



全基因组关联分析在大豆中的研究进展

李廷雨, 黎永力, 甘卓然, 石文茜, 董利东, 刘宝辉

(广州大学 生命科学院, 广东 广州 510006)

摘要:全基因组关联分析(GWAS)是目前发现复杂农艺性状相关遗传基础最有力和最有效的研究方法。随着遗传学的发展,基因分型、统计学方法等许多关键性技术的进步,以连锁不平衡为基础的全基因组关联分析应运而生。近年来 GWAS 在大豆复杂性状研究和作物育种中得到了广泛的应用。本研究在简要介绍 GWAS 的原理、流程的基础上,总结其在大豆重要农艺性状研究中取得的最新进展,最后讨论 GWAS 存在的问题和未来发展趋势,以期为进一步利用 GWAS 进行大豆遗传育种研究提供理论依据。

关键词:大豆;全基因组关联分析;连锁不平衡

Research Progress of Genome-Wide Association Studies in Soybean

LI Ting-yu, LI Yong-li, GAN Zhuo-ran, SHI Wen-qian, DONG Li-dong, LIU Bao-hui

(School of Life Sciences, Guangzhou University, Guangzhou 510006, China)

Abstract: Genome wide association studies (GWAS) is the most powerful and effective method to find the genetic basis of complex agronomic traits. With the development of genetics, gene typing and statistical methods, genome-wide association analysis based on linkage disequilibrium came into being. In recent years, GWAS has been widely used in the research of soybean complex characters and crop breeding. Based on the brief introduction of GWAS's principle and process, this paper summarizes its latest progress in the research of important agronomic characters of soybean, and finally discusses the existing problems and future development trend of GWAS, in order to provide theoretical basis for further research of soybean genetic breeding with GWAS.

Keywords: Soybean; Genome-wide association studies(GWAS); Linkage disequilibrium

大豆(*Glycine max*)是一类重要的豆类作物,其种子中含有丰富的蛋白质和油脂,占世界蛋白质消费的 68% 和植物油消费的 29%^[1]。大豆是人类食物、食用油和家畜饲料的重要来源,在生物柴油生产、工业和医药方面也有着广泛的应用^[2]。然而,当前中国大豆的产量仅占世界总产量的较小部分,远远不能满足国内消费者对大豆的需求。因此,揭示大豆产量、品质相关性状遗传基础并培育优良大豆品种是中国大豆育种家的研究目标,而全基因组关联分析(genome-wide association studies, GWAS)技术为实现这一目标创造了重要条件。

1996 年, Risch 等^[3]首次提出了全基因组关联分析的概念。2001 年, Hansen 等^[4]最早报道了 GWAS 在野生甜菜生长习性方面的研究。随后, Aranzana 等^[5]通过对拟南芥开花时间和病原菌抗性的研究,验证了全基因组关联分析在拟南芥中的可行性,这是在植物中第一次完全意义的全基因组关联分析。随着测序技术和统计模型的快速发展, GWAS 逐渐成为解析植物复杂性状遗传基础的重要

工具。目前在玉米^[6]、高粱^[7]、大麦^[8]、水稻^[9]、小麦^[10]、油菜^[11]和拟南芥^[12]中,通过全基因组关联分析的方法已经找到了多个与重要农艺性状有关的基因。2010 年栽培大豆基因组序列的公布^[13],使得大豆基因组学得到了飞速的发展^[14],也促进了全基因组关联分析在大豆遗传分析中的应用,现已成功通过全基因组关联分析技术挖掘出许多控制大豆重要性状的遗传位点。本研究综述 GWAS 在大豆遗传改良中的研究原理、流程、最新进展及其在大豆遗传育种改良中的应用要求。

1 全基因组关联分析

1.1 全基因组关联分析的含义

关联分析(association analysis)又称关联作图(association mapping),通过将群体内的遗传标记与目的性状表型数据相结合进行统计分析,以此发掘出与目的性状相关联的遗传位点^[15]。从而在分子水平揭示作物表型多样性的遗传基础,为作物性状的改良提供理论基础,为分子育种提供新的途径。

