

基于简化基因组的花生 InDel 标记开发和功能解析

王娟¹, 刘宇^{1,2}, 李春娟¹, 闫彩霞¹, 赵小波¹, 单世华¹

(¹山东省花生研究所, 青岛 266000; ²中国海洋大学食品科学与工程学院, 青岛 266101)

摘要: InDel 在基因组中的分布密度仅次于 SNP, 可作为动植物群体遗传分析、分子辅助育种等研究领域的有效分子标记。花生是世界范围内重要的油料作物之一。目前, 花生栽培种全基因组已经公布, 为准确挖掘花生基因组信息提供了重要参考。本研究通过 169 份花生核心种质的 GBS (Genotyping-by-sequencing) 测序和比对, 共获得大小分布在 1~14 bp 范围内的 10,401 个 InDels。染色体 Arah16 上分布 InDels 最多, 达 741 个; 而在染色体 Arah08 分布最少, 有 263 个。参考基因组注释信息, 仅有 1,167 个 InDels 分布在功能基因相关区域。经 GO 注释, InDel 分子功能主要包括催化活性 (catalytic activity) 和结合 (binding); 生物过程主要涉及代谢过程 (metabolic process)、单组织过程 (single-organism process) 和细胞过程 (cellular process)。经 KEGG 通路分析发现, InDels 所在的基因区域的功能主要与代谢相关。本研究开发出全基因组水平的 InDel 标记, 并做了相应功能分类和注释, 为进一步分子验证和利用提供丰富的基因资源。

关键词: InDels 检测; InDels 分布; 功能分析

Development and Functional Analysis of Peanut Insertion and Deletion (InDel) Markers Based on Genotyping-by-sequencing (GBS)

WANG Juan¹, LIU Yu^{1,2}, LI Chun-juan¹, YAN Cai-xia¹, ZHAO Xiao-bo¹, SHAN Shi-hua¹

(¹Institute of Shandong Peanut Research, Qingdao 266000; ²College of Food Science and Engineering, Ocean University of China, Qingdao 266101)

Abstract: Insertion/deletion (InDel) is the second largest molecular markers, which provide the valuable reference for population genetics and marker-assisted selection study. Peanut is one of the most important oil crops worldwide. The whole genome sequence of cultivated peanut has been released and therefore provided the important genome reference for functional genomic studies. In this study, by taking use of the genotyping-by-sequencing datasets on 169 peanut core collections, we identified 10,401 InDels with the size difference of 1 to 14 bp. The highest number of InDels was observed on chromosome Arah16, while the lowest number was

收稿日期: 2018-05-17

修回日期: 2018-07-13

网络出版日期:

URL (网络出版地址):

第一作者研究方向为花生遗传育种研究。E-mail: wangjuan_1984@163.com

通信作者: 单世华, 研究方向为花生种质资源评价与育种研究。E-mail: shansh1971@163.com

基金项目: 中央引导地方科技发展专项资金; 山东省博士基金 (ZR2017BC082); 泰山学者特聘专家 (ts201712080); 山东省农业科学院农业科技创新工程 (CXGC2018E21)

Foundation project: Central Guidance for Local Science and Technology; Natural Science Foundation of Shandong Province (ZR2017BC082); Taishan Scholars Project (ts201712080); Agricultural Science and Technological Innovation Project of Shandong Academy of Agricultural Science (CXGC2018E21).

found on Arahy.08. Although only 1,167 Indels were located in coding regions, obviously these Indels might show the specific functions, including catalytic activity and binding compared with other regions. The KEGG analysis suggested that Indels were mainly detected in the metabolism process. These Indels explored in this study might serve as the important genetic resources being useful for the further validation of functional genes.

Key words: InDel detection; InDel distribution; GO and KEGG annotation

栽培种花生 (*Arachis hypogaea* L.) 是我国重要的经济作物和油料作物。为了应对市场多样化的需求, 传统的育种方法在花生遗传改良上取得了很大进展^[1]。但是, 传统育种存在着时间周期长、工作量大、效果不突显等缺点。DNA 分子变异是区别作物不同类型和不同品种最可靠的指标。分子标记开发有利于花生品种目标基因的定位和克隆、遗传图谱的构建及分子标记辅助育种等。随着二代测序技术不断提高和测序成本逐渐降低, 为花生的分子遗传改良提供了更大的机遇^[2]。全基因组水平的分子标记可实现遗传改良的精准化。

插入或缺失 (Insertion/Deletion, InDel) 标记在基因组内分布和密度仅次于单核苷酸多态性标记 (Single Nucleotide Polymorphism, SNPs)。InDel 标记兼备了简单重复序列 (SSRs) 和单核苷酸多态性 (SNPs) 变异的特征, 有变异稳定、多态性强、检测容易等优点, 可作为动植物群体遗传分析和分子育种等研究有效分子标记, 受到了越来越多研究者的关注^[3]。InDel 变异是通过同一物种不同个体之间基因组的同一位点核苷酸片段大小来识别^[4]。例如, 郭广君^[5]通过辣椒全基因组重测序数据检测到 40 个 InDel 标记, 35 个 InDel 所在区域具备多态性, 表明重测序数据开发辣椒 InDel 标记的可行性。同样, 李斯更^[6]根据黄瓜基因组重测序的数据得到了

