

基于 SLAF-BSA 技术挖掘大豆酸性磷酸酶 候选基因及标记开发

刘 渊¹, 孔佑宾¹, 李喜焕¹, 张彩英^{1,2}

(¹河北农业大学农学院/分子遗传育种实验室,保定 071001;²河北农业大学生命科学学院/植物资源开发与利用实验室,保定 071001)

摘要: 磷利用效率与大豆产量密切相关,根尖酸性磷酸酶活性是筛选大豆品种磷效率的重要指标。挖掘酸性磷酸酶活性候选基因并开发其功能标记对获得磷高效功能基因、解析磷利用分子机制和培育磷高效大豆新品种意义重大。本研究利用酸性磷酸酶活性重组自交系群体 F₁₂ 构建了 2 个极端性状混池 DNA 文库,通过 SLAF-BSA 技术,获得了 268 个与大豆酸性磷酸酶活性相关的 SNP,包括 12 个非同义突变,其中亲本间 7 个,后代混池间 5 个;在 2 个关联候选区域,获得 79 个酸性磷酸酶活性相关基因,其中第 3 号染色体的 20138271~20268154 间 4 个,17 号染色体的 14368648~15526449 间 75 个;对该区域内基因进行了功能注释。开发了非同义突变基因 *Glyma.17G166200.1* 功能标记 GMsnp-B,用该标记检测 169 份大豆栽培品种基因型,与表型符合率达到 82.8%。

关键词: 大豆;酸性磷酸酶;SLAF-BSA;候选基因;功能标记

Mining Acid Phosphatase Candidate Genes and Development of Functional Markers Based on SLAF-BSA in Soybean

LIU Yuan¹, KONG You-bin¹, LI Xi-huan¹, ZHANG Cai-ying^{1,2}

(¹College of Agronomy, Hebei Agricultural University/Laboratory of Molecular Genetics And Breeding, Baoding 071001;

²College of Life Sciences, Hebei Agricultural University/Laboratory of Plant Resource Development and Utilization, Baoding 071001)

Abstract: The activity of acid phosphatase in root tips is an important index for screening phosphorus efficiency of soybean varieties. Mining for candidate genes with acid phosphatase activity and exploiting their functional markers will bring great significance for identifying the phosphorus-efficient genes, unlocking the molecular mechanism of phosphorus utilization as well as breeding for new phosphorus-efficient soybean varieties. In this study, two DNA libraries with bulked pools of segregants from a F₁₂ recombination inbred line population were constructed for next generation sequencing. 268 significant SNPs were obtained by SLAF-BSA technique. That included 12 gene-based non-synonymous mutations, seven of which were found between parents and five of which were found between offspring pools. Particularly, two candidate regions were identified, where 79 acid phosphatase-related genes were resided. Four genes are allocated from 20138271 to 20268154 on chromosome 3 and 75 genes are allocated from 1436648 to 15526449 on chromosome 17. Furthermore, a genetic marker GMsnp-B, which targets to the non-synonymous mutation gene *Glyma.17G166200.1*, was developed and tested using 169 soybean cultivars. The coincidence rate with phenotype activities in root tips was 82.8%.

Key words: soybean; acid phosphatase; SLAF-BSA; candidate gene; functional marker

收稿日期: 2019-04-28 修回日期: 2019-05-09 网络出版日期: 2019-05-30

URL: <http://doi.org/10.13430/j.cnki.jpgr.20190428001>

第一作者研究方向为作物育种目标性状的遗传与改良, E-mail: liuyuan_2005bd@126.com

通信作者: 张彩英, 研究方向为植物资源开发与利用, E-mail: cyzhang_60@126.com

基金项目: 现代农业科技奖励性后补助资金项目(17927670H); 河北省科技计划项目(16227516D)

Foundation project: Modern Agricultural Science and Technology Award Post-subsidy Fund Project(17927670H), Hebei Science and Technology Plan Project(16227516D)

大豆是重要的粮油作物,其生长期对土壤有效磷的吸收量远远高于水稻、小麦和玉米^[1]。同时,土壤缺磷还会抑制大豆根瘤形成,降低固氮效率,促使花荚脱落,最终严重影响其产量和品质,因此土壤有效磷含量成为大豆生长及产量形成最重要的限制因素^[2]。酸性磷酸酶(AP, acid phosphatase)是植物体内一种重要的诱导酶和水解酶,能够分解土壤中的有机磷释放无机磷供植物生长发育。关于磷胁迫下酸性磷酸酶活性显著提高的研究结果在玉米^[3-4]、花生^[5]、小麦^[6]、甜菜^[7]等作物中均有报道。Liang 等^[8]认为缺磷条件下,普通菜豆酸性磷酸酶 *PvPAP1* 和 *PvPAP3* 基因在叶和根内表达活性升高,增加了磷的吸收利用。作者前期研究了低磷处理后大豆叶片和根尖酸性磷酸酶活性、根冠比、干物质量、全磷含量及磷利用率,表明根尖酸性磷酸酶活性(APA, acid phosphatase activity)与相应的磷效率存在显著正相关,根尖酸性磷酸酶活性可以作为筛选大豆品种磷效率的重要指标^[9]。

大豆是自花授粉作物,基因组复杂且重复序列多, QTL 精细定位及图位克隆程序繁琐。随着测序技术的发展, SLAF-BSA 使得发掘功能基因和构建高密度连锁图谱成为可能,其利用遍布整个基因组的海量信息,可实现候选功能区的精细定位,迅速准确获得大批候选基因和高通量分子标记^[10-13]。关联分析被广泛应用于作物重要性状功能基因和连锁标记的挖掘^[14-16], Ning 等^[17]应用关联分析对大豆耐低磷 5 个性状(株高、茎干重、磷浓度等)进行研究,检测到 19 个显著关联的 SNPs。Zhang 等^[18-19]克隆与磷效率关联 QTL (qpe8) 的基因 *GmACP1*, 进而利用关联分析检测到 4 个与磷效率相关的 QTL 及酸性磷酸酶基因 *GmACP2*。功能标记是基于功能基因序列开发的分子标记,可准确鉴定品种的基因型,从而更准确地对育种后代进行早代选择。Juwattanasomran

等^[20]发现大豆芳香气味的等位基因 *GmBADH2* 外显子缺失 10 bp 核苷酸序列,根据基因差异设计共显性分子标记,并开发了可以区分不同芳香气味大豆的功能标记 *Gmbadh2-1* 和 *Gmbadh2-2*。Bachleda 等^[21]在大豆 N87-2122-4 中发现降低棕榈酸含量的基因 *FAT1a*,根据其在基因组区域的缺失信息,开发荧光功能标记并在 F₂ 分离群体获得验证。

