

李敏敏,何小春,张继雨,等. 基于全基因组重测序开发鲜食大豆通酥 1 号 InDel 分子鉴别标记[J]. 江苏农业科学,2025,53(22):73-80.
doi:10.15889/j.issn.1002-1302.2025.22.009

基于全基因组重测序开发鲜食大豆通酥 1 号 InDel 分子鉴别标记

李敏敏¹, 何小春¹, 张继雨², 朱治佳³, 邱牧¹, 赵杨¹, 李磊⁴, 张正睿¹

(1. 聊城市农业科学院, 山东聊城 252000; 2. 菏泽市农业科学院, 山东菏泽 274000;

3. 黑龙江省农业科学院齐齐哈尔分院, 黑龙江齐齐哈尔 161000; 4. 聊城市财金能源发展有限公司, 山东聊城 252000)

摘要:为从全基因组序列水平解析鲜食大豆品种通酥 1 号的遗传构成,采用 Illumina 公司的 NovaSeq 测序平台进行全基因组测序,平均覆盖深度为 27.08 ×,系统分析通酥 1 号单核苷酸多态性(SNP)、插入/缺失(InDel)、拷贝数变异(CNV)和染色体结构变异(SV)等序列变异。结果显示,在鲜食大豆品种通酥 1 号中共检测到 1 744 120 个 SNP,其中 73.66% 的 SNP 位于基因间区,而在位于外显子区域的 4.08% 变异中,53.83% 为非同义突变,可能导致 2 883 个基因蛋白质功能改变。鉴定出 232 213 个 InDel 变异,其中短片段(1~4 bp)变异占比明显高于长片段(≥5 bp),可能是由于短片段变异对基因组稳定性的破坏较小,被自然选择容忍。在外显子区域的 2 986 个 InDel 变异中,约 1.28% (39 个)引起 *Glyma.02g123100* 等 22 个基因提前终止密码子,可能显著影响基因功能;约 0.13% (5 个)导致 *Glyma.15g187700* 等 5 个基因终止密码子丢失,可能通过延长蛋白翻译产生功能获得性突变。检测到 8 732 个 CNV 和 5 749 个 SV,分别导致 5 982、24 594 个基因发生变异。最后基于 14 个 InDel 变异开发鉴别标记,其中标记 InDel_1、InDel_8、InDel_12 分别导致 *Glyma.02g290500*、*Glyma.10g103100*、*Glyma.16g194600* 基因发生序列变异,可能通过调节风味、影响细胞壁结构与质地等途径影响大豆鲜食特性。这些标记的进一步转化将提升育种效率,加速优质鲜食大豆品种的选育进程。

关键词:全基因组;重测序;鲜食大豆;InDel 分子标记

中图分类号:S188;S565.101 **文献标志码:**A **文章编号:**1002-1302(2025)22-0073-08

目前,规模种植的栽培大豆是在野生大豆的基础上经过长期选育和系统改良而来的,因采收时间不同可分为鲜食型大豆和籽粒型大豆,其中鲜食型大豆在生育期 R6 和 R7 之间收获,作为新鲜蔬菜食用,质地柔软、风味俱佳、营养丰富,维生素、蔗糖、淀粉含量较高,而难消化的低聚糖和抗营养物质的含量则较低。

近年来,基因组测序和重测序的技术成本大幅下降,为全面分析不同植物品种的遗传变异[如单核苷酸多态性(SNP)、结构变异(SV)等]提供了便利。通过对 278 个来自不同育种历史阶段的温带玉米自交系进行测序研究,科学家们评估了选择对玉米遗传多样性的影响,结果显示,现代育种技术已经通过 SNP、InDel 和拷贝数变异(CNVs)的形式引

入了丰富多样的遗传变异,这些变异影响了玉米基因组中的许多基因和非基因区域^[1]。小麦 *Ppd-1* 基因上游 50 bp 区域的 InDel 多态性与小麦品种的抽穗时间相关^[2]。小麦 *Ppd-B1* 的 CNV 影响其光周期敏感性,与抗病性和生物胁迫反应相关的 CNV 也已在拟南芥、水稻和大豆中被鉴定出来,表明这些基因的可变拷贝数可能在应对不断变化的环境条件和不断进化的害虫及病原体时具有优势^[3-6]。在多种植物物种中,结构变异已被鉴定出来,包括拟南芥、大麦、谷子、玉米、高粱、大豆和小麦等^[2,6-14]。在某些情况下,SV 与表型变异相关,例如,在高粱中鉴定出 1 组近 1 500 个区分甜高粱和粒用高粱的基因均携带 SV^[14]。上述发现强调了基因组变异在植物育种和适应性中的重要性。

随着分子生物学的发展,分子标记技术应运而生。分子标记技术通过检测植物基因组中的特定 DNA 序列来识别和选择优良品种,这种方法不受环境因素的影响,能够更准确地反映植物的遗传特性。因此,分子标记技术在农作物的遗传改良中具

收稿日期:2025-05-07

基金项目:聊城市重点研发计划政策引导类项目(编号:2024YD91)。

作者简介:李敏敏(1991—),女,山东聊城人,博士,高级农艺师,主要从事大豆遗传育种与栽培研究。E-mail:liminmin527@163.com。

有重要的应用价值。结合运用形态标记技术和分子标记技术,可实现农作物性状的精准鉴定与优良品种的高效选育,进而显著提升遗传改良效率与种质资源利用水平。具体到鲜食大豆,通过开发分子标记,分析其遗传多样性,不仅有助于挖掘和利用优异的种质资源,还能拓宽遗传背景,筛选出优良的品种(系),为鲜食大豆的育种和生产提供重要指导。

本研究通过高通量重测序技术对通酥 1 号基因组的变异进行系统性解析,揭示其基因组结构特征与遗传多样性,旨在为解析其表型特性及分子育种提供重要的理论依据;同时,开发一套多态性 InDel 标记,并利用其对 29 份代表性种质进行遗传多样性分析,以评估其在分子标记辅助育种中的应用潜力。

