

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公开说明书

[21] 申请号 200480022218.1

[51] Int. Cl.

C12N 15/09 (2006.01)
A01H 1/00 (2006.01)
A01H 5/00 (2006.01)
C12N 5/04 (2006.01)

[43] 公开日 2006年10月4日

[11] 公开号 CN 1842592A

[22] 申请日 2004.7.30

[21] 申请号 200480022218.1

[30] 优先权

[32] 2003.7.31 [33] US [31] 60/491,837

[86] 国际申请 PCT/JP2004/011307 2004.7.30

[87] 国际公布 WO2005/012520 日 2005.2.10

[85] 进入国家阶段日期 2006.2.5

[71] 申请人 本田技研工业株式会社

地址 日本东京都

共同申请人 国立大学法人名古屋大学

[72] 发明人 西村明日香 松冈信 芦苺基行

[74] 专利代理机构 永新专利商标代理有限公司
代理人 林晓红

权利要求书 2 页 说明书 19 页 序列表 31 页
附图 8 页

[54] 发明名称

赋予植物再分化能力的基因及其应用

[57] 摘要

本发明应用连锁分析成功地分离并鉴定到与植物的再生能力有关的基因。此外，本发明还公开了用于培育高再生性品种的方法、用于转化不可栽培品种的方法和用于选择转化细胞的方法，这些方法都应用了这些基因。本发明在例如栽培品种改良和使用转化方法的基因分析等领域中是有用的。

1. 一种与植物再生能力有关的 DNA，所述 DNA 是以下(a)-(d)中的任一种：
 - (a) 编码包含 SEQ ID NO: 3 的氨基酸序列的蛋白质的 DNA；
 - (b) 包含 SEQ ID NO: 1 或 2 的核苷酸序列的编码区的 DNA；
 - (c) 编码包含在 SEQ ID NO: 3 的氨基酸序列中具有一个或多个氨基酸取代、缺失、添加和/或插入的氨基酸序列的蛋白质的 DNA；和
 - (d) 在严格条件下与包含 SEQ ID NO: 1 或 2 的核苷酸序列的 DNA 杂交的 DNA。
2. 编码包含 SEQ ID NO: 3 的氨基酸序列的蛋白质的部分肽的 DNA。
3. 包含 SEQ ID NO: 1 或 2 的核苷酸序列的启动子区域的 DNA。
4. 包含权利要求 1 或 2 的 DNA 的载体。
5. 包含权利要求 3 的 DNA 的载体。
6. 携带权利要求 4 的载体的宿主细胞。
7. 携带权利要求 4 的载体的植物细胞。
8. 包含权利要求 7 的植物细胞的植物转化体。
9. 一种植物转化体，其是权利要求 8 的植物转化体的后代或克隆。
10. 权利要求 8 或 9 的植物转化体的繁殖材料。
11. 一种用于制备植物转化体的方法，该方法包括将权利要求 1 或 2 的 DNA 导入植物细胞和从所述植物细胞再生植物的步骤。
12. 由权利要求 1 或 2 的 DNA 所编码的蛋白质。
13. 一种用于制备权利要求 12 的蛋白质的方法，该方法包括培养权利要求 6 的宿主细胞并从所述细胞或其培养物上清中收集重组蛋白质的步骤。
14. 与权利要求 12 的蛋白质结合的抗体。
15. 包含与 SEQ ID NO: 1 或 2 的核苷酸序列互补的至少 15 个连续核苷酸的多核苷酸，或与其互补的序列。

16. 一种用于提高植物的再生能力的方法，该方法包括在植物的细胞中表达权利要求 1 或 2 的 DNA 的步骤。

17. 一种用于改变植物的再生能力的制剂，该制剂包含权利要求 1 或 2 的 DNA 或权利要求 4 的载体作为活性成分。

18. 一种用于测定植物细胞的再生能力的方法，该方法包括检测权利要求 1 的 DNA 或权利要求 12 的蛋白质在植物细胞中的表达的步骤。

19. 一种用于测定植物细胞的再生能力的方法，该方法包括检测权利要求 12 的蛋白质在植物细胞中的活性的步骤。

20. 一种用于提高植物的再生能力的方法，该方法包括调节该植物中的内源性的权利要求 12 的蛋白质的活性的步骤。

21. 一种用于选择转化的植物细胞的方法，该方法包括以下步骤：

(a) 将包含用作选择标记的权利要求 1 或 2 的 DNA 的载体导入植物细胞中；和

(b) 培养所述植物细胞并选择已经获得再生能力的植物细胞。

22. 一种用于改变植物的再生能力的方法，该方法包括通过杂交取代植物中的内源性的权利要求 1 或 2 的 DNA 的步骤。

赋予植物再分化能力的基因及其应用

技术领域

本发明涉及赋予植物再生能力的基因的分离和鉴定，以及应用这些基因提高再生能力和选择转化细胞的方法。本发明可使得植物的培养特性得到改良并可基于安全性方面的特别考虑而开发转化方法。

背景技术

在适当的条件下，分化的植物组织去分化并在经过细胞分裂后形成愈伤组织（去分化细胞群）。依赖于所述的条件，愈伤组织可以进一步再分化以再生为完整的植物体。这种分化细胞或去分化细胞再生为个体的能力被称为分化全能性，在 1930-1950 年代对烟草、番茄等的培养研究中最初证实了这一点。组织培养技术是基于这一分化全能性，且已被广泛应用，尤其是在植物育种方面。例如，组织培养技术已经被用于通过细胞融合和胚珠培养生产新品种、缩短育种年限和固定遗传特性。近年来，组织培养技术作为以基因功能分析为目的的人工基因转移（转化方法）中的关键技术已成为分子育种和植物的基础研究所必需的。

分化全能性通常被认为是所有植物都拥有的能力。事实上，已知对于某些植物来说展示这一能力是容易的，对于其它植物来说则是困难的，这取决于植物种类、品种或器官。与双子叶植物相比，单子叶植物包括主要农作物例如水稻、小麦和玉米的组织培养和再生是困难的，因此对于培养包括转化方法的分析而言，再三反复实验是必需的。在水稻中，已经利用特定品种的成熟种子确立了一个相对简单的培养系统，但是具有足够的再生能力的品种是有限的。特别地，可食品种例如 Koshihikari 和 Sasanishiki，以及在热带地区广泛栽培的 IR 品系品种的再生能力低且通过组织培养再生为植物体困难。提高这些品种

的再生能力将不仅对于选择育种和基因特性的研究有用，而且还有可能阐明再生过程的机制。此外，其它不可培养的植物种和品种的再生能力也可能得到提高。

此外，近年来已开发了大量经遗传修饰的农产品（GMO），且它们的栽培面积正逐年增加。同时，许多消费者担心这些农产品的安全性。在 GMO 的安全性问题的讨论中最主要的问题是它们掺入了抗生素抗性基因。因此，开发不使用抗生素抗性基因的转化方法将减轻目前消费者对 GMO 的担忧，同时作为无需昂贵的抗生素的简单转化方法对于研究人员可能也是有利的。

发明内容

再生能力取决于作为数量性状（QTL：数量性状基因座）的众多基因间的相互作用，但是迄今为止未有从那样的基因的基因座成功分离出具有再生能力的基因的报道。本发明的一个目的是分离和鉴定与植物的再生能力有关的基因、提供应用这些基因改良植物的方法以及应用这些基因作为选择标记的转化方法。

在繁殖用于检测具有再生能力的 QTL 的杂种群体之前，本发明人选择了将用做该杂种群体的亲本的品种。本发明人选择了在再生能力上有明显差异的两个品种：日本水稻（japonica rice）“Koshihikari”和印度水稻（indica rice）“Kasalath”（图 1）。将这两个栽培种杂交产生 F1 个体，然后用 Koshihikari 作为回归亲本对它们进行回交，并进行自体受精。生成了 BCF1 群体的 99 个品系，并收集 BC1F2 种子。用每一品系的 20 粒 BC1F2 种子在诱导培养基中培养愈伤组织 30 天后，将生长的愈伤组织转移到再生培养基中并再培养 30 天。30 天之后，测量每粒种子的愈伤组织重量和苗数，用每一品系的 20 粒种子测定平均值。该平均值被当作再生能力（图 1）。采用 262 个 PCR 标记得出每一品系的基因型。当基于这些数据进行关于再生能力的 QTL 分析时，检出了 4 个具有再生能力增加效应的 QTL（图 2）。在其中一个 QTL 中，靠近 1 号染色体的短臂上的 TGS2451 标记（*PSRI*；

苗再生启动子 1) 处, 成功地发现 *Kasalath* 基因组在 *Koshihikari* 的再生能力上具有很大的增加效应 (图 2)。其次, 为了鉴定出 *PSR1* 基因的大致基因座, 从 BC2F1 群体中选择出其 *PSR1* 区域已被 *Kasalath* 的 *PSR1* 区域取代的 30 株个体, 用来自这些个体的每一株的 10 粒种子 (BC2F2 种子) 诱导产生愈伤组织。从生长的愈伤组织提取 DNA 以利用分子标记测定基因型, 通过研究再生能力进行连锁分析。此外, 为了对该基因座进行详细说明, 采用分子标记对大约 3,800 粒其中 *PSR1* 被隔离的 BC3F2 种子进行基因型调查, 并进行高分辨连锁分析。结果, 发现 *PSR1* 定位于分子标记 3132 和 P182 间的大约 50.8 kb 的区域 (图 3)。在该区域的基因预测表明在该区域存在 4 个基因, 包括一假定的蛋白质。为了测定这些基因中的哪一个是具有再生能力的基因, 构建了 *Kasalath* BAC 文库 (平均长度 120 kb), 通过 PCR 筛选分离出了一个包含 *PSR1* 区域 (BHAL15) 的 BAC 克隆。将位于 BHAL15 克隆中的合适的限制性酶切位点用以制备包含每一候选基因区域的 *Kasalath* 基因组片段, 并将它们导入到 *Koshihikari* 中。结果发现仅当导入了包含被认为编码铁氧还蛋白亚硝酸还原酶 (*NiR*) 的基因的 *Kasalath* 基因组片段 (图 3 中的 3F) 时, *Koshihikari* 的再生能力才提高 (图 4)。铁氧还蛋白亚硝酸还原酶是一种亚硝酸还原酶, 其利用铁氧还蛋白作为电子供体发挥功能, 并具备将亚硝酸盐转换为氨的作用。测定了被认为是铁氧还蛋白亚硝酸还原酶基因的遗传区域及其大约 2 kb 的上游区的核苷酸序列并与 *Kasalath* 和 *Koshihikari* 进行比较, 在该核苷酸序列内发现了许多的突变 (图 5) 此外, 当通过半定量 RT-PCR 和实时定量 PCR 对位于愈伤组织中的这一基因的 mRNA 表达水平进行测定时, *Kasalath* 中的 mRNA 的量大约是 *Koshihikari* 中的该量的 2.5 倍 (图 6 中左侧照片的顶部和中部和右侧照片)。使用特异于 *NiR* 蛋白的抗体进行的 Western 印迹分析也表明该 *NiR* 蛋白在 *Kasalath* 中的储存量比在 *Koshihikari* 中的储存量更大 (图 6 中左侧照片的底部)。此外, 使用萘乙二胺法和在 *E. coli* 中表达的 *NiR* 重组蛋白比较每单位蛋白的 *NiR* 酶活性, *Kasalath* 的 *NiR*

酶活性显示出高于 Koshihikari 的 NiR 酶活性大约 1.6 倍（图 7）。上述结果表明 Koshihikari 和 Kasalath 之间在再生能力上的差异首先归因于两者在 *NiR* 基因的转录调节水平上的差异，其次归因于每分子该合成蛋白在活性上的差异。

将 Kasalath *PSRI* 基因的基因组区域导入中赋予了不能再生的 Koshihikari 以再生能力。这表明当转化 Koshihikari 时 Kasalath *PSRI* 基因可以被用做一个选择标记。更特别地，当将已同时插入了 Kasalath *PSRI* 基因和一靶基因的载体导入到 Koshihikari 中时，仅已被导入了 *PSRI* 基因的那些细胞将获得再生能力，因此再生的植物体将同时具有被掺入的靶基因。为了证明这一观点，构建了携带有于 pBI101 二元载体的 T-DNA 区域中的 Kasalath *NiR* 基因组+35S 启动子 *GUS*、Kasalath *NiR* 启动子:: *NiR* cDNA :: *NiR* 终止子+35S 启动子 *GUS*、水稻 *Actin1* 启动子:: *NiR* cDNA :: *NiR* 终止子+35S 启动子 *GUS* 的载体和未携带 *NiR* 基因的载体，并将其导入到 Koshihikari 中。当导入 3 种类型的包含 *NiR* 基因的载体时，在所有情况下都获得了大量的再生个体，且在这些再生个体自其衍生而来的愈伤组织中观察到由于 *GUS* 基因所产生的着色（图 8）。此外，*NiR* 基因具有代谢对植物有毒性的亚硝酸的特性，应用这一特性也使得 *NiR* 基因能被用作高再生性品种的转化标记。更特别地，位于肌动蛋白启动子（该启动子是水稻中的一个高表达启动子）控制下的过表达 *NiR* 基因的载体被导入到一个高再生性的 Kasalath 品种中，并在添加有亚硝酸的培养基中进行培养，所添加的亚硝酸的浓度能抑制普通野生型的生长。只有被转化的细胞由于过表达的 *NiR* 基因的效应而生长，且 *GUS* 染色仅在这些生长的细胞中被观测到（图 9）。这一选择方法的使用使得生产不使用衍生自微生物的抗生素抗性基因（转化细胞的选择标记）的更安全的重组植物成为可能，这被认为是传统的遗传修饰的农产品中的问题所在。此外，因为昂贵的抗生素是不必要的，开发转化体的费用被降低。

更特别地，本发明涉及提高植物再生能力的基因的分离和鉴定和通过应用这些基因提高植物的培养特性，以及使用这些基因作为选择

标记的转化。本发明提供了下述[1]-[22]个方面：

[1] 涉及植物的再生能力的 DNA，其中所述 DNA 是(a) - (d)中的任一个：

(a) 编码一种包含 SEQ ID NO: 3 的氨基酸序列的蛋白质的 DNA；

(b) 包含 SEQ ID NO: 1 或 2 的核苷酸序列的编码区的 DNA；

(c) 编码包含 SEQ ID NO: 3 的氨基酸序列中的一个或多个氨基酸取代、缺失、添加和/或插入的氨基酸序列的蛋白质的 DNA；和

(d) 在严格条件下与包含 SEQ ID NO: 1 或 2 的核苷酸序列的 DNA 杂交的 DNA；

[2] 编码包含 SEQ ID NO: 3 的氨基酸序列的蛋白质的部分肽的 DNA；

[3] 包含 SEQ ID NO: 1 或 2 的核苷酸序列的启动子区域的 DNA；

[4] 包含[1]或[2]的 DNA 的载体；

[5] 包含[3]的 DNA 的载体；

[6] 携带[4]的载体的宿主细胞；

[7] 携带[4]的载体的植物细胞；

[8] 包含[7]的植物细胞的植物转化体；

[9] 是[8]的植物转化体的后代或克隆的植物转化体；

[10] [8]或[9]的植物转化体的繁殖材料；

[11] 一种用于制备植物转化体的方法，该方法包括将[1]或[2]的 DNA 导入植物细胞和从所述植物细胞再生植物的步骤；

[12] 由[1]或[2]的 DNA 所编码的蛋白质；

[13] 一种用于制备[12]的蛋白质的方法，该方法包括培养[6]所述的宿主细胞并从所述细胞或其培养物上清收集重组蛋白质的步骤；

[14] 与[12]的蛋白质结合的抗体；

[15] 包含至少 15 个与 SEQ ID NO: 1 或 2 的核苷酸序列互补的连续核苷酸的多核苷酸，或与其互补的序列；

[16] 一种用于提高植物的再生能力的方法，该方法包括在植物

的细胞中表达 [1]或[2]的 DNA 的步骤;

[17] 一种用于改变植物的再生能力的制剂, 该制剂包含[1]或[2]的 DNA 或[4]的载体做为活性成分;

[18] 一种用于测定植物细胞的再生能力的方法, 该方法包括检测[1] 的 DNA 或[12]的蛋白质在植物细胞中的表达的步骤;

[19] 一种用于测定植物细胞的再生能力的方法, 该方法包括检测[12]的蛋白质在植物细胞中的活性的步骤;

[20] 一种用于提高植物的再生能力的方法, 该方法包括调节[12]的内源蛋白在该植物中的活性的步骤;

[21] 一种用于选择转化的植物细胞的方法, 该方法包括以下步骤:

(a) 将包含用做选择标记的[1]或[2]的 DNA 的载体导入植物细胞中; 和 (b) 培养所述植物细胞并选择已经获得再生能力的植物细胞; 和

[22] 一种用于改变植物的再生能力的方法, 该方法包括通过杂交取代植物中的[1]或[2]的内源 DNA 的步骤。

本发明提供了编码水稻来源的 NiR 蛋白的 DNA。“Kasalath”的基因组 DNA 的核苷酸序列示于 SEQ ID NO: 1 中, “Kasalath”的 cDNA 的核苷酸序列示于 SEQ ID NO: 2 中, 由该 DNA 所编码的蛋白质的氨基酸序列示与 SEQ ID NO: 3 中。“Koshihikari”的基因组 DNA 的核苷酸序列示于 SEQ ID NO: 4 中, “Koshihikari”的 cDNA 的核苷酸序列示于 SEQ ID NO: 5 中, 由该 DNA 所编码的蛋白质的氨基酸序列示与 SEQ ID NO: 6 中。

本发明表明了植物的再生能力可以通过调节 *PSRI* 基因在植物中的表达或活性而提高。这使得不可培养品种例如 Koshihikari 的培养成为可能, 并使得生产稳定且高再生性的品种成为可能。

术语“再生能力的提高”在本发明中仅指在培养条件下植物的再生能力被提高, 且再生个体的形式未改变。这一在再生能力上的提高

使得可以将所期望的品种用于各种培养试验中，结果使得新品种和基因的功能性分析得到有效的发展。

在本发明中，术语“植物的 *PSR1* 基因”指编码植物的铁氧还蛋白亚硝酸还原酶的 *NiR* 基因。“植物的 *PSR1* 基因”包含水稻的 *PSR1* 基因（图 5）和来源于其他植物的 *PSR1* 基因。编码本发明的 *PSR1* 蛋白的 DNA 包括基因组 DNA、cDNA 和化学合成的 DNA。基因组 DNA 和 cDNA 可以根据本领域技术人员所熟知的常规方法制备。更特别的，基因组 DNA 可以例如按照如下制备：(1) 从具有 *PSR1* 基因的水稻品种（例如 Koshihikari）提取基因组 DNA；(2) 构建一个基因组文库（应用载体例如质粒、噬菌体、粘粒、BAC 和 PAC）；(3) 发展所述文库；和 (4) 用基于编码本发明的蛋白质的 DNA（例如 SEQ ID NO: 1 或 2）制备的探针进行克隆杂交或噬菌斑杂交。或者，可以通过使用特异于编码本发明的蛋白质的 DNA（例如 SEQ ID NO: 1 或 2）的引物进行 PCR 植被基因组 DNA。另一方面，可以例如按照如下制备 cDNA：(1) 基于从具有 *PSR1* 基因的水稻品种（例如 Koshihikari）提取的 mRNA 合成 cDNA；(2) 通过将该合成的 cDNA 插入到载体例如 λ ZAP 中制备一个 cDNA 文库；(3) 发展所述 cDNA 文库；和 (4) 如上述进行克隆杂交或噬菌斑杂交。或者，也可以通过 PCR 制备 cDNA。

本发明包括编码 SEQ ID NO:3 的 *PSR1* 蛋白的功能性等价物的蛋白（Kasalath）的 DNA。在本申请中，术语“*PSR1* 蛋白的功能性等价物”是指对目标蛋白的表达或活性的修饰使得再生能力提高。

这样的 DNA 的例子包括编码突变体、衍生物、等位基因、变异体和包含 SEQ ID NO: 3 中一个或多个氨基酸被取代、删除、添加和/或插入的氨基酸序列的同源物的那些 DNA。

本领域技术人员所熟知的用于制备编码包含被改变的氨基酸的 DNA 的方法的例子包括定点诱变（Kramer, W.和 Fritz, H. -J., (1987) “Oligonucleotide-directed construction of mutagenesis via gapped duplex DNA.” *Methods in Enzymology*, 154: 350-367)。蛋白质的氨基酸序列也

可以在自然界中由于核苷酸序列的突变发生突变。编码其中具有一个或多个氨基酸被取代、缺失和/或添加的天然 PSR1 蛋白的氨基酸序列的蛋白质的 DNA 也包括在本发明的 DNA 中，只要它们编码与天然 PSR1 蛋白 (SEQ ID NO: 3) 功能性等价的蛋白质。此外，不增加蛋白质的氨基酸序列的改变的核苷酸突变体 (退行性突变体) 也包括在本发明的 DNA 的范围内。

编码与 SEQ ID NO: 3 中所述的 PSR1 蛋白的功能性等价性蛋白的 DNA 可以通过例如本领域技术人员所熟知的方法来制备，包括使用杂交技术的方法 (Southern, E.M., *Journal of Molecular Biology*, Vol. 98, 503, 1975.) 和聚合酶链式反应 (PCR) 技术 (Saiki, R. K. *et al. Science*, vol.230, 1350-1354, 1985; Saiki, R. K. *et al. Science*, vol.239, 487-491, 1988)。也就是说，通过使用 PSR1 基因的核苷酸序列 (SEQ ID NO: 2) 或其一部分作为探针，并使用与 PSR1 基因的核苷酸序列 (SEQ ID NO: 2) 特异性杂交的寡核苷酸作为引物，从水稻和其他植物分离出与 PSR1 基因具有高度同源性的 DNA 对于本领域技术人员来说是很常规的。这样的可通过杂交技术或 PCR 技术获得的编码与 PSR1 蛋白功能性等价的蛋白质的 DNA 也包括在本发明的 DNA 的范围内。

优选在严格条件下进行杂交反应以分离这样的 DNA。本发明的严格杂交条件包括例如 6M 尿素、0.4% SDS 和 0.5x SSC 这样的条件，和那些产生相似的严格性的条件。期望当在具备较高的严格性的条件下，例如 6M 尿素、0.4% SDS、和 0.1x SSC 的条件下进行杂交时获得具有较高同源性的 DNA。在这种条件下分离出的那些 DNA 被期望编码与 PSR1 蛋白 (SEQ ID NO:3 或 6) 具有高水平的氨基酸同源性蛋白质。在本申请中，高同源性的意思是在整个氨基酸序列上具有至少 50% 或更高的同一性，更优选 70% 或更高，更加优选 90% 或更高 (例如 95%、96%、97%、98%、99% 或更高)。一条氨基酸序或核苷酸序列与另一条氨基酸序列或者核苷酸序列的同源性的程度可以遵循由 Karlin 和 Altschul 建立的 BLAST 算法进行测定 (Proc. Natl. Acad.

