



(12)发明专利



(10)授权公告号 CN 105143247 B

(45)授权公告日 2019.11.08

(21)申请号 201480003586.5

(22)申请日 2014.11.21

(65)同一申请的已公布的文献号

申请公布号 CN 105143247 A

(43)申请公布日 2015.12.09

(30)优先权数据

61/908444 2013.11.25 US

(85)PCT国际申请进入国家阶段日

2015.06.18

(86)PCT国际申请的申请数据

PCT/US2014/066772 2014.11.21

(87)PCT国际申请的公布数据

W02015/077538 EN 2015.05.28

(73)专利权人 孟加拉朱特研究所

地址 孟加拉国达卡

(72)发明人 M.阿拉姆 M.S.伊斯拉姆

B.艾哈迈德 M.S.哈克

M.M.阿拉姆

(74)专利代理机构 中国专利代理(香港)有限公

司 72001

代理人 李波 黄希贵

(51)Int.Cl.

C07K 14/415(2006.01)

C12N 15/29(2006.01)

(56)对比文件

CN 1352194 A,2002.06.05,

CN 1860231 A,2006.11.08,

WO 2012149009 A2,2012.11.01,

WO 2012177418 A1,2012.12.27,

HARRIS JOHN C.Modulation of plant growth by HD-Zip class 1 and 11

transcription factors in response to environmental stimuli.《THE NEW

PHYTOLOGIST》.2011,第190卷(第4期),823-837.

Kellmer.PhD Thesis: Funktionelle

Charakterisierung von CKX7 und

cytokininregulierten

Transkriptionsfaktorgenen in Arabidopsis thaliana.《Frei Universität Berlin》.2009,

审查员 吕健

权利要求书1页 说明书17页

序列表7页 附图1页

(54)发明名称

包含HAT22多核苷酸和多肽的新颖组合物以及使用方法

(57)摘要

本发明涉及分离的多核苷酸和源自其的相应多肽,所述多核苷酸编码来自植物长蒴黄麻和圆果种黄麻的同源框-亮氨酸拉链蛋白HAT22(HD-ZIP蛋白22)。本发明进一步提供了包括和/或由同源框-亮氨酸拉链蛋白HAT22(HD-ZIP蛋白22)的核苷酸序列组成的载体、表达构建体和宿主细胞。本发明也提供了产生所述蛋白质的方法和修饰所述蛋白质以便改善它们的期望特征的方法。本发明的所述蛋白质可用于各种途径,包括增加/增强植物的纤维长度、高度和生物质以

及纤维产率。

1. 源自植物长蒴黄麻 (*Corchorus olitorius*) 的分离的多核苷酸, 其由SEQ ID NO:2所示的核苷酸序列组成, 其中所述多核苷酸编码同源框-亮氨酸拉链HAT22蛋白。
2. 根据权利要求1所述的分离的多核苷酸, 其中所述植物长蒴黄麻是品种0-4。

包含HAT22多核苷酸和多肽的新颖组合物以及使用方法

[0001] 相关申请

[0002] 本申请要求2013年11月25日提交的美国临时申请号61/908,444 的益处;其内容通过引用以其整体并入本文。

技术领域

[0003] 本发明涉及分子生物学领域,和通过调节编码同源框-亮氨酸拉链 HAT22多肽的核酸表达来增加或增强植物的纤维长度、高度和/或生物量的方法。而且,具体地,本发明提供了同源框-亮氨酸拉链蛋白HAT22 同系物,其分离自两个黄麻植物物种,即,长蒴黄麻(“*C. olitorius*”)和圆果种黄麻(“*C. capsularis*”),和它们在增加黄麻的纤维长度、植物高度和/或植物生物量中的应用通过调节编码同源框-亮氨酸拉链蛋白HAT22的基因的表达。

背景技术

[0004] 商业上栽培黄麻主要产生自两个物种,即,长蒴黄麻(通常称为“纤维黄麻”)和圆果种黄麻(通常称为“白黄麻”)。两个物种是二倍体($2n = 14$),并且仅仅这两个用在农业中。这两个物种构成了东南亚国家和巴西的重要经济作物,提供生物可降解的木质纤维素纤维。纤维通常用作工业产品的包装材料,和生产其他附加值产品,比如纱和纺织品、绳索、细绳和网、非织造织物、纸巾、纸浆和纸、土工布、复合材料和家用纺织品。由于黄麻纤维的环境友好和生物可降解的特征,全球市场中的需求增加。为了满足不断增加的全球需求,需要进一步改善黄麻纤维生产。

[0005] 黄麻纤维获得自茎部并且纤维分组成束。布置为厚壁组织纤维细胞的三角形楔形物的纤维与其他软组织的髓线交替。这些纤维细胞与韧皮部不同并且以四个按时间顺序的阶段发育——启动、延伸、第二细胞壁增厚和成熟阶段。关于黄麻纤维启动和延伸的遗传控制知道的较少。但是,纤维长度和均匀性是大部分工业使用的常规要求。例如,对于纺织工业的细织物生产,长纤维和均匀细胞是理想的。在纸浆工业中,纸浆的强度特征部分由纤维长度、纸浆产率和碱消耗量决定,原因是它们的强度和粘合特性。为满足工业需要,开发具有期望纤维长度以及强度的黄麻品种是必要的。所以,纤维的生物合成和与纤维生物合成相关的分子生物学是非常重要的。

[0006] 韧皮纤维主要开始于顶端分生组织,并且纤维逐渐延伸,主要通过侵入生长,直到出现第二个壁发育。纤维细胞延伸是源自内部膨压和细胞壁的机械强度之间相互作用的过程,其是内生控制的。但是,还未完全确定纤维细胞延伸相关的机制和基因。

[0007] 已经鉴定了与棉花纤维的延伸和形成相关的数个候选基因。例如,通过差别筛选和展示方法鉴定了来自棉花植物专门在棉花纤维延伸阶段表达的五个基因[美国专利号5,880,100和美国专利号5,932,713, 6,225,536和6,166,294;其所有通过参考并入],但是,在黄麻和/或韧皮纤维启动和延伸领域,没有这样的报道。

[0008] 同源框-亮氨酸拉链(HD-Zip)蛋白是植物特有的转录因子,由植物基因组中的多

拷贝的基因编码。例如,拟南芥基因组具有大于25个该家族的基因。基因中,同源框-亮氨酸拉链蛋白HAT22 (HD-ZIP蛋白 22) 参与脉管纤维细胞的延伸。

[0009] HD-ZIP蛋白22在其中纤维分化的束间区域中表达,与其在束间纤维分化的作用一致。此外,其也在第二生长阶段期间在韧皮部中表达 (Tornerio等1996), 指示其在调节脉管组织形成中的作用 (Zhong和 Ye,1999)。不像其他植物同源框基因,HD-ZIP蛋白22不在分生组织中表达,并且HD-ZIP蛋白22的功能可能是参与第二阶段的脉管发育期间韧皮部组织的同一性和/或活性的调节 (Tornerio等1996)。

[0010] 鉴于同源框-亮氨酸拉链蛋白HAT22可在韧皮部纤维分化的空间控制中起重要的作用,期望采用遗传方法来通过采用和利用与同源框-亮氨酸拉链蛋白HAT22相关的分子生物和遗传信息,理解植物中的纤维的生物合成。另外,因为植物每个物种的纤维生物合成途径和遗传组成通常不同,所以物种特异性方法也是优选的以便优化从黄麻植物的纤维生产率,并且获得兼容的结果,以确保工业中的使用。

发明内容

[0011] 本发明在某些实施方式中涉及参与确定或塑造纤维长度、植物高度和/或植物生物量的多核苷酸和多肽,以及使用其的方法。

[0012] 本发明也涉及鉴定参与同源框-亮氨酸拉链蛋白HAT22介导的过程的基因,使用这样的基因和蛋白质改变植物特征的用途。

[0013] 本发明提供了源自两个黄麻植物长蒴黄麻和圆果种黄麻的同源框-亮氨酸拉链蛋白HAT22、包括和/或由调节同源框-亮氨酸拉链HAT22 多肽在植物细胞中表达组成的方法,和调节植物中改善的纤维长度和/或高度和/或生物质的方法。

[0014] 本发明的一个方面提供了编码下述蛋白质的基因,所述蛋白质源自长蒴黄麻和圆果种黄麻,其参与改善的纤维长度、植物高度,和/或植物生物量产生,并且具有同源框-亮氨酸拉链蛋白HAT22样序列。更具体而言,本发明提供了能够诱导和/或增强植物中纤维长度、植物高度和/或植物生物量产生的基因;和从而编码的蛋白质,以及其用途。

[0015] 本发明的另一目的是提供遗传信息,其是关于待采用/利用以诱导、改善或增强长蒴黄麻和圆果种黄麻植物,以及其他韧皮纤维生产植物的纤维长度和/或植物高度和/或生物量产生的同源框-亮氨酸拉链蛋白 HAT22。

[0016] 本发明的仍另一目的是获得长蒴黄麻和圆果种黄麻的转基因植物,其通过调节植物中同源框-亮氨酸拉链蛋白HAT22的生物合成而具有增强的纤维长度和/或植物高度/生物量。

[0017] 本发明的仍另一目的是提供分离的多核苷酸,其可用于利于本文公开的方法的表现和开发转基因长蒴黄麻和圆果种黄麻植物。

[0018] 本发明的进一步目的是提供商业上可行的方式,以增加黄麻纤维基产品的纤维长度和/或植物高度。

[0019] 本发明进一步提供了分离自长蒴黄麻的基因,其编码具有如阐释 SEQ ID NO:3的氨基酸序列的蛋白质,其参与诱导、改善或增强纤维长度和/或植物高度和/或生物量,并且具有同源框-亮氨酸拉链蛋白 HAT22样序列。本发明也提供了编码蛋白质的基因,所述蛋白质具有通过添加或缺失一个或多个氨基酸和/或用如SEQ ID NO:3中阐释的氨基酸序列的

其他氨基酸替换而修饰的氨基酸序列,其参与诱导、改善或增强纤维长度和/或植物高度/生物量,并且具有同源框-亮氨酸拉链蛋白HAT22样序列。本发明另外提供了与如SEQ ID NO:1中所阐释核酸杂交的基因,尤其它的DNA或其一部分,并且其编码参与诱导、改善或增强纤维长度和/或植物高度/生物量的蛋白质,并且具有同源框-亮氨酸拉链蛋白HAT22样序列。

