

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

C12N 15/82 (2006.01)

C12N 15/29 (2006.01)



# [12] 发明专利申请公布说明书

[21] 申请号 200780046706.X

[43] 公开日 2009年11月11日

[11] 公开号 CN 101578370A

[22] 申请日 2007.12.20

[21] 申请号 200780046706.X

[30] 优先权

[32] 2006.12.20 [33] US [31] 60/871,061

[86] 国际申请 PCT/BR2007/000358 2007.12.20

[87] 国际公布 WO2008/074116 英 2008.6.26

[85] 进入国家阶段日期 2009.6.17

[71] 申请人 阿莱利克斯有限公司

地址 巴西坎皮纳斯

[72] 发明人 P·阿鲁达 I·R·格哈特

[74] 专利代理机构 北京市金杜律师事务所

代理人 陈文平 尚继栋

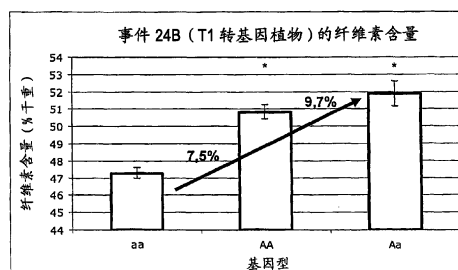
权利要求书 3 页 说明书 27 页 序列表 4 页  
附图 7 页

## [54] 发明名称

编码 C3HC4 家族植物蛋白质的核酸分子和改变植物纤维素和木素含量的方法

## [57] 摘要

本发明公开了用于改变植物组织中纤维素和/或木素含量的多核苷酸、核酸构建体和方法。植物用编码 C3HC4 蛋白的基因进行基因工程化，编码 C3HC4 蛋白的基因在植物维管系统中超表达时导致纤维素含量提高。带有 C3HC4 蛋白基因的植物转化体显示提高的纤维素含量和/或降低的木素含量，该性状被认为改善了阔叶树在制浆和造纸过程中的纤维素提取。



1. 一种分离的核酸序列，其包含选自下组的序列：
  - (a) SEQ ID NO:1 所示的核酸序列，或其互补链；
  - (b) 编码 SEQ ID NO: 2 所示的氨基酸序列的核酸序列；
  - (c) 在严格条件下能够与(a)或(b)的核酸序列杂交的核酸序列，其中所述杂交序列编码 C3HC4 多肽；
  - (d) 作为(a)或(b)的核酸序列的等位基因变体或选择性剪接变体的核酸；和
  - (e) 与(a)或(b)的序列有至少 50%、60%、70%、75%、80%、85%、90%或更高的序列同一性的核酸序列。
2. 一种分离的 C3HC4 蛋白，其选自：
  - (a) SEQ ID NO: 2 所示的多肽；
  - (b) 具有与 SEQ ID NO: 2 所示的氨基酸序列有至少 85%或更高序列同一性的氨基酸序列的多肽；和
  - (c) 如(a)或(b)中定义的多肽的变体。
3. 一种核酸构建体，其包含与一个或多个合适的启动子可操作地连接的分离的 C3HC4 多核苷酸序列，该启动子驱动该 C3HC4 多核苷酸序列的表达。
4. 如权利要求 3 所述的核酸构建体，其中所述启动子是木质部优先的启动子。
5. 如权利要求 4 所述的核酸构建体，其中所述木质部优先的启动子选自 TUB 基因启动子、SuSy 基因启动子、COMT 基因启动子和 C4H 基因启动子。
6. 一种植物细胞，其包含如权利要求 3 所述的核酸构建体。
7. 一种由权利要求 6 的植物细胞产生的转基因植物，其中所述植物与同种的非转基因植物相比，具有改变的纤维素和/或木素含量。
8. 如权利要求 6 所述的植物细胞，其中所述启动子是木质部优先的启动子。

9. 如权利要求 8 所述的植物细胞, 其中所述木质部优先的启动子选自 TUB 基因启动子、SuSy 基因启动子、COMT 基因启动子和 C4H 基因启动子。

10. 如权利要求 7 所述的转基因植物, 其中所述植物是双子叶植物。

11. 如权利要求 7 所述的转基因植物, 其中所述植物是单子叶植物。

12. 如权利要求 7 所述的转基因植物, 其中所述植物是裸子植物。

13. 如权利要求 7 所述的转基因植物, 其中所述植物是阔叶树。

14. 如权利要求 13 所述的转基因植物, 其中所述阔叶树是桉属植物。

15. 如权利要求 13 所述的转基因植物, 其中所述阔叶树是杨属植物。

16. 如权利要求 11 所述的转基因植物, 其中所述针叶树是松属植物。

17. 如权利要求 7 所述的转基因植物的一部分, 其选自叶、茎、花、子房、果实、种子和愈伤组织。

18. 如权利要求 7 所述的转基因植物的后代。

19. 如权利要求 18 所述的后代, 其中所述后代是杂种植物。

20. 一种改变植物中的纤维素和/或木素含量的方法, 包括:

(a) 向分离的植物细胞中引入核酸构建体, 该核酸构建体包含与一个或多个合适的启动子可操作地连接的分离的 C3HC4 多核苷酸序列, 该启动子驱动该 C3HC4 多核苷酸序列的表达; 和

(b) 在促进植物生长的条件下培养所述植物细胞, 其中所述植物超表达 C3HC4 蛋白, 并且与同种的非转基因植物相比, 具有提高的纤维素含量和/或减少的木素含量。

21. 如权利要求 20 所述的方法, 其中所述启动子是木质部优先的启动子。

22. 如权利要求 21 所述的方法，其中所述木质部优先的启动子选自 TUB 基因启动子、SuSy 基因启动子、COMT 基因启动子和 C4H 基因启动子。

## 编码 C3HC4 家族植物蛋白质的核酸分子 和改变植物纤维素和木素含量的方法

### 相关申请的交叉引用

本申请要求 2006 年 12 月 20 日提交的美国临时申请 60/871,061 的优先权，该临时申请的公开内容完整引入作为参考。

### 技术领域

本发明涉及植物生物技术领域。更具体地，本发明涉及通过调节编码 C3HC4 蛋白的基因的表达改变纤维素含量。

### 背景技术

在高等植物的木材形成过程中，大多数在碳水化合物代谢过程中产生的葡萄糖被传递给纤维素用于次生壁沉积。Djerbi 等, *Cellulose* 11: 301-12 (2004)。纤维素是由结晶形成微原纤维的直链  $\beta$ -(1,4)-连接的葡聚糖分子组成的纤维状聚合物。微原纤维提供纤维素特有的柔性强度。纤维素在高等植物中是由大的多聚体质膜结合复合物合成的，这些复合物在微原纤维的末端形成玫瑰花形结构。Somerville, *Ann. Rev. Cell Dev. Biol.* 22: 53-78 (2006)。

纤维素作为纸浆、纤维以及作为合成商业上重要的聚合物的起点是有价值的。可增加纤维素沉积的改变很可能对木素沉积具有抑制效应。Hu 等, *Nature Biotech.* 17: 808-19 (1999)。木本植物中木素含量的减少是人们所希望的，因为纤维素的工业生产和木素的化学去除是昂贵的，并且对环境是一个巨大的挑战。

纤维素的生物合成途径在分子水平上了解很少。来自纤维素合酶(CESA)家族的基因和编码用于 N-聚糖合成和加工的蛋白质的基因已经在大量生物体中分离。Nicol 等, *EMBO J.* 17: 5562-76 (1998)。

一个大量用于研究木材形成，特别是纤维素合成和沉积的实验系统是弯曲树木，以使得在茎的拉紧侧形成应拉木(TW)。Andersson-Gunneras 等, *Plant J.* 45: 144-65 (2006)。在桉属(*Eucalyptus*)和杨属(*Populus*)的种中，应拉木典型地发生在倾斜的茎的上侧，并且使树重新定向其轴。应拉木的主要特征是在其腔中具有特别厚的胶质次生层(G层)的木质部纤维。这种G层几乎只含有具有高结晶度的纤维素微原纤维。Déjardin 等, *Plant Biol.* 6: 55-64 (2004)。由于应拉木富含纤维素但是缺乏木素和半纤维素，它可以用来检测和分析参与控制向木素、纤维素和半纤维素的碳流的基因。

Andersson-Gunneras 等(2006)鉴定了在TW中高表达的与细胞壁形成有关的基因，例如参与碳水化合物代谢和细胞骨架形成的基因，以及管家基因和两种在欧洲山杨(L.)北美颤杨(Michx)杂交种中具有未知功能的基因。C3HC4-型锌指(环指)蛋白在TW中显示差异表达，但是在该研究或其它研究中没有数据暗示其在纤维素生物合成中的作用。

锌指结构域是结合一个或多个锌原子的相对较小的蛋白质基序。它们最初被鉴定为光滑爪蟾(*Xenopus laevis*)转录因子TFIIIA中的DNA-结合基序，但是现在被认为结合DNA、RNA、蛋白质和/或脂类物质。环指是一种特殊类型的锌指，有40-60个残基，结合两个锌原子，且可能参与介导蛋白质-蛋白质相互作用。存在两种不同的变体：C3HC4-型和C3H2C3-型。C3HC4-型环指基序在许多细胞和病毒蛋白质中被发现，其中一些已证明在体内和体外都具有泛素E3连接酶活性。Laity 等, *Curr. Opin. Struct. Biol.* 11:39-46 (2001)。

鉴于传统林木育种的困难，例如由于较长的世代周期导致的缓慢发展，和产生具有理想性状的植物的困难，基因技术的发展可以明显缩短产生植物新品种所需的时间，并且允许在特定树种中更准确地靶向被认为是林业和制浆工业所希望的性状。

## 发明内容

一方面，本发明提供一种分离的核酸序列，其包含选自下组的序列：(a) SEQ ID NO: 1 所示的核酸序列或其互补链；(b) 编码 SEQ ID NO: 2 所示的氨基酸序列的核酸序列；(c) 在严格条件下能够与(a)或(b)的核酸序列杂交的核酸序列，其中所述杂交序列编码 C3HC4 多肽；(d) 作为(a)或(b)的核酸序列的等位基因变体或选择性剪接变体的核酸；和(e) 与(a)或(b)的序列有至少 50%、60%、70%、75%、80%、85%、90%或更高的序列同一性的核酸序列。

另一方面，本发明提供一种分离的 C3HC4 蛋白，其选自：(a) SEQ ID NO: 2 所示的多肽；(b) 具有与 SEQ ID NO: 2 所示的氨基酸序列有至少 85%或更高序列同一性的氨基酸序列的多肽；和(c) 如(a)或(b)中定义的多肽的变体。

另一方面，本发明提供一种核酸构建体，其包含与一个或多个合适的启动子可操作地连接的分离的 C3HC4 多核苷酸序列，该启动子驱动该 C3HC4 多核苷酸序列的表达。在一个实施方案中，植物细胞包含该核酸构建体。在进一步的实施方案中，由该植物细胞产生转基因植物，并且该植物与同种的非转基因植物相比，具有改变的纤维素和/或木素含量。在再进一步的实施方案中，植物是双子叶植物、单子叶植物、裸子植物或阔叶树。进一步的实施方案包括该转基因植物的后代，包括杂种植物。