综上所述,磷对大豆生长发育至关重要,利用 SLAF-BSA 技术发掘磷高效基因、开发其功能标记对解析磷对大豆生长发育作用的分子机制及培育磷高效品种至关重要。鉴于此,本研究基于 SLAF-BSA 技术,利用 2 个亲本和 RIL 群体 F₁₂ 2 个极端性状混池 DNA 文库对大豆酸性磷酸酶活性进行关联分析,获得酸性磷酸酶活性相关基因的候选区域,结合候选基因的功能注释和富集分析预测基因功能;利用发掘的功能基因,开发其功能标记并对其实用性进行评价。研究结果为大豆磷高效分子机制解析和磷高效分子育种奠定理论和物质基础。

1 材料与方法

1.1 试验材料

以冀豆 12 (APA 0.0197 $\mu\text{mol}/\mu\text{g} \cdot \text{min}$) 为母本,冀 NF58 (APA 0.1777 $\mu\text{mol}/\mu\text{g} \cdot \text{min}$) 为父本构建重组自交系 F₁₂ 群体(由河北省农林科学院粮油所提供),包含 133 个株系,分别选取其中 35 个具有高 APA 和低 APA 的单株组成 2 个混池,用于发掘酸性磷酸酶关联候选基因和分子标记。根据候选基因 *Glyma.17G166200.1* 序列设计引物(上游引物 5'-TTGGACTCTATTGCCCTTGCC-3', 下游引物 5'-AGGCTTCAATTTCCCGATTT-3');根据基因 *Glyma.17G166200.1* 的位点 15166211bp 标记 GMsnp01 (T) 和位点 15166218bp 标记 GMsnp02 (C) 的差异设计 3 对引物。设计出的引物序列(表 1),利用 169 个

表 1 SNP 引物序列

Table 1 Sequence of SNP marker

标记 Marker	引物 Primer	序列 Sequence	退火温度(°C) Temperature	引物大小(bp) Length	产物(bp) Product
GMsnp01 (T)	GMsnp-A	5'-CCAGATGATAGTTATATCTC-3'	59.83	20	140
GMsnp02 (C)		5'-GTGCCGTACACATCACTTGG-3'			
	GMsnp-B	5'-CCAGATGATAGTTATATCTC-3'	59.96	20	217
		5'-GGACAACAACACTGATCATC-3'			
	GMsnp-C	5'-CCAGATGATAGTTATATCTC-3'	59.93	20	109
		5'-TATTTGCGTTTGGACCATCA-3'			

栽培品种,其中包括 26 个磷高效品种、69 个磷中效品种和 74 个磷低效品种,进行功能标记验证。

1.2 试验方法

1.2.1 BSA 混池构建 分别于 2015 年 6 月 21 日和 2016 年 6 月 19 日连续 2 年重组自交系 RIL F₁₂ 进行低磷(2 μ mol/L)处理,测定 133 个株系的根尖酸性磷酸酶活性^[22]。分别选取 35 个具有高、低根尖酸性磷酸酶活性单株,采用 CTAB 法提取高质量基因组 DNA,将 35 个低酸性磷酸酶活性单株 DNA 样品等量混合,组成 F₁₂ 低活性池(B1 混池,编号为 B1);将 35 个高酸性磷酸酶活性单株 DNA 样品等量混合,组成 F₁₂ 高活性池(B2 混池,编号为 B2);同时提取父本冀 NF58(编号为 P)和母本冀豆 12(编号为 M)的基因组 DNA。

1.2.2 SLAF 文库的构建 根据选定的最适酶切方案,对检测合格的各样品基因组 DNA 分别进行酶切。对得到的酶切片段(SLAF 标签)进行 3' 端加 A 处理、连接 Dual-index 测序接头、PCR 扩增、纯化、混样、切胶选取目的片段,文库质检合格后用 IlluminaHiSeqTM, PE125bp 进行测序。为评估酶切实验的准确性,选用品种冀豆 12 作为对照(Control)进行测序。通过将 reads 与参考基因组比对,在亲本和混池中开发 SLAF 标签,寻找在亲本中存在多态性的 SLAF 标签和有 reads 覆盖区域的 SNP。SNP 的检测主要使用 GATK^[23] 软件工具包实现。利用得到的 SLAF 标签和 SNP 标记位点对大豆根尖酸性磷酸酶活性进行关联分析,获得与之紧密关联的位点。

1.2.3 关联分析 通过欧氏距离(ED, euclidean distance)算法^[24],利用测序数据寻找混池间存在显著差异的 SNP 位点,统计各个碱基在不同混池中的深度,并计算每个位点 ED 值,以此评估与性状关联区域。采用 SNPNUM 方法可对同一条染色体上标

记的 Δ SNP-index 值进行拟合^[25],选择关联阈值以上的区域作为与性状相关的区域^[26-27]。

1.2.4 候选区域的 SNP 和基因功能注释 应用 BLAST^[23] 软件对候选区间内的编码基因进行多个数据库(NR^[25]、Swiss-Prot、GO^[26]、KEGG^[27]、COG^[28])的深度注释。

1.2.5 候选区域非同义突变基因功能标记的开发 利用 Primer Premier 5.0 软件,根据非同义突变等位基因的特异性差异设计引物,对高、低活性亲本及 169 个自然材料基因组 DNA 进行 PCR 扩增。利用温度梯度 PCR 选择最佳解链温度,经琼脂糖凝胶电泳分离差异条带,获得功能标记并验证其实用性。

2 结果与分析

2.1 SLAF 标签和 SNP 标记开发

经 Illumina HiSeq TM 2500 测序平台进行测序,对父本(P)、母本(M)、低活性(B1)和高活性(B2)的 SLAF 文库样品测序数据的 reads 数量、测序质量值(Q30)和 GC(guanine and cytosine)含量进行统计,共获得约 442.18 万 reads 数据,具体结果见表 2。Q30 是评估高通量测序单碱基错误率的重要指标,测序质量值越高对应的碱基测序错误率越低,本研究中测序质量值 Q30 的范围在 83.52%~85.74%,说明测序碱基错误率低,所获数据合格。测序获得平均 GC 含量为 39.54%,说明达到测序要求。

利用大豆基因组,参照表 2 可知,本研究共开发 156730 个 SLAF 标签,SLAF 标签亲本平均测序深度(平均每个 SLAF 上对应亲本的测序 reads 数)为 55.845,混池平均测序深度(平均每个 SLAF 上对应混池样品的测序 reads 数)为 59.075。