Sci. USA 87:2264-2268, 1990; Proc. Natl. Acad. Sci. USA, 90: 5873, 1993)。基于 BLAST 算法开发了例如 BLASTN 和 BLASTX 这样的程序 (Altschul SF, *et al.* J. Mol. Biol. 215: 403, 1990)。为了根据 BLASTN 对核苷酸序列进行测定, 参数被设置为例如得分=100 和字长=12。另一方面, 用于通过 BLASTX 对氨基酸序列进行分析的参数包括例如得分=50 和字长=3。当应用 BLAST 和缺口 BLAST 程序时, 使用每一程序的缺省值。用于这种分析的特定技术是本领域中已知的。

可以按照如下评估某一特定的 DNA 是否编码与植物的再生能力有关的蛋白质。最为常规的方法包括检测 DNA 的功能, 然后进行培养并对再生的能力进行调查。更特别地, 所述方法包括在维持了 DNA 的功能的条件下和在 DNA 失去功能的条件下进行培养, 并比较结果产生的再生能力。如果再生能力不发生变化或几乎相同, 则该 DNA 与再生能力无关。当 DNA 与再生能力相关时, 再生率被进一步提高, 这一差别被认为是再生能力的程度。

本发明的 DNA 可以被用于例如制备重组蛋白和生产再生能力被改变的植物转化体。通常通过以下步骤制备重组蛋白: 将编码本发明的蛋白质的 DNA 插入到一合适的表达载体中; 将该载体导入到一合适的细胞中; 培养该转化细胞; 使得所述细胞能够表达所述重组蛋白并纯化所表达的蛋白质。重组蛋白可以与其他蛋白质一起作为融合蛋白被表达以使得能够容易地被纯化, 例如作为与麦芽糖结合蛋白在大肠杆菌 (*Escherichia coli*) 中的融合蛋白 (New England Biolabs, USA, Pmal 载体系列)、作为与谷胱甘肽-S-转移酶 (GST) 的融合蛋白 (Amersham Pharmacia Biotech, pGEX 载体系列) 或用组氨酸标签 (Novagen, pET 系列)。对宿主细胞没有限制, 只要该细胞适于表达该重组蛋白即可。有可能利用酵母或各种动物、植物或昆虫细胞以及上述的大肠杆菌。可以通过多种本领域技术人员已知的方法将载体导入到宿主细胞当中。例如, 利用钙离子的转化方法可以被用于将载体导入 *E. coli* 中 (Mandel, M. 和 Higa, A. (1970) *Journal of Molecular Biology*, 53, 158-162, Hanahan, D. (1983) *Journal of Molecular Biology*,

166, 557-580)。在宿主细胞中表达的重组蛋白可以通过已知的方法进行纯化并从宿主细胞或其培养物上清进行回收。当重组蛋白作为与麦芽糖结合蛋白或其他配偶体 (partner) 的融合蛋白被表达时, 所述重组蛋白能够通过亲和层析容易地被纯化。本发明的蛋白质可以从按照下述通过将本发明的 DNA 导入到植物中所产生的转化植物来制备。因此, 如下述, 本发明的转化植物不仅包括导入了本发明的 DNA 以改变其再生能力的植物, 而且包括了导入了本发明的 DNA 以制备本发明的蛋白质的植物。

所得蛋白质可以用于制备与所述蛋白质结合的抗体。例如, 可以通过用本发明的纯化蛋白或其一部分对免疫动物例如兔进行免疫、在一定时期后采集血液并除去血液凝块来制备多克隆抗体。可以通过将骨髓瘤细胞与被上述蛋白质或其一部分免疫的动物的抗体生成细胞相融合、分离表达所需抗体的单克隆细胞 (杂交瘤) 并从所述细胞除去回收所述抗体来制备单克隆抗体。所获得的抗体可以被用于纯化或检测本发明的蛋白质。因此, 本发明包括了与本发明的蛋白质相结合的抗体。这些抗体的使用使得可以将与植物体的再生能力相关的蛋白质的表达位点区分出来, 或能够测定某一植物种是否表达与再生能力相关的蛋白质。

当应用本发明的 DNA 来制备再生能力被提高的转化植物时, 将编码本发明的蛋白质的 DNA 被插入到一合适的载体中, 随后将该载体导入到植物细胞中。通过这些步骤所获得的转化的植物细胞随后被再生。导入了所述载体的植物细胞优选具有本发明的 DNA 的低表达的植物细胞。因此, 术语“植物细胞”包括各种形式的植物细胞, 例如悬浮培养细胞、原生质体、叶的切片 (leaf sections) 和愈伤组织。

对用于植物细胞转化的载体没有特别的限制, 只要它们能在细胞中表达被插入的基因即可。实例包括“pBI121”、“pBI221”和“pBI101”质粒 (全部来自 Clontech)。

本发明的载体可以包含一个用于组成型或诱导型表达本发明的蛋白质的启动子。用于组成型表达的启动子的例子包括花椰菜花叶病

毒的 35S 启动子 (Odell *et al.* 1985 Nature 313:810), 水稻的肌动蛋白启动子 (Zhang *et al.* 1991 Plant Cell 3:1155) 和玉米的遍在蛋白启动子 (Cornejo *et al.* 1993 Plant Mol. Biol. 23:567)。

用于诱导型表达的启动子的实例包括已知的由于外因启动表达的启动子, 例如丝状真菌、细菌和病毒的感染和入侵、低温、高温、干燥、紫外线照射和特殊化合物的喷射。这样的启动子的实例包括通过丝状真菌、细菌和病毒的感染和入侵导入的几丁质基因启动子 (Xu *et al.* 1996 Plant Mol. Biol. 30:387) 和烟草 PR 蛋白基因启动子 (Ohshima *et al.* 1990 Plant Cell 2:95)、低温所诱导的水稻“lip19”基因启动子 (Aguan *et al.* 1993 Mol. Gen Genet. 240:1)、高温所诱导的水稻“hsp 80”基因和“hsp 72”基因启动子 (Van Breusegem *et al.* 1994 Planta 193:57)、干燥所诱导的拟南芥“rab 16”基因启动子 (Nundy *et al.*, 1990 Proc. Natl. Acad. Sci. USA 87:1406)、紫外线照射所诱导的查耳酮合酶基因启动子 (Schulze-Lefert *et al.* 1989 EMBO J. 8:651) 和厌氧条件所诱导的玉米乙醇脱氢酶基因启动子 (Walker *et al.*, 1987 Proc. Natl. Acad. Sci. USA 84:6624)。此外, 水稻的几丁质基因启动子和烟草的 PR 蛋白基因启动子也能被特定的化合物例如水杨酸所诱导, “rab 16”也能通过喷射脱落酸, 一种植物激素被诱导。

此外, 所述载体可以包含编码本发明的蛋白质的 DNA 的启动子。编码本发明的蛋白质的 DNA 的启动子区域可以通过例如用包含 SEQ ID NO: 1 或 2 的核苷酸序列或其一部分的 DNA 作为探针对基因组文库进行筛选获得。

此外, 本发明提供了导入了本发明的载体的转化细胞。除了上述用于制备重组蛋白的细胞外, 导入了本发明的载体的细胞包括用于制备转化植物的植物细胞。对于所述植物细胞的类型没有特殊的限制, 其实例有拟南芥、水稻、玉米、马铃薯和烟草的细胞。除培养的细胞外, 本发明的植物细胞包括植物中的细胞以及原生质体、shoot primordia、multiple shoots 和发根 (hairy root) 中的细胞。可以通过已知的方法将载体导入植物细胞中, 例如聚乙烯乙二醇方法、电穿孔、

农杆菌介导的转移和微粒轰击。根据植物细胞的类型，可以通过已知的方法从转化的植物细胞再生植物（Toki *et al.*, (1995) *Plant Physiol.* 100:1503-1507）。例如，用于水稻类植物的转化和再生方法包括：(1) 用聚乙烯乙二醇将基因导入原生质体中并再生植物体（适于 *indica* 水稻品种）（Datta, S.K. (1995) in “Gene Transfer To Plants”, Potrykus I and Spangenberg Eds., pp66-74); (2) 用电脉冲将基因导入原生质体中并再生植物体(适于 *japonica* 水稻品种)(Toki *et al.* (1992) *Plant Physiol.* 100, 1503-1507); (3) 通过微粒轰击直接将基因导入细胞中并再生植物体（Christou *et al.* (1991) *Bio/Technology*, 9: 957-962); 和(4) 用农杆菌将基因导入并再生植物体（Hiei *et al.* (1994) *Plant J.* 6: 271-282)。这些方法是现有技术中已经建立起来的方法并被广泛地应用于本发明所属的技术领域中。这样的方法可以被适当地用于本发明。

已经获得了在其基因组中含有本发明的 DNA 的转化植物，因而有可能通过有性繁殖或无性繁殖获得该植物的后代。也有可能从所述植物或其后代或克隆获得繁殖材料（例如种子、果实、穗、块茎、块茎状的根、残株、愈伤组织、和原生质体）以基于这样的材料大量生产所述植物。因此，本发明包括导入了本发明的 DNA 的植物细胞、含有这些细胞的植物、这些植物的后代或克隆以及该植物及其后代和克隆的繁殖材料。

以这种方式制备的其再生能力被修饰的植物相对于野生型植物其再生能力和产量发生变化。例如，其中导入了位于水稻肌动蛋白启动子控制下的编码 PSR1 蛋白的 DNA 的植物被期望在其再生能力上有所提高。本发明的方法的使用可以提高水稻的再生能力，这对于农作物而言是有益的。本发明还有益于高再生型水稻品种的开发。

此外，本发明提供了包含与 SEQ ID NO: 1 或 2 的核苷酸序列互补的至少 15 个连续核苷酸的多核苷酸或其互补序列。因此，术语“互补序列”是包含 A:T 和 G:C 碱基对的双链 DNA 的一条链的序列相对于另一条链的序列而言。术语“互补”不限制于某一序列与至少 15 个连续的核苷酸完全互补的情形，还包括其中核苷酸序列同一性为至少

70%、优选至少 80%、更优选 90%、再更优选 95%或更高（例如 96%或更高、97%或更高、98%或更高、或 99%或更高）。这些 DNA 作为探针有益于检测或分离本发明的 DNA, 和用做引物扩增所述 DNA。

本发明还提供了用于测在植物中是否存在再生能力的遗传诊断方法。在本发明中, “测定在植物中是否存在再生能力”不仅对于测定迄今为止所栽培的品中是否存在再生能力是有效的, 而且还包括了测定在通过杂交和遗传工程技术所制备的新品种中是否存在再生能力。这些方法对于测定 japonica 水稻品种中是否存在再生能力由其有效。

本发明的用于评估在植物中是否存在再生能力的方法包括检测编码 PSR1 蛋白的 DNA 和 PSR1 蛋白的植物表达水平。例如, 如果编码 PSR1 的 DNA 或 PSR1 蛋白的表达水平高于在 Koshihikari 中的表达水平, 则所检测的植物被测定为是拥有再生能力的一个品种。

本发明提供了用于在植物的转化中将 *PSR1* 基因作为选择标记的方法。先前被用作转化的植物细胞的选择标记基因的实例包括提供对抗生素潮霉素的抗性的潮霉素磷酸转移酶基因、提供对卡那霉素或庆大霉素的抗性的新霉素磷酸转移酶、提供对除草剂膦丝菌素的抗性的乙酰转移酶基因和提供对双丙氨膦 (bialaphos) 的抗性的双丙氨膦抗性基因。当使用这些基因时, 通过在一种已知的选择培养基中进行培养获得转化的植物细胞, 所述选择性培养基包含一种与选择标记基因的类型匹配的选择剂。当将 *PSR1* 基因而不是这些药物抗性基因用作选择标记时, 如果待转化的植物细胞不具备再生能力, 如在 Koshihikari 中, 可以使用所获得的再生能力作为选择性状对转化体进行选择, 而不使用用于选择的特殊制剂。也就是说, 由于非转化体不能够再生, 由于 *PSR1* 基因的效应而再生的个体被假定为转化体。此外, 当使用 *PSR1* 基因作为具有再生能力的植物细胞的选择标记时, 可以通过向选择培养基中添加一定浓度的亚硝酸盐对具有再生能力的转化细胞进行选择, 该亚硝酸盐将抑制非转化体的生长。上述被用于选择转化体的常规药物抗性基因衍生自微生物; 因此, 保留了这些抗性基因的经遗传修饰的农产品 (GMO) 已经增加了人们对于其对

生态系统和人体所产生的副作用的担忧。但是，使用本发明的 *PSRI* 基因用于选择转化体的方法在减轻此类担忧和开发廉价的遗传修饰农作物方面具有优势。

本申请中所引用的所有现有技术文献通过引用合并到本发明中。

附图说明

图 1 是表示 Koshihikari 和 Kasalath 的基因型的一组照片。左侧的照片表示 Koshihikari, 右侧的照片表示 Kasalath。该图将 Koshihikari 和 Kasalath 的再生能力用每克愈伤组织的再生个体数表示。

图 2 表示染色体上的再生能力 QTL 的位置。

图 3 表示再生能力 QTL 的高精度连锁图谱。

图 4 是表示互补测验结果的一组照片。左侧的照片表示当载体被单独插入 Koshihikari 中时的结果，右侧照片表示当 Kasalath 的 3F 片段被插入到 Koshihikari 中时发生的再生。

图 5 表示与 Koshihikari *NiR* 基因组序列相比 Kasalath *NiR* 基因组的突变位点。示意图中的阿拉伯数字表示插入或缺失的核苷酸数目。黑框表示编码区。垂直线表示取代位点。加外框部分表示在 Koshihikari (顶部) 和 Kasalath (底部) 中的 *NiR* 基因序列的比较。框内围住的部分表示 Koshihikari 和 Kasalath 间的不同氨基酸。粗斜体所示区域表示叶绿体转运肽结构域，虚线下划线所示区域表示铁氧还蛋白结合域，下划线部分表示 4Fe-4S 簇。

图 6 是一组照片和比较 Koshihikari 和 Kasalath 的愈伤组织中的 *NiR* 基因和 *NiR* 蛋白的表达水平的图。在左侧照片中，顶部的一行表示由半定量 RT-PCR 所检测到的 *NiR* 基因，中间一行表示用做表达对照组并由定量 RT-PCR 所检测到的水稻遍在蛋白 1 基因(*Rubq1*)，底部一行表示使用 *NiR* 蛋白抗体进行的 Western 印迹杂交所检测到的 *NiR* 蛋白。右侧图表示将 *Rubq1* 基因的表达水平用做内标通过实时定量 RT-PCR 测量 *NiR* 基因的表达水平的结果。RT-PCR 引物位点示于图 5 中。

图 7 是比较 Koshihikari 和 Kasalath NiR 重组蛋白的酶活性的图。

图 8 是一张示意图和一组表示用于确认 *NiR* 基因作为选择标记的效力的试验结果的照片。该示意图表示用于转化的二元载体的 T-DNA 区域。照片表示当每一载体被导入 Koshihikari 中时的再生状态。表格表示再生个体中 GUS-染色个体的比例。

图 9 是显示当由于肌动蛋白启动子而过表达 *NiR* 基因的载体被导入 Kasalath 中时，选择愈伤组织的结果的照片。顶部的照片表示愈伤组织选择的结果。由于亚硝酸盐被添加到培养基中，转化体“a”由于过表达的 *NiR* 基因效应而生长，而非转化体“b”的愈伤组织生长被抑制。底部的照片表示愈伤组织“a”和“b”的 GUS 染色结果。

具体实施方式

本发明将用以下实施例进行详细描述，但其不应被理解为对本发明的限制。

实施例 1：测验材料的选择和近-同源品系的制备

在培育用于 QTL 分析的杂种群体之前，选择用做群体亲本的品种。首先研究了水稻的几个品种和印度水稻的几个品种的平均再生能力，并选择了在再生能力上具有显著差别的两个品种：日本水稻“Koshihikari”和印度水稻“Kasalath”（图 1）。将日本品种“Koshihikari”和印度品种“Kasalath”杂交制备 F1 个体。然后用 Koshihikari 作为回归亲本对这些个体进行回交，并进行自体受精。生成 BC1F1 群体后，收集 BC1F2 种子。用来自每一品系的 20 粒 BC1F2 种子在诱导培养基中培养愈伤组织 30 天，然后将生长的愈伤组织转移至再生培养基中。转移 30 天之后，测量每粒种子的愈伤组织重量和苗数，用每一品系的 20 粒种子测定平均值。该平均值被当作再生能力（图 1）。采用 262 个 PCR 标记测定出每一品系的基因型。

当基于这些数据进行关于再生能力的 QTL 分析时，检出了 4 个具有再生能力增加效应的 QTL（图 2）。在其中一个 QTL 中，靠近 1

号染色体的短臂上的 TGS2451 标记 (*PSR1*; 苗再生启动子 1) 处, 成功地发现 *Kasalath* 基因组在 *Koshihikari* 的再生能力上具有很大的增加效应。用重复回交和 MAS 制备了一个 *PSR1* 近-同源品系 (Nil-*PSR1*: 用靠近 *Kasalath* 染色体 1 的 TGS2451 标记的一个区域对 *Koshihikari* 染色体进行了取代的一个品系)。对 Nil-*PSR1* 和 *Koshihikari* (对照) 的再生能力进行调查, 确认了 QTL (*PSR1*) 的存在。在其中靠近染色体 1 的短臂上的 TGS2451 的一个区域被 *Kasalath* 的相应区域取代的一个品系中, 其再生能力平均增加了 14.7 倍。

实施例 2 使用 *PSR1* 的隔离群体进行高精度连锁分析

从 BC2F1 群体中选择出 30 个其 *PSR1* 区域已被 *Kasalath* 的相应区域取代的个体。使用每种种子 (BC2F2 种子) 的 10 粒种子, 并从愈伤组织提取 DNA。使用分子标记阐明基因型, 并通过调查再生能力进行连锁分析。此外, 为了对具体基因座进行详细说明, 约 3,800 粒其中 *PSR1* 被分离的 BC3F2 种子被用于使用分子标记调查基因型, 并进行高精度连锁分析。结果, 发现 *PSR1* 定位于分子标记 3132 和 P182 间的约 50.8 kb 区域 (图 3)。在这一区域的基因预测表明存在 4 个基因, 包括一假定的蛋白质。为了测定这些基因中的那些是再生能力型基因, 构建 *Kasalath* BAC 文库 (平均长度 120 kb), 通过 PCR 筛选将一包含 *PSR1* 区域 (BHAL15) 的 BAC 克隆分离出来。BHAL15 克隆中合适的限制性酶切位点被用于制备含有各后选基因区域的 *Kasalath* 基因组片段, 并将它们导入 *Koshihikari* 中。结果发现仅当导入了包含被认为编码了铁氧还蛋白亚硝酸还原酶 (NiR) 的基因的 *Kasalath* 基因组片段 (图 3 中的 3F) 时, *Koshihikari* 的再生能力才提高 (图 4)。测定了被认为编码了铁氧还蛋白亚硝酸还原酶的遗传区域及其约 2 kb 的上游区的核苷酸序列并与 *Kasalath* 和 *Koshihikari* 进行比较, 在该核苷酸序列内发现了许多的突变 (图 5)。

实施例3 提高不可栽培品种的培养特性

Kasalath 的 *PSRI* 基因区域（或使用基因组序列或使用 cDNA 序列）被导入 Koshihikari 中以赋予 Koshihikari 以再生能力，生成高再生性 Koshihikari（图 4、8 和 9）。在这种情形中，PSR 启动子和一种组成型启动子例如肌动蛋白启动子作为用于表达 *PSRI* 基因的启动子是有效的。

实施例4 *PSRI* 基因和 PSR1 蛋白的表达分析

当通过半定量 RT-PCR 和实时定量 PCR 测定愈伤组织中的 *NiR* mRNA 表达水平时，Kasalath 中的 mRNA 的量约为 Koshihikari 中的 2.5 倍（图 6 中左侧照片的顶部和中部的行及右图）。使用特异于 NiR 蛋白的抗体进行的 Western 印迹分析也表明 NiR 蛋白在 Kasalath 中的储存量较 Koshihikari 中的储存量更大（图 6 中左侧照片的底部一行）。此外，用蔡基乙二胺方法和其表达由大肠杆菌所诱导的 NiR 重组蛋白比较每单位蛋白的 NiR 酶活性，Kasalath NiR 蛋白显示出比 Koshihikari 的 NiR 蛋白的酶活性高约 1.6 倍（图 7）。上述结果表明 Koshihikari 和 Kasalath 间在再生能力上的差异主要是由于在 *NiR* 基因的转录调节水平上的差异，其次是由于每分子合成蛋白在活性上的差异。

