

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200780037164. X

[51] Int. Cl.

C12N 15/82 (2006.01)

C12N 15/11 (2006.01)

A01H 5/00 (2006.01)

[43] 公开日 2009 年 9 月 2 日

[11] 公开号 CN 101522901A

[22] 申请日 2007. 10. 5

[21] 申请号 200780037164. X

[30] 优先权

[32] 2006. 10. 5 [33] US [31] 60/849,672

[86] 国际申请 PCT/US2007/021618 2007. 10. 5

[87] 国际公布 WO2008/045460 英 2008. 4. 17

[85] 进入国家阶段日期 2009. 4. 3

[71] 申请人 纳幕尔杜邦公司

地址 美国特拉华州

[72] 发明人 M·奥克曼 H·萨凯

J·蒂斯达尔 J·M·威尔逊

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 李 波 付 磊

权利要求书 1 页 说明书 55 页

[54] 发明名称

玉米微 RNA 序列

[57] 摘要

本申请描述了用于靶序列抑制和靶序列验证的方法和组合物。还描述了用于基因沉默的多核苷酸构建体, 以及包含该多核苷酸的细胞、植物和种子, 和使用微 RNA 来沉默靶序列的方法。

1. 分离的多核苷酸，其包含：选自 SEQ ID NOs: 1-2652 和 SEQ ID NOs: 7959-8114 的微 RNA。

2. 如权利要求 1 所述的微 RNA 的全长互补序列，或能够与权利要求 1 的微 RNA 杂交的核苷酸序列，其中所述核苷酸序列包含至少 21 个核苷酸。

3. 包含 SEQ ID NO: 7597 或 SEQ ID NO: 8427 的用于改变植物基因表达的分离的多核苷酸，其中所述分离的多核苷酸包含至少一个功能结构域，所述功能结构域具有至少 $21(x)$ 个邻接核苷酸，并且 x 是 1-2652 的整数，以及其中所述核苷酸在核苷酸 1 或任何核苷酸 $21(x)+1$ 起始。

4. 如权利要求 3 所述的分离的多核苷酸的功能亚结构域，其中所述功能亚结构域包含至少一个微 RNA。

5. 包含微 RNA 的分离的多核苷酸，所述微 RNA 含有选自 SEQ ID NOs: 5305-7956 和 SEQ ID NOs: 8271-8426 的序列。

6. 重组 DNA 构建体，所述重组 DNA 构建体包含与至少一个调节序列可操纵地连接的权利要求 1、2、3、4 或 5 的分离的多核苷酸。

7. 在其基因组中包含权利要求 6 的重组 DNA 构建体的植物、植物组织或植物细胞。

8. 从权利要求 7 的植物中获得的转基因种子、转基因种子副产品或转基因后代植物。

9. 改变植物中稳定导入的核苷酸序列的表达的方法，所述方法包括：

- a) 制备 DNA 表达构建体，所述构建体包含稳定导入的核苷酸序列以及至少一个能够与权利要求 1 的分离的多核苷酸杂交的序列；
- b) 用部分(a)的 DNA 表达构建体转化植物；和
- c) 选择转化的植物，所述转化的植物在其基因组中包含部分(a)的 DNA 表达构建体，并且当与用部分(a)的 DNA 表达构建体的修饰变型转化的植物相比时，具有稳定导入的核苷酸序列的变化的表达，其中所述修饰的构建体缺乏能够与权利要求 1 的分离的多核苷酸杂交的序列。

玉米微 RNA 序列

本申请要求于 2006 年 10 月 5 日提交的第 60/849,672 号美国临时申请的优先权, 该临时申请的公开内容全文以引用方式并入本文。

发明领域

本发明的领域一般涉及植物分子生物学。更具体地讲, 本发明涉及用于抑制靶基因表达的构建体和方法。

发明背景

微 RNA(miRNA)仅仅在几年前被首次鉴定出来, 但是已清楚的是, 他们在调节基因活性中起重要作用。这些 20-22 核苷酸非编码 RNA 能够经由碱基配对与特异性靶 mRNA 杂交, 并且通过介导 RNA 裂解或翻译阻抑来下调这些转录物的表达。最近的研究已经表明, miRNA 在发育期间具有重要功能。在植物中, 已经表明他们控制多个发育过程, 包括开花时间、叶子形态学、器官极性、花形态学以及根发育(综述于 Mallory 和 Vaucheret (2006) Nat Genet 38: S31-36 中)。由于 miRNA 的确定的调节作用, 它们有可能还参与某些主要作物特征例如抗旱性和抗病性。

植物 miRNA 是从长度为 50 至 500 个核苷酸、称为 pre-miRNA 的较长前体转录物加工的, 并且这些前体能够形成稳定的发夹结构(综述于 Bartel (2004) Cell 116: 281-297 中)。很多 miRNA 发夹前体最初是作为 1-2 kb 或更长的较长转录物存在, 其称为 pri-miRNA, 是聚腺苷酸化的和加帽的。该事实与在表达序列标记(EST)文库中检测到很多 pri-miRNA 结合在一起表明, RNA 聚合酶 II 是负责 miRNA 基因转录的酶。转基因实验表明, 是 pre-miRNA 的结构而不是其序列指导它们的正确加工, 并且对于 miRNA 的生成来说, 不需要其余 pri-miRNA。虽然在后生动物中, pri-miRNA 在细胞核中通过 Drosha 被加工成 pre-miRNA, 并且在细胞质中 Dicer 裂解 pre-miRNA, 但是植物中的 miRNA 成熟与动物中的途径不同, 因为植物缺乏 Drosha 同系物。反而, 除了其在发夹加工中的已知功能以外, 与 Dicer 同源的 RNase III 酶

DICER-LIKE 1 (DCL1)还可具有 Drosha 功能(Kurihara 和 Watanabe (2004) Proc Natl Acad Sci 101: 12753-12758)。

通过几个实验室的克隆努力,已经在拟南芥(*Arabidopsis*)中鉴定出了至少 30 个 miRNA 家族(综述于 Meyers 等人,(2006) Curr Opin Biotech 17: 1-8 中)。很多这些 miRNA 序列由一个以上基因座表现,总数目最高达大约 100 个。因为由一个实验室发现的特定 miRNA 与另一个实验室发现的 miRNA 一般不重叠,所以假定对于由给定植物基因组表达的整组 miRNA - “miRNome”的探求仍然是不完整的。导致这种情况的一个原因可能是,很多 miRNA 仅在非常特定的条件下表达,并且由此被标准克隆工作所遗漏。Sunkar 和 Zhu 的最近研究(2004, Plant Cell 16: 2001-2019)提出,实际上,可通过给文库构建选择“非标准”生长条件来帮助 miRNA 发现。Sunkar 和 Zhu 在由多种不同胁迫诱导的组织组成的文库中鉴定了新的 miRNA。他们继续进行研究,以证实干旱、寒冷和其他胁迫对这些 miRNA 当中的某一些的诱导,这意味着在 miRNA 在应激反应中的作用。完全表征植物 miRNome 的努力可能需要在很多不同组织中以及在很多不同条件下测定小的 RNA 状况。

标准 miRNA 克隆的一个补充方法是使用可获得的基因组和/或 EST 序列计算预测 miRNA,并且已经有几个实验室报道了使用该方法发现了新的拟南芥 miRNA (综述于 Bonnet 等人,(2006) New Phytol 171: 451-468)。使用部分依赖于观察到已知 miRNA 位于发夹前体中的这些计算方法,已经预测了几百种植物 miRNA。然而,只有一小部分已经通过 Northern 印记分析得到实验证实。此外,这些计算方法中的大部分依赖于比较两个典型基因组(例如拟南芥和水稻),以发现保守基因间隔区,因此不适于鉴定物种特异性 miRNA,所述物种特异性 miRNA 可能代表任何给定生物体的 miRNome 的相当大的一部分。

计算方法还已经帮助了 miRNA 靶的预测,并且通常植物 miRNA 与其靶共享高度的互补性(综述于 Bonnet 等人,(2006) New Phytol 171: 451-468 中)。所预测的植物 miRNA 的 mRNA 靶编码多种不同蛋白。这些蛋白有很多是转录因子,因此可能对于发育来说是重要的。然而,还有很多假定被定向的酶,并且这些酶可能在过程例如线粒体代谢、氧化应激反应、蛋白酶体功能和木质化中具有作用。随着另外的 miRNA 被鉴定出来,由 miRNA 调节的过程的目录可能会变得更长,并且可能

最终 miRNA 将涉及对于作物改善来说是至关重要的过程。例如，在硫同化途径中鉴定出了最近鉴定的 miRNA 靶基因，并且表明其在硫酸盐饥饿条件下被诱导(Jones-Rhoades 和 Bartel (2004) Mol Cell 14: 787-799)。于是该特定 miRNA 成为提高硫同化效率的候选基因。这使得有兴趣推测，同化其他化合物例如水和硝酸盐的途径也在 miRNA 控制下。

鉴定新 miRNA 的很多工作是在模型系统拟南芥中进行的，因此对于作物植物例如玉米、水稻和大豆的 miRNome 的了解较不充分。对于作物例如玉米和大豆，没有完全的基因组序列可以利用，这进一步妨碍了 miRNome 分析。很多拟南芥 miRNA 在这些其他物种中具有同系物，然而也存在似乎是拟南芥所特有的 miRNA。同样，预计也存在上述作物物种所特有的非保守 miRNA。有相当一部分非保守 miRNA 可能是与物种特异性生长条件或发育过程有关的调节网络的一部分。这样，在作物物种例如玉米中进行 miRNA 克隆以补充目前正在使用的生物信息学方法并且最终完全表征作物物种的 miRNome 是至关重要的。

表格简述

表 1: 由 6 列 2808 行组成的推定的 miRNA 的表格，所述 6 列是
第 1 列: ID: 单一 ID 数字编号，其代表该行中相应的微 RNA (miRNA)的 SEQ ID NO。

第 2 列: miRNA: 21 核苷酸玉米微 RNA 序列。

第 3 列: TARGETSEQ: 与靶 mRNA 上的潜在裂解位点相应的 miRNA 的精确互补体。

第 4 列: BACKBONE: 形成掺入 miRNA 序列内的发夹结构的基因组 DNA 序列。

第 5 列: FOLD: 基于嵌套括号的上下文无关文法的骨架序列的二级结构信息，其中“(“代表与由”)”表示的下游核苷酸杂交的核苷酸，并且“.”代表未配对核苷酸，所致结构显示了由骨架形成的发夹结构。

第 6 列: TARGET: miRNA 的可能基因组靶的公共 ID，所检索的数据库包括来自 TIGR Version 4 水稻数据集 all.cds 和 all_small_genes.cds 的水稻基因；以及 TIGR 玉米基因索引文件 ZMGI.101205。

表 4: 有关本发明的玉米微 RNA 和公知领域玉米(*Zea mays*) 微 RNA 的小 RNA 表达数据的表, 其由 18 列 2702 行组成, 所述 18 列是

第 1 列: ID: 单一 ID 数字编号, 其代表该行中相应的推定微 RNA (miRNA) 的 SEQ ID NO, 或者与该行中 miRNA 序列有关的公知领域玉米(*Zea mays*) 名称。

第 2 列: MIRNACORE: ~ 21 核苷酸玉米微 RNA 序列, 还称为“查询序列”。

第 3 列: CALL B73_LEAF: 应用到该行中的 miRNA, 基于在源自 B73 叶子组织的 RNA 样本中的其相应的表达谱 (profiling) 测量结果的可检测性评估。只有当有关该 miRNA 的数据满足表 3 中的特异性、再现性和信号强度标准时, 才应用“通过的”、“富集的”或“高富集的”。“富集的”应用于 miRNA, 表明在相同实验中, B73 叶子样本中的完全序列配对(0MM)探针信号强度是其他样本中的 5 倍至 10 倍。”高富集的”应用于 miRNA, 表明在相同实验中, B73 叶子样本中的 0MM-探针信号强度是其他样本中的 10 倍。样本 B73 叶子以及其他样本描述在表 2 中。

第 4 列: 0MM B73_LEAF: 如在基因型 B73 叶子 RNA 的表达谱中测定的该行中 miRNA 的完全序列配对探针的标准化荧光强度评分。

第 5 列: CALL B73_EAR: 应用到该行中的 miRNA, 基于在源自 B73 穗组织的 RNA 样本中的其相应的表达谱测量结果的可检测性评估。

第 6 列: 0MM B73_EAR: 如在基因型 B73 穗 RNA 的表达谱中测定的该行中 miRNA 的完全序列配对探针的标准化荧光强度评分。

第 7 列: CALL B73_SEEDLING: 应用到该行中的 miRNA, 基于在源自 B73 幼苗组织的 RNA 样本中的其相应的表达谱测量结果的可检测性评估。

第 8 列: 0MM B73_SEEDLING: 如在基因型 B73 幼苗 RNA 的表达谱中测定的该行中 miRNA 的完全序列配对探针的标准化荧光强度评分。

第 9 列: CALL 3245_Normal_N: 应用到该行中的 miRNA, 基于在源自 3245 叶子组织的 RNA 样本中的其相应的表达谱测量结果的可检测性评估, 其中所述 3245 叶子组织是在具有标准水平氮的田地中生长的。

第 10 列: 0MM 3245_Normal_N: 如在来自基因型 3245 植物叶子的 RNA 的表达谱中测定的该行中 miRNA 的完全序列配对探针的标准化荧光强度评分, 其中所述 3245 植物是在具有标准水平氮的田地中生长的。

第 11 列: CALL 3245_Low_N: 应用到该行中的 miRNA, 基于在源自 3245 叶子组织的 RNA 样本中的其相应的表达谱测量结果的可检测性评估, 其中所述 3245 叶子组织是在具有低水平氮的田地中生长的。

第 12 列: 0MM 3245_Low_N: 如在来自基因型 3245 植物叶子的 RNA 的表达谱中测定的该行中 miRNA 的完全序列配对探针的标准化荧光强度评分, 其中所述 3245 植物是在具有低水平氮的田地中生长的。

第 13 列: CALL 33B50_Low_N: 应用到该行中的 miRNA, 基于在源自 33B50 叶子组织的 RNA 样本中的其相应的表达谱测量结果的可检测性评估, 其中所述 33B50 叶子组织是在具有低水平氮的田地中生长的。

第 14 列: 0MM 33B50_Low_N: 如在来自基因型 33B50 植物叶子的 RNA 的表达谱中测定的该行中 miRNA 的完全序列配对探针的标准化荧光强度评分, 其中所述 33B50 植物是在具有低水平氮的田地中生长的。

第 15 列: CALL B73_Low_N: 应用到该行中的 miRNA, 基于在源自 B73 叶子组织的 RNA 样本中的其相应的表达谱测量结果的可检测性评估, 其中所述 B73 叶子组织是在最小氮补充的条件下生长的。

第 16 列: 0MM B73_Low_N: 如在来自基因型 B73 植物叶子的 RNA 的表达谱中测定的该行中 miRNA 的完全序列配对探针的标准化荧光强度评分, 其中所述 B73 植物是在最小氮补充的条件下生长的。

第 17 列: TARGETSEQ: 与靶 mRNA 上的潜在裂解位点相应的 miRNA 的精确互补体。

第 18 列: TARGET: 推定的 miRNA 的可能基因组靶的公共 ID, 所检索的数据库包括来自 TIGR Version 4 水稻数据集 all.cds 和 all_small_genes.cds 的水稻基因; 以及 TIGR 玉米基因索引文件 ZMGI.101205。

序列表简述

本申请在光盘上提供序列表和表格。根据 37 CFR 1.52(e), PCT Rule 5.2, 和 PCT Administrative Instruction 802(a)的规定, 将含有序列表和表格的光盘的内容以引用方式并入本文。光盘以一式三份提交, 并且彼此相同。光盘标记为“副本 1 - 序列表和表格”, “副本 2 - 序列表和表格”, 和 CRF。光盘含有以下文件: 具有以下大小的 BB1593 US PRV 序列表: 1,471,000 个字节, 和具有以下大小的 BB1593 US PRV 表格: 4,430,000 个字节, 是在 2006 年 10 月 5 日产生的。

SEQ ID NOs: 1-2652 代表来自玉米的个体 21 核苷酸微 RNA 序列。个体微 RNA 在表 1 中作为“miRNAcore”列(第 2 列)显示, 每一序列的 SEQ ID NO 显示在表 1 的第 1 列(ID)中。

SEQ ID NOs: 2653-5304 代表 SEQ ID NOs: 1-2652 的互补序列, 还在表 1 的第 3 列中作为“targetseq”显示。

SEQ ID NOs: 5305-7956 分别代表玉米基因组中环绕 SEQ ID NOs: 1-2652 中的 miRNA 的“骨架”发夹序列。该信息还可参见表 1 的第 4 列。

SEQ ID NO: 7957 是在 SEQ ID NOs: 1-2652 中发现的所有 21 核苷酸 miRNA 的顺序串联($21 \times 2652 = 55,692$)。据信该代表性实例是用于改变植物基因表达的很多不同形式中的一个。其他实例将包括个体 miRNA 序列的不同排序, 一些 miRNA 序列的重复, 一些 miRNA 序列的消除, 以及 miRNA 序列之间的改变的间距。

SEQ ID NO: 7958 是与拟南芥(Arabidopsis)脂肪酸脱饱和酶 2 (FAD2)基因互补的 21 核苷酸序列。

SEQ ID NOs: 7959-8114 代表来自玉米的个体 21 核苷酸微 RNA 序列。个体微 RNAs 在表 1 中作为“miRNAcore”列(第 2 列)显示,每一序列的 SEQ ID NO 在表 1 的第 1 列(ID)中显示。