在另一方面，本发明提供一种改变植物中的纤维素和/或木素含量的方法，包括：(a) 向分离的植物细胞中引入核酸构建体，该核酸构建体包含与一个或多个合适的启动子可操作地连接的分离的 C3HC4 多核苷酸序列，该启动子驱动该 C3HC4 多核苷酸序列的表达；和(b) 在促进植物生长的条件下培养所述植物细胞，其中所述植物超表达 C3HC4 蛋白，并且与同种的非转基因植物相比，具有提高的纤维素含量和/或降低的木素含量。

#### 附图说明

图 1 显示一组桉树组织中 C3HC4 cDNA 的表达概况。由从美洲

黑杨(*Populus deltoides*)木质部分离的 mRNA 克隆直向同源基因。

图 2 示意性地显示了本发明的植物表达质粒载体 pALELLYX-C3HC4, 其包含驱动本发明的 C3HC4 核苷酸序列表达的 形成层/木质部优先的启动子。

图 3 显示几种用 pALELLYX-C3HC4 转化的转基因株系和相应的对照非转基因植物的纤维素含量。星号表示统计学显著较高的平均纤维素含量值。

图 4 显示用 pALELLYX-C3HC4 转化的 T1 转基因植物(6B 株系)的两种基因型的纤维素含量。星号表示统计学显著较高的平均纤维素含量值( $P < 0.05$ , t-检验)。

图 5 显示用本发明的植物表达质粒载体 pALELLYX-C3HC4 转化的 T1 转基因植物(6B 株系)的两种基因型的木素含量。星号表示统计学显著较低的平均木素含量值( $P < 0.05$ , t-检验)。

图 6 显示用本发明的植物表达质粒载体 pALELLYX-C3HC4 转化的 T1 转基因植物(24B 株系)的三种基因型的纤维素含量。星号表示统计学显著较高的平均纤维素含量值( $P < 0.05$ , t-检验)。

图 7 显示用本发明的植物表达质粒载体 pALELLYX-C3HC4 转化的 T1 转基因植物(25B 株系)的三种基因型的纤维素含量。星号表示统计学显著较高的平均纤维素含量值( $P < 0.05$ , t-检验)。

## 发明详述

本发明涉及遗传操作的植物, 其特征在于提高的 C3HC4 蛋白表达。在这点上, 本发明集中于遗传操作的植物, 该植物超表达包含 C3HC4 基因的核酸分子, 由此调节所述遗传操作的植物的纤维素含量。提高基因的转录速率可以导致蛋白质产物增多, 从而提高该基因所参与的代谢过程的速率。

在这点上, 本发明人已经确定, 超表达 C3HC4 基因的遗传操作的植物显示提高的纤维素含量。因此, 本发明人确定 C3HC4 蛋白直接或间接地控制与纤维素合成有关的基因和/或蛋白质。



因此，本发明的申请包括但不限于通过提高木本树木中的纤维素含量改进造纸过程中的纤维素纤维生产，以及通过提高棉花纤维中的纤维素含量为纺织品生产改进纤维素纤维的提取。另外，增加的纤维素沉积可能对木素沉积具有抑制效应。由木本树木工业生产纤维素需要在制浆工艺中化学去除木素，这要使用大量的浓缩化学试剂。为了生产高质量的纸张，需要通过额外的漂白步骤进一步除去残余的木素，该漂白步骤涉及使用非常有害的物质。这整个过程花费多，并且对环境是一个巨大的挑战。由于这个原因，预计减少木本植物(典型地为树)中的木素含量可减少这些非常昂贵的提取过程的化学试剂和能量的需要，也应当能够减少排出物的量，排出物是一种主要的潜在环境污染物，不仅难以处理，而且处理的费用很高。Campbell等, *Plant Physiol.* 110: 3-13 (1996)。因此，纤维素生物合成的基因工程可以提供提高转基因植物的纤维素质量和产量、同时减少木素含量的策略。

本说明书中的技术术语符合生物化学、分子生物学和农学中通常的用法。这种用法和这些技术术语在以下文献中有详细说明：MOLECULAR CLONING: A LABORATORY MANUAL (3rd ed.), vol. 1-3, Cold Spring Harbor Laboratory Press (2001); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Greene Publishing Associates and Wiley-Interscience (1988), 定期更新；SHORT PROTOCOLS IN MOLECULAR BIOLOGY: A COMPENDIUM OF METHODS FROM CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (5<sup>th</sup> ed.), vol. 1-2, John Wiley & Sons, Inc. (2002); GENOME ANALYSIS: A LABORATORY MANUAL, vol. 1-2, Cold Spring Harbor Laboratory Press (1997)。如此处所述的合适的植物生物学技术进一步在诸如以下的方法学文献中详细说明：METHODS IN PLANT MOLECULAR BIOLOGY: A LABORATORY COURSE MANUAL, Cold Spring Harbor Laboratory Press (1995)。例如, Innis等, PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS, Academic Press (1990) 以及 Dieffenbach 和 Dveksler, PCR PRIMER: A

LABORATORY MANUAL (2<sup>nd</sup> ed.), Cold Spring Harbor Laboratory Press (2003)中描述了各种使用 PCR 的方法。PCR 引物对可以通过已知的技术, 例如使用用于该目的的计算机程序, 如 Primer, 0.5 版, 1991 (Whitehead Institute for Biomedical Research, Cambridge, MA), 由已知的序列得到。例如, Beaucage 和 Caruthers, *Tetra. Letts.* 22: 1859-62 (1981)以及 Matteucci 和 Caruthers, *J. Am. Chem. Soc.* 103:3185(1981)中描述了示例性的化学合成核酸的方法。

限制性酶消化、磷酸化、连接和转化如 Sambrook 等, MOLECULAR CLONING: A LABORATORY MANUAL (2<sup>nd</sup> ed.), Cold Spring Harbor Laboratory Press (1989)所述进行。除非另外说明, 所有用于细菌细胞生长和维持的试剂和材料都来自 Aldrich Chemicals (Milwaukee, Wis.)、DIFCO Laboratories (Detroit, Mich.)、Invitrogen (Gaithersburg, Md.)或 Sigma Chemical Company (St. Louis, Mo.)。

术语“编码”是指一个基因通过转录和翻译机制给一个细胞提供信息的过程, 从该细胞中一系列氨基酸可以装配为特定的氨基酸序列, 从而产生活性酶。由于遗传密码的简并性, DNA 序列中的某些碱基变化不改变蛋白质的氨基酸序列。因此可以认为, 可以预期涉及基本上不影响蛋白质的功能特性的编码 C3HC4 蛋白的 DNA 序列的修饰。

在本说明书中, “表达”表示由基因编码的蛋白质产物或多肽的产生。可选择地或者另外, “表达”表示为了产生多肽, 编码 DNA 分子(如结构基因)所经历的细胞内过程的组合, 包括转录和翻译。“超表达”是指特定基因序列的表达, 其中转基因生物体中 mRNA 或多肽的产量超过了非转基因生物体中的产量水平。

术语“异源核酸”是指已通过人工努力导入细胞(或细胞的祖先)中的核酸、DNA 或 RNA。这种外源核酸可以是在所导入的细胞中天然发现的序列的拷贝, 或其片段。

相反, 术语“内源核酸”是指存在于将要基因工程化的植物或生物体中的核酸、基因、多核苷酸、DNA、RNA、mRNA 或 cDNA 分

子。内源序列对于将要基因工程化的植物或生物体来说是“天然的”，即，固有的。

术语“同源序列”是指由于共同的祖先和序列保守性而相似的多核苷酸或多肽序列。

对于本发明，“旁系同源物(paralog)”是通过基因复制产生的同源物。它们代表在生物体内复制然后趋异的来源于共同祖先基因的基因。“直向同源物(ortholog)”是通过物种形成产生的同源物。它们代表由于相关的生物体趋异而趋异的来源于共同祖先的基因。参见 Brinkman 和 Leipe, *In: BIOINFORMATICS, A PRACTICAL GUIDE TO THE ANALYSIS OF GENES AND PROTEINS* 323-58, Wiley-Interscience (2001)。

术语“功能同源物”是指由于共同的祖先和序列保守性而相似，并且在催化、细胞或生物体水平上具有相同或相似功能的多核苷酸或多肽序列。

### C3HC4 序列

生物体代谢和发育所需的许多生物学过程由基因家族控制。C3HC4 蛋白就是这样，它属于一个包括许多基因的基因家族。C3HC4-结构域蛋白质属于所谓的“锌指”蛋白质家族，其特征在于“环指”结构域，该结构域包括 8 个由共有基序 C3HC4 形成的金属配体。环指结构域通常以独特的横拉条(cross brace)排列结合两个锌离子，并且基本上可以被认为是一个蛋白质相互作用结构域。Jackson 等, *Trends Cell Biol.* 10: 429-39 (2000)。

环指蛋白与从转录和翻译调节到发育和定向蛋白质水解的许多不同的生物学过程有关。因此，其它 C3HC4 家族成员也将提高纤维素含量。SEQ ID NO:1 中显示的一个示例性的 C3HC4 基因是从白杨中分离的。

预计来自任一编码环指基因的生物体的、编码具有类似于 C3HC4 基因翻译产物的结构和生物学性质的蛋白质的任一基因，对

纤维素含量的影响将与实施例中所述发明人对转基因植物所证明的影响相同。这些基因可以通过序列对比进行鉴定和功能注释。有见识的分子生物学家可以借助于常规方法，例如使用合适的杂交探针筛查 cDNA 或基因组文库，或者检索公共数据库如 NCBI 的 Genbank，容易地鉴定与 C3HC4 序列有关的序列。也可以借助于简并寡核苷酸，利用已知的基于 PCR 的技术分离同源序列。

也可以使用允许利用各种已知技术通过序列对比鉴定基因的计算程序。例如，Innis 等, PCR Protocols: A GUIDE TO METHODS AND APPLICATIONS, Academic Press (1990) 和 GENOME ANALYSIS: A LABORATORY MANUAL, 第 1 卷和第 2 卷, Cold Spring Harbor Laboratory Press (1997) 描述了这方面的技术的例子。

在与本发明相关的研究中，本发明人在种间杂种巨尾桉杂交种 (*E. grandis* x *E. urophylla*) 植物中诱导了 TW。因此鉴定了大量在 TW 和正常木中差异表达的基因。在应拉木中表达发生改变的基因中，确定有许多编码在桉树 TW 中高表达的 C3HC4-型锌指家族的成员。

由于 C3HC4 蛋白家族的成员在 TW (主要由高度结晶的纤维素组成的组织) 中高表达，且 C3HC4 成员似乎与维管分生组织和次生生长的控制有关，因此可以预期用包含编码 C3HC4 家族成员的分子的核酸构建体对木本树木进行基因转化将会改变维管模式及纤维素合成和沉积。因此，当 C3HC4 基因超表达时将发生纤维素合成和沉积的增加。相反，当 C3HC4 基因下调时，本发明人预期纤维素合成将减少。由于纤维素合成和沉积的改变通常导致木素合成和沉积的改变，参见 Hu 等, *Nature Biotech* 17: 808-19 (1999)，本发明人同样理解，提高纤维素合成和沉积将降低木本树木的木素含量。如果纤维素合成减少，则出现相反的情况。