实施例5 使用再生能力作为选择性状进行转化

将 Kasalath *PSRI* 基因导入 Koshihikari 中能够赋予不能再生的 Koshihikari 以再生能力。这表明当对 Koshihikari 进行转化时，Kasalath *PSRI* 基因可以被用做一个选择标记。更特别地，当串联插入了 Kasalath *PSRI* 基因和一靶基因的载体被导入 Koshihikari 中时，仅那些导入了 *PSRI* 基因的细胞将获得再生能力。因此，再生的植物体应该已经同时掺入了靶基因。为了证明这一观点，构建了携带 Kasalath *NiR* 基因组+35S 启动子 *GUS*、Kasalath *NiR* 启动子:: *NiR* cDNA :: *NiR* 终止子+35S 启动子 *GUS*、水稻 *Actin1* 启动子:: *NiR* cDNA :: *NiR* 终止

子+位于 pBI101 二元载体中的 T-DNA 区域中的 35S 启动子 *GUS* 的载体和不携带 *NiR* 基因的载体，并将它们导入 Koshihikari 中。结果，当导入 3 种类型的包含 *NiR* 基因的载体时，获得了大量的再生个体，且在这些再生个体自其衍生的愈伤组织中观察到由于 *GUS* 基因所产生的着色（图 8）。

此外，*NiR* 基因具有代谢对植物有毒性的亚硝酸的特性，应用这一特性也使得 *NiR* 基因能被用作高再生性品种的转化标记。更特别地，位于肌动蛋白启动子（该启动子是水稻中的一个高表达启动子）控制下的过表达 *NiR* 基因的载体被导入到一个高再生性的 Kasalath 品种中，并在添加有亚硝酸的培养基中进行培养，所添加的亚硝酸的浓度能抑制普通野生型的生长。只有被转化的细胞由于过表达的 *NiR* 基因的效应而生长，且 *GUS* 染色仅在这些生长的细胞中被观测到（图 9）。这一选择方法的使用相对于其中使用衍生自微生物的抗生素抗性基因作为选择标记的传统方法抗生素的费用减低。此外，该方法使得可以生产更加环保的重组植物，因为该再生植物不含有微生物基因。

工业实用性

近来，应用转化方法开发有用的植物和对基因进行功能分析进展迅速。由于转化方法允许超越生理种类的界限使用基因从而可以制备新的植物，这在基于杂交和选择的传统育种中是不可能实现的。此外，由于基因组序列被接连阐明，转化方法也正被用于基因破裂、表达调节分析等以阐明每一基因的功能。通常，当制备植物转化体时，将同时包含待导入的基因和药物抗性标记基因例如抗生素抗性基因的质粒载体通过农杆菌方法或通过电穿孔导入植物细胞中，并通过药物处理选择出转化细胞。选择出来的转化细胞经过细胞生长再生为植物体。因此，为了应用这样的转化方法，必须建立组织培养技术。组织培养技术不仅在转化方法中尤其有用，而且在使用体细胞克隆变异制备突变体、使用细胞融合或胚珠培育栽培品种、遗传特性的固定、缩短育种年限等方面也尤其有用。

培育技术使用得最多的谷物是水稻，但是在品种间存在的在培育特性上的巨大差异被认为是一个问题。特别地，很难在日本培育主要的品种例如 Koshihikari 和 Akitakomachi, 以及许多在热带地区栽培的印度品种，因此这些品种不能被用做组织培养的材料。在品种之间在培育特性上的这些差异是在众多的植物中普遍观察到的现象而并非仅限于水稻，但在阐明其原因方面还没有进展。

本申请的发明人分离了与再生能力有关的基因，使得可以通过使用分子标记对高再生性性状进行有效的选择（标记选择育种），并使得可以应用分子生物学方法提高再生能力（分子育种）。此外，将 *PSRI* 基因用做选择标记已经使得生产廉价且兼顾环境的植物转化体成为可能。

谷物例如水稻、玉米、小麦和大麦是人类的主要能量来源，也是对人类最重要的植物。这些谷物都属于 *Poaceae* 家族，且可能是从一个共同的祖先进化而来。它们相互间具有高度的遗传同源性（基因组同线性）。在这些谷物中，水稻的基因组最小，这也就是为什么水稻被用做谷物的模型植物的原因。水稻基因存在于水稻的亲缘植物例如小麦和玉米的基因组中，分离自水稻的基因可以容易地分离自小麦和玉米。此外，水稻基因可以被直接用于小麦、玉米等的培育。因此，本发明的基因不仅可以应用于水稻还可应用于多种植物品种。

<110> 本田技研工业株式会社
<120> 赋予植物再生能力的基因及其应用

<130> H3-A0301P

<150> US 60/491837
<151> 2003-07-31

<160> 6

<170> PatentIn version 3.1

<210> 1
<211> 12161
<212> DNA
<213> *Oryza sativa*

<220>
<221> exon
<222> (6010)..(6418)
<223>

<220>
<221> exon
<222> (10247)..(10601)
<223>

<220>
<221> exon
<222> (10703)..(10991)
<223>

<220>
<221> exon
<222> (11076)..(11813)
<223>

<220>
<221> misc_feature
<222> (4429)..(4429)
<223> "n"=a, t, g or c.

<220>
<221> misc_feature
<222> (4479)..(4479)
<223> "n"=a, t, g or c.

<220>
<221> misc_feature
<222> (4520)..(4520)
<223> "n"=a, t, g or c.

<220>
<221> misc_feature

<222> (4555)..(4555)
 <223> "n"=a, t, g or c.

<220>
 <221> misc_feature
 <222> (4561)..(4561)
 <223> "n"=a, t, g or c.

<220>
 <221> misc_feature
 <222> (4563)..(4563)
 <223> "n"=a, t, g or c.

<220>
 <221> misc_feature
 <222> (4577)..(4578)
 <223> "n"=a, t, g or c.

<220>
 <221> misc_feature
 <222> (4823)..(4823)
 <223> "n"=a, t, g or c.

<220>
 <221> misc_feature
 <222> (7186)..(7186)
 <223> "n"=a, t, g or c.

<220>
 <221> misc_feature
 <222> (11994)..(11994)
 <223> "n"=a, t, g or c.

<400> 1
 ctcgagcttt tttgactgcc ctaatcaggc gggttccttg tgggaccac ataatgcttt 60
 tttaaatcgc cttcacgggc tgcattgcaa ctatcaggca tgggacttcc actactagaa 120
 aaaacgggcg gtcgaaacac gttttcgcag gcaggcaaac cttccacatg tatcttaacg 180
 accgtaaaaa tctccaattt tcacaggtgg accacagcac cgttttcgcg ggctacattt 240
 cgaatcttcc tgggtgctac agtaaaccac ctgcaaaaat actcacggcg ccaaaaaaaaa 300
 tttccgccag ccccgcccc tccctattca aatcacaat cacaattct cacaattctc 360
 atccaaaaac aaaatccaat ccaaaaatcc ataatcaac acaaagcatt ggattcaaat 420
 ccacaacatc aatttacaag ttaacatcaa tcaacatgta agctttaaaa cgaaacgctc 480
 tcgtcgccgg caaactcctt tgcattgagg gccgctgccc cccccctccc cctctgtccg 540
 gatttggggag ggagggagg aggtgtttgc cgccaccacc gccctcccct ctctctgtag 600
 ggccggatct cgggagggag gagaggggag ccgcctccgc acagccatca acgtccgtgc 660
 cgccgtcgcc tcgttcgcac caccgcccgtt gcttcccctc ctccggccag atctaggagc 720

ggggaggaag	agagggggag	ccaccgccac	cgtcgcccc	tcgctccgc	gccgtcgtca	780
ccgtccacgc	cgccgcgtcc	gtgcccgcgc	tgctcgtccc	cctcctctgg	cgaggaggga	840
gagagaggga	gccgtcgcgc	cgccgtcgtc	cccctccttc	ggcgaggagg	gagagagggg	900
gaggggaagag	ggatggaggg	gaggagagtg	gcgctgagag	agagagagag	agacgctgag	960
gagaggaaat	gagtgggtgg	gaggggtgga	ggagaagata	aggaggactt	agattttttt	1020
ttgggtaggt	atgatttttg	caggcggacc	acataagggt	ccgcctgcga	aatcaattt	1080
tttcacgcag	accacttaag	aggtcgcgat	gcgaaaataa	aggtattttt	ttaggcagac	1140
ctcttaagtg	gtccgcctgg	aaaaattgat	tttcacaagc	agatgacgaa	aattcaccct	1200
ggtttatatt	ttcgaagatg	cttcacgcac	gacatcgacc	gcgtcctcta	tgacggcaac	1260
gaccgcgtca	ccgacaacgg	catcgatcac	gtcatctacg	atgacaacga	ctgcatcaac	1320
tccgcatcac	tattgtgatg	actgttacat	ggcgtagaag	aaccaaccaa	agtgggtggc	1380
tcatcgccaa	cgacgtcctc	tgacatatgc	aagacgtccc	caatggcatc	ctcagacatc	1440
tacaagggtc	aagatgctaa	caattacagt	ttttgtcttc	acactgtggc	ataaatattt	1500
tttttcgctc	tcggctatat	tcggctacac	ctacaaccac	ggttactaca	tgatcggctc	1560
catcaacgaa	catctataac	aacaatcatt	gacggaaact	ccagtcaaga	gcgtctgtgt	1620
categctatc	ttccatgaca	ctcccgtat	gactacgtga	gggaatagag	gagagtcaag	1680
ggacgacacg	gaaggagacg	taggcaccag	gtggaggacc	gtccatcaaa	gatgcaattg	1740
atgatgggtg	gttgaagaag	atgaagaaat	aaaagatttc	aaatccagtc	gcaatcgttc	1800
gcttcgctcc	cgttacgact	gagggggaat	gttagaagca	tagatatatt	aattggagat	1860
aagagtcata	caaatataga	gataagatat	catcctagag	atagaattct	atagataaaa	1920
tagagtccta	gagataaatc	tactcttact	tgtacccta	tatatacccc	atgagaggat	1980
caatgcaata	caccgagaat	acaacaatta	gattttttta	cagttgtaac	tatgatacgt	2040
tgtaatatgc	tggatcgggg	aagagcgcct	gtaatcagtg	ccccagagat	gtaggtctcg	2100
gctgaactcc	attatcaaat	accgtacctc	ggtgttgtca	tcatgtttga	atcttctatg	2160
acgtttcttt	tgcatcgggt	tttcgatgtg	acttcagggc	tggttttata	ataatgatta	2220
tagtgctgtg	acggcaatcg	gttgtgagaa	ttagctattc	gggtccctcc	atgtgatttt	2280
cttgtgattg	ggatgtatgg	taatgctagg	gttttaagggt	gtaggattgg	tgcatgagag	2340
atcatcactt	cacttgatg	accttctctc	cttttatatt	tttttatcat	tctctccttt	2400
tttttataat	gctactgaac	tagtggaaata	caggggacta	atgcaaaaata	aaagaaaagt	2460

atcactggtc	acggcataca	athtagaaag	tgtgtgattt	aggcatagag	ctgaccacga	2520
ccctttacga	cttggtcgct	cggtttgta	gacgatagat	caaccaaca	aagctacgat	2580
acatgatgta	cgtgtcagga	tacaaatcct	tacaaataac	aacagttatt	gttcgataac	2640
ttttatcagt	tgtctaggct	taccaatgta	taatagaaga	tgaaaattcc	atattactgg	2700
tatcgatcaa	tgctagtaac	tctttgagct	ttgtctaggt	taaaaaaaaat	tatggatcca	2760
ccatcacaaa	aatgaaaaac	accggggaaa	acaaaaaac	atttaataac	agcacaagac	2820
aaaatgatgt	taccgtctac	ccgagctcct	actccgtacc	agcacaacca	aacgaacagt	2880
acccgccggg	tcaggggcac	gttcgtaa	ttccctcccg	tggctggctg	gctgccatct	2940
ctctcagcca	gggttggtaa	tttcggccgt	ttcgggtgggt	cccgatagta	aatgagctcc	3000
agtcaaaaacg	ccctctgcct	ccctcattg	cgccacacgc	acaccgcac	tagatccaga	3060
tcgaaaaaat	cgccatctcg	ccgagtcgcc	agtcgccgcc	tcaacgccgg	tcgccgtacc	3120
gccggcgctg	cacgcccccc	tccaagccgt	cgccccatcg	ccccagccg	cccggtgggtg	3180
gggcagcggga	tgccgagctt	ggcgaggttg	ccgaggacga	accaggcgag	gaggacgagg	3240
atcttgtcga	cgagccagag	cgggagccac	gcatgagca	acacggcgag	ctcgaacgtg	3300
gacttgccga	gcacctcgcc	agggaggacg	tggacggcgt	cgcgaccac	catcgccggg	3360
agggcgctgt	ggtcgcagag	gtcgcgcgac	accaccatgc	cggagttgcc	gcaccgcagc	3420
acgagcacct	tcttgccgcg	gtacgcctcg	ccggacttgt	agaccgcgac	atgcatcacc	3480
tcgctgctat	atgtgttctt	ggactgtgga	gacttgctgt	cagtgggtgt	gttcagaatt	3540
gctgctgcag	cttgccagca	atgtgtgatg	cagcagctac	agcttgatg	gctgccgagt	3600
agagcgagtg	ttgctatctg	ttttttgttc	tctttttcag	aaatttcgcc	cgcaaatttt	3660
aaatttgaat	tcaaattttt	aaaagaacta	gcaaatatgc	ccgtgcgttg	caccgggtga	3720
atatcaaca	aatattgatg	gtaagattg	cttgtgtact	tataacacat	atgcacaaaa	3780
atattgaata	tgtacatacc	tcgcaaatat	ctccaaattt	tatacatatg	agttgtgtaa	3840
atcgtgtgag	ttccatattg	tcatgttgat	atggagtatt	actgatgagc	ccatctatgg	3900
tgataatttt	ggaggttgta	gctcaacgaa	tttgtatttg	ctatgtatct	caacgttgat	3960
aagtcactac	cacaaccatc	ggcgacctt	ctcgggatcc	aagcatgttg	accccgccaa	4020
cgtggcgctc	gtgcagggca	ccgagatgaa	caccacgggg	ctatgtgcct	gtccagggtc	4080
atcctaggct	taaggccaacg	acactcaagg	acgtgggtggg	cggcgctcgcg	gaggtgctcc	4140
aagcgaacaa	gctggccacc	aaggaggacg	ccgacaaggt	ggcggccacc	gctatgcaga	4200

acgatgggag	gcacgccggt	gacgacaagg	agctaacacg	atccatttag	tcccgatccg	4260
agttgatcag	gaattcaatc	ctgcacctg	cggttacggt	tttcttctcc	gcgggaaaag	4320
caatcaccga	tggtagggac	aaagtgtgtg	tgagaacgga	ggccaggcca	aagtgcgtgc	4380
gagaacggag	gctaggccat	cgctggattg	gatttacgaa	tgaaatatng	atgtgacgaa	4440
cagaaaatta	tcagtttgat	ttaattttca	taatcgganc	tctttaatag	gaaaaaaaaat	4500
tacatgtacg	ttccttcatn	gtgcccattg	ccatccggga	gtccagggtt	attcncaaag	4560
ncncaatcaa	cagctannaa	tccatgtcct	tccccgccgt	tccctactct	gctttttttt	4620
ctttcatttg	aaaccttccg	ctatgaattt	ctagtcgttc	ctagcatcca	cgcacacaaa	4680
atagatttcc	ctcgcaaggc	aaaacataca	aatatgagtg	catgcaagat	attacaaacc	4740
caatccatta	aaaatagaac	ataattaact	ttagecctacc	tatctcaata	ttgggtatatg	4800
cccaaactca	aaaggagaaa	aancaaacta	aaacttttaa	taaagtgacc	ccaagagata	4860
aaaaggtgat	agtaacaaca	aatctcact	tgacaatgtc	gttgatcagc	actattttta	4920
aatattactt	aaaaatcttt	atatttacct	attaaaacaa	tgaaaaacag	aagatgtttc	4980
ttttttat	acaacagcgt	tgtatttagt	catgtcctat	ctaagagaga	aaaatgaatt	5040
taacgaaaag	aagctcagaa	aaaaaaaaaga	gaacagggcc	accacaccag	taatccctat	5100
gttatcaatg	aaaaaaaaatt	tcaatgctag	gttttttata	agaaaagggtg	ataaagtgtt	5160
gaaaaataca	gcaggaaatt	atatactctg	ctggtttaac	attaattcaa	gcatatagat	5220
ataaaaatat	atcaggctag	gaaaggaaaa	ggataaaatt	ggagagaaaa	aggaaaagaa	5280
cagtagagga	taaccagcaa	aaagatgaaa	ggattcgaac	ccatgacct	gcgttacaat	5340
tgtttcacag	gctaaccaat	cgagaatcat	cgacgtagtg	taatcttggtg	tagctacatt	5400
tgaaaaaata	tgttttgagc	tgaacgttgg	tgtgtccgcc	cctgcatccg	atacatgttg	5460
gagcgtggag	cgcggttaata	tctccttctc	tctcgtcgtc	ttctgctctc	ccccgtctct	5520
ccttcgccaa	cagccgagaa	gaggcagaga	gagcgcgcgc	ccccgtccct	ctctctccct	5580
ctcgtcctcg	ccccatccc	tctcgtcttt	cccttgccgg	cagcagagga	ggcggcagcg	5640
acggcttcag	ctgctcccac	gggccggatc	gggcagtggc	ggtggcgtcg	gcggcttccg	5700
ctggcgaatc	cggcgggtga	atcgggtgaa	atctgggtga	ccccgatac	aatcagtggt	5760
tccgataggt	aataccctgc	tctcagcatc	tgcctttttg	aattcgccaa	gagccagcat	5820
ctgccctttt	gaattcgcca	agggccagca	tctgccatt	tgattttgaa	ttcgccaaga	5880
gccagcaaca	gcgccccgcg	gccccctccc	tcctccgcaa	taaacagcca	cacgcgccgc	5940

ccccatgtcc accctcatcg ccacagcgca ccaccaccac caccaccacc accaccaccg 6000
tctccagcc atg gcc tcc tcc gcc tcc ctg cag cgc ttc ctc ccc ccg tac 6051
Met Ala Ser Ser Ala Ser Leu Gln Arg Phe Leu Pro Pro Tyr
1 5 10
ccc cac gcg gca gca tcc cgc tgc cgc cct ccc ggc gtc cgc gcc cgc 6099
Pro His Ala Ala Ala Ser Arg Cys Arg Pro Pro Gly Val Arg Ala Arg
15 20 25 30
ccc gtg cag tcg tcg acg gtg tcc gca ccg tcc tcc tcg act ccg gcg 6147
Pro Val Gln Ser Ser Thr Val Ser Ala Pro Ser Ser Ser Thr Pro Ala
35 40 45
gcg gac gag gcc gtg tcg gcg gag cgg ctg gag ccg cgg gtg gag cag 6195
Ala Asp Glu Ala Val Ser Ala Glu Arg Leu Glu Pro Arg Val Glu Gln
50 55 60
cgg gag ggc cgg tac tgg gtg ctc aag gag aag tac cgg acg ggg ctg 6243
Arg Glu Gly Arg Tyr Trp Val Leu Lys Glu Lys Tyr Arg Thr Gly Leu
65 70 75
aac ccg cag gag aag gtg aag ctg ggg aag gag ccc atg tca ttg ttc 6291
Asn Pro Gln Glu Lys Val Lys Leu Gly Lys Glu Pro Met Ser Leu Phe
80 85 90
atg gag ggc gcc atc aag gag ctc gcc aag atg ccc atg gag gag atc 6339
Met Glu Gly Gly Ile Lys Glu Leu Ala Lys Met Pro Met Glu Glu Ile
95 100 105 110
gag gcc gac aag ctc tcc aag gag gac atc gac gtg cgg ctc aag tgg 6387
Glu Ala Asp Lys Leu Ser Lys Glu Asp Ile Asp Val Arg Leu Lys Trp
115 120 125
ctc ggc ctc ttc cac cgc cgc aag cat cag t gtatgcctct cttctcttgc 6438
Leu Gly Leu Phe His Arg Arg Lys His Gln
130 135
tcctctgac aacacatttt cttgctttcg ttcggttatt tgtcgcgccg aggaagttaa 6498
ttcgccaaga tattctgcag ttttttttct cgatgcacat tcagcaacct aattaagact 6558
gattaagttg ctgtgatttt tatagcttaa ttacggctctc gtgggtaatg actatattata 6618
ttgagtaaac atggttacct ttgatccaat cacttcacct ccatgtgcca tatatagcca 6678
caggctctac caagtaacac tagtaatatg cccgtgctac gacacgggtgg cataataaat 6738
cattaaattt tattataatc aaattaagga tcctaaaatt ggtccaattg ggtgttaatt 6798
cgatgcaggt catataaaaa tatatttttag gcaaggtgca attcaagagc atcaaccatt 6858
atatccaatc actttaatat atatttgaag ataacatatg tcggaaaaaa aatgatggag 6918
agctattttca ttaacttgtg agcataaaca gatcaccaga tgatgccacc ataagtcccg 6978
ccacagtaag tgatgcagct catcttggcc taggcgttcg gtctaaccag tagatagaaa 7038

