



(12)发明专利申请

(10)申请公布号 CN 108291234 A

(43)申请公布日 2018.07.17

(21)申请号 201680064830.8

(74)专利代理机构 永新专利商标代理有限公司
72002

(22)申请日 2016.09.05

代理人 区斌

(30)优先权数据

2015398 2015.09.04 NL

(51)Int.Cl.

C12N 15/82(2006.01)

(85)PCT国际申请进入国家阶段日

2018.05.04

C12Q 1/6895(2018.01)

(86)PCT国际申请的申请数据

PCT/NL2016/050617 2016.09.05

(87)PCT国际申请的公布数据

W02017/039452 EN 2017.03.09

(71)申请人 主基因有限公司

地址 荷兰瓦格宁根

(72)发明人 P·J·范戴克 D·里格拉

M·W·普林森 A·J·范杜能

权利要求书4页 说明书42页

序列表34页 附图2页

(54)发明名称

倍数孢子体形成基因

(57)摘要

本发明提供了Dip基因的核苷酸序列和氨基酸序列以及其(功能性)同源物、片段和变体,其提供了倍数孢子体形成作为无融合生殖的一部分。还提供了倍数孢子体形成植物及其制备方法以及使用它们的方法,和制备无融合生殖种子的方法。

1. 分离的多核苷酸,其包含SEQ ID NO:1的核酸序列或与SEQ ID NO:1的核酸序列具有至少50%或70%,优选至少80%,更优选至少90%,甚至更优选至少95%,甚至更优选至少96%或97%,最优选至少98%或99%序列同一性的核酸序列。

2. 分离的多核苷酸,其包含SEQ ID NO:2的核酸序列或与SEQ ID NO:2的核酸序列具有至少50%或70%,优选至少80%,更优选至少90%,甚至更优选至少95%,甚至更优选至少96%或97%,最优选至少98%或99%序列同一性的核酸序列。

3. 根据权利要求1-2中任一项所述的多核苷酸,其中所述多核苷酸包含或具有SEQ ID NO:1或SEQ ID NO:2的核酸序列。

4. 根据权利要求1-3中任一项所述的多核苷酸,其中所述多核苷酸和/或所述多核苷酸的表达产物和/或由所述多核苷酸编码的蛋白质能够向植物或植物细胞提供倍数孢子体形成功能。

5. 根据权利要求4所述的多核苷酸,其中当引入进植物或植物细胞时,所述多核苷酸和/或所述多核苷酸的表达产物和/或由所述多核苷酸编码的蛋白质向植物或植物细胞提供倍数孢子体形成功能。

6. 根据权利要求4-5中任一项所述的多核苷酸,其中所述表达产物是RNA分子,优选为mRNA分子、siRNA或miRNA分子。

7. 根据权利要求1-3中任一项的多核苷酸的片段,其中所述片段和/或所述片段的表达产物和/或由所述片段编码的蛋白质能够向植物或植物细胞提供倍数孢子体形成功能。

8. 根据权利要求7所述的片段,其中当引入进植物或植物细胞时,所述片段和/或所述片段的表达产物和/或由所述片段编码的蛋白质向植物或植物细胞提供倍数孢子体形成功能。

9. 根据权利要求7-8中任一项所述的片段,其中所述表达产物是RNA分子,优选为mRNA分子、siRNA或miRNA分子。

10. 权利要求9所述的片段,其中所述表达产物包含或具有SEQ ID NO:5的核酸序列。

11. 根据权利要求7-8中任一项所述的片段,其中所述片段包含或具有SEQ ID NO:4或6的核酸序列。

12. 根据权利要求7-11中任一项所述的片段,其中所述片段的长度为所述多核苷酸的至少20、30、40、50、100、150、200、250、300、400、500、600、700、800、900、1000、1500、2000或3000个连续核苷酸。

13. 分离的多肽,其包含SEQ ID NO:3的氨基酸序列或与SEQ ID NO:3的氨基酸序列具有至少50%或70%,优选至少80%,更优选至少90%,甚至更优选至少95%,甚至更优选至少96%或97%,最优选至少98%或99%序列同一性的氨基酸序列。

14. 根据权利要求13所述的多肽,其中所述多肽包含或具有SEQ ID NO:3的氨基酸序列。

15. 根据权利要求13-14中任一项所述的多肽,其中所述多肽能够向植物或植物细胞提供倍数孢子体形成功能。

16. 根据权利要求15所述的多肽,其中所述多肽在引入进植物或植物细胞时向植物或植物细胞提供倍数孢子体形成功能。

17. 根据权利要求13-14中任一项所述的多肽的片段,其中所述片段能够向植物或植物

细胞提供倍数孢子体形成功能。

18. 根据权利要求17所述的片段,其中所述片段包含SEQ ID NO:7或SEQ ID NO:12的氨基酸序列。

19. 根据权利要求17-18中任一项所述的片段,其中所述片段的长度为所述多肽的至少10、20、30、40、50、100、150、200、250、300、400或500个连续氨基酸。

20. 一种嵌合基因,其包含根据权利要求1-6中任一项所述的多核苷酸或根据权利要求7-12中任一项所述的片段。

21. 一种遗传构建体,其包含根据权利要求1-6中任一项所述的多核苷酸、根据权利要求7-12中任一项所述的片段或根据权利要求20所述的嵌合基因。

22. 基因构建体,其在5'-3'方向上包含编码根据权利要求13-19中任一项的多肽或片段的开放读码框多核苷酸。

23. 一种核酸载体,其包含在植物细胞中有活性的启动子序列,所述启动子序列可操作地连接于根据权利要求1-6中任一项的多核苷酸、根据权利要求7-12中任一项的片段、根据权利要求20的嵌合基因或根据权利要求21-22中任一项的遗传构建体。

24. 权利要求23的核酸载体,其中所述启动子序列包含:

e) SEQ ID NO:2的核酸序列的天然启动子序列;

f) a) 的启动子序列的功能性片段;或者

g) 包含与SEQ ID NO:2的核酸序列的天然启动子序列具有至少70%,优选至少80%,更优选至少90%,最优选至少95%序列同一性的核酸序列;

h) SEQ ID NO:6的核酸序列的天然启动子序列;

i) d) 的启动子序列的功能性片段;或者

j) 包含与SEQ ID NO:6的核酸序列的天然启动子序列具有至少70%,优选至少80%,更优选至少90%,最优选至少95%序列同一性的核酸序列。

25. 权利要求23的核酸载体,其中所述启动子是雌性子房特异性启动子,优选大孢子母细胞特异性启动子和/或雌性配子特异性启动子。

26. 植物、植物部分或植物细胞,其包含权利要求20的嵌合基因、权利要求21-22中任一项的遗传构建体和/或权利要求23-25中任一项的核酸载体,其中所述基因、构建体和/或载体至少在雌性子房中存在和/或表达,优选在大孢子母细胞中和/或在雌性配子中表达。

27. 根据权利要求26所述的植物的种子,其中所述种子是所述植物的无融合生殖种子。

28. 根据权利要求27所述的种子,其中所述种子是产生该种子的植物的克隆。

29. 根据权利要求26-28中任一项所述的植物、植物部分、植物细胞或种子,其中所述植物、植物部分、植物细胞或种子来自选自以下的物种:蒲公英属、莴苣属、豌豆属、辣椒属、茄属、黄瓜属、玉蜀黍属、棉属、大豆属、小麦属、稻属、葱属、芸苔属、向日葵属、甜菜属、菊苣属、菊属、狼尾草属、黑麦属、大麦属、苜蓿属、菜豆属、蔷薇属、百合属、咖啡属、亚麻属、大麻属、木薯属、胡萝卜属、南瓜属、西瓜属和高粱属。

30. 一种生产无融合生殖种子的方法,其包括以下步骤:

a) 用根据权利要求1-6中任一项的多核苷酸、根据权利要求7-12中任一项的片段、根据权利要求20的嵌合基因、根据权利要求21-22中任一项的基因构建体和/或根据权利要求23-25中任一项的核酸载体转化植物、植物部分或植物细胞,以产生初级转化体;

b) 从所述初级转化体生长开花植物和/或花,其中所述多核苷酸、片段、基因、构建体和/或载体至少在雌性子房中存在和/或表达,优选在大孢子母细胞中和/或在雌性配子中存在和/或表达;和

c) 使所述初级转化体授粉以诱导产生种子,优选使用四倍体植物的花粉或所述初级转化体的自身花粉授粉。

31. 根据权利要求30所述的方法,其中所述植物、植物部分或植物细胞能够进行孤雌生殖。

32. 一种生产杂交植物的克隆的方法,其包括以下步骤:

a) 用根据权利要求26、29、37或42中任一项所述的植物的花粉使有性生殖的植物异花受精以产生F1杂种种子;

b) 选择F1植物,其至少在雌性子房,优选在大孢子母细胞和/或雌性配子中,包含和/或表达根据权利要求1-6中任一项所述的多核苷酸、根据权利要求7-12中任一项所述的片段或根据权利要求13-19中任一项所述的多肽;

c) 任选地,使所述选择的F1植物授粉以诱导产生种子,优选使用四倍体植物的花粉授粉;和

d) 收获种子;和

e) 任选地,从所述种子生长杂种克隆植物。

33. 根据权利要求32所述的方法,其中所述植物、植物部分或植物细胞能够进行孤雌生殖,并且其中所述克隆是无融合生殖克隆。

34. 杂种植物,其可通过根据权利要求32-33中任一项所述的方法而获得或通过根据权利要求32-33中任一项所述的方法而获得。

35. 一种赋予植物、植物部分或植物细胞倍数孢子体形成的方法,其包括以下步骤:

a) 用根据权利要求1-6中任一项所述的多核苷酸、根据权利要求7-12中任一项所述的片段、根据权利要求20所述的嵌合基因、根据权利要求21-22中任一项所述的基因构建体、和/或根据权利要求23-25中任一项所述的核酸载体转化所述植物、植物部分或植物细胞;和

b) 任选地,使植物再生,其中所述多核苷酸、片段、基因、构建体和/或载体至少在雌性子房中存在和/或表达,优选在大孢子母细胞中和/或在雌配子中存在和/或表达。

36. 根据权利要求35所述的方法,其中将根据权利要求1-6中任一项所述的多核苷酸或根据权利要求7-12中任一项所述的片段整合到所述植物、植物部分或植物细胞的基因组中。

37. 倍数孢子体形成植物、植物部分或植物细胞,其可通过根据权利要求35-36中任一项所述的方法而获得或通过根据权利要求35-36中任一项所述的方法而获得。

38. 一种赋予植物、植物部分或植物细胞倍数孢子体形成的方法,其包括以下步骤:

a) 修饰植物、植物部分或植物细胞中的内源性多核苷酸或多核苷酸片段,优选为液泡蛋白质分选相关蛋白质基因的多核苷酸或片段,使得经修饰后的植物、植物部分或植物细胞包含根据权利要求1-6中任一项所述的多核苷酸或根据权利要求7-12中任一项所述的片段;和

b) 任选地,再生植物。

39. 根据权利要求38所述的方法,其中所述经修饰的多核苷酸或多核苷酸的片段被表达和/或编码多肽。

40. 根据权利要求38-39中任一项所述的方法,其中所述经修饰的多核苷酸或多核苷酸的片段至少存在于雌性子房中,优选存在于大孢子母细胞中和/或雌性配子中。

41. 根据权利要求38所述的方法,其中所述修改是通过:

a) 在所述植物、植物部分或植物细胞中引入或表达至少一种位点特异性核酸酶,优选其中所述核酸酶选自:Cas/RNA CRISPR核酸酶、锌指核酸酶、大范围核酸酶和TAL效应子核酸酶;和/或通过

b) 使用寡核苷酸进行寡核苷酸定向诱变,优选其中所述寡核苷酸是单链寡核苷酸;和/或通过

c) 化学诱变,优选用甲磺酸乙酯。

42. 倍数孢子体形成植物,其可通过根据权利要求38-41中任一项所述的方法而获得或通过根据权利要求38-41中任一项所述的方法而获得。

43. 根据权利要求1-6中任一项所述的多核苷酸或根据权利要求7-12中任一项所述的多核苷酸的片段在植物中诱导倍数孢子体形成的用途。

44. 根据权利要求1-6中任一项所述的多核苷酸或根据权利要求7-12中任一项所述的多核苷酸的片段用于预防多个基因、QTL或转基因的分离的用途。

45. 根据权利要求1-6中任一项所述的多核苷酸或根据权利要求7-12中任一项所述的多核苷酸的片段用于基因堆叠的用途。

46. 根据权利要求1-6中任一项所述的多核苷酸或其片段用于发育和/或鉴定所述倍数孢子体形成性状的标记物的用途。

47. 根据权利要求1-6中任一项所述的多核苷酸或根据权利要求7-12中任一项所述的多核苷酸的片段用于在植物中诱导多倍体化的用途。

48. 根据权利要求1-6中任一项所述的多核苷酸或根据权利要求7-12中任一项所述的多核苷酸的片段用于在植物中诱导无融合生殖的用途。

倍数孢子体形成基因

技术领域

[0001] 本发明涉及生物技术领域,并具体涉及包括无性植物育种的植物生物技术。具体而言,本发明涉及鉴定与无融合生殖(特别是通过倍数孢子体形成(diplospory)的配子体无融合生殖)过程相关的基因、其变体或片段以及它们编码的蛋白质和肽。本发明还涉及使用本发明的基因、蛋白质、变体和片段在植物和农作物中通过倍数孢子体形成诱导配子体无融合生殖的方法,以及产生倍数孢子体形成植物和无融合生殖种子的方法。

背景技术

[0002] 在植物学中,无融合生殖(也称为不完全无配生殖)是指通过无性过程形成种子。无融合生殖通过一系列发育过程而发生,这些发育过程共同将植物的有性发育程序转化为无性发育程序。据报道,在400多种开花植物物种中存在复发性无融合生殖(Bicknell和Koltunow,2004)。无融合生殖可以以不同的形式出现,包括被称为配子体无融合生殖和孢子体无融合生殖(也称为不定胚形状(adventives embryony))的至少两种形式。其中配子体无融合生殖发生的植物实例包括蒲公英(*Taraxacum* sp.)、山柳菊(*Hieracium* sp.)、草地早熟禾(*Poa pratensis*)、鸭茅状摩擦禾(*Tripsacum dactyloides*)等。发生孢子体无融合生殖的植物的例子包括柑橘属(*Citrus* sp.)、山竹(*Garcinia mangostana*)等。

[0003] 对无融合生殖总体上的兴趣(尤其是对配子体无融合生殖的兴趣)由于它对农业(特别是为了克隆种子生产的)的潜在用途而在过去的几十年里增加了。配子体无融合生殖的特点是至少有两个发育过程:(1)避免减数分裂(不完全减数分裂);和(2)卵细胞发育成胚而不受精(孤雌生殖)。由配子体无融合生殖过程产生的种子被称为无融合生殖种子。

[0004] 由于无融合生殖种子在遗传上与母本植物相同,它们被认为是母本植物的克隆,并因此产生这种种子的过程被称为克隆种子生产。长期以来人们认识到无融合生殖在植物育种中非常有用(Asker 1979,Hermsen,JG Th.1980.Breeding for apomixis in potato: Pursuing a utopian scheme.*Euphytica* 29:595-607,Asker and Jerling 1990,DeVielle Calzada等,1995)。无融合生殖的优势是能够对具有杂种优势的F1杂种实现真实遗传(true breeding)(即具有统一遗传品质的F1杂种的无限增殖)。在大多数作物中,F1杂种是最好的品种,因为它们往往与较高的产量相关,这种现象通常称为“杂种优势”。由于F1杂种的自花授粉导致F2有性作物通过重组丧失杂种优势,F1代杂种的每代必须通过杂交近交纯合亲本再次产生。生产有性F1种子是一个复杂而昂贵的过程,需要不断重复。相反,无融合生殖F1杂种是真实遗传生物,即其能够纯育。

[0005] 在农业中,对无融合生殖具有极大的兴趣,因为它具有固定有利基因型的能力(无论其遗传复杂性如何),并且允许生产能够一步纯育的生物体。这意味着无融合生殖可用于即刻固定感兴趣的多基因的数量性状。应该指出的是,大多数产量性状是多基因的。无性生殖可用于多种性状(例如各种抗性、几种转基因或多种数量性状基因座)的堆积(或堆叠)。如果没有无融合生殖,为了固定这样的一系列性状,每个性状基因座必须单独制成纯合子,

且然后再组合成杂种。随着涉及性状的基因座数目的增加,通过杂交产生纯合性状基因座是费力的、耗时的和具有逻辑挑战性的。同样,为F1杂种选择合适的亲本种系需要投入大量时间和精力。此外,等位基因之间的特定上位相互作用在纯合(亲本系)阶段中丧失,并且可能不会在F1杂种中组合时恢复。通过无融合生殖,使固定这种非加性遗传变异成为可能。

[0006] 除了即时固定任何基因型(无论其复杂性如何)外,无性生殖还有其他重要的农业用途。由于减数分裂问题,有性种间杂种和自聚倍体经常发生不育。由于无融合生殖体跳过减数分裂,所以在种间杂交种和自聚倍体中出现的这些问题将被解决。由于无融合生殖阻止雌性杂交,因此已经提出将无融合生殖与雄性不育相结合用于对转基因加以控制,从而防止在转基因作物野生亲属中出现转基因渐渗杂交(Daniell, H. 2002. *Molecular strategies for gene containment in transgenic crops. Nature biotechnology* 20: 581-586)。在昆虫授粉作物(例如芸苔属)中,无融合生殖种子形成不会因传粉者服务不足而受到限制。鉴于授粉蜜蜂种群日益严重的健康问题(瓦螨感染、非洲杀戮蜂等),这一点变得越来越重要。由于大多数病毒不通过种子传播,对于块茎繁殖作物(如马铃薯)可利用无融合生殖以克隆的方式保持优良基因型,而消除通过块茎传播病毒的风险。无融合生殖种子的储存成本也会低于块茎。在观赏植物中,无融合生殖可取代劳动密集型和昂贵的组织培养繁殖。众所周知,无融合生殖通常会大大降低品种发育和繁殖的成本。

[0007] 主要农作物不发生无融合生殖,其中大部分是有性的种子作物。已有众多尝试试图在有性作物中引入无融合生殖。具体而言,由于无融合生殖受遗传调控,许多人试图鉴定参与无融合生殖过程的基因。在天然无融合生殖植物中,已经发现了无融合生殖为无融合生殖基因的来源(Ozias-Akins, P. and P. J. van Dijk. 2007 in: *Annu. Rev. Genet.* 41: 509-537)。然而,对于无融合生殖的遗传学和分子背景仍然知之甚少,并且对鉴定无融合生殖基因的尝试迄今尚未产生适用于农业的基因。这主要是由于鉴定和分离无融合生殖基因已被证明是一项困难的任務。天然无融合生殖植物通常是多倍体,而在多倍体植物中难以进行定位克隆。其他复杂因素是抑制无融合生殖特异性染色体区域中的重组、重复序列和杂交中的分离畸变。此外,无融合生殖植物的基因组还没有被测序,这使整个无融合生殖基因的搜索变得复杂。因此,无融合生殖基因尚未被克隆和/或分离。尝试在有性作物中引入无融合生殖可概括如下:

[0008] a) 通过广泛杂交将野生无融合生殖植物中的无融合生殖(无融合生殖的)基因渐渗杂交入作物物种迄今尚未成功,例如,试图将无融合生殖从鸭茅状摩擦禾(*Tripsacum dactyloides*)转移到玉米和小米[Savidan, Y. (2001). *Transfer of apomixis through wide crosses. In Flowering of Apomixis: From Mechanisms to Genetic Engineering*, Y.; apomixis from *Pennisetum squamulatum* into pearl millet. Savidan, J.G. Carman, and T. Dresselhaus, eds (Mexico: CIMMYT, IRD, European Commission DG VI), pp. 153-167; Morgan, R., Ozias-Akins, P., and Hanna, W.W. (1998). *Seed set in an apomictic BC3 pearl millet. Int. J. Plant Sci.* 159, 89-97.; W097/10704.]。

[0009] b) 有性模式物种的突变体,特别是在拟南芥中。例如, W02007066214描述了在拟南芥中使用称为Dyad的不完全减数分裂突变体。然而, Dyad是一种外显率很低的隐性突变。将该突变体实际用于作物物种中,其实际用途是非常有限的。

[0010] c) 通过两种有性生态型之间的杂交从头生成无融合生殖并未导致在农学上感兴

趣的无融合生殖体(US20040168216和US20050155111)。

[0011] d) 通过玉米中的转座子标签克隆候选的无融合生殖基因。US20040148667公开了elongate基因的同源物,其被假定为诱导无融合生殖。然而,据Barrell和Grossniklaus(2005)在Plant Journal Vol:34,pp 309-320中所述,elongate基因跳过了第二次减数分裂,并因此不能维持母本的基因型。

[0012] 此外,在US20060179498中已经描述了所谓的“反向育种”可用作无融合生殖的替代选择。然而,与不需要任何实验室方法的无融合生殖(因为它是由植物本身在没有任何外部(人类)干预的情况下进行的体内方法)相比,反向育种代表了复杂且费力的体外实验室方法。而且,在反向育种中,一旦亲本系被重建(双倍配子纯合子),则必须进行杂交。

[0013] 因此,需要用于诱导有性作物无融合生殖的替代方法,其至少不具有现有技术的一些局限性。特别地,需要用于生产倍数孢子体形成植物和无融合生殖种子的方法。还需要揭示无融合生殖过程中(特别是倍数孢子体形成)涉及的替代基因和蛋白质,这些基因和蛋白质适用于上述方法,并且可基本模仿有性作物中的无融合生殖途径。

[0014] 发明概述

[0015] 本发明提供了Dip基因的核酸序列和氨基酸序列以及其(功能性)同源物、片段和变体,其提供了倍数孢子体形成作为无融合生殖的一部分的倍数孢子体形成。还提供了倍数孢子体形成植物及其制备方法以及使用它们的方法,和制备无融合生殖种子的方法。

[0016] 发明详述

[0017] 定义

[0018] 如本文所用,术语“有性植物繁殖”是指称为“大孢子母细胞”的二倍体体细胞经历减数分裂以产生四个减数的大孢子(reduced megaspore)的发育途径。这些大孢子中的一个经有丝分裂以形成含有减数的卵细胞(即与母体相比具有减少的染色体数量的细胞)和两个减数的极核的雌配子体(也称为胚囊)。一个花粉粒精子细胞使卵细胞受精产生二倍体胚,而第二个精细胞使两个极核受精产生三倍体胚乳(过程被称为双受精)。

[0019] 如本文所用,术语“大孢子母细胞”是指通过减数(通常是减数分裂)产生大孢子的二倍体细胞,以产生将要发育成雌配子体的四个单倍体大孢子。在被子植物(也称为开花植物)中,大孢子母细胞通过包括大孢子发生(在珠心或大孢子囊中形成大孢子)和雌配子体发生(大孢子发育成雌配子体)两个不同的过程,产生发育为雌配子体的大孢子。

[0020] 如本文所用,术语“无性植物繁殖”是通过在没有受精并且没有配子融合实现植物繁殖的过程。无性生殖产生新的个体,除非发生突变,其在遗传上与亲本植物相同并且彼此相同。植物有两种无性繁殖的主要类型,包括营养繁殖(即涉及芽殖分蘖等的初始植物的营养片)和无融合生殖。

[0021] 如本文所用,术语“无融合生殖”是指通过无性过程形成种子。

[0022] 如本文所用,术语“倍数孢子体形成”是指直接通过有丝分裂或通过中止的减数分裂事件从大孢子母细胞衍生出未减数的胚囊的情况。已经报道了三种主要类型的倍数孢子体形成,以它们出现的植物命名,它们是蒲公英、苦苣菜(Ixeris)和蝶须属(Antennaria)型。在蒲公英型中,减数分裂前期开始,但随后该过程中止,形成两个未减数的二分体,其中一个通过有丝分裂产生胚囊。在苦苣菜型中,减数分裂前期之后,核按均等分裂进一步进行两次有丝分裂以产生八核胚囊。蒲公英和苦苣菜型被称为减数分裂倍数孢子体形成,因为

它们涉及减数分裂的修饰。相反,在蝶须属型,称为有丝分裂倍数孢子体形成,大孢母细胞不启动减数分裂,并直接分裂三次以产生未减数的胚囊。在通过倍数孢子体形成的配子体无融合生殖中,未减数的大孢子产生未减数的配子体。这个未减数的大孢子来自有丝分裂样分裂(有丝分裂倍数孢子体形成)或经修饰的减数分裂(减数分裂倍数孢子体形成)。在无孢子生殖的配子体无融合生殖和倍数孢子体形成的配子体无融合生殖中,未减数的卵细胞孤雌生殖地发育成胚。蒲公英的无融合生殖体是二倍体型,这意味着第一个雌性减数分裂(减数分裂I)被跳过,导致两个未减数的大孢子具有与母本植物相同的基因型。这些大孢子中的一个退化,且其他存活的未减数的大孢子产生未减数的雌配子体(或胚囊),其含有未减数的卵细胞。这种未减数的卵细胞在没有受精的情况下发育成与母本植物具有相同基因型的胚。由配子体无融合生殖过程产生的种子被称为无融合生殖种子。

[0023] 术语“倍数孢子体形成功能”是指在植物中,优选在雌性子房中,优选在大孢子母细胞和/或雌配子中,诱导倍数孢子体形成的能力。因此,其中引入了倍数孢子体形成功能的植物能够进行倍数孢子体形成过程,即通过第一次减数分裂恢复产生未减数的配子。

[0024] 术语“作为配子体无融合生殖的一部分的倍数孢子体形成”是指无融合生殖过程的倍数孢子体形成部分,即,在无性生殖过程中倍数孢子体形成在种子形成中所起的作用。特别是,除了倍数孢子体形成功能外,还需要无融合生殖功能以建立无融合生殖过程。因此,倍数孢子体形成和孤雌生殖功能的组合可能导致无融合生殖。

[0025] 已知无融合生殖以不同形式发生,包括至少两种称为配子体无融合生殖和孢子体无融合生殖(也称为不定胚形状)的形式。其中发生配子体无融合生殖的植物的实例包括蒲公英(*Taraxacum* sp.)、山柳菊(*Hieracium* sp.)、草地早熟禾(*Poa pratensis*)、鸭茅状摩擦禾(*Tripsacum dactyloides*)等。发生孢子体无融合生殖的植物的实例包括柑橘属(*Citrus* sp.)、山竹(*Garcinia mangostana*)等。

[0026] 如本文所用,术语“倍数孢子体形成植物”是指通过倍数孢子体形成的配子体无融合生殖的植物,或已经诱导(例如通过遗传修饰)以进行通过倍数孢子体形成的配子体无融合生殖的植物。在这两种情况下,当与无融合生殖因子结合时,倍数孢子体形成植物产生无融合生殖种子。

[0027] 如本文所用,术语“无融合生殖种子”是指从无融合生殖植物物种或通过诱导经历无融合生殖(特别是通过倍数孢子体形成的配子体无融合生殖)的植物或作物获得的种子。无性生殖种子的特征在于它们是克隆且在遗传上与亲本植物相同,并且能萌发具有真实遗传能力的植物。

[0028] 细胞、植物、植物的部分或种子的“克隆”的特征在于它们在遗传上与它们的同胞以及它们所来源的亲本植物相同。单个克隆的基因组DNA序列几乎相同,但是,突变可能会导致细微的差异。

[0029] 如本文所用,术语“真实遗传”或“真实遗传生物体”(也称为纯种生物体)是指总是将某种表型性状不变或几乎不变地遗传给其后代的生物体。生物体被称作针对每种性状的真实遗传(其中每种性状是真是遗传的),且“真实遗传”一词也被用来描述单个遗传性状。

[0030] 如本文所用,术语“F1杂种”(或子代1杂种)指代明显不同于亲本类型的后代的第一代子代。F1杂种被用于遗传学,在选择育种中它可能出现为F1杂交。明显不同亲本类型的后代产生具有亲本特征组合的新的统一表型。“F1杂种”与例如杂种优势的独特优势相关,

且因此在农业实践中非常需要。在本发明的一个实施方案中,本文所教导的方法、基因、蛋白质、其变体或片段可用于固定F1杂种的基因型,无论其遗传复杂性,并且允许在一个步骤中产生纯育的生物体。

[0031] 如本文所用,术语“等位基因”是指在特定基因座处的基因的任何一种或多种可变形式。在生物体的二倍体细胞中,给定基因的等位基因位于染色体上的特定位置或基因座。一对同源染色体的每条染色体上存在一个等位基因。二倍体或多倍体植物物种可在特定位置包含大量不同的等位基因。在一个实施方案中,如本文所教导的野生蒲公英种(accession)的Dip基因座可包含各种Dip或dip等位基因,其在核酸和/或编码的氨基酸序列中可略有不同。

[0032] 如本文所用,术语“基因座”是指染色体上的一个或多个特定位置或位点,其中例如一个或多个基因或遗传标记物位于其中。例如,如本文所教导的“Dip基因座”是指在其中发现如本文所教导的Dip基因(和两个(或更多个)Dip等位基因)在基因组中的位置。

[0033] 如本文所用,术语“显性等位基因”是指一个基因的等位基因之间的关系,其中一个等位基因的表型(即显性等位基因)的影响掩盖了在相同基因座处第二等位基因(隐性等位基因)的贡献。第一个等位基因是显性的,且第二个等位基因是隐性的。对于常染色体(除性染色体以外的任何染色体)上的基因,等位基因及其相关性状是常染色体显性或常染色体隐性。显性是孟德尔遗传学和经典遗传学中的一个关键概念。例如,显性等位基因可编码功能性蛋白质,而隐性等位基因则不编码。在一个实施方案中,如本文所教导的基因和其片段或变体指Dip基因的显性等位基因。

[0034] 如本文所用,术语“雌性子房”是指其中形成孢子的外壳。它可由单个细胞组成或可以是多细胞的。所有植物、真菌和许多其他谱系在其生命周期的某时间点形成子房。子房可通过有丝分裂或减数分裂产生孢子。通常,在每个子房内,大孢子母细胞的减数分裂产生四个单倍体大孢子。在裸子植物和被子植物中,这四个大孢子中只有一个在成熟时起作用,且另外三个大孢子退化。余留的大孢子有丝分裂并发育成雌配子体,其最终产生一个卵细胞。

[0035] 如本文所用,术语“雌配子”是指在有性生殖的生物体中受精(受孕)期间与另一种(“雄性”)细胞融合的细胞。在产生两种形态上不同类型的配子并且其中每个个体只产生一种类型的物种中,雌性是产生更大类型配子(称为胚珠(卵子)或卵)的任何个体。在植物中,雌性胚珠由花的子房产生。当成熟时,单倍体胚珠产生准备受精的雌性配子。雄性细胞是(主要是单倍体)花粉,且由花药产生。

[0036] 如本文所用,术语“授粉”是指花粉从花药(雄性部分)转移至植物的柱头(雌性部分)从而使之受精和繁殖的过程。它是作为花卉植物的被子植物所特有的。每个花粉粒是一个雄性单倍体配子体,适于被运送到雌性配子体,在双重受精过程中它可通过产生雄性配子(或配子)来影响受精。一个成功的被子植物花粉粒(配子体)含有雄性配子被运送到柱头,在那里萌发且其花粉管沿花柱下降生长到子房。它的两个配子沿着管下行至被保持在心皮内的含有雌配子的配子体。一个核与极核融合产生胚乳组织,且另一个核与胚珠产生胚。即使是大多数的天然无融合生殖体也需要授粉来进行胚乳的有性发育。然而,在少量的无融合生殖体中,例如在蒲公英和山柳菊中,胚乳的发育没有通过被称为自发性胚乳发育的极核受精过程。在拟南芥中,已知一些引起自发性胚乳发育的突变。

[0037] 如本文所用,术语“无融合生殖”是指不完全减数分裂的形式,其中胚的生长和发育不通过受精而发生。本发明的基因和蛋白质可与无融合生殖因子(例如基因或化学因子)组合产生无融合生殖后代。

[0038] 如本文所用,术语“液泡蛋白分选相关蛋白类型13”(缩写为VPS13),是指由所述Vps13基因编码的蛋白质,其参与通过反式高尔基体网络到液泡和细胞膜的蛋白质循环步骤的控制。

[0039] 如本文所用,术语“遗传标记”或“多态性标记”是指基因组DNA上的区域,其可用于“标记”染色体上的特定位置。如果一个遗传标记与一个基因紧密关联,或是遗传标记“位于”基因中(在基因标记中),它“标记”在DNA中发现所述基因的所述DNA,并因此可用于如本文所教导的(分子)标记分析以选择或不选择所述基因的存在,例如在标记辅助的育种/选择(MAS)方法中。遗传标记的非限制性实例是AFLP(扩增片段长度多态性,EP534858)、微卫星、RFLP(限制性片段长度多态性)、STS(序列标记位点)、SNP(单核苷酸多态性)、SFP(单特征多态性;参见Borevitz等人(2003) In: Genome Research Vol:13, pp 513-523)、SCAR(序列表征的扩增区)、CAPS标记(酶切扩增多态序列)等。标记离基因越远,标记与基因之间发生重组(互换)的可能性越大,由此失去了关联(以及标记和基因的共分离)。遗传位点之间的距离用重组频率来衡量,并以cM给出(centiMorgans;1cM是1%的两个标记之间的减数分裂重组频率)。由于物种之间的基因组大小差别很大,因此由1cM表示的实际物理距离(即两个标记之间的千碱基,kb)在物种之间也大不相同。应该理解的是,当提及本文中的“连接”标记时,这也包括“位于”基因本身的标记。

[0040] 如本文所用,术语“标记辅助选择”(缩写为“MAS”)指的是筛选植物以筛选一种或多种遗传和/或表型标记的存在和/或不存在,以加速转移包含标记(和任选地缺乏侧翼区域)的DNA区域导入(精英)育种品系的过程。

[0041] 如本文所用,术语“分子标记物测定”(或测试)是(直接或间接地)指植物或植物部分中特定等位基因(例如Dip等位基因)存在或不存在的(基于DNA的)测定。优选地,它允许确定在任何单个植物中特定的等位基因在Dip基因座处是纯合的还是杂合的。例如,在一个实施方案中,使用PCR引物扩增与Dip基因座连接的核酸,扩增产物被酶促消化,并基于扩增产物的电泳分析模式,人们可确定哪个Dip等位基因存在于任何单个植物中,以及等位基因在Dip位点的接合性(即每个基因座的基因型)。分子标记测定的非限制性实例包括序列表征的扩增区域(SCAR)标记测定、切割的扩增多态性序列(CAPS)标记测定等。

