

研究报告

Research Report

扩增子测序结合高分辨率溶解曲线鉴定花生单核苷酸多态



李杏瑜^{1,2,3} , 刘颖^{1,2,3} , 朱方何^{1,2,3} , 洪彦彬^{1,2,3} , 刘洪⁴ , 陈小平^{1,2,3} , 李海芬^{1,2,3} , 梁炫强^{1,2,3} 

1 广东省农业科学院作物研究所, 广州, 510640

2 国家油料作物改良中心南方花生分中心, 广州, 510640

3 广东省农作物遗传改良重点实验室, 广州, 510640

4 华南农业大学农学院, 广州, 510642

 通讯作者, liang804@yahoo.com;  作者

豆科基因组学与遗传学, 2015 年, 第 6 卷, 第 6 篇 doi: 10.5376/lgg.cn.2015.06.0006

收稿日期: 2015 年 11 月 04 日 接受日期: 2015 年 11 月 04 日 发表日期: 2015 年 11 月 04 日


© 2015 BioPublisher 生命科学中文期刊出版平台

本文首次发表在《分子植物育种》2015 年, 第 13 卷, 第 9 期上。现依据版权所有人授权的许可协议, 采用 [Creative Commons Attribution License](#) 协议对其进行授权, 再次发表与传播。只要对原作有恰当的引用, 版权所有人允许并同意第三方无条件的使用与传播。

摘要 异源四倍体花生(*Arachis hypogaea* L.) 包含两套基因组, 分别来自二倍体祖先 *A. duranensis* (A 基因组) 和 *A. ipaensis* (B 基因组)。相对于二倍体, 异源四倍体 SNP 的鉴定和分析面临更多挑战, 因为在 SNP 的鉴定和分析过程中, 通常需要同时分析两套基因组中相同位点的 DNA 序列。本研究以 12 个花生品种和 2 个二倍体祖先为材料, 通过扩增子重测序 EST 和 GSS (各 100 条序列) 开发 SNP。结果共检测出 18 个 EST-SNPs 和 44 个 genomic-SNPs, 出现频率分别为 1 SNP/2577 bp 和 1 SNP/1011 bp。为了进一步评估和应用所开发的 SNP, 采用高分辨率溶解曲线方法对 96 个花生品种进行 SNP 基因分型。EST-SNP 在供试品种中的多态性信息量介于 0.021~0.413, 平均为 0.172。Genomic-SNP 的多态性信息量介于 0.08~0.478, 平均为 0.249。本研究表明采用扩增子测序和 HRM 方法能够从异源四倍体花生中准确鉴定 SNP, 且所开发的 SNP 信息量丰富, 能够用于花生遗传育种研究。

关键词 花生; 单核苷酸多态; 扩增子; 高分辨率溶解曲线

Identification and Evaluation of Single-nucleotide Polymorphisms in Peanut (*Arachis hypogaea* L.) Based on Amplicon Sequencing Combined with High Resolution Melting Analysis



Li Xingyu^{1,2,3} , Liu Ying^{1,2,3} , Zhu Fanghe^{1,2,3} , Hong Yanbin^{1,2,3} , Liu Hong⁴ , Chen Xiaoping^{1,2,3} , Li Haifen^{1,2,3} , Liang Xuanqiang^{1,2,3} 

1 Crops Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, 510640

2 South China Peanut Sub-Center, National Center of Oilseeds Crop Improvement, Guangzhou, 510640

3 Guangdong Provincial Key Laboratory of Crops Genetics and Improvement, Guangzhou, 510640

4 College of Agriculture, South China Agricultural University, Guangzhou, 510642

 Corresponding author, liang804@yahoo.com;  Authors

Abstract Peanut (*Arachis hypogaea* L.) is an allotetraploid species derived from the A genome diploid species, *A. duranensis* and the B genome diploid species, *A. ipaensis*. The presence of A and B genomes makes it more challenging to identify single nucleotide polymorphism (SNP) in peanut than in diploid species. In this study, 100 DNA fragments derived from expressed sequence tags (ESTs) and genome survey sequences (GSS) were amplified and sequenced for the identification of SNPs among 12 allotetraploid cultivated peanut varieties and two diploid wild species representing the ancestral genomes. A total of 18 EST-SNPs and 44

genomic-SNPs were identified in 12 peanut varieties. The average frequency of genomic-SNPs was one per kilobase (kb), while only 0.39 per kb for EST-SNPs. To validate the identified SNPs, 96 peanut varieties were used for genotyping using high resolution melting (HRM) method. Polymorphism information content (PIC) values of EST-SNPs ranged from 0.021 to 0.413 with a mean of 0.172, while PIC values of genomic-SNPs ranged from 0.080 to 0.478 with a mean of 0.249. Our results showed that SNPs can be identified in allotetraploid peanut with high accuracy through sequencing and HRM. The identified SNPs were very informative and would be used for genetic improvement and breeding programs in peanut.

