

CAPS 与熔解曲线技术相结合的菜薹 SNP 基因分型体系的建立与验证

李平霞^{1,2}, 黄英金¹, 刘博², 梁建丽², 程锋², 武剑², 王晓武^{2,*}

(¹江西农业大学农学院, 南昌 330045; ²中国农业科学院蔬菜花卉研究所, 北京 100081)

摘要: 根据不同长度或不同碱基序列的 DNA 片段的熔解温度 (T_m) 不同的原理, 将酶切扩增多态性序列 (cleaved amplified polymorphic sequence, CAPS) 与熔解曲线 (melting curve, MC) 技术相结合, 建立了 CAPS-MC 技术, 从而实现 SNP 分子标记基因分型。通过菜薹 ‘L58’ 和紫菜薹 ‘ZCT095’ 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, 预测了 719 个 *Hind*III 内切酶识别碱基序列的 SNP 位点, 并将其设计成 CAPS-MC 分子标记, 从中随机选择 10 对标记进行体系建立与优化, 结果显示本体系有灵敏度强、精准性高、快速、高通量、经济实用等优势。

关键词: 菜薹; 酶切扩增多态性序列; 熔解曲线; 单个核苷酸多态性; 高通量基因分型

中图分类号: S 634

文献标志码: A

文章编号: 0513-353X (2013) 10-1916-11

Establishment and Validation of SNP Genotyping System by CAPS Combining Melting Curve in *Brassica campestris*

LI Ping-xia^{1,2}, HUANG Ying-jin¹, LIU Bo², LIANG Jian-li², CHENG Feng², WU Jian², and WANG Xiao-wu^{2,*}

(¹College of Agronomy, Jiangxi Agricultural University, Nanchang 330045, China; ²Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China)

Abstract: Melting temperature (T_m) was different according to the different lengths or nucleotide sequences of DNA fragments. In this report, we established a novel single nucleotide polymorphism (SNP) genotyping system referred as CAPS-MC approach. The CAPS-MC approach combined cleaved amplified polymorphic sequence (CAPS) with melting curve (MC) analysis. By aligning whole genome re-sequencing data from two *Brassica campestris* accessions, ‘L58’ and ‘ZCT095’, to the reference genome sequence of ‘Chiifu-401-42’, a total of 719 *Hind*III endonuclease recognition nucleotide sequence SNP polymorphisms were converted to CAPS-MC markers. We established and optimized CAPS-MC system by randomly selecting 10 pairs of these markers. Our results clearly demonstrated that CAPS-MC approach was sensitive, accurate, rapid and inexpensive. Because the melting curve of 96 samples can be measured in 22 mins once, it is well suited for high-throughput genotyping.

Key words: *Brassica campestris*; cleaved amplified polymorphic sequence (CAPS); melting curve;

收稿日期: 2013-05-11; 修回日期: 2013-08-13

基金项目: 国家重点基础研究发展计划 (‘973’ 计划) 项目 (2012CB113906)

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

single nucleotide polymorphism (SNP); high-throughput genotyping.

白菜在十字花科芸薹属芸薹种 (*Brassica campestris* syn. *rapa*) 作物中占据非常重要的地位。单个核苷酸多态性 (single nucleotide polymorphisms, SNPs) 主要是指由于单个核苷酸转换 (即以 一种嘧啶置换另一种嘧啶或一种嘌呤置换另一种嘌呤) 或颠换 (即嘌呤与嘧啶互换) 而引起 DNA 序列多态性 (陈秋玲 等, 2010)。目前已发展了多种 SNP 基因分型技术: 如特异等位基因 PCR (Myakishev et al., 2001)、*Taq* Man 技术 (Holland et al., 1991)、基因特异性杂交 (Howell et al., 1999)、双色荧光偏振技术 (Gut, 2001; Hsu et al., 2001)、MALDI-TOF 质谱分析 (Ross et al., 1998; Stoerker et al., 2000; Bray et al., 2001) 以及毛细管电泳仪高通量测序分型等。CAPS (cleaved amplified polymorphic sequence, CAPS) 标记也常被用来检测 SNP, 它是一种检测 PCR 片段限制性内切酶多态性的方法, 通常是将 PCR 产物用一组内切酶进行酶切筛选, 直到找到具多态性的内切酶, 然后通过电泳的方法检测片段是否被酶切。白菜 SNP 本身密度高、分布广, 研究过程常以遗传群体为对象, 数量较大。此开发过程需要检测很多内切酶, 成本高, 工作量大, 而且检测过程需要使用自动化程度低, 通量小的电泳技术。熔解曲线 (melting curve) 反应的是随温度升高 DNA 双螺旋结构的降解程度, DNA 片段在加热过程中达到熔解温度 (T_m) 时, 双链分开, 引起插入双链的荧光染料释放, 导致荧光信号丢失 (Ririe et al., 1997), 可通过荧光定量 PCR 仪检测荧光并搜集曲线。高分辨率熔解曲线 (high resolution melting, HRM) 技术是近年来发展起来的一种用于突变扫描和基因分型的新技术, 需用昂贵的饱和荧光染料, 如 LC Green (Wittwer et al., 2003), EvaGreen (Li et al., 2010) 或 Syto 9 (Erali & Wittwer, 2010)。虽然以上各种 SNP 分型方法有着各自的优势, 但因成本昂贵, 或对所用试剂和环境要求较高, 操作繁琐, 容易产生误差等应用不普遍 (Chen et al., 2000; Sawyer et al., 2000; Kwok, 2001; Beatriz et al., 2005; Jiannis, 2009)。

本研究中开发了 CAPS 和熔解曲线技术相结合的 SNP 分型技术。通过菜薹 ‘L58’ 和紫菜薹 ‘ZCT095’ 的基因组重测序数据和参考基因组序列的比较, 利用生物信息学方法从全基因组水平预测限制性内切酶识别序列中 SNP 位点并转化成 CAPS 标记。根据限制性内切酶酶切产生的不同长度序列的熔解温度不同的原理, 加入不饱和荧光染料, 通过荧光定量 PCR 仪搜集熔解曲线并自动化对 SNP 基因分型, 从而建立了 CAPS 熔解曲线技术 (CAPS-Melting Curve, CAPS-MC)。利用菜薹和紫菜薹及其 F_2 群体对所开发的若干标记进行了体系验证。该体系简化了标记开发过程, 可直接选择相应的内切酶, 避免了优化内切酶的繁琐, 是一套快速、稳定, 易自动化操作, 高通量又经济的 SNP 基因分型标准体系。该方法也为构建白菜类蔬菜高密度 SNP 分子标记遗传图谱提供了一种有效方法, 也适用于任何 DNA 片段酶切多态性的检测。