收稿日期:2019-12-25

基金项目:国家自然科学基金(31771815)。

第一作者简介:李廷雨(1992-),男,硕士,主要从事光周期调控大豆开花分子机理研究。E-mail:1689763520@qq.com。

通讯作者:刘宝辉(1964-),男,博士,教授,博导,主要从事光周期调控大豆开花和重要农艺性状分子机理研究。E-mail:liubh@gzhu.edu.cn。

全基因组关联分析是在全基因组范围内进行整体研究,大大增加了分子标记的数量^[5]。作为一项开创性的研究方法,它不仅适合研究单基因控制的性状,还适用于复杂性状的研究。随着高通量测序成本的降低,GWAS 在农作物的研究上表现出了巨大的优势。

1.2 全基因组关联分析的理论基础

连锁不平衡(linkage disequilibrium, LD),指不同基因座位等位基因的关联频率高于随机关联的频率^[16-17],是全基因组关联分析的主要理论基础。目前,有很多统计方法能表示 LD,但通常主要采用 D' (standardized disequilibrium correlation) 和 r^2 (squared allele-frequency correlations) 两种统计量^[18]。 D' 和 r^2 的取值范围为 0 ~ 1,当 D' 和 r^2 均为 0 时,表示两个位点间没有连锁不平衡;当 D' 或 r^2 增大到 1 时说明两个位点处于完全的连锁不平衡状态。 D' 能够解释位点间的重组史,但小样本量对 D' 的影响较大,尤其是在群体中有较低的等位基因频率时。而 r^2 在统计学上有着更加严格的定义,受群体大小影响较小,因此一般选择 r^2 作为关联分析中的度量值^[19]。群体 LD 的水平受突变和重组等多种因素的综合影响,位点间连锁的紧密程度决定达到新的连锁平衡的时间^[20]。

1.3 全基因组关联分析的基本方法和步骤

GWAS 主要包括以下 5 步:(1)选择遗传材料:种质资源是进行 GWAS 分析的基础,研究的群体内要有足够大的样本量,尽可能地包括该物种全部遗传变异信息,这样才能达到较高的统计学效力,发现一些对性状影响较小的遗传变异。在研究过程中可以根据自身的情况适当的调整群体大小,群体中品种数目一般推荐要大于 300 个。(2)表型考察:目标性状的评估会对 GWAS 的结果产生很大的影响。表型性状不仅受基因型的控制,还受到周围环境的影响。因此,制定良好的试验计划^[21]、从时间、空间上增加表型的重复性显得非常重要^[22]。(3)基因型分型:随着测序技术的快速发展,研究者能够更方便的获得高通量、低成本的测序服务。单核苷酸多态性(single nucleotide polymorphism, SNP)标记因其遗传稳定性高、数量丰富等特点,在 GWAS 研究中广泛应用。(4)群体结构和亲缘关系分析:种群结构会导致一些等位基因频率在不同的亚群之间明显不同,从而导致结果的假阳性^[23]。因此在进行 GWAS 分析之前要对群体结构和亲缘关系进行有效的评估。通过构建群体系统进化树、主成分分析、Structure 分析,研究样本间的群体结构和亲缘关系,这 3 种分析结果可以相互验证。(5)关联分

析:随着 GWAS 研究的不断深入,关联分析的算法模型和分析软件不断得到优化,目前有很多可以用于全基因组关联分析的软件。常用的分析软件有 PLINK^[24] 和 TASSEL^[25] 等。

