

白菜型油菜和菜薹的 InDel 标记开发及其 RILs 群体遗传连锁图谱的构建

孟霖^{1,2,*}, 刘博^{2,*}, 林良斌¹, 程峰², 王晓武², 武剑^{2,**}

(¹ 云南农业大学农学与生物技术学院, 昆明 650201; ² 中国农业科学院蔬菜花卉研究所, 北京 100081)

摘要: 通过白菜型油菜 ‘R-O-18’ 和菜薹 ‘L58’ 的基因组重测序数据与白菜基因组的参考序列 (‘Chiifu-401-42’ 的基因组序列) 的比对, 在全基因组范围内检测到了 18 479 个短插入缺失变异位点 (≤ 5 bp)。从中挑选了 500 个插入/缺失片段为 4 ~ 5 bp 的 InDel 变异位点, 将其设计成 InDel 分子标记并进行试验验证, 结果有 452 个标记通过 PCR 扩增出单一条带, 但仅有 106 个在 ‘R-O-18’ 和 ‘L58’ 间表现出多态性, 346 个没有多态性, 48 个在 PCR 中没有扩增。亲本间具有多态性的 106 个 InDel 标记用来检测以 ‘R-O-18’ 和 ‘L58’ 为亲本构建的 RILs 基因型, 并构建了一张包含 99 个标记的遗传连锁图谱。

关键词: 菜薹; 油菜; 白菜; 插入/缺失; 全基因组重测序; Illumina 测序; 遗传连锁图

中图分类号: S 634

文献标识码: A

文章编号: 0513-353X (2012) 08-1491-10

Development of InDel Markers for *Brassica campestris* and Genetic Linkage Map Construction of the RILs Population

MENG Lin^{1,2,*}, LIU Bo^{2,*}, LIN Liang-bin¹, CHENG Feng², WANG Xiao-wu², and WU Jian^{2,**}

(¹ College of Agronomy and Biotechnology, Yunnan Agricultural University, Kunming 650201, China; ² Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China)

Abstract: Genome-wide *Brassica campestris* short Insertion/Deletion length polymorphisms (InDels) (≤ 5 bp) were detected by aligning whole genome re-sequencing data from two *B. campestris* accessions, ‘R-O-18’ and ‘L58’, to the reference genome sequence of ‘Chiifu-401-42’. In total, we identified 18 479 InDel polymorphisms between ‘R-O-18’ and ‘L58’. A total of 500 InDel polymorphisms (4 ~ 5 bp in length) were converted to InDel markers and experimentally validated. Of the selected 500 InDel polymorphisms, 452 yielded a single PCR fragment, but only 106 showed polymorphism between ‘R-O-18’ and ‘L58’, 48 did not amplify a product, 346 showed no polymorphism. These 106 InDel markers were used to screen the genotype of RILs developed from the crossing of ‘R-O-18’ and ‘L58’. A genetic linkage map contained 99 InDel markers was constructed.

Key words: *Brassica campestris*; insertion/deletions (InDels); whole genome re-sequencing; Illumina sequencing; genetic linkage map

收稿日期: 2012-04-23; 修回日期: 2012-06-15

基金项目: 中央级公益性科研院所基本科研业务费专项 (1610032011011)

* 对文章同等贡献。

** 通信作者 Author for correspondence (E-mail: wujian@caas.net.cn)

白菜类作物归芸薹属芸薹种 (*Brassica campestris* syn. *rapa*), 是中国栽培分布最广、种植面积最大的蔬菜作物。白菜基因组测序计划 (The Multi-national *B. rapa* Genome Sequencing Project, BrGSP) 于 2003 年启动, 并于 2011 年完成了大白菜栽培种 ‘Chiifu-401-42’ 基因组的测序组装工作。作为 A 基因组的参考序列, 其至少覆盖了 98% 的基因区 (Wang et al., 2011a)。这一成果大大促进了基于序列的白菜分子标记的发展, 并为芸薹属作物的品种改良奠定了坚实的基础。

分子标记在基因型确定、遗传多样性分析、标记辅助育种和系统发育分析 (Vos et al., 1995; McCouch et al., 1997; Nagaraju et al., 2002) 等研究中是非常有价值的工具。基于非重复 DNA 序列变异开发的分子标记有多种: RFLP (Restriction Fragment Length Polymorphism) (Botstein et al., 1980), RAPD (Random Amplified Polymorphism) (Williams et al., 1990) 和 AFLP (Amplified Fragment Length Polymorphism) (Vos et al., 1995)。基于简单重复序列开发的标记 SSR (Simple Sequence Repeat) (Tautz, 1989) 在重复性、多等位基因性、共显性、丰富性和基因组覆盖度等方面都优于早期的分子标记 (Powell et al., 1996)。虽然丰富的文库和基因组序列在一定程度上促进了 SSR 标记的开发, 但在植物中其多态性比率一直被限制在 30%~40% 以内 (Hisano et al., 2007; Kim et al., 2009; Ren et al., 2009)。

基于全基因组序列的大规模开发的新一代分子标记 SNP (single nucleotide polymorphism) 和 InDel (insertion/deletion) 为高密度分子连锁图谱构建提供了新的途径。随着人类、水稻和拟南芥的基因组测序工作的完成, 其基因组内 InDel 分布频率分别达到 138.89、1 050 和 151 InDels · Mb⁻¹ (Jander et al., 2002; Shen et al., 2004; Mills et al., 2006)。Park 等 (2010) 用大白菜 ‘Chiifu-401-42’ 的 1 398 个 STSs 和 557 个 BAC 序列进行了比对, 结果表明白菜的 InDel 分布频率为 4.83 InDels · kb⁻¹。虽然 SNP 的频率要远高于 InDel, 但其在基因分型技术上的缺陷限制了其应用。目前虽有很多 SNP 基因分型技术 (Syvanen, 2001), 但应用在中小规模试验中大多成本较高, 操作较复杂, 而应用在大规模试验中又需要专用仪器。与之相反, 以片段长度多态性进行基因分型的 InDel 标记是可以直接应用 SSR 标记的基因分型试验技术 (Bhatramakki et al., 2002)。现在, 可以通过白菜 (*B. campestris* L. ssp. *pekinensis*) 基因组参考序列 (Wang et al., 2011a) 与重测序数据的比较, 在不同白菜品种间快速而高效地设计基于 PCR 的 InDel 标记以供高分辨率的遗传分析使用。