SNP 的检测主要使用 GATK 软件工具包实现,样品间的 SNP 注释具体统计结果共得到 136473

表 2 样品测序数据评估与 SLAF 标签统计

Table 2 Statistics of sample sequencing data assessment and SLAF tag

样品编号 Sample ID	过滤后的 reads 数 Clean_reads	过滤后的 碱基数 Clean_base	测序质量值百分比 (%) The percentage of Q30	GC 含量(%) Guanine and cytosine content	SLAF 标签数 SLAF number	测序总深度 Total depth of sequence	测序平均深度 Average depth of sequence
父本 P	10908819	2179494610	84.63	38.92	150760	8852443	58.72
母本 M	10467438	2091194012	83.52	40.56	149234	7904962	52.97
低活性 B1	11585919	2314419934	85.48	39.22	154939	9517970	61.43
高活性 B2	11255795	2247167080	85.74	39.47	155484	8818299	56.72
合计 Total	44217971	2208068909	84.84	39.54	156730	35093674	57.46

个 SNP, 亲本间得到 92123 个 SNP, 后代混池间得到 47916 个 SNP。亲本间非同义编码突变 SNP 有 2269 个, 后代混池间非同义编码突变 SNP 有 1221 个。不同染色体上的 SLAF 标签与 SNP 标记的分布见图 1。由图可见, 开发的 SLAF 标签和 SNP 标记分布较均匀。

在关联分析前, 总的 SNP 位点为 136473 个, 过滤掉有多个等位基因型的 SNP 位点 117 个, read 支

持度小于 4 的 SNP 位点 83311 个, 混池基因型一致的 SNP 位点 3802 个以及利用亲本过滤的 SNP 位点 5796 个, 最终得到高质量的可靠 SNP 位点 43447 个进行后续分析。

2.2 酸性磷酸酶活性关联分析

2.2.1 ED 方法关联分析 取原始 ED 的 2 次方作为关联值以达到消除背景噪音的功能, 然后采用 LOESS 方法对 ED 值进行拟合, 分布如图 2 所示。

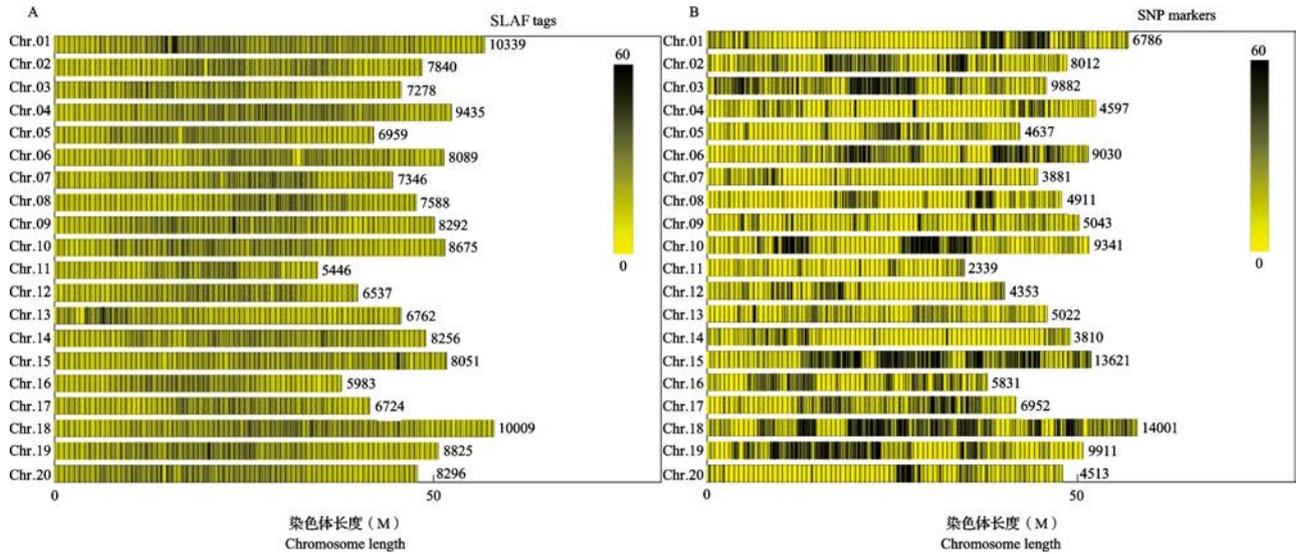
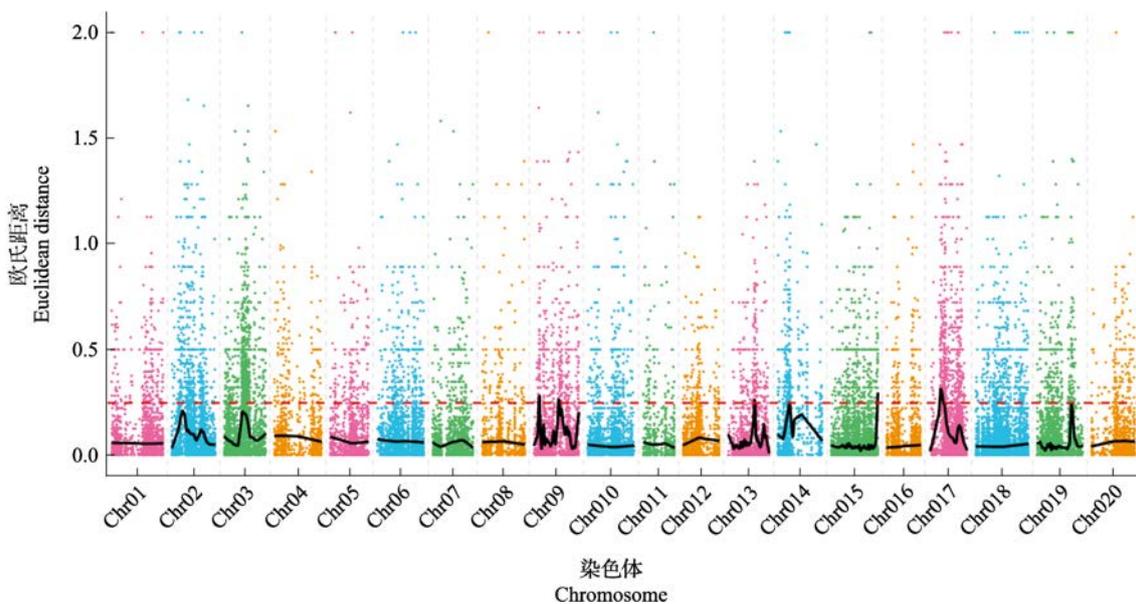


图 1 SLAF 标签 (A) 和 SNP 标记 (B) 在基因组各染色体上的分布
Fig.1 Distribution of SLAF tag (A) and SNP marker (B) on chromosomes



彩色的点代表每个 SNP 位点的 ED 值, 黑色的线为拟合后的 ED 值, 红色的虚线代表显著性关联阈值, ED 值越高, 代表该点关联效果越好
Color spots represents ED of SNP site, and black line is fitting for ED value, the red dotted line represents the relationship between the threshold, the higher the ED value, and the better on behalf of the association