134 对 InDel 引物, 利用 16 份黄瓜典型种质检测出 116 对引物具有多态性, 表明 InDel 标记能够为黄瓜种质资源鉴定和分子遗传育种研究提供有力依据。同样, 刘栓桃^[7]采用 86 对 InDel 引物对 105 份大白菜育种材料进行了鉴定, 共得到 189 个多态性位点。此外, 还筛选到 3 对适宜 4 个杂交种纯度鉴定的通用引物, 为大白菜杂种优势育种提供分子证据, 也为杂交组合高效选配提供理论依据。但是相比于其它经济作物 InDel 标记的开发利用情况, 目前花生的 InDel 标记相关报道相对较少, 应用也较为有限。

简化基因组测序 (Reduced-Representation Genome Sequencing, RRGs) 是对部分基因组应用二代测序 (Next generation sequencing, NGS) 技术获得海量标签序列来代表目标物种全基因组水平信息的测序方法。简化基因组测序包括了 GBS (Genotyping-by-Sequencing), RAD (Restriction-site Associated DNA), 2b-RAD 和 SLAF-seq (Specific-Locus Amplified Fragment sequencing) 等多种测序技术。相比于全基因组重测序技术 (Whole Genome Resequencing, WGR), 简化基因组测序方法具有高效率、低成本和高准确性等技术优势, 可实现群体水平上全基因组层面的分子标记的有效开发。本研究采用的 GBS 方法, 是一种基于酶切的简化基因组测序方法^[8], 具备高通量、高效和简单特

性, 可实现群体全基因组水平遗传变异标记检测。目前, 栽培种花生全基因组序列 (<http://peanutgr.fafu.edu.cn/index.php>; https://www.peanutbase.org/peanut_genome) 已经释放和共享, 为基因资源挖掘提供了重要参考。在此, 我们选用 169 个花生核心种质材料, 进行 GBS 测序, 开发全基因组水平的 InDel 标记, 统计 InDel 的分布, 进行相关功能分析。这些丰富的 InDel 标记将为种质遗传资源研究和分子遗传改良提供基础。

1 材料与方法

1.1 试验材料

本研究选用的是基于花生资源目录信息中的近 3300 份农家种和育成品种的 13 个表型性状数据构

建的 169 份中国栽培花生核心种质材料^[9] (见表 1)。核心种质中包括 14 份多粒型 (8.28%), 51 份珍珠豆型 (30.18%), 24 份龙生型 (14.79%), 71 份普通型 (42.01%) 和 9 份中间型 (5.32%)。每份材料随机选取 3~5 粒完整种子, 培养皿中浸种 ~24 小时, 萌发后置入培养盒, 28°C 下培养 ~10 天。使用植物基因组 DNA 提取试剂盒 (TIANGEN, Beijing) 提取花生叶片 DNA。利用 Super GelRed (S-2001) 进行凝胶电泳检测 (US Everbright Inc, Suzhou, China) 后, 进一步使用 Nanodrop™ 2000 (Thermo Scientific) 分光光度计检测 DNA 的完整性和纯度。每份 DNA 样品保证单位浓度不少于 50 ng/μL, 总质量大于 2μg, 置于 -80°C 保存备用。

表 1 花生材料信息

Table 1 Information of accessions used in this study

原始编号	种质编号	备注	原始编号	种质编号	备注
Original Code	Germ Number	Note	Original Code	Germ Number	Note
zh.h0858	SD11_27	普通型	zh.h4795	GD23_112	珍珠豆型
zh.h2405	SD11_45	普通型	zh.h2604	GD13_113	普通型
zh.h2764	SD51_50	中间型	zh.h0577	JS42_78	龙生型
zh.h2250	SC42_82	龙生型	zh.h2330	GX44_144	龙生型
zh.h0949	SD11_34	普通型	zh.h5020	CQ12_62	普通型
zh.h0764	SD11_23	普通型	zh.h4716	ZJ22_94	珍珠豆型
zh.h2464	HeN11_10	普通型	zh.h2294	HuN42_77	龙生型
zh.h2591	HuN12_74	普通型	zh.h2618	GD13_114	普通型
zh.h0861	SD11_28	普通型	zh.h0481	GX23_129	珍珠豆型
zh.h2406	SD11_46	普通型	zh.h1602	LN36_163	多粒型
zh.h5034	HuB12_66	普通型	zh.h4754	JX23_140	珍珠豆型
zh.h5327	SD11_52	普通型	zh.h4701	Anh22_60	珍珠豆型
zh.h0537	SD41_21	龙生型	zh.h2608	HuN44_150	普通型
zh.h0977	SD11_35	普通型	zh.h0162	JX22_79	珍珠豆型
zh.h4749	JX23_139	珍珠豆型	zh.h5075	ZJ12_95	普通型
zh.h2462	HeN11_9	普通型	zh.h2076	GX23_130	珍珠豆型
zh.h1401	SD51_41	中间型	zh.h2258	SC42_89	龙生型
zh.h1395	SD51_39	中间型	zh.h2665	GD13_115	普通型
zh.h2702	LN16_158	普通型	zh.h0125	Anh21_1	珍珠豆型