SEQ ID NOs: 8115-8270 代表 SEQ ID NOs: 7959-8114 的互补序列,还在表 1 的第 3 列中作为“targetseq”显示。

SEQ ID NOs: 8271-8426 分别代表玉米基因组中环绕 SEQ ID NOs: 7959-8114 中的 miRNA 的“骨架”发夹序列。该信息还可参见表 1 的第 4 列。

SEQ ID NO: 8427 是在 SEQ ID NOs: 7959-8114 中发现的所有 21 核苷酸 miRNA 的顺序串联($21 \times 156 = 3,276$)。据信该代表性实例是用于改变植物基因表达的很多不同形式中的一个。其他实例将包括个体 miRNA 序列的不同排序,一些 miRNA 序列的重复,一些 miRNA 序列的消除,以及 miRNA 序列之间的改变的间距。

SEQ ID NOs: 8428-8429 是 RNA 衔接子序列。

发明概述

本发明特别涉及分离的多核苷酸,其包含:选自 SEQ ID NOs: 1-2652 和 SEQ ID NOs: 7959-8114 的微 RNA。

在第二个实施方案中,本发明涉及选自 SEQ ID NOs: 1-2652 和 SEQ ID NOs: 7959-8114 的微 RNA 的全长互补序列,或能够与这些上述微 RNA 杂交的核苷酸序列,其中所述可杂交的核苷酸序列包含至少 21 个核苷酸。

在第三个实施方案中,本发明涉及包含 SEQ ID NO: 7597 或 SEQ ID NO: 8427 的用于改变植物基因表达的分离的多核苷酸,其中所述分离的多核苷酸包含至少一个功能结构域,所述功能结构域具有至少 $21(x)$ 个邻接核苷酸,并且 x 是 1-2652 的整数,以及其中所述核苷酸在核苷酸 1 或任何核苷酸 $21(x)+1$ 起始。

在第四个实施方案中,本发明涉及这样的分离的多核苷酸的功能亚结构域,其中所述功能亚结构域包含至少一个微 RNA。

在第五个实施方案中,本发明涉及包含微 RNA 的分离的多核苷酸,所述微 RNA 含有选自 SEQ ID NOs: 5305-7956 和 SEQ ID NOs: 8271-8426 的序列。

在第六个实施方案中,本发明涉及 DNA 表达构建体,所述 DNA 表达构建体包含与至少一个调节序列可操纵地连接的任何本文所述的分离的多核苷酸。

在第七个实施方案中,本发明涉及在其基因组中包含本文所述的 DNA 表达构建体的植物。这样的植物可以选自玉米、水稻、高粱、向日葵、小米、大豆、卡诺拉(canola)、小麦、大麦、燕麦、菜豆(bean)和坚果。

在第八个实施方案中,本发明涉及从在其基因组中包含本文所述的 DNA 表达构建体的植物中获得的转基因种子。还包括在本发明范围内的是在其基因组中包含本文所述的 DNA 表达构建体的转化的植物组织或植物细胞。

在第九个实施方案中,本发明涉及从这样的转基因种子获得的副产品,或从这样的转基因种子获得的后代植物。

在第十个实施方案中,本发明涉及改变植物中稳定导入的核苷酸序列的表达的方法,所述方法包括:

- a) 制备 DNA 表达构建体,所述构建体包含稳定导入的核苷酸序列以及至少一个能够与本发明分离的多核苷酸杂交的序列;
- b) 用部分(a)的 DNA 表达构建体转化植物; 和
- c) 选择转化的植物,所述转化的植物在其基因组中包含部分(a)的 DNA 表达构建体,并且当与用部分(a)的 DNA 表达构建体的修饰变型转化的植物相比时,具有稳定导入的核苷酸序列的改变的表达,其中所述修饰的构建体缺乏能够与本发明分离的多核苷酸杂交的序列。

发明详述

与本申请有关的信息可参见于 2004 年 10 月 12 日提交的美国专利申请 10/963,238 和 10/963,394。上述专利申请全文以引用方式并入本文。

可用于理解本发明的其他参考文献包括于 2004 年 7 月 1 日提交的美国专利申请 10/883,374; 于 2004 年 8 月 6 日提交的美国专利申请 10/913,288; 和于 2006 年 1 月 6 日提交的美国专利申请 11/334,776。

最近发现的小 RNA 在控制基因表达中起重要作用。包括开花在内的很多发育过程是由小 RNA 控制的。现在能够通过使用在植物中产生小 RNA 的转基因构建体来以工程手段改变植物基因的基因表达。

本发明提供了用于抑制靶序列的方法和组合物。本发明组合物可用于任何类型的植物细胞，以及包含适当的加工成分(例如 RNA 干扰成分)的其他细胞，包括无脊椎动物和脊椎动物细胞。本发明组合物和方法是基于在拟南芥中发现的内源性 miRNA 沉默过程，类似策略也用于扩大组合物和使用这些方法的生物体的数量。可以让本发明方法适于在任何真核细胞系统中实施。此外，本文所述组合物和方法可用于培养物中的个体细胞、细胞或组织，或者在体内用于生物体，或者用于器官或生物体的其他部分。

本发明组合物通过编码与靶序列区域具有显著互补性的 miRNA 而选择性地抑制靶序列。miRNA 在核酸构建体中提供。当转录成 RNA 时，预计所述核酸构建体会形成发夹结构，细胞加工该发夹结构以产生 miRNA，而 miRNA 抑制靶序列的表达。

本申请公开了编码来自玉米的 miRNA 的核酸序列。还公开了含有个体 miRNA 序列的骨架发夹。描述了用于 miRNA 及其骨架的转基因表达的构建体。或者，描述了构建体，其中骨架序列与 miRNA 序列交换，由此改变转基因宿主中 miRNA 及其随后的特定靶序列的表达模式。任何 miRNA 可以与任何其他骨架交换，以产生新的 miRNA/骨架杂合体。

本申请提供了抑制靶序列的方法，所述方法利用任何上述构建体，其中设计 miRNA 来识别靶序列区域，并且插入到构建体内。在导入到细胞内后，产生的 miRNA 抑制靶序列的表达。靶序列可以是内源性植物序列，或者是植物中的异源转基因。

本文还提及了靶基因，例如来自植物病原体，例如致病病毒、线虫、昆虫、霉菌或真菌的基因。

本发明的另一个方面涉及包含构建体和/或 miRNA 的植物、细胞和种子。通常，细胞是得自植物的细胞，但是也可以是其他原核细胞或真核细胞，包括但不限于病毒、细菌、酵母、昆虫、线虫或动物细胞。植物细胞包括来自单子叶植物(monocots)和双子叶植物(dicots)的细胞。本发明还提供了包含构建体和/或 miRNA 的植物和种子。

“植物”包括整个植株、植物器官、植物组织、种子和植物细胞以及同一植物的后代。植物细胞包括但不限于得自下列的细胞：种子、悬浮培养物、胚芽、分生区域、愈伤组织、叶、根、芽、配子体、孢子体、花粉和小孢子。

术语“植物部分”包括分化和未分化的组织，包括但不限于下列部分：根、茎、芽、叶、花粉、种子、肿瘤组织和多种形式的细胞和培养物(例如单个细胞、原生质体、胚和愈伤组织)。植物组织可存在于植株或植物器官、组织或细胞培养物中。

术语“植物器官”是指构成植株的形态和功能不同部分的植物组织或组织群。

术语“基因组”是指：(1)存在于生物体的每一个细胞或者病毒或细胞器中的全套遗传物质(基因和非编码序列)；(2)作为(单倍体)单位从一个亲本遗传的全套染色体。

“后代”包括植物的任何后面的世代。后代将从其亲本植物遗传并且稳定地分离基因和转基因。

单位、前缀和符号可以以其 SI 接受的形式表示。除非另外说明，核酸是以 5'到 3'方向从左往右写；氨基酸序列分别是以氨基到羧基方向从左往右写。本说明书中提及的数值范围包括限定该范围的数字，并且包括在该限定范围内的每一个整数。在本文中氨基酸可用通常已知的三字母符号或 IUPAC-IUB Biochemical Nomenclature Commission 推荐的单字母符号表示。同样，核苷酸可以用其通常接受的单字母代码表示。除非另有说明，在本文中使用的软件、电和电子术语是如在 The New IEEE Standard Dictionary of Electrical and Electronics Terms (第 5 版，1993)中所定义的。把说明书作为一个整体时，下面定义的术语就定义得更全面。

术语“重组构建体”、“表达构建体”、“嵌合构建体”、“构建体”和“重组 DNA 构建体”在本文中可互换使用。重组构建体包含核酸片段例如在天然条件下不会一起存在的调节序列和编码序列的人工组合。例如，嵌合构建体可包含源自不同来源的调节序列和编码序列，或源自相同来源但以不同于天然存在的方式排列的调节序列和编码序列。这样的构建体可以单独使用或与载体联合使用。如果使用载体，则载体的选择取决于本领域技术人员熟知的将被用于转化宿主细胞的

方法。例如，可以使用质粒载体。本领域技术人员十分了解必须存在于载体上以成功地转化、选择和繁殖包含本发明任何分离的核酸片段的宿主细胞的遗传元件。本领域技术人员还将认识到，不同的独立转化事件将导致不同水平和模式的表达(Jones 等人, EMBO J. 4: 2411-2418 (1985); De Almeida 等人, Mol. Gen. Genetics 218: 78-86 (1989)), 因此, 必须筛选多次事件以获得显示期望表达水平和模式的品系。这样的筛选可以通过 DNA 的 Southern 印迹分析、mRNA 表达的 Northern 印迹分析、蛋白表达的免疫印迹分析或表型分析等完成。

该构建体可以包含脱氧核糖核苷酸、核糖核苷酸和/或修饰核苷酸的任何组合。构建体可以转录以形成 RNA, 其中 RNA 可以能够形成双链 RNA 和/或发夹结构。构建体可以在细胞中表达, 或者分离, 或者合成制得。构建体还可以包含启动子或帮助构建体操作或表达的其他序列。

本文所用“阻抑”或“沉默”或“抑制”可互换使用, 是指相对于野生型生物体中其正常表达水平, 靶序列的产物表达的下调。抑制包括, 相对于野生型表达水平, 表达降低了约 10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95% 或 100%。

本文所用术语“编码”是指可被加工以产生 RNA 和/或多肽的 DNA 序列。

本文所用术语“表达”是指产生功能产物, 例如由导入的构建体、内源 DNA 序列或稳定导入的异源 DNA 序列产生 RNA 转录物。该术语还可以指由产生自任何上述 DNA 前体的 mRNA 产生多肽。因此, 核酸片段的表达可以指核酸片段的转录(例如生成 mRNA 或其他功能 RNA 的转录)和/或 RNA 翻译成前体或成熟蛋白(多肽)。

本文在序列方面所用的“异源”是指, 源自外来物种的序列, 或者, 如果源自相同物种, 通过人为介入而从其天然形式在组成和/或基因组基因座方面显著修饰的序列。例如, 对于核酸, 其可以是源自外来物种的核酸, 或者是合成设计的核酸, 或者, 如果源自相同物种, 通过人为介入而从其天然形式在组成和/或基因组基因座方面显著修饰的核酸。异源蛋白可以源自外来物种, 或者, 如果源自相同物种, 通过人为介入而从其起源形式显著修饰。

术语“宿主细胞”是指含有或者其中导入核酸构建体，并且支持该构建体复制和/或表达的细胞。宿主细胞可以是原核细胞例如大肠杆菌(*E. coli*)，或真核细胞例如真菌、酵母、昆虫、两栖动物、线虫或哺乳动物细胞。或者，宿主细胞是单子叶植物或双子叶植物细胞。单子叶植物宿主细胞的一个实例是玉米宿主细胞。

术语“导入”是指将核酸(例如表达构建体)或蛋白提供至细胞中。导入包括指将核酸整合进真核或原核细胞中，在该细胞中核酸可以整合进细胞的基因组内，以及包括指将核酸或蛋白暂时提供给细胞。导入包括指稳定的或瞬时的转化方法以及有性杂交。因此，将核酸片段(例如重组 DNA 构建体/表达构建体)插入细胞内的情况下的“导入”是指“转染”或“转化”或“转导”，并且包括指将核酸片段整合进真核或原核细胞中，在该细胞中核酸片段可以整合进细胞的基因组(如染色体、质粒、质体或线粒体 DNA)内、转变成自主的复制子或瞬时表达(例如转染的 mRNA)。

在应用于植物细胞时，术语“基因组”不仅包括在细胞核内发现的染色体 DNA，并且还包括在细胞的亚细胞成分(例如线粒体、质体)内发现的细胞器 DNA。

术语“分离的”是指材料例如核酸或蛋白，其：(1)基本上或本质上不含有通常伴随在其天然存在环境中发现的材料或者与在其天然存在环境中发现的材料相互作用的成分，或者(2)如果材料在其天然环境中，则已经通过人为介入改变了材料的组成和/或在细胞基因座上放置了不是该材料的天然基因座的基因座。

本文使用的微 RNA 或“miRNA”是指调节包含靶序列的多核苷酸的表达的寡核糖核酸。“成熟 miRNA”是指从 miRNA 前体的加工产生的 miRNA。“miRNA 模板”是核酸构建体中编码 miRNA 的区域。miRNA 的“背面”区域是与 miRNA 模板基本上互补并且预计与 miRNA 模板碱基配对的多核苷酸构建体的部分。miRNA 模板和背面可以形成包括发夹结构的双链多核苷酸。

本文所用的“结构域”或“功能结构域”是指能够引起植物中的生物反应的核酸序列。本发明涉及由单独起作用或者与其他 miRNA 序列协同起作用的至少 21 核苷酸序列组成的 miRNA，由此，结构域可以指个体 miRNA 或 miRNA 群。而且，与其骨架序列有关的 miRNA 序

列可以被认为用于将 miRNA 加工成其活性形式的结构域。本文所用的“亚结构域”或“亚功能结构域”是指能够引起植物中的生物反应的结构域的亚序列。miRNA 可以被认为骨架序列的亚结构域。“邻接”序列或结构域是指顺序连接的，在结构域之间没有加入的核苷酸插入的序列。邻接结构域串的一个实例存在于 SEQ ID NO: 7957 中，其代表作为连续串的 SEQ ID NOs: 1-2652，据信该连续串可以是以顺序串联而连接在一起的 2652 miRNA 序列。

本文所用的短语“靶序列”和“所关注序列”在本文中可互换使用。靶序列是用于指选择用来改变(例如阻抑)表达的核酸序列，并且其不限于编码多肽的多核苷酸。靶序列包括与 miRNA 基本上或完全互补的序列。靶序列包括但不限于包含靶序列的 RNA、DNA 或多核苷酸。如 Bartel 和 Bartel (2003) *Plant Phys.* 132: 709-719 中所讨论的那样，大部分微 RNA 序列是 20-22 个核苷酸，当与其靶序列比较时，在任何位置具有 0-3 个错配。

应当理解，本发明的微 RNA 序列例如 21 核苷酸序列作为更短(20 核苷酸)或更长序列(22 核苷酸)可以仍然具有功能。此外，某些核苷酸取代，特别是在微 RNA 序列的 3' 末端的最后两个核苷酸上的核苷酸取代，可用于保留至少某些微 RNA 功能。

本文所用的“核酸”是指多核苷酸，并包括单链或双链的脱氧核糖核苷酸或核糖核苷酸碱基聚合物。核酸也可以包括片段和经修饰的核苷酸。因此，术语“多核苷酸”、“核酸序列”、“核苷酸序列”或“核酸片段”可互换使用，并且是作为单链或双链的 RNA 或 DNA 聚合物，任选含有合成的、非天然的或改变的核苷酸碱基。核苷酸(通常以它们的 5'-单磷酸形式存在)可以用如下它们的单字母名称指代：“A”表示腺苷酸或脱氧腺苷酸(分别针对 RNA 或 DNA)，“C”表示胞苷酸或脱氧胞苷酸，“G”表示鸟苷酸或脱氧鸟苷酸，“U”表示尿苷酸，“T”表示脱氧胸苷酸，“R”表示嘌呤(A 或 G)，“Y”表示嘧啶(C 或 T)，“K”表示 G 或 T，“H”表示 A 或 C 或 T，“I”表示肌苷，“N”表示任何核苷酸。

“核酸文库”是指包含并且基本上代表特定生物体或来自该生物体的组织的基因组的整个转录部分的分离的 DNA 或 RNA 的集合。示例性核酸文库，例如基因组和 cDNA 文库的构建教导于标准分子生物

学参考文献中, 例如 Berger 和 Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology, 第 152 卷, Academic Press, Inc., San Diego, CA (Berger); Sambrook 等人, Molecular Cloning - A Laboratory Manual, 第二版, 第 1-3 卷, (1989); 和 Current Protocols in Molecular Biology, F.M. Ausubel 等人, Eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc. (1994)。

本文所用的“可操纵地连接”包括指至少两个序列的功能性连接。可操纵地连接包括在启动子与另一个序列之间的连接, 其中启动子序列引发和介导与另一个序列相应的 DNA 序列的转录。