2 利用全基因组关联分析研究大豆农艺性状的进展

2.1 GWAS 在大豆产量相关性状研究中的应用

2.1.1 株型性状 株高、主茎节数和分枝数等是影响大豆产量的主要株型性状。良好的大豆株型对提高大豆产量能起到重要的作用。近些年利用 GWAS 分析的方法鉴定出了一些与大豆株型相关的重要 SNP 位点。例如,Lyu 等^[26] 对 192 份大豆材料进行基因分型共获得了 1 536 个单核苷酸多态性标记,并对株高性状进行了 GWAS 分析,结果显示有 3 个 SNP 与生长后期植株高度相关,2 个 SNP 与生长前期植株高度相关。Jing 等^[27] 以 185 份优良大豆种质为材料,利用简化基因组测序技术共获得 33 149 个 SNP,随后通过 GWAS 检测到 14 个与株高相关的 SNP 信号,在每个 SNP 位点附近的基因组区域发现了许多功能基因,根据转录分析预测了 6 个可能与株高相关的新基因。Chang 等^[28] 以 368 个大豆种质资源为材料,基于 2 种算法模型对株高和主茎节数 2 个性状进行 GWAS 分析,检测到 45 和 43 个位点分别与株高和主茎节数相关。Shim 等^[29] 采用全基因组关联分析和连锁分析相结合的方法研究大豆分枝数性状,经分析发现有 5 个数量性状核苷酸(QTN)与大豆核心种质的分枝数相关,进一步分析推测 *Glyma. 06G210600* 可能作为调控大豆分枝数的候选基因。以上的研究表明利用 GWAS 分析能够有效的鉴定大豆株高、节数和分枝的关键遗传位点。对大豆株高和分枝遗传位点的鉴定,将有助于大豆理想株型性状分子调控网络的解析,为大豆分子育种提供理论基础。

2.1.2 生育期性状 生育期性状是影响大豆产量的重要性状,深入了解这些性状的遗传基础,可以为大豆品种的种植面积划分、合理布局和分子育种提供理论依据。尽管利用经典遗传学和反向遗传学的方法已经鉴定出了一些控制大豆生育期的关键基因,但是仍然有一些重要的控制大豆生育期基因未被鉴定。利用 GWAS 分析的方法能够简单,高效的鉴定出这些控制大豆生育期的重要基因。例如,Li 等^[30] 对 1 938 个不同地方品种和 97 个野生大豆品种进行了基因分型,随后以 99 085 个优质 SNP 为基础,通过 GWAS 鉴定出 17 个与开花时间相关的 SNP。Pan 等^[31] 对 427 份大豆自交系群体进行重

测序分析,共获得了576 874个 SNP,通过 GWAS 和一些后续分析确定了52个和开花期相关的候选基因。这些候选基因中包括一些已知的基因,如 *E1*、*E2*、*E9* 和 *J*。Zhang 等^[32]对309份大豆进行了基因分型,并在多种环境下对生育期表型做了统计,通过 GWAS 分别检测到27和6个与开花期和成熟期相关的位点。Liu 等^[33]利用4 032个单核苷酸多态性(SNP)标记,对146个主要来自东北地区的品种进行了全基因组关联分析,发掘与开花天数、开花期至成熟期天数和成熟期天数相关的数量性状位点,基于 GWAS 检测出分布在11条大豆染色体上的19个显著相关的 SNP,其中7个位于或接近先前报道的控制生育期性状的 QTL 或基因。这些研究所鉴定的高分辨率 QTL 将丰富对生育期性状的基因组学理解,并可作为遗传标记应用于大豆育种中。

2.1.3 荚粒性状 百粒重、单株荚数等荚粒性状是大豆产量的重要组成因子,通过挖掘优异遗传位点可以进一步提高大豆的产量。Zhao 等^[34]以东北地区185份具有代表性的大豆为样本,通过 GWAS 检测到34个与百粒重相关的位点,其中19个位点是新发现的,另外15个位点位于已知的基因附近。Contreras 等^[35]对巴西169个大豆品种进行了基因分型并调查了百粒重和种子产量性状,通过 GWAS 分别鉴定出11和17个与百粒重和种子产量显著相关的位点。Wen 等^[36]对2个大豆群体进行了基因分型共获得52 041个 SNP,通过 GWAS 检测到1个与单株荚数相关的 SNP (*Gm01-55794390*),位于已知的与单株荚数相关的 SNP 附近。Hu 等^[37]以211份来自不同环境的大豆为材料,通过全基因组关联分析,检测并鉴定出163个与荚果炸裂显著相关的 SNP。此外,实时定量 PCR (qRT-PCR) 结果和多态性分析表明 *Glyma09g06290* 基因可能与荚果炸裂有关。尽管,利用 GWAS 分析发现了大量与控制大豆荚粒性状相关的位点,为大豆荚粒性状分子机制的解析和分子育种的应用提供了理论依据,但是,大多数的位点缺乏候选基因的鉴定和验证,所以后续研究中应该着重候选基因的鉴定和功能研究。