Keywords Peanut (*Arachis hypogaea* L.), SNPs, Amplicon, HRM

单核苷酸多态(Single nucleotide polymorphisms, SNPs)是动植物基因组中最主要的多态类型, 现已被大量用于遗传作图、基因定位和关联分析等研究。人类单倍体计划通过 SNP 进行标记-性状关联分析, 为人类基因组中复杂性状的研究开辟新途径。近年来, 一系列低成本、高通量的 SNP 基因分型平台正加快遗传学、基因组学等研究的发展。然而, 当前 SNP 基因分型平台以及相应的数据分析软件主要针对人和二倍体生物进行设计, 将其应用到植物上存在很多问题, 因此自然界中多数植物为多倍体或古多倍体。

异源多倍体中, SNP 被定义为不同个体同一亚基因组间的单核苷酸变异。部分同源序列变异(Homoeologous sequence variants, HSVs)被定义为亚基因组之间相应的单核苷酸变异 (Somers et al., 2003)。旁系同源序列变异 (Paralogous sequence variants, PSVs) 则定义为从一个相同的祖先序列分化出来, 并存在于同一基因组中的序列拷贝间的单核苷酸差异 (Fredman et al., 2004)。在多倍体上的遗传应用研究上, 如遗传作图、遗传多样性分析、关联分析等, 上述三种类型的核苷酸变异, 仅 SNP 具有应用价值。对 SNP 的开发需要测序不同品系/品种并准确区分同源和部分同源染色体之间的单核苷酸变异。除非使用基因组特异标记, 否则几乎所有 SNP 标记分析平台必须面对这样的事实: 即在 SNP 分析过程中, 部分同源基因组上的序列拷贝容易与同源序列一起被取样分析。在这种情况下, 如果异源四倍体两个基因组同一位点均出现多态, 其后代的分离模式将与二倍体的不同(AA, AB 和 BB), 按照 AAAA、AAAB、AABB、ABBB 和 BBBB 五种类型进行分离。此外, 当异源四倍体某一个体中的 SNP 与另一个个体上的 HSV 位点重合时, 两个体的杂交后代将无法产生正常的等位分离模式, 例

如 CT(A 基因组)TT(B 基因组) × CC(A 基因组)TT(A 基因组)。

栽培花生(*Arachis hypogaea* L.)属闭花授粉异源四倍体(AABB, $2n = 4 \times = 40$), 作为全球干旱和半干旱地区主要的经济作物, 是人类营养中油脂和蛋白的重要来源。研究表明栽培花生形成的时间很短, 大约在 3500 年通过野生种 *A. duranensis* (A genome, $n = 10$) 和 *A. ipaensis* (B genome, $n = 10$) 融合产生(Seijo et al., 2004)。由于这个物种遗传基础狭窄, 导致其 DNA 多态性极低, 使其分子标记的开发和基因组学资源的挖掘一直以来面临很大的挑战(Paterson et al., 2004)。当前花生上开发利用的标记以 SSR 为主。而关于花生上 SNP 标记的开发只有零星报道。最近, Zhou et al., (2014) 利用 ddRADseq (double-digest restriction-site-associated DNA sequencing) 技术为两个栽培花生品系及其 166 重自交系(RIL) 后代构建简化文库(reduced-representation libraries, RRLs), 并从两个亲本间检测出 53,257 个候选 SNPs, 其中 14,663 个在 RIL 群体中同样被检出, 但最终只有 1,765 个在随后的遗传图谱构建中得到验证。显然, 在缺乏参考基因组序列以及序列变异参数条件下, 利用第二代测序技术开发 SNP 的假性率高。虽然利用第二代测序技术挖掘 SNP 可节省单个标记的开发成本, 但一旦所开发的 SNP 假阳性率高, 将显著增加 SNP 的应用难度及成本。因此, 弄清 SNP、HSV 和 PSV 在花生基因组中的频率, 据此制定合适的 SNP 鉴定方法比仓促进行大规模 SNP 开发更稳妥可靠。

本研究的目的是: (1)利用扩增子测序方法鉴定 EST-SNPs 和 genomic-SNPs; (2)估算花生基因组中 SNPs, HSVs 和 PSVs 的频率; (3)估算我国南方花生品种 SNP 的分布频率。

结果与分析

1.1 SNPs、HSVs 和 PSVs 的鉴定

分析 12 个花生品种和 2 个花生二倍体祖先的 EST 和 GSS 扩增子表明, 约 90% 的 EST 和 GSS 能够成功扩增出 PCR 产物。由于 EST 序列对应的基因组

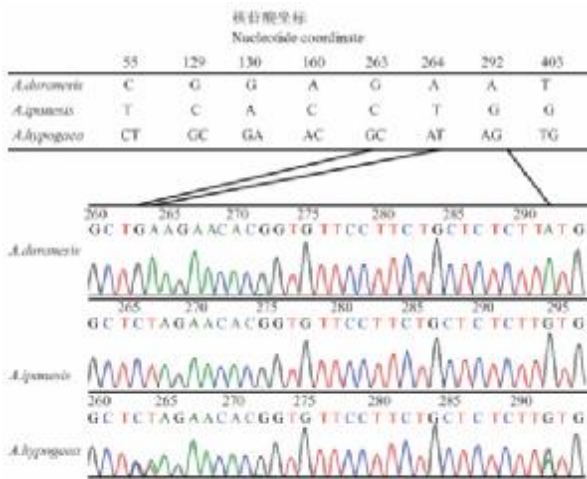


图 1 EST-48 在四倍体花生及其二倍体祖先上的扩增子序列
注: EST-48 在 *A. duranensis* 和 *A. ipaensis* 基因组中各保存一份拷贝, 且两者均保留在栽培种基因组中

Figure 1 Amplicon Sequence of EST-48 in *A. hypogaea* and putative diploid progenitor

Note: *A. duranensis* and *A. ipaensis* genome each contained one copy of sequence "EST-48", and both the copies retained in *A. hypogaea* genome

序列可能包含内含子, 导致部分基因组扩增片段超出预期长度。为保证测序质量及避免内含子的干扰, 本研究中片段长度超出预期的 EST 扩增子不用于后续的测序分析。此外, 为保证测序的特异性, 琼脂糖凝胶电泳检测出现多条扩增带的扩增子也不用于进一步测序分析。最终从 EST 和 GSS 扩增子中各挑选 100 个用于测序分析。结果表明, 82% 的 EST 扩增子和 77% 的 GSS 扩增子被成功测序。来自花生二倍体祖先 *A. duranensis* 和 *A. ipaensis* 的扩增子多为纯化子, 其中 EST 扩增子的纯化率 (87.8%) 高于 GSS 扩增子 (75.3%)。相反, 四倍体花生的扩增子多为杂合子, 其中 GSS 扩增子的杂合率 (76.6%) 高于 EST 扩增子 (62.2%)。

