

InDel 和 SNP 标记在水稻图位克隆中的应用

潘存红 王子斌 马玉银 殷跃军 张亚芳 左示敏 陈宗祥 潘学彪 *

(扬州大学 江苏省作物遗传生理重点实验室/植物功能基因组学教育部重点实验室,江苏 扬州 225009 ; * 通讯联系人 ,E mail :shuidao@yzu .edu .cn)

InDel and SNP Markers and Their Application in Map Based Cloning of Rice Genes

PAN Cun hong , WANG Zi bin , MA Yu yin , YIN Yue jun , ZHANG Ya fang , ZUO Shi min , CHEN Zong xiang , PAN Xue biao *
(Jiangsu Province Key Laboratory of Crop Genetics and Physiology/ Key Laboratory of Plant Functional Genomics , Ministry of Education , Yangzhou University , Yangzhou 225009 , China ; * Corresponding author , E mail :shuidao@yzu .edu .cn)

Abstract : The high density markers are necessary for map based cloning of rice genes , but the available marker could not reach the satisfying level . InDel (insertion deletion length polymorphism) and SNP (single nucleotide polymorphism) are new molecular markers , which could basically meet the need of the fine mapping . InDel and SNP markers could be developed through bioinformatics approaches . InDel and SNP markers are valuable markers with the characteristics of low cost , high specificity and stability . The methods for designing InDel and SNP markers were introduced and the example of mapping of the rice rolled leaf gene was used to describe the design procedures in practice . In addition , some key factors for improving the design efficiency were discussed .
Key words : insertion deletion length polymorphism ; single nucleotide polymorphism ; molecular marker ; bioinformatics ; map based cloning ; rice ; rolled leaf gene

摘 要 : 在水稻功能基因的图位克隆中 , 常需要高密度的分子标记 , 而目前公布的各种标记的密度还远未达到令人满意的程度。InDel 和 SNP 标记是新型的分子标记类型 , 基本可以满足精细定位目的基因的需要。InDel 和 SNP 标记可以较容易地通过生物信息学的方法获得 , 而且经济实用、特异性高、稳定性好。介绍了设计 InDel 和 SNP 标记的方法 , 并以一个水稻卷叶基因的研究为实例 , 介绍了在基因精细定位研究中利用这两种标记的方法和提高标记设计成功率的注意点。
关键词 : 插入缺失长度多态性 ; 单核苷酸多态性 ; 分子标记 ; 生物信息学 ; 图位克隆 ; 水稻 ; 卷叶基因
中图分类号 : Q943 .2 ; S511 .03 文献标识码 : A 文章编号 : 1001-7216(2007)05-0447-07

水稻是最重要的粮食作物 , 也是单子叶模式植物。水稻重要基因的精细定位和克隆已成为当前植物功能基因组学研究的热点之一 , 而功能基因克隆的最基本和最重要的方法是图位克隆。图位克隆的关键是寻找与目标基因紧密连锁的分子标记 , 然而 , 目标基因周围已有的分子标记密度常常不能满足需要。因此 , 合理地利用亲本之间 DNA 序列的多态性来发展新的分子标记 , 是提高图位克隆效率的重要方法。双亲间 DNA 序列的多态性是发展分子标记的基础 , 基于 PCR 的共显性分子标记是现在最常用的标记类型。通常用于发展这类分子标记的 DNA 序列多态性有 3 种基本的形式 : 简单序列重复长度多态性 (simple sequence repeat length polymorphism , SSR)、插入缺失长度多态性 (insertion deletion length polymorphism , InDel) 和单核苷酸多态性 (single nucleotide polymorphism , SNP) , 由这 3 种基本的多态性产生的分子标记分别为 SSR 标记、InDel 标记和 SNP 标记。

SSR 标记应用最为广泛 , 在水稻基因组中已经公布了 2740 个 SSR 标记 , 平均每 157 kb 就有 1 个

SSR 标记^[1-2] , 而水稻基因组中平均每 10.36 kb 有 1 个基因^[3]。由此看来 , 尽管 SSR 标记增加了遗传图谱的密度 , 但就某个具体需要克隆的基因而言 , 这样的标记密度还远未达到图位克隆所需的密度。
InDel 和 SNP 在水稻基因组中都有较高的分布频率。Shen 等^[4] 将日本晴 (Oryza sativa L . ssp . japonica) 和 9311 (Oryza sativa L . ssp . indica) 的全基因组序列用软件 MUMmer3.0 进行了比对 , 构建了反映这两个基因组序列多态性差异的 SNP 和 InDel 数据库 , 结果表明平均每 268 bp 有 1 个 SNP , 每 953 bp 有 1 个 InDel。Feltus 等^[5] 也对日本晴和 9311 的全基因组序列进行了 BLAST 比较 , 发现这两个基因组之间的 SNP 和单碱基的 InDel 的多态率分别是每 1 kb 有 1.70 个和 0.11 个 , 而且它们在基因组中的分布是不均匀的 , SNP 的富集区域比稀缺区域少 4 × 10⁶ bp。这就需要研

收稿日期 : 2007-02-09 ; 修改稿收到日期 : 2007-03-22。
基金项目 : 长三角两省一市联合科技攻关项目 (BE2004399)。
第一作者简介 : 潘存红 (1977 -) , 男 , 博士研究生。