2.2 GWAS 在大豆品质相关性状研究中的应用

2.2.1 蛋白质含量 大豆以其优质蛋白质受到广泛欢迎,蛋白质含量是由众多基因相互作用决定的数量遗传性状。为了更好地了解大豆蛋白质分子基础,Zhang 等^[38]以388份重组自交系(3 454个 SNP)和200份大豆自然群体(94 462个 SNP 和 indel)为材料,同时运用连锁分析和 GWAS 方法定位到和大豆蛋白含量相关的遗传位点,在2个 QTL 区

间(*qPro15-1*, *qPro20-1*)内,共确定25个可能参与蛋白质合成的候选基因。Zhang 等^[39]为了鉴定控制蛋白质含量的数量性状核苷酸(QTN),利用基因芯片对144份重组自交系材料进行基因分型并获得了包含109 676个 SNP 的基因型数据,用 GWAS 方法共鉴定出129个与蛋白质含量相关的显著位点,最后确定了8个可能参与大豆种子蛋白质合成和代谢的潜在候选基因。最后以365份蛋白质含量变化范围为37.51%~50.46%的具有代表性的大豆材料为研究对象,采用新的限制性两阶段多位点全基因组关联研究方法发掘与种子蛋白质含量相关的 QTL,分析鉴定出89个 QTL,在这些 QTL 附近共有255个基因^[40]。Li 等^[41]对185份大豆种质资源进行重测序,共获得12 072个优质 SNP,然后通过 GWAS 分析检测到3个 SNP 与种子蛋白质含量显著相关。这些结果将有助于发现大豆中控制蛋白质合成的多基因网络,并有助于高蛋白大豆品种的分选育种。

2.2.2 油脂含量 大豆油是世界上生产最广泛的植物油,其在大豆种子中的含量是育种中重要的品质性状。饱和脂肪酸(FA)是大豆油的重要组成部分,Zhao 等^[42]以185份不同大豆材料为研究对象,利用 GWAS 技术进行 FA 相关基因挖掘,通过 GWAS 检测到了65个与脂肪酸含量相关的 SNPs,其中49个基因被鉴定为与饱和脂肪酸相关的潜在候选基因。Zhang 等^[43]评估了219个不同大豆品种在6个不同环境下的种子油脂含量,并使用 GWAS 分析了其潜在遗传机制,发现了1个在各个环境都能被稳定检出的与油脂含量显著相关的位点。单倍型和表达分析表明油酸蛋白编码基因(*GmOLEO1*)与种子油含量显著相关。Li 等^[44]采用限制性两阶段多位点全基因组关联分析方法,对中国江淮地区279份大豆种质资源的种子油脂含量遗传结构进行了分析,鉴定到23个与种子含油量相关的 QTL,在这些检测到的 QTL 附近发现55个与油脂含量相关的候选基因。Hwang 等^[45]采用全基因组关联分析对298份大豆种质资源的种子油脂含量进行了分析,结果检测到了位于13个不同基因组区域中的25个 SNP 与油脂含量显著相关。这些新发现的遗传位点可以显著改善大豆分子育种程序,所鉴定的候选基因有助于阐明大豆油脂生物合成的机制。

2.3 GWAS 在大豆抗逆相关性状研究中的应用

2.3.1 非生物胁迫 干旱、盐碱、矿物质缺乏等非生物胁迫会严重影响大豆产量,近年来科学家利用 GWAS 分析的方法挖掘到了大量的大豆抗逆相关的

位点。例如,Dhanapal 等^[46]在 4 个环境下种植 373 个不同的大豆品种,并用碳同位素比值(delta C-13) 作为水分利用效率的替代指标来评估表型,最后利用 GWAS 分析共鉴定出 39 个与 delta C-13 相关联的 SNP,其中有 15 个 SNP 位于同一个基因内。除此之外,Zhang 等^[47]利用 GWAS 也发现 8 号染色体上与耐盐性显著相关的 SNP,进一步研究发现 *GmC-DF1* 在大豆耐盐性中起负调控作用。缺铁性黄化病是石灰质土壤条件下大豆产区的产量限制问题之一,Mamidi 等^[48]利用 GWAS 分析发现了与缺铁性黄化病性状相关的 7 个 QTL,在这些 QTL 附近的与铁代谢相关的 12 个基因被确定为候选基因。低磷胁迫是制约植物生长和产量的主要因素,在植物中,磷效率是一个涉及多个基因的复杂数量性状,其机制尚不清楚。Zhang 等^[49]通过全基因组关联分析,鉴定了一个与磷效率相关的大豆基因,确定了其有利的单倍型并开发了有价值的功能标记。这些重要的分子标记的发现,将为大豆抗逆分子育种提供理论依据。