花生二倍体祖先出现杂合扩增子说明存在旁系同源序列, 而由于四倍体花生同时存在 A 和 B 基因组, 其杂合扩增子的序列组成将更加复杂, 可能包含部分同源基因, 也可能包含旁系同源基因。序

列比对发现, 大部分扩增片段来自 A 和 B 基因组(图 1), 但有部分只来自其中一个基因组。图 2 说明序列 EST-86 在花生 A 基因组中出现, 但在 B 基因组过程中基因组序列的重组与缺失(Adams et al., 2005)往往增加了单核苷酸变异类型的区分难度。但通过比较四倍体花生及其二倍体祖先的扩增片段序列, 多数单核苷酸变异类型能够被区分出来。如图 3 所示, 通过序列比对可推断序列 EST-87 在栽培种上的两个拷贝来自 A 基因组, 因此可确定拷贝间的序列变异为 PSV, 并非 HSV, 此外还可确定栽培品种 YY7 和 TSSLR 在第 213 碱基处的 SNP 位于 A 基因组。

1.2 SNPs、HSVs 和 PSVs 频率

82 个 EST 扩增子在 12 个花生品种中只有 12 个产生多态(至少包含 1 个 SNP), 共检测出 18 个 EST-SNP, 平均 2557 bp 出现一个 EST-SNP。C/T (G/A) 和 A/C (T/G) (均占 38.9%) 为 EST-SNP 中最丰富的核苷酸变异类型, 其次是 A/T (T/A) 和 C/G (G/C), 分别占 16.7% 和 5.5%。77 个 GSS 扩增子, 有 20 个产生多态, 共产生 44 个 genomic-SNPs。genomic-SNP 出现频率为 1 SNP/1011 bp, 远高于 EST-SNP。genomic-SNP 的核苷酸变异类型与 EST-SNP 相似, C/T (G/A) 和 A/C (T/G) 最多(分别占 68.2% 和 22.7%), 其次是 A/T (T/A) 和 C/G (G/C), 分别为 6.8% 和 2.3%。

1.3 等位基因频率

本研究中共有 32 个包含 SNP 的扩增子可用于分析 96 个南方花生品种的多态性, 其中 30 个扩增子只需分析 1 个 SNP 就足以对整个片段进行基因分型, 因为扩增子包含的 SNP 均处于连锁不平衡状态。另外两个扩增子 GSS-56 和 GSS-67, 各包含 2 个连锁平衡的独立 SNP。最终共有 34 个 SNP 可用于计算 PIC 值。对 34 个 SNP 位点设计巢式 PCR 引物, 随后进行 HRM 分析。结果表明, 除了 GSS-35, 其余 33 个扩增产物溶解曲线能够很好对调查材料进行基因分型(图 4), 其基因分型结果与前面的测序结果相吻合。SNPs 是典型的双等位基因标记, 因此单个 SNP 的 PIC 值不可能超过 0.5。本研究中 EST-SNP 的 PIC 值介于 0.021~0.413, 平均为 0.172(表 1)。由于供试材料中只有一个品种在 EST-12 位点上产生等位变异, 因此 EST-12 的 PIC 值仅为 0.021。Genomic-SNP 的 PIC 值显著高于 EST-SNP, 最低为 0.08, 最高为 0.478, 平均为 0.249。

2 讨论

本研究以四倍体花生及其二倍体祖先为材料, 通过重测序 EST 和 GSS 开发 SNP。如何鉴定出可真正用于基因分型的可靠 SNP 是当前多倍体 SNP 开发的最大挑战。为确保所开发 SNP 的可信度, 首先必须将其与 HSV 和 PSV 区分开。这对于通过比较来自不同品种的 EST 或 GSS 直接开发 SNP 在生物信息学上无疑是一个难题。首先需要每个品种有大量高度冗余的序列, 而目前 NCBI 数据库上面提供的花生 EST 或 GSS 序列尚未达到此测序深度。最近, Zhou et al. (2014) 通过利用 Illumina 第二代测序平台大量平行测序四倍体花生基因组解决上述问题, 共开发了 53 257 个候选 SNP, 但仅有极小部分在后续的遗传图谱构建中得到验证。因此, 尚不清楚该研究中多大比例的 SNP 为假阳性。其他复杂的高通量测序技术的境遇也可能于此相同。

本研究显示花生 EST-SNP 和 Genomic-SNP 频率分别为 1 SNP/2557bp 和 1 SNP/1011bp, 显著低于其他自花授粉作物, 如大豆、小麦、水稻、油菜。与此同时, 花生 HSV 与 PSV 的频率则显著高于 SNP, 是其几十倍。因此, 在序列拼接时, 一旦将部分同源序列或旁系同源序列误判为同源序列, 将显著增加 SNP 的假阳性率。因此, 利用第二代测序技术开发 SNP, 适当提高序列的拼接标准, 将有助于防止部分同源基因和旁系同源基因与同源基因错误拼接, 降低 SNP 的假阳性率。

本研究所采用的扩增子测序是一个复杂的方法, 其中光测序每个个体的扩增子就需要很大的工作量。针对异源多倍体植物, 此法的优点是只有基因组上特异的序列才被用于后续的测序分析。在一个或两个基因组中扩增出多个位点的扩增子测序往往失败, 因为扩增片段之间非常有可能出现插入缺失导致无法读取测序峰图。相反, 只来自基因组内一个位点的扩增片段通常能被高质量测序。因此, 结合二倍体祖先, 通过对四倍体花生品种扩增产物的测序, 最终能够开发出可靠的 SNP。然而, 由于花生基因组 SNP 发生频率偏低, 单独采用扩增子测序存在低效率的问题。为兼顾 SNP 开发的效率和准确性, 将第二代测序技术与扩增子测序相结合不失为 SNP 鉴定的好方法。首先利用第二代测序技术筛选大量的候选 SNP, 然后利用扩增子测序进行验证。

