



(12) 发明专利申请

(10) 申请公布号 CN 114255822 A

(43) 申请公布日 2022.03.29

(21) 申请号 202210079030.0

C12N 15/11 (2006.01)

(22) 申请日 2022.01.24

(71) 申请人 广东省林业科学研究院

地址 510520 广东省广州市天河区广汕一路233号

(72) 发明人 白青松 何波祥 汪迎利 连辉明 陈杰连

(74) 专利代理机构 广州市华学知识产权代理有限公司 44245

代理人 刘瑜

(51) Int. Cl.

G16B 20/20 (2019.01)

G16B 20/30 (2019.01)

G16B 30/10 (2019.01)

C12Q 1/6895 (2018.01)

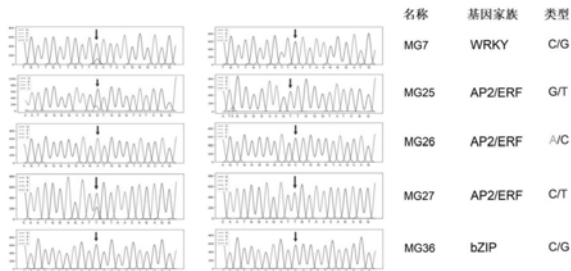
权利要求书3页 说明书13页
序列表12页 附图4页

(54) 发明名称

一种马尾松高产脂功能SNP标记的筛选方法及其应用

(57) 摘要

本发明中公开了一种马尾松高产脂功能SNP标记的筛选方法及其应用。该方法包括如下步骤:采集无性系马尾松的松脂并分别测定其产脂力大小,再从产脂力高、中、低的无性系马尾松样本中收集组织材料,提取RNA后进行PacBio全长转录组测序,同时开展二代转录组、代谢组、蛋白组测序并进行差异表达分析筛选候选基因,以此开发SNP功能标记。本发明方法只需针对少数样品进行转录组等多组学测序,具有较高的SNP位点开发效率,开发成本较低,研发的SNP标记有助于实现马尾松产脂力性状早期选择,可用在分子辅助选择育种领域。



1. 一种马尾松高产脂功能SNP标记的筛选方法,其特征在于,包括如下步骤:

(1) 依据表型选择材料

采集m个无性系马尾松的松脂,每个无性系马尾松选择n个植株,然后根据产脂力计算公式分别计算其产脂力,并计算每个植株产脂力的平均值;接着分别以每个无性系马尾松的产脂力最大的植株作为高产脂力植株,以产脂力最小的植株作为低产脂力植株,以产脂力等于或接近产脂力平均值的植株作为中产脂力植株;再分别采集每个无性系马尾松的高、中、低产脂力植株的次生木质部组织材料;其中, $m \geq 3$, $n \geq 3$;产脂力RYC的计算公式如下:

$$RYC = \frac{Wt}{(D \times Wd/C)}$$

式中:Wt表示松脂总重量;D表示每棵树切割树脂的次数;Wd表示剖面宽度;C表示树皮被切割处的树干周长;

(2) 基于多组学数据联合分析筛选候选基因

对步骤(1)中采集的每个无性系马尾松的高、中、低产脂力植株的组织材料分别进行转录组和代谢组测序,同时将所有组织材料混合后进行PacBio全长转录组测序;然后利用多组学数据进行差异表达分析,并根据转录组与代谢组联合分析初步筛选得到候选基因;

(3) qRT-PCR验证基因表达量

分别提取每个无性系马尾松的高、中、低产脂力植株的组织材料的RNA,并将其反转录为cDNA,然后采用qRT-PCR方法验证步骤(2)中筛选得到的候选基因的表达量,选择差异表达的基因用于SNP位点开发;

(4) 开发SNP位点

分别提取每个无性系马尾松的高、中、低产脂力植株的组织材料的DNA,然后进行PCR全长扩增,再直接进行高、中、低产脂力植株的序列相互比较开发SNP分子标记,筛选得到SNP位点;

(5) 开发产脂力性状功能SNP位点

根据步骤(4)中筛选得到的SNP位点,对每个无性系马尾松的高、中、低产脂力植株的基因型进行关联分析,得到与产脂力性状显著相关的SNP标记,进而筛选得到马尾松高产脂功能SNP标记。

2. 根据权利要求1所述的马尾松高产脂功能SNP标记的筛选方法,其特征在于:

步骤(1)中所述的m的取值范围为: $m \geq 20$;

步骤(1)中所述的n的取值范围为: $n \geq 5$;

步骤(2)中所述的候选基因为选择3个以上的候选基因。

3. 根据权利要求2所述的马尾松高产脂功能SNP标记的筛选方法,其特征在于:

步骤(1)中所述的m的取值范围为: $m \geq 50$;

步骤(1)中所述的n的取值范围为: $5 \leq n \leq 20$;

步骤(2)中所述的候选基因为选择5个以上的候选基因。

4. 根据权利要求3所述的马尾松高产脂功能SNP标记的筛选方法,其特征在于:

步骤(2)中所述的候选基因为MG7、MG25、MG26、MG27、MG36基因,其核苷酸序列如SEQ ID NO.1~5所示。

5. 根据权利要求1所述的马尾松高产脂功能SNP标记的筛选方法,其特征在于:
 - 步骤(1)中所述的松脂的采集时间为每年的7~9月;
 - 步骤(1)中所述的组织材料的采集部位高度与割脂部位相同,采集时间以中午12点~1点晴朗天气;
 - 步骤(1)中所述的中产脂力植株选择产脂力为中位数对应的植株作为中产脂力植株;
 - 步骤(1)中所述的无性系马尾松的树龄10年生以上或胸径达到16cm以上的无性系马尾松;
 - 步骤(1)中所述的组织材料包括次生木质部、成熟针叶、成熟枝条或未成熟枝条。
6. 根据权利要求1所述的马尾松高产脂功能SNP标记的筛选方法,其特征在于:
 - 步骤(3)中所述的qRT-PCR方法中所用的引物序列如SEQ ID NO.19~28所示;
 - 步骤(3)中所述的qRT-PCR方法中的扩增反应程序为:变性温度95℃持续10秒,退火温度57℃~62℃持续30秒,延伸温度72℃持续20秒;
 - 步骤(4)中所述的PCR全长扩增所用的引物序列如SEQ ID NO.29~38所示;
 - 步骤(4)中所述的PCR全长扩增的反应程序为:预变性温度94℃持续5分钟,变性温度94℃持续1分钟,退火温度55℃~63℃持续30秒,延伸温度72℃持续时间1~3分钟。
7. 根据权利要求1所述的马尾松高产脂功能SNP标记的筛选方法,其特征在于,在步骤(2)之后、步骤(3)之前还包括如下至少一个步骤:
 - a、利用GO、KEGG数据库对差异基因进行功能和代谢通路注释,并选择功能基因并使用候选基因FPKM值进行表达分析;
 - b、采集的每个无性系马尾松的高、中、低产脂力植株的其他组织材料,分析候选基因的表达量,即对候选基因进行表达验证;
 - c、根据PacBio全长转录组测序数据库获得候选基因的全长序列,利用NCBI保守结构域搜索软件和Pfam保守结构域数据库分析确定差异基因的所属基因家族。
8. 根据权利要求1所述的马尾松高产脂功能SNP标记的筛选方法,其特征在于:在步骤(4)之后、步骤(5)之前还包括SNP位点在群体的多态性验证的步骤,具体如下:

在SNP标记所在位点的两端序列设计引物,然后提取马尾松种质资源的基因组DNA,进行PCR扩增与序列比对,验证SNP标记与产脂力性状的相关性。
9. 权利要求1~8任一项所述的马尾松高产脂功能SNP标记的筛选方法在筛选马尾松高产脂功能SNP标记或马尾松育种方面的应用。
10. 一种马尾松高产脂功能SNP标记,其特征在于:所述的SNP标记为如下任一种:
 - (1) 位于SEQ ID NO.6所示的核苷酸序列,其中自5'端起第76位碱基中的等位基因为T/C;
 - (2) 位于SEQ ID NO.7所示的核苷酸序列,其中自5'端起第78位碱基和第1073位碱基中的等位基因均为T/C;
 - (3) 位于SEQ ID NO.8所示的核苷酸序列,其中自5'端起第169位碱基中的等位基因为C/T;
 - (4) 位于SEQ ID NO.9所示的核苷酸序列,其中自5'端起第128位碱基中的等位基因为T/G;
 - (5) 位于SEQ ID NO.10所示的核苷酸序列,其中自5'端起第103位碱基中的等位基因为

G/T,或第194位碱基中的等位基因A/G;

(6) 位于SEQ ID NO.11所示的核苷酸序列,其中自5'端起第130位碱基中的等位基因为G/T;

(7) 位于SEQ ID NO.12所示的核苷酸序列,其中自5'端起第29位碱基中的等位基因为C/T;

(8) 位于SEQ ID NO.13所示的核苷酸序列,其中自5'端起第73位碱基中的等位基因为T/A,或第486位碱基中的等位基因为G/C;

(9) 位于SEQ ID NO.14所示的核苷酸序列,其中自5'端起第109位碱基中的等位基因为A/G;

(10) 位于SEQ ID NO.15所示的核苷酸序列,其中自5'端起第143位碱基中的等位基因为C/T;

(11) 位于SEQ ID NO.16所示的核苷酸序列,其中自5'端起第125位碱基中的等位基因为G/T,第527位碱基中的等位基因为G/T,第895位碱基中的等位基因为G/C,或第1087位碱基中的等位基因为T/G;

(12) 位于SEQ ID NO.17所示的核苷酸序列,其中自5'端起第138位碱基中的等位基因为A/G;

(13) 位于SEQ ID NO.18所示的核苷酸序列,其中自5'端起第97位碱基中的等位基因为A/C。

一种马尾松高产脂功能SNP标记的筛选方法及其应用

技术领域

[0001] 本发明属于林业技术领域,特别涉及一种马尾松高产脂功能SNP标记的筛选方法及其应用。

背景技术

[0002] 研发目标性状功能标记是植物分子辅助育种的常用方法,也是实现优良个体早期选择的重要手段。现在常用的SNP开发技术包括:基于酶切的简化基因组测序(SLAF-seq、GBS、RAD等)、全基因组重测序(genome re-sequencing)等。这些方法通常需要先收集大量的群体遗传资源,在提取基因组DNA后,采用酶切或随机打断的方式对群体全部个体DNA进行测序,并通过序列比较的方式开发群体SNP分子标记。随后,利用GWAS方法开发目标性状关联标记是目前常用的手段。

[0003] 由于SNP开发技术要对群体内的每个个体开展测序,通常需要消耗大量的研发费用。而基于酶切的简化基因组测序所开发的SNP位点通常位于非编码区,不利于挖掘功能基因。全基因组重测序要根据基因组大小和测序深度计费,在面对如马尾松等具有较大基因组(通常在20GB以上)的物种时所消耗研发费用过高,不具有广泛的适用性。此外,利用GWAS方法所得到的关联标记通常需要群体PCR方法验证,并且所得到的关联标记假阳性率较高。

发明内容

[0004] 本发明的首要目的在于克服现有技术的缺点与不足,提供一种马尾松高产脂功能SNP标记的筛选方法。

[0005] 本发明的另一目的在于提供所述马尾松高产脂功能SNP标记的筛选方法的应用。

[0006] 本发明的目的通过下述技术方案实现:

[0007] 一种马尾松高产脂功能SNP标记的筛选方法,包括如下步骤:

[0008] (1) 依据表型选择材料

[0009] 采集m个无性系马尾松的松脂,每个无性系马尾松选择n个植株,然后根据产脂力计算公式分别计算其产脂力,并计算每个植株产脂力的平均值;接着分别以每个无性系马尾松的产脂力最大的植株作为高产脂力植株,以产脂力最小的植株作为低产脂力植株,以产脂力等于或接近产脂力平均值的植株作为中产脂力植株;再分别采集每个无性系马尾松的高、中、低产脂力植株的次生木质部组织材料;其中, $m \geq 3$, $n \geq 3$;产脂力RYC的计算公式如下:

$$[0010] \quad RYC = \frac{Wt}{(D \times Wd/C)}$$

[0011] 式中:Wt表示松脂总重量;D表示每棵树切割树脂的次数;Wd表示剖面宽度;C表示树皮被切割处的树干周长;

[0012] (2) 基于多组学数据联合分析筛选候选基因

[0013] 对步骤(1)中采集的每个无性系马尾松的高、中、低产脂力植株的组织材料分别进

行转录组和代谢组测序,同时将所有组织材料混合后进行PacBio全长转录组测序;然后利用多组学数据进行差异表达分析,并根据转录组与代谢组联合分析初步筛选得到候选基因;

[0014] (3) qRT-PCR验证基因表达量

[0015] 分别提取每个无性系马尾松的高、中、低产脂力植株的组织材料的RNA,并将其反转录为cDNA,然后采用qRT-PCR方法验证步骤(2)中筛选得到的候选基因的表达量,选择差异表达的基因用于SNP位点开发;

