

花生全基因组变异的鉴定和 InDel 标记的开发及应用

张雪姣¹, 季木灯^{1,2}, 王丽媛¹, 李启彪¹, 徐磊¹, 江日东³, 徐志军¹

(¹广东省省级现代农业(耕地保育与节水农业)产业技术研发中心/中国热带农业科学院湛江实验站, 湛江 524013; ²云南农业大学农学与生物技术学院, 昆明 650201; ³湛江市农业科学研究院, 广东湛江 524094;)

摘要: 利用 13 个花生品种的重测序数据进行花生全基因组变异鉴定, 分析 InDel 在花生基因组上的分布特点, 开发、验证并评估 InDel 标记在花生亲缘关系和品种鉴定中的效率。结果表明, 从 13 个花生品种中共检测到 313432 个高质量 SNP 和 38777 个高质量 InDel, 平均分布密度分别为 123 个/Mb 和 15.23 个/Mb; InDel 和 SNP 主要分布于基因间区域, 分别占 52.35% 和 60.08%。利用插入或缺失长度 ≥ 10 bp 的 InDel 位点进行标记开发, 共有 3675 个位点可以进行引物设计, 这些位点在花生 20 条染色体上不均匀分布, 平均分布密度为 1.48 个/Mb。e-PCR 检测表明, InDel 引物扩增位点以 1 个位点为主, 有效引物数为 2561 对 (69.69%), 可检测出 3133 个位点, 根据引物扩增位点在基因组上的位置信息绘制了扩增位点的物理图谱。在 100 对随机引物中, 共有 31 对在 4 个亲缘关系较远的花生品种中存在差异条带, 这些引物在 47 份花生品种 (或育种系) 中共检测到 62 个等位变异, 主基因频率范围为 0.51~0.98, 平均为 0.77; PIC 信息量范围为 0.04~0.37, 平均为 0.24。聚类分析和群体结构分析均可将 47 份花生品种 (或育种系) 分为 2 个类群, 且结果基本一致; 47 份材料最少可用 7 个标记进行区分, 表明开发的 InDel 标记可以有效的用于花生遗传多样性的评估和品种鉴定。研究结果丰富了花生的分子标记, 为 InDel 标记在花生资源遗传多样性、品种鉴定、指纹图谱构建等遗传研究中的利用提供了理论基础。

关键词: 花生; 全基因组变异; InDel; 分子标记; 遗传多样性

Genome-wide Variation Identification in Peanut and Development of InDel Markers for Genetic Research

ZHANG Xuejiao¹, JI Mudeng^{1,3}, WANG Liyuan¹, LI Qibiao¹, XU Lei¹, JIANG Ridong², XU Zhijun¹

(¹Guangdong Modern Agriculture (Cultivated Land Conservation and Water-saving Agriculture) Industrial Technology Research and Development Center/Zhanjiang Experiment Station, Chinese Academy of Tropical Agricultural Sciences, Zhanjiang 524013; ²College of Agronomy and Biotechnology, Yunnan Agricultural University, Kunming 650201; ³Zhanjiang Academy of Agricultural Sciences, Zhanjiang 524094)

Abstract: This study aimed to identify genetic variations on the whole genome level using genome resequencing data of 13 peanut cultivars, clarify their distribution characteristic, develop and verify InDel markers, and evaluate the efficiency of InDel markers in peanut cultivar identification. A total of 313432 high-quality SNPs and 38777 high-quality InDel were detected, with an average distribution density of 123 and 15.23 /Mb, respectively. The InDels and SNPs were mainly distributed in intergenic regions, with a frequency of 52.35% and 60.08%, respectively. Primers were designed using InDel with insertion or deletion length ≥ 10 bp, and 3675 InDel could be used to develop InDel markers. These InDel locus were unevenly distributed on the 20 chromosomes of peanut with an average density of 1.48 /Mb. Using electronic PCR, the InDel primers amplified mainly with 1 loci and 2561 effective primers (69.69%) could amplified 3133 effective loci in the peanut reference genome. The physical map of amplification loci was drawn according to the loci position in cultivated peanut genome. Among 100 pairs of random primers, 31 pairs amplified different bands in the 4 varieties with distant relatives. The 31 InDel primer pairs amplified 62 alleles in 47 peanut cultivars (or breeding lines), the

第一作者研究方向为作物遗传育种, E-mail: zhangxj@catas.cn

通信作者: 徐志军, 研究方向为花生种质资源学, E-mail: zhijunxu1990@163.com

基金项目: 海南省自然科学基金项目(321QN348); 中央级公益性科研院所基本科研业务费专项(1630102024002, 1630102023001)

Foundation projects: Hainan Provincial Natural Science Foundation of China (321QN348); Central Public-interest Scientific Institution Basal Research Fund(1630102024002, 1630102023001)

frequency of major genes ranged from 0.51 to 0.98 with an average of 0.77, and the PIC information ranged from 0.04 to 0.37, with an average of 0.24. Both cluster analysis and population structure analysis could divide the 47 peanut cultivars (or breeding lines) into two groups, and the 47 materials could be distinguished by at least 7 markers, indicating that the developed InDel markers could be effectively used for the assessment of genetic diversity and variety identification of peanut. The research results enriched the molecular markers of peanut, and provided a theoretical basis for the use of InDel markers in genetic studies of peanut resource genetic diversity, variety identification, fingerprint construction.

Key words: peanut; genome-wide variation; InDel; molecular markers; genetic diversity

花生 (*Arachis hypogaea* L.) 是我国重要的油料作物、经济作物、食用作物和饲料兼用作物, 在保证我国食用油安全和供给方面具有重要意义^[1]。花生在我国广泛种植, 2022 年全国的花生种植面积为 4.68×10^6 hm², 种植面积在国内大宗作物中位居第七, 种植业产值仅次于水稻、小麦和玉米, 种植效益明显 (<https://data.stats.gov.cn/>)。随着栽培种花生二倍体祖先种^[2] (*Arachis duranensis* (AA) 和 *Arachis ipaensis* (BB)) , 四倍体近缘野生种^[3] (*Arachis monticola*, AABB) 和 4 个栽培种花生^[4-7] (Tifrunner、狮头企、伏花生和 BaileyII, AABB) 的全基因组数据发布, 基于全基因组数据的花生分子标记快速发展, 大量的 SSR、SNP 位点被鉴定出来, 已广泛的应用于花生品种鉴定、种质资源遗传多样性、重要性状的 QTL 定位、基因挖掘和育种研究^[8-11]。InDel 是基因组序列中发生的小片段的插入或缺失, 其分布密度仅此于 SNP。利用 InDel 位点碱基序列差异开发的分子标记, 因具有变异稳定、准确性高、易于分型, 所需检测设备简单和成本低廉的特点, 已经广泛的应用于许多作物的遗传研究中。在花生的研究上, Zhuang 等^[6]从 52 份花生资源的重测序数据中鉴定到 1716 万个无重复 SNP 和 452 万个无重复 InDel; 王娟等^[12]通过对 169 份花生核心种质进行 GBS 测序鉴定到 10401 个 InDel, 发现仅有少数分布在功能基因相关区域; 刘宇等^[13]利用 13 个多态性 InDel 标记检测 4 个典型生物学类型的 90 份中国花生地方品种材料, 共检测出 43 个等位变异, 标记的多样性指数平均为 0.443, 可以有效揭示栽培种花生的遗传变异; 陶顺玉等^[14]利用 2 个含油量差异显著的亲本徐花 13 和中花 6 号的三代测序数据, 开发了 84 个位于含油量 QTL 区间的 InDel 标记, 使用其中的 9 个标记对近等基因系群体进行基因分型, 通过连锁分析将含油量 QTL 定位到 1.2Mb 区间; 黄莉等^[15]利用 InDel 标记加密, 将根部结瘤 QTL *qPNA08* 和 *qPNB07* 的区间分别由 4.7 Mb 和 9.9 Mb 缩小至 1.6 Mb 和 1.8 Mb, 两个 QTL 的遗传变异解释率分别由 9.1%和 7.1%增加至 16.4%和 9.9%, 证实了利用 InDel 标记进行 QTL 定位的可行性。这些研究表明花生基因组中存在着丰富的 InDel 变异, 基于 InDel 位点的分子标记在花生的遗传研究中具有丰富的应用场景。然而, 到目前为止可利用的花生 InDel 标记的数量还十分有限, 基于花生全基因组 InDel 位点的分子标记开发还未见报道, 限制了花生 InDel 标记在花生品种鉴定和遗传研究中的应用。因此, 本项目拟利用 13 个花生品种的全基因组重测序数据, 鉴定全基因组变异, 分析 SNP 和 InDel 变异在基因组上的分布规律, 开发易于检测的 InDel 标记, 同时利用随机挑选的 InDel 标记对 8 份亲缘关系较远的花生品种和 39 份亲缘关系较近的的育种系进行遗传多样性分析、群体结构分析和指纹图谱构建, 来评估开发的 InDel 标记在花生资源鉴定和遗传研究中的效率, 研究结果将进一步丰富花生的分子标记, 同时为 InDel

标记在花生品种鉴定、资源评价和 QTL 定位等遗传研究上的应用奠定基础。

1 材料与方法

1.1 基因组数据

13 个花生品种的重测序数据来源于 NCBI 数据库 (<https://www.ncbi.nlm.nih.gov/>)，品种信息和数据库编号如表 1 所示。

表 1 13 个重测序花生品种的信息和数据编号

Table 1 Information and genome resequencing data accessions of 13 peanut cultivars