[0042] 如本文所用,术语“杂合的”是指一种遗传条件,其中存在如果两个(或更多在多倍体情况下)不同等位基因存在于特定基因座如Dip基因座(例如显性Dip等位基因/隐性dip等位基因)中,但存在单独定位在细胞中相应的同源染色体对上。

[0043] 如本文所用,术语“纯合的”是指一种遗传条件,其当两个(或多个,在多倍性的情况下)相同的等位基因位于特定基因座(例如Dip基因座,例如对于显性Dip等位基因/隐性dip等位基因),但分别位于细胞中相对应的同源染色体上时存在。

[0044] 如本文所用,术语“品种”是指符合UPOV惯例,并且是指在最低已知等级的单一植物学分类内的植物分组,该分组可通过由给定基因型或基因型的组合产生的特征的表达来定义,其可通过至少一种所述特征的表达来区别于任何其他植物分组,并且鉴于它具有在繁殖时保持不变(稳定)的适合性而作为单位被考虑。

[0045] 如本文所用,术语“多肽”和“蛋白质”可互换使用,并且指由氨基酸链组成的分子,而不涉及特定的作用模式、尺寸、三维结构或起源。

[0046] 如本文所用,术语“分离的多肽”或“分离的蛋白质”可互换使用,并且指不再存在于其天然环境中的蛋白质,例如存在于试管中(体外)或存在于重组细菌或植物宿主细胞中的蛋白质是分离的蛋白质。

[0047] 如本文所用,术语“核酸”是指嘧啶和嘌呤碱基的任何聚合物或低聚物,优选分别为胞嘧啶、胸腺嘧啶和尿嘧啶以及腺嘌呤和鸟嘌呤(参见Albert L. Lehninger, Principles of Biochemistry, at 793-800 (Worth Pub. 1982), 为了所有目的通过引用将其全文引入)。本发明考虑任何脱氧核糖核酸、核糖核酸或肽核酸组分及其任何化学变体,例如这些碱基的甲基化、羟基甲基化或糖基化形式等。聚合物或低聚物在组成上可以是异质或同质的,并且可从天然发生的来源分离或可人工产生或合成产生。

[0048] 根据本发明,术语“多核苷酸”、“核酸分子”、“核苷酸序列”或“核酸序列”是指单链或双链形式的聚合DNA或RNA分子,特别是编码蛋白质的DNA、其变体或片段。

[0049] 如本文所用,术语“分离的多核苷酸”、“分离的核酸分子”、“分离的核苷酸序列”或“分离的核酸序列”是指不再处于天然环境中的多核苷酸,即与天然地伴随自然序列或蛋白质(例如核糖体、聚合酶、许多其他序列和蛋白质)的其他细胞组分基本上分离。该术语包括已从其天然存在的环境中分离的多核苷酸,并包括重组或克隆的核酸分离物和化学合成的类似物或通过异源系统生物合成的类似物,例如在细菌宿主细胞或植物核中或质体基因组中的核酸序列。

[0050] 如本文所用,术语“功能性DIP基因或蛋白质”或“功能性DIP基因或蛋白质变体或片段”(例如直系同源物或突变体,和基因的部分)是指基因和/或编码的蛋白质修饰的能力或在植物中通过在所述植物中改变一种或多种基因的表达水平(例如通过过表达或沉默),(定量和/或定性的)诱导无融合生殖(特别是通过倍数孢子体形成的配子体无融合生殖)的过程的能力。例如,从植物物种X获得的推定的DIP蛋白质的功能性可通过各种方法来测试。优选地,如果蛋白质是功能性的,在植物物种X中使用例如基因沉默载体来沉默编码蛋白质的Dip基因,将引起倍数孢子体形成的减少(即染色体数量将减少)或抑制,而在易感植物中过表达将导致增强的倍数孢子体形成。同样,对功能性DIP蛋白的补足将能够恢复或赋予倍数孢子体形成。技术人员在测试功能性方面将没有困难。

[0051] 如本文所用,术语“基因”是指包含可操作地连接至合适的调控区(例如启动子)的区域(转录区)的DNA序列,所述区域在细胞中被转录成RNA分子(例如mRNA)。因此基因可包含几个可操作地连接的序列,例如启动子、包含例如翻译起始相关序列的5'前导序列、(蛋白质)编码区(cDNA或基因组DNA)和包含例如转录终止位点的3'非翻译序列。

[0052] 如本文所用,术语“嵌合基因”或“重组基因”是指在自然界中通常不存在于一个物种中,特别是一个基因中存在的核酸序列的一个或多个核苷酸序列的部分在自然界中以无相互关联的形式存在的任何基因。例如,启动子在自然界中与部分或全部的转录区或与另一个调控区不相关。术语“嵌合基因”被理解为包括在其中启动子或转录调节序列可操作地连接至一个或多个编码序列或反义序列(有义链的反向互补)或反向重复序列(有义和反义,由此RNA转录物在转录时形成双链RNA)的表达构建体。

[0053] 术语“3' UTR”或“3'非翻译序列”(也称为“3'非翻译区”或“3'末端”)是指在基因的

编码序列的下游发现的核酸序列,其包含例如转录终止位点和(在大多数但非全部真核 mRNA 中)聚腺苷酸化信号(例如 AAUAAA 或其变体)。在转录终止之后,mRNA 转录物可在聚腺苷酸化信号的下游切割并且可添加参与将 mRNA 向细胞质(翻译发生处)转运的 poly(A) 尾。

[0054] 如本文所用,术语“5' UTR”或“前导序列”或“5' 非翻译区”是指 mRNA 转录的区域和相应的 DNA,在 mRNA 转录开始的+1 位与编码区的翻译起始密码子(通常为 mRNA 上的 AUG 或 DNA 上的 ATG)之间。5' UTR 通常含有对翻译、mRNA 稳定性和/或降解(turnover)以及其他调控元件重要的位点。

[0055] 如本文所用,术语“所述基因或变体或其片段的表达”是指将与适当的调控区(特别是启动子)可操作地连接的 DNA 区域被转录成具有生物活性的 RNA(即其能够被翻译成具有生物活性的蛋白质或肽(或活性肽片段)或其本身具有活性(例如在转录后基因沉默或 RNAi))的过程。某些实施方案中的活性蛋白质是指具有持续活性的蛋白质。编码序列优选以有义方向并编码期望的、具有生物活性的蛋白质或肽或活性肽片段。在基因沉默方法中,DNA 序列优选以反义 DNA 或反向重复 DNA 的形式存在的,其包含反义方向或有义和反义方向的靶基因的短序列。“异位表达”是指在通常不表达该基因的组织中的表达该基因。

[0056] 如本文所用,术语“转录调控序列”是指能够调节与转录调控序列可操作连接的(编码)序列的转录速率的核酸序列。因此,本文所定义的转录调控序列将包含启动转录所必需的所有序列元件(启动子元件),用于维持和调节转录,包括例如衰减子或增强子。虽然主要指编码序列的上游(5')转录调控序列,但是在该定义中也包括编码序列的下游(3')发现的调控序列。

[0057] 如本文所用,术语“启动子”是指用于控制一个或多个基因的转录的核酸片段,其位于该基因的转录起始位点的转录方向的上游,并且在结构上以 DNA 依赖性 RNA 聚合酶的结合位点、转录起始位点和任何其他 DNA 序列(包括但不限于转录因子结合位点、阻遏物和激活物蛋白质结合位点,以及本领域技术人员已知的直接或间接起作用以调节来自启动子的转录量的任何其他核酸序列)的存在为特征。任选地,术语“启动子”还可包括 5' UTR 区(例如,启动子可包括基因的翻译起始密码子的上游(5')的一个或多个部分,因为该区域可在调节转录和/或翻译中发挥作用)。

[0058] 如本文所用,术语“组成型启动子”是指在大多数生理和发育条件下在大多数组织中有活性的启动子。

[0059] 如本文所用,术语“诱导型启动子”是指生理上(例如通过外部施用某些化合物)或发育上调节的启动子。

[0060] 如本文所用,术语“组织特异性启动子”是指仅在特定类型的组织或细胞中有活性的启动子。“在植物或植物细胞中有活性的启动子”是指启动子驱动植物或植物细胞内转录的一般能力。它不对启动子的时空活性有任何暗示。

[0061] 如本文所用,术语“可操作地连接”是指多核苷酸元件以功能性关系连接。当核酸置于与另一核酸序列的功能关系中时,核酸“可操作地连接”。例如,如果启动子或转录调控序列影响编码序列的转录,则其与编码序列可操作地连接。可操作地连接意味着被连接的 DNA 序列通常是连续的,并且在必要时连续地并在阅读框中连接两个蛋白质的编码区以产生嵌合蛋白质。

[0062] 如本文所用,术语“嵌合蛋白”或“杂合蛋白”是指由各种蛋白质结构域或基序组成

的蛋白质,其在自然界中并未如此发现,但其连接形成功能性蛋白质,所述功能性蛋白质显示出连接的结构域的功能性。嵌合蛋白也可以是天然存在的两种或更多种蛋白质的融合蛋白。

[0063] 如本文所用,术语“结构域”是指具有特定结构或功能的蛋白质的任何部分或区域,其可被转移到另一种蛋白质以提供至少具有结构域的功能性特征的新杂合蛋白质。

[0064] 如本文所用的术语“靶肽”是指将蛋白质或蛋白质片段靶向至细胞内细胞器例如质体,优选叶绿体、线粒体或细胞外空间或非原质体(分泌信号肽)的氨基酸序列。编码靶肽的核酸序列可与编码蛋白质或蛋白质片段的氨基末端(N-末端)的核酸序列融合(框内),或者可用于替换天然靶肽。

[0065] 如本文所用,术语“核酸构建体”或“载体”是指由使用重组DNA技术产生并用于将外源DNA或RNA递送到宿主细胞中的人造核酸分子。载体骨架可以是例如双元的或超二元载体(参见例如US5591616、US2002138879和W09506722),如本领域已知的并且如本文别处所述的共整合载体或T-DNA载体,嵌合基因被整合入其中,或者如果已经存在合适的转录调控序列,则只将期望的核酸序列(例如编码序列、反义或反向重复序列)整合到转录调控序列的下游。载体通常包含其他遗传元件以促进其用于分子克隆,例如,选择标记、多克隆位点等(见下文)。

[0066] 如本文所用,术语“宿主细胞”或“重组宿主细胞”或“转化细胞”或“转基因细胞”指由于至少一种核酸分子特别是包含编码所期望的蛋白质或核酸序列的嵌合基因,而产生的新个体细胞(或生物体),所述核酸序列在转录后产生用于沉默靶基因/基因家族的反义RNA或反向重复RNA(或发夹RNA)或siRNA或miRNA,其已被引入所述细胞。宿主细胞优选是植物细胞或细菌细胞。宿主细胞可含有作为染色体外(附加型)复制分子的核酸构建体,或者更优选包含整合在宿主细胞的核或质体基因组中的嵌合基因。

[0067] 如本文所用,术语“重组植物”或“重组植物部分”或“转基因植物”是指在所有细胞和植物部分在相同的基因座处包含如本文所教导的嵌合基因的植物或植物部分(例如种子或果实或叶子),尽管基因可能不会在所有细胞中表达。

[0068] 如本文所用,术语“精英事件”是指已选择在基因组中使得该植物具有良好的或期望的表型和/或农学特性的位置处包含重组基因的重组植物。可对整合位点的侧翼DNA进行测序以表征整合位点并区别于在基因组中其他位置包含相同嵌合基因的其他转基因植物。

[0069] 如本文所用,术语“可选择标记”是指本领域中公知的术语,并且在本文中用于描述任何遗传实体,其在表达时可用于选择含有可选择标记的细胞。可选择标记基因产物赋予例如抗生素抗性 or 更优选除草剂抗性 or 另一种可选择的性状如表型性状(例如色素沉着的变化)或营养需求。术语“报告基因”主要用于指可见标记,如绿色荧光蛋白(GFP)、eGFP、萤光素酶、GUS等。

[0070] 如本文所用,术语“基因直系同源物”或“蛋白质直系同源物”是指在另一物种中发现的同源基因或同源蛋白质,其具有与基因或蛋白质相同的功能,但是(通常)从此物种具有基因分化的时候起发生序列分化(即从物种的共同祖先进化而来的基因)。在一个实施方案中,基于二者序列比较(例如基于整个序列或特定结构域上的序列同一性的百分比)和功能性分析,可在其他植物物种中鉴定蒲公英属Dip基因的直系同源物。

[0071] 本文所用的表达“同线性区域”是指用于比较基因组学的术语,并且指的是两个相

关物种的染色体上的相同区域。

[0072] 如本文所用,术语“严格杂交条件”是指可用于鉴定与给定核酸序列基本相同的核酸序列的条件。严格条件是序列依赖性的,并且在不同情况下会有所不同。通常,严格条件被选择为比限定离子强度和pH下特定序列的热溶解温度(T_m)低约 5°C 。 T_m 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的温度。通常选择的严格条件为其中盐浓度在pH 7时约为0.02摩尔且温度至少为 60°C 。降低盐浓度和/或提高温度增加了严格性。RNA-DNA杂交的严格条件(使用例如100nt的探针的Northern印迹)是例如包括在 63°C 0.2X SSC中至少一次洗涤20分钟或等同条件的那些条件。用于DNA-DNA杂交的严格条件(使用例如100nt的探针的Southern印迹)例如包括在至少 50°C ,通常约 55°C 的温度下,在0.2X SSC中至少一次洗涤20分钟(通常2次)或等同条件的那些条件。另见Sambrook et al. (1989)和Sambrook and Russell (2001)。

[0073] 如本文所用,术语“高严格条件”是指可通过例如在 65°C 下在含有 $6\times\text{SSC}$ ($20\times\text{SSC}$ 含有3.0M NaCl、0.3M柠檬酸钠, pH 7.0)、 $5\times\text{Denhardt's}$ ($100\times\text{Denhardt's}$ 含有2% Ficoll、2% 聚乙烯吡咯烷酮、2% 牛血清白蛋白)、0.5% 十二烷基磺酸钠(SDS)和 $20\mu\text{M}$ 巯基磺变性载体DNA(单链鱼精子DNA,平均长度为120-3000个核苷酸)作为非特异性竞争剂的水溶液中杂交而可实现的条件。杂交后,可在几个步骤中进行高度严格洗涤,在 $0.2-0.1\times\text{SSC}$ 、0.1% SDS中在杂交温度下进行最终洗涤(约30分钟)。

[0074] 如本文所用,术语“中等严格性”是指等效于上述溶液中杂交的条件,但在约 $60-62^{\circ}\text{C}$ 的条件下进行。在这种情况下,最终洗涤在杂交温度下在 $1\times\text{SSC}$ 、0.1% SDS中进行。

[0075] 如本文所用,术语“低严格性”是指在等效于上述溶液中杂交在约 $50-52^{\circ}\text{C}$ 下的条件。在这种情况下,最终洗涤在杂交温度下在 $2\times\text{SSC}$ 、0.1% SDS中进行。另见Sambrook et al. (1989)和Sambrook and Russell (2001)。

[0076] 当在氨基酸序列或核酸序列的上下文中使用时,本文所用的术语“基本相同的”或“基本相同”或“基本上相似”或“基本相似性”或“变体”或“序列同一性”指两条氨基酸序列或两条核酸序列当例如通过程序GAP或BESTFIT使用默认参数进行最佳比对时,共享至少一定百分比的序列同一性。GAP使用Needleman和Wunsch全局比对算法将两个序列在其全长上进行比对,从而最大化匹配数量并最小化空位数量。通常,使用GAP默认参数,其缺口产生罚分= 50 (核酸)/ 8 (蛋白质)和缺口延长罚分= 3 (核酸)/ 2 (蛋白质)。对于核酸,使用的默认评分矩阵是nwsgapdna,并且对于蛋白质,默认评分矩阵是Blosom62 (Henikoff&Henikoff, 1992, PNAS 89, 915-919)。很明显,当RNA序列被认为与DNA序列基本相似或具有一定程度的序列同一性时,DNA序列中的胸腺嘧啶(T)被认为与RNA序列中的尿嘧啶(U)等同。序列比对和百分比序列同一性的评分可使用计算机程序来确定,如GCG Wisconsin Package, Version 10.3, 获取自Accelrys Inc., 9685 Scranton Road, San Diego, CA 92121-3752 USA。或在EmbossWIN(版本2.10.0)中使用程序“needle”,使用与上述相同的GAP参数或使用缺口开启罚分10.0和缺口延长罚分0.5,使用DNAFULL作为矩阵。为了比较不同长度序列之间的序列同一性,优选使用局部比对算法,如Smith Waterman算法(Smith TF, Waterman MS (1981) J. Mol. Biol 147 (1); 195-7),使用例如在EmbossWIN中的程序“water”。默认参数是缺口开启罚分10.0和缺口延长罚分0.5,对于蛋白质使用Blosom62和对于核算使用DNAFULL矩阵。

[0077] 这里使用的术语“包含”和“包括”以及它们的同义词是指这样一种情况,其中所述术语在其非限制性意义上被用来表示包括该单词之后的项目,但未被具体提及的项目不被排除。它还包含更限制性的动词“由……组成”。此外,不定冠词“一”或“一种”对元件的提及并不排除存在多于一个该元件的可能性,除非上下文明确要求存在一个且仅有一个该元件。因此,不定冠词“一”或“一种”通常意味着“至少一(种)”。进一步理解的是,当提及本文中的“序列”时,通常是指具有特定亚基序列(例如氨基酸)的实体分子。

[0078] 如本文所用,术语“植物”包括植物细胞、植物组织或器官、植物原生质体、可再生植物的植物细胞组织培养物、植物愈伤组织、植物细胞团和植物中完整的植物细胞,或植物的部分如胚、花粉、胚珠、孢子囊、果实、花、叶(如收获的莴苣作物)、种子、根、根尖等。

[0079] 如本文所用,术语“基因沉默”是指一个或多个靶基因(例如内源性Dip基因)的基因表达的下调或完全抑制。使用抑制性RNA来减少或消除基因表达在本领域中已经成熟,并且是一些综述的论题(例如Baulcombe 1996、Stam et al.1997、Depicker and Van Montagu,1997)。有许多技术可用于在植物中实现基因沉默,例如产生全部或部分靶基因的反义RNA的嵌合基因(参见例如EP 0140308B1、EP 0240208B1和EP 0223399B1),或者产生有义RNA(也称为共抑制),参见EP 0465572B1。然而迄今为止最成功的方法是产生目标基因的有义和反义RNA二者(“反向重复序列”),其在细胞中形成双链RNA(dsRNA)并沉默靶基因。用于dsRNA的生产和基因沉默的方法和载体已经在EP 1068311、EP 983370 A1、EP 1042462 A1、EP 1071762 A1和EP 1080208 A1中描述。因此,本发明中的载体可包含在植物细胞中具有活性的转录调控区,所述转录调控区可操作地连接至本发明中的DIP基因的有义和/或反义DNA片段。通常,靶基因序列的短(有义和反义)区段,例如编码或非编码序列的17、18、19、20、21、22或23个核苷酸是足够的。也可使用更长的序列,例如50、100、200或250个核苷酸或更多。优选地,短有义和反义片段被间隔序列(例如内含子)隔开,其在形成dsRNA时形成环(或发夹)。SEQ ID NO:4和/或SEQ ID NO:5的任何短片段或其片段或变体可用于制备DIP基因来源的沉默载体,和在所有或一些组织或器官中(取决于所使用的启动子)一个或多个靶基因被沉默的转基因植物。

[0080] 产生发夹构建体的便利方式是使用通用载体如基于Gateway® technology的载体pHANNIBAL和pHELLSGATE(参见Wesley et al.2004,Methods Mol Biol.265:117-30; Wesley et al.2003,Methods Mol Biol.236:273-86and Helliwell&Waterhouse 2003,Methods 30(4):289-95.),所有这些文献通过引用并入本文。

附图说明

[0081] 图1A.在没有异花授粉的情况下,完全无融合生殖的三倍体植物A68(野生型)的种子头部。注意完全发育的种子的黑暗中心。

[0082] 图1B.在没有异花授粉的情况下,A68倍数孢子体形成缺失(LoD)缺失突变体的典型种子头部。典型的LoD突变体在这些条件下具有比A68野生型小的种子头部。注意有斑点的中心,没有孤雌生殖基因的许多白色未发育的种子和有孤雌生殖基因的少许发育中的种子。孤雌生殖是配子体表达的基因,且因此当倍数孢子体形成缺失时分离并被减数分裂取代。

[0083] 图2.大型蒲公英种质资源组内的关联序列多态性和倍数孢子体形成表型。有性

(dip)和二倍体等位基因(Dip)之间的差异用灰色表示。

[0084] 发明详述

[0085] 在第一方面,本发明涉及分离的多核苷酸,其包含SEQ ID NO:1的核酸序列或与SEQ ID NO:1的核酸序列具有至少50%或70%,优选至少80%,更优选至少90%,甚至更优选至少95%,甚至更优选至少96%或97%,最优选至少98%或99%序列同一性的核酸序列。

[0086] 第二方面,本发明涉及分离的多核苷酸,其包含SEQ ID NO:2的核酸序列或与SEQ ID NO:2的核酸序列具有至少50%或70%,优选至少80%,更优选至少90%,甚至更优选至少95%,甚至更优选至少96%或97%,最优选至少98%或99%序列同一性的核酸序列。

[0087] 包含核酸序列SEQ ID NO:1或SEQ ID NO:2的分离的多核苷酸被鉴定为广义上的蒲公英属的推定的液泡蛋白质分选相关蛋白基因(Vps13)的一部分。Vps13基因是一个大基因。因此,SEQ ID NO:1和SEQ ID NO:2的所述核酸序列可包含在单个分离的核酸序列中,即作为相同核酸序列的一部分。因此分离的核酸序列可包含SEQ ID NO:1和SEQ ID NO:2或其变体。可以理解的是,Vps13基因可包含许多外显子和内含子以及其它基因相关序列,例如包含在SEQ ID NO:1中的启动子和终止子序列,延伸至所示蛋白质编码序列的5'和3'端(开放阅读框;ORF)(SEQ ID NO:2),并因此可大于SEQ ID NO:2。因此,序列同一性的百分比因此可能不涉及分离的核酸序列的完整序列。相反,仅包含在所述分离的核酸序列中的核酸序列可与SEQ ID NO:1或SEQ ID NO:2具有所述序列同一性的百分比。因此应理解,序列同一性的百分比应相对于包含在分离的核酸序列中的核酸序列来计算,其中核酸序列的第一个和最后一个核苷酸与SEQ ID NO:1和/或SEQ ID NO:2的核酸序列相匹配。因此,当要计算序列同一性的百分比时,优选仅相对于对应于SEQ ID NO:1和/或SEQ ID NO:2的序列。还应理解的是,SEQ ID NO:1和SEQ ID NO:2或其变体是编码序列,即编码氨基酸的序列。因此,DNA中的这种编码序列中可散布有内含子序列。因此,在从DNA序列计算序列同一性的情况下,不与SEQ ID NO:1和/或SEQ ID NO:2显示比对的部分序列(例如内含子)不被考虑。

[0088] 在一个实施方案中,如本文所教导的分离的多核苷酸具有如本文所教导的SEQ ID NO:1或SEQ ID NO:2的核酸序列或其变体。

[0089] 在一个实施方案中,如本文所教导的包含本文中教导的SEQ ID NO:1或SEQ ID NO:2的核酸序列或其变体或片段的分离的多核苷酸可称为“Dip或DIP多核苷酸”或“Dip或DIP基因”或“无融合生殖多核苷酸或无融合生殖基因”或“倍数孢子体形成多核苷酸或倍数孢子体形成基因”。

[0090] 在一个实施方案中,如本文所教导的分离的多核苷酸包含如本文所教导的SEQ ID NO:1或SEQ ID NO:2的核酸序列或其变体和/或所述多核苷酸和/或由所述多核苷酸编码的蛋白质的表达产物,所述多核苷酸能够向植物或植物细胞提供倍数孢子体形成功能,或能够诱导倍数孢子体形成或以倍数孢子体形成作为配子体无融合生殖的一部分,优选通过倍数孢子体形成进行的类型,优选在目前被认为是无性作物的作物中。通过倍数孢子体形成的配子体无融合生殖产生与亲本植物遗传上相同的后代。因此,在一个实施方案中,如本文所教导的分离的多核苷酸或其变体可用于产生遗传上与亲代植物相同的后代,而不需要受精和杂交。

[0091] 在一个优选的实施方案中,如上文所教导的Dip多核苷酸或基因及其变体和/或所述多核苷酸的表达产物和/或由所述多核苷酸编码的蛋白质能够向植物或植物细胞提供倍

数孢子体形成功能,优选的类型是在引入植物或植物细胞时在有性作物中通过倍数孢子体形成发生。

[0092] 应该理解,术语“分离的多核苷酸”或其变体(例如基因组DNA cDNA或mRNA)包括天然发生的、人造的或合成的核酸分子。核酸分子可编码本文所教导的任何多肽或其变体。所述核酸分子可用于产生如本文所教导的多肽或蛋白质或其变体。由于遗传密码的简并性,多种核酸分子可编码相同的多肽(例如本文所教导的包含SEQ ID NO:3和/或SEQ ID NO:7或12的氨基酸序列的多肽或蛋白质或其变体)。

[0093] 在一个实施方案中,如本文所教导的分离的多核苷酸包括任何变体核酸分子,其涵盖包含与SEQ ID NO:1或SEQ ID NO:2的核酸序列具有超过50%,优选超过55%、优选超过60%、优选超过65%、优选超过70%、优选超过75%、优选超过80%、优选超过85%、优选超过90%、优选超过95%、优选超过96%、优选超过97%、优选超过98%、和优选超过99%的序列同一性的核苷酸序列的任何核酸分子。变体还包括从具有SEQ ID NO:1或SEQ ID NO:2的核酸序列的核酸分子通过一个或多个核酸取代、缺失或插入衍生的核酸分子。优选地,与SEQ ID NO:1或SEQ ID NO:2相比,此类核酸分子包含1、2、3、4、5、6、7、8、9、10个或更多个至约100、90、80、70、60、50、45、40、35、30、25、20、15个核酸取代、缺失或插入。序列同一性可通过本领域可获得的任何合适手段来确定。例如,可使用生物信息学来执行核酸序列之间的成对比对以鉴定可能因序列之间的功能、结构或进化关系而具有相似性的区域。还应理解的是,可使用许多方法来鉴定、合成或分离本文所教导的多核苷酸的变体,例如核酸杂交、PCR技术、经电脑模拟分析和核酸合成等。

[0094] 在一个实施方案中,术语“变体”还包括自然界中(例如在其他蒲公英物种或其他植物中)发现的天然变体。从其他蒲公英属物种或其他植物中分离的所述变体核苷酸序列可包括显性Dip等位基因以及来自不同植物物种(例如包括不同的蒲公英品种、品系、种质或育种品系)的隐性dip等位基因。例如,不受理论的束缚,实施例中鉴定的EMS突变是可视作隐性dip的变体,因为倍数孢子体形成功能丧失,而野生型序列可被认为是显性Dip,因为野生型序列提供了倍数孢子体形成功能。

[0095] 在一个实施方案中,本发明所述的变体分离的多核苷酸,例如同源或直系同源物,也可在除了属于蒲公英属的那些之外的植物中发现和/或从中分离。所述分离的多核苷酸可使用已知方法如PCR、严格杂交方法等从其他野生或栽培的无融合生殖或非无融合生殖植物和/或其他植物中分离。因此,SEQ ID NO:1和/或SEQ ID NO:2的变体还包括天然发现于其他蒲公英植物、品系或品种中,和/或天然存在于其他物种的其他植物中的核苷酸序列。这种核苷酸可例如在Blast搜索中被鉴定,或者通过重新在植物体中鉴定相应序列。

[0096] 在一个实施方案中,如本文所教导的分离的多核苷酸变体包括例如衍生自与SEQ ID NO:1和/或SEQ ID NO:2不同“来源”的本发明所述的分离的多核苷酸,其为蒲公英来源。因此具体地讲,本发明包括来自一种植物(例如野生或栽培植物和/或来自其他植物)的基因或等位基因,其中存在倍数孢子体形成(作为通过倍数孢子体形成的配子体无融合生殖的一部分)。使用所提供的核苷酸序列和/或其互补序列或其部分作为引物或探针可容易地分离这些同系物。例如,可使用中度严格、严格或高度严格的核酸杂交方法。例如,可使用SEQ ID NO:1和/或SEQ ID NO:2序列的片段,或其互补序列。所述用于这种杂交方法的片段可包含SEQ ID NO:1和/或SEQ ID NO:2中的至少10、20、30、40、50、100、150、200、250、300、

400、500、600、700、800、900、1000、1500、2000、3000或更多个连续的核酸。

[0097] 应该理解,由于遗传密码的简并性,多种核酸序列可编码相同的氨基酸序列。为了在宿主中最佳表达,根据本发明的分离的核酸序列可通过调整密码子使用至植物基因中最优选的密码子来进行密码子优化,特别是使用可获得的密码子使用表(例如,更适应于目的植物中的表达)优化对于目的植物属或种天然的基因(Bennetzen&Hall,1982,J.Biol.Chem.257,3026-3031;Itakura et al.,1977Science 198,1056-1063)。各种植物物种的密码子使用表由例如Ikemura(1993,In“Plant Molecular Biology Labfax”,Croy, ed.,Bios Scientific Publishers Ltd.)和Nakamura等人(2000,Nucl.Acids Res.28,292.)和Nakamura等人(2000,Nucl.Acids Res.28,292.)发表,并在主要的DNA序列数据库(例如德国Heidelberg的EMBL)中。因此,可构建合成的DNA序列从而产生相同或基本相同的蛋白质。在专利和科学文献中可找到将密码子使用修改为被宿主细胞偏好的多种技术。修改密码子使用的确切方法对于本发明不是关键的。

[0098] 可常规地对上述DNA序列进行小的如上所述的修饰,即通过PCR介导的诱变(Ho et al.,1989,Gene 77,51-59.,White et al.,1989,Trends in Genet.5,185-189)。对DNA序列的修饰也可通过使用可获得的技术对所期望的编码区进行重新DNA合成而常规性引入。

[0099] 在一个实施方案中,通过添加或缺失蛋白质的N-末端的一个或多个氨基酸,可修饰本发明中的分离的多核苷酸或其变体使得DIP蛋白的N-末端具有最佳翻译起始环境。通常为了最佳翻译起始而优选的是,待在植物细胞中表达的本发明蛋白质以Met-Asp或Met-Ala二肽起始。因此可在现有的Met之后插入Asp或Ala密码子,或者第二密码子Val可被Asp(GAT或GAC)或Ala(GCT、GCC、GCA或GCG)的密码子置换。DNA序列也可被修饰以除去非法剪接位点。

[0100] 根据本发明分离的多核苷酸或其变体优选是“功能性的”,即它们优选能够向植物提供倍数孢子体形成功能,优选作为配子体无融合生殖的一部分,优选通过倍数孢子体形成发生于植物或植物细胞或有性作物中的类型。在一个实施方案中,提供了分离的多核苷酸或其变体,其与包含源自蒲公英属的核酸序列SEQ ID NO:1和/或SEQ ID NO:2的多核苷酸同源,所述分离的多核苷酸分离自无融合生殖植物。因此,在该实施方案中,根据本发明分离的多核苷酸或其变体是从无融合生殖植物中分离的。此类分离的多核苷酸或其变体可特别能够在植物或植物细胞或(有性)作物中向植物提供倍数孢子体形成功能。

[0101] 应理解的是,如本文所教导的多核苷酸的变体发挥与包含如本文所教导的SEQ ID NO:1或SEQ ID NO:2的核酸序列的多核苷酸相同的功能,即能够向植物或植物细胞提供倍数孢子体形成功能,优选作为在植物或植物细胞或有性作物中诱导倍数孢子体形成或配子体无融合生殖的一部分,特别是当导入植物或植物细胞或有性作物时。进一步理解的是,如本文所教导的任何分离的多核苷酸及其变体可编码如本文所教导的任何多肽及其变体。

[0102] 在一个实施方案中,如本文所教导的多核苷酸及其变体的表达产物是RNA分子,优选为mRNA分子或siRNA或miRNA分子。

[0103] 在一个实施方案中,如本文所教导的多核苷酸及其变体的片段和/或所述片段的表达产物和/或由所述片段编码的蛋白质能够向植物或植物细胞提供倍数孢子体形成功能,优选作为诱导配子体无融合生殖的一部分。

[0104] 在一个优选的实施方案中,如本文所教导的片段和/或由所述片段编码的蛋白质

能够提供倍数孢子体形成功能,优选诱导倍数孢子体形成或作为诱导配子体无融合生殖的一部分。

[0105] 在一个实施方案中,如本文所教导的片段的表达产物是RNA分子,优选为mRNA分子或siRNA或miRNA分子。

[0106] 在一个实施方案中,如本文所教导的片段可为具有包含本文所教导的SEQ ID NO:1或SEQ ID NO:2及其变体核酸序列的至少20、30、40、50、100、150、200、250、300、400、500、600、700、800、900、1000、1500、2000、或3000个连续核苷酸的长度的分离的多核苷酸。

[0107] 在一个优选的实施方案中,本文所教导的片段具有SEQ ID NO:4、6或11的核酸序列。

[0108] 在更加优选的实施方案中,本文所教导的片段的表达产物具有SEQ ID NO:5的核酸序列。

[0109] 在一个实施方案中,如本文所教导的片段的表达产物编码包含如SEQ ID NO:7和/或12中所述的氨基酸序列的多肽。

[0110] 嵌合基因和载体

[0111] 在一个实施方案中,嵌合基因可包含本文所教导的任何多核苷酸、其片段和变体。

[0112] 在一个实施方案中,当包含在本文所教导的载体中时,本文所教导的任何多核苷酸、其片段和变体可以可操作地连接至启动子。可使用本领域已知的任何启动子,并且其适用于与本文所教导的多核苷酸、其片段和变体连接。合适的启动子的非限制性实例包括允许组成型或调控型表达、弱表达和强表达等的启动子。本领域的任何已知方法可用于将本文所教导的多核苷酸、其变体或片段整合入嵌合基因中。

[0113] 在某些实施方案中,将本文所教导的多核苷酸、其片段和变体可操作地连接到所谓的“组成型启动子”可能是有利的。或者,将本文所教导的多核苷酸、其片段和变体可操作地连接到所谓的“诱导型启动子”可能是有利的。诱导型启动子可以是生理上(例如通过外部施用某些化合物)调节的启动子。