本文所用的“植物”包括植物和植物部分, 包括但不限于植物细胞, 植物组织例如叶、茎、根、花和种子。

本文所用的“多肽”是指蛋白、蛋白片段、修饰的蛋白、氨基酸序列以及合成氨基酸序列。多肽可以是糖基化的或者未糖基化的。

本文所用的“启动子”是指涉及 RNA 聚合酶和其他蛋白的识别以及结合以引发转录的核酸片段, 例如 DNA 区域。换句话说, 该核酸片段能够控制另一核酸片段的转录。

术语“选择性杂交”包括指在严格杂交条件下, 核酸序列与特定核酸靶序列以比其与非靶核酸序列的杂交更高的可检测程度(例如至少两倍于背景)杂交, 并指基本排除了非靶核酸。选择性杂交的序列通常彼此具有约至少 80% 的序列同一性, 或 90% 的序列同一性, 最多并包括 100% 的序列同一性(即完全互补)。

术语“严格条件”或“严格杂交条件”包括指探针将选择性杂交至其靶序列的条件。严格条件是序列依赖性的, 并将因不同的环境而异。通过控制杂交和/或洗涤条件的严格性, 可鉴定与探针 100% 互补的靶序列(同源探测)。作为另一种选择, 可调节严格条件以允许序列中的一些错配, 以便检测到更低程度的相似性(异源探测)。通常, 探针的长度少于约 1000 个核苷酸, 任选长度少于 500 个核苷酸。

一般地, 严格条件将是如下那些条件: 在 pH7.0-8.3 下盐浓度低于约 1.5M 钠离子, 通常约 0.01-1.0M 钠离子浓度(或其它盐), 并且对于短探针(例如 10-50 个核苷酸)温度为至少约 30℃, 而对于长的探针(例如多于 50 个核苷酸)温度为至少约 60℃。严格条件也可以通过加入诸

如甲酰胺之类的去稳定剂来实现。示例性的低严格条件包括在 37℃ 于含有 30-35% 甲酰胺、1M NaCl、1% SDS(十二烷基硫酸钠)的缓冲液中杂交,以及在 50-55℃ 用 1×至 2×SSC(20×SSC=3.0M NaCl/0.3M 柠檬酸三钠)洗涤。示例性的中等严格条件包括在 37℃ 于 40-45% 甲酰胺、1M NaCl、1% SDS 中杂交,以及在 55-60℃下用 0.5×至 1×SSC 洗涤。示例性的高严格条件包括在 37℃ 于 50% 甲酰胺、1M NaCl、1% SDS 中杂交,以及在 60-65℃用 0.1×SSC 洗涤。

特异性通常取决于杂交后洗涤,最后的洗涤溶液的离子强度和温度是关键因素。对于 DNA-DNA 杂交体, T_m 可以用 Meinkoth 和 Wahl, Anal. Biochem. 138: 267-284(1984)中的等式估算: $T_m = 81.5^\circ\text{C} + 16.6(\log M) + 0.41(\%GC) - 0.61(\%\text{form}) - 500/L$; 其中 M 是单价阳离子的体积摩尔浓度, %GC 是 DNA 中鸟苷和胞嘧啶核苷酸的百分比, %form 是杂交溶液中甲酰胺的百分比, 而 L 是以碱基对表示的杂交体的长度。 T_m 是(在确定的离子强度和 pH 下) 50%的互补靶序列与完全匹配的探针杂交时的温度。每出现 1% 的错配, T_m 降低约 1℃; 因此, 可调节 T_m 、杂交和/或洗涤条件以与具有期望同一性的序列杂交。例如, 如果要寻求具有 $\geq 90\%$ 同一性的序列, 则可将 T_m 降低 10℃。通常, 在确定的离子强度和 pH 下, 严格条件选择为比特定序列及其互补序列的热解链温度(T_m)低约 5℃。然而, 极严格条件可采用在比热解链温度(T_m)低 1、2、3 或 4℃ 的温度下杂交和/或洗涤; 中等严格条件可采用在比热解链温度(T_m)低 6、7、8、9 或 10℃ 的温度下杂交和/或洗涤; 低严格条件可采用在比热解链温度(T_m)低 11、12、13、14、15 或 20℃ 温度下杂交和/或洗涤。利用所述等式、杂交和洗涤组合物以及理想的 T_m , 本领域技术人员将会理解, 杂交和/或洗涤溶液的严格性的变化已经在本质上得以描述。如果所需的错配程度导致 T_m 低于 45℃ (水溶液)或 32℃ (甲酰胺溶液), 则优选增加 SSC 的浓度以便能使用更高的温度。对于核酸杂交的详尽指导可在以下文献中找到: Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes, 第 I 部分, 第 2 章 “Overview of principles of hybridization and the strategy of nucleic acid probe assays”, Elsevier, New York (1993); 和 Current Protocols in Molecular Biology, 第 2 章, Ausubel

等人, Eds., Greene Publishing and Wiley-Interscience, New York (1995)。杂交和/或洗涤条件可进行至少 10、30、60、90、120 或 240 分钟。

术语“可靠检测”和“可靠地检测的”在本文中定义为是指在背景噪音以上的可测量的序列特异性信号强度的可再现检测。

本文所用的“转基因”是指其基因组内含有异源多核苷酸的植物或细胞。优选地, 将异源多核苷酸稳定地整合进基因组内以便该多核苷酸连续传代。异源多核苷酸可以单独地或作为表达构建体的部分整合进基因组中。本文所用的转基因包括因异源核酸存在而已经改变了基因型的任何细胞、细胞系、愈伤组织、组织、植物部分或植物, 包括初始进行此类改变的转基因以及从初始的转基因通过有性杂交或无性繁殖产生的那些转基因。本文所用的术语“转基因”不包括通过常规植物育种方法或通过诸如随机异花受精、非重组病毒感染、非重组细菌转化、非重组转座或自发突变之类的自然发生事件导致的基因组(染色体或染色体外)改变。

本文所用的“载体”是指可用于将本发明的多核苷酸递送到宿主细胞内的小核酸分子(质粒、病毒、噬菌体、人工或剪切的 DNA 分子)。载体能够复制, 并且含有用于导入外来多核苷酸的克隆位点。因此, 表达载体容许插入其内的核酸转录。

多核苷酸序列可以与靶基因的所选区域具有基本上相同、基本上同源或基本上互补。本文使用的“基本上相同”和“基本上同源”是指彼此具有序列同一性或同源性的序列。通常, 基本上相同或基本上同源的序列将具有约 75%、80%、85%、90%、95% 或 100% 的序列同一性, 其中百分比序列同一性是基于整个序列, 并且是通过使用缺省参数的 GAP 比对(GCG, GAP version 10, Accelrys, San Diego, CA)测定的。GAP 使用 Needleman and Wunsch 的算法(J. Mol. Biol. 48: 443-453, 1970)以发现将匹配数目最大化以及将序列缺口数目最小化的两个完全序列的比对。具有 100% 同一性的序列是相同的。“基本上互补”是指彼此互补, 并且能够彼此碱基配对的序列。在描述互补序列时, 如果第一个序列中的所有核苷酸将与第二个序列碱基配对, 那么这些序列就完全或全部互补。

本发明特别涉及分离的多核苷酸, 其包含: 选自 SEQ ID NOs: 1-2652 和 SEQ ID NOs: 7959-8114 的微 RNA。

在第二个实施方案中,本发明涉及选自 SEQ ID NOs: 1-2652 和 SEQ ID NOs: 7959-8114 的微 RNA 的全长互补序列,或能够与这些上述微 RNA 杂交的核苷酸序列,其中所述可杂交的核苷酸序列包含至少 21 个核苷酸。

在第三个实施方案中,本发明涉及包含 SEQ ID NO: 7597 或 SEQ ID NO: 8427 的用于改变植物基因表达的分离的多核苷酸,其中所述分离的多核苷酸包含至少一个功能结构域,所述功能结构域具有至少 $21(x)$ 个邻接核苷酸,并且 x 是 1-2652 的整数,以及其中所述核苷酸在核苷酸 1 或任何核苷酸 $21(x)+1$ 起始。

在第四个实施方案中,本发明涉及这样的分离的多核苷酸的功能亚结构域,其中所述功能亚结构域包含至少一个微 RNA。

在第五个实施方案中,本发明涉及包含微 RNA 的分离的多核苷酸,所述微 RNA 含有选自 SEQ ID NOs: 5305-7956 和 SEQ ID NOs: 8271-8426 的序列。

miRNA 的计算鉴定是由按大小选择的小 RNA 文库实现的,所述小 RNA 文库来自叶、干旱胁迫的叶、种子以及各种其他组织。

在某些实施方案中,相对于靶序列,miRNA 模板(即编码 miRNA 的多核苷酸)以及由此 miRNA 可包含某些错配。在某些实施方案中,与靶序列相比,miRNA 模板具有 ≥ 1 个核苷酸错配,例如,与靶序列相比,miRNA 模板可具有 1、2、3、4、5 或更多个错配。错配程度还可以通过确定 miRNA 模板与靶序列的互补序列的同一性百分比来描述。例如,与靶序列的互补序列相比,miRNA 模板可具有包括约至少 70%、75%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99% 或 100% 的同一性百分比。

在某些实施方案中,相对于 miRNA 背面,miRNA 模板(即编码 miRNA 的多核苷酸)以及由此 miRNA 可包含某些错配。在某些实施方案中,与 miRNA 背面相比,miRNA 模板具有 ≥ 1 个核苷酸错配,例如,与 miRNA 背面相比,miRNA 模板可具有 1、2、3、4、5 或更多个错配。错配程度还可以通过确定 miRNA 模板与 miRNA 背面的互补序列的同一性百分比来描述。例如,与 miRNA 背面的互补序列相比,miRNA 模板可具有包括约至少 70%、75%、77%、78%、79%、80%、

81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99% 或 100% 的同一性百分比。

在某些实施方案中，靶序列选自植物病原体。预计，包含针对病原体的靶序列的 miRNA 的植物或细胞具有降低的敏感性和/或提高的病原体抗性。在某些实施方案中，miRNA 由还包含可操纵地连接的启动子的核酸构建体编码。在某些实施方案中，启动子是病原体诱导型启动子。

在另一个实施方案中，提供了用于抑制靶序列的核酸构建体。所述核酸构建体编码与靶基本上互补的 miRNA。在某些实施方案中，核酸构建体还包含与编码 miRNA 的多核苷酸可操纵地连接的启动子。在某些实施方案中，设计缺乏启动子的核酸构建体，并且导入，使其在宿主基因组中整合后变得与启动子可操纵地连接。在某些实施方案中，使用重组，包括位点特异性重组来整合核酸构建体。参见例如 WO 99/25821，其以引用方式并入本文。在某些实施方案中，核酸构建体是 RNA。在某些实施方案中，核酸构建体包含至少一个重组位点，包括位点特异性重组位点。在某些实施方案中，核酸构建体包含至少一个重组位点以帮助构建体的整合、修饰或克隆。在某些实施方案中，核酸构建体包含在 miRNA 前体侧翼的两个位点特异性重组位点。在某些实施方案中，位点特异性重组位点包括 FRT 位点、lox 位点或 att 位点，包括 attB、attL、attP 或 attR 位点。参见例如 WO 99/25821 和美国专利 5,888,732、6,143,557、6,171,861、6,270,969 和 6,277,608，其以引用方式并入本文。

在第六个实施方案中，本发明涉及 DNA 表达构建体，所述 DNA 表达构建体包含与至少一个调节序列可操纵地连接的任何本文所述分离的多核苷酸。

在第七个实施方案中，本发明涉及在其基因组中包含本文所述 DNA 表达构建体的植物。这样的植物可以选自玉米、水稻、高粱、向日葵、小米、大豆、卡诺拉(canola)、小麦、大麦、燕麦、菜豆(bean)和坚果。

在第八个实施方案中，本发明涉及从在其基因组中包含本文所述 DNA 表达构建体的植物中获得的转基因种子。还包括在本发明范围内

的是在其基因组中包含本文所述DNA表达构建体的转化的植物组织或植物细胞。

在第九个实施方案中，本发明涉及从这样的转基因种子获得的副产品和后代植物。

在另一个实施方案中，核酸构建体包含分离的多核苷酸，所述分离的多核苷酸包含编码修饰的植物 miRNA 前体的多核苷酸，所述修饰的前体包含第一和第二寡核苷酸，其中至少一个第一或第二寡核苷酸是与前体异源的，其中所述第一寡核苷酸与第二寡核苷酸基本上互补，并且所述第二寡核苷酸包含与靶序列基本上互补的 miRNA，其中所述前体能够形成发夹。

在某些实施方案中，提供了包含导入的多核苷酸和/或通过本发明方法产生的细胞、植物和种子。细胞包括原核细胞和真核细胞，包括但不限于细菌细胞、酵母细胞、真菌细胞、病毒细胞、无脊椎动物细胞、脊椎动物细胞和植物细胞。本发明的植物、植物细胞和种子包括裸子植物(gymnosperms)、单子叶植物和双子叶植物，包括但不限于例如水稻、小麦、燕麦、大麦、小米、高粱、大豆、向日葵、红花、卡诺拉(canola)、苜蓿、棉花、拟南芥(*Arabidopsis*)和烟草。

本文所用的“副产品”是指由种子加工产生的任何产物、组分或材料。对玉米种仁(种子)进行湿研磨和干研磨。这两个加工的目的是分离出胚芽、胚乳和果皮(外壳)。湿研磨将玉米的化学成分分离成淀粉、蛋白、油和纤维组分。

本发明涉及用于抑制靶序列和/或验证功能的方法和组合物。本发明还涉及使用微RNA(miRNA)介导的RNA干扰(RNAi)来沉默或抑制靶序列，以在表型作用和/或特征发展方面评估功能或验证靶序列的方法。具体来说，本发明涉及包含能够诱导沉默的小核酸分子，miRNA的构建体，以及使用这些miRNA来选择性地沉默靶序列的方法。

RNA干扰是指由短干扰性RNA(siRNA)介导的动物中序列特异性转录后基因沉默的过程(Fire等人, *Nature* 391: 806 1998)。在植物中的相应过程通常称为转录后基因沉默(PTGS)或RNA沉默，并且在真菌中还称为镇压。据信转录后基因沉默过程是用于防止外来基因表达的进化保守的细胞防御机制，并且通常由不同植物区系和门所共享(Fire等人, *Trends Genet.* 15: 358 1999)。这种防止外来基因表达的保护作用

可能已经发展以响应源自病毒感染或源自转座子随机整合到宿主基因组内的双链 RNA (dsRNAs)的生成,是经由特异性地破坏病毒基因组 RNA 的同源单链 RNA 的细胞反应。细胞中 dsRNA 的存在经由还没有完全表征的机制来引发 RNAi 反应。

细胞中长 dsRNA 的存在刺激称为“dicer”的核糖核酸酶 III 酶的活性。Dicer 参与 dsRNA 加工成称为短干扰性 RNA (siRNA)的 dsRNA 的短片段(Berstein 等人, Nature 409: 363 2001)和/或 pre miRNA 加工成 miRNA。源自 dicer 活性的短干扰性 RNA 的长度通常是约 21 至约 23 个核苷酸,并且包含约 19 个碱基对双链体(Elbashir 等人, Genes Dev. 15: 188 2001)。还有人提出 Dicer 参与从保守结构的前体 RNA 上切下 21-和 22-核苷酸小的时序 RNA (stRNA),所述小的时序 RNA 参与翻译控制(Hutvagner 等人, 2001, Science 293: 834)。RNAi 反应还涉及内切核酸酶复合物,通常称为 RNA 诱导沉默复合物(RISC),其介导具有与 siRNA 双链体的反义链互补的序列的单链 RNA 的裂解。靶 RNA 的裂解在与 siRNA 双链体的反义链互补的区域中间发生(Elbashir 等人, Genes Dev. 15: 188 2001)。此外, RNA 干扰还涉及小 RNA (例如微 RNA 或 miRNA)介导的基因沉默,假定是经由调节染色质结构并由此防止靶基因序列转录的细胞机制(参见,例如 Allshire, Science 297: 1818-1819 2002; Volpe 等人, Science 297: 1833-1837 2002; Jenuwein, Science 297: 2215-2218 2002; 和 Hall 等人, Science 297: 2232-2237 2002)。这样,本发明的 miRNA 分子可用于通过与 RNA 转录物相互作用或者通过与特定基因序列相互作用来介导基因沉默,其中这样的相互作用导致在转录或转录后水平上的基因沉默。

已经在多种系统中研究了 RNAi。Fire 等人, (Nature 391: 806 1998)首次秀丽隐杆线虫(*C. elegans*)中观察到 RNAi。Wianny 和 Goetz (Nature Cell Biol. 2: 70 1999)描述了在小鼠胚胎中由 dsRNA 介导的 RNAi。Hammond 等人, (Nature 404: 293 2000)描述了在用 dsRNA 转染的果蝇(*Drosophila*)细胞中的 RNAi。Elbashir 等人, (Nature 411: 494 2001)描述了通过在包括人胚胎肾和 HeLa 细胞的培养的哺乳动物细胞中导入合成 21-核苷酸 RNA 的双链体而诱导的 RNAi。

小 RNA 在控制基因表达中起重要作用。很多发育过程，包括开花的调节，是由小 RNA 控制的。现在能够通过使用在植物中产生小 RNA 的转基因构建体来以工程手段改变植物基因的基因表达。

小 RNA 似乎是通过与互补 RNA 或 DNA 靶序列碱基配对来行使功能的。当与 RNA 结合时，小 RNA 引发靶序列的 RNA 裂解或翻译抑制。当与 DNA 靶序列结合时，据信小 RNA 可介导靶序列的 DNA 甲基化。无论具体机制是什么，这些事件的后果是基因表达被抑制。