1 材料与方法

1.1 材料及 DNA 提取

鲜食大豆品种通酥 1 号购自市场,为春播中熟菜用大豆品种,由江苏沿江地区农业科学研究所于 2002 年育成,在淮北地区生产中表现出产量高、品质优、适应性强等特点^[15]。2024 年 6 月将种子播种于聊城市农业科学院茌平基地(115.98°E,36.45°N),7 d 后取幼嫩叶片,使用 DNA 提取试剂盒(Hi-DNAsecure Plant Kit)提取其 DNA,并使用 0.8% 琼脂糖凝胶电泳及 Nanodrop 2000 检测 DNA 质量,于 -20 °C 保存备用。

1.2 文库构建和质量控制

构建插入片段为 400 bp 的文库,利用第二代测序技术(NGS),基于 Illumina NovaSeq 测序平台,对这些文库进行双末端(PE)测序。采用 fastp(v0.20.0)对原始数据进行过滤:(1)去除 3' 端接头污染;(2)采用滑动窗口法进行质量过滤,窗口大小设置为 5 bp,步长设置为 1 bp,若窗口平均 Q 值 ≤ 20 ,仅保留该窗口倒数第 2 个碱基及之前的碱基;(3)长度过滤,去处任意一条 reads 长度 ≤ 50 bp 的双末端 reads^[16]。

1.3 序列比对及变异分析

使用 BWA(0.7.12-r1039)的 mem 程序将过滤后的高质量数据比对至参考基因组 Williams 82(Glycine max v4.0),获得 sam 文件,然后使用 Picard 1.107 软件对 sam 文件进行排序并转换为 bam 文件^[17]。

SNP 检测:采用 GATK 软件检测 SNP,使用

RealignerTargetCreator 命令输出包含所有可能的 InDels 文件;利用 IndelRealigner 命令对所有 InDels 附近的 reads 进行重新比对,以提高 SNP 预测的准确率。使用 UnifiedGenotyper 程序获得 SNP 位点,将 stand_call_conf 设置为 30^[18]。进一步过滤 SNP 位点,过滤标准:(1) Fisher test of strand bias (FS) ≤ 60 ;(2) HaplotypeScore ≤ 13.0 ;(3) Mapping Quality (MQ) ≥ 40 ;(4) Quality Depth (QD) ≥ 2 ;(5) ReadPosRankSum ≥ -8.0 ;(6) MQRankSum > -12.5 ;(7) DP ≥ 4 reads。

InDel 检测:使用 GenomeAnalysisTK v3.8 软件的 UnifiedGenotyper 程序获得样品所有突变位点,将 stand_call_conf 设置为 30,利用 SelectVariants 命令提取 InDel 突变信息。进一步过滤 InDel 位点,过滤标准:(1) Fisher Test of Strand Bias (FS) ≤ 200 ;(2) Read depth (DP) > 4 ;(3) Quality Depth (QD) ≥ 2 ;(4) ReadPosRankSum ≥ -20 。

CNV 检测:采用 CNVnator v0.2.7 检测全基因组中存在的 CNV,其中 bin_size 设置为 100,同时对通过 CNVnator 获得的结果进行筛选,将基因组中多拷贝区域去除,同时设置 Read Depth 值(将染色体正常区域的 Read Depth 定义为 1),保留 RD 值 ≤ 0.05 或 ≥ 1.8 的区域。

SV 检测:采用 Breakdancer1.3.7 检测可能存在的染色体结构变异。

1.4 变异位点注释分析

使用 ANNOVAR 软件对 SNP 及 InDel 位点进行 GO 和 KEGG 注释^[19]。

1.5 基因组变异图谱绘制

使用 R circlize 软件将 SNP、InDel、CNV 和 SV 的变异分析结果展示到环形图上^[20]。

1.6 e-PCR 扩增及进化分析

通过 TBtools 软件的 Batch Target Region Primer Design 插件对 InDel 分子标记进行引物设计,然后通过 Primer Check 插件基于全基因组序列(SoyOmics, <https://ngdc.cnbc.ac.cn/soyomics/index>; LIS, <https://www.legumeinfo.org/collections/glycine/>)进行 e-PCR 扩增,最后通过 R 语言的 ggtree 软件包进行种质资源进化分析,并构建进化树。