在本发明的上下文中，一个序列可以通过如上所述的方法鉴定，并且由此功能注释为属于 C3HC4 家族。在本说明书中，短语“C3HC4 多核苷酸序列”和“C3HC4 核酸序列”表示任何编码 C3HC4 多肽的核酸、基因、多核苷酸、DNA、RNA、mRNA、cDNA 分子，当其超表

达时导致植物的纤维素含量提高和/或木素含量降低。短语“C3HC4 多核苷酸序列”和“C3HC4 核酸序列”也包括任何具有在严格条件下能够与此处所述的任一序列杂交的核苷酸序列并且编码 C3HC4 多肽的核酸分子，当其超表达时导致植物的纤维素含量提高和/或木素含量降低。该短语也表示与 SEQ ID NO:1 交叉杂交的序列，优选地具有与 SEQ ID NO: 1 至少 40%、更优选至少 60%、甚至更优选至少 80%、最优选至少 90%的同源性或同一性的序列。

本发明的核苷酸序列也可以编码与 SEQ ID NO: 2 所示的预测基因产物同源的蛋白质。

短语“C3HC4 多核苷酸序列”和“C3HC4 核酸序列”也表示那些由 SEQ ID NO: 1 的片段或变体代表的序列，它们具有与 SEQ ID NO: 1 至少 50%、优选至少 60%、更优选至少 70%、甚至更优选至少 80%、最优选至少 90%的同一性，并且编码 C3HC4 多肽，当其超表达时导致植物的纤维素含量提高和/或木素含量降低。基于同样的理由，本发明的核苷酸序列包括那些编码包含 SEQ ID NO: 2 的氨基酸序列或与 SEQ ID NO: 2 至少 50%、优选至少 60%、更优选至少 70%、甚至更优选至少 80%、最优选至少 90%相同的氨基酸序列的多肽的序列，其超表达导致植物的纤维素含量提高和/或木素含量降低。

短语“严格条件”在此表示本领域中熟悉的参数。当单链多核苷酸基于多种良好表征的物理化学力如氢键、溶剂排斥和碱基堆积而结合时，它们发生杂交。杂交的严格性反映了所涉及的核酸的序列同一性的程度，因此严格性越高，两条多核苷酸链越相似。严格性受多种因素的影响，包括杂交和洗涤溶液的温度、盐浓度和组成、有机和无机添加剂、溶剂等以及孵育(和次数)。本领域普通技术人员通过改变杂交反应和洗涤过程中的温度、杂交反应和洗涤过程中的盐浓度等，能够容易地选择这些条件。

对于具有 100 个以上互补残基的互补核酸在 Southern 或 Northern 印迹中在滤膜上的杂交，“严格”杂交条件的例子是在限定的离子强度和 pH 下，比特定序列的热解链温度( $T_m$ )低大约 5°C 至 20°C 的温

度。T<sub>m</sub>是在限定的离子强度和 pH 下，50%的靶序列与完全匹配的探针杂交时的温度。在严格条件下杂交的核酸分子一般基于整个 cDNA 或选择的部分与探针杂交。更优选地，此处的“严格条件”是指本领域熟悉的参数，例如在 3.5xSSC、1xDenhardt's 溶液、25mM 磷酸钠缓冲液(pH 7.0)、0.5% SDS 和 2mM EDTA 中 65°C 杂交 18 小时，然后于 65°C 在 2xSSC、0.1% SDS 中洗涤滤膜 4 次，各 20 分钟，且为了获得更高的严格性，在 0.5xSSC、0.1% SDS 或 0.3xSSC 和 0.1% SDS 中进行最多 20 分钟的最后一次洗涤，为了获得甚至更高的严格性，在 0.1xSSC、0.1% SDS 中洗涤。可以用其它条件代替，只要严格性程度基本上等同于此处使用 0.5xSSC 的最后洗涤提供的程度。

需要指出，本说明书中的短语“C3HC4 核酸序列”是指具有在严格条件下能够与此处公开的序列杂交的核苷酸序列的任何核酸分子，其编码与具有此处 SEQ ID NO: 2 公开的氨基酸序列的蛋白质等同的多肽。该短语也包括与 SEQ ID NO:1 交叉杂交的序列，优选地具有与 SEQ ID NO: 1 至少 55%、优选地至少 65%、更优选地至少 75%、甚至更优选地至少 85%、最优选地至少 90%的同源性或同一性。本发明的核苷酸序列可以编码与 SEQ ID NO: 2 的预测基因产物同源的蛋白质。

进一步的实施方案包括包含具有一个或多个缺失、置换、插入或添加的碱基的上述任何碱基序列，并且编码与 SEQ ID NO: 2 编码的蛋白质同源的多肽的任何核酸分子。这些核酸分子包括 SEQ ID NO: 1 的等位基因变体和选择性剪接变体。

相应地，术语“变体”是与特定基因或蛋白质的标准的或者给定的核苷酸或氨基酸序列偏离的核苷酸或氨基酸序列。变体可以具有“保守性”改变，其中置换的氨基酸具有相似的结构或化学性质，例如，用异亮氨酸替换亮氨酸。变体可以具有“非保守性”改变，例如，用色氨酸替换甘氨酸。类似的微小变化也可以包括氨基酸缺失或插入，或缺失和插入。确定哪些氨基酸残基可以被置换、插入或删除的指导可以使用本领域公知的计算机程序来获得，例如 Vector NTI

Suite (InforMax, MD)软件。“变体”也可以指“改组的基因”，例如，如美国专利 6,506,603、6,132,970、6,165,793 和 6,117,679 所述。

此处提到的“具有一个或多个缺失、置换、插入或添加的碱基的碱基序列”是本领域普通技术人员公知的，甚至当通常具有生理活性的蛋白质的氨基酸序列具有一个或多个置换、缺失、插入或添加的氨基酸时，也保留其生理活性。具有这些修饰并且编码 C3HC4 蛋白的核苷酸序列包括在本发明的范围内。例如，可以删除 poly A 尾或 5'-或 3'-端非翻译区，并且可以删除碱基到删除氨基酸的程度。碱基也可以被置换，只要不导致移码。也可以“添加”碱基到添加氨基酸的程度。但是，重要的是这些修饰不导致 C3HC4 蛋白功能的丧失。这些修饰的核酸可以通过以下方法获得，例如，修饰本发明的碱基序列，使得特定位点的氨基酸通过定点诱变被置换、删除、插入或添加。参见 Zoller 和 Smith, *Nucleic Acid Res.* 10: 6487-500 (1982)。

应当理解，氨基酸和核酸序列可以包括额外的残基，例如额外的 N-或 C-末端氨基酸或 5'或 3'序列。只要得到的序列编码保持相同或相当的生物蛋白质活性的多肽，这些添加就是合适的。

本发明提供编码 C3HC4 蛋白的核苷酸序列。该序列可以来源于 cDNA，例如美洲黑杨 cDNA，或者来源于基因组 DNA。一个示例性的 cDNA 克隆如以上 SEQ ID NO: 1 所示，其编码 C3HC4 蛋白。根据本发明的一个方面，通过控制 C3HC4 蛋白的表达，改变植物组织（例如木质部或棉籽的纤维细胞）中的纤维素含量。因此，植物细胞或整个植物（例如）用 C3HC4 蛋白编码序列基因工程化，该编码序列例如来源于美洲黑杨，在纤维细胞中优先表达，并且导致纤维素合成和沉积增加。

另外，本发明提供包含选自下组的核苷酸序列的核酸分子：(a) SEQ ID No: 1 或其部分或其互补序列；(b) 在相当于 0.1X SSC 至 1.0X SSC、0.1 % SDS、50-65°C 的洗涤严格性下与所述(a)的核苷酸序列杂交的核苷酸序列；(c) 编码具有与(a)的核苷酸序列编码的蛋白质相同的氨基酸序列的蛋白质，但是根据遗传密码简并性而简并的核苷

酸序列；和(d) 编码与(b)的核苷酸序列所编码的相同的氨基酸序列，但是根据遗传密码简并性而简并的核苷酸序列。

本发明的一个进一步的特征是本发明的核酸分子编码的蛋白质和多肽，其例子包括但不限于具有由 SEQ ID NO: 2 组成的氨基酸序列的多肽。优选地，本发明的多肽具有含有与上述序列至少 60%相同的区域的氨基酸序列。相对于上述序列具有大于 70%的同一性是优选的，而大于 80%、90%或甚至 95%的同一性是最优选的。

本发明的核酸分子可以“干净地”使用，或者优选地在表达载体构建体中使用，用于导入细胞，例如植物细胞中。可以采用熟练技术人员公知的标准分子生物学技术。

### 核酸构建体

重组核酸构建体可以用标准技术制备。例如，用于转录的核苷酸序列可以通过用限制性内切酶处理含有所述序列的载体以切下适当的片段而获得。用于转录的核苷酸序列也可以通过退火和连接合成寡核苷酸或者通过在聚合酶链反应(PCR)中使用合成寡核苷酸在每一末端得到合适的限制酶切位点而产生。然后将核苷酸序列克隆到含有诸如上游启动子和下游终止子序列的适当调节元件的载体中。植物转化载体一般包括一个或多个处于 5'和 3'调节序列的转录控制下的克隆的植物编码序列(基因组或 cDNA)以及选择性标记。这样的植物转化载体一般也含有启动子、转录起始位点、RNA 加工信号(如剪接信号序列)、转录终止位点和/或多腺苷酸化信号。也可以存在增强子和靶向序列。

可用于表达 C3HC4 蛋白序列的合适的组成型植物启动子包括但不限于花椰菜花叶病毒(CaMV) 35S 启动子、玉米和杨树聚泛素启动子，它们在大多数植物组织中引起组成型的、高水平的表达(参见，例如，WO 2007/00611，美国专利 5,510,474; Odell 等, *Nature*, 1985, 313: 810-812); 胭脂氨酸合酶启动子(An 等, 1988, *Plant Physiol.* 88: 547-552); 来自玄参花叶病毒的 FMV 启动子(美国专利 5,378,619);



和章鱼氨酸合酶启动子(Fromm 等, 1989, *Plant Cell* 1: 977-984)。

载体也可以含有终止序列, 终止序列位于本发明的核酸分子的下流, 使得 mRNA 的转录终止, 并且添加 polyA 序列。终止子的例子有花椰菜花叶病毒(CaMV) 35S 终止子和胭脂氨酸合酶基因(NOS) 终止子。

表达载体也可以含有选择标记, 通过该选择标记可以在培养中鉴别出被转化的细胞。标记可以结合异源核酸分子, 即与启动子可操作地连接的基因。此处使用的“标记”是指编码一种性状或一种表型的基因, 该性状或表型允许选择或筛选含有该标记的植物或细胞。在植物中, 例如, 标记基因可编码抗生素或除草剂抗性。这允许从未被转化或转染的细胞中选择出转化细胞。

合适的选择性标记的例子包括腺苷脱氨酶、二氢叶酸还原酶、潮霉素 B 磷酸转移酶、胸苷激酶、黄嘌呤-鸟嘌呤磷酸核糖基转移酶、草甘膦和草丁膦抗性和氨基-糖苷 3'-O-磷酸转移酶(卡那霉素、新霉素和 G418 抗性)。这些标记可以包括对 G418、潮霉素、博来霉素、卡那霉素和庆大霉素的抗性。该构建体也可以含有选择性标记基因 *Bar*, *Bar* 赋予对除草剂膦丝菌素类似物如草丁膦铵的抗性。Thompson 等, *EMBO J.* 6: 2519-23 (1987)。其它合适的选择标记也是已知的。