序号	品种名称	来源	重测序数据编号
No.	Cultivar name	Origin	Resequencing data accession
1	狮头企	广东省澄海县	SRR7617991
2	伏花生	山东省烟台市	SRR8433136
3	小白沙	广东省澄海县白沙农场	SRR8439009
4	鲁花 11	莱阳农学院	SRR8441995
5	海花 1 号	山东省花生研究所	SRR8441996
6	汕油 27	汕头市农业科学研究所	SRR8442014
7	中花 16	中国农业科学院油料作物研究所	SRR8442017
8	闽花 6	福建农林大学作物科学学院	SRR8442644
9	豫花 23	河南省农业科学院经济作物研究所	SRR8442723
10	Sunoleic95	美国	SRR8446494
11	四粒红	辽宁省	SRR8447032
12	远杂 9102	河南省农业科学院经济作物研究所	SRR8759104
13	徐州 68-4	江苏省徐州地区农业科学研究所	SRR8759105

1.2 全基因组变异鉴定

从花生数据库 (<https://peanutbase.org/>) 中下载 Tiffrunner 基因组 Version 2.0 作为参考基因组，使用 Fastp 软件^[16]对测序数据进行质控获得 Clean reads，使用 BWA 软件(<https://bio-bwa.sourceforge.net/>)中的 mem 程序将质控后的数据比对到花生参考基因组，然后使用 samtools 软件(<https://www.htslib.org/>)进行数据格式转换，排序并去除标记 PCR 重复序列。然后使用 GATK 软件 (<https://gatk.broadinstitute.org/hc/en-us>) 中的 HaplotypeCaller 进行变异检测，按照 “QD<2.0 || MQ<40.0 || FS>60.0 || SOR>3.0 || MQRankSum<-12.5 || ReadPosRankSum<-8.0” 进行质控。使用 SnpEff 软件(<https://pcingola.github.io/SnpEff/>)对获得的变异进行注释和统计。

1.3 InDel 标记开发

对获得的 InDel 位点按照插入或缺失 ≥ 10 bp 的标准进行筛选，对筛选后的位点使用 Primer3.0 软件进行引物设计。

1.4 引物质量检测

使用 e-PCR Version: 2.3.9 (<http://www.ncbi.nlm.nih.gov/tools/epcr/>) 对设计的引物进行电子 PCR 检测，参数设置按照 DENG 等^[17]方法进行。分析 InDel 位点在栽培种花生 (Tifrunner) 全基因组中的扩增情况，统计并记录引物在基因组上的扩增次数，剔除扩增产物长度小于 100 bp 或大于 500 bp 及扩增位点超过 3

个的引物。使用 Tbtools^[18]软件对扩增位点在基因组上的位置进行可视化。

1.5 InDel 标记的通用性检验和花生资源的遗传多样性分析

表 2 47 份花生品种（或育种系）的基本信息

Table 2 Basic information of test materials

编号	品种名称	来源	亲本	编号	品种名称	来源	亲本
No.	Name	Origin	Parents	No.	Name	Origin	Parents
1	中花 10 号	中国农业科学院油料作物研究所	四会大株× 86-14007	25	YI6H1	中国热带农业科学院湛江实验站	豫花 40×ICG12625
2	ICG12625	中国农业科学院油料作物研究所	-	26	YI6H23G2	中国热带农业科学院湛江实验站	豫花 40×ICG12625
3	四粒红	吉林省农业科学院	-	27	热花 18	中国热带农业科学院湛江实验站	中花 10×18
4	远杂 9102	河南省农业科学院	白沙 1016× <i>A. chacoense</i>	28	YI6H21	中国热带农业科学院湛江实验站	豫花 40×ICG12625
5	花育 51	山东省农业科学院	P76×鲁花 15 号	29	YI6H30K	中国热带农业科学院湛江实验站	豫花 40×ICG12625
6	中花 16	中国农业科学院油料作物研究所	8130×中花 5 号	30	YI6H21	中国热带农业科学院湛江实验站	豫花 40×ICG12625
7	YI6H29	中国热带农业科学院湛江实验站	豫花 40×ICG12625	31	YI6H13	中国热带农业科学院湛江实验站	豫花 40×ICG12625
8	H51R	中国热带农业科学院湛江实验站	H15002（自然突变）	32	YI6H22K	中国热带农业科学院湛江实验站	豫花 40×ICG12625
9	YI6H2F	中国热带农业科学院湛江实验站	豫花 40×ICG12625	33	YI6H61	中国热带农业科学院湛江实验站	豫花 40×ICG12625
10	YI6H2H	中国热带农业科学院湛江实验站	豫花 40×ICG12625	34	YI6H62	中国热带农业科学院湛江实验站	豫花 40×ICG12625
11	YI6H621	中国热带农业科学院湛江实验站	豫花 40×ICG12625	35	YI6H63	中国热带农业科学院湛江实验站	豫花 40×ICG12625
12	YI6H16	中国热带农业科学院湛江实验站	豫花 40×ICG12625	36	YI6H611	中国热带农业科学院湛江实验站	豫花 40×ICG12625
13	TS23H21	中国热带农业科学院湛江实验站	四粒红（自然突变）	37	YI6H26H	中国热带农业科学院湛江实验站	豫花 40×ICG12625
14	热花 57	中国热带农业科学院湛江实验站	中花 10 号×57	38	热黑 1 号	中国热带农业科学院湛江实验站	中花 10 号×157
15	JH67H	中国热带农业科学院湛江实验站	冀花 7 号（诱变）	39	RH12	中国热带农业科学院湛江实验站	热红 1 号（诱变）
16	YI6H13	中国热带农业科学院湛江实验站	豫花 40×ICG12625	40	热红 1 号	中国热带农业科学院湛江实验站	25-6（诱变）
17	YI6H62	中国热带农业科学院湛江实验站	豫花 40×ICG12625	41	花育 60	山东省农业科学院	开农 49×E16
18	JH67-1	中国热带农业科学院湛江实验站	冀花 7 号（诱变）	42	濮科花 4 号	濮阳市农业科学院	豫花 11 号×濮 9321
19	YI6H1-2	中国热带农业科学院湛江实验站	豫花 40×ICG12625	43	YI6H25	中国热带农业科学院湛江实验站	豫花 40×ICG12625
20	热花 144	中国热带农业科学院湛江实验站	中花 10 号×144	44	YI6H22	中国热带农业科学院湛江实验站	豫花 40×ICG12625
21	YI6H23G	中国热带农业科学院湛江实验站	豫花 40×ICG12625	45	YI6H111	中国热带农业科学院湛江实验站	豫花 40×ICG12625
22	RT1	中国热带农业科学院湛江实验站	冀花甜 2 号（诱变）	46	YI6H1X1	中国热带农业科学院湛江实验站	豫花 40×ICG12625
23	RH11	中国热带农业科学院湛江实验站	热红 1 号（诱变）	47	YI6H1X2	中国热带农业科学院湛江实验站	豫花 40×ICG12625
24	YI6H131	中国热带农业科学院湛江实验站	豫花 40×ICG12625				

- 表示种质资源的亲本未知

- indicates that the parents of germplasm resources are unknown.

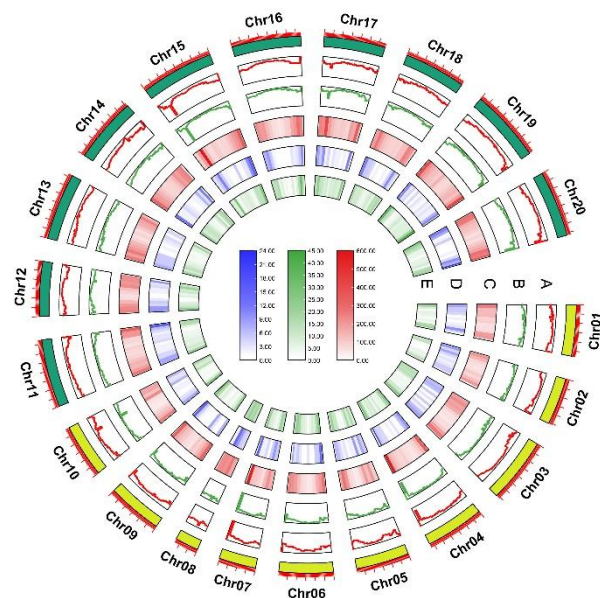
100 对随机选取的引物由生工生物工程（上海）股份有限公司合成（详见 <https://doi.org/10.13430/j.cnki.jpgr.20240206001>,附表 1）。47 份花生品种（或育种系）的信息如表 2 所示，中花 10 号、ICG12625、四粒红、远杂 9102、花育 51、中花 16、花育 60 和濮科花 4 号为本单位收集的亲缘关系较远的品种，其余 39 个为本单位选育的亲缘关系较近的花生育种系，其中 27 个为以豫花 40 为母本、ICG12625 为父本选育的育种系。DNA 提取选择花生幼嫩叶片，采用改良 CTAB 法进行，利用 1%的琼脂糖凝胶电泳检测 DNA 浓度和纯度，于-20℃保存备用。PCR 反应体系和 PCR 程序按照 Huang 等^[8]的方法进行。对 100 对引物首先利用中花 10 号、ICG12625、远杂 9102 和四粒红进行引物筛选，然后利用获得的多态性引物对花生品种(或育种系)进行检测，PCR 产物使用 2.5%的琼脂糖凝胶电泳检测，电压 60 V，电泳 120 min，使用凝胶成像系统拍照保存。带型统计按照第一条带记为 1/1，第二条带记为 2/2，杂合带记为 1/2，缺失记为 -，所有缺失条带进行三次独立 PCR 验证。使用 Powermarker 软件