[0016] (4) 开发SNP位点

[0017] 分别提取每个无性系马尾松的高、中、低产脂力植株的组织材料的DNA,然后进行PCR全长扩增(基因全长扩增),再直接进行高、中、低产脂力植株的序列相互比较开发SNP分子标记,筛选得到SNP位点(SNP/Indel等位变异位点);

[0018] (5) 开发产脂力性状功能SNP位点

[0019] 根据步骤(4)中筛选得到的SNP位点,对每个无性系马尾松的高、中、低产脂力植株的基因型进行关联分析,得到与产脂力性状显著相关的SNP标记,进而筛选得到马尾松高产脂功能SNP标记。

[0020] 步骤(1)中所述的无性系马尾松的个数(m的取值)可以根据实际需要进行选择;优选为 $m \geq 20$;进一步优选为 $m \geq 50$;再进一步优选为 $m \geq 150$ 。

[0021] 步骤(1)中所述的每个无性系马尾松的植株数(n的取值)可以根据实际需要进行选择;优选为 $n \geq 5$;更优选为: $5 \leq n \leq 20$ 。

[0022] 步骤(1)中所述的高、中、低产脂力植株可根据实际需要选择生物学重复的次数,优选为3次以上(即各个无性系马尾松各3个以上单株作为生物学重复)。

[0023] 步骤(1)中所述的松脂的采集时间优选为每年的7~9月。

[0024] 步骤(1)中所述的组织材料的采集部位高度与割脂部位相同(每个单株需收集约15克),采集时间以中午12点~1点晴朗天气为宜。

[0025] 步骤(1)中所述的无性系马尾松的树龄10年生以上或胸径达到16cm以上的无性系马尾松;优选为29年生无性系马尾松。

[0026] 步骤(1)中所述的中产脂力植株优选为选择产脂力为中位数对应的植株作为中产脂力植株。

[0027] 步骤(1)中所述的组织材料包括次生木质部、成熟针叶、成熟枝条或未成熟枝条;优选为次生木质部的组织材料。

[0028] 步骤(2)中所述的转录组筛选差异表达基因的条件为: $\log_2^{\text{Ratio}} \geq 1.0$ 和 $P\text{-value} < 0.05$ 。

[0029] 步骤(2)中所述的代谢组筛选差异代谢物的条件为: $\log_2^{\text{(fold change)}} > 0.5$ 和 $q\text{-value} < 0.01$ 。

[0030] 步骤(2)中所述的蛋白组筛选差异蛋白的条件为: $\text{Fold change} \geq 1.2$ 和 $P < 0.5$ 。

[0031] 步骤(2)中所述的联合分析筛选差异基因的条件为不同组学数据的差异倍数相关性 $R^2 > 0.9$ 。

[0032] 步骤(2)中所述的候选基因可以从转录因子及其他基因中进行选择,优选为选择3个以上的候选基因;进一步优先为选择5个以上的候选基因;更优选为MG7、MG25、MG26、

MG27、MG36基因,其核苷酸序列如SEQ ID NO.1~5所示。

[0033] 所述的马尾松高产脂功能SNP标记的筛选方法,在步骤(2)之后、步骤(3)之前还包括如下至少一个步骤:

[0034] a、利用GO、KEGG等数据库对差异基因进行功能和代谢通路注释,并选择功能基因并使用候选基因FPKM值进行表达分析;

[0035] b、采集的每个无性系马尾松的高、中、低产脂力植株的其他组织材料,分析候选基因的表达量,即对候选基因进行表达验证;

[0036] c、根据PacBio全长转录组测序数据库获得候选基因的全长序列,利用NCBI保守结构域搜索软件和Pfam保守结构域数据库分析确定差异基因的所属基因家族。

[0037] 步骤(3)中所述的qRT-PCR方法中所用的引物序列优选为如SEQ ID NO.19~28所示。

[0038] 步骤(3)中所述的qRT-PCR方法中的扩增反应程序为:变性温度95℃持续10秒,退火温度57℃~62℃持续30秒,延伸温度72℃持续20秒。

[0039] 步骤(4)中所述的PCR全长扩增(基因全长扩增)的反应程序为:预变性温度94℃持续5分钟,变性温度94℃持续1分钟,退火温度55℃~63℃持续30秒,延伸温度72℃持续时间(约1kb/分钟)1~3分钟。

[0040] 步骤(4)中所述的PCR全长扩增所用的引物序列优选为如SEQ ID NO.29~38所示。

[0041] 步骤(4)中所述的开发SNP分子标记优选为在转录因子基因序列上进行开发。

[0042] 所述的马尾松高产脂功能SNP标记的筛选方法,在步骤(4)之后、步骤(5)之前还包括SNP位点在群体的多态性验证的步骤,具体如下:在SNP标记所在位点的两端特意序列设计引物,然后提取马尾松种质资源的基因组DNA,进行PCR扩增与序列比对,验证SNP标记与产脂力性状的相关性。

[0043] 所述的马尾松高产脂功能SNP标记的筛选方法在筛选马尾松高产脂功能SNP标记或马尾松育种方面的应用,该方法获得的马尾松高产脂功能SNP标记有助于实现马尾松产脂力性状的早期选择,(如筛选出高产脂马尾松),可用在分子辅助选择育种领域。

[0044] 一种马尾松高产脂功能SNP标记,所述的SNP标记为如下任一种:

[0045] (1)位于SEQ ID NO.6所示的核苷酸序列,其中自5'端起第76位碱基中的等位基因为T/C;

[0046] (2)位于SEQ ID NO.7所示的核苷酸序列,其中自5'端起第78位碱基和第1073位碱基中的等位基因均为T/C;

[0047] (3)位于SEQ ID NO.8所示的核苷酸序列,其中自5'端起第169位碱基中的等位基因为C/T;

[0048] (4)位于SEQ ID NO.9所示的核苷酸序列,其中自5'端起第128位碱基中的等位基因为T/G;

[0049] (5)位于SEQ ID NO.10所示的核苷酸序列,其中自5'端起第103位碱基中的等位基因为G/T,或第194位碱基中的等位基因为A/G;

[0050] (6)位于SEQ ID NO.11所示的核苷酸序列,其中自5'端起第130位碱基中的等位基因为G/T;

[0051] (7)位于SEQ ID NO.12所示的核苷酸序列,其中自5'端起第29位碱基中的等位基

因为C/T;

[0052] (8) 位于SEQ ID NO.13所示的核苷酸序列,其中自5'端起第73位碱基中的等位基因为T/A,或第486位碱基中的等位基因为G/C;

[0053] (9) 位于SEQ ID NO.14所示的核苷酸序列,其中自5'端起第109位碱基中的等位基因为A/G;

[0054] (10) 位于SEQ ID NO.15所示的核苷酸序列,其中自5'端起第143位碱基中的等位基因为C/T;

[0055] (11) 位于SEQ ID NO.16所示的核苷酸序列,其中自5'端起第125位碱基中的等位基因为G/T,第527位碱基中的等位基因为G/T,第895位碱基中的等位基因为G/C,或第1087位碱基中的等位基因为T/G;

[0056] (12) 位于SEQ ID NO.17所示的核苷酸序列,其中自5'端起第138位碱基中的等位基因为A/G;

[0057] (13) 位于SEQ ID NO.18所示的核苷酸序列,其中自5'端起第97位碱基中的等位基因为A/C。

[0058] 所述的SNP标记中,(2)、(3)、(6)、(7)以及(5)中第194位碱基中的等位基因为A/G的马尾松无性系植株的产脂力相对高于其他SNP标记中的马尾松无性系植株。

[0059] 本发明相对于现有技术具有如下的优点及效果:

[0060] (1) 本发明提供了一种开发目标性状功能SNP位点的技术,以马尾松高产脂性状为例,研发与目标性状关联的功能标记,实现了多组学测序技术与极端性状表型的有效结合,大幅降低研发成本,且材料选择上只需选择具有最高、中等、最低产脂力的马尾松品系收集组织材料,提取RNA后等量混合用来进行PacBio全长转录组测序,同时开展二代转录组、代谢组、蛋白组测序并进行差异表达分析,测序成本较低,该方法适用于不同物种不同性状,具有较高的适用性,尤其在面对较大基因组物种时具有明显优势。

[0061] (2) 本发明简化了功能SNP研发技术流程,现有技术常采用:表型测定+全部样本重测序+群体开发SNP位点+全基因组关联分析开发SNP功能标记的方法,而本发明采用:表型测定+少量样本多组学测序+筛选候选基因+开发SNP功能标记的技术流程。

[0062] (3) 本发明针对目标性状差异表型,利用多组学精准定位候选基因,并通过qRT-PCR验证差异表达,接着利用PCR针对特异片段开展群体验证,即先选择差异基因,再进行序列扩增比对,具有较高的SNP位点开发效率。

[0063] (4) 本发明对新选择基因开发SNP时,仅需选择高、中、低产脂力马尾松各3个无性系就可开发与目标性状相关SNP位点,降低SNP研发成本,本技术只需针对少数样品进行转录组等多组学测序,成本较低,研发的SNP标记有助于实现马尾松产脂力性状早期选择。

[0064] (5) 本发明适用性更广,研发费用更低,现有技术常采用全基因组重测序开发群体SNP标记,测序个体数量大,所需研发费用高,并且该技术不适于拥有较大基因组的物种,而大基因组物种常用的简化基因组测序技术不利于挖掘功能基因,而本发明仅需对少量拥有极端性状表型的个体进行多组学测序,直接定位候选功能基因,所需研发费用较低,且适于多种性状和不同大小基因组的物种,尤其具有较大基因组的物种。

附图说明

- [0065] 图1是本发明技术流程图。
- [0066] 图2是多组学(转录组、代谢组、蛋白组)所定位的候选基因在数量上的差异结果图。
- [0067] 图3是候选基因FPKM表达分析图。
- [0068] 图4是候选基因的保守结构域的鉴定结果图。
- [0069] 图5是候选基因的qRT-PCR验证结果图。
- [0070] 图6是SNP标记等位基因类型图。
- [0071] 图7是SNP位点基因型频率图。

具体实施方式

[0072] 下面结合实施例对本发明作进一步详细的描述,但本发明的实施方式不限于此。除非特别说明,本发明采用的试剂、方法和设备为本技术领域常规试剂、方法和设备。下列实施例中未注明具体实验条件的试验方法,通常按照常规实验条件。

[0073] 实施例1

[0074] (1) 技术流程

[0075] 本发明主要采用多组学筛选候选基因,接着用PCR对候选基因开发SNP位点,并经群体验证获得功能标记,技术流程图见图1。

[0076] (2) 依据表型选择材料

[0077] 利用收集和保存在马尾松种子园的150份马尾松无性系种质资源(共150个无性系,每个无性系5~20个单株,这些单株相当于生物学重复;这些实验材料均为常规材料,保存于广东省林业科学研究院的种子园)开展产脂力(Resin-yielding capacity,RYC)计算,可针对各树龄(一般马尾松10年生或胸径达到16cm以上的就可以采脂,本实验采用的是29年生马尾松测定的数据,这些树的胸径虽大小不一,但产脂力计算公式可消除这一差异)马尾松开展,根据马尾松采脂特点(采集部位一般低于胸径的位置,采脂的时候会测量割取位置的周长,用于计算产脂力(具体参考文献:张谦,曾令海,何波祥,等.马尾松自由授粉家系产脂力的年度变化及遗传分析[J].林业科学,2013,49(001):48-52.)),以每年的7-9月为最佳产脂力测定和样品收集时间,RYC计算公式如下:

$$[0078] \quad RYC = \frac{Wt}{(D \times Wd/C)}$$

[0079] 公式中:Wt表示松脂总重量,D表示每棵树切割树脂的次数,Wd表示剖面宽度,C表示树皮被切割处的树干周长。

[0080] 根据上述计算的各无性系产脂力平均值选择表型极端的无性系(高、中、低产脂力,表性差异显著(F检验,P<0.05)),其中,以每个无性系马尾松的产脂力最高的植株作为高产脂力植株,以产脂力最低的植株作为低产脂力植株,以产脂力等于或接近产脂力平均值的植株作为中产脂力植株(本实验选择这组数中的中位数)。本实验筛选得到的3个无性系植株为GW2(高产脂)、GW9(中产脂)、GW92(低产脂),已在参考文献(Qingsong Bai, Yanling Cai, Boxiang He, et al. Core set construction and association analysis of *Pinus massoniana* from Guangdong province in southern China using SLAF-

seq.Scientific Reports.2019.(9):13157.doi:10.1038/s41598-019-49737-2.)中公开),各选3个单株采集次生木质部组织材料(高、中、低产脂力马尾松各3个单株,相当于3次生物学重复),材料采集部位高度与割脂部位相同,采集材料时需先将树皮剥去,待露出次生木质部组织时用小刀刮取组织材料,每个单株共需收集约15克,采集时间以中午12点-1点晴朗天气为宜。

[0081] 随后,开展多组学测序,分别对上述采集的3个无性系高、中、低产脂力共9个单株的次生木质部组织材料进行测序,包括:转录组、代谢组、蛋白组测序,每个无性系3个单株作为生物学重复独立测序;同时,将3个无性系共9个单株RNA样品等量混合开展PacBio全长转录组测序。本发明委托北京诺禾致源生物科技有限公司进行转录组、代谢组、蛋白组测序。

