

全基因组关联分析在蔬菜育种研究中的应用

曾美娟 刘建汀 卓玲玲 陈敏敏 叶新如 王 彬 朱海生* 温庆放*

(福建省蔬菜遗传育种重点实验室, 福建省农业科学院作物研究所, 福建省蔬菜工程技术研究中心, 福州 350013)

摘 要: 全基因组关联分析 (GWAS) 是一种高效地将表型和基因型进行关联并用于遗传作图和搜寻相关性状候选基因的方法, 可同时对多个复杂性状进行关联分析, 在蔬菜育种中应用日益广泛。本文对全基因组关联分析方法进行概述, 对近年来全基因组关联分析在蔬菜生长发育过程相关性状、品质和产量性状以及抗性性状上的应用研究进行总结, 并展望了其在蔬菜育种领域的应用前景。

关键词: 全基因组关联分析 (GWAS); 单核苷酸多态性 (SNP); 蔬菜育种; 综述

随着 20 世纪 80 年代初基于 DNA 的分子标记和 80 年代末先进的统计工具的出现, 挖掘控制数量性状的基因组区域成为可能。挖掘控制数量性状基因组的方法包括数量性状座位 (quantitative trait locus, QTL) 区间定位和全基因组关联分析 (genome-wide association study, GWAS)。传统的 QTL 定位高度依赖于双亲的遗传多样性, 检测到 QTL 效率因群体而异。QTL 区域也可能相当大, 包含太多的基因, 很难作为潜在的候选基因进行研究。此外, QTL 定位往往需要构建作图群体, 耗时较长且定位精度不高 (Rafalski, 2010)。GWAS 可以在一定程度上克服 QTL 分析的局限性, 它可以利用自然群体缩小候选区域, 同时对多个性状进行分析 (Yu & Buckler, 2006; Huang & Han,

2014), 以降低分析误差 (曹英杰 等, 2019), 极大地提高育种效率。全基因组关联分析高效地将表型和基因型进行关联并用于遗传作图和搜寻相关性状候选基因 (Gajardo et al., 2015; 阙青敏 等, 2019), 可同时对多个复杂性状进行关联, 检测多个等位基因, 适用于定位性状关联区间, 功能基因研究, 开发性状选育标记等, 具有高分辨率和高通量等优点, 在蔬菜育种研究中的应用日益广泛。全基因组关联分析在揭示蔬菜复杂性状的分子机理和蔬菜分子育种中起到重要的作用 (李廷雨 等, 2020)。

1 全基因组关联分析

1.1 概述

全基因组关联分析是近年来兴起的遗传分析方法, 其以连锁不平衡 (linkage disequilibrium, LD) 为基础, 通过识别数百个或数千个个体定位群体中高密度的分子标记, 一般是上万个甚至上百万个单核苷酸多态性 (single nucleotide polymorphism, SNP) 标记, 筛选出与复杂性状表现型变异相关联的分子标记 (Du et al., 2018)。换言之, 通过关联分析分子标记与性状变异, 对群体中的个体进行大规模的基因分型和表型分析, 从而识别导致个体之间表型差异的基因组区域 (Hirschhorn & Daly,

曾美娟, 女, 博士, 助理研究员, 主要从事蔬菜分子育种研究, E-mail: 860905zzmr@163.com

* 通信作者 (Corresponding authors): 朱海生, 男, 研究员, 主要从事蔬菜育种研究, E-mail: zhs0246@163.com; 温庆放, 男, 研究员, 主要从事蔬菜育种研究, E-mail: lfjvrc@163.com

收稿日期: 2020-06-08; 接受日期: 2020-12-07

基金项目: 福建省科技计划项目-省属公益类科研院所基本科研专项 (2020R1031002), 福建省自然科学基金项目 (2019J01112), 福建省农业科学院科技创新平台专项 (CXPT202001), 国家大宗蔬菜产业技术体系福州综合试验站项目 (CARS-23-G-53), 福建省科技重大专项专题 (2018NZ0002-3)

