



(12) 发明专利申请

(10) 申请公布号 CN 112204147 A

(43) 申请公布日 2021. 01. 08

(21) 申请号 201880090026.6

(22) 申请日 2018.12.21

(30) 优先权数据

62/609,508 2017.12.22 US

62/758,068 2018.11.09 US

(85) PCT国际申请进入国家阶段日
2020.08.21

(86) PCT国际申请的申请数据

PCT/EP2018/086725 2018.12.21

(87) PCT国际申请的公布数据

W02019/122394 EN 2019.06.27

(71) 申请人 科沃施种子欧洲股份两合公司

地址 德国艾恩贝克

(72) 发明人 M·拉布斯 A·赫梅尔 梅钰

(74) 专利代理机构 永新专利商标代理有限公司
72002

代理人 左路 区斌

(51) Int.Cl.

C12N 15/82 (2006.01)

C12N 15/63 (2006.01)

A01H 4/00 (2006.01)

A01H 1/08 (2006.01)

权利要求书7页 说明书70页
序列表200页 附图14页

(54) 发明名称

基于Cpf1的植物转录调控系统

(57) 摘要

本发明涉及基因表达的靶向调控,更具体地涉及合成转录因子(STF),其包含至少一个基于CRISPR/Cpf1系统的高靶向特异性工程化识别结构域,还包含至少一个激活或沉默结构域以调控目的基因的表达,优选地调控真核生物特别是植物的形态发生基因的转录。本发明还公开了使用STF以提高转化频率,优化成功的基因组编辑方法,提供单倍体或双单倍体生物,以及/或者提供适于一般转化但也适于育种目的的组合物的方法。

1. 一种合成转录因子或编码其的核苷酸序列,所述合成转录因子包含至少一个识别结构域和至少一个激活结构域,其中所述合成转录因子配置为调控细胞系统中形态发生基因的表达。

2. 一种合成转录因子或编码其的核苷酸序列,所述合成转录因子包含至少一个识别结构域和至少一个激活结构域,其中所述合成转录因子配置为激活细胞系统中内源基因的表达。

3. 根据权利要求1或2所述的合成转录因子,其中所述至少一个识别结构域是至少一个解除武装的CRISPR/核酸酶系统,或是至少一个解除武装的CRISPR/核酸酶系统的片段。

4. 根据权利要求3所述的合成转录因子,其中所述至少一个解除武装的CRISPR/核酸酶系统是CRISPR/dCpf1系统,其中所述至少一个解除武装的CRISPR/核酸酶系统包含至少一个向导RNA。

5. 根据权利要求1或2所述的合成转录因子,其中所述至少一个激活结构域选自酸性转录激活结构域,优选地,其中所述至少一个激活结构域来自水稻白叶枯病菌(*Xanthomonas oryzae*)的无毒基因、单纯疱疹病毒的VP16或四聚体VP64、VPR、SAM、Scaffold、Suntag、P300、VP160或其任何组合。

6. 根据权利要求1或2所述的合成转录因子,其中所述至少一个激活结构域相对于所述至少一个识别结构域位于N-末端和/或C-末端。

7. 根据权利要求1所述的合成转录因子,其中所述形态发生基因选自BBM、WUS(包括WUS2)、WOX基因、WUS或BBM同源物、Lec1、Lec2、WIND1、ESR1、PLT3、PLT5、PLT7、IPT、IPT2、Knotted1和RKD4。

8. 根据权利要求1所述的合成转录因子,其中所述合成转录因子配置为通过结合位于相对于起始密码子一定距离处的调控区域而调控所述形态发生基因的表达,优选地调控所述形态发生基因的转录。

9. 根据权利要求2所述的合成转录因子,其中,所述内源基因选自:编码对非生物胁迫的抗性或耐受性的基因,所述非生物胁迫包括干旱胁迫、渗透胁迫、热胁迫、冷胁迫、氧化胁迫、重金属胁迫、氮缺乏、磷缺乏、盐胁迫或渍水;编码除草剂抗性的基因,所述除草剂抗性包括草甘膦抗性、草铵膦/磷丝菌素抗性、潮霉素抗性、2,4-D抗性或耐受性、原卟啉原氧化酶(PPO)抑制剂抗性或耐受性、ALS抑制剂抗性或耐受性和麦草畏抗性或耐受性;编码对生物胁迫的抗性或耐受性的基因,包括病毒抗性基因、真菌抗性基因、细菌抗性基因、昆虫抗性基因;或编码产量相关性状的基因,所述产量相关性状包括抗倒伏性、开花时间、抗落粒性、种子颜色、胚乳成分或营养含量。

10. 根据权利要求1或2所述的合成转录因子,其中所述合成转录因子和/或所述至少一个识别结构域包含SEQ ID NO:276、277、282、283、284、288、289、290中任一个所示的序列,或包含与SEQ ID NO:276、277、282、283、284、288、289、290中的任一个在全长上具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性的序列。

11. 根据权利要求1或2所述的合成转录因子,其中所述细胞系统选自至少一个真核细胞或真核生物体,优选地,其中所述至少一个真核细胞是至少一个植物细胞,和/或其中所述至少一个真核生物体是植物或植物的一部分。

12. 如权利要求11所述的合成转录因子,其中所述植物的至少一部分选自叶、茎、根、初生胚根、花、花部分、花瓣、果实、花粉、花粉管、花药花丝、胚珠、胚囊、卵细胞、子房、合子、胚、合子胚、体细胞胚、顶端分生组织、维管束、中柱鞘、种子、根和插枝。

13. 根据权利要求12所述的合成转录因子,其中所述至少一个植物细胞、所述至少一个植物或所述植物的至少一部分源自选自以下的植物物种:大麦(*Hordeum vulgare*)、球茎大麦(*Hordeum bulbosum*)、高粱(*Sorghum bicolor*)、甘蔗(*Saccharum officinarum*)、玉米(*Zea mays*)、小米(*Setaria italica*)、小粒野生稻(*Oryza minuta*)、水稻(*Oryza sativa*)、澳洲野生稻(*Oryza australiensis*)、高秆野生稻(*Oryza alta*)、小麦(*Triticum aestivum*)、黑麦(*Secale cereale*)、苹果(*Malus domestica*)、二穗短柄草(*Brachypodium distachyon*)、滨海大麦草(*Hordeum marinum*)、节节麦(*Aegilops tauschii*)、胡萝卜(*Daucus glochidiatus*)、甜菜(*Beta vulgaris*)、小胡萝卜(*Daucus pusillus*)、*Daucus muricatus*、野胡萝卜(*Daucus carota*)、巨桉(*Eucalyptus grandis*)、美花烟草(*Nicotiana sylvestris*)、绒毛状烟草(*Nicotiana tomentosiformis*)、普通烟草(*Nicotiana tabacum*)、番茄(*Solanum lycopersicum*)、马铃薯(*Solanum tuberosum*)、中果咖啡(*Coffea canephora*)、葡萄(*Vitis vinifera*)、*Erythraea guttata*、*Genlisea aurea*、黄瓜(*Cucumis sativus*)、川桑(*Morus notabilis*)、*Arabidopsis arenosa*、琴叶拟南芥(*Arabidopsis lyrata*)、拟南芥(*Arabidopsis thaliana*)、近缘种须弥芥(*Crucihimalaya himalaica*)、卵叶须弥芥(*Crucihimalaya wallichii*)、弯曲碎米荠(*Cardamine flexuosa*)、北美独行菜(*Lepidium virginicum*)、荠菜(*Capsella bursa pastoris*)、*Olm arabidopsis pumila*、硬毛南芥(*Arabis hirsute*)、甘蓝型油菜(*Brassica napus*)、甘蓝(*Brassica oleracea*)、芜菁(*Brassica rapa*)、萝卜(*Raphanus sativus*)、芥菜(*Brassica juncea*)、黑芥菜(*Brassica nigra*)、芝麻菜(*Eruca vesicaria* subsp. *sativa*)、甜橙(*Citrus sinensis*)、麻疯树(*Jatropha curcas*)、毛果杨(*Populus trichocarpa*)、蒺藜苜蓿(*Medicago truncatula*)、*Cicer yamashitae*、*Cicer bijugum*、鹰嘴豆(*Cicer arietinum*)、*Cicer reticulatum*、*Cicer judaicum*、木豆(*Cajanus cajanifolius*)、蔓草虫豆(*Cajanus scarabaeoides*)、菜豆(*Phaseolus vulgaris*)、大豆(*Glycine max*)、紫云英(*Astragalus sinicus*)、百脉根(*Lotus japonicus*)、夏堇(*Torenia fournieri*)、洋葱(*Allium cepa*)、大葱(*Allium fistulosum*)、大蒜(*Allium sativum*)和韭菜(*Allium tuberosum*)。

14. 一种用于提高细胞系统中转化效率的方法,其中所述方法包括以下步骤:

- (a) 提供细胞系统;
- (b) 在所述细胞系统中引入至少一个合成转录因子或编码其的核苷酸序列;以及
- (c) 在所述细胞系统中引入至少一个目的核苷酸序列;
- (d) 任选地,在获得所述细胞系统的转化的后代的条件下培养所述细胞系统;

其中所述至少一个合成转录因子或编码其的所述核苷酸序列包含至少一个识别结构域和至少一个激活结构域,其中所述合成转录因子配置为调控所述细胞系统中的至少一个形态发生基因的表达,优选地为调控所述细胞系统中的至少一个形态发生基因的转录;并且

其中所述至少一个合成转录因子或编码其的核苷酸序列的引入与所述至少一个目的

核苷酸序列的引入平行进行或序贯进行。

15. 根据权利要求14所述的方法, 其中

(a) 所述至少一个合成转录因子或编码其的所述序列, 或者所述至少一个合成转录因子的至少一个组分或编码其的序列; 以及

(b) 所述至少一个目的核苷酸序列

通过独立地选自生物方法和/或物理方法的方法引入所述细胞系统中, 所述方法包括转染、转化, 包括通过农杆菌属物种 (优选地为根癌农杆菌)、病毒载体、生物弹轰击进行转化, 使用化学试剂进行转染 (包括聚乙二醇转染) 或其任何组合。

16. 一种用于提高细胞系统中至少一个内源基因的表达的方法, 其中所述方法包括以下步骤:

(a) 提供细胞系统;

(b) 在所述细胞系统中引入至少一个合成转录因子或编码其的核苷酸序列;

其中所述至少一个合成转录因子或编码其的所述核苷酸序列包含至少一个识别结构域和至少一个激活结构域, 其中所述合成转录因子配置为调控所述细胞系统中至少一个内源基因的表达, 优选地为调控所述细胞系统中至少一个内源基因的转录。

17. 根据权利要求16所述的方法, 其中所述至少一个合成转录因子或编码其的所述序列, 或者所述至少一个合成转录因子的至少一个组分或编码其的序列通过独立地选自生物和/或物理方法的方法引入所述细胞系统中, 所述方法包括转染、转化, 包括通过农杆菌属物种 (优选地为根癌农杆菌)、病毒载体、生物弹轰击进行转化, 使用化学试剂进行转染 (包括聚乙二醇转染), 或其任何组合。

18. 根据权利要求14或16所述的方法, 其中所述至少一个识别结构域是至少一个解除武装的非功能性CRISPR/核酸酶系统, 或是至少一个解除武装的非功能性CRISPR/核酸酶系统的片段。

19. 根据权利要求18所述的方法, 其中所述至少一个解除武装的CRISPR/核酸酶系统是CRISPR/dCpf1系统, 其中所述至少一个解除武装的CRISPR/核酸酶系统包括至少一个向导RNA。

20. 根据权利要求14或16所述的方法, 其中所述至少一个合成转录因子的所述至少一个激活结构域选自酸性转录激活结构域, 优选地, 其中所述至少一个激活结构域来自水稻白叶枯病菌的无毒基因、单纯疱疹病毒的VP16或四聚体VP64、VPR、SAM、Scaffold、Suntag、P300、VP160或其任何组合。

21. 根据权利要求14或16所述的方法, 其中所述至少一个合成转录因子的所述至少一个激活结构域相对于所述至少一个合成转录因子的所述至少一个识别结构域位于N-末端和/或C-末端。

22. 根据权利要求14所述的方法, 其中所述至少一个形态发生基因选自BBM、WUS (包括WUS2)、WOX基因、WUS或BBM同源物、Lec1、Lec2、WIND1、ESR1、PLT3、PLT5、PLT7、IPT、IPT2、Knotted1和RKD4。

23. 根据权利要求14所述的方法, 其中所述合成转录因子配置为通过结合位于相对于所述起始密码子一定距离处的调控区域来调控所述形态发生基因的表达, 优选地为调控所述形态发生基因的转录。

24. 根据权利要求16所述的方法,其中所述内源基因选自:编码对非生物胁迫的抗性或耐受性的基因,所述非生物胁迫包括干旱胁迫、渗透胁迫、热胁迫、冷胁迫、氧化胁迫、重金属胁迫、氮缺乏、磷缺乏、盐胁迫或渍水;编码除草剂抗性的基因,所述除草剂抗性包括草甘膦抗性、草铵膦/磷丝菌素抗性、潮霉素抗性、2,4-D抗性或耐受性、原卟啉原氧化酶(PPO)抑制剂抗性或耐受性、ALS抑制剂抗性或耐受性和麦草畏抗性或耐受性;编码对生物胁迫的抗性或耐受性的基因,包括病毒抗性基因、真菌抗性基因、细菌抗性基因、昆虫抗性基因;或编码产量相关性状的基因,所述产量相关性状包括抗倒伏性、开花时间、抗落粒性、种子颜色、胚乳成分或营养含量。

25. 根据权利要求14或16所述的方法,其中所述合成转录因子和/或所述至少一个识别结构域包含SEQ ID NO:276、277、282、283、284、288、289、290中任一个所示的序列,或包含与SEQ ID NO:276、277、282、283、284、288、289、290中的任一个在全长上具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性的序列。

26. 根据权利要求14或16所述的方法,其中所述细胞系统选自至少一个真核细胞或真核生物体,优选地,其中所述至少一个真核细胞是至少一个植物细胞,和/或其中所述至少一个真核生物体是植物或植物的一部分。

27. 根据权利要求26所述的方法,其中所述植物的至少一部分选自叶、茎、根、初生胚根、花、花部分、花瓣、果实、花粉、花粉管、花药花丝、胚珠、胚囊、卵细胞、子房、合子、胚、合子胚、体细胞胚、顶端分生组织、维管束、中柱鞘、种子、根和插枝。

28. 根据权利要求27所述的方法,其中所述至少一个植物细胞、所述至少一个植物或所述植物的至少一部分源自选自以下的植物物种:大麦(*Hordeum vulgare*)、球茎大麦(*Hordeum bulbosum*)、高粱(*Sorghum bicolor*)、甘蔗(*Saccharum officinarum*)、玉米(*Zea mays*)、小米(*Setaria italica*)、小粒野生稻(*Oryza minuta*)、水稻(*Oryza sativa*)、澳洲野生稻(*Oryza australiensis*)、高秆野生稻(*Oryza alta*)、小麦(*Triticum aestivum*)、黑麦(*Secale cereale*)、苹果(*Malus domestica*)、二穗短柄草(*Brachypodium distachyon*)、滨海大麦草(*Hordeum marinum*)、节节麦(*Aegilops tauschii*)、胡萝卜(*Daucus glochidiatus*)、甜菜(*Beta vulgaris*)、小胡萝卜(*Daucus pusillus*)、*Daucus muricatus*、野胡萝卜(*Daucus carota*)、巨桉(*Eucalyptus grandis*)、美花烟草(*Nicotiana sylvestris*)、绒毛状烟草(*Nicotiana tomentosiformis*)、普通烟草(*Nicotiana tabacum*)、番茄(*Solanum lycopersicum*)、马铃薯(*Solanum tuberosum*)、中果咖啡(*Coffea canephora*)、葡萄(*Vitis vinifera*)、*Erythraea guttata*、*Genlisea aurea*、黄瓜(*Cucumis sativus*)、川桑(*Morus notabilis*)、*Arabidopsis arenosa*、琴叶拟南芥(*Arabidopsis lyrata*)、拟南芥(*Arabidopsis thaliana*)、近缘种须弥芥(*Crucihimalaya himalaica*)、卵叶须弥芥(*Crucihimalaya wallichii*)、弯曲碎米荠(*Cardamine flexuosa*)、北美独行菜(*Lepidium virginicum*)、荠菜(*Capsella bursa pastoris*)、*Olm arabidopsis pumila*、硬毛南芥(*Arabis hirsute*)、甘蓝型油菜(*Brassica napus*)、甘蓝(*Brassica oleracea*)、芜菁(*Brassica rapa*)、萝卜(*Raphanus sativus*)、芥菜(*Brassica juncea*)、黑芥菜(*Brassica nigra*)、芝麻菜(*Eruca vesicaria* subsp. *sativa*)、甜橙(*Citrus sinensis*)、麻疯树(*Jatropha curcas*)、毛果杨(*Populus*

trichocarpa)、蒺藜苜蓿 (*Medicago truncatula*)、*Cicer yamashitae*、*Cicer bijugum*、鹰嘴豆 (*Cicer arietinum*)、*Cicer reticulatum*、*Cicer judaicum*、木豆 (*Cajanus cajanifolius*)、蔓草虫豆 (*Cajanus scarabaeoides*)、菜豆 (*Phaseolus vulgaris*)、大豆 (*Glycine max*)、紫云英 (*Astragalus sinicus*)、百脉根 (*Lotus japonicas*)、夏堇 (*Torenia fournieri*)、洋葱 (*Allium cepa*)、大葱 (*Allium fistulosum*)、大蒜 (*Allium sativum*) 和韭菜 (*Allium tuberosum*)。

29. 一种在预定位置修饰细胞系统的遗传物质的方法,其中所述方法包括以下步骤:

(a) 提供细胞系统;

(b) 将至少一个合成转录因子或编码其的序列引入所述细胞系统中,

(c) 进一步向所述细胞系统中引入:

(i) 至少一个位点特异性核酸酶或编码其的序列,其中所述位点特异性核酸酶在所述预定位置诱导双链断裂;

(ii) 任选地:至少一个目的核苷酸序列,优选地侧翼为与所述细胞系统的遗传物质中所述预定位置相邻的一个或多个核苷酸序列互补的一个或多个同源序列;以及

(e) 任选地:确定修饰存在于所述细胞系统的遗传物质中的所述预定位置;以及

(f) 获得在所述细胞系统的遗传物质的所述预定位置处包含修饰的细胞系统;

其中所述至少一个合成转录因子或编码其的所述核苷酸序列包含至少一个识别结构域和至少一个激活结构域,其中所述至少一个合成转录因子配置为调控所述细胞系统中的至少一个形态发生基因的表达,优选地为调控所述细胞系统中的至少一个形态发生基因的转录;并且

其中所述至少一个合成转录因子或编码其的所述核苷酸序列的引入与所述至少一个位点特异性核酸酶或编码其的序列和所述任选的至少一个目的核苷酸序列的引入平行进行或序贯进行。

30. 根据权利要求29所述的方法,其中所述方法还包括在获得所述修饰的细胞系统的遗传修饰后代的条件下培养所述细胞系统的步骤。

31. 根据权利要求29所述的方法,其中,

(i) 所述至少一个合成转录因子或编码其的所述序列,或者所述至少一个合成转录因子的至少一个组分或编码其的序列;以及

(ii) 所述至少一个位点特异性核酸酶或包含其的序列;以及任选的

(iii) 所述至少一个目的核苷酸序列

通过独立地选自生物方法和/或物理方法的方法引入所述细胞系统中,所述方法包括转染、转化,包括通过农杆菌属物种(优选地为根癌农杆菌)、病毒载体、生物弹轰击进行转化,使用化学试剂进行转染(包括聚乙二醇转染),或其任何组合。

32. 根据权利要求29所述的方法,其中所述至少一个识别结构域是至少一个解除武装的CRISPR/核酸酶系统,或是至少一个解除武装的CRISPR/核酸酶系统的片段。

33. 根据权利要求32所述的方法,其中所述至少一个解除武装的CRISPR/核酸酶系统是CRISPR/dCpf1系统,其中所述至少一个解除武装的CRISPR/核酸酶系统包括至少一个向导RNA。

34. 根据权利要求29所述的方法,其中所述至少一个合成转录因子的所述至少一个激

活结构域选自酸性转录激活结构域,优选地,其中所述至少一个激活结构域来自水稻白叶枯病菌的无毒基因、单纯疱疹病毒的VP16或四聚体VP64、VPR、SAM、Scaffold、Suntag、P300、VP160或其任何组合。

35. 根据权利要求29所述的方法,其中所述至少一个合成转录因子的所述至少一个激活结构域相对于所述至少一个合成转录因子的所述至少一个识别结构域位于N-末端和/或C-末端。

36. 根据权利要求29所述的方法,其中所述至少一个形态发生基因选自BBM、WUS (包括WUS2)、WOX基因、WUS或BBM同源物、Lec1、Lec2、WIND1、ESR1、PLT3、PLT5、PLT7、IPT、IPT2、Knotted1和RKD4。

37. 根据权利要求29所述的方法,其中所述合成转录因子配置为通过结合位于相对于所述起始密码子一定距离处的调控区域来调控形态发生基因的表达,优选地为调控形态发生基因的转录。

38. 根据权利要求29所述的方法,其中所述合成转录因子和/或所述至少一个识别结构域包含SEQ ID NO:276、277、282、283、284、288、289、290中任一个所示的序列,或包含与SEQ ID NO:276、277、282、283、284、288、289、290中的任一个在全长上具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性的序列。

39. 根据权利要求29所述的方法,其中所述细胞系统选自至少一个真核细胞或真核生物体,优选地,其中所述至少一个真核细胞是至少一个植物细胞,和/或其中所述至少一个真核生物体是植物或植物的一部分。

40. 根据权利要求29所述的方法,其中位于所述预定位置处的所述至少一个目的核苷酸序列侧翼的一个或多个核苷酸序列与邻近所述预定位置,位于所述预定位置的上游和/或下游的一个或多个核苷酸序列在各个邻近区域的全长上至少85%至100%互补。

41. 一种产生单倍体或双单倍体生物体的方法,其中所述方法包括以下步骤:

(a) 提供单倍体细胞系统;

(b) 在所述单倍体细胞系统中引入至少一个合成转录因子或编码其的核苷酸序列;

(c) 在获得至少一个单倍体或双单倍体生物体的条件下培养所述单倍体细胞系统;以及

(d) 任选地:选择步骤(c)中获得的所述至少一个单倍体或双单倍体生物体,

其中所述至少一个合成转录因子或编码其的所述核苷酸序列包含至少一个识别结构域和至少一个激活结构域,其中所述至少一个合成转录因子配置为调控所述单倍体细胞系统中的至少一个形态发生基因的表达,优选地为调控所述单倍体细胞系统中的至少一个形态发生基因的转录。

42. 根据权利要求41所述的方法,其中步骤(a)的所述单倍体细胞系统是单倍体胚,或者其中步骤(c)中定义的所述至少一个单倍体或双单倍体生物体通过由(b)的所述单倍体细胞系统生成至少一个单倍体胚的中间步骤获得。

43. 根据权利要求41或42所述的方法,其中所述至少一个合成转录因子或编码其的序列,或者所述至少一个合成转录因子的至少一个组分或编码其的序列通过独立地选自生物方法和/或物理方法的方法引入所述单倍体细胞系统中,所述方法包括转染、转化,包括通

过农杆菌属物种(优选地为根癌农杆菌)、病毒载体、生物弹轰击进行转化,使用化学试剂进行转染(包括聚乙二醇转染),或其任何组合。

44. 根据权利要求41所述的方法,其中所述至少一个识别结构域是至少一个解除武装的CRISPR/核酸酶系统,或是至少一个解除武装的CRISPR/核酸酶系统的片段。

45. 根据权利要求44所述的方法,其中所述至少一个解除武装的CRISPR/核酸酶系统是CRISPR/dCpf1系统,其中所述至少一个解除武装的CRISPR/核酸酶系统包括至少一个向导RNA。

46. 根据权利要求41所述的方法,其中所述至少一个合成转录因子的所述至少一个激活结构域选自酸性转录激活结构域,优选地,其中所述至少一个激活结构域来自水稻白叶枯病菌的无毒基因、单纯疱疹病毒的VP16或四聚体VP64、VPR、SAM、Scaffold、Suntag、P300、VP160或其任何组合。

47. 根据权利要求41所述的方法,其中所述至少一个合成转录因子的所述至少一个激活结构域相对于所述至少一个合成转录因子的所述至少一个识别结构域位于N-末端和/或C-末端。

48. 根据权利要求41所述的方法,其中所述至少一个形态发生基因选自BBM、WUS(包括WUS2)、WOX基因、WUS或BBM同源物、Lec1、Lec2、WIND1、ESR1、PLT3、PLT5、PLT7、IPT、IPT2、Knotted1和RKD4。

49. 根据权利要求41所述的方法,其中所述合成转录因子配置为通过结合位于相对于所述起始密码子一定距离处的调控区域来调控所述形态发生基因的表达,优选地为调控所述形态发生基因的转录。

50. 根据权利要求41所述的方法,其中所述合成转录因子和/或所述至少一个识别结构域包含SEQ ID NO:276、277、282、283、284、288、289、290中任一个所示的序列,或包含与SEQ ID NO:276、277、282、283、284、288、289、290中的任一个在全长上具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性的序列。

51. 根据权利要求41所述的方法,其中所述至少一个单倍体细胞系统选自至少一个真核细胞或真核生物体,优选地,其中所述至少一个真核细胞是至少一个植物细胞,和/或其中所述至少一个真核生物体是植物或植物的一部分。

52. 一种通过权利要求14或16所述的方法获得的细胞系统或其后代。

53. 一种通过权利要求29所述的方法获得的细胞系统或其后代。

54. 一种通过权利要求41所述的方法获得的单倍体或双单倍体生物体。

55. 权利要求1所述的合成转录因子或编码其的序列在权利要求14所述的方法中的用途。

56. 权利要求1所述的合成转录因子或编码其的序列在权利要求29所述的方法中的用途。

57. 权利要求1所述的合成转录因子或编码其的序列在权利要求41所述的方法中的用途。

58. 权利要求2所述的合成转录因子或编码其的序列在权利要求16所述的方法中的用途。

基于Cpf1的植物转录调控系统

技术领域

[0001] 本发明涉及基因表达的靶向调控,更具体地涉及合成转录因子(STF),其包含至少一个基于CRISPR/Cpf1系统的高靶向特异性工程识别结构域,还包含至少一个激活或沉默结构域以调控目的基因的表达,优选地调控真核生物特别是植物的形态发生基因的转录。本发明还公开了使用STF提高转化频率,优化成功的基因组编辑方法,提供单倍体或双单倍体生物,以及/或者提供适于一般转化但也适于育种目的的组合物的方法。这些方法和用途依赖于包含基因表达调控结构域(例如,激活结构域或沉默结构域)的STF的协同增效作用,使得可以重编程细胞并诱导细胞分裂和/或再生,同时转化所述细胞或编辑所述细胞的基因组。

背景技术

[0002] 在真核细胞中有效转化和精确修饰遗传物质的能力使其在农产品开发、基础研究和其它技术领域中具有广泛的高价值应用。从根本上讲,基因组工程或基因编辑(GE)通过在真核基因组和原核基因组中特定位置引入预定义的遗传变异来提供此功能。同时,存在大量的用于在特定发育阶段转化不同真核或原核细胞的方法。此外,某些细胞类型或基因型的转化或转染效率有时仍然非常低,必须建立针对源自不同基因型的不同细胞进行微调的高特异性方法。

[0003] 此外,以高靶向方式修饰且特异性调控(即,激活或抑制)基因表达的能力在植物生物技术方面具有很高的价值。

[0004] 例如,虽然目前可以对主要的单子叶作物进行转化,但该方法通常局限于每个物种的一种或两种基因型,往往在农艺学上较差,且效率较差是这些方法难于农业实施。

[0005] 全球人口增长将迫使世界粮食产量在未来几十年内翻一番,同时气候变化也为植物育种者带来了新的挑战,因此非常需要具有生物胁迫或非生物胁迫抗性(例如,对新兴植物病原体的抗性或干旱抗性)的优化的作物植物。依靠经典育种和选择技术可能无法有效应对急剧增长的需求,也无法在未来几十年应对生态社会学变化实现持续供应。因此,必须开发新的策略和生物技术措施,建立植物可以更好地适应不利环境条件的性状。

[0006] 目前,玉米是世界上最重要的粮食和饲料作物之一,也是生物能源。同时,自20世纪90年代中期建立首例转基因苏云金芽孢杆菌(Bt)玉米产品以来,玉米已成为生物技术创新的最重要目标作物之一。尽管玉米基因组具有复杂性(与模型植物相比),但同时市场上玉米的生物技术性状比其他作物都多。自首次成功报道使用劳动密集型和耗费时间的原生质体转化方法(Rhodes等,1988a)以来,转基因玉米的生产已取得了巨大进展。微粒轰击转化(Fromm等,1990;Gordon-Kamm等,1990)和农杆菌介导转化(Ishida等,1996)技术的发展使转基因玉米的生产变得更简单、更可靠。在Hi-II中以BAR作为选择性标记(Frame等,2000)以及在优良自交系CG00526中以PMI作为选择性标记(Wright等,2001),建立高产生物弹射击转化系统。利用自交系A188(Ishida等,1996;Negrotto等,2000)、Hi-II(Zhao等,2001)和A188/Hi-II杂交(Li等,2003)报道了有效的农杆菌介导转化系统。在过去的几年

中,基因组工程技术的进步使得在玉米基因组中的特异性染色体靶位点进行修饰和插入转基因成为可能(Shukla等,2009;Gao等,2010;Liang等,2014;综述:Que等,《植物科学前沿》(Front.Plant.Sci.),2014,5,379)。然而,以上技术都不能提供适用于不同基因型的可靠且可转移的结果,更不用说用于不同植物。

[0007] 过去几十年来,植物生物技术领域的进步是在转基因农作物建立的基础上取得的。但是,社会经济和监管因素越来越表明,对于某些国家和地区而言,非转基因植物和植物产品的开发变得越来越重要。

[0008] 形态发生通常是指导致生物体发育的生物过程。它与控制细胞生长和细胞分化一样,是发育生物学的三个基本方面之一,三者进化发育生物学中是统一的。参与形态发生的重要的一类分子是转录因子蛋白,其通过与DNA相互作用来决定细胞的命运。这些分子可以由主调节基因编码,并且可以激活或去激活其它基因的转录;反过来,这些次级基因产物可以在基因调控网络的调控级联反应中调控其他基因的表达。在该级联的末尾,是控制细胞行为(例如,细胞迁移)或更普遍地控制其特性(例如,细胞粘着或细胞运动、细胞增殖和细胞凋亡)的分子类别。

[0009] 目前,Lowe等人的小组(Lowe等,“形态发生调控子Baby boom和Wuschel改善单子叶转化”(Morphogenic Regulators Baby boom and Wuschel Improve Monocot Transformation),《植物细胞》(Plant Cell),2016,第28卷:1998-2015)报道了一种涉及过表达玉米(*Zea mays*)形态发生基因Baby boom(BBM)和玉米Wuschel(WUS)基因的转化方法,在许多之前不可转化的玉米自交系中产生了高转化频率。Lowe等发现,以前难以转化的自交系中BBM和WUS的过表达导致转基因愈伤组织的再生能力增强。WUS和BBM在植物发育中的作用已经在前面进行了描述(US7,256,322B2或US 2013/0254935A1)。

[0010] 但是,目前,上述方法和其它方法都依赖于形态发生基因的异源过表达,例如在此类系统通常在其内不表达的细胞区室中,或者依赖于提供携带稳定地掺入其基因组的各个基因的转基因作物植物。另一种策略是例如使用诱导型和/或组织特异性启动子在时间或空间上调控靶基因的表达。但是,不受控制的过表达会导致表型变化,从而可能影响作物植物的适应性和产量效率,因此在农业中使用这种方法不具吸引力。因此,仍然非常需要确定新的策略来利用内源基因(包括形态发生因子)的功能,从而避免在目的细胞或细胞系统中过表达异源基因。

[0011] 许多植物细胞都具有仅从单个细胞或组织就可再生完整生物的能力。此过程通常称为全能性。整个植物的再生过程似乎与形态发生过程密切相关。体外培养的植物组织和细胞进行形态发生的能力会导致形成离散器官甚至整个植物,这为体外植物生物学在基础植物学、生物化学、育种和新作物植物发育的研究方面的众多应用提供了机会。

[0012] 单倍体是含有配子染色体数(n)的植物。它们可以自然产生,也可以是各种诱导技术的结果。自1992年Blakeslee首次描述曼陀罗(*Datura stramonium*)的这种现象(Blakeslee等,1922)以来,单倍体植物的自发发育就已经是公知了;随后在普通烟草(*Nicotiana tabacum*)、小麦(*Triticum aestivum*)和几个其它物种中也有类似的报道(Forster等,2007)。然而,单倍体的自发发生较为罕见,因此实用价值有限。

[0013] 由二倍体物种产生的单倍体,又称一倍体,在孢子体期仅包含一组染色体。与供体植物相比,它们更小并且表现出更低的植物生命力,并且由于它们的染色体在减数分裂过

程中无法配对,因此是不育的。为了通过种子繁殖它们并将其纳入育种计划,必须通过自发或诱导的染色体加倍来恢复它们的繁殖力。所获得的加倍单倍体或双单倍体在所有位点都是纯合的,可代表新品种(自花授粉作物)或用于生产杂交品种的亲本自交系(异花授粉作物)。实际上,异花授粉物种通常表现出高度的近交衰退。对于这些物种,诱导过程本身不仅可以作为生产纯合品系的快速方法,而且可以作为消除表达强近交衰退的基因型的选择工具。可以预期选择与营养生长相关的隐性有害基因引起的性状。因此,单倍体和双单倍体植物系统对于植物育种策略非常重要,但是对诸如形态发生途径的发育途径之间的相互作用及其在单倍体植物系统产生中的潜在影响却知之甚少。

[0014] 此外,在转化携带高价值基因型的优良种质时,存在着严重的问题,因为可从所述优良植物衍生的各植物或植物部分或可体外培养细胞通常是非常难以进行转化和/或转染的。这一事实使得目标植物的发育或育种变得非常复杂、耗时且昂贵,因为必须采用许多附加的育种和/或分子生物学步骤才能成功地将优良事件转移到目的遗传背景中。

[0015] 因此,本发明的目的是开发新的策略,以在其天然细胞环境中诱导内源基因,优选地为形态发生基因,从而通过已知技术改善转化困难或甚至非常难以进行转化/转染的作物植物的再生。此外,其目的是统一最新基因编辑技术可实现的高精度,以提供可调节方法来调控形态发生基因(优选地以瞬时的方式),从而使得靶细胞或组织的转化和再生能力更好,而不会过度影响细胞的内源性形态发生系统,其中该方法应配置为可以在与基因型无关的情况下提高转化/转染率。

[0016] 基于人工调控基因表达(主要是转录调控)的开发,另一个目的是提供针对转录控制具有沉默能力的合成转录因子,以提供有效的组合物来控制异常表达基因的转录和表达。

[0017] 再一个目的是基于形态发生基因的靶向修饰,建立提供单倍体和双单倍体植物细胞、细胞系统和整个生物体的新策略,从而为各种相关作物植物提供用于产生双单倍体的起始材料,所述双单倍体为完全纯合品系,表示植物育种和植物生物技术中的宝贵工具。

[0018] 已经利用具有已知可以在被募集到启动子区域时激活或抑制基因转录的转录效应物结构域的失活CRISPR核酸内切酶融合构建体,开发了转录调控工具。到目前为止,基于CRISPR/Cas9的转录激活和抑制系统已经可用于哺乳动物细胞和植物细胞系统(Chen等(2013),RNA向导的转录激活子系统CRISPR-on对内源基因的多重激活(Multiplexed activation of endogenous genes by CRISPR-on,an RNA-guided transcriptional activator system),《细胞研究》(Cell Research),23:1163-1171;Lowder等(2015),用于植物基因组多重编辑与转录调控的CRISPR/Cas9工具箱(A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation),《植物生理学》(Plant Physiology),169:971-985;Lowder等(2017),多重CRISPR-Act2.0和mTALE-Act系统对植物进行稳健的转录激活(Robust transcriptional activation in plants using multiplexed CRISPR-Act2.0 and mTALE-Act systems),《分子植物》(Molecular Plant),11:245-256;以及Li等(2017),对植物和动物细胞有效的Cas9驱动基因激活子(A potent Cas9-driven gene activator for plant and animal cells),《自然植物》(Nature Plants),3:930-936)。

[0019] 与基于Cas9的转录激活系统相比,基于Cpf1的转录激活系统具有多个优势。它们

可用于靶向富含AT的启动子区域,而基于Cas9的系统则对富含GC的区域具有特异性。由于Cpf1的RNase活性能够从单一转录物加工多个crRNA,因此基于Cpf1的转录调控系统相对于众所周知的基于Cas9的系统具有优势,即,它可以轻松应用于多重基因调控。

[0020] 然而,基于Cpf1的转录激活系统目前仅适用于哺乳动物细胞系统(Tak等(2017),使用基于CRISPR/Cpf1的转录因子进行诱导和多重基因调控(Inducible and multiplex gene regulation using CRISPR/Cpf1 based transcription factors),《自然方法》(Nature Methods),14(12):1163-1166;以及Liu等(2017),利用基于CRISPR/Cpf1的可调控转录因子的工程细胞信号传导(Engineering cell signaling using tunable CRISPR/Cpf1 based transcription factors),《自然-通信》(Nature Communications),8(1):2095),尽管已经在拟南芥(Arabidopsis)中证明了基于Cpf1的转录抑制(Tang等(2017),用于植物中有效基因组编辑和转录阻遏的CRISPR/Cpf1系统(A CRISPR/Cpf1 system for efficient genome editing and transcriptional repression in plants),《自然植物》(Nature Plants),3:17018)。到目前为止,在植物中尚未显示出基于Cpf1的转录激活,这表明,转录抑制结构域(如Tang等中使用的那个)简单地由转录激活结构域替换是不可能的,并且需要精心配置和测试正确的接头和激活结构域序列。因此,根据现有技术并不能确定在基于Cpf1的系统中简单地用激活结构域替换抑制结构域是否会引起内源基因表达的激活。相反,现有技术表明,需要大量的修饰和实验来提供可用于植物细胞的基于Cpf1的转录激活子。

[0021] 特别地,因此本发明的目的是提供一种基于Cpf1的转录激活(或抑制)系统,其可用于多种作物植物中以靶向富含AT的启动子区域,优选内源基因的启动子区域。该系统应易于用于多重化,即,通过使用向导RNA阵列同时靶向多个基因组区域。此外,应可以在无转基因环境中暂时使用该系统。此外,本发明的另一个目的是,建立通过特异性靶向形态发生基因以增强表达来改善转化效率和基因组修饰技术的方法。

发明内容

[0022] 一方面,上述目的通过提供一种合成转录因子或编码其的核苷酸序列而实现,其包含至少一个识别结构域和至少一个基因表达调控结构域,尤其是激活结构域,其中,该合成转录因子配置为调控细胞系统中形态发生基因的表达。

[0023] 还提供了一种合成转录因子,其中该至少一个识别结构域是至少一个解除武装(disarmed)的CRISPR/核酸酶系统,或是至少一个解除武装的CRISPR/核酸酶系统的片段。

[0024] 在一个实施方案中,提供了一种合成转录因子,其中该至少一个解除武装的CRISPR/核酸酶系统是CRISPR/dCpf1系统,其中该至少一个解除武装的CRISPR/核酸酶系统包含至少一个向导RNA。

[0025] 在另一个实施方案中,提供了一种合成转录因子,其中该至少一个激活结构域选自自由酸性转录激活结构域所构成的组,优选地,其中该至少一个激活结构域来自水稻白叶枯病菌(*Xanthomonas oryzae*)的无毒基因、单纯疱疹病毒的VP16(SEQ ID NO:259)或四聚体VP64(SEQ ID NO:260)、VPR(SEQ ID NO:261)、SAM(SEQ ID NO:262;SEQ ID NO:263)、Scaffold(SEQ ID NO:264;SEQ ID NO:265)、Suntag(SEQ ID NO:266;SEQ ID NO:267)、P300(SEQ ID NO:268)、VP160(SEQ ID NO:269)或其任何组合。在本发明的优选实施方案

中,该激活结构域是VPR。

[0026] 在又一个实施方案中,提供了一种合成转录因子,其中该至少一个激活结构域相对于该至少一个识别结构域位于N-末端和/或C-末端。

[0027] 在一个实施方案中,提供了一种合成转录因子,其中该形态发生基因选自BBM、WUS (包括WUS2)、WOX基因、WUS或BBM同源物、Lec1、Lec2、WIND1、ESR1、PLT3、PLT5、PLT7、IPT、IPT2、Knotted1和RKD4所构成的组。

[0028] 在另一个实施方案中,提供了一种合成转录因子,其中该形态发生基因包含选自以下各项所构成的组的核苷酸序列:(i) SEQ ID NO:199-237中任一个所示的核苷酸序列,(ii) 具有SEQ ID NO:199-237中任一个所示的核苷酸序列的编码序列的核苷酸序列,(iii) 与(i)或(ii)的核苷酸序列互补的核苷酸序列,(iv) 与(i)、(ii)或(iii)的核苷酸序列具有至少50%、60%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性(优选在全长上)的核苷酸序列,(v) 在严格条件下与(iii)的核苷酸序列杂交的核苷酸序列,(vi) 编码包含SEQ ID NO:238-258中任一个所示的氨基酸序列的蛋白的核苷酸序列,(vii) 编码包含与SEQ ID NO:238-258中任一个所示的序列至少50%、60%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的氨基酸序列的蛋白的核苷酸序列,或(viii) 编码包含SEQ ID NO:238-258中任一个所示的氨基酸序列的蛋白的同源物、类似物或直系同源物的核苷酸序列。

[0029] 在另一个实施方案中,提供了一种合成转录因子,其中该合成转录因子配置为通过结合至位于相对于起始密码子一定距离处的调控区域来调控形态发生基因的表达,优选地为调控形态发生基因的转录。

[0030] 在又一个实施方案中,提供了一种合成转录因子,其中该合成转录因子和/或该至少一个识别结构域包含SEQ ID No:276、277、282、283、284、288、289、290中任一个所示的序列,或包含与SEQ ID No:276、277、282、283、284、288、289、290中的任一个在全长上具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性的序列。

[0031] 在另一个实施方案中,提供了一种合成转录因子,其中该细胞系统选自至少一个真核细胞或真核生物体,优选地,其中该至少一个真核细胞是至少一个植物细胞,并且/或者其中该至少一个真核生物体是植物或植物的一部分。

[0032] 在一个实施方案中,提供了一种合成转录因子,其中,该植物的至少一部分选自叶、茎、根、初生胚根(emerged radicles)、花、花部分、花瓣、果实、花粉、花粉管、花药花丝、胚珠、胚囊、卵细胞、子房、合子、胚、合子胚、体细胞胚、顶端分生组织、维管束、中柱鞘、种子、根和插枝所构成的组。

[0033] 在另一个实施方案中,提供了一种合成转录因子,其中,该至少一个植物细胞、该至少一个植物或植物的至少一部分源自选自以下的植物物种:大麦(*Hordeum vulgare*)、球茎大麦(*Hordeum bulbosum*)、高粱(*Sorghum bicolor*)、甘蔗(*Saccharum officinarum*)、玉米(*Zea mays*)、小米(*Setaria italica*)、小粒野生稻(*Oryza minuta*)、水稻(*Oryza sativa*)、澳洲野生稻(*Oryza australiensis*)、高秆野生稻(*Oryza alta*)、小麦(*Triticum aestivum*)、黑麦(*Secale cereale*)、苹果(*Malus domestica*)、二穗短柄草(*Brachypodium distachyon*)、滨海大麦草(*Hordeum marinum*)、节节麦(*Aegilops tauschii*)、胡萝卜

(*Daucus glochidiatus*)、甜菜 (*Beta vulgaris*)、小胡萝卜 (*Daucus pusillus*)、*Daucus muricatus*、野胡萝卜 (*Daucus carota*)、巨桉 (*Eucalyptus grandis*)、美花烟草 (*Nicotiana sylvestris*)、绒毛状烟草 (*Nicotiana tomentosiformis*)、普通烟草 (*Nicotiana tabacum*)、番茄 (*Solanum lycopersicum*)、马铃薯 (*Solanum tuberosum*)、中果咖啡 (*Coffea canephora*)、葡萄 (*Vitis vinifera*)、*Erythrante guttata*、*Genlisea aurea*、黄瓜 (*Cucumis sativus*)、川桑 (*Morus notabilis*)、*Arabidopsis arenosa*、琴叶拟南芥 (*Arabidopsis lyrata*)、拟南芥 (*Arabidopsis thaliana*)、近缘种须弥芥 (*Crucihimalaya himalaica*)、卵叶须弥芥 (*Crucihimalaya wallichii*)、弯曲碎米荠 (*Cardamine flexuosa*)、北美独行菜 (*Lepidium virginicum*)、荠菜 (*Capsella bursa pastoris*)、*Olm arabidopsis pumila*、硬毛南芥 (*Arabis hirsute*)、甘蓝型油菜 (*Brassica napus*)、甘蓝 (*Brassica oleracea*)、芜菁 (*Brassica rapa*)、萝卜 (*Raphanus sativus*)、芥菜 (*Brassica juncea*)、黑芥菜 (*Brassica nigra*)、芝麻菜 (*Eruca vesicaria* subsp. *sativa*)、甜橙 (*Citrus sinensis*)、麻疯树 (*Jatropha curcas*)、毛果杨 (*Populus trichocarpa*)、蒺藜苜蓿 (*Medicago truncatula*)、*Cicer yamashitae*、*Cicer bijugum*、鹰嘴豆 (*Cicer arietinum*)、*Cicer reticulatum*、*Cicer judaicum*、木豆 (*Cajanus cajanifolius*)、蔓草虫豆 (*Cajanus scarabaeoides*)、菜豆 (*Phaseolus vulgaris*)、大豆 (*Glycine max*)、紫云英 (*Astragalus sinicus*)、百脉根 (*Lotus japonicas*)、夏堇 (*Torenia fournieri*)、洋葱 (*Allium cepa*)、大葱 (*Allium fistulosum*)、大蒜 (*Allium sativum*) 和韭菜 (*Allium tuberosum*)。

[0034] 一方面,提供了一种用于提高细胞系统中转化效率的方法,其中,该方法包括以下步骤:(a) 提供细胞系统;(b) 在该细胞系统中引入至少一个合成转录因子或编码其的核苷酸序列;以及(c) 在该细胞系统中引入至少一个目的核苷酸序列;(d) 可选地:在获得该细胞系统后代的条件下培养该细胞系统;其中该至少一个合成转录因子或编码其的核苷酸序列包含至少一个识别结构域和至少一个基因表达调控结构域,特别是至少一个激活结构域,其中该合成转录因子配置为调控该细胞系统中形态发生基因的表达,优选地是调控该细胞系统中形态发生基因的转录;并且其中,该至少一个合成转录因子或编码其的核苷酸序列的引入与该至少一个目的核苷酸序列的引入平行进行,或序贯进行。

[0035] 在一个实施方案中,提供了一种方法,其中,(a) 该至少一个合成转录因子或编码其的序列,或该至少一个合成转录因子的至少一种组分或编码其的序列;以及(b) 该至少一个目的核苷酸序列,通过独立地选自生物和/或物理方法的方法被引入该细胞系统中,该方法包括转染、转化,其中包括通过农杆菌属物种(优选地为根癌农杆菌)、病毒载体、生物弹轰击进行转化、使用化学试剂进行转染(包括聚乙二醇转染)、电穿孔、细胞融合或其任何组合。

[0036] 在又一个实施方案中,提供了一种方法,其中该至少一个识别结构域是至少一个解除武装的CRISPR/核酸酶系统,或是至少一个解除武装的CRISPR/核酸酶系统的片段。

[0037] 在另一个实施方案中,提供了一种方法,其中该至少一个解除武装的CRISPR/核酸酶系统是CRISPR/dCpf1系统,其中该至少一个解除武装的CRISPR/核酸酶系统包含至少一个向导RNA。

[0038] 在另一个实施方案中,提供了一种方法,其中该至少一个合成转录因子的至少一

个激活结构域选自由酸性转录激活结构域所构成的组,优选地,其中该至少一个激活结构域来自水稻白叶枯病菌的无毒基因、单纯疱疹病毒的VP16或四聚体VP64、VPR、SAM、Scaffold、Suntag、P300、VP160或其任何组合。在本发明的优选实施方案中,该激活结构域是VPR (SEQ ID NO:276)。

[0039] 在又一个实施方案中,提供了一种方法,其中该至少一个合成转录因子的至少一个激活结构域相对于该至少一个合成转录因子的至少一个识别结构域位于N-末端和/或C-末端。

[0040] 在另一个实施方案中,提供了一种方法,其中该至少一个形态发生基因选自由BBM、WUS (包括WUS2)、WOX基因、WUS或BBM同源物、Lec1、Lec2、WIND1、ESR1、PLT3、PLT5、PLT7、IPT、IPT2、Knotted1和RKD4所构成的组。

[0041] 在另一个实施方案中,提供了一种方法,其中该至少一个形态发生基因包含选自由以下各项所构成的组的核苷酸序列:(i) SEQ ID NO:199-237中任一个所示的核苷酸序列,(ii) 具有SEQ ID NO:199-237中任一个所示的核苷酸序列的编码序列的核苷酸序列,(iii) 与(i)或(ii)的核苷酸序列互补的核苷酸序列,(iv) 与(i)、(ii)或(iii)的核苷酸序列具有至少50%、60%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性(优选在全长上)的核苷酸序列,(v) 在严格条件下与(iii)的核苷酸序列杂交的核苷酸序列,(vi) 编码包含SEQ ID NO:238-258中任一个所示的氨基酸序列的蛋白的核苷酸序列,(vii) 编码包含与SEQ ID NO:238-258中任一个所示的序列至少50%、60%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的氨基酸序列的蛋白的核苷酸序列,或(viii) 编码包含SEQ ID NO:238-258中任一个所示的氨基酸序列的蛋白的同源物、类似物或直系同源物的核苷酸序列。

[0042] 在另一个实施方案中,提供了一种方法,其中该合成转录因子配置为通过结合至位于相对于起始密码子一定距离处的调控区域来调控形态发生基因的表达,优选地为调控形态发生基因的转录。

[0043] 在一个实施方案中,提供了一种方法,其中该合成转录因子和/或该至少一个识别结构域包含SEQ ID NO:276、277、282、283、284、288、289、290中任一个所示的序列,或包含与SEQ ID NO:276、277、282、283、284、288、289、290中的任一个在全长上具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性的序列。

[0044] 在另一个实施方案中,提供了一种方法,其中该细胞系统选自至少一个真核细胞或真核生物体,优选地,其中该至少一个真核细胞是至少一个植物细胞,并且/或者其中该至少一个真核生物体是植物或植物的一部分。

[0045] 在另一个实施方案中,提供了一种方法,其中,该植物的至少一部分选自由叶、茎、根、初生胚根(emerged radicles)、花、花部分、花瓣、果实、花粉、花粉管、花药花丝、胚珠、胚囊、卵细胞、子房、合子、胚、合子胚、体细胞胚、顶端分生组织、维管束、中柱鞘、种子、根和插枝所构成的组。

[0046] 在又一个实施方案中,提供了一种方法,其中,该至少一个植物细胞、该至少一个植物或植物的至少一部分源自选自以下的植物物种:大麦(*Hordeum vulgare*)、球茎大麦(*Hordeum bulbosum*)、高粱(*Sorghum bicolor*)、甘蔗(*Saccharum officinarum*)、玉米

(*Zea mays*)、小米(*Setaria italica*)、小粒野生稻(*Oryza minuta*)、水稻(*Oriza sativa*)、澳洲野生稻(*Oryza australiensis*)、高秆野生稻(*Oryza alta*)、小麦(*Triticum aestivum*)、黑麦(*Secale cereale*)、苹果(*Malus domestica*)、二穗短柄草(*Brachypodium distachyon*)、滨海大麦草(*Hordeum marinum*)、节节麦(*Aegilops tauschii*)、胡萝卜(*Daucus glochidiatus*)、甜菜(*Beta vulgaris*)、小胡萝卜(*Daucus pusillus*)、*Daucus muricatus*、野胡萝卜(*Daucus carota*)、巨桉(*Eucalyptus grandis*)、美花烟草(*Nicotiana sylvestris*)、绒毛状烟草(*Nicotiana tomentosiformis*)、普通烟草(*Nicotiana tabacum*)、番茄(*Solanum lycopersicum*)、马铃薯(*Solanum tuberosum*)、中果咖啡(*Coffea canephora*)、葡萄(*Vitis vinifera*)、*Erythrante guttata*、*Genlisea aurea*、黄瓜(*Cucumis sativus*)、川桑(*Morus notabilis*)、*Arabidopsis arenosa*、琴叶拟南芥(*Arabidopsis lyrata*)、拟南芥(*Arabidopsis thaliana*)、近缘种须弥芥(*Crucihimalaya himalaica*)、卵叶须弥芥(*Crucihimalaya wallichii*)、弯曲碎米荠(*Cardamine flexuosa*)、北美独行菜(*Lepidium virginicum*)、荠菜(*Capsella bursa pastoris*)、*Olm arabidopsis pumila*、硬毛南芥(*Arabis hirsute*)、甘蓝型油菜(*Brassica napus*)、甘蓝(*Brassica oleracea*)、芜菁(*Brassica rapa*)、萝卜(*Raphanus sativus*)、芥菜(*Brassica juncea*)、黑芥菜(*Brassica nigra*)、芝麻菜(*Eruca vesicaria* subsp. *sativa*)、甜橙(*Citrus sinensis*)、麻疯树(*Jatropha curcas*)、毛果杨(*Populus trichocarpa*)、蒺藜苜蓿(*Medicago truncatula*)、*Cicer yamashitae*、*Cicer bijugum*、鹰嘴豆(*Cicer arietinum*)、*Cicer reticulatum*、*Cicer judaicum*、木豆(*Cajanus cajanifolius*)、蔓草虫豆(*Cajanus scarabaeoides*)、菜豆(*Phaseolus vulgaris*)、大豆(*Glycine max*)、紫云英(*Astragalus sinicus*)、百脉根(*Lotus japonicas*)、夏堇(*Torenia fournieri*)、洋葱(*Allium cepa*)、大葱(*Allium fistulosum*)、大蒜(*Allium sativum*)和韭菜(*Allium tuberosum*)。

[0047] 另一方面,提供了一种在预定位置修饰细胞系统的遗传物质的方法,其中,该方法包括以下步骤:(a) 提供细胞系统;(b) 将至少一个合成转录因子或编码其的序列引入该细胞系统中,(c) 再将以下各项引入该细胞系统中:(i) 至少一个位点特异性核酸酶或编码其的序列,其中该位点特异性核酸酶在预定位置诱导双链断裂;(ii) 可选的:至少一个目的核苷酸序列,优选地侧翼为与细胞系统遗传物质中预定位置相邻的一个或多个核苷酸序列互补的一个或多个同源序列;以及(e) 可选地:确定修饰存在于细胞系统的遗传物质中的预定位置;以及(f) 获得在细胞系统遗传物质的预定位置处包含修饰的细胞系统;其中该至少一种合成转录因子或编码其的核苷酸序列包含至少一个识别结构域和至少一个激活结构域,其中该至少一个合成转录因子配置为调控该细胞系统中的至少一个形态发生基因的表达,优选地为调控该细胞系统中的至少一个形态发生基因的转录;并且其中该至少一个合成转录因子或编码其的核苷酸序列的引入,与该至少一个位点特异性核酸酶或编码其的序列以及该可选的至少一个目的核苷酸序列的引入平行进行,或序贯进行。

[0048] 在该方面的另一个实施方案中,提供了一种方法,其中该方法还包括在获得所修饰的细胞系统的遗传修饰后代的条件下培养该细胞系统的步骤。

[0049] 在预定位置修饰细胞系统遗传物质的方法的另一个实施方案中,提供了一种方法,其中,(i) 至少一种合成转录因子或编码其的序列,或该至少一种合成转录因子的至少

一种组分或编码其的序列;以及(ii)该至少一个位点特异性核酸酶或包含其的序列;以及可选的(iii)该至少一个目的核苷酸序列通过独立地选自生物方法和/或物理方法的方法引入该细胞系统,包括转染、转化,包括通过农杆菌属物种(优选地为根癌农杆菌)、病毒载体、生物弹轰击进行转化,使用化学试剂进行转染(包括聚乙二醇转染),电穿孔,细胞融合或其任何组合。

[0050] 在一个实施方案中,提供了一种方法,其中该至少一个识别结构域是至少一个解除武装的CRISPR/核酸酶系统,或是至少一个解除武装的CRISPR/核酸酶系统的片段。

[0051] 在另一个实施方案中,提供了一种方法,其中该至少一个解除武装的CRISPR/核酸酶系统是CRISPR/dCpf1系统,其中该至少一个解除武装的CRISPR/核酸酶系统包含至少一个向导RNA。

[0052] 还提供了上述方法的实施方案,其中该至少一个合成转录因子的至少一个激活结构域选自由酸性转录激活结构域所构成的组,优选地,其中该至少一个激活结构域来自水稻白叶枯病菌的无毒基因、单纯疱疹病毒的VP16或四聚体VP64、VPR、SAM、Scaffold、Suntag、P300、VP160或其任何组合。在本发明的优选实施方案中,该激活结构域是VPR (SEQ ID NO:276)。

[0053] 在一个实施方案中,提供了一种方法,其中该至少一个合成转录因子的至少一个激活结构域相对于该至少一个合成转录因子的至少一个识别结构域位于N-末端和/或C-末端。

[0054] 在另一个实施方案中,提供了一种方法,其中该至少一个形态发生基因选自由BBM、WUS(包括WUS2)、WOX基因、WUS或BBM同源物、Lec1、Lec2、WIND1、ESR1、PLT3、PLT5、PLT7、IPT、IPT2、Knotted1和RKD4所构成的组。

[0055] 在另一个实施方案中,提供了一种方法,其中该至少一个形态发生基因包含选自由以下各项所构成的组的核苷酸序列:(i) SEQ ID NO:199-237中任一个所示的核苷酸序列,(ii)具有SEQ ID NO:199-237中任一个所示的核苷酸序列的编码序列的核苷酸序列,(iii)与(i)或(ii)的核苷酸序列互补的核苷酸序列,(iv)与(i)、(ii)或(iii)的核苷酸序列具有至少50%、60%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性(优选在全长上)的核苷酸序列,(v)在严格条件下与(iii)的核苷酸序列杂交的核苷酸序列,(vi)编码包含SEQ ID NO:238-258中任一个所示的氨基酸序列的蛋白的核苷酸序列,(vii)编码包含与SEQ ID NO:238-258中任一个所示的序列至少50%、60%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的氨基酸序列的蛋白的核苷酸序列,或(viii)编码包含SEQ ID NO:238-258中任一个所示的氨基酸序列的蛋白的同源物、类似物或直系同源物的核苷酸序列。

[0056] 在另一个实施方案中,提供了一种方法,其中该合成转录因子配置为通过结合位于相对于起始密码子一定距离处的调控区域来调控形态发生基因的表达,优选地为调控形态发生基因的转录。

[0057] 在又一个实施方案中,提供了一种方法,其中该合成转录因子和/或该至少一个识别结构域包含SEQ ID No:276、277、282、283、284、288、289、290中任一个的序列,或包含与SEQ ID No:276、277、282、283、284、288、289、290中的任一个在全长上具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性的序

列。

[0058] 在另一个实施方案中,提供了一种方法,其中该细胞系统选自至少一个真核细胞或真核生物体,优选地,其中该至少一个真核细胞是至少一个植物细胞,并且/或者其中该至少一个真核生物体是植物或植物的一部分。

[0059] 在又一个实施方案中,提供了一种方法,其中预定位置处的至少一个目的核苷酸序列侧翼的一个或多个核苷酸序列与邻近预定位置(位于预定位置的上游和/或下游)的一个或多个核苷酸序列在各个邻近区域的全长上至少85%至100%互补。

[0060] 在本发明的另一个方面,提供了一种产生单倍体或双单倍体细胞系统或生物体的方法,其中该方法包括以下步骤:(a)提供单倍体细胞系统;(b)在该单倍体细胞系统中引入至少一个合成转录因子或编码其的核苷酸序列;(c)在获得至少一种单倍体或双单倍体生物体的条件下培养该单倍体细胞系统;以及(d)可选地,选择步骤(c)中获得的至少一个单倍体或双单倍体生物体,其中该至少一个合成转录因子或编码其的核苷酸序列包含至少一个识别结构域和至少一个激活结构域,其中该至少一个合成转录因子配置为调控该单倍体细胞系统中的至少一个形态发生基因的表达,优选地为调控该单倍体细胞系统中的至少一个形态发生基因的转录。

[0061] 在一个实施方案中,提供了一种方法,其中上述方法的步骤(a)的单倍体细胞系统是单倍体胚,或者其中,上述方法的步骤(c)的至少一个单倍体或双单倍体生物体是通过从(b)的单倍体细胞系统产生至少一个单倍体胚的中间步骤获得的。

[0062] 在一个实施方案中,提供了一种方法,其中该至少一个合成转录因子或编码其的序列,或者该至少一个合成转录因子的至少一个组分或编码其的序列通过独立选自生物和或物理方法的方法引入该单倍体细胞系统中,该方法包括转染、转化,其中包括通过农杆菌属物种(优选地为根癌农杆菌)、病毒载体、生物弹轰击进行转化,使用化学试剂进行转染(包括聚乙二醇转染),电穿孔,细胞融合或其任何组合。

[0063] 在另一个实施方案中,提供了一种方法,其中该至少一个识别结构域是至少一个解除武装的CRISPR/核酸酶系统,或是至少一个解除武装的CRISPR/核酸酶系统的片段。

[0064] 在又一个实施方案中,提供了一种方法,其中该至少一个解除武装的CRISPR/核酸酶系统是CRISPR/dCpf1系统,其中该至少一个解除武装的CRISPR/核酸酶系统包含至少一个向导RNA。

[0065] 在另一个实施方案中,提供了一种方法,其中该至少一个合成转录因子的至少一个激活结构域选自由酸性转录激活结构域所构成的组,优选地,其中该至少一个激活结构域来自水稻白叶枯病菌的无毒基因、单纯疱疹病毒的VP16或四聚体VP64、VPR、SAM、Scaffold、Suntag、P300、VP160或其任何组合。在本发明的优选实施方案中,该激活结构域是VPR(SEQ ID NO:276)。

[0066] 在另一个实施方案中,提供了一种方法,其中该至少一个合成转录因子的至少一个激活结构域相对于该至少一个合成转录因子的至少一个识别结构域位于N-末端和/或C-末端。

[0067] 在又一个实施方案中,提供了一种方法,其中该至少一个形态发生基因选自由BBM、WUS(包括WUS2)、WOX基因、WUS或BBM同源物、Lec1、Lec2、WIND1、ESR1、PLT3、PLT5、PLT7、IPT、IPT2、Knotted1和RKD4所构成的组。

[0068] 在另一个实施方案中,提供了一种方法,其中该至少一个形态发生基因包含选自以下各项所构成的组的核苷酸序列:(i) SEQ ID NO:199-237中任一个所示的核苷酸序列,(ii) 具有SEQ ID NO:199-237中任一个所示的核苷酸序列的编码序列的核苷酸序列,(iii) 与(i)或(ii)的核苷酸序列互补的核苷酸序列,(iv) 与(i)、(ii)或(iii)的核苷酸序列具有至少50%、60%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性(优选在全长上)的核苷酸序列,(v) 在严格条件下与(iii)的核苷酸序列杂交的核苷酸序列,(vi) 编码包含SEQ ID NO:238-258中任一个所示的氨基酸序列的蛋白的核苷酸序列,(vii) 编码包含与SEQ ID NO:238-258中任一个所示的序列至少50%、60%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的氨基酸序列的蛋白的核苷酸序列,或(viii) 编码包含SEQ ID NO:238-258中任一个所示的氨基酸序列的蛋白的同源物、类似物或直系同源物的核苷酸序列。

[0069] 在一个实施方案中,提供了一种方法,其中该合成转录因子配置为通过结合位于相对于起始密码子一定距离处的调控区域来调控形态发生基因的表达,优选地为调控形态发生基因的转录。

[0070] 在另一个实施方案中,提供了一种方法,其中该合成转录因子和/或该至少一个识别结构域包含SEQ ID No:276、277、282、283、284、288、289、290中任一个所示的序列,或包含与SEQ ID No:276、277、282、283、284、288、289、290中的任一个在全长上具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性的序列。

[0071] 在又一个实施方案中,提供了一种方法,其中该至少一个单倍体细胞系统选自至少一个真核细胞或真核生物体,优选地,其中该至少一个真核细胞是至少一个植物细胞,并且/或者其中该至少一个真核生物体是植物或植物的一部分。

[0072] 还提供了通过本文提供的任何一种方法获得的细胞系统或其后代。

[0073] 在另一方面,提供了一种通过本文提供的任何一种方法获得的单倍体或双单倍体细胞系统或生物体。

[0074] 在另一方面,提供了本文提供的合成转录因子或编码其的序列在本文提供的任何方法中的用途。

[0075] 在另一方面,提供了一种合成转录因子或编码其的核苷酸序列,其包含至少一个识别结构域和至少一个激活结构域,其中该合成转录因子配置为调控细胞系统中内源基因的表达,优选地为调控细胞系统中内源基因的转录。

[0076] 在另一方面,提供了一种用于提高细胞系统中至少一个内源基因表达的方法,其中该方法包括以下步骤:

[0077] (a) 提供细胞系统;

[0078] (b) 在该细胞系统中引入至少一个合成转录因子或编码其的核苷酸序列;

[0079] 其中该至少一个合成转录因子或其编码的核苷酸序列包含至少一个识别结构域和至少一个激活结构域,其中,该合成转录因子配置为提高该细胞系统中的至少一个内源基因的表达,优选地为提高该细胞系统中的至少一个内源基因的转录。

[0080] 本发明的其他方面和实施方案可以从随后的具体实施方式、附图、序列表以及所附权利要求中得出。

[0081] 附图及序列的简要说明

[0082] 图1.用于靶向基因激活修饰的合成转录因子(STF)的示例性实施例。(A)示出了经由TAL转录因子的靶向基因激活。TAL转录因子由与例如转录激活因子样效应物(TALE)的DNA结合结构域融合的激活结构域(例如,VP64)组成。(B)示出了经由CRISPR/dCas9和/或CRISPR/dCpf1转录系统的靶向基因激活。CRISPR/dCas9和CRISPR/dCpf1转录因子系统包含与激活结构域(例如,VP64)融合的解除武装的核酸酶(例如,dCas9或dCpf1)。DNA结合由与解除武装的核酸酶相关的向导RNA介导。在与目的形态发生基因的转录起始位点非常接近的基因组靶位点结合后,STF通过激活结构域将RNA聚合酶II复合物(即,转录复合物)募集到形态发生基因的启动子区域(基因的转录在此进行)中。

[0083] 图2.通过基因编辑机器与对形态发生基因具有特异性的示例性合成转录因子(STF)共转染改进基因编辑的示意图。通过基因编辑机器(例如,CRISPR/Cpf1或CRISPR/Cas9)的修饰(诸如INDEL或用修复模板替换靶基因)会产生基因修饰的植物细胞。基因编辑机器与对BBM和WUS具有特异性的一种或多种STF的瞬时共转染可确保靶细胞的恢复并增加被编辑植物的再生。

[0084] 图3.靶向内源性Wuschel(WUS)和Babyboom(BBM)基因的Tal效应物结合位点的设计。这些位点与起始密码子的距离不同。(A)内源性WUS的结合位点(其示出部分如SEQ ID NO:315所示)的长度为18个碱基对,并且还包含起始的T核碱基(TALE 1、2、3)。(B)内源性BBM的结合位点(其示出部分如SEQ ID NO:316所示)的长度为24个碱基对,并且还包含起始的T核碱基(TALE 4、5、6)。

[0085] 图4.TALE转录因子对内源性WUS和BBM的瞬时表达。在玉米原生质体测定系统中测试了由TAL转录因子对基因表达的诱导。通过使用基于PEG的转化系统,用包含靶向WUS或BBM的TALE转录因子的载体构建体转化玉米原生质体。实验一式三份进行,并重复四次作为生物学重复。24小时后,使用市售试剂盒从提取的原生质体RNA中产生cDNA。使用SYBR Green qRT-PCR方法确定内源性WUS和BBM的表达。(A)结果表明,合成转录因子TALE1是内源性WUS的最强诱导物,显示出内源性WUS基因表达的平均倍数变化为60。(B)结果表明,合成转录因子TALE5是内源性BBM的最强诱导物,显示出内源性BBM基因表达平均变化490倍。

[0086] 图5.瞬时TALE转录因子诱导的内源性ZmWUS的表型功能的评估。为了评估合成转录因子对再生和胚胎形成的影响,用荧光标记tdTomato(tdT)、TALE1和PLT7通过粒子轰击转化玉米A188的愈伤组织。将构建体递送至单个细胞,并且在检测到tdT的红色荧光信号时通过荧光显微镜确认细胞增殖的诱导(参见白色圆圈和箭头)。

[0087] 图6.实施例13中制备的pGEP767(A)、pGEP761(B)和pGEP772(C)的质粒图。

[0088] 图7:实施例14中的ZmBBM基因(A)(其示出部分如SEQ ID NO:317所示)和ZmWUS2基因(B)(其示出部分如SEQ ID NO:318所示)的向导RNA设计。选定的TTTV、TYCV和TATV PAM用相应的箭头标记。设计的向导RNA如黑色箭头所示。在转录激活中测试的那些以圆圈突出显示。

[0089] 图8:表达向导RNA的最终构建体的代表pGEP667的质粒图(此处为:crGEP186)。

[0090] 图9:如实施例15中确定的WUS2和BBM表达的转录激活。通过使用靶向WUS2启动子区域的向导RNA,所测试的向导(crGEP186和crGEP201)引起WUS2表达的显著激活(A)。同样地,靶向BBM启动子区域的两个向导RNA(crGEP210和crGEP211)引起BBM表达的显著激活

(B)。将仅用LbCpf1-VPR表达载体转化的样品中BBM和WUS2的表达水平用作对照。

[0091] 图10:如实施例14中设计的靶向ZmBBM和ZmWUS2的向导RNA序列。

[0092] 表1:序列表中公开序列的简要说明

[0093]

序列标识符 [SEQ ID NO]:	说明	序列标识符 [SEQ ID NO]:	说明
1-3	靶向玉米 BBM 的启动子区域的 Cas9 的 gRNA	277	5xGS 接头
4-6	靶向玉米 WUS 的启动子区域的 Cas9 的 gRNA		
7-9	靶向玉米 BBM 的启动子区域的 Cpf1 的 crRNA	278	质粒 pKWS20 的序列
10-12	靶向玉米 WUS 的启动子区域的 Cpf1 的 crRNA		
13-51	靶向玉米 BBM 的启动子区域的 TAL 识别结构域	279	表达质粒 pGEP754 的序列
52-94	靶向玉米 WUS 的启动子区域的 TAL 识别结构域	280	表达质粒 pGEP755 的序列
95	玉米 BBM 的靶启动子区域	281	表达质粒 pGEP756 的序列
96	玉米 WUS 的靶启动子区域	282	野生型 LbCpf1
97-99	玉米 BBM 的启动子区域中 Cas9 的 gRNA 的靶位点	283	LbCpf1 的 RR 变体
100-102	玉米 BBM 的启动子区域中 Cpf1 的 crRNA 的靶位点	284	LbCpf1 的 RVR 变体
103-105	玉米 WUS 的启动子区域中 Cas9 的 gRNA 的靶位点	285	表达质粒 pGEP767 的序列
106-108	玉米 WUS 的启动子区域中 Cpf1	286	表达质粒 pGEP772 的序列

[0094]

	的 crRNA 的靶位点		
109-147	玉米 BBM 的启动子区域中 TAL 效应物的靶位点	287	表达质粒 pGEP761 的序列
148-190	玉米 WUS 的启动子区域中 TAL 效应物的靶位点	288	dLbCpf1-VPR
191-198	引物	289	dLbCpf1(RR)-VPR
199-216	来自不同物种的多种形态发生基因的 cDNA	290	dLbCpf1(RVR)-VPR
217-237	来自玉米的多种形态发生基因的 cDNA	291-294	靶向 WUS2 的 gRNA
238-258	来自不同物种的多种形态发生基因的氨基酸序列		
259-269	编码激活结构域或其部分的各种示例性核苷酸序列	295-298	靶向 BBM 的 gRNA
270-272	BBM 靶序列		
273	表达质粒 pGEP362 的序列	299-306	gRNA 的表达质粒
274	表达质粒 pGEP487 的序列	307	玉米 BBM
275	表达质粒 pGEP488 的序列	308	玉米 WUS2
276	VPR 转录激活结构域		

[0095] 定义

[0096] 术语“位点特异性DNA修饰酶”、“序列特异性DNA修饰酶”、“基因编辑酶”、“基因组编辑酶”和“基因组工程化酶”在本文中可互换使用,是指用于在至少一个位置对活生物的任何遗传或表观遗传信息或基因组进行靶向特异性修饰或靶向随机修饰的酶或酶复合物。酶的序列特异性性质是指,它们可以被靶向以编辑基因,还可以编辑除基因组的基因编码区域以外的区域。它还包括细胞的细胞核(如果存在)以及其他遗传信息的编辑或工程化。此外,遗传信息的修饰包括对包含在核或核外基因组(包括DNA或RNA基因组)中的核酸碱基进行编辑、工程改造、突变或破坏的靶向修饰。它还可以包括从基因组表达的信息的靶向修饰,例如RNA信息。此类酶包括但不限于,核酸外切酶、核酸内切酶、切口酶、解旋酶、聚合酶、连接酶和脱氨酶,包括胞苷、腺嘌呤或其他碱基编辑器。表观遗传信息的修饰包括甲基化、组蛋白修饰或可能导致基因表达遗传变化的非编码RNA的靶向修饰。

[0097] 如本文所使用的,“碱基编辑器”是指包含至少一种蛋白或其片段的蛋白或复合物,其具有介导靶向碱基修饰的能力,即,目的碱基的转换(conversion)导致目的点突变。优选地,在本发明的上下文中,至少一个碱基编辑器包括用于将碱基编辑器靶向核酸序列的特异性位点的至少一个核酸识别结构域和在特异性靶位点执行至少一个核碱基转换的至少一个核酸编辑结构域。该核酸识别结构域还可包含至少一个核酸分子(例如,向导RNA)或任何其他单链或双链核酸分子。因此,“碱基编辑”是指携带不同于之前的核碱基的至少一个特异性核苷酸。基于以上所述,根据本发明的“预定位置”是指细胞系统的基因组材料中或待修饰的目的细胞的基因组内的位置或位点,其中将引入靶向编辑。除核酸识别结构域和核酸编辑结构域外,碱基编辑器还可包含除核酸识别结构域和核酸编辑结构域之外的其它组分,例如间隔子、定位信号和抑制自然发生的DNA或RNA修复机制的组分,以确保期望的编辑结果。术语“核酸识别结构域”是指碱基编辑器的组分,其通过将碱基编辑器引导至预定位置内的靶位点来确保碱基编辑器的位点特异性。利用向导RNA(gRNA)或单链向导RNA(sgRNA)特异性识别细胞系统的核酸分子内的靶序列的核酸识别结构域可以是CRISPR RNA(crRNA)和反式激活crRNA(tracrRNA)的合成融合体。

[0098] 如本文所使用的,“CRISPR核酸酶”是已经在天然CRISPR系统中鉴定出的任何核酸酶,其随后已从其天然环境中分离出来,并且优选地已被修饰或组合成目的重组构建体以适合作为靶向基因组工程的工具。只要原始野生型CRISPR核酸酶提供DNA识别,即结合特性,就可以使用并可选地重编程或突变以适合于根据本发明的各种实施方案。所述DNA识别可以是PAM(前间区临近基序)依赖性的。具有优化和工程设计的PAM识别模式的CRISPR核酸酶可用于特定应用并创建。PAM识别代码的扩展可适合于将位点特异性效应物复合物靶向目标靶位点,而与野生型基于CRISPR的核酸酶的原始PAM特异性无关。CRISPR核酸酶还包含天然存在的CRISPR效应物序列或编码它们的相应序列的突变体或催化活性片段或融合体。CRISPR核酸酶尤其还可以指,在其天然环境中具有核酸内切酶功能的CRISPR多肽的CRISPR切口酶或甚至核酸酶缺陷变体。

[0099] 术语“核酸编辑结构域”是指碱基编辑器的组件,其启动核苷酸转换以产生所需的编辑。核酸编辑结构域的催化功能可以是胞苷脱氨酶或腺嘌呤脱氨酶功能。

[0100] 通常,碱基编辑器由至少一个核酸识别结构域和使胞苷或腺嘌呤脱氨基的至少一个核酸编辑结构域组成。使胞苷脱氨基的核酸编辑结构域能够将C转换为T(G转换为A),它

们被称为BE;使腺嘌呤脱氨基的核酸编辑结构域可以将A转换为G(T转换为C),它们被称为ABE。

[0101] 碱基编辑器通常由胞苷脱氨酶结构域(例如,APOBEC1、APOBEC3A、APOBEC3G、PmCDA1、AID)、接头(通常为XTEN)、CRISPR结构域(d/nCas9、dCpf1、CasX、CasY或其他合适的结构域)和尿嘧啶DNA糖基化酶抑制剂(UGI)组成。在经修饰的系统中,UGI结构域或NLS的数量可以变化,接头的长度也可以变化。其还可包括其他结构域,例如,Gam(例如,在BE4中)。存在胞苷脱氨酶结构域中可能存在具有氨基酸点突变的变体,例如YE-BE3、YEE-BE3,其用于不同的编辑窗口,也可能存在用于不同PAM识别的CRISPR结构域突变的变体,例如VQR-BE3、EQR-BE3、VRER-BE3和SaKKH-BE3。在BE-PLUS系统中,CRISPR结构域和胞苷脱氨酶结构域不被表达为融合蛋白,而是使用Suntag系统连接在一起以扩大编辑范围。有关优选碱基编辑器的更多详细信息,包括基于胞苷脱氨酶的DNA碱基编辑器、基于腺嘌呤脱氨酶的DNA碱基编辑器,可以从Eid A等人那里获得(Ayman Eid,Sahar Alshareef和Magdy M.Mahfouz (2018),CRISPR碱基编辑器:无双链断裂的基因组编辑(CRISPR base editors:genome editing without double-strand breaks),《生物化学杂志》(Biochemical Journal) (2018) 475 1955-1964)。

[0102] 根据本公开,术语“与……缔合”应被广义地解释,因此,根据本发明暗示了,提供一个分子(包含天然存在和/或合成的构建基块的DNA、RNA、氨基酸)与另一个分子物理缔合,该缔合具有共价或非共价性质。例如,修复模板可以与CRISPR核酸酶的gRNA缔合,其中该缔合可具有非共价性质(互补碱基配对),或者该分子可以通过共价键彼此物理连接。

[0103] 如本文所用的,术语“催化活性片段”是指氨基酸序列,是指衍生自给定模板氨基酸序列的核心序列或编码其的核酸序列,其包含模板序列的全部或部分活性位点,前提是所得到的催化活性片段仍然具有表征模板序列的活性,天然酶或其变体的活性位点负责该模板序列。所述修饰适合于产生较小体积的氨基酸序列,该氨基酸序列仍具有与模板序列相同的活性,从而使催化活性片段成为对空间要求较低的更多功能或更稳定的工具。

[0104] “共价连接”或“共价键”是化学键,涉及彼此共价连接的分子或序列的原子之间电子对的共享。“非共价”相互作用与共价键的不同之处在于它不涉及电子共享,而是涉及分子/序列之间或分子/序列内电磁相互作用的更分散的变化。因此,非共价相互作用或附着包括静电相互作用、范德华力、 π 效应和疏水效应。在核酸分子的上下文中特别重要的是作为静电相互作用的氢键。氢键(H键)是偶极-偶极相互作用的一种特定类型,其涉及部分正氢原子与未共价键合至所述氢原子的高负电性,部分负氧,氮,硫或氟原子之间的相互作用。因此,本文所用的任何“缔合”或“物理缔合”表示共价或非共价相互作用或附着。对于分子复合物,例如,由CRISPR核酸酶、gRNA和修复模板(RT)形成的复合物,可以存在更多的共价和非共价相互作用,以连接并缔合目标分子复合物的不同成分。

[0105] 术语“CRISPR多肽”、“CRISPR核酸内切酶”、“CRISPR核酸酶”、“CRISPR蛋白”、“CRISPR效应物”或“CRISPR酶”在本文中可互换使用,是指任何天然存在的或人工的氨基酸序列,或编码该序列的核酸序列,作为位点特异性DNA核酸酶或切口酶,其中,“CRISPR多肽”来自任何生物的CRISPR系统,其可以克隆并用于靶向基因组工程。术语“CRISPR核酸酶”或“CRISPR多肽”还包含天然存在的CRISPR效应物序列或编码该序列的相应序列的突变体或催化活性片段或融合体。因此,“CRISPR核酸酶”或“CRISPR多肽”也可以是指例如在其天然环

境中具有核酸内切酶功能的CRISPR多肽的CRISPR切口酶或甚至核酸酶缺陷变体。优选地，本发明的公开内容依赖于核酸酶缺陷型CRISPR核酸酶，其仍然具有其固有的由同源CRISPR RNA辅助的DNA识别和结合特性。

[0106] 本文公开的核酸序列可以是“密码子优化的”。“密码子优化”表示，从供体生物合成产生或分离的DNA或RNA适应于不同受体生物的密码子使用，以提高转录速率、mRNA加工和/或稳定性，以及/或翻译速率，和/或所述重组核酸在目的细胞或生物体中的后续蛋白质折叠。技术人员充分意识到以下事实：由于密码子简并性，靶核酸可以在一个位置被修饰，而这种修饰在翻译后仍会在该位置导致相同的氨基酸序列，这是通过密码子优化实现的，同时考虑了靶细胞或生物体的物种特定密码子使用情况。继而，本文定义的核酸序列可与编码相同蛋白质但已被密码子优化的不同序列具有一定程度的相同性。

[0107] 如本文所用的，“互补”或“互补性”描述了两个(c)DNA之间、两个RNA之间或RNA与(c)DNA核酸区域之间的关系。由DNA或RNA的核碱基定义，两个核酸区域可以根据锁与钥匙模型相互杂交。为此，Watson-Crick碱基配对的原理分别具有碱基腺嘌呤和胸腺嘧啶/尿嘧啶以及鸟嘌呤和胞嘧啶作为互补碱基。此外，本文所用的术语“互补”也包括非Watson-Crick配对，如反式Watson-Crick、Hoogsteen、反式Hoogsteen和Wobble配对，只要各个碱基对可彼此建立氢键，即，基于所述互补性，两条不同的核酸链可以彼此杂交。

[0108] 如在本申请的上下文中使用的，术语“约”可以表示所列数值的 $\pm 10\%$ ，优选地是所列数值的 $\pm 5\%$ 。例如，然后将约100个核苷酸(nt)理解为介于90和110nt之间的值，优选地介于95和105之间。

[0109] 本文在原核或真核细胞，优选动物细胞且更优选植物或植物细胞或植物材料的上下文中使用的术语“衍生物”或“子代”或“后代”涉及此细胞或材料的子代，其由自然生殖繁殖(包括有性和无性繁殖)产生。本领域技术人员众所周知，所述繁殖可以导致由自然现象导致的生物体基因组突变，所述自然现象产生的子代或后代在基因组上与亲本有机体或细胞不同，其仍然属于同一属/种，并具有与亲本重组宿主细胞基本相同的特性。由繁殖或再生过程中的自然现象产生的此类衍生物或子代或后代因此被本公开的术语所涵盖，并且当将“衍生物”或“子代”或“后代”与各自的父母或祖先进行比较时，本领域技术人员可以容易地识别。此外，在物质或核酸或氨基酸分子的上下文中且不指复制中的细胞或生物体的术语“衍生物”可表示通过化学和/或生物技术手段衍生自原始物质或分子的物质或分子。所得的衍生物将具有允许技术人员清楚地定义该衍生物源自的原始或母体分子的特征。此外，该衍生物可能具有其他或不同的生物学功能，但原始分子的衍生物或“活性片段”仍将共享母体分子的至少一个生物学功能，及时该衍生物或活性片段可能比亲本序列短/长，并且与各个亲本序列相比可包含某些突变、缺失或插入。

[0110] 如本文所用的，“真核细胞”是指具有真核、核膜和细胞器的细胞，所述细胞属于原生生物界、植物界、真菌界或动物界中的任何一个。真核生物体可包括单细胞和多细胞生物体。根据本发明的优选真核细胞和生物体是植物细胞。

[0111] 如本文所用的，“融合体”可以指包含一个或多个非天然序列(例如，部分)的蛋白质和/或核酸。因此，可以以融合分子的形式提供根据本发明的任何核酸序列或氨基酸序列。融合体可以在修饰蛋白的N-末端或C-末端，或两者兼而有之，也可以在分子内作为单独的结构域。对于核酸分子，融合分子可以连接在5'或3'端，或两者之间的任何合适位置。融

合体可以是转录和/或翻译融合体。融合体可包含一个或多个相同的非天然序列。融合体可包含一个或多个不同的非天然序列。融合体可以是嵌合体。融合体可包含核酸亲和标签。融合体可包含条形码。融合体可包含肽亲和标签。融合体可以提供如本文所公开的至少一个合成转录因子的亚细胞定位(例如,用于靶向(例如,位点特异性核酸酶)至细胞核的核定位信号(NLS)、用于靶向线粒体的线粒体定位信号、用于靶向叶绿体的叶绿体定位信号、内质网(ER)滞留信号等)。融合体可以提供可用于追踪或纯化的非天然序列(例如,亲和标签)。融合体可以是小分子,例如生物素;也可以是染料,例如alexa fluor染料、Cyanine3染料、Cyanine5染料。融合体可以提高或降低稳定性。在一些实施方案中,融合体可包含可检测标记,包括可提供可检测信号的部分。可提供可检测信号的合适的可检测标记和/或部分可包括但不限于,酶、放射性同位素、特异性结合对的成员;荧光基团;荧光报告因子或荧光蛋白;量子点等。融合体可包含FRET对的成员,或荧光基团/量子点供体/受体对。融合体可包含酶。合适的酶可包括但不限于,辣根过氧化物酶、荧光素酶、 β -25半乳糖苷酶等。融合体可包含荧光蛋白。合适的荧光蛋白可包括但不限于,绿色荧光蛋白(GFP)(例如,来自维多利亚多管发光水母(*Aequoria victoria*)的GFP、来自鳗鲡(*Anguilla japonica*)的荧光蛋白,或其突变体或衍生物)、红色荧光蛋白、黄色荧光蛋白、黄绿色荧光蛋白(例如,衍生自cephalochordate *Branchiostoma lanceolatum*的四聚体荧光蛋白的mNeonGreen),其是多种荧光和彩色蛋白中的任何一种。融合体可包含纳米颗粒。合适的纳米颗粒可包括荧光或发光纳米颗粒,以及磁性纳米颗粒或纳米金刚石,其任选地连接至纳米颗粒。可以检测纳米颗粒的任何光学或磁力特性或特征。融合体可包含解旋酶、核酸酶(例如,FokI)、核酸内切酶、核酸外切酶(例如,5'核酸外切酶和/或3'核酸外切酶)、连接酶、切口酶、核酸酶-解旋酶(例如,Cas3)、DNA甲基转移酶(例如,Dam)或DNA去甲基化酶、组蛋白甲基转移酶、组蛋白去甲基化酶、乙酰基转移酶(包括例如但不限于,组蛋白乙酰基转移酶)、去乙酰化酶(包括例如但不限于,组蛋白去乙酰化酶)、磷酸酶、激酶、转录(共)激活子、转录(共)因子、RNA聚合酶亚基、转录阻遏物、DNA结合蛋白、DNA结构蛋白、长非编码RNA、DNA修复蛋白(例如,涉及单链和/或双链断裂修复的蛋白,例如,参与碱基切除修复、核苷酸切除修复、错配修复、NHEJ、HR、微同源介导的末端连接(MMEJ)和/或替代性非同源末端连接(ANHEJ)的蛋白,例如但不限于,HR调节子和HR复合物组装信号)、标记蛋白、报告蛋白、荧光蛋白、配体结合蛋白(例如,mCherry或重金属结合蛋白)、信号肽(例如,Tat信号序列)、靶向蛋白或肽、亚细胞定位序列(例如,核定位序列、叶绿体定位序列)和/或抗体表位或其任何组合。

[0112] 如本文所用的,“基因”是指编码基因产物的DNA区域,以及调控该基因产物的产生的所有DNA区域,无论这些调控序列是否与编码和/或转录序列相邻。因此,基因包括但不限于,启动子序列、终止子、翻译调控序列(如核糖体结合位点和内部核糖体进入位点)、增强子、沉默子、绝缘子、边界元件、复制起始点、基质附着位点和基因座控制区域。

[0113] 本文所用的术语“基因表达”或“表达”是指将基因中包含的信息转换为“基因产物”。“基因产物”可以是基因的直接转录产物(例如,mRNA、tRNA、rRNA、反义RNA、核酶、结构RNA或任何其他类型的RNA),或是通过翻译mRNA产生的蛋白。基因产物还包括通过诸如加帽、多聚腺苷酸化、甲基化和编辑等过程修饰的RNA,以及通过例如甲基化、乙酰化、磷酸化、泛素化、ADP-核糖基化、肉豆蔻酰化(myristylation)和糖基化修饰的蛋白。

[0114] 术语“基因激活”或“基因表达(的)增强/激活/上调”是指导致基因产物产量增加

的任何过程。基因产物可以是RNA (包括但不限于mRNA、rRNA、tRNA和结构RNA) 或蛋白。因此, 基因激活包括那些提高基因转录和/或mRNA翻译的过程。提高转录的基因激活过程的实例包括但不限于, 促进转录起始复合物形成的那些、提高转录起始速率的那些、提高转录持续性的那些, 以及减轻转录阻遏 (例如通过阻断转录阻遏物的结合) 的那些。基因激活可以构成例如对阻遏的抑制, 以及在现有水平之上表达的刺激。提高翻译的基因激活过程的实例包括提高翻译起始的那些、提高翻译延伸的那些, 以及提高mRNA稳定性的那些。通常, 基因激活包括基因产物产量的任何可检测增加, 优选地为基因产物的产量增加约2倍, 更优选地为约2倍至约5倍或它们之间的任何整数值, 更优选地为约5倍至约10倍或它们之间的任何整数值, 更优选地为约10倍至约20倍或它们之间的任何整数值, 再更优选地为约20倍至约50倍或它们之间的任何整数值, 更优选地为约50倍至约100倍或它们之间的任何整数值, 更优选地为100倍或以上。

[0115] 相反, 术语“基因阻遏”或“抑制/阻遏/沉默/下调基因表达”是指导致基因产物产量减少的任何过程。基因产物可以是RNA (包括但不限于, mRNA、rRNA、tRNA和结构RNA) 或蛋白。因此, 基因阻遏包括减少基因转录和/或mRNA翻译的那些过程。减少转录的基因阻遏过程的实例包括但不限于, 抑制转录起始复合物形成的那些、降低转录起始速率的那些、降低转录伸长率的那些、降低转录持续性的那些, 以及拮抗转录激活 (例如通过阻断转录激活子的结合) 的那些。基因阻遏可以造成例如防止激活, 以及将表达抑制在现有水平之上。降低翻译的基因阻遏过程的实例包括但不限于, 降低翻译起始的那些、降低翻译延伸的那些, 以及降低mRNA稳定性的那些。转录阻遏包括基因转录的可逆和不可逆失活。通常, 基因阻遏包括基因产物产量的任何可检测的减少, 优选地基因产物地产量减少约2倍, 更优选地为约2倍至约5倍或它们之间的任何整数值, 更优选地为约5倍至约10倍或它们之间的任何整数值, 更优选地为约10倍至约20倍或它们之间的任何整数值, 再更优选地为约20倍至约50倍或它们之间的任何整数值, 更优选地为约50倍至约100倍或它们之间的任何整数值, 更优选地为100倍或以上。最优选地, 基因阻遏导致基因表达的完全抑制, 因此没有可检测的基因产物。

[0116] 本文使用的术语“遗传构建体”或“重组构建体”、“载体”或“质粒(载体)” (例如, 在待引入细胞系统中的至少一个核酸序列的上下文中) 是指包括以下内容的构建体: 质粒或(质粒) 载体、粘粒、人工酵母染色体或细菌人工染色体(YAC和BAC)、噬菌粒、基于细菌噬菌体的载体、表达盒、分离的单链或双链核酸序列(包括DNA和RNA序列(呈线性或环状形式) 或氨基酸序列)、病毒载体(包括修饰的病毒), 及其组合或混合物, 用于引入或转化、转染或转导至根据本发明的任何原核或真核靶细胞(包括植物、植物细胞、器官或材料)。因此, 在生物材料(如细胞或载体) 中的“重组”是指人工生产的材料。根据本公开的重组构建体可包含核酸或氨基酸序列形式的效应物结构域, 其中, 效应物结构域代表可以在靶细胞中发挥作用的分子, 并且包括转基因、单链或双链RNA分子(包括向导RNA((s) gRNA)、miRNA或siRNA) 或氨基酸序列, 尤其包括酶或其催化活性片段、结合蛋白、抗体、转录因子、核酸酶(优选位点特异性核酸酶) 等。此外, 重组构建体可包含调控序列和/或定位序列。重组构建体可以整合到载体(包括质粒载体) 中, 并且/或者它可以从载体结构中分离出来, 例如以多肽序列的形式或作为非载体连接的单链或双链核酸形式存在。引入后, 例如通过生物或物理方法进行转化或转染, 遗传构建体可以在染色体外存在, 即, 未整合到靶细胞的基因组中, 例如以

双链或单链DNA、双链或单链RNA形式或作为氨基酸序列存在。可替换地,根据本公开的遗传构建体或其部分可以稳定地整合到靶细胞的基因组中,包括靶细胞的核基因组或其它遗传元件,包括质体(如线粒体或叶绿体)的基因组。在此使用的术语质粒载体是指最初从质粒获得的遗传构建体。质粒通常是指呈双链核酸序列形式的循环自主复制染色体外元件。在基因工程领域中,通过插入例如编码对抗生素或除草剂具有抗性的基因,以及编码靶核酸序列、定位序列、调控序列、标签序列、标记基因(包括抗生素标记或荧光标记)、可选编码的序列、易于识别的序列等的基因等,对这些质粒进行常规修饰。原始质粒的结构组分(如复制起点)得以保留。根据本发明的某些实施方案,定位序列可包括核定位序列(NLS)、质体定位序列,优选地为线粒体定位序列或叶绿体定位序列。所述定位序列对于植物生物技术领域的技术人员而言是可用的。用于不同目的靶细胞的多种质粒载体是可商购的,并且其修饰是相应领域技术人员公知的。

[0117] 本文所用的“基因组”既包括基因(编码区域)、非编码DNA,也包括线粒体和/或叶绿体的遗传物质(如果存在)或编码病毒或病毒一部分的基因组材料。生物体的“基因组”或“遗传物质”通常由DNA组成,其中,病毒的基因组可以由RNA(单链或双链)组成。

[0118] 术语“基因组编辑”、“基因编辑”和“基因组工程”在本文可互换使用,并且是指用于在至少一个位置上对活生物的任何遗传信息或基因组进行靶向特异性修饰的策略和技术。因此,这些术语既包括基因编辑,也包括对除基因组的基因编码区域以外的区域的编辑。它还包括细胞核(如果存在)以及细胞其他遗传信息的编辑或工程化。此外,术语“基因组编辑”、“基因编辑”和“基因组工程”还包括表观遗传编辑或工程化,即,例如非编码RNA进行靶向修饰,例如甲基化、组蛋白修饰,可能引起基因表达的遗传变化。

[0119] 如本文所用的,“种质”是用于描述遗传资源或更准确地描述生物体的DNA和该物质的集合的术语。在育种技术中,“种质”一词用来表示可以产生新植物或植物品种的遗传物质的集合。

[0120] 术语“向导RNA”、“gRNA”、“CRISPR核酸序列”、“单链向导RNA”或“sgRNA”在本文中可互换使用,并且是指CRISPR RNA(crRNA)和反式激活crRNA(tracrRNA)的合成融合体,或者该术语是指仅由crRNA和/或tracrRNA组成的单个RNA分子,或者该术语是指只包含crRNA或tracrRNA部分的gRNA。因此,tracr和crRNA部分(如果按相应的CRISPR多肽的需要存在)不一定必须存在于一个共价连接的RNA分子上,但是它们也可以由两个单独的RNA分子组成,它们可以缔合或可以通过非共价或共价相互作用而缔合,以提供根据本发明的gRNA。例如,在单个RNA引导的核酸内切酶(如Cpf1)的情况下(参见Zetsche等,2015),crRNA作为单链向导核酸序列可足以介导DNA靶向。

[0121] 本文使用的术语“杂交”是指互补核酸(即,DNA和/或RNA)的配对,其使用核酸链通过碱基配对与互补链结合以形成杂交复合物的任何方法实现。杂交和杂交强度(即核酸之间缔合的强度)受以下因素影响:核酸之间的互补性程度和长度、所涉及条件的严格性、所形成的杂交体的 T_m ,以及核酸中的G:C比。术语杂交复合物是指由于互补的G和C碱基之间以及互补的A和T/U碱基之间的氢键的形成而在两个核酸序列之间形成的复合物。可以在两个DNA核酸分子之间、两个RNA核酸分子之间,或DNA和RNA核酸分子之间形成杂交复合物或相应的杂交构建体。对于所有基因座,核酸分子可以是在体外或体内产生的天然存在的核酸分子和/或人工或合成的核酸分子。如上所述的杂交,例如可以在DNA、RNA和DNA/RNA序列之

间形成的Watson-Crick碱基对,是由特定的氢键模式决定的,因此该氢键模式表示根据本发明的非共价连接形式。在杂交的情况下,术语“严格杂交条件”应理解是指主要仅在同源核酸分子之间发生杂交的那些条件。在这方面,术语“杂交条件”不仅是指核酸实际团聚过程中普遍存在的实际条件,而且还指后续洗涤步骤期间普遍存在的条件。严格杂交条件的实例是主要仅那些具有至少70%,优选至少75%、至少80%、至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.50%序列相同性的核酸分子进行杂交。严格的杂交条件为:例如,在65℃下4xSSC,随后在65℃下在0.1xSSC下多次洗涤约1小时。如本文所用的,术语“严格杂交条件”还可以意指:在68℃下在0.25M磷酸钠、pH 7.2、7% SDS、1mM EDTA和1% BSA中杂交16小时,随后用2×SSC和0.1% SDS在68℃下洗涤两次。优选地,杂交在严格条件下进行。

[0122] 术语“形态发生”和“形态形成”在本文中通常在基因的上下文中可互换使用,其中,由所述基因编码的基因产物参与形态发生,即,引起生物体形成其形状的生物过程。该术语还用于直接或间接参与细胞或生物形态发生过程的任何因素,包括合成或天然存在的转录因子。此外,这些术语在导致整个植物再生的细胞途径的上下文中使用。

[0123] 关于序列或分子的术语“核苷酸”和“核酸”在本文可互换使用,是指天然或合成来源的单链或双链DNA或RNA。因此,术语核苷酸序列用于任何与其长度无关的DNA或RNA序列,因此该术语包括任何核苷酸序列,该核苷酸序列包含至少一个核苷酸,但也包括任何种类的较大的寡核苷酸或多聚核苷酸。因此,该术语是指天然和/或合成的脱氧核糖核酸(DNA)和/或核糖核酸(RNA)序列,其可以任选地包含合成的核酸类似物。根据本公开的核酸可以任选地被密码子优化。密码子优化意味着,DNA或RNA的密码子使用适于目的细胞或生物体的密码子使用,以提高所述重组核酸在目的细胞或生物体中的转录速率。技术人员应充分意识到以下事实:由于密码子简并性,靶核酸可以在一个位置被修饰,而这种修饰在翻译后仍会在该位置导致相同氨基酸序列,这是通过密码子优化考虑靶细胞或生物体的物种特异性密码子使用实现的。根据本申请的核酸序列可对以下非限制性生物体列表进行特定的密码子优化:大麦(*Hordeum vulgare*)、高粱(*Sorghum bicolor*)、黑麦(*Secale cereale*)、黑小麦(*Triticale*)、甘蔗(*Saccharum officinarum*)、玉米(*Zea mays*)、小米(*Setaria italica*)、水稻(*Oryza sativa*)、小粒野生稻(*Oryza minuta*)、澳洲野生稻(*Oryza australiensis*)、高秆野生稻(*Oryza alta*)、小麦(*Triticum aestivum*)、硬粒小麦(*Triticum durum*)、球茎大麦(*Hordeum bulbosum*)、二穗短柄草(*Brachypodium distachyon*)、滨海大麦草(*Hordeum marinum*)、节节麦(*Aegilops tauschii*)、苹果(*Malus domestica*)、甜菜(*Beta vulgaris*)、向日葵(*Helianthus annuus*)、胡萝卜(*Daucus glochidiatus*)、小胡萝卜(*Daucus pusillus*)、*Daucus muricatus*、野胡萝卜(*Daucus carota*)、巨桉(*Eucalyptus grandis*)、*Erythranthe guttata*、*Genlisea aurea*、美花烟草(*Nicotiana sylvestris*)、普通烟草(*Nicotiana tabacum*)、绒毛状烟草(*Nicotiana tomentosiformis*)、本氏烟(*Nicotiana benthamiana*)、番茄(*Solanum lycopersicum*)、马铃薯(*Solanum tuberosum*)、中果咖啡(*Coffea canephora*)、葡萄(*Vitis vinifera*)、黄瓜(*Cucumis sativus*)、川桑(*Morus notabilis*)、拟南芥(*Arabidopsis thaliana*)、琴叶拟南芥(*Arabidopsis lyrata*)、*Arabidopsis arenosa*、近缘种须弥芥(*Crucihimalaya himalaica*)、卵叶须弥芥(*Crucihimalaya wallichii*)、弯曲碎米荠(*Cardamine*

flexuosa)、北美独行菜 (*Lepidium virginicum*)、荠菜 (*Capsella bursa-pastoris*)、*Olmrabidopsis pumila*、硬毛南芥 (*Arabis hirsuta*)、甘蓝型油菜 (*Brassica napus*)、甘蓝 (*Brassica oleracea*)、芜菁 (*Brassica rapa*)、芥菜 (*Brassica juncacea*)、黑芥菜 (*Brassica nigra*)、萝卜 (*Raphanus sativus*)、芝麻菜 (*Eruca vesicaria sativa*)、甜橙 (*Citrus sinensis*)、麻疯树 (*Jatropha curcas*)、大豆 (*Glycine max*)、*Gossypium ssp.* 或毛果杨 (*Populus trichocarpa*)。

[0124] 如本文所用的,“非天然的”或“非天然存在的”或“人工的”或“合成的”可指未在天然核酸或蛋白中发现的核酸或多肽序列,或任何其他生物分子(如生物素或荧光素)。非天然可以指代亲和标签。非天然可以指融合。非天然可以指包含突变、插入和/或缺失的天然存在的核酸或多肽序列。非天然序列可以显示和/或编码也可以通过与非天然序列融合的核酸和/或多肽序列显示的活性(例如,酶活性、转移酶活性、乙酰基转移酶活性、激酶活性、泛素化酶活性等)。非天然核酸或多肽序列可以通过基因工程与天然存在的核酸或多肽序列(或其变体)连接,以产生嵌合核酸,和/或编码嵌合核酸和/或多肽的多肽序列。非天然序列可以指3'杂交延伸序列,或附着于分子的核定位信号(NLS)。因此,本文所用的“合成转录因子”是指包含并非自然界中天然存在的至少两个结构域—识别结构域和激活结构域分子。

[0125] 本文所用的“生物体”是指单个的真核或原核生命形式,尤其包括动物、植物、真菌或单细胞生命形式。在本发明的上下文中,生物体优选地是植物或植物的一部分。

[0126] 如本文所用的,术语“粒子轰击”,也称为“生物弹转染”或“生物弹轰击”或“微粒介导的基因转移”,是指用于将包含目的核酸或遗传构建体的包被微粒或纳米颗粒转移到靶细胞或组织中的物理递送方法。微粒或纳米颗粒起射弹的作用,并使用合适的装置(通常称为“基因枪”)在高压下发射到目的靶结构上。通过粒子轰击进行的转化使用覆盖有目的基因的金属微粒,然后使用称为“基因枪”的设备(Sandford等,1987)以足以穿透靶组织的细胞壁但不足以造成细胞死亡的力道和速度将其发射到靶细胞上。对于原生质体,其细胞壁被完全去除,其条件在逻辑上是不同的。至少一个微粒上的沉淀的核酸或遗传构建体在轰击后释放到细胞中,并整合到基因组中,或根据上述定义瞬时表达。微粒的加速是通过高压放电或压缩气体(氦气)来实现的。关于所用的金属颗粒,其必须无毒、无反应,并且其直径小于靶细胞的直径。最常用的是金或钨。从基因枪及涉及到其一般用途的相关系统的制造商和供应商可以公开获得大量信息。

[0127] 本文所用的术语“植物”或“植物细胞”是指植物生物体、植物器官、分化和未分化的植物组织、植物细胞、种子及其衍生物和后代。植物细胞包括但不限于,例如来自种子的细胞、来自成熟和未成熟细胞或器官的细胞,该成熟和未成熟细胞或器官包括处于不同分化状态的胚、分生组织、幼苗、愈伤组织、叶、花、根、枝、雄或雌配子体、孢子体、花粉、花粉管和小孢子、原生质体、大型藻类和微藻类。不同的真核细胞(例如,植物细胞)可具有任何程度的倍性,即,它们可以是单倍体、二倍体、四倍体、六倍体或多倍体。优选地,本文所使用的植物细胞、植物或植物的一部分源自或属于选自以下各项所构成的组的植物物种:大麦 (*Hordeum vulgare*)、球茎大麦 (*Hordeum bulbosom*)、高粱 (*Sorghum bicolor*)、甘蔗 (*Saccharum officinarium*)、玉米 (*Zea mays*)、小米 (*Setaria italica*)、小粒野生稻 (*Oryza minuta*)、水稻 (*Oryza sativa*)、澳洲野生稻 (*Oryza australiensis*)、高秆野生稻

(*Oryza alta*)、小麦(*Triticum aestivum*)、黑麦(*Secale cereale*)、苹果(*Malus domestica*)、二穗短柄草(*Brachypodium distachyon*)、滨海大麦草(*Hordeum marinum*)、节节麦(*Aegilops tauschii*)、胡萝卜(*Daucus glochidiatus*)、甜菜(*Beta vulgaris*)、小胡萝卜(*Daucus pusillus*)、*Daucus muricatus*、野胡萝卜(*Daucus carota*)、巨桉(*Eucalyptus grandis*)、美花烟草(*Nicotiana sylvestris*)、绒毛状烟草(*Nicotiana tomentosiformis*)、普通烟草(*Nicotiana tabacum*)、番茄(*Solanum lycopersicum*)、马铃薯(*Solanum tuberosum*)、中果咖啡(*Coffea canephora*)、葡萄(*Vitis vinifera*)、*Erythrante guttata*、*Genlisea aurea*、黄瓜(*Cucumis sativus*)、川桑(*Morus notabilis*)、*Arabidopsis arenosa*、琴叶拟南芥(*Arabidopsis lyrata*)、拟南芥(*Arabidopsis thaliana*)、近缘种须弥芥(*Crucihimalaya himalaica*)、卵叶须弥芥(*Crucihimalaya wallichii*)、弯曲碎米荠(*Cardamine flexuosa*)、北美独行菜(*Lepidium virginicum*)、芥菜(*Capsella bursa pastoris*)、*Olm arabidopsis pumila*、硬毛南芥(*Arabis hirsute*)、甘蓝型油菜(*Brassica napus*)、甘蓝(*Brassica oleracea*)、芜菁(*Brassica rapa*)、萝卜(*Raphanus sativus*)、芥菜(*Brassica juncea*)、黑芥菜(*Brassica nigra*)、芝麻菜(*Eruca vesicaria subsp. sativa*)、甜橙(*Citrus sinensis*)、麻疯树(*Jatropha curcas*)、毛果杨(*Populus trichocarpa*)、蒺藜苜蓿(*Medicago truncatula*)、*Cicer yamashitae*、*Cicer bijugum*、鹰嘴豆(*Cicer arietinum*)、*Cicer reticulatum*、*Cicer judaicum*、木豆(*Cajanus cajanifolius*)、蔓草虫豆(*Cajanus scarabaeoides*)、菜豆(*Phaseolus vulgaris*)、大豆(*Glycine max*)、紫云英(*Astragalus sinicus*)、百脉根(*Lotus japonicas*)、夏堇(*Torenia fournieri*)、洋葱(*Allium cepa*)、大葱(*Allium fistulosum*)、大蒜(*Allium sativum*)和韭菜(*Allium tuberosum*)。

[0128] “启动子”是指能够控制编码序列(即,基因或其部分)或功能性RNA(即,不经翻译而具有活性的RNA)(例如miRNA、siRNA、反向重复RNA或发夹形成RNA)的表达的DNA序列。启动子通常位于基因的5'部分。启动子结构存在于整个生物界中,即,存在于细菌、古细菌和真核生物中,它们具有不同的结构。启动子序列通常由相对于调控序列的近端和远端元件组成,后者通常被称为增强子。启动子可具有广谱活性,但是它们也可具有组织或发育阶段特异性活性。例如,它们可以在根、种子和分生组织细胞等细胞中有活性。启动子可以以组成型的方式有活性,或者可以被诱导。诱导可以被多种环境条件和刺激物刺激。存在强启动子和弱启动子,强启动子可以使调控序列高度转录。通常,启动子受到高度调控。本公开的启动子可包括天然存在于细胞中的内源启动子,或来自另一物种的人工或转基因启动子,或人工或嵌合启动子,即,该组合物中不是天然存在且由不同启动子元件组成的启动子。转录过程始于RNA聚合酶(RNAP)与启动子区域中DNA的结合,而启动子区域紧邻转录起始位点(TSS)。典型的启动子序列被认为包含一些相对于TSS位于特异性位点的序列基序。例如,观察到原核启动子具有两个六聚体基序,其相对于TSS位于-10(Pribnow box)或-35位置处或附近。此外,在-35区域的上游可以有富含AT的UP(上游)元素。原核启动子被 σ 因子识别为转录因子。真核启动子的结构通常更复杂,并且它们具有几个不同的序列基序,例如TATA box、INR box、BRE、CCAAT-box和GC-box(Bucher P.,《分子生物学杂志》(J.Mol.Biol.), 1990年4月20日;212(4):563-78)。真核细胞具有三个RNAP,分别为聚合酶I、II、III。RNAP I产生核糖体RNA(rRNA),RNAP II产生信使RNA(mRNA)和小核RNA(snRNA),RNAP III产生转移

RNA (tRNA)、snRNA和5S-RNA。

[0129] 如本文所用的,术语“调控序列”是指可指导目的核酸序列的转录和/或翻译和/或修饰的核酸或氨基酸序列。调控序列可包括顺式或反式序列。示例性调控序列包含启动子、增强子、终止子、操纵子、转录因子、转录因子结合位点、内含子等。

[0130] 本文所用的术语“终止子”是指位于编码序列下游(即,3'方向)的DNA序列,可包括多聚腺苷酸化信号和其他序列,即,编码能够影响mRNA加工和/或基因表达的调控序列的其它序列。多聚腺苷酸化信号的特征通常在于,它在mRNA前体的3'末端添加了聚A-核苷酸。

[0131] 如本文所用,术语“瞬时”或“瞬时引入”是指根据本公开的至少一个核酸和/或氨基酸序列(优选地并入递送载体和/或重组构建体中)借助或不借助递送载体瞬时引入目标结构,例如植物细胞或细胞系统,其中该至少一个核酸或核苷酸序列在适当的反应条件下引入从而使得该至少一个核酸序列不会整合到目标结构的内源性核酸材料中,基因组作为一个整体使得该至少一个核酸序列不会整合到靶细胞的内源性DNA中。结果,在瞬时导入的情况下,所导入的遗传构建体将不会遗传至目标结构(例如,植物细胞)的后代。该至少一个核酸和/或氨基酸序列或由其转录、翻译、加工、翻译后修饰或复合物构建产生的产物仅是暂时存在的,即以瞬时的方式,以组成型或诱导型形式存在,因此,仅在靶细胞中活跃以在有限的时间内发挥其作用。因此,通过瞬时引入而引入的至少一个序列将不会遗传于细胞的后代。然而,由以瞬时方式引入的至少一个序列或效应物介导的效果可以潜在地遗传给靶细胞的后代。因此,“稳定”的引入意味着将核酸或核苷酸序列整合到目的靶细胞或细胞系统的基因组中,其中,该基因组包括核基因组以及其他细胞器所包含的基因组。

