

[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 200880018512.3

[51] Int. Cl.

*C12N 5/10 (2006.01)*  
*A01H 1/00 (2006.01)*  
*A01H 1/04 (2006.01)*  
*A01H 3/00 (2006.01)*  
*A01H 5/00 (2006.01)*  
*C07K 16/16 (2006.01)*

[43] 公开日 2010年3月31日

[11] 公开号 CN 101688180A

[51] Int. Cl. (续)

*C12N 15/82 (2006.01)*

*C12N 15/29 (2006.01)*

*C12Q 1/68 (2006.01)*

*G01N 33/53 (2006.01)*

[22] 申请日 2008.4.16

[21] 申请号 200880018512.3

[30] 优先权

[32] 2007.4.17 [33] CA [31] 2,584,934

[86] 国际申请 PCT/CA2008/000688 2008.4.16

[87] 国际公布 WO2008/124933 英 2008.10.23

[85] 进入国家阶段日期 2009.12.2

[71] 申请人 圭尔夫大学

地址 加拿大安大略省

[72] 发明人 S·罗斯坦 Y-M·毕

[74] 专利代理机构 北京市中咨律师事务所

代理人 黄革生 史文静

权利要求书5页 说明书80页 序列表8页  
附图9页

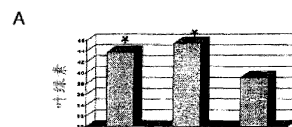
[54] 发明名称

调节碳和氮的基因与蛋白质及其调节

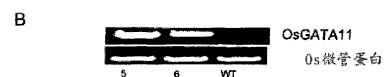
[57] 摘要

本发明涉及针对糖和氮的积累所需要的调节氮的 GATA 转录因子基因以及为调节植物中的特性而对这种基因的表达的调节。本发明的 GATA 转录因子与调解植物中糖和氮的积累有关。提高这个基因或基本相似的基因的表达能够产生具有改善的氮利用以及提高的产量和提高的胁迫耐受性的植物。

过量表达OsGATA11的植物的表型



\* P < 0.01 (与6株wt植物比较三到六株PMII阳性的植物的平均水平)



1.调节植物或植物细胞中的特性的方法，其包括调节植物或植物细胞中的 GATA 转录因子的表达，其中 GATA 转录因子包括：

- (a) SEQ ID NO: 1 的核苷酸序列；
- (b) 编码 SEQ ID NO:2 的多肽的核苷酸序列；或
- (c) 能够与 (a) 或 (b) 进行杂交的核苷酸序列，

并且其中所述特性选自：叶绿素合成、种子产量、胁迫耐受性、硝酸盐水平、氨基酸水平以及糖积累。

2.如权利要求 1 所述的方法，其中 GATA 转录因子基因的表达是通过给予植物或植物细胞有效量的活性物质来进行调节，所述活性物质能够调节在植物细胞中 GATA 转录因子基因的表达水平。

3.如权利要求 2 所述的方法，其中活性物质增强了在植物或植物细胞中 GATA 转录因子的表达水平。

4.如权利要求 3 所述的方法，其中受调节的特性是以下各项中的一项或多项的提高或改善：叶绿素合成、种子产量、胁迫耐受性、硝酸盐水平、氨基酸水平以及糖积累。

5.如权利要求 3 或 4 所述的方法，其中增强植物或植物细胞中的 GATA 转录因子基因的表达水平的活性物质包括编码 GATA 转录因子的核酸分子，其中所述核酸分子表达于植物或植物细胞中。

6.如权利要求 5 所述的方法，其中核酸分子包括：

- (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、其片段或结构域；
- (c) 与 (a) 或 (b) 具有基本相似性的核苷酸序列；
- (d) 能够与 (a)、(b) 或 (c) 进行杂交的核苷酸序列；
- (e) 与 (a)、(b)、(c) 或 (d) 互补的核苷酸序列；或
- (f) 核苷酸序列，它是 (a)、(b)、(c) 或 (d) 的反向互补序列。

7.如权利要求 5 所述的方法，其中核酸分子包括 SEQ ID NO: 1 的

OsGATA11 基因的序列或其功能片段。

8.如权利要求 5 所述的方法，其中核酸分子包括序列，所述序列在中度严格性条件下与 SEQ ID NO: 1 的 OsGATA11 基因或其功能片段进行杂交。

9.如权利要求 5 所述的方法，其中核酸分子包括核酸序列，所述核酸序列衍生自 SEQ ID NO: 1 的 OsGATA11 基因的核苷酸序列，并且具有包括对植物中的表达特异的密码子的核苷酸序列。

10. 如权利要求 5 至 9 中的任何一项所述的方法，其中所述核酸序列表达于植物的特定的位置或组织中。

11. 如权利要求 10 所述的方法，其中所述位置或组织选自下述中的一项或多项：种子、表皮、根、维管组织、分生组织、形成层、皮层、髓、叶、以及花。

12. 如权利要求 11 所述的方法，其中所述位置或组织是种子。

13. 如权利要求 5 至 12 中的任何一项所述的方法，其中增强植物细胞中的 GATA 转录因子基因的表达水平的活性物质包括表达盒，所述表达盒包括启动子序列，所述启动子序列与编码 GATA 转录因子的核酸有效连接。

14. 如权利要求 3 所述的方法，其中能够调节植物或植物细胞中的 GATA 转录因子的表达水平的活性物质包括：

(a) 在 SEQ ID NO: 2 中所示出的多肽序列，或其功能片段、结构域、重复、或嵌合体；

(b) 与 (a) 具有基本相似性的多肽序列；

(c) 由核苷酸序列所编码的多肽序列，所述核苷酸序列与在 SEQ ID NO: 1 中所列出的核苷酸序列、或其功能片段或结构域、或与它互补的序列相同或具有基本相似性；或

(d) 由核苷酸序列所编码的多肽序列，所述核苷酸序列能够在中度严格性条件下与在 SEQ ID NO: 1 中所列出的核苷酸序列、或与之互补的序列进行杂交。

15. 如权利要求 2 所述的方法, 其中所述活性物质减少了在植物或植物细胞中 GATA 转录因子的表达水平。

16. 如权利要求 15 所述的方法, 其中受调节的特性是以下各项中的一项或多项的降低或减少: 叶绿素合成、种子产量、胁迫耐受性、硝酸盐水平、氨基酸水平以及糖积累。

17. 如权利要求 15 或 16 所述的方法, 其中降低了植物或植物细胞中的 GATA 转录因子基因的表达水平的活性物质包括核酸分子, 其阻抑或抑制 GATA 转录因子。

18. 如权利要求 17 所述的方法, 其中核酸是干扰 RNA (RNAi)。

19. 如权利要求 1 至 18 中的任何一项所述的方法, 其中植物细胞是双子叶植物、裸子植物、或单子叶植物。

20. 如权利要求 19 所述的方法, 其中单子叶植物是选自: 玉米、小麦、大麦、燕麦、裸麦、粟、高粱、黑小麦、黑麦属、单粒小麦、斯佩耳特小麦、双粒小麦、画眉草、蜀黍、亚麻、格兰马草、磨擦草属物种和玉米草。

21. 如权利要求 20 所述的方法, 其中双子叶植物是选自: 大豆、烟草或棉花。

22. 产生转基因植物的方法, 其包括:

(1) 提供具有下述序列的分离的核酸, 所述序列包括:

(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、其片段或结构域;

(c) 与 (a) 或 (b) 具有基本相似性的核苷酸序列;

(d) 能够与 (a)、(b) 或 (c) 进行杂交的核苷酸序列;

(e) 与 (a)、(b)、(c) 或 (d) 互补的核苷酸序列; 或

(f) 核苷酸序列, 它是 (a)、(b)、(c) 或 (d) 的反向互补序列; 以及

(2) 将所述核酸引入植物之中, 其中所述核酸在植物中表达。

23. 如权利要求 22 所述的方法, 其中植物表现出在以下的一项或多项中的提高或改善: 叶绿素合成、种子产量、胁迫耐受性、硝酸盐水平、

氨基酸水平以及糖积累。

24. 如权利要求 23 所述的方法，其中植物在叶绿素合成、种子产量和/或胁迫耐受性上具有提高。

25. 如权利要求 22 至 24 中的任何一项所述的方法，其中使用方法将核酸引入植物之中，所述方法选自：微粒轰击、农杆菌介导的转化、以及颈须-介导的转化。

26. 由权利要求 22 至 25 中的任何一项产生的植物的植物细胞。

27. 核酸分子的用途，所述核酸分子包括具有至少 10 个碱基的核苷酸序列，该序列与 SEQ ID NO: 1、或其功能片段的任何区域相同、互补、或基本上相似，并且其中所述用途选自：

(i)作为染色体标记物用于鉴定天然或人工染色体上相应或互补的多核苷酸位置的用途；

(ii)作为 RFLP 分析标记物的用途；

(iii)作为数量性状关联育种的标记物的用途；

(iv)作为标记物-辅助育种的标记物的用途；

(v)作为钓饵序列在双杂交体系中用于鉴定编码多肽的序列的用途，所述多肽与钓饵序列编码的多肽相互作用；

(vi)作为对个体或个体群进行基因分型或鉴定的诊断指示物的用途；和

(vii)用于鉴定基因或外显子边界的遗传分析的用途。

28. 针对分离的多肽所产生的抗体，所述多肽包括：

(a) SEQ ID NO: 2 的多肽序列，或它的片段、结构域、重复、或嵌合体；

(b) 与 (a) 具有基本相似性的多肽序列；

(c) 多肽序列，所述多肽序列由核苷酸序列编码，该核苷酸序列与 SEQ ID NO: 2 中所列出的核苷酸序列、或其片段或结构域、或与它互补的序列，相同或具有基本相似性；

(d) 多肽序列，该多肽序列由核苷酸序列编码，该核苷酸序列能够在中度严格性条件下与 SEQ ID NO: 1 中所列出的核苷酸序列或与之互补的

序列进行杂交；或者

(e) (a)、(b)、(c) 或 (d) 的功能片段。

29. 如权利要求 28 所述的抗体，其中所述多肽包括 SEQ ID NO: 2 的序列或其变体，所述变体具有保守的氨基酸修饰。

30. 免疫测定试剂盒，所述试剂盒包括如权利要求 28 或 29 所述的抗体及其使用说明书。

## 调节碳和氮的基因与蛋白质及其调节

### 发明领域

本发明涉及通过调节植物细胞中的 GATA 转录因子的表达来调节植物中的农学性状的方法。具体而言，这种 GATA 转录因子是从稻(*Oryza sativa*)中分离的，并且在叶绿素合成、糖积累、氮状况、胁迫耐受性以及籽粒(grain)产量方面是重要的，并且最后可以调节氮摄取和总的碳代谢。

### 发明背景

对作物植物农学特征的改良从农业开始就一直在进行。大部分适合作物生产的土地目前已被使用。因为人口持续增长，所以需要改良的作物品种以充分地提供我们的食物和饲料(Trewavas (2001) *Plant Physiol.* 125: 174-179)。为了避免灾难性的饥荒和营养不良，未来的作物栽培种将需要使用等量的农业投入获得改进的产量。这些栽培种将需要更有效地抵抗不利条件如干旱、土壤盐碱化或疾病，当贫瘠的土地进行耕种时这会是特别重要的。最后，我们将需要下述栽培种，其具有改变的营养组成以增强人和动物营养并使得能够进行有效的食物和饲料加工。对所有这些性状而言，鉴定控制目的性状表型表达的基因对于通过常规或转基因手段加速优良作物种质的开发是决定性的。

可获得大量高效的方法来帮助鉴定在农学重要性状的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。遗传学是遗传机制的科学研究。通过鉴定改变目的途径或应答的突变，经典遗传学(或正向遗传学)能够帮助鉴定涉及这些途径或应答的基因。例如，对疾病具有增加的易感性的突变体可鉴定从病原体识别通向疾病抗性的植物信号转导途径的重要组件。遗传学也是通过育种改良种质的中心组件。通过遗传杂交的分子和表型分析，控制目的性状的基因座可以被绘图并在

随后的世代中被跟踪。获知作物增加物(accession)之间表型变异下潜在的基因可使得能够开发下述标记物,所述标记物大幅提高种质改良方法的效率,并且打开了发现其他优良等位基因的通道。

基因组学是对生物基因组的系统水平研究,所述生物基因组包括基因和相应的基因产物——RNA和蛋白质。在初级水平上,基因组方法提供了来自不同植物物种序列信息的巨大数据集,包括模式植物物种拟南芥的全长和部分cDNA序列和全基因组序列。最近,也可以获得作物植物基因组稻(*Oryza sativa*)基因组的最初序列草案(draft sequence)。全基因组序列的可用性使得可能开发在系统水平上研究其他分子互补物的工具,如阵列和芯片,其用于测定生物在特定条件下表达的基因的互补物。这些数据可被用作某些基因在不同植物表型的表达中起关键作用的潜能的最初指示。

生物信息学方法与初级水平的基因组数据集直接结合,允许通过注解(annotative)或其他手段处理以揭示目的序列。使用例如相似性搜索、比对和种系发生分析,生物信息学通常可鉴定目的基因产物的同源物(homolog)。非常类似的同源物(例如在蛋白质全长上具有>~90%的氨基酸同一性)非常可能是直向同源物,即在不同的生物中具有同样的功能。

功能基因组学可被定义为对基因及其产物的功能指定。功能基因组学从遗传学、基因组学和生物信息学中得到鉴定下述基因的途径,所述基因在具体的目的途径或应答中是重要的。表达分析例如使用高密度DNA微阵列(长来自基因组规模的生物测序)在单次实验中监测数千基因的mRNA表达。实验处理可包括引发目的应答的处理,所述目的应答例如用病原体感染的植物中的疾病抗性应答。为了给出微阵列用途的额外实例,可在一段发育时程内的不同组织中或在受目的应答影响的突变体中监测mRNA表达水平。蛋白组学也可通过在单次实验中检验数百个蛋白质的表达和翻译后修饰来帮助指定功能。

蛋白组学方法在许多情况下与在微阵列实验中采取的监测mRNA表达的方法类似。蛋白质-蛋白质相互作用也可通过鉴定与途径或应答的已知组件相互作用的蛋白质,帮助将蛋白质指定至给定的途径或应答。对于功

能基因组学而言,通常使用大规模酵母双杂交实验研究蛋白质-蛋白质相互作用。指定基因功能的另一方法是在异源宿主例如细菌大肠杆菌(*Escherichia coli*)中表达相应的蛋白质,然后进行纯化和酶检验。

目的基因控制给定性状的能力的证实可来自例如在目的植物物种中的实验验证。转基因植物目的基因的产生和分析可用于植物功能基因组学,这具有若干优点。基因通常可被超量表达和低表达(*underexpressed*) (“敲除”),从而增加观察到下述表型的机会,所述表型将该基因与目的途径或应答联系在一起。转基因功能基因组学的两个方面有助于给予通过该途径的功能指定以高置信度水平。首先,在生活植物的背景下进行表型观察。其次,可以检查观察到的表型范围并与观察到的引入的转基因表达水平相关联。转基因功能基因组学在改良栽培种的开发中特别有价值。只有下述基因作为作物改良成就的候选基因被促进,所述基因在目的途径或应答中发挥作用并且能够另外赋予以期望的性状为基础的表型。在一些情况下,针对功能基因组学研究开发的转基因株系可在产品开发的初期直接使用。

朝向植物功能基因组学的另一途径首先鉴定在特定目的基因中具有突变的植物株系,然后在所研究的性状上对这类基因敲除的结果进行表型评价。这样的途径揭示了特定性状表达必需的基因。

通过功能基因组学鉴定的基因可在如上所述通过转基因手段改良种质的努力中直接使用,或被用于开发在作图和繁殖种群中鉴定目的等位基因踪迹的标记物。获知这类基因也可使得能够通过大量分子方法中的任意方法来构建自然中不存在的优良等位基因。

在过去80年中,行栽作物(*row crop*)中产量的快速增加在大致相等的程度上归因于改进的遗传学和改进的农学实践。具体地,在作物如玉米中,高产杂种和大量氮肥使用的组合在理想的条件下允许大于440蒲式耳/英亩(*bu/acre*)的产量。然而,大量氮肥的使用具有负面的副作用,主要在于增加的该农民投入成本和增加的环境成本,因为硝酸盐污染是在许多农业地区显著促成淡水和海洋环境退化的主要问题。通过理解基因型对氮使用的作用来开发更有效利用氮的作物遗传学在降低生产者投入成本以及环境负

荷中会是高度有利的。这对使用高水平氮肥栽培的作物如玉米而言尤其重要。

氮使用效率可以以若干种方式定义，尽管最简单的是产量/应用的N。该方法中存在两个阶段：首先，被吸收、储存和同化为氨基酸和其他重要含氮化合物的可获得的氮量；其次，被分配至种子而导致最终产量的氮比例。已对多种农业上重要的作物进行了多种产量研究以研究该问题(Lawlor DW 等, 2001 在 Lea PJ, Morot-Gaudry JF, 编著 *Plant Nitrogen*. Berlin: Springer-Verlag 343-367; Lafitte HR 和 Edmeades GO 1994 *Field Crops Res* 39, 15-25; Lawlor DW 2002 *J Exp Bot.* 53, 773-87; Moll RH 等, 1982 *Agron J* 74, 562-564)。这些实验已证明存在氮使用效率的遗传组件，但就确定何种基因对该过程是重要的而言，尚未证明这些实验是令人满意的。另外，玉米种植者一般不把在有限的氮肥下维持产量作为目标。这些类型的对氮使用的产量研究就多种使得实验难以解释的原因而言是困难的，所述原因包括在测试田 (test field) 中或在任何处理制度下的田间位点(field sites)之间缺乏可得氮的均一性，和其他环境因素的互相影响。

因此，尽管存在这种性状的遗传变异的实验证据，但从这些实验中做出任何什么引起这一变异的结论仍是困难的。发展多种方法以便在田间条件下在作物植物中研究这种性状，这应该是可行的并且无疑是重要的。然而，通过使用模式系统（像拟南芥属(*Arabidopsis*））可进行针对鉴定、理解、以及操作重要的性状的重大进步。至少，这些实验将给出关于在重要的田间作物中有待评估的潜在靶基因的重要线索。此外，还存在相当多的可供用来研究稻的遗传资源和基因组资源，并且这一物种还将作为与拟南芥属相比更类似于玉米的一个物种用于所提出的实验中的一些实验。

硝酸盐是田间可供使用的氮的主要形式，并且存在大量的文献涉及硝酸盐摄取和还原的基因（Forde BG 2000 *Biochimica et Biophysica Acta* 1465, 219-235; Howitt SM 和 Udvardi MK 2000 *Biochimica et Biophysica Acta* 1465, 152-170; Stitt M 等人 2002 *J Exp Bot.* 53, 959-70）以及参与氮代谢的其他方面的基因（Lea PJ, Morot-Gaudry JF, 编著 2001 *Plant*

Nitrogen. Berlin: Springer-Verlag; Morot-Gaudry JF 2001 Nitrogen assimilation by plants Science Publishers Inc. NH, US)。同样,清楚的是碳代谢产物的利用率对田间硝酸盐的有效利用至关重要,并且针对碳和氮代谢之间的联系存在良好的实验证据(Coruzzi GM 和 Zhou L 2001 Curr Opin Plant Biol. 4, 247-53)。此外,一些实验提示GS以及GOGAT参与了将N从开始衰老的器官中重新流通至库器官(sink organ)中(Brouquisse R 等人 2001 in Lea PJ, Morot-Gaudry JF, 编著 Plant Nitrogen. Berlin: Springer-Verlag 275-293; Yamaya T 等人 2002 J Exp Bot. 53, 917-925)。然而,这些基因的调节的大多数方面仍然是不清楚的,并且对这种调节如何影响氮使用效率仍没有概念。

植物能够感知(sense)碳和氮代谢产物的水平并从而调整生长和发育。这种感觉机制是控制基因表达以适应营养依赖性的细胞活性不断进行变化的复杂的调节网络。拥有糖感知机制使植物能够在碳骨架充足时关闭光合作用。氮感知机制使植物能够在还原氮或有机氮处于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)。

在植物中存在着多种糖的信号转导途径。葡萄糖已显现为光合植物中许多生命过程(如在光合作用中以及在碳和氮代谢中)的一种关键调节物(Rolland, F., Moore, B. & Sheen, J. (2002) Plant Cell S185-S205)。己糖激酶(HXK)是葡萄糖代谢的一个重要的控制点。它们不仅催化葡萄糖的磷酸化,而且作为一种葡萄糖感受器发挥功能,使养分、光以及激素信号传导网络相互联系,用于响应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., Cheng, W., Liu, Y., Hwang, I., Jones, T. & Sheen, J. (2003) Science 300, 332-336)。在其他生物中已经表明己糖运输分子也充当糖感受器。

植物中也存在多种N信号和传感(sensing)途径。植物具有多种机制

来感知作为无机 N 状况的信号的硝酸盐（氮肥的主要形式），并感知作为还原性 N 或有机 N 状况的信号的来源于硝酸盐的代谢产物。硝酸还原酶（NR）和亚硝酸还原酶（NiR）是硝酸盐还原过程中最初的两种酶，并且它们的表达可受到硝酸盐的存在的刺激并被其他生理因素（包括一些含氮化合物、蔗糖、光以及激素）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., 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. 编著 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密切相连并被紧密地调节（Coruzzi, G. & Bush, D.R. (2001) *Plant Physiol* 125, 61-64）。碳代谢产物的利用率对于有效地使用硝酸盐而言是至关重要的，并且氮状况对光合作用是非常敏感的。尽管参与碳和氮代谢的结构性基因的知识得到了增加，但涉及到 C/N 基因表达的转录调节的反式作用因子还没有得到表征。

GATA 转录因子是在真核生物中广泛分布的一类转录调节物。通常，GATA 的 DNA 结合结构域识别共有序列 WGATAR（W = T 或 A；R = G 或 A）（Lowry, J. & Atchley, W. (2000) *J Mol Evol* 50, 103-115）。已经在许多光应答基因的调节区域中鉴定出 GATA 基序（Arguello-Astorga, G. & Herrera-Estrella, L. (1998) *Annu Rev Plant Physiol Plant Mol Biol* 49, 525-555），这些光应答基因包括参与光合作用或与光合作用有关的许多基因，如 RBCS、CAB（叶绿素 A/B 结合蛋白）以及 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作用的基因，如硝酸还原酶、亚硝酸

还原酶以及 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)。在 N 代谢中总体调节基因的一些已知的反式作用调节蛋白是 GATA 转录因子基因。在酵母中, 四个总氮调节因子 GLN3、NIL1、NIL2 以及 DAL80 是包含单个 GATA 锌指的 DNA 结合蛋白, 其识别共有基序 GATA( Hofman-Bang, J. (1999) *Mol Biotech* 12, 35-73)。在真菌中, 粗糙链孢霉(*Neurospora crassa*)NIT2( Tao Y 和 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 因子的体内功能仍没有得到很好的定义, 其中拟南芥属基因组具有 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., 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)。以前申请人鉴定出拟南芥属的 GATA 转录因子基因 GNC ( *At5g56860* ), 它在叶绿素合成和糖敏感性中是重要的 ( WO 2006/074547)。在稻 ( *Oryza sativa* ) 基因组中, 存在 28 个 GATA 转录因子基因, 其中一个基因 *OsGATA16* 与拟南芥属 GATA 基因 *At5g56860* 共有相似性 ( Reyes, J.C., Muro-Pastor, M.I. & Florencio, F.J. (2004) *Plant Physiol.* 134, 1718-1732 和 WO 2006/074547) 。

### 发明概述

诸位发明人已经从稻中分离出一种新的 GATA 转录因子, 命名为

OsGATA11, 它是来自拟南芥属的 At4g26150 基因的一种直向同源物。At4g26150 基因是在 30 个拟南芥属的 GATA 转录因子基因的系统树中的一种 GNC 旁系同源物 (Reyes, J.C., Muro-Pastor, M.I. & Florencio, F.J. (2004) Plant Physiol. 134, 1718-1732), 并且已被发现具有与 GNC 重叠的功能。诸位发明人已经确定 OsGATA11 基因的表达调节了叶绿素合成、种子产量以及针对低氮水平的胁迫应答。而且, OsGATA11 基因的过量表达能够对硝酸盐和氨基酸水平以及对糖的积累产生积极的作用。OsGATA11 基因中的失功能突变体植物导致叶绿素水平降低、氨基酸和蛋白质水平降低、连同糖积累减少。

具体而言, 产生了通过 RNAi 将 OsGATA11 基因沉默的转基因稻植物以及过量表达该稻基因的转基因植物。用 OsGATA11 基因转化的植物具有提高的叶绿素水平和提高的种子产量, 并且具有针对低氮水平的改善的胁迫应答。在高 N 下生长的植物在从生长室转移到温室后经历了胁迫, 并且这些过量表达 OsGATA11 的转基因植物对该胁迫的应答更好的多。

糖是光合作用的植物中许多生命过程 (如光合作用以及碳和氮代谢) 的中心调节物。这种调节是通过调节基因表达来激活或抑制有关基因而实现的。对糖控制基因表达的机理还不是很了解。这里公开的 GATA 转录因子参与调解了糖水平 (包括蔗糖、葡萄糖以及果糖水平), 连同调节硝酸盐、氨基酸以及蛋白质水平。这个基因的表达提高能够产生具有提高产量的植物, 特别是当糖信号传导途径的操纵能够导致提高的光合作用以及提高的氮同化作用并改变种子、块茎 (tuber)、根、以及其他贮藏器官中的源库关系时。

因此, 本发明涉及调节植物或植物细胞中的特性的方法, 包括调节该植物或植物细胞中的 GATA 转录因子基因的表达。在一个优选的实施方案中, 本发明提供了调节植物或植物细胞中的特性的方法, 包括调节该植物或植物细胞中的 GATA 转录因子的表达, 其中该 GATA 转录因子包括:

- (a) SEQ ID NO: 1 的核苷酸序列;
- (b) 编码 SEQ ID NO: 2 的多肽的核苷酸序列; 或者

(c) 能够与 (a) 或 (b) 进行杂交的核苷酸序列，

并且其中该特性是选自：叶绿素合成、种子产量、胁迫耐受性、硝酸盐水平、氨基酸水平以及糖积累。

在本发明的一个实施方案中，这种 GATA 转录因子基因的表达是通过给予该细胞有效量的活性物质 (agent) 来进行调节的，该活性物质可以调节该植物细胞中 GATA 转录因子基因的表达水平。在本发明的一个具体的实施方案中，该活性物质增强了该植物或植物细胞中 GATA 转录因子基因的表达水平。在本发明的另一个具体的实施方案中，该活性物质减小了该植物或植物细胞中 GATA 转录因子基因的表达水平。

在该植物中有待调节的特性可以是任何所感兴趣的农学性状。在本发明的一个实施方案中，该特性是受以下各项影响的任何特性：氮、碳和/或硫代谢，脂类的生物合成，养分的感知、营养的适应，电子传递和/或与膜相关的能量守恒。在本发明的另外一个实施方案中，该特性是选自以下各项中的一个或多个：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、非生物胁迫耐受性、以及营养组成。在一个优选的实施方案中，该特性是选自：叶绿素合成、种子产量、胁迫耐受性、硝酸盐水平、氨基酸水平以及糖积累。在本发明的仍又另外一个实施方案中，这种受调节的特性是以下各项中的一项或多项的提高或改善：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、非生物胁迫耐受性、以及营养组成。在一个优选的实施方案中，这种受调节的特性是以下各项的提高或改善：叶绿素合成、种子产量、胁迫耐受性、硝酸盐水平、氨基酸水平以及糖积累。在另一个实施方案中，这种受调节的特性是以下各项的降低或减少：叶绿素合成、种子产量、胁迫耐受性、硝酸盐水平、氨基酸水平以及糖积累。

