苷酸的序列。对于抑制 ERECTA 表达, 22 个核苷酸的序列选自 ERECTA 转录物序列, 并含有所述 ERECTA 序列按有义方向的 22 个核苷酸以及对应于与有义序列互补的反义序列的 21 个核苷酸。miRNA 分子在抑制内源基因表达中非常有效, 且它们诱导的 RNA 干扰被遗传给植物后代。

## 2. 基于多肽的基因表达的抑制

在一个实施方案中，多核苷酸编码与 ERECTA 多肽编码基因结合的锌指蛋白，导致基因表达降低。在具体实施方案中，锌指蛋白与 ERECTA 基因的调节区结合。在其它实施方案中，锌指蛋白与编码 ERECTA 多肽的信使 RNA 结合并阻止其进行翻译。选择被锌指蛋白靶向的位点的方法参见例如美国专利第 6,453,242 号，利用锌指蛋白抑制植物中基因表达的方法参见例如美国专利公布号 20030037355；所述每个专利均通过引用结合到本文中。

## 3. 基于多肽的蛋白质活性的抑制

在本发明的一些实施方案中，多核苷酸编码与至少一种 ERECTA 多肽结合的抗体，并降低 ERECTA 多肽的植物生长调节活性。在另一个实施方案中，抗体的结合通过细胞质量控制机制使抗体-ERECTA 复合体的更新加快。植物细胞中抗体的表达，以及植物细胞中通过抗体表达及抗体与蛋白质的结合来进行的抑制分子途径是本领域众所周知的。参见例如 Conrad 和 Sonnewald (2003) *Nature Biotech.* 21:35-36, 该文献通过引用结合到本文中。

## 4. 基因破坏

在本发明的一些实施方案中，通过破坏编码 ERECTA 多肽的基因来降低或消除 ERECTA 多肽的活性。可以通过本领域已知方法破坏编码 ERECTA 多肽的基因。例如，在一个实施方案中，基因被转座子标记破坏。在另一个实施方案中，采用随机诱变或定向诱变对植物进行诱变来破坏基因，并选出植物生长调节活性降低的植物。



### **i. 转座子标记**

在本发明的一个实施方案中，转座子标记用来降低或消除一种或多种 ERECTA 多肽的 ERECTA 活性。转座子标记包括将转座子插入内源 ERECTA 基因中以降低或消除 ERECTA 多肽的表达。“ERECTA 基因”是指编码本发明 ERECTA 多肽的基因。

在本实施方案中，通过将转座子插入 ERECTA 多肽编码基因的调节区或编码区来降低或消除一种或多种 ERECTA 多肽的表达。可以使用在外显子、内含子、5'或 3'非翻译序列、启动子或 ERECTA 基因的任何其它调节序列内的转座子来降低或消除编码 ERECTA 多肽的表达和/或活性。

植物中用于特定基因转座子标记的方法是本领域众所周知的。参见例如 Maes 等(1999) *Trends Plant Sci.* 4:90-96; Dharmapuri 和 Sonti (1999) *FEMS Microbiol. Lett.* 179:53-59; Meissner 等(2000) *Plant J.* 22:265-274; Phogat 等(2000) *J. Biosci.* 25:57-63; Walbot (2000) *Curr. Opin. Plant Biol.* 2:103-107; Gai 等(2000) *Nucleic Acids Res.* 28:94-96; Fitzmaurice 等(1999) *Genetics* 153:1919-1928)。另外，在选定基因中选出  $\mu$  插入的 TUSC 法参见 Bensen 等(1995) *Plant Cell* 7:75-84; Mena 等(1996) *Science* 274:1537-1540; 以及美国专利第 5,962,764 号; 所述每个文献均通过引用结合到本文中。

### **ii. 活性降低的突变型植物**

用于在植物中降低或消除内源基因表达的其它方法也是本领域已知的，同样可适用于本发明。这些方法包括其它诱变形式，例如甲磺酸乙酯诱导的诱变、缺失诱变和用于在(用 PCR)反向遗传学有义(reverse genetics sense)的快速中子缺失诱变(fast neutron deletion mutagenesis)，以鉴定其中内源基因已缺失的植物系。这些方法的实例参见 Ohshima 等

(1998) *Virology* 243:472-481; Okubara 等(1994) *Genetics* 137:867-874; 以及 Quesada 等(2000) *Genetics* 154:421-436; 所述每个文献均通过引用结合到本文中。另外, 用于筛选化学诱导突变的自动化快速方法 TILLING (定向诱导基因组局部突变法(Targeting Induced Local Lesions In Genomes)), 利用变性 HPLC 或选择性内切核酸酶消化选出的 PCR 产物, 同样适用于本发明。参见 McCallum 等(2000) *Nat. Biotechnol.* 18:455-457, 该文献通过引用结合到本文中。

影响基因表达或者干扰编码蛋白质的功能(植物生长调节活性)的突变是本领域众所周知的。在基因外显子中的插入突变常常导致无效突变体。保守残基的突变在抑制编码蛋白质的植物生长调节活性时特别有效。有研究指出适于诱变以达消除植物生长调节活性目的的植物 ERECTA 多肽的保守残基。这类突变型可按照众所周知的方法分离出来, 而且不同 ERECTA 基因座中的突变可以通过遗传杂交堆积。参见例如 Gruis 等(2002) *Plant Cell* 14:2863-2882。

在本发明的另一个实施方案中, 显性突变型可用来引发由复制基因座的基因倒置和重组基因所致的 RNA 沉默。参见例如 Kusaba 等(2003) *Plant Cell* 15:1455-1467。

本发明包括用于降低或消除一种或多种 ERECTA 多肽活性的其它方法。用于使植物基因组核苷酸序列改变或突变的其它方法的实例是本领域已知的, 包括但不限于使用 RNA:DNA 载体、RNA:DNA 突变载体、RNA:DNA 修复载体、混合双链体寡核苷酸(mixed-duplex oligonucleotide)、自身互补的 RNA:DNA 寡核苷酸(self-complementary RNA:DNA oligonucleotide)和诱重组的寡核苷酸碱基(recombinogenic oligonucleobase)。这类载体和使用方法是本领域已知的。参见例如美国专利号 5,565,350、5,731,181、5,756,325、5,760,012、5,795,972 和 5,871,984; 所述每个专利均通过引用结合到本文中。另参见 WO 98/49350、WO

99/07865、WO 99/25821 和 Beetham 等(1999) *Proc. Natl. Acad. Sci. USA* 96:8774-8778; 所述每个专利均通过引用结合到本文中。

### iii. 调节植物生长和/或器官发育活性

在具体方法中，植物中特定组织生长的水平和/或活性随植物中 ERECTA 多肽水平或活性的提高而提高。本文其它部分论述了用于提高植物中 ERECTA 多肽的水平或活性的方法。简单来说，这类方法包括向植物提供本发明的 ERECTA 多肽从而提高 ERECTA 多肽的水平或活性。在其它实施方案中，可按如下方式提供编码 ERECTA 多肽的 ERECTA 核苷酸序列：将包含本发明 ERECTA 核苷酸序列的多核苷酸导入植物，使 ERECTA 序列进行表达，提高 ERECTA 多肽的活性，从而增加植物或植物组成部分中的组织细胞数。在其它实施方案中，导入植物的 ERECTA 核苷酸构建体被稳定掺入植物的基因组。

在其它方法中，通过提高植物中 ERECTA 多肽的水平或/或活性来增加细胞数和植物组织的生物量。本文其它部分详细公开了这类方法。在这类的一种方法中，ERECTA 核苷酸序列被导入植物，且所述 ERECTA 核苷酸序列的表达降低 ERECTA 多肽的活性，从而加快植物或植物组成部分的植物生长和/或器官发育。在其它实施方案中，导入植物的 ERECTA 核苷酸构建体被稳定掺入植物的基因组。

如上所述，技术人员应当了解用于调节植物中植物生长和/或器官发育多核苷酸和多肽的水平/活性的合适的启动子。本文其它部分公开了该实施方案的示例性启动子。

因此，本发明还提供当与对照植物组织的植物生长和/或器官发育相比时，植物生长和/或器官发育得到改进的植物。在一个实施方案中，本发明植物中的本发明 ERECTA 多肽的水平/活性提高，从而植物组织中植物生长和/或器官发育加快。在其它实施方案中，本发明植物中的本发

明 ERECTA 多肽水平降低或消除，从而植物组织中植物生长和/或器官发育减缓。在其它实施方案中，这类植物具有稳定掺入其基因组的核酸分子，该核酸分子包含与在植物细胞中驱动表达的启动子有效连接的本发明 ERECTA 核苷酸序列。

#### iv. 调节根发育

本发明提供用于调节植物根发育的方法。所谓“调节根发育”是指当与对照植物相比时，植物根发育中产生的任何变化。这类根发育中的变化包括但不限于初生根生长率、新鲜根重、侧根和不定根形成程度、维管系统、分生组织发育或辐射延伸的变化。

本发明提供用于调节植物根发育的方法。该方法包括调节植物 ERECTA 多肽的水平和/或活性。在一种方法中，向植物提供本发明的 ERECTA 序列。在另一种方法中，按如下方式提供 ERECTA 核苷酸序列：将包含本发明 ERECTA 核苷酸序列的多核苷酸导入植物，使 ERECTA 序列进行表达，从而改善根发育。在另外的其它方法中，导入植物的 ERECTA 核苷酸构建体稳定掺入植物基因组。

在其它方法中，通过改变植物中 ERECTA 多肽的水平或活性来调节根发育。当与对照植物相比时，ERECTA 活性的提高可引起至少一种或多种根发育的下列变化，包括但不限于根分生组织变大、根生长加快、辐射延伸加强、维管系统增加、根分支增加、不定根较多和/或新鲜根重增加。

本文所用“根生长”包括在单子叶植物和双子叶植物根系发育不同阶段构成根系不同部分的生长的所有方面。要了解的是，根生长的加快可以由根的一个或多个组成部分(包括初生根、侧根、不定根等)的生长加快引起。

测定根系的这类发育变化的方法是本领域已知的。参见例如美国专

利申请号 20030074698 和 Werner 等(2001) *PNAS* 18:10487-10492, 这两个文献均通过引用结合到本文中。

如上所述,技术人员应当了解用来调节植物根发育的合适的启动子。本实施方案的示例性启动子包括组成型启动子和根优选的启动子(root-preferred promoter)。本文其它部分公开了示例性的根优选的启动子。

通过提高 ERECTA 多肽的活性和/或水平以刺激根生长和增加根质量还用于改善植物的抗倒伏性(standability)。术语“抗倒伏”或“抗倒伏性”是指植物将自身固定在土壤中的能力。对于具有直立或半直立生长习性的植物,该术语还指不利(环境)条件下保持竖立的能力。该性状与根系的大小、深度和形态有关。另外,通过提高 ERECTA 多肽的水平和/或活性来刺激根生长和增加根质量还用于体外促进外植体的繁殖。

此外,由 ERECTA 活性的水平和/或活性提高所致的较高的根生物产量对产量有直接作用,并对由根细胞或转基因根细胞或所述转基因根细胞的细胞培养物产生的化合物产量有间接作用。根培养物中产生有价值的化合物的一个实例是紫草宁,可以通过所述方法有利地提高其产量。

因此,本发明还提供当与对照植物的根发育相比时根发育受到调节的植物。在一些实施方案中,本发明植物中的本发明 ERECTA 多肽的水平/活性提高,而且根生长增强和/或根生物量增加。在其它实施方案中,这类植物具有稳定掺入其基因组的核酸分子,该核酸分子包含与在植物细胞中驱动表达的启动子有效连接的本发明 ERECTA 核苷酸序列。

#### v. 调节茎干和叶发育

本发明还提供用于调节植物茎干(shoot)和叶发育的方法。所谓“调节茎干和/或叶发育”是指植物茎干和/或叶发育中的任何变化。茎干和/或叶发育中的这类变化包括但不限于以下变化:茎干分生组织(shoot meristem)

发育、叶数、叶大小、叶和茎维管、节间长度和叶衰老。本文所用“叶发育”和“茎干发育”分别包括在单子叶植物和双子叶植物的叶和茎干发育的不同阶段，构成叶系和茎轴系不同组成部分生长的所有方面。用于测定茎轴系和叶系的这类发育变化的方法是本领域已知的。参见例如 Werner 等(2001) *PNAS* 98:10487-10492 和美国专利申请号 20030074698，所述每个文献均通过引用结合到本文中。

用于调节植物茎干和/或叶发育的方法包括调节本发明 ERECTA 多肽的活性和/或水平。在一个实施方案中，提供本发明的 ERECTA 序列。在其它实施方案中，可通过以下方式提供 ERECTA 核苷酸序列：将包含本发明 ERECTA 核苷酸序列的多核苷酸导入植物，使 ERECTA 序列进行表达，从而改善茎干和/或叶发育。在其它实施方案中，导入植物的 ERECTA 核苷酸构建体被稳定掺入植物基因组。

在具体的实施方案中，通过降低植物中 ERECTA 多肽的水平或活性来调节茎干或叶发育。当与对照植物相比时，ERECTA 活性的降低可引起至少一种或多种茎干和/或叶发育的下列变化，包括但不限于叶数减少、叶表面缩小、维管减少、节间变短和生长受阻和叶衰老延迟。

如上所述，技术人员应当了解用于调节植物茎干和叶发育的合适的启动子。本实施方案的示例性启动子包括组成型启动子、茎干优选的启动子、茎干分生组织优选的启动子和叶优选的启动子。本文其它部分公开了示例性的启动子。

降低植物的 ERECTA 活性和/或水平导致节间变短且生长受阻。因此，本发明的方法可用于产生矮生植物。另外，如上所述，调节植物的 ERECTA 活性可调节根和茎干的生长。因此，本发明还提供用于改变根冠比的方法。还可通过降低植物的 ERECTA 多肽的水平或活性来调节茎干或叶发育。

因此，本发明还提供当与对照植物相比时茎干和/或叶发育受到调节

的植物。在一些实施方案中，本发明植物中的本发明 ERECTA 多肽的水平/活性提高，茎干和/或叶发育发生改变。这类变化包括但不限于当与对照植物相比时叶数增多、叶表面扩大、维管增加、节间变长和植物高度增高以及叶衰老发生改变。在其它实施方案中，本发明植物中的本发明 ERECTA 多肽的水平/活性降低。

#### vi 调节生殖组织发育

本发明提供用于调节生殖组织发育的方法。在一个实施方案中，提供调节植物花发育的方法。所谓“调节花发育”是指当与其中 ERECTA 多肽的活性或水平未受调节的对照植物相比时，植物生殖组织结构的任何变化。“调节花发育”还包括当与其中 ERECTA 多肽活性或水平未受调节的对照植物相比时，植物生殖组织发育时间上的任何变化(即花发育的时间延迟或加快)。宏观变化可包括以下方面的改变：生殖器官的大小、形状、数量或位置、这些结构形成的发育时期或者在环境胁迫期间维持或继续整个开花过程的能力。微观变化可包括构成生殖器官的细胞类型或形状的改变。

用于调节植物花发育的方法包括调节植物中的 ERECTA 活性。在一种方法中，提供本发明的 ERECTA 序列。可按如下方式提供 ERECTA 核苷酸序列：将包含本发明 ERECTA 核苷酸序列的多核苷酸导入植物，使 ERECTA 序列进行表达，从而改善花发育。在其它实施方案中，导入植物的 ERECTA 核苷酸构建体被稳定掺到植物基因组。

在具体方法中，通过降低植物中 ERECTA 多肽的水平或活性来调节花发育。ERECTA 活性的降低可导致花发育的至少一种或多种下列变化，包括但不限于当与对照植物相比时开花延迟、花的数量减少、部分雄性不育和结籽减少。诱导开花延迟或抑制开花可用来提高饲料作物(例如苜蓿)的产量。用于测定花发育中的这类发育变化的方法是本领域

已知的。参见例如 Mouradov 等(2002) *The Plant Cell* S111-S130, 该文献通过引用结合到本文中。

如上所述,技术人员应当了解用于调节植物花发育的合适的启动子。本实施方案的示例性启动子包括组成型启动子、诱导型启动子、茎干优选的启动子和花序优选的启动子。

在其它方法中,提高本发明 ERECTA 序列的水平和/或活性来调节花的发育。这类方法可包括将 ERECTA 核苷酸序列导入植物及提高 ERECTA 多肽的活性。在其它方法中,导入植物的 ERECTA 核苷酸构建体被稳定掺到植物基因组。增加本发明 ERECTA 序列的表达可在胁迫期间调节花的发育。这类方法见本文其它部分。因此,本发明还提供与对照植物的花发育相比,花的发育受到调节的植物。组合物(composition)包括本发明 ERECTA 多肽的水平/活性提高且花的发育受到调节的植物。组合物还包括本发明 ERECTA 多肽的水平/活性得到提高的植物,其中植物在胁迫期间维持或继续整个开花过程。

本发明还提供使用本发明的 ERECTA 序列以增加种子大小和/或种子重量的方法。该方法包括提高植物或植物组成部分(例如种子)中 ERECTA 序列的活性。种子大小和/或重量的增加包括种子大小或重量的增加和/或一个或多个种子组成部分大小或重量的增加,包括例如胚、胚乳、种皮、糊粉或子叶。

如上所述,技术人员应当了解用来增加种子大小和/或种子重量的合适的启动子。本实施方案的示例性启动子包括组成型启动子、诱导型启动子、种子优选的启动子、胚优选的启动子和胚乳优选的启动子。

用于减小植物种子大小和/或减轻种子重量的方法包括降低植物的 ERECTA 活性。在一个实施方案中,可按如下方式提供 ERECTA 核苷酸序列:将包含本发明 ERECTA 核苷酸序列的多核苷酸导入植物,使 ERECTA 序列进行表达,从而减轻种子重量和/或减小种子大小。在其它



实施方案中，导入植物的 ERECTA 核苷酸构建体被稳定掺入到植物基因组。

还要了解的是，增加种子大小和/或重量还伴随着幼苗生长速度加快或早期活力增强。本文所用术语“早期活力(early vigor)”是指植物在早期发育期快速生长的能力，并且与萌发之后成功建立起发育完善的根系和发育充分的光合器有关。另外，当与对照相比时，种子大小和/或重量的增加还可导致植物产量提高。

因此，本发明还提供当与对照植物相比时种子重量和/或种子大小增加的植物。在其它实施方案中，还提供活力和植物产量提高的植物。在一些实施方案中，本发明植物中的本发明 ERECTA 多肽的水平/活性提高，且种子重量和/或种子大小增加。在其它实施方案中，这类植物具有稳定掺入其基因组的核酸分子，该核酸分子包含与在植物细胞中驱动表达的启动子有效连接的本发明 ERECTA 核苷酸序列。

#### **vii. ERECTA 启动子多核苷酸的使用方法**

包含本发明所公开的 ERECTA 启动子及其变体和片段的多核苷酸当与 DNA 构建体进行装配使得启动子序列与包含目标多核苷酸的核苷酸序列有效连接时，可用于任何宿主细胞、优选植物细胞的遗传操控。按这种方式，在表达盒中提供连同用于在目标宿主细胞中表达的目标多核苷酸序列的本发明 ERECTA 启动子多核苷酸。如下面的实施例 2 中所论述的一样，本发明的 ERECTA 启动子序列在多种组织中表达，因此该启动子序列可用于调节目标多核苷酸的时序表达和/或空间表达。

合成杂合启动子区是本领域已知的。这类区域包含与另一种多核苷酸的启动子元件有效连接的一种多核苷酸的上游启动子元件。在本发明的一个实施方案中，异源序列表达受合成杂合启动子的控制，该杂合启动子包含与得自异源启动子的上游启动子元件有效连接的本发明

ERECTA 启动子序列或其变体或片段。已鉴定出参与植物防御系统的上游启动子元件，并可用来产生合成启动子。参见例如 Rushton 等(1998) *Curr. Opin. Plant Biol.* 1:311-315。或者，合成 ERECTA 启动子序列可包含 ERECTA 启动子序列内存在的上游启动子元件的重复。

要了解的是，本发明的启动子序列可与其天然 ERECTA 编码序列一起使用。包含与其天然 ERECTA 基因有效连接的 ERECTA 启动子的 DNA 构建体可用来转化任何目标植物以产生所需要的表型变化，例如调节细胞数、调节根、茎干、叶、花和胚发育、胁迫耐受性和本文其它部分所述的其它任何表型。

为了改变植物的表型，本文所公开的启动子核苷酸序列和方法可用来调节宿主植物中任何异源核苷酸序列的表达。各种表型变化是有益的，包括改善植物中的脂肪酸组成、改变植物的氨基酸含量、改变植物的病原体防御机制等。可通过提供表达的异源产物或增加植物中内源产物的表达，来达到这类效果。或者，可通过降低植物一种或多种内源产物(特别是酶或辅因子)的表达，来达到这类效果。这些变化导致转化植物表型发生改变。

目标基因反映出商业市场和参与作物开发的各方的利益。目标作物和目标市场不断改变，而且因为发展中国家向世界开放市场，所以新的作物和技术也将涌现。另外，当我们对农艺性状和特征(例如产量和杂种优势)的了解加深时，对用于转化的基因的选择也会因此发生改变。目标基因的通用种类包括例如参与信息的基因(例如锌指)、参与通讯的基因(例如激酶)和参与持家的基因(例如热激蛋白)。转基因更特别的种类例如包括重要农艺性状、抗虫性、抗病性、除草剂抗性、不育性、谷类特征和商业产物的编码基因。目标基因一般包括参与油、淀粉、糖或营养代谢的基因以及影响谷粒大小、蔗糖负载等的基因。

在某些实施方案中，为了产生具有所需表型的植物，可将本发明的

核酸序列与其它目标多核苷酸序列组合使用(“堆积(stacked)”)。所形成的组合可包括目标多核苷酸任一种或多种的多个拷贝。本发明的多核苷酸还可与任何基因或组合基因堆积以产生具有多种所需性状组合的植物,包括但不限于用于动物饲料的所需性状,例如高油基因(例如美国专利第6,232,529号);平衡型氨基酸(例如 hordothionins (美国专利号 5,990,389、5,885,801、5,885,802 和 5,703,409); 大麦高赖氨酸(Williamson 等(1987) *Eur. J. Biochem.* 165:99-106 和 WO 98/20122); 高甲硫氨酸蛋白(Pedersen 等(1986) *J. Biol. Chem.* 261:6279; Kirihara 等(1988) *Gene* 71:359 和 Musumura 等(1989) *Plant Mol. Biol.* 12:123); 提高可消化性(例如改性贮藏蛋白(2001年11月7日提交的美国专利申请顺序号 10/053,410); 以及硫氧还蛋白(2001年12月3日提交的美国专利申请顺序号 10/005,429)), 所述文献的公开内容通过引用结合到本文中。本发明的多核苷酸还可与抗虫性、抗病性或除草剂抗性等所需要的性状堆积(例如苏云金芽孢杆菌(*Bacillus thuringiensis*)毒蛋白(美国专利号 5,366,892、5,747,450、5,737,514、5,723,756、5,593,881; Geiser 等(1986) *Gene* 48:109); 凝集素(Van Damme 等(1994) *Plant Mol. Biol.* 24:825); 串珠镰孢黄菌素解毒基因(美国专利第 5,792,931 号); 无毒基因和抗病性基因(Jones 等(1994) *Science* 266:789; Martin 等(1993) *Science* 262:1432; Mindrinis 等(1994) *Cell* 78:1089); 导致除草剂抗性的乙酰乳酸合酶(ALS)突变型, 例如 S4 和/或 Hra 突变; 谷氨酰胺合酶抑制剂, 例如膦丝菌素或 basta (例如 bar 基因); 草甘膦抗性(EPSPS 基因)); 用于处理或加工产品所需要的性状, 例如高油(例如美国专利第 6,232,529 号); 改良油(例如脂肪酸去饱和酶基因(美国专利第 5,952,544 号、WO 94/11516)); 改性淀粉(例如 ADPG 焦磷酸化酶(AGPase)、淀粉合酶(SS)、淀粉分支酶(SBE)和淀粉脱支酶(SDBE)); 聚合物或生物塑料(bioplastics) (例如美国专利第 5,602,321 号; 促进聚羟基链烷酸酯(PHA)表达的  $\beta$ -酮硫解酶、聚羟基丁酸合酶和乙酰乙酰辅酶 A

还原酶(Schubert 等(1988) *J. Bacteriol.* 170:5837-5847)), 所述文献的公开内容通过引用结合到本文中。还可以将本发明的多核苷酸与影响以下农艺性状的多核苷酸组合: 例如雄性不育(例如参见美国专利第 5,583,210 号)、柄(stalk)强度、开花时间或转化技术性状(例如细胞周期调节或基因寻靶)(例如 WO 99/61619、WO 00/17364、WO 99/25821), 所述文献的公开内容通过引用结合到本文中。

在一个实施方案中, 目标序列改进植物生长和/或提高农作物产量。例如, 目标序列包括导致初生根系或侧根系改善的具有农艺重要性的基因。这类基因包括但不限于养分/水转运蛋白基因和生长诱导基因。这类基因的实例包括但不限于玉米质膜  $H^+$ -ATP 酶(MHA2) (Frias 等(1996) *Plant Cell* 8:1533-44); AKT1, 一种拟南芥钾摄取器(potassium uptake apparatus)的组分(Spalding 等(1999) *J Gen Physiol* 113:909-18); 在根端细胞中激活细胞分裂周期的 RML 基因(Cheng 等(1995) *Plant Physiol* 108:881); 玉米谷氨酰胺合成酶基因(Sukanya 等(1994) *Plant Mol Biol* 26:1935-46)和血红蛋白(Duff 等(1997) *J. Biol. Chem* 27:16749-16752, Arredondo-Peter 等(1997) *Plant Physiol.* 115:1259-1266; Arredondo-Peter 等(1997) *Plant Physiol* 114:493-500 及其中所引用的参考文献)。目标序列还可用于表达对根发育有负面影响的基因的反义核苷酸序列。

另外, 除采用传统育种方法以外, 还可从遗传上改变具有农艺重要性的性状(例如油、淀粉和蛋白质含量)。有关改进包括增加油酸、饱和与不饱和油的含量、提高赖氨酸和硫的水平、提供必需氨基酸, 还包括淀粉改性。Hordothionin 蛋白改进参见美国专利号 5,703,049、5,885,801、5,885,802 和 5,990,389, 这些专利均通过引用结合到本文中。另一个实例是由大豆 2S 白蛋白编码的富含赖氨酸和/或硫的种子蛋白(参见美国专利第 5,850,016 号)和得自大麦的胰凝乳蛋白酶抑制剂(参见 Williamson 等(1987) *Eur. J. Biochem.* 165:99-106, 所述文献的公开内容通过引用结合到

本文中)。

可通过位点定向诱变制备编码序列的衍生物以提高编码多肽中预选定氨基酸的水平。例如,大麦高赖氨酸多肽(BHL)编码基因得自大麦胰凝乳蛋白酶抑制剂(1996年11月1日提交的美国专利专利申请顺序号08/740,682,以及WO 98/20133,所述专利的公开内容通过引用结合到本文中)。其它蛋白质包括例如得自向日葵种子的富含甲硫氨酸的植物蛋白(Lilley等(1989) *Proceedings of the World Congress on Vegetable Protein Utilization in Human Foods and Animal Feedstuffs* (植物蛋白质在人类食物和动物饲料中利用世界大会会议录), Applewhite 主编(美国油类化学家学会(American Oil Chemists Society), Champaign, Illinois), 第497-502页; 该文献通过引用结合到本文中); 玉米(Pedersen等(1986) *J. Biol. Chem.* 261:6279; Kiriara等(1988) *Gene* 71:359; 所述两份文献均通过引用结合到本文中); 以及水稻 (Musumura等(1989) *Plant Mol. Biol.* 12:123, 该文献通过引用结合到本文中)。其它有农艺重要性的基因编码乳胶、Floury 2、生长因子、种子贮藏因子(seed storage factor)和转录因子。

抗虫性基因可编码害虫抗性,所述害虫显著拉低产量,例如根虫、切根虫、欧洲玉米螟等。这类基因包括例如苏云金芽孢杆菌毒蛋白基因(美国专利号5,366,892、5,747,450、5,736,514、5,723,756、5,593,881; 以及 Geiser等(1986) *Gene* 48:109)等。

抗病性性状编码基因包括解毒基因,例如针对串珠镰孢菌素的解毒基因(美国专利第5,792,931号); 无毒(avr)基因和抗病(R)基因(Jones等(1994) *Science* 266:789; Martin等(1993) *Science* 262:1432; 以及 Mindrinos等(1994) *Cell* 78:1089)等。

除草剂抗性性状可包括起抑制乙酰乳酸合酶(ALS)作用的除草剂、特别是磺酰脲型除草剂的抗性编码基因(例如含有导致这类抗性的突变的乙酰乳酸合酶(ALS)基因,特别是S4和/或Hra突变); 起抑制谷氨酰

胺合酶作用的除草剂(例如膦丝菌素或 basta)的抗性编码基因(例如 bar 基因)或者本领域已知的其它这类基因。*bar* 基因编码除草剂 basta 抗性, *nptII* 基因编码抗生素卡那霉素和遗传霉素抗性, ALS-基因突变型编码除草剂氯磺隆抗性。

还可在表达盒中编码不育基因, 并提供物理去雄的替代方法。用于这类方式的基因的实例包括雄性组织优选的基因和具有雄性不育表型的基因, 例如 QM, 参见美国专利第 5,583,210 号。其它基因包括激酶和编码对雄性或雌性配子体发育有毒的化合物的基因。

谷类品质反映在性状上, 例如饱和及不饱和油的水平和类型、必需氨基酸的品质和数量以及纤维素的水平。玉米中, 改良的 hordothionin 蛋白可参见美国专利号 5,703,049、5,885,801、5,885,802 和 5,990,389。

商业性状还可编码到一个或多个基因上, 从而可增加例如用于生产乙醇的淀粉或者提供蛋白质表达。转化植物的另一种重要的商业用途是生产聚合物和生物塑料, 例如参见美国专利第 5,602,321 号。基因例如  $\beta$ -酮硫解酶、PHBase (聚羟基丁酸酯合酶)和乙酰乙酰辅酶 A 还原酶(参见 Schubert 等(1988) *J. Bacteriol.* 170:5837-5847)可促进聚羟基链烷酸酯(PHA)表达。

外源产物包括植物酶和植物产物以及得自包括原核生物和其它真核生物在内的其它来源的产物。这些产物包括酶、辅因子、激素等。可以提高蛋白质、特别是修饰蛋白质的水平, 所述修饰蛋白质具有改进的能提高植物营养价值的氨基酸分布。这可通过表达这类氨基酸含量提高的蛋白质来实现。

可参照下列非限制性实施例来更好地理解本发明。本领域技术人员应当了解的是, 可以实施本发明的其它实施方案但不偏离本文所公开及要求保护的本发明的精神和范围。

## 实施例

### 实施例 1 ERECTA 序列的分离

采用确定基因家族所有成员的常规方法以检索目标 ERECTA 基因。准备一组各式的所有基因家族已知成员作为蛋白质序列。该数据包括来自其它品种的序列。对照专有玉米序列数据集对这些品种进行检索，确定出一组非冗余(nonredundant)的重叠命中数(hits)。分别对任何目标基因的核苷酸序列进行处理，并对照数据库进行检索，重新得到一组非冗余的所有重叠命中数。然后将蛋白质命中数组与核苷酸命中数相比较。如果基因家族是完整的，则所有蛋白质命中数都包含在核苷酸命中数内。ERECTA 家族的基因由 2 个拟南芥基因、2 个水稻基因、2 个玉米基因、3 个高粱基因和 4 个大豆基因组成。图 4 提供表示由这些基因编码的蛋白质的相互关系的系统树图。

### 实施例 2 ERECTA 序列分析

本发明的 ZmERECTA 多肽与多种植物品种的 ERECTA 基因有共同特征。比对分析显示得自多个植物品种的基因、保守区和共有序列间之间的关系，参见图 2。

### 实施例 3 转基因植物的转化和再生

用质粒轰击得自温室供体植物的未成熟玉米胚，该质粒含有与干旱诱导型启动子 RAB17 启动子有效连接的 ZmERECTA 序列(Vilardell 等(1990) *Plant Mol Biol* 14:423-432)和赋予除草剂双丙氨磷抗性的选择标记基因 PAT。或者，选择标记基因用单独的质粒提供。转化如下进行。培养基配方见下。

### 靶组织的制备:

剥开玉米穗的苞叶, 在 30% Clorox<sup>®</sup>漂白剂加上 0.5% Micro 洗涤剂中进行表面灭菌 20 分钟, 用无菌水漂洗 2 次。切取未成熟胚, 将胚轴端向下(盾盖端向上), 按 25 个胚/板接种到 560Y 培养基中 4 小时, 然后排列在 2.5 cm 靶区内以备轰击。

### DNA 的制备:

制备包含与泛素启动子有效连接的 ERECTA 序列的质粒载体。如下应用 CaCl<sub>2</sub> 沉淀法将该质粒 DNA 加上含有 PAT 选择标记的质粒 DNA 沉淀在 1.1 μm (平均直径)钨沉淀上:

100 μl 在水中制备的钨颗粒

10 μl 含(1 μg) DNA 的 Tris EDTA 缓冲液(总 DNA 1 μg)

100 μl 2.5 M CaCl<sub>2</sub>

10 μl 0.1 M 亚精胺

依次将每种试剂加到钨颗粒悬液中, 同时保持在多管涡旋仪中。对最终的混合物进行短时超声处理后, 使之在恒定涡旋下温育 10 分钟。在沉淀期之后, 各管进行短时离心, 弃去液体, 用 500 ml 100%乙醇洗涤后, 离心 30 秒钟。再次弃去液体, 将 105 μl 100%乙醇加到最终的钨颗粒沉淀。对于基因枪轰击, 对钨/DNA 颗粒进行短时超声处理, 将 10 μl 点到每个巨载体的中心, 轰击前使之干燥约 2 分钟。

### 基因枪处理:

用基因枪#HE34-1 或#HE34-2 按 4 号水平对样品板进行轰击。所有样品受到单次射击(650 PSI), 从每管制备的颗粒/DNA 中取等分量, 共 10 份。



### 后续处理:

轰击后,使胚在 560Y 培养基中培养 2 天,然后转移到含有 3 mg/升双丙氨酸的 560R 选择培养基中,每 2 周进行继代培养。在大约 10 周选择后,将选择-抗性愈伤组织克隆转移到 288J 培养基以开始使植物再生。在体细胞胚成熟后(2-4 周),将发育良好体细胞胚转移到培养基以便出芽,并转移到有照明的培养室。大约 7-10 天后,将发育中的小植株转移到管中的 272V 无激素培养基内达 7-10 天,直到小植株完全形成。然后将植物移栽到装有盆栽土的平台的插入物(inserts)中(等同于 2.5"花盆),并在生长室中生长 1 周,随后在温室中再生长 1-2 周,然后移栽到典型 600 盆(1.6 加仑)中,生长至成熟。对植物进行监测,并对耐旱性的提高进行评分。测定耐旱性改进的实验是本领域的常规方法,包括例如当与在相同环境条件下的对照玉米植物相比时,在干旱条件下玉米粒-抽穗能力产量(kernel-earring capacity yield)增加。或者,可监测转化植物对分生组织发育的调节(即玉米穗上小穗形成减少)。参见例如 Bruce 等(2002) *Journal of Experimental Botany* 53:1-13。