[0132] 如本文在氨基酸或核酸序列的上下文中所使用的术语“变体”旨在表示基本上相似的序列。对于核酸序列,变体包括在天然多聚核苷酸内的一个或多个内部位点缺失和/或添加一个或多个核苷酸,以及/或者在天然多聚核苷酸内的一个或多个位点取代一个或多个核苷酸。如本文所用的,“天然”多聚核苷酸或多肽分别包含天然存在的核苷酸序列或氨基酸序列。对于核酸序列,保守变体包括那些由于遗传密码的简并性而编码与本公开的参考序列相同的氨基酸序列的序列。因此,给定核酸序列的变体也将包括合成衍生的核酸序列,例如通过使用定点诱变产生,但仍编码与参考序列相同的蛋白的那些。通常,本公开的特定多聚核苷酸的变体将与由本节下面部分进一步描述的序列比对程序和参数确定的特定核酸序列具有至少约60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高的序列相同性。

[0133] “变体”氨基酸序列、多肽或蛋白(所述术语在本文中可互换使用)是指通过在天然蛋白的一个或多个内部位点处缺失或添加一个或多个氨基酸,和/或在天然蛋白的一个或多个位点处取代一个或多个氨基酸,而衍生自天然氨基酸序列的氨基酸序列。根据本公开的变体氨基酸序列具有生物活性,即,它们仍然拥有天然蛋白的所需生物活性。本公开的天然氨基酸序列的活性变体将与由本节下面部分进一步描述的序列比对程序和参数确定的天然氨基酸序列的氨基酸序列具有至少约65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高的序列相同性。

[0134] 每当本公开涉及核酸或氨基酸序列彼此的相同性百分比时,这些值都限定为通过用于核酸的EMBOSS Water Pairwise Sequence Alignment(核苷酸)程序(www.ebi.ac.uk/Tools/psa/emboss_water/nucleotide.html)或用于氨基酸序列的EMBOSS Water Pairwise

Sequence Alignment(蛋白)程序(www.ebi.ac.uk/Tools/psa/emboss_water/)获得的那些值。如本文所使用的比对或序列比较是指彼此比较的两个序列在全长上的比对。由欧洲生物信息研究所(European Bioinformatics Institute, EBI)欧洲分子生物学实验室(European Molecular Biology Laboratory, EMBL)提供的那些局部序列比对工具使用改良的Smith-Waterman算法(参见www.ebi.ac.uk/Tools/psa/,以及Smith, T.F.&Waterman, M.S., “常见分子子序列的鉴定(Identification of common molecular subsequences)”, 《分子生物学杂志》(Journal of Molecular Biology), 1981 147(1):195-197)。进行比对时,会使用EMBL-EBI定义的默认参数。这些参数是(i)用于氨基酸序列:矩阵=BLOSUM62, 缺口开放罚分=10, 缺口延伸罚分=0.5, 或(ii)用于核酸序列:矩阵=DNAfull, 缺口开放罚分=10, 缺口延伸罚分=0.5。本领域技术人员应充分意识到以下事实:例如,如果相应的序列要在相比于与分子所来源的原始生物体的另一个生物体中使用,则编码蛋白的序列可以“密码子优化”。

具体实施方式

[0135] 本领域技术人员应当理解,本文描述的方面和实施方案不应被解释为限于公开它们的特定上下文,而是整个说明书中描述的方面和实施方案可以彼此组合,与具体上下文无关。

[0136] 本发明基于以下发现:通过使用明确定义的合成转录因子(STF)对内源基因的基因表达进行选择调控,为针对目的基因的特定时间和空间调控提供了合适的工具。反过来,这为优化转化和基因组编辑方法提供了基础,因此在转化/编辑中提供了更高的频率,从而可以改进农业生物技术中的方法。

[0137] 例如,不使用编码形态发生基因的核苷酸序列(例如BBM和WUS)作为分离的或异源的表达盒,可以使用专门设计的合成转录调控因子,例如TAL效应物或解除武装的CRISPR/核酸酶系统等,诱导内源性形态发生基因的表达以重新编程细胞,并在特定时间点以瞬时方式诱导细胞分裂和再生,而无需将转基因形态发生效应物或编码其的序列引入目的细胞或植物中。这些主要发现被扩展到建立包含至少一个激活或沉默结构域的合成转录因子(STF),从而以可诱导的方式特异性地上调或下调靶基因的表达。接着,所述专门设计的人工STF的直接作用随后被用于多种分子生物学方法中,以协同地从调控作用中受益,以优化转化、基因编辑或靶向沉默,其中,这些方法可用于植物育种和潜在的治疗应用。

[0138] 在本发明的一个方面,建立了通过使用对BBM和WUS具有特异性的合成转录因子诱导细胞分裂和植物细胞再生从而产生植物的方法,这些发现随后被外推到基于各种合成转录因子的其它方法和用途中。反过来,这些特异性转录因子使得可以提供通过利用对内源性形态发生基因具有特异性的合成转录因子(其在大多数植物物种(包括已知难以进行转化和再生的那些物种或变体)中可重编程细胞并诱导细胞分裂)提高转基因植物转化和/或再生效率的方法,以显著提高各种物种以及各种不同细胞类型(包括那些在标准设置下非常难以进行转化的细胞类型)的转化效率。因此,本发明涉及对目的形态发生基因具有特异性,靶向调控,优选地靶向激活的分子工具,即,本发明设计特异性合成转录因子和编码其的序列,本发明还涉及使用这些特异性合成或人工转录因子以靶向的方式优化基于转化和转染的植物生物技术方法(特别是基于基因组编辑的方法)的方法,或用于优化难以转化的

植物细胞的转化速率的方法。

[0139] 在本发明的上下文中首次证明,基于Cpf1的转录激活系统可以成功地用于植物中以调节内源靶基因的表达。有利地,所提供的手段和方法使得可以靶向具有富含AT的启动子区域的靶内源基因,这在以前是不可能的。通过提供专门设计的向导RNA阵列,该系统易于用于同时靶向多个基因组区域,并且可以在不引入转基因的情况下瞬时调控表达。

[0140] 一方面,公开了一种合成转录因子(STF)或编码其的核苷酸序列,其可包含至少一个识别结构域和至少一个基因表达调控结构域,特别是至少一个激活结构域,其中,该合成转录因子可配置为调控细胞系统中形态发生基因的表达。

[0141] 如本文所公开的,对任何内源基因(优选地为形态发生基因)的表达的“调控”包括如上所述的基因激活和基因阻遏。可以通过确定靶基因的表达间接或直接影响任何参数来测定这种调控。此类参数包括例如,RNA或蛋白水平的变化;蛋白活性的变化;产物水平的变化;下游基因表达的变化;诸如例如荧光素酶、CAT、 β -半乳糖苷酶或GFP的报道基因的转录或活性的变化(参见,例如,Mistili&Spector(1997),《自然生物技术》(Nature Biotechnology) 15:961-964)。对于形态发生基因,还可以通过视觉手段监测基因表达的调控,包括显微镜检查、植物发育观察等,监测基因表达的任何功能作用的变化。根据本发明的各个方面,如本文所公开的合成转录因子将优选地在转录水平上起作用,并因此将调控至少一个目的基因(优选目的形态发生基因)的转录。在某些实施方案中,可以将至少一个合成转录因子专门设计为上调目的基因(优选目的形态发生基因)的转录。

[0142] 如本文所用的,“细胞系统”是指包含待修饰目的细胞的全部或部分基因组的至少一个元件。因此,该细胞系统可以是任何体内或体外系统,还包括无细胞系统。因此,该细胞系统包含并提供待以合适方式修饰的靶基因组或基因组序列,即以可遗传修饰或操作的形式。因此,该细胞系统可选自例如真核细胞,包括植物细胞,或者该细胞系统可包含如上定义的遗传构建体,该遗传构建体包含待以高靶向方式修饰的真核细胞的基因组的全部或一部分。该细胞系统可以提供为分离的细胞或载体,或者该细胞系统可由组织、器官、材料或整个生物体中的细胞网络组成(以体内方式或者作为分离的体外系统)。因此,在本文中,细胞系统的“遗传物质”可以理解为生物体基因组的全部或一部分(生物体作为整体或一部分存在于待修饰的细胞系统中)。

[0143] 一方面,本发明提供了一种细胞系统,其可通过根据上述方面和实施方案中的任何一个所述的方法来获得。

[0144] 在根据本发明各个方面的一个实施方案中,可以将合成转录因子设计为调控形态发生基因的转录,其中,该形态发生基因可选自由BBM、WUS(Zuo等,2002,《植物学杂志》(Plant J.), 30(3):349-359)(包括WUS2(Nardmann和Werr,2006,《分子生物学与进化》(Mol.Biol.Evol.), 23:22492-22502))、WOX基因、WUS或BBM同源物、Lec1、Lec2、WIND1、ESR1、PLT3、PLT5或PLT7、IPT、IPT2、Knotted1和RKD4所构成的组。

[0145] 根据本发明的各个方面和实施方案,该形态发生基因可选自具有以下编码序列的序列:NM_001112491.1(SEQ ID NO:199)、NM_127349.4(SEQ ID NO:200)、NC_025817.2、KT285832.1(SEQ ID NO:201)、KT285833.1(SEQ ID NO:202)、KT285834.1(SEQ ID NO:203)、KT285835.1(SEQ ID NO:204)、KT285836.1(SEQ ID NO:205)、KT285837.1(SEQ ID NO:206)、XM_008676474.2(SEQ ID NO:207)、CM007649.1、NM_103997.4(SEQ ID NO:208)、

XM_010675298.2 (SEQ ID NO:209)、XM_010675704.2 (SEQ ID NO:210)、AB458519.1 (SEQ ID NO:211)、AB458518.1 (SEQ ID NO:212)、AK451358.1 (SEQ ID NO:213)、AK335319.1 (SEQ ID NO:214)、KU593504.1 (SEQ ID NO:215) 或 KU593503.1 (SEQ ID NO:216)。

[0146] 在另一个实施方案中,提供了一种合成转录因子,其中该形态发生基因包含选自由以下各项所构成的组的核苷酸序列:(i) SEQ ID NO:199-237中任一项所示的核苷酸序列,(ii) 具有SEQ ID NO:199-237中任一项所示的核苷酸序列的编码序列的核苷酸序列,(iii) 与(i)或(ii)的核苷酸序列互补的核苷酸序列,(iv) 与(i)、(ii)或(iii)的核苷酸序列具有至少50%、60%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性(优选在全长上)的核苷酸序列,(v) 在严格条件下与(iii)的核苷酸序列杂交的核苷酸序列,(vi) 编码包含SEQ ID NO:238-258中任一个所示的氨基酸序列的蛋白的核苷酸序列,(vii) 编码包含与SEQ ID NO:238-258中任一个所示的序列至少50%、60%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的氨基酸序列的蛋白的核苷酸序列,或(viii) 编码包含SEQ ID NO:238-258中任一个所示的氨基酸序列的蛋白的同源物、类似物或直系同源物的核苷酸序列。

[0147] 特别地,已将Wuschel (WUS) 多肽鉴定为在顶端分生组织的起始和维持中发挥关键角色,该分生组织含有一群多能干细胞(Endrizzi等,1996,《植物学杂志》(Plant Journal) 10:967-979)。WUS基因的拟南芥植物突变体包含干细胞,这些干细胞被错误指定并且似乎正在分化。WUS编码同源域蛋白,其充当转录调控因子(Mayer等,1998,《细胞》(Cell) 95:805-815,US 2004/166563A1)。一般认为,拟南芥芽分生组织的干细胞群体通过促进器官启动的CLAVATA (CLV) 基因与干细胞身份鉴定所需的WUS基因之间的调控环得以维持的,其中CLV基因在转录水平上抑制WUS。WUS表达可能足以诱导分生组织细胞鉴定和干细胞标志物CLV3的表达(Brand等(2000),《科学》(Science),289:617-619;Schoof等(2000),《细胞》(Cell) 100:635-644)。拟南芥中WUS的组成型表达已经示出会导致叶片不定芽增殖(植物中)(US 2004/166563A1)。

[0148] 其他WUS/WOX同源盒多肽及编码其的基因是技术人员公知的,并且可以被合成转录因子和/或使用本文公开的方法靶向。WUS同源盒多肽可选自WUS1、WUS2、WUS3、WOX2A、WOX4、WOX5或WOX9多肽(van der Graaff等,2009,《基因组生物学》(Genome Biology),10:248)或其同系物。WUS同源盒多肽可以是单子叶植物WUSAVOX同源盒多肽。在各个方面,WUS同源盒多肽可以是大麦、玉米、小米、燕麦、大米、黑麦、*Setaria* sp.、高粱、甘蔗、柳枝稷、黑小麦、草皮草或小麦WUSAVOX同源盒多肽。可替换地,WUS同源盒多肽可以是双子叶植物WUS同源盒多肽(参见WO 2017/074547 A1)。

[0149] 此外,AP2/ERF蛋白质家族是一类植物特异性推定转录因子,已被证明可调控多种发育过程,并且其特征是存在AP2/ERF DNA结合结构域。根据AP2/ERF蛋白包含一个(ERF亚族)还是两个(AP2亚族)DNA结合域,可以将其细分为两个不同的亚族。涉及到多种关键植物细胞功能的AP2家族成员之一是Baby Boom (BBM) 蛋白。来自拟南芥的BBM蛋白优先在种子中表达,并且已显示出在调控胚胎特异性途径中起着核心作用。已经表明,BBM的过表达会诱导幼苗上的体细胞胚和子叶样结构的自发形成。参见Boutiller等(2002),《植物细胞》(The Plant Cell),14:1737-1749。因此,AP2 (APETALA2) 蛋白家族的成员在胚胎形成过程中促进细胞增殖和形态发生。这种活性在促进植物无融合生殖中具有潜在用途。

[0150] 根据本发明的另一个形态发生靶标是胚珠发育蛋白2 (ODP2)。它也是AP2蛋白家族的成员。本发明的ODP2多肽包含两个预测的APETALA2 (AP2) 结构域, 并且是AP2蛋白家族的成员 (PFAM登录号为PF00847)。玉米ODP2多肽的AP2结构域位于SEQ ID NO:2的约氨基酸S273到N343和约S375到R437之间。已证明AP2家族的推定转录因子可调控广泛的发育过程, 并且该家族成员的特征是存在AP2 DNA结合结构域。预计该保守核将形成结合DNA的两亲性 α 螺旋。AP2结构域首先在APETALA2中鉴定, APETALA2是一种调控分生组织身份、花器官规格、种皮发育和花同源异型基因表达的拟南芥蛋白。现在已经在多种蛋白中发现了AP2结构域。

[0151] 因此, AP2家族的形态发生效应物在各种重要的生物学事件 (包括发育、植物再生、细胞分裂等) 中起着至关重要的作用, 这些形态发生效应物在鉴定和表征新型AP2家族成员并开发调控胚胎形成、转化效率和产量相关性状 (包括植物中油含量、淀粉含量等) 的新方法的农学发展领域具有重要价值, 并且这些形态发生效应物是本发明所述的合成转录因子和相关方法的相关靶标。

[0152] 已经进行了许多尝试来利用WUS、BBM和其他形态发生基因的调控提高转化效率、刺激植物细胞 (包括干细胞) 生长、刺激器官发生、刺激体细胞胚形成、诱导无融合生殖, 以及为细胞提供阳性选择。刺激器官发生和/或体细胞胚形成的能力可用于产生无融合生殖的植物。无融合生殖具有经济潜力, 因为它可以使任何基因型 (无论其杂合程度如何) 都能繁殖出真正的后代, 这个生殖过程可以绕过雌性减数分裂和配子配合, 而产生与母本在遗传上相同的胚胎。通过无融合生殖, 自适应或杂种基因型的后代将在整个生命周期中保持其遗传保真度。除固定杂种优势外, 无融合生殖还可以使无法在生产杂种的有效雄性不育或繁殖力恢复系统的农作物中实现商业化杂种生产。无融合生殖可以使杂交发育更加有效。它还简化了杂交生产, 并增加了雄性不育性良好的植物物种的遗传多样性。

[0153] 尽管如此, 目前调控植物细胞内源性形态发生基因库的所有方法仍依赖于提供编码目的形态发生基因的基因来过表达相应的形态发生基因。因此, 当前的方法依赖于目的形态发生基因的稳定或瞬时引入和/或过表达。相反, 本发明确定了一种解决方案, 专门设计合成转录因子来调控目的形态发生基因的转录水平, 优选以瞬时和/或可调节的方式, 而无需引入形态发生基因产物的外源性转基因序列或编码该产物的序列。这提供用于提高植物 (甚至难以进行转化的植物) 转化效率的方法, 例如, 复杂基因组编辑方法, 以及提供用于提高单倍体或双单倍体生物体或细胞系统的方法奠定了基础。

[0154] 各种不同的分子均可用作根据本发明的所述至少一个识别结构域。根据本文公开的各个方面和实施方案, 识别结构域代表蛋白结构域 (任选地作为融合分子), 其具有位点特异性DNA识别并因此具有结合和/或相互作用活性。识别结构域可以是天然存在的蛋白的结构域, 或者该识别结构域可以是此蛋白的片段。优选地, 所述至少一个识别结构域已经被特异性地工程化, 以优化其结合至目标形态发生基因的区域或目标形态发生基因周围的区域的靶标特异性。

[0155] 根据本发明, 可以使用一个以上的识别结构域来提高靶标特异性和/或结合特征, 从而优化对至少一个目的形态发生基因的调控。

[0156] 在一个实施方案中, 合成转录因子可包含选自由至少一个TAL效应物、至少一个解除武装的CRISPR/核酸酶系统、至少一个锌指结构域和至少一个解除武装的归巢核酸内切

酶或其任何组合所构成的组的分子的至少一个识别结构域或片段。

[0157] 在另一个实施方案中,合成转录因子可包含选自CRISPR/dCas9系统、CRISPR/dCpf1系统、CRISPR/dCasX系统或CRISPR/dCasY系统或其任何组合的至少一个解除武装的CRISPR/核酸酶系统,其中该至少一个解除武装的CRISPR/核酸酶系统(如果存在)包含至少一个向导RNA。

[0158] 天然存在的DNA结合转录因子通常至少包含两个结构域:DNA结合结构域(DBD)和转录激活结构域(TAD) (Latchman,2008;Ptashne和Gann,2002)。

[0159] 黄单胞菌属(genus *Xanthomonas*)中植物病原细菌的TAL效应物,通过结合宿主DNA并激活效应物特异性宿主基因,在疾病或触发防御中起重要作用(参见例如,Gu等(2005),《自然》(Nature),435:1122;Römer等(2007),《科学》(Science),318:645)。特异性取决于不完整的效应物可变数目,通常是34个氨基酸重复(Schornack等(2006),《植物生理学》(J.Plant Physiol.),163:256)。多态性主要在重复位置12和13处,其在本文中称为重复可变双残基(RVD)。TAL效应物的RVD以直接、线性的方式对应于其靶位点的核苷酸,一个RVD对应一个核苷酸,但具有一定的简并性,并且没有明显的背景依赖性。这一发现代表了一种蛋白质-DNA识别的有价值机制,该机制能够预测新的靶标特异性TAL效应物的靶位点。因此,TAL效应物可以在研究和生物技术中用作靶向嵌合核酸酶,促进GE方法的同源重组。TAL效应物本身不包含核酸酶结构域。所谓的转录激活因子样效应物核酸酶(TALEN)代表结合TAL效应物功能和核酸酶功能以允许插入位点特异性DNA裂解的人工或合成分子。例如,TAL效应物可以通过C-末端核定位结构域进入宿主细胞核,并且可以通过与宿主基因的启动子区域中的效应物结合元件结合而特异性地激活相应的宿主基因。高度保守的33-35个氨基酸重复序列的中央结构域(每个重复序列均在位置12和13包含高变二核苷酸或RVD)负责识别特异性宿主基因启动子序列。每个TAL效应物都以右旋超螺旋缠绕在DNA周围,从而将每个RVD的第二个残基置于主沟中,并与正向链中的单个核苷酸接触。这些相互作用限定了每种TAL效应物的特异性。然后,C-末端酸性激活结构域可能通过直接结合宿主RNA聚合酶复合物来激活或增强相应内源基因的表达。

[0160] TAL效应物识别特异性DNA序列的模块化机制使得可以在TAL效应物的识别结构域中鉴定和设计人工重复阵列,从而设计能够特异性诱导目的内源基因表达的TAL效应物。

[0161] 天然TALE的基因组靶位点的计算分析表明,转录起始位点(TSS)周围-300至+200bp的表观核心启动子区域优先出现(Grau等,PLoS Comput Biol.,2013;9)。先前基于TALE AvrBs3、AvrXa7和AvrXa27的研究表明,它们将靶基因的天然TSS移至TALE与DNA结合的位置下游40-60bp附近。将Bs3启动子中的AvrBs3-box移到更上游的位置会导致TSS随之向上游移动。这些观察结果表明,TALE控制着与TATA结合蛋白功能类似的转录的发生和位置(Kay等,《科学》(Science),2007;318:648-651)。

[0162] 因此,根据本发明的各个方面和实施方案,TAL效应物结合结构域表示合适的识别结构域,因为可以针对目的靶位点微调结合和识别特异性。因此,可以以高靶向性的方式调控目的形态发生基因的表达,优选地为转录,因为可以将至少一种定制的TAL效应物设计为合成转录因子的至少一个识别结构域。

[0163] TAL效应物在其自然环境中起异源转录因子作用(Yang等,2006),通过细菌III型分泌系统递送到宿主细胞中(Szurek等,2002),在宿主细胞中C-末端核定位信号将它们引

导至细胞核 (Gurlebeck等,2005;Szurek等,2001、2002;Van den Ackerveken等,1996;Yang和Gabriel,1995)。高度保守的33-35个氨基酸重复序列的中央结构域(每个重复序列在12和13位含有高变残基(RVD))引导识别称为效应物结合元件(EBE)的特异性宿主基因启动子序列(Boch等,2009;Moscou和Bogdanove,2009)。每个TAL效应物将DNA包裹在右手超螺旋中,将每个RVD的第二个残基定位在大沟中,并与正向链中的单个核苷酸接触(Deng等,2012;Mak等,2012)。这些相互作用以可预测的方式共同定义了构成EBE的相邻核苷酸的数量和身份。然后,C-末端酸性激活结构域(AD)可能通过直接与宿主RNA聚合酶复合物结合来激活或增强转录(参见Hummel等,《分子植物病理学》(Molecular Plant Pathology),2017,18(1),55-66)。

[0164] 与现有技术的教导相反,本发明部分基于以下发现:对位于相对于目的基因(优选为形态发生基因,例如BBM和WUS)的起始密码子的特异性上游或下游位置的内源性核苷酸序列具有特异性的基于合成TAL效应物的转录因子、基于解除武装的ZFP的转录因子,或基于解除武装的CRISPR的转录因子,可诱导植物细胞在所述基因的转录和表达,从而提高此植物的再生频率。值得注意的是,如果靶向TATA-box以外的非经典调控区域或启动子区域,则可以提高该效率,而天然存在的转录因子以及可商购的转录因子通常通过结合至目的基因的启动子区域内的区域来发挥其功能。有证据表明,与直接靶向TATA区域相比,靠近TATAbox的转录激活更高。因此,基于各种不同的TAL效应物、CRISPR、锌指或归巢核酸内切酶识别结构域的本发明的转录因子包括不同的结构,从而使得可以更好和更精确地调控和调节目的形态发生基因。

[0165] 因此,本发明的合成转录因子和方法的一个优点是,如果正确地设计为包括最佳识别和激活区域,则合成转录因子也可以作用于无TATA的基因上或TATA区域之外。在某些实施方案中,至少一个识别结构域也可靶向目的基因的TATA区域。

[0166] 例如,TAL效应物DNA结合结构域可以对靶DNA具有特异性,其中,该DNA结合结构域包含多个DNA结合重复序列,每个重复序列包含决定靶DNA中碱基对识别的RVD,其中,每个DNA结合重复序列负责识别靶DNA中的一个碱基对,并且其中,TALEN包含以下RVD中的一个或多个:HD,用于识别C;NG,用于识别T;NI,用于识别A;NN,用于识别G或A;NS,用于识别A或C或G或T;N*,用于识别C或T;HG,用于识别T;H*,用于识别T;IG,用于识别T;NK,用于识别G;HA,用于识别C;ND,用于识别C;HI,用于识别C;HN,用于识别G;NA,用于识别G;SN,用于识别G或A;以及YG,用于识别T。TALEN可包含以下RVD中的一个或多个:HA,用于识别C;ND,用于识别C;HI,用于识别C;HN,用于识别G;NA,用于识别G;SN,用于识别G或A;YG,用于识别T;和NK,用于识别G,以及以下中的一个或多个:HD,用于识别C;NG,用于识别T;NI,用于识别A;NN,用于识别G或A;NS,用于识别A或C或G或T;N*,用于识别C或T;HG,用于识别T;H*,用于识别T;以及IG,用于识别T。

[0167] 锌指蛋白(ZFP)是可以以序列特异性方式与DNA结合的蛋白。锌指首次在非洲爪蟾(*Xenopus laevis*)的卵母细胞的转录因子TFIIIA中鉴定出。表征这些蛋白质的一类(Cys2His2类)的示例性基序是Xaa-Cys-Xaa-Cys-Xaa-His-Xaa-His(SEQ ID NO:313),其中Xaa是任何氨基酸。这些蛋白的单个指具有简单的 $\beta\beta\alpha$ 结构,可围绕中心锌离子折叠,而串联的指可沿DNA的大沟接触3-4个碱基对的相邻亚位点(Pabo等(2001),“新型Cys2His2锌指蛋白的设计与选择(Design and selection of novel Cys2His2 zinc finger proteins)”。

《生物化学年鉴》(Ann.Rev.Biochem.),70:313-40)。单个锌指结构域的长度约为30个氨基酸,一些结构研究表明,其包含一个 β 转角(包含两个不变的半胱氨酸残基)和一个 α 螺旋(包含两个不变的组氨酸残基),并由两种胱氨酸和两种组氨酸通过锌原子的配位而保持一种特殊的构象。几种其它类别的锌指蛋白也是已知的,例如,包含由位于N-末端的 β 发夹和位于C-末端的 α -螺旋(它们各自为锌结合提供两个配体)组成的基序的高音谱号类(treble-clef class),尽管N-末端的 β 发夹和C-末端的 α -螺旋之间可能存在不同长度或构型的环和第二个 β 发夹,或褶皱特征为两个 β 发夹形成两个结构相似的锌结合亚位点的锌带状ZFP。

[0168] 出于基因组编辑(GE)的目的,分子生物学技术可用于改变锌指的DNA结合特异性,而这种工程化锌指的串联重复序列可用于靶向所需的基因组DNA序列(Jamieson等,“工程化锌指蛋白的药物发现(Drug discovery with engineered zinc-finger proteins)”,《自然综述·药物发现》(Nature Reviews.Drug Discovery),2(5):361-8)。将第二个蛋白结构域(如转录激活子或阻遏物)融合到结合在给定基因启动子附近的一系列工程化锌指上,可用于改变该基因的转录。工程化锌指阵列与切割或以其他方式修饰DNA的蛋白结构域之间的融合也可用于将那些活性靶向所需的基因组位点。工程化锌指阵列的最常见应用包括锌指转录因子和锌指核酸酶。典型的工程化锌指阵列具有3至6个单独的锌指基序,并结合长度为9个碱基对(bp)到18bp的靶位点。

[0169] 大范围核酸酶是特征为具有较大识别位点(12至40个碱基对的双链DNA序列)的内切脱氧核糖核酸酶。因此,该位点在任何给定的基因组中通常仅出现一次。大范围核酸酶可用于在哺乳动物细胞和植物中实现很高水平的基因靶向效率(Rouet等,《分子细胞生物学》(Mol.Cell.Biol.),1994,14,8096-106;Chouluka等,《分子细胞生物学》(Mol.Cell.Biol.)1995,15,1968-73)。在大范围核酸酶中,归巢核酸内切酶的LAGLIDADG家族在过去几年中已成为研究基因组和基因组工程的重要工具。

[0170] 解除武装(即,核酸酶缺陷)的归巢核酸内切酶(HE)代表根据本发明的合适类别的识别结构域。HE是广泛的天然大范围核酸酶家族,包括数百种蛋白(Chevalier和Stoddard,《核酸研究》(Nucleic Acids Res.),2001,29,3757-74)。这些蛋白由可移动的遗传元件编码,并通过称为“归巢”的过程传播:核酸内切酶切割不存在可移动元件的同源等位基因,从而刺激同源重组事件,将可移动的DNA复制到受体基因座中(Kostriken等,《细胞》(Cell);1983,35,167-74;Jacquier和Dujon,《细胞》(Cell),1985,41,383-94)。鉴于它们的天然功能以及在功效和特异性方面的卓越裂解特性,HE提供了理想的支架以衍生出用于基因组工程的新型核酸内切酶。HE的一个家族称为LAGLIDADG家族。LAGLIDADG(SEQ ID NO:314)指整个家族中实际上保守的唯一序列,并且在该蛋白中以一个或(更经常地)两个拷贝被发现。具有单一基序的蛋白(例如,I-Cre1)形成同型二聚体,并切割回文或假回文DNA序列,而较大的双基序蛋白(例如,I-Sce1)是单体,并裂解非回文靶标。已经结晶了七个不同的LAGLIDADG蛋白,它们表现出非常惊人的核心结构保守性,这与在一级序列水平上缺乏相似性形成了鲜明的对比(Jurica等,《分子细胞》(Mol.Cell.),1998,2,469-76;Chevalier等,《自然结构与分子生物学》(Nat.Struct.Biol.),2001,8,312-6;Chevalier等,《分子生物学杂志》(J.Mol.Biol.),2003,329,253-69)。对与其天然靶标结合的I-Cre结构的分析表明,在每个单体中,八个残基(Y33、Q38、N30、K28、Q26、Q44、R68和R70)与七个碱基在 ± 3 、4、5、6、7、9和10处建立直接相互作用(Jurica等,1998)。此外,一些残留与几个碱基建立了水介导

接触;例如,位置8和-8处的S40和N30和碱基对(Chevalier等,2003)。催化核心是中心,由两个对称单体/结构域组成。具有修饰的切割位点的HE是本领域技术人员公知的,并且可用于将解除武装的HE定义为根据本发明的至少一个识别结构域。

[0171] 根据本发明的各个方面和实施方案,锌指蛋白和从其衍生的结构域可以用作至少一个识别结构域,该至少一个识别结构域可设计为实现根据本发明的合成转录因子的识别特性。

[0172] 除TAL效应物、解除武装的ZFP和大范围核酸酶外,非功能性CRISPR/核酸酶系统也可用于特异性靶向形态发生基因,并促进植物细胞的再生。在这些系统中,使用CRISPR核酸酶,例如Cas9、Cfp1、CasX和/或CasY,其中,核酸酶活性已关闭,以避免切割目标基因组序列。非功能性CRISPR/核酸酶系统的靶特异性由对目的内源性形态发生基因的上游核苷酸启动子区域具有特异性的crRNA和/或sgRNA确定。然后,与CRISPR/核酸酶系统融合的激活结构域将转录机制募集到基因基因组,从而诱导目的内源性形态发生基因的表达。值得注意的是,使用至少一个向导RNA可以显著提高靶特异性,因为这种CRISPR核酸序列还有助于识别目的基因组靶DNA。此外,解除武装的CRISPR核酸酶和向导RNA的双重识别特性使得在设计根据本发明的合成转录因子识别结构域时可以具有更灵活性的灵活性,这反过来又提供了对目的形态发生基因的更好的识别以及因此更好的调控活性。

[0173] 在本发明的优选实施方案中,该至少一个识别结构域是至少一个解除武装的CRISPR/核酸酶系统,或是至少一个解除武装的CRISPR/核酸酶系统的片段。

[0174] CRISPR系统在其天然环境中描述了一种分子复合物,该分子复合物包含与Cas核酸酶或另一个CRISPR核酸酶(如Cpf1核酸酶(Zetsche等,2015,同上))组合的至少一个小型单个非编码RNA,其可产生特异性DNA双链断裂。目前,CRISPR系统分为2类,包括五种类型的CRISPR系统,例如,使用Cas9作为效应物的II型系统和使用Cpf1作为效应物分子的V型系统(Makarova等,《然微生物学综述》(Nature Rev.Microbiol.),2015)。在人工CRISPR系统中,合成非编码RNA和CRISPR核酸酶和/或可选地修饰的CRISPR核酸酶(进行修饰以充当切口酶或缺乏任何核酸酶功能),可与结合crRNA和/或tracrRNA的功能的至少一个合成或人工向导RNA或gRNA一起使用(Makarova等,2015,同上)。在天然系统中由CRISPR/Cas介导的免疫应答需要CRISPR-RNA(crRNA),其中,这种控制CRISPR核酸酶特异性激活的向导RNA的成熟度在迄今已被鉴定的各种CRISPR系统之间存在显著差异。首先,将侵袭性DNA(也称为间隔子)整合到CRISPR基因座近端的两个相邻重复区域之间。例如,II型CRISPR系统可以将Cas9核酸酶编码为干扰步骤的关键酶,该系统既包含crRNA,也包含反式激活RNA(tracrRNA)作为向导基序。这些杂交并形成双链(ds)RNA区域,其被RNaseIII识别,并可以被切割以形成成熟的crRNA。然后,这些反过来又与Cas分子缔合,以便将核酸酶特异性地引导至靶核酸区域。重组gRNA分子既可包含可变DNA识别区域,也可包含Cas相互作用区域,因此可以独立于特异性靶核酸和所需的Cas核酸酶进行专门设计。作为另一个安全机制,靶核酸区域中必须存在PAM(与前间隔序列相邻的基序);这些是直接来自Cas9/RNA复合物识别的DNA的DNA序列。酿脓链球菌Cas9的PAM序列已被描述为“NGG”或“NAG”(标准IUPAC核苷酸代码)(Jinek等,“适应性细菌免疫中的可编程双RNA向导的DNA核酸内切酶(Aprogrammable dual-RNA-guided DNAendonuclease in adaptive bacterial immunity)”,《科学》(Science)2012,337:816-821)。金黄色葡萄球菌Cas9的PAM序列是“NNGRRT”或“NNGRR(N)”。其它变体

CRISPR/Cas9系统也是已知的。因此,脑膜炎双球菌Cas9在PAM序列NNNGATT处裂解。嗜热链球菌Cas9在PAM序列NNAGAAW处裂解。最近,已经描述了用于弯曲杆菌的CRISPR系统的另一种PAM基序NNNNRYAC (WO 2016/021973A1)。对于Cpf1核酸酶,已经描述了,与Cas9系统识别的通常富含G的PAM相比,没有tracrRNA的Cpf1-crRNA复合物可有效识别和切割由富含短T的PAM进行的靶DNA (Zetsche等,同上)。此外,通过使用修饰的CRISPR多肽,可获得特定的单链断裂。Cas切口酶与各种重组gRNA的组合使用还可通过双DNA切口诱导高度特异性DNA双链断裂。此外,通过使用两个gRNA,可以优化DNA结合以及DNA切割的特异性。同时,最初针对细菌描述的其它CRISPR效应物(如CasX和CasY效应物)是可用的,并且代表了其它效应物,其可用于基因组工程目的 (Burstein等,“来自未培养微生物的新型CRISPR-Cas系统 (New CRISPR-Cas systems from uncultivated microbes)”,《自然》(Nature),2017,542,237-241)。

[0175] 目前,例如,已经修饰了依赖Cas9的II型系统或其变体或任何嵌合形式作为核酸内切酶,用于基因组工程。合成CRISPR系统由两个部分组成,一个是“向导RNA”(gRNA),在本文中也称为“单链向导RNA”(sgRNA)或“CRISPR核酸序列”;另一个是非特异性CRISPR相关核酸内切酶,该合成CRISPR系统可用于通过共表达对待靶向基因具有特异性且能够与核酸内切酶Cas9缔合的gRNA而产生敲除细胞或动物。值得注意的是,gRNA是一种人工分子,其包含与Cas或任何其它CRISPR效应蛋白或其变体或催化活性片段相互作用的一个结构域,以及与目的靶核酸相互作用并因此代表crRNA和tracrRNA的合成融合体(作为“单链向导RNA”(sgRNA)或简称“gRNA”)的另一个结构域。该基因组靶标可以是约20个核苷酸DNA序列,前提是该靶标应紧邻PAM序列的上游。PAM序列对于目标结合非常重要,确切的序列取决于Cas9的种类,并例如读取酿脓链球菌衍生的Cas9的5'NGG 3'或5'NAG3'(标准IUPAC核苷酸代码)(Jinek等,《科学》(Science),2012,同上)。金黄色葡萄球菌Cas9的PAM序列为NNGRRT或NNGRR(N)。许多其他变体CRISPR/Cas9系统也是已知的,尤其包括切割PAM序列NNNGATT的脑膜炎双球菌Cas9。切割PAM序列NNAGAAW的嗜热链球菌Cas9。使用修饰的Cas核酸酶,可以将靶向单链断裂引入目标靶序列中。可以使用双切口系统引入这种Cas切口酶与不同重组gRNA高位点特异性DNA双链断裂的组合使用。使用一种或多种gRNA还可提高总体特异性并减少脱靶效应。

[0176] 为了本发明的目的,特别感兴趣的Cas或Cpf1核酸酶的第三变体是核酸酶缺陷的Cas9(dCas9)或dCpf1(Qui等,2013,《细胞》(Cell),154,442-451)。Cas9的HNH结构域中的H840A突变和RuvC结构域中的D10A突变使切割活性失活,但不能阻止DNA结合(Gasiunas等,2012,《美国科学院院报》(Proc.Natl.Acad.Sci.U.S.A.) 111,E2579-2586)。因此,如果正确配置,这些变体可以修改用于序列特异性靶向基因组的区域而不切割。

[0177] 如Tang等所述,Cpf1可衍生自例如氨基酸球菌属(Acidaminococcus sp.) BV3L6(AsCpf1),或衍生自Lachnospiraceae bacterium ND2006(LbCpf1)(Tang等(2017),用于植物中的有效基因组编辑和转录阻遏的CRISPR/Cpf1系统(ACRISPR/Cpf1 system for efficient genome editing and transcriptional repression in plants),《自然植物》(Nature Plants),3:17018)。优选的dLbCpf1变体由SEQ ID NO:282-284和SEQ ID NO:288-290表示。

[0178] CRISPR/Cpf1系统可靶向富含AT的启动子区域,并且可用于多种作物植物。由于

Cpf1的RNase活性能够处理单个转录物的多个crRNA,因此基于Cpf1的转录调控系统比众所周知的基于Cas9的系统具有优势,可以轻松应用于多重基因调控。

[0179] 在本发明的各方面的优选实施方案中,因此,该至少一个解除武装的CRISPR/核酸酶系统是CRISPR/dCpf1系统,其中该至少一个解除武装的CRISPR/核酸酶系统包含至少一个向导RNA。

[0180] 基于Cpf1的转录调控系统具有高度特异性和灵活性,并可以通过使用靶向多个基因组区域的向导RNA阵列来同时激活/抑制多个基因。此外,基于Cpf1的系统无需引入目的基因的外源多聚核苷酸或多肽序列即可实现提高基因表达。因此有可能在无转基因的环境中瞬时诱导内源基因的基因表达。此外,基于Cpf1的系统提供了靶向富含AT的序列的手段,这是已知的对富含GC的区域表现出强烈偏好的基于Cas9的转录调控系统迄今所无法实现的。因此,该系统为植物细胞中目的内源性靶基因的转录激活和/或抑制提供了强大的工具。其使用方便,且适于同时靶向多个基因。重要的是,这是首次示出基于Cpf1的转录激活在植物细胞中起作用。尽管现有技术描述了拟南芥中基于Cpf1的基因抑制,但尚未在植物中显示出基于Cpf1的转录激活,这表明,由转录激活结构域替换转录抑制结构域并不简单,并且需要对正确的接头和激活结构域序列进行精心的配置和测试。

[0181] 在根据本发明的各个方面的一个实施方案中,该识别结构域可包含CRISPR复合物的至少一个gRNA。在某些实施方案中,可以存在一个以上gRNA,例如,可以使用一系列gRNA。单个细胞或细胞系统中多个向导RNA的表达,例如两个、三个、四个、五个或更多个gRNA的表达,可实现内源性基因靶标的协同调控,从而实现在宽动态范围上内源基因表达的组合调控,因为如果根据本发明的STF可为STF提供额外的靶特异性并减少脱靶效应,则该至少一个gRNA作为识别部分,特别是在STF设计为靶向较大真核基因组中的基因时。每个gRNA均可靶向独立的调控/识别区域。

[0182] 在根据本发明的各个方面的一个实施方案中,合成转录因子可配置为通过结合位于相对于起始密码子一定距离处的调控区域来调控形态发生基因的表达,优选地为调控形态发生基因的转录。

[0183] 如本文所用的,“调控区域”是指至少一个识别结构域与基因组中目的形态发生基因处或附近的靶序列的结合位点。取决于本文进一步公开的至少一个激活结构域和至少一个识别结构域的性质,可存在两个离散调控区域,或可存在重叠调控区域,本发明的合成转录因子的不同结构域可以以模块化的方式进行组装。

[0184] 在某些实施方案中,该至少一个识别结构域可靶向相对于目的基因的起始密码子的至少一个序列(识别位点),该序列可以是至少1000bp,相对于目的基因起始密码子的上游(-)或下游(+)-700bp至+700bp、-550bp至+500bp或-550bp至+425bp。在某些实施方案中,识别启动子附近的识别结构域可能是优选的,而这表示了本发明的特异性STF的优点,即,STF的靶向范围比常规或天然存在的TF进行了高度扩展。由于识别和/或激活结构域可以被专门设计和构建以特异性识别和靶向调控的热点。

[0185] 在某些实施方案中,该至少一个识别位点可以是相对于目的基因的起始密码子-169bp至-4bp、-101bp至-48bp、-104至-42bp或-175至+450bp(上游(-)或下游(+)),以提供最佳的空间结合环境,从而实现最佳的调控(优选地是转录激活)活性。特别是对于根据本发明的基于CRISPR的合成转录因子,其与作为识别部分的向导RNA一起起作用,该结合位点

也可位于目的基因的编码区域内(目的基因的起始密码子的下游)。

[0186] 在本发明的合成转录因子的其他实施方案中,识别结构域可结合至目的基因的5'和/或3'非翻译区域(UTR)。在实施方案中,在采用不同的识别结构域的情况下,至少两个识别结构域可结合至目的形态发生基因的不同靶区域,包括5'和/或3'UTR,但是它们也可以在基因区域之外结合,但是仍然在最多1到1500bp的特定距离内。可以结合识别结构域的一个优选区域位于目的形态发生基因的起始密码子上游约-4bp至约-300bp,优选地约-40bp至约-170bp。值得注意的是,由于至少一种gRNA在所述STF中的附加功能,因此本文公开的某些STF,尤其是基于CRISPR的STF,具有更大的识别位点灵活性。

[0187] 根据本文提出的各个方面和实施方案,因此,目的基因组中的识别结构域和相应地识别位点的长度可取决于STF和所应用的识别结构域的性质而变化。基于至少一个识别结构域的分子特征,这还将确定相应的至少一个识别位点的长度。例如,在单个锌指可以为约8bp至约20bp,其中三至六个锌指基序之间的阵列可能是优选的,在这种情况下,单个TALE识别位点可以为约11至约30bp,或更多。基于CRISPR的STF的gRNA的识别位点包含与目的基因组区域杂交的gRNA的靶向序列或“间隔子”序列,而gRNA包含其他结构域,包括与根据本发明的解除武装的CRISPR效应物相互作用的区域。基于解除武装的CRISPR效应物的STF的识别位点将包含一个PAM基序,因为PAM序列对于任何CRISPR效应物的靶标结合都是必需的,确切的序列取决于CRISPR效应物(即,本文公开的解除武装的CRISPR效应物)的种类。

[0188] 在本发明的各个方面的一个实施方案中,该合成转录因子可包含至少一个激活结构域,其中,该至少一个激活结构域可选自由酸性转录激活结构域所构成的组,优选地,其中,该至少一个激活结构域可来自水稻白叶枯病菌的无毒基因、单纯疱疹病毒的VP16或四聚体VP64、VPR、SAM、Scaffold、Suntag、P300、VP160或其任何组合。为了增强对至少一个目的形态发生基因的调控,可存在两个、三个、四个、五个或五个以上激活结构域。在本发明的优选实施方案中,该激活结构域是VPR(SEQ ID NO:276)。

[0189] VP16是最初在1型单纯疱疹病毒(HSV)中发现的转录因子,与病毒即早基因的激活有关(Flint和Shenk,1997;Wysocka和Herr,2003)。VP16野生型序列具有490个氨基酸,在其中心区域具有间接DNA结合所需的核心结构域,以及位于其最后81个氨基酸内的羧基末端TAD(Greaves和O'Hare,1989;Triezenberg等,1988)。VP16最初包含在HSV的病毒体(病毒颗粒)中,并在感染后释放到动物细胞中。VP16首先通过其核心结构域与宿主核蛋白HCF结合,然后与另一个宿主核蛋白Oct-1结合以形成三组分蛋白复合物。然后,该复合物在即早基因的启动子中与其靶DNA序列TAATGARAT(R为嘌呤)结合。这是通过Oct-1与靶DNA序列或重叠该序列的5'部分的共有八聚体基序之间的相互作用实现的。然后,HCF稳定VP16和Oct1之间的相互作用。一旦被募集为即早基因中,VP16通过TAD和其他转录因子之间的相互作用激活基因(Hirai等,《国际发育生物学杂志》(Int.J.Dev.Biol.),2010,54(11-12):1589-1596)。同时,原始的VP16结构域已被广泛用于使用人工或合成转录因子进行各种研究。通常,以单个形式或例如以三联(VP48)或作为VP16的10x串联拷贝(VP160)的形式包含VP16的最小激活结构域的核心结构域被用于这些目的。

[0190] 水稻白叶枯病菌的TAL效应物基因的天然激活结构域是与TAL转录因子一起使用的最明显的激活结构域,并且也代表根据本发明的各个方面可单独或组合使用的一个激活

结构域,但也已用于其他设置。它们属于酸性(转录)激活结构域家族。

[0191] SAM(协同激活介体)激活结构域通常由三个组分组成:通常与VP64融合体组合的溶核失活/灭活的CRISPR核酸酶、在四环和茎环处结合了两个MS2 RNA适体的向导RNA,以及MS2-P65-HSF1激活辅助蛋白(Konermann等,2015,“工程CRISPR-Cas9复合物进行的基因组规模转录激活(Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex)”,《自然》(Nature)517:583-588)。因此,向导RNA可包含两个拷贝的MS2噬菌体RNA发夹,该拷贝与RNA结合蛋白(RBP)MCP(MS2外壳蛋白)相互作用。

[0192] SAM系统采用多种转录激活子来产生协同效应,这使SAM系统成为单独使用或与用于根据本发明的合成转录因子的其他激活结构域结合使用的高通用性激活结构域。在优选的实施方案中,其中,该合成转录因子使用基于CRISPR的识别结构域,可以进一步工程化向导RNA,以优化激活和识别结构域之间的相互作用。

[0193] 根据本发明单独或组合使用的另一个激活结构域是与串联连接的目的识别结构域融合的三方效应物VPR(VP64、p65和Rta)(Rusca和Qi,《分子细胞生物学》(Mol.Cell.Biol.),2015年11月;35(22):3800-3809)。已显示使用VPR激活结构域会导致哺乳动物细胞中GFP表达的转录激活超过20倍(Liu等(2017),使用基于可调节CRISPR/Cpf1的转录因子进行细胞信号传导(Engineering cell signaling using tunable CRISPR/Cpf1 based transcription factors),《自然-通信》(Nature Communications),8(1):2095)。

[0194] 根据本发明单独或组合使用的又一个激活结构域是将例如VP64的多个拷贝募集到特定向导RNA(可选地与其它激活子一起)的“Scaffold”(Chavez等,《自然方法》(Nat.Methods),2016,13(7),563-567)。

[0195] 根据本发明单独或组合使用的另一个激活结构域是包含重复肽阵列的“Suntag”,其可通过将转录激活结构域的多个拷贝募集到本发明的合成转录因子的核酸酶缺陷型识别结构域中而募集抗体融合蛋白的多个拷贝,从而产生强有效的合成转录因子(Tanenbaum等,《细胞》(Cell),2014,159(3):635-46)。

[0196] 在另一个实施方案中,可以使用SAM激活结构域系统,特别是SAM修饰的向导RNA,与suntag激活结构域一起,将与例如VP64连接的具有所需特异性的单链可变片段(scFv)募集到识别结构域的一端,同时将p65-hsf1募集到基于CRISPR的合成转录因子的向导RNA。scFvs,本身不代表激活子,具有极高的特异性和靶标识别多功能性,可以进行工程改造,所以非常适于将目的激活子的多个拷贝募集到目的位置,即,根据本发明的各个方面和实施方案,scFv可与本文公开的激活结构域一起用作放大器。

[0197] 根据本发明单独或组合使用的又一个激活结构域是p300或EP300或E1A(本文中可互换使用)或CBP(也称为CREB结合蛋白或CREBBP)。p300和CBP都与众多转录因子相互作用,并起着提高其靶基因表达的作用(Kasper等,2006,《分子细胞生物学》(Mol.Cell.Biol.),26(3),789-809)。P300和CBP具有相似的结构。两者都包含五个蛋白相互作用结构域:核受体相互作用结构域(RID)、KIX域结构域(CREB和MYB相互作用结构域)、半胱氨酸/组氨酸区域(TAZ1/CH1和TAZ2/CH3),以及干扰素反应结合结构域(IBiD)。p300的最后四个结构域KIX、TAZ1、TAZ2和IBiD分别与跨转录因子p53的两个反式激活结构域9aaTAD的序列紧密结合。此外,p300和CBP各自包含蛋白或组蛋白乙酰转移酶(PAT/HAT)结构域和结合乙酰化赖氨酸与功能未知的PHD指基序的溴结构域。保守的结构域通过长段的非结构化接头相连。

P300和CBP可以通过三种方式提高基因表达:通过其固有的组蛋白乙酰转移酶(HAT)活性放松基因启动子上的染色质结构;通过将包括RNA聚合酶II的基础转录机制募集至启动子;以及/或通过充当接合体分子。

[0198] 根据本发明的各个实施方案,可以分别优化本发明的合成转录因子的至少一个识别结构域和至少一个激活结构域,以实现完美的结合和调控活性。因此,特定数目的激活结构域可能适于适当定位在合成转录因子构建体中的给定识别结构域,以实现最佳的调控活性,优选地为转录激活。因此,根据本发明的各个方面的至少一个激活结构域可包括某些修饰以优化至少一个激活结构域以最优方式与至少一个识别结构域相互作用,使得两个结构域都可以到达要调控的目标靶位点。

[0199] 在一个实施方案中,该至少一个激活结构域可以在本发明的合成转录因子内位于相对于该至少一个识别结构域的N-末端和/或C-末端。对于至少一个识别结构域和至少一个激活结构域之间的融合分子,该构型可能是最佳构型。根据各个实施方案,该至少一个识别结构域和该至少一个激活结构域可通过合适的接头序列分开,以达到最佳的灵活性并避免结构域实现其功能的空间障碍。

[0200] 在一个实施方案中,该合成转录因子可包含至少一个其他元件,包括至少一个核定位信号(NLS)、细胞器定位信号(包括例如线粒体定位信号或叶绿体定位信号),以将STF靶向细胞或细胞系统的区室,STF可以在其中发挥其功能。此外,该合成转录因子可包含至少一个标签,例如以可视化该合成转录因子,从而跟踪转录因子的亚细胞定位,以及/或在合成转录因子内提供活性部分,例如scFv结合位点,以将其他分子连接至该合成转录因子、转位结构域(例如TALE分子中存在的转位结构域),以及本文还公开的和技术人员已知的。该至少一个其它结构域可以定位在相对于所该至少一个识别结构域的N-末端和/或C-末端,包括定位在该至少一个识别结构域和该至少一个激活结构域之间,例如,至少一个NLS可位于一个识别结构域和另一个识别结构域和/或激活结构域之间。如果提供为可转录/可翻译载体,则STF可包含至少一个启动子,用于在目的靶细胞或细胞系统内进行最佳转录。技术人员能够根据目的细胞系统定义具有诱导型或组成型表达的合适的启动子,优选强启动子。例如,BdUbi10是植物系统(例如玉米)中一个非常强的组成型启动子。例如,BdEF1是较弱的启动子。诱导型植物启动子是四环素、地塞米松和水杨酸诱导型启动子。根据本发明合适的其他启动子是CaMV(花椰菜花叶病毒)35S或双35S启动子。其他组成型真核启动子是CMV(巨细胞病毒)、EF1a、TEF1、SV40、PGK1(人或小鼠)、Ubc(泛素1)、人类 β -肌动蛋白、GDS、GAL1或2(对于酵母系统)、CAG(包含CMV增强子、鸡 β 肌动蛋白启动子和兔 β -珠蛋白剪接受体)、H1或U6。多种诱导型启动子是技术人员公知的。

[0201] 因此,在根据本发明的STF中可以存在多种不同的架构。由于本申请的STF具有模块化特征,因此可以针对给定靶标设计几个具有不同结构域架构的STF,并且可以在体外以比较的方式对其进行评估以推断出提供最佳调控效果的架构。

[0202] 在本发明的一个实施方案中,STF包含N-末端TAL识别结构域和C-末端VP64激活结构域,其中,STF还包括在N-末端识别结构域和C-末端激活结构域之间的SV40核定位信号(NLS)。

[0203] 在本发明的另一个实施方案中,STF包含与其C-末端的SV40核定位信号(NLS)缔合的N-末端CRISPR/dCas9或CRISPR/dCpf1识别结构域以及C-末端VP64激活结构域,其中,该

STF还包含位于该N-末端识别结构域和该C-末端激活结构域之间的两个SV40 NLS。

[0204] 在本发明各个方面的优选实施方案中, STF的识别结构域是至少一个解除武装的CRISPR/Cpf1系统, 或是至少一个解除武装的CRISPR/Cpf1系统的片段, 并且该激活结构域是VPR结构域 (SEQ ID NO:276), 优选地在该识别结构域和该激活结构域之间具有接头, 优选为5xGS接头 (SEQ ID NO:277)。在本发明各个方面的另一个优选实施方案中, STF的识别结构域包括解除武装的LbCpf1结构域 (SEQ ID NO:282)、解除武装的LbCpf1_RR结构域 (SEQ ID NO:283) 和/或解除武装的LbCpf1_RVR结构域 (SEQ ID NO:284)。为了提高转录调控 (优选地为激活) 的效率, 优选靶向转录起始位点上游高达250bp区域的CRISPR/Cpf1系统的gRNA。在本发明的本文所述方面的一个实施方案中, 优选的gRNA靶向转录起始位点上游1-250、1-200、1-150、1-100、1-50、50-250、100-250、150-250或200-250bp范围内或本文所公开范围之间的任何范围的区域。

[0205] 在某些实施方案中, 根据本发明的STF或编码其的序列可以作为多重系统提供, 以靶向多个目的基因。例如, 可以通过提供多重载体, 或通过提供待在目的细胞或细胞系统在进行转化或转染的体外组装的多重STF, 设计TALE和基于解除武装的CRISPR的STF, 使其能够靶向2-7个或更多个目的基因基因座, 或利用专门设计为调控一个目的基因的两个或更多个不同的STF实现对所述一个目的基因的靶向。

[0206] 在一个实施方案中, 本发明的合成转录因子或编码其的序列可包含共价或非共价连接至合成转录因子的至少一个氨基酸序列的至少一个非天然存在的核苷酸、氨基酸或合成序列, 或其组合。该实施方案特别适合于将合成转录因子作为预组装复合物递送至目的细胞系统中的情况, 特别是对于基于解除武装的CRISPR的合成转录因子而言, 其中, 该识别结构域还包含gRNA组分。由于核糖核酸相当不稳定, 因此可以通过非天然存在的部分 (例如, 硫代磷酸酯骨架或任何其他稳定核苷酸) 来稳定gRNA识别部分。此外, 优选地在其中预组装蛋白复合物被递送到目的细胞或细胞系统中的实施方案中, 合成转录因子可包含化学修饰以稳定、衍生化或功能化该复合物, 并且/或者在旨在提供用于以靶向方式修饰细胞系统遗传物质的方法的实施方案中, 向该复合物中添加至少一个DNA修复模板。

[0207] 任何基于CRISPR的方法所面临的挑战是, 必须将RNA部分 (gRNA) 和相应的CRISPR多肽以功能性 (未降解) 方式转运至包含基因组DNA (即, DNA靶序列) 的细胞核或任何其他区室中。由于RNA不如多肽或双链DNA稳定且周转率更高, 特别是因为它很容易被核酸酶降解, 因此, 在一些实施方案中, CRISPR RNA序列和/或DNA修复模板核酸序列 (在本发明的某些实施方案中如果存在的话) 包含至少一个非天然存在的核苷酸。如果存在的话, 根据本发明的优选的骨架修饰 (提高CRISPR RNA的稳定性和/或提高DNA修复模板核酸序列的稳定性) 选自以下各项所构成的组: 硫代磷酸酯修饰、甲基膦酸酯修饰、锁核酸修饰、O-(2-甲氧基乙基) 修饰、二硫代磷酸酯修饰, 以及肽核酸修饰。值得注意的是, 所有所述骨架修饰仍然可以在两条核酸链之间形成互补碱基配对, 但是对內源核酸酶的切割的抗性更强。根据与根据本发明的RNA/DNA核酸序列组合使用的解除武装的CRISPR效应物, 可能有必要不修饰CRISPR核酸序列的那些核苷酸位置, 所述核苷酸位置与CRISPR多肽的序列非依赖性相互作用有关。所述信息可来自可用于CRISPR核酸酶/CRISPR核酸序列复合物和解除武装的CRISPR效应物 (例如, dCas9) 的可以结构信息。

[0208] 在本发明的某些实施方案中, 设想至少一个CRISPR核酸序列 (gRNA) 和/或至少一

个可选存在的DNA修复模板核酸序列可包含核苷酸和/或碱基修饰,优选地在所选择的并非所有核苷酸序列位置处。这些修饰选自由以下各项所构成的组:吡啶、胺、生物素、喀斯喀特蓝(cascade blue)、胆固醇、Cy3、Cy5、Cy5.5、Daboyl、地高辛(digoxigenin)、二硝基苯基、Edans、6-FAM、荧光素、3'-甘油基、HEX、IRD-700、IRD-800、JOE、磷酸补骨脂素(phosphate psoralen)、若丹明(rhodamine)、ROX、硫醇(SH)、间隔子、TAMRA、TET、AMCA-S⁺、SE、BODIPY[®]、Marina Blue[®]、Pacific Blue[®]、Oregon Green[®]、Rhodamine Green[®]、Rhodamine Red[®]、Rhodol Green[®]和Texas Red[®]。优选地,所述添加被掺入到CRISPR核酸序列和/或DNA修复模板核酸序列的3'或5'端。这种修饰具有有利的效果,即,可以可视化CRISPR核酸序列和/或可选存在的DNA修复模板核酸序列在细胞中的细胞定位,以研究各序列的分布、浓度和/或可用性。此外,可以研究目的合成转录因子与结合行为的相互作用。研究此类相互作用或可视化如上所述的修饰或标记的核苷酸序列的方法对于相应领域的技术人员是可用的。

[0209] 在一个实施方案中,至少一个CRISPR核酸序列的任何核苷酸或编码本发明的至少一个合成转录因子的序列的任何其他组分,可包含上述修饰之一作为标记或接头。因此,如本文所用的,“核苷酸”通常可以指碱基-糖-磷酸酯的组合。核苷酸可包含合成核苷酸。核苷酸可包含合成核苷酸类似物。核苷酸可以是核酸序列的单体单元(例如,脱氧核糖核酸(DNA)和核糖核酸(RNA))。术语核苷酸可包括腺嘌呤核苷三磷酸(ATP)、尿苷三磷酸(UTP)、三磷酸胞嘧啶(CTP)、鸟苷三磷酸(GTP)和去氧核苷三磷酸(例如,dATP、dCTP、dITP、dUTP、dGTP、dTTP或其衍生物)。此类衍生物可包括例如但不限于,[α S]dATP、7-去氮-dGTP和7-去氮-dATP,以及在含有它们的核酸分子上赋予核酸酶抗性的核苷酸衍生物。本文所用的术语核苷酸可以指双脱氧核苷三磷酸(ddNTP)及其衍生物。双脱氧核苷三磷酸的说明性实例可包括但不限于,ddATP、ddCTP、ddGTP、ddITP和ddTTP。核苷酸可以是未标记的,或通过众所周知的技术进行可检测标记的。标记也可以用量子点进行。可检测的标记可以包括,例如放射性同位素、荧光标记、化学发光标记、生物发光标记和酶标记。核苷酸的荧光标记可包括但不限于,荧光素、5-羧基荧光素(FAM)、2',7'-二甲氧基-4',5-二氯-6-羧基荧光素(JOE)、若丹明、6-羧基若丹明(R6G)、N,N,N',N'-四甲基-6-羧基若丹明(TAMRA)、6-羧基-X-若丹明(ROX)、4-(4'二甲基氨基苯基偶氮)苯甲酸(DABCYL)、喀斯喀特蓝(Cascade Blue)、俄勒冈绿(Oregon Green)、德克萨斯红(Texas Red)、花菁和5-(2'-氨基乙基)氨基萘-1-磺酸(EDANS)。

[0210] 标记或接头还可包含适于点击化学的部分,以使至少一个CRISPR向导的核酸序列或其一部分和/或DNA修复模板核酸序列和/或合成转录因子的至少一个识别结构域和/或合成转录因子的至少一个激活结构域彼此连接。

[0211] 在包含适于根据本发明修饰任何核酸或氨基酸以在体外或体内构建分子复合物的点击化学领域的反应中,一个实例是炔烃与叠氮化物进行Huisgen1,3-偶极环加成反应,形成1,4-二取代-1,2,3-三氮唑。铜(I)催化的反应温和且非常有效,在许多情况下不需要保护基,也不需要纯化。叠氮化物和炔烃官能团通常对生物分子和水性环境呈惰性。三氮唑与自然界中普遍存在的酰胺部分相似,但与酰胺不同的是它不易裂解。另外,它们几乎不可能被氧化或还原。

[0212] 如本领域技术人员所知的,适于体内反应的某些点击化学反应依赖于反应性基

团,例如叠氮化物、末端炔烃或应变炔烃(例如,二苯并环辛基(DBCO)),这些反应性基团可以通过相应修饰的核苷酸(而不是其天然对应物)引入到任何形式的RNA或DNA中。标记可以通过酶或化学方法引入。随后可以通过Cu(I)催化的炔叠氮化物(CuAAC)或无Cu(I)的应变炔叠氮化物(SPAAC)单击化学反应来处理所得的CLICK官能化DNA,其中,优选将无铜反应用于细胞或生命系统中。根据本发明,这些反应可用于引入生物素基团用于随后的纯化任务(通过叠氮化物、生物素的炔烃或含DBCO的生物素化试剂),引入荧光基团用于随后的显微成像(通过荧光叠氮化物、荧光炔烃或含有DBCO的荧光染料),或与生物分子交联,例如本发明的至少一个合成转录因子的至少一个结构域或至少一个合成转录因子,以及可选地DNA修复模板(如果存在),以共价连接和/或提供功能化的生物分子。

[0213] 在一个实施方案中,可以通过任何转化或转染方法将根据本发明的可选纯化的且在功能上缔合的5'或3'端点击化学标记的CRISPR核酸序列递送至稳定或瞬时表达相应的解除武装的CRISPR多肽的细胞或细胞系统中。因此,CRISPR核酸序列与CRISPR多肽相互作用并由此指导CRISPR多肽充当根据本发明的识别结构域。这使得激活结构域可以精确地调控至少一个目的形态发生基因的表达。

[0214] 技术人员可以使用多种其他化学反应和相应的修饰,以共价方式彼此连接本发明的核酸,或与任何氨基酸识别和/或激活结构域连接。这些修饰包括各种交联剂,例如巯基修饰,如硫辛酸N-羟基琥珀酰亚胺(NHS)酯、与伯胺(-NH₂)反应的化学基团。这些伯胺在生理pH下带正电;因此,它们主要发生在天然蛋白质三级结构的外表面,在这里它们很容易与引入水性介质的结合试剂接触。此外,在典型的生物或蛋白样品中可用的官能团中,伯胺尤其亲核;这使得它们易于靶向与几个反应性基团结合。有许多合成化学基团会与伯胺形成化学键。这些包括异硫氰酸酯、异氰酸酯、酰基叠氮化物、NHS酯、含磺酸根(-SO₃)基团的磺基-NHS酯(例如,双(磺基琥珀酰亚胺基)硫酸盐(BS₃))、磺酰氯、醛、乙二醛、环氧化物、环氧乙烷、碳酸酯、芳基卤化物、酰亚胺酯、碳二亚胺(例如,1-乙基-3-(3-二甲基氨基丙基)碳二亚胺(EDC)或二环己基碳二亚胺(DCC))、酸酐和氟苯基酯。

[0215] 在某些实施方案中,可以对根据本发明各个方面的任何核酸序列进行密码子优化,以使该序列适应于在目的靶生物体或细胞中的最佳性能。例如,可以对序列进行密码子优化,以在目的植物属的目的植物细胞中实现高转录率,或者可以对序列进行密码子优化以用于哺乳动物(例如鼠或人)细胞中。

[0216] 根据本发明的各个实施方案,该合成转录因子和/或该至少一个识别结构域可包含SEQ ID NO:1-94中任一个所示的序列,或包含与SEQ ID NO:1-94中的任一个在全长上具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性的序列,或者其中该合成转录因子和/或至少一个识别结构域结合SEQ ID NO:95-190中任一个所示的调控区域或与SEQ ID NO:95-190中的任一个在全长上具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性的序列。

[0217] 在本发明的各个方面的一个实施方案中,该合成转录因子和/或该至少一个识别结构域包含SEQ ID No:276、277、282、283、284、288、289、290中任一个所示的序列,或包含与SEQ ID No:276、277、282、283、284、288、289、290中的任一个在全长上具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性

的序列。

[0218] 优选地对WUS和/或BBM具有特异性的根据本发明的合成转录激活子可以容易地与基因编辑机器和/或T-DNA一起共递送,以改善植物细胞中的转化效率,并诱导转基因植物的再生。因此,本发明还涉及用于通过促进生长刺激基因(形态发生基因)(例如,BBM和WUS)的表达来诱导转化的植物细胞再生的方法。

[0219] 根据本文公开的各个实施方案和方面,该细胞系统可选自由至少一个真核细胞或真核生物体所构成的组,优选地,其中该至少一个真核细胞可以是至少一个植物细胞,并且/或者其中该至少一个真核生物体可以是植物或植物的一部分。

[0220] 在本文公开的某些实施方案中,待进行调控、转化和/或转染的细胞系统可选自由至少一个真核细胞或真核生物体所构成的组,优选地,其中该至少一个真核细胞可以是至少一个植物细胞,并且/或者其中该至少一个真核生物体是植物或植物的一部分。

[0221] 在根据本文公开的各个实施方案和方面的某些实施方案中,该植物的至少一部分可选自由叶、茎、根、初生胚根(emerged radicles)、花、花部分、花瓣、果实、花粉、花粉管、花药花丝、胚珠、胚囊、卵细胞、子房、合子、胚、合子胚、体细胞胚、顶端分生组织、维管束、中柱鞘、种子、根和插枝所构成的组。

[0222] 在细胞系统是或者源自植物细胞的实施方案中,该至少一个植物或植物的至少一部分可源自以下各项所构成的组的植物物种:大麦(*Hordeum vulgare*)、球茎大麦(*Hordeum bulbosum*)、高粱(*Sorghum bicolor*)、甘蔗(*Saccharum officinarum*)、玉米(*Zea mays*)、小米(*Setaria italica*)、小粒野生稻(*Oryza minuta*)、水稻(*Oryza sativa*)、澳洲野生稻(*Oryza australiensis*)、高秆野生稻(*Oryza alta*)、小麦(*Triticum aestivum*)、黑麦(*Secale cereale*)、苹果(*Malus domestica*)、二穗短柄草(*Brachypodium distachyon*)、滨海大麦草(*Hordeum marinum*)、节节麦(*Aegilops tauschii*)、胡萝卜(*Daucus glochidiatus*)、甜菜(*Beta vulgaris*)、小胡萝卜(*Daucus pusillus*)、*Daucus muricatus*、野胡萝卜(*Daucus carota*)、巨桉(*Eucalyptus grandis*)、美花烟草(*Nicotiana sylvestris*)、绒毛状烟草(*Nicotiana tomentosiformis*)、普通烟草(*Nicotiana tabacum*)、番茄(*Solanum lycopersicum*)、马铃薯(*Solanum tuberosum*)、中果咖啡(*Coffea canephora*)、葡萄(*Vitis vinifera*)、*Erythraea guttata*、*Genlisea aurea*、黄瓜(*Cucumis sativus*)、川桑(*Morus notabilis*)、*Arabidopsis arenosa*、琴叶拟南芥(*Arabidopsis lyrata*)、拟南芥(*Arabidopsis thaliana*)、近缘种须弥芥(*Crucihimalaya himalaica*)、卵叶须弥芥(*Crucihimalaya wallichii*)、弯曲碎米荠(*Cardamine flexuosa*)、北美独行菜(*Lepidium virginicum*)、荠菜(*Capsella bursa pastoris*)、*Olm arabidopsis pumila*、硬毛南芥(*Arabis hirsute*)、甘蓝型油菜(*Brassica napus*)、甘蓝(*Brassica oleracea*)、芜菁(*Brassica rapa*)、萝卜(*Raphanus sativus*)、芥菜(*Brassica juncea*)、黑芥菜(*Brassica nigra*)、芝麻菜(*Eruca vesicaria* subsp. *sativa*)、甜橙(*Citrus sinensis*)、麻疯树(*Jatropha curcas*)、毛果杨(*Populus trichocarpa*)、蒺藜苜蓿(*Medicago truncatula*)、*Cicer yamashitae*、*Cicer bijugum*、鹰嘴豆(*Cicer arietinum*)、*Cicer reticulatum*、*Cicer judaicum*、木豆(*Cajanus cajanifolius*)、蔓草虫豆(*Cajanus scarabaeoides*)、菜豆(*Phaseolus vulgaris*)、大豆(*Glycine max*)、紫云英(*Astragalus sinicus*)、百脉根(*Lotus japonicas*)、夏堇(*Torenia*

fournieri)、洋葱 (*Allium cepa*)、大葱 (*Allium fistulosum*)、大蒜 (*Allium sativum*) 和韭菜 (*Allium tuberosum*)。

[0223] 在本发明的另一方面,提供了一种用于提高细胞系统中的转化效率的方法,其中,该方法可包括以下步骤:(a) 提供细胞系统;(b) 在该细胞系统中引入至少一个合成转录因子或编码其的核苷酸序列;以及(c) 在该细胞系统中引入至少一个目的核苷酸序列;(d) 可选地:在获得该细胞系统的转化后代的条件下培养该细胞系统;其中该至少一个合成转录因子或编码其的核苷酸序列包含至少一个识别结构域和至少一个激活结构域,其中该合成转录因子配置为调控该细胞系统中的至少一个形态发生基因的表达,优选地为调控该细胞系统中的至少一个形态发生基因的转录;并且其中该至少一个合成转录因子或编码其的核苷酸序列的引入与至少一个目的核苷酸序列的引入平行进行或序贯进行。

[0224] 因此,本发明公开了通过使用对可以重编程细胞并诱导多种植物物种细胞分裂的内源性形态发生基因具有特异性的合成转录因子来提高植物转化或转染以及/或植物再生效率的方法,从而提供可靠的细胞系统转化方法,包括那些公知难以通过当前可用的方法进行修饰和/或转化的细胞系统。特别地,包括极有价值的优良事件(即,很少发生的事件,并且如果有的话,源于非同寻常的事件以及令人惊讶的事件)的某些优良品系和所述优良品系的种质可能对体外培养和转化高度抵制。此类基因型通常在用于由通常合适的外植体(例如未成熟胚)引起这种反应的培养基上不产生适当的胚胎形成或器官发生培养反应。此外,当将外源性DNA或其他生物分子引入这些未成熟胚中时,在繁琐的选择轮次之后可能无法恢复成功的修饰事件,或者仅仅非常少的事件可以恢复,因此此类基因型的转化不切实际。

[0225] 在一个实施方案中,该方法可包括,使(a) 该至少一个合成转录因子或编码其的序列,或该至少一个合成转录因子的至少一个组分或编码其的序列;以及(b) 该至少一个目的核苷酸序列,通过独立地选自生物方法和/或物理方法的方法引入该细胞系统中,该方法包括转染、转化,包括通过农杆菌属物种(优选地为根癌农杆菌)、病毒载体、生物弹轰击进行转化,使用化学试剂进行转染(包括聚乙二醇转染),电穿孔,细胞融合或其任何组合。

[0226] 因此,“引入”或“引入”的过程可包括将生物分子引入或递送到目的细胞系统中的任何生物、化学和/或物理方法。值得注意的是,可以使用引入或递送技术的任何组合。此外,可以通过相同的技术同时或随后例如通过共轰击来引入待引入到目的细胞系统中的不同组分,或者可以通过不同的引入技术同时或随后将它们引入。

[0227] 在本发明的上下文中首次证明,基于Cpf1的转录调控系统是用于植物中内源性靶基因的转录激活或抑制的强大工具,并且如上所述,与其他方法相比,它具有许多优点。因此,其可用于,通过使用对内源性形态发生基因具有特异性的合成转录因子提供转化细胞系统(包括已知难以进行修饰和/或难以通过当前可以的方法进行转化的那些细胞系统)的方法,提高植物转化或转染和/或植物再生的效率。

[0228] 在本发明的用于提高细胞系统中转化效率的方法的优选实施方案中,该至少一个识别结构域是至少一个解除武装的非功能性CRISPR/核酸酶系统,或是至少一个解除武装的非功能性CRISPR/核酸酶系统的片段。

[0229] 在本发明方法的另一个优选实施方案中,该至少一个解除武装的CRISPR/核酸酶系统是CRISPR/dCpf1系统,其中,该至少一个解除武装的CRISPR/核酸酶系统包含至少一个

向导RNA。

[0230] 在一个实施方案中,该至少一个合成转录因子的至少一个激活结构域选自由酸性转录激活结构域所构成的组,优选地,其中该至少一个激活结构域来自水稻白叶枯病菌的无毒基因、单纯疱疹病毒的VP16或四聚体VP64、VPR、SAM、Scaffold、Suntag、P300、VP160或其任何组合。优选地,该激活结构域是VPR结构域(SEQ ID NO:276)。

[0231] 在另一个实施方案中,该至少一个合成转录因子的至少一个激活结构域相对于该至少一个合成转录因子的至少一个识别结构域位于N-末端和/或C-末端。

[0232] 在本发明方法的优选实施方案中,STF的识别结构域是至少一个解除武装的CRISPR/Cpf1系统,或是至少一个解除武装的CRISPR/Cpf1系统的片段,并且该激活结构域是VPR结构域,可选地在该识别结构域和该激活结构域之间具有接头,优选地为5xGS接头。

[0233] 根据本发明的各个方面和实施方案的转化效率提高可包括与对照植物或细胞系统相比较时的任何统计学上的显著增加。例如,转化效率的提高可包括与对照植物或对照植物部分或对照细胞系统相比较,增加约0.2%、0.5%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、120%、125%或更多。可替换地,转化效率的提高可包括,与对照植物或植物部分或细胞系统相比,该植物、植物部分或细胞系统中的转化效率增加约0.2倍、0.5倍、1倍、2倍、4倍、8倍、16倍或32倍或更多。

[0234] 在一个实施方案中,本发明的方法可包括,提供至少一个目的核苷酸序列作为至少一个载体的一部分或作为至少一个线性分子。

[0235] 在本文公开的方法的一个实施方案中,该至少一个目的核苷酸序列可以选自由转基因、修饰的内源基因、合成序列、内含子序列、编码序列或调控序列所构成的组。

[0236] 在本文公开的方法的一个实施方案中,该至少一个目的核苷酸序列可以是转基因,其中,该转基因可包含编码目的生物体的基因组的基因或所述基因的至少一部分的核苷酸序列。

[0237] 在一个实施方案中,根据本发明的调控序列可以是启动子序列,其中,该启动子的编辑或突变或调控包括用不同的启动子(也称为替换启动子)或启动子片段(也称为替换启动子片段)替换启动子或启动子片段,其中,启动子替换会造成以下任一项或以下任意组合的结果:启动子活性提高、启动子组织特异性提高、启动子活性降低,启动子组织特异性降低、新的启动子活性、诱导型启动子活性、基因表达窗口延长、改变同一细胞层或其他细胞层中基因表达的时间或发育进度(例如,延长花药绒毡层中基因表达的时间)、DNA结合元件突变,以及/或者DNA结合元件缺失或添加。待修饰的启动子(或启动子片段)可以是对于正在编辑的细胞是内源性的、人工的、预先存在的或转基因的启动子(或启动子片段)。替换启动子(或启动子片段)可以是对于正在编辑的细胞是内源性的、人工的、预先存在的或转基因的启动子(或启动子片段)。根据本公开的任何其它调控序列都可以按照针对上述启动子或启动子片段所详述的内容进行修饰。

[0238] 特别是在待修饰的植物基因组的情况下,可能希望的是,由本发明的方法介导的修饰不会通过将外源性DNA以不精确的方式(例如必须要关注的环境、监管和政治问题)整合到亲本基因组中而产生遗传修饰的生物体。因此,根据本发明提供用于以瞬时的方式将目的遗传物质引入细胞系统的方法的实施方案特别适用于无需插入外源性DNA从而不需提

供视为转基因生物体的细胞或生物体就可提供在预定位置进行修饰的细胞系统,因为执行本发明方法所必需的所有工具都可以以活性形式以瞬时的方式提供给细胞系统。

[0239] 在本文所述的方法的一个实施方案中,转录激活以完全瞬时的方式与植物基因组的修饰结合,从而无需将外源性DNA插入植物基因组中,因而也不需提供被视为经基因修饰生物体的植物生物体,就可获得在预定遗传位置处包含修饰的植物生物体。因此,本文所述的方法提供了无需劳动密集型调控程序即可修饰植物基因组的方法。在本文所述方法的又一个实施方案中,STF和/或位点特异性核酸酶是无DNA的,例如蛋白或RNP,从而提供监管益处。在本文公开的各种方法的一个实施方案中,该方法可以以完全瞬时的方式执行。在其他实施方案中,可以通过稳定和瞬时方法的组合来执行该方法。在又一个实施方案中,该方法还可以通过将合适的递送工具稳定地引入到目的细胞或细胞系统中来执行。

[0240] 在本发明各个方面的另一个实施方案中,待引入细胞系统的至少一个目的核苷酸序列可以是目的生物体的转基因,其中,该转基因或转基因的一部分可选自由以下各项所构成的组:编码对非生物胁迫的抗性或耐受性的基因,所述非生物胁迫包括干旱胁迫、渗透胁迫、热胁迫、冷胁迫、氧化胁迫、重金属胁迫、氮缺乏、磷缺乏、盐胁迫或渍水;编码除草剂抗性的基因,所述除草剂抗性包括草甘膦抗性、草铵膦(glufosinate/phosphinotricin)抗性、潮霉素抗性、2,4-D抗性或耐受性、原卟啉原氧化酶(PPO)抑制剂抗性或耐受性、ALS抑制剂抗性或耐受性和麦草畏抗性或耐受性;编码对生物胁迫的抗性或耐受性的基因,包括病毒抗性基因、真菌抗性基因、细菌抗性基因、昆虫抗性基因;或编码产量相关性状的基因,所述产量相关性状包括抗倒伏性、开花时间、抗落粒性、种子颜色、胚乳组成或营养含量。

[0241] 在本发明各个方面的另一个实施方案中,该至少一个目的核苷酸序列可以是目的生物体的修饰的内源基因的至少一部分,其中该修饰的内源基因,与未经修饰的内源基因的核苷酸序列相比,可包含至少一个核苷酸的至少一个缺失、插入和/或取代。

[0242] 在本发明各个方面的又一个实施方案中,该至少一个目的核苷酸序列可以是目的生物体的修饰的内源基因的至少一部分,其中该修饰的内源基因可包含编码该修饰的内源基因的结构域的至少一个核苷酸位置的截短、重复、取代和/或缺失中的至少一种。

[0243] 在一个实施方案中,该至少一个目的核苷酸序列可以是调控序列的至少一部分,其中该调控序列可包括核心启动子序列、近端启动子序列、顺式调控序列、反式调控序列、基因座控制序列、绝缘子序列、沉默子序列、增强子序列、终止子序列或其任何组合中的至少一个。

[0244] 本文下文公开的任何合成转录因子均可用于根据本发明的不同方法中,作为介体,以特异性调控目的形态发生基因的转录。这种调控(优选地为转录上调)使得可以实现更好的细胞系统(优选地为目的植物或植物部分)转化效率。

[0245] 根据本文公开方法的各个实施方案,该待调控的优选形态发生基因可选自由BBM、WUS(包括WUS2)、WOX基因、WUS或BBM同源物、Lec1、Lec2、WIND1、ESR1、PLT3、PLT5、PLT7、IPT、IPT2、Knotted1和RKD4所构成的组。

[0246] 优选地,该形态发生基因包含选自由以下各项所构成的组的核苷酸序列:(i) SEQ ID NO:199-237中任一个所示的核苷酸序列,(ii) 具有SEQ ID NO:199-237中任一个所示的核苷酸序列的编码序列的核苷酸序列,(iii) 与(i)或(ii)的核苷酸序列互补的核苷酸序列,(iv) 与(i)、(ii)或(iii)的核苷酸序列具有至少50%、60%、70%、75%、80%、85%、

90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性(优选在全长上)的核苷酸序列,(v)在严格条件下与(iii)的核苷酸序列杂交的核苷酸序列,(vi)编码包含SEQ ID NO:238-258中任一个所示的氨基酸序列的蛋白的核苷酸序列,(vii)编码包含与SEQ ID NO:238-258中的任一个所示的序列至少50%、60%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的氨基酸序列的蛋白的核苷酸序列,或(viii)编码包含SEQ ID NO:238-258中任一个所示的氨基酸序列的蛋白的同源物、类似物或直系同源物的核苷酸序列。

[0247] 在某些实施方案中,本发明方法中使用的合成转录因子可配置为通过结合位于相对于起始密码子一定距离处的调控区域来调控形态发生基因的表达,优选地为调控形态发生基因的转录。

[0248] 在某些实施方案中,本发明的方法中使用的合成转录因子和/或至少一个识别结构域可包含SEQ ID No:1-94中任一个所示的序列,或包含与SEQ ID No:1-94中任一个在全长上具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性的序列,或者其中该合成转录因子和/或至少一个识别结构域结合SEQ ID NO:95-190中任一个所示的调控区域或与SEQ ID NO:95-190中的任一个在全长上具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性的序列。

[0249] 在本发明方法的一个实施方案中,该合成转录因子和/或该至少一个识别结构域包含SEQ ID No:276、277、282、283、284、288、289、290中任一个所示的序列,或包含与SEQ ID No:276、277、282、283、284、288、289、290中的任一个在全长上具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性的序列。

[0250] 在本发明方法的某些实施方案中,该细胞系统可选自至少一个真核细胞或真核生物体,优选地,其中该至少一个真核细胞是至少一个植物细胞,并且/或者其中该至少一个真核生物体是植物或植物的一部分。

[0251] 在本发明方法的其它实施方案中,该植物的至少一部分选自由叶、茎、根、初生胚根(emerged radicles)、花、花部分、花瓣、果实、花粉、花粉管、花药花丝、胚珠、胚囊、卵细胞、子房、合子、胚、合子胚、体细胞胚、顶端分生组织、维管束、中柱鞘、种子、根和插枝所构成的组。

[0252] 在本发明方法的其它实施方案中,该至少一个植物细胞、该至少一个植物或植物的至少一部分可源自选自以下的植物物种:大麦(*Hordeum vulgare*)、球茎大麦(*Hordeum bulbosum*)、高粱(*Sorghum bicolor*)、甘蔗(*Saccharum officinarum*)、玉米(*Zea mays*)、小米(*Setaria italica*)、小粒野生稻(*Oryza minuta*)、水稻(*Oryza sativa*)、澳洲野生稻(*Oryza australiensis*)、高秆野生稻(*Oryza alta*)、小麦(*Triticum aestivum*)、黑麦(*Secale cereale*)、苹果(*Malus domestica*)、二穗短柄草(*Brachypodium distachyon*)、滨海大麦草(*Hordeum marinum*)、节节麦(*Aegilops tauschii*)、胡萝卜(*Daucus glochidiatus*)、甜菜(*Beta vulgaris*)、小胡萝卜(*Daucus pusillus*)、*Daucus muricatus*、野胡萝卜(*Daucus carota*)、巨桉(*Eucalyptus grandis*)、美花烟草(*Nicotiana sylvestris*)、绒毛状烟草(*Nicotiana tomentosiformis*)、普通烟草(*Nicotiana*

tabacum)、番茄(*Solanum lycopersicum*)、马铃薯(*Solanum tuberosum*)、中果咖啡(*Coffea canephora*)、葡萄(*Vitis vinifera*)、*Erythrante guttata*、*Genlisea aurea*、黄瓜(*Cucumis sativus*)、川桑(*Morus notabilis*)、*Arabidopsis arenosa*、琴叶拟南芥(*Arabidopsis lyrata*)、拟南芥(*Arabidopsis thaliana*)、近缘种须弥芥(*Crucihimalaya himalaica*)、卵叶须弥芥(*Crucihimalaya wallichii*)、弯曲碎米荠(*Cardamine flexuosa*)、北美独行菜(*Lepidium virginicum*)、荠菜(*Capsella bursa pastoris*)、*Olm**arabidopsis pumila*、硬毛南芥(*Arabis hirsute*)、甘蓝型油菜(*Brassica napus*)、甘蓝(*Brassica oleracea*)、芜菁(*Brassica rapa*)、萝卜(*Raphanus sativus*)、芥菜(*Brassica juncea*)、黑芥菜(*Brassica nigra*)、芝麻菜(*Eruca vesicaria* subsp.*sativa*)、甜橙(*Citrus sinensis*)、麻疯树(*Jatropha curcas*)、毛果杨(*Populus trichocarpa*)、蒺藜苜蓿(*Medicago truncatula*)、*Cicer yamashitae*、*Cicer bijugum*、鹰嘴豆(*Cicer arietinum*)、*Cicer reticulatum*、*Cicer judaicum*、木豆(*Cajanus cajanifolius*)、蔓草虫豆(*Cajanus scarabaeoides*)、菜豆(*Phaseolus vulgaris*)、大豆(*Glycine max*)、紫云英(*Astragalus sinicus*)、百脉根(*Lotus japonicas*)、夏堇(*Torenia fournieri*)、洋葱(*Allium cepa*)、大葱(*Allium fistulosum*)、大蒜(*Allium sativum*)和韭菜(*Allium tuberosum*)。

[0253] 在本发明的另一方面，独立地或与本文公开的其他方面和实施方案一起，提供了一种在预定位置修饰细胞系统的遗传物质的方法，其中该方法可以包括以下步骤：(a) 提供细胞系统；(b) 将至少一个合成转录因子或编码其的序列引入该细胞系统中，(c) 进一步向该细胞系统引入：(i) 至少一个位点特异性核酸酶或编码其的序列，其中，该位点特异性核酸酶在预定位置诱导双链断裂；(ii) 可选的：至少一个目的核苷酸序列，优选地侧翼为与细胞系统遗传物质中预定位置相邻的一个或多个核苷酸序列互补的一个或多个同源序列；以及(e) 可选地：确定修饰存在于细胞系统遗传物质中的预定位置；以及(f) 获得在该细胞系统遗传物质的预定位置处进行修饰的细胞系统；其中该至少一个合成转录因子或编码其的核苷酸序列包含至少一个识别结构域和至少一个激活结构域，其中该合成转录因子配置为调控该细胞系统中的至少一个形态发生基因的表达，优选地为调控该细胞系统中的至少一个形态发生基因的转录；并且其中该至少一个合成转录因子或编码其的核苷酸序列的引入与至少一个位点特异性核酸酶或编码其的序列以及可选的至少一个目的核苷酸序列的引入平行，或序贯进行。

[0254] 因此，该方面和相关实施方案协同地整合了细胞系统中至少一个目的形态发生基因的转录率的靶向调控与将某些效应物引入细胞的高度定点基因组编辑(GE)方法的优势。通过在包含至少一个根据本发明的合成转录因子的细胞系统内提供环境，可以在引入至少一个位点特异性核酸酶(SSN)(即，包含DNA双链或DNA单链切割能力的酶)或编码其的序列，以及可选地提供环境的其它工具(如修复模板(RTs))之前或同时，特异性调控细胞系统中至少一个形态发生基因的转录，其中该细胞系统具有高度的转化能力，并具有较高的再生能力。这些因子保证了目的细胞系统中此类编辑的遗传物质的成功编辑和再生，还使得可以有经修饰的细胞系统再生植物或植物材料，因为基于与至少一个合成转录因子或编码其的序列的共处理或预处理，该细胞系统在GE事件期间具有更大的耐受性和生存能力。

[0255] 在一个实施方案中，该方法还包括以下步骤：在获得经修饰细胞系统的基因修饰

后代的条件下培养该细胞系统。

[0256] 本文在预定位置和一个或多个同源区域的上下文中使用的术语“相邻”或“相邻于”可包括上游和下游相邻区域,或两者。因此,相邻区域根据待被修饰的细胞系统的遗传物质来确定,所述物质包括预定位置。

[0257] 在预定位置附近可能存在上游和/或下游相邻区域。对于诱导钝性双链断裂(DSB)的位点特异性核酸酶(SSN)，“预定位置”代表DSB在目的细胞系统的遗传物质内被诱导的位点。对于在DSB诱导后留下突出端的SSN,预定位置是指一条链上5'端的切割与另一链上3'端的切割之间的区域。因此,对于粘性末端SSN,可以使用两条不同的DNA链作为参考来计算相邻区域。因此,术语“与预定位置相邻”可以表示待修饰的遗传物质中的上游和/或下游核苷酸位置,其中在诱导DSB或修饰之前,基于细胞系统的遗传物质来定义相邻区域。根据SSN诱导DSB的不同机制,“预定位置”意味着在目的遗传物质中进行修饰的位置可能因此暗示了对于钝DSB在同一链上的一个特定位置,或对于粘性切割DSB在不同链上的两个切割位点之间的区域,或对于用作SSN的切口酶在一条链中5'位置的切口与另一链上3'位置的切口之间。

[0258] 如果存在,则上游相邻区域是指在启动双链断裂之前,例如在靶向基因组工程化期间,参照预定位置,定义目的位点特异性核酸酶切位点5'端的直接上游区域。相应地,下游相邻区域在启动双链断裂之前(例如,在靶向基因组工程化期间),参照预定位置定义目的SSN切割位点3'端的直接下游区域。5'端和3'端可以相同,这取决于目的位点特异性核酸酶。

[0259] 在某些实施方案中,在远离要诱导的DSB的距离处设计至少一个同源区域也可能是有利的,即,不直接位于预定位置/DSB位置的两侧。在这种情况下,在预定的位置和同源序列(同源臂)之间的基因组序列将在发生同源重组后“缺失”,这对于某些策略可能是优选的,因为这使得可以在DSB附近有针对性地删除序列。因此,根据本发明,对于那些依赖于RT的实施方案,可以设想不同种类的RT配置和设计。RT可以用于引入位点特异性突变,或者RT可以用于目的核酸序列的位点特异性整合,或者RT可以用于辅助靶向的缺失。

[0260] 引入的“同源序列”和相应的“相邻区域”各自可具有从约15bp至约15000bp的变化且不同的长度,即,上游同源区域与下游同源区域相比可以具有不同的长度。可能仅存在一个同源区域。同源区域的长度没有真正的上限,该长度由实际和技术问题决定。根据某些实施方案,依据RT的性质和待引入的靶向修饰,可以优选不对称同源区域,即,其中上游和下游侧翼区域具有可变长度的同源区域。在某些实施方案中,可以仅存在一个上游和下游侧翼区。

[0261] 在根据本发明方法的一个实施方案中,该至少一种位点特异性核酸酶可包含锌指核酸酶、转录激活因子样核酸酶、CRISPR/Cas系统(包括CRISPR/Cas9系统、CRISPR/Cfp1系统、CRISPR/CasX系统、CRISPR/CasY系统)、工程化的归巢核酸内切酶和大范围核酸酶,和/或其任何组合,及其变体或催化活性片段。

[0262] 一旦表达,则Cas9蛋白和gRNA通过gRNA“支架”结构域与Cas9表面暴露的带正电荷的凹槽之间的相互作用形成核糖核蛋白复合物。Cas9在gRNA结合后经历构象变化,其将分子从非活性的非DNA结合构象转变为活性的DNA结合构象。重要的是,gRNA的“间隔子(spacer)”序列仍可自由与靶DNA相互作用。Cas9-gRNA复合物将使与PAM结合任何基因组序

列,但gRNA间隔子与靶DNA匹配的程度决定了Cas9是否会切割。一旦Cas9-gRNA复合物与推定DNA靶标结合,则在gRNA靶序列3'端的“种子”序列开始退火至靶DNA。如果种子和靶DNA序列匹配,则gRNA将继续沿3'至5'方向(相对于gRNA的极性)退火至靶DNA。

[0263] 当正确设计gRNA时,CRISPR/Cas(例如,CRISPR/Cas9)以及同样的CRISPR/Cpf1或CRISPR/CasX或CRISPR/CasY和其它CRISPR系统都是高度特异性的,但特别地,特异性仍然是一个主要问题,特别是对于临床用途或基于该CRISPR技术的靶向植物GE而言。CRISPR系统的特异性在很大程度上取决于与基因组其余部分相比,gRNA靶向序列对基因组靶标的特异性。因此,当与使用至少一种CRISPR核酸酶作为位点特异性核酸酶结合并进一步与使用合适的CRISPR核酸结合时,根据本发明的方法可提供GE的显著更可预测的结果。尽管CRISPR复合物可以在特定位点介导细胞或细胞系统的基因组或遗传物质的高精度切割,但是本文提出的方法提供了保证可编程和可预测的修复机制的附加控制机制。

[0264] 根据本发明的各个实施方案,关于共价和非共价缔合或附着的上述公开内容也适用于CRISPR核酸序列,其可以包含一个以上的部分,例如crRNA和tracrRNA部分,如上所述,它们可以彼此缔合。在一个实施方案中,可以将本发明的RT核酸序列置于目的CRISPR核酸序列内,以形成根据本发明的杂合核酸序列,该杂合体可以通过共价和非共价缔合形成。

[0265] 在根据本发明各个方面的又一个实施方案中,预定位置处的至少一个目的核酸序列侧翼的一个或多个核酸序列与邻近预定位置(位于预定位置的上游和/或下游)的一个或多个核酸序列在各个邻近区域的全长上至少85%至100%互补。

[0266] 值得注意的是,可以使用具有较低程度的同源性或互补性的至少一个侧翼区域,例如,与目的遗传物质中的至少一个相邻区域的同源性/互补性为至少70%、至少75%、至少80%、至少81%、至少82%、至少83%或至少84%。对于依赖于HDR模板的高精度GE(即,本文所公开的RT)而言,超过95%的同源性/互补性有利于实现高度特异性的修复事件。如Runnitz等,《分子生物学杂志》(Mol. Cell Biol.), 1984, 4 (11), 2253-2258, 序列同源性非常低也可能足以获得同源重组。如技术人员所知,互补程度取决于待修饰的遗传物质、计划编辑的性质、基因组的复杂性和大小、潜在脱靶位点的数量、遗传背景,以及待修饰的细胞或细胞系统中的环境。

[0267] 在一个实施方案中,该方法还包括在在获得该经修饰的细胞系统的遗传修饰后代的条件下培养该细胞系统的步骤。

[0268] 在根据本发明的各个方面的又一个实施方案中,细胞系统的遗传物质可以选自由以下各项所构成的组:原生质体、转移于重组宿主细胞中的病毒基因组、真核细胞、组织或器官(优选植物细胞、植物组织或植物器官),以及真核生物体(优选植物生物体)。

[0269] 在本发明方法的一个实施方案中,(i)至少一个合成转录因子或编码其的序列,或该至少一个合成转录因子的至少一个组分或编码其的序列;以及(ii)该至少一个位点特异性核酸酶或包含其的序列;以及可选地(iii)该至少一个目的核苷酸序列,可通过独立地选自生物方法和/或物理方法的方法引入该细胞系统中,该方法包括转染、转化,其中包括通过农杆菌属物种(优选地为根癌农杆菌)、病毒载体、生物弹轰击进行转化,使用化学试剂进行转染(包括聚乙二醇转染)或其任何组合。

[0270] 在本发明的用于在预定位置修饰细胞系统遗传物质的方法的实施方案中,该至少一个识别结构域可以是选自由以下各项多构成的组的分子,或分子的片段:至少一个TAL效

应物、至少一个解除武装的CRISPR/nuclease系统、至少一个锌指结构域和至少一个解除武装的归巢核酸内切酶或其任意组合。

[0271] 在本发明的用于在预定位置修饰细胞系统的遗传物质的方法的一个实施方案中,该至少一个解除武装的CRISPR/核酸酶系统可选自CRISPR/dCas9系统、CRISPR/dCpf1系统、CRISPR/dCasX系统或CRISPR/dCasY系统,或其任何组合,其中该至少一个解除武装的CRISPR/核酸酶系统可包含至少一个向导RNA,优选是针对特异性的解除武装的CRISPR/核酸酶系统和形态发生系统内或附近的特异性靶位点优化的向导RNA,以增加本发明合成转录因子的识别和/或结合特性。

[0272] 在本发明的用于在预定位置修饰细胞系统的遗传物质的方法的优选实施方案中,该至少一个识别结构域是至少一个解除武装的CRISPR/核酸酶系统,或是至少一个解除武装的CRISPR/核酸酶系统的片段。

[0273] 由于上述优点,特别优选的是,在本发明的在预定位置处修饰细胞系统的遗传物质的方法中,该至少一个解除武装的CRISPR/核酸酶系统是CRISPR/dCpf1系统,其中该至少一个解除武装的CRISPR/核酸酶系统包含至少一个向导RNA。

[0274] 在本发明的用于在预定位置修饰细胞系统的遗传物质的方法的另一个实施方案中,该至少一个合成转录因子的至少一个激活结构域可选自由酸性转录激活结构域所构成的组,优选地,其中该至少一个激活结构域来自水稻白叶枯病菌的无毒基因、单纯疱疹病毒的VP16或四聚体VP64、VPR、SAM、Scaffold、Suntag、P300、VP160或其任何组合。在本发明的在预定位置修饰细胞系统的遗传物质的方法的优选实施方案中,该至少一个激活结构域是VPR (SEQ ID NO:276)。在本发明的另一个优选实施方案中,可以使用不同激活结构域的组合,例如,VP64-p65-Rita或本领域通常已知的激活结构域的任何组合。

[0275] 用于本文所述的CRISPR/Cpf1系统的合适的接头包括柔性接头,例如,5GS或XTEN,而可体内裂解的接头不适合于本文描述的本发明的各个方面。

[0276] 为了提高转录调控(优选地为激活)的效率,优选靶向转录起始位点上游高达250bp的区域的CRISPR/Cpf1系统的gRNA。在本文所述的本发明各方面的一个实施方案中,优选的gRNA靶向转录起始位点上游1-250、1-200、1-150、1-100、1-50、50-250、100-250、150-250或200-250bp,或本文公开范围之间的任何范围的区域。

[0277] 在本发明的用于在预定位置修饰细胞系统的遗传物质的方法的另一个实施方案中,该至少一个合成转录因子的至少一个激活结构域可相对于该至少一个合成转录因子的至少一个识别结构域位于N-末端和/或C-末端。

[0278] 在本发明的用于修饰细胞系统遗传物质的方法的优选实施方案中,STF的识别结构域是至少一个解除武装的CRISPR/Cpf1系统,或是至少一个解除武装的CRISPR/Cpf1系统的片段,并且该激活结构域是VPR结构域,优选地在该识别结构域和该激活结构域之间存在接头,优选地为5xGS接头。

[0279] 在本发明的用于在预定位置修饰细胞系统的遗传物质的方法的再一个实施方案中,该至少一个形态发生基因可选自由BBM、WUS (包括WUS2)、WOX基因、WUS或BBM同源物、Lec1、Lec2、WIND1、ESR1、PLT3、PLT5、PLT7、IPT、IPT2、Knotted1和RKD4所构成的组。

[0280] 在另一个实施方案中,提供了本发明的用于在预定位置修饰细胞系统的遗传物质的方法,其中该至少一个形态发生基因包含选自由以下各项所构成的组的核苷酸序列:(i)

SEQ ID NO:199-237中任一个所示的核苷酸序列, (ii) 具有SEQ ID NO:199-237中任一个所示的核苷酸序列的编码序列的核苷酸序列, (iii) 与(i) 或(ii) 的核苷酸序列互补的核苷酸序列, (iv) 与(i)、(ii) 或(iii) 的核苷酸序列具有至少50%、60%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性(优选在全长上)的核苷酸序列, (v) 在严格条件下与(iii) 的核苷酸序列杂交的核苷酸序列, (vi) 编码包含SEQ ID NO:238-258中任一个所示的氨基酸序列的蛋白的核苷酸序列, (vii) 编码包含与SEQ ID NO:238-258中任一个所示的序列至少50%、60%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的氨基酸序列的蛋白的核苷酸序列, 或(viii) 编码包含SEQ ID NO:238-258中任一个所示的氨基酸序列的蛋白的同源物、类似物或直系同源物的核苷酸序列。

[0281] 在本发明的用于在预定位置修饰细胞系统的遗传物质的方法的另一个实施方案, 该合成转录因子可配置为通过结合位于相对于起始密码子一定距离处的调控区域来调控形态发生基因的表达, 优选地为调控形态发生基因的转录。

[0282] 在本发明的用于在预定位置修饰细胞系统遗传物质的方法的一个实施方案中, 该合成转录因子和/或该至少一个识别结构域包含SEQ ID NO:1-94中任一个所示的序列, 或包含与SEQ ID NO:1-94中的任一个在全长上具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性的序列, 或者其中该合成转录因子和/或至少一个识别结构域结合SEQ ID NO:95-190中任一个所示的调控区域, 或与SEQ ID NO:95-190中的任一个在全长上具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性的序列。

[0283] 在本发明的用于在预定位置修饰细胞系统的遗传物质的方法的一个实施方案中, 该合成转录因子和/或该至少一个识别结构域包含SEQ ID NO:276、277、282、283、284、288、289、290中任一个所示的序列, 或包含与SEQ ID NO:276、277、282、283、284、288、289、290中的任一个在全长上具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性的序列。

[0284] 在本发明的用于在预定位置修饰细胞系统的遗传物质的方法的另一个实施方案中, 该细胞系统可选自由至少一个真核细胞或真核生物体所构成的组, 优选地, 其中该至少一个真核细胞可以是至少一个植物细胞, 并且/或者其中该至少一个真核生物体可以是植物或植物的一部分。

[0285] 在一个实施方案中, 该植物的至少一部分选自由叶、茎、根、初生胚根(emerged radicles)、花、花部分、花瓣、果实、花粉、花粉管、花药花丝、胚珠、胚囊、卵细胞、子房、合子、胚、合子胚、体细胞胚、顶端分生组织、维管束、中柱鞘、种子、根和插枝所构成的组。

[0286] 在另一个实施方案中, 该至少一个植物细胞、该至少一个植物或植物的至少一部分源自选自以下的植物物种: 大麦(*Hordeum vulgare*)、球茎大麦(*Hordeum bulbosum*)、高粱(*Sorghum bicolor*)、甘蔗(*Saccharum officinarum*)、玉米(*Zea mays*)、小米(*Setaria italica*)、小粒野生稻(*Oryza minuta*)、水稻(*Oryza sativa*)、澳洲野生稻(*Oryza australiensis*)、高秆野生稻(*Oryza alta*)、小麦(*Triticum aestivum*)、黑麦(*Secale cereale*)、苹果(*Malus domestica*)、二穗短柄草(*Brachypodium distachyon*)、滨海大麦草(*Hordeum marinum*)、节节麦(*Aegilops tauschii*)、胡萝卜(*Daucus glochidiatus*)、甜菜

(*Beta vulgaris*)、小胡萝卜 (*Daucus pusillus*)、*Daucus muricatus*、野胡萝卜 (*Daucus carota*)、巨桉 (*Eucalyptus grandis*)、美花烟草 (*Nicotiana sylvestris*)、绒毛状烟草 (*Nicotiana tomentosiformis*)、普通烟草 (*Nicotiana tabacum*)、番茄 (*Solanum lycopersicum*)、马铃薯 (*Solanum tuberosum*)、中果咖啡 (*Coffea canephora*)、葡萄 (*Vitis vinifera*)、*Erythrante guttata*、*Genlisea aurea*、黄瓜 (*Cucumis sativus*)、川桑 (*Morus notabilis*)、*Arabidopsis arenosa*、琴叶拟南芥 (*Arabidopsis lyrata*)、拟南芥 (*Arabidopsis thaliana*)、近缘种须弥芥 (*Crucihimalaya himalaica*)、卵叶须弥芥 (*Crucihimalaya wallichii*)、弯曲碎米荠 (*Cardamine flexuosa*)、北美独行菜 (*Lepidium virginicum*)、荠菜 (*Capsella bursa pastoris*)、*Olm arabidopsis pumila*、硬毛南芥 (*Arabis hirsute*)、甘蓝型油菜 (*Brassica napus*)、甘蓝 (*Brassica oleracea*)、芜菁 (*Brassica rapa*)、萝卜 (*Raphanus sativus*)、芥菜 (*Brassica juncea*)、黑芥菜 (*Brassica nigra*)、芝麻菜 (*Eruca vesicaria* subsp. *sativa*)、甜橙 (*Citrus sinensis*)、麻疯树 (*Jatropha curcas*)、毛果杨 (*Populus trichocarpa*)、蒺藜苜蓿 (*Medicago truncatula*)、*Cicer yamashitae*、*Cicer bijugum*、鹰嘴豆 (*Cicer arietinum*)、*Cicer reticulatum*、*Cicer judaicum*、木豆 (*Cajanus cajanifolius*)、蔓草虫豆 (*Cajanus scarabaeoides*)、菜豆 (*Phaseolus vulgaris*)、大豆 (*Glycine max*)、紫云英 (*Astragalus sinicus*)、百脉根 (*Lotus japonicas*)、夏堇 (*Torenia fournieri*)、洋葱 (*Allium cepa*)、大葱 (*Allium fistulosum*)、大蒜 (*Allium sativum*) 和韭菜 (*Allium tuberosum*)。

[0287] 在本发明的用于在预定位置修饰细胞系统的遗传物质的方法的又一个实施方案中，预定位置处的至少一个目的核苷酸序列侧翼的一个或多个核苷酸序列与邻近预定位置（位于预定位置的上游和/或下游）的一个或多个核苷酸序列在各个邻近区域的全长上至少85%至100%互补。

[0288] 在本发明的用于在预定位置修饰细胞系统的遗传物质的方法的一个实施方案中，该至少一个目的核苷酸序列可选自由以下各项所构成的组：转基因、修饰的内源基因、合成序列、内含子序列、编码序列或调控序列。如果该至少一个目的核苷酸序列是转基因，则该转基因可包含编码目的生物体的基因组的基因或所述基因的至少一部分的核苷酸序列。

[0289] 在本发明的用于在预定位置修饰细胞系统遗传物质的方法的另一个实施方案中，该至少一个目的核苷酸序列可以是目的生物体的转基因，其中，该转基因或转基因的一部分可选自由以下各项所构成的组：编码对非生物胁迫的抗性或耐受性的基因，所述非生物胁迫包括干旱胁迫、渗透胁迫、热胁迫、冷胁迫、氧化胁迫、重金属胁迫、氮缺乏、磷缺乏、盐胁迫或渍水；编码除草剂抗性的基因，所述除草剂抗性包括草甘膦抗性、草铵膦 (glufosinate/phosphinotricin) 抗性、潮霉素抗性、2,4-D抗性或耐受性、原卟啉原氧化酶 (PPO) 抑制剂抗性或耐受性、ALS抑制剂抗性或耐受性以及麦草畏抗性或耐受性；编码生物胁迫抗性或耐受性的基因，包括病毒抗性基因、真菌抗性基因、细菌抗性基因、昆虫抗性基因；或编码产量相关性状的基因，所述产量相关性状包括抗倒伏性、开花时间、抗落粒性、种子颜色、胚乳组成或营养成分。

[0290] 在本发明的用于在预定位置修饰细胞系统的遗传物质的方法的又一个实施方案中，其中该经修饰的内源基因，与未经修饰的内源基因的核苷酸序列相比，可包含至少一个核苷酸的至少一个缺失、插入和/或取代，并且/或者该至少一个目的核苷酸序列可以是目

的生物体的经修饰的内源基因的至少一部分,其中该经修饰的内源基因可包含编码该经修饰的内源基因的结构域的至少一个核苷酸位置的截短、重复、取代和/或缺失中的至少一种。

[0291] 在本发明的用于在预定位置修饰细胞系统遗传物质的方法的又一个实施方案中,该至少一个目的核苷酸序列可以是调控序列的至少一部分,其中该调控序列可包括核心启动子序列、近端启动子序列、顺式调控序列、反式调控序列、基因座控制序列、绝缘子序列、沉默子序列、增强子序列、终止子序列或其任何组合中的至少一个。

[0292] 还提供了根据本文公开的各个方面的方法的实施方案,其中可以将至少一种位点特异性核酸酶或其催化活性片段作为编码位点特异性核酸酶及其催化活性片段的核酸序列引入细胞系统中,其中该核酸序列是至少一个载体的一部分,或者其中该至少一个位点特异性核酸酶或其催化活性片段作为至少一种氨基酸序列引入到细胞系统中。在一个实施方案中,可以将该至少一个位点特异性核酸酶作为可翻译RNA引入。在又一个实施方案中,可以将该至少一个位点特异性核酸酶与至少一个其他生物分子(例如gRNA)一起作为复合物的一部分引入,所述gRNA可选地与RT缔合,所述RT包含待引入细胞系统中的至少一个核酸序列或与待引入细胞系统中的至少一个核酸序列缔合。

[0293] 在本发明的另一方面,提供了一种选择最佳合成转录因子(STF)以调控(优选地为激活)至少一个目的基因(优选地为形态发生基因)的表达的方法,其中,该方法包括(i)定义目的基因;(ii)定义并提供至少一个识别结构域,其中该识别结构域设计为识别目的基因处或附近的识别位点;(iii)定义并提供至少一个激活结构域;(iv)可选地:提供至少一个其他元件,该元件选自至少一个启动子、至少一个NLS、至少一个反式激活结构域和/或至少一个标签;(v)提供靶向相同目的基因的至少两个STF;(vi)测量每个测试的单个STF的调控比率;(vii)为给定的目的基因选择具有最佳调控比率的STF。此外,本文所述的方法还可用于选择至少两个用于进行调控的最佳STF以微调至少两个目的形态发生基因的转录并提高转化和再生。

[0294] 根据本文提供的各个实施方案,并且由于STF的模块化本质,可以设计多个STF以调控给定的目的基因。由于复杂真核基因组中存在的空间问题和潜在的脱靶效应,因此提供具有不同数量的结构域和不同结构域构型的不同的STF可能是有利的,例如通过结构域改组或通过测试基于TALE和基于CRISPR的STF,最终为所选的靶基因选择最佳的STF。

[0295] 在本发明的另一方面,提供了一种产生单倍或双单倍体生物体或细胞系统的方法,其中,该方法可包括以下步骤:(a)提供单倍体细胞系统;(b)在该单倍体细胞系统中引入至少一个合成转录因子或编码其的核苷酸序列;(c)在获得至少一个单倍体或双单倍体生物体的条件下培养该单倍体细胞系统;以及(d)可选地:选择步骤(c)中获得的至少一个单倍体或双单倍体生物体,其中该至少一种合成转录因子或编码其的核苷酸序列包含至少一个识别结构域和至少一个激活结构域,其中该合成转录因子配置为调控该单倍体细胞系统中的至少一个形态发生基因的表达,特别地为调控该单倍体细胞系统中的至少一个形态发生基因的转录。