2005; Huang et al., 2010; Zhao et al., 2011)。连锁不平衡是指群体内不同位点上等位基因间的非随机关联(曹英杰等, 2019)。连锁不平衡是 GWAS 分析的基础(Abecasis & Cookson, 2000), 受多因素的影响, 不同物种间基因组中连锁不平衡也存在显著差异。高 LD 水平的群体, 能够缩减 GWAS 分析时所需的群体数量。近年来, 随着以 SNP 为代表的第 3 代分子标记技术的发展, 大大加快了 GWAS 的发展速度, 几千甚至几百万个标记被用于 1 个 GWAS, 在人类和动植物复杂性状遗传研究中已取得初步成果(段忠取和朱军, 2015), 亦已成功应用于多种作物的重要农艺性状的遗传研究(Elshire et al., 2011; 赵振卿等, 2014; He et al., 2014)。

1.2 分析策略和常用软件

GWAS 应用于植物育种的研究策略主要包括以下几方面: ① 选择植物群体材料。群体规模太大,

性状调查费用和基因型检测费用均会相应增多。选择表型齐全和遗传变异丰富的研究群体可以减少所需群体内的个体数目, 提高关联分析的分辨率。② 表型鉴定。设计合理的田间试验, 多年多点种植试验材料和多区域重复随机调查表型性状, 并将调查结果进行整合(刘坤等, 2018)。③ 基因型的测定。获取选定试验材料的全基因组 SNP 位点, 从而实现基因型分析(Seki et al., 2005; Rutkoski et al., 2013; 韩德鹏等, 2018)。④ 关联分析。对选定试验群体进行群体结构分析, 选用合适的统计分析模型对基因型与表型进行分析(Raman et al., 2019; 孙程明等, 2020)。涉及质量性状关联分析时, 通常可以采用 Logistic 回归模型进行分析; 涉及数量性状时, 通常可采用普通线性回归模型(卜李那和赵毅强, 2019), 但数量性状通常受到多种因素的共同影响, 在研究过程中通常采用不同的混合模型(表 1)。⑤ 候选基因筛选(姜洪真等, 2018)。

表 1 近年来部分 GWAS 中采用的混合模型方法及其特点

方法类型	特点	参考文献
混合线性模型 (MLM)	处理多个级别的相关性; 固定效应; 随机效应	Yu et al., 2006; Lamara et al., 2016; de la Torre et al., 2019
多位点混合模型 (MLMM)	在 MLM 上进行改进; 可有效地控制群体结构误差率	Segura et al., 2012
压缩混合线性模型 (CMLM)	基于个体之间的亲缘关系, 将个体聚类为较少的群体; 成对群体之间的亲缘关系可替换为成对个体之间的亲缘关系; 大大减少了计算量	Zhang et al., 2010
高效混合模式关联 (EMMA)	计算种群结构和遗传关系的精确方法; 计算速度快; 结果的可靠性高	Kang et al., 2008; Uchiyama et al., 2013
全基因组高效混合模型关联 (GEMMA)	比 EMMA 快; 即使存在庞大的群体结构, 也能产生准确的 p 值; 适合于大型关联分析	Zhou & Stephens, 2012; Bdeir et al., 2019
高效混合模式加速关联 (EMMAX)	短时间内可对海量数据进行关联作图	Kang et al., 2010

目前分析过程涉及的软件很多, 其中 PLINK 软件(Purcell et al., 2007)是较早使用的关联分析软件, 它可用于复杂数量性状、关联作图、数据转化与处理、LD 分析、单倍型检验等。采用 SAMTOOLS 软件(Li et al., 2009)检测、过滤 SNP, 采用 ANNOVAR 软件(Wang et al., 2010)对检测到的 SNP 进行注释, 采用 GCTA 软件(Yang et al., 2011)对群体进行主成分分析、单性状和两相关性状关联分析, 采用 STRUCTURE 软件(Evanon et al., 2005; 王艳玲等, 2017; Volante et al., 2017)对群体结构进行分析。采用 GEMMA 软件(Zhou & Stephens, 2012)基于 SNP 进行关联分析。

采用 TASSEL 软件(Bradbury et al., 2007)进行各种模型的关联分析, 估算 LD 值和作图、估测群体结构和绘制基于遗传距离的树状图等。GAPIT 软件(Lipka et al., 2012)用于 BLUP 基因组预测关联分析和 BLUP 基因组预测。