本研究中, 通过白菜型油菜 ‘R-O-18’ 和菜薹 ‘L58’ 的基因组重测序数据与基因组参考序列 (Wang et al., 2011a) 的比较, 在全基因组共检测到 18 479 个短 InDel 变异位点 (≤ 5 bp)。将其中 500 个位点设计成基于 PCR 的 InDel 分子标记, 并用这两个品种及以二者为亲本构建的 RILs 群体对这些标记进行了试验验证。这些新开发的基于 PCR 的 InDel 分子标记将在白菜的遗传学研究和分子辅助育种中发挥重要作用。

1 材料与方法

1.1 试验材料及基因组序列数据集

‘R-O-18’ 为白菜型油菜 (*B. campestris* L. ssp. *tricoloris*), ‘L58’ 为菜薹 (*B. campestris* L. ssp. *parachinesis*)。以 ‘R-O-18’ 和 ‘L58’ 为亲本构建的 RILs 群体 (由荷兰瓦赫宁根大学提供) 共有 160 个株系, 选取其中的 156 个株系作为供试材料。

所有供试材料于 2010 年 9 月播种于中国农业科学院蔬菜花卉研究所的温室大棚内。待其长出 5~6 片真叶时, 分别选取幼嫩叶片置于自封袋中, 于 -80 °C 超低温冰柜中保存, 用于基因组 DNA 的提取。

白菜品种 ‘Chiifu-401-42’ 的基因组序列从网站 BRAD (<http://brassicadb.org/brad/>) 上获得, 并以该序列作为分析的参考基因组序列。通过 sequencing-by-synthesis 的方法, 对 ‘L58’ 进行重测序, 一共得到 3 Gb 的原始 reads。另外, 大约 2.8 Gb 的 ‘R-O-18’ 重测序数据是通过网站 (<http://www.brassica.info/resource/sequencing.php>) 下载。

1.2 插入缺失 (InDel) 变异位点的检测及 InDel 标记的引物设计

预测 ‘R-O-18’ 和 ‘L58’ 间的 InDel 变异位点的步骤为: 以白菜基因组参考序列 (‘Chiifu-401-42’ 基因组序列) 为桥梁, 通过 ‘L58’ 的重测序数据与参考基因组序列的比对, 提取出二者间的 InDel 变异位点; 将 ‘R-O-18’ 的重测序数据比对到提取出的 InDel 位点上; ‘R-O-18’ 与 ‘Chiifu-401-42’ 序列相同的位点被认为是 ‘R-O-18’ 与 ‘L58’ 间具有多态性的位点。上述过程主要以短读长比对 (Short reads alignment) 的方式进行, 即使用 SOAP 2.21 (Li et al., 2008) 将重测序 PE reads 比对到白菜品种 Chiifu-401-42 的基因组参考序列上, 在选择比对上的 reads 结果时需满足以下条件: (1) PE 测序插入片段长度与比对插入片段长度一致, (2) 一条 read 上最多允许 2 个核苷酸的错配且不容许含有 gap, (3) PE reads 必须成对地比对到参考序列上。

小片段 InDels 的预测 (Identification of short indels) 主要是使用 SOAP 2.21 (Li et al., 2008) 将上述比对不到基因组参考序列上的 PE reads 重新比对到基因组序列上。主要条件为: (1) PE 测序插入片段长度与比对插入片段长度一致, (2) 一条 read 上最多允许 2 个核苷酸的错配且至多允许 5 个核苷酸的 gap, (3) PE reads 必须成对的比对到参考序列上。在这种条件下可以得到长度为 1 ~ 5 bp 的短的 InDels。为提高预测小片段 InDels 的准确性, 要求至少 3 条以上的 PE reads 来确定 InDel 的存在, 并且只允许一对 PE reads 中的一个 reads 上存在 InDel, 如果一对 PE reads 上的两个 reads 都存在 InDel 则就不考虑该 PE reads。提取 InDel 位点两侧各 150 bp 序列, 并把该序列比对到白菜基因组参考序列上, 把比对到多个位点的序列过滤掉。利用 Primer 3 (Rozen & Skaletsky, 2000) 设计引物序列, 引物扩增片段长度为 60 ~ 180 bp, 引物长度 19 ~ 22 bp, T_m 值为 52 ~ 58 °C, 由上海捷瑞生物有限公司合成。

1.3 验证 InDel 标记的多态性和 RILs 群体基因型检测

‘R-O-18’ 和 ‘L58’ 用于验证 InDel 标记的多态性。‘R-O-18’、‘L58’ 和 RILs 全基因组 DNA 的提取和纯化采用改良 CTAB 方法 (Wang et al., 2005)。PCR 反应体系: 总体积 20 μL , 1 \times PCR 缓冲液 2 μL (北京天根生化生物科技有限公司), 2.5 U $\cdot \mu\text{L}^{-1}$ 的 *Taq* 酶 0.2 μL , 40 ng $\cdot \mu\text{L}^{-1}$ 的 DNA 模板 5 μL , 正、反向引物 (5 $\mu\text{mol} \cdot \text{L}^{-1}$) 各 0.4 μL , 2.5 mmol $\cdot \text{L}^{-1}$ 的 dNTP 1.6 μL , 其余以 ddH₂O 补足 20 μL 。PCR 程序为: 94 °C 预变性 5 min; 94 °C 变性 40 s, 57 °C 退火 40 s, 72 °C 延伸 40 s, 35 个循环; 72 °C 保温 10 min。扩增产物用 8% 非变性聚丙烯酰胺凝胶 120 V 恒压电泳分离, 银染显色后在观察灯箱上进行数据统计分析。

1.4 遗传连锁图谱的构建

将群体内各株系的带型按亲本类型分类, 与亲本 ‘R-O-18’ 带型相同者赋值为 a, 与亲本 ‘L58’ 带型相同者赋值为 b, 杂合带型者赋值为 h, 由各种原因造成的数据不清或者缺失者赋值为 “-”。标记间的连锁分析及遗传图谱的构建均使用软件 JoinMap 4.0 (<http://www.kyazma.nl>) (van Ooijen, 2006)。LOD 值的范围设定为 “5” 到 “10”, 并选用 Kosambi (1943) 作图功能将重组率转化为遗传距离, 最后绘制出包括 10 个连锁群的图谱。