1 材料与方法

1.1 试验材料及基因组序列数据集

以菜薹 (*Brassica campestris* L. ssp. *parachinensis*) ‘L58’ 为父本, 紫菜薹 (*Brassica campestris* L. ssp. *chinensis* var. *purpurea* Bailey) ‘ZCT095’ 为母本, 杂交获得 F_1 后进行自交得到 F_2 代群体, 以其中 93 个株系为供试材料。

白菜基因组测序项目 (The Multi-national *B. rapa* Genome Sequencing Project, BrGSP) 所得的大白菜栽培种 ‘Chiifu-401-42’ 基因组作为芸薹属 A 基因组的参考序列, 其至少覆盖了 98% 的基因区 (Wang et al., 2011)。

从芸薹属数据库 BRAD (<http://brassicadb.org/brad/>) 获得白菜品种 ‘Chiifu-401-42’ 的基因组序列。通过 sequencing-by-synthesis 的方法, 对亲本材料 ‘L58’ 和紫菜薹 ‘ZCT095’ 进行重测序, 分别得到约 3 Gb 原始 reads。

1.2 SNP 预测及 CAPS-MC 标记的引物设计

预测 ‘L58’ 和 ‘ZCT095’ 间 *Hind*III 内切酶识别序列的 SNP 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, 通过菜薹 ‘L58’ 的全基因组重测序数据与参考基因组序列的比对, 提取二者间的 SNP 变异位点; 将紫菜薹 ‘ZCT095’ 的重测序数据与提取出的 SNP 位点比对; ‘ZCT095’ 与参考基因组序列相同的位点被认为是 ‘ZCT095’ 与 ‘L58’ 间具有多态性的位点。

同理, ‘ZCT095’ 重测序数据与参考基因组序列比对, 提取出二者间不同的 SNP 位点; 将 ‘L58’ 的重测序数据与提取出的 SNP 位点比对; ‘L58’ 与参考基因组序列相同的位点也被认为是 ‘L58’ 与 ‘ZCT095’ 间具有多态性的位点。在这些多态性位点处搜索限制性内切酶, 如 *Hind*III 酶切序列 (AAGCTT), 该 6 个碱基的任何单碱基的突变都导致无法被切开, 最后根据突变位点上下游 150 bp 序列利用 Primer 3 软件设计合适引物。

1.3 溶解温度 (T_m) 预测、PCR 及限制性内切酶酶切体系

利用在线软件 (<http://biophysics.idtdna.com/>) 分别预测 PCR 产物以及酶切产物序列的 T_m 。

DNA 提取采用 CTAB 法 (Wang et al., 2005)。PCR 反应体系为 15 μ L, 包括模板 DNA 50 ng、1 \times PCR buffer (天根, Tiangen)、 Mg^{2+} 2.5 mmol \cdot L $^{-1}$ 、dNTP 0.25 mmol \cdot L $^{-1}$ 、引物 0.2 μ mol \cdot L $^{-1}$ 、*Taq* DNA polymerase (天根, Tiangen) 0.05 U \cdot μ L $^{-1}$ 和 ddH $_2$ O。PCR 程序: 94 $^{\circ}$ C 预变性 5 min; 然后 94 $^{\circ}$ C 30 s, 55 $^{\circ}$ C 30 s, 72 $^{\circ}$ C 30 s, 循环 30 次; 72 $^{\circ}$ C 延伸 5 min。

*Hind*III 酶切反应体系 15 μ L, 包含 5 μ L PCR 产物 (\leq 1 μ g), 2.5 U 限制性内切酶 *Hind*III (New England Biolabs, USA), 1 \times NEBuffer 2, 无菌双蒸水补足。37 $^{\circ}$ C 保温 2~4 h。

1.4 溶解曲线分析

溶解曲线分析体系包含 20 \times SYBR Green I 荧光染料 2.5 μ L, 甲酰胺 2.5 μ L, 酶切产物 5 μ L, 无菌水补足至 25 μ L。于 Eppendorf Mastercycler ep realplex4 型荧光定量 PCR 仪器上操作, 以 0.04 $^{\circ}$ C \cdot s $^{-1}$ 从溶解温度 T_m 升温至 95 $^{\circ}$ C 过程中收集溶解曲线, 反应时间约为 22 min, 结束后导出数据。试验所获 T_m 值 (T_m -observed) 取 3 个重复的平均值。

2 结果与分析

2.1 CAPS-MC 标记多态性位点的预测

亲本菜薹 ‘L58’ 重测序数据产生了 3 900 万条 reads, 约 3 Gb 数据, 其中高质量 reads 约占 71.79%。亲本紫菜薹 ‘ZCT095’ 含 2 780 万条 reads, 约 2.7 Gb 数据。利用 SOAP 软件把两亲本重测序数据分别比对到白菜基因组参考序列上, 平均能覆盖全基因组长度的 55%。按照 1.2 中所描述预测 SNP 的方法, 预测了亲本 ‘L58’ 和 ‘ZCT095’ 间具有多态性的位点, 每个 SNP 位点至少有 3 对 reads, 并且 reads 比对到基因组序列唯一位置上。在这些多态位点附近搜索限制性内切酶 *Hind*III 识别序列 (AAGCTT), 共预测到 719 个 *Hind*III 内切酶识别序列的 SNP 位点 (表 1)。

表 1 在各染色体中预测的 *Hind*III 内切酶识别碱基序列的 SNP 位点Table 1 SNP polymorphisms of *Hind*III recognition sequence identified in each chromosome

染色体 Chromosome	数量 Count	染色体 Chromosome	数量 Count
A01	65	A07	63
A02	71	A08	51
A03	92	A09	98
A04	55	A10	41
A05	73	组装序列片段 Scaffold	37
A06	73	共计 Total	719