究者在某个具体基因的图位克隆中除利用已公布的多态性数据外,还要对日本晴和 93 11 在所涉及的染色体区间的基因组序列^[6 7]进行比对,寻找 InDel 和 SNP,并据此设计分子标记。通过比较 PCR 产物的长度差异能较容易地发现 InDel 标记,而且技术要求低、针对性强、稳定性高。SNP 标记的检测成本和技术相对高些,但通过一些衍生的检测手段同样能使它成为一种实用的分子标记。本文拟就 InDel 和 SNP 多态性检测、InDel 标记和 SNP 标记的设计方法以及这两种标记在图位克隆中的应用进行论述。

1 InDel 标记

1.1 从水稻公共数据库中寻找 InDel

研究者可利用生物信息学的方法来设计 InDel 标记。

水稻图位克隆中的亲本一般都具有籼粳差异,因而可根据粳稻日本晴和籼稻 93 11 的基因组序列差异设计 InDel 标记。日本晴的 BAC (bacterial artificial chromosome) 序列已在染色体上定位,用鸟枪法测序的 93 11 基因组序列也通过与日本晴的序列比对而被锚定在各个染色体上。设计分子标记需要先从 RGP 网站 (<http://rgp.dna.affrc.go.jp>) 或

TIGR 网站 (<http://www.tigr.org>) 找到目标区段所包括的日本晴 BAC,根据需要进行选择目标 BAC,然后在 NCBI 网站 (<http://www.ncbi.nlm.nih.gov/blast/>) 的 PLANT BLAST 网页,将此 BAC 序列与 93 11 基因组序列进行 BLAST 比较^[8],通过分析 BLAST 结果,就可以找到理想的 InDel。第 1 种方法是寻找带有“-”的位置(“-”表示缺失的碱基),第 2 种方法是比较每一段序列前后的数字差异(数字表示在 BAC 克隆中的碱基位置),再计算得出差异大小。我们在精细定位一个水稻卷叶基因时,就是利用这两种方法设计出了 InDel 标记 P112.9 4 和 P112.9 6 (图 1)。

InDel 长度变化较大,主要集中在 1 ~ 640 bp,其中 1 ~ 2 bp 的 InDel 占 56%,小于 20 bp 的 InDel 占 92%^[9]。寻找 InDel 时,应优先选择多态性差异在 5 ~ 20 bp 的 InDel,这样的 InDel 设计的成功率高,而且便于检测。对于一个具体的研究群体而言,双亲基因组之间的多态性情况可能有别于日本晴与 93 11 之间的多态性,如双亲之间并不具有明显的籼粳差异,就需要在参考日本晴/93 11 之间基因组序列多态性的同时,通过测序来寻找 InDel。