[0082] (3) 基于多组学数据筛选精准定位候选基因

[0083] 利用多组学数据开展差异表达分析与联合分析筛选候选基因,其中:转录组筛选差异表达基因的条件为 $\log_2^{\text{Ratio}} \geq 1.0$ 和 $P\text{-value} < 0.05$;代谢组筛选差异代谢物的条件为 $\log_2^{\text{(fold change)}} > 0.5$ 和 $q\text{-value} < 0.01$;蛋白组筛选差异蛋白的条件为 $\text{Fold change} \geq 1.2$ 和 $P < 0.5$;联合分析筛选差异基因的条件为不同组学数据的差异倍数相关性 $R^2 > 0.9$ 。由于结合了转录组、代谢组、蛋白组所定位的候选基因数量大幅降低(图2),只依靠转录组可得到28075个差异表达基因,转录组与蛋白组联合分析则将差异基因数量缩减至2821个,转录组与代谢组联合分析的差异基因数量为226个。

[0084] 利用GO、KEGG等数据库对差异基因进行功能和代谢通路注释,选择转录因子等功能基因MG7、MG25、MG26、MG27、MG36(利用这5个基因主要是想说明在群体间存在较高的多样性)并使用转录组得到的候选基因FPKM值进行表达分析(图3)。在分析表达量时也从上述3个无性系高、中、低产脂力共9个单株中选用无病虫害的已发育成熟的针叶、未成熟的枝条和成熟的枝条等其他组织用于候选基因的表达验证,MG7、MG25、MG26、MG27、MG36等候选基因在高、中、低产脂力马尾松不同组织差异表达。其中,候选基因序列如下所示:

[0085] MG7 (SEQ ID NO.1):

[0086] ATGCAGAAGATTGTGGATCAAACCTGATGCAACTATAAGAAAAGCCCGAGTTTCAGTCCGAGCTAGAACCGAGTCACCCATGATAAGTGATGGTTGCCAATGGAGGAAATATGGACAAAAGATGGCCAAGGGCAATCCATGCCCCAGAGCTTACTACAGATGCACCATGTCACCGTGCTGTCCAGTGCAGCAAGCAGGTGCAACGCCTGGCCGAAGACAGATCGATTCTGATAACAACATACGAGGGCAGCCATAACCACATGCTTCTCCAGCAGCCACAGCCATGGCATCCACTACAGCAGCAGCTGCCTCAATGCTTCTGTCCGGTCTTCCACATCAGCTGATAACATCGCACTAAATGCAAGCTTCATGGCAGGTGCCCTCATGCAGCATCCTTGCAACACATCAACTGCTAGCATCTCGGCCTCAGCTCCATTCCCCACTATCACTACTGGATCTCACACACAATCCCAACCAATGGCCAACGCTCCAGGTCACATGGCAGCCTCAAACCAAGAGCACTAGCAGGACTCCCAGCTCATGCCATGCCCTTTGTAGGCATGCCACACCAATTCCCCACCAACACACCCAGGGGGCAATTATTCCATGGGCAATCAATTTACAATGCTCCTTCCATGTTTGCACCCTTAGCAGGCCAGCTGCAACGCCCAACAAACCCATGATGCCACACCACCAAGCTCAATATTCAAGCTGGCCAACCACCAACACCACCACAACAACAACCATCATTCATAGACACAGTCAGTGCAGCTACAGCTGCCATAAATTCTGATCCCAATTTACAGCAGCCCTTGCAGCAGCCATCACATCCCTTATGAACAACAACAATGCTGCCAATGCCATTTCTAAACCAACTCCTCTCAGGCACTCAATCTAGTCCCCACGTTGCAGCTCATGTCAAGACAGATCACACCCAA;

[0087] MG25 (SEQ ID NO.2):

[0088] ATGGAAAACCTCCCCAATCAGCAACCTGACCTTCAAATTGCTCAAACACACGAGGATCCCGGGTGCCG
CCGATTTAAGGGAATTCGACTGCGAAAATGGGGAAGGTGGGTATCGGAAATCCGGATGCCAAAATCTCGAGAGAAA
ATATGGCTGGGCTCTTATACGACTCCCGAGCAGGCTGCCCGTGTCTTACGACGCCGCAGTGTATTGTCTGAGAGGGC
CCAACGCCAAATTTAACTTTCCGGAATCCGTGCACGACATTCCGTCTGTGACTTCTGTTTCCCGTCAGGAAATTCA
GCACGCCGCTTCAAATATGCCTTGGGCCAGCCCCCTCCGAGTTTGCAGTCTCTGGAAGGGCAGCCGCCCTCAA
TATGCCTTGGGCCAGCCCCCTCCGAGTTTGCAGTCTCTGGAAGAAGGGCATGCCGCCCTCAAATATGCCTTGGGCC
AGCCTCTCCGAGTTTGCAGTCTCTGGAAGGGCAGCGTCCCGTCACAGTCGTCTACGGTTTTCGAAACGGAGTT
ATCGGGAGAACAGCTGAAGATATCGGAAGAGTGCCCCGACTTAGCACTGTGTTGGTCGCTGTTTGCGGCAGACGAC
ACTGGGATTCCAATTCGAAAAAGTCCCGTCGATTGACGAATATTTTCAGTGCAGTCTTGCAGGAGCAGCGGGAGG
AGGGTTACATTTTACAGATTTGTGGAATTTCCAAGATCAAGATGTT;

[0089] MG26 (SEQ ID NO.3) :

[0090] ATGGAGAAATCCTCCCAACAGGAGGATGACCATGCCATACTCCAGAAGAAGAAGTTCGCGGGCAAAA
GTGCCGTCAATTTAAGGGAACCCGATTGCGAAAGTGGGGGAGATGGGTAGCAGAAATCGAATGCCAAAATCTCGA
GAGAAGTTATGGCTGGGGTCATACAAAAAGCCCGAGCAGGCCGCCCGCCTACGACGCCGCAGTGTATTGTCTGA
GAGGGCCGAACGCCAAATTTCAATTTACCCAATTCTCTACCTGACATTCCGTCTGCGTCTTCTCTTTCCCGCCGGCA
GATTCAACTCGCTGCTGCCAAATGTGCGTTGGATCAATTCCTTCGAGTGCGCCCCCTCTGCAGAATTTAATAAT
AAGGCCATGGACGAGGCCGCATCGCCGTCAAGACTGGATCCGGTATCAGAACTGAGTTGTCGAGCGATGGTCATC
AAATATCAGAGGAAGGGGAGTTGGATTTGTGGAAAGTCCGTTTGAGGTATCAGGCGCAATTATGAAGGGCGCAT
GAACCTGAATTTAGAGAGAATGCCATCGATTGAGGAGTTCTCGGCCCTTGAAATTTTACAGTATTTGTCAGCAG
CATGAGGAGGAGGAACACATAAACCTTTTCTCGACCCACAGAGTTGTGGAACCTT;

[0091] MG27 (SEQ ID NO.4) :

[0092] ATGGATAATGGATCCTCTCTCGTGCCCATCGCCATGCCAATTCCTTGACAGACATTGAAGCAATAAG
CAATTCTCCTTTTTCGGGATAAGGTGGAAACAAGAGAATTCGAACGCAAGACGAAGCTGCCTCTTCGCCTTCACAG
CAAAGCAGGCTGAACCTGCAAAATCCAGCATAACAGAGGCGTGCGCCGTCGTAGCTGGGGCAAATGGGTGTCTGAAA
TTCGGGAACCAAAAAAGAAAAACCGAATCTGGCTCGGCTCCTACGATACACCAGAGATGGCCGCTCGAGCTCACGA
TGTCGCTGCATTCTACTTGAAAGGAAAGAAACATTGCTTGCTCAATTTTCCAGAGCTCATTGATCAACTTCAGAA
CCAATTTCTTCGGCTCCGCCCCACATTCAAGCTGCCCGGCAGCAGCAGCCGTCGCTTCAATTCTGCATCCCGTA
GTGCACAGAACTCTGGAATGTCAAGCGATATCAATAACAAGACAGCGGAAGGCCAAGAAATACTCCAGTAAGCAA
TGAGCACGCAGGTATCTCCTCTTCGAATCAACTATCGGCAGTAATTAGCTCAGAGTTAACTTGGAGACGGCTAAT
GTCGAGTTATTAGGAAACCGAATTATACTAATTCATCATCAATGGAGATCGTAACAGAGGAGGACTTGTTCAAT
CCACCAACTTCTATACGAATTTGGCAGAAGGCCTTATGCTTCTCCACCTCTGTTTCAGTATTCCTGAATTAGACAT
AGAGGAGCAGAGATTGGAGGAAGGATTTCTTTGGTCTGGTTTT;

[0093] MG36 (SEQ ID NO.5) :

[0094] ATGGCTGGCATGGATGACGGAGACATTAATTTTAGCAGCAATATAGTAGATGATTTTGGCAATGGGTC
CTCCATGGAAAGCTTTTTCGAGGAGATTTTGGGGATACTACTCATGCCTGCACTCATAACACACCTGCAACCCT
CCCGGGCCAGATAACACACATACACACATGCTTCCACACGCACACAAAAATCCTTGCTGCTCCCGATGACGAGA
AGTCTGCAGACACTGCTGAGTCTCCACAAAACAGCTTTCGAAGCAAAAAACGACCAGTAGGTAATCGAGAGGC
AGTTAGAAAATACAGGGAAAAAAGGCCCGACGGCCTCCCTGGAGGAGCAGGTTGTTCAACTGACCACTGTT
AATCAGCAATTGCATAGGAGATTACAGGGTCAAGCAGCTTTAGAGGCTGAGATTGCAAGATTGAAGTGTCTGCTGG

CTGACTTTAGGGGCCGGATCGATGGGGAATTGGGGTCTATCCTTACCAAAAGTCAATTAGAATGGATAAACTTG
CAATGATGCACCATTCGCGCAACCGATGCCTGGGGGATATGTCCTGGATCCCTGCAATATCTGGTGAATGCAGAT
GCAGCTTGCCGTGAACCGACTCTGGCATCCAACAGTGAGGGTGGTGTGCAGCACGAACGTGATAGTGCTGCACGTT
GGAATGGCGATTGTGGCCAGATCGCAGGCCATTGTCAAGGTTTGAAGGGTGTATGGCAGTAACCTCTAGTGGACT
TTCTGGATGCTCAGAGGTACAGCAACAAAAACAGTGCCTGCTGCCATGGCTTCTTCTGGAAAAGAGAGAAAAGGT
GCATTTGGTGTG。

[0095] 除了GO和KEGG注释外,根据PacBio全长转录组测序数据库获得候选基因的全长序列,利用NCBI保守结构域搜索软件和Pfam保守结构域数据库分析确定差异基因的所属基因家族(图4),为后续研究其生物功能提供基础。

[0096] (4) qRT-PCR验证基因表达量

[0097] 利用TriZol法分别提取高、中、低产脂力马尾松不同组织(成熟针叶、成熟枝条、未成熟枝条、次生木质部)的RNA,用反转录试剂盒获得cDNA,采用qRT-PCR方法验证初筛差异基因的表达量。根据获得的候选基因全长序列并利用Primer 3设计引物(表1)。选择Actin作为内参基因,扩增反应程序设置如下:变性温度95℃持续10秒,退火温度57℃~62℃持续30秒,延伸温度72℃持续20秒,每个基因每个样品均为三次重复。

[0098] 表1候选基因qRT-PCR引物

| 基因名称 | 扩增产物长度 | 上游(F)/下游(R)引物 | 引物序列(5'-3') |
|------|--------|---------------|--------------------------------------|
| MG7 | 153 | F | CCACACCACCAAAGCTCAAT (SEQ ID NO.19) |
| | | R | TAAGGGATGTGATGGCTGCT (SEQ ID NO.20) |
| MG25 | 168 | F | CATGCCGCCCTCAAATATGC (SEQ ID NO.21) |
| | | R | AAACAGCGACCAACACAGTG (SEQ ID NO.22) |
| MG26 | 187 | F | TCAGAGGAAGGGGAGTTGGA (SEQ ID NO.23) |
| | | R | TGGGGTCGAGGAAAAGGTTT (SEQ ID NO.24) |
| MG27 | 152 | F | GAGTTAAACTTGGAGACGGCT (SEQ ID NO.25) |
| | | R | AGAGGTGGAGGAAGCATAAGG (SEQ ID NO.26) |
| MG36 | 163 | F | GAACCGACTCTGGCATCCA (SEQ ID NO.27) |
| | | R | CTGTACCCTCTGAGCATCCA (SEQ ID NO.28) |

[0100] 反应结束后,利用 $2^{-\Delta\Delta Ct}$ 方法对候选基因进行相对表达量分析,以次生木质部等不同组织高产脂力马尾松作为对照,分析MG7、MG25、MG26、MG27、MG36在中、低产脂力马尾松的表达量(图5),选择差异表达的基因用于SNP位点开发。