图 2 ED 值在染色体上的分布整合图
Fig.2 The integrated map of the distribution of ED association values on chromosome

取所有位点拟合值的 median 和 3SD 作为分析的关联阈值^[23], 计算得 0.19。根据关联阈值判定, 共得到 11 个关联区域, 总长度为 27.19 Mb, 共包含 2012 个基因, 其中非同义突变 SNP 位点的基因共 46 个(表 3)。

表 3 关联区域信息统计表

Table 3 Information table of association area

染色体编号 Chromosome_ID	起始位置 Start	终止位置 End	大小 (Mb) Size	基因数量 Gene_number
Chr.02	10769072	13792035	3.02	349
Chr.03	19492903	23684087	4.19	87
Chr.09	5377858	7048565	1.67	175
Chr.09	26935629	29455875	2.52	80
Chr.09	30002545	31362349	1.36	38
Chr.09	49764885	50072564	0.31	47
Chr.13	28512724	30195760	1.68	261
Chr.14	10890557	13971841	3.08	155
Chr.15	50881356	51739708	0.86	90
Chr.17	11361908	17778751	6.42	496
Chr.19	37471924	39552267	2.08	234
合计 Total	—	—	—	2,012

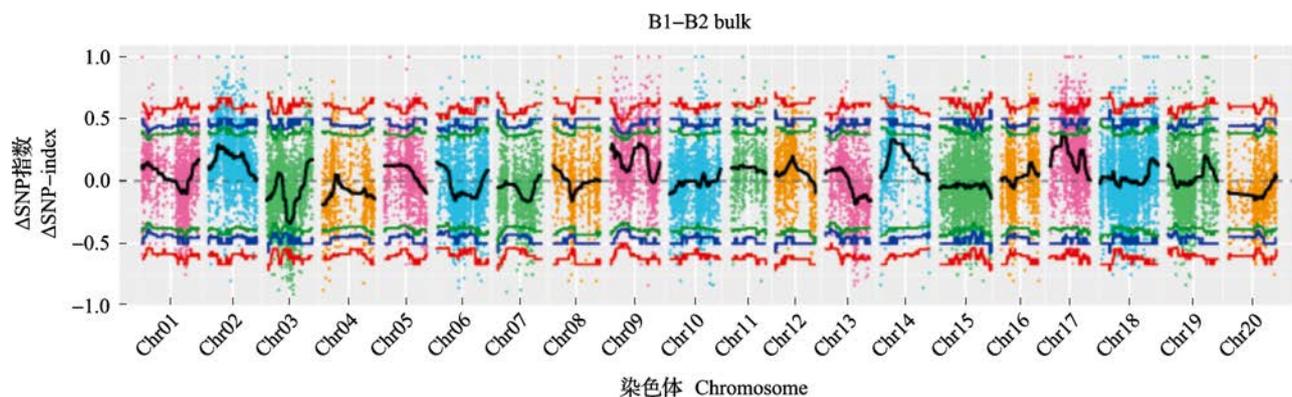
2.2.2 SNP 指数方法关联分析 低活性 B1、高活性 B2 两个后代混池分别的 SNP 指数, 以及采用 SNPNUM 方法对 Δ SNP 指数进行拟合后 Δ SNP 指数的分布图见图 3, 根据计算机模拟实验^[25]计算结果, 当置信度为 0.90 时, 2 个关联区域位置拟合后 Δ SNP 指数值无限靠近绿色线, 分别位于第 3 号

染色体和第 17 号染色体上, 总长度为 1.29 Mb, 共包含 79 个基因, 其中非同义突变基因 2 个。

2.2.3 关联候选区域筛选 对 ED 和 SNP 指数两种关联分析方法得到的关联区域进行综合分析, 取交集得到的结果与 SNP 指数的关联分析结果基本一致, 在外显子区域两个亲本以及后代之间差异共获得 268 个 SNP(表 4), 获得 2 个关联区域, 分别分布在第 3 号染色体的 20138271~20268154 且跨度 0.13 Mb 的区域和第 17 号染色体的 14368648~15526449 且跨度 1.158 Mb 的区域。共获得 79 个基因, 其中第 3 号染色体的关联区域内有 4 个基因, 第 17 号染色体的关联区域内有 75 个基因, 2 个非同义突变基因都处于第 17 号染色体关联区域内, 包括 *Glyma.17G165600.1* 位于 15082556~15084107 的区域, 基因长度 1551bp, *Glyma.17G166200.1* 位于 15165279~15168381 的区域, 基因长度 3103bp。结果表明, SLAF-BSA 将控制大豆酸性磷酸酶活性的基因定位在第 3 号染色体和第 17 号染色体上。

2.3 关联候选区域功能注释与富集

亲本间、后代混池间在候选区域内的 SNP 注释, 针对关联区域内的基因, 发现亲本间存在非同义突变的 SNP 共 7 个, 后代混池间 5 个 SNP 存在非同义突变, 有 2 个是重复检测到的, 这些 SNP 很可能是与大豆酸性磷酸酶活性直接相关的连锁分子标记。应用 BLAST^[23]软件对候选区间内的编码基因进行多个数据库(NR^[25]、Swiss-Prot、GO^[26]、KEGG^[27]、COG^[28])的深度注释。通过详细的注释, 快速筛选候选基因。候选区域内共注释到 78 个基因, 其中在亲本间存在非



彩色的点代表 Δ SNP 指数值, 黑色的线为拟合后的 Δ SNP 指数值。B1-B2 bulk 是低活性混池和高活性混池的 Δ SNP 指数值的分布图, 其中红色的线代表置信度为 0.99 的阈值线, 蓝色的线代表置信度为 0.95 的阈值线, 绿色的线代表置信度为 0.90 的阈值线

The color points represent Δ SNP-index values, and the black lines are fitted Δ SNP-index. B1-B2 bulk is the distribution map of Δ SNP-index value with low-activity mixing pool and high-activity mixing pool, where the red line represents the threshold line of confidence 0.99, the blue line represents the threshold line of confidence 0.95, and the green line represents the threshold line of confidence 0.90

图 3 Δ SNP 指数值在染色体上的分布整合

Fig.3 The distribution and integration of Δ SNP-index association values on chromosome

同义突变基因注释到 2 个,与第 17 号染色体关联区域内的 2 个非同义突变基因一致。注释结果见表 5。

2.4 酸性磷酸酶活性关联候选基因功能标记开发

2.4.1 酸性磷酸酶活性关联候选区等位基因的克隆及序列差异分析

亲本间存在非同义突变注释基因 *Glyma.17G165600.1* 和 *Glyma.17G166200.1*, 分

别位于这两个基因上的亲本间非同义突变的 7 个被注释 SNP,与后代混池间非同义突变的 5 个被注释 SNP 比对,存在 2 个相同的 SNP(表 6)。说明亲本及后代混池关联的候选基因序列的 2 个相同 SNP 与酸性磷酸酶活性高低直接相关,并均在基因 *Glyma.17G166200.1* 上。