2.3.2 生物胁迫 病毒、害虫等引起的生物胁迫严重影响着大豆的产量,了解大豆对生物胁迫的应答机制将对大豆的育种工作提供巨大帮助。GWAS 分析已经在大豆白霉病、茎溃疡病抗性相关基因,大豆胞囊线虫抗性相关基因等大豆生物胁迫相关基因的挖掘中得到应用。为了解析大豆对于白霉病抗性的遗传结构,Wen 等^[50]结合田间和温室环境中的抗性数据,进行了全基因组相关分析,发现分别有 16 和 11 个位点在大田和温室环境下与抗性显著相关,其中 8 个位点定位于先前报道过的 QTL 区间内,有 1 个位点在 2 个环境中均与抗性显著相关,最后确定一系列不同功能的基因作为白霉病抗性的候选基因。大豆胞囊线虫(SCN)是大豆根中一种极难消除的寄生虫,其对产量造成的损失大于其它害虫,为了鉴定抗大豆 SCN 关键的基因,Zhang 等^[51]对 120 个大豆材料进行基因分型,获得了含有 60 109个 SNP 的基因型数据,通过 GWAS 分析,在 7 个不同的基因组区域鉴定出 13 个与 SCN 抗性显著相关的 SNP,其中 3 个 SNP 定位于先前报道的 QTL 区间内,此研究还鉴定了与 SCN 抗性相关的多个新基因座和已知基因座的精细染色体区域。南部茎溃疡病(SSC)是一种重要的大豆病害,曾造成严重损失,Santos 等^[52]对来自世界不同地区的 295 份材料的 SSC 感染反应进行了研究,通过 GWAS 分析在 14 号染色体上找到 19 个与南方茎腐病抗性相关的 SNP。异黄酮在植物防御中起着重要作用,异黄酮含量是一个由多个基因控制的复杂数量性状,其生

物合成的遗传机制目前尚不清楚,Chu 等^[53]通过全基因组关联分析鉴定出 28 个 SNP 与大豆中异黄酮浓度显著相关,进一步研究表明 *GmMYB29* 可能部分影响表型变异。以上的研究表明利用 GWAS 分析的方法能够简单高效的鉴定出一些与大豆生物胁迫相关的位点,但这些位点是否能够有效的应用到大豆分子育种中,还需要进一步利用分子遗传学和分子生物学的手段进行验证。

3 问题与展望

GWAS 在大豆遗传研究中取得了一些进展,但目前研究的范围局限在少数性状上^[54]。与传统的 QTL 作图相比,全基因组关联分析具有检测范围广^[55]、分辨率高^[56-57]、材料来源多^[58]等优势。但是 GWAS 分析手段也有其自身的局限性,关联作图仍然是对连锁作图和其它基因识别和验证方法的补充,并不能替代它们。基因与环境的互作效应会对 GWAS 分析结果造成影响,因此要制定严格的试验标准^[58]。随着研究规模的不断扩大,新的基因座与性状表型间的关联效应会比较微弱^[59]。种群结构是关联研究中的一个很大的影响因素,需要仔细选择种质资源和进行种群结构评估。此外,GWAS 只是对候选遗传位点的一种预测,后续还需要结合其它实验方法进行深入研究以探究其生物学功能。

GWAS 在大豆遗传改良中的应用潜力巨大,尽管 GWAS 当前还存在着一些不足,但随着研究的不断深入,其在实际应用中的局限性也会不断得以克服。GWAS 必将成为大豆遗传研究中一种流行和常规的方法,也必将在揭示大豆复杂性状的遗传基础和 大豆品种改良中发挥更大的作用。

参考文献

[1] Zhang Y H, Liu M F, He J B, et al. Marker-assisted breeding for transgressive seed protein content in soybean [*Glycine max* (L.) Merr.] [J]. Theoretical and Applied Genetics, 2015, 128 (6): 1061-1072.