2 结果与分析

2.1 ‘R-O-18’ 和 ‘L58’ 间的短 InDel 变异位点检测

‘L58’ 重测序产生大约 3 Gb 的原始数据 (3 973 万条 reads), 是由 500 bp 的插入片段库产生的长 75 bp 的 pair-end reads, 其中高质量的 reads 大约为 2 896 万条。另一份重测序样本 ‘R-O-18’, 大约 2.8 Gb (2 896 万条 reads), 是由大约 300 bp 的插入片段库产生的 76 bp 长度的 pair-end reads。利用 SOAP 软件把 ‘L58’ 重测序的原始 reads 比对到白菜基因组参考序列上, 其中 47% 的 pair-end reads 比对到基因组序列上, 并覆盖全基因组长度的 55%。

对于重测序材料 ‘L58’ 与参考序列之间的 InDel 位点的预测, 每个 InDel 位点至少有 3 对 reads, 并且 reads 比对到基因组序列唯一位置上。由于测序 reads 的长度较短, 因此在允许 gap 比对时, 为了降低假阳性, 一条 read 上允许存在的插入缺失至多为 5 bp。通过上述策略, 检测到 ‘L58’ 与参考序列之间的 InDel 多态性位点 (1 ~ 5 bp 长度) 共 18 479 个, 其中大多数 (58%) 是 1 bp 的插入缺失, 19% 为 2 bp 的插入缺失, 仅有 10% 是大于 3 bp 的插入缺失 (表 1)。对于位于白菜参考序列基因区的 InDel 位点, 发现 9% 位于外显子, 36% 位于内含子, 55% 位于基因间区。在位于外显子区的 1 736 个 InDel 多态性位点, 40% 的是非 3 bp 的插入缺失, 这些 InDel 位点很可能导致移码突变并导致基因功能的丧失。

表 1 在全基因组预测的短 InDel 的数量
Table 1 The number of short InDels identified in genome

插入或缺失片段大小/bp Indel size	数量 Count	所占比率/% Proportion
1	10 663	57.7
2	3 442	18.6
3	2 573	13.9
4	1 252	6.8
5	549	3.0
总计 Total	18 479	

对于预测重测序材料 ‘R-O-18’ 与 ‘L58’ 之间的 InDel 多态性位点, 用上述找到的 ‘L58’ 与参考基因组的 InDel 位点作为一个桥梁, 用比对软件 SOAP 把 ‘R-O-18’ 重测序原始 reads 比对到这些 ‘L58’ 与参考基因组之间的 InDel 位点序列上, 如果 ‘R-O-18’ 的序列与基因组参考序列完全相同 (不存在插入缺失), 那么就认为该位点是 ‘L58’ 与 ‘R-O-18’ 之间的 InDel 多态性位点。通过这种方法, 共预测 18 479 个 InDel 多态性位点, 在白菜 10 条染色体上的分布见表 2。

表 2 在各染色体中预测的短片段插入/缺失位点
Table 2 Short InDel polymorphisms identified in each chromosome

染色体 Chromosome	数量 Count	物理距离/Mb Physical distance	频率/ (InDel · Mb ⁻¹) Frequency
A01	1 798	28.6	62.9
A02	1 879	27.9	67.3
A03	2 459	31.7	77.6
A04	1 386	19.0	72.9
A05	1 645	23.9	68.8
A06	1 870	26.3	71.1
A07	1 966	22.6	87.0
A08	1 597	21.6	73.9
A09	2 615	37.1	70.5
A10	1 264	17.6	71.8

2.2 短 InDel 分子标记多态性的试验验证

为了验证短 InDel 变异位点的多态性,选取了其中 500 个均匀分布于基因组的位点,将其转化为基于 PCR 的分子标记。同时,为了便于利用聚丙烯酰胺凝胶电泳检测,只选用插入缺失片段大小为 4 ~ 5 bp 的位点。为了减少非特异性扩增,落在多同源序列区域的 InDel 变异位点被剔除掉。基于上述选择,设计的引物所扩增片段只包含一个 InDel 位点,并且大小控制在 80 ~ 140 bp 之间。分别以 ‘R-O-18’ 和 ‘L58’ 基因组 DNA 为模板,通过 PCR 扩增,在 500 个 InDel 标记中共有 452 个能够得到稳定的扩增,有 148 个标记在二者之间表现出明显的多态性,其中 14 对 InDel 引物在双亲中检测的结果如图 1。

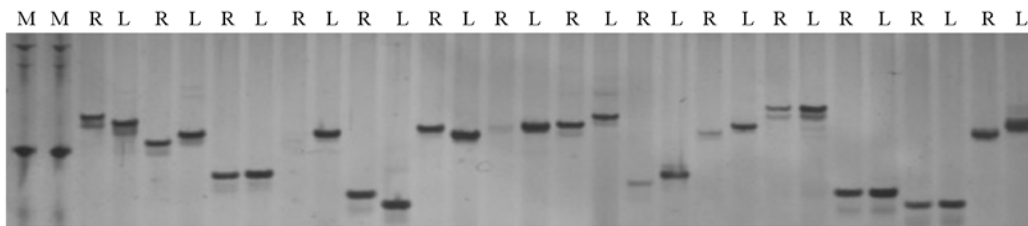


图 1 14 对 InDel 引物在双亲中检测的电泳图

Fig. 1 InDel profiling of 14 primer combinations in two parents

R: R-O-18; L: L58.