2.2 CAPS-MC 标记多态性试验验证

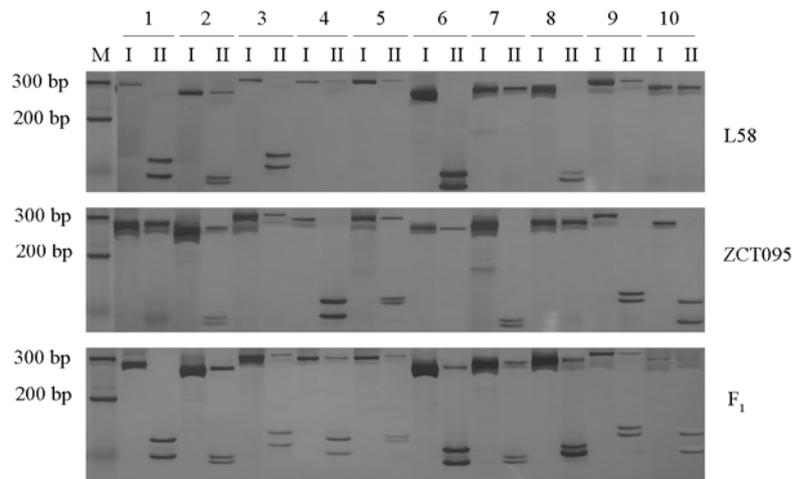
将所预测到的 SNP 位点, 设计成 CAPS-MC 分子标记并进行试验验证。从所有标记结果中随机选择 10 对 (表 2), 在亲本 ‘ZCT095’、‘L58’ 及其杂交一代 (F_1) 材料中进行标记验证, 先用标记引物扩增目标序列, 再进行酶切, PCR 产物大小和酶切产物大小见表 2。

表 2 试验所用 10 对 CAPS-MC 标记名称及引物序列

Table 2 CAPS-MC markers and primers sequences used in the study

编号 Number	引物名称 Primer name	产物大小/bp Product size	正向引物序列 Forward primer sequence	反向引物序列 Reverse primer sequence	酶切产物大小/bp Enzyme-digested products size
1	BrHCAP001	277	TGAATTGCAAGTATGCTCAC	CCATCTCCAAGTCTTTTA	129; 148
2	BrHCAP002	258	CAACTCAATTGCATACACAGG	TACTTTTGGCGCTGAGAT	120; 138
3	BrHCAP003	291	CGATCGTTTTAGCACAACTT	CTCACTTATGCGAGGGAC	139; 152
4	BrHCAP004	284	TCAGTCATGAATCAAAGCA	GATCTGGTCCCATCTAIT	134; 150
5	BrHCAP005	295	TCTACGATCTTTCTTCTGGC	TACCTGACTTTTGACATC	145; 150
6	BrHCAP006	263	CCGAGACTCATGAGGTAAA	CCCAGTTTTCACATCAG	124; 139
7	BrHCAP007	250	TCAGATCAGATTGACTGCAA	GCATGTGAAATTGGCATC	130; 120
8	BrHCAP008	278	GGCAAGGAATGAGAAATTA	GGTCTTCGTTAGAGATTG	126; 152
9	BrHCAP009	299	TCACATACCTAACCAAAC	GAAGTTCTCAAAACCGTC	139; 160
10	BrHCAP010	261	TCAGATTCTTAGCATACCCA	GCCATCGTCTCGATATCTT	128; 133

聚丙烯酰胺凝胶电泳结果 (图 1) 显示, 7 对标记 (BrHCAP003、BrHCAP004、BrHCAP006、

图 1 亲本材料 L58 和 ZCT095 及 F_1 中 10 对标记的 PCR 产物和酶切后产物电泳图

1~10 分别为表 2 中 BrHCAP001~BrHCAP010 共 10 对标记; I 为 PCR 产物; II 为酶切产物; M 为 DNA 标记 I。

Fig. 1 Electrophoresis of PCR products and enzyme-digested products amplified by the markers of BrHCAP001-BrHCAP010 respectively

1 - 10 represent BrHCAP001 - BrHCAP010 marker listed in Table 2; I: PCR products; II: Enzyme-digested products; M: DNA marker I.

BrHCAP007、BrHCAP008、BrHCAP009、BrHCAP010) 能有效区分两种纯合体及杂合体间基因型差异, 而 2 对 (BrHCAP001、BrHCAP005) 标记位点处 F_1 带型分别与亲本 ‘L58’, ‘ZCT095’ 一致, BrHCAP002 在亲本及 F_1 中均是杂合子, 说明 BrHCAP001、BrHCAP002、BrHCAP005 可能为假阳性标记。

2.3 CAPS-MC 标记基因分型体系的建立

结合图 1 结果, 随机选出 4 对能有效区分 3 种基因型 (两种亲本和杂合子基因型) 的 CAPS-MC 标记: BrHCAP004、BrHCAP007、BrHCAP008、BrHCAP010, 分别代表不同位点的 SNP。按照方法中所述的 PCR、酶切、溶解曲线分析等步骤, 上机后获得相应的溶解曲线峰值图 (图 2, 图 3)。