表 4 候选区域内 SNP 统计

Table 4 Statistics of SNP in candidate regions

SNP 所在区域类型 Type of SNP regions	亲本间在关联区域内对应的 SNP 数 Number of SNPs in the associated area between the parents	后代混池间在关联区域内对应的 SNP 数 Number of SNPs in the associated area between the descendant pools
基因前体 5' 端 Utr_5_Prime	1	1
基因上游区域(5 K 以内)Upstream	24	22
非同义编码突变 Non_Synonymous_Coding	7	5
内含子 Intron	5	4
基因内(无转录本信息)Intergenic	99	54
基因下游区域(5 K 以内)Downstream	29	17
全部 All	165	103

表 5 候选区域内基因功能注释结果统计

Table 5 Statistical results of gene function annotation in candidate regions

功能注释数据库 Annotated_databases	有注释信息的候选区域基因数 Number of candidate region genes with annotation information	候选区域内非同义突变的基因数 Number of genes with non synonymous mutations in candidate regions
非冗余蛋白数据库 NR	76	2
非冗余 DNA 序列数据库 NT	74	1
蛋白结构域注释的分类系统数据库 Pfam	60	1
高质量的蛋白数据库 SwissProt	52	2
基因本体论注释数据库 GO	46	2
代谢通路注释数据库 KEGG	32	0
蛋白相邻类的聚簇数据库 COG	24	0
真核生物蛋白相邻类的聚簇数据库 KOG	30	1
合计 Total	78	2

表 6 亲本及后代混池关联的非同义突变基因 SNP 及在染色体上的位置

Table 6 The non synonymous mutation gene SNP and its position on chromosome in the mixed pool of parents and offspring

非同义突变基因 Non synonymous mutation gene	染色体上位置 Position on Chromosome	突变位点 Mutation site			
		母本 Maternal	父本 Paternal	低活性后代 Low activity progeny	高活性后代 Highly active progeny
<i>Glyma.17G165600.1</i>	15083449	G	C	—	—
	15166089	T	C	—	—
<i>Glyma.17G166200.1</i>	15166091	G	A	—	—
	15166100	A	G	—	—
	15166102	C	G	—	—
	15166211	T	C	T	C
	15166218	C	T	C	T

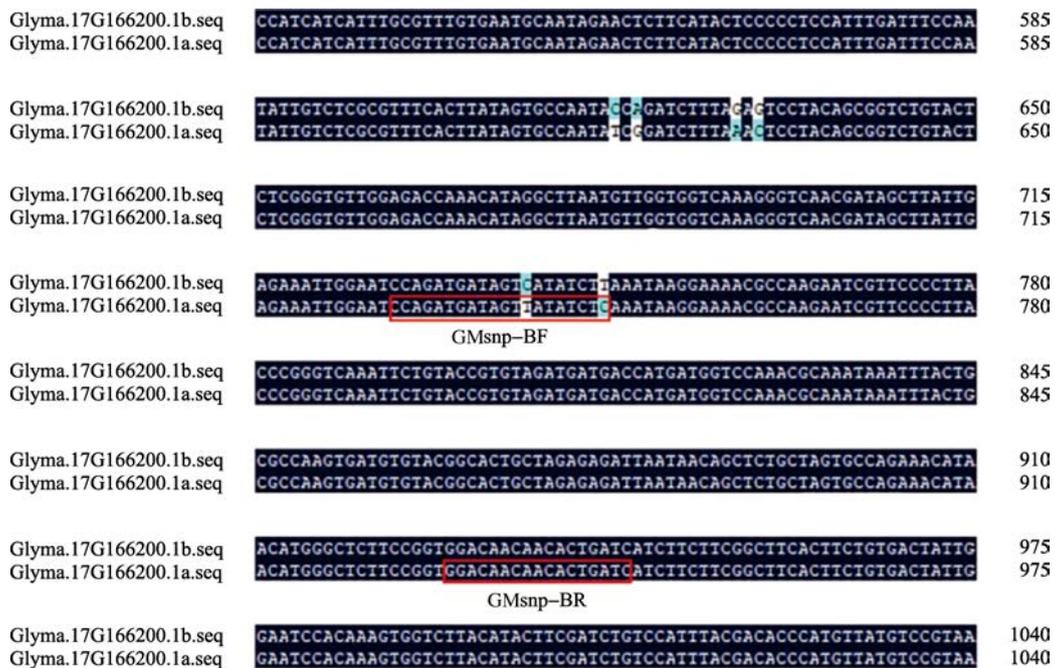
故通过已知目标基因 *Glyma.17G166200.1* 序列设计引物(上游引物 5' -TTGGACTCTATTG-CCTTGCC-3', 下游引物 5' -AGGCTTCAATTC-CCGATTT-3'), 分别对冀 NF58 和冀豆 12(亲本)材料进行扩增, 获得产物 3557bp。经切胶回收、连接反应及转化培养, 菌液送测序, 测序结果与目标基因序列比对, *Glyma.17G166200.1* 的基因片段为 3103bp, 保证 *Glyma.17G166200.1* 基因克隆的正确完整。

基因 *Glyma.17G166200.1* 的 DNA 序列长度为 3103bp, 包括 212 bp 的 5' 非翻译区 (Untranslated region, UTR) 和 359bp 的 3' 非翻译区, 包含一个 2532bp 的编码序列 (CDS, coding sequence)、1 个外显子。通过对冀 NF58 和冀豆 12 的等位序列进行比对, 其中 *Glyma.17G166200.1* 在冀 NF58 中的等位基因序列称为 *Glyma.17G166200.1b*, 在冀豆 12 中的等位基因序列称为 *Glyma.17G166200.1a*。由图 4 可知, 冀 NF58 在 17 号染色体第 15166089bp、15166100bp、15166102bp、15166211bp 和 15166218bp 处对应的碱基为 C、A、G、G、C 和 T, 冀豆 12 在 17 号染色体第 15166089bp、15166100bp、15166102bp、15166211bp 和 15166218bp 处对应的碱基替换为 T、G、A、C、T 和 C, 且突变位点均在其

编码序列上, 此结果与表 6 亲本关联的非同义突变基因 SNP 及在染色体上的位置一致, 进一步确定以亲本及后代混池关联的候选基因序列的 2 个相同 SNP 作为研究重点。