在一个具体的实施方案中，本发明涉及改善植物或植物细胞中氮利用的方法，该方法包括增强该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允许减少施用至植物的氮肥的量，并伴随着农

民成本和环境成本的降低，因为硝酸盐污染是许多农业区中的一个主要问题，显著地导致了淡水和海洋环境的退化。

植物或植物细胞可以来自人们希望在其中调节特性的任何植物。在本发明的一个实施方案中，植物细胞为双子叶植物、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。在本发明的又一实施方案中，单子叶植物选自玉米、小麦、大麦、燕麦、裸麦(rye)、粟、高粱(sorghum)、黑小麦(triticale)、黑麦属(secale)、单粒小麦(einkorn)、斯佩耳特小麦(spelt)、双粒小麦(emmer)、画眉草(teff)、蜀黍(milo)、亚麻、格兰马草(gramma grass)、磨擦草属物种(Tripsacum sp.)和玉米草(teosite)。

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

在本发明的一个实施方案中，对植物细胞中的GATA转录因子基因的表达水平进行调节的活性物质包括：

- (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、其片段或结构域；
- (c) 与(a)或(b)具有基本相似性的核苷酸序列；
- (d) 能够与(a)、(b)或(c)进行杂交的核苷酸序列；
- (e) 与(a)、(b)、(c)或(d)互补的核苷酸序列；或
- (f) 核苷酸序列，它是(a)、(b)、(c)或(d)的反向互补序列。

在一个具体的实施方案中，该活性物质增强了GATA转录因子的表达水平；并包括核酸分子，该核酸分子编码了GATA转录因子。在本发明的另外一个实施方案中，这种核酸分子包括SEQ ID NO: 1的OsGATA11基因的序列或其功能片段。在本发明的仍又另外一个实施方案中，这种核酸分子包括序列，该序列在中度严格性条件下与SEQ ID NO: 1的OsGATA11基因或其功能片段进行杂交。在本发明的另一个实施方案中，这种核酸分子来源于SEQ ID NO: 1的GATA转录因子的核苷酸序列；并且具有包括在植物中对表达特异的密码子的核苷酸序列。

在另一个具体的实施方案中，该活性物质抑制或降低了 GATA 转录因子的表达水平；并包括核酸分子，该核酸分子抑制了该 GATA 转录因子的表达。在一个实施方案中，这种核酸分子是用于 RNA 干扰 (RNAi) 的干扰性 RNA 分子。在另一个实施方案中，这种核酸分子是反义分子。

在本发明的另外一个实施方案中，对植物细胞中的 GATA 转录因子基因的表达水平进行调节的活性物质包括：

- (a) SEQ ID NO: 2 中所列出的多肽序列，或其功能片段、结构域、重复、或嵌合体；
- (b) 与 (a) 具有基本相似性的多肽序列；
- (c) 多肽序列，所述多肽序列由核苷酸序列进行编码，该核苷酸序列与 SEQ ID NO: 1 中所列出的核苷酸序列、或其功能片段或结构域、或与其它互补的序列，相同或具有基本相似性；或
- (d) 多肽序列，该多肽序列由核苷酸序列进行编码，该核苷酸序列能够在中度严格性条件下与 SEQ ID NO: 1 中所列出的核苷酸序列或与之互补的序列进行杂交。

在本发明的一个实施方案中，当活性物质是核酸序列时，该核酸序列表达于植物的特定位置或组织中。该位置或组织是（例如，但不限于）表皮、根、维管组织、分生组织、形成层、皮层、髓、叶和/或花。在一个备选的实施方案中，位置或组织为种子

本发明的实施方案还涉及用于调节植物细胞中特性的改组的核酸分子的用途，所述改组的核酸分子含有大量核苷酸序列片段，其中至少一条片段编码 GATA 转录因子且其中大量序列片段中至少两条是从 5' 到 3' 的顺序，这不是核酸中天然存在的大量片段的顺序。在一个具体的实施方案中，含有大量核苷酸序列片段的改组的核酸分子中所有片段来自单个基因。在一个更具体的实施方案中，大量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，改组的核酸与启动子序列有效连接。另一更具体的实施方案是使用嵌合的多核苷酸用于调节植物细胞中的特征，所述嵌合的多核苷酸包含与改组的核酸有效连接的启动子序列。在一个更具体的实施方

案中，改组的核酸包含在宿主细胞内。在本发明的另外一个具体的实施方案中，对 GATA 转录因子进行编码的片段由以下各项构成或包括以下各项：

- (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、其片段或结构域；
- (c) 与 (a) 或 (b) 具有基本相似性的核苷酸序列；
- (d) 能够与 (a)、(b) 或 (c) 进行杂交的核苷酸序列；
- (e) 与 (a)、(b)、(c) 或 (d) 互补的核苷酸序列；或
- (f) 核苷酸序列，它是 (a)、(b)、(c) 或 (d) 的反向互补序列。

本发明的实施方案还考虑了表达盒用于调节植物细胞中的特性的用途，该表达盒包括启动子序列，该启动子序列与分离的核酸有效连接，该分离的核酸编码了 GATA 转录因子。在本发明的实施方案中，对 GATA 转录因子进行编码的分离的核酸由以下各项构成或包括以下各项：

- (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、其片段或结构域；
- (c) 与 (a) 或 (b) 具有基本相似性的核苷酸序列；
- (d) 能够与 (a)、(b) 或 (c) 进行杂交的核苷酸序列；
- (e) 与 (a)、(b)、(c) 或 (d) 互补的核苷酸序列；或
- (f) 核苷酸序列，它是 (a)、(b)、(c) 或 (d) 的反向互补序列。

进一步涵盖于本发明之内的是重组载体用于调节植物细胞中的特性的用途，该重组载体包含表达盒，该表达盒包括启动子序列，该启动子序列与分离的核酸有效连接，该分离的核酸编码了 GATA 转录因子。在本发明的实施方案中，对 GATA 转录因子进行编码的分离的核酸由以下各项构成或包括以下各项：

- (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、其片段或结构域；
- (c) 与 (a) 或 (b) 具有基本相似性的核苷酸序列；
- (d) 能够与 (a)、(b) 或 (c) 进行杂交的核苷酸序列；

- (e) 与 (a)、(b)、(c) 或 (d) 互补的核苷酸序列；或
- (f) 核苷酸序列，它是 (a)、(b)、(c) 或 (d) 的反向互补序列。

还涵盖了根据本公开的包含表达盒的植物细胞的用途，以及包含这些植物细胞的植物的用途。

在一个实施方案中，该表达盒表达于整个植物中。在另一个实施方案中，该表达盒表达于植物的一个特定位置或组织中。在一个具体的实施方案中，该位置或组织可以是（例如）表皮、根、维管组织、分生组织、形成层、皮层、髓、叶和花。在一个备选的具体的实施方案中，该位置或组织是种子。

本发明的实施方案还提供了种子以及从植物分离出的产物用于调节植物细胞中的特性的用途，它们包含表达盒，该表达盒包括启动子序列，该启动子序列与分离的核酸有效连接，该分离的核酸编码了根据本发明的 GATA 转录因子基因。

在一个具体的实施方案中，表达载体包含一个或多个元件，例如但不限于启动子增强子序列、选择标记物序列，复制起点、表位标签编码序列，或亲和纯化标签编码序列。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、遍在蛋白和菜豆蛋白启动子。在另一实施方案中，启动子可在植物中工作，更特别地为组成型或诱导性启动子。在另一具体的实施方案中，选择标记物序列编码抗生素抗性基因。在另一具体的实施方案中，表位标签序列编码 V5、肽 Phe-His-His-Thr-Thr、血凝素或谷胱甘肽-S-转移酶。在另一具体的实施方案中，亲和纯化标签序列编码多聚氨基酸序列或多肽。在一个更具体的实施方案中，多聚氨基酸序列为多聚组氨酸。在一个更具体的实施方案中，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，亲和纯化标签序列包含内含肽编码序列。

在一个具体的实施方案中，表达载体是真核生物表达载体或原核生物表达载体。在一个更具体的实施方案中，真核生物表达载体包含组织特异的启动子。更特别地，表达载体可在植物中工作。

本发明的实施方案还涉及通过下述方法修饰的植物，所述方法包括向植物中引入核酸，其中核酸可以在植物中以有效影响修饰的量表达。该修饰可以是以下一种或多种目的性状的提高或降低。该修饰可包括基因的过量表达、低表达、反义调节、有义阻抑、诱导性表达、诱导性阻遏、或诱导性调节。在本发明的一个实施方案中，修饰涉及目的性状例如氮利用的提高或改进。

本发明的实施方案提供了从拟南芥中分离的核苷酸和氨基酸序列。特别地，本发明涉及糖感知所需要的调节氮的 GATA 转录因子基因。

本发明的实施方案涉及分离的核酸，该核酸包括核苷酸序列或由其构成，该核苷酸序列包括：

- (a) SEQ ID NO: 1 中所列出的核苷酸序列，或其片段或结构域；
- (b) 与 (a) 具有基本相似性的核苷酸序列；
- (c) 能够与 (a) 进行杂交的核苷酸序列；
- (d) 与 (a)、(b)、或(c) 互补的核苷酸序列；或
- (e) 核苷酸序列，它是 (a)、(b)、或(c) 的反向互补序列。

在一个具体的实施方案中，这种基本相似性是与如 SEQ ID NO: 1 所列出的核苷酸序列、其片段或结构域相比至少约 65% 同一性，特别地约 80% 同一性，特别地 90%，以及更特别地至少约 95% 序列同一性。

在一个实施方案中，与 SEQ ID NO: 1 的核苷酸序列、其片段或结构域具有基本相似性的序列来自植物。在一个具体的实施方案中，这种植物是双子叶植物。在一个更具体的实施方案中，这种双子叶植物是选自：大豆、烟草或棉花。在另一个具体的实施方案中，这种植物是裸子植物。在另一个具体的实施方案中，这种植物是单子叶植物。在一个更具体的实施方案中，这种单子叶植物是谷物。在一个更具体的实施方案中，这种谷物可以是（例如）玉米、小麦、大麦、燕麦、裸麦(rye)、粟、高粱(sorghum)、黑小麦(triticale)、黑麦属(secale)、单粒小麦(einkorn)、斯佩耳特小麦(spelt)、双粒小麦(emmer)、画眉草(teff)、蜀黍(milo)、亚麻、格兰马草(gramma grass)、磨擦草属物种(Tripsacum sp.) 和玉米草(teosinte)。

在一个实施方案中，该核酸表达于植物的特定位置或组织中。该位置或组织是（例如，但不限于）表皮、根、维管组织、分生组织、形成层、皮层、髓、叶和花。在一个备选的实施方案中，该位置或组织是种子。在另一个实施方案中，该核酸编码了多肽，该多肽涉及功能，例如但不限于，碳、氮和/或硫代谢，氮利用、氮同化作用、光合作用、信号转导、细胞生长、繁殖、疾病抗性、非生物胁迫耐受性、营养组成、基因调节、和/或分化。

在一个具体的实施方案中，这种分离的核酸包括核苷酸序列或由其构成，该核苷酸序列能够与 SEQ ID NO: 1 中所列出的核苷酸序列或其片段或结构域进行杂交。在一个具体的实施方案中，杂交使该序列在中度或高度严格性的条件下形成一个双链体。本发明的实施方案还涵盖了核苷酸序列，该核苷酸序列与 SEQ ID NO: 1 的核苷酸序列或其片段或结构域互补。本发明的实施方案进一步涵盖了核苷酸序列，该核苷酸序列互补于与 SEQ ID NO: 1 的核苷酸序列或其片段或结构域具有基本相似性或能够与其进行杂交的核苷酸序列。

在一个具体的实施方案中，具有基本相似性的核苷酸序列是 SEQ ID NO: 1 的核苷酸序列、其片段或结构域的等位基因的变体。在一个备选的实施方案中，具有基本相似性的序列是天然存在的变体。在另一个备选的实施方案中，具有基本相似性的序列是 SEQ ID NO: 1 的核苷酸序列或其片段或结构域的多态变体。

在一个具体的实施方案中，这种分离的核酸包含多个区域，这些区域具有 SEQ ID NO: 1 的核苷酸序列或其外显子或结构域。

在一个具体的实施方案中，这种分离的核酸包含多肽的编码序列。在一个更具体的实施方案中，这种多肽的编码序列包含 20 个碱基对的核苷酸部分，该部分在序列上与 SEQ ID NO: 1 的核酸序列的一个连续的 20 个碱基对核苷酸部分相同。在一个更具体的实施方案中，这种多肽包含 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，这种多肽是植物多肽。在一个更具体的实施方案中，这种植物是双子叶植物。在一个

更具体的实施方案中，这种植物是裸子植物。在一个更具体的实施方案中，这种植物是单子叶植物。在一个更具体的实施方案中，这种单子叶植物是谷物。在一个更具体的实施方案中，这种谷物可以是（例如）玉米、小麦、大麦、燕麦、裸麦(rye)、粟、高粱(sorghum)、黑小麦(triticale)、黑麦属(secale)、单粒小麦(einkorn)、斯佩耳特小麦(spelt)、双粒小麦(emmer)、画眉草(teff)、蜀黍(milo)、亚麻、格兰马草(gramma grass)、磨擦草属物种(Tripsacum sp.)和玉米草(teositne)。

在一个实施方案中，这种多肽表达于整个植物中。在一个更具体的实施方案中，这种多肽表达于植物的特定位置或组织中。在一个更具体的实施方案中，该位置或组织可以是（例如）表皮、根、维管组织、分生组织、形成层、皮层、髓、叶和花。在一个最具体的实施方案中，该位置或组织是种子。

在一个具体的方案中，该分离的核酸的序列编码了多肽，该多肽可用于产生抗体，该抗体具有针对由 SEQ ID NO: 2 的核苷酸序列、或其片段或结构域所编码的多肽的免疫反应性。

在一个具体的实施方案中，具有基本相似性的序列包含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，这种缺失或插入小于约三十个核苷酸。在一个最具体的实施方案中，这种缺失或插入小于约五个核苷酸。

在一个具体的实施方案中，具有基本相似性的分离的核酸的序列包含至少一个密码子的取代或由其构成。在一个具体的实施方案中，这种取代是保守性的。

本发明的实施方案还涉及分离的核酸分子，该分离的核酸分子包括以下序列是由其构成：核苷酸序列、它的互补序列、或它的反向互补序列，所述序列编码了多肽，该多肽包括：

(a) SEQ ID NO: 2 的多肽序列，或它的片段、结构域、重复、或嵌合体；

(b) 与 (a) 具有基本相似性的多肽序列；

(c) 多肽序列, 所述多肽序列由核苷酸序列进行编码, 该核苷酸序列与 SEQ ID NO: 1 中的核苷酸序列、或其片段或结构域、或与它互补的序列, 相同或具有基本相似性;

(d) 多肽序列, 该多肽序列由核苷酸序列进行编码的, 该核苷酸序列能够在中度严格性条件下与 SEQ ID NO: 1 的核苷酸序列或与之互补的序列进行杂交; 或者

(e) (a)、(b)、(c) 或 (d) 的功能片段。

在另一个具体的实施方案中, 具有基本相似性的多肽是 SEQ ID NO: 2 的多肽序列、或其片段、结构域、重复或嵌合体的等位变体。在另一个具体的实施方案中, 这种分离的核酸包括多个区域, 这些区域来自多肽序列, 该多肽序列由核苷酸序列进行编码, 该核苷酸序列与 SEQ ID NO: 1 的核苷酸序列、或它的片段或结构域、或与它互补的序列相同或具有基本相似性。

在另一个具体的实施方案中, 这种多肽是 SEQ ID NO: 2 的多肽序列。在另一个具体的实施方案中, 这种多肽是功能片段或结构域。在又一个具体的实施方案中, 这种多肽是嵌合体, 其中该嵌合体可包括功能性蛋白质结构域, 包括结构域、重复、翻译后修饰位点、或其他特征。在一个更具体的实施方案中, 这种多肽是植物多肽。在一个更具体的实施方案中, 这种植物是双子叶植物。在一个更具体的实施方案中, 这种植物是裸子植物。在一个更具体的实施方案中, 这种植物是单子叶植物。在一个更具体的实施方案中, 这种单子叶植物是谷物。在一个更具体的实施方案中, 这种谷物可以是(例如)玉米、小麦、大麦、燕麦、裸麦(rye)、粟、高粱(sorghum)、黑小麦(triticale)、黑麦属(secale)、单粒小麦(einkorn)、斯佩耳特小麦(spelt)、双粒小麦(emmer)、画眉草(teff)、蜀黍(milo)、亚麻、格兰马草(gramma grass)、磨擦草属物种(Tripsacum sp.)和玉米草(teosinte)。

在一个具体的实施方案中, 这种多肽表达于植物的特定位置或组织中。在一个更具体的实施方案中, 该位置或组织可以是(例如)表皮、根、维管组织、分生组织、形成层、皮层、髓、叶和花。在另一个具体的实施方

案中，该位置或组织是种子。

在一个具体的实施方案中，这种多肽序列由核苷酸序列进行编码，该核苷酸序列与 SEQ ID NO: 1 的核苷酸序列或其片段或结构域或与它互补的序列具有基本相似性；所述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中，这种缺失或插入小于约三十个核苷酸。在一个最具体的实施方案中，这种缺失或插入小于约五个核苷酸。

在一个具体的实施方案中，多肽序列由核苷酸序列进行编码，该核苷酸序列与 SEQ ID NO: 1 的核苷酸序列、或其片段或结构域、或与它互补的序列具有基本相似性；该多肽序列包括至少一个密码子的取代。在一个更具体的实施方案中，这种取代是保守性的。

在一个具体的实施方案中，与 SEQ ID NO: 2 的多肽序列、或它的片段、结构域、重复、或嵌合体具有基本相似性的多肽序列包括至少一个氨基酸的缺失或插入。

在一个具体的实施方案中，与 SEQ ID NO: 2 的多肽序列、或它的片段、结构域、重复、或嵌合体具有基本相似性的多肽序列包括至少一个氨基酸的取代。

本发明的实施方案还涉及改组的核酸，该改组的核酸包含多个核苷酸序列片段，其中这些片段中的至少一个对应于 SEQ ID NO: 1 的核苷酸序列的区域，并且其中这些多个序列片段中的至少两个处于从 5' 到 3' 的顺序，该顺序不是这些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，在包含多个核苷酸序列片段的改组的核酸中的所有片段均来自单一基因。在一个更具体的实施方案中，这些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，这种改组的核酸被有效连接至启动子序列上。另一个更具体的实施方案是嵌合的多核苷酸，该多核苷酸包括有效连接至该改组的核酸的启动子序列。在一个更具体的实施方案中，这种改组的核酸包含于宿主细胞之内。

本发明的实施方案还考虑了表达盒，该表达盒包括启动子序列，该启动子序列有效连接至分离的核酸上，该分离的核酸包含核苷酸序列，该核

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、其片段或结构域;
- (c) 与 (a) 或 (b) 具有基本相似性的核苷酸序列;
- (d) 能够与 (a)、(b) 或 (c) 进行杂交的核苷酸序列;
- (e) 与 (a)、(b)、(c) 或 (d) 互补的核苷酸序列; 或
- (f) 核苷酸序列, 它是 (a)、(b)、(c) 或 (d) 的反向互补序列。

进一步涵盖于本发明之内的是重组载体, 该重组载体包括根据本发明的实施方案的表达盒。还涵盖了根据本公开的包含表达盒的植物细胞, 以及包含这些植物细胞的植物。在一个具体的实施方案中, 这种植物是双子叶植物。在一个更具体的实施方案中, 这种双子叶植物是选自: 大豆、烟草或棉花。在另一个具体的实施方案中, 这种植物是裸子植物。在另一个具体的实施方案中, 这种植物是单子叶植物。在一个更具体的实施方案中, 这种单子叶植物是谷物。在一个更具体的实施方案中, 这种谷物可以是(例如)玉米、小麦、大麦、燕麦、裸麦(rye)、粟、高粱(sorghum)、黑小麦(triticale)、黑麦属(secale)、单粒小麦(einkorn)、斯佩耳特小麦(spelt)、双粒小麦(emmer)、画眉草(teff)、蜀黍(milo)、亚麻、格兰马草(gramma grass)、磨擦草属物种(Tripsacum sp.)和玉米草(teosinte)。

在一个实施方案中, 这种表达盒表达于整个植物中。在另一个实施方案中, 这种表达盒表达于植物的特定位置或组织中。在一个具体的实施方案中, 该位置或组织可以是(例如)表皮、根、维管组织、分生组织、形成层、皮层、髓、叶和花。在一个备选的具体的实施方案中, 该位置或组织是种子。

在一个实施方案中, 这种表达盒涉及功能, 例如但不限于, 碳、氮和/或硫代谢, 氮利用、氮同化作用、光合作用、信号转导、细胞生长、繁殖、疾病抗性、非生物胁迫耐受性、营养组成、基因调节、和/或分化。在一个更具体的实施方案中, 这种嵌合的多肽涉及功能, 例如氮利用、非生物胁迫耐受性、增强的产量、疾病抗性和/或营养组成。

在一个实施方案中，这种植物包含针对植物的表型或可测量的特性的修饰，该修饰可归因于表达盒中所包含的至少一个基因的表达。在一个具体的实施方案中，这种修饰可以是（例如）碳、氮和/或硫代谢，氮利用、氮同化作用、光合作用、信号转导、细胞生长、繁殖、疾病抗性、非生物胁迫耐受性、营养组成、基因调节、和/或分化。

本发明的实施方案还提供了种子以及从植物分离出的产物，它们包含表达盒，该表达盒包括启动子序列，该启动子序列有效连接至分离的核酸上，该分离的核酸包含核苷酸序列，该核苷酸序列包括：

- (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、其片段或结构域；
- (c) 与 (a) 或 (b) 具有基本相似性的核苷酸序列；
- (d) 能够与 (a)、(b) 或 (c) 进行杂交的核苷酸序列；
- (e) 与 (a)、(b)、(c) 或 (d) 互补的核苷酸序列；或
- (f) 核苷酸序列，它是根据本公开的 (a)、(b)、(c) 或 (d) 的反向互补序列。

在一个具体的实施方案中，分离的产物包括酶、营养蛋白质、结构蛋白质、氨基酸、脂类、脂肪酸、多糖、糖、醇、生物碱、类胡萝卜素、丙素（propanoid）、类固醇、色素、维生素以及植物激素。

本发明的实施方案还涉及分离的产物，这些分离的产物通过分离的核酸的表达而产生，该分离的核酸包含核苷酸序列，该核苷酸序列包括：

- (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) 具有基本相似性的核苷酸序列；
- (d) 能够与 (a) 或 (b) 进行杂交的核苷酸序列；
- (e) 与 (a)、(b)、(c) 或 (d) 互补的核苷酸序列；或
- (f) 核苷酸序列，它是根据本公开的 (a)、(b)、(c) 或 (d) 的反向互补序列。

在一个具体的实施方案中，这种产物在植物中产生。在另一个具体的

实施方案中，这种产物在细胞培养物中产生。在另一个具体的实施方案中，这种产物在无细胞的系统中产生。在另一个具体的实施方案中，这种产物包括酶、营养蛋白质、结构蛋白质、氨基酸、脂类、脂肪酸、多糖、糖、醇、生物碱、类胡萝卜素、丙素、类固醇、色素、维生素以及植物激素。

在一个具体的实施方案中，这种产物是包含 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，这种蛋白是转录因子。

本发明的实施方案进一步涉及分离的多核苷酸，该分离的多核苷酸包括具有至少 10 个碱基的核苷酸序列，该序列与 SEQ ID NO: 1 的任何序列的区域相同、互补、或基本相似，并且其中该多核苷酸适于许多用途中的任何。

在一个具体的实施方案中，这种多核苷酸作为染色体标记物而使用。在另一个具体的实施方案中，这种多核苷酸作为 RFLP 分析的标记物而使用。在另一个具体的实施方案中，这种多核苷酸作为数量性状关联育种的标记物而使用。在另一个具体的实施方案中，这种多核苷酸作为标记物辅助育种的标记物而使用。在另一个具体的实施方案中，这种多核苷酸作为双杂交系统中的一个钓饵(bait)序列而使用，以鉴定多肽的编码序列，所述多肽与由该钓饵序列所编码的多肽进行相互作用。在另一个具体的实施方案中，这种多核苷酸用作为基因分型或鉴定个体或个体的群体的诊断性指示物。在另一个具体的实施方案中，这种多核苷酸用于遗传分析以鉴定基因或外显子的边界。

本发明的实施方案还涉及表达载体，该表达载体包括核酸分子或由其构成，该核酸分子包括：

- (a) 编码如 SEQ ID NO:2 中所列出的多肽的核酸；
- (b) SEQ ID NO: 1 的片段、一个或多个结构域、或特征性区域；

或

(c) 与异源序列相组合的在 SEQ ID NO:1 中所列出的完整的核酸序列、或它的片段。

在一个具体的实施方案中，表达载体包含一个或多个元件，例如但不

限于启动子增强子序列、选择标记物序列、复制起点、表位标签编码序列，或亲和纯化标签编码序列。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、遍在蛋白和菜豆蛋白启动子。在另一实施方案中，启动子可在植物中工作，更特别地为组成型或诱导性启动子。在另一具体的实施方案中，选择标记物序列编码抗生素抗性基因。在另一具体的实施方案中，表位标签序列编码 V5、肽 Phe-His-His-Thr-Thr、血凝素或谷胱甘肽-S-转移酶。在另一具体的实施方案中，亲和纯化标签序列编码多聚氨基酸序列或多肽。在一个更具体的实施方案中，多聚氨基酸序列为多聚组氨酸。在一个更具体的实施方案中，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，亲和纯化标签序列包含内含肽编码序列。

在一个具体的实施方案中，表达载体是真核生物表达载体或原核生物表达载体。在一个更具体的实施方案中，真核生物表达载体包含组织特异的启动子。更特别地，表达载体可在植物中工作。

本发明的实施方案还涉及包含核酸构建体或由其组成的细胞，所述核酸构建体包含表达载体和与异源序列组合的下述核酸，所述核酸包括：编码 SEQ ID NO: 2 所列多肽的核酸，或 SEQ ID NO:1 所列的核酸，或其区段。

在一个具体的实施方案中，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，细胞是植物细胞。在一个更具体的实施方案中，多肽在植物的特异位置或组织中表达。在一个最具体的实施方案中，该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、叶和花。在备选的最具体的实施方案中，该位置或组织是种子。在一个具体的实施方案中，多肽涉及下述功能，例如碳、氮和/或硫代谢、氮利用、氮同化作用、光合作用、信号转导、细胞生长、繁殖、疾病抗性、非生物胁迫耐受性、营养组成、基因调控和/或分化。

本发明的实施方案还涉及由本公开的分离的核酸分子所编码的多肽，这些多肽包括下述多肽，该多肽包含多肽序列，该多肽序列由分离的核酸

进行编码，该分离的核酸包含核苷酸序列，该核苷酸序列包括：

- (a) SEQ ID NO: 1 中所列出的核苷酸序列、或其外显子或结构域；
- (b) 与 (a) 具有基本相似性的核苷酸序列；
- (c) 能够与 (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) 核苷酸序列，它是 (a)、(b)、或 (c) 的反向互补序列；
- (f) 或其功能片段。