也可以使用可视标记, 例如绿色荧光蛋白(GFP)。基于细胞分裂的控制来鉴别或选择转化植物的方法也已描述。参见 John 和 Van Mellaert, WO 2000/052168 和 Fabijansk 等, WO 2001/059086。

也可以包含细菌或病毒来源的复制序列, 以允许载体在细菌或噬菌体宿主中克隆。优选地, 使用广宿主范围的原核生物复制起点。可以包含用于细菌的选择性标记, 以允许筛选带有所需构建体的细菌细胞。合适的原核生物选择性标记也包括对抗生素如卡那霉素或四环素的抗性。

本领域中公知, 载体中也可以存在编码另外的功能的其它核酸序列。例如, 当土壤杆菌是宿主时, 可以包含 T-DNA 序列, 以利于

后续向植物染色体中的转移和整合。

根据本发明的进一步的方面，提供了包含如上所述的处于在植物中工作的转录起始区的控制下的 C3HC4 DNA 序列的核酸构建体，使得该构建体可以在植物细胞中产生 RNA。优选地，该转录起始区是器官或组织特异性的植物启动子的一部分，如 WO 2005/096805 公开申请中所述的任一种。更优选地，当组织特异性启动子与 C3HC4 DNA 序列可操作地连接时，该启动子确保在特定细胞类型、组织或器官中转录，使得纤维素合成可以特异性地定向，而不影响其它植物功能。

本发明的转基因植物的特征可以是提高的纤维素含量和/或降低的木素含量。基因工程植物中提高的纤维素含量优选地通过发生纤维素沉积的植物组织中 C3HC4 表达的提高来实现。因此，在一个优选实施方案中，本发明的转基因植物含有一种核酸构建体，该核酸构建体包含与编码 C3HC4 蛋白的基因可操作地连接的形成层/木质部优先的启动子，如以上引用的'805 公开国际申请中所述的那些，从而导致 C3HC4 基因在植物维管系统中的表达提高，这又实现了这些组织中纤维素合成和沉积的增加，而不影响其它植物功能。

如上所述，植物部分的纤维素含量和相关特性可以通过用本发明的核酸构建体进行基因工程来改变。本发明也提供了含有这些构建体或用这些构建体基因工程化的植物细胞、由这些细胞产生的 C3HC4 基因表达发生改变的植物，和这些植物的种子。

本发明的核酸构建体可以包含最小长度的碱基序列，以产生 mRNA，并因此产生保留 C3HC4 功能的多肽。为了方便起见，通常发现使用长度为大约 100 至大约 1000 个碱基的序列是适当的，但是碱基序列的长度没有理论上限。这些构建体的制备在下面更详细地描述。

本发明的分离的核酸分子可以整合到核酸构建体内，使得它们与启动子可操作地连接。优选地，启动子是已知在植物细胞中起作用的启动子，更优选地在诸如根、苗、叶、木质部等特定植物器官

或组织的细胞中起作用。本发明的核酸分子可以与组成型或诱导型启动子可操作地连接。或者，本发明的核酸分子可以与将下游基因的表达优选地或特异性地定向于植物器官或组织(如木质部和形成层)的启动子可操作地连接。

另外，维管系统特异性的、木质部特异性的或木质部优先的启动子可以用于促进本发明的核酸分子特别在维管组织、特别是木质部组织中的表达。使用组成型启动子通常影响在植物的所有部分中的蛋白质水平和功能，而使用组织优先的启动子允许将改变的基因表达定向于特定的植物部分，导致更易控制的表型。因此，在应用本发明时，发现可以方便地使用将在木质部发育过程中引起表达的启动子，由此本发明的蛋白质将只在此处公开的应用需要其作用的器官或组织或细胞类型中过量产生。可以使用的维管组织特异性的、木质部特异性的、维管组织优先的和木质部优先的启动子包括但不限于上述'805 公布国际申请中描述的木质部优先的香豆酸-4-羟化酶(C4H)基因启动子、木质部优先的微管蛋白(TUB)基因启动子和木质部优先的脂质转移蛋白(LTP)基因启动子。所选择的特定启动子应当能够引起足够的表达，以使本发明的蛋白质超表达，以改变木质部的大小或改变植物木质部的化学组成，或这些效应的组合。

尽管基因表达速率主要由启动子调节，但是也可以通过鉴定和使用增强子序列，例如基因的内含子部分，来实现表达的提高，增强子序列以独立的方式定向提高位置靠近的基因的表达水平。对于植物，在基因构建体中启动子和基因编码序列之间的位置处包含某些内含子将导致 mRNA 和蛋白质积累增加。已知可提高植物中表达的内含子在玉米基因中已经鉴定，例如，hsp70、tubA1、Adh1、Sh1、UbH (Brown 和 Santino, 美国专利 5,424,412 和 5,859,347; Jeon 等, 2000, *Plant Physiol.* 123: 1005-1014; Callis 等, 1987, *Genes Dev.* 1: 1183-1200; Vasil 等, 1989, *Plant Physiol.* 91: 1575-1579), 并且在双子叶植物基因中已经鉴定，例如来自矮牵牛的 rbcS (Dean 等, 1989, *Plant Cell* 1: 201-208)、来自马铃薯的 ST-LS1 (Leon 等, 1991, *Plant*

*Physiol* 95: 968-972)和 *UBQ3* (Norris 等,1993, *Plant Mol. Biol.* 21: 895-906) 和来自拟南芥(*Arabidopsis thaliana*)的 *PAT1* (Rose 和 Last, 1997, *Plant J.* 11: 455-464)。

另外, 重组表达载体包含在植物细胞中起作用的启动子、作为上述核苷酸序列的同源物的核酸分子, 该核酸分子编码其氨基酸序列含有与上述 SEQ ID NO: 2 所示的序列至少 60%相同的区域的多肽。更优选地, 该核酸分子编码一种多肽, 该多肽的氨基酸序列含有与上述序列至少 70%、80%或者甚至 90%相同的区域。

本发明的构建体可以用来利用任何合适的技术基因工程化任何植物。单子叶和双子叶被子植物或裸子植物细胞都可以用本领域公知的各种方法基因工程化。Klein 等, *Biotechnology* 4: 583-90 (1993); Bechtold 等, *C. R. Acad. Sci. Paris* 316: 1194-99 (1993); Bent 等, *Mol Gen. Genet.* 204: 383-96 (1986); Paszowski 等, *EMBO J.* 3: 2717-2722 (1984); Sagi 等, *Plant Cell Rep.* 13: 262-66 (1994)。

### 用于基因工程的植物

本发明一般地涉及表达编码此处公开的新型多肽组成的基因或基因区段的转基因植物。此处使用的术语“转基因植物”是指已经整合有核酸序列的植物, 该核酸序列包括但不限于正常情况下也许不存在的基因、正常情况下不转录为 RNA 或翻译为蛋白质(“表达”)的核酸序列, 或者希望导入植物中的其它任何基因或核酸序列, 例如正常情况下可能存在于植物中但是希望基因工程化或改变其表达的基因。可以设想, 在某些情况下本发明的转基因植物的基因组将通过稳定地导入转基因而增加。但是在另外一些情况下, 导入的基因或序列将代替内源序列。可以导入的优选基因包括但不限于来自美洲黑杨的 C3HC4 核酸序列。

可以按照本发明工程化的植物包括但不限于如下的树木: 桉属的种(白桉(*E. alba*)、白花桉(*E. albens*), 杏仁桉(*E. amygdalina*)、*E. aromaphloia*、圆叶桉(*E. baileyana*)、*E. balladoniensis*、双脉桉(*E.*

*bicostata*)、葡萄桉(*E. botryoides*)、短蕊桉(*E. brachyandra*)、褐桉(*E. brassiana*)、短柱桉(*E. brevistylis*)、布罗韦桉(*E. brockwayi*)、钝盖赤桉(*E. camaldulensis*)、*E. ceracea*、大花序桉(*E. cloeziana*)、聚果桉(*E. coccifera*)、异心叶桉(*E. cordata*)、角蕾桉(*E. cornuta*)、*E. corticosa*、常桉(*E. crebra*)、*E. croajingolensis*、*E. curtisii*、山桉(*E. dalrympleana*)、剥桉(*E. deglupta*)、大桉(*E. delegatensis*)、*E. delicata*、卡瑞桉(*E. diversicolor*)、*E. diversifolia*、*E. dives*、*E. dolichocarpa*、*E. dundasii*、邓恩桉(*E. dunnii*)、滨河白桉(*E. elata*)、*E. erythrocorys*、*E. erythrophloia*、*E. eudesmoides*、*E. falcata*、*E. gamophylla*、*E. glaucina*、蓝桉(*E. globulus*)、双脉蓝桉(*E. globulus subsp. bicostata*)、蓝桉原种(*E. globulus subsp. globulus*)、*E. gongylocarpa*、巨桉(*E. grandis*)、巨尾桉杂交种(*E. grandis x urophylla*)、*E. guilfoylei*、西达桉(*E. gunnii*)、*E. hallii*、*E. houseana*、*E. jacksonii*、*E. lansdowneana*、*E. latisinensis*、*E. leucophloia*、白藓叶桉(*E. leucoxylon*)、*E. lockyeri*、*E. lucasii*、直杆蓝桉(*E. maidenii*)、边缘桉(*E. marginata*)、*E. megacarpa*、蜜味桉(*E. melliodora*)、*E. michaeliana*、小帽桉(*E. microcorys*)、小套桉(*E. microtheca*)、缪勒纤皮桉(*E. muelleriana*)、亮果桉(*E. nitens*)、*E. nitida*、斜叶桉(*E. oblique*)、*E. obtusiflora*、西方桉(*E. occidentalis*)、*E. optima*、卵叶桉(*E. ovata*)、*E. pachyphylla*、雪桉(*E. pauciflora*)、粗皮桉(*E. pellita*)、穿叶桉(*E. perriniana*)、*E. petiolaris*、弹丸桉(*E. pilularis*)、*E. piperita*、阔叶桉(*E. platyphylla*)、多花桉(*E. polyanthemos*)、*E. populnea*、*E. preissiana*、*E. pseudoglobulus*、*E. pulchella*、*E. radiata*、*E. radiata subsp. radiata*、王桉(*E. regnans*)、*E. risdonii*、*E. robertsonii*、*E. rodwayi*、河红桉(*E. rubida*)、赤褐桉(*E. rubiginosa*)、柳叶桉(*E. saligna*)、红皮桉(*E. salmonophloia*)、*E. scoparia*、银顶白蜡桉(*E. sieberi*)、*E. spathulata*、*E. staeri*、*E. stoatei*、*E. tenuipes*、*E. tenuiramis*、细叶桉(*E. tereticornis*)、*E. tetragona*、达尔文纤皮桉(*E. tetradonta*)、*E. tindaliae*、*E. torquata*、*E. umbra*、尾叶桉(*E. urophylla*)、*E. vernicosa*、多枝桉(*E. viminalis*)、*E. wandoo*、韦塔桉(*E. wetarensis*)、*E. willisii*、

