

[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 200680008289.5

[51] Int. Cl.

C12N 15/29 (2006.01)

A01H 5/10 (2006.01)

G01N 33/53 (2006.01)

C12N 15/82 (2006.01)

C12Q 1/68 (2006.01)

C12N 5/10 (2006.01)

[43] 公开日 2008 年 3 月 12 日

[11] 公开号 CN 101142317A

[51] Int. Cl. (续)

A01H 5/00 (2006.01)

C12N 5/04 (2006.01)

C07K 16/16 (2006.01)

[22] 申请日 2006.1.13

[21] 申请号 200680008289.5

[30] 优先权

[32] 2005. 1. 14 [33] US [31] 60/643,575

[86] 国际申请 PCT/CA2006/000037 2006.1.13

[87] 国际公布 WO2006/074547 英 2006.7.20

[85] 进入国家阶段日期 2007.9.14

[71] 申请人 圭尔夫大学

地址 加拿大安大略

[72] 发明人 S·罗斯坦 Y-m·毕

[74] 专利代理机构 永新专利商标代理有限公司

代理人 林晓红

权利要求书 6 页 说明书 92 页 序列表 4 页
附图 4 页

[54] 发明名称

氮调节的糖感应基因和蛋白质及其调节

[57] 摘要

本发明涉及糖感应所需的氮调节的 GATA 转录因子基因, 及调节这种基因的表达以调节植物中的特征。本发明的 GATA 转录因子参与调节植物中的糖感应并且其表达受氮状态的影响。这种基因或基本相似的基因的表达增加可产生氮利用被改善并且产量增加的植物。

1.一种调节植物细胞中的特征的方法，包括调节植物细胞中 GATA 转录因子基因的表达。

5 2.权利要求 1 的方法，其中 GATA 转录因子基因的表达是通过给予细胞有效量的、可以调节植物细胞中 GATA 转录因子基因表达水平的物质而调节的。

3.权利要求 1 或 2 的方法，其中所述特征是农学性状。

4.权利要求 3 的方法，其中所述特征是这样的特征，其受氮、碳
10 和/或硫代谢、脂质的生物合成、营养素的感受、营养适应、电子转运和/或膜相关能量守恒的影响。

5.权利要求 3 的方法，其中所述特征选自如下一或多种特征：氮利用、产量、细胞生长、繁殖、光合作用、氮同化作用、疾病抗性、分化、信号转导、基因调节、非生物应激耐受和营养组成。

15 6.权利要求 5 的方法，其中所述特征是氮利用。

7.权利要求 1—6 任一项的方法，其中所述植物细胞是双子叶植物、裸子植物或单子叶植物。

8.权利要求 7 的方法，其中所述单子叶植物选自由玉米、小麦、大麦、稷燕麦、裸麦、稷、高粱、黑小麦、黑麦、单粒小麦、斯佩耳
20 特小麦、双粒小麦、画眉草、买罗高粱、亚麻、gramma 草、摩擦草和类蜀黍组成的组。

9.权利要求 8 的方法，其中所述双子叶植物选自大豆、烟草或棉花。

10.权利要求 2—9 任一项的方法，其中所述物质增强植物细胞中
25 GATA 转录因子基因的表达水平。

11.权利要求 10 的方法，其中所述被调节的特征是选自如下的一或多个特征的增加或改善：氮利用、产量、细胞生长、繁殖、光合作用

用、氮同化、疾病抗性、分化、信号转导、基因调节、非生物应激耐受和营养组成。

12.权利要求 10 或 11 的方法，其中所述增强植物细胞中 GATA 转录因子基因表达水平的物质包含编码 GATA 转录因子的核酸分子。

5 13.权利要求 12 的方法，其中所述核酸分子包含 SEQ ID NO:1 所示 AT5g56860 基因的序列或其功能片段。

14.权利要求 12 的方法，其中所述核酸分子包含在中等严格条件下与 SEQ ID NO:1 所示 AT5g56860 基因或其功能片段杂交的序列。

15.权利要求 12 的方法，其中所述核酸分子包含衍生自 SEQ ID
10 NO:1 所示 At5g56860 基因的核苷酸序列的核酸序列并且所述核酸分子具有包含特异于在植物中表达的密码子的核苷酸序列。

16.权利要求 12 的方法，其中所述核酸分子包含 SEQ ID NO:3。

17.权利要求 1—12 任一项的方法，其中所述能调节植物细胞中 GATA 转录因子基因的表达水平的物质包含：

- 15 (a) SEQ ID NO:1 所示核苷酸序列或其片段或结构域；
 (b) 编码 SEQ ID NO:2 所示多肽的核苷酸序列、其片段或结构域；
 (c) 与(a)或(b)基本相似的核苷酸序列；
 (d) 能与(a)、(b)或(c)杂交的核苷酸序列；
 (e) 与(a)、(b)、(c)或(d)互补的核苷酸序列；或者
20 (f) 是(a)、(b)、(c)或(d)的反向互补序列的核苷酸序列。

18.权利要求 1—12 任一项的方法，其中所述可调节植物细胞中 GATA 转录因子基因表达水平的物质包含：

- (a) SEQ ID NO:2 所列多肽序列，或其功能片段、结构域、重复
 序列或嵌合序列；
25 (b)与(a)基本相似的多肽序列；
 (c) 多肽序列，其由与 SEQ ID NO:1 的核苷酸序列相同或基本
 相似的核苷酸序列或其功能片段或结构域或其互补序列编码；

或

(d) 多肽序列，其由能在中等严格条件下与 SEQ ID NO:1 的核苷酸序列或其互补序列杂交的核苷酸序列编码。

5 19.权利要求 12—18 任一项的方法，其中所述核酸序列在植物的特定位置或组织中表达。

20.权利要求 19 的方法，其中所述位置或组织选自种子、表皮、根、维管组织、分生组织、形成层、皮层、髓层、叶和花组成的组的一或多种位置或组织。

21.权利要求 20 的方法，其中所述位置或组织是种子。

10 22.权利要求12—21任一项的方法，其中增强植物细胞中GATA转录因子基因表达水平的物质包含调节植物细胞中的特征的表达盒，其包含与编码GATA转录因子的分离的核酸可操作地连接的启动子序列。

23.一种分离的核酸，其包含：

- 15 (a) SEQ ID NO:1 所示核苷酸序列，或其片段或结构域；
(b) 与(a)基本相似的核苷酸序列；
(c) 能与(a)杂交的核苷酸序列；
(d) 与(a)、(b)或(c)互补的核苷酸序列；或者
(e) 是(a)、(b)或(c)反向互补序列的核苷酸序列。

20 24. 编码包含如下多肽序列的多肽的分离的核酸分子：

- (a) SEQ ID NO:2 所列多肽序列，或其片段、结构域、重复序列或嵌合序列；
(b)与(a)基本相似的多肽序列；
(c) 多肽序列，其由与 SEQ ID NO:2 的核苷酸序列相同或基本相似的核苷酸序列或其功能片段或结构域或其互补序列编码；
25 (d) 多肽序列，其由能在中等严格条件下与 SEQ ID NO:1 的核苷酸序列或其互补序列杂交的核苷酸序列编码；或者

(e) (a)、(b)、(c)或(d)的功能片段。

25.权利要求 23 的分离的核酸，其具有 SEQ ID NO:1 所示 At5g56860 基因的序列。

26. 分离的核酸，其在中等严格条件下与权利要求 25 的核酸杂交。

27. 分离的核酸，其衍生自权利要求 25 的核苷酸序列，并具有这样的核苷酸序列，所述核苷酸序列中包含特异于在植物中的表达的密码子。

28.一种表达载体，其包含与权利要求 23—27 任一项的分离的核酸分子可操作地连接的启动子序列。

29.一种载体，包含权利要求 23—27 任一项的核酸分子。

30. 植物细胞，其用权利要求 23—27 任一项的核酸转化。

31. 产生转基因植物的方法，包括：

(1) 提供权利要求 23—27 任一项的分离的核酸；
(2) 将所述核酸导入所述植物中，其中所述核酸在植物中表达。

32.权利要求 31 的方法，其中所述植物表现出如下一或多个特征的
的增加或改善：氮利用、产量、细胞生长、繁殖、光合作用、氮同化、
疾病抗性、分化、信号转导、基因调节、非生物应激耐受和营养组成。

33.权利要求 31 或 32 的方法，其中使用选自如下的方法将所述核酸导入植物中：微粒轰击，农杆菌介导的转化，或 whiskers-介导的转化。

34.使用权利要求 31—33 任一项的方法产生的植物。

35.权利要求 34 的植物的种子。

36.权利要求 34 的植物的植物细胞。

37.包含至少 10 个碱基的核苷酸序列的核酸分子的应用，所述序列与 SEQ ID NO:1 或其功能片段中任一种的区域相同、互补或基本相似，其中所述应用选自

- (i) 用作染色体标记,以鉴别天然或人工染色体上相应或互补多核苷酸的位置;
- (ii) 用作 RFLP 分析的标记;
- (iii) 用作数量性状相关育种的标记;
- 5 (iv) 用作标记辅助的育种的标记;
- (v) 在双杂交系统中用作诱饵序列,以鉴别编码与诱饵序列编码的多肽相互作用的多肽的序列;
- (vi) 用作诊断标记以对个体或个体的群体进行基因分型或鉴别; 及
- 10 (vii) 用于遗传分析以鉴别基因或外显子的边界。

38. 一种抗分离的多肽的抗体, 包含:

- (a) SEQ ID NO:2 所示多肽序列, 或其片段、结构域、重复序列或嵌合序列;
- (b) 与(a)基本相似的多肽序列;
- 15 (c) 多肽序列, 其由与 SEQ ID NO:2 的核苷酸序列相同或基本相似的核苷酸序列或其功能片段或结构域或其互补序列编码;
- (d) 多肽序列, 其由能在中等严格条件下与 SEQ ID NO:1 的核苷酸序列或其互补序列杂交的核苷酸序列编码; 或者
- (e) 所述(a)、(b)、(c)、(d)的功能片段。

- 20 39. 权利要求 38 的抗体, 其中所述多肽包含 SEQ ID NO:2 所示序列或者其具有保守氨基酸修饰的变体。

40. 一种免疫测定试剂盒, 其包含权利要求 38 或 39 的抗体及其使用说明书。

- 25 41. 权利要求 23—27 任一项的核酸分子在调节植物细胞中的特征中应用。

42. 权利要求 41 的应用, 其中所述特征选自如下一或多种特征: 氮利用、产量、细胞生长、繁殖、光合作用、氮同化作用、疾病抗性、

分化、信号转导、基因调节、非生物应激耐受及营养组成。

43. 权利要求 41 的应用，其中所述特征是氮利用。

44. 权利要求 41—43 任一项的应用，其中所述植物细胞是双子叶植物、裸子植物或单子叶植物细胞。

5 45. 权利要求 44 的应用，其中所述单子叶植物选自由玉米、小麦、大麦、稷燕麦、裸麦、稷、高粱、黑小麦、黑麦、单粒小麦、斯佩耳特小麦、双粒小麦、画眉草、买罗高粱、亚麻、gramma 草、摩擦草和类蜀黍组成的组。

10 46. 权利要求 44 的应用，其中所述双子叶植物选自大豆、烟草或棉花。

47. 权利要求 41—46 的应用，其中所述被调节的特征是选自如下的一或多个特征得以增加或改善：氮利用、产量、细胞生长、繁殖、光合作用、氮同化、疾病抗性、分化、信号转导、基因调节、非生物应激耐受和营养组成。

15

氮调节的糖感应基因和蛋白质及其调节

发明领域

- 5 本发明涉及通过调节植物细胞中的GATA转录因子的表达而调节植物中的农学性状的方法。本发明特别涉及改善植物中的氮利用的方法。本发明还涉及分离自拟南芥 (*Arabidopsis thaliana*) 的核酸分子, 这些核酸分子包含编码糖感应型并最终可以调节氮摄取和整体碳代谢的蛋白质的核苷酸序列。

10

发明背景

- 作物的农学特征的改良从农业的开始就一直在进行。目前大多数适于作物生产的土地都在被利用。随着人口持续增加, 将需要改良的作物品种以提供足够的食物和饲料 (Trewavas (2001) Plant Physiol.
- 15 125: 174-179)。为了避免灾难性饥荒和营养不良, 未来的作物栽培品种需要具有改良的产量和等价的农场投入。这些栽培品种需要更有效地经受不利的条件如干旱、土壤盐度或疾病, 这在边缘土地被用于耕作时尤为重要。最终, 我们需要具有改变的营养组成的栽培品种以增进人类和动物营养和使得能更有效地进行食品和饲料加工。对于所有
- 20 有这些性状, 鉴别控制感兴趣的性状的表型表达的基因对于加速通过常规或转基因手段开发优异的作物种质是关键的。

- 有一些高效方法可用于帮助鉴别在农学重要性状表达中起关键作用的基因。这些包括遗传学、基因组学、生物信息学和功能基因组学。遗传学是遗传基质的科学研究。通过鉴别改变感兴趣的途径或应
- 25 答的突变, 经典(或正向)遗传学可以帮助鉴别参与这些途径或应答的基因。例如, 具有增强的疾病易感性的突变体可以鉴别植物信号转导途径的重要成分, 从而从病原体识别到疾病抗性。遗传学还是通过育种改良种质的关键成分。通过遗传杂交的分子学和表型分析, 控制

感兴趣性状的基因座可以被定位和在后续代次中被追踪。关于登记作物（crop accessions）之间的表型差异的基因的知识可以用来开发标记，由此大大增加种质改良过程的效率，并且为发现另外的优异等位基因开启了途径。

5 基因组学是生物体基因组的系统水平的研究，包括基因和相应的基因产物-RNA 和蛋白质。在第一种水平，基因组方法提供了来自各种植物物种的序列信息的大数据集，包括全长和部分 cDNA 序列，以及模式植物物种拟南芥（*Arabidopsis thaliana*）的完整基因组序列。最佳，作物植物基因组的第一份草图序列，水稻（*Oryza sativa*）的草图序列也可以利用了。完整基因组序列的可利用性使得开发其它分子互补序列的系统水平研究的工具变得可能，所述工具如用于确定特定条件下在生物体中表达的基因的互补序列的阵列和芯片。这种数据可以用作某些基因在不同植物表型的表达中起关键作用的可能性的最初指示。

10

15 生物信息学方法与第一水平的基因组数据集直接连接，允许进行加工以通过注释或其它手段发现感兴趣序列。使用例如相似性检索、序列对比和系统发生学分析，生物信息学可以经常鉴别感兴趣基因产物的同系物。非常相似的同系物（例如在蛋白质全长上有>~90%氨基酸相同性）非常可能是直向同源物（ortholog），即在不同生物体中具有相同功能。

20

功能基因组学可以被定义为将功能赋予基因及其产物。功能基因组学从遗传学、基因组学和生物信息学中衍生出鉴别在特定途径或感兴趣应答中起重要作用的基因的途径。表达分析，例如，使用高密度 DNA 微阵列（通常衍生自基因组规模的生物体测序）以在一个实验中监测几千个基因的 mRNA 的表达。实验处理可以包括引发感兴趣的应答、如在被病原体感染的植物中的疾病抗性应答的那些处理。为了给出使用微阵列的另外的例子，可以在不同组织中或者在于感兴

25

趣应答中被影响的突变体中随发育时程监测 mRNA 表达水平。蛋白质组学也可以有助于赋予功能,这可以通过在一个实验中测定几百个蛋白质的表达和翻译后修饰而进行。

蛋白质组学方法在许多情况中类似于用来在微阵列实验中监测 mRNA 表达的方法。蛋白质-蛋白质相互作用也有助于将蛋白质归于给定的途径或应答,这是通过鉴别与该途径或应答的已知成分相互作用的蛋白质而进行。对于功能基因组学,蛋白质-蛋白质相互作用经常使用大规模酵母双杂交测定法进行研究。另一种赋予基因功能的方法是在异源宿主例如大肠杆菌中表达相应的蛋白质,随后进行纯化和酶学测定。

对感兴趣基因控制给定性状的能力的证实可以例如通过在感兴趣的植物物种中进行实验测试而进行。对于感兴趣的基因而言是转基因型的植物的产生和分析可以被用于植物功能基因组学,其具有一些优势。基因可以经常被过表达和表达不足(“敲除”),由此增加观察到将该基因与感兴趣的途径或应答联系起来的表型的机会。转基因功能基因组学的两个方面有助于使得通过这一方法进行的功能分配(functional assignment)具有高置信水平。首先,在活植物中进行表型观察。其次,观察到的表型范围可以被检查并与导入的转基因的观察到的表达水平相关。转基因功能基因组学在改良的栽培品种开发中特别有价值。只有在感兴趣的途径或应答中有功能并且另外还能赋予所需基于性状的表型的基因被作为作物改良工作的候选基因。在一些情况中,被开发用于功能基因组学研究的转基因品系可以直接用于产品开发的初始阶段。

植物功能基因组学的另一种方法包括首先鉴别在感兴趣的特异基因中有突变的植物品系,然后对这些基因敲除对所研究的性状的后果进行表型评价。这种方法揭示特异性状的表达所必需的基因。

通过功能基因组学鉴别的基因可以被直接用于如上所述的通过

转基因手段进行的种质改良的工作, 或者被用于开发标记以鉴别在作图和育种群中感兴趣的等位基因。关于这些基因的知识也使得能够通过分子学方法构建自然界中不存在的优异等位基因。

过去80年行间作物产量的快速增加大致相等地归因于改良的遗传学和改良的农学实践。特别是, 在玉米作物中, 高产杂种组合和大量氮肥的使用在理想条件下使得产量大于440bu/英亩。但是, 大量氮肥的使用有不利的副作用, 主要是农夫对于这项投入的成本增加以及环境成本增加, 因为硝酸盐污染在许多农业区是一个主要问题, 导致新鲜水和海洋环境等级显著下降。通过了解基因组对氮利用的作用而更有效利用氮的作物遗传学的发展对于降低生产者投入成本以及环境负荷是高度有利的。这对于像玉米这样的利用高水平氮肥生长的作物来说是特别重要的。

氮利用效率可以以几种方式定义, 但是最简单的是产量/N供应。这一方法有两个阶段: 首先, 被摄取、储存和同化成氨基酸和其它重要的含氮化合物的可利用的氮的量; 第二, 被隔离在种子中从而造成最终产量的氮的比例。已经在各种农业重要作物上进行了各种田间研究以研究这一问题 (Lawlor DW et al 2001 in Lea PJ, Morot-Gaudry JF, eds. Plant Nitrogen. Berlin: Springer-Verlag 343-367; Lafitte HR and Edmeades GO 1994 Field Crops Res 39, 15-25; Lawlor DW 2002 J Exp Bot. 53, 773-87; Moll RH et al 1982 Agron J 74, 562-564)。这些实验已证实存在与氮利用效率相关的遗传成分, 但是未能证实哪些基因对于这一过程是重要的。另外, 玉米育种者的目标通常不是在限制氮肥的条件下保持产量。这些类型的氮利用田间实验很难, 原因较多, 包括在测试田地中或在以任何处理方案处理的田地位点之间缺乏可利用氮的均匀性以及其它环境因素的作用使得实验很难被解释等。

因此, 尽管存在关于这一性状的遗传差异的实验证据, 但是很难从这些实验得出是什么导致这一差异的结论。开发在田间条件下在

作物中研究这一性状的方法应该是可行的并当然是重要的。但是，通过使用模式系统如鼠耳芥(*Arabidopsis*)可使得鉴别、了解和操纵重要的性状出现明显的进展。至少，这些实验会给出关于潜在靶基因的重要线索以便在重要的田间作物中进行评价。另外，也有相当多的遗传和基因组资源可以用来研究水稻，这一物种也可作为与鼠耳芥相比更类似于玉米的物种而用于一些提议的实验中。

硝酸盐是田间可利用氮的主要形式并且有许多文献报道了参与硝酸盐摄取和还原的基因(Forde BG 2000 *Biochimica et Biophysica Acta* 1465, 219-235; Howitt SM and Udvardi MK 2000 *Biochimica et Biophysica Acta* 1465, 152-170; Stitt M et al 2002 *J Exp Bot.* 53, 959-70)以及参与氮代谢的其它方面的基因(Lea PJ, Morot-Gaudry JF, eds. 2001 *Plant Nitrogen*. Berlin: Springer-Verlag; Morot-Gaudry JF 2001 *Nitrogen assimilation by plants* Science Publishers Inc. NH, US)。另外，还清楚碳代谢物的利用度对于有效使用田间硝酸盐是关键，并且有许多良好的实验证据表明碳和氮代谢之间的联系 (Coruzzi GM and Zhou L 2001 *Curr Opin Plant Biol.* 4, 247-53)。另外，一些实验提示GS和GOGAT参与N从老化 (senescing) 器官向库 (sink) 器官的再动员 (Brouquisse R et al 2001 in Lea PJ, Morot-Gaudry JF, eds. *Plant Nitrogen*. Berlin: Springer-Verlag 275-293; Yamaya T et al 2002 *J Exp Bot.* 53, 917-925)。但是，这些基因的调节的大多数方面仍是不清楚的并且仍未意识到这一调节是如何影响氮利用效率。

植物可以感应碳和氮代谢物的水平并因此调节生长和发育。感受 (perception) 机制是复杂的调节网络，其控制基因表达以适应营养依赖性细胞活性的恒定变化。拥有糖感应机制使得植物能在C—骨架丰富时关闭光合作用。N感应机制使得植物能在还原的或有机N水平高时关闭硝酸盐摄取和还原(Coruzzi, G.M. & Zhou, L. (2001) *Curr Opin Plant Biol.* 4, 247-53)。

植物中存在多种糖信号转导途径。葡萄糖在光合植物中的许多重要过程如在光合作用和在碳和氮代谢中是关键的调节物 (Rolland, F., Moore, B. & Sheen, J. (2002) *Plant Cell* S185-S205)。己糖激酶 (HXK)是葡萄糖代谢的重要控制点。它们不仅催化葡萄糖的磷酸化, 5 而且作为葡萄糖感应物 (sensor) 使营养、光和激素信号转导网络相互关联以应答变化的环境来控制生长和发育(Jang, J., Leon, P, Zhou, L. & Sheen, J. (1997) *Plant Cell* 9, 5-19; Dai, N., Schaffer, A., Petreikov, M., Shahak, Y., Giller, Y., Ratner, K., Levine, A. & Granot, D. (1999) *Plant Cell* 11, 1253-1266; Moore, B., Zhou, L., Rolland, F., Hall, Q., 10 Cheng, W., Liu, Y., Hwang, I., Jones, T. & Sheen, J. (2003) *Science* 300, 332-336)。在其它生物体中, 已经表明己糖转运分子也作为糖感应物。

多种N信号和感应途径也在植物中存在。植物具有感应氮肥主要形式硝酸盐作为无机N状态的信号的机制, 以及感应衍生自硝酸盐的代谢物作为还原的或有机N状态的信号的机制。硝酸盐还原酶(NR) 15 和亚硝酸盐还原酶(NiR)是在硝酸盐还原过程中的最初的两个酶, 它们的表达可以在硝酸盐存在下被刺激并且被其它生理因子包括一些含氮化合物、蔗糖、光和激素调节 (Forde, B.G. (2000) *Biochimica et Biophysica Acta* 1465, 219-235; Howitt, S.M. & Udvardi, M.K. (2000) *Biochimica et Biophysica Acta* 1465, 152-170; Stitt, M., Müller, M., 20 Matt, M., Gibon, Y., Carillo, P., Morcuende, R., Scheible, W. & Krapp, A. (2002) *J Exp Bot.* 53, 959-970; Lea, P.J. & Morot-Gaudry, J.F. eds. 2001 *Plant Nitrogen*. Berlin: Springer-Verlag; Morot-Gaudry JF 2001 *Nitrogen assimilation by plants* Science Publishers Inc. NH, US)。

清楚的是碳和氮代谢是紧密连接并紧密调节的(Coruzzi, G. & 25 Bush, D.R. (2001) *Plant Physiol* 125, 61-64)。碳代谢物的利用度对于有效的硝酸盐利用是关键的, 而氮状态对光合作用非常敏感。尽管有关参与碳和氮代谢的结构基因的知识增加, 但是还未鉴定参与C/N基因

表达的转录调节的反式作用因子。

GATA转录因子是一组在真核生物中广泛分布的转录调节物。GATA DNA结合结构域正常情况下识别共有序列WGATAR (W = T或A; R = G或A) (Lowry, J. & Atchley, W. (2000) *J Mol Evol* 50, 103-115)。GATA基序已经在许多光应答基因的调节区中被鉴别 (Arguello-Astorga, G. & Herrera-Estrella, L. (1998) *Annu Rev Plant Physiol Plant Mol Biol* 49, 525-555), 其中包括许多参与光合作用或与光合作用相关的基因如RBCS、CAB (叶绿素A/B结合蛋白)和GAP (甘油醛-3-磷酸脱氢酶) (Terzaghi, W.B. & Cashmore, A.R. (1995) *Annu Rev Plant Physiol Plant Mol Biol* 46, 445-474; Koch, K.E. (1996) *Carbohydrate-modulated gene expression in plants. Annu Rev Plant Physiol Plant Mol Biol* 47, 509-540; Jeong, M.J. & Shih, M.C. (2003) *Biochem Biophys Res Commun* 300, 555-562) 以及参与硝酸盐同化的基因如硝酸盐还原酶、亚硝酸盐还原酶和Gln合成酶(Jarai, G., Truong, H., Daniel-Vedele, F. & Marzluf, G. (1992) *Curr Genet* 21, 37-41; Rastogi, R., Bate, N., Sivasankar, S & Rothstein, S. (1997) *Plant Mol Biol.* 34, 465-76; Oliveira, I.C. & Coruzzi, G.M. (1999) *Plant Physiol* 121, 301-309)。一些已知的全局调节N代谢中的基因的反式作用调节蛋白是GATA转录因子基因。在酵母中, 4种全局氮调节因子 GLN3, NIL1, NIL2 和DAL80是含有单个GATA锌指的DNA结合蛋白, 它们识别共有基序GATA (Hofman-Bang, J. (1999) *Mol Biotech* 12, 35-73)。在真菌中, 粗糙脉孢霉(*Neurospora crassa*) NIT2 (Tao Y and Marzluf GA 1999 *Curr Genet* 36, 153-158) 和构巢曲霉 (*Aspergillus nidulans*) AREA (Caddick MX Arst HN Jr Taylor LH Johnson RI Brownlee AG 1986 Cloning of the regulatory gene areA mediating nitrogen metabolite repression in *Aspergillus nidulans*. *EMBO J* 5, 1087-1090)是GATA转录因子基因。

在植物中，GATA因子的体内功能仍然是非常不了解的，鼠耳芥基因具有30个GATA成员(Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z.,
5 Ghandehari, D., Sherman, B.K. & Yu, G. (2000) Science 290, 2105-2110; Reyes, J.C., Muro-Pastor, M.I. & Florencio, F.J. (2004) Plant Physiol. 134, 1718-1732)。