[0020] 本发明此外提供了分离自圆果种黄麻的基因,其编码具有如SEQ ID NO:6中阐释的氨基酸序列的蛋白质,其参与诱导、改善或增强纤维长度和/或植物高度/生物量,并且具有同源框-亮氨酸拉链蛋白 HAT22样序列。本发明也提供了编码蛋白质的基因,所述蛋白质具有通过添加或缺失一个或多个氨基酸和/或用如SEQ ID NO:6中阐释的氨基酸序列的其他氨基酸替换而修饰的氨基酸序列,其参与诱导、改善或增强纤维长度和/或植物高度/生物量,并且具有同源框-亮氨酸拉链蛋白HAT22样序列。本发明进一步提供了与SEQ ID NO:4中阐释的核酸杂交的基因,尤其其DNA或其一部分,并且其编码参与诱导、改善或增强纤维长度和/或植物高度/生物量的蛋白质,并且具有同源框-亮氨酸拉链蛋白HAT22样序列。

[0021] 根据本发明的一种实施方式,长蒴黄麻的植物是品种0-4,和圆果种黄麻的植物是品种CVL-1。

[0022] 本发明的另一实施方式公开重组基因构建体,其包含和/或由具有 SEQ ID NO:2和/或SEQ ID NO:5中阐释的核苷酸序列的多核苷酸组成,其中多核苷酸在宿主细胞中是可表达的,以分别在长蒴黄麻的植物和圆果种黄麻中产生同源框-亮氨酸拉链蛋白HAT22的同系物。

[0023] 本发明的进一步实施方式是包含和/或由重组基因构建体组成的转化株,所述重组基因构建体能够表达具有SEQ ID NO:2和/或SEQ ID NO:5中阐释的核苷酸序列的多核苷酸,以产生同源框-亮氨酸拉链 HAT22蛋白的同系物。

[0024] 本发明也提供了通过培养和/或栽培上述宿主产生蛋白质的方法,所述蛋白质参与诱导、改善或增强纤维长度和/或植物高度/生物量,并且具有同源框-亮氨酸拉链蛋白HAT22样序列。

[0025] 本发明也提供了诱导、改善或增强植物或植物细胞的纤维长度和/或植物高度/生物量的方法。此外也公开了一种方法,其包括和/或由下述组成的方法:将上述基因引入植物或植物细胞,并且驱动所述基因的表达。

[0026] 本领域技术人员容易认识到,本发明非常适于进行目标并且获得提到的目的和优势,以及其中固有的那些。本文所述的实施方式不旨在限制本发明的范围。

[0027] 参考下述说明书和权利要求,将更容易理解本发明的这些和其他特征、方面和优势。

附图说明

[0028] 图1显示系统发生树,其比较来自长蒴黄麻的SEQ ID NO.3和来自圆果种黄麻的SEQ ID NO.6,以及产生同源框-亮氨酸拉链HAT22 蛋白的其他氨基酸序列。

[0029] 发明详述

[0030] 从下述形成本申请一部分的详细说明书和附图可更充分理解本发明。

[0031] 本文提供的定义和/或方法限定本发明并且在本发明的实践中指导本领域技术人

员。除非以其他方式指明,应根据相关领域技术人员的常规的用途理解术语。在任何定义和/或方法与本文并入的任何专利或非专利参考文献提供的或其他地方出现的任何参考文献的任何定义和/或方法不一致的情况下,应理解,本文使用在本申请中已经清楚提供/改编的所述定义和/或方法。单数术语“一个(a/an)”和“所述”包括复数参照物,除非上下文清楚得另外指出。类似地,词语“或”旨在包括“和”,除非上下文清楚得另外指出。因此,“包括A或B”意思是包括A或B,或A和B。进一步理解,为核酸或多肽提供的所有基础尺寸或氨基酸尺寸,和所有分子量或分子量值是近似的,并且提供为了说明。尽管与本文所述的那些类似或相同的方法和材料可用于本公开的实践或测试,但是下面描述了适当的方法和材料。

[0032] 本发明包括分离的多核苷酸,其编码提取自长蒴黄麻和圆果种黄麻的同源框-亮氨酸拉链蛋白HAT22,和源自其的相应多肽。更具体地,本发明提供了同源框-亮氨酸拉链蛋白HAT22同系物和它们在长蒴黄麻和圆果种黄麻中诱导、改善或增强纤维长度和/或植物高度/生物量中的的应用,以及其转基因长蒴黄麻和圆果种黄麻植物。主要通过比较长蒴黄麻和圆果种黄麻基因组DNA的核苷酸序列与其他植物的已知酶基因的核苷酸序列,鉴定编码酶的本发明的基因组序列。在本发明之前,长蒴黄麻和圆果种黄麻基因的核苷酸序列、阅读框、外显子和内含子的位置、酶的结构,和他们在开发高纤维质量中的用途是未知的,尤其就纤维长度、植物高度/生物量和黄麻植物的产率潜能而言。

[0033] 商业上栽培的黄麻物种长蒴黄麻和圆果种黄麻的基因组序列的分析,揭示两个物种具有编码具有在黄麻中诱导、改善或增强纤维长度和/或植物高度的催化活性的酶的单个基因。起初由软件程序,比如 Augustus,基于半HMM的核酸分析程序(SNAP),和Geneid (Genome BioInformatics Research Lab)注释核苷酸序列,其可鉴定推定编码区域,内含子,和拼接接头。进一步进行自动和手动管理核苷酸序列,以优化和确立编码区域和其他基因特征的精确表征。

[0034] 部分或完全测序来自长蒴黄麻的超过30,096cDNA和来自圆果种黄麻的37,031cDNA。从它们中,我们开发了来自每个物种编码新酶的单个cDNA,具有推定的作用,优选的催化活性,在黄麻中诱导、改善或增强纤维长度和/或植物高度。

[0035] 在对源自长蒴黄麻和圆果种黄麻mRNA的cDNA文库的克隆完全或部分测序之后分析开放阅读框(ORF),并且使用序列分析软件进一步分析,并且通过测定与数据库中已知序列的同源性(公共的/私人的)。

[0036] 本公开的上下文中,遍及说明书使用的许多术语具有指示的意思,除非明确指出具有不同的意思。

[0037] 如本文所使用,“多核苷酸”是比如核酸片段的核苷酸序列。多核苷酸可以是单链或双链的RNA或DNA的聚合物,其任选地包含合成的、非天然的或改变的核苷酸碱基。为DNA聚合物形式的多核苷酸可包括和/或由cDNA、基因组DNA、合成DNA,或其混合物/组合的一个或多个区段组成。本发明分离的多核苷酸可包括源自SEQ ID NO.1 和SEQ ID NO.4的至少一条150个连续核苷(上游和下游二者),或这些序列的互补序列。

[0038] 如本文所使用,“多肽”是通过肽键结合在一起的线性单链氨基酸,并且通常长度具有大于100个氨基酸的序列。

[0039] “分离的”意思是“通过手动”改变天然状态。如果组合物或物质在自然中存在,如果已经被或去除从其初始环境,或二者,其已经是被“分离的”。例如,天然存在于活植物或

动物中的多核苷酸或多肽不是“分离的”，但是从其天然状态的共存材料分开的相同多核苷酸或多肽是“分离的”，如本文所采用的术语。

[0040] 如本文所使用，术语“基因”定义为植物长蒴黄麻和圆果种黄麻的基因组序列，尤其是编码同源框-亮氨酸拉链HAT22酶多肽序列的多核苷酸序列，所述酶参与黄麻中增强纤维长度和/或植物高度的优选的催化活性。术语可进一步包括核酸分子，包括上游、下游和/或内含子核苷酸序列。

[0041] “编码序列”或“编码区域”指当序列表达时，具有足以产生基因产物，比如氨基酸或多肽的序列信息的核酸分子。编码序列可包括和/或由下述组成：翻译区域中的非翻译序列（包括内含子或5'或3'非翻译区），或可缺少这种中间非翻译序列（例如，如在cDNA中）。

[0042] 如本文所使用，术语“寡核苷酸”，是短多核苷酸或一部分多核苷酸，其长度可优选地包括10-1000，最优选地12至50个核苷。在本发明的实施方式中，寡核苷酸中包含的核苷可以是天然存在的核苷的类似物或衍生物。

[0043] 如本文所使用，术语“引物”是能够结合靶核酸序列和启动核酸合成的寡核苷酸。如本文定义的扩增寡核苷酸可优选地长度是10至50，最优选地15至25个核苷。此外，本发明的扩增寡核苷酸可以是化学合成的并且这样的寡核苷酸不是天然存在的核酸。

[0044] 遍及说明书使用的缩写指包括和/或由核苷酸序列组成的核酸是常规的单字符缩写。因此当包括在核酸中时，天然存在的编码核苷的缩写如下：腺嘌呤(A)、鸟嘌呤(G)、胞嘧啶(C)、胸嘧啶(T)和尿嘧啶(U)。而且，除非另外指出，本文存在的核酸序列是5'→3'方向。

[0045] 如本文所使用，使用术语“互补的”和其衍生词指核酸通过熟知的规则配对，A与T或U配对和C与G配对。互补可以是“部分的”或“完全的”。在部分互补中，仅仅一些核酸碱基根据碱基对规则匹配；而在完全的或全部互补中，所有的碱基根据配对规则匹配。核酸链之间的互补程度可对核酸链之间的杂交的效率和强度有显著作用，如本领域熟知的。所述杂交的效率和强度取决于探测方法。

[0046] 如本文所使用，术语“宿主细胞”包括易于用核酸构建体或表达载体转化、转染、转导、表达等的任何细胞类型，所述核酸构建体或表达载体包括和/或由本发明的多核苷酸组成。适当的宿主细胞包括真菌和/或植物细胞，尤其生产韧皮纤维的植物细胞。

[0047] 术语“可操作地连接”在本文通常指其中控制序列相对于多核苷酸序列的编码序列布置在适当位置的构造，使得控制序列引导多肽的编码序列的表达。例如，当启动子影响编码序列的表达时，即该编码序列在启动子的转录控制下，启动子可以可操作地连接编码序列。