过去十年, 新一代荧光双链 DNA 染料直接推动了 HRM 技术的发展(Wittwer et al., 2003)。新染料

例如 LCGreen 和 CYTO®9 等具有毒性低, 能够高浓度饱和双链 DNA 等优点。相比非饱和染料, 新染料对 PCR 的抑制性低, 与 DNA 结合力更强, 熔解曲线分辨率更高(Wittwer et al., 2003; Monis et al., 2005), 最终能检测出扩增产物中的 SNP, 甚至能发现体细胞突变和甲基化(Vossen et al., 2009; Mastoraki et al., 2015; Chang et al., 2014; Kristensen et al., 2008)。目前 HRM 在植物上的研究应用仍较少。最新的研究表明 HRM 是检测 SNP、INDELS 和 SSR 的有效工具(Chateigner-Boutin et al., 2007; Lehmsiek et al., 2008; Wu et al., 2008)。当溶解峰清晰明显时, HRM 分析通常效果最好。为了使 HRM 分析到达更好的效果, 需要充分考虑多个因素, 包括(1)PCR 产物长度和特异性; (2)GC 含量; (3)SNP 类型和密度(注意避开 SNP 密集的区域); (4)PCR 反应体系的优化等(Comai et al., 2006)。HRM 分析能够检测到全部单碱基突变, 其中对单核苷酸变异类型 C/T (G/A)灵敏度高, 而对 A/T 和 G/C 的灵敏度低。本研究中单核苷酸变异类型 C/T (G/A)最为丰富, 因此 HRM 适合用于检测花生 SNP。

本研究表明花生 genomic-SNP 发生频率显著高于 EST-SNP, 几乎是 EST-SNP 的三倍, 且 genomic-SNP 的 PIC 值也显著高于 EST-SNP, 说明 genomic-SNP 在花生基因组中分布广且均匀, 因此, 在进行遗传作图, 关联分析等研究, GSS-SNP 的效率将比 EST-SNP 高。

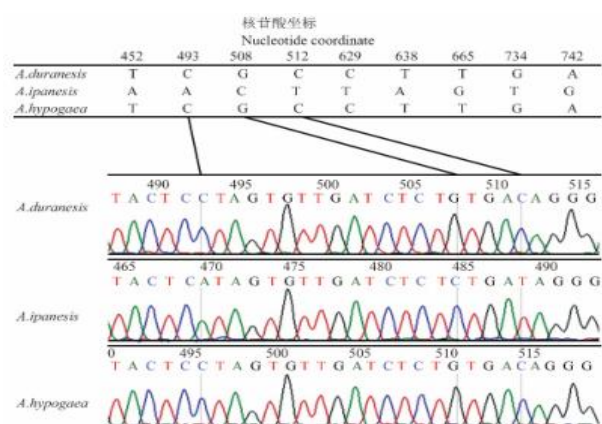


图 2 EST-86 在四倍体花生及其二倍体祖先上的扩增子序列
注: EST-86 在 *A. duranensis* 和 *A. ipanensis* 基因组中均存在一份拷贝, 但栽培种只保留 *A. duranensis* 基因组拷贝

Figure 2 Amplicon sequence of EST-186 in *A. hypogaea* and putative diploid progenitor

Note: *A. duranensis* and *A. ipanensis* genome each contained one copy of sequence "EST-86", but one the copy from *A. duranensis* genome retained in *A. hypogaea* genome