### 轰击和培养基:

轰击培养基(560Y)含有 4.0 g/l N6 基础盐(SIGMA C-1416)、1.0 ml/l Eriksson 维生素混合物(1000X SIGMA-1511)、0.5 mg/l 盐酸硫胺、120.0 g/l 蔗糖、1.0 mg/l 2,4-D 和 2.88 g/l L-脯氨酸(用 D-I H<sub>2</sub>O 调整体积后,用 KOH 调节至 pH 5.8); 2.0 g/l 脱乙酰吉兰糖胶(Gelrite) (用 D-I H<sub>2</sub>O 调整体积后加入); 以及 8.5 mg/l 硝酸银(将培养基灭菌并冷却至室温后加入)。选择培养基(560R)含有 4.0 g/l N6 基础盐(SIGMA C-1416)、1.0 ml/l Eriksson 维生素混合物(1000X SIGMA-1511)、0.5 mg/l 盐酸硫胺、30.0 g/l 蔗糖和 2.0 mg/l 2,4-D (用 D-I H<sub>2</sub>O 调整体积后用 KOH 调节至 pH 5.8); 3.0 g/l 脱

乙酰吉兰糖胶(用 D-I H<sub>2</sub>O 调整体积后加入); 以及 0.85 mg/l 硝酸银和 3.0 mg/l 双丙氨磷(两者均在将培养基灭菌并冷却至室温后加入)。

植物再生培养基(288J)含有 4.3 g/l MS 盐(GIBCO 11117-074)、5.0 ml/l MS 维生素母液(0.100 g 烟酸、0.02 g/l 盐酸硫胺、0.10 g/l 盐酸吡多醇和 0.40 g/l 甘氨酸,用精制 D-I H<sub>2</sub>O调整体积) (Murashige 和 Skoog (1962) *Physiol. Plant.* 15:473)、100 mg/l 肌醇、0.5 mg/l 玉米素、60 g/l 蔗糖和 1.0 ml/l 0.1 mM 脱落酸(调节至 pH 5.6 后,用精制 D-I H<sub>2</sub>O调整体积); 3.0 g/l 脱乙酰吉兰糖胶(用 D-I H<sub>2</sub>O调整体积后加入); 以及 1.0 mg/l 吲哚乙酸和 3.0 mg/l 双丙氨磷(将培养基灭菌并冷却至 60°C后加入)。无激素培养基(272V)含有 4.3 g/l MS 盐(GIBCO 11117-074)、5.0 ml/l MS 维生素母液(0.100 g/l 烟酸、0.02 g/l 盐酸硫胺、0.10 g/l 盐酸吡多醇和 0.40 g/l 甘氨酸,用精制 D-I H<sub>2</sub>O调整体积)、0.1 g/l 肌醇和 40.0 g/l 蔗糖(调节 pH 至 5.6 后,用精制 D-I H<sub>2</sub>O调整体积); 以及 6 g/l 细菌培养用琼脂(bacto-agar) (用精制 D-I H<sub>2</sub>O调整体积后加入), 灭菌后冷却至 60°C。

#### 实施例 4 土壤杆菌介导的转化

对于用本发明 ZmERECTA 序列的反义序列的土壤杆菌介导的玉米转化, 优选采用 Zhao 的方法(美国专利第 5,981,840 号和 PCT 专利公布号 WO98/32326; 所述专利的内容通过引用结合到本文中)。简单地讲, 从玉米中分离出未成熟胚, 使胚与土壤杆菌悬液接触, 其中该细菌能够将 ERECTA 序列转移到至少一个未成熟胚的至少一个细胞中(步骤 1: 感染步骤)。在这一步骤中, 优选将未成熟胚浸入用于开始接种的土壤杆菌悬液。使胚与土壤杆菌共同培养一段时间(步骤 2: 共培养步骤)。优选使未成熟胚在感染步骤后在固体培养基上培养。在该共培养期后, 包括了任选的“静息”步骤。在该静息步骤中, 在至少一种已知抑制土壤杆菌生长的抗生素存在下无需加入植物转化体的选择剂, 使胚孵育(步骤 3: 静息

步骤)。优选将未成熟胚在有抗生素但无选择剂的固体培养基上培养，为感染细胞清除土壤杆菌，并提供静息阶段。随后，将经接种的胚在含有选择剂的培养基上培养，回收生长中的转化愈伤组织(步骤4:选择步骤)。优选将未成熟胚在有选择剂的固体培养基上培养，所述选择剂可引起转化细胞进行选择生长。然后使愈伤组织再生成植物(步骤5:再生步骤)，优选将在选择性培养基上生长的愈伤组织在固体培养基上培养以再生出植物。对植物进行监测，并对在分生组织发育中的调节进行评分。例如，大小的改变、茎干和花分生组织出现和/或叶、花和/或果实产量的提高。

#### 实施例5 ZmERECTA 的过量表达影响植物生长率、花序发育、器官大小和耐旱性。

ZmERECTA 基因在其中细胞分裂活跃的茎端分生组织和玉米穗特异性分生组织中表达(图1)。该基因的功能与细胞增殖有关。我们通过在组成型启动子控制下使 ZmERECTA 基因过量表达来测试此基因的功能。将产生出转基因植物和非转基因同胞，并对转基因效果进行比较。可通过转基因特异性 RT-PCR 证实表达转基因的转基因植物。预期 ZmERECTA 基因影响植物生长率。与转基因阴性同胞相比，预期转基因阳性植物显示生长率提高、植物和器官大小增大和生物量累积增加。

有研究表明 ERECTA 基因控制拟南芥的花发育。如图1所示，ZmERECTA 基因无疑优先在玉米穗花序分生组织(ear inflorescence meristem)中表达，并在茎端分生组织以略低的水平表达。转基因植物过量表达 ZmERECTA 基因，预期对花序发育产生影响。与非转基因同胞相比，预期转基因植物的玉米穗和玉米粒生长加快，总玉米粒产量提高。

ERECTA 基因调节拟南芥的蒸腾效率，并影响耐旱性。通过用组成型或胁迫诱导型启动子使 ZmERECTA 基因过量表达，来测试转基因植物中 ZmERECTA 基因的这个功能。可对转基因植物及其非转基因同胞

进行干旱胁迫处理，通过测量胁迫环境下的植物生长、生物量累积和产量，来测定其耐受性。预期转基因植物通过影响蒸腾效率而改善了耐旱性，并在胁迫或非胁迫生长条件下促进植物生长，提高谷粒产量。

### 实施例 6 大豆胚转化

如下用含有与泛素启动子有效连接的 ERECTA 序列的质粒轰击大豆胚。为了诱导体细胞胚，从经表面灭菌的大豆栽培种 A2872 未成熟种子中剖出长度 3-5 mm 的子叶，在 26°C、光或暗中，在合适的琼脂培养基中培养 6-10 周。然后切取体细胞胚产生的次生胚，接种到合适的液体培养基中。在对繁殖成早期球形期胚的体细胞胚聚簇进行重复选择后，按下述方法维持悬液。

将大豆胚胎发生悬浮培养物维持在旋转振荡器(150 rpm)上的 35 ml 液体培养基中，26°C，荧光灯 16:8 小时昼/夜周期。将约 35 mg 组织接种到 35 ml 液体培养基中，每两周对培养物进行继代培养。

然后，可通过基因枪轰击方法转化大豆胚胎发生悬浮培养物(Klein 等(1987) *Nature* (London) 327:70-73; 美国专利第 4,945,050 号)。Du Pont 生物射弹 PDS1000/HE 仪(helium retrofit)可用于这些转化。

可用来促进大豆转化的选择标记基因是由以下元件组成的转基因：得自花椰菜花叶病毒的 35S 启动子(Odell 等(1985) *Nature* 313:810-812)、得自质粒 pJR225 (得自大肠杆菌; Gritz 等(1983) *Gene* 25:179-188)的潮霉素磷酸转移酶基因和得自根癌土壤杆菌 Ti 质粒 T-DNA 的胭脂碱合酶基因的 3'区。包含与泛素启动子有效连接的 ERECTA 有义序列的表达盒可被分离成限制片段。然后，可将该片段插入携带标记基因的载体的独特限制位点。

向 50  $\mu$ l 的 60 mg/ml 1  $\mu$ m 金颗粒悬液中加入(依次): 5  $\mu$ l DNA (1  $\mu$ g/ $\mu$ l)、20  $\mu$ l 亚精胺(0.1 M)和 50  $\mu$ l CaCl<sub>2</sub> (2.5 M)。然后搅拌颗粒制备物

3 分钟，在台式微量离心机中离心 10 秒钟，弃去上清液。然后将 DNA 包被的颗粒在 400  $\mu$ l 70%乙醇中洗涤一次后，重新悬浮于 40  $\mu$ l 无水乙醇中。可对 DNA/颗粒悬液进行超声处理三次，每次一秒钟。然后将 5 微升 DNA 包被的金颗粒加到每个巨载体盘中。

将约 300-400 mg 生长两周的悬浮培养物接种到 60x15 mm 空培养皿中，用移液枪从组织中弃去残余液体。对于每次转化实验，通常对约 5-10 块板的组织进行轰击。破膜压力设定为 1100 psi，将室(chamber)抽真空至 28 英寸汞柱。将组织置于距阻挡网(retaining screen)约 3.5 英寸处，轰击三次。轰击后，可将组织分成两半，如上所述，接种回液体中后按上述方法培养。

轰击后 5-7 天，液体培养基可用新鲜培养基更换，轰击后 11-12 天用含有 50 mg/ml 潮霉素的新鲜培养基更换。可以每周更换该选择性培养基。轰击后 7-8 周，可以观察到从未转化的坏死胚胎发生聚簇中生长出绿色的转化组织。取出分离的绿色组织，并接种到各个培养瓶中以产生新的无性繁殖的转化胚胎发生悬浮培养物。每个新株系可作为独立的转化事件来处理。然后，可将这些悬液进行继代培养，并作为未成熟胚聚簇维持，或者随每个体细胞胚的成熟和出芽再生出整株植物。

### 实施例 7 向日葵分生组织的转化

如下用含有与泛素启动子有效连接的 ERECTA 序列的表达盒转化向日葵分生组织(另参见欧洲专利号 EP 0 486233，该专利通过引用结合到本文中，以及 Malone-Schoneberg 等(1994) *Plant Science* 103:199-207)。成熟向日葵种子(向日葵(*Helianthus annuus* L.))用单麦穗脱粒机(single wheathead thresher)去壳。用每 50 ml 溶液 2 滴吐温(Tween) 20 的 20% Clorox<sup>®</sup>漂白溶液中进行表面灭菌种子 30 分钟。用无菌蒸馏水漂洗种子两次。

用 Schrammeijer 等人所述方法的改进方法制备剖开的胚轴外植体 (Schrammeijer 等(1990) *Plant Cell Rep.* 9:55-60)。在表面灭菌步骤后,使种子在蒸馏水中吸胀 60 分钟。然后折断每粒种子的子叶,在胚的轴线平面产生平滑的断面。切除根尖后,将外植体在初叶之间纵切成两半。将两半(切面向上)接种到 GBA 培养基中,培养基由 Murashige 和 Skoog 矿物质元素(Murashige 等(1962) *Physiol. Plant.*, 15:473-497)、Shepard 维生素添加物(Shepard (1980), 载于 *Emergent Techniques for the Genetic Improvement of Crops* (University of Minnesota Press, St. Paul, Minnesota))、40 mg/l 硫酸腺嘌呤、30 g/l 蔗糖、0.5 mg/l 6-苄基-氨基嘌呤(BAP)、0.25 mg/l 吲哚-3-乙酸(IAA)、0.1 mg/l 赤霉素(GA<sub>3</sub>) (pH 5.6)和 8 g/l 植物琼脂(Phytagar)组成。

在土壤杆菌处理之前,对外植体进行微粒轰击(Bidney 等(1992) *Plant Mol. Biol.* 18:301-313)。将 30-40 个外植体成环形置于 60 X 20 mm 板的中心以进行该处理。将大约 4.7 mg 的 1.8 mm 钨微弹重新悬浮于 25 ml 无菌 TE 缓冲液(10 mM Tris HCl, 1 mM EDTA, pH 8.0)中,每次轰击使用 1.5 ml 的等分量。在 PDS 1000<sup>®</sup>颗粒加速装置中,通过置于样品上方 2 cm 的 150 mm nytex 网(nytex screen)轰击各板两次。

在所有转化实验中使用卸甲根瘤土壤杆菌菌株 EHA105。按照 Holsters 等人所述方法(Holsters 等(1978) *Mol. Gen. Genet.* 163:181-187),将包含表达盒的双元质粒载体(binary plasmid vector)通过冻融导入土壤杆菌菌株 EHA105,所述表达盒含有与泛素启动子有效连接的 ERECTA 基因。该质粒还包含卡那霉素选择标记基因(即 *nptII*)。使用于植物转化实验的细菌在液体 YEP 培养基(10 gm/l 酵母提取物、10 gm/l 细菌用蛋白胨(Bactopeptone)和 5 gm/l NaCl, pH 7.0)与细菌菌株和双元质粒维持所需要的合适抗生素中生长过夜(28°C, 100 RPM 连续搅拌)。当悬液 OD<sub>600</sub> 达到约 0.4-0.8 时便可使用。使土壤杆菌细胞沉淀,并按最终 OD<sub>600</sub> 为 0.5

重新悬浮于接种培养基中，所述接种培养基由 12.5 mM MES (pH 5.7)、1 gm/l  $\text{NH}_4\text{Cl}$  和 0.3 gm/l  $\text{MgSO}_4$  组成。

将新鲜轰击的外植体接种到土壤杆菌悬液中，混合，静置 30 分钟。然后将外植体转移到 GBA 培养基上，切面向下，在 26°C 和 18 小时白昼下共培养。共培养 3 天后，将外植体转移到补充了 250 mg/l 头孢噻肟和 50 mg/l 硫酸卡那霉素的 374B (缺乏生长调节剂且蔗糖水平降低 1% 的 GBA 培养基) 中。外植体培养 2-5 周进行选择后，转移到缺乏卡那霉素的新鲜 374B 培养基上继续发育 1-2 周。将具有分化的、尚未出苗适于切割的抗生素抗性生长区的外植体转移到 GBA 培养基(含有 250 mg/l 头孢噻肟)进行第二次为期 3 天的植物激素处理。通过 ELISA，分析得自卡那霉素抗性绿色苗的叶样品中存在的 NPTII，并通过分析对分生组织发育的调节(即大小改变以及茎干和花分生组织出现)，分析存在转基因表达。

将 NPTII 阳性苗嫁接到 Pioneer<sup>®</sup> 杂合 6440 体外生长的向日葵幼苗砧木上。使经表面灭菌的种子在 48-0 培养基(一半浓度的 Murashige 和 Skoog 盐、0.5% 蔗糖、0.3% 脱乙酰吉兰糖胶，pH 5.6) 中萌发，并在用于外植体培养的所述条件下生长。摘除幼苗上部分，将下胚轴纵切开 1 cm，将转化苗插入切口。整个部位用石蜡膜包裹以保护苗。体外培养一周后，可将嫁接的植物移栽到土壤中。将土壤中的嫁接体维持在高湿度条件下，随后慢慢适应温室环境。通过 NPTII ELISA 和/或通过叶提取物的 ERECTA 活性分析来鉴定在温室中成熟的  $T_0$  植物(亲代)的转化区(transformed sector)，而通过少部分干种子子叶的 ERECTA 活性分析来鉴定由 NPTII 阳性  $T_0$  植物收获的转基因种子。

向日葵转化的替代方案可供回收转基因子代而无需使用化学选择压力。将种子去壳，在每 100 ml 溶液加入 2-3 滴吐温 20 的 20% Clorox<sup>®</sup> 漂白溶液中，进行表面灭菌 20 分钟，然后用蒸馏水漂洗 3 次。在 26°C 下，使灭菌种子在用水湿润的滤纸上避光吸胀 20 小时。去除子叶和根，将分

生组织外植体在 374E (由 MS 盐、Shepard 维生素、40 mg/l 硫酸腺嘌呤、3%蔗糖、0.5 mg/l 6-BAP、0.25 mg/l IAA、0.1 mg/l GA 和 0.8%植物琼脂组成的 GBA 培养基(pH 为 5.6))上避光培养 24 小时。摘除初生叶暴露出顶端分生组织, 将大约 40 个外植体顶端圆顶面向上接种到 374M (具有 1.2%植物琼脂的 GBA 培养基)中部 2 cm 的环上, 然后在培养基中避光培养 24 小时。

将约 18.8 mg 的 1.8  $\mu\text{m}$  钨颗粒重新悬浮于 150  $\mu\text{l}$  无水乙醇中。超声处理之后, 将其中 8  $\mu\text{l}$  滴到巨载体表面中心。每板用 650 psi 可裂圆片 (rupture discs)在第一搁板(shelf)中按 26 mm Hg 氦枪真空(helium gun vacuum)下轰击两次。

按上述方法通过冻融将目标质粒导入根癌土壤杆菌菌株 EHA105。将在 50  $\mu\text{g/l}$  卡那霉素存在下的液体 YEP 培养基(10 g/l 酵母提取物, 10 g/l 细菌用蛋白胨, 和 5 g/l NaCl, pH 7.0)中于 28 $^{\circ}\text{C}$ 过夜生长的细菌沉淀重新悬浮于接种培养基(12.5 mM 2-mercaptoethanol、MES、1 g/l  $\text{NH}_4\text{Cl}$  和 0.3 g/l  $\text{MgSO}_4$  (pH 5.7))中, 以达到 OD 600 下终浓度为 4.0。将颗粒轰击的外植体转移到 GBA 培养基(374E)上, 将细菌悬液滴直接接种到分生组织顶端。使外植体在培养基上共培养 4 天, 之后将外植体转移到 374C 培养基(具有 1%蔗糖, 无 BAP、IAA、GA3, 并补充 250  $\mu\text{g/ml}$  头孢噻肟的 GBA)上。在 16 小时白昼和 26 $^{\circ}\text{C}$  孵育条件下, 将小植株在培养基上培养约 2 周。

根据分生组织发育的调节情况(即大小改变及苗和花分生组织出现), 对得自在 374C 培养基中培养 2 周的外植体(长约 2 cm)进行筛选。鉴定阳性(即 ERECTA 表达改变)外植体, 弃去没有显示 ERECTA 活性变化的苗, 将每个阳性外植体再分成结节外植体(nodal explant)。一个结节外植体含有至少一个潜在结节。将该成结区段在 GBA 培养基上培养 3-4 天以促进从每结节中形成副芽(auxiliary bud)。然后将它们转移到 374C



培养基上,使之再发育4周。分离出发育中的芽,在374C培养基上再培养4周。通过合适的蛋白质活性测定法,再次对得自每个重获的新苗的合并叶样品进行筛选。此时,从一个单节重获的阳性苗一般都富含在结节培养之前的最初测定中所检测到的转基因区。

将对于ERECTA表达改变为阳性的再生苗嫁接到Pioneer杂合6440体外生长的向日葵幼苗砧木上。砧木按以下方式制备。将种子去壳,并在每100 ml溶液加入2-3滴吐温20的20%Clorox<sup>®</sup>漂白溶液中进行表面灭菌20分钟,用蒸馏水漂洗3次。使灭菌种子在用水湿润的滤纸上萌发3天,然后将它们转移到48培养基(一半浓度的MS盐、0.5%蔗糖、0.3%脱乙酰吉兰糖胶,pH 5.0)上,在26℃下避光生长3天,然后在16小时白昼培养条件下培育。摘除选出小植株的上部分,纵切开各下胚轴,将转化苗插入V型切口。切口部位用石蜡膜包裹。在培养基上培养1周后,将嫁接植物移栽到土壤中。在头两周内,维持在高湿度条件下以适应温室环境。

### 实施例8 干旱胁迫下玉米营养性能的测定方案

可以利用生长条件、胁迫处理和确定的诊断方法,以高通量方式在FAST-玉米(参见美国专利申请号10/367,417)背景中鉴定出赋予转基因玉米耐旱性的基因。

#### 处理与材料:

采用2个处理组,一个是充分浇水处理(对照),另一个是干旱胁迫处理。

植物材料由T1期的转基因玉米幼苗组成。

### 处理详述:

可以自 10-14 天或移栽后 7 天开始施以胁迫,并在穗丝生长(silking)整个过程中持续。可通过安装定时器的自动系统给花盆浇水,以便在整个干旱胁迫处理期间按田间持水量的 25%或 50%浇水。该胁迫可供鉴定对植物生长以及雌雄穗开花间隔的影响。

**盆栽混合物:** 建议使用 1/3 *Turface MVP* (Profile Products LLC, IL, USA)、1/3 沙和 1/3 *SB300 Universal* (Sun Gro Horticulture, WA, USA) 的混合物。*SB300 Universal* 可用 *Fafard* 特级萌发混合物(*Superfine Germinating Mix*) (Conrad Fafard, Inc., MA, USA)替换。同样,可以降低混合物中沙的比例。因此,最终的盆栽混合物可以是 3/8 (37.5%) *turface*、3/8 (37.5%) *Fafard* 和 1/4 (25%)沙。

**田间持水量测定:** 测定用在一个 S200 花盆中的土壤混合物重量(w1) (减去花盆重量)。在测定 w1 之前,可以将土壤在 100°C 下干燥至恒重(如果土壤混合物的所有组分已经变干的,则不再将其干燥)。给花盆中的土壤浇水至完全饱和,使所有重力水排干。在所有重力水渗出后,测定土壤重量(w2) (减去花盆重量)。田间持水量是保持在土壤中的水份重量,按 w2-w1 得到。技术上可写作烘干土壤重量的百分比。对于我们的目的, w2-w1 便足够。

**胁迫处理:** 在植物生长早期(10 天至 21 天),给充分浇水对照每天浇水 75%田间持水量和给干旱胁迫处理每天浇水 25%田间持水量,两者都在每天上午 10 点按一次日浇水量浇水。当植物长得较大时,增加充分浇水对照的每天浇水量至最大田间持水量,增加干旱胁迫处理的至 50%田间持水量。这种增加以对代表性花盆的重量测量为基础,以确定当植物长得较大时植物的排水程度。

**营养液:** 使用改进的 Hoagland 溶液,用灌溉用自来水按 1/16 稀释。用下列配方配制 20 L 改进的 Hoagland 溶液:

10X 微量营养液	16 ml
KH <sub>2</sub> PO <sub>4</sub> (分子量: 136.02)	22 g
MgSO <sub>4</sub> (分子量: 120.36)	77 g
KNO <sub>3</sub> (分子量: 101.2)	129.5 g
Ca(NO <sub>3</sub> ) <sub>2</sub> ·4H <sub>2</sub> O (分子量: 236.15)	151 g
NH <sub>4</sub> NO <sub>3</sub> (分子量: 80.04)	25.6 g
Sprint 330 (铁螯合物)	32 g

用下列配方配制 1L 10X 微量营养液:

10X 微量营养液	ml/L
H <sub>3</sub> BO <sub>3</sub> - 30 mM	1854
MnCl <sub>2</sub> ·4H <sub>2</sub> O - 10 mM	1980
ZnSO <sub>4</sub> ·7H <sub>2</sub> O - 10 mM	2874
CuSO <sub>4</sub> ·5H <sub>2</sub> O - 1 mM	250
H <sub>2</sub> MoO <sub>4</sub> ·H <sub>2</sub> O - 1 mM	242

使用肥料级 KNO<sub>3</sub>。

注意: 在移栽时或在出苗后, 将半茶匙奥绿肥(*Osmocote*) (NPK 15:9:12)加到花盆是有益的(Scotts Miracle-Gro Company, OH, USA)。

**边界植物:** 在温室玻璃壁附近或者微环境变化潜在因素(例如冷却风扇)附近的育苗台边缘有一排边界植物是十分关键的。

**自动化:** 自动化可如下进行: 采用虹吸装置, 利用有钻孔的 PVC 管向系统性摆放的花盆供水。灌溉时间安排可用定时器控制。

**统计分析:** 可采用因素随机区组设计(用事件或构建方法和胁迫处理作为因素), 将数据输入 Spotfire (Spotfire, Inc., MA, USA), 最终用于每个干旱转基因程序组内的 ANOVA。

重复: 每事件每次处理 8-10 株单株植物。

花盆大小: S200

**观察结果:**

(1) 在整个生长期一周三次进行 LemnaTec 测量以获取植物生长率(LemnaTec GmbH, Wurselen, Germany)。

(2) 在整个胁迫期用 Lemnatec 一周三次测定叶色。

(3) 使用 Hansatech FMS2 仪,一周两次自上午 11 点开始记录叶绿素荧光为 PhiPSII。在胁迫处理第 0 天开始测定直到实验结束,记录最嫩最完全伸展叶的测量值(Hansatech Instruments, Norfolk, England)。

(4) 记录各株植物抽穗和长出穗丝的日期,并计算雌雄穗开花间隔(ASI)。通过确定脱落的生长度日(GDU)至与长出穗丝的生长度日之差,计算出 ASI。脱落时间是当观察到第一次脱落的天数,其中第一次脱落定义为至少一个花药脱落。穗丝生长时间是当观察到第一条穗丝长出的天数,其中第一条穗丝长出定义为出现至少 1 毫米的一根穗丝。

#### 实施例 9 ERECTA 的表达及与耐旱性和农艺性能的关联性。

根据(1)已公布的有关直向同源物的信息,所述直向同源物可表示在胁迫感受或耐受性中的作用,以及(2)表示胁迫相关基因表达的内部谱型分析(profiling study),鉴定出一组蛋白质激酶。它们包括 2 个组分信号转导系统的基因,以及根据胁迫相关表达模式鉴定出的基因 - 包括 ERECTA-A。

#### **ERECTA-A**

**背景信息:** ERECTA 是对农业、尤其是对耐旱性和农艺性能领域具有意义的推定的富含亮氨酸重复序列受体样激酶(LRR-RLK)。拟南芥中,已知该基因通过包括气孔密度、表皮细胞膨胀、叶肉细胞增殖和细胞间接触在内的机制对植物蒸腾效率产生影响。另还了解到该基因影响花序发育。对于该蛋白质,还没有激酶活性实验的报告。

**序列信息:** 克隆出全长 CDS。该基因作图位于染色体箱

(chromosome-bin)位置 6.04, 已知大致在其附近出现干旱 QTL。该基因优选在分生组织、未成熟穗中表达, 在如图 3 中所述的多种组织中进行表达。

## 实施例 10 ERECTA 序列的变体

### A. 不改变编码氨基酸序列的 ERECTA 的变体核苷酸序列

用 ERECTA 核苷酸序列来产生当与相应 SEQ ID NO 的起始未改变的 ORF 核苷酸序列相比时, 可读框的核苷酸序列具有约 70%、75%、80%、85%、90%和 95%核苷酸序列同一性的变体核苷酸序列。用标准密码子表可产生这些功能性变体。虽然变体核苷酸序列改变, 但是由可读框编码的氨基酸序列不改变。

### B. ERECTA 多肽的变体氨基酸序列

制备 ERECTA 多肽的变体氨基酸序列。在本实施例中, 一个氨基酸发生了改变。具体地讲, 检查可读框以确定合适的氨基酸变化。通过查阅蛋白质比对(与得自不同品种的其它直向同源物和其它基因家族成员), 选出待改变的氨基酸。选出一般认为不在高选择压力下(不是高度保守的)且更易于被化学特征相似(即相似的官能侧链)的氨基酸取代的氨基酸。用图 2 中所列的蛋白质比对, 可改变合适的氨基酸。一旦鉴定出靶氨基酸, 则随后进行以下 C 部分中所述方法。用该方法产生具有约 70%、75%、80%、85%、90%和 95%核酸序列同一性的变体。

### C. ERECTA 多肽的其它变体氨基酸序列

在本实施例中, 制备相对于参比蛋白质序列具有 80%、85%、90%和 95%同一性的人工蛋白序列。这一最新努力需要鉴定得自图 2 中所列比对的保守区和可变区, 然后审慎应用氨基酸取代表。下面将更详细论

述这些部分。

很大程度上，根据 ERECTA 蛋白之中或根据其它 ERECTA 多肽之中的保守区来测定被改变的氨基酸序列。根据序列比对，很可能被改变的 ERECTA 多肽各个区用小写字母表示，而保守区则用大写字母表示。要了解的是，下列保守区可进行保守取代但不改变功能。另外，技术人员应当了解的是，本发明 ERECTA 序列的功能性变体在保守结构域中可具有较少的非保守氨基酸变化。

因此产生以 80-85%、85-90%、90-95%和 95-100%同一性为区间不同于源蛋白序列的人工蛋白序列。以这些区间的中点为目标，具有例如+或-1%的自由范围。可通过自定义 Perl script 实现氨基酸取代。取代表参见下表 2。

表 2. 取代表

氨基酸	极相似和最佳取代	改变的分级顺序	注释
I	L,V	1	50:50 取代
L	I,V	2	50:50 取代
V	I,L	3	50:50 取代
A	G	4	
G	A	5	
D	E	6	
E	D	7	
W	Y	8	
Y	W	9	
S	T	10	
T	S	11	
K	R	12	
R	K	13	
N	Q	14	
Q	N	15	
F	Y	16	
M	L	17	第一个甲硫氨酸不能改变
H		不适用	无良好取代
C		不适用	无良好取代
P		不适用	无良好取代

首先，鉴定出蛋白质中不能改变的任何保守氨基酸，“标出”以便与取代隔离开。起始甲硫氨酸当然被自动加到该列表中。然后，进行改变。

H、C 和 P 在任何情况下都不得改变。改变先由异亮氨酸开始，从 N 端扫描到 C 端。然后是亮氨酸，如此沿列表进行直到达到所需目标。

可进行中数取代(Interim number substitution), 从而不引起改变逆转。列表顺序为 1-17, 因此, 从按与需要一样多的异亮氨酸开始, 之后是亮氨酸, 并如此进行到甲硫氨酸。显然按这种方式许多氨基酸不必改变。L、I 和 V 将包括两种替代性最佳取代的 50:50 取代。

书面输出变体氨基酸序列。运用 Perl script 计算百分比同一性。采用该方法, 将产生 ERECTA 多肽的变体, 其与起始未改变的 SEQ ID NO: 27-39 ORF 核苷酸序列的氨基酸同一性约为 80%、85%、90%和 95%。

### 实施例 11 表达 ZmERECTA A 的转基因植物证实生长加快

在田间, 对过量表达 ZmERECTA A 基因的转基因植物进行了评价。对转基因阳性事件及其无效对照与植物和器官生长有关的性状进行了表征。在早期生长季节, 转基因阳性植物显示生长加快, 特别是生长率较高。转基因植物比非转基因对照早约 2 天到达开花期。转基因植物还显示株冠加大, 这与叶尺寸增大有关。叶长度和宽度两者都有加大, 这是引起转基因植物叶面积显著增加的原因。发现与非转基因对照相比, 叶面积增加高达 34%。ZmERECTA A 对生长率和器官大小的转基因作用与其所预计的在控制植物和器官大小及促进细胞增殖中的作用一致 (Shpak, E.D.等, *Plant Cell* (2003) 15:1095-1110; *Development* (2004) 131:1491-501)。这些数据支持这样的见解, 即 ZmERECTA A 可用于控制玉米或其它作物整株植物或特定器官的大小。

测定气孔密度(气孔数/mm<sup>2</sup>), 收集数据。在每叶 3 个区域且每事件 2 片叶(植物)的基础上, 观察到与阴性对照相比, 在所有 5 项事件中, 转基因阳性植物的气孔密度都降低。与转基因阴性对照相比, 降低幅度为 5-22%。转基因对气孔密度的作用与该基因在拟南芥中的作用一致, 拟南芥中该基因显示可降低气孔密度和改善蒸腾效率, 从而改善耐旱性 (Masle, Gilmore 和 Farquhar, (2005) *Nature* 436:866)。该转基因在玉米



气孔密度中的作用的一致性提高了赋予农作物品种耐旱性的潜力。

转基因阳性植物中，叶片毛状体(tricome)生长受到影响，因为转基因阳性植物具有更多的毛状体或更浓密分布的毛状体。如各项事件所测一样，发现毛状体生长提高高达 28%。一般认为毛状体生长通过提供阻水性(hydro-repellency)和叶的反射性质以及对昆虫摄食的物理和化学阻碍，而与植物耐旱性和抗虫性有关(Esau, K. (1977) *Anatomy of Seed Plants*, 第 2 版)。转基因 ERECTA 的这种表型作用可对耐旱性和抗虫性产生影响。

本说明书中的所有出版物和专利申请可表示本发明所属领域的基本技术水平。所有出版物和专利申请均通过引用结合到本文中，其程度与每个独立出版物或专利申请具体而单独指明通过引用结合到本文中一样。

参照各个具体和优选的实施方案和技术对本发明进行了描述。然而，应当了解的是，可进行多种更改和修饰而仍保持在本发明的精神和范围内。

<110> 先锋高级育种国际公司

<120> 改善农作物的植物生长、蒸腾效率和耐旱性的玉米 ERECTA 基因

<130> 2290-PCT

<150> 60/847,304

<151> 2006-09-25

<160> 39

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 3191

<212> DNA

<213> 拟南芥(*Arabidopsis thaliana*)

<400> 1

```

gtttcttctt catggagact tgaagcttt taaagtatat ctaaaaacgc agtcgtttta 60
agactgtgtg tgagaaatgg ctctgtttag agatattggt cttcttgggt ttctcttctg 120
cttgagctta gtactactg tgacttcaga ggaggagca acgttgctgg agattaagaa 180
gtcattcaaa gatgtgaaca atgttcttta tgactggaca acttcacctt cttcggatta 240
ttgtgtctgg agaggtgtgt cttgtgaaaa tgcaccttc aatgttggtg ctcttaattt 300
gtcagatttg aatcttgatg gagaaatctc acctgctatt ggagatctca agagtctctt 360
gtcaattgat ctgcgaggta atcgcttgtc tggacaaatc cctgatgaga ttggtgactg 420
ttcttctttg caaaacttag acttatectt caatgaatta agtggtgaca taccgttttc 480
gatttcgaag ttgaagcaac ttgagcagct gattctgaag aataaccaat tgataggacc 540
gatcccttca acactttcac agattccaaa cctgaaaatt ctggacttgg cacagaataa 600
actcagtggg gagataccaa gacttattta ctggaatgaa gttcttcagt atcttgggtt 660
gcgaggaaac aacttagtgc gtaacatttc tccagatttg tgtcaactga ctggctcttg 720
gtatcttgac gtaagaaaca acagtttgac tggtagtata cctgagacga taggaaattg 780
cactgccttc caggttttgg acttgtccta caatcagcta actggtgaga tcccttttga 840
catcggcttc ctgcaagttg caacattatc attgcaaggc aatcaactct ctgggaagat 900
tccatcagtg attggtctca tgcaagcctt tgcagctcta gatctaagtg gcaacttgtt 960
gagtggatct attctccga ttctcggaaa tcttactttc accgagaaat tgtatttga 1020
cagtaacaag ctgactgggt caattccacc tgagcttga aacatgtcaa aactccatta 1080
cctggaactc aatgataatc atctcacggg tcatatacca ccagagcttg ggaagcttac 1140
tgacttggtt gatctgaatg tggccaacaa tgatctggaa ggacctatac ctgatcatct 1200