[0048] “载体”一般指复制子，比如质粒、噬菌体、黏粒、酵母或病毒，另一核酸区段可操作地插入其中，以便使得该区段复制或表达。术语“载体”也旨在指能够运输其已经连接的另一核酸的核酸分子。一种类型的载体是“质粒”，其指其中可连接另外的DNA区段的环形双链DNA环。另一类型的载体是病毒载体，其中另外DNA区段可连接至病毒基因组。某些载体能够在引入它们的宿主细胞中自主复制（例如，具有细菌复制起点的细菌载体和附加体哺乳类的载体）。当引入宿主细胞时，其他载体可整合至宿主细胞的基因组，并且从而与宿主基因组一起复制。而且，某些载体能够引导它们可操作连接的基因的表达。这样的载体在本文

称为“重组表达载体”(或简单地称为“表达载体”)。一般而言,在重组DNA技术中使用的表达载体通常为质粒的形式。在本说明书中,“质粒”和“载体”可交替使用,因为质粒是最常用的载体形式。但是,本发明旨在包括这样的其他形式的表达载体,比如病毒载体(例如,复制缺陷逆转录病毒、腺病毒和腺伴随病毒),其具有相同的功能。

[0049] 术语“核酸构建体”或“DNA构建体”有时用于指编码序列或可操作地连接至适当的调节序列并且插入载体用于转化细胞的序列。该术语可与术语“转化DNA”或“转基因”交替使用。

[0050] 如本文所使用,术语“启动子”指用于引导下游基因转录的核酸序列。启动子一般适于其中表达靶基因的宿主细胞。启动子与其他转录和翻译调节核酸序列(也称为“控制序列”)一起对于表达给定基因是必要的。一般而言,转录和翻译调节序列包括但不限于启动子序列、核糖体结合位点、转录起始序列和终止子序列、翻译起始序列和终止子序列,以及增强子或活化序列。

[0051] 如本文所使用,“同源框-亮氨酸拉链HAT22多肽”意思是具有同源框-亮氨酸拉链HAT22活性,即,参与植物中诱导、改善或增强纤维长度和/或植物高度/生物量的多肽。同源框-亮氨酸拉链蛋白HAT22活性刺激纤维细胞生长和/或植物高度/生物量。同源框-亮氨酸拉链蛋白HAT22的DNA-结合特性显示它们与DNA作为同二聚体相互作用并且识别两个独特的9bp(约)回文序列,CAAT(G/C)ATTG,如通过从随机序列DNA选择高亲和性结合位点所测定。术语也包括具有任何一种上述功能的片段、变体和同系物。

[0052] 如本文所使用,术语“同系物”指蛋白质包括肽、寡肽、多肽、蛋白质和酶,其相对于讨论的未修饰蛋白质具有氨基酸替换、缺失和/或插入并且具有与它们从中衍生的未修饰的蛋白质类似的生物和功能活性。

[0053] 缺失指从蛋白质去除一个或多个氨基酸。

[0054] 插入指一个或多个氨基酸残基引入蛋白质中的预定位点。插入可包括N-末端和/或C-末端融合以及序列内插入单个或多个氨基酸。一般而言,在氨基酸序列内插入比N-或C-末端融合小约1至10个残基的级别。

[0055] 置换指用具有类似特性(比如类似的疏水性、亲水性、抗原性、偏好,以形成或破坏 α -螺旋结构或 β -折叠结构)的其他氨基酸替换蛋白质的氨基酸。氨基酸置换通常为单个残基,但是取决于对多肽所起的作用功能约束可聚簇并且范围可为1至10个氨基酸;插入通常为约1至10个氨基酸残基的级别。氨基酸置换优选地为保守氨基酸置换。保守置换表是本领域熟知的(见例如Creighton(1984)Protein.W.H.Freeman 和Company (Eds) 和下面表1)。

[0056] 表1:保守氨基酸置换的例子

[0057]

残基	保守置换	残基	保守置换
Ala	Ser	Leu	Ile;Val
Arg	Lys	Lys	Arg;Gln
Asn	Gln;His	Met	Leu;Ile
Asp	Glu	Phe	Met;Leu;Tyr
Gln	Asn	Ser	Thr;Gly
Cys	Ser	Thr	Ser;Val

Glu	Asp	Trp	Tyr
Gly	Pro	Tyr	Trp;Phe
His	Asn;Gln	Val	Ile;Leu
Ile	Leu,Val		

[0058] 氨基酸置换、缺失和/或插入可容易使用本领域熟知的肽合成技术,比如固相肽合成和/或任何其他合成技术,或通过重组DNA操作进行。操作DNA序列以产生蛋白质的置换、插入或缺失变体的方法是本领域熟知的。例如,在DNA的预定位点形成置换突变的技术是本领域技术人员熟知的并且包括M13诱变、T7-Gen体外诱变(USB,Cleveland, OH)、Quick Change位点定向诱变(Stratagene, San Diego, CA)、PCR 介导的位点定向诱变或其他位点定向诱变方案。

[0059] 如本文所使用,“生物活性部分”可指同源框-亮氨酸拉链HAT22 蛋白的片段,其具有催化启动、形成、增强或改变长蒴黄麻或圆果种黄麻植物的韧皮部纤维的组成的生物活性。同源框-亮氨酸拉链HAT22 蛋白的生物活性部分包括下述肽或多肽,其包括的氨基酸序列足够与同源框-亮氨酸拉链HAT22蛋白的氨基酸序列相同或源自同源框-亮氨酸拉链HAT22蛋白的氨基酸序列,例如SEQ ID NO:3或SEQ ID NO: 6中阐释的氨基酸序列,其包括与全长同源框-亮氨酸拉链HAT22蛋白不同的少数氨基酸,并且展示同源框-亮氨酸拉链HAT22蛋白的至少一种活性。典型地,生物活性部分包括具有同源框-亮氨酸拉链HAT22 蛋白的至少一种活性的结构域或基序。同源框-亮氨酸拉链HAT22蛋白的生物活性部分可以是例如长度为254、255、256、257、258、259、260、261、262、263、264、265、266或267个氨基酸的多肽。

[0060] 同源框-亮氨酸拉链HAT22蛋白可具有SEQ ID NO:3或SEQ ID NO:6中阐释的氨基酸序列。在其他实施方式中,同源框-亮氨酸拉链 HAT22蛋白基本上与SEQ ID NO:3或SEQ ID NO:6相同,并且,尽管由于天然等位基因变化或诱变而氨基酸序列不同,但保持SEQ ID NO:3或SEQ ID NO:6蛋白的功能活性。在另一实施方式中,同源框-亮氨酸拉链HAT22蛋白包括与SEQ ID NO:3或SEQ ID NO:6有至少约95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高同一性的氨基酸序列。

[0061] 如本文所使用,术语“结构域:指沿着进化上亲缘的序列比对在特定位置保守的一组氨基酸。尽管在其他位置上的氨基酸可能在同系物之间不同,但是在特定位置高度保守的氨基酸指示可能对于蛋白质的结构、稳定性或功能是必须的氨基酸。通过在蛋白质同系物家族的对比序列中它们高度的保守鉴定,它们可用作标识符,以确定任何讨论的多肽是否属于先前鉴定的多肽家族。

[0062] 如本文所使用,术语“基序”指进化上亲缘蛋白质的序列中短的保守区域。基序通常是结构域的高度保守部分,但是也可包括结构域的仅仅一部分,或位于保守结构域的外部(如果基序的所有的氨基酸落在限定的结构域的外部)。

[0063] 为了鉴定结构域而存在的专用数据库,例如,SMART (Schultz等 (1998) Proc.Natl.Acad.Sci.USA 95,5857-5864;Letunic等(2002) Nucleic Acids Res 30,242-244), InterPro (Mulder等,(2003) Nucl.Acids. Res.31,315-318), Prosite (Bucher和Bairoch (1994),用于自动序列阐释的生物分子序列基序和其功能的一般轮廓特征。(In) ISMB-94;第二届分子生物学智能系统的国际大会的会议记录.Altman R., Brutlag D.,

Karp P., Lathrop R., Searls D., Eds., pp53-61, AAAI Press, Menlo Park; Hulo等, Nucl. Acids. Res. 32:D134-D137, (2004)), 或 Pfam (Bateman 等, Nucleic Acids Research 30 (1):276-280 (2002))。用于计算机分析蛋白质序列的一组工具在 ExPASy 蛋白质组服务器中是可用的 (Swiss Institute of Bioinformatics (Gasteiger 等, ExPASy: 用于深度蛋白质知识和分析的蛋白质组服务器, Nucleic Acids Res. 31:3784-3788 (2003)))。可使用常规技术, 比如通过序列比对来鉴定结构域或基序也。

[0064] 为了本发明的目的, “转基因的”、“转基因”或“重组”意思是例如, 核酸序列、表达盒、基因构建体或载体包括和/或由核酸序列组成或生物体用根据本发明的核酸序列、表达盒或载体转化, 所有的那些构建体通过重组方法产生, 其中: (a) 编码蛋白质的核酸序列用于本发明的方法, 或 (b) 遗传控制序列 (一条或多条) 可操作地连接根据本发明的核酸序列, 例如启动子, 或 (c) a) 和 b) 不在它们的天然遗传环境中或已经通过重组方法修饰。修饰可为, 例如置换、添加、缺失、倒置或插入一个或多个核苷酸残基的形式。天然遗传环境理解为意思是在起始植物中的天然基因组或染色体基因座或存在于基因组文库中。在基因组文库的情况下, 优选地至少部分保持核酸序列的天然遗传环境。环境在核酸序列至少一侧的侧翼并且序列长度为约 50bp, 优选地为约 500bp。天然存在的表达盒-例如天然存在的核酸序列的天然启动子与编码用于如上定义的本发明方法的多肽的相应的核酸序列组合-当该表达盒通过非天然合成的 (“人工”) 方法比如, 例如, 诱变处理修饰时变成转基因表达盒。适当的方法描述在, 例如 US 5,565,350 或 WO 00/15815 中 (两篇通过引用并入)。