[0114] 在一个实施方案中,与本文所教导的分离的多核苷酸、其变体或片段可操作连接的启动子可以是例如组成型活性启动子,例如:花椰菜花叶病毒(CaMV)分离株CM 1841(Gardner et al.,1981,Nucleic Acids Research 9,2871-2887)的强组成型35S启动子或增强型35S启动子(“35S启动子”)、CabbB-S(Franck et al.,1980,Cell 21,285-294)和CabbB-JI(Hull and Howell,1987,Virology 86,482-493);Ode11等人(1985,Nature 313,810-812)或US5164316中描述的35S启动子,来自泛素家族的启动子(例如Christensen et al.,1992,Plant Mol.Biol.18,675-689,EP 0 342 926,也参见Cornejo et al.1993,Plant Mol.Biol.23,567-581的玉米泛素启动子)、gos2启动子(de Pater et al.,1992Plant J.2,834-844)、emu启动子>Last et al.,1990,Theor.Appl.Genet.81,581-588)、拟南芥肌动蛋白启动子(如An et al.描述的启动子(1996,Plant J.10,107.))、水稻肌动蛋白启动子(如Zhang et al.描述的启动子(1991,The Plant Cell 3,1155-1165))和US 5,641,876中描述的启动子或如W0070067中所述的水稻肌动蛋白2启动子;木薯叶脉花叶病毒启动子(W0 97/48819,Verdaguier et al.1998,Plant Mol.Biol.37,1055-1067)、来自地三叶草矮缩病毒的pPLEX系列启动子(W0 96/06932,特别是S7启动子)、乙醇脱氢酶启动子(例如pAdh1S(GenBank登录号X04049、X00581))和分别驱动表达T-DNA的1' 和2' 基因的

TR1' 启动子和TR2' 启动子(分别为“TR1' 启动子”和“TR2' 启动子”)(Velten et al.,1984, EMBO J 3,2723-2730)、US6051753和EP426641中描述的玄参花叶病毒启动子、组蛋白基因启动子(如来自拟南芥的Ph4a748启动子(PMB 8:179-191)),或其他。

[0115] 由于嵌合基因、基因构建体或载体在植物中的组成型表达可能对植物的健康高度不利(high cost),因此在一个实施方案中优选使用其活性可被诱导的启动子。诱导型启动子的实例是创伤诱导型启动子,如Cordera等人(1994,The Plant Journal 6,141)描述的MPI启动子,其由创伤(例如由昆虫或物理伤害引起的)或COMPTII启动子(W00056897)或US6031151中描述的PR1启动子诱导。或者,启动子可通过化学品如Aoyama和Chua(1997, Plant Journal 11:605-612)和US6063985中所述的地塞米松或通过四环素(TOPFREE或TOP 10启动子,参见Gatz,1997,Annu Rev Plant Physiol Plant Mol Biol.48:89-108and Love et al.2000,Plant J.21:579-88)而诱导。

[0116] 可使用非组成型的、而对植物的一种或多种组织或器官特异性的启动子。优选的启动子是组织特异性的。启动子可优选在发育上受到调节,例如叶优选的或表皮优选的,由此所述核酸序列仅在或优先在特定组织或器官的细胞中表达,和/或仅在特定发育阶段期间表达,优选在雌性子房、大孢子母细胞和/或雌配子中。例如,通过将编码序列置于光诱导启动子如植物本身或另一植物(如豌豆(如美国专利5,254,799中公开的)或如US5034322及其他专利公开的拟南芥属植物)的核酮糖-1,5-二磷酸羧化酶小亚基基因启动子的控制下,可在植物叶片中选择性表达Dip基因。

[0117] 术语“诱导型”不一定要求启动子在没有诱导物刺激时完全失活。可能存在低水平的非特异性活性,只要这不会导致植物的产量或品质的大量损失。因此,诱导型优选是指增加启动子活性,导致与诱导物接触后下游编码区的转录增加。

[0118] 在一个优选的实施方案中,内源基因的启动子用于表达包含本文所教导的SEQ ID NO:3的氨基酸序列或其变体或片段(例如SEQ ID NO:7和/或12)的蛋白质。例如,根据本发明,蒲公英Dip等位基因的启动子或来自另一植物物种的相应的启动子可被分离并且可操作地连接至编码本发明的蛋白质的核酸序列。所述蛋白质优选能够提供倍数孢子体形成功能,优选作为倍数孢子体形成或配子体无融合生殖的一部分。所述启动子,即通常在编码包含本文所教导的SEQ ID NO:3的氨基酸序列或其片段或变体(例如SEQ ID NO:7和/或12)(例如其他蒲公英属来源的和/或其他植物的同源物)的蛋白质的多核苷酸的转录起始位点和/或翻译起始密码子的上游约2000个碱基对(bp)内的上游转录调节区,可使用已知方法如TAIL-PCR(Liu et al.1995,Genomics 25(3):674-81;Liu et al.2005,Methods Mol.Biol.286:341-8)、Linker-PCR或反向PCR(IPCR)从无融合生殖植物和/或其他植物中分离。应该理解的是,由于所述基因序列是广义蒲公英属的推定的液泡蛋白质分选相关蛋白基因Vps13(SEQ ID 1)的一部分,所述启动子包含位于基因编码区5'端(SEQ ID 2)的SEQ ID 1或位于一个表达为mRNA、miRNA或siRNA的表达的亚基因组区域的5'端的SEQ ID 1的其他区域的序列。表达的mRNA、siRNA或miRNA将包括雌性配子体阶段,即其表达活性可追溯到表达倍数孢子体形成表型的位置和时间或导致该阶段的发育阶段。

[0119] 在本发明的一个实施方案中,可使用源自编码包含本文中教导的SEQ ID NO:3的氨基酸序列或其片段或变体的蛋白质的多核苷酸的内源启动子,如其他蒲公英起源的和/或其他植物的同源物。也可使用比这些序列更长的序列。对于任何所述核酸序列,编码区的

翻译起始密码子上游至约2000bp的区域可包含转录调节元件。因此,在一个实施方案中,可分离所述多核苷酸的翻译或转录起始位点上游2000bp、1500bp、1000bp、800bp、500bp、300bp或更少的核苷酸序列,并且可测试其启动子活性,且如果具有功能性,该序列可以可操作地连接至编码包含本文所教导的SEQ ID NO:3的氨基酸序列或其片段或变体(例如SEQ ID NO:7和/或12)的蛋白质的多核苷酸。整个序列及其片段的启动子活性可通过例如缺失分析来测试,通过删除转录起始位点区域的5'和/或3'端,并使用已知方法(例如将启动子与报告基因的缺失或多个缺失可操作地连接)测试启动子活性。

[0120] 在另一个实施方案中,所述启动子驱动本发明的miRNA和siRNA分子的表达。

[0121] 在本发明所述的植物或植物细胞或有性作物中,源自具有倍数孢子体形成功能的Dip等位基因是否具有提供或诱导倍数孢子体形成(优选作为配子体无性生殖的一部分)的能力,可取决于多肽的分子功能或如本文所教导的由分离的多核苷酸编码的蛋白质的分子功能。在一个实施方案中,由本文所教导的分离的多核苷酸、其片段和变体编码的蛋白质可具有显性功能,其功能通过表达或过表达包含SEQ ID NO:3的氨基酸序列或其变体或片段的蛋白质(例如SEQ ID NO:7和/或12)提供。当在植物中表达时,编码所述蛋白质的所述分离的多核苷酸能够向植物提供倍数孢子体形成功能或增强植物中的倍数孢子体形成功能,或者能够在植物或植物细胞或作物中诱导或增强倍数孢子体形成。

[0122] 例如,当包含SEQ ID NO:1和/或SEQ ID NO:2或其片段或变体(例如SEQ ID NO:4、5、6或11)的核酸序列的多核苷酸通过来自合适的植物启动子在植物中表达、并且具有功能性数量的编码蛋白质被合成时,与缺乏所述蛋白质的植物相比,优选作为配子体无融合生殖一部分的倍数孢子体形成功能或倍数孢子体形成的发生可被诱导或显著增强。可通过将这样的核酸序列引入合适的宿主植物(例如非二倍体蒲公英品系)中使其在宿主植物中表达,来测试功能(即本文所教导的多核苷酸、其变体或片段在植物中诱导或引起倍数孢子体形成的能力),以及分析转化体在生物测定(如本文所教导的实例中所述)中对倍数孢子体形成功能的影响。

[0123] 在一个实施方案中,沉默本文所教导的能够编码包含SEQ ID NO:3的氨基酸序列或其变体或片段(例如SEQ ID NO:7和/或12)的蛋白质的表达的多核苷酸、其变体或片段,可能导致功能丧失,即减少的倍数孢子体形成或缺失的倍数孢子体形成或不发生通过倍数孢子体形成的无配子无融合生殖。因此,技术人员可以容易地确定编码包含SEQ ID NO:3的氨基酸序列或其片段或变体(例如SEQ ID NO:7和/或12)和/或其(如本文所述能够优选提供倍数孢子体形成作为植物或植物细胞或作物中配子体无融合生殖的一部分的)片段的蛋白质的多核苷酸或其变体或片段。

[0124] 在一个实施方案中,提供了如本文所教导的嵌合基因,其包含任何一种分离的多核苷酸(SEQ ID NO:1或SEQ ID NO:2)、如本文所教导的其变体或片段(例如SEQ ID NO:4、5、6或11)。所述嵌合基因优选能够向本发明所述的植物或植物细胞或作物中的植物提供倍数孢子体形成功能。

[0125] 在一个实施方案中,如本文所教导的多核苷酸(例如SEQ ID No:1或SEQ ID NO:2)、如本文所教导的其变体或片段(例如SEQ ID NO:4、5、6或11)或如本文所教导的嵌合基因可被包含在遗传构建体中。

[0126] 在一个优选的实施方案中,本文所教导的遗传构建体可包含本发明的分离的多核

昔酸(例如SEQ ID NO:2)、如本文所教导的其变体或片段(例如SEQ ID NO:4、5、6或11)的开放阅读框。

[0127] 在一个实施方案中,本文所教导的分离的多核苷酸(例如SEQ ID NO:1或SEQ ID NO:2)、如本文所教导的其变体或片段(例如SEQ ID NO:4、5、6或11)可被包含在核酸载体内。

[0128] 本发明所述的嵌合基因、遗传构建体和载体的构建通常是本领域已知的。所述嵌合基因、遗传构建体和载体优选能够向植物提供倍数孢子体形成功能,或者能够在植物、植物细胞或作物中诱导倍数孢子体形成或诱导通过倍数孢子体形成的配子体无融合生殖。嵌合基因可通过修饰内源基因序列而产生。例如,可修饰隐性等位基因(即dip),使其变为显性等位基因(即Dip),从而显性等位基因能够提供倍数孢子体形成功能或能够在植物、植物细胞或作物中的诱导倍数孢子体形成或诱导通过倍数孢子体形成的配子体无融合生殖。或者,替代地,可修饰例如能够提供倍数孢子体形成功能或能够诱导倍数孢子体形成或诱导通过倍数孢子体形成的配子体无融合生殖但不表达的内源基因,即通过修饰内源启动子序列使得内源基因被表达。此类修饰可包括(靶向)诱变,由此内源基因的至少1、5、10、20、50、100、200、500或1000个核苷酸发生突变。在实施例5中可找到这种修饰的一个实例,其中在EMS突变中发现的四个突变赋予倍数孢子体形成表型的丧失,因此,逆转所述突变可提供倍数孢子体形成表型的获得。

[0129] 在一个实施方案中,可通过使用标准分子生物学技术将编码本发明所述的蛋白质(或变体或片段)的核酸序列可操作地连接至适于在宿主细胞中表达的启动子序列来产生本文所教导的嵌合基因。启动子序列可能已经存在于载体中,因此核酸序列被简单地插入载体中的启动子序列下游。在一个实施方案中,嵌合基因包含用于在植物细胞或微生物细胞(例如细菌)中表达的合适启动子,其可操作地连接至本发明中的核酸序列,任选地之后接着一个3'非翻译核酸序列。本发明所述的核酸序列任选地位于5'非翻译序列区(UTR)之后。如下所述,启动子、3'UTR和/或5'UTR可例如来自内源性的Dip基因,或者可来自其他来源。此外,本发明所述的核酸序列还可包含内含子序列,其可包含在3'UTR或5'UTR序列中,但也可被引入本发明的核酸序列的编码序列中。

[0130] 在一个实施方案中,如本文所教导的嵌合基因、遗传构建体和载体优选能够表达编码本发明的氨基酸序列的核酸序列,其中本发明的所述氨基酸序列优选能够向植物提供倍数孢子体形成功能,优选作为植物或植物细胞或作物中配子体无融合生殖的一部分。因此,所述嵌合基因、遗传构建体和载体优选包含本发明的显性Dip等位基因。

[0131] 在一个实施方案中,如本文所教导的核酸载体可包含在植物细胞中有活性的启动子序列,所述启动子序列与分离的多核苷酸(例如SEQ ID NO:1或SEQ ID NO:2)、如本文所教导的其变体或片段(例如SEQ ID NO:4、5、6或11),或如本文所教导的嵌合基因或如本文所教导的遗传构建体可操作地连接。

[0132] 在一个优选的实施方案中,如本文所教导的核酸载体的启动子序列可包含:

[0133] a) SEQ ID NO:1和/或SEQ ID NO:2的核酸序列的天然启动子序列;

[0134] b) a)的启动子序列的功能性片段;或者

[0135] c)包含与SEQ ID NO:1和/或SEQ ID NO:2的核酸序列的天然启动子序列具有至少70%、优选至少80%、更优选至少90%、最优选至少95%序列同一性的核酸序列;

[0136] d) SEQ ID NO:6的核酸序列的天然启动子序列;

[0137] e) d)的启动子序列的功能性片段;或者

[0138] f) 包含与SEQ ID NO:6的核酸序列的天然启动子序列具有至少70%、优选至少80%、更优选至少90%、最优选至少95%序列同一性的核酸序列。

[0139] 在优选的实施方案中,如本文所教导的核酸载体的启动子是雌性子房特异性启动子,优选大孢子母细胞特异性启动子和/或雌性配子特异性启动子。

[0140] 分离的多肽

[0141] 在第三方面,本发明涉及分离的多肽,其包含SEQ ID NO:3的氨基酸序列或与SEQ ID NO:3的氨基酸序列具有至少50%或70%、优选至少80%、更优选至少90%、甚至更优选至少95%、甚至更优选至少96%或97%、最优选至少98%或99%序列同一性的氨基酸序列。

[0142] 在一个优选的实施方案中,本文所教导的多肽具有SEQ ID NO:3的氨基酸序列或其变体或片段。

[0143] 在一个实施方案中,如本文所教导的分离的多肽包含如上教导的SEQ ID NO:3的氨基酸序列及其变体或片段,其可称为DIP多肽或蛋白质或无融合生殖相关的多肽或蛋白质。

[0144] 在一个实施方案中,如上文所教导的DIP多肽或蛋白质及其变体或片段能够向植物或植物细胞提供倍数孢子体形成功能,优选作为在作物中诱导倍数孢子体形成或配子体无融合生殖的一部分。因此,在一个实施方案中,如本文所教导的分离的多肽或蛋白质可用于产生遗传上与亲代植物相同的后代,而不需要受精和杂交。

[0145] 在一个优选的实施方案中,如上文所教导的优选作为配子体无融合生殖体的一部分的DIP多肽或蛋白质及其变体或片段能够为植物或植物细胞提供倍数孢子体形成功能或能够在作物中(特别是当引入植物或植物细胞时)诱导倍数孢子体形成。

[0146] 具有本文所教导的SEQ ID NO:3的氨基酸序列或其变体的多肽或蛋白质被鉴定为广义蒲公英属的推定的液泡蛋白质分选相关蛋白基因Vps13或其部分。Vps13基因是一个大基因。因此,SEQ ID NO:3的所述氨基酸序列可包含在单个分离的蛋白质中,即作为同一个氨基酸序列中的一部分,或者该同一个氨基酸序列中的多个部分。因此分离的蛋白质可包含SEQ ID NO:3或其变体二者。应当理解,由于Vps13基因可构成大蛋白质,当与SEQ ID NO:3的氨基酸序列或其变体的大小相比时,序列同一性的百分比可不相对于分离的蛋白质的完整序列。相反,仅包含在所述分离的蛋白质中的氨基酸序列可与SEQ ID NO:3具有所述百分比的序列同一性。因此应理解,应相对于包含在分离的蛋白质中的氨基酸序列计算序列同一性的百分比,其中氨基酸序列的第一个和最后一个氨基酸与SEQ ID NO:3氨基酸序列对齐。因此,当要计算序列同一性百分比时,优选仅相对于对应于SEQ ID NO:3的序列。

[0147] 应理解的是,本文所教导的多肽还包括具有SEQ ID NO:3的氨基酸序列的变体多肽,所述变体的氨基酸序列与SEQ ID NO:3的氨基酸序列具有超过50%、优选超过55%、超过60%、超过65%、超过70%、优选超过75%、超过80%、超过85%、超过90%、超过95%、优选超过96%、优选超过97%、优选超过98%、优选超过99%的序列同一性。具有SEQ ID NO:3的氨基酸序列的变体多肽还包括已经通过一个或多个氨基酸取代、缺失或插入从而具有SEQ ID NO:3的氨基酸序列的多肽衍生的多肽。优选地,与具有SEQ ID NO:3的氨基酸序列的多肽相比,这样的多肽包含1、2、3、4、5、6、7、8、9、10或更多至100、90、80、70、60、50、45、

40、35、30、25、20、15个氨基酸取代、缺失或插入。

[0148] 在一个实施方案中,本文所教导的变体多肽可通过一个或多个氨基酸缺失、插入和/或取代而与所提供的氨基酸序列不同,并且包括天然和/或合成/人造的变体。

[0149] 在一个实施方案中,术语“变体多肽”还包括自然界中存在的(例如在栽培的或野生的莴苣植物和/或其他植物中)的天然变体多肽。分离的蛋白质还包括分离的蛋白质的片段,即非全长肽。片段包括包含或由SEQ ID NO:3编码的氨基酸序列的至少10、20、30、40、50、100、150、200、250、300、400、500、600、700、800、900、1000或更多个连续氨基酸组成的肽或其变体,尤其包含或由SEQ ID NO:3的10、20、30、40、50、100、150、200、250、300、400、500、600、700、800、900、1000或更多个连续氨基酸组成或其变体。

[0150] 如本文所教导的分离的多肽或其变体优选能够向植物提供倍数孢子体形成功能,优选能够在植物或植物细胞或作物中诱导倍数孢子体形成或配子体无融合生殖。倍数孢子体形成是。这意味着根据本发明的分离的多肽、片段和变体能够诱导倍数孢子体形成。本发明的倍数孢子体形成功能包括跳过第一个雌性减数分裂(减数分裂I),导致两个未减数的与母本植物具有相同基因型的大孢子。其中一个大孢子退化,且另一个存活的未减数的大孢子产生未减数的雌配子体(或胚囊),其含有未减数的卵细胞。这种未减数的卵细胞在未受精的情况下发育成胚,并具有与母本植物相同的基因型,即母本植物的克隆。

[0151] 在一个实施方案中,如本文所教导的分离的多肽或其变体可从天然来源分离,其通过化学合成重新合成(使用例如肽合成仪,例如由Applied Biosystems提供)或通过重组宿主细胞通过表达编码如本文所教导的分离的多肽、其片段和变体的核酸序列产生。

[0152] 在一个实施方案中,如本文所教导的分离的多肽或其变体可包含以下类别内的保守氨基酸取代:

[0153] 基本的(例如Arg、His、Lys);

[0154] 酸性的(例如Asp、Glu);

[0155] 非极性的(例如Ala、Val、Trp、Leu、Ile、Pro、Met、Phe、Trp);或者

[0156] 极性的(例如Gly、Ser、Thr、Tyr、Cys、Asn、Gln)。

[0157] 另外,非保守氨基酸取代也可落入本发明的范围内。

[0158] 在一个实施方案中,如本文所教导的分离的多肽或其变体也可以是嵌合多肽,例如由至少两个不同结构域组成的多肽。由于SEQ ID NO:3来源于Vps13基因或部分来源于Vps13基因,所以SEQ ID NO:3或其变体可与Vps13蛋白中的相应序列交换,该Vps13蛋白不提供或以较差的能力提供倍数孢子体形成功能或不能在植物或植物细胞或作物中通过倍数孢子体形成诱导配子体无融合生殖。这样,可获得能够提供倍数孢子体形成功能或改善的功能,或者在植物或植物细胞或作物中能够进行倍数孢子体形成或改良的倍数孢子体形成的嵌合多肽或蛋白质。如本文所教导的嵌合多肽还可具有SEQ ID NO:3的一部分或多部分氨基酸序列。此外,如本文所教导的嵌合多肽可包含一种蛋白质(例如获自蒲公英属或另一种植物物种)的N-末端和另一种蛋白质(例如获自蒲公英属或另一种植物物种)的中间结构域和/或C-末端结构域。这种嵌合蛋白质可能比天然蛋白质具有改善的倍数孢子体形成功能,或有助于改善诱导或可帮助改善植物或植物细胞或作物中的倍数孢子体形成。

[0159] 氨基酸序列同一性可通过本领域可获得的任何合适的手段来确定。例如,氨基酸序列同一性可通过使用如上定义的Needleman和Wunsch算法和GAP默认参数的成对比对来

确定。还应理解的是,可使用许多方法来鉴定、合成或分离本文所教导的多肽的变体,诸如免疫印迹、免疫组织化学、ELISA、氨基酸合成等。

[0160] 还应理解,如本文所教导的DIP多肽的任何变体或片段与本文所教导的DIP多肽发挥相同的功能和/或具有与DIP多肽相同的活性。任何DIP多肽或其变体的功能性或活性可通过本领域中任何已知的、本领域技术人员认为适用于这些目的的方法来确定,。

[0161] 在一个实施方案中,如本文所教导的多肽(SEQ ID NO:3)或其变体的片段能够向能够诱导倍数孢子体形成或配子体无融合生殖的植物或植物细胞提供倍数孢子体形成功能。

[0162] 在一个实施方案中,如本文所教导的多肽及其变体的片段可具有所述多肽的至少10、20、30、40、50、100、150、200、250、300、400或500个连续氨基酸。

[0163] 在一个实施方案中,如本文所教导的多肽及其变体的片段具有SEQ ID NO:7和/或12的氨基酸序列。

[0164] 方法

[0165] 另一方面,本发明涉及一种用于生产无融合生殖种子的方法,其包括以下步骤:

[0166] a) 用任何本文所教导的多核苷酸(例如SEQ ID NO:1或SEQ ID NO:2)或其变体或片段(例如SEQ ID NO:4、5、6或11),或本文所教导的嵌合基因,或本文所教导的遗传构建体和/或本文所教导的核酸载体转化植物、植物部分或植物细胞,以产生初级转化体;

[0167] b) 从所述初级转化体生长开花植物和/或花,由此如上所述的多核苷酸、变体或片段、嵌合基因、构建体和/或载体至少在雌性子房中存在和/或表达,优选在大孢子母细胞和/或雌性配子中存在和/或表达;和

[0168] c) 授粉所述初级转化体以诱导产生种子,优选使用四倍体植物的花粉或所述初级转化体的自花花粉。

[0169] 应该理解,当所述初级转化体发育自主胚乳时,可省略步骤c)。

[0170] 在一个实施方案中,通过本文所教导的方法获得的无融合生殖种子是如本文所教导的初级转化体的克隆。

[0171] 在步骤(a)的一个实施方案中,植物或植物部分可用包含任何如本文所教导的多核苷酸(例如SEQ ID NO:1或SEQ ID NO:2)或其变体或片段的(例如SEQ ID NO:4或SEQ ID NO:5)的嵌合基因转化。

[0172] 在一个优选的实施方案中,嵌合基因包含SEQ ID NO:2。

[0173] 在一个实施方案中,嵌合基因可包含在本发明的遗传构建体或载体中。

[0174] 在另一个实施方案中,嵌合基因还可包含已被修饰的内源基因。这种修饰可包括通过靶向诱变或使用核酸酶如Crispr/Cas进行修饰,但不限于此。当被引入所述植物、植物部分或植物细胞时,所述嵌合基因优选能够提供倍数孢子体形成功能或能够在植物、植物部分或植物细胞中诱导倍数孢子体形成或诱导通过倍数孢子体形成的配子体无融合生殖。载体可用于转化在核基因组中或在质体、线粒体或叶绿体DNA中插入嵌合基因的宿主细胞,并且使用合适的启动子可表达该载体(例如Mc Bride et al.,1995Bio/Technology 13, 362;US 5,693,507)。质体基因组转化的一个优点是转基因(一种或多种)扩散的风险可降低。质体基因组转化可按照本领域已知的方式进行,参见例如Sidorov VA et al.1999, Plant J.19:209-216or Lutz KA et al.2004,Plant J.37(6):906-13。

[0175] 在一个实施方案中,如上文所教导的包含在嵌合基因中的本文所教导的多核苷酸或变体或片段可操作地连接至启动子序列,其中启动子序列包含:

[0176] (a) SEQ ID NO:1和/或SEQ ID NO:2的核酸序列的内源启动子序列;

[0177] (b) 所述天然启动子序列的功能性片段;

[0178] (c) 与SEQ ID NO:1或SEQ ID NO:2的核酸序列的内源启动子序列包含至少70%序列同一性的核酸序列;或者

[0179] (d) (c) 的核酸序列的功能性片段;

[0180] e) SEQ ID NO:6的核酸序列的天然启动子序列;

[0181] f) d) 的启动子序列的功能性片段;

[0182] g) 包含与SEQ ID NO:6的核酸序列的天然启动子序列具有至少70%,优选至少80%,更优选至少90%,最优选至少95%序列同一性的核酸序列;或者

[0183] h) g) 的核酸序列的功能性片段。

[0184] 应该理解,如上所述,本发明的嵌合基因可代表显性等位基因。因此,用这种显性嵌合基因转化植物、植物部分或植物细胞将足以为所述植物、植物部分或植物细胞提供倍数孢子体形成功能或在所述植物、植物部分或植物细胞中诱导倍数孢子体形成或通过倍数孢子体形成的配子体无融合生殖。

[0185] 在一个实施方案中,提供了能够编码如本文所教导的蛋白质(SEQ ID NO:3)或其变体或片段(例如SEQ ID NO:7和/或12)的多核苷酸,并且其能够向植物、植物部分或植物细胞提供倍数孢子体形成功能,或者如上所述在所述植物、植物部分或植物细胞中通过倍数孢子体形成诱导配子体无融合生殖。此类多核苷酸可用于制备嵌合基因,并且包含这些的载体用于将嵌合基因转移入宿主细胞并在宿主细胞(例如细胞、组织、器官或源自转化的细胞的生物体)中产生蛋白质。用于在植物细胞中产生所述蛋白质(或蛋白质片段或变体)的载体在本文中称为“表达载体”。宿主细胞优选植物细胞。

[0186] 任何植物都可能是合适的宿主,但最优选的宿主是可从增强或减低的倍数孢子体形成中受益的植物物种。特别优选具有其他良好农艺学特性的品种或育种品系。通过产生转基因植物并诱导倍数孢子体形成以及合适的对照植物,容易测试本文提供的基因和/或蛋白质(或其变体或片段)是否赋予宿主植物以所要求的倍数孢子体形成的增加。

[0187] 在一个实施方案中,合适的宿主植物可选自玉米/玉米(玉蜀黍属)、小麦(小麦属)、大麦(例如大麦芽)、燕麦(例如燕麦)、高粱(*Sorghum bicolor*)、黑麦(*Secale cereale*)、大豆(大豆属,例如G.max)、棉花(棉属,例如陆地棉、海岛棉)、芸苔属(例如甘蓝型油菜、芥菜、甘蓝、白菜等)、向日葵(油葵)、红花、山药、木薯、苜蓿(紫花苜蓿)、稻(稻属,例如籼稻种群或粳稻种群)、牧草、珍珠稗(狼尾草属,例如P.glaucum)、树种(松树、杨树、杉木、车前草等)、茶、咖啡、油棕、椰子、蔬菜种(如豌豆、西葫芦、豆类(例如菜豆)、辣椒、黄瓜、朝鲜蓟、芦笋、茄子、西兰花、大蒜、韭葱、生菜、洋葱、萝卜、芜菁、番茄、马铃薯、甘蓝、胡萝卜、花椰菜、菊苣、芹菜、菠菜、莴苣菜、茴香、甜菜)、肉质果实植物(葡萄、桃子、李子、草莓、芒果、苹果、李子、樱桃、杏、香蕉、黑莓、蓝莓、柑橘、猕猴桃、无花果、柠檬、酸橙、油桃、覆盆子、西瓜、桔子、葡萄柚等)、观赏植物(如玫瑰、矮牵牛、菊花、百合、非洲菊属)、草药(薄荷、欧芹、罗勒、百里香等)、木本植物(如杨属、柳属、栎属、桉属)、纤维种例如亚麻(*Linum usitatissimum*)和大麻(*Cannabis sativa*)。

[0188] 在一个优选的实施方案中,宿主植物可以是选自蒲公英属、莴苣属、豌豆属、辣椒属、茄属、黄瓜属、玉蜀黍属、棉属、大豆属、蓖麻属、稻属和高粱属的植物物种。

[0189] 在一个实施方案中,多核苷酸(SEQ ID NO:1或SEQ ID NO:2),其变体或片段(例如SEQ ID NO:4、5、6或11)优选包含于本发明的嵌合基因中,并且能够编码蛋白质(SEQ ID NO:3)或其变体或片段(例如SEQ ID NO:7和/或12),并且能够向植物、植物部分或植物细胞提供倍数孢子体形成功能,或在植物、植物部分或植物细胞中诱导倍数孢子体形成或通过倍数孢子体形成的配子体无融合生殖,其能够以常规方式稳定地插入到单个植物细胞的核基因组中,并且如此转化的植物细胞可以传统方式用于产生转化植物,所述转化植物由于在特定时间所述蛋白质出现在特定细胞中而具有改变的表型。就此而言,包含如本文所教导的多核苷酸、其变体或片段的T-DNA载体能够编码如本文所教导的蛋白质或变体或片段,其能够提供倍数孢子体形成功能或诱导倍数孢子体形成或诱导通过倍数孢子体形成的配子体无融合生殖,在根癌土壤杆菌中可用于转化植物细胞,且其后转化的植物可使用例如在EP 0 116 718、EP 0 270 822、PCT公开文本W084/02913和公布的欧洲专利申请EP 0 242 246和在Gould et al. (1991, Plant Physiol. 95, 426-434) 中描述的步骤从转化的植物细胞再生转化植物。用于农杆菌介导的植物转化的T-DNA载体的构建是本领域公知的。T-DNA载体可以是如EP 0 120 561和EP 0 120 515中所述的双元载体,或可通过同源重组整合到农杆菌Ti-质粒中的共整合载体,如EP 0 116 718中所述。莴苣转化方案已经在例如Michelmore等人,1987和Chupeau等人,1989中描述。

[0190] 优选的T-DNA载体各自含有可操作地连接至功能性地编码能够提供倍数孢子体形成的蛋白质(例如编码SEQ ID NO:3或其变体或片段(例如SEQ ID NO:7和/或12))的核酸序列的启动子。启动子与所述核苷酸序列或T-DNA边界序列之间的序列可操作地连接,或者至少位于右边界序列的左侧。边界序列在Gielen等人(1984, EMBO J 3, 835-845)中描述。当然,可使用其他类型的载体转化植物细胞,例如使用诸如直接基因转移(如在EP 0 223 247中所述)、花粉介导的转化(例如在EP 0 270 356和W085/01856中所描述的)、例如在US 4, 684, 611中描述的原生质体转化、植物RNA病毒介导的转化(例如在EP 0 067 553和US 4, 407, 956中所描述的)、脂质体介导的转化(例如在US 4, 536, 475中所描述的)的方法和其他方法。可使用已知的方法如电穿孔或三亲杂交来将T-DNA载体导入农杆菌。

[0191] 同样,从转化细胞中选择和再生转化植物是本领域众所周知的。显然,对于不同的物种,甚至对于单个物种的不同品种或栽培种,可对操作程序进行具体调整以便以高频率再生转化体。

[0192] 通过本文所教导的方法可获得的植物或植物部分或植物细胞具有改变的倍数孢子体形成水平,特别是包含显著提高的倍数孢子体形成水平的转基因植物。如下文进一步描述的,这样的植物可使用不同的方法制备。

[0193] 通过本发明的方法获得的或通过本发明的方法可获得的植物可用于常规植物育种方案中以产生如本文所教导的含有转基因的更多的转化植物。单拷贝转化植物可使用例如Southern印迹分析或基于PCR的方法或Invader®技术测定(Third Wave Technologies, Inc.)选择。通过嵌合基因的存在可将转化的细胞和植物与未转化的细胞和植物区分开来。也可对转基因插入位点的植物DNA的侧翼序列进行测序,由此可开发“事件特异性”检测方法用于常规使用。参见例如W00141558,其描述了基于例如整合序列和侧翼(基因组)序列的

精英事件检测试剂盒(例如PCR检测试剂盒)。

[0194] 在一个实施方案中,本文所教导的多核苷酸、其变体或片段能够在植物、植物部分或植物细胞中,对植物、植物部分或植物细胞提供倍数孢子体形成功能,或诱导倍数孢子体形成或诱导通过倍数孢子体形成的配子体无融合生殖,例如通过表达本发明的能够在植物、植物部分或植物细胞中提供倍数孢子体形成功能或诱导通过倍数孢子体形成的配子体无融合生殖的蛋白质、其变体或片段插入植物细胞基因组中,从而插入的编码序列位于一个可指导植物细胞中的表达的启动子的下游(即3')并处于其控制之下。这可优选地通过将嵌合基因插入植物细胞基因组中,特别是核或质体(例如叶绿体)基因组中完成。

[0195] 能够向植物提供倍数孢子体形成功能的本发明的核酸序列或与其相对应的序列优选插入植物基因组中,从而使得编码序列位于合适的3'末端非翻译区("3'末端"或3'UTR)上游(即5')。合适的3'末端包括CaMV 35S基因("3' 35S")、胭脂碱合酶基因("3' nos")(Depicker et al.,1982,J.Mol.Appl.Genetics 1,561-573.)、章鱼碱合酶基因("3' ocs")(Gielen et al.,1984,EMBO J 3,835-845)和T-DNA基因7("3' 基因7")(Velten and Schell,1985,Nucleic Acids Research 13,6981-6998),其在转化的植物细胞中起到3'-非翻译DNA序列的作用等。在一个实施方案中,蒲公英等位基因的3'UTR和/或5'UTR能够提供倍数孢子体形成功能,即使用包含SEQ ID NO:1和/或SED ID NO:2(或其变体或片段)。3'UTR和/或5'UTR也可在另一个实施方案中使用,因为它也可与其他编码区或其他核酸构建体组合使用。