2.4.2 酸性磷酸酶活性关联基因 *Glyma.17G166200.1*

功能标记的开发与验证 观察 PCR 电泳结果, 筛选条带最丰富的退火温度为该引物的最优退火温度, 观察 PCR 电泳结果筛选最优引物组合。若有扩增产物, 大豆品种根尖酸性磷酸酶活性低; 若无扩增产物, 大豆品种根尖酸性磷酸酶活性高。经过 PCR 扩增, 3 对引物都能够扩增出条带, 根据条带的明亮程度确定每个标记的最优引物, GMsnp-B 能扩增出特异性明亮清晰条带 217bp(冀豆 12), 上下游引物参照图 4 中红色方框内标记。将最优产物开发为大豆酸性磷酸酶活性的功能标记 GMsnp-B, 其中各选取 8 个大豆酸性磷酸酶高活性和低活性品种检测, 其中低活性品种为铁丰 3 号、中作 04-717、晋遗 55、邯 6147、冀豆 12、郑 92116、09B7、东豆 99-11, 高活性品种为中黄 15、五星 2 号、山东山宁 11 号、中作 J5044、齐交 03-9、石 H570、豫豆 29、冀 NF58, 结果见图 5。在 169 个大豆自然品种材料中进行检测, 由功能标记 GMsnp-B 可扩增出 217bp 片段, 属于



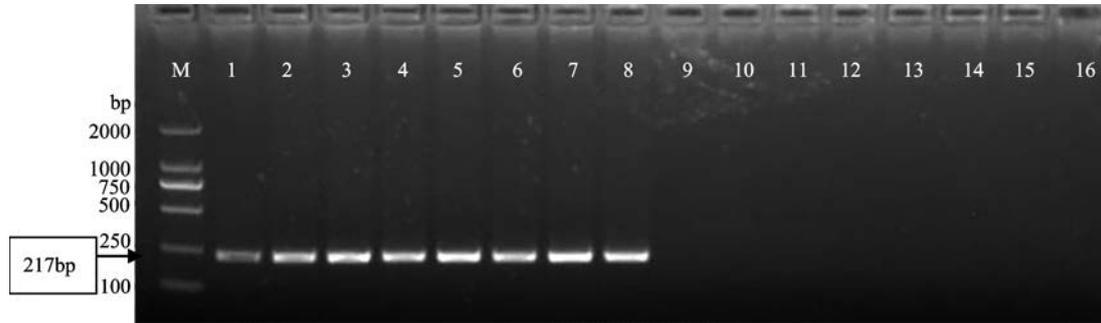
出现在 *Glyma.17G166200.1a* 和 *Glyma.17G166200.1b* 序列中蓝白标记为非同义突变 SNP, 用来开发

Glyma.17G166200.1 功能标记的上游引物与下游引物的反向互补序列加红色方框标出

Blue and white markers appeared in sequences of *Glyma.17G166200.1a* and *Glyma.17G166200.1b* as non-synonymous mutation SNP, the sequence of the forward primer and complementary sequence of the reverse primer for being used to develop *Glyma.17G166200.1* functional marker were redly boxed

图 4 *Glyma.17G166200.1a* (冀豆 12) 与 *Glyma.17G166200.1b* (冀 NF58) 序列比对 (部分)

Fig.4 Alignment of partial DNA sequences of *Glyma.17G166200.1a* (JD12) and *Glyma.17G166200.1b* (NF58)



M: DNA 分子量标准 DL2000; 1: 铁丰 3 号; 2: 中作 04-717; 3: 晋遗 55; 4: 邯 6147; 5: 冀豆 12; 6: 郑 92116; 7: 09B7; 8: 东豆 99-11; 9: 中黄 15; 10: 五星 2 号; 11: 山东山宁 11 号; 12: 中作 J5044; 13: 齐交 03-9; 14: 石 H570; 15: 豫豆 29; 16: 冀 NF58

M: DNA ladder DL2000, 1: TieFengNo.3, 2: ZhongZuo04-717, 3: JinYi55, 4: Han6147, 5: JiDou12, 6: Zheng92116, 7: 09B7, 8: DongDou99-11, 9: ZhongHuang15, 10: WuXingNo.2, 11: ShanDongShanNingNo.11, 12: ZhongZuoJ5044, 13: QiJiao03-9, 14: ShiH570, 15: YuDou29, 16: JiNF58

图 5 功能标记 GMsnp-B 对 8 个低 AP 活性材料和 8 个高 AP 活性材料的多态性检测

Fig.5 Polymorphic test of PCR fragments amplified by GMsnp-B in 8 cultivars with lower AP activity and 8 cultivars with higher AP activity

大豆酸性磷酸酶活性低的品种;无扩增产物的,属于大豆酸性磷酸酶活性高的品种。与大豆酸性磷酸酶活性检测结果统计分析(图 6),其符合率达到 82.8%。

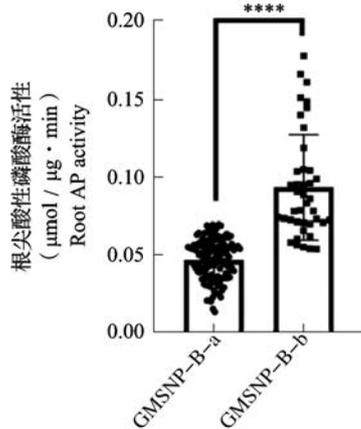


图 6 169 份大豆栽培品种以标记 GMsnp-B 检测结果的统计分析

Fig.6 Association between PCR profile and root AP activity in 169 soybean cultivars tested with GMsnp-B

3 讨论

SLFA-BSA 可以高通量地开发与性状相关的 SNP 标记,这种类型的标记在整个基因组上数量众多、分布均匀,与性状的遗传距离非常近,将其定位到参考基因组上,结合全基因组序列中对应的未知功能的基因序列设计引物,可以有效、准确、快速地克隆目标基因序列。Pan 等^[10]从玉米自交系 501 的愈伤组织再生的玉米植株后代中分离得到 bsd-pg (淡绿色缺陷型维管束鞘)突变基因,利用 SLAF-

Seq 技术,父母本之间 1771 个纯合等位基因分析出 5783 个多态性 SLAF 标签,在 1 号染色体上大小为 3945 Mb 的区域获得 3 个性状相关的候选区域,包括 35 个候选基因。Qin 等^[11]利用 SSR 和 SLAF 序列结合 BSA 的方法,将 HvSGRA 定位在大麦两个相邻的 FP 之间,在 whs18 上 fructokinase-1-like 基因突变可能导致大麦阶段性白化转绿, HvSGRA 的分层级图位克隆和利用基因 / 性状为未来分子育种的可视化研究奠定了基础。Han 等^[12]通过 QTL 精细定位并结合 BSA 和深度测序 SLAF-sequence 研究获得干茎高粱地方品种茎部水分控制位点 qSW6,定位于 6 号染色体的 SSR 标记 ch6-2 和 gpsb069 区间含有 38 个候选基因。Wen 等^[13]在黄瓜叶斑病 (TLS) 研究中,在黄瓜 6 号染色体 2.5 Mb 区域精细定位一个抗叶斑病基因 *cca-3* 的候选基因 *Csa6M375730*,并发现在候选基因序列内单核苷酸多态性 (SNP) 可能导致氨基酸的变化。近年来该项技术已被用于高通量分子标记开发、高密度遗传图谱构建、关联分析及植物目标性状的基因定位等领域。