发明概述

10 在了解GATA转录因子基因家族的生物学功能的尝试中，研究了GATA转录因子基因以了解其在调节碳和氮代谢中的作用。本发明人确定At5g56860基因的表达由氮状态影响，并且其表达调节控制碳代谢的基因和参与其它生物学过程的基因的表达。At5g56860基因丧失功能的突变植物产生降低的叶绿素水平并且这些植物对于外源葡
15 萄糖是高敏的。相反，功能获得性(gain-of-function)转基因植物对于外源葡萄糖不太敏感。

糖是光合植物的许多重要过程如光合作用以及碳和氮代谢的关键调节物。这一调节作用是通过调节基因表达以激活或阻遏所涉及的基因而实现。糖控制基因表达的机制还不太清楚。本文公开的 GATA
20 转录因子参与调节糖感应，该因子自身的表达由 N 状态的变化影响。这一基因增加的表达可以产生产量增加的植物，特别是因为对糖信号转导途径的操作可以导致增加的光合作用和增加的氮同化并改变种子、块茎、根和其它储存器官中的源-库(source-sink)关系。

因此，本发明涉及一种调节植物或植物细胞中的特征的方法，
25 包括调节植物或植物细胞中 GATA 转录因子基因的表达。在本发明的一个实施方案中，GATA 转录因子基因的表达通过给予细胞有效量的下述物质而调节，所述物质能调节植物细胞中的 GATA 转录因子基因

表达水平。在本发明的另一个实施方案中，所述物质增强了植物细胞中 GATA 转录因子基因的表达水平。

植物中被调节的特征可以是任何感兴趣的农学性状。在本发明的一个实施方案中，所述特征是可被氮、碳和/或硫代谢、脂质的生物合成、营养素的感受、营养适应、电子转运和/或膜相关能量守恒影响的任何特征。在本发明的另一个实施方案中，所述特征选自氮利用、产量(yield)、细胞生长、繁殖、光合作用、氮同化(nitrogen assimilation)、疾病抗性、分化、信号转导、基因调节、非生物应激耐受(abiotic stress tolerance)和营养组成(nutritional composition)中的一或多种。在本发明的另一个实施方案中，被调节的特征是氮利用、产量、细胞生长、繁殖、光合作用、氮同化、疾病抗性、分化、信号转导、基因调节、非生物应激耐受和营养组成中的一或多种特征的增加或改善。

在一特定的实施方案中，本发明涉及改善植物或植物细胞中的氮利用的方法，包括增强植物或植物细胞中的 GATA 转录因子基因的表达。改善植物中的氮利用将使得施用于植物的氮肥的量降低并且伴随农夫成本和环境成本的降低，因为硝酸盐污染在许多农业区是一个主要问题，显著导致新鲜水和海洋环境等级下降。

所述植物或植物细胞可以来自任何希望调节特征的植物。在本发明的一个实施方案中，所述植物细胞是双子叶植物、裸子植物或单子叶植物。在一个实施方案中，双子叶植物选自大豆、烟草或棉花。在本发明的另一个实施方案中，单子叶植物选自玉米(maize)、小麦(wheat)、大麦(barley)、稷燕麦(oat)、裸麦(rye)、稷(millet)、高粱(sorghum)、黑小麦(triticale)、黑麦(secale)、单粒小麦(einkorn)、斯佩耳特小麦(spelt)、双粒小麦(emmer)、画眉草(teff)、买罗高粱(milo)、亚麻(flax)、gramma 草(grass)，摩擦草(*Tripsacum sp.*)和类蜀黍(teosinte)。

在本发明的一个实施方案中，增强植物细胞中的 GATA 转录因子基因表达水平的物质包括编码 GATA 转录因子的核酸分子。

在本发明的一个实施方案中，可以调节植物细胞中的 GATA 转录因子基因的表达水平的物质包括：

- 5 (a) SEQ ID NO:1 的核苷酸序列或其片段或结构域;
- (b) 编码 SEQ ID NO:2 的多肽的核苷酸序列或其片段或结构域;
- (c) 与(a)或(b)基本相似的核苷酸序列;
- (d) 能与(a), (b)或(c)杂交的核苷酸序列;
- 10 (e) 与(a), (b), (c)或(d)互补的核苷酸序列; 或
- (f) 是(a), (b), (c)或(d)的反向互补序列的核苷酸序列。

在本发明的另一个实施方案中，所述核酸分子包括 SEQ ID NO:1 的 AT5g56860 基因的序列或其功能片段。在本发明的另一个实施方案中，所述核酸分子包括在中等严格杂交条件下与 SEQ ID NO:1 的 AT5g56860 基因杂交的序列或其功能片段。在本发明的另一个实施方案中，所述核酸分子衍生自 SEQ ID NO:1 的 At5g56860 基因的核苷酸序列并具有包含在植物中特异性表达的密码子。在本发明的另一个实施方案中，所述核酸分子是 AT5g56860 基因的水稻直向同源物，其包含 SEQ ID NO:3 的序列。

20 在本发明的另一个实施方案中，可以调节植物细胞中的 GATA 转录因子基因的表达水平的物质包括：

- (a) SEQ ID NO:2 所列多肽序列，或其功能片段、结构域、重复序列或嵌合序列;
- (b) 与(a)基本相似的多肽序列;
- 25 (c) 多肽序列，其由与 SEQ ID NO:1 的核苷酸序列相同或基本相似的核苷酸序列、或其功能片段或结构域、或其互补序列编码; 或

(d)多肽序列,其由能在中等严格条件下与 SEQ ID NO:1 的核苷酸序列或其互补序列杂交的核苷酸序列编码。

在本发明的一个实施方案中,当所述物质是核酸序列时,所述核酸序列在植物的特定位置或组织中表达。所述位置或组织例如是但不限于表皮、根、维管组织、分生组织、形成层、皮层、髓层、叶和/或花。在另一个实施方案中,所述位置或组织是种子。

本发明的实施方案还涉及使用改组的(shuffled)核酸分子来调节植物细胞中的特征,所述改组的核酸分子含有多个核苷酸序列片段,其中至少一个片段编码 GATA 转录因子,并且其中多个序列片段中的至少两个序列片段从 5'到 3'的顺序不是所述多个片段天然存在于核酸中的顺序。在一个具体的实施方案中,含有多个核苷酸序列片段的改组的核酸分子中的全部片段来自单个基因。在一个更具体的实施方案中,所述多个片段源自至少两个不同基因。在一个更具体的实施方案中,所述改组的核酸可操作地与启动子序列连接。另一个更具体的实施方案是使用嵌合多核苷酸以调节植物细胞中的特征,所述嵌合多核苷酸包括可操作地与改组的核酸连接的启动子序列。在一个更具体的实施方案中,所述改组的核酸包含在宿主细胞内。在本发明的另一个具体的实施方案中,编码 GATA 转录因子的片段由以下序列组成或包含以下序列:

- 20 (a) SEQ ID NO:1 的核苷酸序列或其片段或结构域;
- (b) 编码 SEQ ID NO:2 的多肽的核苷酸序列或其片段或结构域;
- (c) 与(a)或(b)基本相似的核苷酸序列;
- (d) 能与(a), (b)或(c)杂交的核苷酸序列;
- (e) 与(a), (b), (c)或(d)互补的核苷酸序列; 或
- 25 (f) 是(a), (b), (c)或(d)的反向互补序列的核苷酸序列。

本发明的实施方案还涉及使用表达盒来调节植物细胞中的特征,所述表达盒包括与编码GATA转录因子的分离的核酸可操作地连

接的启动子序列。在本发明的实施方案中，所述编码GATA转录因子的分离的核酸由以下序列组成或包含以下序列：

- (a) SEQ ID NO:1的核苷酸序列或其片段或结构域；
- (b) 编码SEQ ID NO:2的多肽的核苷酸序列或其片段或结构域；
- 5 (c) 与(a)或(b)基本相似的核苷酸序列；
- (d) 能与(a), (b)或(c)杂交的核苷酸序列；
- (e) 与(a), (b), (c)或(d)互补的核苷酸序列；或
- (f) 是(a), (b), (c)或(d)的反向互补序列的核苷酸序列。

本发明还包含使用重组载体调节植物细胞中的特征，所述重组载体包含表达盒，所述表达盒包括与编码GATA转录因子的分离的核酸可操作地连接的启动子序列。在本发明的实施方案中，所述编码GATA转录因子的分离的核酸由以下序列组成或包含以下序列：

- (a) SEQ ID NO:1的核苷酸序列或其片段或结构域；
- (b) 编码SEQ ID NO:2的多肽的核苷酸序列或其片段或结构域；
- 15 (c) 与(a)或(b)基本相似的核苷酸序列；
- (d) 能与(a), (b)或(c)杂交的核苷酸序列；
- (e) 与(a), (b), (c)或(d)互补的核苷酸序列；或
- (f) 是(a), (b), (c)或(d)的反向互补序列的核苷酸序列。

还包含的是含有本发明的表达盒的植物细胞的用途以及含有这些植物细胞的植物的用途。

在一个实施方案中，表达盒在整个植物中表达。在另一个实施方案中，表达盒在植物的特定位置或组织中表达。在一个具体的实施方案中，所述位置或组织可以是例如表皮、根、维管组织、分生组织、形成层、皮层、髓层、叶和花。在另一个具体的实施方案中，所述位置或组织是种子。

本发明的实施方案还提供了来自植物的种子和分离的产物在调节植物细胞中的特征中的应用，所述种子和分离的产物含有表达盒，

所述表达盒包括与编码本发明的 GATA 转录因子基因的分离的核酸可操作地连接的启动子序列。

在一个具体的实施方案中，表达载体包括一或多个元件，例如但不限于启动子-增强子序列、选择标记序列、复制起点、表位标记编码序列或亲和纯化标记编码序列。在一个更具体的实施方案中，启动子-增强子序列可以是例如 CaMV 35S 启动子、CaMV 19S 启动子、烟草 PR-1a 启动子、遍在蛋白或菜豆蛋白启动子。在另一个实施方案中，启动子是在植物中可操作的，更具体地，所述启动子是组成型或诱导型启动子。在另一个具体的实施方案中，选择标记序列编码抗生素抗性基因。在另一个具体的实施方案中，表位标记序列编码 V5，肽 Phe-His-His-Thr-Thr 血凝素或谷胱甘肽-S-转移酶。在另一个具体的实施方案中，亲和纯化标记序列编码多氨酸序列或多肽。在一个更具体的实施方案中，所述聚氨基酸序列是聚组氨酸。在一个更具体的实施方案中，所述多肽是壳多糖结合结构域或谷胱甘肽-S-转移酶。在一个更具体的实施方案中，所述亲和纯化标记序列包含蛋白内含子 (intein) 编码序列。

在一个具体的实施方案中，表达载体是真核表达载体或原核表达载体。在一个更具体的实施方案中，真核表达载体包括组织特异性启动子。更具体地，表达载体是在植物中可操作的。

本发明的实施方案还涉及通过一种方法修饰的植物，所述方法包括向植物中导入有效实现所述修饰的量的核酸，其中所述核酸是在植物中可表达的。所述修饰可以是基因的过表达、表达不足、反义调节、有义抑制、诱导抑制、诱导阻遏或诱导调节。在本发明的一个实施方案中，所述修饰涉及感兴趣的性状例如氮利用的增加或改善。

本发明的实施方案提供了分离自拟南芥 (*Arabidopsis thaliana*) 的核苷酸和氨基酸序列。具体地，本发明涉及糖感应所需的氮调节的 GATA 转录因子基因。

本发明的实施方案涉及分离的核酸，其包含或由如下核苷酸序列组成，所述核苷酸序列包含：

- (a) SEQ ID NO:1 所列的核苷酸序列，或其片段或结构域；
- (b) 与(a)基因基本相似性的核苷酸序列；
- 5 (c) 能与(a)杂交的核苷酸序列；
- (d) 与(a), (b)或(c)互补的核苷酸序列；或
- (e) 是(a), (b)或(c)的反向互补序列的核苷酸序列。

在一个具体的实施方案中，所述基本相似性是与 SEQ ID NO:1 所列的核苷酸序列或其片段或结构域有至少约 65%相同性、特别是约
10 80%相同性、特别是 90%、更具体是至少约 95%序列相同性。

在一个实施方案中，与 SEQ ID NO:1 的核苷酸序列或其片段或结构域基本相似的序列来自植物。在一个具体的实施方案中，所述植物是双子叶植物。在一个更具体的实施方案中，所述双子叶植物选自大豆、烟草或棉花。在另一个具体的实施方案中，所述植物是裸子植物。
15 在另一个具体的实施方案中，所述植物是单子叶植物。在一个更具体的实施方案中，所述单子叶植物是谷类。在一个更具体的实施方案中，所述谷类可以是例如玉米、小麦、大麦、稷燕麦、裸麦(rye)、稷、高粱、黑小麦、黑麦、单粒小麦、斯佩耳特小麦、双粒小麦、画眉草、买罗高粱、亚麻、gramma 草、摩擦草和类蜀黍。

20 在一个实施方案中，核酸在植物的特定位置或组织中病毒。所述位置或组织例如是但不限于表皮、根、维管组织、分生组织、形成层、皮层、髓层、叶和花。在另一个实施方案中，所述位置或组织是种子。在另一个实施方案中，核酸编码参与如下功能的多肽，例如但不限于碳、氮和/或硫代谢、氮利用、氮同化、光合作用、信号转导、
25 细胞生长、繁殖、疾病抗性、非生物应激耐受、营养组成、基因调节和/或分化。

在一个具体的实施方案中，所述分离的核酸包含或由能与 SEQ

ID NO:1 所列的核苷酸序列或其片段或结构域杂交的核苷酸序列组成。在一个具体的实施方案中，杂交使得序列在中等或高严格条件下形成双螺旋。本发明的实施方案还包含与 SEQ ID NO:1 的核苷酸序列互补的核苷酸序列或其片段或结构域。本发明的实施方案进一步包含这样的核苷酸序列或其片段或结构域，所述核苷酸序列互补于与 SEQ ID NO:1 的核苷酸序列基本相似或能与其杂交的核苷酸序列。

在一个具体的实施方案中，基本相似的核苷酸序列是 SEQ ID NO:1 的核苷酸序列或其片段或结构域的等位基因变体。在另一个实施方案中，基本相似的序列是天然存在的变体。在另一个实施方案中，基本相似的序列是 SEQ ID NO:1 的核苷酸序列或其片段或结构域的多态性变体。

在一个具体的实施方案中，分离的核酸含有多个具有 SEQ ID NO:1 的核苷酸序列或其外显子或结构域的区域。

在一个具体的实施方案中，分离的核酸含有多肽编码序列。在一个更具体的实施方案中，所述多肽编码序列含有 20 个碱基对的核苷酸部分，其序列与 SEQ ID NO:1 的核苷酸序列的连续 20 个碱基对的核苷酸部分相同。在一个更具体的实施方案中，所述多肽含有 SEQ ID NO:2 的多肽序列或其片段。在一个更具体的实施方案中，所述多肽是植物多肽。在一个更具体的实施方案中，所述植物是双子叶植物。在一个更具体的实施方案中，所述植物是裸子植物。在一个更具体的实施方案中，所述植物是单子叶植物。在一个更具体的实施方案中，所述单子叶植物是谷类。在一个更具体的实施方案中，谷类可以是例如玉米、小麦、大麦、稷燕麦、裸麦、稷、高粱、黑小麦、黑麦、单粒小麦、斯佩耳特小麦、双粒小麦、画眉草、买罗高粱、亚麻、gramma 草、摩擦草和类蜀黍。

在一个实施方案中，所述多肽在整个植物中表达。在一个更具体的实施方案中，所述多肽在植物的特定位置或组织中表达。在一个

更具体的实施方案中，所述位置或组织可以是例如表皮、根、维管组织、分生组织、形成层、皮层、髓层、叶和花。在最具体的实施方案中，所述位置或组织是种子。

在一个具体的实施方案中，所述分离的核酸编码用于产生抗体的多肽，所述抗体具有抗由 SEQ ID NO:2 或其片段或结构域编码的多肽的免疫反应性。

在一个具体的实施方案中，所述基本相似的序列含有至少一个核苷酸的缺失或插入。在一个更具体的实施方案中，所述缺失或插入少于约 30 个核苷酸。在最具体的实施方案中，所述缺失或插入少于约 5 个核苷酸。

在一个具体的实施方案中，基本相似的分离的核酸的序列包含或由至少一个密码子中的取代至少一个密码子中的取代组成。在一个具体的实施方案中，取代是保守性的。

本发明的实施方案还涉及分离的核酸分子，其包含或由编码如下多肽的核苷酸序列、其互补序列或其反向互补序列组成，所述多肽包括：

(a) SEQ ID NO:2 的多肽序列，或其片段、结构域、重复序列或嵌合序列；

(b) 与(a)基本相似的多肽序列；

(c) 多肽序列，由与 SEQ ID NO:1 的核苷酸序列相同或基本相似的核苷酸序列、或其功能片段或结构域、或其互补序列编码；

(d) 由一种能在中等严格条件下与 SEQ ID NO:1 的核苷酸序列或其互补序列杂交的核苷酸序列编码的多肽序列；或

(e) (a), (b), (c)或(d)的功能片段。

在另一个具体的实施方案中，基本相似的多肽是 SEQ ID NO:2 的多肽序列或其片段、结构域、重复序列或嵌合序列的等位基因变体。

在另一个具体的实施方案中,所述分离的核酸包括多个来自由如下核苷酸序列编码的多肽序列的区域,所述核苷酸序列与 SEQ ID NO:1 的核苷酸序列或其外显子或结构域或其互补序列相同或基本相似。

在另一个具体的实施方案中,所述多肽是 SEQ ID NO:2 的多肽序列。在另一个具体的实施方案中,所述多肽是功能片段或结构域。在另一个具体的实施方案中,所述多肽是嵌合体,其中所述嵌合体可以包括功能蛋白质结构域,包括结构域、重复序列、翻译后修饰位点或其它特征。在一个更具体的实施方案中,所述多肽是植物多肽。在一个更具体的实施方案中,所述植物是双子叶植物。在一个更具体的实施方案中,所述植物是裸子植物。在一个更具体的实施方案中,所述植物是单子叶植物。在一个更具体的实施方案中,所述单子叶植物是谷类。在一个更具体的实施方案中,谷类可以是例如玉米、小麦、大麦、燕麦、裸麦、稷、高粱、黑小麦、黑麦、单粒小麦、斯佩耳特小麦、双粒小麦、画眉草、买罗高粱、亚麻、gramma 草、磨擦草和类蜀黍。

在一个具体的实施方案中,所述多肽在植物的特定位置或组织中表达。在一个更具体的实施方案中,所述位置或组织可以是例如表皮、根、维管组织、分生组织、形成层、皮层、髓层、叶和花。在另一个具体的实施方案中,所述位置或组织是种子。

在一个具体的实施方案中,由以下核苷酸序列编码的多肽包括至少一个核苷酸的缺失或插入,所述核苷酸序列与 SEQ ID NO:1 的核苷酸序列或其片段或结构域或其互补序列基本相似。在一个更具体的实施方案中,所述缺失或插入少于约 30 个核苷酸。在最具体的实施方案中,所述缺失或插入少于约 5 个核苷酸。

在一个具体的实施方案中,由以下核苷酸序列编码的多肽包括至少一个密码子中的取代,所述核苷酸序列与 SEQ ID NO:1 的核苷酸序列或其片段或结构域或其互补序列基本相似。在一个更具体的实

施方案中，所述取代是保守性的。

在一个具体的实施方案中，与 SEQ ID NO:2 的多肽序列或其片段、结构域、重复序列或嵌合序列基本相似的多肽序列包括至少一个氨基酸的缺失或插入。

- 5 在一个具体的实施方案中，与 SEQ ID NO:2 的多肽序列或其片段、结构域、重复序列或嵌合序列基本相似的多肽序列包括至少一个氨基酸的取代。

- 10 本发明的实施方案还涉及含有多个核苷酸序列片段的改组的核酸，其中至少一个所述片段相应于 SEQ ID NO:1 的核苷酸序列的一个区域，且其中所述多个序列片段中的至少两个片段的 5'-3' 的顺序不是所述多个片段在核酸中天然存在的顺序。在一个更具体的实施方案中，含有多个核苷酸序列片段的改组的核酸中的全部片段均来自单个基因。在一个更具体的实施方案中，所述多个片段源自至少两个不同的基因。在一个更具体的实施方案中，改组的核酸可操作地与启动子序列连接。另一个更具体的实施方案是包括与改组的核酸可操作地连接的启动子序列的嵌合多核苷酸。在一个更具体的实施方案中，改组的核酸含在宿主细胞中。

本发明的实施方案还涉及表达盒，其包括与含有以下核苷酸序列的分离的核酸可操作地连接的启动子序列，所述核苷酸序列包括：

- 20 a) SEQ ID NO:1 的核苷酸序列或其片段或结构域；
 (b) 编码 SEQ ID NO:2 的多肽的核苷酸序列或其片段或结构域；
 (c) 与(a)或(b)基本相似的核苷酸序列；
 (d) 能与(a), (b)或(c) 杂交的核苷酸序列；
 (e) 与(a), (b), (c)或(d)互补的核苷酸序列；或
25 (f) 是(a), (b), (c) 或(d)的反向互补序列的核苷酸序列。

本发明还包含包含根据本发明的实施方案的表达盒的重组载体。还包含含有本发明的表达盒的植物细胞和含有这些植物细胞的植物。

在一个具体的实施方案中，所述植物是双子叶植物。在一个更具体的实施方案中，所述双子叶植物选自大豆、烟草或棉花。在另一个具体的实施方案中，所述植物是裸子植物。在另一个具体的实施方案中，所述植物是单子叶植物。在一个更具体的实施方案中，所述单子叶植物是谷类。在一个更具体的实施方案中，谷类可以例如是玉米、小麦、大麦、稷燕麦、裸麦、稷、高粱、黑小麦、黑麦、单粒小麦、斯佩耳特小麦、双粒小麦、画眉草、买罗高粱、亚麻、gramma 草、摩擦草和类蜀黍。

在一个实施方案中，表达盒在整个植物中表达。在另一个实施方案中，表达盒在植物的特定位置或组织中表达。在一个具体的实施方案中，所述位置或组织可以是例如表皮、根、维管组织、分生组织、形成层、皮层、髓层、叶和花。在另一个具体的实施方案中，所述位置或组织是种子。

在一个实施方案中，表达盒参与如下功能，例如但不限于碳、氮和/或硫代谢、氮利用、氮同化、光合作用、信号转导、细胞生长、繁殖、疾病抗性、非生物应激耐受、营养组成、基因调节和/或分化。在一个更具体的实施方案中，所述嵌合多肽参与如下功能，如氮利用、非生物应激耐受、增加的产量、疾病抗性和/或营养组成。

在一个实施方案中，植物含有对植物的表型或可测量特征的修饰，所述修饰可归因于表达盒中所含的至少一个基因的表达。在一个具体的实施方案中，修饰可以是例如碳、氮和/或硫代谢、氮利用、氮同化、光合作用、信号转导、细胞生长、繁殖、疾病抗性、非生物应激耐受、营养组成、基因调节和/或分化。

本发明的实施方案还提供了来自植物的种子和分离的产物，所述植物含有表达盒，所述表达盒包括可操作地与一种分离的核酸连接的启动子序列，所述分离的核酸含有如下核苷酸序列，所述核苷酸序列包括：

- (a) SEQ ID NO:1 的核苷酸序列或其片段或结构域;
- (b) 编码 SEQ ID NO:2 的多肽的核苷酸序列或其片段或结构域;
- (c) 与(a)或(b)基本相似的核苷酸序列;
- (d) 能与(a), (b)或(c)杂交的核苷酸序列;
- 5 (e) 与(a), (b), (c)或(d)互补的核苷酸序列; 或
- (f) 是本发明的(a), (b), (c) 或(d)的反向互补序列的核苷酸序列。

在一个具体的实施方案中, 所述分离的产物包括酶、营养蛋白、结构蛋白、氨基酸、脂质、脂肪酸、多糖、糖、醇、生物碱、类胡萝卜素、类丙烷(propanoid)、类固醇、色素、维生素和植物激素。

- 10 本发明的实施方案还涉及由一种分离的核酸表达而产生的分离的产物, 所述分离的核酸含有如下核苷酸序列, 所述核苷酸序列包括:

- (a) SEQ ID NO:1 的核苷酸序列或其片段或结构域;
- (b) 编码 SEQ ID NO:2 的多肽的核苷酸序列或其片段或结构域;
- (c) 与(a)或(b)基本相似的核苷酸序列;
- 15 (d) 能与(a), (b)或(c)杂交的核苷酸序列;
- (e) 与(a), (b), (c)或(d)互补的核苷酸序列; 或
- (f) 是本发明的(a), (b), (c) 或(d)的反向互补序列的核苷酸序列。

- 在一个具体的实施方案中, 所述产物是在植物中产生的。在另一个具体的实施方案中, 所述产物是在细胞培养物中产生的。在另一个具体的实施方案中, 所述产物是在无细胞系统中产生的。在另一个具
- 20 体的实施方案中, 所述产物包括酶、营养蛋白、结构蛋白、氨基酸、脂质、脂肪酸、多糖、糖、醇、生物碱、类胡萝卜素、类丙烷、类固醇、色素、维生素和植物激素。

- 在一个具体的实施方案中, 所述产物是含有 SEQ ID NO:2 的氨基酸序列的多肽。在一个更具体的实施方案中, 所述蛋白质是转录因子。
- 25

本发明的实施方案进一步涉及分离的多核苷酸, 其包括至少 10

个碱基的核苷酸序列，所述序列与 SEQ ID NO:1 序列的一个区域相同、互补或基本相似，其中所述多核苷酸适应各种用途。

在一个具体的实施方案中，所述多核苷酸用作染色体标记。在另一个具体的实施方案中，所述多核苷酸用作 RFLP 分析的标记。在另一个具体的实施方案中，所述多核苷酸用作数量性状连锁育种的标记。在另一个具体的实施方案中，所述多核苷酸用作标记辅助育种的标记。在另一个具体的实施方案中，所述多核苷酸用作双杂交系统中的诱饵序列以鉴别与由诱饵序列编码的多肽相互作用的序列编码多肽。在另一个具体的实施方案中，所述多核苷酸用作诊断指示剂以基因分型或鉴别个体或个体群。在另一个具体的实施方案中，所述多核苷酸用于遗传分析以鉴别基因或外显子边界。

本发明的实施方案还涉及包含或由如下核酸分子组成的表达载体，所述核酸分子包括：

- (a) 编码 SEQ ID NO:2 的多肽的核酸；
- (b) SEQ ID NO:1 的片段、一或多个结构域或特征区；或
- (c) 与异源序列组合的 SEQ ID NO:1 的完整核酸序列或其片段。