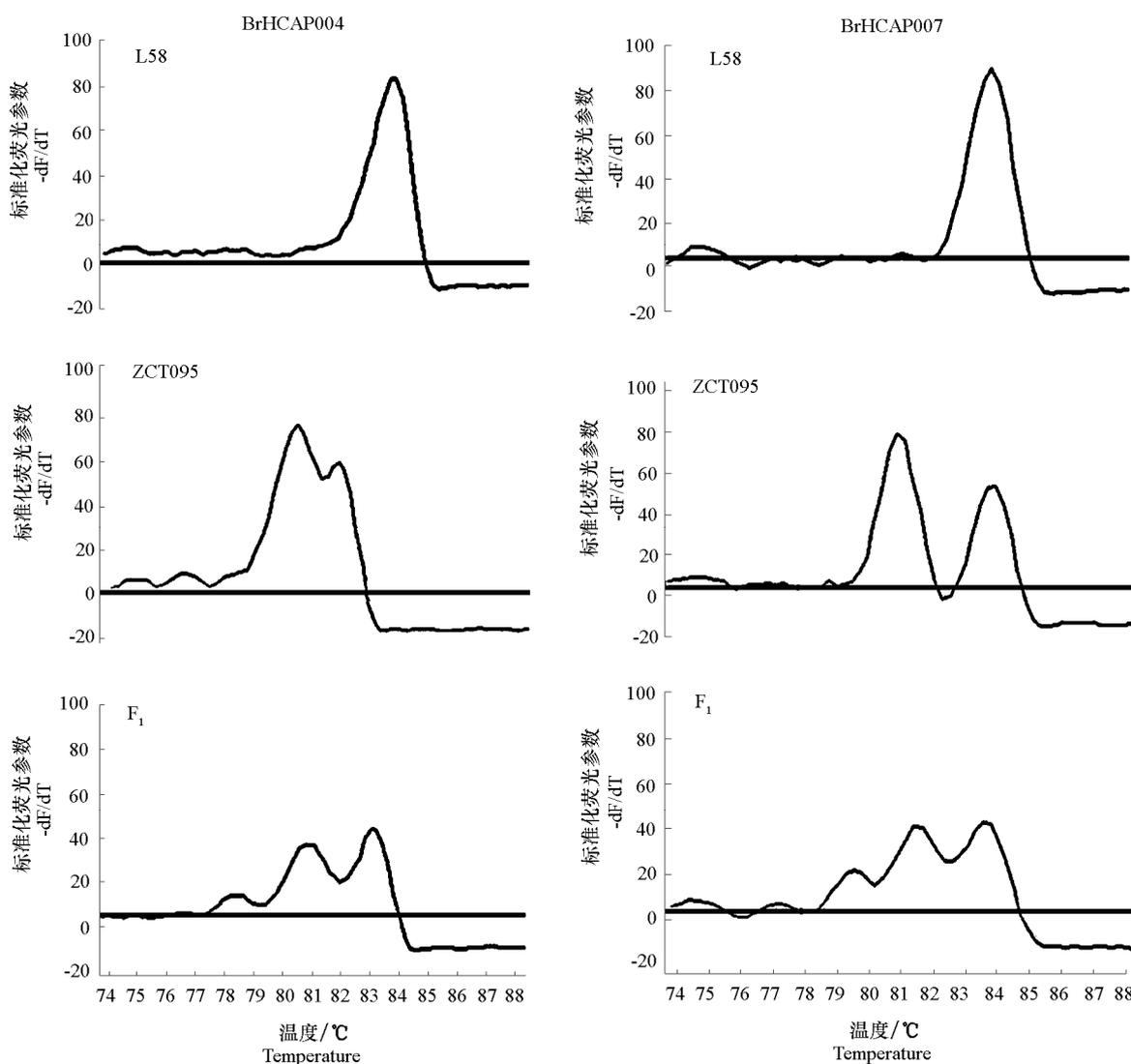


图 2 菜薹 ‘L58’、紫菜薹 ‘ZCT095’ 及其 F_1 SNPs 基因座上两对 CAPS-MC 标记 (BrHCAP004、BrHCAP007) 溶解曲线图

Fig. 2 Melting curve profiles of four CAPS-MC markers (BrHCAP004, BrHCAP007) of SNPs locus in ‘L58’, ‘ZCT095’ and F_1

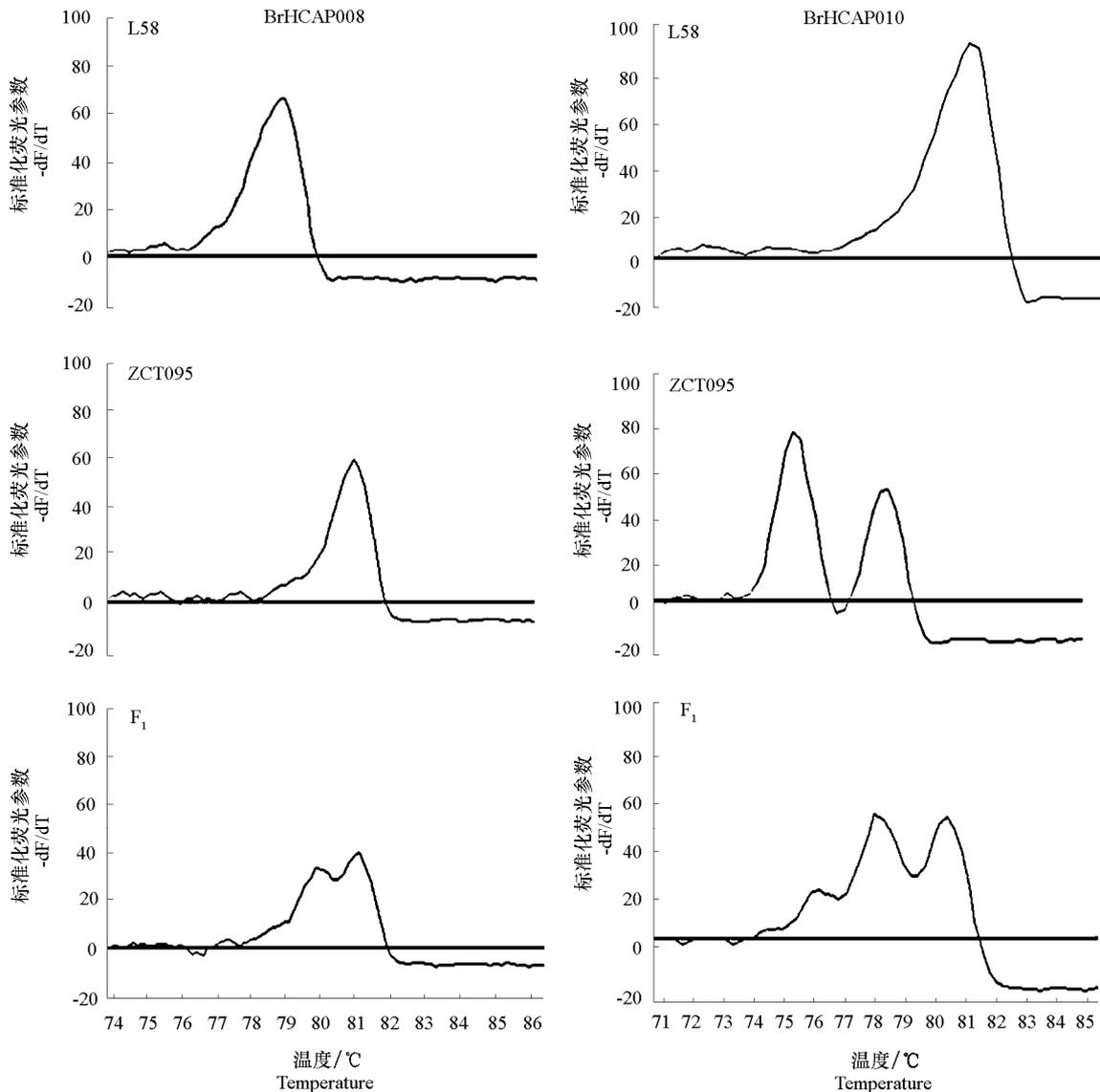


图3 菜蓿‘L58’、紫菜蓿‘ZCT095’及其F₁SNPs基因座上两对CAPS-MC标记(BrHCAP008、BrHCAP010)熔解曲线图

Fig. 3 Melting curve profiles of four CAPS-MC markers (BrHCAP008, BrHCAP010) of SNPs locus in ‘L58’, ‘ZCT095’ and F₁