本研究以重组自交系的 2 个 35+35 极端混池为材料,利用 SALF-BSA 简化基因组测序技术,共发现与控制酸性磷酸酶活性的 2 个关联区域,分别分布在第 3 号染色体的 20138271~20268154 且跨度 0.13 Mb 的区域和第 17 号染色体的 14368648~15526449 且跨度 1.158 Mb 的区域。其中第 3 号染色体的关联区域内有 4 个基因,第 17 号染色体的关联区域内有 75 个基因,2 个非同义突变基因都处于第 17 号染色体关联区域内。

Glyma.17G165600.1 和 *Glyma.17G166200.1*, 分别位于这 2 个基因上的亲本间非同义突变的 7 个被注释 SNP, 与后代混池间非同义突变的 5 个被注释 SNP 比对, 存在 2 个相同的 SNP, 说明亲本及后代混池关联的候选基因序列的 2 个相同 SNP 与酸性磷酸酶活性高低直接相关。鉴于根尖的酸性磷酸酶活性与相应的磷效率存在显著正相关, 可进一步推断出亲本及后代混池关联的非同义突变的 SNP 及其候选基因序列与大豆磷效率高低直接相关。King 等^[29]通过对大豆籽粒中磷累积的候选基因进行数量性状位点 (QTL) 分析, 在第 7 和第 17 染色体上发现了 2 个附加的 QTL, 其中 7 号和 17 号染色体上既有磷酸转运基因, 也有运输基因, 这些区域有更多的基因参与磷酸盐代谢和运输。对 17 号染色体热点区域内基因, 与本研究中的该染色体关联区域的基因对照, 经过 GO、COG 和 KEGG 数据库基因功能预测, 发现这些基因可能参与磷酸盐代谢和运输, 这一结果说明在热点区域可能存在多个控制磷利用率 (酸性磷酸酶活性) 的候选基因。在这些基因中, 本研究克隆了作用最显著的 *Glyma.17G166200.1* 基因, 经 NCBI 比对发现, 该基因表达蛋白酪氨酸激酶, 含有内质网内葡萄糖结合蛋白的区域, 是大豆蛋白激酶家族蛋白, 对蛋白质中的氨基酸残基的磷酸化反应, 是在蛋白质上引入磷酸基团的过程。Zhang 等^[30]发现第 17 号染色体上潜在影响大豆酸性磷酸酶活性和磷效率的连锁位点 q17, LOD 值 4.24, 标记区间为 1984881~2031528, 与本研究中的酸性磷酸酶活性关联区域距离 1.35 Mb, 可进一步推断出本研究中获得的影响酸性磷酸酶活性关联区域很可能是大豆酸性磷酸酶活性和磷效率的连锁位点。

功能标记是利用与表型相关的功能基因序列中功能性单核苷酸多态性位点来开发的新型分子标记。由于是来自基因内的功能性基序, 功能标记不需要进一步验证就可以在不同的遗传背景下确定目标等位基因的有无。与其他遗传标记相比, 功能标记在应用上占有更大的优势。它能够准确地跟踪、定位群体中的目标基因, 高效率地筛选自然群体及育种群体中的有利基因, 进行分子辅助选择育种^[31]。本试验利用 169 份栽培材料用于对 *Glyma.17G166200.1* 的功能标记 GMsnp-B 进行分析, 优异等位基因 *Glyma.17G166200.1b* 具有较高的根尖酸性磷酸酶活性, 尤其在品种中符合率高达 82.8%, 也验证了通过 SALF-BSA 技术筛选酸性磷酸酶活性功能基因的可行性。根据已获得与酸性磷

酸酶活性以及磷效率紧密关联的 SNP 和候选基因, 可进一步对相关基因做生物功能研究, 寻找基因表达的作用机理, 通过品种和群体检验其功能有效性。

参考文献

- [1] 李喜焕, 常文锁, 张彩英. 中国大豆磷素营养及磷高效品种筛选最新进展. 大豆科学, 2011, 30(2): 322-327
Li X H, Chang W S, Zhang C Y. Advances of soybean (*Glycine max* L.) phosphorus nutrition and high P-efficient germplasm screening. Soybean Science, 2011, 30(2): 322-327
- [2] 黄宇, 张海伟, 徐芳森. 植物酸性磷酸酶的研究进展. 华中农业大学学报, 2008, 27(1): 148-154
Huang Y, Zhang H W, Xu F S. Research progress on plant acid phosphatase. Journal of Huazhong Agricultural University, 2008, 27(1): 148-154
- [3] 樊明寿, 徐冰, 王艳. 缺磷条件下玉米根系酸性磷酸酶活性的变化. 中国农业科学导报, 2001, 3(3): 33-36
Fan M S, Xu B, Wang Y. Acid phosphatase activities of intact roots and ground root tissues of maize grown in high P or low P nutrient solution. Review of China Agricultural Science and Technology, 2001, 3(3): 33-36
- [4] Song J Y, Shen M K. Induction of maize acid phosphatase activities under phosphorus starvation. Plant and Soil, 2001, 237: 109-115
- [5] 魏志强, 史衍玺, 孔凡美. 缺磷胁迫对花生磷酸酶活性的影响. 中国油料作物学报, 2002, 24(3): 44-46
Wei Z Q, Shi Y X, Kong F M. The effect of phosphorus deficiency stress on acid phosphatase in peanut. Chinese Journal of Oil Crop Sciences, 2002, 24(3): 44-46
- [6] 孙海国, 张福锁. 缺磷条件下的小麦根系酸性磷酸酶活性研究. 应用生态学报, 2002, 13(3): 379-381
Sun H G, Zhang F S. Effect of phosphorus deficiency on activity of acid phosphatase exuded by wheat roots. Chinese Journal of Applied Ecology, 2002, 13(3): 379-381
- [7] 周建朝, 韩晓日, 奚红光. 磷营养水平对不同基因型甜菜根磷酸酶活性的效应. 植物营养与肥料学报, 2006, 12(2): 233-239
Zhou J C, Han X R, Xi H G. Effects of phosphorus rate on root phosphatase activities of different sugar beet genotypes. Plant Nutrition and Fertilizer Science, 2006, 12(2): 233-239
- [8] Liang C Y, Tian J, Lam H M, Lim B L, Yan X, Liao H. Biochemical and molecular characterization of *PvPAP3*, a novel purple acid phosphatase isolated from common bean enhancing extracellular ATP utilization. Plant Physiology, 2010, 152: 854-865
- [9] 刘渊, 李喜焕, 孙星, 张彩英. 磷胁迫下大豆酸性磷酸酶活性变化及磷效率基因型差异分析. 植物遗传资源学报, 2012, 13(4): 521-528
Liu Y, Li X H, Sun X, Zhang C Y. Variation of acid phosphatase activity and analysis of genotypic difference in P efficiency of soybean under phosphorus stress. Journal of Plant Genetic Resources, 2012, 13(4): 521-528
- [10] Pan Y, Chen X Q, Xie H, Deng L, Li X L, Zhang X D, Han L X, Yang F P, Xue J, Zhang L Q. A maize bundle sheath defective mutation mapped on chromosome 1 between SSR markers umc1395 and umc1603. Journal of Integrative Agriculture,