(<https://brcwebportal.cos.ncsu.edu/powermarker/downloads.htm>)计算主等位基因频率、等位基因数、基因多样性、杂合性和 PIC 信息含量和遗传距离。使用 MEGA 7.0 分析软件, 根据遗传距离进行聚类。利用 Structure 2.3.4 软件 (https://web.stanford.edu/group/pritchardlab/structure_software/release_versions/v2.3.4/html/structure.html) 进行群体结构分析, 通过计算 ΔK 值分布确定最佳群体组群数, 使用 Tbtools 软件^[18]进行可视化。利用 SNPT 软件(<http://www.shigatox.net/stec/cgi-bin/snpt>)筛选最优标记组合, 构建指纹图谱。

2 结果与分析

2.1 花生全基因组变异的鉴定和分布特点

13 个花生品种的重测序数据经过过滤质控后比对到花生参考基因组, 测序深度均超过 10×, 平均比率为 96.74%, 表明 13 个品种的测序质量良好。经过严格的质控和检测, 分别获得 313432 和 38777 个高质量 SNP 和 InDel 位点, SNP 和 InDel 位点在基因组上的密度和分布如图 1 所示。SNP 在基因组上的平均密度为 123 个/Mb, 其中 73.80% (231311) 为转换 (Ts, Transitions; G/A 和 C/T) 和 26.20% (82121) 为颠换 (Tv, Transversions; A/C, A/T, C/G 和 G/T), Ts/Tv 比值为 2.82 (图 2A)。InDel 在基因组上的平均密度为 15.28 个/Mb, 其中 53.94% (20915) 为插入型, 46.06% (17862) 为缺失型; InDel 的长度主要分布在 1~20 bp 之间 (96.74%), 随着插入或缺失的核苷酸增多, 数量逐渐减少, 其中单个核苷酸插入或缺失的 InDel 数量最多, 占总数的 58.40% (22548) (图 2B)。根据参考基因组的基因结构和注释, InDel 和 SNP 主要分布于基因间区域, 分别占 52.35% (33473) 和 60.08% (286276) 的, 其次是上游基因和下游基因序列内, 共占 36.12% (23098) 和 32.49% (154818), 小部分位于基因内区, 分别占 7.90% (5050) 和 4.38% (20806), 其中在外显子上共检测到 761 个 InDel (1.19%) 和 8741 个 SNP (1.18%) (图 3)。



A 和 B 分别表示 SNP 和 InDel 在花生基因组上的分布密度曲线, 按照 1Mb 滑动窗口进行统计; CDE 分别表示全部 InDel、插入型和缺失型 InDel 在花生基因组上的分布, 按照 10Mb 滑动窗口进行统计

A and B represent the distribution density curves of SNP and InDel loci on peanut genome respectively, according to the 1Mb sliding windows. C, D and E represent the distribution of InDel markers, delete type markers and insert type markers on chromosomes respectively, according to the 1Mb sliding windows

图 1 花生全基因组变异的分布

Fig. 1 Density and distribution of peanut SNP/InDel markers on each chromosome of peanut

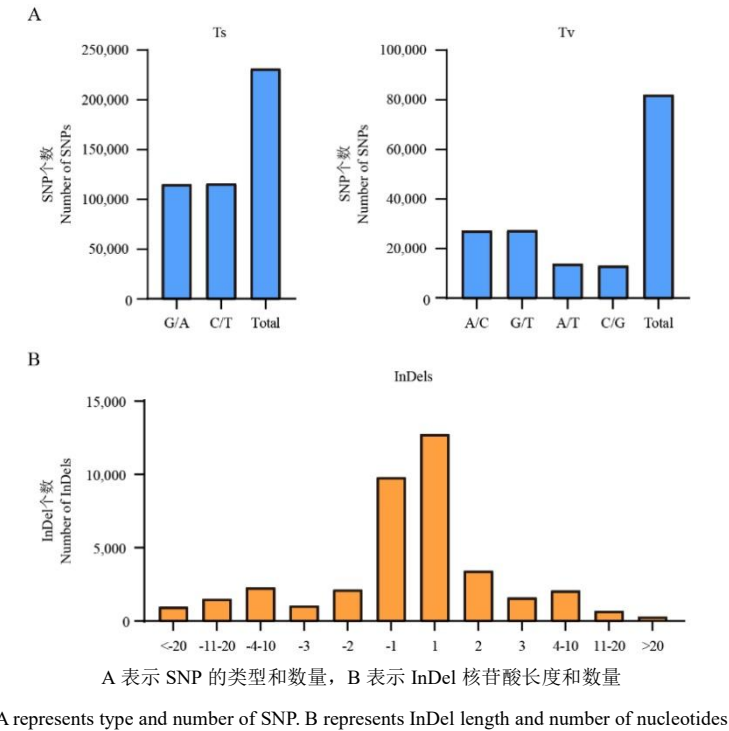


图 2 SNPs/InDels 分类和数量

Fig. 2 SNPs/InDels classification and number

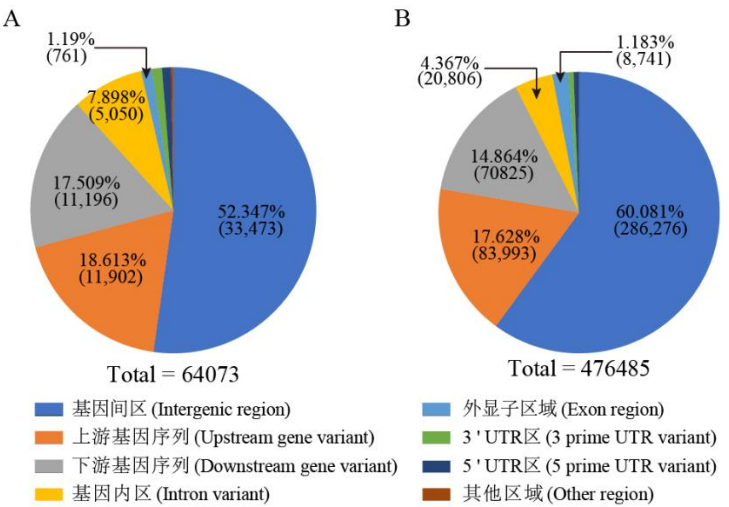


图 3 结构变异发生在基因组不同区域的比例

Fig. 3 Proportion of structural variation occurring in different regions of the genome

2.2 InDel 标记的开发

对获得的 InDel 位点按照插入或缺失长度 ≥ 10 bp 进行筛选，共获得了 4082 个 InDel，对获得的 InDel 位点进行引物设计，发现共有 3675 个 InDel 位点可以进行引物设计。可进行引物设计的 InDel 位点在 20 条染色体上不均匀分布，其中 Chr20 染色体上分布最多，为 233 个；缺失型范围为 64 (Chr02)~190 (Chr19)，共 2606 个 (70.90%)；插入型范围为 38 (Chr08)~77 (Chr20)，共 1069 个 (29.10%)；分布密度范围为 1.08 (Chr02)~2.45 (Chr08)，平均 1.48 个/Mb (表 3)。

表 3 可进行引物设计的 InDel 位点在染色体上的数量统计

Table 3 Statistics of primer design InDel loci on chromosomes

染色体 Chromosome	InDel 标记 InDel Marker			标记平均密度 Average marker density
	缺失型	插入型	合计	
	Insert type	Delete type	Total	
Chr01	97	45	142	1.26
Chr02	64	48	112	1.08
Chr03	150	63	213	1.49
Chr04	148	60	208	1.61
Chr05	130	54	184	1.58
Chr06	151	47	198	1.66
Chr07	90	43	133	1.63
Chr08	88	38	126	2.45
Chr09	109	44	153	1.27
Chr10	106	42	148	1.26
Chr11	125	70	195	1.31
Chr12	124	54	178	1.48
Chr13	119	48	167	1.14
Chr14	150	55	205	1.43
Chr15	165	65	230	1.44
Chr16	148	53	201	1.33
Chr17	158	51	209	1.56
Chr18	138	53	191	1.41
Chr19	190	59	249	1.56
Chr20	156	77	233	1.61
合计 Total	2606	1069	3675	1.48

2.3 InDel 标记质量检测

以栽培种花生 Tifrunner 基因组为模板进行电子 PCR 检测，结果表明 InDel 标记在花生基因组上具有较高的扩增效率，2834（77.12%）对引物可以在基因组上扩增出 6005 位点，扩增片段长度为 39~599 bp，扩增的位点主要以单个位点为主(57.31%)，其次是 2 个位点（9.20%）（表 4）。扩增位点数为 1~3 的引物为有效引物，共 2561 对（69.69%），可检测出 3133 个（52.71%）有效位点，其中 1045 对（40.80%）引物仅能在 A 亚基因组（Chr01~Chr10）上扩增，1253 对（48.93%）引物仅能在 B 亚基因组（Chr11~Chr20）上扩增，263 对（10.27%）引物在 A 和 B 亚基因组上均能扩增。根据有效扩增位点在基因组上的位置，绘制了 InDel 标记扩增位点图谱，为 InDel 标记扩增位点的基因组位置信息提供参考（图 4）。