gagtacaaca	tagatcgaat	gaaaaaaaaa	atctccagaa	gaaagctcaa	ccacattgag	7098
taaattagag	caacaatcaa	atcgagtcag	catatcgtta	tgttagcaga	accaatcacc	7158
acaatttggt	tctcctcttt	atctaagngt	tttggccagg	ttaaagcat	atatactat	7218
gtccaagca	aacatcggca	atggacacgt	caaaaataaa	tgatcaattg	tttctttgag	7278
tacaaaattg	acaatggaca	ctatgttctt	ttgttagaat	tctatttgtc	agggtaggat	7338
gtagaaaaac	ttaactttta	gaggaagctt	aatatccgg	cataaacttg	ctttttcagc	7398
gctctataaa	ataattcaac	agtgaattgt	ccatcttttc	taagtgctcc	aaaagacact	7458
aagttgaaaa	accaggtgaa	ccaacagatt	gatccacaaa	atcttattat	tagattattc	7518
acttaaagc	ctgtctttat	ttcaaacata	taaaaacaga	agttattaat	caggggaagcg	7578
cttatggcag	cctgagcgaa	ccagtgatag	caagtgggtga	aacagtaaa	taggatacat	7638
aaaattata	caaggtttct	actgtttatc	gaaaaaaaaat	atgtgaaaac	agtaaatagg	7698
atacataatc	gacttccaac	ttgtccttat	cataacatcc	agaatcacia	caagaattgc	7758
aacgaataca	tagtcgactt	gagctaagaa	gtcacaagac	ctgtcaaagt	aagctgcctt	7818
tgatcttgaa	gtgaaaggca	tattttattg	tcttccttgg	caaacagata	tcactgtctt	7878
cagcagttca	gtagataat	ccaagatttc	tcacggagaa	gagcatatca	ctcacatcag	7938
tgttgtgccc	tccaaatact	gagataaact	gaattttggt	ctctttgaag	catctgcagg	7998
cattaacaat	aataatactt	tacaaagttt	cattgggtct	aaactattgt	ttgcacatca	8058
tatatatgcc	cagaactttt	tagcatgata	caagggtcct	gttcataact	catgcctaaa	8118
tctgacaaat	ttgtcaaacg	acaatataag	togaattata	atgctgtttta	gaattgacgc	8178
caaaactttt	gctagcgtaa	gtaactcttc	cacctcccag	catgcataca	accaacaagc	8238
taaacttttg	ttcaaaaaa	tgtacattta	tttccttgaa	cacagccttt	gtagaatatg	8298
atataaaaact	catggatgaa	tgaaataatg	taaaagaatg	gtcaaaaatga	tgaatagtac	8358
aagaagcaac	tgtgaacatt	tcacctttac	ctgactgttc	gcaagaaggc	cacgtggcag	8418
aaaagccaga	aatgcaagaa	gcttccttaa	ttgatacacc	atcaagaaat	caatggactc	8478
aacaccagcg	tccgcccaga	caaatgaat	gcaggcacct	aaaatataga	accattgact	8538
tttcaacact	gaattatata	acctgaatat	cttgttttgt	taacacatct	gacaaaatca	8598
gtgcattctg	ttccatatag	atgtatgcat	agctcccata	tgttagtga	tcgatgagca	8658
tgcaaacat	acacacctta	cgttactccc	tctgtcaaaa	aaaatataag	cttgtctaga	8718
tacatagcta	caaatgctta	tatttttggg	ttctcttaa	gctgtagaaa	cttttatcgc	8778

```

cccgccatgg caagtcgagc tgccatcccc aatgaaagcc cccacacagg tttcatgccc 8838
tgctgcacaa tattgagcaa ccaaaaatat aataatattt gtgtcagaat ttgaatcaac 8898
cttacagata ctgggtggcc agaaaatcta gtccaagtaa taccctgaaa aatagcaact 8958
ggcaaatact aaaggcagtg aagagtttcc tttagatcag atgataaaaa aaaatcatat 9018
gttcaatagc aataatcact cacatTTTTT ttgctgttta gaatttagat aaatagtagt 9078
taaacttcta tagcttgcgt agctaagatc aatgggtgatt attagttgaa aaaataatca 9138
aatcatcaaa ctgaggagac ttatacctgc cataagttct gaaatttcaa tgatcctagt 9198
caatatttac tgtatatata gaattaggtc caaaagatga tacttacaat taaggatggt 9258
gtattgatcg gttcataact caagcttcta tttatcatta atcaaaagct ggatcattca 9318
tgcataatac tttgccgcac tcaacatagc agctcggagt cttctttggt cagaagcgag 9378
gaaggagtca acaaataagt actgcaatgt taaacaaacc gacatatcaa atcccaaatt 9438
aagaatgcat gatttattaa tacaggaaat atatgatcaa gtcccaaaaa gtgagtcatg 9498
ttatgtacac tcagtcatca atttcaataa gaatattaac ttgctcattg gtatatggat 9558
ttgattatga cataatttga caatacattt acagaataaa cttgcagtgct tgtgagcata 9618
tgttactaac atgtaaggac cttgttttgc tctgttcaat actcatgttg atcttgatct 9678
gtgtccacat atacctaaat gaaatgaaat caaagaatga ggtttgtagg agtggagttg 9738
gtgaattata gggtagataa tgcgggcaca accgtttgat aagtagtacg agtactttat 9798
ttggcgccac cgcgccagca tcagatgtgt ggctttgca ctgattgaac ccaaaagaaa 9858
aaaaaaagtc gttttggtec cacacaattc tacttcatct gcaggatgta cagaaggtta 9918
catatctatt ctgttctatg ctctgtttac atttataagg gctcacttgg tggctgtcat 9978
tggttggctg gtgcggtata ttactaatag gttttttaat ggcatatatg ttcttaaaat 10038
aaaccagaaa agcaaaagat caactatctt agccacacca atgaaatgga atatactgaa 10098
ctgtcacggc taaaattctc ttcagtcacc tggcccagct ggagccgtgg gctcgtcgtc 10158
ttttctaaac atgtactagt attttggggg cccacagtga atttggccca aatgctgac 10218
agccgctcta cggtctctac ctgtgcag at   ggg cgg ttc atg atg cgg ctg 10269
                               Tyr Gly Arg Phe Met Met Arg Leu
                               140

aag ctg cca aac ggt gtg acg acg agc gag cag acg agg tac ctg gcg 10317
Lys Leu Pro Asn Gly Val Thr Thr Ser Glu Gln Thr Arg Tyr Leu Ala
145                               150                               155                               160

agc gtg atc gag gcg tac ggc aag gag ggc tgc gcc gac gtg aca acc 10365
Ser Val Ile Glu Ala Tyr Gly Lys Glu Gly Cys Ala Asp Val Thr Thr

```

	165	170	175	
cgc cag aac tgg cag atc cgc ggc gtc acg ctc ccc gac gtg ccg gcc				10413
Arg Gln Asn Trp Gln Ile Arg Gly Val Thr Leu Pro Asp Val Pro Ala				
	180	185	190	
atc ctc gac ggg ctc aac gcc gtc ggc ctc acc agc ctc cag agc ggc				10461
Ile Leu Asp Gly Leu Asn Ala Val Gly Leu Thr Ser Leu Gln Ser Gly				
	195	200	205	
atg gac aac gtc cgc aac ccc gtc ggc aac ccg ctc gcc ggc atc gac				10509
Met Asp Asn Val Arg Asn Pro Val Gly Asn Pro Leu Ala Gly Ile Asp				
	210	215	220	
ccc gac gag atc gtc gac acg cga tcc tac acc aac ctc ctc tcc tcc				10557
Pro Asp Glu Ile Val Asp Thr Arg Ser Tyr Thr Asn Leu Leu Ser Ser				
	225	230	235	240
tac atc acc agc aac ttc cag ggc aac ccc acc atc acc aac ct				10601
Tyr Ile Thr Ser Asn Phe Gln Gly Asn Pro Thr Ile Thr Asn Leu				
	245	250		
gtgagtgatc gaatcaactt gatcatgctc tgtgctgtgc tgttcgtgtc gtctctgacg				10661
acatgtttgt tgaatttggt gttgctgcgt gctgttggca g g ccg agg aag tgg				10715
			Pro Arg Lys Trp	
aac gtg tgc gtg atc ggg tcg cac gat ctg tac gag cac ccg cac atc				10763
Asn Val Cys Val Ile Gly Ser His Asp Leu Tyr Glu His Pro His Ile				
	260	265	270	275
aac gac ctc gcg tac atg ccg gcg gtg aag ggc ggc aag ttc ggg ttc				10811
Asn Asp Leu Ala Tyr Met Pro Ala Val Lys Gly Gly Lys Phe Gly Phe				
	280	285	290	
aac ctc ctt gtc ggc ggg ttc atc agc ccc aag agg tgg gag gag gcg				10859
Asn Leu Leu Val Gly Gly Phe Ile Ser Pro Lys Arg Trp Glu Glu Ala				
	295	300	305	
ctg ccg ctg gac gcc tgg gtc ccc ggc gac gac atc atc ccg gtg tgc				10907
Leu Pro Leu Asp Ala Trp Val Pro Gly Asp Asp Ile Ile Pro Val Cys				
	310	315	320	
aag gcc gtt ctc gag gcg tac cgc gac ctc ggc acc agg ggc aac cgc				10955
Lys Ala Val Leu Glu Ala Tyr Arg Asp Leu Gly Thr Arg Gly Asn Arg				
	325	330	335	
cag aag acc cgc atg atg tgg ctc atc gac gaa ctt gtgagcctcc				11001
Gln Lys Thr Arg Met Met Trp Leu Ile Asp Glu Leu				
	340	345	350	
attcatccac gccattgact gaattacgta tgtcccaatg ttcttatcag ttaattgcgg				11061
tgttggcatt gcag gga atg gag gct ttt cgg tcg gag gtg gag aag agg				11111
	Gly Met Glu Ala Phe Arg Ser Glu Val Glu Lys Arg			
	355	360		

atg	ccg	aac	ggc	gtg	ctg	gag	cgc	gct	gcg	ccg	gac	gac	ctc	atc	gac	11159
Met	Pro	Asn	Gly	Val	Leu	Glu	Arg	Ala	Ala	Pro	Asp	Asp	Leu	Ile	Asp	
	365					370					375					
aag	aaa	tgg	cag	agg	agg	gac	tac	ctc	ggc	gtg	cac	ccg	cag	aag	cag	11207
Lys	Lys	Trp	Gln	Arg	Arg	Asp	Tyr	Leu	Gly	Val	His	Pro	Gln	Lys	Gln	
	380				385					390					395	
gaa	ggg	atg	tcc	tac	gtc	ggc	ctg	cac	gtg	ccc	gtc	ggc	cgg	gtg	cag	11255
Glu	Gly	Met	Ser	Tyr	Val	Gly	Leu	His	Val	Pro	Val	Gly	Arg	Val	Gln	
				400					405						410	
gcg	gcg	gac	atg	ttc	gag	ctc	gcc	cgc	ctt	gcc	gac	gag	tat	ggc	tcc	11303
Ala	Ala	Asp	Met	Phe	Glu	Leu	Ala	Arg	Leu	Ala	Asp	Glu	Tyr	Gly	Ser	
			415					420					425			
ggc	gag	ctc	cgc	ctc	acc	gtg	gag	cag	aac	atc	gtg	atc	ccg	aac	gtc	11351
Gly	Glu	Leu	Arg	Leu	Thr	Val	Glu	Gln	Asn	Ile	Val	Ile	Pro	Asn	Val	
		430					435					440				
aag	aac	gag	aag	gtg	gag	gcg	ctg	ctc	gcc	gag	ccg	ctg	ctt	cag	aag	11399
Lys	Asn	Glu	Lys	Val	Glu	Ala	Leu	Leu	Ala	Glu	Pro	Leu	Leu	Gln	Lys	
	445					450					455					
ttc	tcc	ccg	cag	ccg	tcg	ctg	ctg	ctc	aag	ggc	ctg	gtc	gcg	tgc	acc	11447
Phe	Ser	Pro	Gln	Pro	Ser	Leu	Leu	Leu	Lys	Gly	Leu	Val	Ala	Cys	Thr	
	460				465					470					475	
ggc	aac	cag	ttc	tgc	ggc	cag	gcc	atc	atc	gag	acg	aag	cag	cgg	gcg	11495
Gly	Asn	Gln	Phe	Cys	Gly	Gln	Ala	Ile	Ile	Glu	Thr	Lys	Gln	Arg	Ala	
				480					485					490		
ctg	ctg	gtg	acg	tcg	cag	gtg	gag	aag	ctc	gtg	tcg	gtg	ccc	cgg	gcg	11543
Leu	Leu	Val	Thr	Ser	Gln	Val	Glu	Lys	Leu	Val	Ser	Val	Pro	Arg	Ala	
			495					500					505			
gtg	cgg	atg	cac	tgg	acc	ggc	tgc	ccc	aac	agc	tgc	ggc	cag	gtg	cag	11591
Val	Arg	Met	His	Trp	Thr	Gly	Cys	Pro	Asn	Ser	Cys	Gly	Gln	Val	Gln	
		510					515					520				
gtc	gcc	gac	atc	ggc	ttc	atg	ggc	tgc	ctc	acc	aag	gat	agc	gcc	ggc	11639
Val	Ala	Asp	Ile	Gly	Phe	Met	Gly	Cys	Leu	Thr	Lys	Asp	Ser	Ala	Gly	
		525				530					535					
aag	atc	gtc	gag	gcg	gcc	gac	atc	ttc	gtc	ggc	ggc	cgc	gtc	ggc	agc	11687
Lys	Ile	Val	Glu	Ala	Ala	Asp	Ile	Phe	Val	Gly	Gly	Arg	Val	Gly	Ser	
	540				545					550					555	
gac	tcg	cac	ctc	gcc	ggc	gcg	tac	aag	aag	tcc	gtg	ccg	tgc	gac	gag	11735
Asp	Ser	His	Leu	Ala	Gly	Ala	Tyr	Lys	Lys	Ser	Val	Pro	Cys	Asp	Glu	
				560					565					570		
ctg	gcg	ccg	atc	gtc	gcc	gac	atc	ctg	gtc	gag	cgg	ttc	ggg	gcc	gtg	11783
Leu	Ala	Pro	Ile	Val	Ala	Asp	Ile	Leu	Val	Glu	Arg	Phe	Gly	Ala	Val	
			575					580					585			
cgg	agg	gag	agg	gag	gag	gac	gag	gag	tag	gagc	cagac	tgggg	gtggtt			11833
Arg	Arg	Glu	Arg	Glu	Glu	Asp	Glu	Glu								

590 595

tgcttgctcc ggtgatctct cgccgtcctt gtaaagtaga cgacaatatg ccttcgccca 11893

tggcacgctt gtactgtcac gttttggttt gatctttag ccctaaagtt gtgttcattc 11953

tcgttacagt cttacagagg atgattgatt gataaataaa naagaaacag attctgcaac 12013

tgttcatcgc tgttcctaaa tctgatttcg cgatagtatc ttgtctgacc tgtcccaatc 12073

gcagtgctaa aaccatataa tcttgcaagc aatgaaatt gaaagagttc aatgcaacca 12133

ctaacgggtct aacaacatga taaggcct 12161

<210> 2
 <211> 2519
 <212> DNA
 <213> Oryza sativa

<220>
 <221> CDS
 <222> (532)..(2322)
 <223>

<400> 2

tatctccttc tctctcgtcg ctttctgcgt ctccccgtct ctcttcgcc aacagccgag 60

aagaggcaga gagagcgccg cccccgtcc ctctctctcc ctctcgtcct cgcccccatc 120

cctctcgtct ttccttgcc ggcagcagag gaggcggcag cgacggcttc agctgctccc 180

acgggcccga tcgggcagtg gcggtggcgt cggcggcttc cgctggcgaa tccggcgggt 240

gaatcgggtg aaatttggtg gacccccgat acaaatcagt gttccgatag gtaataccct 300

gctctcagca tctgcccttt tgaattcgcc aagagccagc atctgccctt ttgaattcgc 360

caagggccag catctgcccc tttgattttg aattcgccaa gagccagcaa cagcgccccc 420

gcgccccctc cctcctcgc aataaacagc cacacgcgcc gccccatgt ccaccctcat 480

cgccacagcg caccaccacc accaccacca ccaccaccac cgtctccagc c atg gcc 537
 Met Ala
 1

tcc tcc gcc tcc ctg cag cgc ttc ctc ccc ccg tac ccc cac gcg gca 585
 Ser Ser Ala Ser Leu Gln Arg Phe Leu Pro Pro Tyr Pro His Ala Ala
 5 10 15

gca tcc cgc tgc cgc cct ccc ggc gtc cgc gcc cgc ccc gtg cag tcg 633
 Ala Ser Arg Cys Arg Pro Pro Gly Val Arg Ala Arg Pro Val Gln Ser
 20 25 30

tcg acg gtg tcc gca ccg tcc tcc tcg act ccg gcg gcg gac gag gcc 681
 Ser Thr Val Ser Ala Pro Ser Ser Ser Thr Pro Ala Ala Asp Glu Ala
 35 40 45 50

gtg	tcg	gcg	gag	cgg	ctg	gag	ccg	cgg	gtg	gag	cag	cgg	gag	ggc	cgg	729
Val	Ser	Ala	Glu	Arg	Leu	Glu	Pro	Arg	Val	Glu	Gln	Arg	Glu	Gly	Arg	
				55					60					65		
tac	tgg	gtg	ctc	aag	gag	aag	tac	cgg	acg	ggg	ctg	aac	ccg	cag	gag	777
Tyr	Trp	Val	Leu	Lys	Glu	Lys	Tyr	Arg	Thr	Gly	Leu	Asn	Pro	Gln	Glu	
			70					75					80			
aag	gtg	aag	ctg	ggg	aag	gag	ccc	atg	tca	ttg	ttc	atg	gag	ggc	ggc	825
Lys	Val	Lys	Leu	Gly	Lys	Glu	Pro	Met	Ser	Leu	Phe	Met	Glu	Gly	Gly	
		85					90					95				
atc	aag	gag	ctc	gcc	aag	atg	ccc	atg	gag	gag	atc	gag	gcc	gac	aag	873
Ile	Lys	Glu	Leu	Ala	Lys	Met	Pro	Met	Glu	Glu	Ile	Glu	Ala	Asp	Lys	
	100					105					110					
ctc	tcc	aag	gag	gac	atc	gac	gtg	cgg	ctc	aag	tgg	ctc	ggc	ctc	ttc	921
Leu	Ser	Lys	Glu	Asp	Ile	Asp	Val	Arg	Leu	Lys	Trp	Leu	Gly	Leu	Phe	
115					120					125					130	
cac	cgc	cgc	aag	cat	cag	tat	ggg	cgg	ttc	atg	atg	cgg	ctg	aag	ctg	969
His	Arg	Arg	Lys	His	Gln	Tyr	Gly	Arg	Phe	Met	Met	Arg	Leu	Lys	Leu	
				135					140					145		
cca	aac	ggt	gtg	acg	acg	agc	gag	cag	acg	agg	tac	ctg	gcg	agc	gtg	1017
Pro	Asn	Gly	Val	Thr	Thr	Ser	Glu	Gln	Thr	Arg	Tyr	Leu	Ala	Ser	Val	
			150					155					160			
atc	gag	gcg	tac	ggc	aag	gag	ggc	tgc	gcc	gac	gtg	aca	acc	cgc	cag	1065
Ile	Glu	Ala	Tyr	Gly	Lys	Glu	Gly	Cys	Ala	Asp	Val	Thr	Thr	Arg	Gln	
		165					170					175				
aac	tgg	cag	atc	cgc	ggc	gtc	acg	ctc	ccc	gac	gtg	ccg	gcc	atc	ctc	1113
Asn	Trp	Gln	Ile	Arg	Gly	Val	Thr	Leu	Pro	Asp	Val	Pro	Ala	Ile	Leu	
	180					185					190					
gac	ggg	ctc	aac	gcc	gtc	ggc	ctc	acc	agc	ctc	cag	agc	ggc	atg	gac	1161
Asp	Gly	Leu	Asn	Ala	Val	Gly	Leu	Thr	Ser	Leu	Gln	Ser	Gly	Met	Asp	
195				200						205				210		
aac	gtc	cgc	aac	ccc	gtc	ggc	aac	ccg	ctc	gcc	ggc	atc	gac	ccc	gac	1209
Asn	Val	Arg	Asn	Pro	Val	Gly	Asn	Pro	Leu	Ala	Gly	Ile	Asp	Pro	Asp	
				215					220					225		
gag	atc	gtc	gac	acg	cga	tcc	tac	acc	aac	ctc	ctc	tcc	tcc	tac	atc	1257
Glu	Ile	Val	Asp	Thr	Arg	Ser	Tyr	Thr	Asn	Leu	Leu	Ser	Ser	Tyr	Ile	
			230					235					240			
acc	agc	aac	ttc	cag	ggc	aac	ccc	acc	atc	acc	aac	ctg	ccg	agg	aag	1305
Thr	Ser	Asn	Phe	Gln	Gly	Asn	Pro	Thr	Ile	Thr	Asn	Leu	Pro	Arg	Lys	
			245				250					255				
tgg	aac	gtg	tgc	gtg	atc	ggg	tgc	cac	gat	ctg	tac	gag	cac	ccg	cac	1353
Trp	Asn	Val	Cys	Val	Ile	Gly	Ser	His	Asp	Leu	Tyr	Glu	His	Pro	His	
	260					265					270					
atc	aac	gac	ctc	gcg	tac	atg	ccg	gcg	gtg	aag	ggc	ggc	aag	ttc	ggg	1401
Ile	Asn	Asp	Leu	Ala	Tyr	Met	Pro	Ala	Val	Lys	Gly	Gly	Lys	Phe	Gly	