- 2015, 14(10): 1949–1957
- [11] Qin D D , Dong J , Xu F C , Guo G G , Ge S T , Xu Q , Xu Y X , Li M F . Characterization and fine mapping of a novel barley stage green-revertible albino gene (*HvSGRA*) by bulked segregant analysis based on SSR assay and specific length amplified fragment sequencing. *BMC Genomics*, 2015, 16: 838
- [12] Han Y C , Lv P , Hou S H , Li S Y , Ji G S , Ma X , Du R H , Liu G Q . Combining next generation sequencing with bulked segregant analysis to fine map a stem moisture locus in sorghum (*Sorghum bicolor* L. Moench). *PLoS One*, 2015, 10(5): e0127065
- [13] Wen C L , Mao A J , Dong C J , Liu H Y , Yu S C , Guo Y D , Weng Y Q , Xu Y . Fine genetic mapping of target leaf spot resistance gene *cca* in cucumber, *Cucumis sativus* L. *Theoretical and Applied Genetics*, 2015, 128(12): 2495–2506
- [14] Sonah H , O'Donoghue L , Cober E , Rajcan I , Belzile F . Identification of loci governing eight agronomic traits using a GBS–GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnology*, 2015, 13, 211–221
- [15] Hwang E Y , Song Q , Jia G , Specht J E , Hyten D L , Costa J , Cregan P B . A genome-wide association study of seed protein and oil content in soybean. *BMC Genomics*, 2014, 15: 1
- [16] Mamidi S , Lee R K , Goos J R , McClean P E . Genome-wide association studies identifies seven major regions responsible for iron deficiency chlorosis in soybean (*Glycine max*). *PLoS One*, 2014, 9: e107469
- [17] Ning L H , Zhen K G , Du W K , Guo S W , Wang Q , Zhang G Z , Cheng H , Yu D Y . Association analysis for detecting significant single nucleotide polymorphisms for phosphorus-deficiency tolerance at the seedling stage in soybean. *Breeding Science*, 2016, 66: 191–203
- [18] Zhang D , Song H , Cheng H , Hao D , Wang H , Kan G , Jin H , Yu D . The acid phosphatase-encoding gene *GmACP1* contributes to soybean tolerance to low phosphorus stress. *PLoS Genetics*, 2014, 10: e1004061
- [19] Zhang D , Zhang H Y , Chu S S , Li H Y , Chi Y J , Triebwasser-Freese D , Lv H Y , Yu D Y . Integrating QTL mapping and transcriptomics identifies candidate genes underlying QTLs associated with soybean tolerance to low-phosphorus stress. *Plant Molecular Biology*, 2017, 93(1–2): 137–150
- [20] Juwattanasomran R , Somta P , Kaga A , Chankaew S , Shimizu T , Sorajjapinun W , Srinives P . Identification of a new fragrance allele in soybean and development of its functional marker. *Molecular Breeding*, 2012, 29(1): 13–21
- [21] Bachleda N , Pham A , Li Z L . Identifying *FATB1a* deletion that causes reduced palmitic acid content in soybean N87-2122-4 to develop a functional marker for marker-assisted selection. *Molecular Breeding*, 2016, 36: 45
- [22] Ciereszko I , Janonis A , Kociakowska M . Growth and metabolism of cucumber in phosphate-deficient conditions. *Plant Nutrition*, 2002, 25: 1115–1127
- [23] Hill J T , Demarest B L , Bisgrove B W , Gorski B , Su Y C , Yost H J . MMAPPR: Mutation mapping analysis pipeline for pooled RNA-seq. *Genome Research*, 2013, 23: 687–697
- [24] Takagi H , Abe A , Yoshida K , Kosugi S , Natsume S , Mitsuoka C , Uemura A , Utsushi H , Tamiru M , Takuno S , Innan H , Cano L M , Kamoun S , Terauchi R . QTL-seq: Rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant Journal*, 2013, 74: 174–183
- [25] Li H , Handsaker B , Wysoker A , Fennell T , Ruan J , Homer N , Marth G , Abecasis G , Durbin R . The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, 2009, 25: 2078–2079
- [26] Fekih R , Takagi H , Tamiru M , Abe A , Natsume S , Yaegashi H , Sharma S , Sharma S , Kanzaki H , Matsumura H , Saitoh H , Mitsuoka C , Utsushi H , Uemura A , Kanzaki E , Kosugi S , Yoshida K , Cano L , Kamoun S , Terauchi R . MutMap+: Genetic mapping and mutant identification without crossing in rice. *PLoS One*, 2013, 8(7): e68529
- [27] Altschul S F , Madden T L , Schäffer A A , Zhang J , Zhang Z , Miller W , Lipman D J . Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 1997, 25(17): 3389–3402
- [28] Deng Y Y , Li J Q , Wu S F , Zhu Y P , Chen Y W , He F C . Integrated nr database in protein annotation system and its localization. *Computer Engineering*, 2006, 32(5): 71–74
- [29] King K E , Lauter N , Lin S F , Scott M P , Shoemaker R C . Evaluation and QTL mapping of phosphorus concentration in soybean seed. *Euphytica*, 2013, 189(2): 261–269
- [30] Zhang D , Li H Y , Wang J S , Zhang H Y , Chu S S , Lv H Y , Yu D Y . High-density genetic mapping identifies new major loci for tolerance to low-phosphorus stress in soybean. *Frontiers in Plant Science*, 2016, 7: 372–383
- [31] 王昊龙, 韩俊杰, 李森森, 南富波, 李卫华 . 功能标记的开发及在禾谷类作物中的应用. *核农学报*, 2014, 28(11): 1963–1971
- Wang H L , Han J J , Li M M , Nan F B , Li W H . Development and application of functional markers in cereal crops. *Journal of Nuclear Agricultural Sciences*, 2014, 28(11): 1963–1971