[2] Juhi C, Patil G B, Humira S, et al. Expanding omics resources for improvement of soybean seed composition traits [J]. Frontiers in Plant Science, 2015, 6: 1021.

[3] Risch N, Merikangas K. The future of genetic studies of complex human diseases [J]. Science, 1996, 273 (5281): 1516-1517.

[4] Hansen M, Kraft T, Ganestam S, et al. Linkage disequilibrium mapping of the bolting gene in sea beet using AFLP markers [J]. Genetical Research, 2001, 77 (1): 61-66.

[5] Aranzana M J, Kim S, Zhao K Y, et al. Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes [J]. PLoS Genetics, 2005, 1 (5): e60.

- [6] Wang M, Yan J, Zhao J, et al. Genome-wide association study (GWAS) of resistance to head smut in maize[J]. *Plant Science*, 2012, 196: 125-131.
- [7] Morris G P, Ramu P, Deshpande S P, et al. Population genomic and genome-wide association studies of agroclimatic traits in sorghum[J]. *Proceedings of the National Academy of Sciences*, 2013, 110(2): 453-458.
- [8] Alqudah A M, Sharma R, Pasam R K, et al. Genetic dissection of photoperiod response based on GWAS of pre-anthesis phase duration in spring barley[J]. *PLoS One*, 2014, 9(11): 113-120.
- [9] Wu J H, Feng F J, Lian X M, et al. Genome-wide association study (GWAS) of mesocotyl elongation based on resequencing approach in rice[J]. *BMC Plant Biology*, 2015, 15(1): 218.
- [10] Zhou Y, Tang H, Cheng M P, et al. Genome-wide association study for pre-harvest sprouting resistance in a large germplasm collection of Chinese wheat landraces[J]. *Frontiers in Plant Science*, 2017, 8 (93): 401.
- [11] Hatzig S V, Frisch M, Breuer F, et al. Genome-wide association mapping unravels the genetic control of seed germination and vigor in *Brassica napus*[J]. *Frontiers in Plant Science*, 2015, 6(221): 221.
- [12] Si W, Saleh A, Ivaro C. I., et al. Combined use of genome-wide association data and correlation networks unravels key regulators of primary metabolism in *Arabidopsis thaliana*[J]. *PLoS Genetics*, 2016, 12(10): 1-36.
- [13] Schmutz J, Cannon S B, Chlueter J, et al. Genome sequence of the palaeopolyploid soybean[J]. *Nature*, 2010, 463 (7278): 178-183.
- [14] Song Q, Hyten D L, Jia G, et al. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean[J]. *PLoS One*, 2013, 8(1): e54985.
- [15] Hsiao C F, Chiu Y F, Chiang F T, et al. Genome-wide linkage analysis of lipids in nondiabetic Chinese and Japanese from the SAPHIRE family study[J]. *American Journal of Hypertension*, 2006, 19(12): 1270-1277.
- [16] Gaut B S, Long A D. The lowdown on linkage disequilibrium[J]. *The Plant Cell*, 2003, 15(7): 1502-1506.
- [17] 王荣焕, 王天宇, 黎裕. 关联分析在作物种质资源分子评价中的应用[J]. *植物遗传资源学报*, 2016, 8(3): 366-372. (Wang R H, Wang T Y, Li Y. Application of association analysis in molecular evaluation of crop germplasm resources[J]. *Journal of Plant Genetic Resources*, 2016, 8(3): 366-372.)
- [18] Gaut B S, Long A D. The lowdown on linkage disequilibrium[J]. *The Plant Cell*, 2003, 15(7): 1502-1506.
- [19] Zondervan K T, Cardon L R. The complex interplay among factors that influence allelic association[J]. *Nature Reviews Genetics*, 2004, 5(2): 89-100.
- [20] 谭贤杰, 吴子恺, 程伟东, 等. 关联分析及其在植物遗传学研究中的应用[J]. *植物学报*, 2011, 46(1): 108-118. (Tan X J, Wu Z F, Cheng D W, et al. Association analysis and its application in plant genetics[J]. *Chinese Bulletin of Botany*, 2011, 46(1): 108-118.)