2 全基因组关联分析在蔬菜育种研究中的应用

2.1 蔬菜生长发育过程相关性状应用研究

前人在蔬菜生长发育过程相关性状方面, 如黄瓜发芽期、黄瓜幼苗阶段、白菜类作物抽薹开花、菠菜雌雄同株、普通菜豆相关性状的 control 基因等方

面都开展了全基因组关联分析。张松等(2019)对黄瓜发芽期进行全基因组关联分析,检测到5个与相对发芽率关联的位点,2个与相对发芽势关联的位点,2个与相对发芽指数关联的位点,2个与相对胚根长度关联的位点。蔡和序等(2020)对黄瓜幼苗下胚轴长度进行全基因组关联分析,通过分析关联 SNP 位点的 LD 区间序列,获得 *Csa1G074930*、*Csa1G475980*、*Csa2G381650*、*Csa3G141820*、*Csa4G051570*、*Csa3G627150*、*Csa5G174640*、*Csa6G362970* 等8个与黄瓜下胚轴长度有关的候选基因,其中既有光形态建成、泛素化、激素信号通路等调控基因,也有调控网络下游参与细胞生长发育,调节细胞大小,直接调控黄瓜下胚轴长度的基因。抽薹开花是白菜类作物关键的农艺性状。白菜类作物主要包括大白菜、普通白菜、芜菁等。挖掘白菜类作物抽薹开花调控位点和基因,对白菜类作物的培育具有重要的意义。龚振平(2016)对182份大白菜自然群体晚抽薹性状进行全基因组关联分析,获得5个与耐抽薹性状显著关联的位点,为进一步发掘相关性状的候选基因提供了依据。高宝祯等(2017)通过全基因组关联分析鉴定出33个与白菜类作物开花时间相关的显著关联信号。通过定位出的开花时间候选位点,再根据白菜类作物与同源物种拟南芥的基因共线性关系以及基因功能注释结果初步鉴定出与14个白菜类作物开花时间相关的候选基因。汪豪英等(2019)通过对82份菠菜高代自交系的全基因组关联分析,采用压缩混合线性模型在菠菜4号染色体上检测到1个强关联区域,并将控制菠菜雌雄同株的基因 *Xm* 定位在64.6 kb 的区间内。该范围内存在3个基因: *Spo24600*、*Spo24601* 和 *Spo24602*。最近, Wu 等(2020)对来自19个国家的683份普通菜豆资源的全基因组进行重测序,发掘出超过480万个 SNP,构建出国际首张精细的普通菜豆单倍型图谱,鉴定出505个与主要农艺性状紧密相关的遗传位点。

上述研究表明采用全基因组关联分析能够有效鉴定蔬菜生长发育过程相关农艺性状的关键遗传位点。因为蔬菜作物的农艺性状往往都是由多个基因控制且受环境影响,与单基因控制的性状相比,其遗传基础更为复杂。而全基因组关联分析是经典的定量遗传理论的拓展,采用全基因组关联分析的方

法能够既简单且快速地鉴定出蔬菜作物控制发育过程相关性状的重要基因,对蔬菜作物农艺性状相关基因的研究以及辅助育种具有重大意义。

2.2 蔬菜品质和产量性状应用研究

在蔬菜品质和产量性状方面,如番茄的果实质量、果实硬度、心室数目、果形指数、代谢物含量以及辣椒的辣椒素含量等相关性状的控制基因都开展了全基因组关联分析。Xu 等(2013)利用关联分析的方法对44份栽培种番茄、127份樱桃番茄和17份醋栗番茄的果实质量、果实硬度、心室数目等性状进行了初步研究,共检测到40个位点,其中果实质量、心室数目、可溶性固形物等性状的主效位点信号区域都有相关的功能基因。祝光涛(2015)利用843 316个位点对253个番茄栽培种亚群中的果实颜色、果实质量和果形指数等5个性状进行关联分析,发现了11个明显的关联信号位点,其中1个果皮颜色位点、2个果形指数位点和2个心室数目位点和前人的研究结果一致,另外6个位点为新发现的位点。Bauchet 等(2017)借助10 000个 SNP 标记对300份番茄的60种初级和次级代谢产物开展了全基因组关联分析,确定了79个与13个初级代谢产物和19个次级代谢产物高度关联的位点。同时,还发现了4个基因组区域可控制几种代谢物变异,并发掘了决定代谢物含量的候选基因,揭示了番茄亚种复杂而独特的代谢物调控机理。赵建涛(2016)利用混合线性模型对番茄果实中主要的17种糖酸组分进行全基因组关联分析,共检测到139个显著关联位点,除了蔗糖醇外,在其他16个糖酸物质上至少检测到1个显著关联位点。Sauvage 等(2014)利用多位点混合模型对163份番茄种质进行关联分析,利用遍布全基因组的5 995个 SNP 位点对影响番茄品质的76种代谢物质进行了分析,检测到了控制19个性状的44个显著位点。Nimmakayala 等(2016)使用7 331个 SNP 标记对辣椒性状进行研究,发现72个 SNP 标记与辣椒素含量相关,包括1个候选基因,该基因编码一种具有与CS相似的酰基转移酶功能的锚蛋白样蛋白。Han 等(2018)通过QTL定位和GWAS挖掘控制辣椒中辣椒素含量的候选基因,共检测到69个QTL区域,其中10个区域与2个双亲群体的QTL位于同一位置。在这些区域中,鉴