1.2 InDel 标记的设计

将包含 InDel 的约 1000 bp 的日本晴序列导入

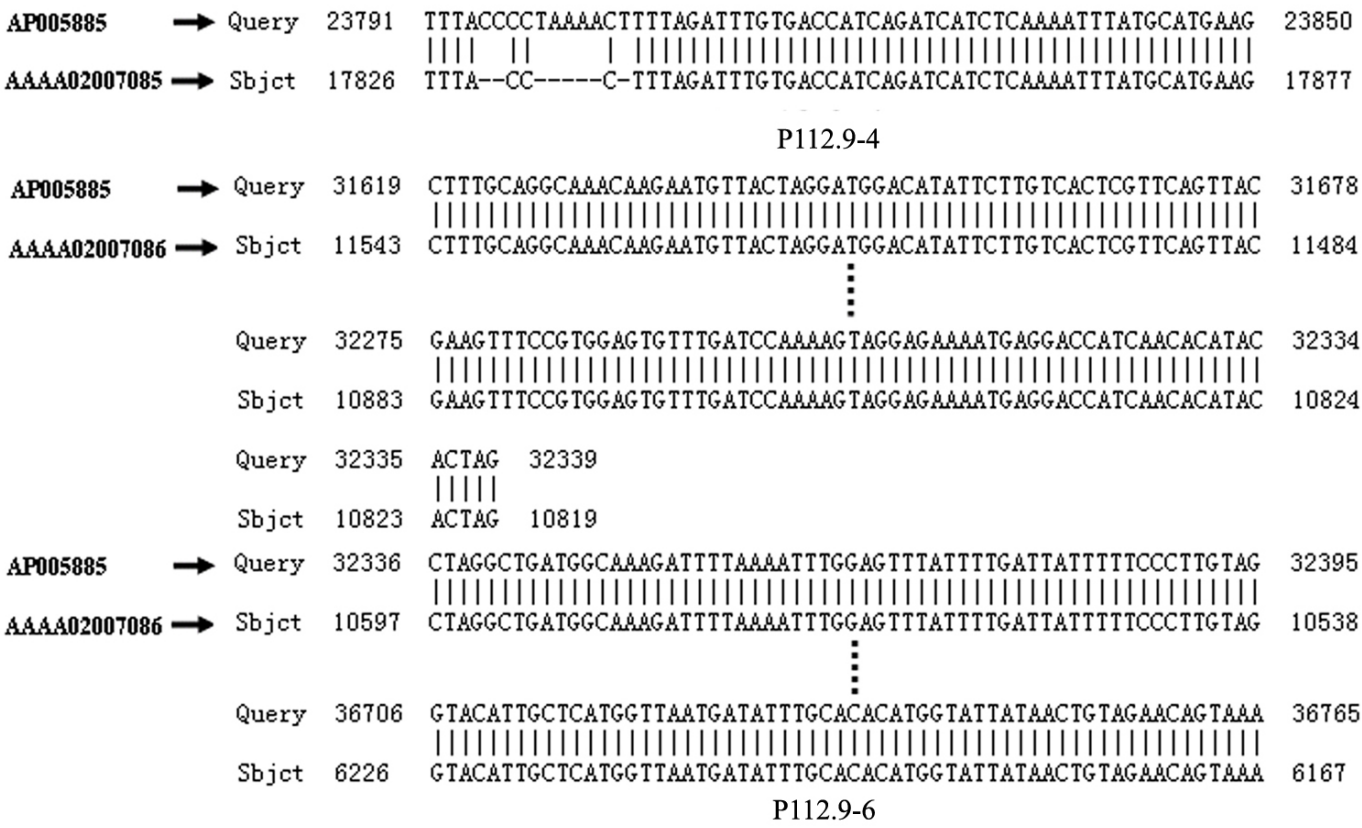


图 1 InDel 标记示意图

Fig. 1. Schematic representation of InDel markers.

InDel 标记 P112.9 4 和 P112.9 6 是由 AP005885 (日本晴克隆) 与 93 11 基因组序列 BLAST 比较得出的。P112.9 4 所示日本晴基因组比 93 11 基因组多 8 bp (AP005885 :23798 - 23805), 据此差异设计出标记 P112.9 4; AP005885 与 AAAA02007086 的两个片段比较可以发现 93 11 基因组比日本晴基因组多 225 bp (AAAA02007086 :10594 - 10818), 据此差异设计出标记 P112.9 6。

The InDel markers P112.9 4 and P112.9 6 were designed on the basis of polymorphism between the Nipponbare genome and 93 11 genome, marker P112.9 4 had an eight nucleotide insertion in the Nipponbare genome (AP005885 :23798 - 23805); Marker P112.9 6 had 225 nucleotide insertion in the 93 11 genome (AAAA02007086 :10594 - 10818).

软件 Primer Premier 5 .0 中设计引物^[10]。引物长度设置在 18 ~ 22 bp ,正反向引物长度差异小于 3 bp ,一个引物的 3 端应尽量避免与另一引物任何位点产生互补 ,以免形成“引物二聚体” ;引物的 T_m 值控制在 50 ~ 60 ,且正反向引物 T_m 值差异应小于 5 ;正反向引物的各种发卡、错配、二级结构的参数 G 最好小于 29.288 kJ/mol(7 kcal/s/mol) ,引物末端应无回文结构 ;PCR 产物长度控制在 500 bp 以内为宜 ,正向和反向引物序列与 93-11 基因组序列应完全匹配。若 InDel 正好处于某个限制性内切酶位点上 ,也可转化为 CAPS (cleaved amplified polymorphic sequence) 标记^[11] ,即扩增片段酶切长度多态性。

2 SNP 标记

SNP 标记是由 DNA 序列中因单个碱基的变异而引起的遗传多态性 ,它是所有标记中密度最高的 ,是克隆水稻功能基因的标记之一 ,其高检测成本和技术曾一度限制了它在水稻遗传育种中的应用。近年来 ,检测 SNP 的方法越来越多 ,而且更加经济实用 ,使它成为继 SSR 之后的新一代分子标记。SNP 标记在水稻遗传图谱的构建、图位克隆和 QTL 定位、标记辅助选择育种、物种进化等方面已被广泛应用。

2.1 从水稻公共数据库中寻找 SNP

同上述 InDel 标记一样 ,将定位区间内的日本晴基因组的 BAC 序列与 93-11 的基因组序列进行 BLAST 比对 ,寻找 SNP 位点。由于这两个水稻基因组测序结果中都存在间隙 ,可能某些区段无法匹配到 93-11 基因组序列 ,则要依照日本晴的序列设计 PCR 引物扩增双亲 DNA 模板 ,再通过测序的方法获得。用基于 PCR 的方法对 SNP 进行检测 ,使它成为简单实用而且高效的分子标记。

2.2 SNP 标记的设计

2.2.1 将 SNP 标记转化为 CAPS 标记

SNP 正好处于限制性内切酶位点上 ,可将它转化为 CAPS 标记(图 2 A) ,即在 SNP 两侧设计引物 ,经限制性内切酶对 PCR 产物进行酶切后 ,电泳检测产物的多态性。

2.2.2 将 SNP 标记转化为 dCAPS (derived cleaved amplified polymorphic sequence) 标记

SNP 所处的位置不是限制性内切酶识别位点 ,但可以经改造引入错配碱基而变为酶切位点 ,将它转化为 dCAPS 标记(即衍生的扩增片段酶切长度

多态性标记)^[12-13]。可用 dCAPS Finder^[14]、SNP2CAPS^[15] 等软件来设计引物 ;或者凭经验对序列进行目测来设计引物。其基本原则是先确定引入错配碱基的正向引物 ,以紧靠 SNP 位点外侧的碱基为正向引物的 3 末端向 5 端延伸约 20 bp 作为正向引物的基序 ,再在 3 端引入 1 ~ 2 个错配碱基 ,使在亲本之一的 PCR 扩增产物上产生一个常用限制性内切酶识别位点 ,而在另一个亲本的 PCR 扩增产物上并没有新增这个限制性内切酶识别位点 ,即双亲的 PCR 产物具有酶切长度差异。然后再按引物设计的一般原则确定反向引物。产物大小应控制在 500 bp 以下(图 2 B)。

dCAPS 标记现已是水稻图位克隆中不可或缺的分子标记 ,在水稻功能基因 *sd 1*^[16]、*Rf 1*^[17]、*Pi d2*^[18] 的图位克隆中都发展了一批新的 dCAPS 标记。