*E. willisii* subsp. *falciformis*、*E. willisii* subsp. *willisii*、*E. woodwardii*), 杨属的种(银白杨(*P. alba*)、银白杨大齿白杨杂交种(*P. alba* x *P. grandidentata*)、银白杨欧洲山杨杂交种(*P. alba* x *P. tremula*)、银白杨欧洲山杨杂交种(变种)(*P. alba* x *P. tremula* var. *glandulosa*)、银白杨美洲山杨杂交种(*P. alba* x *P. tremuloides*)、香脂杨(*P. balsamifera*)、毛果香脂杨(*P. balsamifera* subsp. *trichocarpa*)、毛果香脂杨美洲黑杨杂交种(*P. balsamifera* subsp. *trichocarpa* x *P. deltoids*)、缘毛杨(*P. ciliate*)、美洲黑杨(*P. deltoids*)、胡杨(*P. euphratica*)、欧美黑杨(*P. euramericana*)、杂交颤杨(*P. kitakamiensis*)、大叶杨(*P. lasiocarpa*)、苦杨(*P. laurifolia*)、马氏杨(*P. maximowiczii*)、毛果马氏杨香脂杨杂交种(*P. maximowiczii* x *P. balsamifera* subsp. *trichocarpa*)、黑杨(*P. nigra*)、西氏杨大齿白杨杂交种(*P. sieboldii* x *P. grandidentata*)、甜杨(*P. suaveolens*)、川杨(*P. szechuanica*)、毛白杨(*P. tomentosa*)、欧洲山杨(*P. tremula*)、欧洲山杨北美颤杨杂交种(*P. tremula* x *P. tremuloides*)、北美颤杨(*P. tremuloides*)、椅杨(*P. wilsonii*)、加拿大杨(*P. Canadensis*)、滇杨(*P. yunnanensis*)、松类如火炬松(*Pinus taeda*)、湿地松(*Pinus elliotii*)、美国黄松(*Pinus ponderosa*)、小干松(*Pinus contorta*)和辐射松(*Pinus radiata*)、花旗松(*Pseudotsuga menziesii*)、美国西部铁杉(加拿大铁杉(*Tsuga Canadensis*))、北美云杉(*Picea glauca*)、红杉(*Sequoia sempervirens*)、真杉(true fir)如银杉(*Abies amabilis*)和香脂冷杉(*Abies balsamea*)、及雪松类如大侧柏(北美乔柏(*Thuja plicata*))和黄扁柏(*Chamaecyparis nootkatensis*)。

本发明还涉及产纤维植物的修饰, 如棉花(棉属的种(*Gossipium* spp.))、亚麻(*Linum usitatissimum*)、小荨麻(异株荨麻(*Urtica dioica*))、蛇麻(*Humulus lupulus*)、椴树类(欧洲小叶椴(*Tilia cordata*)、欧洲椴(*T. x. europaea*)和阔叶椴(*T. platyphyllus*))、鹰爪豆(*Spartium junceum*)、苧麻(*Boehmeria nivea*)、楮(*Broussonetya papyrifera*)、新西兰麻(*Phormium tenax*)、罗布麻(磁麻(*Apocynum cannabinum*))、鳶尾属的种(道氏鳶尾(*I. douglasiana*)、*I. macrosiphon* 和 *I. purdyi*)、乳草类(马

利筋属的种(*Asclepia species*))、菠萝、香蕉和其它植物。还包括饲料作物,如紫花苜蓿、黑麦草、羊茅和三叶草。

在本说明书中,“植物”概括地表示任何可以被遗传操作的含有纤维素的植物材料,包括但不限于分化的或未分化的植物细胞、原生质体、完整的植物、植物组织和植物器官,以及植物的任何部分,例如叶、茎、根、芽、块茎、果实、根茎等。

在本说明书中,“转基因植物”是指已整合有核酸序列的植物,该核酸序列包括但不限于正常情况下不存在于宿主植物基因组中的基因,正常情况下不转录为RNA或翻译为蛋白质的核酸序列,或其它任何希望导入野生型植物中的基因或核酸序列,例如正常情况下可能存在于野生型植物中但是希望基因工程化或希望改变其表达的基因。“转基因植物”类别包括原代转化体和例如,通过标准基因渐渗或另外一种育种程序而在其谱系中包括转化体的植物。相反,没有进行遗传操作的植物是对照植物,被称为“非转基因”植物。非转基因植物可以是其基因组未通过导入包含本发明的多核苷酸序列或其片段的构建体而修饰的植物。也可以是由培养的细胞或组织再生的,而没有通过导入包含本发明的多核苷酸序列的构建体事先修饰的植物,或者可以包括由转基因植物自花受精产生的纯合隐性后代(即,不具有任何转基因拷贝)。此处使用的“杂种植物”是指由两个亲本植物之间杂交获得的植物或其部分,其中一个亲本是本发明的基因工程植物。这可以通过(例如)有性繁殖自然发生,或者可以通过(例如)体外核融合人工发生。

本发明的转基因植物含有如此处所述的在植物中有效启动子的控制下表达的核酸序列,使得该植物的特征为,例如,降低的木素含量和纤维素含量提高。

### 植物基因工程方法

可以使用合适的工程技术将本发明的构建体导入任何植物细胞中。单子叶和双子叶被子植物或裸子植物细胞都可以用本领域所知

的各种方法基因工程化。例如，参见 Klein 等, 1993, *Biotechnology* 4: 583-590; Bechtold 等, 1993, *C. R. Acad. Sci. Paris* 316: 1194-1199; Koncz 和 Schell, 1986, *Mol. Gen. Genet.* 204: 383-396; Paszkowski 等, 1984, *EMBO J.* 3: 2717-2722; Sagi 等, 1994, *Plant Cell Rep.* 13: 262-266。

例如，可以按照 Nagel 等, 1990, *Microbiol Lett* 67: 325 所述，使用土壤杆菌属的种，如根癌土壤杆菌(*A. tumefaciens*)和发根土壤杆菌(*A. rhizogenes*)。简言之，土壤杆菌可以用植物表达载体转化，例如通过电穿孔转化，之后通过例如公知的叶盘法(leaf-disk method)将土壤杆菌引入植物细胞中。

其它实现这一目的的方法包括但不限于：用根瘤菌属(*Rhizobium*)、中华根瘤菌属(*Sinorhizobium*)或中慢生根瘤菌属(*Mesorhizobium*)转化(Broothaerts 等, 2005, *Nature* 433: 629-633)、电穿孔、粒子枪轰击、磷酸钙沉淀和聚乙二醇融合、向萌芽的花粉粒中转移、直接转化(Lorz 等, 1985, *Mol. Genet.* 199: 179-82)，及其它已知的方法。如果使用选择性标记，如卡那霉素抗性，则更易于确定哪些细胞被成功转化。

已知以上所述的土壤杆菌转化方法可用于转化双子叶植物。另外，de la Pena 等, 1987, *Nature* 325: 274-76; Rhodes 等, 1988, *Science* 240: 204-207; 和 Shimamoto 等, 1989, *Nature* 328: 274-76 公开了使用土壤杆菌转化谷类单子叶植物。另外参见 Bechtold 和 Pelletier, 1998, *Methods Mol. Biol.* 82: 259-66，其证明了真空渗入在土壤杆菌介导的转化中的应用。

可以检测特定细胞中蛋白质、多肽或核酸分子的存在，以确定，例如，细胞是否已被成功地基因工程化。进行这种试验的能力是公知的，在此不需赘述。

### 定量纤维素/木素含量

在此用来描述本发明植物的短语“提高的纤维素含量”是指与野生型植物中的纤维素量相比，本发明植物中纤维素量的定量增加。



纤维素的定量增加可以通过几种方法测定，例如在磨碎的茎木中多糖酸水解后基于总糖进行定量。Chiang 和 Sarkanen, *Wood Sci. Technol.* 17: 217-26 (1983); Davis, *J. Wood Chem. Technol* 18: 235-52 (1988)。

本发明的工程植物中的纤维素含量可以升高野生型植物纤维素含量的大约 30%至大约 50%、优选大约 25%至大约 45%，甚至更优选大约 20%至大约 40%的水平。本发明植物的最优选的实施方案具有为野生型纤维素含量的大约 10%至大约 15%的纤维素含量。

短语“减少的木素含量”和“降低的木素含量”在此用来描述本发明的植物的一个方面，分别是指植物中的木素量比野生型或非转化植物中的木素量定量减少。木素的定量减少可以通过 Klason 木素测定(Kirk 等, *Method in Enzymol.* 161: 87-101 (1988))和乙酰溴木素测定(Iiyama 等, *Wood Sci. Technol.* 22: 271-80 (1988))说明的常规方法测定。

本发明的工程植物中的木素含量可以降至野生型植物的木素含量的大约 5%至大约 90%，优选地大约 10%至大约 75%，更优选地大约 15%至大约 65%干重的水平。本发明的植物的最优选的实施方案具有野生型木素含量的大约 20%至大约 60%的木素含量。

以下提供了用于获得美洲黑杨 C3HC4 基因的方法的实施例，并且下面给出了使用土壤杆菌导入靶基因从而产生植物转化体的技术。这些只是实施例，而不是对本发明的限制。

### 实施例 1

在应拉木、应力木和正常木中优先表达的基因的表达概况

如 Huang 和 Madan, *Genome Res.* 9: 868-77 (1999)(引入作为参考)所述，利用 CAP3 程序，将来自巨尾桉杂交种(*Eucalyptus grandis* x *Eucalyptus urophylla*)的表达序列标签(ESTs)聚类。从代表以下组织的文库中获得一组 53,522 个 EST：来自野外生长的高 6.5m 的桉树(巨尾桉杂交种)的应拉木、应力木和正常木。在这样产生的簇组中搜索

至少由 90%的来自代表应拉木组织的文库的 EST 读数组成的簇。另外, 在簇组中搜索由至少 3 个来自应拉木组织的 EST 读数和优选地少于两个来自其它文库的读数组成的簇。

这样选择的一个簇, 由 14 个来自应拉木 cDNA 文库的 EST 读数和 0 个来自其它文库(应力木和正常木)的读数组成, 代表 C3HC4 蛋白家族的成员(图 1)。

然后使用 Blast-X 算法, 采用  $\leq 1e-5$  的截止 e-值(参见 Altschul 等, *Nucleic Acids Res.* 25: 3389-402 (1997)), 将使用这些参数选择的簇与来自专门杨树数据库的序列进行比对, 该数据库由从 JGI 毛果杨 (*Populus trichocarpa*) v1.0 数据库 (<http://genome.jgi-psf.org/Poptrl/Poptrl.home.html>)获得的序列组成。比较结果存储在杨树序列本地数据库中。通过这一过程, 取得编码与选自桉树文库的序列直向同源的 C3HC4 蛋白的簇。该簇中最长读数的序列如本文中 SEQ ID NO: 1 所示, 其编码本文中 SEQ ID NO: 2 所示的多肽。

## 实施例 2

### 从美洲黑杨中分离 C3HC4 DNA 序列

#### (a) 从美洲黑杨形成层/木质部制备 RNA 和合成 cDNA

从一年的美洲黑杨树的枝插(stem cutting)上除去树皮。将包括形成层、木质部和木髓的茎的内部切成小片, 在液氮中冷冻, 用溴化十六烷基三甲基铵(CTAB)提取法(Aldrich 和 Cullis, *Plant Mol. Biol. Report.*, 11:128-141 (1993))进行 RNA 提取。在 RT-PCR 实验中使用 cDNA 库, 其中使用分离的总 RNA 作为模板, 且使用 Superscript II 逆转录酶(Invitrogen)和 oligo (dT)引物合成第一链 cDNA。如下所述, 使用基因特异性引物, 通过随后的聚合酶反应获得双链 cDNA。