2 结果与分析

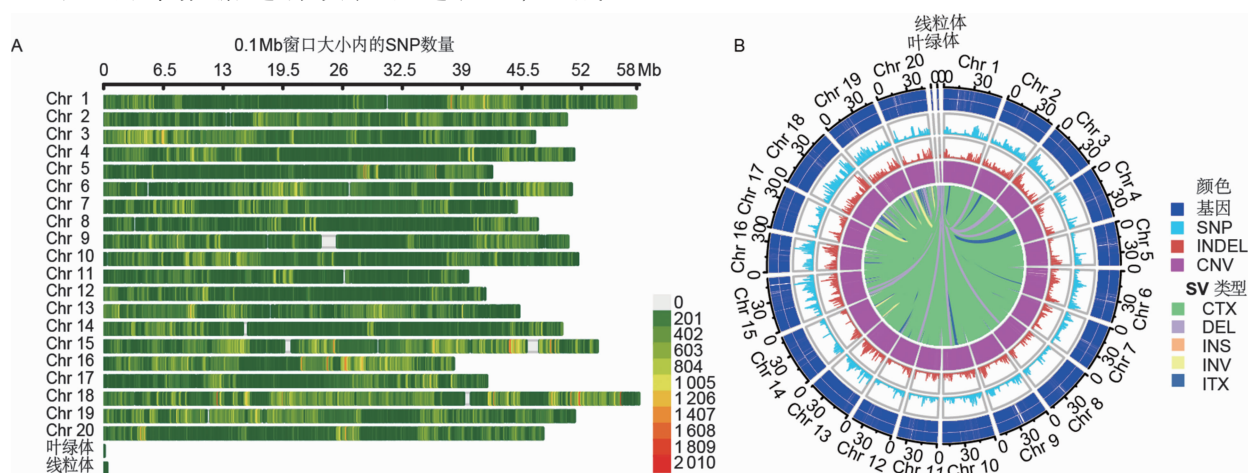
2.1 重测序原始数据质量评估

对通酥 1 号重测序数据进行质量评估,由结果

可知,GC 含量为 39.28%,模糊碱基(N)所占百分比为 0.01%,碱基识别准确率在 99% 以上的碱基(Q20)所占百分比为 98.33%,碱基识别准确率在 99.9% 以上的碱基(Q30)所占百分比为 95.78%。R1、R2 端测序 reads 单碱基平均质量值范围为 35~40,平均预估错误率低于 0.016%,说明测序数据不存在 AT、GC 分离现象,基本保持 AT、GC 数目一致。

对重测序数据进行质控和过滤后,生成

172 796 170 条高质量 reads,占原始 reads 条数的 96.57%;高质量 reads 的碱基数为 25.81 Gb,占原始碱基总数的 95.52%。平均覆盖深度为 27.08 ×,其中 1 ×、4 ×、10 ×、20 × 的覆盖深度在基因组的覆盖度分别为 93.04%、85.35%、71.06%、38.70%。变异在不同染色体上均有分布(图 1),以上结果表明,测序质量较好,可用于后续分析。



A—通酥 1 号染色体覆盖深度分布; B—通酥 1 号各类型变异在染色体的分布

图 1 通酥 1 号重测序数据分析

2.2 SNP 检测与注释分析

检测结果(图 1)表明,从通酥 1 号基因组中共获得 1 744 120 个 SNP,其中嘌呤和嘌呤之间或嘧啶和嘧啶之间的转换类型 T_s (T:A > C:G、C:G > T:A)为主要 SNP 突变型(1 150 576 个),所占比例为 65.97%,嘌呤和嘧啶之间的颠换类型 T_v (T:A > G:C、T:A > A:T、C:G > G:C、C:G > A:T)所占比例为 34.03%(593 544 个),转换/颠换比值(T_s/T_v)为 1.94。由 SNP 突变碱基偏好性统计分析结果可知,与 SNP 突变位点前后的碱基序列相比,SNP 突变位点处 A、T 碱基减少,G、C 碱基增多。SNP 杂合基因型数目为 354 610 个,纯合基因型数目为 1 389 510 个,总体杂合率为 20.33%。

统计每条染色体上 SNP 的分布情况,结果(图 1)显示,核基因组染色体的 SNP 分布密度为 1 051.26~3 001.07 个/Mb,叶绿体基因组的 SNP 分布密度为 295.63 个/Mb,线粒体基因组的 SNP 分布密度为 111.79 个/Mb。由 SNP 注释结果(图 2)可知,通酥 1 号的 SNP 变异主要分布在基因间区,共鉴定出 1 291 521 个(74.05%);其次是内含子区域,为 188 016 个(10.78%);转录起始位点上游

1 kb 区域、下游 1 kb 区域以及外显子区域的 SNP 数量类似,分别为 78 485 个(4.5%)、76 044 个(4.36%)、71 160 个(4.08%);在 3' UTR 区域、5' UTR 区域也存在少量的 SNP 变异,分别为 19 011 个(1.09%)、17 790 个(1.02%)。其中,在外显子区域内的 71 135 个 SNP 中,30 794 个(43.29%)为同义编码突变,38 292 个(53.83%)为非同义编码突变,832 个(1.17%)导致翻译提前终止,156 个(0.22%)导致终止密码子丢失。

对通酥 1 号 SNP 变异进行 GO 和 KEGG 数据库注释分析(图 3),根据 GO 分类数据,突变基因显著富集于细胞过程(4 114 个)、代谢过程(3 912 个)等 17 个生物学过程以及细胞解剖实体(2 151 个)等 2 个细胞组分和结合(6 239 个)、催化活性(4 421 个)等 15 个分子功能。KEGG 注释结果表明,变异基因涉及包含碳水化合物代谢(525 个)、信号转导(492 个)、翻译(306 个)、运输与分解代谢(304 个)、蛋白质折叠、分选和降解(300 个)、氨基酸合成、分解和相互转化(268 个)、环境适应(236 个)等在内的 31 个代谢通路。