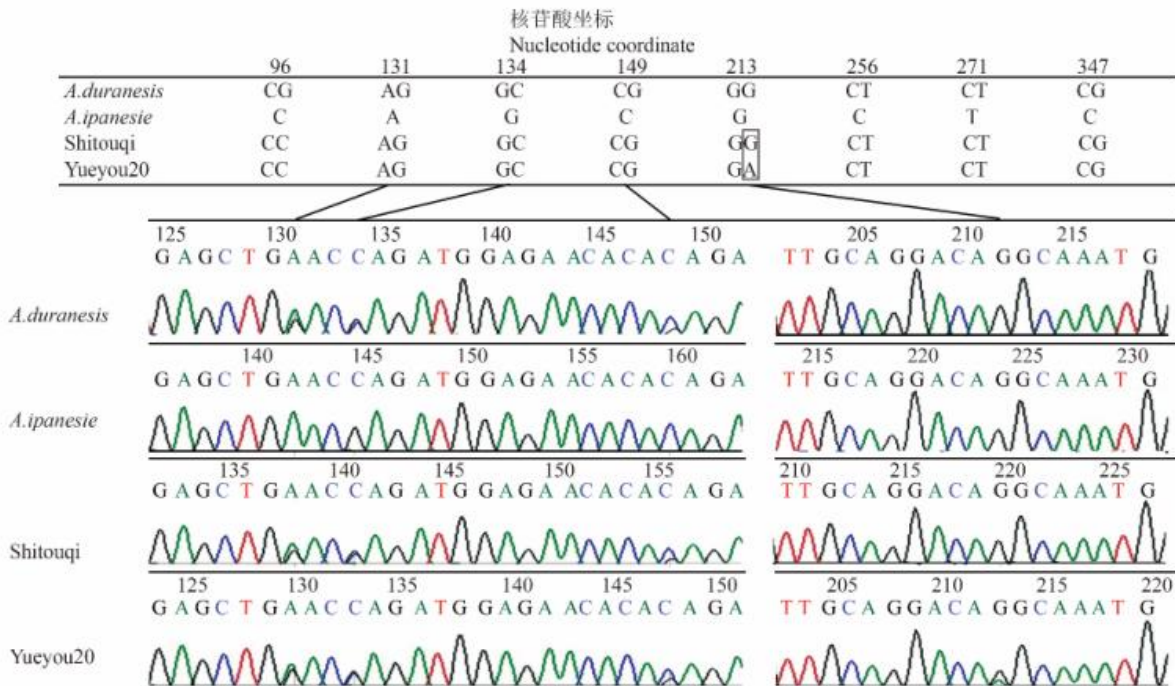


图 3 EST-87 在四倍体花生及其二倍体祖先上的扩增子序列

注: EST-87 在祖先 A 基因组上有两个拷贝,但在祖先 B 基因组上只有一个拷贝,栽培种只保留祖先 A 基因组上的两个拷贝;方框代表花生品种间的 SNP

Figure 3 Amplicon sequence of EST-87 in *A. hypogaea* and putative diploid progenitor

Note: *A. duranesis* genome contained two copy of sequence "EST-87", while *A. ipanesie* genome contained only one copy and only the copies from *A. duranesis* genome retained in *A. hypogaea* genome; SNP between peanut varieties is boxed

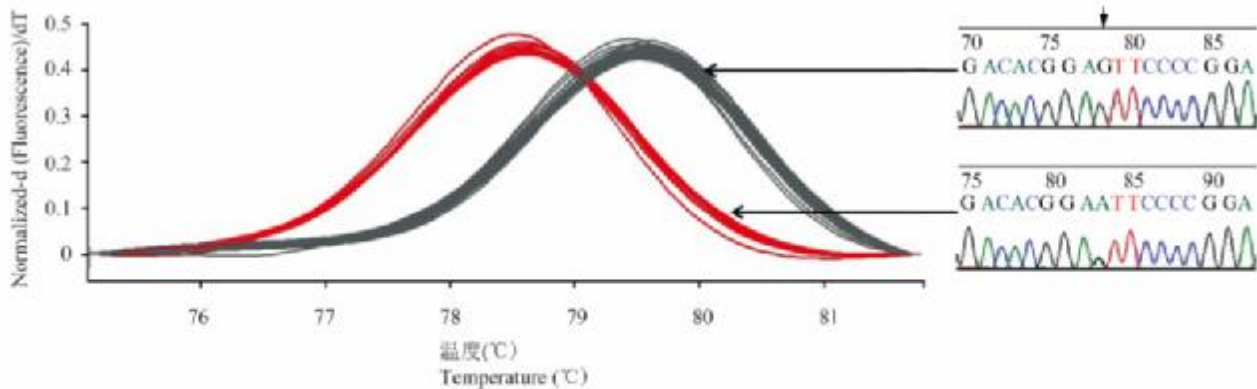


图 4 扩增子 EST-66 的溶解曲线分析

注: 扩增子 EST-66 在 96 花生品种上的 HRM 分析, 28 个样本显示为红线, 68 个显示为灰线; 垂直箭头指向 SNP 位置

Figure 4 Amplicon melting analysis of amplicon EST-66

Note: 96 peanut varieties were analyzed by HRM in amplicon EST-66, 28 samples were showed in red line, and 68 samples were showed in grey lines; Vertical arrows show the SNP position