微 RNA (miRNA) 是长度为约 19 至约 24 个核苷酸(nt)的已经在动物和植物中鉴定出的非编码 RNA (Lagos-Quintana 等人, Science 294: 853-858 2001, Lagos-Quintana 等人, Curr. Biol. 12: 735-739 2002; Lau 等人, Science 294: 858-862 2001; Lee 和 Ambros, Science 294: 862-864 2001; Llave 等人, Plant Cell 14: 1605-1619 2002; Mourelatos 等人, Genes. Dev. 16: 720-728 2002; Park 等人, Curr. Biol. 12: 1484-1495 2002; Reinhart 等人, Genes. Dev. 16: 1616-1626 2002)。它们是由大小为约 70 至 200 nt 的较长前体转录物加工生成的，并且这些前体转录物能够形成稳定的发夹结构。在动物中，涉及 miRNA 前体加工的酶称为 Dicer，这是一种 RNase III-样蛋白(Grishok 等人, Cell 106: 23-34 2001; Hutvagner 等人, Science 293: 834-838 2001; Ketting 等人, Genes. Dev. 15: 2654-2659 2001)。植物也具有 Dicer-样酶，DCL1 (以前称为 CARPEL FACTORY/SHORT INTEGUMENTS1/ SUSPENSOR1)，并且最近的证据表明，其象 Dicer 一样也涉及发夹前体的加工以产生成熟 miRNA (Park 等人, Curr. Biol. 12: 1484-1495 2002; Reinhart 等人, Genes. Dev. 16: 1616-1626 2002)。此外，最近的研究已经清楚地表明，至少某些 miRNA 发夹前体最初是作为较长的聚核苷酸化转录物存在，并且在单个转录物中可以存在几个不同的 miRNA 以及相关发夹(Lagos-Quintana 等人, Science 294: 853-858 2001; Lee 等人, EMBO J 21: 4663-4670 2002)。最近的研究还测定了从 dsRNA 产物的 miRNA 链选择，所述 dsRNA 产物是通过 DICER 加工发夹而产生的(Schwartz 等人, 2003, Cell 115: 199-208)。似乎所加工的 dsRNA 的两个末端的稳定性(即 G: C vs. A: U 含量，和/或错配)影响链选择，低稳定性末端更易于通过解旋酶解旋。将在低稳定性末端的 5'末端链掺入到 RISC 复合物内，而另一条链降解。

在动物中,有直接的证据表明特异性 miRNA 在发育中的作用。基于当产生 lin-4 和 let-7 miRNA 的基因突变时所产生的表型,已经发现,秀丽隐杆线虫(*C. elegans*)中的 lin-4 和 let-7 miRNA 控制时序发育(Lee 等人, *Cell* 75: 843-854 1993; Reinhart 等人, *Nature* 403:901-906 2000)。此外,这两种 miRNA 都表现出与其在发育定时中的作用一致的时序表达模式。其他动物 miRNA 表现出发育调节的表达模式,是时序和组织特异性的(Lagos-Quintana 等人, *Science* 294: 853-853 2001, Lagos-Quintana 等人, *Curr. Biol.* 12: 735-739 2002; Lau 等人, *Science* 294: 858-862 2001; Lee 和 Ambros, *Science* 294: 862-864 2001),导致这样的假设,在很多情况下,miRNA 可能涉及重要发育过程的调节。同样,在植物中,很多 miRNA 的差异表达模式意味着其在发育中的作用(Llave 等人, *Plant Cell* 14: 1605-1619 2002; Park 等人, *Curr. Biol.* 12: 1484-1495 2002; Reinhart 等人, *Genes. Dev.* 16: 1616-1626 2002)。然而,miRNA 的发育作用尚未在植物中直接证实,因为迄今为止没有关于特异性植物 miRNA 相关发育表型的报道。

微 RNA 似乎通过与位于由这些基因产生的转录物中的互补序列结合来调节靶基因。对于 lin-4 和 let-7,靶位点位于靶 mRNA 的 3' UTR 中(Lee 等人, *Cell* 75: 843-854 1993; Wightman 等人, *Cell* 75: 855-862 1993; Reinhart 等人, *Nature* 403: 901-906 2000; Slack 等人, *Mol. Cell* 5: 659-669 2000),并且在 lin-4 和 let-7 miRNA 与其靶位点之间有几个错配。结合 lin-4 或 let-7 miRNA 似乎引起由靶 miRNA 编码的蛋白的稳态水平的下调,而不影响自身的转录物(Olsen 和 Ambros, *Dev. Biol.* 216: 671-680 1999)。另一方面,最近的证据表明,在某些情况下,miRNA 可以引起靶位点内的靶转录物的特异性 RNA 裂解,并且该裂解步骤似乎需要 miRNA 与靶转录物之间具有 100% 的互补性(Hutvagner 和 Zamore, *Science* 297: 2056-2060 2002; Llave 等人, *Plant Cell* 14: 1605-1619 2002)。miRNA 似乎可能进入至少两条靶基因调节途径:当靶互补性 < 100% 时,蛋白下调,当靶互补性是 100% 时, RNA 裂解。进入 RNA 裂解途径的微 RNA 与在动物中 RNA 干扰(RNAi)期间以及在植物中转录后基因沉默(PTGS)期间产生的 21-25 nt 短干扰性 RNAs (siRNAs)类似(Hamilton 和 Baulcombe 1999; Hammond 等人, 2000; Zamore 等人, 2000; Elbashir 等人, 2001),并且可能掺入与在 RNAi

情况中观察到的复合物类似或相同的 RNA-诱导的沉默复合物(RISC)内。

用生物信息学鉴定 miRNA 的靶在动物中没有成功,这可能是因为动物 miRNA 与它们的靶具有低程度的互补性。另一方面,生物信息学方法已经成功地用于预测植物 miRNA 的靶(Llave 等人, Plant Cell 14: 1605-1619 2002; Park 等人, Curr. Biol. 12: 1484-1495 2002; Rhoades 等人, Cell 110: 513-520 2002),因此,似乎植物 miRNA 与其推定靶的整个互补性高于动物 miRNA。植物 miRNA 的这些预测靶当中的大部分编码参与植物发育模式或细胞分化的转录因子家族的成员。虽然如此,对于任何植物 miRNA,没有直接证实生物功能。虽然 Llave 等人, (Science 297: 2053-2056 2002)已经表明了 SCARECROW-样转录因子的转录物是拟南芥(*Arabidopsis*) miRNA mir171 的靶,但是这些研究是在异源物种中进行的,并且没有报道与 mir171 有关的植物表型。

本文提供的方法可以在其中可以利用转化方法的任何生物体中实施,并且对于生物体,至少某些序列信息是有关靶序列的信息,或者有关在所关注靶序列侧翼的区域的信息。应当理解,两个或更多个序列可以通过顺序转化、用一个以上定向载体共转化、或者构建包含一个以上 miRNA 序列的 DNA 构建体来导向。本发明方法还可以通过组合核酸文库构建来实施以产生针对随机靶序列的 miRNA 文库。miRNA 文库可用于高通量筛选以进行基因功能验证。

所关注序列的一般种类包括例如涉及调节或信息的基因,例如锌指、转录因子、同源异型基因、或细胞周期和细胞死亡调节剂,设计通讯例如激酶,以及涉及持家的基因,例如热激蛋白。

靶序列还包括编码区和非编码区例如启动子、增强子、终止子、内含子等,其可以被修饰以改变所关注基因的表达。例如,可以将内含子序列加到 5' 区域以增加聚集的成熟信息的量(参见例如 Buchman 和 Berg, Mol. Cell Biol. 8: 4395-4405 (1988); 和 Callis 等人, Genes Dev. 1: 1183-1200 (1987))。

靶序列可以是内源性序列,或者可以是导入的异源序列或转基因。本发明方法可用于改变转基因的调节或表达,或用于除去转基因或其他导入的序列例如导入的位点特异性重组位点。靶序列还可以是来自病原体的序列,例如,靶序列可以来自植物病原体例如病毒、霉菌或

真菌，昆虫或线虫。miRNA 可以在植物中表达，这样在感染或侵染之后，植物将靶向病原体并且给植物赋予一定程度的抗性。

在植物中，靶序列的其他种类包括影响农学特征、抗昆虫性、抗病性、抗除草剂性、不育、谷粒特征和商品的基因。所关注基因应当还包括涉及油、淀粉、糖类或营养物代谢的那些，以及影响例如种仁大小、蔗糖载量等的那些。谷粒质量在特征中反应，例如油的水平和类型，饱和与不饱和，必需氨基酸的质量和量，以及纤维素水平。例如，可以抑制植酸生物合成途径的基因以产生高可利用磷表型。参见例如：在 WO 02/059324 中公开的植酸生物合成酶，包括肌醇多磷酸激酶-2 多核苷酸，在 WO 03/027243 中公开的肌醇 1,3,4-三磷酸 5/6-激酶多核苷酸，以及在 WO 99/05298 中公开的肌醇 1-磷酸合成酶和其他植酸生物合成多核苷酸，所有这些专利都以引用方式并入本文。可以抑制木质化途径中的基因以提高消化性或能量利用率。可以抑制影响细胞周期或细胞死亡的基因以抑制生长或应激反应。可以抑制影响 DNA 复制和/或重组的基因以提高遗传变异性。可以抑制影响开花时间的基因以及影响能育性的基因。可以抑制任何靶序列以评估或证实其在特定特征或表型中的作用，或切开分子、调节、生化或蛋白组学(proteomic)途径或网络。

可使用多种启动子，可以基于所希望的结果来选择这些启动子。应当认识到，通过在植物表达盒中使用不同启动子来提高不同应用以调节 miRNA 表达的时间、位置和/或水平。如果需要的话，这样的植物表达盒可以含有启动子调节区(例如赋予诱导型、组成型、环境或发育调节的、或细胞或组织特异性/选择性表达的调节区)，转录起始开始位点、核糖体结合位点、RNA 加工信号、转录终止位点和/或聚腺苷酸化信号。

可以采用组成型、组织优先型或诱导型启动子。组成型启动子的实例包括花椰菜花叶病毒(CaMV) 35S 转录起始区，源自根癌土壤杆菌(*Agrobacterium tumefaciens*)的 T-DNA 的 1'-或 2'-启动子，遍在蛋白 1 启动子，Smas 启动子，肉桂醇脱氢酶启动子(美国专利 5,683,439)，Nos 启动子，pEmu 启动子，rubisco 启动子，GRP1-8 启动子，以及来自本领域技术人员已知的各种植物基因的其他转录起始区。如果需要低水平的表达，可以使用弱启动子。弱组成型启动子包括例如 Rsyn7 启动

子的核心启动子(WO 99/43838 和美国专利 6,072,050), 核心 35S CaMV 启动子等。其他组成型启动子包括例如美国专利 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; 和 5,608,142。还参见美国专利 6,177,611, 其以引用方式并入本文。

诱导型启动子的实例是可以通过低氧或寒冷胁迫诱导的 *Adh1* 启动子, 可通过热胁迫诱导的 *Hsp70* 启动子, *PPDK* 启动子和磷酸烯醇丙酮酸羧化酶(*pep*carboxylase)启动子, 二者都是可通过光诱导的。还有用的启动子有, 是可化学诱导的启动子, 例如是安全剂诱导的 *In2-2* 启动子(美国专利 5,364,780), 是雌激素诱导的 *ERE* 启动子, 以及是植物生长素诱导的, 并且是绒毡层(*tapetum*)特异性的, 但是在愈伤组织中也有活性的 *Axig1* 启动子(PCT US01/22169)。

处于发育控制下的启动子的实例包括优先在一些组织例如叶、根、果实、种子或花中启动转录的启动子。示例性启动子是花药特异性启动子 5126 (美国专利 5,689,049 和 5,689,051)。种子优先的启动子的实例包括但不限于 27 kD γ 玉米醇溶蛋白启动子和蜡状启动子, Boronat, A. 等人, (1986) *Plant Sci.* 47: 95-102; Reina, M. 等人, *Nucl. Acids Res.* 18(21): 6426; 和 Kloesgen, R.B. 等人, (1986) *Mol. Gen. Genet.* 203: 237-244。在胚芽、果皮和胚乳中表达的启动子公开在美国专利 6,225,529 和 PCT 出版 WO 00/12733 中。这些专利的公开内容全文以引用方式并入本文。

在某些实施方案中, 由诱导型启动子, 特别是由病原体诱导型启动子表达基因是有益的。这样的启动子包括源自发病机理相关蛋白(PR 蛋白)的那些, 其是在被病原体感染后诱导的, 所述蛋白是例如 PR 蛋白、SAR 蛋白、 β -1,3-葡聚糖酶, 壳多糖酶等。参见例如 Redolfi 等人, (1983) *Neth. J. Plant Pathol.* 89: 245-254; Uknes 等人, (1992) *Plant Cell* 4: 645-656; 和 Van Loon (1985) *Plant Mol. Virol.* 4: 111-116。还参见 WO 99/43819, 其以引用方式并入本文。

所关注的启动子是在病原体感染部位或接近病原体感染部位局部表达的启动子。参见例如 Marineau 等人, (1987) *Plant Mol. Biol.* 9: 335-342; Matton 等人, (1989) *Molecular Plant-Microbe Interactions* 2: 325-331; Somsisch 等人, (1986) *Proc. Natl. Acad. Sci. USA* 83: 2427-2430; Somsisch 等人, (1988) *Mol. Gen. Genet.* 2: 93-98; 和 Yang

(1996) *Proc. Natl. Acad. Sci. USA* 93: 14972-14977。还参见 Chen 等人, (1996) *Plant J.* 10: 955-966; Zhang 等人, (1994) *Proc. Natl. Acad. Sci. USA* 91: 2507-2511; Warner 等人, (1993) *Plant J.* 3: 191-201; Siebertz 等人, (1989) *Plant Cell* 1: 961-968; 美国专利 5,750,386 (线虫诱导型); 以及其中所引用的参考文献。特别关注的是有关玉米 PRms 基因的诱导型启动子, 其表达是由病原体串珠镰孢菌(*Fusarium moniliforme*)诱导的(参见例如 Cordero 等人, (1992) *Physiol. Mol. Plant Path.* 41: 189-200)。

此外, 在病原体发现经由伤口或昆虫损害而找到进入植物的入口时, 伤口诱导型启动子可用于多核苷酸的构建。这样的伤口诱导型启动子包括马铃薯蛋白酶抑制剂(pin II)基因(Ryan (1990) *Ann. Rev. Phytopath.* 28: 425-449; Duan 等人, (1996) *Nature Biotech.* 14: 494-498); wun1 和 wun2, 美国专利 5,428,148; win1 和 win2 (Stanford 等人, (1989) *Mol. Gen. Genet.* 215: 200-208); 系统素(McGurl 等人, (1992) *Science* 225: 1570-1573); WIP1 (Rohmeier 等人, (1993) *Plant Mol. Biol.* 22: 783-792; Eckelkamp 等人, (1993) *FEBS Lett.* 323: 73-76); MPI 基因(Corderok 等人, (1994) *Plant J.* 6(2): 141-150); 等等, 其以引用方式并入本文。

化学调节的启动子可用于通过施用外来化学调节剂来调节基因在植物中的表达。根据这一目标, 启动子可以是化学诱导型启动子, 其中施用化学物质来诱导基因表达, 或化学抑制型启动子, 其中施用化学物质来抑制基因表达。化学诱导型启动子是本领域已知的, 并且包括但不限于玉米 In2-2 启动子, 其是通过苯磺酰胺除草剂安全剂来激活的, 玉米 GST 启动子, 其是通过用作芽前除草剂的疏水亲电性化合物而激活的, 以及烟草 PR-1a 启动子, 其是通过水杨酸激活的。所关注的其他化学调节的启动子包括甾类化合物反应性启动子(参见例如糖皮质激素诱导型启动子, Schena 等人, (1991) *Proc. Natl. Acad. Sci. USA* 88: 10421-10425 和 McNellis 等人, (1998) *Plant J.* 14(2): 247-257), 以及四环素诱导型和四环素抑制型启动子(参见例如 Gatz 等人, (1991) *Mol. Gen. Genet.* 227: 229-237, 和美国专利 5,814,618 和 5,789,156), 其以引用方式并入本文。

组织优先的启动子可用于定向提高特定植物组织内的序列的表达。组织优先的启动子包括 Yamamoto 等人, (1997) *Plant J.* 12(2):

255-265; Kawamata 等人, (1997) *Plant Cell Physiol.* 38(7): 792-803; Hansen 等人, (1997) *Mol. Gen Genet.* 254(3): 337-343; Russell 等人, (1997) *Transgenic Res.* 6(2): 157-168; Rinehart 等人, (1996) *Plant Physiol.* 112(3): 1331-1341; Van Camp 等人, (1996) *Plant Physiol.* 112(2): 525-535; Canevascini 等人, (1996) *Plant Physiol.* 112(2): 513-524; Yamamoto 等人, (1994) *Plant Cell Physiol.* 35(5): 773-778; Lam (1994) *Results Probl. Cell Differ.* 20: 181-196; Orozco 等人, (1993) *Plant Mol Biol.* 23(6): 1129-1138; Matsuoka 等人, (1993) *Proc Natl. Acad. Sci. USA* 90(20): 9586-9590; 和 Guevara-Garcia 等人, (1993) *Plant J.* 4(3): 495-505。如果需要的话, 可修饰这样的启动子以进行弱表达。