2.2.3 等位基因特异 PCR (allele specific PCR , AS PCR) 策略

基因组中只有 30% ~ 40% SNP 能转化为 CAPS 或 dCAPS 标记 ,其他的 SNP 经改造后也不能成为常用的限制性内切酶识别位点 ,但可用 AS PCR 的方法来检测^[19]。同 CAPS 和 dCAPS 标记一样 ,用 AS PCR 方法检测的 SNP 标记也是共显性标记 ,而且同样经济实用、易于操作。

与 dCAPS 标记引入错配碱基的引物后在 SNP 位点都能扩增出 PCR 产物不同的是 ,AS PCR 利用的是自然的碱基错配 ,并不能扩增出 PCR 产物。AS PCR 方法设计了两条特异的正向引物和一条反向引物 ,两条特异的正向引物的 3 端碱基分别与 SNP 的两个碱基相同(图 2- C) ,其中一条正向特异引物(Specific Primer 1)与一个亲本 (Allele 1) 能完全匹配 ,能与反向引物一起扩增出 PCR 产物 ,而与另一亲本 (Allele 2) 不能完全匹配 ,也不能扩增出 PCR 产物 ;另一特异引物同样只能与反向引物从亲本 (Allele 2) 扩增出 PCR 产物。这两种引物组合都能从杂合体中扩增出 PCR 产物(图 2- C1)。为了提高 PCR 检测的效率 ,郑艳萍等^[120] 将其中一条特异引物的 5 端加上了 20 bp 的其他物种的 DNA 序列 (M13) ,两条特异引物与反向引物配对扩增出的 PCR 产物就有了长度差异 ,3 条引物可以同时在一个 PCR 反应中检测样品的标记值 ,如图 2 C2 所示。由于特异引物、*Taq* 酶等反应体系和 PCR 反应程序的原因 ,AS PCR 并不能完全抑制由于错配引起的非特异扩增而出现假阳性的结果。为了克服这个

缺点 ,Drenkard 等^[21]在特异引物 3 端的 3 bp 的范围内又引入 1 个错配碱基来提高 AS PCR 反应的特异性 ,并编写了程序 SNAPER (<http://patho.mgh.harvard.edu/ausubelweb>)来设计 AS PCR 引物 ,示意图见图 2 C3。 Hayashi 等^[22]用改良的 AS PCR 的方法设计 SNP 标记对水稻抗稻瘟病基因 *Piz*进行了定位 ,还提出了在两个 SNP 位点相隔不远的情况下 ,可以将两条特异引物分别设计在两个 SNP 位点而且在不同亲本上 ,将 2 条特异的正向引物与反向引物放在同一 PCR 反应体系中进行 PCR 扩增 ,这样既保证了反应的特异性 ,又提高了检测效率 (图 2 C4)。但是 ,有时由于两个 SNP 位点可能

相距太远或两个特异引物之间可能会形成互补、发夹结构等原因而不能在同一 PCR 反应体系中进行扩增。综合以上方法的优缺点 ,我们在研究中提出 ,在增加另一个错配碱基的基础上 ,将其中一条特异引物的 5 端接上 20 bp 其他物种的序列 ,即可将 3 条引物置于同一 PCR 体系中对群体进行标记检测 ,其引物设计见图 2 C5 ,PCR 产物的电泳示意图见图 2 D。

为了增加 AS PCR 的特异性 ,在引物设计和 PCR 反应时需遵循以下原则 :由于在正向引物 3 端引入错配碱基的种类和数量不同 ,AS PCR 的特异性也不同 ,研究者需要用不同的 PCR 反应条件来检