表 4 InDel 引物 e-PCR 扩增位点统计

Table 4 Statistics of InDel primers amplified in peanut genome by e-PCR

DNA 模板 DNA template	扩增位点 Amplified loci	有效扩增位点 Effective amplified loci	扩增引物数 Number of amplified primers	引物扩增位点统计 Primer amplified loci statistics				有效引物数 Number of effective primer pairs
				1	2	3	>3	
Tifrunner ^a	6005	3133 (52.17%) ^b	2834 (77.12%) ^c	2106 (57.31%) ^d	338 (9.20%) ^d	117 (3.18%) ^d	273 (7.43%) ^d	2561 (69.69%) ^e

a 表示 DNA 模板使用的是 Tifrunner 的基因组，b 表示扩增产物长度为 100~500bp 位点在所有 6005 扩增位点中的比例，c 表示能够扩增出条带的引物

在全部引物中的比例；d 表示扩增位点数为 1，2，3 和>3 的引物在全部引物中的比例，e 表示有效引物在全部引物中的比例

a represents the DNA template used genome of Tifrunner. b represents the proportion of sites with 100-500bp amplification product length in all 6005 amplification sites. c represents the proportion of primers that could amplify products in all the primers. d represents the proportion of primers with amplification site number 1, 2, 3 and more than 3 in all primers. e represents the proportion of effective primers in all the primers

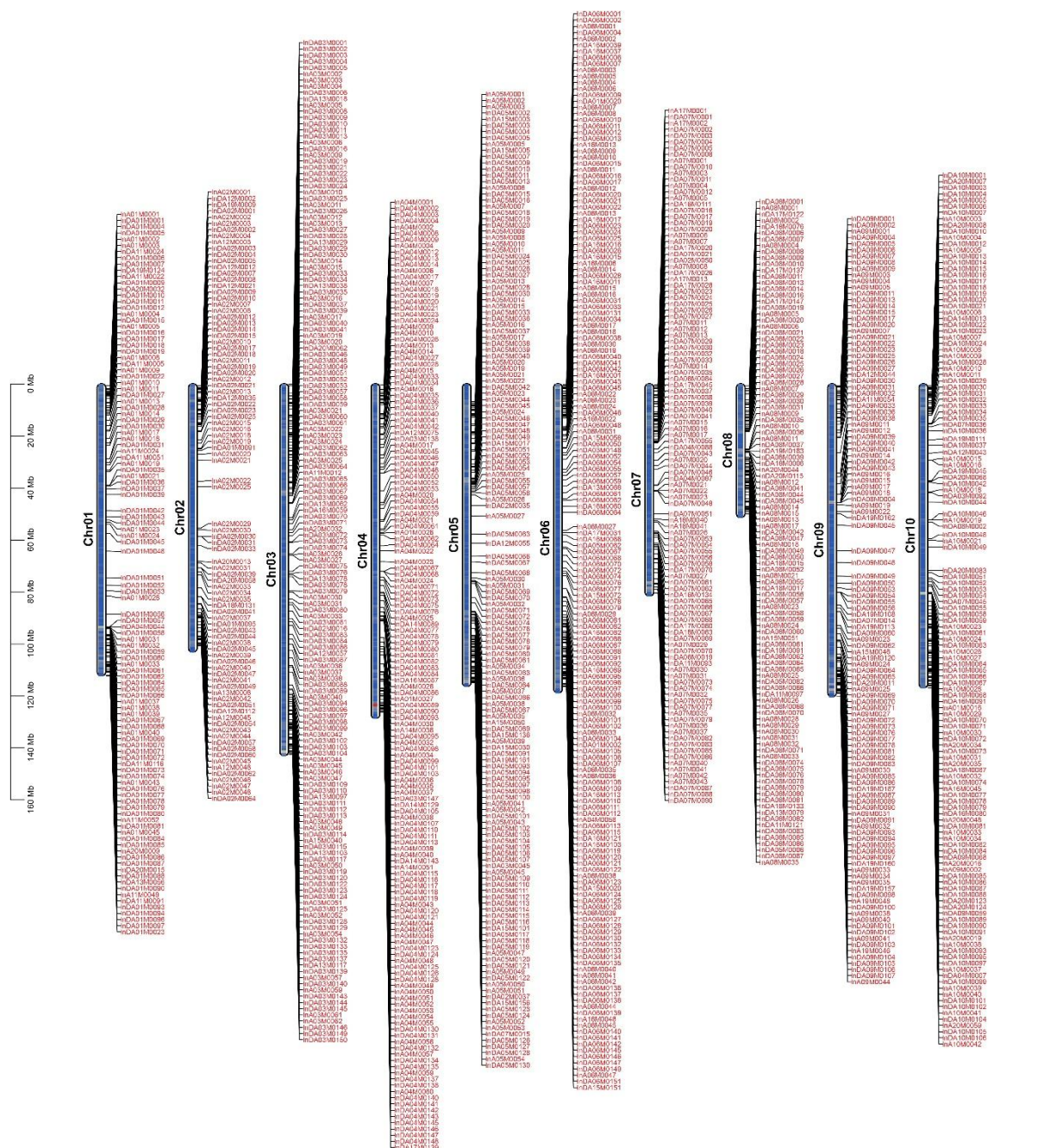


图 4 InDel 标记在花生基因组上的扩增位点图谱（部分）

Fig. 4 Physical map of amplified loci of InDel markers in peanut genome (part)

2.4 InDel 标记多态性的验证

从开发的 InDel 标记中随机挑选了 100 对引物，对中花 10 号、ICG12625、远杂 9102 和四粒红 4 个亲缘关系较远的花生品种进行 PCR 扩增表明，100 对引物均可扩增出清晰的条带，其中 31 对引物在 4 个品种间扩增出差异条带，差异引物均扩增出 2 个条带，表明开发的引物具有较好的扩增效率和分辨率。利用筛选的 31 个多态性 InDel 标记对 47 份花生品种（或育种系）进行检测，结果表明 31 个 InDel 标记多态性存在差异（图 5，表 5）。31 个标记共检测到 62 个等位变异，每个标记可检测到 2~3 种基因型；主基因频

率范围为 0.51~0.98，平均为 0.77；基因多样性范围为 0.04~0.50，平均为 0.30；杂合度范围为 0~0.58，平均为 0.11；PIC 信息量范围为 0.04~0.37，平均为 0.24。

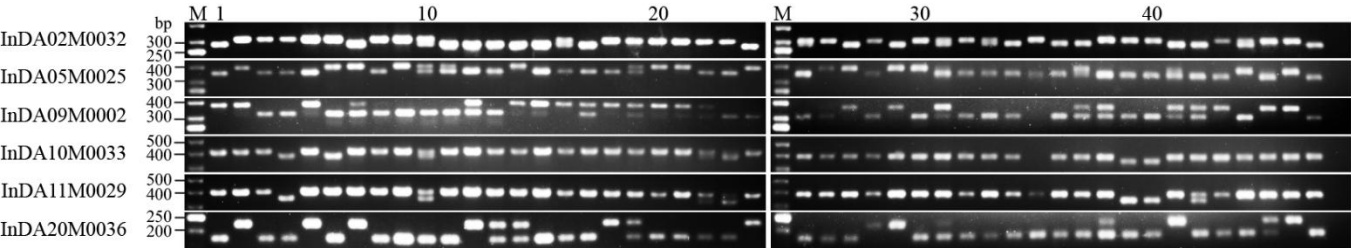


图 5 InDel 标记在 47 个花生品种（或育种系）中的扩增情况（部分）

Fig. 5 Amplification of InDel markers in 47 peanut cultivars (or breeding-lines) (part)