叶优先的启动子是本领域已知的。参见例如 Yamamoto 等人, (1997) *Plant J.* 12(2): 255-265; Kwon 等人, (1994) *Plant Physiol.* 105: 357-67; Yamamoto 等人, (1994) *Plant Cell Physiol.* 35(5): 773-778; Gotor 等人, (1993) *Plant J.* 3: 509-18; Orozco 等人, (1993) *Plant Mol. Biol.* 23(6): 1129-1138; 和 Matsuoka 等人, (1993) *Proc. Natl. Acad. Sci. USA* 90(20): 9586-9590。此外, 可以使用 *cab* 和 *rubisco* 的启动子。参见例如 Simpson 等人, (1958) *EMBO J* 4: 2723-2729 和 Timko 等人, (1988) *Nature* 318: 57-58。

根优先的启动子是已知的, 并且可以选自文献中记载的很多启动子或者从不同相容物种中重新分离的启动子。参见例如 Hire 等人, (1992) *Plant Mol. Biol.* 20(2): 207-218 (大豆根特异性谷氨酰胺合成酶基因); Keller 和 Baumgartner (1991) *Plant Cell* 3(10): 1051-1061 (法国菜豆的 GRP 1.8 基因的根特异性控制元件); Sanger 等人, (1990) *Plant Mol. Biol.* 14(3): 433-443 (根癌土壤杆菌(*Agrobacterium tumefaciens*)的甘露氨酸合酶(MAS)基因的根特异性启动子); 和 Miao 等人, (1991) *Plant Cell* 3(1): 11-22 (编码胞质谷氨酰胺合成酶(GS)的全长 cDNA 克隆, 其在大豆的根和根瘤中表达)。还参见 Bogusz 等人, (1990) *Plant Cell* 2(7): 633-641, 其中描述了两种根特异性启动子, 它们是从得自固氮非豆科(nonlegume) *Parasponia andersonii* 以及相关非固氮非豆科(nonlegume) 山麻黄(*Trema tomentosa*)的血红蛋白基因分离的。将这些基因的启动子与 β -葡萄糖醛酸酶报道基因连接, 并且导入非豆科(nonlegume)烟草(*Nicotiana tabacum*)和豆科百脉根(*Lotus corniculatus*)内, 在这两种情况

下,都保持了根特异性启动子活性。Leach 和 Aoyagi (1991)描述了他们对于发根农杆菌(*Agrobacterium rhizogenes*)的高度表达的 rolC 和 rolD 根诱导基因的启动子的分析(参见 Plant Science (Limerick) 79(1): 69-76)。他们推断,增强子和组织特异性 DNA 决定子在这些启动子中解离。Teeri 等人,(1989)使用与 lacZ 的基因融合以表明编码章鱼碱合酶的农杆菌属(*Agrobacterium*)菌 T-DNA 基因在根尖的表皮中具有特别活性,以及 TR2'基因在完整植物中是根特异性的,并且通过叶组织中的伤害来刺激,这是用来与杀虫或杀幼虫基因一起使用的特别期望的特征组合(参见 EMBO J. 8(2): 343-350)。与 nptII (新霉素磷酸转移酶 II)融合的 TR1'基因表现出类似特征。另外的根优先的启动子包括 VfENOD-GRP3 基因启动子(Kuster 等人,(1995) Plant Mol. Biol. 29(4): 759-772); 和 rolB 启动子(Capana 等人,(1994) Plant Mol. Biol. 25(4): 681-691。还参见美国专利 5,837,876; 5,750,386; 5,633,363; 5,459,252; 5,401,836; 5,110,732; 和 5,023,179。云扁豆蛋白(Murai 等人,(1983) Science 23: 476-482 和 Sengopta-Gopalen 等人,(1988) PNAS 82: 3320-3324。

转化方案以及用于将核苷酸序列导入植物内的方案可以根据被定向转化的植物或植物细胞的类型,例如单子叶植物或双子叶植物而改变。导入 DNA 构建体的合适的方法包括微注射(Crossway 等人,(1986) Biotechniques 4: 320-334; 和美国专利 6,300,543),性杂交,电穿孔(Riggs 等人,(1986) Proc. Natl. Acad. Sci. USA 83: 5602-5606),农杆菌属菌介导的转化(Townsend 等人,美国专利 5,563,055; 和美国专利 5,981,840),直接基因转移(Paszowski 等人,(1984) EMBO J. 3: 2717-2722),和弹道离子加速(参见例如 Sanford 等人,美国专利 4,945,050; Tomes 等人,美国专利 5,879,918; Tomes 等人,美国专利 5,886,244; Bidney 等人,美国专利 5,932,782; Tomes 等人,(1995) "Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment," Plant Cell, Tissue, and Organ Culture: Fundamental Methods, ed. Gamborg 和 Phillips (Springer-Verlag, Berlin); 和 McCabe 等人,(1988) Biotechnology 6: 923-926)。还参见 Weissinger 等人,(1988) Ann. Rev. Genet. 22: 421-477; Sanford 等人,(1987) Particulate Science and Technology 5: 27-37 (洋葱); Christou 等人,(1988) Plant Physiol. 87: 671-674 (大豆); Finer 和

McMullen (1991) *In Vitro Cell Dev. Biol.* 27P: 175-182 (大豆); Singh 等人, (1998) *Theor. Appl. Genet.* 96: 319-324 (大豆); Datta 等人, (1990) *Biotechnology* 8: 736-740 (水稻); Klein 等人, (1988) *Proc. Natl. Acad. Sci. USA* 85: 4305-4309 (玉米); Klein 等人, (1988) *Biotechnology* 6: 559-563 (玉米); Tomes, 美国专利 5,240,855; Buising 等人, 美国专利 5,322,783 和 5,324,646; Klein 等人, (1988) *Plant Physiol.* 91: 440-444 (玉米); Fromm 等人, (1990) *Biotechnology* 8: 833-839 (玉米); Hooykaas-Van Slogteren 等人, (1984) *Nature (London)* 311: 763-764; Bowen 等人, 美国专利 5,736,369 (谷类); Bytebier 等人, (1987) *Proc. Natl. Acad. Sci. USA* 84: 5345-5349 (百合科); De Wet 等人, (1985) *The Experimental Manipulation of Ovule Tissues*, ed. Chapman 等人, (Longman, New York), pp. 197-209 (花粉); Kaeppler 等人, (1990) *Plant Cell Reports* 9: 415-418 and Kaeppler 等人, (1992) *Theor. Appl. Genet.* 84: 560-566 (颈须介导的转化); D'Halluin 等人, (1992) *Plant Cell* 4: 1495-1505 (电穿孔); Li 等人, (1993) *Plant Cell Reports* 12: 250-255 and Christou and Ford (1995) *Annals of Botany* 75: 407-413 (rice); Osjoda 等人, (1996) *Nature Biotechnology* 14: 745-750 (经由根癌土壤杆菌的玉米); 和美国专利 5,736,369 (分生组织转化), 所有这些文献都以引用方式并入本文。

可以通过将植物与病毒或病毒核酸接触来将核苷酸构建体导入植物内。通常, 这样的方法包括将本发明的核苷酸构建体掺入病毒 DNA 或 RNA 分子内。此外, 应当认识到, 有用的启动子包括用于通过病毒 RNA 聚合酶转录的启动子。涉及病毒 DNA 或 RNA 分子的用于将核苷酸构建体导入植物内并且表达其中编码的蛋白的方法是本领域已知的。参见例如美国专利 5,889,191、5,889,190、5,866,785、5,589,367 和 5,316,931; 这些专利以引用方式并入本文。

在某些实施方案中, 瞬时表达可能是期望的。在这些情况下, 可使用标准瞬时转化技术。这样的方法包括但不限于病毒转化方法, 和 DNA 或 RNA 的微注射, 以及本领域众所周知的其他方法。

可根据常规方法, 将得自已经稳定掺入核苷酸序列的植物的细胞生长成植物。参见例如 McCormick 等人, (1986) *Plant Cell Reports* 5: 81-84。然后可让这些植物生长, 并且用相同的转化株或不同的株授粉, 所得杂合体具有所希望的表型特征的组成型表达, 所述表型特征是由

包含在靶位点或转移盒内的所关注核苷酸序列盒/或遗传标记赋予的。可以生长两代或多代以保证所希望的表型特征的表达稳定地保持和遗传,然后收获种子以保证所希望的表型特征的表达已经实现。

在第十个实施方案中,本发明涉及改变植物中稳定导入的核苷酸序列的表达的方法,所述方法包括:

a)制备 DNA 表达构建体,所述构建体包含稳定导入的核苷酸序列以及至少一个能够与本发明分离的多核苷酸杂交的序列;

b)用部分(a)的 DNA 表达构建体转化植物;和

c)选择转化的植物,所述转化的植物在其基因组中包含部分(a)的 DNA 表达构建体,并且当与用部分(a)的 DNA 表达构建体的修饰变型转化的植物相比时,具有稳定导入的核苷酸序列的变化的表达,其中所述修饰的构建体缺乏能够与本发明分离的多核苷酸杂交的序列。

包含 DNA 构建体的细胞和/或植物的初始鉴定和选择可以通过使用标记基因来促进。如果有用于鉴定重组体的敏感方法,例如,如果定向基因修饰可以通过 PCR 分析而容易地检测,或者如果其导致一些表型,则基因导向可以不用选择来进行。然而,在大多数情况下,基因导向事件的鉴定是通过使用标记来促进。有用的标记包括阳性和阴性选择性标记以及促进筛选的标记,例如病毒标记。选择性标记包括携带针对抗生素的抗药性的基因,例如壮观霉素(例如 aada 基因, Svab 等人, 1990 Plant Mol. Biol. 14: 197),链霉素(例如 aada 或 SPT, Svab 等人, 1990 Plant Mol. Biol. 14: 197; Jones 等人, 1987 Mol. Gen. Genet. 210: 86),卡那霉素(例如 nptII, Fraley 等人, 1983 PNAS 80: 4803),潮霉素(例如 HPT, Vanden Elzen 等人, 1985 Plant Mol. Biol. 5: 299),庆大霉素(Hayford 等人, 1988 Plant Physiol. 86: 1216),腐草霉素, zeocin 或博莱霉素(Hille 等人, 1986 Plant Mol. Biol. 7: 171),或携带针对除草剂的抗药性的基因,例如 phosphinothricin (bar 基因),或磺酰脲类(乙酰乳酸合酶(ALS)) (Charest 等人, (1990) Plant Cell Rep. 8: 643),在不完全培养基上满足生长需求的基因例如酵母中的 HIS3、LEU2、URA3、LYS2 和 TRP1 基因,以及本领域已知的其他这样的基因。阴性选择性标记包括胞嘧啶脱氨基酶(codA) (Stougaard 1993 Plant J. 3: 755-761), tms2 (DePicker 等人, 1988 Plant Cell Rep. 7: 63-66),硝酸还原酶(Nussame 等人, 1991 Plant J. 1: 267-274), SU1 (O'Keefe 等人, 1994 Plant

Physiol. 105: 473-482), 得自农杆菌的 Ti 质粒的 aux-2 和胸苷激酶。可筛选的标记包括荧光蛋白例如绿荧光蛋白(GFP) (Chalfie 等人, 1994 Science 263: 802; US 6,146,826; US 5,491,084; 和 WO 97/41228), 报道基因酶例如 β -葡糖醛酸酶(GUS) (Jefferson R.A. 1987 Plant Mol. Biol. Rep. 5: 387; US 5,599,670; 和 US 5,432,081), β -半乳糖苷酶(lacZ), 碱性磷酸酶(AP), 谷胱甘肽 S-转移酶(GST)和荧光素酶(US 5,674,713; 和 Ow 等人, 1986 Science 234(4778): 856-859), 可视标记例如花青苷如 CRC (Ludwig 等人, (1990) Science 247(4841): 449-450) R 基因家族(例如 Lc、P、S)、A、C、R-nj, 果蝇(Drosophila)中的身体和/或眼睛色基因, 哺乳动物系统中的毛色基因, 以及本领域已知的其他基因。

可使用一种或多种标记以选择和筛选基因导向事件。用于基因分裂的一种常用方法包括使用靶修饰多核苷酸, 其中靶被没有启动子的选择性标记分裂。因为选择性标记缺乏启动子, 所以随机整合事件不太可能导致基因转录。基因导向时间将使得选择性标记处于靶基因的启动子的控制下。通过针对选择性标记的表达的选择来鉴定基因靶向事件。另一常用方法使用阳性-阴性选择方案。该方案使用两种选择性标记, 一种标记赋予抗性(R+), 而与之偶联的另一种标记赋予敏感性(S+), 每种标记都具有启动子。当该多核苷酸随机插入时, 所产生的表型是 R+/S+。当产生基因导向事件时, 这两种标记是未偶联的, 并且所产生的表型是 R+/S-。使用阳性-阴性选择的实例参见 Thykjaer 等人, (1997) Plant Mol. Biol. 35: 523-530; 和 WO 01/66717, 其以引用方式并入本文。

实施例

下列非限制性实施例是为了举例说明本发明。虽然为了清楚地理解而已经通过举例说明和实例较详细地描述了本发明, 但是显然可以在权利要求书范围内实施一些改变和修饰。

实施例 1

植物转化

本实施例描述了可用于将多核苷酸或多肽导入植物细胞内的方法。

A. 玉米颗粒介导的 DNA 递送

可以将DNA构建体导入能够在合适的玉米培养基中生长的玉米细胞内。这样的感受态细胞可得自玉米悬浮培养物，在固体培养基上的愈伤组织培养物，新分离的未成熟胚芽或分生组织细胞。可以使用 Hi-II 基因型的未成熟胚芽作为靶细胞。在授粉后大约 10 天收获穗，并且从种仁分离出 1.2-1.5mm 未成熟胚芽，并且放置使得盾片侧朝下在玉米培养物上。

从穗上收获之后 18-72 小时，将未成熟胚芽轰击。在轰击之前 6-18 小时，将未成熟胚芽放置在具有另外渗透性(osmoticum)的培养基上(MS 基础培养基，Musashige 和 Skoog, 1962, *Physiol. Plant* 15: 473-497, 含有 0.25 M 山梨醇)。将在高渗透性培养基上的胚芽用作轰击靶，并且在轰击后，在该培养基上再放置 18 小时。

对于粒子轰击，使用标准 CaCl_2 -亚精胺化学(参见例如 Klein 等人, 1987, *Nature* 327: 70-73)将质粒 DNA (上述)在 1.8 mm 钨粒子上沉淀。使用 DuPont Helium Gun (Lowe 等人, 1995, *Bio/Technol* 13: 677-682)，将每个平板以 600 PSI 轰击一次。对于用于玉米未成熟胚芽分离、愈伤组织引发、愈伤组织增长以及植物再生的典型培养基制剂，参见 Armstrong, C., 1994, In “The Maize Handbook”, M. Freeling 和 V. Walbot, eds. Springer Verlag, NY, pp 663-671。

在粒子轰击之后 1-7 天内，将胚芽移动到含有 3mg/L 选择试剂 bialaphos 的 N6-基培养基上。每两周将胚芽以及晚期愈伤组织转移到新选择平板上。对于所需表型，筛选从未成熟胚芽上发育的愈伤组织。6-8 周后，回收转化的愈伤组织。

B. 大豆转化

在放置于旋转摇动器上的 250 ml 锥形瓶内，将大豆胚芽发生悬浮液培养物在 35mL 液体培养基 SB196 或 SB172 培养基中于 150rpm 和 26℃ 保持，采用冷的白色荧光，以 16: 8 小时白天/黑夜光周期，和 30-35uE/m²s 的光强度。每两周，通过将大约 35mg 组织接种到 35mL 新鲜液体培养基内来将培养物传代培养。或者，将培养物引发，并且在 6 孔 Costar 平板上保持。

制备如下所述的 SB 172 培养基: (每升), 1 瓶 Murashige and Skoog Medium (Duchefa # M 0240), 1mL B5 维生素 1000 × 贮备液, 1mL 2,4-D

贮备液(Gibco 11215-019), 60g 蔗糖, 2g MES, 0.667g 无水 L-天冬酰胺(GibcoBRL 11013-026), pH 5.7。制备如下所述的 SB 196 培养基: (每升) 10mL MS FeEDTA, 10ml MS 硫酸盐, 10mL FN-Lite 卤化物, 10mL FN-Lite P, B, Mo, 1mL B5 维生素 1000×贮备液, 1mL 2,4-D 贮备液(Gibco 11215-019), 2.83g KNO₃, 0.463g (NH₄)₂SO₄, 2g MES, 1g 无水天冬酰胺粉末(Gibco 11013-026), 10g 蔗糖, pH 5.8。如下所述制备浓度为 10mg/mL 的 2,4-D 贮备液: 将 2,4-D 溶解在 0.1 N NaOH 中, 过滤灭菌, 并且在 -20℃ 贮存。如下所述制备 B5 维生素 1000×贮备液: (每 100mL) - 在 -20℃ 的贮存等份式样, 10g 肌醇, 100mg 烟酸, 100mg 盐酸吡哆醇, 1g 硫胺。