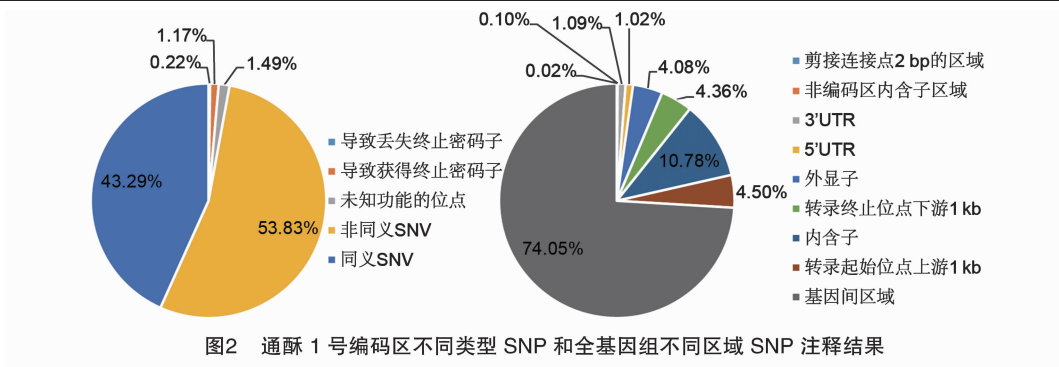


图2 通酥 1 号编码区不同类型 SNP 和全基因组不同区域 SNP 注释结果

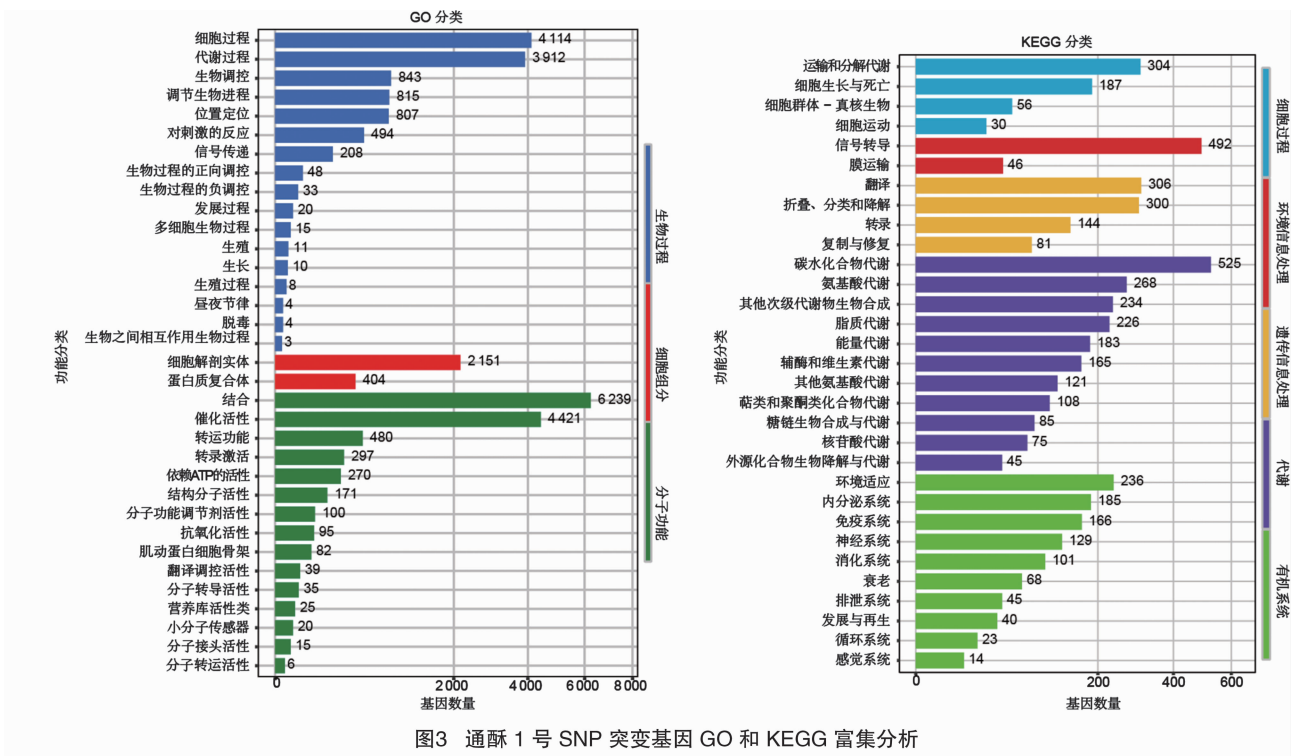


图3 通酥 1 号 SNP 突变基因 GO 和 KEGG 富集分析

2.3 InDel 检测分析与注释

经过过滤共筛选得到 232 213 个 InDel 变异,其中杂合位点 28 149 个(12.12%),纯合位点 204 064 个(87.88%); Insertion 插入位点 113 091 个(48.7%), Deletion 缺失位点 119 122 个(51.3%)。在全基因组范围内相同长度的 Insertion 插入位点和 Deletion 缺失位点比例基本一致(图 4),均存在较多的 1~4 bp 类型变异,而 ≥ 5 bp 类型变异相对占比较小。

对通酥 1 号的 InDel 变异进行注释统计,结果(图 5)发现,InDel 变异主要分布在基因间区域,共鉴定出 150 195 个(64.68%);其次是内含子区域,包括 36 504 个(15.72%);在转录起始位点上游 1 kb 区域、下游 1 kb 区域数量相似,分别为 15 024 个(6.47%)、15 280 个(6.58%),另外,在 5'UTR 区域、3'UTR 区域、外显子区域也存在少量的 SNP 变

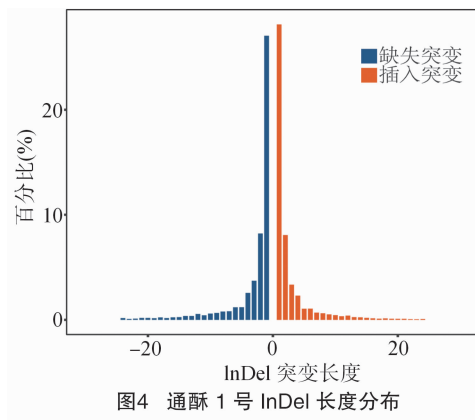


图4 通酥 1 号 InDel 长度分布

异,分别为 5 829 个(2.51%)、4 853 个(2.09%)、2 972 个(1.28%)。其中,在外显子区域内的 2 986 个 InDel 变异中,分别有 901 个(30.18%)、811 个(27.16%)导致非移码缺失、插入,652 个(21.85%)、521 个(17.45%)分别导致移码缺失、

插入, 38 个 (1.28%) 导致编码提前终止, 4 个 (0.13%) 导致终止密码子丢失。

对通酥 1 号的 InDel 变异进行 GO 和 KEGG 数据库注释分析(图 6), 根据 GO 分类数据, 突变基因显著富集于 GO:0009987 (565 个)、GO:0008152 (510 个) 等 14 个生物学过程以及 GO:0110165 (232

个) 等 2 个细胞组分和 GO:0005488 (953 个)、GO:0003824 (508 个) 等 16 个分子功能。KEGG 注释结果表明, 变异基因涉及包含信号转导(63 个)、碳水化合物代谢(39 个)、环境适应(35 个)、蛋白质折叠、分类和降解(30 个)、翻译(30 个) 等在内的 31 个代谢通路。