275	280							285							290			
ttc aac ctc ctt gtc ggc ggg ttc atc agc ccc aag agg tgg gag gag	gac gcc tgg gtc ccc ggc gac gac atc atc ccg gtg	gac gac atc atc ccg gtg	1449															
Phe Asn Leu Leu Val Gly Gly Phe Ile Ser Pro Lys Arg Trp Glu Glu	295	300	305															
gcg ctg ccg ctg gac gcc tgg gtc ccc ggc gac gac atc atc ccg gtg	gac gcc tgg gtc ccc ggc gac gac atc atc ccg gtg	gac gcc tgg gtc ccc ggc gac gac atc atc ccg gtg	1497															
Ala Leu Pro Leu Asp Ala Trp Val Pro Gly Asp Asp Ile Ile Pro Val	310	315	320															
tgc aag gcc gtt ctc gag gcg tac cgc gac ctc ggc acc agg ggc aac	gag gcg tac cgc gac ctc ggc acc agg ggc aac	gag gcg tac cgc gac ctc ggc acc agg ggc aac	1545															
Cys Lys Ala Val Leu Glu Ala Tyr Arg Asp Leu Gly Thr Arg Gly Asn	325	330	335															
cgc cag aag acc cgc atg atg tgg ctc atc gac gaa ctt gga atg gag	atg atg tgg ctc atc gac gaa ctt gga atg gag	atg atg tgg ctc atc gac gaa ctt gga atg gag	1593															
Arg Gln Lys Thr Arg Met Met Trp Leu Ile Asp Glu Leu Gly Met Glu	340	345	350															
gct ttt cgg tcg gag gtg gag aag agg atg ccg aac ggc gtg ctg gag	gtg gag aag agg atg ccg aac ggc gtg ctg gag	gtg gag aag agg atg ccg aac ggc gtg ctg gag	1641															
Ala Phe Arg Ser Glu Val Glu Lys Arg Met Pro Asn Gly Val Leu Glu	355	360	365	370														
cgc gct gcg ccg gac gac ctc atc gac aag aaa tgg cag agg agg gac	gac gac ctc atc gac aag aaa tgg cag agg agg gac	gac gac ctc atc gac aag aaa tgg cag agg agg gac	1689															
Arg Ala Ala Pro Asp Asp Leu Ile Asp Lys Lys Trp Gln Arg Arg Asp	375	380	385															
tac ctc ggc gtg cac ccg cag aag cag gaa ggg atg tcc tac gtc ggc	gtg cac ccg cag aag cag gaa ggg atg tcc tac gtc ggc	gtg cac ccg cag aag cag gaa ggg atg tcc tac gtc ggc	1737															
Tyr Leu Gly Val His Pro Gln Lys Gln Glu Gly Met Ser Tyr Val Gly	390	395	400															
ctg cac gtg ccc gtc ggc cgg gtg cag gcg gcg gac atg ttc gag ctc	gtg cag gcg gcg gac atg ttc gag ctc	gtg cag gcg gcg gac atg ttc gag ctc	1785															
Leu His Val Pro Val Gly Arg Val Gln Ala Ala Asp Met Phe Glu Leu	405	410	415															
gcc cgc ctt gcc gac gag tat ggc tcc ggc gag ctc cgc ctc acc gtg	gac gag tat ggc tcc ggc gag ctc cgc ctc acc gtg	gac gag tat ggc tcc ggc gag ctc cgc ctc acc gtg	1833															
Ala Arg Leu Ala Asp Glu Tyr Gly Ser Gly Glu Leu Arg Leu Thr Val	420	425	430															
gag cag aac atc gtg atc ccg aac gtc aag aac gag aag gtg gag gcg	gtg atc ccg aac gtc aag aac gag aag gtg gag gcg	gtg atc ccg aac gtc aag aac gag aag gtg gag gcg	1881															
Glu Gln Asn Ile Val Ile Pro Asn Val Lys Asn Glu Lys Val Glu Ala	435	440	445	450														
ctg ctc gcc gag ccg ctg ctt cag aag ttc tcc ccg cag ccg tcg ctg	gag ccg ctg ctt cag aag ttc tcc ccg cag ccg tcg ctg	gag ccg ctg ctt cag aag ttc tcc ccg cag ccg tcg ctg	1929															
Leu Leu Ala Glu Pro Leu Leu Gln Lys Phe Ser Pro Gln Pro Ser Leu	455	460	465															
ctg ctc aag ggc ctg gtc gcg tgc acc ggc aac cag ttc tgc ggc cag	gtc gcg tgc acc ggc aac cag ttc tgc ggc cag	gtc gcg tgc acc ggc aac cag ttc tgc ggc cag	1977															
Leu Leu Lys Gly Leu Val Ala Cys Thr Gly Asn Gln Phe Cys Gly Gln	470	475	480															
gcc atc atc gag acg aag cag cgg gcg ctg ctg gtg acg tcg cag gtg	atc gag acg aag cag cgg gcg ctg ctg gtg acg tcg cag gtg	atc gag acg aag cag cgg gcg ctg ctg gtg acg tcg cag gtg	2025															
Ala Ile Ile Glu Thr Lys Gln Arg Ala Leu Leu Val Thr Ser Gln Val	485	490	495															
gag aag ctc gtg tcg gtg ccc cgg gcg gtg cgg atg cac tgg acc ggc	gtg tcg gtg ccc cgg gcg gtg cgg atg cac tgg acc ggc	gtg tcg gtg ccc cgg gcg gtg cgg atg cac tgg acc ggc	2073															
Glu Lys Leu Val Ser Val Pro Arg Ala Val Arg Met His Trp Thr Gly	500	505	510															

```

tgc ccc aac agc tgc ggc cag gtg cag gtc gcc gac atc ggc ttc atg      2121
Cys Pro Asn Ser Cys Gly Gln Val Gln Val Ala Asp Ile Gly Phe Met
515                               520                               525                               530

ggc tgc ctc acc aag gac agc gcc ggc aag atc gtc gag gcg gcc gac      2169
Gly Cys Leu Thr Lys Asp Ser Ala Gly Lys Ile Val Glu Ala Ala Asp
                    535                               540                               545

atc ttc gtc ggc ggc cgc gtc ggc agc gac tcg cac ctc gcc ggc gcg      2217
Ile Phe Val Gly Gly Arg Val Gly Ser Asp Ser His Leu Ala Gly Ala
                    550                               555                               560

tac aag aag tcc gtg ccg tgc gac gag ctg gcg ccg atc gtc gcc gac      2265
Tyr Lys Lys Ser Val Pro Cys Asp Glu Leu Ala Pro Ile Val Ala Asp
                    565                               570                               575

atc ctg gtc gag cgg ttc ggg gcc gtg cgg agg gag agg gag gag gac      2313
Ile Leu Val Glu Arg Phe Gly Ala Val Arg Arg Glu Arg Glu Glu Asp
                    580                               585                               590

gag gag tag gaggcacagac tgggggtggtt tgcttgctcc ggtgatctct      2362
Glu Glu
595

cgccgtcctt gtaaagtaga cgacaatatg ccttcgcccga tggcacgctt gtactgtcac      2422

gttttggttt gatctttag cccaaaagtt gtgttcattc tcgttacagt cttacagagg      2482

atgattgatt gataaataaa gaagaaacag attctgc      2519

<210> 3
<211> 596
<212> PRT
<213> Oryza sativa

<400> 3
Met Ala Ser Ser Ala Ser Leu Gln Arg Phe Leu Pro Pro Tyr Pro His
1                               5                               10                               15

Ala Ala Ala Ser Arg Cys Arg Pro Pro Gly Val Arg Ala Arg Pro Val
                20                               25                               30

Gln Ser Ser Thr Val Ser Ala Pro Ser Ser Ser Thr Pro Ala Ala Asp
                35                               40                               45

Glu Ala Val Ser Ala Glu Arg Leu Glu Pro Arg Val Glu Gln Arg Glu
                50                               55                               60

Gly Arg Tyr Trp Val Leu Lys Glu Lys Tyr Arg Thr Gly Leu Asn Pro
65                               70                               75                               80

Gln Glu Lys Val Lys Leu Gly Lys Glu Pro Met Ser Leu Phe Met Glu
                85                               90                               95

Gly Gly Ile Lys Glu Leu Ala Lys Met Pro Met Glu Glu Ile Glu Ala
                100                               105                               110

```

Asp Lys Leu Ser Lys Glu Asp Ile Asp Val Arg Leu Lys Trp Leu Gly
 115 120 125
 Leu Phe His Arg Arg Lys His Gln Tyr Gly Arg Phe Met Met Arg Leu
 130 135 140
 Lys Leu Pro Asn Gly Val Thr Thr Ser Glu Gln Thr Arg Tyr Leu Ala
 145 150 155 160
 Ser Val Ile Glu Ala Tyr Gly Lys Glu Gly Cys Ala Asp Val Thr Thr
 165 170 175
 Arg Gln Asn Trp Gln Ile Arg Gly Val Thr Leu Pro Asp Val Pro Ala
 180 185 190
 Ile Leu Asp Gly Leu Asn Ala Val Gly Leu Thr Ser Leu Gln Ser Gly
 195 200 205
 Met Asp Asn Val Arg Asn Pro Val Gly Asn Pro Leu Ala Gly Ile Asp
 210 215 220
 Pro Asp Glu Ile Val Asp Thr Arg Ser Tyr Thr Asn Leu Leu Ser Ser
 225 230 235 240
 Tyr Ile Thr Ser Asn Phe Gln Gly Asn Pro Thr Ile Thr Asn Leu Pro
 245 250 255
 Arg Lys Trp Asn Val Cys Val Ile Gly Ser His Asp Leu Tyr Glu His
 260 265 270
 Pro His Ile Asn Asp Leu Ala Tyr Met Pro Ala Val Lys Gly Gly Lys
 275 280 285
 Phe Gly Phe Asn Leu Leu Val Gly Gly Phe Ile Ser Pro Lys Arg Trp
 290 295 300
 Glu Glu Ala Leu Pro Leu Asp Ala Trp Val Pro Gly Asp Asp Ile Ile
 305 310 315 320
 Pro Val Cys Lys Ala Val Leu Glu Ala Tyr Arg Asp Leu Gly Thr Arg
 325 330 335
 Gly Asn Arg Gln Lys Thr Arg Met Met Trp Leu Ile Asp Glu Leu Gly
 340 345 350
 Met Glu Ala Phe Arg Ser Glu Val Glu Lys Arg Met Pro Asn Gly Val
 355 360 365
 Leu Glu Arg Ala Ala Pro Asp Asp Leu Ile Asp Lys Lys Trp Gln Arg
 370 375 380
 Arg Asp Tyr Leu Gly Val His Pro Gln Lys Gln Glu Gly Met Ser Tyr
 385 390 395 400
 Val Gly Leu His Val Pro Val Gly Arg Val Gln Ala Ala Asp Met Phe
 405 410 415
 Glu Leu Ala Arg Leu Ala Asp Glu Tyr Gly Ser Gly Glu Leu Arg Leu

```

                420                425                430
Thr Val Glu Gln Asn Ile Val Ile Pro Asn Val Lys Asn Glu Lys Val
      435                440                445
Glu Ala Leu Leu Ala Glu Pro Leu Leu Gln Lys Phe Ser Pro Gln Pro
      450                455                460
Ser Leu Leu Leu Lys Gly Leu Val Ala Cys Thr Gly Asn Gln Phe Cys
465                470                475                480
Gly Gln Ala Ile Ile Glu Thr Lys Gln Arg Ala Leu Leu Val Thr Ser
      485                490                495
Gln Val Glu Lys Leu Val Ser Val Pro Arg Ala Val Arg Met His Trp
      500                505                510
Thr Gly Cys Pro Asn Ser Cys Gly Gln Val Gln Val Ala Asp Ile Gly
      515                520                525
Phe Met Gly Cys Leu Thr Lys Asp Ser Ala Gly Lys Ile Val Glu Ala
      530                535                540
Ala Asp Ile Phe Val Gly Gly Arg Val Gly Ser Asp Ser His Leu Ala
545                550                555                560
Gly Ala Tyr Lys Lys Ser Val Pro Cys Asp Glu Leu Ala Pro Ile Val
      565                570                575
Ala Asp Ile Leu Val Glu Arg Phe Gly Ala Val Arg Arg Glu Arg Glu
      580                585                590
Glu Asp Glu Glu
      595

```

```

<210> 4
<211> 12179
<212> DNA
<213> Oryza sativa

```

```

<220>
<221> exon
<222> (6001)..(6409)
<223>

```

```

<220>
<221> exon
<222> (10255)..(10609)
<223>

```

```

<220>
<221> exon
<222> (10712)..(11000)
<223>

```

```

<220>
<221> exon

```

<222> (11094)..(11831)
<223>

<400> 4

ctcgagcttt	tttgactgcc	ctaatacaggc	gggttccttg	tgggaccac	ataatgcttt	60
ttttaatcgc	cttcacgggc	tgcatgcaaa	ctatacggcg	tggtaacttcc	actactagaa	120
aaaacgggct	tttcgcaggc	gggcaaacct	tccgcatgta	tattaacgac	cgtaaaaatc	180
tccaattttc	acaggtggac	cccagcaccg	cctgcgaaaa	taattttcgc	aggctgcatt	240
tcgaatcttc	ctgggtgcta	cagtaaacca	cctgcgaaaa	tactcacggc	gccccaaaa	300
aaattttcgc	cagccccgcc	ccctccctat	tcaaatcaca	aattctcaca	aatctcatcc	360
aaaaacaaaa	ttcaatccaa	aatccatac	atcaacacaa	agcattggat	tcaaatccac	420
aacatcaatt	tacaagttaa	catcaatcaa	catgtaagct	ttaaaacgaa	acgtcgtcgt	480
cgccggcaaa	ctccttttgc	atgcggtgcc	gccgccgcc	ccctcccccc	tctgtccgga	540
tttgggaggg	agggaggtgt	ttgccgccac	caccgccctc	ccctctctctc	gtagggccgg	600
atctcgggag	ggaggagagg	ggagccgcct	ccgcacagcc	atcaacgtcc	gtgccgccgt	660
cgctcgttc	gcaccaccgc	cgttgcttcc	cctcctccgg	ccagatctag	gagcggggag	720
gaagagaggg	ggagccaccg	ccaccgtcgc	ccctcgcgt	ccgcgccgtc	gtcaccgtcc	780
acgccgccgc	gtccgtgccg	ccgctgtcgc	tccccctct	ctggcgagga	gggagagaga	840
gggagccgtc	gcgccgccgt	cgctccccctc	cttcggcgag	gagggagaga	gggggagggga	900
agagggatgg	aggggaggag	agtggcgctg	agagagagag	agagagacgc	tgaggagagg	960
aatgagtgg	tggggagggg	tggaggagaa	gataaggagg	acttagattt	tttttttggg	1020
taagtatgat	ttttgcaggc	ggaccacata	aggttccgcc	tgcgaaaatc	aattttttctg	1080
cgcagaccac	ttaagaggtc	cgcatgcgaa	aataaaggta	tttttttagg	cggacctctt	1140
aagtggtcgc	cctggaaaaa	ttgattttctg	caagcggatg	acgaaaatc	accccggttt	1200
atattttcga	agatgcttca	tcgacgacat	cgactgcgtc	ctctatgaca	gcaacgaccg	1260
cgtcaccgac	gacggcatcg	atcacgtcat	ctacgatgac	aatgactgca	tcaactccgc	1320
atcactattg	tgatgactgt	tacacggcgt	agaagaacca	accaaagtgg	tggcttcac	1380
gccaacgacg	tcctctaaca	tatgcaagac	gtccccaatg	gcacccctctg	acatctacaa	1440
ggtgcaagat	gctaacaatt	acagtttttg	tcttcacact	gtggcataaa	tatttttttt	1500
caccttcggc	tatatgcggc	tacacctaca	accacggtta	ctacatgatc	ggctccatca	1560
acgaacatct	ataacaacaa	tcattgatgg	aaactctagt	caaagcgtct	gtgtcatcgc	1620

tatcatccat	gacactcccg	ctatgactac	gtgaggggaat	agataagagt	caagggacga	1680
cacggaagga	gacgtaggca	ccaggtggag	gaccatccat	caaagatgca	attgatgatg	1740
gtgagttgaa	gaagatgaag	aaataaaata	tttcaaatcc	agtcgcaatc	attcgcttcg	1800
ctcccgttac	gactgagggg	gaatgttaga	agcatagata	tattaattgg	agataagagt	1860
catacaaata	tagagataag	atatcatcct	agagatagaa	tcctagagat	aaaatatagt	1920
cctagagata	aatctactct	tacttgtacc	cctatatata	ccccatgaga	ggatcaatgc	1980
aatacaccca	gaatacaaca	attagatfff	tctacggttg	taactataat	acgctgtaat	2040
atgctggatc	ggggaagagc	gcccgtaatc	agtgccccag	agatgtaggt	ctcggttgaa	2100
ctccattatc	aaataccgta	cctcgggtgc	gtcatcatgt	ttgaatcttc	tatgacgttt	2160
cttttgcatt	cggttttcga	tgtgacttcg	gggctggttt	tataacaatg	attatagtgc	2220
tgttgacggc	aatcggttgt	gagaattagc	tattcgggtc	cctccatgtg	atftttcttgt	2280
gattgggatg	tatggtaatg	ctagggtttt	aagggtgtagg	attgggtgcat	gagagatcat	2340
cacttcactt	gtatgacctt	ctctcctttt	atattttttt	atcattctct	cctttttttt	2400
ataatgctac	tgaactagtg	gaatacaggg	gactaatgca	aaataaaaga	aaagtatcac	2460
tggtcacggc	atataattta	gaaagtgtgt	gatttaggca	tagggctgac	catgaccctt	2520
tacgacttgg	tcgctcgggt	tgttagacga	tagatcaacc	aacaaaagct	acgatacatg	2580
atgtacgtgt	caggatacaa	atccttacia	ataacaacag	ttattgttcg	ataactatca	2640
gttgtctagg	cttaccaatg	tataatagaa	gatgaaaatt	ccatattact	ggtatcgttc	2700
aatgctagta	actctttgag	ctttgtctag	gttaaaaaaa	aaattatgga	tccaccatca	2760
caaaaatgaa	aacaccggg	gaaaacaaaa	aaccatttga	tagcagcaca	agacaaaatg	2820
atgttaccgt	ctaccggagc	tctactccg	taccagcaca	accaaacgaa	cagtaccgc	2880
cggaccaggg	gcacgttcgt	aaatttccct	cccgtggctg	gctggctgcc	atctctctca	2940
accagggttg	gtaatttcgg	ccgtttcgg	gggtcccgat	agtaaatgag	ctccgggtcaa	3000
aacgccctcc	gcctcccctc	attgcgccgc	acgcacaccg	catctagatc	cagatcgaaa	3060
aaatcgctat	ctcgccgagt	cgccagtcac	cgctcgcagc	ccggtcgccg	taccgccggc	3120
gctgcacgcc	cccctccaag	cgctcgcgcc	atcgccccca	gccgccaggt	ggtggggcgg	3180
cggatgccga	gcttggcgag	gcttggcgag	acgaaccagg	cgaggaggac	gaggatcttg	3240
tcgacgagcc	agagcgggag	ccacgccatg	agcaacacgg	cgagctcgaa	cgtggacttg	3300
ccgagcacct	cgccagggag	gacgtggacg	gcgtcgcgca	ccaccatcgc	cgggagggcg	3360

ctgtggtcgc	acaggtcgag	cgacaccacc	atgccggagt	tgccgcaccc	gacgacgagc	3420
accttcttgc	cgcggtacgc	ctcgccggac	ttgtagaccg	cgacatgcat	cacctcgctg	3480
ctatatattgt	tcttggactg	tggagacttg	ctgtcagtgg	gtgtgttcag	aattgctgct	3540
gcagcttgca	gccaatttgt	gatgcagcag	ctgcagcttg	tatggctgcc	gagtagagcg	3600
agtgttgcta	tctgtttttg	ttctcttttt	cagaaatttc	gcccgcaa	tttaaatttg	3660
aattcaaatt	tttaaagaa	ctagaaaata	tgcccgtgcg	ttgcaccggg	tgaatatcaa	3720
acaaatattg	atgggtaaga	ttgcttgtgt	acttataaca	catatgcaca	aaaatattga	3780
atatgtacat	acctcgcaa	tatctccaaa	ttttatacat	atgagttgtg	taaatcatgt	3840
gagttccata	ttgtcatggt	aatatggagt	attactgatg	agcccatcta	tggtgataat	3900
tttggagggt	gtagctcaac	gaatttgtat	ttgctatgta	tctcaacggt	gataagtcac	3960
tactacaacc	atcggcgacc	tttctcggga	tccaagcatg	tcgaccccg	caacgtggcg	4020
tcggtgcagg	gcaccgagat	gaacaccacg	gggctatttg	cctgtccagg	gtcatcctag	4080
gcttaaggcc	acgacactca	aggacgtggt	aggcggcgtc	acagagggtg	tcccagcgaa	4140
caagctggcc	accaaggagg	acgccgacaa	ggtggcggcc	accgctatgc	agaaacgatg	4200
ggaggcatgc	cggtgacgac	aaggagctaa	cacgatccat	ttagtcccga	tccgagttta	4260
tcaggaattc	aatcctgcac	cgtgcgggta	cgtttttctt	ttccgcgga	aaagcaatca	4320
ccgatggtag	ggacaaagtg	cgtgtgagaa	cagaggccag	gccaaagtgc	gtgcgagaac	4380
ggaggctagg	ccatcgctgg	attggattta	cgaatgaaat	atcgatgtga	cgaacagaaa	4440
attatcagtt	tgatttaatt	ttcataatca	gaactcttta	ataggaaaaa	aattacatgt	4500
acgttccttc	atcgtgceca	tgtccatctg	ggagtccagg	tttattcaca	aagacccaat	4560
caacagccag	gaatccatgt	ccttccccgc	cgttccctac	tctgcttttt	tttctttcat	4620
ttgaaacctt	ccgctatgaa	tttctagtcg	ttcctagcat	ccacgcacac	aaaatagatt	4680
tcctcgcaa	ggcaaaacat	acaaatatga	gtgcatgcaa	gatattacaa	acccaatcca	4740
ttaaaaatag	aaaataatta	actttagcct	acctatctca	atattggtat	atgcccaaac	4800
tcaaaaggag	aaaaaccaa	ctaaaacttt	taataaagtg	aaccaagag	ataaaaagg	4860
gatagtaaca	acaaaatctc	acttgacaat	gtcgttaatc	aacctgttt	ttaaataatta	4920
cttaaaaatc	tttatattta	cctattaaaa	caatgaaaa	cagaagatgt	ttctttttta	4980
tttacaacag	cgttgtattt	agtcatgtcc	tatctaagag	agaaaaatga	atttaacgaa	5040
aagaagctca	gaaaaaaaa	gagaacaggg	ccaccacacc	agtaatccct	atgttatcaa	5100