通过粒子枪轰击方法(Klein 等人, 1987 Nature 327: 70)用各种制粒转化大豆胚芽发生悬浮液培养物。为了制备用于轰击的组织, 将大约两烧瓶悬浮液培养物组织放置于在底部含有 1 片无菌滤纸以帮助吸收水分的 60×20mm 培养皿中, 由于其最近的传代培养, 所述悬浮液培养物组织已经经由大约 1-2 周来回收。将组织(即大小为约 3-5mm 的悬浮液簇)越过每一培养基均匀地铺开。在轰击之间, 用吸移管将残余液体从组织中除去, 或者让残余液体蒸发以除去过量水分。每一个实验, 轰击 4-6 个组织平板。每一个平板取自两个烧瓶。

为了制备用于轰击的金离子, 将 30mg 金在乙醇中洗涤, 离心, 并且重悬在 0.5mL 无菌水中。对于用于轰击的每个制粒组合(处理), 制备单独的微离心管, 用 50μl 上面制备的金离子开始。还向每一个管中加入以下物质: 5μl 质粒 DNA (以 1μg/μl)、50μl CaCl₂ 和 20μl 0.1 M 亚精胺。将该混合物在涡旋摇动器上搅拌 3 分钟, 然后使用设置为 14,000 RPM 的微离心管离心 10 秒。将该悬浮液倾析, 并且将具有粘附沉淀 DNA 的金离子用 400 μl 等份式样的乙醇洗涤两次(在每次洗涤之间进行如上所述的短暂离心)。将 100% 乙醇的最终体积调节至 40μl, 把该离子/DNA 悬浮液在冰上保持直至用于轰击。

在施加离子/DNA 悬浮液之前, 将管在超声器浴中短暂浸泡以分散离子, 然后将 5μL DNA 制备物吸移到每个飞盘上, 并且让其干燥。然后将飞盘置于 DuPont Biolistics PDS1000/HE 内。使用用于将离子介导的 DNA 递送到大豆悬浮液簇内的 DuPont Biolistic PDS1000/HE 装置, 采用以下设定。膜破裂压力为 1100psi。将室抽空至 27-28 英寸汞

柱的真空。将组织放置在距离保留/停止筛网大约 3.5 英寸处(从底物开始的第三个架子)。每个平板轰击两次，并且在射击之间使用无菌铲将组织簇重排。

轰击后，将组织重悬在液体培养基中，每个平板用 2 个烧瓶分开，这两个烧瓶装有新鲜的 SB196 或 SB172 培养基并且如上所述进行培养。轰击后 4-7 天，将培养基用含有选择试剂的新鲜培养基替换。对于前四周，将选择培养基每周用新鲜培养基替换一次，并且在第 6 周再用新鲜培养基替换一次。4 周后，如果细胞密度和培养基和培养基浊度很高，可能需要每周一次新鲜培养基替换。

轰击后 4-8 周，可观察到从未转化的坏死胚芽发生簇中生长出绿色转化组织。取出分离的绿色组织，并且接种到含有液体培养基的 6-孔微量滴定板中，以产生克隆繁殖的、转化的胚芽发生悬浮液培养物。

将每一胚芽发生簇置于 Costar 6-孔平板的一个孔中，所述孔具有包含选择试剂的 5mL 新鲜 SB196 培养基。将培养物保持 2-6 周，每两周进行一次新鲜培养基替换。当获得足够组织时，将一部分存活的转化克隆转移到另一个 6-孔平板上进行传代培养，作为备份以进行抗污染保护。

为了促进体外成熟，将转化的胚芽发生簇从液体 SB196 中取出，并且在固体琼脂培养基 SB 166 上放置两周。将 2 - 4 mm 大小的组织块以 10-15 个簇/平板的组织密度铺在平板上。将平板在弥漫的低光(< 10 μ E)中于 26 $\pm$ 1 $^{\circ}$ C 培养。两周后，将组织簇转移到 SB 103 培养基上进行传代培养，培养 3-4 周。

制备如下所述的 SB 166: (每升)，1 pkg. MS 盐(Gibco/ BRL - Cat# 11117-017)，1mL B5 维生素 1000 $\times$ 贮备液，60g 麦芽糖，750mg MgCl₂ 六水合物，5g 活性炭，pH 5.7，2g gelrite。制备如下所述的 SB 103 培养基: (每升)，1 pkg. MS 盐(Gibco/BRL - Cat# 11117-017)，1mL B5 维生素 1000 $\times$ 贮备液，60g 麦芽糖，750mg MgCl₂ 六水合物，pH 5.7，2g gelrite。成熟 5-6 周后，通过将胚芽置于具有 1cm² 部分 SB103 培养基的 100 $\times$ 15 培养皿内来将个体胚芽干燥，以产生具有足以促进部分干燥，但是不死亡的湿度的室。

每个平板将大约 25 个胚芽干燥。将平板用几层 parafilm 密封，并且再次置于低光条件下。干燥步骤的最佳持续时间是凭经验确定的，

并且其取决与每个平板放置的胚芽的大小和量。例如，小的胚芽或很少的胚芽/平板需要较短的干燥时间，而大的胚芽或很多的胚芽/平板需要较长的干燥时间。最好在大约 3 天后检查胚芽，但是充分的干燥最可能需要 5-7 天。在该期间，胚芽的大小将减小。

将干燥的胚芽植入 SB 71-1 或 MSO 培养基中，让它们在上述有关悬浮液培养物的相同培养条件下发芽。当小植株具有两个完全发育的三叶型叶子时，将发芽并且生根的胚芽转移到无菌土壤中，并且用 MS 肥料浇水。让植株生长至成熟期以进行种子收集和分析。让健康、能繁殖的植株在温室中生长。

制备如下所述的 SB 71-1: 1 瓶 Gamborg 的 B5 盐 w/蔗糖 (Gibco/BRL - Cat# 21153-036), 10g 蔗糖, 750mg MgCl₂ 六水合物, pH 5.7, 2g gelrite。制备如下所述的 MSO 培养基: 1 pkg Murashige and Skoog 盐(Gibco 11117-066), 1mL B5 维生素 1000 × 贮备液, 30g 蔗糖, pH 5.8, 2g Gelrite。

C. 使用农杆菌属(Agrobacterium)细菌转化玉米

农杆菌属细菌介导的玉米转化基本上按照以下文献中描述的方法进行: Zhao 等人, Meth. Mol. Biol. 318: 315-323 (2006) (还参见 Zhao 等人, Mol. Breed. 8: 323-333 (2001)和 1999 年 11 月 9 日公布的美国专利 5,981,840, 以引用方式并入本文)。该转化过程涉及细菌接种, 共培养, 静息, 选择和植物再生。

1. 未成熟胚芽制备:

将未成熟胚芽从颖果上切下来, 并且放置在含有 2mL PHI-A 培养基的 2mL 胃管中。

2. 未成熟胚芽的农杆菌属细菌感染和共胚芽:

2.1 感染步骤:

用 1mL 微吸移管将(1)的 PHI-A 培养基取出, 并且加入 1mL 农杆菌属细菌悬浮液。将该管轻轻地倒置以混合。将该混合物在室温培养 5 分钟。

2.2 共培养步骤:

用 1mL 微吸移管将农杆菌属细菌悬浮液从感染步骤中取出。使用无菌铲将胚芽从管中刮下来, 并且转移到在 100 × 15mm 培养皿中的 PHI-B 培养基的平板中。确定胚芽的

朝向，使得胚轴向下朝着培养基的表面上。将具有胚芽的平板在 20℃ 于黑暗中培养 3 天。L-半胱氨酸可用于共培养阶段。采用标准二分载体，补充 100-400 mg/L L-半胱氨酸的共培养培养基对于回收稳定转基因事件是至关重要的。

3. 选择推定的转基因事件：

向在 100 × 15mm 培养皿中的 PHI-D 培养基的平板中转移 10 个胚芽，保持朝向，并且用 parafilm 将培养皿密封。将平板在黑暗中于 28℃ 培养。预计在 6-8 周将看见作为黄色胚芽组织的主动生长推定事件。不产生事件的胚芽可以是棕色和坏死的，并且易碎组织生长是明显的。以 2-3 周的间隔将推定的转基因胚芽组织转移到新鲜的 PHI-D 平板上进行传代培养，时间间隔取决于生长速度。记录事件。

4. T0 植株的再生：

将在 PHI-D 培养基上繁殖的胚芽组织转移到在 100 × 25mm 培养皿中的 PHI-E 培养基(体细胞胚芽成熟培养基)中进行传代培养，在 28℃ 于黑暗中培养直至体细胞胚芽成熟，培养大约 10-18 天。将具有良好限定的盾片和胚芽鞘的个体成熟体细胞胚芽转移到 PHI-F 胚芽发芽培养基中，并且在 28℃ 于光中(约 80μE，来自冷光灯或同等荧光灯)培养。在 7-10 天，将约 10cm 高的再生的植株置于盆中的园艺混合物中，并且使用标准园艺方法进行耐寒锻炼(hardened-off)。

用于植物转化的培养基：

1. PHI-A: 4g/L CHU 基础盐，1.0mL/L 1000 × Eriksson's 维生素混合物，0.5mg/L 盐酸硫胺，1.5mg/L 2,4-D，0.69g/L L-脯氨酸，68.5g/L 蔗糖，36g/L 葡萄糖，pH 5.2。加入 100μM 乙酰丁香酮(过滤灭菌的)。
2. PHI-B: PHI-A，不含有葡萄糖，将 2,4-D 增加至 2mg/L，将蔗糖降低至 30g/L，并且补充 0.85mg/L 硝酸银(过滤灭菌的)，3.0g/L Gelrite[®]，100μM 乙酰丁香酮(过滤灭菌的)，pH 5.8。

3. PHI-C: PHI-B, 不含 Gelrite[®] 和乙酰丁香酮, 将 2,4-D 降低至 1.5 mg/L, 并且补充 8.0g/L 琼脂, 0.5g/L 2-[N-吗啉代]乙烷-磺酸(MES)缓冲液, 100mg/L 羧苄西林(过滤灭菌的)。
4. PHI-D: PHI-C 补充 3 mg/L bialaphos (过滤灭菌的)。
5. PHI-E: 4.3g/L Murashige and Skoog (MS) 盐(Gibco, BRL 11117-074), 0.5mg/L 烟酸, 0.1mg/L 盐酸硫胺, 0.5mg/L 盐酸吡哆醇, 2.0mg/L 甘氨酸, 0.1g/L 肌醇, 0.5mg/L 玉米素(Sigma, Cat. No. Z-0164), 1mg/L 吲哚乙酸(IAA), 26.4μg/L 脱落酸(ABA), 60g/L 蔗糖, 3mg/L bialaphos (过滤灭菌的), 100mg/L 羧苄西林(过滤灭菌的), 8g/L 琼脂, pH 5.6。
6. PHI-F: PHI-E, 不含有玉米素、IAA、ABA; 将蔗糖降低至 40g/L; 用 1.5g/L Gelrite[®] 替代琼脂; pH 5.6。

可以通过以下方法由愈伤组织再生出植物: 首先将组织簇转移到 N6 培养基中, 所述培养基补充了 0.2mg 2,4-D/升。两周后, 可将组织转移到再生培养基中(Fromm 等人, Bio/Technology 8: 833-839 (1990))。

转基因 T0 植株可以再生, 并且可以确定其表型。可以收集 T1 种子。

此外, 可通过直接转化或者从单独转化的品系基因渗入而将含有证实的拟南芥基因的重组 DNA 构建体导入玉米近交系内。

近交或杂交的转基因植物可通过更有力的大田试验来研究表达效果。

实施例 2

分离小 RNA 以用于 MPSS 或 454 测序

使用 Trizol 试剂(Invitrogen)从以下组织中提取 RNA 样本: 干旱胁迫的玉米叶子, 良好浇灌的玉米叶子, 未受精的玉米原卵, 早期阶段的玉米种仁(授粉后 2 天), 混合后期阶段的玉米种仁(授粉后 7、14 和 21 天)。将总 RNA 在 15% 聚丙烯酰胺 TBE/尿素凝胶上分离, 并且在单独的泳道中还包括 21-nt RNA 标记。在电泳之后, 将凝胶用溴化乙锭染色, 并且将与 20-22 个核苷酸相对应的凝胶区域切下来。将小的 RNA 片段洗脱过夜, 乙醇沉淀, 然后使用 T4 RNA 连接酶依次与 5'和 3' RNA 衍接子连接 (5' RNA 衍接子

GGUCUUAGUCGCAUCCUGUAGAUGGAUC 和 3' RNA 衔接子 pAUGCACACUGAUGCUGACACCUGCidT, 其中 p = 磷酸; idT = 反向脱氧胸苷; 分别是 SEQ ID NOs: 8428 和 8429)。将每一连接的产物在 10% 变性聚丙烯酰胺凝胶上进行凝胶纯化以除去未连接的衔接子。然后对最终的连接产物进行 RT-PCR, 使用与 5'和 3'衔接子序列互补的引物。使用以下三种方法当中的一种测定与小 RNA 相对应的扩增的 cDNA 的序列: 串联, 然后进行标准二脱氧测序(Elbashir 等人, 2001 Genes & Dev. 15: 188-200), Massively Parallel Signature Sequencing (MPSS™)标记测序(Solexa)或 454™ 测序(454 Life Sciences), 或通过 SBS (Sequencing-By-Synthesis; Illumina™ Inc.San Diego, CA)的小 RNA 测序。

实施例 3

鉴定玉米 miRNA 序列

通过叫做“mirna”的程序分析实验数据, 所述程序称为“该程序”。该程序是用 Perl 程序语言写成的。其主要从 CGI 网界面上运行, 但是也可以从命令行来运行。

对于玉米(Zea mays)基因组数据与实验结果的任何特定组合, 来自 MPSS 或 454 测序的实验序列数据位于玉米基因组数据库中; 表明实验序列与基因组序列之间的配对的那些序列称为“标记”。将专有和公共基因组数据装配成一组 contigs。单独检索任何非邻接基因组序列或未装配的基因组序列。此外, 来自 B73 和 MO17 品种的数据保持独立。所有随后的分析都是在标记上进行, 所述标记是通过将来自小 RNA MPSS 或 454 实验的序列与位于基因组数据库中的序列进行配对而鉴定的。

给这些标记施加可选择的一系列筛选以计算确定哪些标记可能代表 miRNA。每一筛选可以由运行该分析的用户选择或跳过。筛选包括: 将存在于重复区域中的标记除去; 将配对已知 RNA 的标记除去; 将在基因组数据中具有显著重叠的标记合并; 将配对公共已知 miRNA 的标记除去; 将在簇中显示彼此接近的标记除去; 将在实验数据中具有低 PPM (百万份数, 频率的度量单位)值的标记除去; 将在基因组中

附近不具有容易鉴定的碱基配对序列的标记除去；和测试围绕标记的基因组区域的良好 RNA 折叠特征。

如下所述进行 RNA 折叠测试：首先对含有标记的 500bp 的基因组区域采用公众可获得算法(Vienna RNA Package)。将根据缺省参数不折叠的区域(包括 $-30\Delta G$ 的 Gibbs 自由能；Hofacker 等人, 1994 *Monatsh.Chem.* 125: 167-188)除去不进一步考虑。用可调节的参数进行可选择的一系列测试以鉴定其折叠似乎是 miRNA 前体的标记。这些参数包括对于以下方面的几个限定：在推定折叠中错配碱基对的大小和数目；折叠的发夹区域的大小和构象；通常在 miRNA 中存在的一些碱基的配对特征；以及与 miRNA 前体相邻的区域的折叠特征。

然后将通过这些计算筛选的标记组在几个图形显示中检验，所述图形显示表现折叠、推定 miRNA 以及折叠测定和其他实验结果的详细结果。该实验可带来被已经通过计算筛选的专家(expert)拒绝的标记。然后可将由这些专家检查提出的另外的计算筛选并入到该程序内。

通过使用 WU-BLAST (Gish, Washington University, St. Louis; Altschul 等人, 1990 *J Mol Biol* 215: 403-410)分析来进一步测定 miRNA 序列的生物有效性，以鉴定推定基因组靶。基于该靶测定，玉米基因组标记的给定不完全性质未被拒绝。然而，靶的计算鉴定增加了 miRNA 序列有生物活性的可能性。水稻与玉米基因组数据的组合用于鉴定推定靶。

实施例 4

玉米 miRNA 序列

在实施例 3 中描绘的方案用于从以下玉米组织中分离的玉米 RNA：干旱胁迫的玉米叶子，良好浇灌的玉米叶子，未受精的玉米原卵，早期阶段的玉米种仁(授粉后 2 天)，混合后期阶段的玉米种仁(授粉后 7、14 和 21 天)。

分析产生了 2808 个玉米微 RNA 序列(参见表 1，其显示了记载于在 2006 年 10 月 5 日提交的第 60/849,672 号美国临时申请中的 SEQ ID NOs: 1-2652, 以及加到本文中 156 个另外的微 RNA 序列 SEQ ID NOs: 7959-8114)。

本申请在光盘上提供序列列表和表格。根据 37 CFR 1.52(e), PCT Rule 5.2, 和 PCT Administrative Instruction 802(a)的规定, 将含有序列表和表格的光盘的内容以引用方式并入本文。光盘以一式三份提交, 并且彼此相同。光盘标记为“副本 1 - 序列列表和表格”, “副本 2 - 序列列表和表格”, 和 CRF。光盘含有以下文件: 具有以下大小的 BB1593 US PRV 序列列表: 1,471,000 个字节, 和具有以下大小的 BB1593 US PRV 表格: 4,430,000 个字节, 是在 2006 年 10 月 5 日产生的。