定出 5 个已知参与辣椒素生物合成的候选基因。

随着人们生活水平逐步提高,多样化的蔬菜品种给予了大众更多的选择,大众对蔬菜品质的要求则越来越高。在蔬菜育种中,品质高低影响着该品种能否适应市场以及满足消费者的需求。蔬菜品质育种也是蔬菜遗传改良的重点。同时,蔬菜产量的高低也影响着其经济价值,通过全基因组关联分析来开展蔬菜品质和产量性状的相关研究对蔬菜品质和产量的提升具有重要意义。

2.3 蔬菜抗性性状应用研究

为有效控制蔬菜病害的发生与危害,选育和利用具有相关抗性的蔬菜品种也是育种的关键。随着全基因组学的发展,全基因组关联分析也在马铃薯抗晚疫病、大白菜抗霜霉病、瓠瓜抗白粉病、黄瓜抗低温等蔬菜抗性相关基因的挖掘中得到应用。为了明确马铃薯晚疫病抗性的稳定性与标记间的关联,Lindqvist-Kreuze 等(2014)结合田间晚疫病表型数据,对适应热带高地的马铃薯群体的基因型进行全基因组关联分析,发现第 9 号染色体上的 SNP 标记与马铃薯晚疫病抗性相关,且与稳定性有关。龚振平(2016)对 182 份自交系材料组成的大白菜自然群体开展 5 种病害的全基因组关联分析,分别获得与霜霉病(2 个)、病毒病(5 个)、黑腐病(2 个)、黄萎病(5 个)和根肿病(8 个)抗性显著关联的 22 个位点或热点区。吴晓花等(2020)利用两年的抗病表型数据,对 117 份瓠瓜微核心种质的白粉病抗性进行研究,通过全基因组关联分析,分别获得 22 个和 13 个与白粉病抗性相关的 SNP 标记。王伟平等(2019)以黄瓜核心种质为材料开展苗期耐低温鉴定和筛选,并进行全基因组关联分析,挖掘耐低温相关位点。在 1、3、4、5 号染色体上分别检测到苗期耐低温位点 *gLTS1.1*、*gLTS3.1*、*gLTS4.1* 和 *gLTS5.1*。魏爽等(2019)进行了黄瓜苗期耐热性筛选,通过全基因组关联分析共检测到 7 个与苗期耐热性相关位点 *gHII4.1*、*gHII5.1*、*gHII5.2*、*gHII6.1*、*gHII7.1*、*gHII4.2*、*gHII6.2*。

施用化学药剂能够在一定程度上控制相应的病虫害危害,但也带来环境污染等问题,同时增加农民种植成本,产生蔬菜安全问题。通过全基因组关联分析,将这些性状关联标记用于蔬菜分子辅助育种,有助于选育出具有抗性的蔬菜品种。而选育具

有抗性的蔬菜品种又是目前防治相应病害、适应不佳种植环境及选育反季蔬菜较为经济、有效的方法。

3 展望

传统的 QTL 定位高度依赖于双亲的遗传多样性,检测到的 QTL 效应因群体而异。QTL 区域也可能相当大,包含太多的基因,而利用全基因组关联分析(GWAS)可以利用自然群体缩小候选区域,在一定程度上克服 QTL 分析的局限性。当然,GWAS 的应用也存在一定的局限性,例如在自花授粉作物中,用 GWAS 定位性状相关基因难以达到单基因水平,一些群体结构因素易导致假阳性,非遗传因素产生的表型变化导致的假阳性以及基因与环境的互作效应亦会影响 GWAS 结果(Stacey & Joanna, 2013)。为获得更精准的 GWAS 结果,研究人员需对群体结构加以准确分析,通过增大群体规模来尽可能减少假阳性。相关技术的不断更新也将在一定程度上降低群体结构的干扰。同时,也有必要对结果进行验证(Zhu et al., 2008; Korte & Farlow, 2013)。由于每种方法都有其局限性,将不同方法进行联合,取长补短可在一定程度上提高分析结果的准确性。已有研究表明,QTL 定位联合 GWAS 是鉴定控制复杂性状的基因位点的一种强有力的组合方法(Han et al., 2018)。