#### (b) PCR 引物的设计和 RT-PCR 反应

合成基于 SEQ ID NO: 1 的寡聚体作为 PCR 引物, 包括编码该多肽的主要 ORF 的第一个 ATG 密码子周围的区域或终止密码子周围

的区域，以扩增该主要 ORF 的完整编码区。引物序列为：

C3HC4NDE: 长度: 30

catatgaata cgcggtaccc ctttccaatg (SEQ ID NO: 3)

C3HC4XBA: 长度: 31

tctagactat ctctccaatc cttgtttaca g (SEQ ID NO: 4)

在 PCR 反应中，使用(a)中获得的 cDNA 库作为模板，使用 SEQ ID NOs: 5、6、7 和 8 的引物。PCR 包括 94°C 1 分钟、51°C 1 分钟和 72°C 2 分钟的 40 个循环，随后是 72°C 7 分钟的一个额外的延伸步骤。PCR 产物通过 1.0%琼脂糖凝胶电泳分离，然后用溴化乙锭染色电泳凝胶，并在 UV 透射仪上检测扩增的条带。验证检测到的扩增条带，并用刀片从琼脂糖凝胶上切下。将凝胶片转移到 1.5mL 微型管中，分离 DNA 片段，并用 GFX PCR 净化和凝胶条带纯化试剂盒(Amersham)纯化。将回收的 DNA 片段亚克隆到可商购的克隆载体中，转化大肠杆菌，然后用来制备质粒 DNA，然后采用标准方法通过二脱氧法(Messing, *Methods in Enzymol* 101, 20-78 (1983))测序。得到核苷酸序列 SEQ ID NO. 1，其编码此处公开的 SEQ ID NO: 2 的多肽。

### 实施例 3

转基因本氏烟草(*Nicotiana benthamiana*)植物的制备

将以上实施例 2 中获得的来自美洲黑杨的核酸分子导入植物宿主中以产生转基因烟草植物。

将实施例 2 中获得的从美洲黑杨中分离的核酸分子克隆到表达载体中木质部优先的香豆酸-4-羟化酶基因(C4H)启动子的下游(图 2)。得到的表达构建体在大肠杆菌中扩增，然后通过化学转化法转化到根癌土壤杆菌 LBA4404 株中。

土壤杆菌介导的本氏烟草的转化利用 Horsch 等, *Science* 227: 1229 (1985)的叶盘法实现。简言之，LBA4404 土壤杆菌株生长过夜，直到达到对数中期生长。用无菌水 1:10 稀释培养物，并与来自无菌

生长的幼小本氏烟草植物的叶盘共培养 20 分钟。这些叶盘在 Murashige-Skoog 培养基中在暗处培养。48 小时后，将叶盘颠倒置于添加有 0.4 mg/L 吲哚乙酸(IAA)、2 mg/L 苄基氨基嘌呤(hOBAP)、1mg/L Finale 和 500 mg/L 羧苄西林的相同生长培养基的新鲜平板上。当形成幼苗时，从叶盘上取下，置于只添加 1 mg/L Finale 的新鲜培养基中。让转基因杂合的本氏烟草原代转化体的幼苗在 Murashige 和 Skoog 培养基上生根，然后转移到土壤中，并在温室中生长。按照此处提供的说明，条件(~50 pM/m<sup>2</sup>/sec 光, 27°C)足以鉴定那些显示改变的木质部结构和/或木质部化学组成或这些效应的组合的转基因植物。

#### 实施例 4

##### 外源基因插入宿主植物基因组的 PCR 验证

利用 PCR 验证基因构建体在转基因植物的基因组中的整合。合成一对引物，用于从选择性标记基因 *Bar* 扩增 400bp DNA 序列。另外，合成另一对引物，用于扩增内源本氏烟草查耳酮合酶(*CHS*)基因。这些引物组在公开的国际申请 WO 2006/096951 中均有描述。

Bar 35: 长度: 20

tctacatga gccagaacg

Bar 36: 长度: 23

aattcggggg atctggattt tag

CHS 150: 长度: 24

gccagcccaa atccaagatt actc

CHS 151: 长度: 23

aatgttagcc caacttcacg gag

利用 *Bar* 引物 PCR 扩增含有本发明核酸分子的表达构建体的 T-DNA 部分的一部分，即从本氏烟草转化体的基因组 DNA 扩增。

PCR 反应混合物含有使用溴化十六烷基三甲基铵(CTAB)提取法 (Aldrich 和 Cullis, *Plant Mol. Biol. Report* 11: 128-41 (1993))制备的转

化植物的 100 ng 基因组 DNA、0.2  $\mu$ M 用于 *Bar* 基因的各引物、0.2  $\mu$ M 用于内源 *CHS* 控制基因的各引物、100  $\mu$ M 各脱氧核糖核苷三磷酸、1xPCR 缓冲液和 2.5 单位 AmpliTaq DNA 聚合酶 (Applied Biosystems), 总体积为 50  $\mu$ L。循环参数如下: 94°C 1 分钟, 57°C 1 分钟, 72°C 1 分钟, 40 个循环, 加上 72°C 5 分钟延伸。PCR 产物在 1%琼脂糖凝胶上电泳。

### 实施例 5

#### 转基因植物中转基因表达水平的确定

利用半定量 RT-PCR 检测转基因植物的茎组织中美洲黑杨 C3HC4 转录物的积累。利用 CTAB 法(Aldrich 和 Cullis, *Plant Mol. Biol. Report.* 11:128-141 (1993))从 4 个月大的转基因烟草 T0 植物的茎切段中分离总 RNA。使用 Superscript II RNase H-RT (Invitrogen, USA)由 500 ng 总 RNA 合成 cDNA。上述引物与用于编码查耳酮合酶(*CHS*)的组成型基因的引物一起使用, 后者作为标准化每个样品中使用的总 RNA 的量的内部控制。使用第一链 cDNA 的 12.5 倍稀释液在下列条件下进行 PCR: 94°C 3 分钟, 和 94°C 1 分钟、51°C 1 分钟、72°C 1 分 30 秒的 27 个循环。

上述公开内容和实施例描述了本发明的各个特征, 基本上需要分离和克隆可用于产生基因工程植物的、编码 C3HC4 蛋白家族成员的核酸分子。已经用这种分离的核酸分子转化或转染的重组植物可显示纤维素和/或木素含量的定量改变。

### 实施例 6

#### 转基因植物的组织化学分析

将转基因烟草和对照非转基因植物的茎切段, 并在 4%低聚甲醛中固定 24 小时。然后在切片机(Leica RM2255)上将固定的组织切片, 之后用星蓝/番红染色。组织学染色的切片在 Leica DM1L 倒置显微镜下采用亮视野和暗视野照明进行观察。

### 实施例 7

在维管组织中超表达 C3HC4 的转基因植物中纤维素含量的提高

收集用包含在木质部优先的美洲黑杨 C4H 启动子控制下的美洲黑杨 C3HC4 基因的构建体转化的烟草转基因事件(event)和非转基因对照植物的主茎，风干两周。将干燥的茎切片，使用 30 目的筛在切碎机上制成粉末。然后对茎粉末样品进行化学分析，以确定纤维素和木素含量。简言之，在酸水解这些从茎中提取的多糖后基于总糖确定纤维素和半纤维素含量。碾碎的茎于 45°C 真空干燥，并用 H<sub>2</sub>SO<sub>4</sub> 水解。高 pH 阴离子交换层析后，基于水解物的组成对葡聚糖和其它多糖(半纤维素)进行定量。Chiang 和 Sarkanen (1983)和 Davis (1988)，同上。按照实施例 5 详述的程序得知表达转基因的三个 C3HC4 转基因事件，显示统计学显著的纤维素含量升高(图 3)。与对照非转基因植物的 50.00%相比，转基因事件 6B 显示 54.09%的纤维素，表明纤维素含量显著提高了 8.18% (P<0.05, t-检验)。与对照植物的 50.00%相比，转基因事件 24B 显示 53.90%的纤维素，表明纤维素含量提高了 7.80% (图 3; P<0.05, t-检验)。与对照非转基因植物的 50.00%相比，转基因事件 4B 显示 53.26%的纤维素含量，表明纤维素含量提高了 6.52% (图 3; P<0.05, t-检验)。

生长成熟后，T0 事件自花受精产生 T1 株。这里显示了已分析其 T1 后代的三个事件的相关结果。

对来自事件 6B 的 T1 群体的分析表明，C3HC4 基因的纯合显性条件是致死性的，因为在分离群体中没有检测到纯合显性植物。植物发育可能受到影响。尽管如此，与纯合隐性植物相比，C3HC4 基因半合植物显示纤维素含量显著提高 8.4% (P<0.05, t-检验) (图 4)。与纯合隐性植物相比，它们也显示木素含量降低 18% (P<0.05, t-检验) (图 5)。

在事件 24B 和 25B 的分离群体中，可以鉴别出纯合显性植物。但是与纯合隐性植物相比，在半合植物组中观察到纤维素含量的较

大提高。与纯合隐性植物相比，来自事件 24B 的半合植物显示纤维素含量提高 9.7%，纯合显性植物显示纤维素含量提高 7.5% (图 6;  $P \leq 0.05$ , t-检验)。与纯合隐性植物组相比，来自事件 25B 的半合植物显示提高 10.4% (图 7;  $P < 0.05$ , t-检验)。对于这两个事件，没有观察到木素含量的显著改变。

<110> GERHARDT, Isabel Rodrigues  
ARRUDA, Paulo

<120> 编码 C3HC4 家族植物蛋白质的核酸分子和改变植物纤维素和木素含量的方法

<130>

<140>

<141>

<160> 4

<210> 1

<211> 1029

<212> DNA

<213> 美洲黑杨(Populus deltoides)

<221> DNA

<220>

<221> CDs

<222> (1)...(1029)

<223> 美洲黑杨 C3HC4 蛋白, cDNA 克隆

<400> 1

```

atg aat acg cgg tac ccc ttt cca atg aat tca ttc tgt ggt tca agt      48
Met Asn Thr Arg Tyr Pro Phe Pro Met Asn Ser Phe Cys Gly Ser Ser
          5                10                15

cct gca gca tct cca gca gga gaa gag cgt gga act gtg tct ctc aga      96
Pro Ala Ala Ser Pro Ala Gly Glu Glu Arg Gly Thr Val Ser Leu Arg
          20                25                30

aac agg tct gct cgc act ccc act tcg tcc tta ttg gta aga aca gcg      144
Asn Arg Ser Ala Arg Thr Pro Thr Ser Ser Leu Leu Val Arg Thr Ala
          35                40                45

atg aga ata tct aga gca aga tgg ttc acc ttc ttg aga aga gta ttc      192
Met Arg Ile Ser Arg Ala Arg Trp Phe Thr Phe Leu Arg Arg Val Phe
          50                55                60

cac tac cag aat ggc tca agg tct aat ctt ggg tct aat cct ttc aac      240
His Tyr Gln Asn Gly Ser Arg Ser Asn Leu Gly Ser Asn Pro Phe Asn
          65                70                75                80

tcc agc act tgg atg atg ctg gag ttt gta gct ttg gtc ctt caa ata      288
Ser Ser Thr Trp Met Met Leu Glu Phe Val Ala Leu Val Leu Gln Ile
          85                90                95

agt att acc atg ttc acc ctg gct att tcc aag gca gag aag cca gtt      336
Ser Ile Thr Met Phe Thr Leu Ala Ile Ser Lys Ala Glu Lys Pro Val
          100                105                110