已知的 miRNA 序列从该列中排除。发现很多已知的 miRNA 序列在序列文库中大量呈现, 并且通过在本申请中描绘的方案鉴定。已知的 miRNA 序列不限于玉米, 尽管起始 RNA 都是来自玉米。因此, 预计在本申请中公开的 2808 个 miRNA 序列是富集玉米特异性和较低丰度 miRNA 物种的群体。应当注意, “较低丰度”是相对术语, 并且在任何给定组织或发育状态中的任何给定微 RNA 序列的绝对量可能有宽的变化。

实施例 5

玉米 miRNA 的表达分析

表达分析可用于提供进一步的证据以表明给定的小 RNA 序列相当于 miRNA。通过 Northern 印迹和杂交检测 miRNA 的能力是研究人员用于验证候选 miRNA 的标准准则(Ambros 等人, 2003)。或者, 已经开发了 RT-PCR 方案, 其容许检测具有较低丰度的小 RNA (Shi 和 Chiang, 2005; Chen 等人, 2005)。我们将使用其中一种方法或这两种方法来验证从我们的列中选择的候选 miRNA。此外, 大部分(如果不是所有的话)我们公开的候选 miRNA 将作为有关微阵列芯片的特征包括在内, 其将容许与各种不同组织和处理的小 RNA 片段杂交。通过微阵列分析的 miRNA 的表达谱是快速发展并且有力的技术, 其应当容许进一步验证候选 miRNA, 以及提供有关特定 miRNA 的组织特异性和/或条件表达的信息。

特定候选 miRNA 可具有基于生物信息分析可清楚地识别的靶转录物, 这提供了进一步验证 miRNA 的机会。在所选的候选 miRNA 的推定靶的修饰 RACE-PCR 将用于鉴定 RISC 裂解产物, 所述裂解产物具有位于 miRNA 的推定靶位点的中央的 5'末端(Llave 等人, 2002)。该结

果将强烈地表明给定的候选 miRNA 进入了 RISC 复合物, 并且其功能将象所预计的真实 miRNA 的功能那样。

对于具有农学所关注的表达模式和/或靶的选择的 miRNA, 例如干旱诱导的或种仁特异性 miRNA, 开始进行功能研究。为了确定所选 miRNA 的种植功能, 可以通过以下方法将 miRNA 过表达: 将包含 miRNA 的发夹前体加最小量侧翼序列的序列(cDNA 或基因组序列)与组成型启动子(例如玉米遍在蛋白启动子)融合, 并且产生具有该构建体的转基因玉米系。根据所研究的农学特征, 可以在过表达 miRNA 的独立 T0 个体上进行表型研究。此外, miRNA 过表达对于相关靶基因的影响可以通过 Northern 印迹和/或 RT-PCR 来评估, 以证实预测的靶事实上是真实的靶。给定 miRNA 控制给定靶基因的表达的独立证明, 将使靶基因内存在的靶位点突变(不改变氨基酸序列), 以及再转化靶基因的“miRNA 抗性”变型。该方法使得能够评估由于 miRNA 调节的失去而带来的表型, 在大部分情况下, 这样的表型不能通过直接剔除 miRNA 来评估(由于基因过多和/或缺乏可利用的剔除品系)。

实施例 6

通过表达谱验证小 RNA

由于所公开的候选 miRNA 的数目达到数千, 所以必须使用高通量方法来验证。微阵列技术可用于该目的。微阵列的适应性和可靠性已经得到良好阐述(Davison 等人, (2006), *Methods in Enzymology* 411: 14-34; Axtell 和 Bartel (2005), *Plant Cell* 17: 1658-1673), 它们适于分析很多不同组织和条件的表达。在过去几年, 对于微阵列, 已经能够获得小 RNA 杂交的专门方案。构建微阵列来表现候选 miRNA, 以及已经在文库中的已知 miRNA。用从多种不同玉米组织纯化的标记的小 RNA 片段来进行杂交, 所述玉米组织包括不同器官例如根、叶和花, 并且包括环境变量例如干旱/病原体胁迫和低养分。杂交数据的统计学分析使得能够鉴定表现出显著表达的候选 miRNA, 以及以该方式验证内源性表达。该谱方法还提供了关于特定 miRNA 的体内功能的初始线索, 尤其是如果它们的表达是组织特异性或受环境调节时。功能验证将包括在有希望候选者上进行的 Northern 印迹分析和/或 qRT-PCR (参见实施例 8)。

通过与常规设计的植物 miRNA 表达阵列杂交来分析小 RNA 表达。阵列是通过 Agilent 技术以 2-包装 11,000 特征格式合成的。来自多个物种的已知 miRNA 和候选 miRNA 序列通过一系列专门设计的报道基因 oligos 来代表。设计有义和反义方向的包含 0、1 或 2 个与靶序列的错配的报道基因。此外，加入含有反义靶序列的两个串联拷贝的双重 (duplexed) 报道基因。使用反义双重和 0-错配报道基因来检测 miRNA 表达。反义 1-和 2-错配报道基因提供了有关报道基因特异性的信息。有义报道基因数据有助于将 miRNA 与小的干扰性 RNA 区分开。

使用 Trizol 试剂 (Invitrogen) 制备总 RNA，并且使用玻璃纤维滤器 (miRvana 小 RNA 分离试剂盒，Ambion) 富集小 RNA (<200 nt)。将小 RNA 样本用 Cy-5 (Label IT miRNA 标记试剂盒，Mirus Bio) 直接标记。将 4 微克标记的 RNA 与常规阵列在 50℃ 杂交。按照 Agilent One-Color Microarray-Based Gene Expression Analysis 方案进行杂交、洗涤、扫描和特征提取方案。为了保证再现性和可靠性，将样本以一式三份进行分析。使用 Rosetta Resolver, Spotfire DecisionSite 和定制的点呼叫 (spot-calling) 和误差计算程序分析数据。

如果对于所有三个样本，反义双重和 0-错配报道基因记发器 (register) 强度超出背景 2 个标准偏差以上，但是其他报道基因记发器信号为背景强度或者在背景强度以下，则测定欲表达的小 RNA。通过其他技术例如小 RNA Northern 印迹 (Plant Cell. 2003; 15(11): 2730–2741)，凝胶内 miRNA 检测 (miR-tect-IT miRNA 标记和检测试剂盒，USB Corporation)，常规 Taqman miRNA 分析 (Applied Biosystems)，和聚腺苷酸化辅助的 RT-PCR 对这些序列进行另外的验证。

实施例 7

通过表达谱验证小 RNA

组织源和 RNA 提取

表 2 列出了用于表达谱而收获的组织基因型、生长条件和发育阶段。

表 2: 测定的样本的生长和收获条件

样本名称	组织	基因型	阶段	生长条件	实验条件
B73 叶	叶	B73	VT	盆种植; 屏风帐: Newark, DE	组织测量
穗	未成熟穗	B73	VT	盆种植; 屏风帐: Newark, DE	组织测量
B73 幼苗	出土幼苗	B73	V4	在水中的发芽纸; 生长室(28℃, 14 小时光照, 10 小时黑暗)	组织测量
3245 标准氮	叶	3245	V18	田地种植: Johnston, IA	标准氮
3245 低氮	叶	3245	V18	田地种植: Johnston, IA	低氮
33B50 标准氮	叶	33B50	V18	田地种植: Johnston, IA	标准氮
B73 低氮	叶	B73	V18	盆种植, 温室	低氮

将收获的组织立即在液氮中冷冻, 并且在 -80℃ 贮存。将组织匀化成粉末形式, 同时保持冷冻。使用生产商有关 Trizol™ Reagent (Invitrogen; Carlsbad, CA) 或 E.Z.N.A. DNA/RNA Isolation System™ (Omega Bio-Tek; Doraville, GA) 的说明书, 通过有机提取从 2-5 克匀化组织中分离总 RNA。使用 mirVana™ miRNA Isolation Kit (Ambion; Austin TX), 方案 IV.A: 从总 RNA 样本分离小 RNA, 从 100 微克总 RNA 分离出富集长度低于 200 个核苷酸的 RNA 分子的级份。将总 RNA 和小 RNA 样本分别在 RNA 6000 Nano Chip 和 Small RNA Chip (2100 Bioanalyzer, Agilent Technologies) 上分析, 以确定样本的质量和浓度。

阵列设计

常规基因表达微阵列是由 Agilent Technologies 以 4 × 44k 格式生产的。每个六十核苷酸探针编码两个串联饵(bait)序列, 这两个序列通过三核苷酸间隔区隔开, 并且后面跟着具有 15 个核苷酸的 Agilent's 一般间隔序列。每个饵序列的前面是鸟嘌呤残基。整个探针设计是按照以下格式: 5'-G[BAIT]_{21nt}ATAG[BAIT]_{21nt}GCGTTCCGTATGTGG-3'。

对于每个测定的小 RNA, 设计三个探针: 完全序列配对探针(0MM)使用与查询序列反义的饵序列; 单错配探针(1MM)使用反义查询序列, 该反义查询序列在从作为饵的有义查询序列的 5' 末端开始计第 10 个位置具有取代; 双错配探针(2MM)使用反义查询序列, 该反义查询序列在从作为饵的有义查询序列的 5' 末端开始计第 7 和 15 个位置具有取

代。取代是通过用 T 替换 A，用 A 替换 T，用 C 替换 G，以及用 G 替换 C。

小 RNA 标记

在标记之前，将等份试样的掺杂(spike-in)对照混合物加到 5 微克每一小 RNA 样本中。如下所述进行 RNA 的直接标记：将 4 微克每一样本与 4 微升 LabelIT Cy3 Reagent 根据 LabelIT miRNA Labeling Kit 方案 (Mirus Bio Corporation; Madison WI) 进行培养。然后使用分光光度计定量测定样本，以测定浓度以及 Cy3 掺入速度。

阵列杂交

在旋转杂交烘箱中，将每一样本与常规 Agilent 寡核苷酸阵列于 50 °C 杂交过夜。根据 Agilent Technologies One-Color Microarray-Based Gene Expression Analysis Protocol 将微阵列载片杂交并且洗涤，并且立即用 Agilent's DNA Microarray Scanner 在两个激光设置(100% 和 10% PMT)进行扫描。

数据分析

目测检查图像的图像伪影(artifacts)，并且提取特征强度，过滤，并且用 Agilent's Feature Extraction Software (9.5.3.1 版本)进行归一化。使用 Rosetta's Resolver Database (NBIC Rosetta Resolver System) 中的数据分析工具进行进一步的质量控制和下游分析。

样本作为生物一式两份或一式三份进行分析。将特征强度数据合并，并且使用最小二乘方方法，采用误差的加权回归，来在评选测定间以及实验内进行归一化。现有技术中报道了 miRNA 微阵列数据分析的统计学方法的简短描述(Luck, (2001) Proceedings of SPIE 4266: 153-157)。该分析是基于微阵列和 miRNA 的荧光强度的向量空间表示。分析的第一个阶段包括将微阵列强度归一化，以调节平行测定之间装置反应的差异。通过应用加权回归来表征平行测定之间的协方差。将平行测定的归一化强度平均以获得给定组织的每一 miRNA 探针的强度。使平行测定的强度相关联，然后将 miRNA 向量在平均线周围分布，所述平均线具有等于平行测定数目的维数。通过分析平行测定向量偏离平均线的平方偏差来估计强度测定中的噪音。通过平方偏差对平均强度的加权二次回归获得方差函数。然后将经验置信区间与平均强度值关联起来。置信区间表征整个噪音，包括生物和装置带来的噪

音。使用置信区间来评估所观察到的偏离平均值的偏差的显著水平，以及确定平行测定间的异常值向量。类似地，使用平均值之间差异的显著水平来确定由于错配探针而表现出显著杂交变化的 miRNA。使用 t 分布来评估差异统计学的 p-值，其中自由度等于平行测定的数目。

在每一样本的生物平行测定之间，通过 ANOVA 或 t-检验计算来确定每一归一化信号的再现性。平行测定 P-值大于 0.05 的序列被认为具有足够的再现性。

进行 t-检验来比较每一探针的归一化信号与该样本中的背景信号。0MM-背景 P-值小于 0.05 的查询序列被认为表达显著在背景以上。

对于每一查询序列，使用两个计算来确保 0MM 信号的特异性。再次使用 t-检验，将 0MM 归一化信号与每一查询序列的错配探针的归一化信号进行两两比较。因为 P-值仅表明信号之间差异的显著性，但是不表明哪个信号更大，所以每一 0MM 信号还需要比其相应的 1MM 和 2MM 信号都大。其 0MM 信号大于其两个相应的错配信号，并且其 0MM-1MM 和 0MM-2MM P-值都小于 0.05 的任何查询序列被认为表现出合适的信号特异性。标准总结在表 3 中。

表 3：用于确定查询序列的可靠检测的标准。

值	标准
0MM 平行测定 P-值	>0.05
0MM-背景 P-值	<0.05
0MM-1MM P-值	<0.05
0MM-2MM P-值	<0.05
0MM – 1MM	>0
0MM – 2MM	>0

给每一样本独立地施用检测标准。满足所有上述标准的查询序列被认为已经通过表达阵列可靠地检测，并且在表 4 中标为“通过的”、“富集的”或“高富集的”。在给定样本中，如果查询序列满足所有可靠检测标准，并且在该样本中的其 0MM 信号强度是相同实验中任何其他样本的 0MM 强度的 5 倍至 10 倍，则该查询序列称为“富集的”。“高富集的”序列满足所有可靠检测标准，并且其表现出的 0MM 信号强度

是相同实验中任何其他样本的 0MM 强度的至少 10 倍。在仅含有单一样本的两个实验(“33B50 标准氮”和“B73 低氮”)中,只分析查询序列是否满足可靠检测标准 - 不进行有关富集的询问(call)。在任意和可靠检测评估中的 0MM 强度评分在表 4 中通过每一查询序列的样本列出。

有关表 1 中前 2652 个玉米微 RNA 的表达谱结果总结在表 4 中。

本申请在光盘上提供序列表和表格。根据 37 CFR 1.52(e), PCT Rule 5.2, 和 PCT Administrative Instruction 802(a)的规定,将含有序列表和表格的光盘的内容以引用方式并入本文。光盘以一式三份提交,并且彼此相同。光盘标记为“副本 1 - 序列表和表格”,“副本 2 - 序列表和表格”,和 CRF。光盘含有以下文件:具有以下大小的 BB1593 US PRV 序列表: 1,471,000 个字节,和具有以下大小的 BB1593 US PRV 表格: 4,430,000 个字节,是在 2006 年 10 月 5 日产生的。

检测公共领域玉米(Zea mays) miRNA

公共领域玉米(Zea mays) miRNA, Public domain Zea mays miRNA 包括在查询序列中间以用作检测标准。由于 miRNA 家族成员当中的序列冗余性,在阵列设计时公布的 116 个玉米用 50 个独特的 miRNA 序列完全表示(Griffiths-Jones 等人, Nuc Acids Res (2006) 34: D140-D144. Release 9.0; Zhang, 等人, FEBS Lett. (2006) 580: 3753-62; Sunkar, 等人, The Plant Cell (2005) 17: 1397-1411; Johnson, 等人, Nuc Acids Res (2007) 35: D829-D833)。使用表 3 中给出的可靠检测标准,检测至少一个样本中的 88% 公共领域 miRNA 序列。

在任何样本中的不能满足可靠检测标准的 6 个公共领域 miRNA 序列当中,只有 miR437 和 miR444 不能超过背景噪音。因为没有在玉米中测定 miR437 表达,并且已知 miR444 在玉米中的表达比其在水稻中的表达弱,所以能够预计这些序列的低信号强度(Sunkar 等人, The Plant Cell (2005) 17: 1397-1411)。只有 4% 的公共领域 miRNA 序列不能超过背景噪音,这意味着在这些数据库中具有优良的信噪比。MiR168a/b、miR390 和 miR395a/b/c 超过背景噪音,但是不能表现出在 0MM 与 1MM 信号之间的显著差异。虽然 0MM 信号超过了背景,并且与其相应的错配不同,但是 miR171g 的 2MM 信号始终高于其 0MM signal 信号。对于缺乏信号分辨率的 4 个 miRNA 序列,不能推断查询序列是否不存在,

以及 OMM 信号是否表明交叉杂交, 或者除了与错配探针特异性地杂交的第二个序列之外, 是否存在查询序列。

很多公共领域 miRNA 在几乎所有组织中都具有可检测的基础表达 (Johnson 等人, Nuc Acids Res (2007) 35: D829-D833)。与此一致的是, 在所有检测样本中可靠地检测到了 46% 的公共领域 miRNA。在组织比较实验内, 任何样本中存在 78% 或更多的公共领域 miRNA。在单一组织或实验条件下的近遍在基础表达也是这么少公共领域 miRNA 被评为“富集的”或“高富集的”的原因。表 5 中总结了公共领域 miRNA 的可靠检测。

表 5: 公共领域 miRNA 的可靠检测的总结。

样本	通过所有标准	单一通过	富集的	高富集的
最大可能	100% (50)			
B73 叶	84% (42)	84% (42)	0% (0)	0% (0)
B73 穗	78% (39)	76% (38)	2% (1)	0% (0)
B73 幼苗	82% (41)	82% (41)	0% (0)	0% (0)
测定的所有组织	74% (37)			
一种或多种组织	88% (44)			
3245 标准氮	70% (35)	70% (35)	0% (0)	0% (0)
3245 低氮	62% (31)	62% (31)	0% (0)	0% (0)
两个样本	62% (31)			
33B50 标准氮	56% (28)			
B73 低氮	78% (39)			
所有样本	46% (23)			
一个或多个样本	88% (44)			