在一个具体的实施方案中，表达载体包括一或多个元件例如但不限于启动子-增强子序列、选择标记序列、复制起点、表位标记编码序列或亲和纯化标记编码序列。在一个更具体的实施方案中，启动子-增强子序列可以是例如，CaMV 35S 启动子、CaMV 19S 启动子、烟草 PR-1a 启动子、遍在蛋白或菜豆蛋白启动子。在另一个实施方案中，启动子是在植物中可操作的，更具体地，是组成型或诱导型启动子。在另一个具体的实施方案中，选择标记序列编码抗生素抗性基因。在另一个具体的实施方案中，表位标记序列编码 V5，肽 Phe-His-His-Thr-Thr，血凝素或谷胱甘肽-S-转移酶。在另一个具体的实施方案中，亲和纯化标记编码聚氨基酸序列或多肽。在一个更具体

的实施方案中，聚氨基酸序列是聚组氨酸。在一个更具体的实施方案中，所述多肽是壳多糖结合结构域或谷胱甘肽-S-转移酶。在一个更具体的实施方案中，亲和纯化标记序列包含蛋白内含子编码序列。

在一个具体的实施方案中，表达载体是真核表达载体或原核表达载体。在一个更具体的实施方案中，真核表达载体包括组织特异性启动子。更具体地，表达载体是在植物中可操作的。

本发明的实施方案还涉及包含或由一种核酸构建体组成的细胞，所述核酸构建体包含表达载体和如下核酸，所述核酸包括编码 SEQ ID NO:2 的多肽的核酸或与异源序列组合的 SEQ ID NO:1 的核苷酸序列或其片段。

在一个具体的实施方案中，所述细胞是细菌细胞、真菌细胞、植物细胞或动物细胞。在一个具体的实施方案中，所述细胞是植物细胞。在一个更具体的实施方案中，所述多肽在植物的特定位置或组织中表达。在最具体的实施方案中，所述位置或组织可以是例如表皮、根、维管组织、分生组织、形成层、皮层、髓层、叶和花。在另一个最具体的实施方案中，所述位置或组织是种子。在一个具体的实施方案中，所述多肽参与如下功能，如碳、氮和/或硫代谢、氮利用、氮同化、光合作用、信号转导、细胞生长、繁殖、疾病抗性、非生物应激耐受、营养组成、基因调节和/或分化。

本发明的实施方案还涉及由本发明的分离的核酸分子编码的多肽，包括含有由一种分离的核酸编码的多肽序列的多肽，所述分离的核酸含有如下核苷酸序列，包括：

- (a) SEQ ID NO:1 的核苷酸序列或其外显子或结构域;
- (b) 与(a)基本相似的核苷酸序列;
- (c) 能与(a)杂交的核苷酸序列;
- (d) 与(a), (b), (c)或(d)互补的核苷酸序列; 或
- (e) 是 (a), (b), 或(c)的反向互补序列的核苷酸序列;

(f) 或其功能片段。

含有由一种分离的核酸编码的多肽序列的多肽，所述分离的核酸含有编码如下多肽的核苷酸序列、其互补序列或其反向互补序列，所述多肽包括如下多肽序列：

5 (a) SEQ ID NO:2 的多肽序列或其结构域、重复序列或嵌合序列；

 (b) 与 (a) 基本相似的多肽序列；

 (c) 多肽序列，由与 SEQ ID NO:1 的核苷酸序列相同或基本相似的核苷酸序列、或其外显子或结构域、或其互补序列编
10 码；

 (d) 多肽序列，由在中等严格条件下能与 SEQ ID NO:1 的核苷酸序列或其互补序列杂交的核苷酸序列编码；或者

 (e) (a), (b), (c) 或(d) 的功能片段；

 (f) 或其功能片段。

15 本发明的实施方案涉及含有由一种分离的核酸编码的多肽序列的多肽，所述分离的核酸包括含有多个核苷酸序列片段的改组的核酸，其中至少一个片段相应于 SEQ ID NO:1 的核苷酸序列的一个区域，并且其中所述多个序列片段的至少两个片段的 5'-3' 顺序与所述多个片段在核酸或其功能片段中天然存在的顺序不同。

20 本发明的实施方案涉及含有由一种分离的多核苷酸编码的多肽序列的多肽，所述分离的多核苷酸含有至少 10 个碱基的核苷酸序列，所述序列与 SEQ ID NO:1 的序列或其功能片段的一个区域相同、互补或基本相似，其中所述多核苷酸适应于如下用途：

 (a) 用作染色体标记以鉴别相应的或互补的多核苷酸在天
25 然或人工染色体上的位置；

 (b) 用作 RFLP 分析的标记；

 (c) 用作定量性状连锁育种 (quantitative trait linked

breeding)的标记;

(d)用作标记辅助育种的标记;

(e)用作双杂交系统中的诱饵序列以鉴别编码与由诱饵序列编码的多肽相互作用的多肽的序列;

- 5 (f)用作诊断指示剂以基因分型或鉴别个体或个体群; 或
 (g)用于遗传分析以鉴别基因或外显子边界。

本发明的实施方案还涉及分离的多肽, 其含有如下多肽序列:

(a) SEQ ID NO:2 的多肽序列或其外显子或结构域;

(b)与(a)基本相似的多肽序列;

- 10 (c)多肽序列, 由与 SEQ ID NO:1 的核苷酸序列相同或基本相似的核苷酸序列、或其外显子或结构域、或其互补序列编码;

(d)多肽序列, 由在中等严格条件下能与 SEQ ID NO:1 的核苷酸序列或其互补序列杂交的核苷酸序列编码; 或者

- 15 (e) (a), (b), (c)或(d)的功能片段。

在一个具体的实施方案中, 基本相似是至少 65% 相同性。在一个更具体的实施方案中, 基本相似是至少约 80% 相同性。在最具体的实施方案中, 基本相似是至少约 95% 相同性。在一个具体的实施方案中, 基本相似比与序列列表中列出的最接近的同源序列的相同性百分比高至少 3%。

- 20 在一个具体的实施方案中, 基本相似的序列来自植物。在一个更具体的实施方案中, 所述植物是双子叶植物。在一个更具体的实施方案中, 所述植物是裸子植物。在一个更具体的实施方案中, 所述植物是单子叶植物。在一个更具体的实施方案中, 所述单子叶植物是谷类。在一个更具体的实施方案中, 谷类可以是例如玉米、小麦、大麦、
25 稷燕麦、裸麦、稷、高粱、黑小麦、黑麦、单粒小麦、斯佩耳特小麦、双粒小麦、画眉草、买罗高粱、亚麻、gramma 草, 摩擦草和类蜀黍。

在一个具体的实施方案中，所述多肽在植物的特定位置或组织中表达。在一个更具体的实施方案中，所述位置或组织可以是例如表皮、根、维管组织、分生组织、形成层、皮层、髓层、叶和花。在另一个具体的实施方案中，所述位置或组织是种子。在一个具体的实施方案中，所述多肽参与如下功能，例如 碳、氮和/或硫代谢、氮利用、氮同化、光合作用、信号转导、细胞生长、繁殖、疾病抗性、非生物应激耐受、营养组成、基因调节和/或分化。

在一个具体的实施方案中，下述多肽序列的杂交使得序列在中等或高严格条件下形成双螺旋：由与 SEQ ID NO:1 的核苷酸序列相同或基本相似的核苷酸序列或其外显子或结构域或其互补序列编码的多肽序列，或者由在中等严格条件下能与 SEQ ID NO:1 的核苷酸序列或其互补序列杂交的核苷酸序列编码的多肽序列。

在一个具体的实施方案中，与 SEQ ID NO:2 或其外显子或结构域基本相似的多肽是 SEQ ID NO:2 的多肽序列的等位基因变体。在另一个具体的实施方案中，与 SEQ ID NO:2 或其外显子或结构域基本相似的多肽是 SEQ ID NO:2 的多肽序列的天然存在的变体。在另一个具体的实施方案中，与 SEQ ID NO:2 或其外显子或结构域基本相似的多肽是 SEQ ID NO:2 的多肽序列的多态性变体。

在一个替代的具体的实施方案中，基本相似的序列含有至少一个氨基酸的缺失或插入。在一个更具体的实施方案中，所述缺失或插入少于约 10 个氨基酸。在最具体的实施方案中，所述缺失或插入少于约 3 个氨基酸。

在一个具体的实施方案中，基本相似的序列编码至少一个氨基酸中的取代。

本发明还涉及产生包含修饰的植物的方法，包括如下步骤：(1) 提供一种核酸，其是含有如下核苷酸序列的分离的核酸，所述核苷酸序列包括：

- (a) SEQ ID NO:1 的核苷酸序列或其外显子或结构域;
- (b) 与(a)基本相似的核苷酸序列;
- (c) 能与(a)杂交的核苷酸序列;
- (d) 与(a), (b)或(c)互补的核苷酸序列; 或
- 5 (e) 是(a), (b)或(c)的反向互补序列的核苷酸序列;

和(2) 将有效实现所述修饰的量的所述核酸导入植物, 其中所述核酸在植物中是可表达的。在一个实施方案中, 所述修饰包括植物中改变的特征, 其中所述特征相应于导入植物中的核酸。在其它具体的实施方案中, 所述特征相应于碳、氮和/或硫代谢、氮利用、氮同化、光合作用、信号转导、细胞生长、繁殖、疾病抗性、非生物应激耐受、
10 营养组成、基因调节和/或分化。

在另一个实施方案中, 所述修饰包括植物产物的增加的或降低的表达或积累。具体地, 所述产物是植物的天然产物。同样具体地, 所述产物是新的或改变的植物产物。具体地, 所述产物包含 GATA 转录
15 因子。

本发明还包括生产重组蛋白的方法, 包括如下步骤:

- (a) 在合适的生长条件下培养包含核酸构建体的重组细胞, 所述构建体包含表达载体和如下核酸, 所述核酸包括: 编码 SEQ ID NO:2 的蛋白质的核酸或 SEQ ID NO:1 的核酸序列或其片段; 以及
20 (b) 从所述重组细胞分离由其表达的重组蛋白。

本发明的实施方案提供了生产重组蛋白的方法, 其中表达载体包括一或多个元件, 所述元件包括启动子—增强子序列、选择标记序列、复制起点、表位标记编码序列和亲和纯化标记编码序列。在一个具体的实施方案中, 核酸构建体包括表位标记编码序列并且分离步骤
25 包括使用特异于所述表位标记的抗体。在另一个具体的实施方案中, 核酸构建体含有聚氨基酸编码序列并且分离步骤包括使用包含聚氨基酸结合物质的树脂, 特别是当聚氨基酸是聚组氨酸时聚氨基酸结合

树脂是镍琼脂糖树脂。在另一个具体的实施方案中，核酸构建体含有多肽编码序列并且分离步骤包括使用含有多肽结合物质的树脂，特别是当所述多肽是壳多糖结合结构域时树脂含有壳多糖—琼脂糖。

本发明的实施方案还涉及通过一种方法修饰的植物，所述方法包括向植物中导入有效实现所述修饰的量的可在植物中表达的核酸。所述修饰可以是例如碳、氮和/或硫代谢、氮利用、氮同化、光合作用、信号转导、细胞生长、繁殖、疾病抗性、非生物应激耐受、营养组成、基因调节和/或分化。在一个实施方案中，修饰的植物具有增加的或降低的对除草剂、应激或病原体的抗性。在另一个实施方案中，修饰的植物具有增强的或降低的对光、水、氮或微量元素的需
10 要。在另一个实施方案中，修饰的植物中必需氨基酸作为植物的蛋白质级分的比例而言是富集的。所述蛋白质级分可以是例如总种子蛋白、可溶性蛋白、不溶性蛋白、可水提取蛋白和脂质相关蛋白。修饰可以包括基因的过表达、表达不足、反义调节、有义抑制、诱导表达、诱导阻
15 遏或诱导调节。

本发明还涉及来自修饰的植物的种子或修饰的植物的分离的产物，其中所述产物可以是酶、营养蛋白质、结构蛋白、氨基酸、脂质、脂肪酸、多糖、糖、醇、生物碱、类胡萝卜素、类丙烷、类固醇、色素、维生素和植物激素。

上述发明概述列出了若干本发明的实施方案，在许多情况中列出了这些实施方案的变异和改变。概述仅仅是无数的变化的实施方案的举例。一个给定的实施方案中提到一或多个特异的特征也是举例性的。这种实施方案可以通常具有或不具有所提到的特征；类似地，那些特征可以被用于概述中列出或未列出的本发明的其它实施方案。为
20 了避免过多的重复，概述没有列出或提示这些特征的所有可能的组合。

为了总结本发明和相对于现有技术的优点的目的，本发明的一

些目的和优点已经在以上描述。当然，应理解的是，不是所有这些目的或优点均可以根据本发明的任何特定实施方案实现。因此，例如，本领域技术人员能认识到本发明可以以实现或优化本文教导的一个优点或一组优点的方式实施，而不必要实现本文可能教导或提示的其它目的或优点。

本发明的其它方面、特征和优点根据下述特异实施方案的详细描述而变得明显。

附图简述

10 图 1 和 SEQ ID NO:1 显示了全长 At5g56860 的 CDS 序列。

图 2 和 SEQ ID NO:2 显示了 At5g56860 的氨基酸序列。

图 3 和 SEQ ID NO:3 显示了 At5g56860 的水稻直向同源物的核酸序列。

15 图 4 是质粒构建体，其 T-DNA 的右边界 (RB) 和左边界 (LB) 之间含有 CaMV 35S 启动子、GNC cDNA 和胭脂氨酸合酶终止子 (nos)。

定义

为清楚起见，说明书中使用的一些术语的定义如下：

20 “结合/可操作地连接 (Associated with / operatively linked)”是指两个核酸序列物理地或功能地相连。例如，在如下情况时启动子和调节 DNA 序列被称为与编码 RNA 或蛋白质的 DNA 序列“相连”：两个序列是可操作地连接的，或者位置关系使得调节 DNA 序列影响编码或结构 DNA 序列的表达。

25 “嵌合构建体”是一种重组核酸序列，其中启动子或调节核酸序列与编码 mRNA 的核酸序列可操作地连接或结合，或者其表达为蛋白质，从而调节核酸序列能调节所结合的核酸序列的转录或表达。

嵌合构建体的调节核酸序列不是以通常在自然界中发现的那样与相关核酸序列可操作地连接。

“辅因子”是酶催化反应所需的天然反应物，如有机分子或金属离子。辅因子是例如 NAD(P)、核黄素（包括 FAD 和 FMN）、叶酸、
5 molybdopterin、硫胺素、生物素、硫辛酸、泛酸和辅酶 A、S-腺苷甲
硫氨酸、吡哆醛磷酸、泛醌、甲基萘醌。任选地，辅因子可以再生和
重复使用。

“编码序列”是转录成 RNA 如 mRNA, rRNA, tRNA, snRNA, 有
义 RNA 或反义 RNA 的核酸序列。具体地，RNA 随后在生物体中被
10 翻译产生蛋白质。

互补：“互补”是指两个核苷酸序列包含反平行的核苷酸序列，
当反平行的核苷酸序列中的互补碱基残基之间形成氢键时，这两个序
列能互相配对。

酶活性：在本文中是指酶催化底物转化成产物的能力。酶的底
15 物包括酶的天然底物，也包括天然底物的类似物，所述天然底物的类
似物也可以被酶转化成产物或产物类似物。酶活性通过例如确定在一
定时间后反应中的产物量或通过测定一定时间后反应混合物中的剩
余的底物量而测量。酶活性也可以通过测定一定时间后反应混合物中
剩余的未使用的辅因子的量或通过测定一定时间后反应混合物中使
20 用的辅因子的量而测量。酶活性还可以通过测定一定时间后反应混合
物中剩余的自由能供体或富含能量的分子（例如 ATP、磷酸烯醇丙酮
酸、乙酰磷酸或磷酸肌酸）的量或通过测定一定时间后反应混合物中
被使用的自由能供体或富含能量的分子（例如 ADP、丙酮酸、乙酸
盐或肌酸）的量而测量。

25 表达盒：本文所用术语“表达盒”是指能够指导特定的核苷酸
序列在合适的宿主细胞中表达的核酸分子，其包含与感兴趣的核苷酸
序列可操作地连接的启动子，所述感兴趣的核苷酸序列与终止信号可

操作地连接。其还通常包括核苷酸序列的正确翻译所需的序列。编码区通常编码感兴趣的蛋白质，但是也可以编码感兴趣的功能 RNA，例如反义 RNA 或非翻译 RNA，它们呈有义或反义方向。包含感兴趣的核苷酸序列的表达盒可以是嵌合的，意味着其至少一种成分相对于其至少一种其它成分是异源的。表达盒还可以是天然存在的但已经以重组形式获得而用于异源表达的表达盒。但是，通常，表达盒对于宿主是异源的，即表达盒的特定 DNA 序列不是在宿主细胞中天然存在的而是必须通过转化事件而被导入宿主细胞或宿主细胞的祖先中。表达盒中的核苷酸序列的表达可以在组成型启动子或诱导型启动子的控制下，诱导型启动子仅当宿主细胞被暴露于一些特定外部刺激时才启动转录。在多细胞生物如植物的情况中，启动子也可以特异于特定的组织或器官或发育阶段。

本文所用术语“功能片段”当与核酸或蛋白质序列相关时是指保持了全长序列功能的序列片段或部分。

15 基因：术语“基因”广义地指与生物学功能相关的任何 DNA 节段。因此，基因包括编码序列和/或其表达所需的调节序列。基因还包括未表达的 DNA 节段，其例如形成针对其它蛋白质的识别序列。基因可以得自各种来源，包括从感兴趣的来源克隆或从已知的或预测的序列信息合成，并且可包括设计具有希望的参数的序列。

20 异源的/外源的：本文所用术语“异源的”和“外源的”是指核酸序列（例如 DNA 序列）或基因，指起源于对特定宿主细胞而言是外来的来源的序列，或者如果来源于相同来源，则是从其原始形式修饰而来的。因此，宿主细胞中的异源基因包括对于特定宿主细胞而言是内源的但已被通过例如 DNA 改组而修饰。所述术语还包括天然存在的 DNA 序列的非天然存在的多拷贝形式。因此，所述术语是指 DNA 节段对于细胞是外来的或异源的或者对于细胞是同源的但在宿主细胞核酸中的位置是该元件通常所不在的位置。外源 DNA 节段被

表达以产生外源多肽。

“同源的”核酸（例如 DNA）序列是与其被导入其中的宿主细胞天然相关的核酸（例如 DNA）序列。

杂交：术语“特异性杂交”是指当特定序列存在于复合混合物（例如总细胞）DNA 或 RNA 中时，分子仅与该特定核苷酸序列在严格条件下结合、形成双螺旋或杂交。“实质上结合（Bind(s) substantially）”是指探针核酸和靶核酸之间的互补杂交，并包含少量错配，其可以通过降低杂交介质的严格性而适应，以实现靶核酸序列的预期的检测。

10 抑物质：一种失活蛋白质的酶活性的化学物质，所述蛋白质如生物合成酶、受体、信号转导蛋白、结构基因产物或转运蛋白。术语“除草剂”（或“除草化合物”）在本文中用于定义一种在发育的任何阶段被施于植物的抑物质，其中所述除草剂抑制植物的生长或杀死植物。

15 相互作用：相互作用的性质或状态，从而一种蛋白质或化合物对另一种蛋白质的有效性或毒性是抑制性的（拮抗剂）或增强性的（激动剂）。

当一种核酸序列编码具有与参照核酸序列所编码的多肽相同的氨基酸序列时，所述核酸序列与所述参照核酸序列同质异构（“isocoding with”）。

20 等基因的：遗传学相同的植物，但可能由于异源 DNA 序列的存在或不存在而不同。

分离的：在本发明中，分离的 DNA 分子或分离的酶是这样的 DNA 分子或酶，其通过人工介入而与其天然环境分离地存在并因此不是天然产物。分离的 DNA 分子或酶可以以纯化的形式存在，或者

25 可以存在于非天然环境中如在转基因宿主细胞中。

成熟蛋白：转运肽、信号肽和/或前肽部分已被除去的蛋白质。

最小启动子：启动子的能支持任何转录的最小片段，如 TATA 元件。最小启动子通常在不存在上游激活时具有大大降低的启动子活性。在存在合适的转录因子时，最小启动子的功能允许转录。

修饰的酶活性：与植物中天然存在的酶活性（即在不存在对酶活性的直接或间接的人工操作时天然存在的酶活性）不同的酶活性，其耐受抑制天然存在的酶活性的抑物质。

天然：是指基因存在于未转化的植物细胞的基因组中。

天然存在：术语“天然存在”是用于描述一种对象，其可以在自然界中发现，与人工产生的不同。例如，可以从自然界来源中分离并且未被人在实验室中有意修饰的、存在于生物体（包括病毒）中的蛋白质或核苷酸序列是天然存在的。

核酸：术语“核酸”是指脱氧核糖核苷酸或核糖核苷酸及其单链或双链形式的聚合物。除非特别限定，否则该术语涵盖含有天然核苷酸的已知类似物的核酸，其具有与参照核酸相似的结合性质并且以与天然存在的核苷酸类似的方式被代谢。除非特别指明，否则特定的核酸序列也包括其保守修饰的变体（例如简并密码子取代）和互补序列以及所明确指明的序列。具体地，简并密码子取代可以通过产生这样的序列而实现，其中一或多个选定的（或全部）密码子的第三位被混合碱基和/或脱氧肌苷残基取代(Batzer et al., *Nucleic Acid Res.* 19: 5081 (1991); Ohtsuka et al., *J. Biol. Chem.* 260: 2605-2608 (1985); Rossolini et al., *Mol. Cell. Probes* 8: 91-98 (1994))。术语“核酸”或“核酸序列”也可与基因、cDNA 和基因编码的 mRNA 互换使用。

“ORF”是指可读框。

相同性百分比：术语“相同性百分比”指核酸或蛋白质序列时是指当对两个或多个序列或亚序列进行最大相应性比较和对比时，通过下述序列比较算法中的一种或用肉眼观察进行检测，具有例如 60%、特别是 70%、更具体是 80%、更具体是 90%、更具体是 95%、

最特别是至少 99%核苷酸或氨基酸残基相同性。具体地, 相同性百分比在至少约 50 个残基长度的序列区域存在, 更具体地在至少约 100 个残基的区域存在, 最具体地相同性百分比在至少约 150 个残基的区域存在。在一个特别具体的实施方案中, 相同性百分比在编码区的整个长度上存在。

为进行序列比较, 通常利用一个序列作为参考序列以与测试序列进行比较。当使用序列比较算法时, 测试序列和参考序列被输入计算机, 如果需要, 指定亚序列坐标, 并指定序列算法程序参数。序列比较算法然后基于指定的程序参数计算测试序列相对于参考序列的序列相同性百分比。

用于比较的序列的最佳对比可以例如用 Smith & Waterman, Adv. Appl. Math. 2: 482 (1981)的局部同源性算法、用 Needleman & Wunsch, J. Mol. Biol. 48: 443 (1970)的同源性对比算法、用 Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85: 2444 (1988)的相似性检索方法、用这些算法的计算机化实现 (GAP, BESTFIT, FASTA 和 TFASTA, 在 Wisconsin Genetics Software Package 中, Genetics Computer Group, 575 Science Dr., Madison, WI)或通过检视(参见 Ausubel et al., *infra*)而进行。

适于确定序列相同性百分比和序列相似性百分比的算法的一个例子是 BLAST 算法, 其描述见 Altschul et al., J. Mol. Biol. 215: 403-410 (1990)。进行 BLAST 分析的软件可在国家生物工程信息中心获得(<http://www.ncbi.nlm.nih.gov/>)。这一算法包括首先通过在查询序列中鉴别短字长 W 而鉴别高评分序列对 (high scoring sequence pairs, HSPs), 所述字长 W 当与数据库序列中的相同长度的字对比时匹配或满足一些正值阈值分 T 。 T 被称为相邻字符分值阈值 (Altschul et al., 1990)。这些最初的相邻字符击中(hit)作为种子起始搜索以发现含有它们的更长的 HSP。只要累积对比分可以增加, 所述字

符击中随后被沿每个序列在两个方向上延伸。对于核苷酸序列，累积分使用参数 M （对于匹配残基对的奖分，总是 > 0 ）和 N （错配残基的罚分，总是 < 0 ）进行计算。对于氨基酸序列，使用评分矩阵计算累积分。当累积对比分从其最大实现值下降 X 时，字符击中在每个方向上的延伸停止，由于一或多个负分残基对比的累积或者达到每个序列的末端，累积分变为 0 或低于 0。BLAST 算法参数 W , T 和 X 决定了对比的敏感性和速度。BLASTN 程序（针对核苷酸序列）使用如下默认参数：字长（ W ）为 11，期望值（ E ）为 10，截断值为 100， $M=5$ ， $N=-4$ ，比较两条链。对于氨基酸序列，BLASTP 程序使用如下默认参数：字长（ W ）为 3，期望值（expectation, E ）为 10，和 BLOSUM62 评分矩阵(参见 Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89: 10915 (1989))。

除了计算序列相同性百分比外，BLAST 算法还进行两个序列之间的相似性的统计学分析（参见例如 Karlin & Altschul, Proc. Nat'l. Acad. Sci. USA 90: 5873-5787 (1993))。BLAST 算法提供的相似性的一个测量是最小和概率 ($P(N)$)，其提供了两个核苷酸或氨基酸序列之间发生匹配的机会的概率的指示。例如，如果测试核酸序列与参照核酸序列相比的最小和概率低于约 0.1、更具体地低于约 0.01、最具体地低于约 0.001，则测试核酸序列被认为与参照序列相似。

20 前蛋白（Pre-protein）：被正常导向细胞器如叶绿体并仍包含其天然转运肽的蛋白质。

纯化的：术语“纯化的”当用于核酸或蛋白质时，是指所述核酸或蛋白质基本上不含天然状态下与其相关的其它细胞成分。其特别是在均质状态，当然也可以在干或水溶液中。纯度和均质性通常用分析化学技术如聚丙烯酰胺凝胶电泳或高效液相色谱确定。在物质中是主要成分的蛋白质是基本上纯化的。术语“纯化的”是指核酸或蛋白质在电泳凝胶中基本上是一条带。具体地，其是指核酸或蛋白质是至

少约 50% 纯的，更具体是至少约 85% 纯的，最特别是至少约 99% 纯的。

- 当来自两个核酸中每一个的序列组合在后代核酸中时这两个核酸是“重组的”。当两个核酸均是重组的底物时这两个序列是“直接”
- 5 重组的。当两个序列用中间体如交换(cross-over)寡核苷酸重组时，这两个序列是“间接重组的”。对于间接重组，不超过一个序列是重组的真正底物，在一些情况中，无一序列是重组的底物。

- “调节元件”是指参与控制核苷酸序列的表达的序列。调节元件包括与感兴趣的核苷酸序列可操作地连接的启动子以及终止信号。
- 10 它们还通常包括核苷酸序列的正确翻译所需的序列。

显著增加：比测量技术固有的误差限度更大的酶活性增加，特别是在抑制剂存在下比野生型酶活性增加约 2 倍或更多，更具体是增加约 5 倍或更多，最特别是增加约 10 倍或更多。