zh.h1843	HuN22_75	珍珠豆型	zh.h2683	GZ24_146	普通型
zh.h1374	XJ17_168	普通型	zh.h4756	JX23_141	珍珠豆型
zh.h0930	SD11_32	普通型	zh.h1075	HeN12_65	普通型
zh.h2461	HeN11_8	普通型	zh.h2483	SC12_90	普通型
zh.h1388	LN16_159	普通型	zh.h0477	GX23_131	珍珠豆型
zh.h1416	HuB52_67	中间型	zh.h2374	GD43_116	龙生型
zh.h5030	SC12_83	普通型	zh.h0436	GX23_132	珍珠豆型
zh.h2426	SD11_49	普通型	zh.h0024	SD31_17	多粒型
zh.h2477	HeN11_13	普通型	zh.h2682	GZ24_148	普通型
zh.h0680	BJ11_2	普通型	zh.h0608	JX42_80	龙生型
zh.h0531	SD41_20	龙生型	zh.h2176	LN26_164	珍珠豆型
zh.h5060	Anh12_55	普通型	zh.h2105	GX23_133	珍珠豆型
zh.h2561	Anh12_56	普通型	zh.h0042	LN36_165	多粒型
zh.h2413	SD11_53	普通型	zh.h2292	ZJ42_96	龙生型
zh.h0719	HeB11_4	普通型	zh.h2069	GX23_134	珍珠豆型
zh.h2456	HeN11_7	普通型	zh.h2508	SC12_91	普通型
zh.h2499	SC12_84	普通型	zh.h4431	SD31_51	多粒型
zh.h2404	SD11_44	普通型	zh.h0378	FJ23_104	珍珠豆型
zh.h1399	SD51_40	中间型	zh.h1934	GD23_117	珍珠豆型
zh.h4923	HuN24_151	龙生型	zh.h0519	LN26_166	珍珠豆型
zh.h2466	HeN11_11	普通型	zh.h2208	SD41_43	龙生型
zh.h4820	LN26_160	珍珠豆型	zh.h5065	Anh12_61	普通型
zh.h0525	LN26_161	珍珠豆型	zh.h2380	GD43_118	龙生型
zh.h2410	SD11_47	普通型	zh.h2378	GD43_119	龙生型
zh.h1331	GX13_126	普通型	zh.h1590	ZJ32_97	多粒型
zh.h1714	SC22_85	珍珠豆型	zh.h4911	SC42_92	龙生型
zh.h5399	SX55_154	中间型	zh.h2654	GD13_120	普通型
zh.h0934	SD11_33	普通型	zh.h1683	HeN21_6	珍珠豆型
zh.h2044	FJ23_99	珍珠豆型	zh.h0074	SD21_19	珍珠豆型
zh.h0540	SD41_22	龙生型	zh.h1777	HuB22_71	珍珠豆型
zh.h1816	Anh22_57	珍珠豆型	zh.h0678	XJ47_169	龙生型
zh.h2550	HuB12_68	普通型	zh.h1306	JX12_81	普通型
zh.h2231	SC42_86	龙生型	zh.h0008	SD31_16	多粒型
zh.h0868	SD11_29	普通型	zh.h1689	SC22_93	珍珠豆型
zh.h4779	GD23_105	珍珠豆型	zh.h0003	HeB31_3	多粒型
zh.h2671	GD13_106	普通型	zh.h2248	CQ42_63	龙生型
zh.h2006	GD23_107	珍珠豆型	zh.h2621	GD13_121	普通型
zh.h1771	HuB22_69	珍珠豆型	zh.h1597	YN54_153	多粒型
zh.h0636	GX44_142	龙生型	zh.h0041	LN36_167	多粒型
zh.h1961	GD23_108	珍珠豆型	zh.h0417	GX23_135	珍珠豆型
zh.h2609	HaiN13_138	普通型	zh.h0441	GX23_136	珍珠豆型
zh.h2475	HeN11_12	普通型	zh.h2603	GD13_122	普通型
zh.h2177	LN26_162	珍珠豆型	zh.h2387	GD43_123	龙生型
zh.h1324	GX14_143	普通型	zh.h1045	HeN11_5	普通型

zh.h2100	GX23_127	珍珠豆型	zh.h2160	GZ14_149	珍珠豆型
zh.h1315	FJ13_100	普通型	zh.h1672	SX55_156	珍珠豆型
zh.h5365	SX25_155	中间型	zh.h4673	HuB22_72	珍珠豆型
zh.h0099	JS21_14	珍珠豆型	zh.h4681	HuB22_73	珍珠豆型
zh.h1585	SX31_54	多粒型	zh.h0102	GX32_64	多粒型
zh.h2360	GX43_128	龙生型	zh.h1591	ZJ32_98	多粒型
zh.h2562	Anh12_58	普通型	zh.h1596	GX33_137	多粒型
zh.h2045	FJ23_101	珍珠豆型	zh.h0062	SD21_18	珍珠豆型
zh.h2652	GD13_109	普通型	zh.h2651	GD13_124	普通型
zh.h2224	SC42_87	龙生型	zh.h0875	SD11_30	普通型
zh.h1830	HuN22_76	珍珠豆型	zh.h0998	SD11_37	普通型
zh.h0500	GZ24_145	珍珠豆型	zh.h2664	GD13_125	普通型
zh.h2376	GD43_110	龙生型	zh.h5376	SX55_157	中间型
zh.h2152	GZ14_147	珍珠豆型	zh.h0856	SD11_26	普通型
zh.h0033	FJ33_102	多粒型	zh.h0844	SD11_24	普通型
zh.h1797	HuB22_70	珍珠豆型	zh.h0980	SD11_36	普通型
zh.h2425	SD11_48	普通型	zh.h1001	SD11_38	普通型
zh.h0127	Anh22_59	珍珠豆型	zh.h0853	SD11_25	普通型
zh.h5031	SC12_88	普通型	zh.h0897	SD11_31	普通型
zh.h2056	FJ23_103	珍珠豆型	zh.h1411	JS51_15	中间型
zh.h2637	GD13_111	普通型	zh.h1649	SD21_42	珍珠豆型
zh.h0490	YN34_152	珍珠豆型			