检测候选 miRNA 序列

还将表 3 中描述的可靠检测标准应用于得自代表本申请中公开的 2652 个候选 miRNA 的探针的数据(见表 6)。在所有 7 个样本中仅可靠地检测了 136 个或 5.1% 的这些候选者。因为它们是从小 RNA 文库中

克隆和测序的，所以最丰富以及广泛表达的 miRNA 将已经在最早被检测和公开的那些当中。因此，遍在地检测的候选 miRNA 的比例远小于公共领域 miRNA 的比例是合理的。在严格术语下，136 个 miRNA 的完全验证将大约是玉米 miRNA 的已知核心(cadre)的两倍。

因为预计最丰富的 miRNA 已经在早期 RNA 文库检索中发现和出版，所以在所有或大部分样本当中表现出极强 OMM 强度分数的候选者应当谨慎对待。因为直接标记技术将荧光分子与每一个鸟嘌呤残基缀合，所以更长序列一般比短序列亮。大小分级试剂盒用于富集 0-200 个核苷酸的小 RNA 的样本制备。因此，在大部分样本中，OMM 强度分数大于 20,000 的候选者可以是由其长度远大于 21 个核苷酸的 RNA 物种引起的人工产物。

41 个候选序列或 1.5% 的候选序列被可靠地检测为在特定样本中富集或高富集的。强的组织或条件特异性检测模式表明这些候选者可尤其用于定向基因沉默。

B73 是用于 MPSS 深度测序实验的一种玉米株，所述实验产生 2652 个候选 miRNA。用于该阵列分析的组织测定样本还得自 B73 植株，同样也是未配对的低氮样本。在 B73 样本中，公共领域和候选 miRNA 满足可靠检测标准的比例高于在其他基因型中的比例。因为公共领域 miRNA 在 B73 与其他基因型之间表现出不同的通过比例，所以某些差异可归因于生长条件或样本处理中的差异。虽然如此，候选 miRNA 在 B73 与其他基因型之间的可靠检测中表现出大得多的差异，从而提高了 miRNA 可在不同遗传背景中差异表达的可能性。

表 6: 候选 miRNA 的可靠检测的总结。

样本	通过所有标准	单一通过	富集的	高富集的
最大可能	100.0% (2652)			
B73 叶	31.7% (841)	31.1% (824)	0.1% (3)	0.5% (14)
B73 穗	62.0% (1645)	61.3% (1626)	0.8% (20)	0.2% (6)
B73 幼苗	42.2% (1120)	42.2% (1118)	0.1% (2)	0.0% (1)
测定的所有组织	27.6% (731)			
一种或多种组织	64.9% (1722)			
3245 标准氮	12.9% (343)	12.9% (342)	0.0% (0)	0.0% (1)
3245 低氮	9.4% (250)	9.4% (250)	0.0% (0)	0.0% (0)
两个样本	9.2% (243)			
33B50 标准氮	8.1% (216)			
B73 低氮	20.4% (540)			
所有上述样本	5.1% (136)			
一个或多个样本	65.6% (1739)			

实施例 8

通过聚腺苷酸化辅助的 RT-PCR 验证小 RNA

在实施例 7 中描述的 miRNA 表达阵列和分析方法可用于评估候选 miRNA 在另外样本中的可检测性以及表达水平, 所述另外样本包括但不限于: 另外发育阶段的器官或组织; 得自在 miRNA 和小 RNA 生源论方面有缺陷的突变株的组织; 得自暴露于生物胁迫例如病原体、害虫或杂草压力的植株的组织; 得自暴露于非生物胁迫例如干旱、热或冷的植株的组织; 以及得自施加了营养物胁迫例如低氮的植株的组织。

可以基于在其中候选者被可靠检测的样本的数目, 推定靶的潜在用途, 以及时间、空间和条件表达模式, 将候选者以优先顺序排列以进行个体表征。众所周知的表征方法包括但不限于定量 RT-PCR、Northern 印迹、体内杂交、报道基因沉默阵列以及在推定靶上的修饰 5' RACE。

作为实例, 使用 Shi & Chiang 方案的修饰版本来分析小 RNA, 该方案容许扩增不具有天然聚(A)尾部的 RNA 片段(Biotechniques. 2005; 39(4): 519-25)。通过用重组 DNaseI (Ambion)消化将 DNA 从用 Trizol 试剂(Invitrogen)制备的总 RNA 样本中除去。用乙醇将 RNA 沉淀之后, 使用大肠杆菌(E. coli)聚(A)聚合酶(Poly(A) Tailing Kit, Ambion)将材料聚腺苷酸化。

使用 SuperScriptIII 逆转录酶和提供的聚(T)引物(GeneRacer Kit, Invitrogen)完成第一链 DNA 合成。随后通过终点 PCR 的扩增使用低的退火温度(56 °C)和短的延伸时间(30 秒)。使用 GeneRacer 3' Primer (GeneRacer Kit, Invitrogen)作为方向引物, 并且使用所关注的小 RNA 的有义 DNA 变型用作正向引物。通过在 10% 丙烯酰胺 TBE 凝胶 (Criterion gels, BioRad)上用溴化乙锭将 PCR 产物显色来评估诊断性 ~ 81mer PCR 产物的存在。

实施例 9

用微 RNA 选择性调节转基因

因为每一候选 miRNA 都有可能识别和负调节互补序列(即“靶位点”), 所以据信这些候选 miRNA 的靶位点将用作转基因构建体中的有用顺式作用调节序列(Parizotto 等人, (2004) Genes Dev 18: 2237-2242)。当 miRNA 表现出精确的组织特异性或条件表达模式时这是尤其正确的。例如, 为了表达在芽中而不是在根中所关注的特定基因, 将根特异性 miRNA 的靶位点与所述所关注的基因的编码序列融合。以该方式, 即使所关注的基因不被组成型启动子驱动, 转录物也应当仅在芽中而不是在根中聚集, 因为内源性根特异性 miRNA 将识别附着的靶位点, 并且沉默基因在根中表达。

为了测试该观点, 鉴定表现出组织特异性表达的 miRNA (参见实施例 7)。理想地, 生物相关靶可通过生物信息手段发现, 并且通过 RACE 裂解分析验证; 然而, 与 miRNA 完全互补的序列应当也满足。将指定为靶位点的序列与由组成型启动子取代的适当报道基因的 5'或 3'末端融合, 并且生成转基因植物。把与组织特异性 miRNA 模式相反的表达模式作为表明靶位点是作为顺式调节元件工作的证据。这种顺式调节容许进一步精化转基因表达模式, 这样“漏的”启动子(例如在不需要

的组织或条件中具有少量基因表达)被良好调节, 并且实现真正的组织特异性。

具有阶段特异性表达模式的 miRNA 的一个实例是 miR172。在幼小的拟南芥幼苗中, miR172 是不能检测到的, 而在更成熟的幼苗中, 其具有较高水平, 并且该较高表达持续到开花后(Aukerman 和 Sakai (2003) Plant Cell 15: 2730-41)。因此, 给转基因构建体中的异源基因加上 miR172 靶位点应当容许转基因在早期幼苗发育阶段转变, 稍后下调。为了测试该观念, 将 miR172 的靶序列与珊瑚黄(coral reef yellow)荧光蛋白(ZS-yellow, Clontech)的编码序列的 3'末端融合。作为阴性对照, 将 miR172 靶位点的突变变型单独融合到 ZS-yellow 上, 在标准二元载体中将两个基因盒单独置于组成型 35S 启动子的控制下。将这些构建体转化到拟南芥(生态型 Col-0)内, 并且如果转基因沉默水平(独立于 miRNA 调节)太高的话, 转转化成 rdr6 突变体。预计的结果是, ZS-yellow 的水平在含有 ZS-yellow/野生型靶位点的幼小转基因幼苗中将很高, 而在植物成熟时将低得多。另一方面, 含有 ZS-yellow/突变的靶位点的转基因植株应当不表现出这种阶段特异性下调。

为了将该分析延伸到玉米上, 本申请序列表中一个玉米 miRNA 的靶位点(SEQ ID NO: 116; 5'-TTAGATGACCATCAGCAAAC-3')是有用的。该特定 miRNA 在玉米胚乳中以高水平表达, 并且在其他组织中以低得多的水平表达。靶基因是 EB158863, 并且靶位点与 miRNA 在 miRNA 的 3'末端具有两个错配。将该靶位点与 ZS-yellow 的 3'末端融合, 并且将该构建体置于玉米遍在蛋白启动子的控制下。将表达构建体插入到用于玉米转化的合适的载体内。还制备含有该靶位点的突变变型的阴性对照构建体。将这两种构建体转化到玉米胚芽发生愈伤组织内, 并且选择转基因植物。预计的结果是, ZS-yellow 的水平在含有 ZS-yellow/野生型靶位点的植物的大多数部分中是高的, 但是在胚乳中低得多。另一方面, 含有 ZS-yellow/突变的靶位点的转基因植物应当不表现出这样的胚乳优先的下调。

在这些类型的实验中, 相反结果也是可能的。实施例 7 中显示的富集/高富集询问是基于条件特异性或提高。因此, 将总结结果在检测对未检测方面进行加权。刚好可能是特定微 RNA 在具体组织中耗尽而

不是在具体组织中富集。这将使得靶基因通常被抑制，然后在其中特定微 RNA 是不足表达的组织中以受限方式表达。

在上述两个实验中，把与组织特异性 miRNA 模式相反的表达模式当作表明所研究的靶位点是作为顺式调节元件工作的证据。联合使用这些新的顺式元件与多种不同启动子元件将容许在植物生长和发育期间转基因以非常特定的模式或阶段表达。例如，通过 miRNA 靶的这种顺式调节将容许进一步精化转基因表达模式，这样“漏的”启动子(例如在不需要的组织或条件中具有少量基因表达)被良好调节，并且实现真正的组织特异性。还预想，可将多个 miRNA 靶掺入，以与单个表达的转基因一致。具有不同调节和控制模式的多个 miRNA 可以给转基因表达带来新的调节模式。

实施例 10

转基因丰度可影响通过微 RNA 的调节

为了测试通过多个 miRNA 来联合控制转基因的可能性，进行以下实验。将 ZS-yellow 编码区域单独与 AtmiR827 的靶序列在 5'末端或者与 AtmiR397 的靶序列在 3'末端融合(“单位点构建体”)。在第 3 种构建体中，在与单位点构建体相同的位置上将两个靶序列与 ZS-yellow 编码序列融合，由此产生“双位点构建体”。基于 AtmiR827 启动子-GUS 组织化学测定的分析，以前的数据表明 AtmiR827 在排水器中特异性地表达。因此，含有 AtmiR827 的靶位点的单位点转基因构建体将处于内源性 miRNA 的控制下，并且可预计，当与“无位点”对照构建体相比时，会引起转基因在排水器中的表达降低。同样，微阵列分析已经表明，AtmiR397 优先在根中表达，因此，可预计，当与“无位点”对照相比时，含有 AtmiR397 的靶位点的构建体会引起转基因在根中的表达降低。预计双位点构建体表现为两种单位点构建体的加合，从而引起在排水器和根中的表达降低。

在二元载体 pBC (Promega; Madison, WI)中，将所有三种修饰的 ZS-yellow 序列单独置于组成型 35S 启动子的控制之下，并且还将未修饰的 ZS-yellow 序列单独与 35S 融合，作为“无位点”对照。使用“花浸”转化(Clough 和 Bent (1998) Plant J 16: 735-43)，将两种单位点构建体、双位点构建体和“无位点”对照构建体转化到拟南芥内，并且

鉴定从 ZS-yellow 转基因表达 YFP 的 T1 种子。让 T1 种子在琼脂平板上发芽，并且在发芽后 5 天开始监测 YFP 表达。在这四种构建体之间没有观察到表达模式的差异。更具体来说，当与单独的 ZS-yellow 比较时，在含有与 AtmiR827 的靶位点融合的 ZS-yellow 的植株的排水器中，没有观察到 YFP 表达降低，在含有与 AtmiR397 的靶位点融合的 ZS-yellow 的植株的根中，也没有观察到 YFP 表达降低。相对于单独的 ZS-yellow，含有与 ZS-yellow 融合的两个位点的植株同样没有表现出任何影响。相对于靶连接的 ZS-yellow 的水平，内源性 AtmiR827 和 AtmiR397 的水平可能太低而不能在相关组织中已经达到所希望的下调。与通常研究的 miRNA 例如 miR171 或 miR156 相比，这两种 miRNA 通常不是非常丰富；而在该实验中的靶基因序列是由强的病毒启动子 (35S) 驱动。应当注意，具有报道的由内源性 miRNA 介导的转基因下调的唯一组 (Parizotto 等人, (2004) Genes Dev 18: 2237–2242) 使用了 miR171——一种高丰度 miRNA，以下调含有相关靶位点的 35S-驱动的转基因。

实施例 11

选择性调节具有相差不大丰度的微 RNA 的转基因

掺入 AtmiR827 和 AtmiR397 靶位点的 YFP 转录物可能没有表现出改变的表达，这是由于这些 miRNA 在植物中的极低丰度，尤其是当与可能由 35S 启动子产生的高水平 YFP mRNA 相比时更是如此。因此，来自更高丰度 miRNA 的靶位点可能绕过该可能的问题。为此，可以将 ZS-yellow 编码区域单独与 AtmiR171 的靶序列在 5' 末端或者与 AtmiR156 的靶序列在 3' 末端融合(“单位点构建体”)。在第 3 种构建体中，在与单位点构建体相同的位置上将两个靶序列与 ZS-yellow 编码序列融合，由此产生“双位点构建体”。AtmiR171 和 AtmiR156 都是丰度较大的微 RNA，并且在多种组织中表达。在雌蕊中，AtmiR171 在心皮瓣中表达，但是胚座框(replum)、花柱和柱头中不存在；该模式与观察到的含有 AtmiR171 靶位点的报道基因(“传感器构建体”)的模式 (Parizotto 等人, (2004) Genes Dev 18: 2237–2242) 相反。而 AtmiR156 在雌蕊中的表达则限于花柱 (Rebecca Schwab, 2006, 来自 University of Tübingen 的论文)，因此，含有与 ZS-yellow 融合的 AtmiR156 靶位点的

单位点构建体应当在除了花柱之外的雌蕊的所有部分中具有 YFP 表达。预计，具有 ZS-yellow 的双位点构建体将仅在两种 miRNA 都不存在的区域，即胚座框和柱头中引起 YFP 表达。在雌蕊中的 YFP 表达方面，预计双位点构建体表现为两种单位点构建体的加合。

实施例 12

反式作用 miRNA 沉默

可以料想，使用组织或阶段特异性 miRNA，利用人工反式作用 siRNA (ta-siRNA)构建体，以仅在生物体特定组织或发育阶段沉默所关注基因。这种情况的目标是，通过使用与组织或阶段特异性 miRNA 相对应的 miRNA 靶作为引发子(trigger)序列(“引发子”是指引起下游 ta-siRNA 产生的靶序列)，以组织或阶段特异性方式产生人工 ta-siRNA。例如，如果希望沉默植物种子中的基因，可以选择仅在种子中存在的 miRNA 的靶序列作为引发子序列。如下所述产生这样的构建体：

构建嵌合多核苷酸，其中将组织或阶段特异性 miRNA 的靶位点用作引发子序列，并且与沉默子序列的 5'末端可操纵地连接。沉默子序列包含合成 DNA 片段，该合成 DNA 片段含有具有序列 [5'-TTGCTTTCTTCAGATCTCCCA-3'; SEQ ID NO: 7958]的与拟南芥脂肪酸脱饱和酶 2 (FAD2)基因互补的 21 核苷酸片段的 5 个重复拷贝。与 miRNA 互补的引发子序列的后面是 11 个核苷酸，这样 miRNA 裂解位点通过 21 个核苷酸与第一个 21 核苷酸 FAD2 片段隔开。在引发子和沉默子侧翼的序列源自 TAS1c 基因座(Allen 等人, (2005) Cell 121: 207-21)。35S 启动子和前导序列(Odell (1985) Nature 313: 810-812)连接在该嵌合构建体的 5' 末端，并且云扁豆蛋白转录终止子(Barr 等人, (2004) Molecular Breeding 13: 345-356)与 3'末端连接。将整个嵌合多核苷酸插入到标准二元载体 pBE851 (Aukerman 和 Sakai (2003) Plant Cell 15: 2730-41)内，并且使用 Clough 和 Bent (1998) Plant Journal 16: 735-43 的方法转化到拟南芥内。制备精确相同但是 miRNA 靶位点的 3 个核苷酸突变的构建体作为对照。使用脂肪酸分析(Browse 等人, (1986) Analytical Biochemistry 152: 141-145)监测含有实验构建体的转基因植株的 FAD2 基因沉默，并且与对照植株比较。沉默作为可检测的转录物水平降低和/或由这些转录物产生的基因产物减少来评估。

据信,本发明微 RNA 的任何靶都可以用作组织或阶段特异性引发子序列。此外,miRNA 序列的多联体(例如 SEQ ID NO: 7957 或 SEQ ID NO: 8427 中存在的)也可用作沉默子序列,其以由引发子序列指导的模式下调它们的内源性靶和/或这些靶的基因产物。使用 miRNA 靶序列的不同组合,可以想象得到地用单一构建体改变多个基因的表达模式。