[0196] DIP编码核酸序列可任选地作为杂合基因序列插入植物基因组中,由此能够为植物提供倍数孢子体形成功能的序列与编码可选择的或可评分的标记物的基因在可读框内连接(US 5,254,799;Vaeck et al.,1987,Nature 328,33-37),例如编码卡那霉素抗性的neo(或nptII)基因(EP 0 242 236),从而使该植物表达易于检测的融合蛋白。

[0197] 优选地,为了选择的目的但也为了控制杂草的选择,本发明的转基因植物还可用编码赋予对除草剂的抗性的蛋白质的DNA进行转化,所述除草剂例如广谱除草剂,例如基于活性成分为草铵膦的除草剂(例如Liberty®或BASTA;抗性由PAT或bar基因赋予;参见EP 0 242 236和EP 0 242 246)或草甘膦(例如RoundUp®;抗性由EPSPS基因赋予,参见例如EP0 508 909和EP 0 507 698)。使用除草剂抗性基因(或赋予所期望的表型的其他基因)作为选择性标记还具有可避免引入抗生素抗性基因的优点。

[0198] 或者,可使用其他可选择标记物基因,例如抗生素抗性基因。由于在转化的宿主植物中保留抗生素抗性基因可能不被接受,因此在选择转化体后可再次去除这些基因。存在用于去除转基因的不同技术。实现去除的一种方法是通过将嵌合基因与lox位点侧接,并且在选择后,用表达CRE重组酶的植物杂交转化的植物(参见例如EP506763B1)。位点特异性重组导致标记基因的切除。另一个位点特异性重组系统是EP686191和US5527695中描述的FLP/FRT系统。位点特异性重组系统如CRE/LOX和FLP/FRT也可用于基因叠加的目的。此外,已经描述了单组分切除系统,参见例如W09737012或W09500555。

[0199] 全部或部分本发明的核酸序列能够向植物提供倍数孢子体形成功能,例如因为它编码本发明的蛋白质,还可用于转化微生物,例如细菌(例如大肠杆菌、假单胞菌、土壤杆菌属、芽孢杆菌属等)、真菌或藻类或昆虫,或用于制备重组病毒。用整合入合适的克隆载体中的本发明的全部或部分核酸序列转化细菌以常可规方式进行,优选使用Maillon等人

(1989, FEMS Microbiol. Letters 60, 205-210) 和 WO 90/06999 中所述的常规电穿孔技术。为了在原核宿主细胞中表达, 可相应地优化核酸序列的密码子使用。应该去除内含子序列, 并且可按照已知的方式进行其他调整以获得最佳表达。

[0200] 本发明的核酸序列的 DNA 序列可以翻译中性方式进一步改变, 即对于氨基酸序列, 修饰存在于基因部分中的潜在抑制性 DNA 序列和/或通过引入密码子用法的改变, 例如, 如上所述, 调整密码子使用使之对于植物最优化, 优选特定的相关植物属。

[0201] 如上所述, 根据本发明的一个实施方案, 能够向植物提供倍数孢子体形成功能的本发明的蛋白质或嵌合蛋白靶向细胞内细胞器如质体, 优选叶绿体、线粒体, 并且也可以是从细胞分泌, 潜在地优化蛋白质稳定性和/或表达。同样, 蛋白质可能被定位到空泡。为此目的, 在本发明的一个实施方案中, 本发明的嵌合基因包含编码与本发明的蛋白质编码区连接的信号肽或靶肽的编码区。包含在本发明蛋白质中的特别优选的肽是用于叶绿体或其他质体靶向的转运肽, 特别是来自其基因产物被靶向至质体的植物基因的重复转运肽区域, Capellades 等人的优化的转运肽 (US 5,635,618)、来自菠菜的铁氧还蛋白-NADP+氧化还原酶的转运肽 (Oelmuller et al., 1993, Mol. Gen. Genet. 237, 261-272)、Wong 等人描述的转运肽 (1992, Plant Molec. Biol. 20, 81-93) 和公开的 PCT 专利申请 WO 00/26371 中的靶向肽。同样优选的是与此种胞外肽相连的蛋白质的信号分泌肽, 例如马铃薯蛋白酶抑制剂 II 的分泌信号 (Keil et al., 1986, Nucl. Acids Res. 14, 5641-5650)、水稻的阿尔法-淀粉酶 3 基因的分泌信号 (Sutliff et al., 1991, Plant Molec. Biol. 16, 579-591) 和烟草 PR1 蛋白的分泌信号 (Cornelissen et al., 1986, EMBO J. 5, 37-40)。本发明的特别有用的信号肽包括叶绿体转运肽 (例如 Van Den Broeck et al., 1985, Nature 313, 358), 或 US 5,510,471 和 US 5,635,618 的优化的引起蛋白质转运到叶绿体的叶绿体转运肽。也可使用分泌信号肽或将蛋白质靶向至其他质体、线粒体、内质网或另一种细胞器的肽。用于靶向细胞内细胞器或用于植物细胞外或向细胞壁分泌的信号序列见于天然靶向或分泌的蛋白质中, 优选 Klösgen 等人 (1989, Mol. Gen. Genet. 217, 155-161)、Klösgen 和 Weil (1991, Mol. Gen. Genet. 225, 297-304)、Neuhaus 和 Rogers (1998, Plant Mol. Biol. 38, 127-144)、Bih 等人 (1999, J. Biol. Chem. 274, 2284-22894)、Morris 等人 (1999, Biochem. Biophys. Res. Commun. 255, 328-333)、Hesse 等人 (1989, EMBO J. 8, 2453-2461)、Tavladoraki 等人 (1998, FEBS Lett. 426, 62-66.)、Terashima 等人 (1999, Appl. Microbiol. Biotechnol. 52, 516-523)、Park 等人 (1997, J. Biol. Chem. 272, 6876-6881)、Shcherban 等人 (1995, Proc. Natl. Acad. Sci USA 92, 9245-9249) 中描述的那些蛋白质。

[0202] 在一个实施方案中, 本发明的能够向植物提供倍数孢子体形成功能的几种编码蛋白质得核酸序列, 选择性地受不同启动子的控制在单一宿主中共表达。通过转化已经表达本发明蛋白质的植物, 或通过用本发明的不同蛋白质转化的植物进行杂交, 可容易地获得共表达宿主植物。因此, 本发明还提供了具有本发明的相同或不同的分离的核酸序列的多个核酸序列的植物或植物部分, 其中每个核酸序列都能够向植物提供倍数孢子体形成功能。可以理解的是, 在这方面术语多个是指每细胞。或者, 本发明的几种核酸序列 (其中的每种核酸序列都能够向植物提供倍数孢子体形成功能) 可存在于单个转化载体上, 或者可使用分开的载体同时共转化并选择包含多个嵌合基因的转化体。类似地, 编码能够提供本发明的倍数孢子体形成功能的蛋白质的一个或多个基因可与其他嵌合基因一起在单个植物

中表达,例如编码增强或抑制倍数孢子体形成的或者涉及无融合生殖的其他蛋白质。可以理解,不同的蛋白质可在相同的植物中表达,或者每种蛋白质可在单个植物中表达,并然后通过将单个植物相互杂交而在相同的植物中组合。例如,在杂交种子生产中,每个亲本植物可表达单一蛋白质。在杂交亲本植物以产生杂种后,两种蛋白质在杂交植物中组合。

[0203] 优选在不同启动子的控制下的多个本发明的嵌合基因产生的植物也是一个实施方案。这样,通过在合适的时间和位置表达合适量的本发明的能够向植物提供倍数孢子体形成功能的蛋白质可微调倍数孢子体形成表型的增强或抑制。这种微调可通过确定最合适的启动子和/或通过选择显示所期望的表达水平的转化“事件”来完成。

[0204] 本发明的表达所期望水平的能够提供倍数孢子体形成功能的蛋白质的转化体通过例如分析拷贝数(Southern印迹分析)、mRNA转录水平(例如使用引物对或侧翼引物的RT-PCR)或通过分析所述倍数孢子体形成蛋白在各种组织中的存在和水平(例如SDS-PAGE; ELISA测定等)而被选择。出于调控原因,优选选择单拷贝转化体,并分析嵌合基因插入位点侧翼序列,优选测序,以表征转化结果。选择表达高或中等表达DIP的转基因事件用于进一步发育,直到获得具有稳定Dip转基因的高性能精英事件。

[0205] 此外,可设想具有几个嵌合基因的植物可具有编码能够提供倍数孢子体形成功能的蛋白质的第一嵌合基因,和能够抑制或沉默第一嵌合基因的第二嵌合基因。所述第二嵌合基因优选处于诱导型启动子的控制下。这种植物可能是特别有利的,因为它允许控制倍数孢子体形成功能。通过诱导所述启动子的表达,植物中的倍数孢子体形成功能可能丧失。此外,也可通过在倍数孢子体形成植物中导入根据本发明的嵌合基因来获得或可获得这种控制,所述嵌合基因也能够抑制或沉默向植物提供倍数孢子体形成功能的内源基因(即其自然编码本发明的氨基酸)。

[0206] 通过选择本发明的核酸序列的保守核酸序列部分,宿主植物或植物部分中的等位基因可被沉默。如上所述,所述沉默可能导致抑制植物的倍数孢子体形成功能。因此,本文还涵盖包含嵌合基因的植物,所述嵌合基因包含与本发明的核酸序列的有义和/或反义DNA片段可操作连接的转录调控元件,并且所述嵌合基因能够表现出抑制性或增强的倍数孢子体形成。所述转录调节元件可以是合适的启动子,其可以是诱导型启动子。

[0207] 表达能够向本发明的植物提供倍数孢子体形成功能的一种或多种蛋白质的转化植物还可包含其他转基因,例如赋予疾病抗性或赋予对其他生物和/或非生物胁迫的耐受性的基因。为了获得具有“叠加的”转基因的这种植物,可将其它转基因引入转化的植物中,或者可随后用一种或多种其他基因转化已转化的植物,或者可使用几种嵌合基因来转化植物品系或品种。例如,几种嵌合基因可存在于单个载体上,或者可存在于共转化的不同载体上。

[0208] 在一个实施方案中,将以下基因与本发明所述的一种或多种嵌合基因组合:已知的疾病抗性基因(尤其是赋予对于坏死性病原体增强的抗性的基因)、病毒抗性基因、昆虫抗性基因、非生物胁迫抗性基因(例如耐旱性、耐盐性、耐热或耐冷性等)、除草剂抗性基因等。因此,叠加的转化体可具有对病原体抗性、昆虫抗性、线虫抗性、盐度、冷胁迫、热胁迫、水分胁迫等更广泛的生物和/或非生物压力的耐受性。另外,如上所述,在该实施方案中,沉默或抑制倍数孢子体形成功能方式可与基因表达方法结合在单一植物在中。

[0209] 应当理解,包含本发明嵌合基因的植物或植物部分优选不显示非期望的表型,例

如产量降低、对疾病(特别是对于坏死菌)增加的易感性或不希望的结构变化(矮化、变形)等,并且如果在初级转化植物中看到这些表型,则可通过常规方法去除这些表型。本文所述的任何植物对于本发明的嵌合基因可以是纯合的或半合的。

[0210] 另一方面,本发明涉及产生杂交植物的克隆的方法,其包括以下步骤:

[0211] a) 用本文所教导的植物的花粉异花受精有性繁殖的植物以产生F1杂种种子;

[0212] b) 至少在雌性子房中,优选在大孢子母细胞中和/或在雌配子中,选择包含和/或表达如本文所教导的多核苷酸或其变体或片段或本文所教导的多肽或其变体或片段的F1植物;

[0213] c) 任选地,使所述选定的F1植物授粉以诱导产生种子,优选与四倍体植物的花粉产生种子;和

[0214] d) 收获种子;和

[0215] e) 任选地,从所述种子生长杂种克隆植物。

[0216] 当选定的F1植物发育自主胚乳时,可省略步骤c)。

[0217] 在一个实施方案中,本文所教导的方法的步骤(e)的克隆是无融合生殖克隆。

[0218] 在一个实施方案中,本文所教导的方法包括获得所述杂种植物。

[0219] 另一方面,本发明涉及一种赋予植物、植物部分或植物细胞倍数孢子体形成或在植物、植物部分或植物细胞中诱导通过倍数孢子体形成的配子体无融合生殖的方法,其包括以下步骤:

[0220] a) 用本文所教导的任何多核苷酸、其变体或片段、本文所教导的嵌合基因、本文所教导的遗传构建体和/或本文所教导的核酸载体转化所述植物、植物部分或植物细胞;和

[0221] b) 任选地再生植物,其中所述多核苷酸、其变体或片段、基因、构建体和/或载体至少在雌性子房中,优选在大孢子母细胞中和/或在雌配子中存在和/或表达。

[0222] 在一个实施方案中,本文所教导的多核苷酸、其变体或片段整合到所述植物、植物部分或植物细胞的基因组中。

[0223] 在一个实施方案中,如本文所教导的方法包括获得倍数孢子体形成植物。

[0224] 另一方面,本发明涉及在植物、植物部分或植物细胞上赋予倍数孢子体形成或在植物、植物部分或植物细胞中诱导倍数孢子体形成或诱导通过倍数孢子体形成的配子体无融合生殖的方法,其包括步骤:

[0225] a) 在植物、植物部分或植物细胞中修饰内源多核苷酸、多核苷酸的变体或片段,优选为液泡蛋白质分选相关蛋白基因,从而使得修饰后的植物、植物部分或植物细胞包含如本文所教导的任何一种多核苷酸、其变体或片段;并且

[0226] b) r) 任选地,再生植物。

[0227] 在一个实施方案中,本文所教导的方法的步骤(a)的经修饰的多核苷酸、多核苷酸的变体或片段被表达和/或编码多肽。

[0228] 在一个实施方案中,本文教导的方法的步骤(a)的经修饰的多核苷酸或多核苷酸片段至少存在于雌性子房中,优选存在于大孢子母细胞中和/或雌性配子中。

[0229] 在一个实施方案中,本文教导的方法的步骤(a)的修改通过以下进行:

[0230] a) 在所述植物、植物部分或植物细胞中引入或表达至少一种位点特异性核酸酶,优选其中所述核酸酶选自:Cas9/RNA CRISPR核酸酶、锌指核酸酶、大范围核酸酶和TAL效应

子核酸酶;和/或通过

[0231] b) 使用寡核苷酸进行寡核苷酸定向诱变,优选其中所述寡核苷酸是单链寡核苷酸;和/或通过

[0232] c) 化学诱变,优选用甲磺酸乙酯。

[0233] 在一个实施方案中,本文所教导的方法包括获得倍数孢子体形成植物。

[0234] 在一个实施方案中,所述修饰,特别是在蒲公英属中,包含编码对应于SEQ ID NO:10所示的内源dip氨基酸序列的位置96-100的氨基酸残基GGGGW的核苷酸的缺失和/或编码对应于如SEQ ID NO:10所示的内源dip氨基酸序列的位置108-110的氨基酸残基PPT的核苷酸的缺失。在其他生物体中,编码对应于在蒲公英中发现的氨基酸残基GGGGW或PPT的氨基酸残基的核酸可能缺失。技术人员将能够鉴定待缺失的正确氨基酸残基以及编码这些氨基酸残基的相应核苷酸序列。

[0235] 在一个实施方案中,所述修改包括一个或多个,例如,如图2所示的dip(有性等位基因;SEQ ID NO:13)和Dip(倍数孢子体形成等位基因)核苷酸序列之间的所有差异。

[0236] 在一个实施方案中,通过本文所教导的方法可获得的任何转化植物的完整植物、种子、细胞、组织和后代包括在本文中,并且可在DNA中通过检测鉴定本文所教导的嵌合基因、遗传构建体或载体的存在,例如通过使用全基因组DNA作为模板和使用特定PCR引物对的PCR分析,例如,如上所述的针对序列SEQ ID NO:1和/或SEQ ID NO:2和/或SEQ ID NO:4、6或11或其变体设计的特异性引物对。还可开发“事件特异性”PCR诊断方法,其中PCR引物基于插入的嵌合基因的侧翼植物DNA,参见US6563026。类似地,可开发事件特异性AFLP指纹或RFLP指纹,其识别经转化或经修饰的植物或从其衍生的任何植物、种子、组织或细胞。

[0237] 植物和种子

[0238] 另一方面,本发明涉及包含本文教导的嵌合基因、本文教导的遗传构建体和/或本文教导的核酸载体的植物、植物部分或植物细胞,其中所述基因、构建体和/或载体至少存在于和/或至少在雌性子房中表达,优选在大孢子母细胞和/或雌性配子中存在和/或表达。

[0239] 在一个实施方案中,本文所教导的植物种子是无融合生殖种子。

[0240] 在一个实施方案中,本文所教导的种子是其发育的本文教导的植物的克隆。

[0241] 在一个优选的实施方案中,本文所教导的植物、植物部分、植物细胞或种子来自:蒲公英属、莴苣属、豌豆属、辣椒属、茄属、黄瓜属、玉蜀黍属、棉属、大豆属、小麦属、稻属、葱属、芸苔属、向日葵属、甜菜属、菊苣属、菊属、狼尾草属、黑麦属、大麦属、苜蓿属、菜豆属、蔷薇属、百合属、咖啡属、亚麻属、大麻属、木薯属、胡萝卜属、南瓜属、西瓜属和高粱属。

[0242] 用途

[0243] 另一方面,本发明涉及如本文所教导的任何分离的多核苷酸、其变体或片段用于在植物中诱导倍数孢子体形成的用途。

[0244] 另一方面,本发明涉及如本文所教导的任何分离的多核苷酸或其片段或变体用于预防多个基因、QTLs或转基因的分离的用途。

[0245] 另一方面,本发明涉及如本文所教导的任何分离的多核苷酸或其片段或变体用于基因叠加的用途。

[0246] 另一方面,本发明涉及如本文所教导的任何分离的多核苷酸或其片段或变体用于发育和/或鉴定倍数孢子体形成性状的标记物的用途。

[0247] 在一个实施方案中,如本文所教导的多核苷酸(SEQ ID NO:1或SEQ ID NO:2)、其变体或片段(例如SEQ ID NO:4、5、6或11)、其能够编码的蛋白质(SEQ ID NO:3)或其变体或片段(例如SEQ ID NO:7和/或12)以及编码能够提供倍数孢子体形成功能或诱导通过倍数孢子体形成的配子体无融合生殖的任何蛋白质和其变体的多核苷酸序列,可用作遗传标记物用于标记物辅助选择以下:能够提供蒲公英属物种(和/或其他植物物种)的倍数孢子体形成功能的等位基因,以及用于将不同或相同的倍数孢子体形成等位基因转移到感兴趣的植物和/或组合在感兴趣的植物中,和/或转移到和/或组合在可被用于与其中存在倍数孢子体形成等位基因(或变异体)的植物产生种内或种间杂种的植物中。

[0248] 基于这些序列可开发大量不同的标记物测定。标记物测定的开发通常涉及鉴定等位基因之间的多态性,以使得多态性是“标记”特定等位基因的遗传标记。然后将多态性用于标记物测定中。例如,根据本发明的SEQ ID NO:1和/或SEQ ID NO:2和/或SEQ ID NO:4、5、6或11或其变体的核酸序列可与倍数孢子体形成的存在、不存在、减少、抑制或增强相关。例如通过对倍数孢子体形成植物材料和/或非倍数孢子体形成植物材料筛选一个或多个此类序列以使将特定等位基因与不存在或存在倍数孢子体形成功能相关联来完成。因此,可合成PCR引物或探针,其从植物材料获得的样品(例如RNA、cDNA或基因组DNA样品)中检测SEQ ID NO:1和/或SEQ ID NO:2和/或SEQ ID NO:4、5、6或11或其变体或片段的存在或不存在。将序列或其部分进行比较,并且可鉴定可能与倍数孢子体形成相关的多态性标记物。然后可将多态性标记物(例如与Dip或dip等位基因连接的SNP标记)开发成用于在植物材料中筛选存在或不存在倍数孢子体形成等位基因的快速分子测定法。因此,这些“遗传标记物”的存在或不存在指示与其连接的Dip等位基因的存在,并且人们可用检测遗传标记物代替检测Dip等位基因。这些标记物的例子在实施例部分中被公开。

[0249] 优选地,使用简单和快速的标记物测定法,其能够快速检测样本中(例如DNA样本)的特定Dip或dip等位基因(例如赋予倍数孢子体形成的等位基因如Dip,对比无此作用的等位基因,例如dip)或等位基因的组合。因此,在一个实施方案中,提供了SEQ ID NO:1和/或SEQ ID NO:2的核酸序列或与其具有至少70%、80%、90%、95%、98%、99%或更多的核酸同一性的其变体或片段(SEQ ID NO:4或SEQ ID NO:5、6或11)、或其一个或多个片段在分子测定中的用途,用于在样本中确定Dip等位基因和/或dip等位基因的存在或不存在,和/或样品相对于所述等位基因是纯合的还是杂合的。

[0250] 这种测定可例如包括以下步骤:

[0251] (a) 提供倍数孢子体形成和非倍数孢子体形成植物材料和/或其核酸样品;

[0252] (b) 确定源自Vps13基因的核苷酸序列,例如,包括来自(a)的材料中的对应于SEQ ID NO:1和/或SEQ ID NO:2或其变体和/或片段(SEQ ID NO:4、5、6或11)的序列,从而鉴定核苷酸序列之间的多态性;

[0253] (c) 将多态性与植物的倍数孢子体形成特征相关联,从而将多态性与Dip基因座的倍数孢子体形成和非倍数孢子体形成等位基因相关联;

[0254] 所标识的相关多态性可任选地进一步用于步骤(d)

[0255] (d) 使用所述多态性标记物开发用于种质筛选或表征和MAS的标记物测定。

[0256] 因此,在本发明的一个实施方案中,提供了PCR引物和/或探针、分子标记物和用于检测来源于倍数孢子体形成基因的等位基因的DNA或RNA序列的试剂盒(即Dip和/或dip等

位基因)。能够基于在本领域广为人知的所述序列(或其变体)合成可从样品中扩增Dip和/或dip DNA(例如来自SEQ ID NO:1和/或SEQ ID NO:2的核酸序列或其变体或片段(例如,SEQ ID NO:4、5、6或11))的简并的或特异性PCR引物(参见Dieffenbach and Dveksler (1995)PCR Primer:A Laboratory Manual,Cold Spring Harbor Laboratory Press和McPherson et al.(2000)PCR-Basics:From Background to Bench,First Edition, Springer Verlag,Germany)。例如,那些序列(或补体链)中任何长度为9、10、11、12、13、14、15、16、18或更长的连续核苷酸可用作引物或探针。本发明的多核苷酸序列也可用作杂交探针。Dip基因/等位基因检测试剂盒可包含Dip和/或dip等位基因特异性引物和/或Dip和/或dip等位基因特异性探针。可使用相关方案用引物和/或探针检测样品中的Dip和/或dip DNA。例如,这样的检测试剂盒可用于确定植物是否已经用本发明的Dip基因(或其部分或变体)转化,或者用于筛选蒲公英种质和/或其他植物物种种质中Dip等位基因(或Dip同源物或直系同源物)的存在和任选存在的配合物测定。

[0257] 因此,在一个实施方案中,提供了检测植物组织中(例如在蒲公英组织或其核酸样品中)编码DIP蛋白的核苷酸序列是否存在的方法。该方法包括:

[0258] a) 获得植物组织样品,例如,蒲公英组织样品或其核酸样品,

[0259] b) 使用分子标记物测定分析核酸样品中与Dip等位基因关联的一个或多个标记物的存在或不存在,其中所述标记物测定检测SEQ ID NO:1和/或SEQ ID NO:2和/或SEQ ID NO:4、5、6或11,或与所述样品中包含至少70%核苷酸同一性的序列中的任一个,和任选地

[0260] c) 选择包含一种或多种所述标记物的植物(例如蒲公英植物)。

[0261] 倍数孢子体形成的进一步应用

[0262] 倍数孢子体形成是无融合生殖的元素,且倍数孢子体形成的基因可与孤雌生殖基因结合使用以产生无融合生殖,并将其用于上述列出的应用中。这些基因可通过转化引入有性作物。无融合生殖基因的结构和功能知识也可用于修饰内源性有性生殖基因,使其成为无融合生殖基因。优选的用途是将无融合生殖基因置于诱导型启动子下,以便当有性生殖产生新的基因型时可关闭无融合生殖,并且当需要无融合生殖来繁殖精英基因型时启动无融合生殖。

[0263] 然而,本发明的倍数孢子体形成多核苷酸或基因也可以全新的方式使用,而不是直接作为无融合生殖的元件。倍数孢子体形成基因可用于有性多倍体化,从二倍体植物中产生多倍体后代。多倍体植物通常具有杂种优势并产生比二倍体植物更高的产量(Bingham,E.T.,R.W.Groose,D.R.Woodfield&K.K.Kidwell,1994.Complementary gene interactions in alfalfa are greater in autopolyploids than diploids.Crop Sci 34:823-829.;Mendiburu,A.O.&S.J.Peloquin,1971.High yielding tetraploids from 4x-2x and 2x-2x matings.Amer Potato J 48:300-301)。Dip基因,即能够向本发明的植物提供倍数孢子体形成功能的基因(或嵌合基因、或载体或遗传构建体),避免了雌性减数分裂I,并因此产生了第一分裂恢复(FDR)卵细胞,其传递全部母系基因组包括所有杂合性和上位性基因的相互作用(Mok,D.W.S.and S.J.Peloquin.1972.Three mechanisms of 2n pollen formation in diploid potatoes.Am.Potato J.49:362-363.;Ramanna,M.S.,1979.A re-examination of the mechanisms of 2n gamete formation in potato and its implications for breeding.Euphytica 28:537-561)。FDR配子产生的后代优于第二

分裂恢复 (SDR) 配子产生的后代,后者只将亲本杂合性和上位性的一部分传递给后代。FDR和SDR类型的未减数配子在杂交后产生杂种后代,与通过化学处理(例如秋水仙碱)进行的体细胞多倍化相比,杂合性大大增加。因此FDR配子,像Dip基因诱导的那些,是有性多倍体化最优选的配子类型。FDR配子已被证明可用于改良多倍体作物,如马铃薯、苜蓿、蓝莓和一些草料饲料(Ramanna, M.S. and Jacobsen E. 2003. Relevance of sexual polyploidization for crop improvement-a review. *Euphytica* 133:3-8; Mariani, A. & S. Tavoletti, 1992. Gametes with Somatic Chromosome Number in the Evolution and Breeding of Polyploid Polysomic Species. Proc Workshop, Perugia, Tipolithographia Porziuncola-Assisi (PG) Italy, pp.1-103; Veilleux, R., 1985. Diploid and polyploid gametes in crop plants: Mechanisms of formation and utilization in plant breeding. *Plant Breed Rev* 3:252-288)。在这些应用中, Dip基因仅在雌性大孢子发生期间表达,并且雄性发生减数分裂,这是非常有益的。这可通过减少花粉粒使Dip基因渐渗入二倍体基因池,通过杂交创建新的有益基因组合。用于植物育种的Dip-基因的另一个非常有用的特性是它的显性,因此杂合子表达倍数孢子体形成表型。这显著简化了育种方案中Dip基因的使用。

[0264] 有性多倍体化的一个具体应用是用于产生三倍体,其可用于产生无籽果实。三倍体也可作为三体组学的来源,这对定位研究非常有用。

[0265] 然而在无融合生殖中,倍数孢子体形成和孤雌生殖二者在单一植物中结合,在一代中使用倍数孢子体形成和在下一代使用孤雌生殖将在二倍体和多倍体水平通过无融合生殖引起的倍数增加和孤雌生殖引起的倍数水平降低连接作物的有性基因库。这是非常实际的,因为多倍体种群可能更适合突变诱导,因为它们可耐受更多的突变。多倍体植物也可更强健。然而,二倍体群体更适合选择,且二倍体杂交更有利于遗传定位、BAC文库等的构建。多倍体中的孤雌生殖产生可与二倍体杂交的双单倍体。二倍体中的倍数孢子体形成产生未减数的FDR卵细胞,其可通过来自多倍体的花粉受精以产生多倍体后代。因此,在不同繁殖世代中倍数孢子体形成和孤雌生殖的交替,连接二倍体和多倍体基因池。

[0266] 以下非限制性实施例说明了本发明的不同实施方案。除非在实施例中另有说明,否则所有重组DNA技术均按照如下所述的标准方案进行,例如描述于 Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, and Sambrook and Russell (2001) *Molecular Cloning: A Laboratory Manual*, Third Edition, Cold Spring Harbor Laboratory Press, NY; and in Volumes 1 and 2 of Ausubel et al. (1994) *Current Protocols in Molecular Biology*, Current Protocols, USA. Standard materials and methods for plant molecular work are described in *Plant Molecular Biology Labfax* (1993) by R.D.D. Croy, jointly published by BIOS Scientific Publications Ltd (UK) and Blackwell Scientific Publications, UK。

实施例

[0267] 实施例1: DIP基因座的遗传定位

[0268] 1.1 无融合生殖重组群

[0269] 对于倍数孢子体形成 (Dip) 基因座的遗传定位,在二倍体有性蒲公英植物TJX3-20和三倍体单性繁殖植物A68之间进行杂交。选择TJX3-20作为雄性不育(无花粉生产)种子亲本,以防止产生高比例的自交后代,这是花粉蒙导效应的结果,在二倍体×三倍体蒲公英杂交中是常见的情况(Tas en Van Dijk 1999)。TJX3-20×A68杂交中的平均种子形成低,在1-3%之间。大量的杂交产生了共190个后代。仅产生有生存力的整倍体后代:97个二倍体、92个三倍体和1个四倍体(倍数水平用PARTEC流式细胞仪测定, Van Dijk et al. 2003)。没有任何二倍体是无融合生殖的,与分为无融合生殖/非无融合生殖的三倍体相反。

[0270] 1.2倍数孢子体形成表型分型

[0271] 为了从遗传学上对DIP基因座进行定位,针对倍数孢子体形成与非倍数孢子体形成(减数分裂)对三倍体后代植株进行了表型分型。产生三倍体种子而没有交叉授粉的三倍体后代植物是无融合生殖的,且因此也是倍数孢子体形成的。对于非无融合生殖植物的倍数孢子体形成表型分型,进行所谓的拟测交(Ozias Akins and Van Dijk 2007)。来自TJX3-20×A68杂交的三倍体后代与二倍体有性花粉供体杂交。收获种子并萌发,且通过流式细胞术测定后代的倍性水平(Partec倍性分析仪, van Dijk et al. 2003)。如果后代仅由四倍体植物组成,则通过减法得出三倍体母本植物是倍数孢子体形成的,因为二倍体花粉供体产生单倍体花粉粒。如果后代由具有三倍体或更低倍性水平的植物组成,则可得出结论,母本植物的卵细胞具有减少的染色体数目,并且母本植物本身是非倍数孢子体形成的。

[0272] 1.3DIP染色体区域的遗传图谱

[0273] 根据Wu等人(1992)所述的方法,可在同源多倍体植物中定位单剂显性标记物(单纯性例如001)。在来自TJX3-20xA68杂交的76个三倍体后代植物中定位与Dip基因座紧密关联的7个AFLP(Vos et al. 1995)标记物(来自Vijverberg et al 2004):(对于AFLP引物代码,参见表1)E40M60-505(505表示碱基对片段的大小;短代码:S4)、E38M48-215(S8)、E42M50-440(S7)、E35M52-235(S10)、E38M48-215(S9)、E45M53-090(A4)和E37M59-135(A5)。为了定位Dip基因座,使用如上所述的拟测交法对三倍体后代植物进行倍数孢子体形成表型分型。表2显示了在DIP染色体区域具有重组事件的四种三倍体后代植物(AS99、AS112、AS193和AS196)的基因型。

[0274] EcoRI

[0275]

EcoRI	选择性核苷酸
E35	ACT
E37	ACG
E38	ACT
E40	AGC
E42	AGT
E43	ATA
E45	ATG
E49	CAG
E60	CTC

MseI	
M40	AGC
M42	AGT
M48	CAC
M50	CAT
M52	CCC
M53	CCG
M59	CTA
M60	CTC

[0276] 表1. 使用的AFLP引物的选择性核苷酸。

[0277]

标记物/基因	植物				
	AS196	AS99	AS112	AS193	I124 缺失
S4	+	-	-	-	+
S8	+	-	-	-	+
S7	+	-	-	-	+
S10	+	+	-	-	+
S9	+	+	-	-	-
DD1	-	+	-	-	-
DD2	-	+	-	-	-
DD3	-	+	-	-	-
<i>Dip</i>	-	+	-	-	-
A4	-	+	+	-	-
A5	-	+	+	+	+

[0278] 表2. *Dip*区域的重组 (TJX320×A68) 和缺失 (A68_i124) 标记物图谱。(+) 号表示标记物存在; (-) 标志表示标记物不存在。

[0279] 实施例2. DIP基因座的缺失定位

[0280] 2.1 无融合生殖缺失群体

[0281] 由于TJX3-20×A68杂交中的种子数量太低而不能产生遗传精细定位所需的数千种子,因此需要一种替代方法。因此,使用缺失定位方法来精确定位该染色体区域。伽马辐射导致整个基因组中不同尺寸的随机缺失,而不管重组的热点或冷点。 γ 辐射缺失已成功地用于定位禾本属物种中的无融合生殖基因 (Catanach AS, Erasmuson SK, Podivinsky E, Jordan BR, Bicknell RA (2006) Deletion mapping of genetic regions associated with apomixis in Hieracium. Proc Natl Acad Sci USA 103 (49):18650-18655)。首先,通过⁶⁰Co源 (在Isotron BV, Ede, The Netherlands) 产生的100-800格雷范围的一系列测试剂量辐照干燥蒲公英种子测定对于克隆A68而言 γ 辐射的最佳剂量 (50% 幼苗存活)。对于最后的实验,用3种不同剂量辐射3×2000颗种子:三分之一使用250Gy,三分之一使用300Gy,且三分之一使用400Gy。在室温下使种子在培养皿中的湿滤纸上萌发。总共有3075株植物在温室中的盆中生长 (200Gy处理的350株,300Gy处理的1600株和400Gy处理的1125株)。植物在加热的温室中生长两个月 (日间21℃,光照16小时,夜间18℃)。接下来,将植物在2-10℃保