1.2 高通量测序

首先，双酶切是 GBS 建库的第一步。选取合适的酶切组合 (*EcoR I* 和 *Nia III*) 对质量合格的 DNA 进行酶切^[10]。其次，GBS 建库，酶切好的每个样本加不同的一级接头 (adapter)，再通过 PCR 加入二级接头，以获取大小~350 bp 的扩增片段。最后，确认建库质量后，在 Illumina HiSeq4000 平台上进行 PE150 双末端测序。

1.3 原始测序数据过滤

利用 Q30 标准对序列质量进行评估,采用 FASTX-Toolkit 软件对原始数据进行过滤 (http://hannonlab.cshl.edu/fastx_toolkit/index.html), 具体内容如下: (1) 去掉带接头的 paired reads; (2) 去掉 N 的含量超过该条 read 长度比例 10% 的

paired reads; (3) 去除当单端序列低质量碱基 ($Q \leq 5$) 比例超过 read 长度 50% 的双端 reads。

1.4 InDel calling 和统计分析

首先，以栽培种花生基因组序列为参考，使用 BWA^[11]对过滤后的数据进行比对。根据 Clean Reads 在参考基因组的定位，通过 SAMTOOLS 软件^[12]去重复，接着使用 GATK 软件进行局部比对和碱基质量值校正等预处理，以保证检测结果的准确性，最后使用 GATK 软件进行变异检测，获得可靠的 InDel 位点^[13]。

1.5 InDel 功能分类，GO 注释和 KEGG 通路分析

先利用 SnpEff 软件获取 InDel 位点的基因组位置信息，进一步获取 InDel 所在 DNA 片段的 GO 注释和 KEGG (Kyoto Encyclopedia of Genes and

Genomes) 信息, 最后利用 R 软件包 (<https://www.r-project.org/>) 进行统计和图示。

2 结果与分析

2.1 GBS 测序和数据过滤

169 份花生核心种质经过 GBS 测序和质控, 共产生了 1,695 million 高质量 reads, 包含了 254.3 G 测序数据。与原始 reads 相比, 保留下来的 clean reads 占 95.4%。其中, phred score > 20 的 reads 占 97.6%, phred score > 30 的 reads 占 93.1%。GC 含量较稳定, 在 37.6%~ 40.2%之间浮动。

2.2 InDel 检测和分布

共有 10,401 个 InDels 保留下来。虽然 InDel 的长度从 1~14bp 均有分布, 但是主要还是以小片段 (1-5bp) 分布为主。其中, 单碱基 InDels 最多, 有 8196 个。接着是双碱基 InDels, 有 1038 个 (图 1)。随着 InDel 碱基数增加, InDel 数目减少。在各个染色体上, InDel 数目的分布在 263~741 个之间。总体上, 分布于花生各个染色体上 InDel 标记比较均匀, 平均每条染色上每 Mb 距离内分布 3.23~5.07 个 InDels (图 2, 3 和表 2)。