- 显著低于：是指酶反应的产物量比测定技术固有的误差的限度
- 15 低更多，特别是在不存在抑制剂时比野生型酶活性降低约 2 倍或更多，更具体是降低约 5 倍或更多，最特别是降低约 10 倍或更多。

- 特异的结合/免疫学交叉反应性：两个核酸序列或蛋白质基本上相同的指征是由第一个核酸编码的蛋白质与由第二个核酸编码的蛋白质免疫学交叉反应或特异性结合。因此，例如当两个蛋白质的不同
- 20 仅在于保守性取代时，一个蛋白质通常与第二个蛋白质基本上相同。
- 术语“与抗体特异性（或选择性）结合”或“特异性（或选择性）与...发生免疫反应”当指蛋白质或肽时，是指结合反应是在蛋白质和群体生物物质的异质群体的存在下存在该蛋白质的决定因素。因此，在指定的免疫测定条件，所指明的抗体与特定的蛋白质结合，而不以显著
- 25 量结合样品中存在的其它蛋白质。在这种条件下与抗体的特异性结合可能需要针对其对一特定蛋白质的特异性而选择的抗体。例如，针对具有由本发明的任一核酸序列编码的氨基酸序列的蛋白质产生的抗

- 体可以被选择以获得与该蛋白质特异性免疫反应但不与除多态性变体之外的其它蛋白质反应的抗体。可以使用各种免疫测定来选择与特定蛋白质特异性免疫反应的抗体。例如，固相 ELISA 免疫测定，Western 印迹或免疫组织化学被常规用于选择与蛋白质特异性免疫反应
- 5 应的单克隆抗体。免疫测定形式和可用于确定特异性免疫反应性的条件的描述参见 Harlow and Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York “Harlow and Lane”)。通常，特异性或选择性反应是背景信噪的至少两倍，更通常是高于背景的 10—100 倍。
- 10 “严格杂交条件”和“严格杂交洗涤条件”在核酸杂交实验如 Southern 和 Northern 杂交中是序列依赖性的，并且在不同环境参数下是不同的。较长的序列在较高的温度下特异杂交。核酸杂交的详细指导见 Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes part I chapter*
- 15 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays" Elsevier, New York。通常，高严格杂交和洗涤条件的选择是低于特定序列在规定的离子强度和 pH 的解链温度 (T_m) 约 5°C。通常，在“严格条件”下，探针将与其靶亚序列杂交，但不与其它序列杂交。
- 20 T_m 是 50%靶序列与完全匹配的探针杂交的温度（在规定的离子强度和 pH 下）。对于特定的探针选择非常严格的条件相等于 T_m 。杂交 Southern 或 northern 印迹滤膜上的具有多于 100 个互补残基的互补核酸的严格杂交条件的一个例子是具有 1mg 肝素的 50%甲酰胺，42°C，杂交过夜。高严格洗涤条件的一个例子是 0.1 5M NaCl、72°C
- 25 洗涤约 15 分钟。严格洗涤条件的一个例子是在 65°C 在 0.2x SSC 中洗涤 15 分钟 (SSC 缓冲液的描述参见 Sambrook, *infra*)。通常，高严格性洗涤之前为低严格性洗涤以除去背景探针信号。对于例如超过

100 个核苷酸的双链的中等严格洗涤的例子是 1x SSC、45°C 洗涤 15 分钟。对于例如超过 100 个核苷酸的双链的低严格洗涤的例子是 4-6x SSC、40°C 洗涤 15 分钟。对于短探针（例如约 10—50 个核苷酸），严格条件通常包括小于约 1.0 M Na 离子、通常约 0.01—1.0 M Na 离子浓度（或其它盐）pH7.0-8.3 的盐浓度，温度通常是至少约 30°C。严格条件也可以通过添加去稳定剂如甲酰胺而实现。通常，在特定杂交测定中是针对无关探针观察到的信噪比的 2 倍（或更高）表明检测到特异杂交。在严格条件下不互相杂交的核酸仍是基本上相同的，如果它们编码的蛋白质是基本上相同的。这例如当核酸的拷贝是用遗传密码所允许的最大密码简并性产生时。

以下是杂交/洗涤条件组合的例子，其可以被用于克隆是本发明的参照核苷酸序列的同系物的核苷酸序列：参照核苷酸序列特异性与参照核苷酸序列在 7%十二烷基硫酸钠(SDS), 0.5 M NaPO₄, 1 mM EDTA 中于 50°C 杂交并在 50°C 于 2X SSC, 0.1% SDS 中洗涤，更希望的是在 7%十二烷基硫酸钠(SDS), 0.5 M NaPO₄, 1 mM EDTA 中于 50°C 杂交并在 50°C 于 1X SSC, 0.1% SDS 中洗涤，更期望的是在 7%十二烷基硫酸钠(SDS), 0.5 M NaPO₄, 1 mM EDTA 中于 50°C 杂交并在 50°C 于 0.5X SSC, 0.1% SDS 中洗涤，特别是在 7%十二烷基硫酸钠(SDS), 0.5 M NaPO₄, 1 mM EDTA 中于 50°C 杂交并在 50°C 于 0.1X SSC, 0.1% SDS 中洗涤，更具体是在 7%十二烷基硫酸钠(SDS), 0.5 M NaPO₄, 1 mM EDTA 中于 50°C 杂交并在 65°C 于 0.1X SSC, 0.1% SDS 中洗涤。

“亚序列 (subsequence)”是指分别包含更长的核酸或氨基酸序列（例如蛋白质）的一部分的核酸或氨基酸序列。

基本相似：术语“基本相似”在指两个核酸或蛋白质序列时，是指两个或多个序列或亚序列基本相似，例如具有 50%、特别是 60%、更具体是 70%、更具体是 80%、更具体是 90%、更具体是 95

%、最特别是 99% 序列相同性。

底物：底物是酶天然识别并转化为酶天然实现其功能的生物化学途径中的产物的分子，或者是所述分子的修饰版本，其也被酶识别并且在类似于天然存在反应的酶反应中被酶转化为产物。

- 5 转化：将异源 DNA 导入植物细胞、植物组织或植物中的过程。应理解转化的植物细胞、植物组织或植物不仅包含转化过程的终产物，也包括其转基因后代。

“转化的”、“转基因的”和“重组的”是指宿主生物体如导入了异源核酸分子的细菌或植物。所述核酸分子能被稳定整合进宿主的基因组中，或者所述核酸分子也能以染色体外分子的形式存在。这种染色体外分子可以自身复制。应理解转化的细胞、组织或植物不仅包含转化过程的终产物，也包括其转基因后代。“未转化的”、“非转基因的”或“非重组的”宿主是指野生型生物体例如细菌或植物，其不含异源核酸分子。

- 15 生存力：本文所用术语“生存力”是植物的健康(fitness)参数。测定植物发育的纯合性能，由此指明哪些蛋白质对于植物生长是必需的。

发明的详细描述

20

I. 性状功能基因组学的一般描述

功能基因组学的目的是鉴别控制生物体表型表达的基因，并采用许多方法学，包括但不限于生物信息学、基因表达研究、基因和基因产物相互作用、遗传学、生物化学和分子遗传学。例如，生物信息学可以通过鉴别异源生物体中的具有氨基酸或核苷酸水平高度相似性(同源性)的基因而确定给定基因的功能。mRNA 或蛋白质水平的基因表达可以通过将基因表达与环境应答、发育过程或者遗传（突变）

或分子遗传（基因过表达或表达不足）扰动相关联而确定一给定基因的功能。mRNA 水平的基因表达可以单独确定（Northern 分析）或者与其它基因一起确定（微阵列分析），而蛋白质水平的基因表达可以单独确定（天然或变性蛋白质凝胶或免疫印迹分析）或与其它基因一起确定（蛋白质组学分析）。有关蛋白质/蛋白质和蛋白质/DNA 相互作用的 5 知识可以通过鉴别在同一生物学过程中一起起作用的蛋白质和核酸序列而确定功能。遗传学可以通过证实基因中的 DNA 损害（突变）具有对生物体的可量化的效果而确定基因的功能，对生物体的可量化的效果包括但不限于：其发育；激素生物合成和应答；生长和生长习性（植物体系结构）；mRNA 表达谱；蛋白质表达谱；抵抗疾病的能力；非生物应激的耐受；获得营养素的能力；光合作用效率；改变的初级和次级代谢；和各种植物器官的组成。生物化学方法可以通过证实通常当在异源生物体中表达时，由基因编码的蛋白质单独或与其它蛋白质组合具有一定的酶活性而确定功能。分子遗传学方法可以通过在天然植物或在异源生物体中过表达基因或使基因表达不足，并 10 观察上述遗传学功能确定中描述的可量化效果而确定功能。在功能基因组学中，可以使用任何或全部这些方法，通常一起使用，以针对各种生物表型中的任意表型确定基因功能。

本领域技术人员了解这些不同的方法学可以各自提供数据作为 20 特定基因的功能的证据，这种证据随着用于功能确定的数据量的增加而更强有力：特别是来自一种方法学，更具体是来自两种方法学，更具体是来自超过两种方法学。另外，本领域技术人员能意识到不同的方法学在关于基因功能确定中的证据力度方面有不同。通常，但不是总是这样，生物化学、遗传学和分子遗传学证据的数据被认为是比生物信息学或基因表达证据的数据更有力。最终，本领域技术人员能认识到对于不同的基因，来自一种方法学的单一数据可能在证据力度方面不同，所述证据由用于确定不同基因的功能的每种独立的数据提 25

供。

作物性状功能基因组学的目的是鉴别作物性状基因，即能在作物中赋予有用的农学性状的基因。这类农学性状包括但不限于：增加的产量，可以是数量上也可以是质量上；增加的营养素获得和增加的代谢效率；增强的或改变的用于食品、饲料、纤维或加工的植物组织的营养组成；增加的农业或工业加工用途；增加的对植物疾病的抗性；增加的对不利环境条件（非生物应激）包括但不限于干旱、过冷、过热或过度土壤盐度或极端酸度或碱度的耐受；以及植物系统结构或发育的改变，包括发育时间的变化。通过转基因或非转基因手段运用这类被鉴别的性状基因可以从本质上改良作物，对农业有利。

在人类和动物消费方面，谷类是地球上最重要的作物。在水稻、玉米、小麦、大麦、黑麦、燕麦和其它农业重要的单子叶植物中观察到了基因组同线性(syntenly)（在大染色体节段内基因顺序保守），其促进了基于单个谷类基因对来自不同谷类物种的直向同源基因的作图和分离。水稻在谷类中具有最小的 (~ 420 Mb)基因组，并在最近成为公共和私有基因组和 EST 测序努力的主要焦点。

为了鉴别水稻[小麦]基因组中控制[性状]的作物性状基因，来自水稻草图基因组序列[小麦 EST 数据库]的基因被基于一或多种功能基因组方法学分出优先次序。例如，利用稻瘟病真菌(*Magnaporthe grisea*)感染的水稻植物的基因组范围的表达研究来区分控制疾病抗性的候选基因的优先次序。随后可基于水稻全基因组序列的分析预测水稻性状基因候选物的全长或部分 cDNA，并通过利用商业可获得的 PCR 引物挑选程序设计和使用 PCR 扩增引物而分离。引物被用于来自水稻 cDNA 文库的全长或部分 cDNA 或第一链 cDNA 的 PCR 扩增。从每种途径产生的 cDNA 克隆被用于用下述植物分子遗传方法学来构建载体，所述载体设计用于改变这些基因在转基因植物中的表达。通过在转基因植物中过表达关键性状基因或使这些基因表达不足而

改变植物表型是用于确定植物基因功能的已建立的方法。鉴别具有改变的感兴趣性状的转基因植物的测定被毫无疑问地用于利用这些基因通过转基因或经典育种方法来改良水稻，并可延及到其它谷类。

5 II. 鉴别、克隆和测序cDNA

本发明的cDNA的克隆和测序在实施例1中描述。

本发明的分离的核酸和蛋白质可以用于一系列植物，即单子叶植物和双子叶植物，特别是单子叶植物如水稻、小麦、大麦和玉米。在一个更具体的实施方案中，所述单子叶植物是谷类。在一个更具体的实施方案中，谷类可以是例如玉米、小麦、大麦、稷燕麦、裸麦、稷、高粱、黑小麦、黑麦、单粒小麦、斯佩耳特小麦、双粒小麦、画眉草、买罗高粱、亚麻、gramma草、摩擦草和类蜀黍。在最具体的实施方案中，谷类是水稻。其它的植物属包括但不限于葫芦属(Cucurbita)，蔷薇属(Rosa)，葡萄属(Vitis)，核桃(Juglans)，Gragaria，莲属(Lotus)，苜蓿(Medicago)，驴食草属(Onobrychis)，胡卢巴属(Trigonella)，豇豆属(Vigna)，柑桔属(Citrus)，亚麻属(Linum)，天竺葵属(Geranium)，木薯属(Manihot)，胡萝卜属(Daucus)，鼠耳芥属，芸苔属(Brassica)，萝卜属(Raphanus)，百芥属(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)，Solanum，矮牵牛(Petunia)，洋地黄(Digitalis)，Majorana，Ciahorium，向日葵属(Helianthus)，莴苣属(Lactuca)，雀麦属(Bromus)，芦笋属(Asparagus)，金鱼草属(Antirrhinum)，Heterocallis，Nemesis，天竺葵属(Pelargonium)，Panieum，狼尾草属(Pennisetum)，毛茛属(Ranunculus)，千里光属(Senecio)，Salpiglossis，黄瓜属(Cucumis)，Browaalia，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，黑麦属，葱属(Allium)和小麦属(Triticum)。

本发明还提供了对包含本发明的核酸分子的植物或植物部分进行基因分型的方法。任选地，植物是单子叶植物如但不限于水稻或小麦。基因分型提供了区分染色体 *pari* 的同系物的手段并可以被用于区分植物群中的分离分子。分子标记方法可以用于系统发生学研究，鉴定作物品种间的遗传关系，鉴别杂交或体细胞杂种，定位影响单基因(mongenic)性状的染色体区段，基于图谱克隆以及研究数量遗传(see Plant Molecular Biology: A Laboratory Manual, Chapter 7, Clark ed., Springer-Verlag, Berlin 1997; Paterson, A.H., "The DNA Revolution", chapter 2 in Genome Mapping in Plants, Paterson, A.H. ed., Academic Press/R.G. Lands Co., Austin, Texas 1996)。

基因分型的方法可以采用任意数量的分子标记分析技术如但不限于限制长度多态性 (RFLP)。如本领域熟知的，RFLP 通过 DNA 限制片段长度的差异而产生，而所述长度的差异是由于同一基因的等位基因之间的核苷酸差异导致。因此，本发明提供了通过使用 RFLP 分析跟踪本发明的基因或核酸或者遗传连锁的染色体序列的分离的方法。连锁的染色体序列为 50 厘摩(centiMorgans)(50 cM) 以内、40 或 30 cM 以内、特别是 20 或 10cM 以内，更具体是 5、3、2 或 1cM 以内的本发明的核酸。

20 III. 感兴趣的性状

本发明涵盖了多核苷酸的鉴别和分离，所述多核苷酸编码参与糖感应和最终氮摄取和碳代谢的蛋白质。改变与这些性状相关的基因的表达可以被用于根据需要改良或修饰植物和/或谷物。实施例描述了感兴趣的分离的基因以及分析表达的改变及其对植物特征的作用效果的方法。

本发明的一个方面提供了用于改变(即增加或降低)本发明的核酸分子和多肽在植物中的水平的组合物和方法。具体地，本发明的核

酸分子和多肽组成性、时间性或空间性表达，例如在发育阶段、在某些组织中和/或以一定数量表达，这些不是非重组改造的植物的特征。因此，本发明提供了在示例性应用如改变上述指明的特征中的用途。

5 VI. 在转基因植物中控制基因表达

本发明进一步涉及包含核酸分子的转化的细胞，转化的植物，种子和植物部分，以及通过改变本发明的基因的表达而修饰感兴趣的表型性状的方法。

10 A. 修饰编码序列和相邻序列

衍生自异源来源的基因在植物中的转基因表达可涉及修饰这些基因以实现和优化它们在植物中的表达。特别是，编码不同的酶但由天然微生物中的同一转录物编码的细菌 ORF 在植物中最好是在不同的转录物上表达。为此，逐个分离每种微生物 ORF 并克隆在表达盒
15 内，该表达盒在 ORF 的 5'末端提供植物启动子序列，在 ORF 的 3'末端提供植物转录终止子。分离的 ORF 序列特别包括起始 ATG 密码子和终止 STOP 密码子，但可以包括在起始 ATG 和 STOP 密码子之外的额外序列。另外，ORF 可以被截短，但仍保留所需的活性；对于特别长的 ORF，保留活性的截短版本可能对于在转基因生物体中的表达是优选的。“植物启动子”和“植物转录终止子”是指在植物
20 细胞内起作用的启动子和转录终止子。这包括可以衍生自非植物来源如病毒（一个例子是花椰菜花叶病毒）的启动子和转录终止子。

在一些情况中，对 ORF 编码序列和相邻序列的修饰是不需要的。分离含有感兴趣的 ORF 的片段并将其插入植物启动子的下游就
25 足够了。例如，Gaffney et al. (Science 261: 754-756 (1993)) 成功地在转基因植物中表达了在 CaMV 35S 启动子和 CaMV tm 终止子控制下的假单胞菌 nahG 基因，其没有修饰编码序列并且 ATG 上游的假单胞

菌基因的核苷酸和 STOP 密码子下游的核苷酸仍然附着于 nahG ORF。具体地, 应该有尽可能少的相邻微生物序列附着在 ATG 上游和 STOP 密码子下游。在实践中, 这种构建可依赖于限制位点的可用性。

5 在其它情况中, 衍生自微生物来源的基因的表达可能会在表达中产生问题。这些问题在本领域中已有鉴定并且对于衍生自一些来源如芽孢杆菌的基因是特别常见的。这些可能适用于本发明的核苷酸序列, 这些基因的修饰可以用本领域熟知的技术进行。可能遇到下面这些问题:

10 1. 密码子选择

植物中的特异密码子选择不同于某些微生物中的特异密码子选择。将克隆的微生物 ORF 中的密码子选择与植物基因(特别是来自靶植物的基因)中的密码子选择进行比较将使得能鉴别 ORF 中应该被特别改变的密码子。通常, 植物进化是在单子叶植物的第 3 碱基位置强烈优选核苷酸 C 和 G, 而双子叶植物通常在这一位置使用核苷酸 A 或 T。通过修饰基因以掺入针对特定靶转基因物种的特异密码子选择, 下述许多关于 GC/AT 含量和不正常剪接的问题会被克服。

2. GC/AT 含量

20 植物基因通常具有超过 35% 的 GC 含量。富含 A 和 T 核苷酸的 ORF 序列可在植物中导致严重问题。首先, ATTTA 基序据信导致信使不稳定, 并在许多短寿命 mRNA 的 3' 末端被发现。其次, 在信使内的不合适位置出现聚腺苷化信号如 AATAAA 据信导致转录的过早截短。另外, 单子叶植物可识别富 AT 序列作为剪接位点(见下述)。

3. 相邻于起始甲硫氨酸的序列

25 植物有别于微生物之处在于它们的信使不具有限定的核糖体结合位点。相反, 据信核糖体附着于信使的 5' 末端并扫描第一个可利用的 ATG 以开始翻译。不过, 据信, ATG 相邻的一些核苷酸是有优先

选择的，并且微生物基因的表达可以通过在 ATG 包括真核共有翻译起始子而增强。Clontech (1993/1994 catalog, page 210, incorporated herein by reference)已经提示一种序列作为在植物中表达大肠杆菌 uidA 基因的共有翻译起始子。另外,Joshi (N.A.R. 15: 6643-6653 (1987),
5 包含在此作为参考) 比较了相邻于 ATG 的许多植物序列并提示了另一种共有序列。在微生物 ORF 在植物中的表达遇到困难的情况下，在起始 ATG 处包括这些序列中的一种可能会改善翻译。在这种情况下，共有序列的最后 3 个核苷酸可能因为它们在第二个 AA 残基的修饰而不适合包括在修饰的序列中。在不同植物物种中相邻于起始甲硫
10 氨酸的具体序列可能不同。对 GenBank 数据库中的 14 个玉米基因的调查提供了如下结果：

在 14 个玉米基因中起始 ATG 之前的位置：：

		-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
15	C	3	8	4	6	2	5	6	0	10	7
	T	3	0	3	4	3	2	1	1	1	0
	A	2	3	1	4	3	2	3	7	2	3
	G	6	3	6	0	6	5	4	6	1	5

20 可以对掺入核苷酸序列的希望的植物物种进行这一分析，并且修饰 ATG 相邻的序列以掺入特异的核苷酸。

4. 除去不正常剪接位点

从非植物来源克隆的并且没有进行表达优化的基因可能还含有在植物中可被识别为 5’或 3’剪接位点的基序，并且所述基序可被裂
25 解，由此产生截短的或缺失的信使。这些位点可以用本领域熟知的技术除去。

用于修饰编码序列和相邻序列的技术是本领域熟知的。当微生

物 ORF 的初始表达低并且认为对序列进行上述改变是合适的时, 可以根据本领域熟知的方法进行合成基因的构建。这些例如在如下公布的专利文献中描述: EP 0 385 962 (to Monsanto), EP 0 359 472 (to Lubrizol)和 WO 93/07278 (to Ciba-Geigy), 所有文献并入本文作参考。

- 5 在大多数情况中, 优选的是在将基因转移进转基因植物中之前用瞬时测定方法(这是本领域熟知的)测定基因构建体的表达。

B. 植物表达盒的构建

- 首先将意欲在转基因植物中表达的编码序列组装在表达盒中可在植物中表达的合适的启动子之后。表达盒还可包含转基因的表达所需的或所选择的序列。这些序列包括但不限于转录终止子、增加表达的额外序列如内含子、关键(vital)序列以及用于将基因产物导向特异细胞器和细胞区室的序列。这些表达盒随后可以被容易地转移到下述植物转化载体。以下是常规表达盒的各种成分的描述。
- 10

15 1. 启动子

- 表达盒中所用的启动子的选择将决定转基因在转基因植物中的空间和时间表达模式。选择的启动子将在特异的细胞类型(如叶表皮细胞、叶肉细胞、根皮层细胞)或特异的组织或器官(例如根、叶或花)中表达转基因, 选择将反映希望的基因产物积累位置。或者, 选择的启动子可以在各种诱导条件下驱动基因表达。启动子的长度不同, 即启动转录的能力不同。根据所用宿主细胞系统, 可以使用许多合适启动子中的任一种, 包括基因的天然启动子。以下是可以用于表达盒的启动子的非限制性例子。
- 20

a. 组成型表达, 遍在蛋白启动子:

- 25 遍在蛋白是已知在许多细胞类型中积累的基因产物, 已经从几种物种中克隆了其启动子用于转基因植物中(例如向日葵- Binet et al. Plant Science 79: 87-94 (1991); 玉米 - Christensen et al. Plant Molec.

Biol. 12: 619-632 (1989); 和鼠耳芥 - Callis et al., J. Biol. Chem. 265:12486-12493 (1990) 和 Norris et al., Plant Mol. Biol. 21:895-906 (1993)。玉米遍在蛋白启动子已在转基因单子叶植物系统中被开发，其序列和构建用于单子叶植物转化的载体在专利公开 EP 0 342 926
5 (to Lubrizol)中描述，该专利公开引入本文作参考。Taylor et al. (Plant Cell Rep. 12: 491-495 (1993)) 描述了一种(pAHC25)，其包含玉米遍在蛋白启动子和第一内含子，当经微粒轰击导入时其在各种单子叶植物的细胞悬液中有高活性。鼠耳芥遍在蛋白启动子理想地与本发明的核苷酸序列一起使用。该遍在蛋白启动子适于在单子叶和双子叶的转基因植物中进行基因表达。合适的载体是 pAHC25 的衍生物或这一申
10 请中描述的任一转化载体，其通过导入合适的遍在蛋白启动子和/或内含子序列而修饰。

b. 组成型表达，CaMV 35S 启动子:

质粒 pCGN1761 的构建在公开的专利申请 EP 0 392 225 (实施例
15 23)中描述，该专利申请引入本文作参考。pCGN1761 含有“双” CaMV 35S 启动子和 tml 转录终止子并具有 pUC 型骨架，在启动子和终止子之间有唯一的 EcoRI 位点。构建 pCGN1761 的衍生物，其具有修饰的多接头，所述多接头除了包含已存在的 EcoRI 位点之外还包括 NotI 和 XhoI 位点。这一衍生物被命名为 pCGN1761ENX。pCGN1761ENX
20 可用于将 cDNA 序列或编码序列（包括微生物 ORF 序列）克隆在其多接头内，以在 35S 启动子控制下在转基因植物中表达它们。这种构建体的完整 35S 启动子-编码序列-tml 终止子表达盒可以在启动子 5' 的 HindIII, SphI, SalI 和 XbaI 位点以及终止子 3' 的 XbaI, BamHI 和 BglII 位点被切割，以转移至如下所述的转化载体。另外，双 35S 启动子片段可以通过用 HindIII, SphI, SalI, XbaI 或 PstI 进行 5'切割以及利用多接头的任何限制位点 (EcoRI, NotI 或 XhoI) 进行 3'切割，从而
25 用另一种启动子置换。如果需要，可以通过导入可增强翻译的序列而

围绕克隆位点进行修饰。这在希望过表达时特别有用。例如，pCGN1761ENX 可以通过如美国专利 5,639,949 的实施例 37 所述优化翻译起始位点而被修饰，该美国专利引入本文作参考。

c. 组成型表达，肌动蛋白启动子：

5 已知几种肌动蛋白同种型在大多数细胞类型中表达，因此肌动蛋白启动子是组成型启动子的一个好的选择。具体地，来自水稻 ActI 基因的启动子已被克隆和鉴定 (McElroy et al. Plant Cell 2: 163-171 (1990))。发现该启动子的 1.3kb 片段含有在水稻原生质体中表达所需的全部调节元件。另外，已经构建了基于 ActI 启动子的许多表达载体以特别用于单子叶植物 (McElroy et al. Mol. Gen. Genet. 231: 150-160 (1991))。这些载体掺入了 ActI-内含子 1, AdhI 5'侧翼序列和 AdhI-内含子 1 (来自玉米醇脱氢酶基因)和来自 CaMV 35S 启动子的序列。显示最高表达的载体是 35S 和 ActI 内含子或 ActI 5'侧翼序列和 ActI 内含子的融合体。(GUS 报道基因的)起始 ATG 周围的序列的优化也增强表达。McElroy et al. (Mol. Gen. Genet. 231: 150-160 (1991))所述的启动子表达盒可以被容易地修饰用于基因表达并且特别适于用在单子叶植物宿主中。例如，从 McElroy 构建体中取出含有启动子的片段并用于置换 pCGN1761ENX 中的双 35S 启动子，其随后可用于插入特异的基因序列。如此构建的融合基因可以随后被转移至合适的转化载体。在另一个报道中，含有其第一个内含子的水稻 ActI 启动子已被发现在培养的大麦细胞中指导高表达 (Chibbar et al. Plant Cell Rep. 12: 506-509 (1993))。

d. 诱导型表达，PR-1 启动子：

25 pCGN1761ENX 中的双 35S 启动子可以用选择的产生合适的高表达水平的任何其它启动子置换。例如，美国专利 5,614,395 中描述的一种化学可调节的启动子如烟草 PR-1a 启动子可以置换双 35S 启动子。或者，可以使用 Lebel et al., Plant J. 16:223-233 (1998) 描述的鼠