[0065] 为了本发明的目的, 转基因植物理解为包括其中在本发明的方法中使用的核酸不在它们植物的基因组的天然基因座的那些植物, 并且因此核酸同源或异源表达是可能的。但是, 转基因也意思是, 尽管根据本发明的或在本发明的方法中使用的核酸在它们植物基因组的天然位置, 相对于天然序列, 序列已经被修饰, 和/或天然序列的调节序列已经被修饰。转基因理解为意思是根据本发明的核酸在基因组的非天然基因座的表达, 其中核酸发生同源的或, 优选地, 异源表达。

[0066] 如本文所使用, 术语“引入”、或“转化”包括外源性多核苷酸转化至宿主细胞, 无论用于转化的方法如何。能够随后克隆繁殖的植物组织, 无论通过器官发生或胚芽发生, 可用本发明的遗传构建体转化并且从其再生完整植物。取决于可用的克隆繁殖系统, 和最适于待转化的具体物种, 选择的具体组织不同。示例性组织靶包括叶盘、花粉、胚芽、子叶、下胚轴、雌配子体、愈伤组织、存在的分生组织 (例如, 顶端分生组织、腋芽和根分生组织), 和诱导分生组织 (例如, 子叶分生组织和下胚轴分生组织)。多核苷酸可瞬时或稳定引入宿主细胞并且可非整合保持为, 例如, 质粒。可选地, 其可整合至宿主基因组。所得转化植物细胞可然后用于以本领域熟知的方式再生转化植物。

[0067] 外源基因转移至植物的基因组称为转化。

[0068] 植物物种的转化现在是相当常规的技术。有利地, 任何数个转化方法可用于将感兴趣的基因引入适当的祖先细胞。为从植物组织或植物细胞转化和再生植物描述的方法可用于瞬时或稳定转化。转化方法包括使用脂质体、电穿孔和增加游离 DNA 吸收的化学药品、DNA 直接注入植物、粒子枪轰击, 和使用病毒或花粉和微投射的转化。转基因植物, 包括转基因作物植物优选地经土壤杆菌介导的转化产生。有利的转化方法是在植物学中的转化 (Sajib 等 2008)。

[0069] 通常,转化之后,选择存在由与感兴趣的基因一起转化,植物可表达的基因编码的一个或多个标记物的植物细胞或细胞组,其后将转化材料再生至完整植物。为了选择转化的植物,转化的获得的植物材料通常经受选择性条件,从而转化的植物可与未转化的植物不同。例如,可种植以上述方式获得的种子,并且在初始生长阶段之后,通过喷雾经受适当的选择。进一步可能性由如果需要,在消毒之后,在使用适当的选择试剂的琼脂平板上的种子生长组成,从而仅仅转化的种子可生长为植物。可选地,用存在的选择标记比如上述那些筛选转化的植物。

[0070] 在DNA转移和再生之后,也可评估推定转化的植物,例如使用 Southern分析,感兴趣的基因的存在、拷贝数和/或基因组组织。可选地或另外,可使用Northern和/或Western分析监测新引入的DNA的表达水平,两种技术是本领域技术人员熟知的。

[0071] 产生的转化植物可通过各种方式繁殖,比如通过克隆繁殖或经典繁殖技术。例如,第一代(或T1)转化植物可自授粉的并且可选择纯合子第二代(或T2)转化株,和T2植物可然后进一步通过经典繁殖技术繁殖。产生的转化生物体可采用各种形式。例如,它们可以是转化细胞和非转化细胞的嵌合体;克隆转化株(例如,包含表达盒的所有细胞转化);转化的和未转化的组织的嫁接体(例如,植物中,嫁接至未转化的接穗的转化根茎)。

[0072] 如本文所使用,术语“增加的表达”或“过表达”指比初始野生型表达水平更多的任何形式的表达。

[0073] 本领域充分记载了用于增加基因或基因产物表达的方法并且包括,例如,通过适当的启动子驱动的过表达、使用转录增强子或翻译增强子。用作启动子或增强子元件的分离的核酸可在非异源形式的多核苷酸的适当位置(通常上游)引入,以便上调编码感兴趣多肽的核酸的表达。例如,可通过突变、缺失和/或置换在体内改变内源启动子(见, Kmiec,US 5,565,350(通过参考并入);Zarling等W093/22443(通过参考并入)),或分离的启动子可以以适当的定向和与本发明基因适当的距离引入植物细胞,以便控制基因的表达。

[0074] 如果期望多肽表达,一般期望在多核苷酸编码区域的3'端包括多腺苷酸化区域。多腺苷酸化区域可源自天然基因、各种其他植物基因或T-DNA。待添加的3'端序列可源自,例如,胭脂氨酸合酶或章鱼碱合酶基因,或可选地源自另一植物基因,或次优选地源自任何其他真核基因。

[0075] 内含子序列也可添加至5'非翻译区(UTR)或部分编码序列的编码序列,以增加积聚在胞质溶胶中成熟的信使的量。已经显示,在植物和动物表达构建体的转录单元中包括可拼接的内含子增加在mRNA和蛋白质水平的基因表达至1000倍(Buchman和Berg(1988) Mol.Cell Biol.8:4395-4405;Callis等(1987) Genes Dev 1:1183-1200)。当布置在转录单元的5'端附近时,这样的内含子增强基因表达通常是最大的。玉米内含子Adh1-S内含子1、2和6,Bronze-1内含子的使用是本领域已知的。对于一般的信息见:The Maize Handbook, 116章, Freeling 和Walbot,Eds.,Springer,N.Y.(1994)。

[0076] 为了测定两个氨基酸序列或两个核酸序列的同一性百分数,可为了最佳比较目的而比对序列(例如,可在第一和第二氨基酸或核酸序列的之一或二者都引入空隙,以为了最佳比对,并且为了比较目的可忽略不同的序列)。为了比较目的,比对的参考序列的长度可以是参考序列的长度的至少95%。可然后比较相应的氨基酸位置或核苷酸位置的氨基酸残基或核苷。当第一序列中的位置被第二序列中相应位置的相同氨基酸残基或核苷酸占据

时,分子在该位置是相同的(如本文所使用,氨基酸或核酸“同一性”与氨基酸或核酸“同源性”相同)。两条序列之间的同一性百分数是序列共享的相同位置数量的函数,考虑需要为了两条序列的最佳比对引入空隙的数量,和空隙的长度。

[0077] 序列的比较和两条序列之间的同一性百分数的测定可使用数学算法完成。在一种实施方式中,两条氨基酸序列之间的同一性百分数可使用Needleman和Wunsch(J.Mol.Biol.48:444-453(1970))算法测定,其已经并入GCG软件包装的GAP程序(在<http://www.gcg.com>可用),使用Blosum 62矩阵或PAM250矩阵,和空隙权重为16、14、12、10、8、6或4和长度权重为1、2、3、4、5或6。在仍另一优选的实施方式中,两条核苷酸序列之间的同一性百分数可使用GCG软件包装中的GAP程序(在<http://www.gcg.com>可用)测定,使用NWSgapdna.CMP矩阵并且空隙权重为40、50、60、70或80和长度权重为1、2、3、4、5或6。在另一实施方式中,两条氨基酸或核苷酸序列之间的同一性百分数可使用E.Meyers和W.Miller(Comput)Appl.Biosci.4:11-17(1988))测定,其已经并入ALIGN程序(版本2.0或2.0U),使用PAM120权重残基表,空隙长度罚分为12和空隙罚分为4。

[0078] 可用于测定两条序列之间同一性的示例性计算机程序包括但不限于BLAST程序套装,例如,BLASTN、BLASTX和TBLASTX,BLASTP和TBLASTN,在www.ncbi.nlm.nih.gov/BLAST是公众可用的。

[0079] 当评估给定的核酸序列相对于GenBank DNA序列和其他公开数据库的核酸序列时,通常使用BLASTN程序进行序列搜索。对于针对GenBank蛋白质序列和其他公开数据库中的氨基酸序列搜索在所有阅读框中翻译的核酸序列,优选BLASTX程序。

[0080] 为了测定两条或更多条序列之间的“%同一性”,使用例如CLUSTAL-W程序进行选择序列的比对。

[0081] 本发明的一种实施方式是编码长蒴黄麻和圆果种黄麻中出现的同源框-亮氨酸拉链HAT22多肽的分离的多核苷酸,其包括和/或由分别在SEQ ID NO 2和SEQ ID NO 5中阐释的核苷酸序列组成。相应地,由这些核苷酸序列编码的各自的同源框-亮氨酸拉链HAT22多肽应具备SEQ ID NO 3和SEQ ID NO 6中阐释的氨基酸序列。根据本发明的一种实施方式,SEQ ID NO 3指源自长蒴黄麻的同源框-亮氨酸拉链蛋白HAT22同系物的多肽序列,和SEQ ID NO 6指源自圆果种黄麻的同源框-亮氨酸拉链蛋白HAT22同系物的多肽序列。这两种酶出现在长蒴黄麻和圆果种黄麻中纤维的生物合成途径中,用于催化增强植物中的纤维长度、植物高度和/或植物生物量产生。

[0082] 本发明也提供了编码来自植物长蒴黄麻和圆果种黄麻的同源框-亮氨酸拉链蛋白HAT22同系物的基因序列。

[0083] 在一种实施方式中,SEQ ID NO.1中阐释的1296bp长多核苷酸是分离自长蒴黄麻的全长基因。该基因序列包括来自基因上游和下游的至少150个连续核苷。这也提供了基因的基因内序列。

[0084] 在另一实施方式中,SEQ ID NO.4中阐释的1337bp长多核苷酸是分离自圆果种黄麻的全长基因。该基因序列包括来自基因上游和下游的至少150个连续核苷。这也提供了基因的基因内序列。

[0085] 在本发明的仍另一实施方式中,提供了编码多肽包括和/或由SEQ ID NO 2和/或SEQ ID NO 5中阐释的核苷酸序列组成的分离的多核苷酸。SEQ ID NO 2指源自长蒴黄麻

的同源框-亮氨酸拉链HAT22蛋白同系物序列的多核苷酸序列和SEQ ID NO 5指源自圆果种黄麻的同源框-亮氨酸拉链HAT22蛋白同系物序列的多核苷酸序列。

[0086] 在仍另一实施方式中,能够编码同源框-亮氨酸拉链HAT22多肽,或其生物活性片段的分离的核酸分子,包括的核苷酸序列与SEQ ID NO:2,SEQ ID NO:5阐释全长的核苷酸序列,或其任何互补物有至少约95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高的同一性。