[0101] (5) 开发SNP位点

[0102] 对经qRT-PCR验证的差异表达基因MG7、MG25、MG26、MG27、MG36开发SNP分子标记(图6)。收集不同无性系马尾松的成熟针叶,提取基因组DNA(高、中、低各选一个单株分别提取DNA),利用转录组数据获得候选基因的5'端序列、编码序列、3'端序列并进行PCR全长扩增,首先设计扩增引物(表2),PCR反应程序包括:预变性温度94℃持续5分钟,变性温度94℃持续1分钟,退火温度根据引物特性一般在55℃~63℃之间持续30秒,延伸温度72℃持续时间(约1kb/分钟)根据序列长度而定一般在1~3分钟。根据不同序列特征,一些序列需经过多次PCR扩增加拼接的到完整序列。初步分析显示,在WRKY、AP2/ERF、bZIP等转录因子基因序列开发得到SNP分子标记,等位基因类型包括C/G、G/T、A/C、C/T等。

[0103] 表2候选基因扩增引物

| 基因名称 | 扩增产物长度 | 上游(F)/下游(R)引物 | | 引物序列 (5'-3') |
|------|--------|---------------|---|--|
| | | F | R | |
| MG7 | 948 | F | | ATGCAGAAGATTGTGGATCAAAC TG (SEQ ID NO.29) |
| | | | R | TTGGGTGTGATCTGTCTTGACATG (SEQ ID NO.30) |
| MG25 | 723 | F | | ATGGAAAACCTCCCCAATCAGC (SEQ ID NO.31) |
| | | | R | AACATCTTGATCTTGAAATTCCACAAATC (SEQ ID NO.32) |
| MG26 | 657 | F | | ATGGAGAAATCCTCCCAACAGGAG (SEQ ID NO.33) |
| | | | R | AAAGTTCCACAACCTCTGTGGG (SEQ ID NO.34) |
| MG27 | 795 | F | | ATGGATAATGGATCCTCTCTCGTG (SEQ ID NO.35) |
| | | | R | AAAACCAGACCAAAGAAATCCTTCCTC (SEQ ID NO.36) |
| MG36 | 510 | F | | ATGCCGAATCCGAAGAAGAATGTG (SEQ ID NO.37) |
| | | | R | ATTGACGATGAGTTGTGCGAGGG (SEQ ID NO.38) |

[0104] (6) SNP位点在群体的多态性验证

[0105] 根据SNP标记所在位点两端特意序列设计引物(表2),在150份马尾松种质资源的基因组DNA开展PCR扩增与序列比对,分析表明,在MG7、MG25、MG26、MG27、MG36等5个基因的SNP位点具有较高的扩增效率与成功率。以上5个基因均含有纯合型与杂合型等位基因(图7)。其中,MG7基因的基因型包括C/G、C/C、G/G;MG25基因的基因型包括G/T、G/G、T/T;MG26基因的基因型包括A/C、A/A、C/C;MG27基因的基因型包括C/T、C/C、T/T;MG36基因的基因型包括G/C、C/C、G/G。各基因的基因型频率分析表明,MG7和MG36两个基因的杂合度较高,其C/G基因型在群体中的占比分别为48%和57%;MG25和MG26两个基因的纯合度较高,其纯合基因型(G/G、T/T)和(A/A、C/C)在群体中的占比和分别为83%和94%。因此,在马尾松重要差异基因开发SNP位点及研发重要性状关联标记具有极高的可行性。

[0106] (7) 产脂力性状功能SNP位点研发

[0107] 基于多组学测序数据在高、中、低产脂力马尾松共筛选得到1014个含有单核苷酸多态性(SNP)/插入缺失(Indel)(简称SNP/Indel)等位变异位点,其中10个为Indel变异位点。利用序列信息设计引物并对其开展PCR扩增与序列比对验证,结果在根据差异表达信息和序列比较信息筛选出来的13个基因中共开发得到19个SNP位点(表3),其等位变异类型包括C/T、T/G、C/G、A/G、A/C、T/A等。其中,Cluster-35640.0的SNP位点最多为4个,其次为Cluster-19799.0、Cluster-24038.0与Cluster-32834.3均含有2个SNP位点,剩余9个基因仅含1个SNP位点。

[0108] 在高、中、低产脂力马尾松各选3个株系,利用PCR与序列比对方法对SNP分子标记基因型进行验证(表3)。结果表明,除了Cluster-19799.0(C/T)、Cluster-31423.0(T/C)两个基因在高产脂马尾松中为杂合基因型,其余基因在高产脂马尾松的基因型均为纯合型;15个SNP位点在低产脂马尾松中为杂合基因型。在高、中、低产脂力马尾松的基因型关联分析显示,本技术所开发的SNP位点均与产脂力表现出较高的相关性。位于Cluster-19799.0(2个)、Cluster-24038.0(2个)、Cluster-32834.3(2个)、Cluster-35640.0(4个)等4个基因的多个变异位点为连锁遗传的SNP标记;其中,Cluster-24038.0、Cluster-29744.0、Cluster-32834.3、Cluster-35640.0在高产脂马尾松中为纯合基因型,在低产脂马尾松中为杂合基因型。因此,本研究得到的与产脂力性状紧密相关的SNP功能标记在分子辅助选择育种领域均具有较高的应用价值。

[0109] 表3与产脂力性状紧密关联的19个功能SNP位点

| 基因 ID | 位置 | 等位变异 | 高 | 高 | 高 | 中 | 中 | 中 | 低 | 低 | 低 |
|---------------------|------|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Cluster-19758.0 | 76 | T/C | T/T | T/T | T/T | T/T | T/T | T/T | T/C | T/C | T/C |
| Cluster-19799.0 | 78 | T/C | C/T | C/T | C/T | C/C | C/C | C/C | C/C | C/C | C/C |
| | 1073 | T/C | C/T | C/T | C/T | C/C | C/C | C/C | C/C | C/C | C/C |
| Cluster-20708.0 | 169 | C/T | C/C | C/C | C/C | C/T | C/T | C/T | T/T | T/T | T/T |
| Cluster-23480.0 | 128 | T/G | T/T | T/T | T/T | T/T | T/T | T/T | T/G | T/G | T/G |
| Cluster-24038.0 | 103 | G/T | G/G | G/G | G/G | G/G | G/G | G/G | G/T | G/T | G/T |
| | 194 | A/G | A/A | A/A | A/A | A/G | A/G | A/G | A/G | A/G | A/G |
| Cluster-29744.0 | 130 | G/T | G/G | G/G | G/G | G/T | G/T | G/T | G/T | G/T | G/T |
| Cluster-31423.0 | 29 | C/T | T/C | T/C | T/C | T/T | T/T | T/T | T/T | T/T | T/T |
| Cluster-32834.3 | 73 | T/A | T/T | T/T | T/T | T/T | T/T | T/T | T/A | T/A | T/A |
| | 486 | G/C | G/G | G/G | G/G | G/G | G/G | G/G | G/C | G/C | G/C |
| Cluster-32938.2 | 109 | A/G | A/A | A/A | A/A | A/A | A/A | A/A | A/G | A/G | A/G |
| Cluster-34813.0 | 143 | C/T | C/C | C/C | C/C | C/C | C/C | C/C | C/T | C/T | C/T |
| Cluster-35640.0 | 125 | G/T | G/G | G/G | G/G | G/G | G/G | G/G | G/T | G/T | G/T |
| | 527 | G/T | G/G | G/G | G/G | G/G | G/G | G/G | G/T | G/T | G/T |
| | 895 | G/C | G/G | G/G | G/G | G/G | G/G | G/G | G/C | G/C | G/C |
| | 1087 | T/G | T/T | T/T | T/T | T/T | T/T | T/T | T/G | T/G | T/G |
| Cluster-39957.3 | 138 | A/G | A/A | A/A | A/A | A/A | A/A | A/A | A/G | A/G | A/G |
| Cluster-40412.10076 | 97 | A/C | A/A | A/A | A/A | A/A | A/A | A/A | A/C | A/C | A/C |

[0111] Cluster-19758.0

[0112] 该SNP位点在Cluster-19758.0所示核苷酸序列,其中自5'端起第76位碱基中的等位基因为T/C,核苷酸如下所示(SEQ ID NO.6):

[0113] GTTTTGGACATATCCAGCAATCATGACAGTCCAGGACACAACACTCCTCTCAGACATTTTGTCAAACA
ATTGTCT (T/C) GCAATTTCTATATATCCACATCTAGCATAATGTCAATAAGAGCACCCCTAACATAGATATCGG
ACTCAAACCAGTTTTGATTACATAATCATGAACTTCTTTGCCCTTTTCCAGGGACTTGAAGTAAGCACATGCTGA
GAG;

[0114] Cluster-19799.0:

[0115] 该SNP位点在Cluster-19799.0所示核苷酸序列,其中自5'端起第78位碱基和第
1073位碱基中的等位基因均为T/C,核苷酸如下所示(SEQ ID NO.7):

[0116] GTTTCTGCTCTTTCTAACAGTTTCCAAGCTTCATCAACATTGCATGCCTTACAGAAGCCACCAATAAG
TGTATTTAAA (T/C) AAAGAAGCATTAGGGGGGATGCCTTTCTGGTACATCTCATTAAAAGGGTATGAGCTTCATA
TACTCTGCCTCCTTACAAAAGACAGTTCATTAGCAGACTATAGCTAAAGGCATTAGGCACACAACCCTTCTTCTGC
ATGTAATTGAGTAATTTATAGGCCCTTATCTATTTTGCCTCCAATACTAAGGCCTCGCATCAGTAAATTGTAGGTTT
CAGTTGTAGGAGCACACTCCCTTGTGGTGCATCTCACTTAAGAGTTTGAAGCCATTGACAATTCTCCATGCTTACA
AAAGCCATCTATTAGAGTACTATAAGAGATAGCATCTGGGTGAAGGCCTTTCTCCAACATCTGAATAAAAAACCTC
TTTGCTTCATCCAATCTGCCCTCTTGCAGAGGCCATCAATGACTGTTGTATATGTGACAACATCAGGAAAGCAGT
TGTCATGTTAACTGAATTTACCAGCCCCAAAGAAGCATCGTCAGTCTCATAGAAATCACCATAGCAGCACTGCC
ATGAAGCCATATGCGATCCATCATTTTAATCGCTTTGTTTACTTTGCCATTTTACAAAGGCCTTCTATCATAATA
TTACAAGTTATCGTATCTGGAGTGAGACCTTTTTCCCTCATTTTGTGTAGAAGATGTTCTGCTTCTGAAATCCTGC
CCTCCTTGCAGACACTTTGGACAAAATGTTGTAAGTAACCGTATTAGCAAAACAGCCTTTGCTCAACATTTACC
AAGCAATTGATTAGCTCGTTCAACATTACCTCTCTTGAATACCCATGAAGAAGCGTGCTGTATGTTATGCTGTCT
GGAACAACCTCATTTTGAAGCATCGCATTACAGGATCCTCTCTGCATCAGATAGCATACCCTCCCTACACATGCCGT

CGACTAATACATTGTAGGAGATCACATCGGGAGTGATGTTAGCTGTCATCATCGCTTCCAGGAGCTTGCGAGCCTC
AGAAATTCTCCCTAGCTTGG (T/C) CAGTCCACTTAAAAGAATGTTGTAGGACACCACATTGCAGGAATGACAATT
ATCCTTCATGTGCTCCATAACCTTGAAAGCTTCGTCAAGTTTGCCTTC;

[0118] Cluster-20708.0:

[0119] 该SNP位点在Cluster-20708.0所示核苷酸序列,其中自5'端起第169位碱基中的
等位基因为C/T,核苷酸如下所示 (SEQ ID NO.8):

[0120] TTAAGGCCTACATAACCTCCGAAGCTCTCTTAAGCGTTCTATGTCGAAATTCCTTTTCTTCATCAGTC
TGCTCTGCCTGGCACTGAAGCGAGTATTGAAGAAGCTCTTGAATGAAGGATTTGTGGTTGTGTTAAGGAAGAATAG
ATACCCCGCCATGAAGTATATAGA (C/T) GCGAGAAAGTAGCATAACGGTTCCATCACATCCCATGACAGCTCCCA
GAATGTCAGCCTCATTAAACCCGCAGTTTGAAGCCCGAAAAATCCCAAGCCCAACCGTAGTTCAGTCATGGCGTCA
CTCTTAGCTTTCTGA;

[0121] Cluster-23480.0:

[0122] 该SNP位点在Cluster-23480.0所示核苷酸序列,其中自5'端起第128位碱基中的
等位基因为T/G,核苷酸如下所示 (SEQ ID NO.9):

[0123] TGCCCATGTAGAGCATAGCCTACAATCATCGCAGTCCACGAAACCACATCTGTCTTCGGTATCATT
AAAGACTTCACATGCATCACCTATGGAACACTATTTAGCATAACATTGTAATAAGAGCAG (T/G) CCCCACAGAGGC
ATCCTCCTGAAATCCATTTTGTGGTCTGGCAATGAAACTGCTTGCTTGATCAATAATTGCCAGGTCCGCACAC
ACG;