```



tgg cct gtg aga ata tgg att att ggc tat aat att ggt tgt gtc ctt	384
Trp Pro Val Arg Ile Trp Ile Ile Gly Tyr Asn Ile Gly Cys Val Leu	
115 120 125	
agt ctg ctg ctg ttc tat ggg cga tac cgg caa att aac acg act caa	432
Ser Leu Leu Leu Phe Tyr Gly Arg Tyr Arg Gln Ile Asn Thr Thr Gln	
130 135 140	
ggc gat ggt ttt ggc cta cct gat ttg gaa caa cag agg ggc agt gag	480
Gly Asp Gly Phe Gly Leu Pro Asp Leu Glu Gln Gln Arg Gly Ser Glu	
145 150 155 160	
gaa tcc agt gta tgc agg tgc tca att ttg atg cac aag tgc cgg act	528
Glu Ser Ser Val Cys Arg Cys Ser Ile Leu Met His Lys Cys Arg Thr	
165 170 175	
tca ctc gag ctc ttc ttt gca ata tgg ttt gtg atg ggt aat gtt tgg	576
Ser Leu Glu Leu Phe Phe Ala Ile Trp Phe Val Met Gly Asn Val Trp	
180 185 190	
gtc ttt gat tct cgc ttt gga tcc tac cac cgt gct cca aaa ctc cat	624
Val Phe Asp Ser Arg Phe Gly Ser Tyr His Arg Ala Pro Lys Leu His	
195 200 205	
gtg cta tgc atc tct cta ctt gcc tgg aat gct ctc agt tac tcc ttc	672
Val Leu Cys Ile Ser Leu Leu Ala Trp Asn Ala Leu Ser Tyr Ser Phe	
210 215 220	
cca ttt ctc ttg ttt cta ctc tta tgt tgt tgc gtg cca ctc att agc	720
Pro Phe Leu Leu Phe Leu Leu Leu Cys Cys Cys Val Pro Leu Ile Ser	
225 230 235 240	
act gtc ctt ggt tac aac atg aac atg ggt tct gcc gag aga ggg gct	768
Thr Val Leu Gly Tyr Asn Met Asn Met Gly Ser Ala Glu Arg Gly Ala	
245 250 255	
tcc gac gat caa atc tcc agt cta cca agt tgg agg tac aaa gca gct	816
Ser Asp Asp Gln Ile Ser Ser Leu Pro Ser Trp Arg Tyr Lys Ala Ala	
260 265 270	
gac acc aac tcg gag ttt cgc aat aat gct gac tgc aat tca acc att	864
Asp Thr Asn Ser Glu Phe Arg Asn Asn Ala Asp Cys Asn Ser Thr Ile	
275 280 285	
gca agt gaa gat ctg gaa tgt tgt atc tgc cta gcc aag tat aaa gac	912
Ala Ser Glu Asp Leu Glu Cys Cys Ile Cys Leu Ala Lys Tyr Lys Asp	
290 295 300	
aaa gaa gaa gtt agg aaa ttg cca tgt tcc cat atg ttc cac ctc aag	960
Lys Glu Glu Val Arg Lys Leu Pro Cys Ser His Met Phe His Leu Lys	
305 310 315 320	
tgt gta gat caa tgg ctt cga att ata tcc tgc tgc cct ctc tgt aaa	1008
Cys Val Asp Gln Trp Leu Arg Ile Ile Ser Cys Cys Pro Leu Cys Lys	
325 330 335	
caa gga ttg gag aga tag	1029
Gln Gly Leu Glu Arg	

&lt;210&gt; 2

&lt;211&gt; 342

&lt;212&gt; PRT

&lt;213&gt;美洲黑杨

&lt;220&gt;

&lt;400&gt; 2

Met Asn Thr Arg Tyr Pro Phe Pro Met Asn Ser Phe Cys Gly Ser Ser  
5 10 15

Pro Ala Ala Ser Pro Ala Gly Glu Glu Arg Gly Thr Val Ser Leu Arg  
20 25 30

Asn Arg Ser Ala Arg Thr Pro Thr Ser Ser Leu Leu Val Arg Thr Ala  
35 40 45

Met Arg Ile Ser Arg Ala Arg Trp Phe Thr Phe Leu Arg Arg Val Phe  
50 55 60

His Tyr Gln Asn Gly Ser Arg Ser Asn Leu Gly Ser Asn Pro Phe Asn  
65 70 75 80

Ser Ser Thr Trp Met Met Leu Glu Phe Val Ala Leu Val Leu Gln Ile  
85 90 95

Ser Ile Thr Met Phe Thr Leu Ala Ile Ser Lys Ala Glu Lys Pro Val  
100 105 110

Trp Pro Val Arg Ile Trp Ile Ile Gly Tyr Asn Ile Gly Cys Val Leu  
115 120 125

Ser Leu Leu Leu Phe Tyr Gly Arg Tyr Arg Gln Ile Asn Thr Thr Gln  
130 135 140

Gly Asp Gly Phe Gly Leu Pro Asp Leu Glu Gln Gln Arg Gly Ser Glu  
145 150 155 160

Glu Ser Ser Val Cys Arg Cys Ser Ile Leu Met His Lys Cys Arg Thr  
165 170 175

Ser Leu Glu Leu Phe Phe Ala Ile Trp Phe Val Met Gly Asn Val Trp  
180 185 190

Val Phe Asp Ser Arg Phe Gly Ser Tyr His Arg Ala Pro Lys Leu His  
195 200 205

Val Leu Cys Ile Ser Leu Leu Ala Trp Asn Ala Leu Ser Tyr Ser Phe  
210 215 220

Pro Phe Leu Leu Phe Leu Leu Leu Cys Cys Cys Val Pro Leu Ile Ser  
225 230 235 240

Thr Val Leu Gly Tyr Asn Met Asn Met Gly Ser Ala Glu Arg Gly Ala  
245 250 255

---

Ser Asp Asp Gln Ile Ser Ser Leu Pro Ser Trp Arg Tyr Lys Ala Ala  
 260 265 270

Asp Thr Asn Ser Glu Phe Arg Asn Asn Ala Asp Cys Asn Ser Thr Ile  
 275 280 285

Ala Ser Glu Asp Leu Glu Cys Cys Ile Cys Leu Ala Lys Tyr Lys Asp  
 290 295 300

Lys Glu Glu Val Arg Lys Leu Pro Cys Ser His Met Phe His Leu Lys  
 305 310 315 320

Cys Val Asp Gln Trp Leu Arg Ile Ile Ser Cys Cys Pro Leu Cys Lys  
 325 330 335

Gln Gly Leu Glu Arg  
 340

<210> 3  
 <211> 30  
 <212> DNA  
 <213> 人工序列  
 <220>  
 <223> 合成寡核苷酸 (C3HC4NDE 引物)  
 <400> 3  
 catatgaata cgcggtaccc ctttccaatg

30

<210> 4  
 <211> 31  
 <212> DNA  
 <213> 人工序列  
 <220>  
 <223> 合成寡核苷酸 (C3HC4XBA 引物)  
 <400> 4  
 tctagactat ctctccaatc cttgtttaca g