[0296] 由于单倍体在所有基因座都是纯合的,并且单倍体可以代表新品种(自花授粉作物)或亲本自交系,用于生产杂交品种(异花授粉作物),这使其成为植物育种计划中有吸引力的细胞类型。与野生型供体植株相比,单倍体通常较小,植株活力较低,由于减数分裂期

间染色体不能配对而不育。因此,本文中提供的合成转录因子和方法可用于单倍体细胞、细胞系统和植物的发育中,因为将至少一个本发明的合成转录因子或编码本发明的合成转录因子的核苷酸序列引入到单倍体细胞系统中,可以显著提高单倍体细胞系统的生殖能力,从而发育成单倍体胚,进而可以用作单倍体和双单倍体植物的基础。

[0297] “双单倍体”细胞、细胞系统或生物体是通过在培养单倍体细胞或细胞系统的步骤中进行自发染色体倍增或通过在选所获得的单倍体生物体后通过诱导染色体倍增而获得的。术语“双单倍体”和“加倍单倍体”在本文可互换使用。

[0298] 在一个实施方案中,在产生单倍体或双单倍体生物体的方法中,步骤(a)的单倍体细胞系统是单倍体胚,或者其中步骤(c)中获得的至少一个单倍体或双单倍体生物体是通过从(b)的单倍体细胞系统产生至少一个单倍体胚的中间步骤获得的。

[0299] 许多植物细胞都具有仅从单个细胞或组织就可再生完整生物体的能力。此过程通常称为全能性。各种各样的细胞都有可能发育成胚胎,包括单倍体配子体细胞,例如花粉和胚囊细胞(见Forster,B.P.等(2007),《植物科学发展趋势》(Trends Plant Sci.)12:368-375,以及Segui-Simarro,J.M.(2010),《植物学回顾》(Bot.Rev.76:377-404),以及源自植物所有三个组织层的体细胞(Gaj,M.D.(2004)Plant Growth Regul.43:27-47,或Rose,R.等(2010),《体细胞胚胎发生的发育生物学》:《植物发育生物学-生物技术观点》,Pua EC和Davey MR,Eds.(施普林格(Berlin Heidelberg:Springer)),第3-26页)。无融合生殖(无性种子发育的一种)在没有卵细胞受精的情况下也会发生胚胎发育。无融合生殖植物的全能性仅限于配子体和孢子体细胞,这些细胞通常有助于种子及其前体的发育,包括未受精卵细胞和周围的孢子体组织(参见Bicknell,R.A.和Koltunow,A.M.(2004)《植物细胞》(Plant Cell)16:S228-S245)。

[0300] 值得注意的是,植物细胞的全能性现象在组织培养(即,体外)达到其最高表达。因此,用于产生单倍体的相关步骤从体外不成熟细胞的培养开始,其必须在合适的条件下进行处理以诱导胚胎形成。这些步骤通常是耗时并且通常是相当低效率的,因为只有一小部分培养的单倍体细胞系统以期望的方式成熟到形态和细胞状态,任选地包括任何进一步的GE事件。在本文公开的合成转录因子和方法的辅助下,单倍体和/或双倍单倍体系统的产生可显著增强,因为该方法提供了具有更高再生能力的细胞系统,从而保证了更高频率的阳性事件。

[0301] 在产生单倍体或双单倍体细胞系统或生物体的方法的一个实施方案中,该方法可包括诱导小孢子来源的胚胎形成的附加步骤。小孢子来源的胚胎形成是一个独特的过程,其中单倍体、未成熟花粉(小孢子)通过一种或多种胁迫处理而诱导以在培养物中形成胚胎。然后,这些小孢子来源的胚胎可以通过染色体加倍剂和/或通过自发加倍而发芽并转化为纯合的双倍单倍体植物。如上所述,双单倍体生产是植物育种和性状发现计划中的主要工具,因为它可以实现以单代形式生产纯合品系。这种快速的纯合性途径不仅大大缩短了繁殖期,而且还揭示了由隐性等位基因控制的性状。双倍单倍体作为F1杂种种子生产的亲本广泛用于作物改良中,以促进用于突变育种的回交转化,并产生永生种群进行分子作图研究。

[0302] 本文在细胞系统的上下文中所使用的术语“未成熟的”旨在表示可从植物获得的任何未成熟的细胞或遗传物质。“未成熟”细胞或细胞系统可包括雄性或雌性未成熟细胞,

或未成熟的营养细胞。未成熟的雌性或雄性细胞或细胞系统可选自未成熟的胚胎或未成熟愈伤组织、雄性配子体(例如,小孢子)或营养,花粉粒的生殖或精子细胞,或雌性配子体,包括大孢子及其衍生物,包括卵细胞、极核、中央细胞、助细胞、反足细胞(antipodal)。雌性配子体材料可包含在胚珠中,并且胚珠可以代表根据本发明的细胞系统。当使用小孢子作为本发明的单倍体细胞系统时,可以形成愈伤组织,然后该愈伤组织可以进行器官发生以形成胚胎。

[0303] 用于使用化学方法获得单倍体和双单倍体细胞系统和生物体的方法是技术人员已知的(参见,例如,WO 2015/044199A1)。根据用于产生单倍体细胞系统的方法的某些实施方案,因此,该方法在将至少一个合成转录因子或编码本发明的合成转录因子的核苷酸序列引入该单倍体细胞系统之前,可包括处理或培育单倍体细胞系统的额外步骤,其中,该处理或培养的额外步骤可包括向正在发育的细胞系统中添加组蛋白去乙酰化酶抑制剂或至少一种化学物质。组蛋白去乙酰化酶抑制剂(HDACi)优选地是能够与组蛋白去乙酰化酶相互作用并抑制其酶活性从而降低组蛋白去乙酰化酶从组蛋白中除去乙酰基的能力的化合物,并且可包括例如异羟肟酸(水杨异羟肟酸除外)、环状四肽、脂族酸、苯甲酰胺、多酚或亲电性酮、曲古抑菌素A(TSA)、丁酸、丁酸盐、丁酸钾、丁酸钠、丁酸铵、丁酸锂、丁酸苯酯、钠苯丁酸酯或正丁酸钠,其中在本说明书中使用的术语丁酸不包括异丁酸或 α,β -二氯丁酸或辛二酰苯胺异羟肟酸,以上所有化合物均可商购获得。

[0304] 在另一个实施方案中,可以将物理胁迫施加到单倍体细胞系统或生物体。例如,物理胁迫可以是温度、黑暗、光或电离辐射中的任何一种。该光可以是全光谱太阳光,或者是选自可见光谱、红外光谱或紫外光谱的一个或多个频率。可以使用一种或多种物理胁迫,或胁迫组合。胁迫可以是连续的或间断的(周期性的);随时间变化规律或随机变化。当胁迫随着时间而组合时,它们可能是同时的(连续的或部分重叠的)或分离的。

[0305] 在另一个实施方案中,可以在本发明的方法中施加增加化学胁迫的附加步骤。因此,可以通过在体外培养期间,将花药或分离的配子体暴露于非生物胁迫或化学胁迫下,来另外诱导单倍体胚胎发育或小孢子胚胎形成、花粉胚胎形成或雄性生殖(Touraev, A.等(1997),《植物科学发展趋势》(Trends Plant Sci.)2:297-302)。

[0306] 在另一个实施方案中,该产生单倍体细胞系统或生物体的方法可包括从单倍体细胞系统产生至少一个双单倍体细胞系统或生物体的附加步骤。

[0307] 在又一个实施方案中,该产生单倍体或双单倍体细胞系统或生物体的方法可包括从至少一个单倍体细胞系统或生物体,或从至少一个双单倍体细胞系统或生物体产生幼苗的附加步骤。单倍体胚胎自发地或在用染色体加倍剂处理后转化为双单倍体植物的能力已被广泛利用,并且是技术人员公知的(Touraev, A.等(1997),《植物科学发展趋势》(Trends Plant Sci.)2:297-302;Forster等(2007)同上)。在某些实施方案中,单倍体胚胎形成和染色体加倍基本可以同时发生。在其他实施方案中,在单倍体胚胎形成和染色体加倍之间可能存在时间延迟。时间延迟可能与单倍体胚、幼苗或小植株生长达到的发育阶段有关。如果单倍体幼苗、植物或小植株的生长不涉及自发的染色体加倍事件,则可以根据普通技术人员熟悉的程序使用化学染色体加倍剂。根据本发明的各个方面和实施方案合适的染色体加倍和染色体加倍剂在以下文献中进行了描述:Segui-Simarro J.M.和Nuez F. (2008)《细胞遗传与基因组研究》(Cytogenet. Genome Res.) 120:358-369)。合适的染色体加倍剂包括,

例如,秋水仙碱、抗微管剂或抗微管除草剂(如戊炔草胺)、一氧化二氮或任何有丝分裂抑制剂。在使用秋水仙碱的情况下,培养基中的浓度通常可以是0.01%-0.2%,或约0.05%或APM(5-225 μ M)。秋水仙碱的浓度范围可以为约400-600mg/L,或约500mg/L。在使用戊炔草胺的情况下,培养基浓度可以为约0.5-20 μ M。诸如DMSO、佐剂或表面活性剂的其他试剂可以与有丝分裂抑制剂一起使用,以提高加倍效率。合适的染色体加倍剂的通用名称或商品名称包括:秋水仙碱、乙酰三甲基秋水仙碱酸衍生物(acetyltrimethylcolchicinic acid derivative)、草长灭、氯苯胺灵、苯胺灵、戊炔草胺(pronamide/propyzamide tebutam)、二甲基敌草索(DCPA)、麦草畏(Dicamba/dianat/disugran)(BANVEL,CLARITY)、氟草胺(benfluralin/benefin,BALAN)、仲丁灵、chloralin、敌乐胺(dinitramine)、乙丁烯氟灵(ethalfluralin,Sonalan)、氟消草(fluchloralin)、异乐灵(isopropalin)、methalpropalin、磺乐灵(nitralin)、黄草消(oryzalin,SURFLAN)、二甲戊乐灵(pendimethalin,PROWL)、氨氟乐灵(prodiamine)、环丙氟灵(profluralin)、氟乐灵(trifluralin,TREFLAN,TRIFIC,TRILLIN)、AMP(甲基胺草磷,Amiprofos methyl);甲基胺草磷抑草磷(amiprofos-methyl Butamifos)、氟硫草定(Dithiopyr),以及噻唑烟酸(Thiazopyr)。施用所述试剂的结果是纯合双单倍体细胞或细胞系统生物体。

[0308] 在上述方法的一个实施方案中,该至少一个合成转录因子或编码其的序列,或者该至少一个合成转录因子的至少一个组分或编码其的序列,通过独立地选自生物和/或物理方法的方法引入该单倍体细胞系统中,该方法包括转染、转化,其中包括通过农杆菌属物种(优选地为根癌农杆菌)、病毒载体、生物弹轰击进行转化,使用化学试剂进行转染(包括聚乙二醇转染)或其任何组合。

[0309] 在上述方法的另一个实施方案中,该至少一个识别结构域是至少一个解除武装的CRISPR/核酸酶系统,或是至少一个解除武装的CRISPR/核酸酶系统的片段。

[0310] 在上述方法的一个实施方案中,该至少一个解除武装的CRISPR/核酸酶系统是CRISPR/dCpf1系,其中该至少一个解除武装的CRISPR/核酸酶系统包含至少一个向导RNA。

[0311] 在本发明各个方面的另一个优选实施方案中,STF的识别结构域包括解除武装的LbCpf1结构域(SEQ ID NO:282)、解除武装的LbCpf1_RR结构域(SEQ ID NO:283)和/或解除武装的LbCpf1_RVR结构域(SEQ ID NO:284)。为了提高转录调控(优选地为激活)的效率,优选靶向转录起始位点上游高达250bp的区域的CRISPR/Cpf1系统gRNA。在本文所述的本发明各方面的一个实施方案中,优选的gRNA靶向该转录起始位点上游1-250、1-200、1-150、1-100、1-50、50-250、100-250、150-250或200-250bp范围内,或本文所公开范围之间的任何范围的区域。

[0312] 在一个优选实施方案中,提供单倍体或双单倍体细胞系统或生物体的方法可以利用如本文上文所述的至少一个包含至少一个识别结构域和至少一个激活结构域的合成转录因子,其中与本发明的合成转录因子相关的所述实施方案和方面可用于提供获得单倍体或双倍单倍体细胞系统或生物体的优化方法。

[0313] 在提供单倍体或双单倍体细胞系统或生物体的方法的另一个实施方案中,该至少一个合成转录因子的至少一个激活结构域选自酸性转录激活结构域所构成的组,优选地,其中该至少一个激活结构域来自水稻白叶枯病菌的无毒基因、单纯疱疹病毒的VP16或四聚体VP64、VPR、SAM、Scaffold、Suntag、P300、VP160或其任何组合。在本发明的优选实施

方案中,该至少一个激活结构域是VPR(SEQ ID NO:276)。在本发明的另一个优选实施方案中,可以使用不同激活结构域的组合,例如,VP64-p65-Rita或本领域通常已知的激活结构域的任何组合。

[0314] 用于本文所述的CRISPR/Cpf1系统的合适的接头包括柔性接头,例如5GS或XTEN,而体内可裂解的接头不适合于本文所述的本发明的各个方面。

[0315] 在提供单倍体或双单倍体细胞系统或生物体的方法的另一个实施方案中,该至少一个合成转录因子的至少一个激活结构域相对于该至少一个合成转录因子的至少一个识别结构域位于N-末端和/或C-末端。

[0316] 在本发明的提供单倍体或双单倍体细胞系统或生物体的方法的优选实施方案中,STF的识别结构域是至少一个解除武装的CRISPR/Cpf1系统,或是至少一个解除武装的CRISPR/Cpf1系统的片段,并且该激活结构域是VPR结构域,可选地在该识别结构域和该激活结构域之间具有接头,优选地为5xGS接头。

[0317] 根据本文公开方法的待修饰的优选形态发生基因可选自由BBM、WUS、WOX基因、WUS或BBM同源物、Lec1、Lec2、WIND1、ESR1、PLT3、PLT5、PLT7、IPT、IPT2、Knotted1和RKD4所构成的组。根据本文公开方法的待修饰的更优选形态发生基因可以是包含选自以下各项所构成的组的核苷酸序列的基因:(i) SEQ ID NO:199-237中任一个所示的核苷酸序列,(ii) 具有SEQ ID NO:199-237中任一个所示的核苷酸序列的编码序列的核苷酸序列,(iii) 与(i)或(ii)的核苷酸序列互补的核苷酸序列,(iv) 与(i)、(ii)或(iii)的核苷酸序列具有至少50%、60%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性(优选在全长上)的核苷酸序列,(v) 在严格条件下与(iii)的核苷酸序列杂交的核苷酸序列,(vi) 编码包含SEQ ID NO:238-258中任一个所示的氨基酸序列的蛋白的核苷酸序列,(vii) 编码包含与SEQ ID NO:238-258中任一个所示的序列至少50%、60%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同的氨基酸序列的蛋白的核苷酸序列,或(viii) 编码包含SEQ ID NO:238-258中任一个所示的氨基酸序列的蛋白的同源物、类似物或直系同源物的核苷酸序列。

[0318] 在提供单倍体或双单倍体细胞系统或生物体的方法的一个实施方案中,该至少一个合成转录因子配置为通过结合位于相对于起始密码子一定距离处的调控区域来调控形态发生基因的表达,优选地为转录。

[0319] 在提供单倍体或双单倍体细胞系统或生物体的方法的另一个实施方案中,该合成转录因子和/或该至少一个识别结构域包含SEQ ID No:276、277、282、283、284、288、289、290中任一个所示的序列,或包含与SEQ ID No:276、277、282、283、284、288、289、290中的任一个在全长上具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性的序列。

[0320] 在一个实施方案中,该至少一个单倍体细胞系统可选自由至少一个真核细胞或真核生物体所构成的组,优选地,其中该至少一个真核细胞可以是至少一个植物细胞,并且/或者其中,该至少一个真核生物体可以是植物或植物的一部分。

[0321] 在另一个实施方案中,该植物的至少一部分可选自由叶、茎、根、初生胚根(emerged radicles)、花、花部分、花瓣、果实、花粉、花粉管、花药花丝、胚珠、胚囊、卵细胞、子房、合子、胚、合子胚、体细胞胚、顶端分生组织、维管束、中柱鞘、种子、根和插枝所构成的

组。

[0322] 在另一个实施方案中,该植物细胞、该至少一个植物或植物的一部分源自可选自由以下各项所构成的组的植物物种:大麦 (*Hordeum vulgare*)、球茎大麦 (*Hordeum bulbosum*)、高粱 (*Sorghum bicolor*)、甘蔗 (*Saccharum officinarum*)、玉米 (*Zea mays*)、小米 (*Setaria italica*)、小粒野生稻 (*Oryza minuta*)、水稻 (*Oryza sativa*)、澳洲野生稻 (*Oryza australiensis*)、高秆野生稻 (*Oryza alta*)、小麦 (*Triticum aestivum*)、黑麦 (*Secale cereale*)、苹果 (*Malus domestica*)、二穗短柄草 (*Brachypodium distachyon*)、滨海大麦草 (*Hordeum marinum*)、节节麦 (*Aegilops tauschii*)、胡萝卜 (*Daucus glochidiatus*)、甜菜 (*Beta vulgaris*)、小胡萝卜 (*Daucus pusillus*)、*Daucus muricatus*、野胡萝卜 (*Daucus carota*)、巨桉 (*Eucalyptus grandis*)、美花烟草 (*Nicotiana sylvestris*)、绒毛状烟草 (*Nicotiana tomentosiformis*)、普通烟草 (*Nicotiana tabacum*)、番茄 (*Solanum lycopersicum*)、马铃薯 (*Solanum tuberosum*)、中果咖啡 (*Coffea canephora*)、葡萄 (*Vitis vinifera*)、*Erythrante guttata*、*Genlisea aurea*、黄瓜 (*Cucumis sativus*)、川桑 (*Morus notabilis*)、*Arabidopsis arenosa*、琴叶拟南芥 (*Arabidopsis lyrata*)、拟南芥 (*Arabidopsis thaliana*)、近缘种须弥芥 (*Crucihimalaya himalaica*)、卵叶须弥芥 (*Crucihimalaya wallichii*)、弯曲碎米荠 (*Cardamine flexuosa*)、北美独行菜 (*Lepidium virginicum*)、荠菜 (*Capsella bursa pastoris*)、*Olmarabidopsis pumila*、硬毛南芥 (*Arabis hirsute*)、甘蓝型油菜 (*Brassica napus*)、甘蓝 (*Brassica oleracea*)、芜菁 (*Brassica rapa*)、萝卜 (*Raphanus sativus*)、芥菜 (*Brassica juncea*)、黑芥菜 (*Brassica nigra*)、芝麻菜 (*Eruca vesicaria* subsp. *sativa*)、甜橙 (*Citrus sinensis*)、麻疯树 (*Jatropha curcas*)、毛果杨 (*Populus trichocarpa*)、蒺藜苜蓿 (*Medicago truncatula*)、*Cicer yamashitae*、*Cicer bijugum*、鹰嘴豆 (*Cicer arietinum*)、*Cicer reticulatum*、*Cicer judaicum*、木豆 (*Cajanus cajanifolius*)、蔓草虫豆 (*Cajanus scarabaeoides*)、菜豆 (*Phaseolus vulgaris*)、大豆 (*Glycine max*)、紫云英 (*Astragalus sinicus*)、百脉根 (*Lotus japonicas*)、夏堇 (*Torenia fournieri*)、洋葱 (*Allium cepa*)、大葱 (*Allium fistulosum*)、大蒜 (*Allium sativum*) 和韭菜 (*Allium tuberosum*)。

[0323] 一方面,本发明涉及一种细胞系统或其后代,其通过根据上述实施方案中任一项的用于提高细胞系统中转化效率的方法获得。

[0324] 另一方面,本发明涉及一种细胞系统或其后代,其通过根据上述实施方案中任一项的在预定位置修饰细胞系统的遗传物质的方法获得。

[0325] 在另一方面,本发明涉及一种单倍体或双单倍体生物体,其通过根据以上实施方案中任一项的生产单倍体或双单倍体生物体的方法获得。

[0326] 在本发明的一个方面,可提供至少一个细胞系统、至少一个单倍体细胞系统和/或至少一个单倍体或双单倍体细胞系统或生物体,其可通过本文所公开的方法,利用特异性调控至少一个目的形态发生基因的转录的至少一个合成转录因子获得。这样获得的细胞系统然后可以用于本文中使用的其他基因组编辑方法,或用于从修饰的细胞系统再生植物。

[0327] 在本发明的一个方面,根据本文公开的各种方法,提供了一种基于合成转录因子或编码该转录因子的序列的方法或用途。

[0328] 一方面,本发明还提供了根据上述任何实施方案的合成转录因子或编码其的序列在根据上述任何实施方案的用于提高细胞系统中转化效率的方法中的用途。

[0329] 另一方面,本发明还提供了根据上述任何实施方案的合成转录因子或编码其的序列在根据上述任何实施方案的在预定位置修饰细胞系统的遗传物质的方法中的应用。

[0330] 另一方面,本发明还提供了根据上述任何实施方案的合成转录因子或编码其的序列在根据上述任何实施方案的产生单倍体或双单倍体生物体的方法中的用途。

[0331] 通过使用本发明的合成转录因子,可以激活细胞系统中内源基因的表达。可以特异性地靶向多个内源基因,以瞬时的方式在无转基因的环境中增强表达。因此,本文所述的手段和方法具有广泛的潜在应用。

[0332] 一方面,公开了一种合成转录因子或编码其的核苷酸序列,其包含至少一个识别结构域和至少一个激活结构域,其中该合成转录因子可配置为调控细胞系统中内源基因的表达。

[0333] 在优选实施方案中,该至少一个识别结构域是至少一个解除武装的CRISPR/核酸酶系统,或是至少一个解除武装的CRISPR/核酸酶系统的片段。

[0334] 在另一个优选实施方案中,该至少一个解除武装的CRISPR/核酸酶系统是CRISPR/dCpf1系统,其中该至少一个解除武装的CRISPR/核酸酶系统包含至少一个向导RNA。

[0335] 在本发明各个方面的另一个优选实施方案中,STF的识别结构域包括解除武装的LbCpf1结构域(SEQ ID NO:282)、解除武装的LbCpf1_RR结构域(SEQ ID NO:283)和/或解除武装的LbCpf1_RVR结构域(SEQ ID NO:284)。为了提高转录调控(优选地为激活)的效率,优选靶向转录起始位点上游高达250bp的区域的CRISPR/Cpf1系统的gRNA。在本文所述的本发明各个方面的一个实施方案中,优选的gRNA靶向转录起始位点上游1-250、1-200、1-150、1-100、1-50、50-250、100-250、150-250或200-250bp范围内,或本文所公开范围之间的任何范围的区域。

[0336] 在一个实施方案中,该至少一个激活结构域选自由酸性转录激活结构域所构成的组,优选地,其中该至少一个激活结构域来自水稻白叶枯病菌的无毒基因、单纯疱疹病毒的VP16或四聚体VP64、VPR、SAM、Scaffold、Suntag、P300、VP160或其任何组合。在优选实施方案中,该至少一个激活结构域是VPR(SEQ ID NO:276)。在本发明的另一个优选实施方案中,可以使用不同激活结构域的组合,例如,VP64-p65-Rita或本领域通常已知的激活结构域的任何组合。

[0337] 用于本文所述的CRISPR/Cpf1系统的合适的接头包括柔性接头,例如5GS或XTEN,而体内可裂解的接头不适用于本文描述的本发明的各个方面。在另一个实施方案中,该至少一个激活结构域相对于该至少一个识别结构域位于N-末端和/或C-末端。

[0338] 在本发明的合成转录因子的优选实施方案中,STF的识别结构域是至少一个解除武装的CRISPR/Cpf1系统,或是至少一个解除武装的CRISPR/Cpf1系统的片段,并且该激活结构域是VPR结构域,可选地在该识别结构域和该激活结构域之间具有接头,优选地为5xGS接头。

[0339] 在另一个实施方案中,该内源基因选自由以下各项所构成的组:编码对非生物胁迫的抗性或耐受性的基因,该非生物胁迫包括干旱胁迫、渗透胁迫、热胁迫、冷胁迫、氧化胁迫、重金属胁迫、氮缺乏、磷缺乏、盐胁迫或渍水;编码除草剂抗性的基因,该除草剂抗性包

括草甘膦抗性、草铵膦/磷丝菌素抗性、潮霉素抗性、2,4-D抗性或耐受性、原卟啉原氧化酶(PPO)抑制剂抗性或耐受性、ALS抑制剂抗性或耐受性和麦草畏抗性或耐受性;编码对生物胁迫的抗性或耐受性的基因,该生物胁迫包括病毒抗性基因、真菌抗性基因、细菌抗性基因、昆虫抗性基因;或编码产量相关性状的基因,该产量相关性状包括抗倒伏性、开花时间、抗落粒性、种子颜色、胚乳成分或营养含量。具体的优选实例是ZmZEP1 (SEQ ID NO:309)、ZmRCA-beta (SEQ ID NO:310)、BvEPSPS (SEQ ID NO:311)和BvFT2 (SEQ ID NO:312)。

[0340] 本发明的其它优选实施方案包括,增加Na⁺/H⁺反转运蛋白的表达以诱导番茄植物中的盐耐受性(Zhang HX和Blumwald E(2001),转基因耐盐番茄植物在叶子中积累盐而不在果实中积累盐(Transgenic salt-tolerant tomato plants accumulate salt in foliage but not in fruit),《自然生物技术》(Nature Biotechnology)19,765-768); BvTST2.1过表达以提高直根中的蔗糖产量(Jung等(2015),鉴定甜菜直根中负责蔗糖积累的转运蛋白(Identification of the transporter responsible for sucrose accumulation in sugar beet taproots),《自然植物》(Nature Plants)1,14001); Rubisco组装伴侣Rubisco组装因子1(RAF1)对Rubisco大小亚基的过表达对提高玉米产量的作用(overexpression of small and large subunits from Rubisco with the Rubisco assembly chaperone RUBISCO ASSEMBLY FACTOR 1(RAF1)for improving corn productivity)(Salesse-Smith CE等(2018),RAF1过量表达Rubisco亚基会增加玉米中Rubisco的含量(Overexpression of Rubisco subunits with RAF1increases Rubisco content in maize),《自然植物》(Nature Plants)2,802-810);过表达ZmArgos以增加抗旱性(Shi J等(2015),ARGOS基因的过度表达修饰植物对乙烯的敏感性,从而提高拟南芥和玉米的耐旱性(Overexpression of ARGOS genes modifies plant sensitivity to ethylene,leading to improved drought tolerance in both Arabidopsis and maize),《植物生理学》(Plant Physiology)169(1),266-282);以及激活HPPD基因表达以诱导除草剂抗性(Nakka S等(2017),莧菜(Amaranthus palmeri S.Wats)对羟基苯丙酮酸双酯酶(HPPD)抑制剂抗性的生理和分子生物学特性(Physiological and molecular characterization of hydroxyphenylpyruvate dixygenase(HPPD)-inhibitor resistance in Palmer Amaranth(Amaranthus palmeri S.Wats)),《植物科学前沿》(Frontiers in Plant Science)8,555)。

[0341] 在一个实施方案中,该合成转录因子配置为通过结合位于相对于起始密码子一定距离处的调控区域来调控内源基因的表达,优选地为调控内源基因的转录。

[0342] 在另一个实施方案中,该合成转录因子和/或该至少一个识别结构域包含SEQ ID No:276、277、282、283、284、288、289、290中任一个所示的序列,或包含与SEQ ID NO:276、277、282、283、284、288、289、290中的任一个在全长上具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性的序列。

[0343] 在一个实施方案中,该细胞系统选自由至少一个真核细胞或真核生物体所构成的组,优选地,其中该至少一个真核细胞是至少一个植物细胞,并且/或者其中该至少一个真核生物体是植物或植物的一部分。

[0344] 在另一个实施方案中,该植物的至少一部分选自由叶、茎、根、初生胚根(emerged radicles)、花、花部分、花瓣、果实、花粉、花粉管、花药花丝、胚珠、胚囊、卵细胞、子房、合

子、胚、合子胚、体细胞胚、顶端分生组织、维管束、中柱鞘、种子、根和插枝所构成的组。

[0345] 在另一个实施方案中,该至少一个植物细胞、该至少一个植物或植物的至少一部分源自选自以下各项所构成的组的植物物种:大麦 (*Hordeum vulgare*)、球茎大麦 (*Hordeum bulbosum*)、高粱 (*Sorghum bicolor*)、甘蔗 (*Saccharum officinarum*)、玉米 (*Zea mays*)、小米 (*Setaria italica*)、小粒野生稻 (*Oryza minuta*)、水稻 (*Oryza sativa*)、澳洲野生稻 (*Oryza australiensis*)、高秆野生稻 (*Oryza alta*)、小麦 (*Triticum aestivum*)、黑麦 (*Secale cereale*)、苹果 (*Malus domestica*)、二穗短柄草 (*Brachypodium distachyon*)、滨海大麦草 (*Hordeum marinum*)、节节麦 (*Aegilops tauschii*)、胡萝卜 (*Daucus glochidiatus*)、甜菜 (*Beta vulgaris*)、小胡萝卜 (*Daucus pusillus*)、*Daucus muricatus*、野胡萝卜 (*Daucus carota*)、巨桉 (*Eucalyptus grandis*)、美花烟草 (*Nicotiana sylvestris*)、绒毛状烟草 (*Nicotiana tomentosiformis*)、普通烟草 (*Nicotiana tabacum*)、番茄 (*Solanum lycopersicum*)、马铃薯 (*Solanum tuberosum*)、中果咖啡 (*Coffea canephora*)、葡萄 (*Vitis vinifera*)、*Erythraea guttata*、*Genlisea aurea*、黄瓜 (*Cucumis sativus*)、川桑 (*Morus notabilis*)、*Arabidopsis arenosa*、琴叶拟南芥 (*Arabidopsis lyrata*)、拟南芥 (*Arabidopsis thaliana*)、近缘种须弥芥 (*Crucihimalaya himalaica*)、卵叶须弥芥 (*Crucihimalaya wallichii*)、弯曲碎米荠 (*Cardamine flexuosa*)、北美独行菜 (*Lepidium virginicum*)、荠菜 (*Capsella bursa pastoris*)、*Olmarabidopsis pumila*、硬毛南芥 (*Arabis hirsute*)、甘蓝型油菜 (*Brassica napus*)、甘蓝 (*Brassica oleracea*)、芜菁 (*Brassica rapa*)、萝卜 (*Raphanus sativus*)、芥菜 (*Brassica juncea*)、黑芥菜 (*Brassica nigra*)、芝麻菜 (*Eruca vesicaria* subsp. *sativa*)、甜橙 (*Citrus sinensis*)、麻疯树 (*Jatropha curcas*)、毛果杨 (*Populus trichocarpa*)、蒺藜苜蓿 (*Medicago truncatula*)、*Cicer yamashitae*、*Cicer bijugum*、鹰嘴豆 (*Cicer arietinum*)、*Cicer reticulatum*、*Cicer judaicum*、木豆 (*Cajanus cajanifolius*)、蔓草虫豆 (*Cajanus scarabaeoides*)、菜豆 (*Phaseolus vulgaris*)、大豆 (*Glycine max*)、紫云英 (*Astragalus sinicus*)、百脉根 (*Lotus japonicus*)、夏堇 (*Torenia fournieri*)、洋葱 (*Allium cepa*)、大葱 (*Allium fistulosum*)、大蒜 (*Allium sativum*) 和韭菜 (*Allium tuberosum*)。

[0346] 在另一方面,提供了一种用于提高细胞系统中至少一个内源基因表达的方法,其中该方法包括以下步骤:

[0347] (a) 提供细胞系统;

[0348] (b) 在该细胞系统中引入至少一个合成转录因子或编码其的核苷酸序列;

[0349] 其中该至少一个合成转录因子或编码其的核苷酸序列包含至少一个识别结构域和至少一个激活结构域,其中该合成转录因子配置为提高该细胞系统中的至少一个内源基因的表达,特别地为提高该细胞系统中的至少一个内源基因的转录。

[0350] 在优选实施方案中,该至少一个识别结构域是至少一个解除武装的CRISPR/核酸酶系统,或是至少一个解除武装的CRISPR/核酸酶系统的片段。

[0351] 在另一个优选实施方案中,该至少一个解除武装的CRISPR/核酸酶系统是CRISPR/dCpf1系统,其中该至少一个解除武装的CRISPR/核酸酶系统包含至少一个向导RNA。

[0352] 在本发明各个方面的另一个优选实施方案中,STF的识别结构域包括解除武装的

LbCpf1结构域(SEQ ID NO:282)、解除武装的LbCpf1_RR结构域(SEQ ID NO:283)和/或解除武装的LbCpf1_RVR结构域(SEQ ID NO:284)。为了提高转录调控(优选地为激活)的效率,优选靶向转录起始位点上游高达250bp的区域的CRISPR/Cpf1系统的gRNA。在本文所述的本发明各个方面的一个实施方案中,优选的gRNA靶向转录起始位点上游1-250、1-200、1-150、1-100、1-50、50-250、100-250、150-250或200-250bp范围内,或本文所公开范围之间的任何范围内的区域。

[0353] 在一个实施方案中,该至少一个激活结构域选自由酸性转录激活结构域所构成的组,优选地,其中该至少一个激活结构域来自水稻白叶枯病菌的无毒基因、单纯疱疹病毒的VP16或四聚体VP64、VPR、SAM、Scaffold、Suntag、P300、VP160或其任何组合。在优选实施方案中,该至少一个激活结构域是VPR(SEQ ID NO:276)。在优选实施方案中,该至少一个激活结构域是VPR(SEQ ID NO:276)。在本发明的另一个优选实施方案中,可以使用不同激活结构域的组合,例如,VP64-p65-Rita或本领域通常已知的激活结构域的任何组合。

[0354] 用于本文所述的CRISPR/Cpf1系统的合适的接头包括柔性接头,例如5GS或XTEN,而体内可裂解的接头不适合于本文所述的本发明的各个方面。在另一个实施方案中,该至少一个激活结构域相对于该至少一个识别结构域位于N-末端和/或C-末端。

[0355] 在本发明的用于提高细胞系统中至少一种内源基因表达的方法的优选实施方案中,STF的识别结构域是至少一个解除武装的CRISPR/Cpf1系统,或是至少一个解除武装的CRISPR/Cpf1系统的片段,并且该激活结构域是VPR结构域,可选地在该识别结构域和该激活结构域之间具有接头,优选地为5xGS接头。

[0356] 在一个实施方案中,该内源基因选自由以下各项所构成的组:编码对非生物胁迫的抗性或耐受性的基因,该非生物胁迫包括干旱胁迫、渗透胁迫、热胁迫、冷胁迫、氧化胁迫、重金属胁迫、氮缺乏、磷缺乏、盐胁迫或渍水;编码除草剂抗性的基因,所述除草剂抗性包括草甘膦抗性、草铵膦/磷丝菌素抗性、潮霉素抗性、2,4-D抗性或耐受性、原卟啉原氧化酶(PPO)抑制剂抗性或耐受性、ALS抑制剂抗性或耐受性和麦草畏抗性或耐受性;编码对生物胁迫的抗性或耐受性的基因,该生物胁迫包括病毒抗性基因、真菌抗性基因、细菌抗性基因、昆虫抗性基因;或编码产量相关性状的基因,所述产量相关性状包括抗倒伏性、开花时间、抗落粒性、种子颜色、胚乳成分或营养含量。。

[0357] 在一个实施方案中,该合成转录因子配置为通过结合位于相对于起始密码子一定距离处的调控区域来调控内源基因的表达,优选地为调控内源基因的转录。

[0358] 在另一个实施方案中,该合成转录因子和/或该至少一个识别结构域包含SEQ ID NO:276、277、282、283、284、288、289、290中任一个所示的序列,或包含与SEQ ID NO:276、277、282、283、284、288、289、290中的任一个在全长上具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%相同性的序列。

[0359] 在一个实施方案中,该细胞系统选自由至少一个真核细胞或真核生物体所构成的组,优选地,其中,该至少一个真核细胞是至少一个植物细胞,并且/或者其中,该至少一个真核生物体是植物或植物的一部分。

[0360] 在一个实施方案中,该植物的至少一部分选自由叶、茎、根、初生胚根(emerged radicles)、花、花部分、花瓣、果实、花粉、花粉管、花药花丝、胚珠、胚囊、卵细胞、子房、合子、胚、合子胚、体细胞胚、顶端分生组织、维管束、中柱鞘、种子、根和插枝所构成的组。

[0361] 在另一个实施方案中,该至少一个植物细胞、该至少一个植物或植物的至少一部分源自选自以下各项所构成的组的植物物种:大麦 (*Hordeum vulgare*)、球茎大麦 (*Hordeum bulbosum*)、高粱 (*Sorghum bicolor*)、甘蔗 (*Saccharum officinarum*)、玉米 (*Zea mays*)、小米 (*Setaria italica*)、小粒野生稻 (*Oryza minuta*)、水稻 (*Oryza sativa*)、澳洲野生稻 (*Oryza australiensis*)、高秆野生稻 (*Oryza alta*)、小麦 (*Triticum aestivum*)、黑麦 (*Secale cereale*)、苹果 (*Malus domestica*)、二穗短柄草 (*Brachypodium distachyon*)、滨海大麦草 (*Hordeum marinum*)、节节麦 (*Aegilops tauschii*)、胡萝卜 (*Daucus glochidiatus*)、甜菜 (*Beta vulgaris*)、小胡萝卜 (*Daucus pusillus*)、*Daucus muricatus*、野胡萝卜 (*Daucus carota*)、巨桉 (*Eucalyptus grandis*)、美花烟草 (*Nicotiana sylvestris*)、绒毛状烟草 (*Nicotiana tomentosiformis*)、普通烟草 (*Nicotiana tabacum*)、番茄 (*Solanum lycopersicum*)、马铃薯 (*Solanum tuberosum*)、中果咖啡 (*Coffea canephora*)、葡萄 (*Vitis vinifera*)、*Erythraea guttata*、*Genlisea aurea*、黄瓜 (*Cucumis sativus*)、川桑 (*Morus notabilis*)、*Arabidopsis arenosa*、琴叶拟南芥 (*Arabidopsis lyrata*)、拟南芥 (*Arabidopsis thaliana*)、近缘种须弥芥 (*Crucihimalaya himalaica*)、卵叶须弥芥 (*Crucihimalaya wallichii*)、弯曲碎米荠 (*Cardamine flexuosa*)、北美独行菜 (*Lepidium virginicum*)、荠菜 (*Capsella bursa pastoris*)、*Olm arabidopsis pumila*、硬毛南芥 (*Arabis hirsute*)、甘蓝型油菜 (*Brassica napus*)、甘蓝 (*Brassica oleracea*)、芜菁 (*Brassica rapa*)、萝卜 (*Raphanus sativus*)、芥菜 (*Brassica juncea*)、黑芥菜 (*Brassica nigra*)、芝麻菜 (*Eruca vesicaria* subsp. *sativa*)、甜橙 (*Citrus sinensis*)、麻疯树 (*Jatropha curcas*)、毛果杨 (*Populus trichocarpa*)、蒺藜苜蓿 (*Medicago truncatula*)、*Cicer yamashitae*、*Cicer bijugum*、鹰嘴豆 (*Cicer arietinum*)、*Cicer reticulatum*、*Cicer judaicum*、木豆 (*Cajanus cajanifolius*)、蔓草虫豆 (*Cajanus scarabaeoides*)、菜豆 (*Phaseolus vulgaris*)、大豆 (*Glycine max*)、紫云英 (*Astragalus sinicus*)、百脉根 (*Lotus japonicus*)、夏堇 (*Torenia fournieri*)、洋葱 (*Allium cepa*)、大葱 (*Allium fistulosum*)、大蒜 (*Allium sativum*) 和韭菜 (*Allium tuberosum*)。

[0362] 由于本文公开的合成转录因子的模块特性,还可以提供至少一个合成转录因子,其包含至少一个本文所公开的识别结构域,并且还包含沉默结构域。因此,沉默结构域取代了激活结构域,以提供高度特异性的合成转录因子,用于在这种情况下降低目的基因的转录。

[0363] 真核生物体的转录阻遏是通过“沉默子”实现的,沉默子有不同的类型,即“沉默子元件”和“负调控元件”(NRE)。沉默子元件是经典的,与位置无关的元件,可控制主动的阻遏机制,而NRE是与位置有关的元件,可引导被动的阻遏机制。另外,“阻遏物”是与沉默子直接相互作用的DNA结合转录因子。沉默子本身及其在给定启动子中的环境(而不是相互作用的阻遏物)通常决定了阻遏的机制。沉默子构成许多真核启动子的固有部分,因此对于真核生物体(包括动植物细胞)的基因调控极为重要。沉默子元件可位于相对于转录起始位点的5'或3'方向。

[0364] 因此,本发明的合成转录因子或编码其的核苷酸序列还可包含至少一个识别结构域和至少一个沉默结构域,其中该合成转录因子配置为调控目的细胞或细胞系统(优选地

为植物细胞)中形态发生基因的表达。

[0365] 一方面,提供了一种用于产生转基因细胞系统或生物体的方法,其包括进行本文详述的任何方法,其中,该方法还包括再生包含至少一个目的核苷酸序列作为转基因的细胞系统或生物体。在本文中,“转基因”是指人工引入到细胞、细胞系统或生物体中的任何核苷酸序列。

[0366] 根据某些实施方案,用于产生转基因细胞系统或生物体的方法可以优选地使用本文公开的合成转录因子来获得这种转化材料的更高转化频率和/或再生速率。

[0367] 在另一方面,提供了一种用于产生遗传修饰的细胞系统或生物体的方法,其中,该方法可包括在上文详述的预定位置处执行修饰细胞系统遗传物质的方法,其中,该方法还包括以下步骤:使在细胞系统或生物体遗传物质中的预定位置包含修饰的细胞系统或生物体再生。同样,所述方法依赖于使用根据本发明各个方面和实施方案的合成转录因子。该方面可以有利地用于将至少一个构建体或遗传物质瞬时引入目的细胞或细胞系统中,以靶向的方式修饰目的基因(优选地为形态发生基因)的转录,从而增强有可能携带插入和/或缺失和/或编辑的靶向细胞或细胞系统的再生能力。反过来,这大大减少了要针对阳性遗传修饰或编辑进行筛选的细胞数量。

[0368] 在根据本发明的各个方面的一个实施方案中,该至少一个目的核酸序列作为至少一个载体的一部分,或作为至少一种线性分子提供。在另一方面,该至少一个目的核酸序列可以复合物的形式提供,优选地为物理缔合至少一种核酸序列与另一种RT和/或与gRNA和/或与位点特异性核酸酶的复合物。该至少一个目的核酸序列还可包含允许所述目的序列(例如标签(包括荧光标签))的快速可追溯性(包括视觉可追溯性)的序列。该至少一个目的核酸序列可以是双链、单链或其混合物。此外,该至少一个目的核酸序列可包含DNA和RNA核苷酸的混合物,还包括合成的,即非天然存在的,核苷酸。

[0369] 递送和分析方法:

[0370] 可以使用将至少一个生物分子引入细胞或细胞系统的任何合适的递送方法,具体取决于目的细胞或细胞系统。因此,本文所用术语“引入”意指,将生物分子或遗传构建体(DNA、RNA、单链或双链蛋白(包括天然和/或合成组分)或其混合物)功能性转移到至少一个细胞或细胞系统中,这使得可以实现转录和/或翻译和/或催化活性,和或结合活性(包括将核酸分子与另一个核酸分子(包括DNA或RNA)结合、将蛋白与该至少一个细胞或细胞系统内的靶结构结合),和/或转录和/或翻译后引入酶的催化活性。在适当的情况下,遗传构建体的功能整合可能发生在至少一个细胞的某个细胞区室中,包括细胞核、胞质溶胶、线粒体、叶绿体、液泡、膜、细胞壁等。因此,“功能整合”一词意指,通过利用生物手段(包括农杆菌转化)或物理手段(包括粒子轰击)进行转化、转染或转导的方法以及后续步骤将目的分子复合物引入该至少一个细胞或细胞系统中,其中该分子复合物可在其引入的至少一个细胞或细胞系统上或内发挥效用,而不管该构建体或复合物是以稳定方式还是以瞬时方式引入。

[0371] 因此,根据各个实施方案,可以以至少一种载体(例如质粒载体)的形式,作为至少一个线性分子或至少一个复合物提供根据本发明的至少一个STF。

[0372] 取决于要待引入的遗传构建体或生物分子的性质,所述作用自然可以变化,并且单独或组合地包括由遗传构建体编码的DNA向核糖核酸的转录、RNA向氨基酸序列的翻译、细胞内RNA分子的活性,包括用于RNA干扰的向导RNA、crRNA、tracrRNA或miRNA或siRNA的活

性和/或结合活性(包括核酸分子与另一种核酸分子(包括DNA或RNA)的结合,或蛋白质与至少一个细胞内靶结构的结合,或包括通过载体或遗传构建体递送序列的整合(无论是瞬时还是稳定的方式)。所述作用还可以包括在至少一个细胞中代表酶或其催化活性部分的氨基酸序列的催化活性。在根据本发明的分子复合物的功能整合后实现的所述效果可取决于调控序列或定位序列的存在,该调控序列或定位序列是由所属领域技术人员已知的目的遗传结构组成的。

[0373] 根据本发明的方法的各种合适的瞬时和稳定的递送技术,用于将遗传物质、生物分子(包括任何种类的单链和双链DNA和/或RNA或氨基酸)、合成或化学物质引入真核细胞(优选植物细胞),或引入包含目的遗传物质的细胞系统中,是技术人员已知的,并且尤其包括选择直接递送技术,其范围从原生质体的聚乙二醇(PEG)处理(Potrykus,1985);电穿孔(D'Halluin等,1992)、显微注射(Neuhaus等,1987)、碳化硅纤维晶须技术(Kaeppler等,1992)、病毒载体介导的方法(Gelvin,《自然生物技术》(Nature Biotechnology) 23,“病毒介导的植物转化得到促进”(Viral-mediated plant transformation gets a boost), 684-685 (2005))和粒子轰击(参见Sood等,2011,《植物生物学》(Biologia Plantarum), 55, 1-15)等方法。用PEI瞬时转染哺乳动物细胞的方法在Longo等,《酶学方法》(Methods Enzymol.), 2013, 529:227-240中进行了公开。转化哺乳动物细胞的方案在《分子生物学方法》-核酸或蛋白质,ed. John M. Walker, Springer Protocols中进行了公开。

[0374] 对于要修饰的植物细胞,尽管发展了基于农杆菌转化或病毒载体介导的植物转化等生物转化方法,基于粒子轰击或显微注射等的物理传递方法仍是将遗传物质引入目的植物细胞或组织的重要技术。Helenius等(“使用Helios™基因枪将基因递送到完整植物中(Gene delivery into intact plants using the Helios™ Gene Gun)”,《植物分子生物学报道》,2000,18(3):287-288)公开了一种粒子轰击作为将材料引入植物细胞的物理方法。

[0375] 当前,因此,存在多种植物转化方法,以遗传构建体的形式将遗传物质引入目标植物细胞或细胞系统中,包括植物生物技术领域技术人员已知的生物和物理手段,均可应用于根据本发明的生物分子或其复合物的各种引入技术。值得注意的是,所述用于转化和转染的递送方法可以应用于同时引入本发明的工具。常见的生物学手段是用农杆菌属物种转化,几十年来,它已用于多种不同的植物材料。病毒载体介导的植物转化代表了将遗传物质引入目的细胞的另一种策略。如本文所用的,术语“粒子轰击”,也称为“生物转染”或“生物轰击”或“微粒介导的基因转移”,是指用于将包含目的核酸或遗传构建体的包被微粒或纳米颗粒转移到靶细胞或细胞中的物理递送方法。物理引入方法适合于引入核酸(即RNA和/或DNA)以及蛋白质。同样,存在用于将目的核酸或氨基酸构建体特异性引入植物细胞的特异性转化或转染方法,包括电穿孔、显微注射、纳米颗粒和细胞穿透肽(CPP)。此外,存在基于化学的转染方法以引入遗传构建体和/或核酸和/或蛋白质,尤其包括用磷酸钙转染,使用脂质体(例如阳离子脂质体)进行转染,或使用阳离子聚合物(包括DEAD-葡聚糖或聚乙烯亚胺)进行转染,或其组合。因此,所述递送方法和递送载体或产物固有地不同于用于其他真核细胞(包括动物和哺乳动物细胞)的递送工具,并且每种递送方法可能必须针对引入和/或修饰至少一个细胞系统、植物细胞、组织、器官或整个植物的遗传物质目的构建体进行专门的微调和优化;并/或可以以完全功能性和活性的方式引入到目标靶细胞的特定区

室中。

[0376] 单独或组合使用以上递送技术,可用于体内(植物体内)或体外方法。根据本发明的各个实施方案,不同的递送技术可以同时或相继地彼此组合使用,例如,使用用于至少合成转录因子或编码其的序列以及位点特异性核酸酶或编码其的mRNA或DNA以及可选的其他分子(例如gRNA)的化学转染,而这与使用基于农杆菌的技术瞬时提供(部分)失活相结合。

[0377] 因此,本发明的合成转录因子可以与用于诱导靶向的基因组编辑的相关工具的转化和/或转染、和/或与诱导单倍体或双倍单倍体发育的其他化学物质一起、在之前或之后引入。

[0378] 同样地,根据本发明的用于分析成功的转化或转染事件的方法是本领域技术人员已知的,并且包括但不限于,聚合酶链反应(PCR),尤其包括实时定量PCR、多重PCR、RT-PCR、巢式PCR、分析PCR等,显微镜(包括明视野和暗视野显微镜)、分散体染色、相衬、荧光、共焦、微分干涉衬度、解卷积、电子显微镜、UV显微镜、IR显微镜、扫描探针显微镜检查、植物或植物细胞代谢产物的分析、RNA分析、蛋白质组分析、用于确定功能整合的功能测定(例如标记基因或目的转基因,或敲除)、Southern-Blot分析、测序(包括下一代测序,包括深度测序或多重测序等)及其组合。

[0379] 在根据本发明的上述方面的又一个实施方案中,目的构建体的引入是使用物理和/或生物学方法进行的,所述物理和/或生物学方法选自由以下各项所构成的组:适用于粒子轰击的装置,包括基因枪(包括手持式基因枪(例如Helios®基因枪系统,BIO-RAD)或固定式基因枪;转化,包括使用农杆菌属物种,或使用病毒载体、显微注射、电穿孔、晶须技术(包括碳化硅晶须技术)进行的转化;以及转染;或其组合。

[0380] 除非另有说明,否则所公开方法的实施采用分子生物学、生物化学、遗传学、计算化学、细胞培养、重组DNA和相关领域中的常规技术,如本领域技术人员所知的。这些技术在文献中已得到充分解释。参见,例如,Sambrook等,分子克隆:实验室手册(MOLECULAR CLONING:A LABORATORY MANUAL),第二版,冷泉港实验室出版社(Cold Spring Harbor Laboratory Press),1989;Ausubel等,《分子生物学最新方案》(CURRENT PROTOCOLS IN MOLECULAR BIOLOGY),约翰·威利父子出版社(John Wiley&Sons),纽约,1987;定期更新;以及《酶学方法》系列(series METHODS IN ENZYMOLOGY),学术出版社(Academic Press),圣地亚哥。

[0381] 参考以下非限制性实施例进一步描述本发明。

[0382] 实施例

[0383] 实施例1:用于瞬时表达玉米(Zm)的内源性形态发生基因的TAL转录因子

[0384] 在一个实施例中,商业设计和构建的TAL转录因子用于暂时性地增强BBM和WUS的表达。TAL转录因子设计为与BBM的SEQ ID NO:95、109-147、270-272所示约24bp的调控区域和/或WUS的SEQ ID NO:96、148-190所示约18bp的调控区域结合(参见图3A和3B)。BBM的TAL转录因子识别结构域包含SEQ ID NO:13-51所示的序列,并且/或者WUS的TAL转录因子识别结构域包含SEQ ID NO:52-94所示的序列。

[0385] 可以设计和克隆TAL效应物序列,可以以融合蛋白样方式将单纯疱疹病毒(VP16或四聚体VP64)的激活结构域添加到构建体中。

[0386] 首先通过PEG介导的转化和定量逆转录酶PCR或分别针对ZmBBM和ZmWUS mRNA或蛋

白质的蛋白质免疫印迹(western blot)在玉米原生质体中测试表达的瞬时诱导。为实现这一点,将20 μ g编码TALE转录因子的质粒DNA通过本领域公知的基于PEG的转化系统输送到约600000个原生质体中(参见图4)。一式三份进行实验,并重复四次(生物学重复)。转化后24小时,使用市售试剂盒提取RNA,并将其转化为cDNA。使用SYBR Green qRT-PCR方法确定内源性ZmWUS和ZmBBM的表达。结果清楚地表明,合成转录因子TALE1 (SEQ ID NO:151) 和TALE5 (SEQ ID NO:271) 能够诱导WUS (60倍诱导) 和BBM (490倍诱导) 的内源基因表达(参见图4A和4B)。

[0387] 接下来,在可再生组织中测试由TALE转录因子诱导的瞬时ZmWUS表达的表型功能(参见图5)。因此,用荧光标记tdT、TALE1和PLT7通过基因枪法转化玉米A188愈伤组织的单细胞。在监测到tdTomato的红色荧光信号后,通过荧光显微镜确认细胞增殖的诱导(参见图5,白色圆圈和箭头)。结果清楚地表明,TALE转录因子能够通过WUS和/或BBM的瞬时表达诱导再生和胚胎形成。

[0388] 此外,定量逆转录酶PCR或使用分别针对ZmBBM和ZmWUS mRNA或蛋白质的特异性抗体的蛋白质免疫印迹表明了表达与胚胎形成表型之间的联系。表达的瞬时行为可以通过逆转录酶PCR或分别针对ZmBBM和ZmWUS mRNA或蛋白的蛋白质免疫印迹随时间进行检测。

[0389] 实施例2:非功能性CRISPR-核酸酶和激活结构域之间的融合蛋白,用于在玉米中瞬时表达内源性形态发生基因

[0390] 与实施例1相似,设计用于瞬时递送的构建体,在这种情况下,表达dCas9(可用PAM变体)或dCpf1(可用PAM变体)作为具有激活结构域(例如,VP16或VP64)的融合蛋白。潜在的靶位点/调控区域包括:ZmBBM的SEQ ID No:97-99所示的Cas9靶序列;ZmBBM的SEQ ID No:100-102所示的Cpf1靶序列;ZmWUS2的SEQ ID NO:103-105所示的Cas9靶序列;ZmWUS2的SEQ ID No:106-108所示的Cpf1靶序列。

[0391] 基于上述CRISPR/dCas9和CRISPR/dCpf1的调控区域,可以设计并商业上获得基于CRISPR的转录因子系统,其具有包含SEQ ID NO:1-12所示序列的识别结构域。

[0392] 首先通过PEG介导的转化和定量逆转录酶PCR或分别针对ZmBBM和ZmWUS mRNA或蛋白质的蛋白质免疫印迹在玉米原生质体中测试表达的瞬时诱导。然后,通过颗粒递送或农杆菌介导的转化,在诸如愈伤组织或未成熟胚的可再生组织中测试了瞬时ZmBBM和ZmWUS表达的表型功能。技术人员可以成功地诱导胚胎形成。此外,定量逆转录酶PCR或分别针对ZmBBM和ZmWUS mRNA或蛋白质的蛋白质免疫印迹表明了表达与胚胎形成表型之间的联系。

[0393] 表达的瞬时行为可以通过反转录酶PCR或分别针对ZmBBM和ZmWUS mRNA或蛋白质的蛋白质免疫印迹随时间进行检测。

[0394] 实施例3:替换激活结构域以优化形态发生基因的表达

[0395] 本实施例旨在以系统的方式测试不同的前述激活结构域的行为。这足以用来评估它们对ZmWUS和ZmBBM的表达水平的影响。如上所述,目的特异性靶基因的不同STF可包含不同的激活和识别结构域以及其他元件。因此,其非常适合为一个靶标和同一靶标设计不同的STF,从而最终定义用于调控目的基因的最佳STF。

[0396] 水稻白叶枯病菌的TAL效应物基因的天然激活结构域是与TAL转录因子一起使用的最明显的激活结构域,并且还代表一个根据本发明的各个方面可单独或组合使用但也已经用于其它设定的激活结构域。它们属于酸性(转录)激活结构域家族。

[0397] 先前已经在哺乳动物和昆虫细胞系统中测试了其他可用的激活结构域(Chavez、Alejandro等,“跨物种Cas9激活子的比较分析”(Comparative Analysis of Cas9 Activators Across Multiple Species),《自然方法》(Nature methods) 13.7 (2016):563-567,PMC.Web.2017年9月22日),但对于用于调控目的形态发生基因转录的特定用途的植物系统中使用的合成转录因子的最佳激活结构域却知之甚少。

[0398] 在此实施例中,实施例1和2中的VP16或VP64在由氨基酸链的N-末端或C-末端或两个末端的VPR、SAM、Scaffold、Suntag、P300、VP160,或这些因子中至少两个的组合或VP16和VP64替代。

[0399] 通过定量逆转录酶PCR或针对激活的基因ZmBBM和ZmWUS的蛋白质免疫印迹评估与TAL或dCas9结合的激活结构域的功效,但最终通过愈伤组织或未成熟胚的表型反应进行评估。

[0400] 实施例4:替换识别结构域以提高目标可变性和灵活性

[0401] 在该实施例中,将实施例1、2和3的TAL、dCas9或dCpf1替换为序列特异性锌指结构域或归巢核酸内切酶。作为实施例3中鉴定的具有最佳激活结构域的融合蛋白,可以组合多种导致不同基因表达强度不同的转录激活子。例如,由于dCas9或dCpf1不能在sgRNA结合中提供足够的特异性,因此仅依靠dCas9系统可能不足以特异性靶向激活结构域(至少对于某些目的基因而言)。具体来说,dCas9和dCpf1系统在靶位点特异性方面受到限制,因为它们在靶基因的调控区域中需要特异性PAM基序,而在至少某些目的基因中可能并不存在该基序(Gao、L.等(2017),“具有改变的PAM特异性的工程化Cpf1变体(Engineered Cpf1 variants with altered PAM specificities)”,《自然生物技术》(Nat Biotech);以及Kleinstiver、B.P.等(2015),“具有改变的PAM特异性的工程化CRISPR-Cas9核酸酶(Engineered CRISPR-Cas9 nucleases with altered PAM specificities)”,《自然》(Nature),523(7561):481-485)。相反,TAL转录因子通常需要初始T才能识别靶位点。因此,为了改善与难以与例如TAL STF连接的特定性目的靶基因的调控区域的结合,可以用基于dCpf1的系统替换TAL识别结构域,以便能够缩小到ATG的最佳距离,或者确定实现增强转录激活的更宽的靶向范围。此外,通过本文描述的实验获得的信息可用于设计和组合用于不同内源性调控区域的不同STF系统,以改善至少一种目的靶基因的转录激活。

[0402] 改善靶位点特异性和转录激活的另一种选择是组合使用至少两个对相同目的靶基因的相同调控区域具有特异性的识别结构域(Bolukbasi、M.F.等(2015),“DNA结合结构域融合提高Cas9的靶向范围和精确度(DNA-binding-domain fusions enhance the targeting range and precision of Cas9)”,《自然方法》(Nat Meth) 12(12):1150-1156)。

[0403] 与实施例3的激活子结合的其它识别结构域将通过定量逆转录酶PCR或针对激活的基因ZmBBM和ZmWUS的蛋白质免疫印迹再次进行评估。最终,其通过愈伤组织或未成熟胚中的表型反应来评估。

[0404] 实施例5:除ZmBBM和ZmWUS外的形态发生和胚胎形成靶基因

[0405] 已经描述了由胚胎形成的诱导导致的在愈伤组织或未成熟胚以及叶片或其它组织中瞬时过表达的多种基因。这些基因或其同源物可以单独或以组合的方式与实施例1至4中的转录激活子一起使用。该列表包括但不限于,WOX基因、其它WUS和BBM同源物、Lec1和

Lec2、WIND1、ESR1、PLT3、PLT5、PLT7、IPT和IPT2、Knotted1和RKD4。优选地,设计为调控本文所公开的形态发生基因之一的合成转录因子包含至少两个激活结构域的融合,以提供最佳的识别特性,而单独使用一个激活结构域(例如,dCas9或dCpf1)则无法实现。此外,存在至少两个适当定位以避免空间位阻并实现高激活率的激活结构域。

[0406] 实施例6:转录激活子在甜菜和小麦的形态发生和胚胎形成基因中的应用

[0407] 实施例1至5中所述的方法可以转移至具有包含体外再生或组织培养步骤的转化方案的所有相关农作物。包括实施例1至5中所述的评估方案在内的所有操作和优化步骤以及靶基因及其同源物都可以转移至其他作物系统。必须知道形态发生和胚胎形成基因的基因组序列,以便可以设计和测试dCas9、dCpf1(两者均可用的PAM变体)、TAL效应物、锌指和归巢内切核酸酶的靶标。优选地,合成转录因子包含至少两个激活结构域的融合以提供最佳识别特性,而单独使用一个激活结构域(例如,dCas9或dCpf1)则无法实现。此外,存在至少两个适当定位以避免空间位阻并实现高激活率的激活结构域。

[0408] 实施例7:定量分析增加的ZmBBM和ZmWUS转录

[0409] 可以通过简单的PCR系统或定量逆转录酶PCR来测量BBM和WUS转录的诱导。后者的优点在于对转录的绝对定量具有更高的标准化程度。优选将简单的PCR系统用于针对野生型或转化事件之间的转录的相对比较。

[0410] 使用简单PCR测定法测量BBM的转录激活。引物是SEQ ID NO:191所示的BBM-1和SEQ ID NO:192所示的BBM-2。Hot-Fire聚合酶用于34循环PCR中。

[0411] 使用qRT-PCR(Taq-Man检测)测量WUS的转录激活。EF1基因用作参考。在40循环qPCR中,ZmEF1使用SEQ ID NO:193所示的引物ZmEF1xxxr01和SEQ ID NO:194所示的引物ZmEF1xxxr01进行扩增,并用SEQ ID NO:195所示的ZmEF1xxxMGB.1进行检测。ZmWUS使用SEQ ID NO:196所示的引物WUSxxxFw1和SEQ ID NO:197所示的引物WUSxxxRv1进行扩增,并用SEQ ID NO:198所示的WUSxxxMGB进行检测。

[0412] 可以通过既定和先前公布的方法进行统计分析。

[0413] 实施例8:玉米和甜菜愈伤组织和未成熟胚中合成转录因子的递送及形态发生提高的验证

[0414] 实施例1至6中所述的合成转录因子可以以DNA、RNA或蛋白质的形式递送。已经描述了使用DNA转化玉米或甜菜愈伤组织和未成熟胚,并且这可以通过根癌农杆菌或粒子递送来完成。DNA的转化可以是瞬时的,这意味着表达盒未整合到基因组中,因此无法遗传或稳定,这意味着转化的目的是插入转基因盒。合成的或体外转录的RNA可以通过轰击来递送。通过根癌农杆菌的修饰菌株或颗粒递送已经完成了蛋白质递送。

[0415] 可以在有标记基因或没有标记基因的情况下,引入瞬时或稳定转化的基因或基因片段或任何其它合成构建体(例如,包括合适的标签)。标记基因可助于选择或筛选转化细胞或组织。其范围可以从检测转化细胞的荧光标记(例如tdTomato)到有利于阳性选择的除草剂抗性基因。

[0416] 有知识的技术人员可以通过眼睛或各种形式的显微镜检查(即,通过目视检查)来确定玉米或甜菜组织中形态发生提高的影响。通常,可以通过受影响的组织中细胞分裂的增加和胚胎形成的诱导来区分。胚胎形成可以使受影响的细胞被重新编程为胚胎早期发育阶段,即使它们之前是体细胞。

[0417] 根据检测到的效果,可能有必要对转录强度和表达谱进行修改,以达到期望的效果。这种优化可能涉及鉴定最佳的转录激活子(实施例3)、靶位点(实施例1和2)、驱动表达的启动子、递送方法(实施例8和10)、递送时间(使用可诱导系统的可能性)和其它因素。

[0418] 实施例9:将合成转录因子与基因编辑结合以提高具有编辑的再生植物的比率

[0419] 可以将实施例1至8中所述的优化的转录激活子与基因编辑试剂或T-DNA载体共递送。典型的转化方法(例如,粒子轰击和农杆菌)可能对转化的或暴露的细胞不利。根据形态发生基因的瞬时激活的最新进展,可以共递送T-DNA盒和包含上述转录因子的质粒。这使得转化的或暴露的细胞具有优势而非劣势。

[0420] 在该实施例中,可以和包含Cpf1基因和专门设计的crRNA(例如,为相关性状基因设计的)的表达盒一起通过粒子轰击来递送实施例1至8所示的质粒编码的瞬时转录激活子。该表达盒不包含供选择的抗性基因。从该愈伤组织再生的所有植物都在靶位点进行INDEL筛选。与未接受转录激活子的未选择组织相比,我们预期INDEL效率会大大降低。

[0421] 将成功编辑的植物带入下一代并再次确认用Cpf1或其他定点核酸酶进行的修饰,我们预期与对照相比,编辑过的T1植物数量更高。

[0422] 实施例10:基于蛋白质共递送合成转录激活子和定点核酸酶RNP,用于改善瞬时基因编辑

[0423] 在该实施例中,将实施例9的组分作为纯化蛋白递送到诸如愈伤组织或未成熟胚的植物组织中。实施例1至8中所述的转录因子在原核或真核细胞系统中表达,并从中纯化。Cpf1同等地生产并与合成转录或体外转录的crRNA进行孵育以形成核糖核蛋白(RNP)。已经通过粒子轰击或融合至细胞穿透肽证明了蛋白质的递送。与实施例9相比,可以预期获得的编辑过的T1植物数量更少。但是,由于完全不存在可遗传物质,因此这种方法是极为需要的。

[0424] 实施例11:将合成转录因子与碱基编辑结合以提高携带编辑的再生植物的速率

[0425] 将实施例1至8中所述的优化的转录激活子与碱基编辑试剂一起共递送到轰击的DNA盒或携带其表达盒的一个或多个T-DNA载体上。典型的转化方法(例如,粒子轰击和农杆菌)可能对转化或暴露的细胞不利。根据形态发生基因的瞬时激活的最新进展,可以共递送T-DNA盒和包含上述转录因子的质粒。这使得转化或暴露的细胞具有优势而非劣势。

[0426] 在该实施例中,可以和包含碱基编辑器基因和专门设计的向导RNA(例如,为相关性状基因设计的)的表达盒一起通过粒子轰击来递送实施例1至8所示的质粒编码的转录激活子,从而将碱基编辑器靶向适当的位点。该表达盒可包含或不包含供选择的抗性基因。碱基编辑器基因可编码胞苷脱氨酶、腺嘌呤脱氨酶或另一种脱氨酶或其他适于进行碱基转化的催化活性。碱基编辑器还可基于适合于将碱基编辑功能递送至靶位点的任何CRISPR结构域。这可包括但不限于,Cas9、Cpf1、CasX、CasY或其他合适的结构域。从该愈伤组织再生的所有植物都在靶位点进行碱基取代筛选。与未接受转录激活子的细胞相比,我们预期该再生效率会更高。

[0427] 实施例12:基于蛋白质共递送合成转录激活子和碱基编辑器RNP,用于改善瞬时基因编辑

[0428] 在该实施例中,将实施例11的组分作为纯化蛋白和RNA递送到诸如愈伤组织或未成熟胚的植物组织中。实施例1至8中所述的转录因子在原核或真核细胞系统中表达并从中

纯化。将碱基编辑器与合成转录或体外转录的crRNA进行同等的生产和孵育以形成核糖核蛋白(RNP)。通过粒子轰击或融合至细胞穿透肽已证明了蛋白质的递送。与实施例11相比,可以预期获得的编辑过的T1植物数量更少。但是,由于完全不存在可遗传物质,因此这种方法是极为需要的。

[0429] 实施例13:基于Cpf1的转录激活子的生成

[0430] 为了生成基于Cpf1的转录激活子,使用了LbCpf1表达质粒,包括识别原始TTTV PAM基序(pGEP362,SEQ ID NO:273)的野生型Lbcpf1以及两个分别识别TYCV和TATV PAM基序(pGEP487,SEQ ID NO:274;和pGEP488,SEQ ID NO:275)的LbCpf1变体(RR和RVR)。除了编码多聚核苷酸的LbCpfs之外,这些构建体还包含荧光标记mNeoGreen(参见图6A至图6C)。为了获得基于Cpf1的转录激活子,首先将VPR转录激活结构域(SEQ ID NO:276)融合到LbCpf1的C-末端。在哺乳动物细胞中显示,当用于激活GFP表达时,dAsCpf1-VP64融合仅导致最小激活,而使用VPR激活结构域导致20倍以上的转录激活(参见Liu等(2017),同上)。此外,dCas9-VP64融合构建体在植物和动物细胞也只显示了对具有单个sgRNA(在一些情况下甚至具有多个sgRNA)的靶基因的弱激活。基于这些观察结果,使用了VPR激活结构域,VPR激活域经证实可在具有dCpf1-VPR融合系统的哺乳动物细胞中诱导强转录激活(Liu等人(2017),同上;以及Tak等(2017),同上)。

[0431] Tak等(2017)使用的VPR结构域的序列(SEQ ID NO:276)进行了调整,并且在LbCpf1和VPR结构域之间使用在基于Cas9的植物转录激活系统(Lowder等(2017),同上)中使用的5xGS接头(SEQ ID NO:277)。编码5xGS接头和VPR结构域的DNA序列已针对玉米进行了密码子优化(Genscript提供)。为促进克隆过程,在EcoRI和HindIII限制性位点之间的pUC57克隆载体中,通过侧翼为位于5'端的LbCpf1编码区域的3'端和位于3'端的Nos终止子的Genscript合成密码子优化的序列。所得的质粒命名为pKWS20,并示于(SEQ ID NO:278)。

[0432] 接下来,pKWS20中带有VPR结构域的5xGS接头片段(随后是Nos终止子)通过EcoRI和HindIII双酶切反应释放,并用Gibson assembly方法克隆到MscI和XmaI双重消化的pGEP362(SEQ ID NO:273)、pGEP487(SEQ ID NO:274)或pGEP488(SEQ ID NO:275)的骨架(backbone)中,生成pGEP754(SEQ ID NO:279)、pGEP755(SEQ ID NO:280)和pGEP756(SEQ ID NO:281),携带与VPR激活结构域融合的野生型LbCpf1(SEQ ID NO:282)或LbCpf1的RR变体(LbCpf1(RR),SEQ ID NO:283)或LbCpf1的RVR变体(LbCpf1-RVR,SEQ ID NO:284)。将D832A突变进一步引入pGEP754、pGEP755和pGEP756中,以产生pGEP767(SEQ ID NO:285)、pGEP772(SEQ ID NO:286)和pGEP761(SEQ ID NO:287),其分别包含dLbCpf1-VPR(SEQ ID NO:288)或dLbCpf1(RR)-VPR(SEQ ID NO:289)或dLbCpf1(RVR)-VPR(SEQ ID NO:290)表达盒。质粒pGEP767、pGEP772和pGEP761(图6A、B、C)在以下转录激活实验中与不同的向导RNA表达质粒结合使用。

[0433] 实施例14:靶向BBM和WUS的向导RNA设计

[0434] 玉米Babyboom(BBM,SEQ ID NO:307)和Wuschel 2(WUS2,SEQ ID NO:308)基因是形态发生基因,据报道其通过异源过表达在许多之前不可转化的玉米自交系中实现高转化效率(Lowe,2016,同上)。为了测试内源性BBM和WUS2基因表达的激活是否具有相似的作用,设计了向导RNA,靶向BBM(SEQ ID NO:295-298)和WUS2(SEQ ID NO:291-294)启动子区域,以与LbCpf1-VPR融合蛋白结合使用。

[0435] 据报道,通过在哺乳动物细胞中使用dCpf1-VPR融合系统,用转录起始位点上游~600bp和下游~400bp之间的靶标检测到转录激活(Tak等(2017),同上)。基于此,扫描ZmBBM和ZmWUS2的启动子区域中从转录起始位点上游~500bp到翻译起始位点的所有可能的PAM,并且利用不同的PAM,设计了跨越整个区域的总共的针对BBM的4个向导RNA(SEQ ID NO: 295-298)和针对WUS2的4个向导RNA(SEQ ID NO:291-294)(图7和图10)。对于每个向导RNA序列,互补寡核苷酸集由IDT合成,然后退火并通过Golden Gate Assembly方法克隆到pGEP296(SEQ ID NO:299-306)中的LbCpf1 crRNA支架和丁型肝炎病毒(HDV)核酶之间(参见图8的代表性质粒图)。

[0436] 实施例15:使用LbCpf1-VPR系统对ZmBBM和ZmWUS2的转录调控

[0437] 首先通过PEG介导的转化,然后通过定量逆转录PCR,在玉米原生质体中测试内源基因表达的瞬时激活。为此,将15μg的编码LbCpf1-VPR融合蛋白的质粒DNA和8μg的表达向导RNA的质粒DNA通过本领域公知的基于PEG的转化系统共递送至约600000个玉米原生质体中。转化后24小时,使用市售试剂盒收集原生质体样品用于RNA提取和cDNA合成。然后使用SYBR Green qRT-PCR方法确定内源性ZmBBM和ZmWUS2的表达。如图9所示,所测试的靶向WUS2的启动子区域的向导RNAcrGEP186(SEQ ID NO:291)和crGEP201(SEQ ID NO:294)引起WUS2表达的显著激活(图9A)。类似地,发现靶向BBM启动子区域的向导RNA crGEP210(SEQ ID NO:297)和crGEP211(SEQ ID NO:298)会引起内源性BBM的强激活(图9B)。由于此实验仅进行了一次生物重复(三次技术重复),因此需要进一步确认,并且实验正在进行中。然而,本文首次提供的数据清楚地表明,可以使用基于Cpf1的转录激活系统来刺激植物基因激活。

序列表

<110> 科沃施种子欧洲股份两合公司

<120> 基于Cpf1的植物转录调控系统

<130> KWS0315PCT

<150> US62/609,508

<151> 2017-12-22

<150> US62/758,068

<151> 2018-11-09

<160> 318

<170> PatentIn version 3.5

<210> 1

<211> 100

<212> RNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 1

caccgcucug aucacaagca guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaaagu ggcaccgagu cggugcuuuu 100

<210> 2

<211> 100

<212> RNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 2

cccauguguu guucuaucac guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaaagu ggcaccgagu cggugcuuuu 100

<210> 3

<211> 100

<212> RNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 3

acacaugggu cagugugaag guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaaagu ggcaccgagu cggugcuuuu 100

<210> 4

<211> 100

<212> RNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 4

gucuauggca agagaggcga guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaaagu ggcaccgagu cggugcuuuu 100

<210> 5

<211> 100

<212> RNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 5

uuuauaagga gggagugcau guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaaagu ggcaccgagu cggugcuuuu 100

<210> 6

<211> 100

<212> RNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 6

uagcaugcag agagcgagag guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaaagu ggcaccgagu cggugcuuuu 100

<210> 7

<211> 45

<212> RNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 7

uaauuucuaac uaaguguaga uaccgcucug aucacaagca aggca 45

<210> 8

<211> 45

<212> RNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 8

uaauuucuac uaaguguaga uuggaaagcu auaccuccuu acccc 45

<210> 9

<211> 45

<212> RNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 9

uaauuucuac uaaguguaga uugcccucuu cacacugacc caugu 45

<210> 10

<211> 45

<212> RNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 10

uaauuucuac uaaguguaga ugcaagagag gcgaaggagg guucc 45

<210> 11

<211> 45

<212> RNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 11

uaauuucuac uaaguguaga uuaaggaggg agugcauugg accua 45

<210> 12

<211> 45

<212> RNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 12

uaauuucuac uaaguguaga ugcucucgcu cucugcaugc uagcu 45

<210> 13

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 13

```
His Asp His Asp Asn Gly Asn His His Asp His Asp His Asp Asn Gly
1          5          10          15
His Asp Asn Gly Asn Gly His Asp Asn Ile His Asp Asn Ile His Asp
          20          25          30
Asn Gly Asn His
          35
```

<210> 14

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 14

```
His Asp Asn Gly Asn Gly Asn Gly Asn Ile Asn Gly His Asp His Asp
1          5          10          15
Asn Gly Asn Gly Asn Ile Asn Ile Asn Ile Asn Gly Asn Ile Asn Ile
          20          25          30
Asn His Asn Ile
          35
```

<210> 15

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 15

```
Asn His His Asp His Asp His Asp Asn Gly His Asp Asn Gly Asn Gly
1          5          10          15
His Asp Asn Ile His Asp Asn Ile His Asp Asn Gly Asn His Asn Ile
          20          25          30
His Asp His Asp
          35
```

<210> 16

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 16

```
Asn Ile Asn Gly His Asp His Asp Asn Gly Asn Gly Asn Ile Asn Ile
1           5           10           15
Asn Ile Asn Gly Asn Ile Asn Ile Asn His Asn Ile Asn Ile Asn His
           20           25           30
His Asp Asn Ile
           35
```

<210> 17

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 17

```
His Asp His Asp Asn Gly Asn Gly Asn Ile Asn Ile Asn Ile Asn Gly
1           5           10           15
Asn Ile Asn Ile Asn His Asn Ile Asn Ile Asn His His Asp Asn Ile
           20           25           30
Asn Gly Asn Ile
           35
```

<210> 18

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 18

```
His Asp Asn Gly Asn Gly His Asp Asn Ile His Asp Asn Ile His Asp
1           5           10           15
Asn Gly Asn His Asn Ile His Asp His Asp His Asp Asn Ile Asn Gly
           20           25           30
Asn His Asn Gly
           35
```

<210> 19

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 19

Asn His Asn Gly Asn Gly His Asp Asn Gly Asn Ile Asn Gly His Asp
1 5 10 15
Asn Ile Asn Ile His Asp Asn His His Asp His Asp His Asp His Asp
 20 25 30
Asn Gly His Asp
 35

<210> 20

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 20

His Asp Asn Gly Asn Ile Asn Gly His Asp Asn Ile Asn Ile His Asp
1 5 10 15
Asn His His Asp His Asp His Asp His Asp Asn Gly His Asp His Asp
 20 25 30
His Asp Asn Gly
 35

<210> 21

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 21

Asn Ile Asn Gly His Asp Asn Ile Asn Ile His Asp Asn His His Asp
1 5 10 15
His Asp His Asp His Asp Asn Gly His Asp His Asp His Asp Asn Gly
 20 25 30
Asn Gly Asn Ile
 35

<210> 22

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 22

Asn	His	Asn	Gly	Asn	Gly	Asn	His	Asn	Gly	Asn	Gly	His	Asp	Asn	Gly
1				5				10						15	
Asn	Ile	Asn	Gly	His	Asp	His	Asp	His	Asp	Asn	Gly	Asn	His	Asn	His
				20				25						30	
Asn	Ile	Asn	Ile												
				35											

<210> 23

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 23

Asn	His	Asn	Gly	Asn	Gly	His	Asp	Asn	Gly	Asn	Ile	Asn	Gly	His	Asp
1				5				10						15	
His	Asp	His	Asp	Asn	Gly	Asn	His	Asn	His	Asn	Ile	Asn	Ile	Asn	Ile
				20				25						30	
Asn	His	His	Asp												
				35											

<210> 24

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 24

His	Asp	Asn	Gly	Asn	Ile	Asn	Gly	His	Asp	His	Asp	His	Asp	Asn	Gly
1				5				10						15	
Asn	His	Asn	His	Asn	Ile	Asn	Ile	Asn	Ile	Asn	His	His	Asp	Asn	Gly
				20				25						30	
Asn	Ile	Asn	Gly												
				35											

<210> 25

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 25

Asn Ile Asn Gly His Asp His Asp His Asp Asn Gly Asn His Asn His
1 5 10 15
Asn Ile Asn Ile Asn Ile Asn His His Asp Asn Gly Asn Ile Asn Gly
 20 25 30
Asn Ile His Asp
 35

<210> 26

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 26

His Asp Asn Gly His Asp Asn Ile Asn His His Asp His Asp Asn Ile
1 5 10 15
Asn His Asn Gly Asn Gly His Asp Asn Gly Asn Gly Asn Ile Asn Ile
 20 25 30
His Asp Asn Gly
 35

<210> 27

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 27

Asn His Asn His Asn Ile Asn Ile Asn Ile Asn His His Asp Asn Gly
1 5 10 15
Asn Ile Asn Gly Asn Ile His Asp His Asp Asn Gly His Asp His Asp
 20 25 30
Asn Gly Asn Gly
 35

<210> 28

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 28

```
Asn Ile Asn Gly Asn Ile His Asp His Asp Asn Gly His Asp His Asp
1           5           10           15
Asn Gly Asn Gly Asn Ile His Asp His Asp His Asp His Asp Asn Gly
           20           25           30
Asn Ile Asn Gly
           35
```

<210> 29

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 29

```
Asn Ile His Asp His Asp Asn Gly His Asp His Asp Asn Gly Asn Gly
1           5           10           15
Asn Ile His Asp His Asp His Asp His Asp Asn Gly Asn Ile Asn Gly
           20           25           30
His Asp Asn Ile
           35
```

<210> 30

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 30

```
His Asp His Asp Asn Gly Asn Gly Asn Ile His Asp His Asp His Asp
1           5           10           15
His Asp Asn Gly Asn Ile Asn Gly His Asp Asn Ile Asn His His Asp
           20           25           30
Asn Gly Asn Gly
           35
```

<210> 31

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 31

His Asp Asn Gly His Asp Asn Gly Asn Gly Asn Ile Asn Gly Asn Ile
1 5 10 15
Asn Ile Asn Ile Asn Gly Asn Ile His Asp Asn Ile Asn His Asn Ile
 20 25 30
His Asp His Asp
 35

<210> 32

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 32

His Asp Asn Gly Asn Gly Asn Ile Asn Gly Asn Ile Asn Ile Asn Ile
1 5 10 15
Asn Gly Asn Ile His Asp Asn Ile Asn His Asn Ile His Asp His Asp
 20 25 30
Asn Gly Asn Gly
 35

<210> 33

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 33

Asn Ile His Asp His Asp His Asp His Asp Asn Gly Asn Ile Asn Gly
1 5 10 15
His Asp Asn Ile Asn His His Asp Asn Gly Asn Gly His Asp Asn Gly
 20 25 30
His Asp His Asp
 35

<210> 34

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 34

```
Asn Ile Asn Gly Asn Ile Asn Ile Asn Ile Asn Gly Asn Ile His Asp
1           5           10           15
Asn Ile Asn His Asn Ile His Asp His Asp Asn Gly Asn Gly Asn His
           20           25           30
Asn Gly Asn Ile
           35
```

<210> 35

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 35

```
Asn Ile Asn Gly His Asp Asn Ile Asn His His Asp Asn Gly Asn Gly
1           5           10           15
His Asp Asn Gly His Asp His Asp Asn Gly His Asp Asn Ile His Asp
           20           25           30
Asn Ile Asn Gly
           35
```

<210> 36

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 36

```
Asn Ile His Asp Asn Ile Asn His Asn Ile His Asp His Asp Asn Gly
1           5           10           15
Asn Gly Asn His Asn Gly Asn Ile His Asp Asn Ile Asn Ile His Asp
           20           25           30
Asn Ile His Asp
           35
```

<210> 37

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 37

His Asp Asn Gly His Asp His Asp Asn Gly His Asp Asn Ile His Asp
1 5 10 15
Asn Ile Asn Gly His Asp Asn Gly His Asp His Asp Asn Gly His Asp
 20 25 30
Asn Gly His Asp
 35

<210> 38

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 38

His Asp His Asp Asn Gly His Asp Asn Ile His Asp Asn Ile Asn Gly
1 5 10 15
His Asp Asn Gly His Asp His Asp Asn Gly His Asp Asn Gly His Asp
 20 25 30
Asn His Asn Gly
 35

<210> 39

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 39

Asn His Asn Gly Asn Ile His Asp Asn Ile Asn Ile His Asp Asn Ile
1 5 10 15
His Asp Asn Gly Asn Gly Asn Gly His Asp Asn Ile His Asp His Asp
 20 25 30
Asn Gly His Asp
 35

<210> 40

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 40

Asn Ile His Asp Asn Ile Asn Ile His Asp Asn Ile His Asp Asn Gly
1 5 10 15
Asn Gly Asn Gly His Asp Asn Ile His Asp His Asp Asn Gly His Asp
 20 25 30
His Asp Asn Gly
 35

<210> 41

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 41

His Asp Asn Gly His Asp His Asp Asn Gly His Asp Asn Gly His Asp
1 5 10 15
Asn His Asn Gly His Asp Asn His His Asp His Asp Asn Ile His Asp
 20 25 30
His Asp His Asp
 35

<210> 42

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 42

His Asp His Asp Asn Gly His Asp Asn Gly His Asp Asn His Asn Gly
1 5 10 15
His Asp Asn His His Asp His Asp Asn Ile His Asp His Asp His Asp
 20 25 30
Asn Ile Asn Gly
 35

<210> 43

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 43

His Asp Asn Gly His Asp Asn His Asn Gly His Asp Asn His His Asp
1 5 10 15
His Asp Asn Ile His Asp His Asp His Asp Asn Ile Asn Gly Asn His
 20 25 30
His Asp Asn Gly
 35

<210> 44

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 44

His Asp Asn His Asn Gly His Asp Asn His His Asp His Asp Asn Ile
1 5 10 15
His Asp His Asp His Asp Asn Ile Asn Gly Asn His His Asp Asn Gly
 20 25 30
Asn Ile Asn Gly
 35

<210> 45

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 45

His Asp Asn His His Asp His Asp Asn Ile His Asp His Asp His Asp
1 5 10 15
Asn Ile Asn Gly Asn His His Asp Asn Gly Asn Ile Asn Gly His Asp
 20 25 30
Asn Ile His Asp
 35

<210> 46

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 46

His Asp His Asp His Asp Asn Gly Asn His Asn His Asn Ile Asn Ile
1 5 10 15
Asn Ile Asn His His Asp Asn Gly Asn Ile Asn Gly Asn Ile His Asp
 20 25 30
His Asp Asn Gly
 35

<210> 47

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 47

Asn His His Asp Asn Gly Asn Ile Asn Gly His Asp Asn Ile His Asp
1 5 10 15
His Asp Asn His His Asp Asn Gly His Asp Asn Gly Asn His Asn Ile
 20 25 30
Asn Gly His Asp
 35

<210> 48

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 48

Asn Ile Asn Gly His Asp Asn Ile His Asp His Asp Asn His His Asp
1 5 10 15
Asn Gly His Asp Asn Gly Asn His Asn Ile Asn Gly His Asp Asn Ile
 20 25 30
His Asp Asn Ile
 35

<210> 49

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 49

His Asp Asn Gly Asn His Asn Ile Asn Gly His Asp Asn Ile His Asp
1 5 10 15
Asn Ile Asn Ile Asn His His Asp Asn Ile Asn Ile Asn His Asn His
 20 25 30
His Asp Asn Ile
 35

<210> 50

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 50

His Asp Asn Gly Asn His Asn His His Asp His Asp His Asp His Asp
1 5 10 15
Asn Gly Asn Gly His Asp His Asp Asn Gly Asn His His Asp His Asp
 20 25 30
His Asp Asn Gly
 35

<210> 51

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 51

Asn His Asn His His Asp His Asp His Asp His Asp Asn Gly Asn Gly
1 5 10 15
His Asp His Asp Asn Gly Asn His His Asp His Asp His Asp Asn Gly
 20 25 30
His Asp Asn Gly
 35

<210> 52

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 52

His Asp Asn Gly Asn Gly His Asp Asn Gly His Asp His Asp His Asp
1 5 10 15
Asn His His Asp Asn Gly His Asp Asn Gly His Asp Asn His His Asp
 20 25 30
Asn Gly His Asp
 35

<210> 53

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 53

His Asp Asn Gly His Asp His Asp His Asp Asn His His Asp Asn Gly
1 5 10 15
His Asp Asn Gly His Asp Asn His His Asp Asn Gly His Asp Asn Gly
 20 25 30
His Asp Asn Gly
 35

<210> 54

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 54

Asn His His Asp Asn Ile Asn Gly Asn His His Asp Asn Gly Asn Ile
1 5 10 15
Asn His His Asp Asn Gly Asn Ile His Asp His Asp Asn Gly Asn Gly
 20 25 30
His Asp Asn Gly
 35

<210> 55

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 55

His	Asp	His	Asp	His	Asp	Asn	His	His	Asp	Asn	Gly	His	Asp	Asn	Gly
1				5					10					15	
His	Asp	Asn	His	His	Asp	Asn	Gly	His	Asp	Asn	Gly	His	Asp	Asn	Gly
				20				25						30	
Asn	His	His	Asp												
				35											

<210> 56

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 56

His	Asp	Asn	Gly	His	Asp	Asn	His	His	Asp	Asn	Gly	His	Asp	Asn	Gly
1				5					10					15	
His	Asp	Asn	Gly	Asn	His	His	Asp	Asn	Ile	Asn	Gly	Asn	His	His	Asp
				20				25						30	
Asn	Gly	Asn	Ile												
				35											

<210> 57

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 57

His	Asp	Asn	His	His	Asp	Asn	Gly	His	Asp	Asn	Gly	His	Asp	Asn	Gly
1				5					10					15	
Asn	His	His	Asp	Asn	Ile	Asn	Gly	Asn	His	His	Asp	Asn	Gly	Asn	Ile
				20				25						30	
Asn	His	His	Asp												
				35											

<210> 58

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 58

His	Asp	Asn	Gly	His	Asp	Asn	Gly	Asn	His	His	Asp	Asn	Ile	Asn	Gly
1				5				10					15		
Asn	His	His	Asp	Asn	Gly	Asn	Ile	Asn	His	His	Asp	Asn	Gly	Asn	Ile
			20				25						30		
His	Asp	His	Asp												
			35												

<210> 59

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 59

His	Asp	Asn	Gly	Asn	His	His	Asp	Asn	Ile	Asn	Gly	Asn	His	His	Asp
1				5				10					15		
Asn	Gly	Asn	Ile	Asn	His	His	Asp	Asn	Gly	Asn	Ile	His	Asp	His	Asp
			20				25					30			
Asn	Gly	Asn	Gly												
			35												

<210> 60

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 60

Asn	Ile	Asn	Gly	Asn	His	Asn	Gly	His	Asp	Asn	Ile	Asn	Ile	His	Asp
1				5				10					15		
Asn	Gly	Asn	Gly	His	Asp	Asn	Ile	His	Asp	Asn	Gly	Asn	Gly	Asn	His
			20				25					30			
Asn	Gly	His	Asp												
			35												

<210> 61

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 61

Asn	His	Asn	Gly	His	Asp	Asn	Ile	Asn	Ile	His	Asp	Asn	Gly	Asn	Gly
1				5				10						15	
His	Asp	Asn	Ile	His	Asp	Asn	Gly	Asn	Gly	Asn	His	Asn	Gly	His	Asp
			20				25						30		
Asn	Gly	His	Asp												
			35												

<210> 62

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 62

Asn	His	His	Asp	Asn	Gly	Asn	Ile	Asn	His	His	Asp	Asn	Gly	Asn	Ile
1				5				10						15	
His	Asp	His	Asp	Asn	Gly	Asn	Gly	His	Asp	Asn	Gly	Asn	Ile	Asn	His
			20				25						30		
His	Asp	Asn	Gly												
			35												

<210> 63

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 63

Asn	Ile	Asn	His	His	Asp	Asn	Gly	Asn	Ile	His	Asp	His	Asp	Asn	Gly
1				5				10						15	
Asn	Gly	His	Asp	Asn	Gly	Asn	Ile	Asn	His	His	Asp	Asn	Gly	Asn	Ile
			20				25						30		
Asn	Gly	His	Asp												
			35												

<210> 64

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 64

Asn	Ile	His	Asp	His	Asp	Asn	Gly	Asn	Gly	His	Asp	Asn	Gly	Asn	Ile
1				5				10						15	
Asn	His	His	Asp	Asn	Gly	Asn	Ile	Asn	Gly	His	Asp	Asn	Gly	Asn	Ile
			20				25						30		
Asn	His	His	Asp												
			35												

<210> 65

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 65

Asn	His	Asn	Gly	His	Asp	Asn	Gly	His	Asp	Asn	Gly	His	Asp	Asn	Gly
1				5				10						15	
His	Asp	His	Asp	Asn	Ile	Asn	Ile	Asn	Ile	Asn	Ile	Asn	His	Asn	Ile
			20				25						30		
Asn	Gly	Asn	Ile												
			35												

<210> 66

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 66

His	Asp	Asn	Gly	Asn	Ile	Asn	His	His	Asp	Asn	Gly	Asn	Ile	Asn	Gly
1				5				10						15	
His	Asp	Asn	Gly	Asn	Ile	Asn	His	His	Asp	His	Asp	Asn	Gly	His	Asp
			20				25						30		
Asn	Gly	Asn	Ile												
			35												

<210> 67

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 67

His	Asp	Asn	Gly	His	Asp	Asn	Gly	His	Asp	Asn	Gly	His	Asp	His	Asp
1				5				10						15	
Asn	Ile	Asn	Ile	Asn	Ile	Asn	Ile	Asn	His	Asn	Ile	Asn	Gly	Asn	Ile
				20				25						30	
Asn	Gly	His	Asp												
				35											

<210> 68

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 68

Asn	Ile	Asn	His	His	Asp	Asn	Gly	Asn	Ile	Asn	Gly	His	Asp	Asn	Gly
1				5				10						15	
Asn	Ile	Asn	His	His	Asp	His	Asp	Asn	Gly	His	Asp	Asn	Gly	Asn	Ile
				20				25						30	
Asn	His	Asn	His												
				35											

<210> 69

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 69

His	Asp	Asn	Gly	His	Asp	Asn	Gly	His	Asp	His	Asp	Asn	Ile	Asn	Ile
1				5				10						15	
Asn	Ile	Asn	Ile	Asn	His	Asn	Ile	Asn	Gly	Asn	Ile	Asn	Gly	His	Asp
				20				25						30	
Asn	His	Asn	Gly												
				35											

<210> 70

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 70

His	Asp	Asn	Gly	His	Asp	His	Asp	Asn	Ile	Asn	Ile	Asn	Ile	Asn	Ile
1				5					10					15	
Asn	His	Asn	Ile	Asn	Gly	Asn	Ile	Asn	Gly	His	Asp	Asn	His	Asn	Gly
				20					25					30	
Asn	Ile	Asn	Gly												
				35											

<210> 71

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 71

Asn	Ile	Asn	Gly	His	Asp	Asn	Gly	Asn	Ile	Asn	His	His	Asp	His	Asp
1				5					10					15	
Asn	Gly	His	Asp	Asn	Gly	Asn	Ile	Asn	His	Asn	His	Asn	Gly	His	Asp
				20					25					30	
His	Asp	Asn	Ile												
				35											

<210> 72

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 72

His	Asp	His	Asp	Asn	Ile	Asn	Ile	Asn	Ile	Asn	Ile	Asn	His	Asn	Ile
1				5					10					15	
Asn	Gly	Asn	Ile	Asn	Gly	His	Asp	Asn	His	Asn	Gly	Asn	Ile	Asn	Gly
				20					25					30	
His	Asp	Asn	Ile												
				35											

<210> 73

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 73

```
His Asp Asn Gly Asn Ile Asn His His Asp His Asp Asn Gly His Asp
1           5           10           15
Asn Gly Asn Ile Asn His Asn His Asn Gly His Asp His Asp Asn Ile
           20           25           30
Asn Ile Asn Gly
           35
```

<210> 74

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 74

```
Asn Ile Asn His His Asp His Asp Asn Gly His Asp Asn Gly Asn Ile
1           5           10           15
Asn His Asn His Asn Gly His Asp His Asp Asn Ile Asn Ile Asn Gly
           20           25           30
Asn His His Asp
           35
```

<210> 75

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 75

```
His Asp Asn Gly Asn Ile Asn His Asn His Asn Gly His Asp His Asp
1           5           10           15
Asn Ile Asn Ile Asn Gly Asn His His Asp Asn Ile His Asp Asn Gly
           20           25           30
His Asp His Asp
           35
```

<210> 76

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 76

```
Asn Ile Asn Gly His Asp Asn His Asn Gly Asn Ile Asn Gly His Asp
1           5           10           15
Asn Ile His Asp His Asp His Asp Asn Ile Asn Gly Asn His Asn His
           20           25           30
Asn His His Asp
           35
```

<210> 77

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 77

```
Asn Ile Asn His Asn His Asn Gly His Asp His Asp Asn Ile Asn Ile
1           5           10           15
Asn Gly Asn His His Asp Asn Ile His Asp Asn Gly His Asp His Asp
           20           25           30
His Asp Asn Gly
           35
```

<210> 78

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 78

```
His Asp Asn His Asn Gly Asn Ile Asn Gly His Asp Asn Ile His Asp
1           5           10           15
His Asp His Asp Asn Ile Asn Gly Asn His Asn His Asn His His Asp
           20           25           30
Asn Ile Asn Ile
           35
```

<210> 79

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 79

His	Asp	His	Asp	Asn	Ile	Asn	Ile	Asn	Gly	Asn	His	His	Asp	Asn	Ile
1				5					10					15	
His	Asp	Asn	Gly	His	Asp	His	Asp	His	Asp	Asn	Gly	His	Asp	His	Asp
			20					25					30		
Asn	Gly	Asn	Gly												
			35												

<210> 80

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 80

His	Asp	Asn	Gly	Asn	Gly	Asn	His	His	Asp	His	Asp	Asn	Ile	Asn	Gly
1				5					10					15	
Asn	Ile	Asn	His	Asn	Ile	His	Asp	His	Asp	Asn	His	Asn	His	Asn	Ile
			20					25					30		
His	Asp	Asn	Ile												
			35												

<210> 81

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 81

Asn	His	His	Asp	Asn	Ile	His	Asp	Asn	Gly	His	Asp	His	Asp	His	Asp
1				5					10					15	
Asn	Gly	His	Asp	His	Asp	Asn	Gly	Asn	Gly	Asn	Ile	Asn	Gly	Asn	Ile
			20					25					30		
Asn	Ile	Asn	Ile												
			35												

<210> 82

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 82

```
His Asp His Asp His Asp Asn Gly His Asp His Asp Asn Gly Asn Gly
1          5          10          15
Asn Ile Asn Gly Asn Ile Asn Ile Asn Ile His Asp Asn Ile Asn Ile
          20          25          30
Asn His Asn His
          35
```

<210> 83

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 83

```
His Asp His Asp Asn Gly Asn Gly Asn Ile Asn Gly Asn Ile Asn Ile
1          5          10          15
Asn Ile His Asp Asn Ile Asn Ile Asn His Asn His Asn Ile Asn Ile
          20          25          30
His Asp His Asp
          35
```

<210> 84

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 84

```
Asn His Asn His His Asp His Asp Asn Ile Asn Gly Asn His Asn Ile
1          5          10          15
His Asp His Asp His Asp His Asp His Asp His Asp Asn Gly His Asp
          20          25          30
His Asp His Asp
          35
```

<210> 85

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 85

Asn Ile Asn Gly Asn Ile Asn Ile Asn Ile His Asp Asn Ile Asn Ile
1 5 10 15
Asn His Asn His Asn Ile Asn Ile His Asp His Asp His Asp Asn Gly
 20 25 30
His Asp His Asp
 35

<210> 86

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 86

His Asp His Asp His Asp Asn Ile Asn His His Asp His Asp His Asp
1 5 10 15
His Asp Asn Ile Asn Ile His Asp His Asp Asn Gly Asn Ile Asn Gly
 20 25 30
Asn Ile Asn Gly
 35

<210> 87

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 87

His Asp His Asp Asn Gly Asn Gly His Asp Asn His His Asp His Asp
1 5 10 15
Asn Gly His Asp Asn Gly His Asp Asn Gly Asn Gly Asn His His Asp
 20 25 30
His Asp Asn Ile
 35

<210> 88

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 88

His Asp Asn His His Asp His Asp Asn Gly His Asp Asn Gly His Asp
1 5 10 15
Asn Gly Asn Gly Asn His His Asp His Asp Asn Ile Asn Gly Asn Ile
 20 25 30
Asn His Asn Ile
 35

<210> 89

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 89

His Asp Asn Gly His Asp Asn Gly Asn Gly Asn His His Asp His Asp
1 5 10 15
Asn Ile Asn Gly Asn Ile Asn His Asn Ile His Asp His Asp Asn His
 20 25 30
Asn His Asn Ile
 35

<210> 90

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 90

Asn Ile Asn Gly Asn Ile Asn Gly His Asp Asn Ile His Asp His Asp
1 5 10 15
Asn Gly Asn Ile Asn His His Asp Asn His His Asp Asn Ile Asn His
 20 25 30
His Asp Asn Gly
 35

<210> 91

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 91

```
Asn Ile Asn Gly His Asp Asn Ile His Asp His Asp Asn Gly Asn Ile
1           5           10           15
Asn His His Asp Asn His His Asp Asn Ile Asn His His Asp Asn Gly
           20           25           30
Asn Ile His Asp
           35
```

<210> 92

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 92

```
Asn Ile Asn His His Asp Asn His His Asp Asn Ile Asn His His Asp
1           5           10           15
Asn Gly Asn Ile His Asp Asn His His Asp Asn Gly His Asp Asn Gly
           20           25           30
His Asp Asn Gly
           35
```

<210> 93

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 93

```
Asn Ile His Asp Asn His His Asp Asn Gly His Asp Asn Gly His Asp
1           5           10           15
Asn Gly Asn Gly His Asp Asn Gly His Asp His Asp His Asp Asn His
           20           25           30
His Asp Asn Gly
           35
```

<210> 94

<211> 36

<212> PRT

<213> 人工序列

<220>

<223> 识别结构域

<400> 94

```

His Asp Asn Gly His Asp Asn Gly Asn Gly His Asp Asn Gly His Asp
1           5           10           15
His Asp His Asp Asn His His Asp Asn Gly His Asp Asn Gly His Asp
           20           25           30
Asn His His Asp
           35

```

<210> 95

<211> 303

<212> DNA

<213> Zea mays

<400> 95

```

cctctttatc cttaaataag aagcataaaa cgggatttct cagccagttc ttaacttctc 60
ttataaatac agaccttgta caacactttc acctcctctc aggtggccag gatatttttt 120
ctggccccctt cctgccctct tcacactgac ccatgtgttg ttctatccct ggaaagctat 180
acctccttac ccctatcagc ttctcctcac atctcctctc gtcgccaccc atgctatcac 240
cgctctgatc acaagcaagg caaacctca ctgttctatc aagccccctc ccttagctag 300
atg 303

```

<210> 96

<211> 303

<212> DNA

<213> Zea mays

<400> 96

```

gcaagagcca gcccccgcc gtatgtcaac ttcacttgct tctctccaaa agatatcgta 60
tcacccatgg gcaatggcca tgacccccct ccagcccca acctatatca cctagcgcag 120
ctacgctctc ttctcccgtc ctgctctctc gcattgtagc taccttctag ctatctagcc 180
tctaggtcca atgcactccc tccttataaa caaggaacce tccttcgcct ctcttgccat 240
agaccggaca ccggagagct aggtcacagg agcgtcagg aaggccgctg agatagaggc 300
atg 303

```

<210> 97

<211> 23

<212> DNA

<213> Zea mays

<400> 97

```

caccgctctg atcacaagca agg 23

```

<210> 98

<211> 23

<212> DNA
<213> Zea mays
<400> 98
cccatgtgtt gttctatccc tgg 23
<210> 99
<211> 23
<212> DNA
<213> Zea mays
<400> 99
acacatgggt cagtgtgaag agg 23
<210> 100
<211> 27
<212> DNA
<213> Zea mays
<400> 100
atcacccgctc tgatcacaag caaggca 27
<210> 101
<211> 28
<212> DNA
<213> Zea mays
<400> 101
tccctggaaa gctatacctc cttacccc 28
<210> 102
<211> 28
<212> DNA
<213> Zea mays
<400> 102
ttcctgccct cttcacactg acccatgt 28
<210> 103
<211> 23
<212> DNA
<213> Zea mays
<400> 103
gtctatggca agagaggcga agg 23
<210> 104
<211> 23
<212> DNA
<213> Zea mays
<400> 104

tttataagga gggagtgcac tgg 23
<210> 105
<211> 23
<212> DNA
<213> Zea mays
<400> 105
tagcatgcag agagcgagag cgg 23
<210> 106
<211> 28
<212> DNA
<213> Zea mays
<400> 106
tatggcaaga gaggcgaagg agggttcc 28
<210> 107
<211> 28
<212> DNA
<213> Zea mays
<400> 107
tttataagga gggagtgcac tggaccta 28
<210> 108
<211> 28
<212> DNA
<213> Zea mays
<400> 108
tcccgctctc gctctctgca tgctagct 28
<210> 109
<211> 19
<212> DNA
<213> Zea mays
<400> 109
tcctgccctc ttcacactg 19
<210> 110
<211> 19
<212> DNA
<213> Zea mays
<400> 110
tctttatcct taaataaga 19
<210> 111
<211> 19

<212> DNA
<213> Zea mays
<400> 111
tgccctcttc acactgacc 19
<210> 112
<211> 19
<212> DNA
<213> Zea mays
<400> 112
tatccttaaa taagaagca 19
<210> 113
<211> 19
<212> DNA
<213> Zea mays
<400> 113
tccttaaata agaagcata 19
<210> 114
<211> 19
<212> DNA
<213> Zea mays
<400> 114
tcttcacact gacccatgt 19
<210> 115
<211> 19
<212> DNA
<213> Zea mays
<400> 115
tgttttatca acgcccctc 19
<210> 116
<211> 19
<212> DNA
<213> Zea mays
<400> 116
tctatcaacg cccctccct 19
<210> 117
<211> 19
<212> DNA
<213> Zea mays
<400> 117

tatcaacgcc cctccctta 19

<210> 118

<211> 19

<212> DNA

<213> Zea mays

<400> 118

tgttgttcta tccctggaa 19

<210> 119

<211> 19

<212> DNA

<213> Zea mays

<400> 119

tgttctatcc ctggaaagc 19

<210> 120

<211> 19

<212> PRT

<213> Zea mays

<400> 120

Thr Cys Thr Ala Thr Cys Cys Cys Thr Gly Gly Ala Ala Ala Gly Cys

1

5

10

15

Thr Ala Thr

<210> 121

<211> 19

<212> DNA

<213> Zea mays

<400> 121

tatccctgga aagctatac 19

<210> 122

<211> 19

<212> DNA

<213> Zea mays

<400> 122

tctcagccag ttcttaact 19

<210> 123

<211> 19

<212> DNA

<213> Zea mays

<400> 123

tggaagcta tacctcctt 19

<210> 124
<211> 19
<212> DNA
<213> Zea mays
<400> 124
tatacctcct taccctat 19
<210> 125
<211> 19
<212> PRT
<213> Zea mays
<400> 125
Thr Ala Cys Cys Thr Cys Cys Thr Thr Ala Cys Cys Cys Cys Thr Ala
1 5 10 15
Thr Cys Ala
<210> 126
<211> 19
<212> DNA
<213> Zea mays
<400> 126
tccttaccct taccagctt 19
<210> 127
<211> 19
<212> DNA
<213> Zea mays
<400> 127
tctcttataa atacagacc 19
<210> 128
<211> 19
<212> DNA
<213> Zea mays
<400> 128
tcttataaat acagacctt 19
<210> 129
<211> 19
<212> DNA
<213> Zea mays
<400> 129
taccctatc agcttctcc 19
<210> 130

<211> 19
<212> DNA
<213> Zea mays
<400> 130
tataaatata gaccttgta 19
<210> 131
<211> 19
<212> DNA
<213> Zea mays
<400> 131
tatcagcttc tcctcacat 19
<210> 132
<211> 19
<212> DNA
<213> Zea mays
<400> 132
tacagacctt gtacaacac 19
<210> 133
<211> 19
<212> DNA
<213> Zea mays
<400> 133
tctcctcaca tctcctctc 19
<210> 134
<211> 19
<212> DNA
<213> Zea mays
<400> 134
tcctcacatc tcctctcgt 19
<210> 135
<211> 19
<212> DNA
<213> Zea mays
<400> 135
tgtacaacac tttcacctc 19
<210> 136
<211> 19
<212> DNA
<213> Zea mays

<400> 136
tacaacactt tcacctct 19
<210> 137
<211> 19
<212> DNA
<213> Zea mays
<400> 137
tctcctctcg tcgccaccc 19
<210> 138
<211> 19
<212> DNA
<213> Zea mays
<400> 138
tcctctcgtc gccacccat 19
<210> 139
<211> 19
<212> DNA
<213> Zea mays
<400> 139
tctcgtcgcc acccatgct 19
<210> 140
<211> 19
<212> DNA
<213> Zea mays
<400> 140
tcgtcgccac ccatgctat 19
<210> 141
<211> 19
<212> DNA
<213> Zea mays
<400> 141
tcgccaccca tgctatcac 19
<210> 142
<211> 19
<212> DNA
<213> Zea mays
<400> 142
tccctggaaa gctatacct 19
<210> 143

<211> 19
<212> DNA
<213> Zea mays
<400> 143
tgctatcacc gctctgac 19
<210> 144
<211> 19
<212> DNA
<213> Zea mays
<400> 144
tatcaccgct ctgatcaca 19
<210> 145
<211> 19
<212> DNA
<213> Zea mays
<400> 145
tctgatcaca agcaaggca 19
<210> 146
<211> 19
<212> DNA
<213> Zea mays
<400> 146
tctggcccct tcttgcct 19
<210> 147
<211> 19
<212> DNA
<213> Zea mays
<400> 147
tggccccttc ctgccctct 19
<210> 148
<211> 19
<212> DNA
<213> Zea mays
<400> 148
tcttctcccg ctctcgctc 19
<210> 149
<211> 19
<212> DNA
<213> Zea mays

<400> 149
tctcccgcctc tcgctctct 19
<210> 150
<211> 19
<212> DNA
<213> Zea mays
<400> 150
tgcatgctag ctaccttct 19
<210> 151
<211> 19
<212> DNA
<213> Zea mays
<400> 151
tcccgcctctc gctctctgc 19
<210> 152
<211> 19
<212> DNA
<213> Zea mays
<400> 152
tctcgctctc tgcatgcta 19
<210> 153
<211> 19
<212> DNA
<213> Zea mays
<400> 153
tcgctctctg catgctagc 19
<210> 154
<211> 19
<212> DNA
<213> Zea mays
<400> 154
tctctgcatg ctagctacc 19
<210> 155
<211> 19
<212> DNA
<213> Zea mays
<400> 155
tctgcatgct agctacctt 19
<210> 156

<211> 19
<212> DNA
<213> Zea mays
<400> 156
tatgtcaact tcacttgtc 19
<210> 157
<211> 19
<212> DNA
<213> Zea mays
<400> 157
tgtcaacttc acttgtctc 19
<210> 158
<211> 19
<212> DNA
<213> Zea mays
<400> 158
tgctagctac cttctagct 19
<210> 159
<211> 19
<212> DNA
<213> Zea mays
<400> 159
tagctacctt ctagctatc 19
<210> 160
<211> 19
<212> DNA
<213> Zea mays
<400> 160
taccttctag ctatctagc 19
<210> 161
<211> 19
<212> DNA
<213> Zea mays
<400> 161
tgtctctctc caaaagata 19
<210> 162
<211> 19
<212> DNA
<213> Zea mays

<400> 162
tctagctatc tagcctcta 19
<210> 163
<211> 19
<212> DNA
<213> Zea mays
<400> 163
tctctctcca aaagatatc 19
<210> 164
<211> 19
<212> DNA
<213> Zea mays
<400> 164
tagctatcta gcctctagg 19
<210> 165
<211> 19
<212> DNA
<213> Zea mays
<400> 165
tctctccaaa agatatcgt 19
<210> 166
<211> 19
<212> DNA
<213> Zea mays
<400> 166
tctccaaaag atatcgtat 19
<210> 167
<211> 19
<212> DNA
<213> Zea mays
<400> 167
tatctagcct ctaggtcca 19
<210> 168
<211> 19
<212> DNA
<213> Zea mays
<400> 168
tccaaaagat atcgtatca 19
<210> 169