对在两亲本间表现出多态性的 106 个标记进一步用 RILs 群体验证,其中在双亲及 30 个 RIL 株系中的检测结果如图 2, 106 个 InDel 标记均表现出多态性,各标记信息见表 3,其中最后 7 个标记未定位到图谱中。

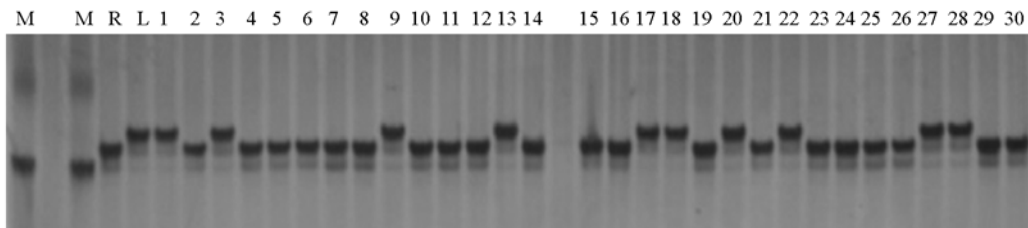


图 2 InDel 标记在双亲及群体的 30 个 RIL 株系中检测的电泳图

M: Marker; R: R-O-18; L: L58; 1 ~ 30: RILs 群体中的 30 株系。

Fig. 2 InDel profiling in two parents and 30 lines of the RILs population

M: Marker; R: R-O-18; L: L58; 1 - 30: 30 lines of the RILs population.

表 3 在 RILs 群体中表现多态性的 106 个 InDel 标记的信息

Table 3 Information of the 106 InDel markers showing polymorphism among the RILs population

标记名称 Marker name	染色体 Chromosome	正向引物序列 Sequence of forward primer	反向引物序列 Sequence of reverse primer
BrID11699	A01	TCGTCAGATAAAGCTCCAAT	GAATCTTGTGCGTCATTTTC
BrID11693	A01	ATGCGCTTACATAATGGACT	CATCTTCTGCCTAAGTTACGA
BrID11059	A01	CCTCTACTAATCACCAGCTA	CATTCTTGAGCAAGTAAGCC
BrID11345	A01	GCTTATGGTGAAGATCATGG	TAGGCTACGGGTAGAGCATA
BrID11385	A01	ATAGGAAGCAGACCATGAAA	TCCGAATCCTAACCATACAG
BrID11337	A01	GCTCATTGAAATGCTAAACC	GCTGACATTGTAGCTCTTCC
BrID11087	A01	ATGGTCATCTCCTACAAACG	CTCACTTTCTCATGCACCTT
BrID11095	A02	TAGGCTCTTGAGCCACATTA	CTTCCATCGTTATCCTCTG
BrID11411	A02	TCTTCTGGAGAAAAGACCGTA	GTCAAATTGCTAAACCGAGA
BrID11415	A02	TAGGATTCATCTCATCGTCC	GAGGACTTCGATATTTGCAT

续表 3

标记名称 Marker name	染色体 Chromosome	正向引物序列 Sequence of forward primer	反向引物序列 Sequence of reverse primer
BrID11907	A02	AGCTACACCGTCTCTATCCA	TCTTCGCTCTTCATTAGCTC
BrID11121	A02	CATGTTCAACCACAGCATCT	GCCCTATGCCCTAAGACTAT
BrID11895	A02	TCGGAGATTCTGGAACCTTA	GCTATTTTGATGCTAAACGG
BrID11113	A02	GTCTCGATTCTTCAGCAAC	CAATCTTCAGAACCGTGACT
BrID11905	A02	CCACAACACCATAACACAAG	TCTGATGATGAAGATGCTGA
BrID11431	A02	CCATACACCGATTCTTGTT	GTCTCTTTTCTTTCCGGTT
BrID11129	A02	CTTAGCTAGAAAGGACCGTG	AGGTGAGTTCATATTCGCAG
BrID11407	A02	TCTGGGAGAGATTTAGACCA	GGAAATCACATGTTTTGGC
BrID11139	A03	CACAACACATGATCGAAAAC	ATAAGATATATGCCGAACCG
BrID11901	A03	CGAACTTGGTTCGTATGTG	TTTGAGGTCTACCAAAATG
BrID11893	A03	CCGTTCCGTATTTCCAGACA	AAGTCTTCACGGTCTCCCA
BrID11853	A03	TGGGTTCTTCTGGAATGCT	TTACCAGGATCTGAAGGAGGA
BrID11973	A03	TGGGGAAAAGAACTGTACT	CGATCTAAAGCAGAGAAAAGTCT
BrID11611	A03	AATAGACCGGTAAAACCTTGC	CCCATTAACTCACTCACACA
BrID11597	A03	TGTTCATATATATTGCGTGTCTATTG	TTTAAGCATACTAAGGAGATGGACA
BrID11663	A03	AATGGACTAGTTGCTGCTTC	AGCCCTTGAAAGTAATAGA
BrID11899	A04	GGTTAGGGAAGAAAGTGAGC	ATCCCTAAACAAAAGCCTTG
BrID11915	A04	GCTTCGGTTATTTTCTCAGT	AACAAACCGAACCAACC
BrID11643	A04	GTATTTTGGGAGGCTTCTTT	AGACTCCAGCATATTGTCTATC
BrID11439	A04	AAGCTTTCCTCGTGTCATTA	CTTCCGAGTTCCCATAGTG
BrID11635	A04	CACAAAGCTCACTTCTTAGGCA	TATAGCCTTGCGGTGGAGTT
BrID11617	A04	GAGAACATCGATCATCCAGTA	CTAGGCTAACTCAAAGTTGGTA
BrID11453	A04	GTTTAACAATGTATGTGGGG	AGAAAATTGACAAGAGCGTC
BrID11911	A04	ACATTGCCTCTAACATCCAC	ACATGCCAGTGAGTAAGAAGA
BrID11035	A04	CTAATAACGTGGCAGCTCTT	GTCGATGTGTTTTTGTTCG
BrID11457	A04	GCATCTTTTGGAAAGTGAGAC	CCCATATGGTTCCAAGAGTA
BrID11177	A05	TGGAGAGCATTGAGGTACA	TGCCTTTTCTAGACACAACC
BrID11321	A05	ACGTTCCCTTCACATGCTTAT	AGAAACGCCAAGAACCTC
BrID11327	A05	GATAAACTTTGATGTTTGC	GACTGATTTTCGATGATCTC
BrID11165	A05	AGTGATGGGAAGAGAAAACA	CATCTCTCGTCTACCCTC
BrID11753	A05	GTTGGTTGGTTGTTTGAATC	TATGGAACACGAACGAGTA
BrID11169	A05	AGAAGATCATCAAAGCGTGT	GATATGCCTCGCTGTTTAAT
BrID11641	A05	GAACAGAAACAGCATTACCC	TTATTGTTCTCTGTTTCGT
BrID11639	A05	GTAATCGAATGGGACCATAA	GCATAATTGAAGTTGAAGGG
BrID11629	A05	CGGACGATATGAATGTAGTG	CCATGTGGGGAAGTAAAAT
BrID11623	A05	AGCTAAAGAGAAAGGGAAGG	GACTTTACCAATTCAGTCG
BrID11181	A06	CAAACACCACAACACAAATC	CTATTTCCCCTTGGCTAGTT
BrID11501	A06	AGCAAGCTAGCAACATTTTC	TATGTTGGAGGAATCAGGAG
BrID11187	A06	AGAAACCTGTGGGTCTACT	CCCCAAAAGAAAAGAGAAG
BrID111041	A06	ACACCATTAACCAAGCATC	GAAACATTTTCCACGTACCA
BrID111037	A06	TCGAAGAAACAGTCTGAACA	GCTCAGGTGACATGTGTTTT
BrID11491	A06	GCATGAGCCCAACATTA	GGAGACTTGTTTGTCTTCAG
BrID11199	A06	GCTTAAAGCCACAGTCAAGT	ACACTTTTTGTTGGAGAGGA
BrID11207	A06	TAGACATACCACCGTATCCC	TGGATATAGTTGGAGTGGTGT
BrID11209	A06	TTGGAGTGATAGGTGAGACC	GATCGAAGAGAACTAGAGC
BrID11217	A06	TATTAGGCTAACAGGAACGG	ACATTGGGAGCTCATAGAGA
BrID11489	A06	AAGTTCCTTGCTTAAGCCTC	GGTTTGCATATACCAAAAGG
BrID11495	A06	ACTACTCCCAACTATTGAG	GCTCTTTGTTTGGGTCTA
BrID11493	A06	GGACCAACCTTTAGATGG	AAAAACCTCGAACACAGTCT
BrID11597	A07	ATAAGGGGACCCTGTAAAAA	TTCTGTCAACTAACATCCCA
BrID11671	A07	AAGGATCTTTGGCGATTGTG	ACCTACCCATCGTCTTCCT
BrID11615	A07	ATGACTGTTAGGAATCCCAT	GCTAATACATGAATCCCTCG
BrID11605	A07	AACATGGAAATGTCGGTAAC	ACCTAAATGCGCTTACTCCT
BrID11601	A07	CGGCTCAAAATGATTAGAAG	GTTACAAAATGAACCACCGT
BrID11593	A07	AGAAGCTAACAAAGACGGTGA	ACAAGAAGAAAAGACCAAGG