多肽，该多肽包含由分离的核酸编码的多肽序列，该分离的核酸包含核苷酸序列、它的互补序列、或它的反向互补序列，它们编码包括多肽序列的多肽，所述多肽序列包括：

- (a) SEQ ID NO: 2 中所列出的多肽序列，或其结构域、重复、或嵌合体；
- (b) 与 (a) 具有基本相似性的多肽序列；
- (c) 多肽序列，该多肽序列由核苷酸序列进行编码，该核苷酸序列与 SEQ ID NO: 1 所列出的核苷酸序列、或其外显子或结构域、或与它互补的序列相同或具有基本相似性；
- (d) 多肽序列，该多肽序列由核苷酸序列进行编码，该核苷酸序列能够在中度严格性条件下与在 SEQ ID NO: 1 所列出的核苷酸序列或与之互补的序列进行杂交；或
- (e) (a)、(b)、(c) 或 (d) 的功能片段；
- (f) 或其功能片段。

本发明的实施方案考虑了多肽，该多肽包含多肽序列，该多肽序列由分离的核酸进行编码，该分离的核酸包括改组的核酸，该改组的核酸包含多个核苷酸序列片段，其中这些片段中的至少一个对应于 SEQ ID NO: 1 所列出的核苷酸序列的区域，并且其中这些多个序列片段中的至少两个处于从 5' 到 3' 的顺序，该顺序不是这些多个片段在核酸中、或核酸的功能片段中天然存在的顺序。

本发明的实施方案考虑了多肽，该多肽包含多肽序列，该多肽序列由

分离的多核苷酸进行编码,该分离的多核苷酸包含具有至少 10 个碱基的核苷酸序列,该序列与 SEQ ID NO: 1 的序列或其功能片段中的任何区域、相同、互补、或基本上相似,并且其中该多核苷酸适于下述用途,包括:

(a)作为染色体标记物用于鉴定天然或人工染色体上相应或互补的多核苷酸位置的用途;

(b)作为 RFLP 分析标记物的用途;

(c)作为数量性状关联育种的标记物的用途;

(d)作为标记物-辅助育种的标记物的用途;

(e)作为钓饵序列在双杂交体系中用于鉴定编码多肽的序列的用途,所述多肽与钓饵序列编码的多肽相互作用;

(f)作为对个体或个体群进行基因分型或鉴定的诊断指示物的用途;或

(g)用于鉴定基因或外显子边界的遗传分析的用途。

本发明的实施方案还考虑了分离的多肽,该分离的多肽包含多肽序列,该多肽序列包括:

(a) SEQ ID NO: 2 所列出的多肽序列、或它的外显子或结构域;

(b)与 (a) 具有基本相似性的多肽序列;

(c)多肽序列,该多肽序列由核苷酸序列进行编码,该核苷酸序列与 SEQ ID NO: 1 的核苷酸序列、或其外显子或结构域、或与它互补的序列相同或具有基本相似性;

(d)多肽序列,该多肽序列由核苷酸序列进行编码,该核苷酸序列能够在中度严格性条件下与 SEQ ID NO: 1 中所列出的核苷酸序列、或与其互补的序列进行杂交;或

(e) (a)、(b)、(c) 或 (d) 的功能片段。

在一个具体的实施方案中,这种基本相似性是至少约 65% 的同一性。在一个更具体的实施方案中,这种基本相似性是至少约 80% 的同一性。在一个最具体的实施方案中,这种基本相似性是至少约 95% 的同一性。在一个具体的实施方案中,这种基本相似性与列表中的任何一个所列出的最接近的同源序列的百分比同一性相比至少大出三个百分点。

在一个具体的实施方案中，具有基本相似性的序列来自植物。在一个更具体的实施方案中，这种植物是双子叶植物。在一个更具体的实施方案中，这种植物是裸子植物。在一个更具体的实施方案中，这种植物是单子叶植物。在一个更具体的实施方案中，这种单子叶植物是谷物。在一个更具体的实施方案中，这种谷物可以是（例如）玉米、小麦、大麦、燕麦、裸麦(rye)、粟、高粱(sorghum)、黑小麦(triticale)、黑麦属(secale)、单粒小麦(einkorn)、斯佩耳特小麦(spelt)、双粒小麦(emmer)、画眉草(teff)、蜀黍(milo)、亚麻、格兰马草(gramma grass)、磨擦草属物种(Tripsacum sp.)和玉米草(teosinte)。

在一个具体的实施方案中，多肽表达于植物的特定位置或组织中。在一个更具体的实施方案中，这种位置或组织可以是（例如）表皮、根、维管组织、分生组织、形成层、皮层、髓、叶、和花。在另一个具体的实施方案中，这种位置或组织是种子。在一个具体的实施方案中，这肽涉及功能，例如碳、氮和/或硫代谢、氮利用、氮同化作用、光合作用、信号转导、细胞生长、繁殖、疾病抗性、非生物胁迫耐受性、营养组成、基因调节、和/或分化。

在一个具体的实施方案中，以下多肽序列的杂交使该序列在中度或高度严格条件下形成双链体，所述多肽序列由核苷酸序列进行编码，该核苷酸序列与 SEQ ID NO: 1 中所列出的核苷酸序列、或其外显子或结构域、或与它互补的序列相同或具有基本相似性；或者所述多肽序列由核苷酸序列进行编码，该核苷酸序列在中度严格条件下能够与 SEQ ID NO: 1 中所列出的核苷酸序列、或与其互补的序列进行杂交。

在一个具体的实施方案中，与 SEQ ID NO: 2 中所列出的多肽序列、或其外显子或结构域具有基本相似性的多肽是 SEQ ID NO: 2 中所列出的多肽序列的等位变体。在另一个具体的实施方案中，与 SEQ ID NO: 2 中所列出的多肽序列、或其外显子或结构域具有基本相似性的多肽是 SEQ ID NO: 2 中所列出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 中所列出的多肽序列、或其外显子或结构域具有基

本相似性的多肽是 SEQ ID NO: 2 中所列出的多肽序列的多态变体。

在一个备选的具体实施方案中，具有基本相似性的序列包含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，这种缺失或插入小于约十个氨基酸。在一个最具体的实施方案中，这种缺失或插入小于约三个氨基酸。

在一个具体的实施方案中，具有基本相似性的序列编码了至少一个氨基酸中的取代。

还包括生产包含修饰的植物的方法，其包括步骤：(1)提供核酸，其为含有核苷酸序列的分离的核酸，所述核苷酸序列包括：

- (a)如 SEQ ID NO:1 所示的核苷酸序列，或其外显子或结构域，
- (b)与(a)具有基本相似性的核苷酸序列；
- (c) 能够与(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；

和(2)将核酸引入植物，其中所述核酸在所述植物中以有效影响修饰的量表达。在一个实施方案中，修饰包括植物中改变的特性，其中所述特性对应于被引入植物的核酸。在其他具体的实施方案中，该特征对应于碳、氮和/或硫代谢、氮利用、氮同化作用、光合作用、信号转导、细胞生长、繁殖、疾病抗性、非生物胁迫耐受、营养组成、基因调节和/或分化。

在另一实施方案中，修饰包括提高或降低的表达，或植物产物的累积。特别地，该产物是植物的天然产物。同等特别地，该产物是植物的新产物或改变的产物。特别地，该产物包括 GATA 转录因子。

本文公开的发明还包括制备重组蛋白质的方法，其包括步骤：

- (a)在合适的培养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核酸，所述核酸包括：编码如 SEQ ID NO: 2 所示蛋白质的核酸，或 SEQ ID NO:1 所示核酸序列，或其区段；和
- (b)从重组细胞中分离其表达的重组蛋白质。

本发明的实施方案提供了制备重组蛋白质的方法，其中表达载体包含

一个或多个元件，包括启动子增强子序列、选择标记物序列、复制起点、表位标签编码序列，和亲和纯化标签编码序列。在一个具体的实施方案中，核酸构建体包含表位标签编码序列，且分离步骤包括使用对该表位标签特异的抗体。在另一具体的实施方案中，核酸构建体含有多聚氨基酸编码序列，且分离步骤包括使用包含多聚氨基酸结合物质的树脂，特别是其中多聚氨基酸为多聚组氨酸且多聚氨基酸结合树脂为镍-带电琼脂糖树脂。在另一具体的实施方案中，核酸构建体含有多肽编码序列，且分离步骤包括使用含多肽结合物质的树脂，特别是当多肽为壳多糖结合结构域且树脂含有壳多糖-琼脂糖凝胶（sepharose）时。

本发明的实施方案还涉及通过下述方法修饰的植物，所述方法包括向植物中引入核酸，其中该核酸可在植物中以有效影响修饰的量表达。该修饰可以是例如碳、氮和/或硫代谢、氮利用、氮同化作用、光合作用、信号转导、细胞生长、繁殖、疾病抗性、非生物胁迫耐受、营养组成、基因基因调节和/或分化。在一个实施方案中，经修饰的植物具有对除草剂、胁迫或病原体的提高或降低的抗性。在另一实施方案中，经修饰的植物具有对光、水、氮或痕量元素的增加或减轻的需求。在另一实施方案中，以植物蛋白质级分的比例计，经修饰的植物富含必需氨基酸。该蛋白质级分可以是例如总种子蛋白、可溶蛋白质、不溶蛋白质、可用水提取的蛋白质和脂质结合蛋白质。修饰可包括基因的过量表达、低表达、反义调节、有义阻抑、诱导性表达、诱导性阻遏、或诱导性调节。

本发明还涉及来自经修饰的植物的种子，或经修饰的植物的分离产物，其中该产物可以是酶、营养蛋白质、结构蛋白质、氨基酸、脂质、脂肪酸、多糖、糖、醇、生物碱、类胡萝卜素、丙素（propanoid）、类固醇、色素、维生素和植物激素。

上述“发明概述”列举了本发明的若干个实施方案，并且在许多情况下列举了这些实施方案的变更和置换。该概述仅是大量和变化的实施方案的示例。提到给定实施方案的一个或多个特别特征同样是示例性的。一般可存在具有或不具有所述一个或多个特征的这样的实施方案；同样，这些

特征可应用于本发明的其他实施方案，无论所述实施方案在概述中是否列出。为了避免过度重复，该概述不列举或提出这类特征的所有可能的组合。

为了概述本发明和达到的超出现有技术的优点，上文已描述了本发明的某些目标和优点。当然，应当理解对本发明的任何具体的实施方案而言，不必须达到所有这些目标和优点。因此，例如本领域技术人员会知道本发明可以以下述方式进行，所述方式达到或最优化本文教导的一个优点或一组优点，而不必须达成本文可教导或提出的其他目标或优点。

下面具体实施方案的详细描述使得本发明的其他方面、特征和优点变得显而易见。

### 附图说明

图 1 和 SEQ ID NO: 1 示出了全长 OsGATA11 的核酸序列。

图 2 和 SEQ ID NO: 2 示出了 OsGATA11 的氨基酸序列。

图 3 示出了 At4g26150 的氨基酸序列 (SEQ ID NO: 7) 与它的稻直向同源物 OsGATA11 (SEQ ID NO: 2) 的比对。

图 4A 和 4B 示出了过量表达 OsGATA11 的植物的表型。

图 5A 和 B 示出了受 *OsGATA11* 基因的表达所影响的叶绿素水平。

图 6A 和 B 示出了过量表达 OsGATA11 的植物的种子产量。

图 7 是一些图片，这些图片示出了过量表达 *OsGATA11* 的植物中对胁迫更强的抗性。

图 8 是一些图，这些图示出了糖积累是如何通过调节 OsGATA11 的表达而受到影响的：8A：葡萄糖水平；8B：果糖水平；8C：蔗糖水平。

图 9 是一些图，这些图示出了在 OsGATA11 转基因植物中氮状况如何受到调节：9A：硝酸盐水平；9B：氨基酸水平；9C：蛋白质水平。

### 定义

为了清楚起见，如下定义说明书中使用的某些术语：

“关联/有效连接”是指两条核酸序列物理上或功能上相关联。例如，

如果启动子或调节 DNA 序列与编码 RNA 或蛋白质的 DNA 序列有效连接，或位置使得调节 DNA 序列会影响编码或结构 DNA 序列的表达水平，则该启动子或调节 DNA 序列被称作与该 DNA 序列“关联”。

“嵌合构建体”是重组的核酸序列，其中启动子或调节核酸序列与核酸序列（所述核酸序列编码 mRNA 或被表达为蛋白质）有效连接或关联，使得调节核酸序列能够调节关联的核酸序列的转录或表达。嵌合构建体的调节核酸序列通常不与天然发现的关联的核酸序列有效连接。

“辅因子”是酶催化的反应中所需的天然反应物，如有机分子或金属离子。辅因子为例如 NAD(P)、维生素 B<sub>2</sub>（包括 FAD 和 FMN）、叶酸、钼蝶呤、维生素 B<sub>1</sub>(thiamin)、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、甲基萘醌类。任选地，辅因子可以再生和再使用。

“编码序列”是被转录为 RNA 如 mRNA、rRNA、tRNA、snRNA、有义 RNA 或反义 RNA 的核酸序列。特别地，该 RNA 随后在生物中被翻译产生蛋白质。

互补的：“互补的”是指包含反向平行核苷酸序列的两条核苷酸序列，其能够通过反向平行的核苷酸序列中互补的碱基残基之间形成氢键而彼此配对。

酶活性：在本文中表示酶催化底物转化为产物的能力。酶的底物包括酶的天然底物，但是也包括天然底物的类似物，所述类似物也可以被酶转化为产物或转化为产物的类似物。例如通过测定某时间段后反应中的产物量，或通过测定某时间段后反应混合物中剩余的底物量来测量酶活性。还通过测定某时间段后反应混合物中剩余的未使用的反应辅因子的量，或通过测定某时间段后反应混合物中使用的辅因子的量来测量酶活性。还通过测定某时间段后反应混合物中剩余的自由能供体或能量富集分子（例如 ATP、磷酸烯醇丙酮酸、乙酰磷酸或磷酸肌酸）的量，或通过测定某时间段后反应混合物中使用的自由能供体或能量富集分子（例如 ADP、丙酮酸、乙酸或肌酸）的量来测量酶活性。

**表达盒:** 本文使用的“表达盒”表示能够指导具体核苷酸序列在适当宿主细胞组中表达的核酸分子, 其包含与目的核苷酸序列有效连接的启动子, 所述目的核苷酸序列与终止信号有效连接。其一般还包含正确翻译核苷酸序列所需的序列。编码区通常编码目的蛋白质, 但是也可编码有义或反义方向上的目的功能性 RNA, 例如反义 RNA 或非翻译的 RNA。包含目的核苷酸序列的表达盒可以是嵌合的, 这表示其至少一个组件对于其至少另一个组件而言是异源的。表达盒也可以是下述表达盒, 所述表达盒是天然存在的, 但是以适用于异源表达的重组形式获得。然而, 表达盒相对于宿主一般是异源的, 即表达盒的具体 DNA 序列在宿主细胞中天然不存在, 并且必须已通过转化事件被引入宿主细胞或宿主细胞的祖先中。表达盒中核苷酸序列的表达可以位于组成型启动子或诱导性启动子的控制下, 所述诱导性启动子仅在所述宿主细胞暴露于一些具体的外部刺激时起始转录。在多细胞生物如植物的情况下, 启动子也可以对具体的组织或器官或发育阶段是特异的。

本文与核酸或蛋白质序列相关使用的术语“功能片段”表示保留全长序列功能的序列片段或部分。

**基因:** 术语“基因”被广泛用于表示与生物学功能相关联的任何 DNA 区段。因此, 基因包括编码序列和/或其表达所需的调节序列。基因还包括非表达的 DNA 区段, 例如形成其他蛋白质的识别序列的 DNA 区段。基因可得自多种来源, 包括从目的来源克隆或从已知或预测的序列信息合成, 并可包括被设计为具有期望参数的序列。

**异源的/外源的:** 术语“异源的”和“外源的”在本文中涉及核酸序列(例如 DNA 序列)或基因使用时表示来自具体宿主细胞的外部来源, 或如果来自相同来源时, 表示对其原始形式进行了修饰。因此, 宿主细胞中的异源基因包括对具体宿主细胞是内源的, 但是已通过例如 DNA 改组的使用进行了修饰的基因。该术语还包括天然存在的 DNA 序列的非天然存在的多个拷贝。因此, 该术语是指对细胞是外源或异源的 DNA 区段, 或对细胞是同源的但是位于宿主细胞核酸中下述位置的 DNA 区段, 该元件

通常不存在于所述位置。表达外源 DNA 区段得到外源多肽。

“同源”核酸（例如 DNA）序列是与引入该序列的宿主细胞天然关联的核酸（例如 DNA）序列。

杂交：短语“与……特异杂交”是指在严格性条件下，当下述序列存在于复杂混合物（例如总细胞）DNA 或 RNA 中时，一个分子只与该具体核苷酸序列结合、形成双链体或杂交。“基本结合”是指探针核酸和靶核酸之间包含小量错配的互补杂交，可以通过降低杂交介质的严格性来调节所述错配，以达到靶核酸序列的期望检测。

抑制剂：使蛋白质如生物合成酶、受体、信号转导蛋白质、结构基因产物或运输蛋白的酶活性失活的化学物质。在本文中使用术语“除草剂”（或“除草化合物”）定义下述抑制剂，对任何发育阶段的植物应用该抑制剂，藉此该除草剂抑制植物的生长或杀死植物。

相互作用：相互作用的品质或状态使得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（拮抗剂）或增强（激动剂）的。

当核酸序列编码的多肽与参考核酸序列编码的多肽具有相同氨基酸序列时，该核酸序列与参考核酸序列是“同类编码”的。

等基因的：在遗传上等同的植物，只是因为存在或不存在异源 DNA 序列而不同。

分离的：在本发明的上下文中，分离的 DNA 分子或分离的酶是通过人的介入远离其天然环境并因此不是天然产物的 DNA 分子或酶。分离的 DNA 分子或酶可以以纯化的形式存在，或可存在于非天然的环境中，例如存在于转基因宿主细胞中。

成熟蛋白质：其中转运肽、信号肽和/或前肽部分已被去除的蛋白质。

最小启动子：可支持任何转录的最小的启动子部分，如 TATA 元件。在缺失上游激活时，最小启动子一般具有被大幅降低的启动子活性。存在合适的转录因子时，最小启动子发挥允许转录的作用。

修饰的酶活性：与植物中天然存在的酶活性不同的酶活性（即在缺失人对这类活性的直接或间接操作时天然存在的酶活性），其对抑制天然存

在的酶活性的抑制剂耐受。

**固有的(native):** 是指存在于未经转化的植物细胞基因组中的基因。

**天然存在的:** 术语“天然存在的”用于描述可在自然中发现的物体，其与由人工生产的不同。例如，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，所述蛋白质或核苷酸序列可从天然来源分离，并且未由人在实验室中有意地修饰。

**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多聚体。除非特别地限制，该术语包括含有天然核苷酸的已知类似物的核酸，其与参考核酸具有类似的结合特性，并以类似于天然存在的核苷酸的方式代谢。除非另有说明，具体的核酸序列还含蓄地包括其经保守修饰（例如简并密码子取代）的变体和互补序列以及明确指出的序列。特别地，可通过产生下述序列达成简并密码子取代，所述序列中一个或多个选定的（或所有）密码子的第三个位置被混合性碱基和/或脱氧肌苷残基取代(Batzer 等, *Nucleic Acid Res.* 19: 5081 (1991); Ohtsuka 等, *J. Biol. Chem.* 260: 2605-2608 (1985); Rossolini 等, *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”表示可读框。

**百分比同一性:** 在两条核酸或蛋白质序列的上下文中，短语“百分比同一性”或“百分比相同”是指针对最大对应进行比较和比对时，具有例如 60%，特别是 70%，更特别是 80%，仍然更特别是 90%，进一步更特别是 95%和最特别是至少 99%的核苷酸或氨基酸残基同一性的两条或多条序列或亚序列 (subsequence)，其使用以下序列比较算法之一测量或通过视觉检查测量。特别地，百分比同一性存在于长度为至少约 50 个残基的序列区域中，更特别地存在于至少约 100 个残基的区域中，最特别地，百分比同一性存在于至少约 150 个残基中。在一个特别特定的实施方案中，百分比同一性存在于编码区的全长中。

为了进行序列比较，通常一条序列发挥参考序列的作用，测试序列与

该参考序列比较。使用序列比较算法时，将测试和参考序列输入计算机中，如果需要的话指定亚序列坐标，并指定序列算法程序参数。序列比较算法随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同一性百分比。

可例如通过 Smith & Waterman, *Adv. Appl. Math.* 2: 482 (1981)的局部同源性算法，通过 Needleman & Wunsch, *J. Mol. Biol.* 48: 443 (1970)的同源性比对算法，通过 Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85: 2444 (1988)的搜索相似性方法，通过这些算法的计算机化执行 (Wisconsin Genetics 软件包中的 GAP、BESTFIT、FASTA 和 TFASTA, Genetics Computer Group, 575 Science Dr., Madison, WI) 或通过视觉检查 (一般参见 Ausubel 等, 下文)，对用于比较的序列进行最佳比对。

适用于测定序列同一性百分比和序列相似性百分比的算法的一个实例是 BLAST 算法, 其描述于 Altschul 等, *J. Mol. Biol.* 215: 403-410 (1990)中。公众可通过 National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>)获得运行 BLAST 分析的软件。该算法涉及首先通过鉴定查询序列中长度为 W 的短字长 (short word) 鉴定高分序列对(HSP)，所述短字长与数据库序列中相同长度的字长比对时匹配或满足一些正值的阈值分数 T。T 是指邻域字长分数阈值(Altschul 等, 1990)。这些初始的邻域字长匹配(hit)发挥起始下述搜索的种子的作用，所述搜索寻找含有它们的更长的 HSP。该字长匹配随后在两个方向上沿着各序列扩展，直到累积的比对分数能够被提高为止。对核苷酸序列而言，使用参数 M (一对匹配残基的奖励分数; 始终 $> 0$ ) 和 N (错配残基的罚分; 始终 $< 0$ ) 计算累积的分数。对氨基酸序列而言，使用评分矩阵计算累积分数。当累积的比对分数从其达到的最大值跌落数量 X、累积的分数由于一个或多个负分残基比对的累积而达到或低于零、或达到任一序列的末端时，停止字长匹配在各方向上的扩展。BLAST 算法参数 W、T 和 X 确定比对的灵敏度和速度。BLASTN 程序 (对核苷酸序列而言) 使用 11 的字长(W)、10 的预期(E)、100 的截断、M=5、N=-4 和两条链的比较作为默认值。对氨基

酸序列而言, BLASTP 程序使用 3 的字长(W)、10 的预期(E)和 BLOSUM62 评分矩阵作为默认值 (见 Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89: 10915 (1989))。

除了计算序列同一性百分比外, BLAST 算法还在两个序列之间进行相似性的统计学分析 (见例如 Karlin & Altschul, Proc. Nat'l. Acad. Sci. USA 90: 5873-5787 (1993))。由 BLAST 提供的相似性的一种度量为最小概率和(P(N)), 其提供了两条核苷酸或氨基酸序列之间偶然发生匹配的概率的指示。例如, 如果在测试核酸序列与参考核酸序列的比较中, 最小概率和小于约 0.1, 更特别地小于约 0.01, 和最特别地小于约 0.001, 则认为测试核酸序列与参考序列相似。

**前蛋白质:** 通常靶向细胞器 (如叶绿体) 并且仍包含其天然转运肽的蛋白质。

**纯化的:** 应用于核酸或蛋白质时, 术语“纯化的”表示核酸或蛋白质基本上不含其他分子组件, 所述分子组件在天然状态下与所述核酸或蛋白质相连。尽管其可以是干燥的或在水性溶液中, 但是其特别地处于同质状态(homogeneous state)。通常使用分析化学技术如聚丙烯酰胺凝胶电泳或高效液相层析测定纯度和同质性。在制品中是优势种类的蛋白质是基本纯化的。术语“纯化的”表示核酸或蛋白质在电泳凝胶中基本上给出一个条带。具体地, 这表示核酸或蛋白质至少约 50%纯净, 更特别地至少约 85%纯净, 和最特别地至少约 99%纯净。

当来自两条核酸各自的序列在子代核酸中组合时, 该两条核酸是“重组的”。当核酸均为重组的底物时, 这两条序列是“直接”重组的。当序列使用中间体如交换(cross-over)寡核苷酸重组时, 两条序列是“间接重组的”。对间接重组而言, 不多于一条序列是重组的真实底物, 并且在一些情况下, 序列均不是重组的底物。

“调节元件”是指涉及控制核苷酸序列表达的序列。调节元件包括与目的核苷酸序列有效连接的启动子和终止信号。它们一般还包括核苷酸序列适当翻译所需的序列。

**显著的提高:** 大于测量技术中固有误差的限度的酶活性提高, 特别是在存在抑制剂时野生型酶活性提高约 2 倍或更大, 更特别地提高约 5 倍或更大, 最特别地提高约 10 倍或更大。

**显著更少:** 表示酶反应的产物量被减少得多于测量技术中固有误差的限度, 特别是在缺失抑制剂时野生型酶活性减少约 2 倍或更大, 更特别地减少约 5 倍或更大, 最特别地减少约 10 倍或更大。

**特异的结合/免疫交叉反应性:** 两条核酸序列或蛋白质基本相同的指标是第一核酸编码的蛋白质与由第二核酸编码的蛋白质免疫杂交反应或特异结合。因此, 例如当两个蛋白质仅由保守取代区别时, 蛋白质一般与第二蛋白质是基本相同的。涉及蛋白质或肽时, 短语“与抗体特异(或选择性)结合”或“与……特异(或选择性)免疫反应”是指在存在蛋白质的异源群体和其他生物制品时决定蛋白质存在的结合反应。因此, 在指定的免疫测定条件下, 特定的抗体与具体的蛋白质结合, 并且不以显著的量与样品中存在的其他蛋白质结合。在这类条件下与抗体的特异结合可需要下述抗体, 所述抗体因其对具体蛋白质的特异性而被选择。例如, 可选择针对下述蛋白质产生的抗体获得与该蛋白质特异免疫反应而不与其他蛋白质(除多态变体以外)特异免疫反应的抗体, 所述蛋白质具有本发明的任何核酸序列编码的氨基酸序列。可使用多种免疫测定方式选择与具体蛋白质特异免疫反应的抗体。例如, 常规地使用固相 ELISA 免疫测定、Western 印迹、或免疫组织化学选择与蛋白质特异免疫反应的单克隆抗体。可用于测定特异免疫反应性的免疫测定方式和条件的描述见 Harlow 和 Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York “Harlow and Lane”)。特异或选择性的反应通常会背景信号或噪音的至少两倍, 更通常是背景的多于 10 到 100 倍。

在核酸杂交实验如 Southern 和 Northern 杂交的语境中, “严格性杂交条件”和“严格性杂交洗涤条件”是序列依赖性的, 并在不同的环境参数下不同。更长的序列在更高的温度下特异杂交。核酸杂交的广泛指南见 Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular*

**Biology-Hybridization with Nucleic Acid Probes**, 第 1 部分第 2 章, "Overview of principles of hybridization and the strategy of nucleic acid probe assays" Elsevier, New York. 一般地, 高严格性杂交和洗涤条件被选择为比确定的离子强度和 pH 下特异序列的热解链温度( $T_m$ )低约  $5^\circ\text{C}$ 。通常在“严格性条件”下, 探针会与其靶亚序列杂交, 但不与其他序列杂交。