存两个月以诱导开花。在此春化期后,植物再次在上述条件下在加热的温室中生长。超过90%的植物开花并产生种子。植物以是否表现出无融合生殖缺失表型(LoA)分类。无融合生殖A68植物自发产生种子并形成大的白色种子穗(seed head),具有深棕色中心,其中种子(以植物学术语是瘦果:单种子果实)附着于花托(参见图1A)。

[0282] 在丧失无融合生殖表型的情况下,种子头部的中心较轻,并且由于种子不能正常发育,因此种子穗的直径通常会减小。在超过13000个种子穗中筛选无融合生殖缺失表型。最后,将102株植物鉴定为无融合生殖缺失表型。大多数这些植物产生无融合生殖缺失和无融合生殖种子头部二者,表明它们是嵌合体。这是因为这样的事实:经辐射的种子的芽分生组织是多细胞的(M1代)。

[0283] 2.2倍数孢子体形成缺失型表型分型

[0284] 经辐射的植物中的无融合生殖缺失可能是由于倍数孢子体形成缺失、孤雌生殖缺失或其他原因引起的。通过拟测交检测无融合生殖株中的倍数孢子体形成缺失(参见上文)。由于这些非Dip植物保留了孤雌生殖表型,所以倍数孢子体形成缺失植物也自发产生(因此没有任何异花传粉;参见图1B)数量较少的三倍体和低倍性三倍体后代。由于孤雌生殖是配子体表达的,它在非倍数孢子体形成植物的卵细胞中分离。

[0285] 2.3低分辨率缺失定位

[0286] 当三条同源染色体之一的一部分被缺失时,位于缺失区域的单剂量AFLP/SCAR标记物将丢失。为了确定102个无融合生殖缺失植物中有哪些丢失了部分的Dip基因座,研究了以下Dip关联的单剂量标记物的存在/不存在:S8、S7、S9、S10、A4和A5。

[0287] 共有23株无融合生殖缺失植物丢失了这些标记物中的两个或更多个。大多数这些植物也可被表型分型为倍数孢子体形成缺失,证实Dip-基因通过缺失而丢失。丢失的标记物数目是缺失尺寸的指示物(Catanach et al 2006)。除S9和A4之外,植物i124保留了所有这些标记物,表明该植物在Dip-基因座中具有最小的缺失。将具有最小缺失的五株植物(包括i124)通过组织培养制成非嵌合体。叶子被灭菌并且外植体在体外生长以再生整个植物。AFLP分析证实这些植物是纯合的,且仍然携带DIP缺失。

[0288] 实施例3.DIP基因座的DNA测序

[0289] 3.1在缺失群体中使用AFLP标记物精确定位DIP基因座

[0290] 为了在检测到的最小Dip缺失(i124)内找到新的AFLP标记物,开发了新的标记物筛选策略Bulked Deletion Analysis(BDA),其类似于Bulked分离分析(Michelmore et al. 1991)。比较三种DNA样品中是否存在AFLP片段:样品A:来自具有最小Dip缺失(i124)的植物的DNA,样品B:在Dip区域中具有较大缺失的三种植物的DNA池,和样品C:来自A68克隆的未经辐射的DNA。只有样品A和样品B都缺乏的AFLP位于最小的缺失中。考虑汇集的样本B阻止了选择在Dip基因座之外的缺失。来自BDA的候选AFLP在每个倍数孢子体形成缺失的缺失植物上进行了验证。筛选966个不同的AFLP引物组合得到位于植物i124的Dip缺失内的三个新的Dip缺失标记物(DD1:E43M40-68、DD2:E49M42-215和DD3:E60M42-76)。基于用966个AFLP引物组合筛选的AFLP标记物的数量和三个标记物缺失,估计在植物i124中Dip缺失的大小小于450kb。DD2标记物被成功克隆并测序(SEQ ID NO:12)。

[0291] 3.2通过BAC登陆和步移分离基因

[0292] 为了构建无融合生殖克隆A68的Dip基因座的完整物理BAC重叠群,筛选了BAC文

库。A68的BAC文库由Arizona Genome Institute构建,可通过AGI网站(<http://www.genome.arizona.edu/orders/>)作为T0_Ba获得。该BAC文库的平均插入大小为113kb,覆盖10个基因组当量(Taraxacum基因组大小:835Mb/1C)。它构建于pAGIBAC1载体的HindIII位点,且含有73728个克隆。将BAC插入文库双重点样于四张尼龙滤纸上。并汇集来自BAC文库中的克隆的DNA(192个超级池;384个BAC的板池;每个板也汇集在96个BAC DNA的4个池中)。通过对汇集的BAC DNA进行AFLP分析,在BAC插入文库中筛选含有S10、A4、DD1、DD2和DD3标记物的BAC。还使用DD2序列(SEQ ID NO:12)(Ross et al.1999),通过尼龙滤膜的重叠义寡核苷酸杂交(overgo hybridization)也筛选了BAC文库。对于每个标记物,选择一个BAC插入片段,其使用GS-FLX序列技术对其进行完全测序。通过使用种子BAC的末端来开发新的重叠义寡核苷酸探针,可延长BAC重叠群(BAC步行)。

[0293] 除BAC步移之外,使用基于序列的标签(Whole Genome Profiling-Van Oeveren et al.2011)制作A68BAC文库的物理图谱。BAC步移和WGP定位为DIP区域提供了一致的BAC重叠群。使用Finger Printed Contig(FPC)软件,基于共享的WGP标签构建最小的BAC覆瓦式路径(tiling path)。使用GS FLX技术对具有最小的覆瓦式路径的BAC进行测序。使用Newbler软件来组装单个454读段。在大多数情况下,发现两种BAC变体,之间的序列同一性在95-99%之间变化。这些变体被解释为不同的等位基因或单倍型。DD2标记物(SEQ ID NO:12)的存在区分Dip和dip BAC最小覆瓦式路径。

[0294] 3.3 BAC最小覆瓦式路径上缺失断裂位点的定位

[0295] 为了定位缺失断裂位点并确定覆盖了i124中最小的Dip缺失的最小覆瓦式路径,针对每个BAC序列设计一个基因的PCR引物。PCR扩增基因,并将DNA在ABI 3730XL上直接进行Sanger测序。这在A68的ABI跟踪文件中产生了一个复杂的原始测序数据,其中有许多双峰。然而,在i124中,模式经常被简化,并且是A68模式的子集,当等位基因之一(最不不同的)缺失时,预期这种模式。当基因的序列模式在A68和i124中都具有双峰时,推断该基因在i124中未缺失。在最小覆瓦式路径中间的BAC通常显示被缺失的基因,而在末端的BAC未显示缺失的迹象。因此得出的结论是,最小覆瓦式路径跨越了i124中的缺失。

[0296] 实施例4.Dip基因座内倍数孢子体形成基因的无偏差鉴定

[0297] 4.1 EMS无融合生殖敲除的产生

[0298] 我们推断,当蒲公英的无融合生殖受到遗传控制时,应该可能通过诱变剂如甲基磺酸乙酯(EMS)产生敲除突变。由于我们可预测Dip基因座内的基因,因此应该可以通过对倍数孢子体形成缺失突变体的Dip基因座中的基因重新测序来鉴定Dip基因。当我们发现若干倍数孢子体形成突变体时,它们应该在同一个基因(Dip基因)中有突变,而在Dip基因座中的基因中但与倍数孢子体形成表型不相关的突变不会被富集。从而,这将确定功能性的Dip基因。

[0299] 为了产生EMS无融合生殖敲除,生长自A68种子的1800株植物在室温下用0.35% EMS处理16小时。种子形成后,对植物进行倍数孢子体形成缺失表型的筛选(参见上文的描述)。共检测到6种假定的DIP缺失突变体(LoD1至LoD6),尽管其中两种LoD3和LoD5在拟测交中不产生种子。由于分离LoD植物用于孤雌生殖,产生一些有活性的M2种子并且从其生长M2植物。据我们所知,这是第一次通过EMS处理成功地产生了无融合生殖缺失突变体。试图在其他物种中通过EMS处理产生无融合生殖缺失没有成功(Asker and Jerling 1990,

Praekelt and Scott 2001)。

[0300] 4.2无融合生殖缺失EMS突变体中的倍数孢子体形成缺失物理间隔图中所预测的基因的高通量重新测序

[0301] 使用Augustus基因预测软件(Stanke M., R Steinkamp, S Waack and B Morgenstern(2004) "AUGUSTUS: a web server for gene finding in eukaryotes" Nucleic Acids Research, Vol. 32, W309-W312), 使用拟南芥基因模型预测在Dip和dip BAC最小覆瓦式路径(参见上文)中的基因。通过对预测的蛋白质序列与来自NCBI的非冗余数据库进行BLAST来进行基因注释, 以40%蛋白质同一性作为阈值。在Dip和dip BAC最小覆瓦式路径中共预测了129个蒲公英基因。

[0302] 从蒲公英A68、A68LoD1-6EMS突变体和LoD缺失系(A68i124)收集叶材料。使用CTAB程序提取基因组DNA(Rogstad 1992)。使用标准程序, 在FLUOstar Omega (BGM LABTHEC) 上用Quant-iT™ TmPicoGreen® dsDNA reagent (Invitrogen) 对DNA样品进行定量。将DNA样品稀释至20ng/μl的浓度, 并随后合并LoD样品以产生2个池(池A=LoD1+LoD2+LoD3; 池B=LoD4+LoD5+LoD6)。

[0303] 设计特异性引物用于129个预测基因的PCR扩增, 以主要靶向其编码序列。总共设计了295对引物。选择蒲公英无融合生殖A68克隆、A68_i124缺失系(LoD表型)和A68LoD EMS突变体池A和B作为扩增子筛选的目标, 目的是将EMS突变表型与EMS突变相关联, 并从而鉴定DIP基因。

[0304] 从每个选定的目标通过PCR反应产生295扩增子。进行50μl PCR反应, 每个样品含80ng DNA、50ng正向引物、50ng反向引物、0.2mM dNTP、1U Herculase H II融合DNA聚合酶(Stratagene)和1X Herculase H II反应缓冲液。使用以下热谱进行PCR: 95°C 2分钟; 然后95°C 30秒, 30秒55°C和30秒72°C, 35个循环; 随后冷却至4°C。来自样品的等量PCR产物用于GS FLX片段文库样品。

[0305] 使用基因组测序仪(GS) FLX+PLATFORM (Roche Applied Science) 进行扩增子筛选, 其允许单个DNA分子的大规模并行的皮升级扩增和焦磷酸测序。使用标准Roche方案构建扩增子样品文库。条形码(多重标识符, MID) 在实验准备期间添加。合并MID标记的样品用于同时扩增和测序(多重化)。使用具有两个区域的一个完整的皮升级微量滴定板(PTP) (70x75mm) 对扩增子文库(A68、A68_i124、A68_EMS池A和B) 进行测序。根据制造商的说明书(Roche Applied Science) 进行测序。

[0306] 突变筛查的生物信息学分析由5部分组成:

[0307] (1) GS FLX+数据处理, 使用Roche GS FLX+软件。对碱基描述的(base-called)的读取进行修剪并就质量进行过滤, 并转换成FASTA格式。

[0308] (2) 样品处理。序列阅读的起始是基于特定的条形码识别的。对条形码序列进行修剪并将每个样品的序列读数分别保存到数据库中。

[0309] (3) 扩增子处理。扩增子的起始是基于靶特异性的引物序列识别的。使用CAP3将每个扩增子的序列读数聚类(95%同源性, 40个核苷酸重叠)。

[0310] (4) 多态性检测。在每个聚类扩增子中鉴定所有潜在的SNP和INDELS。

[0311] (5) 检测EMS SNP。EMS处理诱导的SNP的鉴定。预期这种SNP仅在EMS突变植物中(EMS池A或B)。考虑六个独立的EMS突变体汇集在一起(池A中3个, 池B中3个), 并且在池A或B

中检测到EMS诱导的SNP,但不能在两者中均检测到。如果匹配以下参数,SNP被认为是真正的EMS-SNP:(a)不存在于A68和A68_i124中;(b)在池A或池B中被检测到。

[0312] 鉴定了总共6种推定的EMS突变(C→T或G→A),其中在单个基因中发现了4种,其与拟南芥蛋白质的液泡蛋白质分选(VPS)13样蛋白质具有非常高的蛋白质BLAST同源性(gi|10129653|emb|CAC08248.1|)(表3)。

[0313]

氨基酸起始	氨基酸终止	Blast分数	E值
643	785	204.91	1.9e-054
1020	1393	237.65	2.6e-064
1645	2097	256.91	4.1e-070
2142	2618	303.91	3.0e-084
2608	3384	728.78	3.7e-212
3390	3737	327.79	1.9e-091
3621	3931	273.09	5.6e-075

[0314] 表3.SEQ ID NO:3和拟南芥VPS13样蛋白质(gi|10129653|emb|CAC08248.1|)之间的蛋白质同源性。Tera-BLASTP搜索蛋白质查询(DeCypher,TimeLogic™标准设置)。

[0315] 这是一个巨大的基因,代表了测序的295个外显子的34个,相当于总重新测序核苷酸的11%。通过对倍数孢子体形成缺失表型进行选择,Dip基因中突变的富集是预期的。所有四种ToVps13EMS突变都在Dip单倍型中,没有一种在dip单倍型中。我们计算这种基因序列上的突变分布的概率是由于以下几率引起的。预测的ToVps13的大小是总重新测序区域的11%。由于有三种单倍型,单个ToVps13单倍型的大小为总重新测序区域的3.7%。第一个EMS突变体位于Dip单倍型的概率是0.33。第二、第三和第四个EMS突变位于相同单倍型的同一基因中的概率为 $0.037 \times 0.037 \times 0.037 = 5.1 \times 10E-5$ 。第一个EMS突变位于正确的单倍型且第二个、第三个和第四个位于同一单倍型中的同一个DNA区域的组合概率为 $0.33 \times 5.1 \times 10E-5 = 1.67E-5$ 。由于这也可能发生在其他DNA区域,整个重新测序区域的概率是 $100/11 \times 1.67 \times E-5 = 1.54 \times 10E-4$ 。因此,这种分配随机发生的概率是 $1.54/10,000$ 。因此,Vps13序列很可能包含于倍数孢子体形成。

[0316] 在两个LoD植物中,在重新测序的区域发现第二个EMS突变,一个在寡肽转运蛋白中,且另一个在推定的转运蛋白基因中。在这两种情况下,突变都不在Dip单倍型中,而是在dip单倍型中。因此我们得出这样的结论:这两个EMS突变与倍数孢子体形成表型无关。在推定的LoD3和LoD5植物中,在重新测序的区域中未检测到EMS突变。这些植物不会在拟测交中产生后代(见上文),并且可能是雌性不育突变,而不是无融合生殖缺失突变。

[0317] 实施例5.在多种不相关的有性和无融合生殖蒲公英的DIP基因座的关联定位

[0318] 为了提供SEQ ID NO:1涉及倍数孢子体形成表型的进一步证据,在一组无融合生殖(=倍数孢子体形成)植物和一组有性(=二倍体)植物中研究了序列SEQ ID NO:4和倍数孢子体形成的联系。两个小组由13个不相关的植物组成,就地理起源和分类学组(蒲公英属中的不同部分(section)和不同物种)而言尽可能多样化。根据Tas和Van Dijk(1999, Heredity 83:707-714)所述的方法,通过流式细胞术测定倍性水平。通过与传粉媒介隔离的种子形成确定育种系统:无融合生殖隔离产生完整的种子形成,有性组隔离不产生种子。

部分SEQ ID NO:4重新测序,1-300nt或7-586nt,第一次通过Illumia配对末端测序,第二次通过在基因组测序仪(GS)FLX+PLATFORM(Roche Applied Science)上测序。使用具有标准设置的Decypher(TimeLogic),用针对SEQ ID NO:4的核苷酸BLAST分析序列。在表4中给出了每株植物中最高的核苷酸序列同一性和最小的E值。从该表中清楚地看出,所有的无融合生殖体携带SEQ ID NO:4的测序区域,而没有任何有性体携带该DNA片段。因此,在这个序列和倍数孢子体形成之间存在最大的关联不平衡。如果核苷酸区域在功能上不参与倍数孢子体形成,则重组和诱变将随时间侵蚀核苷酸区域和倍数孢子体形成之间的关联不平衡。因此,无融合生殖与SEQ ID NO:4之间的完美关联证实了该序列对于倍数孢子体形成是必不可少的。

	部分	物种	起源	倍性
	<u>A. 有性生物体(减数分裂)</u>			
	1. Biennia	<i>T. mutans</i>	中国, 山西	2x
	2. Ceratoidea	<i>T. koksaghyz 1</i>	哈萨克斯坦	2x
	3. Ceratoidea	<i>T. koksaghyz 2</i>	哈萨克斯坦	2x
	4. Ceratoidea	<i>T. koksaghyz 3</i>	哈萨克斯坦	2x
[0319]	5. Ceratoidea	<i>T. koksaghyz 4</i>	哈萨克斯坦	2x
	6. Ceratoidea	<i>T. koksaghyz 5</i>	哈萨克斯坦	2x
	7. Ceratoidea	<i>T. koksaghyz 6</i>	哈萨克斯坦	2x
	8. Mongolica	<i>T. hallaisanense</i>	韩国	2x
	9. Obliqua	<i>T. pyrenaicum</i>	法国	2x
	10. Piesis/Primigenia	<i>T. cylleneum</i>	希腊	2x
	11. Piesis	<i>T. bessarabicum</i>	乌克兰	2x
	12. Piesis	<i>T. stenocephalum</i>	俄罗斯, 高加索	4x
[0320]	13. Ruderalia	<i>T. officinale 3 (FCH72)</i>	瑞士	2x
[0321]	<u>B. 无融合生殖体(倍数孢子体形成)</u>			
	1. Borealia	indet.	中国, 山西	indet.
	2. Ceratoidea	<i>T. brevicorniculatum</i>	哈萨克斯坦	3x
	3. Erythrocarpa	<i>T. gratum</i>	高加索	indet.
	4. Erythrosperma	<i>T. lacistophylloides</i>	IBOT	indet.
	5. Erythrosperma	<i>T. brachyglossum</i>	智利	3x
	6. Palustria	<i>T. validum</i>	IBOT	indet.
[0322]	7. Ruderalia	<i>T. officinale 1(Ron)</i>	法国	4x
	8. Scariosa	<i>T. minimum</i>	马耳他	5x
	9. Stenoloba	indet.	西伯利亚, 雅库特	indet.
	10. Mongolica	<i>T. aurantiacum</i>	中国	4x
	11. Nevosa	<i>T. richardsiamum</i>	英国, 威尔士	4x
	12. Ruderalia	<i>T. officinale 2 (A68)</i>	荷兰	3x
	13. Scariosa	<i>T. Hybirmum</i>	俄罗斯, 克里米亚	3x

部分	<i>nt</i> 同一性	BLAST E_值	区域 SEQ ID NO:4	
<u>A. 有性生物体(减数分裂)</u>				
	1. Biennia	93	6.4e-128	1-300
	2. Ceratoidea	95	8.9e-081	1-300
	3. Ceratoidea	96	1.5e-085	1-300
	4. Ceratoidea	95	8.9e-081	1-300
	5. Ceratoidea	93	5.3e-076	1-300
[0323]	6. Ceratoidea	95	8.9e-081	1-300
	7. Ceratoidea	96	1.5e-085	1-300
	8. Mongolica	96	1.3e-150	1-300
	9. Obliqua	99	7.1e-100	1-300
	10. Piesis/Primigenia	95	3.7e-083	1-300
	11. Piesis	97	2.5e-090	1-300
	12. Piesis	91	5.5e-116	1-300
	13. Ruderalia	96	1.3e-150	1-300
[0324]	<u>B. 无融合生殖体(倍数孢子体形成)</u>			
	1. Borealia	100	4.4e-172	1-300
	2. Ceratoidea	100	4.4e-172	1-300
	3. Erythrocarpa	100	4.4e-172	1-300
	4. Erythrosperma	100	4.4e-172	1-300
	5. Erythrosperma	100	4.4e-172	1-300
[0325]	6. Palustria	100	4.4e-172	1-300
	7. Ruderalia	100	4.4e-172	1-300
	8. Scariosa	100	4.4e-172	1-300
	9. Stenoloba	100	4.4e-172	1-300
	10. Mongolica	100	0.000000	7-586
	11. Nevosa	100	0.000000	7-586
	12. Ruderalia	100	0.000000	7-586
[0326]	13. Scariosa	100	0.000000	7-586

[0327] 表4. 无融合生殖与SEQ ID NO:4之间的关联作图。使用具有标准设置的Decypher (TimeLogic), 用针对SEQ ID NO:4的核苷酸BLAST分析序列。给出了每株植物的最高核苷酸同一性和最小E值。Indet. 表示无法确定。IBOT指由捷克共和国植物学研究所提供, 地理来源不详。

[0328] 实施例6. DIP基因在无融合生殖植物的大孢子母细胞中的表达以及近等基因的倍数孢子体形成缺失突变体

[0329] 为了研究DIP候选基因的表达, 从分离的大孢子母细胞(MMC)和无融合生殖(A68)的雌配子体(FG)及其同基因缺失系(i124)进行RNAseq。初步研究清楚地表明, 蒲公英的大孢子发生在非常年轻的花序(直径约0.5厘米)的芽中, 其发生于茎伸长之前、当芽仍然在植物的莲座丛中时。对于后期(雌性配子体; FG), 茎长1cm时收集芽。

[0330] 将新鲜子房切开并在果胶酶、果胶裂解酶、半纤维素酶和纤维素酶的甘露醇混合物中浸渍。通过使用针手动显微解剖将胚珠与周围组织分离。使用CellTram® 0i1装置(Eppendorf)在20个胚珠批次中收集分离的胚珠并立即在-80℃冰箱中冷冻直至进一步处

理。用 Arcturus®Picopure® RNA 分离试剂盒从 20 个胚珠的池中提取 RNA。使用 AmbionMessageAmp™ II aRNA 扩增试剂盒通过体外逆转录来线性扩增 RNA。来自相同基因型和组织的 20 个胚珠的不同池被认为是生物学重复。

[0331] 总共 10 个样品在 6 个 Illumina HiSeq 泳道 (A68MMC 的 3 个生物重复、FG 的 3 个生物重复和 MMC i124 的 4 个生物重复) 中测序。根据样品, 使用 FLASH 软件合并重叠读取对 (<http://ccb.jhu.edu/software/FLASH/>)。合并 (未经过滤) 的读取使用 Trinity 软件 (<http://trinityrnaseq.sourceforge.net/>) 进行组装。对于每个样品, 根据 Trinity 的“使用 RSEM 的丰度估计”方案 (http://trinityrnaseq.sourceforge.net/analysis/abundance_estimation.html) 估计转录本丰度。然后根据“鉴定差异表达的三位一体转录物”方案 (http://trinityrnaseq.sourceforge.net/analysis/diff_expression_analysis.html) 鉴定差异表达的同种型。

[0332] 在重新装配的表达基因中, 检测到超过 40 种减数分裂基因 (例如 Dmc1、Spo11、Rad50), 表明研究了正确的发育胚珠阶段的 MMC 和 FG。SEQ ID NO:4 是重新组装的, 并且显示在 MMC 和 FG 阶段以中等表达水平在无融合生殖植物 A68 中表达。在表 5 中, 表达被定量为 FPKM 值 (每特征千碱基每百万个定位的读段的片段)。在缺失突变体 i124 中, SEQ ID NO:4 不表达, 但在其倍数孢子体形成同源体 A68 中表达。因此, 表达数据证实 Vps13 基因缺失中并且在 MMC 和 FG 发育阶段表达。

[0333] 到目前为止进行的表达与关联作图分析表明, 目前注释为 Vps13 基因的 3 主要末端的如 SEQ ID NO:4 所示的核酸分子独立转录, 作为新基因或作为 Vps13 基因的差异剪接变体, 与 *Saccharomyces pombe* 的孢子形成基因 Spo2 相似。Spo2 基因编码由 133 个氨基酸残基组成的 15-kDa 蛋白质, 其被错误地注释为裂殖酵母 Vps13 基因的最后一个外显子。实际上, Spo2 基因直接位于 Vps13 基因的下游并且独立转录 (Nakase et al 2008, *Molecular Biology of the Cell*. Vol.19, 2476-2487)。

[0334] 值得注意的是, SEQ ID NO:5 的 mRNA 序列不含 ATG 起始密码子, 并且可能翻译的开放阅读框短。然而, 在芽殖酵母 (酿酒酵母 (*Saccharomyces cerevisiae*)) 中使用核糖体谱分析时, Brar 实验室 (加利福尼亚大学伯克利分校 <http://www.unal-and-brar-labs.org/brar-sorfs>) 已经鉴定了在减数分裂期间的数千个新短肽的非经典翻译。这些减数分裂特异性表达的短开放阅读框 (sORFs) 没有 ATG 起始密码子, 且它们翻译的肽短于 80 个氨基酸, 并因此不能用标准基因软件预测。sORF 位于以前不知道含有表达序列的区域。sORF 也可以是具有已知功能的蛋白质的短的替代亚型。经典方法已证实减数分裂期间这些短肽的存在。然而, 数千个这些短的减数分裂特异性肽的功能仍然是一个谜。

植物	组织	FPKM	S.E.
A68	MMC1	6.80	2.51
	MMC2	5.96	
	MMC3	8.54	
	平均	7.10	
i124	MMC1	0.00	0.00
	MMC2	0.00	
	MMC3	0.00	
	MMC4	0.00	
	平均	0.00	
A68	FG1	6.45	0.69
	FG2	5.56	
	FG3	6.45	
	平均	6.15	

[0336] 表5.SEQ ID NO:4在大孢子母细胞中的表达和无融合生殖体A68和Dip缺失品系i124的雌配子体中的表达。绝对表达是按照每特征千碱基每百万个读段的片段 (FPKM) 作图测量的。计算平均值和标准误差。指出了等位基因特异性表达的百分比。

[0337] 实施例7. 在拟南芥中过表达ToDIP和Todip

[0338] 将前有人工ATG起始密码子的ToDIP序列片段 (SEQ ID NO:11) 和前有人工ATG起始密码子的Todip序列片段 (SEQ ID NO:9) 克隆到具有35S启动子的载体中。用这些组成型过表达载体进行三个独立的拟南芥花dip转化实验。在每个实验中, 获得每个等位基因15到30个To植株。

[0339] 35S::Todip过表达转化体与野生型植物无法区分并且完全可育。相反, 在所有三个实验中的35S::ToDIP过表达转化体中, 一些植物是部分不育的 (第一个实验中20%、第二和第三个实验中10%的转化体)。

[0340] 使用Yadegari, R., et al. (1994) (Cell differentiation and morphogenesis are uncoupled in Arabidopsis raspberry embryos. Plant Cell, 6, 1713-1729) 的方法清除胚珠, 通过Nomarski显微镜研究大孢子母细胞 (MMC) 和雌配子体 (FG) 发育。所有研究的35S::dip转化体中的MMC和FG发育看起来都是正常的, 如在野生型拟南芥植物中一样。然而, 35S::ToDIP植物经常表现出大孢子母细胞异常、大孢子附近的额外小核以及FG发育受阻, 例如在FG1阶段停滞、没有空泡和破裂的胚囊。在影响雌性和雄性减数分裂的拟南芥dyad突变中 (Ravi M et al. (2008) Gamete formation without meiosis in Arabidopsis. Nature 451:1121-1124), 观察到FG发育的类似紊乱。因此, 观察到的35S::ToDIP异常的MMC和FG表型可能表明存在被干扰的雌性减数分裂。

[0341] 这些ToDIP表型是显性的, 因为它们在半合子T₀中被观察到。这与蒲公英DIP等位基因的显性一致。在第一个实验中, 在一些植物中, 花粉发育也受到影响 (额外的核), 但在第二和第三实验中, 花粉发育看起来正常。至少在第二和第三个实验中, DIP构建体的表型效应是雌性减数分裂特异性的, 这与蒲公英的DIP功能一致。

[0342] 总之, 发现蒲公英DIP等位基因在异源植物物种中产生雌性特异性减数分裂显性。蒲公英dip等位基因没有发现这种效应。拟南芥过表达表型提供了强有力的支持性证据, 表

明DIP序列在蒲公英中引起了倍数孢子体形成表型。

[0343] 实施例8. 蒲公英中的DIP基因功能

[0344] 为了进一步确认SEQ ID NO:4的倍数孢子体形成功能,将其中DIP等位基因缺失的蒲公英i124植物用含有SEQ ID NO:4的质粒转化,其与适当载体中的不同启动子和调控元件融合。使用以下启动子序列:

[0345] 1. SEQ ID NO:4的天然蒲公英启动子(SEQ ID NO:1的约1500bp,在SEQ ID NO:4的上游)

[0346] 2. 拟南芥DMC1 (At3g22880) 的蒲公英直系同源物的启动子(Klimyuk V.I. and Jones J.D.1997. *AtDMC1, the Arabidopsis homologue of the yeast DMC1 gene: characterization, transposon-induced allelic variation and meiosis-associated expression.* *Plant J.*:11:1-14)。该基因具有减数分裂特异性启动子。

[0347] 3. 35S启动子。该启动子导致SEQ ID NO:4的过表达。

[0348] Wahler等人2009 (*Plant Phys.*151, pp.334-346) 已经公布了转化蒲公英植物的方案。由于i124携带无融合生殖的所有其他元素,因此倍数孢子体形成的互补将恢复无融合生殖,这可通过该三倍体植物中的高种子形成以及T1子代中的遗传标记物而容易地确定。后代植物含有完整的母体基因组,没有母体标记物的分离。

[0349] 实施例9. 通过转化在有性作物中引入倍数孢子体形成

[0350] 根据Dreni, L等人2011 (*Plant Cell* 23:2850-2863) 和Dias, B.B.A. 等人2006 (*Plant Pathology* 55:187-193) 的方案,分别使用水稻和莴苣的二倍性有性植物进行转化。使用如实施例8中公开的具有启动子和SEQ ID NO:4的相同构建体。将T₀倍数孢子体形成植物与二倍体花粉供体杂交后,产生三倍体后代。三倍体可通过根尖染色体计数或通过流式细胞术确定。两者都是标准方法(Tas and Van Dijk 1999, *Heredity* 83:707-714)。进一步证明倍数孢子体形成可在后代植物分析中通过遗传标记物的分析找到。除了父本标记物之外,后代将携带完整的母本基因型。

[0351] 实施例10. 通过基因组编辑将倍数孢子体形成引入有性作物中

[0352] 靶向性的基因组编辑技术,如CRISPR-CAS9、TALENs和ZFN (锌指核酸酶) 在本领域中通常用于在现有基因中产生突变。不仅通过产生敲除等位基因,而且通过引入由所谓的“修复DNA”编码的突变(例如Doudna J.A. and Gersbach C.A.2015 *Genome editing: the end of the beginning* *Genome Biology* (2015) 201516:292,以及其中引用的参考文献)。

[0353] 这种DNA片段通常编码一个(靶)基因序列的片段,其中引入了导致基因功能改变的变化。典型地,这样的序列通过同源重组代替基因组编辑事件中靶向的基因序列,从而以靶向的方式在宿主细胞(例如一个植物细胞)的基因组中引入选择性突变。

[0354] 该实施例包括引入给定植物物种中对dip同系物的改变,其导致功能改变为DIP,即通过显性倍数孢子体形成(DIP)等位基因改变天然发生的隐性非倍数孢子体形成等位基因的功能。

[0355] Dip同系物很容易在许多植物物种中鉴定出来。CRISPR CAS介导的使用基于蒲公英的“修复”质粒设计的基因组编辑可通过符合蒲公英DIP和dip等位基因之间的差异简单修饰SNP和插入缺失突变(indel)来将天然dip同源物转化为其DIP姊妹体(sibling)。

序列表

<110> 主基因有限公司
 <120> 倍数孢子体形成基因
 <130> P29610PC00
 <150> NL2015398
 <151> 2015-09-04
 <160> 13
 <170> PatentIn version 3.5
 <210> 1
 <211> 23361
 <212> DNA
 <213> *Taraxacum officinale*
 <400> 1

```

ttttcaccta gaagagcaca accetettca agtgaagcac aactcgaag tgtatttccg 60
tatttcagtc cagctggaaa cagaagacca ttaatgatat tttagcaaga aagctttcat 120
tcatttggtc atcatagcag catcetaact cctacaacat tgtatthttga aaaatctacc 180
aattaaatag caatatcadc caaatacaaaa gcaagtttca ctgctatata ttgctaaacc 240
tctctttaat ccctagatat aataagggca aaaaagtgga acatatgtac ctttactctg 300
tcatttgatg ataaatgttg tttttaagtt agtatcaaac tgaaattgtg ctaatcacia 360
tggtgttgaa ttgttgcaa attctattta ttaataaat cataggtaaa acaatgacat 420
ataaactatt ttactctacc atttgacaac cactettgat gtatcaaac gattatatga 480
caccacacc catataatca gataatcact tccaaaaage atatccaacc acccctctat 540
atthtcaacta actatcccaa ttgctgtct ataaattcag aattaataaa tctactcagc 600
tgattgattt tgaataaata ataatgacac attgcatact gatagataaa gttctgcagc 660
acccaaatca aatgtcctaa aaccttaagg tgaagaacac agcccaaagg aattgaagat 720
cgacataac tctttaacag ttccttaata tcaatgacia aactgaatcg aatttaacgg 780
ggggaaaaat aaatacacca atttcataac ttaaacacc tggttaatcc tgcgactac 840
aaatcaaac agctgaatat ttaaccaaga gaaaaataa aaaccctagt tgatcgcatt 900
gacatctgag ctacacacat ccaattacat aacgaaaaat acgaatcaga agtttacgaa 960
acgaaacaag atctagtcaa taaagaaaac tgaactaag acagatcaga aaagtggttt 1020
ctcgaacgta cctgatcgaa gtgaagatgt cagatctgat aaactggtat cttgtagtaa 1080
aagatgaatc gagagataaa aatgaagaac gaatgcagaa ttgtaataa aaatacacia 1140
aactcaagat ttatttctgt tttatgttga aaatttgaat gaaaatgatt ctaacagaga 1200
aaggtttgg tctttccttt ctgtttctct gtcaacaaca atthaggaa ggcgtagaga 1260
gatgagagca gacagcgcac ctttgcgcgg agaatttctg ggatatact atthatttcc 1320
ctthtacctt tctthtcatt atgtthtttt thttthcttc agaaaattca aactthcttt 1380
ttcaatattc tgagtacget aatttctagc tggcatgtag caagtcttaa thtctagctg 1440
gcatgtagca agtctthgtt ggggtgtata gcatttccct cthtttaata gtaagtgaaa 1500