```

```

gagctcttgc acaaatctaa acagcttaaa tgttcatggg aacaagtta gtggcactat 1260
accccgagca tttcaaaagc tagaaagtat gacttacctt aatctgtcca gcaacaatat 1320
caaaggtcca atccccggtg agctatctcg tatcggtaac ttagatacat tggatctttc 1380
caacaacaag ataaatggaa tcattccttc ttcccttggg gatitggagc atcttctcaa 1440
gatgaacttg agtagaaatc atataactgg tgtagtcca ggcgactttg gaaatctaag 1500
aagcatcatg gaaatagatc tttcaataa tgatatctct ggccaattc cagaagagct 1560
taaccaatta cagaacataa ttttgctgag actggaaaat aataacctga ctggtaatgt 1620
tggttcatta gccaaactgc tcagtctcac tgtattgaat gtatctcata acaacctcgt 1680
aggtgatatc cctaagaaca ataacttctc aagatittca ccagacagct tcattggcaa 1740
tcctggctct tgcggtagtt ggctaaactc accgigtcat gattctcgtc gaactgtacg 1800
agtgtcaatc tctagagcag ctattcttgg aatagctatt gggggacttg tgatccttct 1860
catggctcta atagcagctt gccgaccgca taatctcct ccttttcttg atggatcact 1920
tgacaaacca gtaacttatt cgacaccgaa gctcgtcac ctcatatga acatggcact 1980
ccacgtttac gaggatata tgagaatgac agagaatcta agtgagaagt atatcattgg 2040
gcacggagca tcaagcactg tatacaaatg tgttttgaag aattgtaaac cggttcgat 2100
taagcggctt tactctcaca acccacagtc aatgaaacag ttgaaacag aactcgagat 2160
gctaagtagc atcaagcaca gaaatcttgt gaggctaca gcttattccc tctctcactt 2220
ggggagtctt ctgttctatg actatttggg aaatggtagc ctctgggac ttcttcatgg 2280
ccctacgaag aaaaagactt ttgattggga cacacggctt aagatagcat atggtgcagc 2340
acaaggttta gcttatctac accatgactg tagtccaagg atcattcaca gagacgtgaa 2400
gtcgtccaac attctcttgg acaagactt agaggctcgt ttgacagatt ttggaatagc 2460
gaaaagcttg tgtgtgtcaa agtcacatac ttcaacttac gtgatgggca cgataggtta 2520
catagacccc gagtatgctc gcacttcacg gctcactgag aaatccgatg tctacagtta 2580
tggaatagtc cttcttgagt tgtaaccgg aaggaaagcc gttgatgacg aatccaatct 2640
ccaccatctg ataatgtcaa agacggggaa caatgaagtg atggaaatgg cagatccaga 2700
catcacatcg acgtgtaaag atctcgggtg ggtgaagaaa gttttccaac tggcactcct 2760
atgcacaaa agacacccga atgatcgacc cacaatgcac caggtgactc gtgttctcgg 2820
cagttttatg ctatcggaac aaccacctgc tgcgactgac acgtcagcga cgctggctgg 2880
ttcgtgctac gtcgatgagt atgcaaatct caagactcct cattctgtca attgctcttc 2940
catgagtgct tctgatgctc aactgtttct tcggtttgga caagttattt ctcagaacag 3000
tgagttagtt ttcgttagga ggagaatctt taaaacggta tcttttcggt gcgttaagct 3060
gtagaaaaa ttaatgtctc atgtaaagta ttatgactg ccttattatt attagacaag 3120
tgtgtggtgt gaatatgtct tcagactggc acttagactt cctataagtt cttgcctatc 3180
taagtttttc t 3191

```

<210> 2

<211> 976

<212> PRT

<213> 拟南芥(*Arabidopsis thaliana*)

<400> 2

```

Met Ala Leu Phe Arg Asp Ile Val Leu Leu Gly Phe Leu Phe Cys Leu
  1           5           10           15
Ser Leu Val Ala Thr Val Thr Ser Glu Glu Gly Ala Thr Leu Leu Glu
           20           25           30

```

Ile Lys Lys Ser Phe Lys Asp Val Asn Asn Val Leu Tyr Asp Trp Thr  
 35 40 45  
 Thr Ser Pro Ser Ser Asp Tyr Cys Val Trp Arg Gly Val Ser Cys Glu  
 50 55 60  
 Asn Val Thr Phe Asn Val Val Ala Leu Asn Leu Ser Asp Leu Asn Leu  
 65 70 75 80  
 Asp Gly Glu Ile Ser Pro Ala Ile Gly Asp Leu Lys Ser Leu Leu Ser  
 85 90 95  
 Ile Asp Leu Arg Gly Asn Arg Leu Ser Gly Gln Ile Pro Asp Glu Ile  
 100 105 110  
 Gly Asp Cys Ser Ser Leu Gln Asn Leu Asp Leu Ser Phe Asn Glu Leu  
 115 120 125  
 Ser Gly Asp Ile Pro Phe Ser Ile Ser Lys Leu Lys Gln Leu Glu Gln  
 130 135 140  
 Leu Ile Leu Lys Asn Asn Gln Leu Ile Gly Pro Ile Pro Ser Thr Leu  
 145 150 155 160  
 Ser Gln Ile Pro Asn Leu Lys Ile Leu Asp Leu Ala Gln Asn Lys Leu  
 165 170 175  
 Ser Gly Glu Ile Pro Arg Leu Ile Tyr Trp Asn Glu Val Leu Gln Tyr  
 180 185 190  
 Leu Gly Leu Arg Gly Asn Asn Leu Val Gly Asn Ile Ser Pro Asp Leu  
 195 200 205  
 Cys Gln Leu Thr Gly Leu Trp Tyr Phe Asp Val Arg Asn Asn Ser Leu  
 210 215 220  
 Thr Gly Ser Ile Pro Glu Thr Ile Gly Asn Cys Thr Ala Phe Gln Val  
 225 230 235 240  
 Leu Asp Leu Ser Tyr Asn Gln Leu Thr Gly Glu Ile Pro Phe Asp Ile  
 245 250 255  
 Gly Phe Leu Gln Val Ala Thr Leu Ser Leu Gln Gly Asn Gln Leu Ser  
 260 265 270  
 Gly Lys Ile Pro Ser Val Ile Gly Leu Met Gln Ala Leu Ala Val Leu  
 275 280 285  
 Asp Leu Ser Gly Asn Leu Leu Ser Gly Ser Ile Pro Pro Ile Leu Gly  
 290 295 300  
 Asn Leu Thr Phe Thr Glu Lys Leu Tyr Leu His Ser Asn Lys Leu Thr  
 305 310 315 320  
 Gly Ser Ile Pro Pro Glu Leu Gly Asn Met Ser Lys Leu His Tyr Leu  
 325 330 335  
 Glu Leu Asn Asp Asn His Leu Thr Gly His Ile Pro Pro Glu Leu Gly  
 340 345 350  
 Lys Leu Thr Asp Leu Phe Asp Leu Asn Val Ala Asn Asn Asp Leu Glu  
 355 360 365  
 Gly Pro Ile Pro Asp His Leu Ser Ser Cys Thr Asn Leu Asn Ser Leu  
 370 375 380  
 Asn Val His Gly Asn Lys Phe Ser Gly Thr Ile Pro Arg Ala Phe Gln

385                                    390                                    395                                    400  
 Lys Leu Glu Ser Met Thr Tyr Leu Asn Leu Ser Ser Asn Asn Ile Lys  
    405                                    410                                    415  
 Gly Pro Ile Pro Val Glu Leu Ser Arg Ile Gly Asn Leu Asp Thr Leu  
    420                                    425                                    430  
 Asp Leu Ser Asn Asn Lys Ile Asn Gly Ile Ile Pro Ser Ser Leu Gly  
    435                                    440                                    445  
 Asp Leu Glu His Leu Leu Lys Met Asn Leu Ser Arg Asn His Ile Thr  
    450                                    455                                    460  
 Gly Val Val Pro Gly Asp Phe Gly Asn Leu Arg Ser Ile Met Glu Ile  
 465                                    470                                    475                                    480  
 Asp Leu Ser Asn Asn Asp Ile Ser Gly Pro Ile Pro Glu Glu Leu Asn  
    485                                    490                                    495  
 Gln Leu Gln Asn Ile Ile Leu Leu Arg Leu Glu Asn Asn Asn Leu Thr  
    500                                    505                                    510  
 Gly Asn Val Gly Ser Leu Ala Asn Cys Leu Ser Leu Thr Val Leu Asn  
    515                                    520                                    525  
 Val Ser His Asn Asn Leu Val Gly Asp Ile Pro Lys Asn Asn Asn Phe  
 530                                    535                                    540  
 Ser Arg Phe Ser Pro Asp Ser Phe Ile Gly Asn Pro Gly Leu Cys Gly  
 545                                    550                                    555                                    560  
 Ser Trp Leu Asn Ser Pro Cys His Asp Ser Arg Arg Thr Val Arg Val  
    565                                    570                                    575  
 Ser Ile Ser Arg Ala Ala Ile Leu Gly Ile Ala Ile Gly Gly Leu Val  
    580                                    585                                    590  
 Ile Leu Leu Met Val Leu Ile Ala Ala Cys Arg Pro His Asn Pro Pro  
    595                                    600                                    605  
 Pro Phe Leu Asp Gly Ser Leu Asp Lys Pro Val Thr Tyr Ser Thr Pro  
    610                                    615                                    620  
 Lys Leu Val Ile Leu His Met Asn Met Ala Leu His Val Tyr Glu Asp  
 625                                    630                                    635                                    640  
 Ile Met Arg Met Thr Glu Asn Leu Ser Glu Lys Tyr Ile Ile Gly His  
    645                                    650                                    655  
 Gly Ala Ser Ser Thr Val Tyr Lys Cys Val Leu Lys Asn Cys Lys Pro  
    660                                    665                                    670  
 Val Ala Ile Lys Arg Leu Tyr Ser His Asn Pro Gln Ser Met Lys Gln  
    675                                    680                                    685  
 Phe Glu Thr Glu Leu Glu Met Leu Ser Ser Ile Lys His Arg Asn Leu  
    690                                    695                                    700  
 Val Ser Leu Gln Ala Tyr Ser Leu Ser His Leu Gly Ser Leu Leu Phe  
 705                                    710                                    715                                    720  
 Tyr Asp Tyr Leu Glu Asn Gly Ser Leu Trp Asp Leu Leu His Gly Pro  
    725                                    730                                    735  
 Thr Lys Lys Lys Thr Leu Asp Trp Asp Thr Arg Leu Lys Ile Ala Tyr  
    740                                    745                                    750

Gly Ala Ala Gln Gly Leu Ala Tyr Leu His His Asp Cys Ser Pro Arg  
           755                          760                          765  
 Ile Ile His Arg Asp Val Lys Ser Ser Asn Ile Leu Leu Asp Lys Asp  
           770                          775                          780  
 Leu Glu Ala Arg Leu Thr Asp Phe Gly Ile Ala Lys Ser Leu Cys Val  
 785                          790                          795                          800  
 Ser Lys Ser His Thr Ser Thr Tyr Val Met Gly Thr Ile Gly Tyr Ile  
                           805                          810                          815  
 Asp Pro Glu Tyr Ala Arg Thr Ser Arg Leu Thr Glu Lys Ser Asp Val  
                           820                          825                          830  
 Tyr Ser Tyr Gly Ile Val Leu Leu Glu Leu Leu Thr Arg Arg Lys Ala  
           835                          840                          845  
 Val Asp Asp Glu Ser Asn Leu His His Leu Ile Met Ser Lys Thr Gly  
           850                          855                          860  
 Asn Asn Glu Val Met Glu Met Ala Asp Pro Asp Ile Thr Ser Thr Cys  
 865                          870                          875                          880  
 Lys Asp Leu Gly Val Val Lys Lys Val Phe Gln Leu Ala Leu Leu Cys  
                           885                          890                          895  
 Thr Lys Arg Gln Pro Asn Asp Arg Pro Thr Met His Gln Val Thr Arg  
           900                          905                          910  
 Val Leu Gly Ser Phe Met Leu Ser Glu Gln Pro Pro Ala Ala Thr Asp  
           915                          920                          925  
 Thr Ser Ala Thr Leu Ala Gly Ser Cys Tyr Val Asp Glu Tyr Ala Asn  
           930                          935                          940  
 Leu Lys Thr Pro His Ser Val Asn Cys Ser Ser Met Ser Ala Ser Asp  
 945                          950                          955                          960  
 Ala Gln Leu Phe Leu Arg Phe Gly Gln Val Ile Ser Gln Asn Ser Glu  
                           965                          970                          975

<210> 3

<211> 2901

<212> DNA

<213> 拟南芥(*Arabidopsis thaliana*)

<400> 3

atgaaggaga agatgcagcg aatggtttta tctttagcaa tgggtgggttt tatggttttt 60  
 ggtgttgctt cgctatgaa caacgaaggg aaagctctga tggcgataaa aggctctttc 120  
 agcaacttag tgaatatgct tttggattgg gacgatgttc acaacagtga cttgtgttct 180  
 tggcgagggtg ttttctgcca caacgtagc tactccgttg tctctctgaa tttgtccagt 240  
 ctgaatcttg gaggggagat atctccagct attggagacc tacggaattt gcaatcaata 300  
 gacttgcaag gtaataaact agcagggtcaa attocagatg agattggaaa ctgtgcttct 360  
 cttgtttatc tggatttgct cgagaatctg ttatatggag acataacctt ctcaatctct 420  
 aaactcaagc agcttgaaac tctgaatctg aagaacaatc agctcacagg tcctgtacca 480  
 gcaaccttaa cccagattcc aaaccttaag agacttgatc ttgctggcaa tcatctaacg 540

ggtgagatat cgagattgct ttactggaat gaagttttgc agtatcttgg attacgaggg 600  
 aatatgttga ctggaacggt atcttctgat atgtgtcagc taaccggttt gtggtacttt 660  
 gatgtgagag gaaataatct aactggaacc atccccgaga gcatcgaaa tgcacaagc 720  
 tttcaaatcc tggacatatc ttataatcag ataacaggag agattcctta caatatcggc 780  
 ttctccaag ttgctactct gtcacttcaa ggaaacagat tgacgggtag aattccagaa 840  
 gttattggtc taatgcaggc tcttgctgtt ttggatttga gtgacaatga gcttgttgg 900  
 cctatcccac cgatacttgg caatctctca tttaccggaa agttgtatct ccatggcaat 960  
 atgctcactg gtccaatccc ctctgagctt gggaatatgt cacgtctcag ctatttgcag 1020  
 ctaaacgaca ataaactagt gggaaactatt ccacctgagc ttggaaagct ggagcaattg 1080  
 tttgaactga atcttgccaa caaccgttta gtagggccca taccatcaa cattagtcca 1140  
 tgtgcagcct tgaatcaatt caatgttcat gggaacctct tgagtggatc tattccactg 1200  
 gcgtttcgca atctcgggag cttgacttat ctgaaatctt cgtcgaacaa tttcaaggga 1260  
 aaaataccag ttgagcttgg acatataatc aatcttgaca aactagatct gtctggcaat 1320  
 aacttctcag ggtctatacc attaacgctt ggcgatcttg aacaccttct catattaaat 1380  
 cttagcagaa accatcttag tggacaatta cctgcagagt ttgggaacct tccaagcatt 1440  
 cagatgattg atglatcatt caatctgctc tccggagtta ttccaactga acttggccaa 1500  
 ttgcagaatt taaactctt aatattgaac aacaacaagc ttcatgggaa aattccagat 1560  
 cagcttacga actgcttcac tctgtcaat ctgaaatgtc cctcaacaa tctctcggg 1620  
 atagtccac caatgaaaaa ctctcacgt tttgctccag ccagctttgt tggaaatcca 1680  
 taictttgtg gaaactgggt tggatctatt tgtggctctt taccgaaatc tggagtattc 1740  
 tccagagggt ctttgcactg cattgttctt ggcgatcaca ctctctatg tatgattttc 1800  
 cttgcagttt acaaatcaat gcagcagaag aagattctac aaggctcctc aaaacaagct 1860  
 gaagggttaa ccaagctagt gattctccac atggacatgg caattcatac atttgatgat 1920  
 atcatgagag tgactgagaa tctaacgaa aagtttataa ttggatatgg tgcttctagc 1980  
 acggtataca aatgtgcatt aaaaagttcc cgacctatg ccattaagcg actctacaat 2040  
 cagtatccgc ataacttggg ggaatttggag acagaacttg agaccattgg gagcattagg 2100  
 cacagaacaa tagtcagctt geatggatat gccttgtctc ctactggcaa ccttctttc 2160  
 tatgactaca tggaaaatgg atcactttgg gaccttctc atgggtcatt gaagaaagt 2220  
 aagcttgatt gggagacaag gttgaagata gcggttggag ctgcacaagg actagcctat 2280  
 cttcaccacg attgtactcc tccaatcatt caccgtgaca tcaagtcac gaacatactt 2340  
 cttgatgaga atttcaagc acatttatct gatttcggga ttgctaagag cataccagct 2400  
 agcaaaacc atgcctcgac ttatgttttg ggaacaattg gttatataga cccagagtat 2460  
 gctcgtactt cacgaatcaa tgagaaatcc gatatatata gcttcggtat tgttcttctt 2520  
 gagcttctca ctgggaagaa agcagtgatg aacgaagcta acttgcatac actgatattg 2580  
 tcaaaggctg atgataatc tgtgatggaa gcagttgatc cagaggttac tgtgacttgt 2640  
 atggacttgg gacatatcag gaagacatt cagctggctc tcttatgcac aaagcgaaac 2700  
 ccttagaga gaccacaat gcttgaagtc tctagggttc tgctctctct tgtccatct 2760  
 ctgcaagtag caaagaagct acctctctt gatcactcaa ccaaaaagct gcagcaagag 2820  
 aatgaagtta ggaatctga tgcagaagca tctcaatggt ttgttcagtt ccgtgaagtc 2880  
 atctccaaa gtagcatata a 2901

<210> 4

<211> 966

<212> PRT

<213> 拟南芥(*Arabidopsis thaliana*)

&lt;400&gt; 4

Met Lys Glu Lys Met Gln Arg Met Val Leu Ser Leu Ala Met Val Gly  
 1 5 10 15  
 Phe Met Val Phe Gly Val Ala Ser Ala Met Asn Asn Glu Gly Lys Ala  
 20 25 30  
 Leu Met Ala Ile Lys Gly Ser Phe Ser Asn Leu Val Asn Met Leu Leu  
 35 40 45  
 Asp Trp Asp Asp Val His Asn Ser Asp Leu Cys Ser Trp Arg Gly Val  
 50 55 60  
 Phe Cys Asp Asn Val Ser Tyr Ser Val Val Ser Leu Asn Leu Ser Ser  
 65 70 75 80  
 Leu Asn Leu Gly Gly Glu Ile Ser Pro Ala Ile Gly Asp Leu Arg Asn  
 85 90 95  
 Leu Gln Ser Ile Asp Leu Gln Gly Asn Lys Leu Ala Gly Gln Ile Pro  
 100 105 110  
 Asp Glu Ile Gly Asn Cys Ala Ser Leu Val Tyr Leu Asp Leu Ser Glu  
 115 120 125  
 Asn Leu Leu Tyr Gly Asp Ile Pro Phe Ser Ile Ser Lys Leu Lys Gln  
 130 135 140  
 Leu Glu Thr Leu Asn Leu Lys Asn Asn Gln Leu Thr Gly Pro Val Pro  
 145 150 155 160  
 Ala Thr Leu Thr Gln Ile Pro Asn Leu Lys Arg Leu Asp Leu Ala Gly  
 165 170 175  
 Asn His Leu Thr Gly Glu Ile Ser Arg Leu Leu Tyr Trp Asn Glu Val  
 180 185 190  
 Leu Gln Tyr Leu Gly Leu Arg Gly Asn Met Leu Thr Gly Thr Leu Ser  
 195 200 205  
 Ser Asp Met Cys Gln Leu Thr Gly Leu Trp Tyr Phe Asp Val Arg Gly  
 210 215 220  
 Asn Asn Leu Thr Gly Thr Ile Pro Glu Ser Ile Gly Asn Cys Thr Ser  
 225 230 235 240  
 Phe Gln Ile Leu Asp Ile Ser Tyr Asn Gln Ile Thr Gly Glu Ile Pro  
 245 250 255  
 Tyr Asn Ile Gly Phe Leu Gln Val Ala Thr Leu Ser Leu Gln Gly Asn  
 260 265 270  
 Arg Leu Thr Gly Arg Ile Pro Glu Val Ile Gly Leu Met Gln Ala Leu  
 275 280 285  
 Ala Val Leu Asp Leu Ser Asp Asn Glu Leu Val Gly Pro Ile Pro Pro  
 290 295 300  
 Ile Leu Gly Asn Leu Ser Phe Thr Gly Lys Leu Tyr Leu His Gly Asn  
 305 310 315 320  
 Met Leu Thr Gly Pro Ile Pro Ser Glu Leu Gly Asn Met Ser Arg Leu  
 325 330 335  
 Ser Tyr Leu Gln Leu Asn Asp Asn Lys Leu Val Gly Thr Ile Pro Pro



340 345 350  
 Glu Leu Gly Lys Leu Glu Gln Leu Phe Glu Leu Asn Leu Ala Asn Asn  
 355 360 365  
 Arg Leu Val Gly Pro Ile Pro Ser Asn Ile Ser Ser Cys Ala Ala Leu  
 370 375 380  
 Asn Gln Phe Asn Val His Gly Asn Leu Leu Ser Gly Ser Ile Pro Leu  
 385 390 395 400  
 Ala Phe Arg Asn Leu Gly Ser Leu Thr Tyr Leu Asn Leu Ser Ser Asn  
 405 410 415  
 Asn Phe Lys Gly Lys Ile Pro Val Glu Leu Gly His Ile Ile Asn Leu  
 420 425 430  
 Asp Lys Leu Asp Leu Ser Gly Asn Asn Phe Ser Gly Ser Ile Pro Leu  
 435 440 445  
 Thr Leu Gly Asp Leu Glu His Leu Leu Ile Leu Asn Leu Ser Arg Asn  
 450 455 460  
 His Leu Ser Gly Gln Leu Pro Ala Glu Phe Gly Asn Leu Arg Ser Ile  
 465 470 475 480  
 Gln Met Ile Asp Val Ser Phe Asn Leu Leu Ser Gly Val Ile Pro Thr  
 485 490 495  
 Glu Leu Gly Gln Leu Gln Asn Leu Asn Ser Leu Ile Leu Asn Asn Asn  
 500 505 510  
 Lys Leu His Gly Lys Ile Pro Asp Gln Leu Thr Asn Cys Phe Thr Leu  
 515 520 525  
 Val Asn Leu Asn Val Ser Phe Asn Asn Leu Ser Gly Ile Val Pro Pro  
 530 535 540  
 Met Lys Asn Phe Ser Arg Phe Ala Pro Ala Ser Phe Val Gly Asn Pro  
 545 550 555 560  
 Tyr Leu Cys Gly Asn Trp Val Gly Ser Ile Cys Gly Pro Leu Pro Lys  
 565 570 575  
 Ser Arg Val Phe Ser Arg Gly Ala Leu Ile Cys Ile Val Leu Gly Val  
 580 585 590  
 Ile Thr Leu Leu Cys Met Ile Phe Leu Ala Val Tyr Lys Ser Met Gln  
 595 600 605  
 Gln Lys Lys Ile Leu Gln Gly Ser Ser Lys Gln Ala Glu Gly Leu Thr  
 610 615 620  
 Lys Leu Val Ile Leu His Met Asp Met Ala Ile His Thr Phe Asp Asp  
 625 630 635 640  
 Ile Met Arg Val Thr Glu Asn Leu Asn Glu Lys Phe Ile Ile Gly Tyr  
 645 650 655  
 Gly Ala Ser Ser Thr Val Tyr Lys Cys Ala Leu Lys Ser Ser Arg Pro  
 660 665 670  
 Ile Ala Ile Lys Arg Leu Tyr Asn Gln Tyr Pro His Asn Leu Arg Glu  
 675 680 685  
 Phe Glu Thr Glu Leu Glu Thr Ile Gly Ser Ile Arg His Arg Asn Ile  
 690 695 700



aatgtcactc ccccgatgcc tgtgcgcagc tcagtggcca tgacgacgac ggccgcccgt 240  
gctctcgccg ccctcgtgct cgtcaccgcc gcccgcccg cgcgcccggt cgccgacgat 300  
ggggcggcgc tgggtggagat caagaagtcc ttccgcaacg tcggcaacgt actgtacgat 360  
tgggcccggc acgactactg ctcttgccgc ggcgtcctgt gcgacaacgt cacgttcgcc 420  
gtcggcggc tcaacctctc tggcctcaac cttgagggtg agatctctcc agccgtcggc 480  
agcctcaaga gcctcgtctc catcgacctc aagtcaaatg gcctatccgg gcagatccct 540  
gatgagattg gcgattgttc gtcacttagg acgctggact ttctattcaa caacttggac 600  
ggcgacatac cttttcaat atcaaagctg aagcacctgg agaacttgat attgaagaac 660  
aaccggctga ttggtgcgat ccctcaaca ttgtcacagc tcccaaattt gaagattctg 720  
gacttggcac aaaacaaact gactggggag ataccgaggc ttatctattg gaacgaggtt 780  
cttcaatact tgggtttgcg gggaaatcat ttagaaggaa gcctctctcc tgatatgtgc 840  
cagcttactg gcctttggta ctttgatgtg aagaacaata gtttgactgg ggcgatacca 900  
gacaccattg ggaactgtac aagttttcag gtcttgattg tgtcttacia cgcctttact 960  
ggaccaatcc cattcaacat tggtttccta caagtggcta cactatcctt gcaagggaac 1020  
aagttcactg goccaattcc ttcagtaatt ggccttatgc aggcctctgc tgccttagat 1080  
ctgagttaca accaattatc tggcctata ccatctatac taggcaactt gacatacact 1140  
gagaagctgt acatgcaagg caacaggtta actggatcga taccaccaga gctaggaat 1200  
atgcaacac ttattacct agaactgaat gataatcaac ttactgggtc aattccacca 1260  
gagcttgaa ggctaacagg ctgtttgac ctgaaccctg cgaataacca ccttgaagga 1320  
ccaattcctg acaacctaaag ttcagtgtgt aatctcaata gcttcaatgc ttatggcaac 1380  
aagttaaatg gaaccattcc tcgttcgctg cggaaacttg aaagcatgac ctatttfaat 1440  
ctttcatcaa atttcataag tggttctatt cctattgagc tatcaaggat caacaatttg 1500  
gacacgttgg gcttatcctg taacatgatg acgggtccaa ttccatcatc cattggcaac 1560  
ctagagcacc tattgaggct taacttgagc aagaatgac tagttggatt catccctgcg 1620  
gagtttggta atttgggaag tgcacaggag attgatttat cctataatca tcttgggtgt 1680  
ctgattcctc aagaacttgg aatgctgcaa aacctgatgt tgctaaaact ggaaaacaac 1740  
aatataactg gcgatgtctc ttctctgatg aactgcttca gcctcaatat cttaaatgtg 1800  
tcatacaata atttagctgg tgcgtccct actgacaaca acttcacacg gttttcacat 1860  
gacagctttt taggtaatcc tggactctgt ggatattggc ttggttcttc atgctgttcc 1920  
actggccacc gagacaaacc gccaatctca aaggctgcca taattgggtg tgctgtgggt 1980  
ggacttgta tcctctgat gatcttagta gctgtatgca ggccacacca tccacctgct 2040  
tttaaagatg ccactgtaag caagccagtg agcaatggtc cacccaagct gatgatcctt 2100  
catatgaaca tggctcttca tgcctttgat gatataatga ggatgactga gaacttgagt 2160  
gagaataca tcattggata cggggcatca agtacagttt ataaatgtgt tctaaagaat 2220  
tgcaaacacc tggcaataaa aaagctgtat gccactacc ctacagacct taaggaattt 2280  
gaaactgagc tcgagactgt tggtagcatc aaacaccgga atctagtcag cttcaaggg 2340  
tactcgttgt cacctgttgg gaacctcctc ttttatgatt atatggagag tggcagctta 2400  
tgggatgttt tacatgaagg ctcatccaag gagaacaac ttgactgggt gactcgccta 2460  
cggatcctc ttggtgcagc tcaaggcctc gcttaccttc accatgactg cagcccacga 2520  
ataattcacc gggacgtaaa atcaaagaat atactcctc acaaagatta tgaggcccat 2580  
cttacagact tcggcatcgc taagagctta tgtgtctcga agactcacac gtcaacctac 2640  
gtcatgggca ctattggta cattgatccc gagtacgccc gcacctccg cctcaacgag 2700  
aagtctgatg tctacagcta cggcatcggt ctgctggagc tgctgaccgg caagaagcca 2760  
gtggacaacg agtgcaatct ccatcacttg atcctatcga agacggcgag caacgaggtc 2820  
atggagacgg tggacccgga cgtgggagac acctgcaagg acctgggcca ggtgaagaag 2880

ctgttccagc tggcgtcct ctgcaccaag cggcagccct cggaccggcc gacgatgcac 2940  
 gagtggtg gcgtccttga ctgctggtg aacccggagc cgcgcgcga gccgcagcag 3000  
 cagcagcaga aggcgcacgc gcaccaccag ctgccgccgc agccgtcgcc gccggcctac 3060  
 gtcgacgagt acgtcagcct gggggcacc ggcgccctct cctgcgcaa ctcgtccagc 3120  
 acctcggacg ccgagctggt cctcaagttc ggcgaggcca tctcgcagaa catggtgtag 3180  
 ggaagacgt agacgttcgg tgaggcgcct tgagtgaggc cgattgcagg gggagtagtt 3240  
 tgactgacat tttgtgggac gcagcgcagg agattaacat gggactcagt agctagggtg 3300  
 ttgttagctg taaaaaagt catgtgacgc aagagcagcg gagcttcttc ctcttcttta 3360  
 tccccctcc ccatttttct ttggtggtct aacttactag gaggctgtat tgatccatca 3420  
 tcctctctcc cgttctctg cttgatgac ttttgtgact ttcgcggtcc tctgtattga 3480  
 tccatgatct tttgtgacct ctgctttatt ttctgggac tttacctcag aac 3533

<210> 6

<211> 994

<212> PRT

<213> 玉米(Zea mays)

<400> 6

Met Pro Val Arg Ser Ser Val Ala Met Thr Thr Thr Ala Ala Arg Ala  
 1 5 10 15  
 Leu Ala Ala Leu Val Leu Val Thr Ala Ala Ala Ala Ala Ala Ala Val  
 20 25 30  
 Ala Asp Asp Gly Ala Ala Leu Val Glu Ile Lys Lys Ser Phe Arg Asn  
 35 40 45  
 Val Gly Asn Val Leu Tyr Asp Trp Ala Gly Asp Asp Tyr Cys Ser Trp  
 50 55 60  
 Arg Gly Val Leu Cys Asp Asn Val Thr Phe Ala Val Ala Ala Leu Asn  
 65 70 75 80  
 Leu Ser Gly Leu Asn Leu Glu Gly Glu Ile Ser Pro Ala Val Gly Ser  
 85 90 95  
 Leu Lys Ser Leu Val Ser Ile Asp Leu Lys Ser Asn Gly Leu Ser Gly  
 100 105 110  
 Gln Ile Pro Asp Glu Ile Gly Asp Cys Ser Ser Leu Arg Thr Leu Asp  
 115 120 125  
 Phe Ser Phe Asn Asn Leu Asp Gly Asp Ile Pro Phe Ser Ile Ser Lys  
 130 135 140  
 Leu Lys His Leu Glu Asn Leu Ile Leu Lys Asn Asn Arg Leu Ile Gly  
 145 150 155 160  
 Ala Ile Pro Ser Thr Leu Ser Gln Leu Pro Asn Leu Lys Ile Leu Asp  
 165 170 175  
 Leu Ala Gln Asn Lys Leu Thr Gly Glu Ile Pro Arg Leu Ile Tyr Trp  
 180 185 190  
 Asn Glu Val Leu Gln Tyr Leu Gly Leu Arg Gly Asn His Leu Glu Gly  
 195 200 205  
 Ser Leu Ser Pro Asp Met Cys Gln Leu Thr Gly Leu Trp Tyr Phe Asp

210	215	220
Val Lys Asn Asn Ser Leu Thr Gly Ala Ile Pro Asp Thr Ile Gly Asn		
225	230	235
Cys Thr Ser Phe Gln Val Leu Asp Leu Ser Tyr Asn Arg Phe Thr Gly		
	245	250
Pro Ile Pro Phe Asn Ile Gly Phe Leu Gln Val Ala Thr Leu Ser Leu		
	260	265
Gln Gly Asn Lys Phe Thr Gly Pro Ile Pro Ser Val Ile Gly Leu Met		
	275	280
Gln Ala Leu Ala Val Leu Asp Leu Ser Tyr Asn Gln Leu Ser Gly Pro		
	290	295
Ile Pro Ser Ile Leu Gly Asn Leu Thr Tyr Thr Glu Lys Leu Tyr Met		
305	310	315
Gln Gly Asn Arg Leu Thr Gly Ser Ile Pro Pro Glu Leu Gly Asn Met		
	325	330
Ser Thr Leu His Tyr Leu Glu Leu Asn Asp Asn Gln Leu Thr Gly Ser		
	340	345
Ile Pro Pro Glu Leu Gly Arg Leu Thr Gly Leu Phe Asp Leu Asn Pro		
	355	360
Ala Asn Asn His Leu Glu Gly Pro Ile Pro Asp Asn Leu Ser Ser Cys		
	370	375
Val Asn Leu Asn Ser Phe Asn Ala Tyr Gly Asn Lys Leu Asn Gly Thr		
385	390	395
Ile Pro Arg Ser Leu Arg Lys Leu Glu Ser Met Thr Tyr Leu Asn Leu		
	405	410
Ser Ser Asn Phe Ile Ser Gly Ser Ile Pro Ile Glu Leu Ser Arg Ile		
	420	425
Asn Asn Leu Asp Thr Leu Gly Leu Ser Cys Asn Met Met Thr Gly Pro		
	435	440
Ile Pro Ser Ser Ile Gly Asn Leu Glu His Leu Leu Arg Leu Asn Leu		
	450	455
Ser Lys Asn Asp Leu Val Gly Phe Ile Pro Ala Glu Phe Gly Asn Leu		
465	470	475
Gly Ser Val Met Glu Ile Asp Leu Ser Tyr Asn His Leu Gly Gly Leu		
	485	490
Ile Pro Gln Glu Leu Gly Met Leu Gln Asn Leu Met Leu Leu Lys Leu		
	500	505
Glu Asn Asn Asn Ile Thr Gly Asp Val Ser Ser Leu Met Asn Cys Phe		
	515	520
Ser Leu Asn Ile Leu Asn Val Ser Tyr Asn Asn Leu Ala Gly Ala Val		
	530	535
Pro Thr Asp Asn Asn Phe Thr Arg Phe Ser His Asp Ser Phe Leu Gly		
545	550	555
Asn Pro Gly Leu Cys Gly Tyr Trp Leu Gly Ser Ser Cys Arg Ser Thr		
	565	570
		575