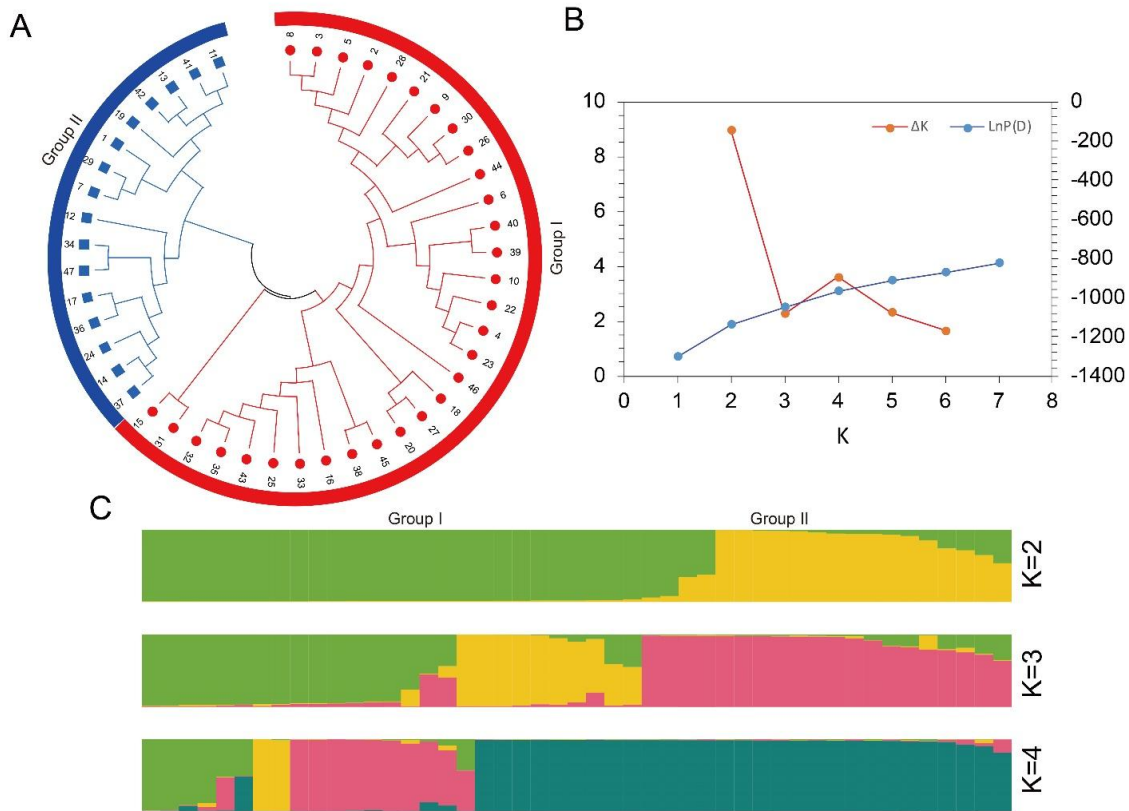
表 5 31 个 InDel 标记的遗传参数

Table 5 Genetic parameters of 31 InDel markers

标记	主基因频率	等位基因数	基因型数	基因多样性	杂合度	多态性信息量
Marker	MAF	Allele No	Genotype No	Gene Diversity	Heterozygosity	PIC
InDA01M0046	0.61	2.00	2.00	0.48	0.00	0.36
InDA01M0057	0.56	2.00	3.00	0.49	0.13	0.37
InDA02M0012	0.51	2.00	3.00	0.50	0.19	0.37
InDA02M0032	0.55	2.00	3.00	0.49	0.17	0.37
InDA03M0011	0.95	2.00	2.00	0.09	0.00	0.08
InDA03M0022	0.90	2.00	3.00	0.18	0.07	0.16
InDA03M0034	0.95	2.00	2.00	0.09	0.00	0.08
InDA04M0001	0.98	2.00	2.00	0.04	0.00	0.04
InDA04M0053	0.96	2.00	2.00	0.08	0.00	0.08
InDA05M0025	0.65	2.00	3.00	0.46	0.11	0.35
InDA06M0001	0.94	2.00	3.00	0.10	0.02	0.10
InDA06M0020	0.68	2.00	3.00	0.44	0.20	0.34
InDA08M0007	0.95	2.00	3.00	0.09	0.05	0.08
InDA08M0058	0.86	2.00	3.00	0.24	0.02	0.21
InDA09M0002	0.59	2.00	3.00	0.48	0.22	0.37
InDA10M0033	0.88	2.00	3.00	0.21	0.07	0.19
InDA11M0029	0.90	2.00	3.00	0.17	0.06	0.16
InDA16M0148	0.96	2.00	2.00	0.08	0.00	0.08
InDA20M0036	0.72	2.00	3.00	0.40	0.13	0.32
InDA01M0001	0.65	2.00	3.00	0.45	0.12	0.35
InDA02M0043	0.51	2.00	3.00	0.50	0.26	0.37
InDA05M0013	0.96	2.00	2.00	0.08	0.00	0.08
InDA10M0002	0.87	2.00	3.00	0.23	0.24	0.21
InDA10M0018	0.70	2.00	2.00	0.42	0.00	0.33
InDA12M0101	0.51	2.00	3.00	0.50	0.31	0.37
InDA18M0002	0.70	2.00	2.00	0.42	0.00	0.33
InDA18M0018	0.55	2.00	3.00	0.50	0.58	0.37
InDA18M0045	0.89	2.00	2.00	0.20	0.00	0.18
InDA19M00161	0.96	2.00	2.00	0.08	0.00	0.08
InDA19M0190	0.81	2.00	3.00	0.31	0.11	0.26
InDA20M0002	0.57	2.00	3.00	0.49	0.39	0.37
平均 Average	0.77	2.00	2.65	0.30	0.11	0.24

2.5 花生品种（或育种系）聚类分析、群体结构分析和指纹图谱构建

基于遗传距离的聚类分析表明，利用 31 个 InDel 标记可以有效的将 47 份材料区分开。47 份花生品种（或育种系）在遗传距离为 0.25 时可以分为两个类群，其中类群 I 共有 31 份材料，包括四粒红、花育 51、中花 16、远杂 9102 和 ICG12625 5 份品种和 26 份育种系；类群 II 包含 16 份材料，包括中花 10 号、濮科花 4 号和花育 60 3 份品种及 13 个育种系(图 6A)。以 ICG12625 为父本的 27 份育种系，有 17 份与 ICG12625 聚在类群 I，10 份位于类群 II，表明 InDel 标记可以有效区分亲缘关系较近的花生材料。群体结构分析表明，随着 K 值增加，LnP(D)持续增加，无明显拐点； ΔK 在 K=2 是有最大值，表明 47 份材料的最佳分群为 2 个类群，分别包含 32 份材料和 15 份材料，群体结构与聚类结果基本一致（图 6B 和图 6C）。利用 SNPT 软件进行标记筛选发现，最少用 7 个标记（InDA01M0001、InDA01M0046、InDA02M0012、InDA06M0020、InDA09M0002、InDA12M0101、InDA20M0002）就可将 47 份材料区分开，表明 InDel 标记具有较高的分辨率。利用这些标记构建了 47 份的材料的 InDel 指纹图谱（图 7）。

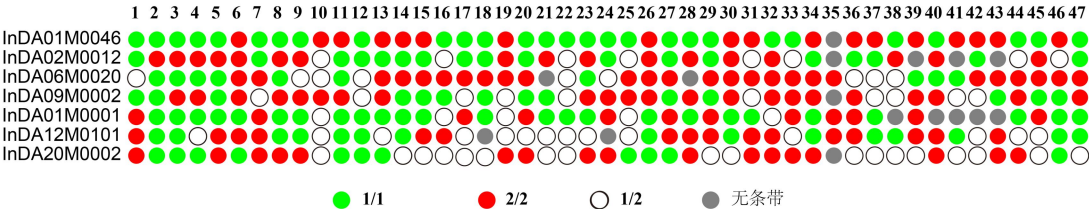


A, 47 个花生品种（或育种系）的聚类图；B, ΔK 和 LnP(D)的变化趋势；C, 47 个花生品种（或育种系）在 K=2、3、4 时的的群体结构

A, the cluster tree of the 47 peanut cultivars (or breeding-lines). B, the distribution of ΔK and LnP(D). C, the population structure of the 47 peanut cultivars (or breeding-lines) with the K value of 2, 3 and 4

图 6 47 个花生品种（或育种系）的聚类分析和群体结构分析

Fig. 6 Cluster analysis and population structure analysis of the 47 peanut cultivars (or breeding-lines)



绿色圆点表示检测的基因型 1，红色圆点表示检测的基因型 2，空白圆点表示基因型 1 和 2 的杂合型，灰色圆点表示无条带缺失型

The green dot represents the detected genotype 1, the red dot represents the detected genotype 2, the blank dot represents the heterozygous genotype 1 and 2, and the gray dot represents the no banding and no genotypes detected.

图 7 47 个花生品种（或育种系）的 InDel 指纹图谱

Fig. 7 The fingerprint of the 47 peanut cultivars (or breeding-lines)

3 讨论

分子标记是开展作物遗传研究和高效育种的重要工具，自 9 个花生基因组公布以来，大量基于花生全基因组的序列变异被鉴定出来并用于分子标记开发，主要表现在花生 SSR 位点的鉴定和开发上，如 Luo 等^[19]从野生花生基因组 *A. duranensis* 和 *A. ipaensis* 中分别鉴定了 264135 和 392107 个 SSR 位点，利用鉴定的位点开发了 84383 和 120056 个 SSR 标记；Lu 等^[20]从栽培种花生伏花生基因组中鉴定了 8329496 个 SSR 位点，并新开发了 973394 个新的 SSR 标记，其中 44.15% 的可以实际检测到；徐志军等^[21]从花生的全转录组数据中鉴定到 19143 个 SSR 位点，开发了 5859 个转录组 SSR 标记；王玉龙等^[22]从四倍体野生花生基因组中鉴定了 676878 个 SRR 位点，设计了 192303 对 SSR 引物，极大的丰富了花生的分子标记。但 SSR 标记的多态率通常在 10%~30% 之间，带型复杂，读带困难，需要借助聚丙烯酰胺凝胶电泳或者价格昂贵的毛细管电泳检测进行基因分型。与 SSR 标记相比，InDel 标记在基因组中分布广，密度大，丰富性远高于 SSR，且具有多态性高、稳定性好、分型简单的特点。本研究利用 13 个花生品种重测序数据进行了全基因组变异检测，共鉴定到 313432 和 38777 个高质量 SNP 和 InDel 位点，利用插入或缺失 ≥ 10 bp 的 3675 个 InDel 位点首次在全基因组水平上进行了 InDel 引物设计和标记质量检测，获得了 2561 个 InDel 标记，进一步丰富了花生标记，为基于分子标记的遗传研究提供了可利用的资源。且所开发的 InDel 标记主要为 1~3 个扩增位点，特异性高，易于分型，具有较好的准确性。