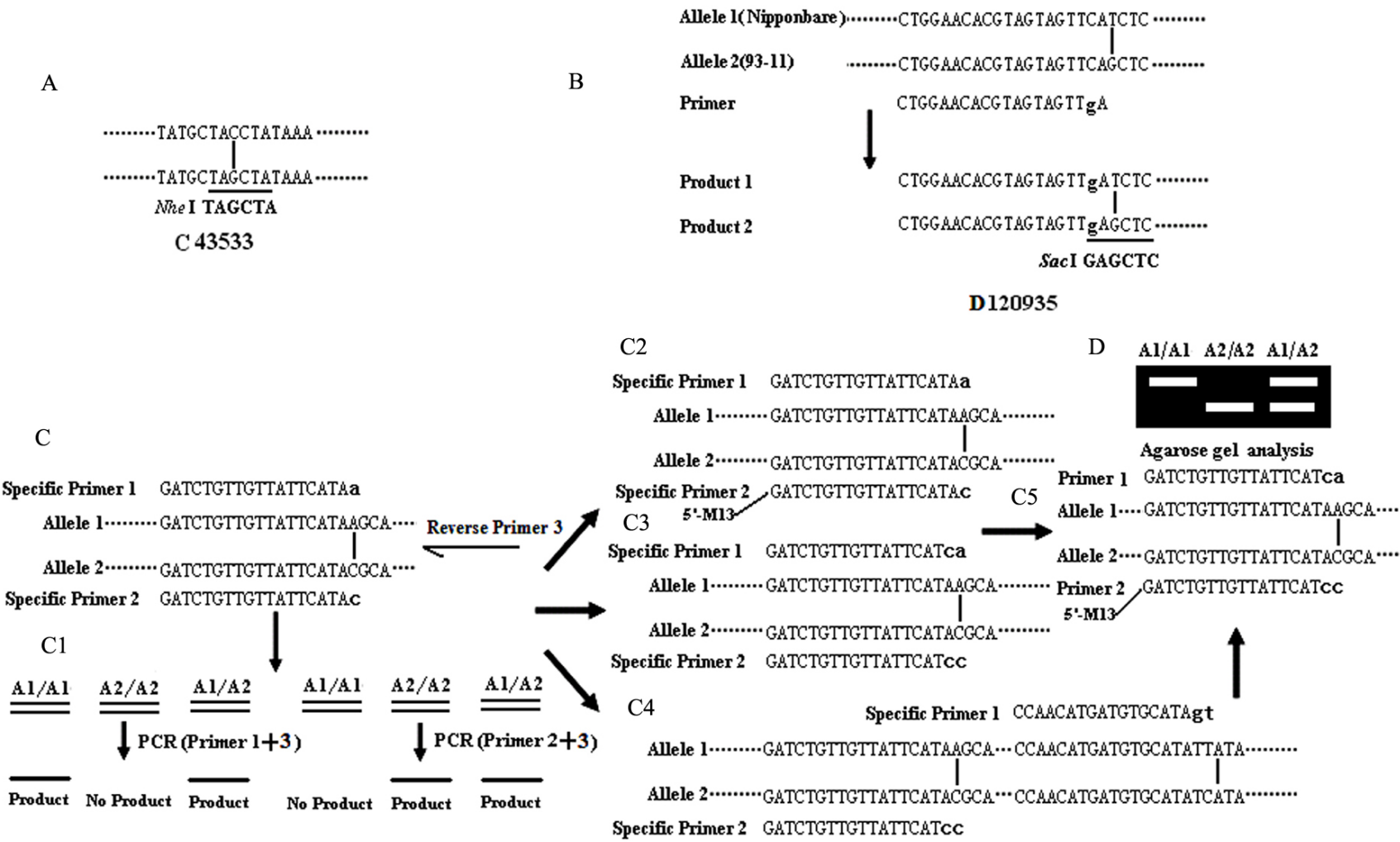


图 2 SNP 标记示意图

Fig 2 . Schematic representation of SNP markers .

SNP 是由日本晴克隆 AP005885 与 93 11 基因组序列通过 BLAST 比较而来。 A - CAPS 标记 ,标记 C43533 具有限制性内切酶 *Nhe* 识别位点差异 ; B - dCAPS 标记 ,标记 D120935 具有 T/G 的差异 ,在正向引物中引入 1 个错配碱基使两亲本的 PCR 产物具有限制性内切酶 *Sac* 识别位点差异 ; C - AS PCR 标记 ,日本晴与 93 11 基因组具有 A/C 的差异 ,设计两条仅在 3 末端有差异的正向引物和一条反向引物通过两次 PCR 扩增使之成为共显性标记 ; C1 - AS PCR 标记的 PCR 扩增示意图 ; C2 - 在一条正向引物 5 端加 20 bp 左右的 M13 序列 ; C3 - 在两条正向引物中引入 1 个错配碱基 ; C4 - 在相距 500 bp 以内的两个 SNP 分别设计一条正向引物 ; C5 - 在正向引物中引入错配碱基 ,在另一条引物 5 端加入 M13 序列 ; D - AS PCR 电泳示意图。

SNP was derived from polymorphism between the 93 11 genome and Nipponbare genome . A , CAPS marker . Marker C43533 had polymorphism of the restriction enzyme site (*Nhe*) ; B , dCAPS marker . Marker D120935 had a single nucleotide change from T to G , and a mismatch was incorporated into the forward primer to create novel restriction site (*Sac*) in one of the alleles . C , AS PCR marker . A codominant marker was designed with two forward and one reverse primers according to a change from A to C . C1 , Schematic representation of AS PCR strategy ; C2 , A twenty- nucleotide sequence of M13 were incorporated into the 5 end of one forward primer ; C3 , A mismatch was incorporated into two forward primers ; C4 , Two forward primers were designed in two near SNPs ; C5 , Mismatches were incorporated into two forward primers , furthermore , sequences of M13 were incorporated into the 5 end of one forward primer . D , Schematic of agarose gel analysis of AS PCR marker .

测引物的特异性 ,选择特异性高的引物用于基因定位。每个 SNP 可设计 2~4 组引物 ,在 PCR 反应中设立高退火温度和低退火温度 (根据具体引物来定)、高浓度和低浓度的模板 DNA (相差 10 倍以上)、多循环数和少循环数 (38 个和 28 个循环)的对比试验 ,选择在各对比试验中都表现有特异性的引物组 ,即根据亲本之一设计的特异引物只能在这个特异亲本中扩增出产物 ,而在另一个非特异亲本中并不能扩增出产物。

2.2.4 其他检测方法

在 SNP 处设计的 dCAPS 和 AS-PCR 引物并不都能有效地扩增出 PCR 产物 ,或者扩增出的产物中带有非目标产物而不能正确地进行标记检测分析。这时 ,可用其他方法来对个体进行 SNP 分型^[23]。

3 InDel 和 SNP 标记在水稻基因精细定位中应用的一个实例

3.1 研究的基因和精细定位策略

我们研究的是一个卷叶基因。该基因是一个不完全隐性主基因 ,杂合体表型为叶片中度卷曲 ,纯合

体表现为叶片卷曲。以卷叶亲本为卷叶基因供体亲本 ,与平展叶亲本杂交和回交 ,构建了近等基因系。BC₄F₂ 群体用于初步定位 ;BC₄F₂ 中选卷叶杂合单株 (经 PCR 检测在定位区间两侧标记为杂合带型) ,其自交后代共 3650 个单株 ,用于精细定位。