续表 3

标记名称 Marker name	染色体 Chromosome	正向引物序列 Sequence of forward primer	反向引物序列 Sequence of reverse primer
BrID11587	A07	TGCTGTTGATCAAGAAAGAC	CTTGCTCTCCTTCACCTTC
BrID111243	A07	CATATCTTTCAGAGGGCTTG	GTCCTTCCAAGACAAGACAG
BrID11667	A07	ATTGAAGCAGTGGTGTGCT	GAAACAAAGCATTGTACAC
BrID11965	A08	GTGAGGCTGAGTTTCTTGAC	TAAGCCCATATGCACTCAG
BrID11975	A08	GATTGGAAGAAGCAAAAGTC	ATATCATGTCTGAACCGAGG
BrID11717	A08	GCTTAATGTATGGAATGGTG	CAGTAACTGCTAACGCCACT
BrID11689	A08	ACTACATATGCTTTTTGCC	CAAAACACCTCCCTTCTACA
BrID11685	A08	TACTTGCGTCAGCCATTAG	TGAGTCTTGTGTCAATGTCG
BrID11239	A08	GATTTCATGATGCTTGGTCT	ACAGAAACAACATTTCTGTC
BrID11235	A08	ATTCAAGACTTGCAAGAAGC	CTCCATCATTCCTCACTCAT
BrID11417	A08	CGATATCGTATAGGACCTTCA	CTTCATTCAATGTGTATGC
BrID11275	A09	TGGTCGACGATTAGACATAA	TCTTGCTCTTCTGCATGTT
BrID11781	A09	AGAATAAATGACACCAGCGT	CATTTTGAGGGTCACTCTCT
BrID111405	A09	GCAAAATGCACAACAAACAC	TTCTATACTGGAGCAATGCC
BrID11757	A09	ACTTTAACAACAGTGAGGCA	ATGGGAGAAAAAGTGTGATG
BrID11247	A09	CTATTTGACCAACAGTCGC	TAAGGGGGAGATTACAGGTT
BrID11285	A09	GGTTTTGTTGAAAGAGTTGC	CAGAGTTCGGTTTTATCAG
BrID111395	A09	TTCCGTTACTCCAATCTGAC	TGAAGATCGGTCTCAAAGTT
BrID11287	A09	CCATTGAGTAAGGTTATGGA	CCTCTTTACACCATTTGCAT
BrID111389	A09	TAAGAGATTGGATGGGCTAA	GCCAGTTCAAAATCAAACCTC
BrID11523	A09	ACCTAAATGGGCTACTCTC	GGTTTAGATTCAACCGTGAC
BrID11867	A09	ACAAATCCAGTTTGTCTCAC	TCAGACAGAAGACAACATGG
BrID11039	A09	AAGAGGTCCAGAGAAACACA	ATCTCATGACCACCTCAAAG
BrID111409	A09	TACTGGACAAGCTTTAGCAA	GGACCAGCTTTTATTCTCTT
BrID11011	A10	CAGCTAAAGCTGAAGATTGC	TGTGAAAGGAAAGTCTCAAC
BrID11563	A10	GTCGACACAACCTTAAACCC	CATTGGCAAGGTAAAGAAGA
BrID11007	A10	ACCATTTCGTGTATTCTTGG	CCTGACTTGCTATCCAAACA
BrID11315	A10	GCTATGACTGTTTTTGGCTT	TACCTAACGCTGAAGTATG
BrID11537	A10	CTGCCTAATACTGTTTCTCCT	ACATAGCGTATATCGGTGGA
BrID11539	A10	TTCAAGAAGAGGAAGAACCA	TTTGGTCTTTCTCCAACCTA
BrID11003	A10	CAATAGATTTATGCAGGGGA	TTACGCTTCCATTGTCTAC
BrID11579	A10	GTTTAAACGAGGCTAAGGGT	GCTGAGAGTTCAGCACAAGT
BrID11027	A10	CCACTTGCTTTGAGTATTCC	GTTTTGGTTGTGTACCTCT
BrID11581	A10	TGGTTCGATGATAATGGC	GGTAAGAGCAAGAAGGGATT
BrID11367	A01	ATTGTGACAAACCAATCTCC	GCGATTTAGGGTTTCGAT
BrID11377	A01	GTTTCATGAAGGATGCTAGG	ATACACACTACGTTCCCTCC
BrID11375	A01	TGTATCTTCGTTTCATCTCAGC	TCATTCACTATCAAGCAACC
BrID11067	A01	AAGGAGATTGGAACCTTGCT	CTACTAGCTGTGCTGTGGAG
BrID11093	A02	TTGCTCTTCAAATTCTCCAG	GACTTGTGCGCCAAACAA
BrID11917	A04	TTATGATGTTCCCTTCATC	TACTTGACGCTGGAAGCTAT
BrID11191	A06	AAACAGACTGAGCCGATTTA	CAAACATCCCATCCTCTCTA