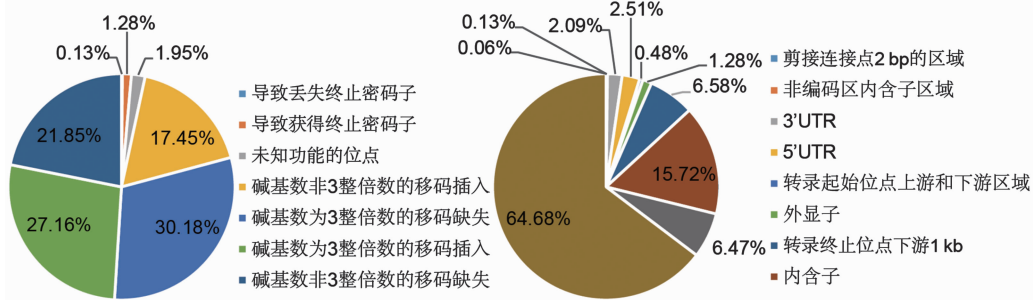


图5 通酥 1 号编码区上不同类型 InDel 和全基因组不同区域 InDel 注释结果

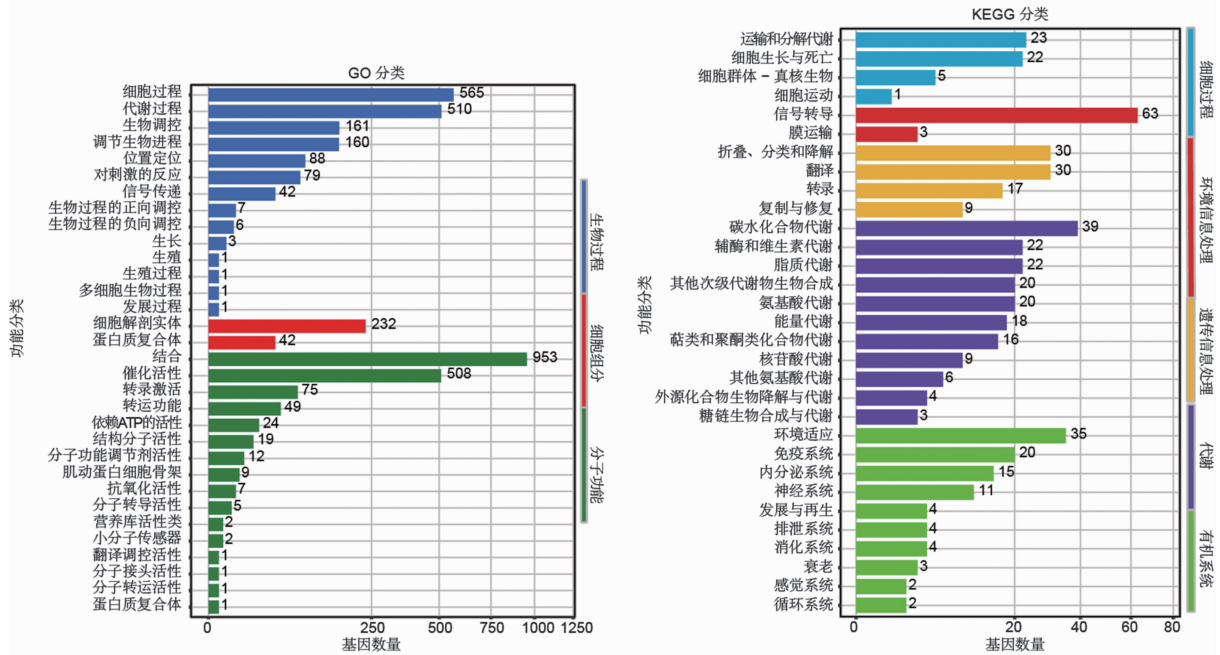


图6 通酥 1 号 InDel 突变基因 GO 和 KEGG 富集分析

2.4 CNV 与 SV 检测分析与注释

检测得出, 总 CNV 变异数为 8 732 个, 拷贝数缺失、增加个数分别为 739、7 993 个, 共涉及 5 982 个变异基因。KEGG 注释结果表明, 高频注释到的基因主要对应 ABC 转运蛋白(k14488)、酰基 - CoA 脱氢酶(k09422)、细胞色素 P450(k00430) 等功能, 表明可能参与糖等小分子跨膜运输、活跃核心能量代谢及对压力环境的适应等。总 SV 变异数为 5 749 个, 包括 1 926 个 (33.5%) 缺失、178 个 (3.1%) 倒位、2 255 个 (39.22%) 染色体内部易位、1 390 个 (24.18%) 染色体间易位, 共涉及 24 594 个

变异基因。KEGG 注释结果表明, 代谢通路主要涉及代谢活性增强、抗生素耐药性机制及环境适应相关通路等。

2.5 InDel 分子标记开发

为验证这些位点的多态性特征, 筛选分布在通酥 1 号 15 条染色体中的 31 个长度为 30 ~ 54 bp 的纯合 InDel 变异位点, 采用电子 PCR(e-PCR) 技术对 29 份已完成全基因组测序的大豆种质资源(包括冀豆 17、合丰 25、齐黄 34、豫豆 22、中黄 35、Fiskeby III、铜山天鹅蛋等代表性品种) 进行分子标记有效性验证。通过聚类分析发现, 14 个 InDel 位

点(占筛选位点 45.2%)在不同种质间呈现稳定的扩增多态性(表 1),可以有效解析 29 份大豆种质资源的遗传多样性特征(图 7),进一步将其转化为分子标记,用于进行群体鉴别。其中,InDel_1 变异导致 *Glyma. 02g290500* 基因产生非移码缺失,KEGG 注释途径为 k14776(β -葡萄糖苷酶),可以通过直接水解糖苷键,增加可溶性糖含量(甜度)、降解细胞壁多糖,改善质地(柔软度)、释放次生代谢物,调