耳芥 PR-1 启动子。选择的启动子通过限制酶从其来源特别切下，但是也可以用携带合适的末端限制位点的引物经 PCR 扩增。如果进行 PCR 扩增，则在将扩增的启动子克隆在靶载体中之后应对启动子重新测序以检查扩增错误。化学/病原体可调节的烟草 PR-1a 启动子是从质粒 pCIB1004 (构建参见 EP 0 332 104 的实施例 21，该文献引入本文作参考) 中切下并转移进质粒 pCGN1761ENX (Uknes et al., Plant Cell 4: 645-656 (1992)) 中。pCIB1004 用 NcoI 切割，所得线性化片段的 3' 突出端用 T4 DNA 聚合酶处理而平末端化。该片段然后用 HindIII 裂解，所得含有 PR-1a 启动子的片段用凝胶纯化并克隆进除去了双 35S 启动子的 pCGN1761ENX 中。这如下实现：用 XhoI 切割并用 T4 聚合酶平末端化，随后用 HindIII 切割，分离含有载体-终止子的大片段并将 pCIB1004 启动子片段克隆进其中。这产生了 pCGN1761ENX 衍生物，其具有 PR-1a 启动子和 tml 终止子以及具有独特的 EcoRI 和 NotI 位点的间插多接头。选择的编码序列可以被插入这一载体，融合产物（即启动子-基因-终止子）可以随后被转移至任何选择的转化载体，包括下述那些。可采用各种化学调节剂诱导选择的编码序列在根据本发明转化的植物中的表达，包括美国专利 5,523,311 和 5,614,395 描述的苯并噻二唑(benzothiadiazole)，异烟酸和水杨酸化合物。

20 e. 诱导型表达，乙醇诱导型启动子：

由某些醇或酮类如乙醇诱导的启动子也可用于赋予本发明的编码序列的诱导型表达。这种启动子例如是来自 *m Aspergillus nidulans* 的 *alcA* 基因启动子 (Caddick et al. (1998) Nat. Biotechnol 16:177-180)。在 *A. nidulans* 中，*alcA* 基因编码醇脱氢酶 I，其表达由 AlcR 转录因子在化学诱导物的存在下调节。为本发明目的，质粒 *palcA*:CAT 中的 CAT 编码序列被本发明的编码序列置换以形成具有在 *alcA* 基因启动子控制下的编码序列的表达盒，所述质粒 *palcA*:CAT

包含与最小 35S 启动子融合的 *alcA* 基因启动子序列 (Caddick et al. (1998) *Nat. Biotechnol* 16:177-180)。这用本领域熟知的方法实现。

f. 诱导型表达，糖皮质激素诱导型启动子：

用基于类固醇激素的系统诱导本发明的核酸序列的表达也包括
5 在本发明内。例如，使用糖皮质激素介导的诱导系统(Aoyama and Chua (1997) *The Plant Journal* 11: 605-612)，基因表达通过应用糖皮质激素而诱导，所述糖皮质激素例如合成的糖皮质激素，具体是地塞米松，浓度范围特别是 0.1mM—1mM，更具体是 10mM—100mM。为了本发明的目的，萤光素酶基因序列被本发明的核酸序列置换以形成
10 具有在六拷贝的与 35S 最小启动子融合的 GAL4 上游激活序列的控制下的本发明的核酸序列的表达盒。这使用本领域熟知的方法进行。反式作用因子包含 GAL4 DNA 结合结构域(Keegan et al. (1986) *Science* 231: 699-704)，该结构域与疱疹病毒蛋白 VP16 的反式激活结构域融合 (Triezenberg et al. (1988) *Genes Devel.* 2: 718-729)，而后者与大鼠
15 糖皮质激素受体的激素结合结构域 (Picard et al. (1988) *Cell* 54: 1073-1080) 融合。融合蛋白的表达由本领域已知的启动子或本文描述的启动子控制。这种表达盒还包含在植物中，所述植物包含与 6xGAL4/最小启动子融合的本发明的核酸序列。因此，融合蛋白的组织或器官特异性得以实现，导致杀昆虫毒素的诱导型组织或器官特异
20 性。

g. 根特异性表达：

基因表达的另一种模式是根表达。合适的根启动子是玉米金属硫蛋白样(MTL)基因的启动子，如 de Framond 所述(FEBS 290: 103-106 (1991))，以及如美国专利 5,466,785 所述，这些文献引入本文作参考。
25 这种“MTL”启动子被转移至合适的载体如 pCGN1761ENX，用以插入选择的基因并随后将完整的启动子-基因-终止子盒转移至感兴趣的转化载体。

h. 创伤诱导型启动子:

创伤诱导型启动子也适合基因表达。已经描述了许多这样的启动子(例如 Xu et al. *Plant Molec. Biol.* 22: 573-588 (1993), Logemann et al. *Plant Cell* 1: 151-158 (1989), Rohrmeier & Lehle, *Plant Molec. Biol.* 22: 783-792 (1993), Firek et al. *Plant Molec. Biol.* 22: 129-142 (1993), Warner et al. *Plant J.* 3: 191-201 (1993)), 所有均适用于本发明。Logemann et al.描述了双子叶马铃薯 *wunI* 基因的 5' 上游序列。Xu et al.显示来自双子叶马铃薯的创伤诱导型启动子 (*pin2*) 在单子叶水稻中有活性。另外, Rohrmeier & Lehle 描述了玉米 *WipI* cDNA 的克隆, 其是创伤诱导型的并且可以用于经标准技术分离协同启动子。类似地, Firek et al.和 Warner et al.描述了来自单子叶植物 *Asparagus officinalis* 的创伤诱导型基因, 其在局部创伤和病原体侵入部位表达。使用本领域熟知的克隆技术, 这些启动子可被转移至合适的载体, 与本发明的基因融合, 并用于在植物创伤部位表达这些基因。

15 i. 髓层特异性表达:

引入本文作参考的专利申请 WO 93/07278 描述了玉米 *trpA* 基因的分离, 该基因优先在髓层细胞中表达。给出了基因序列以及从转录起点向上延伸至-1726bp 的启动子。使用标准分子生物学技术, 这种启动子或其部分可以被转移至载体如 pCGN1761, 其可以置换 35S 启动子并被用于驱动外来基因以髓层特异性方式表达。事实上, 含有髓层特异性启动子或其部分的片段可以被转移至任何载体并被修饰以用于转基因植物。

20 j. 叶特异性表达:

编码磷酸烯醇羧化酶(PEPC)的玉米基因已经由 Hudspeth & Grula 描述 (*Plant Molec Biol* 12: 579-589 (1989))。使用标准分子生物学技术, 这一基因的启动子可以被用于任何基因在转基因植物中以叶特异性方式表达。

k. 花粉特异性表达:

WO 93/07278 描述了玉米钙依赖性蛋白激酶 (CDPK) 基因的分离, 该基因在花粉细胞中表达。基因序列和启动子从转录起点向上延伸至 1400bp。使用标准分子生物学技术, 这种启动子或其部分可以
5 被转移至载体如 pCGN1761, 其可以置换 35S 启动子并被用于驱动本发明的核酸序列以花粉特异性方式表达。

2. 转录终止子

许多转录终止子可用于表达盒中。这些终止子与转基因和正确的 mRNA 聚腺苷酸化之外的转录的终止相关。合适的转录终止子是已知
10 在植物中起作用的那些终止子, 包括 CaMV 35S 终止子、tml 终止子、胭脂氨酸合酶终止子及 pea rbcS E9 终止子。这些终止子既可用于单子叶植物也可用于双子叶植物中。另外, 可以使用基因天然的转录终止子。

3. 增强或调节表达的序列

15 已经发现许多序列增强基因从转录单位中的表达, 这些序列可与本发明的基因联合应用以增加其在转基因植物中的表达。

许多内含子序列已经示出增强表达, 特别是在单子叶植物细胞中。例如, 已经发现玉米 AdhI 基因的内含子在其同源启动子导入玉米细胞中时显著增强野生型基因的表达。发现内含子 1 在具有氯霉素
20 乙酰转移酶基因的融合构建体中特别有效并增强表达 (Callis et al., Genes Develop. 1: 1183-1200 (1987))。在相同的实验系统中, 玉米 bronze1 基因的内含子具有相似的增强表达作用。已经将内含子序列常规掺入植物转化载体中, 特别是非翻译的前导序列中。

也已知衍生自病毒的许多非翻译的前导序列增强表达, 这些序列
25 在双子叶植物细胞中特别有效。具体地, 来自烟草花叶病毒(TMV, W-序列)、玉米 Chlorotic Mottle 病毒(MCMV)及苜蓿花叶病毒(AMV)的前导序列已经示出对于增强表达有效(例如 Gallie et al. Nucl. Acids

Res. 15: 8693-8711 (1987); Skuzeski et al. Plant Molec. Biol. 15: 65-79 (1990)). 本领域已知的其它前导序列包括但不限于: 小核糖核酸病毒前导序列, 例如 EMCV 前导序列(脑心肌炎病毒 5'非编码区) (Elroy-Stein, O., Fuerst, T. R., 及 Moss, B. PNAS USA 86:6126-6130 (1989)); 马铃薯 Y 病毒(potyvirus)前导序列, 例如 TEV 前导序列(烟草 Etch 病毒) (Allison et al., 1986); MDMV 前导序列(玉米矮小花叶病毒); Virology 154:9-20); 人免疫球蛋白重链结合蛋白(BiP) 前导序列(Macejak, D. G., and Sarnow, P., Nature 353: 90-94 (1991); 来自苜蓿花叶病毒的外壳蛋白 mRNA 的未翻译的前导序列(AMV RNA 4), (Jobling, S. A., and Gehrke, L., Nature 325:622-625 (1987); 烟草花叶病毒前导序列(TMV), (Gallie, D. R. et al., Molecular Biology of RNA, 237-256 页(1989); 及玉米 Chlorotic Mottle 病毒前导序列(MCMV) (Lommel, S. A. et al., Virology 81:382-385 (1991)。也见 Della-Cioppa et al., Plant Physiology 84:965-968 (1987)所述。

除了向本发明的靶表达盒的 5'调节区中掺入一或多种前述元件之外, 也可以掺入靶表达盒特有的其它元件。这种元件包括但不限于最小启动子。最小启动子是指无上游活化而失活或接近失活的基本启动子元件。当不存在反式激活物(transactivator)或者当增强子或应答元件结合位点不存在时, 这种启动子在植物中背景活性较低。对于植物中靶基因特别有效的一种最小启动子是 Bz1 最小启动子, 其得自玉米的 bronze1 基因。Bz1 核心启动子通过在第-53 至-58 位切割而得自 myc 突变体 Bz1-荧光素酶构建体 pBz1LucR98, Roth et al., Plant Cell 3: 317 (1991)。因此衍生的 Bz1 核心启动子片段是从-53 延伸至+227, 在 5'未翻译区中包含 Bz1 内含子-1。对于本发明也有效的是通过使用合成的 TATA 元件产生的最小启动子。TATA 元件使得所述启动子可以由 RNA 聚合酶识别, 并在无活化的条件下赋予基因表达的基础水

平(见 Mukumoto (1993) *Plant Mol Biol* 23: 995-1003; Green (2000) *Trends Biochem Sci* 25: 59-63)。

4. 基因产物在细胞内的靶向

已知植物中存在靶向基因产物的各种机制,控制这些机制功能的序列在一定程度上已经鉴定。例如,将基因产物靶向于叶绿体是通过各种蛋白质的氨基末端发现的信号序列控制的,所述信号序列在叶绿体输入期间被切割以产生成熟蛋白质(例如 Comai et al. *J. Biol. Chem.* 263: 15104-15109 (1988))。这些信号序列可以与异源基因产物融合,以实现异源产物输入叶绿体中(van den Broeck, et al. *Nature* 313: 358-363 (1985))。编码合适的信号序列的 DNA 可分离自编码 RUBISCO 蛋白、CAB 蛋白、EPSP 合酶、GS2 蛋白及已知位于叶绿体的许多其它蛋白质的 cDNA 的 5'末端。见美国专利 No. 5,639,949 的实施例 37 中题目为“叶绿体靶向表达”章节所述。

其它基因产物位于其它细胞器如线粒体和过氧化物酶体(例如 Unger et al. *Plant Molec. Biol.* 13: 411-418 (1989))。对编码这些产物的 cDNA 也可以进行操纵以实现异源基因产物靶向这些细胞器。这种序列例如是核编码的 ATPases 及线粒体特异性的天冬氨酸氨基转移酶同种型。靶向细胞蛋白质体已经由 Rogers et al. (*Proc. Natl. Acad. Sci. USA* 82: 6512-6516 (1985))描述。

另外,已经鉴定了导致基因产物靶向其它细胞区室的序列。氨基末端序列与靶向 ER、质外体及自糊粉细胞的细胞外分泌相关(Koehler & Ho, *Plant Cell* 2: 769-783 (1990))。另外,氨基末端序列联合羧基末端序列与基因产物的空泡靶向相关(Shinshi et al. *Plant Molec. Biol.* 14: 357-368 (1990))。

通过将上述合适的靶向序列与感兴趣的转基因序列融合,可指导转基因产物靶向任何细胞器或细胞区室。例如对于叶绿体靶向而言,将来自 RUBISCO 基因、CAB 基因、EPSP 合酶基因、或者 GS2 基因

的叶绿体信号序列在读框内与转基因的氨基末端 ATG 融合。选择的信号序列应包括已知切割位点，融合构建体应考虑到切割所需的在切割位点之后的任何氨基酸。在一些情况中，这种需要可通过在切割位点与转基因 ATG 之间加入少量氨基酸，或者用转基因序列置换一些氨基酸而实现。为进行叶绿体输入而构建的融合体可以通过体外转录的构建体的体外翻译及随后的体外叶绿体摄取而测试其叶绿体摄取效力，这使用 Bartlett et al. In: Edelman et al. (Eds.) *Methods in Chloroplast Molecular Biology*, Elsevier pp 1081-1091 (1982) 及 Wasmann et al. *Mol. Gen. Genet.* 205: 446-453 (1986)所述技术进行。这些构建技术为本领域所熟知，可等同地用于线粒体和过氧化物酶体。

上述细胞靶向机制不仅可与其同源启动子联合应用，也可以与异源启动子联合应用，以在启动子的转录调节下实现特异性细胞靶向的目的，所述启动子的表达方式不同于靶向信号衍生自其中的启动子的表达方式。

15

C. 植物转化载体的构建

植物转化领域技术人员已知许多转化载体可用于进行植物转化，本发明的基因可与任何这种载体联合应用。载体的选择依赖于具体的转化技术和用于转化的具体靶物种。对于某些靶物种而言，不同的抗生素或除草剂选择标记可以是特异性的。转化方法中常用的选择标记包括 nptII 基因，其赋予卡那霉素和相关抗生素抗性 (Messing & Vierra. *Gene* 19: 259-268 (1982); Bevan et al., *Nature* 304:184-187 (1983)), bar 基因，其赋予除草剂抗草丁膦(phosphinothricin)抗性 (White et al., *Nucl. Acids Res* 18: 1062 (1990), Spencer et al. *Theor. Appl. Genet* 79: 625-631 (1990)), hph 基因，其赋予抗生素潮霉素抗性 (Blochinger & Diggelmann, *Mol Cell Biol* 4: 2929-2931), 及 dhfr 基因，其赋予氨甲喋呤抗性 (Bourouis et al., *EMBO J.* 2(7): 1099-1104

(1983)), EPSPS 基因, 其赋予草甘膦(glyphosate)抗性(U.S. Patent Nos. 4,940,935 和 5,188,642), 及甘露糖-6-磷酸异构酶基因, 其提供使甘露糖代谢的能力(U.S. Patent Nos. 5,767,378 和 5,994,629)。

1. 适于农杆菌(*Agrobacterium*)转化的载体

5 许多载体可用于转化根癌农杆菌(*Agrobacterium tumefaciens*)。这些载体典型携带至少一个 T-DNA 边界序列, 包括载体如 pBIN19 (Bevan, Nucl. Acids Res. (1984))。下文描述了适于农杆菌转化的两种典型的载体的构建。

a. pCIB200 和 pCIB2001

10 二元载体 pCIB200 和 pCIB2001 用于构建与农杆菌一起使用的重组载体, 以如下方式构建。pTJS75kan 是通过用 *NarI* 消化 pTJS75 (Schmidhauser & Helinski, J. Bacteriol. 164: 446-455 (1985)), 使得四环素抗性基因切除, 随后插入携带 NPTII 的 pUC4K 的 *AccI* 片段(Messing & Vierra, Gene 19: 259-268 (1982); Bevan et al., Nature 304: 184-187
15 (1983); McBride et al., Plant Molecular Biology 14: 266-276 (1990))而产生的。将 *XhoI* 接头与含有左侧和右侧 T-DNA 边界的 PCIB7 的 *EcoRV* 片段连接, 将植物可选择的 nos/nptII 嵌合基因及 pUC 多接头 (Rothstein et al., Gene 53: 153-161 (1987))和 *XhoI*-消化的片段克隆进 *Sall*-消化的 pTJS75kan 中产生 pCIB200 (也见 EP 0 332 104, 实施例
20 19)。pCIB200 含有如下独特的多接头限制位点: *EcoRI*、*SstI*、*KpnI*、*BglII*、*XbaI* 和 *Sall*。pCIB2001 是 pCIB200 的衍生物, 通过插入额外的限制位点的多接头而产生。pCIB2001 的多接头中的独特限制位点是 *EcoRI*、*SstI*、*KpnI*、*BglII*、*XbaI*、*Sall*、*MluI*、*BclI*、*AvrII*、*ApaI*、*HpaI* 和 *StuI*。pCIB2001 除了含有这些独特的限制位点之外, 还具有
25 植物和细菌卡那霉素抗性选择, 转化的左侧和右侧 T-DNA 边界用于进行农杆菌介导, RK2-衍生的 *trfA* 功能以在大肠杆菌与其它宿主之间移动, 及也来自 RK2 的 *OriT* 和 *OriV* 功能。pCIB2001 多接头适于

克隆含有其自身调节信号的植物表达盒。

b. pCIB10 及其潮霉素选择衍生物

二元载体 pCIB10 含有编码卡那霉素抗性以在植物中选择的基因及 T-DNA 右侧和左侧边界序列，并掺入来自广泛宿主范围的质粒 pRK252 的序列，使其在大肠杆菌和农杆菌中均表达。其构建由 Rothstein et al. (Gene 53: 153-161 (1987))描述。构建 pCIB10 的各种衍生物，其掺入潮霉素 B 磷酸转移酶基因，如 Gritz et al. (Gene 25: 179-188 (1983))所述。这些衍生物使得可以仅基于潮霉素(pCIB743)或者基于潮霉素和卡那霉素(pCIB715, pCIB717)而选择转基因植物。

10 2. 适于非农杆菌转化的载体

不使用根癌农杆菌的转化不需要所选择转化载体中的 T-DNA 序列，因此除了如上述含有 T-DNA 序列的载体之外也可以利用没有这些序列的载体。不依赖于农杆菌的转化技术包括通过粒子轰击、原生质体摄取(例如 PEG 和电穿孔)及显微注射。载体的选择很大程度上依赖于转化的物种。下文描述了适于非农杆菌转化的典型载体的构建。

a. pCIB3064

pCIB3064 是适于指导基因转移技术以及通过除草剂 basta(或者抗草丁膦)的选择的 pUC-衍生的载体。质粒 pCIB246 包含与大肠杆菌 GUS 基因可操作地融合的 CaMV 35S 启动子及 CaMV 35S 转录终止子，所述质粒在 PCT 公开 WO 93/07278 中描述。这个载体的 35S 启动子含有起始位点的两个 5'ATG 序列。使用标准 PCR 技术对这些位点进行突变，由此除去 ATG 并产生限制位点 SspI 和 PvuII。新的限制位点与独特的 SalI 的位点相隔 96 和 37 bp，与实际的起始位点相隔 101 和 42 bp。所得 pCIB246 衍生物称作 pCIB3025。然后将 GUS 基因通过用 SalI 和 SacI 消化而从 pCIB3025 中切下，将末端平末端化并再连接以产生质粒 pCIB3060。质粒 pJIT82 得自 John Innes Centre, Norwich, 从绿色产色链霉菌(*Streptomyces viridochromogenes*)中切下

含有 bar 基因的 400 bp SmaI 片段, 插入 pCIB3060 的 HpaI 位点中 (Thompson et al. EMBO J 6: 2519-2523 (1987))。这样产生 pCIB3064, 其包含在 CaMV 35S 启动子控制下的 bar 基因及进行除草剂选择的终止子, 氨苄青霉素抗性基因(在大肠杆菌中选择)和具有独特位点

5 SphI、PstI、HindIII 和 BamHI 的多接头。这种载体适于克隆含有其自身调节信号的植物表达盒。

b. pSOG19 和 pSOG35

pSOG35 是利用赋予氨甲喋呤抗性的大肠杆菌基因二氢叶酸还原酶(DFR)作为可选择标记的转化载体。使用 PCR 扩增 35S 启动子(-800

10 bp), 玉米 Adh1 基因的内含子 6(-550 bp)和来自 pSOG10 的 18 bp 的 GUS 未翻译的前导序列。将编码大肠杆菌二氢叶酸还原酶 II 型基因的 250-bp 片段也通过 PCR 扩增, 将这两个 PCR 片段与包含 pUC19 载体主链和胭脂氨酸合酶终止子的 pB1221 (Clontech)的 SacI-PstI 片段装配在一起。这些片段的装配产生 pSOG19, 其含有与内含子 6 序

15 列融合的 35S 启动子, GUS 前导序列, DHFR 基因和胭脂氨酸合酶终止子。pSOG19 中 GUS 前导序列由玉米褪绿斑驳病毒(Chlorotic MottleVirus)(MCMV)的前导序列置换, 产生载体 pSOG35。pSOG19 和 pSOG35 携带氨苄青霉素抗性 pUC 基因并具有可用于克隆外源物质的 HindIII、SphI、PstI 和 EcoRI 位点。

20 3. 适于叶绿体转化的载体

为了在植物质体中表达本发明的核苷酸序列, 使用质体转化载体 pPH143 (WO 97/32011, 实施例 36)。将所述核苷酸序列插入 pPH143 中, 从而置换 PROTOX 编码序列。然后将这个载体用于质体转化和壮观霉素抗性转化体的选择。或者, 将所述核苷酸序列插入 pPH143

25 中, 由此其取代 aadH 基因。在这种情况下, 针对 PROTOX 抑物质抗性选择转化体。

D. 转化

一旦已经将本发明的核酸序列克隆进表达系统中, 则其被转化进植物相比中。本发明的受体和靶表达盒可以本领域公认的众多方式导入植物细胞中。植物的再生方法也为本领域所熟知。例如, 利用 Ti 质粒输送外源 DNA, 以及直接 DNA 摄取, 脂质体, 电穿孔, 显微注射和显微发射。另外, 可利用细菌转化植物细胞。在下文描述了转化双子叶植物和单子叶植物的代表性技术以及代表性的质体转化技术。

1. 双子叶植物的转化

双子叶植物的转化技术为本领域所熟知, 包括基于农杆菌的技术及不需要农杆菌的技术。非农杆菌技术包括由原生质体或细胞直接摄取外源遗传物质。这种技术可通过 PEG 或者电穿孔介导的摄取、粒子轰击介导的输送或者显微注射而实现。这些技术例如由 Paszkowski et al., EMBO J 3: 2717-2722 (1984)、Potrykus et al., Mol. Gen. Genet. 199: 169-177 (1985)、Reich et al., Biotechnology 4: 1001-1004 (1986)及 Klein et al., Nature 327: 70-73 (1987)所描述。在每种情况中, 使用本领域已知的标准技术将转化的细胞再生为完整植物。

农杆菌介导的转化是转化双子叶植物的具体技术, 因为其转化效力高并可用于许多不同物种的广泛效用。农杆菌转化典型包括将携带感兴趣的外源 DNA 的二元载体(例如 pCIB200 或 pCIB2001)转移至合适的农杆菌菌株中, 这依赖于宿主农杆菌菌株在 Ti 质粒或染色体上携带的 vir 基因的互补体(例如针对 pCIB200 和 pCIB2001 的菌株 CIB542 (Uknes et al. Plant Cell 5: 159-169 (1993))). 重组二元载体转移至农杆菌中是通过三亲株杂交方法实现的, 其中使用携带重组二元载体大肠杆菌, 携带质粒如 pRK2013 并能使所述重组二元载体向靶农杆菌菌株迁移的辅助大肠杆菌菌株。或者, 可以将重组的二元载体通过 DNA 转化转移至农杆菌(Höfgen & Willmitzer, Nucl. Acids Res. 16: 9877 (1988))。

通过重组农杆菌对靶植物物种的转化通常包括将农杆菌与来自所述植物的外植体共培养，根据本领域熟知的方法进行。将转化的组织在二元质粒 T-DNA 边界之间携带抗生素或除草剂抗性标记的可选择的培养基上再生。

- 5 另一种用基因转化植物细胞的方法包括在植物组织和细胞推进惰性或生物学活性粒子。这种技术在 Sanford 的美国专利 No. 4,945,050、5,036,006 和 5,100,792 中描述。通常地，这种方法包括在有效穿透细胞外表面并使得掺入细胞内部的条件向细胞推进惰性或生物学活性粒子。当利用惰性粒子时，载体可以通过用含有希望的
- 10 基因的载体包被所述粒子而导入细胞中。或者，靶基因可以环绕在载体周围，以便通过粒子激发将所述载体带进细胞中。生物学活性粒子(例如含有被导入的 DNA 的干燥的酵母细胞、干燥的细菌或噬菌体)也可以被推进植物细胞组织中。

2. 单子叶植物的转化

- 15 大多数单子叶植物的转化目前也已经成为常规方法。具体的技术包括使用 PEG 或电穿孔技术将基因直接转移进原生质体中及通过粒子轰击进入愈伤组织组织中。可以利用单一 DNA 种类或者多个 DNA 种类(即共转化)进行转化，这两种技术均适用于本发明。共转化的优势是避免完全的载体构建及产生具有未连接的感兴趣基因的基因座
- 20 及选择标记的转基因植物，使得可以在随后的世代中除去选择标记，这也该是希望的。然而，使用共转化的缺点是分离的 DNA 物种不是 100%地整合进基因组中(Schocher et al. Biotechnology 4: 1093-1096 (1986))。

- 25 专利申请 EP 0 292 435、EP 0 392 225 和 WO 93/07278 描述了从原种近交系玉米中制备愈伤组织组织和原生质体、使用 PEG 或电穿孔转化原生质体、及从转化的原生质体中再生玉米植物的技术。Gordon-Kamm et al. (Plant Cell 2: 603-618 (1990))和 Fromm et al.