GWAS 只识别与目标性状相关的基因组区域,而不是发现基因,一个性状相关的基因组区域内会有许多标记(例如,多个 SNP 标记)。随着下一代测序技术的发展,几千甚至几百万个标记被用于 1 个 GWAS,每个 GWAS 可用的标记数量大幅增加,SNP 集被应用于 GWAS 是当前的关键研究领域,也将促进 GWAS 的发展。近年来,GWAS 研究取得了重大进展,这些性状关联标记也被有效地用于标记辅助选择,以补充传统的蔬菜育种方法,改良简单和复杂的数量性状。GWAS 将极大推动蔬菜由传统育种向高效、定向的分子设计育种转变,亦将为揭示蔬菜主要农艺性状、品质性状和抗性性状的分子机理发挥重要的作用。

参考文献

- 卜李那,赵毅强. 2019. 全基因组关联分析及其扩展方法的研究进展. 农业生物技术学报, 27 (1): 150-158.
蔡和序,薄凯亮,周琪,苗晗,董邵云,顾兴芳,张圣平. 2020.

- 黄瓜幼苗下胚轴长度 GWAS 分析及候选基因挖掘. 中国农业科学, 53 (1): 122-132.
- 曹英杰, 杨剑飞, 王宇. 2019. 全基因组关联分析在作物育种研究中的应用. 核农学报, 33 (8): 1508-1518.
- 韩德鹏, 周灿, 郑伟, 李亚贞, 付东辉, 周庆红. 2018. 甘蓝型油菜主花序性状全基因组关联分析. 核农学报, 32 (3): 463-476.
- 段忠取, 朱军. 2015. 全基因组关联分析研究进展. 浙江大学学报 (农业与生命科学版), 41 (4): 385-393.
- 高宝祯, 刘博, 李石开, 梁建丽, 程锋, 王晓武, 武剑. 2017. 白菜类作物开花时间的全基因组关联分析. 中国农业科学, 50 (17): 3375-3385.
- 龚振平. 2016. 大白菜抗病和晚抽薹性状的 GWAS 分析及其优异资源发掘 (博士学位论文). 北京: 中国农业科学院.
- 姜洪真, 马伯军, 钱前, 高振宇. 2018. 全基因组关联分析 (GWAS) 在作物农艺性状研究中的应用. 农业生物技术学报, 26 (7): 1244-1257.
- 李廷雨, 黎永力, 甘卓然, 石文茜, 董利东, 刘宝辉. 2020. 全基因组关联分析在大豆中的研究进展. 大豆科学, 39 (3): 479-484.
- 刘坤, 张雪海, 孙高阳, 闫鹏帅, 郭海平, 陈思远, 薛亚东, 郭战勇, 谢惠玲, 汤继华, 李卫华. 2018. 玉米株型相关性状的 GWAS 分析. 中国农业科学, 51 (5): 821-834.
- 阙青敏, 欧阳昆唏, 李培, 陈晓阳. 2019. 全基因组关联分析 (GWAS) 在林木育种中的应用. 植物生理学报, 55 (11): 1555-1562.
- 孙程明, 陈锋, 陈松, 彭琦, 张维, 易斌, 张洁夫, 傅廷栋. 2020. 甘蓝型油菜每角粒数的全基因组关联分析. 作物学报, 46 (1): 147-153.
- 汪豪英, 刘志远, 王晓武, 武剑, 张合龙, 夏志兰, 徐兆生, 钱伟. 2019. 菠菜雌雄同株性状的全基因组关联分析. 园艺学报, 46 (8): 1495-1502.
- 王伟平, 宋子超, 薄凯亮, 董邵云, 魏爽, 苗晗, 李锦斌, 张圣平, 顾兴芳. 2019. 黄瓜核心种质幼苗耐低温性评价及 GWAS 分析. 植物遗传资源学报, 20 (6): 1606-1612.
- 王艳玲, 许彦宾, 李琼, 王盼乔, 胡建斌, 杨路明, 朱华玉, 孙守, 马长生. 2017. 基于 GWAS 发掘甜瓜核心种质果实性状的关联位点. 农业生物技术学报, 25 (9): 1434-1442.
- 魏爽, 张松, 薄凯亮, 王伟平, 苗晗, 董邵云, 顾兴芳, 张圣平. 2019. 黄瓜核心种质幼苗耐热性评价及 GWAS 分析. 植物遗传资源学报, 20 (5): 1223-1231.
- 吴晓花, 汪颖, 吴新义, 徐沛, 汪宝根, 鲁忠富, 李国景. 2020. 瓠瓜 (*Lagenaria siceraria*) 白粉病抗性的全基因组关联分析. 分子植物育种, 18 (3): 759-764.
- 张松, 苗晗, 宋子超, 薄凯亮, 王烨, 顾兴芳, 张圣平. 2019. 黄瓜发芽期耐热性评价及全基因组关联分析. 植物遗传资源学报, 20 (2): 335-346.
- 赵建涛. 2016. 番茄果实主要糖酸成分全基因组关联分析 (硕士论文). 杨凌: 西北农林科技大学.
- 赵振卿, 顾宏辉, 盛小光, 虞慧芳, 王建升, 曹家树. 2014. 作物数量性状位点研究进展及其育种应用. 核农学报, 28 (9): 1615-1624.
- 祝光涛. 2015. 番茄重要农艺性状的全基因组关联分析及野生种质在栽培种的渐渗分析 (博士学位论文). 北京: 中国农业科学院.