Gly His Arg Asp Lys Pro Pro Ile Ser Lys Ala Ala Ile Ile Gly Val  
 580 585 590  
 Ala Val Gly Gly Leu Val Ile Leu Leu Met Ile Leu Val Ala Val Cys  
 595 600 605  
 Arg Pro His His Pro Pro Ala Phe Lys Asp Ala Thr Val Ser Lys Pro  
 610 615 620  
 Val Ser Asn Gly Pro Pro Lys Leu Met Ile Leu His Met Asn Met Ala  
 625 630 635 640  
 Leu His Val Phe Asp Asp Ile Met Arg Met Thr Glu Asn Leu Ser Glu  
 645 650 655  
 Lys Tyr Ile Ile Gly Tyr Gly Ala Ser Ser Thr Val Tyr Lys Cys Val  
 660 665 670  
 Leu Lys Asn Cys Lys Pro Val Ala Ile Lys Lys Leu Tyr Ala His Tyr  
 675 680 685  
 Pro Gln Ser Leu Lys Glu Phe Glu Thr Glu Leu Glu Thr Val Gly Ser  
 690 695 700  
 Ile Lys His Arg Asn Leu Val Ser Leu Gln Gly Tyr Ser Leu Ser Pro  
 705 710 715 720  
 Val Gly Asn Leu Leu Phe Tyr Asp Tyr Met Glu Ser Gly Ser Leu Trp  
 725 730 735  
 Asp Val Leu His Glu Gly Ser Ser Lys Glu Asn Lys Leu Asp Trp Val  
 740 745 750  
 Thr Arg Leu Arg Ile Ala Leu Gly Ala Ala Gln Gly Leu Ala Tyr Leu  
 755 760 765  
 His His Asp Cys Ser Pro Arg Ile Ile His Arg Asp Val Lys Ser Lys  
 770 775 780  
 Asn Ile Leu Leu Asp Lys Asp Tyr Glu Ala His Leu Thr Asp Phe Gly  
 785 790 795 800  
 Ile Ala Lys Ser Leu Cys Val Ser Lys Thr His Thr Ser Thr Tyr Val  
 805 810 815  
 Met Gly Thr Ile Gly Tyr Ile Asp Pro Glu Tyr Ala Arg Thr Ser Arg  
 820 825 830  
 Leu Asn Glu Lys Ser Asp Val Tyr Ser Tyr Gly Ile Val Leu Leu Glu  
 835 840 845  
 Leu Leu Thr Gly Lys Lys Pro Val Asp Asn Glu Cys Asn Leu His His  
 850 855 860  
 Leu Ile Leu Ser Lys Thr Ala Ser Asn Glu Val Met Glu Thr Val Asp  
 865 870 875 880  
 Pro Asp Val Gly Asp Thr Cys Lys Asp Leu Gly Glu Val Lys Lys Leu  
 885 890 895  
 Phe Gln Leu Ala Leu Leu Cys Thr Lys Arg Gln Pro Ser Asp Arg Pro  
 900 905 910  
 Thr Met His Glu Val Val Arg Val Leu Asp Cys Leu Val Asn Pro Glu  
 915 920 925  
 Pro Pro Pro Gln Pro Gln Gln Gln Gln Gln Lys Ala His Ala His His

930		935		940
Gln Leu Pro Pro Gln Pro Ser Pro Pro Ala Tyr Val Asp Glu Tyr Val				
945		950		955
Ser Leu Arg Gly Thr Gly Ala Leu Ser Cys Ala Asn Ser Ser Ser Thr				
		965		970
Ser Asp Ala Glu Leu Phe Leu Lys Phe Gly Glu Ala Ile Ser Gln Asn				
		980		985
Met Val				990

&lt;210&gt; 7

&lt;211&gt; 3224

&lt;212&gt; DNA

&lt;213&gt; 玉米(Zea mays)

&lt;400&gt; 7

```

gagaaacact gctggagatc aagaaatcct tccgcgacgg cggcaacgcg ctgtacgatt 60
gttccggcga cggcgcgtcg cgggctact gctcgtggcg cggcgtgcta tgcgacaacg 120
tcaccttgcg tgtcgcggcg ctcaacctct ctgggctcaa tctcagggtg gaaatctcag 180
cggccatcgg gagtctgcaa cgtcttgtct caatcgattt gaagtcgaat ggactctctg 240
gacagatccc cgatgagatt ggtgattggt ctttgctcga aactttggat ttgtcatcta 300
acaatctaga aggagacata ccattctcca tgtccaagct gaagcacctt gagaacttga 360
tttgaagaa caacaaactg gtgggagtga tcccatcgac actctctcaa cttccaaatt 420
tgaagatatt ggacttggct caaaacaagt taagtgggta aattccgaat ctaatatatt 480
ggaatgaggt tcttcaatac ttgggattgc gaagcaatag tttagaagga agcctctctc 540
ccgatatgtg ccagtttaact ggtctgtggt actttgatgt gaagaacaat agcttgacgg 600
gtgcaatacc agaaaccata gggaaactgt cgagctttca ggtcttagat ttgtcaaaca 660
atcatcttac tggagaaatc ccgttcaata ttggtttctt gcaagtggct acgttatctt 720
tgcaagggaa caagtctctt ggtcctatac catcagtgat tggccttatg caggcgttg 780
cagtgtcaga tctgagtctt aatgagctat ctggcccaat accctctata ctgggcaact 840
tgacatacac tgagaaatta tacctgcaag gcaataggtt aactggattg ataccgccag 900
agcttggtaa tatgtcgaca ctgcattacc tggaaactgaa cgacaatctg ttgactgggt 960
tcattcctcc tgatcttggg aaacttacag aattggttga attgaacctt gcaacaaca 1020
accttatagg acctatccct gagaatttaa gttcatgtgc aaatctcatt agtttcaatg 1080
cttatggcaa taaattgaat ggaaccattc cacgttcatt tcacaagctt gagagtctga 1140
cttatctgaa tctgtcatca aatcatctca gtggagcact tccaattgag gttgcaagaa 1200
tgagaaattt ggacactctg gacttatcct gtaacatgat cactggttca attccctcgg 1260
ctattgggaa actagagcat cttttgaggc tcaacttaag caaaaataat gtggctggac 1320
acattcctgc tgaatttggg aacttaagga gcatcatgga gattgatttg tcttacaacc 1380
acctcagtgg cctgattcct caagaggttg ggatgctaca aaatttgata ctgttaaaat 1440
tagaaaagca taatattact ggagatgtct cttcacttat ttactgcttg agtctcaata 1500
tcttaaagt atcatacaac catctttatg gtactgtacc tacagacaac aacttctcac 1560
gatlttcacc cgacagcttc ttgggtaacc ctggactttg tggctattgg cttcactctg 1620
cttcatgcac acaattatcc aatgcagagc aatgaagag atcctctagc gcaaaggcct 1680

```

```

caatgtttgc agctattggt gttggtgccg tattgcttgt tattatgctc gttatcctag 1740
tagttatttg ctggccacat aactctccag tgctcaaaga tgtctctgta aacaaaccag 1800
ataaccttgc ttcagcatca aacaacattc atcccaagct tgtgatcctc cacatgaaca 1860
tggccctcta tgtatatgat gatataatga ggatgactga aaacttgagc gaaaaataca 1920
ttattggtta tggagcctca agtacagtct acagatgcga cctgaagaac tgcaagccaa 1980
ttgcgattaa aaagtgtat gtcactacc ctcagagctt gaaggaattc gagactgaac 2040
ttgagactgt tggaagcatc aaacaccgga atcttgaag ccttcagggg tactccctgt 2100
caccatctgg gaatctcctc ttctatgatt acatggaaaa tggcagcctc tgggacattt 2160
tacatgcttc atcgaagaaa aagaaactcg attgggaggc tcgcctcaag attgctctcg 2220
gagctgctca aggctggct tatcttcacc atgaatgcag tccacgaata atccacaggg 2280
atgtgaagtc aaagaatata ctctagaca aagactacga ggctcatctt gctgacttcg 2340
gtattgccaa gagcttgtgt gtgtcgaaga cgcacacatc aacgtacgtg atgggcacca 2400
ttggctacat tgacctgag tatgcacgga catcccggat caacgagaaa tcggatgtgt 2460
acagctacgg cattgtcttg ctggagctgc ttaccggcaa aaagcctgtc gacgacgagt 2520
gcaaccttea ccacttgatc ctatccaaag ccgcagaaaa cacggtcatg gagacgtag 2580
accaggacat caccgacacg tgcaaggacc tcggcagagt caagaaggtg ttccagctgg 2640
cgctcctttg cagcaagagg cagccgtcgg atcgaccgac catgcacgag gtcgcgcgcg 2700
tcctggacag cctcgtctgc ccagcaggcc cgccccgaa gcaggcgcag ggcagggcac 2760
aggcacaggc gtcggagaag ccgtccacca cggcgcgag ctatgtcagc gactacgtcg 2820
gcctacgagg cggcggcgcc ggcagcggcc tctcctgcac caactcgtcg agcgcgtccg 2880
acgccgagct cttcatgaag tttggcgagg tgatctcgcg gagcacggaa tagtccaata 2940
gatcgacgat cttttgtgt tacagtttac aataggtgtc tggccggcgc cctgtgatgt 3000
gactcgtctg aaatatgctg ctgccttctc gttgggactc cactaggcac tagtagctat 3060
ttttctcctg tatgcacgtc tctctctttt tttttgctgg gaaactcga cgtgtaactaa 3120
aggaatggat ttatgtgcag gctgcgttta ctgacaaaaa aaaaaaaaaa aaaaaaaaaa 3180
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaa 3224

```

<210> 8

<211> 976

<212> PRT

<213> 玉米(Zea mays)

<400> 8

```

Glu Thr Leu Leu Glu Ile Lys Lys Ser Phe Arg Asp Gly Gly Asn Ala
 1           5           10          15
Leu Tyr Asp Trp Ser Gly Asp Gly Ala Ser Pro Gly Tyr Cys Ser Trp
          20           25           30
Arg Gly Val Leu Cys Asp Asn Val Thr Phe Ala Val Ala Ala Leu Asn
          35           40           45
Leu Ser Gly Leu Asn Leu Glu Gly Glu Ile Ser Ala Ala Ile Gly Ser
          50           55           60
Leu Gln Arg Leu Val Ser Ile Asp Leu Lys Ser Asn Gly Leu Ser Gly
65           70           75           80
Gln Ile Pro Asp Glu Ile Gly Asp Cys Ser Leu Leu Glu Thr Leu Asp
          85           90           95

```



Leu Ser Ser Asn Asn Leu Glu Gly Asp Ile Pro Phe Ser Met Ser Lys  
 100 105 110  
 Leu Lys His Leu Glu Asn Leu Ile Leu Lys Asn Asn Lys Leu Val Gly  
 115 120 125  
 Val Ile Pro Ser Thr Leu Ser Gln Leu Pro Asn Leu Lys Ile Leu Asp  
 130 135 140  
 Leu Ala Gln Asn Lys Leu Ser Gly Glu Ile Pro Asn Leu Ile Tyr Trp  
 145 150 155 160  
 Asn Glu Val Leu Gln Tyr Leu Gly Leu Arg Ser Asn Ser Leu Glu Gly  
 165 170 175  
 Ser Leu Ser Pro Asp Met Cys Gln Leu Thr Gly Leu Trp Tyr Phe Asp  
 180 185 190  
 Val Lys Asn Asn Ser Leu Thr Gly Ala Ile Pro Glu Thr Ile Gly Asn  
 195 200 205  
 Cys Thr Ser Phe Gln Val Leu Asp Leu Ser Asn Asn His Leu Thr Gly  
 210 215 220  
 Glu Ile Pro Phe Asn Ile Gly Phe Leu Gln Val Ala Thr Leu Ser Leu  
 225 230 235 240  
 Gln Gly Asn Lys Phe Ser Gly Pro Ile Pro Ser Val Ile Gly Leu Met  
 245 250 255  
 Gln Ala Leu Ala Val Leu Asp Leu Ser Phe Asn Glu Leu Ser Gly Pro  
 260 265 270  
 Ile Pro Ser Ile Leu Gly Asn Leu Thr Tyr Thr Glu Lys Leu Tyr Leu  
 275 280 285  
 Gln Gly Asn Arg Leu Thr Gly Leu Ile Pro Pro Glu Leu Gly Asn Met  
 290 295 300  
 Ser Thr Leu His Tyr Leu Glu Leu Asn Asp Asn Leu Leu Thr Gly Phe  
 305 310 315 320  
 Ile Pro Pro Asp Leu Gly Lys Leu Thr Glu Leu Phe Glu Leu Asn Leu  
 325 330 335  
 Ala Asn Asn Asn Leu Ile Gly Pro Ile Pro Glu Asn Leu Ser Ser Cys  
 340 345 350  
 Ala Asn Leu Ile Ser Phe Asn Ala Tyr Gly Asn Lys Leu Asn Gly Thr  
 355 360 365  
 Ile Pro Arg Ser Phe His Lys Leu Glu Ser Leu Thr Tyr Leu Asn Leu  
 370 375 380  
 Ser Ser Asn His Leu Ser Gly Ala Leu Pro Ile Glu Val Ala Arg Met  
 385 390 395 400  
 Arg Asn Leu Asp Thr Leu Asp Leu Ser Cys Asn Met Ile Thr Gly Ser  
 405 410 415  
 Ile Pro Ser Ala Ile Gly Lys Leu Glu His Leu Leu Arg Leu Asn Leu  
 420 425 430  
 Ser Lys Asn Asn Val Ala Gly His Ile Pro Ala Glu Phe Gly Asn Leu  
 435 440 445  
 Arg Ser Ile Met Glu Ile Asp Leu Ser Tyr Asn His Leu Ser Gly Leu

450		455		460
Ile Pro Gln Glu Val Gly Met Leu Gln Asn Leu Ile Leu Leu Lys Leu				
465		470		475
Glu Ser Asn Asn Ile Thr Gly Asp Val Ser Ser Leu Ile Tyr Cys Leu				
	485		490	495
Ser Leu Asn Ile Leu Asn Val Ser Tyr Asn His Leu Tyr Gly Thr Val				
	500		505	510
Pro Thr Asp Asn Asn Phe Ser Arg Phe Ser Pro Asp Ser Phe Leu Gly				
	515		520	525
Asn Pro Gly Leu Cys Gly Tyr Trp Leu His Ser Ala Ser Cys Thr Gln				
	530		535	540
Leu Ser Asn Ala Glu Gln Met Lys Arg Ser Ser Ser Ala Lys Ala Ser				
545		550		555
Met Phe Ala Ala Ile Gly Val Gly Ala Val Leu Leu Val Ile Met Leu				
	565		570	575
Val Ile Leu Val Val Ile Cys Trp Pro His Asn Ser Pro Val Leu Lys				
	580		585	590
Asp Val Ser Val Asn Lys Pro Asp Asn Leu Ala Ser Ala Ser Asn Asn				
	595		600	605
Ile His Pro Lys Leu Val Ile Leu His Met Asn Met Ala Leu Tyr Val				
	610		615	620
Tyr Asp Asp Ile Met Arg Met Thr Glu Asn Leu Ser Glu Lys Tyr Ile				
625		630		635
Ile Gly Tyr Gly Ala Ser Ser Thr Val Tyr Arg Cys Asp Leu Lys Asn				
	645		650	655
Cys Lys Pro Ile Ala Ile Lys Lys Leu Tyr Ala His Tyr Pro Gln Ser				
	660		665	670
Leu Lys Glu Phe Glu Thr Glu Leu Glu Thr Val Gly Ser Ile Lys His				
	675		680	685
Arg Asn Leu Val Ser Leu Gln Gly Tyr Ser Leu Ser Pro Ser Gly Asn				
	690		695	700
Leu Leu Phe Tyr Asp Tyr Met Glu Asn Gly Ser Leu Trp Asp Ile Leu				
705		710		715
His Ala Ser Ser Lys Lys Lys Lys Leu Asp Trp Glu Ala Arg Leu Lys				
	725		730	735
Ile Ala Leu Gly Ala Ala Gln Gly Leu Ala Tyr Leu His His Glu Cys				
	740		745	750
Ser Pro Arg Ile Ile His Arg Asp Val Lys Ser Lys Asn Ile Leu Leu				
	755		760	765
Asp Lys Asp Tyr Glu Ala His Leu Ala Asp Phe Gly Ile Ala Lys Ser				
	770		775	780
Leu Cys Val Ser Lys Thr His Thr Ser Thr Tyr Val Met Gly Thr Ile				
785		790		795
Gly Tyr Ile Asp Pro Glu Tyr Ala Arg Thr Ser Arg Ile Asn Glu Lys				
	805		810	815

Ser Asp Val Tyr Ser Tyr Gly Ile Val Leu Leu Glu Leu Leu Thr Gly  
820 825 830  
Lys Lys Pro Val Asp Asp Glu Cys Asn Leu His His Leu Ile Leu Ser  
835 840 845  
Lys Ala Ala Glu Asn Thr Val Met Glu Thr Val Asp Gln Asp Ile Thr  
850 855 860  
Asp Thr Cys Lys Asp Leu Gly Glu Val Lys Lys Val Phe Gln Leu Ala  
865 870 875 880  
Leu Leu Cys Ser Lys Arg Gln Pro Ser Asp Arg Pro Thr Met His Glu  
885 890 895  
Val Ala Arg Val Leu Asp Ser Leu Val Cys Pro Ala Gly Pro Pro Pro  
900 905 910  
Lys Gln Ala Gln Ala Gln Ala Gln Ala Gln Ala Ser Glu Lys Pro Ser  
915 920 925  
Thr Thr Ala Pro Ser Tyr Val Ser Glu Tyr Val Gly Leu Arg Gly Gly  
930 935 940  
Gly Gly Gly Ser Ala Leu Ser Cys Thr Asn Ser Ser Ser Ala Ser Asp  
945 950 955 960  
Ala Glu Leu Phe Met Lys Phe Gly Glu Val Ile Ser Arg Ser Thr Glu  
965 970 975

<210> 9

<211> 2937

<212> DNA

<213> 水稻(Oryza sativa)

<400> 9

atgacgccgg cgccggcggc cgctcctac cgcgctctcg tcgctcct gtcgctgcc 60  
gtcgcggttg cogatgatgg gtcgacgctg ctggagatca agaagtcctt ccgcaatgtg 120  
gacaacgtac tgtacgattg ggccggcggc gactactgct cgtggcggcg cgctcctctgc 180  
gacaacgtca ccttcgccgt cgccgcgctc aacctatccg ggctcaacct cggaggcgag 240  
atctctccgg ccgctcggcag gttgaagggc atcgtctcga ttgacttgaa gtcgaatggg 300  
ctgtctgggc agatccctga tgagattggc gattgttcat cactaaaaac tctggatttg 360  
tctttcaata gcttggatgg ggacattccg ttctcagtat cgaagctgaa gcacattgag 420  
agcttgatat tgaagaacaa ccaactgacg ggagtgatcc catcaacgct ctcacagctc 480  
ccaaatttga agattttgga cttggcacag aacaaactga gtggagagat accaagactg 540  
atatattgga acgaggttct tcaactactg ggattacgcg gtaataattt agaaggcagc 600  
atctcccag atatatgcca gttgactggg ctttggtact ttgacgtaaa gaacaacagc 660  
ttgactgggc cgataaccaga aaccttggg aactgtacaa gttttcaggt cttggatttg 720  
tcttacaata aactttctgg atcaattcct ttcaacattg gtttctaca agttgctaca 780  
ctatctttgc aagggaacat gtttactggt cctattccat cagttattgg acttatgcag 840  
gctctcgtg tactggatct gagttacaac caattgtctg gtcctattcc atcgatacta 900  
ggcaatttaa catacactga gaagctgtat atgcaaggca ataagtaac aggtccaata 960  
ccacctgagc ttggaaatat gtcaaccctt cattacttag aacttaacga taatcaactt 1020

agcgggttca ttctccaga gttcggaaag ctaacagggt tatttgactt aaaccttgca 1080  
 aacaacaact ttgaaggicc aatccctgat aacataagct catgtgtgaa tctcaatagc 1140  
 ttcaatgctt atggcaacag attaatggg accattcctc cttcattgca taaacttgag 1200  
 agcatgactt atttgaattt gtcacaaat tttctaagtg gttctattcc tattgagcta 1260  
 tcgagaatca acaatttggg caccttggat ttatcctgta acatgattac tggcccaatt 1320  
 ccatcaacca ttgggagttt ggagcatcta ttaagactta acttgagcaa caatggctta 1380  
 gtaggattca ttctgcaga aattggcaac ttgaggagta tcatggagat tgatatgtcc 1440  
 aacaatcatc ttggcggttt gattcctcaa gaactcggaa tgctgcaaaa tctgatgttg 1500  
 ttaaactctca aaaacaacaa cataactggg gatgtctctt cactgatgaa ctgcttcagc 1560  
 ctcaatatct taaatgtatc ctataataat ttggctggtg ttgtacctac tgataacaac 1620  
 ttctcacggg tttgcctga cagctttttg ggtaatccag gactttgtgg atattggctt 1680  
 ggttcttcgt gccgttcac tggccatcaa cagaaaccac taatctcaaa ggctgcaata 1740  
 cttggaattg ccggtgggtg gcttgttata ctctgatga tcttagtagc ggtctgcagg 1800  
 cctcatagtc cacctgtttt caaagatgtc tctgttagca aaccagttag caatgttccc 1860  
 cccaagctgg ttatccttca tatgaacctt tcccttcttg tatacagagga tataatgacg 1920  
 atgactgaaa acctgagtga gaagtacatc attgggtacg gagcatccag cacggtttat 1980  
 aatgtgttt cgaagaaccg caaacctagt gcagtaaaaa agctatatgc ccactatcca 2040  
 cagagcttca aggaatttga aactgagctt gagactgttg gtagcatcaa acaccggaat 2100  
 ctagttagtc ttcaaggata ttccctatct cctgttggaa atcttctctt ctacgattac 2160  
 atggaaaatg gaagcctctg ggatgttttg catgaaggtc caactaagaa gaaaaaactt 2220  
 gattgggaaa ctctctacg aattgtctta ggtgcggccc aaggccttgc ttatcttcat 2280  
 catgactgta gcccacggat aatacacagg gatgtgaaat caaaaaatat actccttgat 2340  
 aaagattatg aggcacatct tacagacttt ggcatgtcta agagtttgtg tgtttcaaaa 2400  
 actcacacgt ccacctatgt catgggaact attggctata tcgacctga gtatgctcgc 2460  
 acctccgctc tcaatgaaaa gtctgatgtc tacagctatg gcattgttct gcttgagctg 2520  
 ctgaccggaa aaaagccagt ggacaacgag tgcaatctcc atcacttgat cttgtcaaaag 2580  
 acggctaaca atgctgtcat ggagacagtc gaccgggaca ttgcagacac ttgcaaggat 2640  
 cttggtgagg tcaagaagg tttccagctg gcgctccttt gcaccaagag acaacctcgc 2700  
 gatggccga caatgcacga gttgtgcgc gtcctggact gcctagtctg tcccgaccgc 2760  
 ccaccgaagt cgcacacga gctggccatg ccgcagcggc ctgctgtccc gagctacatc 2820  
 aacgagtatg tcagcttaag aggcaccagc gtgctctcct gcgccaactc gtcgtgtact 2880  
 tccgatgctg agctgtttct caagtttggc gaggtcattt ctcagaacac agagtag 2937

<210> 10

<211> 978

<212> PRT

<213> 水稻(Oryza sativa)

<400> 10

Met	Thr	Pro	Ala	Pro	Ala	Ala	Ala	Ser	Tyr	Arg	Ala	Leu	Val	Ala	Leu
1			5					10					15		
Leu	Leu	Val	Ala	Val	Ala	Val	Ala	Asp	Asp	Gly	Ser	Thr	Leu	Leu	Glu
		20						25					30		
Ile	Lys	Lys	Ser	Phe	Arg	Asn	Val	Asp	Asn	Val	Leu	Tyr	Asp	Trp	Ala
		35					40						45		

Gly Gly Asp Tyr Cys Ser Trp Arg Gly Val Leu Cys Asp Asn Val Thr  
 50 55 60  
 Phe Ala Val Ala Ala Leu Asn Leu Ser Gly Leu Asn Leu Gly Gly Glu  
 65 70 75 80  
 Ile Ser Pro Ala Val Gly Arg Leu Lys Gly Ile Val Ser Ile Asp Leu  
 85 90 95  
 Lys Ser Asn Gly Leu Ser Gly Gln Ile Pro Asp Glu Ile Gly Asp Cys  
 100 105 110  
 Ser Ser Leu Lys Thr Leu Asp Leu Ser Phe Asn Ser Leu Asp Gly Asp  
 115 120 125  
 Ile Pro Phe Ser Val Ser Lys Leu Lys His Ile Glu Ser Leu Ile Leu  
 130 135 140  
 Lys Asn Asn Gln Leu Ile Gly Val Ile Pro Ser Thr Leu Ser Gln Leu  
 145 150 155 160  
 Pro Asn Leu Lys Ile Leu Asp Leu Ala Gln Asn Lys Leu Ser Gly Glu  
 165 170 175  
 Ile Pro Arg Leu Ile Tyr Trp Asn Glu Val Leu Gln Tyr Leu Gly Leu  
 180 185 190  
 Arg Gly Asn Asn Leu Glu Gly Ser Ile Ser Pro Asp Ile Cys Gln Leu  
 195 200 205  
 Thr Gly Leu Trp Tyr Phe Asp Val Lys Asn Asn Ser Leu Thr Gly Pro  
 210 215 220  
 Ile Pro Glu Thr Ile Gly Asn Cys Thr Ser Phe Gln Val Leu Asp Leu  
 225 230 235 240  
 Ser Tyr Asn Lys Leu Ser Gly Ser Ile Pro Phe Asn Ile Gly Phe Leu  
 245 250 255  
 Gln Val Ala Thr Leu Ser Leu Gln Gly Asn Met Phe Thr Gly Pro Ile  
 260 265 270  
 Pro Ser Val Ile Gly Leu Met Gln Ala Leu Ala Val Leu Asp Leu Ser  
 275 280 285  
 Tyr Asn Gln Leu Ser Gly Pro Ile Pro Ser Ile Leu Gly Asn Leu Thr  
 290 295 300  
 Tyr Thr Glu Lys Leu Tyr Met Gln Gly Asn Lys Leu Thr Gly Pro Ile  
 305 310 315 320  
 Pro Pro Glu Leu Gly Asn Met Ser Thr Leu His Tyr Leu Glu Leu Asn  
 325 330 335  
 Asp Asn Gln Leu Ser Gly Phe Ile Pro Pro Glu Phe Gly Lys Leu Thr  
 340 345 350  
 Gly Leu Phe Asp Leu Asn Leu Ala Asn Asn Asn Phe Glu Gly Pro Ile  
 355 360 365  
 Pro Asp Asn Ile Ser Ser Cys Val Asn Leu Asn Ser Phe Asn Ala Tyr  
 370 375 380  
 Gly Asn Arg Leu Asn Gly Thr Ile Pro Pro Ser Leu His Lys Leu Glu  
 385 390 395 400  
 Ser Met Thr Tyr Leu Asn Leu Ser Ser Asn Phe Leu Ser Gly Ser Ile

405 410 415  
 Pro Ile Glu Leu Ser Arg Ile Asn Asn Leu Asp Thr Leu Asp Leu Ser  
 420 425 430  
 Cys Asn Met Ile Thr Gly Pro Ile Pro Ser Thr Ile Gly Ser Leu Glu  
 435 440 445  
 His Leu Leu Arg Leu Asn Leu Ser Asn Asn Gly Leu Val Gly Phe Ile  
 450 455 460  
 Pro Ala Glu Ile Gly Asn Leu Arg Ser Ile Met Glu Ile Asp Met Ser  
 465 470 475 480  
 Asn Asn His Leu Gly Gly Leu Ile Pro Gln Glu Leu Gly Met Leu Gln  
 485 490 495  
 Asn Leu Met Leu Leu Asn Leu Lys Asn Asn Asn Ile Thr Gly Asp Val  
 500 505 510  
 Ser Ser Leu Met Asn Cys Phe Ser Leu Asn Ile Leu Asn Val Ser Tyr  
 515 520 525  
 Asn Asn Leu Ala Gly Val Val Pro Thr Asp Asn Asn Phe Ser Arg Phe  
 530 535 540  
 Ser Pro Asp Ser Phe Leu Gly Asn Pro Gly Leu Cys Gly Tyr Trp Leu  
 545 550 555 560  
 Gly Ser Ser Cys Arg Ser Ser Gly His Gln Gln Lys Pro Leu Ile Ser  
 565 570 575  
 Lys Ala Ala Ile Leu Gly Ile Ala Val Gly Gly Leu Val Ile Leu Leu  
 580 585 590  
 Met Ile Leu Val Ala Val Cys Arg Pro His Ser Pro Pro Val Phe Lys  
 595 600 605  
 Asp Val Ser Val Ser Lys Pro Val Ser Asn Val Pro Pro Lys Leu Val  
 610 615 620  
 Ile Leu His Met Asn Leu Ser Leu Leu Val Tyr Glu Asp Ile Met Thr  
 625 630 635 640  
 Met Thr Glu Asn Leu Ser Glu Lys Tyr Ile Ile Gly Tyr Gly Ala Ser  
 645 650 655  
 Ser Thr Val Tyr Lys Cys Val Ser Lys Asn Arg Lys Pro Val Ala Val  
 660 665 670  
 Lys Lys Leu Tyr Ala His Tyr Pro Gln Ser Phe Lys Glu Phe Glu Thr  
 675 680 685  
 Glu Leu Glu Thr Val Gly Ser Ile Lys His Arg Asn Leu Val Ser Leu  
 690 695 700  
 Gln Gly Tyr Ser Leu Ser Pro Val Gly Asn Leu Leu Phe Tyr Asp Tyr  
 705 710 715 720  
 Met Glu Asn Gly Ser Leu Trp Asp Val Leu His Glu Gly Pro Thr Lys  
 725 730 735  
 Lys Lys Lys Leu Asp Trp Glu Thr Arg Leu Arg Ile Ala Leu Gly Ala  
 740 745 750  
 Ala Gln Gly Leu Ala Tyr Leu His His Asp Cys Ser Pro Arg Ile Ile  
 755 760 765

His Arg Asp Val Lys Ser Lys Asn Ile Leu Leu Asp Lys Asp Tyr Glu  
 770 775 780  
 Ala His Leu Thr Asp Phe Gly Ile Ala Lys Ser Leu Cys Val Ser Lys  
 785 790 795 800  
 Thr His Thr Ser Thr Tyr Val Met Gly Thr Ile Gly Tyr Ile Asp Pro  
 805 810 815  
 Glu Tyr Ala Arg Thr Ser Arg Leu Asn Glu Lys Ser Asp Val Tyr Ser  
 820 825 830  
 Tyr Gly Ile Val Leu Leu Glu Leu Leu Thr Gly Lys Lys Pro Val Asp  
 835 840 845  
 Asn Glu Cys Asn Leu His His Leu Ile Leu Ser Lys Thr Ala Asn Asn  
 850 855 860  
 Ala Val Met Glu Thr Val Asp Pro Asp Ile Ala Asp Thr Cys Lys Asp  
 865 870 875 880  
 Leu Gly Glu Val Lys Lys Val Phe Gln Leu Ala Leu Leu Cys Thr Lys  
 885 890 895  
 Arg Gln Pro Ser Asp Arg Pro Thr Met His Glu Val Val Arg Val Leu  
 900 905 910  
 Asp Cys Leu Val Arg Pro Asp Pro Pro Pro Lys Ser Ala Gln Gln Leu  
 915 920 925  
 Ala Met Pro Gln Arg Pro Ala Val Pro Ser Tyr Ile Asn Glu Tyr Val  
 930 935 940  
 Ser Leu Arg Gly Thr Ser Val Leu Ser Cys Ala Asn Ser Ser Cys Thr  
 945 950 955 960  
 Ser Asp Ala Glu Leu Phe Leu Lys Phe Gly Glu Val Ile Ser Gln Asn  
 965 970 975  
 Thr Glu

<210> 11

<211> 3000

<212> DNA

<213> 水稻(Oryza sativa)