- [21] Myles S, Peiffer J, Brown P J, et al. Association mapping: Critical considerations shift from genotyping to experimental design[J]. *Plant Cell*, 2009, 21(8): 2194.
- [22] Flint-Garcia S A, Thuillet A C, Yu J M, et al. Maize association population: A high-resolution platform for quantitative trait locus dissection[J]. *Plant Journal*, 2005, 44(6): 1054-1064.
- [23] Kriz A L, Larkins B A. Molecular genetic approaches to maize improvement[M]. Germany: Springer Berlin Heidelberg, 2009: 365-369.
- [24] Purcell S, Neale B, Todd-Brown K, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses[J]. *American Journal of Human Genetics*, 2007, 81(3): 559-575.
- [25] Bradbury P J, Zhang Z, Kroon D E, et al. TASSEL: Software for association mapping of complex traits in diverse samples[J]. *Bioinformatics*, 2007, 23(19): 2633-2635.
- [26] Lyu H Y, Li H W, Fan R, et al. Genome-wide association study of dynamic developmental plant height in soybean[J]. *Canadian Journal of Plant Science*, 2017, 97(2): 308-315.
- [27] Jing Y, Zhao X, Wang J, et al. Identification of loci and candidate genes for plant height in soybean (*Glycine max*) via genome wide association study[J]. *Plant Breeding*, 2019, 138(6): 721-732.
- [28] Chang F G, Guo C, Zhang J et al. Genome-wide association studies for dynamic plant height and number of nodes on the main stem in summer sowing soybeans[J]. *Frontiers in Plant Science*, 2018, 9: 1184.
- [29] Shim S, Ha J, Kim M Y, et al. GmBRC1 is a candidate gene for branching in soybean [*Glycine max* (L.) Merrill][J]. *Plant Genetics and Molecular Breeding*, 2019, 20(1): 135.
- [30] Li Y H, Li D, Jiao Y Q, et al. Identification of loci controlling adaptation in Chinese soybean landraces via a combination of conventional and bioclimatic GWAS[J]. *Plant Biotechnology Journal*, 2019, 18(2): 389-401.
- [31] Pan L Y, He J B, Zhao T J, et al. Efficient QTL detection of flowering date in a soybean RIL population using the novel restricted two-stage multi-locus GWAS procedure[J]. *Theoretical and Applied Genetics*, 2018, 131(12): 2581-2599.
- [32] Zhang J P, Song Q J, Cregan P B, et al. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm[J]. *BMC Genomics*, 2015, 16(1): 217.
- [33] Liu Z X, Li H H, Fan X H, et al. Phenotypic characterization and genetic dissection of growth period traits in soybean (*Glycine max*) using association mapping[J]. *PLoS One*, 2017, 256: 76-87.
- [34] Zhao X, Dong H R, Chang H, et al. Genome wide association mapping and candidate gene analysis for hundred seed weight in soybean [*Glycine max* (L.) Merrill][J]. *BMC Genomics*, 2019, 20: 648.
- [35] Contreras R, Mora F, Mar O, et al. A genome-wide association study for agronomic traits in soybean using SNP markers and SNP-based haplotype analysis[J]. *PLoS One*, 2017, 12(2): e0171105.
- [36] Wen Z X, Boyse J F, Song Q J, et al. Genomic consequences of selection and genome-wide association mapping in soybean[J]. *BMC Genomics*, 2015, 16(1): 671.
- [37] Hu Z D, Kan G Z, Hu W, et al. Identification of loci and candidate genes responsible for pod dehiscence in soybean via genome-wide association analysis across multiple environments[J]. *Frontiers in Plant Science*, 2019, 10: 1184.