<211> 19
<212> DNA
<213> Zea mays
<400> 169
tctagcctct aggtccaat 19
<210> 170
<211> 19
<212> DNA
<213> Zea mays
<400> 170
tagcctctag gtccaatgc 19
<210> 171
<211> 19
<212> DNA
<213> Zea mays
<400> 171
tctaggtcca atgcactcc 19
<210> 172
<211> 19
<212> DNA
<213> Zea mays
<400> 172
tatcgtatca cccatgggc 19
<210> 173
<211> 19
<212> DNA
<213> Zea mays
<400> 173
taggtccaat gcactccct 19
<210> 174
<211> 19
<212> DNA
<213> Zea mays
<400> 174
tcgtatcacc catgggcaa 19
<210> 175
<211> 19
<212> DNA
<213> Zea mays

<400> 175
tccaatgcac tccctcctt 19
<210> 176
<211> 19
<212> DNA
<213> Zea mays
<400> 176
tcttgccata gaccggaca 19
<210> 177
<211> 19
<212> DNA
<213> Zea mays
<400> 177
tgcactccct ccttataaa 19
<210> 178
<211> 19
<212> DNA
<213> Zea mays
<400> 178
tccctcctta taaacaagg 19
<210> 179
<211> 19
<212> DNA
<213> Zea mays
<400> 179
tccttataaa caaggaacc 19
<210> 180
<211> 19
<212> DNA
<213> Zea mays
<400> 180
tggccatgac cccctccc 19
<210> 181
<211> 19
<212> DNA
<213> Zea mays
<400> 181
tataaacaag gaaccctcc 19
<210> 182

<211> 19
<212> DNA
<213> Zea mays
<400> 182
tcccagcccc aacctatat 19
<210> 183
<211> 19
<212> DNA
<213> Zea mays
<400> 183
tccttcgcct ctcttgcca 19
<210> 184
<211> 19
<212> DNA
<213> Zea mays
<400> 184
tcgcctctct tgccataga 19
<210> 185
<211> 19
<212> DNA
<213> Zea mays
<400> 185
tctcttgcca tagaccgga 19
<210> 186
<211> 19
<212> DNA
<213> Zea mays
<400> 186
tatatcacct agcgagct 19
<210> 187
<211> 19
<212> DNA
<213> Zea mays
<400> 187
tatcacctag cgagctac 19
<210> 188
<211> 19
<212> DNA
<213> Zea mays

<400> 188
tagcgcagct acgctctct 19
<210> 189
<211> 19
<212> DNA
<213> Zea mays
<400> 189
tacgctctct tctcccgt 19
<210> 190
<211> 19
<212> DNA
<213> Zea mays
<400> 190
tgggcaatgg ccatgaccc 19
<210> 191
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 引物
<400> 191
ggtacagctg gtgatggta 19
<210> 192
<211> 18
<212> DNA
<213> 人工序列
<220>
<223> 引物
<400> 192
gactcttctt cctccctt 18
<210> 193
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 引物
<400> 193
cgtctccccc ttcaggatgt 20
<210> 194

<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 引物
<400> 194
gtccaacagg gacagttcca a 21
<210> 195
<211> 13
<212> DNA
<213> 人工序列
<220>
<223> 引物
<400> 195
accaccaatc ttg 13
<210> 196
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 引物
<400> 196
caggatgctg aaggagctct acta 24
<210> 197
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 引物
<400> 197
tggaaccagt agaagacgtt cttg 24
<210> 198
<211> 15
<212> DNA
<213> 人工序列
<220>
<223> 引物
<400> 198
atccggtcgc ccagc 15

<210> 199

<211> 978

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea mays WUS2 protein (wus2)

<400> 199

```

atggcggcca atgcgggcgg cgggtggagcg ggaggaggca gcggcagcgg cagcgtggct 60
gcgccggcgg tgtgccccc cagcggctcg cgggtggacgc cgacgccgga gcagatcagg 120
atgctgaagg agctctacta cggctgcggc atccggtcgc ccagctcgga gcagatccag 180
cgcatcaccg ccatgctgcg gcagcacggc aagatcgagg gcaagaacgt cttctactgg 240
ttccagaacc acaaggcccc cgagcgccag aagcgccgcc tcaccagcct cgacgtcaac 300
gtgcccggcg ccggcgcggc cgacgccacc accagccaac tcggcgctct ctcgtgtcg 360
tcgccgcctt caggcgcggc gcctccctcg cccaccctcg gcttctacgc cgccggcaat 420
ggcggcggat cggctgggct gctggacacg agttccgact ggggcagcag cggcgctgcc 480
atggccaccg agacatgctt cctgcaggac tacatgggcg tgacggacac gggcagctcg 540
tcgcagtggc catgcttctc gtcgtcggac acgataatgg cggcggcggc ggccgcggcg 600
cgggtggcga cgacgcgggc gcccagaca ctccctctct tcccacctg cggcgacgac 660
gacgacgacg acagccagcc cccgccgcgg ccgcggcacg cagtcccagt cccggcaggc 720
gagaccatcc gcggcgcgcg cggcagcagc agcagctact tgccgttctg ggggtgccgt 780
gccgcgtcca caactgccgg cgccacttct tccgttgcca tccagcagca acaccagctg 840
caggagcagt acagcttita cagcaacagc acccagctgg ccggcaccgg cagccaagac 900
gtatcggctt cagcggccgc cctggagctg agcctcagct catggtgctc cccttacct 960
gctgcaggga gcatgtga 978

```

<210> 200

<211> 879

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Arabidopsis thaliana Homeodomain-like superfamily protein (WUS)

<400> 200

```

atggagccgc cacagcatca gcatcatcat catcaagccg accaagaaag cggcaacaac 60
aacaacaaca agtccggctc tgggtggttac acgtgtcgcc agaccagcac gaggtggaca 120
ccgacgacgg agcaaatcaa aatcctcaaa gaactttact acaacaatgc aatccggtca 180
ccaacagccg atcagatcca gaagatcact gcaaggctga gacagttcgg aaagattgag 240
ggcaagaacg tcttttactg gttccagaac cataaggctc gtgagcgta gaagaagaga 300
ttcaacggaa caaacatgac cacaccatct tcatcaccca actcggttat gatggcggct 360
aacgatcatt atcatcctct acttcacat catcacggtg ttcccatgca gagacctgct 420

```

aattccgtca acgttaaact taaccaagac catcatctct atcatcataa caagccatat 480
 cccagcttca ataacgggaa tttaaatcat gcaagctcag gtactgaatg tgggtgttgtt 540
 aatgcttcta atggctacat gagtagccat gtctatggat ctatggaaca agactgttct 600
 atgaattaca acaacgtagg tggaggatgg gcaaacaatgg atcatcatta ctcatctgca 660
 ccttacaact tcttcgatag agcaaagcct ctgtttggtc tagaaggta tcaagaagaa 720
 gaagaatgtg gtggcgatgc ttatctggaa catcgacgta cgcttcctct cttccctatg 780
 cacggtgaag atcacatcaa cggtggtagt ggtgccatct ggaagtatgg ccaatcgaa 840
 gttcgccctt gcgcttctct tgagctacgt ctgaactag 879

<210> 201

<211> 795

<212> DNA

<213> 人工序列

<220>

<223> cDNA of *Triticum aestivum* cultivar Avalon WUSCHEL-Like-B1
 (WUSCHELL-B1) gene

<400> 201

atggcggcga cggcgactgc gacggcggcg gcgacgagcg tggtgacggg gacgacgcgg 60
 tggtgcccga cgccggagca gctgatgac ctggaggaga tgtaccgcgg cgggctgcgc 120
 accccaacg cgtcgcagat ccagcagatc acggcgcacc tggcccacta cggccgcac 180
 gagggcaaga acgtcttcta ctggttccag aaccacaagg cccgggaccg ccagaagctc 240
 cgccgcaggc tctgcatgag ccaccacctc ctctcctgcg cccactacta cgccgcccgc 300
 aacgccggcc agtaccacca ccagcagcag ctctcggcg cggcgcggt tccccctccg 360
 ctgctgcagc accagcagca gcagcagta tactccgcct cctgcgccgg cggcagctac 420
 gaccagcacc tgctcccgc gaccgtccca gcttccgctt atgctgctgc tgctgctggg 480
 tacgcctacc ccttcgccgc cgtgccggcg agccggtgcg ccgacccctc gccgccaac 540
 acgccgctgt ccttccatca ccagggtgga ggcgtagtag gatcgccgga gtactcactg 600
 gggaggctgg gcaacttcgg cgtggtggac gacacgtgcc ggccgtcgcg gtgcgagcag 660
 cagccacagc agctggccgt ggcgacggaa gatcaggcgg cgccggtgac ggcgacgggg 720
 ctgttctgcc ggccgctgaa gacgtggac ctcttccccg gcgcatcaa ggaggagcag 780
 cgcgatgtcg cctag 795

<210> 202

<211> 795

<212> DNA

<213> 人工序列

<220>

<223> cDNA of *Triticum aestivum* cultivar Cadenza WUSCHEL-Like-B1
 (WUSCHELL-B1) gene

<400> 202

atggcggcga cggcgactgc gacggcggcg gcgacgagcg tggtgacggg gacgacgcgg 60

tgggtgcccga cgccggagca gctgatgatc ctggaggaga tgtaccgcgg cgggctgcgc 120
 acccccaacg cgtcgcagat ccagcagatc acggcgcacc tggcccacta cggccgcac 180
 gagggcaaga acgtcttcta ctggttccag aaccacaagg cccgggaccg ccagaagctc 240
 cgccgcaggc tctgcatgag ccaccacctc ctctcctgcg cccactacta cgccgccgcc 300
 aacgccggcc agtaccacca ccagcagcag ctctcctggcg ccggcgcggt tccccctccg 360
 ctgctgcagc accagcagca gcagcagtac tactccgcct cctgcgccgg cggcagctac 420
 gaccagcacc tgctcccgac gaccgtccca gtttccgctt atgctgctgc tgctgctggg 480
 tacgcctacc ctttcgccgc cgtgccggcg agccggtgcg ccgaccctc gccgccaac 540
 acgccgctgt cttccatca ccagggtgga ggcgtagtag gatcgccgga gtactcactg 600
 gggaggctgg gcaacttcgg cgtggtggac gacacgtgcc ggccgtcgcg gtgcgagcag 660
 cagccacagc agctggccgt ggcgacggaa gatcaggcgg cgccggtgac ggcgacgggg 720
 ctgttctgcc ggccgtgaa gacgtggac ctcttccccg gcgcgatcaa ggaggagcag 780
 cgcatgtcg cctag 795

<210> 203

<211> 795

<212> DNA

<213> 人工序列

<220>

<223> cDNA of *Triticum aestivum* cultivar Badger WUSCHEL-Like-B1
 (WUSCHELL-B1) gene

<400> 203

atggcggcga cggcgactgc gacggcgcg gcgacgagcg tggtgacggg gacgacgcgg 60
 tgggtgcccga cgccggagca gctgatgatc ctggaggaga tgtaccgcgg cgggctgcgc 120
 acccccaacg cgtcgcagat ccagcagatc acggcgcacc tggcccacta cggccgcac 180
 gagggcaaga acgtcttcta ctggttccag aaccacaagg cccgggaccg ccagaagctc 240
 cgccgcaggc tctgcatgag ccaccacctc ctctcctgcg cccactacta cgccgccgcc 300
 aacgccggcc agtaccacca ccagcagcag ctctcctggcg ccggcgcggt tccccctccg 360
 ctgctgcagc accagcagca gcagcagtac tactccgcct cctgcgccgg cggcagctac 420
 gaccagcacc tgctcccgac gaccgtccca gtttccgctt atgctgctgc tgctgctggg 480
 tacgcctacc ctttcgccgc cgtgccggcg agccggtgcg ccgaccctc gccgccaac 540
 acgccgctgt cttccatca ccagggtgga ggcgtagtag gatcgccgga gtactcactg 600
 gggaggctgg gcaacttcgg cgtggtggac gacacgtgcc ggccgtcgcg gtgcgagcag 660
 cagccacagc agctggccgt ggcgacggaa gatcaggcgg cgccggtgac ggcgacgggg 720
 ctgttctgcc ggccgtgaa gacgtggac ctcttccccg gcgcgatcaa ggaggagcag 780
 cgcatgtcg cctag 795

<210> 204

<211> 795

<212> DNA

<213> 人工序列

<220>

<223> cDNA of *Triticum aestivum* cultivar Charger WUSCHEL-Like-B1
(WUSCHELL-B1) gene

<400> 204

```

atggcggcga cggcgactgc gacggcggcg gcgacgagcg tggtagacggg gacgacgcgg 60
tggtagcccg cggcgagca gctgatgac ctggaggaga tgtaccgcgg cgggctgcgc 120
acccccaaacg cgtcgagat ccagcagatc acggcgcacc tggcccaacta cggccgcac 180
gagggcaaga acgtcttcta ctggttccag aaccacaagg cccgggaccg ccagaagctc 240
cgccgcaggc tctgcatgag ccaccacctc ctctctgcg cccactacta cgccgccgcc 300
aacgccggcc agtaccacca ccagcagcag ctctctggcg cggcgcggt tccccctccg 360
ctgctgcagc accagcagca gcagcagtac tactccgct cctgcgccgg cggcagctac 420
gaccagcacc tgctcccgac gaccgtccca gttctcgtt atgtgtctgc tgctgtggg 480
tagcctacc ccttcgccgc cgtgccggcg agccggtgcg ccgaccctc gccgccaac 540
acgccgtgt ccttccatca ccagggtgga ggcgtagtag gatcgccgga gtactactg 600
gggaggctgg gcaacttcgg cgtggtggac gacacgtgcc ggccgtcgcg gtacgagcag 660
cagccacagc agctggccgt ggcgacggaa gatcaggcgg cgccggtgac ggcgacgggg 720
ctgttctgcc ggccgtgaa gacgtggac ctcttccccg gcgcatcaa ggaggagcag 780
cgcatgtcg cctag 795

```

<210> 205

<211> 795

<212> DNA

<213> 人工序列

<220>

<223> cDNA of *Triticum aestivum* cultivar Claire WUSCHEL-Like-B1
(WUSCHELL-B1) gene

<400> 205

```

atggcggcga cggcgactgc gacggcggcg gcgacgagcg tggtagacggg gacgacgcgg 60
tggtagcccg cggcgagca gctgatgac ctggaggaga tgtaccgcgg cgggctgcgc 120
acccccaaacg cgtcgagat ccagcagatc acggcgcacc tggcccaacta cggccgcac 180
gagggcaaga acgtcttcta ctggttccag aaccacaagg cccgggaccg ccagaagctc 240
cgccgcaggc tctgcatgag ccaccacctc ctctctgcg cccactacta cgccgccgcc 300
aacgccggcc agtaccacca ccagcagcag ctctctggcg cggcgcggt tccccctccg 360
ctgctgcagc accagcagca gcagcagtac tactccgct cctgcgccgg cggcagctac 420
gaccagcacc tgctcccgac gaccgtccca gttctcgtt atgtgtctgc tgctgtggg 480
tagcctacc ccttcgccgc cgtgccggcg agccggtgcg ccgaccctc gccgccaac 540
acgccgtgt ccttccatca ccagggtgga ggcgtagtag gatcgccgga gtactactg 600
gggaggctgg gcaacttcgg cgtggtggac gacacgtgcc ggccgtcgcg gtacgagcag 660
cagccacagc agctggccgt ggcgacggaa gatcaggcgg cgccggtgac ggcgacgggg 720
ctgttctgcc ggccgtgaa gacgtggac ctcttccccg gcgcatcaa ggaggagcag 780

```

cgcgatgtcg cctag 795

<210> 206

<211> 795

<212> DNA

<213> 人工序列

<220>

<223> cDNA of *Triticum aestivum* cultivar Spark WUSCHEL-Like-B1
(WUSCHELL-B1) gene

<400> 206

atggcggcga cggcgactgc gacggcggcg gcgacgagcg tgggtgacggg gacgacgcgg 60
tgggtccccga cgccggagca gctgatgatc ctggaggaga tgtaccgcgg cgggctgcgc 120
acccccaaacg cgtcgcagat ccagcagatc acggcgcacc tggcccacta cggccgcac 180
gagggcaaga acgtcttcta ctggttcag aaccacaagg cccgggaccg ccagaagctc 240
cgccgcaggc tctgcatgag ccaccacctc ctctctgcg cccactacta cgccgccgcc 300
aacgccggcc agtaccacca ccagcagcag ctctctggcg ccggcgcggt tccccctccg 360
ctgctgcagc accagcagca gcagcagtac tactccgcct cctgcgccgg cggcagctac 420
gaccagcacc tgctcccgac gaccgtccca gcttccgctt atgctgctgc tgctgctggg 480
tacgcctacc ccttcgccgc cgtgccggcg agccggtgcg ccgaccctc gccgccaac 540
acgccgctgt ccttccatca ccagggtgga ggcgtagtag gatcgccgga gtactcactg 600
gggaggctgg gcaacttcgg cgtggtggac gacacgtgcc ggccgtcgcg gtacgagcag 660
cagccacagc agctggccgt ggcgacggaa gatcaggcgg cgccggtgac ggcgacgggg 720
ctgttctgcc ggccgctgaa gacgctggac ctcttccccg gcgcatcaa ggaggagcag 780
cgcgatgtcg cctag 795

<210> 207

<211> 2130

<212> DNA

<213> 人工序列

<220>

<223> cDNA of *Zea mays* AP2-like ethylene-responsive transcription
factor BBM2 (LOC103650883)

<400> 207

atggccactg tgaacaactg gctcgtttc tccctctccc cgcaggagct gccgccctcc 60
cagacgacgg actccacgct catctcggcc gccaccgccg accatgtctc cggcgatgtc 120
tgcttcaaca tcccccaaga ttggagcatg aggggatcag agctttcggc gctcgtcgcg 180
gagccgaagc tggaggactt cctcggcggc atctccttct ccgagcagca tcacaagtcc 240
aactgcaact tgataccag cactagcagc acagtttgcg acgcgagctc agctgctagc 300
accggctacc atcaccagct gtaccagccc accagctccg cgtccactt cgcgactcc 360
gtcatggtgg cctcctcggc cgggtgccac gacggcggtt ccatgctcag cgcgccgcc 420
gctaacggtg tcgctggcg tgccagtgcc aacggcggcg gcatcgggct gtccatgatc 480

aagaactggc tgcggagcca accggcgccc atgcagccga gggcgggcggc ggctgagggc 540
gcgcaggggc tctctttgtc catgaacatg gcggggacga cccaaggcgc tgctggcatg 600
ccacttctcg ctggagagcg cgcacgggcg cccgagagtg tatcgacgtc agcacagggt 660
ggtgccgtcg tcgtcacggc gccgaaggag gatagcgggtg gcagcgggtg tgccgggtgct 720
ctagtagccg tgagcacgga cacgggtggc agcggcgggc cgtcggctga caacacggca 780
aggaagacgg tggacacgtt cgggcagcgc acgtcgattt accgtggcgt gacaaggcat 840
agatggactg ggagatatga ggcacatctt tgggataaca gttgcagaag ggaaggacaa 900
actcgtaagg gtcgtcaagt ctatttaggt ggctatgata aagaggagaa agctgctagg 960
gcttatgata ttgctgctct gaagtactgg ggtgccacaa caacaacaaa ttttccagt 1020
agtaactacg aaaaggagct cgaggacatg aagcacatga caaggcagga gtttgtagcg 1080
tctctgagaa ggaagagcag tggtttctcc agaggtgcat ccatttacag gggagtgact 1140
aggcatcacc aacatggaag atggcaagca cggattggac gagttgcagg gaacaaggat 1200
ctttacttgg gcaccttcag caccaggag gaggcagcg aggcgtagc catcgcgcg 1260
atcaagttcc gcggcctcaa cgccgtcacc aacttcgaca tgagccgcta cgacgtgaag 1320
agcatcctgg acagcagcgc cctccccatc ggcagcgccg ccaagcgtct caaggaggcc 1380
gaggccgcag cgtccgcgca gcaccaccac gccggcgtgg tgagctacga cgtcggccgc 1440
atgcctcgc agctcggcga cggcggagcc ctagcggcg cgtacggcg gcactaccac 1500
ggcgccgcct ggccgacat cgcgttccag ccgggcgccg ccaccacagg cctgtaccac 1560
ccgtacgcgc agcagccaat gcgcggcggc ggggtggtga agcaggagca ggaccacgcg 1620
gtgatcgcgg ccgcgcacag cctgcaggac ctccaccact tgaacctggg cgcgcccggc 1680
gcgcacgact ttttctcggc agggcagcag gccgccgccg cagctcgat gcacggcctg 1740
gctagcatcg acagtgcgtc gctcgagcac agcaccggct ccaactccgt cgtctacaac 1800
ggcggggtcg gcgatagcaa cggcgccagc gccgttggca gcggcggtgg ctacatgatg 1860
ccgatgagcg ctgccggagc aaccactaca tcggcaatgg tgagccacga gcagatgcat 1920
gcacgggcct acgacgaagc caagcaggct gtcagatgg ggtacgagag ctacctggtg 1980
aacgcggaga acaatggtgg cggaaggatg tctgcatggg ggaccgtcgt ctctgcagcc 2040
gcggcggcag cagcaagcag caacgacaac attgccgccg acgtcggcca tggcggcgcg 2100
cagctcttca gtgtctggaa cgacacttaa 2130

<210> 208

<211> 1707

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Arabidopsis thaliana Integrase-type DNA-binding
superfamily protein (PLT2)

<400> 208

atgaattcta acaactggct cgcgttcct ctatcacaa ctcactcttc tttgccgcct 60
cacattcact cttcacaaaa ttctcatttc aatctaggtt tggtaacga caatatcgac 120
aacccttttc aaaaccaagg atggaatatg atcaatccac atggtggagg cggcgaagg 180

ggagaggttc caaaagtggc tgatttctta ggagtgcga aatcggggga tcatcacacc 240
 gatcacacc tcgtacctta taacgacatt catcaaacca acgcctccga ctactacttt 300
 caaaccaata gcttgttacc tacagtcgtc acttgtgcct ctaatgctcc taataattat 360
 gagcttcaag agagtgcaca caatttgcaa tctctcactc tctctatggg aagtactgga 420
 gctgccgctg cagaagtcgc cactgtgaaa gcctcgccgg ctgagactag tgccgataat 480
 agtagcagca ctaccaacac aagtggagga gccatcgttg aggctacacc gagacggact 540
 ttggaaaactt ttggacaacg aacctctatc tatcgtggag ttacaagaca tagatggacc 600
 ggtagatatg aagctcatct ttgggataat agctgtagaa gagaaggaca atcaaggaaa 660
 ggaagacaag tctacttagg tgggtatgac aaagaagaga aagcagccag agcatatgat 720
 ctagctgcac ttaaataattg gggtccctct actactacca actttccgat aactaactac 780
 gagaaggaag tagaggagat gaaaaacatg acgagacaag agtttgtggc ttctataaga 840
 aggaaaagta gcggtattct gcgtggtgca tccatgtatc gtggagtaac aaggcatcat 900
 caacatggaa gatggcaagc aaggatcggc cgagttgctg gaaacaaaga tctctacttg 960
 ggaacattca gcacggagga agaagcagca gaagcttatg acatagctgc gataaagttt 1020
 cgaggtctaa acgcggttac aaactttgag ataaatcggg atgatgtgaa agccatcctg 1080
 gagagcaaca cacttcctat aggaggtggg gcggctaaac ggctcaaaga agctcaagct 1140
 ctagaatcat caagaaaacg agaggaaatg atagccctcg gatcaaattt ccatcaatat 1200
 ggtgcagcga gcggctcgag ctctgttgct tccagctcta ggcttcagct tcaaccttac 1260
 cctctaagca ttcaacaacc ttttgagcat cttcatcatc atcagccttt acttactcta 1320
 cagaacaaca acgatatctc tcagtatcat gattccttta gttacattca gacgcagctt 1380
 catcttcacc aacaacaaac caacaattac ttgcagtctt ctagtcacac ttcacagctc 1440
 tacaatgctt atcttcagag taaccctggg ctgcttcatg gatttgtctc tgataataac 1500
 aacacttcag ggtttcttgg aaacaatggg attgggtattg ggtcaagctc taccgttgga 1560
 tcatcggctg aggaagagtt tccagccgtg aaagtcgatt acgatatgcc tccttccggg 1620
 ggagctacag ggtatggagg atggaatagt ggagagtctg ctcaaggatc gaatccagga 1680
 ggtgttttca cgatgtggaa tgaataa 1707

<210> 209

<211> 1818

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Beta vulgaris subsp. vulgaris AP2-like

ethylene-responsive transcription factor PLT2 (LOC104889956)

<400> 209

atgggctcaa tgaattcaaa caattggttg tcttttctc tttctctac acatccttca 60
 cttcaatcac atcttcaaac caatgattca caacctcatc aacaattctc cttgggtctt 120
 gtatctgacc acattgacaa cccctttggg caagcgcaag aatggaactt gctcaatcca 180
 caagggccaa atgaagtacc caaaatagca gatttcttag gagtagggaa ttcagaaact 240
 catcattcac cagaccttac agcgttcagt gacatgagcc aaggtggtga atcagattat 300

cttttctccg gcaacggcgg cggcttaatg gcggtgcaaa acaccgtagc agcagctact 360
 aatagtagcc aatatgatca ataccaagag aactctaata attgcttgca atctttgact 420
 ctatcaatgg gaagtagtgg acaacagcct caacaacagc aacaaccacc ttcaagcact 480
 aataattgtg agactagtgg tgacaataat agcaccgcta gtgtcgccgc ctctactgcc 540
 gccactgtca ccaccgcgat tactcctgtg gttgaagcca cccctaggag aaccttgat 600
 acttttggcc aaaggacttc tatttataga ggtgttaciaa ggcataggatg gacaggaaga 660
 tatgaagctc atctttggga taatagtgtt agaagggaag gacagtcaag gaagggtcgt 720
 caagtgtatc ttggagggtg tgataaggaa gagaaggccg ctaggtctta tgatttagct 780
 gcaatcaagt attggggaac ttcaactact acaaattttc caataagcaa ctatgagaaa 840
 gaaatagaag acatgaaaca catgactaga caagaatttg tagcagctat tagaaggaag 900
 agtagtgat tctctagagg tgcatacaat tatcgtggtg taacaagaca ccatcaacat 960
 gggagatggc aagcaagaat tggaagggtg gcaggaaaca aggatctcta cttaggaaca 1020
 ttttagcacag aggaagaggc tgcagaagct tatgatatcg cggctatcaa gtttagaggc 1080
 cttaatgctg tgacaaatth tgacatgagc cggatgatg ttaaagccat cctagagagc 1140
 aacactcttc ccataggagg aggggcggcg aagcgcctta aggaagctca agctatagaa 1200
 tcctctagga agagggaaga aatgcttgcc ctaagcaata gtagctacc atatggagct 1260
 agtagctcga gctcgactcg atatggagcc catcaacaag caacaactca tgcataccct 1320
 ttgttaccat accaccatca agaccatcaa ccacaacctt tgctaaccct aaaaaataac 1380
 catggtcaag aaagcaatat ttccctatca cattactctc aagaggctca attccttcag 1440
 ttgtaccaac aatcaagtta ctcaaacct agtagcatgt acaacaatta cctccaaact 1500
 aaccctagtt tgcttcatgg gttcatgaac atgggctcaa actcttgtgg tgttattgat 1560
 actaacaata ctaatggaag ttcaagtggg agttatagt gtggagggtg ccttggtggt 1620
 ggggctggga tcaatgccat gggtgccgcc tcgacaacga gcaatgcggt gggttccggt 1680
 gaaccggagc cacttgcat ggtgaagggt gactatgata tgccttctgc tgggtggtggt 1740
 ggaggaagtt atgaggggtg gtcaactgag acggttcaag gacctaataa tgggggttttt 1800
 acaatgtgga atgactaa 1818

<210> 210

<211> 2157

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Beta vulgaris subsp. vulgaris AP2-like

ethylene-responsive transcription factor BBM (LOC104890283)

<400> 210

atgggttcaa tgaattggtt aggtttctct ttatctctc aagaacttcc ttcacaaact 60
 cctgatcatg gtagtaatca agatccat catcatcact ttacaagcaa caacaatgga 120
 gagtgtttcg atctcgggcc cggtcaacg cctcattctt ctctcaatca catcccttct 180
 tcctttggaa tccttgaggc cttccataga tcaactaat atcaatccca agattggaac 240
 aatatgaagg gaaactcaga gcttagtatg ctaatgggaa accaagaagt tgaagaggag 300

```

ccaaaactag aaaactttct agggagtagt cactctttta gagagaatca tcatcaaaat 360
aatggagatc tctacatggt taatactaca catgataaca acaataatag tactatgtca 420
aaccctaagg atattactag tcctgctagt aataataata ataataataa taacggactc 480
aatgtttcaa tgatcaagac atggttgaga tcaaaccacc ctcctcaatc aaatatagtg 540
gatggtggtg gtggcagtgg tggcggcggg gcgaatgcac aaacattatc cttttcaatg 600
ggaactggtg tgtcccaatc cgccttgccg ctactagcgg caggaggagg aggtggtggt 660
ggtggaggag agatagagag tagtttgtct gagaatagta gtagtaataa taaacaacaa 720
ttaagtata caacggccgg gatatgtaat aacacagcta gtactattac tgctatcggt 780
gatgttcaaa gtagtgcact agaaagcggt ctaggaaat ctattgatac atttgacaa 840
cgtacatcca tttaccgtgg tgtaacaaga cataggtgga ctgggagata tgaagctcat 900
ctatgggata atagctgtag gagagaaggg cagactcgta agggcagaca agttttatttg 960
gggggttatg acaaagaaga aaaagcggt agagcttatg atttggtgc acttaaatat 1020
tggggtacca ctaccaccac caactttcct attactgatt atgaaaagga agttgaggat 1080
atgaagcata tgacacgcca agaatatgtg gcactcttac gaaggaaaag tagtggttt 1140
tctcgtggtg catcaattta tcgaggagta acaaggcatc atcagcatgg tcgttggtcaa 1200
gcaaggatag gtagggttgc aggcaacaaa gacctctacc tgggaacttt cagtacacaa 1260
gaagaagcag cagaagcata tgatatagca gcaataaagt ttaggggatt aaatgcagta 1320
acaaactttg agataaacag gtatgatgtg aaagccatac ttgatagcac cacacttcct 1380
ataggaggag cagcaaagag gttaaaagat gtggaggatt taaccacaat tactccagat 1440
aaacagatta ttagggcaat tacttcgagt aatgataata atcatgaaa ttctcagctt 1500
actaattttg gtaatgggac tccaatttc cattcctggc ctggaatcgc attcccacaa 1560
gctcaaccac ttgcaatgca ttacccttat gcaacttctc aacaacaaca acaacaacaa 1620
caaaggtttt ggtgtaagca agaagttcaa gatactacta atgattacca agatcatctt 1680
aatcagcagc ttcaaatgaa taatgggaca cataatttct ttcagatgca taatttgatg 1740
gggttggtgaga attcttctac tagtttggtg catagtcttg ggtcgaattc cgctgtttat 1800
gggaatggga atgggaatgg gaatggaaat gatcatggtg ttgggaatgg gtatggatta 1860
ccctttggga tgtcaacagt aattgctcat gatgggaatg ggaatggaag tgggaatggg 1920
aatgaacaaa gtgggtatga gaattattac tatctttcac accaaggaaa taataataat 1980
catggtaatg ctgctggtgt aagaggagct gttgggactt atgatcaagg gtcagcttgt 2040
aacaattggg tccaacggc gattccgaca ctggttccga ggccgaataa tatggcggct 2100
gttggtggtc atggtggagg aggaatccct actttcactg tgtggaatga cacctaa 2157

```

<210> 211

<211> 1884

<212> DNA

<213> 人工序列

<220>

<223> cDNA of *Triticum aestivum* WANT1-2 mRNA for AP2 transcription factor

<400> 211

atgagagcga tggccagcgg cggcggcaac tggttaggct tctccctctc cccgcacatg 60
 gccatggagg tgccctcctc ctctgaaccc gaccacgctc agcctgctag cgctagtgtc 120
 atgtctgctt ctcccaccaa cgccgcgacc tgcaacctcc tattctccca acccgcgcaa 180
 atggccgctc cacctcctgg atactactac gtccggcgcg cctatgggga tggcaccagc 240
 accgctggcg tctactactc ccaccacccc gtcatgcccc tcacgtccga tggatctctg 300
 tgcacatggg aagggatgat gccgtcgtcc tcgccgaagc tcgaggactt cttgggtggc 360
 ggcaatggca gcggacatga cgcggtcacc tactacagcc accagcagca ggaccaacaa 420
 gaccaggagg caagcagaat ctaccagcac catcaacagc agcagcagca gctagcgccc 480
 tacaacttcc agcacttgac ggaagcagag gcgatctacc aagaggccac ggcgccgatg 540
 gacgaggcaa tggccgctgc caagaaccag ctggtgacga gctacggctc atgctacagc 600
 aacgcgggga tgcagccgct gagcctgtcc atgagcccca ggtcccagtc cagcagctgc 660
 gtcagcgcag ctctcagca gcatcagatg gctgcggctg ctgctgctgc ctcttggct 720
 gcttcccagg gaggcagtaa tgggtggtggg gagcaggagc agtgcgtggg gaagaagagg 780
 ggcaactggga agggaggcca gaagcagccc gttcatcgca agtccatcga cacgtttggg 840
 cagaggacct cccagtatag gggcgtcacc aggcacaggt ggactgggag atatgaagcc 900
 cacctctggg acaacagctg caagaaggat gggcagacaa ggaaaggag gcaagtttat 960
 ctaggtggtt atgacaatga agacaaggct gccagggctt atgatctggc tgctctgaaa 1020
 tattgggggc cgtcgacgaa caccaatttc ccgctagaaa attatcgaga ggaggtcgag 1080
 gagatgaaaa gcatgacaag gcaggaattc gttgcacact tgagaaggag aagcagcggg 1140
 ttttctcgtg gtgcttcgat atatcgagga gtaacaggc atcatcagca tggaagatgg 1200
 caagctagga ttggcagggt tgctggcaac aaagacttgt atctcggcac ttccaccact 1260
 caggaagaag cagccgaggc ctacgacgta gcccgcatca agttccgtgg cctgaacgcc 1320
 gtgaccaact tcgacataac cagatacgac gtggacaaga tcatggagag cagctctctg 1380
 ctgcccgggt acgaagcgcg caaggtcaag gcggtcgagg cagccaacca cgtgcctgcc 1440
 atgcacaacg gcggcgggga gatcagccat gccgaagaag gaagctccgg cgtctggagg 1500
 atggtactcc atggaacacc gcagcaagct gcacagtga ccccgagggt ggcagacctt 1560
 cagaagggtc tcatgggcgg cggcgaccct cgctcgtccc tgcattggcat cgccgggttc 1620
 gacgtcgagt cggcggcgca tgacatcgac gtctcaggca agatcaacta ctccaacccg 1680
 tcctccctgg tgaccagcct cagcaactcg agagagggga gccagagag gttcagcctg 1740
 ccctcgctgt acgccaagca tcccaacgcc gtcagcgtcg ccagcatgag cccgtggatg 1800
 gcgatgccag cgccggccgc cgcccacgtg ttaagggggc cgaattctc catgcctgtg 1860
 ttcgtgcct ggacggacgc atag 1884
 <210> 212
 <211> 1896
 <212> DNA
 <213> 人工序列
 <220>
 <223> cDNA of *Triticum aestivum* WANT1 mRNA for AP2 transcription factor
 <400> 212

atgagagcga tggccagcgg cggcggcaac tggttagggt tctccctctc cccgcacatg 60
 gccatggagg tggcctctc tgaacccgac cacgctcagg ctcaacctgc tagcgctagc 120
 gctatgtccg cttctccac aaacgccgac acctgcaacc tcctattctc ccaacccgag 180
 caaatggccg ctccacctcc tggctactac tacgtcggcg gcgcctatgg ggatggcacc 240
 agcaccgccg gcgtctacta ctcccaccac tccgtcatgc ccatcacgtc cgatggatcc 300
 ctgtgcatca tgggaaggat gatgccatcg tctcggcga agctcgagga cttcttgggt 360
 ggcggcaatg gaagtgggca cgacgcggtc acctactaca gccaccacca gcagcagcag 420
 gaccaacagg accaggaggc aagcagaatc taccagcacc atcagcagca gctagcgccc 480
 tacaacttcc agcacttgac ggaaacggag gcgatctacc aagagaccac ggcgccgatg 540
 gatgaggcaa tggccgctgc caagaacctg ctctgacga gctatggctc atgctacagc 600
 aacgcgggga tgcagccgct gagcctgtcc atgagccca ggtcccagtc cagcagctgc 660
 gtcaccgcag ctctcagca gcatcagatg gctgcggctg ctgctgctgc tgctgcctct 720
 atggctgctt cccaggagg cagtaatggt ggtggggagc agtgcgtggg gaagaagagg 780
 ggcaactggga agggaggcca gaagcagccc gttcaccgca agtccatcga cacgtttggg 840
 cagaggacct cccagtatag gggcgtcacc aggcacaggt ggactgggag atatgaagcc 900
 cacctgtggg acaacagttg caagaaggat gggcagacaa ggaaaggag gcaagtttat 960
 ctaggtggtt atgataatga agacaaggct gccagggctt atgatctggc tgctctgaaa 1020
 tactgggggc cgtcgacgaa caccaatttc ccgctagaaa attatcgaga ggaggtcgag 1080
 gagatgaaaa gcatgacaag gcaggaattc gttgcacact tgagaaggag aagcagcggg 1140
 ttttctcgtg gtgcttcgat atatcgagga gtaacaggc atcatcagca tggaagatgg 1200
 caagctagga ttggcagggt tgctggcaac aaagacttgt atctcggcac ttccaccact 1260
 caagaagaag cagccgaggc ctatgacgta gccgcgatca agttccgtgg cctgaacgcc 1320
 gtgaccaact tcgacataac cagatacgac gtggacaaga tcatggagag cagctctctg 1380
 ctgcccgggg acgaagcgcg caaggtcagg ccgatcgagg cggccaacca cgtgccttcc 1440
 atgcacaacg gcggcgggga gctcagccat gccgaagaag gaagctcagg cgtctggagg 1500
 atgggtgctc atggaacacc gcagcaagct gcacagtga ccccgagggt ggccgacctt 1560
 cagaagggtc tcatggacgg cgaccctcgc tcgtccctgc atggcaatgg cattgccggg 1620
 ttgcagctcg agtctgccg gcacgacatc gacgtttcag gcaagattaa ctactccaac 1680
 tcgtcttccc tggtagaccag cctcagcaac tcgagagagg ggagccccga gaggttcagc 1740
 ctgccctcgc tgtacgcaa gcacccaac gccgtcagcc tcgccaccat gagcccgagg 1800
 atggcgatgc cggcgccgac cgccaccac gcgttagagg ggccgaattc ctccatccct 1860
 cccatgcctg tgtttgctgc ctggacagac gcatag 1896

<210> 213

<211> 2382

<212> DNA

<213> 人工序列

<220>

<223> cDNA of *Triticum aestivum* clone: tplb0046e23, cultivar Chinese Spring

<400> 213

```

gggttggtccc ctccctctca ttccttttgc tcagctcacg ggtccctctc gcccgctctc 60
ctcgtagtctc acttctcttt taccaccact gcctccatct ccatgtcgtc gctcggacaa 120
gggtagtggg gccgcagtag cagtagagct cagctcagag tgaaagcgaa gcaagaagcg 180
ttttcgtctg tgtttgtttg ttgatgagag cgatggccag cggcggcaac tggtaggct 240
tctccctctc cccgcacatg gccatggagg tgccctctc ctctgagccc gaccacgctc 300
agcctgctag cgctagcgct atgtccgctt ctcccacaa cgcgcgccacc tgcaacctcc 360
tcttctcccc tccctcgcaa atggccgctc cacctcctgg ctactactac gtcggcgggg 420
cctacgggga tggcaccagc accgcggcg tttactactc ccaccacccc gtcatgcccc 480
tcacgtccga tggatccctg tgcacatgg aagggatgat gccgtcgtcc tcgccgaagc 540
tcgaggactt cttgggtggc ggcaatggca gtgcgcacga cgcggtcacc tactacagcc 600
accaccagca gcagcagcag gaccaacagg accaggaggt aagcagaatc taccagcacc 660
atcagcagca gctagcgcgc tacaacttcc agcacttgac ggaggcagag gcgatctacc 720
aagaggccac ggcgccgacg gatgaggcaa tggccgctgc caagaacctg ctcgtgacga 780
gctatggctc atgtctacagc aacgcgggga tgcagccgt gagcctgtcc atgagcccca 840
gggtcccagtc cagcagctgc gtcagcgag ctctcagca gcatcagatg gctgcggttg 900
ctgctgcggc tgctgcctct atggttgctt cccaggagg cagtaatggt ggtggggagc 960
agtgcgtggg gaagaagagg ggactggga agggaggcca gaagcagccc gttcatcgca 1020
agtccatcga cacgtttggg cagaggacct cccagtatag ggcggtcacc aggcacaggt 1080
ggactgggag atatgaagcc cacctgtggg acaacagttg caagaaggat gggcagacaa 1140
ggaaagggag gcaagtttat ctaggtggtt atgacaatga agacaaggct gccagggctt 1200
atgatctggc tgctctgaaa tattgggggc catcgacgaa caccaatttc ccgctagaaa 1260
attatcgaga ggaggctcag gagatgaaaa gcatgacaag acaggaattc gttgcacact 1320
tgagaaggag aagcagcggg ttttctctg gtgcttcgat atatcgagga gtaacaggc 1380
atcatcagca tggaagatgg caagctagga ttggcagggt tgctggcaac aaagacttgt 1440
atctcggcac ttccaccact caggaagaag cagctgaggc ctacgacgta gcggcgatca 1500
agttccgtgg cctgaacgcc gtgaccaact tcgacataac cagatacgac gtggacaaga 1560
tcatggagag cagctctctg ctgcccgggg acgaagcgcg caaggtcagg ccgatcgagg 1620
cagccagcca cgtgtctccc atgcacaacg gcggcgggga gctcagccat gccgaagaag 1680
gaagctccgg cgtctggagg atggtgctcc atggaacacc gcagcaagct gcgccgtgca 1740
cccccgaggg ggccgacctt cagaagggtt tcatggacgg cgaccctcgc tcgtccctgc 1800
atggcaatgg cattgccggg ttcgacgtgg agtctgcggc gcatgacatc gacgtctcag 1860
gcaagatcaa ctactccaac tcgtcttccc tggtagaccag cctcagcaac tcgagagagg 1920
ggagccccga gaggttcagc ctaccctcgc tgtacgcaa gcatcccaac gccgtcagcc 1980
tcgccagcat gagcccgtgg atggcgatgc cggcgccgac cgcgcgccac acgttgaggg 2040
gaccgaattc ctccatccct tctatgcctg tgtttgtctg ctggacggac gcatagccgt 2100
gttgacagctg ctcaaatctt gctgtcactg gccatgttgt agtaaatgg agctggatta 2160
gtagcgtcgt tgctcatgtc gcttaagttt aatctgggaa ggctggttaa ttggttatca 2220
cgaaggcggg gtagtggtag tggtagtgg acgtaggaga agcatgcatt agtctctagc 2280

```

tcaccgaact tgtagcagta cgtagtgttc ttacttactt tcttttgagc ctataacaat 2340
gcatggaagg aggctgtccc aagaaaaaaa aaaaaaaac ga 2382
<210> 214
<211> 2528
<212> DNA
<213> 人工序列
<220>
<223> cDNA of *Triticum aestivum* clone: WT012_J17, cultivar: Chinese Spring
<400> 214
gacacacgcg cgcacagacc aaagtccccc ttcaaaccgg ctgagcttgc aatggagagc 60
agcggcatca ttgcgacatg tgtcccccaa tgattgatec tctcattccc atctaagcta 120
gatcttcttg aatcttgaga ccaccacagc ctcatcccca gtcgtgctcg tgcgcccttg 180
ctcccatccg ctccgcccga tgaccaacgg cggccacagc atgagcggcg ccagcatcgc 240
gagcgggtgct ggcggtctggc tgggtttctc gctgtcgctt cacgtcgcca tggaggcggc 300
ggccggctcc ggcatcgtcg acgtggccgg ccaccaccac ggcgagcacg gcgggggtcta 360
ctatcacctt gacgcggtcg cctcctcccc catgtccttc tacttcggtg ggagcgacaa 420
tgtcggcgcc gcgagcggcg ggtactactc cgggatctcc gactgcctc tcaggtccga 480
cggctccctc tgctcgcggc acgcgctccg gaggagcgag cagaaacacc acggggcgga 540
ggtgtcggcg ccgccgaagc tcgaggactt cctgggcgcg agtcccgcga tggcgctgag 600
cctggacaac tcgggctact actacggcgg ccaaggccat ggccatggcg acgcaggagg 660
cggccagcac cagctgccgt acgcatgat gcctggctcc ggtggccacc acatgtacta 720
cgacgcccac gcggcggtgc tggacgagca ggctgcagcc acgtcggccg cgatggaagc 780
ggccggctgg atggcgctg ccggagacgt ctacgacgtg gacgccggca acggcgagga 840
cgccatcgtg gcgaccggcc acgacaaccc cgggtgggtac gtacaccgcg tgacgctgtc 900
catgagctcc ggggtcccagt ccagctgcgt caccatgcag caggcggctg cacacgccc 960
cgcctacgtc ggtgccggcg gcgagtgcgt cggccaggcg accgcggcca gcaagaagcg 1020
cggcgcgggc gccgggcaga acaagcagcc ggtcgtgcac cgcaagtgca tcgacacctt 1080
cggccagcgc acgtccaagt accggggcgt caccaggcat aggtggacgg ggaggtatga 1140
ggcgcacctc tgggacaaca gctgccggaa ggaaggccag accaggaaag gccggcaagt 1200
ttatcttggg gggatatgaca tggaggagaa ggcggcgagg gcgtatgacc tcgcggcgct 1260
caagtactgg gcgcgtcca cgcacatcaa ctccccggtg gaggactacc aggaggagct 1320
ggaggtgatg aagaacatga ccaggcagga gtatgtggct cacctcagaa ggaagagcag 1380
cgggttctcg cgcggcgctt cgggtgtacc gggagtcacc aggcaccacc agcaggggcg 1440
gtggcaggcg cgcatcggcc gcgtctccgg caacaaggac ctctacctcg gcacattcag 1500
cgcggaggcg gacgcggcgg aggcgtacga cgtggcgggc atcaagttec gcggcctcaa 1560
cgcggtcacc aacttcgaca tcaaccgcta cgacgtggac aagatcatgg agagcagcac 1620
gctcctgccc ggcgaccagg tgcggcgcag gaaggacggc cccgacgaga gcgccgccgt 1680
ggtggcaagc gcggcgggcg ccctcgtgca ggccggcagc gccgcggact actggaggca 1740

gcctgcggcg gtagaccacgg aagagcacag ccgccaccac ctggaccttc tgtcgagcga 1800
gtccttctcc ctgctgcgcg gcgtggtgtc cctggacggc gacgcggctg gtgctcaggg 1860
gcagggcaac cgcatgtcgg gcgcgtcgtc cctggccacg agcctgagca actcccggga 1920
gcagagccccg gaccagggag gcggcctggc catgctgttc gcccggccccg aggcgccgaa 1980
gctggcgagc tcgctgccc tgggcacctg ggtctcatcg ccggcgccgg ccaggcccgg 2040
tgtgtccgtg gcgcacatgc cagtgttcgc gcgtggggc gacgcctgac ttgctcgact 2100
acagcgtcgt ccttttggcc ctgcatccac gaggagatag caaggttggt taactaggac 2160
tggttaccta gcattagtag ctgcgttagc aaggaactgt aaggtgggtt tattagccat 2220
agctggtagc ttagcggcgc atgcatgcat ctgcctgggc tctcgtgggt ccttccccag 2280
ctgcgtctgg gacgaagggt tttttagta tcgagccatg gcacggcagc agcagcgtcg 2340
cctccggccc ggcgagagc cgccgccgt gatcgagct ggatgggtag ctgtagctcc 2400
tgtctctaga cctcctaact ttcacaaac caaatgttg gaccttcgtg ttcgtgtggc 2460
ctcgcggcgc gtctgaacat ctgatttttt tatttttttt gagggtaagc aaaaaaaaaa 2520
aaaaacga 2528

<210> 215

<211> 1803

<212> DNA

<213> 人工序列

<220>

<223> cDNA of *Triticum aestivum* PARG-2D

<400> 215

atgaccaaca acaacggcaa tggcaatggc ggcagcaacg cggcgggcag tggctggctg 60
ggcttctcgc tctcgccgca catggacgaa cacaaccacg tgcagcagca gcaacagcac 120
cagggcctat tctaccccag ctccgtcgcc gccgcctaca gcctcggcgg cgacgtcgcc 180
accgacgggt actattcgca gctagcctcc atgcctctca agtcagacgg ctccctctgc 240
atcatggaag ctctacgccg aaccgatcaa caagatcacc acggtccgaa gctggaggac 300
tttctgggcg cggggcaacc ggcgatggcg ctgagcctgg acaacacctc caacttctat 360
tactacggcg gcggtggcgg agccggtggg caacacggac agagccacgg cggcagcttc 420
ctgcagcaag catacgacgt gtacagcggg cccgcaacgg catcggtgct ggcgcccaat 480
gaggacgccg cggcagccac ggccatggcg aactgggtgc aggtcgcgcg cggtgccacc 540
gcgtacgcca cagccgagaa cgtcttgctc gcggcggcgg accggcagca gcatcttcac 600
caccaccctc tggcactctc catgagctcc gccgggtcgc tctccagctg cgttaccgcg 660
ggggccgagt acggcggcgt cggggcgacg gtggacggcg ggcgaaagcg cggcgggcgcg 720
acggcggggc agaagcagcc ggtgcaccac cgcaagtcca tcgacacgtt cgggcagcgc 780
acgtcgcagt accgtggcgt caccaggcat aggtggacgg ggcggtatga ggcgcacctg 840
tgggacaaca gctgcaagaa ggaaggccag accaggaaag ggaggcaagt ttacctcgga 900
ggatatgaca tggaggagaa ggcggcgaga gcctacgacc aggcggcgct caagtactgg 960
ggcccttcca cccatatcaa cttcccgtc gagactacc agcaggagct ggaggagatg 1020
aagaacatga cgaggcagga gtacgtggca caccttagaa ggaagagcag cggcttctcg 1080

cgtggcgcgt ccatgtaccg tggcgtgacc cggcaccacc agcacgggcg gtggcaggcg 1140
 cgcatcggcc gcgtctccgg caacaaggac ctctacctcg gcactttcgg caccagagag 1200
 gaggccgcgg aggcgtacga catcgccgcc atcaagttcc ggggcctcaa cgccgtcacc 1260
 aacttcgaca tcacccgcta cgacgtcgac aagatcatgg ccagcaaacac gtcctctccc 1320
 ggcgagcacg ccaggcgcaa caaggacgac aacgccgcgc ccctgcccct ccccgcccc 1380
 gacgactgcg ccgcctctgc cctggtgccc gtgtccactc cggggacgga caccggcggc 1440
 agcggccagc accgctacca cgacgtcatg tcctcgggcg aggccttctc ggcgctacac 1500
 gacctggtca ccgtggacgg ccacaccgcg caggcgggga acggcgcgca cgtgcacatg 1560
 tcgatgtcgg gcgcatcgtc gctggtgacg agcctgagca actcccgaga ggagagccca 1620
 gaccggggcg gcgggctgtc catgctcttc gccaaagccgc cgcagcagcc ggccacgaca 1680
 acggcggcgt ccccgaaagt gatgagcact ctgaagccgc tgggctcctg ggcgtcgtcg 1740
 gcgaggccgg ccgccgtttc catcgtcac atgcccattg tcgccgcgtg gagcgacgca 1800
 tga 1803

<210> 216

<211> 1806

<212> DNA

<213> 人工序列

<220>

<223> cDNA of *Triticum aestivum* PARG-2A

<400> 216

atgaccaaca acaacggcaa tgggaatggc ggcagcaacg cggcggcgag tggctggctg 60
 ggcttctcgc tctcgccgca catggacgaa cacaaccacg tgcagcagca gcagcaacaa 120
 caccagggcc tattctaccc cagctccgtc gccgccgctt acagcctcgg cagcgacgtc 180
 gccaccggcg ggtactattc gcagctagcc tccatgcctc tcaagtcaga cggctccctc 240
 tgcatcatgg aagctctacg ccgaaccgat caacaagatc accacggtcc gaagctggag 300
 gactttcttg gcgcggggca accggcgatg gcgctgagcc tggacaacac ctccaacttc 360
 tattactaca gcggcggttg cggagcaggt gggcaaacac gacagagcca cggcggcggc 420
 ttctgcagc aagcatacga cgtgtacggc gggcccgcaa cggcatcggg gctggcggcc 480
 gatgaggacg ccgcggcagc cacggccatg gcgaactggg tgcaggtcgc gcgcgggtgc 540
 accgcgtacg ccacagccga gaacgtcttg tccgcggcgg cggaccggca gcagcatctt 600
 caccaccacc ctctggcact ctccatgagc tccgccgggt cgctctccag ctgcgttacc 660
 gcggggggccg agtacggcgg cgtcgtggcg acggtggacg gcgggcgaaa acgcggtggc 720
 gcgacggcgg ggcagaagca gccggtgcac caccgcaagt ccatcgacac gttcgggcag 780
 cgcacgtcgc agcaccgtgg cgtcaccagg cataggtgga cggggcggtg tgaggcgcac 840
 ctgtgggaca acagctgcaa gaaggaaggc cagaccagga aaggaggca agtttacctc 900
 ggagggtatg acatggagga gaaggcggcg agagcctacg accaggcggc gctcaagtac 960
 tgggggcctt ccacccatat caacttcccg ctcgaggact accagcagga gctggaggag 1020
 atgaagaaca tgacgaggca ggagtacgtg gcacacctta gaaggaagag cagcggcttc 1080
 tcgcgtggcg cgtccatgta ccgtggcgtg acccggcacc accagcacgg gcggtggcag 1140

gcgcgcatcg gccgcgtctc cggcaacaag gacctctatc tcggcacttt cggcaccag 1200
gaggaggccg cggaggcgta cgacatcgcc gccatcaagt tccggggact caacgccgtc 1260
accaacttcg acatcacccg ctacgacgtc gacaagatca tggccagcaa cacgtcctc 1320
ccgggcgagc tcgccaggcg caacaaggac gccaacgccg cggccctgcc cctccccgcc 1380
cccgcagact gcgccgcctc tgccctgggtg cccgtgtcta ctccggggac ggacaccggc 1440
ggcagcggcc agcaccgaaa ccaggacgtc atgtcctcgg gcgaggcctt ctgggcgctg 1500
cacgacctgg tcaccgtgga cggccacacc gcgcaggggc gcaacggcgc gcgcgtgcac 1560
atgtcgatgt cgggcgcatc gtcgctgggtg acgagcctga gcaactcccg cgaggagagc 1620
ccagaccggg gcggtggcct gtctatgtc ttcgccaagc cgccgcagca gccggccacg 1680
acaacggcgg cgtccccgaa gctgatgagc actctggcgc cgctgggttc ctgggcgctg 1740
tcggcgaggc cggccgccgt ttccatcgct cacatgccc tgttcgccgc gtggagcgac 1800
gcatga 1806

<210> 217

<211> 2040

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea mays BBM

<400> 217

atggcttcag cgaacaactg gctgggcttc tcgctctcgg gccaggataa cccgcagcct 60
aaccaggata gctcgctgc cgccgtatc gacatctccg gcgccagcga cttctatggc 120
ctgcccacgc agcagggtc cgacgggcat ctggcggtgc cgggcctgcg ggacgatcac 180
gcttcttatg gtatcatgga ggcctacaac agggttcctc aagaaacca agattggaac 240
atgaggggct tggactacaa cggcggtggc tcggagctct cgatgcttgt ggggtccagc 300
ggcggcggcg ggggcaacgg caagagggcc gtggaagaca gcgagcccaa gctcgaagat 360
ttcctcggcg gcaactcgtt cgtctccgat caagatcagt ccggcggtta cctgttctct 420
ggagtccga tagccagcag cgccaatagc aacagcggga gcaacacat ggagctctcc 480
atgatcaaga cctggctacg gaacaaccag gtggcccagc ccagccgcc agctccacat 540
cagccgcagc ctgaggaaat gagcaccgac gccagcggca gcagctttgg atgctcgat 600
tcgatgggaa ggaacagcat ggtggcggct ggtgggagct cgcagagcct ggcgtctcg 660
atgagcacgg gctcgcacct gccatgggtt gtgcccagcg gcgccgccag cggagcggcc 720
tcggagagca catcgtcgga gaacaagcga gcgagcgggt ccatggattc gcccggcagc 780
gcggtagaag ccgtaccgag gaagtccatc gacacgttcg ggcaaaggac ctctatatat 840
cgagggttaa caaggcatag atggacagg cggtatgagg ctcatctatg ggataatagt 900
tgtagaagg aagggcagag tcgcaagggt aggcaagttt accttggtgg ctatgacaag 960
gaggacaagg cagcaagggc ttatgatttg gcagctctca agtattgggg cactacgaca 1020
acaacaaatt tccctataag caactacgaa aaggagctag aagaaatgaa acatatgact 1080
agacaggagt acattgcata cctaagaaga aatagcagtg gattttctcg tggggcgta 1140
aagtatcgtg gagtaactag acatcatcag catgggagat ggcaagcaag gatagggaga 1200

gttgcaggaa acaaggatct ctacttgggc acattcagca ccgaggagga ggcggcggag 1260
gcctacgaca tcgccgcgat caagttccgc ggtctcaacg ccgtcaccaa cttcgacatg 1320
agccgctacg acgtgaagag catcctcgag agcagcacac tgcctgtcgg cgggtgcggcc 1380
aggcgccctca aggacgccgt ggaccacgtg gaggccggcg ccaccatctg gcgcgccgac 1440
atggacggcg ccgtgatctc ccagctggcc gaagccggga tgggcggcta cgcctcgtac 1500
ggccaccacg gctggccgac catcgcgttc cagcagccgt cgccgctctc cgtccactac 1560
ccgtacggcc agccgtcccc cgggtgggtg aaacccgagc aggacgcggc cgccgccgcg 1620
gcgcacagcc tgcaggacct ccagcagctg cacctcggca gcgcggccca caacttcttc 1680
caggcgctgt cgagctccac agtctacaac ggcggcggcg gcgccagtgg tgggtaccag 1740
ggcctcgggtg gtggcagctc tttcctcatg ccgtcgagca ctgtcgtggc ggcgcccgac 1800
caggggcaca gcagcacggc caaccagggg agcacgtgca gctacgggga cgaccaccag 1860
gaggggaagc tcatcggtta cgacgccgcc atggtggcga ccgcagctgg tggagaccgc 1920
tacgtgcgg cgaggaacgg gtaccagttc tcgcagggt cgggatccac ggtgagcatc 1980
gcgagggcga acgggtacgc taacaactgg agctctcctt tcaacaacgg catggggtga 2040

<210> 218

<211> 963

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea mays WUS1

<400> 218

atggcggcca acgtgggcgc gggcaggagt gctggcggcg gcggagccgg cactggcact 60
ggcactgctg ctggcagcgg cggcgtgtcg acggccgtgt gccgccctag cggctcgcgg 120
tggacgccga cgccggagca gatcaggatc ctcaaggagc tctactacgg ctgcggcatc 180
cggtcgcccc actcggagca gatccagcgc atcaccgcca tgctgcggca gcacggcaag 240
atcgagggca agaactctt ctactggttc cagaaccaca aggcccgcga gcgccagaag 300
cgccgcctca ccaacctcga cgtcaacgtg cccgtcgccg ccgacgacag cgcccaccgc 360
cttggcgtcc tctcgttgtc gccttcttca ggttgttcag gcgcggcgcc tccgtcgccc 420
accctcggct tctacgccgg cggcaatggc tccgtgtga tgctggacac gagttccgat 480
tggggcagcg ctgctgccat ggccactgag gcatgttca tgcaggacta catgggcgtg 540
atgggcggcg cgtcaccgtg ggcatgtctc tctcgtcgt cggaggaccg gatggcggcg 600
ctggcgctgg cgccgaaggt gaccggggcg cccgagacgc tccctctctt ccgaccggc 660
ggcggcggag acgataggca gccccgcgg ccgcggcagt ctgtcccagc aggcgaggcc 720
atccgcggcg gcagcagcag cagcagctac ctccgttct ggggtgccgc gccacccca 780
actggcagtg ccacttccgt tgcgatccag cagcaacacc agctgatgca gatgcaagag 840
cagtacagct ttacagcaa cgcccagctg ctgcccggca ccggcagcca ggatgcagca 900
gcaacatccc tggagctgag cctcagctcc tgggtgctcc cttaccctgc agggaccatg 960

tga 963

<210> 219

<211> 978

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea mays WUS2

<400> 219

```

atggcggcca atgcgggcgg cgggtggagcg ggaggaggca gcggcagcgg cagcgtggct 60
gcgccggcgg tgtgccgccc cagcggctcg cgggtggacgc cgacgccgga gcagatcagg 120
atgctgaagg agctctacta cggctgcggc atccggctcg ccagctcgga gcagatccag 180
cgcatcaccg ccatgctgcg gcagcacggc aagatcgagg gcaagaacgt cttctactgg 240
ttccagaacc acaaggcccc cgagcgccag aagcgccgcc tcaccagcct cgacgtcaac 300
gtgcccggcg ccggcgcggc cgacgccacc accagccaac tcggcgctct ctcgctgtcg 360
tcgccgcctt caggcgcggc gcctccctcg cccaccctcg gcttctacgc cgccggcaat 420
ggcggcggat cggctgggct gctggacacg agttccgact ggggcagcag cggcgctgcc 480
atggccaccg agacatgctt cctgcaggac tacatgggcg tgacggacac gggcagctcg 540
tcgcagtggc catgcttctc gtcgtcggac acgataatgg cggcggcggc ggccgcggcg 600
cgggtggcga cgacgcgggc gcccagaca ctccctctct tcccgacctg cggcgacgac 660
gacgacgacg acagccagcc cccgccgcgg ccgcggcacg cagtcccagt cccggcaggc 720
gagaccatcc gcggcgggcg cggcagcagc agcagctact tgccgttctg ggggtgccgt 780
gccgcgtcca caactgccgg cgccacttct tccgttgcca tccagcagca acaccagctg 840
caggagcagt acagctttta cagcaacagc acccagctgg ccggcacggc cagccaagac 900
gtatcggctt cagcggccgc cctggagctg agcctcagct catggtgctc cccttacct 960
gctgcaggga gcatgtga 978

```

<210> 220

<211> 975

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea mays WOX2

<400> 220

```

atggagacgc cacagcagca atccgccgcc gccgccgccg ccgccgcca cgggcaggac 60
gacggcgggt cgccgccgat gtcgccggcc tcgccgcggc cggcggcgct ggccaacgcg 120
cgggtggaacc cgaccaagga gcaggtggcc gtgctggagg ggctgtacga gcacggcctg 180
cgcccccca gcgcggagca gatacagcag atcacgggca ggctgcggga gcacggcgcc 240
atcgagggca agaacgtctt ctactggttc cagaaccaca agggccgcca gcgccagagg 300
cagaagcagg acagcttcgc ctacttcagc aggtctctcc gccggcccc gccgctgccc 360
gtgctctcca tgcccccgcc gccaccgtac catcacgcc gcgtcccggc gccgccgcg 420
ataccgatgc cgatggcgcc gccgccgcc gctgcatgca acgacaacgg cggcgcgcg 480
gtgatctaca ggaaccatt ctacgtggct gcgccgcagg cgccccctgc aaatgccgcc 540

```

tactactacc cacagccaca gcagcagcag cagcagcagg tgacagtcac gtaccagtac 600
 ccgagaatgg aggtagccgg ccaggacaag atgatgacca gggccgcggc gcaccagcag 660
 cagcagcaca acggcgccgg gcaacaaccg ggacgcggcg gccaccccag ccgcgagacg 720
 ctccagctgt tcccgtctca gcccaccttc gtgctgcggc acgacaaggg gcgcgcccgc 780
 aacggcagta ataacgactc cctgacgtcg acgtcgacgg cgactgcgac agcgacagcg 840
 acagcgacag cgtccgcttc catctccgag gactcggatg gcctggagag cggcagctcc 900
 ggcaagggcg tcgaggaggc gcccgcgctg ccgttctatg acttcttcgg gctccagtcc 960
 tccggaggcc gctga 975

<210> 221

<211> 666

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea mays WOX5

<400> 221

atggaggcgc tgagcgggcg ggtaggcgtc aagtgcgggc ggtggaaccc tacggcggag 60
 caggtgaagg tcctgacgga gctcttccgc gcggggctgc ggacgcccag cacggagcag 120
 atccagcgca tctccacca cctcagcgcc ttcggaaggc tggagagcaa gaacgtcttc 180
 tactggttcc agaaccacaa ggcccgcgag cgccaccacc acaagaagcg acgccgcggc 240
 gcgtcgtcgt cctccccga cagcggcagc ggcaggggaa gcaacaacga ggaagacggc 300
 cgtggtgccg cctcgcagtc gcacgacgcc gacgccgacg ccgacctcgt gctgcaaccg 360
 ccagagagca agcgggaggc cagaagctat ggccaccatc accggctcgt gacatgctac 420
 gtcagggacg tgggtggagca gcaggaggcg tcgccgtcgt gggagcggcc gacgaggag 480
 gtggagacgc tagagctctt cccctcaag tcgtacggcg acctcgaggc ggcggagaag 540
 gtccggctcgt acgtcagagg aagcggcgcc accagcgagc agtcagggga gttgtccttc 600
 ttgcacgtcg tctccgccgg ccgggatccg ccgctcgagc tcaggctctg cagcttcggt 660
 ccctag 666

<210> 222

<211> 1521

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea Mays WOX8

<400> 222

atggcgtcct cgaacaggca ctggccgagc atgtacaggt ccagtctcgc ctgcaacttc 60
 cagcagccgc agccgcagcc tgacatgaac aacggcggca agtcctcact catgtcctca 120
 aggtgcgagg agaacggcgg aaggaacccg gagccgaggc cgcggtggaa cccgcggccg 180
 gagcagatca ggatcctgga agggatcttc aactccggca tgggtgaaccc gccgcgcgac 240
 gagatccgcc gcatccgcct ccaactgcag gactacgggc ccgtcggcga cgccaacgtc 300

ttctactggt tccagaaccg caagtcccg accaagcaca agctgcgcgc cgcggggcag 360
 ctgcagccgt cgggctcggg ccgctccgcc ctgcaggcgc gcgctgctgc cccggcgccc 420
 gtgacgcctc ccaggaacct gcagctcgcg gccgctgctc ccgtggcgcc gccacgtcc 480
 tcgtcctcgt cgtcctccga ccggtcctcg gggatcatcat cgagcaagtc ggtgaccgtg 540
 accccgacga ccgccgtcgc gcttgcttct cccgcaggcg ccgcgccggc tgctgtcttc 600
 cgccagcagg gcgtgatgcc gacgacggcc atggacctgc ttacgccgct gccgtcgtcg 660
 tcggccgctc tggccgcgcg ccagctctac tatcagtacc acagccagat catggcgctt 720
 gccgcgccgc cgatgcccga tacggtgatc gcctctccgg agcagttcct tccgagtggt 780
 cagcagggcg gacagcagca ttattacctg ccggccaccg agctcggtgg cgtcctcgac 840
 ggccactccc accacacaca cgagccccg gcggccatac accggcccgt ctgctctcta 900
 cccagcgtgc tctttggcct gtgcaacgaa gctctaaggc aagactactg cgccgacatc 960
 agcgtcgtcc ccaccaaggg actcggccat ggccaccagt tctggaacag caccacctgc 1020
 ggctctgata tgggcaatag caatagcaag atcgacgccg tgagcgccgt gatcagggac 1080
 gacgagaagt ccaggctggg gttactccac tactacggct tggcgggcgc gacgacgacc 1140
 gctgctgcgg ctgtcgtcc ggcccccttc gctgcagatg ccgccgccgg tacggccacg 1200
 ctgcttccaa gctctgcggc gagcgaccag ttgcaagggc tgttgatgc tgctgggctg 1260
 ctgatggggg agacgccgcc gacgccgacg gcgacggtgg tggccgtggc ccgggacgcc 1320
 gtgacgtgcg cggccaccgc caccgcgcag ttcagcgtgc cggcgtcgat gcgcctggac 1380
 gtgaggctgg cgttcggcga ggccgccctt ctggcgccgc acaccggcga ggcggtcccc 1440
 gtcgacgagt ccggcgctac ggtggagccg ctccagcagg acactctcta ctacgtgctc 1500
 atgcaggcga ctaataactg a 1521

<210> 223

<211> 822

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea mays WOX10

<400> 223

atggagtggg tggacaggac caaggcctcc gccgccgccg ccgcagcggc ggccgacgag 60
 agggctgggg gagcggaagg gctcgcggga tacgtcaagg tcatgaccga cgaacagatg 120
 gaggtgctcc gcaagcagat ctccatctac gccaccatct gcgagcagct tgtcgagatg 180
 caccgcgccc tcaccgagca ccaggacacc attgcaggaa ttaggttttag taatctgtac 240
 tgtgatcctc aaattatccc tggaggccac aagatcacag caaggcaacg atggcaacca 300
 acaccaatgc agctgcagat cttggagaac atctttgacc aaggcaatgg aacaccaagc 360
 aagcagagga taaaggagat aacggcagag ctctcgcacc atggccaaat ctcgagaca 420
 aatgtgtaca actggttcca gaacagacgg gcacggtcaa agcgggaagca ggccgcttct 480
 ttaccgaaca atgctgaatc tgaagctgag gtggacgagg agtctctcac cgataagaag 540
 ccgaagttag atcggtcgtc ccaggacaac aaggctatgg gcgctcaca cgctgacagg 600
 atatctggga tgcatcactt ggacactgat catgaccaa tccgtggcat gatgtatgga 660

tgcaatgaca acggcttgag atcgtctggc agttctggcc agatgtcctt ctacgggaac 720
 atcatgccga atccaagaat cgatcatttc ccggggaagg tggagagctc ccggagcttc 780
 tcccatctcc aacacgggga aggctttgac atgtttggat ga 822

<210> 224

<211> 849

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea mays WOX13

<400> 224

atggactggg ggaacaggac caaggccgcc gccgccgctg cggcgccgga cgagagggcc 60
 gggggagggg aagggtcgg aggatacgtc aaggtcatga ccgacgaaca gatggaggtg 120
 ctccgcaagc agatctccat ctacgccacc atctgcgagc agcttgtcga gatgcatcgc 180
 gtcctcaccg agcaccagga caccattgca ggattgaggt ttagcaatct gtactgtgac 240
 cctctaataca tccccggcgg tcacaagatc acggcaaggc agcgggtggca accaacaccg 300
 atgcagctgc agatcctgga gagcatcttc gaccagggca acgggacacc gagcaagcag 360
 aagataaagg agataacagc ggagctctcg cagcacggcc agatctcgga gacgaacgtg 420
 tacaactggg tccagaacag gcgggcacgg tcgaagcgga agcaggccgc tgcttcctta 480
 ccgaacaacg ccgaatccga agccgaggcg gacgaggagc ctctcgccga caagaagccg 540
 aagtcagaca ggccgccgcc gccgccgccg ccgatccagg ataataccaa ggctacgggc 600
 gctctcagcg ccgacagggt ctctgggtggg acgcgtcact tggacacggg tcatgaccag 660
 accagtggcg tgatgtatgg gtgcaacgac agtggcttgt tgagatcgtc cggcagttcg 720
 ggccagatgt ccttgtacga gaacttcatg tcgaatccaa gaatcgatcg tttcccgcg 780
 aaggtggaga gctcccggag cttcccccat ctccaacaac acggggaagg ctttggcatg 840
 tttggatga 849

<210> 225

<211> 795

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea mays Lec1

<400> 225

atggactcca gcttcctccc tgccggcgcg gacaatggct cggcgggagg cgccaacaat 60
 ggcgggcgcg ctgctcagca ggcgccgccg atccgcgagc aggaccggct gatgccgac 120
 gcgaacgtca tccgcatcat gcggcgcgctg ctgccggcg cgcgaagat ctcgacgac 180
 gccaaaggaga cgatccagga gtgctgtctg gactacatca gcttcatcac gggggaggcc 240
 aacgagcggg gccagcggga gcagcgcaag accatcaccg ccgaggacgt gctgtgggcc 300
 atgagccgcc tcggcttcga cgactacgtc gagccgctca gcgtctacct ccaccgtac 360
 cgcgagttcg agggcgaggc gcggggcgctc ggctctgccc cggcccctcc gcgcggcgac 420

caccaccacc accaccactc cgtgccgcca tcgatgctca acaagtcccg cgggcccggc 480
tccggagccg tcatgctacc gcaccaccac caccacgaca tgcacgcctc catgtacggg 540
ggcgccgtgc ccccgccgcc gcaccacggc ttcctcatgc cacaccaca gggcgggcac 600
tacctgcctt acccctacga gcccacgtcg tacggcgggc agcacgcctt ggccagcggg 660
tactatggag gggccgcgta cgcgccgggc aacaacggcg ggagcggcga tggcagcggc 720
gggagcgcgt cgcacgcacc gccggggcggc agcggcgggc gcttcgacca cccgcacacg 780
ttcgcgtaca agtag 795

<210> 226

<211> 1179

<212> DNA

<213> 人工序列

<220>

<223> cDNA od Zea mays Lec2

<400> 226

atgccagccc gcgcctccca cccggcgctt gccacctcgc gcgcgcgcgg ttggccgcgc 60
ctgcgcgccc tcggcatcgc ccccgacggg gggcgttggc gttgcctccc ccactttgca 120
cccatttcag agcccggccc acaacttgta ccgcgcgccc ccgcctccgc gtctccgccc 180
gcccgcccc atccggctat aaaagcctcg ccctctccaa ccctagccgc cgctgccgct 240
gccgcgcgcg ccgctacctc ctcccttcct tccttctccg ctctctctcg ttctaccggc 300
atggccggca ttaccaagcg ccgcacctcc ccggcctcca cctcctcttc gtccggcgac 360
gtcttgccgc agcgggtcac ccggaagcgt cggctccgcc gccgcggggc ccggagcacc 420
gcccgtaggc cgtcggcgcc tccacctatg aatgaactgg acttgaatac agctgctctt 480
gatccggatc attatgctac aggattgaga gttcttcttc agaaggagct ccgaaatagc 540
gatgtaagcc agcttgggag aattgttctc ccaaagaagg aggcggagtc ttacctccct 600
attctgatgg caaaggatgg aaagagtta tgcatgcatg acttgctaaa ttcacaactg 660
tggaacctca agtatagata ttggttcaac aacaaaagca ggatgtatgt gcttgaaaat 720
accggagatt atgtaaaagc tcatgacctt cagcaaggag acttcatcgt gatctacaag 780
gacgacgaga acaaccgctt tgtcatagga gcaaagaagg caggagatga gcagaccgcc 840
actgtacctc aagtccatga acacatgcac atctctgccg cactgccagc tccacaagcg 900
ttccatgact atgcaggccc cgtcgcagca gaagctggta tgctcgcgat cgtgccacag 960
ggtgacgaga tattcgacgg catactgaac tccttgccgg agataccagt ggcgaaacgtg 1020
aggtaactccg acttcttcga cccgttcggg gactccatgg acatggcgaa tccgctgagc 1080
tcctccaata acccctcggt caacctggct acgcatttcc atgacgagag gatcggggagc 1140
tgctcgtttc cctacccaaa atccgggcct cagatgtga 1179

<210> 227

<211> 1026

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea mays WIND1_1

<400> 227

```

atggccgcag ccatcgacat gtacaagtac tacaatacca gcgcacacca gatccccctcc 60
tcatccccct cggatcagga gctcgcgaaa gactcgcagc cttttataac gagggtcttc 120
tcctcttcat cctcctcccc ctaccatggc tactcgtcct ctccatccat gtcccaagat 180
tcttacatgc ctacaccctc ttacaccage tacgccacct cgcctcttcc cactccccgc 240
gccgcctcct cctcgcagct tccgccgctc tactcgtcgc cttatgcggc gccgtgcatg 300
gccggccaga tgggcctgaa ccagctcggc ccggcccaga tccagcagat ccaggcccag 360
ttcatgttcc agcagcagca gcagcagcag aggggcctgc acgcggcggt cctgggccccg 420
cgggcgcagc cgatgaagca gtcagggtcg ccgtcgcgcg cgcgcgcgct ggcgccggcg 480
cagtcgaagc tgtaccgcgg cgtgcggcag gccactggg gcaagtgggt ggcgagatc 540
cggctccccg agaaccgcac gcggtgtgtg ctcggcacct tcgacaccgc ggaggacgcg 600
gcgctgcctt acgacaaggc ggcttccgc ctccgcggcg acacggcgcg cctcaacttc 660
ccggccctcc ggcgcgggcg cgcgcacctc gccggcccgc tgcacgcctc cgtggacgcc 720
aagctgaccg ccatctgcca gtccctgtcg gaggccaagt ccaagagcgg ctgctccggc 780
gacgagtcgg ccgctcccc gccggactcc cccaagtgtc cggcgtcgac gacggaggga 840
gagggggagg aggagtcggg ctccgccggc tcccctctc ctctctctc tccccgacg 900
ctggcgccgc ccgtgccgga gatggcgaag ctggacttca cggaggcgcc gtgggacgag 960
acggaggcct tccacctgcg caagtaccgc tcctgggaga tcgactggga ttccatctg 1020
tcatga 1026

```

<210> 228

<211> 951

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea mays WIND1_2

<400> 228

```

atggccgcag ccatagacat gtacaagtac tgcaatacca gcgcacacct tatcgccctcc 60
tcgtccccct cggatcagga gctcgcgaaa gactcgcagc cttttataac gagggtcttc 120
tccccctacc atcgctactc gttggcccca gattcttaca tgccctacac ctctctctac 180
accacctcgc ctcttccca cccacacctc tcgcctttct cgcagcttcc gccactctac 240
tcgtgcctt acgcggcttc gacggcgctc ggctgtggtg ggccgatggg cctgaaccag 300
ctcgccccgg ccagatcca gcagatccag gccagctca tgttccagca ccagcagcag 360
aggggcctgc acgcggcggt cctgggcccc cgggcgcagc cgatgaagca gtccgggtcg 420
ccgccggcgc agtcgaagct gtaccgcggc gtgcgccagc gccactgggg caagtgggtg 480
gcggagatcc gcctccccaa gaaccgcacg cggctgtggc tcggcacctt cgacaccgcc 540
gagggcgcg cgtggccta cgacgaggcg gccttccgcc tccgcggcga cacggcgcg 600
ctcaacttcc cgtccctccg ccgcggcggc ggcgcgcgcc tcgccggccc gctccacgcc 660
tccgtggacg ccaagctcac cgccatctgc cagtccctgg cggggtccaa gaacagctcg 720

```

tccagcgacg agtcggccgc gtccctgccg gactccccc a gtcgacggag 780
 ggggatgagg actcggcctc cgccggctcc cctccttccc cgacgcaggc gccgccctg 840
 ccggagatgg cgaagctgga cttcaccgag gcgccgtggg acgaaacgga ggccttccac 900
 ctgcgcaagt acccgctcctg ggagatcgac tgggattcca tcctctcatg a 951

<210> 229

<211> 702

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea mays ESR1_1

<400> 229

atggcgccga gaacgtcaga gaaaacccatg gcaccggcgg cggccgctgc cacggggctc 60
 gcgctcagcg tcggcgccgg cggcgggggc ggcgggccgc actacagagg cgtgaggaag 120
 cggccgtggg gccggtacgc ggcgagatc cgcgaccgg cgaagaagag ccgggtgttg 180
 ctcggcacct acgacacggc cgaggacgcc gcgcgggcct acgacgccgc cgcgcgcgag 240
 taccgcggcg ccaaggccaa gaccaacttc cttaccctt cgtgcgtgcc cctctccgca 300
 gccggttgcc ggagcagcaa cagcagcacc gtcgagtcct tcagcagcga cgcgcaggcg 360
 cccatgcagg ccatgccgct cccgccgtcg ctcgagctgg acctgttcca ccgcgcggcg 420
 gccgcggcca cgggcacggg cgctgccgcc gtacgcttcc ctttcggcag cateccccgtt 480
 acgcacccgt actacttctt cgggcaggcc gcagccgcag ccgcggaagc aggggtgccgt 540
 gtgctcaagc tggcgccggc ggtcaccgtg gcgcagagcg actccgactg ttcgtcggtta 600
 gtggatctgt cgccgtcgcc accggccgct gtgtcggcga ggaagccgcg cgcgttcgat 660
 ctcgacctga actgctcacc gccgacggag gcggaagcct ag 702

<210> 230

<211> 885

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea mays ESR1_2

<400> 230

atggaggacg tggccaacgc acacatctac gcccacgccc accggagcaa gcgtccccag 60
 tcggccgcga tcaaagacgg ggacggggac gtcgacctgt ccatgaaagg cgcgcggtac 120
 cgcggcgtgc ggcgccggcc gtggggccgg ttgcggcgag agatccgcga ccccatgtcc 180
 aaggagcggc ggtggctcgg cacccttcgac accgccgagc aggccgcctg cgcctacgac 240
 atcgcgggcg gcgccatgcg cggcaacaag gcgcgcacca acttcccggg ccacgccacg 300
 gcgggctact ggccgtgggg cgcgccgcag ccggcgccgg tggcgacccc gateaacctt 360
 ttctctctgc acaacctcat catgagctcc tccaaccacg gctgccgcct gctcaaccac 420
 gcaggccacg gacacgtcca ctccgcagcc cccagacctc cggcgccggc ggcgagcgcc 480
 acgtccacga ccatcgcagc gcccttcctt gtcgccgcac accccgccgt agcgatggac 540

gaggacgtgg acgactggga cggcgtcctg cggagcgagc ccgcggacgc cgggctgctg 600
 caggacgcgc tgcacgactt ctaccctttc acgcgtccgc gcgccggcgg gggcaggcgc 660
 ggccgtgtccg cggccggaac cgacgccagg gcggcagctg cgttggtggc gccggtaaag 720
 ccggatgctt tcgtcgttcc cagccctttc gccggcgtcg agggggacgg tgaatacccg 780
 atgatgccgc agggcctgct cgaggacgtg atccactccc cggcgttcgt ggaggttggtg 840
 gccgcgccgc cgtccgtccc cacgcgccgc ggccgccggg gctga 885

<210> 231

<211> 2130

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea mays PLT3

<400> 231

atggccactg tgaacaactg gctcgtttc tccctctccc cgcaggagct gccgccctcc 60
 cagacgacgg actccacgct catctcggcc gccaccgccg accatgtctc cggcgatgtc 120
 tgcttcaaca tcccccaaga ttggagcatg aggggatcag agctttcggc gctcgtcgcg 180
 gagccgaagc tggaggactt cctcggcggc atctccttct ccgagcagca tcacaagtcc 240
 aactgcaact tgatacccag cactagcagc acagtttgtc acgcgagctc agctgctagc 300
 accggctacc atcaccagct gtaccagccc accagctccg cgctccactt cgcggactcc 360
 gtcatggttg cctcctcggc cgggtgtccac gacggcgggt ccatgctcag cgcggccgcc 420
 gctaacggtg tcgtcggcgc tgccagtgcc aacggcggcg gcacgcgggt gtccatgac 480
 aagaactggc tcgagagcca accggcggcc atgcagccga gggcggcggc ggctgagggc 540
 gcgcaggggc tctctttgtc catgaacatg gcggggacga cccaaggcgc tgctggcatg 600
 ccacttctcg ctggagagcg cgcacgggcg cccgagagtg tatcgacgtc agcacagggt 660
 ggtgccgtcg tcgtcacggc gccgaaggag gatagcgggt gcagcgggtg tgccgggtgt 720
 ctagtagccg tgagcacgga cacgggtggc agcggcggcg cgtcggctga caacacggca 780
 aggaagacgg tggacacgtt cgggcagcgc acgtcgattt accgtggcgt gacaaggcat 840
 agatggactg ggagatatga ggcacatctt tgggataaca gttgcagaag ggaaggacaa 900
 actcgtaagg gtcgtcaagt ctatttaggt ggctatgata aagaggagaa agctgctagg 960
 gcttatgacg ttgctgctct gaagtactgg ggtgccacaa caacaacaaa ttttcagtg 1020
 agtaactacg aaaaggagct cgaggacatg aagcacatga caaggcagga gtttgtagcg 1080
 tctctgagaa ggaagagcag tggtttctcc agaggtgcat ccatttacag gggagtgact 1140
 aggcatcacc aacatggaag atggcaagca cggattggac gagttgcagg gaacaaggat 1200
 ctttacttgg gcaccttcag caccaggag gaggcagcgg aggcgtacga catcgcgcg 1260
 atcaagttcc gcggcctcaa cgccgtcacc aacttcgaca tgagccgcta cgacgtgaag 1320
 agcatcctgg acagcagcgc cctccccatc ggcagcgccg ccaagcgtct caaggaggcc 1380
 gaggccgcag cgtccgcgca gcaccaccac gccggcgtgg tgagctacga cgtcggccgc 1440
 atgcctcgc agctcggcga cggcggagcc ctagcggcgg cgtacggcgc gactaccac 1500
 ggccgccgct ggccgacat cgcgttcag ccgggcgccg ccaccacagg cctgtaccac 1560

ccgtacgcgc agcagccaat gcgcggcggc ggggtggtgca agcaggagca ggaccacgcg 1620
 gtgatcgcg cgcgcacag cctgcaggac ctccaccact tgaacctggg cgcggccggc 1680
 gcgcacgact ttttctcggc agggcagcag gccgccggcg cagctgcgat gcacggcctg 1740
 gctagcatcg acagtgcgtc gctcgagcac agcaccggct ccaactccgt cgtctacaac 1800
 ggcggggtcg gcgatagcaa cggcgccagc gccgttggca gcggcggtgg ctacatgatg 1860
 ccgatgagcg ctgccggagc aaccactaca tcggcaatgg tgagccacga gcagatgcat 1920
 gcacgggcct acgacgaagc caagcaggct gctcagatgg ggtacgagag ctacctggtg 1980
 aacgcggaga acaatggtgg cggaaggatg tctgcatggg ggaccgtcgt ctctgcagcc 2040
 gcggcggcag cagcaagcag caacgacaac attgccggcg acgtcggcca tggcggcgcg 2100
 cagctcttca gtgtctggaa cgacacttaa 2130

<210> 232

<211> 1479

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea mays PLT5

<400> 232

atggacacct cgcaccacta tcatccatgg ctcaacttct ccctcgccca ccaactgtgac 60
 ctcgaggagg aggagagggg cgcggccggc gagctggccg cgatagccgg cgccgcggcg 120
 ccgccgaagc tggaggactt cctcggcgga ggctcgcca ccggtggtcc ggaggcgggtg 180
 gcgcccgcgg agatgtacga ctcggaacct aagttcatag ccgcccggcg gttccttggc 240
 ggctcggcgg cgcgggcggc gacgtcgccg ctgtcctccc tcgaccaggc cggttccaag 300
 ctggccttgc ctgcggcggc ggctgctccg gcgcccggagc agaggaaggc cgtcgactcc 360
 tttgggcagc gcacgtccat ctaccgcggc gtcacacggc accggtggac tggcaggtac 420
 gaggcacatc tgtgggacaa cagctgccga cggaagggc agagccgcaa gggccgcaa 480
 gtatatattg gtggctatga taaggaggag aaggctgcca gggcgatga tcttgcagct 540
 ttgaagtact ggggttctag caccaccacc aactttccgg ttgctgagta tgagaaggag 600
 gtcgaggaga tgaagaacat gacgcgacaa gagtttgttg cttcccttcg aaggaagagc 660
 agtggattct ctcggggtgc ttccatctac cgaggtgtaa ccagacatca ccagcatgga 720
 cggtggcagg cgaggatcgg aagggtggcc ggtaacaagg acctctacct tgggacgttc 780
 agcaccgagg aggaagctgc agaggcctac gacatagcgg ccatcaagtt cagaggcctg 840
 aacccgtca caaacttcga gatcagccgg tacaacgtgg agaccataat gagcagcaac 900
 cttccagtcg cgagcatgtc gtcgtcggcg gcggcgccgg cgggtggccg gagcagcaag 960
 gcgctggagt cccctccgtc cggtcgtt gacggcgcg gcggcatgcc agtcgtcgaa 1020
 gccagcacgg caccgccgt gttcattccg gtgaagtacg accagcagca gcaggagtac 1080
 ctgtcgatgc tcgcttgca gcagcaccac cagcagcaac aagcaggga cctgttgag 1140
 gggccgctag tagggttcgg cggcctctac tcctccgggg tgaacctgga tttcgccaac 1200
 tcccacggca cggcggctcc gtcgtcgatg gccaccact gctacgcaa tggcaccgcc 1260
 tccgcctcgc atgagcacca gcaccagatg cagcaggcg gcgagaacga gacgcagccg 1320

cagccgcagc agagctccag cagctgctcc tccctgccat tcgccacccc ggctcgctttc 1380
aatgggtcct atgaaagctc catcacggcg gcaggcccct ttggatactc ctacccaaat 1440
gtggcagcct ttcagacgcc gatctatgga atggaatga 1479

<210> 233

<211> 1467

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea mays PLT7

<400> 233

atggacatgg acatgagctc agcttateccc caccattggc tctccttctc cctctccaac 60
aactaccacc atggcctact cgaagccttc tctaactcct ccggtactcc tcttggagac 120
gagcagggcg cagtggagga gtccccgagg acggtggagg acttctctcg cggcgtcggt 180
ggcgccggcg ccccgccgca gccggcgggcg gctgcagatc aggatcacca gcttgtgtgc 240
ggcgagctgg gcagcatcac agccaggttc ttgcgccact acccggcggc gccagctggg 300
acgacggtgg agaacccccg cgcggtgacc gtggcgggcca tgtcgtcgac ggacgtggcc 360
ggggcgaggat ccgaccaggc gaggcggccc gccgagacgt tcggccagcg cacatccatc 420
taccgtggcg tcaccaggca ccggtggacg gggagatatg aggcgcacct gtgggacaac 480
agctgccgcc gggaggggcca aagccgcaaa ggacggcaag tctacctagg aggctatgac 540
aaggaggaga aggcggctag agcttacgac ctcccgcgcg tcaagtactg ggggcctaca 600
accacgacca acttcccgtt gtccaactac gagaaggagc tggaggagat gaagtccatg 660
acgcggcagg agttcatcgc gtcgttgcg aggaagagca gcggcttctc acgaggcgcc 720
tccatctaca gaggagtcac aaggcatcat cagcacggcc ggtggcaggc gaggatcggc 780
agggtggccg gaaacaagga cctgtacttg ggcactttca gtactcagga agaggcggcg 840
gaggcgtagc acatcgctgc gatcaagttc cgcgggctca acgccgtcac caactttgac 900
atgagccgct acgacgtgga gagcatcctc agcagcgacc tccccgtcgg gggcgagact 960
agcggtcgcg ccccgcccaa gttcccgttg gactcgctgc agccggggag cgctgccgcc 1020
atgatgctcg ccggggctgc tgccgcttcg caggccacca tgccgccgtc cgagaaggac 1080
tactggctctc tgctcgccct gcactaccag cagcagcagg agcaggagcg gcagttcccg 1140
gcttctgctt acgaggctta cggtccggc ggcgtgaacg tggacttcac gatgggcacc 1200
agtagcggca acaacaacaa caacaccggc agcggcgctc tgtggggcgcg caccactggt 1260
gcagtagtag tgggacagca agacagcagc ggcaagcagg gcaacggcta tgccagcaac 1320
attccttatg ctgctgctgc tatggtttct ggatctgctg gctacgaggg ctccaccggc 1380
gacaatggaa cctgggttac tacgactacc agcagcaaca ccggcacggc tccccactac 1440
tacaactatc tcttcgggat ggagtag 1467

<210> 234

<211> 1413

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea mays IPT

<400> 234

```
atggcccacc cctccgccgc cgccgccgcc gtatcctcca cggcgcccgc tgcaaaccct 60
agttctggcg cccgcgagga aggaggcgcc cgctctccgc cgtcgccgctc tccgtctcag 120
agggggcggg ccaaggtggt gatcgttatg ggcgccacgg gcgcccggcaa gtcgcggtg 180
gccgtcgacc tcgcggccca cttcgccggc gtcgaagtgg tcagcgccga ctccatgcag 240
ctctaccgcg gcctcgacgt cctcaccaac aaggtcccc tccacgagca gaacggtgtt 300
cctcatcatc tacttagcgt gattgatecc tctgtcgagt tcacttgccg tgatttccgc 360
gaccgtgccg tgccgattat acaggaaata gtggaccgcg gtggcctccc tgtggtgtgc 420
ggcggcacaa acttctacat ccaggtctc gttagcccat tcctcttgga tgatatggca 480
gaagaaatgc agggctgtac tctgagagat cacatagatg atggtcttac tgatgaagat 540
gaaggcaatg ggtttgaacg cttgaaggag atcgatcctg tggctgcgca gaggatccat 600
ccaaacgacc atagaaaaat caaacgtac ctcgagttgt atgcaaccac gggtgcccta 660
cccagcgatc tgttccaagg agaggccgct aagaaatggg gtcggcctag taactccaga 720
ctcgactgct gtttcctgtg ggtagatgct gatcttcaag tcctggacag ttatgtcaac 780
aaaagggctg attgcatgat ggatggtggc ctgctggacg aagtatgcag catatatgat 840
gcggatgctg tctataccca ggggctgcgg caggctattg gggttcgtga gtttgacgag 900
tttttcagag catatttacc cagaaaagaa tctggtgagg gttcctgtgc aagcctgtta 960
ggtatgcatg acgatcagct taagagcttg ttggacgaag ctgtttccca gctgaaggca 1020
aacactcgta gactagtctg acgtcaaaga cggagattgc atcggctgag taaagatttt 1080
gggtggaact tgcatcgtgt tgacgcaacc gaagcattct tctgtgccac tgacgactca 1140
tggaacaaaga aagttgtcaa accatgtgtg gatgtcgtaa gaaggttttt gtcggacaat 1200
tccactgttt tgccaagcac aagcgcaagt gacccctctt caagagagct gtggacgcaa 1260
tatgtgtgcg aggcctgcgg caaccgggtg ctgcgaggtg cgcacgagtg ggagcagcac 1320
aggcaagggc gaggccaccg gaagcgagtg cagcgctga agcagaagag cctgaggcca 1380
tgccatcgct tgctgcccc aagaccgcagc tga 1413
```

<210> 235

<211> 1080

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea mays Knotted1

<400> 235

```
atggaggaga tcaccaaca ctttggagtt ggcgcaagca gccacggcca tggccacggc 60
cagcaccacc atcatcacca ccaccaccac ccgtgggcat cctccctcag cgccgtcgta 120
gcgccgctgc cgccgcaacc gccaaagcga ggctgcccgc tgaccctgaa cacggtggcg 180
gccactggga acagcggcgg tagcggcaac ccggtgctgc agcttgccaa cgggtggcggc 240
ctcctcgacg catgcgtcaa ggcaaggag ccctcgtcgt cgtctcccta cgcaggcgac 300
```

gtcgaggcca tcaaggccaa gatcatctcg caccacact actactcgct cctcactgcc 360
 tacctcgagt gcaacaaggt gggggcacca ccggaggtgt cggcgaggct gacggagata 420
 gcgcaggagg tggaggcgcg gcagcgcacg gcgctcggcg gcctggccgc tgcgacggag 480
 ccggagctgg accagttcat ggaggcgtac cacgagatgc tggatgaagtt cagggaggag 540
 ctgacgaggc cgctgcagga ggcgatggag ttcatgcgaa gggatggagtc gcagctgaac 600
 tcgctttcca tctccggaag gtcgctgcgc aacatccttt catctggctc ttctgaggag 660
 gatcaagaag gtagcggagg agagaccgag ctccctgaag ttgatgcaca tggatggac 720
 caagagctga agcaccatct cctgaagaaa tacagtggct atctaagctc gctcaagcaa 780
 gaactgtcaa agaagaagaa gaaagggaag ctccccaagg aggctcgcca gcagctcctt 840
 agctgggtggg atcagcacta caaatggcct taccctcag agactcagaa ggtggcactg 900
 gctgagtcta ccgggcttga cctgaagcag atcaacaact ggttcatcaa ccagcggaag 960
 cggcactgga agccatccga ggagatgcac cacctgatga tggacgggta ccacaccacc 1020
 aatgccttct acatggacgg ccacttcate aacgacggcg ggctgtaccg gctcggctag 1080

<210> 236

<211> 936

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea mays RKD4_1

<400> 236

atgacgggcc tcgacgaggc gctcatgctg ccgttcaccg acatcgatct tgaggccttc 60
 gacaacgccg aagagcaaaa gcctcctgtc gaccaaattg ttatgatgcc gccgacgggt 120
 gaacaccccc ccgccgccgg gacgcgagcc ccaatcatca ttgatggtag ggcgaccgtt 180
 ggccaaaatg taggtgggtg tgtcgtccac gctcatcaga aggcggccat gacgaccata 240
 gaggactcca gctgcttccg acgaggagcc agctgtgtcg acgacgacat ggccgtcgtc 300
 attcaccatg tcgagcgtcg tcgtcaagca ggctctaccg ccgtggcgct attgccgccg 360
 ccgcagccgt cactgccgcg gccgcgtgca agggcgagcg gcggcgcggg cgagcggta 420
 gctccggcgg ccgccgggaa gacgaggatg gaccacatcg gcttcgacga gctgcgcaag 480
 tactttctaca tgcccatcac caggcgggcc agggagatga acgtggggct caccgtgctc 540
 aagaagcgct gccgcgagct cggcgtggcg cggatggcctc accggaagat gaagagcctc 600
 aagtccctca tggccaacgt acaggaaatg gggaacggca tgtcgccggg ggctgtgcag 660
 catgagcttg cggcgctgga gacgtactgc gcgctcatgg aggagaacc atggatcgag 720
 ctacacggacc ggacgaagag gctgcggcag gcctgcttca aggagagcta caagcggagg 780
 aaggcggccg caggcaacgc tatcgagacg gatcacattg tctacagctt tggacagcat 840
 cgtcggtaca agcagcagct gctgcctccg ccaactgcgg gtagtaccag tgctgacgac 900
 cgccatggcc agagcagccg tttcttttgc tactga 936

<210> 237

<211> 1176

<212> DNA

<213> 人工序列

<220>

<223> cDNA of Zea mays RKD4_2

<400> 237

```

atggcgatgg tgccatgtgg cggtagacgac gcggaatggt gcaatatgat ggaggccatc 60
aaccacctga tgatgtcttc catgtcctcg ccgcacgtcg ccatggggcg cagcagttgc 120
agggaagagg acgacgacag tttgtacttg cccatgtact actcatctgc gccaccgcca 180
gccgtcgtca gcgatcagta ctgccccgaa caactcccac cgctgcctgc tgccggtgca 240
atgacggggc tcgacgaggc gctcatgctg ccgttcaccg acatcgatct tgaggccttc 300
gacaacgccg aagagcaaaa gcctcctgtc gaccaaatgg ttatgatgcc gccgacggtt 360
gaacaccccc cgcgcgccgg gacgcgagcc ccaatcatca ttgatggtac ggcgaccgtt 420
ggccaaaatg taggtggtgg tgtcgtccac gctcatcaga aggcggccat gacgaccata 480
gaggactcca gctgcttccg acgaggagcc agctgtgtcg acgacgacat ggccgtcgtc 540
attcaccatg tcgagcgtcg tcgtcaagca ggctctaccg ccgtggcgct attgccgccg 600
ccgcagccgt cactgccgcg gccgcgtgca agggcgagcg gcggcgcggg cgagcggtca 660
gctccggcgg ccgccgggaa gacgaggatg gaccacatcg gcttcgacga gctgcgcaag 720
tacttctaca tgcccatcac cagggcggcc agggagatga acgtggggct caccgtgctc 780
aagaagcgct gccgcgagct cggcgtggcg cgggtggcctc accggaagat gaagagcctc 840
aagtccctca tggccaacgt acaggaaatg gggaacggca tgtcgccggt ggctgtgcag 900
catgagcttg cggcgtgga gacgtactgc gcgctcatgg aggagaacct atggatcgag 960
ctcacggacc ggacgaagag gctgcggcag gcctgcttca aggagagcta caagcggagg 1020
aaggcggccg caggcaacgc tatcgagacg gatcacattg tctacagctt tggacagcat 1080
cgtcgttaca agcagcagct gctgcctccg ccaactgcgg gtagtaccag tgctgacgac 1140
cgccatggcc agagcagccg tttcttttgc tactga 1176

```

<210> 238

<211> 679

<212> PRT

<213> Zea mays

<400> 238

```

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

```

Met	Arg	Gly	Leu	Asp	Tyr	Asn	Gly	Gly	Gly	Ser	Glu	Leu	Ser	Met	Leu	
				85					90					95		
Val	Gly	Ser	Ser	Gly	Gly	Gly	Gly	Gly	Asn	Gly	Lys	Arg	Ala	Val	Glu	
				100					105					110		
Asp	Ser	Glu	Pro	Lys	Leu	Glu	Asp	Phe	Leu	Gly	Gly	Asn	Ser	Phe	Val	
				115					120					125		
Ser	Asp	Gln	Asp	Gln	Ser	Gly	Gly	Tyr	Leu	Phe	Ser	Gly	Val	Pro	Ile	
				130					135					140		
Ala	Ser	Ser	Ala	Asn	Ser	Asn	Ser	Gly	Ser	Asn	Thr	Met	Glu	Leu	Ser	
145					150					155					160	
Met	Ile	Lys	Thr	Trp	Leu	Arg	Asn	Asn	Gln	Val	Ala	Gln	Pro	Gln	Pro	
				165					170					175		
Pro	Ala	Pro	His	Gln	Pro	Gln	Pro	Glu	Glu	Met	Ser	Thr	Asp	Ala	Ser	
				180					185					190		
Gly	Ser	Ser	Phe	Gly	Cys	Ser	Asp	Ser	Met	Gly	Arg	Asn	Ser	Met	Val	
				195					200					205		
Ala	Ala	Gly	Gly	Ser	Ser	Gln	Ser	Leu	Ala	Leu	Ser	Met	Ser	Thr	Gly	
				210					215					220		
Ser	His	Leu	Pro	Met	Val	Val	Pro	Ser	Gly	Ala	Ala	Ser	Gly	Ala	Ala	
225					230					235					240	
Ser	Glu	Ser	Thr	Ser	Ser	Glu	Asn	Lys	Arg	Ala	Ser	Gly	Ala	Met	Asp	
				245					250					255		
Ser	Pro	Gly	Ser	Ala	Val	Glu	Ala	Val	Pro	Arg	Lys	Ser	Ile	Asp	Thr	
				260					265					270		
Phe	Gly	Gln	Arg	Thr	Ser	Ile	Tyr	Arg	Gly	Val	Thr	Arg	His	Arg	Trp	
				275					280					285		
Thr	Gly	Arg	Tyr	Glu	Ala	His	Leu	Trp	Asp	Asn	Ser	Cys	Arg	Arg	Glu	
				290					295					300		
Gly	Gln	Ser	Arg	Lys	Gly	Arg	Gln	Val	Tyr	Leu	Gly	Gly	Tyr	Asp	Lys	
305					310					315					320	
Glu	Asp	Lys	Ala	Ala	Arg	Ala	Tyr	Asp	Leu	Ala	Ala	Leu	Lys	Tyr	Trp	
				325					330					335		
Gly	Thr	Thr	Thr	Thr	Thr	Asn	Phe	Pro	Ile	Ser	Asn	Tyr	Glu	Lys	Glu	
				340					345					350		
Leu	Glu	Glu	Met	Lys	His	Met	Thr	Arg	Gln	Glu	Tyr	Ile	Ala	Tyr	Leu	
				355					360					365		
Arg	Arg	Asn	Ser	Ser	Gly	Phe	Ser	Arg	Gly	Ala	Ser	Lys	Tyr	Arg	Gly	
				370					375					380		
Val	Thr	Arg	His	His	Gln	His	Gly	Arg	Trp	Gln	Ala	Arg	Ile	Gly	Arg	

385	390	395	400
Val Ala Gly Asn Lys Asp Leu Tyr Leu Gly Thr Phe Ser Thr Glu Glu			
	405	410	415
Glu Ala Ala Glu Ala Tyr Asp Ile Ala Ala Ile Lys Phe Arg Gly Leu			
	420	425	430
Asn Ala Val Thr Asn Phe Asp Met Ser Arg Tyr Asp Val Lys Ser Ile			
	435	440	445
Leu Glu Ser Ser Thr Leu Pro Val Gly Gly Ala Ala Arg Arg Leu Lys			
	450	455	460
Asp Ala Val Asp His Val Glu Ala Gly Ala Thr Ile Trp Arg Ala Asp			
465	470	475	480
Met Asp Gly Ala Val Ile Ser Gln Leu Ala Glu Ala Gly Met Gly Gly			
	485	490	495
Tyr Ala Ser Tyr Gly His His Gly Trp Pro Thr Ile Ala Phe Gln Gln			
	500	505	510
Pro Ser Pro Leu Ser Val His Tyr Pro Tyr Gly Gln Pro Ser Arg Gly			
	515	520	525
Trp Cys Lys Pro Glu Gln Asp Ala Ala Ala Ala Ala Ala His Ser Leu			
	530	535	540
Gln Asp Leu Gln Gln Leu His Leu Gly Ser Ala Ala His Asn Phe Phe			
545	550	555	560
Gln Ala Ser Ser Ser Ser Thr Val Tyr Asn Gly Gly Ala Gly Ala Ser			
	565	570	575
Gly Gly Tyr Gln Gly Leu Gly Gly Gly Ser Ser Phe Leu Met Pro Ser			
	580	585	590
Ser Thr Val Val Ala Ala Ala Asp Gln Gly His Ser Ser Thr Ala Asn			
	595	600	605
Gln Gly Ser Thr Cys Ser Tyr Gly Asp Asp His Gln Glu Gly Lys Leu			
	610	615	620
Ile Gly Tyr Asp Ala Ala Met Val Ala Thr Ala Ala Gly Gly Asp Pro			
625	630	635	640
Tyr Ala Ala Ala Arg Asn Gly Tyr Gln Phe Ser Gln Gly Ser Gly Ser			
	645	650	655
Thr Val Ser Ile Ala Arg Ala Asn Gly Tyr Ala Asn Asn Trp Ser Ser			
	660	665	670
Pro Phe Asn Asn Gly Met Gly			
	675		
<210> 239			
<211> 320			

<212> PRT

<213> Zea mays

<400> 239

```

Met Ala Ala Asn Val Gly Ala Gly Arg Ser Ala Gly Gly Gly Gly Ala
1           5           10           15
Gly Thr Gly Thr Gly Thr Ala Ala Gly Ser Gly Gly Val Ser Thr Ala
           20           25           30
Val Cys Arg Pro Ser Gly Ser Arg Trp Thr Pro Thr Pro Glu Gln Ile
           35           40           45
Arg Ile Leu Lys Glu Leu Tyr Tyr Gly Cys Gly Ile Arg Ser Pro Asn
           50           55           60
Ser Glu Gln Ile Gln Arg Ile Thr Ala Met Leu Arg Gln His Gly Lys
65           70           75           80
Ile Glu Gly Lys Asn Val Phe Tyr Trp Phe Gln Asn His Lys Ala Arg
           85           90           95
Glu Arg Gln Lys Arg Arg Leu Thr Asn Leu Asp Val Asn Val Pro Val
           100          105          110
Ala Ala Asp Asp Ser Ala His Arg Leu Gly Val Leu Ser Leu Ser Pro
           115          120          125
Ser Ser Gly Cys Ser Gly Ala Ala Pro Pro Ser Pro Thr Leu Gly Phe
           130          135          140
Tyr Ala Gly Gly Asn Gly Ser Ala Val Met Leu Asp Thr Ser Ser Asp
145          150          155          160
Trp Gly Ser Ala Ala Ala Met Ala Thr Glu Ala Cys Phe Met Gln Asp
           165          170          175
Tyr Met Gly Val Met Gly Gly Ala Ser Pro Trp Ala Cys Ser Ser Ser
           180          185          190
Ser Ser Glu Asp Pro Met Ala Ala Leu Ala Leu Ala Pro Lys Val Thr
           195          200          205
Arg Ala Pro Glu Thr Leu Pro Leu Phe Pro Thr Gly Gly Gly Gly Asp
           210          215          220
Asp Arg Gln Pro Pro Arg Pro Arg Gln Ser Val Pro Ala Gly Glu Ala
225          230          235          240
Ile Arg Gly Gly Ser Ser Ser Ser Ser Tyr Leu Pro Phe Trp Gly Ala
           245          250          255
Ala Pro Thr Pro Thr Gly Ser Ala Thr Ser Val Ala Ile Gln Gln Gln
           260          265          270
His Gln Leu Met Gln Met Gln Glu Gln Tyr Ser Phe Tyr Ser Asn Ala
           275          280          285

```

Gln Leu Leu Pro Gly Thr Gly Ser Gln Asp Ala Ala Ala Thr Ser Leu
 290 295 300
 Glu Leu Ser Leu Ser Ser Trp Cys Ser Pro Tyr Pro Ala Gly Thr Met
 305 310 315 320
 <210> 240
 <211> 325
 <212> PRT
 <213> Zea mays
 <400> 240
 Met Ala Ala Asn Ala Gly Gly Gly Gly Ala Gly Gly Gly Ser Gly Ser
 1 5 10 15
 Gly Ser Val Ala Ala Pro Ala Val Cys Arg Pro Ser Gly Ser Arg Trp
 20 25 30
 Thr Pro Thr Pro Glu Gln Ile Arg Met Leu Lys Glu Leu Tyr Tyr Gly
 35 40 45
 Cys Gly Ile Arg Ser Pro Ser Ser Glu Gln Ile Gln Arg Ile Thr Ala
 50 55 60
 Met Leu Arg Gln His Gly Lys Ile Glu Gly Lys Asn Val Phe Tyr Trp
 65 70 75 80
 Phe Gln Asn His Lys Ala Arg Glu Arg Gln Lys Arg Arg Leu Thr Ser
 85 90 95
 Leu Asp Val Asn Val Pro Ala Ala Gly Ala Ala Asp Ala Thr Thr Ser
 100 105 110
 Gln Leu Gly Val Leu Ser Leu Ser Ser Pro Pro Ser Gly Ala Ala Pro
 115 120 125
 Pro Ser Pro Thr Leu Gly Phe Tyr Ala Ala Gly Asn Gly Gly Gly Ser
 130 135 140
 Ala Gly Leu Leu Asp Thr Ser Ser Asp Trp Gly Ser Ser Gly Ala Ala
 145 150 155 160
 Met Ala Thr Glu Thr Cys Phe Leu Gln Asp Tyr Met Gly Val Thr Asp
 165 170 175
 Thr Gly Ser Ser Ser Gln Trp Pro Cys Phe Ser Ser Ser Asp Thr Ile
 180 185 190
 Met Ala Ala Ala Ala Ala Ala Ala Arg Val Ala Thr Thr Arg Ala Pro
 195 200 205
 Glu Thr Leu Pro Leu Phe Pro Thr Cys Gly Asp Asp Asp Asp Asp Asp
 210 215 220
 Ser Gln Pro Pro Pro Arg Pro Arg His Ala Val Pro Val Pro Ala Gly
 225 230 235 240

Glu Thr Ile Arg Gly Gly Gly Gly Ser Ser Ser Ser Tyr Leu Pro Phe
 245 250 255
 Trp Gly Ala Gly Ala Ala Ser Thr Thr Ala Gly Ala Thr Ser Ser Val
 260 265 270
 Ala Ile Gln Gln Gln His Gln Leu Gln Glu Gln Tyr Ser Phe Tyr Ser
 275 280 285
 Asn Ser Thr Gln Leu Ala Gly Thr Gly Ser Gln Asp Val Ser Ala Ser
 290 295 300
 Ala Ala Ala Leu Glu Leu Ser Leu Ser Ser Trp Cys Ser Pro Tyr Pro
 305 310 315 320
 Ala Ala Gly Ser Met
 325