<400> 11

atggcggcgg cgagggcgcc gtggctgtgg tggagggtgg tggaggttgt tgggtgtggcg 60  
 gtggcggagg cggcctcgg aggaggagga gggggagatg gggaggggaa ggcgctgatg 120  
 ggcgtgaagg cgggtttcgg gaacgcggcc aacgcgctcg tcgactggga cggcggcgcc 180  
 gaccactgcg cgtggcgcgg cgtcacctgc gacaacgcct ccttcgccgt cctgcacctg 240  
 aacttgtcaa atctaaacct aggaggtgag atctcgccgg ccatcggaga gtcaagaat 300  
 ctacagttcg ttgatctcaa ggggaacaag ctcactggcc aaatcccaga tgagattggg 360  
 gactgcatct ccttaaaata tttgatttg tctggcaact tgctgtatgg agacatcccc 420  
 ttctccatct ccaagctcaa gcagcttgag gagctgattt tgaagaacaa ccagctcacg 480  
 ggacctatcc ctccacatt gtcccaaatt ccaaatctca agacattgga cctggcacag 540

aaccagctta caggcgatat cccaaggctc atatactgga atgaagtict gcaataccta 600  
ggtttgaggg gtaactcaact gactggaact ttgtcacctg acatgtgcca actgactggc 660  
ctgtggtagt ttgatgtaag gggaaacaat ctcacaggga ccattccaga gagcataggg 720  
aactgcacca gctttgagat tctggacatt togtataacc aaatctctgg agaaatacct 780  
tacaacatag gctttcttca agtagccaca ctgtcacttc aaggaaatag actgactggg 840  
aaaattccag atgtgattgg cctgatgcaa gctcttgctg ttctagacct gagtgagaac 900  
gagctggtag ggcccatcc ttctatactg ggcaatctat cctatactgg aaaactatat 960  
ttacatggga acaaacttac tggagtcata ccgcccggagc ttgggaacat gagtaaactt 1020  
agctacctac aactgaatga taatgaattg gtgggcacaa ttccagcaga gcttggcaaa 1080  
cttgaagagc tttttgaaact aaatcttgcc aacaacaatc ttcaaggtec tattctctgca 1140  
aacatcagtt cttgcaactgc tctaaacaaa ttcaatgitt atggcaataa gctaaatggt 1200  
tctattctctg ctggtttcca gaagtggag agtctgactt acttgaacct atcttcaaac 1260  
aatttcaaag gcaatattcc ttctgagcct ggtcacatca tcaacttggga cacattggat 1320  
cttctctaca atgaattctc tggaccagtt cctgctacca ttggtgatct agagcacctt 1380  
cttgaactga atttgagtaa gaaccatctt gatgggcccag ttctgtctga gtttggaaac 1440  
ttgagaagcg tccaagtaat tgatgatgtec aacaacaact tatctggtag tctgcccagag 1500  
gaacttggac aacttcaaaa ccttgatagc ctgattctta acaacaacaa tttggttggg 1560  
gagatccctg ctcaattggc caactgcttc agcttaaata accttgcatt tcaggaattt 1620  
gtcatacaac aatttatctg gacatgtccc gatggcaaag aacttctcga aattcccaat 1680  
ggaaagcacc ttctaatttc tgattgcaac cagtacataa atcataaatg cagcttcttg 1740  
ggtaatccat tactgcatgt ttactgcca gattccagct gtggacactc tcatggacaa 1800  
agagttaata tttcaaagac agcaattgct tgcattatct taggctttat catattgctc 1860  
tgcgttctgc tgttggctat atataaaaca aatcaaccac agccacttgt caaaggatcc 1920  
gataagccag tgcaaggacc tccaaagcta gttgttctcc agatggacat ggctatccat 1980  
acttacgagg acatcatgag gctgacagag aatttgagcg agaaatacat cattggctat 2040  
ggcgcctcaa gcactgtcta caaatgtgaa ctcaagagcg gcaaggccat tgctgtcaag 2100  
cggttttaca gtcagtataa ccatagcctc cgagagtgtg aacagaact agagacaatt 2160  
ggcagcatac ggcacaggaa tcttgtagc ctccatggct tctcgtatc tccacatgga 2220  
aacttgctct tctatgatta catggaaaat ggttcttctg gggatcttct ccacgtcca 2280  
tcaaagaaag tgaagctcaa ctgggacaça agactgagga tgcggtcgg agctgcacaa 2340  
gggctggcct atctccacca tgactgcaac cctcgcataa tccacagaga tgtcaagtcc 2400  
tccaacatcc tgctcgacga gaacttcgaa gcgcacctct cagatttcgg catagccaaa 2460  
tgtgtcccct ctgccaagtc ccatgcctcc acttatgtgc taggaacct cggtacatt 2520  
gatccggagt atgccaggac ttccaggctc aatgagaaat ctgatgtgta cagcttcggc 2580  
atcgtccttc tggaaattgct cacagggaag aaggccgtcg acaacgaatc gaacttgcatt 2640  
caattgatac tctccaaagc tgatgacaac acagtcatgg aggcagtgga ctcgagggtg 2700  
tcagtgacgt gcacggacat gggactggct aggaaggcct tccagctcgc ctttctgtgc 2760  
accaagaggc acccttcaga ccggccgacc atgcacgagg ttgcaagggt gctgctctcc 2820  
ctgctgccgg cctccgccat gacaacgccc aagacggtgg actactccc gttgctggcg 2880  
tcgacgacga cggcggcoga catgagagg cagcagctga ccgacatcgg cgacaacagc 2940  
tctccgacg agcagtggtt cgtcaggttc ggcgagggtca tatccaagca cacaatgtga 3000

<210> 12

<211> 999



&lt;212&gt; PRT

<213> 水稻(*Oryza sativa*)

&lt;400&gt; 12

Met Ala Ala Ala Arg Ala Pro Trp Leu Trp Trp Trp Val Val Val Val  
 1 5 10 15  
 Val Gly Val Ala Val Ala Glu Ala Ala Ser Gly Gly Gly Gly Gly Gly  
 20 25 30  
 Asp Gly Glu Gly Lys Ala Leu Met Gly Val Lys Ala Gly Phe Gly Asn  
 35 40 45  
 Ala Ala Asn Ala Leu Val Asp Trp Asp Gly Gly Ala Asp His Cys Ala  
 50 55 60  
 Trp Arg Gly Val Thr Cys Asp Asn Ala Ser Phe Ala Val Leu Ala Leu  
 65 70 75 80  
 Asn Leu Ser Asn Leu Asn Leu Gly Gly Glu Ile Ser Pro Ala Ile Gly  
 85 90 95  
 Glu Leu Lys Asn Leu Gln Phe Val Asp Leu Lys Gly Asn Lys Leu Thr  
 100 105 110  
 Gly Gln Ile Pro Asp Glu Ile Gly Asp Cys Ile Ser Leu Lys Tyr Leu  
 115 120 125  
 Asp Leu Ser Gly Asn Leu Leu Tyr Gly Asp Ile Pro Phe Ser Ile Ser  
 130 135 140  
 Lys Leu Lys Gln Leu Glu Glu Leu Ile Leu Lys Asn Asn Gln Leu Thr  
 145 150 155 160  
 Gly Pro Ile Pro Ser Thr Leu Ser Gln Ile Pro Asn Leu Lys Thr Leu  
 165 170 175  
 Asp Leu Ala Gln Asn Gln Leu Thr Gly Asp Ile Pro Arg Leu Ile Tyr  
 180 185 190  
 Trp Asn Glu Val Leu Gln Tyr Leu Gly Leu Arg Gly Asn Ser Leu Thr  
 195 200 205  
 Gly Thr Leu Ser Pro Asp Met Cys Gln Leu Thr Gly Leu Trp Tyr Phe  
 210 215 220  
 Asp Val Arg Gly Asn Asn Leu Thr Gly Thr Ile Pro Glu Ser Ile Gly  
 225 230 235 240  
 Asn Cys Thr Ser Phe Glu Ile Leu Asp Ile Ser Tyr Asn Gln Ile Ser  
 245 250 255  
 Gly Glu Ile Pro Tyr Asn Ile Gly Phe Leu Gln Val Ala Thr Leu Ser  
 260 265 270  
 Leu Gln Gly Asn Arg Leu Thr Gly Lys Ile Pro Asp Val Ile Gly Leu  
 275 280 285  
 Met Gln Ala Leu Ala Val Leu Asp Leu Ser Glu Asn Glu Leu Val Gly  
 290 295 300  
 Pro Ile Pro Ser Ile Leu Gly Asn Leu Ser Tyr Thr Gly Lys Leu Tyr  
 305 310 315 320  
 Leu His Gly Asn Lys Leu Thr Gly Val Ile Pro Pro Glu Leu Gly Asn



Cys Glu Leu Lys Ser Gly Lys Ala Ile Ala Val Lys Arg Leu Tyr Ser  
 690 695 700  
 Gln Tyr Asn His Ser Leu Arg Glu Phe Glu Thr Glu Leu Glu Thr Ile  
 705 710 715 720  
 Gly Ser Ile Arg His Arg Asn Leu Val Ser Leu His Gly Phe Ser Leu  
 725 730 735  
 Ser Pro His Gly Asn Leu Leu Phe Tyr Asp Tyr Met Glu Asn Gly Ser  
 740 745 750  
 Leu Trp Asp Leu Leu His Gly Pro Ser Lys Lys Val Lys Leu Asn Trp  
 755 760 765  
 Asp Thr Arg Leu Arg Ile Ala Val Gly Ala Ala Gln Gly Leu Ala Tyr  
 770 775 780  
 Leu His His Asp Cys Asn Pro Arg Ile Ile His Arg Asp Val Lys Ser  
 785 790 795 800  
 Ser Asn Ile Leu Leu Asp Glu Asn Phe Glu Ala His Leu Ser Asp Phe  
 805 810 815  
 Gly Ile Ala Lys Cys Val Pro Ser Ala Lys Ser His Ala Ser Thr Tyr  
 820 825 830  
 Val Leu Gly Thr Ile Gly Tyr Ile Asp Pro Glu Tyr Ala Arg Thr Ser  
 835 840 845  
 Arg Leu Asn Glu Lys Ser Asp Val Tyr Ser Phe Gly Ile Val Leu Leu  
 850 855 860  
 Glu Leu Leu Thr Gly Lys Lys Ala Val Asp Asn Glu Ser Asn Leu His  
 865 870 875 880  
 Gln Leu Ile Leu Ser Lys Ala Asp Asp Asn Thr Val Met Glu Ala Val  
 885 890 895  
 Asp Ser Glu Val Ser Val Thr Cys Thr Asp Met Gly Leu Val Arg Lys  
 900 905 910  
 Ala Phe Gln Leu Ala Leu Leu Cys Thr Lys Arg His Pro Ser Asp Arg  
 915 920 925  
 Pro Thr Met His Glu Val Ala Arg Val Leu Leu Ser Leu Leu Pro Ala  
 930 935 940  
 Ser Ala Met Thr Thr Pro Lys Thr Val Asp Tyr Ser Arg Leu Leu Ala  
 945 950 955 960  
 Ser Thr Thr Thr Ala Ala Asp Met Arg Gly His Asp Val Thr Asp Ile  
 965 970 975  
 Gly Asp Asn Ser Ser Ser Asp Glu Gln Trp Phe Val Arg Phe Gly Glu  
 980 985 990  
 Val Ile Ser Lys His Thr Met  
 995

<210> 13

<211> 3317

<212> DNA

<213> 两色蜀黍(*Sorghum bicolor*)

## &lt;400&gt; 13

cctactggat cgcctatcgc ccagccatac agctactcgc tgtcactgag aaccagcggc 60  
gcaagtgcgc aaagccgagc gcgagaaccc acgaaaccaa ccaagctctg cccccattaa 120  
ggcagggcag gcagaggagg caggctcgcg gcagcacgct tcgcgctcca tcgtcgtgt 180  
ccttttctcc ctgtaatgtc actccccga tgcctgtccg cagctcagtg gccatgacga 240  
cgacggccgc ccgtgctctc gtcgccctcc tcctcgtcgc cgtcgcctgc gccgacgatg 300  
gggcgacgct ggtggagatc aagaagtcct tccgcaacgt cggcaacgta ctgtacgatt 360  
gggccggcga cgactactgc tcttggcgcg gcgtcctgtg cgacaacgtc acattcggccg 420  
tcgctcgcgt caacctctct ggcccaacc ttgaggcga gatctctcca gccgtcggca 480  
gcctcaagag cctcgtctcc atcgatctga agtcaaagtg gctatccggg cagatccctg 540  
atgagattgg tgattgttca tcaacttagga cgctggactt ttctttcaac aacttggatg 600  
gcgacatacc attctctata tcaaagctga agcacctgga gaacttgata ttgagaaga 660  
accagctgat tgggtgcgat ccatcaacat gtgcacagct cccaaatttg aagattttgg 720  
atttggcaca aaacaactg actggggaga taccaaggct tatctactgg aatgaggttc 780  
ttcaatatct tgatgtgaag aacaatagct tgaccggggt gataccagac accattggga 840  
actgtacaag ttttcaagtc ttggatttgt cttacaaccg ctttactgga ccaatccat 900  
tcaacattgg tttctacaa gtggctacac tatecttgca agggaacaag ttcaccggtc 960  
caattccttc agtaattggt ctatgcagg ctctcgtctg tctagatctg agttacaacc 1020  
aattatctgg tctatacca tcaatactag gcaactgac atacactgag aagctgtaca 1080  
tccaaggcaa taagttaact gggtcgatc caccagagtt aggaaatatg tcaacacttc 1140  
attacctaga actgaacgat aatcaactta ctgggtcaat tccaccagag ctgggaaggc 1200  
taacaggctt gtttgacctg aaccttgcca ataaccacct ggaaggacca attcctgaca 1260  
acctaagttc atgtgtgaat ctcaatagct tcaatgotta tggcaacaag ttaaatggga 1320  
ccattcctcg ttcgttcggg aaacttgaaa gcatgacctt ttaaatctg tcatcaaact 1380  
tcataagtgg ctctatctct attgagttat caaggatcaa caatttggac acgctggatt 1440  
tatcctgtaa catgatgact ggtccaattc catcatcaat tggcagccta gagcatctat 1500  
tgagacttaa cttgagcaag aatggtctag ttggattcat ccccgaggag tttggtaatt 1560  
tgaggagtgt catggagatt gatttatctc ataatacact tgggtggcctg attcctcaag 1620  
aacttgaat gctgcaaac ctgatgttg taaatgtgc gtacaataat ttggctggtg 1680  
ttgtccctgc tgacaacaac ttcacacggt tttcacctga cagcttttta ggtaatcctg 1740  
gactctgtgg atactggctt ggttcgtcgt gtcgttccac tggccaccac gagaaaccgc 1800  
ctatctcaa ggctgccata attggtgttg ctgtgggtgg acttgttacc ctcttgatga 1860  
tcttagtagc tgtttgcagg ccacatcgtc cacctgcttt taaagatgac actgtaagca 1920  
agccagtgag aaatgctccc ccaagctgg tgatcctca tatgaacatg gcccttcacg 1980  
tatacgatga cataatgagg atgactgaga acttgagtga gaaatacatc attggatacg 2040  
ggcgtcaag tacagtttat aatgtgtcc taaagaattg caaacgggtg gcaataaaaa 2100  
agctgtatgc ccactacca cagagcctta aggaatttga aactgagctt gagactgttg 2160  
gtagcatcaa gcaccggaat ctagtacgcc ttcaagggtg ctattatca cctgttggga 2220  
acctcctctt ttatgattat atggaatgtg gcagcttatg ggatgtttta catgaaggtt 2280  
catccaagaa gaaaaactt gactgggaga ctgcctacg gattgctctt ggtgcagctc 2340  
aaggccttgc ttacctcac catgactgca gtccacggat aattcatcgg gatgtaaaat 2400  
caaagaatat actccttgac aaagattatg aggcccatct tacagacttt ggaattgcta 2460  
agagcttatg tgtctcaaaa actcacacat caacctatgt catgggaact attggctaca 2520

```

ttgatcctga gtacgccgc acttcccgtc tcaacgaaa gtctgatgtc tacaggctat 2580
ggcattgttc tgctggagct gctgactggc aagaagccag tggacaacga atcctatcga 2640
agacggcaag caacgaggtc atggataccg tggaccctga catcggggac acctgcaagg 2700
acctcgcgga ggtgaagaag ctgttccagc tggcgctcct ttgcaccaag cggcaaccct 2760
cggaccgacc gacgatgcac gaggtggtgc gcgtcctgga ctgcctggtg aaccgggacc 2820
cgccgcaaaa gccgtcggcg caccagctgc cgcagccgtc gccagccgtg ccaagctaca 2880
tcaacgagta cgtcagcctg cggggcaccg gcgctctctc ctgcgccaac tcgaccagca 2940
cctcggacgc cgagctgttc ctcaagttcg gcgaggccat ctgcgagaac atggagtagg 3000
gaagagagag aggtctgggg agttgaagga tgcagggagt agtgggagta gctgactgac 3060
atittgctgg atgcaggagg agattaacat ggggaactca gtaggggtgtt ggttaactgt 3120
aaaaaagtc atgtgectct agagcagcag tagagcttct tcctcatctt tctttttttt 3180
atcaccccc atttttcctt ggtggtctct aacttgtag gaggtgtat tggatcatctc 3240
tcacagttgt gtgcctgatg atcttttgtg tgacttttgc cgtgttgaa tcatgggagg 3300
tgactttctt gatctgc 3317

```

<210> 14

<211> 929

<212> PRT

<213> 两色蜀黍(Sorghum bicolor)

<400> 14

```

Met Pro Val Arg Ser Ser Val Ala Met Thr Thr Thr Ala Ala Arg Ala
 1           5           10           15
Leu Val Ala Leu Leu Leu Val Ala Val Ala Val Ala Asp Asp Gly Ala
 20           25           30
Thr Leu Val Glu Ile Lys Lys Ser Phe Arg Asn Val Gly Asn Val Leu
 35           40           45
Tyr Asp Trp Ala Gly Asp Asp Tyr Cys Ser Trp Arg Gly Val Leu Cys
 50           55           60
Asp Asn Val Thr Phe Ala Val Ala Ala Leu Asn Leu Ser Gly Leu Asn
 65           70           75           80
Leu Glu Gly Glu Ile Ser Pro Ala Val Gly Ser Leu Lys Ser Leu Val
 85           90           95
Ser Ile Asp Leu Lys Ser Asn Gly Leu Ser Gly Gln Ile Pro Asp Glu
 100          105          110
Ile Gly Asp Cys Ser Ser Leu Arg Thr Leu Asp Phe Ser Phe Asn Asn
 115          120          125
Leu Asp Gly Asp Ile Pro Phe Ser Ile Ser Lys Leu Lys His Leu Glu
 130          135          140
Asn Leu Ile Leu Lys Asn Asn Gln Leu Ile Gly Ala Ile Pro Ser Thr
 145          150          155          160
Leu Ser Gln Leu Pro Asn Leu Lys Ile Leu Asp Leu Ala Gln Asn Lys
 165          170          175
Leu Thr Gly Glu Ile Pro Arg Leu Ile Tyr Trp Asn Glu Val Leu Gln
 180          185          190

```

Tyr Leu Asp Val Lys Asn Asn Ser Leu Thr Gly Val Ile Pro Asp Thr  
 195 200 205  
 Ile Gly Asn Cys Thr Ser Phe Gln Val Leu Asp Leu Ser Tyr Asn Arg  
 210 215 220  
 Phe Thr Gly Pro Ile Pro Phe Asn Ile Gly Phe Leu Gln Val Ala Thr  
 225 230 235 240  
 Leu Ser Leu Gln Gly Asn Lys Phe Thr Gly Pro Ile Pro Ser Val Ile  
 245 250 255  
 Gly Leu Met Gln Ala Leu Ala Val Leu Asp Leu Ser Tyr Asn Gln Leu  
 260 265 270  
 Ser Gly Pro Ile Pro Ser Ile Leu Gly Asn Leu Thr Tyr Thr Glu Lys  
 275 280 285  
 Leu Tyr Ile Gln Gly Asn Lys Leu Thr Gly Ser Ile Pro Pro Glu Leu  
 290 295 300  
 Gly Asn Met Ser Thr Leu His Tyr Leu Glu Leu Asn Asp Asn Gln Leu  
 305 310 315 320  
 Thr Gly Ser Ile Pro Pro Glu Leu Gly Arg Leu Thr Gly Leu Phe Asp  
 325 330 335  
 Leu Asn Leu Ala Asn Asn His Leu Glu Gly Pro Ile Pro Asp Asn Leu  
 340 345 350  
 Ser Ser Cys Val Asn Leu Asn Ser Phe Asn Ala Tyr Gly Asn Lys Leu  
 355 360 365  
 Asn Gly Thr Ile Pro Arg Ser Leu Arg Lys Leu Glu Ser Met Thr Tyr  
 370 375 380  
 Leu Asn Leu Ser Ser Asn Phe Ile Ser Gly Ser Ile Pro Ile Glu Leu  
 385 390 395 400  
 Ser Arg Ile Asn Asn Leu Asp Thr Leu Asp Leu Ser Cys Asn Met Met  
 405 410 415  
 Thr Gly Pro Ile Pro Ser Ser Ile Gly Ser Leu Glu His Leu Leu Arg  
 420 425 430  
 Leu Asn Leu Ser Lys Asn Gly Leu Val Gly Phe Ile Pro Ala Glu Phe  
 435 440 445  
 Gly Asn Leu Arg Ser Val Met Glu Ile Asp Leu Ser Tyr Asn His Leu  
 450 455 460  
 Gly Gly Leu Ile Pro Gln Glu Leu Glu Met Leu Gln Asn Leu Met Leu  
 465 470 475 480  
 Leu Asn Val Ser Tyr Asn Asn Leu Ala Gly Val Val Pro Ala Asp Asn  
 485 490 495  
 Asn Phe Thr Arg Phe Ser Pro Asp Ser Phe Leu Gly Asn Pro Gly Leu  
 500 505 510  
 Cys Gly Tyr Trp Leu Gly Ser Ser Cys Arg Ser Thr Gly His His Glu  
 515 520 525  
 Lys Pro Pro Ile Ser Lys Ala Ala Ile Ile Gly Val Ala Val Gly Gly  
 530 535 540  
 Leu Val Ile Leu Leu Met Ile Leu Val Ala Val Cys Arg Pro His Arg

---

545                    550                    555                    560  
 Pro Pro Ala Phe Lys Asp Val Thr Val Ser Lys Pro Val Arg Asn Ala  
                           565                    570                    575  
 Pro Pro Lys Leu Val Ile Leu His Met Asn Met Ala Leu His Val Tyr  
                           580                    585                    590  
 Asp Asp Ile Met Arg Met Thr Glu Asn Leu Ser Glu Lys Tyr Ile Ile  
                           595                    600                    605  
 Gly Tyr Gly Ala Ser Ser Thr Val Tyr Lys Cys Val Leu Lys Asn Cys  
                           610                    615                    620  
 Lys Pro Val Ala Ile Lys Lys Leu Tyr Ala His Tyr Pro Gln Ser Leu  
 625                    630                    635                    640  
 Lys Glu Phe Glu Thr Glu Leu Glu Thr Val Gly Ser Ile Lys His Arg  
                           645                    650                    655  
 Asn Leu Val Ser Leu Gln Gly Tyr Ser Leu Ser Pro Val Gly Asn Leu  
                           660                    665                    670  
 Leu Phe Tyr Asp Tyr Met Glu Cys Gly Ser Leu Trp Asp Val Leu His  
                           675                    680                    685  
 Glu Gly Ser Ser Lys Lys Lys Lys Leu Asp Trp Glu Thr Arg Leu Arg  
                           690                    695                    700  
 Ile Ala Leu Gly Ala Ala Gln Gly Leu Ala Tyr Leu His His Asp Cys  
 705                    710                    715                    720  
 Ser Pro Arg Ile Ile His Arg Asp Val Lys Ser Lys Asn Ile Leu Leu  
                           725                    730                    735  
 Asp Lys Asp Tyr Glu Ala His Leu Thr Asp Phe Gly Ile Ala Lys Ser  
                           740                    745                    750  
 Leu Cys Val Ser Lys Thr His Thr Ser Thr Tyr Val Met Gly Thr Ile  
                           755                    760                    765  
 Gly Tyr Ile Asp Pro Glu Tyr Ala Arg Thr Ser Arg Leu Asn Glu Lys  
                           770                    775                    780  
 Ser Asp Val Tyr Arg Leu Trp His Cys Ser Ala Gly Ala Ala Asp Trp  
 785                    790                    795                    800  
 Gln Glu Ala Ser Gly Gln Arg Ile Leu Ser Lys Thr Ala Ser Asn Glu  
                           805                    810                    815  
 Val Met Asp Thr Val Asp Pro Asp Ile Gly Asp Thr Cys Lys Asp Leu  
                           820                    825                    830  
 Gly Glu Val Lys Lys Leu Phe Gln Leu Ala Leu Leu Cys Thr Lys Arg  
                           835                    840                    845  
 Gln Pro Ser Asp Arg Pro Thr Met His Glu Val Val Arg Val Leu Asp  
                           850                    855                    860  
 Cys Leu Val Asn Pro Asp Pro Pro Pro Lys Pro Ser Ala His Gln Leu  
 865                    870                    875                    880  
 Pro Gln Pro Ser Pro Ala Val Pro Ser Tyr Ile Asn Glu Tyr Val Ser  
                           885                    890                    895  
 Leu Arg Gly Thr Gly Ala Leu Ser Cys Ala Asn Ser Thr Ser Thr Ser  
                           900                    905                    910

Asp Ala Glu Leu Phe Leu Lys Phe Gly Glu Ala Ile Ser Gln Asn Met  
 915 920 925  
 Glu

<210> 15

<211> 1514

<212> DNA

<213> 两色蜀黍(Sorghum bicolor)

<400> 15

```

atggcagcga gggcggcggc ggcggtggtg ctgcttattg cggctgtggt gtcggtgtcg 60
gcgggaggag gtgaagggga cggagacggg cagacgctga tggcgggtaa ggcgggggtc 120
gggaacgcgg ccaacgcgct ggcggactgg gacggcggcc gcgaccactg cgcttgccgc 180
ggcgtcgccg ggcgacggcc ctctttcgcc gtcgtcggcc tgaacctgtc aaatctaaac 240
ctcggagggg agatctcgcc ggctataggg cagctcaaga gcctacagt cgtggatctc 300
aagctgaaca agctcacagg ccaaatccca gatgagattg gggattgtgt ctcttaaaa 360
tatttgattg tgtctggaaa ctgtctgtat ggagacatcc cttctccat ctccaagctc 420
aaacagcttg aggacctgat ttgagaagc aaccaactca cgggacccat ccttccaca 480
ctgtcccaga ttccaatct caagacctg gatctggcgc agaacaagct caccggagac 540
attcccaggc tcatctactg gaatgaagta ctgcaatacc taggcttgag gggcaattca 600
ctgactggaa ctctgtcacc tgatatgtgc caactgactg gcctgtggta tttgatgta 660
agggggaaca atctcacagg aacaattcca gagggcatag ggaactgcac tagctttgag 720
attctggata ttctatacaa ccaatctct ggagaaatac cttacaacat aggttacctt 780
caagtagcca cacttgatct tagcgagaat gaacttggg gaccaattcc tccgatactt 840
ggcaacctgt cctacacagg caaactctat ttacatggca acaaactcac gggacatata 900
ccaccagaac tggggaacat gagtaactt agctacctgc agctgaatga caatgaacta 960
gtgggcacaa tcccagctga gcttggcaaa ctacagagt tatttgaatt gaacttggc 1020
aacaacaatc ttgagggtca tattcctgca aacatcagct cttgcagtgc actgaacaaa 1080
ttcaatgtgt atggcaatag actgaatggc tctatccctg ctggtttcca ggagttggag 1140
agtttgacat acctgaacct ttcttcaaac aatttcaaag gccagattcc ctctgagctt 1200
ggtcacatag tcaacttggc cacactagat ctttctaca atgaattctc tggaccagtt 1260
cctcactacta ttggtgatct cgagcatctt cttgaattga atttgagtaa aaacctctt 1320
actggatctg tgctgctga atttggaaac ttgagaagtg tccaagtaat tgacatatcc 1380
agcaacaact tgactgggta tctccctgaa gaacttggac agctgcaaaa ccttgatagc 1440
ttgattctta acaacaacaa tttggttggg gagatccctg ctgagctggc taactgcttc 1500
agcttaatta cctt 1514

```

<210> 16

<211> 504

<212> PRT

<213> 两色蜀黍(Sorghum bicolor)

<400> 16



Met Ala Ala Arg Ala Ala Ala Ala Val Val Leu Leu Ile Ala Ala Val  
 1 5 10 15  
 Val Ser Val Ser Ala Gly Gly Gly Glu Gly Asp Gly Asp Gly Gln Thr  
 20 25 30  
 Leu Met Ala Val Lys Ala Gly Phe Gly Asn Ala Ala Asn Ala Leu Ala  
 35 40 45  
 Asp Trp Asp Gly Gly Arg Asp His Cys Ala Trp Arg Gly Val Ala Cys  
 50 55 60  
 Asp Ala Ala Ser Phe Ala Val Val Gly Leu Asn Leu Ser Asn Leu Asn  
 65 70 75 80  
 Leu Gly Gly Glu Ile Ser Pro Ala Ile Gly Gln Leu Lys Ser Leu Gln  
 85 90 95  
 Phe Val Asp Leu Lys Leu Asn Lys Leu Thr Gly Gln Ile Pro Asp Glu  
 100 105 110  
 Ile Gly Asp Cys Val Ser Leu Lys Tyr Leu Asp Leu Ser Gly Asn Leu  
 115 120 125  
 Leu Tyr Gly Asp Ile Pro Phe Ser Ile Ser Lys Leu Lys Gln Leu Glu  
 130 135 140  
 Asp Leu Ile Leu Lys Asn Asn Gln Leu Thr Gly Pro Ile Pro Ser Thr  
 145 150 155 160  
 Leu Ser Gln Ile Pro Asn Leu Lys Thr Leu Asp Leu Ala Gln Asn Lys  
 165 170 175  
 Leu Thr Gly Asp Ile Pro Arg Leu Ile Tyr Trp Asn Glu Val Leu Gln  
 180 185 190  
 Tyr Leu Gly Leu Arg Gly Asn Ser Leu Thr Gly Thr Leu Ser Pro Asp  
 195 200 205  
 Met Cys Gln Leu Thr Gly Leu Trp Tyr Phe Asp Val Arg Gly Asn Asn  
 210 215 220  
 Leu Thr Gly Thr Ile Pro Glu Gly Ile Gly Asn Cys Thr Ser Phe Glu  
 225 230 235 240  
 Ile Leu Asp Ile Ser Tyr Asn Gln Ile Ser Gly Glu Ile Pro Tyr Asn  
 245 250 255  
 Ile Gly Tyr Leu Gln Val Ala Thr Leu Asp Leu Ser Glu Asn Glu Leu  
 260 265 270  
 Val Gly Pro Ile Pro Pro Ile Leu Gly Asn Leu Ser Tyr Thr Gly Lys  
 275 280 285  
 Leu Tyr Leu His Gly Asn Lys Leu Thr Gly His Ile Pro Pro Glu Leu  
 290 295 300  
 Gly Asn Met Ser Lys Leu Ser Tyr Leu Gln Leu Asn Asp Asn Glu Leu  
 305 310 315 320  
 Val Gly Thr Ile Pro Ala Glu Leu Gly Lys Leu Thr Glu Leu Phe Glu  
 325 330 335  
 Leu Asn Leu Ala Asn Asn Asn Leu Glu Gly His Ile Pro Ala Asn Ile  
 340 345 350  
 Ser Ser Cys Ser Ala Leu Asn Lys Phe Asn Val Tyr Gly Asn Arg Leu

```

      355                360                365
Asn Gly Ser Ile Pro Ala Gly Phe Gln Glu Leu Glu Ser Leu Thr Tyr
   370                375                380
Leu Asn Leu Ser Ser Asn Asn Phe Lys Gly Gln Ile Pro Ser Glu Leu
385                390                395                400
Gly His Ile Val Asn Leu Asp Thr Leu Asp Leu Ser Tyr Asn Glu Phe
      405                410                415
Ser Gly Pro Val Pro Pro Thr Ile Gly Asp Leu Glu His Leu Leu Glu
      420                425                430
Leu Asn Leu Ser Lys Asn His Leu Thr Gly Ser Val Pro Ala Glu Phe
      435                440                445
Gly Asn Leu Arg Ser Val Gln Val Ile Asp Ile Ser Ser Asn Asn Leu
      450                455                460
Thr Gly Tyr Leu Pro Glu Glu Leu Gly Gln Leu Gln Asn Leu Asp Ser
465                470                475                480
Leu Ile Leu Asn Asn Asn Asn Leu Val Gly Glu Ile Pro Ala Gln Leu
      485                490                495
Ala Asn Cys Phe Ser Leu Ile Thr
      500

```

&lt;210&gt; 17

&lt;211&gt; 752

&lt;212&gt; PRT

&lt;213&gt; 两色蜀黍(Sorghum bicolor)

&lt;400&gt; 17

```

Cys Thr Thr Cys Ala Ala Ala Thr Gly Gly Ala Cys Ala Thr Gly Gly
  1                5                10                15
Cys Thr Ala Cys Cys Cys Ala Thr Ala Cys Cys Thr Ala Thr Gly Ala
      20                25                30
Ala Gly Ala Cys Ala Thr Thr Ala Thr Gly Ala Gly Gly Thr Thr Gly
      35                40                45
Ala Cys Thr Gly Ala Gly Ala Ala Thr Thr Thr Gly Ala Gly Cys Gly
      50                55                60
Ala Gly Ala Ala Ala Thr Ala Cys Ala Thr Cys Ala Thr Thr Gly Gly
65                70                75                80
Thr Thr Ala Thr Gly Gly Gly Gly Cys Ala Thr Cys Ala Ala Gly Thr
      85                90                95
Ala Cys Thr Gly Thr Gly Thr Ala Cys Ala Ala Ala Thr Gly Thr Gly
      100                105                110
Ala Thr Cys Thr Cys Ala Ala Gly Gly Gly Cys Gly Gly Cys Ala Ala
      115                120                125
Ala Gly Cys Cys Ala Thr Cys Gly Cys Thr Gly Thr Cys Ala Ala Ala
      130                135                140

```

Cys Gly Gly Cys Thr Thr Thr Ala Cys Ala Gly Thr Cys Ala Gly Thr  
 145 150 155 160  
 Ala Thr Ala Ala Cys Cys Ala Cys Ala Gly Cys Cys Thr Cys Cys Gly  
 165 170 175  
 Thr Gly Ala Gly Thr Thr Thr Gly Ala Gly Ala Cys Ala Gly Ala Ala  
 180 185 190  
 Cys Thr Gly Gly Ala Gly Ala Cys Gly Ala Thr Cys Gly Gly Thr Ala  
 195 200 205  
 Gly Cys Ala Thr Cys Cys Gly Ala Cys Ala Cys Ala Gly Gly Ala Ala  
 210 215 220  
 Thr Cys Thr Cys Gly Thr Cys Ala Gly Cys Cys Thr Thr Cys Ala Thr  
 225 230 235 240  
 Gly Gly Cys Thr Thr Cys Thr Cys Ala Cys Thr Cys Thr Cys Cys Cys  
 245 250 255  
 Cys Thr Cys Ala Thr Gly Gly Ala Ala Ala Cys Cys Thr Gly Cys Thr  
 260 265 270  
 Thr Thr Thr Cys Thr Ala Cys Gly Ala Thr Thr Ala Cys Ala Thr Gly  
 275 280 285  
 Gly Ala Ala Ala Ala Thr Gly Gly Thr Thr Cys Cys Cys Thr Gly Thr  
 290 295 300  
 Gly Gly Gly Ala Thr Cys Thr Thr Cys Thr Thr Cys Ala Thr Gly Gly  
 305 310 315 320  
 Thr Cys Cys Ala Thr Cys Ala Ala Ala Gly Ala Ala Gly Gly Thr Gly  
 325 330 335  
 Ala Ala Gly Cys Thr Thr Gly Ala Thr Thr Gly Gly Gly Ala Cys Ala  
 340 345 350  
 Cys Ala Ala Gly Gly Cys Thr Thr Ala Ala Gly Ala Thr Thr Gly Cys  
 355 360 365  
 Gly Gly Thr Ala Gly Gly Thr Gly Cys Thr Gly Cys Gly Cys Ala Ala  
 370 375 380  
 Gly Gly Ala Cys Thr Gly Gly Cys Cys Thr Ala Thr Cys Thr Thr Cys  
 385 390 395 400  
 Ala Cys Cys Ala Thr Gly Ala Cys Thr Gly Cys Ala Ala Cys Cys Cys  
 405 410 415  
 Thr Cys Gly Cys Ala Thr Ala Ala Thr Cys Cys Ala Cys Ala Gly Gly  
 420 425 430  
 Gly Ala Thr Gly Thr Cys Ala Ala Gly Thr Cys Cys Thr Cys Ala Ala  
 435 440 445  
 Ala Cys Ala Thr Cys Cys Thr Gly Cys Thr Cys Gly Ala Cys Gly Ala  
 450 455 460  
 Gly Ala Ala Cys Thr Thr Cys Gly Ala Ala Gly Cys Gly Cys Ala Cys  
 465 470 475 480  
 Cys Thr Cys Thr Cys Thr Gly Ala Thr Thr Thr Cys Gly Gly Cys Ala  
 485 490 495  
 Thr Cys Gly Cys Cys Ala Ala Ala Thr Gly Thr Gly Thr Cys Cys Cys