[0087] 在一种实施方式中,SEQ ID NO.2中阐释的804bp长多核苷酸是全长cDNA克隆,其编码同源框-亮氨酸拉链HAT22蛋白,展示编码如SEQ ID NO.3阐释的267个氨基酸多肽的开放阅读框,计算的分子量为约29.99kD。通过SEQ ID NO.3的SMART分析,其揭示序列中同源框结构域的存在。这是DNA-结合因子,其参与植物关键发育过程的转录调节。这是同源框-亮氨酸拉链HAT22蛋白,其参与增强植物中纤维长度、植物高度和/或植物生物量产生。

[0088] 在一种实施方式中,SEQ ID NO.5中阐释的807bp长多核苷酸是全长cDNA克隆,其编码同源框-亮氨酸拉链HAT22蛋白,展示编码如SEQ ID NO.6阐释的268个氨基酸多肽的开放阅读框,计算的分子量为约30.10kD。通过SEQ ID NO.6的SMART分析,揭示序列中存在同源框结构域。这是DNA-结合因子,其参与植物的关键发育过程的转录调节。这是同源框-亮氨酸拉链HAT22蛋白,其参与植物中增强纤维长度、植物高度和/或植物生物量产生。

[0089] 根据本发明的实施方式,SEQ ID NO 2中阐释的分离的多核苷酸可通过使用分离自长蒴黄麻植物的总RNA的PCR扩增基因的保守区域而获得,并且SEQ ID NO 5可通过PCR使用分离自圆果种黄麻的植物的总RNA扩增该基因的保守区域而获得。如在之前的说明书中所阐释,使用的长蒴黄麻的植物是品种0-4,和使用的圆果种黄麻是品种 CVL-1。

[0090] 本发明的另一实施方式,公开了重组基因构建体,其包括和/或由具有SEQ ID NO 2和/或SEQ ID NO 4中阐释的核苷酸序列的多核苷酸组成,其中多核苷酸在宿主细胞中是可表达的,并且是可翻译的以在长蒴黄麻的植物和圆果种黄麻中。用于扩增、克隆和测序来自长蒴黄麻的植物和圆果种黄麻的同源框-亮氨酸拉链HAT22的程序在实施例2 中进一步详述。在某些实施方式中,重组基因构建体进一步包括可操作地连接的启动子区域,以增强多核苷酸模板的表达。在特定启动子的转录控制下,可接着改善包含SEQ ID NO 2和/或SEQ ID NO 4多核苷酸的编码区域在重组基因构建体中的表达,阐释更高产率的同源框-亮氨酸拉链HAT22蛋白。

[0091] 根据本发明的特征,调制的表达是增加的表达或活性,例如同源框-亮氨酸拉链HAT22多肽编码核酸分子,例如编码SEQ ID NO 2和 SEQ ID NO 5的核酸分子的过表达。本领域充分记载了用于增加核酸或基因,或基因产物表达的方法并且实施例提供在定义部分。

[0092] 本发明也提供了用于产生相对于对照植物具有改善的纤维长度和产率的转基因植物的方法,包括和/或由在植物中引入和表达编码如本文上面定义的同源框-亮氨酸拉链HAT22多肽的任何核酸组成。

[0093] 更具体而言,本发明提供了用于产生转基因植物的方法,所述转基因植物与无效对照植物相比具有改善的纤维产率,该方法包括:

[0094] (i) 在植物或植物细胞中引入和表达编码同源框-亮氨酸拉链HAT22 多肽的核酸或遗传构建体,其包括和/或由编码同源框-亮氨酸拉链 HAT22多肽的核酸组成;和

[0095] (ii) 在促进纤维细胞生长和发育的条件下培养植物细胞。

[0096] 本发明的另一方面涉及分离的多核苷酸,其编码同源框-亮氨酸拉链HAT 22蛋白并且源自植物长蒴黄麻,其包括选自下述的核酸分子:

[0097] a) 包括SEQ ID NO:2中阐释的核苷酸序列的核酸分子,或其互补物;和

[0098] b) 包括与SEQ ID NO:2中阐释的核苷酸序列具有至少95%序列同一性的核苷酸序列的核酸分子,或其互补物。

[0099] 在某些实施方式中,长蒴黄麻的植物是品种0-4。

[0100] 本发明的另一方面涉及分离的多核苷酸,其编码同源框-亮氨酸拉链HAT22蛋白并且源自植物圆果种黄麻,其包括选自下述的核酸分子:

[0101] a) 包括SEQ ID NO:5中阐释的核苷酸序列的核酸分子,或其互补物;和

[0102] b) 包括与SEQ ID NO:5中阐释的核苷酸序列具有至少95%序列同一性的核苷酸序列的核酸分子,或其互补物。

[0103] 在某些实施方式中,圆果种黄麻的植物是品种CVL-1。

[0104] 本发明的另一方面涉及分离的同源框-亮氨酸拉链HAT22多肽,其包括SEQ ID NO:3中阐释的氨基酸序列,或其生物活性片段,所述多肽包括选自下述的一个或多个功能:催化长蒴黄麻的植物中启动、形成、改善、增强,和改变从而修饰纤维长度、植物高度、生物量或其任何组合。

[0105] 在某些实施方式中,长蒴黄麻的植物是品种0-4。

[0106] 在某些实施方式中,所述多肽包括与SEQ ID NO:3中阐释的氨基酸序列至少95%的序列同一性。

[0107] 本发明的另一方面涉及分离的同源框-亮氨酸拉链HAT22蛋白多肽,其包括SEQ ID NO:6中阐释的氨基酸序列,或其生物活性片段,所述多肽包括选自下述的一个或多个功能:催化圆果种黄麻的植物中启动、形成、改善、增强和改变从而修饰纤维长度、植物高度、生物量,或其任何组合。

[0108] 在某些实施方式中,圆果种黄麻的植物是品种CVL-1。

[0109] 在某些实施方式中,所述多肽包括与SEQ ID NO:6中阐释的氨基酸序列至少95%的序列同一性。

[0110] 本发明的另一方面涉及重组基因构建体,其包括选自下述的多核苷酸:

[0111] a) 包括SEQ ID NO:2或SEQ ID NO:5中阐释的核苷酸序列的核酸分子,或其互补物;和

[0112] b) 包括与SEQ ID NO:2或SEQ ID NO:5中阐释的核苷酸序列具有至少95%序列同一性的核苷酸序列的核酸分子,或其互补物;其中多核苷酸在宿主细胞中是可表达的以在长蒴黄麻的植物和圆果种黄麻中产生同源框-亮氨酸拉链HAT22多肽的同系物。

[0113] 在某些实施方式中,重组基因构建体进一步包括可操作地连接至 a) 或b) 中阐释的核酸分子的启动子区域,其中所述启动子增强多核苷酸的转录或表达。

[0114] 本发明的另一方面涉及转化株,其包括能够表达选自下述多核苷酸的重组基因构建体:

[0115] a) 包括SEQ ID NO:2或SEQ ID NO:5中阐释的核苷酸序列的核酸分子,或其互补物;和

[0116] b) 包括与SEQ ID NO:2或SEQ ID NO:5中阐释的核苷酸序列具有至少95%序列同一性的核苷酸序列的核酸分子,或其互补物;其中所述转化株产生同源框-亮氨酸拉链HAT22多肽的同系物。

[0117] 本发明的另一方面涉及产生包括选自下述一个或多个特征的植物或转基因植物的方法:相对于对照植物增加、改善或修饰纤维产率、纤维长度和植物高度,所述方法包括下述步骤:

[0118] a) 将重组基因构建体引入植物细胞,所述重组基因构建体包括选自下述的多核苷酸:i) 包括SEQ ID NO:2或SEQ ID NO:5中阐释的核苷酸序列的核酸分子,或其互补物;和ii) 包括与SEQ ID NO:2或SEQ ID NO:5中阐释的核苷酸序列具有至少95%序列同一性的核苷酸序列的核酸分子,或其互补物;和

[0119] b) 表达选自下述的多肽:i) 包括SEQ ID NO:3或SEQ ID NO:6 中阐释的氨基酸序列的多肽,或其生物活性片段;和ii) 与SEQ ID NO:3或SEQ ID NO:6中阐释的氨基酸序列具有至少95%序列同一性的多肽;和

[0120] c) 在促进植物生长和发育的条件下培养植物细胞。

[0121] 本发明提供的序列可也用作合理修饰或设计新酶的预备材料,所述新酶具有确保酶在要求过程中表现更好的特征。

[0122] 本公开包括如所附的权利要求书中包含的,以及前述说明书中包含的。尽管已经以其一定程度特殊的优选形式描述了本发明,但是应理解,优选形式的公开仅仅作为例子并且可采用部件的构造和组合以及布置细节的许多改变,而不背离本发明和权利要求的范围。

[0123] 本文引用了许多参考文献,其内容通过参考以它们的整体并入。

实施例

[0124] 下述实施例旨在进一步阐释本发明,而决不将本发明限于其中描述的具体的实施方式。

[0125] 实施例1引物的设计和合成

[0126] 在研究中使用的引物是手动精选的转录组设计并且本发明不限于长蒴黄麻和圆果种黄麻的具体的实施方式,通过手动选择具有完全 ORF的序列或使用类似的基因已经成功从其他植物分离的数据库。使用NCBI BLAST、BLASTP、RPS-BLAST、BLASTX和PSI-BLAST进行获得自转录组的核苷酸序列的比较生物信息学分析,以鉴定相关基因的同系物并且用于基因的适当鉴定。当多个序列出现在“基因库”中时,通过clustalW版本1.82进行核苷酸序列比对。然后编辑比对。手动选择基因特异性引物(正向和反向)或通过Primer 3plus工具并且定制合成引物。

[0127] 合成在该研究中使用的所有寡核苷酸并且通过供应商HPLC纯化以及从整合DNA技术(IDT)获得。在高压灭菌ddH₂O中制备约100 pmol的原料液并且储存在约-20℃,以等分试样待用。

[0128] 用作PCR引物的寡核苷酸序列

[0129]

引物名	SEQ ID NO	寡核苷酸序列	来自 cDNA 的扩增产品
COLHLZ F	1	TCTTGCTCGCTACCAACATT	883
COLHLZ R	1	GCACTGTTCCTATGGCCTCA	
CCAHLZ F	4	ACCCATTTCCTTGTTTCCTCC TT	926
CCAHLZ R	4	GCACTGTTCCTATGGCCTCAT	