[0124] Cluster-24038.0:

[0125] 该SNP位点在Cluster-24038.0所示核苷酸序列,其中自5'端起第103位和第194位
碱基中的等位基因为G/T和A/G,核苷酸如下所示 (SEQ ID NO.10):

[0126] ATGGTGCCTTCTCTAATTCAAATTGATTCATCACAGGTGATGGAGTCAGAAACAGCAGCAGTAACAGG
GTCTCTACTTTTTGGAACAGATAAAATGACTATG (G/T) GACCAGGATCTGAAAATGTTTGGCCCTGTTTTTTCCTC
TTTAACAGGCCTGCAATCAAACCAGATGCTTATTTCTCATTCCAACCAAATGG (A/G) CTTCCACAAAAGTTCTGT
AGTCCCAATCCATGCAGCAGAATTAAGCAGAGAAATATATGATATGGAAAGTACCCAATACAACAATATTCACTT
TT;

[0127] Cluster-29744.0:

[0128] 该SNP位点在Cluster-29744.0所示核苷酸序列,其中自5'端起第130位碱基中的
等位基因为G/T,核苷酸如下所示 (SEQ ID NO.11):

[0129] GGTTGTTCCGCAAATCCAGTTGGATTACTAAACCCTTGCCATTTTAAAGCACCCCTGCTTAAATTCTCG
TCATATTCTTCAAGAGGGGACCATTCCTCTGATTCTTCTTCTATTTCACTCGTCCAATA (G/T) CAAAATCAA
TCTCCTATGCTCACCCCTGCGCGCGTGCACAAATTAATTGCTAAGCAGACAGATGCATTGTTAGCAATGGAAATTT
TTG;

[0130] Cluster-31423.0:

[0131] 该SNP位点在Cluster-31423.0所示核苷酸序列,其中自5'端起第29位碱基中的等
位基因为C/T,核苷酸如下所示 (SEQ ID NO.12):

[0132] TTATATGCGAGACCTTGATTACCATTG (C/T) TTGGTGTGCACGGTAGGAGTAGACGCCCTTCCCA
CTCTTCCGACCAAGGTACCCTGCATCAACATATTGCACAAGCAGTGGGCATGGGGAATATTTGCTTTCTCCAAGCC
CGTGATGGAGAATCTTCATAATTGAGAGGCAAACATCTAATCCAATGAAGTCTGCAAGTTCAGGGGTCCCATGG;

[0133] Cluster-32834.3

[0134] 该SNP位点在Cluster-32834.3所示核苷酸序列,其中自5'端起第73位和第486位碱基中的等位基因为T/A和G/C,核苷酸如下所示(SEQ ID NO.13):

[0135] CAACAACATACCCAAACCAGTCTCAGTTCCCAAGGAGAATCTAACCGACTAACGAGGTCGTCATCCTC
CAGA (T/A) CCAAAGCAAACACAATACCTGCAAGTGCTTTTCATAGATGGGGTGGCCACCAAAGACATCAACAT
TGATCCTAGTACTGGTTCCCTTCGAATTTTTCTGCCAGAACCAGAATTAGCAGTAGCAGCCCTCCCACAGGGC
CAGCAGAGAGATTCAAGGGATGTGGATCTCACTCTCAGAGCTTGGGGTTTGGCCAAGAGAGAAACCATACATGAAG
TGGATTCAAGTTCCAGGAAATTGAACAGGAATGCCAGTACTAATGCAAAAAGGGTTCAGTTACGAATCAAG
TTCAACACACAGGAGGAAAAGTTATGGTGGTATTGTAGAAAACATGAGAAGCAGCTATGGTGGTATTGGGGACAAG
CAGCAGCGAGTTTGTCAAGATTCGTTGCTCAGACAAGGAAT (G/C) GACGAAGACGTGAATCCATCAGAGAAGGAT
AGTTTGTCTCAGGTGTACGGATAAGACATCAGGATCAGGCCAAATAAATGAT;

[0136] Cluster-32938.2

[0137] 该SNP位点在Cluster-32938.2所示核苷酸序列,其中自5'端起第109位碱基中的等位基因为A/G,核苷酸如下所示(SEQ ID NO.14):

[0138] AGACAAAACACCAATAAAGGTGACATGGTTTGGTTTTCACATGGGAGTTCCTCATATGTTCAAACAGCT
TGAGAGCCTTCTTTCCTTGACCATGTATGGCATATCCTAC (A/G) ATCATGGCAGTCCATGAAACAACATTTCCGC
TAGGCATTTTGTCAAATACTCCGCATGCATCCTCTAGGCACCCACATTTGGCATAACATATCTATAAGGGCATTCCC
CAC;

[0139] Cluster-34813.0

[0140] 该SNP位点在Cluster-34813.0所示核苷酸序列,其中自5'端起第143位碱基中的等位基因为C/T,核苷酸如下所示(SEQ ID NO.15):

[0141] GGTATCTAACTCTGTATTGTCAACAATTTTTGCCTGCTGTTTGGATGTCAAAAAGACCTGCATCAGAAG
ATTCATTGCAAGTTTGTGCAGCTGTGAAATCCTCATTCTGACCACCATCTCTGAATCCCTTTTGAATGATAT (C/
T) CATGCTTTCCAGTATGGGATACGCCCTCCAGTCTGTAAATTTTTGGTTCCAGAACAACAGTTGTTTCCCCTTTA
TCTG;

[0142] Cluster-35640.0

[0143] 该SNP位点在Cluster-35640.0所示核苷酸序列,其中自5'端起第125位、第527位、第895位、第1087位碱基中的等位基因为G/T、G/T、G/C、T/G,核苷酸如下所示(SEQ ID NO.16):

[0144] GGAATGTGGTGCATGGAATGCAATAATTGCGGCATCTATACAGCATGGTTGGGTTAAGCAGGCATTG
GAATTATTTGCTGAAATGCAATTAGCAGGCATAAAGGCGAGCCACGTTACTTTGG (G/T) CTCGTTCTCAACGCG
TTTGCGATTCTAGGAAACCTCGAACAGGGTAAGCAACTCCATGCCTGTGTGTTAGAAATGGATTTGGATCGGATT
TGCTTGTGGGTAGCTCTGTTATTCTAATGTATATAAAGTGTGGAACATAGATGGTGCCCGCAAGTGTGTTGACAA
AATGCCTATGGTAGACCTGGGGTTATTCAATGCTACAATTCAAGGATATAGCAGCGAGGGCCATAATAATGAGGCA
ATGGAACATTTGGTCAACTACTACGAACGGGTTTGAACCAAATGATATTACCTTACCATTGTTCTTAGAGCCT
GTGCCAGCTTTGAAACTGCCCTGAACAGGGCAGGCAACTCCATGTACAGATAATGAAATCTGTGTTTTCAGCACAA
TGTTTC (G/T) GTGATTAGTTCTCTCATCACTGTGTATGCTAAAAGTGGTTGCATAGCTGATGCACAAAAAGTGT
TGACAAGATGACTGTACAAAACAATGTAGTTTTATGGACTGCATTGATAGCTGGTTACACCCAGAAAGGGTATACA
GACGAAGCCTTGAAACTCTTCTCTTAATGCAACGTGCAGGTGTTAAACCTAATCAATCACCTACCCAACCTGTTT

TCCGTGCTTGTGCAAACCTTAGCTTCTATAGAAGAAGGAAAACAAGTCCATTGTCATATTATCAAAGCTGGGGCTGC
GTCGGATACTTTTCGTTGCCAGCGCCATTGTTGACATGTATATTAAGTGCGGTAGTCTAGAGGATGGCCAACGAG (G/
C) GTTTGATAGAATTCCTAGACGAGACATTGTATCGTGGAATACAATGATTGCAGGATATGCTCAACATGGGTATG
TTGACAAGGCACTTTAATCTTTGAAGAAATGCAACAGTATGGCATGAAACCCAACCACGTAACCTTTTGTTCAGT
TCTCTCTGCATGCAGCCATGGAGGACTTGTTCGTTTAGGGC (T/G) CCGATACTTTAGTTCCTTGAGCCGAACCTCA
TGGCATTATTCCAAGAATGGAGCATTATGCTTGCATAGTTGACCTCCTTGGCCGTGTTGGGCACTTGTATGAGGCA
GAAGACTTTATCAACAATATGCCTTTTCGAGCCA;

[0145] Cluster-39957.3

[0146] 该SNP位点在Cluster-39957.3所示核苷酸序列,其中自5'端起第138位碱基中的等位基因为A/G,核苷酸如下所示(SEQ ID NO.17):

[0147] TTGCAGCGCATGCCTTAAGAACACAGGGAAATGTGAAATTATCCGGTCGCAAACCCATCCCTTGCATC
TGATCATATAATGCAAGCGCCTCTCTGCAGCCACCCTGCCTAGCATACGCTCCAATCATGGCATTCCAC (A/G) TA
AACTTATTCCTTTGAGCCATTTGTCAAACACTTCGCGTGCATATTCCAAGTTCCACACTTTCATACATGGTGA
TG;

[0148] Cluster-40412.10076

[0149] 该SNP位点在Cluster-40412.10076所示核苷酸序列,其中自5'端起第97位碱基中的等位基因为A/C,核苷酸如下所示(SEQ ID NO.18):

[0150] AAATTGGTTTTTTGTTTTAGAAAAGACATAATAGAAGAACTGAGAATCATTGAAATACATAAAGAGCA
GTTTTGTGGAAGACTGATGTCTGCTCGT (A/C) TAGTTACTCTAATAACACAGATCATGGTAATGGCGGTGGTAT
ATCAGTTATTTGCTTTTCCTTGAATTGGAATTCTGGTTCAATTCATCTGTGTATTGTGGGTAGAGCACTCAG
GA。

[0151] 上述实施例为本发明较佳的实施方式,但本发明的实施方式并不受上述实施例的限制,其他的任何未背离本发明的精神实质与原理下所作的改变、修饰、替代、组合、简化,均应为等效的置换方式,都包含在本发明的保护范围之内。

序列表

| | | |
|-------|--|-----|
| <110> | 广东省林业科学研究院 | |
| <120> | 一种马尾松高产脂功能SNP标记的筛选方法及其应用 | |
| <160> | 38 | |
| <170> | SIPOSequenceListing 1.0 | |
| <210> | 1 | |
| <211> | 948 | |
| <212> | DNA | |
| <213> | 人工序列(Artificial Sequence) | |
| <400> | 1 | |
| | atgcagaaga ttgtggatca aactgatgca actataagaa aagcccgagt ttcagtccga | 60 |
| | gctagaaccg agtcacccat gataagtgat gtttgccaat ggaggaaata tggacaaaag | 120 |
| | atggccaagg gcaatccatg ccccagagct tactacagat gcacatgtc accgtgctgt | 180 |
| | ccagtgcgca agcaggtgca acgcctggcc gaagacagat cgattctgat aacaacatac | 240 |
| | gagggcagcc ataaccacat gtttctcca gcagccacag ccatggcatc cactacagca | 300 |
| | gcagctgcct caatgcttct gtccggttct tccacatcag ctgataacat cgcactaaat | 360 |
| | gcaagcttca tggcaggtgc cctcatgcag catccttgca acacatcaac tgctagcatc | 420 |
| | tcggcctcag ctccattccc cactatcaca ctggatctca cacacaatcc caaccaaag | 480 |
| | gccaacgctc caggtcacat ggccagcctca aaccaagag cactagcagg actcccagct | 540 |
| | catgccatgc cttttagtag catgccacac caattccca ccaacacacc ccagggggca | 600 |
| | ttattccatg ggcaatcaat ttacaatgct ctttccatgt ttgcaccctt agcaggccag | 660 |
| | ctgcaacgcc cccaacaacc catgatgccc acaccacaa agctcaatat tcaagctggc | 720 |
| | caaccaccaa caccaccaca acaacaacca tcattcatag acacagtcag tgcagctaca | 780 |
| | gctgccataa cttctgatcc caatttcaaca gcagcccttg cagcagccat cacatccctt | 840 |
| | atgaacaaca acaacaatgc tgccaatgcc atttctaaac ccaactctc tcaggcactc | 900 |
| | aatcctagtc cccacgttgc agctcatgtc aagacagatc acaccaa | 948 |
| <210> | 2 | |
| <211> | 723 | |
| <212> | DNA | |
| <213> | 人工序列(Artificial Sequence) | |
| <400> | 2 | |
| | atggaaaacc tcccacatca gcaacctgac ctgaaattg ctcaaacaca cgaggatccc | 60 |
| | gggtgccgcc gatttaaggg aattcgactg cgaaaatggg gaaggtgggt atcggaatc | 120 |
| | cggatgccaa aatctcgaga gaaaatattg ctgggctctt atacgactcc cgagcaggct | 180 |
| | gcccgtgctt acgacgccgc agtgtattgt ctgagagggc ccaacgcaa atttaacttt | 240 |
| | ccggaatccg tgcacgacat tccgtctgtg acttctgttt cccgtcagga aattcagcac | 300 |
| | gccgccttca aatatgcctt gggccagccc cctccgagtt tgcagtctct ggaagggcac | 360 |
| | gccgccttca aatatgcctt gggccagccc cctccgagtt tgcagtctct ggaagaaggg | 420 |