在初步定位区间^[25]内发展新的分子标记 ,并对分离群体进行检测。针对在这两个标记之间发生交换的单株 ,缩小定位区间 ;再用亚区间内新发展的分子标记进行检测 ,不断缩小区间。最后 ,通过对小区间的 PCR 扩增和测序 ,进一步发现多态性并发展了新的分子标记 ,达到精细定位的目的。

3.2 定位区间内 InDel 和 SNP 标记的发展

由于该卷叶基因所覆盖的物理区间内没有 Shen 等^[4]和 Feltus 等^[5]公布的 InDel 和 SNP ,故从 RGP 下载此区间日本晴克隆 (AP004188、AP004160、AP005885)的序列 ,通过与籼稻品种 9311 的序列进行 BLAST 比较来获得 InDel 和 SNP 多态性信息 ,发展新的分子标记。研究中先后共发展了 25 个 InDel 标记和 15 个 SNP 标记 ,其中 33 个标记在亲本间表现多态性 (表1 ,表2) ,成功率高达

表 1 新发展的亲本间多态性 InDel 标记

Table 1 . Primer sequences of the polymorphic InDel markers developed in mapping rolled leaf gene .

标记 Marker	BAC	正向引物序列 Forward primer	反向引物序列 Reverse primer	产物长度 Length of product/bp	InDel 长度 Insertion deletion length/bp
P112.8	AP004160	5-CAAGTCTAAAGAGGGAGTGGG-3	5-AAGCAGGGTAGTAAACAGCGT-3	354	56
P112.9.1	AP004160	5-TGATGACGGGT TCTTCTG-3	5-TATAATGTGGGGGCCTTT-3	497	20
P112.9.2	AP004160	5-GCTCTCCCGCCCATCCTA-3	5-ACTCGTCCGAACGTGCCA-3	205	11
P112.9.3	AP004160	5-TCCACAATCATAGCACCTG-3	5-GTGTTGGAGGGT TACTTTG-3	585	26
P112.9.4	AP005885	5-CATTGACCTTGCCAGACT-3	5-CGATGCTTTGGTTAGATAG-3	332	8
P112.9.5	AP005885	5-ATAGGTGTCCTATTCTGTTTG-3	5-AGTATGTTTGT TTTGTGTTGA-3	203	6
P112.9.6	AP005885	5-GATTAGGAGGCATCAACTG-3	5-ATAACCGATAAGCCGTACA-3	686	225
P112.9.7	AP005885	5-TGAGATTGCGGTGACGAA-3	5-AGCATCTAAGGTCCTGGGT-3	711	70
P112.9.8	AP005885	5-CTACACCATAGCATTCCTACT-3	5-CCAACAAGTATCAGCTAGTAG-3	312	20
P113.1	AP005885	5-TCAAACCCTATTTTACACG-3	5-TTTATCATAACCCCATCCA-3	548	25
P113.2	AP005885	5-TAGCACAGTCGATTAGGC-3	5-ATGTTCTTCCGTGTTATG-3	346	6
P113.2.4	AP005885	5-TCTTAGAGGCTCATCCAG-3	5-ACAGAGTAGGCACATTTCG-3	512	40
P113.3	AP005885	5-ATAAAGGCACAGTTTGGA T-3	5-TCTGTCTTGTTTGGTTTTG-3	320	6
P113.4	AP005885	5-AGGCTTATTTAGTTGAAGG-3	5-AAGTTGAAAGGAGATGCTA-3	410	11
P113.6	AP005885	5-TACAGTAGGTGGAAGGGAA-3	5-TCCTACAAGACTGGAAAGG-3	329	13
P113.7	AP005885	5-TTGTTTCGTAGAGGTGGTG-3	5-GATTTGTGGGTTTGCTTG-3	394	26
P114	AP006060	5-CTGTAGTTGTATAGGGGAGT-3	5-CAGATTGACTGGGAAGAG-3	301	13
P115	AP006060	5-GCTTGACAAATGGTGAGGT-3	5-GGCTGAAAGTCTGAAACG-3	441	21
P115.5	AP006060	5-GGCATTCAAATACTCCGTG-3	5-GGGTCTCCTGGGCTTCT-3	238	12
P116	AP006060	5-TCCAAGTAGAGGCAACAG-3	5-ATATCCGGGATAAAGACG-3	351	20
P117	AP006060	5-CTAGAGGTTATGGAATGGGTT-3	5-GAGACGGTGGTGGGAGAG-3	344	10
P118	AP005775	5-GTATTTGCGGATAATGTCTG-3	5-TTTGGGATGTTAGTTTGAT-3	508	34

表 2 新发展的亲本间有多态的 SNP 标记

Table 2 . Primer sequences of the polymorphic SNP markers developed in mapping rolled leaf gene .