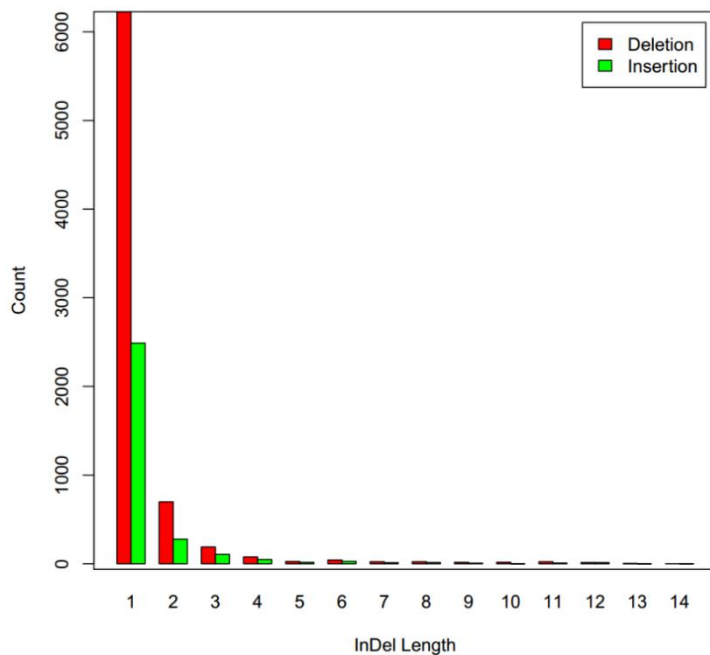


图 1 不同长度的 InDel 在花生基因组的分布

Fig 1 The distribution of different size InDel in peanut genome

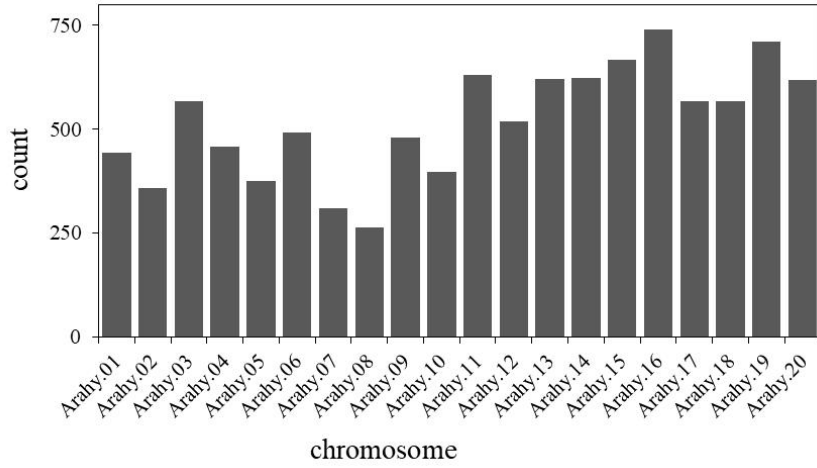


图 2 每个染色体上分布的 InDels 数目

Fig 2 The number of InDels on each chromosome

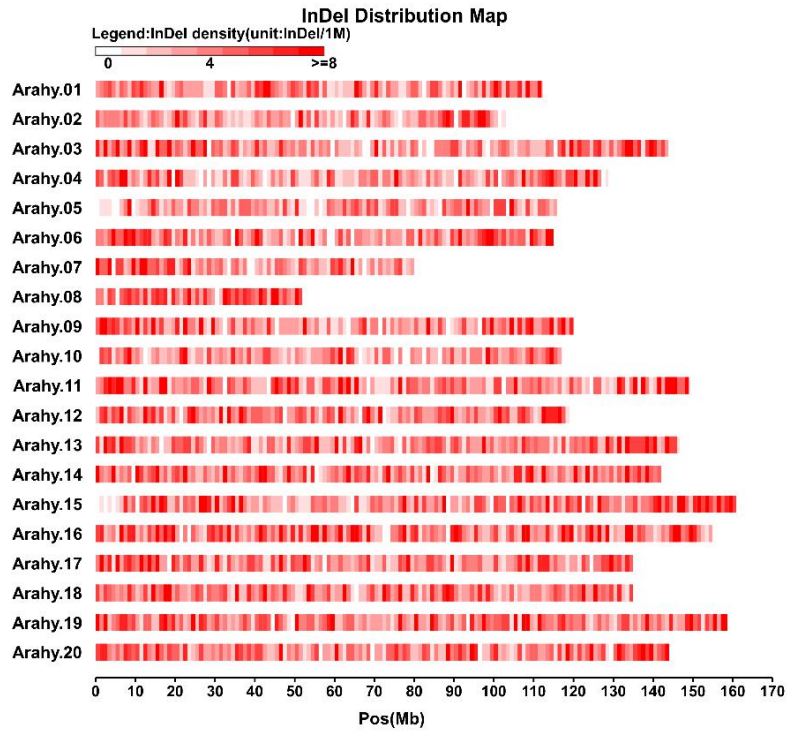


图 3 InDels 在各个染色体上的分布情况。横轴为染色体长度（以 Mb 为单位）；垂直轴依次为染色体编号

Fig 3 InDel distribution on the 20 chromosomes of peanut. The horizontal axis shows chromosome length (Mb). The vertical axis are the serial number of 20 chromosomes

表 2 每个染色体上 InDel 分布频率情况

Table 2 Distribution and frequency of InDel in each chromosome

染色体	物理长度	数目	频率
Chromosome	Physical length (bp)	Total numbers	Frequency (InDels/1Mb)
Arahy.01	112420854	442	3.93

Arahy.02	102981163	358	3.48
Arahy.03	143813506	566	3.94
Arahy.04	128801742	458	3.56
Arahy.05	115930344	375	3.23
Arahy.06	115504342	491	4.25
Arahy.07	81119488	310	3.82
Arahy.08	51897010	263	5.07
Arahy.09	120519698	479	3.97
Arahy.10	117088237	398	3.40
Arahy.11	149299306	630	4.22
Arahy.12	120579088	518	4.30
Arahy.13	146725006	620	4.23
Arahy.14	143237272	624	4.36
Arahy.15	160879708	668	4.15
Arahy.16	154808347	741	4.79
Arahy.17	134922436	567	4.20
Arahy.18	135150084	566	4.19
Arahy.19	158625764	710	4.48
Arahy.20	143980330	617	4.29