$T_m$  是 50% 的靶序列与优选的匹配的探针杂交的温度 (在确定的离子强度和 pH 下)。非常严格的条件被选择为等于具体探针的  $T_m$ 。在 Southern 或 Northern 印迹的滤纸上用于互补核酸杂交的严格性杂交条件的实例是  $42^\circ\text{C}$  下含 1 mg 肝素的 50% 甲酰胺中过夜进行杂交, 所述互补核酸具有多于 100 个互补的残基。高度严格洗涤条件的实例是  $72^\circ\text{C}$  下 0.15M NaCl 约 15 分钟。严格洗涤条件的实例是  $65^\circ\text{C}$  下 0.2x SSC 洗涤 15 分钟 (SSC 缓冲液的描述见 Sambrook, 下文)。通常, 在高严格性洗涤之前进行低严格性洗涤, 去除背景探针信号。用于例如多于 100 个核苷酸双链体的中严格性洗涤的实例是  $45^\circ\text{C}$  下 1x SSC 15 分钟。例如多于 100 个核苷酸双链体的低严格性洗涤的实例是  $40^\circ\text{C}$  下 4-6x SSC 15 分钟。对短探针 (例如约 10 到 50 个核苷酸) 而言, 严格条件通常涉及在 pH 7.0 到 8.3 下少于约 1.0 M Na 离子的盐浓度, 通常约 0.01 到 1.0 M Na (或其他盐) 离子浓度, 而温度通常至少约  $30^\circ\text{C}$ 。也可通过去稳定剂如甲酰胺的添加达到严格条件。一般地, 是具体杂交实验中针对无关探针观察到的信噪比 2 倍 (或更高) 的信噪比表示检测到特异杂交。如果其所编码的蛋白质是基本相同的, 则在严格条件下彼此不杂交的核酸仍然是基本相同的。这发生于例如使用遗传密码子允许的最大密码子简并性产生核酸拷贝时。

以下是杂交/洗涤条件集合的实例, 其可用于克隆与本发明的参考核苷酸序列同源的核苷酸序列: 参考核苷酸序列与所述参考核苷酸序列在 7% 十二烷基硫酸钠(SDS)、0.5 M  $\text{NaPO}_4$ 、1 mM EDTA 中于  $50^\circ\text{C}$  下特异杂交, 在 2X SSC、0.1% SDS 中于  $50^\circ\text{C}$  洗涤; 更期望在 7% 十二烷基硫酸钠(SDS)、0.5 M  $\text{NaPO}_4$ 、1 mM EDTA 中于  $50^\circ\text{C}$  下特异杂交, 在 1X SSC、0.1% SDS 中于  $50^\circ\text{C}$  下洗涤; 进一步更期望在 7% 十二烷基硫酸钠(SDS)、0.5 M

NaPO<sub>4</sub>、1 mM EDTA 中于 50℃ 下特异杂交，在 0.5X SSC、0.1% SDS 中于 50℃ 下洗涤；特别地在 7% 十二烷基硫酸钠(SDS)、0.5 M NaPO<sub>4</sub>、1 mM EDTA 中于 50℃ 下特异杂交，在 0.1X SSC、0.1% SDS 中于 50℃ 下洗涤；更特别地在 7% 十二烷基硫酸钠(SDS)、0.5 M NaPO<sub>4</sub>、1 mM EDTA 中于 50℃ 下特异杂交，在 0.1X SSC、0.1% SDS 中于 65℃ 下洗涤。

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

基本相似性：在两条核酸或蛋白质序列的语境中，术语“基本相似性”是指基本相似的两条或更多序列或亚序列，例如具有 50%，特别是 60%，更特别是 70%，进一步更特别是 80%，仍然更特别是 90%，还更特别是 95%和最特别是 99%序列同一性。

底物：底物是酶天然识别并在酶天然发挥其功能的生物化学途径中转化为产物的分子，或是该分子经修饰的版本，该版本也被酶识别并在于天然发生的反应相似的酶反应中被酶转化为产物。

转化：用于将异源 DNA 引入植物细胞、植物组织或植物中的方法。转化的植物细胞、植物组织或植物理解为不仅包括转化过程的终产物，而且包括其转基因后代。

“转化的”、“转基因的”和“重组的”是指其中已引入异源核酸分子的宿主生物，如细菌或植物。核酸分子可被稳定整合进宿主的基因组中，或核酸分子也可作为染色体外分子存在。这类染色体外分子可以自我复制。转化的细胞、组织或植物理解为不仅包括转化过程的终产物，而且包括其转基因后代。“非转化的”、“非转基因的”或“非重组的”宿主是指不含有异源核酸分子的野生型生物，例如细菌或植物。

存活力：本文使用“存活力”是指植物的适应度(fitness)参数。针对植物发育的纯合表现对其进行测定，指出何种蛋白质对于植物生长是必要的。

## 发明详述

### 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定控制生物表型表达的基因，并使用多种方法学，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，生物信息学能够通过异源生物中鉴定在氨基酸或核苷酸水平上具有高相似性（同源性）程度的基因，为给定的基因指定功能。基因在 mRNA 或蛋白质水平上的表达能够通过将基因的表达与环境应答、发育过程或遗传（突变）或分子遗传（基因过量表达或低表达）干扰相关联来指定功能。基因在 mRNA 水平上的表达可单独（Northern 分析）或与其他基因一起（微阵列分析）检查，而基因在蛋白质水平上的表达能够单独（天然或变性的蛋白质凝胶或免疫印迹分析）或与其他基因一起（蛋白质组分析）检查。对蛋白质/蛋白质和蛋白质/DNA 相互作用的了解能够通过鉴定在相同生物学过程中一起发挥作用的蛋白质和核酸序列来指定功能。遗传学可通过证明基因中的 DNA 损伤（突变）对生物具有可计量的影响来对基因指定功能，所述影响包括但不限于：其发育；激素生物合成和应答；生长和生长习性（植物结构）；mRNA 表达概况；蛋白质表达概况；抗病能力；对非生物胁迫的耐受；获得营养物的能力；光合作用效率；改变的初级和次级代谢；和多种植物器官的组成。生物化学可通过证明由基因编码的蛋白质（特别是在异源生物中表达时）单独或与其他蛋白质一起具有某种酶活性来指定功能。分子遗传学能够通过天然植物或在异源生物中过量表达或低表达基因，并观察上文遗传学功能指定中所述的可定量影响来指定功能。在功能基因组学中，使用任何或所有这些方法（通常一起使用）；基于大量生物表型中的任意表型为基因指定功能。

本领域技术人员明白，所有这些不同的方法学均可提供证明具体基因功能的数据，并且这类证据随着递增数量的数据而更强大，所述数据用于功能指定：特别是来自一种方法学，更特别地来自两种方法学，并且进一步更特别地来自多于两种方法学。另外，本领域技术人员明白，不同的方法学在证明基因功能指定的证明力度中可不同。通常生物化学、遗传学和分子遗传学证据的资料被认为比生物信息学或基因表达证据的资料更有

力，但不总是如此。最后，本领域技术人员明白，对不同的基因而言，来自一种方法学的一种资料在证据的力度方面可以是不同的，所述证据由用于这些不同基因功能指定的各种不同资料提供。

作物性状功能基因组学(crop trait functional genomics)的目的是鉴定作物性状基因，即能够在作物植物中赋予有用的农业性状的基因。这类农业性状包括但不仅限于：增加的产量，无论是质量还是品质；增加的营养物获取和增加的代谢效率；用于食品、饲料或加工的植物组织增强或改变的营养物组成；增加的农业或工业加工实用性；增加的植物抗病性；增强的不良环境条件（非生物胁迫）耐受，所述不良环境条件包括但不仅限于干旱、过冷、过热、或过量的土壤盐度或极端酸度或碱度；和植物结构或发育中的改变，包括发育时间的改变。通过转基因或非转基因手段鉴定的这类性状基因的运用可以为了农业的利益显著地改良作物植物。

对人和动物消耗而言，谷物均是地球上最重要的作物植物。在稻、玉米、小麦、大麦、裸麦、燕麦和其他农业重要的单子叶植物中观察到遗传同线性（大染色体区段内基因顺序的保守），这有助于以单个谷物基因的序列为基础对来自不同谷物物种的直向同源基因进行绘图和分离。稻在谷物（cereal grain）中具有最小(~ 420 Mb)的基因组，并且近期是公众和私人的基因组测序与 EST 测序努力的主要焦点。

为了在稻[小麦]基因组中鉴定控制[性状]的作物性状基因，以一种或多种功能基因组方法为基础对来自稻基因组草图[小麦 EST 数据库]的基因进行优先顺序处理(prioritize)。例如，使用被稻瘟菌(*Magnaporthe grisea*)感染的稻植物的全基因组表达研究对控制疾病抗性的候选基因进行优先顺序处理。然后可以以稻全基因组序列的分析为基础，预测稻性状基因候选者的全长和部分 cDNA，并使用可商业获得的 PCR 引物挑选程序通过设计和使用 PCR 扩增引物将其分离。引物被用于从稻 cDNA 文库或第一链 cDNA 中 PCR 扩增全长或部分 cDNA。使用植物分子遗传学方法将得自任一方法的 cDNA 克隆用于构建载体，所述载体被设计为改变这些基因在转基因植物中的表达，所述分子遗传学方法在下文详述。通过在转基因植物

中过量表达或低表达关键性状基因来改变植物表型是对植物基因指定功能的一种有力和确定的方法。鉴定具有改变的目的性状的转基因植物的实验被用于明确地指定这些基因用于通过转基因方法或经典育种方法改良稻（并扩展至其他谷物）的实用性。

## II. cDNA 的鉴定、克隆和测序

本发明 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在一定范围的植物、单子叶植物和双子叶植物中使用，尤其是在单子叶植物如稻、小麦、大麦和玉米中使用。在一个更具体的实施方案中，单子叶植物是谷物。在一个更具体的实施方案中，谷物可以是例如玉米、小麦、大麦、燕麦、裸麦、粟、高粱、黑小麦、黑麦属(*secale*)、单粒小麦、斯佩耳特小麦、双粒小麦、画眉草、蜀黍、亚麻、格兰马草、磨擦草属物种或玉米草。在一个最具体的实施方案中，谷物是稻。其他植物的属包括，但不仅限于南瓜属 (*Cucurbita*)、蔷薇属 (*Rosa*)、葡萄属 (*Vitis*)、胡桃属 (*Juglans*)、*Gragaria*、百脉根属 (*Lotus*)、苜蓿属 (*Medicago*)、驴食草属 (*Onobrychis*)、胡卢巴属 (*Trigonella*)、豇豆属 (*Vigna*)、柑橘属 (*Citrus*)、亚麻属 (*Linum*)、老鹳草属 (*Geranium*)、木薯属 (*Manihot*)、胡萝卜属 (*Daucus*)、拟南芥属 (*Arabidopsis*)、芸苔属 (*Brassica*)、萝卜属 (*Raphanus*)、白芥属 (*Sinapis*)、颠茄属 (*Atropa*)、辣椒属 (*Capsicum*)、曼陀罗属 (*Datura*)、天仙子属 (*Hyoscyamus*)、番茄属 (*Lycopersicon*)、烟草属 (*Nicotiana*)、茄属 (*Solanum*)、碧冬茄属 (*Petunia*)、毛地黄属 (*Digitalis*)、花薄荷属 (*Majorana*)、菊苣属 (*Cichorium*)、向日葵属 (*Helianthus*)、莴苣属 (*Lactuca*)、雀麦属 (*Bromus*)、天门冬属 (*Asparagus*)、金鱼草属 (*Antirrhinum*)、*Heterocallis*、*Nemesis*、天竺葵属 (*Pelargonium*)、黍属 (*Panicum*)、狼尾草属 (*Pennisetum*)、毛茛属 (*Ranunculus*)、千里光属 (*Senecio*)、喇叭舌属 (*Salpiglossis*)、香瓜属 (*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 (*Hordeum*)、黑麦属 (*Secale*)、葱属 (*Allium*)

和小麦属(*Triticum*)。

本发明还提供了对包含本发明核酸分子的植物或植物部分进行基因分型的方法。任选地，该植物是单子叶植物，例如但不仅限于稻或小麦。基因分型提供了区分染色体对(*pari*)同源物的手段，并可用于在植物种群中区分分离体。分子标记物方法可用于系统发生研究，表征作物变种间的亲缘关系，鉴定杂交种或体细胞杂种(*somatic hybrid*)，定位影响单基因性状的染色体区段，图位克隆，和研究定量的遗传(参见 *Plant Molecular Biology: A Laboratory Manual*, 第7章, Clark 编著, Springer-Verlag, Berlin 1997; Paterson, A.H., "The DNA Revolution", *Genome Mapping in Plants* 中的第2章, Paterson, A.H. 编著, Academic 出版/R.G. Lands Co., Austin, Texas 1996)。

基因分型方法可使用任何数量的分子标记物分析技术，例如但不仅限于限制长度多态性(RFLP)。如本领域所公知的，RFLP 由 DNA 限制片段长度中的差异产生，所述差异得自相同基因的等位基因之间的核苷酸差异。因此，本发明提供了以下通过使用 RFLP 分析分离本发明的基因或核酸或遗传上连锁的染色体序列的方法。连锁的染色体序列在本发明核酸的 50 厘摩(50 cM)内，40 或 30 cM 内，特别地在 20 或 10 cM 内，更特别地在 5、3、2 或 1 cM 内。

### III. 目的性状

本发明包括编码下述蛋白质的多核苷酸的鉴定和分离，所述蛋白质涉及糖感知，以及最终的氮摄取和碳代谢。改变与这些性状相关的基因的表达能够用于根据期望改良或修饰植物和/或谷类(*grain*)。实施例描述了分离的目的基因，和分析表达改变及其对植物特性的影响的方法。

本发明的一个方面提供了用于调节或改变(即提高或降低)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被组成型地、时间或空间(例如在发育阶段、某些组织中和/或以一定数量)地表达，这对于非重组改造的植物是不典型的。因此，本

发明提供了在这类示范性应用中改变上文鉴定的特定特性的实用性。

## VI. 在转基因植物中控制基因表达

本发明还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，和通过改变本发明基因的表达来修饰目的表型性状的方法。

### A. 修饰编码序列和相邻序列

来自异源来源的基因在植物中的转基因表达可涉及对这些基因的修饰，以达到和最优化其在植物中的表达。具体地，在植物中在独立的转录物上表达下述细菌 ORF 是最佳的，所述细菌 ORF 编码独立的酶但是由天然微生物中同一转录物编码。为了达成该目的，将各微生物 ORF 各自分离并克隆在盒中，所述盒在 ORF 的 5' 末端提供植物启动子序列并在 ORF 的 3' 末端提供植物转录终止子。分离的 ORF 序列特别包含起始 ATG 密码子和终止 STOP 密码子，但是可包含除起始 ATG 和终止 STOP 密码子之外的额外序列。另外，ORF 可以是截短的，但仍然保留所需的活性；对尤其长的 ORF 而言，保留活性的截短的版本对于在转基因生物中表达可以是优选的。“植物启动子”和“植物转录终止子”旨在表示在植物细胞中工作的启动子和转录终止子。这包括可来自非植物来源如病毒（一个实例是花椰菜花叶病毒）的启动子和转录终止子。

在一些情况下，对 ORF 编码序列和相邻序列的修饰不是必需的。分离含目的 ORF 的片段并将其插入植物启动子下游就是足够的。例如，Gaffney 等 (Science 261: 754-756 (1993)) 在转基因植物中成功地表达了位于 CaMV 35S 启动子和 CaMV tml 终止子控制下的假单胞杆菌 (*Pseudomonas*) nahG 基因，而未修饰编码序列，且假单胞菌基因 ATG 上游的核苷酸和 STOP 密码子下游的核苷酸仍然与 nahG ORF 相连。特别地，应该留下尽可能少的相邻微生物序列连接在 ATG 上游和 STOP 密码子下游。事实上，这类构建可取决于限制性位点的可用性。

在其他情况下，来自微生物来源的基因的表达可在表达中产生问题。

这些问题已在本领域中充分表征，并且对来自某些来源如芽孢杆菌 (*Bacillus*) 的基因而言尤其常见。这些问题可适用于本发明的核苷酸序列，且这些基因的修饰可使用本领域目前公知的技术进行。可遇到以下的问题：

### 1. 密码子选择。

植物中特定的密码子选择与某些微生物中特定的密码子选择不同。将克隆的微生物 ORF 中的密码子选择与植物基因（尤其是来自靶植物的基因）中的选择进行比较，会使得能够识别 ORF 中应当被特别改变的密码子。通常，植物进化趋向于在单子叶植物的第三个碱基位置中对核苷酸 C 和 G 的有力偏好，而双子叶植物常在该位置使用核苷酸 A 或 T。通过修饰基因以掺入具体靶转基因物种的特定密码子选择，会解决下文所述关于 GC/AT 含量和不合理剪接的许多问题。

### 2. GC/AT 含量。

植物基因通常具有多于 35% 的 GC 含量。富含 A 和 T 核苷酸的 ORF 序列能够在植物中引起若干问题。首先，ATTTA 基序被认为引起信号的去稳定化并存在于在许多短寿命 mRNA 的 3' 末端。其次，多腺苷酸化信号如 AATAAA 在信息中不适当位置的发生被认为引起转录的过早截断。另外，单子叶植物可将富含 AT 的序列识别为剪接位点（见下文）。

### 3. 与起始的甲硫氨酸相邻的序列。

植物与微生物的区别在于其信息不具有确定的核糖体结合位点。更确切地说，认为核糖体与信息 5' 末端结合并扫描第一个可获得的 ATG，在这里开始翻译。然而，认为存在对与 ATG 相邻的某些核苷酸的偏好，并且可通过在 ATG 处包含真核生物共有的翻译起始子来增加微生物基因的该表达。Clontech (1993/1994 目录, 第 210 页, 引入本文作为参考) 提出一条序列作为大肠杆菌 *uidA* 基因在植物中表达的共有翻译起始子。另外, Joshi (N.A.R. 15: 6643-6653 (1987), 引入本文作为参考) 比较了与 ATG 相邻的许多植物序列, 并提出另一共有序列。在植物中表达微生物 ORF 时遇到困难的情况下, 在起始 ATG 处包含这些序列中的一条可促进翻译。在这些情况下, 共有序列的最后三个核苷酸可由于其对第二个 AA 残基的

修饰而不适合包含在修饰的序列中。与起始甲硫氨酸相邻的特定序列可在不同的植物物种间不同。对位于 GenBank 数据库中 14 个玉米基因的调查提供了以下的结果：

14 个玉米基因中起始 ATG 前的位置：

	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
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

可对核苷酸要掺入其中的期望的植物物种和为了掺入特定核苷酸而被修饰的与 ATG 相邻的序列进行该分析。

#### 4. 去除不合理的剪接位点。

从非植物来源克隆并且未针对在植物中表达进行最优化的基因也可含有在植物中被识别为 5' 或 3' 剪接位点并被切割的基序，因此产生截短的或删除的信息。可使用本领域公知的技术去除这些位点。

修饰编码序列和相邻序列的技术是本领域公知的。在微生物 ORF 的原始表达低并认为如上文所述改变序列是合适的的情况下，可根据本领域公知的方法完成合成基因的构建。这些方法例如描述于公开的专利公开文本 EP 0 385 962 (属于 Monsanto)、EP 0 359 472 (属于 Lubrizol) 和 WO 93/07278 (属于 Ciba-Geigy) 中，其均引入本文作为参考。在大部分情况下，优选使用瞬时测定方案 (这是本领域公知的) 在基因构建体转移至转基因植物之前测定其表达。

#### B. 构建植物表达盒

旨在转基因植物中表达的编码序列首先被装配在表达盒中的可在植物中表达的合适启动子之后。该表达盒也可包含转基因表达所需或选择的任何其他序列。这类序列包括但不限于转录终止子、增加表达的外来序列如内含子、关键序列(vital sequence)，和旨在将基因产物靶向特定细胞器

和细胞区室的序列。然后可将这些表达盒容易地转移至下文所述的植物转化载体。以下描述了典型的表达盒的多种组件。

### 1. 启动子

用于表达盒中的启动子的选择会确定转基因在转基因植物中的空间和时间表达模式。选择的启动子会在特异的细胞类型（如叶表皮细胞、叶肉细胞、根皮层细胞）或在特异的组织或器官（例如根、叶或花）中表达转基因，且该选择会反映期望的基因产物累积位置。或者，选择的启动子可驱动基因在不同的诱导条件下表达。启动子的强度（即启动转录的能力）不同。根据使用的宿主细胞体系，可使用大量合适启动子之任一，包括基因固有的启动子。以下是可用于表达盒中的启动子的非限制性实例。

#### a. 组成型表达，遍在蛋白启动子：

遍在蛋白是已知在许多细胞类型中累积的基因产物，且其启动子从若干物种中被克隆用于在转基因植物中使用（例如向日葵-Binet 等，*Plant Science* 79: 87-94 (1991); 玉米- Christensen 等，*Plant Molec. Biol.* 12: 619-632 (1989); 和拟南芥- Callis 等，*J. Biol. Chem.* 265:12486-12493 (1990) 和 Norris 等，*Plant Mol. Biol.* 21:895-906 (1993)）。已在转基因单子叶植物体系中开发了玉米遍在蛋白启动子，并且其构建用于单子叶植物转化的序列和载体在专利公开 EP 0 342 926（属于 Lubrizol）中公开，该文献引入本文作为参考。Taylor 等 (*Plant Cell Rep.* 12: 491-495 (1993))描述了包含玉米遍在蛋白启动子和第一内含子的载体(pAHC25)，以及通过微粒轰击被引入后其在大量单子叶生物细胞悬浮液中的高活性。对于用于本发明核苷酸序列，拟南芥遍在蛋白启动子是理想的。遍在蛋白启动子适用于在转基因植物（单子叶植物和双子叶植物二者）中的基因表达。合适的载体是 pAHC25 的衍生物或该申请中描述的任何转化载体，其通过引入适当的遍在蛋白启动子和/或内含子序列被修饰。

#### b. 组成型表达，CaMV 35S 启动子：

质粒 pCGN1761 的构建描述于公开的专利申请 EP 0 392 225（实施例 23）中，该文献引入本文作为参考。pCGN1761 含有“双” CaMV 35S 启

动子和 tml 转录终止子，在启动子和终止子之间具有特有的 EcoRI 位点，并具有 pUC-型主链。构建了具有修饰的多接头的 pCGN1761 的衍生物，其除了已有的 EcoRI 位点外还包含 NotI 和 XhoI 位点。该衍生物被命名为 pCGN1761ENX。pCGN1761ENX 适用于为了在转基因植物中在 35S 启动子的控制下表达的目的，在其多接头内克隆 cDNA 序列或编码序列（包括微生物 ORF 序列）。这类构建体的整个 35S 启动子 - 编码序列 - tml 终止子盒可通过启动子 5' 的 HindIII、SphI、Sall 和 XbaI 位点和终止子 3' 的 XbaI、BamHI 和 BglII 位点切割用于转移至转化载体，如下文所述的转化载体。另外，可通过 HindIII、SphI、Sall、XbaI 或 PstI 的 5' 切割或任何多接头限制性位点（EcoRI、NotI 或 XhoI）的 3' 切割去除双 35S 启动子片段，用于置换另一启动子。如果期望的话，可通过引入可增加翻译的序列在克隆位点附近进行修饰。这尤其适用于期望过量表达时。例如，可通过如美国专利 No. 5,639,949 的实施例 37 所述最优化翻译起点来修饰 pCGN1761ENX，所述文献引入本文作为参考。

### c. 组成型表达，肌动蛋白启动子：

已知若干种肌动蛋白同种型在大部分细胞类型中表达，因此肌动蛋白启动子是组成型启动子的良好选择。具体地，已克隆和表征了来自稻 Act1 基因的启动子 (McElroy 等, Plant Cell 2: 163-171 (1990))。发现启动子的 1.3kb 片段含有在稻原生质体中表达所需的所有调节元件。另外，已构建了基于 ActI 启动子的大量表达载体，它们尤其用于单子叶植物 (McElroy 等, Mol. Gen. Genet. 231: 150-160 (1991))。这些载体整合了 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 与 ActI 内含子或 ActI 5' 侧翼序列与 ActI 内含子的融合物。(GUS 报告子基因) 起始 ATG 附近序列的最优化也增加了表达。McElroy 等 (Mol. Gen. Genet. 231: 150-160 (1991)) 所述的启动子表达盒可被容易地修饰用于基因表达，并尤其适合在单子叶植物宿主中使用。例如，从 McElroy 构建体中取出含启动子的片段并用于置换 pCGN1761ENX 中的双 35S 启动子，然后所述 pCGN1761ENX

可用于插入特异的基因序列。然后可将由此构建的融合基因转移至适当的转化载体。在独立的报道中，也发现稻 ActI 启动子与其第一个内含子指导在培养的大麦细胞中的高表达(Chibbar 等, *Plant Cell Rep.* 12: 506-509 (1993))。

d. 诱导性表达, PR-1 启动子:

pCGN1761ENX 中的双 35S 启动子可用选择的任何另一启动子代替, 所述另一启动子会导致适当的高表达水平。举例而言, 美国专利 No. 5,614,395 中所述的可化学调节的启动子之一(如烟草 PR-1a 启动子)可代替双 35S 启动子。或者, 可使用 Lebel 等, *Plant J.* 16:223-233 (1998) 中所述的拟南芥 PR-1 启动子。通过限制性酶将选择的启动子从其来源特异切割, 但是也可以使用带有适当末端限制性位点的引物进行 PCR 扩增从而从其来源特异切割。要进行 PCR-扩增时, 应当在靶载体中克隆扩增的启动子之后对启动子再测序以检验扩增错误。将可化学/病原体调节的烟草 PR-1a 启动子从质粒 pCIB1004 (用于构建体, 见 EP 0 332 104 的实施例 21, 该文献引入本文作为参考) 中切割并转移至质粒 pCGN1761ENX (Uknes 等, *Plant Cell* 4: 645-656 (1992))。用 NcoI 切割 pCIB1004, 并通过用 T4 DNA 聚合酶处理使得到的线性片段的 3' 突出端成为平端。然后用 HindIII 切割该片段, 对得到的含 PR-1a 启动子片段进行凝胶纯化并克隆进 pCGN1761ENX 中, 所述 pCGN1761ENX 已去除了双 35S 启动子。这如下完成: 用 XhoI 切割并用 T4 聚合酶变平端, 然后用 HindIII 切割, 并分离含较大载体-终止子的片段, pCIB1004 启动子片段被克隆进上述片段中。这产生了 pCGN1761ENX 衍生物, 其具有 PR-1a 启动子和 tml 终止子和具有特有的 EcoRI 和 NotI 位点的插入多接头。可向该载体中插入选择的编码序列, 随后将融合产物(即启动子-基因-终止子)转移至任何选择的转化载体, 包括下文所述的转化载体。可使用多种化学调节剂在根据本发明转化的植物中诱导选择的编码序列表达, 所述化学调节剂包括美国专利第 5,523,311 号和第 5,614,395 号中公开的苯并噻二唑、异烟酸和水杨酸化合物。

e.诱导性表达, 乙醇诱导性启动子:

也可使用可由某醇或酮(如乙醇)诱导的启动子赋予本发明编码序列的诱导性表达。这样的启动子例如为来自无冠构巢曲霉(*Aspergillus nidulans*)的 *alcA* 基因启动子(Caddick 等, (1998) *Nat. Biotechnol* 16:177-180)。在构巢曲霉中, *alcA* 基因编码醇脱氢酶 I, 其表达在存在化学诱导剂时由 AlcR 转录因子调节。就本发明的目的而言, 将质粒 *palcA:CAT* 中的 CAT 编码序列替换为本发明的编码序列, 以形成编码序列处于 *alcA* 基因启动子控制下的表达盒, 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 等, (1998) *Nat. Biotechnol* 16:177-180)。这使用本领域公知的方法完成。

f.诱导性表达, 糖皮质激素诱导性启动子:

还涉及使用基于类固醇激素的系统诱导本发明核酸序列的表达。例如, 使用糖皮质激素介导的诱导体系(Aoyama 和 Chua (1997) *The Plant Journal* 11: 605-612)并通过应用糖皮质激素诱导基因表达, 所述糖皮质激素例如是合成的糖皮质激素, 特别是地塞米松, 特别是以 0.1mM 到 1mM, 更特别是从 10mM 到 100mM 范围内的浓度。就本发明的目的而言, 用本发明的核酸序列代替萤光素酶基因序列, 形成本发明的核酸序列处于与 35S 最小启动子融合的六拷贝 GAL4 上游激活序列的控制下的表达盒。这使用本领域公知的方法完成。反式作用因子包括与疱疹病毒蛋白 VP16 (Triezenberg 等, (1988) *Genes Devel.* 2: 718-729)融合的 GAL4 DNA-结合结构域(Keegan 等, (1986) *Science* 231: 699-704), 所述疱疹病毒蛋白 VP16 与大鼠糖皮质激素受体的激素结合结构域融合(Picard 等, (1988) *Cell* 54: 1073-1080)。融合蛋白的表达由本领域已知的或本文所述的启动子控制。该表达盒还包含在下述植物中, 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, 融合蛋白的组织特异性或器官特异性的实现, 导致杀虫毒素的可诱导的组织或器官特异性。

g.根特异表达:

另一种基因表达模式是根表达。合适的根启动子是由 de Framond

(FEBS 290: 103-106 (1991))和美国专利第 5,466,785 号所述的玉米金属硫蛋白样(MTL)基因启动子, 所述文献引入本文作为参考。将该“MTL”启动子转移至用于插入选定基因的合适载体如 pCGN1761ENX, 并随后将整个启动子-基因-终止子盒转移至目的转化载体。

#### h. 创伤诱导性启动子:

创伤诱导性启动子也适用于基因表达。已描述了大量这类启动子(例如 Xu 等, *Plant Molec. Biol.* 22: 573-588 (1993), Logemann 等, *Plant Cell* 1: 151-158 (1989), Rohrmeier & Lehle, *Plant Molec. Biol.* 22: 783-792 (1993), Firek 等, *Plant Molec. Biol.* 22: 129-142 (1993), Warner 等, *Plant J.* 3: 191-201 (1993))且其均适用于本发明。Logemann 等描述了双子叶马铃薯 *wunI* 基因的 5'上游序列。Xu 等显示来自双子叶植物马铃薯的创伤诱导性启动子(*pin2*)在单子叶植物稻中有活性。另外, Rohrmeier & Lehle 描述了使用标准技术克隆玉米 *WipI* cDNA, 该 cDNA 是创伤诱导的并可用于分离同族(cognate)启动子。类似地, Firek 等, 和 Warner 等描述了来自单子叶植物石刁柏(*Asparagus officinalis*)的创伤诱导基因, 该基因在局部创伤和病原体侵入位点表达。使用本领域公知的克隆技术, 可将这些启动子转移至合适的载体, 与关于本发明的基因融合, 并用于在植物创伤位点表达这些基因。

#### i. 髓特异表达:

专利申请 WO 93/07278 描述了优先在髓细胞中表达的玉米 *trpA* 基因的分离, 该文献引入本文作为参考。提出了扩展至转录起点-1726bp 的基因序列和启动子。使用标准分子生物学技术, 可将该启动子或其部分转移至载体如 pCGN1761, 其中该启动子能够替换 35S 启动子, 并用于驱动外源基因以髓特异的方式表达。事实上, 含有髓特异启动子或其部分的片段可被转移至任何载体, 并针对在转基因植物中的实用性进行修饰。

#### j. 叶特异表达:

Hudspeth & Grula (*Plant Molec Biol* 12: 579-589 (1989))已描述了编码磷酸烯醇羧化酶(PEPC)的玉米基因。使用标准分子生物学技术, 可使用

该基因的启动子驱动任何基因以叶特异的方式在转基因植物中表达。

#### k.花粉特异表达:

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。基因序列和启动子扩展至从转录起点 1400 bp。使用标准分子生物学技术,可将该启动子或其部分转移至载体如 pCGN1761,其中该启动子能够替换 35S 启动子,并用于驱动本发明的核酸序列以花粉特异的方式表达。

#### 2.转录终止子

可获得大量在表达盒中使用的转录终止子。它们负责终止超出转基因的转录并校正 mRNA 多聚腺苷酸化。适当的转录终止子是已知在植物中发挥作用的那些,并包括 CaMV 35S 终止子、tml 终止子、胭脂碱合酶终止子和豌豆 rbcS E9 终止子。其可用于单子叶植物和双子叶植物。另外,可使用基因的固有转录终止子。

#### 3.用于增加或调节表达的序列

发现来自转录单位内的大量序列增加基因表达,并且这些序列可与本发明的基因组合使用,提高本发明的基因在转基因植物中的表达。

多种内含子序列已显示增加表达,尤其是在单子叶植物细胞中。例如,发现当引入玉米细胞中时,玉米 AdhI 基因的内含子显著增加位于其同族启动子下的野生型基因的表达。发现内含子 1 是尤其有效的,并在与氯霉素乙酰转移酶基因的融合构建体中增加表达(Callis 等, *Genes Develop.* 1: 1183-1200 (1987))。在相同的实验体系中,来自玉米 bronze1 基因的内含子具有类似的增加表达的效果。内含子序列已被常规地整合进植物转化载体中,通常在非翻译的前导区内。

也已知来自病毒的大量非翻译前导序列增加表达,并且它们在双子叶植物细胞中尤其有效。特别地,来自烟草花叶病毒(TMV,“W-序列”)、玉米萎黄病斑点病毒(MCMV)和苜蓿花叶病毒(AMV)的前导序列已显示有效增加表达(例如 Gallie 等, *Nucl. Acids Res.* 15: 8693-8711 (1987); Skuzeski 等, *Plant Molec. Biol.* 15: 65-79 (1990))。本领域已知的其他前

导序列包括但不限于：小核糖核酸病毒前导区，例如 EMCV 前导区（脑心肌炎 5'非编码区）(Elroy-Stein, O.、Fuerst, T. R.和 Moss, B. PNAS USA 86:6126-6130 (1989)); 马铃薯 Y 病毒(potyvirus)前导区，例如 TEV 前导区（烟草蚀斑病毒）(Allison 等, 1986); MDMV 前导区（玉米矮花叶病毒）; Virology 154:9-20); 人免疫球蛋白重链结合蛋白(BiP)前导区(Macejak, D. G. 和 Sarnow, P., Nature 353: 90-94 (1991)); 来自苜蓿花叶病毒外壳蛋白 mRNA(AMV RNA 4)的非翻译前导区(Jobling, S. A.和 Gehrke, L., Nature 325:622-625 (1987)); 烟草花叶病毒前导区(TMV), (Gallie, D. R. 等, Molecular Biology of RNA, 第 237-256 页 (1989)); 和玉米萎黄病斑点病毒(MCMV)前导区(Lommel, S. A.等, Virology 81:382-385 (1991))。还参见 Della-Cioppa 等, Plant Physiology 84:965-968 (1987)。

除了将一个或多个上述元件整合进本发明的靶表达盒的 5'调节区内以外，也可整合靶表达盒特有的其他元件。这类元件包括但不限于最小启动子。最小启动子旨在表示无上游激活时是无活性的或几乎无活性的基本启动子元件。当不存在反式作用子或当不存在增强子或应答元件结合位点时，这样的启动子在植物中具有低背景活性。尤其适用于植物中靶基因的一种最小启动子是 Bz1 最小启动子，其得自玉米的 *bronze1* 基因。该 Bz1 核心启动子通过在位于-53 和-58 的 *NheI* 位点切割，得自"myc"突变体 Bz1-萤光素酶构建体 pBz1LucR98。Roth 等, Plant Cell 3: 317 (1991)。衍生的 Bz1 核心启动子片段因此从-53 扩展至+227，并在 5'非翻译区包含 Bz1 内含子-1。还适用于本发明的是通过使用合成的 TATA 元件创建的最小启动子。该 TATA 元件允许 RNA 聚合酶因子识别启动子，并在不缺失激活时赋予基底水平的基因表达（一般参见 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

已知在植物中存在靶向基因产物的多种机制，并且已较详细地表征了控制这些机制功能的序列。例如，基因产物到叶绿体的靶向由存在于多种蛋白质氨基端的信号序列控制，该信号序列在叶绿体输入时被切割，产生

成熟的蛋白质(例如 Comai 等, *J. Biol. Chem.* 263: 15104-15109 (1988))。这些信号序列可与异源基因产物融合,影响异源产物进入叶绿体的输入(van den Broeck 等, *Nature* 313: 358-363 (1985))。编码适当信号序列的 DNA 可从下述 cDNA 的 5'端分离,所述 cDNA 编码 RUBISCO 蛋白质、CAB 蛋白质、EPSP 合酶、GS2 蛋白质和已知定位于叶绿体的许多其他蛋白质。还参见美国专利第 5,639,949 号的实施例 37 中题为“Expression With Chloroplast Targeting”的部分。

其他基因产物定位于其他细胞器如线粒体和过氧化物酶体中(例如 Unger 等, *Plant Molec. Biol.* 13: 411-418 (1989))。也可利用编码这些产物的 cDNA 影响异源基因产物到这些细胞器的靶向。这类序列的实例是核编码的 ATP 酶和线粒体特异的天冬氨酸氨基转移酶同种型。靶向细胞蛋白质体已由 Rogers 等(*Proc. Natl. Acad. Sci. USA* 82: 6512-6516 (1985))描述。

另外,已表征了引起基因产物靶向到其他细胞区室的序列。氨基端序列负责到 ER、质外体的靶向和从糊粉细胞的胞外分泌(Koehler & Ho, *Plant Cell* 2: 769-783 (1990))。另外,氨基端序列与羧基端序列组合负责基因产物的液泡靶向(Shinshi 等, *Plant Molec. Biol.* 14: 357-368 (1990))。

通过将上述适当的靶向序列与目的转基因序列融合,可能指导转基因产物去往任何细胞器或细胞区室。对叶绿体靶向而言,例如将来自 RUBISCO 基因、CAB 基因、EPSP 合酶基因或 GS2 基因的叶绿体信号序列与转基因的氨基端 ATG 融合。选择的信号序列应包括已知的切割位点,构建的融合物应考虑切割所需的切割位点后的任何氨基酸。在一些情况下,可通过在切割位点和转基因 ATG 之间添加少量氨基酸,或者通过替换转基因序列中的一些氨基酸来满足该需要。可如下测试针对叶绿体输入构建的融合物的叶绿体吸收效率:将体外转录的构建体体外翻译,然后使用 Bartlett 等,在 Edelman 等(编著) *Methods in Chloroplast Molecular Biology*, Elsevier, 第 1081-1091 页 (1982)和 Wasmann 等, *Mol. Gen. Genet.* 205: 446-453 (1986)中所述的技术进行体外叶绿体吸收。这些构建技术是本领域公知的,并同样适用于线粒体和过氧化物酶体。

细胞靶向的上述机制不仅可与其同族启动子组合使用，而且也可以与异源启动子组合使用，从而完成在启动子的转录调节下特异的细胞靶向目标，所述启动子具有与靶向信号来源的启动子不同的表达模式。

### C.构建植物转化载体

可用于植物转化的大量转化载体是植物转化领域的普通技术人员已知的，并且与本发明有关的这些基因可与任何这类载体组合使用。载体的选择应取决于特异的转化技术和转化的靶物种。对某些靶物种而言，不同的抗生素或除草剂选择标记物可以是特异的。转化中常规使用的选择标记物包括赋予卡那霉素和相关抗生素抗性的 *nptII* 基因 (Messing & Vierra, *Gene* 19: 259-268 (1982); Bevan 等, *Nature* 304:184-187 (1983))、赋予除草剂膦丝菌素抗性的 *bar* 基因(White 等, *Nucl. Acids Res* 18: 1062 (1990), Spencer 等, *Theor. Appl. Genet* 79: 625-631 (1990))、赋予抗生素潮霉素抗性的 *hph* 基因(Blochinger & Diggelmann, *Mol Cell Biol* 4: 2929-2931)、和赋予对甲氨蝶呤 (methotrexate) 抗性的 *dhfr* 基因(Bourouis 等, *EMBO J.* 2(7): 1099-1104 (1983))、赋予草甘膦抗性的 EPSPS 基因 (美国专利第 4,940,935 号和第 5,188,642 号) 和提供代谢甘露糖的能力的甘露糖-6-磷酸异构酶基因 (美国专利第 5,767,378 号和第 5,994,629 号)。

#### 1.适用于农杆菌(*Agrobacterium*)转化的载体

许多载体可用于使用根瘤农杆菌(*Agrobacterium tumefaciens*)的转化。这些载体通常带有至少一个 T-DNA 边界序列并包括载体如 pBIN19 (Bevan, *Nucl. Acids Res.* (1984))。下文描述了适用于农杆菌转化的两种典型载体的构建。

##### a. pCIB200 和 pCIB2001:

二元载体 pCIB200 和 pCIB2001 用于构建与农杆菌一起使用的重组载体，并以如下方式构建。如下创建 pTJS75kan: *NarI* 消化 pTJS75 (Schmidhauser & Helinski, *J. Bacteriol.* 164: 446-455 (1985))，该消化允许切除四环素抗性基因，然后插入来自 pUC4K 的带有 NPTII 的 *AccI* 片段

(Messing & Vierra, *Gene* 19: 259-268 (1982); Bevan 等, *Nature* 304: 184-187 (1983); McBride 等, *Plant Molecular Biology* 14: 266-276 (1990))。将 XhoI 接头与 PCIB7 的 EcoRV 片段连接, 所述 EcoRV 片段含有左侧和右侧 T-DNA 边界、植物可选择的 nos/nptII 嵌合基因和 pUC 多接头 (Rothstein 等, *Gene* 53: 153-161 (1987)), 并将 XhoI-消化的片段克隆进 Sall-消化的 pTJS75kan 中, 创建 pCIB200 (也见 EP 0 332 104, 实施例 19)。pCIB200 含有以下的特有多接头限制性位点: EcoRI、SstI、KpnI、BglII、XbaI 和 Sall。pCIB2001 是通过向多接头中插入额外的限制性位点创建的 pCIB200 的衍生物。pCIB2001 多接头中特有的限制性位点是 EcoRI、SstI、KpnI、BglII、XbaI、Sall、MluI、BclI、AvrII、ApaI、HpaI 和 StuI。除了含有这些特有的限制性位点以外, pCIB2001 还具有植物和细菌卡那霉素选择、用于农杆菌介导的转化的左侧和右侧 T-DNA 边界、用于在大肠杆菌和其他宿主间流通的来自 RK2 的 trfA 功能, 且 OriT 和 OriV 功能也来自 RK2。pCIB2001 多接头适用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 含有用于在植物中选择的编码卡那霉素抗性的基因和 T-DNA 右侧和左侧边界序列, 并掺入来自广泛宿主范围质粒 pRK252 的序列, 这允许其在大肠杆菌和农杆菌二者中复制。其构建由 Rothstein 等(*Gene* 53: 153-161 (1987))描述。构建了 pCIB10 的多种衍生物, 其掺入由 Gritz 等(*Gene* 25: 179-188 (1983))所述的潮霉素 B 磷酸转移酶基因。这些衍生物使得能够在只有潮霉素(pCIB743)时, 或在潮霉素和卡那霉素(pCIB715, pCIB717)上选择转基因植物。

2.适用于非农杆菌转化的载体

不使用根瘤农杆菌的转化绕过了选择的转化载体中对 T-DNA 序列的需要, 因此除了如上所述含有 T-DNA 序列的载体外可使用缺乏这些序列的载体。不依赖于农杆菌的转化技术包括通过粒子轰击、原生质体吸收(例如 PEG 和电穿孔)和显微注射的转化。载体的选择主要取决于对被转化的物种的特异选择。下文描述了适用于非农杆菌转化的典型载体的构建。

**a. pCIB3064:**

pCIB3064 是适用于直接基因转移技术的来自 pUC 的载体, 所述直接基因转移技术与除草剂 basta (或膦丝菌素) 选择组合。质粒 pCIB246 包含与大肠杆菌 GUS 基因和 CaMV 35S 转录终止子有效融合的 CaMV 35S 启动子, 并描述于 PCT 已公开的申请 WO 93/07278 中。该载体的 35S 启动子在起点 5' 含有两条 ATG 序列。以去除 ATG 并产生限制性位点 SspI 和 PvuII 的方式使用标准 PCR 技术突变这些位点。新的限制性位点距离特有的 SalI 位点 96 和 37 bp, 并距离真实的起点 101 和 42 bp。得到的 pCIB246 衍生物被称作 pCIB3025。然后通过用 SalI 和 SacI 消化从 pCIB3025 中切除 GUS 基因, 使末端变平端并重新连接, 产生质粒 pCIB3060。质粒 pJIT82 得自 John Innes Centre, Norwich, 切下含有来自绿色链霉菌(*Streptomyces viridochromogenes*)的 bar 基因的 400 bp SmaI 片段并插入 pCIB3060 (Thompson 等, EMBO J 6: 2519-2523 (1987))的 HpaI 位点中。这产生了 pCIB3064, 其包含处于 CaMV 35S 启动子和终止子控制下的用于除草剂选择的 bar 基因、氨苄青霉素抗性基因 (用于在大肠杆菌中选择) 和具有特有位点 SphI、PstI、HindIII 和 BamHI 的多接头。该载体适用于克隆含有其自身调节信号的植物表达盒。

**b. pSOG19 和 pSOG35:**

pSOG35 是一种转化载体, 其利用大肠杆菌基因二氢叶酸还原酶(DFR) 作为赋予甲氧蝶呤抗性的可选择标记物。使用 PCR 扩增 35S 启动子、来自玉米 Adh1 基因的内含子 6(-550 bp), 和来自 pSOG10 的 18 bp GUS 非翻译前导序列。还通过 PCR 扩增编码大肠杆菌 II 型二氢叶酸还原酶基因的 250-bp 片段, 并将这两条 PCR 片段用来自 pB1221 (Clontech) 的 SacI-PstI 片段装配, 所述 SacI-PstI 片段包含 pUC19 载体主链和胭脂碱合酶终止子。这些片段的装配产生 pSOG19, 其含有与内含子 6 序列、GUS 前导区、DHFR 基因和胭脂碱合酶融合的 35S 启动子。用来自玉米萎黄病斑点病毒(MCMV)的前导序列替换 pSOG19 中的 GUS 前导区, 产生载体 pSOG35。pSOG19 和 pSOG35 带有氨苄青霉素抗性的 pUC 基因, 并具有

可用于克隆外源物质的 HindIII、SphI、PstI 和 EcoRI 位点。

### 3.适用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)。将核苷酸序列插入 pPH143 中，从而替换 PROTOX 编码序列。然后将该载体用于质体转化，并选择具有壮观霉素抗性的转化体。或者，将核苷酸序列插入 pPH143 中，使其替换 aadH 基因。在该情况下，针对 PROTOX 抑制剂抗性选择转化体。

### D.转化

一旦本发明的核酸被克隆进表达体系内，将其转化进植物细胞中。可以通过大量本领域认可的方式将本发明的受体和靶表达盒引入植物细胞中。再生植物的方法也是本领域公知的。例如，Ti 质粒载体已用于递送外源 DNA，以及直接 DNA 吸收、脂质体、电穿孔法、显微注射和微粒。另外，来自农杆菌属的细菌可用于转化植物细胞。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是本领域公知的，并包括基于农杆菌的技术和不需要农杆菌的技术。非农杆菌技术涉及原生质体或细胞对外源遗传材料的直接吸收。这可通过 PEG 或电穿孔介导的吸收、粒子轰击介导的递送或显微注射完成。这些技术的实例由 Paszkowski 等, EMBO J 3: 2717-2722 (1984)、Potrykus 等, Mol. Gen. Genet. 199: 169-177 (1985)、Reich 等, Biotechnology 4: 1001-1004 (1986)和 Klein 等, Nature 327: 70-73 (1987)描述。在各情况下，使用本领域已知的标准技术将转化的细胞再生为整株植物。

农杆菌介导的转化是转化双子叶植物的特异技术，因为其转化的高效率和针对许多不同物种的广泛实用性。农杆菌转化通常涉及带有目的外来 DNA 的二元载体(例如 pCIB200 或 pCIB2001)到适当农杆菌菌株的转移，该转移可取决于宿主农杆菌菌株(例如对 pCIB200 和 pCIB2001 而言是菌株 CIB542(Uknes 等, Plant Cell 5: 159-169 (1993))在共同存在的 Ti 质粒

或染色体上带有的 *vir* 基因的互补序列。通过三亲本交配步骤，使用带有重组二元载体的大肠杆菌、辅助大肠杆菌菌株完成重组二元载体到农杆菌的转移，所述助手大肠杆菌菌株包含质粒如 pRK2013 并能够将重组二元载体移动至靶农杆菌菌株。或者，可通过 DNA 转化将重组二元载体转移至农杆菌(Höfgen & Willmitzer, Nucl. Acids Res. 16: 9877 (1988))。

通过重组农杆菌转化靶植物物种常涉及农杆菌与来自植物的外植体的共同培养，并且遵循本领域公知的方案。在可选择培养基上再生带有抗生素或除草剂抗性标记物的转化的组织，所述标记物存在于二元质粒 T-DNA 边界之间。

用基因转化植物细胞的另一途径涉及在植物组织和细胞中推进惰性或有生物学活性的颗粒。该技术在均属于 Sanford 等的美国专利第 4,945,050 号、第 5,036,006 号和第 5,100,792 号中公开。一般地，该步骤涉及在有效穿透细胞外表面并在其内部提供整合的条件下在细胞中推进惰性或有生物学活性的颗粒。使用惰性颗粒时，可通过用含有期望基因的载体包裹颗粒而将载体引入细胞中。或者，可用载体包围靶细胞，使得载体通过颗粒的活跃(wake)被带入细胞中。也可向植物细胞组织中推进生物学活性颗粒（例如干燥的酵母细胞、干燥的细菌或噬菌体，分别含有想要引入的 DNA）。

## 2. 单子叶植物的转化

现在大部分单子叶植物物种的转化也已成为常规。特异的技术包括使用 PEG 或电穿孔技术将基因直接转移进入原生质体，和粒子轰击进入愈伤组织中。转化可用单个 DNA 物种或多种 DNA 物种（即共转化）进行，这两种技术均适用于本发明。共转化可具有下述优点：避免完整载体构建和产生下述转基因植物，所述转基因植物中目的基因和可选择标记物是未连锁的基因座，使得能够在随后的世代中去除可选择标记物，这被认为是期望的。然而，使用共转化的缺点是独立的 DNA 物种被整合进基因组中的频率低于 100%(Schocher 等, Biotechnology 4: 1093-1096 (1986))。

专利申请 EP 0 292 435、EP 0 392 225 和 WO 93/07278 描述了从玉米的良种自交系 (elite inbred line) 制备愈伤组织和原生质体、使用 PEG 或

电穿孔转化原生质体，和从转化的原生质体再生玉米植物的技术。Gordon-Kamm 等，(Plant Cell 2: 603-618 (1990))和 Fromm 等，(Biotechnology 8: 833-839 (1990))已公开了使用粒子轰击转化来自 A188-的玉米株系的技术。另外，WO 93/07278 和 Koziel 等，(Biotechnology 11: 194-200 (1993))描述了通过粒子轰击转化玉米良种自交系的技术。该技术利用在授粉后 14-15 天从玉米穗中切下的 1.5-2.5 mm 长的未成熟玉米胚和 PDS-1000He 生物射弹设备用于轰击。

稻的转化也可利用原生质体或粒子轰击通过直接基因转移技术进行。已针对 Japonica-型和 Indica-型描述了原生质体介导的转化(Zhang 等，Plant Cell Rep 7: 379-384 (1988); Shimamoto 等，Nature 338: 274-277 (1989); Datta 等，Biotechnology 8: 736-740 (1990))。这两种类型常规地也可使用粒子轰击转化(Christou 等，Biotechnology 9: 957-962 (1991))。另外，WO 93/21335 描述了通过电穿孔转化稻的技术。

专利申请 EP 0 332 581 描述了用于生产、转化和再生早熟禾亚科(Pooideae)原生质体的技术。这些技术允许转化鸭茅属(Dactylis)和小麦。另外，Vasil 等 (Biotechnology 10: 667-674 (1992))已描述了使用粒子轰击进 C 型长期可再生愈伤组织中的小麦转化，Vasil 等 (Biotechnology 11: 1553-1558 (1993))和 Weeks 等 (Plant Physiol. 102: 1077-1084 (1993))还描述了使用粒子轰击未成熟胚和来自未成熟胚的愈伤组织的小麦转化。然而，用于小麦转化的特异技术涉及通过粒子轰击未成熟胚转化小麦，并且在基因递送之前包括高蔗糖或高麦芽糖步骤。在轰击之前，将任何数量的胚(长度为 0.75-1 mm)涂布于含 3%蔗糖(Murashiga & Skoog, Physiologia Plantarum 15: 473-497 (1962))和 3 mg/l 2,4-D 的 MS 培养基上用于诱导体细胞胚，这允许在暗中继续。在选择轰击日，将胚从诱导培养基上取下并置于渗透剂(即以期望浓度添加蔗糖或麦芽糖的诱导培养基，所述期望浓度通常为 15%)上。允许胚质壁分离 2-3 小时，然后进行轰击。典型的是每个靶平板二十个胚，尽管不是必须的。使用标准步骤将带有适当基因的质粒(如 pCIB3064 或 pSG35)沉淀在微米大小的金颗粒上。使用~1000

psi 的爆发压, 使用标准 80 筛目, 用 DuPont Biolistics® 氦设备轰击各个胚平板。在轰击后, 将胚放置回暗中并恢复约 24 小时 (仍然在渗透剂上)。24 小时后, 将胚从渗透剂上取出并放回诱导培养基上, 在那里保持一个月后再生。约一个月后, 将具有发育中胚胎发生愈伤组织的胚外植体转移至再生培养基 (MS + 1 mg/升 NAA, 5 mg/升 GA), 所述再生培养基还含有适当的选择剂 (在 pCIB3064 的情况下为 10 mg/l basta, 在 pSOG35 的情况下为 2 mg/l 甲氨基蝶呤)。约一个月后, 将发育的茎转移至已知为 "GA7s" 的更大无菌容器中, 所述容器含有半浓度 MS、2% 蔗糖和相同浓度的选择剂。