tgaaaaaaaa	tttcaatgct	aggtttttta	taagaaaagg	tgataaagtg	ttgaaaaaat	5160
acagcaggaa	attatatatc	ttgctggttt	aacatgaatt	caagcatata	gatataaaaa	5220
tatatcaggc	taggaaagga	aaaggataaa	attggagaga	aaaaggaaaa	gaacagtaga	5280
ggataaccag	caaaaagatg	aaaggattcg	aacccatgac	ctagcgggtac	aattgtttca	5340
caggctaacc	aattgagaat	catcgacggt	gtgtcatctt	gtgtagctac	at ttgaaaaa	5400
atatgttttg	agctgaacgt	tgggtgtgtcc	gcccctgcat	ccgatacatg	ttggagcgtg	5460
gagcgcggta	aagaaaaaat	cctatcgaac	cttatctcct	tctctctcgt	cgctttctgc	5520
gtctccccgt	ctctccttcg	ccaacagccg	agaagaggca	gagagagcgc	cgccccccgt	5580
ccctctctct	ccctctcgtc	ctcgccccca	tccctctcgt	ctttcccttg	ccggcagcag	5640
aggaggcggc	agcgaacggt	tcagctgctc	ccacgggccg	gatcgggcag	tggcggtgcc	5700
gtcggcggct	tccgctggcg	aatccggcgg	gtggatacaa	atcagtgttc	cgataggtaa	5760
aaccctgctc	tcagcatctg	cccttttgaa	ttcgccaaga	gccagcatct	gcccttttga	5820
attcgccaag	ggccagcatc	tgcccatttg	at ttttgaatt	cgccaagagc	cagcaacagc	5880
gccccgcgc	cccctccctc	ctccgcaata	aacagccaca	cgcgcgcgcc	ccatgtccac	5940
cctcatcgcc	acagcgcacc	accaccacca	ccaccaccac	caccaccacc	gtctccagcc	6000
atg gcc tcc tcc gcc tcc ctg cag cgc ttc ctc ccc ccg tac ccc cac	6048					
Met Ala Ser Ser Ala Ser Leu Gln Arg Phe Leu Pro Pro Tyr Pro His						
1 5 10 15						
gcg gca gca tcc cgc tgc cgc cct ccc ggc gtc cgc gcc cgc ccc gtg	6096					
Ala Ala Ala Ser Arg Cys Arg Pro Pro Gly Val Arg Ala Arg Pro Val						
20 25 30						
cag tcg tcg acg gtg tcc gca ccg tcc tcc tcg act ccg gcg gcg gac	6144					
Gln Ser Ser Thr Val Ser Ala Pro Ser Ser Ser Thr Pro Ala Ala Asp						
35 40 45						
gag gcc gtg tcg gcg gag ccg ctg gag ccg cgg gtg gag cag ccg gag	6192					
Glu Ala Val Ser Ala Glu Arg Leu Glu Pro Arg Val Glu Gln Arg Glu						
50 55 60						
ggc cgg tac tgg gtg ctc aag gag aag tac cgg acg ggg ctg aac ccg	6240					
Gly Arg Tyr Trp Val Leu Lys Glu Lys Tyr Arg Thr Gly Leu Asn Pro						
65 70 75 80						
cag gag aag gtg aag ctg ggg aag gag ccc atg tca ttg ttc atg gag	6288					
Gln Glu Lys Val Lys Leu Gly Lys Glu Pro Met Ser Leu Phe Met Glu						
85 90 95						
ggc ggc atc aag gag ctc gcc aag atg ccc atg gag gag atc gag gcc	6336					
Gly Gly Ile Lys Glu Leu Ala Lys Met Pro Met Glu Glu Ile Glu Ala						
100 105 110						

gac aag ctc tcc aag gag gac atc gac gtg cgg ctc aag tgg ctc ggc	6384
Asp Lys Leu Ser Lys Glu Asp Ile Asp Val Arg Leu Lys Trp Leu Gly	
115 120 125	
ctc ttc cac cgc cgc aag cat cag t gtatgcctct cttctcttgc	6429
Leu Phe His Arg Arg Lys His Gln	
130 135	
tcctctgatac aacacatttt cttgcttttcg ttcggttatt tgtcgcgcgcg aggaagttaa	6489
ttcgccaaga tattctgcag ttttttttct cgatgcacat tcagcaacct aattaagact	6549
gattaagttg ctgtgatttt tatagcttaa ttacggctctc gtgggtaatg actatttata	6609
ttgagtaaac atggttacct ttgatccaat cacttcacct ccatgtgccca tatatagcca	6669
caggctctac caagtaacac tagtaatatg cctgtgatac gccacgggtgg cataataaat	6729
cattaaattt tattataatc aaattaagga tcctaaaatt ggtccaattg ggtgttaatt	6789
cgatgcaggt catataaaaa tatattttag gcaaggtgca attcaagagc atcaaccatt	6849
atatccaatc actttaatat atatttgaag ataacatatg tcggaaaaaa aatgatggag	6909
agctatttca ttaacttggtg agcataaaca gatcaccaga tgatgccacc ataagtcccg	6969
ccacagtaag tgatgcagct catcttgccc taggcgtttcg gtctaaccag tagatagaaa	7029
gagtacaaca tagatcgaat gaaaaaaaa atctccagaa gaaagctcaa ccacattgag	7089
taaattagag caacaatcaa atcgagtcag catatcgтта tgttagcaga accaatcacc	7149
acaatttggt tctcctcttt atctaagtgt tttgccaggt taaaagcata tatcactatg	7209
ttccaagcaa acatcggcaa tggacatgtc aaaaataaat gatcaattgt ttctttgagt	7269
acaaaattga caatggacac tatgttctct tgttagaatt ctatttgtca gggtaggatg	7329
tagaaaaact taacttttag aggaagctta aatatccggc ataaacttgc tttttcagcg	7389
ctctataaaa taattcaaca gtgaattgtc catcttttct aagtgtcca aaagacacta	7449
agttgaaaaa ccaggtgaac caacagattg atccacaaaa tcttattatt agattattca	7509
cttaaaagcc tgtctttatt tcaaacatat aaaaacagaa gttattaatc agggaagcgc	7569
ttatggcagc ctgagcgaac cagtgatagc aagtggtgaa aacagtaaat aggatacata	7629
aaaattatac aaggtttcta ctgtttatca aaaaaaata tttgaaaaca gtaaatagga	7689
tacataatcg acttccaact tgccttatc ataacatcca gaatcacaac aagaattgca	7749
acgaatacat agtcgacttg agctaagaag tcacaagacc tgtcaaagta agctgccctt	7809
gatcttgaag tgaaaggcat attttattgt cttccttggc aacagatat cactgtcttc	7869
agcagttcag ttagataatc caagatttct cacggagaag agcatatcac tcgcatcagt	7929

gttgtgccct	caaataactg	agataaactg	aatthttgttc	tctttgaagc	atctgcaggc	7989
attaacaatt	ataatacttt	acaaagtthc	attgggtcta	aactattggt	tgcacatcat	8049
atatatgccc	agaactthtt	agcatgatac	aagggtcctg	ttcataactc	atgcctaaat	8109
ctgacaaatt	tgtcaaacga	caatataagt	cgaattataa	tgcgttttag	aattgacgcc	8169
aaaactthttg	ctagcgttaag	taactcttcc	acctcccagc	atgcatacaa	ccaacaagct	8229
aaactthttgt	tcaaaaaaat	gtacatttat	ttccttgaac	acagcctttg	tagaatatga	8289
ttaaaaactc	atggatgaat	gaaataatgt	aaaagaatgg	tcaaatgat	gaatagtaca	8349
agaagcaact	gtgaacattt	cacctttacc	tgactgttcg	caagaaggcc	acgtggcaga	8409
aaagccagaa	atgcaagaag	cttcctaat	tgatacacca	tcaagaaatc	aatggactca	8469
acaccagcgt	ctgcccagac	aaaatgaatg	caggcaccta	aatatagaa	ccattgactt	8529
ttcaacactg	aattatataa	cctgaatatc	ttgtthtttt	aacacatctg	acaaaatcag	8589
tgcattctgt	tccatataga	tgtatgcata	gctcccatat	gttagttgat	cgatgagcat	8649
gcaaactata	cacaccttac	gttactccct	ctgtcaaaaa	aatataagc	ttgtctagat	8709
acatagctac	aatgcttat	atthttggat	tctcttaaag	ctgtagaaac	thttatcgcc	8769
ccgccatggc	aagtcgagat	gccatcccca	atgaaagccc	ccacacaggt	ttcatgcctt	8829
gctgcacaat	attgagcaac	caaaaatata	ataatatttg	tgtcagaatt	tgaatcaacc	8889
ttacagatac	tgggtggcca	gaaaatctag	tccaagtaat	atcctgaaaa	atagcaactg	8949
gcaaatacta	aaggcagtga	agagthttct	ttagatcaga	tgataaaaaa	aatcatatg	9009
ttcaatagca	ataatcactc	acathttthtt	tgctgttttag	aathtagata	attagtagtt	9069
aaactthctat	agcttgcgta	gctaagatca	atggtgatta	ttagttgaaa	aaataatcaa	9129
atcatcaaac	tgaggagact	tatacctgcc	ataagthctg	aaaththcaat	gatcctagtc	9189
aataththact	gtatatatag	aathaggtcc	aaaagatgat	acttacaatt	aaggatgthg	9249
tattgatcgg	ttcataactc	aagctthctat	ttatcattaa	tcaaaagctg	gatcattcat	9309
gcatatacct	ttgccgcact	caacgtagca	gctcggagtc	thctthgttc	agaagcgagg	9369
aaggagtcaa	caaataagta	ctgcaatgth	aaacaaaccg	acatatcaaa	tcccaaatta	9429
agaatgcatg	atthattaat	acaggaaata	tatgatcaag	tcccaaaaag	tgagtcatgt	9489
tatgtacact	cagtcatcaa	thtcaataag	aataththact	tgctcattgg	tatatggatt	9549
tgattatgac	ataaththgac	aatacattta	cagaataaac	thgcagtgct	gtgagcatat	9609
gttactaaca	tgtaaggacc	thgtththgct	ctgtthcaata	ctcatgtthga	thcttgatctg	9669

tgtccacata tacctaaatg aatgaaatc aaagaatgag gttttagga gtggagttgg	9729
tgaattatag ggtagataat gtcggcacia cggtttgata agtagtacga gtactttatt	9789
tggcgccacc gcgccagcat cagatgtgtg gcctttgcac tgattgaatc caaaagaaaa	9849
aaaaagtcgt tttgggtccca cacaattcta cttcatctgc aggatgtaca gaaggttaca	9909
tatctattct gttctatgct ctgtttacat ttatatttat agtactaggt tgaaagggct	9969
cacttggtgg ctgtcattgg ttggctgggt cggtatatta ctaatagggt ttttaatggc	10029
atatatgttc ttaaaataaa ccagaaaagc aaaagatcaa ctatcttagc cacaccaatg	10089
aatggaata tactgaactg tcacggctaa aattctcttc agtcacctgg cccaactgga	10149
gccgtgggct cgtcgtcttt tctaaacatg tactagtatt ttgggggccc acagtgaatt	10209
tggcccaaaa tgctgacagc cgctctacgg ctctacgctg tgcag at ggg cgg ttc	10265
	Tyr Gly Arg Phe 140
atg atg cgg ctg aag ctg cca aac ggt gtg acg acg agc gag cag acg	10313
Met Met Arg Leu Lys Leu Pro Asn Gly Val Thr Thr Ser Glu Gln Thr	
	145 150 155
agg tac ctg gcg agc gtg atc gag gcg tac ggc aag gag ggc tgc gcc	10361
Arg Tyr Leu Ala Ser Val Ile Glu Ala Tyr Gly Lys Glu Gly Cys Ala	
	160 165 170
gac gtg aca acc cgc cag aac tgg cag atc cgc ggc gtc acg ctc ccc	10409
Asp Val Thr Thr Arg Gln Asn Trp Gln Ile Arg Gly Val Thr Leu Pro	
	175 180 185
gac gtg ccg gcc atc ctc gac ggg ctc aac gcc gtc ggc ctc acc agc	10457
Asp Val Pro Ala Ile Leu Asp Gly Leu Asn Ala Val Gly Leu Thr Ser	
	190 195 200
ctc cag agc ggc atg gac aac gtc cgc aac ccc gtc ggc aac ccg ctc	10505
Leu Gln Ser Gly Met Asp Asn Val Arg Asn Pro Val Gly Asn Pro Leu	
	205 210 215 220
gcc ggc atc gac ccc gac gag atc gtc gac acg cga tcc tac acc aac	10553
Ala Gly Ile Asp Pro Asp Glu Ile Val Asp Thr Arg Ser Tyr Thr Asn	
	225 230 235
ctc ctc tcc tcc tac atc acc agc aac ttc cag ggc aac ccc acc atc	10601
Leu Leu Ser Ser Tyr Ile Thr Ser Asn Phe Gln Gly Asn Pro Thr Ile	
	240 245 250
acc aac ct gtgagtgatc gaatcaaatt gatcatgctc tgtgctgtgc	10649
Thr Asn Leu	
tgtttcgtgt cgtctctgac gacatgtttg ttgaatttgt tgttgctgcg tgctgttggc	10709
ag g ccg agg aag tgg aac gtg tgc gtg atc ggg tcg cac gat ctg tac	10757
Pro Arg Lys Trp Asn Val Cys Val Ile Gly Ser His Asp Leu Tyr	

					260						265						270	
gag	cac	cca	cac	atc	aac	gac	ctc	gcg	tac	atg	ccg	gcg	gtg	aag	ggc	10805		
Glu	His	Pro	His	Ile	Asn	Asp	Leu	Ala	Tyr	Met	Pro	Ala	Val	Lys	Gly			
				275					280					285				
ggc	aag	ttc	ggg	ttc	aac	ctc	ctc	gtc	ggc	ggg	ttc	ata	agc	ccc	aag	10853		
Gly	Lys	Phe	Gly	Phe	Asn	Leu	Leu	Val	Gly	Gly	Phe	Ile	Ser	Pro	Lys			
			290				295				300							
agg	tgg	gag	gag	gcg	ctg	ccg	ctc	gac	gcc	tgg	gtc	ccc	ggc	gac	gac	10901		
Arg	Trp	Glu	Glu	Ala	Leu	Pro	Leu	Asp	Ala	Trp	Val	Pro	Gly	Asp	Asp			
		305					310				315							
atc	atc	ccg	gtg	tgc	aag	gcc	ggt	ctc	gag	gcg	tac	cgc	gac	ctc	ggc	10949		
Ile	Ile	Pro	Val	Cys	Lys	Ala	Val	Leu	Glu	Ala	Tyr	Arg	Asp	Leu	Gly			
		320				325				330								
acc	agg	ggc	aac	cgc	cag	aag	acc	cgc	atg	atg	tgg	ctc	atc	gac	gaa	10997		
Thr	Arg	Gly	Asn	Arg	Gln	Lys	Thr	Arg	Met	Met	Trp	Leu	Ile	Asp	Glu			
335					340					345				350				
ctt	gtgaaccatt	tttttctcca	ttcatccacg	ccattgactg	aattacgtat	11050												
Leu																		
gtcccaatgt	tcttatcagt	taattgocggt	gttggcattg	cag	gga	atg	gag	gct	11105									
								Gly	Met	Glu	Ala							
											355							
ttt	cgg	tcg	gag	gtg	gag	aag	agg	atg	ccg	aac	ggc	gtg	ctg	gag	cgc	11153		
Phe	Arg	Ser	Glu	Val	Glu	Lys	Arg	Met	Pro	Asn	Gly	Val	Leu	Glu	Arg			
				360					365					370				
gcg	gcg	ccg	gag	gac	ctc	atc	gac	aag	aaa	tgg	cag	agg	agg	gac	tac	11201		
Ala	Ala	Pro	Glu	Asp	Leu	Ile	Asp	Lys	Lys	Trp	Gln	Arg	Arg	Asp	Tyr			
			375					380					385					
ctc	ggc	gtg	cac	ccg	cag	aag	cag	gaa	ggg	atg	tcc	tac	gtc	ggc	ctg	11249		
Leu	Gly	Val	His	Pro	Gln	Lys	Gln	Glu	Gly	Met	Ser	Tyr	Val	Gly	Leu			
			390					395				400						
cac	gtg	ccc	gtc	ggc	cgg	gtg	cag	gcg	gcg	gac	atg	ttc	gag	ctc	gca	11297		
His	Val	Pro	Val	Gly	Arg	Val	Gln	Ala	Ala	Asp	Met	Phe	Glu	Leu	Ala			
		405					410					415						
cgc	ctc	gcc	gac	gag	tac	ggc	tcc	ggc	gag	ctc	cgc	ctc	acc	gtg	gag	11345		
Arg	Leu	Ala	Asp	Glu	Tyr	Gly	Ser	Gly	Glu	Leu	Arg	Leu	Thr	Val	Glu			
420					425					430				435				
cag	aac	atc	gtg	atc	ccg	aac	gtc	aag	aac	gag	aag	gtg	gag	gcg	ctg	11393		
Gln	Asn	Ile	Val	Ile	Pro	Asn	Val	Lys	Asn	Glu	Lys	Val	Glu	Ala	Leu			
				440					445					450				
ctc	tcc	gag	ccg	ctg	ctt	cag	aag	tcc	tcc	ccg	cag	ccg	tcg	ctg	ctg	11441		
Leu	Ser	Glu	Pro	Leu	Leu	Gln	Lys	Phe	Ser	Pro	Gln	Pro	Ser	Leu	Leu			
			455					460					465					

```

ctc aag ggc ctc gtc gcg tgc acc ggc aac cag ttc tgc ggc cag gcc      11489
Leu Lys Gly Leu Val Ala Cys Thr Gly Asn Gln Phe Cys Gly Gln Ala
      470                      475                      480

atc atc gag acg aag cag cgg gcg ctg ctg gtg acg tcg cag gtg gag      11537
Ile Ile Glu Thr Lys Gln Arg Ala Leu Leu Val Thr Ser Gln Val Glu
      485                      490                      495

aag ctc gtg tcg gtg ccc cgg gcg gtg cgg atg cac tgg acc ggc tgc      11585
Lys Leu Val Ser Val Pro Arg Ala Val Arg Met His Trp Thr Gly Cys
      500                      505                      510                      515

ccc aac agc tgc ggc cag gtg cag gtc gcc gac atc ggc ttc atg ggc      11633
Pro Asn Ser Cys Gly Gln Val Gln Val Ala Asp Ile Gly Phe Met Gly
      520                      525                      530

tgc ctc acc aag gac agc gcc ggc aag atc gtt gag gcg gcc gac atc      11681
Cys Leu Thr Lys Asp Ser Ala Gly Lys Ile Val Glu Ala Ala Asp Ile
      535                      540                      545

ttc gtc ggc ggc cgc gtc ggc agc gac tcg cac ctc gcc ggc gcg tac      11729
Phe Val Gly Gly Arg Val Gly Ser Asp Ser His Leu Ala Gly Ala Tyr
      550                      555                      560

aag aag tcc gtg ccg tgc gac gag ctg gcg ccg atc gtc gcc gac atc      11777
Lys Lys Ser Val Pro Cys Asp Glu Leu Ala Pro Ile Val Ala Asp Ile
      565                      570                      575

ctg gtc gag cgg ttc ggg gcc gtg cgg agg gag agg gag gag gac gag      11825
Leu Val Glu Arg Phe Gly Ala Val Arg Arg Glu Arg Glu Glu Asp Glu
      580                      585                      590                      595

gag tag gaacacagac tgggggtgttt tgcttgctcc ggtgatctct cgccgtcctt      11881
Glu

gtaaagtaga cgacaatatg ccttcgcca tggcacgctt gtactgtcac gttttggttt      11941

gatctttag cccaaaagtt gtgttcattc tcgttacagt cttacagagg atgattgatt      12001

gataaataaa gaagaaacag attctgcaac tgttcatcgc tgttcctaaa tctgatttag      12061

cgaaagtatc ttgctgacc tgtcccaatc gcagtgctaa aaccatataa tcttgcaagc      12121

aatgaaatt gaaagagttc aatgcaacca ctaacagtct aacaacatga taaggcct      12179