(Biotechnology 8: 833-839 (1990))已经公开了使用粒子轰击转化 A188-衍生的玉米品系的技术。另外, WO 93/07278 和 Koziel et al. (Biotechnology 11: 194-200 (1993))描述了通过粒子轰击转化原种近交系玉米的技术。这种技术利用从授粉后 14—15 天的玉米穗切下的长度为 1.5—2.5 mm 的未成熟的玉米胚及 PDS-1000He Biolistics 设备进行轰击。

水稻的转化也可以利用原生质体或粒子轰击的直接基因转移技术。本领域已经有了关于日本产和印度产植物类型进行原生质体介导的转化的描述(Zhang et al. Plant Cell Rep 7: 379-384 (1988); Shimamoto et al. Nature 338: 274-277 (1989); Datta et al. Biotechnology 8: 736-740 (1990))。这两种类型也可使用粒子轰击常规转化(Christou et al. Biotechnology 9: 957-962 (1991))。另外, WO 93/21335 描述了通过电穿孔转化水稻的技术。

专利申请 EP 0 332 581 描述了早熟禾亚科(Pooideae)原生质体的产生、转化和再生的技术。这些技术可以转化鸭茅(Dactylis)和小麦。另外,小麦的转化已经由 Vasil et al. (Biotechnology 10: 667-674 (1992))描述,其中使用粒子轰击进入 C 型长期可再生的愈伤组织中,也在 Vasil et al. (Biotechnology 11: 1553-1558 (1993))和 Weeks et al. (Plant Physiol. 102: 1077-1084 (1993))中所述,其中使用粒子轰击未成熟胚和未成熟胚衍生的愈伤组织。然而,对于小麦转化的具体技术包括通过用粒子轰击不成熟的胚对小麦进行转化,并包括在基因递送之前的高蔗糖或高麦芽糖步骤。在轰击之前,将任意数目的胚(长度为 0.75—1 mm)铺板于具有 3%蔗糖和 3 mg/l 2,4-D 的 MS 培养基(Murashiga & Skoog, Physiologia Plantarum 15: 473-497 (1962))上诱导体细胞胚,所述过程可在黑暗中进行。在选择进行攻击当天,将胚从诱导培养基中取出,置于渗透剂(即具有希望浓度的蔗糖或麦芽糖的诱导培养基,所述浓度通常为 15%)上。允许所述胚质壁分离 2—3 小时,然后进行轰击。尽管不是关键的,但是通常为每个平板 20 个胚。使用标准技

术将合适的携带基因的质粒(如 pCIB3064 或 pSG35)在微米大小的金颗粒上沉淀。每个平板的胚用 DuPont Biolistics® 氦设备射击, 使用爆破压力为~1000 psi, 标准 80 网孔。在轰击后, 将胚置于黑暗处恢复大约 24 小时(仍在渗透剂上)。24 小时后, 将胚从渗透剂上取出, 置于诱导培养基上, 保持大约 1 个月, 之后进行再生。大约一个月后, 将具有发育中的胚发生愈伤组织的胚外植体转移至含有合适的选择剂(在 pCIB3064 情况中为 10 mg/l basta, 在 pSOG35 情况中为 2 mg/l 氨基嘌呤)的再生培养基(MS + 1 mg/L NAA, 5 mg/L GA)中。大约 1 个月后, 将发育的芽转移至较大的无菌容器“GA7s”中, 该容器中含有 1/2(half-strength)MS、2%蔗糖及相同浓度的选择剂。

使用农杆菌的单子叶植物的转化也已经有描述。见 WO 94/00977 及美国专利 No. 5,591,616 所述, 这两篇文献在此均并入作参考。也见并入作参考的 Negrotto et al., Plant Cell Reports 19: 798-803 (2000)所述。例如, 水稻(*Oryza sativa*)用于产生转基因植物。可以使用各种水稻栽培种(Hiei et al., 1994, Plant Journal 6:271-282; Dong et al., 1996, Molecular Breeding 2:267-276; Hiei et al., 1997, Plant Molecular Biology, 35:205-218)。同样, 下文描述的各种培养基成分可以在数量上改变或代替。起始胚发生应答和/或通过 MS-CIM 培养基(MS 基础盐, 4.3 g/L; B5 维生素(200 x), 5 ml/L; 蔗糖, 30 g/L; 脯氨酸, 500 mg/L; 谷氨酰胺, 500 mg/L; 酪蛋白水解产物, 300 mg/L; 2,4-D (1 mg/ml), 2 ml/L; 用 1 N KOH 将 pH 调节为 5.8; Phytigel, 3 g/L)上培养而从成熟胚中建立培养物。接种培养应答初始阶段或者建立的培养系的成熟胚并与含有希望的载体构建体的根癌农杆菌菌株 LBA4404 (农杆菌)共培养。将获自甘油原种的农杆菌在固体 YPC 培养基(100 mg/L 壮观霉素及任何其它合适的抗生素)上在 28℃ 培养大约 2 天。将农杆菌再悬浮于液体 MS-CIM 培养基中。将农杆菌培养物稀释为 OD600 为 0.2—0.3, 加入乙酰丁香酮至终浓度为 200 μ M。加入乙酰丁香酮之后, 将所述溶液与水稻培养物混合以诱导农杆菌将 DNA 转移至植物细胞。为了进行接种, 将植物培养物浸入细菌悬浮

液中。除去液体细菌悬浮液并将接种的培养物置于共培养基上，在 22℃ 培育两天。然后将培养物转移至具有替卡西林二钠(400 mg/L)的 MS-CIM 培养基，以抑制农杆菌生长。对于利用 PMI 选择标记基因的构建体(Reed et al., In Vitro Cell. Dev. Biol.-Plant 37:127-132)，7 天后

5 将培养物转移至含有甘露糖作为碳源的选择培养基(MS 具有 2%甘露糖，300 mg/L Ticarcillin)中，在黑暗中培养 3—4 周。然后将抗性集落转移至再生诱导培养基(MS，没有 2,4-D，具有 0.5 mg/L IAA、1 mg/L 玉米素、200 mg/L 替卡西林钠克拉维酸钾(timentin)、2%甘露糖和 3% 山梨糖醇)中，在黑暗中生长 14 天。然后将组织的集落转移至另一再

10 生诱导培养基中，转移至光亮生长室中。将再生的芽转移至具有 GA7-1 培养基(MS，无激素，具有 2%山梨糖醇)的 GA7 容器中培养 2 周，然后当其足够大并具有适当的根时转移至温室中。将植物移植在温室土壤中 (T0 世代)，生长至成熟，收获 T1 种子。

3. 质体的转化

15 使烟草(*Nicotiana tabacum*)c.v. 'Xanthi nc' 的种子在 T 琼脂培养基上以 1" 环形阵列在平板 (7 个/平板) 中发芽，在播种后 12—14 天用质粒 pPH143 和 pPH145 的 DNA (Svab, Z. and Maliga, P. (1993) PNAS 90, 913—917) 包被的 1 μm 钨粒子(M10, Biorad, Hercules, CA)轰击。将被轰击的秧苗在 T 培养基上培育 2 天，之后切下叶并置于含有

20 500 μg/ml 二盐酸壮观霉素(Sigma, St. Louis, MO)的 RMOP 培养基 (Svab, Z., Hajdukiewicz, P. and Maliga, P. (1990) PNAS 87, 8526—8530) 的平板上，环境为光亮的(350-500 μmol 光子/m²/s)，背轴侧向上。将在轰击 3—8 周后在变白的叶下面出现的抗性芽亚克隆在相同的选择培养基上，使其新成愈伤组织，分离并亚克隆二级芽。在独立的亚克

25 隆中转化的质体基因组拷贝(同源原生质性(homoplasmicity))的完全分离通过标准 Southern 印迹评定(Sambrook et al., (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor)。将 BamHI/EcoRI-消化的全部细胞 DNA (Mettler, I. J.

(1987) Plant Mol Biol Reporter 5, 346–349)在 1% Tris-硼酸盐(TBE)琼脂糖凝胶上分离, 转移至尼龙膜 (Amersham)上, 用 ^{32}P -标记的随机引发的 DNA 序列探查, 所述 DNA 序列相应于含有部分质体靶向序列的 pC8 的 0.7 kb BamHI/HindIII DNA 片段。将同源原生质型的芽在
5 含有壮观霉素的 MS/IBA 培养基(McBride, K. E. et al. (1994) PNAS 91, 7301-7305)上无菌生根并转移至温室中。

V. 育种和种子产生

A. 育种

10 通过用本发明的核酸序列转化而获得的植物可以是任何植物物种, 包括单子叶植物和双子叶植物; 然而, 本发明方法中使用的植物特别选自前述农业学重要的靶作物。本发明基因的表达与对于生产和质量重要的其它特征的组合可以通过育种而掺入植物品系中。育种方法和技术为本领域所已知。见例如 Welsh J. R., Fundamentals of Plant
15 Genetics and Breeding, John Wiley & Sons, NY (1981); Crop Breeding, Wood D. R. (Ed.) American Society of Agronomy Madison, Wisconsin (1983); Mayo O., The Theory of Plant Breeding, Second Edition, Clarendon Press, Oxford (1987); Singh, D.P., Breeding for Resistance to Diseases and Insect Pests, Springer-Verlag, NY (1986); 及 Wricke and
20 Weber, Quantitative Genetics and Selection Plant Breeding, Walter de Gruyter and Co., Berlin (1986)所述。

上述工程化进入转基因种子和植物中的遗传性质是通过有性生殖或者无性生长而传递的, 并因此可以在后代植物中保持和增殖。通常所述保持和增殖是利用符合特定目的如耕种、播种或收获的已知农
25 业学方法。也可以应用专门的方法如水培法或温室技术。由于生长中的作物易受由昆虫或感染的攻击和损害以及与杂草植物竞争, 因此要采取措施以控制杂草、植物疾病、昆虫、线虫及其它有害情形以改良

产量。这些措施包括机械措施如耕耘土壤或除去杂草和受到感染的植物，以及应用农用化学品如除草剂、杀真菌剂、杀配子剂、杀线虫剂、生长调节剂、催熟剂和杀虫剂。

5 本发明的转基因植物和种子的有利的遗传性质可进一步用于植物育种中，目标是产生具有改良性质的植物，所述改良的性质例如是害虫、除草剂或应激抗性，改良的营养值，增加的产量，或者使得倒伏或脱粒减少的经改良的结构。各种育种步骤的特征在于非常明确的人工干预，如选择杂交的品系、指导亲代品系的授粉，或者选择合适的后代植物。根据希望的性质，采取不同的育种措施。本领域熟知相关的技术，包括但不限于杂交、近交、回交、多品系育种、品种混合、
10 种间杂交、非整倍体技术等。杂交技术还包括植物的不育，通过机械、化学或者生物化学方式产生雄性或雌性不育植物。雄性不育植物与不同品系的花粉交叉授粉保证了雄性不育而非雌性不育植物的基因组均一地获得这两个亲代品系的性质。因此，本发明的转基因种子和植物可用于育种改良的植物品系，例如增加常规方法如除草剂或杀虫剂
15 处理的效力，或者由于其修饰的遗传性质而可以省却所述方法。或者，可以获得改良的应激耐受性的新作物，由于其优化的遗传“装备”而产生与未能耐受相似不利发育条件的产物相比具有更好的品质的收获物。

20 B. 种子产生

在种子的产生中，种子的发芽质量和均匀性是基本的产物特征。由于难以保持农作物没有其它作物和杂草的种子，难以控制种子传播的疾病，及产生具有良好发芽性能的种子，在纯种子的生长、条件和销售领域均富有经验的种子生产者已经相当广泛和充分地进行了种子
25 生产实践。因此，农场主通常是购买经鉴定符合特定质量标准的种子，而不是使用从其自己的作物中收获的种子。用作种子的繁殖材料通常被处理为具有包含除草剂、杀昆虫剂、杀真菌剂、杀菌剂、杀线

虫剂、杀软体动物剂或其混合物的杀虫剂涂层。通常使用的杀虫剂涂层包含化合物如克菌丹(captan)、萎锈灵(carboxin)、双硫胺甲酰(thiram)(TMTD®)、methalaxyl (Apron®) 和 pirimiphos-methyl (Actellic®)。如果需要, 这些化合物与配制领域常用的其它载体、表面活性剂或者促进应用的佐剂一起配制, 以提供对抗由细菌、真菌或动物害虫导致的损害的保护作用。所述杀虫剂涂层可以通过将繁殖材料浸入液体配制品中或者通过用组合的湿或干的配制品涂层而施加。其它施加方法也可以是例如针对芽或果实处理。

10 VI. 核酸分子表达的改变

本发明的核酸分子表达的改变通过如下的一种方式实现:

A. “有义”抑制

本发明的核苷酸序列表达的改变、特别是其表达降低是通过“有义”抑制获得的(参见例如 Jorgensen et al. (1996) Plant Mol. Biol. 31, 957-973)。在这种情况下, 将本发明的全部或部分核苷酸序列包含于 DNA 分子中。将所述 DNA 分子与在包含靶基因的细胞、特别是植物细胞中起作用的启动子特异性可操作地连接, 并导入细胞中, 在其中所述核苷酸序列可被表达。将所述核苷酸序列以“有义方向”插入 DNA 分子中意味着所述核苷酸序列的编码链可以被转录。在一个具体的实施方案中, 所述核苷酸序列是可充分翻译的, 所述核苷酸序列或其一部分中包含的所有遗传信息均翻译成多肽。在另一个具体的实施方案中, 所述核苷酸序列是可部分翻译的, 被翻译成较短的肽。在一个具体的实施方案中, 这可通过将至少一个不成熟的终止密码子插入所述核苷酸序列中, 其使得翻译停止来实现。在另一个更具体的实施方案中, 所述核苷酸序列被转录但无翻译产物产生。这通常通过除去由所述核苷酸序列编码的多肽的起始密码子例如 ATG 而实现。在另一个具体的实施方案中, 将包含所述核苷酸序列或其一部分的

DNA 分子稳定整合进植物细胞的基因组中。在另一个具体的实施方案中，将包含所述核苷酸序列或其一部分的 DNA 分子包含于染色体外复制型分子中。

在包含上述 DNA 分子之一的转基因植物中，相应于所述 DNA
5 分子包含的核苷酸序列的核苷酸序列的表达被特异性降低。具体地，所述 DNA 分子中的核苷酸序列与其表达被降低的核苷酸序列至少 70 % 相同，更具体地至少 80%、90%、95% 相同，更特别至少 99% 相同。

B. “反义”抑制

在另一个具体的实施方案中，本发明的核苷酸序列的表达的改
10 变、特别是其表达降低通过“反义”抑制获得的。将本发明的全部或部分核苷酸序列包含在 DNA 分子中。将所述 DNA 分子与在植物细胞中起作用的启动子特异性可操作地连接，并导入细胞中，在其中所述核苷酸序列可被表达。核苷酸序列以“反义方向”插入 DNA 分子中，意味着所述核苷酸序列的反向互补序列(有时也称作非编码链)可
15 被转录。在一个具体的实施方案中，将包含所述核苷酸序列或其一部分的 DNA 分子稳定整合在植物细胞的基因组中。在另一个具体的实施方案中，将包含所述核苷酸序列或其一部分的 DNA 分子包含在染色体外复制分子中。本发明引用了描述这种方法的一些出版物一进一步例证(Green, P. J. et al., Ann. Rev. Biochem. 55:569-597 (1986); van
20 der Krol, A. R. et al, Antisense Nuc. Acids & Proteins, pp. 125-141 (1991); Abel, P. P. et al., PNAS Proc. Natl. Acad. Sci. USA 86:6949-6952 (1989); Ecker, J. R. et al., Proc. Natl. Acad. Sci. USANAS 83:5372-5376 (Aug. 1986))。

在含有上述 DNA 分子之一的转基因植物中，相应于所述 DNA
25 分子中包含的核苷酸序列的核苷酸序列的表达被特异性降低。具体地，所述 DNA 分子中的核酸序列与其表达被降低的核苷酸序列至少 70 % 相同，更具体地至少 80%、90%、95% 相同，更具体地至少 99

%相同。

C. 同源重组

在另一个具体的实施方案中,将植物基因组中相应于本发明核苷酸序列的至少一个基因组拷贝通过同源重组进行修饰,如 Paszkowski
5 et al., EMBO Journal 7:4021-26 (1988)所进一步例证。这种技术使用同源序列彼此识别及通过本领域称作同源重组的技术彼此交换核苷酸序列的性质。同源重组可发生在细胞中核苷酸序列的染色体拷贝与通过转化导入细胞中的核苷酸序列的引入拷贝之间。因此在核苷酸序列的染色体拷贝中正确产生特异性修饰。在一个实施方案中,本发明的
10 核苷酸序列的调节元件被修饰。这种调节元件易于通过使用本发明的核苷酸序列或其一部分作为探针筛选基因组文库而获得。将现有的调节元件由不同的调节元件置换,因此改变所述核苷酸序列的表达,或者它们被突变或缺失,因此消除所述核苷酸序列的表达。在另一个实施方案中,所述核苷酸序列通过缺失一部分核苷酸序列或者全部的核苷酸序列或者通过突变而被修饰。植物细胞中突变的多肽的表达也涵盖在本发明中。这种破坏内源植物基因的技术的最新改进在(Kempin
15 et al., Nature 389:802-803 (1997)和 Miao and Lam, Plant J., 7:359-365 (1995)中描述。

在另一个具体的实施方案中,核苷酸序列的染色体拷贝中的突
20 变是通过用嵌合的寡核苷酸转化细胞而产生的,所述嵌合的寡核苷酸包含双链体构象中 RNA 和 DNA 残基的连续片段,其中在末端上具有双发夹帽。所述寡核苷酸的一个额外特征是例如在 RNA 残基存在 2'-O-甲基化。设计所述 RNA/DNA 序列,使其与本发明核苷酸序列的染色体拷贝的序列排成一行,并含有希望的核苷酸改变。例如,
25 这种技术在美国专利 No.5,501,967 和 Zhu et al. (1999) Proc. Natl. Acad. Sci. USA 96: 8768-8773 中进一步例证。

D. 核酶

在另一个实施方案中，编码本发明多肽的RNA通过特异于这种RNA的催化RNA或者核酶裂解。将所述核酶在转基因植物中表达，导致植物细胞中编码本发明多肽的RNA数量减少，由此导致在细胞中积聚的多肽的数量减少。这种方法在美国专利No.4,987,071中进一步例证。

E. 显性-阴性突变体

在另一个具体的实施方案中，改变由本发明的核苷酸序列编码的多肽的活性。这个目的通过将所述蛋白质的显性阴性突变体在转基因植物中表达，导致内源蛋白质的活性丧失而实现。

F. 适体 (Aptamers)

在另一个实施方案中，本发明的多肽的活性通过在转基因植物中表达称作适体的与蛋白质特异性结合的核酸配体而被抑制。适体优选通过 SELEX (Systematic Evolution of Ligands by EXponential Enrichment)方法获得。在 SELEX 方法中，将具有随机序列区域的单链核酸的候选混合物与蛋白质接触，从候选混合物的剩余物中分离与靶位亲和性增加的那些核酸。将分离出的核酸扩增，产生配体富集的混合物。在几次反复之后，获得与所述多肽具有最佳亲和性的核酸，将其用于在转基因植物中表达。这种方法在美国专利 No.5,270,163 中进一步举例说明。

G. 锌指蛋白

结合本发明的核苷酸序列或其调节区的锌指蛋白也用于改变所述核苷酸序列的表达。具体地，所述核苷酸序列的转录被降低或增加。锌指蛋白例如在 Beerli et al. (1998) PNAS 95:14628-14633，或者 WO 95/19431, WO 98/54311，或者 WO 96/06166 中描述，所有文献在此均以其全文并入作参考。

H. dsRNA

本发明的核苷酸序列表达的改变也通过 dsRNA 干扰而获得，例

如 WO 99/32619、WO 99/53050 或 WO 99/61631 所述，所述文献在此均以其全文并入作参考。在另一个具体的实施方案中，本发明的核苷酸序列表达的改变，特别是其表达降低，这是通过双链 RNA (dsRNA) 干扰获得的。将本发明的核苷酸序列的全部或者一部分包含在 DNA 分子中。所述 DNA 分子的大小为 100—1000 个核苷酸或更多；根据经验确定最佳大小。连接相同 DNA 分子的两个拷贝，以间隔 DNA 分子分隔，由此所述第一个和第二个拷贝是相反方向。在这个具体的实施方案中，所述 DNA 分子的第一个拷贝是反向互补序列 (也称作非编码区)，第二个拷贝是编码链；在最具体的实施方案中，第一个拷贝是编码链，第二个拷贝是反向互补序列。间隔 DNA 分子的大小 (长度) 为 200—10,000 个核苷酸，更具体是 400—5000 个核苷酸，更具体是 600—1500 个核苷酸。所述间隔是随机 DNA 片段，特别是与进行 dsRNA 干扰的靶生物体不同源的随机 DNA 片段，更具体是由所述靶生物体有效剪接的功能内含子。将由所述间隔分隔的 DNA 分子的两个拷贝与在植物细胞中起作用的启动子可操作地连接，并导入植物细胞中，在其中所述核苷酸序列可被表达。在一个具体的实施方案中，将包含所述核苷酸序列或其一部分的 DNA 分子稳定整合在植物细胞的基因组中。在另一个具体的实施方案中，将包含所述核苷酸序列或其一部分的 DNA 分子包含在染色体外复制分子中。本发明引用了一些描述这种方法的出版物以进一步例证 (Waterhouse et al. (1998) PNAS 95:13959-13964 ; Chuang and Meyerowitz (2000) PNAS 97:4985-4990; Smith et al. (2000) Nature 407:319-320)。通过 dsRNA 干扰对核苷酸序列表达的改变也在例如 WO 99/32619、WO 99/53050 或 WO 99/61631 中描述，所述文献在此以其全文并入作参考。

在含有上述 DNA 分子之一的转基因植物中，相应于所述 DNA 分子中包含的核苷酸序列的核苷酸序列的表达被特别降低。具体地，所

述DNA分子中的核苷酸序列与其表达被降低的核苷酸序列至少70%相同,更具体地至少80%、90%、95%及更具体地至少99%相同。

I. DNA分子的插入(插入诱变)

在另一个具体的实施方案中,将DNA分子插入本发明核苷酸序列的染色体拷贝中,或者插入其调节区域中。具体地,这种DNA分子包含能在植物细胞中转座的可换位的元件,例如Ac/Ds、Em/Spm、增变基因。或者,所述DNA分子包含农杆菌T-DNA的T-DNA边缘。所述DNA分子也可包含重组酶或整合酶识别位点,其可用于从植物细胞的染色体中除去DNA分子的一部分。本发明也包含使用T-DNA、转座子、寡核苷酸的插入诱变方法或者本领域技术人员已知的其它方法。使用T-DNA和转座子的插入诱变方法在Winkler et al. (1989) Methods Mol. Biol. 82:129-136 和 Martienssen (1998) PNAS 95:2021-2026 中描述,所述文献在此以其全文并入作参考。

J. 缺失诱变

在另一个实施方案中,本发明核酸分子的突变是通过在细胞或植物中在基因组序列拷贝中缺失一部分核苷酸序列或调节序列而产生。缺失诱变的方法为本领域技术人员所已知。见例如Miao et al, (1995) Plant J. 7:359。

在另一个实施方案中,这种缺失是在大量植物中通过化学诱变或照射随机产生的,在本发明的基因中具有缺失的植物通过正向或反求遗传学方法分离。已知用快中子或 γ 射线照射在植物中导致缺失突变(Silverstone et al, (1998) Plant Cell, 10:155-169; Bruggemann et al., (1996) Plant J., 10:755-760; Redei and Koncz in Methods in 鼠耳芥 Research, World Scientific Press (1992), pp. 16-82)。本发明基因中的缺失突变可以在反向遗传学策略中使用多组基因组DNA进行PCR而恢复,如已经在C. elegans 中示出(Liu et al., (1999), Genome Research,

9:859-867.)。正向遗传学策略包括诱变显示 PTGS 的品系,随后筛选没有 PTGS 的 M2 后代。在这些突变体中,预期有一些破坏本发明基因的突变体。这可以通过对具有这些突变体的基因组 DNA 的本发明的基因进行 Southern 印迹或 PCR 而评定。

5 K. 在植物细胞中过表达

在另一个具体的实施方案中,本发明的编码多肽的核苷酸序列被过表达。核酸分子及过表达本发明的核酸分子的和表达盒例如在上文描述。本领域技术人员已知的过表达核酸分子的方法也涵盖在本发明中。

10 在一个具体的实施方案中,本发明的核苷酸序列的表达在植物的每一个细胞中均改变。这个结果例如通过同源重组或者通过插入染色体中而获得。这也可以通过在能在植物的每个细胞中表达有义或反义 RNA、锌指蛋白或核酶的启动子的控制下表达有义或反义 RNA、锌指蛋白或者核酶而获得。组成型表达、诱导型的、组织特异性或者发育调节的表达也在本发明范围内,导致本发明的核苷酸序列在植物细胞中的表达发生组成型、诱导型的、组织特异性或发育调节性的改变。根据本发明如前述的教导制备表达有义或反义 RNA、锌指蛋白或核酶或者过表达本发明的核苷酸序列的构建体,并转化进植物细胞中。

20 VII. 多肽

本发明进一步涉及包含 SEQ ID NO:2 所示氨基酸序列的分离的多肽。具体地,分离的多肽包含 SEQ ID NO:2 所示氨基酸序列,以及具有保守氨基酸修饰的变体。本领域技术人员意识到在编码的序列中,改变、添加或缺失一个氨基酸或少量氨基酸的核酸、肽、多肽或蛋白质序列的各个取代、缺失或添加是都保守修饰,其中所述修饰导致氨基酸由化学相似的氨基酸取代。保守修饰的变体提供了与未修饰

的多肽相似的生物学活性。本领域已知列出功能相似的氨基酸的保守取代表。见Crichton (1984) Proteins, W.H. Freeman and Company。

5 在一个具体的实施方案中，与 SEQ ID NO:2 所示多肽序列、或其外显子或结构域基本相似的多肽是 SEQ ID NO:2 所示多肽序列的等位基因变体。在另一个具体的实施方案中，与 SEQ ID NO:2 所示多肽序列、或其外显子或结构域基本相似的多肽是 SEQ ID NO:2 所示多肽序列的天然存在的变体。在另一个具体的实施方案中，与 SEQ ID NO:2 所示多肽序列、或其外显子或结构域基本相似的多肽是 SEQ ID NO:2 所示多肽序列的多态性变体。

10 在另一个具体的实施方案中，基本相似的序列含有至少一个氨基酸的缺失或插入。在一个更具体的实施方案中，所述缺失或插入少于大约 10 个氨基酸。在最具体的实施方案中，所述缺失或插入少于大约 3 个氨基酸。