还描述了使用农杆菌对单子叶植物的转化。见 WO 94/00977 和美国专利第 5,591,616 号, 均引入本文作为参考。还参见 Negrotto 等, *Plant Cell Reports* 19: 798-803 (2000), 其引入本文作为参考。对该实例而言, 使用稻 (*Oryza sativa*) 产生转基因植物。可使用多种稻栽培种 (Hiei 等, 1994, *Plant Journal* 6:271-282; Dong 等, 1996, *Molecular Breeding* 2:267-276; Hiei 等, 1997, *Plant Molecular Biology*, 35:205-218)。另外, 下文所述的多种培养基组分可在数量上变化或被取代。通过在 MS-CIM 培养基 (MS 基础盐, 4.3 g/升; B5 维生素 (200 x), 5 ml/升; 蔗糖, 30 g/升; 脯氨酸, 500 mg/升; 谷氨酰胺, 500 mg/升; 酪蛋白水解产物, 300 mg/升; 2,4-D (1 mg/ml), 2 ml/升; 用 1 N KOH 调节 pH 至 5.8; Phytigel, 3 g/升) 上起始胚胎发生应答和/或从成熟的胚建立培养物。接种培养物应答起始阶段的成熟胚或确立的培养物株系, 并与含有期望的载体构建体的根瘤农杆菌菌株 LBA4404 (农杆菌属) 共同培养。在 28°C 将来自于甘油储存液的农杆菌在固体 YPC 培养基 (100 mg/L 壮观霉素和任何其他适当的抗生素) 上培养~2 天。将农杆菌重悬于液体 MS-CIM 培养基中。将农杆菌培养物稀释至 OD600 为 0.2-0.3 并添加乙酰丁香酮至 200  $\mu$ M 的终浓度。在溶液与稻培养物混合之前添加乙酰丁香酮, 以诱导农杆菌将 DNA 转移至植物细胞。为了接种, 将植物培养物浸入细菌悬浮液中。去除液体细菌悬浮液, 将接种的培养物置于共同培养培养基上并在 22°C 孵育两天。然后将培养物转移至含替卡西林 (400

mg/升)的 MS-CIM 培养基上以抑制农杆菌的生长。对使用 PMI 可选择标记物基因的构建体(Reed 等, *In Vitro Cell. Dev. Biol.-Plant* 37:127-132)而言,在 7 天后将培养物转移至含甘露糖作为糖类来源的选择培养基(含 2%甘露糖、300 mg/升替卡西林的 MS),并在暗中培养 3-4 周。然后将抗性菌落转移至再生诱导培养基(不含 2,4-D,含 0.5 mg/升 IAA、1 mg/升玉米素、200 mg/升特美汀、2%甘露糖和 3%山梨糖醇的 MS),并在暗中培养 14 天。然后将增殖的菌落转移至另一轮再生诱导培养基,并移至光照培养室。将再生的茎转移至含 GA7-1 培养基(不含激素并含有 2%山梨糖醇的 MS)的 GA7 容器中保持 2 周,然后当它们足够大并具有足够的根时移至温室。将植物在温室中移植至土壤(T0 世代),生长至成熟,并收获 T1 种子。

### 3.质体的转化

将珊西烟(*Nicotiana tabacum* c.v. 'Xanthi nc')的种子在 T 琼脂培养基上以 1"的环形阵列每盘七粒播种在 T 琼脂培养基上,并在播种后 12-14 天用 1  $\mu\text{m}$  钨颗粒(M10, Biorad, Hercules, CA)轰击,所述颗粒如(Svab, Z. 和 Maliga, P. (1993) *PNAS* 90, 913-917)所述主要来自质粒 pPH143 和 pPH145 的 DNA 包裹。将轰击的幼苗在 T 培养基上孵育两天,之后切下叶,并远轴向上置于光明中(350-500  $\mu\text{mol}$  光子/ $\text{m}^2/\text{s}$ )含 500  $\mu\text{g}/\text{ml}$  盐酸壮观霉素(Sigma, St. Louis, MO)的 RMOP 培养基平板上(Svab, Z., Hajdukiewicz, P. 和 Maliga, P. (1990) *PNAS* 87, 8526-8530)。将轰击后三到八周在漂白的叶下出现的抗性茎亚克隆至相同的选择培养基上,允许形成愈伤组织,并对次生茎进行分离和亚克隆。通过 Southern 印迹的标准技术(Sambrook 等, (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor)评价独立的亚克隆中转化的质体基因组拷贝(同质性)的完全分离。在 1% Tris-硼酸盐(TBE)琼脂糖凝胶上分离 BamHI/EcoRI-消化的总细胞 DNA(Mettler, I. J. (1987) *Plant Mol Biol Reporter* 5, 346-349),转移至尼龙膜(Amersham)并用  $^{32}\text{P}$ -标记的随机引物 DNA 序列作为探针检测,该序列对应于 0.7 kb BamHI/HindIII

DNA 片段, 所述片段来自含有部分 rps7/12 质体靶向序列的 pC8。使同质性茎在含壮观霉素的 MS/IBA 培养基(McBride, K. E.等, (1994) PNAS 91, 7301-7305)上无菌生根, 并转移至温室。

## V. 育种和种子产生

### A. 育种

通过用本发明的核酸序列转化获得的植物可以是大量植物物种中任一, 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地选自上文公开的具有农业重要性的靶作物。本发明基因的表达与具有生产和品质重要性的其他特性组合, 可通过育种整合进植物中。育种方法和技术是本领域已知的。参见例如 Welsh J. R., *Fundamentals of Plant Genetics and Breeding*, John Wiley & Sons, NY (1981); *Crop Breeding*, Wood D. R. (编著) American Society of Agronomy Madison, Wisconsin (1983); Mayo O., *The Theory of Plant Breeding*, 第二版, Clarendon 出版, Oxford (1987); Singh, D.P., *Breeding for Resistance to Diseases and Insect Pests*, Springer-Verlag, NY (1986); 和 Wricke 和 Weber, *Quantitative Genetics and Selection Plant Breeding*, Walter de Gruyter 和 Co., Berlin (1986)。

操作进入上述转基因种子和植物中的遗传特性通过有性繁殖或营养生长传递, 并因此可在后代植物中保留和繁殖。一般地, 所述保留和繁殖利用被开发为适应特定目的的已知农业方法, 例如耕耘、播种或收获。也可以使用专门化的方法如水培法或温室技术。因为生长中的作物容易受到由昆虫或感染引起的攻击和损伤以及与杂草植物的竞争, 所以采取手段控制杂草、植物疾病、昆虫、线虫和其他不利条件以提高产量。这些包括机械手段如耕耘土壤或去除杂草和感染的植物, 以及应用农药如除草剂、杀真菌剂、杀配子剂、杀线虫剂、生长调节剂、熟化剂和杀昆虫剂。

在植物育种中可进一步利用转基因植物和根据本发明的种子的有利遗传特性, 其目的在于开发具有改良的特性的植物, 所述改良特性例如对害虫、除草剂或胁迫的耐受, 改进的营养价值, 提高的产量, 或引起更少倒

伏或落粒损失的改良的结构。多种育种步骤由明确确定的人为介入表征，例如选择待杂交的株系、指导亲本株系的传粉、或选择适当的后代植物。根据期望的特性，采取不同的育种手段。相关技术是本领域公知的，包括但不限于杂交、近交、回交育种、多株系育种(multiline breeding)、变种掺合(variety blend)、种间杂交、非整倍体技术等等。杂交技术还包括通过机械、化学或生物化学手段对植物不育化，产生雄性或雌性不育植物。用不同株系的花粉对雄性不育植物异花传粉确保雄性不育但雌性能育植物的基因组会均一地获得两种亲本株系的特性。因此，根据本发明的转基因种子和植物可用于改良的植物株系育种，例如提高常规方法如除草剂或杀虫剂处理的有效性，或允许植物由于其修饰的遗传特性而进行所述方法。或者，可获得具有改进的胁迫耐受的新作物，其由于最优化的遗传“装置”产生收获的产物，该产物比不能耐受同等不利发育条件的产物具有更好的品质。

#### B.种子生产

在种子生产中，萌发品质和种子的均一性是关键的产品特性。因为难以使作物不含其他作物和杂草的种子，所以为了控制种子带有的疾病和生产具有良好萌发的种子，在培育、调节(conditioning)和销售纯净种子领域有经验的种子生产者开发了相当广泛和明确定义的种子生产实践。因此对农民而言，购买满足特定品质标准的经过检验的种子而不是使用从其自己作物收获的种子是常见的实践。作为种子使用的繁殖材料通常用防护涂层处理，所述防护涂层包含除草剂、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。通常使用的保护性涂层包含化合物如环己烯亚胺、萎锈灵(carboxin)、塞仑(TMTD $\text{\textcircled{O}}$ )、methalaxyl (Apron $\text{\textcircled{O}}$ )和甲基嘧啶磷(Actellic $\text{\textcircled{O}}$ )。如果期望的话，将这些化合物与制剂领域常用的其他运载体、表面活性剂或促进应用的佐剂一起配制，提供针对细菌、真菌或动物害虫引起的损伤的保护。可通过用液体制剂浸渍遗传材料，或通过用组合的湿或干制剂涂布来应用保护性涂层。其他应用方法如在芽或果实处定向处理也是可能的。

## VI. 改变核酸分子的表达

以如下途径之一实现本发明核酸分子表达的改变：

### A. “有义” 阻抑

通过“有义”阻抑获得本发明核苷酸序列表达的改变，特别是其表达的降低（参考例如 Jorgensen 等，(1996) *Plant Mol. Biol.* 31, 957-973）。在该情况下，本发明核苷酸序列的整体或部分包含在 DNA 分子中。该 DNA 分子与在包含靶基因的细胞（特别是植物细胞）中有功能的启动子特异地有效连接，并被引入该细胞，其中该核苷酸序列可表达。核苷酸序列以“有义方向”插入 DNA 分子中，这意味着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，核苷酸序列是可完全翻译的，且核苷酸序列中包含的所有遗传信息或其部分被翻译为多肽。在另一具体的实施方案中，核苷酸序列是可部分翻译的，并且翻译了一条短肽。在一个具体的实施方案中，这通过在核苷酸序列中插入至少一个过早的终止密码子实现，所述终止密码子使翻译停止。在另一个更具体的实施方案中，核苷酸序列被转录，但是未产生翻译产物。这通常通过去除核苷酸序列编码的多肽的起始密码子例如“ATG”实现。在一个更具体的实施方案中，包含核苷酸序列或其部分的 DNA 分子被稳定整合在植物细胞的基因组中。在另一具体的实施方案中，包含核苷酸序列或其部分的 DNA 分子包含在染色体外复制分子中。

在含有上段所述 DNA 分子之一的转基因植物中，对应于 DNA 分子中包含的核苷酸序列的核苷酸序列的表达被特异降低。特别地，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，更特别地至少 80% 相同，还更特别地至少 90% 相同，还更特别地至少 95% 相同，还更特别地至少 99% 相同。

### B. “反义” 阻抑

在另一具体的实施方案中，通过“反义”阻抑获得发明核苷酸序列表达的改变，特别是其表达的降低。本发明核苷酸序列的整体或部分包含在 DNA 分子中。该 DNA 分子与在植物细胞中有功能的启动子特异地有效连

接, 并被引入该细胞, 其中该核苷酸序列可表达。核苷酸序列以“反义方向”插入 DNA 分子中, 这意味着核苷酸序列的反向互补序列(有时也称为非编码链)可以被转录。在一个具体的实施方案中, 包含核苷酸序列或其部分的 DNA 分子被稳定整合在植物细胞的基因组中。在另一具体的实施方案中, 包含核苷酸序列或其部分的 DNA 分子包含在染色体外复制分子中。引用了描述该途径的若干出版物用于进一步阐述(Green, P. J.等, *Ann. Rev. Biochem.* 55:569-597 (1986); van der Krol, A. R.等, *Antisense Nuc. Acids & Proteins*, 第 125-141 页 (1991); Abel, P. P.等, *PNAS Proc. Natl. Acad. Sci. USA* 86:6949-6952 (1989); Ecker, J. R.等, *Proc. Natl. Acad. Sci. USANAS* 83:5372-5376 (1986 年 8 月))。

在含有上段所述 DNA 分子之一的转基因植物中, 对应于 DNA 分子中包含的核苷酸序列的核苷酸序列的表达被特异降低。特别地, 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, 更特别地至少 80%相同, 还更特别地至少 90%相同, 还更特别地至少 95%相同, 还更特别地至少 99%相同。

### C. 同源重组

在另一具体的实施方案中, 通过同源重组在基因组中修饰了对应于本发明核苷酸序列的至少一个基因组拷贝, 所述同源重组如 Paszkowski 等, *EMBO Journal* 7:4021-26 (1988)中所进一步阐述的。该技术利用同源序列的下述特性: 识别彼此并通过本领域已知为同源重组的过程互相交换核苷酸序列。同源重组可在细胞中核苷酸序列的染色体拷贝和通过转化引入细胞的核苷酸序列的进入拷贝之间发生。特异的修饰因此被精确地引入核苷酸序列的染色体拷贝中。在一个实施方案中, 修饰了本发明核苷酸序列的调节元件。这类调节元件可通过使用本发明核苷酸序列或其部分作为探针筛选基因组文库容易地获得。现存的调节元件被不同的调节元件代替从而改变核苷酸序列的表达, 或其被突变或缺失从而消除核苷酸序列的表达。在另一实施方案中, 通过缺失部分核苷酸序列或整体核苷酸序列或通过突变修饰核苷酸序列。本发明还涉及突变的多肽在植物细胞中的表达。已描

述了对破坏内源植物基因的这些技术更近期的改进(Kempin 等, Nature 389:802-803 (1997), 以及 Miao 和 Lam, Plant J., 7:359-365 (1995)。

在另一具体的实施方案中, 通过用嵌合寡核苷酸转化细胞在核苷酸序列的染色体拷贝中引入突变, 所述嵌合寡核苷酸由双链体构型的一段连续的 RNA 和 DNA 残基组成, 其末端上带有双重发夹帽。寡核苷酸的另一特性是例如 RNA 残基处 2'-O-甲基化的存在。RNA/DNA 序列被设计为与本发明核苷酸序列的染色体拷贝序列排列在一条链上 (align), 并含有期望的核苷酸改变。例如, 该技术在美国专利 5,501,967 和 Zhu 等, (1999) Proc. Natl. Acad. Sci. USA 96: 8768-8773 中进一步阐述。

#### D. 核酶

在另一实施方案中, 用对编码本发明多肽的 RNA 特异的催化性 RNA 或核酶切割编码本发明多肽的 RNA。核酶在转基因植物中表达并导致植物细胞中编码本发明多肽的 RNA 数量减少, 从而导致细胞中累积的多肽数量减少。该方法在美国专利 4,987,071 中进一步阐述。

#### E. 显性失活突变体

在另一具体的实施方案中, 改变了本发明核苷酸序列编码的多肽活性。这通过在转基因植物中表达蛋白质的显性失活突变体导致内源蛋白质活性丧失来实现。

#### F. 适体

在又一实施方案中, 通过在转基因植物中表达与蛋白质特异结合的核酸配体 (即所谓的适体) 抑制本发明多肽的活性。优选地通过 SELEX (通过指数式富集法对配体的系统性进化) 方法获得适体。在 SELEX 方法中, 具有随机化序列区域的单链核酸的候选混合物与蛋白质接触, 并从候选混合物的剩余部分分离对靶具有提高的亲力的那些核酸。扩增分离的核酸产生富含配体的混合物。若干次重复后获得对多肽具有最优亲和力的核酸, 并用于在转基因植物中表达。该方法在美国专利 5,270,163 中进一步阐述。

#### G. 锌指蛋白

也使用与本发明核苷酸序列或其调节区结合的锌指蛋白改变核苷酸序

列的表达。特别地，核苷酸序列的转录被降低或提高。锌指蛋白例如描述于 Beerli 等，(1998) PNAS 95:14628-14633. 中或 WO 95/19431、WO 98/54311 或 WO 96/06166 中，所述文献均通过参考整体并入本文。

#### H. dsRNA

还通过例如 WO 99/32619、WO 99/53050 或 WO 99/61631 中所述的 dsRNA 干扰获得本发明核苷酸序列表达的改变，所述文献均通过参考整体并入本文。在另一具体的实施方案中，通过双链 RNA(dsRNA)干扰获得本发明核苷酸序列表达的改变，特别是表达的降低。本发明核苷酸序列的整体，特别是其部分包含在 DNA 分子中。该 DNA 分子的大小特别地为 100 到 1000 个核苷酸或更多；最佳大小可根据经验确定。同一 DNA 分子的两个拷贝是连接的，由间隔 DNA 分子隔开，使得第一拷贝和第二拷贝处于相反的方向。在具体的实施方案中，DNA 分子的第一拷贝是反向互补序列（也已知为非编码链），第二拷贝是编码链；在最具体的实施方案中，第一拷贝是编码链，第二拷贝是反向互补序列。间隔 DNA 分子的大小特别地为 200 到 10,000 个核苷酸，更特别地为 400 到 5000 个核苷酸，最特别地为 600 到 1500 个核苷酸长。间隔特别地是 DNA 的随机片段，更特别地是与 dsRNA 干扰的靶生物无同源性的 DNA 随机片段，最特别地是被靶生物有效剪接的功能性内含子。被间隔分开的 DNA 分子的两个拷贝与在植物细胞中有功能的启动子有效连接，并被引入植物细胞中，核苷酸序列可在该细胞中表达。在一个具体的实施方案中，包含核苷酸序列或其部分的 DNA 分子被稳定整合进植物细胞的基因组中。在另一具体的实施方案中，包含核苷酸序列或其部分的 DNA 分子包含在染色体外复制的分子中。引用了描述该方法的若干出版物用于进一步阐述(Waterhouse 等，(1998) PNAS 95:13959-13964; Chuang 和 Meyerowitz (2000) PNAS 97:4985-4990; Smith 等，(2000) Nature 407:319-320)。通过 dsRNA 干扰改变核苷酸序列的表达还描述于例如 WO 99/32619、WO 99/53050 或 WO 99/61631 中，其均通过参考整体并入本文。

在含有上段所述 DNA 分子之一的转基因植物中，对应于 DNA 分子中

包含的核苷酸序列的核苷酸序列的表达被特异降低。特别地，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，更特别地至少 80%相同，还更特别地至少 90%相同，还更特别地至少 95%相同，还更特别地至少 99%相同。

#### I. 插入 DNA 分子（插入诱变）

在另一具体的实施方案中，DNA 分子被插入本发明核苷酸序列的染色体拷贝或其调节区中。特别地，这类 DNA 分子包含能够在植物细胞中转录的转座元件，例如 Ac/Ds、Em/Spm、突变子(mutator)。或者，DNA 分子包含农杆菌 T-DNA 的 T-DNA 边界。T-DNA 分子也可包含重组酶或整合酶识别位点，该位点可用于从植物细胞的染色体去除 DNA 分子的部分。也包括使用 T-DNA、转座子、寡核苷酸的插入诱变方法或本领域技术人员已知的其他方法。使用 T-DNA 和转座子用于插入诱变的方法描述于 Winkler 等, (1989) *Methods Mol. Biol.* 82:129-136 和 Martienssen (1998) *PNAS* 95:2021-2026 中，其通过参考整体并入本文。

#### J. 缺失诱变

在另一实施方案中，通过缺失核苷酸序列或调节序列的一部分，在细胞或植物中序列的基因组拷贝中创建本发明核酸分子的突变。缺失诱变的方法是本领域技术人员已知的。参见例如 Miao 等, (1995) *Plant J.* 7:359。

在另一实施方案中，通过化学诱变或辐射在大的植物群体中随机产生缺失，并通过正求或反求遗传学分离在本发明的基因中具有缺失的植物。已知用快中子或  $\gamma$  射线辐照引起植物中的缺失突变(Silverstone 等, (1998) *Plant Cell*, 10:155-169; Bruggemann 等, (1996) *Plant J.*, 10:755-760; Redei 和 Koncz, 在 *Methods in Arabidopsis Research*, World Scientific 出版 (1992), 第 16-82 页)。如在秀丽新小杆线虫(*C. elegans*)中所示(Liu 等, (1999), *Genome Research*, 9:859-867.)，可使用 PCR 在反求遗传学策略中恢复本发明基因中的缺失突变，所述 PCR 使用合并的基因组 DNA 集合。正求遗传学策略应涉及展示 PTGS 的株系的诱变，随后针对 PTGS 的缺失筛选 M2 后代。可期望这些突变体中的一些破坏了本发明的基因。这可通过针对本

发明基因的 Southern 印迹或 PCR, 用来自这些突变体的基因组 DNA 进行评价。

#### K. 在植物细胞中过量表达

在另一具体的实施方案中, 本发明的编码多肽的核苷酸序列被过量表达。用于过量表达本发明核酸分子的核酸分子和表达盒的实例在上文描述。本发明还包括本领域技术人员已知用于过量表达核酸分子的方法。

在一个具体的实施方案中, 在植物的每个细胞中改变了本发明核苷酸序列的表达。这例如通过同源重组或通过染色体中插入获得。也可通过例如在下述启动子的控制下表达有义或反义 RNA、锌指蛋白或核酶获得, 所述启动子能够在植物的每个细胞中表达有义或反义 RNA、锌指蛋白或核酶。组成型表达, 诱导性、组织特异性或发育调节性表达也在本发明的范围内, 并导致本发明的核苷酸序列在植物细胞中表达的组成型、诱导性、组织特异性或发育调节性改变。根据本发明的教导(例如如下文所述)制备下述构建体并转化进植物细胞中, 所述构建体用于表达有义或反义 RNA、锌指蛋白或核酶, 或用于过量表达本发明的核苷酸序列。

#### VII. 多肽

本发明还涉及包含氨基酸序列 SEQ ID NO: 2 的分离的多肽。具体地, 包含氨基酸序列 SEQ ID NO: 2 的分离的多肽和具有保守氨基酸修饰的变体。本领域技术人员会明白, 对核酸、肽、多肽或蛋白质序列的各个取代、缺失或添加(其改变、添加或缺失编码的序列中单个氨基酸或小比例的氨基酸)是“保守修饰”, 所述修饰导致用化学上类似的氨基酸取代氨基酸。保守修饰的变体提供与未经修饰的多肽相似的生物学活性。列出功能类似的氨基酸的保守取代的表格是本领域已知的。参见 Crighton (1984) *Proteins*, W.H. Freeman and Company。

在一个具体的实施方案中, 与多肽序列 SEQ ID NO:2 或其外显子或结构域具有基本相似性的多肽是 SEQ ID NO:2 所示多肽序列的等位变体。在另一具体的实施方案中, 与 SEQ ID NO:2 所示多肽序列具有基本相似性的多肽或其外显子或结构域, 是 SEQ ID NO:2 所示多肽序列的天然存在的变

体。在另一具体的实施方案中，与 SEQ ID NO:2 所示多肽序列具有基本相似性的多肽或其外显子或结构域，是 SEQ ID NO:2 所示多肽序列的多态变体。

在一个备选的具体的实施方案中，具有基本相似性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缺失或插入为少于约 10 个氨基酸。在一个最具体的实施方案中，缺失或插入为少于约三个氨基酸。

在一个具体的实施方案中，具有基本相似性的序列编码至少一个氨基酸处的取代。

本发明的实施方案还考虑了分离的多肽，该多肽包含多肽序列，该多肽序列包括

- (a) SEQ ID NO: 2 中所列出的多肽序列、或它的外显子或结构域；
- (b) 与 (a) 具有基本相似性的多肽序列；
- (c) 多肽序列，该多肽序列由核苷酸序列进行编码，该核苷酸序列与 SEQ ID NO: 1 中所列出的核苷酸序列、或其外显子或结构域、或与其它互补的序列相同或具有基本相似性；
- (d) 多肽序列，该多肽序列由核苷酸序列进行编码，该核苷酸序列能够在中度严格性条件下与 SEQ ID NO: 1 中所列出的核苷酸序列或与之互补的序列进行杂交；或
- (e) (a)、(b)、(c) 或 (d)的功能片段。

在另一具体的实施方案中，具有基本相似性的多肽是 SEQ ID NO:2 所示多肽序列或其片段、结构域、重复或嵌合体的等位变体。在另一具体的实施方案中，分离的核酸包含来自下述核苷酸序列编码的多肽序列的大量区域，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与之互补的序列相同或具有基本相似性。

在另一具体的实施方案中，多肽是 SEQ ID NO:2 所示多肽。在另一具体的实施方案中，多肽是功能性片段或结构域。在另一具体的实施方案中，多肽是嵌合体，其中该嵌合体可包含功能性蛋白质结构域（包括结构域、

重复、翻译后修饰位点)或其他特性。在一个更具体的实施方案中,多肽是植物多肽。在一个更具体的实施方案中,植物是双子叶植物。在一个更具体的实施方案中,植物是裸子植物。在一个更具体的实施方案中,植物是单子叶植物。在一个更具体的实施方案中,该单子叶植物是谷物。在一个更具体的实施方案中,谷物可以是例如玉米、小麦、大麦、燕麦、裸麦、粟、高粱、黑小麦、黑麦属、单粒小麦、斯佩耳特小麦、双粒小麦、画眉草、蜀黍、亚麻、格兰马草、磨擦草属物种和玉米草。在另一具体的实施方案中,谷物是稻。

在一个具体的实施方案中,多肽在植物的特异位置或组织中表达。在一个更具体的实施方案中,该位置或组织为例如但不限于表皮、维管组织、分生组织、形成层、皮层或髓。在一个最具体的实施方案中,该位置或组织为叶或鞘、根、花和发育中的胚珠或种子。在一个更具体的实施方案中,该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、叶和花。在一个更具体的实施方案中,该位置或组织为种子。

在一个具体的实施方案中,由下述核苷酸序列编码的多肽序列包含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 所示核苷酸序列,或其片段或结构域或与之互补的序列具有基本相似性。在一个更具体的实施方案中,缺失或插入少于约三十个核苷酸。在一个最具体的实施方案中,缺失或插入少于约五个核苷酸。

在一个具体的实施方案中,由下述核苷酸序列编码的多肽序列包含至少一个密码子的取代,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与之互补的序列具有基本的相似性。在一个更具体的实施方案中,该取代是保守的。

在一个具体的实施方案中,与 SEQ ID NO:2 所示多肽序列或其片段、结构域、重复或嵌合体具有基本相似性的多肽序列包含至少一个氨基酸的缺失或插入。

本发明的多肽、其片段或变体可包含来自本发明多肽的任何数量的连续氨基酸残基,其中所述残基数量选自自由 10 到本发明全长多肽的残基数组

成的整数组。特别地，多肽的部分或片段是功能性蛋白质。本发明包括下述活性多肽，其具有天然（非合成的）内源多肽比活性的至少 20%、30% 或 40%，特别地至少 50%、60% 或 70%，最特别地至少 80%、90% 或 95%。另外，底物特异性(kcat/Km)任选地与天然（非合成的）的内源多肽基本相似。通常，Km 会是天然内源多肽的至少 30%、40% 或 50%；更特别地是至少 60%、70%、80% 或 90%。测定和定量测定活性和底物特异性的方法是本领域技术人员公知的。

作为免疫原存在时，本发明的分离的多肽会引发与本发明多肽特异反应的抗体的生产。因此，为了例如但不限于免疫测定或蛋白质纯化技术的目的，本发明的多肽可用作免疫原，用于构建与本发明的蛋白质免疫反应的抗体。测定结合的免疫测定法是本领域技术人员公知的，例如但不限于 ELISA 或竞争性免疫测定法。

本发明的实施方案还涉及由本公开内容的分离的核酸分子编码的嵌合多肽，其包括含有由下述分离的核酸编码的多肽序列的嵌合多肽，所述分离的核酸含有下述核苷酸序列，其包括：

- (a) 如 SEQ ID NO:1 所示的核苷酸序列或其外显子或结构域，
- (b) 与(a)具有基本相似性的核苷酸序列；
- (c) 能够与(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；或
- (f) 其功能性片段。

含有由分离的核酸编码的多肽序列的多肽，所述核酸含有编码下述多肽的核苷酸序列、其互补序列或其反向互补序列，所述多肽包含下述多肽序列：

- (a) SEQ ID NO:2 所示的多肽序列，或其结构域、重复或嵌合体；
- (b) 与(a)具有基本相似性的多肽序列；
- (c) 由下述核苷酸序列编码的多肽序列，所述核苷酸序列与 SEQ ID NO:1 所示核苷酸序列或其外显子或结构域或与之互补的序列相同或具有