[0130] 实施例2来自长蒴黄麻和圆果种黄麻的同源框-亮氨酸拉链 HAT22的扩增、克隆和测序

[0131] 从MS培养基上生长的三天龄秧苗分离总RNA,如先前 Chomczynski P和Sacchi N描述,通过酸胍盐硫氰酸盐-苯酚-氯仿提取的RNA分离的单步方法。(Anal Biochem 1987, 162:156-159)。通过琼脂糖凝胶电泳检查RNA的质量或完整性并且使用Thermo Scientific Nano Drop 2000根据标准程序来量化。使用SuperScript III逆转录酶 (Invitrogen) 按照制造商的指导合成cDNA第一链。通过PCR使用基因特异性引物从cDNA扩增基因。PCR反应 (50 μ L) 包含1 μ L的cDNA, 20pmoles的每个引物, 5 μ L的10X PCR缓冲液, 5 μ L的2.5mM dNTP 混合物和1.0单位的PfuTaq DNA聚合酶。在Thermal Cycler (Applied Biosystems) 中使用下述条件进行PCR: 在约95 $^{\circ}$ C下初始变性约5min 随后35个下述循环: 在约95 $^{\circ}$ C下变性约30sec、在约58 $^{\circ}$ C下退火约 30sec和在约72 $^{\circ}$ C下延伸约1min, 在约72 $^{\circ}$ C下最终延伸约7min。通过1%琼脂糖凝胶使用1X TAE缓冲液分析PCR产品并且从凝胶使用 QIAGEN凝胶提取试剂盒根据制造商的指导洗脱扩增子。纯化的PCR 产品连接至pCR[®]8/GW/TOPO[®] TA克隆试剂盒 (Invitrogen) 和转化至感受态大肠杆菌细胞 (Invitrogen)。从推定克隆使用QIAprep Spin Miniprep 试剂盒 (QIAGEN) 按照制造商指导分离质粒, 在MS培养基上生长的龄秧苗, 如先前 Choe描述, 对基因特异性引物和阳性质粒进行测序。

[0132] 实施例3序列的分析

[0133] 分别通过BLASTN和BLASTP程序分析核苷酸序列和氨基酸序列。从其他植物报道的序列用ClustalW比对。使用Neighbour Joining (NJ) 进行系统发生分析。

[0134] 实施例4纤维生物合成的途径构建

[0135] 显示同源框-亮氨酸拉链HAT22在纤维生物合成中作用的自动代谢途径重建通过鉴定来自长蒴黄麻和圆果种黄麻的直向同源基因构建, 预测的蛋白质与Arabidopsis基因组比较。使用酶反应构建长蒴黄麻和圆果种黄麻基因组中编码的同源框-亮氨酸拉链HAT 22催化酶反应, 所述酶反应在用于途径工作室的Resnet-Plant 3.0数据库以及来自代谢途径数据库可用。

[0136] 通过参考并入

[0137] 本文引用的所有的美国专利、美国公开专利申请, 和公开的PCT 申请通过参考并

入本文。

[0138] 等同原则

[0139] 尽管本文已经描述和阐释了本发明的数个实施方式,本领域技术人员容易想到进行功能和/或获得本文所述的结果和/或一种或多种优势的各种其他方式和/或结构,并且每个这样的变型和/或修饰视为在本发明的范围内。本领域技术人员不使用过多的常规实验认识到,或能够确定本文所述的本发明具体实施方式的许多等价方式。所以,应理解前述实施方式仅作为例子呈现,并且在所附的权利要求和其等价方式的范围内;本发明可以如具体描述和要求的其他方式实践。

[0140] SEQ ID NO:1

[0141] 长度:1296bp

[0142] 类型:DNA

[0143] 特征名/关键字:内含子,外显子包括150bp 5'UTR和150bp 3' UTR

[0144] 生物体:黄麻,长蒴黄麻

[0145] CACCCCCCTAATTAACACTCCTTCCTCTCATTTCTCTCTAAGAAT TTTTTTTAAAAAAATTACA
ACCCTTTTCCTTGTTTCCTCCTTAG CATCAACATTTTCTTGCTCGCTACCAACATTTTCTCAGGTTTGAAG TTT
GAAACTACCATGGGTTTTGATGATATTTGTAACACTGGCCTTG GTCTTGGAAGTGGCTGCCTTGTGAAGCAAGAA
AAGTATTCGCAG TCTGATCATCAGCAGCAAAAGAAGAAGAAATTGTTCTTGAAAGA TGATCAGTTTTTCCCATC
TCTTACTTTAGGTCCATCAGATGACATAT ACCAATTATCTCCTGCTAATAAAATTCATGGGGAATCAACTGATTT
GCACCAACAACAACAAGCCTCTTCTATAAGTGGAGTATCTTCATT TTCTAACTCTAGTGTCAAGAAGGAGAAAGA
TATTTGTGGTGAAGA GGTAGAATTGGAAAGAATGTCTTCAAGGGTCAGTGATGAAGATG ATGAAGGCAGCCCCA
GGAAGAAACTTCGGCTTAGCAAAGAACA AGCTGCCATTTTAGAAGACAGCTTCAAAGAACACAGCACTCTCA AT
CCTGTAAATCATCTTCCCTTGTCCTTTTAATATGTTATTTTCATGTT TTATATACAATTTTATTGCATATAATCTT
AATATTTACTTGTTTCTATT GTTAACAGAAACAGAAGCAGGCTTTAGCAGAACATCTGAATCTT AGGCCACGGC
AAGTGGAAGTATGGTTCCAAAACAGGAGAGCCA GGTAAATTTTCCTTTCAATTTTGACCCTATATTCATAGCTTG
GCCTT ATATGAGTTTATTCCTTACTTTGGATTCTCTTTTGAACAACAG GACAAAGCTGAAACAGACTGAAG
TAGATTGTGAGTTACTGAAGA AATGTTGTGAGACATTAACAGAAGAGAACAACGGTTGCAGAAG GAACTGCAG
GAGCTTAAATCATTGAAATTGACAGCATCTTATTATA TGCAGCTGCCGGCGGCCACCCTCACCATGTGCCCTTCC
TGTGAG AGGGTTGCCGGCGCCGGTGAAGGTCCTTCCAGTACAAGCCCTTT TACAATTGGACAGAAATCTCACTT
CTTCAACCCCTTCACTCACCC ATCTGCAGCATGCTAGGCAACTTGTTGATGAGGCCATAGGAACAG TGCTATAT
TTTTGTTCTAGGCTTTATTTATGTATGTCAAAATTATATA TATGATCAGTATGGAGTAATTAAATCTTAAAGTAT
AAGTAGGCTAA TATCTTTATATTTCTTGTATAAATTAA

[0146] SEQ ID NO:2

[0147] 长度:804bp

[0148] 类型:DNA

[0149] 生物体:黄麻,长蒴黄麻

[0150] 特征名/关键字:CDS

[0151] 位置:(1) ... (804)

[0152] ATGGGTTTTGATGATATTTGTAACACTGGCCTTGGTCTTGGACTAG GCTGCCTTGTGAAGCAAGAAA
AGTATTCGCAGTCTGATCATCAGC AGCAAAAGAAGAAGAAATTGTTCTTGAAAGATGATCAGTTTTTC CCATCT

CTTACTTTAGGTCCATCAGATGACATATACCAATTATCTCC TGCTAATAAAATTCATGGGGAATCAACTGATTG
CACCAACAACA ACAAGCCTCTTCTATAAGTGAGTATCTTCATTTTCTAACTCTAGT GTCAAGAAGGAGAAAGA
TATTTGTGGTGAAGAGGTAGAATTGGA AAGAATGTCTTCAAGGGTCAGTGATGAAGATGATGAAGGCAGCC CCA
GGAAGAAACTTCGGCTTAGCAAAGAACAAGCTGCCATTTTA GAAGACAGCTTCAAAGAACACAGCACTCTCAATC
CTAAACAGAA GCAGGCTTTAGCAGAACATCTGAATCTTAGGCCACGGCAAGTGG AAGTATGGTTCCAAAACAGG
AGAGCCAGGACAAAGCTGAAACA GACTGAAGTAGATTGTGAGTTACTGAAGAAATGTTGTGAGACATT AACAGA
AGAGAACAAACGGTTGCAGAAGGAAGTGCAGGAGCTT AAATCATTGAAATTGACAGCATCTTATTATATGCAGCT
GCCGGCGG CCACCCTCACCATGTGCCCTTCCTGTGAGAGGGTTGCCGGCGCC GGTGAAGGTCCTTCCAGTACAA
GCCCTTTTACAATTGGACAGAAA TCTCACTTCTTCAACCCCTTCACTCACCCATCTGCAGCATGCTAG

[0153] SEQ ID NO:3

[0154] 长度:267

[0155] 类型:PRT

[0156] 生物体:黄麻,长蒴黄麻

[0157] MGFDDICNTGLGLGLCLVKQEKYSQSDHQQKKKKLFLKDDQFF PSLTLGPSDDIYQLSPANKIHG
ESTDLHQQQQASSISGVSSFNSSVK KEKDICGEEVELERMSSRVSDDEDDEGSPRKKLRLSKEQAAILEDSEFK E
HSTLNPKQKQALAEHLNLRPRQVEVWFQNRRTKLKQTEVDCE LLKKCCETLTEENKRLQKELQELKSLKLTAS
YYMQLPAATLTMCPSCERVAGAGEGPSSTSPFTIGQKSHFFNPFTHPAAC