节风味与营养等途径影响大豆鲜食特性。InDel_8 变异导致 *Glyma. 10g103100* 基因产生移码缺失,KEGG 注释途径为 k13412(天冬氨酸蛋白酶),可能在种子成熟后期水解大豆贮藏蛋白,释放游离氨基酸和小肽,影响豆粒质地软化和风味形成。InDel_12 变异导致 *Glyma. 16g194600* 基因产生移码缺失,KEGG 注释途径为 k10664(糖基转移酶),影响细胞壁结构与质地,直接影响豆粒的嫩度和咀嚼感。

表 1 14 个 InDel 分子标记信息

编号	染色体	位置	Williams 82	通酥 1 号	基因
1	Gm02	48831944	TTCACCCTCGTCAGAGTCACTATCAAAGTATACCTTTACTCTTT TTATGCAGTCTA	T	<i>Glyma. 02g290500</i>
2	Gm02	50061608	T	TCATTTGTCCGCACCAT TCTAACAGCTTTAC	<i>Glyma. 02g308600</i>
3	Gm03	3866902	ATAAAATTACAAGAAGAGAGTGGGTTGCCTTGAAAGG	A	-
4	Gm04	1014450	CGATGCTGGCGCGACGAATTATCCTGAGAT	C	<i>Glyma. 04g013300</i>
5	Gm07	20149006	T	TCCACTGCACCCCTTTGA GTGTTACAGCTTCACACC	<i>Glyma. 07g159600</i>
6	Gm08	44385355	CAATTGTGTATGCTGGCGTGAAATGGATACTTTGTTTTACAGT	C	-
7	Gm09	12267638	AATTGGAACAGAATGGGGAGAAGACCAACTTCCATTCTAGG	A	-
8	Gm10	21414835	AAGTCACTATATTTTCACTTTTTTTTATTCAATTTGATTTTAT	A	<i>Glyma. 10g103100</i>
9	Gm13	28033266	A	ATGAGGCACGAGAAG GTAATGCCAACCCAGG	<i>Glyma. 13g172100</i>
10	Gm13	32370867	GAGGAACAGGTGTTGGTATCTGGAAGTTCAAATC	G	<i>Glyma. 13g216400</i>
11	Gm13	42180045	ATCATTATTAGAGTTTTAGGATTTATAAAGAGAAATACAAGTACT	A	<i>Glyma. 13g334400</i>
12	Gm16	35889588	GTGAAATCCGTTAGACAGACGGCGCATTCGAGCGTTTGGTTGGCC TTATT	G	<i>Glyma. 16g194600</i>
13	Gm17	39474344	ATTTTAGCATGAATCTGGGTGGTGTAAACAGGTTTT	A	<i>Glyma. 17g238300</i>
14	Gm18	2089656	AGCCACAGGAATGATCAACATCCTCGATGCAT	A	-

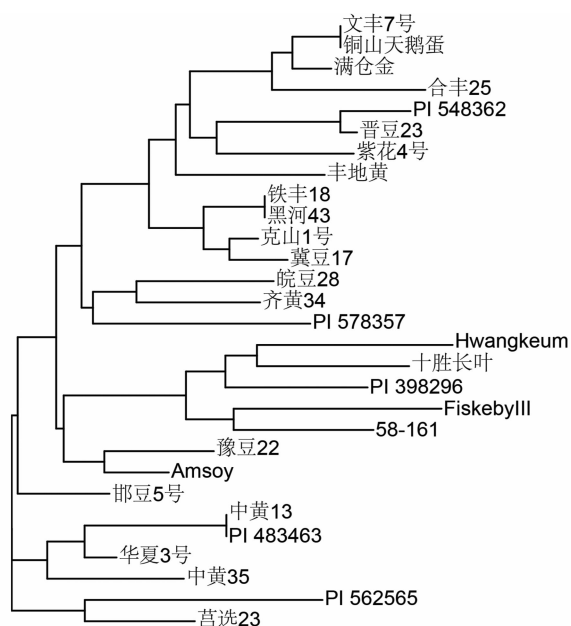


图 7 29 个品种基于 14 个 InDel 标记扩增的进化树分析

3 结论与讨论

随着大豆种质资源全基因组测序工作的完成,相关遗传图谱的解析显著推进了新一代分子育种技术的发展进程^[21-23]。本研究对鲜食大豆优异品种通酥 1 号开展高深度全基因组重测序(平均覆盖深度 27.08 \times),较已报道的鲜食大豆品种“浙鲜豆 8 号”“浙鲜 9 号”的全基因组重测序深度提升约 3.5 倍,显著提高了 CNV 和 SV 的检测精度,为精准定位调控豆荚发育与籽粒风味形成的关键基因组区段提供了数据依据,为鲜食大豆品质定向改良的分子标记开发奠定了技术基础^[24]。

以 Williams 82 基因组序列为参考基因组,在鲜食大豆品种通酥 1 号中共检测到 1 744 120 个 SNP,转换/颠换比(T_s/T_v)为 1.94,SNP 位点中的 G/C 碱基比例增加,暗示突变可能受 DNA 修复机制或环境

适应性选择驱动^[25]。功能注释结果显示,73.66% 的 SNP 位于基因间区,而在位于外显子区域的 4.08% (71 135 个) 变异中,53.83% (38 288 个) 为非同义突变,可能导致 2 883 个基因的蛋白质功能改变。另外,686 个基因终止密码子提前发生,154 个基因终止密码子丢失,可能导致蛋白延长,这些基因中可能存在与鲜食特性(如糖分积累、质地调控)相关的候选基因,需进一步进行验证。