在花生研究上，InDel 标记多应用于 QTL 的精细定位研究中，如种皮色^[23]、开花习性^[24]、根部结瘤性状^[15]和含油量 QTL^[14]的精细定位，而在花生品种鉴定和遗传多样性方面的研究还较少。本研究从 100 对随机合成引物中筛选到 31 对多态性引物，在 47 个花生品种（或育种系）中共检测到 62 个等位变异；主基因频率范围为 0.51~0.98，平均为 0.77；基因多样性范围为 0.04~0.50，平均为 0.30；杂合度范围为 0~0.58，平均为 0.11；PIC 信息量范围为 0.04~0.37，平均为 0.24，表明 47 个花生品种（或育种系）的遗传多样性较低，这可能与本研究使用了 39 份亲缘关系较近的育种系有关。白冬梅等^[25]利用 90 对 SSR 对 72 份山西地方种质检测表明，共检测到 317 个等位变异，变化范围为 2~8 个，SSR 位点的主基因频率变化幅度为 0.29~0.90，平均为 0.68；基因多样性指数变化幅度为 0.18~0.77，平均为 0.45；PIC 变异范围为 0.17~0.74，平均为 0.40。与 SSR 标记相比，InDel 标记由于扩增位点较少，基因多样性和 PIC 信息含量较低，刘宇等^[26]利用 21 个花生品种进行 InDel 标记筛选也有类似发现。聚类分析和群体遗传结构分析能反映出材料间基因交流和渗入情况，能更好地说明材料间的遗传结构，有助于育种者准确地掌握种质间的遗传关系。本研究中 8 个品种和 39 个育种系在可以聚为 2 个类群，聚类结果与群体结构分析结果基本一致，来源于 ICG12625 的 27 个育种系可以分布在 2 个类群中，促进了对这些育种系的分子评估。在遗传研究中，亲缘关系较近的

材料通常难于区分,在本研究中 InDel 标记在亲缘关系较近的材料中也表现出较好的多态性,利用 7 个 InDel 标记就可区分使用的全部材料,为利用 InDel 标记鉴定亲缘关系较近的花生材料提供了依据。综上所述,InDel 标记在研究花生品种的遗传多样性和品种鉴定方面均具有较高的效率,可以广泛的应用于花生的遗传研究。

参考文献

- [1] 廖伯寿. 我国花生生产发展现状与潜力分析. 中国油料作物学报, 2020, 42 (2): 161-166
- Liao B C. A review on progress and prospects of peanut industry in China. Chinese Journal of Oil Crop Sciences, 2020, 42 (2): 161-166
- [2] Bertoli D J, Cannon S B, Froenicke L, Huang G, Farmer AD, Cannon EK, Liu X, Gao D, Clevenger J, Dash S, Ren L, Moretzsohn M C, Shirasawa K, Huang W, Vidigal B, Abernathy B, Chu Y, Niederhuth C E, Umale P, Araújo A C, Kozik A, Kim K D, Burow M D, Varshney R K, Wang X, Zhang X, Barkley N, Guimarães PM, Isobe S, Guo B, Liao B, Stalker HT, Schmitz RJ, Scheffler B E, Leal-Bertioli S C, Xun X, Jackson SA, Micheltore R, Ozias-Akins P. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nature Genetics, 2016, 48(4): 438-446.
- [3] Yin D, Ji C, Ma X, Li H, Zhang W, Li S, Liu F, Zhao K, Li F, Li K, Ning L, He J, Wang Y, Zhao F, Xie Y, Zheng H, Zhang X, Zhang Y, Zhang J. Genome of an allotetraploid wild peanut *Arachis monticola*: a de novo assembly. Gigascience, 2018, 7(6): giy066.
- [4] Bertoli D J, Jenkins J, Clevenger J, Dudchenko O, Gao D, Seijo G, Leal-Bertioli S C M, Ren L, Farmer A D, Pandey M K, Samoluk S S, Abernathy B, Agarwal G, Ballén-Taborda C, Cameron C, Campbell J, Chavarro C, Chitkineni A, Chu Y, Dash S, El Baidouri M, Guo B, Huang W, Kim K D, Korani W, Lanciano S, Lui CG, Mirouze M, Moretzsohn MC, Pham M, Shin JH, Shirasawa K, Sinharoy S, Sreedasyam A, Weeks N T, Zhang X, Zheng Z, Sun Z, Froenicke L, Aiden EL, Micheltore R, Varshney R K, Holbrook C C, Cannon E K S, Scheffler B E, Grimwood J, Ozias-Akins P, Cannon S B, Jackson S A, Schmutz J. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. Nature Genetics, 2019, 51(5): 877-884.
- [5] Chen X, Lu Q, Liu H, Zhang J, Hong Y, Lan H, Li H, Wang J, Liu H, Li S, Pandey MK, Zhang Z, Zhou G, Yu J, Zhang G, Yuan J, Li X, Wen S, Meng F, Yu S, Wang X, Siddique KHM, Liu ZJ, Paterson AH, Varshney R K, Liang X. Sequencing of cultivated peanut, *Arachis hypogaea*, yields insights into genome evolution and oil improvement. Molecular Plant, 2019, 12(7): 920-934.
- [6] Zhuang W, Chen H, Yang M, Wang J, Pandey MK, Zhang C, Chang WC, Zhang L, Zhang X, Tang R, Garg V, Wang X, Tang H, Chow CN, Wang J, Deng Y, Wang D, Khan AW, Yang Q, Cai T, Bajaj P, Wu K, Guo B, Zhang X, Li J, Liang F, Hu J, Liao B, Liu S, Chitkineni A, Yan H, Zheng Y, Shan S, Liu Q, Xie D, Wang Z, Khan SA, Ali N, Zhao C, Li X, Luo Z, Zhang S, Zhuang R, Peng Z, Wang S, Mamadou G, Zhuang Y, Zhao Z, Yu W, Xiong F, Quan W, Yuan M, Li Y, Zou H, Xia H, Zha L, Fan J, Yu J, Xie W, Yuan J, Chen K, Zhao S, Chu W, Chen Y, Sun P, Meng F, Zhuo T, Zhao Y, Li C, He G, Zhao Y, Wang C, Kavikishor P B, Pan R L, Paterson A H, Wang X, Ming R, Varshney R K. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nature Genetics, 2019, 51(5): 865-876.
- [7] Newman C S, Andres R J, Youngblood RC, Campbell J D, Simpson S A, Cannon SB, Scheffler B E, Oakley A T, Hulse-Kemp A M, Dunne J C. Initiation of genomics-assisted breeding in Virginia-type peanuts through the generation of a de novo reference genome and informative markers. Frontiers in Plant Sciences, 2023, 13: 1073542.
- [8] Huang L, Ren X, Wu B, Li X, Chen W, Zhou X, Chen Y, Pandey M K, Jiao Y, Luo H, Lei Y, Varshney R K, Liao B, Jiang H. Development and deployment of a high-density linkage map identified quantitative trait loci for plant height in peanut (*Arachis hypogaea* L.). Scientific Reports, 2016, 6: 39478.
- [9] 姜慧芳, 任小平, 张晓杰, 黄家权, 雷永, 晏立英, 廖伯寿, Upadhyaya H D, Holbrook C C. 中国花生小核心种质与 ICRISAT 微核心种质的 SSR 遗传多样性比较. 作物学报, 2010, 36 (7): 1084-1091
- Jiang H F, Ren X P, Zhang X J, Huang J Q, Lei Y, Yan L Y, Liao B S, Upadhyaya H D, Holbrook C C. Comparison of genetic diversity between peanut mini core collections from China and ICRISAT by SSR markers. Acta Agronomica Sinica, 2010, 36 (7): 1084-1091