[0158] SEQ ID NO:4

[0159] 长度:1337bp

[0160] 类型:DNA

[0161] 特征名/关键字:内含子,外显子包括150bp 5'UTR和150bp 3' UTR

[0162] 生物体:黄麻,圆果种黄麻

[0163] CTCAAACAAGTTCCTAATTAACACTCCTTCCTCTCATTTCTCTCT AAGAAAAAAAAAATACAACCC
ATTTCTTGTTTCCTCCTTAGCAT CAAGATTAATTTTCTTGCTAGCTACCAACATTTTCTAGGTTTGAA GTTT
GAAACTACCATGGGTTTTGATGATATTTGTAACACTGGCCTT AATCTTGGACTCGGATGCCTTGTGAAGCAAGAA
AAGTATTCGCCG TCCGATCATCAGCAGCAAAAGAAGAAGAAATTGTTGTTGAAAGA TGATCAGTTTTTCCCATC
TCTTACTTTAGGTCCATCAGATGACATAT ACCAATTATCTGCTGCTAATAAATTAGCTCATGGGGAATCAATTGA
TTTGCACCAACAACAAGCCTCTTCTATCAGTGGAGTATCTTC ATTTTCTAACTCTAGTGTCAAGAAGGAGAA
AGATATTTGTGGTGA AGAGGTAGAATTGGAGAGAATGTCTTCAAGGGTCAGTGATGAAG ATGATGAAGGCAGCC
CCAGGAAGAACTTCGGCTTAGCAAAGAA CAAGCTGCCATTTTAGAAGACAGCTTCAAAGAACACAGCACTCT C
AATCCTGTAAATCATCTTCTAATTAATTCCTTGTCTTTTTCTATAT CATTTTCATGTTTTATATTAATTTTATTG
CACATAATCTTAATATTTAC TTGTTTCTCATGTTAACAGAAACAGAAGCAGGCTTAGCAGAACA TCTGAATCT
TAGGCCACGGCAAGTGGAAGTATGGTTCCAAAACA GGAGAGCCAGGTAAATTTTCCTTTCAATTTTGACCATATA
TTATATA CATAGCTTGGCCTTGATGATTTTATATTCTCTTACTTTGGATTC TCTTTTTGAAACAACAGGA
CAAAGCTGAAACAGACTGAAGTAGA TTGTGAGTTACTGAAGAAATGTTGTGAGACATTAACAGAAGAGA AAAAA
CGGTTGCAGAAGGAAGTGCAGGAGCTTAAATCATTGAAA TTGACAACATCTTATTATATGCAGCTGCCGGCTGCC
ACCCTCACCA TGTGCCCTTCCTGTGAGAGGGTTGCCGGCACCGGTGAAGGCCCT TCCAGTACAAGCCCTTTTAC
AATTGGACAGAAATCTCACTTCTTC AACCCCTTCACTCACCCATCAGCCGCATGCTAGGCAACTTGTTGA TGAG
GCCATAGGAACAGTGCTATTTTGTCTAGGCTTTTTATTTA TGTCAAATATATATATGATCAGTATGGAGTAAT

TAAATCCTGAAGTA TTATAAGTAGGCTAATATCTTTATATTTCTTGTATAAATTAAGAT

[0164] SEQ ID NO:5

[0165] 长度:807

[0166] 类型:DNA

[0167] 生物体:黄麻,圆果种黄麻

[0168] 特征名/关键字:CDS

[0169] 位置:(1) ... (807)

[0170] ATGGGTTTTGATGATATTTGTAACACTGGCCTTAATCTTGGACTCG GATGCCTTGTGAAGCAAGAAA
AGTATTCCCGTCCGATCATCAGC AGCAAAAGAAGAAGAAATTGTTGTTGAAAGATGATCAGTTTTTC CCATCT
CTTACTTTAGGTCCATCAGATGACATATACCAATTATCTGC TGCTAATAAATTAGCTCATGGGGAATCAATTGAT
TTGCACCAACA ACAACAAGCCTCTTCTATCAGTGGAGTATCTTCATTTTCTAACTCT AGTGTCAAGAAGGAGAA
AGATATTTGTGGTGAAGAGGTAGAATT GGAGAGAATGTCTTCAAGGGTCAGTGATGAAGATGATGAAGGCA GCC
CCAGGAAGAACTTCGGCTTAGCAAAGAACAAGCTGCCATT TTAGAAGACAGCTTCAAAGAACACAGCACTCTCA
ATCCTAAACA GAAGCAGGCTTTAGCAGAACATCTGAATCTTAGGCCACGGCAAG TGGAAGTATGGTTCCAAAAC
AGGAGAGCCAGGACAAAGCTGAA ACAGACTGAAGTAGATTGTGAGTTACTGAAGAAATGTTGTGAGA CATTAAAC
AGAAGAGAACAACGGTTGCAGAAGGAAGTGCAGGA GCTTAAATCATTGAAATTGACAACATCTTATTATATGCA
GCTGCCG GCTGCCACCCTCACCATGTGCCCTTCCTGTGAGAGGGTTGCCGGC ACCGGTGAAGGCCCTTCCAGTA
CAAGCCCTTTTACAATTGGACA GAAATCTCACTTCTTCAACCCCTTCACTCACCCATCAGCCGCATG CTAG

[0171] SEQ ID NO:6

[0172] 长度:268

[0173] 类型:PRT

[0174] 生物体:黄麻,圆果种黄麻

[0175] MGFDDICNTGLNLGLCLVKQEKYSPSDHQQQKKKKLLKDDQFF PSLTLGPSDDIYQLSAANKLAH
GESIDLHQQQQASSISGVSSFSNSSV KKEKDICGEEVELERMSSRVSDDEDEGSPRKKLRLSKEQAILEDSEF K
EHSTLNPQKQKQALAEHLNLRPRQVEVWFQNRRTARTKLKQTEVDC ELLKKCCETLTEENKRLQKELQELKSLKLT
SYMQLPAAATLTMCPSCERVAGTGEGPSSTSPFTIGQKSHFFNPFTHPSAAC

[0001]

序列表

<110> 孟加拉国黄麻研究所

<120> 编码来自长蒴黄麻和圆果种黄麻的同源框-亮氨酸拉链蛋白HAT22 (HD-ZIP蛋白22)的核苷酸序列和使用方法

<130> JGX-009.25

<140> PCT/US2014/066772

<141> 2014-11-21

<150> 61/908,444

<151> 2013-11-25

<160> 10

<170> PatentIn version 3.5

<210> 1

<211> 1296

<212> DNA

<213> 长蒴黄麻

<400> 1

caccccccta attaacactc cttcctctca tttcctctct aagaatTTTT ttttaaaaaa	60
aattacaacc cttttccttg tttcctcctt agcatcaaca ttttcttgct cgctaccaac	120
atTTTctcag gtttgaagtt tgaaactacc atgggttttg atgatatttg taacactggc	180
cttggtcttg gactaggctg cttgtgaag caagaaaagt attcgcagtc tgatcatcag	240
cagcaaaaga agaagaaatt gttcttgaaa gatgatcagt ttttcccatc tcttacttta	300
ggTccatcag atgacatata ccaattatct cctgctaata aaattcatgg ggaatcaact	360
gatttgcacc aacaacaaca agcctcttct ataagtggag tatcttcatt ttctaactct	420
agtgtcaaga aggagaaaga tatttgtggt gaagaggtag aattggaaag aatgtcttca	480
agggtcagtg atgaagatga tgaaggcagc cccaggaaga aacttcggct tagcaaagaa	540
caagctgccA ttttagaaga cagcttcaaa gaacacagca ctctcaatcc tgtaaAcat	600
cttcccttgt ccttttaata tgttatttca tgTTTTatat acaatTTtat tgcatataat	660
cttaatatTT acttgTTtct attgttaaca gaaacagaag caggctttag cagaacatct	720
gaatcttagg ccacggcaag tggaagtatg gttccaaaac aggagagcca ggTTaatttt	780
cctttcaatt ttgaccctat attcatagct tggccttata tgagtttatt cccttacttt	840
ggattctctt tttgaaacaa caggacaaag ctgaaacaga ctgaagtaga ttgtgagtta	900
ctgaagaaat gttgtgagac attaacagaa gagaacaaac ggttgcagaa ggaactgcag	960

[0002]

gagcttaaatt cattgaaatt gacagcatct tattatatgc agctgccggc ggccaccctc	1020
accatgtgcc cttcctgtga gagggttgcc ggcgccggtg aaggtccttc cagtacaagc	1080
ccttttaciaa ttggacagaa atctcacttc ttcaaccctt tcactcacc c atctgcagca	1140
tgctaggcaa cttgttgatg aggccatagg aacagtgcta tatttttggt ctaggcttta	1200
tttatgtatg tcaaaattat atatatgatc agtatggagt aattaaatct taaagtataa	1260
gtaggctaatt atcctttatat ttcttgtata aattaa	1296

<210> 2

$\langle 211 \rangle$ 804

<212> DNA

〈213〉 长蒴黄麻

 $\langle 400 \rangle$ 2

atggggttttg	atgatatattg	taacactggc	cttgggtcttg	gactaggcctg	ccttgtgaag	60
caagaaaagt	attcgcagtc	tgatcatcag	cagcaaaaga	agaagaaatt	gttcttgaag	120
gatgatcagt	ttttcccatc	tcttacttta	gggccatcag	atgacatata	ccaattatct	180
cctgctaata	aaattcatgg	ggaatcaact	gatttgcacc	aacaacaaca	agcctcttct	240
ataagtggag	tatcttcatt	ttctaactct	agtgtcaaga	aggagaaaga	tatttgtggt	300
gaagaggtag	aattggaaag	aatgtcttca	agggtcagtg	atgaagatga	tgaaggcagc	360
cccaggaaga	aacttcggct	tagcaaagaa	caagctgcca	ttttagaaga	cagcttcaaa	420
gaacacagca	ctctcaatcc	taaacagaag	caggcttttag	cagaacatct	gaatcttagg	480
ccacggcaag	tggaagtatg	gttccaaaac	aggagagcca	ggacaaagct	gaaacagact	540
gaagtagatt	gtgagttact	gaagaaatgt	tgtgagacat	taacagaaga	gaacaaacgg	600
ttgcagaagg	aactgcagga	gcttaaataca	ttgaaattga	cagcatctta	ttatatgcag	660
ctgccggcgg	ccaccctcac	catgtgccct	tcctgtgaga	gggttgccgg	cgccgggtgaa	720
gttccttcca	gtacaagccc	ttttacaatt	ggacagaaat	ctcacttctt	caacccttcc	780
actcacccat	ctgcagcatg	ctag				804