- Abecasis G R, Cookson W O C. 2000. GOLD-Graphical overview of linkage disequilibrium. Bioinformatics, 16: 182-183.
- Bauchet G, Grenier S, Samson N, Bonnet J, Grivet L, Causse M. 2017. Use of modern tomato breeding germplasm for deciphering the genetic control of agronomical traits by genome wide association study. Theoretical & Applied Genetics, 130 (5): 875-889.
- Bdeir R, Muchero W, Yordanov Y, Tuskan G, Busov V, Gailing O. 2019. Genome-wide association studies of bark texture in *Populus trichocarpa*. Tree Genet Genomes, 15: 14.
- Bradbury P J, Zhang Z, Kroon D E, Casstevens T M, Ramdoss Y, Buckler E S. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics, 23: 2633-2635.
- de la Torre A R, Puiu D, Crepeau M W, Stevens K, Salzberg S L, Langley C H, Neale D B. 2019. Genomic architecture of complex traits in loblolly pine. New Phytol, 221 (4): 1789-1801.
- Du Q, Lu W, Quan M, Xiao L, Song F, Li P, Zhou D, Xie J, Wang L, Zhang D. 2018. Genome-wide association studies to improve wood properties: challenges and prospects. Front Plant Sci, 9: 1912.
- Elshire R J, Glaubitz J C, Sun Q, Poland J A, Kawamoto K, Buckler E S, Mitchell S E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One, 6: e19379.
- Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology, 14: 2611-2620.
- Gajardo H A, Wittkop B, Soto-Cerda B, Higgins E E, Parkin I A P, Snowdon R J, Federico M L, Iniguez-Luy F. 2015. Association mapping of seed quality traits in *Brassica napus* L. using GWAS and candidate QTL approaches. Molecular Breeding, 35: 143.
- Han K, Lee H Y, Ro N Y, Hur O S, Lee J H, Kwon J K, Kang B C. 2018. QTL mapping and GWAS reveal candidate genes controlling capsaicinoid content in *Capsicum*. Plant Biotechnology Journal, 16 (9): 1546-1558.
- He J, Zhao X, Laroche A, Lu Z X, Liu H, Li Z. 2014. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in Plant Science, 5: 48.
- Hirschhorn J N, Daly M J. 2005. Genome-wide association studies for common diseases and complex traits. Nature Reviews Genetics, 6 (2): 95-108.
- Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L,

- Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler E, Qian Q, Zhang Q, Li J, Han B. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics*, 42: 961-967.
- Huang X, Han B. 2014. Natural variations and genome-wide association studies in crop plants. *Annual Review of Plant Biology*, 65: 531-551.
- Kang H M, Zaitlen N A, Wade C M, Kirby A, Heckerman D, Daly M J, Eskin E. 2008. Efficient control of population structure in model organism association mapping. *Genetics*, 178: 1709-1723.
- Kang H M, Sul J H, Service S K, Zaitlen N A, Kong S K, Freimer N B, Sabatti C, Eskin E. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, 42: 348-354.
- Korte A, Farlow A. 2013. The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, 9: 29.
- Lamara M, Raherison E, Lenz P. 2016. Genetic architecture of wood properties based on association analysis and co-expression networks in white spruce. *New Phytol*, 210 (1): 240-255.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25 (16): 2078-2079.
- Lindqvist-Kreuzer H, Gastelo M, Perez W, Forbes G A, de Koeijer D, Bonierbale M. 2014. Phenotypic stability and genome-wide association study of late blight resistance in potato genotypes adapted to the tropical highlands. *Phytopathology*, 104 (6): 624-633.
- Lipka A E, Tian F, Wang Q. 2012. GAPIT: genome association and prediction integrated tool. *Bioinformatics*, 28: 2397-2399.
- Nimmakayala P, Abburi V L, Saminathan T, Alaparthi S B, Almeida A, Davenport B, Nadimi M, Davidson J, Tonapi K, Yadav L, Malkaram S, Vajja G, Hankins G, Harris R, Park M, Choi D, Stommel J, Reddy U. 2016. Genome-wide diversity and association mapping for capsaicinoids and fruit weight in *Capsicum annuum* L. *Scientific Reports*, 6: 38081.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M A R, Bender D, Maller J, Sklar P, Bakker P W, Daly M J, Sham P C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81 (3): 559-575.
- Rafalski J A. 2010. Association genetics in crop improvement. *Current Opinion in Plant Biology*, 13 (2): 174-180.
- Raman H, Raman R, Qiu Y, Yadav A S, Sureshkumar S, Borg L, Rohan M, Wheeler D A, Owen Q, Menz L, Balasubramanian S. 2019. GWAS hints at pleiotropic roles for FLOWERING LOCUS T in flowering time and yield-related traits in canola. *BMC Genomics*, 20: 636.
- Rutkoski J E, Poland J, Jannink J L, Sorrells M E. 2013. Imputation of unordered markers and the impact on genomic selection accuracy. *G3 (Bethesda)*, 3 (3): 427-439.
- Sauvage C, Segura V, Bauchet G, Stevens R, Do P T, Nikoloski Z, Fernie A R, Causse M. 2014. Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. *Plant physiology*, 165: 1120-1132.
- Segura V, Vilhjalmsen B J, Platt A, Korte A, Seren U, Long Q, Nordborg M. 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics*, 44: 825-830.
- Seki S, Kawaguchi Y, Chiba K, Mikami Y, Kizawa H, Oya T, Mio F, Mori M, Miyamoto Y, Masuda I, Tsunoda T, Kamata M, Kubo T, Toyama Y, Kimura T, Nakamura Y, Ikegawa S. 2005. A functional SNP in CILP, encoding cartilage intermediate layer protein, is associated with susceptibility to lumbar disc disease. *Nat Genet*, 37 (6): 607-612.
- Stacey J W, Joanna M B. 2013. Gene-environment interactions in genome-wide association studies: current approaches and new directions. *Journal of Child Psychology and Psychiatry*, 54 (10): 1120-1134.
- Uchiyama K, Iwata H, Moriguchi Y, Ujino-Ihara T, Ueno S, Taguchi Y, Tsubomura M, Mishima K, Iki T, Watanabe A, Futamura N, Shinohara K, Tsumura Y. 2013. Demonstration of genome-wide association studies for identifying markers for wood property and male strobili traits in *Cryptomeria japonica*. *PLoS One*, 8 (11): e7986611.
- Volante A, Tondelli A, Aragona M, Valente MT, Biselli C, Desiderio F, Bagnaresi P, Matic S, Gullino M L, Infantino A, Spadaro D, Vale G. 2017. Identification of bakanae disease resistance loci in japonica rice through genome wide association study. *Rice*, 10 (1): 29.
- Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*, 38 (16): e164.
- Wu J, Wang L, Fu J, Chen J, Wei S, Zhang S, Zhang J, Tang Y, Chen M, Zhu J, Lei L, Geng Q, Liu C, Wu L, Li X, Wang X, Wang Q, Wang Z, Xing S, Zhang H, Blair M W, Wang S. 2020. Resequencing of 683 common bean genotypes identifies yield component trait associations across a north-south cline. *Nature Genetics*, 52: 118-125.
- Xu J, Ranc N, Muñoz S, Rolland S, Bouchet J P, Desplat N, Paslier M C L, Liang Y, Brunel D, Causse M. 2013. Phenotypic diversity and association mapping for fruit quality traits in cultivated tomato and related species. *Theoretical and Applied Genetics*, 126 (3): 567-581.
- Yang J, Lee S H, Goddard M E, Visscher P M. 2011. GCTA: a tool for genome-wide complex trait analysis. *Am J Human Genet*, 88: 76-82.
- Yu J, Buckler E S. 2006. Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology*, 17(2):

155–160.