2.3 构建的遗传连锁图谱

利用在两亲本及群体中均表现出多态性的 106 个 InDel 标记构建遗传连锁图, 其中 99 个标记均匀分布于 10 个连锁群 (A1 ~ A10) 上, 得到一张包含 10 个连锁群的分子标记遗传连锁图谱 (图 3), 根据 InDel 标记所在的物理位置和白菜参考图谱 (Wang et al., 2011b) 进行一一对应, 将 10 个连锁群分别锚定在对应的 10 条染色体上。99 个标记中, A1 上 7 个; A2 上 11 个; A3 上 8 个; A4 上 10 个; A5 上 10 个; A6 上 13 个; A7 上 9 个; A8 上 8 个; A9 上 13 个; A10 上 10 个。10 个连锁群跨越基因组遗传距离为 873.7 cM, 密度为每标记 8.8 cM。

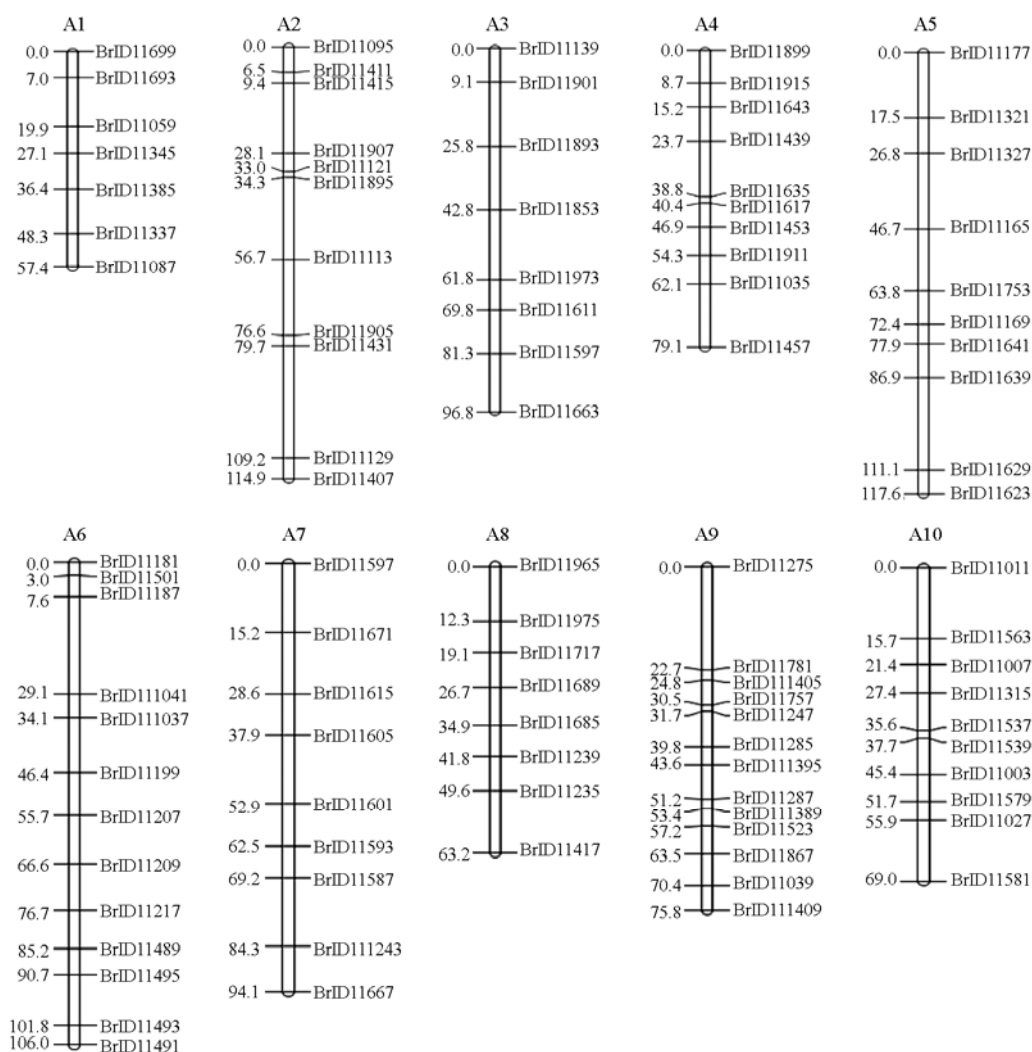


图3 R-O-18 × L58 的 RILs 群体的遗传连锁图谱

图谱左侧为遗传图距 (cM)，右侧为标记名称。

Fig. 3 Genetic linkage map of the R-O-18 × L58 RILs population

Genetic distance in cM is indicated on the left side of linkage groups and loci names on the right.