2.3 GO 注释及 KEGG 通路分析

根据栽培种花生的全基因组序列注释信息，共得到了 1,167 条已知蛋白及其相关信息。其中，分布在外显子区域的 InDel 标记有 91 个。5'和 3'端非

翻译区域（UTR）内有 244 个 InDels。距离基因上游或/和下游 1kb 的距离内有 828 个 InDels。此外，还有 4 个 InDels 存在于剪切区域（图 4 和表 3）。

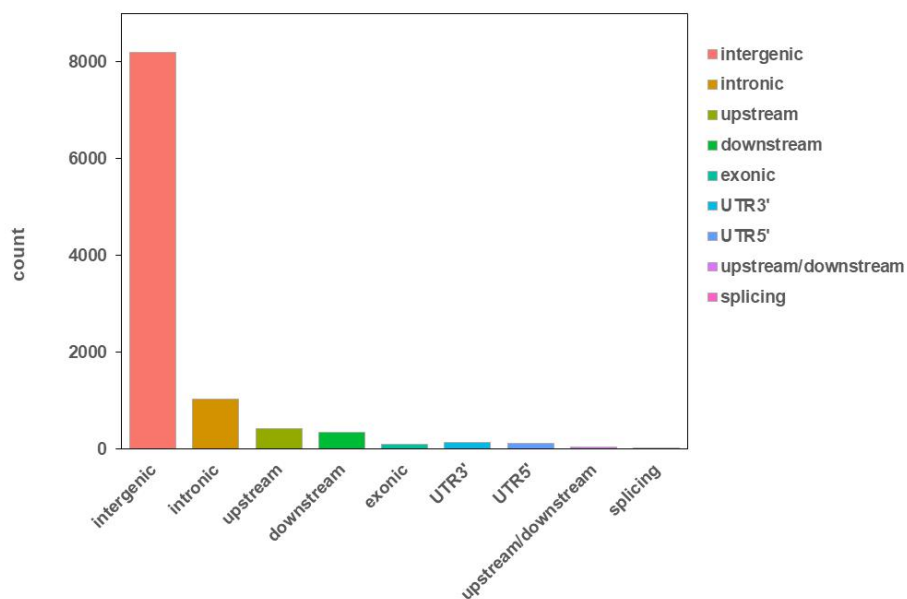


图 4 InDels 区域的功能分类

Fig 4 The distribution of InDels surrounding the gene regions

表 3 InDel 所在区域的功能分类情况

Table 3 The functional classification of InDels

DNA 区域		总数
DNA region		Total
Intergenic		8196
Intronic		1038
Exonic	Frameshift deletion	38
	Frameshift insertion	32
	Non-frameshift deletion	13
	Non-frameshift insertion	4
	Stop-gain	3
	Stop-loss	1
UTR3'		130
UTR5'		114
Upstream		427
Downstream		350
Upstream/downstream		51
Splicing		4
Total		10401

对 InDel 标记所在的 DNA 片段做 GO 功能注释，生物过程 (Biological Process) 中占比例较多的是代谢过程 (Metabolic Process)，含 200 个 InDel 区域。细胞过程 (Cellular Process) 次之，有 177 个 InDel 区域。在分子功能 (Molecular Function) 这一类中，

催化活性 (catalytic activity) 最高，包括 215 个含 InDel 的基因区域。蛋白结合 (binding) 次之，包括 149 个 InDel 区域。在细胞组分 (Cellular Component) 这一类中，细胞 (cell) 及细胞组成 (cell part) 均含有 131 个 InDel 的基因区域 (图 5)。

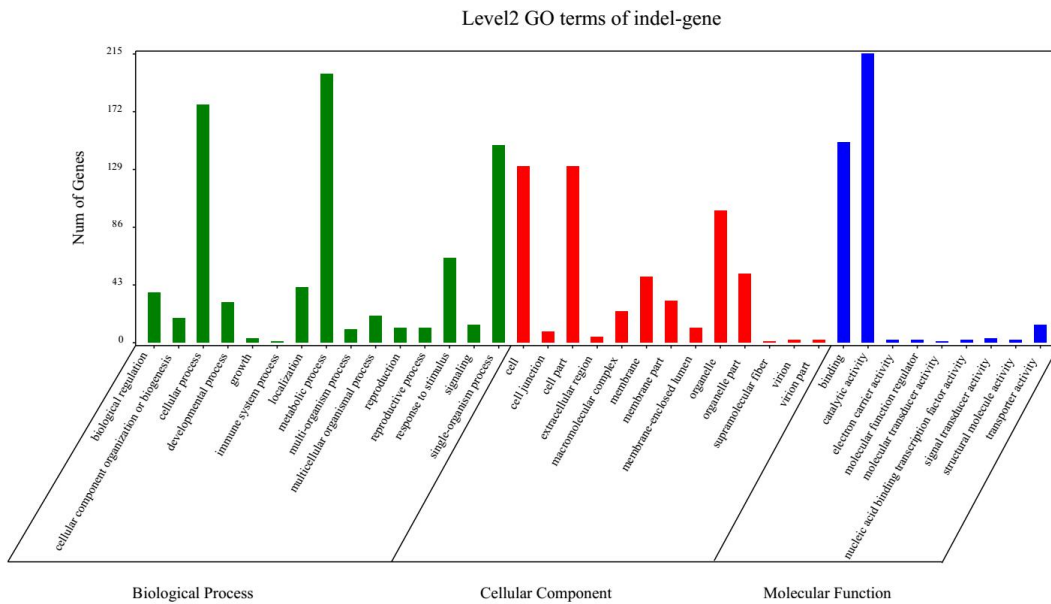


图 5 InDels 的 GO 注释结果

Fig 5 GO annotation of InDels

根据 KEGG 注释, 我们得到了 108 个基因通路信息。代谢相关通路占了 75%。其中, 均有 13 个含 InDel 的基因片段参与了氨基酸代谢和碳水化合物