Yu J, Pressoir G, Briggs W H, Bi I V, Yamasaki M, Doebley J F, McMullen M D, Gaut B S, Nielsen D M, Holland J B, Kresovich S, Buckler E S. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, 38, 203–208.

Zhang Z, Ersoz E, Lai C Q, Todhunter R J, Tiwari H K, Gore M A, Bradbury P J, Yu J, Arnett D K, Ordovas J M, Buckler E S. 2010. Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, 42, 355–360.

Zhao K Y, Tung C W, Eizenga G C, Wright M H, Ali M L, Price A H, Norton G J, Islam M R, Reynolds A, Mezey J, McClung A M, Bustamante C D, McCouch S R. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nature Communications*, 2 (1): 467.

Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, 44: 821–824.

Zhu C, Gore M, Buckler E S, Yu J. 2008. Status and Prospects of association mapping in plants. *Plant Genome*, 1: 5–20.

Application of Genome-wide Association Study in Vegetable Breeding

ZENG Meijuan, LIU Jianting, ZHUO Lingling, CHEN Mindong, YE Xinru, WANG Bin, ZHU Haisheng*, WEN Qingfang*

(Key Laboratory of Vegetable Genetics and Breeding of Fujian Province, Crops Research Institute, Fujian Academy of Agricultural Sciences, Vegetables Engineering Research Center of Fujian Province, Fuzhou 350013, Fujian, China)

Abstract: Genome-wide association study (GWAS) is a method efficiently associating phenotype with genotype, and being used in genetic mapping and searching for candidate genes with related traits. It can be used to carry out associated analysis on multiple complex traits at the same time, and is widely used in vegetable breeding. This paper reviewed the method of whole genome association analysis; summarized its application recently in studying the related traits, including quality, yield and resistance during vegetable growth and development process. The paper also looked forward to its application prospect in vegetable breeding.

Keywords: genome-wide association study (GWAS); single nucleotide polymorphism (SNP); vegetable breeding; review

· 信息 ·

Cell 封面论文揭示“超级害虫”烟粉虱为害 600 多种植物的奥秘

2021 年 3 月 25 日,《细胞 (Cell)》以封面论文的形式发表了中国农业科学院蔬菜花卉研究所张友军团队的最新研究成果。该研究在国际上首次揭示了烟粉虱在距今 3 500 万年~8 600 万年之间,通过水平基因转移方式“偷窃”植物源解毒基因,进而分解掉植物的防御性物质,使自身可直接取食 600 多种植物。该研究不但发现昆虫“以子之矛,攻子之盾”的进化现象,还为探索昆虫适应性进化规律开辟了新的视角,是多食性昆虫广泛寄主适应性进化机制研究的重大突破,创新了害虫种群暴发危害理论。该研究将为研发烟粉虱精准绿色防控技术提供全新思路。

据张友军研究员介绍,在漫长的植物和昆虫的共进化过程中,植物产生多种次生物质抵御昆虫的为害,酚糖类物质就是其中典型的代表。酚糖广泛存在于各种植物中,它能够抑制昆虫取食,影响昆虫的生长发育,甚至对昆虫有致死作用。烟粉虱是能够为害 600 多种植物的“超级害虫”,因此,必然存在能够解毒酚糖的机制。研究团队通过比较基因组学和生物信息学的方法发现,在烟粉虱的基因组中存在 1 个植物源基因——酚糖丙二酰基转移酶。研究人员利用分子生物学和进化分析等手段证实该基因是植物源的水平转移基因。在植物中,酚糖丙二酰基转移酶能够分解酚糖,从而使植物免受酚糖过量积累对自身造成的伤害。进一步研究也证实烟粉虱活体能够分解酚糖。研究团队还发现,干扰该基因能够显著增加烟粉虱成虫的死亡率,并抑制烟粉虱产卵。更为重要的是,持续干扰该基因能够实现对烟粉虱的完全防控,而对非靶标的节肢动物蚜虫和斑叶螨没有任何影响。

(中国农业科学院蔬菜花卉研究所)