```

atgacttaat agggtaaaaa cgtttcgaac ttgtccatat ttggtcatta tcctatTTTT 1560
tgtacacaat acaccattaa gatttatatg accggttaag tttcgtccat tgccgatttt 1620
cataggaat catcgtgaaa tcagtaagca acctcatgaa tcaccactc ctgccgccac 1680
aaccacagcc gaagcttcat ttgagcccct ggatgaaaat ctcttctcgg atctaacccc 1740
actctgttat ttactccaat ccacctcccc caaatcacc ccaacattaa tctttcctcc 1800
gttctaccac caccaatcct ctcaagattc atcctttctc atcgccgtga cgttttccag 1860
caaaacccat gtccaccacc gccgccgtca aaccagtccg gaaattcccg ccatcgtgcc 1920
cggatgtaac cccgccgtca aaccagtccg gaaatcaaca acacaacccc ctgttttgca 1980
tcaccagagc ttgaatccgg gtcgtggagt tcgatttaaa ctatcttcaa ctcaaaatct 2040
ggtaacatta gcagctagtt ccaccaatca taccctgat ttcgattcta ggtttccaaa 2100
tcacgattcc tcatacatga acaactggtc taatcaagaa gaagacaacg acgatgggtg 2160
tcttgatgac tctacaatta atgaaacctc gattcatccg ggtaccaaaa tcacaagagt 2220
gtgaaaggag gaattcggat aaaatcctca attccatccc acctcgggcc gagaaagttg 2280
agcaaggctg ctcatgaaca tgaagtcgat gaggatgggtg aatgtggat taaagtgcct 2340
agaaaacaaa atttgTTTT cggaacctc caaaacggca atagTTTTaa tctaattgg 2400
gatcaaggga agaaaagaag aatgaaggg atataggagg tggttcaatc gattaagcta 2460
ttaggagatg ggttcatgaa gatgctgact gatttcagga ggttgctgac tgatttcacg 2520
atgattacct attaaaaccg gtaaatggac gaaacttaac cggtcataata aaccttaatg 2580
gtgtattgtg tacaaaaaaa taggataatg accaaatggt gacaagtctg aaacgtTTTT 2640
acctattaa gtcattttcc ctaatagtta actaattatt cttttcgct ctaaattatt 2700
ttttgttata atttcttcta aatagactgg tttatcttcc ttaattactg tatgaaacat 2760
ttcacaact tattactttt taaaaatata tcatatttct caagataaaa agtaagataa 2820
aagaattaat atatctatgt gttacaaatc aaaaaatgat aattgtttgg atgccgtagt 2880
atTTTgccac gactaaaagt taacatatta tttctagggc atTTTtaata ataaagaac 2940
tgacttatat atatgtaaca tattatttgt tcatatttag tcattatata tcttttgtca 3000
tattattaaa tttgttacia ttttttatat ttagccacta caatttgcaa ttgcaggtca 3060
tagtgaccaa atgtaaacia ttgcaacaag ttcgatggta taatgtgaca aaaaaatata 3120
atgaataaat atgttttcaa acattctttt ttatatttag ccactacaat ttgtcatctg 3180
agttggaata aatatgtttt caaacattct ttttttaaat tcaaaatgaa tgttctcgtg 3240
taaatcacag caatcactta ttatgatttt acattattgt gaactaaaat aaaacttctc 3300
gacgaaatac atatattgta tataatatct ttacttccgg tagtctgtgtt tttaaaatat 3360
attcaatctt attatatgta caattaaata ggTTaattaa taaaacataa ttattatatt 3420
aaatatattt attgatagat acaaaaaata gaagttagta tagctaaatg gaaaaaac 3480
ataaaaaaat gattaatagg ttattgggtg catatttatt ctaaacgcta aacatatatg 3540
gttctaaaaa aatgtctgat gaatgttcaa aatatggtac aaataatgt ctcaagaact 3600
gtaacatcct taatttttag aagttatgtt ttaagttaaa attaactaat taggacctaa 3660
tttgTTaagt ttaggataa taatgaaaga aataaagaga ctcaaggTTa ttttgcattt 3720
ttgacaagta gagggaccac taatgaaaaa tgtggaatTT aaaagcctgg aactctcttc 3780
cagcttctaa gatcgcgaga gaaagagagc agtgaagggt aggggcgatt ttgagagcaa 3840

accaagaaca agagctaaat tctgatcatt gaagagatta ggagtgatac taaagaggat 3900
 tcaagcacia aaaggaaga aatcatcttc tatttacaag aaattcgttt ttgatgttga 3960
 gggtagaaac ccaaaatcga aatcagatga aattgaaggt ttaagggtc gattctaacc 4020
 tcttttggtc attagaagt atttctagct tttgctcacc tttgaattga ttatttgaag 4080
 gttttgggga agatcaccat gaacttagaa tttgctattg gggtttgatt atatgctttg 4140
 attccatccc ttagacatgt cataaagcct gtaaagagt gataagaagt attgctaaaa 4200
 ggtaggact ttagaccaaa agtttgaag ttgcaagaaa ttgaaaaga gaagctgcct 4260
 agactgcgga cacgggtcgt gtcccacagg ctggaatcat tttgagttc gatttcttta 4320
 tttttttatc caattccaat aattccaagt gcaataagct tagatatgat gataaatgag 4380
 ttgcacaaaa gttgaatata tttgaaata cccaagggtta ttttggctat tttaatgacc 4440
 tttagactat tgttttgtag aaaacaaacg aaggagcaa taggaaggag gtctagctgg 4500
 gggttgcaag tatctgaaaa cctctacaac ctaaggtgag tacgtgtgat tcatttcccc 4560
 tctttgaggg tatgtttatg tgataggaat ctatgtatgc aaagtaggtt gttattgtaa 4620
 agggaaatgt tagtctttct agttgcaggt actatgttgc tatgatgatg tatagaatga 4680
 tatgaagata tgaatgctag aaagaatgtg tacgaaatga atgagaatgt tgcctggact 4740
 atgagagcct cgcaaggggg cctagaggat tctattggga cggtaacctc cttcgcgaaa 4800
 tagaactcct aatgtgatg agagccttgg tggactgatt tgttcttttc acctggcct 4860
 agacttgggg tggttcctaa taggacatag accctaagt taaaggaatg atcattgaag 4920
 gggatcccga tatattgtgg ataatggcac aagaattaaa tagaattgat gatgttataa 4980
 attgaattta aatgaatgtt ttagattata tccgcgtatc tcaccagacc ttgtctgata 5040
 tgttgttttg tgccatgat tccaggtagt agttctcatg cataggagat ttggatattt 5100
 tgaagtata ttagcaagt gaatgaagac ccggttgta tcattctcc aattcagttt 5160
 atactatggg tgttttgat acaagttggt aaataccca aacttaatta catcttttga 5220
 aagtaaatgt ttatgatatg taaatagttt taaacttgat atgttttctt ggttaaacca 5280
 tagtcataaa aaaatatttt aaaaagttat gaaaggttgg ggtgttcaa gaacagtcctg 5340
 acagcggccc tctccgacg gtgttcccg gcgagttcct ttctactccg gctagcttcc 5400
 ccattattgg attaacctat tactgcacct ctccctcctc ctcgaacatg tcacattggt 5460
 ctccggttctc tctgcatga ccataacaac aaacatatgt atagttgtgt gtaggcattg 5520
 agtttaaacg gaggagaaaa tatgcatgg ccagctcacc cgctgctgcc atcccagcca 5580
 gcattagccc agccgtcgt cacatccacc gaaaagctga tttccacatc gtcgctcccg 5640
 gccgtcacct acacgtaggt tcgtaagttt ccaatatctt agttgcgttt tcggatggct 5700
 gataaaaatg catccgttga caaggaagat attgtgattt gtgagtaaac tattttgctg 5760
 gaggaaggtg agttagatca catggcgagc atggtaagat gtgaggtggg taaatttcca 5820
 ttcaattact tggggttcc cataggggca aacatgaaat tatctaaaca ttggaacatt 5880
 atagttgata agtttgagaa aagattatca aactggaaag ctgagaattt gtcttttagg 5940
 gggcggtgta ctttgataaa atcgggtgat gggagtcttc cgttgtttta tttctcgatg 6000
 tttagggtc cgaagaaagt ggtcgataag ctagaaggga taagaagaag gttcttatgg 6060
 ggcgtaaga agtcgaaaa aaagattcac tgggttctgt gggagaaggt gataaaatca 6120
 aaagataaag aaggtttagg ggtgaatgga ttgagcagca tgaatatggc cttgctagtg 6180

aaatggtttt ggcggctcaa aactgaaagg gacagcctgt gggttagatg tgtcacggct 6240
 tgtcataata tcaaacttat tgatgggaaa cgggtggcta aagcttcctt gaaaggagta 6300
 tgggtgaaca tcatgagctg tgttgaagag ttaaaaacga aaggaatttc tgtggagtca 6360
 aagttagtaa ggcaactggg aaacggcaaa cacacgcatt tttggaagga tagatggtta 6420
 cacaacaaag ttttaaaaga tgaccttccg gagttgtaca aaatagaagg ggacaaaaat 6480
 tgtatggtaa accaaagact ggtttgggac aacaacgaga aaatattcaa gcaagcttgg 6540
 gactggaaaa gaccgatcag gagaggaaga gaaaccaaag aactcgaaac tttataaatt 6600
 ttgacaaatg ggatacaatt aaaggaaata gaagataatt ggagatggaa ggagggatcg 6660
 gacgggaaat tttcgggtggg aaaattgagg aaactctttg cttatcagga gcaggccgag 6720
 gttgatggtg gattcgattg gatcaattgg gtccccttga aggtgaattg ctttgcctgg 6780
 aggttgaaac aagagcgagt tctgtgatg tgtaaattag caaagagagg ggtatatgtg 6840
 gaatccaaaa tatgtaaaat ctgtcaacag gaagaagagg aaaccgaaca tgcttttttc 6900
 aggtgtgcac atgcgcatca ggtgtgggac tggttcaaga tgtggtcggg tctgatgcgg 6960
 gaaatccctc taaacttcag atccatggag gcggagatca aggctgggtc tggtgacaaa 7020
 aaatcgggta aactaggaat ggctttggct tatgtgatgc tgtggactat ctggaaaatc 7080
 aggaatggtg cagtcttcaa caacagaaaa gcgagggcaa tgaacacgac ggatgaggtt 7140
 caattaatct cctttaattg gataaaaaat agaagtagat gcaattggat caaatggtgt 7200
 gattgggggtg ttaagccttg catgaactgt aatatgtagc tgtactcttt gtttctctag 7260
 catcttgcta gaagttttgt tttgctttta taataccaac gccgttcaaa aaaaaaaaaa 7320
 ctattttgct gagtttttga cgaaatgagt gttattttta aacattttt cgctttgtta 7380
 ttttaaaaca tttacaaaaa gggtagtaaa catatttaca aaaagagtgt ttttaaaat 7440
 taacaaccag tgcgtaaaac acccccggtt ctttctctc tgcgtcgact tcagaaatcc 7500
 atctcctacc ggccgacttc accagcggcg gctccctctc catctccggc gccctctcca 7560
 gcgaccgctc cctctccacc ggcgacgtct tctccagagc cggcgacca caggtagggt 7620
 tcttgtttct ttaggggacg tctttgaatt tgctgaaat tgctgaaac aacagtgaga 7680
 tttcaagtca aatagttggt gacctgaaat tgctttttgt gtttcaaatg gaaaggaata 7740
 ggtgttgatt taaatcaata gcttgcatag cctacagtgg gattcaagtc acaggtaggg 7800
 ttctttcttt agaaattgcc tgaaacacct atttgacctt ttcctttcca ttcgaaattg 7860
 cttgcatagc ctacagtggg attcaagtca aagctattgt gatttaaata atataaaggt 7920
 cgtacagcag acaacttgta aaggaacatg ctctgatcat gatccatccc ctgtcagtca 7980
 tttgatcctg tagtagcccc atgctaattg gcactgaggt gactgtcttc aggtaacaca 8040
 ttattgtgta ttacttcaat taatcatgtc cattttatgt tttcatctat gttaatcatt 8100
 tcccgtatt tattattccg tgettactgt actaacattg cagttttata attggtatag 8160
 gaaatgtttg aaggtttagt acggcagctg atattaggtt atcttggcca atatattaaa 8220
 gatatacaca gagaacaact caagatcaca ctgtggaatg gtaagtcacc actctctatc 8280
 atattcatga caaatagttc acgtagtttg ttattcatgg cttgctactt tcattaata 8340
 tcagcacaat atatttactt ttagtcattt acatttatgt ggtttttaag ggtataacag 8400
 cataagctga atactcatat cactcacatt tatggtgaac accaatatgc catttcttct 8460
 atgccccctt cctctttatt tatttatttg taacttggtc tgaggcatgt tttctcgtct 8520

ttgatctctt tatgactcct tcaatatttc gccatattat atattegtat caatatccag 8580
 tgaatttctt atatctgatg tcatttttat gtgaagaaca tactgatcac aatctatgct 8640
 attgaaattt tgctctcatt ttaatgcaga ggaagtgttt ttggaaaatg tggagttaat 8700
 tctggaagct tttgattatc ttgaacttcc gtttgctcta aagcaagggtg atcaataata 8760
 tccgtaaact ttcaagtttc aatactttgt ttatattggt atttttgatg tggtttctca 8820
 tttttgggtt ggttgtcatt atttgacaca tgtgagtggg tgataggacg ggttgggagg 8880
 ctaagcatta gaattccttg gaaaaagctt ggttgggatc ctattataat aatcttagag 8940
 gatataatag tttgtgcctc tcaacgtgag gatgaagagg taagtagtat ctatttcaa 9000
 ggtataataa gtattgtgca tctccattat attatataac ttttactaa tgtgaacagt 9060
 tatccttttt gtagtggagt gttgatgatg ttgaaagacg agaattttct ggaaaaaagg 9120
 ccaaacttgc tgcagcagaa ttggcaaagt tatcgcaacg tgtatgtggg gagtgactat 9180
 tttccatttc tatgtttcat tgttataact cctgttatte attgtggact atgataatat 9240
 gcattaattt aaagaagttg aagactacca ttgctggtag gctttgtatt ttagtgaat 9300
 agtaactatg ttttgccagt cctatatgta aattaaacag ctttttcagc tttcttact 9360
 gttggcagtc tttttaatgc tcagatcttc cagagtttta taagatcacc tttttactat 9420
 acttttttat tgtttagggg aaaagtcagt gttagttaat catagctata tacttatatt 9480
 cactgttaat gcatttaatt ctttcatccc aaaagatgca tttttttgac acaatacata 9540
 ataaaaatgt tggttattgg tcaaatataa aaaatattca tttatgaggt caaaataaag 9600
 atttttattt tttgaaccag aatcaaaatt tcggatatta aataacaata gttaatggga 9660
 aaacacatta agaaatgttg cttgtcgtat atgtatatgg atatatggga atatgggatc 9720
 agtgacattc ttggatacac cctttttttt ttttgcataa aaccagtaaa atctttttca 9780
 cattatttgt gctaatatgt tgcctttttc tgcagataat cagactggga aatcatttat 9840
 gtcatacatt actgccaagg tagtgagatt tcatgttget tgaacttggt aaaaatgtca 9900
 aaaagtcata tgccgttttt tccgttttac ttatcatgta atcttagttt tcttaatcct 9960
 catttgaaat caggtttgat tttacgtgta aatggctgta tagttgtgtc actagtattt 10020
 atttatctat ttattcccat ccagattatt gatggcattc aagtcacat caggaatgtc 10080
 catatcgat atagagatat ttcaaatgaa aatcccaaa ctgtatttgg tgtgaagtgg 10140
 gctagtttga ctgcaatgaa gcaaaactat gctgggtata cctctttttc ttactctaca 10200
 gtcaatcaca tacttaaat ttacaaaaac tcttaatgac tttgatatat gtatatatat 10260
 cttcagggtt ttaagtggaa aggtgagagt tggccaagta aacaaaattg ttgagataca 10320
 aggtttggaa atatactgta aaaccttca tggatcttca acagacatcc atactgaaaa 10380
 tggngaagac tccatggcaa tgggtgctgc aagttatgat aatgatgaac atgctcactt 10440
 gttggcccca gtcaacgtat ctgettctct ttcggtatgc atcatgcaact tgtaccatct 10500
 cataacaaat tttctagatt tttattgtta tgttgaactt tgcaggtgaa taggtctgga 10560
 aggctggaga ataatgcagc acaatactcg gttgatattg agttgtctgg cttggttagt 10620
 tagttcctta acttttctaa tcataattaa ttattaagtt attatgtcag attatctaac 10680
 atgttagtat gtcaccattt ttaggtattg tccctagatg aagatcagtt gcagcaaata 10740
 ctgtatctat atgaatatct atgcacatgt cggctaagag agaagtgagt ggattgcttt 10800
 ttcaccacct tatcctttag aattacattt tcattatcct tttattgttt ttttaagata 10860

tggacgatat cgtccttggg ggaaacctat atcagagaga caattgggat ggcagataca 10920
 gtgggtggcat tatgctcaac actctgtggt atctgatggt cgtaaaagac tgaagaaaac 10980
 ttcatggaaa taccttggag aacgtctgta agtttaattc tttatttaat tttaacatat 11040
 acctgcaagt tttttatgat gaaatcttga aatctgttac atgtacagag gcagaaaacg 11100
 acggtatgta aatctgtaca aattgaaact cgaatgtctt cgaaaagaac aggtgagttt 11160
 tcatgattta aacatcaagg tgtatatgat ttaaaccatta ttttaaatta tgaattatga 11220
 tcagcctttg gatgatgaaa ttgtaatgga gttggaccaa atggagaaaag tgtctgatat 11280
 agaagatata ttgagttaca gatctgctgc tgagaatgaa cttcaggtat gattaagatc 11340
 attgctattg gtatctataa cacatgcatt tatttaactt atacctttta ttaaagacta 11400
 tgaaaatgta ggagttcttg gtggattcac cttctgggat tggaggtagt gaagtgaata 11460
 ctaccattga caagtcaatg gatgatgacc aaacatctgg caaacgcaa ggatggttga 11520
 aatggctgtc ccgtgggatg ctaggtgctg gaggtacaga cgattccagc cagttttctg 11580
 gtgttgtttc agatgaagta atcaaggtaa ggtgaaatga tttcaattcc aattgaatta 11640
 attaatacatt atattaattt gatacaggat atttatgagg caacaaagt tcatctgct 11700
 ccttcccctg tcttggatgc ttctggaact gatagggttc ttttgacctc catcaaatgc 11760
 tctatacatc aaatttctgc aacacttcgc aataagtatg acgtattaat taattaatta 11820
 aatatctaata tatctaatta tctttatagt tttgtaagac ttacttcttt tttctaggaa 11880
 gttggatcga gctattggtg aagtggtttt tgaggggaaat gttgtggagt gcatgatttg 11940
 ggaggaatct gctgttgta ctgcatcaat caattctgta gagatgatta atccattaa 12000
 caatcaagcc attttactta ttaaaagggt gtgtgttctt tttgtttttg ttacatggaa 12060
 agtcttgagt tatttatttt tgaatctttt tctatttctt tcaggtcate tctgaggaga 12120
 gttttcttga agaggagaaa ccgtctttaa atatccaagc ttacattcca caagcaaate 12180
 gtgagggtga cttgacattg aaggtagccc accttgacatt ttttgaaaaa tattttcttg 12240
 tgctataaat ttctgataat tacagtgact cttaaacttc attttaggtt ttgcttgagc 12300
 cgattgaagt gacgtgtgat ccaacatata ttgttaattt catggagcta tatactgtgt 12360
 tgggttccta tacctctcat gaagaaaggg tcagttttta catttctcaa agtagatttc 12420
 ttttaagaac cttctggaag atgcagattg gaagttatat ctagatagtg aattcaatgt 12480
 ttctcatata aaaacactga acatacttta cttatgaga ttttgaataa caacttccat 12540
 tagtttaatg ttacaggtcc taaactcgtc taatgggata aatgatgtaa agtcacgtct 12600
 aatatccaaa gccaaagtat tcatctacta tacctttagg gtctagtaat gggattgata 12660
 gcaacttcat gtttaaagct atatctatct cttctcttct tttgttttgc aggtatattt 12720
 tgtcaggccg gaagagaatg atgtgggata ttagtttgat aaacatcaag ataaatattc 12780
 catgggagaa tgggaactca gagatgcata aattggtaat ttaatttctt attacatggt 12840
 ctacaaaactc acattatttg cattaaactt aaacaaaaac atccaggtac ttgaattaac 12900
 agctgtcacc tttgcatcca agcgcgatat cggetctttt gcaccagata tcaatgtacc 12960
 atctcaattc atgaggaate tgattgatga caattctca aacgagcttc tagaaggaac 13020
 tcacattcaa gatctgtacg atctcttggg aatcaaaaata atcgacttcc aggtcagcat 13080
 atcccatcat ctctcatctt tcacatcaact gatgtttgac tttgactttg actttgactt 13140
 cctgcagata aacttgtttg tgcctttcta tccataact tttctatct tggagaaact 13200

caatgcctct tctgctttat catgggtgcat tgttcaggac gaatctttac taaaagcact 13260
agaggtaggt aaacggtaaa tctcacattt actaattact gttgattttt ttggcacact 13320
ttttaccatc aataattggt ttcaggttta tgtattgggtg gctacccttt tggccatgt 13380
gtcaccatca ataattggct cttttttaga actagttgaa agcatgaaca tgctgcatca 13440
tacttcacaa ttgggcgcca catcagcaac ctctgtcaatt gaaccaagga actctagcag 13500
tatctctggt attgctaatt tggagtctgc tagcatcatt gttgacctg aaaatggctt 13560
agaagctagc tgcacactaa ctgtgtctct tcaggacttg gatatgagga tgggtagtat 13620
gaaatccaca caatctttct ggatatgtac aagggtatta aaagtaactt ctcggttggt 13680
ggaaagtggg gatgacctgg acctataat atgtctgctt caaagtactt cacctaata 13740
tgggtgtctt gtgctacatt atgatggtaa tttgagcata tgtttgagtg atttggatct 13800
tcattgctat ccacatattg ttggattgct ggttgagttt tctggaaagc tatccacata 13860
tagtccttca aatgccaaaa atcaagattt tgtggacagt aacagtaaca ctatactctc 13920
agattcttat atagacttcc agaggtttgg ttgttccaaa atttcagtta atcattacce 13980
ttttgttaca atatacaatg acagatctct tettaacctc gatacttcc ttattaacat 14040
caagaagggt cataagacaa atagttcaaa gttgagggca aaaaaggata atcatcaagt 14100
agggtctcta gttgtaatga atcttgatct caacagcatt agactacatc ttcatgactc 14160
ttcatccatt gttgcatctg ttacacttcc tgtttccaaa tctctttttg ctatccatga 14220
gaactttctg gatgtgttat tttcaactga gggattgagt ctttcatcac agtggtatcc 14280
tcagactcta caagactctc tgtggggccc tgettcactg aatctttctc cagtgatcaa 14340
cattcgtgtg agaaaaggga accatggaat cgaactggat tttagtgttc aaaatgtttc 14400
ctgcatattg ccaactgagt tctggctgc actcattggt tacttctcat tgectgattg 14460
gagctactca aatccaaatg agtcatcacc tactactacc aacaccaata ccaataacaa 14520
cagcatcagt ttacttaca agtttgagat attggactcg gttttattta cacctgtggc 14580
taatcctgat cacgagttta taaagcttaa tattccacag atgtactgca cattcattga 14640
tagcattgat tcagacactc tgttgaaaga aatcccttta gagtgtctctg ttccagttgg 14700
atttactgga aatcagaatt attgtctgaa tgtatttggg agggatttat ctctacatca 14760
tattatttgt cgaaaagaca atgcttctga agtgacaagt gtcagcttga tgcgccttt 14820
tagtggcgat atatggatca caataccata tgaatctaat tcttcttatg caacatgtat 14880
catgtcaagg gttagtaaat gtcagtttac tgttgaagggt aatattttc taccaaactc 14940
tgagttttgt tcatgtttga tacttgatta atgcaatggt gtttcatttt tccaggaaga 15000
gaaatacttg gctgcattgg agcattacag gatgttgctg accaattctc atctgttggc 15060
aatctatcca cttgtttcac ttctgatgtc tcagagtttc tcaatttaaa agaaaattat 15120
gtggttccag ttccattga atettcaact gtcagcttta cagaaatcag atgctctggt 15180
caatccatgt cagtagaact ttactctgac aagatgaatg gtaacagact tatcgccaaa 15240
tccgacatga aattcgcatg tggaaataca atgaaaacag acaagcctct ttctcttgat 15300
atatcattca cttgtttcac accttctca cttcttacct ctgttctctt actggaatgc 15360
acatcatgta ccaaaaatgt accggttctc aacatgcagt tcttgatgtc agatgatgg 15420
aaaaaccacc tgcgattctc cttccttgt gttaacattt ggttgttctt gtctgaatgg 15480
agtcaagttg ttgacctggt caattcttgc tgtgaacctg caatccagaa tgaggaaccg 15540

gaaaaatcca catcagctcc ggtttctcgt gttgatactg cagaaaactc acctcaatcc 15600
accactgtct ccagctatcc atctttagaa gatcggtttt cattaaccgt aaagtcggat 15660
cttattgggtg taaaaatccg tattcctggt caagtttctg gagaagtagt taaatacttt 15720
ggggcccccac aagttcgaga gcagagttta gttacaggaa gagatcacgg aagttttctg 15780
tttatttatac tacaaagtag gtgactgag gtgaacatga agggcgaaac ggtaaatttg 15840
aagtcgaatc tggggaaagc aatgggaaca gttgaactgt ttcagaacaa gagtgtccat 15900
tcttggcccc tttttcagtt actggaaatc gatatagaag ccgaatctgg taatgatgat 15960
atggaccgca tgcatttaaa gacagagatt cactgtgata atcttgatgt ttggctctca 16020
catcatgcat tttacttctg gcaaactatg ctgtttatgt ttccggaagg ttccggatcc 16080
gaatctcctc aacttccagt tggcagtgtc aatttcagat tccacttacg aaaactctcc 16140
attcttttaa cagatgaaaa ggtatgtaaa ctgtaaactg taaagtaaac tgttttctca 16200
tattaaaaaa tataatgaaa ctaatttcaa gtttttaatt agtggagctc caatggctct 16260
ttattggaga ttctcatggg aagtttgta tttcacggaa tcataactgc aaacatgatg 16320
gaagggtcaa tcgacagtga cettcaagtc aactacaaca acatccataa agtcctatgg 16380
gagccttttc tcgaacctg gaagttccaa ataacctta gaagacaaca aggaaaaagc 16440
accctcgaga acagtccagt catgaccgat atccgctcg agtcatcaac aaacctcaac 16500
atcaatgtca ccgaatcctt catcgaggtc gctttcagaa cattcgacat gatcaaagac 16560
gcctcagatc tcataagtct taacattctt cctgaaaaca attcaagatt gacaaaacc 16620
catacaaatg aaaacacact cgcgaaataga tacgccctt acacactcga aaacttgaca 16680
tcaattcctc tggattttta catctcaaaa ggtgaggggt tcaatatgac atcattgaaa 16740
gatggaaaac acgtgcagcc aggctgttca taccctgta atattgataa caaccggaa 16800
gaacaaacgt ttgggttag gcctagtcatt tctactgaca atctaggtgg cgacatgcaa 16860
ttcgtgatg ctcaacatca ttatatagtt gttcaacttg aaggaacttc tacactttcg 16920
gctccagttt ctattgatct tgttgggtc agcttctttg aggtggattt ttctactaat 16980
ataggacttg atgtttctaa aggtggctat gttgttctg tagtaatcga tgtttcagtc 17040
caacggtaca caaagctagt ccgctgttac tcaacggtac gtttatttct ttccactcct 17100
aatatatata tacatatata tatatatata tattgatgtg atgacgtggc aggtcact 17160
gacgaatgca acatcaatgc catttgaagt acggttcgat atcccatttg gtgtatctcc 17220
caagattcta gacctgtat accccggcca tgagtttctt ctctctctac atttagcaga 17280
atcagggcga ataagatggc gtccattagg aagcacttac ctatggagtg aagcttacag 17340
catttcaaat attctctcaa atgaaagcaa gattggacat ttgcggctgt ttgtctgtta 17400
tccttctctt ccaagtagcg accctttctg gtgttgtgtg tccgttcatg atgtgtgttt 17460
gccatctgct ggaagggtaa taaacaagag gggatcatct tcactctctgt ataataata 17520
tactcatgat catggtgaca agatagagaa ccaggatctc tcgaataaga ggtgtattca 17580
cttgattact ttgagtaatc ctttgatagt gaaaaactat ttgccagttg aagtgtctgt 17640
gggtattgag agtgggtggag tttcacgtag catgetgctt tcagaggttg ttatcagtta 17700
tcttattata aactttgaca ttttcttaat gtattgtttt ataacatatg caggttgaaa 17760
ctttttttta tcatattgat tcttcacatg atctttcgtt aacttttgag atacacgggt 17820
ttagaccttc tcttttgaag tttccacgtg ctgaaaagtt cagtggaatt gctaaattca 17880

gtggaacaaa gttttcttct tcagaatcca tcaactttgc tgctgataat tctaaaggtg 17940
 ctgcacactt acttttctca tctccctttt tctctettac attaaaagaa gctgatgtgt 18000
 ctttgtttag gtccattata tgtgacaatg gaaaaggtga tggatgcatt ctctggggct 18060
 cgtgaaatct gcatatttgt gcccttctta ttgtacaact gctgtggttt tccccttacg 18120
 attgcaaatt cgactaatga tctcacaatg cgtgacactt tgccttcatg ttatgatttg 18180
 gatgaagagg acccgTTTTT gggcaaaaaa gatgggtctaa gccttttggt ttccaaccaa 18240
 gtttccaata atgatcctat gagtaatctt gtttccacta gaaaggaatc tttcacctca 18300
 tctggatcaa ccaaaaaaaaa catcggcaca caaaaaccgt ctttacacga tcaggaaaaa 18360
 agtcaacttt cgcatagtca acaacttgac tttgacgaaa caactcgcaa aaaagtcaac 18420
 ttccgcatgt attctcccga ccctaacatc gcttcgagtg agatcatggt gagagttagt 18480
 cgatgccggt ctgatgctga catggcaagt acctcagact acacgtggtc cagtcaattc 18540
 ttcttggtcc caccagccgg ttcaaccaca gtcttgttc cccgatcacc aaccaatgct 18600
 tcatatgtat tatcagttgc ctctagtget atttccgggc catattctgg aaggacaagg 18660
 attatcaatt tccagcctag atacgttate agtaatgcat gcagtcagga cttgtgctat 18720
 aggcagaaag gtccgactt tatataccat ttgaaagcag gacaacactc ccatatccat 18780
 tggacagaca taacaaggta catcaatcat tttcccctaa ctttctgtta ccaaattcac 18840
 atgcttatgt ggcagttgca taaaataaaa caggagttta ctggtgtctg ttcgTTTTGA 18900
 tgagccaggg tggcagtggt caggctgctt ttttcagaa catctaggtg atacacagct 18960
 gaagatgaga aactatgtta gtggtgcagt tagtatggtt cgtgtagagg tccaaaatgc 19020
 tgacgatgca atcagagatg acaagattgt tgggaaccct catggtgaat ctgggacaaa 19080
 tttgattctt ttgtctgatg atgatactgg ctttatgcca tacagggttg ataatttctc 19140
 aaaggagggt agttacttac tccattcttc atttctaatt ctttccattt gatctatgat 19200
 ataagtttgg tgaatgttg ccagaggtta cgtatttacc aacaaaaatg tgaagcattt 19260
 gagactgtta tacattcata tacgtcttgc ctttacgcat gggatgaacc ctccaccaca 19320
 catcgTTTTA ctgttgaggt atgcatttga tttaaaaca gttggctatt aaattattca 19380
 tatattcatt ctgttgTTAA atctccaagg tgtttgctga gagggtagta ggtcttaca 19440
 ctctagatga tgctaaagag tacaacctg tggTTTTACC ttcaacctcc gaggtaaatt 19500
 tgtaattttt acaaacatgg atgggttaaa tctgtaacta actactctt catgatttct 19560
 tttggtagaa acctgaaaga aggttgctaa tatctgtcca tgctgaagga gcattgaagg 19620
 tctgagcat cattgattca agctatcata tatttgatga tgtaaaaatc ccacgttctc 19680
 ctcggttaac tgagaaaaga gaatacgacc aaaaacaaga aagttcactt ctttatcaag 19740
 aaagagtatc gatttccatt ccattcattg ggatttctgt tatgagctct caaccacagg 19800
 taccattact ttatacaaat atcttgteta tatgtggtga atttccat tcttataaaa 19860
 tgtattttgt aggagttgct ttttgcatgt gcaaggaaca caatgattga cctggttcaa 19920
 agtctggacc agcaaaatct ttcttgaaa atctctctt tacagattga taaccaactg 19980
 ccaaccacac cctaccccgT tatattatct ttcgacctg agtataaaca actcccaacc 20040
 tctcagataa aaaacaaaga tgagcctgtg ttctcattgg ctgcagcaaa atggaggaat 20100
 aaagatagag ctttgcttcc attcgagcat ataaacttaa ggtacatgat tataacaata 20160
 tgtcaataaa tagacagtca gaaattagtc aattactgag aattttggtt gttgcagaat 20220

ggcagatttc catcttgagc ttgaacagga tgtgatttta agtctgtttg atttctccaa 20280
 ggcagtatcc tcaaggttcc atagcagagg aatgccacat atggattcag tctgcatcc 20340
 tctttcctca aacttgagtg gaaataaagc aactaaattg gctgaaaaga ctgaaattga 20400
 gggtgaaagt ttccctcttt taccatcaat agtgccaatt ggtgcacat ggcagaaaat 20460
 atatctcctg gcaagaaaac agaagaaaat atatgtggaa cttcttgaag tggcccccacat 20520
 cactttaacc ctaaggtaag acgatcgaca ctgagagggg catatgtgta aatgtgttga 20580
 attggtcttt ttttctgca gcttttcgag cagtccatgg atgctaagga atggaatact 20640
 tacatcagga gaatatctta tccatgtaag tgtgaccatt ggaatagggt tactgatagt 20700
 agtaaagagt aatttgtatg ggattgcaga gaggtctgat ggctcttgct gacgtggagg 20760
 gagcacggat ccatctaagg cggttaacaa tctctcatca gttggccagc ttggaatcca 20820
 tacgagagat cttaatcata cattataccc gccaaactct ccacgaaatg tacaaggtta 20880
 gattacatct ctacatataat aaatettcaa ttgtgattat ttttcttttt cttcttcagg 20940
 tatttggttc agctggggta ataggcaacc ccatgggttt cgcaagaagt gtaggacttg 21000
 gcatccgaga ctctctctca gttccagcca gaagcttcat gcaggtata aataaataat 21060
 taaatattaa atccaagtta attagtaatg tcaattttaa ttatatcgca gageccccga 21120
 ggacttatca cggggatggc acagggaact acaagtcttc taagcaatac ggtttacgcc 21180
 ataagtgatg ctgccacca agtcagtaga gccgcacaca aggtatata atatacat 21240
 tatattatac ttttctcaat tctaagattt catcatttac ttgaaggga ttgttgcat 21300
 tacaatggac gacccccat ctgcagcaga aatgggcaaa ggtgtaata atgaagtttt 21360
 ggaggggctg actggtcttc tccaatcacc aataagagga gctgaaaaac acgggcttcc 21420
 aggagtctt tcaggtatag cactaggagt aacgggteta gtggcaaggc cagccgccag 21480
 cactactggaa gtaacagaaa aaaccgcccg cagcataaga aaccgaagca aactctacca 21540
 catccgcctc cgggtccgcc tcccagacc gtaaccccc aaccaccacc cgttaaaacc 21600
 ctactcgtgg gaccaagcag tcggcctctc cgtcctcacc aataccaatt ccaattccga 21660
 ttccgacctc aaagacgaaa ccctcgtcct ctccaaatcc ctcaaaaa aggcaaat 21720
 cgtcattatc acccaacggt tactcctcat tgttacctc tcgagcctaa cgaatttagg 21780
 tcaacccaat ttcaaaggcg tcctgcgga ccccgattgg gtggttgaag ccgagataac 21840
 gttggatagt gtgatacacg tggatgttga tggagagggt gtgcatattg tcgggagtag 21900
 ttctgatgtg gtggttagac agaattgttg tgggaagcag cgggtgtata atccgttgcc 21960
 gctgtttcag acgaatttgg agtgtttagg gaaggaggag gcgggggagt tgttgaagg 22020
 gttgttgggt acgattgaga gagggaagga gagagggtgg ggccgggggt gtgtgtaccg 22080
 tctgcatcag agtaatgta ggtgatgtat atttttttc tacatataaa gttactatag 22140
 gagaaaaagg actggatatt atattataca tactgaaac aaggaaact tttctttcaa 22200
 aattttggct gtattattat tttgtcgacc atgttgggt aaaatggcca attatttact 22260
 tatgacatgg ttaaaaaata ttggtgtctt gttttgtata attacaatt atattagtat 22320
 cgatgcaatg taagattgta gaaagcgeta ccgtataaaa caacataagt catgaggtta 22380
 caccctagtg gggtaaggga aacaaaaaca ttttaaacgt ttttcagatt tgcatttcca 22440
 gccgcctata actaccatc taattcttat ccgacatctt aacacgatat agggattttt 22500
 actttacttg tcaaaacttg aagttgacag tttgataata tccagtcctc ttttccaata 22560

gtaacttttc cagccgcca taactatcat ttttttttg ttataactat gttctatcta 22620
 atttagttgt tttttattat ttttgataaa aatatgtttt agttgtttct gattttgttt 22680
 atatttaatt attattgttt tttttgctaa aacatatcaa atttgttttg aattttgtct 22740
 atgtatagtt tttatgataa attatttgta aatttagccg gttaaaatga atttttcggc 22800
 caagcatgaa catttcgatg aacgaacgca catgcttttt ataggttatc tacttgagca 22860
 cgtgcttcaa tttcataatt gcattcaata atccattgaa ctacttgtaa ttttgctcta 22920
 gttccttgcc agagtggttt tctttattat agggatattt tcatttataat ggtgactttt 22980
 tctaaaaaag gtaaacattt ttttagattt gcaaataatc ggtcataagg actggaaaaa 23040
 tcaaaaacaa caatgatgca ttttactaaa aaaaacaata gtgacaaaat atgaacaaaa 23100
 tcagaaataa tgacgtttta cataaaaaat aacgactaaa tatagacaaa acgagaacta 23160
 agttatcctt gtaagttatt atcccctgat tatattatca aacaaatttc gtatgacaca 23220
 atacaggtct aaaaacctac ctcaactact tacatecgec aaaatttaca tcttatctat 23280
 tcatattaaa cctatacttt gaaattttga actccatgaa ttaaagtcta aaatttctaa 23340
 aatgatgaaa accctaattc a 23361