```

```

<210> 5
<211> 2508
<212> DNA
<213> Oryza sativa

<220>
<221> CDS
<222> (519)..(2309)
<223>

<400> 5

```

```

tatcgaacct tatctccttc tctctcgtcg ctttctgctt ctccccgtct ctccctcgcc 60
aacagccgag aagaggcaga gagagcgccg cccccgtcc ctctctctcc ctctcgtctt 120
cgccccatc cctctcgtct tcccttgcc ggcagcagag gaggcggcag cgacggcttc 180
agetgctccc acgggcccga tggggcagtg ggggtggcgt cggcggcttc cgctggcgaa 240
tccggcgggt ggatacaaat cagtgttccg ataggtaaaa cctgctctc agcatctgcc 300
cttttgaatt cgccaagagc cagcatctgc ctttttgaat tcgccaaggg ccagcatctg 360
cccatttgat tttgaattcg ccaagagcca gcaacagcgc ccccgcgccc cctccctcct 420
ccgcaataaa cagccacacg cgccgcccc atgtccacc tcatcgccac agcgcaccac 480
caccaccacc accaccacca ccaccaccgt ctccagcc atg gcc tcc tcc gcc tcc 536
                                   Met Ala Ser Ser Ala Ser
                                   1 5
ctg cag cgc ttc ctc ccc ccg tac ccc cac gcg gca gca tcc cgc tgc 584
Leu Gln Arg Phe Leu Pro Pro Tyr Pro His Ala Ala Ala Ser Arg Cys
                                   10 15 20
cgc cct ccc ggc gtc cgc gcc cgc ccc gtg cag tgc tgc acg gtg tcc 632
Arg Pro Pro Gly Val Arg Ala Arg Pro Val Gln Ser Ser Thr Val Ser
                                   25 30 35
gca ccg tcc tcc tgc act ccg gcg gcg gac gag gcc gtg tgc gcg gag 680
Ala Pro Ser Ser Ser Thr Pro Ala Ala Asp Glu Ala Val Ser Ala Glu
                                   40 45 50
cgg ctg gag ccg cgg gtg gag cag cgg gag ggc cgg tac tgg gtg ctc 728
Arg Leu Glu Pro Arg Val Glu Gln Arg Glu Gly Arg Tyr Trp Val Leu
                                   55 60 65 70
aag gag aag tac ccg acg ggg ctg aac ccg cag gag aag gtg aag ctg 776
Lys Glu Lys Tyr Arg Thr Gly Leu Asn Pro Gln Glu Lys Val Lys Leu
                                   75 80 85
ggg aag gag ccc atg tca ttg ttc atg gag ggc gcc atc aag gag ctc 824
Gly Lys Glu Pro Met Ser Leu Phe Met Glu Gly Gly Ile Lys Glu Leu
                                   90 95 100
gcc aag atg ccc atg gag gag atc gag gcc gac aag ctc tcc aag gag 872
Ala Lys Met Pro Met Glu Glu Ile Glu Ala Asp Lys Leu Ser Lys Glu
                                   105 110 115
gac atc gac gtg ccg ctc aag tgg ctc ggc ctc ttc cac cgc cgc aag 920
Asp Ile Asp Val Arg Leu Lys Trp Leu Gly Leu Phe His Arg Arg Lys
                                   120 125 130
cat cag tat ggg ccg ttc atg atg ccg ctg aag ctg cca aac ggt gtg 968
His Gln Tyr Gly Arg Phe Met Met Arg Leu Lys Leu Pro Asn Gly Val
                                   135 140 145 150
acg acg agc gag cag acg agg tac ctg gcg agc gtg atc gag gcg tac 1016
Thr Thr Ser Glu Gln Thr Arg Tyr Leu Ala Ser Val Ile Glu Ala Tyr

```


155						160						165					
ggc aag gag ggc tgc gcc gac gtg aca acc cgc cag aac tgg cag atc																	1064
Gly Lys Glu Gly Cys Ala Asp Val Thr Thr Arg Gln Asn Trp Gln Ile																	
			170					175					180				
cgc ggc gtc acg ctc ccc gac gtg ccg gcc atc ctc gac ggg ctc aac																	1112
Arg Gly Val Thr Leu Pro Asp Val Pro Ala Ile Leu Asp Gly Leu Asn																	
		185					190					195					
gcc gtc ggc ctc acc agc ctc cag agc ggc atg gac aac gtc cgc aac																	1160
Ala Val Gly Leu Thr Ser Leu Gln Ser Gly Met Asp Asn Val Arg Asn																	
	200					205					210						
ccc gtc ggc aac ccg ctc gcc ggc atc gac ccc gac gag atc gtc gac																	1208
Pro Val Gly Asn Pro Leu Ala Gly Ile Asp Pro Asp Glu Ile Val Asp																	
	215				220					225						230	
acg cga tcc tac acc aac ctc ctc tcc tcc tac atc acc agc aac ttc																	1256
Thr Arg Ser Tyr Thr Asn Leu Leu Ser Ser Tyr Ile Thr Ser Asn Phe																	
			235						240						245		
cag ggc aac ccc acc atc acc aac ctg ccg agg aag tgg aac gtg tgc																	1304
Gln Gly Asn Pro Thr Ile Thr Asn Leu Pro Arg Lys Trp Asn Val Cys																	
		250					255						260				
gtg atc ggg tcg cac gat ctg tac gag cac cca cac atc aac gac ctc																	1352
Val Ile Gly Ser His Asp Leu Tyr Glu His Pro His Ile Asn Asp Leu																	
	265						270						275				
gcg tac atg ccg gcg gtg aag ggc ggc aag ttc ggg ttc aac ctc ctc																	1400
Ala Tyr Met Pro Ala Val Lys Gly Gly Lys Phe Gly Phe Asn Leu Leu																	
	280					285						290					
gtc ggc ggg ttc ata agc ccc aag agg tgg gag gag gcg ctg ccg ctc																	1448
Val Gly Gly Phe Ile Ser Pro Lys Arg Trp Glu Glu Ala Leu Pro Leu																	
	295				300					305						310	
gac gcc tgg gtc ccc ggc gac gac atc atc ccg gtg tgc aag gcc gtt																	1496
Asp Ala Trp Val Pro Gly Asp Asp Ile Ile Pro Val Cys Lys Ala Val																	
			315						320						325		
ctc gag gcg tac cgc gac ctc ggc acc agg ggc aac cgc cag aag acc																	1544
Leu Glu Ala Tyr Arg Asp Leu Gly Thr Arg Gly Asn Arg Gln Lys Thr																	
		330					335						340				
cgc atg atg tgg ctc atc gac gaa ctt gga atg gag gct ttt cgg tcg																	1592
Arg Met Met Trp Leu Ile Asp Glu Leu Gly Met Glu Ala Phe Arg Ser																	
	345						350					355					
gag gtg gag aag agg atg ccg aac ggc gtg ctg gag cgc gcg gcg ccg																	1640
Glu Val Glu Lys Arg Met Pro Asn Gly Val Leu Glu Arg Ala Ala Pro																	
	360					365						370					
gag gac ctc atc gac aag aaa tgg cag agg agg gac tac ctc ggc gtg																	1688
Glu Asp Leu Ile Asp Lys Lys Trp Gln Arg Arg Asp Tyr Leu Gly Val																	
	375				380					385						390	

cac ccg cag aag cag gaa ggg atg tcc tac gtc ggc ctg cac gtg ccc	1736
His Pro Gln Lys Gln Glu Gly Met Ser Tyr Val Gly Leu His Val Pro	
395 400 405	
gtc ggc cgg gtg cag gcg gcg gac atg ttc gag ctc gca cgc ctc gcc	1784
Val Gly Arg Val Gln Ala Ala Asp Met Phe Glu Leu Ala Arg Leu Ala	
410 415 420	
gac gag tac ggc tcc ggc gag ctc cgc ctc acc gtg gag cag aac atc	1832
Asp Glu Tyr Gly Ser Gly Glu Leu Arg Leu Thr Val Glu Gln Asn Ile	
425 430 435	
gtg atc ccg aac gtc aag aac gag aag gtg gag gcg ctg ctc tcc gag	1880
Val Ile Pro Asn Val Lys Asn Glu Lys Val Glu Ala Leu Leu Ser Glu	
440 445 450	
ccg ctg ctt cag aag ttc tcc ccg cag ccg tcg ctg ctg ctc aag ggc	1928
Pro Leu Leu Gln Lys Phe Ser Pro Gln Pro Ser Leu Leu Leu Lys Gly	
455 460 465 470	
ctc gtc gcg tgc acc ggc aac cag ttc tgc ggc cag gcc atc atc gag	1976
Leu Val Ala Cys Thr Gly Asn Gln Phe Cys Gly Gln Ala Ile Ile Glu	
475 480 485	
acg aag cag cgg gcg ctg ctg gtg acg tcg cag gtg gag aag ctc gtg	2024
Thr Lys Gln Arg Ala Leu Leu Val Thr Ser Gln Val Glu Lys Leu Val	
490 495 500	
tcg gtg ccc cgg gcg gtg cgg atg cac tgg acc ggc tgc ccc aac agc	2072
Ser Val Pro Arg Ala Val Arg Met His Trp Thr Gly Cys Pro Asn Ser	
505 510 515	
tgc ggc cag gtg cag gtc gcc gac atc ggc ttc atg ggc tgc ctc acc	2120
Cys Gly Gln Val Gln Val Ala Asp Ile Gly Phe Met Gly Cys Leu Thr	
520 525 530	
aag gac agc gcc ggc aag atc gtt gag gcg gcc gac atc ttc gtc ggc	2168
Lys Asp Ser Ala Gly Lys Ile Val Glu Ala Ala Asp Ile Phe Val Gly	
535 540 545 550	
ggc cgc gtc ggc agc gac tcg cac ctc gcc ggc gcg tac aag aag tcc	2216
Gly Arg Val Gly Ser Asp Ser His Leu Ala Gly Ala Tyr Lys Lys Ser	
555 560 565	
gtg ccg tgc gac gag ctg gcg ccg atc gtc gcc gac atc ctg gtc gag	2264
Val Pro Cys Asp Glu Leu Ala Pro Ile Val Ala Asp Ile Leu Val Glu	
570 575 580	
cgg ttc ggg gcc gtg cgg agg gag agg gag gag gac gag gag tag	2309
Arg Phe Gly Ala Val Arg Arg Glu Arg Glu Glu Asp Glu Glu	
585 590 595	
gaacacagac tggggtgttt tgcttgctcc ggtgatctct cgccgtcctt gtaaagtaga	2369
cgacaatatg ccttcgceca tggcagcgtt gtactgtcac gttttggttt gatctttag	2429
cccaaaagtt gtgttcattc tcgttacagt cttacagagg atgattgatt gataaataaa	2489

gaagaaacag attctgcaa

2508

<210> 6

<211> 596

<212> PRT

<213> *Oryza sativa*

<400> 6

Met Ala Ser Ser Ala Ser Leu Gln Arg Phe Leu Pro Pro Tyr Pro His
 1 5 10 15

Ala Ala Ala Ser Arg Cys Arg Pro Pro Gly Val Arg Ala Arg Pro Val
 20 25 30

Gln Ser Ser Thr Val Ser Ala Pro Ser Ser Ser Thr Pro Ala Ala Asp
 35 40 45

Glu Ala Val Ser Ala Glu Arg Leu Glu Pro Arg Val Glu Gln Arg Glu
 50 55 60

Gly Arg Tyr Trp Val Leu Lys Glu Lys Tyr Arg Thr Gly Leu Asn Pro
 65 70 75 80

Gln Glu Lys Val Lys Leu Gly Lys Glu Pro Met Ser Leu Phe Met Glu
 85 90 95

Gly Gly Ile Lys Glu Leu Ala Lys Met Pro Met Glu Glu Ile Glu Ala
 100 105 110

Asp Lys Leu Ser Lys Glu Asp Ile Asp Val Arg Leu Lys Trp Leu Gly
 115 120 125

Leu Phe His Arg Arg Lys His Gln Tyr Gly Arg Phe Met Met Arg Leu
 130 135 140

Lys Leu Pro Asn Gly Val Thr Thr Ser Glu Gln Thr Arg Tyr Leu Ala
 145 150 155 160

Ser Val Ile Glu Ala Tyr Gly Lys Glu Gly Cys Ala Asp Val Thr Thr
 165 170 175

Arg Gln Asn Trp Gln Ile Arg Gly Val Thr Leu Pro Asp Val Pro Ala
 180 185 190

Ile Leu Asp Gly Leu Asn Ala Val Gly Leu Thr Ser Leu Gln Ser Gly
 195 200 205

Met Asp Asn Val Arg Asn Pro Val Gly Asn Pro Leu Ala Gly Ile Asp
 210 215 220

Pro Asp Glu Ile Val Asp Thr Arg Ser Tyr Thr Asn Leu Leu Ser Ser
 225 230 235 240

Tyr Ile Thr Ser Asn Phe Gln Gly Asn Pro Thr Ile Thr Asn Leu Pro
 245 250 255

Arg Lys Trp Asn Val Cys Val Ile Gly Ser His Asp Leu Tyr Glu His

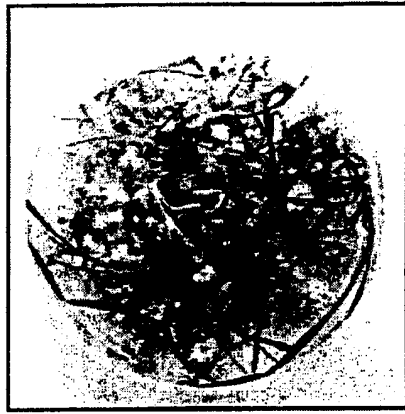
260					265					270					
Pro	His	Ile	Asn	Asp	Leu	Ala	Tyr	Met	Pro	Ala	Val	Lys	Gly	Gly	Lys
		275					280					285			
Phe	Gly	Phe	Asn	Leu	Leu	Val	Gly	Gly	Phe	Ile	Ser	Pro	Lys	Arg	Trp
	290					295					300				
Glu	Glu	Ala	Leu	Pro	Leu	Asp	Ala	Trp	Val	Pro	Gly	Asp	Asp	Ile	Ile
305					310					315					320
Pro	Val	Cys	Lys	Ala	Val	Leu	Glu	Ala	Tyr	Arg	Asp	Leu	Gly	Thr	Arg
				325					330					335	
Gly	Asn	Arg	Gln	Lys	Thr	Arg	Met	Met	Trp	Leu	Ile	Asp	Glu	Leu	Gly
			340					345					350		
Met	Glu	Ala	Phe	Arg	Ser	Glu	Val	Glu	Lys	Arg	Met	Pro	Asn	Gly	Val
		355					360					365			
Leu	Glu	Arg	Ala	Ala	Pro	Glu	Asp	Leu	Ile	Asp	Lys	Lys	Trp	Gln	Arg
	370					375					380				
Arg	Asp	Tyr	Leu	Gly	Val	His	Pro	Gln	Lys	Gln	Glu	Gly	Met	Ser	Tyr
385					390					395					400
Val	Gly	Leu	His	Val	Pro	Val	Gly	Arg	Val	Gln	Ala	Ala	Asp	Met	Phe
			405						410					415	
Glu	Leu	Ala	Arg	Leu	Ala	Asp	Glu	Tyr	Gly	Ser	Gly	Glu	Leu	Arg	Leu
			420					425					430		
Thr	Val	Glu	Gln	Asn	Ile	Val	Ile	Pro	Asn	Val	Lys	Asn	Glu	Lys	Val
		435					440					445			
Glu	Ala	Leu	Leu	Ser	Glu	Pro	Leu	Leu	Gln	Lys	Phe	Ser	Pro	Gln	Pro
	450					455					460				
Ser	Leu	Leu	Leu	Lys	Gly	Leu	Val	Ala	Cys	Thr	Gly	Asn	Gln	Phe	Cys
465				470						475					480
Gly	Gln	Ala	Ile	Ile	Glu	Thr	Lys	Gln	Arg	Ala	Leu	Leu	Val	Thr	Ser
			485					490						495	
Gln	Val	Glu	Lys	Leu	Val	Ser	Val	Pro	Arg	Ala	Val	Arg	Met	His	Trp
			500					505					510		
Thr	Gly	Cys	Pro	Asn	Ser	Cys	Gly	Gln	Val	Gln	Val	Ala	Asp	Ile	Gly
		515					520					525			
Phe	Met	Gly	Cys	Leu	Thr	Lys	Asp	Ser	Ala	Gly	Lys	Ile	Val	Glu	Ala
	530					535					540				
Ala	Asp	Ile	Phe	Val	Gly	Gly	Arg	Val	Gly	Ser	Asp	Ser	His	Leu	Ala
545					550					555					560
Gly	Ala	Tyr	Lys	Lys	Ser	Val	Pro	Cys	Asp	Glu	Leu	Ala	Pro	Ile	Val
				565					570					575	

Ala Asp Ile Leu Val Glu Arg Phe Gly Ala Val Arg Arg Glu Arg Glu
580 585 590

Glu Asp Glu Glu
595

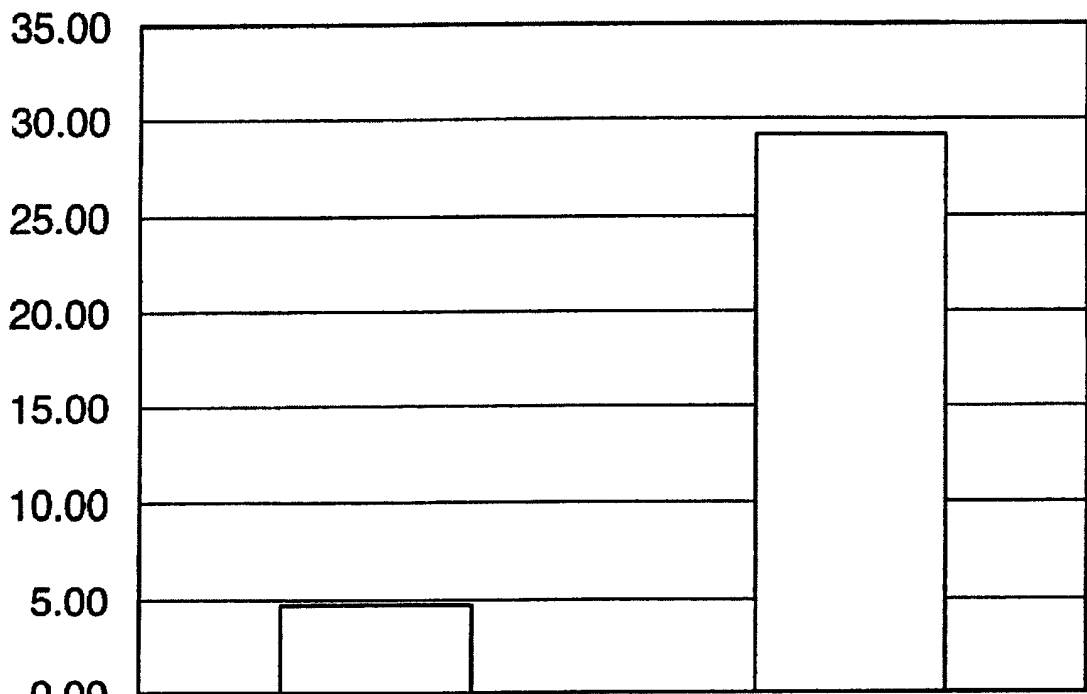


KOSHIHIKARI



KASALATH

苗的数量/愈伤组织重量(g)