15 在一个具体的实施方案中，基本相似的序列编码至少一个氨基酸的取代。

本发明的实施方案还涵盖了含有如下多肽序列的分离的多肽：

- (a) SEQ ID NO:2 所示多肽序列，或其外显子或结构域；
- (b) 与(a)基本相似的多肽序列；
- (c) 多肽序列，其由与 SEQ ID NO:1 所示核苷酸序列或其外显子或结构域或者其互补序列相同或基本相似的核苷酸序列编码；
- 20 (d)多肽序列，其由能在中等严格条件下与 SEQ ID NO:1 所示核苷酸序列或者其互补序列杂交的核苷酸序列编码；或者
- (e) 上述(a)、(b)、(c)或(d)的功能片段。

25 在另一个具体的实施方案中，基本相似的多肽是 SEQ ID NO:2 所示多肽序列的等位基因变体，或其片段、结构域、重复或嵌合体。在另一个具体的实施方案中，分离的核酸包括来自多肽序列的多个区

域, 所述多肽由与 SEQ ID NO:1 所示核苷酸序列、或其片段或结构域或者其互补序列相同或基本相似的核苷酸序列编码。

在另一个具体的实施方案中, 所述多肽是 SEQ ID NO:2 所示多肽序列。在另一个具体的实施方案中, 所述多肽是功能片段或结构域。

- 5 在另一个具体的实施方案中, 所述多肽是嵌合序列, 其中所述嵌合序列可包括功能蛋白质结构域, 包括结构域、重复序列、翻译后修饰位点或者其它特征。在一个更具体的实施方案中, 所述多肽是植物多肽。在一个更具体的实施方案中, 所述植物是双子叶植物。在一个更具体的实施方案中, 所述植物是裸子植物。在一个更具体的实施方案中,
- 10 所述植物是单子叶植物。在一个更具体的实施方案中, 所述单子叶植物是谷类植物。在一个更具体的实施方案中, 所述谷类植物可例如是小麦、大麦、稷燕麦、裸麦、稷、高粱、黑小麦、黑麦、单粒小麦、斯佩耳特小麦、双粒小麦、画眉草、买罗高粱、亚麻、gramma 草、摩擦草和类蜀黍。在另一个具体的实施方案中, 所述谷类植物是水稻。

- 15 在一个具体的实施方案中, 所述多肽在植物的特定位置或组织中表达。在一个更具体的实施方案中, 所述位置或组织例如但不限于表皮、维管组织、分生组织、形成层、皮层或髓层。在最具体的实施方案中, 所述位置或组织是叶或鞘、根、花、及发育中的胚珠或种子。在一个更具体的实施方案中, 所述位置或组织可例如是表皮、根、维
- 20 管组织、分生组织、形成层、皮层、髓层、叶和花。在一个更具体的实施方案中, 所述位置或组织是种子。

- 在一个具体的实施方案中, 由与 SEQ ID NO:1 所示核苷酸序列或其片段或结构域或其互补序列基本相似的核苷酸序列编码的多肽序列包括至少一个核苷酸的缺失或插入。在一个更具体的实施方案
- 25 中, 所述缺失或插入少于大约 30 个核苷酸。在最具体的实施方案中, 所述缺失或插入少于大约 5 个核苷酸。

在一个具体的实施方案中,由下述的核苷酸序列编码的多肽序列包括至少一个密码子中的取代,所述核苷酸序列与 SEQ ID NO:1 所示核苷酸序列或其片段或结构域或其互补序列基本相似。在一个更具体的实施方案中,所述取代是保守性的。

- 5 在一个具体的实施方案中,与 SEQ ID NO:2 所示序列基本相似的多肽序列或其片段、结构域、重复序列或嵌合序列包括至少一个氨基酸缺失或插入。

本发明的多肽、其片段或变体可包含本发明多肽的任何数目的连续氨基酸残基,其中残基的数目选自10至本发明多肽全长残基数目之间的整数的组。具体地,所述多肽的一部分或片段是功能蛋白。本发明包括活性多肽,其具有天然(非合成的)内源多肽至少20%、30%、或者40%、特别是至少50%、60%或70%、更具体是至少80%、90%或95%的的特异性活性。另外,底物特异性 (k_{cat}/K_m)任选与天然(非合成的)内源多肽基本相似。通常, K_m 为天然内源多肽的至少30%、
10 40%或50%; 更特别为至少60%、70%、80%或90%。分析和量化活性和底物特异性的方法为本领域技术人员所熟知。

本发明的分离的多肽当以免疫原呈递时,导致与本发明多肽特异性反应的抗体产生。因此,本发明的多肽可用作免疫原以构建与本发明的蛋白质免疫反应的抗体,用于但不限于免疫测定或蛋白质纯化技术。确定结合的免疫测定技术为本领域所熟知,例如但不限于ELISA
20 或竞争性免疫测定。

本发明的实施方案还涉及由本发明分离的核酸分子编码的嵌合多肽,包括含有由含有如下核苷酸序列的分离的核酸编码的多肽序列的嵌合多肽:

- 25 (a) SEQ ID NO:1 所示核苷酸序列, 或其外显子或结构域;
 (b) 与(a)基本相似的核苷酸序列;
 (c) 能与(a)杂交的核苷酸序列;

- (d) 与(a)、(b)或(c)互补的核苷酸序列;
- (e) 与(a)、(b)或(c)反向互补的核苷酸序列;
- (f) 其功能片段。

含有多肽序列的多肽,所述多肽序列由含有核苷酸序列的分离的
5 核酸、其互补体、或其反向互补序列编码,所述分离的核酸编码选自下组的多肽序列:

- (a) SEQ ID NO:2 所示多肽序列,或其结构域、重复序列或嵌合序列;
- (b) 与(a)基本相似的多肽序列;
- 10 (c) 多肽序列,其由与 SEQ ID NO:1 所示核苷酸序列相同或基本相似的核苷酸序列、或其外显子、或结构域、或者其互补序列编码;
- (d) 核苷酸序列编码的多肽序列,所述核苷酸序列能在中等严格条件下与 SEQ ID NO:1 所示核苷酸序列、或者其互补序列杂交;及(a)、(b)、(c)或(d)的功能片段;或者
- 15 (e) 其功能片段。

本发明的分离的核酸分子可用于在重组工程化的细胞如细菌、酵母、昆虫、哺乳动物或植物细胞中表达本发明的多肽。所述细胞在非自然条件(例如数量、成分、位置和/或时间)下产生所述多肽,因为它们已经发生遗传改变而如此。本领域技术人员已知有许多表达系统
20 可用于表达编码本发明蛋白质的核酸,这不用在下文详细描述。

简而言之,编码本发明多肽的分离的核酸的表达通常例如可通过将所述核酸或cDNA与启动子(组成型或可调节的启动子)可操作地连接及随后掺入表达载体中而实现。所述载体适于在原核细胞或真核细胞中复制和/或整合。常用的表达载体包含转录和翻译终止子,起始
25 序列和调节编码所述多肽的核酸分子表达的启动子。为了获得克隆的核酸分子的高水平表达,希望使用这样的表达载体,其包含强启动子以指导转录、包含核糖体结合位点以起始翻译及包含转录/翻译终止

子。本领域技术人员意识到可以对本发明的多肽进行修饰而不降低其生物学活性。可以进行一些修饰以便于本发明多肽的克隆、表达或掺入融合蛋白中。这种修饰为本领域所熟知，包括但不限于在氨基末端加入甲硫氨酸以提供起始位点，或者在任一末端放置另外的氨基酸
5 (例如聚组氨酸)以产生便于定位的纯化序列。限制位点或终止密码子也可以导入载体中。

在一个具体的实施方案中，所述表达载体包括一或多个元件，例如但不限于启动子-增强子序列，选择标记序列，复制起点，表位标记编码序列，或亲和纯化-标记编码序列。在一个更具体的实施方案
10 中，所述启动子-增强子序列可以例如是 CaMV 35S 启动子、CaMV 19S 启动子、烟草 PR-1a 启动子、遍在蛋白启动子，及菜豆蛋白启动子。在另一个实施方案中，所述启动子在植物中是可操作的，更具体是组成型或诱导型启动子。在另一个具体的实施方案中，所述选择标记序列编码抗生素抗性基因。在另一个具体的实施方案中，所述表位标记
15 序列编码 V5、肽 Phe-His-His-Thr-Thr、血凝素，或者谷胱甘肽-S-转移酶。在另一个具体的实施方案中，所述亲和纯化标记序列编码聚氨基酸序列或者多肽。在一个更具体的实施方案中，所述聚氨基酸序列是聚组氨酸。在一个更具体的实施方案中，所述多肽是壳多糖结合结构域或者谷胱甘肽-S-转移酶。在一个更具体的实施方案中，所述亲
20 和纯化标记序列包含一个蛋白内含子编码序列。

原核细胞可用作宿主细胞，例如但不限于大肠杆菌及本领域已知的那些其它微生物株系。在原核细胞中表达蛋白质的方法为本领域技术人员所熟知，可见于许多实验指导中，如见于Molecular Cloning: A Laboratory Manual, by J. Sambrook et al. (1989, Cold Spring Harbor
25 Laboratory Press)。本领域技术人员可获得控制表达的各种启动子、核糖体结合位点及操纵子，是选择标记如抗生素抗性基因。选择的载体类型是使得可以在选择的细胞类型中最佳生长和表达。

可以利用许多真核表达系统，例如但不限于酵母、昆虫细胞系、植物细胞和哺乳动物细胞。本领域熟知在酵母中异源蛋白质的表达与合成方法(见Sherman et al., *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory Press, 1982)。广泛用于真核蛋白质生产的常用的酵母株是酿酒酵母和巴斯德毕赤氏酵母，用于表达的载体、菌株和方案
5 可得自商业供应商(例如Invitrogen)。

可以用表达载体转染哺乳动物细胞系统以生产蛋白质。本领域技术人员可利用许多合适的宿主细胞系，例如但不限于HEK293、BHK21和CHO细胞系。这些细胞的表达载体可包括表达控制序列如
10 复制起点，启动子(例如CMV启动子、HSV tk启动子或磷酸甘油酸激酶(pgk)启动子)，增强子，及蛋白质加工位点如核糖体结合位点、RNA剪接位点、聚腺苷酸化位点，及转录终止序列。可用于生产蛋白质的其它动物细胞系可商购或得自如美国典型培养物保藏中心(ATCC)。

在昆虫细胞中表达蛋白质的表达载体通常衍生自SF9杆状病毒或者本领域已知的其它病毒。可利用许多合适的昆虫细胞系，包括但不限于蚊子幼虫、蚕、粘虫、蛾及果蝇细胞系。
15

本领域已知转染动物和低等真核细胞的方法。许多方法用于产生能导入DNA的真核细胞，例如但不限于：磷酸钙沉淀方法，受体细胞与含有所述DNA的细菌原生质体融合的方法，用含有所述DNA的脂质体处理受体细胞的方法，DEAE葡聚糖方法，电穿孔，基因枪(biolistics)，及将所述DNA经显微注射方法直接注入细胞中。使用本
20 领域熟知的方法培养转染的细胞(见Kuchler, R.J., *Biochemical Methods in Cell Culture and Virology*, Dowden, Hutchinson and Ross, Inc. 1997)。

一旦本发明的多肽被表达，则可以使用本领域技术人员已知的方法从细胞中分离并纯化。纯化过程可以使用 Western 印迹或者放射性免疫测定或者其它标准免疫测定技术监测。蛋白质纯化技术为本领域
25

所熟知和使用(见 R. Scopes, Protein Purification: Principles and Practice, Springer-Verlag, New York 1982; Deutscher, Guide to Protein Purification, Academic Press (1990)。本发明的实施方案提供了一种产生重组蛋白质的方法, 其中表达载体包括选自如下的一或多个元件:

- 5 启动子-增强子序列, 选择标记序列, 复制起点, 表位标记编码序列, 亲和纯化标记编码序列。在一个具体的实施方案中, 所述核酸构建体包括表位标记编码序列, 分离步骤包括使用特异于所述表位标记的抗体。在另一个具体的实施方案中, 所述核酸构建体含有聚氨基酸编码序列, 分离步骤包括使用包含聚氨基酸结合物的树脂, 特别是其中所述聚氨基酸是聚组氨酸且所述聚氨基酸结合树脂是镍琼脂糖树脂。在
- 10 另一个具体的实施方案中, 所述核酸构建体含有多肽编码序列, 分离步骤包括使用含有多肽结合物的树脂, 特别是其中所述多肽是壳多糖结合结构域, 所述树脂含有壳多糖-琼脂糖。

- 本发明的多肽可以使用本领域已知的非细胞合成方法合成。固相
- 15 合成技术由Barany and Mayfield, Solid-Phase Peptide Synthesis, pp. 3-284 in the Peptides: Analysis, Synthesis, Biology, Vol.2, Special Methods in Peptide Synthesis, Part A; Merrifield, et al., J. Am. Chem. Soc. 85:2149-56 (1963)及Stewart et al., Solid Phase Peptide Synthesis, 2nd ed. Pierce Chem. Co., Rockford, IL (1984)描述。

- 20 本发明进一步提供了一种修饰(即增加或减少)植物或其一部分中本发明多肽的浓度或成分的方法。修饰可以通过增加或降低植物中的浓度和/或成分(即本发明多肽的比率)而实现。所述方法包括向植物细胞中导入包含本发明的核酸分子或者编码At5g56860序列的核酸的表达盒, 以获得转化的植物细胞或组织, 培养所述转化的植物细胞或
- 25 组织。所述核酸分子可在组成型或可诱导型启动子的调节下。所述方法进一步包括足够长时间地诱导或抑制植物中核酸分子序列的表达, 以修饰植物或其一部分中的所述浓度和/或成分。

使用本领域技术人员已知的方法对本发明的核酸分子在其中的表达经修饰的植物或植物的一部分进行分析及选择,所述方法例如但不限于Southern印迹、DNA测序、或者使用特异于所述核酸分子的引物及检测从中产生的扩增产物的PCR分析。

- 5 通常地,所述浓度或成分相对于天然对照植物、植物的一部分或者缺乏所述表达盒的细胞增加或减少至少5%、10%、20%、30%、40%、50%、60%、70%、80%或90%。

- 10 糖是光合作用植物中许多重要过程如光合作用、碳和氮的代谢的关键调节物,这种调节是通过激活或抑制参与的基因以调节基因表达而实现的。糖控制基因表达的机制还未充分明了。本发明所述 GATA 转录因子参与调节糖感应,所述转录因子自身的表达受 N 状态改变的影响。这种基因的表达增加可以产生产量增加的植物,特别是糖信号转导途径的操纵可导致光合作用增加及氮利用增加,并且改变种子、块茎、根及其它贮藏器官的源-库关系。

15

本发明通过参考如下详细描述的实施例得以进一步描述。这些实施例只是例证了本发明,而无限制本发明之意。

实施例

- 20 本发明使用的标准重组 DNA 技术及分子克隆技术为本领域所熟知,由 J. Sambrook, et al., Molecular Cloning: A Laboratory Manual, 3d Ed., Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press (2001)、T.J. Silhavy, M.L. Berman, and L.W. Enquist, Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
25 (1984)及 Ausubel, F.M. et al., Current Protocols in Molecular Biology, New York, John Wiley and Sons Inc., (1988)、Reiter, et al., Methods in

鼠耳芥 Research, World Scientific Press (1992)及 Schultz et al., Plant Molecular Biology Manual, Kluwer Academic Publishers (1998)描述。

实验背景和方法

5 A. 确定在限制性氮条件下水稻和玉米的生长条件

在过去研究参与硝酸盐的吸收和同化作用的基因的实验中, 本发明人及其它成员利用其中存在过量或不存在硝酸盐的生长条件。在后者一种情况中, 通常将硝酸盐在没有硝酸盐的条件下生长的植物, 以了解这些及其它基因的硝酸盐调节作用。这种类型的极限处理可用于解释基因调节的一些方面, 但其不适于获得关于氮的限制的作用的更好的了解。本发明人阐述了对鼠耳芥而言其中氮限制生长的条件。这包括如Rockwool (Hirai et al., 1995 Plant Cell Physiol 36, 1331-1339)所述建立一个系统, 限定三个条件; 一个条件是其中生长为最大值; 一个条件是其中氮限制生长到最大生长水平的70-75%; 一个条件是更严重地将生长限制到最大生长水平的30—35%。氮的限制作用作为“应激”, “应激”的量通过改变硝酸盐的浓度而易于变化。本发明人通过测定硝酸盐、叶绿素(在田间条件下其通常反映氮状态, 见例如Fox RH et al 2001 Agron J. 93, 590-597; Minotti PL et al 1994 Hort Science 29, 1497-1550)、氨基酸水平、硝酸盐还原酶和谷氨酰胺合成酶活性而对生理状态下“氮状态”进行了分析, 以提供评估对突变体品系进行的研究的基础。

B. 鼠耳芥植物在氮限制条件下的表达模式实验

转录表达模式可用于同时测试大量基因的RNA水平。过去已经进行了大量这些类型的实验, 如果所述实验系统经得起检验, 则这些系统可用于在不同条件下查明 (pinpoint) 生物体的“表达状态”, 并使用这种信息推断多个过程中涉及的基因和途径。本发明人发现生长

条件越不同，则在这些条件下生长的植物之间的转录模式的差异越大，并且更难以解释哪种改变更重要。这个领域中唯一公开的全基因组图谱实验是在鼠耳芥中进行的实验，其中研究了硝酸盐的极端变化 (Wang R et al 2003 Plant Physiol. 132, 556-67)。在氮限制的情况中，本
5 发明人研究了在慢性氮应激的条件下生长植物的作用，以及可利用的氮的水平变化。本发明人已经确定了在鼠耳芥中不同水平的氮对生长的影响。

研究了不同水平的氮对转录物图谱的作用：其中氮不限制生长。对于鼠耳芥，本发明人收集了在不同氮水平条件下生长的4周龄的嫩芽。收集三种不同的生物样品(在生物学上为一式三份(biological
10 triplicates))，以获得有统计学意义的结果。使用鼠耳芥 GeneChip®全基因组阵列(Affymetrix)确定转录模式，以研究鼠耳芥中的转录水平。进行研究这些可观的实验数据所需的生物信息分析。通过研究氮的限制对表达模式的作用，本发明人可查明哪些途径参与对营养素应激的
15 应答。

实施例 1：At5g56860 的克隆和测序

从序列数据库中获得基因预测，并用于设计寡核苷酸引物以对水稻第一链 cDNA 的全长(对于转基因基因过表达包含推测的起始密码
20 子和终止密码子)或部分(对于转基因基因敲除)cDNA 克隆进行 PCR 扩增。在一些情况中，这些 PCR 引物包括另外的 5'序列以进行基于 Gateway™ 重组的克隆(Invitrogen)。使用 Advantage II (Clonetech)或 EXPAND (Roche) PCR 试剂盒，根据厂商指导进行 PCR 扩增。将 PCR 产物根据厂商(Invitrogen)说明克隆进 pCR2.1-TOPO 或 pDONR201
25 中。

根据厂商说明从 4—8 个独立克隆中小量制备 DNA。使用 BigDye™ Terminator 试剂盒(ABI)，根据厂商说明对 DNA 进行序列分

析。使用针对推测的基因的两条链设计的引物进行测序。使用 Phred/Phrap/Consed 软件包(University of Washington)对所有的序列数据进行分析并加以组合, 在共有序列水平误差率等于或低于 10^{-4} 。

通过多种方法确认共有序列是完整的和正确的基因。最初, 使用
5 限制分析以证实各个克隆中预期大小的 cDNA 插入体的存在。对于全长克隆而言, 通过对 cDNA 插入体进行测序, 检测编码区不存在中断(推测存在起始密码子和终止密码子, 无内在的终止密码子)。对于全长 cDNA 克隆及部分 cDNA 克隆而言, 使用与基因预测的对比和 BLAST 分析以证实指定的靶基因被扩增。将通过 PCR 产生的所有
10 cDNA 克隆均克隆进定制的二元目标(binary destination)载体中, 这可使用基于 Gateway™ 重组的克隆方法根据厂商(Invitrogen)说明进行。或者, 使用基于限制酶的常规克隆方法克隆 PCR 产物。

表达载体和植物的转化

15 针对植物转化的二元目标载体由二元主链和 T-DNA 的一部分组成。所述二元主链含有在大肠杆菌 DH-5α (Invitrogen)和根癌农杆菌 LBA4404 中选择和生长所必需的序列, 包括大肠杆菌转位子 Tn7 的细菌壮观霉素抗生素抗性 aadA 基因, 大肠杆菌(Cole1)和根癌农杆菌(VS1)的复制起点, 及根癌农杆菌 virG 基因。所述 T-DNA 部分两侧
20 是右和左边界序列, 包括 Positech™ (Syngenta)植物选择标记及根据应用而变化的基因表达盒。这种情况中的 Positech™ 植物选择标记由驱动 PMI (磷酸甘露糖异构酶)基因表达的水稻 ACT1 (肌动蛋白)启动子及随后的花椰菜花叶病毒转录终止子组成, 并赋予甘露糖抗性。

所述二元目标载体的基因表达盒部分根据应用而变化。通常地,
25 所述表达盒包含设计为可在植物某些组织中表达所述基因的启动子及随后的克隆位点(在一些情况中由一个间隔 DNA 序列中断), 及最后的根癌农杆菌 nos 3' 末端转录终止子。所用启动子设计为根据感兴

趣的基因不同而在特定靶组织中表达(例如胚乳: 水稻 RS-4, 小麦谷蛋白, 玉米 ADPgpp 或者 γ -玉米蛋白, 或者大麦 α -硫堇; 例如胚: 玉米球蛋白或油体蛋白(oleosin); 例如糊粉: 大麦 α -淀粉酶; 例如根: 玉米 MSR1 和 MRS3; 例如绿色组织: 玉米 PEPC)或组成型表达(例如玉米 UBI plus 内含子)感兴趣的基因。所述克隆位点含有正向或反向的独特的限制酶位点(对于常规克隆)和/或基于 GatewayTM 重组的克隆盒(Invitrogen)。在针对双链干扰 RNA(dsRNA)的生产而设计的基因表达盒中, 所述克隆位点被间隔区(例如水稻 SH1 基因的第一个内含子)分开。所述间隔允许其中一个为正向一个为反向的两个基因片段的克隆。反义技术(反向表达)是可用于沉默感兴趣的基因的另一种技术。

使用在本文上述详细描述章节中描述的方法将本发明的核酸分子转化进植物中。检测含有本发明的核酸分子的植物的测试试剂盒可以使用常规技术制备, 用来生产ELISA型免疫测定中使用的抗体。或者, 植物DNA的PCR型分析的试剂盒可用于测试本发明的核酸分子的存在。

结果

At5g56860基因表达的丧失导致T-DNA插入品系的叶绿素水平降低

20 鉴别在At5g56860基因中具有T-DNA插入突变的纯合植物并筛选任何表型变化。突变植物的叶由于叶绿素水平降低而呈现淡绿色。At5g56860 基因显示出单个锌指, 在锌指环中具有18个残基(C-X2-C-X18-C-X2-C)。全长蛋白质序列含有398个氨基酸。所述T-DNA插入体(SALK_001778)与第二个外显子的末端接近, 导致

25 GATA结构域的缺失。At5g56860的表达未通过RT-PCR在所述突变体中检测到。

为了确定SALK_001778品系的表型变化是否是由单一基因突变导致，本发明人构建了过表达Atg56860基因的构建体，并将这个基因转化回到所述突变体品系中。突变体中的表型在其中检测到At5g56860基因表达的转化体中得以互补，表明At5g56860基因与

5 SALK_001778突变体品系中降低的叶绿素表型变化相关。

将所述突变体与野生型植物回交，通过PCR鉴别既具有野生型基因也具有插入的等位基因的杂种植物。使这些植物自身繁殖，后代种子示出与仅在At5g56860基因中具有T-DNA插入的那些纯合植物中存在的突变体表型呈现3:1分离，表明所述突变体表型是由单一

10 At5g56860基因突变导致的，这对于野生型等位基因是隐性的。

At5g56860基因的表达是组织特异性的，由硝酸盐状态调节；与硝酸盐同化作用基因的倾向一致

为了确定At5g56860基因的表达模式，从芽、叶和根中提取总RNA并产生cDNA。使用RT-PCR确定这个基因的表达模式，表明At5g56860

15 基因在芽和叶中表达，但在根中不表达。为了确定在发育期间这个基因的表达，收获在MA培养基上生长3天和5天后的秧苗并提取总RNA。发现At5g56860基因在生长后正确表达。

由于叶绿素水平通常用于反映氮状态(29—30参考文献)，因此在

20 不同氮条件下研究所述基因的表达，以确定其表达水平是否受氮利用度的影响。当以与高浓度(3 mM)相比在极低硝酸盐浓度(0.3 mM)的溶液培养方式下生长时，At5g56860基因的基础表达水平较低。为了确定这个基因是否受硝酸盐的调节，将植物从1 mM硝酸盐生长条件转移至3mM硝酸盐生长条件。在2小时后At5g56860基因的表达受增高的

25 硝酸盐浓度的上调，在24小时后表达水平降低，但其仍高于基础表达水平。针对nia1、nia2和nir发现相同的表达模式，其为酶硝酸盐还

原中的前两种酶，nia1、nia2和nir的改变倍数的量级比At5g56860基因大。

当N的供应有限时At5g56860-负(less)的植物积聚较少的总N量，

5 但是At5g56860似乎不直接调节硝酸盐同化作用基因的表达

由于At5g56860的表达受氮状态的影响，因此本发明人测试了不同的氮条件是否影响突变体植物的生长。当植物在土壤(3 mM硝酸盐)中生长时,将所述突变体和野生型植物在N受限的条件下和理想氮(10 mM硝酸盐)条件下生长。收集4周龄的突变体和对照植物的芽，测定
10 生物量和总N量。生物量无明显改变。然而，突变体中的总N量在受限N条件下比野生型植物中总N量的情况少，当N供应充足时与野生型植物相似。使用实时PCR分析突变体植物中nia1、nia2和nir的表达，示出突变体中nia1、nia2和nir的表达与野生型植物中的表达无显著不同。如果At5g56860基因是直接控制nia1、nia2和nir表达的调节基因之一，
15 则nia1、nia2和nir的基础表达水平在缺失At5g56860表达的T-DNA突变体中将变化。

At5g56860基因调节与不同功能类别相关的基因的表达

进行转录模式实验从而在突变体和野生型植物中以全部基因组
20 规模对比基础基因表达，以示出哪些基因在突变体中的表达发生改变。再一次将野生型植物和SALK_001778突变体植物在受限的N条件(3 mM 硝酸盐)下或足够的N条件(10 mM硝酸盐)下生长，收集4周龄的芽。提取RNA并与鼠耳芥 GeneChip®全基因组阵列杂交。分析在受限N的条件(3 mM硝酸盐)下生长的突变体中的表达较野生型植物
25 中的表达低至少1.5倍的基因，并分析在足够N浓度(10 mM硝酸盐)下生长的突变体中的表达较野生型植物中的表达低至少1.5倍的基因。使用基于浏览器(browser-based)的功能分类程序(Provar, N. & Zhu,