标记	类型	正向引物序列	反向引物序列	限制性 内切酶	产物长度 Length of
Marker	Type	Forward primer	Reverse primer	Restriction enzyme	products /bp
C18813	CAPS	5- AAGCGAAGGCATCCACAA 3	5- GCATTTCTCCACCAACG 3	<i>Xba</i>	551
C20250	CAPS	5- GTATCCGGTTTTCTGGTG 3	5- GCCTATTTGAGCCTCTATG 3	<i>Xsp</i>	283
C43533	CAPS	5- ATAAAGGGAAAGAGGAGA 3	5- AGCAGTCTGACGATTCTA 3	<i>Nhe</i>	256
C16237	CAPS	5- TGCCATTCCGTGACAACA 3	5- CCATTTCTGCCCTACGCT- 3	<i>PshB</i>	977
D21991	dCAPS	5- ATATGGAGCGTGCA TGA_ G)TA 3	5- GAATAGTCCTGGTGCAGT 3	<i>Ssp</i>	264
D36046	dCAPS	5- TTCTCTAGTTAATGAAG_ C AA)AT 3	5- CTGAAACCTCTGTTAATAACT- 3	<i>Sph</i>	149
D37366	dCAPS	5- CTAGGCCTCGACCCCTT 3	5- CGACCACGAGCACTGC_ G)- 3	<i>Pvu</i>	278
D35565	dCAPS	5- GGAATAACATGCTTGTTT_ T)AG 3	5- CTGAATTCATGCTGAAGAGAA 3	<i>Hind</i>	143
D120935	dCAPS	5- CTGGAACACGTAGTAGTT_ C)A 3	5- GGAGAGGTTGAACAGGTAAG 3	<i>Sac</i>	195
D128465	dCAPS	5- TGGAGACCACGGTATGG_ C)A 3	5- ACTCGAAACAAGA ACTC_ A)TG 3	<i>Hind / Xho</i>	319
A101289	AS PCR	5- CACGACGTTGTAAAACGAG- GATCTGTTGTTATTCAT_ A)G 3 5- GATCTGTTGTTATTCAT_ A)A 3	5- CAATGTTGCACTCTTGATCGA 3		233

注 :带下划线的碱基为错配碱基 ,括号内碱基为错配碱基的原始碱基。
Note : Mismatches are underlined ,letters in parentheses are original nucleotides .

82.5%。通过发展这些 InDel 和 SNP 标记 ,最终实现了对目的基因的精细定位和克隆(详细结果将另文报道)。

4 InDel 和 SNP 标记利用中的注意点

4.1 InDel 和 SNP 标记在基因图位克隆中的重要性

本研究组在精细定位卷叶基因时共发展了 40 个标记 ,全部为 InDel 标记(25 个)和 SNP 标记(15 个)。其中 33 个标记在亲本间表现多态性 ,同时在日本晴与 93 11 间也存在多态性 ,多态率为 82.5%。这样高的成功率充分说明了这两种标记在基因精细定位中的重要性。

4.2 InDel 标记利用中的几个注意点

本研究组设计的标记中 InDel 的平均大小为 31 bp ,符合 InDel 在近缘种属间平均差异大小为 36 bp 的结论^[24] ,这也进一步说明 ,较小的 InDel 在近缘种属间是广泛分布的。由于 InDel 可能会引起物种的分化 ,在近缘种属中就存在较多的 InDel ,这也给我们应用 InDel 标记提供了可能。InDel 长度变化较大 ,小于 20 bp 的 InDel 占了 92% ,较长的 InDel 可能对生物有不同程度的危害 ,其比例也相对很少^[9]。在 InDel 标记的设计中应多利用较小的 InDel ,1~2 bp 的 InDel 也可依照 SNP 标记的设计方法。较长的 InDel 若经设计的引物检测是存在的 ,则这个 InDel 标记由于其 PCR 产物经较短的电泳时间就能分出多态而在检测大的定位群体时更为

高效和实用。

93 11 的全基因组序列存在少量的测序错误^[25] ,为了提高标记设计的成功率 ,根据经验 ,优先选择多态性位点两侧序列在日本晴和 93 11 基因组中完全匹配的 InDel ,另外插入发生在 93 11 基因组中 ,或缺失发生在日本晴基因组中的 InDel 也是优先考虑的对象。

4.3 SNP 标记利用中的几个注意点

SNP 是物种进化源泉——点突变的表现形式 ,不仅在水稻亚种籼稻与粳稻之间有非常高的分布密度 ,而且在亚种内的品种间也有较多的分布。Nasu 等^[26] 分析了 3 个粳稻品种、2 个籼稻品种、1 个野生稻之间 SNP 的发生频率 ,对 6 个材料两两比较 ,发现每 232 个碱基存在 1 个 SNP ,亲缘关系较远的广陆矮 4 号和 Kasalath 之间的 SNP 频率为 0.49% ,亲缘关系较近的粳稻品种日本晴与 Koshihikari、Kitaake 之间的 SNP 频率分别为 0.03% 和 0.05%。这些数据表明 ,在水稻的图位克隆中 ,当亲本的亲缘关系较近时 ,对 SNP 标记的利用就显得更加重要。在亲缘关系较近的水稻品种间发展 SNP 标记时 ,除利用公布的一些基因组数据外 ,还可利用目标区段内的 EST 和 UniGene 数据库^[27] ,或直接对亲本基因组进行测序来寻找 SNP ,根据 SNP 的具体形式来发展分子标记。

4.4 InDel 和 SNP 标记的优缺点比较

InDel 和 SNP 在水稻基因组中的分布频率较高 ,因而 InDel 和 SNP 标记是图位克隆中不可或缺

的。InDel 标记检测方法简单、经济,在基因的定位区段还比较大时,应优先考虑。SNP 标记对引物设计和检测方法都有着较高的要求,当基因的定位区段较小而找不到 InDel 标记时,利用 SNP 标记可能进一步缩小定位区间。同 InDel 相比,SNP 在基因内部的分布频率也较高^[28],SNP 标记有可能与目标基因共分离。尽管 InDel 和 SNP 在水稻基因组中的分布频率都较高,但它们的分布是不均匀的,它们的位置有一定互补性,可根据目标基因所在区段来设计相应的分子标记。

参考文献：

[1] Temnykh S , DeClerck G , Lukashova A , et al . Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L .) : Frequency , length variation , transposon associations , and genetic marker potential . *Genome Res* , 2001 , 11 : 1441-1452 .