Arg Leu Tyr Ser Gln Tyr Asn His Ser Leu Arg Glu Phe Glu Thr Glu  
 50 55 60  
 Leu Glu Thr Ile Gly Ser Ile Arg His Arg Asn Leu Val Ser Leu His  
 65 70 75 80  
 Gly Phe Ser Leu Ser Pro His Gly Asn Leu Leu Phe Tyr Asp Tyr Met  
 85 90 95  
 Glu Asn Gly Ser Leu Trp Asp Leu Leu His Gly Pro Ser Lys Lys Val  
 100 105 110  
 Lys Leu Asp Trp Asp Thr Arg Leu Lys Ile Ala Val Gly Ala Ala Gln  
 115 120 125  
 Gly Leu Ala Tyr Leu His His Asp Cys Asn Pro Arg Ile Ile His Arg  
 130 135 140  
 Asp Val Lys Ser Ser Asn Ile Leu Leu Asp Glu Asn Phe Glu Ala His  
 145 150 155 160  
 Leu Ser Asp Phe Gly Ile Ala Lys Cys Val Pro Ala Ala Lys Ser His  
 165 170 175  
 Ala Ser Thr Tyr Val Leu Gly Thr Ile Gly Tyr Ile Asp Pro Glu Tyr  
 180 185 190  
 Ala Arg Thr Ser Arg Leu Asn Glu Lys Ser Asp Val Tyr Ser Phe Gly  
 195 200 205  
 Ile Val Leu Leu Glu Leu Leu Thr Gly Lys Lys Ala Val Asp Asn Glu  
 210 215 220  
 Ser Asn Leu His Gln Leu Ile Leu Ser Lys Ala Asp Asp Asn Thr Val  
 225 230 235 240  
 Met Glu Ala Val Asp Ser Glu Val Ser Val  
 245 250

<210> 19

<211> 1402

<212> DNA

<213> 大豆(Glycine max)

<220>

<221> misc\_feature

<222> 1253

<223> n = A, T, C 或 G

<400> 19

cttagtctct ctctacttaa tgtgtcctat aacaaactat ttggtgttat ccccacgagt 60  
 aacaaacttta ccaggtttcc cctgacagt ttcattggaa accctggctt ttgtggtaat 120  
 tggetgaatt tgccatgcca tgggtctcgt ccttcagagc gagttacatt atctaagget 180  
 gccattcttg gaattacttt ggggtcccctt gtgattcttc ttatggtatt ggtggcagct 240  
 tgccgaccac acagcccctc tccttttctt gatggatcat ttgacaaacc aattaatttc 300  
 tcccctccaa agctagtgat tcttcatatg aatatggcac tacatgtgta tgaagatata 360

atgaggatga ctgaaaacct aagtgagaag tatatcattg gatatgggtgc atcaagtaca 420  
 gtttataaat gtgttcttaa gaattgtaag cgggtggcta tcaagaggat ctattctcac 480  
 tatccccaat gtattaaaga atttgaaact gaactcgaga ctgttggcag catcaagcac 540  
 cggaatttgg tcagtctcca aggctactcc ttgtcccat atggccatct cctgttttat 600  
 gactacatgg aaaatggcag tctatgggat cttcttcatg gacctacca gaagaaaaag 660  
 cttgactggg agctgcgtct aaaaatagca cttggagcag cacaagggtc tgcttatcta 720  
 caccatgatt gctgtcctag aatcatccac agagatgtga aatcatctaa cattctattg 780  
 gatgcagact ttgagcctca tctcactgat ttggcattg ccaaaagtct ctgcccctca 840  
 aagtcccata cttctactta cataatgggc acaattggct atatagacc tgagtatgct 900  
 agaacttcac gtctcactga gaagtctgat gtgtacagtt acggtattgt tttacttgag 960  
 ttgctaactg gaaggaaagc tgttgacaat gaatccaacc tccaccatct gattttgtcc 1020  
 aaggcagcaa ccaatgcagt gatggaaaca gttgatccag acattactgc cacatgcaag 1080  
 gacctaggag ctgtaaaaaa gtttatcag cttgctctat tatgcacaaa gaggcagcca 1140  
 gctgataggc cgacaatgca cgaagtgaca cgtgtactcg gaagccttgt gctgtccaac 1200  
 accccaccaa agcaactagc tgcactacca cctgcttcag atccatctgc canagtgcca 1260  
 tgctacgtgg atgagtatgg caaacctcaa gactccacac ttggtgaaac tgcacctcaa 1320  
 tgagccacct aaatgctcaa cttcttctca agtttggaga agtaatctct tcaacaatg 1380  
 aggtgaaaga gtgaagccgg ga 1402

<210> 20

<211> 440

<212> PRT

<213> 大豆(Glycine max)

<220>

<221> 变体

<222> 418

<223> Xaa = 任意氨基酸

<400> 20

Leu Ser Leu Ser Leu Leu Asn Val Ser Tyr Asn Lys Leu Phe Gly Val  
 1 5 10 15  
 Ile Pro Thr Ser Asn Asn Phe Thr Arg Phe Pro Pro Asp Ser Phe Ile  
 20 25 30  
 Gly Asn Pro Gly Leu Cys Gly Asn Trp Leu Asn Leu Pro Cys His Gly  
 35 40 45  
 Ala Arg Pro Ser Glu Arg Val Thr Leu Ser Lys Ala Ala Ile Leu Gly  
 50 55 60  
 Ile Thr Leu Gly Ala Leu Val Ile Leu Leu Met Val Leu Val Ala Ala  
 65 70 75 80  
 Cys Arg Pro His Ser Pro Ser Pro Phe Pro Asp Gly Ser Phe Asp Lys  
 85 90 95  
 Pro Ile Asn Phe Ser Pro Pro Lys Leu Val Ile Leu His Met Asn Met  
 100 105 110  
 Ala Leu His Val Tyr Glu Asp Ile Met Arg Met Thr Glu Asn Leu Ser



<212> DNA

<213> 大豆 (Glycine max)

<400> 21

```

gagtatgcta gaacttcgca tctcactgag aagtctgatg tgtacagtta tggatttgtt 60
ttactcgagt tgctaactgg aaggaaagct gttgacaatg aatccaacct ccaccatctt 120
atlttgcca aggcagcaac caatgctgtg atggaacagc ttgatcccga cattactgcc 180
acatgcaagg acctaggagc tgtaaaaaag gtttatcagc ttgctctatt atgcacaaag 240
aggcagccag ctgataggcc aacaatgcac gaagtgcacac gtgtactcgg aagtctcgtg 300
ccatcaagca tcccacaaa gcaactagct gacctaccac ctgcttcaaa tccatctgcc 360
aaagtgccat gctacgtgga tgagtatgca aacctcaaaa cccacactt agtaaactgc 420
ccctcaatga gcacttcaga tgctcaactc ttcctcaagt ttggagaagt aatctctcaa 480
aacagtgagt ga 492

```

<210> 22

<211> 163

<212> PRT

<213> 大豆 (Glycine max)

<400> 22

```

Glu Tyr Ala Arg Thr Ser His Leu Thr Glu Lys Ser Asp Val Tyr Ser
 1           5           10          15
Tyr Gly Ile Val Leu Leu Glu Leu Thr Gly Arg Lys Ala Val Asp
 20          25          30
Asn Glu Ser Asn Leu His His Leu Ile Leu Ser Lys Ala Ala Thr Asn
 35          40          45
Ala Val Met Glu Thr Val Asp Pro Asp Ile Thr Ala Thr Cys Lys Asp
 50          55          60
Leu Gly Ala Val Lys Lys Val Tyr Gln Leu Ala Leu Leu Cys Thr Lys
 65          70          75          80
Arg Gln Pro Ala Asp Arg Pro Thr Met His Glu Val Thr Arg Val Leu
 85          90          95
Gly Ser Leu Val Pro Ser Ser Ile Pro Pro Lys Gln Leu Ala Asp Leu
100         105         110
Pro Pro Ala Ser Asn Pro Ser Ala Lys Val Pro Cys Tyr Val Asp Glu
115         120         125
Tyr Ala Asn Leu Lys Thr Pro His Leu Val Asn Cys Pro Ser Met Ser
130         135         140
Thr Ser Asp Ala Gln Leu Phe Leu Lys Phe Gly Glu Val Ile Ser Gln
145         150         155         160
Asn Ser Glu

```

<210> 23



&lt;211&gt; 797

&lt;212&gt; DNA

&lt;213&gt; 大豆(Glycine max)

&lt;400&gt; 23

```

ggaaaatggc agtctatggg atcttctgca tggacctacc aagaagaaaa agcttgattg 60
ggatcttcgt ctaaaaatag cactaggatc agcccaaggg cttgcttacc tacaccatga 120
ttgcagtcca ctcacattc acagggacgt gaaatcatct aatattttac tagacaaaga 180
ctttgagccc catctcgtcg atttcggcat tgcaaaaagt ctatgcccat ctaagaccca 240
cacttcaact tacataatgg gcacaattgg ctacatagac cctgagtatg ctagaacttc 300
ccgcctcact gagaagtccg atgtgtatag ctatggatc gtattgcttg agcttctaac 360
tgaggagaaa gctgttgaca acgaatcaaa cctccatcat ctgattttgt ccaagacagc 420
taatgatggc gtaatggaaa ccggtgatcc agatattact accacatgca gggacatggg 480
agcagtaaaa aaggtttttc agcttgctct tttatgcaca aagaagcaac cagtcgatag 540
gctacaatg catgaagtga ctgcgctctg ggaagccttg tgccatccat aaccctacca 600
aaacaaactg actcaacaca agtctgtgta cctgattctc agtcatctgc taaaatgcaa 660
tgctacaaag atgaatatgc aaatctcaaa actcctcact tggtaactg tccttccatg 720
agcacctctg atgcacaact cttcctcaaa tttggggaag taatatcaca aaatagtcac 780
tgaggatttt gataatc                                     797

```

&lt;210&gt; 24

&lt;211&gt; 196

&lt;212&gt; PRT

&lt;213&gt; 大豆(Glycine max)

&lt;400&gt; 24

```

Glu Asn Gly Ser Leu Trp Asp Leu Leu His Gly Pro Thr Lys Lys Lys
 1           5           10          15
Lys Leu Asp Trp Asp Leu Arg Leu Lys Ile Ala Leu Gly Ser Ala Gln
 20          25          30
Gly Leu Ala Tyr Leu His His Asp Cys Ser Pro Leu Ile Ile His Arg
 35          40          45
Asp Val Lys Ser Ser Asn Ile Leu Leu Asp Lys Asp Phe Glu Pro His
 50          55          60
Leu Ala Asp Phe Gly Ile Ala Lys Ser Leu Cys Pro Ser Lys Thr His
 65          70          75          80
Thr Ser Thr Tyr Ile Met Gly Thr Ile Gly Tyr Ile Asp Pro Glu Tyr
 85          90          95
Ala Arg Thr Ser Arg Leu Thr Glu Lys Ser Asp Val Tyr Ser Tyr Gly
100         105         110
Ile Val Leu Leu Glu Leu Leu Thr Gly Arg Lys Ala Val Asp Asn Glu
115         120         125
Ser Asn Leu His His Leu Ile Leu Ser Lys Thr Ala Asn Asp Gly Val
130         135         140
Met Glu Thr Val Asp Pro Asp Ile Thr Thr Thr Cys Arg Asp Met Gly

```

145                                    150                                    155                                    160  
 Ala Val Lys Lys Val Phe Gln Leu Ala Leu Leu Cys Thr Lys Lys Gln  
    165                                    170                                    175  
 Pro Val Asp Arg Pro Thr Met His Glu Val Thr Arg Val Trp Glu Ala  
    180                                    185                                    190  
 Leu Cys His Pro  
    195

<210> 25

<211> 279

<212> DNA

<213> 大豆(Glycine max)

<400> 25

agtacagttt ataaatgtgt ccttaaaaat tgcaagccgg ttgctatcaa gaagctctat 60  
 tcccactacc cacaatactt gaaagagttt gagactgagc ttgagacagt ttgtagcggt 120  
 aagcacagaa atctggtcag tctccaaggc tactctttgt caacgtacgg aaatcttctc 180  
 ttttatgact acatggaaaa tggcagtcta tgggatcttc tgcattggacc taccaagaag 240  
 aaaaagcttg attgggatct tctgtctaaaa atagcacta 279

<210> 26

<211> 93

<212> PRT

<213> 大豆(Glycine max)

<400> 26

Ser Thr Val Tyr Lys Cys Val Leu Lys Asn Cys Lys Pro Val Ala Ile  
 1                                    5                                    10                                    15  
 Lys Lys Leu Tyr Ser His Tyr Pro Gln Tyr Leu Lys Glu Phe Glu Thr  
    20                                    25                                    30  
 Glu Leu Glu Thr Val Gly Ser Val Lys His Arg Asn Leu Val Ser Leu  
    35                                    40                                    45  
 Gln Gly Tyr Ser Leu Ser Thr Tyr Gly Asn Leu Leu Phe Tyr Asp Tyr  
    50                                    55                                    60  
 Met Glu Asn Gly Ser Leu Trp Asp Leu Leu His Gly Pro Thr Lys Lys  
 65                                    70                                    75                                    80  
 Lys Lys Leu Asp Trp Asp Leu Arg Leu Lys Ile Ala Leu  
    85                                    90

<210> 27

<211> 2931

<212> DNA

<213> 拟南芥(Arabidopsis thaliana)

&lt;400&gt; 27

atggctctgt ttagagatat tgttcttctt gggtttctct tctgcttgag cttagtagct 60  
actgtgactt cagaggagg agcaacgttg ctggagatta agaagtcatt caaagatgtg 120  
aacaatgttc tttatgactg gacaacttca ctttcttcgg attattgtgt ctggagaggt 180  
gtgtcttggtg aaaatgtcac ctcaatgtt gttgctctta attgtcaga tttgaatctt 240  
gatggagaaa tctcacctgc tattggagat ctcaagagtc tcttgtcaat tgatctgcga 300  
ggtaatcgct tgtctggaca aatccctgat gagattggg actgttcttc tttgcaaaac 360  
ttagacttat cttcaatga attaagtggt gacataccgt tttcgatttc gaagttgaag 420  
caacttgagc agctgattct gaagaataac caattgatag gaccgatccc ttcaacactt 480  
tcacagattc caaacctgaa aattctggac ttggcacaga ataaactcag tgggtgagata 540  
ccaagactta tttactggaa tgaagttctt cagtatcttg ggttgcgagg aaacaactta 600  
gtcgttaaca tttctccaga tttgtgtcaa ctgactggtc tttggtattt tgacgtaaga 660  
aacaacagtt tgactggtag tatacctgag acgataggaa attgcaactgc cttccagggt 720  
ttggacttgt cctacaatca gctaactggt gagatccctt ttgacatcgg cttcctgcaa 780  
gttgcaacat tatcattgca aggcaatcaa ctctctggga agattccatc agtgattggt 840  
ctcatgcaag cccttgcaag cttagatcta agtggcaact tgttgagtgg atctattcct 900  
ccgattctcg gaaatcttac tttcacaggag aaattgtatt tgcacagtaa caagctgact 960  
ggttcaattc cacctgagct tggaaacatg tcaaaaactcc attacctgga actcaatgat 1020  
aatcatctca cgggtcatat accaccagag cttgggaagc ttactgactt gtttgatctg 1080  
aatgtggcca acaatgatct ggaaggacct atacctgac atctgagctc ttgcacaaat 1140  
ctaaacagct taaatgttca tgggaacaag tttagtggca ctataccccg agcatttcaa 1200  
aagctagaaa gtatgactta ctttaatctg tccagcaaca atatcaaagg tccaatccccg 1260  
gttgagctat ctctgatcgg taacttagat acattggatc tttccaacaa caagataaat 1320  
ggaatcattc cttcttccct tgggtatttg gagcatcttc tcaagatgaa cttgagtaga 1380  
aatcatataa ctgggttagt tccaggcgac tttggaaatc taagaagcat catggaaata 1440  
gatctttcaa ataatgatat ctctggceca attccagaag agcttaacca attacagaac 1500  
ataattttgc tgagactgga aaataataac ctgactggta atgttggttc attagccaac 1560  
tgtctcagtc tcaactgtatt gaatgtatct cataacaacc tcgtaggtga tatccctaag 1620  
aacaataact tctcaagatt ttaccagac agcttcatig gcaatcctgg tctttgcggt 1680  
agttggctaa actcaccgtg tcatgattct cgtcgaactg tacgagtgtc aatctctaga 1740  
gcagctattc ttggaatagc tattggggga ctigtgatcc ttctcatggt cttaatagca 1800  
gcttgccgac cgcataatcc tctcctttt cttgatggat cacttgacaa accagtaact 1860  
tattcgacac cgaagctcgt catccttcat atgaacatgg cactccacgt ttacgaggat 1920  
atcatgagaa tgacagagaa tctaagtgag aagtatatca ttgggcacgg agcatcaagc 1980  
actgtataca aatgtgtttt gaagaattgt aaaccggttg cgattaagcg gctttactct 2040  
cacaaccac agtcaatgaa acagtttgaa acagaactcg agatgctaag tagcatcaag 2100  
cacagaaatc ttgtgagcct acaagcttat tccctctctc acttggggag tcttctgttc 2160  
tatgactatt tggaaaatgg tagcctctgg gatcttcttc atggccctac gaagaaaaag 2220  
actcttgatt gggacacacg gcttaagata gcatatgggt cagcacaagg tttagcttat 2280  
ctacaccatg actgtagtcc aaggatcatt cacagagacg tgaagtcgtc caacattctc 2340  
ttggacaaaag acttagaggc tcgtttgaca gatthttggaa tagcgaaaag cttgtgtgtg 2400  
tcaaagtcac atacttcaac ttacgtgatg ggcacgatag gttacataga ccccgagtat 2460  
gctcgcactt cacggctcac tgagaaatcc gatgtctaca gttatggaat agtccttctt 2520  
gagttgttaa cccgaaggaa agccgttgat gacgaatcca atctccacca tctgataatg 2580

tcaaagacgg ggaacaatga agtgatggaa atggcagatc cagacatcac atcgacgtgt 2640  
 aaagatctcg gtgtggtgaa gaaagttttc caactggcac tcctatgcac caaaagacag 2700  
 ccgaatgatc gaccacaat gcaccaggtg actcgtgttc tcggcagttt tatgctatcg 2760  
 gaacaaccac ctgctgcgac tgacacgtca gcgacgctgg ctggttcgtg ctacgtcgat 2820  
 gagtatgcaa atctcaagac tectcattct gtcaattgct ctccatgag tgcttctgat 2880  
 gctcaactgt ttcttcgggt tggacaagtt atttctcaga acagtgagta g 2931

<210> 28

<211> 2901

<212> DNA

<213> 拟南芥(*Arabidopsis thaliana*)

<400> 28

atgaaggaga agatgcagcg aatggtttta tctttagcaa tggtagggtt tatggttttt 60  
 ggtgttgctt cggctatgaa caacgaaggg aaagctctga tggcgataaa aggctctttc 120  
 agcaacttag tgaatatgct ttggattgg gacgatgttc acaacagtga cttgtgttct 180  
 tggcgagggtg ttttctgcga caacgttagc tactccgttg tctctctgaa tttgtccagt 240  
 ctgaatcttg gaggggagat atctccagct attggagacc tacggaattt gcaatcaata 300  
 gacttgcaag gtaataaact agcaggtcaa attccagatg agattggaaa ctgtgcttct 360  
 cttgtttatc tggatttgtc cgagaatctg ttatatggag acataccttt ctcaatctct 420  
 aaactcaagc agcttgaaac tclgaatctg aagaacaatc agctcacagg tcctgtacca 480  
 gcaaccttaa cccagattcc aaaccttaag agacttgatc ttgctggcaa tcatctaacg 540  
 ggtgagatat cgagattgct ttactggaat gaagtttgc agtatcttgg attacgaggg 600  
 aatatgttga ctggaacgtt atcttctgat atgtgtcagc taaccggtt gtggtacttt 660  
 gatgtgagag gaaataatct aactggaacc atcccggaga gcatcgaaa ttgcacaagc 720  
 ttcaaatcc tggacatate ttataatcag ataacaggag agattcctta caatctcggc 780  
 ttctccaag ttgctactct gtcacttcaa ggaaacagat tgacgggtag aattccagaa 840  
 gttattggtc taatgcaggc tcttgctggt ttggatttga gtgacaatga gcttgttgg 900  
 cctatcccac cgatacttgg caatctctca tttaccgga agttgtatct ccatggcaat 960  
 atgctcactg gtccaatccc ctctgagctt gggaatatgt cacgtctcag ctatttgcag 1020  
 ctaaacgaca ataaactagt gggaactatt ccacctgagc ttggaagct ggagcaattg 1080  
 ttgaaactga atcttgccaa caaccgttta gtagggccca taccatccaa cattagtcca 1140  
 tgtgcagcct tgaatcaatt caatgttcat gggaacctct tgagtggatc tattccactg 1200  
 gcgtttcgca atctcgggag cttagcttat ctgaatcttt cgtcgaacaa tttcaagga 1260  
 aaaaataccag ttgagcttgg acatataatc aatcttgaca aactagatct gtctggcaat 1320  
 aacttctcag ggtctatacc attaacgctt ggcgatcttg aacaccttct catattaaat 1380  
 cttagcagaa accatcttag tggacaatta cctgcagagt ttgggaacct tcgaagcatt 1440  
 cagatgattg atgtatcatt caatctgctc tccggagtta ttccaactga acttggccaa 1500  
 ttgcagaatt taaactcttt aatattgaac aacaacaagc ttcattgggaa aattccagat 1560  
 cagcttacga actgcttcac tcttgtcaat ctgaatgtct ccttcaacaa tctctccggg 1620  
 atagtcccac caatgaaaaa cttctcacgt tttgctccag ccagctttgt tggaaatcca 1680  
 tatctttgtg gaaactgggt tggatctatt tgtgttctt taccgaaatc tcgagtatc 1740  
 tccagaggtg ctttgcctct cattgttctt ggcgtcatca ctctctatg tatgattttc 1800  
 cttgcagttt acaaatcaat gcagcagaag aagattctac aaggctctc aaaacaagct 1860  
 gaagggttaa ccaagctagt gattctccac atggacatgg caattcatac atttgcagat 1920

atcatgagag tgactgagaa tcttaacgaa aagtttataa ttggatatgg tgcttctagc 1980  
 acggtataca aatgtgcatt aaaaagtcc cgacctattg ccattaagcg actctacaat 2040  
 cagtatccgc ataacttgcg ggaatttgag acagaacttg agaccattgg gagcattagg 2100  
 cacagaaaca tagtcagctt gcatggatat gccttgtctc ctactggcaa ccttctttc 2160  
 tatgactaca tggaaaatgg atcactttgg gaccttcttc atgggtcatt gaagaaagt 2220  
 aagcttgatt gggagacaag gttgaagata gcggttgag ctgcacaagg actagcctat 2280  
 cttcaccacg attgtactcc tcgaatcatt caccgtgaca tcaagtcac gaacatactt 2340  
 cttgatgaga atttcgaagc acatttatct gatttcggga ttgctaagag cataccagct 2400  
 agcaaaacc atgcctcgac ttatgttttg ggaacaatg gttatataga cccagagtat 2460  
 gctcgtactt cacgaatcaa tgagaaatcc gatataatac gcttcgggat tgttcttctt 2520  
 gagcttctca ctgggaagaa agcagtggat aacgaagcta acttgcac caactgatattg 2580  
 tcaaaggctg atgataatac tggatggaa gcagttgac cagaggttac tgtgacttgt 2640  
 atggacttgg gacatatcag gaagacattt cagctggctc tcttatgcac aaagcgaac 2700  
 ccttagaga gaccacaat gcttgaagtc tctagggttc tgctctctct tgtccatct 2760  
 ctgcaagtag caaagaagct accttctctt gatcaactca ccaaaaagct gcagcaagag 2820  
 aatgaagtta ggaatcctga tgcagaagca tctcaatggt ttgttcagtt cctggaagtc 2880  
 atctccaaa gtagcatata a 2901

<210> 29

<211> 2985

<212> DNA

<213> 玉米 (*Zea mays*)

<400> 29

atgcctgtgc gcagctcagt ggccatgac acgacggccg cccgtgctct cgccgccctc 60  
 gtgctcgtca ccgccccgc cgccgccgc gccgtgccg acgatggggc ggcgctggg 120  
 gagatcaaga agtccttccg caacgtcggc aacgtactgt acgattgggc cggcgacgac 180  
 tactgtcctt ggcgcgccgt cctgtgcgac aacgtcacgt tcgccgtcgc cgcgctcaac 240  
 ctctctggcc tcaacctga ggtgagatc tctccagccg tcggcagcct caagagcctc 300  
 gtctccatcg acctcaagtc aaatggccta tccgggcaga tcctgatga gattggcgat 360  
 tgttcgtcac ttaggacgct ggacttttca ttcaacaact tggacggcga catacattt 420  
 tcaatatcaa agctgaagca cctggagaac ttgatattga agaacaaccg gctgattggt 480  
 gcgatcccct caacattgtc acagctccca aatttgaaga ttctggactt ggcacaaaac 540  
 aaactgactg gggagatacc gagccttacc tattggaacg aggttcttca atacttgggt 600  
 ttgcggggaa atcatttaga aggaagcctc tctcctgata tgtgccagct tactggcctt 660  
 tggacttttg atgigaagaa caatagtttg actggggcga taccagacac cattgggaac 720  
 tgtacaagtt ttcaggctct ggatttgtct tacaaccgct ttactggacc aatcccattc 780  
 aacattgggt tctacaagt ggctacacta tccttgcaag ggaacaagtt cactggccca 840  
 attccttcag taaltggcct tatgcaggct ctgctgtcc tagatctgag ttacaaccaa 900  
 ttatctggtc ctataaccatc tatactaggc aacttgacat acactgagaa gctgtacatg 960  
 caaggcaaca ggttaactgg atcgatacca ccagagctag gaaatatgtc aacacttcat 1020  
 tacctagaac tgaatgataa tcaacttact gggtaattc caccagagct tggaggccta 1080  
 acaggcttgt ttgacctgaa ccctgcgaat aaccacctg aaggaccaat tctgacaac 1140  
 ctaagttcat gtgtgaatct caatagcttc aatgcttatg gcaacaagtt aatggaacc 1200  
 attcctcgtt cgctgcggaa acttgaagc atgacctatt taaatcttc atcaaattc 1260

ataagtggtt ctattcctat tgagctatca aggatcaaca atttgacac gttgggctta 1320  
 tctgtaca tgatgacggg tccaattcca tcatccattg gcaacctaga gcatctattg 1380  
 aggcctaact tgagcaagaa tgatctagtt ggattcatcc ctgcggagtt tgtaatttg 1440  
 ggaagtgtca tggagattga tttatcctat aatcatcttg gtggtctgat tcctcaagaa 1500  
 cttggaatgc tgcaaacct gatgttgcta aaactggaaa acaacaatat aactggcgat 1560  
 gtctcttctc tgatgaactg cttcagcctc aatatcttaa atgtgcata caataattta 1620  
 gctgggtctg tccctactga caacaacttc acacggtttt cacatgacag ctttttaggt 1680  
 aatcctggac tctgtggata ttggcttggg tcttcatgtc gttccactgg ccaccgagac 1740  
 aaaccgcaa tctcaaaggc tgccataatt ggtgttgctg tgggtggact tgttatcctc 1800  
 ctgatgatct tagtagctgt atgcaggcca caccatccac ctgcttttaa agatgccact 1860  
 gtaagcaagc cagtgagcaa tggccaccc aagctgatga tcttcatat gaacatggct 1920  
 cttcatgtct ttgatgatat aatgaggatg actgagaact tgagtgagaa atacatcatt 1980  
 ggatacgggg catcaagtac agttataaa tgtgttctaa agaattgcaa accagtggca 2040  
 ataaaaagc tgtatgcca ctaccctcag agccttaagg aattgaaac tgagctcgag 2100  
 actgttggtg gcatcaaca cgggaatcta gtcagcctc aagggtactc gttgtcacct 2160  
 gtigggaacc tctctttta tgattatatg gagagtggca gcttatggga tgttttacat 2220  
 gaaggtcat ccaaggagaa caaacttgac tgggtgactc gctacggat cgctcttggg 2280  
 gcagctcaag gcctcgctta cctcaccat gactgcagcc cacgaataat tcaccgggac 2340  
 gtaaatcaa agaataact cctcgacaaa gattatgagg ccatcttac agacttcggc 2400  
 atcgtaaga gcttatgtgt ctogaagact cacacgtcaa cctacgtcat gggcactatt 2460  
 ggttacattg atcccagta cggccgacc tcccgcctca acgagaagtc tgatgtctac 2520  
 agctacggca tegtcttctg ggagctgctg accggcaaga agccagtgga caacgagtgc 2580  
 aatctccatc acttgatcct atcgaagacg gcgagcaacg aggtcatgga gacggtggac 2640  
 cccgacgtgg gagacacctg caaggacctg ggcgaggtga agaagctggt ccagctggcg 2700  
 ctctctgca ccaagcggca gcctcggac cggccgacga tgcacgaggt ggtgcgcgctc 2760  
 cttgactgcc tgggtgaacc ggagccgccc ccgcagccc agcagcagca gcagaaggcg 2820  
 cacgcgcacc accagctgcc gccgcagccc tcgccccgg cctacgtcga cgagtacgct 2880  
 agcctgcggg gcaccggcgc cctctctcgc gccaaactcg ccagcacctc ggacgccgag 2940  
 ctgttctca agttcggcga ggccatctcg cagaacatgg tgtag 2985

<210> 30

<211> 2931

<212> DNA

<213> 玉米 (*Zea mays*)

<400> 30

gaaacactgc tggagatcaa gaaatccttc cgcgacggcg gcaacgcgct gtacgattgg 60  
 tccggcgacg ggcgctcgcg gggctactgc tctgtggcgc gcgtgctatg cgacaacgctc 120  
 accttcgctg tcgcgcgctc caacctctct gggctcaatc tcgaggggtga aatctcagcg 180  
 gccatcggga gtctgcaacg tcttgtctca atcgatttga agtcgaatgg actctctgga 240  
 cagatccccg atgagattgg tgattgttct ttgctcgaaa ctttggattt gtcactaac 300  
 aatctagaag gagacatacc attctccatg tccaagctga agcaccttga gaacttgatt 360  
 ttgaagaaca acaactgggt gggagtgatc ccatcgacac tctctcaact tccaaatttg 420  
 aagatattgg acttggctca aaacaagtta agtgggtgaa ttccgaatct aatatattgg 480  
 aatgaggttc ttcaatactt gggattgcca agcaatagtt tagaaggaag cctctctccc 540

gatatgtgcc agttaactgg tctgtggtac tttgatgtga agaacaatag cttgacgggt 600  
gcaataaccag aaaccatagg gaactgtacg agctttcagg tcttagattt gtcaaacaat 660  
catcttactg gagaaatccc gttcaatatt ggtttcctgc aagtggctac gttatctttg 720  
caaggaaca agttctctgg tcctatacca tcagtgattg gccttatgca ggcgcttgca 780  
gtgctagatc tgagtttcaa tgagctatct ggccaatac cctctatact gggcaacttg 840  
acatacactg agaaattata cctgcaaggc aataggttaa ctggattgat accgccagag 900  
cttggtaata tgcgacact gcattacctg gaactgaacg acaatctggt gactggggtc 960  
attcctcctg atcttgaaa acttacagaa ttgtttgaat tgaacctgac aaacaacaac 1020  
cttataggac ctatccctga gaatttaagt tcatgtgcaa atctcattag tttcaatgct 1080  
tatggcaata aattgaatgg aaccattcca cgttcatttc acaagcttga gagtctgact 1140  
tatctgaatc tgcatacaaa tcatctcagt ggagcacttc caattgaggt tgcaagaatg 1200  
agaaatttgg aactctgga cttatcctgt aacatgatca ctggttcaat tccctcggct 1260  
attgggaaac tagagcatct tttgaggctc aacttaagca aaaataatgt ggctggacac 1320  
attcctgctg aatttgggaa cttaggagc atcatggaga ttgattgtc ttacaaccac 1380  
ctcagtgccc tgattctca agaggttggg atgctacaaa atttgatact gttaaaatta 1440  
gaaagcaata atattactgg agatgtctct tcacttattt actgcttgag tctcaatata 1500  
ttaaatgtat catacaacca tctttatggt actgtaccta cagacaacaa ctctcacga 1560  
ttttcaccgg acagcttctt gggtaacctt ggactttgtg gctattggct tcaactctgct 1620  
tcatgcacac aattatccaa tgcagagcaa atgaagagat cctctagcgc aaaggcctca 1680  
atgtttgcag ctattgggtg tgggtgcocta ttgcttgta ttatgctcgt tatectagta 1740  
gttatttgct ggccacataa ctctccagtg ctcaaagatg tctctgtaaa caaacagat 1800  
aaccttgctt cagcatcaaa caacattcat cccaagcttg tgatctcca catgaacatg 1860  
gccctctatg tatatgatga tataatgagg atgactgaaa acttgagcga aaaatacatt 1920  
attggttatg gagcctcaag tacagtctac agatgagacc tgaagaactg caagccaatt 1980  
gcgattaaaa agctgtatgc tcaactacct cagagcttga aggaattoga gactgaactt 2040  
gagactgttg gaagcatcaa acaccggaat ctgtgaagcc ttcaggggta ctccctgtca 2100  
ccatctggga atctcctctt ctatgattac atggaaaatg gcagcctctg ggacatttta 2160  
catgcttcat cgaagaaaaa gaaactcgat tgggaggctc gcctcaagat tgctctcgga 2220  
gctgctcaag gcctggetta tcttcacat gaatgcagtc cacgaataat ccacagggat 2280  
gtgaagtcaa agaatacct cctagacaaa gactacgagg ctcatcttgc tgacttcggt 2340  
attgccaaga gcttgtgtgt gtcgaagacg cacacatcaa cgtacgtgat gggcaccatt 2400  
ggctacattg accctgagta tgcacggaca tcccggatca acgagaaatc ggatgtgtac 2460  
agctacggca ttgtcttgct ggagctgctt accggcaaaa agcctgtcga cgacgagtgc 2520  
aaccttcacc acttgatcct atccaaagcc gcagaaaaca cggctatgga gacggtagac 2580  
caggacatca ccgacacgtg caaggacctc ggcgaggtca agaaggtgtt ccagctggcg 2640  
ctcctttgca gcaagaggca gccgtcggat cgaccgacca tgcacgaggt cgcgcgcgtc 2700  
ctggacagcc tcgtctgccc agcaggcccg cccccgaagc aggcgcaggc gcaggcacag 2760  
gcacaggcgt cggagaagcc gtccaccacg gcgccgagct atgtcagcga gtacgtcggc 2820  
ctacgaggcg gcggcggcgg cagcgcctc tctgcacca actegtcgag cgcgtccgac 2880  
gccgagctct tcatgaagtt tggcgagggt atctcgcgga gcacggaata g 2931