T. (2003) *Currents in Computational Molecular Biology* 271-272) , 示出下调的基因包括参与氮和硫的代谢中那些基因 (At3g44300 和 At4g27450), 参与调节C-化合物和碳水化合物利用的基因 (At2g18700 和 At1g70290), 参与脂质的生物合成、二级代谢、糖苷的基因
5 (At4g37150、At1g09500、At3g09260), 参与感知营养素和营养适应的基因 (At5g24160), 参与营养素摄取和吸收的基因 (At1g03220 和 At2g37770), 参与电子转运和膜相关的能量守恒的基因 (At5g24160 和 At5g20230)。本发明人通过定量RT-PCR检验了这些基因中的一些基因, PCR和微阵列数据之间的关联性非常高。

10

At5g56860突变体植物对外源葡萄糖更敏感

At5g56860可调节对于不同的生物过程重要的基因的表达, 这些基因有许多参与C的代谢, 如糖和己糖转运子。为了测试在不同C源上突变体植物的生长与野生型植物是否不同, 将这些植物在不同浓度的葡萄糖和蔗糖条件下生长。在6%葡萄糖培养基上, 突变体秧苗的
15 平均生长情况与野生型植物相比而言较矮小。对在含有6%葡萄糖的平板上生长的秧苗表型进行关于糖敏感性的充分的生物分析 (Jang, J., Leon, P, Zhou, L. & Sheen, J. (1997) *Plant Cell* 9, 5-19), 结果示出无 At5g56860表达的突变体对6%葡萄糖高敏感。

20

At5g56860基因可调节HXK1而非HKX2的表达并且也调节己糖转运基因

At5g56860基因影响对葡萄糖的敏感性, 因此预期其影响参与葡萄糖代谢的基因的表达。有两种基因显示影响葡萄糖感应。第一种基因是己糖激酶 (HXK), 因为这种基因中HXK的过表达产生一种与
25 At5g56860 GATA转录因子基因中插入突变的情况极其相似的表型。第二种基因是已显示参与酵母中糖感应的己糖转运基因。通过酵母

hvk1hvk2双突变体的遗传互补鉴别了两种鼠耳芥 HXK基因(Jang, J., Leon, P, Zhou, L. & Sheen, J. (1997) Plant Cell 9, 5-19)。结果示出基础HXK1表达水平在突变体中高于野生型，但是HXK2水平在突变体中根本未改变，提示At5g56860是HXK1而非HXK2的阴性调节子。

- 5 对HXK1和HXK2的500 bp上游序列进行搜索，共有的GATA结合序列(T/A)GATA(G/A)或者互补链上的(T/C)TATC(T/A)在HXK1启动子序列中多次存在，而其在HXK2启动子序列中不存在(序列未示出)。所述己糖转运基因在突变体品系中被下调，在其启动子区域中也存在GATA基序。尽管GATA基序的存在在论证GATA因子对HXK1及己糖
- 10 转运子的调节中未确定，因此其存在至少支持这样的观点，即其也许直接调节这些基因。

功能获得性(Gain-of-function)转基因植物对糖是低敏感性的

- 过表达At5g56860的转基因鼠耳芥植物是通过农杆菌-介导的转
- 15 化产生的(Bechtold, N., Ellis, J. & Pelletier, G. (1993) C R Acad Sci 316, 1194-1199)。转基因植物在含有卡那霉素的培养基上选择，T2品系在卡那霉素上示出3:1分离率，这表示所述转基因的单一插入。这些植物是针对自身授粉加以选择的。将转基因纯合的T3世代转基因品系用于进一步分析。At5g56860在转基因品系中的表达水平通过实时
- 20 RT-PCR确定(数据未示出)。将过表达品系的种子在6%葡萄糖平板上生长，对过表达型转基因秧苗的生长几乎没有抑制效应，而野生型植物受到严重抑制，表明其对外源葡萄糖的敏感性较低。令人感兴趣地注意到HXK1基因在这些转基因品系中未被下调(数据未示出)，而已糖转运基因在这些植物中被上调。

25

讨论

At5g56860是糖感应所必需的

在无At5g56860基因表达的T-DNA突变体中, 所述突变体是糖高敏感的, 而过表达At5g56860基因的转基因植物示出糖的低敏性, 表明At5g56860参与糖感应的调节。对6%葡萄糖的高敏性在过表达AtHXK1和AtHXK2的转基因鼠耳芥植物中已经有报道(Jang, J., Leon, P, Zhou, L. & Sheen, J. (1997) Plant Cell 9, 5-19), 对6%葡萄糖的低敏感性在表达反义AtHXK1和AtHXK2的转基因鼠耳芥植物中报道(Jang, J., Leon, P, Zhou, L. & Sheen, J. (1997) Plant Cell 9, 5-19)。据报道过表达AtHXK1的转基因秧苗在其叶片中叶绿素含量降低(Jang, J., Leon, P, Zhou, L. & Sheen, J. (1997) Plant Cell 9, 5-19; Dai, N., Schaffer, A., Petreikov, M., Shahak, Y., Giller, Y., Ratner, K., Levine, A. & Granot, D. (1999) Plant Cell 11, 1253-1266)。在无At5g56860表达的T-DNA突变体中, AtHXK1基础表达高于野生型对照组, 它们是糖高敏感的。所述突变体示出叶绿素水平降低。在HXK2结构基因的500bp上游序列中没有发现GATA元件。在HXK2启动子的-737区域中有一个GATA元件。然而, 因为位于30bp之内的串联重复或两个GATA元件或核心序列(T/A)GATA(G/A)中是有效结合所需要的(Chiang, T.Y. & Marzluf, G.A. (1994) Biochemistry 33, 576-582; Lin, Y., Hwang, C.F., Brown, J.B. & Cheng, C.L. (1994) Plant Physiol 106, 477-484), 因此不太可能HXK2是由GATA因子直接调节的。相反, HXK1的500bp上游序列具有许多GATA元件。HXK1的表达水平在At5g56860突变体中改变, 但是HXK2的表达水平未改变, 这支持了这样的观点, 即这个GATA因子调节HXK1的表达。

过表达At5g56860的品系对高葡萄糖水平的存在有抗性。这种GATA因子的过表达不导致HXK1表达的改变, 但是导致己糖转运基因的表达增加。已经发现这个基因在酵母中参与糖感应。

At5g56860表达应答于N状态的改变

- At5g56860基因表达显然受植物的N状态的影响。当硝酸盐水平较低时，At5g56860基因的基础表达水平相对较低，当硝酸盐水平较高时，At5g56860基因的基础表达水平相对较高。当将植物从较低硝酸盐条件转变至较高硝酸盐条件时，At5g56860基因的表达被上调。
- 5 这种表达模式与硝酸盐同化作用途径中的关键酶如NR和NiR的表达模式相同，但是当后者暴露于高水平硝酸盐条件时具有较高的表达诱导率。然而，At5g56860基因不直接控制nia1、nia2 和nir的表达，因为所述酶在突变体品系中不受影响。At5g56860基因的表达对氮的代谢确实有影响，因为丧失At5g56860基因导致总的N聚集量降低。令人感兴趣地注意到At5g56860基因与来自所有30个鼠耳芥 GATA转录因子基因的系统发生树的At4g26150基因形成平行(paralogous)关系 (Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z., Ghandehari, D., Sherman, B.K. & Yu, G. (2000) *Science* 290, 2105-2110; Reyes, J.C., Muro-Pastor, M.I. & Florencio, F.J. (2004) *Plant physiol.* 134, 1718-1732)。已经揭示了At4g26150可由硝酸盐诱导(Wang, R., Okamoto, M., Xing, X. & Crawford, N.M. (2003) *Plant Physiol* 132, 556-567; Scheible, W., Morcuende, R., Czechowski, T., Fritz, C., Osuna, D., Palacios-Rojas, D., Schindelasch, D., Thimm, O., Udvardi, M.K. & Stitt, M. (2004) *Plant Physiol* 136, 2483-2499)。然而，其功能似乎不是多余的，因为这两个基因的表达特征不完全一样(Czechowski, T., Bari, R., Stitt, M., Scheible, W. & Udvardi, M. (2004) *Plant J.* 38, 366-379)。在丧失At5g56860基因表达的SALK_001778突变体中，At4g26150基因的表达增加，但是不能互补突变体表型。
- 25

C与N调节之间的干扰

具有挑战性的问题是推定C和N信号途径怎样彼此相互作用，未提及其与其它信号途径如激素、光和应激包含的途径之间的相互作用。对于本发明人而言，这是首次报道的参与调节糖感应的反式作用因子，重要的是注意到所述因子自身的表达受N状态改变的影响。

5

实施例 2

将全长At5g56860 cDNA (GNC)从鼠耳芥叶片cDNA中扩增，使用引物 5'-GCTCTAGATTTCTCTCTCTCTTTGTGTCTTCATTTG-3' (SEQ ID NO:4)和5'-gcgagctctcgggtgactaatgttcgttcc-3' (SEQ ID NO:5)进行。所得~1500 bp片段通过测序检测，然后用XbaI和SacI消化并克隆进XbaI-SacI-消化的表达载体pROK2中，其中含有以组成型、高水平方式驱动GNC表达的花椰菜花叶病毒(CaMV) 35S启动子(图4)。将p35S-GNC二元载体转化进根癌农杆菌菌株EHA105中，所得农杆菌菌株用于转化野生型植物(Col-0)，基于卡那霉素抗性对转化体进行选择。选择具有单个转基因插入的植物用于自身授粉，以产生T3纯合品系用于进一步分析。转基因植物中GNC的过表达通过定量RT-PCR证实。将野生型植物和GNC过表达的植物在足够的N(10mM硝酸盐)和受限的N(3 mM硝酸盐)的条件下生长。对于这两种条件在最初生长阶段观测到无明显的差别。大约28天后，当植物开始进入生殖阶段时，在受限的氮的条件下生长的野生型植物中观测到叶片衰老，但是在GNC过表达的植物中未观测到该现象，这种衰老的差异在此阶段之后的时期中可明显观测到。另外，GNC过表达植物中总叶绿素水平高出大约15%。叶绿素水平是氮利用度和光合作用能力的指标，表明这些过表达的植物品系在氮降低的方案中是较健康的。4周龄的野生型和GNC过表达植物在足够氮的条件下未示出这些差异。在足够氮条件下野生型和GNC过表达植物之间叶绿素水平无显著差异。

20
25

通过前述非限制性实施例对本发明的特殊实施方案进行的描述，本发明在如下权利要求书中得以进一步阐述。本领域技术人员意识到所述权利要求书也允许包含超出权利要求书的文字描述范围的等价物。

<110> 圭尔夫大学 (UNIVERSITY OF GUELPH)
ROTHSTEIN, Steven
BI, Yong-Mei

<120> 氮调节的糖感应基因和蛋白质及其调节
(NITROGEN-REGULATED SUGAR SENSING GENE AND PROTEIN AND MODULATION THEREOF)

<130> 6580-339

<140>

<141>

<150> US 60/643,575

<151> 2005-01-14

<160> 5

<170> PatentIn version 3.3

<210> 1

<211> 1197

<212> DNA

<213> *Arabidopsis thaliana*

<400> 1

```
atggattcaa attttcatta ctcgatagat cttaacgaag atcaaaacca tcacgaacaa      60
cccttttttct atcctcttgg atcctcttcc tcgcttcac atcatcatca tcatcatcat      120
catcaagtcc cttctaattc ttcattcttct tcttcgtcca tttcatcgct ctctctttac      180
ctcccttttct tgatcaactc tcaagaagat caacatgttg cctacaacaa cacttatcac      240
gctgatcatc tccatctttc tcaacccctc aaggccaaga tgtttgtggc taacggtgga      300
tcatcagcat gcgatcacat ggtgccaaag aaggagacaa gactgaaact aacgataagg      360
aaaaaagatc acgaagacca accccatcct cttcatcaaa acccgacaaa acccgattca      420
gactccgaca agtggttgat gtccccaag atgcggttga tcaagaaaac aatcaccaac      480
aataaacagc tcattgatca gactaataat aataatcata aagaaagtga tctactacct      540
ttgaatcata agactaattt cgacgaggat caccatgaag atcttaattt caagaacgct      600
ttgaccagga agaccacggc cgcgaccacc gagaatcgct acaatacaat caacgagaac      660
ggttatagta ataacaatgg cgtgattagg gtttggtcgg attgtaacac caccaagact      720
cctctttggc gaagtggacc tcgaggtccc aagtctcttt gtaacgcatg tggatatacg      780
caaagaaagg caaggcgagc cgctatggcc gcggccgctg cagccggcga ccaagaggtg      840
gcggtagcgc ccgagtgca acaattaccg ctgaaaaaga agttgcaaaa taaaaaaaag      900
agatcaaacg gaggggaaaa atacaatcac tctctccaa tgggtggcaa ggccaaaaag      960
tgcaagatca aagaggaaga ggagaaggaa atggaagcgg aaacggttgc cggagattca     1020
gagatcagca aatctacaac ttcttctaata tcttcgattt cgtcaaacaa attttgcttc     1080
gatgatttga caataatgtt gagcaaaagc tcagcttatc aacaagtgtt cccacaagat     1140
```

gagaaggagg ctgctgtttt gctcatggct ctgtcgtatg gaatggttca cggttga 1197

<210> 2

<211> 398

<212> PRT

<213> Arabidopsis thaliana

<400> 2

Met Asp Ser Asn Phe His Tyr Ser Ile Asp Leu Asn Glu Asp Gln Asn
1 5 10 15

His His Glu Gln Pro Phe Phe Tyr Pro Leu Gly Ser Ser Ser Ser Leu
20 25 30

His His His His His His His His Gln Val Pro Ser Asn Ser Ser
35 40 45

Ser Ser Ser Ser Ser Ile Ser Ser Leu Ser Ser Tyr Leu Pro Phe Leu
50 55 60

Ile Asn Ser Gln Glu Asp Gln His Val Ala Tyr Asn Asn Thr Tyr His
65 70 75 80

Ala Asp His Leu His Leu Ser Gln Pro Leu Lys Ala Lys Met Phe Val
85 90 95

Ala Asn Gly Gly Ser Ser Ala Cys Asp His Met Val Pro Lys Lys Glu
100 105 110

Thr Arg Leu Lys Leu Thr Ile Arg Lys Lys Asp His Glu Asp Gln Pro
115 120 125

His Pro Leu His Gln Asn Pro Thr Lys Pro Asp Ser Asp Ser Asp Lys
130 135 140

Trp Leu Met Ser Pro Lys Met Arg Leu Ile Lys Lys Thr Ile Thr Asn
145 150 155 160

Asn Lys Gln Leu Ile Asp Gln Thr Asn Asn Asn His Lys Glu Ser
165 170 175

Asp His Tyr Pro Leu Asn His Lys Thr Asn Phe Asp Glu Asp His His
180 185 190

Glu Asp Leu Asn Phe Lys Asn Val Leu Thr Arg Lys Thr Thr Ala Ala
195 200 205

Thr Thr Glu Asn Arg Tyr Asn Thr Ile Asn Glu Asn Gly Tyr Ser Asn
210 215 220

Asn Asn Gly Val Ile Arg Val Cys Ser Asp Cys Asn Thr Thr Lys Thr
225 230 235 240

Pro Leu Trp Arg Ser Gly Pro Arg Gly Pro Lys Ser Leu Cys Asn Ala
245 250 255

Cys Gly Ile Arg Gln Arg Lys Ala Arg Arg Ala Ala Met Ala Ala Ala
260 265 270

Ala Ala Ala Gly Asp Gln Glu Val Ala Val Ala Pro Arg Val Gln Gln
275 280 285

Leu Pro Leu Lys Lys Lys Leu Gln Asn Lys Lys Lys Arg Ser Asn Gly
290 295 300

Gly Glu Lys Tyr Asn His Ser Pro Pro Met Val Ala Lys Ala Lys Lys
305 310 315 320

Cys Lys Ile Lys Glu Glu Glu Glu Lys Glu Met Glu Ala Glu Thr Val
325 330 335

Ala Gly Asp Ser Glu Ile Ser Lys Ser Thr Thr Ser Ser Asn Ser Ser
340 345 350

Ile Ser Ser Asn Lys Phe Cys Phe Asp Asp Leu Thr Ile Met Leu Ser
355 360 365

Lys Ser Ser Ala Tyr Gln Gln Val Phe Pro Gln Asp Glu Lys Glu Ala
370 375 380

Ala Val Leu Leu Met Ala Leu Ser Tyr Gly Met Val His Gly
385 390 395

<210> 3

<211> 1044

<212> DNA

<213> Oryza sativa

<400> 3

```

atgtctacca tctacatgag tcagctctca gctgctctcc ctctcatgga gggggagcac      60
caccatcacc accaggatca tcaccaaggc cacttccaag ccttctccct gcagcctaag      120
gatccccag tcttattccc ctttgtgata agtagaagaa gcagcagcag cagccctagc      180
gacagcacca ctctaagcta tggttcagac catcacttga cacagcagca gcagcatcag      240
catcaagcca tgcttgagcc ccaaaatatg attggaggat catccgctgg catctttgcg      300
acgccgttcc cgaccgtcaa gagcatccgc gacgacatga tcgagcggtc gcagttcgat      360
ccatacgata ccgagaagct gcaggcgagc tgcgggttag ccaaggtcgt cgccggcggc      420
aagtggagcg cggtgccagc ggccaagatg aagatcacga ggaagatggg tgagccgtcg      480
tccggtgtca ctggcggggc tgcgacgacg gtggcgccga agaagccgag gaggaggccg      540
gcgcaggcgt acgaggatca cggccatggc ggcgccatgg gccaagcttt tggcgtgatt      600

```

```

agggtgtgct ccgactgcaa caccaccaag actcccttgt ggaggagtgg cccgtgcggc 660
cccaagtgcg ttgcaacgc gtgcggcatc aggcagagga aggcgcggcg ggcgatgatg 720
gcctccggac taccagcgtc cccaacgcc gccggcccca aggcggccgc acatagcggc 780
gccacaaacg cagccgccgc agctgccatg gaggagacgg ccgagtcgcg caccgtcgcc 840
ccgcccccg cgcgcacgac gaggggtggt actctcgtcg acagcatcgg gctcagctgg 900
agcaagacc atgccgccgc caccgcctcc tgcagcttcc ggccgtcacc ggtggctccc 960
ggcttcgcgg cggcggtgca ggacgagatc actgacgccg ccatgctgct catgacgctg 1020
tcctgcgggc ttgtccggag ctga 1044

```

```

<210> 4
<211> 36
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> primer

```

```

<400> 4
gctctagatt tctctctctc ttgtgtctt catttg 36

```

```

<210> 5
<211> 30
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> primer

```

```

<400> 5
gcgagctctc gggtgactaa tgttcgttcc 30

```

At5g56860 全长 CDS (1197 bp): SEQ ID NO: 1

```

1 ATGGATTCAA ATTTTCATTA CTCGATAGAT CTTAACGAAG ATCAAAACCA
51 TCACGAACAA CCCTTTTCT ATCCTCTTGG ATCCTCTTCC TCGCTTCATC
101 ATCATCATCA TCATCATCAT CATCAAGTCC CTTCTAATTC TTCATCTTCT
151 TCTTCGTCCA TTTCATCGCT CTCCTCTTAC CTCCCTTTCT TGATCAACTC
201 TCAAGAAGAT CAACATGTTG CCTACAACAA CACTTATCAC GCTGATCATC
251 TCCATCTTTC TCAACCCCTC AAGGCCAAGA TGTTTGTGGC TAACGGTGGA
301 TCATCAGCAT GCGATCACAT GGTGCCAAG AAGGAGACAA GACTGAAACT
351 AACGATAAGG AAAAAAGATC ACGAAGACCA ACCCATCCT CTTCATCAAA
401 ACCCGACAAA ACCCGATTCA GACTCCGACA AGTGGTTGAT GTCCCCAAG
451 ATGCGGTTGA TCAAGAAAAC AATCACCAAC AATAAACAGC TCATTGATCA
501 GACTAATAAT AATAATCATA AAGAAAGTGA TCACTACCTT TTGAATCATA
551 AGACTAATTT CGACGAGGAT CACCATGAAG ATCTTAATTT CAAGAACGTC
601 TTGACCAGGA AGACCACGGC CGCGACCACC GAGAATCGCT ACAATACAAT
651 CAACGAGAAC GGTATAGTA ATAACAATGG CGTGATTAGG GTTGTTCGG
701 ATTGTAACAC CACCAAGACT CCTCTTTGGC GAAGTGGACC TCGAGGTCCC
751 AAGTCTCTTT GTAACGCATG TGGTATACGG CAAAGAAAGG CAAGGCGAGC
801 CGCTATGGCC GCGGCCGCTG CAGCCGGCGA CCAAGAGGTG GCGGTAGCGC
851 CCCGAGTGCA ACAATTACCG CTGAAAAAGA AGTTGCAAAA TAAAAAAAAG
901 AGATCAAACG GAGGGGAAAA ATACAATCAC TCTCCTCCAA TGGTGGCCAA
951 GGCCAAAAAG TGCAAGATCA AAGAGGAAGA GGAGAAGGAA ATGGAAGCGG
1001 AAACGGTTGC CGGAGATTCA GAGATCAGCA AATCTACAAC TTCTTCTAAT
1051 TCTTCGATTT CGTCAAACAA ATTTTGCTTC GATGATTTGA CAATAATGTT
1101 GAGCAAAAGC TCAGCTTATC AACAAGTGTT CCCACAAGAT GAGAAGGAGG
1151 CTGCTGTTTT GCTCATGGCT CTGTCGTATG GAATGGTTCA CGGTTGA

```

图 1

At5g56860 蛋白序列 (398 aa): SEQ ID NO: 2

```
1 MDSNFHYSID LNEDQNHHEQ PFFYPLGSSS SLHHHHHHHH HQVPSNSSSS
51 SSSISSLSSY LPFLINSQED QHVAYNNTYH ADHLHLSQPL KAKMFVANGG
101 SSACDHMVPK KETRLKLTIR KKDHEQDQPH LHQNPTKPDS DSDKWLMSPK
151 MRLIKKTITN NKQLIDQTNN NNHKEVDHYP LNHKTNFDDE HHEDLNFKNV
201 LTRKTTAATT ENRYNTINEN GYSNNNGVIR VCSDCNTTKT PLWRSGPRGP
251 KSLCNACGIR QRKARRAAMA AAAAAGDQEV AVAPRVQQLP LKKKLQNKKK
301 RSNGGEKYNH SPPMVAKAKK CKIKEEEEEKE MEAETVAGDS EISKSTSSN
351 SSISSNKFCF DDLTIMLSKS SAYQQVFPQD EKEAAVLLMA LSYGMVHG
```

图 2

SEQ ID NO: 3 水稻同系物

>OsGATA16_205599

ATGTCTACCATCTACATGAGTCAGCTCTCAGCTGCTCTCCCTCTCATGGAGGGGGAG
CACCACCATCACCACCAGGATCATCACCAAGGCCACTTCCAAGCCTTCTCCCTGCAG
CCTAAGGATCCCCCAGTCTTATTCCCCTTTGTGATCAGTAGAAGAAGCAGCAGCAGC
AGCCCTAGCGACAGCACCCTCTAAGCTATGGTTCAGACCATCACTTGACACAGCAG
CAGCAGCATCAGCATCAAGCCATGCTTGAGCCCCAAAATATGATTGGAGGATCATCC
GCTGGCATCTTTGCGACGCCGTTCCCGACCGTCAAGAGCATCCGCGACGACATGATC
GAGCGGTTCGAGTTCGATCCATACGATAACGAGAAGCTGCAGGCGAGCTGCGGGTTA
GCCAAGGTCGTCGCCGGCGGCAAGTGGAGCGCGGTGCCAGCGGCCAAGATGAAGATC
ACGAGGAAGATGGGTGAGCCGTCGTCCGGTGTCACTGGCGGGGCTGCGACGACGGTG
GCGCCGAAGAAGCCGAGGAGGAGGCCGCGCAGGCGTACGAGGATCACGGCCATGGC
GGCGCCATGGGCCAAGCTTTTGGCGTGATTAGGGTGTGCTCCGACTGCAACACCACC
AAGACTCCCTTGTGGAGGAGTGGCCCGTGCGGCCCCAAGTCGCTTTGCAACGCGTGC
GGCATCAGGCAGAGGAAGGCGCGGGCGGCGATGATGGCCTCCGGAATACCAGCGTCC
CCCAACGCCGCCGGCCCCAAGGCGGCCGCACATAGCGGCGCCACAAACGCAGCCGCC
GCAGCTGCCATGGAGGAGACGGCCGAGTCCGCCACCGTCGCCCCGCCCCCGGCGCCG
ACGACGAGGGGTGGTACTCTCGTCGACAGCATCGGGCTCAGCTGGAGCAAGACCCAT
GCCGCCGCCACCGCCTCCTGCAGCTTCCGGCCGTCACCGGTGGCTCCCGGCTTCGCG
GCGGCGGTGCAGGACGAGATCACTGACGCCGCCATGCTGCTCATGACGCTGTCCTGC
GGGCTTGTCCGGAGCTGA

图 3

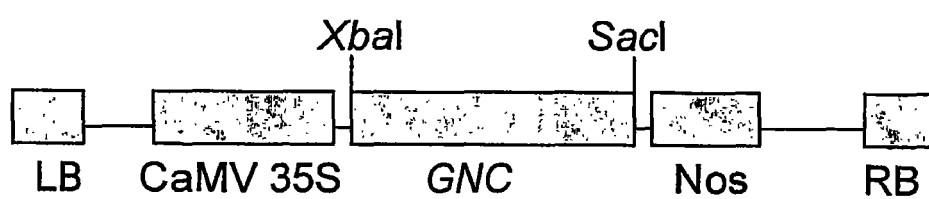


图 4

专利名称(译)	氮调节的糖感应基因和蛋白质及其调节		
公开(公告)号	CN101142317A	公开(公告)日	2008-03-12
申请号	CN200680008289.5	申请日	2006-01-13
[标]申请(专利权)人(译)	圭尔夫大学		
申请(专利权)人(译)	圭尔夫大学		
当前申请(专利权)人(译)	圭尔夫大学		
[标]发明人	S罗斯坦 Y M毕		
发明人	S·罗斯坦 Y-M·毕		
IPC分类号	C12N15/29 A01H5/10 G01N33/53 C12N15/82 C12Q1/68 C12N5/10 A01H5/00 C12N5/04 C07K16/16		
CPC分类号	C07K14/415 C12N15/8245 C12N15/8243 C12N15/8261 C12N15/8242 Y02A40/146		
代理人(译)	林晓红		
优先权	60/643575 2005-01-14 US		
外部链接	Espacenet SIPO		

摘要(译)

本发明涉及糖感应所需的氮调节的GATA转录因子基因，及调节这种基因的表达以调节植物中的特征。本发明的GATA转录因子参与调节植物中的糖感应并且其表达受氮状态的影响。这种基因或基本相似的基因的表达增加可产生氮利用被改善并且产量增加的植物。