在通酥 1 号中共鉴定出 232 213 个 InDel 变异,缺失频率(51.3%)与插入(48.7%)频率接近,但缺失稍占优势,符合植物基因组中 InDel 变异的普遍特征,可能与复制滑动和 DNA 修复机制偏向性有关^[26]。与水稻等大多数物种一致,短片段(1~4 bp)变异占比明显高于长片段(≥ 5 bp),可能与 DNA 复制过程中滑动错配机制偏好短片段插入或缺失有关,且短片段变异对基因组稳定性的破坏较小,尤其在非编码区或冗余密码子区域,可能被自然选择容忍^[27-29]。值得注意的是,在外显子区域内的 2 986 个 InDel 变异中,1 710 个(57.34%)导致非移码缺失、插入,1 173 个(39.3%)导致移码缺失、插入,均可能改变相关蛋白质的结构和功能,影响蛋白质稳定性及其与其他蛋白质和核酸的相互作用,引发多种疾病;约 1.28% (39 个)引起 *Glyma. 02g123100* 等 22 个基因提前终止密码子,可能显著影响基因功能,约 0.13% (5 个)导致 *Glyma. 15g187700* 等 5 个基因终止密码子丢失,可能通过延长蛋白翻译产生功能获得性突变,这类变异在植物适应性进化中具有潜在作用^[30-33]。功能注释分析结果显示,InDel 相关基因显著富集于 GO:0009987(细胞过程调控)和 GO:0005488(结合功能)等条目中,提示变异可能影响基础代谢与信号转导网络。KEGG 通路中信号转导(63 个基因)与环境适应(35 个基因)的富集,进一步支持 InDel 在植物响应生物/非生物胁迫中的潜在贡献^[34]。此外,碳水化合物代谢与蛋白质折叠相关通路的变异可能与通酥 1 号特有农艺性状(如抗逆性或品质形成)存在关联,但需进一步验证。本研究为解析该品种的遗传特性提供了重要数据基础,后续可通过基因编辑或关联分析明确关键 InDel 的功能效应^[35]。

CNV 分散在植物基因组中,在玉米中发现的许多 CNV 与驯化有关,基因 *Vrn-A1* 的 CNV 与中间或晚开花表型相关;SV 可以通过复制事件产生,在

谱系之间存在基因的差异丢失^[36-38]。大豆是第 1 个对其基因组进行 CNV 分析的豆类物种,在测定的基因组中共检测到 267 个平均大小为 18~23 kb 的 CNV^[6]。本研究在鲜食大豆通酥 1 号中检测到的 8 732 个 CNV 和 5 749 个 SV,它们可能通过基因剂量效应或染色体重排影响表型^[39]。比如,马铃薯中受 CNV 影响的 4 个基因在转录中显示出剂量效应,并且可能影响马铃薯植物的生长和发育^[40]。

由于中等长度 InDel 标记易于转化为通用型分子标记(如 CAPS/dCAPS),适配不同实验室技术条件,利于后续大规模群体检测,且可以保证 PCR 扩增效率(片段过短易产生引物二聚体,过长则导致扩增难度增加),通过常规电泳清晰区分多态性,本研究筛选到分布在通酥 1 号 15 条染色体中的 31 个长度为 30~54 bp 的纯合 InDel 变异位点,并对其进一步进行分子标记有效性验证,其中 14 个位点(45.2%)在种质间具有稳定扩增多态性,可以成功解析遗传多样性,将其转化为群体鉴别标记,标记 InDel_1、InDel_8、InDel_12 分别导致 *Glyma. 02g290500*、*Glyma. 10g103100*、*Glyma. 16g194600* 基因发生序列变异,基因功能注释结果表明,它们可能通过调节风味、影响细胞壁结构与质地等途径影响大豆鲜食特性。这些标记的进一步转化将提升育种效率,加速优质鲜食大豆品种的选育进程。

参考文献:

- [1] Jiao Y P, Zhao H N, Ren L H, et al. Genome-wide genetic changes during modern breeding of maize [J]. *Nature Genetics*, 2012, 44 (7): 812-815.
- [2] Nishida H, Yoshida T, Kawakami K, et al. Structural variation in the 5' upstream region of photoperiod-insensitive alleles Ppd-A1a and Ppd-B1a identified in hexaploid wheat (*Triticum aestivum* L.), and their effect on heading time [J]. *Molecular Breeding*, 2013, 31 (1): 27-37.
- [3] Díaz A, Zikhali M, Turner A S, et al. Copy number variation affecting the Photoperiod-B1 and Vernalization-A1 genes is associated with altered flowering time in wheat (*Triticum aestivum*) [J]. *PLoS One*, 2012, 7 (3): e33234.
- [4] Lu P L, Han X W, Qi J, et al. Analysis of *Arabidopsis* genome-wide variations before and after meiosis and meiotic recombination by resequencing *Landsberg erecta* and all four products of a single meiosis [J]. *Genome Research*, 2012, 22 (3): 508-518.
- [5] Xu X, Liu X, Ge S, et al. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes [J]. *Nature Biotechnology*, 2011, 30 (1): 105-111.
- [6] McHale L K, Haun W J, Xu W W, et al. Structural variants in the