<210> 241

<211> 324

<212> PRT

<213> Zea mays

<400> 241

Met Glu Thr Pro Gln Gln Gln Ser Ala Ala Ala Ala Ala Ala Ala
 1 5 10 15
 His Gly Gln Asp Asp Gly Gly Ser Pro Pro Met Ser Pro Ala Ser Ala
 20 25 30
 Ala Ala Ala Ala Leu Ala Asn Ala Arg Trp Asn Pro Thr Lys Glu Gln
 35 40 45
 Val Ala Val Leu Glu Gly Leu Tyr Glu His Gly Leu Arg Thr Pro Ser
 50 55 60
 Ala Glu Gln Ile Gln Gln Ile Thr Gly Arg Leu Arg Glu His Gly Ala
 65 70 75 80
 Ile Glu Gly Lys Asn Val Phe Tyr Trp Phe Gln Asn His Lys Ala Arg
 85 90 95
 Gln Arg Gln Arg Gln Lys Gln Asp Ser Phe Ala Tyr Phe Ser Arg Leu
 100 105 110
 Leu Arg Arg Pro Pro Pro Leu Pro Val Leu Ser Met Pro Pro Ala Pro
 115 120 125
 Pro Tyr His His Ala Arg Val Pro Ala Pro Pro Ala Ile Pro Met Pro
 130 135 140
 Met Ala Pro Pro Pro Pro Ala Ala Cys Asn Asp Asn Gly Gly Ala Arg
 145 150 155 160
 Val Ile Tyr Arg Asn Pro Phe Tyr Val Ala Ala Pro Gln Ala Pro Pro
 165 170 175

Ala Asn Ala Ala Tyr Tyr Tyr Pro Gln Pro Gln Gln Gln Gln Gln		
180	185	190
Gln Val Thr Val Met Tyr Gln Tyr Pro Arg Met Glu Val Ala Gly Gln		
195	200	205
Asp Lys Met Met Thr Arg Ala Ala Ala His Gln Gln Gln Gln His Asn		
210	215	220
Gly Ala Gly Gln Gln Pro Gly Arg Ala Gly His Pro Ser Arg Glu Thr		
225	230	235
Leu Gln Leu Phe Pro Leu Gln Pro Thr Phe Val Leu Arg His Asp Lys		
245	250	255
Gly Arg Ala Ala Asn Gly Ser Asn Asn Asp Ser Leu Thr Ser Thr Ser		
260	265	270
Thr Ala Thr Ala Thr Ala Thr Ala Thr Ala Thr Ala Ser Ala Ser Ile		
275	280	285
Ser Glu Asp Ser Asp Gly Leu Glu Ser Gly Ser Ser Gly Lys Gly Val		
290	295	300
Glu Glu Ala Pro Ala Leu Pro Phe Tyr Asp Phe Phe Gly Leu Gln Ser		
305	310	315
Ser Gly Gly Arg		
<210> 242		
<211> 221		
<212> PRT		
<213> Zea mays		
<400> 242		
Met Glu Ala Leu Ser Gly Arg Val Gly Val Lys Cys Gly Arg Trp Asn		
1	5	10
Pro Thr Ala Glu Gln Val Lys Val Leu Thr Glu Leu Phe Arg Ala Gly		
20	25	30
Leu Arg Thr Pro Ser Thr Glu Gln Ile Gln Arg Ile Ser Thr His Leu		
35	40	45
Ser Ala Phe Gly Lys Val Glu Ser Lys Asn Val Phe Tyr Trp Phe Gln		
50	55	60
Asn His Lys Ala Arg Glu Arg His His His Lys Lys Arg Arg Arg Gly		
65	70	75
Ala Ser Ser Ser Ser Pro Asp Ser Gly Ser Gly Arg Gly Ser Asn Asn		
85	90	95
Glu Glu Asp Gly Arg Gly Ala Ala Ser Gln Ser His Asp Ala Asp Ala		
100	105	110
Asp Ala Asp Leu Val Leu Gln Pro Pro Glu Ser Lys Arg Glu Ala Arg		

115	120	125
Ser Tyr Gly His His His Arg Leu Val Thr Cys Tyr Val Arg Asp Val		
130	135	140
Val Glu Gln Gln Glu Ala Ser Pro Ser Trp Glu Arg Pro Thr Arg Glu		
145	150	155
Val Glu Thr Leu Glu Leu Phe Pro Leu Lys Ser Tyr Gly Asp Leu Glu		
165	170	175
Ala Ala Glu Lys Val Arg Ser Tyr Val Arg Gly Ser Gly Ala Thr Ser		
180	185	190
Glu Gln Cys Arg Glu Leu Ser Phe Phe Asp Val Val Ser Ala Gly Arg		
195	200	205
Asp Pro Pro Leu Glu Leu Arg Leu Cys Ser Phe Gly Pro		
210	215	220
<210> 243		
<211> 506		
<212> PRT		
<213> Zea mays		
<400> 243		
Met Ala Ser Ser Asn Arg His Trp Pro Ser Met Tyr Arg Ser Ser Leu		
1	5	10
Ala Cys Asn Phe Gln Gln Pro Gln Pro Gln Pro Asp Met Asn Asn Gly		
20	25	30
Gly Lys Ser Ser Leu Met Ser Ser Arg Cys Glu Glu Asn Gly Gly Arg		
35	40	45
Asn Pro Glu Pro Arg Pro Arg Trp Asn Pro Arg Pro Glu Gln Ile Arg		
50	55	60
Ile Leu Glu Gly Ile Phe Asn Ser Gly Met Val Asn Pro Pro Arg Asp		
65	70	75
Glu Ile Arg Arg Ile Arg Leu Gln Leu Gln Glu Tyr Gly Pro Val Gly		
85	90	95
Asp Ala Asn Val Phe Tyr Trp Phe Gln Asn Arg Lys Ser Arg Thr Lys		
100	105	110
His Lys Leu Arg Ala Ala Gly Gln Leu Gln Pro Ser Gly Ser Gly Arg		
115	120	125
Ser Ala Leu Gln Ala Arg Ala Cys Ala Pro Ala Pro Val Thr Pro Pro		
130	135	140
Arg Asn Leu Gln Leu Ala Ala Ala Ala Pro Val Ala Pro Pro Thr Ser		
145	150	155
Ser Ser Ser Ser Ser Ser Asp Arg Ser Ser Gly Ser Ser Ser Ser Lys		

				165						170					175
Ser	Val	Thr	Val	Thr	Pro	Thr	Thr	Ala	Val	Ala	Leu	Ala	Ser	Pro	Ala
				180						185					190
Gly	Ala	Ala	Pro	Ala	Ala	Val	Phe	Arg	Gln	Gln	Gly	Val	Met	Pro	Thr
				195						200					205
Thr	Ala	Met	Asp	Leu	Leu	Thr	Pro	Leu	Pro	Ser	Ser	Ser	Ala	Ala	Leu
				210						215					220
Ala	Ala	Arg	Gln	Leu	Tyr	Tyr	Gln	Tyr	His	Ser	Gln	Ile	Met	Ala	Pro
225															240
Ala	Ala	Pro	Pro	Met	Pro	Asp	Thr	Val	Ile	Ala	Ser	Pro	Glu	Gln	Phe
				245						250					255
Leu	Pro	Gln	Trp	Gln	Gln	Gly	Gly	Gln	Gln	His	Tyr	Tyr	Leu	Pro	Ala
				260						265					270
Thr	Glu	Leu	Gly	Gly	Val	Leu	Asp	Gly	His	Ser	His	His	Thr	His	Glu
				275						280					285
Pro	Pro	Ala	Ala	Ile	His	Arg	Pro	Val	Ser	Leu	Ser	Pro	Ser	Val	Leu
				290											300
Phe	Gly	Leu	Cys	Asn	Glu	Ala	Leu	Arg	Gln	Asp	Tyr	Cys	Ala	Asp	Ile
305															320
Ser	Val	Val	Pro	Thr	Lys	Gly	Leu	Gly	His	Gly	His	Gln	Phe	Trp	Asn
				325						330					335
Ser	Thr	Thr	Cys	Gly	Ser	Asp	Met	Gly	Asn	Ser	Asn	Ser	Lys	Ile	Asp
				340						345					350
Ala	Val	Ser	Ala	Val	Ile	Arg	Asp	Asp	Glu	Lys	Ser	Arg	Leu	Gly	Leu
				355											365
Leu	His	Tyr	Tyr	Gly	Leu	Ala	Gly	Ala	Thr	Thr	Thr	Ala	Ala	Ala	Ala
				370						375					380
Val	Ala	Pro	Ala	Pro	Leu	Ala	Ala	Asp	Ala	Ala	Ala	Gly	Thr	Ala	Thr
385															400
Leu	Leu	Pro	Ser	Ser	Ala	Ala	Ser	Asp	Gln	Leu	Gln	Gly	Leu	Leu	Asp
				405						410					415
Ala	Ala	Gly	Leu	Leu	Met	Gly	Glu	Thr	Pro	Pro	Thr	Pro	Thr	Ala	Thr
				420						425					430
Val	Val	Ala	Val	Ala	Arg	Asp	Ala	Val	Thr	Cys	Ala	Ala	Thr	Ala	Thr
				435						440					445
Ala	Gln	Phe	Ser	Val	Pro	Ala	Ser	Met	Arg	Leu	Asp	Val	Arg	Leu	Ala
				450						455					460
Phe	Gly	Glu	Ala	Ala	Leu	Leu	Ala	Arg	His	Thr	Gly	Glu	Ala	Val	Pro
465										470					480

Val Asp Glu Ser Gly Val Thr Val Glu Pro Leu Gln Gln Asp Thr Leu
 485 490 495
 Tyr Tyr Val Leu Met Gln Ala Thr Asn Asn
 500 505
 <210> 244
 <211> 273
 <212> PRT
 <213> Zea mays
 <400> 244
 Met Glu Trp Val Asp Arg Thr Lys Ala Ser Ala Ala Ala Ala Ala Ala
 1 5 10 15
 Ala Ala Asp Glu Arg Ala Gly Gly Ala Glu Gly Leu Ala Gly Tyr Val
 20 25 30
 Lys Val Met Thr Asp Glu Gln Met Glu Val Leu Arg Lys Gln Ile Ser
 35 40 45
 Ile Tyr Ala Thr Ile Cys Glu Gln Leu Val Glu Met His Arg Ala Leu
 50 55 60
 Thr Glu His Gln Asp Thr Ile Ala Gly Ile Arg Phe Ser Asn Leu Tyr
 65 70 75 80
 Cys Asp Pro Gln Ile Ile Pro Gly Gly His Lys Ile Thr Ala Arg Gln
 85 90 95
 Arg Trp Gln Pro Thr Pro Met Gln Leu Gln Ile Leu Glu Asn Ile Phe
 100 105 110
 Asp Gln Gly Asn Gly Thr Pro Ser Lys Gln Arg Ile Lys Glu Ile Thr
 115 120 125
 Ala Glu Leu Ser His His Gly Gln Ile Ser Glu Thr Asn Val Tyr Asn
 130 135 140
 Trp Phe Gln Asn Arg Arg Ala Arg Ser Lys Arg Lys Gln Ala Ala Ser
 145 150 155 160
 Leu Pro Asn Asn Ala Glu Ser Glu Ala Glu Val Asp Glu Glu Ser Leu
 165 170 175
 Thr Asp Lys Lys Pro Lys Ser Asp Arg Ser Leu Gln Asp Asn Lys Ala
 180 185 190
 Met Gly Ala His Asn Ala Asp Arg Ile Ser Gly Met His His Leu Asp
 195 200 205
 Thr Asp His Asp Gln Ile Gly Gly Met Met Tyr Gly Cys Asn Asp Asn
 210 215 220
 Gly Leu Arg Ser Ser Gly Ser Ser Gly Gln Met Ser Phe Tyr Gly Asn
 225 230 235 240

Ile Met Pro Asn Pro Arg Ile Asp His Phe Pro Gly Lys Val Glu Ser
 245 250 255
 Ser Arg Ser Phe Ser His Leu Gln His Gly Glu Gly Phe Asp Met Phe
 260 265 270
 Gly
 <210> 245
 <211> 282
 <212> PRT
 <213> Zea mays
 <400> 245
 Met Asp Trp Gly Asn Arg Thr Lys Ala Ala Ala Ala Ala Ala Pro
 1 5 10 15
 Asp Glu Arg Ala Gly Gly Gly Glu Gly Leu Gly Gly Tyr Val Lys Val
 20 25 30
 Met Thr Asp Glu Gln Met Glu Val Leu Arg Lys Gln Ile Ser Ile Tyr
 35 40 45
 Ala Thr Ile Cys Glu Gln Leu Val Glu Met His Arg Val Leu Thr Glu
 50 55 60
 His Gln Asp Thr Ile Ala Gly Leu Arg Phe Ser Asn Leu Tyr Cys Asp
 65 70 75 80
 Pro Leu Ile Ile Pro Gly Gly His Lys Ile Thr Ala Arg Gln Arg Trp
 85 90 95
 Gln Pro Thr Pro Met Gln Leu Gln Ile Leu Glu Ser Ile Phe Asp Gln
 100 105 110
 Gly Asn Gly Thr Pro Ser Lys Gln Lys Ile Lys Glu Ile Thr Ala Glu
 115 120 125
 Leu Ser Gln His Gly Gln Ile Ser Glu Thr Asn Val Tyr Asn Trp Phe
 130 135 140
 Gln Asn Arg Arg Ala Arg Ser Lys Arg Lys Gln Ala Ala Ala Ser Leu
 145 150 155 160
 Pro Asn Asn Ala Glu Ser Glu Ala Glu Ala Asp Glu Glu Pro Leu Ala
 165 170 175
 Asp Lys Lys Pro Lys Ser Asp Arg Pro Pro Pro Pro Pro Pro Ile
 180 185 190
 Gln Asp Asn Thr Lys Ala Thr Gly Ala Leu Ser Ala Asp Arg Val Ser
 195 200 205
 Gly Gly Thr Arg His Leu Asp Thr Gly His Asp Gln Thr Ser Gly Val
 210 215 220
 Met Tyr Gly Cys Asn Asp Ser Gly Leu Leu Arg Ser Ser Gly Ser Ser

225	230	235	240
Gly Gln Met Ser Leu Tyr Glu Asn Phe Met Ser Asn Pro Arg Ile Asp			
	245	250	255
Arg Phe Pro Ala Lys Val Glu Ser Ser Arg Ser Phe Pro His Leu Gln			
	260	265	270
Gln His Gly Glu Gly Phe Gly Met Phe Gly			
	275	280	
<210> 246			
<211> 264			
<212> PRT			
<213> Zea mays			
<400> 246			
Met Asp Ser Ser Phe Leu Pro Ala Gly Ala Asp Asn Gly Ser Ala Gly			
1	5	10	15
Gly Ala Asn Asn Gly Gly Gly Ala Ala Gln Gln Ala Pro Pro Ile Arg			
	20	25	30
Glu Gln Asp Arg Leu Met Pro Ile Ala Asn Val Ile Arg Ile Met Arg			
	35	40	45
Arg Val Leu Pro Ala His Ala Lys Ile Ser Asp Asp Ala Lys Glu Thr			
	50	55	60
Ile Gln Glu Cys Val Ser Glu Tyr Ile Ser Phe Ile Thr Gly Glu Ala			
65	70	75	80
Asn Glu Arg Cys Gln Arg Glu Gln Arg Lys Thr Ile Thr Ala Glu Asp			
	85	90	95
Val Leu Trp Ala Met Ser Arg Leu Gly Phe Asp Asp Tyr Val Glu Pro			
	100	105	110
Leu Ser Val Tyr Leu His Arg Tyr Arg Glu Phe Glu Gly Glu Ala Arg			
	115	120	125
Gly Val Gly Leu Ala Pro Ala Pro Pro Arg Gly Asp His His His His			
	130	135	140
His His Ser Val Pro Pro Ser Met Leu Asn Lys Ser Arg Gly Pro Gly			
145	150	155	160
Ser Gly Ala Val Met Leu Pro His His His His His Asp Met His Ala			
	165	170	175
Ser Met Tyr Gly Gly Ala Val Pro Pro Pro Pro His His Gly Phe Leu			
	180	185	190
Met Pro His Pro Gln Gly Gly His Tyr Leu Pro Tyr Pro Tyr Glu Pro			
	195	200	205
Thr Ser Tyr Gly Gly Glu His Ala Leu Ala Ser Gly Tyr Tyr Gly Gly			

210	215	220
Ala Ala Tyr Ala Pro Gly Asn Asn Gly Gly Ser Gly Asp Gly Ser Gly		
225	230	235
Gly Ser Ala Ser His Ala Pro Pro Gly Gly Ser Gly Gly Gly Phe Asp		240
	245	250
His Pro His Thr Phe Ala Tyr Lys		255
260		
<210> 247		
<211> 392		
<212> PRT		
<213> Zea mays		
<400> 247		
Met Pro Ala Arg Ala Ser His Pro Ala Leu Ala Thr Ser Arg Ala Arg		
1	5	10
Gly Trp Pro Arg Leu Arg Ala Leu Gly Ile Ala Pro Asp Gly Gly Arg		15
	20	25
Trp Arg Cys Leu Pro His Phe Ala Pro Ile Ser Glu Pro Ala Arg His		30
	35	40
Leu Ser Pro Arg Ala Pro Ala Ser Ala Ser Pro Pro Ala Arg Pro His		45
	50	55
Pro Ala Ile Lys Ala Ser Pro Ser Pro Thr Leu Ala Ala Ala Ala Ala		60
65	70	75
Ala Ala Ala Ala Ala Thr Ser Ser Leu Pro Ser Phe Ser Ala Arg Arg		80
	85	90
Arg Ser Thr Gly Met Ala Gly Ile Thr Lys Arg Arg Thr Ser Pro Ala		95
	100	105
Ser Thr Ser Ser Ser Ser Gly Asp Val Leu Pro Gln Arg Val Thr Arg		110
	115	120
Lys Arg Arg Ser Ala Arg Arg Gly Pro Arg Ser Thr Ala Arg Arg Pro		125
	130	135
Ser Ala Pro Pro Pro Met Asn Glu Leu Asp Leu Asn Thr Ala Ala Leu		140
145	150	155
Asp Pro Asp His Tyr Ala Thr Gly Leu Arg Val Leu Leu Gln Lys Glu		160
	165	170
Leu Arg Asn Ser Asp Val Ser Gln Leu Gly Arg Ile Val Leu Pro Lys		175
	180	185
Lys Glu Ala Glu Ser Tyr Leu Pro Ile Leu Met Ala Lys Asp Gly Lys		190
	195	200
Ser Leu Cys Met His Asp Leu Leu Asn Ser Gln Leu Trp Thr Phe Lys		205

210	215	220
Tyr Arg Tyr Trp Phe Asn Asn Lys Ser Arg Met Tyr Val Leu Glu Asn		
225	230	235
Thr Gly Asp Tyr Val Lys Ala His Asp Leu Gln Gln Gly Asp Phe Ile		240
	245	250
Val Ile Tyr Lys Asp Asp Glu Asn Asn Arg Phe Val Ile Gly Ala Lys		255
	260	265
Lys Ala Gly Asp Glu Gln Thr Ala Thr Val Pro Gln Val His Glu His		270
	275	280
Met His Ile Ser Ala Ala Leu Pro Ala Pro Gln Ala Phe His Asp Tyr		285
	290	300
Ala Gly Pro Val Ala Ala Glu Ala Gly Met Leu Ala Ile Val Pro Gln		
305	310	315
Gly Asp Glu Ile Phe Asp Gly Ile Leu Asn Ser Leu Pro Glu Ile Pro		320
	325	330
Val Ala Asn Val Arg Tyr Ser Asp Phe Phe Asp Pro Phe Gly Asp Ser		335
	340	345
Met Asp Met Ala Asn Pro Leu Ser Ser Ser Asn Asn Pro Ser Val Asn		350
	355	360
Leu Ala Thr His Phe His Asp Glu Arg Ile Gly Ser Cys Ser Phe Pro		365
	370	375
Tyr Pro Lys Ser Gly Pro Gln Met		380
385	390	
<210> 248		
<211> 341		
<212> PRT		
<213> Zea mays		
<400> 248		
Met Ala Ala Ala Ile Asp Met Tyr Lys Tyr Tyr Asn Thr Ser Ala His		
1	5	10
Gln Ile Pro Ser Ser Ser Pro Ser Asp Gln Glu Leu Ala Lys Ala Leu		15
	20	25
Glu Pro Phe Ile Thr Ser Ala Ser Ser Ser Ser Ser Ser Ser Pro Tyr		30
	35	40
His Gly Tyr Ser Ser Ser Pro Ser Met Ser Gln Asp Ser Tyr Met Pro		45
	50	55
Thr Pro Ser Tyr Thr Ser Tyr Ala Thr Ser Pro Leu Pro Thr Pro Ala		60
65	70	75
Ala Ala Ser Ser Ser Gln Leu Pro Pro Leu Tyr Ser Ser Pro Tyr Ala		80

	85		90		95
Ala Pro Cys Met	Ala Gly Gln Met	Gly Leu Asn Gln	Leu Gly Pro	Ala	
100		105		110	
Gln Ile Gln Gln	Ile Gln Ala Gln	Phe Met Phe Gln	Gln Gln Gln	Gln Gln	
115		120		125	
Gln Gln Arg Gly	Leu His Ala Ala	Phe Leu Gly Pro	Arg Ala Gln	Pro	
130		135		140	
Met Lys Gln Ser	Gly Ser Pro Ser	Pro Pro Pro	Pro Leu Ala	Pro Ala	
145		150		155	
Gln Ser Lys Leu	Tyr Arg Gly Val	Arg Gln Arg His	Trp Gly Lys	Trp	
	165		170		175
Val Ala Glu Ile	Arg Leu Pro Lys	Asn Arg Thr Arg	Leu Trp Leu	Gly	
	180		185		190
Thr Phe Asp Thr	Ala Glu Asp Ala	Ala Leu Ala Tyr	Asp Lys Ala	Ala	
195		200		205	
Phe Arg Leu Arg	Gly Asp Thr Ala	Arg Leu Asn Phe	Pro Ala Leu	Arg	
210		215		220	
Arg Gly Gly Ala	His Leu Ala Gly	Pro Leu His Ala	Ser Val Asp	Ala	
225		230		235	
Lys Leu Thr Ala	Ile Cys Gln Ser	Leu Ser Glu Ser	Lys Ser Lys	Ser	
	245		250		255
Gly Ser Ser Gly	Asp Glu Ser Ala	Ala Ser Pro Pro	Asp Ser Pro	Lys	
	260		265		270
Cys Ser Ala Ser	Thr Thr Glu Gly	Glu Gly Glu Glu	Glu Ser Gly	Ser	
	275		280		285
Ala Gly Ser Pro	Pro Pro Pro Pro	Pro Pro Pro	Thr Leu Ala	Pro Pro	
	290		295		300
Val Pro Glu Met	Ala Lys Leu Asp	Phe Thr Glu Ala	Pro Trp Asp	Glu	
305		310		315	
Thr Glu Ala Phe	His Leu Arg Lys	Tyr Pro Ser Trp	Glu Ile Asp	Trp	
	325		330		335
Asp Ser Ile Leu	Ser				
	340				
<210>	249				
<211>	316				
<212>	PRT				
<213>	Zea mays				
<400>	249				
Met Ala Ala Ala	Ile Asp Met Tyr	Lys Tyr Cys Asn	Thr Ser Ala	His	

1	5	10	15
Leu Ile Ala Ser Ser Ser Pro Ser Asp Gln Glu Leu Ala Lys Ala Leu			
20	25	30	
Glu Pro Phe Ile Thr Ser Ala Ser Ser Pro Tyr His Arg Tyr Ser Leu			
35	40	45	
Ala Pro Asp Ser Tyr Met Pro Thr Pro Ser Ser Tyr Thr Thr Ser Pro			
50	55	60	
Leu Pro Thr Pro Thr Ser Ser Pro Phe Ser Gln Leu Pro Pro Leu Tyr			
65	70	75	80
Ser Ser Pro Tyr Ala Ala Ser Thr Ala Ser Gly Val Ala Gly Pro Met			
85	90	95	
Gly Leu Asn Gln Leu Gly Pro Ala Gln Ile Gln Gln Ile Gln Ala Gln			
100	105	110	
Leu Met Phe Gln His Gln Gln Gln Arg Gly Leu His Ala Ala Phe Leu			
115	120	125	
Gly Pro Arg Ala Gln Pro Met Lys Gln Ser Gly Ser Pro Pro Ala Gln			
130	135	140	
Ser Lys Leu Tyr Arg Gly Val Arg Gln Arg His Trp Gly Lys Trp Val			
145	150	155	160
Ala Glu Ile Arg Leu Pro Lys Asn Arg Thr Arg Leu Trp Leu Gly Thr			
165	170	175	
Phe Asp Thr Ala Glu Gly Ala Ala Leu Ala Tyr Asp Glu Ala Ala Phe			
180	185	190	
Arg Leu Arg Gly Asp Thr Ala Arg Leu Asn Phe Pro Ser Leu Arg Arg			
195	200	205	
Gly Gly Gly Ala Arg Leu Ala Gly Pro Leu His Ala Ser Val Asp Ala			
210	215	220	
Lys Leu Thr Ala Ile Cys Gln Ser Leu Ala Gly Ser Lys Asn Ser Ser			
225	230	235	240
Ser Ser Asp Glu Ser Ala Ala Ser Leu Pro Asp Ser Pro Lys Cys Ser			
245	250	255	
Ala Ser Thr Glu Gly Asp Glu Asp Ser Ala Ser Ala Gly Ser Pro Pro			
260	265	270	
Ser Pro Thr Gln Ala Pro Pro Val Pro Glu Met Ala Lys Leu Asp Phe			
275	280	285	
Thr Glu Ala Pro Trp Asp Glu Thr Glu Ala Phe His Leu Arg Lys Tyr			
290	295	300	
Pro Ser Trp Glu Ile Asp Trp Asp Ser Ile Leu Ser			
305	310	315	

<210> 250

<211> 233

<212> PRT

<213> Zea mays

<400> 250

```

Met Ala Pro Arg Thr Ser Glu Lys Thr Met Ala Pro Ala Ala Ala Ala
1           5           10           15
Ala Thr Gly Leu Ala Leu Ser Val Gly Gly Gly Gly Gly Ala Gly Gly
           20           25           30
Pro His Tyr Arg Gly Val Arg Lys Arg Pro Trp Gly Arg Tyr Ala Ala
           35           40           45
Glu Ile Arg Asp Pro Ala Lys Lys Ser Arg Val Trp Leu Gly Thr Tyr
           50           55           60
Asp Thr Ala Glu Asp Ala Ala Arg Ala Tyr Asp Ala Ala Ala Arg Glu
65           70           75           80
Tyr Arg Gly Ala Lys Ala Lys Thr Asn Phe Pro Tyr Pro Ser Cys Val
           85           90           95
Pro Leu Ser Ala Ala Gly Cys Arg Ser Ser Asn Ser Ser Thr Val Glu
           100          105          110
Ser Phe Ser Ser Asp Ala Gln Ala Pro Met Gln Ala Met Pro Leu Pro
           115          120          125
Pro Ser Leu Glu Leu Asp Leu Phe His Arg Ala Ala Ala Ala Ala Thr
           130          135          140
Gly Thr Gly Ala Ala Ala Val Arg Phe Pro Phe Gly Ser Ile Pro Val
145          150          155          160
Thr His Pro Tyr Tyr Phe Phe Gly Gln Ala Ala Ala Ala Ala Glu
           165          170          175
Ala Gly Cys Arg Val Leu Lys Leu Ala Pro Ala Val Thr Val Ala Gln
           180          185          190
Ser Asp Ser Asp Cys Ser Ser Val Val Asp Leu Ser Pro Ser Pro Pro
           195          200          205
Ala Ala Val Ser Ala Arg Lys Pro Ala Ala Phe Asp Leu Asp Leu Asn
           210          215          220
Cys Ser Pro Pro Thr Glu Ala Glu Ala
225           230

```

<210> 251

<211> 294

<212> PRT

<213> Zea mays

<400> 251

Met	Glu	Asp	Val	Ala	Asn	Ala	His	Ile	Tyr	Ala	His	Ala	His	Arg	Ser
1				5					10					15	
Lys	Arg	Pro	Gln	Ser	Ala	Ala	Ile	Lys	Asp	Gly	Asp	Gly	Asp	Val	Asp
			20					25						30	
Leu	Ser	Met	Lys	Gly	Ala	Arg	Tyr	Arg	Gly	Val	Arg	Arg	Arg	Pro	Trp
		35					40						45		
Gly	Arg	Phe	Ala	Ala	Glu	Ile	Arg	Asp	Pro	Met	Ser	Lys	Glu	Arg	Arg
		50				55					60				
Trp	Leu	Gly	Thr	Phe	Asp	Thr	Ala	Glu	Gln	Ala	Ala	Cys	Ala	Tyr	Asp
65					70				75					80	
Ile	Ala	Ala	Arg	Ala	Met	Arg	Gly	Asn	Lys	Ala	Arg	Thr	Asn	Phe	Pro
				85				90						95	
Gly	His	Ala	Thr	Ala	Gly	Tyr	Trp	Pro	Trp	Gly	Ala	Pro	Gln	Pro	Ala
			100					105					110		
Ala	Val	Ala	His	Pro	Ile	Asn	Pro	Phe	Leu	Leu	His	Asn	Leu	Ile	Met
			115				120					125			
Ser	Ser	Ser	Asn	His	Gly	Cys	Arg	Leu	Leu	Asn	His	Ala	Gly	His	Gly
			130				135					140			
His	Val	His	Ser	Ala	Ala	Pro	Arg	Pro	Pro	Ala	Pro	Ala	Ala	Asp	Ala
145					150					155				160	
Thr	Ser	Thr	Thr	Ile	Ala	Ala	Pro	Phe	Pro	Val	Ala	Ala	His	Pro	Ala
				165					170					175	
Val	Ala	Met	Asp	Glu	Asp	Val	Asp	Asp	Trp	Asp	Gly	Val	Leu	Arg	Ser
			180					185					190		
Glu	Pro	Ala	Asp	Ala	Gly	Leu	Leu	Gln	Asp	Ala	Leu	His	Asp	Phe	Tyr
		195					200					205			
Pro	Phe	Thr	Arg	Pro	Arg	Ala	Gly	Gly	Gly	Arg	Arg	Gly	Leu	Ser	Ala
		210				215					220				
Ala	Gly	Thr	Asp	Ala	Arg	Ala	Ala	Ala	Ala	Leu	Val	Ala	Pro	Val	Lys
225					230					235				240	
Pro	Asp	Ala	Phe	Val	Val	Pro	Ser	Pro	Phe	Ala	Gly	Val	Glu	Gly	Asp
				245					250					255	
Gly	Glu	Tyr	Pro	Met	Met	Pro	Gln	Gly	Leu	Leu	Glu	Asp	Val	Ile	His
			260					265					270		
Ser	Pro	Ala	Phe	Val	Glu	Val	Val	Ala	Ala	Pro	Pro	Ser	Val	Pro	Thr
			275					280					285		
Arg	Arg	Gly	Arg	Arg	Gly										
			290												

<210> 252

<211> 709

<212> PRT

<213> Zea mays

<400> 252

```

Met Ala Thr Val Asn Asn Trp Leu Ala Phe Ser Leu Ser Pro Gln Glu
1           5           10           15
Leu Pro Pro Ser Gln Thr Thr Asp Ser Thr Leu Ile Ser Ala Ala Thr
           20           25           30
Ala Asp His Val Ser Gly Asp Val Cys Phe Asn Ile Pro Gln Asp Trp
           35           40           45
Ser Met Arg Gly Ser Glu Leu Ser Ala Leu Val Ala Glu Pro Lys Leu
           50           55           60
Glu Asp Phe Leu Gly Gly Ile Ser Phe Ser Glu Gln His His Lys Ser
65           70           75           80
Asn Cys Asn Leu Ile Pro Ser Thr Ser Ser Thr Val Cys Tyr Ala Ser
           85           90           95
Ser Ala Ala Ser Thr Gly Tyr His His Gln Leu Tyr Gln Pro Thr Ser
           100          105          110
Ser Ala Leu His Phe Ala Asp Ser Val Met Val Ala Ser Ser Ala Gly
           115          120          125
Val His Asp Gly Gly Ser Met Leu Ser Ala Ala Ala Ala Asn Gly Val
           130          135          140
Ala Gly Ala Ala Ser Ala Asn Gly Gly Gly Ile Gly Leu Ser Met Ile
145          150          155          160
Lys Asn Trp Leu Arg Ser Gln Pro Ala Pro Met Gln Pro Arg Ala Ala
           165          170          175
Ala Ala Glu Gly Ala Gln Gly Leu Ser Leu Ser Met Asn Met Ala Gly
           180          185          190
Thr Thr Gln Gly Ala Ala Gly Met Pro Leu Leu Ala Gly Glu Arg Ala
           195          200          205
Arg Ala Pro Glu Ser Val Ser Thr Ser Ala Gln Gly Gly Ala Val Val
           210          215          220
Val Thr Ala Pro Lys Glu Asp Ser Gly Gly Ser Gly Val Ala Gly Ala
225          230          235          240
Leu Val Ala Val Ser Thr Asp Thr Gly Gly Ser Gly Gly Ala Ser Ala
           245          250          255
Asp Asn Thr Ala Arg Lys Thr Val Asp Thr Phe Gly Gln Arg Thr Ser
           260          265          270

```

Ile Tyr Arg Gly Val Thr Arg His Arg Trp Thr Gly Arg Tyr Glu Ala		
275	280	285
His Leu Trp Asp Asn Ser Cys Arg Arg Glu Gly Gln Thr Arg Lys Gly		
290	295	300
Arg Gln Val Tyr Leu Gly Gly Tyr Asp Lys Glu Glu Lys Ala Ala Arg		
305	310	315
Ala Tyr Asp Leu Ala Ala Leu Lys Tyr Trp Gly Ala Thr Thr Thr Thr		
	325	330
Asn Phe Pro Val Ser Asn Tyr Glu Lys Glu Leu Glu Asp Met Lys His		
	340	345
Met Thr Arg Gln Glu Phe Val Ala Ser Leu Arg Arg Lys Ser Ser Gly		
	355	360
Phe Ser Arg Gly Ala Ser Ile Tyr Arg Gly Val Thr Arg His His Gln		
370	375	380
His Gly Arg Trp Gln Ala Arg Ile Gly Arg Val Ala Gly Asn Lys Asp		
385	390	395
Leu Tyr Leu Gly Thr Phe Ser Thr Gln Glu Glu Ala Ala Glu Ala Tyr		
	405	410
Asp Ile Ala Ala Ile Lys Phe Arg Gly Leu Asn Ala Val Thr Asn Phe		
	420	425
Asp Met Ser Arg Tyr Asp Val Lys Ser Ile Leu Asp Ser Ser Ala Leu		
	435	440
Pro Ile Gly Ser Ala Ala Lys Arg Leu Lys Glu Ala Glu Ala Ala Ala		
450	455	460
Ser Ala Gln His His His Ala Gly Val Val Ser Tyr Asp Val Gly Arg		
465	470	475
Ile Ala Ser Gln Leu Gly Asp Gly Gly Ala Leu Ala Ala Ala Tyr Gly		
	485	490
Ala His Tyr His Gly Ala Ala Trp Pro Thr Ile Ala Phe Gln Pro Gly		
	500	505
Ala Ala Thr Thr Gly Leu Tyr His Pro Tyr Ala Gln Gln Pro Met Arg		
	515	520
Gly Gly Gly Trp Cys Lys Gln Glu Gln Asp His Ala Val Ile Ala Ala		
	530	535
Ala His Ser Leu Gln Asp Leu His His Leu Asn Leu Gly Ala Ala Gly		
545	550	555
Ala His Asp Phe Phe Ser Ala Gly Gln Gln Ala Ala Ala Ala Ala		
	565	570
Met His Gly Leu Ala Ser Ile Asp Ser Ala Ser Leu Glu His Ser Thr		

580	585	590
Gly Ser Asn Ser Val Val Tyr Asn Gly Gly Val Gly Asp Ser Asn Gly		
595	600	605
Ala Ser Ala Val Gly Ser Gly Gly Gly Tyr Met Met Pro Met Ser Ala		
610	615	620
Ala Gly Ala Thr Thr Thr Ser Ala Met Val Ser His Glu Gln Met His		
625	630	635
Ala Arg Ala Tyr Asp Glu Ala Lys Gln Ala Ala Gln Met Gly Tyr Glu		
645	650	655
Ser Tyr Leu Val Asn Ala Glu Asn Asn Gly Gly Gly Arg Met Ser Ala		
660	665	670
Trp Gly Thr Val Val Ser Ala Ala Ala Ala Ala Ala Ser Ser Asn		
675	680	685
Asp Asn Ile Ala Ala Asp Val Gly His Gly Gly Ala Gln Leu Phe Ser		
690	695	700
Val Trp Asn Asp Thr		
705		
<210> 253		
<211> 492		
<212> PRT		
<213> Zea mays		
<400> 253		
Met Asp Thr Ser His His Tyr His Pro Trp Leu Asn Phe Ser Leu Ala		
1	5	10
His His Cys Asp Leu Glu Glu Glu Glu Arg Gly Ala Ala Ala Glu Leu		
20	25	30
Ala Ala Ile Ala Gly Ala Ala Pro Pro Pro Lys Leu Glu Asp Phe Leu		
35	40	45
Gly Gly Gly Val Ala Thr Gly Gly Pro Glu Ala Val Ala Pro Ala Glu		
50	55	60
Met Tyr Asp Ser Asp Leu Lys Phe Ile Ala Ala Ala Gly Phe Leu Gly		
65	70	75
Gly Ser Ala Ala Ala Ala Ala Thr Ser Pro Leu Ser Ser Leu Asp Gln		
85	90	95
Ala Gly Ser Lys Leu Ala Leu Pro Ala Ala Ala Ala Ala Pro Ala Pro		
100	105	110
Glu Gln Arg Lys Ala Val Asp Ser Phe Gly Gln Arg Thr Ser Ile Tyr		
115	120	125
Arg Gly Val Thr Arg His Arg Trp Thr Gly Arg Tyr Glu Ala His Leu		

130	135	140
Trp Asp Asn Ser Cys Arg Arg Glu Gly Gln Ser Arg Lys Gly Arg Gln		
145	150	155
Val Tyr Leu Gly Gly Tyr Asp Lys Glu Glu Lys Ala Ala Arg Ala Tyr		
165	170	175
Asp Leu Ala Ala Leu Lys Tyr Trp Gly Ser Ser Thr Thr Thr Asn Phe		
180	185	190
Pro Val Ala Glu Tyr Glu Lys Glu Val Glu Glu Met Lys Asn Met Thr		
195	200	205
Arg Gln Glu Phe Val Ala Ser Leu Arg Arg Lys Ser Ser Gly Phe Ser		
210	215	220
Arg Gly Ala Ser Ile Tyr Arg Gly Val Thr Arg His His Gln His Gly		
225	230	235
Arg Trp Gln Ala Arg Ile Gly Arg Val Ala Gly Asn Lys Asp Leu Tyr		
245	250	255
Leu Gly Thr Phe Ser Thr Glu Glu Glu Ala Ala Glu Ala Tyr Asp Ile		
260	265	270
Ala Ala Ile Lys Phe Arg Gly Leu Asn Ala Val Thr Asn Phe Glu Ile		
275	280	285
Ser Arg Tyr Asn Val Glu Thr Ile Met Ser Ser Asn Leu Pro Val Ala		
290	295	300
Ser Met Ser Ser Ser Ala Ala Ala Ala Ala Gly Gly Arg Ser Ser Lys		
305	310	315
Ala Leu Glu Ser Pro Pro Ser Gly Ser Leu Asp Gly Gly Gly Gly Met		
325	330	335
Pro Val Val Glu Ala Ser Thr Ala Pro Pro Leu Phe Ile Pro Val Lys		
340	345	350
Tyr Asp Gln Gln Gln Gln Glu Tyr Leu Ser Met Leu Ala Leu Gln Gln		
355	360	365
His His Gln Gln Gln Gln Ala Gly Asn Leu Leu Gln Gly Pro Leu Val		
370	375	380
Gly Phe Gly Gly Leu Tyr Ser Ser Gly Val Asn Leu Asp Phe Ala Asn		
385	390	395
Ser His Gly Thr Ala Ala Pro Ser Ser Met Ala His His Cys Tyr Ala		
405	410	415
Asn Gly Thr Ala Ser Ala Ser His Glu His Gln His Gln Met Gln Gln		
420	425	430
Gly Gly Glu Asn Glu Thr Gln Pro Gln Pro Gln Gln Ser Ser Ser Ser		
435	440	445

Cys Ser Ser Leu Pro Phe Ala Thr Pro Val Ala Phe Asn Gly Ser Tyr
 450 455 460
 Glu Ser Ser Ile Thr Ala Ala Gly Pro Phe Gly Tyr Ser Tyr Pro Asn
 465 470 475 480
 Val Ala Ala Phe Gln Thr Pro Ile Tyr Gly Met Glu
 485 490
 <210> 254
 <211> 488
 <212> PRT
 <213> Zea mays
 <400> 254
 Met Asp Met Asp Met Ser Ser Ala Tyr Pro His His Trp Leu Ser Phe
 1 5 10 15
 Ser Leu Ser Asn Asn Tyr His His Gly Leu Leu Glu Ala Phe Ser Asn
 20 25 30
 Ser Ser Gly Thr Pro Leu Gly Asp Glu Gln Gly Ala Val Glu Glu Ser
 35 40 45
 Pro Arg Thr Val Glu Asp Phe Leu Gly Gly Val Gly Gly Ala Gly Ala
 50 55 60
 Pro Pro Gln Pro Ala Ala Ala Asp Gln Asp His Gln Leu Val Cys
 65 70 75 80
 Gly Glu Leu Gly Ser Ile Thr Ala Arg Phe Leu Arg His Tyr Pro Ala
 85 90 95
 Ala Pro Ala Gly Thr Thr Val Glu Asn Pro Gly Ala Val Thr Val Ala
 100 105 110
 Ala Met Ser Ser Thr Asp Val Ala Gly Ala Glu Ser Asp Gln Ala Arg
 115 120 125
 Arg Pro Ala Glu Thr Phe Gly Gln Arg Thr Ser Ile Tyr Arg Gly Val
 130 135 140
 Thr Arg His Arg Trp Thr Gly Arg Tyr Glu Ala His Leu Trp Asp Asn
 145 150 155 160
 Ser Cys Arg Arg Glu Gly Gln Ser Arg Lys Gly Arg Gln Val Tyr Leu
 165 170 175
 Gly Gly Tyr Asp Lys Glu Glu Lys Ala Ala Arg Ala Tyr Asp Leu Ala
 180 185 190
 Ala Leu Lys Tyr Trp Gly Pro Thr Thr Thr Thr Asn Phe Pro Val Ser
 195 200 205
 Asn Tyr Glu Lys Glu Leu Glu Glu Met Lys Ser Met Thr Arg Gln Glu
 210 215 220

Phe Ile Ala Ser Leu Arg Arg Lys Ser Ser Gly Phe Ser Arg Gly Ala
 225 230 235 240
 Ser Ile Tyr Arg Gly Val Thr Arg His His Gln His Gly Arg Trp Gln
 245 250 255
 Ala Arg Ile Gly Arg Val Ala Gly Asn Lys Asp Leu Tyr Leu Gly Thr
 260 265 270
 Phe Ser Thr Gln Glu Glu Ala Ala Glu Ala Tyr Asp Ile Ala Ala Ile
 275 280 285
 Lys Phe Arg Gly Leu Asn Ala Val Thr Asn Phe Asp Met Ser Arg Tyr
 290 295 300
 Asp Val Glu Ser Ile Leu Ser Ser Asp Leu Pro Val Gly Gly Gly Ala
 305 310 315 320
 Ser Gly Arg Ala Pro Ala Lys Phe Pro Leu Asp Ser Leu Gln Pro Gly
 325 330 335
 Ser Ala Ala Ala Met Met Leu Ala Gly Ala Ala Ala Ala Ser Gln Ala
 340 345 350
 Thr Met Pro Pro Ser Glu Lys Asp Tyr Trp Ser Leu Leu Ala Leu His
 355 360 365
 Tyr Gln Gln Gln Gln Glu Gln Glu Arg Gln Phe Pro Ala Ser Ala Tyr
 370 375 380
 Glu Ala Tyr Gly Ser Gly Gly Val Asn Val Asp Phe Thr Met Gly Thr
 385 390 395 400
 Ser Ser Gly Asn Asn Asn Asn Asn Thr Gly Ser Gly Val Met Trp Gly
 405 410 415
 Ala Thr Thr Gly Ala Val Val Val Gly Gln Gln Asp Ser Ser Gly Lys
 420 425 430
 Gln Gly Asn Gly Tyr Ala Ser Asn Ile Pro Tyr Ala Ala Ala Ala Met
 435 440 445
 Val Ser Gly Ser Ala Gly Tyr Glu Gly Ser Thr Gly Asp Asn Gly Thr
 450 455 460
 Trp Val Thr Thr Thr Thr Ser Ser Asn Thr Gly Thr Ala Pro His Tyr
 465 470 475 480
 Tyr Asn Tyr Leu Phe Gly Met Glu
 485

<210> 255

<211> 470

<212> PRT

<213> Zea mays

<400> 255

Met	Ala	His	Pro	Ser	Ala	Ala	Ala	Ala	Ala	Val	Ser	Ser	Thr	Ala	Pro
1				5					10					15	
Ala	Ala	Asn	Pro	Ser	Ser	Gly	Ala	Arg	Glu	Glu	Gly	Gly	Ala	Arg	Ser
			20					25					30		
Pro	Pro	Ser	Pro	Ser	Pro	Ser	Gln	Arg	Gly	Arg	Ala	Lys	Val	Val	Ile
		35					40					45			
Val	Met	Gly	Ala	Thr	Gly	Ala	Gly	Lys	Ser	Arg	Leu	Ala	Val	Asp	Leu
	50					55					60				
Ala	Ala	His	Phe	Ala	Gly	Val	Glu	Val	Val	Ser	Ala	Asp	Ser	Met	Gln
65					70					75				80	
Leu	Tyr	Arg	Gly	Leu	Asp	Val	Leu	Thr	Asn	Lys	Ala	Pro	Leu	His	Glu
				85					90					95	
Gln	Asn	Gly	Val	Pro	His	His	Leu	Leu	Ser	Val	Ile	Asp	Pro	Ser	Val
			100					105					110		
Glu	Phe	Thr	Cys	Arg	Asp	Phe	Arg	Asp	Arg	Ala	Val	Pro	Ile	Ile	Gln
		115					120					125			
Glu	Ile	Val	Asp	Arg	Gly	Gly	Leu	Pro	Val	Val	Val	Gly	Gly	Thr	Asn
	130					135					140				
Phe	Tyr	Ile	Gln	Ala	Leu	Val	Ser	Pro	Phe	Leu	Leu	Asp	Asp	Met	Ala
145					150					155				160	
Glu	Glu	Met	Gln	Gly	Cys	Thr	Leu	Arg	Asp	His	Ile	Asp	Asp	Gly	Leu
				165					170					175	
Thr	Asp	Glu	Asp	Glu	Gly	Asn	Gly	Phe	Glu	Arg	Leu	Lys	Glu	Ile	Asp
		180					185						190		
Pro	Val	Ala	Ala	Gln	Arg	Ile	His	Pro	Asn	Asp	His	Arg	Lys	Ile	Lys
		195					200					205			
Arg	Tyr	Leu	Glu	Leu	Tyr	Ala	Thr	Thr	Gly	Ala	Leu	Pro	Ser	Asp	Leu
	210					215					220				
Phe	Gln	Gly	Glu	Ala	Ala	Lys	Lys	Trp	Gly	Arg	Pro	Ser	Asn	Ser	Arg
225					230					235				240	
Leu	Asp	Cys	Cys	Phe	Leu	Trp	Val	Asp	Ala	Asp	Leu	Gln	Val	Leu	Asp
				245					250					255	
Ser	Tyr	Val	Asn	Lys	Arg	Val	Asp	Cys	Met	Met	Asp	Gly	Gly	Leu	Leu
		260					265						270		
Asp	Glu	Val	Cys	Ser	Ile	Tyr	Asp	Ala	Asp	Ala	Val	Tyr	Thr	Gln	Gly
		275					280					285			
Leu	Arg	Gln	Ala	Ile	Gly	Val	Arg	Glu	Phe	Asp	Glu	Phe	Phe	Arg	Ala
	290					295					300				
Tyr	Leu	Pro	Arg	Lys	Glu	Ser	Gly	Glu	Gly	Ser	Cys	Ala	Ser	Leu	Leu

305	310	315	320
Gly Met His Asp Asp Gln Leu Lys Ser Leu Leu Asp Glu Ala Val Ser			
	325	330	335
Gln Leu Lys Ala Asn Thr Arg Arg Leu Val Arg Arg Gln Arg Arg Arg			
	340	345	350
Leu His Arg Leu Ser Lys Asp Phe Gly Trp Asn Leu His Arg Val Asp			
	355	360	365
Ala Thr Glu Ala Phe Phe Cys Ala Thr Asp Asp Ser Trp Gln Lys Lys			
	370	375	380
Val Val Lys Pro Cys Val Asp Val Val Arg Arg Phe Leu Ser Asp Asn			
385	390	395	400
Ser Thr Val Leu Pro Ser Thr Ser Ala Ser Asp Pro Ser Ser Arg Glu			
	405	410	415
Leu Trp Thr Gln Tyr Val Cys Glu Ala Cys Gly Asn Arg Val Leu Arg			
	420	425	430
Gly Ala His Glu Trp Glu Gln His Arg Gln Gly Arg Gly His Arg Lys			
	435	440	445
Arg Val Gln Arg Leu Lys Gln Lys Ser Leu Arg Pro Trp Pro Ser Leu			
	450	455	460
Leu Pro Gln Asp Arg Ser			
465	470		
<210> 256			
<211> 359			
<212> PRT			
<213> Zea mays			
<400> 256			
Met Glu Glu Ile Thr Gln His Phe Gly Val Gly Ala Ser Ser His Gly			
1	5	10	15
His Gly His Gly Gln His His His His His His His His His Pro Trp			
	20	25	30
Ala Ser Ser Leu Ser Ala Val Val Ala Pro Leu Pro Pro Gln Pro Pro			
	35	40	45
Ser Ala Gly Leu Pro Leu Thr Leu Asn Thr Val Ala Ala Thr Gly Asn			
	50	55	60
Ser Gly Gly Ser Gly Asn Pro Val Leu Gln Leu Ala Asn Gly Gly Gly			
65	70	75	80
Leu Leu Asp Ala Cys Val Lys Ala Lys Glu Pro Ser Ser Ser Ser Pro			
	85	90	95
Tyr Ala Gly Asp Val Glu Ala Ile Lys Ala Lys Ile Ile Ser His Pro			

100	105	110
His Tyr Tyr Ser Leu Leu Thr Ala Tyr Leu Glu Cys Asn Lys Val Gly		
115	120	125
Ala Pro Pro Glu Val Ser Ala Arg Leu Thr Glu Ile Ala Gln Glu Val		
130	135	140
Glu Ala Arg Gln Arg Thr Ala Leu Gly Gly Leu Ala Ala Ala Thr Glu		
145	150	155
Pro Glu Leu Asp Gln Phe Met Glu Ala Tyr His Glu Met Leu Val Lys		
165	170	175
Phe Arg Glu Glu Leu Thr Arg Pro Leu Gln Glu Ala Met Glu Phe Met		
180	185	190
Arg Arg Val Glu Ser Gln Leu Asn Ser Leu Ser Ile Ser Gly Arg Ser		
195	200	205
Leu Arg Asn Ile Leu Ser Ser Gly Ser Ser Glu Glu Asp Gln Glu Gly		
210	215	220
Ser Gly Gly Glu Thr Glu Leu Pro Glu Val Asp Ala His Gly Val Asp		
225	230	235
Gln Glu Leu Lys His His Leu Leu Lys Lys Tyr Ser Gly Tyr Leu Ser		
245	250	255
Ser Leu Lys Gln Glu Leu Ser Lys Lys Lys Lys Lys Gly Lys Leu Pro		
260	265	270
Lys Glu Ala Arg Gln Gln Leu Leu Ser Trp Trp Asp Gln His Tyr Lys		
275	280	285
Trp Pro Tyr Pro Ser Glu Thr Gln Lys Val Ala Leu Ala Glu Ser Thr		
290	295	300
Gly Leu Asp Leu Lys Gln Ile Asn Asn Trp Phe Ile Asn Gln Arg Lys		
305	310	315
Arg His Trp Lys Pro Ser Glu Glu Met His His Leu Met Met Asp Gly		
325	330	335
Tyr His Thr Thr Asn Ala Phe Tyr Met Asp Gly His Phe Ile Asn Asp		
340	345	350
Gly Gly Leu Tyr Arg Leu Gly		
355		
<210> 257		
<211> 311		
<212> PRT		
<213> Zea mays		
<400> 257		
Met Thr Gly Leu Asp Glu Ala Leu Met Leu Pro Phe Thr Asp Ile Asp		

1													5													10													15
Leu	Glu	Ala	Phe	Asp	Asn	Ala	Glu	Glu	Gln	Lys	Pro	Pro	Val	Asp	Gln																								
20								25								30																							
Met	Val	Met	Met	Pro	Pro	Thr	Val	Glu	His	Pro	Ala	Ala	Ala	Gly	Thr																								
35								40								45																							
Arg	Ala	Pro	Ile	Ile	Ile	Asp	Gly	Thr	Ala	Thr	Val	Gly	Gln	Asn	Val																								
50								55								60																							
Gly	Gly	Gly	Val	Val	His	Ala	His	Gln	Lys	Ala	Ala	Met	Thr	Thr	Ile																								
65					70								75				80																						
Glu	Asp	Ser	Ser	Cys	Phe	Arg	Arg	Gly	Ala	Ser	Cys	Val	Asp	Asp	Asp																								
				85								90								95																			
Met	Ala	Val	Val	Ile	His	His	Val	Glu	Arg	Arg	Arg	Gln	Ala	Gly	Ser																								
				100								105								110																			
Thr	Ala	Val	Ala	Leu	Leu	Pro	Pro	Pro	Gln	Pro	Ser	Leu	Pro	Arg	Pro																								
				115								120								125																			
Arg	Ala	Arg	Ala	Ser	Gly	Gly	Ala	Gly	Glu	Arg	Ser	Ala	Pro	Ala	Ala																								
				130								135								140																			
Ala	Gly	Lys	Thr	Arg	Met	Asp	His	Ile	Gly	Phe	Asp	Glu	Leu	Arg	Lys																								
145					150								155				160																						
Tyr	Phe	Tyr	Met	Pro	Ile	Thr	Arg	Ala	Ala	Arg	Glu	Met	Asn	Val	Gly																								
				165								170								175																			
Leu	Thr	Val	Leu	Lys	Lys	Arg	Cys	Arg	Glu	Leu	Gly	Val	Ala	Arg	Trp																								
				180								185								190																			
Pro	His	Arg	Lys	Met	Lys	Ser	Leu	Lys	Ser	Leu	Met	Ala	Asn	Val	Gln																								
				195								200								205																			
Glu	Met	Gly	Asn	Gly	Met	Ser	Pro	Val	Ala	Val	Gln	His	Glu	Leu	Ala																								
				210								215								220																			
Ala	Leu	Glu	Thr	Tyr	Cys	Ala	Leu	Met	Glu	Glu	Asn	Pro	Trp	Ile	Glu																								
225					230								235				240																						
Leu	Thr	Asp	Arg	Thr	Lys	Arg	Leu	Arg	Gln	Ala	Cys	Phe	Lys	Glu	Ser																								
				245								250								255																			
Tyr	Lys	Arg	Arg	Lys	Ala	Ala	Ala	Gly	Asn	Ala	Ile	Glu	Thr	Asp	His																								
				260								265								270																			
Ile	Val	Tyr	Ser	Phe	Gly	Gln	His	Arg	Arg	Tyr	Lys	Gln	Gln	Leu	Leu																								
				275								280								285																			
Pro	Pro	Pro	Thr	Ala	Gly	Ser	Thr	Ser	Ala	Asp	Asp	Arg	His	Gly	Gln																								
				290								295								300																			
Ser	Ser	Arg	Phe	Phe	Cys	Tyr																																	
305					310																																		

<210> 258

<211> 391

<212> PRT

<213> Zea mays

<400> 258

Met	Ala	Met	Val	Pro	Cys	Gly	Gly	Asp	Asp	Ala	Glu	Trp	Cys	Asn	Met
1				5				10						15	
Met	Glu	Ala	Ile	Asn	His	Leu	Met	Met	Ser	Ser	Met	Ser	Ser	Pro	His
			20					25						30	
Val	Ala	Met	Gly	Ala	Ser	Ser	Cys	Arg	Glu	Glu	Asp	Asp	Asp	Ser	Leu
			35					40						45	
Tyr	Leu	Pro	Met	Tyr	Tyr	Ser	Ser	Ala	Pro	Pro	Pro	Ala	Val	Val	Ser
			50					55						60	
Asp	Gln	Tyr	Cys	Pro	Glu	Gln	Leu	Pro	Pro	Leu	Pro	Ala	Ala	Gly	Ala
65								70						75	
Met	Thr	Gly	Leu	Asp	Glu	Ala	Leu	Met	Leu	Pro	Phe	Thr	Asp	Ile	Asp
								85						90	
Leu	Glu	Ala	Phe	Asp	Asn	Ala	Glu	Glu	Gln	Lys	Pro	Pro	Val	Asp	Gln
								100						105	
Met	Val	Met	Met	Pro	Pro	Thr	Val	Glu	His	Pro	Ala	Ala	Ala	Gly	Thr
								115						120	
Arg	Ala	Pro	Ile	Ile	Ile	Asp	Gly	Thr	Ala	Thr	Val	Gly	Gln	Asn	Val
								130						135	
Gly	Gly	Gly	Val	Val	His	Ala	His	Gln	Lys	Ala	Ala	Met	Thr	Thr	Ile
								145						150	
Glu	Asp	Ser	Ser	Cys	Phe	Arg	Arg	Gly	Ala	Ser	Cys	Val	Asp	Asp	Asp
								155						160	
Met	Ala	Val	Val	Ile	His	His	Val	Glu	Arg	Arg	Arg	Gln	Ala	Gly	Ser
								165						170	
Met	Ala	Val	Val	Ile	His	His	Val	Glu	Arg	Arg	Arg	Gln	Ala	Gly	Ser
								180						185	
Thr	Ala	Val	Ala	Leu	Leu	Pro	Pro	Pro	Gln	Pro	Ser	Leu	Pro	Arg	Pro
								190						195	
Arg	Ala	Arg	Ala	Ser	Gly	Gly	Ala	Gly	Glu	Arg	Ser	Ala	Pro	Ala	Ala
								200						205	
Ala	Gly	Lys	Thr	Arg	Met	Asp	His	Ile	Gly	Phe	Asp	Glu	Leu	Arg	Lys
								210						215	
Ala	Gly	Lys	Thr	Arg	Met	Asp	His	Ile	Gly	Phe	Asp	Glu	Leu	Arg	Lys
								220						225	
Tyr	Phe	Tyr	Met	Pro	Ile	Thr	Arg	Ala	Ala	Arg	Glu	Met	Asn	Val	Gly
								230						235	
Leu	Thr	Val	Leu	Lys	Lys	Arg	Cys	Arg	Glu	Leu	Gly	Val	Ala	Arg	Trp
								240						245	
Leu	Thr	Val	Leu	Lys	Lys	Arg	Cys	Arg	Glu	Leu	Gly	Val	Ala	Arg	Trp
								250						255	
Leu	Thr	Val	Leu	Lys	Lys	Arg	Cys	Arg	Glu	Leu	Gly	Val	Ala	Arg	Trp
								260						265	
Leu	Thr	Val	Leu	Lys	Lys	Arg	Cys	Arg	Glu	Leu	Gly	Val	Ala	Arg	Trp
								270							

Pro His Arg Lys Met Lys Ser Leu Lys Ser Leu Met Ala Asn Val Gln
 275 280 285
 Glu Met Gly Asn Gly Met Ser Pro Val Ala Val Gln His Glu Leu Ala
 290 295 300
 Ala Leu Glu Thr Tyr Cys Ala Leu Met Glu Glu Asn Pro Trp Ile Glu
 305 310 315 320
 Leu Thr Asp Arg Thr Lys Arg Leu Arg Gln Ala Cys Phe Lys Glu Ser
 325 330 335
 Tyr Lys Arg Arg Lys Ala Ala Ala Gly Asn Ala Ile Glu Thr Asp His
 340 345 350
 Ile Val Tyr Ser Phe Gly Gln His Arg Arg Tyr Lys Gln Gln Leu Leu
 355 360 365
 Pro Pro Pro Thr Ala Gly Ser Thr Ser Ala Asp Asp Arg His Gly Gln
 370 375 380
 Ser Ser Arg Phe Phe Cys Tyr
 385 390

<210> 259

<211> 33

<212> DNA

<213> herpes simplex

<400> 259

gacgcttttg acgacttcga cttggacatg ttg 33

<210> 260

<211> 183

<212> DNA

<213> herpes simplex

<400> 260

gaagcctctg gatctggcag agccgatgcc ctggatgatt ttgatctgga tatgctggga 60
 agcgacgccc tggatgattt cgatctggat atgctgggat ctgacgccct ggatgatttc 120
 gatctggata tgctgggata tgacgcctg gatgatttcg atctggacat gctgatcaac 180
 agc 183

<210> 261

<211> 1569

<212> DNA

<213> 人工序列

<220>

<223> tripartite effector VPR (VP64, p65, and Rta)

<400> 261

gacgcattgg acgattttga tctggatatg ctgggaagtg acgccctcga tgattttgac 60

cttgacatgc ttggttcggg tggccttgat gactttgacc tcgacatgct cggcagtgac 120
 gcccttgatg atttcgacct ggacatgctg attaaactta gaagttccgg atctccgaaa 180
 aagaaacgca aagttggttag ccagtagctg cccgacaccg acgaccggca ccgtagcgag 240
 gaaaagcggg agcggaccta cgagacattc aagagcatca tgaagaagtc ccccttcagc 300
 ggccccaccg accctagacc tccacctaga agaatcgccg tgcccagcag atccagcgcc 360
 agcgtgccaa aacctgcccc ccagccttac ccttcacca gcagcctgag caccatcaac 420
 tacgacgagt tccctaccat ggtgttcccc agcggccaga tctctcaggc ctctgctctg 480
 gctccagccc ctctcaggt gctgcctcag gctcctgtc ctgcaccagc tccagccatg 540
 gtgtctgcac tggctcaggc accagcacc gtgcctgtgc tggctcctgg acctccacag 600
 gctgtggctc caccagcccc taaacctaca caggccggcg agggcacact gtctgaagct 660
 ctgctgcagc tgcagttcga cgacaggat ctgggagccc tgctgggaaa cagcaccgat 720
 cctgccgtgt tcaccgacct ggccagcgtg gacaacagcg agttccagca gctgctgaac 780
 cagggcaccc ctgtggcccc tcacaccacc gagcccatgc tgatggaata ccccgaggcc 840
 atcaccgggc tcgtgacagg cgctcagagg cctcctgac cagctcctgc cctctggga 900
 gcaccaggcc tgcctaattg actgctgtct ggccagcagg acttcagctc tatcgccgac 960
 atggacttct ccgcaactgt gggtagcggg tcgggatctc gggattccag ggaagggatg 1020
 tttttgccga agcctgaggc cggtccgct attagtgcg tgtttgaggg ccgcgaggtg 1080
 tgccagccaa aacgaatccg gccatttcat cctccaggaa gtccatgggc caaccgccc 1140
 ctccccgcca gcctcgacc aacaccaacc ggtccagtag atgagccagt cgggtcactg 1200
 accccggcac cagtccctca gccactggat ccagcggccg cagtactcc cgaggccagt 1260
 cacctgttgg aggatcccg tgaagagacg agccaggctg tcaaagccct tcgggagatg 1320
 gccgatactg tgattcccca gaaggaagag gctgcaatct gtggccaaat ggacctttcc 1380
 catccgcccc caaggggcca tctggatgag ctgacaacca cacttgagtc catgaccgag 1440
 gatctgaacc tggactcacc cctgaccccg gaattgaacg agattctgga taccttcctg 1500
 aacgacgagt gcctcttgca tgccatgcat atcagcacag gactgtccat ctctgacaca 1560
 tctctgttt 1569

<210> 262

<211> 136

<212> DNA

<213> 人工序列

<220>

<223> SAM Part I (modification to the gRNA adding two ms2 hairpin extensions)

<400> 262

gtttttagagc taggccaaca tgaggatcac ccatgtctgc agggcctagc aagttaaaat 60
 aaggctagtc cgttatcaac ttggccaaca tgaggatcac ccatgtctgc agggccaagt 120
 ggacccgagt cgggtgc 136

<210> 263

<211> 399

<212> DNA

<213> 人工序列

<220>

<223> MCP domain

<400> 263

```
gcggccgctg actacaagga tgacgacgat aaatctagaa tggcttctaa ctttactcag 60
ttcgtttctcg tcgacaatgg cggaactggc gacgtgactg tcgccccaaag caacttcgct 120
aacgggatcg ctgaatggat cagctctaac tcgcgttcac aggcttataa agtaacctgt 180
agcgttcgtc agagctctgc gcagaatcgc aaatacacca tcaaagtcga ggtgcctaaa 240
ggcgccctggc gttcgtactt aaatatggaa ctaaccattc caattttcgc cacgaattcc 300
gactgcgagc ttattgttaa ggcaatgcaa ggtctcctaa aagatggaaa cccgattccc 360
tcagcaatcg cagcaaaactc cggcatctac gaggccagc 399
```

<210> 264

<211> 56

<212> DNA

<213> 人工序列

<220>

<223> Scaffold Part I (modification to the gRNA)

<400> 264

```
gggagcacat gaggatcacc catgtgacgac tcccacagtc actggggagt cttccc 56
```

<210> 265

<211> 353

<212> DNA

<213> 人工序列

<220>

<223> MCP domain

<400> 265

```
agaatggctt ctaactttac tcagttcgtt ctcgtcgaca atggcggaac tggcgacgtg 60
actgtcgccc caagcaactt cgctaacggg atcgctgaat ggatcagctc taactcgcgt 120
tcacaggctt acaaagtaac ctgtagcggt cgtcagagct ctgcgagaa tcgcaaatac 180
accatcaaag tcgaggtgcc taaaggcgcc tggcgttcgt acttaaatat ggaactaacc 240
attccaatth tcgccacgaa ttccgactgc gagcttattg ttaaggcaat gcaaggtctc 300
ctaaaagatg gaaacccgat tccctcagca atcgagcaa actccggcat cta 353
```

<210> 266

<211> 705

<212> DNA

<213> 人工序列

<220>

<223> Suntag Part I (10xGCN4_v4)

<400> 266

gaagaacttt tgagcaagaa ttatcatctt gagaacgaag tggctcgtct taagaaaggt 60
 tctggcagtg gagaagaact gctttcaaag aattaccacc tggaaaatga ggtagctaga 120
 ctgaaaaagg ggagcggaag tggggaggag ttgctgagca aaaattatca tttggagaac 180
 gaagtagcac gactaaagaa aggggtccgga tcgggtgagg agttactctc gaaaaattat 240
 catctcgaaa acgaagtggc tcggctaaaa aagggcagtg gttctggaga agagctatta 300
 tctaaaaact accacctcga aaatgaggtg gcacgcttaa aaaagggaag tggcagtgg 360
 gaagagctac tatccaagaa ttatcatctt gagaacgagg tagcgcggtt gaagaagggt 420
 tccggctcag gagaggaaact gctctcgaag aactatcatc ttgaaaatga ggtcgctcga 480
 ttaaaaaagg gatcgggcag tggtagaggaa ctactttcaa agaattacca ctcgaaaac 540
 gaagtagctc gattaaagaa aggttcaggg tcgggtgaag aattactgag taaaaattat 600
 catctggaac atgaggtagc gagactaaaa aaggggagtg gttctggcga ggaattgcta 660
 tcgaaaaatt atcatcttga gaacgaagtt gctaggctca aaaag 705

<210> 267

<211> 831

<212> DNA

<213> 人工序列

<220>

<223> Suntag Part II (ScFv_GCN4)

<400> 267

atgggccccg acatcgtgat gaccagagc cccagcagcc tgagcgccag cgtgggagac 60
 cgcgtgacca tcacctgccg cagcagcacc ggcgccgtga ccaccagcaa ctacgccagc 120
 tgggtgcagg agaagccccg caagctgttc aagggcctga tcggcgccac caacaaccgc 180
 gccccggcgc tgcccagccg cttcagcggc agcctgatcg gcgacaaggc caccctgacc 240
 atcagcagcc tgcagcccga ggacttcgcc acctacttct gcgccctgtg gtacagcaac 300
 cactgggtgt tcggccaggg caccaagggt gagctgaagc gcggcgccgc cggcagcggc 360
 ggcgccggca gcggcgccgc cggcagcagc ggcgccggca gcgaggtgaa gctgctggag 420
 agcgccggcg gcctgggtga gcccggcggc agcctgaagc tgagctgcgc cgtgagcggc 480
 ttcagcctga ccgactacgg cgtgaactgg gtgcgccagg cccccggccg cggcctggag 540
 tggatcgccg tgatctgggg cgacggcatc accgactaca acagcgccct gaaggaccgc 600
 ttcacatca gcaaggacaa cggcaagaac accgtgtacc tgcagatgag caaggtgcgc 660
 agcgacgaca ccgccctgta ctactgcgtg accggcctgt tcgactactg gggccagggc 720
 accctgggtga ccgtgagcag ctaccatac gatgttccag attacgtgg tggaggcgga 780
 ggttctgggg gaggaggtag tggcggtggt ggttcaggag gcggcggaag c 831

<210> 268

<211> 1851

<212> DNA

<213> 人工序列

<220>

<223> P300

<400> 268

```
attttcaaac cagaagaact acgacaggca ctgatgccaa ctttggaggc actttaccgt 60
caggatccag aatccccttc ctttcgtcaa cctgtggacc ctcagctttt aggaatccct 120
gattactttg atattgtgaa gagcccatg gatctttcta ccattaagag gaagttagac 180
actggacagt atcaggagcc ctggcagtat gtcgatgata tttggcttat gttcaataat 240
gcctgggttat ataaccggaa aacatcacgg gtatacaaat actgctccaa gctctctgag 300
gtctttgaac aagaaattga cccagtgatg caaagccttg gatactgttg tggcagaaag 360
ttggagttct ctccacagac actgtgttgc tacggcaaac agttgtgcac aatacctcgt 420
gatgccactt attacagtta ccagaacagg tatcatttct gtgagaagtg tttcaatgag 480
atccaagggg agagcgtttc tttgggggat gacccttccc agcctcaaac tacaataaat 540
aaagaacaat tttccaagag aaaaaatgac aactggatc ctgaactgtt tgttgaatgt 600
acagagtgcg gaagaaagat gcatcagatc tgtgtccttc accatgagat catctggcct 660
gctggattcg tctgtgatgg ctgtttaaag aaaagtgcac gaactaggaa agaaaataag 720
ttttctgcta aaaggttgcc atctaccaga cttggcacct ttctagagaa tcgtgtgaat 780
gactttctga ggcgacagaa tcaccctgag tcaggagagg tcaactgttag agtagttcat 840
gcttctgaca aaaccgtgga agtaaaacca ggcataaaag caaggtttgt ggacagtgga 900
gagatggcag aatcctttcc ataccgaacc aaagccctct ttgcctttga agaaattgat 960
gggtgttgacc tgtgtcttct tggcatgcat gttcaagagt atggctctga ctgccctcca 1020
cccaaccaga ggagagtata catatcttac ctcgatagtg ttcatttctt ccgtcctaaa 1080
tgcttgagga ctgcagtcta tcatgaaatc ctaattggat atttagaata tgtcaagaaa 1140
ttaggttaca caacagggca tatttgggca tgtccaccaa gtgagggaga tgattatata 1200
ttccattgcc atcctcctga ccagaagata cccaagccca agcgactgca ggaatggtac 1260
aaaaaaatgc ttgacaaggc tgtatcagag cgtattgtcc atgactacaa ggatattttt 1320
aaacaagcta ctgaagatag attaacaagt gcaaaggaat tgccttattt cgagggtgat 1380
ttctggccca atgttctgga agaaagcatt aaggaactgg aacaggagga agaagagaga 1440
aaacgagagg aaaacaccag caatgaaagc acagatgtga ccaagggaga cagcaaaaat 1500
gctaaaaaga agaataataa gaaaaccagc aaaaataaga gcagcctgag taggggcaac 1560
aagaagaaac ccgggatgcc caatgtatct aacgacctct cacagaaact atatgccacc 1620
atggagaagc ataaagaggt cttctttgtg atccgcctca ttgctggccc tgctgccaac 1680
tccctgcctc ccattgttga tcttgatcct ctcacccctt gcgatctgat ggatggctcg 1740
gatgcgtttc tcacgtggc aagggacaag cacctggagt tctcttcact ccgaagagcc 1800
cagtgggtcca ccatgtgcat gctggtggag ctgcacacgc agagccagga c 1851
```

<210> 269

<211> 384

<212> DNA

<213> 人工序列

<220>

<223> VP160

<400> 269

gacgcgctgg acgatttcga tctcgacatg ctgggttctg atgccctcga tgactttgac 60
ctggatatgt tgggaagcga cgcattggat gactttgac tggacatgct cggctccgat 120
gctctggacg atttcgatct cgatatgtta gggtcagacg cactggatga tttcgacctt 180
gatatgttgg gaagcgatgc ctttgatgat ttcgacctgg acatgctcgg cagcgacgcc 240
ctggacgatt tcgatctgga catgctgggg tccgatgcct tggatgattt tgacttggat 300
atgctgggga gtgatgccct ggacgacttt gacctggaca tgctgggctc cgatgcgctc 360
gatgacttcg atttggatat gttg 384

<210> 270

<211> 25

<212> DNA

<213> 人工序列

<220>

<223> BBM target sequence

<400> 270

tggagtgtac cagttgtata aatat 25

<210> 271

<211> 25

<212> DNA

<213> 人工序列

<220>

<223> BBM target sequence

<400> 271

tcctcgaatc attctaagaa gaaac 25

<210> 272

<211> 25

<212> DNA

<213> 人工序列

<220>

<223> BBM target sequence

<400> 272

tggccgtgac aacgtatact attat 25

<210> 273

<211> 11414

<212> DNA

<213> 人工序列

<220>

<223> pGEP362 expression plasmid

<400> 273

agcatgaatg cctgggggag aagaactcga gaggaattg cagatcatga ggcagatggc 60
 tattttttgtg tcacatatgc gcaaaaagag aggctatatt tgtgtcccta ggttcttcgt 120
 tgtattgcag ttcccatatc aatctgactt ggtcgcatga gaaattgatg gttaaataat 180
 ttgaatctct catgtagtat caactattag atattatatt caccaaatat atttccatcg 240
 gagaagaaga ggctacagag gaagcagaag agaggggtgg gagaattttt acacttttgt 300
 acacccactt aaacagcaaa atccgtatga aaacaggccc accaaaacaa tgccacgata 360
 acaatccgta gaaacaaaaag cttcatthta cagcggcgca acaaagcacg cttatccatg 420
 gtagttgtag tccgtatgcg atccaaagat cagattcac gcgtgacgga cggacgacgc 480
 gtgccacacc acaactaacg gcatccatgg tagttgtagt ccgtatgca tccaaagatc 540
 acgattcacg cgtgacggac ggacgacgcg cgccacacca caactaacag cgtgagccag 600
 cgtccaaact ccgatggca acggggacga aaccgctcg gtagtactg cccaaacccg 660
 tccccgaac cttcatccca aaccgctccc cgtttccggt cgcggtttc agttttctac 720
 cagaccgctc cccatcggtt ttttcatccc cgtcgggaaa tccgaacccg ccagcatttc 780
 agcaccaagc caaagttgca gcagcaacat gaataaaaaa caaccgttt caacaccaag 840
 ataaaacaaa acattataat ttagacaaca ttccacacgt ataacaataa catatagttc 900
 tcacatataa caacaccatt tcacacataa aacaacacca tttgggataa aaatatgggc 960
 tatatcaggc catttttatg ggccatattg agttttcgtg ggtttcacag gtaccggatt 1020
 tgtagaatgc tgaaccgggt ttgaaccgta aaatccgcgg gtattgaatt tgaccaatc 1080
 ccgtcgtccc ctggtgggtt aaaaacacca tcttgagtc aaacggccac caaccaaact 1140
 ccgacggcaa caaacaacg gcgttgcttt gctcctcgt atctccgtga ccgctcaatc 1200
 tcccggctgt tccccggaa ttgcgtggac tctctcatcc acacgaaac cgcctctccc 1260
 tctctctcgc tctatccgc cccggtgccg tagcctcacg ggactcttct tctcccttg 1320
 ctataaaatc cccgccccct cccgtctcct ctccacacat ccaaactctc aatcgaccg 1380
 agaaaaatct cctagcgatc gaagcgaagc ctctcccgat cctctcaagg tacgcccgtt 1440
 tcccgtcgat cctcctcctt ccgttcgtgt tctgtagccg atcgattcga ttcccttaca 1500
 cccgttcgtg ttctctcgtg gatcgatcga ttgtttgttg ctagaaggaa ctcgtagatc 1560
 tggcgtttat gaactgtgat tcgggttagt ccagatcgat tcaggtcggc cgtcgttag 1620
 cctctcggct atgtctggat tatcgtgtag atctgctggt tcagttgatt atgttcttct 1680
 aggagtaatt tcgttgggtc agcgcgattt ctgcttaatc tatgctgctt attgcgcctg 1740
 tacctatcta ctaagctatg tgcacctgta attttctag attattcggt catcctcgta 1800
 gttggtttgt cacagtaatc cgtatgggtt ctgacgatgt tattgttggt catacctagg 1860
 cttctccaga ttttattttg ttaaaattgg atagatctgc tactgatagt tgatgatgga 1920
 atttggtgct gaatctatgc tatttattgc gcctatacct gatctatcgg gctatgtacg 1980
 gctgtagttt actggattat tcgttcaccc tcggtagttg gttcatcggt tgggttctga 2040
 cgataatatt gttgattatg cgtaggttc tgcagattgt tgttaaaatt ggatacatcg 2100
 gttactgatg gttgatgata gatttgtgct gaacctatct gtttattgct cctatacctg 2160
 atctataggg ctatgtatgc ctgtaattta ccagattatt cgttcaccc cgtagttggt 2220
 tcatctctat aattcgtatg gggttcttat atgttatcgt tgattatgcc tagtcttata 2280
 cagattattg tgtcaagatt gaatatacct gctactgatc ggtgataatt tggttagtag 2340

tttgcaatct gctaggaaca cgttaccact gtaatctgta aacatggttt gccagagtag 2400
tttgttctac tactcttgat atggttgctg attttagtcg cctccttttg gatcatgtat 2460
tgatgtcctt gcagatttcc gtgtacttac cccggctttt gtgtacttcg tgtaaacagg 2520
tcgggtaccg aagcaaacaat ggcatctagc atggcaccaa agaaaaaag gaaagtttcc 2580
aaacttgaaa aatttacaaa ctgctactcc ctttccaaga cgcttaggtt taaagcgatc 2640
cccgttggca agacccaaga gaatatcgat aacaaaagac ttctggtcga agatgaaaaa 2700
agggccgaag actacaaggg ggtcaagaag ttgctcgatc gctattatct ttcctttatc 2760
aacgatgtgc ttcattcaat caaactgaag aacttgata actacattag ctttttcaga 2820
aagaaaacga ggactgaaaa ggagaacaag gaacttgaga atcttgaaat aaaccttcgc 2880
aaagaaattg caaaaagcctt caaggggaac gaaggatata aatctctttt caaaaaagac 2940
attatagaaa caattttgcc tgagtttctt gacgacaagg atgaaattgc gctcgtcaat 3000
agctttaacg gatttacaac tgccttcaca gggttcttcg acaataggga gaatatgttt 3060
agcgaggagg caaaaagcac atccatcgca ttcagatgca tcaatgaaaa tcttaccggg 3120
tacatatcga atatggacat atttgaaaaa gtggatgcaa tattcgataa gcacgaagtc 3180
caggagataa aggaaaagat actgaatagc gactatgatg tcgaagattt tttcgaaggt 3240
gagttcttca actttgtcct gactcaagaa ggcatgatg tctataatgc aataattgga 3300
ggttttgtga ctgagtctgg cgagaagata aagggttga acgagtatat caatctctac 3360
aaccagaaga ctaagcaaaa gttgcctaaa tttaaaccgc tttaacagca agttttgagc 3420
gaccgggaaa gcctttcctt ttacggtgaa ggatacacga gcgatgaaga agtcctcgaa 3480
gtcttccgca acacactcaa caagaactca gaaatctttt cctcaattaa aaaattggag 3540
aagcttttca agaacttcca tgaatactct tcggcgggga tttttgtgaa gaacggcccc 3600
gcaatttcca caatatctaa agacattttc ggagaatgga acgtgataag agacaagtgg 3660
aatgcggagt atgatgacat acacctgaag aagaaggcag ttgtgactga aaaatacgaa 3720
gatgacagga gaaaaagctt taaaaagatc gggtcctttt cactggaaca gctgcaggag 3780
tatgccgacg ccgatctttc ggttgtcgaa aagctcaaag aaataattat ccagaaggtc 3840
gatgaaatct acaagggtga cggctcaagc gagaagctct ttgatgctga cttcgtgttg 3900
gagaagtctc ttaaaaaaaaa cgacgcagtc gtgcgataa tgaaagattt gctggattca 3960
gtgaaatcct tcgagaatta tatcaaagcc ttcttcggcg aggggaagga gacaaacagg 4020
gatgagtcct tctatggaga cttcgttctg gcttacgaca tccttcttaa ggtcgaccac 4080
atctatgacg caattcggaa ctatgtgacg cagaagccgt attcgaaaga taagtccaag 4140
ctctatttcc aaaaccctca atttatgggt ggggtgggata aagacaaaga gaccgattac 4200
cgggcaacaa ttttgcggtg cgggtctaaa tattacctcg ctataatgga taagaaatac 4260
gctaaatgtc tccagaaaaat tgacaaaagat gacgtcaacg gcaattatga aaaaatcaat 4320
tataaactcc ttcctggccc aaataaaatg ctcccgaagg tgtttttttc caaaaagtgg 4380
atggcctatt ataatccatc agaggatatt cagaaaatct ataaaaatgg gacctttaag 4440
aagggtgaca tgtttaacct gaacgattgc cacaagctta tagatttttt caaagactct 4500
attagccgct atcccaaatg gtctaattgt tatgatttca acttctctga aactgaaaag 4560
tacaaagata ttgcaggatt ctaccgcgaa gttgaagaac aaggttataa ggtttccttt 4620
gagtcctgct ccaagaaaaga ggtcgataag ttggtcgaag aagggaatt gtatatgttt 4680

```

caaatttaca ataaagactt ttccgacaag tcccatggta cacctaattct gcataccatg 4740
tacttcaaac tgctgttcga tgagaataat cacggtcaga ttgccttgag cggaggggcg 4800
gaactcttca tgaggagagc atcgttgaaa aaagaggagc tcgtcgtgca tccggctaac 4860
agccccattg ctaacaagaa tccggataat ccaaagaaga ctactaccct ctcctatgac 4920
gtctataagg ataagagatt ctctgaggac cagtacgagt tgcacatccc tattgcgata 4980
aataaatgcc ctaagaacat ctttaaaatc aatactgagg tcagagtcct gcttaagcac 5040
gacgacaacc cgtatgtgat cgggattgat aggggtgaaa ggaacttgct ttatatgtg 5100
gttgtcgatg gaaaaggtaa tatagtggaa caatactctc tgaatgaaat tatcaacaac 5160
ttcaatggca ttaggatcaa gaccgactat cattctctgt tggacaagaa agagaaagag 5220
cgcttcgagg cacggcaaaa ctggacgtct attgagaaca tcaaggagct taaggctggt 5280
tacatttctc aggttgtgca caaaatttgc gaactggtcg agaaatatga tgccgttatc 5340
gcacttgaag atctcaacag cggatttaag aattctcggg tgaaagtcga aaaacaggtg 5400
tatcaaaaat tcgaaaagat gctgatcgac aagctcaatt atatggttga taaaagagc 5460
aaccatgcg ccacgggggg tgcgcttaag ggctatcaga ttacgaacaa atttgaatcc 5520
ttcaagtcaa tgtcgacgca aaatgggttt atattctata taccggcgtg gcttacatct 5580
aaaatagatc ctagcactgg gttcgtgaac ctgctgaaaa ccaagtacac ttcaatcgca 5640
gattctaaaa aatttataag cagcttcgac agaatcatgt atgtgccga ggaagacctc 5700
ttcgagtttg cccttgatta caaaaatttc tcaagaacgg atgcagacta cataaagaag 5760
tggaagctgt actcttatgg gaaccggatt cggatattca gaaatccgaa aaaaaacaat 5820
gtctttgatt gggaggaagt ttgtcttacc tctgcttaca aagagctgtt caataaatat 5880
ggcattaatt accagcaagg tgatatccgg gcgctccttt gcgaacagtc tgacaaagct 5940
ttctattctt catttatggc gctcatgtca ttgatgctgc agatgaggaa tagcattacg 6000
gggaggactg atgttgactt tctgatctcg cccgtgaaaa attctgatgg aatcttctac 6060
gattccagga attatgaggc ccaggaaaat gctatccttc ccaagaacgc agacgcaaat 6120
ggcgcgtaga atatagctcg caaggttttg tgggctatag gccaattcaa gaaagccgaa 6180
gacgaaaagc tggacaaagt taagattgct atatctaaca aagagtggct tgagtatgcg 6240
caaacatctg ttaaacacaa acgccccgcg gctacaaaga aggctggcca ggcaaagaag 6300
aagaagtgag tcgaccgatc gttcaaacat ttggcaataa agtttcttaa gattgaatcc 6360
tgttgccggc cttgcgatga ttatcatata atttctgttg aattacgtta agcatgtaat 6420
aattaacatg taatgcatga cgttatttat gagatgggtt ttatgatta gagtcccgca 6480
attatacatt taatacgcg tagaaaaaca aatatacgcg gcaaactagg ataaattatc 6540
gcgcgcggtg tcatctatgt tactagatcg atccccggat atcgcgcccg cgtcgttcgg 6600
ctgcggcgag cggtatcagc tcaactcaaag gcggtataac ggttatccac agaatcaggg 6660
gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag 6720
gccgcgttgc tggcgttttt ccataggctc cgccccctg acgagcatca caaaaatcga 6780
cgctcaagtc agaggtggcg aaacccgaca ggactataaa gataccaggc gtttccccct 6840
ggaagctccc tcgtgcgtc tcctgttccg accctgccgc ttaccggata cctgtccgcc 6900
tttctccctt cgggaagcgt ggcgctttct catagctcac gctgtaggta tctcagttcg 6960
gtgtaggtcg ttcgctccaa gctgggctgt gtgcacgaac cccccgttca gcccgaccgc 7020

```

tgcgccttat ccggttaacta tcgtcttgag tccaacccgg taagacacga cttatcgcca 7080
ctggcagcag ccactggtaa caggattagc agagcgaggt atgtaggcgg tgctacagag 7140
ttcttgaagt ggtggcctaa ctacggctac actagaagga cagtatttgg tatctgcgct 7200
ctgctgaagc cagttacctt cggaaaaaga gttggttagct cttgatccgg caaacaacc 7260
accgctggta gcggtggttt ttttgtttgc aagcagcaga ttacgcgcag aaaaaagga 7320
tctcaagaag atcctttgat cttttctacg gggctctgacg ctcagtggaa cgaaaactca 7380
cgtaaggga ttttgggtcat gagattatca aaaaggatct tcacctagat ctttttaaat 7440
taaaaatgaa gttttaaatc aatctaaagt atatatgagt aaacttggtc tgacagttac 7500
caatgcttaa tcagtgaggc acctatctca gcgatctgtc ttttctgttc atccatagtt 7560
gcctgactcc ccgtcgtgta gataactacg atacgggagg gcttaccatc tggccccagt 7620
gctgcaatga taccgcgaga cccacgtca ccggtccag atttatcagc aataaaccag 7680
ccagccggaa gggccgagcg cagaagtggc cctgcaactt tatccgctc catccagtct 7740
attaattgtt gccgggaagc tagagtaagt agttcgccag ttaatagttt gcgcaacgtt 7800
gttgccattg ctacaggcat cgtggtgtca cgctcgtcgt ttggtatggc ttcattcagc 7860
tccggttccc aacgatcaag gcgagttaca tgatcccca tgttgtgcaa aaaagcggtt 7920
agctccttcg gtccctccgat cgttgtcaga agtaagttgg ccgcagtgtt atcactcatg 7980
gttatggcag cactgcataa ttctcttact gtcatgccat ccgtaagatg cttttctgtg 8040
actggtgagt actcaaccaa gtcattctga gaatagtgtg tgcggcgacc gagttgctct 8100
tgcccgcggt caatacggga taataccgcg ccacatagca gaactttaaa agtgcctcatc 8160
attggaaaac gttcttcggg gcgaaaactc tcaaggatct taccgctgtt gagatccagt 8220
tcgatgtaac ccactcgtgc acccaactga tcttcagcat cttttacttt caccagcgtt 8280
tctgggtgag caaaaacagg aaggcaaaat gccgcaaaaa aggaataag ggcgacacgg 8340
aaatgttgaa tactcact cttcctttt caatattatt gaagcattta tcagggttat 8400
tgtctcatga gcggatacat atttgaatgt atttagaaaa ataaacaaat aggggttccg 8460
cgcacatttc cccgaaaagt gccacctgac gcgccctgta gcggcacgtc taattcgggg 8520
gatctggatt ttagtactgg attttggtt taggaattag aaattttatt gatagaagta 8580
ttttacaaat acaaatacat actaagggtt tcttatatgc tcaacacatg agcgaaaccc 8640
tataggaacc ctaattccct tatctgggaa ctactcacac attattatgg agaaactcga 8700
gcttgtcgat cgacatgatc agggagccct agattatttg tatagttcat ccatgcccat 8760
tacgtcggta aatgccttct gccactcctt gaagttaagt tcggtcttgg aatgtttcaa 8820
ctcagtctta cggaacacgt acatgggttg gttcttaagg tagtttagcg ccattgggtt 8880
agcgaatgtg taggtagtcc tggtctgtag gcgatctc ttgccattgc ctgtggtgta 8940
agaccatttg aaggtactaa tgatggtctt gtcgttaggg taggttttct tggaccggca 9000
ccaatcagcg gcagttaagg agttggtcat gacaggtcca tcagcaggaa agcctgtccc 9060
cttcacttgg gcttctcctt tgatgtggct ccttcgtaa gtgtaacgg agttgacgg 9120
gagcgaagca ccgtcctcaa actgcattgt cctgtggact tggatatccg agccatcaac 9180
catggctgct tggaatggac tcattccgtc agggatatga aggtattgat ggaatccgta 9240
gccaatgtgt ggcaccagaa tccatggaga aaactgaaga tcaccttgg tgctcttgag 9300
gttcagctct tcgtatccgt cattagggtt cccagtgcct tgtccgacca tatcgaagtc 9360

aacgccgttg atggaaccga agatgtgaag ctcatgtgtg gctggaagcg aagccatggt 9420
 atctttcttct cctttactca cggaggacgc catggtggcg ggatcgcgcc ctatcgttcg 9480
 taaatgggtga aaattttcag aaaattgctt ttgctttaaa agaaatgatt taaattgctg 9540
 caatagaagt agaatgcttg attgcttgag attcgtttgt tttgtatatg ttgtgttgag 9600
 aggatcctct agagtcgacc tgcagaagta acaccaaaca acagggtgag catcgacaaa 9660
 agaaacagta ccaagcaaat aaatagcgta tgaaggcagg gctaaaaaaaa tccacatata 9720
 gctgctgcat atgccatcat ccaagtatat caagatcaaa ataattataa aacatacttg 9780
 tttattataa tagataggta ctcaaggtta gagcatatga atagatgctg catatgccat 9840
 catgtatatg catcagtaaa acccacatca acatgtatac ctatcctaga tcgatatttc 9900
 catccatctt aaactcgtaa ctatgaagat gtatgacaca cacatacagt tccaaaatta 9960
 ataaatacac caggtagttt gaaacagtat tctactccga tctagaacga atgaacgacc 10020
 gccaaccac accacatcat cacaaccaag cgaacaaaag catctctgta tatgcatcag 10080
 taaaacccgc atcaacatgt atacctatcc tagatcgata tttccatcca tcatcttcaa 10140
 ttcgtaacta tgaatatgta tggcacacac atacagatcc aaaattaata aatccaccag 10200
 gtagtttgaa acagaattct actccgatct agaacgaccg cccaaccaga ccacatcatc 10260
 acaaccaaga caaaaaaaaa catgaaaaga tgacccgaca aacaagtgca cggcatatat 10320
 tgaaataaag gaaaagggca aaccaaaccc tatgcaacga aacaaaaaaaa atcatgaaat 10380
 cgatcccgtc tgcggaacgg ctagagccat cccaggattc cccaagaga aacactggca 10440
 agtttagcaat cagaacgtgt ctgacgtaca ggtcgcatcc gtgtacgaac gctagcagca 10500
 cggatctaac acaaacacgg atctaacaca aacatgaaca gaagtagaac taccgggccc 10560
 taaccatgga ccggaacgcc gatctagaga aggtagagag ggggggggag gacgagcggc 10620
 gtaccttgaa gcggagggtc cgacgggtgg atttggggga gatccactag ttctagagcg 10680
 gccgccaccg cgggtggaatt ctcgagggtc tctccaaatg aaatgaactt ctttatatag 10740
 aggaagggtc ttgcgaagga tagtgggatt gtgcgtcatc ctttacgtca gtggagatat 10800
 cacatcaatc cacttgcttt gaagacgtgg ttggaacgtc ttctttttcc acgatgctcc 10860
 tcgtgggttg ggggtccatct ttgggaccac tgtcggcaga ggcatcttga acgatagcct 10920
 ttcttttatc gcaatgatgg cattttagg tgccacctc cttttctact gtccttttga 10980
 tcaagtgacc gatagctggg caatggaatc cgaggaggtt tcccgatatt accctttgtt 11040
 gaaaagtctc aatagccctt tggctctctg agactgtatc tttgatattc ttggagtaga 11100
 cgagagtgtc gtgtccacc atgttatcac atcaattcac ttgctttgaa gacgtggttg 11160
 gaacgtcttc tttttccacg atgctcctcg tgggtggggg tccatctttg ggaccactgt 11220
 cggcagaggc atcttgaacg atagcctttc ctttatcgca atgatggcat ttgtaggtgc 11280
 caccttcctt ttctactgtc cttttgatca agtgacagat agctgggcaa tggaatccga 11340
 ggaggtttcc cgatattacc ctttgttgaa aagtctcaat agcccttttg tcttctgaga 11400
 cctgcaggca agca 11414

<210> 274

<211> 11414

<212> DNA

<213> 人工序列

<220>

<223> pGEP487 expression plasmid

<400> 274

aagcaagcat gaatgcctgg gggagaagaa ctcgagaggg aattgcagat catgaggcag 60
atggctatatt ttgtgtcaca tatgcgcaaa aagagaggct atatttgtgt ccctaggttc 120
ttcgttgtat tgcagtttcc atatcaatct gacttggctg catgagaaat tgatggttaa 180
ataatttgaa tctctcatgt agtatcaact attagatatt attttcacca aatatatttc 240
catcggagaa gaagaggcta cagaggaagc agaagagagg ggtgggagaa tttttacact 300
tttgtacacc cacttaaaca gcaaaatccg tatgaaaaca ggcccaccaa aacaatgcc 360
cgataacaat ccgtagaaac aaaagcttca tttaacagcg gcgcaacaaa gcacgcttat 420
ccatggtagt tgtagtccgt atgcgatcca aagatcacga ttcacgcgtg acggacggac 480
gacgcgtgcc acaccacaac taacggcatc catggtagtt gtagtccgta tgcgatccaa 540
agatcacgat tcacgcgtga cggacggacg acgcgcgcca caccacaact aacagcgtga 600
gccagcgtcc aaactccgga tggcaacggg gacgaaacc gtcgggtagt cactgccc 660
accggtcccc gcaaccttca tcccaaacc gtccccgttt ccggtcgcgg gtttcagttt 720
tctaccagac ccgtcccat cggtttttt atccccgtcg ggaaatccga acccgccagc 780
atttcagcac caagccaaag ttgcagcagc aacatgaata aaaaacaacc cgtttcaaca 840
ccaagataaa acaaaacatt ataatttaga caacatttca cacgtataac aataacatat 900
agttctcaca tataacaaca ccatttcaca cataaaacaa caccatttgg gataaaaata 960
tgggctatat caggccattt ttatgggcca tattgagttt tcgtgggttt cacaggtacc 1020
ggattttagt aatgctgaac cgggtttgaa ccgtaaaatc cgcgggtatt gaatttgacc 1080
caatcccgtc gtccccgtgt ggggtaaaaa caccatcttg agtccaaacg gccaccaacc 1140
aaactccgac ggcaacaaac aaacggcggt gctttgctcc tcggtatctc cgtgaccgct 1200
caatctcccg gctgtttccc cggaattgcg tggactctct catccacacg caaaccgcct 1260
ctccccctc tctcgtccta tccgccccgg tgccgtagcc tcacgggact cttcttcctc 1320
ccttgctata aaatccccgc cccctccggt ctctctcca cacatccaaa ctctcaatcg 1380
caccgagaaa aatctcctag cgatcgaagc gaagcctctc ccgatcctct caaggtacgc 1440
ccgtttcccg tcgatcctcc tccttcggt cgtgttctgt agccgatcga ttcgattccc 1500
ttacacccgt tcgtgttctc tcgtggatcg atcgattgtt tgttgctaga aggaactcgt 1560
agatctggcg tttatgaact gtgattcggg ttagtccaga tcgattcagg tcggtcgtcg 1620
ttgagcctct cggtatgtc tggattatcg ttagatctg ctggttcagt tgattatgtt 1680
cttctaggag taatttcgtt gggtcagcgc gatttctgct taatctatgc tgcttattgc 1740
gcctgtacct atctactaag ctatgtgcac ctgtaatttt gctagattat tcgttcatcc 1800
tcgtagttgg tttgtcacag taatccgtat gggttctgac gatgttattg ttggtcatac 1860
ctaggcttct ccagatttta ttttgtaaaa attggataga tctgctactg atagttgatg 1920
atggaatttg gtgctgaatc tatgtatatt attgcgcta tacctgatct atcgggctat 1980
gtacggctgt agtttactgg attattcggt catcctcggg agttggttca tcgtttgggt 2040
tctgacgata atattgttga ttatgcgtag gcttctgcag attgttggtta aaattggata 2100
catcggttac tgatggttga tgatagattt gtgctgaacc tatctgttta ttgctcctat 2160

acctgatcta tagggctatg tatgcctgta atttaccaga ttattcgttc atcctcgtag 2220
 ttggttcatc tctataattc gtatgggttc ttatgatgtt atcgttgatt atgcctagtc 2280
 ttatacagat tatttgttca agattgaata tacctgctac tgatcggtga taatttggtt 2340
 agtagtttgc aatctgctag gaacacgtta ccactgtaat ctgtaaacad ggtttgccag 2400
 agtagtttgt tctactactc ttgatatggg tgctgatttt agtcgcctcc ttttgatca 2460
 tgtattgatg tccttgcaaga tttccgtgta cttaccccg cttttgtgta cttcgtgtta 2520
 acaggtcggg taccgaagca aacatggcat ctagcatggc accaaagaaa aaaaggaaag 2580
 tttccaaact tgaaaaattt acaaaactgct actcccttc caagacgctt aggtttaaag 2640
 cgatccccgt tggcaagacc caagagaata tcgataacaa aagacttctg gtcgaagatg 2700
 aaaaaagggc cgaagactac aaggggttca agaagttgct cgatcgctat tatctttcct 2760
 ttatcaacga tgtgttcat tcaatcaaac tgaagaactt gaataactac attagccttt 2820
 tcagaaagaa aacgaggact gaaaaggaga acaaggaact tgagaatctt gaaataaacc 2880
 ttgcgaagaa aattgcaaaa gccttcaagg ggaacgaagg atataaatct cttttcaaaa 2940
 aagacattat agaaacaatt ttgcctgagt ttcttgacga caaggatgaa attgcgctcg 3000
 tcaatagctt taacggattt acaactgcct tcacagggtt cttcgacaat aggagaata 3060
 tgtttagcga ggaggcaaaa agcacatcca tcgcattcag atgcatcaat gaaaatctta 3120
 cccgtacat atcgaatatg gacatatttg aaaaagtgga tgcaatatc gataagcacg 3180
 aagtccagga gataaaggaa aagatactga atagcgacta tgatgtcgaa gattttttcg 3240
 aaggtgagtt ctcaacttt gtcctgactc aagaaggcat tgatgtctat aatgcaataa 3300
 ttggaggttt tgtgactgag tctggcgaga agataaagg cttgaacgag tatatcaatc 3360
 tctacaacca gaagactaag caaaagtgc ctaaatttaa accgctttac aagcaagttt 3420
 tgagcgaccg ggaaagcctt tccttttacg gtgaaggata cacgagcgat gaagaagtcc 3480
 tcgaagtctt ccgcaacaca ctcaacaaga actcagaaat cttttcctca attaaaaaat 3540
 tggagaagct tttcaagaac ttcgatgaat actcttcggc ggggattttt gtgaagaacg 3600
 gcccggaat ttccacaata tctaaagaca ttttcggaga atggaacgtg ataagagaca 3660
 agtggaatgc ggagtatgat gacatacacc tgaagaagaa ggcagttgtg actgaaaaat 3720
 acgaagatga caggagaaaa agctttaaaa agatcgggtc cttttcactg gaacagctgc 3780
 aggagtatgc cgacgccgat ctttcggttg tcgaaaagct caaagaaata attatccaga 3840
 aggtcgatga aatctacaag gtgtacggct caagcgagaa gctctttgat gctgacttcg 3900
 tgtttggagaa gtctcttaaa aaaaacgacg cagtcgtcgc gataatgaaa gatttgctgg 3960
 attcagtgaa atccttcgag aattatatca aagccttctt cggcgagggg aaggagacaa 4020
 acagggatga gtccttctat ggagacttcg ttctggctta cgacatcctt cttaaggtcg 4080
 accacatcta tgacgcaatt cggaactatg tgacgcagaa gccgtattcg aaagataagt 4140
 tcaagctcta tttccaaaac cctcaattta tgcgtgggtg ggataaagac aaagagaccg 4200
 attaccgggc aacaattttg cggtaggggt ctaaataatta cctcgtata atggataaga 4260
 aatacgttaa atgtctccag aaaattgaca aagatgacgt caacggcaat tatgaaaaaa 4320
 tcaattataa actccttcct ggcccaata aatgctccc gaggtgttt ttttcaaaa 4380
 agtggatggc ctattataat ccatcagagg atattcagaa aatctataaa aatgggacct 4440
 ttaagaaggg tgacatgttt aacctgaacg attgccacaa gcttatagat tttttcaaa 4500

actctattag cgcctatccc aaatgggtcta atgcttatga tttcaacttc tctgaaactg 4560
aaaagtacaa agatattgca ggattctacc gcgaagtga agaacaaggt tataaggttt 4620
cctttgagtc tgcgtccaag aaagaggctc ataagttggc cgaagaagg aaattgtata 4680
tgtttcaaat ttacaataaa gacttttccg acaagtccca tggtagacct aatctgcata 4740
ccatgtactt caaactgctg ttcgatgaga ataatcacgg tcagattcgc ctgagcggag 4800
gggcggaact cttcatgagg agagcatcgt tgaaaaaaga ggagctcgtc gtgcatccgg 4860
ctaacagccc cattgctaac aagaatccgg ataatccaaa gaagactact accctctcct 4920
atgacgtcta taaggataag agattctctg aggaccagta cgagttgcac atccctattg 4980
cgataaataa atgccctaag aacatcttta aatcaatac tgaggtcaga gtctgtctta 5040
agcacgacga caaccgctat gtgatcggga ttgatagggg tgaaaggaac ttgctttata 5100
ttgtggttgt cgatggaaaa ggtaatatag tggaacaata ctctctgaat gaaattatca 5160
acaacttcaa tggcattagg atcaagaccg actatcattc tctgttggac aagaaagaga 5220
aagagcgctt cgaggcacgg caaaactgga cgtctattga gaacatcaag gagcttaagg 5280
ctggttacat ttctcaggtt gtgcacaaaa tttgcgaact ggtcgagaaa tatgatgccg 5340
ttatcgcact tgaagatctc aacagcggat ttaagaattc tcgggtgaaa gtcgaaaaac 5400
aggtgtatca aaaattcgaa aagatgctga tcgacaagct caattatatg gttgataaaa 5460
agagcaaccc atgcgccacg ggggggtgcg ttaagggtta tcagattacg aacaaatttg 5520
aatccttcaa gtcaatgtcg acgcaaaatg ggtttatatt ctatataccg gcgtggctta 5580
catctaaaat agatcctagc actgggttcg tgaacctgct gaaaaccaag tacacttcaa 5640
tcgcagattc taaaaaattt ataagcagct tcgacagaat catgtatgtg cccgaggaag 5700
acctcttcga gtttgccctt gattacaaaa atttctcaag aacggatgca gactacataa 5760
agaagtggaa gctgtactct tatgggaacc ggattcggat attcagaaat ccgaaaaaaa 5820
acaatgtctt tgattgggag gaagtttgtc ttacctctgc ttacaaagag ctgttcaata 5880
aatatggcat taattaccag caaggtgata tccgggcgct cctttgcgaa cagtctgaca 5940
aagctttcta ttcttcattt atggcgctca tgtcattgat gctgcagatg aggaatagca 6000
ttacggggag gactgatgtt gactttctga tctcgccgt gaaaaattct gatggaatct 6060
tctacgattc caggaattat gaggcccagg aaaatgctat ccttcccaag aacgcagacg 6120
caaatggcgc gtacaatata gctcgcaagg ttttgtgggc tataggccaa ttcaagaaag 6180
ccgaagacga aaagctggac aaagttaaga ttgctatata taacaaagag tggcttgagt 6240
atgcgcaaac atctgttaaa cacaaacgcc ccgcggctac aaagaaggct ggccaggcca 6300
agaagaagaa gtgagtcgac cgatcgttca aacatttggc aataaagttt cttaagattg 6360
aatcctgttg ccggtcttgc gatgattata atataatttc tgttgaatta cgttaagcat 6420
gtaataatta acatgtaatg catgacgtta tttatgagat gggtttttat gattagagtc 6480
ccgcaattat acatttaata cgcgatagaa aacaaaatat agcgcgcaaa ctaggataaa 6540
ttatcgcgcg cgggtgtcatc tatgttacta gatcgatccc gggatatcgc ggccgcgtcg 6600
ttcggctgcg gcgagcggta tcagctcact caaaggcggc aatacggtta tccacagaat 6660
caggggataa cgcaggaaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta 6720
aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa 6780
atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc 6840

```

ccccctggaag ctccctcgtg cgctctcctg ttccgaccct gccgcttacc ggatacctgt 6900
ccgccttttct cccttcggga agcgtggcgc tttctcatag ctcacgctgt aggtatctca 6960
gttcggtgta ggctggtcgc tccaagctgg gctgtgtgca cgaaccccc gttcagccccg 7020
accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccgtaaga cagcacttat 7080
cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta 7140
cagagttctt gaagtgggtg cctaactacg gctacactag aaggacagta tttggtatct 7200
gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac 7260
aaaccaccgc tggtagcggg ggtttttttg tttgcaagca gcagattacg cgcagaaaaa 7320
aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaacgaaa 7380
actcacgtta agggattttg gtcattgagat tatcaaaaag gatcttcacc tagatccttt 7440
taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact tggctctgaca 7500
gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctatct cgttcattca 7560
tagttgcctg actccccgc gtgtagataa ctacgatacg ggagggtta ccatctggcc 7620
ccagtgtgc aatgataccg cgagaccac gtcaccggc tccagattta tcagcaataa 7680
accagccagc cggaagggcc gagcgagaa gtggtcctgc aactttatcc gcctccatcc 7740
agtctattaa ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca 7800
acgttggtgc cattgtaca ggcatcgtg gtgcacgctc gtcgtttggt atggcttcat 7860
tcagctccgg ttcccaacga tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag 7920
cggtagctc cttcggtcct ccgatcgtt tcagaagtaa gttggccgca gtgttatcac 7980
tcatggttat ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt 8040
ctgtgactgg tgagtactca accaagtcatt tctgagaata gtgtatgcgg cgaccgagtt 8100
gctcttgccc ggctcaata cgggataata ccgcgccaca tagcagaact taaaagtgc 8160
tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat 8220
ccagttcgat gtaaccact cgtgcacca actgatcttc agcatctttt actttcacca 8280
gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccg aaaaaaggga ataaggcgca 8340
cacggaatg ttgaatactc atactcttcc ttttcaata ttattgaagc atttatcagg 8400
gttattgtct catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg 8460
ttccgcgcac atttccccga aaagtgccac ctgacgcgcc ctgtagcggc acgtctaatt 8520
cgggggatct ggatttttagt actggatttt ggttttagga attagaaatt ttattgatag 8580
aagtatttta caaatacaaa tacatactaa gggtttctta tatgctcaac acatgagcga 8640
aaccctatag gaaccctaatt tcccttatct gggaactact cacacattat tatggagaaa 8700
ctcgagcttg tcgatcgaca tgatcaggga gctctagatt atttgtatag ttcatccatg 8760
cccattacgt cggtaaatgc cttctgccac tcttgaagt taagtccggt cttggaatgt 8820
ttcaactcag tcttacggaa cacgtacatg ggttggttct taaggtagtt agcggccatt 8880
ggtttagcga atgtgtaggt agtcctggct gtagagcgat atctcttgcc attgcctgtg 8940
gtgtaagacc atttgaaggt actaatgatg gtcttgctgt tagggtaggt tttcttgac 9000
cggcaccaat cagcggcagt taaggagttg gtcattgacg gtccatcagc aggaaagcct 9060
gtccccctca cttgggcttc tcctttgatg tggtccctt cgtaagtgtg acggtagttg 9120
acggtgagcg aagcaccgtc ctcaaaactgc attgtcctgt ggacttggtg tccggagcca 9180