- [10] Luo H, Pandey M K, Khan A W, Guo J, Wu B, Cai Y, Huang L, Zhou X, Chen Y, Chen W, Liu N, Lei Y, Liao B, Varshney R K, Jiang H. Discovery of genomic regions and candidate genes controlling shelling percentage using QTL-seq approach in cultivated peanut (*Arachis hypogaea* L.). Plant Biotechnology Journal, 2019, 17(7): 1248-1260.
- [11] Varshney R K, Pandey M K, Janila P, Nigam S N, Sudini H, Gowda M V, Sriswathi M, Radhakrishnan T, Manohar S S, Nagesh P. Marker-assisted introgression of a QTL region to improve rust resistance in three elite and popular varieties of peanut (*Arachis hypogaea* L.). Theoretical and Applied Genetics, 2014, 27(8): 1771-1781.
- [12] 王娟, 刘宇, 李春娟, 闫彩霞, 赵小波, 单世华. 基于简化基因组的花生 InDel 标记开发和功能解析. 植物遗传资源学报, 2019, 20(1): 179-187.
- Wang J, Liu Y, Li C J, Yan C X, Zhao X B, Shan S H. Development and functional analysis of peanut insertion and deletion (InDel) markers based on genotyping-by-sequencing (GBS). Journal of Plant Genetic Resources, 2019, 20(1): 179-187.
- [13] 刘宇, 李春娟, 石大川, 闫彩霞, 赵小波, 孔青, 孙全喜, 苑翠玲, 王娟, 单世华. 利用 InDel 标记解析中国花生地方品种的遗传多样性与群体结构. 中国油料作物学报, 2020, 42(5): 743-752.
- Liu Y, Li C J, Shi D C, Yan C X, Zhao X B, Kong Q, Sun Q X, Yuan C L, Wang J, Shan S H. Population genetic diversity and structure of Chinese peanut landraces analyzed by InDel markers. Chinese Journal of Oil Crop Sciences, 2020, 42(5): 743-752.
- [14] 陶顺玉, 吴贝, 刘念, 罗怀勇, 黄莉, 周小静, 陈伟刚, 郭建斌, 喻博伦, 雷永, 廖伯寿, 姜慧芳. 花生 InDel 标记开发及其在含油量 QTL 定位中的应用. 作物学报, 2023, 49(5): 1222-1230.
- Tao S Y, Wu B, Liu N, Luo H Y, Huang L, Zhou X J, Chen W G, Guo J B, Yu B L, Lei Y, Liao B S, Jiang H F. Development and employment of InDel marker in peanut QTL mapping of oil content. Acta Agronomica Sinica, 2023, 49(5): 1222-1230.
- [15] 黄莉, 陈伟刚, 李威涛, 喻博伦, 郭建斌, 周小静, 罗怀勇, 刘念, 雷永, 廖伯寿, 姜慧芳. 花生根部结瘤性状 QTL 定位. 作物学报, 2023, 49(8): 2097-2104.
- Huang L, Chen W G, Li W T, Yu B L, Guo J B, Zhou X J, Luo H Y, Liu N, Lei Y, Liao B S, Jiang H F. Identification of major QTLs for nodule formation in peanut. Acta Agronomica Sinica, 2023, 49(8): 2097-2104.
- [16] Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics, 2018, 34(17): i884-i890.
- [17] Deng P, Wang M, Feng K, Cui L, Tong W, Song W, Nie X. Genome-wide characterization of microsatellites in Triticeae species: abundance, distribution and evolution. Scientific Reports, 2016, 6: 32224.
- [18] Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, Liu Y, Feng J, Chen H, He Y, Xia R. TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. Molecular Plant, 2023, 16(11): 1733-1742.
- [19] Luo H, Xu Z, Li Z, Li X, Lv J, Ren X, Huang L, Zhou X, Chen Y, Yu J, Chen W, Lei Y, Liao B, Jiang H. Development of SSR markers and identification of major quantitative trait loci controlling shelling percentage in cultivated peanut (*Arachis hypogaea* L.). Theoretical and Applied Genetics, 2017, 130(8): 1635-1648.
- [20] Lu Q, Hong Y, Li S, Liu H, Li H, Zhang J, Lan H, Liu H, Li X, Wen S, Zhou G, Varshney R K, Jiang H, Chen X, Liang X. Genome-wide identification of microsatellite markers from cultivated peanut (*Arachis hypogaea* L.). BMC Genomics, 2019, 20(1): 799.
- [21] 徐志军, 赵胜, 徐磊, 胡小文, 安东升, 刘洋. 基于 RNA-seq 数据的栽培种花生 SSR 位点鉴定和标记开发. 中国农业科学, 2020, 53(4): 695-706.
- Xu Z J, Zhao S, Xu L, Hu X W, An D S, Liu Y. Discovery of microsatellite markers from RNA-seq data in cultivated peanut (*Arachis hypogaea*). Scientia Agricultura Sinica, 2020, 53(4): 695-706.
- [22] 王玉龙, 黄冰艳, 王思雨, 杜培, 齐飞艳, 房元瑾, 孙子淇, 郑崢, 董文召, 张新友. 四倍体野生种花生 *A. monticola* 全基因组 SSR 的开发与特征分析. 中国农业科学, 2019, 52(15): 2567-2580.

Wang Y L, Huang B Y, Wang S Y, Du P, Qi F Y, Fang Y J, Sun Z Q, Zheng Z, Dong W Z, Zang X Y. Development and characterization of whole genome SSR in tetraploid wild peanut (*Arachis monticola*). Scientia Agricultura Sinica, 2019, 52(15): 2567-2580.

[23] Zhang K, Yuan M, Xia H, He L, Ma J, Wang M, Zhao H, Hou L, Zhao S, Li P, Tian R, Pan J, Li G, Thudi M, Ma C, Wang X, Zhao C. BSA-seq and genetic mapping reveals *AhRt2* as a candidate gene responsible for red testa of peanut. Theoretical and Applied Genetics, 2022, 135(5): 1529-1540.

[24] Pan J, Zhou X, Ahmad N, Zhang K, Tang R, Zhao H, Jiang J, Tian M, Li C, Li A, Zhang X, He L, Ma J, Li X, Tian R, Ma C, Pandey MK, Varshney R K, Wang X, Zhao C. BSA-seq and genetic mapping identified candidate genes for branching habit in peanut. Theoretical and Applied Genetics, 2022, 135(12): 4457-4468.

[25] 白冬梅, 薛云云, 张鑫, 王鹏冬, 黄莉, 田跃霞, 姜慧芳. 基于 SSR 标记分析山西地方花生种质遗传多样性. 中国油料作物学报, 2020, 42(5): 753-759.

Bai D M, Xue Y, Zhang X, Wang P, Huang L, Tian Y, Jiang H. Genetic diversity of peanut local germplasm resources in Shanxi Province based on SSR markers. Chinese Journal of Oil Crop Sciences, 2020, 42(5): 753-759.

[26] 刘宇, 闫彩霞, 李春娟, 徐洪明, 孔青, 孙全喜, 王娟, 单世华. 花生栽培种 InDel 有效标记筛选与评估. 核农学报, 2020, 34(2): 256-264.

Liu Y, Yan C X, Li C J, Xu K M, Kong Q, Sun Q X, Wang J, Shan S H. Screening and Application of InDel Markers in Cultivated Peanut (*Arachis hypogaea* L.). Journal of Nuclear Agricultural Sciences, 2020, 34(2): 256-264.

附表 1 100 对随机 InDel 引物的序列信息

Supplemental Table 1 Sequence information of 100 random InDel primers

序号	标记名称	染色体	正向引物序列(5'-3')	反向引物序列(5'-3')
Number	Marker	Chromosome	Forward primers sequence (5'-3')	Reverse primers sequence (5'-3')
1	InDA01M0001	1	GCGTCAGTCTAGGACCTTCC	AGTGTGGGCGGTTTAATGT
2	InDA01M0016	1	TGTAGCGTGTTATTGCGTGT	CCCAATGAGAAAGTGACACACA
3	InDA01M0031	1	GTAACAGACGCCGAGAAGGT	ACACCGGTGGGATTGAAACT
4	InDA01M0046	1	AGGAGGAGAGCACATTGAAAA	TGCTTGCTGCTTCTATGCT
5	InDA01M0057	1	GAGATGAACACCAACAACGCA	TCTCTTGACTCTCTGCTCTCA
6	InDA02M0012	2	TTGCCAGTTGCGTTAACGTT	CCACTGTGGTAGGGTGTTACA
7	InDA02M0032	2	CTTCTAGGGCCCACTTGGTG	ATTGGGCCCTGAAGCTCTTC
8	InDA02M0043	2	TGACAAATAGGTCAAAGTCTATCGG	CCCACCAAAGTCCACGGATT
9	InDA02M0048	2	AGCACTAAACAAGCAATAACCACC	TTGCTGGTGATTGCTTGCTT
10	InDA02M0058	2	CACATGGCGGTGATTGTTGGT	GGTAACCTGAAGGTGAGGTAAGC
11	InDA03M0002	3	TCCTGGCTTAAGCGAACGAG	AGCACAGAGGCCTGAATTCC
12	InDA03M0011	3	TGATGTCCTAGGCTCCTAGCT	GGAGTATGTGCACGCAAACCT
13	InDA03M0022	3	ATGGGAGGCTACGGAGGAAT	ACCAACCTTGAGCCTATGCC
14	InDA03M0034	3	CCACAGAAGCACTTTGACGC	CCATTCTGGGCCAAAAACGG
15	InDA03M0053	3	GGGATCAACGGCTGTCAAGA	AGAGAGTAGAGCGGAAGGGA
16	InDA04M0001	4	CGCTAGGGACAGCACTTTTG	GCTTACAGCCGTTTATCCGT
17	InDA04M0025	4	ATCAAGCTCCGGATGCAACC	CGGTGAGCATCTGGAGGTTT
18	InDA04M0037	4	TGTGCTGCTGTGCATTTAGC	ACCGCCGTTAAGTCAAGGAT
19	InDA04M0053	4	ACTCCCTTGATGACAGCACG	TTCGGTCCACGCTTTGCAT
20	InDA04M0076	4	AGCGCGTAAGTTCAAGCTCT	TCTCGTATTCTGTGTTGTATCTCGT
21	InDA04M0105	4	CCCAATCAGCCCCCAAGAAT	TCCTGTGACATCTAAGTCGTTGA
22	InDA05M0001	5	ATGTGGCAACCATCTCGGAG	GTTCTCTGGTGGGGTGATG
23	InDA05M0013	5	GTGTCTCGGTCTATGGCCAA	ACCAGGTATGCTCAGCTTTCA
24	InDA05M0025	5	GAACGCCCAGCAACTACCTA	CCGCTCTATTCTCGATGCA
25	InDA05M0056	5	TGCCTAGTAGTGCTTGAAGGG	AGATCCCAGACCTCAGTCCC
26	InDA05M0099	5	TCGAACCCCGACACTTGTTT	GGTTTAAGGTGGTGTTGTCAACA
27	InDA06M0001	6	AAGGAAGGTCAAGCTACGGC	TGCTTGGTCCAATTCAAACCG
28	InDA06M0020	6	CCAAACCAATTCCCCAACCC	GGTTGGTCGAGTGGTCAACT
29	InDA06M0038	6	ACGCAAGCGAAAAATGTTGGA	TGGATTGGTTTGTCAGCACTG
30	InDA06M0056	6	TTCTTGTCCTCAAGCAAGCA	CCAAAAGGCTGCTGCAACAT
31	InDA06M0078	6	AGCCAGTTTAAAAAGCTCCCG	GTTCTGTTGGGCTGGTCTGA
32	InDA07M0009	7	ACCCCAAACGTTGAGGACAA	TGGAGAGCCACAAAATTGGA
33	InDA07M0025	7	AGGATAACGGTCATCCCCCT	TAGGTGGAATAGGGGACGGG
34	InDA07M0045	7	TGGGTTGTTTCGGTATTGGGA	AGCATCATCATCAGCATCACCA
35	InDA07M0064	7	AGGAGACGGATGAAGAGGCT	TGCGCCACTCTGATAGCAAA
36	InDA07M0090	7	GGCTGCTTTGCTTTGGTCTC	AGGGTGTTGCTGTGAGTCTA
37	InDA08M0007	8	CCCGGCCTGTGGACTTTTAT	TCAGATGACGGTGTTGAGCC
38	InDA08M0058	8	GAGTGTGTGTGGTTGGAGGA	CATGGGTCGGACTAGCTTGG
39	InDA08M0065	8	ATGTGGTGCCAGTTTtagct	ACCAGCATGTAAAAATGGAGTTCG
40	InDA08M0076	8	CTAATCGAACCGGCCAGACA	TTTGGGTTTGGGCCAACTTG