KOSHIHIKARI

KASALATH

图1

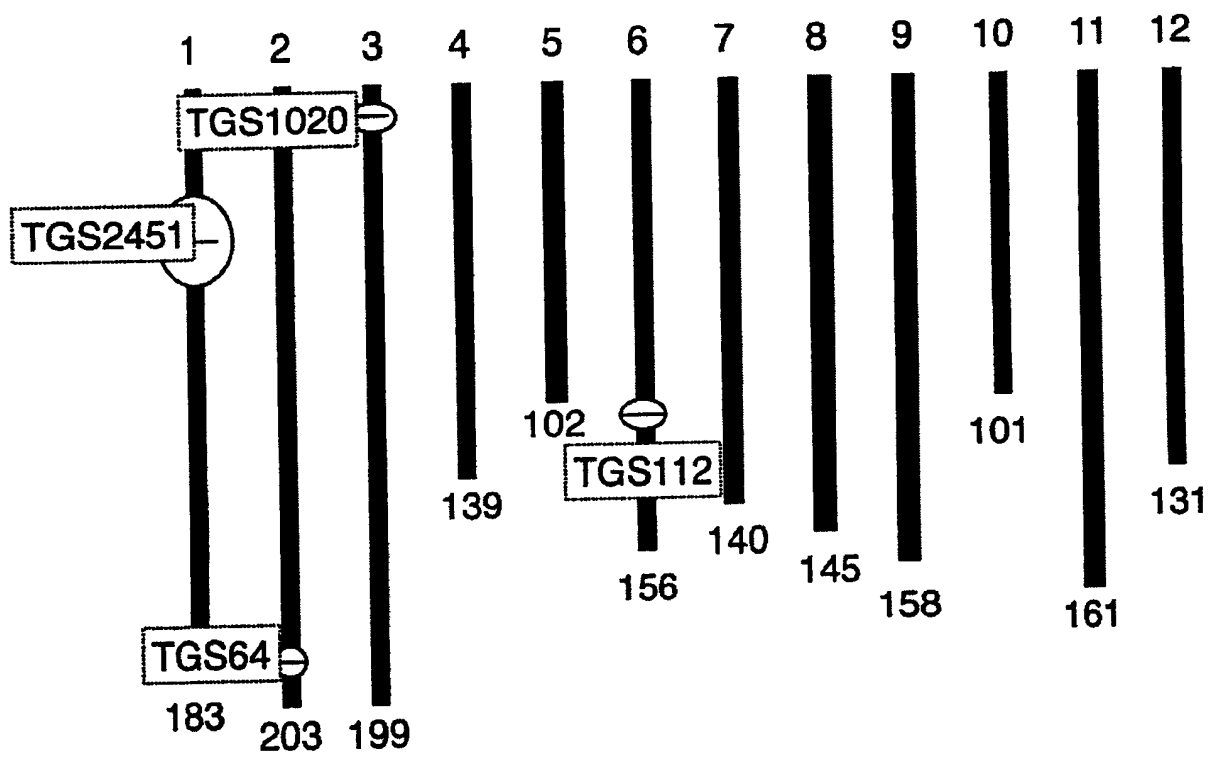
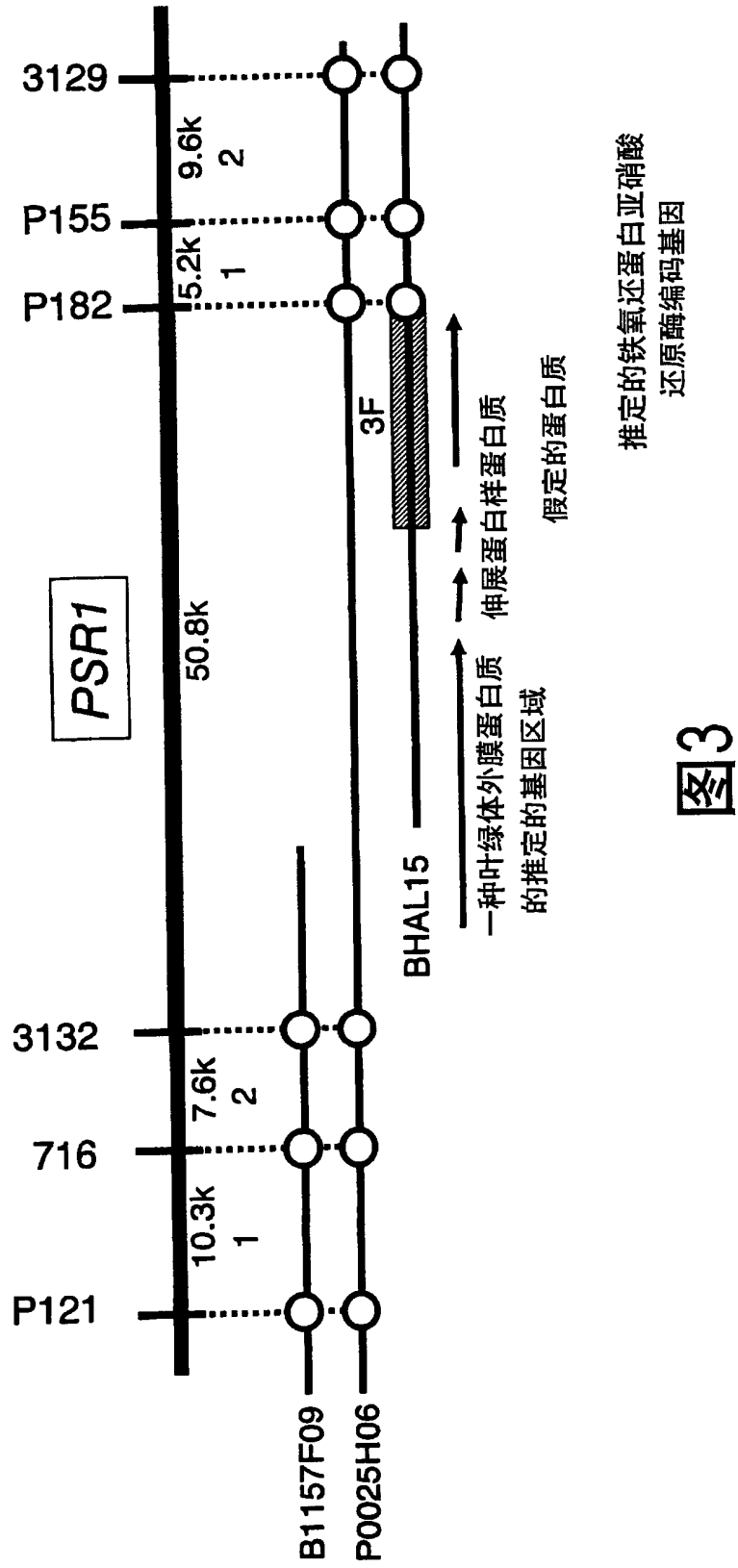
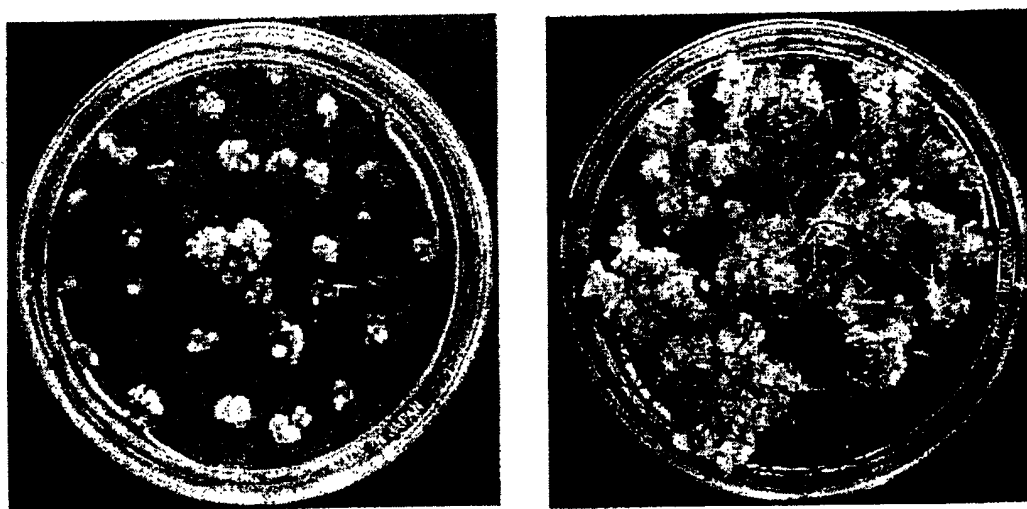


图2





载体

3F

图4

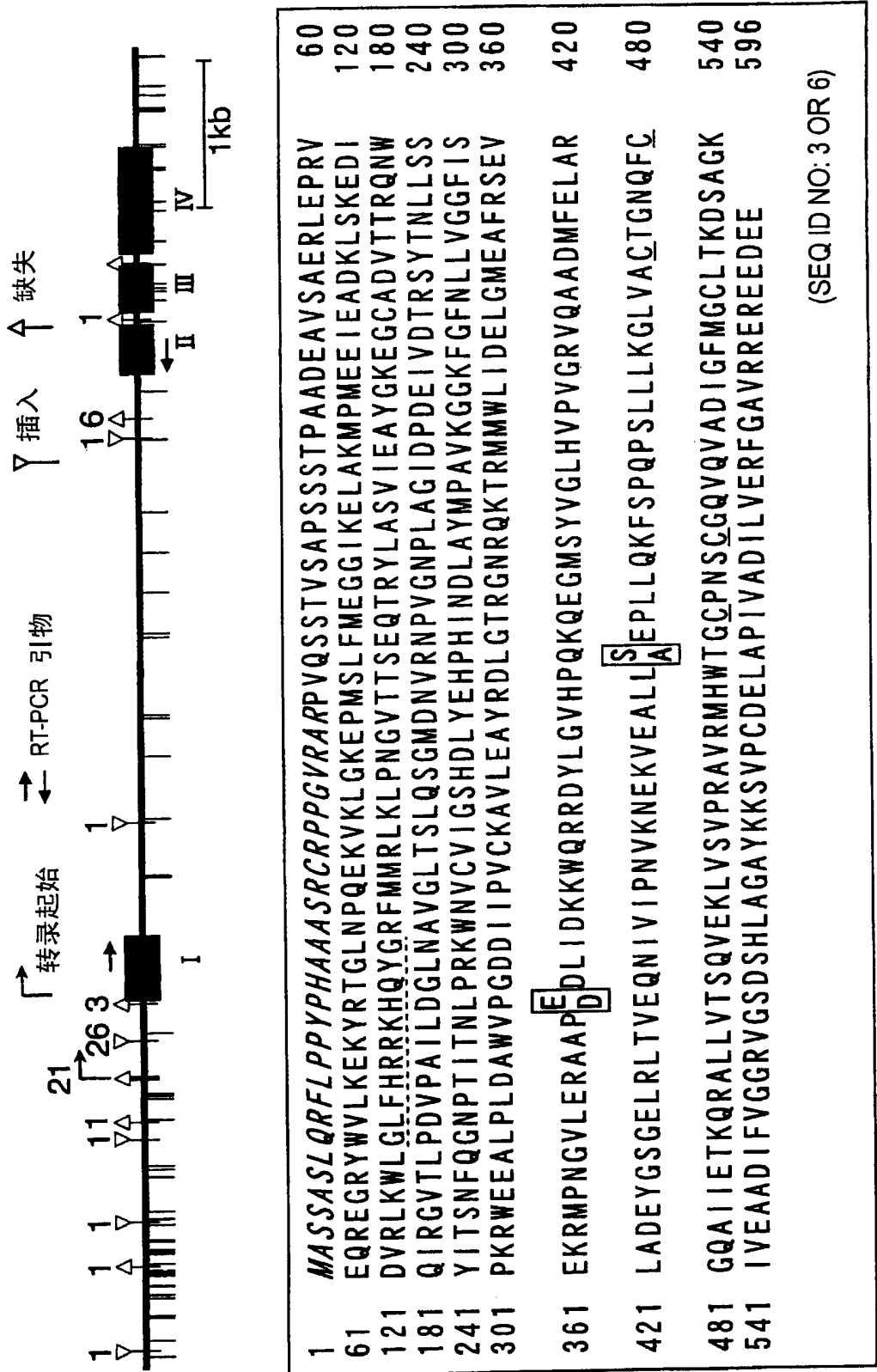


图5

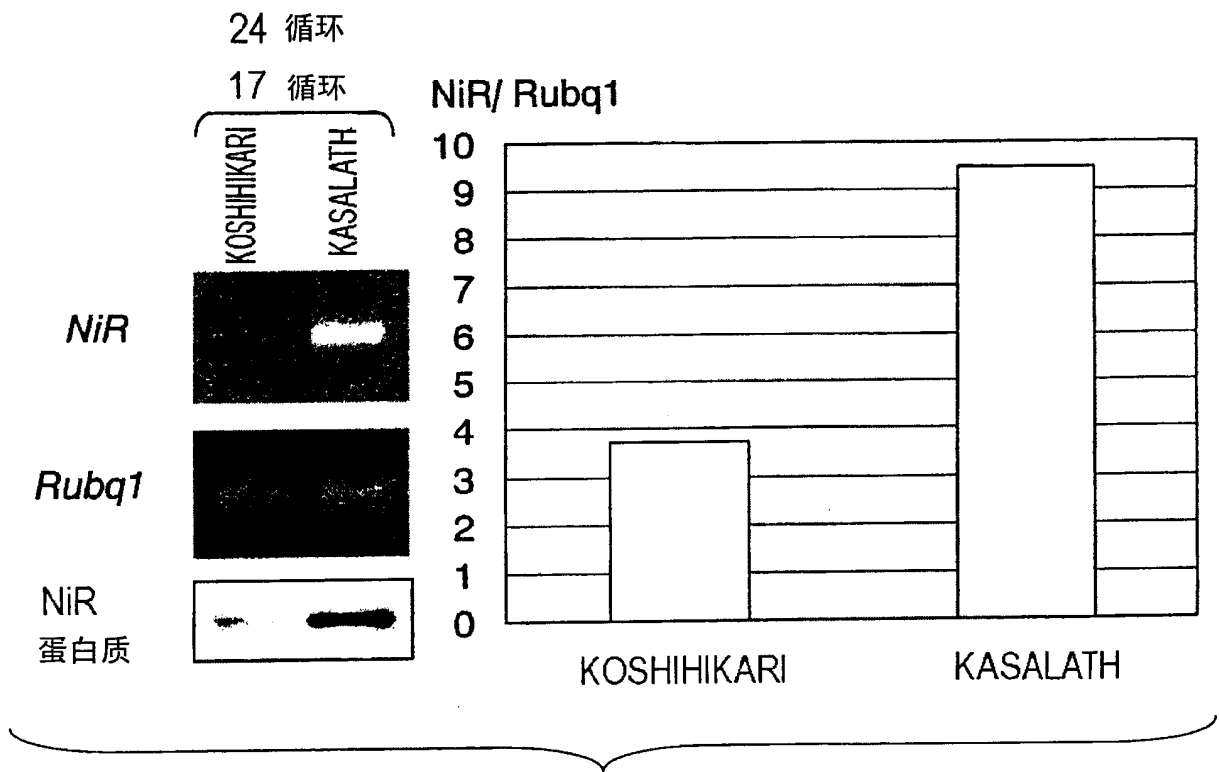


图6

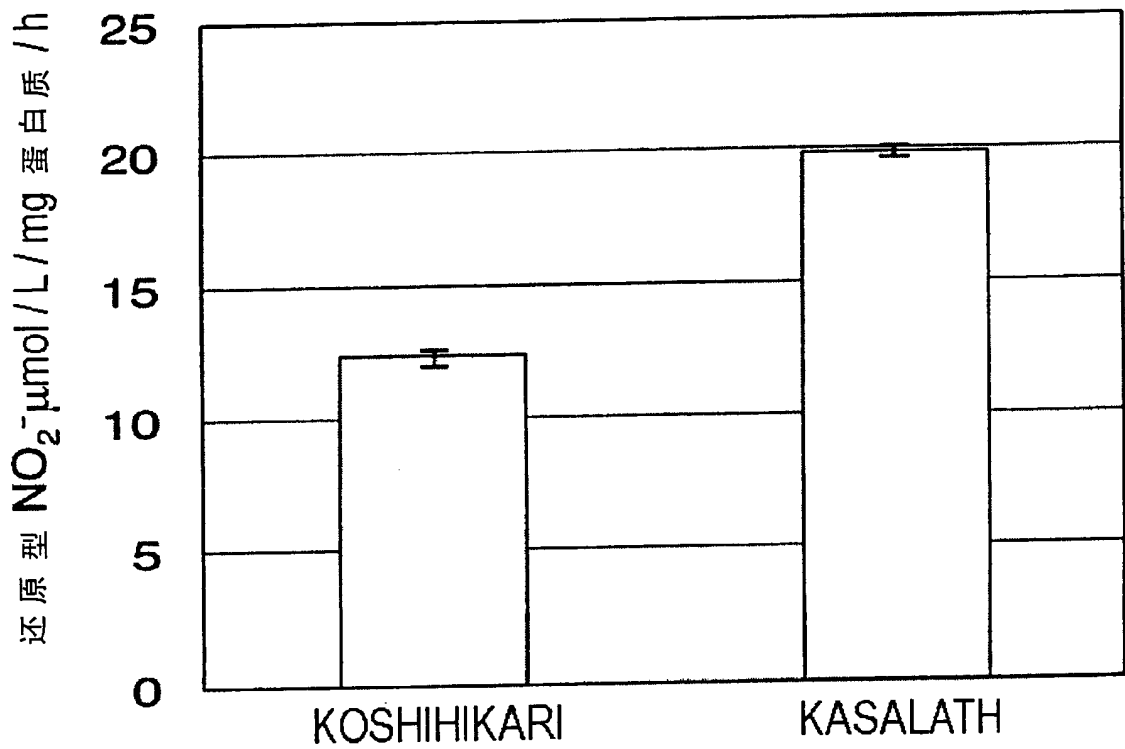
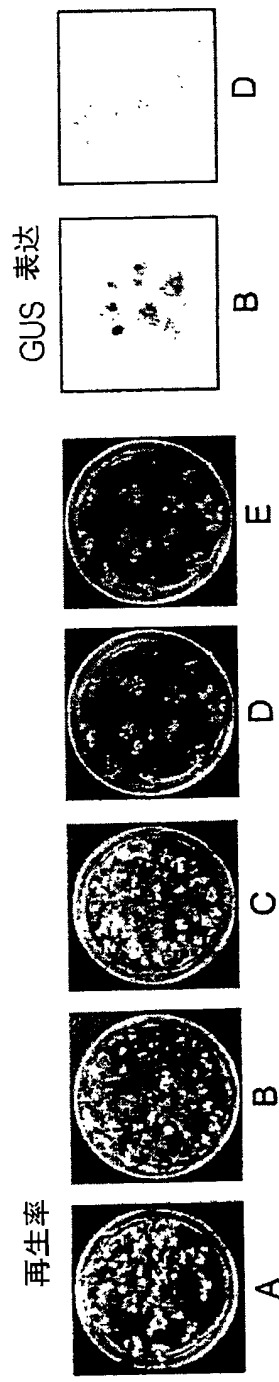
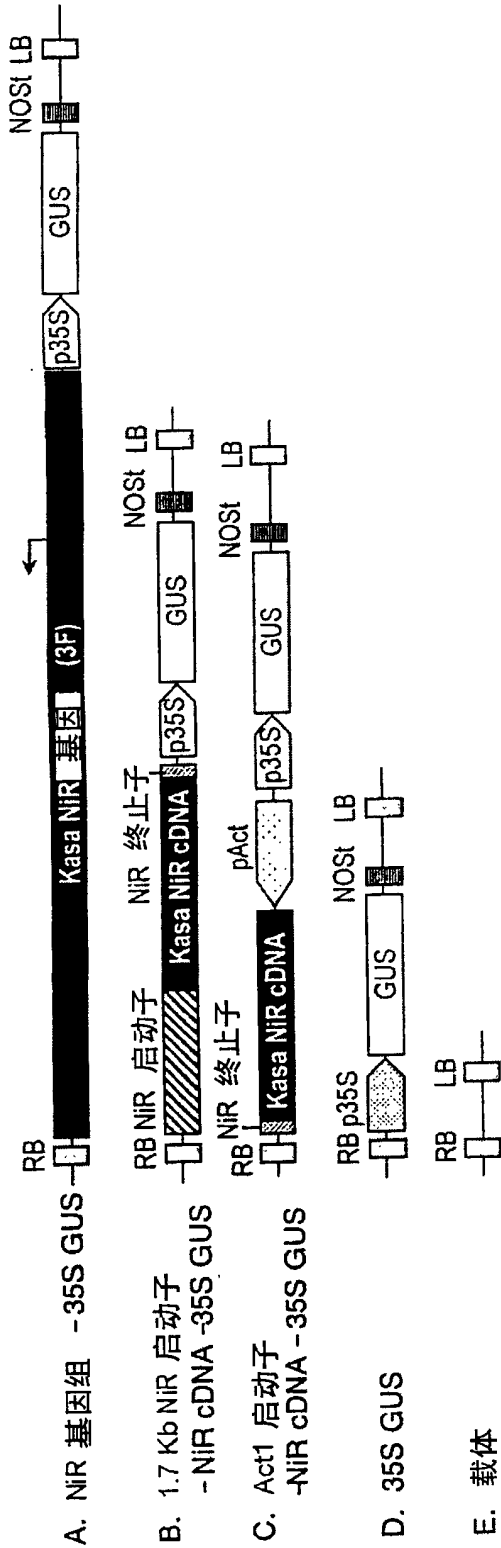


图7



	再生的愈伤组织	GUS染色的愈伤组织	GUS染色的愈伤组织的%
A. NIR 基因组 - 35S GUS	29	28	97
B. 1.7 kb NIR 启动子 - NIR cDNA - 35S GUS	54	48	88
C. Act1 启动子 - NIR cDNA - 35S GUS	42	34	81

图8

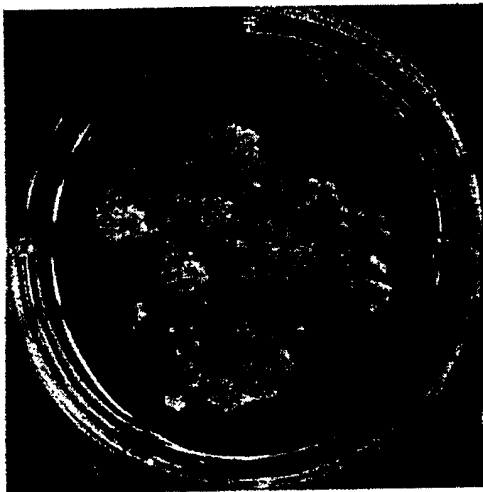
+NaNO₂



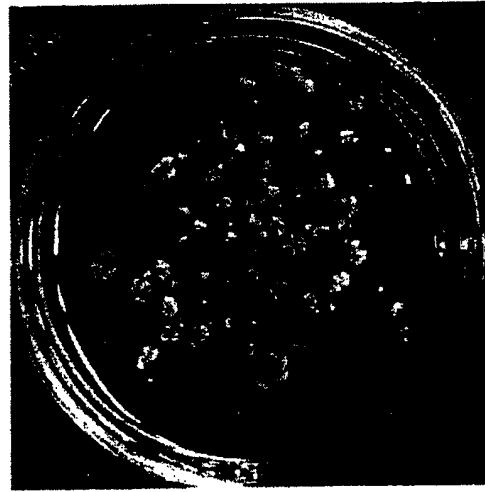
a

b

GUS- 染色



a



b

图9