<210> 31

<211> 2937

<212> DNA

<213> 水稻(Oryza sativa)

&lt;400&gt; 31

atgacgccgg cgccggcggc cgcctcctac cgcgctctcg tcgctctcct gctcgtcgcc 60  
 gtcgccgttg ccgatgatgg gtcgacgctg ctggagatca agaagtcctt ccgcaatgtg 120  
 gacaacgtac tgtacgattg ggccggcggc gactactgct cgtggcggcg cgtcctctgc 180  
 gacaacgtca ccttcgccgt cgccgcgctc aacctatccg ggctcaacct cggaggcgag 240  
 atctctccgg ccgtcggcag gttgaagggc atcgtctcga ttgacttgaa gtcgaatggg 300  
 ctgtctgggc agatccctga tgagattggc gattgttcat cactaaaaac tctggatttg 360  
 tctttcaata gcttggatgg ggacattccg ttctcagtat cgaagctgaa gcacattgag 420  
 agcttgatat tgaagaacaa ccaactgac ggagtgatcc catcaacgct ctcacagctc 480  
 ccaaatttga agattttgga cttggcacag aacaaactga gtggagagat accaagactg 540  
 atatatgga acgaggttct tcaataactg ggattacgcg gtaataattt agaaggcagc 600  
 atctcccag atatatgcca gttgactggg ctttgggtact ttgacgtaaa gaacaacagc 660  
 ttgactgggc cgataccaga aaccattggg aactgtacaa gtttccaggt cttggatttg 720  
 tcttacaata aactttctgg atcaattcct ttcaacattg gtttctaca agttgctaca 780  
 ctatctttgc aagggaaact gtttactggg cctattccat cagtattggg acttatgcag 840  
 gctctcgctg tactggatct gagttacaac caattgtctg gtcctattcc atcgatacta 900  
 ggcaatttaa catacactga gaagctgtat atgcaaggca ataagttaac aggtccaata 960  
 ccacctgagc ttggaaat atgcaaccctt cactacttag aacttaacga taatcaactt 1020  
 agcgggttca ttctccaga gttcggaaag ctaacagggt tatttgactt aaaccttgc 1080  
 aacaacaact ttgaaggctc aatccctgat aacataagct catgtgtgaa tctcaatagc 1140  
 ttcaatgctt atggcaacag attaaatggg accattctc cttcattgca taaacttgag 1200  
 agcatgactt atttgaattt gtcacaaat tttctaagtg gttctattcc tattgagcta 1260  
 tcgagaatca acaatttggg caccitggat ttatcctgta acatgattac tggcccaatt 1320  
 ccatcaacca ttgggagttt ggagcatcta ttaagactta acttgagcaa caatggctca 1380  
 gtaggattca ttctgcaga aattggcaac ttgaggagta tcatggagat tgatattgctc 1440  
 aacaatcacc ttggcggttt gattcctcaa gaactcggaa tgctcaaaa tctgatgttg 1500  
 ttaaacttca aaaacaacaa cataactggg gatgtctctt cactgatgaa ctgcttcagc 1560  
 ctcaatatct taaatgtatc ctataataat ttggctgggt ttgtacctac tgataacaac 1620  
 ttctcaggtt ttctgcctga cagctttttg ggtaatccag gactttgtgg atattggctt 1680  
 ggttcttctg gccgttcac tggccatcaa cagaaaccac taatctcaa ggctgcaata 1740  
 cttggaattg ccgtgggttg gcttgttacc ctctgatga tcttagtagc ggtctgcagg 1800  
 cctcatagtc cacctgtttt caaagatgct tctgttagca aaccagttag caatgttccc 1860  
 cccaagctgg ttatccttca tatgaacctt tcccttcttg tatacagagga tataatgacg 1920  
 atgactgaaa acctgagtga gaagtacatc attgggtacg gagcatccag cacggtttat 1980  
 aatgtgttt cgaagaaccg caaaccagtg gcagtaaaaa agctatatgc ccactatcca 2040  
 cagagcttca aggaatttga aactgagctt gagactgttg gtagcatcaa acaccggaat 2100  
 ctagtcagtc ttcaaggata ttccctatct cctgttggaa atcttctctt ctacgattac 2160  
 atggaaaatg gaagcctctg ggatgttttg catgaaggct caactaagaa gaaaaaactt 2220  
 gattgggaaa ctctgtctacg aattgtctca ggtgcggccc aaggccttgc ttatcttcat 2280  
 catgactgta gccacaggat aatacacagg gatgtgaaat caaaaaatat actccttgat 2340  
 aaagattatg aggcacatct tacagacttt ggcattgcta agagtttggt tgtttcaaaa 2400  
 actcacagct ccacctatgt catgggaact attggctata tcgatcctga gtatgctcgc 2460  
 acctcccgtc tcaatgaaaa gtctgatgct tacagctatg gcattgttct gcttgagctg 2520  
 ctgaccggaa aaaagccagt ggacaacgag tgcaatctcc atcacttgat cttgtcaaag 2580



acggctaaca atgctgtcat ggagacagtc gacccggaca ttgcagacac ttgcaaggat 2640  
 cttggtgagg tcaagaaggt gttccagctg gcgctcctt gcaccaagag acaaccatcg 2700  
 gateggccga caatgcacga ggttgtgcgc gtctggact gcctagtctg tcccgaccg 2760  
 ccaccgaagt cgcacagca gctggccatg ccgcagcggc ctgctgtccc gagctacatc 2820  
 aacgagtatg tcagcttaag aggcaccagc gtgctctcct gcgccaactc gtcgtgtact 2880  
 tccgatgctg agctgtttct caagtttggc gaggtcattt ctcagaacac agagtag 2937

<210> 32

<211> 3000

<212> DNA

<213> 水稻(Oryza sativa)

<400> 32

atggcggcgg cgagggcgcc gtggctgtgg tgggtgggtg tgggtggtgt tgggtgtggcg 60  
 gtggcggagg cggcctccgg aggaggagga gggggagatg gggaggggaa ggcgctgatg 120  
 ggcgtgaagg ccggtttcgg gaacgcggcc aacgcgctcg tcgactggga cggcggcgcc 180  
 gaccactgcg cgtggcgcgcg cgtcacctgc gacaacgcct ccttcgccgt cctcgccctg 240  
 aacttgtcaa atctaaacct aggaggtgag atctcgccgg ccatcggaga gctcaagaat 300  
 ctacagtctg ttgatctcaa ggggaacaag ctcactggcc aaatcccaga tgagattggg 360  
 gactgcatct ccttaaaata tttggatttg tctggcaact tgctgtatgg agacatcccc 420  
 ttctccatct ccaagctcaa gcagcttgag gagctgattt tgaagaacaa ccagctcacg 480  
 ggacccatcc cttccacatt gtcccaaatt ccaaattctca agacattgga cctggcacag 540  
 aaccagctta caggcgatat cccaaggctc atatactgga atgaagtct gcaataccta 600  
 ggtttgaggg gtaactcact gactggaact ttgtcacctg acatgtgcca actgactggc 660  
 ctgtggtagt ttgatgtaag gggaaacaat ctcacagggg ccattccaga gagcataggg 720  
 aactgcacca gctttgagat tctggacatt tcgtataacc aaatctctgg agaaatacct 780  
 tacaacatag gctttcttca agtagccaca ctgtcacttc aaggaaatag actgactggg 840  
 aaaattccag atgtgattgg cctgatgcaa gctcttgctg ttctagacct gaggtagaac 900  
 gagctggtag ggccattcc ttctatactg ggcaatctat cctatactgg aaaactatat 960  
 ttacatggga acaaacttac tggagtcata ccgcccggagc ttgggaacat gagtaaactt 1020  
 agctacctac aactgaatga taatgaattg gtgggcacaa ttccagcaga gcttggcaaa 1080  
 ctigaagagc tttttgaact aaatcttgcc aacaacaatc ttcaaggctc tttcctgca 1140  
 aacatcagtt cttgcaactg tctaaacaaa ttcaatgttt atggcaataa gctaaatggt 1200  
 tctattcctg ctggtttcca gaagttggag agtctgactt acttgaacct atcttcaaac 1260  
 aatttcaaag gcaatattcc tctgagctt ggtcacatca tcaacttggg cacattggat 1320  
 ctttctaca atgaattctc tggaccagtt cctgctacca ttggtgatct agagcacctt 1380  
 cttgaactga atttgagtaa gaaccatctt gatgggccag ttctgctga gtttggaaac 1440  
 ttgagaagcg tccaagtaat tgatagtcc aacaacaact tatctggtag tctgcccagg 1500  
 gaacttggac aacttcaaaa ccttgatagc ctgattctta acaacaacaa ttggttggg 1560  
 gagatccctg ctcaattggc caactgcttc agcttaaata accttgcatc tcaggaattt 1620  
 gtacataaac aatttatctg gacatgtccc gatggcaaag aacttctcga aattcccaat 1680  
 ggaaagcadc ttctaatttc tgattgcaac cagtacataa atcataaatg cagcttcttg 1740  
 ggtaatccat tactgcatgt ttactgocaa gattccagct gtggacactc tcatggacaa 1800  
 agagttaata tttcaaagac agcaattgct tgcattatct taggctttat catattgctc 1860  
 tgcgttctgc tgttggctat atataaaaca aatcaaccac agccacttgt caaaggatcc 1920

gataagccag tgcaaggacc tccaagcta gttgttctcc agatggacat ggctatccat 1980  
acttacgagg acatcatgag gctgacagag aatttgagcg agaaatacat cattggctat 2040  
ggcgctcaa gcaactgtcta caaatgtgaa ctcaagagcg gcaaggccat tgctgtcaag 2100  
cggctttaca gtcagtataa ccatagcctc cgagagtttg aaacagaact agagacaatt 2160  
ggcagcatalc ggcacaggaa tcttgtagc ctccatggct tctcgctatc tccacatgga 2220  
aacttgctct tctatgatta catggaaaat ggttcttgt gggatcttct ccacggteca 2280  
tcaaagaaag tgaagctcaa ctgggacaca agactgagga tcgcggtcgg agctgcacaa 2340  
gggctggcct atctccacca tgactgcaac cctcgcataa tccacagaga tgcaagtcc 2400  
tccaacatcc tgctcgacga gaacttcgaa ggcacacctc cagatttcgg catagccaaa 2460  
tgtgtcccct ctgccaagtc ccatgcctcc acttatgtgc taggaacctc cggctacatt 2520  
gatccggagt atgccaggac ttccaggctc aatgagaaat ctgatgtgta cagcttcggc 2580  
atcgtccttc tggaaattgct cacaggaag aaggccgtcg acaacgaatc gaacttgcatt 2640  
caattgatac tctccaaagc tgatgacaac acagtcatgg aggcagtgga ctcgagggtg 2700  
tcagtacagt gcacggacat gggactggtc aggaaggcct tccagctcgc cttctgtgc 2760  
accaagaggc acccttcaga ccggccgacc atgcacgagg ttgcaagggt gctgctctcc 2820  
ctgctgccgg cctccgccat gacaacgccc aagacgggtg actactcccg gttgctggcg 2880  
tcgacgacga cggcggccga catgcgaggg cagcagctga ccgacatcgg cgacaacagc 2940  
tctccgacg agcagtggtt cgtcaggttc ggcgaggtca tatccaagca cacaatgtga 3000

<210> 33

<211> 2790

<212> DNA

<213> 两色蜀黍 (*Sorghum bicolor*)

<400> 33

atgctgtcc gcagctcagt ggccatgacg acgacggccg cccgtgctct cgtcgccctc 60  
ctctcgtcg ccgtcgccgt cgcgacgat gggcgacgc tggaggagat caagaagtcc 120  
ttccgaacg tcggcaacgt actgtacgat tggccggcg acgactactg ctctggcgc 180  
ggcgtcctgt gcgacaacgt cacattcgcc gtcgctgccc tcaacctctc tggcctcaac 240  
cttgagggcg agatctctcc agccgtcggc agcctcaaga gcctcgtctc catcgatctg 300  
aagtcaaatg ggctatccgg gcagatccct gatgagattg gtgattgttc atcaactagg 360  
acgtggact tttcttcaaa caacttggal ggcgacatalc cattctctat atcaaagctg 420  
aagcacctgg agaacttgat attgaagaac aaccagctga ttggtgcat cccatcaaca 480  
ttgtcacagc tcccaaatg gaagatttg gatttggcac aaaacaaact gactggggag 540  
ataccaaggc ttatctactg gaatgaggtt cttcaatc ttgatgtgaa gaacaatagc 600  
ttgaccgggg tgataccaga cccattggg aactgtacaa gttttcaagt cttggatttg 660  
tcttacaacc gctttactgg accaatccca tcaacattg gtttctaca agtggtaca 720  
ctatccttgc aagggaacaa gttaccggg ccaattcctc cagtaattgg tcttatgcag 780  
gctctcgtg ttctagatct gagttacaac caattatctg gtcctatacc atcaatacta 840  
ggcaacttga catacactga gaagctgtac atccaaggca ataagttaac tgggtcgata 900  
ccaccagagt taggaaatat gtcaacactt cattacctag aactgaacga taatcaactt 960  
actgggtcaa ttccaccaga gcttgaagg ctaacaggct tgtttgacct gaaccttgcg 1020  
aataaccacc tggaggacc aattcctgac aacctaatg catgtgtgaa tctcaatagc 1080  
ttcaatgctt atggcaacaa gttaaatggg accattcctc gttcgttgcg gaaacttgaa 1140

agcatgacct atttaaatct gtcataaac ttcataagt gctctattcc tattgagtta 1200  
 tcaaggatca acaatttggga cacgctggat ttatcctgta acatgatgac tggccaatt 1260  
 ccatcatcaa ttggcagcct agagcatcta ttgagactta acttgagcaa gaatggtcta 1320  
 gttggattca tccccgcgga gtttgtaaat ttgaggagt tcatggagat tgatttatcc 1380  
 tataatcacc ttggtggcct gattcctcaa gaacttgaaa tgctgcaaaa cctgatgttg 1440  
 ctaaagtgt cgtacaataa tttggctggt gttgtccctg ctgacaacaa cttcacacgg 1500  
 ttttcacctg acagcttttt aggtaatcct ggactctgtg gatactggct tggttcgtcg 1560  
 tgtcgttcca ctggccacca cgagaaaccg cctatctcaa aggctgcat aattggtgtt 1620  
 gctgtgggtg gacttgttat cctcttgatg atcttagtag ctgtttgcag gccacatcgt 1680  
 ccacctgctt ttaaagatgt cactgtaagc aagccagtga gaaatgctcc cccaagctg 1740  
 gtgatccttc atatgaacat ggcccttcat gtatacagat acataatgag gatgactgag 1800  
 aacttgagtg agaaatacat cattggatac gggcgctcaa gtacagtta taaatgtgtc 1860  
 ctaaagaatt gcaaaccggt ggcaataaaa aagctgtatg cccactacc acagagcctt 1920  
 aaggaatttg aaactgagct tgagactggt gtagcatca agcaccgga tctagtcagc 1980  
 cttcaagggt actcattatc acctgttggg aacctcctt tttatgatta tatggaatgt 2040  
 ggcagcttat gggatgtttt acatgaaggt tcatccaaga agaaaaact tgactgggag 2100  
 actgccttac ggattgctct tgggtcagct caaggccttg cttacctca ccatgactgc 2160  
 agtccacgga taattcatcg ggatgtaaaa tcaaagaata tactccttga caaagattat 2220  
 gaggccatc ttacagactt tggaaatgct aagagcttat gtgtctcaaa aactcacaca 2280  
 tcaacctatg tcatgggaac tattggctac attgatcctg agtacgccg cacttcccgt 2340  
 ctcaacgaaa agtctgatgt ctacaggcta tggcattgtt ctgctggagc tgctgactgg 2400  
 caagaagcca gtggacaacg aatcctatcg aagacggcaa gcaacgaggt catggatacc 2460  
 gtggaccctg acatcgggga cacctgcaag gacctcggcg aggtgaagaa gctgttccag 2520  
 ctggcgtccc ttgcaacaa gcggcaaccc tggaccgac cgacgatgca cgagggtggtg 2580  
 cgcgtcctgg actgcctggt gaaccggac ccgccgcaa agccgtcggc gcaccagctg 2640  
 ccgcagccgt cgcagccgt gccaaactac atcaacgagt acgtcagcct gcggggcacc 2700  
 ggcgtctct cctgcgcaa ctcgaccagc acctcggacg ccgagctgtt ctcaagttc 2760  
 ggcgaggcca tctcgcagaa catggagtag 2790

<210> 34

<211> 1514

<212> DNA

<213> 两色蜀黍(Sorghum bicolor)

<400> 34

atggcagcga gggcggcggc ggcggtggtg ctgcttattg cggctgtggt gtcggtgtcg 60  
 gcgggaggag gtgaagggga cggagacggg cagacgctga tggcggtaa ggcggggttc 120  
 gggaacgcgg ccaacgcgct ggcggactgg gacggcggcc gcgaccactg cgcttggcgc 180  
 ggcgtcgcct gcgacgccg ctccttccgc gtcgtcggcc tgaacctgtc aatctaaac 240  
 ctcggagggg agatctcgc ggctataggg cagctcaaga gcctacagt cgtggatctc 300  
 aagctgaaca agctcacagg ccaaatecca gatgagattg gggatttgtt ctcttaaaa 360  
 tatttggatt tgcttgaaa cttgctgtat ggagacatcc ccttctccat ctccaagctc 420  
 aaacagcttg aggacctgat tttgaagaac aaccaactca cgggacccat ccctccaca 480  
 ctgtcccaga ttccaaatct caagacctg gatctggcgc agaacaagct caccggagac 540  
 attcccaggc tcatctactg gaatgaagta ctgcaatacc taggcttgag gggcaattca 600

```

ctgactggaa ctctgtcacc tgatatgtgc caactgactg gcctgtggta ttttgatgta 660
agggggaaca atctcacagg aacaattcca gagggcatag ggaactgcac tagctttgag 720
attctggata tttcatacaa ccaaatctct ggagaaatac cttacaacat aggttacctt 780
caagtagcca cacttgatct tagcgagaat gaacttgtgg gaccaattcc tccgatactt 840
ggcaacctgt cctacacagg caaactctat ttacatggca acaactcac gggacatata 900
ccaccagaac tggggaacat gagtaaactt agctacctgc agctgaatga caatgaacta 960
gtgggcacaa tcccagctga gcttggcaaa ctcacagagt tatttgaatt gaatcttgcc 1020
aacaacaatc ttgagggtca tttcctgca aacatcagct cttgcagtgc actgaacaaa 1080
ttcaatgtgt atggcaatag actgaatggc tctatcctg ctggttcca ggagttggag 1140
agttgacat acctgaacct ttctcaaac aatttcaaag gccagattcc ctctgagctt 1200
ggtcacatag tcaacttggg cacactagat ctttctaca atgaattctc tggaccagt 1260
cctcctaacta ttggtgatct cgagcatctt ctigaattga atttgagtaa aaaccatctt 1320
actggatctg tgcctgctga atttggaaac ttgagaagtg tccaagtaat tgacatatcc 1380
agcaacaact tgactgggta tctccctgaa gaacttggac agctgcaaaa cttgatagc 1440
ttgattctta acaacaacaa tttggttggg gagatccctg ctcagctggc taactgcttc 1500
agcttaatta cctt                                     1514

```

<210> 35

<211> 752

<212> DNA

<213> 两色蜀黍(*Sorghum bicolor*)

<400> 35

```

cttcaaatgg acatggctac ccatacctat gaagacatta tgaggttgac tgagaatttg 60
agcgagaaat acatcattgg ttatggggca tcaagtactg tgtacaaatg tgatctcaag 120
ggcggcaaaag ccacgctgt caaacggctt tacagtcagt ataaccacag cctccgtgag 180
tttgagacag aactggagac gatcggtagc atccgacaca ggaatctcgt cagccttcat 240
ggcttctcac tctcccctca tggaaacctg cttttctacg attacatgga aatggttcc 300
ctgtgggatc ttcttcatgg tccatcaaag aaggtgaagc ttgattggga cacaaggctt 360
aagattgctg taggtgctgc gcaaggactg gcctatcttc accatgactg caaccctcgc 420
ataatccaca gggatgtcaa gtectcaaac atcctgctcg acgagaactt cgaagcgcac 480
ctctctgatt tcggcatcgc caaatgtgtc ccggtgcca agtcccatgc ctccacctac 540
gtgcttggaa ccattggcta cattgatcca gagtatgcc gcacgtcgag gctcaatgag 600
aaatcggatg tctacagctt tggcatcgtc cttctggagc tgctcaccgg gaagaaggct 660
gtagaçaatg aatccaactt gcaccaattg atactctcaa aagctgacga caacacggtg 720
atggaggctg tggactcgga ggtgtcagtg ac                                     752

```

<210> 36

<211> 1323

<212> DNA

<213> 大豆(*Glycine max*)

<220>

<221> misc\_feature

<222> 1253

<223> n = A, T, C 或 G

<400> 36

```

ctiagtctct ctctacttaa tgtgtcctat aacaaactat ttggtgttat ccccacgagt 60
aacaacttta ccaggtttcc ccctgacagt ttcattggaa accctggctt ttgtggtaat 120
tggctgaatt tgccatgcca tgggtgctgt ccttcagagc gagttacatt atctaaggct 180
gccattcttg gaattacttt ggggtgcctt gtgattcttc ttatggtatt ggtggcagct 240
tgccgaccac acagcccctc tccttttctt gatggatcat ttgacaaacc aattaatttc 300
tcccctcaa agctagtgat tcttcatatg aatatggcac tacatgtgta tgaagatata 360
atgaggatga ctgaaaacct aagtgagaag tatacattg gatatgggtc atcaagtaca 420
gtttataaat gtgttcttaa gaattgtaag ccggtggcta tcaagaggat ctattctcac 480
tatccccaat gtattaaaga atttgaaact gaactcgaga ctggtggcag catcaagcac 540
cggaatttgg tcagtctcca aggctactcc ttgtcccat atggccatct cctgttttat 600
gactacatgg aaaatggcag tctatgggat cttcttcatg gacctaccaa gaagaaaaag 660
cttgactggg agctgcgtct aaaaatagca cttggagcag cacaagggtt tgcttatcta 720
caccatgatt gctgtcctag aatcatccac agagatgtga aatcatctaa cattctattg 780
gatgcagact ttgagcctca tctactgat tttggcattg ccaaaagtct ctgcccctca 840
aagtccata cttctactta cataatgggc acaattggct atatagaccg tgagtatgct 900
agaactcac gtctactga gaagtctgat ggtacagtt acggtatgt ttacttgag 960
ttgctaactg gaaggaaagc tgttgacaat gaatccaacc tccaccatct gattttgtcc 1020
aaggcagcaa ccaatgcagt gatggaaaca gttgatccag acattactgc cacatgcaag 1080
gacctaggag ctgtaaaaaa ggtttatcag cttgctctat tatgcacaaa gaggcagcca 1140
gctgataggc cgacaatgca cgaagtgaca cgtgtactcg gaagccttgt gctgtccaac 1200
accccaccaa agcaactagc tgcaactacca cctgcttcag atccatctgc canagtgcca 1260
tgctacgtgg atgagtatgg caaacctcaa gactccacac ttggtgaaac tgcccctcaa 1320
tga 1323

```

<210> 37

<211> 492

<212> DNA

<213> 大豆(Glycine max)

<400> 37

```

gagtatgcta gaacttcgca tctcactgag aagtctgatg tgtacagtta tggatttggt 60
ttactcgagt tgctaactgg aaggaaagct gttgacaatg aatccaacct ccaccatctt 120
atthttgtcca aggcagcaac caatgctgtg atggaaacag ttgatcccga cattactgcc 180
acatgcaagg acctaggagc tgtaaaaaag gttttatcagc ttgctctatt atgcacaaaag 240
aggcagccag ctgataggcc aacaatgcac gaagtgcac gtgtactcgg aagtctctgt 300
ccatcaagca tcccacaaa gcaactagct gacctaccac ctgcttcaaa tccatctgcc 360
aaagtgccat gctactgga tgagtatgca aacctcaaaa cccacactt agtaaactgc 420
ccctcaatga gcacttcaga tgctcaactc ttcctcaagt ttggagaagt aatctctcaa 480
aacagtgagt ga 492

```

<210> 38

<211> 591

<212> DNA

<213> 大豆(Glycine max)

<400> 38

```

gaaaatggca gtctatggga tcttctgcat ggacctacca agaagaaaaa gcttgattgg 60
gatcttcgtc taaaaatagc actaggatca gccaagggc ttgcttatct acaccatgat 120
tgcagtccac tcatcattca cagggacgtg aaatcatcta atattttact agacaaagac 180
tttgagcccc atctcgtgga tttcggcatt gcaaaaagtc tatgcccatac taagaccac 240
acttcaactt acataatggg cacaattggc tacatagacc ctgagtatgc tagaacttcc 300
cgcctcactg agaagtcoga tgtgtatagc tatggtatcg tattgcttga gcttctaact 360
gggaggaaag ctggtgacaa cgaatcaaac ctccatcacc tgattttgtc caagacagct 420
aatgatggcg taatggaaac cgttgatcca gatattacta ccacatgcag ggacatggga 480
gcagtaaaaa aggtttttca gcttgctctt ttatgcacaa agaagcaacc agtcgatagg 540
cctacaatgc atgaagtgac tcgctcttgg gaagccttgt gccatccata a 591

```

<210> 39

<211> 279

<212> DNA

<213> 大豆(Glycine max)

<400> 39

```

agtacagttt ataaatgtgt ccttaaaaat tgcaagccgg tggctatcaa gaagctctat 60
tcccactacc cacaatactt gaaagagttt gagactgagc ttgagacagt tggtagcgtt 120
aagcacagaa atctggtcag tctccaaggc tactctttgt caacgtacgg aaatcttctc 180
ttttatgact acatggaaaa tggcagtcta tgggatcttc tgcatggacc taccaagaag 240
aaaaagcttg attgggatct tegtctaaaa atagcacta 279
App. Ref. : 2290-PCT

```

4

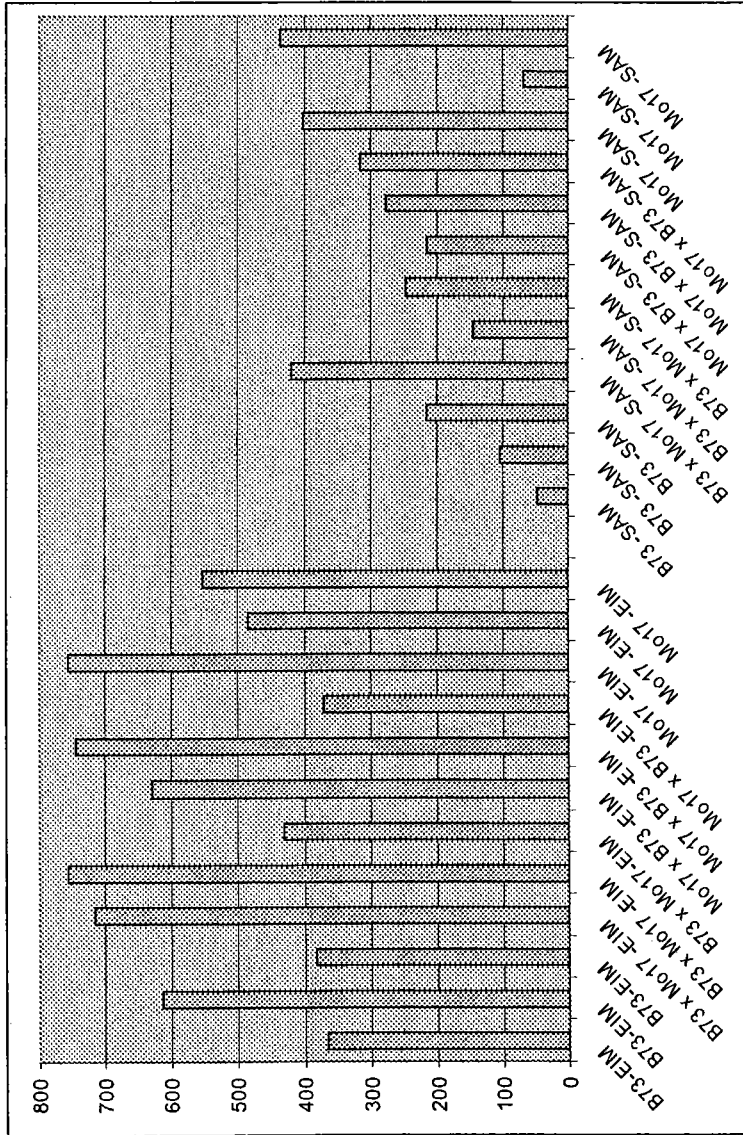


图 1

Ndd

		1		50
SEQ ID NO: 2	(1)	-----MALFRDIVLLGFLFCLSLVAVTVTSEEAGATLLEIKKSKFDVN		
SEQ ID NO: 4	(1)	-----MKEKMQRMVLSLAMVGFVGVASAMNNEGKALMAIKGSFNSLV		
SEQ ID NO: 6	(1)	MPVRSVAMTTTAARALAAVLVTAATAAAVAVADGGALVEIKKSFNRVNG		
SEQ ID NO: 8	(1)	-----ETLLEIKKSFNRDGG		
SEQ ID NO:10	(1)	----MTPAPAAAASYRALVALLV----AVAVADGGSTLLEIKKSFNRVD		
SEQ ID NO:12	(1)	MAAARAPWLWWWVVVVGVAVAEASGGGGGGDGGKALMGVKAGFGNAA		
SEQ ID NO:14	(1)	MPVRSVAMTTTAARALVALLV----AVAVADGGATLVEIKKSFNRVNG		
SEQ ID NO:16	(1)	-----MAARAAAAVVLLIAAVVSVSACGGEGDGGQTLMAVKAGFGNAA		
SEQ ID NO:18	(1)	-----		
SEQ ID NO:20	(1)	-----		
SEQ ID NO:22	(1)	-----		
SEQ ID NO:24	(1)	-----		
SEQ ID NO:26	(1)	-----		
共有序列	(1)		A DG LL IK F	
		51		100
SEQ ID NO: 2	(42)	NVLYDWT-TS-PSSDYCVWRGVSCENVVFNVVALNLSDLNLDGEISPAIG		
SEQ ID NO: 4	(45)	NMLLDWDDVHN--SDLC <sup>W</sup> SRGVFCDNVSYSVVSLNLSLNLGGEISPAIG		
SEQ ID NO: 6	(51)	NVLYDWA----GDDYCSWRGVLCDNVTF <sup>F</sup> AV <sup>A</sup> ALNLSGLNLEGEISPAIG		
SEQ ID NO: 8	(15)	NALYDWSGDG-ASPGYCSWRGVLCDNVTF <sup>F</sup> AV <sup>A</sup> ALNLSGLNLEGEISAAIG		
SEQ ID NO:10	(42)	NVLYDWA----GGDYCSWRGVLCDNVTF <sup>F</sup> AV <sup>A</sup> ALNLSGLNLEGEISPAIG		
SEQ ID NO:12	(51)	NALVDWDGGA----DHC <sup>A</sup> WRGVTCDNAS <sup>F</sup> AV <sup>L</sup> ALNLSNLNLGGEISPAIG		
SEQ ID NO:14	(46)	NVLYDWA----GDDYCSWRGVLCDNVTF <sup>F</sup> AV <sup>A</sup> ALNLSGLNLEGEISPAIG		
SEQ ID NO:16	(45)	NALADWDGGR----DHC <sup>A</sup> WRGVACDAAS <sup>F</sup> AV <sup>V</sup> GLNLSNLNLGGEISPAIG		
SEQ ID NO:18	(1)	-----		
SEQ ID NO:20	(1)	-----		
SEQ ID NO:22	(1)	-----		
SEQ ID NO:24	(1)	-----		
SEQ ID NO:26	(1)	-----		
共有序列	(51)	N L DW	DYCSWRGV CDN TFAV ALNLS LNL GEISPAIG	
		101		150
SEQ ID NO: 2	(90)	DLKSLLSID <sup>L</sup> DKGNRLSGQIPDEIGDCSSLQNLDSLNFNLSGDI <sup>P</sup> FS <sup>I</sup> SKL		
SEQ ID NO: 4	(93)	DLKSLNSID <sup>L</sup> LQGNKLAGQIPDEIGNCASLVYLDLSENLLYGDIPFSISKL		
SEQ ID NO: 6	(96)	SLKSLVSD <sup>L</sup> DKSNGLSGQIPDEIGDCSSLR <sup>T</sup> LD <sup>F</sup> SFNNLDGDI <sup>P</sup> FS <sup>I</sup> SKL		
SEQ ID NO: 8	(64)	SLQRLVSD <sup>L</sup> DKSNGLSGQIPDEIGDCS <sup>L</sup> LETLDLSSNNLEGDIPFSISKL		
SEQ ID NO:10	(87)	RLKGTVSD <sup>L</sup> DKSNGLSGQIPDEIGDCSSLK <sup>T</sup> LDL <sup>S</sup> FN <sup>S</sup> LDGDI <sup>P</sup> FS <sup>I</sup> SKL		
SEQ ID NO:12	(97)	ELK <sup>N</sup> LQFV <sup>D</sup> DKGNKLSGQIPDEIGDCISLKYLDL <sup>S</sup> GNLLYGDIPFSISKL		
SEQ ID NO:14	(91)	SLKSLVSD <sup>L</sup> DKSNGLSGQIPDEIGDCSSLR <sup>T</sup> LD <sup>F</sup> SFNNLDGDI <sup>P</sup> FS <sup>I</sup> SKL		
SEQ ID NO:16	(91)	QLKSLQFV <sup>D</sup> DKLNKLSGQIPDEIGDCVSLKYLDL <sup>S</sup> GNLLYGDIPFSISKL		
SEQ ID NO:18	(1)	-----		
SEQ ID NO:20	(1)	-----		
SEQ ID NO:22	(1)	-----		
SEQ ID NO:24	(1)	-----		
SEQ ID NO:26	(1)	-----		
共有序列	(101)	LK L IDLK N LSGQIPDEIGDC SL LD S N L GDIPFSISKL		
		151		200
SEQ ID NO: 2	(140)	KQLEQLILKNNQLIGPIPTLSQIPNLKILDLAQNKL <sup>T</sup> G <sup>E</sup> I <sup>P</sup> R <sup>L</sup> IY <sup>W</sup> NEV		
SEQ ID NO: 4	(143)	KQLETNLKNNQLTGPIPTLSQIPNLKRLDLAGNHL <sup>T</sup> G <sup>E</sup> I <sup>S</sup> R <sup>L</sup> IY <sup>W</sup> NEV		
SEQ ID NO: 6	(146)	KHLENLILKNNRLGAIPTLSQIPNLKILDLAQNKL <sup>T</sup> G <sup>E</sup> I <sup>P</sup> R <sup>L</sup> IY <sup>W</sup> NEV		
SEQ ID NO: 8	(114)	KHLENLILKNNKLVGVIPSTLSQIPNLKILDLAQNKL <sup>T</sup> G <sup>E</sup> I <sup>P</sup> N <sup>L</sup> IY <sup>W</sup> NEV		
SEQ ID NO:10	(137)	KHLESILKNNQLIGVIPSTLSQIPNLKILDLAQNKL <sup>T</sup> G <sup>E</sup> I <sup>P</sup> R <sup>L</sup> IY <sup>W</sup> NEV		
SEQ ID NO:12	(147)	KQLEELILKNNQLTGPIPTLSQIPNLKTLDLAQNQL <sup>T</sup> G <sup>D</sup> I <sup>P</sup> R <sup>L</sup> IY <sup>W</sup> NEV		
SEQ ID NO:14	(141)	KHLENLILKNNQLGAIPTLSQIPNLKILDLAQNKL <sup>T</sup> G <sup>E</sup> I <sup>P</sup> R <sup>L</sup> IY <sup>W</sup> NEV		
SEQ ID NO:16	(141)	KQLEDLILKNNQLTGPIPTLSQIPNLKTLDLAQNKL <sup>T</sup> G <sup>D</sup> I <sup>P</sup> R <sup>L</sup> IY <sup>W</sup> NEV		
SEQ ID NO:18	(1)	-----		
SEQ ID NO:20	(1)	-----		
SEQ ID NO:22	(1)	-----		
SEQ ID NO:24	(1)	-----		
SEQ ID NO:26	(1)	-----		
共有序列	(151)	K LE LILKNN L G IPSTLSQIPNLK LDLAQN LTGEIPRLIYWNEV		