通过逐渐增加温度同时监测每一步的荧光信号来产生熔解曲线。随着反应中双链DNA变性,荧光染料又回复到游离状态导致荧光信号降低,用荧光信号改变的负的一次导数(即标准化荧光参数,Y轴)与温度(X轴)作图,从熔解峰个数以及 T_m 值可以有效区分两种纯合子和杂合子基因型。

熔解峰个数上看,BrHCAP004、BrHCAP007、BrHCAP010标记中,亲本‘L58’表现为酶切不开的纯合子基因型,熔解曲线只含1个峰,亲本‘ZCT095’表现为可被酶切开的纯合子基因型,熔解曲线有2个峰,杂合子F₁都含3个峰。但是标记BrHCAP008中,两个亲本均只有1个熔解峰,‘ZCT095’的 $T_m = 81\text{ }^\circ\text{C}$,明显高于‘L58’;‘L58’为纯合可切得开基因型(图1),酶切开后两小片段 T_m 相近, T_m 均为 $78.8\text{ }^\circ\text{C}$,相应杂合子F₁只有两个峰(图2,图3)。

溶解温度 T_m 上看, 纯合不能被酶切的基因型 T_m 要高于纯合被酶切后的两个小片段, 如图 2, 图 3 中 BrHCAP004、BrHCAP010 标记在亲本 ‘L58’ 中的 T_m 值要均高于亲本 ‘ZCT095’, BrHCAP008 标记中亲本 ‘ZCT095’ 的 T_m 值要高于亲 ‘L58’。相应地杂合子 F_1 溶解峰中最高温度对应的是不能被酶切开的基因片段, 另外两个偏低溶解温度表示酶切开后相应的两个基因片段。但 BrHCAP007 中 ‘L58’ 为纯合不能被酶切开, ‘ZCT095’ 为纯合能被切开 (图 1), ‘L58’ 溶解峰 T_m 与 ‘ZCT095’ 第 2 个溶解峰 T_m 相似 (图 2), 这是由于被酶切开后片段 T_m 与未被酶切片段 T_m 值差异太小, 可通过优化引物使得片段间 T_m 差异变大。

2.4 自动化基因分型及琼脂糖凝胶电泳验证

采用优化好的 CAPS-MC 标记基因分型体系, 从亲本、 F_1 及 93 个 F_2 株系中进行筛选 BrHCAP010 标记的多态性。结合溶解曲线峰型及自动生成的 T_m 值结果, 可以分为 3 种曲线峰型 (图 4): A 型只出现一个溶解峰, T_m 均值为 $80.8\text{ }^{\circ}\text{C}$ (范围 $80.5\sim 81.2\text{ }^{\circ}\text{C}$), 出现该峰型的材料有亲本 ‘L58’ 以及 F_2 中 25 份株系 (图 4, A); B 型含有两个溶解峰, T_m 均值分别为 $76\text{ }^{\circ}\text{C}$ (范围 $75.8\sim 76.2\text{ }^{\circ}\text{C}$) 和 $78.5\text{ }^{\circ}\text{C}$ (范围 $78.3\sim 78.9\text{ }^{\circ}\text{C}$), 出现该峰型的材料有亲本 ‘ZCT095’ 和 F_2 中 18 份株系 (图 4, B); C 型呈现 3 个溶解峰, 属于杂合基因型, T_m 均值分别为 $76\text{ }^{\circ}\text{C}$ (范围 $75.6\sim 76.4\text{ }^{\circ}\text{C}$)、 $78.5\text{ }^{\circ}\text{C}$ (范围 $78.4\sim 78.7\text{ }^{\circ}\text{C}$) 和 $80.8\text{ }^{\circ}\text{C}$ (范围 $80.6\sim 81.3\text{ }^{\circ}\text{C}$), F_1 以及 F_2 中 50 份株系呈现出该峰型 (图 4, C)。

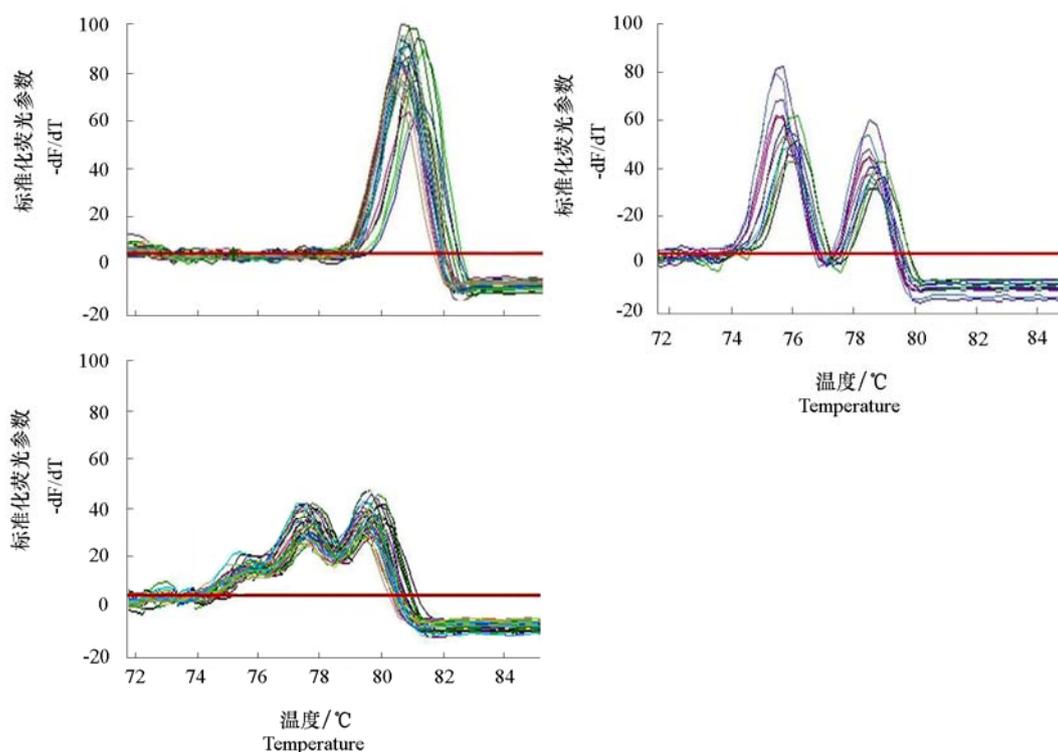


图 4 亲本、 F_1 以及 F_2 群体共 96 份株系中 BrHCAP010 标记溶解曲线图

每条曲线代表每个单株。A: 纯合不能被酶切基因型; B: 纯合可被酶切基因型; C: 杂合子。

Fig. 4 Melting curve profiles of BrHCAP010 marker in 96 accessions including two parents and F_2 population

Each of the curves stands for individual plant. A: Homozygous undigested genotypes;

B: Homozygous digested; C: Heterozygous.