表1 花生SNP 标记等位基因频率和PIC 值

Table 1 Allele frequencies and PIC values of SNP markers in peanut

标记 Marker	引物序列 Primer sequence		变异类型 Variant classes	等位基因频率 Allele frequency		多态性信息量 PIC
	正向 Forward	反向 Reverse		等位基因1 (%)	等位基因2 (%)	
				Allele 1 (%)	Allele 2 (%)	
EST-11SNP	TTCACGCCAGATGAGGAT	CAGGAGTCGGGCTATTGT	C/G	4	92	0.080
EST-12SNP	CATCTCAGCACTCAACTT	GAGTAAGGGTTTGAGGAA	C/T	1	95	0.021
EST-21SNP	TTGAGTGAACCACTTGAGG	GCCGACGAAGAAGAGGAATA	G/T	15	81	0.264
EST-26SNP	TACCCCTGACTTGGTTTATGGC	GCCTTCAGCGAACTGTCCC	A/C	8	88	0.153
EST-33SNP	CAAACAAAGTCATCGCAGTC	CTCTTCATCCATCTCACAG	G/T	92	4	0.080
EST-45SNP	AACTGGCGTCTTTCACCG	GTCCCAAGCCTGCATCCAAA	G/T	88	8	0.153
EST-48SNP	ACCGCTTACTATCACTACCA	TCTAATATGACTGTTCACAGA	G/T	83	13	0.234
EST-46SNP	CATCGTCATCCGTGAGGTGT	TTGGACTTGAGGAGGAGGTAG	A/G	28	68	0.413
EST-70SNP	AGGCGTTGAAATCGAGGTCC	GGCATCCCAGTTCAGAAAGG	C/T	79	17	0.291
EST-80SNP	ACAGAGTCATTGGTGATGGGAGTT	CACCTGCAGGTGCATGGGTT	A/G	5	91	0.099
EST-87SNP	ATGAGCTGGACCAGATGGAG	CCTGAAGTTACACTGTTGAGGC	A/G	13	83	0.234
EST-92SNP	GAGTAATCGTTGAGAACTCGC	TCAGAAGACTTCGCCACCTT	A/T	2	94	0.041
GSS-25SNP	ACTTAGCATAAGAGGGTATTTG	AGCGGAACAGGATAAGCAAT	A/G	14	82	0.249
GSS-105SNP	GCCAGAACTATGCTTCCCTT	AAATCCCTCGAGGACAATC	A/C	21	75	0.234
GSS-265SNP	GACTCAGCGCAGTGCCAACT	GTGCCATGCTTACTCAACAAAA	C/A	85	11	0.203
GSS-355SNP	ATTGAACAAAGGATTGAGAT	CAAATGAAGGTGTAGATGGC	A/T	24	72	0.375
GSS-445SNP	CAACATCCAGACACCCCAAAT	CTCCAGGTCAGGCTGAACAT	A/C	16	80	0.278
GSS-455SNP	TCTAGCCAGATCGCCATAC	GAATACGTGACTGACCCAAG	A/G	87	9	0.170
GSS-465SNP	TCTACCTTTGCCCTATCCAC	ACTTGTTCGAAAAGATGCC	C/T	4	92	0.080
GSS-495SNP	TTCAATGCCTAATCATTCCCACTA	GAGGACTTGTCCGCCACCTT	A/C	18	78	0.305
GSS-525SNP	AAGCCTTAAACAGGGGTAGC	TTGGTAGGAGGTGAGCGATG	C/T	89	7	0.135
GSS-545SNP	AGTGCAAAAGGTTGCGCTTG	TTGCTCCATTTTCATTAGGTTTT	C/T	82	14	0.249
GSS-56-1SNP	ACATTGCCAGGTGGTGGTAA	ATGGGAGAAATGTCACCTTATGTT	A/G	6	90	0.117
GSS-56-2SNP	TTCTGTATGCAAACTACTCCG	ATGAATACTACTGTCTCCCTG	C/T	15	81	0.264
GSS-67-1SNP	ATAATCCACTGCCACCAGAA	TATGCTATGCCCAACAAGG	C/T	87	9	0.170
GSS-67-2SNP	ATGGTTCATCCAAAGATAG	TGAGATTCATACCCAAAGAG	C/T	83	13	0.234
GSS-695SNP	TGGCTAGAGGATGGTTGGAG	TGATAGCCCGGTCTATGGTA	A/C	26	70	0.395
GSS-725SNP	CATTCTACTGGTGGTCTGT	CATGAACTGGATTATTTGCA	A/T	71	15	0.429
GSS-745SNP	ATTTGCCACTAACTCCCTC	TTAGCATACAAGATAGTTGAAAAGT	C/T	89	7	0.135
GSS-765SNP	ACAAGCTATCCCAACTCCAC	TAAGCAGTCCACCAATCAA	A/G	33	61	0.478
GSS-805SNP	CCTAGAATATTCTAAACCTCCAT	GCCTTACACGTACATGCTTT	A/G	17	79	0.291
GSS-815SNP	CCATTACTTGTGTTTGCTCAC	TGTGGTTATAGAACAGAGGG	C/T	83	13	0.234
GSS-995SNP	TAGAAAGCCCTGGATGTTAG	AAAGAGTGCAGATGCTGTCA	C/T	90	6	0.117

表2 96 个被调查的花生品种

Table 2 Ninety six investigated peanut varieties

序号	品种	登录号	序号	品种	登录号
No.	Variety	Accession No.	No.	Variety	Accession No.
1	伏花生	GH03291	25	大桶豆	GH00522
2	Fuhuasheng	GH02437	26	Datongdou	GH00929
	狮头企			信宜四粒红	
3	Shitouqi	GH02986	27	Xinyisilihong	GH02771
	竹丝豆			河源箬壳豆	
4	Zhusidou	GH00654	28	Heyuanbokedou	GH00792
	猪仔豆			十里甫红衣	
5	Zhuzaidou	GH03862	29	Shilipuhongyi	GH00796
	陆屋大花生			二康仔	
6	Luwudahuasheng	GH03801	30	Erkangzai	GH00787
	翁源勾豆			四粒红	

7	Wengyuangoudou 台山三粒红	GH02818	31	Silihong 珠生种	GH02932
8	Taishansanlirou 辽宁四粒红	GH02919	32	Zhushengzhong 潭口小子	GH03254
9	Liaoningsilihong 珍珠黑	GH02940	33	Tankouxiaozi 琅江花生	GH02922
10	Zhenzhuhei 粤油20	GH02739	34	Langjianghuasheng 浸潭九龙豆	GH02982
11	Yueyou20 崖县小粒	GH02809	35	Jintanjiulongdou 长针珠豆	GH02410
12	Yaxianxiaoli 粤油13	GH02742	36	Zhangzhenzhudou 台山花生	GH02846
13	Yueyou13 光明豆柠	GH03092	37	Taishanhuasheng 高埠大珠豆	GH02194
14	Guangmingdouning 木梗糯	GH00497	38	Gaobudazhudou 白肉仔	GH02390
15	Mugengnuo 广柳	GH00283	39	Bairouzai 新会珠豆	GH02432
16	Guangliu 台山特域	GH00488	40	Xinhuizhudou 石龙红	GH02183
17	Taishanteyu 八月豆	GH00644	41	Shilonghong 白延豆柠	GH02412
18	Bayuedou 台山珠	GH02834	42	Baiyandouning 信宜四粒白	GH00930
19	Taishanzhu 罗定大豆	GH00638	43	Xinyisilibai 岑北锣湾	GH00858
20	Luodingdadou 开建铁豆	GH00645	44	Chenbeiluowan 鹤山花生	GH03086
21	Kaijiantiedou 勾鼻豆	GH03088	45	Heshanhuasheng 高要白土	GH02172
22	Goubidou 大埔种	GH00636	46	Gaoyaobaitu 防城	GH02902
23	Dabuzhong 紫狮	GH02125	47	Fangcheng 辐狮	GH03326
24	Zishi 小果峰	GH03384	48	Fushi 临桂大扯子	GH02210
	Xiaoguofeng			Linguidachezi	