| | |
|--|-----|
| catgccgccc tcaaatatgc cttgggccag cctcctccga gtttgcagtc tctggaaggg | 480 |
| cacgcgtcgc cgtcacagtc gtctacggtt tcggaaacgg agttatcggg agaacagctg | 540 |
| aagatatcgg aagagtgcc cgacttagca ctgtgttggc cgctgtttgc ggcagacgac | 600 |
| actgggattc ccaattcgga aaaagtcccc tcgattgacg aatatttcag tgcgactttg | 660 |
| caggagcagc gggaggaggg ttacattttc acagatttgt ggaatttcca agatcaagat | 720 |
| gtt | 723 |
| <210> 3 | |
| <211> 657 | |
| <212> DNA | |
| <213> 人工序列(Artificial Sequence) | |
| <400> 3 | |
| atggagaaat cctcccaaca ggaggatgac catgcccata ctccagaaga agaagttcgc | 60 |
| gggcaaaaagt gccgtcaatt taaggaacc cgattgcgaa agtgggggag atgggtagca | 120 |
| gaaattcgaa tgcccaaatc tcgagagaag ttatggctgg ggtcatacaa aaagcccagag | 180 |
| caggccgccc gcgcctacga cgccgcagtg tattgtctga gagggccgaa cgccaaattc | 240 |
| aatttaccba attctctacc tgacattccg tctgcgtctt ctctttcccg ccggcagatt | 300 |
| caactcgctg ctgccaaatg tgcgttggat caattccctt cgagtgcgcc ccctctgcag | 360 |
| aattttaata ataaggccat ggacgaggcc gcatcgccgt caagactgga tccggtatca | 420 |
| gaaactgagt tgtcgagcga tggatcatcaa atatcagagg aaggggagtt ggatttgtgg | 480 |
| gaaagtccgt ttgaggtatc aggcggcaat tatgaaggc gcatgaacct gaatttagag | 540 |
| agaatgccat cgattgagga gttctcggcc ttggaaatta tttacagtat ttgtcagcag | 600 |
| catgaggagg aggaacacat aaaccttttc ctcgacceca cagagttgtg gaacttt | 657 |
| <210> 4 | |
| <211> 795 | |
| <212> DNA | |
| <213> 人工序列(Artificial Sequence) | |
| <400> 4 | |
| atggataatg gatcctctct cgtgcccata gccatgccc attccttgac agacattgaa | 60 |
| gcaataagca attctccttt tgcggataag ggtggaaca agagaattcg aacgcaagac | 120 |
| gaagctgcct ctctgccttc acagcaaagc aggtgaacc tgcaaaatcc agcatacaga | 180 |
| ggcgtgcgcc gtcgtagctg gggcaaatgg gtgtctgaaa ttcgggaacc aaaaaagaaa | 240 |
| aaccgaatct ggctcggtc ctacgataca ccagagatgg ccgctcgagc tcacgatgct | 300 |
| gctgcattct acttgaaagg aaagaaacat tcgttgetca attttcaga gctcattgat | 360 |
| caacttcag aaccaatttc ttcggetccg cccacattc aagctgccg ggcagcagca | 420 |
| gccgtcgctt tcaattctgc atcccgtagt gcacagaact ctggaatgct aagcgatctc | 480 |
| aataaacaag acagcggaag gcccaagaaat actccagtaa gcaatgagca cgcaggtatc | 540 |
| tcctcttcga atcaactatc ggcagtaatt agctcagagt taaacttga gacggctaat | 600 |
| gtcgagttat tagggaaacc gaattatact aattcatcat caatggagat cgtaacagag | 660 |
| gaggacttgt tcgaatccac caacttctat acgaatttgg cagaaggcct tatgcttct | 720 |

| | |
|--|-----|
| ccacctctgt tcagtattcc tgaattagac atagaggagc agagattgga ggaaggattt | 780 |
| ctttgggtctg gtttt | 795 |
| <210> 5 | |
| <211> 840 | |
| <212> DNA | |
| <213> 人工序列(Artificial Sequence) | |
| <400> 5 | |
| atggctggca tggatgacgg agacattaat tttagcagca atatagtaga tgattttggc | 60 |
| aatgggtcct ccatggaaag ctttttcgag gagattttga gggatactac tcatgcctgc | 120 |
| actcatacac acacctgcaa cctcccggg ccagataaca cacatacaca cacatgcttc | 180 |
| cacacgcaca caaaaaatcct tgetgctccc gatgacgaga agtctgcaga cactgctgag | 240 |
| tctccacaaa acagctcttc caagccaaaa aaacgaccag taggtaatcg agaggcagtt | 300 |
| agaaaataca gggaaaaaaaa aaaggcccgg acggcctccc tggaggagca ggttgttcaa | 360 |
| ctgaccactg ttaatcagca attgcatagg agattacagg gtcaagcagc tttagaggct | 420 |
| gagattgcaa gattgaagtg tctgctggct gactttaggg gccggatcga tggggaattg | 480 |
| gggtcctatc cttaccaaaa gtcaattaga atggataaaa cttgcaatga tgcaccattc | 540 |
| cggcaaccga tgctggggg atatgtcctg gatccctgca atatctggtg caatgcagat | 600 |
| gcagcttgcc gtgaaccgac tctggcatcc aacagtgagg gtggtgtgca gcacgaacgt | 660 |
| gatagtgctg cacgttgga tggcgattgt ggccagatcg caggccattg tcaaggtttg | 720 |
| aagggtgata tggcagtaac ctctagtga ctttctggat gctcagaggg tacagcaaca | 780 |
| aaaacagtgc ctgctgccat ggcttcttct ggaaaagaga gaaaaggtgc atttggtgtg | 840 |
| <210> 6 | |
| <211> 219 | |
| <212> DNA | |
| <213> 人工序列(Artificial Sequence) | |
| <221> misc_feature | |
| <222> (76) .. (76) | |
| <223> n i s c o r t | |
| <400> 6 | |
| gttttggaca tatccagcaa tcatgacagt ccaggacaca aactcctct cagacatttt | 60 |
| gtcaaaacaat tgtctngcaa tttctatata tccacatcta gcatacatgt caataagagc | 120 |
| accctaaca tagatatecg actcaaaacc agttttgatt acataatcat gaacttcttt | 180 |
| gcccttttcc agggacttga agtaagcaca tgctgagag | 219 |
| <210> 7 | |
| <211> 1172 | |
| <212> DNA | |
| <213> 人工序列(Artificial Sequence) | |
| <221> misc_feature | |
| <222> (78) .. (78) | |

| | | |
|-------|--|------|
| <223> | n i s c o r t | |
| <221> | misc_feature | |
| <222> | (1073)..(1073) | |
| <223> | n i s c o r t | |
| <400> | 7 | |
| | gtttctgctc tttctaacag tttccaagct tcatcaacat tgcattgcctt acagaagcca | 60 |
| | ccaataagtg tattaaanaa agaagcatta ggggggatgc ctttctggta catctcattt | 120 |
| | aaaagggtat gagcttcata tactctgect cttcacaaa gacagttcat tagcagacta | 180 |
| | tagctaaagg cattaggcac acaacccttc ttctgcatgt aattgagtaa tttataggcc | 240 |
| | ttatctattt tgccccaat actaaggcct cgcacagta aattgtaggt tccagttgta | 300 |
| | ggagcacact cccttgttg catctcactt aagagtttga aagccattga caattctcca | 360 |
| | tgcttacaaa agccatctat tagagtacta taagagatag catctgggtg aaggcctttc | 420 |
| | tccaacatct gaataaaaaa cctcttttget tcatccaatc tgccctcttt gcagaggcca | 480 |
| | tcaatgactg ttgtatatgt gacaacatca ggaaagcagt tgcattggtt aactgaattt | 540 |
| | accagcccca aagaagcatt gtcagttctc tagaaatcac caatagcagc actgccatga | 600 |
| | agccatatgc gatccatcat tttaatcgtt ttgtttactt tgccattttt acaaaggcct | 660 |
| | tctatcataa tattacaagt tatcgtatct ggagtggagc ctttttcctt cattttgtgt | 720 |
| | agaagatggt ctgcttctga aatcctgccc tccttgcaga cactttggac caaatgttg | 780 |
| | taagtaaccg tattagcaaa acagcctttg ctcaacattt caccaagcaa ttgattagct | 840 |
| | cgttcaacat tacctctctt gcaataccca tgaagaagcg tgctgtatgt tatgctgtct | 900 |
| | ggaacaactc catttcgaag catcgcattc aggatcctct ctgcatcaga tagcataccc | 960 |
| | tcctacaca tgccgtcgac taatacattg taggagatca catcgggagt gatgttagct | 1020 |
| | gtcatcatcg cttccaggag cttgcgagcc tcagaaatc tccttagctt ggncagtcca | 1080 |
| | cttaaaagaa tgtttagga caccacattg caggaatgac aattatcctt catgtgctcc | 1140 |
| | ataaccttga aagcttcgct aagtttgcct tc | 1172 |
| <210> | 8 | |
| <211> | 307 | |
| <212> | DNA | |
| <213> | 人工序列(Artificial Sequence) | |
| <221> | misc_feature | |
| <222> | (169)..(169) | |
| <223> | n i s c o r t | |
| <400> | 8 | |
| | ttaaggccta cataacctcc gaagetctct taagcgttct atgtcgaaat tccttttctt | 60 |
| | catcagtctg ctctgcctgg cactgaagcg agtattgaag aagctcttga atgaaggatt | 120 |
| | tgtggttggt ttaaggaaga atagataccc cgccatgaag tatatagang cgagaaagta | 180 |
| | gcatacgggt tccatcacat cccatgacag ctcccagaat gtcagcctca ttaaaccgcg | 240 |
| | agtttgaagc ccgaaaaatc ccaagcccaa ccgtagttca gtcattggcgt cactcttagc | 300 |
| | tttctga | 307 |

| | | |
|-------|--|-----|
| <210> | 9 | |
| <211> | 219 | |
| <212> | DNA | |
| <213> | 人工序列(Artificial Sequence) | |
| <221> | misc_feature | |
| <222> | (128) .. (128) | |
| <223> | n is g or t | |
| <400> | 9 | |
| | tgcccatgta gagcatagcc tacaatcadc gcagtcacg aaaccacatc tgtcttcggt | 60 |
| | atcatttttaa agacttcaca tgcatacct atggaactac atttagcata cattgtaata | 120 |
| | agagcagncc ccacagagge atcctctga aatccatttt tgatggtctg gcaatgaaac | 180 |
| | tgcttgcctt gatcaataat tgccaggtcc gcacacacg | 219 |
| <210> | 10 | |
| <211> | 289 | |
| <212> | DNA | |
| <213> | 人工序列(Artificial Sequence) | |
| <221> | misc_feature | |
| <222> | (103) .. (103) | |
| <223> | n is g or t | |
| <221> | misc_feature | |
| <222> | (194) .. (194) | |
| <223> | n is a or g | |
| <400> | 10 | |
| | atggtgcctt ctctaattca aattgattca tcacaggtga tggagtcaga aacagcagca | 60 |
| | gtaacaggggt ctctactttt tggaacagat aaaatgacta tngaccagg atctgaaaat | 120 |
| | gtttgccctg ttttttcctc ttaacaggc ctgcaatcaa accagatgct tattttctcat | 180 |
| | tccaacccaaa tggncctcca caaaagttct gtagtcccaa tccatgcagc agaattaagc | 240 |
| | agagaaatat atgatatgga aagtacccaa tacaacaata ttcactttt | 289 |
| <210> | 11 | |
| <211> | 219 | |
| <212> | DNA | |
| <213> | 人工序列(Artificial Sequence) | |
| <221> | misc_feature | |
| <222> | (130) .. (130) | |
| <223> | n is g or t | |
| <400> | 11 | |
| | ggttgttccg caaatccagt tggattacta aacccttgcc atttttaagc accctgctta | 60 |
| | aattctcgtc atattcttca agaggggacc attcctctga ttctttcttct atttcacact | 120 |
| | cgtccaatan ccaaaatcaa tctcctatgc tcaccctgc gcgcgtgcac aaattaattg | 180 |