采用琼脂糖凝胶电泳检测的方法验证上述 SNP 变异位点多态性, 96 份株系的酶切产物凝胶电泳部分结果如图 5。结合凝胶电泳和 CAPS-MC 标记基因分型体系的检测结果发现: 呈现出图 5 中 A 带型 (261 bp) 的株系全表现为只有 1 个熔解峰, 与图 4 中 A 峰型的株系编号相吻合; 呈现出图 5 中 B 带型 (133 bp) 株系均表现为图 4 中 B 峰型, 含两个熔解峰, 且株系编号相吻合; 表现出杂合基因型即图 5 中 C 带型的株系均是呈现 3 个熔解峰, 与图 4 中 C 峰型的株系编号相吻合。

由此可见, CAPS-MC 标记基因分型体系检测效果与凝胶电泳检测结果是一致的。

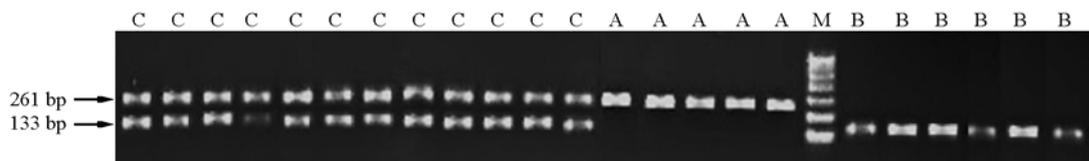


图 5 部分酶切产物凝胶电泳图

A: 纯合不能被 *Hind*III 酶切基因型; B: 纯合可被 *Hind*III 酶切基因型; C: 杂合子; M: Marker I。

Fig. 5 Electrophoresis of digested PCR products amplified by the marker BrHCAP010

A: Homozygous undigested genotypes; B: Homozygous digested; C: Heterozygous; M: Marker I.

2.5 T_m 值

已有研究报道了 DNA 熔解温度预测方法 (Turner, 1996)。本研究中预测酶切前后片段 T_m 值是非常必要的, 有助于自动化基因分型。表 3 将试验中熔解曲线实际所获 T_m 值 (T_m -observed) 与在线软件 (<http://biophysics.idtdna.com/>) 预测 T_m 值 (T_m -predicted) 进行比较。结果显示 (图 6) T_m -observed 与 T_m -predicted 有着高度的相关性 ($R^2 = 0.8079$)。

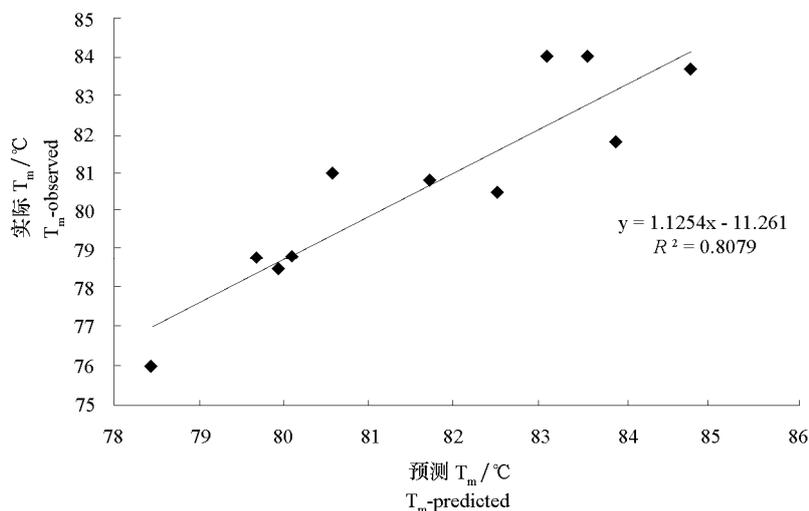


图 6 DNA 片段熔解曲线所获 T_m 值与软件预测 T_m 值相关性分析

Fig. 6 Relationship between observed and predicted T_m values for the DNA fragments

表 3 中, 标记 BrHCAP007 的 PCR 片段大小 250 bp, GC 含量 37.6%, 预测 T_m 为 83.54 °C; 酶切后片段 I 大小 130 bp, GC 含量 44%, 预测 T_m 为 83.07 °C; 两者片段大小相差 120 bp, GC 含量相差 6.4%, 预测 T_m 值相差 0.47 °C; 熔解曲线实际获得 T_m 均为 84.0 °C。标记 BrHCAP008 酶切后

两个片段大小相差 26 bp, GC 含量相差 3.9%, 预测 T_m 相差 0.42 °C, 实际 T_m 无差异(均为 78.8 °C)。因此, 本研究建议 PCR 片段及酶切后片段的 T_m 值, 任何两个 T_m 相差约 1 °C 或 1 °C 以上, 如标记 BrHCAP004、BrHCAP010, 有利于自动化分型, T_m 值相差越大, 效果越好。

表 3 *Hind*III 酶切前后片段大小及 T_m 值的比较

Table 3 Comparison of fragment size and T_m between digested and undigested by *Hind*III enzyme

标记编号 Marker ID	片段类型 Fragment type	大小/bp Size	预测 T_m /°C Predicted T_m		实际 T_m /°C Observed T_m	
			T_m	ΔT_m	T_m	ΔT_m
BrHCAP004	P	284	84.74	ΔT_m (P-I) = 0.87	83.7	ΔT_m (P-I) = 1.90
	I	134	83.87	ΔT_m (P-II) = 2.26	81.8	ΔT_m (P-II) = 3.20
	II	150	82.48	ΔT_m (I-II) = 1.39	80.5	ΔT_m (I-II) = 1.30
BrHCAP007	P	250	83.54	ΔT_m (P-I) = 0.47	84.0	ΔT_m (P-I) = 0
	I	130	83.07	ΔT_m (P-II) = 2.98	84.0	ΔT_m (P-II) = 3.00
	II	120	80.56	ΔT_m (I-II) = 2.51	81.0	ΔT_m (I-II) = 3.00
BrHCAP008	P	278	82.50	ΔT_m (P-I) = 2.42	80.5	ΔT_m (P-I) = 1.70
	I	152	80.08	ΔT_m (P-II) = 2.84	78.8	ΔT_m (P-II) = 1.70
	II	126	79.66	ΔT_m (I-II) = 0.42	78.8	ΔT_m (I-II) = 0
BrHCAP010	P	261	81.69	ΔT_m (P-I) = 1.76	80.8	ΔT_m (P-I) = 2.30
	I	128	79.93	ΔT_m (P-II) = 3.25	78.5	ΔT_m (P-II) = 4.80
	II	133	78.44	ΔT_m (I-II) = 1.49	76.0	ΔT_m (I-II) = 2.50

注: P 指 PCR 片段; I 指 *Hind*III 酶切后片段 I; II 指 *Hind*III 酶切后片段 II。

Note: P refers to PCR product; I, II refer to DNA fragment that result for digested alleles by restriction enzyme digestion, respectively.