tiers in Plant Science, 2019, 10: 811.

[38] Zhang T F, Teng F, Wu T T, et al. A combined linkage and GWAS analysis identifies QTLs linked to soybean seed protein and oil content [J]. International Journal of Molecular Sciences, 2019, 20(23): 5915.

[39] Zhang K X, Liu S L, Li W B, et al. Identification of QTNs controlling seed protein content in soybean using multi-locus genome-wide association studies[J]. Frontiers in Plant Science, 2018, 9: 1690.

[40] Zhang Y H, He J B, Meng S, et al. Identifying QTL-allele system of seed protein content in Chinese soybean landraces for population differentiation studies and optimal cross predictions[J]. Euphytica, 2018, 214(9): 157.

[41] Li D, Zhao X, Han Y, et al. Genome-wide association mapping for seed protein and oil contents using a large panel of soybean accessions[J]. Genomics, 2018, 111(1): 90-95.

[42] Zhao X, Chang H, Feng L, et al. Genome-wide association mapping and candidate gene analysis for saturated fatty acid content in soybean seed[J]. Plant Breeding, 2019, 138: 588-598.

[43] Zhang D, Zhang H Y, Hu Z B, et al. Artificial selection on GmOLEO1 contributes to the increase in seed oil during soybean domestication[J]. PLoS Genetics, 2019, 15(7): e1008267.

[44] Li S G, Xu H F, Yang J Y, et al. Dissecting the genetic architecture of seed protein and oil content in soybean from the Yangtze and Huaihe River Valleys using multi-locus genome-wide association studies [J]. International Journal of Molecular Sciences, 2019, 20(12): 3041.

[45] Hwang E Y, Song Q, Jia G, et al. A genome-wide association study of seed protein and oil content in soybean[J]. BMC Genomics, 2014, 15(1): 1.

[46] Dhanapal A, Ray J, Singh S, et al. Genome-wide association study (GWAS) of carbon isotope ratio ($\delta^{13}C$) in diverse soybean [*Glycine max* (L.) Merr.] genotypes [J]. Theoretical and Applied Genetics, 2015, 128(1): 73-91.

[47] Zhang W, Liu X L, Cui Y M, et al. A cation diffusion facilitator, GmCDF1, negatively regulates salt tolerance in soybean[J]. PLoS Genetics, 2019, 15(1): e1007798.

[48] Mamidi S, Lee R K, Goos J R, et al. Genome-wide association studies identifies seven major regions responsible for iron deficiency chlorosis in soybean (*Glycine max*) [J]. PLoS One, 2014, 9(9): e107469.

[49] Zhang D, Song H, Cheng H, et al. The acid phosphatase-encoding gene *GmACP1* contributes to soybean tolerance to low-phosphorus stress[J]. PLoS Genetics, 2014, 10(1): e1004061.

[50] Wen Z, Tan R, Zhang S, et al. Integrating GWAS and gene expression data for functional characterization of resistance to white mold in soybean [J]. Plant Biotechnology Journal, 2018, 16(11): 1825-1835.

[51] Zhang J, Wen Z, Li W, et al. Genome-wide association study for soybean cyst nematode resistance in Chinese elite soybean cultivars [J]. Molecular Breeding, 2017, 37(5): 60.

[52] Santos J, Ferreira E, Passianotto A, et al. Association mapping of a locus that confers southern stem canker resistance in soybean and SNP marker development[J]. BMC Genomics, 2019, 20(1): 798.

[53] Chu S S, Wang J, Zhu Y, et al. An R2R3-type MYB transcription factor, GmMYB29, regulates isoflavone biosynthesis in soybean[J]. PLoS Genetics, 2017, 13(5): e1006770.

[54] Contreras R, Mora F, Mar O, et al. A Genome-wide association study for agronomic traits in soybean using SNP markers and SNP-based haplotype analysis[J]. PLoS One, 2017, 12(2): e0171105.

[55] Yu J, Holland J B, Mc Mullen M D, et al. Genetic design and statistical power of nested association mapping in maize [J]. Genetics, 2008, 178(1): 539-551.

[56] Yang J, Jiang H, Yeh C T, et al. Extreme-phenotype genome-wide association study (XP-GWAS): A method for identifying trait-associated variants by sequencing pools of individuals selected from a diversity panel [J]. The Plant Journal, 2015, 84(3): 587-596.

[57] Huang X, Wei X, Sang T, et al. Genome-wide association studies of 14 agronomic traits in rice landraces [J]. Nature Genetics, 2010, 42(11): 961-967.

[58] Stacey J W, Joanna M B. Gene-environment interactions in Genome-Wide Association studies: Current approaches and new directions[J]. Journal of Child Psychology and Psychiatry, 2013, 54(10): 1120-1134.

[59] Yang Q, Wu H S, Guo C Y, et al. Analyze multivariate phenotypes in genetic association studies by combining univariate association tests [J]. Genetic Epidemiology, 2010, 34(5): 444-454.