| | |
|---|-----|
| ctaagcagac agatgcattg ttagcaatgg aaatTTTTg | 219 |
| <210> 12 | |
| <211> 215 | |
| <212> DNA | |
| <213> 人工序列(Artificial Sequence) | |
| <221> misc_feature | |
| <222> (29) .. (29) | |
| <223> n i s c o r t | |
| <400> 12 | |
| ttatatgcga gaccttggat taccattgnt tggatgca cggtaggagt agacgccctt | 60 |
| cccactcttc cgaccaaggt accctgcac aacatattgc acaagcagtg ggcattggga | 120 |
| atatTTgctt tctccaagcc cgtgatggag aatcttata attgagagc aaacatctaa | 180 |
| tccaatgaag tctgcaagtt ccaggggtcc catgg | 215 |
| <210> 13 | |
| <211> 567 | |
| <212> DNA | |
| <213> 人工序列(Artificial Sequence) | |
| <221> misc_feature | |
| <222> (73) .. (73) | |
| <223> n i s a o r t | |
| <221> misc_feature | |
| <222> (486) .. (486) | |
| <223> n i s c o r g | |
| <400> 13 | |
| caacaacata cccaaccag tctcagttcc caaggagaat ctaaccgact aacgaggtcg | 60 |
| tcatcctcca ganccaaagc aaacacaata cctgcaagtg ctgctttcat agatggggtg | 120 |
| gccaccaaag acatcaacat tgatcctagt aactgggtt cccttgaat ttttctgcca | 180 |
| gaaccagaat tagcagtagc agccctcca caggccagc agagagattc aaggatgtg | 240 |
| gatctcactc tcagagcttg gggtttggcc aagagagaaa ccatacatga agtggattca | 300 |
| gctcaaagtt ccaggaaatt gaacaggaat gccagtacta atgcaaaagg ggtcagttac | 360 |
| gaatcaagtt caacacacag gaggaaaagt tatggtggtg ttgtagaaaa catgagaagc | 420 |
| agctatggtg gtattgggga caagcagcag cgagtttgtc aggattcgtt gctcagacaa | 480 |
| ggaatngacg aagacgtgaa tccatcagag aaggatagtt tgctcaggtg tacggataag | 540 |
| acatcaggat caggccaaat aatgat | 567 |
| <210> 14 | |
| <211> 219 | |
| <212> DNA | |
| <213> 人工序列(Artificial Sequence) | |
| <221> misc_feature | |

| | | |
|-------|---|-----|
| <222> | (109) .. (109) | |
| <223> | n is a or g | |
| <400> | 14 | |
| | agacaaaaca ccaataaagg tgacatgggt tggtttcaca tgggagttcc tcatatgttc | 60 |
| | aaacagcttg agagccttct ttccttgacc atgtatggca taccctacna tcatggcagt | 120 |
| | ccatgaaaca acatttcgcc taggcatttt gtcaaatact ccgcatgcat cctctaggca | 180 |
| | cccacatttg gcatacatat ctataagggc attccccac | 219 |
| <210> | 15 | |
| <211> | 221 | |
| <212> | DNA | |
| <213> | 人工序列(Artificial Sequence) | |
| <221> | misc_feature | |
| <222> | (143) .. (143) | |
| <223> | n is c or t | |
| <400> | 15 | |
| | ggtatctaac tctgtattgt caacaatttt tgctgctgt ttggatgtca aaagacctgc | 60 |
| | atcagaagat tcattgcaag tttgtgcagc tgtgaaatcc tcatttctga ccaccatctc | 120 |
| | tgaatccctt ttgcaatgat atncatgctt tccagtatgg gatagccct ccagtctgtt | 180 |
| | aaattttggt tccagaacaa cagttgtttc ccctttatct g | 221 |
| <210> | 16 | |
| <211> | 1226 | |
| <212> | DNA | |
| <213> | 人工序列(Artificial Sequence) | |
| <221> | misc_feature | |
| <222> | (125) .. (125) | |
| <223> | n is g or t | |
| <221> | misc_feature | |
| <222> | (527) .. (527) | |
| <223> | n is g or t | |
| <221> | misc_feature | |
| <222> | (895) .. (895) | |
| <223> | n is c or g | |
| <221> | misc_feature | |
| <222> | (1087) .. (1087) | |
| <223> | n is g or t | |
| <400> | 16 | |
| | ggaatgtggt gtcatggaat gcaataattg cggcatctat acagcatggt tgggttaagc | 60 |
| | aggcattgga attatttgcg gaaatgcaat tagcaggcat aaaggcgagc cacgttactt | 120 |
| | ttgnctcgt tctcaacgcg tttgcgattc taggaaacct cgaacagggt aagcaactcc | 180 |

| | |
|--|------|
| atgcctgtgt gtttagaaat ggatttggat cggatttgc tgtgggtagc tctgttattc | 240 |
| taatgtatat aaagtgtgga aacatagatg gtgcccgcc agtgtttgac aaaatgccta | 300 |
| tggtagacct ggggttattc aatgctacaa ttcaaggata tagcagcgag ggccataata | 360 |
| atgaggcaat ggaactattht ggtcaactac tacgaacggg tttgaaacca aatgatatta | 420 |
| ccttcacat tgttcttaga gcctgtgcca gctttgaaac tgcccttgaa cagggcaggc | 480 |
| aactccatgt acagataatg aaatctgtgt ttcagcacia tgtttcngtg attagttctc | 540 |
| tcatcactgt gtatgctaaa agtggttga tagctgatgc acaaaaagtg tttgacaaga | 600 |
| tgactgtaca aaacaatgta gttttatgga ctgcattgat agctggttac acccagaaag | 660 |
| ggtatacaga cgaagccttg aaactctttc tcttaatgca acgtgcaggt gttaaacta | 720 |
| atcaattcac ctacccaact gttctccgtg cttgtgcaaa cttagcttct atagaagaag | 780 |
| gaaaacaagt ccattgtcat attatcaaag ctggggctgc gtcggatact ttcgttgcca | 840 |
| gcgccattgt tgacatgtat attaagtgcg gtagtctaga ggatggccaa cgagngtttg | 900 |
| atagaattcc tagacgagac attgtatcgt ggaatacaat gattgcagga tatgctcaac | 960 |
| atgggtatgt tgacaaggca cttttaatct ttgaagaaat gcaacagtat ggcatgaaac | 1020 |
| ccaaccacgt aacttttgtt gcagttctct ctgcatgcag ccatggagga cttgttcgtt | 1080 |
| tagggcnccg atactttagt tccttgagcc gaactcatgg cattattcca agaatggagc | 1140 |
| attatgcttg catagttgac ctcttggcc gtgttgggca cttgtatgag gcagaagact | 1200 |
| ttatcaacia tatgcctttc gagcca | 1226 |
| <210> 17 | |
| <211> 218 | |
| <212> DNA | |
| <213> 人工序列(Artificial Sequence) | |
| <221> misc_feature | |
| <222> (138)..(138) | |
| <223> n is a or g | |
| <400> 17 | |
| ttgcagcgca tgccttaaga acacaggga atgtgaaatt atccggtcgc aaacccatcc | 60 |
| cttgcactctg atcatataat gcaagcgct ctctgcagcc accctgccta gcatacgctc | 120 |
| caatcatggc attccaanta aacttatcc tttgagccat tttgtcaaac acttcgcgtg | 180 |
| catattccaa gtttcacac tttgcataca tggatgat | 218 |
| <210> 18 | |
| <211> 217 | |
| <212> DNA | |
| <213> 人工序列(Artificial Sequence) | |
| <221> misc_feature | |
| <222> (97)..(97) | |
| <223> n is a or c | |
| <400> 18 | |
| aaattggttt tttgttttag aaaagacata atagaagaac tgagaatcat tgaatacat | 60 |

| | |
|--|-----|
| aaagagcagt tttgtggaag actgatgtct gctcgtntag ttactctaataacacagatc | 120 |
| atggtaatgg cgggtggtgat atcagttatt ttgcttttcc ttggaattgg aattctggtt | 180 |
| tcaattcatc tgtgtattgt gggtagagca ctcagga | 217 |
| <210> 19 | |
| <211> 20 | |
| <212> DNA | |
| <213> 人工序列(Artificial Sequence) | |
| <400> 19 | |
| ccacaccacc aaagctcaat | 20 |
| <210> 20 | |
| <211> 20 | |
| <212> DNA | |
| <213> 人工序列(Artificial Sequence) | |
| <400> 20 | |
| taagggatgt gatggctgct | 20 |
| <210> 21 | |
| <211> 20 | |
| <212> DNA | |
| <213> 人工序列(Artificial Sequence) | |
| <400> 21 | |
| catgccgccc tcaaatatgc | 20 |
| <210> 22 | |
| <211> 20 | |
| <212> DNA | |
| <213> 人工序列(Artificial Sequence) | |
| <400> 22 | |
| aaacagcgac caacacagtg | 20 |
| <210> 23 | |
| <211> 20 | |
| <212> DNA | |
| <213> 人工序列(Artificial Sequence) | |
| <400> 23 | |
| tcagaggaag gggagttgga | 20 |
| <210> 24 | |
| <211> 20 | |
| <212> DNA | |
| <213> 人工序列(Artificial Sequence) | |
| <400> 24 | |
| tggggtcgag gaaaaggttt | 20 |

| | |
|----------------------------------|----|
| <210> 25 | |
| <211> 21 | |
| <212> DNA | |
| <213> 人工序列 (Artificial Sequence) | |
| <400> 25 | |
| gagttaaact tggagacggc t | 21 |
| <210> 26 | |
| <211> 21 | |
| <212> DNA | |
| <213> 人工序列 (Artificial Sequence) | |
| <400> 26 | |
| agaggtggag gaagcataag g | 21 |
| <210> 27 | |
| <211> 19 | |
| <212> DNA | |
| <213> 人工序列 (Artificial Sequence) | |
| <400> 27 | |
| gaaccgactc tggcatcca | 19 |
| <210> 28 | |
| <211> 20 | |
| <212> DNA | |
| <213> 人工序列 (Artificial Sequence) | |
| <400> 28 | |
| ctgtaccctc tgagcatcca | 20 |
| <210> 29 | |
| <211> 25 | |
| <212> DNA | |
| <213> 人工序列 (Artificial Sequence) | |
| <400> 29 | |
| atgcagaaga ttgtggatca aactg | 25 |
| <210> 30 | |
| <211> 24 | |
| <212> DNA | |
| <213> 人工序列 (Artificial Sequence) | |
| <400> 30 | |
| ttgggtgtga tctgtcttga catg | 24 |
| <210> 31 | |
| <211> 22 | |
| <212> DNA | |

| | |
|----------------------------------|----|
| <213> 人工序列 (Artificial Sequence) | |
| <400> 31 | |
| atggaaaacc tcccgaatca gc | 22 |
| <210> 32 | |
| <211> 30 | |
| <212> DNA | |
| <213> 人工序列 (Artificial Sequence) | |
| <400> 32 | |
| aacatcttga tcttggaat tccacaaatc | 30 |
| <210> 33 | |
| <211> 24 | |
| <212> DNA | |
| <213> 人工序列 (Artificial Sequence) | |
| <400> 33 | |
| atggagaaat cctcccaaca ggag | 24 |
| <210> 34 | |
| <211> 21 | |
| <212> DNA | |
| <213> 人工序列 (Artificial Sequence) | |
| <400> 34 | |
| aaagttccac aactctgtgg g | 21 |
| <210> 35 | |
| <211> 24 | |
| <212> DNA | |
| <213> 人工序列 (Artificial Sequence) | |
| <400> 35 | |
| atggataatg gatcctctct cgtg | 24 |
| <210> 36 | |
| <211> 27 | |
| <212> DNA | |
| <213> 人工序列 (Artificial Sequence) | |
| <400> 36 | |
| aaaaccagac caaagaaatc cttctc | 27 |
| <210> 37 | |
| <211> 24 | |
| <212> DNA | |
| <213> 人工序列 (Artificial Sequence) | |
| <400> 37 | |
| atgccgaatc cgaagaagaa tgtg | 24 |

<210> 38

<211> 22

<212> DNA

<213> 人工序列(Artificial Sequence)