3 讨论

插入缺失标记是针对基因组中插入缺失变异位点而设计的 PCR 标记。InDel 变异广泛分布于基因组中 (Feltus et al., 2004; 冯芳君 等, 2005; 潘存红 等, 2007), 在人类的 1 000 万个多态性分子标记中, 至少有 150 万个是 InDel 标记 (南海洋 等, 2009)。水稻中每 953 bp 就存在 1 个 InDel 变异位点 (Shen et al., 2004)。本研究利用白菜基因组重测序数据以及基因组参考数据, 在全基因组范围预测了 18 479 个短的 (≤ 5 bp) 插入缺失变异位点, 平均出现频率为 $72.1 \text{ InDels} \cdot \text{Mb}^{-1}$, 远远低于之前 Park 等 (2010) 报道的 $4 830 \text{ InDels} \cdot \text{Mb}^{-1}$ (1 ~ 100 bp 的 InDel), 这一差异可能来自于检测范围的差异, 作者检测的是全基因组范围的短 InDel, 而 Park 等 (2010) 的研究仅检测了覆盖 1 396 498 bp 的 1 398 个 STSs 中的 InDel; 另外, 本研究检测的是小于 5 bp 的 InDel, 没有统计包括大于 5 bp 的 InDel, 因此检测到的 InDel 也低于拟南芥中的 $151 \text{ InDels} \cdot \text{Mb}^{-1}$ (1 ~ 100 bp 的 InDel) (Jander et al., 2002)。将其中的 500 个位点设计成基于 PCR 的 InDel 标记, 其中有 452 个标记可

以在两亲本上得到稳定的扩增。但是只有 106 个标记在两亲本间表现出多态性, 多态性比率仅为 23.5%, 远低于之前报道的 87% (Wang et al., 2011b)。这主要与 ‘R-O-18’ 和 ‘L58’ 的重测序数据有关, 由于 ‘R-O-18’ 数据是从网站上下载的, 数据的质量较低, 另外虽然命名相同, 但由于构建本群体用的材料来源不同, 有可能存在基因型差异。再则, ‘L58’ 重测序的原始 reads 仅覆盖全基因组长度的 55%, 测序深度不高也是造成上述结果的原因。

与其他标记相比, 在基因组不同位置出现同样大小的 InDel 标记的概率极小, 因此避免了由于基因组复杂性及异质性而导致的后续分析的较大模糊性。在基因组测序的基础上, 可以较容易地通过生物信息学的方法开发 InDel 标记, 现已在基因定位 (Schnabel et al., 2005)、图位克隆 (Jander et al., 2002)、分子标记辅助选择育种 (王岩 等, 2009)、植物遗传多样性 (王明军 等, 2010) 等领域被广泛应用。

本研究中所选的 500 个变异位点是依据物理图距均匀挑选的, 所以从构建的遗传连锁图谱可以看出, 99 个标记在 10 个连锁群中呈均匀分布。先前报道的白菜基因组参考图谱包括 507 个标记, 覆盖基因组物理图距为 235 Mb (Wang et al., 2011b)。在本研究结果中仅 99 个标记覆盖基因组的物理图距达 144 Mb, 证实了依据参考序列和基因组重测序数据在确定的基因组区域设计 InDel 标记的可能性, 同时也大大提高了构建高密度遗传连锁图谱的效率。总之, 本研究中开发的 InDel 标记将在白菜遗传图谱构建, 重要农艺性状的基因定位, 以及分子辅助育种等工作中发挥重要作用。

References

- Bhatramakki D, Dolan M, Hanafey M, Wineland R, Vaske D, Register J C 3rd, Tingey S V, Rafalski A. 2002. Insertion-deletion polymorphisms in 3' regions of maize genes occur frequently and can be used as highly informative genetic markers. *Plant Mol Biol*, 48 (5 - 6): 539 - 547.
- Botstein D, White R L, Skolnick M, Davis R W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet*, 32 (3): 314 - 331.
- Feltus F A, Wan J, Schulze S R, Estill J C, Jiang N, Paterson A H. 2004. An SNP resource for rice genetics and breeding based on subspecies *Indica* and *Japonica* genome alignments. *Genome Research*, 14 (9): 1812 - 1819.
- Feng Fang-jun, Luo Li-jun, Li Ying, Zhou Li-guo, Xu Xiao-yan, Wu Jin-hong, Chen Hong-wei, Chen Liang, Mei Han-wei. 2005. Comparative analysis of polymorphism of InDel and SSR markers in rice. *Molecular Plant Breeding*, 3 (5): 725 - 730. (in Chinese)
- 冯芳君, 罗利军, 李 莹, 周立国, 徐小艳, 吴金红, 陈宏伟, 陈 亮, 梅捍卫. 2005. 水稻 InDel 和 SSR 标记多态性的比较分析. *分子植物育种*, 3 (5): 725 - 730.
- Hisano H, Sato S, Isobe S, Sasamoto S, Wada T, Matsuno A, Fujishiro T, Yamada M, Nakayama S, Nakamura Y. 2007. Characterization of the soybean genome using EST-derived microsatellite markers. *DNA Research*, 14 (6): 271.
- Jander G, Norris S R, Rounsley S D, Bush D F, Levin I M, Last R L. 2002. *Arabidopsis* map-based cloning in the post-genome era. *Plant Physiol*, 129 (2): 440 - 450.
- Kim H R, Choi S R, Bae J, Hong C P, Lee S Y, Hossain M J, van Nguyen D, Jin M, Park B S, Bang J W. 2009. Sequenced BAC anchored reference genetic map that reconciles the ten individual chromosomes of *Brassica rapa*. *BMC Genomics*, 10 (1): 432.
- Kosambi D. 1943. The estimation of map distances from recombination values. *Annals of Human Genetics*, 12 (1): 172 - 175.
- Li R, Li Y, Kristiansen K, Wang J. 2008. SOAP: Short oligonucleotide alignment program. *Bioinformatics*, 24 (5): 713 - 714.
- McCouch S R, Chen X, Panaud O, Temnykh S, Xu Y, Cho Y G, Huang N, Ishii T, Blair M. 1997. Microsatellite marker development, mapping and applications in rice genetics and breeding. *Plant Mol Biol*, 35 (1 - 2): 89 - 99.
- Mills R E, Luttig C T, Larkins C E, Beauchamp A, Tsui C, Pittard W S, Devine S E. 2006. An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res*, 16 (9): 1182 - 1190.
- Nagaraju J, Kathirvel M, Kumar R R, Siddiq E, Hasnain S E. 2002. Genetic analysis of traditional and evolved Basmati and non-Basmati rice varieties by using fluorescence-based ISSR-PCR and SSR markers. *Proceedings of the National Academy of Sciences of the United States of America*, 99 (9): 5836.
- Nan Hai-yang, Li Ying-hui, Chang Ru-zhen, Qiu Li-juan. 2009. Development and identification of InDel marker based on candidate gene *rhg1* for