基本的相似性;

(d) 由下述核苷酸序列编码的多肽, 所述核苷酸序列在中度严格性条件下能够与 SEQ ID NO:1 所示核苷酸序列或与之互补的序列杂交; 和(a)、(b)、(c)或(d)的功能性片段; 或

(e)其功能性片段。

本发明的分离的核酸分子适用于在重组改造的细胞如细菌、酵母、昆虫、哺乳动物或植物细胞中表达本发明的多肽。该细胞在非天然条件(例如数量、组成、定位和/或时间)下产生多肽, 因为其被遗传改变为如此。本领域技术人员知道可用于表达编码本发明蛋白质的核酸的大量表达体系, 并且不在下文中详细描述这些体系。

简言之, 通常例如通过将核酸或 cDNA 与启动子(组成型或可调节的)有效连接, 然后整合进表达载体中实现编码本发明的多肽的分离的核酸的表达。载体适用于在原核生物或真核生物任一中复制和/或整合。常用的表达载体包含转录和翻译终止子、起始序列和用于调节编码多肽的核酸分子表达的启动子。为了获得所克隆的核酸分子的高水平表达, 期望使用指导转录的包含强启动子、用于翻译起始的核糖体结合位点和转录/翻译终止子的表达载体。本领域技术人员应当明白可对本发明的多肽进行修饰而不减小其生物学活性。可进行一些修饰来利于本发明多肽的克隆、表达或进入融合蛋白的整合。此类修饰是本领域公知的, 并包括但不仅限于在氨基端添加以提供起始位点的甲硫氨酸, 或置于任一末端以创建方便定位纯化序列的额外氨基酸(例如多聚组氨酸(poly Histadine))。也可向载体中引入限制性位点或终止密码子。

在一个具体的实施方案中, 表达载体包含一个或多个元件, 例如但不限于启动子增强子序列、选择标记物序列、复制起点、表位标签编码序列, 或亲和纯化标签编码序列。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、遍在蛋白启动子和菜豆蛋白启动子。在另一实施方案中, 启动子可在植物中工作, 更特别地为组成型或诱导性启动子。在另一具体的实施方案中,

选择标记物序列编码抗生素抗性基因。在另一具体的实施方案中，表位标签序列编码 V5、肽 Phe-His-His-Thr-Thr、血凝素或谷胱甘肽-S-转移酶。在另一具体的实施方案中，亲和纯化标签序列编码多聚氨基酸序列或多肽。在一个更具体的实施方案中，多聚氨基酸序列为多聚组氨酸。在一个更具体的实施方案中，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，亲和纯化标签序列包含内含肽编码序列。

可使用原核细胞作为宿主细胞，例如但不仅限于大肠杆菌和本领域已知的其他微生物菌株。用于在原核生物中表达蛋白质的方法是本领域技术人员公知的，并可见于许多实验室手册如 *Molecular Cloning: A Laboratory Manual*, J. Sambrook 等, (1989, Cold Spring Harbor Laboratory 出版)中。本领域技术人员可获得大量控制表达的启动子、核糖体结合位点和操纵子，也可获得可选择标记物如抗生素抗性基因。选择的载体类型是为了允许最佳生长和在选定的细胞类型中表达。

可获得大量真核生物表达体系，例如但不仅限于酵母、昆虫细胞系、植物细胞和哺乳动物细胞。异源蛋白质在酵母中的表达和合成是公知的(见 Sherman 等, *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory 出版, 1982)。广泛用于产生真核生物蛋白质的常用酵母菌株是酿酒酵母 (*Saccharomyces cerevisiae*)和巴斯德毕赤酵母(*Pichia pastoris*), 并且可从商业供应商(例如 Invitrogen)获得用于表达的载体、菌株和方案。

可用于产生蛋白质的表达载体转染哺乳动物细胞体系。本领域技术人员可获得许多合适的宿主细胞系，例如但不限于 HEK293、BHK21 和 CHO 细胞系。用于这些细胞的表达载体可包含表达控制序列如复制起点、启动子(例如 CMV 启动子、HSV tk 启动子或磷酸甘油酸酯激酶(pgk)启动子)、增强子和蛋白质加工位点，如核糖体结合位点、RNA 剪接位点、多聚腺苷酸化位点和转录终止序列。适用于产生蛋白质的其他动物细胞系可商业获得或来自保藏机构如美国典型培养物保藏中心。

用于在昆虫细胞中表达蛋白质的表达载体通常来自 SF9 杆状病毒或本领域已知的其他病毒。可获得大量合适的昆虫细胞系，包括但不仅限于蚊

幼虫、蚕、黏虫(*armyworm*)、蛾和果蝇(*Drosophila*)细胞系。

转染动物和低等真核细胞的方法是已知的。使用大量方法制备真核细胞感受态以引入 DNA，所述方法包括但不限于：磷酸钙沉淀、受体细胞与含有 DNA 的细菌原生质体的融合、用含有 DNA 的脂质体处理受体细胞、DEAE 糊精、电穿孔、生物射弹和将 DNA 直接显微注射进细胞中。转化的细胞使用本领域公知的手段培养（参见 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，可使用本领域技术人员已知的方法将其从细胞中分离和纯化。可使用 Western 印迹技术或放射免疫测定或其他标准的免疫测定技术监测纯化过程。蛋白质纯化技术是本领域技术人员公知和使用的（参见 R. Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag, New York 1982; Deutscher, *Guide to Protein Purification*, Academic 出版 (1990)）。本发明的实施方案提供了产生重组蛋白质的方法，其中表达载体包含一个或多个元件，包括启动子增强子序列、选择标记物序列、复制起点、表位标签编码序列，和亲和纯化标签编码序列。在一个具体的实施方案中，核酸构建体包含表位标签编码序列，且分离步骤包括使用对该表位标签特异的抗体。在另一具体的实施方案中，核酸构建体含有多聚氨基酸编码序列，且分离步骤包括使用包含多聚氨基酸结合物质的树脂，特别是多聚氨基酸为多聚组氨酸且多聚氨基酸结合树脂为镍-带电琼脂糖树脂。在另一具体的实施方案中，核酸构建体含有多肽编码序列，且分离步骤包括使用含多肽结合物质的树脂，特别是多肽为壳多糖结合结构域且树脂含有壳多糖-琼脂糖凝胶。

本发明的多肽可以使用本领域技术人员已知的非细胞合成方法合成。用于固相合成的技术由 Barany 和 Mayfield, *Solid-Phase Peptide Synthesis*, 在 *Peptides: Analysis, Synthesis, Biology*, 第 2 卷第 3-284 页, *Special Methods in Peptide Synthesis, Part A*; Merrifield 等, *J. Am. Chem. Soc.* 85:2149-56 (1963) 和 Stewart 等, *Solid Phase Peptide Synthesis*, 第二版,

**Pierce Chem. Co., Rockford, IL (1984)描述。**

本发明还提供了修饰（即提高或降低）植物或其部分中本发明多肽的浓度或组成的方法。修饰可通过提高或降低植物中的浓度和/或组成（即本发明多肽的比例）实现。该方法包括向植物细胞中引入表达盒以获得转化的植物细胞或组织，并培养转化的植物细胞或组织，所述表达盒含有本发明的核酸分子，或编码如上所述 OsGATA11 序列的核酸。核酸分子可处于组成型或诱导性启动子的调节下。该方法还可包括以足够修饰植物或植物部分中浓度和/或组成的时间诱导或阻遏核酸分子序列在植物中的表达。

可使用本领域技术人员已知的方法对具有修饰的本发明核酸分子表达的植物或植物部分进行分析和选择，所述方法包括但不限于 Southern 印迹、DNA 测序或使用该核酸分子特异引物的 PCR 分析，并检测由其产生的扩增子。

一般，相对于缺乏表达盒的对照植物、植物部分或细胞，浓度或组成提高或降低至少 5%、10%、20%、30%、40%、50%、60%、70%、80% 或 90%。

在光合作用植物中糖是许多生命过程的中心调节物，如光合作用、碳和氮代谢，并且这种调节是通过调节基因表达（激活或阻遏有关基因）而实现的。糖控制基因表达的机理还不是很了解。这里所公开的这种 GATA 转录因子参与了调节糖感知，并且这种因子本身的表达受到 N 状况的变化影响。这个基因的表达提高能够产生具有提高产量的植物，特别是当糖信号传导途径的操作能够导致光合作用提高以及氮利用提高并改变种子、块茎（tuber）、根、以及其他贮藏器官中的源库关系时。

本发明会参考以下的详细实施例进一步描述。除非另有说明，这些实施例仅就说明的目的提供，而非旨在限制。

**实施例**

本文使用的标准重组 DNA 和分子克隆技术是本领域公知的，并由 J. Sambrook 等, *Molecular Cloning: A Laboratory Manual*, 第三版, Cold

Spring Harbor, NY: Cold Spring Harbor Laboratory 出版 (2001); 由 T.J. Silhavy、M.L. Berman 和 L.W. Enquist, Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984)和由 Ausubel, F.M.等, Current Protocols in Molecular Biology, New York, John Wiley and Sons Inc., (1988), Reiter 等, Methods in Arabidopsis Research, World Scientific 出版 (1992)和 Schultz 等, Plant Molecular Biology Manual, Kluwer Academic Publishers (1998)描述。

### 实验背景与操作

#### A. 在限制氮的条件下确定稻和玉米的生长条件

在过去的研究涉及到硝酸盐摄取和同化作用的基因的实验中, 诸位发明人以及其他已经利用了多种生长条件, 在这些生长条件下硝酸盐通常过量存在或整体缺乏。在后情况中, 硝酸盐被通常加至在其缺乏的条件下生长的植物中以便理解这些基因和其他基因的硝酸盐调节。尽管这种类型的极端处理在定义基因调节的某些方面是有用的, 但它不适宜于获得对氮限制的效果的更好的理解。诸位发明人已经定义了拟南芥属中氮限制生长的条件 (Bi 等人 2007 BMC Genomics 8:281)。这涉及发展了使用 Rockwool 的系统 (Hirai 等人, 1995 Plant Cell Physiol 36, 1331-1339) 并定义三种条件: 条件之一是其中生长达到最大; 条件之一是其中氮将生长限制到最大生长水平的 70%到 75%; 条件之一是其中更严重的限制到最大生长水平的 30%到 35%。氮限制作用为“胁迫”, 通过改变硝酸盐的浓度容易地使“胁迫”的量发生变化。诸位发明人通过测量硝酸盐、叶绿素 (它经常作为在田间条件下氮状况的反映而进行使用-参见例如, Fox RH 等人 2001 Agron J. 93, 590-597; Minotti PL 等人 1994 Hort Science 29, 1497-1550)、氨基酸水平、以及硝酸还原酶以及谷氨酰胺合成酶的活性而测定了生理学的“氮状况”, 从而给出评估突变品系的研究的基准线。

#### B. 在氮限制情况下针对拟南芥属植物的表达概况分析实验

转录本的表达概况分析 (expression profiling) 可被用来同时测定大量基因的 RNA 水平。过去已经做了大量的这些类型的实验, 并且如果该实验体系容易控制, 则这些实验可用来准确定位出生物在不同条件下的“表达状况”, 并使用这种信息对哪些基因和途径参与不同的过程做出假设。诸位发明人发现生长条件的差异越深, 在这些生长条件下所生长的植物之间的转录本的概况 (transcript profile) 的差异就越大, 并且破译哪些变化是最重要的就越困难。在这一领域中唯一公布的全基因组概况分析实验是在拟南芥属中的实验, 其中研究了在硝酸盐水平上的极端变化 (Wang R 等人 2003 Plant Physiol. 132, 556-67)。在氮限制的情况中, 诸位发明人研究了生长中的植物在慢性氮胁迫下的效果以及在可供使用的氮的水平上的变化。诸位发明人已经在拟南芥属中确定了不同氮水平对生长的影响。

研究了不同的氮水平对转录本概况的影响: 其中氮没有限制生长。对于拟南芥属而言, 诸位发明人收集了在不同的氮方案下生长的 4 周龄幼苗。为了得到在统计学上显著的结果, 收集了三个不同的样本 (生物学的一式三份)。使用 Arabidopsis GeneChip® 全基因组阵列 (Affymetrix) 进行转录本概况分析, 以便研究拟南芥属中的转录本的水平。进行了对于研究由这些实验所产生的大量数据而言所必需的生物信息学分析。通过研究氮限制对表达模式的作用, 诸位发明人可以准确定位出哪些途径参与了它们对养分胁迫的应答。

### 实施例 1

#### 材料与amp;方法

#### 植物生长条件

将泥炭藓和蛭石 (vermiculate) (1:4) (SunGro Horticulture Canada Ltd. BC, Canada) 用来培育稻的 Kaybonnet 植物, 一周一次添加含有不同量的硝酸盐的营养液直到收获。该营养液含有 4 mM MgSO<sub>4</sub>、5 mM KCl、5 mM CaCl<sub>2</sub>、1 mM KH<sub>2</sub>PO<sub>4</sub>、0.1 mM Fe-EDTA、0.5 mM MES (pH6.0)、9 μM MnSO<sub>4</sub>、0.7 μM Zn SO<sub>4</sub>、0.3 μM CuSO<sub>4</sub>、46 μM NaB<sub>4</sub>O<sub>7</sub> 以及 0.2 μM

(NH<sub>4</sub>)<sub>6</sub>Mo<sub>7</sub>O<sub>2</sub>。对于限制性 N 条件而言,使用 3mM 的 N 溶液,一周一次直至收获。对于充足的 N 条件而言,最初六周使用 10mM N 溶液一周一次,更换成 5mM 的 N 溶液又持续 6 周,并且更换成 3mM 的 N 溶液直至收获。最初四周使植物在生长室生长(在 28 至 30°C 给予 16 小时光照(约 400 μmolm<sup>-2</sup>s<sup>-1</sup>)并且在 22 至 24°C 处于黑暗中 8 小时),并接着有一周短日照的处理(10 小时光照/14 小时黑暗)。之后,将植物移到温室中生长直到收获。

#### 产生转基因稻植物

制备了过量表达 *OsGATA11* 或使 *OsGATA11* 沉默的构建体。RNAi 序列的序列显示在 SEQ ID NO: 8 中。包含 RNAi 序列作为由茎环结构分开的反向重复的构建体的序列显示在 SEQ ID NO: 9 中。对过量表达 *OsGATA11* 或使 *OsGATA11* 沉默 (RNAi) 的 T1 转基因种子进行分析。

#### 转基因植物的基因分型

将叶样品在 300 μl 缓冲液 (Strategic Diagnostics Inc. Part # 7000006) 中磨碎。将一个试纸条(dipstick)(Strategic Diagnostics Inc. Part # 7000052) 插至试管中并放置约 15 分钟,进行一段时间使得试纸条上的线条清晰。在条纹(strip)上出现一条红线(对照)指示阴性结果。在条纹上出现两条红线(对照和测试)指示阳性结果。

#### 通过半定量 RT-PCR 进行表达分析

使用所提取的一微克总 RNA 来制备 cDNA。*OsGATA11* 的引物是 5'-CGTCGAGCACCAAGGGCAAATC-3' (SEQ ID NO: 3) 以及 5'-GGATAGGGTCATGAGCAGCATGG-3' (SEQ ID NO: 4)。Os 微管蛋白 (*OsTubulin*) 的引物是: 5'-AGGAGGATGCCGCTAACAACTTTG-3' (SEQ ID NO: 5) 以及 5'-AAACAGCATTGGTGATTTTCAGGC-3' (SEQ ID NO: 6)。

#### 叶绿素的测量

使用 Minolta SPAD 502DL 叶绿素计 (Tokyo, Japan) 测量总叶绿素,或由乙醇进行提取并根据 Kirk (1968) 通过分光光度计对总叶绿素进

行测量。

### 代谢产物分析

从处于限制氮的条件 (3mM) 下所生长的 4 周龄野生型和转基因植物中收获叶子, 将叶子在液氮中冷冻并储存在 $-80^{\circ}\text{C}$ 用于以下的生物化学分析。将硝酸盐从冷冻的叶子中提取出来并根据 Clothern 等人 (1975) 的方法进行测定。用在 HEPES-KOH 缓冲液 (pH 7.4) 中的 80%、50%、0% 的乙醇成功提取了总氨基酸, 并将集合的上清液用于总氨基酸测定, 如 Rosen (1957) 所说明。为了提取可溶性蛋白质, 将冷冻的叶的粉末重悬于 100 mM HEPES-KOH (pH 7.5) + 0.1% Triton X-100 缓冲液中, 并以 14,000 rpm 离心 10 min。使用商业的蛋白质测定试剂盒 (Bio-Rad, Hercules, CA) 确定上清液中总的可溶性蛋白的含量。如由 Geiger 等人 (1998) 所说明提取可溶性糖, 并使用可商购的试剂盒 (Megazyme, Ireland) 测定葡萄糖、果糖以及蔗糖的含量。

### 结果

#### 对表型转基因植物的策略

用于最初的遗传和表型分析的策略涉及在主要为限制氮 (N) 的条件下生长来自每个构建体的 5 种转基因事件 (transgenic event) (约 18 株植物)。还有一些植物生长在 N 充足的条件下 (约 10 株植物)。将 PMI 条用于基因分型以检测可选择的标记物 PMI。通过半定量 RT-PCR 对转基因表达水平进行测试。记录了叶绿素水平、秆长度、分蘖数目、圆锥花序的数目、开花时间、种子产量以及枝条生物量。

#### 过量表达 *OsGATA11* 的植物的表型

*OsGATA11* 基因在蛋白质水平上与 *AtGATA* 基因 (*At4g26150*, 图 3) 具有约 34% 的相似性。当转基因植物在限制 N 的条件下大约 4 周龄时测量了总的叶绿素水平。与野生型对照植物 (6 株植物) 相比, 从 PMI 阳性植物 (3 至 6 株植物) 的平均值看, 至少两个转基因事件 (事件 5 和事件 6) 具有显著的更高的叶绿素含量 (图 4A)。那些转基因植物的 *OsGATA11* 基因的表达确实提高 (图 4B)。为了确定叶绿素水平能受到 *OsGATA11*

基因的表达水平的影响，对转基因的 RNAi OsGATA11 植物进行了分析。在转基因的 RNAi OsGATA11 植物中 *OsGATA11* 基因的表达水平显著降低（图 5A），并且实际上在那些植物中叶绿素水平显著更低（图 5B）。在限制 N 的条件下与 11 株野生型对照植物的平均值相比，从 10 株 PMI 阳性植物的平均值看，一个事件（事件 6）具有高出约 20% 的种子产量（图 6A）。在充足 N 的条件下与 6 株野生型对照植物的平均值相比，从 4 株 PMI 阳性植物的平均值看，这一相同的事件具有几乎两倍的种子产量（图 6B）。同样，生长在高 N 下的植物在从生长室转移到温室之后经历了胁迫，并且这些转基因植物对该胁迫的应答更好（图 7）。OsGATA11 基因的表达的调节影响了葡萄糖（图 8A）、果糖（图 8B）、以及蔗糖（图 8C）的水平，并且影响了硝酸盐（图 9A）、氨基酸（9B）以及蛋白质（图 9C）的水平。

现在已经通过以上的实例对本发明的具体的实施方案进行了说明，这些实例并非旨在进行限制，本发明将在以下权利要求书中进行进一步阐明。本领域的普通技术人员将会认识到，权利要求书也允许包含除了该权利要求书字面范围以外的等效物。

---

<110>	圭尔夫大学 (University of Guelph)	
	S·罗斯坦 (Rothstein, Steven)	
	Y-M·毕 (Bi, Yong-Mei)	
<120>	调节碳和氮的基因与蛋白质及其调节	
<130>	6580-364	
<150>	CA 2,584,934	
<151>	2007-04-17	
<160>	9	
<170>	PatentIn 版本 3.3	
<210>	1	
<211>	1343	
<212>	DNA	
<213>	稻 ( <i>Oryza sativa</i> )	
<400>	1	
	gaacttctct cccatctctt tcctcctcct cctctctgat atgtctacta tctacatgag	60
	ccagctacct gctactctcc ctctaattgga ggggatcag gatcaggggc tctaccacgc	120
	cttccataga gcaaaggacc ctctatctt gttccctttc atgatcgaca gcgccgtcga	180
	gcaccaaggg caaatctatg gagatcaggg cttgaggagg cagcaggttt tgggtgaatc	240
	caatcaacag ttcaatgatc acatgatgat gggcggatca gatgtcttcc tcacaccgtc	300
	tccgttccga ccaaccatcc aaagcatcgg cagcgacatg atccagcgat catcttatga	360
	tccatacgat atcgagagta acaacaagca gcatgccaat ggatcaacca gcaagtggat	420
	gtcgacgccg ccaatgaaga tgaggatcat aaggaagggg gcggcaaccg atcctgaggg	480
	cggggcggtg agaaagccaa ggagaagagc acaagcgcac caggatgaga gccagcaaca	540
	actgcagcaa gctttgggtg tcgtagagtg gtgctcggac tgcaacacca ccaagacccc	600
	cttgtggaga agtggtcctt gtggccccc aatcccttgc aacgcgtgtg gcatcaggca	660
	aaggaaggcg cggcgggcca tggccgctgc tgccaacggc ggagcggcgg tggcgcggc	720
	aaagagcgtg gccgcggcgc cggatgaaca taagccggcg gcgaagaagg agaagagggc	780
	ggcggacgtc gaccggtcgc tgccgttcaa gaaacggtgc aagatggtcg atcacggtgc	840
	tgctgccgtc gctgccacca agcccacggc tgctggagaa gtagtggccg ccgctccgaa	900

ggaccaagat cacgtcatcg tcgtcgggtgg cgagaacgcc gccgccacct ccatgccggc 960  
 acagaaccgg atatccaagg cggcggcgac cgccgctgcc gccgccgct ctccggcggtt 1020  
 cttccacggc ctccctcgcg acgagatcac cgacgccgcc atgctgctca tgaccctatc 1080  
 ctgtggcctc gtccacagct agctagctag ctgatcaaaa ctagctagct actagtagcg 1140  
 ttaatttgat gagggcaaca accagagtac tatgtaccac tactagcaat attttgtgtg 1200  
 tgccttgtga tcttttgttg ttttgtgttg ttgaggagat cactagatca ggatgaagga 1260  
 gagatagtga tcacatgtct aaggacgaaa taaacgagaa caaactcgct agctagctac 1320  
 tagccgggat caggattata ttt 1343

<210> 2  
 <211> 353  
 <212> PRT  
 <213> 稻

<400> 2

Met Ser Thr Ile Tyr Met Ser Gln Leu Pro Ala Thr Leu Pro Leu Met  
1 5 10 15

Glu Gly Asp Gln Asp Gln Gly Leu Tyr Pro Ala Phe His Arg Ala Lys  
20 25 30

Asp Pro Pro Ile Leu Phe Pro Phe Met Ile Asp Ser Ala Val Glu His  
35 40 45

Gln Gly Gln Ile Tyr Gly Asp Gln Gly Leu Arg Arg Gln Gln Val Leu  
50 55 60

Gly Glu Ser Asn Gln Gln Phe Asn Asp His Met Met Met Gly Gly Ser  
65 70 75 80

Asp Val Phe Leu Thr Pro Ser Pro Phe Arg Pro Thr Ile Gln Ser Ile  
85 90 95

Gly Ser Asp Met Ile Gln Arg Ser Ser Tyr Asp Pro Tyr Asp Ile Glu  
100 105 110

Ser Asn Asn Lys Gln His Ala Asn Gly Ser Thr Ser Lys Trp Met Ser



Thr Asp Ala Ala Met Leu Leu Met Thr Leu Ser Cys Gly Leu Val His  
 340 345 350

Ser

<210> 3  
 <211> 22  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 引物

<400> 3  
 cgtcgagcac caagggcaaa tc 22

<210> 4  
 <211> 23  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 引物

<400> 4  
 ggatagggtc atgagcagca tgg 23

<210> 5  
 <211> 24  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 引物

<400> 5  
 aggaggatgc cgctaacaac tttg 24

<210> 6  
 <211> 23  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 引物

<400> 6  
 aaacagcatt ggtgatttca ggc 23

<210> 7  
 <211> 352  
 <212> PRT  
 <213> 拟南芥(*Arabidopsis thaliana*)  
  
 <400> 7  
  
 Met Gly Ser Asn Phe His Tyr Thr Ile Asp Leu Asn Glu Asp Gln Asn  
 1 5 10 15  
  
 His Gln Pro Phe Phe Ala Ser Leu Gly Ser Ser Leu His His His Leu  
 20 25 30  
  
 Gln Gln Gln Gln Gln Gln Gln Gln His Phe His His Gln Ala Ser Ser  
 35 40 45  
  
 Asn Pro Ser Ser Leu Met Ser Pro Ser Leu Ser Tyr Phe Pro Phe Leu  
 50 55 60  
  
 Ile Asn Ser Arg Gln Asp Gln Val Tyr Val Gly Tyr Asn Asn Asn Thr  
 65 70 75 80  
  
 Phe His Asp Val Leu Asp Thr His Ile Ser Gln Pro Leu Glu Thr Lys  
 85 90 95  
  
 Asn Phe Val Ser Asp Gly Gly Ser Ser Ser Ser Asp Gln Met Val Pro  
 100 105 110  
  
 Lys Lys Glu Thr Arg Leu Lys Leu Thr Ile Lys Lys Lys Asp Asn His  
 115 120 125  
  
 Gln Asp Gln Thr Asp Leu Pro Gln Ser Pro Ile Lys Asp Met Thr Gly  
 130 135 140  
  
 Thr Asn Ser Leu Lys Trp Ile Ser Ser Lys Val Arg Leu Met Lys Lys  
 145 150 155 160  
  
 Lys Lys Ala Ile Ile Thr Thr Ser Asp Ser Ser Lys Gln His Thr Asn  
 165 170 175  
  
 Asn Asp Gln Ser Ser Asn Leu Ser Asn Ser Glu Arg Gln Asn Gly Tyr

	180		185		190	
Asn	Asn Asp Cys Val Ile Arg Ile Cys Ser Asp Cys Asn Thr Thr Lys					
	195		200		205	
Thr	Pro Leu Trp Arg Ser Gly Pro Arg Gly Pro Lys Ser Leu Cys Asn					
	210		215		220	
Ala	Cys Gly Ile Arg Gln Arg Lys Ala Arg Arg Ala Ala Met Ala Thr					
	225		230		235	240
Ala	Thr Ala Thr Ala Val Ser Gly Val Ser Pro Pro Val Met Lys Lys					
		245		250		255
Lys	Met Gln Asn Lys Asn Lys Ile Ser Asn Gly Val Tyr Lys Ile Leu					
		260		265		270
Ser	Pro Leu Pro Leu Lys Val Asn Thr Cys Lys Arg Met Ile Thr Leu					
		275		280		285
Glu	Glu Thr Ala Leu Ala Glu Asp Leu Glu Thr Gln Ser Asn Ser Thr					
	290		295		300	
Met	Leu Ser Ser Ser Asp Asn Ile Tyr Phe Asp Asp Leu Ala Leu Leu					
	305		310		315	320
Leu	Ser Lys Ser Ser Ala Tyr Gln Gln Val Phe Pro Gln Asp Glu Lys					
		325		330		335
Glu	Ala Ala Ile Leu Leu Met Ala Leu Ser His Gly Met Val His Gly					
		340		345		350
<210>	8					
<211>	372					
<212>	DNA					
<213>	人工序列					
<220>						
<223>	RNAi					
<400>	8					
tccatgagcc	agctacctgc tactctcct ctaatggagg gggatcagga tcaggggctc					60
taccagcct	tccatagagc aaaggaccct cctatcttgt tccctttcat gatcgacagc					120

gcccgtcgagc accaagggca aatctatgga gatcagggct tgaggaggca gcaggttttg 180  
 ggtgaatcca atcaacagtt caatgatcac atgatgatgg gcggatcaga tgtcttcctc 240  
 acaccgtctc cgttccgacc aaccatccaa agcatcggca gcgacatgat ccagcgatca 300  
 tcttatgatc catacgatat cgagagtaac aacaagcagc atgccaatgg atcaaccagc 360  
 aagtggctgt ag 372