<400> 38

attgacgatg agttgtcgag gg

22

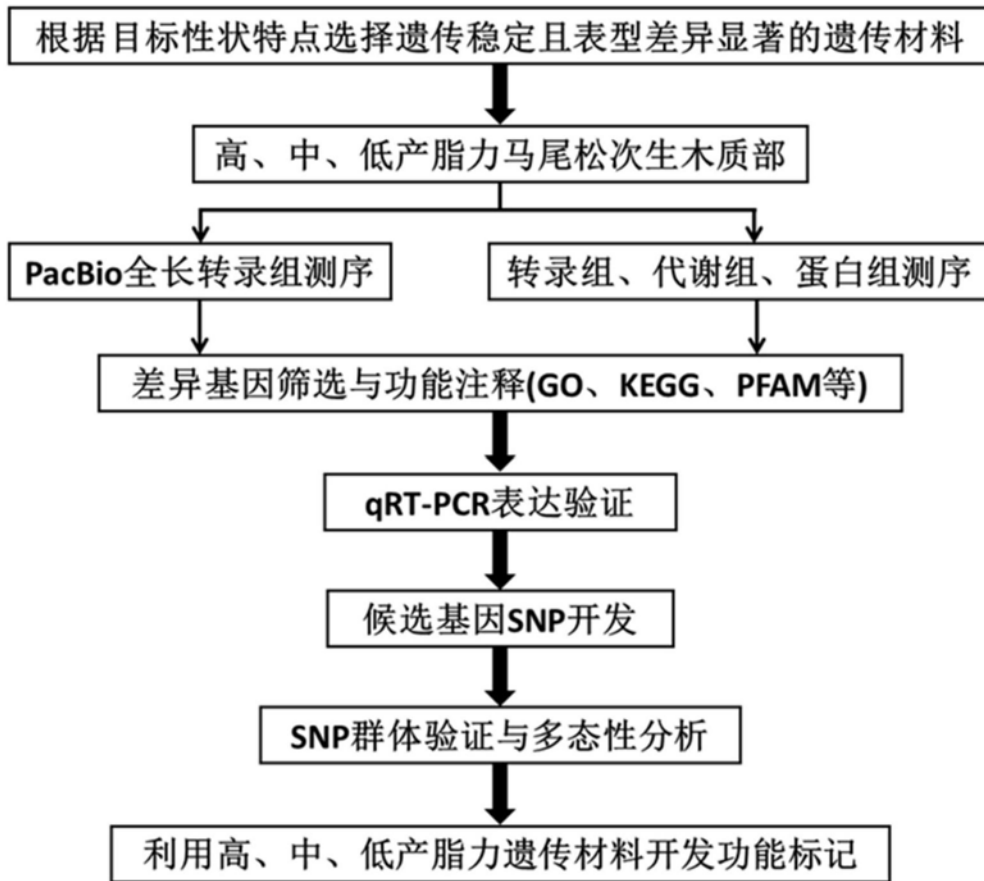


图1

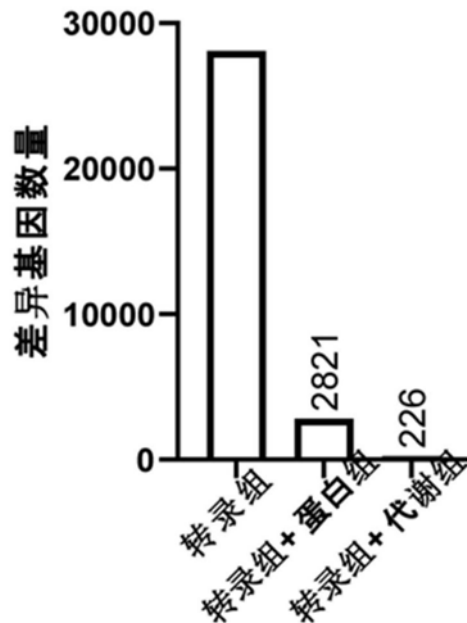


图2

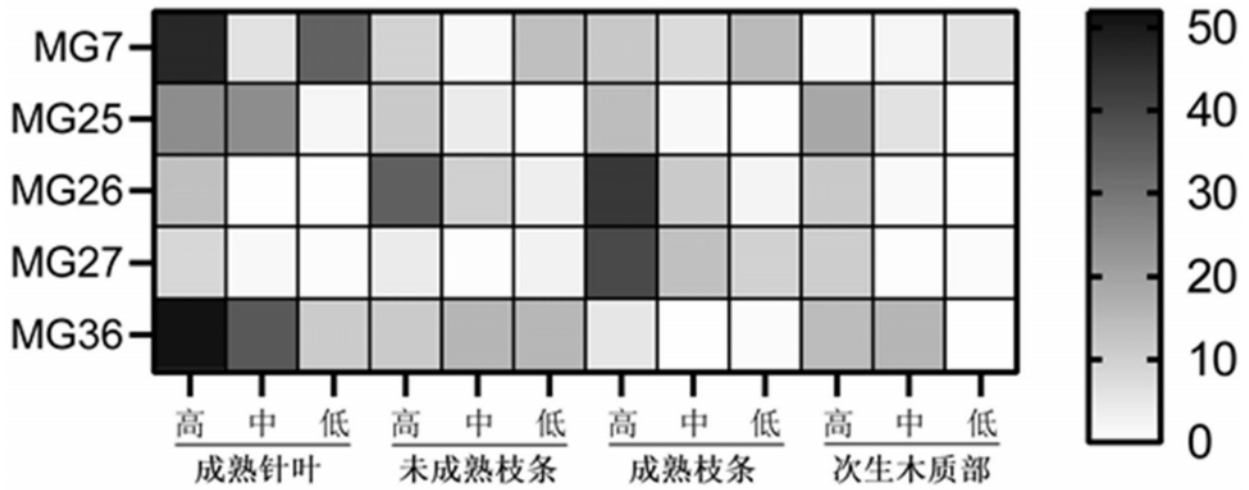


图3

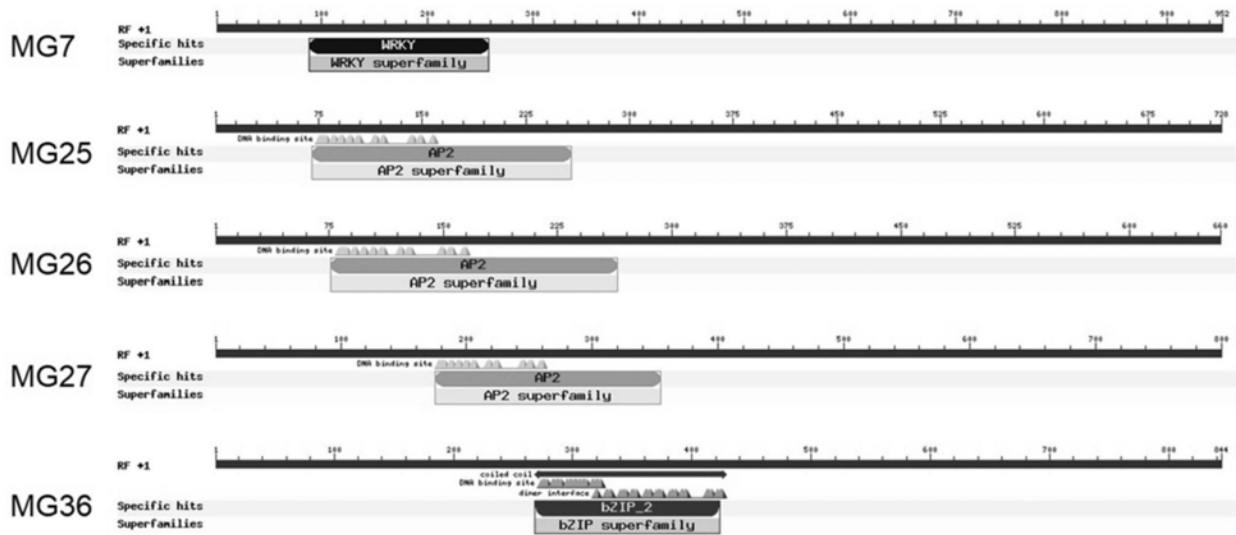


图4

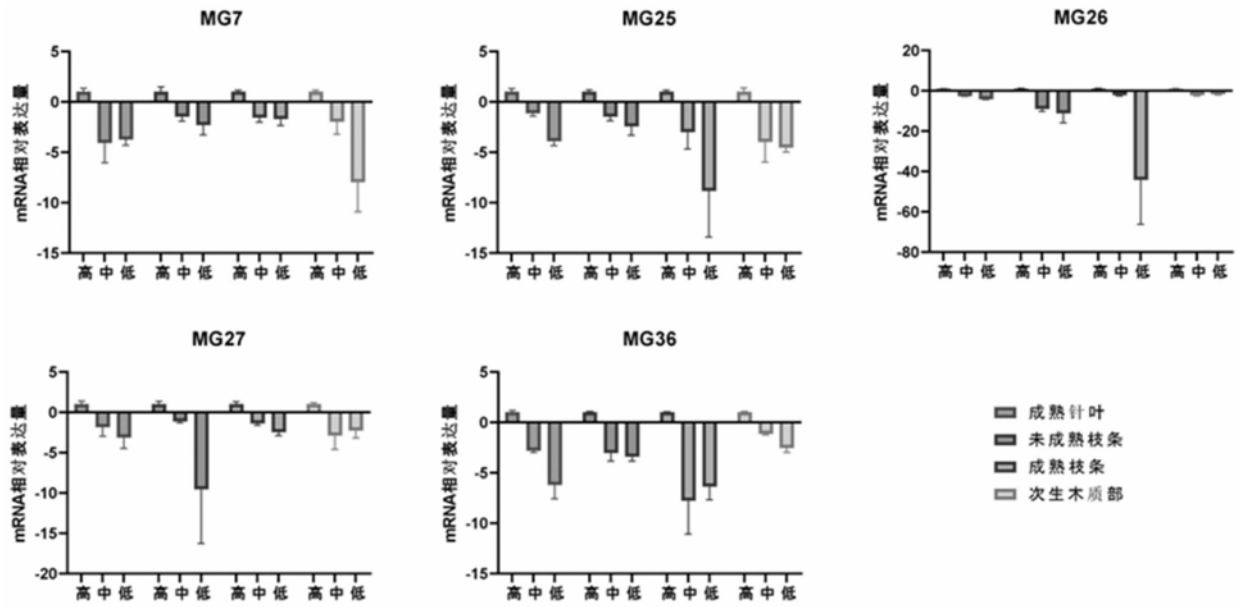


图5

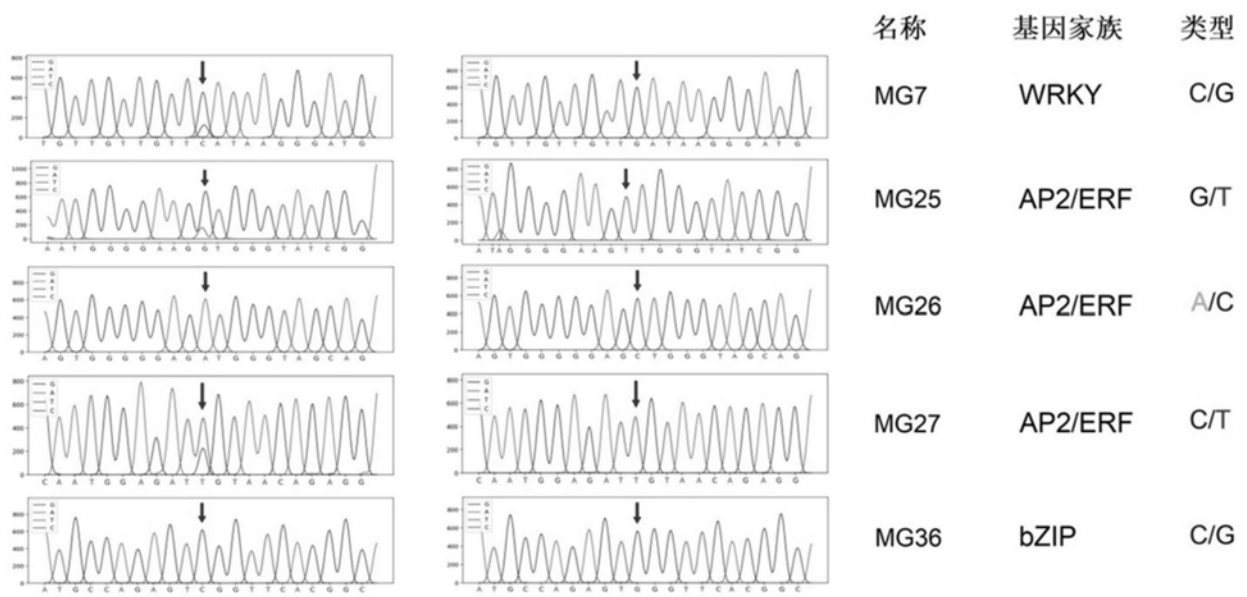


图6

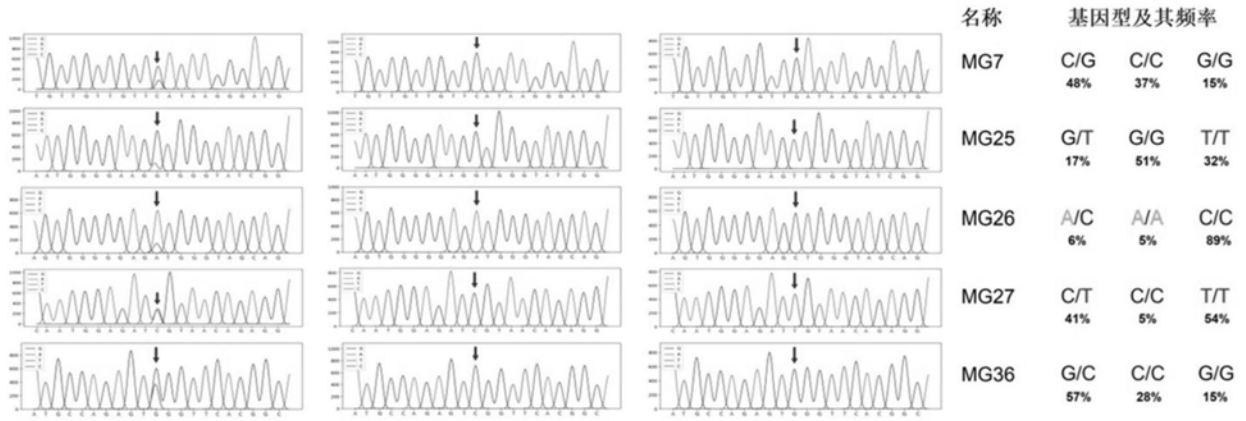


图7