图 2a



		201		250
SEQ ID NO: 2	(190)	LQYLGLRGNNLVGNISPDICQLTGLWYFDVIRNNSLTGSIPEIIGNCTAFQ		
SEQ ID NO: 4	(193)	LQYLGLRGNNLTGTISSDMCQLTGLWYFDVIRGNNLTGTIPESIGNCTSFQ		
SEQ ID NO: 6	(196)	LQYLGLRGNNHLEGSISPDICQLTGLWYFDVIRNNSLTGAIPDIIGNCTSFQ		
SEQ ID NO: 8	(164)	LQYLGLRSNSLEGSISPDICQLTGLWYFDVIRNNSLTGAIPDIIGNCTSFQ		
SEQ ID NO:10	(187)	LQYLGLRGNNLEGSISPDICQLTGLWYFDVIRNNSLTGPIPEIIGNCTSFQ		
SEQ ID NO:12	(197)	LQYLGLRGNSLTGTISPDICQLTGLWYFDVIRGNNLTGTIPESIGNCTSFQ		
SEQ ID NO:14	(191)	LQYL-----DVIRNNSLTGVIPDIIGNCTSFQ		
SEQ ID NO:16	(191)	LQYLGLRGNSLTGTISPDICQLTGLWYFDVIRGNNLTGTIPESIGNCTSFQ		
SEQ ID NO:18	(1)	-----		
SEQ ID NO:20	(1)	-----		
SEQ ID NO:22	(1)	-----		
SEQ ID NO:24	(1)	-----		
SEQ ID NO:26	(1)	-----		
共有序列	(201)	LQYLGLR N L G LS DMCQLTGLWYFDV K N LTG IPETIGNCTSF		
		251		300
SEQ ID NO: 2	(240)	VLDISYNQLTGEIPIEDIGFLQVATLSLQGNQLSGKIPSVIGLMQALAVLD		
SEQ ID NO: 4	(243)	VLDISYNQITGEIPIYNIGFLQVATLSLQGNRLTGRIPEVIGLMQALAVLD		
SEQ ID NO: 6	(246)	VLDISYNRFTGPIPIENIGFLQVATLSLQGNKFTGPIPSVIGLMQALAVLD		
SEQ ID NO: 8	(214)	VLDISNNHLTGEIPIENIGFLQVATLSLQGNKFTGPIPSVIGLMQALAVLD		
SEQ ID NO:10	(237)	VLDISYNKLSGSIPENIGFLQVATLSLQGNMFTGPIPSVIGLMQALAVLD		
SEQ ID NO:12	(247)	VLDISYNQISGEIPIYNIGFLQVATLSLQGNRLTGKIPDVIGLMQALAVLD		
SEQ ID NO:14	(217)	VLDISYNRFTGPIPIENIGFLQVATLSLQGNKFTGPIPSVIGLMQALAVLD		
SEQ ID NO:16	(241)	VLDISYNQISGEIPIYNIGFLQVATLDLS-----		
SEQ ID NO:18	(1)	-----		
SEQ ID NO:20	(1)	-----		
SEQ ID NO:22	(1)	-----		
SEQ ID NO:24	(1)	-----		
SEQ ID NO:26	(1)	-----		
共有序列	(251)	VLDLSYN TG IPFNIGFLQVATLSLQGN TG IP VIGLMQALAVLD		
		301		350
SEQ ID NO: 2	(290)	LSGNLLSGSIPPILGNLSTTEKLYLHNSKLTGSIPPELGNMSKLHYLELN		
SEQ ID NO: 4	(293)	LSDNELVGPIPPILGNLSTTGKLYLHGNMLTGPIPELGNMRSLSYLQLN		
SEQ ID NO: 6	(296)	LSYNQLSGPIPSILGNLSTTEKLYMOGNRLTGSIPPELGNMSTLHYLELN		
SEQ ID NO: 8	(264)	LSFNELSGPIPSILGNLSTTEKLYMOGNRLTGLIPPELGNMSTLHYLELN		
SEQ ID NO:10	(287)	LSYNQLSGPIPSILGNLSTTEKLYMOGNKLTGPIPELGNMSTLHYLELN		
SEQ ID NO:12	(297)	LSENELVGPPIPSILGNLSTTGKLYLHGNKLTGVIPPELGNMSKLSYLQLN		
SEQ ID NO:14	(267)	LSYNQLSGPIPSILGNLSTTEKLYMOGNKLTGSIPPELGNMSTLHYLELN		
SEQ ID NO:16	(269)	--ENELVGPIPPILGNLSTTGKLYLHGNKLTGHIPPELGNMSKLSYLQLN		
SEQ ID NO:18	(1)	-----		
SEQ ID NO:20	(1)	-----		
SEQ ID NO:22	(1)	-----		
SEQ ID NO:24	(1)	-----		
SEQ ID NO:26	(1)	-----		
共有序列	(301)	LS N L GPIP ILGNLTYT KLYL GNKLTG IPPELGNMS L YL LN		
		351		400
SEQ ID NO: 2	(340)	DNHLTGHIPPELGLTDLFDLNVANNLDLEGPIPDHISSTCNLNSLNVHGN		
SEQ ID NO: 4	(343)	DNKLVGTIPPELGLKLEQLFDLNVANNRLVGPIPSNISSCAALNQFNHGN		
SEQ ID NO: 6	(346)	DNQLTGSIPPELGLRTGLFDLNVANNHLEGPIPDNISSCVNLNSFNHGN		
SEQ ID NO: 8	(314)	DNLLTGFIPFDLGLTELFDLNVANNLIGPIPENISSCANLISFNHGN		
SEQ ID NO:10	(337)	DNQLSGFIPPEFGKLTGLFDLNVANNNFEGPIPDNISSCVNLNSFNHGN		
SEQ ID NO:12	(347)	DNELVGTIPAEGLKLEELFDLNVANNLQGPAPANISSCTALNKFNHGN		
SEQ ID NO:14	(317)	DNQLTGSIPPELGLRTGLFDLNVANNHLEGPIPDNISSCVNLNSFNHGN		
SEQ ID NO:16	(317)	DNELVGTIPAEGLKLTLELFDLNVANNLEGGHPAPANISSCSALNKFNHGN		
SEQ ID NO:18	(1)	-----		
SEQ ID NO:20	(1)	-----		
SEQ ID NO:22	(1)	-----		
SEQ ID NO:24	(1)	-----		
SEQ ID NO:26	(1)	-----		
共有序列	(351)	DN L G IP ELGKL LFDLNLANN L GPIP NISSC LN FN YGN		

图 2b

```

401
450
SEQ ID NO: 2 (390) KFSGTIPRAFOKLESMTYLNLSNNIKGPIVELSRIGNLDTLDLSNNKI
SEQ ID NO: 4 (393) LLSCSIPLAFRNLSMTYLNLSNNFKGKIPVELGHIINLDKLDLSSNFF
SEQ ID NO: 6 (396) KLNCTIPRSLRKLESMTYLNLSNNFISGSIPIELSRINNLDLGLSCNMM
SEQ ID NO: 8 (364) KLNCTIPRSFHKLESMTYLNLSNNHLSGAMPIEVARMRNLDLTLSCNMI
SEQ ID NO:10 (387) RLNGTIPPSLHKLESMTYLNLSNNFLSGSIPIELSRINNLDLTLSCNMI
SEQ ID NO:12 (397) KLNCSIIPAGFOKLESMTYLNLSNNFKGNIPELGHIIINLDLTLDSYNEF
SEQ ID NO:14 (367) KLNCTIPRSLRKLESMTYLNLSNNFISGSIPIELSRINNLDLTLSCNMM
SEQ ID NO:16 (367) RLNGSIIPAGFOELESMTYLNLSNNFKGQIPELGHIVNLDLTLDSYNEF
SEQ ID NO:18 (1) -----
SEQ ID NO:20 (1) -----
SEQ ID NO:22 (1) -----
SEQ ID NO:24 (1) -----
SEQ ID NO:26 (1) -----
共有序列 (401) KL GTIP LESLTYLNLSNN G IP EL I NLDLTLDS N
451
500
SEQ ID NO: 2 (440) NGTIPSSLDGDEHLLKLNLSRNHITGVWPGDFGNLRSIMEIDISNNDISG
SEQ ID NO: 4 (443) SGTIPPLLDGDEHLLLNLNRNHSQHPAEFGNLRSIQOMIDISFNLLSG
SEQ ID NO: 6 (446) TGTIPSSITGNLEHLLRLNLSKNDIVGFPAEFGNLSVMEIDISYNHLGG
SEQ ID NO: 8 (414) TGTIPSAIGKLEHLLRLNLSKNNVAGHPAEFGNLRSEMEIDISYNHLSG
SEQ ID NO:10 (437) TGTIPSTIGSLEHLLRLNLSNNGIVGFPAEIGNLRSEMEIDISNNHLGG
SEQ ID NO:12 (447) SGTVPATLDGDEHLLLELNLNLSKNDIVGFPAEFGNLRSVQVIDISSNNLSG
SEQ ID NO:14 (417) TGTIPSSIGSLEHLLRLNLSKNGIVGFPAEFGNLRSEMEIDISYNHLGG
SEQ ID NO:16 (417) SGTVPPTLDGDEHLLLELNLNLSKNDITGSVPAEFGNLRSVQVIDISSNNLTS
SEQ ID NO:18 (1) -----
SEQ ID NO:20 (1) -----
SEQ ID NO:22 (1) -----
SEQ ID NO:24 (1) -----
SEQ ID NO:26 (1) -----
共有序列 (451) TG IP TIG LEHLL LNLNLSK L G IPAEFGNLRSI IDLS N L G
501
550
SEQ ID NO: 2 (490) YTPPEELNQLQNLII LLRLENNNTTGNM-GSLANCLSLTVLNVSHNNLVGDY
SEQ ID NO: 4 (493) VTPTELQQLQNLNLSLILNNNKTHGKTPDQLTNCFTLVNLNVSFNNLSGIV
SEQ ID NO: 6 (496) LTPQELGMLQNLMLLKLNNNTTGDV-SSLMNCFSLNINLVSYNNLAGAV
SEQ ID NO: 8 (464) LTPQELGMLQNLILLLKLESNNNTTGDV-SSLIYCLSLNINLVSYNNHLYGTV
SEQ ID NO:10 (487) LTPQELGMLQNLMLLNLKNNNTTGDV-SSLMNCFSLNINLVSYNNLAGAV
SEQ ID NO:12 (497) STPEELGQLQNLDSLILNNNNTVGEIPAQLANCFSLNNAEQEFVITQOFA
SEQ ID NO:14 (467) LTPQELEMLQNLMLLNVMS-----YNNLAGAV
SEQ ID NO:16 (467) YTPPEELGQLQNLDSLILNNNNTVGEIPAQLANCFSLIT-----
SEQ ID NO:18 (1) -----
SEQ ID NO:20 (1) -----LSSLNLSYNNKLFQVY
SEQ ID NO:22 (1) -----
SEQ ID NO:24 (1) -----
SEQ ID NO:26 (1) -----
共有序列 (501) IP EL LQNL L L N L G V L C S L L N L G V
551
600
SEQ ID NO: 2 (539) PKNNNFSRFSPDSFTGNPGLCGSWINSPCHDSRRTVRVS-----I
SEQ ID NO: 4 (543) PPMKNFSRFAPASFVGNPYLCGNWVGSICGPLK-----S-
SEQ ID NO: 6 (545) PTDNNEFRFSHDSFTGNPGLCGYWTGSSCRSTGHRDKPP-----I
SEQ ID NO: 8 (513) PTDNNEFRFSPPDSFTGNPGLCGYWTLSASCTQLSNAEQMKRSSS---AKA
SEQ ID NO:10 (536) PTDNNEFRFSPPDSFTGNPGLCGYWTGSSCRSSGHQOKPL-----I
SEQ ID NO:12 (547) WTCPDGKELLEIPNGKHLITSDCNQYINHKCSFLGNPLLVHVCQDSSCGH
SEQ ID NO:14 (493) PADNNEFRFSPPDSFTGNPGLCGYWTGSSCRSTGHHEKPP-----I
SEQ ID NO:16 (505) -----
SEQ ID NO:18 (1) -----
SEQ ID NO:20 (18) PTSNNEFRFPDSFTGNPGLCGNWVNLPCHGARPSEVTV-----L
SEQ ID NO:22 (1) -----
SEQ ID NO:24 (1) -----
SEQ ID NO:26 (1) -----
共有序列 (551) P NFSRF SFLGNP LCG WL

```

图 2c

```

601
SEQ ID NO: 2 (579) SRAAIGGIAIGGLVILMLVLAACRPHNPPPELDGSLDKPVTYS-----
SEQ ID NO: 4 (578) -----RVFSRGALICVILGVITLLCMIFLAVYKSMQOKKIQQGSSKQAEQ
SEQ ID NO: 6 (585) SKAAIIGVAVGGLVILMLVLAACRPHHPPPAEKDATVSKPVSNG-----
SEQ ID NO: 8 (560) SMFAAIGVGAVLLVIMLVILVVICWPHNSPVLDVSVNKPDPNLAGSNNI
SEQ ID NO: 10 (576) SKAAIIGVAVGGLVILMLVLAACRPHSPPVAKDVSVSKPVSNV-----
SEQ ID NO: 12 (597) SHGQRVNTSKTATACITLGFVILLCVLLLAIKTNQPOPLVKGSDKPVQG
SEQ ID NO: 14 (533) SKAAIIGVAVGGLVILMLVLAACRPHRPPPAEKDVTVSKPVSNA-----
SEQ ID NO: 16 (505) -----
SEQ ID NO: 18 (1) -----
SEQ ID NO: 20 (58) SKAAIIGITLGLVILMLVLAACRPHSPSPKPDGSDFKPVSNFS-----
SEQ ID NO: 22 (1) -----
SEQ ID NO: 24 (1) -----
SEQ ID NO: 26 (1) -----
共有序列 (601) S I I GLV LLM LV F K V
651
SEQ ID NO: 2 (623) TPKLVLHNMMAEHVYEDIMRMTENLSEKYIIGYGASSTVYKCVLKNCKP
SEQ ID NO: 4 (623) LTKLVILEMDMAHTEEDIMRMTENLSEKYIIGYGASSTVYKCALKSSRF
SEQ ID NO: 6 (629) PPKLVILHNMMAEHVYEDIMRMTENLSEKYIIGYGASSTVYKCVLKNCKP
SEQ ID NO: 8 (610) HPKLVILHNMMAEHVYEDIMRMTENLSEKYIIGYGASSTVYKCDLKNCKP
SEQ ID NO: 10 (620) PPKLVILHNMMAEHVYEDIMRMTENLSEKYIIGYGASSTVYKCVSKNRKF
SEQ ID NO: 12 (647) PPKLVILEMDMAHTEEDIMRMTENLSEKYIIGYGASSTVYKCELEKSGKA
SEQ ID NO: 14 (577) PPKLVILHNMMAEHVYEDIMRMTENLSEKYIIGYGASSTVYKCVLKNCKP
SEQ ID NO: 16 (505) -----
SEQ ID NO: 18 (1) -----LQMDMAHTEEDIMRMTENLSEKYIIGYGASSTVYKCDLKGKGA
SEQ ID NO: 20 (102) PPKLVILHNMMAEHVYEDIMRMTENLSEKYIIGYGASSTVYKCVLKNCKP
SEQ ID NO: 22 (1) -----
SEQ ID NO: 24 (1) -----
SEQ ID NO: 26 (1) -----STVYKCVLKNCKP
共有序列 (651) PKLVILHM MALH YEDIMRMTENLSEKYIIGYGASSTVYKCLKN KP
701
SEQ ID NO: 2 (673) VAIKKLYSHNPQSMKQFETELEMSSIKHRNLVSLQGYSLPHGSLLFYD
SEQ ID NO: 4 (673) VAIKKLYSHNPQSMKQFETELEMSSIKHRNLVSLQGYSLPHGSLLFYD
SEQ ID NO: 6 (679) VAIKKLYSHYPSLKEFETELETVGSIKHRNLVSLQGYSLSPVGNLLFYD
SEQ ID NO: 8 (660) VAIKKLYSHYPSLKEFETELETVGSIKHRNLVSLQGYSLSPVGNLLFYD
SEQ ID NO: 10 (670) VAIKKLYSHYPSLKEFETELETVGSIKHRNLVSLQGYSLSPVGNLLFYD
SEQ ID NO: 12 (697) VAIKKLYSHYPSLKEFETELETVGSIKHRNLVSLQGYSLSPVGNLLFYD
SEQ ID NO: 14 (627) VAIKKLYSHYPSLKEFETELETVGSIKHRNLVSLQGYSLSPVGNLLFYD
SEQ ID NO: 16 (505) -----
SEQ ID NO: 18 (45) VAIKKLYSHYPSLKEFETELETVGSIKHRNLVSLQGYSLSPVGNLLFYD
SEQ ID NO: 20 (152) VAIKKLYSHYPSLKEFETELETVGSIKHRNLVSLQGYSLSPVGNLLFYD
SEQ ID NO: 22 (1) -----
SEQ ID NO: 24 (1) -----
SEQ ID NO: 26 (14) VAIKKLYSHYPSLKEFETELETVGSIKHRNLVSLQGYSLSPVGNLLFYD
共有序列 (701) VAIKKLYSHYPSLKEFETELETVGSIKHRNLVSLQGYSLSPVGNLLFYD
751
SEQ ID NO: 2 (723) YMENGLWDLHGP-SKKKLDWDLRLKIALGAAQGLAYLHHDCSPRIIH
SEQ ID NO: 4 (723) YMENGLWDLHGP-SKKKLDWDLRLKIALGAAQGLAYLHHDCSPRIIH
SEQ ID NO: 6 (729) YMENGLWDLHGP-SKKKLDWDLRLKIALGAAQGLAYLHHDCSPRIIH
SEQ ID NO: 8 (710) YMENGLWDLHGP-SKKKLDWDLRLKIALGAAQGLAYLHHDCSPRIIH
SEQ ID NO: 10 (720) YMENGLWDLHGP-SKKKLDWDLRLKIALGAAQGLAYLHHDCSPRIIH
SEQ ID NO: 12 (747) YMENGLWDLHGP-SKKKLDWDLRLKIALGAAQGLAYLHHDCSPRIIH
SEQ ID NO: 14 (677) YMENGLWDLHGP-SKKKLDWDLRLKIALGAAQGLAYLHHDCSPRIIH
SEQ ID NO: 16 (505) -----
SEQ ID NO: 18 (95) YMENGLWDLHGP-SKKKLDWDLRLKIALGAAQGLAYLHHDCSPRIIH
SEQ ID NO: 20 (202) YMENGLWDLHGP-SKKKLDWDLRLKIALGAAQGLAYLHHDCSPRIIH
SEQ ID NO: 22 (1) -----
SEQ ID NO: 24 (1) --ENGLWDLHGP-SKKKLDWDLRLKIALGAAQGLAYLHHDCSPRIIH
SEQ ID NO: 26 (64) YMENGLWDLHGP-SKKKLDWDLRLKIALGAAQGLAYLHHDCSPRIIH
共有序列 (751) YMENGLWDLHGP-SKKKLDWDLRLKIALGAAQGLAYLHHDCSPRIIH

```

图 2d

		801		850
SEQ ID NO: 2	(772)	RDVKSSNILLDKDLEARLDFGIAKSLCVSKSHTSTYVMGTIGYIDPEYA		
SEQ ID NO: 4	(772)	RDVKSSNILLDENFEAHLSDFGIAKSLPASKHASTYVLTGTIGYIDPEYA		
SEQ ID NO: 6	(779)	RDVKSKNILLDKDYEHLDFGIAKSLCVSKHTSTYVMGTIGYIDPEYA		
SEQ ID NO: 8	(759)	RDVKSKNILLDKDYEHLADFGIAKSLCVSKHTSTYVMGTIGYIDPEYA		
SEQ ID NO:10	(770)	RDVKSKNILLDKDYEHLDFGIAKSLCVSKHTSTYVMGTIGYIDPEYA		
SEQ ID NO:12	(796)	RDVKSSNILLDENFEAHLSDFGIAKSLPASKHASTYVLTGTIGYIDPEYA		
SEQ ID NO:14	(727)	RDVKSKNILLDKDYEHLDFGIAKSLCVSKHTSTYVMGTIGYIDPEYA		
SEQ ID NO:16	(505)	-----		
SEQ ID NO:18	(144)	RDVKSSNILLDENFEAHLSDFGIAKSLPASKHASTYVLTGTIGYIDPEYA		
SEQ ID NO:20	(251)	RDVKSSNILLDADFEPLHDFGIAKSLCPSKSHSTSTYVMGTIGYIDPEYA		
SEQ ID NO:22	(1)	-----EYA		
SEQ ID NO:24	(48)	RDVKSSNILLDKDYEHLADFGIAKSLCPSKHTSTYVMGTIGYIDPEYA		
SEQ ID NO:26	(94)	-----		
共有序列	(801)	RDVKS NILLD DFEAHLDFGIAKSLC SKTHTSTYVMGTIGYIDPEYA		
		851		900
SEQ ID NO: 2	(822)	RTSRLTEKSDVYSYGIVLLELLTRKKAVDDESNLHHLIMSKTGNNEVMEM		
SEQ ID NO: 4	(822)	RTSRLNEKSDVYSYGIVLLELLTGKKAVDNEANLHQLILSKADDNTVMEA		
SEQ ID NO: 6	(829)	RTSRLNEKSDVYSYGIVLLELLTGKPKVDNECNLHHLILSKTASNEVMET		
SEQ ID NO: 8	(809)	RTSRLNEKSDVYSYGIVLLELLTGKPKVDDECNLHHLILSKAAANVMET		
SEQ ID NO:10	(820)	RTSRLNEKSDVYSYGIVLLELLTGKPKVDNECNLHHLILSKTANNVMET		
SEQ ID NO:12	(846)	RTSRLNEKSDVYSYGIVLLELLTGKKAVDNESNLHQLILSKADDNTVMEA		
SEQ ID NO:14	(777)	RTSRLNEKSDVYRLWHCSAGAADWQEASG-----Q-RILSKTASNEVMET		
SEQ ID NO:16	(505)	-----		
SEQ ID NO:18	(194)	RTSRLNEKSDVYSYGIVLLELLTGKKAVDNESNLHQLILSKADDNTVMEA		
SEQ ID NO:20	(301)	RTSRLTEKSDVYSYGIVLLELLTGKKAVDNESNLHHLILSKAATNAVMET		
SEQ ID NO:22	(4)	RTSHLTEKSDVYSYGIVLLELLTGKKAVDNESNLHHLILSKAATNAVMET		
SEQ ID NO:24	(98)	RTSRLTEKSDVYSYGIVLLELLTGKKAVDNESNLHHLILSKTANDGVMET		
SEQ ID NO:26	(94)	-----		
共有序列	(851)	RTSRLNEKSDVYSYGIVLLELLTGKKAVDNESNLHHLILSK A N VMET		
		901		950
SEQ ID NO: 2	(872)	ADPDIITSTCKDLGVVKKMFQ LALLCTKRQPNDRPTMHQVTRVLGSLFSE		
SEQ ID NO: 4	(872)	VDPDIITVTCDLGHTRKTFQ LALLCTKRNP LERPTMLEVSRVLLSLVPSL		
SEQ ID NO: 6	(879)	VDPDIITGDTCKDLGEVKKMFQ LALLCTKRQPSDRPTMHEVVRVLDCLVNP		
SEQ ID NO: 8	(859)	VQDIITDTCKDLGEVKKMFQ LALLCSKRQPSDRPTMHEVARVLDLVCPA		
SEQ ID NO:10	(870)	VDPDIADTCKDLGEVKKMFQ LALLCTKRQPSDRPTMHEVVRVLDCLVRPD		
SEQ ID NO:12	(896)	VDSIEVSVTCTDMGLVTRKAFQ LALLCTKRHPSDRPTMHEVARVLLSLIPAS		
SEQ ID NO:14	(821)	VDPDIGDTCKDLGEVKKMFQ LALLCTKRQPSDRPTMHEVVRVLDCLVNP		
SEQ ID NO:16	(505)	-----		
SEQ ID NO:18	(244)	VDSIEVSV-----		
SEQ ID NO:20	(351)	VDPDIITATCKDLGAVKKVYQ LALLCTKRQPADRPTMHEVTRVLGSLVLSN		
SEQ ID NO:22	(54)	VDPDIITATCKDLGAVKKVYQ LALLCTKRQPADRPTMHEVTRVLGSLVPS		
SEQ ID NO:24	(148)	VDPDIITTCEDMGA VKKMFQ LALLCTKRQPVDRPTMHEVTRVWEALCHP-		
SEQ ID NO:26	(94)	-----		
共有序列	(901)	VDPDIT TCKDLG VKKVFQ LALLCTKRQPSDRPTMHEV RVL SLV		
		951		1000
SEQ ID NO: 2	(922)	QPP-----AATDTSATLAGSCYVDEYANLKTTP---HSMNCS-S		
SEQ ID NO: 4	(922)	QVAKKLP SLDHSTKKLQENEVRNPDAEASQWVQFREVISKSSIT-----		
SEQ ID NO: 6	(929)	PPPQQQQQKAAHHP-PQ-PSPPAYVDEYVSLRGTG---ALSCANS		
SEQ ID NO: 8	(909)	GPPPKQAQ---AQAQAQASEKPTTAPSYVSEYVGLRGGGGGSAVLSCTNS		
SEQ ID NO:10	(920)	PPPKS-----AQLAMPQRPVAVPSYVNEYVSLRGTG---ALSCANS		
SEQ ID NO:12	(946)	AMTPKTVDSRLLASTTAADMGRGHVTDIGDNSSSDEQWVRFGEVIS		
SEQ ID NO:14	(871)	PPPKP-----S-AHQLPQPS-PAVPSYVNEYVSLRGTG---ALSCANS		
SEQ ID NO:16	(505)	-----		
SEQ ID NO:18	(251)	-----		
SEQ ID NO:20	(401)	TPPKQ-----LAALPPASDPSAXVPCYVDEYGPQDSTLGETAPO---		
SEQ ID NO:22	(104)	IPPKQ-----LADLPPASNPSAKVPCYVDEYANLKTTP---HLVNCP-S		
SEQ ID NO:24	(197)	-----		
SEQ ID NO:26	(94)	-----		
共有序列	(951)	PP		YV EY R L S

图 2e

	1001	1021
SEQ ID NO: 2	(956) MSASDAQFLRFQVISONSE	
SEQ ID NO: 4	(967) -----	
SEQ ID NO: 6	(974) SSTSDAELFLKFGEAISQNMV	
SEQ ID NO: 8	(956) SSASDAELFMKFGEVISRSTE	
SEQ ID NO: 10	(958) SCTSDAELFLKFGEVISQNT	
SEQ ID NO: 12	(996) KHTM-----	
SEQ ID NO: 14	(909) TSTSDAELFLKFGEAISQNME	
SEQ ID NO: 16	(505) -----	
SEQ ID NO: 18	(251) -----	
SEQ ID NO: 20	(441) -----	
SEQ ID NO: 22	(143) MSTSDAQLFLKFGEVISQNSE	
SEQ ID NO: 24	(197) -----	
SEQ ID NO: 26	(94) -----	
共有序列	(1001)	

图 2f

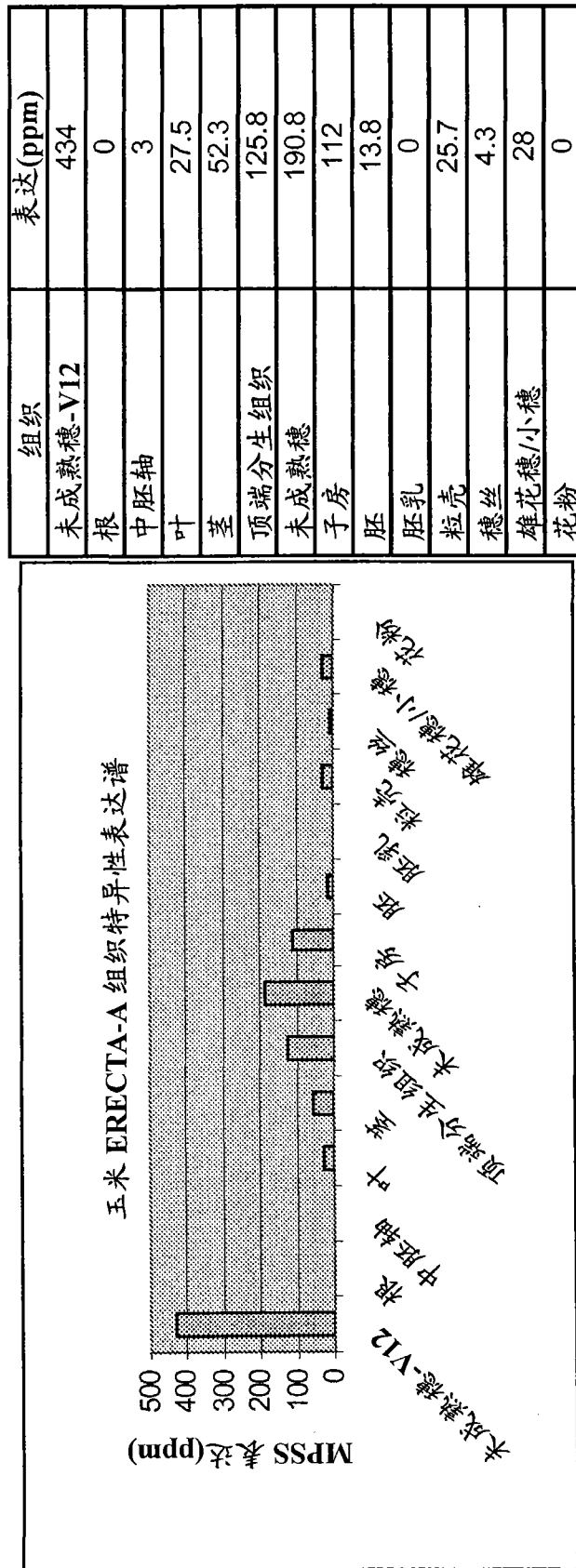


图 3

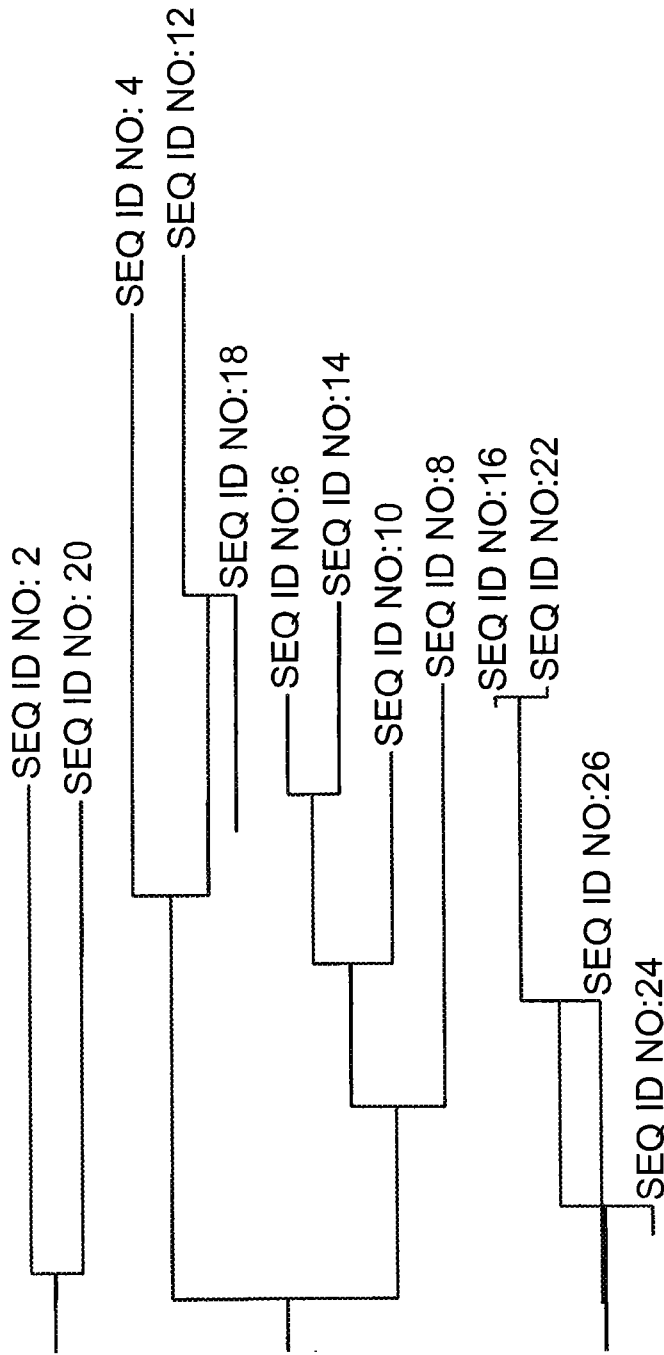


图 4