代谢。脂类代谢次之, 包含了 12 个 InDel 区域 (图 6)。

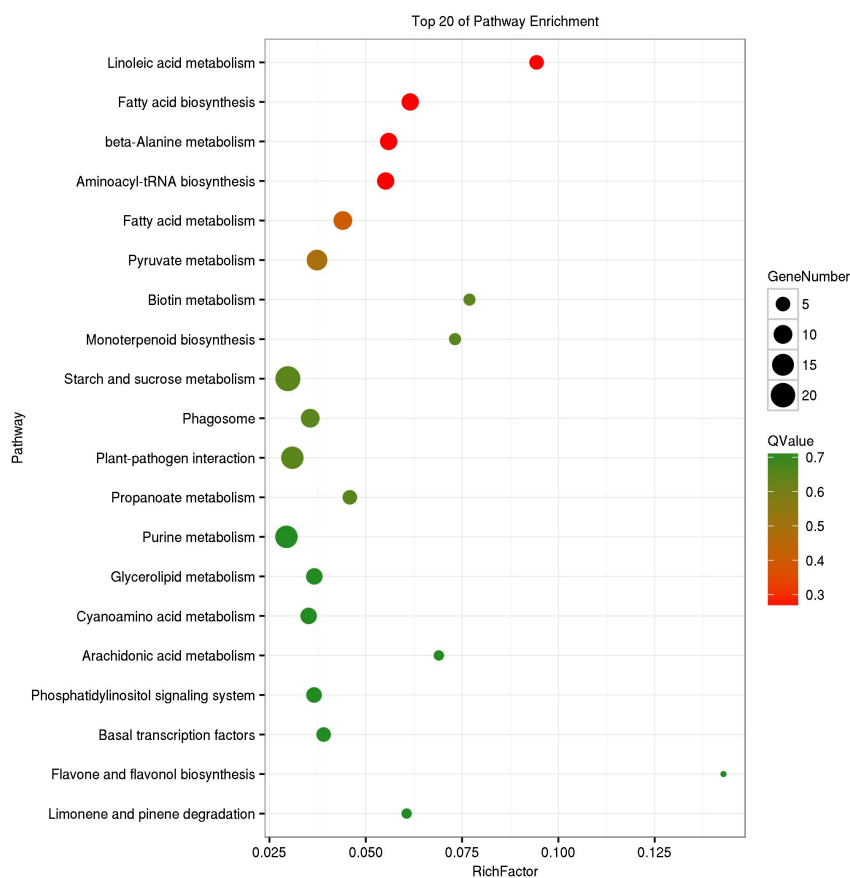


图 6 InDel 标记的 KEGG 分析
Fig 6 KEGG pathway analysis of InDels

3 讨论

近十年来, 随着二代测序技术不断提高, 测序成本不断降低, 越来越多的植物全基因组数据为大家所共享^[4], 这为全基因组水平分子标记的开发和利用提供了重要参考。DNA 片段的遗传多样性一般表现在单核苷酸多态性 (SNPs), 多核苷酸多态性 (MNP), 插入和缺失 (InDels) 差异以及大段的结构差异 (SV)。其中, InDel 标记为共显性标记, 具有较好的稳定性和较丰富的多态性。S. Meng^[15] 利用 48 个 InDel 标记检测中国北方 54 份花生品种的遗传多样性, 结果一共扩增到 61 个多态性位点。

并且, 聚类结果能够较好地分辨花生四个主要生物学类型。这说明 InDel 标记对花生种质资源鉴定的有效性。LM. Liu^[6] 从包含花生六个生物学类型的 118 个种质的 48 个较为保守的功能基因区域开发出 16 个多态 InDel 标记位点。相对于其它类型, var. *aequatoriana*, var. *peruviana* 和 var. *hirsuta* 含有较高的遗传多样性。并且, 5 个 InDels 所在区域与番茄斑萎病毒、叶斑病显著相关, 这些新发现的 InDel 标记可用于花生资源中抗病相关基因型的鉴定。与花生 InDel 开发与多态性检测相关的研究涉及的 InDel 数目较为有限, 限制了花生中 InDel 分子标记

资源的充分挖掘。本研究中,我们通过 GBS 测序技术对 169 份花生核心种质进行高通量测序,成功获得全基因组水平上 InDel 标记。由于编码序列相对保守,仅发现 11.2% InDels 位于基因相关区域。类似的, S. Mahmood^[17]对来自不同生态地区且拥有不同生长习性的 23 个甘蓝型油菜自交系进行基因组重测序,共获得了 538,691 个序列长度为 1-10bp 的 InDel 变异位点,研究指出基因区域的 InDel 标记也仅仅占了 25.17%。后续研究将进一步确认 InDel 多态性位点,揭示花生种质资源的遗传多样性,这对发掘优良种质资源,选配优良亲本和拓宽育成品种的遗传基础有着十分重要的意义。此外,结合表型数据,具备多态性的 InDel 标记可应用于花生产量、品质以及抗性性状等功能基因的筛选,有利于优异等位基因的发现。