<210> 2

<211> 11799

<212> DNA

<213> *Taraxacum officinale*

<400> 2

atgtcagatc tgataaactg taagcaacct catgaatcac ccaactcctgc cgccacaacc 60
 acagccgaag cttcatttga gccctggat gaaaatctct tctcggatct aacccactc 120
 tgttatttac tccaatccac ctccccaaa tcaccccaa cattaatctt tctcctgctc 180
 taccaccacc aatcctctca agattcatcc tttctcatcg ccgtgacggt ttccagcaaa 240
 acccatgtcc accaccgccg ccgtcaaacc agtccggaaa ttcccgccat cgtgcccgga 300
 tgtaaccccc cgtcaaacc agtccggaaa tcaacaacac aacccctgt tttgcatcac 360
 cagagcttga atccgggtcg tggagttcga tttaaactat cttcaactca aaatctggta 420
 acattagcag ctagttccac caatcatacc cctgatttgc attctagggt tccaatcac 480
 gattcctcat acatgaacaa ctgggtctaat caagaagaag acaacgacga tgggtgtctt 540
 gatgactcta caattaatga aacctcgatt catccgggag tgatactaaa gaggattcaa 600
 gcacaaaaag gcattgagtt taaacggagg agaaaatag ccatggccag ctaccgccg 660
 gtcgccatcc cagccagcat tagcccagcc gtcgctcaca tccaccgaaa agctgatttc 720
 cacatcgtcg ctcccggccg tcacctacac gtaggttcgg ctccgaagaa agtggtcgat 780
 aagctagaag ggataagaag aaggttctta tggggcggtg agaagtcgga aaaaagatt 840
 cactggggtt cgtgggagaa ggtgataaaa tcaaaagata aagaaggttt aggggtgaat 900
 ggattgagca gcatgaatat ggcttgcata gtgaaatggt tttggcggct caaaactgaa 960
 agggacagcc tgtgggttag atgtgtcagc gcttgcata atatcaact tattgatggg 1020
 aaacgggtgg ctaaagcttc cttgaaagga gtatggtgga acatcatgag ctgtgttgaa 1080
 gagttaaaaa cgaaaggaat ttctgtggag tcaaagttag taaggcaact gggaaacggc 1140
 aaacacacgc atttttgtaa ggatagatgg ttacacaaca aagttttaa agatgacctt 1200

ccggagttgt acaaaataga aggggacaaa aattgtatgg taaaccaaag actggtttgg 1260
 gacaacaacg agaaaatatt caagcaagct tgggactgga aaagaccgat caggagagga 1320
 agagaaacca aagaactcga aactttaata attttgacia atgggataca attaaaggaa 1380
 atagaagata attggagatg gaaggagggga tcggacggga aattttcggg gggaaaattg 1440
 aggaaactct ttgcttatca ggagcaggcc gaggttgatg gtggattcga ttgatcaat 1500
 tgggtcccct tgaaggaaga agaggaaacc gaacatgctt ttttcagggtg tgcacatgcg 1560
 catcaggtgt gggactgggt caagatgtgg tcgggtctga tgcgggaaat ccctctaac 1620
 ttcagatcca tggaggcgga gatcaaggct ggtgctgggt acaaaaaatc ggtgaaacta 1680
 ggaatggctt tggcttatgt gatgctgtgg actatctgga aatcaggaa tggcgcagtc 1740
 ttcaacaaca gaaaagcgag ggcaatgaac acgacggatg agaaatccat ctctaccgg 1800
 ccgacttcac cagcggcggc tccctctcca tctccggcgc cctctccagc gaccgctccc 1860
 tctccaccgg cgacgtcttc tccagagccg gcgaccaca gcccctgct aattggcact 1920
 gaggaaatgt ttgaaggttt agtacggcag ctgatattag gttatcttgg ccaatatatt 1980
 aaagatatac acagagaaca actcaagatc aactgtgga atgaggaagt gtttttggaa 2040
 aatgtggagt taattctgga agcttttgat tatcttgaac ttcctgttgc tctaaagcaa 2100
 ggacgggttg ggaggctaag cattagaatt ccttgaaaa agcttggttg ggatcctatt 2160
 ataataatct tagaggatat attagtttgt gcctctcaac gtgaggatga agagtggagt 2220
 gttgatgatg ttgaaagacg agaattttct ggaaaaaagg ccaaacttgc tgcagcagaa 2280
 ttggcaaagt tatcgcaacg tgtatgtgat aatcagactg ggaaatcatt tatgtcatac 2340
 attactgcca agattattga tggcattcaa gtcacatca ggaatgtcca tatcgtatat 2400
 agagatatatt caaatgaaaa atcccaaact gattttgggt tgaagttggc tagtttgact 2460
 gcaatgaagc aaaactatgc tgggttatta agtggaaagg tgagagttgg ccaagtaaac 2520
 aaaattgttg agatacaagg tttggaaata tactgtaaaa ctttctatgg atcttcaaca 2580
 gacatccata ctgaaaatgg tgaagactcc atggcaatgg tggctgcaag ttatgataat 2640
 gatgaacatg ctcaactgtt ggccccagtc aacgtatctg cttctcttcc ggtgaatagg 2700
 tctggaaggc tggagaataa tgcagcacia tactcggttg atattgagtt gtctggcttg 2760
 gtattgtccc tagatgaaga tcagttgcag caaatactgt atctatatga atatctatgc 2820
 acatgtcggc taagagagaa atatggacga tatcgtcctt ggggaaacc tatatcagag 2880
 agacaattgg gatggcagat acagtgggtg cattatgctc aacctctgt gttatctgat 2940
 gttcgtaaaa gactgaagaa aacttcatgg aaatacctg gagaacgtct aggcaaaaa 3000
 cgacgggatg taaatctgta caaattgaaa ctcgatgtc ttcgaaaaga acagccttg 3060
 gatgatgaaa ttgtaatgga gttggaccaa atggagaaag tgtctgatat agaagatata 3120
 ttgagttaca gatctgctgc tgagaatgaa cttcaggagt tcttgggtgga ttcaccttct 3180
 ggtattggag gtagtgaagt gaatactacc attgacaagt caatggatga tgaccaaaaca 3240
 tctggcaaac cgcaaggatg gttgaaatgg ctgtcccgtg gtatgctagg tgctggaggt 3300
 acagacgatt ccagccagtt ttctgggtgt gtttcagatg aagtaatcaa ggatatttat 3360
 gaggaacaa agtttcatcc tgetccttcc cctgtcttgg atgcttctgg aactgatagg 3420
 gttcttttga cctccatcaa atgetctata catcaaatct ctgcaacact tcgcaataag 3480
 aagttggatc gagctattgg tgaagtgggt tttgagggga atgttgtgga gtgcatgatt 3540

tgggaggaat ctgctgttgt tactgcatca atcaattctg tagagatgat taatccatta 3600
 aacaatcaag ccattttact tattaaaagg gtcactctg aggagagttt tcttgaagag 3660
 gagaaaccgt ctttaaatat ccaagcttac attccacaag caaatcgtga gggtgacttg 3720
 acattgaagg ttttgcttga gccgattgaa gtgacgtgtg atccaacata tcttgtaaat 3780
 ttcattggagc tataactgt gttgggttcc tatacctctc atgaagaaag ggtcctaaac 3840
 tcgcttaatg ggataaatga tgtaaagtca cgtctaatat ccaaagccaa gtatattttg 3900
 tcaggccgga agagaatgat gtgggatatt agtttgataa acatcaagat aaatattcca 3960
 tgggagaatg ggaactcaga gatgcataaa ttggtacttg aattaacagc tgtcaccttt 4020
 gcatccaagc gcgatatcgg ctcttttgea ccagataatc atgtaccatc tcaattcatg 4080
 aggaatctga ttgatgacaa ttcttcaaac gagcttctag aaggaaactc cattcaagat 4140
 ctgtacgac tcttggaat caaaaataatc gacttccagg acgaatcttt actaaaagca 4200
 ctagaggttt atgtattggt ggetaccett ttggcacatg tgtcaccatc aataattggc 4260
 tcatttttag aactagtga aagcatgaac atgctgcatc atacttcaca attgggcgcc 4320
 acatcagcaa cctcgtcaat tgaaccaagg aactctagca gtatctctgt tattgctaat 4380
 ttggagtctg ctagcatcat tgttgacctt gaaaatggct tagaagctag ctgcacacta 4440
 actgtgtctc ttcaggactt ggatatgagg atgggtagta tgaaatccac acaatctttc 4500
 tggatatgta caagggattt aaaagtaact tctcggttgt tggaaagtgg tgatgacctg 4560
 gacctataa tatgtctgcc tcaaagtact tcacctaatg atgggtgtct tgtgctacat 4620
 tatgatggta atttgagcat atgtttgagt gatttgatc ttcattgcta tccacatatt 4680
 gttggattgc tggttgagtt ttctggaaag ctatccacat atagtccttc aaatgccaaa 4740
 aatcaagatt ttgtggacag taacagtaac actatactct cagattetta tatagacttc 4800
 cagaggtttg gttgttccaa aatttcagtt aatcattacc cttttgttac aatatacaat 4860
 gacagatctc ttcttaacct cgatacttca cttattaaca tcaagaaggt tcataagaca 4920
 aatagttcaa agttgagggc aaaaaaggat aatcatcaag taggtgctct agttgtaatg 4980
 aatcttgatc tcaacagcat tagactacat cttcatgact cttcatccat tgttgcactc 5040
 gttacacttc ctgtttccaa atcctctttt gctatccatg agaactttct ggatgtgtta 5100
 ttttcaactg agggattgag tctttcatca cagtggatc ctcagactct acaagactct 5160
 ctgtggggcc ctgcttact gaatctttct ccagtgatca acattcgtgt gagaaaaggg 5220
 aacatggaa tcgaactgga ttttagtggt caaaatgttt cctgcatatt gccaaactgag 5280
 ttctggctg cactcattgg ttacttctca ttgctgatt ggagctactc aaatccaaat 5340
 gagtcatcac ctactactac caacaccaat accaataaca acagcatcag tttcacttac 5400
 aagtttgaga tattggactc ggttttattt acacctgtgg ctaatctga tcacgagttt 5460
 ataaagctta atattccaca gatgtactgc acattcattg atagcattga ttcagacact 5520
 ctgttgaaag aaatcccttt agagtgetct gttccagttg gatttactgg aaatcagaat 5580
 tattgtctga atgtatttg gagggattta tctctacatc atattatttg tcgaaaagac 5640
 aatgcttctg aagtgacaag tgtcagettg atcgcgctt ttagtggcga tatatggatc 5700
 acaataccat atgaatctaa ttcttcttat gcaacatgta tcatgtcaag ggttagtaaa 5760
 tgtcagttta ctgttgaagg aagagaaata cttggctgca ttggagcatt acaggatggt 5820
 gtcgaccaat tctcatctgt tggcaatcta tccacttgtt tcacttctga tgtctcagag 5880

tttctcaatt taaaagaaaa ttatgtgggt ccagttceca ttgaatcttc aactgtcagc 5940
tttacagaaa tcagatgctc tgttcaatcc atgtcagtag aactttactc tgacaagatg 6000
aatggtaaca gacttatcgc caaatccgac atgaaattcg catgtggaat atcaatgaaa 6060
acagacaagc ctctttctct tgatataca ttactttgtt tcacactttc ctactttctt 6120
acctctgttg tcttactgga atgcacatca tgtaccaaaa atgtaccggt tctcaacatg 6180
cagttcttga tgtcagatga tggtaaaaaac cacctgcgat tctcccttcc ttgtgttaac 6240
atttggttgt tcttgtctga atggagtcaa gttgttgacc tggtaattc ttgctgtgaa 6300
cctgcaatcc agaattgagga accggaaaaa tccacatcag ctccggtttc tctgtttgat 6360
actgcagaaa actcacctca atccaccact gtctccagct atccatcttt agaagatcgg 6420
ttttcattaa ccgtaaagtc ggatcttatt ggtgtaaaaa tccgtattcc tgttcaagtt 6480
tctggagaag tagttaaata ctttggggcc ccacaagttc gagagcagag tttagttaca 6540
ggaagagatc acggaagttt tctgtttatt tatctacaaa gtaggtgcac tgaggtgaac 6600
atgaagggcg aaacggtaaa tttgaagtcg aatctgggga aagcaatggg aacagttgaa 6660
ctgtttcaga acaagagtgt ccattcttgg ccccttttcc agttactgga aatcgatata 6720
gaagccgaat ctggtaatga tgatatggac cgcattgcatt taaagacaga gattcactgt 6780
gataatcttg atgtttgget ctcacatcat gcattttact tctggcaaac tatgctgttt 6840
atgtttccgg aaggttccgg atccgaatct cctcaacttc cagttggcag tgtcaatttc 6900
agattccact tacgaaaact ctccattctt ttaacagatg aaaagtggag ctccaatggt 6960
cctttatttg agattctcat ggggaagttt ttatttcacg gaatcataac tgcaaacatg 7020
atggaagggc caatcgacag tgacctcaa gtcaactaca acaacatcca taaagtecta 7080
tgggagcctt ttctcgaacc atggaagttc caaataacct taagaagaca acaaggaaaa 7140
agcacctcgc agaacagtcc agtcatgacc gatatccgcc tcgagtcate acaaacctc 7200
aacatcaatg tcaccgaatc ctctatcgag gtcgcttcca gaacattcga catgatcaaa 7260
gacgcctcag atctcataag tcttaacatt ctctctgaaa acaattcaag attgacaaaa 7320
ccccatacaa atgaaaacac actcgcgaat agatagccc cttacacact cgaaaacttg 7380
acatcacttc ctctgggtatt ttacatctca aaaggtgagg ggttcaatat gacatcattg 7440
aaagatggaa aacacgtgca gccaggctgt tcatatcccg tgaatattga taacaacccg 7500
gaagaacaaa cgtttgggtt taggcctagt cattctactg acaatctagg tggcgacatg 7560
caattcgctg atgctcaaca tcattatata gttgttcaac ttgaaggaaac ttctacactt 7620
tcggctccag tttctattga tcttgttggg gtcagcttct ttgaggtgga tttttctact 7680
aatataggac ttgatgttcc taaaggtggc tatgttggtc ctgtagtaat cgatgtttca 7740
gtccaacggt acacaaaagct agtccgcttg tactcaacgg tcatactgac gaatgcaaca 7800
tcaatgcat ttgaagtaac gttcgatata ccatttgggt tatctcccaa gattctagac 7860
cctgtatacc ccggccatga gtttctctt cctctacatt tagcagaatc agggcgaata 7920
agatggcgtc cattaggaag cacttaccta tggagtgaag cttacagcat ttcaaatatt 7980
ctctcaaatg aaagcaagat tggacatttg cggctgcttg tctgttatec ttctcttcca 8040
agtagcgacc cctttcgggt ttgtgtgtcc gttcatgatg tgtgtttgcc atctgctgga 8100
agggtaataa acaagagggg atcatcttca tctctgtata atattaatac tcatgatcat 8160
ggtgacaaga tagagaacca ggatctctcg aataagaggt gtattcactt gattactttg 8220

agtaatcctt tgatagtgaa aaactatttg ccagttgaag tgtctgtggt gattgagagt 8280
 ggtggagttt cacgtagcat gctgcttca gaggttgaaa ctttttttta tcatattgat 8340
 tcttcacatg atctttcgtt aacttttgag atacacgggt ttagacctc tcttttgaag 8400
 tttccacgtg ctgaaaagtt cagtggatt gctaaattca gtggaacaaa gttttcttct 8460
 tcagaatcca tcaactttgc tgctgataat tctaaaggct cattatatgt gacaatggaa 8520
 aaggtgatgg atgcattctc tggggctcgt gaaatctgca tttttgtgcc cttcttattg 8580
 tacaactgct gtggttttcc ctttacgatt gcaaattcga ctaatgatct cacaatgcgt 8640
 gacactttgc cttcatgtta tgatttggat gaagaggacc cgtttttggg caaaaaagat 8700
 ggtctaagcc ttttgttttc caaccaagtt tccaataatg atcctatgag taatcttgtt 8760
 tccactagaa aggaatcttt cacctcatct ggatcaacca aaaaaaacat cggcacacaa 8820
 aaaccgtctt tacacgatca ggaaaaaagt caactttcgc atagtcaaca acttgacttt 8880
 gacgaaacaa ctgcacaaaa agtcaacttc cgcattgtatt ctcccgacc taacatcgct 8940
 tcgagtgaga tcatggtgag agttagtcga tgccggtctg atgctgacat ggcaagtacc 9000
 tcagactaca cgtggtccag tcaattcttc ctggctccac cageccggtc aaccacagtc 9060
 cttgtccccc gatcatcaac caatgettca tatgtattat cagttgctc tagtgctatt 9120
 tccgggcatc attctggaag gacaaggatt atcaatttcc agcctagata cgttatcagt 9180
 aatgcatgca gtcaggactt gtgctatagg cagaaaggtt ccgactttat atacatttg 9240
 aaagcaggac aacactccca tatccattgg acagacataa caaggagtt actggtgtct 9300
 gttcgtttcg atgagccagg gtggcagtggt tcaggctgct tttttccaga acatctaggt 9360
 gatacacagc tgaagatgag aaactatggt agtgggtcag ttagtatggt tcgtgtagag 9420
 gtccaaaatg ctgacgatgc aatcagagat gacaagattg ttgggaacce tcatggtgaa 9480
 tctgggacaa atttgattct tttgtctgat gatgatactg gctttatgcc atacagggtt 9540
 gataatttct caaaggagag gttacgtatt taccaacaaa aatgtgaagc atttgagact 9600
 gttatacatt catatacgtc ttgcccttac gcatgggatg aacctccta cccacatcgt 9660
 ttaactgttg aggtgtttgc tgagagggtg gtagggctctt aactctaga tgatgctaaa 9720
 gagtacaaac ctgtggtttt acctcaacc tccgagaaac ctgaaagaag gttgctaata 9780
 tctgtccatg ctgaaggagc attgaaggct ctgagcatca ttgattcaag ctatcatata 9840
 tttgatgatg taaaaatccc acgttctct cggttaactg agaaaagaga atacgaccaa 9900
 aaacaagaaa gttcacttct ttatcaagaa agagtatcga tttccattcc attcattggg 9960
 atttctgtta tgagctctca accacaggag ttgcttttg catgtgcaag gaacacaatg 10020
 attgacctgg ttcaaagtct ggaccagcaa aatcttctct tgaaaatctc ctctttacag 10080
 attgataacc aactgccaac cacacctac cccgttatat tatctttcga ccatgagtat 10140
 aaacaactcc caacctctca gataaaaaac aaagatgagc ctgtgttctc attggctgca 10200
 gcaaaatgga ggaataaaga tagagctttg ctttcattcg agcatataaa cttagaatg 10260
 gcagatttcc atcttgagct tgaacaggat gtgattttaa gtctgtttga tttctccaag 10320
 gcagtatcct caaggttcca tagcagagga atgccacata tggattcagt cgtgcatcct 10380
 ctttctcaa acttgagtgg aaataaagca actaaattgg ctgaaaagac tgaaattgag 10440
 ggtgaaaagt tccctctttt accatcaata gtgccattg gtgcacatg gcagaaaata 10500
 tatctctctg caagaaaaca gaagaaaata tatgtggaac ttcttgaagt ggccccatc 10560

actttaacce taagcttttc gagcagtcca tggatgctaa ggaatggaat acttacatca 10620
 ggagaatatac ttatccatag aggtctgatg gctcttgctg acgtggaggg agcacggatc 10680
 catctaaggc ggtaacaat ctctcatcag ttggccagct tggatccat acgagagatc 10740
 ttaatcatac attatacccg ccaacttctc cacgaaatgt acaaggtatt tggttcagct 10800
 ggggtaatag gcaaccccat gggtttcgca agaagtgtag gacttggcat ccgagacttc 10860
 ctctcagttc cagccagaag cttcatgcag agccccgcag gacttatcac ggggatggca 10920
 cagggaaacta caagtcttct aagcaatacg gtttacgcca taagtgatgc tgccacccaa 10980
 ggcattgttg catttacaat ggacgacccc ccatctgcag cagaaatggg caaaggtgta 11040
 ataaatgaag ttttggaggg gctgactggt cttctccaat caccaataag aggagctgaa 11100
 aaacacgggc ttccaggagt ctttcaggt atagcactag gagtaacggg tctagtggca 11160
 aggccagccg ccagcatact ggaagtaaca gaaaaaacgg cccgcagcat aagaaaccga 11220
 agcaaaactct accacatccg cctccgggtc cgctcccga gaccgctaac cccaaccac 11280
 caccggttaa aaccctactc gtgggaccaa gcagtcggcc tctccgtect caccaatacc 11340
 aattccaatt ccgattccga cctcaaagac gaaacctcg tectctcaa atccctcaa 11400
 caaaagggca aattcgtcat taccaccaa cggttactcc tcattgttac ctctcgagc 11460
 ctaacgaatt taggtcaacc caatttcaa ggcgtccctg cggaccccga ttgggtggtt 11520
 gaagccgaga taacgttga tagtgtgata cacgtggatg ttgatggaga ggtggtgcat 11580
 attgtcggga gtagttctga tgtggtggtt agacagaatg ttggtgggaa gcagcgggtg 11640
 tataatccgt tgccgtgtt tcagacgaat ttggagtgtt tagggaagga ggagcgggg 11700
 gagttgttga aggtgtgtt ggtgacgatt gagagaggga aggagagagg gtggggccgg 11760
 ggggtgtgtg accgtctgca tcagagtaat gttaggtga 11799

<210> 3

<211> 3932

<212> PRT

<213> Taraxacum officinale

<400> 3

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

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

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

	725		730		735										
Glu	Glu	Trp	Ser	Val	Asp	Asp	Val	Glu	Arg	Arg	Glu	Phe	Ser	Gly	Lys
	740		745		750										
Lys	Ala	Lys	Leu	Ala	Ala	Ala	Glu	Leu	Ala	Lys	Leu	Ser	Gln	Arg	Val
	755		760		765										
Cys	Asp	Asn	Gln	Thr	Gly	Lys	Ser	Phe	Met	Ser	Tyr	Ile	Thr	Ala	Lys
	770		775		780										
Ile	Ile	Asp	Gly	Ile	Gln	Val	Thr	Ile	Arg	Asn	Val	His	Ile	Val	Tyr
785			790		795		800								
Arg	Asp	Ile	Ser	Asn	Glu	Lys	Ser	Gln	Thr	Val	Phe	Gly	Val	Lys	Leu
	805		810		815										
Ala	Ser	Leu	Thr	Ala	Met	Lys	Gln	Asn	Tyr	Ala	Gly	Val	Leu	Ser	Gly
	820		825		830										
Lys	Val	Arg	Val	Gly	Gln	Val	Asn	Lys	Ile	Val	Glu	Ile	Gln	Gly	Leu
	835		840		845										
Glu	Ile	Tyr	Cys	Lys	Thr	Phe	His	Gly	Ser	Ser	Thr	Asp	Ile	His	Thr
	850		855		860										
Glu	Asn	Gly	Glu	Asp	Ser	Met	Ala	Met	Val	Ala	Ala	Ser	Tyr	Asp	Asn
865			870		875		880								
Asp	Glu	His	Ala	His	Leu	Leu	Ala	Pro	Val	Asn	Val	Ser	Ala	Ser	Leu
	885		890		895										
Ser	Val	Asn	Arg	Ser	Gly	Arg	Leu	Glu	Asn	Asn	Ala	Ala	Gln	Tyr	Ser
	900		905		910										
Val	Asp	Ile	Glu	Leu	Ser	Gly	Leu	Val	Leu	Ser	Leu	Asp	Glu	Asp	Gln
	915		920		925										
Leu	Gln	Gln	Ile	Leu	Tyr	Leu	Tyr	Glu	Tyr	Leu	Cys	Thr	Cys	Arg	Leu
	930		935		940										
Arg	Glu	Lys	Tyr	Gly	Arg	Tyr	Arg	Pro	Trp	Gly	Lys	Pro	Ile	Ser	Glu
945			950		955		960								
Arg	Gln	Leu	Gly	Trp	Gln	Ile	Gln	Trp	Trp	His	Tyr	Ala	Gln	His	Ser
	965		970		975										
Val	Leu	Ser	Asp	Val	Arg	Lys	Arg	Leu	Lys	Lys	Thr	Ser	Trp	Lys	Tyr
	980		985		990										
Leu	Gly	Glu	Arg	Leu	Gly	Arg	Lys	Arg	Arg	Tyr	Val	Asn	Leu	Tyr	Lys
	995		1000		1005										
Leu	Lys	Leu	Glu	Cys	Leu	Arg	Lys	Glu	Gln	Pro	Leu	Asp	Asp	Glu	
	1010		1015		1020										
Ile	Val	Met	Glu	Leu	Asp	Gln	Met	Glu	Lys	Val	Ser	Asp	Ile	Glu	
	1025		1030		1035										

Asp Ile Leu Ser Tyr Arg Ser Ala Ala Glu Asn Glu Leu Gln Glu 1040	1045	1050
Phe Leu Val Asp Ser Pro Ser Gly Ile Gly Gly Ser Glu Val Asn 1055	1060	1065
Thr Thr Ile Asp Lys Ser Met Asp Asp Asp Gln Thr Ser Gly Lys 1070	1075	1080
Pro Gln Gly Trp Leu Lys Trp Leu Ser Arg Gly Met Leu Gly Ala 1085	1090	1095
Gly Gly Thr Asp Asp Ser Ser Gln Phe Ser Gly Val Val Ser Asp 1100	1105	1110
Glu Val Ile Lys Asp Ile Tyr Glu Ala Thr Lys Phe His Pro Ala 1115	1120	1125
Pro Ser Pro Val Leu Asp Ala Ser Gly Thr Asp Arg Val Leu Leu 1130	1135	1140
Thr Ser Ile Lys Cys Ser Ile His Gln Ile Ser Ala Thr Leu Arg 1145	1150	1155
Asn Lys Lys Leu Asp Arg Ala Ile Gly Glu Val Val Phe Glu Gly 1160	1165	1170
Asn Val Val Glu Cys Met Ile Trp Glu Glu Ser Ala Val Val Thr 1175	1180	1185
Ala Ser Ile Asn Ser Val Glu Met Ile Asn Pro Leu Asn Asn Gln 1190	1195	1200
Ala Ile Leu Leu Ile Lys Arg Val Ile Ser Glu Glu Ser Phe Leu 1205	1210	1215
Glu Glu Glu Lys Pro Ser Leu Asn Ile Gln Ala Tyr Ile Pro Gln 1220	1225	1230
Ala Asn Arg Glu Gly Asp Leu Thr Leu Lys Val Leu Leu Glu Pro 1235	1240	1245
Ile Glu Val Thr Cys Asp Pro Thr Tyr Leu Val Asn Phe Met Glu 1250	1255	1260
Leu Tyr Thr Val Leu Gly Ser Tyr Thr Ser His Glu Glu Arg Val 1265	1270	1275
Leu Asn Ser Leu Asn Gly Ile Asn Asp Val Lys Ser Arg Leu Ile 1280	1285	1290
Ser Lys Ala Lys Tyr Ile Leu Ser Gly Arg Lys Arg Met Met Trp 1295	1300	1305
Asp Ile Ser Leu Ile Asn Ile Lys Ile Asn Ile Pro Trp Glu Asn 1310	1315	1320
Gly Asn Ser Glu Met His Lys Leu Val Leu Glu Leu Thr Ala Val		

1325	1330	1335
Thr Phe Ala Ser Lys Arg Asp Ile Gly Ser Phe Ala Pro Asp Ile		
1340	1345	1350
Asn Val Pro Ser Gln Phe Met Arg Asn Leu Ile Asp Asp Asn Ser		
1355	1360	1365
Ser Asn Glu Leu Leu Glu Gly Thr His Ile Gln Asp Leu Tyr Asp		
1370	1375	1380
Leu Leu Glu Ile Lys Ile Ile Asp Phe Gln Asp Glu Ser Leu Leu		
1385	1390	1395
Lys Ala Leu Glu Val Tyr Val Leu Val Ala Thr Leu Leu Ala His		
1400	1405	1410
Val Ser Pro Ser Ile Ile Gly Ser Phe Leu Glu Leu Val Glu Ser		
1415	1420	1425
Met Asn Met Leu His His Thr Ser Gln Leu Gly Ala Thr Ser Ala		
1430	1435	1440
Thr Ser Ser Ile Glu Pro Arg Asn Ser Ser Ser Ile Ser Val Ile		
1445	1450	1455
Ala Asn Leu Glu Ser Ala Ser Ile Ile Val Asp Leu Glu Asn Gly		
1460	1465	1470
Leu Glu Ala Ser Cys Thr Leu Thr Val Ser Leu Gln Asp Leu Asp		
1475	1480	1485
Met Arg Met Gly Ser Met Lys Ser Thr Gln Ser Phe Trp Ile Cys		
1490	1495	1500
Thr Arg Asp Leu Lys Val Thr Ser Arg Leu Leu Glu Ser Gly Asp		
1505	1510	1515
Asp Leu Asp Leu Ile Ile Cys Leu Pro Gln Ser Thr Ser Pro Asn		
1520	1525	1530
Asp Gly Cys Leu Val Leu His Tyr Asp Gly Asn Leu Ser Ile Cys		
1535	1540	1545
Leu Ser Asp Leu Asp Leu His Cys Tyr Pro His Ile Val Gly Leu		
1550	1555	1560
Leu Val Glu Phe Ser Gly Lys Leu Ser Thr Tyr Ser Pro Ser Asn		
1565	1570	1575
Ala Lys Asn Gln Asp Phe Val Asp Ser Asn Ser Asn Thr Ile Leu		
1580	1585	1590
Ser Asp Ser Tyr Ile Asp Phe Gln Arg Phe Gly Cys Ser Lys Ile		
1595	1600	1605
Ser Val Asn His Tyr Pro Phe Val Thr Ile Tyr Asn Asp Arg Ser		
1610	1615	1620