49	址山大豆	GH02815	73	岭北铜锣湾	GH00255
50	Zhishandadou 朱砂红	GH03139	74	Lingbeitongluowan 五华百日仔	GH00789
51	Zhushahong 电白钩	GH02204	75	Wuhuabairizai 小硫球	GH02132
52	Dianbaigou 北海珍珠	GH02825	76	Xiaoliuqiu 信宜三粒红	GH00773
53	Beihaizhenzhu 崩江种	GH02789	77	Xinyisanlihong 佛州巨人	GH01130
54	Bengjiangzhong 胭脂花生	GH02796	78	Fouzhoujuren 大花大豆	GH03708
55	Yanzhihuasheng 海丰四季豆	GH02699	79	Dahuadadou 罗定牛角豆	GH03765
56	Haifengsijidou 阳春铺地毡	GH03722	80	Luodingniujiadou 乐会大豆	GH03805
57	Yangchunpudizhan 台山附城	GH02827	81	Lehuidadou 泰岭百日豆	GH03817
58	Taishanfucheng 东莞拔豆	GH02839	82	Tailingbairidou 天涯	GH03858
59	Dongguanbadou 陆屋拔豆	GH02927	83	Tianya 英德细介子	GH03936
60	Luwubadou 循58	GH03322	84	Yingdexijiezi 东平立基花生	GH00273
61	Xun58 道明豆拧	GH02799	85	Dongpinglijihuasheng 花县细珠	GH02817
62	Daomingdouning 昌黎大花生	GH00194	86	Huaxianxizhu 遁地雷	GH02935
63	Chanlidahuasheng 小果子	GH00211	87	Dundilei 黄梅花生	GH02950
64	Xiaoguozi 早蔓生	GH00206	88	Huameihuasheng 潮州豆	GH03691
65	Zaomansheng 早花生	GH02404	89	Chaozhoudou 佛州蔓生	GH02946
66	Zaohuasheng 直花蔓生	GH02747	90	Fouzhoumansheng 迟生	GH02961
67	Zhihuamansheng 城东社大豆	GH02814	91	Chisheng 那大大豆	GH03741

68	Chengdongshedadou 紫金种	GH02108	92	Nadadadou 节豆	GH03927
69	Zijinzhong 群毅三月柠	GH02213	93	Jiedou 恩花1 号	GH00502
70	Qunyisanyuening 花县拔	GH02503	94	Enhualhao 水豆	GH00640
71	Huaxianba 细粒种	GH03077	95	Shuidou 新华6 号	GH03143
72	Xilizhong 浠水川德	GH02966	96	Xinhua6hao 六月豆	GH03379
	Xishuichuande			Liuyuedou	

3 材料与试剂

3.1 植物材料

14 个用于 SNP 鉴定的植物材料, 包括(1)12 个花生品种(表 2); (2)2 个花生野生种(二倍体祖先): *A. duranensis* (A 基因组)和 *A. ipaensis* (B 基因组)。96 个南方花生被用于本研究中等位基因频率的调查, 其中包含上述的 12 个品种(表 2-1)。基因组 DNA 采用 Moretzsohn et al. (2005)的方法从花生嫩叶中提取。

3.2 PCR 扩增

分别从 Guo et al., (2008)开发的花生 EST 数据库和 NCBI 数据库(www.ncbi.nlm.nih.gov)中挑选 EST(expressed sequence tags, EST)和 GSS (Genome Survey Sequences)序列。采用 Primer premier 5 软件(Whitehead Institute for Biomedical Research, Cambridge, Mass)设计引物。引物设计基于下列原则: (1)溶解温度(Tm) 50°C~60°C, 最适温度 55°C; (2)扩增产物介于 450 bp~750 bp; (3)引物长度 18 bp 至 24 bp, 扩增率大于 80%; (4)引物 GC 含量 40%~60%。PCR 反应在 PE9700 热循环仪(美国 ABI 公司)上进行。PCR 反应体系(20 μL): 1×PCR buffer (50 mmol L⁻¹ KCl, 10 mmol L⁻¹ Tris-HCl, pH8.3, 1.5 mmol L⁻¹ MgCl₂, 0.1% 明胶), 1 U Taq DNA 聚合酶, 50~100 ng 模板 DNA, 0.15 μmol L⁻¹ 引物, 0.2 mmol L⁻¹ dNTP。反应条件为 94°C 5min; 94°C 1 min, 55°C 30 s, 72°C 1 min, 35 个循环; 72°C 5 min。单一条带 PCR 产物通过琼脂糖凝胶电泳检测确定。

3.3 Sanger 测序和 SNP 挖掘

将 12 个花生品种和 2 个野生种的单一条带 PCR 产物送至北京六合华大基因科技股份有限公司, 采用 Sanger 法进行双向测序。序列比对和 SNP 鉴定采用突变分析软件 Mutation Surveyor®。

3.4 SNP 基因分型

采用高分辨率溶解曲线(High Resolution Melting, HRM)方法对 96 个花生品种进行 SNP 基因分型。基于扩增子测序结果, 在 SNP 侧翼序列设计用于 HRM 分析的巢式 PCR 引物。引物设计遵循以下原则: (1)扩增片段长度控制在 80 bp 至 300 bp; (2)扩增片段内只包含一个 SNP; (3)引物结合区含有 HSV 或 PSV, 以便设计亚基因组特异性引物。PCR 反应容量为 10 μl, 除了在反应液中添加 2.5 μM CYTO®9 (Invitrogen, Carlsbad, CA, USA), 其余反应条件与上述相同。1 μl 100×第一次 PCR 产物稀释液作为 PCR 反应的 DNA 模板。