```

tcaaccatgg ctgcttggaa tggactcatt ccgtcagggt atggaaggtg ttgatggaat 9240
 ccgtagccaa tgtgtggcac cagaatccat ggagaaaact gaagatcacc tttggtgctc 9300
 ttgaggttca gctcttcgta tccgtcatta gggttccag tgccttgtcc gaccatatcg 9360
 aagtcaacgc cgttgatgga accgaagatg tgaagctcat gtgtggctgg aagcgaagcc 9420
 atgttatctt cttctccttt actcacggag gacgccatgg tggcgggatc gcgccctatc 9480
 gttcgtaaat ggtgaaaatt ttcagaaaat tgcttttgc ttaaaagaaa tgattttaat 9540
 tgctgcaata gaagtagaat gcttgattgc ttgagattcg tttgttttgt atatgttgtg 9600
 ttgagaggat cctcaagctt cgacctgcag aagtaacacc aaacaacagg gtgagcatcg 9660
 acaaaagaaa cagtaccaag caaataaata gcgtatgaag gcagggctaa aaaaatccac 9720
 atatagctgc tgcatatgcc atcatccaag tatatcaaga tcaaaataat tataaaacat 9780
 acttgtttat tataatagat aggtactcaa ggtagagca tatgaataga tgctgcatat 9840
 gccatcatgt atatgcatca gtaaaaccca catcaacatg tatacctatc ctagatcgat 9900
 atttccatcc atcttaaact cgtaactatg aagatgtatg acacacacat acagttccaa 9960
 aattaataaa tacaccaggt agtttgaaac agtattctac tccgatctag aacgaatgaa 10020
 cgaccgcca accacaccac atcatcaca ccaagcgaac aaaagcatct ctgtatatgc 10080
 atcagtaaaa cccgcatcaa catgtatacc taccctagat cgatatttcc atccatcatc 10140
 ttcaattcgt aactatgaat atgtatggca cacacataca gatccaaaat taataaatcc 10200
 accaggtagt ttgaaacaga attctactcc gatctagaac gaccgcccaa ccagaccaca 10260
 tcatacaaac caagacaaaa aaaagcatga aaagatgacc cgacaaacaa gtgcacggca 10320
 tatattgaaa taaaggaaaa gggcaaacca aaccctatgc aacgaaacaa aaaaaatcat 10380
 gaaatcgatc cgtctgcgg aacggctaga gccatcccag gattcccaa agagaaacac 10440
 tggcaagtta gcaatcagaa cgtgtctgac gtacaggctc catccgtgta cgaacgctag 10500
 cagcacggat ctaacacaaa cacggatcta acacaaacat gaacagaagt agaactaccg 10560
 ggccctaacc atggaccgga acgccgatct agagaaggta gagagggggg gggaggacga 10620
 gcggcgctacc ttgaagcgga ggtgccgacg ggtggatttg ggggagatcc actagttcta 10680
 gagcgccgc caccgcggtg gaattctcga ggtcctctcc aatgaaatg aacttcctta 10740
 tatagaggaa gggctctgcg aaggatagt ggattgtgcg tcatccctta cgtcagtgga 10800
 gatatcacat caatccactt gctttgaaga cgtggttgga acgtcttctt tttccacgat 10860
 gtcctcgtg ggtgggggtc catctttggg accactgtcg gcagaggcat cttgaacgat 10920
 agcctttcct ttatcgcaat gatggcattt gtaggtgcc a cttcctttt ctactgtcct 10980
 tttgatcaag tgaccgatag ctgggcaatg gaatccgagg aggtttcccg atattaccct 11040
 ttgttgaaaa gtctcaatag ccttttggtc ttctgagact gtatctttga tattcttgga 11100
 gtagacgaga gtgtcgtgct ccaccatgtt atcacatcaa ttcacttgct ttgaagacgt 11160
 ggttggaacg tcttcttttt ccacgatgct cctcgtgggt ggggggtccat ctttgggacc 11220
 actgtcggca gaggcattct gaacgatagc ctttccttta tcgcaatgat ggcatttgta 11280
 ggtgccacct tccttttcta ctgtcctttt gatcaagtga cagatagctg ggcaatggaa 11340
 tccgaggagg tttcccgata ttaccctttg ttgaaaagtc tcaatagccc tttggtcttc 11400
 tgagacctgc aggc 11414
 <210> 275

<211> 11414

<212> DNA

<213> 人工序列

<220>

<223> pGEP488 expression plasmid

<400> 275

```

cgatcttttcg gttgtcgaag agctcaaaga aataattatc cagaaggatc atgaaatcta 60
caagggtgtac ggctcaagcg agaagctctt tgatgctgac ttcgtgttgg agaagtctct 120
taaaaaaaaaac gacgcagtcg tcgcgataat gaaagatttg ctggattcag tgaaatcctt 180
cgagaattat atcaaaagcct tcttcggcga ggggaaggag acaaacaggg atgagtcctt 240
ctatggagac ttcgttcttg cttacgacat cttctttaag gtcgaccaca tctatgacgc 300
aattcggaac tatgtgacgc agaagccgta ttcgaaagat aagttcaagc tctattttcca 360
aaaccctcaa tttatgctg ggtgggataa agacgtagag accgatcgcc gggcaacaat 420
tttgcggtac ggggtctaaat attacctcgc tataatggat aagaaatacg ctaaattgtct 480
ccagaaaatt gacaaagatg acgtcaacgg caattatgaa aaaatcaatt ataaactcct 540
tcctggccca aataaaatgc tcccgaagggt gtttttttcc aaaaagtga tggcctatta 600
taatccatca gaggatattc agaaaatcta taaaaatggg acctttaaga aggggtgacat 660
gtttaacctg aacgattgcc acaagcttat agattttttc aaagactcta ttagccgcta 720
tcccaaatgg tctaattgctt atgatttcaa cttctctgaa actgaaaagt acaagatat 780
tgcaggattc taccgcgaag ttgaagaaca aggttataag gtttcctttg agtctgcgtc 840
caagaaagag gtcgataagt tggtcgaaga agggaaattg tatatgtttc aaattttaca 900
taaagacttt tccgacaagt cccatggtac acctaactctg cataccatgt acttcaaact 960
gctgtttcgat gagaataatc acggtcagat tcgcctgagc ggaggggagg aactcttcat 1020
gaggagagca tcgttgaaaa aagaggagct cgtcgtgcat ccggctaaca gcccatttgc 1080
taacaagaat ccgataatc caaagaagac tactaccctc tcctatgacg tctataagga 1140
taagagattc tctgaggacc agtacgagtt gcacatccct attgagataa ataaatgccc 1200
taagaacatc tttaaaatca atactgaggt cagagtccctg cttaaagcac acgacaaccc 1260
gtatgtgatc gggattgata ggggtgaaag gaacttgctt tatattgttg ttgtcgatgg 1320
aaaaggtaat atagtggaac aatactctct gaatgaaatt atcaacaact tcaatggcat 1380
taggatcaag accgactatc attctctgtt ggacaagaaa gagaaagagc gcttcgaggc 1440
acggcaaaac tggacgtcta ttgagaacat caaggagctt aaggctggtt acattttctca 1500
ggttgtgcac aaaaatttgcg aactggtcga gaaatatgat gccgttatcg cacttgaaga 1560
tctcaacagc ggatttaaga attctcgggt gaaagtcgaa aaacaggtgt atcaaaaatt 1620
cgaaaagatg ctgatcgaca agctcaatta tatggttgat aaaaagagca acccatgcgc 1680
cacgggggggt gcgcttaagg gctatcagat tacgaacaaa tttgaatcct tcaagtcaat 1740
gtcgacgcaa aatgggttta tattctatat accggcgttg cttacatcta aaatagatcc 1800
tagcactggg ttcgtgaacc tgctgaaaac caagtacact tcaatcgagc attctaaaaa 1860
atttataagc agcttcgaca gaatcatgta tgtgcccag gaagacctct tcgagtttgc 1920
ccttgattac aaaaatttct caagaacgga tgcagactac ataaagaagt ggaagctgta 1980

```

ctcttatggg aaccggattc ggatattcag aaatccgaaa aaaaacaatg tcttttgattg 2040
ggaggaagtt tgtcttacct ctgcttacia agagctgttc aataaatatg gcattaatta 2100
ccagcaaggt gatatccggg cgctcctttg cgaacagtct gacaaagctt tctattcttc 2160
atttatggcg ctcatgtcat tgatgctgca gatgaggaat agcattacgg ggaggactga 2220
tgttgacttt ctgatctcgc ccgtgaaaaa ttctgatgga atcttctacg attccaggaa 2280
ttatgaggcc caggaaaaatg ctatccttcc caagaacgca gacgcaaatg gcgcgtacaa 2340
tatagctcgc aaggttttgt gggctatagg ccaattcaag aaagccgaag acgaaaagct 2400
ggacaaaagt aagattgcta tatctaacia agagtggcct gagtatgcgc aaacatctgt 2460
taaacacaaa cgccccgcgg ctacaaaaga ggctggccag gcaaagaaga agaagtgagt 2520
cgaccgatcg ttcaaacatt tggcaataaa gtttcttaag attgaatcct gttgccggctc 2580
ttgcgatgat tatcatataa tttctgttga attacgttaa gcatgtaata attaacatgt 2640
aatgcatgac gttatttatg agatgggttt ttatgattag agtcccga tttatacattt 2700
aatacgcat agaaaaacaaa atatagcgcg caaactagga taaattatcg cgcgcggtgt 2760
catctatgtt actagatcga tccccggata tcgcggccgc gtcgttcggc tgcggcgagc 2820
ggtatcagct cactcaaagg cggtaatagc gttatccaca gaatcagggg ataacgcagg 2880
aaagaacatg tgagcaaaaag gccagcaaaa ggccaggaa cgtaaaaagg ccgcgttgct 2940
ggcggttttt cataggctcc gccccctga cgagcatcac aaaaatcgac gctcaagtca 3000
gaggtggcga aacccgacag gactataaag ataccaggcg tttccccctg gaagctccct 3060
cgtgcgctct cctgttccga ccctgccgct taccggatac ctgtccgcct ttctcccttc 3120
gggaagcgtg gcgctttctc atagctcacg ctgtaggtat ctcagttcgg ttaggtcgt 3180
tcgctccaag ctgggctgtg tgcacgaacc cccggttcag cccgaccgct gcgccttate 3240
cggtaaactat cgtcttgagt ccaaccgggt aagacacgac ttatcgccac tggcagcagc 3300
cactggtaac aggattagca gagcgaggta ttaggcgggt gctacagagt tcttgaagt 3360
gtggcctaac tacggctaca ctagaaggac agtatttgggt atctgcgctc tgctgaagcc 3420
agttaccttc ggaaaaagag ttggtagctc ttgatccggc aaacaaacca ccgctggtag 3480
cggtgggtttt tttgtttgca agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga 3540
tcctttgatc ttttctacgg ggtctgacgc tcagtggaa cgaactcac gtttaaggat 3600
tttggtcatg agattatcaa aaaggatctt cacctagatc cttttaaatt aaaaatgaag 3660
ttttaaatca atctaaagta tatatgagta aacttggctt gacagttacc aatgcttaat 3720
cagtgaggca cctatctcag cgatctgtct atttcgttca tccatagttg cctgactccc 3780
cgctcgtag ataaactacga tacgggaggg cttaccatct ggccccagtg ctgcaatgat 3840
accgcgagac ccacgctcac cggtccaga tttatcgca ataaaccagc cagccggaag 3900
ggccgagcgc agaagtggct ctgcaacttt atccgcctcc atccagtcta ttaattgttg 3960
ccgggaagct agagtaagta gttcgccagt taatagtttg cgcaacgttg ttgccattgc 4020
tacaggcatc gtggtgtcac gctcgtcgtt tggatatggct tcattcagct ccggttccca 4080
acgatcaagg cgagttacat gatcccccat gttgtgcaaa aaagcggtta gctccttcgg 4140
tcctccgac gttgtcagaa gtaagttggc cgcagtgtta tcactcatgg ttatggcagc 4200
actgcataat tctcttactg tcatgccatc cgtaagatgc ttttctgtga ctggtgagta 4260
ctcaaccaag tcattctgag aatagtgtat gcggcgaccg agttgctctt gcccggcgtc 4320

aatacgggat aataccgcgc cacatagcag aactttaaaa gtgctcatca ttgaaaacg 4380
ttcttcgggg cgaaaactct caaggatctt accgctgttg agatccagtt cgatgtaacc 4440
cactcgtgca cccaactgat cttcagcatc ttttactttc accagcgttt ctgggtgagc 4500
aaaaacagga aggcaaaatg ccgcaaaaaa gggaataagg gcgacacgga aatgttgaat 4560
actcatactc ttcttttttc aatattattg aagcatttat cagggttatt gtctcatgag 4620
cggatacata tttgaatgta tttagaaaaa taaacaaata ggggttccgc gcacatttcc 4680
ccgaaaagtg ccacctgacg cgccctgtag cggcacgtct aattcggggg atctggattt 4740
tagtactgga ttttggtttt aggaattaga aattttattg atagaagtat ttacaaaata 4800
caaatacata ctaagggttt cttatatgct caacacatga gcgaaaccct ataggaaccc 4860
taattccctt atctgggaac tactcacaca ttattatgga gaaactcgag cttgtcgatc 4920
gacatgatca gggagctcta gattatttgt atagtccatc catgccatt acgtcggtaa 4980
atgccttctg ccactccttg aagttaagtt cggctcttga atgtttcaac tcagtcttac 5040
ggaacacgta catgggttgg ttcttaaggt agtttagcgc cattggttta gcgaatgtgt 5100
aggtagtcct ggctgtagag cgatatctct tgccattgcc tgttgtgtaa gaccatttga 5160
aggtaactaat gatggtcttg tcgttagggg aggttttctt ggaccggcac caatcagcgg 5220
cagttaagga gttggtcatg acaggtccat cagcaggaaa gcctgtcccc ttcacttggg 5280
cttctccttt gatgtggctc cttcgttaag tgtaacggtg gttgacggtg agcgaagcac 5340
cgctctcaaa ctgcattgtc ctgtggactt ggtatccgga gccatcaacc atggctgctt 5400
ggaatggact cattccgtca gggatggaa ggtattgatg gaatccgtag ccaatgtgtg 5460
gcaccagaat ccatggagaa aactgaagat cacctttggt gctcttgagg ttcagctctt 5520
cgtatccgtc attagggttc ccagtgcctt gtccgacct atcgaagtca acgccgttga 5580
tggaaccgaa gatgtgaagc tcatgtgtgg ctggaagcga agccatgtta tcttcttctc 5640
ctttactcac ggaggacgcc atggtggcgg gatcgcgccc tatcgttcgt aaatggtgaa 5700
aattttcaga aaattgcttt tgctttaaaa gaaatgattt aaattgctgc aatagaagta 5760
gaatgcttga ttgcttgaga ttcttttgtt ttgtatatgt tgtgttgaga ggatcctcaa 5820
gcttcgacct gcagaagtaa caccaaaca cagggtgagc atcgacaaaa gaaacagtac 5880
caagcaaaata aatagcgtat gaaggcagg ctaaaaaaat ccacatatag ctgctgcata 5940
tgccatcatc caagtatatc aagatcaaaa taattataaa acatacttgt ttattataat 6000
agataggtag tcaaggtag agcatatgaa tagatgctgc atatgccatc atgtatatgc 6060
atcagtaaaa cccacatcaa catgtatacc tatcctagat cgatatttcc atccatctta 6120
aactcgtaac tatgaagatg tatgacacac acatacagtt ccaaaattaa taaatacacc 6180
aggtagtttg aaacagtatt ctactccgat ctagaacgaa tgaacgaccg cccaaccaca 6240
ccacatcatc acaaccaagc gaacaaaagc atctctgtat atgcatcagt aaaaccgca 6300
tcaacatgta tacctatcct agatcgatat ttccatccat catcttcaat tcgtaactat 6360
gaatatgtat ggcacacaca tacagatcca aaattaataa atccaccagg tagtttgaaa 6420
cagaattcta ctccgatcta gaacgaccgc ccaaccagac cacatcatca caaccaagac 6480
aaaaaaaaagc atgaaaagat gacccgacaa acaagtgcac ggcatatatt gaaataaagg 6540
aaaagggcaa accaaaccct atgcaacgaa acaaaaaaaaa tcatgaaatc gatcccgctt 6600
gcggaacggc tagagccatc ccaggattcc ccaaagagaa aactggcaa gttagcaatc 6660

agaacgtgtc tgacgtacag gtcgcatccg tgtacgaacg ctagcagcac ggatctaaca 6720
caaacacgga tctaacacaa acatgaacag aagtagaact accgggccct aaccatggac 6780
cggaacgccg atctagagaa ggtagagagg gggggggagg acgagcggcg taccttgaag 6840
cggaggtgcc gacgggtgga tttgggggag atccactagt tctagagcgg ccgccaccgc 6900
ggtggaattc tcgaggtcct ctccaaatga aatgaacttc cttatataga ggaagggtct 6960
tgcgaaggat agtgggattg tgcgtcatcc cttacgtcag tggagatata acatcaatcc 7020
acttgctttg aagacgtggt tggaacgtct tctttttcca cgatgctcct cgtgggtggg 7080
ggtccatctt tgggaccact gtcggcagag gcattctgaa cgatagcctt tccttttatcg 7140
caatgatggc atttgtaggt gccaccttcc ttttctactg tccttttgat caagtgaccg 7200
atagctgggc aatggaatcc gaggaggttt cccgatatta ccctttgttg aaaagtctca 7260
atagcccttt ggtcttctga gactgtatct ttgatattct tggagtagac gagagtgtcg 7320
tgctccacca tgttatcaca tcaattcact tgctttgaag acgtggttgg aacgtcttct 7380
ttttccacga tgctcctcgt ggggtgggggt ccatctttgg gaccactgtc ggcagaggca 7440
tcttgaacga tagcctttcc tttatcgcaa tgatggcatt tgtaggtgcc accttccctt 7500
tctactgtcc ttttgatcaa gtgacagata gctgggcaat ggaatccgag gaggtttccc 7560
gatattaccc tttgttgaag agtctcaata gccctttggt cttctgagac ttgcaggcaa 7620
gcaagcatga atgcctgggg gagaagaact cgagagggaa ttgcagatca tgaggcagat 7680
ggctatTTTT gtgtcacata tgcgcaaaaa gagaggctat atttgtgtcc ctaggttctt 7740
cgttgtattg cagtttccat atcaatctga cttggtcgca tgagaaattg atggttaaat 7800
aatttgaatc tctcatgtag tatcaactat tagatattat tttcaccaa tatatttcca 7860
tcggagaaga agaggctaca gaggaagcag aagagagggg tgggagaatt ttacacttt 7920
tgtacacca cttaaacagc aaaatccgta tgaaaacagg cccacaaaa caatgccacg 7980
ataacaatcc gtagaaacaa aagcttcatt taacagcggc gcaacaaagc acgcttatcc 8040
atggtagttg tagtccgtat gcgatccaaa gatcacgatt cacgcgtgac ggacggacga 8100
cgcgtgccac accacaacta acggcatcca tggtagttgt agtccgtatg cgatccaaag 8160
atcacgattc acgcgtgacg gacggacgac gcgcgccaca ccacaactaa cagcgtgagc 8220
cagcgtccaa actccggatg gcaacgggga cgaaaccgt cgggtagtca ctgccccaac 8280
ccgtccccgc aaccttcac ccaaaccgt ccccgtttc ggtcgcgggt ttcagttttc 8340
taccagacc gtccccatcg ggtttttcat ccccgtcggg aaatccgaac ccgccagcat 8400
ttcagacca agccaaagt gcagcagcaa catgaataaa aaacaaccg tttcaacacc 8460
aagataaaac aaaacattat aatttagaca acatttcaca cgtataacaa taacatatag 8520
ttctcacata taacaacacc atttcacaca taaaacaaca ccatttggga taaaaatatg 8580
ggctatatca ggccattttt atgggccata ttgagttttc gtgggtttca caggtaccgg 8640
atttgtagaa tgctgaaccg ggtttgaacc gtaaaatccg cgggtattga atttgacca 8700
atcccgtcgt cccctggtgg ggtaaaaaca ccatcttgag tccaaacggc caccaacca 8760
actccgacgg caacaaacaa acggcgttgc tttgtctctc ggtatctccg tgaccgtca 8820
atctcccggc tgtttccccg gaattgcgtg gactctctca tccacacgca aaccgcctct 8880
ccctcctctc tcgtcctatc cgccccgtg ccgtagcctc acgggactct tcttctctcc 8940
ttgctataaa atccccgcc cctcccgctt cctctccaca catccaaact ctcaatcgca 9000

ccgagaaaaa tctcctagcg atcgaagcga agcctctccc gatcctctca aggtacgccc 9060
gtttcccgtc gatcctcctc cttccgttcg tgttctgtag ccgatcgatt cgattccctt 9120
acacccggtc gtgttctctc gtggatcgat cgattgtttg ttgctagaag gaactcgtag 9180
atctggcggt tatgaactgt gattcgggtt agtccagatc gattcaggtc ggctcgtcgtt 9240
gagcctctcg gctatgtctg gattatcgtg tagatctgct gggtcagttg attatgttct 9300
tctaggagta atttcgttgg gtcagcgcga tttctgctta atctatgctg cttattgcgc 9360
ctgtacctat ctactaagct atgtgcacct gtaattttgc tagattattc gttcatcctc 9420
gtagttgggt tgtcacagta atccgtatgg gttctgacga tggtattgtt ggtcatacct 9480
aggcttctcc agatttttatt ttgttaaaat tggatagatc tgctactgat agttgatgat 9540
ggaatttgggt gctgaatcta tgctatttat tgcgcctata cctgatctat cgggctatgt 9600
acggctgtag ttactggat tattcgttca tctcggtag ttggttcatc gtttgggttc 9660
tgacgataat attgttgatt atgcgtaggc ttctgcagat tggtgttaaa attggataca 9720
tcggttactg atggttgatg atagatttgt gctgaaccta tctgtttatt gtcctatac 9780
ctgatctata gggctatgta tgcctgtaat ttaccagatt attcgttcat cctcgtagtt 9840
ggttcatctc tataattcgt atgggttctt atgatgttat cggttgattat gcctagtctt 9900
atacagatta ttgtgtcaag attgaatata cctgctactg atcggtgata atttggttag 9960
tagtttgcaa tctgctagga acacgttacc actgtaatct gtaaacaatgg tttgccagag 10020
tagtttggtc tactactctt gatatggttg ctgattttag tcgcctcctt ttggatcatg 10080
tattgatgtc cttgcagatt tccgtgtact taccgccgtt tttgtgtact tcgtgttaac 10140
aggtcgggta ccgaagcaaa catggcatct agcatggcac caaagaaaaa aaggaaagtt 10200
tccaaacttg aaaaatttac aaactgctac tccctttcca agacgcttag gtttaaagcg 10260
atccccgttg gcaagacca agagaatata gataacaaaa gacttctgggt cgaagatgaa 10320
aaaagggccg aagactacaa ggggtcaag aagttgctcg atcgctatta tctttccttt 10380
atcaacgatg tgcttcattc aatcaaactg aagaacttga ataactacat tagccttttc 10440
agaaagaaaa cgaggactga aaaggagaac aaggaaactg agaacttga aataaacctt 10500
cgcaaagaaa ttgcaaaagc cttcaagggg aacgaaggat ataaatctct tttcaaaaaa 10560
gacattatag aaacaatttt gcctgagttt cttgacgaca aggatgaaat tgcgctcgtc 10620
aatagcttta acgattttac aactgccttc acagggttct tcgacaatag ggagaatatg 10680
tttagcgagg aggcaaaaag cacatccatc gcattcagat gcatcaatga aaatcttacc 10740
cggtacatat cgaatatgga catatttgaa aaagtggatg caatattcga taagcacgaa 10800
gtccaggaga taaaggaaaa gatactgaat agcgactatg atgtcgaaga ttttttcgaa 10860
ggtgagttct tcaactttgt cctgactcaa gaaggcattg atgtctataa tgcaataatt 10920
ggaggttttg tgactgagtc tggcgagaag ataaagggtt tgaacgagta tatcaatctc 10980
tacaaccaga agactaagca aaagttgcct aaattttaa acgctttacaa gcaagttttg 11040
agcgaccggg aaagcctttc cttttacggt gaaggataca cgagcgatga agaagtcctc 11100
gaagtcttcc gcaacacact caacaagaac tcagaaatct tttcctcaat taaaaaattg 11160
gagaagcttt tcaagaactt cgatgaatac tcttcggcgg ggatttttgt gaagaacggc 11220
ccggcaattt ccacaatata taaagacatt ttcggagaat ggaacgtgat aagagacaag 11280
tggaatgcgg agtatgatga catacacctg aagaagaagg cagttgtgac tgaaaaatac 11340

gaagatgaca ggagaaaaag ctttaaaaag atcgggtcct tttcactgga acagctgcag 11400
gagtatgccg acgc 11414

<210> 276

<211> 1572

<212> DNA

<213> 人工序列

<220>

<223> VPR transcriptional activation domain

<400> 276

gacgccctgg acgacttcga cctcgacatg ctgggctccg acgccctcga tgatttcgac 60
ctcgatatgc tcggcagcga cgcgctcgat gacttcgacc tcgatatgct ggggagcgac 120
gccctcgacg attttgacct cgatatgctg atcaactccc gctccagcgg cagcccgaag 180
aagaagcgca aagtgggctc gcagtacctg cccgacaccg acgacaggca caggatcgag 240
gagaagcgca agaggacgta cgagaccttc aagtccatca tgaagaagtc cccgttcagc 300
ggcccaacgg acccccgcgc gccgccgagg aggatcgccg tgccgtccag gtccagcgcg 360
tcgggtcccca agccggcccc gcagccctac ccgttcacgt ccagcctcag caccatcaac 420
tacgacgagt tccccaccat ggtgttcccg tccggccaga tctcccaggc cagcgcgctg 480
gccccgcgc ccccgcaggt gctgccccag gctccggccc ccgctccggc cccggccatg 540
gtctccgcgc tggcccaggc gcccgccccg gtgcccgtcc tcgcgccggg cccgcccgag 600
gcggctgccc cgccagcgcc gaagcccacg caggccggcg agggcaccct cagcgaggcg 660
ctcctgcagc tgcagttcga cgacgaggac ctcggcgccc tcctgggcaa ctcgaccgac 720
cccgccgtgt tcaccgacct ggcctccgtc gacaacagcg agttccagca gctgctgaac 780
cagggcatcc cgggtggcgcc gcacaccacg gagcccatgc tgatggagta cccggaggcg 840
atcacgcgcc tcgtcaccgg cgcccagagg ccccccggacc ccgccccggc cccgctcggc 900
gccccaggcc tgccgaacgg cctcctgagc ggcgacgagg acttctccag catcgcgac 960
atggacttct ccgccctcct ggggtcgggc tcgggcagcc gcgacagcag ggagggcag 1020
ttcctcccaa agcccagagg cggtccgcc atctcggacg tgttcgaggg caggaggtc 1080
tgccagccaa agcgcatcag gccgttccac ccgcccgggt ccccggtggg gaaccggccg 1140
ctccccgcca gcctggctcc aaccccgacc ggccccgtgc acgagccggg cggcagcctg 1200
acgcccgcgc cgggtgcccc gccgctcgac cccgcgccgg ccgtcaccac cgaggcctcc 1260
cacctcctgg aggaccccga cgaggagacc tcgcaggccg tgaaggccct gagggagatg 1320
gccgacaccg tcatcccca gaaggaggag gcggccatct gcggccagat ggacctgtcg 1380
caccgcccgc cgcgcgccca cctcgacgag ctgaccacga cctcgagtc catgaccgag 1440
gacctcaacc tggacagccc cctcacgccc gagctgaacg agatcctcga caccttcttg 1500
aacgacgagt gcctcctgca cgccatgcac atctccacgg gcctgagcat cttegcacac 1560
agcctcttct ga 1572

<210> 277

<211> 30

<212> DNA

<213> 人工序列

<220>

<223> 5xGS linker sequence

<400> 277

ggctcgggggt cgggggtcggg ctctgggctcg 30

<210> 278

<211> 4570

<212> DNA

<213> 人工序列

<220>

<223> pKWS20 plasmid

<400> 278

```

tcgcgcgttt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca 60
cagcttgtct gtaagcggat gccgggagca gacaagcccc tcagggcgcg tcagcgggtg 120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180
accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggcgcc 240
attcgccatt caggctgcgc aactgttggg aagggcgatc ggtgcgggcc tcttcgctat 300
tacgccagct ggcgaaaggg ggatgtgctg caaggcgatt aagttgggta acgccagggt 360
tttccagtc acgacgttgt aaaacgacgg ccagtgaatt cgcggctaca aagaaggctg 420
gccaggccaa gaagaagaag ggctcgggggt cgggggtcggg ctctgggctcg gacgccttg 480
acgacttcga cctcgacatg ctgggctccg acgccttcga tgatttcgac ctgatatgc 540
tcggcagcga cgcgctcgat gacttcgacc tcgatatgct ggggagcgac gccctcgacg 600
atittgacct cgatatgctg atcaactccc gctccagcgg cagcccgaag aagaagcgca 660
aagtgggctc gcagtacctg cccgacaccg acgacaggca caggatcgag gagaagcgca 720
agaggacgta cgagacctc aagtccatca tgaagaagtc cccgttcagc ggcccaacgg 780
acccccgccc gccgccgagg aggatcgccg tgccgtccag gtccagcgcg tcggtcccca 840
agccggcccc gcagcccctac ccgttcacgt ccagcctcag caccatcaac tacgacgagt 900
tccccaccat ggtgttcccg tccggccaga tctcccaggc cagcgcgctg gccccgcgc 960
ccccgcaggt gctgccccag gctccggccc ccgctccggc cccggccatg gtctccgcgc 1020
tggtcccaggc gccccccccg gtgcccgtcc tcgcgccggg cccgccgcag gcggtcgccc 1080
cgccagcgcc gaagcccacg caggccggcg agggcaccct cagcgaggcg ctctgcagc 1140
tgcagttcga cgacgaggac ctctggcgccc tcttgggcaa ctcgaccgac cccgccgtgt 1200
tcaccgacct ggccctccgtc gacaacagcg agttccagca gctgctgaac cagggcattc 1260
cggtggcgcc gcacaccacg gagcccatgc tgatggagta cccggaggcg atcacgcgcc 1320
tcgtcaccgg cgcacagagg cccccggacc ccgccccggc cccgctcggc gccccaggcc 1380
tgccgaacgg cctcctgagc ggcgacgagg acttctccag catcgcgga atggaattct 1440
ccgcccctct ggggtcgggc tcgggcagcc gcgacagcag ggagggcatt ttctcccaa 1500
agccccaggc cggctccgcc atctcgagc tggttcaggg cagggaggtc tgccagccaa 1560
agcgcacatc gccgttccac ccgccgggct ccccggtgggc gaaccggccg ctccccgcca 1620

```

gcctggctcc aacccccgacc ggccccgtgc acgagccggt cggcagcctg acgccccgcgc 1680
 cgggtgccccca gccgctcgac cccgcgcccgg ccgtcacccc cgaggcctcc cacctcctgg 1740
 aggacccccga cgaggagacc tcgcaggccg tgaaggccct gagggagatg gccgacaccg 1800
 tcatccccca gaaggaggag gcggccatct gcggccagat ggacctgtcg caccgcgcgc 1860
 cgcgcggcca cctcgacgag ctgaccacga ccctcgagtc catgaccgag gacctcaacc 1920
 tggacagccc cctcacgccg gagctgaacg agatcctcga caccttcctg aacgacgagt 1980
 gcctcctgca cgccatgcac atctccacgg gcctgagcat cttcgacacc agcctcttct 2040
 gagtcgaccg atcggtcaaa catttgcaa taaagtttct taagattgaa tcctgttgcc 2100
 ggtcttgca tgattatcat ataatttctg ttgaattacg ttaagcatgt aataattaac 2160
 atgtaatgca tgacgttatt tatgagatgg gtttttatga ttagagtccc gcaattatac 2220
 atttaatacg cgatagaaaa caaaatatag cgcgcaaact aggataaatt atcgcgcgcg 2280
 gtgtcatcta tgttactaga tcgatcccgg gatatcgcg ccgcgtcgtt aagcttggcg 2340
 taatcatggt catagctgtt tcctgtgtga aattgttate cgctcacaat tccacacaac 2400
 atacgagccg gaagcataaa gtgtaaagcc tggggtgcct aatgagttag ctaactcaca 2460
 ttaattgcgt tgcgctcact gcccgtttc cagtcgggaa acctgtcgtg ccagctgcat 2520
 taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta ttgggcgtc ttccgcttcc 2580
 tcgctcactg actcgtcgcg ctcggtcgtt cggctgcggc gagcggatc agctcactca 2640
 aaggcggtaa tacggttate cacagaatca ggggataacg caggaaagaa catgtgagca 2700
 aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt tgctggcgtt ttccatagg 2760
 ctccgcccc ctgacgagca tcacaaaaat cgacgtcaa gtcagaggtg gcgaaacccg 2820
 acaggactat aaagatacca ggcgtttccc cctggaagct ccctcgtgcg ctctcctgtt 2880
 ccgacctgc cgcttaccgg atacctgtcc gcctttctcc cttcggaag cgtggcgctt 2940
 tctcatagct cacgctgtag gtatctcagt tcggtgtagg tcgttcgctc caagctgggc 3000
 tgtgtgcacg aacccccgt tcagcccgac cgctgcgcct tatccggtaa ctatcgtctt 3060
 gagtccaacc cggtaagaca cgacttatcg ccactggcag cagccactgg taacaggatt 3120
 agcagagcga ggatgtagg cgggtctaca gagttcttga agtggtggcc taactacggc 3180
 tacactagaa gaacagtatt tggtatctgc gctctgctga agccagttac cttcgaaaa 3240
 agagttggta gctcttgatc cggcaaacia accaccgctg gtagcgggtg tttttttgtt 3300
 tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag aagatccttt gatcttttct 3360
 acggggtctg acgctcagt gaacgaaaac tcacgttaag ggatttttgt catgagatta 3420
 tcaaaaagga tcttcaccta gatcctttta aattaaaaat gaagtttta atcaatctaa 3480
 agtatatatg agtaaaactg gtctgacagt taccaatgct taatcagtga ggcacctatc 3540
 tcagcgatct gtctatttcg ttcattcata gttgctgac tccccgtcgt gtagataact 3600
 acgatacggg agggcttacc atctggcccc agtctgcaa tgataccgcg agaccacgc 3660
 tcaccggctc cagatttate agcaataaac cagccagccg gaagggccga gcgcagaagt 3720
 ggtcctgcaa ctttatccgc ctccatccag tctattaatt gttgccggga agctagagta 3780
 agtagttcgc cagttaatag tttgcgcaac gttgttgcca ttgctacagg catcgtggtg 3840
 tcacgctcgt cgtttggtat ggcttcattc agctccggtt cccaacgatc aaggcgagtt 3900
 acatgatccc ccatgttgtg caaaaaagcg gttagctcct tcggtcctcc gatcgttgct 3960

agaagtaagt tggccgcagt gttatcactc atggttatgg cagcactgca taattctctt 4020
 actgtcatgc catccgtaag atgcttttct gtgactgggt agtactcaac caagtcattc 4080
 tgagaatagt gtatgcggcg accgagttgc tcttgcccgg cgtcaatacg ggataatacc 4140
 gcgccacata gcagaacttt aaaagtgtc atcattggaa aacgttcttc ggggcgaaaa 4200
 ctctcaagga tcttaccgct gttgagatcc agttcgatgt aaccactcg tgcacccaac 4260
 tgatcttcag catcttttac tttcaccagc gtttctgggt gagcaaaaac aggaaggcaa 4320
 aatgccgcaa aaaagggaat aagggcgaca cggaatgtt gaatactcat actcttcctt 4380
 tttcaatatt attgaagcat ttatcagggt tattgtctca tgagcggata catatttgaa 4440
 tgtattttaga aaaaataaaca aataggggtt ccgcgcacat ttccccgaaa agtgccacct 4500
 gacgtctaag aaaccattat tatcatgaca ttaacctata aaaataggcg tatcacgagg 4560
 ccctttcgtc 4570

<210> 279

<211> 13012

<212> DNA

<213> 人工序列

<220>

<223> pGEP754 expression plasmid

<400> 279

agcatgaatg cctgggggag aagaactcga gaggaattg cagatcatga ggcagatggc 60
 tatttttgtg tcacatatgc gcaaaaagag aggctatatt tgtgtcccta ggttcttcgt 120
 tgtattgcag tttccatc atctgactt ggtcgcatga gaaattgatg gttaaataat 180
 ttgaatctct catgtagtat caactattag atattatatt caccaaata atttccatcg 240
 gagaagaaga ggctacagag gaagcagaag agaggggtgg gagaattttt acacttttgt 300
 acacccactt aaacagcaaa atccgtatga aaacaggccc accaaaacaa tgccacgata 360
 acaatccgta gaaacaaaag cttcatttaa cagcggcgca acaaagcacg cttatccatg 420
 gtagttgtag tccgtatgcg atccaaagat cagattcac gcgtgacgga cggacgacgc 480
 gtgccacacc acaactaacg gcacccatgg tagttgtagt ccgtatgca tccaaagatc 540
 acgattcacg cgtgacggac ggacgacgcg cgccacacca caactaacag cgtgagccag 600
 cgtccaaact ccgatggca acggggacga aaccgctcgg gtagtcactg cccaaacccg 660
 tccccgcaac cttcatccca aaccgctccc cgtttccggt cgcgggtttc agttttctac 720
 cagacccgtc cccatcgggt ttttcateccc cgtcgggaaa tccgaacccg ccagcatttc 780
 agcaccaagc caaagttgca gcagcaacat gaataaaaaa caaccgttt caacaccaag 840
 ataaaacaaa acattataat ttagacaaca tttcacacgt ataacaataa catatagttc 900
 tcacatataa caacaccatt tcacacataa aacaacacca tttgggataa aaatatgggc 960
 tatatcaggc catttttatg ggccatattg agttttcgtg ggtttcacag gtaccggatt 1020
 tgtagaatgc tgaaccgggt ttgaaccgta aaatccgcgg gtattgaatt tgaccaatc 1080
 ccgtcgtccc ctggtggggt aaaaacacca tcttgagtcc aaacggccac caaccaaact 1140
 ccgacggcaa caaacaacg gcgttgcttt gctcctcgggt atctccgtga ccgtcaatc 1200
 tcccggctgt ttccccgaa ttgcgtggac tctctcatcc acacgcaaac cgcctctccc 1260

```

tcctctctcg tcctatccgc cccggtgccg tagcctcacg ggactcttct tcctcccttg 1320
ctataaaatc cccgccccct cccgtctcct ctccacacat ccaaactctc aatcgaccg 1380
agaaaaatct cctagcgatc gaagcgaagc ctctcccgat cctctcaagg tacgcccgtt 1440
tcccgtcgat cctcctcctt ccgttcgtgt tctgtagccg atcgattcga ttcccttaca 1500
cccgttcgtg ttctctcgtg gatcgatcga ttgtttgttg ctagaaggaa ctcgtagatc 1560
tggcgtttat gaactgtgat tcgggtagt ccagatcgat tcaggtcggt cgtcgttgag 1620
cctctcggct atgtctggat tatcgtgtag atctgctggt tcagttgatt atgttcttct 1680
aggagtaatt tcgttgggtc agcgcgattt ctgcttaatc tatgctgctt attgcgcctg 1740
tacctatcta ctaagctatg tgcacctgta attttgctag attattcggt catcctcgta 1800
gttggtttgt cacagtaatc cgtatgggtt ctgacgatgt tattgttggt catacctagg 1860
cttctccaga ttttattttg ttaaaattgg atagatctgc tactgatagt tgatgatgga 1920
at ttggtgct gaatctatgc tatttattgc gcctatacct gatctatcgg gctatgtacg 1980
gctgtagttt actggattat tcgttcaccc tcggtagttg gttcatcggt tgggttctga 2040
cgataatatt gttgattatg cgtaggttc tgcagattgt tgttaaaatt ggatacatcg 2100
gttactgatg gttgatgata gatttgtgct gaacctatct gtttattgct cctatacctg 2160
atctataggg ctatgtatgc ctgtaattta ccagattatt cgttcaccc cgtagttggt 2220
tcactcttat aattcgtatg gggttcttat atgttatcgt tgattatgcc tagtcttata 2280
cagattattg tgtcaagatt gaatatacct gctactgatc ggtgataatt tggttagtag 2340
tttgcaatct gctaggaaca cgttaccact gtaatctgta aacatggttt gccagagtag 2400
tttgttctac tactcttgat atggttgctg attttagtcg cctccttttg gatcatgtat 2460
tgatgtcctt gcagatttcc gtgtacttac cccggtttt gtgtacttcg tgttaacagg 2520
tcgggtaccg aagcaaacat ggcatctagc atggcaccaa agaaaaaag gaaagtttcc 2580
aaacttgaaa aatttacaaa ctgctactcc cttccaaga cgcttaggtt taaagcgatc 2640
cccgttggca agaccaaga gaatatcgat aacaaaagac ttctggtcga agatgaaaaa 2700
agggccgaag actacaaggg ggtcaagaag ttgctcgatc gctattatct ttcctttatc 2760
aacgatgtgc ttcatcatt caaactgaag aacttgaata actacattag ctttttcaga 2820
aagaaaacga ggactgaaaa ggagaacaag gaacttgaga atcttgaaat aaaccttcgc 2880
aaagaaattg caaaagcctt caaggggaac gaaggatata aatctctttt caaaaaagac 2940
attatagaaa caattttgcc tgagtttctt gacgacaagg atgaaattgc gctcgtcaat 3000
agctttaacg gatttacaac tgccttcaca gggttcttcg acaataggga gaatatgttt 3060
agcgaggagg caaaaagcac atccatcgca ttcagatgca tcaatgaaaa tcttaccggg 3120
tacatatcga atatggacat atttgaaaaa gtggatgcaa tattcgataa gcacgaagtc 3180
caggagataa aggaaaagat actgaatagc gactatgatg tcgaagattt tttcgaaggt 3240
gagttcttca actttgtcct gactcaagaa ggcattgatg tctataatgc aataattgga 3300
ggttttgtga ctgagtctgg cgagaagata aagggttga acgagtatat caatctctac 3360
aaccagaaga ctaagcaaaa gttgcctaaa tttaaaccgc tttacaagca agttttgagc 3420
gaccgggaaa gcctttcctt ttacggtgaa ggatacacga gcgatgaaga agtcctcgaa 3480
gtcttccgca acacactcaa caagaactca gaaatctttt cctcaattaa aaaattggag 3540
aagcttttca agaacttcga tgaatactct tcggcgggga tttttgtgaa gaacggcccg 3600

```

gcaatttcca caatatctaa agacattttc ggagaatgga acgtgataag agacaagtgg 3660
 aatgcggagt atgatgacat acacctgaag aagaaggcag ttgtgactga aaaatacgaa 3720
 gatgacagga gaaaaagcct taaaaagatc gggtcctttt cactggaaca gctgcaggag 3780
 tatgccgacg ccgatctttc ggttgtcgaa aagctcaaag aaataattat ccagaaggctc 3840
 gatgaaatct acaagggtga cggctcaagc gagaagctct ttgatgctga cttcgtgttg 3900
 gagaagtctc ttaaaaaaaaa cgacgcagtc gtgcgcataa tgaaagattt gctggattca 3960
 gtgaaatcct tcgagaatta tatcaaagcc ttcttcggcg aggggaagga gacaaacagg 4020
 gatgagtcct tctatggaga cttcgtttctg gcttacgaca tccttcttaa ggtcgaccac 4080
 atctatgacg caattcggaa ctatgtgacg cagaagccgt attcgaaaga taagttcaag 4140
 ctctatttcc aaaaccctca atttatgggt ggggtgggata aagacaaaga gaccgattac 4200
 cgggcaacaa ttttgcggtta cgggtctaaa tattacctg ctataatgga taagaaatac 4260
 gctaaatgtc tccagaaaaat tgacaaagat gacgtcaacg gcaattatga aaaaatcaat 4320
 tataaactcc ttcttgcccc aaataaaatg ctcccgaagg tgtttttttc caaaaagtgg 4380
 atggcctatt ataatccatc agaggatatt cagaaaatct ataaaaatgg gacctttaag 4440
 aagggtgaca tgtttaacct gaacgattgc cacaagctta tagatttttt caaagactct 4500
 attagccgct atcccaaatg gtctaattgct tatgatttca acttctctga aactgaaaag 4560
 tacaagata ttgcaggatt ctaccgcgaa gttgaagaac aaggttataa ggttttcttt 4620
 gagtctgcgt ccaagaaaga ggtcgataag ttggtcgaag aagggaatt gtatatgttt 4680
 caaatttaca ataaagactt ttccgacaag tcccatggtta cacctaactc gcataccatg 4740
 tacttcaaac tgctgttcga tgagaataat cacggtcaga ttcgcctgag cggaggggcg 4800
 gaactcttca tgaggagagc atcgttgaaa aaagaggagc tcgtcgtgca tccggctaac 4860
 agccccattg ctaacaagaa tccggataat ccaaagaaga ctactaccct ctctatgac 4920
 gtctataagg ataagagatt ctctgaggac cagtacgagt tgcacatccc tattgcgata 4980
 aataaatgcc ctaagaacat ctttaaaatc aatactgagg tcagagtcct gcttaagcac 5040
 gacgacaacc cgtatgtgat cgggattgat aggggtgaaa ggaacttgct ttatatgttg 5100
 gttgtcgatg gaaaaggtaa tatagtggaa caatactctc tgaatgaaat tatcaacaac 5160
 ttcaatggca ttaggatcaa gaccgactat cattctctgt tggacaagaa agagaaagag 5220
 cgcttcgagg cacggcaaaa ctggacgtct attgagaaca tcaaggagct taaggctggt 5280
 tacattttctc aggttgtgca caaaatttgc gaactggtcg agaaatatga tgccgttatc 5340
 gcacttgaag atctcaacag cggatttaag aattctcggg tgaaagtcga aaaacagggtg 5400
 tatcaaaaaat tcgaaaagat gctgatcgac aagctcaatt atatggttga taaaaagagc 5460
 aacccatgcg ccacgggggg tgcgtttaag ggctatcaga ttacgaacaa atttgaatcc 5520
 ttcaagtcaa tgtcgacgca aaatggggtt atattctata taccggcgtg gcttacatct 5580
 aaaatagatc ctagcactgg gttcgtgaac ctgctgaaaa ccaagtacac ttcaatcgca 5640
 gattctaaaa aatttataag cagcttcgac agaatcatgt atgtgcccga ggaagacctc 5700
 ttcgagtttg cccttgatta caaaaatttc tcaagaacgg atgcagacta cataaagaag 5760
 tggaagctgt actcttatgg gaaccggatt cgcatattca gaaatccgaa aaaaaacaat 5820
 gtctttgatt gggaggaagt ttgtcttacc tctgcttaca aagagctgtt caataaatat 5880
 ggcatthaatt accagcaagg tgatatccgg gcgctccttt gcgaacagtc tgacaaagct 5940

ttctattctt catttatggc gctcatgtca ttgatgctgc agatgaggaa tagcattacg 6000
gggaggactg atgttgactt tctgatctcg cccgtgaaaa attctgatgg aatcttctac 6060
gattccagga attatgaggc ccaggaaaat gctatccttc ccaagaacgc agacgcaaatt 6120
ggcgcgtaca atatagctcg caaggttttg tgggctatag gccaatcaa gaaagccgaa 6180
gacgaaaagc tggacaaagt taagattgct atatctaaca aagagtggct tgagtatgcg 6240
caaacatctg ttaaacacaa acgccccgcg gctacaaaga aggctggcca ggccaagaag 6300
aagaagggct cggggctcggg gtcgggctcg ggctcggacg ccctggacga cttcgacctc 6360
gacatgctgg gctccgacgc cctcgatgat ttcgacctcg atatgctcgg cagcgacgcg 6420
ctcgatgact tcgacctcga tatgctgggg agcgacgccc tcgacgattt tgacctcgat 6480
atgctgatca actcccgtc cagcggcagc ccgaagaaga agcgcaaagt gggctcgcag 6540
tacctgcccc acaccgacga caggcacagg atcaggaga agcgcaagag gacgtacgag 6600
accttcaagt ccatcatgaa gaagtccccg ttcagcggcc caacggacct cgcgccgcg 6660
ccgaggagga tcgccgtgcc gtccaggtcc agcgcgtcgg tccccagcc ggccccgcag 6720
ccctacccgt tcacgtccag cctcagcacc atcaactacg acgagttccc caccatggtg 6780
ttcccgctcg gccagatctc ccaggccagc gcgctggccc ccgcgcccc gcaggtgctg 6840
ccccaggctc cggccccgc tccggccccg gccatggtct ccgcgctggc ccaggcgccc 6900
gccccggtgc ccgtcctcgc gccgggcccc ccgcaggcgg tcgccccgc agcgccgaag 6960
cccacgcagg ccggcgaggg caccctcagc gaggcgtcc tgcagctgca gttcgacgac 7020
gaggacctcg gcgcccctc gggcaactcg accgaccccc ccgtgttcac cgacctggcc 7080
tccgtcgaca acagcgagtt ccagcagctg ctgaaccagg gcatccccgt ggcgccgcac 7140
accacggagc ccatgctgat ggagtacccg gagcgatca cgcgcctcgt caccggcgcc 7200
cagaggcccc cggacccccg cccggccccg ctcggcgccc caggcctgcc gaacggcctc 7260
ctgagcggcg acgaggactt ctccagcatc gcggacatgg acttctccgc ctcctgggg 7320
tcgggctcgg gcagccgcga cagcaggag ggcatgttcc tcccaaagcc cgaggccggc 7380
tccgccatct cggacgtgtt cgagggcagg gaggtctgcc agccaaagcg catcaggccg 7440
ttccacccgc cgggctcccc gtgggcgaac cggccgctcc ccgccagcct ggctccaacc 7500
ccgaccggcc ccgtgcacga gccggtcggc agcctgacgc ccgcgccggt gccccagccg 7560
ctcgaccccc cgccggccgt ccccccgag gcctcccacc tcctggagga ccccgacgag 7620
gagacctcgc aggccgtgaa ggccctgagg gagatggccg acaccgtcat cccccagaag 7680
gaggaggcgg ccatctgcgg ccagatggac ctgtcgcacc cgccgccg cggccacctc 7740
gacgagctga ccacgaccct cgagtccatg accgaggacc tcaacctgga cagccccctc 7800
acgccggagc tgaacgagat cctcgacacc ttctgaacg acgagtgcct cctgcacgcc 7860
atgcacatct ccacgggcct gagcatcttc gacaccagcc tcttctgagt cgaccgatcg 7920
ttcaaacatt tggcaataaa gtttcttaag attgaatcct gttgccggtc ttgcgatgat 7980
tatcatataa tttctgttga attacgttaa gcatgtaata attaactgt aatgcatgac 8040
gttatattat agatgggttt ttatgattag agtccccgaa ttatacatct aatacgcgat 8100
agaaaacaaa atatagcgcg caaactagga taaattatcg cgcgcggtgt catctatgtt 8160
actagatcga tcccgggata tcgcggccgg tcgttcggct gcggcgagcg gtatcagctc 8220
actcaaaggc ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt 8280

gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc 8340
ataggctccg ccccccgtac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa 8400
acccgacagg actataaaga taccaggcgt tttcccctgg aagctccctc gtgcgctctc 8460
ctgttccgac cctgccgctt accggatacc tgtccgcctt tctcccttcg ggaagcgtgg 8520
cgcttttctca tagctcacgc ttaggtatc tcagttcggg ttaggtcgtt cgctccaagc 8580
tggtgtgtgt gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc ggtaactatc 8640
gtcttgagtc caaccggtga agacacgact tatcgccact ggcagcagcc actggtaaca 8700
ggattagcag agcgaggtat gtaggcgggt ctacagagtt cttgaagtgg tggcctaact 8760
acggctacac tagaaggaca gtatttggtg tctgcgctct gctgaagcca gttaccttcg 8820
gaaaaagagt tggtagctct tgatccggca acaaaaccac cgctggtagc ggtgggtttt 8880
ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat cttttgatct 8940
tttctacggg gtctgacgct cagtggaaacg aaaactcacg ttaagggtatt ttggtcatga 9000
gattatcaaa aaggatcttc acctagatcc ttttaatta aaaatgaagt tttaaatcaa 9060
tctaaagtat atatgagtaa acttggtctg acagttacca atgcttaatc agtgaggcac 9120
ctatctcagc gatctgtcta tttcgttcat ccatagttgc ctgactcccc gtcgtgtaga 9180
taactacgat acgggagggc ttaccatctg gccccagtgc tgcaatgata ccgcgagacc 9240
cacgctcacc ggctccagat ttatcagcaa taaaccagcc agccggaagg gccgagcgca 9300
gaagtgggcc tgcaacttta tccgcctcca tccagtctat taattgttgc cgggaagcta 9360
gagtaagtag ttcgccagtt aatagtttgc gcaacgttgt tgccattgct acaggcatcg 9420
tggtgtcacg ctcgtcgttt ggtatggctt cattcagctc cggttcccaa cgatcaaggc 9480
gagttacatg atccccatg ttgtgcaaaa aagcggttag ctccttcggt cctccgatcg 9540
ttgtcagaag taagtggcc gcagtgttat cactcatggt tatggcagca ctgcataatt 9600
ctcttactgt catgccatcc gtaagatgct tttctgtgac tgggtgagta tcaaccaagt 9660
cattctgaga atagtgtatg cggcgaccga gttgctcttg cccggcgtca atacgggata 9720
ataccgcgcc acatagcaga actttaaaag tgctcatcat tggaaaacgt tcttcggggc 9780
gaaaactctc aaggatctta ccgctgttga gatccagttc gatgtaacct actcgtgcac 9840
ccaactgac ttcagcatct tttactttca ccagcgtttc tgggtgagca aaaacaggaa 9900
ggcaaaatgc cgcaaaaaag ggaataaggc cgacacggaa atgttgaata ctcatactct 9960
tcctttttca atattattga agcatttatc agggttattg tctcatgagc ggatacatat 10020
ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc cgaaaagtgc 10080
cacctgacgc gccctgtagc ggcacgtcta attcggggga tctggatttt agtactggat 10140
tttggtttta ggaattagaa attttattga tagaagtatt ttacaaatac aaatacatac 10200
taagggtttc ttatatgctc aacacatgag cgaaacccta taggaaccct aattccctta 10260
tctgggaact actcacacat tattatggag aaactcgagc ttgtcgatcg acatgatcag 10320
ggagccctag attatttgta tagttcatcc atgcccatta cgtcggtaaa tgccttctgc 10380
cactccttga agttaagttc ggtcttgga tgtttcaact cagtcttacg gaacacgtac 10440
atgggttggg tcttaaggta gttagcgcc attggtttag cgaatgtgta ggtagtcctg 10500
gctgtagagc gatatctctt gccattgcct gtggtgtaag accatttgaa ggtactaatg 10560
atgggtctgt cgtagggta ggttttcttg gaccggcacc aatcagcggc agttaaggag 10620

```

ttggtcatga caggatccatc agcaggaaaag cctgtcccct tcacttgggc ttctcctttg 10680
atgtggctcc cttcgttaagt gtaacggtag ttgacggtag gcgaagcacc gtcctcaaac 10740
tgcatgtgcc tgtggacttg gtatccggag ccatcaacca tggctgcttg gaatggactc 10800
attccgtcag ggtatggaag gtattgatgg aatccgtagc caatgtgtgg caccagaatc 10860
catggagaaa actgaagatc acctttgggtg ctcttgaggt tcagctcttc gtatccgtca 10920
ttagggttcc cagtgccttg tccgaccata tcgaagtcaa cgccgttgat ggaaccgaag 10980
atgtgaagct catgtgtggc tggaagcgaa gccatgttat cttcttctcc ttactcacg 11040
gaggacgcca tgggtggcggg atcgcgccct atcgttcgta aatgggtgaaa attttcagaa 11100
aattgctttt gctttaaaag aaatgattta aattgctgca atagaagtag aatgcttgat 11160
tgcttgagat tcgtttgttt tgtatatgtt gtgttgagag gatcctctag agtcgacctg 11220
cagaagtaac accaaacaac aggttgagca tcgacaaaag aaacagtacc aagcaaataa 11280
atagcgtatg aaggcagggc taaaaaaatc cacatatagc tgctgcatat gccatcatcc 11340
aagtatatca agatcaaaaat aattataaaa catacttggtt tattataata gataggtact 11400
caaggtaga gcatatgaat agatgctgca tatgccaatc tgtatatgca tcagtaaaac 11460
ccacatcaac atgtatacct atcctagatc gatatttcca tccatcttaa actcgtaact 11520
atgaagatgt atgacacaca catacagttc caaaattaat aaatacacca ggtagtttga 11580
aacagtattc tactccgatc tagaacgaat gaacgaccgc ccaaccacac cacatcatca 11640
caaccaagcg aacaaaagca tctctgtata tgcatcagta aaacccgat caacatgtat 11700
acctatccta gatcgatatt tccatccatc atcttcaatt cgtaactatg aatatgtatg 11760
gcacacacat acagatccaa aattaataaa tccaccaggt agtttgaaac agaattctac 11820
tccgatctag aacgaccgcc caaccagacc acatcatcac aaccaagaca aaaaaagca 11880
tgaaaagatg acccgacaaa caagtgcacg gcatatattg aaataaagga aaagggcaaa 11940
ccaaacccta tgcaacgaaa caaaaaaaat catgaaatcg atcccgtctg cggaacggct 12000
agagccatcc caggattccc caaagagaaa cactggcaag ttagcaatca gaacgtgtct 12060
gacgtacagg tcgcatccgt gtacgaacgc tagcagcacg gatctaacac aaacacggat 12120
ctaacacaaa catgaacaga agtagaacta ccggggcccta accatggacc ggaacgccga 12180
tctagagaag gtagagaggg ggggggagga cgagcggcgt acctgaagc ggaggtgccg 12240
acgggtggat ttgggggaga tccactagtt ctagagcggc cgccaccgcg gtggaattct 12300
cgaggtcctc tccaaatgaa atgaacttcc ttatatagag gaagggtctt gcgaaggata 12360
gtgggattgt gcgtcatccc ttacgtcagt ggagatatca catcaatcca cttgctttga 12420
agacgtggtt ggaacgtctt cttttccac gatgctctc gtgggtgggg gtccatcttt 12480
gggaccactg tcggcagagg catcttgaac gatagccttt cttttatcgc aatgatggca 12540
tttgtaggtg ccaccttcc tttctactgt cttttgatc aagtaccga tagctgggca 12600
atggaatccg aggaggttcc ccgatattac cttttgttga aaagtctcaa tagccctttg 12660
gtcttctgag actgtatctt tgatattctt ggagtagacg agagtgtcgt gctccaccat 12720
gttatcacat caattcactt gctttgaaga cgtggttga acgtcttctt ttccacgat 12780
gtcctcgtg ggtgggggtc catctttggg accactgtcg gcagaggcat cttgaacgat 12840
agcctttcct ttatcgcaat gatggcattt gtagtgcca ccttcctttt ctactgtcct 12900
tttgatcaag tgacagatag ctgggcaatg gaatccgagg aggtttcccg atattaccct 12960

```

ttgttgaaaa gtctcaatag ccctttggtc ttctgagact tgcaggcaag ca 13012
 <210> 280
 <211> 13013
 <212> DNA
 <213> 人工序列
 <220>
 <223> pGEP755 expression plasmid
 <400> 280
 cgatcttttcg gttgtcgaaa agctcaaaga aataattatc cagaaggctg atgaaatcta 60
 caagggtgtac ggctcaagcg agaagctctt tgatgctgac ttcgtgttgg agaagtctct 120
 taaaaaaaaac gacgcagtcg tcgcgataat gaaagatttg ctggattcag tgaaatcctt 180
 cgagaattat atcaaaagcct tcttcggcga ggggaaggag acaaacaggg atgagtcctt 240
 ctatggagac ttcgttctgg cttacgacat cttctttaag gtcgaccaca tctatgacgc 300
 aattcggaac tatgtgacgc agaagccgta ttcgaaagat aagttcaagc tctattttcca 360
 aaaccctcaa tttatgctg ggtgggataa agacaaagag accgattacc gggcaacaat 420
 tttgcggtac ggggtctaaat attacctcgc tataatggat aagaaatacg ctaaattgtct 480
 ccagaaaatt gacaaaagatg acgtcaacgg caattatgaa aaaatcaatt ataaactcct 540
 tcctggccca aataaaatgc tcccaggagg gtttttttcc aaaaagtga tggcctatta 600
 taatccatca gaggatattc agaaaatcta taaaaatggg acctttaaga aggggtgacat 660
 gttaaacctg aacgattgcc acaagcttat agattttttc aaagactcta ttagccgcta 720
 tcccaaatgg tctaattgctt atgatttcaa cttctctgaa actgaaaagt acaagatat 780
 tgcaggattc taccgcgaag ttgaagaaca aggttataag gtttcctttg agtctgcgtc 840
 caagaaagag gtcgataagt tggtcgaaga agggaaattg tatatgtttc aaattttaca 900
 taaagacttt tccgacaagt cccatggtac acctaactctg cataccatgt acttcaaact 960
 gctgttcgat gagaataatc acggtcagat tcgcctgagc ggagggggcg aactcttcat 1020
 gaggagagca tcgttgaaaa aagaggagct cgtcgtgcat ccggctaaca gccccattgc 1080
 taacaagaat ccgataatc caaagaagac tactaccctc tcctatgacg tctataagga 1140
 taagagattc tctgaggacc agtacgagtt gcacatccct attgcgataa ataaatgccc 1200
 taagaacatc tttaaaatca atactgaggt cagagtccctg ctttaagcacg acgacaaccc 1260
 gtatgtgata gggattgata ggggtgaaag gaacttgctt tatattgttg ttgtcgatgg 1320
 aaaaggtaat atagtggaac aatactctct gaattgaaatt atcaacaact tcaatggcat 1380
 taggatcaag accgactatc attctctgtt ggacaagaaa gagaaagagc gcttcgaggc 1440
 acggcaaaac tggacgtcta ttgagaacat caaggagctt aaggctggtt acattttctca 1500
 ggttggtgcac aaaaatttgcg aactggtcga gaaatatgat gccgttatcg cacttgaaga 1560
 tctcaacagc ggatttaaga attctcgggt gaaagtcgaa aaacaggtgt atcaaaaatt 1620
 cgaagagatg ctgatcgaca agctcaatta tatggttgat aaaaagagca acccatgcgc 1680
 cacgggggggt gcgcttaagg gctatcagat tacgaacaaa tttgaatcct tcaagtcaat 1740
 gtcgacgcaa aatgggttta tattctatat accggcgtgg cttacatcta aaatagatcc 1800
 tagcactggg ttcgtgaacc tgctgaaaac caagtacact tcaatcgcag attctaaaaa 1860

atttataagc agcttcgaca gaatcatgta tgtgcccagag gaagacctct tcgagtttgc 1920
 ccttgattac aaaaatttct caagaacgga tgcagactac ataaagaagt ggaagctgta 1980
 ctcttatggg aaccggattc ggatattcag aaatccgaaa aaaaacaatg tctttgattg 2040
 ggaggaagtt tgtcttacct ctgcttacia agagctgttc aataaatatg gcattaatta 2100
 ccagcaaggt gatatccggg cgctcctttg cgaacagtct gacaaagctt tctattcttc 2160
 atttatggcg ctcatgtcat tgatgctgca gatgaggaat agcattacgg ggaggactga 2220
 tgttgacttt ctgatctcgc ccgtgaaaaa ttctgatgga atcttctacg attccaggaa 2280
 ttatgaggcc caggaaaaatg ctatccttcc caagaacgca gacgcaaatg gcgcgtacaa 2340
 tatagctcgc aaggttttgt gggctatagg ccaattcaag aaagccgaag acgaaaagct 2400
 ggacaaaagt aagattgcta tatctaacia agagtggctt gagtatgcgc aaacatctgt 2460
 taaacacaaa cgccccgcgg ctacaaagaa ggctggccag gccaaagaaga agaagggctc 2520
 ggggtcgggg tcgggctcgg gctcggacgc cctggacgac ttcgacctcg acatgctggg 2580
 ctccgacgcc ctcatgatt tcgacctcga tatgctcggc agcgacgcgc tcgatgactt 2640
 cgacctgat atgctgggga gcgacgcctt cgacgatttt gacctcgata tgctgatcaa 2700
 ctcccgtcc agcggcagcc cgaagaagaa gcgcaaagtg ggctcgcagt acctgcccga 2760
 caccgacgac aggcacagga tcgaggagaa gcgcaagagg acgtacgaga ctttcaagtc 2820
 catcatgaag aagtccccgt tcagcgcccc aacggacccc cgccccgcgc cgaggaggat 2880
 cgccgtgccg tccagggtcca gcgcgtcggc ccccaagccg gccccgcagc cctaccggtt 2940
 cacgtccagc ctacgacca tcaactacga cgagttcccc accatggtgt tcccgctccg 3000
 ccagatctcc caggccagcg cgctggcccc cgcgcccccg caggtgctgc cccaggtcc 3060
 ggcccccgct ccggcccccg ccatggtctc cgcgctggcc caggcgcccc ccccggtgcc 3120
 cgctctcgcg ccgggccccg cgagggcggt cgccccgcca gcgccgaagc ccacgcaggc 3180
 cggcgagggc accctcagcg aggcgctcct gcagctgcag ttcgacgacg aggacctcg 3240
 cgccctcctg ggcaactcga ccgacccccg cgtgttcacc gacctggcct ccgtcgacia 3300
 cagcgagttc cagcagctgc tgaaccaggg catcccggtg gcgccgcaca ccacggagcc 3360
 catgctgatg gagtaccggg aggcgatcac gcgcctcgtc accggcgccc agaggcccc 3420
 ggacccccgc ccggccccgc tcggcgcccc aggcctgccg aacggcctcc tgagcggcga 3480
 cgaggacttc tccagcatcg cggacatgga cttctccgcc ctctgggggt cgggctcggg 3540
 cagccgcgac agcaggagg gcatgttctt cccaaagccc gaggcgggt ccgccatctc 3600
 ggacgtgttc gagggcaggg aggtctgcca gccaaagcgc atcaggccgt tccaccgcc 3660
 gggctccccg tgggcgaacc ggccgctccc cgccagcctg gctccaacct cgaccggccc 3720
 cgtgcacgag ccggtcggca gctgacgcc cgcgccggtg cccagccgc tcgacccccg 3780
 gccggccgtc accccccagg cctccacct cctggaggac cccgacgagg agacctcga 3840
 ggccgtgaag gccctgaggg agatggccga caccgtcatc cccagaagg aggaggcggc 3900
 catctgcggc cagatggacc tgtcgacccc gccgcgcgc gccacctcg acgagctgac 3960
 cagaccctc gagtccatga ccgaggacct caacctggac agccccctca cgccggagct 4020
 gaacgagatc ctgcacacct tcctgaacga cgagtgcctc ctgcacgcca tgcacatctc 4080
 caggggcctg agcatcttcg acaccagcct cttctgagtc gaccgatcgt tcaaacattt 4140
 ggcaataaag tttcttaaga ttgaatcctg ttgccggtct tgcgatgatt atcatataat 4200

ttctgttgaa ttacgttaag catgtaataa ttaacatgta atgcatgacg ttattttatga 4260
gatgggtttt tatgattaga gtcccgcaat tatacattta atacgcgata gaaaacaaaa 4320
tatagcgcgc aaactaggat aaattatcgc gcgcggtgtc atctatgtta ctagatcgat 4380
cccgggatat cgcggccgcg tcgttaagct gcggcgagcg gtatcagctc actcaaaggc 4440
ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt gagcaaaagg 4500
ccagcaaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg 4560
ccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg 4620
actataaaga taccaggcgt tccccctgg aagctccctc gtgcgctctc ctgttccgac 4680
cctgccgctt accggatacc tgtccgctt tctcccttcg ggaagcgtgg cgcttttctca 4740
tagctcacgc ttaggtatc tcagttcggg ttaggtcgtt cgctccaagc tgggctgtgt 4800
gcacgaaccc cccgttcagc ccgaccgtg cgccttatcc ggtaactatc gtcttgagtc 4860
caaccggta agacacgact tatcgccact ggcagcagcc actggtaaca ggattagcag 4920
agcgaggat gtaggcggtg ctacagagtt cttgaagtgg tggcctaact acggctacac 4980
tagaaggaca gtatttggtg tctgcgctct gctgaagcca gttaccttcg gaaaaagagt 5040
tggtagctct tgatccggca aacaaaccac cgctggtagc ggtggttttt ttgtttgcaa 5100
gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat cctttgatct tttctacggg 5160
gtctgacgct cagtgaacg aaaactcacg ttaagggtt ttggtcatga gattatcaaa 5220
aaggatcttc acctagatcc ttttaaatta aaaatgaagt tttaaatcaa tctaaagtat 5280
atatgagtaa acttggctctg acagttacca atgcttaatc agtgaggcac ctatctcagc 5340
gatctgtcta tttcgttcac ccatagttgc ctgactcccc gtcgtgtaga taactacgat 5400
acgggagggc ttaccatctg gccccagtgc tgcaatgata ccgcgagacc cacgctcacc 5460
ggctccagat ttatcagcaa taaaccagcc agccggaagg gccgagcgca gaagtgttcc 5520
tgcaacttta tccgctcca tccagtctat taattgttgc cgggaagcta gagtaagtag 5580
ttcgccagtt aatagtttgc gcaacgttgt tgccattgct acaggcatcg tgggtgcacg 5640
ctcgtcgttt ggtaggtt cattcagctc cggttcccaa cgatcaaggc gagttacatg 5700
atcccccatg ttgtgcaaaa aagcggttag ctcttcggt cctccgatcg ttgtcagaag 5760
taagttggcc gcagtgttat cactcatggt tatggcagca ctgcataatt ctcttactgt 5820
catgccatcc gtaagatgct tttctgtgac tggtagtac tcaaccaagt cattctgaga 5880
atagtgtatg cggcgaccga gttgctcttg cccggcgta atacgggata atacgcgcc 5940
acatagcaga actttaaaaag tgctcatcat tggaaaacgt tcttcggggc gaaaactctc 6000
aaggatctta ccgctgttga gatccagttc gatgtaacce actcgtgcac ccaactgac 6060
ttcagcatct tttactttca ccagcgttgc tgggtgagca aaaacaggaa ggcaaaatgc 6120
cgcaaaaaaag ggaataaggg cgacacggaa atgttgaata ctcatactct tcctttttca 6180
atattattga agcatttatc agggttattg tctcatgagc ggatacatat ttgaatgtat 6240
ttagaaaaat aaacaaatag gggttccgcg cacatttccc cgaaaagtgc cacctgacgc 6300
gccctgtagc ggcacgtcta attcggggga tctggatttt agtactggat tttggtttta 6360
ggaattagaa attttattga tagaagtatt ttacaaatac aaatacatac taagggtttc 6420
ttatatgctc aacacatgag cgaaacccta taggaacct aattccctta tctgggaact 6480
actcacacat tattatggag aaactcgagc ttgtcgatcg acatgatcag ggagccctag 6540

attatttgtg tagttcatcc atgcccatta cgtcggtaaa tgccttctgc cactccttga 6600
 agttaagttc ggtcttggaa tgtttcaact cagtcttacg gaacacgtac atgggttggg 6660
 tcttaaggta gttagcggcc attggtttag cgaatgtgta ggtagtcctg gctgtagagc 6720
 gatatctctt gccattgcct gtggtgtaag accatttgaa ggtactaatg atgggtctgt 6780
 cgttagggtg ggttttcttg gaccggcacc aatcagcggc agttaaggag ttgggtcatga 6840
 cagggtccatc agcaggaaaag cctgtcccct tcacttgggc ttctcctttg atgtgggtcc 6900
 cttcgtaagt gtaacggtag ttgacggtag gcgaagcacc gtcctcaaac tgcattgtcc 6960
 tgtggacttg gtatccggag ccatcaacca tggctgcttg gaatggactc attccgtcag 7020
 ggtatggaag gtattgatgg aatccgtagc caatgtgtgg caccagaatc catggagaaa 7080
 actgaagatc acctttggtg ctcttgaggt tcagctcttc gtatccgtca ttagggttcc 7140
 cagtgccttg tccgaccata tcgaagtcaa cgccgttgat ggaaccgaag atgtgaagct 7200
 catgtgtggc tggaagcgaa gccatgttat cttcttctcc ttactcacg gaggacgcca 7260
 tgggtggcggg atcgcgccct atcgttcgta aatggtgaaa attttcagaa aattgctttt 7320
 gctttaaaag aaatgattta aattgctgca atagaagtag aatgcttgat tgcttgagat 7380
 tcgtttgttt tgtatatgtt gtgttgagag gatcctcaag cttcgacctg cagaagtaac 7440
 accaaacaac aggggtgagca tcgacaaaag aaacagtacc aagcaaataa atagcgtatg 7500
 aaggcagggc taaaaaaatc cacatatagc tgctgcatat gccatcatcc aagtatatca 7560
 agatcaaaat aattataaaa catacttggt tattataata gataggtact caaggttaga 7620
 gcatatgaat agatgctgca tatgccatca tgtatatgca tcagtaaaac ccacatcaac 7680
 atgtatacct atcctagatc gatatttcca tccatcttaa actcgtaact atgaagatgt 7740
 atgacacaca catacagttc caaaattaat aaatacacca ggtagtttga aacagtattc 7800
 tactccgac tagaacgaat gaacgaccgc ccaaccacac cacatcatca caaccaagcg 7860
 aacaaaagca tctctgtata tgcacagta aaacccgcac caacatgtat acctatccta 7920
 gatcgatatt tccatccatc atcttcaatt cgtaactatg aatatgtatg gcacacacat 7980
 acagatccaa aattaataaa tccaccaggt agtttgaaac agaattctac tccgatctag 8040
 aacgaccgcc caaccagacc acatcatcac aaccaagaca aaaaaagca tgaaaagatg 8100
 acccgacaaa caagtgcacg gcatatattg aaataaagga aaagggcaaa ccaaacccta 8160
 tgcaacgaaa caaaaaaaat catgaaatcg atcccgtctg cggaacggct agagccatcc 8220
 caggattccc caaagagaaa cactggcaag ttagcaatca gaacgtgtct gacgtacagg 8280
 tcgcatccgt gtacgaacgc tagcagcacg gatctaacac aaacacggat ctaacacaaa 8340
 catgaacaga agtagaacta ccggggcccta accatggacc ggaacgccga tctagagaag 8400
 gtagagaggg ggggggagga cgagcggcgt acctgaagc ggaggtgccg acgggtggat 8460
 ttgggggaga tccactagtt ctagagcggc cgccaccgcg gtggaattct cgaggtcctc 8520
 tccaaatgaa atgaacttcc ttatatagag gaagggtctt gcgaaggata gtgggattgt 8580
 gcgtcatccc ttacgtcagt ggagatatca catcaatcca cttgctttga agacgtgggt 8640
 ggaacgtctt ctttttccac gatgtcctc gtgggtgggg gtccatcttt gggaccactg 8700
 tcggcagagg catcttgaac gatagccttt cttttatcgc aatgatggca tttgtaggtg 8760
 ccaccttctt ttctactgt ctttttgatc aagtgaccga tagctgggca atggaatccg 8820
 aggaggtttc ccgataattac cttttgttga aaagtctcaa tagccctttg gtcttctgag 8880

actgtatcctt tgatattcctt ggagtagacg agagtgtcgt gctccaccat gttatcacat 8940
caattcactt gctttgaaga cgtggttgga acgtcttcctt tttccacgat gtcctcgtg 9000
ggtgggggtc catctttggg accactgtcg gcagaggcat cttgaacgat agcctttcct 9060
ttatcgc aat gatggcattt gtaggtgcc ccttcctttt ctactgtcct tttgatcaag 9120
tgacagatag ctgggcaatg gaatccgagg aggtttcccg atattaccct ttgttgaaaa 9180
gtctcaatag ccctttggtc ttctgagact tgcaggcaag caagcatgaa tgcctggggg 9240
agaagaactc gagagggaat tgcagatcat gaggcagatg gctatTTTTTg tgtcacatat 9300
gcgcaaaaaag agaggctata tttgtgtccc taggtttcttc gttgtattgc agtttccata 9360
tcaatctgac ttggtcgc aat gagaaattga tggttaaata atttgaatct ctcatgtagt 9420
atcaactatt agatattatt ttcaccaa atatttccat cggagaagaa gaggctacag 9480
aggaagcaga agagaggggt gggagaattt ttacactttt gtacaccac ttaaacagca 9540
aaatccgtat gaaaacaggc ccaccaa aatgccacga taacaatccg tagaaacaaa 9600
agcttcattt aacagcggcg caacaaagca cgcttatcca tggtagttgt agtccgtatg 9660
cgatccaaag atcacgattc acgcgtgacg gacggacgac gcgtgccaca ccacaactaa 9720
cggcatccat ggtagttgta gtccgtatgc gatccaaaga tcacgattca cgcgtgacgg 9780
acggacgacg cgcgccacac cacaactaac agcgtgagcc agcgtccaaa ctccggatgg 9840
caacggggac gaaacccgtc gggtagtcac tgcccaaacc cgtccccgca accttcatcc 9900
caaacccgtc cccgtttccg gtcgcgggtt tcagttttct accagaccg tccccatcgg 9960
gtttttcatc cccgtcggga aatccgaacc cgccagcatt tcagcaccaa gccaaagtgt 10020
cagcagcaac atgaataaaa aacaacccgt ttcaacacca agataaaaca aaacattata 10080
atttagacaa catttcacac gtataacaat aacatatagt tctcacatat aacaacacca 10140
tttcacacat aaaacaacac catttgggat aaaaatatgg gctatatcag gccattttta 10200
tgggccatat tgagttttcg tgggtttcac aggtaccgga tttgtagaat gctgaaccgg 10260
gtttgaaccg taaaatccgc gggtattgaa tttgaccaa tcccgtcgtc ccctggtggg 10320
gtaaaaacac catcttgagt ccaaacggcc accaaccaaa ctccgacggc aacaaacaaa 10380
cggcgttgct ttgtcctcgt gtatctccgt gaccgtcaa tctcccggct gtttccccgg 10440
aattgcgtgg actctctcat ccacacgcaa accgctctc cctcctctct cgtcctatcc 10500
gccccgggtc cgtagcctca cgggactctt ctctctcctt tgctataaaa tccccgccc 10560
ctcccgtctc ctctccacac atccaaactc tcaatcgac cgagaaaaat ctctagcga 10620
tcgaagcgaa gcctctccc atcctctcaa ggtacgccg tttcccgtcg atcctctcc 10680
ttccgttcgt gttctgtagc cgatcgattc gattccctta caccgttcg tgttctctcg 10740
tgatcgatc gattgtttgt tgctagaagg aactcgtaga tctggcgttt atgaactgtg 10800
attcgggtta gtccagatcg attcaggtcg gtcgtcgttg agcctctcgg ctatgtctgg 10860
attatcgtgt agatctgctg gttcagttga ttatgttctt ctaggagtaa tttcgttggg 10920
tcagcgcgat ttctgcttaa tctatgtcgc ttattgcgc tgtacctatc tactaagcta 10980
tgtgcacctg taattttgct agattattcg ttcactctcg tagttggttt gtcacagtaa 11040
tccgtatggg ttctgacgat gttattgttg gtcataccta ggcttctcca gattttattt 11100
tgttaaaatt ggatagatct gctactgata gttgatgatg gaatttggtg ctgaatctat 11160
gctattttatt gcgcctatac ctgatctatc gggctatgta cggctgtagt ttactggatt 11220

attcgttcat cctcggtagt tggttcatcg tttgggttct gacgataata ttgttgatta 11280
tgcgtaggct tctgcagatt gttgttaaaa ttggatacat cggttactga tggttgatga 11340
tagatttgtg ctgaacctat ctgtttattg ctcctatacc tgatctatag ggctatgtat 11400
gcctgtaatt taccagatta ttcgttcac ctcgtagttg gttcatctct ataattcgta 11460
tgggttctta tgatgttatac gttgattatg cctagtctta tacagattat tgtgtcaaga 11520
ttgaatatac ctgctactga tcggtgataa tttgggttagt agtttgcaat ctgctaggaa 11580
cacgttacca ctgtaacttg taaacatggg ttgccagagt agtttggttct actactcttg 11640
atatggttgc tgatttttagt cgcctccttt tggatcatgt attgatgtcc ttgcagattt 11700
ccgtgtactt accccggcctt ttgtgtactt cgtgttaaca ggtcgggtac cgaagcaaac 11760
atggcatcta gcatggcacc aaagaaaaaa aggaaagttt ccaaacttga aaaatttaca 11820
aactgctact ccctttccaa gacgcttagg tttaaagcga tccccgttgg caagacccaa 11880
gagaatatcg ataacaaaag acttctggtc gaagatgaaa aaagggccga agactacaag 11940
gggggtcaaga agttgctcga tcgctattat ctttccttta tcaacgatgt gcttcattca 12000
atcaaactga agaacttgaa taactacatt agccttttca gaaagaaaac gaggactgaa 12060
aaggagaaca aggaacttga gaatcttgaa ataaaccttc gcaaagaaat tgcaaaagcc 12120
ttcaagggga acgaaggata taaatctctt ttcaaaaaag acattataga aacaattttg 12180
cctgagtttc ttgacgacaa ggatgaaatt gcgctcgtca atagctttta cggatttaca 12240
actgccttca cagggttctt cgacaatagg gagaatatgt ttagcgagga ggcaaaaagc 12300
acatccatcg cattcagatg catcaatgaa aatcttacc ggtacatatc gaatatggac 12360
atatttgaaa aagtggatgc aatattcgat aagcacgaag tccaggagat aaaggaaaag 12420
atactgaata gcgactatga tgtcgaagat ttttcgaag gtgagttctt caactttgtc 12480
ctgactcaag aaggcattga tgtctataat gcaataattg gaggttttgt gactgagtct 12540
ggcgagaaga taaagggcctt gaacgagtat atcaatctct acaaccagaa gactaagcaa 12600
aagttgccta aatttaaacc gctttacaag caagttttga gcgaccggga aagcctttcc 12660
ttttacggtg aaggatacac gagcgatgaa gaagtcctcg aagtcttccg caacacactc 12720
aacaagaact cagaaatctt ttcctcaatt aaaaaattgg agaagctttt caagaacttc 12780
gatgaatact cttcggcggg gatTTTTgtg aagaacggcc cggcaatttc cacaatatct 12840
aaagacattt tcggagaatg gaacgtgata agagacaagt ggaatgcgga gtatgatgac 12900
atacacctga agaagaagc agttgtgact gaaaaatacg aagatgacag gagaaaaagc 12960
tttaaaaaaga tcgggtcctt ttcactggaa cagctgcagg agtatgccga cgc 13013

<210> 281

<211> 13012

<212> DNA

<213> 人工序列

<220>

<223> pGEP756 expression plasmid

<400> 281

cgatctttcg gttgtcgaaa agctcaaaga aataattatc cagaaggtcg atgaaatcta 60
caaggtgtac ggctcaagcg agaagctctt tgatgctgac ttcgtgttgg agaagtctct 120

```

taaaaaaac gacgcagtcg tcgcgataat gaaagatttg ctggattcag tgaaatcctt 180
cgagaattat atcaaacctt tcttcggcga ggggaaggag acaaacaggg atgagtcctt 240
ctatggagac ttcgttcttg cttacgacat ctttcttaag gtcgaccaca tctatgacgc 300
aattcggaac tatgtgacgc agaagccgta ttcgaaagat aagttcaagc tctatttcca 360
aaaccctcaa tttatgcgtg ggtgggataa agacgtagag accgatcgcc gggcaacaat 420
tttgcggtac ggggtctaaat attacctcgc tataatggat aagaaatacg ctaaattgtct 480
ccagaaaatt gacaaaagat acgtcaacgg caattatgaa aaaatcaatt ataaactcct 540
tcctggccca aataaaatgc tcccgaaggt gtttttttcc aaaaagtgga tggcctatta 600
taatccatca gaggatattc agaaaatcta taaaaatggg acctttaaga aggggtgacat 660
gtttaacctg aacgattgcc acaagcttat agattttttc aaagactcta ttagccgcta 720
tcccaaatgg tctaattgctt atgatttcaa cttctctgaa actgaaaagt acaagatat 780
tgcaggattc taccgcgaag ttgaagaaca aggttataag gtttcctttg agtctgcgtc 840
caagaaaagag gtcgataagt tggtcgaaga agggaaattg tatatgtttc aaatttaca 900
taaagacttt tccgacaagt cccatggtac acctaactg cataccatgt acttcaaact 960
gctgttcgat gagaataatc acggtcagat tcgcctgagc ggaggggagg aactcttcat 1020
gaggagagca tcgttgaaaa aagaggagct cgtcgtgcat ccggctaaca gccccattgc 1080
taacaagaat ccgataatc caaagaagac tactaccctc tcctatgacg tctataagga 1140
taagagattc tctgaggacc agtacgagtt gcacatccct attgcgataa ataaatgccc 1200
taagaacatc tttaaaatca atactgaggt cagagtcctg cttaaagcac acgacaaccc 1260
gtatgtgatc gggattgata ggggtgaaag gaacttgctt tatattgttg ttgtcgatgg 1320
aaaaggtaat atagtggaa aatactctct gaatgaaatt atcaacaact tcaatggcat 1380
taggatcaag accgactatc attctctgtt ggacaagaaa gagaaagagc gcttcgaggc 1440
acggcaaaac tggacgtcta ttgagaacat caaggagctt aaggctgggt acatttctca 1500
ggttgtgcac aaaatttgcg aactggtcga gaaatatgat gccgttatcg cacttgaaga 1560
tctcaacagc ggatttaaga attctcgggt gaaagtcgaa aaacaggtgt atcaaaaatt 1620
cgaaaagatg ctgatcgaca agctcaatta tatggttgat aaaaagagca acccatgcgc 1680
cacggggggt gcgcttaagg gctatcagat tacgaacaaa tttgaatcct tcaagtcaat 1740
gtcgacgcaa aatgggttta tattctatat accggcgttg cttacatcta aaatagatcc 1800
tagcactggg ttcgtgaacc tgctgaaaac caagtacact tcaatcgagc attctaaaaa 1860
atttataagc agcttcgaca gaatcatgta tgtgcccag gaagacctct tcgagtttgc 1920
ccttgattac aaaaatttct caagaacgga tgcagactac ataaagaagt ggaagctgta 1980
ctcttatggg aaccggattc ggatattcag aaatccgaaa aaaaacaatg tctttgattg 2040
ggaggaagtt tgtcttacct ctgcttaca agagctgttc aataaatatg gcattaatta 2100
ccagcaaggt gatatccggg cgtccttttg cgaacagtct gacaaagctt tctattcttc 2160
atttatggcg ctcatgtcat tgatgctgca gatgaggaat agcattacgg ggaggactga 2220
tgttgacttt ctgatctcgc cctgaaaaa ttctgatgga atcttctacg attccaggaa 2280
ttatgaggcc caggaaaatg ctatccttcc caagaacgca gacgcaaatg gcgcgtacaa 2340
tatagctcgc aaggttttgt gggctatagg ccaattcaag aaagccgaag acgaaaagct 2400
ggacaaagtt aagattgcta tatctaacaa agagtggctt gagtatgcgc aaacatctgt 2460

```

taaacacaaa cgccccgcgg ctacaaagaa ggctggccag gccaagaaga agaagggtc 2520
 ggggtcgggg tcgggtcgg gctcggacgc cctggacgac ttcgacctcg acatgctggg 2580
 ctccgacgcc ctcatgatt tcgacctga tatgctcggc agcgacgcgc tcgatgactt 2640
 cgacctcgat atgctgggga gcgacgccct cgacgatttt gacctcgata tgctgatcaa 2700
 ctcccgtcc agcggcagcc cgaagaagaa gcgcaaagt ggctcgcagt acctgcccga 2760
 caccgacgac aggcacagga tcgaggagaa gcgcaagagg acgtacgaga ctttcaagtc 2820
 catcatgaag aagtccccgt tcagcggccc aacggacccc cgccccgcgc cgaggaggat 2880
 cgccgtgccg tccaggtcca gcgcgtcgggt cccaagccg gccccgcgc cctaccggtt 2940
 cacgtccagc ctacgacca tcaactacga cgagttcccc accatggtgt tcccgctcgg 3000
 ccagatctcc caggccagcg cgctggcccc cgcgcccccg caggtgctgc cccaggctcc 3060
 ggcccccgct ccggccccgg ccatggtctc cgcgctggcc caggcgcccc ccccggtgcc 3120
 cgtcctcgcg ccggggccgc cgcaggcgggt cgccccgcca gcgccgaagc ccacgcaggc 3180
 cggcgagggc accctcagcg aggcgtcct gcagctcgag ttcgacgacg aggacctcgg 3240
 cgccctcctg ggcaactcga ccgacccccg cgtgttcacc gacctggcct ccgtcgacaa 3300
 cagcgagttc cagcagctgc tgaaccagg catcccggtg gcgccgcaca ccacggagcc 3360
 catgctgatg gagtaccgg aggcgatcac gcgcctcgtc accggcgccc agaggcccc 3420
 ggacccccgc ccggccccgc tcggcgcccc aggcctgccg aacggcctcc tgagcggcga 3480
 cgaggacttc tccagcatcg cggacatgga cttctccgc ctctgggggt cgggctcggg 3540
 cagccgcgac agcaggagg gcattgtcct ccaaagccc gaggcgggt ccgccatctc 3600
 ggacgtgttc gagggcagg aggtctgcca gccaaagcgc atcaggccgt tccaccgcc 3660
 gggctccccg tgggcgaacc ggccgctccc cgccagcctg gctccaacct cgaccggccc 3720
 cgtgcacgag ccggtcggca gcctgacgcc cgcgccgggtg cccagccgc tcgacccccg 3780
 gccggccgtc acccccgagg cctcccacct cctggaggac cccgacgagg agacctcga 3840
 ggccgtgaag gccctgagg agatggccga caccgtcatc cccagaagg aggaggcggc 3900
 catctgcggc cagatggacc tgtcgcaccc gccgccgcgc ggccacctcg acgagctgac 3960
 cacgaccctc gattccatga ccgaggacct caacctggac agccccctca cgccggagct 4020
 gaacgagatc ctgacacct tcctgaacga cgagtgcctc ctgcacgcca tgcacatctc 4080
 cacgggcctg agcatcttcg acaccagcct cttctgagtc gaccgatcgt tcaaacattt 4140
 ggcaataaag tttcttaaga ttgaatcctg ttgccggtct tgcatgatt atcatataat 4200
 ttctgttgaa ttacgttaag catgtaataa ttaacatgta atgcatgacg ttatttatga 4260
 gatgggtttt tatgattaga gtcccgcaat tatacattta atacgcgata gaaaacaaaa 4320
 tatagcgcgc aaactaggat aaattatcgc gcgcgggtgtc atctatgtta ctagatcgat 4380
 cccgggatat cgcgccgggt cgttcgggtg cggcgagcgg tatcagctca ctcaaaggcg 4440
 gtaatacggc tatccacaga atcaggggat aacgcaggaa agaacatgtg agcaaaaggc 4500
 cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca taggtccgc 4560
 cccctgacg agcatcacia aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga 4620
 ctataaagat accaggcgtt tccccctgga agtcctcctg tgcgctctcc tgttccgacc 4680
 ctgccgctta ccggatacct gtccgccttt ctcccttcgg gaagcgtggc gctttctcat 4740
 agctcacgct gtaggtatct cagttcgggt taggtcgttc gctccaagct gggctgtgtg 4800

cacgaacccc ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgagtcc 4860
aacccggtaa gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga 4920
gcgaggatat taggcgggtgc tacagagttc ttgaagtggg ggcctaacta cggctacact 4980
agaaggacag tatttgggtat ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt 5040
ggtagctctt gatccggcaa acaaaccacc gctggtagcg gtggtttttt tgtttgcaag 5100
cagcagatta cgcgcagaaa aaaaggatct caagaagatc ctttgatctt ttctacgggg 5160
tctgacgctc agtggaacga aaactcacgt taagggtatt tggatcatgag attatcaaaa 5220
aggatcttca cctagatcct tttaaattaa aatgaagtt ttaaatcaat ctaaagtata 5280
tatgagtaaa cttgggtctga cagttaccaaa tgcttaatac gtgaggcacc tatctcagcg 5340
atctgtctat ttcgttcacg catagttgcc tgactccccg tcgtgtagat aactacgata 5400
cgggagggtt taccatctgg ccccgatgct gcaatgatac cgcgagaccg acgctcaccg 5460
gctccagatt tatcagcaat aaaccagcca gccggaaggg ccgagcgcag aagtggctct 5520
gcaactttat ccgcctccat ccagtctatt aattgttgcc gggaagctag agtaagtagt 5580
tcgccagtta atagtttgcg caacgttggt gccattgcta caggcatcgt ggtgtcacgc 5640
tcgtcgtttg gtatggcttc attcagctcc ggttcccaac gatcaaggcg agttacatga 5700
tccccatgt tgtgcaaaaa agcggttagc tccttcggtc ctccgatcgt tgtcagaagt 5760
aagttggccg cagtgttata actcatggtt atggcagcac tgcataattc tcttactgtc 5820
atgccatccg taagatgctt ttctgtgact ggtgagtact caaccaagtc attctgagaa 5880
tagtgtagtc ggcgaccgag ttgctcttgc ccggcgtaaa tacgggataa taccgcgcca 5940
catagcagaa ctttaaaaagt gctcatcatt ggaaaacgtt cttcggggcg aaaactctca 6000
aggatcttac cgctgttgag atccagttcg atgtaacca ctcgtgcacc caactgatct 6060
tcagcatctt ttactttcac cagcgtttct gggtagcaaa aaacaggaag gcaaaatgcc 6120
gcaaaaaagg gaataagggc gacacggaaa tgttgaatac tcatactctt cttttttcaa 6180
tattattgaa gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt 6240
tagaaaaata aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc acctgacgcg 6300
ccctgtagcg gcacgtctaa ttcgggggat ctggatttta gtactggatt ttggtttttag 6360
gaattagaaa ttttattgat agaagtattt tacaataca aatacatact aagggtttct 6420
tatatgctca acacatgagc gaaaccctat aggaacccta attcccttat ctgggaacta 6480
ctcacacatt attatggaga aactcgagct tgctgatcga catgatcagg gagccctaga 6540
ttatttgtat agttcatcca tgccattac gtcggtaaat gccttctgcc actccttgaa 6600
gttaagtctg gtcttggaat gtttcaactc agtcttacgg aacacgtaca tgggttggtt 6660
cttaaggtag ttagcggcca ttggtttagc gaatgtgtag gtagtcctgg ctgtagagcg 6720
atatctcttg ccattgcctg tgggtgaaga ccatttgaag gtactaatga tgggtctgtc 6780
gttaggtag gttttcttgg accggcacca atcagcgca gtttaaggagt tggatcatgac 6840
aggtccatca gcaggaaagc ctgtccctt cacttgggtt tctcctttga tgtggctccc 6900
ttcgtaaagt taacggtagt tgacggtgag cgaagcaccg tcctcaaact gcattgtcct 6960
gtggacttgg tatccggagc catcaacct ggctgcttgg aatggactca ttccgtcagg 7020
gtatggaagg tattgatgga atccgtagcc aatgtgtggc accagaatcc atggagaaaa 7080
ctgaagatca cctttggtgc tcttgaggtt cagctcttcg tatccgtcat tagggttccc 7140

```

agtgccttgt ccgaccatat cgaagtcaac gccgttgatg gaaccgaaga tgtgaagctc 7200
atgtgtggct ggaagcgaag ccatgttatac ttcttctcct ttactcacgg aggacgcat 7260
gggtggcggga tcgcgcccta tcgttcgtaa atggtgaaaa ttttcagaaa attgcttttg 7320
ctttaaaaga aatgatttaa attgctgcaa tagaagtaga atgcttgatt gcttgagatt 7380
cgtttgtttt gtatatgttg tgttgagagg atcctcaagc ttcgacctgc agaagtaaca 7440
ccaaacaaca ggggtgagcat cgacaaaaga aacagtacca agcaaataaa tagcgtatga 7500
aggcagggct aaaaaaatcc acatatagct gctgcatatg ccatcatcca agtatatcaa 7560
gatcaaaaata attataaaac atacttgttt attataatag ataggtactc aaggttagag 7620
catatgaata gatgctgcat atgccatcat gtatatgcat cagtaaaacc cacatcaaca 7680
tgtataccta tcctagatcg atatttccat ccatcttaaa ctcgtaacta tgaagatgta 7740
tgacacacac atacagttcc aaaattaata aatacaccag gtagtttgaa acagtattct 7800
actccgatct agaacgaatg aacgaccgcc caaccacacc acatcatcac aaccaagcga 7860
acaaaagcat ctctgtatat gcatcagtaa aaccgcacac aacatgtata cctatcctag 7920
atcgatatat ccatccatca tcttcaattc gtaactatga atatgtatgg cacacacata 7980
cagatccaaa attaataaat ccaccaggta gtttgaaaca gaattctact ccgatctaga 8040
acgaccgccc aaccagacca catcatcaca accaagacaa aaaaaagcat gaaaagatga 8100
cccgacaaac aagtgcacgg catatatatga aataaaggaa aagggcaaac caaacctat 8160
gcaacgaaac aaaaaaaaaatc atgaaatcga tcccgtctgc ggaacggcta gagccatccc 8220
aggattcccc aaagagaaac actggcaagt tagcaatcag aacgtgtctg acgtacaggt 8280
cgcatccgtg tacgaacgct agcagcacgg atctaacaca aacacggatc taacacaaac 8340
atgaacagaa gtagaactac cgggccctaa ccatggaccg gaacgccgat ctagagaagg 8400
tagagagggg gggggaggac gagcggcgta ccttgaagcg gaggtgccga cgggttgatt 8460
tgggggagat ccactagtgc tagagcggcc gccaccgcgg tggaattctc gaggtcctct 8520
ccaaatgaaa tgaacttcct tatatagagg aagggtcttg cgaaggatag tgggattgtg 8580
cgtcatccct tacgtcagtg gagatatcac atcaatccac ttgctttgaa gacgtggttg 8640
gaacgtcttc tttttccacg atgctcctcg tgggtggggg tccatctttg ggaccactgt 8700
cggcagaggc atcttgaacg atagcctttc ctttatcgca atgatggcat ttgtaggtgc 8760
caccttcctt ttctactgtc cttttgatca agtgaccgat agctgggcaa tggaatccga 8820
ggaggtttcc cgatattacc ctttgttgaa aagtctcaat agccctttgg tcttctgaga 8880
ctgtatcttt gatattcttg gagtagacga gagtgtcgtg ctccaccatg ttatcacatc 8940
aattcacttg ctttgaagac gtggttgga cgtcttcttt ttccacgatg ctctctgtgg 9000
gtgggggtcc atctttggga ccactgtcgg cagaggcacc ttgaacgata gcctttcctt 9060
tatcgcaatg atggcatttg taggtgccac ctctcttttc tactgtcctt ttgatcaagt 9120
gacagatagc tgggcaatgg aatccgagga ggtttccga tattaccctt tgttgaaaag 9180
tctcaatagc cctttggtct tctgagactt gcaggcaagc aagcatgaat gcctggggga 9240
gaagaactcg agagggaatt gcagatcatg aggcatgg ctatTTTTgt gtcacatatg 9300
cgcaaaaaga gaggtatat ttgtgtccct aggttcttcg ttgtattgca gtttccatat 9360
caatctgact tggctcgcag agaaattgat ggttaaataa tttgaatctc tcatgtagta 9420
tcaactatta gatattattt tcaccaaata tatttccatc ggagaagaag aggctacaga 9480

```

ggaagcagaa gagaggggtg ggagaatttt tacacttttg tacaccact taaacagcaa 9540
 aatccgtatg aaaacaggcc caccaaaaca atgccacgat aacaatccgt agaaacaaaa 9600
 gcttcattta acagcggcgc aacaaagcac gcttatccat ggtagttgta gtccgtatgc 9660
 gatccaaaga tcacgattca cgcgtgacgg acggacgacg cgtgccacac cacaactaac 9720
 ggcatccatg gtagttgtag tccgtatgcg atccaaagat cacgattcac gcgtgacgga 9780
 cggacgacgc gcgccacacc acaactaaca gcgtgagcca gcgtccaaac tccgatggc 9840
 aacggggacg aaacccgtcg ggtagtcact gcccaaacc gtccccgcaa cttcatccc 9900
 aaacccgtcc ccgtttccgg tcgcggggtt cagttttcta ccagaccgt ccccatcggg 9960
 tttttcatcc ccgtcgggaa atccgaacc gccagcattt cagcaccaag ccaaagttgc 10020
 agcagcaaca tgaataaaaa acaacccgtt tcaacaccaa gataaaaca aacattataa 10080
 tttagacaac atttcacacg tataacaata acatatagtt ctacatatata acaacaccat 10140
 ttcacacata aaacaacacc atttgggata aaaatatggg ctatatcagg ccatttttat 10200
 gggccatatt gagttttcgt gggtttcaca ggtaccgat ttgtagaatg ctgaaccggg 10260
 tttgaaccgt aaaaatccgcg ggtattgaat ttgaccaat cccgtcgtcc cctggtgggg 10320
 taaaaacacc atcttgagtc caaacggcca ccaacaaac tccgacggca acaacaaac 10380
 ggcgttgctt tgctcctcgg tatctccgtg accgtcaat ctcccggtg tttccccgga 10440
 attgcgtgga ctctctcatc cacacgcaaa ccgcctctcc ctctctctc gtctatccg 10500
 ccccggtgcc gtagcctcac gggactctt tcctccctt gctataaaat ccccgcccc 10560
 tcccgctcc tctccacaca tccaaactct caatcgcacc gagaaaaatc tcctagcgat 10620
 cgaagcgaag cctctcccga tcctctcaag gtacgcccggt tcccgctcga tcctctctct 10680
 tccgttcgtg ttctgtagcc gatcgattcg attcccttac acccgttcgt gttctctcgt 10740
 ggatcgatcg attgtttgtt gctagaagga actcgtagat ctggcgttta tgaactgtga 10800
 ttcgggttag tccagatcga ttcaggctcg tcgtcgttga gcctctcggc tatgtctgga 10860
 ttatcgtgta gatctgctgg ttcagttgat tatgttctt taggagtaat ttcgttgggt 10920
 cagcgcgatt tctgcttaat ctatgctgct tattgcgcct gtacctatct actaagctat 10980
 gtgcacctgt aattttgcta gattattcgt tcatcctcgt agttggtttg tcacagtaat 11040
 ccgtatgggt tctgacgatg ttattgttgg tcatacctag gttctccag attttatttt 11100
 gttaaaattg gatagatctg ctactgatag ttgatgatgg aatttggtgc tgaatctatg 11160
 ctattttattg cgcctatacc tgatctatcg ggctatgtac ggctgtagtt tactggatta 11220
 ttcgttcac ctcggtagtt ggttcatcgt ttgggttctg acgataatat tgttgattat 11280
 gcgtaggctt ctgcagattg ttgttaaaat tggatacatc ggttactgat ggttgatgat 11340
 agatttgtgc tgaacctatc tgtttattgc tcctatacct gatctatagg gctatgtatg 11400
 cctgtaattt accagattat tcgttcaccc tcgtagttgg ttcactctta taattcgtat 11460
 gggttcttat gatgttatcg ttgattatgc ctagtcttat acagattatt gtgtcaagat 11520
 tgaatatacc tgctactgat cggtgataat ttggtttagta gtttgcaatc tgctaggaac 11580
 acgttaccac tgtaatctgt aaacatgggt tgccagagta gtttgttcta ctactcttga 11640
 tatggttgct gattttagtc gcctcctttt ggatcatgta ttgatgtcct tgcagatttc 11700
 cgtgtactta ccccggtttt tgtgtacttc gtgttaacag gtcgggtacc gaagcaaaca 11760
 tggcatctag catggcacca aagaaaaaaa ggaaagtttc caaacttgaa aaatttaca 11820

actgctactc cctttccaag acgcttaggt ttaaagcgat ccccgttggc aagacccaag 11880
 agaatatcga taacaaaaga cttctggctg aagatgaaaa aagggccgaa gactacaagg 11940
 ggggtcaagaa gttgctcgat cgctattatc tttcctttat caacgatgtg cttcattcaa 12000
 tcaaaactgaa gaacttgaat aactacatta gccttttcag aaagaaaacg aggactgaaa 12060
 aggagaacaa ggaacttgag aatcttgaaa taaaccttcg caaagaaatt gcaaaaagcct 12120
 tcaaggggaa cgaaggatat aaatctcttt tcaaaaaaga cattatagaa acaattttgc 12180
 ctgagtttct tgacgacaag gatgaaattg cgctcgtcaa tagctttaac ggattttaca 12240
 ctgccttcac aggggttctt gacaataggg agaatatgtt tagcgaggag gcaaaaagca 12300
 catccatcgc attcagatgc atcaatgaaa atcttaccgc gtacatatcg aatatggaca 12360
 tatttgaaaa agtggatgca atattcgata agcacgaagt ccaggagata aaggaaaaga 12420
 tactgaatag cgactatgat gtcgaagatt ttttcgaagg tgagttcttc aactttgtcc 12480
 tgactcaaga aggcattgat gtctataatg caataattgg aggttttgtg actgagtctg 12540
 gcgagaagat aaagggttg aacgagtata tcaatctcta caaccagaag actaagcaaa 12600
 agttgcctaa atttaaaccg ctttacaagc aagttttgag cgaccgggaa agcctttcct 12660
 tttacggtga aggatacacg agcgatgaag aagtctctga agtcttccgc aacacactca 12720
 acaagaactc agaaatcttt tcctcaatta aaaaattgga gaagcttttc aagaacttcg 12780
 atgaatactc ttcggcgggg atttttgtga agaacggccc ggcaatttcc acaatatcta 12840
 aagacatttt cggagaatgg aacgtgataa gagacaagtg gaatgcggag tatgatgaca 12900
 tacacctgaa gaagaaggca gttgtgactg aaaaatacga agatgacagg agaaaaagct 12960
 ttaaaaagat cgggtccttt tcaactggaac agctgcagga gtatgccgac gc 13012

<210> 282

<211> 3768

<212> DNA

<213> Lachnospiracea bacterium

<400> 282

atggcatcta gcatggcacc aaagaaaaaa aggaaagttt ccaaacttga aaaatttaca 60
 aactgctact ccctttccaa gacgcttagg tttaaagcga tccccgttg caagacccaa 120
 gagaatatcg ataacaaaag acttctggct gaagatgaaa aaagggccga agactacaag 180
 ggggtcaaga agttgctcga tcgctattat ctttccttta tcaacgatgt gcttcattca 240
 atcaaactga agaacttgaa taactacatt agccttttca gaaagaaaac gaggactgaa 300
 aaggagaaca aggaacttga gaatcttgaa ataaaccttc gcaaagaaat tgcaaaagcc 360
 ttcaagggga acgaaggata taaatctctt ttcaaaaaag acattataga aacaattttg 420
 cctgagtttc ttgacgacaa ggatgaaatt gcgctcgtca atagctttta cggattttaca 480
 actgccttca cagggttctt cgacaatagg gagaatatgt ttagcgagga ggcaaaaagc 540
 acatccatcg cattcagatg catcaatgaa aatcttacc ggtacatatc gaatatggac 600
 atatttgaaa aagtggatgc aatattcgat aagcacgaag tccaggagat aaaggaaaag 660
 atactgaata gcgactatga tgtcgaagat tttttcgaag gtgagttctt caactttgtc 720
 ctgactcaag aaggcattga tgtctataat gcaataattg gaggttttgt gactgagtct 780
 ggcgagaaga taaagggtt gaacgagtat atcaatctct acaaccagaa gactaagcaa 840

```

aagttgccta aatttaaacc gctttacaag caagttttga gcgaccggga aagcctttcc 900
ttttacgggtg aaggatacac gagcgatgaa gaagtcctcg aagtcctccg caacacactc 960
aacaagaact cagaaatctt ttcctcaatt aaaaaattgg agaagctttt caagaacttc 1020
gatgaatact cttcggcggg gatttttgtg aagaacggcc cggcaatttc cacaatatct 1080
aaagacattt tcggagaatg gaacgtgata agagacaagt ggaatgcgga gtatgatgac 1140
atacacctga agaagaaggc agttgtgact gaaaaatacg aagatgacag gagaaaaagc 1200
tttaaaaaaga tcgggtcctt ttcactggaa cagctgcagg agtatgccga cgccgatctt 1260
tcggttgtcg aaaagctcaa agaaataatt atccagaagg tcgatgaaat ctacaagggtg 1320
tacggctcaa gcgagaagct ctttgatgct gacttcgtgt tggagaagtc tcttaaaaaa 1380
aacgacgcag tcgtcgcgat aatgaaagat ttgctggatt cagtgaatc cttcgagaat 1440
tatatcaaag ccttcttcgg cgaggggaag gagacaaaca gggatgagtc cttctatgga 1500
gacttcgttc tggttacga catccttctt aaggtcgacc acatctatga cgcaattcgg 1560
aactatgtga cgcagaagcc gtattcgaaa gataagttca agctctatct ccaaaacct 1620
caatttatgg gtgggtggga taaagacaaa gagaccgatt accgggcaac aattttgcgg 1680
tacgggtcta aatattacct cgctataatg gataagaaat acgctaaatg tctccagaaa 1740
attgacaaaag atgacgtcaa cggcaattat gaaaaaatca attataaact ctttctggc 1800
ccaaataaaa tgctcccga ggtgtttttt tccaaaaagt ggatggccta ttataatcca 1860
tcagaggata ttcagaaaat ctataaaaat gggaccttta agaagggtga catgtttaac 1920
ctgaacgatt gccacaagct tatagatttt ttcaaagact ctattagccg ctatcccaaa 1980
tggtctaata cttatgattt caacttctct gaaactgaaa agtacaaga tattgcagga 2040
ttctaccgag aagtgaaga acaaggttat aaggtttcct ttgagtctgc gtccaagaaa 2100
gaggtcgata agttggtcga agaagggaat ttgtatatgt ttcaaattta caataaagac 2160
ttttccgaca agtcccatgg tacaccta atctgcatacca tgtacttcaa actgctgttc 2220
gatgagaata atcacgggtca gattcgcttg agcggagggg cggaactctt catgaggaga 2280
gcatcggtga aaaaagagga gctcgtcgtg catccggcta acagcccat tgctaacaag 2340
aatccggata atccaaagaa gactactacc ctctcctatg acgtctataa ggataagaga 2400
ttctctgagg accagtacga gttgcacatc cctattgcga taaataaatg ccctaagaac 2460
atcttttaaaa tcaatactga ggtcagagtc ctgcttaagc acgacgacaa cccgtatgtg 2520
atcgggattg ctaggggtga aaggaacttg ctttatattg tggttgtcga tggaaaaggt 2580
aatatagtg aacaatactc tctgaatgaa attatcaaca acttcaatgg cattaggatc 2640
aagaccgact atcattctct gttggacaag aaagagaaag agcgcttcga ggcacggcaa 2700
aactggacgt ctattgagaa catcaaggag ctttaaggctg gttacatttc tcaggttgtg 2760
cacaaaattt gcgaactggt cgagaaatat gatgccgtta tcgcacttga agatctcaac 2820
agcggattta agaattctcg ggtgaaagtc gaaaaacagg tgtatcaaaa attcgaaaag 2880
atgctgatcg acaagctcaa ttatatggtt gataaaaaga gcaacccatg cgccacgggg 2940
ggtgcgctta agggctatca gattacgaac aaatttgaat cttcaagtc aatgtcgacg 3000
caaaatgggt ttatatctta tataccggcg tggcttacat ctaaaataga tcttagcact 3060
gggttcgtga acctgctgaa aaccaagtac acttcaatcg cagattctaa aaaatttata 3120
agcagcttcg acagaatcat gtatgtgcc gaggaagacc tcttcgagtt tgcccttgat 3180

```

tacaaaaatt tctcaagaac ggatgcagac tacataaaga agtggaagct gtactcttat 3240
gggaaccgga ttcggatatt cagaaatccg aaaaaaaca atgtctttga ttgggaggaa 3300
gtttgtctta cctctgctta caaagagctg ttcaataaat atggcattaa ttaccagcaa 3360
ggtgatatcc gggcgctcct ttgcgaacag tctgacaaag ctttctattc ttcatttatg 3420
gcgctcatgt cattgatgct gcagatgagg aatagcatta cggggaggac tgatgttgac 3480
tttctgatct cgcccgtgaa aaattctgat ggaatcttct acgattccag gaattatgag 3540
gcccaggaaa atgctatcct tcccaagaac gcagacgcaa atggcgcgta caatatagct 3600
cgcaaggttt tgtgggctat aggccaattc aagaaagccg aagacgaaaa gctggacaaa 3660
gttaagattg ctatatctaa caaagagtgg cttgagtatg cgcaaacatc tgttaaacac 3720
aaacgccccg cggctacaaa gaaggctggc caggccaaga agaagaag 3768