Leu Leu Asn Leu Asp Thr Ser Leu Ile Asn Ile Lys Lys Val His 1625	1630	1635
Lys Thr Asn Ser Ser Lys Leu Arg Ala Lys Lys Asp Asn His Gln 1640	1645	1650
Val Gly Ala Leu Val Val Met Asn Leu Asp Leu Asn Ser Ile Arg 1655	1660	1665
Leu His Leu His Asp Ser Ser Ser Ile Val Ala Ser Val Thr Leu 1670	1675	1680
Pro Val Ser Lys Ser Ser Phe Ala Ile His Glu Asn Phe Leu Asp 1685	1690	1695
Val Leu Phe Ser Thr Glu Gly Leu Ser Leu Ser Ser Gln Trp Tyr 1700	1705	1710
Pro Gln Thr Leu Gln Asp Ser Leu Trp Gly Pro Ala Ser Leu Asn 1715	1720	1725
Leu Ser Pro Val Ile Asn Ile Arg Val Arg Lys Gly Asn His Gly 1730	1735	1740
Ile Glu Leu Asp Phe Ser Val Gln Asn Val Ser Cys Ile Leu Pro 1745	1750	1755
Thr Glu Phe Leu Ala Ala Leu Ile Gly Tyr Phe Ser Leu Pro Asp 1760	1765	1770
Trp Ser Tyr Ser Asn Pro Asn Glu Ser Ser Pro Thr Thr Thr Asn 1775	1780	1785
Thr Asn Thr Asn Asn Asn Ser Ile Ser Phe Thr Tyr Lys Phe Glu 1790	1795	1800
Ile Leu Asp Ser Val Leu Phe Thr Pro Val Ala Asn Pro Asp His 1805	1810	1815
Glu Phe Ile Lys Leu Asn Ile Pro Gln Met Tyr Cys Thr Phe Ile 1820	1825	1830
Asp Ser Ile Asp Ser Asp Thr Leu Leu Lys Glu Ile Pro Leu Glu 1835	1840	1845
Cys Ser Val Pro Val Gly Phe Thr Gly Asn Gln Asn Tyr Cys Leu 1850	1855	1860
Asn Val Phe Gly Arg Asp Leu Ser Leu His His Ile Ile Cys Arg 1865	1870	1875
Lys Asp Asn Ala Ser Glu Val Thr Ser Val Ser Leu Ile Ala Pro 1880	1885	1890
Phe Ser Gly Asp Ile Trp Ile Thr Ile Pro Tyr Glu Ser Asn Ser 1895	1900	1905
Ser Tyr Ala Thr Cys Ile Met Ser Arg Val Ser Lys Cys Gln Phe		

1910	1915	1920
Thr Val Glu Gly Arg Glu	Ile Leu Gly Cys Ile	Gly Ala Leu Gln
1925	1930	1935
Asp Val Val Asp Gln Phe	Ser Ser Val Gly Asn	Leu Ser Thr Cys
1940	1945	1950
Phe Thr Ser Asp Val Ser	Glu Phe Leu Asn Leu	Lys Glu Asn Tyr
1955	1960	1965
Val Val Pro Val Pro Ile	Glu Ser Ser Thr Val	Ser Phe Thr Glu
1970	1975	1980
Ile Arg Cys Ser Val Gln	Ser Met Ser Val Glu	Leu Tyr Ser Asp
1985	1990	1995
Lys Met Asn Gly Asn Arg	Leu Ile Ala Lys Ser	Asp Met Lys Phe
2000	2005	2010
Ala Cys Gly Ile Ser Met	Lys Thr Asp Lys Pro	Leu Ser Leu Asp
2015	2020	2025
Ile Ser Phe Thr Cys Phe	Thr Leu Ser Ser Leu	Leu Thr Ser Val
2030	2035	2040
Val Leu Leu Glu Cys Thr	Ser Cys Thr Lys Asn	Val Pro Val Leu
2045	2050	2055
Asn Met Gln Phe Leu Met	Ser Asp Asp Gly Lys	Asn His Leu Arg
2060	2065	2070
Phe Ser Leu Pro Cys Val	Asn Ile Trp Leu Phe	Leu Ser Glu Trp
2075	2080	2085
Ser Gln Val Val Asp Leu	Val Asn Ser Cys Cys	Glu Pro Ala Ile
2090	2095	2100
Gln Asn Glu Glu Pro Glu	Lys Ser Thr Ser Ala	Pro Val Ser Arg
2105	2110	2115
Val Asp Thr Ala Glu Asn	Ser Pro Gln Ser Thr	Thr Val Ser Ser
2120	2125	2130
Tyr Pro Ser Leu Glu Asp	Arg Phe Ser Leu Thr	Val Lys Ser Asp
2135	2140	2145
Leu Ile Gly Val Lys Ile	Arg Ile Pro Val Gln	Val Ser Gly Glu
2150	2155	2160
Val Val Lys Tyr Phe Gly	Ala Pro Gln Val Arg	Glu Gln Ser Leu
2165	2170	2175
Val Thr Gly Arg Asp His	Gly Ser Phe Leu Phe	Ile Tyr Leu Gln
2180	2185	2190
Ser Arg Cys Thr Glu Val	Asn Met Lys Gly Glu	Thr Val Asn Leu
2195	2200	2205

Lys Ser Asn Leu Gly Lys Ala Met Gly Thr Val Glu Leu Phe Gln 2210	2215	2220
Asn Lys Ser Val His Ser Trp Pro Leu Phe Gln Leu Leu Glu Ile 2225	2230	2235
Asp Ile Glu Ala Glu Ser Gly Asn Asp Asp Met Asp Arg Met His 2240	2245	2250
Leu Lys Thr Glu Ile His Cys Asp Asn Leu Asp Val Trp Leu Ser 2255	2260	2265
His His Ala Phe Tyr Phe Trp Gln Thr Met Leu Phe Met Phe Pro 2270	2275	2280
Glu Gly Ser Gly Ser Glu Ser Pro Gln Leu Pro Val Gly Ser Val 2285	2290	2295
Asn Phe Arg Phe His Leu Arg Lys Leu Ser Ile Leu Leu Thr Asp 2300	2305	2310
Glu Lys Trp Ser Ser Asn Gly Pro Leu Leu Glu Ile Leu Met Gly 2315	2320	2325
Ser Leu Leu Phe His Gly Ile Ile Thr Ala Asn Met Met Glu Gly 2330	2335	2340
Ser Ile Asp Ser Asp Leu Gln Val Asn Tyr Asn Asn Ile His Lys 2345	2350	2355
Val Leu Trp Glu Pro Phe Leu Glu Pro Trp Lys Phe Gln Ile Thr 2360	2365	2370
Leu Arg Arg Gln Gln Gly Lys Ser Thr Leu Glu Asn Ser Pro Val 2375	2380	2385
Met Thr Asp Ile Arg Leu Glu Ser Ser Thr Asn Leu Asn Ile Asn 2390	2395	2400
Val Thr Glu Ser Phe Ile Glu Val Ala Phe Arg Thr Phe Asp Met 2405	2410	2415
Ile Lys Asp Ala Ser Asp Leu Ile Ser Leu Asn Ile Leu Pro Glu 2420	2425	2430
Asn Asn Ser Arg Leu Thr Lys Pro His Thr Asn Glu Asn Thr Leu 2435	2440	2445
Ala Asn Arg Tyr Ala Pro Tyr Thr Leu Glu Asn Leu Thr Ser Leu 2450	2455	2460
Pro Leu Val Phe Tyr Ile Ser Lys Gly Glu Gly Phe Asn Met Thr 2465	2470	2475
Ser Leu Lys Asp Gly Lys His Val Gln Pro Gly Cys Ser Tyr Pro 2480	2485	2490
Val Asn Ile Asp Asn Asn Pro Glu Glu Gln Thr Phe Gly Phe Arg		

2495	2500	2505
Pro Ser His Ser Thr Asp	Asn Leu Gly Gly Asp	Met Gln Phe Ala
2510	2515	2520
Asp Ala Gln His His Tyr	Ile Val Val Gln Leu	Glu Gly Thr Ser
2525	2530	2535
Thr Leu Ser Ala Pro Val	Ser Ile Asp Leu Val	Gly Val Ser Phe
2540	2545	2550
Phe Glu Val Asp Phe Ser	Thr Asn Ile Gly Leu	Asp Val Ser Lys
2555	2560	2565
Gly Gly Tyr Val Val Pro	Val Val Ile Asp Val	Ser Val Gln Arg
2570	2575	2580
Tyr Thr Lys Leu Val Arg	Leu Tyr Ser Thr Val	Ile Leu Thr Asn
2585	2590	2595
Ala Thr Ser Met Pro Phe	Glu Val Arg Phe Asp	Ile Pro Phe Gly
2600	2605	2610
Val Ser Pro Lys Ile Leu	Asp Pro Val Tyr Pro	Gly His Glu Phe
2615	2620	2625
Pro Leu Pro Leu His Leu	Ala Glu Ser Gly Arg	Ile Arg Trp Arg
2630	2635	2640
Pro Leu Gly Ser Thr Tyr	Leu Trp Ser Glu Ala	Tyr Ser Ile Ser
2645	2650	2655
Asn Ile Leu Ser Asn Glu	Ser Lys Ile Gly His	Leu Arg Ser Phe
2660	2665	2670
Val Cys Tyr Pro Ser Leu	Pro Ser Ser Asp Pro	Phe Arg Cys Cys
2675	2680	2685
Val Ser Val His Asp Val	Cys Leu Pro Ser Ala	Gly Arg Val Ile
2690	2695	2700
Asn Lys Arg Gly Ser Ser	Ser Ser Leu Tyr Asn	Ile Asn Thr His
2705	2710	2715
Asp His Gly Asp Lys Ile	Glu Asn Gln Asp Leu	Ser Asn Lys Arg
2720	2725	2730
Cys Ile His Leu Ile Thr	Leu Ser Asn Pro Leu	Ile Val Lys Asn
2735	2740	2745
Tyr Leu Pro Val Glu Val	Ser Val Val Ile Glu	Ser Gly Gly Val
2750	2755	2760
Ser Arg Ser Met Leu Leu	Ser Glu Val Glu Thr	Phe Phe Tyr His
2765	2770	2775
Ile Asp Ser Ser His Asp	Leu Ser Leu Thr Phe	Glu Ile His Gly
2780	2785	2790

Phe Arg Pro Ser Leu Leu Lys Phe Pro Arg Ala Glu Lys Phe Ser	2795	2800	2805
Gly Ile Ala Lys Phe Ser Gly Thr Lys Phe Ser Ser Ser Glu Ser	2810	2815	2820
Ile Asn Phe Ala Ala Asp Asn Ser Lys Gly Pro Leu Tyr Val Thr	2825	2830	2835
Met Glu Lys Val Met Asp Ala Phe Ser Gly Ala Arg Glu Ile Cys	2840	2845	2850
Ile Phe Val Pro Phe Leu Leu Tyr Asn Cys Cys Gly Phe Pro Leu	2855	2860	2865
Thr Ile Ala Asn Ser Thr Asn Asp Leu Thr Met Arg Asp Thr Leu	2870	2875	2880
Pro Ser Cys Tyr Asp Leu Asp Glu Glu Asp Pro Phe Leu Gly Lys	2885	2890	2895
Lys Asp Gly Leu Ser Leu Leu Phe Ser Asn Gln Val Ser Asn Asn	2900	2905	2910
Asp Pro Met Ser Asn Leu Val Ser Thr Arg Lys Glu Ser Phe Thr	2915	2920	2925
Ser Ser Gly Ser Thr Lys Lys Asn Ile Gly Thr Gln Lys Pro Ser	2930	2935	2940
Leu His Asp Gln Glu Lys Ser Gln Leu Ser His Ser Gln Gln Leu	2945	2950	2955
Asp Phe Asp Glu Thr Thr Arg Lys Lys Val Asn Phe Arg Met Tyr	2960	2965	2970
Ser Pro Asp Pro Asn Ile Ala Ser Ser Glu Ile Met Val Arg Val	2975	2980	2985
Ser Arg Cys Arg Ser Asp Ala Asp Met Ala Ser Thr Ser Asp Tyr	2990	2995	3000
Thr Trp Ser Ser Gln Phe Phe Leu Val Pro Pro Ala Gly Ser Thr	3005	3010	3015
Thr Val Leu Val Pro Arg Ser Ser Thr Asn Ala Ser Tyr Val Leu	3020	3025	3030
Ser Val Ala Ser Ser Ala Ile Ser Gly Pro Tyr Ser Gly Arg Thr	3035	3040	3045
Arg Ile Ile Asn Phe Gln Pro Arg Tyr Val Ile Ser Asn Ala Cys	3050	3055	3060
Ser Gln Asp Leu Cys Tyr Arg Gln Lys Gly Ser Asp Phe Ile Tyr	3065	3070	3075
His Leu Lys Ala Gly Gln His Ser His Ile His Trp Thr Asp Ile			

3080	3085	3090
Thr Arg Glu Leu Leu Val	Ser Val Arg Phe Asp	Glu Pro Gly Trp
3095	3100	3105
Gln Trp Ser Gly Cys Phe	Phe Pro Glu His Leu	Gly Asp Thr Gln
3110	3115	3120
Leu Lys Met Arg Asn Tyr	Val Ser Gly Ala Val	Ser Met Val Arg
3125	3130	3135
Val Glu Val Gln Asn Ala	Asp Asp Ala Ile Arg	Asp Asp Lys Ile
3140	3145	3150
Val Gly Asn Pro His Gly	Glu Ser Gly Thr Asn	Leu Ile Leu Leu
3155	3160	3165
Ser Asp Asp Asp Thr Gly	Phe Met Pro Tyr Arg	Val Asp Asn Phe
3170	3175	3180
Ser Lys Glu Arg Leu Arg	Ile Tyr Gln Gln Lys	Cys Glu Ala Phe
3185	3190	3195
Glu Thr Val Ile His Ser	Tyr Thr Ser Cys Pro	Tyr Ala Trp Asp
3200	3205	3210
Glu Pro Ser Tyr Pro His	Arg Leu Thr Val Glu	Val Phe Ala Glu
3215	3220	3225
Arg Val Val Gly Ser Tyr	Thr Leu Asp Asp Ala	Lys Glu Tyr Lys
3230	3235	3240
Pro Val Val Leu Pro Ser	Thr Ser Glu Lys Pro	Glu Arg Arg Leu
3245	3250	3255
Leu Ile Ser Val His Ala	Glu Gly Ala Leu Lys	Val Leu Ser Ile
3260	3265	3270
Ile Asp Ser Ser Tyr His	Ile Phe Asp Asp Val	Lys Ile Pro Arg
3275	3280	3285
Ser Pro Arg Leu Thr Glu	Lys Arg Glu Tyr Asp	Gln Lys Gln Glu
3290	3295	3300
Ser Ser Leu Leu Tyr Gln	Glu Arg Val Ser Ile	Ser Ile Pro Phe
3305	3310	3315
Ile Gly Ile Ser Val Met	Ser Ser Gln Pro Gln	Glu Leu Leu Phe
3320	3325	3330
Ala Cys Ala Arg Asn Thr	Met Ile Asp Leu Val	Gln Ser Leu Asp
3335	3340	3345
Gln Gln Asn Leu Ser Leu	Lys Ile Ser Ser Leu	Gln Ile Asp Asn
3350	3355	3360
Gln Leu Pro Thr Thr Pro	Tyr Pro Val Ile Leu	Ser Phe Asp His
3365	3370	3375

Glu Tyr Lys Gln Leu Pro Thr Ser Gln Ile Lys Asn Lys Asp Glu 3380	3385	3390
Pro Val Phe Ser Leu Ala Ala Ala Lys Trp Arg Asn Lys Asp Arg 3395	3400	3405
Ala Leu Leu Ser Phe Glu His Ile Asn Leu Arg Met Ala Asp Phe 3410	3415	3420
His Leu Glu Leu Glu Gln Asp Val Ile Leu Ser Leu Phe Asp Phe 3425	3430	3435
Ser Lys Ala Val Ser Ser Arg Phe His Ser Arg Gly Met Pro His 3440	3445	3450
Met Asp Ser Val Val His Pro Leu Ser Ser Asn Leu Ser Gly Asn 3455	3460	3465
Lys Ala Thr Lys Leu Ala Glu Lys Thr Glu Ile Glu Gly Glu Ser 3470	3475	3480
Phe Pro Leu Leu Pro Ser Ile Val Pro Ile Gly Ala Pro Trp Gln 3485	3490	3495
Lys Ile Tyr Leu Leu Ala Arg Lys Gln Lys Lys Ile Tyr Val Glu 3500	3505	3510
Leu Leu Glu Val Ala Pro Ile Thr Leu Thr Leu Ser Phe Ser Ser 3515	3520	3525
Ser Pro Trp Met Leu Arg Asn Gly Ile Leu Thr Ser Gly Glu Tyr 3530	3535	3540
Leu Ile His Arg Gly Leu Met Ala Leu Ala Asp Val Glu Gly Ala 3545	3550	3555
Arg Ile His Leu Arg Arg Leu Thr Ile Ser His Gln Leu Ala Ser 3560	3565	3570
Leu Glu Ser Ile Arg Glu Ile Leu Ile Ile His Tyr Thr Arg Gln 3575	3580	3585
Leu Leu His Glu Met Tyr Lys Val Phe Gly Ser Ala Gly Val Ile 3590	3595	3600
Gly Asn Pro Met Gly Phe Ala Arg Ser Val Gly Leu Gly Ile Arg 3605	3610	3615
Asp Phe Leu Ser Val Pro Ala Arg Ser Phe Met Gln Ser Pro Ala 3620	3625	3630
Gly Leu Ile Thr Gly Met Ala Gln Gly Thr Thr Ser Leu Leu Ser 3635	3640	3645
Asn Thr Val Tyr Ala Ile Ser Asp Ala Ala Thr Gln Gly Ile Val 3650	3655	3660
Ala Phe Thr Met Asp Asp Pro Pro Ser Ala Ala Glu Met Gly Lys		

3665	3670	3675
Gly Val Ile Asn Glu Val	Leu Glu Gly Leu Thr	Gly Leu Leu Gln
3680	3685	3690
Ser Pro Ile Arg Gly Ala	Glu Lys His Gly Leu	Pro Gly Val Leu
3695	3700	3705
Ser Gly Ile Ala Leu Gly	Val Thr Gly Leu Val	Ala Arg Pro Ala
3710	3715	3720
Ala Ser Ile Leu Glu Val	Thr Glu Lys Thr Ala	Arg Ser Ile Arg
3725	3730	3735
Asn Arg Ser Lys Leu Tyr	His Ile Arg Leu Arg	Val Arg Leu Pro
3740	3745	3750
Arg Pro Leu Thr Pro Asn	His His Pro Leu Lys	Pro Tyr Ser Trp
3755	3760	3765
Asp Gln Ala Val Gly Leu	Ser Val Leu Thr Asn	Thr Asn Ser Asn
3770	3775	3780
Ser Asp Ser Asp Leu Lys	Asp Glu Thr Leu Val	Leu Ser Lys Ser
3785	3790	3795
Leu Lys Gln Lys Gly Lys	Phe Val Ile Ile Thr	Gln Arg Leu Leu
3800	3805	3810
Leu Ile Val Thr Ser Ser	Ser Leu Thr Asn Leu	Gly Gln Pro Asn
3815	3820	3825
Phe Lys Gly Val Pro Ala	Asp Pro Asp Trp Val	Val Glu Ala Glu
3830	3835	3840
Ile Thr Leu Asp Ser Val	Ile His Val Asp Val	Asp Gly Glu Val
3845	3850	3855
Val His Ile Val Gly Ser	Ser Ser Asp Val Val	Val Arg Gln Asn
3860	3865	3870
Val Gly Gly Lys Gln Arg	Trp Tyr Asn Pro Leu	Pro Leu Phe Gln
3875	3880	3885
Thr Asn Leu Glu Cys Leu	Gly Lys Glu Glu Ala	Gly Glu Leu Leu
3890	3895	3900
Lys Val Leu Leu Val Thr	Ile Glu Arg Gly Lys	Glu Arg Gly Trp
3905	3910	3915
Gly Arg Gly Cys Val Tyr	Arg Leu His Gln Ser	Asn Val Arg
3920	3925	3930

<210> 4

<211> 909

<212> DNA

<213> Taraxacum officinale

<400> 4

gaaaccgaag caaactctac cacatccgcc tccgggtccg cctcccgaga ccgctaacce 60
 ccaaccacca cccgttaaaa ccctactcgt gggaccaagc agtcggcctc tccgtcctca 120
 ccaataccaa ttccaattcc gattccgacc tcaaagacga aaccctcgtc ctctccaaat 180
 ccctcaaaca aaagggcaaa ttcgtcatta tcaccaacg gttactcctc attgttacct 240
 cctcgagcct aacgaattta ggtcaacca atttcaaagg cgtccctgcg gaccccgatt 300
 ggggtggttga agccgagata acgttgata gtgtgataca cgtggatggt gatggagagg 360
 tgggtcatat tgtcgggagt agttctgatg tgggtggttag acagaatggt ggtgggaagc 420
 agcggtggtta taatccgttg ccgctgtttc agacgaattt ggagtgttta gggaaggagg 480
 aggcgggggga gttgttgaag gtgttggttg tgacgattga gagaggggaag gagagagggt 540
 ggggccgggg gtgtgtgtac cgtctgcatc agagtaatgt taggtgatgt atattttttt 600
 tctacatata aagtactat aggagaaaaa ggactggata ttatattata catacctgaa 660
 acaaggaaac gttttcttc aaaattttgg ctgtattatt attttgcga ccatgttggg 720
 ctaaaatggc caattattta cttatgacat ggttaaaaaa tattggtgtc ttgttttgta 780
 taattacaat ttatattagt atcgatgcaa tgtaagattg tagaaagcgc taccgtataa 840
 aacaacataa gtcatgaggt tacaccctag tgggggtcaag ggaacaaaaa cattttaaac 900
 gtttttcag 909

<210> 5

<211> 909

<212> RNA

<213> Taraxacum officinale

<400> 5

gaaaccgaag caaacucuac cacauccgcc uccggguccg ccucccgaga ccgcuaacce 60
 ccaaccacca cccguuaaaa ccuacucgu gggaccaagc agucggccuc uccguccuca 120
 ccaauaccaa uucaauucc gauuccgacc ucaaagacga aaccucguc cucuccaaau 180
 ccucuaaaca aaagggcaaa uucgucauuu ucaccaacg guuacuccuc auuguuaccu 240
 ccucgagccu aacgaauuuu ggucaacca auuucaaagg cgucccugcg gaccccgauu 300
 gggugguuga agccgagaua acguuggaua gugugauaca cguggauguu gauggagagg 360
 uggugcauau ugucgggagu aguucugaug uggugguuag acagaauuuu ggugggaagc 420
 agcgguggua uaauccguug ccgcuguuuc agacgaauuu ggaguguuuu gggaaggagg 480
 aggcgggggga guuguugaag guguuguugg ugacgauuga gagaggggaag gagagagggu 540
 ggggccgggg guguguguac cgucugcauc agaguaaugu uaggugaugu auauuuuuuu 600
 ucuacauaua aaguuaauu aggagaaaaa ggacuggaua uuauuuuaua cauaccugaa 660
 acaaggaaac guuuucuuuc aaaauuuugg cuguauuuuu auuuugucga ccauguuggg 720
 cuaaaauggc cauuuuuuu cuuaugacau gguuaaaaaa uauugguguc uuguuuugua 780
 uauuuacaau uuauuuuagu aucgaugcaa uguaagauug uagaaagcgc uaccguauaa 840
 aacaacauaa gucaugaggu uacaccuag uggggucaag ggaacaaaaa cauuuuaaac 900
 guuuuucag 909

<210> 6

<211> 582

<212> DNA

<213> Taraxacum officinale

<400> 6

```
aaccgaagca aactctacca catccgctc cgggtccgcc tcccgagacc gctaaccccc 60
aaccaccacc cgtaaacc ctactcgtgg gaccaagcag tcggcctctc cgtectcacc 120
aataccaatt ccaattccga ttccgacctc aaagacgaaa ccctcgtcct ctccaaatcc 180
ctcaaacaaa agggcaaatt cgtcattatc acccaacggg tactcctcat tgttacctcc 240
tcgagcctaa cgaatttagg tcaacceaat tcaaaaggcg tcctcgcgga ccccgattgg 300
gtggttgaag ccgagataac gttggatagt gtgatacacg tggatgttga tggagagggt 360
gtgcatattg tcgggagtag ttctgatgtg gtggttagac agaatgttgg tgggaagcag 420
cggtggtata atccgttccc gctgtttcag acgaatttgg agtgtttagg gaaggaggag 480
gcgggggagt tgttgaaggt gttgttggtg acgattgaga gagggaagga gagagggtgg 540
ggccgggggt gtgtgtaccg tctgcatcag agtaatgtta gg 582
```

<210> 7

<211> 194

<212> PRT

<213> Taraxacum officinale

<400> 7

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

Ala Gly Glu Leu Leu Lys Val Leu Leu Val Thr Ile Gln Arg Gly Lys
 165 170 175

Glu Arg Gly Trp Gly Arg Gly Cys Val Tyr Arg Leu His Gln Ser Asn
 180 185 190

Val Arg

<210> 8

<211> 51

<212> DNA

<213> Taraxacum officinale

<400> 8

gaattcctct atgttttggga accttcttgt tgttattecg caccacttta a 51

<210> 9

<211> 479

<212> DNA

<213> Taraxacum officinale

<400> 9

aattccaatt ccgattccga cctcaaagac gaaaccctcg tectctccaa atccctcaaa 60
 caaaagggca aattcgatc taccaccaa cggttactcc tcattgttac ctctcgagc 120
 ctaacgaatt taggtcaacc cgatttcaaa ggcgtccctg cggaccccgga ttgggtggtt 180
 gaagccgaga taacgttga tagtgtgata cacgtggatg ttgatggaga ggtggtgcat 240
 attgtcggga gtagttctga tgtggtggtt agacagaatg ttggtggtgg tgggtggtgg 300
 gggaaagcagc ggtggtataa tccgccgacg ccgttgccgc tgtttcagac gaatttgag 360
 tgtttagggga aggaggaggc gggggagtgt ttgaaggtgt tgttggtgac gattcagaga 420
 gggaaaggaga gaggggtgggg ccgggggtgt gtgtaccgtc tgcacagag taatgttag 479

<210> 10

<211> 159

<212> PRT

<213> Taraxacum officinale

<400> 10

Asn Ser Asn Ser Asp Ser Asp Leu Lys Asp Glu Thr Leu Val Leu Ser
 1 5 10 15

Lys Ser Leu Lys Gln Lys Gly Lys Phe Val Ile Ile Thr Gln Arg Leu
 20 25 30

Leu Leu Ile Val Thr Ser Ser Ser Leu Thr Asn Leu Gly Gln Pro Asp
 35 40 45

Phe Lys Gly Val Pro Ala Asp Pro Asp Trp Val Val Glu Ala Glu Ile
 50 55 60

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

Ile Val Gly Ser Ser Ser Asp Val Val Val Arg Gln Asn Val Gly Gly
85 90 95
Gly Gly Gly Trp Gly Lys Gln Arg Trp Tyr Asn Pro Pro Thr Pro Leu
100 105 110
Pro Leu Phe Gln Thr Asn Leu Glu Cys Leu Gly Lys Glu Glu Ala Gly
115 120 125
Glu Leu Leu Lys Val Leu Leu Val Thr Ile Gln Arg Gly Lys Glu Arg
130 135 140
Gly Trp Gly Arg Gly Cys Val Tyr Arg Leu His Gln Ser Asn Val
145 150 155

<210> 11

<211> 455

<212> DNA

<213> Taraxacum officinale

<400> 11

aattccaatt ccgattccga cctcaaagac gaaaccctcg tcctctccaa atccctcaaa 60
caaaagggca aattcgteat taccaccaa cggttactcc tcattgttac ctctcgagc 120
ctaacgaatt taggtcaacc caattcaaa ggcgtccctg cggaccccga ttgggtggtt 180
gaagccgaga taacgttga tagtgtgata cacgtggatg ttgatggaga ggtggtgcat 240
attgtcggga gtagttctga tgtggtggtt agacagaatg ttggtgggaa gcagcgggtg 300
tataatccgt tgccgtgtt tcagacgaat ttggagtgtt tagggaagga ggaggcgggg 360
gagttgttga aggtgttgtt ggtgacgatt gagagagga aggagagagg gtggggccgg 420
gggtgtgtgt accgtctgca tcagagtaat gttag 455

<210> 12

<211> 151

<212> PRT

<213> Taraxacum officinale

<400> 12

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

	85		90		95										
Lys	Gln	Arg	Trp	Tyr	Asn	Pro	Leu	Pro	Leu	Phe	Gln	Thr	Asn	Leu	Glu
	100		105		110										
Cys	Leu	Gly	Lys	Glu	Glu	Ala	Gly	Glu	Leu	Leu	Lys	Val	Leu	Leu	Val
	115		120		125										
Thr	Ile	Glu	Arg	Gly	Lys	Glu	Arg	Gly	Trp	Gly	Arg	Gly	Cys	Val	Tyr
	130		135		140										
Arg	Leu	His	Gln	Ser	Asn	Val									
	145		150												

<210> 13

<211> 938

<212> DNA

<213> Taraxacum officinale

<400> 13

```

gaaaccgaag caaactctac cacatccgcc tccgggtccg cctcccaaga ccgctaacce 60
ccaaccaccc gttaaaaccc tactcgtggg accaagcagt cggectctec gtctcacca 120
attccaattc cgattccgac ctcaaagacg aaaccctcgt cctctccaaa tcctcaaac 180
aaaagggcaa attcgtcatt atcacccaac ggttactcct cattgttacc tctcagacc 240
taacgaattt aggtcaacce gatttcaaag gcgtccctgc ggaccccgat tgggtggttg 300
aagccgagat aacgttggat agtgtgatac acgtggatgt tgatggagag gtggtgcata 360
ttgtcgggag tagttctgat gtgggtggtta gacagaatgt tgggtggtgt ggtgggtggg 420
ggaagcagcg gtggtataat ccgccgacgc cgttgccgct gtttcagacg aatttggagt 480
gtttagggaa ggaggaggcg ggggagtgtg tgaagggtgt gttggtgacg attcagagag 540
ggaaggagag aggggtggggc cgggggtgtg tgtaccgtct gcatcagagt aatgtaggt 600
gatgtatatt tttttgtac atataaagtt tactatagga gaaaaggac tggatattat 660
attatacata cctgaaaaca aggaaacgtt ttctttcaa attttggctg tattattatt 720
ttgtcgacca tgttgggcta aaatggccaa ttatttactt atgacatggt taaaaaatat 780
tggtgtcttg ttttgataa ttacaattta tattccttt tagactataa cttatgcaat 840
gtaagattgt gtataaaaca acataagtca tgaggttaca ccctagtggg atcaaggggg 900
ctacaccccg gaacaaaaac attttaaagc tttttcag 938

```

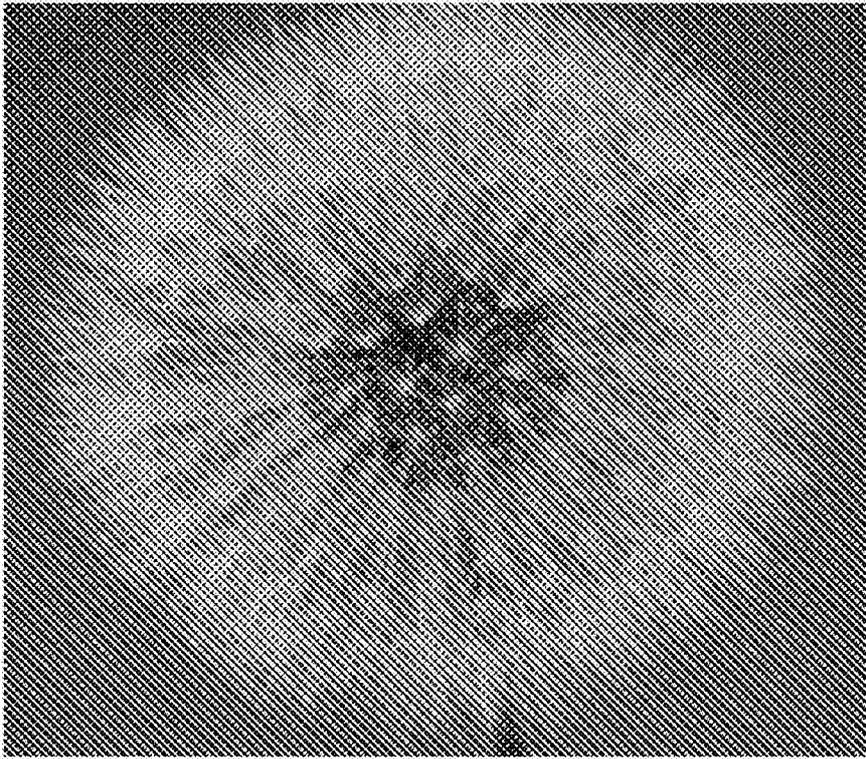


图1A

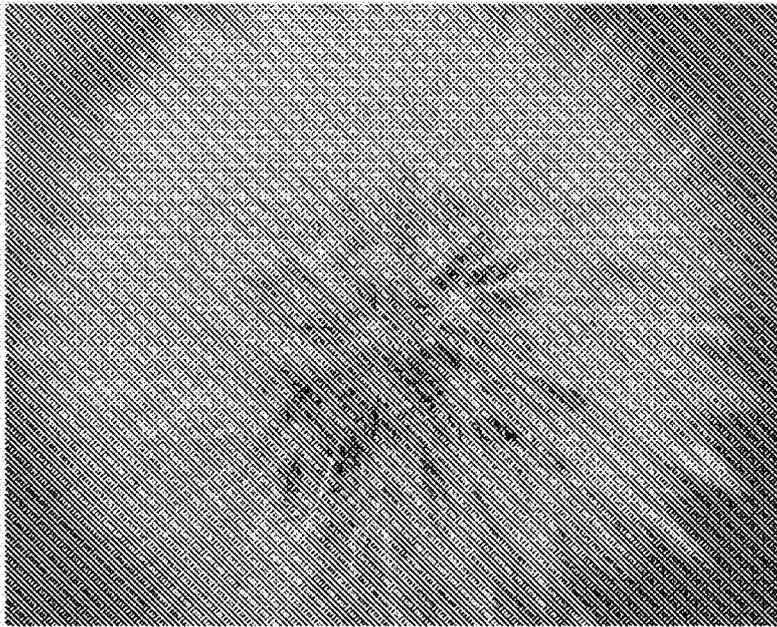


图1B