 $\langle 210 \rangle$ 3

$\langle 211 \rangle$ 267

<212> PRT

〈213〉 长蒴黄麻

<400> 3

Met Gly Phe Asp Asp Ile Cys Asn Thr Gly Leu Gly Leu Gly Leu Gly
1 5 10 15

Cys Leu Val Lys Gln Glu Lys Tyr Ser Gln Ser Asp His Gln Gln Gln
20 25 30

Lys Lys Lys Lys Leu Phe Leu Lys Asp Asp Gln Phe Phe Pro Ser Leu
35 40 45

Thr Leu Gly Pro Ser Asp Asp Ile Tyr Gln Leu Ser Pro Ala Asn Lys
50 55 60

Ile His Gly Glu Ser Thr Asp Leu His Gln Gln Gln Gln Ala Ser Ser
65 70 75 80

Ile Ser Gly Val Ser Ser Phe Ser Asn Ser Ser Val Lys Lys Glu Lys
85 90 95

Asp Ile Cys Gly Glu Glu Val Glu Leu Glu Arg Met Ser Ser Arg Val
100 105 110

Ser Asp Glu Asp Asp Glu Gly Ser Pro Arg Lys Lys Leu Arg Leu Ser
115 120 125

[0003]

Lys Glu Gln Ala Ala Ile Leu Glu Asp Ser Phe Lys Glu His Ser Thr
130 135 140

Leu Asn Pro Lys Gln Lys Gln Ala Leu Ala Glu His Leu Asn Leu Arg
145 150 155 160

Pro Arg Gln Val Glu Val Trp Phe Gln Asn Arg Arg Ala Arg Thr Lys
165 170 175

Leu Lys Gln Thr Glu Val Asp Cys Glu Leu Leu Lys Lys Cys Cys Glu
180 185 190

Thr Leu Thr Glu Glu Asn Lys Arg Leu Gln Lys Glu Leu Gln Glu Leu
195 200 205

Lys Ser Leu Lys Leu Thr Ala Ser Tyr Tyr Met Gln Leu Pro Ala Ala
210 215 220

Thr Leu Thr Met Cys Pro Ser Cys Glu Arg Val Ala Gly Ala Gly Glu
225 230 235 240

[0004]

Gly Pro Ser Ser Thr Ser Pro Phe Thr Ile Gly Gln Lys Ser His Phe
 245 250 255

Phe Asn Pro Phe Thr His Pro Ser Ala Ala Cys
 260 265

<210> 4

<211> 1317

<212> DNA

<213> 圆果种黄麻

<400> 4

```

ctcaaacaag ttctaatta acactccttc ctctcatttc ctctctaaga aaaaaaaaaat      60
acaacccatt tccttgtttc ctcccttagca tcaagattaa ttttcttgct agctaccaac      120
atctttctag gtttgaagtt tgaaactacc atgggttttg atgatatttg taacactggc      180
cttaatcttg gactcggatg ccttggtgaag caagaaaagt attcgccgtc cgatcatcag      240
cagcaaaaga agaagaaatt gttgttgaaa gatgatcagt ttttcccatc tcttacttta      300
gggccatcag atgacatata ccaattatct gctgctaata aattagetca tggggaatca      360
attgatttgc accaacaaca acaagcctct tctatcagtg gagtatcttc attttctaac      420
tctagtgtca agaaggagaa agatatttgt ggtgaagagg tagaattgga gagaatgtct      480
tcaagggtca gtgatgaaga tgatgaaggc agccccagga agaaacttcg gcttagcaaa      540
gaacaagctg ccattttaga agacagcttc aaagaacaca gcactctcaa tcctgtaaat      600
catcttctaa ttaattccct tgtctttttc tatatcattt catgttttat attaatttta      660
ttgcacataa tcttaatat tacttgtttc tcatgttaac agaaacagaa gcaggcttta      720
gcagaacatc tgaatcttag gccacggcaa gtggaagtat ggttccaaaa caggagagcc      780
aggttaatth tcctttcaat ttgaccata tattatatac atagcttggc cttgtatgat      840
tttatattat tctcttactt tggattctct ttttgaaaca acaggacaaa gctgaaacag      900
actgaagtag attgtgagtt actgaagaaa tgttggtgaga cattaacaga agagaacaaa      960
cggttgcaga aggaactgca ggagcttaaa tcattgaaat tgacaacatc ttatttatatg     1020
cagctgccgg ctgccaccct caccatgtgc cttcctgtg agagggttgc cggcaccggt     1080
gaaggccctt ccagtacaag cccttttaca attggacaga aatctcactt cttcaacccc     1140
ttcactcacc catcagccgc atgctaggca acttggtgat gaggccatag gaacagtget     1200
atctttgttc taggtttttt atttatgtca aatatatata tgatcagtat ggagtaatta     1260
aatcctgaag tattataagt aggctaatat ctttatatth cttgtataaa ttaagat      1317

```

<210> 5
 <211> 807
 <212> DNA
 <213> 圆果种黄麻

<400> 5	
atggggttttg atgatatttg taacactggc cttaatcttg gactcggatg ccttgtgaag	60
caagaaaagt attcgccgtc cgatcatcag cagcaaaaga agaagaaatt gttgttgaaa	120
gatgatcagt ttttcccatc tcttacttta ggtccatcag atgacatata ccaattatct	180
gctgctaata aattagctca tggggaatca attgatttgc accaacaaca acaagcctct	240
tctatcagtg gagtatcttc attttctaac tctagtgtca agaaggagaa agatatttgt	300
ggtgaagagg tagaattgga gagaatgtct tcaagggtca gtgatgaaga tgatgaaggc	360
agccccagga agaaaacttcg gcttagcaaa gaacaagctg ccattttaga agacagcttc	420
aaagaacaca gcactctcaa tcctaaacag aagcaggctt tagcagaaca tctgaatctt	480
aggccacggc aagtggaagt atggttccaa aacaggagag ccaggacaaa gctgaaacag	540
actgaagtag attgtgagtt actgaagaaa tgttgtgaga cattaacaga agagaacaaa	600
cggttgcaga aggaactgca ggagcttaaa tcattgaaat tgacaacatc ttattatatg	660
cagctgccgg ctgccaccct caccatgtgc ctttcctgtg agagggttgc cggcaccggt	720
gaaggccctt ccagtacaag cccttttaca attggacaga aatctcactt cttcaacccc	780
ttcactcacc catcagccgc atgctag	807

<210> 6
 <211> 268
 <212> PRT
 <213> 圆果种黄麻

<400> 6
Met Gly Phe Asp Asp Ile Cys Asn Thr Gly Leu Asn Leu Gly Leu Gly
1 5 10 15
Cys Leu Val Lys Gln Glu Lys Tyr Ser Pro Ser Asp His Gln Gln Gln
20 25 30
Lys Lys Lys Lys Leu Leu Leu Lys Asp Asp Gln Phe Phe Pro Ser Leu
35 40 45
Thr Leu Gly Pro Ser Asp Asp Ile Tyr Gln Leu Ser Ala Ala Asn Lys
50 55 60

Leu Ala His Gly Glu Ser Ile Asp Leu His Gln Gln Gln Gln Ala Ser
65 70 75 80

Ser Ile Ser Gly Val Ser Ser Phe Ser Asn Ser Ser Val Lys Lys Glu
85 90 95

Lys Asp Ile Cys Gly Glu Glu Val Glu Leu Glu Arg Met Ser Ser Arg
100 105 110

Val Ser Asp Glu Asp Asp Glu Gly Ser Pro Arg Lys Lys Leu Arg Leu
115 120 125

Ser Lys Glu Gln Ala Ala Ile Leu Glu Asp Ser Phe Lys Glu His Ser
130 135 140

Thr Leu Asn Pro Lys Gln Lys Gln Ala Leu Ala Glu His Leu Asn Leu
145 150 155 160

Arg Pro Arg Gln Val Glu Val Trp Phe Gln Asn Arg Arg Ala Arg Thr
165 170 175

[0006]

Lys Leu Lys Gln Thr Glu Val Asp Cys Glu Leu Leu Lys Lys Cys Cys
180 185 190

Glu Thr Leu Thr Glu Glu Asn Lys Arg Leu Gln Lys Glu Leu Gln Glu
195 200 205

Leu Lys Ser Leu Lys Leu Thr Thr Ser Tyr Tyr Met Gln Leu Pro Ala
210 215 220

Ala Thr Leu Thr Met Cys Pro Ser Cys Glu Arg Val Ala Gly Thr Gly
225 230 235 240

Glu Gly Pro Ser Ser Thr Ser Pro Phe Thr Ile Gly Gln Lys Ser His
245 250 255

Phe Phe Asn Pro Phe Thr His Pro Ser Ala Ala Cys
260 265

<210> 7
<211> 20
<212> DNA

[0007]

<213> 人工序列

<220>

<223> 人工序列的描述：合成启动子

<400> 7

tcttgctcgc taccaacatt

20

<210> 8

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成启动子

<400> 8

gcactgttcc tatggcctca

20

<210> 9

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成启动子

<400> 9

acccatttcc ttgtttcctc ctt

23

<210> 10

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成启动子

<400> 10

gcactgttcc tatggcctca t

21

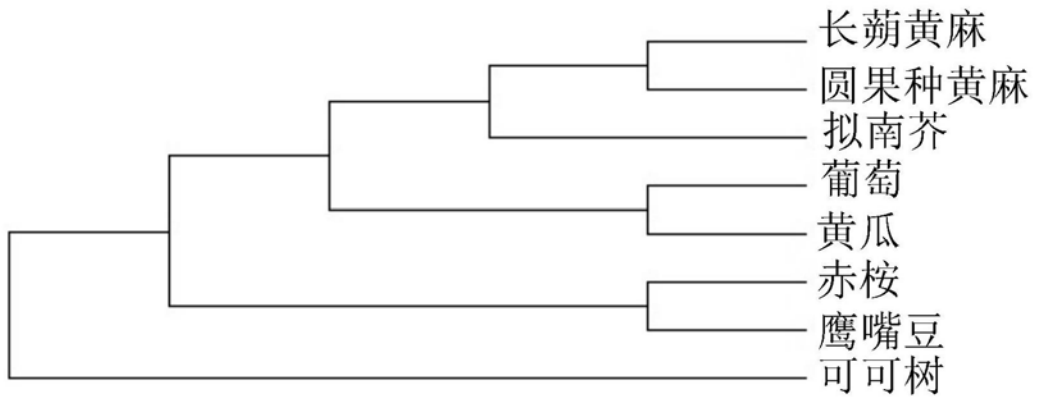


图1