值得注意的是,GO 注释到 93.6%的 InDels 均与催化活性(215 个)和蛋白结合(149 个)相关,而 InDel 标记涉及的其它分子功能不超过 25 个。通过分析表明含 InDel 标记倾向于出现在氨基酸、碳水化合物和脂类代谢等生理过程紧密相关基因区域中。类似的,吴磊^[18]通过拟斯卑尔脱山羊草 EST (Expressed Sequence Tag) 与小麦 Uniene 序列的比对分析,发现山羊草插入/缺失(InDel)位点 137 个,在这些位点两端序列设计 24 对引物中,发现 11 对引物具多态性。这些包含突变位点的基因也主要涉及到亚细胞定位、蛋白质结合与催化以及代谢等过程。这些 InDels 主要相关的基因功能是植物体内重要性状形成的基础。加之, InDels 突变容易造成蛋白质读码框的移位,导致编码蛋白功能缺失和改变,从而直接影响基因功能的发挥。上述 InDels 基

因功能的倾向性也进一步多态性检测和分子生物学验证,以确认 InDel 标记在花生功能基因中的重要作用,为花生资源研究和遗传改良提供重要参考。

参考文献

- [1] 禹山林. 中国花生品种及其系谱.上海:科学技术出版社,2008
- [2] Rajeev K. Varshney, Spurthi N. Nayak, Gregory D. May,Scott A. Jackson.Next- generation sequencing technologies and their implications for crop genetics and breeding. Trends in biotechnology, 2009, 27 (9) :522 -530
- [3] 杨洁. InDel 标记的研究和应用进展. 生物多样性,2016,24(2):237-243
- [4] Alkan Can, Coe Bradley P, Eichler Evan E. Genome structural variation discovery and genotyping. Nature reviews. Genetics, 2011, 12 (5): 363 - 376
- [5] 郭广君, 孙茜, 刘金兵, 潘宝贵, 刁卫平, 戈伟, 高长洲, 王述彬. 基于辣椒基因组重测序的 InDel 标记开发及应用. 江苏农业学报,2015,31(06):1400-1406
- [6] 李斯更, 沈楠, 刘博, 邱杨, 张晓辉, 张忠华, 王海平, 李锡香. 基于黄瓜基因组重测序的 InDel 标记开发及其应用. 植物遗传资源学报, 2013, 14(2):278-283
- [7] 刘栓桃, 张志刚, 司立英, 王荣花, 李巧云, 王立华, 赵智中, 梁水美, 张全芳, 步迅. 基于 InDels 标记的大白菜育种材料的亲缘关系鉴定. 植物遗传资源学报,2018, 19 (4) 657-667
- [8] Elshire R J, Glaubitz J C, Sun Q, Poland JA, Kawamoto K, Buckler E S, Robert J. Elshire. A Robust, Simple Genotyping-by- Sequencing (GBS) Approach for High Diversity Species. Plos One, 2013, 6(5): e19379
- [9] 张浩. 中国栽培花生 (*Arachis hypogaea* L.) 遗传多样性分析与核心种质构建.山东: 山东农业大学, 2013

- [10] 王娟, 李春娟, 闫彩霞, 赵小波, 单世华. 栽培种花生 ddGBS 测序中酶切组合的选择. 花生学报, 2017, 46(4):48-51
- [11] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 2009, 25(14), 1754–1760
- [12] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Marth H N, Abecasis G, Durbin R, and 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMTOOLS. *Bioinformatics*, 2009; 25(16): 2078–2079
- [13] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 2010, 20(9):1297-303
- [14] Bertoli D J, Cannon S B, Froenicke L, Huang G, Farmer A D, Cannon E K, 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 AC, Kozik A, Kim K D, Burow M D, Varshney R K, Wang X, Zhang X, Barkley N, Guimarães P M, Isobe S, Guo B, Liao B, Stalker H T, Schmitz R J, Scheffler B E, Leal-Bertioli S C, Xun X, Jackson S A, Michelmore 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
- [15] Meng S, Yang X L, Dang P M, Cui S L, Mu G J, Chen C Y, Liu L F. Evaluation of insertion-deletion markers suitable for genetic diversity studies and marker-trait correlation analyses in cultivated peanut (*Arachis hypogaea* L.). *Genetics & Molecular Research Gmr*, 2016, 15(3)
- [16] Liu L, Dang P M, Chen C Y. Development and Utilization of InDel Markers to Identify Peanut (*Arachis hypogaea*) Disease Resistance. *Frontiers in Plant Science*, 2015, 6
- [17] Mahmood S. Genome-wide characterization of insertion/deletion variations and marker development in oilseed rape (*Brassica napus* L.) by valorizing resequencing genome data. 武汉: 华中农业大学, 2016
- [18] 吴磊, 王丹, 苏文悦, 郭长虹, 束永俊. 利用比较基因组学开发山羊草属 InDel 分子标记. 作物学报, 2012, 38(7):1334-1338