<210> 9  
 <211> 2065  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> RNAi 构建体

<400> 9  
 acaagtttgt acaaaaaagc aggcttccat gagccagcta cctgctactc tccttctaata 60  
 ggaggggggat caggatcagg ggctctacc agccttccat agagcaaagg accctcctat 120  
 cttgttccct ttcattgatcg acagcgccgt cgagcaccaa gggcaaactc atggagatca 180  
 gggcttgagg aggcagcagg ttttgggtga atccaatcaa cagttcaatg atcacatgat 240  
 gatggggcga tcagatgtct tcctcacacc gtctccgttc cgaccaacca tccaaagcat 300  
 cggcagcgac atgatccagc gatcatctta tgatccatac gatatcgaga gtaacaacaa 360  
 gcagcatgcc aatggatcaa ccagcaagtg gctgtagacc cagctttctt gtacaaagtg 420  
 gttggcgccc tgcaaagaaa ggttacactt agcaacacgt accaaaacca ctcaacttgca 480  
 caagaataat tagtcaacag ccatcactaa gcattgcaag actatctctg aacaggaaag 540  
 ccatgctaaa tcaacactaa taacatcaca caaagcatt ggaagatcaa aacataacta 600  
 aaaacagctg tttcatctac acaactgaaa gcatctatgg tttacgaagc agagtgcgag 660  
 tactgattca aaataaatca acctgaacca atatactctg acaatgtttt caaagggata 720  
 aaagaaccag ctttatcaaa tggatttggt gggttttagt aagtatcatt gagataccga 780  
 tggcatacct caaactttgc aaaattataa tggcatgggt ccaaattacc ctttagtatt 840  
 agcaccagtt agatcctaata tcctaaatcc gataggacag agcgaaagat ccctggagat 900  
 atgaagattt ggctacagat taagcagagc caacatgaag ttccgaatat tatgaatccg 960  
 caagcgggga gatcaaagag aagaatacgg aaggctcgca ctccatgaaa gaatccaacc 1020

---

aaaaaccaa agatTTTTct cagttcaaaa aacaaaaacc cttcattttt ggttcgccat	1080
ccaccgacag gcaccaagac attcctcagg aagcaaaaaa gattaagcag aacaagtgat	1140
aagcaagaca cagtatcaac ggactacgag tcgagaaaat cactgaggcg cgattcttac	1200
tgcaccaagt aaaaaaaaaat ttgggggcaa aaagaactct gcaatggggc ggagcaacgt	1260
ggcagcaaaa ctaaaggtcg aggatttgag gttttttgcc ggttttcctc gaaacccccga	1320
atccgctcat agtaaacca ctaaactgca gcagaaacc cctccttggt tcagatttac	1380
cgaaagcagt aaacccaaga acatgtcagc aaaaactcct gcaagattca gctgacgacc	1440
caccaaagaa tcgcaagaaa tgggcgagcc aagaccacga tcaagaggct gcaagcgccg	1500
ctgaatcagc tcaaaaacc aagcaaaaaa tcacacagcc aagatcaatc agcctatggg	1560
taaaagggaa cagtttggca atggaagcaa catacctgcg gccgcgggccc caagcttact	1620
agtgaattcg ccctttggcg ccaaccactt tgtacaagaa agctgggtct acagccactt	1680
gctggttgat ccattggcat gctgcttggt gttactctcg atatcgatg gatcataaga	1740
tgatcgctgg atcatgtcgc tgccgatgct ttggatgggt ggtcggaacg gagacggtgt	1800
gaggaagaca tetgatccgc ccatcatcat gtgatcattg aactgttgat tggattcacc	1860
caaacctgc tgctctctca agccctgac tccatagatt tgcccttggt gctcgacggc	1920
gctgtcgatc atgaaaggg acaagatagg agggtccttt gctctatgga aggctgggta	1980
gagcccctga tctgatccc cctccattag agggagagta gcaggtagct ggctcatgga	2040
agcctgcttt tttgtacaaa cttgt	2065

## OsGATA11全长cDNA (1343bp)

```

1  gaacttctct cccatctctt tcctcctcct cctctctgat atgtctacta tctacatgag
61  ccagctacct gctactctcc ctctaattgga gggggatcag gatcaggggc tctaccagc
121 cttccataga gcaaaggacc ctctatcttt gttccctttc atgatcgaca gcgccgtcga
181 gcaccaaggg caaatctatg gagatcaggg cttgaggagg cagcaggttt tgggtgaatc
241 caatcaacag ttcaatgatc acatgatgat gggcggatca gatgtcttcc tcacaccgtc
301 tccgttccga ccaaccatcc aaagcatcgg cagcgacatg atccagcgat catcttatga
361 tccatacgat atcgagagta acaacaagca gcatgccaat ggatcaacca gcaagtggat
421 gtcgacgccg ccaatgaaga tgaggatcat aaggaagggg gcggcaaccg atcctgaggg
481 cggggcgggtg agaaagccaa ggagaagagc acaagcgcac caggatgaga gccagcaaca
541 actgcagcaa gctttgggtg tcgttagagt gtgctcggac tgcaacacca ccaagacccc
601 cttgtggaga agtggtcctt gtggcccca a gtcccttgc aacgcgtgtg gcatcaggca
661 aaggaaggcg cggcgggcca tggccgctgc tgccaacggc ggagcggcgg tggcgccggc
721 aaagagcgtg gccgcggcgc cggatgaaca taagccggcg gcgaagaagg agaagagggc
781 ggcggaogtc gaccggtcgc tgccgttcaa gaaacggtgc aagatggtcg atcacgttgc
841 tgctgccgtc gctgccacca agcccacggc tgctggagaa gtagtggccg ccgctccgaa
901 ggaccaagat cacgtcatcg tcgtcgggtg cgagaacgcc gccgccacct ccatgccggc
961 acagaacccg atatccaagg cggcggcgac cgccgctgcc gccgccgct ctccggcggt
1021 cttccacggc ctccctcgcg acgagatcac cgacgccgcc atgctgctca tgaccctate
1081 ctgtggcctc gtccacagct agctagctag ctgatcaaaa ctagctagct actagtaccg
1141 ttaatttgat gagggcaaca accagagtac tatgtaccac tactagcaat attttgtgtg
1201 tgccttgtga tcttttgttg ttttgtgtg ttgaggagat cactagatca ggatgaagga
1261 gagatagtga tcacatgtct aaggacgaaa taaacgagaa caaactcgct agctagctac
1321 tagccgggat caggattata ttt

```

图1

OsGATA11蛋白质序列 (353个氨基酸)

MSTIYMSQLPATLPLMEGDQDQGLYPAFHRAKDPPILFPFMIDSAVEHQGQIYGDQGL  
RRQQVLGESNQQFNDHMMGGSDVFLTPSPFRPTIQSIGSDMIQRSSYDPYDIESNNKQH  
ANGSTSKWMSTPPMKMRIIRKGAATDPEGGAVRKPRRRAQAHQDESQQQLQQALGVVRC  
SDCNTTKTPLWRSGPCGPKSLCNACGIRQRKARRAMAAAANGGAAPA KSVAAAPVNNK  
PAAKKEKRAADVDRSLPFFKRCKMVDHVAAAVAATKPTAAGEVVAAAPKDQDHVIVVGGE  
NAAATSMPAQNPI SKAAATAAAAAASPAFFHGLPRDEITDAAMLLMTLSCGLVHS

图 2

At4g26150与其稻直向同源物的比对

```

AT4G26150.1_1 (SEQ ID NO:7)  MGSNFHYTIDLNEDQNHQP  FFA SLGSSLHHHLQ QQQQQQH FHHQASSNPSSLMSPSLSY
OsGATA11_21044_2_1
(SEQ ID NO:2)  ----- MSTI YMS QL PAT L PLMEG DQDQGLYPA FHRAKDPP---- I ----- L
AT4G26150.1_1  FPF L INSRQDVVVG YNNN TFH DVLDTHI SQPLE TKNFV SDGGSSSDQWV KKETRLKL
OsGATA11_210442_1  FPF MIDSAVEHQGI YGDQ GLR RQVLGE SNOQF NDHMMGGSDVFLTPSPF RPTI QSIG
AT4G26150.1_1  TIKK KDNHQDQTDLPQSPI KDMTGTNSLKI SSKVRLMKK KKA IITTS DSSKQHTNNDQS
OsGATA11_210442_1  SDMI QRSSYDPYDIE SNNKQHANGSTS KWS T PPMKMR I I RKG AAT DPEGGAVR KPRRRA
AT4G26150.1_1  SNLSNSE RQNGYNDC VTR ICSDCNTTKT PLWRSGP RGPKSLCNACCI RQKARRA AMAT
OsGATA11_210442_1  CAHQDES QQQIQQALG VVRVCSDCNTTKT PLWRSGF CGPKSLCNACCI RQKARRA MA AA
AT4G26150.1_1  ATATAVSGVS PPVMKKKMQNKNI SNGVYKILSPL PLKVNTCK RMI TLEE TALAEDLE TQ
OsGATA11_210442_1  ANGGAAVAPAKSVAAPV NKPAA KKEKRAADVDR SLPFK KRC KVV DHVA AAVAATKP TA
AT4G26150.1_1  SNSTMLSSDNIYF----- DDL ALLL SKSSAYQQVF PQDE -K
OsGATA11_210442_1  AGEV VAAA PK DQD HVIVVGGENA AAT SMPAQN PISKAAAT AAAA AASPFFHGLP RDE IT
AT4G26150.1_1  EAA I LLMALS HGMVHG
OsGATA11_210442_1  DAAMLLMTLSCGLVHS

```



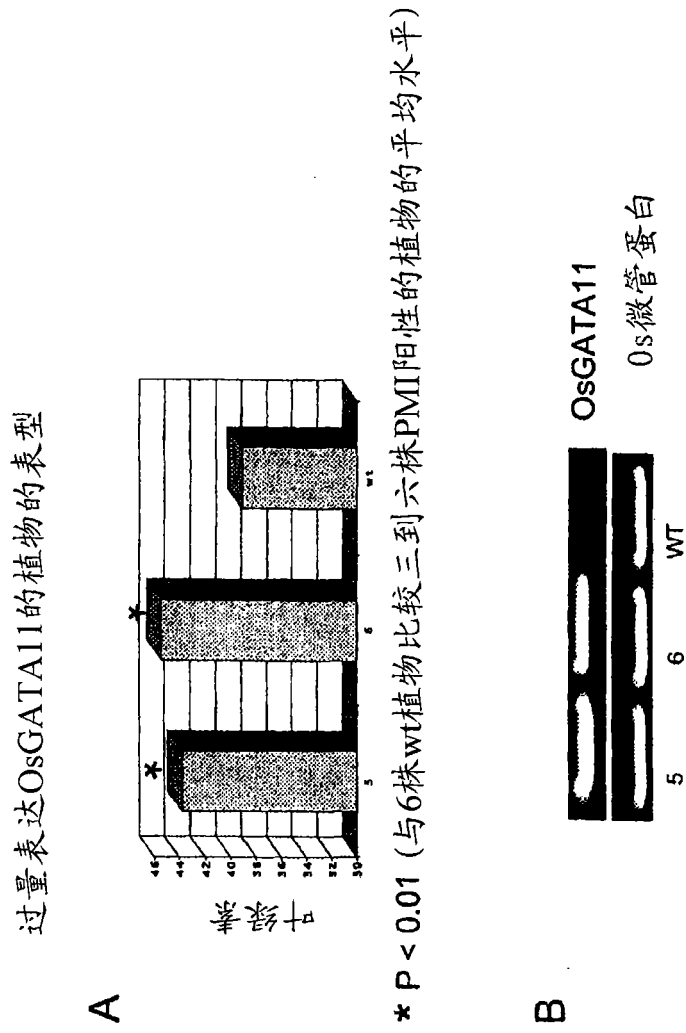


图 4

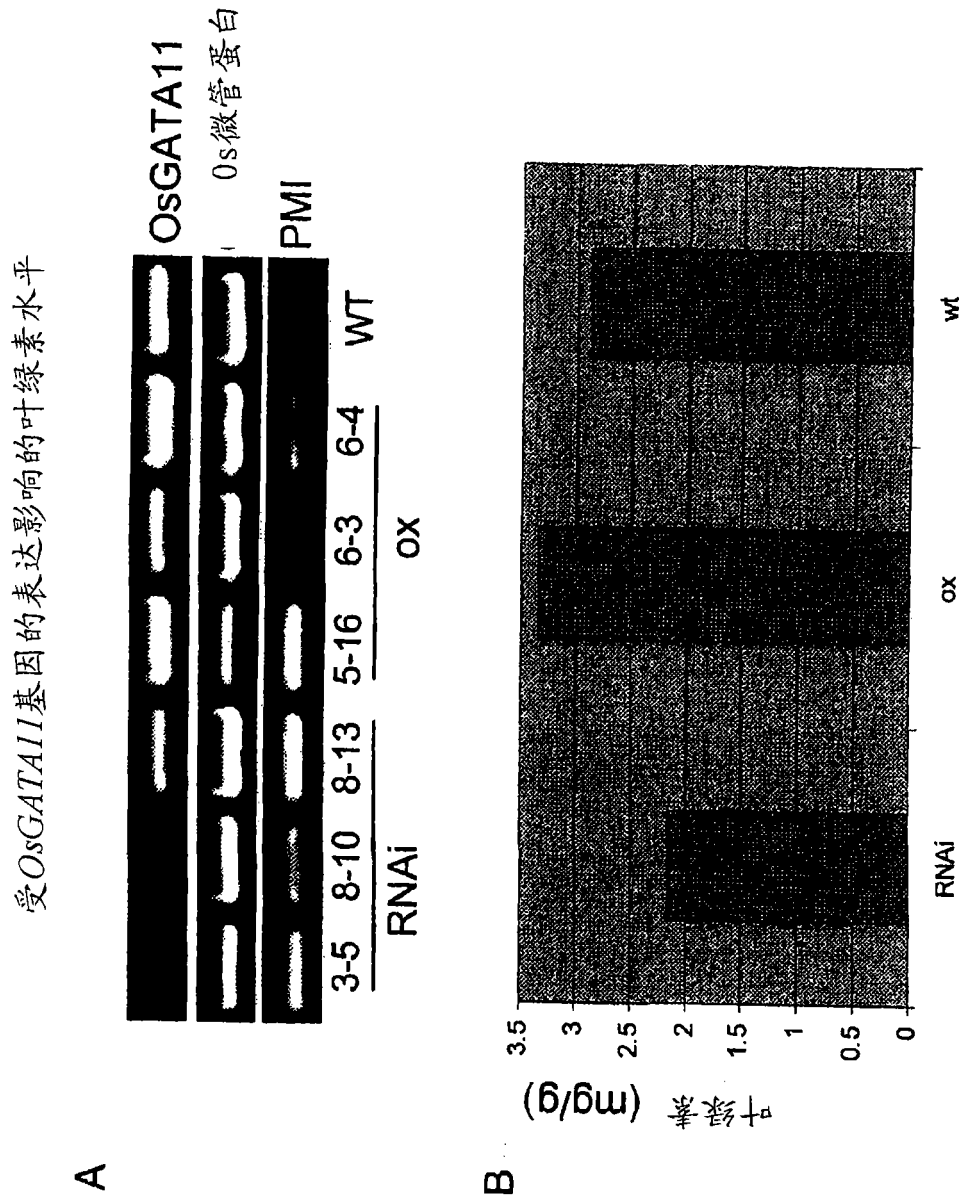
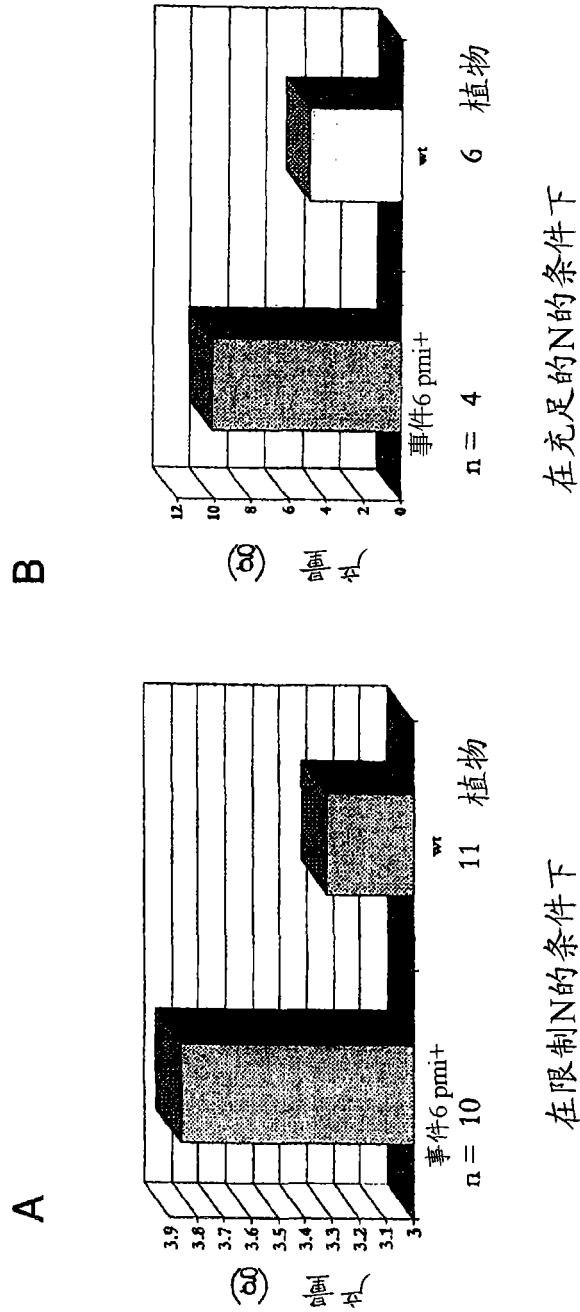


图 5

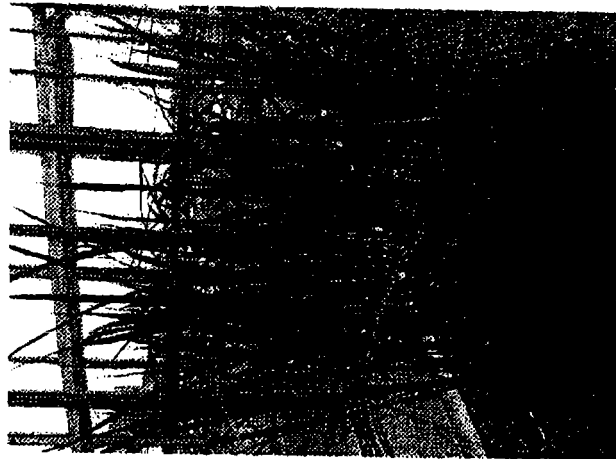
过量表达OsGATA11的植物的种子产量



pmi+植物的平均值具有比wt对照的平均值更高的产量.

图 6

在过量表达OsGATA11的植物中更强的胁迫抗性



OX

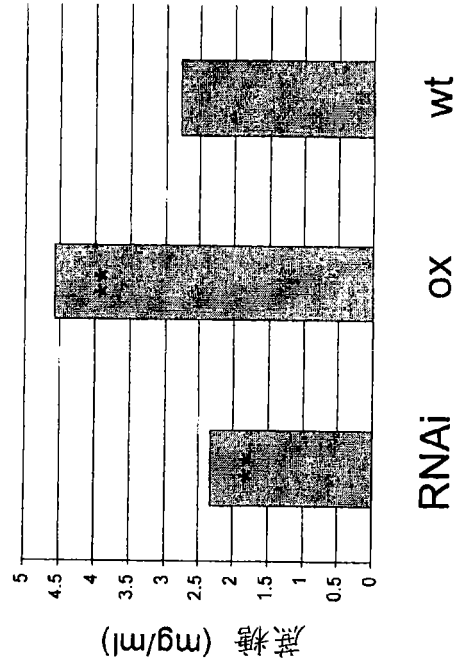


对照

图 7

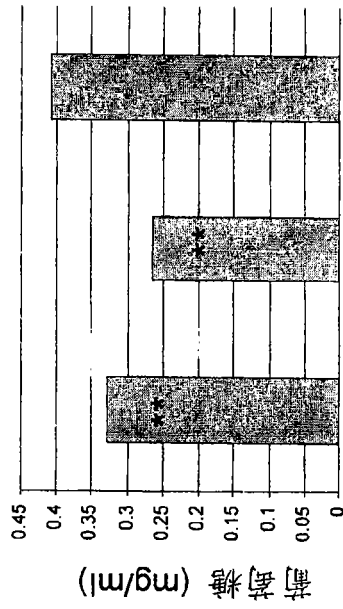
OsGATA11碳状况的生化测定

C

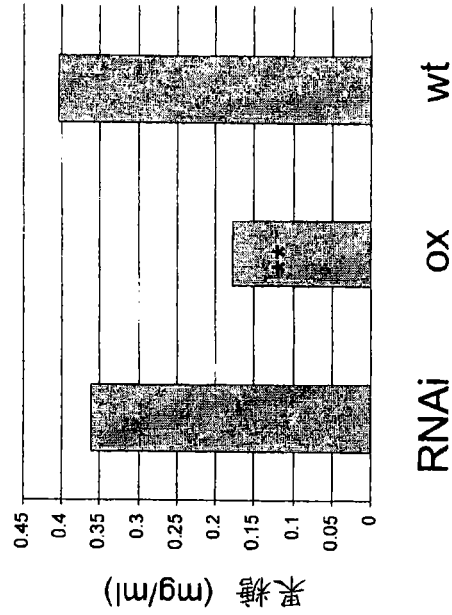


\*\*P<0.01来自7至9株植物的平均值

A



B



图

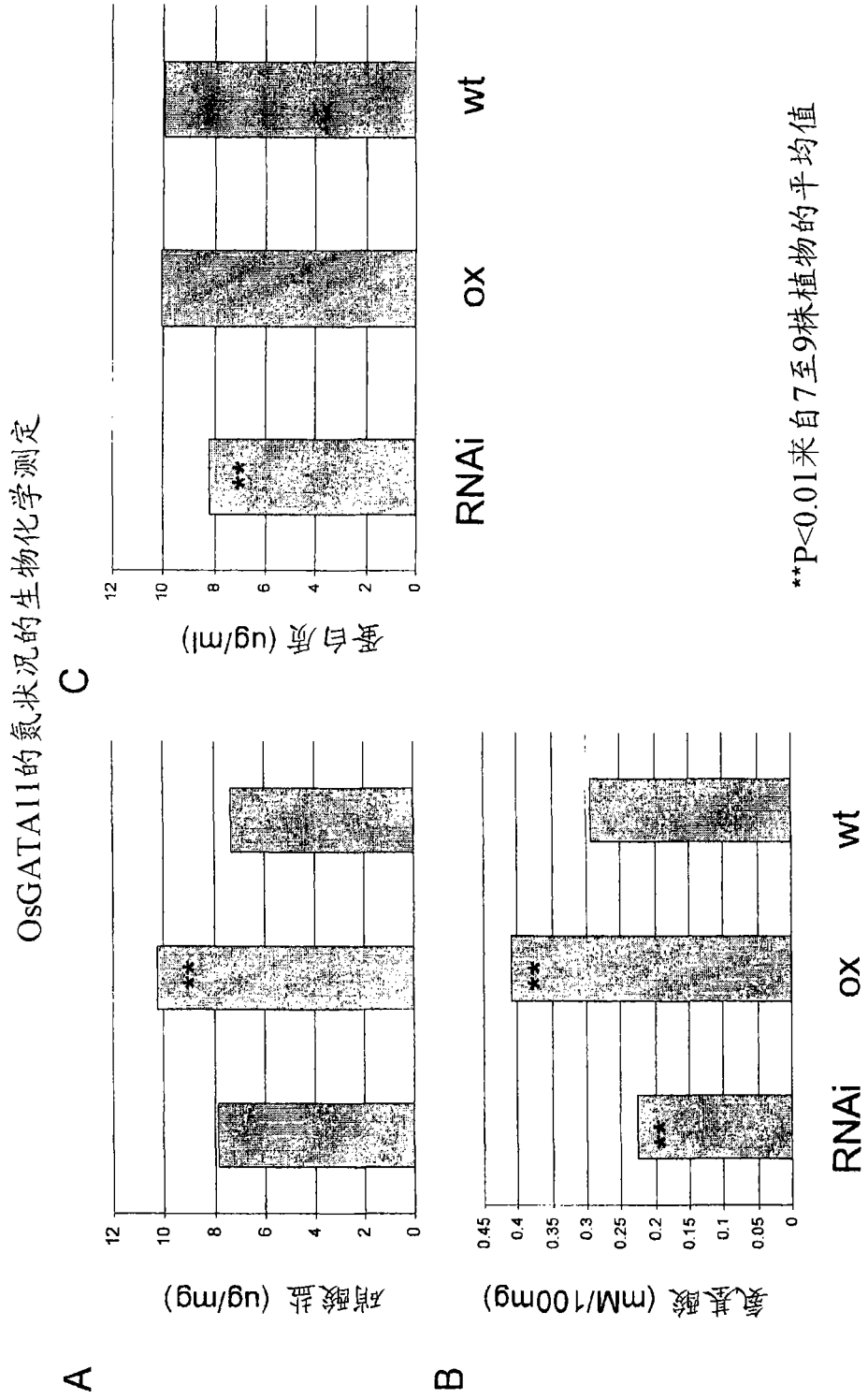


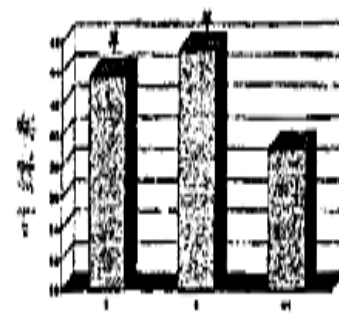
图 9

专利名称(译)	调节碳和氮的基因与蛋白质及其调节		
公开(公告)号	<a href="#">CN101688180A</a>	公开(公告)日	2010-03-31
申请号	CN200880018512.3	申请日	2008-04-16
[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 Y - M·毕		
IPC分类号	C12N5/10 A01H1/00 A01H1/04 A01H3/00 A01H5/00 C07K16/16 C12N15/82 C12N15/29 C12Q1/68 G01N33/53		
CPC分类号	G01N33/6872 G01N33/5097 C12N15/8238		
代理人(译)	史文静		
优先权	2584934 2007-04-17 CA		
外部链接	<a href="#">Espacenet</a> <a href="#">SIPO</a>		

摘要(译)

本发明涉及针对糖和氮的积累所需要的调节氮的GATA转录因子基因以及为调节植物中的特性而对这种基因的表达的调节。本发明的GATA转录因子与调解植物中糖和氮的积累有关。提高这个基因或基本相似的基因的表达能够产生具有改善的氮利用以及提高的产量和提高的胁迫耐受性的植物。

过量表达OsGATA11的植物的表型



\*  $P < 0.01$  (与6株wt植物比较三到六株PMI阳性的植物的平均水平)