<210> 283

<211> 3768

<212> DNA

<213> 人工序列

<220>

<223> LbCpf1_RR

<400> 283

atggcatcta gcatggcacc aaagaaaaaa aggaaagttt ccaaacttga aaaatttaca 60
aactgctact ccctttccaa gacgcttagg tttaaagcga tccccgttg caagacccaa 120
gagaatatcg ataacaaaag acttctggtc gaagatgaaa aaagggccga agactacaag 180
ggggtcaaga agttgctcga tcgctattat ctttccttta tcaacgatgt gcttcattca 240
atcaaactga agaacttgaa taactacatt agccttttca gaaagaaaac gaggactgaa 300
aaggagaaca aggaacttga gaatcttgaa ataaaccttc gcaaagaaat tgcaaaagcc 360
ttcaagggga acgaaggata taaatctctt ttcaaaaaag acattataga aacaattttg 420
cctgagtttc ttgacgacaa ggatgaaatt gcgctcgtca atagctttaa cggatttaca 480
actgccttca cagggttctt cgacaatagg gagaatatgt ttagcgagga ggcaaaaagc 540
acatccatcg cattcagatg catcaatgaa aatcttaccg ggtacatatc gaatatggac 600
atatattgaaa aagtggatgc aatattcgat aagcacgaag tccaggagat aaaggaaaag 660
atactgaata gcgactatga tgtcgaagat ttttcgaag gtgagttctt caactttgtc 720
ctgactcaag aaggcattga tgtctataat gcaataattg gaggttttgt gactgagtct 780
ggcgagaaga taaagggtt gaacgagtat atcaatctct acaaccagaa gactaagcaa 840
aagttgccta aatttaaacc gctttacaag caagttttga gcgaccgga aagcctttcc 900
ttttacggtg aaggatacac gagcgatgaa gaagtctcgc aagtcttccg caacacactc 960
aacaagaact cagaaatctt ttcttcaatt aaaaaatttg agaagctttt caagaacttc 1020
gatgaatact cttcggcggg gattttttgtg aagaacggcc cggcaatttc cacaatatct 1080
aaagacattt tcggagaatg gaacgtgata agagacaagt ggaatgcgga gtatgatgac 1140
atacacctga agaagaaggc agttgtgact gaaaaatacg aagatgacag gagaaaaagc 1200
tttaaaaaga tcgggtcctt ttcactggaa cagctgcagg agtatgccga cgccgatctt 1260
tcggttgtcg aaaagctcaa agaaataatt atccagaagg tcgatgaaat ctacaaggtg 1320

tacggctcaa gcgagaagct ctttgatgct gacttcgtgt tggagaagtc tcttaaaaaa 1380
aacgacgcag tcgtcgcgat aatgaaagat ttgctggatt cagtgaaatc cttcgagaat 1440
tatatcaaag ccttcttcgg cgaggggaag gagacaaaca gggatgagtc cttctatgga 1500
gacttcgttc tggcttacga catccttctt aaggtcgacc acatctatga cgcaattcgg 1560
aactatgtga cgcagaagcc gtattcgaaa gataagttca agctctatct ccaaaaccc 1620
caatttatgc gtgggtggga taaagacaaa gagaccgatt accgggcaac aattttgctg 1680
tacgggtcta aatattacct cgctataatg gataagaaat acgctaaatg tctccagaaa 1740
attgacaaaag atgacgtcaa cggcaattat gaaaaaatca attataaact ctttcctggc 1800
ccaaataaaa tgctcccag ggtgtttttt tccaaaaagt ggatggccta ttataatcca 1860
tcagaggata ttcagaaaaat ctataaaaaat gggaccttta agaagggtga catgtttaac 1920
ctgaacgatt gccacaagct tatagatttt ttcaaagact ctattagccg ctatcccaaa 1980
tggtctaattg cttatgattt caacttctct gaaactgaaa agtacaaga tattgcagga 2040
ttctaccgcg aagtgaaga acaagggtat aaggtttctt ttgagtctgc gtccaagaaa 2100
gaggtcgata agttggtcga agaagggaat ttgtatatgt ttcaaattta caataagac 2160
ttttccgaca agtcccatgg tacacctaat ctgcatacca tgtacttcaa actgctgttc 2220
gatgagaata atcacggtca gattcgctg agcggagggg cggaactctt catgaggaga 2280
gcatcggtga aaaaagagga gctcgtcgtg catccggcta acagcccat tgctaacaag 2340
aatccggata atccaaagaa gactactacc ctctcctatg acgtctataa ggataagaga 2400
ttctctgagg accagtacga gttgcacatc cctattgcga taaataaatg ccctaagaac 2460
atctttaaaa tcaatactga ggtcagagtc ctgcttaagc acgacgacaa cccgtatgtg 2520
atcgggattg ctaggggtga aaggaaactg ctttatattg tggttgtcga tggaaaaggt 2580
aatatagtg aacaatactc tctgaatgaa attatcaaca acttcaatgg cattaggatc 2640
aagaccgact atcattctct gttggacaag aaagagaaag agcgcttca ggcacggcaa 2700
aactggacgt ctattgagaa catcaaggag cttaggctg gttacatttc tcaggttgtg 2760
cacaaaattt gcgaactggt cgagaaatat gatgccgta tcgcacttga agatctcaac 2820
agcggattta agaattctcg ggtgaaagtc gaaaaacagg tgtatcaaaa attcgaaaag 2880
atgctgatcg acaagctcaa ttatatggtt gataaaaaga gcaacccatg cgccacgggg 2940
ggtgcgctta agggctatca gattacgaac aaatttgaat cttcaagtc aatgtcgacg 3000
caaatgggt ttatatctta tataccggcg tggtttacat ctaaaataga tcctagcact 3060
gggttcgtga acctgctgaa aaccaagtac acttcaatcg cagattctaa aaaatttata 3120
agcagcttcg acagaatcat gtatgtgccc gaggaagacc tcttcgagtt tgcccttgat 3180
tacaaaaatt tctcaagaac ggatgcagac tacataaaga agtggaagct gtactcttat 3240
gggaaccgga ttcggatatt cagaaatccg aaaaaaaca atgtctttga ttgggaggaa 3300
gtttgtctta cctctgctta caaagagctg ttcaataaat atggcattaa ttaccagcaa 3360
ggtgatatcc gggcgctcct ttgcgaacag tctgacaaag ctttctatc ttcatctatg 3420
gcgctcatgt cattgatgct gcagatgagg aatagcatta cggggaggac tgatgttgac 3480
tttctgatct cgcccgtaa aaattctgat ggaatcttct acgattccag gaattatgag 3540
gccaggaaa atgctatcct tccaagaac gcagacgcaa atggcgcgta caatatagct 3600
cgcaaggttt tgtgggctat aggccaattc aagaaagccg aagacgaaaa gctggacaaa 3660

gttaagattg ctatatctaa caaagagtgg cttgagtatg cgcaaacatc tgttaaacac 3720
aaacgccccg cggctacaaa gaaggctggc caggccaaga agaagaag 3768
<210> 284
<211> 3768
<212> DNA
<213> 人工序列
<220>
<223> LbCpf1_RVR
<400> 284
atggcatcta gcatggcacc aaagaaaaaa aggaaagttt ccaaacttga aaaattttaca 60
aactgctact cccitttccaa gacgcttagg tttaaagcga tccccgttgg caagacccaa 120
gagaatatcg ataacaaaag acttctggtc gaagatgaaa aaagggccga agactacaag 180
ggggtcaaga agttgctcga tcgctattat ctttcttta tcaacgatgt gcttcattca 240
atcaaactga agaacttgaa taactacatt agccttttca gaaagaaaac gaggactgaa 300
aaggagaaca aggaacttga gaatcttgaa ataaaccttc gcaaagaaat tgcaaaagcc 360
ttcaagggga acgaaggata taaatctctt ttcaaaaaag acattataga aacaattttg 420
cctgagtttc ttgacgacaa ggatgaaatt gcgctcgtca atagctttta cggattttaca 480
actgccttca cagggttctt cgacaatagg gagaatatgt ttagcgagga ggcaaaaagc 540
acatccatcg cattcagatg catcaatgaa aatcttaccg ggtacatatc gaatatggac 600
atatattgaaa aagtggatgc aatattcgat aagcacgaag tccaggagat aaaggaaaag 660
atactgaata gcgactatga tgtcgaagat tttttcgaag gtgagttctt caactttgtc 720
ctgactcaag aaggcattga tgtctataat gcaataattg gaggttttgt gactgagtct 780
ggcgagaaga taaagggcctt gaacgagtat atcaatctct acaaccagaa gactaagcaa 840
aagttgccta aatttaaacc gctttacaag caagttttga gcgaccggga aagcctttcc 900
ttttacggtg aaggatacac gagcgatgaa gaagtcctcg aagtcttccg caacacactc 960
aacaagaact cagaaatctt ttcctcaatt aaaaaattgg agaagctttt caagaacttc 1020
gatgaatact cttcggcggg gatTTTTgtg aagaacggcc cggcaatttc cacaatatct 1080
aaagacattt tcggagaatg gaacgtgata agagacaagt ggaatgcgga gtatgatgac 1140
atacacctga agaagaagcc agttgtgact gaaaaatacg aagatgacag gagaaaaagc 1200
tttaaaaaaga tcgggtcctt ttcaactggaa cagctgcagg agtatgccga cgccgatctt 1260
tcggttgtcg aaaagctcaa agaaataatt atccagaagg tcgatgaaat ctacaaggtg 1320
tacggctcaa gcgagaagct ctttgatgct gacttcgtgt tggagaagtc tcttaaaaaa 1380
aacgacgcag tcgtcgcgat aatgaaagat ttgctggatt cagtgaatc cttcgagaat 1440
tatatcaaag ctttcttcgg cgaggggaag gagacaaaca gggatgagtc cttctatgga 1500
gacttcgttc tggcttacga catccttctt aaggtcgacc acatctatga cgcaattcgg 1560
aactatgtga cgcagaagcc gtattcgaaa gataagtta agctctatct ccaaaacctt 1620
caatttatgc gtgggtggga taaagacgta gagaccgatc gccgggcaac aattttgcgg 1680
tacgggtcta aatattacct cgctataatg gataagaaat acgctaaatg tctccagaaa 1740
attgacaaaag atgacgtcaa cggcaattat gaaaaaatca attataaact ctttctggtc 1800

```

ccaaataaaa tgctcccgaa ggtgtttttt tccaaaaagt ggatggccta ttataatcca 1860
tcagaggata ttcagaaaat ctataaaaat gggaccttta agaagggtga catgtttaac 1920
ctgaacgatt gccacaagct tatagatttt ttcaaagact ctattagccg ctatcccaaa 1980
tggtctaata cttatgattt caacttctct gaaactgaaa agtacaaaga tattgcagga 2040
ttctaccgcg aagttgaaga acaagggttat aaggtttctt ttgagtctgc gtccaagaaa 2100
gaggtcgata agttggtcga agaagggaaa ttgtatatgt ttcaaattta caataaagac 2160
ttttccgaca agtcccatgg tacacctaat ctgcatacca tgtacttcaa actgctgttc 2220
gatgagaata atcacgggtca gattcgcttg agcggagggg cggaactctt catgaggaga 2280
gcacggttga aaaaagagga gctcgtcgtg catccggcta acagcccat tgctaacaag 2340
aatccggata atccaaagaa gactactacc ctctctatg acgtctataa ggataagaga 2400
ttctctgagg accagtacga gttgcacatc cctattgcga taaataaatg ccctaagaac 2460
atctttaaaa tcaatactga ggtcagagtc ctgcttaagc acgacgaca cccgtatgtg 2520
atcgggattg ctaggggtga aaggaaactg ctttatattg tggttgctga tggaaaaggt 2580
aatatagtgg aacaatactc tctgaatgaa attatcaaca acttcaatgg cattaggatc 2640
aagaccgact atcattctct gttggacaag aaagagaaag agcgcttcga ggcacggcaa 2700
aactggacgt ctattgagaa catcaaggag cttaggctg gttacatttc tcaggttgtg 2760
cacaaaattt gcgaactggt cgagaaatat gatgccgtta tcgcacttga agatctcaac 2820
agcggattta agaattctcg ggtgaaagtc gaaaaacagg tgtatcaaaa attcgaaaag 2880
atgctgatcg acaagctcaa ttatatggtt gataaaaaga gcaacccatg cgccacgggg 2940
ggtgcgctta agggctatca gattacgaac aaatttgaat cettcaagtc aatgtcgacg 3000
caaaatgggt ttatatctta tataccggcg tggtttacat ctaaaataga tcctagcact 3060
gggttcgtga acctgctgaa aaccaagtac acttcaatcg cagattctaa aaaatttata 3120
agcagcttcg acagaatcat gtatgtgccc gaggaagacc tcttcgagtt tgcccttgat 3180
tacaaaaatt tctcaagaac ggatgcagac tacataaaga agtggaagct gtactcttat 3240
gggaaccgga ttcggatatt cagaaatccg aaaaaaaca atgtctttga ttgggaggaa 3300
gtttgtctta cctctgctta caaagagctg ttcaataaat atggcattaa ttaccagcaa 3360
ggtgatatac gggcgctcct ttgcgaacag tctgacaaag ctttctattc ttcatthtg 3420
gcgctcatgt cattgatgct gcagatgagg aatagcatta cggggaggac tgatgttgac 3480
tttctgatct cgcccgtaaa aaattctgat ggaattctt acgattccag gaattatgag 3540
gcccaggaaa atgctatcct tcccaagaac gcagacgcaa atggcgcgta caatatagct 3600
cgcaagggtt tgtgggctat aggccaattc aagaaagccg aagacgaaaa gctggacaaa 3660
gttaagattg ctatatctaa caaagagtgg cttgagtatg cgcaaacatc tgtaaacac 3720
aaacgccccg cggtacaaa gaaggtggc caggccaaga agaagaag 3768

```

<210> 285

<211> 13012

<212> DNA

<213> 人工序列

<220>

<223> pGEP767 expression plasmid

<400> 285

agcatgaatg cctgggggag aagaactcga gaggaattg cagatcatga ggcagatggc 60
tattttttgtg tcacatatgc gcaaaaagag aggctatatt tgtgtcccta ggttcttcgt 120
tgtattgcag tttccatatc aatctgactt ggtcgcatga gaaattgatg gttaaataat 180
ttgaatctct catgtagtat caactattag atattatatt caccaaataat atttccatcg 240
gagaagaaga ggctacagag gaagcagaag agaggggtgg gagaattttt acacttttgt 300
acacccactt aaacagcaaa atccgtatga aaacaggccc accaaaacaa tgccacgata 360
acaatccgta gaaacaaaaag cttcatttaa cagcggcgca acaaagcacg cttatccatg 420
gtagttgtag tccgtatgcg atccaaagat cagcattcac gcgtgacgga cggacgacgc 480
gtgccacacc acaactaacg gcatccatgg tagttgtagt ccgtatgcga tccaaagatc 540
acgattcacg cgtgacggac ggacgacgcg cgccacacca caactaacag cgtgagccag 600
cgtccaaact ccgatggca acggggacga aaccgctcgg gtagtcactg cccaaacccg 660
tccccgaac cttcatccca aaccgctccc cgtttccggt cgcgggtttc agttttctac 720
cagaccgctc cccatcgggt ttttcatccc cgtcgggaaa tccgaacccg ccagcatttc 780
agcaccaagc caaagttgca gcagcaacat gaataaaaaa caaccgttt caacaccaag 840
ataaaacaaa acattataat ttagacaaca tttcacacgt ataacaataa catatagttc 900
tcacatataa caacaccatt tcacacataa aacaacacca tttgggataa aaatatgggc 960
tatatcaggc catttttatg ggccatattg agttttcgtg ggtttcacag gtaccggatt 1020
tgtagaatgc tgaaccgggt ttgaaccgta aaatccgcgg gtattgaatt tgaccaatc 1080
ccgtcgtccc ctggtggggt aaaaacacca tcttgagtcc aaacggccac caaccaaact 1140
ccgacggcaa caaacaacg gcgttgcttt gctcctcgtt atctccgtga ccgtcaatc 1200
tcccggctgt tccccggaa ttgcgtggac tctctcatcc acacgcaaac cgcctctccc 1260
tcctctctcg tcctatccgc cccggtgccg tagcctcacg ggactcttct tcctcccttg 1320
ctataaaatc cccgccccct cccgtctcct ctccacacat ccaaactctc aatcgcaccg 1380
agaaaaatct cctagcgatc gaagcgaagc ctctcccgat cctctcaagg tacgcccggt 1440
tcccgtcgat cctcctcctt ccgttcgtgt tctgtagccg atcgattcga ttcccttaca 1500
cccgttcgtg ttctctcgtg gatcgatcga ttgtttgttg ctagaaggaa ctcgtagatc 1560
tggcgtttat gaactgtgat tcgggttagt ccagatcgat tcaggtcggg cgtcgttgag 1620
cctctcggct atgtctggat tatcgtgtag atctgctggt tcagttgatt atgttcttct 1680
aggagtaatt tcgttgggtc agcgcgattt ctgcttaatc tatgctgctt attgcgcctg 1740
tacctatcta ctaagctatg tgcacctgta attttgctag attattcgtt catcctcgta 1800
gttggtttgt cacagtaatc cgtatgggtt ctgacgatgt tattgttggt catacctagg 1860
cttctccaga ttttattttg ttaaaattgg atagatctgc tactgatagt tgatgatgga 1920
atttggtgct gaatctatgc tatttattgc gcctatacct gatctatcgg gctatgtacg 1980
gctgtagttt actggattat tcgttcatcc tcggtagttg gttcatcggt tgggttctga 2040
cgataatatt gttgattatg cgtaggttc tgcagattgt tgttaaaatt ggatacatcg 2100
gttactgatg gttgatgata gatttgtgct gaacctatct gtttattgct cctatacctg 2160
atctataggg ctatgtatgc ctgtaattta ccagattatt cgttcatcct cgtagttggt 2220
tcactcttat aattcgtatg ggttcttatg atgttatcgt tgattatgcc tagtcttata 2280

cagattattg tgtcaagatt gaatatacct gctactgac ggtgataatt tggttagtag 2340
 tttgcaatct gctaggaaca cgttaccact gtaatctgta aacatggttt gccagagtag 2400
 tttgttctac tactcttgat atggttgctg attttagtcg cctccttttg gatcatgtat 2460
 tgatgtcctt gcagatttcc gtgtacttac cccggctttt gtgtacttcg tgtaaacagg 2520
 tcgggtaccg aagcaaacaat ggcatctagc atggcaccaa agaaaaaag gaaagtttcc 2580
 aaacttgaaa aatttacaaa ctgctactcc ctttccaaga cgcttaggtt taaagcgatc 2640
 cccgttggca agacccaaga gaatatcgat aacaaaagac ttctggtcga agatgaaaaa 2700
 agggccgaag actacaaggg ggtcaagaag ttgctcgatc gctattatct ttcctttatc 2760
 aacgatgtgc ttcatcatt caaactgaag aacttgaata actacattag ctttttcaga 2820
 aagaaaacga ggactgaaaa ggagaacaag gaacttgaga atcttgaaat aaaccttcgc 2880
 aaagaaattg caaaaagcctt caaggggaac gaaggatata aatctctttt caaaaaagac 2940
 attatagaaa caattttgcc tgagtttctt gacgacaagg atgaaattgc gctcgtcaat 3000
 agctttaacg gatttacaaac tgccttcaca gggttcttcg acaataggga gaatatgttt 3060
 agcgaggagg caaaaagcac atccatcgca ttcagatgca tcaatgaaaa tcttaccggg 3120
 tacatatcga atatggacat atttgaaaaa gtggatgcaa tattcgataa gcacgaagtc 3180
 caggagataa aggaaaagat actgaatagc gactatgatg tcgaagattt tttcgaaggt 3240
 gagttcttca actttgtcct gactcaagaa ggcattgatg tctataatgc aataattgga 3300
 ggttttgtga ctgagtctgg cgagaagata aagggttga acgagtatat caatctctac 3360
 aaccagaaga ctaagcaaaa gttgcctaaa tttaaaccgc tttacaagca agttttgagc 3420
 gaccgggaaa gcctttcctt ttacggtgaa ggatacacga gcgatgaaga agtcctcgaa 3480
 gtcttccgca acacactcaa caagaactca gaaatctttt cctcaattaa aaaattggag 3540
 aagcttttca agaacttcca tgaatactct tcggcgggga tttttgtgaa gaacggcccc 3600
 gcaatttcca caatatctaa agacatttcc ggagaatgga acgtgataag agacaagtgg 3660
 aatgcggagt atgatgacat acacctgaag aagaaggcag ttgtgactga aaaatacgaa 3720
 gatgacagga gaaaaagcct taaaaagatc gggtcctttt cactggaaca gctgcaggag 3780
 tatgccgacg ccgatctttc ggttgtcgaa aagctcaaag aaataattat ccagaaggtc 3840
 gatgaaatct acaagggtga cggtcgaagc gagaagctct ttgatgctga cttcgtgttg 3900
 gagaagtctc ttaaaaaaaaa cgacgcagtc gtgcgataa tgaaagattt gctggattca 3960
 gtgaaatcct tcgagaatta tatcaaagcc ttcttcggcg aggggaagga gacaaacagg 4020
 gatgagtcct tctatggaga cttcgttctg gcttacgaca tccttcttaa ggtcgaccac 4080
 atctatgacg caattcggaa ctatgtgacg cagaagccgt attcgaaaga taagttcaag 4140
 ctctatttcc aaaaccctca atttatgggt ggggtgggata aagacaaaga gaccgattac 4200
 cgggcaacaa ttttgcggtg cgggtctaaa tattacctcg ctataatgga taagaaatac 4260
 gctaaatgtc tccagaaaaa tgacaaagat gacgtcaacg gcaattatga aaaaatcaat 4320
 tataaactcc ttcctggccc aaataaaatg ctcccgaagg tgtttttttc caaaaagtgg 4380
 atggcctatt ataatccatc agaggatatt cagaaaatct ataaaaatgg gacctttaag 4440
 aagggtgaca tgtttaacct gaacgattgc cacaagctta tagatttttt caaagactct 4500
 attagccgct atcccaaatg gtctaagtct tatgatttca acttctctga aactgaaaag 4560
 taaaaagata ttgcaggatt ctaccgcgaa gttgaagaac aaggttataa ggtttccttt 4620

gagtctgcgt ccaagaaaga ggtcgataag ttggtcgaag aagggaatt gtatatgttt 4680
 caaatattaca ataaagactt ttccgacaag tcccatggta cacctaatt gcataccatg 4740
 tacttcaaac tgctgttcga tgagaataat cacggtcaga ttgcctgag cggaggggag 4800
 gaactcttca tgaggagagc atcgttgaaa aaagaggagc tcgtcgtgca tccggctaac 4860
 agccccattg ctaacaagaa tccggataat ccaaagaaga ctactaccct ctctatgac 4920
 gtctataagg ataagagatt ctctgaggac cagtacgagt tgcacatccc tattgagata 4980
 aataaatgcc ctaagaacat ctttaaaatc aatactgagg tcagagtcct gcttaagcac 5040
 gacgacaacc cgtatgtgat cgggattgct aggggtgaaa ggaacttgct ttatattgtg 5100
 gttgtcgatg gaaaaggtaa tatagtggaa caatactctc tgaatgaaat tatcaacaac 5160
 ttcaatggca ttaggatcaa gaccgactat cattctctgt tggacaagaa agagaaagag 5220
 cgcttcgagg cacggcaaaa ctggacgtct attgagaaca tcaaggagct taaggctggt 5280
 tacatttctc aggttgtgca caaaatttgc gaactggctg agaaatatga tgccgttatt 5340
 gcacttgaag atctcaacag cggatttaag aattctcggg tgaaagtcga aaaacaggtg 5400
 tatcaaaaat tcgaaaagat gctgatcgac aagctcaatt atatggttga taaaagagc 5460
 aacccatgag ccacgggggg tgctgcttaag ggctatcaga ttacgaacaa atttgaatcc 5520
 ttcaagtcaa tgcgacgca aaatgggttt atattctata taccggcgtg gcttacatct 5580
 aaaatagatc ctagcactgg gttcgtgaac ctgctgaaaa ccaagtacac ttcaatcgca 5640
 gattctaaaa aatttataag cagcttcgac agaactcatgt atgtgcccga ggaagacctc 5700
 ttcgagtttg cccttgatta caaaaatttc tcaagaacgg atgcagacta cataaagaag 5760
 tggaagctgt actcttatgg gaaccggatt cggatattca gaaatccgaa aaaaaacaat 5820
 gtctttgatt gggaggaagt ttgtcttacc tctgcttaca aagagctggt caataaatat 5880
 ggcatataat accagcaagg tgatatccgg gcgctccttt gcgaacagtc tgacaaagct 5940
 ttctattctt catttatggc gctcatgtca ttgatgctgc agatgaggaa tagcattacg 6000
 gggaggactg atgttgactt tctgatctcg cccgtgaaaa attctgatgg aatcttctac 6060
 gattccagga attatgaggc ccaggaaaat gctatccttc ccaagaacgc agacgcaaat 6120
 ggcgcgtaca atatagctcg caaggttttg tgggctatag gccaattcaa gaaagccgaa 6180
 gacgaaaagc tggacaaaag taagattgct atatctaaca aagagtggct tgagtatgag 6240
 caaacatctg ttaaaccaca acgccccgag gctacaaaaga aggctggcca ggccaagaag 6300
 aagaagggtc cggggctcgg gtcgggctcg ggctcggacg ccctggacga cttcgacctc 6360
 gacatgctgg gctccgacgc cctcgatgat ttgacctcg atatgctcgg cagcgacgag 6420
 ctcgatgact tcgacctcga tatgctgggg agcgacgccc tcgacgattt tgacctcgat 6480
 atgctgatca actcccgtc cagcggcagc ccgaagaaga agcgcaaagt gggctcgcag 6540
 tacctgcccc acaccgacga caggcacagg atcgaggaga agcgcaagag gacgtacgag 6600
 accttcaagt ccatcatgaa gaagtccccg ttacgcggcc caacggaccc ccgccccgag 6660
 ccgaggagga tcgccgtgcc gtccaggtcc agcgcgtcgg tccccagcc ggccccgcag 6720
 ccctaccgtg tcacgtccag cctcagcacc atcaactacg acgagttccc caccatggtg 6780
 ttcccgtccg gccagatctc ccaggccagc gcgctggccc ccgcgcccc gcaggtgctg 6840
 cccaggtc cgcccccg tccggccccg gccatggtct ccgcgctggc ccaggcgcgc 6900
 gccccggtgc ccgtcctcgc gccgggcccc ccgcaggcgg tcgccccgcc agcgccgaag 6960

```

cccacgcagg ccggcgaggg caccctcagc gaggcgctcc tgcagctgca gttcgacgac 7020
gaggacctcg gcgcccctct gggcaactcg accgaccccc ccgtgttcac cgacctggcc 7080
tccgtcgaca acagcgagtt ccagcagctg ctgaaccagg gcatccccgt ggcgccgcac 7140
accacggagc ccatgctgat ggagtacccg gaggcgatca cgcgcctcgt caccggcgcc 7200
cagaggcccc cggacccccg cccggccccg ctcggcgccc caggcctgcc gaacggcctc 7260
ctgagcggcg acgaggactt ctccagcatc gcggacatgg acttctccgc cctcctgggg 7320
tcgggctcgg gcagccgcga cagcaggag ggcatgttcc tcccaaagcc cgaggccggc 7380
tccgccatct cggacgtgtt cgagggcagg gaggtctgcc agccaaagcg catcaggccg 7440
ttccacccgc cgggctcccc gtgggcgaac cggccgctcc ccgccagcct ggctccaacc 7500
ccgaccggcc ccgtgcacga gccggtcggc agcctgacgc ccgcgccgt gccccagccg 7560
ctcgaccccc cgccggccgt ccccccgag gcctcccacc tcttgagga ccccgacgag 7620
gagacctcgc aggccgtgaa ggccctgagg gagatggccg acaccgtcat cccccagaag 7680
gaggaggcgg ccatctgcgg ccagatggac ctgtcgcacc cgccgccgcg cggccacctc 7740
gacgagctga ccacgaccct cgagtccatg accgaggacc tcaacctgga cagccccctc 7800
acgccggagc tgaacgagat cctcgacacc ttctgaacg acgagtgcct cctgcacgcc 7860
atgcacatct ccacgggcct gagcatcttc gacaccagcc tcttctgagt cgaccgatcg 7920
ttcaaacatt tggcaataaa gtttcttaag attgaatcct gttgccggtc ttgcgatgat 7980
tatcatataa tttctgttga attacgttaa gcatgtaata attaacatgt aatgcatgac 8040
gttatttatg agatgggttt ttatgattag agtcccgcaa ttatacattt aatacgcat 8100
agaaaacaaa atatagcgcg caaactagga taaattatcg cgcgcggtgt catctatgtt 8160
actagatcga tcccgggata tcgcgccgg tcgttcggct gcggcgagcg gtatcagctc 8220
actcaaaggc ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt 8280
gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc 8340
ataggctccg cccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa 8400
acccgacagg actataaaga taccaggcgt ttccccctgg aagctccctc gtgcgctctc 8460
ctgttccgac cctgccgctt accggatacc tgtccgcctt tctcccttcg ggaagcgtgg 8520
cgctttctca tagctcacgc ttaggtatc tcagttcggg ttaggtcgtt cgctccaagc 8580
tgggctgtgt gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc ggtaactatc 8640
gtcttgagtc caacccggtg agacacgact tatcgccact ggcagcagcc actggtaaca 8700
ggattagcag agcgaggtat gtaggcggtg ctacagagtt cttgaagtgg tggcctaact 8760
acggctacac tagaaggaca gtatttggtg tctgcgctct gctgaagcca gttaccttcg 8820
gaaaaagagt tggtagctct tgatccggca aacaaaccac cgctggtagc ggtgggtttt 8880
ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat cttttgatct 8940
tttctacggg gtctgacgct cagtggaacg aaaactcacg ttaagggtt ttggtcatga 9000
gattatcaaa aaggatcttc acctagatcc ttttaaatta aaaatgaagt tttaaatcaa 9060
tctaaagtat atatgagtaa acttggtctg acagttacca atgcttaatc agtgaggcac 9120
ctatctcagc gatctgtcta tttcgttcat ccatagttgc ctgactcccc gtcgtgtaga 9180
taactacgat acgggagggc ttaccatctg gcccagtgcc tgcaatgata ccgcgagacc 9240
cacgctcacc ggctccagat ttatcagcaa taaaccagcc agccggaagg gccgagcgca 9300

```

gaagtgggtcc tgcaacttta tccgcctcca tccagtctat taattgttgc cgggaagcta 9360
gagtaagtag ttcgccagtt aatagtttgc gcaacgttgt tgccattgct acaggcatcg 9420
tggtgtcacg ctcgtcgttt ggtatggctt cattcagctc cggttcccaa cgatcaaggc 9480
gagttacatg atcccccatg ttgtgcaaaa aagcggttag ctccttcggt cctccgatcg 9540
ttgtcagaag taagtgggcc gcagtgttat cactcatggt tatggcagca ctgcataatt 9600
ctcttactgt catgccatcc gtaagatgct tttctgtgac tggtagtac tcaaccaagt 9660
cattctgaga atagtgtatg cggcgaccga gttgctcttg cccggcgtca atacgggata 9720
ataccgcgcc acatagcaga actttaaaag tgctcatcat tggaaaacgt tcttcggggc 9780
gaaaactctc aaggatctta ccgctgttga gatccagttc gatgtaacct actcgtgcac 9840
ccaactgac ttcagcatct tttactttca ccagcgtttc tgggtgagca aaaacaggaa 9900
ggcaaaatgc cgcaaaaaag ggaataaggg cgacacggaa atgttgaata ctcatactct 9960
tcctttttca atattattga agcatttatc agggttattg tctcatgagc ggatacatat 10020
ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc cgaaaagtgc 10080
cacctgacgc gccctgtagc ggacagctta attcggggga tctggatttt agtactggat 10140
tttggtttta ggaattagaa attttattga tagaagtatt ttacaaatac aaatacatac 10200
taagggtttc ttatatgctc aacacatgag cgaaacccta taggaaccct aattccctta 10260
tctgggaact actcacacat tattatggag aaactcgagc ttgtcgatcg acatgatcag 10320
ggagccctag attatttcta tagttcatcc atgcccatta cgtcggtaaa tgccttctgc 10380
cactccttga agttaagttc ggtcttgga tgtttcaact cagtcttacg gaacacgtac 10440
atgggttggg tcttaaggta gttagcggcc attggtttag cgaatgtgta ggtagtcctg 10500
gctgtagagc gatatctctt gccattgcct gtgggtgaag accatttgaa ggtactaatg 10560
atggctctgt cgtagggta ggttttcttg gaccggcacc aatcagcggc agttaaggag 10620
ttggctcatga caggtcctc agcaggaaa cctgtccctt tcaattgggc ttctcctttg 10680
atgtggctcc cttcgtaagt gtaacggtag ttgacggtga gcgaagcacc gtcctcaaac 10740
tgcatgtgcc tgtggacttg gtatccggag ccatcaacca tggctgcttg gaatggactc 10800
attccgtcag ggtaggaag gtattgatgg aatccgtagc caatgtgtgg caccagaatc 10860
catggagaaa actgaagatc acctttggtg ctcttgaggt tcagctcttc gtatccgtca 10920
ttaggggtcc cagtgccttg tccgaccata tcgaagtcaa cgccgttgat ggaaccgaag 10980
atgtgaagct catgtgtggc tggaagcgaa gccatgttat cttcttctcc ttactcacg 11040
gaggacgcca tgggtggcggg atcgcgcctt atcgttcgta aatgggtgaa attttcagaa 11100
aattgctttt gctttaaaaag aaatgattta aattgctgca atagaagtag aatgcttgat 11160
tgcttgagat tcgtttgttt tgtatatgtt gtgttgagag gatcctctag agtcgacctg 11220
cagaagtaac accaaacaac agggtagca tcgacaaaag aaacagtacc aagcaaataa 11280
atagcgtatg aaggcagggc taaaaaaatc cacatatagc tgctgcatat gccatcatcc 11340
aagtatatca agatcaaaaat aattataaaa catacttggt tattataata gataggtact 11400
caaggtaga gcatatgaat agatgctgca tatgcatca tgtatatgca tcagtaaaac 11460
ccacatcaac atgtatacct atcctagatc gatatttcca tccatcttaa actcgtaact 11520
atgaagatgt atgacacaca catacagttc caaaattaat aaatacacca ggtagtttga 11580
aacagtattc tactccgatc tagaacgaat gaacgaccgc ccaaccacac cacatcatca 11640

caaccaagcg aacaaaagca tctctgtata tgcattcagta aaacccgcat caacatgtat 11700
 acctatccta gatcgatatt tccatccatc atcttcaatt cgtaactatg aatatgtatg 11760
 gcacacacat acagatccaa aattaataaa tccaccaggt agtttgaaac agaattctac 11820
 tccgatctag aacgaccgcc caaccagacc acatcatcac aaccaagaca aaaaaaagca 11880
 tgaaaagatg acccgacaaa caagtgcacg gcatatattg aaataaagga aaagggcaaa 11940
 ccaaacccta tgcaacgaaa caaaaaaaat catgaaatcg atcccgtctg cggaacggct 12000
 agagccatcc caggattccc caaagagaaa cactggcaag ttagcaatca gaacgtgtct 12060
 gacgtacagg tcgcatccgt gtacgaacgc tagcagcacg gatctaacac aaacacggat 12120
 ctaacacaaa catgaacaga agtagaacta ccggggcccta accatggacc ggaacgccga 12180
 tctagagaag gtagagaggg ggggggagga cgagcggcgt accttgaagc ggaggtgccg 12240
 acgggtggat ttgggggaga tccactagtt ctagagcggc cgccaccgcg gtggaattct 12300
 cgaggtcctc tccaaatgaa atgaacttcc ttatatagag gaagggtctt gcgaaggata 12360
 gtgggattgt gcgtcatccc ttacgtcagt ggagatatca catcaatcca cttgctttga 12420
 agacgtggtt ggaacgtctt ctttttccac gatgctctc gtgggtgggg gtccatcttt 12480
 gggaccactg tcggcagagg catcttgaac gatagccttt cttttatcgc aatgatggca 12540
 tttgtaggtg ccaccttctt ttttactgt ctttttgatc aagtaccga tagctgggca 12600
 atggaatccg aggaggtttc ccgatattac cttttgttga aaagtctcaa tagccctttg 12660
 gtcttctgag actgtatctt tgatattctt ggagtagacg agagtgtcgt gctccaccat 12720
 gttatcacat caattcactt gctttgaaga cgtggttga acgtcttctt tttccacgat 12780
 gctcctcgtg ggtgggggtc catctttggg accactgtcg gcagaggcat cttgaacgat 12840
 agcctttcct ttatcgcaat gatggcattt gtaggtgcc a ctttctttt ctactgtcct 12900
 tttgatcaag tgacagatag ctgggcaatg gaatccgagg aggtttcccg atattaccct 12960
 ttgttgaaaa gtctcaatag ccctttggtc ttctgagact tgcaggcaag ca 13012

<210> 286

<211> 13013

<212> DNA

<213> 人工序列

<220>

<223> pGEP772 expression plasmid

<400> 286

cgatctttcg gttgtcgaag agtcaaaga aataattatc cagaaggctc atgaaatcta 60
 caagggttac ggctcaagcg agaagctctt tgatgtgac ttcgtgttgg agaagtctct 120
 taacaaaaaac gacgcagtcg tcgcgataat gaaagatttg ctggattcag tgaaatcctt 180
 cgagaattat atcaaagcct tcttcggcga ggggaaggag acaaacaggg atgagtcctt 240
 ctatggagac ttcgttctgg cttacgacat cttctttaag gtcgaccaca tctatgacgc 300
 aattcggaac tatgtgacgc agaagccgta ttcgaaagat aagttcaagc tctattttcca 360
 aaaccctcaa tttatgcgtg ggtgggataa agacaaagag accgattacc gggcaacaat 420
 tttgcggtac ggggtctaaat attacctcgc tataatggat aagaaatacg ctaaattgtct 480
 ccagaaaatt gacaaagatg acgtcaacgg caattatgaa aaaatcaatt ataaactcct 540

tccctggccca aataaaatgc tcccgagggt gtttttttcc aaaaagtga tggcctatta 600
taatccatca gaggatatc agaaaatcta taaaaatggg acctttaaga aggggtgacat 660
gtttaacctg aacgattgcc acaagcttat agattttttc aaagactcta ttagccgcta 720
tcccaaatgg tctaattgctt atgatttcaa cttctctgaa actgaaaagt acaaagatat 780
tgcaggattc taccgcgaag ttgaagaaca aggttataag gtttcctttg agtctgcgtc 840
caagaaagag gtcgataagt tggtcgaaga agggaaattg tatatgtttc aaatttacia 900
taaagacttt tccgacaagt cccatggtac acctaactctg cataccatgt acttcaaact 960
gctgttcgat gagaataatc acggtcagat tcgcctgagc ggaggggagg aactcttcat 1020
gaggagagca tcgttgaaaa aagaggagct cgtcgtgcat ccggctaaca gccccattgc 1080
taacaagaat ccgataatc caaagaagac tactaccctc tcctatgacg tctataagga 1140
taagagattc tctgaggacc agtacgagtt gcacatccct attgcgataa ataaatgccc 1200
taagaacatc tttaaaatca atactgaggt cagagtcttg ctttaagcag acgacaaccc 1260
gtatgtgatc gggattgcta ggggtgaaag gaacttgctt tatattgttg ttgtcgatgg 1320
aaaaggtaat atagtggaa aatactctct gaatgaaatt atcaacaact tcaatggcat 1380
taggatcaag accgactatc attctctgtt ggacaagaaa gagaaagagc gcttcgaggc 1440
acggcaaaac tggacgtcta ttgagaacat caaggagctt aaggctggtt acattttctca 1500
ggttgtgcac aaaaatttgcg aactggtcga gaaatatgat gccgttatcg cacttgaaga 1560
tctcaacagc ggatttaaga attctcgggt gaaagtcgaa aaacaggtgt atcaaaaatt 1620
cgaaaagatg ctgatcgaca agtcaatta tatggttgat aaaaagagca acccatgcgc 1680
cacggggggg gcgcttaagg gctatcagat tacgaacaaa tttgaatcct tcaagtcaat 1740
gtcgacgcaa aatgggttta tattctatat accggcgttg cttacatcta aaatagatcc 1800
tagcactggg ttcgtgaacc tgctgaaaac caagtacact tcaatcgcag attctaaaaa 1860
atttataagc agcttcgaca gaatcatgta tgtgcccag gaagacctct tcgagtttgc 1920
ccttgattac aaaaatttct caagaacgga tgcagactac ataaagaagt ggaagctgta 1980
ctcttatggg aaccggattc ggatattcag aaatccgaaa aaaaacaatg tctttgattg 2040
ggaggaagtt tgtcttacct ctgcttacia agagctgttc aataaatatg gcattaatta 2100
ccagcaaggt gatatccggg cgctcctttg cgaacagtct gacaaagctt tctattcttc 2160
atttatggcg ctcatgtcat tgatgtgca gatgaggaat agcattacgg ggaggactga 2220
tgttgacttt ctgatctcgc ccgtgaaaaa ttctgatgga atcttctacg attccaggaa 2280
ttatgaggcc caggaaaaat ctatccttcc caagaacgca gacgcaaatg gcggtacaa 2340
tatagctcgc aaggttttgt gggctatagg ccaattcaag aaagccgaag acgaaaagct 2400
ggacaaaagt aagattgcta tatctaacia agagtggctt gagtatgcgc aaacatctgt 2460
taaacacaaa cgccccgcgg ctacaaagaa ggctggccag gccagaaga agaagggtc 2520
ggggtcgggg tcgggctcgg gctcggacgc cctggacgac ttcgacctcg acatgctggg 2580
ctccgacgcc ctcatgatt tcgacctcga tatgctcggc agcgacgcgc tcgatgactt 2640
cgacctgat atgctgggga gcgacgcct cgacgatttt gacctcgata tgctgatcaa 2700
ctcccgtcc agcggcagcc cgaagaagaa gcgcaaagt ggctcgcagt acctgcccga 2760
caccgacgac aggcacagga tcgaggagaa gcgcaagagg acgtacgaga ctttcaagtc 2820
catcatgaag aagtccccgt tcagcggccc aacggacccc cgcccgcgc cgaggaggat 2880

cgccgtgccg tccaggtcca gcgcgtcggc ccccaagccg gccccgcagc cctacccgtt 2940
 cacgtccagc ctcagcacca tcaactacga cgagttcccc accatggtgt tcccggtccg 3000
 ccagatctcc caggccagcg cgctggcccc cgcgcccccg caggtgctgc cccaggtcc 3060
 ggcccccgct ccggccccgg ccatggtctc cgcgctggcc caggcgcccc ccccggtgcc 3120
 cgtcctcgcg ccggggccgc cgcaggcggc cgccccgcca gcgccgaagc ccacgcaggc 3180
 cggcgagggc accctcagcg aggcgctcct gcagctgcag ttcgacgacg aggacctcg 3240
 cgccctcctg ggcaactcga ccgacccgc cgtgttcacc gacctggcct ccgtcgacaa 3300
 cagcgagttc cagcagctgc tgaaccaggc catccccggt gcgccgcaca ccacggagcc 3360
 catgctgatg gagtaccgag aggcgatcac gcgcctcgtc accggcgccc agaggcccc 3420
 ggacccccgc ccggccccgc tcggcgcccc aggcctgccg aacggcctcc tgagcggcga 3480
 cgaggacttc tccagcatcg cggacatgga cttctccgc ctctggggc cgggctcggg 3540
 cagccgcgac agcagggagg gcatgttctt cccaaagccc gaggcggct ccgccatctc 3600
 ggacgtgttc gagggcaggg aggtctgcca gccaaagcgc atcaggccgt tccaccgcc 3660
 gggctccccg tgggcgaacc ggccgtccc cgccagcctg gctccaacc cgaccggccc 3720
 cgtgcacgag ccggtcggca gcctgacgcc cgcgccggc cccagccgc tcgacccgc 3780
 gccggccgtc acccccgagg cctccacact cctggaggac cccgacgagg agacctgca 3840
 ggccgtgaag gccctgagg agatggccga caccgtcatc cccagaagg aggaggcggc 3900
 catctgcggc cagatggacc tgtcgcacc gccgccgcgc ggccacctc acgagctgac 3960
 cagcacctc gagtccatga ccgaggacct caacctggac agccccctca cgccggagct 4020
 gaacgagatc ctcgacacct tcctgaacga cgagtgcctc ctgcacgcca tgcacatctc 4080
 caggggcctg agcatcttcg acaccagcct cttctgagtc gaccgatcgt tcaaacattt 4140
 ggcaataaag tttcttaaga ttgaatcctg ttgccggtct tgcgatgatt atcatataat 4200
 ttctgttgaa ttacgttaag catgtaataa ttaacatgta atgcatgacg ttatttatga 4260
 gatgggtttt tatgattaga gtcccgcaat tatacattt atacgcgata gaaaacaaaa 4320
 tatagcgcgc aaactaggat aaattatcgc gcgcggtgtc atctatgtta ctagatcgat 4380
 cccgggatat cgcgccgcgc tcgttaagct gcggcgagcg gtatcagctc actcaaaggc 4440
 ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt gagcaaaagg 4500
 ccagcaaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg 4560
 cccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg 4620
 actataaaga taccaggcgt tccccctgg aagctccctc gtgcgctctc ctgttccgac 4680
 cctgccgctt accggatacc tgtccgctt tctcccttcg ggaagcgtgg cgctttctca 4740
 tagctcacgc ttaggtatc tcagttcggc gtaggtcgtt cgctccaagc tgggctgtgt 4800
 gcacgaaccc ccggttcagc ccgaccgtg cgcttatcc ggtaactatc gtcttgagtc 4860
 caaccggta agacacgact tatcgccact ggcagcagc actggtaaca ggattagcag 4920
 agcgaggat gtaggcggtg ctacagagtt cttgaagtgg tggcctaact acggctacac 4980
 tagaaggaca gtatttggtg tctgcgtctc gctgaagcca gttaccttcg gaaaaagagt 5040
 tggtagctct tgatccggca aacaaaccac cgctggtagc ggtggttttt ttgtttgcaa 5100
 gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat cctttgatct tttctacggg 5160
 gtctgacgct cagtggaacg aaaactcacg ttaagggatt ttggtcatga gattatcaaa 5220

aaggatcttc acctagatcc ttttaaatta aaaatgaagt tttaaatcaa tctaaagtat 5280
atatgagtaa acttggctctg acagttacca atgcttaatc agtgaggcac ctatctcagc 5340
gatctgtcta tttcgttcat ccatagtgc ctgactcccc gtcgtgtaga taactacgat 5400
acgggagggc ttaccatctg gccccagtgc tgcaatgata ccgcgagacc cacgctcacc 5460
ggctccagat ttatcagcaa taaaccagcc agccggaagg gccgagcgca gaagtgggcc 5520
tgcaacttta tccgcctcca tccagtctat taattgttgc cgggaagcta gagtaagtag 5580
ttcgccagtt aatagtttgc gcaacgttgt tgccattgct acaggcatcg tgggtgcacg 5640
ctcgtcgttt ggtatggctt cattcagctc cggttcccaa cgatcaaggc gagttacatg 5700
atcccccatg ttgtgcaaaa aagcggtttag ctcttcgggt cctccgatcg ttgtcagaag 5760
taagttggcc gcagtgttat cactcatggt tatggcagca ctgcataatt ctcttactgt 5820
catgccatcc gtaagatgct tttctgtgac tggtagtac tcaaccaagt cattctgaga 5880
atagtgtatg cggcgaccga gttgctcttg cccggcgtea atacgggata ataccgcgcc 5940
acatagcaga actttaaaaag tgctcatcat tggaaaacgt tcttcggggc gaaaactctc 6000
aaggatctta ccgctgttga gatccagttc gatgtaacct actcgtgcac ccaactgac 6060
ttcagcatct tttactttca ccagcgttcc tgggtgagca aaaacaggaa ggcaaaatgc 6120
cgcaaaaaag ggaataaggc cgacacggaa atgttgaata ctcatactct tcctttttca 6180
atattattga agcatttatc agggttattg tctcatgagc ggatacatat ttgaatgtat 6240
ttagaaaaat aaacaaatag gggttccgcg cacatttccc cgaaaagtgc cacctgacgc 6300
gccctgtagc ggcacgtcta attcggggga tctggatttt agtactggat tttggtttta 6360
ggaattagaa attttattga tagaagtatt ttacaaatac aaatacatac taagggtttc 6420
ttatatgctc aacacatgag cgaaacccta taggaacct aattccctta tctgggaact 6480
actcacacat tattatggag aaactcgagc ttgtcgatcg acatgatcag ggagccctag 6540
attatttgta tagttcatcc atgccatta cgtcggtaaa tgccttctgc cactccttga 6600
agttaagttc ggtcttgga tgtttcaact cagtcttacg gaacacgtac atgggttgg 6660
tcttaaggta gttagcggcc attggttttag cgaatgtgta ggtagtcctg gctgtagagc 6720
gatattctct gccattgcct gtggtgtaag accatttgaa ggtactaatg atggcttctgt 6780
cgttagggtg ggttttcttg gaccggcacc aatcagcggc agttaaggag ttggtcatga 6840
caggtccatc agcaggaaaag cctgtcccct tcaattgggc ttctcctttg atgtggctcc 6900
cttcgtaagt gtaacggtag ttgacggtga gcgaagcacc gtcctcaaac tgcattgtcc 6960
tgtggacttg gtatccggag ccatcaacca tggtgcttg gaatggactc attccgtcag 7020
ggtatggaag gtattgatgg aatccgtagc caatgtgtgg caccagaatc catggagaaa 7080
actgaagatc acctttgggt ctcttgaggt tcagctcttc gtatccgtca ttagggttcc 7140
cagtgccctg tccgaccata tcgaagtcaa cgccgttgat ggaaccgaag atgtgaagct 7200
catgtgtggc tggaagcgaa gccatgttat cttcttctcc tttactcacg gaggacgcca 7260
tggtggcggg atcgcgccct atcgttcgta aatggtgaaa attttcagaa aattgctttt 7320
gctttaaaaag aaatgattta aattgctgca atagaagtag aatgcttgat tgcttgagat 7380
tcgtttgttt tgtatatgtt gtgttgagag gatcctcaag cttcgacctg cagaagtaac 7440
accaaacaac agggtagca tcgacaaaag aaacagtacc aagcaaataa atagcgtatg 7500
aaggcagggc taaaaaaatc cacatatagc tgctgcatat gccatcatcc aagtatatca 7560

```

agatcaaaat aattataaaa cataacttggt tattataata gataggtact caaggttaga 7620
gcatatgaat agatgctgca tatgccatca tgtatatgca tcagtaaaac ccacatcaac 7680
atgtatacct atcctagatc gatattttcca tccatcttaa actcgtaact atgaagatgt 7740
atgacacaca catacagttc caaaattaat aaatacacca ggtagtttga aacagtattc 7800
tactccgatac tagaacgaat gaacgaccgc ccaaccacac cacatcatca caaccaagcg 7860
aacaaaagca tctctgtata tgcatacagta aaacccgcat caacatgtat acctatccta 7920
gatcgatatt tccatccatc atcttcaatt cgtaactatg aatatgtatg gcacacacat 7980
acagatccaa aattaataaaa tccaccaggt agtttgaaac agaattctac tccgatctag 8040
aacgaccgcc caaccagacc acatcatcac aaccaagaca aaaaaaagca tgaaaagatg 8100
acccgacaaa caagtgcacg gcatatattg aaataaagga aaagggcaaa ccaaacccta 8160
tgcaacgaaa caaaaaaaat catgaaatcg atcccgtctg cggaacggct agagccatcc 8220
caggattccc caaagagaaa cactggcaag ttagcaatca gaacgtgtct gacgtacagg 8280
tcgcatccgt gtacgaacgc tagcagcacg gatctaacac aaacacggat ctaacacaaa 8340
catgaacaga agtagaacta ccggggcccta accatggacc ggaacgccga tctagagaag 8400
gtagagaggg ggggggagga cgagcggcgt acctgaagc ggaggtgccg acgggtggat 8460
ttgggggaga tccactagtt ctagagcggc cgccaccgcg gtggaattct cgaggtcctc 8520
tccaaatgaa atgaacttcc ttatatagag gaagggtctt gcgaaggata gtgggattgt 8580
gcgtcatccc ttacgtcagt ggagatatca catcaatcca cttgctttga agacgtggtt 8640
ggaacgtctt ctttttccac gatgtcctc gtgggtgggg gtccatcttt gggaccactg 8700
tcggcagagg catcttgaac gatagccttt cttttatcgc aatgatggca tttgtaggtg 8760
ccaccttctt tttctactgt ctttttgatc aagtgaccga tagctgggca atggaatccg 8820
aggaggtttc ccgatattac cttttgttga aaagtctcaa tagccctttg gtcttctgag 8880
actgtatctt tgatatctt ggagtagacg agagtgtcgt gctccaccat gttatcacat 8940
caattcactt gctttgaaga cgtggttga acgtcttctt tttccacgat gtcctcgtg 9000
gggtgggggtc catctttggg accactgtcg gcagaggcat cttgaacgat agcctttcct 9060
ttatcgcaat gatggcattt gtaggtgcca ctttccttt ctactgtcct tttgatcaag 9120
tgacagatag ctgggcaatg gaatccgagg aggtttccc atattaccct ttgttgaaaa 9180
gtctcaatag ccctttggtc ttctgagact tgcaggcaag caagcatgaa tgcctggggg 9240
agaagaactc gagagggaat tgcagatcat gaggcagatg gctatTTTTg tgtcacatat 9300
gcgcaaaaaag agaggctata tttgtgtccc taggttcttc gttgtattgc agtttccata 9360
tcaatctgac ttggtcgcat gagaaattga tggttaaata atttgaatct ctcatttagt 9420
atcaactatt agatattatt ttcaccaa atatttccat cggagaagaa gaggtacag 9480
aggaagcaga agagaggggt gggagaattt ttacactttt gtacaccac ttaaacagca 9540
aaatccgtat gaaaacaggc ccaccaa aatgccacga taacaatccg tagaaacaaa 9600
agcttcattt aacagcggcg caacaaagca cgcttatcca tggtagttgt agtccgtatg 9660
cgatccaaag atcacgattc acgcgtgacg gacggacgac gcgtgccaca ccacaactaa 9720
cggcattcat ggtagttgta gtccgtatgc gatccaaaga tcacgattca cgcgtgacgg 9780
acggacgacg cgcgccacac cacaactaac agcgtgagcc agcgtccaaa ctccggatgg 9840
caacgggggac gaaacccgtc gggtagtcac tgcccaaacc cgtccccgca accttcatcc 9900

```

caaaccgctc cccgtttccg gtcgcgggtt tcagttttct accagaccgg tccccatcgg 9960
gtttttcatc cccgtcggga aatccgaacc cgccagcatt tcagcaccaa gccaaagtgt 10020
cagcagcaac atgaataaaa aacaaccggt ttcaacacca agataaaaca aaacattata 10080
atttagacaa catttcacac gtataacaat aacatatagt tctcacatat aacaacacca 10140
tttcacacat aaaacaacac catttgggat aaaaatatgg gctatatcag gccattttta 10200
tgggccatat tgagttttcg tgggtttcac aggtaccgga tttgtagaat gctgaaccgg 10260
gtttgaaccg taaaatccgc gggtattgaa tttgaccaa tcccgctcgt ccctgggtggg 10320
gtaaaaacac catcttgagt ccaaacggcc accaacaaa ctccgacggc aacaaacaaa 10380
cggcgttgct ttgtcctcgt gtatctcgtt gaccgctcaa tctcccggtt gtttccccgg 10440
aattgcgtgg actctctcat ccacacgcaa accgctctc cctcctctct cgtcctatcc 10500
gccccggtgc cgtagcctca cgggactctt ctctctcctt tgctataaaa tccccgcccc 10560
ctccccgtct ctctccacac atccaaactc tcaatcgac cgagaaaaat ctcttagcga 10620
tcgaagcgaa gcctctcccc atcctctcaa ggtacggcgg tttcccgctg atcctcctcc 10680
ttccgttcgt gttctgtagc cgatcgattc gattccctta caccggttcg tgttctctcg 10740
tggtatcgatc gattgtttgt tgctagaagg aactcgtaga tctggcggtt atgaactgtg 10800
attcgggtta gtccagatcg attcaggtcg gtcgtcgttg agcctctcgg ctatgtctgg 10860
attatcgtgt agatctgctg gttcagttga ttatgttctt ctaggagtaa tttcgttggg 10920
tcagcgcgat ttctgcttaa tctatgctgc ttattgcgcc tgtacctatc tactaagcta 10980
tgtgcacctg taattttgct agattattcg ttcatcctcg tagttgggtt gtcacagtaa 11040
tccgtatggg ttctgacgat gttattgttg gtcataccta ggcttctcca gattttattt 11100
tgtttaaatt ggatagatct gctactgata gttgatgatg gaatttggtg ctgaatctat 11160
gctattttatt gcgcctatac ctgatctatc gggctatgta cggctgtagt ttactggatt 11220
attcgttcac cctcggtagt tggttcatcg tttgggttct gacgataata ttgttgatta 11280
tgcgtaggct tctgcagatt gttgttaaaa ttggatacat cggttactga tggttgatga 11340
tagattttgt ctgaacctat ctgtttattg ctctataacc tgatctatag ggctatgtat 11400
gcctgtaatt taccagatta ttcgttcac ctcgtagttg gttcatctct ataattcgta 11460
tggtttctta tgatgttatt gttgattatg ctagtctta tacagattat tgtgtcaaga 11520
ttgaatatac ctgctactga tcggtgataa tttgggttagt agtttgcaat ctgctaggaa 11580
cacgttacca ctgtaatctg taaacatggg ttgccagagt agtttggtct actactcttg 11640
atatggttgc tgatttttagt cgctcctttt tggatcatgt attgatgtcc ttgcagattt 11700
ccgtgtactt accccgggtt ttgtgtactt cgtgttaaca ggtcgggtac cgaagcaaac 11760
atggcatcta gcatggcacc aaagaaaaaa aggaaagttt ccaaacttga aaaatttaca 11820
aactgctact ccctttccaa gacgcttagg tttaaagcga tccccgttgg caagacccaa 11880
gagaatatcg ataacaaaag acttctggtc gaagatgaaa aaagggccga agactacaag 11940
gggggtcaaga agttgctcga tcgctattat ctttctttta tcaacgatgt gcttcattca 12000
atcaaactga agaacttgaa taactacatt agccttttca gaaagaaaac gaggactgaa 12060
aaggagaaca aggaacttga gaatcttgaa ataaaccttc gcaaagaaat tgcaaaagcc 12120
ttcaaggggga acgaaggata taaatctctt ttcaaaaaag acattataga aacaattttg 12180
cctgagtttc ttgacgacaa ggatgaaatt gcgctcgtca atagctttta cggattttaca 12240

actgccttca cagggttctt cgacaatagg gagaatatgt ttagcgagga ggcaaaaagc 12300
 acatccatcg cattcagatg catcaatgaa aatcttaccg ggtacatatc gaatatggac 12360
 atattttgaaa aagtggatgc aatattcgat aagcacgaag tccaggagat aaaggaaaag 12420
 atactgaata gcgactatga tgtcgaagat tttttcgaag gtgagttctt caactttgtc 12480
 ctgactcaag aaggcattga tgtctataat gcaataattg gaggttttgt gactgagtct 12540
 ggcgagaaga taaagggcctt gaacgagtat atcaatctct acaaccagaa gactaagcaa 12600
 aagttgccta aattttaaacc gctttacaag caagttttga gcgaccggga aagcctttcc 12660
 ttttacgggtg aaggatacac gagcgatgaa gaagtcctcg aagtcttccg caacacactc 12720
 aacaagaact cagaaaatctt ttcctcaatt aaaaaattgg agaagctttt caagaacttc 12780
 gatgaatact cttcggcgagg gattttttgtg aagaacggcc cggcaatttc cacaatatct 12840
 aaagacattt tcggagaatg gaacgtgata agagacaagt ggaatgcgga gtatgatgac 12900
 atacacctga agaagaaggc agttgtgact gaaaaatacg aagatgacag gagaaaaagc 12960
 tttaaaaaga tcgggtcctt ttcactggaa cagctgcagg agtatgccga cgc 13013

<210> 287

<211> 13012

<212> DNA

<213> 人工序列

<220>

<223> pGEP761 expression plasmid

<400> 287

cgatctttcg gttgtcgaag agtcaaaga aataattatc cagaaggctg atgaaatcta 60
 caagggtgac ggctcaagcg agaagctctt tgatgctgac ttcgtgttgg agaagtctct 120
 taaaaaaaaac gacgcagtcg tcgcgataat gaaagatttg ctggattcag tgaaatcctt 180
 cgagaattat atcaaagcct tcttcggcga ggggaaggag acaaacaggg atgagtcctt 240
 ctatggagac ttcgttctgg cttacgacat cttcttaag gtcgaccaca tctatgacgc 300
 aattcggaac tatgtgacgc agaagccgta ttcgaaagat aagttcaagc tctattttcca 360
 aaaccctcaa tttatgctg ggtgggataa agacgtagag accgatcgcc gggcaacaat 420
 tttgcggtac ggggtctaaat attacctgc tataatggat aagaaatacg ctaaattgtct 480
 ccagaaaatt gacaaaagat acgtcaacgg caattatgaa aaaatcaatt ataaactcct 540
 tcctggccca aataaaatgc tcccgaagggt gtttttttcc aaaaagtga tgccctatta 600
 taatccatca gaggatattc agaaaatcta taaaaatggg acctttaaga aggggtgacat 660
 gttaaacctg aacgattgcc acaagcttat agattttttc aaagactcta ttagccgcta 720
 tcccaaatgg tctaattgctt atgattttcaa cttctctgaa actgaaaagt acaaagatat 780
 tgcaggattc taccgcgaag ttgaagaaca aggttataag gtttcctttg agtctgcgtc 840
 caagaaagag gtcgataagt tggtcgaaga agggaaattg tatatgtttc aaattttaca 900
 taaagacttt tccgacaagt cccatggtac acctaatctg cataccatgt acttcaaact 960
 gctgttcgat gagaataatc acggtcagat tcgcctgagc ggaggggcgg aactcttcat 1020
 gaggagagca tcgttgaaaa aagaggagct cgtcgtgcat ccggctaaca gccccattgc 1080
 taacaagaat ccggataatc caaagaagac tactaccctc tcctatgacg tctataagga 1140

```

taagagattc tctgaggacc agtacgagtt gcacatccct attgcgataa ataaatgccc 1200
taagaacatc tttaaaatca atactgaggt cagagtcctg cttaaagcacg acgacaaccc 1260
gtatgtgatc gggattgcta ggggtgaaag gaacttgctt tatattgttg ttgtcgatgg 1320
aaaaggtaat atagtggaac aatactctct gaatgaaatt atcaacaact tcaatggcat 1380
taggatcaag accgactatc attctctgtt ggacaagaaa gagaaagagc gcttcgaggc 1440
acggcaaaaac tggacgtcta ttgagaacat caaggagctt aaggctgggt acatttctca 1500
ggttgtgcac aaaaatttgcg aactggtcga gaaatatgat gccgttatcg cacttgaaga 1560
tctcaacagc ggatttaaga attctcgggt gaaagtcgaa aaacaggtgt atcaaaaatt 1620
cgaaaagatg ctgacgaca agctcaatta tatggttgat aaaaagagca acccatgcgc 1680
cacgggggggt gcgcttaagg gctatcagat tacgaacaaa tttgaatcct tcaagtcaat 1740
gtcgacgcaa aatgggttta tattctatat accggcgttg cttacatcta aaatagatcc 1800
tagcactggg ttcgtgaacc tgetgaaaac caagtacact tcaatcgagc attctaaaaa 1860
atttataagc agcttcgaca gaatcatgta tgtgcccgag gaagacctct tcgagtttgc 1920
ccttgattac aaaaatttct caagaacgga tgcagactac ataaagaagt ggaagctgta 1980
ctcttatggg aaccggattc ggatattcag aaatccgaaa aaaaacaatg tctttgattg 2040
ggaggaagtt tgtcttacct ctgcttaca agagctgttc aataaatatg gcattaatta 2100
ccagcaaggt gatatccggg cgctcctttg cgaacagtct gacaaagctt tctattcttc 2160
atttatggcg ctcatgtcat tgatgctgca gatgaggaat agcattacgg ggaggactga 2220
tgttgacttt ctgatctcgc ccgtgaaaaa ttctgatgga atcttctacg attccaggaa 2280
ttatgaggcc caggaaaatg ctatccttcc caagaacgca gacgcaaatg gcgcgtacaa 2340
tatagctcgc aaggttttgt gggctatagg ccaattcaag aaagccgaag acgaaaagct 2400
ggacaaagtt aagattgcta tatctaaca agagtggctt gagtatgcgc aaacatctgt 2460
taaacacaaa cgccccgcg ctacaaagaa ggctggccag gccaaagaaga agaagggtc 2520
ggggctcgggg tcgggctcgg gctcggacgc cctggacgac ttcgacctcg acatgctggg 2580
ctccgacgcc ctgatgatt tcgacctga tatgctcggc agcgacgcgc tcgatgactt 2640
cgacctcgat atgctgggga gcgacgcct cgacgatttt gacctgata tgctgatcaa 2700
ctcccgtcc agcggcagcc cgaagaagaa gcgcaaagt ggctcgagt acctgcccga 2760
caccgacgac aggcacagga tcgaggagaa gcgcaagagg acgtacgaga cttcaagtc 2820
catcatgaag aagtccccgt tcagcggccc aacggacccc cgcccgccgc cgaggaggat 2880
cgccgtgccg tccaggtcca gcgcgtcggg cccaagccg gccccgcgc cctaccggtt 2940
cacgtccagc ctgacacca tcaactacga cgagtcccc accatggtgt tcccgctcgg 3000
ccagatctcc caggccagcg cgctggcccc cgcgcccccg caggtgctgc ccaggtctc 3060
ggcccccgct ccggccccgg ccatggtctc cgcgctggcc caggcgcccc ccccggtgcc 3120
cgtcctcgcg ccgggccccg cgagggcggt cgcccccca gcgccgaagc ccacgcaggc 3180
cggcgagggc accctcagcg aggcgtcct gcagctgcag ttcgacgacg aggacctcg 3240
cgccctcctg ggcaactcga ccgacccccg cgtgttcacc gacctggcct ccgtcgacaa 3300
cagcgagttc cagcagctgc tgaaccagg catccccgtg gcgccgaca ccacggagcc 3360
catgctgatg gagtaccgg aggcgatcac gcgcctcgtc accggcgccc agaggcccc 3420
ggacccccgc ccggccccgc tcggcgcccc aggcctgccg aacggcctcc tgagcggcga 3480

```

cgaggacttc tccagcatcg cggacatgga cttctccgcc ctcttggggt cgggctcggg 3540
cagccgcgac agcagggagg gcatgttcct cccaaagccc gaggcggct ccgccatctc 3600
ggacgtgttc gagggcaggg aggtctgcca gccaaagcgc atcaggccgt tccaccgcc 3660
gggctccccg tgggcgaacc ggccgctccc cgccagcctg gctccaaccc cgaccggccc 3720
cgtgcacgag ccggtcggca gcctgacgcc cgcgccgggtg ccccagccgc tcgacccgc 3780
gccggccgtc acccccaggg cctcccacct cctggaggac cccgacgagg agacctcgca 3840
ggccgtgaag gccctgaggg agatggccga caccgtcatc cccagaagg aggaggcggc 3900
catctgcggc cagatggacc tgtcgcaccc gccgccgcgc ggccacctcg acgagctgac 3960
cacgaccctc gagtccatga ccgaggacct caacctggac agccccctca cgccggagct 4020
gaacgagatc ctgcacacct tctgaacga cgagtgcctc ctgcacgcca tgcacatctc 4080
cacgggcctg agcatcttcg acaccagcct cttctgagtc gaccgatcgt tcaaacattt 4140
ggcaataaag tttcttaaga ttgaatcctg ttgccggtct tgcgatgatt atcatataat 4200
ttctgttgaa ttacgttaag catgtaataa ttaacatgta atgcatgacg ttattttatga 4260
gatgggtttt tatgattaga gtcccgcaat tatacattta atacgcgata gaaaacaaaa 4320
tatagcgcgc aaactaggat aaattatcgc gcgcggtgtc atctatgtta ctagatcgat 4380
cccgggatat cgcggccggt cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg 4440
gtaatacggc tatccacaga atcaggggat aacgcaggaa agaacatgtg agcaaaaggc 4500
cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca taggctccgc 4560
ccccctgacg agcatcacia aaatcgacgc tcaagtcaga ggtggcgaaa ccgacagga 4620
ctataaagat accaggcgtt tccccctgga agctccctcg tgcgctctcc tgttccgacc 4680
ctgccgctta ccgataacct gtccgccttt ctcccttcgg gaagcgtggc gctttctcat 4740
agctcacgct gtaggtatct cagttcgggt taggtcgttc gctccaagct gggctgtgtg 4800
cacgaacccc ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgagtcc 4860
aaccgcgtaa gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga 4920
gcgaggtatg taggcgggtc tacagagttc ttgaagtggg ggcctaacta cggctacact 4980
agaaggacag tatttggtat ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt 5040
ggtagctctt gatccggcaa acaaaccacc gctggtagcg gtgggttttt tgtttgcaag 5100
cagcagatta cgcgcagaaa aaaaggatct caagaagatc ctttgatctt ttctacgggg 5160
tctgacgctc agtggaaacga aaactcacgt taagggattt tggatcatgag attatcaaaa 5220
aggatcttca cctagatcct tttaaattaa aatgaagtt ttaaataaat ctaaagtata 5280
tatgagtaaa cttggtctga cagttaccaa tgcttaata gtaggacacc tatctcagcg 5340
atctgtctat ttcgttcac ctagttgcc tgactccccg tcgtgtagat aactacgata 5400
cgggagggct taccatctgg cccagtgct gcaatgatac cgcgagacc acgctcaccg 5460
gctccagatt tatcagcaat aaaccagcca gccggaaggg ccgagcgcag aagtggctct 5520
gcaactttat ccgcctccat ccagtctatt aattgttgcc gggaagctag agtaagtagt 5580
tcgccagtta atagtttgcg caacgttggt gccattgcta caggcatcgt ggtgtcacgc 5640
tcgtcgtttg gtatggcttc attcagctcc ggttcccaac gatcaaggcg agttacatga 5700
tccccatgt tgtgcaaaaa agcggtttagc tccttcggtc ctccgatcgt tgtcagaagt 5760
aagttggccg cagtgttatc actcatggtt atggcagcac tgcataattc tcttactgtc 5820

atgccatccg taagatgctt ttctgtgact ggtgagtact caaccaagtc attctgagaa 5880
tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtcaa tacgggataa taccgcgcca 5940
catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt cttcggggcg aaaactctca 6000
aggatccttac cgctgttgag atccagttcg atgtaaccca ctcgtgcacc caactgatct 6060
tcagcatctt ttactttcac cagcgtttct gggtagcaaa aaacaggaag gcaaaatgcc 6120
gcaaaaaagg gaataagggc gacacggaaa tgttgaatac tcatactctt cttttttcaa 6180
tattattgaa gcatttatca gggttattgt ctcagagcg gatacatatt tgaatgtatt 6240
tagaaaaata aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc acctgacgcg 6300
ccctgtagcg gcacgtctaa ttcgggggat ctggatttta gtactggatt ttggtttttag 6360
gaattagaaa ttttattgat agaagtattt tacaataca aatacatact aagggtttct 6420
tatatgctca acacatgagc gaaaccctat aggaacccta attcccttat ctgggaacta 6480
ctcacacatt attatggaga aactcgagct tgtcgatcga catgatcagg gagccctaga 6540
ttatttgtat agttcatcca tgcccattac gtcggtaaat gccttctgcc actccttgaa 6600
gttaagttcg gtcttggaaat gtttcaactc agtcttacgg aacacgtaca tgggttggtt 6660
cttaaggtag ttagcggcca ttggtttagc gaatgtgtag gtagtcctgg ctgtagagcg 6720
atatctcttg ccattgcctg tgggtgaaga ccatttgaag gtactaatga tggctctgtc 6780
gttagggtag gttttcttgg accggcacca atcagcgga gttaaggagt tggctatgac 6840
aggtccatca gcaggaaagc ctgtcccctt cacttgggct tctcctttga tgtggctccc 6900
ttcgtaagtg taacggtagt tgacggtgag cgaagcaccg tcctcaaact gcattgtcct 6960
gtggacttgg tatccggagc catcaacat ggctgcttgg aatggactca ttccgtcagg 7020
gtatggaagg tattgatgga atccgtagcc aatgtgtggc accagaatcc atggagaaaa 7080
ctgaagatca cctttggtgc tcttgagggt cagctcttcg tatccgtcat tagggttccc 7140
agtgccttgt ccgaccatat cgaagtcaac gccgttgatg gaaccgaaga tgtgaagctc 7200
atgtgtggct ggaagcgaag ccatgttatc ttcttctcct ttactcacgg aggacgcat 7260
ggtagcgagg tcgcgcccta tcgttcgtaa atgggtgaaaa ttttcagaaa attgcttttg 7320
ctttaaaaga aatgatttaa attgctgcaa tagaagtaga atgcttgatt gcttgagatt 7380
cgtttgtttt gtatatgttg tgttgagagg atcctcaagc ttcgacctgc agaagtaaca 7440
ccaaacaaca gggtagcat cgacaaaaga aacagtacca agcaaataaa tagcgtatga 7500
aggcagggct aaaaaaatcc acatatagct gctgcatatg ccatcatcca agtatatcaa 7560
gatcaaaata attataaaac atacttgttt attataatag ataggtactc aaggttagag 7620
catatgaata gatgctgcat atgccatcat gtatatgcat cagtaaaacc cacatcaaca 7680
tgtataccta tcctagatcg atatttccat ccatcttaaa ctcgtaacta tgaagatgta 7740
tgacacacac atacagttcc aaaattaata aatacaccag gtagtttgaa acagtattct 7800
actccgatct agaacgaatg aacgaccgcc caaccacacc acatcatcac aaccaagcga 7860
acaaaagcat ctctgtatat gcatcagtaa aacccgcac aacatgtata cctatcctag 7920
atcgatatatt ccatccatca tcttcaattc gtaactatga atatgtatgg cacacacata 7980
cagatccaaa attaataaat ccaccaggta gtttgaaaca gaattctact ccgatctaga 8040
acgaccgccc aaccagacca catcatcaca accaagacaa aaaaaagcat gaaaagatga 8100
cccgacaaac aagtgcacgg catatatattga aataaaggaa aagggcaaac caaacctat 8160

gcaacgaaac aaaaaaaatc atgaaatcga tcccgtctgc ggaacggcta gagccatccc 8220
aggattcccc aaagagaaac actggcaagt tagcaatcag aacgtgtctg acgtacaggt 8280
cgcatccgtg tacgaacgct agcagcacgg atctaacaca aacacggatc taacacaaac 8340
atgaacagaa gtagaactac cgggccctaa ccatggaccg gaacgccgat ctagagaagg 8400
tagagagggg gggggaggac gagcggcgta ccttgaagcg gaggtgccga cgggtggatt 8460
tgggggagat ccactagttc tagagcggcc gccaccgcgg tggaattctc gaggtcctct 8520
ccaaatgaaa tgaacttcct tatatagagg aagggtcttg cgaaggatag tgggattgtg 8580
cgtcatccct tacgtcagtg gagatatcac atcaatccac ttgctttgaa gacgtggttg 8640
gaacgtcttc tttttccacg atgctcctcg tgggtggggg tccatctttg ggaccactgt 8700
cggcagaggc atcttgaacg atagcctttc ctttatcgca atgatggcat ttgtaggtgc 8760
caccttcctt ttctactgtc cttttgatca agtgaccgat agctgggcaa tggaatccga 8820
ggaggtttcc cgatattacc ctttgttgaa aagtctcaat agccctttgg tcttctgaga 8880
ctgtatcttt gatattcttg gagtagacga gagtgtcgtg ctccaccatg ttatcacatc 8940
aattcacttg ctttgaagac gtggttgga cgtcttcttt ttccacgatg ctctcctgtg 9000
gtgggggtcc atctttggga ccactgtcgg cagaggcatc ttgaacgata gcctttcctt 9060
tatcgcaatg atggcatttg taggtgccac cttccttttc tactgtcctt ttgatcaagt 9120
gacagatagc tgggcaatgg aatccgagga ggtttccga tattaccctt tgttgaaaag 9180
tctcaatagc cctttggtct tctgagactt gcaggcaagc aagcatgaat gcctggggga 9240
gaagaactcg agagggaatt gcagatcatg aggcatatg ctatTTTTgt gtcacatatg 9300
cgcaaaaaga gaggtatat ttgtgtccct aggttcttcg ttgtattgca gtttccatat 9360
caatctgact tggctcgtg agaaattgat ggttaaataa tttgaatctc tcatgtagta 9420
tcaactatta gatattattt tcaccaaata tatttccatc ggagaagaag aggctacaga 9480
ggaagcagaa gagaggggtg ggagaatttt tacacttttg tacaccact taaacagcaa 9540
aatccgtatg aaaacaggcc caccaaaaca atgccacgat aacaatccgt agaaacaaaa 9600
gcttcattta acagcggcgc aacaaagcac gcttatccat ggtagttgta gtccgtatgc 9660
gatccaaaga tcacgattca cgcgtgacgg acggacgacg cgtgccacac cacaactaac 9720
ggcatccatg gtagttgtag tccgtatgcg atccaaagat cacgattcac gcgtgacgga 9780
cggacgacgc gcgccacacc acaactaaca gcgtgagcca gcgtccaaac tccgatggc 9840
aacggggacg aaaccctcg ggtagtcact gcccaaacc gtccccgcaa ctttcatccc 9900
aaaccctcc ccgtttccgg tcgcgggttt cagttttcta ccagaccgt ccccatcggg 9960
tttttcatcc ccgtcgggaa atccgaacc gccagcattt cagcaccaag ccaaagttgc 10020
agcagcaaca tgaataaaaa acaaccgtt tcaacaccaa gataaaaca aacattataa 10080
tttagacaac atttcacacg tataacaata acatatagtt ctacatatata acaacaccat 10140
ttcacacata aaacaacacc atttgggata aaaatatggg ctatatcagg cattttttat 10200
gggcatatt gagttttcgt gggtttcaca ggtaccgat ttgtagaatg ctgaaccggg 10260
tttgaaccgt aaaaatccgc ggtattgaat ttgacccaat cccgtcgtcc cctggtgggg 10320
taaaaacacc atcttgagtc caaacggcca ccaaccaaac tccgacggca acaacaaac 10380
ggcgttgctt tgctcctcg tatctccgtg accgtcaat ctcccggctg tttccccgga 10440
attgcgtgga ctctctcatc cacacgcaaa ccgcctctcc ctctctctc gtcctatccg 10500

ccccggtgcc gtagcctcac gggactcttc ttctccctt gctataaaat ccccgcccc 10560
tcccgctctcc tctccacaca tccaaactct caatcgcacc gagaaaaatc tcctagcgat 10620
cgaagcgaag cctctcccca tcctctcaag gtacgcccgt ttcccgtcga tcctcctcct 10680
tccggttcgtg ttctgtagcc gatcgattcg attcccttac acccgttcgt gttctctcgt 10740
ggatcgatcg attgtttgtt gctagaagga actcgtagat ctggcgttta tgaactgtga 10800
ttcgggtagg tccagatcga ttcaggtcgg tcgtcggtga gcctctcggc tatgtctgga 10860
ttatcgtgta gatctgctgg ttcagttgat tatgttcttc taggagtaat ttcggtgggt 10920
cagcgcgatt tctgcttaat ctatgctgct tattgcgcct gtacctatct actaagctat 10980
gtgcacctgt aattttgcta gattattcgt tcctctcgt agttggtttg tcacagtaat 11040
ccgtatgggt tctgacgatg ttattgttgg tcatacctag gcttctccag attttatttt 11100
gttaaaattg gatagatctg ctactgatag ttgatgatgg aatttggtgc tgaatctatg 11160
ctatttattg cgcctatacc tgatctatcg ggctatgtac ggctgtagtt tactggatta 11220
ttcgttcac ctcggtagtt gggttcacgt ttgggttctg acgataatat gtgtgattat 11280
gcgtaggctt ctgcagattg ttgttaaaat tggatacatc ggttactgat ggttgatgat 11340
agatttggtc tgaacctatc tgtttattgc tcctatacct gatctatagg gctatgtatg 11400
cctgtaattt accagattat tcgttcaccc tcgtagttgg ttcactctta taattcgtat 11460
gggttcttat gatgttatcg ttgattatgc ctagtcttat acagattatt gtgtcaagat 11520
tgaatatacc tgctactgat cggtgataat ttggttagta gtttgcaatc tgctaggaac 11580
acgttaccac tgtaatctgt aaacatgggt tgccagagta gtttgttcta ctactcttga 11640
tatggttgct gattttagtc gcctcctttt ggatcatgta ttgatgtcct tgcagatttc 11700
cgtgtactta ccccggtttt tgtgtacttc gtgttaacag gtcgggtacc gaagcaaaca 11760
tggcatctag catggcacca aagaaaaaaa ggaaagtctc caaacttgaa aaatttaca 11820
actgtactc cctttccaag acgcttaggt ttaaagcgat ccccggtggc aagacccaag 11880
agaatatcga taacaaaaga cttctggtcg aagatgaaaa aagggccgaa gactacaagg 11940
gggtcaagaa gttgctcgat cgctattatc tttcctttat caacgatgtg cttcattcaa 12000
tcaaactgaa gaacttgaat aactacatta gccttttcag aaagaaaacg aggactgaaa 12060
aggagaacaa ggaacttgag aatcttgaaa taaaccttcg caaagaaatt gcaaaagcct 12120
tcaaggggaa cgaaggatat aaatctcttt tcaaaaaaga cattatagaa acaattttgc 12180
ctgagtttct tgacgacaag gatgaaattg cgctcgtaaa tagctttaac ggatttaca 12240
ctgccttcac agggttcttc gacaataggg agaatatgtt tagcgaggag gcaaaaagca 12300
catccatcgc attcagatgc atcaatgaaa atcttaccg gtacatatcg aatatggaca 12360
tatttgaaaa agtggatgca atattcgata agcacgaagt ccaggagata aaggaaaaga 12420
tactgaatag cgactatgat gtcgaagatt ttttcgaagg tgagttcttc aactttgtcc 12480
tgactcaaga aggcattgat gtctataatg caataattgg aggttttgtg actgagtctg 12540
gcgagaagat aaagggttg aacgagtata tcaatctcta caaccagaag actaagcaaa 12600
agttgcctaa atttaaaccg ctttacaagc aagttttgag cgaccgggaa agcctttcct 12660
tttacggtga aggatacacg agcgatgaag aagtcctcga agtcttccgc aacacactca 12720
acaagaactc agaaatcttt tcctcaatta aaaaattgga gaagcttttc aagaacttcg 12780
atgaatactc ttcggcgggg atttttgtga agaacggccc ggcaatttcc acaatatcta 12840

aagacatttt cggagaatgg aacgtgataa gagacaagtg gaatgcggag tatgatgaca 12900
 tacacctgaa gaagaaggca gttgtgactg aaaaatacga agatgacagg agaaaaagct 12960
 ttaaaaagat cgggtccttt tctactggaac agctgcagga gtatgccgac gc 13012

<210> 288

<211> 5370

<212> DNA

<213> 人工序列

<220>

<223> dLbCpf1-VPR

<400> 288

atggcatcta gcatggcacc aaagaaaaaa aggaaagttt ccaaacttga aaaattttaca 60
 aactgctact ccctttccaa gacgcttagg tttaaagcga tccccgttgg caagacccaa 120
 gagaatatcg ataacaaaag acttctggtc gaagatgaaa aaagggccga agactacaag 180
 ggggtcaaga agttgctcga tcgctattat ctttccttta tcaacgatgt gcttcattca 240
 atcaaactga agaacttgaa taactacatt agccttttca gaaagaaaac gaggactgaa 300
 aaggagaaca aggaacttga gaatcttgaa ataaaccttc gcaaagaaat tgcaaaagcc 360
 ttcaagggga acgaaggata taaatctctt ttcaaaaaag acattataga aacaattttg 420
 cctgagtttc ttgacgacaa ggatgaaatt gcgctcgtca atagctttta cggattttaca 480
 actgccttca cagggttctt cgacaatagg gagaatatgt ttagcgagga ggcaaaaagc 540
 acatccatcg cattcagatg catcaatgaa aatcttaccg ggtacatatc gaatatggac 600
 atatttgaaa aagtggatgc aatattcgat aagcacgaag tccaggagat aaaggaaaag 660
 atactgaata gcgactatga tgtcgaagat ttttcgaag gtgagttctt caactttgtc 720
 ctgactcaag aaggcattga tgtctataat gcaataattg gaggttttgt gactgagtct 780
 ggcgagaaga taaagggtt gaacgagtat atcaatctct acaaccagaa gactaagcaa 840
 aagttgccta aatttaaacc gctttacaag caagttttga gcgaccggga aagcctttcc 900
 ttttacggtg aaggatacac gagcgatgaa gaagtcctcg aagtcttccg caacacactc 960
 aacaagaact cagaaatctt ttcctcaatt aaaaaattgg agaagctttt caagaacttc 1020
 gatgaatact cttcggcggg gattttttgt aagaacggcc cggcaatttc cacaatatct 1080
 aaagacattt tcggagaatg gaacgtgata agagacaagt ggaatgcgga gtatgatgac 1140
 atacacctga agaagaaggc agttgtgact gaaaaatacga aagatgacag gagaaaaagc 1200
 tttaaaaaga tcgggtcctt ttcactggaa cagctgcagg agtatgccga cgccgatctt 1260
 tcggttgctg aaaagctcaa agaaataatt atccagaagg tcgatgaaat ctacaagggtg 1320
 tacggctcaa gcgagaagct ctttgatgct gacttcgtgt tggagaagtc tcttaaaaaa 1380
 aacgacgcag tcgtcgcat aatgaaagat ttgctggatt cagtgaaatc cttcgagaat 1440
 tatatcaaag cttcttcgga cgaggggaag gagacaaaca gggatgagtc cttctatgga 1500
 gacttcgttc tggttacga catccttctt aaggtcgacc acatctatga cgcaattcgg 1560
 aactatgtga cgcagaagcc gtattcgaaa gataagttca agctctatct ccaaaacctt 1620
 caatttatgg gtgggtggga taaagacaaa gagaccgatt accgggcaac aattttgcgg 1680
 tacgggtcta aatattacct cgctataatg gataagaaat acgctaaatg tctccagaaa 1740

attgacaaag atgacgtcaa cggcaattat gaaaaaatca attataaact ccttcctggc 1800
 ccaataaaaa tgctcccga ggtgtttttt tccaaaaagt ggatggccta ttataatcca 1860
 tcagaggata ttcagaaaat ctataaaaat gggaccttta agaagggtga catgtttaac 1920
 ctgaacgatt gccacaagct tatagatttt ttcaaagact ctattagccg ctatcccaaa 1980
 tggctctaatg cttatgattt caacttctct gaaactgaaa agtacaaga tattgcagga 2040
 ttctaccgcg aagttgaaga acaagggttat aaggtttctt ttgagtctgc gtccaagaaa 2100
 gaggtcgata agttgggtcga agaagggaat ttgtatatgt ttcaaattta caataaagac 2160
 ttttccgaca agtcccatgg tacacctaata ctgcatacca tgtacttcaa actgctgttc 2220
 gatgagaata atcacgggtca gattcgcttg agcggagggg cggaactctt catgaggaga 2280
 gcatcggttg aaaaagagga gctcgtcgtg catccggcta acagcccat tgctaacaag 2340
 aatccggata atccaaagaa gactactacc ctctctatg acgtctataa ggataagaga 2400
 ttctctgagg accagtacga gttgcacatc cctattgcga taaataaatg ccctaagaac 2460
 atctttaaaa tcaatactga ggtcagagtc ctgcttaagc acgacgaca cccgtatgtg 2520
 atcgggattg ctaggggtga aaggaaactg ctttatattg tggttgtcga tggaaaaggt 2580
 aatatagtgg aacaatactc tctgaatgaa attatcaaca acttcaatgg cattaggatc 2640
 aagaccgact atcattctct gttggacaag aaagagaaag agcgcttcga ggcacggcaa 2700
 aactggacgt ctattgagaa catcaaggag ctttaaggctg gttacatttc tcaggttgtg 2760
 cacaaaattt gcgaactggt cgagaaatat gatgccgtta tcgcacttga agatctcaac 2820
 agcggattta agaattctcg ggtgaaagtc gaaaaacagg tgtatcaaaa attcgaaaag 2880
 atgctgatcg acaagctcaa ttatatggtt gataaaaaga gcaacccatg cgccacgggg 2940
 ggtgcgctta agggctatca gattacgaac aaatttgaat cttcaagtc aatgtcgacg 3000
 caaatgggtt ttatatctta tataccggcg tggcttacat ctaaaataga tcctagcact 3060
 gggttcgtga acctgctgaa aaccaagtac acttcaatcg cagattctaa aaaatttata 3120
 agcagcttcg acagaatcat gtatgtgccc gaggaagacc tcttcgagtt tgcccttgat 3180
 tacaaaaatt tctcaagaac ggatgcagac tacataaaga agtggaagct gtactcttat 3240
 gggaaccgga ttcggatatt cagaaatccg aaaaaaaca atgtctttga ttgggaggaa 3300
 gtttgtctta cctctgctta caaagagctg ttcaataaat atggcattaa ttaccagcaa 3360
 ggtgatatcc gggcgtcctt ttgcgaacag tctgacaaag ctttctattc ttcatthtatg 3420
 gcgctcatgt cattgatgct gcagatgagg aatagcatta cggggaggac tgatgttgac 3480
 tttctgatct cgcccgtgaa aaattctgat ggaatcttct acgattccag gaattatgag 3540
 gcccaggaaa atgctatcct tccaagaac gcagacgcaa atggcgcgta caatatagct 3600
 cgcaaggttt tgtgggctat aggccaattc aagaaagccg aagacgaaaa gctggacaaa 3660
 gttaagattg ctatatctaa caaagagtgg cttgagtatg cgcaaacatc tgttaaacac 3720
 aaacgccccg cggtacaaa gaaggtggc caggccaaga agaagaaggg ctcggggctcg 3780
 gggtcgggct cgggctcgga cgccctggac gacttcgacc tcgacatgct gggctccgac 3840
 gccctgatg atttcgacct cgatatgctc ggcagcgacg cgctcgatga cttecgacctc 3900
 gatatgctgg ggagcgacgc cctcgacgat ttgacctcg atatgctgat caactcccgc 3960
 tccagcgga gcccgaaagaa gaagcgcaaa gtgggctcgc agtacctgcc cgacaccgac 4020
 gacaggcaca ggatcgagga gaagcgcaag aggacgtacg agaccttcaa gtccatcatg 4080

aagaagtccc cggttcagcgg cccaacggac ccccgccccg cgccgaggag gatcgccgtg 4140
ccgtccaggt ccagcgcgtc ggtccccaag ccggccccgc agccctaccc gtacacgtcc 4200
agcctcagca ccatcaacta cgacgagttc cccacatggt tgttcccgtc cggccagatc 4260
tcccaggcca gcgcgctggc ccccgcgccc ccgcaggtgc tgccccaggc tccggcccc 4320
gctccggccc cggccatggt ctccgcgctg gccagggcgc ccgccccggt gcccgctctc 4380
gcgcccggcc cgccgcaggc ggtcgccccg ccagcgccga agcccacgca ggccggcgag 4440
ggcacccctca gcgaggcgct cctgcagctg cagtctgacg acgaggacct cggcgccctc 4500
ctgggcaact cgaccgaccc cgccgtgttc accgacctgg cctccgtcga caacagcgag 4560
ttccagcagc tgctgaacca gggcatcccc gtggcgccgc acaccacgga gcccatgctg 4620
atggagtacc cggaggcgat cacgcgcctc gtcaccggcg cccagaggcc cccggacccc 4680
gccccggccc cgctcggcgc cccaggcctg ccgaacggcc tcttgagcgg cgacaggagc 4740
ttctccagca tcgcggacat ggactttctc gccctcctgg ggtcgggctc gggcagccgc 4800
gacagcaggg agggcatggt cctcccaaag ccgaggccg gctccgcat ctcgacgtg 4860
ttcgagggca gggagggtct ccagccaaag cgcacaggc cgttccacc gccgggtctc 4920
ccgtgggcca accggccgct ccccgccagc ctggctccaa ccccgaccgg ccccgctcac 4980
gagccggctc gcagcctgac gcccgcgccg gtgccccagc cgctcgacc cgcgccggcc 5040
gtcacccccg aggcctccca cctcctggag gaccccgacg aggagacctc gcaggccgtg 5100
aaggccctga gggagatggc cgacaccgtc atccccaga aggaggaggc ggccatctgc 5160
ggccagatgg acctgtcgca cccgcccgcg cgcggccacc tcgacgagct gaccacgacc 5220
ctcgagtcca tgaccgagga cctcaacctg gacagcccc tcacgccgga gctgaacgag 5280
atcctcgaca ccttctgaa cgacgagtgc ctctgcacg ccatgcacat ctccacgggc 5340
ctgagcatct tcgacaccag cctcttctga 5370

<210> 289

<211> 5370

<212> DNA

<213> 人工序列

<220>

<223> LbCpf1 (RR) -VPR

<400> 289

atggcatcta gcatggcacc aaagaaaaaa aggaaagttt ccaaacttga aaaattttaca 60
aactgctact ccctttccaa gacgcttagg tttaaagcga tccccgttgg caagacccaa 120
gagaatatcg ataacaaaag acttctggtc gaagatgaaa aaagggccga agactacaag 180
gggggtcaaga agttgctcga tcgctattat ctttcttta tcaacgatgt gcttcattca 240
atcaaactga agaacttgaa taactacatt agccttttca gaaagaaaac gaggactgaa 300
aaggagaaca aggaacttga gaatcttgaa ataaaccttc gcaaagaaat tgcaaaagcc 360
ttcaagggga acgaaggata taaatctctt ttcaaaaaag acattataga aacaattttg 420
cctgagtttc ttgacgacaa ggatgaaatt gcgctcgta atagctttta cggattttaca 480
actgccttca cagggttctt cgacaatagg gagaatatgt ttagcgagga ggcaaaaagc 540
acatccatcg cattcagatg catcaatgaa aatcttacc ggtacatatc gaatatggac 600

atatttgaaa aagtggatgc aatattcgat aagcacgaag tccaggagat aaaggaaaag 660
atactgaata gcgactatga tgtcgaagat ttttcgaag gtgagttctt caactttgtc 720
ctgactcaag aaggcattga tgtctataat gcaataattg gaggttttgt gactgagtct 780
ggcgagaaga taaagggcctt gaacgagtat atcaatctct acaaccagaa gactaagcaa 840
aagttgccta aatttaaacc gctttacaag caagttttga gcgaccggga aagcctttcc 900
ttttacgggtg aaggatacac gagcgatgaa gaagtcctcg aagtcctccg caacacactc 960
aacaagaact cagaaaatctt ttcctcaatt aaaaaattgg agaagctttt caagaacttc 1020
gatgaatact cttcggcggg gatttttgtg aagaacggcc cggcaatttc cacaatatct 1080
aaagacattt tcggagaatg gaacgtgata agagacaagt ggaatgcgga gtatgatgac 1140
atacacctga agaagaaggc agttgtgact gaaaaatacg aagatgacag gagaaaaagc 1200
tttaaaaaaga tcgggtcctt ttcaactggaa cagctgcagg agtatgccga cgccgatctt 1260
tcggttgtcg aaaagctcaa agaaataatt atccagaagg tcgatgaaat ctacaagggtg 1320
tacggctcaa gcgagaagct ctttgatgct gacttcgtgt tggagaagtc tcttaaaaaa 1380
aacgacgcag tcgtcgcgat aatgaaagat ttgctggatt cagtgaaatc cttcgagaat 1440
tatatcaaaag ctttcttcgg cgaggggaag gagacaaaca gggatgagtc cttctatgga 1500
gacttcgttc tggcttacga catccttctt aaggtcgacc acatctatga cgcaattcgg 1560
aactatgtga cgcagaagcc gtattcgaaa gataagttca agctctattt ccaaaacctt 1620
caatttatgc gtgggtggga taaagacaaa gagaccgatt accgggcaac aattttgcgg 1680
tacgggtcta aatattacct cgctataatg gataagaaat acgctaaatg tctccagaaa 1740
attgacaaaag atgacgtcaa cggcaattat gaaaaaatca attataaact ctttctggc 1800
ccaaataaaa tgctcccag ggtgtttttt tccaaaaagt ggatggccta ttataatcca 1860
tcagaggata ttcagaaaat ctataaaaat gggaccttta agaagggtga catgtttaac 1920
ctgaacgatt gccacaagct tatagatttt ttcaaagact ctattagccg ctatcccaaa 1980
tggtctaattg cttatgattt caacttctct gaaactgaaa agtacaaga tattgcagga 2040
ttctaccgcg aagtgaaga acaaggttat aaggtttcct ttgagtctgc gtccaagaaa 2100
gaggtcgata agttggtcga agaagggaat ttgtatatgt ttcaaattta caataaagac 2160
ttttccgaca agtcccatgg tacacctaatt ctgcatacca tgtacttcaa actgctgttc 2220
gatgagaata atcacgggtca gattcgctg agcggagggg cggaactctt catgaggaga 2280
gcatcggtga aaaaagagga gtcgtcgtg catccggcta acagcccat tgctaacaag 2340
aatccgata atccaaagaa gactactacc ctctctatg acgtctataa ggataagaga 2400
ttctctgagg accagtacga gttgcacatc cctattgcga taaataaatg ccctaagaac 2460
atctttaaaa tcaatactga ggtcagagtc ctgcttaagc acgacgaca cccgtatgtg 2520
atcgggattg ctaggggtga aaggaacttg ctttatattg tggttgtcga tggaaaaggt 2580
aatatagtg aacaatactc tctgaatgaa attatcaaca acttcaatgg cattaggatc 2640
aagaccgact atcattctct gttggacaag aaagagaaag agcgcttcga ggcacggcaa 2700
aactggacgt ctattgagaa catcaaggag cttaaggctg gttacatttc tcaggttgtg 2760
cacaaaattt gcgaactggt cgagaaatat gatgccgtta tcgcacttga agatctcaac 2820
agcggattta agaattctcg ggtgaaagtc gaaaaacagg tgtatcaaaa attcgaaaag 2880
atgctgatcg acaagctcaa ttatatggtt gataaaaaa gcaacccatg cgccacgggg 2940

```

ggtgcgctta agggctatca gattacgaac aaatttgaat cttcaagtc aatgtcgacg 3000
caaaatgggt ttatatctta tataccggcg tggcttacat ctaaaataga tcctagcact 3060
gggttcgtga acctgctgaa aaccaagtac acttcaatcg cagattctaa aaaatttata 3120
agcagcttcg acagaatcat gtatgtgccc gaggaagacc tcttcgagtt tgcccttgat 3180
tacaanaaatt tctcaagaac ggatgcagac tacataaaga agtggaagct gtactcttat 3240
gggaaccgga ttcggatatt cagaaatccg aaaaaaaca atgtctttga ttgggaggaa 3300
gtttgtctta cctctgctta caaagagctg ttcaataaat atggcattaa ttaccagcaa 3360
ggtgatatacc gggcgctcct ttgcgaacag tctgacaaag ctttctattc ttcatttatg 3420
gcgctcatgt cattgatgct gcagatgagg aatagcatta cggggaggac tgatgttgac 3480
tttctgatct cgcccgtaga aaattctgat ggaatcttct acgattccag gaattatgag 3540
gccaggaana atgctatcct tccaagaac gcagacgcaa atggcgcgta caatatagct 3600
cgcaaggttt tgtgggctat aggccaattc aagaaagccg aagacgaaaa gctggacaaa 3660
gttaagattg ctatatctaa caaagagtgg cttgagtatg cgcaaacatc tgttaaacac 3720
aaacgccccg cggctacaaa gaaggctggc caggccaaga agaagaaggg ctcggggtcg 3780
gggtcgggct cgggctcgga cgccctggac gacttcgacc tcgacatgct gggctccgac 3840
gccctcgatg atttcgacct cgatatgctc ggcagcgacg cgctcgatga cttcgacctc 3900
gatatgctgg ggagcgacgc cctcgacgat ttgacctcg atatgctgat caactccgcg 3960
tccagcggca gcccgaagaa gaagcgcaaa gtgggctcgc agtacctgcc cgacaccgac 4020
gacaggcaca ggatcgagga gaagcgcaag aggacgtacg agaccttcaa gtccatcatg 4080
aagaagtccc cgctcagcgg cccaacggac ccccgccgcg cgccgaggag gatcgccgtg 4140
ccgtccaggt ccagcgctc ggtcccaag ccggccccgc agccctacct gtccacgtcc 4200
agcctcagca ccatcaacta cgacgagttc cccacctggt tgttcccgtc cggccagatc 4260
tcccaggcca gcgcgctggc cccgcgccc ccgcaggtgc tgccccaggc tccggcccc 4320
gctccggccc cggccatggt ctccgcgctg gccagggcgc ccgccccggt gcccgctcctc 4380
gcgcccggcc cgccgcaggc ggtcgcccc ccagcgccga agcccacgca ggccggcgag 4440
ggcacccctca gcgaggcgct cctgcagctg cagttcgacg acgaggacct cggcgccctc 4500
ctgggcaact cgaccgacct cgccgtgtt accgacctgg cctccgtcga caacagcgag 4560
ttccagcagc tgctgaacca gggcatcccg gtggcgccgc acaccacgga gcccatgctg 4620
atggagtacc cggaggcgat cacgcgctc gtcaccggcg cccagaggcc cccggacccc 4680
gccccggccc cgctcggcgc cccaggcctg ccgaacggcc tcctgagcgg cgacgaggac 4740
ttctccagca tcgcgacat ggactttctc gccctcctgg ggtcgggctc gggcagccgc 4800
gacagcaggg agggcatgtt cctccaaaag ccgaggcgcg gctccgcat ctcggacgtg 4860
ttcgagggca gggaggtctg ccagccaaag cgcacaggc cgttccacc gccgggtctc 4920
ccgtgggcca accggccgct ccccgccagc ctggtccaa ccccgaccgg ccccgtcac 4980
gagccggtcg gcagcctgac gcccgcgccc gtgccccagc cgctcgacc cgcgccggcc 5040
gtaccccccg aggcctccca cctcctggag gaccccgacg aggagacctc gcaggcctg 5100
aaggccctga gggagatggc cgacaccgtc atccccaga aggaggaggc ggccatctgc 5160
ggccagatgg acctgtcgca cccgcccgcg cgcgccacc tcgacgagct gaccacgacc 5220
ctcgagtcca tgaccgagga cctcaacctg gacagcccc tcacgccgga gctgaacgag 5280

```

atcctcgaca ccttcctgaa cgacgagtgc ctctgcacg ccatgcacat ctccacgggc 5340
 ctgagcatct tcgacaccag cctcttctga 5370
 <210> 290
 <211> 5370
 <212> DNA
 <213> 人工序列
 <220>
 <223> dLbCpf1 (RVR) -VPR
 <400> 290
 atggcatcta gcatggcacc aaagaaaaaa aggaaagttt ccaaacttga aaaattttaca 60
 aactgctact ccctttccaa gacgcttagg tttaaagcga tccccgttgg caagacccaa 120
 gagaatatcg ataacaaaag acttctggtc gaagatgaaa aaagggccga agactacaag 180
 ggggtcaaga agttgctcga tcgctattat ctttccttta tcaacgatgt gcttcattca 240
 atcaaactga agaacttga taactacatt agccttttca gaaagaaaac gaggactgaa 300
 aaggagaaca aggaacttga gaatcttgaa ataaaccttc gcaaagaaat tgcaaaagcc 360
 ttcaagggga acgaaggata taaatctctt ttcaaaaaag acattataga aacaattttg 420
 cctgagtttc ttgacgacaa ggatgaaatt gcgctcgtca atagctttta cggattttaca 480
 actgccttca cagggttctt cgacaatagg gagaatatgt ttagcgagga ggcaaaaagc 540
 acatccatcg cattcagatg catcaatgaa aatcttaccg ggtacatatc gaatatggac 600
 atatttgaaa aagtggatgc aatattcgat aagcacgaag tccaggagat aaaggaaaag 660
 atactgaata gcgactatga tgtcgaagat ttttcgaag gtgagttctt caactttgtc 720
 ctgactcaag aaggcattga tgtctataat gcaataattg gaggttttgt gactgagtct 780
 ggcgagaaga taaagggctt gaacgagtat atcaatctct acaaccagaa gactaagcaa 840
 aagttgccta aatttaaacc gctttacaag caagttttga gcgaccggga aagcctttcc 900
 ttttacggtg aaggatacac gagcgatgaa gaagtcctcg aagtcttccg caacacactc 960
 aacaagaact cagaaatctt ttcctcaatt aaaaaattgg agaagctttt caagaacttc 1020
 gatgaatact cttcggcggg gatTTTTgtg aagaacggcc cggcaatttc cacaatatct 1080
 aaagacattt tcggagaatg gaacgtgata agagacaagt ggaatgcgga gtatgatgac 1140
 atacacctga agaagaagc agttgtgact gaaaaatacg aagatgacag gagaaaaagc 1200
 tttaaaaaga tcgggtcctt ttacttgaa cagctgcagg agtatgccga cgccgatctt 1260
 tcggttgctg aaaagctcaa agaaataatt atccagaagg tcgatgaaat ctacaaggtg 1320
 tacggctcaa gcgagaagct ctttgatgct gacttcgtgt tggagaagtc tcttaaaaaa 1380
 aacgacgcag tcgtcgcgat aatgaaagat ttgctggatt cagtgaaatc cttcgagaat 1440
 tatatcaaag cttcttcgg cgaggggaag gagacaaaca gggatgagtc cttctatgga 1500
 gacttcgttc tggcttacga catccttctt aaggtcgacc acatctatga cgcaattcgg 1560
 aactatgtga cgcagaagcc gtattcgaaa gataagtta agctctatct ccaaaacctt 1620
 caatttatgc gtgggtggga taaagacgta gagaccgatc gccgggcaac aattttgcgg 1680
 tacgggtcta aatattacct cgctataatg gataagaaat acgctaaatg tctccagaaa 1740
 attgacaaag atgacgtcaa cggcaattat gaaaaaatca attataaact ctttcctggc 1800

ccaaataaaa tgctcccgaa ggtgtttttt tccaaaaagt ggatggccta ttataatcca 1860
tcagaggata ttcagaaaat ctataaaaat gggaccttta agaagggtga catgtttaac 1920
ctgaacgatt gccacaagct tatagatttt ttcaaagact ctattagccg ctatcccaaa 1980
tggtctaatag cttatgattt caacttctct gaaactgaaa agtaciaaaga tattgcagga 2040
ttctaccgcg aagttgaaga acaagggttat aaggtttctt ttgagtctgc gtccaagaaa 2100
gaggtcgata agttggtcga agaagggaat ttgtatatgt ttcaaattta caataaagac 2160
ttttccgaca agtcccatgg tacacctaat ctgcatacca tgtacttcaa actgctgttc 2220
gatgagaata atcacgggtca gattcgcttg agcggagggg cggaactctt catgaggaga 2280
gcatcggttg aaaaaagagga gctcgtcgtg catccggcta acagcccat tgctaacaag 2340
aatccggata atccaaagaa gactactacc ctctctatg acgtctataa ggataagaga 2400
ttctctgagg accagtacga gttgcacatc cctattgcga taaataaatg ccctaagaac 2460
atctttaaaa tcaatactga ggtcagagtc ctgcttaagc acgacgaca cccgtatgtg 2520
atcgggattg ctagggtga aaggaaactg ctttatattg tggttgtcga tggaaaaggt 2580
aatatagtg aacaatactc tctgaatgaa attatcaaca acttcaatgg cattaggatc 2640
aagaccgact atcattctct gttggacaag aaagagaaag agcgcttcga ggcacggcaa 2700
aactggacgt ctattgagaa catcaaggag cttaggctg gttacatttc tcaggttgtg 2760
cacaaaattt gcgaactggt cgagaaatat gatgccgta tcgcacttga agatctcaac 2820
agcggattta agaattctcg ggtgaaagtc gaaaaacagg tgtatcaaaa attcgaaaag 2880
atgctgatcg acaagctcaa ttatatggtt gataaaaaga gcaacccatg cgccacgggg 2940
ggtgcgctta agggctatca gattacgaac aaatttgaat cttcaagtc aatgtcgacg 3000
caaaatgggt ttatatctta tataccggcg tggtttacat ctaaaataga tcctagcact 3060
gggttcgtga acctgctgaa aaccaagtac acttcaatcg cagattctaa aaaatttata 3120
agcagcttcg acagaatcat gtatgtgccc gaggaagacc tcttcgagtt tgcccttgat 3180
tacaaaaatt tctcaagaac ggatgcagac tacataaaga agtggaagct gtactcttat 3240
gggaaccgga ttcggatatt cagaaatccg aaaaaaaca atgtctttga ttgggaggaa 3300
gtttgtctta cctctgctta caaagagctg ttcaataaat atggcattaa ttaccagcaa 3360
ggtgatatcc gggcgctcct ttgcgaacag tctgacaaag ctttctattc ttcatattatg 3420
gcgctcatgt cattgatgct gcagatgagg aatagcatta cggggaggac tgatgttgac 3480
tttctgatct cgcccgtaaa aaattctgat ggaattctt acgattccag gaattatgag 3540
gcccaggaaa atgctatcct tcccaagaac gcagacgcaa atggcgcgta caatatagct 3600
cgcaagggtt tgtgggctat aggccaattc aagaaagccg aagacgaaaa gctggacaaa 3660
gttaagattg ctatatctaa caaagagtgg cttgagtatg cgcaaacatc tgtaaacac 3720
aaacgccccg cggtacaaa gaaggtggc caggccaaga agaagaaggg ctcggggtcg 3780
gggtcgggct cgggctcgga cgccctggac gacttcgacc tcgacatgct gggctccgac 3840
gccctcgatg atttcgacct cgatatgctc ggcagcgacg cgctcgatga cttcgacctc 3900
gatatgctgg ggagcgacgc cctcgacgat ttgacctcg atatgctgat caactccgcg 3960
tccagcgga gcccgaaaga gaagcgcaaa gtgggctcgc agtacctgcc cgacaccgac 4020
gacaggcaca ggatcgagga gaagcgcaag aggacgtacg agaccttcaa gtccatcatg 4080
aagaagtccc cgttcagcgg cccaacggac ccccgccgcg cgccgaggag gatcgccgtg 4140

ccgtccaggt ccagcgcgtc ggtccccaag ccggccccgc agccctaccc gttcacgtcc 4200
 agcctcagca ccatcaacta cgacgagttc cccaccatgg tgttcccgtc cggccagatc 4260
 tcccaggcca gcgcgctggc ccccgcgccc ccgcaggtgc tgccccaggc tccggcccc 4320
 gctccggccc cggccatggt ctccgcgctg gcccaggcgc ccgccccggt gcccgtcctc 4380
 gcgcccgggccc cgccgcaggc ggtcgccccg ccagcgccga agcccacgca ggccggcgag 4440
 ggcaccctca gcgaggcgct cctgcagctg cagttcgacg acgaggacct cggcgccctc 4500
 ctgggcaact cgaccgaccc cgccgtgttc accgacctgg cctccgtcga caacagcgag 4560
 ttccagcagc tgctgaacca gggcatcccc gtggcgccgc acaccacgga gcccatgctg 4620
 atggagtagc cggaggcgat cacgcgcctc gtcaccggcg cccagaggcc cccggacccc 4680
 gccccggccc cgctcggcgc cccaggcctg ccgaacggcc tcctgagcgg cgacgaggac 4740
 ttctccagca tcgcggacat ggactttctc gccctcttgg ggtcgggctc gggcagccgc 4800
 gacagcaggg agggcatgtt cctcccaaag cccgaggcgc gctccgcat ctcgacgtg 4860
 ttcgagggca gggaggtctg ccagccaaag cgcatcaggc cgttccaccc gccgggctcc 4920
 ccgtgggcca accggccgct ccccgccagc ctggctccaa ccccgaccgg ccccgctcac 4980
 gagccggtcg gcagcctgac gcccgcgccc gtgccccagc cgctcgaccc cgcgccggcc 5040
 gtcacccccg aggcctccca cctcctggag gaccccgacg aggagacctc gcaggccgtg 5100
 aaggccctga gggagatggc cgacaccgtc atccccaga aggaggaggc ggccatctgc 5160
 ggccagatgg acctgtcga cccgccgccg cgcggccacc tcgacgagct gaccacgacc 5220
 ctcgagtcca tgaccgagga cctcaacctg gacagcccc tcacgccgga gctgaacgag 5280
 atcctcgaca ccttcctgaa cgacgagtgc ctctgcacg ccatgcacat ctccacgggc 5340
 ctgagcatct tcgacaccag cctcttctga 5370