常见的 DNA 变性剂有甲酰胺、尿素、二甲基亚砜 (Dimethyl sulfoxide, DMSO) 等。在本研究预试验中分别加入了该 3 种 DNA 变性剂, 比较它们对溶解曲线的影响。结果发现三者均能促使 DNA 解链, 降低 T_m , 但甲酰胺促使解链过程最为稳定, 溶解曲线峰型最为尖锐, 而尿素和 DMSO 所形成的溶解曲线峰型宽而低, 降低了样品的荧光丰度。因此本研究体系选择了甲酰胺作为变性剂。此外, 对于甲酰胺浓度, 也进行了梯度试验, 发现甲酰胺浓度为体系的 10% 效果最佳。

3 讨论

SNP 标记作为目前最具发展潜力的分子标记, 因其遗传稳定性高, 在人以及动植物基因组中广泛分布, 数量巨多, 已被广泛运用到医学, 生物, 农学, 生物进化等众多领域 (Anthony, 1999; Kwok & Gu, 1999)。传统的 CAPS 标记技术, 常被用于检测 SNP, 是一种结合 PCR 与限制性内切酶酶切产生的分子标记技术, 具有共显性、位点特异性等特点, 广泛被用于植物基因的分型、定位、图位克隆和分子鉴定等研究 (Weiland & Yu, 2003; Spaniolas et al., 2006; Palomino et al., 2009)。但是, 此开发过程需要检测很多内切酶, 成本高, 工作量大, 而且需要通过电泳技术进行基因分型, 操作繁琐、通量低, 因此 CAPS 标记难以被大规模的开发应用。

本研究中将 CAPS 和溶解曲线技术相结合, 建立了 CAPS-MC 技术, 从而实现 SNP 分子标记基因分型。CAPS-MC 技术主要分为几个步骤: 利用重测序数据及参考基因组通过生物信息学方法预测限制性内切酶识别序列中含 SNP 的位点并设计成 CAPS-MC 标记; 利用软件分别预测 PCR 产物以及酶切产物碱基序列 T_m ; PCR 扩增; 限制性内切酶酶切; 收集溶解曲线; 自动化基因分型。由于溶解温度 T_m 受到多种因素的影响主要包括 DNA 分子的组成, 如 DNA 分子的 4 个碱基的排列顺序和其中的 GC/AT 含量; 生化反应条件如反应物浓度、溶液的 pH 值、 Na^+ 浓度; 变性剂等。CAPS-MC 技术进行基因分型的先决条件就是 PCR 扩增片段和酶切后片段之间须存在可明确分辨的 T_m , 这就

决定 CAPS-MC 技术应用前应进行 T_m 值的预测。Umelt 软件 (Dwight et al., 2011) 以及在线软件 (<http://biophysics.idtdna.com>) 可以对设计完引物后的 PCR 产物及酶切后产物片段分别进行 T_m 值评估, 熔解温度 (T_m) 值之差 (ΔT_m) 越大则不同基因型间熔解曲线差异越明显, 基因分型越容易; 其同型样本间变异系数 (CV) 越小则熔解曲线峰漂移越小, 标准化曲线拟合越均一, 软件自动分型率越高。此外, SYBR Green I 非特异嵌合于 DNA 双螺旋结构中的小沟内, 结合状态的荧光强度较游离状态增强千倍, 由于它可以和任何 DNA 双螺旋结合 (如引物二聚体), 使试验可能产生假阳性。因此扩增反应完成后可进行熔解曲线的测定, 以判定是否有非特异性产物并据此对引物进行优化。

本研究是利用生物信息学方法基于全基因组重测序数据开发的 CAPS-MC 标记, 能够高效、精确挖掘出 SNP 等变异位点是非常重要的。对于一些经历了高度基因组复制过程的物种, 如玉米、小麦、大豆、白菜等作物, 鉴定 SNP 位点工作更为复杂 (Swaminathan et al., 2007; Wang et al., 2011)。本试验结果中 BrHCAP001、BrHCAP002、BrHCAP005 标记不能够有效区分两种纯合体及杂合体间基因型差异, 可能为假阳性标记, 即表示真实不存在而在生物信息学分析过程中算法找到的 SNP, 这与测序深度和覆盖度有关。本研究中的亲本重测序数据仅覆盖了全基因组总长度的 55%, 此外, 亲本重测序深度为 5 重, 这些都可能导致假阳性标记。相似的, van Orsouw 等 (2007) 对玉米基因组进行了重测序, 75% 预测的 SNPs 位点能够被试验证实存在。

本研究中采用菜薹和紫菜薹为亲本, 杂交 F_1 及其 F_2 群体为供试材料, 取得理想效果, 同样在其他白菜类作物中也有着广泛的应用前景。优势在于: 针对遗传群体亲本重测序数据和白菜参考基因组序列, 从全基因组水平利用生物信息学方法预测限制性内切酶识别序列中 SNP 位点并转化成 CAPS-MC 标记, 不仅挖掘了大量的 SNP 位点, 而且简化了标记开发过程, 大大缩短了标记开发的时间, 降低了标记开发的成本, 并可以直接选择相应的内切酶, 避免了优化内切酶种类的工作。此外, 检测体系灵敏度强、精准性高; 快速、高通量, 从 PCR (55 min)、限制性内切酶酶切 (4 h) 及熔解曲线搜集到基因分型 (22 min) 总计约为共 5.5 h, 一次可检测 96 个样品。本检测体系无需凝胶电泳, 操作过程对人体无害; 采用价格低廉的 SYBR Green I 不饱和染料, 节约了大量的成本; 对仪器设备要求不苛刻, 只需普通荧光定量 PCR 仪便可以进行分析。