41	InDA08M0088	8	TGTGCCATGACAGCTCAACT	GGTATTGTTGCTGGTTGCTGG
42	InDA09M0002	9	TGCCCTTAAAGCGTGTGTGA	TGGCCCTCTCAGCTCTATGT
43	InDA09M0011	9	GAGTTGCTCGGGGGAATCAA	CCTTTTCGATCCTGTCCGCT
44	InDA09M0024	9	CAACAACCTCCGAGCAGACCT	CTCTCGAAGGTCTGATCGGC
45	InDA09M0041	9	TGCGTGGCATGTGTATCTCA	TGCAGTTTTTCGACGGTTCAG
46	InDA09M0109	9	TCTAGGGGTTTTCTGGGGGT	ACAGAAAAACGCTAAAAAGCCCT
47	InDA10M0002	10	TGGTCCAGTGAATAGGATTAACCA	TGCTTGTTCTCATGGCAGTG
48	InDA10M0018	10	TGTTGATGGCAATGTTGGAGA	AGGTCGCCTCATGTTTTCGT
49	InDA10M0033	10	ACGCGTACAGGTACAGTGAA	CCTGGGCTCCTGCTATGAAG
50	InDA10M0081	10	TTAGCGACGCTGATGTGTGT	GTGTGATTATTCGTCCGCGC
51	InDA10M0106	10	ACATCATTCGAGGTCACGAGA	AGAGTTTACCCACCACACATGA
52	InDA11M0002	11	ACTGCTGCACTGGGTATTCA	AGGTGCTATTGGTCCACACA
53	InDA11M0029	11	AGCACTTCCCATAGCACCAC	CATCGTTCCTCTCCCCATCG
54	InDA11M0051	11	GGGGGCTTCCATCTTGTGAA	CATGTGCAAGTAGGGCATGC
55	InDA11M0100	11	ATCGTTCCTCCTCTCCTCCC	GCGGGGTGAGTGTCTTGTTA
56	InDA11M0124	11	TAAGCAAGTGTGCGGGGTTT	TTACAGTGAAGTCGACTGCA
57	InDA12M0001	12	CACCAGCAAGTTTTGGGAGC	CTTGAGCTCTCCCCAAGTGG
58	InDA12M0022	12	TGCCAAAGCTCTTAAGTGCA	TCTTACATCGTGCACGAC
59	InDA12M0081	12	CCTCTCACTGGCGGAGTTTT	CTACCAGTGAGAGGAGGCCT
60	InDA12M0101	12	TTGGCATAGTGGATCGACCC	TTCTCGTCACCGTCTCTCT
61	InDA12M0123	12	TGGATGATGAAGTAGGGTGCG	GAATTGGCAACCTTACCCGA
62	InDA13M0042	13	TGCACGGTTGTCATACCTCC	CAGGTGTCGAAACAGGTGA
63	InDA13M0078	13	CGCGTATAGTGAGGCTTGCT	GAAAGCGCCAACTAAGGGC
64	InDA13M0090	13	ACCTTGGCGGGAAAAGCTTA	GAGGATGAACCCTAGTGAAGCT
65	InDA13M0119	13	AAGGGTCTAAGGCTTCGGC	GACCTGGTCCTCTTCAAGCA
66	InDA14M0002	14	GGGACGATTCAAGAGACGTGT	CCATGTTTGTAAGCGTGCCA
67	InDA14M0020	14	AGCCATATCTCAGCACGCTC	CCCCAAGGGCAGAGAAAGAG
68	InDA14M0097	14	CGGCTAGATCGTCAGGAACC	CTTGC GCGCCGTATTATTT
69	InDA14M0111	14	CGTTCAGGTGTGTCCGAAGA	AGAAACTCACCCAAACAAGCT
70	InDA14M0150	14	GGCACC GATACCTTTCAACC	CCAGCGTCCACCGTACAATA
71	InDA15M0002	15	GCGAGTTGTGTCCCTCAGAA	TCGGACCATGCGAGTTCTTT
72	InDA15M0044	15	AAGTGAGCACGGGCATTCTT	CCGGTTCCATCATCACTTCGA
73	InDA15M0101	15	CACCATCCTTGTCGGAACA	TGTTGAGGGTTTCTGGCCTG
74	InDA15M0125	15	AAACAGAACGCGTTTTGGGG	AGTGGCGAAGGATCCAAAGG
75	InDA15M0165	15	ATCCACATGGCCCATTCGTT	GCCAAGAGAGCATCCCTTGT
76	InDA16M0002	16	GTTAGGGGCCACTCTCAACC	TCTCAATCTCTCAGTCTCTGTTCA
77	InDA16M0016	16	TGCAAGCCCAACACAGTACA	GTGGCTCAAGGTACAGGACA
78	InDA16M0039	16	TCGCACTTTCACTCAAAAGGA	CTGGATAAGCTGGGTTGCCA
79	InDA16M0148	16	TGTTGGTGGGCTTTGGAGTT	AACCCACCCATTTCCGAA
80	InDA17M0002	17	TTCGTCGCGTATCTGAGCTC	GTCAACTCACTCGTCCGCTT
81	InDA17M0017	17	GCAGCCTTTTGTCTCCCTCT	GTGTGGCGCACTTGTCAAAT
82	InDA17M0043	17	CATGCAGTTCATGTAGTTGCTAACT	TTAGAGAGGTGCGCAACTGG
83	InDA17M0158	17	GTCCTCTTGTCAGTTGGGCA	ATGCCACGTGTGACTGACAT
84	InDA18M0002	18	TGCTGTTGTGTGATTCTTCGT	AGACAGGGGTTGATGTGTCT

85	InDA18M0018	18	GCGCGTGTAAATCAACGTACC	AAGAAGAAAGGGCGCGAGTT
86	InDA18M0045	18	GCGGTGGCTCTTCGATTAGA	GTGCACGAAACTAGTTCCGC
87	InDA18M0105	18	CCCAAATCCCATCCCATCCC	CCCTTTCTCATGCCCCCTCTC
88	InDA18M0138	18	CCAATCCTACCATGCCCTCC	GGTGGATGTCTCGACCAAGA
89	InDA19M0002	19	AGGTAGCCGTAAGTGAAGGT	CGAAATCAAATACAGTCCTACACAC
90	InDA19M0036	19	TGGCGCACCTACAAATTCA	ATGTGCACCCAAGGTTTCGAA
91	InDA19M0104	19	GCCCATGGCCAGTATAGGAC	CCAATCATGGTCTTTGGCCC
92	InDA19M0161	19	GAGGTGACTGGTCGAAGTGG	TCAAGGGTCACTGGGTCAGA
93	InDA19M0190	19	AGCTGGCATGCATCTGAACT	CAACCAATAGGGCACGACCT
94	InDA20M0002	20	AACACGCCTCAGATGACACC	TACGGTTGGACTGGTTGGAC
95	InDA20M0035	20	GGTTAGACCGTTGGACCAG	GTCCAAAGACTTGAAGGACCT
96	InDA20M0036	20	CCACCTTCTTCACCACCTCC	AGGTGGTGAAAAGAGCGGTT
97	InDA20M0058	20	TGTAGATGCATGAAACGACGA	TGGACAATCCAGGCTTTACAA
98	InDA20M0059	20	AGGAGTTGCTTGCAGTGGTT	TAAGCCCCTAAGGTCGAGCT
99	InDA20M0155	20	CCAGATGAGAGGTCCCAGGA	TGCAGAGTCTAAACCCGCAA
100	InDA20M0156	20	TGGAGAGCTGAAAATGCAAAGC	GGCCACAACCTGAGATAGGCA