- soybean genome localize to clusters of biotic stress – response genes [J]. *Plant Physiology*, 2012, 159(4): 1295 – 1308.
- [7] DeBolt S. Copy number variation shapes genome diversity in *Arabidopsis* over immediate family generational scales [J]. *Genome Biology and Evolution*, 2010, 2: 441 – 453.
- [8] Sutton T, Baumann U, Hayes J, et al. Boron – toxicity tolerance in barley arising from efflux transporter amplification [J]. *Science*, 2007, 318(5855): 1446 – 1449.
- [9] Muñoz – Amatriáin M, Eichten S R, Wicker T, et al. Distribution, functional impact, and origin mechanisms of copy number variation in the barley genome [J]. *Genome Biology*, 2013, 14(6): R58.
- [10] Bai H, Cao Y H, Quan J Z, et al. Identifying the genome – wide sequence variations and developing new molecular markers for genetics research by re – sequencing a Landrace cultivar of foxtail millet [J]. *PLoS One*, 2013, 8(9): e73514.
- [11] Springer N M, Ying K, Fu Y, et al. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content [J]. *PLoS Genetics*, 2009, 5(11): e1000734.
- [12] Swanson – Wagner R A, Eichten S R, Kumari S, et al. Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor [J]. *Genome Research*, 2010, 20(12): 1689 – 1699.
- [13] Beló A, Beatty M K, Hondred D, et al. Allelic genome structural variations in maize detected by array comparative genome hybridization [J]. *Theoretical and Applied Genetics*, 2010, 120(2): 355 – 367.
- [14] Zheng L Y, Guo X S, He B, et al. Genome – wide patterns of genetic variation in sweet and grain Sorghum (*Sorghum bicolor*) [J]. *Genome Biology*, 2011, 12(11): R114.
- [15] 包明艳, 汪洪洋, 宋昌博. 菜用大豆通酥1号高产栽培技术 [J]. *现代农业科技*, 2006(4): 56.
- [16] Chen S F. Ultrafast one – pass FASTQ data preprocessing, quality control, and deduplication using fastp [J]. *iMeta*, 2023, 2(2): e107.
- [17] Li H, Durbin R. Fast and accurate short read alignment with Burrows – Wheeler transform [J]. *Bioinformatics*, 2009, 25(14): 1754 – 1760.
- [18] Zhu P Y, He L Y, Li Y Q, et al. Correction: OTG – snpcaller: an optimized pipeline based on TMAP and GATK for SNP calling from ion torrent data [J]. *PLoS One*, 2015, 10(9): e0138824.
- [19] Wang K, Li M Y, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high – throughput sequencing data [J]. *Nucleic Acids Research*, 2010, 38(16): e164.
- [20] Gu Z G, Gu L, Eils R, et al. *circize* Implements and enhances circular visualization in R [J]. *Bioinformatics*, 2014, 30(19): 2811 – 2812.
- [21] Lam H M, Xu X, Liu X, et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection [J]. *Nature Genetics*, 2010, 42(12): 1053 – 1059.
- [22] Schmutz J, Cannon S B, Schlueter J, et al. Genome sequence of the Palaeopolyploid soybean [J]. *Nature*, 2010, 463(7278): 178 – 183.
- [23] 张中起, 梁邦平, 王俊涛, 等. 大豆 VLN 基因家族的全基因组鉴定及组织表达分析 [J]. *江苏农业科学*, 2023, 51(3): 35 – 45.
- [24] 郭丹丹, 袁凤杰, 郁晓敏. 基于重测序的籽粒型和鲜食型大豆的全基因组变异分析 [J]. *分子植物育种*, 2019, 17(22): 7306 – 7312.
- [25] Eyre – Walker A, Hurst L D. The evolution of isochores [J]. *Nature Reviews. Genetics*, 2001, 2(7): 549 – 555.
- [26] Cao J, Schneeberger K, Ossowski S, et al. Whole – genome sequencing of multiple *Arabidopsis thaliana* populations [J]. *Nature Genetics*, 2011, 43(10): 956 – 963.
- [27] Zhou X, Li J L, Chen L, et al. The genomic pattern of insertion/deletion variations during rice improvement [J]. *BMC Genomics*, 2024, 25(1): 1263.
- [28] Alkan C, Coe B P, Eichler E E. Genome structural variation discovery and genotyping [J]. *Nature Reviews (Genetics)*, 2011, 12(5): 363 – 376.
- [29] Ellegren H. Microsatellites: simple sequences with complex evolution [J]. *Nature Reviews (Genetics)*, 2004, 5(6): 435 – 445.
- [30] 褚昕璐. 非移码插入缺失突变对蛋白质相互作用影响分析及其致病性预测方法研究 [D]. 合肥: 安徽大学, 2021.
- [31] 陈华夏, 周成博, 邢永忠. 水稻 *Dwarf1* 移码突变的新突变体鉴定 [J]. *遗传*, 2011, 33(4): 397 – 403.
- [32] Lu S J, Zhao X H, Hu Y L, et al. Natural variation at the soybean J locus improves adaptation to the tropics and enhances yield [J]. *Nature Genetics*, 2017, 49(5): 773 – 779.
- [33] Yuan X B, Jiang X Y, Zhang M Z, et al. Integrative omics analysis elucidates the genetic basis underlying seed weight and oil content in soybean [J]. *The Plant Cell*, 2024, 36(6): 2160 – 2175.
- [34] 马进, 郑钢. 南方型紫花苜蓿叶片盐胁迫应答基因鉴定与分析 [J]. *农业生物技术学报*, 2015, 23(12): 1531 – 1541.
- [35] Hicks K G, Cluntun A A, Schubert H L, et al. Protein – metabolite interactomics of carbohydrate metabolism reveal regulation of lactate dehydrogenase [J]. *Science*, 2023, 379(6636): 996 – 1003.
- [36] Chia J M, Song C, Bradbury P J, et al. Maize HapMap2 identifies extant variation from a genome in flux [J]. *Nature Genetics*, 2012, 44(7): 803 – 807.
- [37] Lyra D H, Galli G, Alves F C, et al. Modeling copy number variation in the genomic prediction of maize hybrids [J]. *Theor Appl Genet*, 2019, 132(1): 273 – 288.
- [38] Feuk L, Marshall C R, Wintle R F, et al. Structural variants: changing the landscape of chromosomes and design of disease studies [J]. *Human Molecular Genetics*, 2006, 15: R57 – R66.
- [39] Hollox E J, Zuccherato L W, Tucci S. Genome structural variation in human evolution [J]. *Trends in Genetics*, 2022, 38(1): 45 – 58.
- [40] Iovene M, Zhang T, Lou Q F, et al. Copy number variation in potato – an asexually propagated autotetraploid species [J]. *The Plant Journal*, 2013, 75(1): 80 – 89.