References

- Anthony J Brookes. 1999. The essence of SNPs. *Gene*, 234: 177 - 186.
- Beatriz Sobrino, MaríaBrión, Angel Carracedo. 2005. SNPs in forensic genetics: A review on SNP typing methodologies. *Forensic Science International*, 154: 181 - 194.
- Bray M S, Boerwinkle E, Doris P A. 2001. High-throughput multiplex SNP genotyping with MALDI-TOF mass spectrometry: Practice, problems and promise. *Hum Mutat*, 17: 296 - 304.
- Chen J W, Iannone M A, Li M S, Taylor J D, Rivers P, Nelsen A J, Slentz-Kesler K A, Roses A, Weiner M P. 2000. A microsphere-based assay for multiplexed single nucleotide polymorphism analysis using single base chain extension. *Genome Res*, 10 (4): 549 - 557.
- Chen Qiu-ling, Gao Jian-ming, Luo Feng. 2010. Research and development of molecular marker technologies for gene mapping of gramineous crops. *Chinese Agricultural Science Bulletin*, 26 (9): 42 - 48. (in Chinese)
- 陈秋玲, 高建明, 罗峰. 2010. 分子标记技术在禾本科作物基因定位上的研究进展. *中国农学通报*, 26 (9): 42 - 48.
- Dwight Z, Palais R, Wittwer C T. 2011. UMELT: prediction of high-resolution melting curves and dynamic melting profiles of PCR products in a rich web application. *Bioinformatics*, 27 (7): 1019 - 1020.
- Erali M, Wittwer C T. 2010. High resolution melting analysis for gene scanning. *Methods*, 50 (4): 250 - 261.
- Gut Ivo G. 2001. Automation in genotyping of single nucleotide polymorphisms. *Hum Mutat*, 17 (6): 475 - 492.
- Holland P M, Abramson R D, Watson R, Gelfand D H. 1991. Detection of specific polymerase chain reaction product by utilizing the 5'→3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc Natl Sci USA*, 88: 7276 - 7280.

- Howell W M, Jobs M, Gyllensten U, Brookes A J. 1999. Dynamic allele-specific hybridization. A new method for scoring single nucleotide polymorphisms. *Nat Biotechnol*, 17: 87 - 88.
- Hsu Tony M, Law Scott M, Duan Shenghui, Neri Bruce P, Kwok Pui-Yan. 2001. Genotyping single-nucleotide polymorphisms by the invader assay with dual-color fluorescence polarization detection. *Clinical Chemistry*, 47 (8): 1373 - 1377.
- Jiannis Ragoussis. 2009. Genotyping technologies for genetic research. *Annu Rev Genomics Hum Genet*, 10: 117 - 133.
- Kwok Pui-yan, Gu Zhijie. 1999. Single nucleotide polymorphism libraries: Why and how are we building them? *Trends in Mol Med*, 5: 538 - 543.
- Kwok Pui-yan. 2001. Methods for genotyping single nucleotide polymorphisms. *Annu Rev Genomics Hum Genet*, 2: 235 - 258.
- Li Y D, Chu Z Z, Liu X G, Jing H C, Liu Y G, Hao D Y. 2010. A Cost-effective high-resolution melting approach using the EvaGreen dye for DNA polymorphism detection and genotyping in plants. *Journal of Integrative Plant Biology*, 52 (12): 1036 - 1042.
- Myakishev M V, Khripin Y, Hu S, Hamer D H. 2001. High-throughput SNP genotyping by allele-specific PCR with universal energy-transfer-labeled primers. *Genome Research*, 11 (1): 163 - 169.
- Palomino C, Fernández-Romero M, Rubio J, Torres A, Moreno M, Millán T. 2009. Integration of new CAPS and dCAPS-RGA markers into a composite chickpea genetic map and their association with disease resistance. *Theor Appl Genet*, 118: 671 - 682.
- Ririe Kirk M, Rasmussen Randy P, Wittwer Carl T. 1997. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal Biochem*, 245: 145 - 160.
- Ross P, Hall L, Smirnov I, Haft L. 1998. High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nat Biotechnol*, 16: 1347 - 1351.
- Sascha Sauer, Doris Lechner, Kurt Berlin, Hans Lehrach, Jean-Louis Escary, Nick Fox, Ivo Glynne Gut. 2000. A novel procedure for efficient genotyping of single nucleotide polymorphisms. *Nucleic Acid Res*, 28 (5): e13.
- Spaniolas S, May S T, Bennett M J, Toker G A. 2006. Authentication of coffee by means of PCR-RFLP analysis and lab-on-a chip capillary electrophoresis. *J Agric Food Chem*, 54: 7466 - 7470.
- Stoerker J, Mayo J D, Tetzlaff C N, Sarracino D A, Schwoppe I, Richert C. 2000. Rapid genotyping by MALDI-monitored nuclease selection from probe libraries. *Nat Biotechnol*, 18: 1213 - 1216.
- Swaminathan K, Varala K, Hudson M E. 2007. Global repeat discovery and estimation of genomic copy number in a large, complex genome using a high-throughput 454 sequence survey. *BMC Genomics*, 8: 1471 - 2164.
- Turner D H. 1996. Thermodynamics of base pairing. *Curr Opin Struct Biol*, 6: 299 - 304.
- van Orsouw N J, Hogers R C J, Janssen A, Yalcin F, Snoeijsers S, Verstege E, Schneiders H, van der Poel Hein, van Oeveren J, Verstege H, van Eijk M J T. 2007. Complexity reduction of polymorphic sequences (CRoPS): A novel approach for large-scale polymorphism discovery in complex genomes. *PLoS One*, 2 (11): 1 - 15.
- Wang X W, Lou P, Bonnema G, Yang B J, He H J, Zhang Y G, Fang Z Y. 2005. Linkage mapping of a dominant male sterility gene *Ms-cd1* in *Brassica oleracea*. *Genome*, 48: 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, Chalhoub 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. 2011. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet*, 43: 1035 - 1039.
- Weiland J J, Yu M H. 2003. A cleaved amplified polymorphic sequence (CAPS) marker associated with root-knot nematode resistance in sugarbeet. *Crop Sci*, 43: 1814 - 1818.
- Wittwer C T, Reed G H, Gundry C N, Vandersteen J G, Pryor R J. 2003. High-resolution genotyping by amplicon melting analysis using LC Green. *Clinical Chemistry*, 49 (6): 853 - 860.