[2] McCouch S , Teytelman L , Xu Y B , et al . Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L .) . *DNA Res* , 2002 , 9 : 199-207 .

[3] International Rice Genome Sequencing Project . The map based sequence of the rice genome . *Nature* , 2005 , 436 : 793-800 .

[4] Shen Y J , Jiang H , Jin J P , et al . Development of genome wide DNA polymorphism database for map based cloning of rice genes . *Plant Physiol* , 2004 , 135 : 1198-1205 .

[5] Feltus F A , Wan J , Schulze S R , et al . An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments . *Genome Res* , 2004 , 14 : 1812-1819 .

[6] Goff S A , Ricke D , Lan T H , et al . A draft sequence of the rice genomes (*Oryza sativa* L . ssp . *japonica*) . *Science* , 2002 , 296 : 92-100 .

[7] Yu J , Hu S , Wang J , et al . A draft sequence of the rice genome (*Oryza sativa* L . ssp . *indica*) . *Science* , 2002 , 296 : 79-92 .

[8] Altschul S F , Gish W , Miller W , et al . Basic local alignment search tool . *J Mol Biol* , 1990 , 215 : 403-410 .

[9] Tenaillon M I , Sawkins M C , Anderson L K , et al . Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays* ssp . *mays* L .) . *Genetics* , 2002 , 162 : 1401-1413 .

[10] Singh V K , Mangalam A K , Dwivedi S , et al . Primer premier : Program for design of degenerate primers from a protein sequence . *Biotechniques* , 1998 , 24(2) : 318-319 .

[11] Konieczny A , Ausubel F M . A procedure for mapping *Arabidopsis* mutations using co dominant ecotype specific PCR-based markers . *Plant J* , 1993 , 4(2) : 403-410 .

[12] Michaels S , Amasino R . A robust method for detecting single

nucleotide changes as polymorphic markers by PCR . *Plant J* , 1998 , 14(3) : 381-385 .

[13] Neff M , Neff J , Chory J , et al . dCAPS , a simple technique for the genetic analysis of single nucleotide polymorphisms : Experimental applications in *Arabidopsis thaliana* genetics . *Plant J* , 1998 , 14(3) : 387-392 .

[14] Thiel T , Kota R , Grosse I , et al . SNP2CAPS : A SNP and INDEL analysis tool for CAPS marker development . *Nucl Acids Res* , 2004 , 32 : e5 .

[15] Neff M , Turk E , Kalishman M . Web based primer design for single nucleotide polymorphism analysis . *Trends Genet* , 2002 , 18(12) : 613-615 .

[16] Monna L , Kitazawa N , Yoshinl R , et al . Positional cloning of rice semidwarfing gene , *sd 1* : Rice “Green Revolution Gene” encodes a mutant enzyme involved in gibberellin synthesis . *DNA Res* , 2002 , 9 : 11-17 .

[17] Komori T , Ohta S , Murai N , et al . Map based cloning of a fertility restorer gene , *Rf 1* , in rice (*Oryza sativa* L .) . *Plant J* , 2004 , 37(3) : 315-325 .

[18] Chen X W , Shang J J , Chen D X , et al . A B lectin receptor kinase gene conferring rice blast resistance . *Plant J* , 2006 , 46(5) : 794-804 .

[19] Ugozzoli L , Wallace R B . Allele specific polymerase chain reaction . *Methods Enzymol* , 1991 , 2 : 42-48 .

[20] 郑艳萍 , Chander Subhash , 杨小红 , 等 . 一种基于聚合酶链式反应检测 SNP 的方法 . 中国农业大学学报 , 2006 , 11(3) : 51-55 .

[21] Drenkard E , Richter B G , Rozen S , et al . A simple procedure for the analysis of single nucleotide polymorphisms facilitates map based cloning in *Arabidopsis* . *Plant Physiol* , 2000 , 124 : 1483-1492 .

[22] Hayashi K , Yoshida H , Ashikawa I . Development of PCR based allele specific and InDel marker sets for nine rice blast resistance genes . *Theor Appl Genet* , 2006 , 113 : 251-260 .

[23] 高秀丽 , 景奉香 , 杨剑波 , 等 . 单核苷酸多态性检测分析技术 . 遗传 , 2005 , 27(1) : 110-122 .

[24] Britten R J , Rowen L , Williams J , et al . Majority of divergence between closely related DNA samples is due to indels . *PNAS* , 2003 , 100(8) : 4661-4665 .

[25] Han B , Xue Y B . Genome wide intraspecific DNA sequence variations in rice . *Curr Opin Plant Biol* , 2003 , 16 : 134-138 .

[26] Nasu S , Suzuki J , Ohta R , et al . Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa* , *Oryza rufipogon*) and establishment of SNP markers . *DNA Res* , 2002 , 9 : 163-171 .

[27] 杨 仑 , 沈文飏 , 陈 虹 , 等 . 基于生物信息学的水稻候选 SNP 发掘 . 中国水稻科学 , 2004 , 18(3) : 185-191 .

[28] Feng Q , Zhang Y , Hao P , et al . Sequence and analysis of rice chromosome 4 . *Nature* , 2002 , 420 : 316-320 .