31

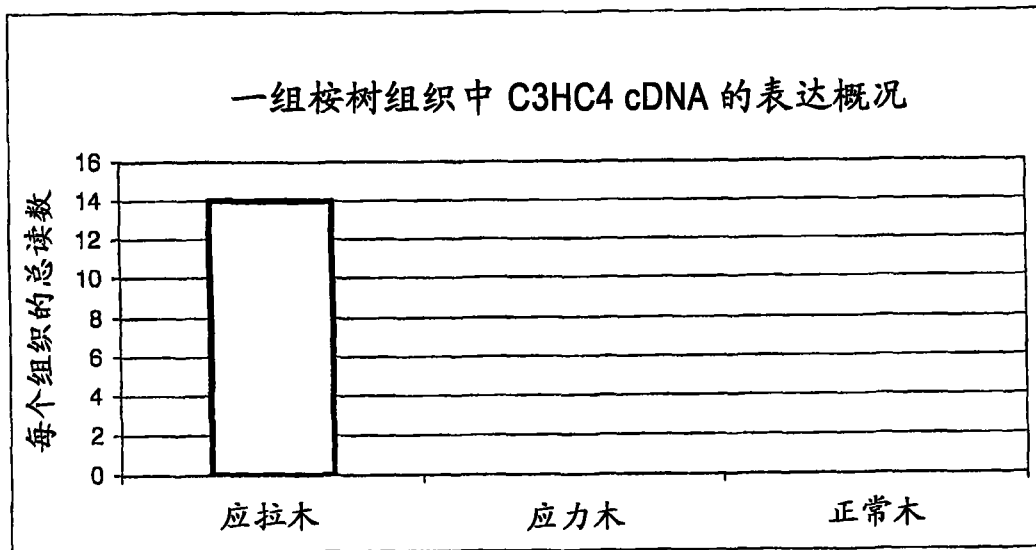


图 1

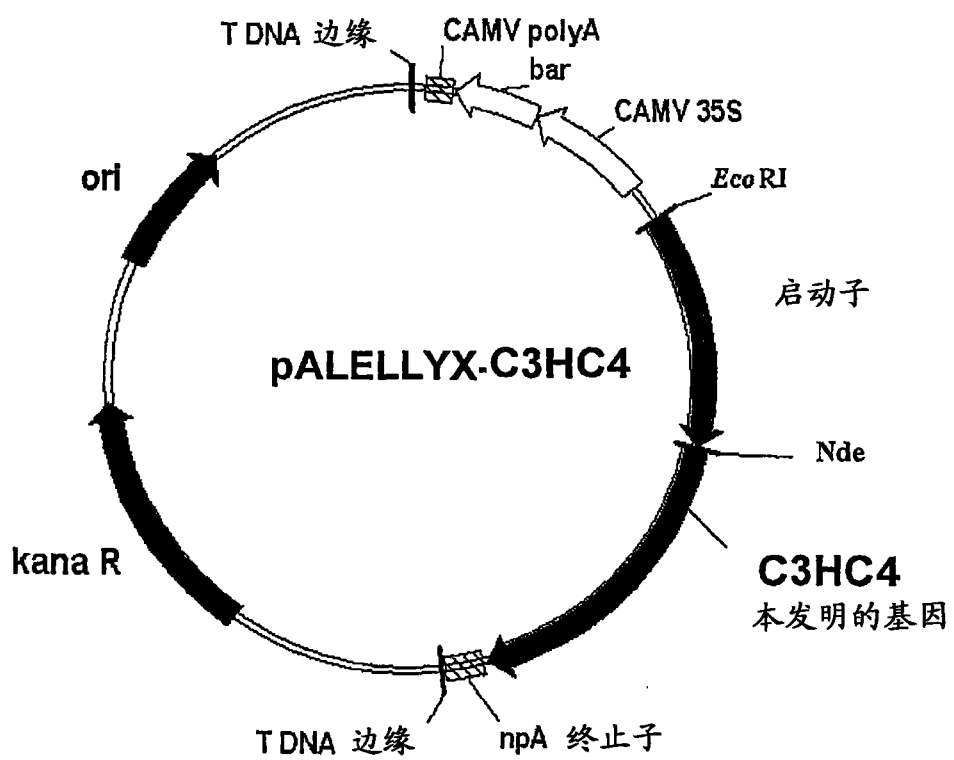


图 2

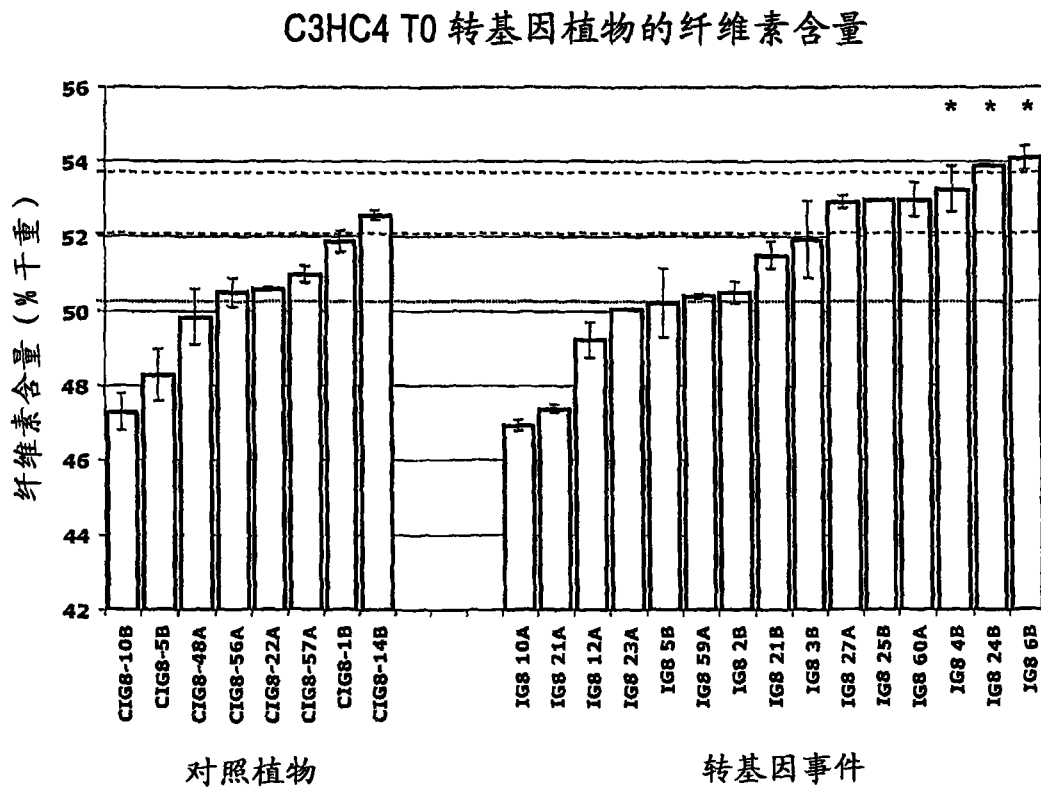


图 3

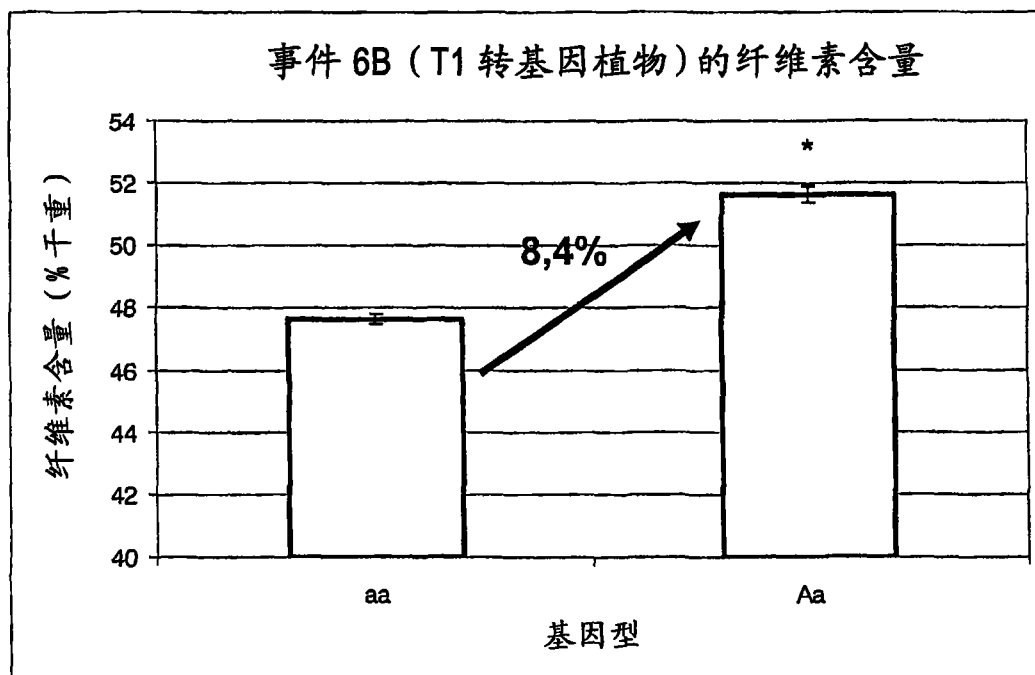


图 4

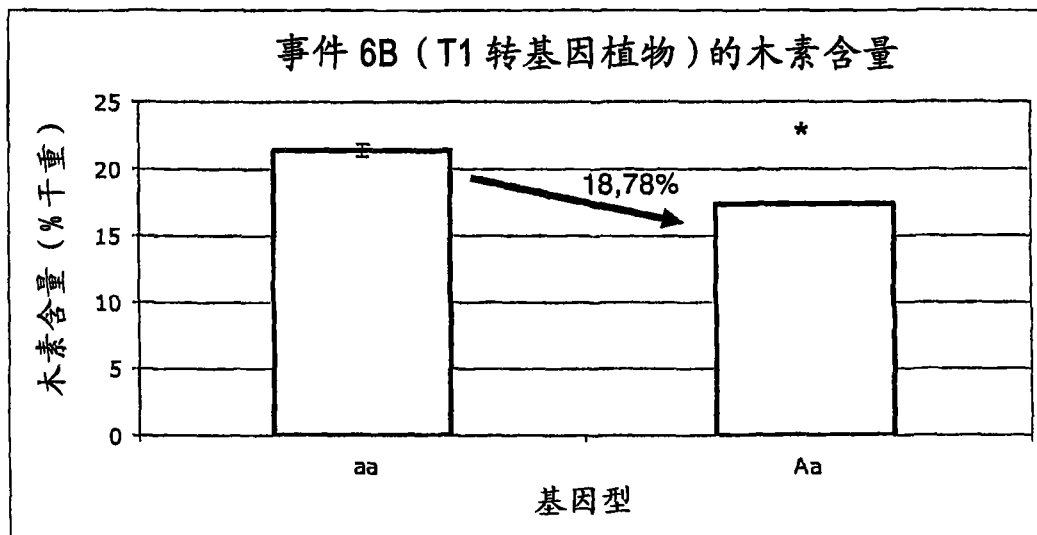


图 5



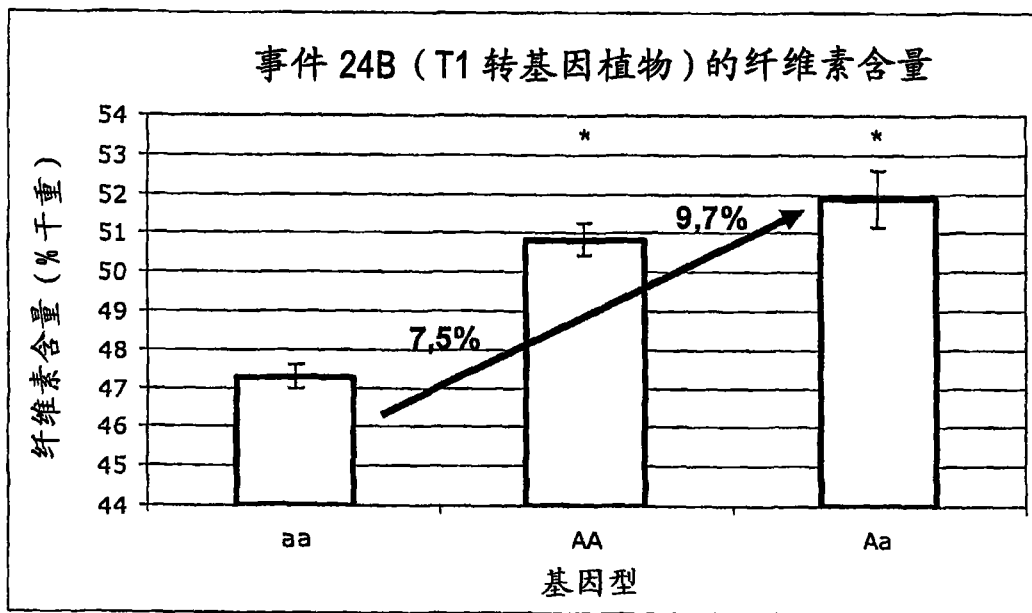


图 6

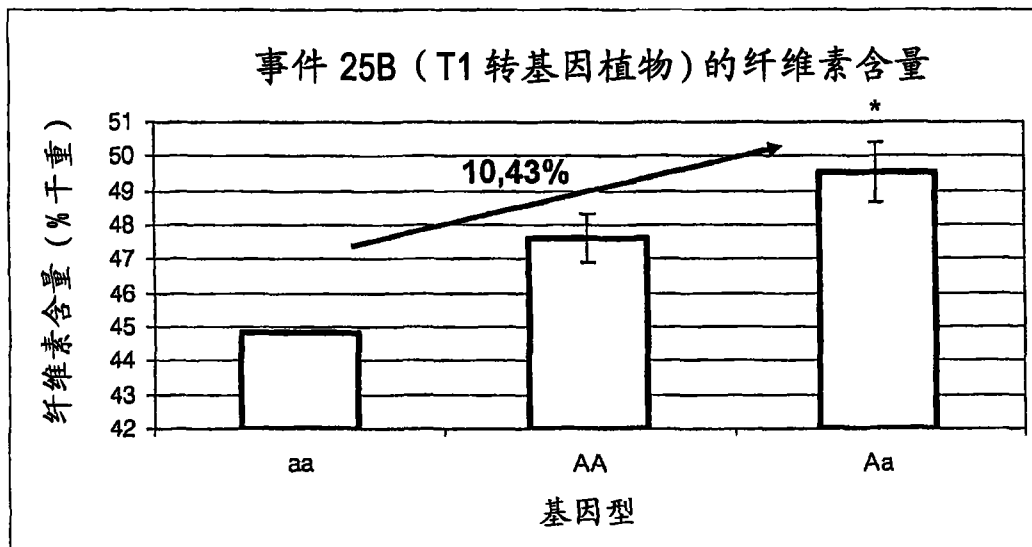


图 7