<210> 291

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> crGEP186 gRNA

<400> 291

gcaagagagg cgaaggagg ttcc 24

<210> 292

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> crGEP187 gRNA

<400> 292

taaggaggga gtgcattgga ccta 24

<210> 293

<211> 24

<212> DNA
<213> 人工序列
<220>
<223> crGEP188 gRNA
<400> 293
gctctcgctc tctgcatgct agct 24
<210> 294
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> crGEP201 gRNA
<400> 294
gtatcaccca tgggcaatgg ccat 24
<210> 295
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> crGEP208 gRNA
<400> 295
ctcacttcct cgaatcattc taag 24
<210> 296
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> crGEP209 gRNA
<400> 296
ctgaataccc caaaactctc tgct 24
<210> 297
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> crGEP210 gRNA
<400> 297
tgatagcgag atactctata ctta 24
<210> 298

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> crGEP211 gRNA

<400> 298

gtaagtatag agtatctcgc tatc 24

<210> 299

<211> 3841

<212> DNA

<213> 人工序列

<220>

<223> crGEP186 expression plasmid

<400> 299

ctgacgcgcc ctgtagcggc ctgcagtgca gcgtgacccg gtcgtgcccc tctctagaga 60
taatgagcat tgcattgtcta agttataaaa aattaccaca tttttttttt gtcacacttg 120
tttgaagtgc agtttatcta tctttataca tatattttaa ctttactcta cgaataatat 180
aatctatagt actacaataa tatcagtgtt ttagagaatc atataaatga acagtttagac 240
atgggtctaaa ggacaattga gtattttgac aacaggactc tacagtttta tcttttttagt 300
gtgcatgtgt tctccttttt ttttgcaaat agcttcacct atataaact tcatccattt 360
tattagtaca tccatttagg gtttaggggt aatgggtttt atagactaat ttttttagta 420
catctatttt attctatttt agcctctaaa ttaagaaaac taaaactcta ttttagtttt 480
tttatttaat aatttagata taaaatagaa taaaataaag tgactaaaaa ttaaacaat 540
accctttaag aaattaaaaa aactaaggaa acatttttct tgtttcgagt agataatgcc 600
agcctgttaa acgccgtcga tcgacgagtc taacggacac caaccagcga accagcagcg 660
tcgctgcggg ccaagcgaag cagacggcac ggcattctctg tcgctgcctc tggacccctc 720
tcgagagtgc cgctccaccg ttggacttgc tccgctgtcg gcatccagaa attgcgtggc 780
ggagcggcag acgtgagccg gcacggcagg cggcctcctc ctctctcac ggcaccggca 840
gctacggggg attcctttcc caccgtcctc tcgctttccc ttctctgccc gccgtaataa 900
atagacaccc cctccacacc ctctttcccc aacctcgtgt tgttcggagc gcacacacac 960
acaaccagat ctcccccaaa tccaccgctc ggcacctcgc cttcaaggta cgccgctcgt 1020
cctccccccc cccccctctc taccttctct agatcggcgt tccgggtccat ggtaggggcc 1080
cggtagttct acttctgttc atgtttgtgt tagatccgtg tttgtgttag atccgtgctg 1140
ctagcgttcg tacacggatg cgacctgtac gtcagacacg ttctgattgc taacttgcca 1200
gtgtttctct ttggggaatc ctgggatggc tctagccgtt ccgcagacgg gatcgatcta 1260
ggataggtat acatgttgat gtgggtttta ctgatgcata tacatgatgg catatgcagc 1320
atctattcat atgctctaac cttgagtacc tatctattat aataaacaag tatgttttat 1380
aattattttg atcttgatat acttgatga tggcatatgc agcagctata tgtggatttt 1440
tttagccctg ccttcatacg ctatttattt gcttggtact gtttcttttg tcgatgctca 1500

ccctgttggt tgggtgttact tctgcaggga tccaaattac tgatgagtcc gtgaggacga 1560
aacgagtaag ctctgtctaat ttctactaag tgtagatgca agagaggcga aggagggttc 1620
cggccggcat ggtcccagcc tcctcgctgg cgccggctgg gcaacatgct tcggcatggc 1680
gaatgggacc gatcgttcaa acatttggca ataaagtttc ttaagattga atcctgttgc 1740
cggctcttgcg atgattatca tataatttct gttgaattac gttaagcatg taataattaa 1800
catgtaatgc atgacgttat ttatgagatg ggtttttatg attagagtcc cgcaattata 1860
catttaatac gcgatagaaa acaaaatata gcgcgcaaac taggataaat tatcgcgcg 1920
ggtgtcatct atgttactag atcgatcgtc gttcggctgc ggcgagcggg atcagctcac 1980
tcaaaggcgg taatacgggt atccacagaa tcaggggata acgcaggaaa gaacatgtga 2040
gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat 2100
aggctccgcc cccctgacga gcatcacaaa aatcgacgt caagtcagag gtggcgaaac 2160
ccgacaggac tataaagata ccaggcggtt cccctggaa gctccctcgt gcgctctcct 2220
gttccgaccc tgccgcttac cggatactg tccgccttcc tcccttcggg aagcgtggcg 2280
ctttctcata gctcacgctg taggtatctc agttcgggtg aggtcgttcg ctccaagctg 2340
ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg ccttatccgg taactatcgt 2400
cttgagtcca acccggttaag acacgactta tcgccactgg cagcagccac tggtaacagg 2460
attagcagag cgaggtatgt aggcggtgct acagagtctt tgaagtggg gcctaactac 2520
ggctacacta gaagaacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga 2580
aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggttttttt 2640
gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt 2700
tctacggggg ctgacgctca gtggaacgaa aactcacgtt aagggatttt ggtcatgaga 2760
ttatcaaaaa ggatcttcac ctagatcctt ttaaattaa aatgaagttt taaatcaatc 2820
taaagtatat atgagtaaac ttggtctgac agttaccaat gcttaatcag tgaggcacct 2880
atctcagcga tctgtctatt tcgttcatcc atagttgcct gactccccgt cgtgtagata 2940
actacgatac gggaggggctt accatctggc ccagtgctg caatgatacc gcgagaccca 3000
cgctcaccgg ctccagattt atcagcaata aaccagccag ccggaagggc cgagcgcaga 3060
agtggtcctg caactttatc cgctccatc cagtctatta attgttgccg ggaagctaga 3120
gtaagtagtt cgccagttaa tagtttgcgc aacgttggtg ccattgctac aggcatcgtg 3180
gtgtcacgct cgtcgtttgg tatggcttca ttcagctccg gttcccaacg atcaaggcga 3240
gttacatgat ccccatgtt gtgcaaaaaa gcggttagct ccttcgggtc tccgatcgtt 3300
gtcagaagta agttggccgc agtgttatca ctcatggta tggcagcact gcataattct 3360
cttactgtca tgccatccgt aagatgctt tctgtgactg gtgagtactc aaccaagtca 3420
ttctgagaat agtgatatgc gcgaccgagt tgctcttgcc cggcgtcaat acgggataat 3480
accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgttc ttcggggcga 3540
aaactctcaa ggatcttacc gctgttgaga tccagttcga tgtaaccac tcgtgcaccc 3600
aactgatctt cagcatcttt tactttcacc agcgtttctg ggtgagcaaa aacaggaagg 3660
caaaatgccg caaaaaaggg aataaggcg acacggaaat gttgaatact catactcttc 3720
ctttttcaat attattgaag catttatcag ggttattgtc tcatgagcgg atacatattt 3780
gaatgtattt agaaaaataa acaaataggg gttccgcga catttccccg aaaagtgcc 3840

c 3841

<210> 300

<211> 3841

<212> DNA

<213> 人工序列

<220>

<223> crGEP187 expression plasmid

<400> 300

```

ctgacgcgcc ctgtagcggc ctgcagtgca gcgtgacccg gtcgtgcccc tctctagaga 60
taatgagcat tgcatgtcta agttataaaa aattaccaca tatttttttt gtcacacttg 120
tttgaagtgc agtttatcta tctttataca tatatttaaa ctttactcta cgaataatat 180
aatctatagt actacaataa tatcagtgtt ttagagaatc atataaatga acagtttagac 240
atgggtctaaa ggacaattga gtattttgac aacaggactc tacagtttta tcttttttagt 300
gtgcatgtgt tctccttttt ttttgcaaat agcttcacct atataatact tcatccattt 360
tattagtaca tccatttagg gtttaggggt aatgggtttt atagactaat ttttttagta 420
catctatttt attctatttt agcctctaaa ttaagaaaac taaaactcta ttttagtttt 480
tttatttaat aatttagata taaaatagaa taaaataaag tgactaaaaa ttaaacaat 540
accctttaag aaattaaaaa aactaaggaa acatttttct tgtttcgagt agataatgcc 600
agcctgttaa acgccgtcga tcgacgagtc taacggacac caaccagcga accagcagcg 660
tcgcgtcggg ccaagcgaag cagacggcac ggcatctctg tcgctgcctc tggacccctc 720
tcgagagtgc cgctccaccg ttggacttgc tccgctgtcg gcattccaga attgcgtggc 780
ggagcggcag acgtgagccg gcacggcagg cggcctcctc ctctctcac ggcaccggca 840
gctacggggg attcctttcc caccgtcctc tcgttttccc ttctctgccc gccgtaataa 900
atagacaccc cctccacacc ctctttcccc aacctcgtgt tgttcggagc gcacacacac 960
acaaccagat ctcccccaaa tccaccgctc ggcacctccg cttcaaggta cgccgctcgt 1020
cctccccccc cccccctctc taccttctct agatcggcgt tccggtccat ggtagggcc 1080
cggtagtctt acttctgttc atgtttgtgt tagatccgtg tttgtgttag atccgtgctg 1140
ctagcgttcg tacacggatg cgacctgtac gtcagacacg ttctgattgc taacttgcca 1200
gtgtttctct ttggggaatc ctgggatggc tctagccgtt ccgcagacgg gatcgatcta 1260
ggataggtat acatgttgat gtgggtttta ctgatgcata tacatgatgg catatgcagc 1320
atctattcat atgctctaac cttgagtacc tatctattat aataaacaag tatgttttat 1380
aattattttg atcttgatat acttggtatg tggcatatgc agcagctata tgtggatttt 1440
tttagccctg ccttcatacg ctatttattt gcttgggtact gtttcttttg tcgatgctca 1500
ccctgttggt tgggtgttact tctgcaggga tccaaattac tgatgagtcc gtgaggacga 1560
aacgagtaag ctcgtctaata ttctactaag tgtagattaa ggaggagtg cattggacct 1620
aggccggcat ggccccagcc tcctcgtctg cgccggctgg gcaacatgct tcggcatggc 1680
gaatgggacc gatcgttcaa acatttggca ataaagtttc ttaagattga atcctgttgc 1740
cggctcttgc atgattatca tataatttct gttgaattac gttaagcatg taataattaa 1800
catgtaatgc atgacgttat ttatgagatg ggtttttatg attagagtcc cgcaattata 1860

```

catttaatac gcgatagaaa acaaaatata gcgcgcaaac taggataaat tatcgcgcg 1920
gggtgtcatct atgttactag atcgatcgtc gttcggctgc ggcgagcggg atcagctcac 1980
tcaaaggcgg taatacgggt atccacagaa tcaggggata acgcaggaaa gaacatgtga 2040
gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat 2100
aggctccgcc cccctgacga gcatcacaaa aatcgacgct caagtcagag gtggcgaaac 2160
ccgacaggac tataaaagata ccaggcggtt cccctggaa gctccctcgt gcgctctcct 2220
gttccgaccc tgccgcttac cggatacctg tccgcctttc tcccttcggg aagcgtggcg 2280
ctttctcata gctcacgctg taggtatctc agttcgggtg aggtcgttcg ctccaagctg 2340
ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg ccttatccgg taactatcgt 2400
cttgagtcca acccggttaag acacgactta tcgccactgg cagcagccac tggtaacagg 2460
attagcagag cgaggtatgt aggcggtgct acagagttct tgaagtggtg gcctaactac 2520
ggctacacta gaagaacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga 2580
aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggttttttt 2640
gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt 2700
tctacggggt ctgacgctca gtggaacgaa aactcacgtt aagggatttt ggtcatgaga 2760
ttatcaaaaa ggatcttcac ctagatcctt ttaaattaaa aatgaagttt taaatcaatc 2820
taaagtatat atgagtaaac ttggtctgac agttaccaat gcttaatcag tgaggcacct 2880
atctcagcga tctgtctatt tcgttcaccc atagttgcct gactccccgt cgtgtagata 2940
actacgatac gggagggtt accatctggc ccagtgctg caatgatacc gcgagaccca 3000
cgctcaccgg ctccagattt atcagcaata aaccagccag ccggaagggc cgagcgcaga 3060
agtggtcctg caactttatc cgcctccatc cagtctatta attgttgccg ggaagctaga 3120
gtaagtagtt cgccagttaa tagtttgcg aacgttggtt ccattgctac aggcatcgtg 3180
gtgtcacgct cgtcgtttgg tatggcttca ttcagctccg gttcccaacg atcaaggcga 3240
gttacaatgat ccccatggt gtgcaaaaaa gcggttagct ccttcggtcc tccgatcgtt 3300
gtcagaagta agttggccgc agtggtatca ctcatggta tggcagcact gcataattct 3360
cttactgtca tgccatccgt aagatgcttt tctgtgactg gtgagtactc aaccaagtca 3420
ttctgagaat agtgtatgcg gcgaccgagt tgctcttgcc cggcgtcaat acgggataat 3480
accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgttc ttcggggcga 3540
aaactctcaa ggatcttacc gctgttgaga tccagttcga tgtaaccac tcgtgcaccc 3600
aactgatctt cagcatcttt tactttcacc agcgtttctg ggtgagcaaa aacaggaagg 3660
caaaatgccg caaaaaagg aataagggcg acacggaaat gttgaatact catactcttc 3720
ctttttcaat attattgaag cttttatcag gtttattgtc tcatgagcgg atacatattt 3780
gaatgtattt agaaaaataa acaaataggg gttccgcga catttccccg aaaagtgcc 3840

c 3841

<210> 301

<211> 3841

<212> DNA

<213> 人工序列

<220>

<223> crGEP188 expression plasmid

<400> 301

```

ctgacgcgcc ctgtagcggc ctgcagtgca gcgtgacccg gtcgtgcccc tctctagaga 60
taatgagcat tgcattgtcta agttataaaa aattaccaca tttttttttt gtcacacttg 120
tttgaagtgc agtttatcta tctttataca tatatttaaa ctttactcta cgaataatat 180
aatctatagt actacaataa tatcagtgtt ttagagaatc atataaatga acagtttagac 240
atgggtctaaa ggacaattga gtattttgac aacaggactc tacagtttta tcttttttagt 300
gtgcatgtgt tctccttttt ttttgcaaat agcttcacct atataatact tcatccattt 360
tattagtaca tccatttagg gtttaggggt aatgggtttt atagactaat ttttttagta 420
catctatttt attctatttt agcctctaaa ttaagaaaac taaaactcta ttttagtttt 480
tttatttaat aatttagata taaaatagaa taaaataaag tgactaaaaa ttaaacaat 540
accctttaag aaattaaaaa aactaaggaa acatttttct tgtttcgagt agataatgcc 600
agcctgttaa acgccgtcga tcgacgagtc taacggacac caaccagcga accagcagcg 660
tcgcgtcggg ccaagcgaag cagacggcac ggcattctctg tcgctgcctc tggacccctc 720
tcgagagttc cgctccaccg ttggacttgc tccgctgtcg gcatccagaa attgcgtggc 780
ggagcggcag acgtgagccg gcacggcagg cggcctcctc ctctctcac ggcaccggca 840
gctacggggg attcctttcc caccgtcctc tcgctttccc ttctcgcgc gccgtaataa 900
atagacaccc cctccacacc ctctttcccc aacctcgtgt tgttcggagc gcacacacac 960
acaaccagat ctcccccaaa tccaccgctc ggcacctccg cttcaaggta cgccgctcgt 1020
cctccccccc cccccctctc taccttctct agatcggcgt tccggtccat ggttagggcc 1080
cggtagtctt acttctgttc atgtttgtgt tagatccgtg tttgtgttag atccgtgctg 1140
ctagcgttcg tacacggatg cgacctgtac gtcagacacg ttctgattgc taacttgcca 1200
gtgtttctct ttggggaatc ctgggatggc tctagccgtt ccgcagacgg gatcgatcta 1260
ggatagggtat acatgttgat gtgggtttta ctgatgcata tacatgatgg catatgcagc 1320
atctattcat atgctctaac cttagtacc tatctattat aataaacaag tatgttttat 1380
aattattttg atcttgatat acttgatga tggcatatgc agcagctata tgtggatttt 1440
tttagccctg ccttcatacg ctatttattt gcttggtact gtttcttttg tcgatgctca 1500
ccctgttggt tgggtgttact tctgcaggga tccaaattac tgatgagtcc gtgaggacga 1560
aacgagtaag ctctgctaata ttctactaag tgtagatgct ctgctctctc gcatgctagc 1620
tggccggcat ggtcccagcc tctcgtctgg cgccggctgg gcaacatgct tcggcatggc 1680
gaatgggacc gatcgttcaa acatttggca ataaagtttc ttaagattga atcctgttgc 1740
cggtcttgcg atgattatca tataatttct gttgaattac gtttaagcatg taataattaa 1800
catgtaatgc atgacgttat ttatgagatg ggtttttatg attagagtcc cgcaattata 1860
catttaatac gcgatagaaa acaaaatata gcgcgcaaac taggataaat tatcgcgcg 1920
ggtgtcatct atgttactag atcgatcgtc gttcggctgc ggcgagcgg atcagctcac 1980
tcaaaggcgg taatacgggt atccacagaa tcaggggata acgcaggaaa gaacatgtga 2040
gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat 2100
aggctccgcc cccctgacga gcatcacaaa aatcgacgtc caagtcagag gtggcgaaac 2160
ccgacaggac tataaagata ccaggcgttt ccccttgaa gctccctcgt gcgctctcct 2220

```

gttccgaccc tgccgcttac cggataacctg tccgcctttc tcccttcggg aagcgtggcg 2280
 ctttctcata gctcacgctg taggtatctc agttcggtgt aggtcgttcg ctccaagctg 2340
 ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg ccttatccgg taactatcgt 2400
 cttgagtcca acccggttaag acacgactta tcgccactgg cagcagccac tggtaacagg 2460
 attagcagag cgaggatatgt aggcgggtgct acagagtctt tgaagtgggt gcctaactac 2520
 ggctacacta gaagaacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga 2580
 aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggttttttt 2640
 gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt 2700
 tctacggggg ctgacgctca gtggaacgaa aactcacgtt aagggtttt ggtcatgaga 2760
 ttatcaaaaa ggatcttcac ctagatcctt ttaaattaaa aatgaagttt taaatcaatc 2820
 taaagtatat atgagtaaac ttggtctgac agttaccaat gcttaatcag tgaggcacct 2880
 atctcagcga tctgtctatt tcgttcatec atagttgect gactccccgt cgtgtagata 2940
 actacgatac gggagggtt accatctggc ccagtgctg caatgatacc gcgagacca 3000
 cgctcaccgg ctccagattt atcagcaata aaccagccag ccggaagggc cgagcgcaga 3060
 agtggtcctg caactttatc cgctccatc cagtctatta attgttgccg ggaagctaga 3120
 gtaagtagtt cgccagttaa tagtttgccg aacgttggtt ccattgctac aggcacgtg 3180
 gtgtcacgct cgtcgtttgg tatggcttca ttcagctccg gttcccaacg atcaaggcga 3240
 gttacatgat ccccatgtt gtgcaaaaaa gcggttagct ccttcggtcc tccgatcgtt 3300
 gtcagaagta agttggccgc agtggttatca ctcatggtta tggcagcact gcataattct 3360
 cttactgtca tgccatccgt aagatgcttt tctgtgactg gtgagtactc aaccaagtca 3420
 ttctgagaat agtgatgcg gcgaccgagt tgctcttgcc cggcgtcaat acgggataat 3480
 accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgttc ttcggggcga 3540
 aaactctcaa ggatcttacc gctgttgaga tccagttcga tgtaaccac tcgtgcaccc 3600
 aactgatctt cagcatcttt tactttcacc agcgtttctg ggtgagcaaa aacaggaagg 3660
 caaatgccg caaaaaaggg aataaggcg acacggaaat gttgaatact catactcttc 3720
 ctttttcaat attattgaag catttatcag ggttattgtc tcatgagcgg atacatattt 3780
 gaatgtattt agaaaaataa acaaataggg gttccgcgca catttccccg aaaagtgcc 3840
 c 3841

<210> 302

<211> 3841

<212> DNA

<213> 人工序列

<220>

<223> crGEP201 expression plasmid

<400> 302

ctgacgcgcc ctgtagcggc ctgcagtgcg gcgtgaccgc gtcgtgcccc tctctagaga 60
 taatgagcat tgcattgcta agttataaaa aattaccaca tttttttttt gtcacacttg 120
 tttgaagtgc agtttatcta tctttataca tatatttaaa ctttactcta cgaataatat 180
 aatctatagt actacaataa tatcagtgtt ttagagaatc atataaatga acagtttagac 240

```

atggtctaaa ggacaattga gtattttgac aacaggactc tacagtttta tcttttttagt 300
gtgcatgtgt tctccttttt ttttgcaa atgttcacct atataatact tcatccattt 360
tattagtaca tccatttagg gtttaggggt aatggttttt atagactaat ttttttagta 420
catctatttt attctatttt agcctctaaa ttaagaaaac taaaactcta ttttagtttt 480
tttatttaat aatttagata taaaatagaa taaaataaag tgactaaaaa ttaaacaat 540
accctttaag aaattaaaaa aactaaggaa acatttttct tgtttcgagt agataatgcc 600
agcctgttaa acgccgtcga tcgacgagtc taacggacac caaccagcga accagcagcg 660
tcgctgcggg ccaagcgaag cagacggcac ggcctctctg tcgctgcctc tggacccctc 720
tcgagagtgc cgctccaccg ttggacttgc tccgctgtcg gcatccagaa attgcgtggc 780
ggagcggcag acgtgagccg gcacggcagg cggcctctc ctcctctcac ggcaccggca 840
gctacggggg attcctttcc caccgtcct tcgttttccc ttctcgcgc gccgtaataa 900
atagacaccc cctccacacc ctctttcccc aacctcgtgt tgttcggagc gcacacacac 960
acaaccagat ctcccccaaa tccaccgtc ggcacctccg cttcaaggta cgcgcgtcgt 1020
cctccccccc cccccctctc taccttctct agatcggcgt tccggtccat ggttagggcc 1080
cggtagttct acttctgttc atgtttgtgt tagatccgtg tttgtgttag atccgtgctg 1140
ctagcgttcg tacacggatg cgacctgtac gtcagacacg ttctgattgc taacttgcca 1200
gtgtttctct ttggggaatc ctgggatggc tctagccgtt ccgcagacgg gatcgatcta 1260
ggataggtat acatgttgat gtgggtttta ctgatgcata tacatgatgg catatgcagc 1320
atctattcat atgctctaac cttgagtacc tatctattat aataaacaag tatgttttat 1380
aattattttg atcttgatat acttgatga tggcatatgc agcagctata tgttgatttt 1440
tttagccctg ccttcatacg ctatttattt gcttggtact gtttcttttg tcgatgtcta 1500
ccctgttggt tgggtgtact tctgcaggga tccaaattac tgatgagtcc gtgaggacga 1560
aacgagtaag ctgctcta atttactaag tgtagatgta tcacccatgg gcaatggcca 1620
tggccggcat ggtcccagcc tcctcgtcgg cgcggcgtgg gcaacatgct tcggcatggc 1680
gaatgggacc gatcgttcaa acatttggca ataaagtttc ttaagattga atcctgttgc 1740
cggctcttgc atgattatca tataatttct gttgaattac gtttagcatg taataattaa 1800
catgtaatgc atgacgttat ttatgagatg ggtttttatg attagagtcc cgcaattata 1860
catttaatac gcgatagaaa acaaaatata gcgcgcaaac taggataaat tatcgcgcg 1920
ggtgtcatct atgttactag atcgatcgtc gttcggtcgc ggcgagcggg atcagctcac 1980
tcaaaggcgg taatacgggt atccacagaa tcaggggata acgcaggaaa gaacatgtga 2040
gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat 2100
aggctccgcc cccctgacga gcatcaciaa aatcgacgt caagtcagag gtggcgaaac 2160
ccgacaggac tataaagata ccaggcgttt cccctggaa gctccctcgt gcgctctcct 2220
gttccgacct tgccgcttac cggatactg tccgctttc tcccttcggg aagcgtggcg 2280
ctttctcata gctcacgctg taggtatctc agttcgtgt aggtcgttcg ctccaagctg 2340
ggctgtgtgc acgaaccccc cgttcagccc gaccgtcgc cttatccgg taactatcgt 2400
cttgagtcca acccgtaag acacgactta tcgccactgg cagcagccac tggtaacagg 2460
attagcagag cgaggtatgt aggcggtgct acagagttct tgaagtgtg gcctaactac 2520
ggctacacta gaagaacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga 2580

```

aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggttttttt 2640
 gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt 2700
 tctacggggg ctgacgctca gtggaacgaa aactcacgtt aagggatttt ggtcatgaga 2760
 ttatcaaaaa ggatcttcac ctagatcctt ttaaattaaa aatgaagttt taaatcaatc 2820
 taaagtatat atgagtaaac ttggtctgac agttaccaat gcttaatcag tgaggcacct 2880
 atctcagcga tctgtctatt tcgttcatcc atagtgcct gactccccgt cgtgtagata 2940
 actacgatac gggagggtt accatctggc cccagtgtg caatgatacc gcgagaccca 3000
 cgctcaccgg ctccagattt atcagcaata aaccagccag ccggaagggc cgagcgcaga 3060
 agtggtcctg caactttatc cgctccatc cagtctatta attgttgccg ggaagctaga 3120
 gtaagtagtt cgccagttaa tagtttgccg aacgttggtt ccattgctac aggcattcgtg 3180
 gtgtcacgct cgtcgtttgg tatggttca ttcagctccg gttcccaacg atcaaggcga 3240
 gttacatgat ccccatgtt gtgcaaaaaa gcggttagct ctttcggtcc tccgatcgtt 3300
 gtcagaagta agttggccgc agtggtatca ctcatggtta tggcagcact gcataattct 3360
 cttactgtca tgccatccgt aagatgcttt tctgtgactg gtgagtactc aaccaagtca 3420
 ttctgagaat agtgtatgcg gcgaccgagt tgctcttgcc cggcgtcaat acgggataat 3480
 accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgttc ttcggggcga 3540
 aaactctcaa ggatcttacc gctgttgaga tccagttcga tgtaaccac tcgtgcaccc 3600
 aactgatctt cagcatcttt tactttcacc agcgtttctg ggtgagcaaa aacaggaagg 3660
 caaatgccg caaaaaaggg aataaggcg acacggaaat gttgaatact catactcttc 3720
 ctttttcaat attattgaag catttatcag ggttattgtc tcatgagcgg atacatattt 3780
 gaatgtattt agaaaaataa acaaataggg gttccgcga catttccccg aaaagtgcc 3840
 c 3841

<210> 303

<211> 3841

<212> DNA

<213> 人工序列

<220>

<223> crGEP208 expression plasmid

<400> 303

ctgacgcgcc ctgtagcggc ctgcagtgc gcgtgaccgc gtcgtgcccc tctctagaga 60
 taatgagcat tgcattgtta agttataaaa aattaccaca tttttttttt gtcacacttg 120
 tttgaagtgc agtttatcta tctttataca tatatttaaa ctttactcta cgaataatat 180
 aatctatagt actacaataa tatcagtgtt ttagagaatc atataaatga acagtttagac 240
 atgggtctaaa ggacaattga gtattttgac aacaggactc tacagtttta tcttttttagt 300
 gtgcatgtgt tctccttttt ttttgcaaat agcttcacct atataatact tcatccattt 360
 tattagtaca tccatttagg gtttagggtt aatggttttt atagactaat ttttttagta 420
 catctatttt attctatttt agcctctaaa ttaagaaaac taaaactcta ttttagtttt 480
 tttatttaaat aatttagata taaaatagaa taaaataaag tgactaaaaa ttaaacaat 540
 accctttaag aaattaaaaa aactaaggaa acatttttct tgtttcgagt agataatgcc 600

```

agcctgttaa acgccgtcga tcgacgagtc taacggacac caaccagcga accagcagcg 660
tcgcgtcggg ccaagcgaag cagacggcac ggcattctctg tcgctgcctc tggacccctc 720
tcgagagttc cgctccaccg ttggacttgc tccgctgtcg gcatccagaa attgcgtggc 780
ggagcggcag acgtgagccg gcacggcagg cggcctcctc ctctctcac ggcaccggca 840
gctacggggg attcctttcc caccgtcctc tcgctttccc ttctcgcgc gccgtaataa 900
atagacaccc cctccacacc ctctttcccc aacctcgtgt tgttcggagc gcacacacac 960
acaaccagat ctcccccaaa tccacccgtc ggcacctccg cttcaaggta cgccgctcgt 1020
cctccccccc cccccctctc taccttctct agatcggcgt tccgggtccat ggtaggggc 1080
cggtagttct acttctgttc atgttttgtt tagatccgtg tttgtgttag atccgtgctg 1140
ctagcgttcg tacacggatg cgacctgtac gtcagacacg ttctgattgc taacttgcca 1200
gtgtttctct ttggggaatc ctgggatggc tctagccgtt ccgcagacgg gatcgatcta 1260
ggataggtat acatgttgat gtgggtttta ctgatgcata tacatgatgg catatgcagc 1320
atctattcat atgctctaac cttgagtacc tatctattat aataaacaag tatgttttat 1380
aattattttg atcttgatat acttgatga tggcatatgc agcagctata tgtggatttt 1440
tttagccctg ccttcatacg ctatttattt gcttggtact gtttcttttg tcgatgtca 1500
ccctgttggt tgggtgtact tctgcaggga tccaaattac tgatgagtcc gtgaggacga 1560
aacgagtaag ctctctaat ttctactaag tgtagatctc acttcctcga atcattctaa 1620
gggccggcat ggtcccagcc tcctcgtcgg cgccggctgg gcaacatgct tcggcatggc 1680
gaatgggacc gatcgttcaa acatttggca ataaagtttc ttaagattga atcctgttgc 1740
cggtcttgcg atgattatca tataatttct gttgaattac gttaagcatg taataattaa 1800
catgtaatgc atgacgttat ttatgagatg ggtttttatg attagagtcc cgcaattata 1860
catttaatac gcgatagaaa acaaaatata gcgcgcaaac taggataaat tatcgcgcg 1920
gggtgtcatct atgttactag atcgatcgtc gttcggctgc ggcgagcgg atcagctcac 1980
tcaaaggcgg taatacgggt atccacagaa tcaggggata acgcaggaaa gaacatgtga 2040
gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat 2100
aggctccgcc cccctgacga gcatcacaaa aatcgacgct caagtcagag gtggcgaaac 2160
ccgacaggac tataaagata ccaggcgtt cccctggaa gctccctcgt gcgctctcct 2220
gttccgaccc tgccgcttac cggatactg tccgccttc tcccttcggg aagcgtggcg 2280
ctttctcata gctcacgtg taggtatctc agttcgggtg aggtcgttcg ctccaagctg 2340
ggctgtgtgc acgaaccccc cgttcagccc gaccgtgcg cttatccgg taactatcgt 2400
cttgagtcca acccggttaag acacgactta tcgccactgg cagcagccac tggtaacagg 2460
attagcagag cgaggtatgt aggcggtgct acagagttct tgaagtggg gcctaactac 2520
ggctacacta gaagaacagt atttggtatc tgcgtctgc tgaagccagt taccttcgga 2580
aaaagagttg gtagctcttg atccggcaaa caaacaccg ctggtagcgg tggttttttt 2640
gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt 2700
tctacggggt ctgacgtca gtggaacgaa aactcacgtt aagggtttt ggtcatgaga 2760
ttatcaaaaa ggatcttcac ctagatcctt ttaaattaaa aatgaagttt taaatcaatc 2820
taaagtatat atgagtaaac ttggtctgac agttaccaat gcttaatcag tgaggcacct 2880
atctcagcga tctgtctatt tcgttcatcc atagttgcct gactccccgt cgtgtagata 2940

```

actacgatac gggagggcctt accatctggc cccagtgtg caatgatacc gcgagaccca 3000
 cgctcaccgg ctccagattt atcagcaata aaccagccag ccggaaggcg cgagcgcaga 3060
 agtggtcctg caactttatc cgcctccatc cagtctatta attgttgccg ggaagctaga 3120
 gtaagtagtt cgccagttaa tagtttgccg aacgttggtt ccattgctac aggcattcgtg 3180
 gtgtcacgct cgctgtttgg tatggcttca ttcagctccg gttcccaacg atcaaggcga 3240
 gttacatgat ccccatgtt gtgcaaaaaa gcggttagct ccttcgggtc tccgatcgtt 3300
 gtcagaagta agttggccgc agtgttatca ctcatggta tggcagcact gcataattct 3360
 cttactgtca tgccatccgt aagatgcttt tctgtgactg gtgagtactc aaccaagtca 3420
 ttctgagaat agtgtatgcg gcgaccgagt tgctcttgcc cggcgtcaat acgggataat 3480
 accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgttc ttcggggcga 3540
 aaactctcaa ggatcttacc gctgttgaga tccagttcga tgtaaccac tcgtgcaccc 3600
 aactgatctt cagcatcttt tactttcacc agcgtttctg ggtgagcaaa aacaggaagg 3660
 caaaatgccg caaaaaaggg aataagggcg acacggaaat gttgaatact catactcttc 3720
 ctttttcaat attattgaag catttatcag ggttattgtc tcatgagcgg atacatattt 3780
 gaatgtattt agaaaaataa acaaataggg gttccgcgca catttccccg aaaagtgcc 3840

c 3841

<210> 304

<211> 3841

<212> DNA

<213> 人工序列

<220>

<223> crGEP209 expression plasmid

<400> 304

ctgacgcgcc ctgtagcggc ctgcagtga gcgtgaccg gtcgtgcccc tctctagaga 60
 taatgagcat tgcatgtcta agttataaaa aattaccaca ttttttttt gtcacacttg 120
 tttgaagtgc agtttatcta tctttataca tatatttaaa ctttactcta cgaataatat 180
 aatctatagt actacaataa tatcagtgtt ttagagaatc atataaatga acagtttagac 240
 atgggtctaaa ggacaattga gtattttgac aacaggactc tacagtttta tcttttttagt 300
 gtgcatgtgt tctccttttt ttttgcaaat agcttcacct atataatact tcatccattt 360
 tattagtaca tccatttagg gtttaggggtt aatggttttt atagactaat ttttttagta 420
 catctatttt attctatttt agcctctaaa ttaagaaaac taaaactcta ttttagtttt 480
 tttatttaaat aatttagata taaaatagaa taaaataaag tgactaaaaa ttaaacaat 540
 accctttaag aaattaaaaa aactaaggaa acatttttct tgtttcgagt agataatgcc 600
 agcctgttaa acgccgtcga tcgacgagtc taacggacac caaccagcga accagcagcg 660
 tcgcgtcggg ccaagcgaag cagacggcac ggcattctctg tcgctgcctc tggacccctc 720
 tcgagagttc cgctccaccg ttggacttgc tccgctgtcg gcatccagaa attgcgtggc 780
 ggagcggcag acgtgagccg gcacggcagg cggcctctc ctctctcac ggcaccggca 840
 gctacggggg attcctttcc caccgtcct tcgctttccc ttctcgcgc gccgtaataa 900
 atagacaccc cctccacacc ctctttcccc aacctcgtgt tgttcggagc gcacacacac 960

acaaccagat ctccccaaa tccaccgctc ggcacctccg cttcaaggta cgccgctcgt 1020
cctccccccc cccccctctc taccttctct agatcggcgt tccggtccat ggtagggcc 1080
cggtagtctt acttctgttc atgtttgtgt tagatccgtg tttgtgttag atccgtgctg 1140
ctagcggttcg tacacggatg cgacctgtac gtcagacacg ttctgattgc taacttgcca 1200
gtgtttctct ttggggaatc ctgggatggc tctagccgtt ccgcagacgg gatcgatcta 1260
ggataggtat acatgttgat gtgggtttta ctgatgcata tacatgatgg catatgcagc 1320
atctattcat atgctctaac cttgagtacc tatctattat aataaacaag tatgttttat 1380
aattattttg atcttgatat acttggtatg tggcatatgc agcagctata tgtggatttt 1440
tttagccctg ccttcatacg ctatttattt gcttggtact gtttcttttg tcgatgctca 1500
ccctgttggt tgggtgttact tctgcaggga tccaaattac tgatgagtcc gtgaggacga 1560
aacgagtaag ctctgtcta tttactaag tgtagatctg aatacccaa aactctctgc 1620
tggccggcat ggtcccagcc tcctcgtggt cgccggctgg gcaacatgct tcggcatggc 1680
gaatgggacc gatcggtcaa acatttggca ataaagtttc ttaagattga atcctgttgc 1740
cggtcttgcg atgattatca tataatttct gttgaattac gttaagcatg taataattaa 1800
catgtaatgc atgacgttat ttatgagatg ggtttttatg attagagtcc cgcaattata 1860
catttaatac gcgatagaaa acaaaatata gcgcgcaaac taggataaat tatcgcgcg 1920
gggtgtcatct atgttactag atcgatcgct gttcggctgc ggcgagcgg atcagctcac 1980
tcaaaggcgg taatacgggt atccacagaa tcaggggata acgcaggaaa gaacatgtga 2040
gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat 2100
aggctccgcc cccctgacga gcatcacaaa aatcgacgt caagtcagag gtggcgaaac 2160
ccgacaggac tataaagata ccaggcgtt cccctggaa gctccctcgt gcgctctcct 2220
gttccgaccc tgccgcttac cggatacctg tccgccttc tcccttcggg aagcgtggcg 2280
ctttctcata gctcacgtg taggtatctc agttcgggtg aggtcgttcg ctccaagctg 2340
ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg ccttatccgg taactatcgt 2400
cttgagtcca acccggttaag acacgactta tcgccactgg cagcagccac tggtaacagg 2460
attagcagag cgaggatatg aggcgggtgct acagagtct tgaagtgggt gcctaactac 2520
ggctacacta gaagaacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga 2580
aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggttttttt 2640
gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt 2700
tctacggggt ctgacgtca gtggaacgaa aactcacgtt aagggtttt ggtcatgaga 2760
ttatcaaaaa ggatcttcac ctatgacctt ttaaattaaa aatgaagttt taaatcaatc 2820
taaagtatat atgagtaaac ttggtctgac agttaccaat gcttaatcag tgaggcacct 2880
atctcagcga tctgtctatt tcgttcaccc atagttgcct gactccccgt cgtgtagata 2940
actacgatac gggagggtt accatctggc ccagtgctg caatgatacc gcgagaccca 3000
cgctcaccgg ctccagattt atcagcaata aaccagccag ccggaagggc cgagcgcaga 3060
agtggtcctg caactttatc cgctccatc cagtctatta attgttgccg ggaagctaga 3120
gtaagtagtt cgccagttaa tagtttgccg aacgttggtt ccattgctac aggcacgtg 3180
gtgtcacgct cgtcgtttgg tatggcttca ttcagctccg gttcccaacg atcaaggcga 3240
gttacatgat ccccatggt gtgcaaaaaa gcggttagct ccttcggtcc tccgatcgtt 3300

gtcagaagta agttggccgc agtgttatca ctcatggtta tggcagcact gcataattct 3360
 cttactgtca tgccatccgt aagatgcttt tctgtgactg gtgagtactc aaccaagtca 3420
 ttctgagaat agtgtatgcg gcgaccgagt tgctcttgcc cggcgtcaat acgggataat 3480
 accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgttc ttcggggcga 3540
 aaactctcaa ggatcttacc gctgttgaga tccagttcga tgtaaccacac tcgtgcaccc 3600
 aactgatctt cagcatcttt tactttcacc agcgtttctg ggtgagcaaa aacaggaagg 3660
 caaaatgccg caaaaaaggg aataagggcg acacggaaat gttgaatact catactcttc 3720
 ctttttcaat attattgaag catttatcag ggttattgtc tcatgagcgg atacatattt 3780
 gaatgtattt agaaaaataa acaaataggg gttccgcgca catttccccg aaaagtgcc 3840
 c 3841

<210> 305

<211> 3841

<212> DNA

<213> 人工序列

<220>

<223> crGEP210 expression plasmid

<400> 305

ctgacgcgcc ctgtagcggc ctgcagtga gcgtgacccg gtcgtgcccc tctctagaga 60
 taatgagcat tgcattgcta agttataaaa aattaccaca tatttttttt gtcacacttg 120
 tttgaagtgc agtttatcta tctttataca tatatttaaa ctttactcta cgaataatat 180
 aatctatagt actacaataa tatcagtgtt ttagagaatc atataaatga acagtttagac 240
 atgggtctaaa ggacaattga gtattttgac aacaggactc tacagtttta tcttttttagt 300
 gtgcatgtgt tctccttttt ttttgcaaat agcttcacct atataatact tcatccattt 360
 tattagtaca tccatttagg gtttaggggt aatggttttt atagactaat ttttttagta 420
 catctatttt attctatttt agcctctaaa ttaagaaaac taaaactcta ttttagtttt 480
 tttattttaat aatttagata taaaatagaa taaaataaag tgactaaaaa ttaaacaat 540
 accctttaag aaattaaaaa aactaaggaa acatttttct tgtttcgagt agataatgcc 600
 agcctgttaa acgccgtcga tcgacgagtc taacggacac caaccagcga accagcagcg 660
 tcgctgcggg ccaagcgaag cagacggcac ggcattctctg tcgctgcctc tggacccctc 720
 tcgagagttc cgctccaccg ttggacttgc tccgtgtcg gcatccagaa attgcgtggc 780
 ggagcggcag acgtgagccg gcacggcagg cggcctcctc ctctctcac ggcaccggca 840
 gctacggggg attcctttcc caccgtcct tcgttttccc ttctcgcgc gccgtaataa 900
 atagacaccc cctccacacc ctctttcccc aacctcgtgt tgttcggagc gcacacacac 960
 acaaccagat ctcccccaaa tccaccgctc ggcacctccg cttcaaggta cgccgctcgt 1020
 cctccccccc cccccctctc taccttctct agatcggcgt tccgggtccat ggttagggcc 1080
 cggtagttct acttctgttc atgtttgtgt tagatccgtg tttgtgttag atccgtgctg 1140
 ctagcgttcg tacacggatg cgacctgtac gtcagacacg ttctgattgc taacttgcca 1200
 gtgtttctct ttggggaatc ctgggatggc tctagccgtt ccgcagacgg gatcgatcta 1260
 ggataggtat acatgttgat gtgggtttta ctgatgcata tacatgatgg catatgcagc 1320

atctattcat atgctctaac cttgagtacc tatctattat aataaacaag tatgttttat 1380
aattatitttg atcttgatat acttggatga tggcatatgc agcagctata tgtggatttt 1440
tttagccctg ccttcatacg ctatttattt gcttgggtact gtttcttttg tcgatgctca 1500
ccctgttggt tgggtgttact tctgcaggga tccaaattac tgatgagtcc gtgaggacga 1560
aacgagtaag ctcgtctaata ttctactaag tgtagattga tagcgagata ctctatactt 1620
aggccggcat ggtcccagcc tcctcgtggt cgccgggtgg gcaacatgct tcggcatggc 1680
gaatgggacc gatcgttcaa acatttggca ataaagtctt ttaagattga atcctgttgc 1740
cggctcttgcg atgattatca tataatttct gttgaattac gttaagcatg taataattaa 1800
catgtaatgc atgacgttat ttatgagatg ggtttttatg attagagtcc cgcaattata 1860
catttaatac gcgatagaaa acaaaatata gcgcgcaaac taggataaat tatcgcgcg 1920
ggtgtcatct atgttactag atcgatcgtc gttcgggtgc ggcgagcgg atcagctcac 1980
tcaaaggcgg taatacgggt atccacagaa tcaggggata acgcaggaaa gaacatgtga 2040
gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat 2100
aggctccgcc cccctgacga gcatcacaaa aatcgacgt caagtcagag gtggcgaaac 2160
ccgacaggac tataaagata ccaggcgtt ccccttgaa gctccctcgt gcgtctctct 2220
gttccgaccc tgccgcttac cggatacctg tccgccttcc tcccttcggg aagcgtggcg 2280
ctttctcata gctcacgtg taggtatctc agttcgggtg aggtcgttcg ctccaagctg 2340
ggctgtgtgc acgaaccccc cgttcagccc gaccgtgcg ccttatccgg taactatcgt 2400
cttgagtcca acccggttaag acacgactta tcgccactgg cagcagccac tggtaacagg 2460
attagcagag cgaggtatgt aggcggtgct acagagttct tgaagtgtg gcctaactac 2520
ggctacacta gaagaacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga 2580
aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggttttttt 2640
gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt 2700
tctacggggg ctgacgctca gtggaacgaa aactcacgtt aagggtttt ggtcatgaga 2760
ttatcaaaaa ggatcttcac ctatgcctt ttaaattaa aatgaagttt taaatcaatc 2820
taaagtatat atgagtaaac ttggtctgac agttaccaat gcttaatcag tgaggcacct 2880
atctcagcga tctgtctatt tcgttcaccc atagttgcct gactccccgt cgtgtagata 2940
actacgatac gggagggtt accatctggc ccagtgctg caatgatacc gcgagaccca 3000
cgctcaccgg ctccagattt atcagcaata aaccagccag ccggaagggc cgagcgcaga 3060
agtggctctg caactttatc cgctccatc cagtctatta attgttgccg ggaagctaga 3120
gtaagtagtt cgccagttaa tagtttgccg aacgttggtt ccattgctac aggcatcgtg 3180
gtgtcacgct cgtcgtttgg tatggcttca ttcagctccg gttcccaacg atcaaggcga 3240
gttacatgat ccccatgtt gtgcaaaaaa gcggttagct ccttcgggtcc tccgatcgtt 3300
gtcagaagta agttggccgc agtggtatca ctcatggtta tggcagcact gcataattct 3360
cttactgtca tgccatccgt aagatgcttt tctgtgactg gtgagtactc aaccaagtca 3420
ttctgagaat agtgatgcg gcgaccgagt tgctcttgcc cggcgtcaat acgggataat 3480
accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgtt ttcggggcga 3540
aaactctcaa ggatcttacc gctgttgaga tccagttcga tgtaaccac tcgtgcaccc 3600
aactgatctt cagcatcttt tactttcacc agcgtttctg ggtgagcaaa aacaggaagg 3660

caaaatgccg caaaaaaggg aataagggcg acacggaaat gttgaatact catactcttc 3720
 ctttttcaat attattgaag catttatcag ggttattgtc tcatgagcgg atacatattt 3780
 gaatgtattt agaaaaataa acaaataggg gttccgcgca catttccccg aaaagtgcc 3840
 c 3841

<210> 306

<211> 3841

<212> DNA

<213> 人工序列

<220>

<223> crGEP211 expression plasmid

<400> 306

ctgacgcgcc ctgtagcggc ctgcagtgc gctgacccg gtcgtgcccc tctctagaga 60
 taatgagcat tgcatgtcta agttataaaa aattaccaca tttttttttt gtcacacttg 120
 tttgaagtgc agtttatcta tctttataca tatatttaaa ctttactcta cgaataatat 180
 aatctatagt actacaataa tatcagtgtt ttagagaatc atataaatga acagtttagac 240
 atgggtctaaa ggacaattga gtattttgac aacaggactc tacagtttta tcttttttagt 300
 gtgcatgtgt tctccttttt ttttgcaaag agcttcacct atataatact tcatccattt 360
 tattagtaca tccatttagg gtttaggggt aatgggtttt atagactaat ttttttagta 420
 catctatttt attctatttt agcctctaaa ttaagaaaac taaaactcta ttttagtttt 480
 tttatttaaat aatttagata taaaatagaa taaaataaag tgactaaaaa ttaaacaat 540
 accctttaag aaattaaaaa aactaaggaa acatttttct tgtttcgagt agataatgcc 600
 agcctgttaa acgccgtcga tcgacgagtc taacggacac caaccagcga accagcagcg 660
 tcgctgcggg ccaagcgaag cagacggcac ggcctctctg tcgctgcctc tggacccctc 720
 tcgagagttc cgctccaccg ttggacttgc tccgctgtcg gcatccagaa attgcgtggc 780
 ggagcggcag acgtgagccg gcacggcagg cggcctcctc ctctctcac ggcaccggca 840
 gctacggggg attcctttcc caccgctcct tcgctttccc ttcctcgccc gccgtaataa 900
 atagacaccc cctccacacc ctctttcccc aacctcgtgt tgttcggagc gcacacacac 960
 acaaccagat ctcccccaaa tccaccgctc ggcacctccg cttcaaggta cgccgctcgt 1020
 cctccccccc cccccctctc taccttctct agatcggcgt tccgggtccat ggtagggcc 1080
 cggtagttct acttctgttc atgtttgtgt tagatccgtg tttgtgttag atccgtgctg 1140
 ctagcgttcg tacacggatg cgacctgtac gtcagacacg ttctgattgc taacttgcca 1200
 gtgtttctct ttggggaatc ctgggatggc tctagccgtt ccgcagacgg gatcgatcta 1260
 ggataggtat acatgttgat gtgggtttta ctgatgcata tacatgatgg catatgcagc 1320
 atctattcat atgctctaac cttgagtacc tatctattat aataaacaag tatgttttat 1380
 aattattttg atcttgatat acttgatga tggcatatgc agcagctata tgtggatttt 1440
 tttagccctg ccttcatacg ctattttatt gcttggtact gtttcttttg tcgatgctca 1500
 ccctgttggt tgggtgtact tctgcaggga tccaaattac tgatgagtcc gtgaggacga 1560
 aacgagtaag ctcgtctaag ttctactaag tgtagatgta agtatagagt atctcgtat 1620
 cggccggcat ggtcccagcc tcctcgtcgt cgccggctgg gcaacatgct tcggcatggc 1680

gaatgggacc gatcgttcaa acatttggca ataaagtttc ttaagattga atcctgttgc 1740
cggtcttgcg atgattatca tataatttct gttgaattac gttaagcatg taataattaa 1800
catgtaatgc atgacgttat ttatgagatg ggtttttatg attagagtcc cgcaattata 1860
catttaatac gcgatagaaa acaaaatata gcgcgcaaac taggataaat tatcgcgcg 1920
ggtgtcatct atgttactag atcgatcgtc gttcggctgc ggcgagcgg atcagctcac 1980
tcaaaggcgg taatacggtt atccacagaa tcaggggata acgcaggaaa gaacatgtga 2040
gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat 2100
aggctccgcc cccctgacga gcatcaciaa aatcgacgt caagtcagag gtggcgaaac 2160
ccgacaggac tataaagata ccaggcgttt cccctggaa gctccctcgt gcgctctcct 2220
gttccgaccc tgccgcttac cggatacctg tccgccttcc tcccttcggg aagcgtggcg 2280
ctttctcata gctcacgtg taggtatctc agttcgggtg aggtcgttcg ctccaagctg 2340
ggctgtgtgc acgaaccccc cgttcagccc gaccgtgcg ccttatccg taactatcgt 2400
cttgagtcca acccggttaag acacgaetta tcgccactgg cagcagccac tggtaacagg 2460
attagcagag cgaggtatgt aggcggtgct acagagttct tgaagtgtg gcctaactac 2520
ggctacacta gaagaacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga 2580
aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggttttttt 2640
gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt 2700
tctacggggc ctgacgctca gtggaacgaa aactcacgtt aagggatttt ggtcatgaga 2760
ttatcaaaaa ggatcttcac ctatgcctt ttaaattaa aatgaagttt taaatcaatc 2820
taaagtatat atgagtaaac ttggtctgac agttaccaat gcttaatcag tgaggcacct 2880
atctcagcga tctgtctatt tcgttcaccc atagttgcct gactccccgt cgtgtagata 2940
actacgatac gggagggtt accatctggc ccagtgctg caatgatacc gcgagacca 3000
cgctcaccgg ctccagattt atcagcaata aaccagccag ccggaaggcg cgagcgcaga 3060
agtggctctg caactttatc cgctccatc cagtctatta attgttgccg ggaagctaga 3120
gtaagtagtt cgccagttaa tagtttgccg aacgttggtt ccattgctac aggcatcgtg 3180
gtgtcacgct cgtcgtttgg tatggcttca ttcagctccg gttcccaacg atcaaggcga 3240
gttacatgat ccccatgtt gtgcaaaaaa gcggttagct ccttcggtcc tccgatcgtt 3300
gtcagaagta agttggccgc agtggtatca ctcatggta tggcagcact gcataattct 3360
cttactgtca tgccatccgt aagatgcttt tctgtgactg gtgagtactc aaccaagtca 3420
ttctgagaat agtgtatgcg gcgaccgagt tgctcttgcc cggcgtcaat acgggataat 3480
accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgttc ttcggggcga 3540
aaactctcaa ggatcttacc gctgttgaga tccagttcga tgtaaccac tcgtgcaccc 3600
aactgatctt cagcatcttt tactttcacc agcgtttctg ggtgagcaaa aacaggaagg 3660
caaaatgccg caaaaaaggg aataagggcg acacggaaat gttgaatact catactcttc 3720
ctttttcaat attattgaag cttttatcag gtttattgtc tcatgagcgg atacatattt 3780
gaatgtattt agaaaaataa acaaataggg gttccgcgca catttccccg aaaagtgcc 3840
c 3841

<210> 307

<211> 2040

<212> DNA

<213> Zea mays

<400> 307

atggcttcag cgaacaactg gctgggcttc tcgctctcgg gccaggataa cccgcagcct 60
aaccaggata gctcgcctgc cgccggtatc gacatctccg gcgccagcga cttctatggc 120
ctgcccacgc agcagggctc cgacgggcat ctcggcgtgc cgggcctgcg ggacgatcac 180
gcttcttatg gtatcatgga ggcctacaac agggttcctc aagaaacca agattggaac 240
atgaggggct tggactacaa cggcgggtggc tcggagctct cgatgcttgt ggggtccagc 300
ggcggcggcg ggggcaacgg caagagggcc gtggaagaca gcgagcccaa gctcgaagat 360
ttcctcggcg gcaactcggt cgtctccgat caagatcagt ccggcggtta cctgtttctt 420
ggagtcccga tagccagcag cgccaatagc aacagcggga gcaacacat ggagctctcc 480
atgatcaaga cctggctacg gaacaaccag gtggcccagc cccagccgcc agctccacat 540
cagccgcagc ctgaggaaat gagcaccgac gccagcggca gcagctttgg atgctcggat 600
tcgatgggaa ggaacagcat ggtggcggct ggtgggagct cgcagagcct ggcgctctcg 660
atgagcacgg gctcgcacct gcccattggt gtgcccagcg gcgccgccag cggagcggcc 720
tcggagagca catcgtcga gaacaagcga gcgagcgggt ccatggattc gcccggcagc 780
gcggtagaag ccgtaccgag gaagtccatc gacacgttcg ggcaaaggac ctctatatat 840
cgaggtgtaa caaggcatag atggacaggc cggtatgagg ctcatctatg ggataatagt 900
tgtagaaggg aagggcagag tcgcaagggt aggcaagttt accttggtgg ctatgacaag 960
gaggacaagg cagcaagggc ttatgatttg gcagctctca agtattgggg cactacgaca 1020
acaacaaatt tccctataag caactacgaa aaggagctag aagaaatgaa acatatgact 1080
agacaggagt acattgcata cctaagaaga aatagcagtg gattttctcg tggggcgctca 1140
aagtatcgtg gagtaactag acatcatcag catgggagat ggcaagcaag gatagggaga 1200
gttgcaggaa acaaggatct ctacttgggc acattcagca ccgaggagga ggcggcggag 1260
gcctacgaca tcgcccgcat caagttccgc ggtctcaacg ccgtcaccaa cttcgacatg 1320
agccgctacg acgtgaagag catcctcgag agcagcacac tgcctgtcgg cgggtcggcc 1380
aggcgctca aggacgccgt ggaccacgtg gaggccggcg ccaccatctg gcgcgccgac 1440
atggacggcg ccgtgatctc ccagctggcc gaagccggga tgggcggcta cgcctctac 1500
ggccaccacg gctggccgac catcgcgttc cagcagccgt cgccgctctc cgtccactac 1560
ccgtacggcc agccgtcccc cggttggtgc aaacccgagc aggacgcggc cgccgccgcg 1620
gcgcacagcc tgcaggacct ccagcagctg cacctcggca gcgcggccca caacttcttc 1680
caggcgctgt cgagctccac agtctacaac ggcggcgccg gcgccagtgg tgggtaccag 1740
ggcctcgggt gtggcagctc tttctcatg ccgtcgagca ctgtcgtggc ggcggccgac 1800
caggggcaca gcagcacggc caaccagggg agcacgtgca gctacgggga cgaccaccag 1860
gaggggaagc tcatcggtta cgacgccgcc atggtggcga ccgcagctgg tggagaccgc 1920
tacgtgcgg cgaggaacgg gtaccagttc tcgcagggt cgggatccac ggtgagcatc 1980
gcgagggcga acgggtacgc taacaactgg agctctcctt tcaacaacgg catggggtga 2040

<210> 308

<211> 978

<212> DNA

<213> Zea mays

<400> 308

atggcggcca atgcgggcgg cggtggagcg ggaggaggca gcggcagcgg cagcgtggct 60
gcgccggcgg tgtgccgccc cagcggctcg cggtggacgc cgacgccgga gcagatcagg 120
atgctgaagg agctctacta cggtgcggc atccggtcgc ccagctcgga gcagatccag 180
cgcatcaccg ccatgctgcg gcagcacggc aagatcgagg gcaagaacgt cttctactgg 240
ttccagaacc acaaggcccc cgagcgccag aagcgccgcc tcaccagcct cgacgtcaac 300
gtgcccggcg ccggcgcggc cgacgccacc accagccaac tcggcgtcct ctgctgtcg 360
tcgccgcctt caggcgcggc gcctccctcg cccaccctcg gcttctacgc cgccggcaat 420
ggcgcgcgat cggtggggt gctggacacg agttccgact ggggcagcag cggcgctgcc 480
atggccaccg agacatgctt cctgcaggac tacatgggcg tgacggacac gggcagctcg 540
tcgcagtggc catgcttctc gtcgtcggac acgataatgg cggcgggcgg gcgcgcggcg 600
cgggtggcga cgacgcgggc gcccagaca ctccctctct tcccgaactg cggcgacgac 660
gacgacgacg acagccagcc cccgccgcgg ccgcggcacg cagtcccagt cccggcaggc 720
gagaccatcc gcggcgggcg cggcagcagc agcagctact tgccgttctg gggtgccggt 780
gccgcgtcca caactgccgg cgccacttct tccgttgca tccagcagca acaccagctg 840
caggagcagt acagctttta cagcaacagc acccagctgg ccggcaccgg cagccaagac 900
gtatcggtt cagcggccgc cctggagctg agcctcagct catggtgctc cccttacct 960
gctgcaggga gcatgtga 978

<210> 309

<211> 1754

<212> DNA

<213> Zea mays

<400> 309

atcggaccca aatcatagac acatgatgat ataataacag acaacaaaaa ttgagagtgg 60
caaaatagca aatttctgat agtcatgtga tagagaatag tagacaattt tgacataata 120
tatgtacact aattagtcaa caaaagcgat attgcggtta aaacagtgat tgccagtgtt 180
ttgacccgag tgcctaacc aaccaataaa gtaaatttat gctatgtgtc ctgctccaga 240
tggtatgatgc aagaagacac aagatttatt ttggttcgga caatagaagg cctactttca 300
gcggaggggg atgggattta tattatcttg cacctaagtg cttgtagtag aaggtacaag 360
ttagtcgaga gagagagaga atcccaactc tctgcggatg attgaggcaa gtgtcaatat 420
cggccgcgga gggcaatagg tgaagtgtat tgtctctctc ccttgcaagc cttggactcc 480
ttttatagcc ttaatgaggg aatcaaggag taataattag ttgaagactg attaagaaac 540
agtccatctg ttagtttttt tgtttaaata ggctaaagct aattttatct agttcttaat 600
tagctaataa ttattatttc gtaggatcca aaccattcct aagctatagt gctattatat 660
caagtgtaga tctatatgta ctcaaggcca tgatgtttgc aaaccaacaa tgaaatttat 720
cgcacacatt ggtcatggca gatcaacttt tttgccacaa aacaacaag aatagtgcaa 780
acgaagttgc ataaaaatgaa acaatatatt atgtgaatag ttgcatggtt tatcttgcta 840

gttccatttt aacacacaca catatcttgc tagttccatt ttaacttcta cttgcacaat 900
 tccaaaagga acctaaatth catttacga tgagtcacaa gaaacttaga tctaattaaa 960
 tttaaagaaa aatagcaata ttatatthtt taaatatatt tattataaaa atttatctca 1020
 tattctagct aatgatatth attatgcac ataactatta aatatatagc tatatatata 1080
 tatattthcat aagthttcatg ttgtthtaact taatagagat ttatatthttt agggctagtt 1140
 tggcaaaacta thttthccaaa ggattthcat ttctataaag aaaattthtt thttaaaaaa 1200
 aatagaaatc tcttgaaaaga atagaattgt taaactactc ttagacaaat aaagagtatc 1260
 cttggthtcgt ggctaaccgt atcatathtt atctaagtta gthgtthccaa thaaagaact 1320
 aaththttatac acaaaaagtta agtaaagtat agcaaattag tccgcgaacc aaatatgacc 1380
 gaaatatcga ggagtgagga ggcttaaccc thcccatgtg tgtatctact gttacaccgt 1440
 gagctacaaa gthactggca caaacgtata gaggatgggtg aggacatggg aagataaaat 1500
 cctggthccag caagatccgt thttccaaat gggatcaggt gattggctcc agthtctctc 1560
 cccctcagca ccaccagtct cctccagthc agthcccgthc thtctccgct caagagthctc 1620
 agaccaacgg caaagthtcta gaagcacggt tgcacgggca gcacggcata acacctccct 1680
 ccactgatcc agthtccagthc gcccacgccc ccaacgtctt ctcttgcaaa tcgcaagcaa 1740
 acttctctgtt cacg 1754

<210> 310

<211> 658

<212> DNA

<213> *Zea mays*

<400> 310

gttggctact tgagthtagat thtggthgtg thtcatcccc acgtacgtcc agcaaagaaa 60
 aattgaagct agthcatgca tggthcgtca tcaaatgcat ggccggccgg atacaaatth 120
 gaactgtagc tatcgacgta cgcathgtatt aaththtatc agagaagaca aggaacacag 180
 atacatacat gthcgaacaa thaththtcta tggcacttga gctagctagc atacaathth 240
 gththtaaatg aathgaaact gaagacgathc gatcgaattg aaggtthgtg thcgtgagca 300
 atgcaatgca gththcacaga acgthgccaa tgcaacaagc caccaagaaa agagaagthct 360
 actcgathctt gcaatgatta ggctthggatg atgcgtgggg ccacgtacgt atggacathc 420
 aagaacccca thctcagcgt gtggcctgag ggtgatggca aagctgatcc acacathgctg 480
 gccccctthc cccctcaga gacctgacc thccgagcac agccagccac cgcgcaacgc 540
 cggccaccac caccaccacc atacctgcta gcgctagctc ththththth acgccgctg 600
 gtgcgtgcct cgacgacctc actactthtga gctgcaaggt ccgaactaaa aagcacgc 658

<210> 311

<211> 1700

<212> DNA

<213> *Beta vulgaris*

<400> 311

tataagthtca aactthcaata caggtaththt cggtatgtga thacctthaca atthtctatt 60
 thcaaagaat ththacctgtg cagctatgtt ggataacctg tgcgagathc cgtthtcagta 120

ggacactttt tttttttacc aataaaaaaa aacttataag ttcatgagct aatttttata 180
 gatagtttta agtaccgggt ggaggatgaa tagttgagtt ttttcttcaa aattagatac 240
 ttcttccgtt ttttattaga tgttacactt ttcaaatcac ggactcctag gtaatttttg 300
 gagaggagag agatagagag aatgaaaaac aaaagggtcc catgtgagta tgtgatagga 360
 gagagataga gagaatttat tacccaaaat aaaagtgtaa catctaattc aaaacttcct 420
 aaaatagaaa gtgtaacatc taaaaaaaac ggaggaagta tttgaatttg atatagatat 480
 tgtgtctttg tgtgtgttga atttcaattc ccagttccct aaaaaaaatt tacaattgca 540
 atttcgagat tatgatgtaa attaaatttg agagactaga aagtatttg tcaacccaaa 600
 aaaaaaatat caatacttat ataaatcaaa aacataatag agaatccaat ttactaaaa 660
 atattagtaa ttttgattaa aataatctat taaatgaac tctaaccctc acataatttc 720
 cacatattat taatcaacaa aataagcatc acaattattt agaataggcg atctaatttt 780
 aacataaaat tagacgaatt caaattgaat ttttctaaca agctcattcc atttcacgca 840
 acccaaaatt atcctagtca gtagtcatcc attcttttct cattccttta ttcttgatta 900
 tcgaactaca acagataatt tcaaaaaaaa actaaattgg tagtcttaac tgattaaact 960
 acttactaaa tggattaaag aatgtcatta ctgaatagat taaactgatt acgaaataga 1020
 ttaacttggc ccctaaatag attaaattag ttactatatt aaaattaggc gatctcttac 1080
 aaaaccaact gaataagcat agctctgtat attacctaga tttcaactaa atcaaaaccc 1140
 cttacagttc aatctagagc tgatcatttt ggctcggccc gtcccatttt tgggccgggt 1200
 tttagtcaga tttttttggc ccgcggtcgg gcccgccccg attttttttg ctttgggcaa 1260
 gccaaaaacg acttttcagt ttattttttg gcccgaccg tttttaccg caaaagcccg 1320
 ctaatttagg tccgcacttt gggcacaaaa atttagcccc aacttaaacc tggccccgacc 1380
 catgatcacc tctagttaa tccaaactaa aaaactacac aagttagcca aaaattatgt 1440
 ctactttgta caactttata aaatacacac agtagttgat atcttgatga ttaactcctt 1500
 ttgaagtttg actacacacc aaccccaaac acaccactt tttccccct cttgtcacca 1560
 accccccctc ctcttttagc accaaagttt ggttggtgag tcctccataa ctgctaaatt 1620
 ctctcttttt tctctctcct aaaaaactaa aaccaccaa aatttcagac atcaaaaaaa 1680
 ttacaagtga aggaacaat 1700

<210> 312

<211> 991

<212> DNA

<213> Beta vulgaris

<400> 312

aaagaaggaa aggaaggaat ttgaacatgt gacctatcgt tcacagcacc tcaatcttaa 60
 tcactagacc aaaacatcct tggttcttgc gcaagaaggt tggctagaaa tttttgttaa 120
 aaacactagc cccgctcagt tcataatgag aatgtcgatg tcaccaaagg gatattaaat 180
 gaatggaatt gggatatgga tggaatataa tgaaatagag ccactttgag gttccctatg 240
 aaatgaggca tggaaggag ccactacgaa aaagtccgg gagttacgaa ggaagcttcg 300
 agctcatatt ggtcatgaac ccgattactg agtctaataa gttcaattga aaagaaaaag 360
 tcttatgttc taaaagaact tttcgtgcgg tttgcatgag ttcatagtcc atataatata 420

atgcaggaat gaagtctctca gttgattctt ccacacccgt ccctcacccc ctaggcccc 480
ccctcacccc gccgaaaaaa ataaagaaaa tccaacgtta tttttcttag aaatgacagt 540
ttgatataga aaggaaaaat aataataaaa aaaaaaagtg ttggcgtttt cattttcaac 600
ctcagtatgt tggtttgccc caacaagttc tgaaccaatt ggcgatgtaa tcttataaga 660
agaatctaac gttgggtccat tttgcttcta cagttttgaa agttaggtgg gccccattat 720
tatgttgatc ctagaataat taattttggg aggctgagaa gaggaaaaat aaagaacaat 780
gctaaaaaca agtgaaaaat atagttgcaa ctcatgatgc aacatgagat gcgatgaaat 840
atgatagtaa cttgagctca caactctgta tataagtgct catttggaca cttattttct 900
acaatttcct agtaactcag cttagcttca ttcccgaactt ttttataaaa gtcaggacga 960
tcaatatcta tctatttata tgtctgtctg t 991

<210> 313

<211> 8

<212> PRT

<213> 人工序列

<220>

<223> Cys2His2 motif

<220>

<221> VARIANT

<222> (1) .. (1)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> VARIANT

<222> (3) .. (3)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> VARIANT

<222> (5) .. (5)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> VARIANT

<222> (7) .. (7)

<223> Xaa can be any naturally occurring amino acid

<400> 313

Xaa Cys Xaa Cys Xaa His Xaa His

1

5

<210> 314

<211> 9

<212> PRT

<213> 人工序列

<220>

<223> LAGLIDADG motif

<400> 314

Leu Ala Gly Leu Ile Asp Ala Asp Gly

1

5

<210> 315

<211> 360

<212> DNA

<213> Zea mays

<400> 315

cgaccgatg ccgcagccgt agtagagctc cttcagcadc ctgatctgct ccggcgctcg 60
 cgtccaccgc gagccgctgg ggccggcacac cgccggcgca gccacgctgc cgctgcctcc 120
 tcccgctcca ccgccgcccc cattggccgc catgcctcta tctcagcggc cttcctgagc 180
 gctcctgtga cctagctctc cgggtgtccg tctatggcaa gagaggcgaa ggagggttcc 240
 ttgtttataa ggaggagtg cattggacct agaggctaga tagctagaag gtagctagca 300
 tgcagagagc gagagcggga gaagagagcg tagctgcgct aggtgatata ggttggggct 360

<210> 316

<211> 900

<212> DNA

<213> Zea mays

<400> 316

taatcgttct tgacagcaac ctgccagtca aatggccgtg acaacgtata ctattatcga 60
 gtaaaaggct gccactttag tagtacatgt acatgcatgc gcagatacat catcaggtac 120
 tcataatatg gcacacatat agacatgttt tgaggaaaat gagacaaagt atagtggaga 180
 cttccctaga aagcagaaga aaaagaagtg gtttatgttc cgttaaatca tactacaact 240
 tttttttatt atactctcca ttttgtcatc attaggtact catatatggg cacacatata 300
 gtactgccaa tttttcttgc taaaaaaagt tccactatat atatgtatgt atgcacaaat 360
 aaactaatTT tcttagaaaa gaaaaccggt gtaatacata ctaagggcta gtttgggaac 420
 cctggtttcc taaggaattt ttttttcca aaaaaatag tttatttttc cttcggaat 480
 taggaatctc ttataaaaatt cgagttccca aactattcct aatatatata tcatactctc 540
 catcagtcta tatatagatt acatatagta agtatagagt atctcgctat cacatagtgc 600
 cactaatctt ctggagtgtt ccagttgtat aaatatctat cagtatcagc actactgttt 660
 gctgaatacc ccaaaaactct ctgcttgact tctcttccct aacctttgca ctgtccaaaa 720
 tggcttcctg atccccctac ttctctgaat cattctaaga agaaactcaa gccgctacca 780
 ttaggggcag attaatgtct gcactttcag ataatctacc atggccactg tgaacaactg 840
 gctcgttttc tccctctccc cgcaggagct gccgcctcc cagacgacgg actccacgt 900

<210> 317

<211> 281

<212> DNA

<213> Zea mays

<400> 317

atatatagat tacatatagt aagtatagag tatctcgcta tcacatagtg ccactaatct 60
tctggagtgt accagttgta taaatatcta tcagtatcag cactactggt tgctgaatac 120
cccaaaactc tctgcttgac ttctcttccc taacctttgc actgtccaaa atggcttcct 180
gatccccctca cttcctcgaa tcattctaag aagaaactca agccgctacc attaggggca 240
gattaattgc tgcactttca gataatctac catggccact g 281

<210> 318

<211> 372

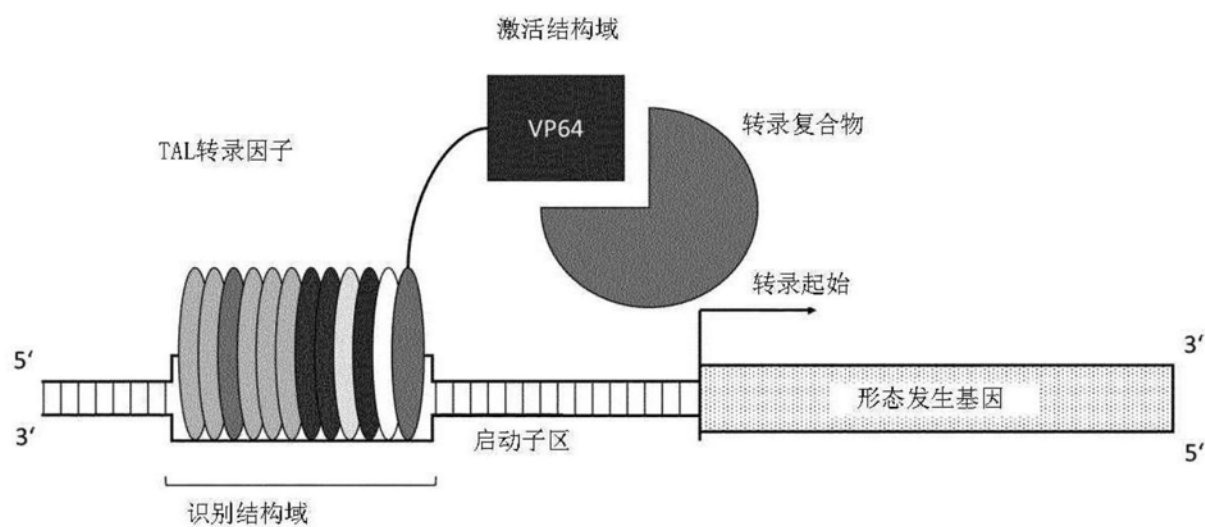
<212> DNA

<213> Zea mays

<400> 318

gatctgctcc ggcgtcggcg tccaccgcga gccgctgggg cggcacaccg ccggcgcagc 60
cacgctgccg ctgcctcctc ccgctccacc gccgcccga ttggccgcca tgcctctatc 120
tcagcggcct tcctgagcgc tcctgtgacc tagctctccg gtgtccggtc tatggcaaga 180
gaggcgaagg agggttcctt gtttataagg agggagtga ttggacctag aggctagata 240
gcatgaaggt agctagcatg cagagagcga gagcgggaga agagagcgta gctgcgctag 300
gtgatatagg ttggggctgg gaggggggtc atggccattg cccatgggtg atacgatatc 360
ttttggagag ag 372

A



B

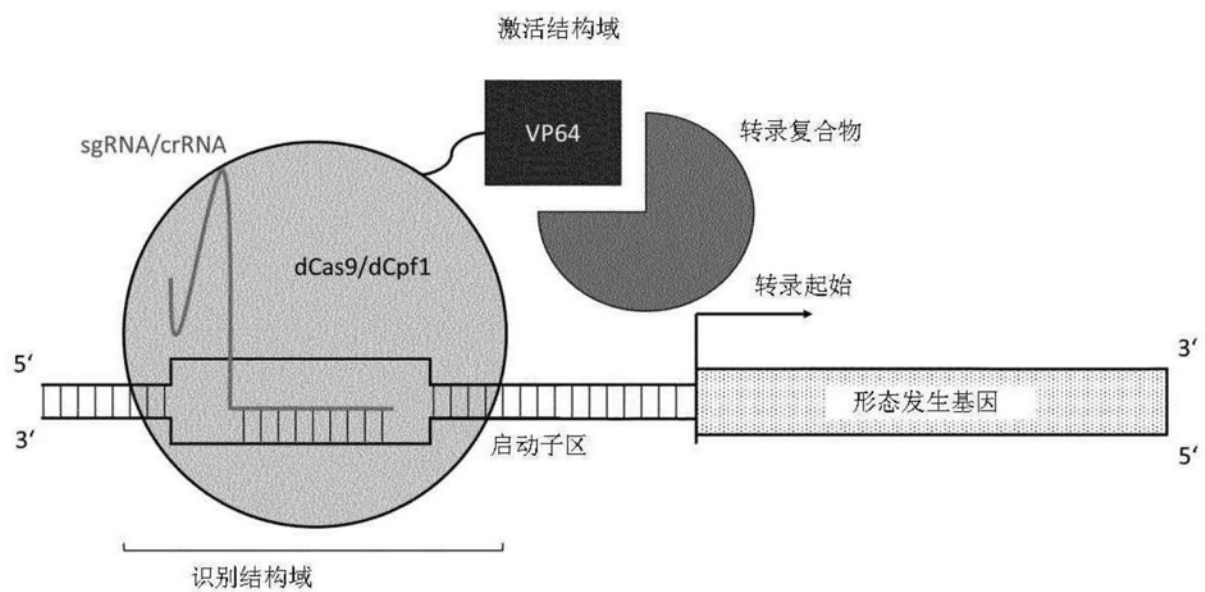


图1

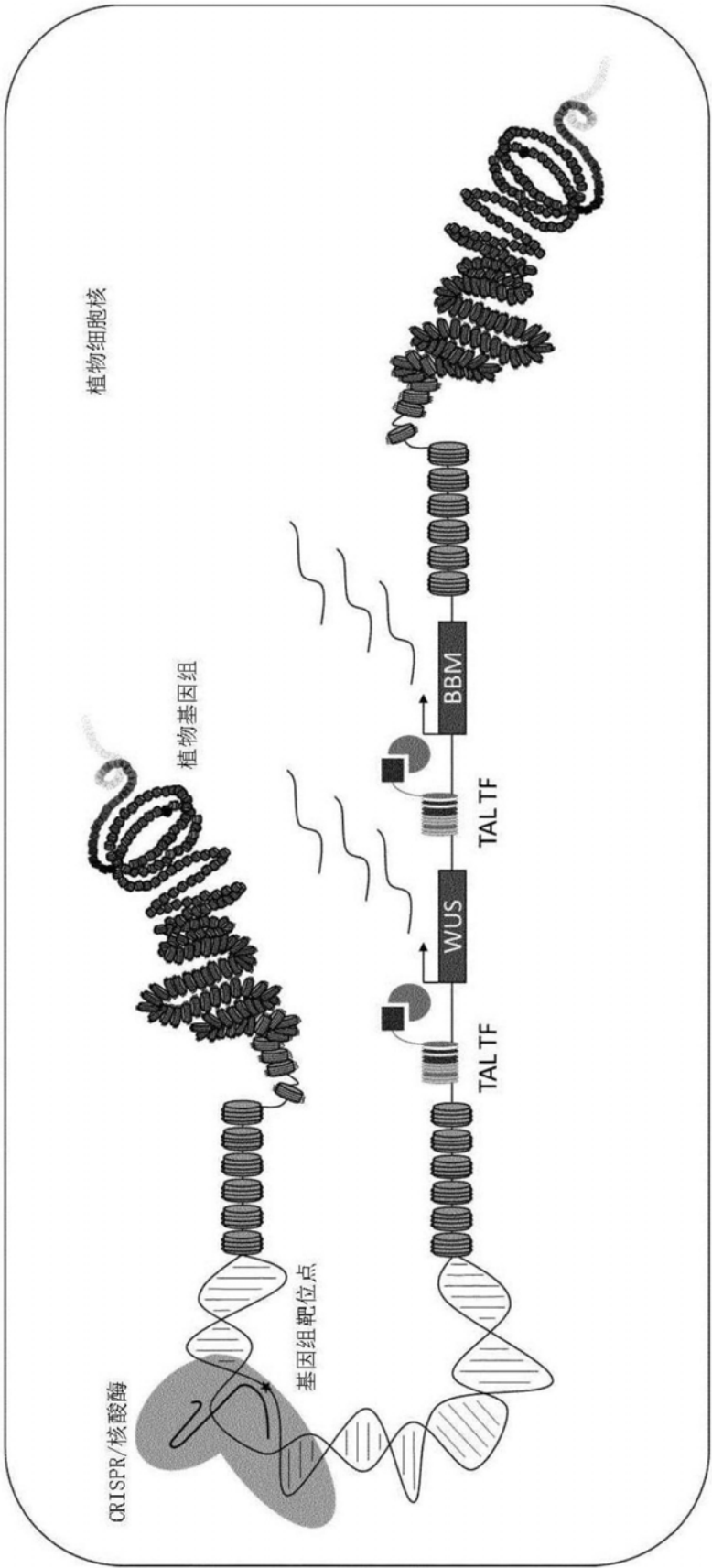


图2

A

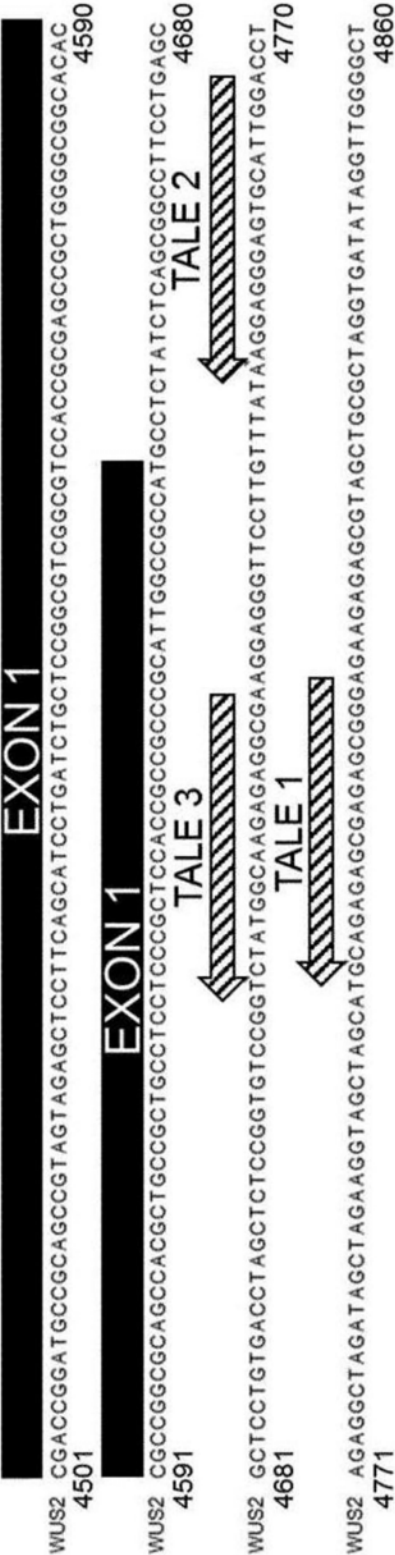


图3 (1/2)

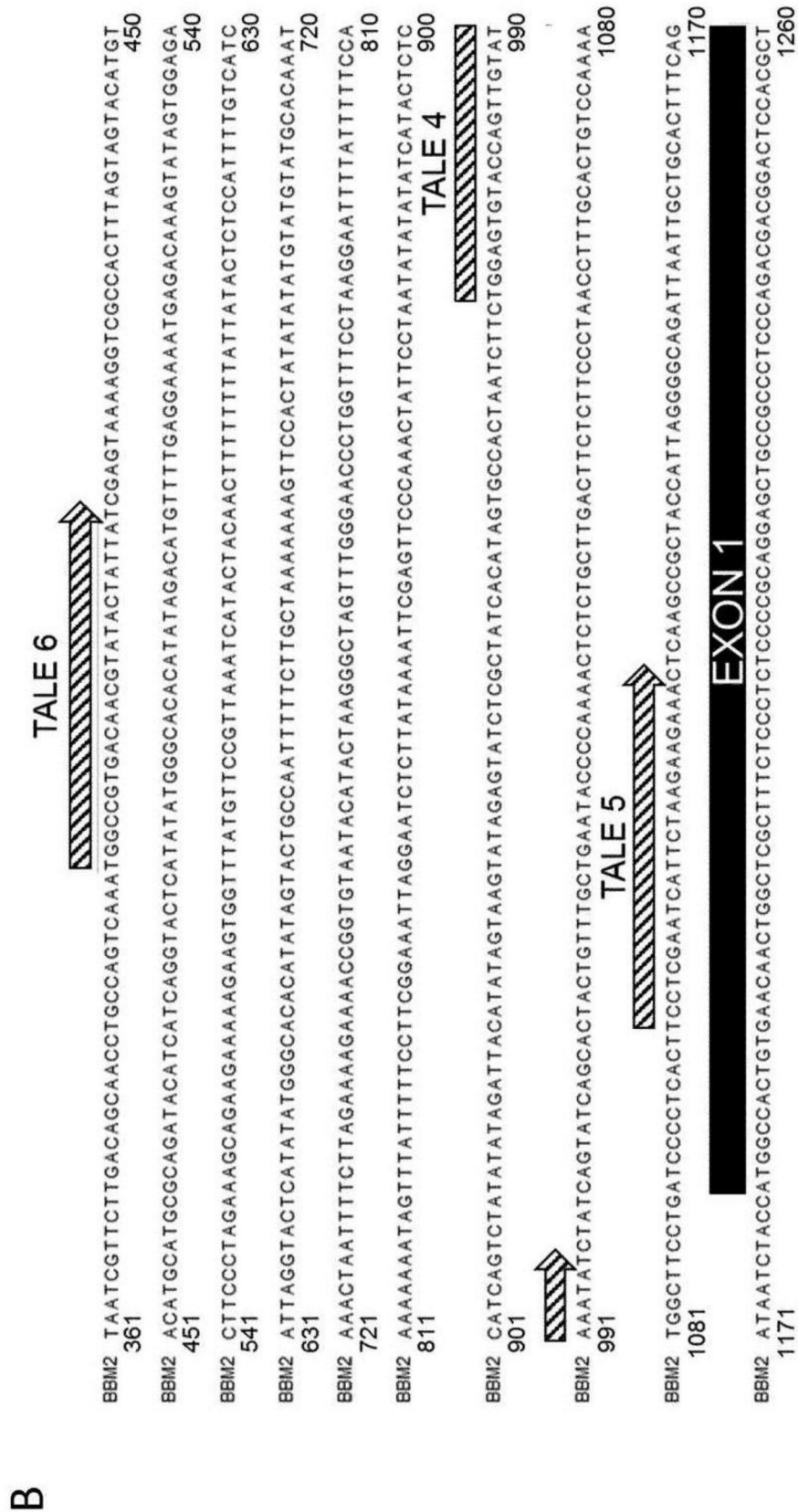


图3 (2/2)

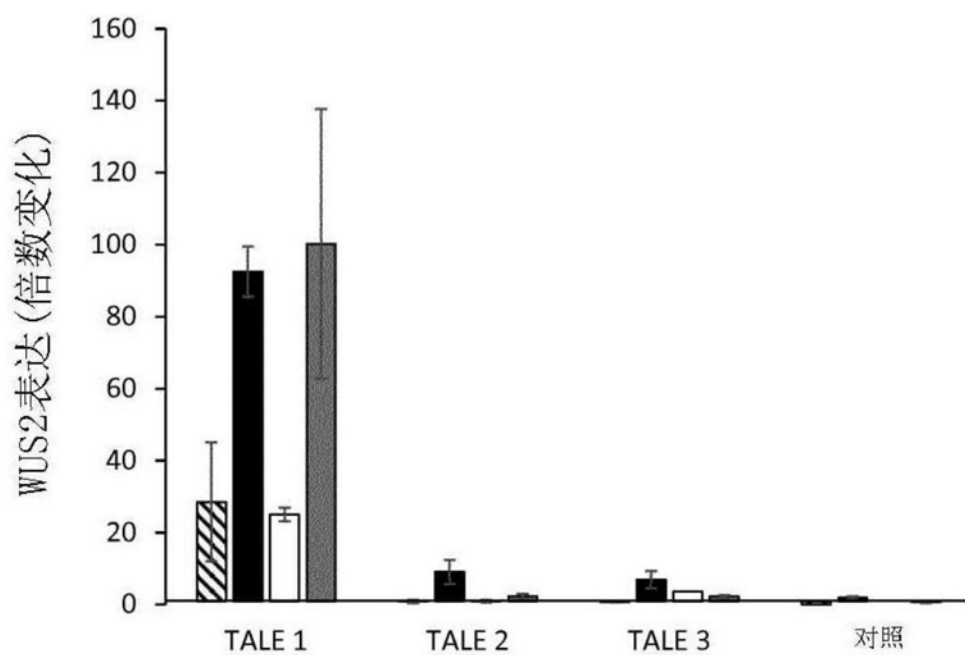
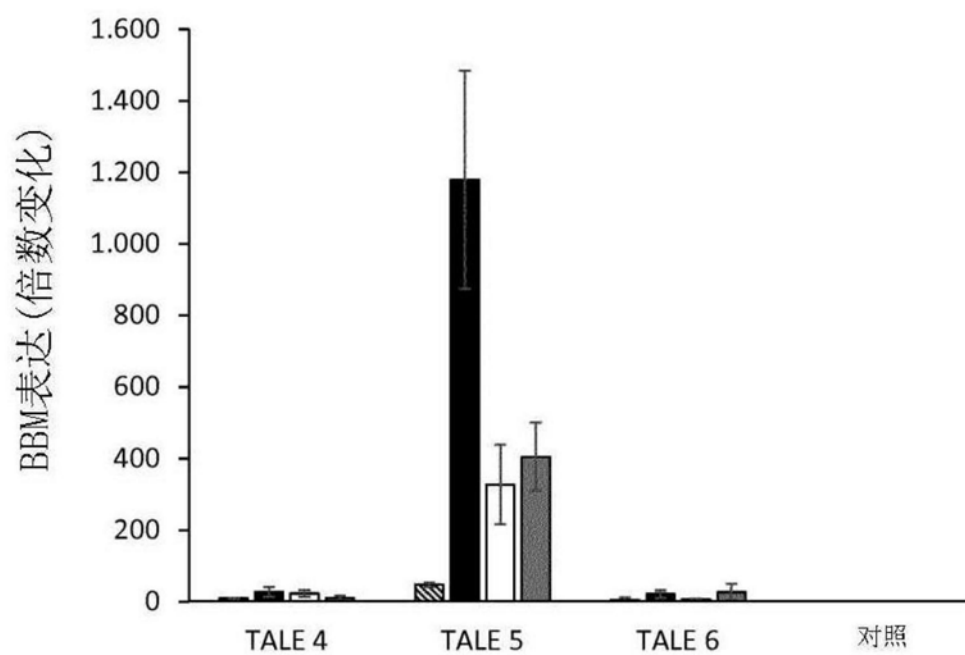
A**B**

图4

tDT + TALEN + PLT7

对照

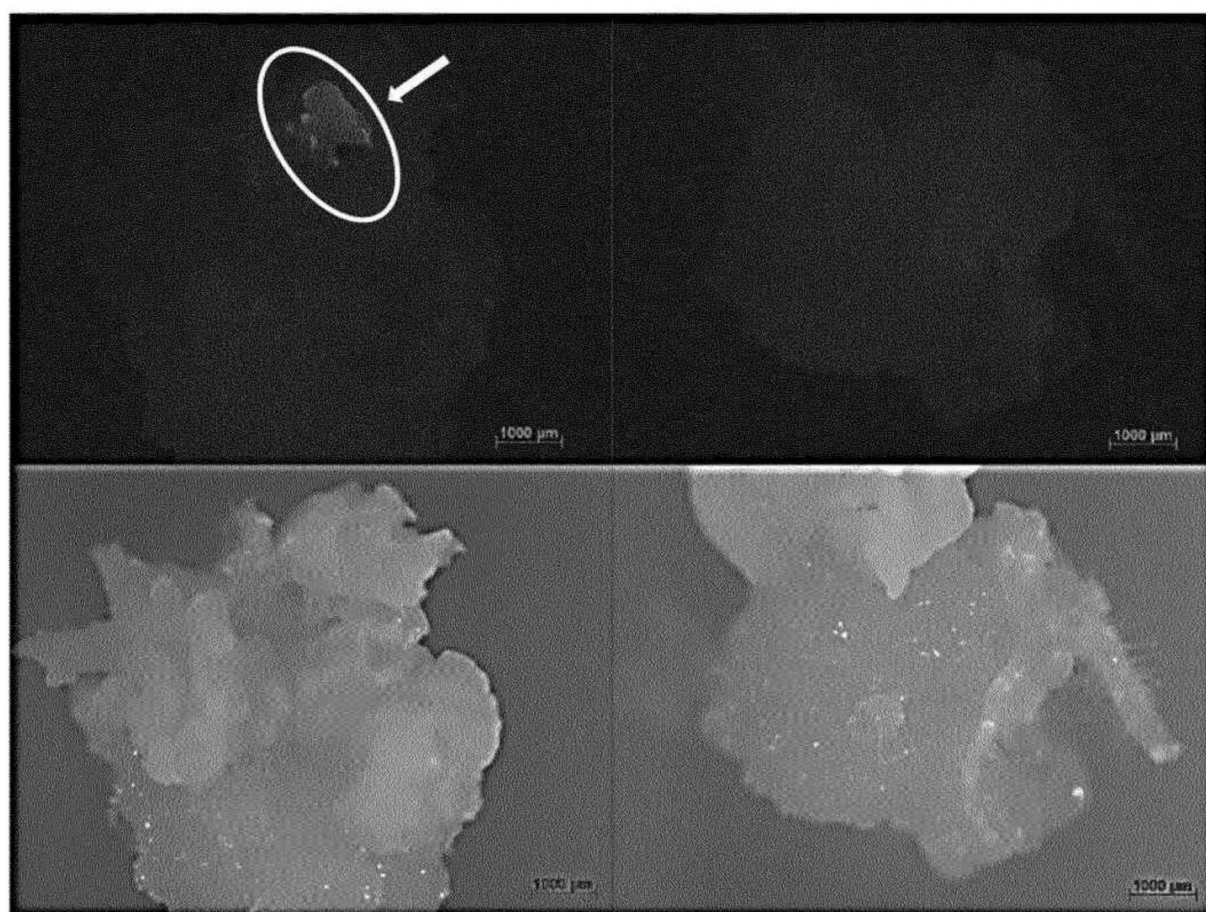


图5

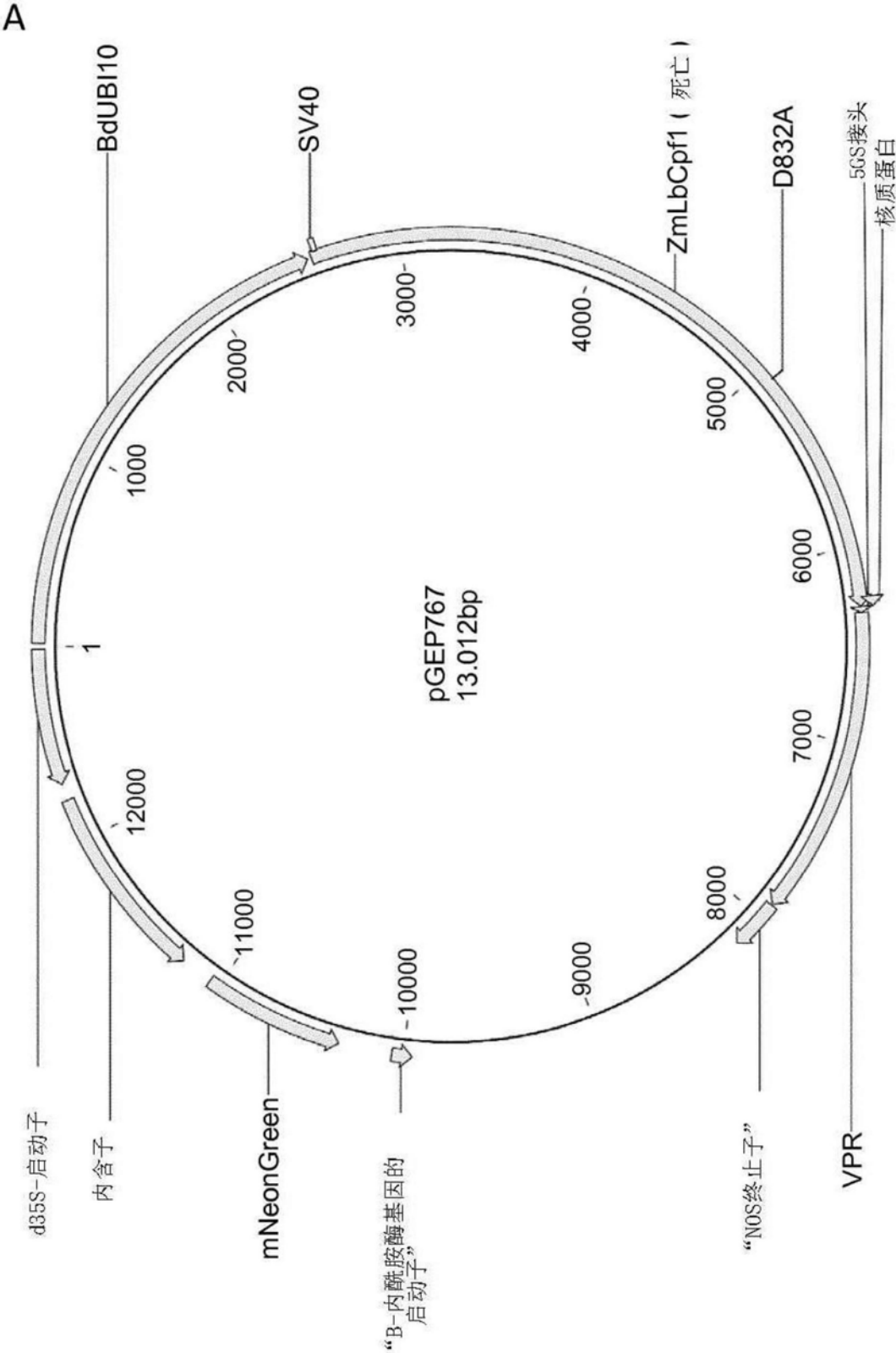


图6 (1/3)

B

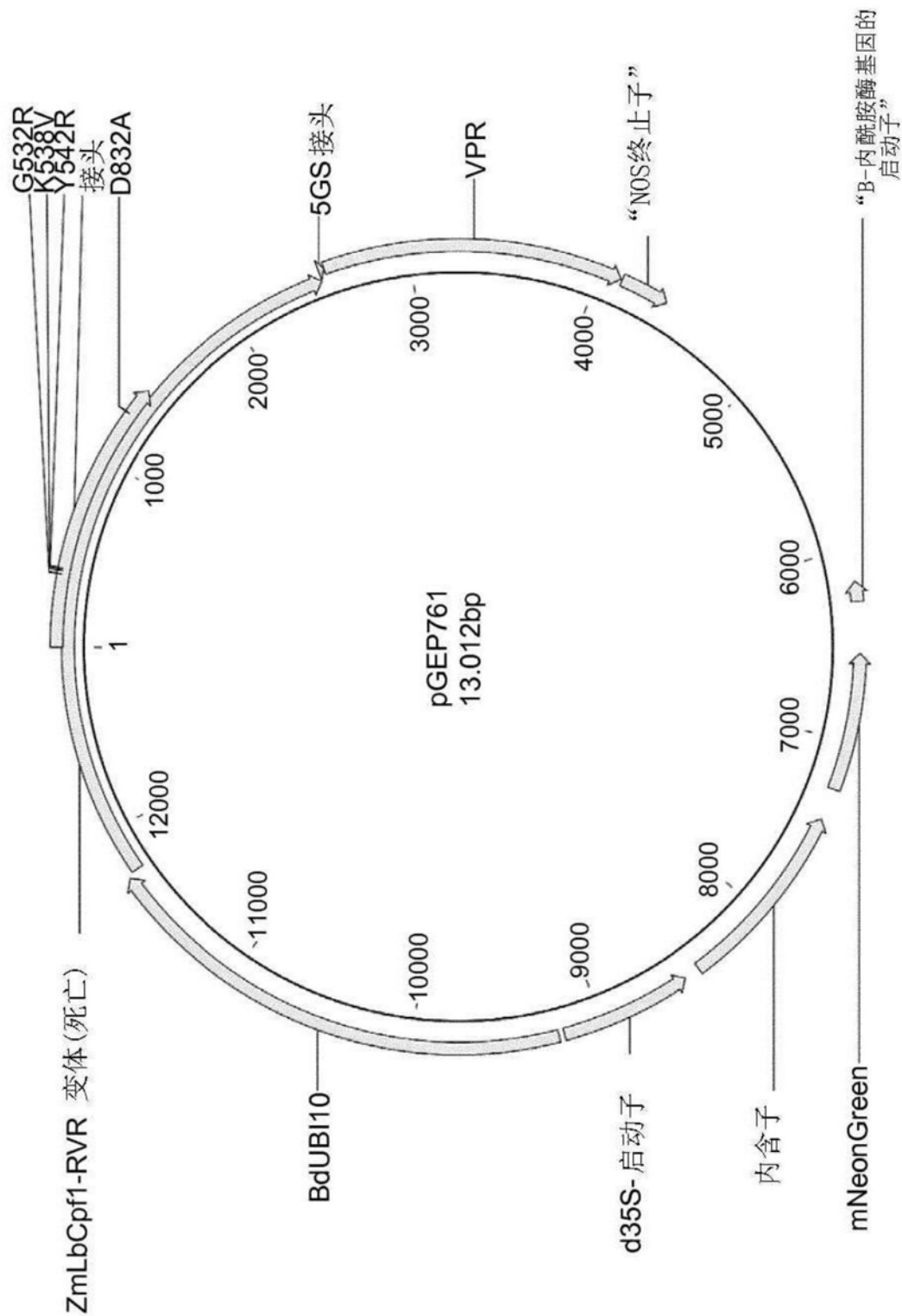


图6 (2/3)

C

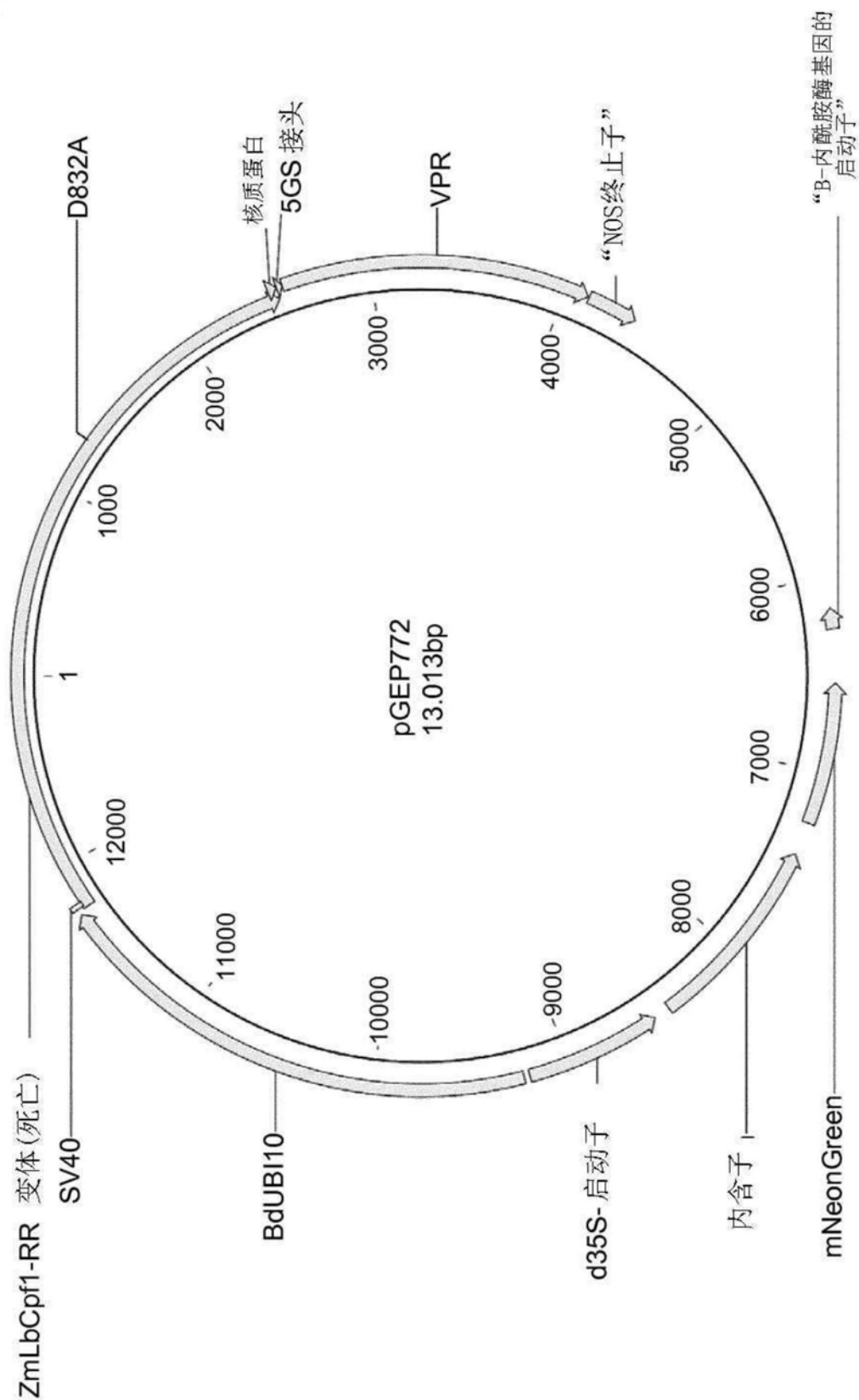


图6 (3/3)

B

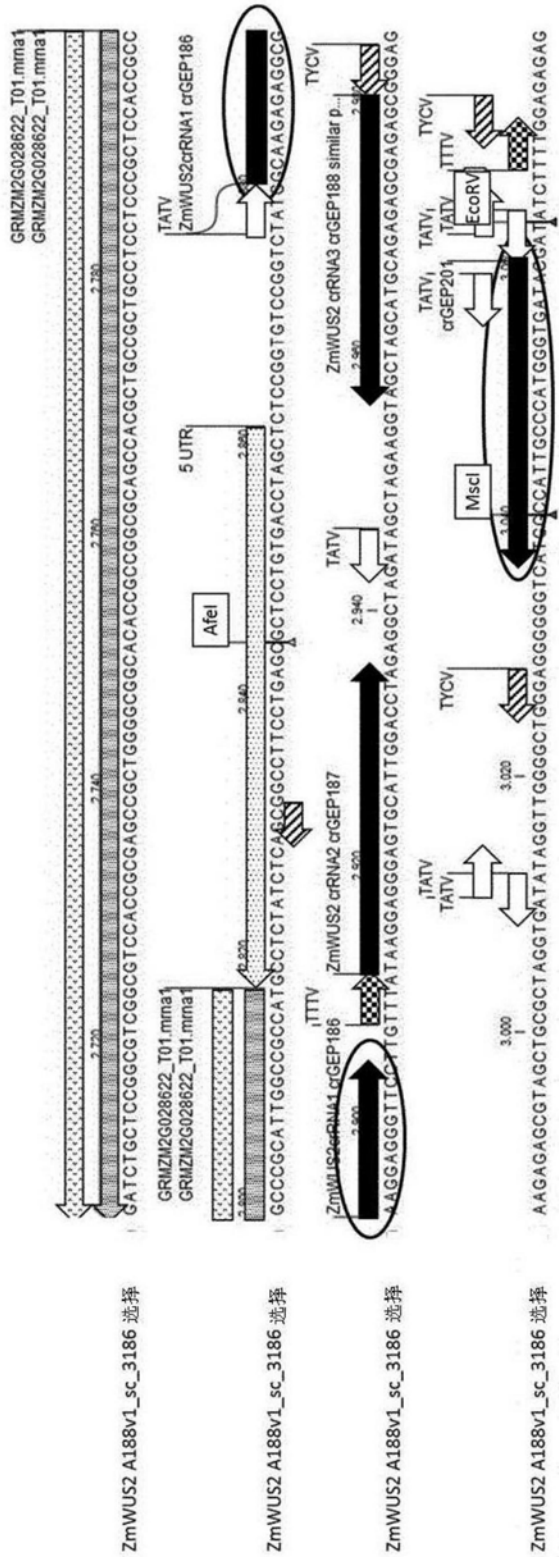


图7 (2/2)

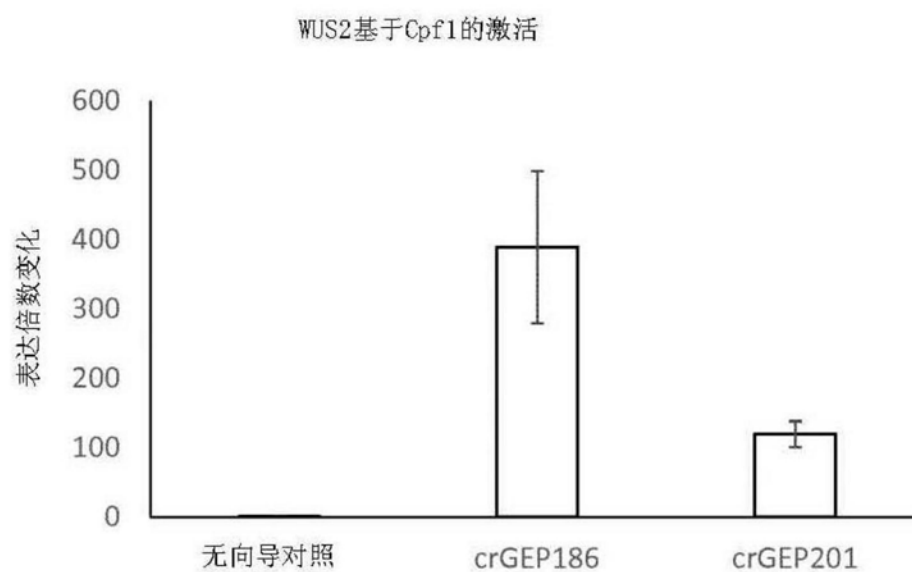
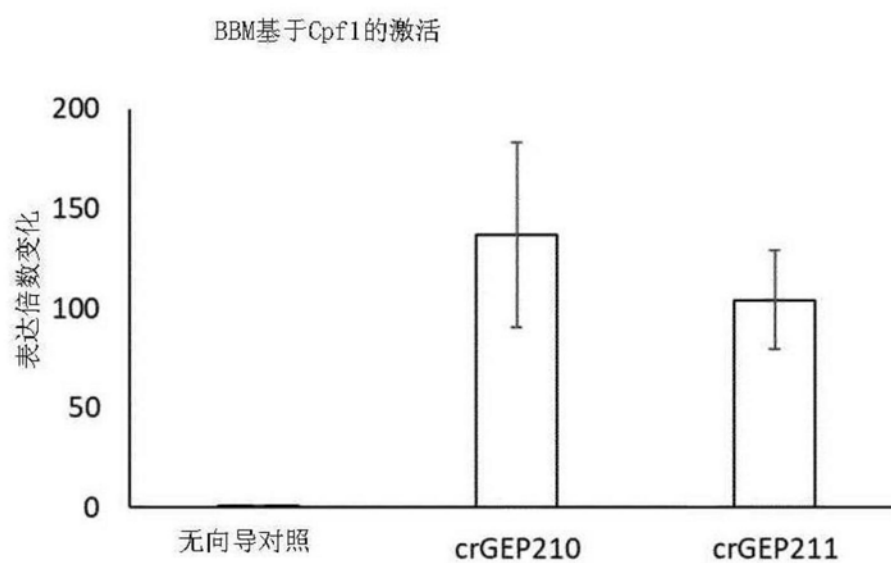
A**B**

图9

向导RNA名称	靶基因	PAM	靶序列
crGEP186	ZmWUS2	TATG	GCAAGAGAGGCGAAGGAGGGTTCC (SEQ ID NO: 291)
crGEP187	ZmWUS2	TTTA	TAAGGAGGGAGTGCATTGGACCTA (SEQ ID NO: 292)
crGEP188	ZmWUS2	TCCC	GCTCTCGCTCTCTGCATGCTAGCT (SEQ ID NO: 293)
crGEP201	ZmWUS2	TATC	GTATCACCCATGGGCAATGGCCAT (SEQ ID NO: 294)
crGEP208	ZmBBM	TCCC	CTCACTTCCTCGAATCATCTTAAG (SEQ ID NO: 295)
crGEP209	ZmBBM	TTTG	CTGAATACCCCAAAACTCTCTGCT (SEQ ID NO: 296)
crGEP210	ZmBBM	TATG	TGATAGCGAGATACTCTATACTTA (SEQ ID NO: 297)
crGEP211	ZmBBM	TATA	GTAAGTATAGAGTATCTCGCTATC (SEQ ID NO: 298)

图10