3.5 统计分析

SNP 标记的多态性信息量计算采用公式:

$$PIC=1-\sum_{i=1}^k p_i^2$$

其中 k 代表等位基因总数量, p 为某一既定位点第 i 个等位基因的频率。

作者贡献

刘颖和朱方何负责扩增子测序; 李杏瑜、洪彦彬和刘洪负责 HRM 分析; 李杏瑜完成数据分析和论文初稿的写作; 陈小平和李海芬参与论文修改; 梁炫强是项目的构思者及负责人, 指导实验设计,

数据分析, 论文写作与修改。全体作者都阅读并同意最终的文本。

致谢

本研究由现代农业产业技术体系建设专项(CARS-14)、广州市珠江科技新星专项(2013J2200088)、国家自然科学基金(31271767)、广东省自然科学基金(S2013020012647)、国家科技支撑计划(2013BAD01B05-4)、广东省科技厅团队项目(2011A020102010)共同资助。

参考文献

- Anderson J.A., Churchill G.A., Autrique J.E., Tanksley S.D., and Sorrells M.E., 1993, Optimizing parental selection for genetic linkage maps, *Genome*, 36(1): 181-186
- Botstein D., White R.L., Skolnik M., and 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
- Chang C.C., Chang Y.S., Chan W.L., Yeh K.T., Wei R.J., and Chang J.G., 2014, Detection of SF3B3 gene mutations in oral cancer by high resolution melting analysis, *Clin. Lab.*, 60 (12): 2023-2029
- Chateigner-Boutin A.L., and Small I., 2007, A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons, *Nucleic Acids Res.*, 35(17): e114
- Comai L., and Henikoff S., 2006, TILLING: practical single-nucleotide mutation discovery, *Plant J.*, 45(4): 684-694
- Fredman D., White S.J., Potter S., Eichler E.E., Den Dunnen J.T., and Brookes A.J., 2004, Complex SNP-related sequence variation in segmental genome duplications, *Nat. Genet.*, 36 (8): 861-866
- Guo B., Chen X., Dang P., Scully B.T., Liang X., Holbrook C. C., Yu J., and Culbreath A.K., 2008, Peanut gene expression profiling in developing seeds at different reproduction stages during *Aspergillus parasiticus* infection, *BMC Dev. Biol.*, 8: 12
- Kristensen L.S., Mikeska T., Krypuy M., and Dobrovic A., 2008, Sensitive melting analysis after real time-methylation specific PCR (SMARTMSP): high-throughput and probe-free quantitative DNA methylation detection, *Nucleic Acids Res.*, 36 (7): e42
- Lehmensiek A., Sutherland M.W., and McNamara R.B., 2008, The use of high resolution melting (HRM) to map single nucleotide polymorphism markers linked to a covered smut resistance gene in barley, *Theor. Appl. Genet.*, 117(5): 721-728
- Mastoraki S., Chimonidou M., Dimitrakopoulos L., Kounelis S., Malamos N., Georgoulas V., and Lianidou E., 2015, A rapid and accurate closed-tube methylation-sensitive high resolution melting analysis assay for the semi-quantitative determination of SOX17 promoter methylation in clinical samples, *Clin. Chim. Acta*, 444: 303-309
- Monis P.T., Giglio S., and Saint C.P., 2005, Comparison of SYTO9 and SYBR Green I for real-time polymerase chain reaction and investigation of the effect of dye concentration on amplification and DNA melting curve analysis, *Anal. Biochem.*, 340(1): 24-34
- Moretzsohn M.C., Leoi L., Proite K., Guimaraes P.M., Leal-Bertioli S.C., Gimenes M.A., Martins W.S., Valls J.F., Grattapaglia D., and Bertioli D.J., 2005, A microsatellite-based, gene-rich linkage map for the AA genome of *Arachis* (Fabaceae), *Theor. Appl. Genet.*, 1(6): 1060-1071
- Paterson A.H., Stalker H.T., Gallo-Meagher M., Burow M.D., Dwivedi S.L., Crouch J.H., and Mace E.S., 2004, Genomics and genetic enhancement of peanut, In: Wilson R.F., Stalker H.T., and Brummer C.E. (eds.), *Genomics for legume crops*, AOCS Press, Champaign, USA, pp.97-109
- Seijo J.G., Lavia G.I., Fernández A., Krapovickas A., Ducasse D., and Moscone E.A., 2004, Physical mapping of the 5S and 18S~25S rRNA genes by FISH as evidence that *Arachis duranensis* and *A. ipaensis* are the wild diploid progenitors of *A. hypogaea* (Leguminosae), *Am.J.Bot.*, 91(9):1294-1303
- Somers D.J., Kirkpatrick R., Moniwa M., and Walsh A., 2003, Mining single-nucleotide polymorphisms from hexaploid wheat ESTs, *Genome*, 49(3): 431-437
- Vossen R.H., Aten E., Roos A., den Dunnen J.T., 2009, High-resolution melting analysis (HRMA): more than just sequence variant screening, *Hum. Mutat.*, 30(6): 860-866
- Wittwer C.T., Reed G.H., Gundry C.N., Vandersteen J.G., Pryor R.J., 2003, High-resolution genotyping by amplicon melting analysis using LCGreen, *Clin. Chem.*, 49(6 Pt 1): 853-860

- Wu S.B., Wirthensohn M., Hunt P., Gibson J., and Sedgley M.,
2008, High resolution melting analysis of almond SNPs de-
rived from ESTs, *Theor. Appl. Genet.*, 118(1): 1-14
- Zhou X., Xia Y., Ren X., Chen Y., Huang L., Huang S., Liao B.,
Lei Y., Yan L., and Jiang H., 2014, Construction of a
SNP-based genetic linkage map in cultivated peanut based
on large scale marker development using next-generation
double-digest restriction-site-associated DNA sequencing
(ddRADseq), *BMC Genomics*, 15: 351