- resistance to soybean cyst nematode (*Heterodera glycines* Ichinohe). *Acta Agronomica Sinica*, 35 (7): 1236 – 1243. (in Chinese)
- 南海洋, 李英慧, 常汝镇, 邱丽娟. 2009. 基于大豆胞囊线虫病抗性候选基因 *rhg1* 的 InDel 标记开发与鉴定. *作物学报*, 35 (7): 1236 – 1243.
- Pan Cun-hong, Wang Zi-bin, Ma Yu-yin, Yin Yue-jun, Zhang Ya-fang, Zuo Shi-min, Chen Zong-xiang, Pan Xue-biao. 2007. InDel and SNP markers and their application in map-based cloning of rice genes. *Chinese Journal of Rice Sciences*, 21 (5): 447 – 453. (in Chinese)
- 潘存红, 王子斌, 马玉银, 殷跃军, 张亚芳, 左士敏, 陈宗祥, 潘学彪. 2007. InDel 和 SNP 标记在水稻图位克隆中的应用. *中国水稻科学*, 21 (5): 447 – 453.
- Park S, Yu H. J, Mun J H, Lee S C. 2010. Genome-wide discovery of DNA polymorphism in *Brassica rapa*. *Mol Genet Genomics*, 283 (2): 135 – 145.
- Powell W, Machray G C, Provan J. 1996. Polymorphism revealed by simple sequence repeats. *Trends in Plant Science*, 1 (7): 215 – 222.
- Ren Y, Zhang Z, Liu J, Staub J E, Han Y, Cheng Z, Li X, Lu J, Miao H, Kang H, Xie B, Gu X, Wang X, Du Y, Jin W, Huang S. 2009. An integrated genetic and cytogenetic map of the cucumber genome. *PLoS One*, 4 (6): e5795.
- Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*, 132: 365 – 386.
- Schnabel R D, Kim J J, Ashwell M S, Sonstegard T S, Van Tassel C P, Connor E E, Taylor J F. 2005. Fine-mapping milk production quantitative trait loci on BTA6: Analysis of the bovine osteopontin gene. *Proceedings of the National Academy of Sciences of the United States of America*, 102 (19): 6896.
- Shen Y J, Jiang H, Jin J P, Zhang Z B, Xi B, He Y Y, Wang G, Wang C, Qian L, Li X, Yu Q B, Liu H J, Chen D H, Gao J H, Huang H, Shi T L, Yang Z N. 2004. Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. *Plant Physiol*, 135 (3): 1198 – 1205.
- Syvanen A C. 2001. Accessing genetic variation: Genotyping single nucleotide polymorphisms. *Nat Rev Genet*, 2 (12): 930 – 942.
- Tautz D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res*, 17 (16): 6463 – 6471.
- van Ooijen J W. 2006. JoinMap 4, software for the calculation of genetic linkage maps in experimental populations. Wageningen, Netherlands: Kyazma B V.
- Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M. 1995. AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res*, 23 (21): 4407 – 4414.
- Wang X, Lou P, Bonnema G, Yang B, He H, Zhang Y, Fang Z. 2005. Linkage mapping of a dominant male sterility gene *Ms-cd1* in *Brassica oleracea*. *Genome*, 48 (5): 848 – 854.
- Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun J H, Bancroft I, Cheng F, Huang S, Li X, Hua W, Freeling M, Pires J C, Paterson A H, Chalhouh B, Wang B, Hayward A, Sharpe A G, Park B S, Weisshaar B, Liu B, Li B, Tong C, Song C, Duran C, Peng C, Geng C, Koh C, Lin C, Edwards D, Mu D, Shen D, Soumpourou E, Li F, Fraser F, Conant G, Lassalle G, King G J, Bonnema G, Tang H, Belcram H, Zhou H, Hirakawa H, Abe H, Guo H, Jin H, Parkin I A, Batley J, Kim J S, Just J, Li J, Xu J, Deng J, Kim J A, Yu J, Meng J, Min J, Poulain J, Hatakeyama K, Wu K, Wang L, Fang L, Trick M, Links M G, Zhao M, Jin M, Ramchiary N, Drou N, Berkman P J, Cai Q, Huang Q, Li R, Tabata S, Cheng S, Zhang S, Sato S, Sun S, Kwon S J, Choi S R, Lee T H, Fan W, Zhao X, Tan X, Xu X, Wang Y, Qiu Y, Yin Y, Li Y, Du Y, Liao Y, Lim Y, Narusaka Y, Wang Z, Li Z, Xiong Z, Zhang Z. 2011a. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet*, 43: 1035 – 1039.
- Wang Y, Sun S, Liu B, Wang H, Deng J, Liao Y, Wang Q, Cheng F, Wang X, Wu J. 2011b. A sequence-based genetic linkage map as a reference for *Brassica rapa* pseudochromosome assembly. *BMC Genomics*, 12 (1): 239.
- Wang Yan, Fu Xin-min, Gao Guan-jun, He Yu-qing. 2009. Improving the grain quality of Minghui 63, a restorer line of rice with good quality of marker-assisted selection. *Molecular Plant Breeding*, 7 (4): 661 – 665. (in Chinese)
- 王 岩, 付新民, 高冠军, 何予卿. 2009. 分子标记辅助选择改良优质水稻恢复系明恢 63 的稻米品质. *分子植物育种*, 7 (4): 661 – 665.
- Wang Ming-jun, Wang Yun-yue, Lu Chun-ming, Yang Hui, Wang Yun-tao. 2010. Characteristic of Yunnan glutinous rice varieties revealed by the *Indica-Japonica* specified Insertion/Deletion (InDel) molecular markers. *Journal of Yunnan Agricultural University*, 25 (3): 333 – 337. (in Chinese)
- 王明军, 王云月, 陆春明, 杨 慧, 王云涛. 2010. 利用籼粳稻特异 InDel 标记分析云南糯稻品种的籼粳特性. *云南农业大学学报: 自然科学版*, 25 (3): 333 – 337.
- Williams J G, Kubelik A R, Livak K J, Rafalski J A, Tingey S V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res*, 18 (22): 6531 – 6535.