

房婉萍, 雷小刚, 杨彬, 等. 全基因组关联分析在茶叶研究中的应用进展[J]. 华中农业大学学报, 2022, 41(5): 33-40.
DOI: 10.13300/j.cnki.hnlkxb.2022.05.005

全基因组关联分析在茶叶研究中的应用进展

房婉萍, 雷小刚, 杨彬, 王亚, 马媛春

南京农业大学园艺学院, 南京 210095

摘要 全基因组关联分析(genome-wide association analysis, GWAS)是以高通量测序技术为基础, 结合生物信息学和统计学方法, 在全基因组水平上鉴定调控复杂性状的基因变异, 是研究复杂农艺性状和遗传变异最有力和最有效的研究方法, 其核心是研究遗传变异和目标性状之间的关联。GWAS研究检测到的关联位点一般很少, 而且关联的位点仅能解释很少一部分性状变异。本文介绍了影响GWAS的主要因素, 总结了GWAS在茶叶饮料消费、茶树重要农艺性状、茶叶品质和茶树群体结构研究中取得的一系列进展, 提出了茶树GWAS研究中遇到的问题和未来的发展方向。

关键词 茶树; 全基因组关联分析; 农艺性状; 茶叶消费; 品质; 育种

中图分类号 S571.103 **文献标识码** A **文章编号** 1000-2421(2022)05-0033-08

茶树(*Camellia sinensis*(L.)O.Kuntze)是重要的经济作物, 具有数千年种植历史。全球超50个国家种植茶树, 茶饮料深受消费者的喜爱^[1-2]。据统计, 人类每天饮茶超20亿杯, 全球茶叶年产量约为500万t, 每年由茶带来的产值约为57亿美元^[3]。茶叶中富含茶多酚、儿茶素和咖啡碱等对人体健康有益的次生代谢物, 是良好的养生保健产品^[4]。目前, 对于氨基酸、儿茶素、茶多酚、咖啡碱等茶叶次生代谢物的研究体系较为成熟并且研究得比较深入, 但茶树分子育种研究相对滞后^[5]。茶树基因组的公布不仅加快了茶树育种和分子生物学研究进程, 还使茶树的分子育种研究进入一个“新时代”。以基因组学为基础的许多分子育种技术如数量性状基因座(quantitative trait locus, QTL)定位、转录组学研究、分子标记选择、全基因组关联分析(genome-wide association analysis, GWAS)等在茶树基因组发表之后得到了快速发展。鉴于茶树约3 Gb大小的基因组、超过一半的重复序列及高度杂合特性, 茶树基因组学研究进展较为缓慢。2017年, 我国研究者公布了基于二代测序技术获得的大叶种‘云抗10号’茶树基因组^[6], 之后陆续公布了9个茶树基因组, 其中‘舒茶早’、‘黄旦’和‘铁观音’等6个茶树品种的基因组达到参考基因组水平^[2, 5-12]。茶树基因组的相继公布加速了茶树

茶树生物学如分子标记辅助选择(marker assisted selection, MAS)、转录组学和全基因组关联分析等研究水平和进程。在茶树基因组公布之前, 茶树GWAS研究领域始终处于空白, 但是在茶树基因组公布后2 a, 以茶树表型性状和茶叶次生代谢物为基础的GWAS得以广泛开展。

GWAS是一种基于连锁不平衡(linkage disequilibrium, LD)原理的分析方法, 用于检测遗传变异和群体样本特征之间的联系, 挖掘与目标性状相关联的遗传位点, 为作物目标性状的改良提供基础, 为分子育种提供新的研究路径与思路^[13]。虽然GWAS在茶树多表型性状、茶树育种、茶叶品质成分和茶叶饮用等方面取得了一些研究进展, 但远滞后于玉米、大豆、水稻等常见作物。本文介绍了影响全基因组分析结果的主要因素, 总结了GWAS在茶叶饮料消费、茶树重要农艺性状、茶叶品质和茶树群体结构研究中取得的一系列进展, 以期为茶树遗传育种研究中利用GWAS提供理论依据。

1 全基因组关联分析

1.1 影响关联分析结果的主要因素

虽然GWAS在人类、动物和植物中已经被广泛

收稿日期: 2022-01-05

基金项目: 国家自然科学基金项目(31972460); 国家现代茶叶产业技术体系专项(CARS-19); 江苏省一带一路创新合作项目(BZ2019012); 贵州省科技计划项目(ZK[2021]154)

房婉萍, E-mail: fangwp@njau.edu.cn

通信作者: 马媛春, E-mail: myc@njau.edu.cn

应用,但是仍然有许多因素限制着GWAS的发展。GWAS研究检测到的关联位点一般很少,而且关联的位点仅能解释很少一部分性状变异。而未能检测到的遗传关联位点占有很大一部分比例^[14],丢失的部分遗传变异位点主要由以下3个因素导致。

1)结构变异。一般情况下,复杂性状的差异不是由一个位点的遗传变异引起,还可能受到结构变异的影响,如拷贝数变异(copy number variations, CNV)^[15]。最初统计方法的不完备和粗糙的基因分型方法等条件限制了GWAS研究,在许多GWAS研究中忽略了结构变异的功能。随着GWAS研究的不断发展和传播,众多GWAS研究开始关注拷贝数引起的变异。但是,研究主要集中在人类和动物研究中的CNV与性状关联性。但是,除水稻、玉米等模式植物中关于CNV的研究较多,有关CNV对植物性状变异的研究报道比较少。

2)上位效应和环境效应。据统计,大多数GWAS研究没有关注不同遗传变异位点之间的相互作用以及遗传位点和环境之间的相互作用^[16]。Zuk等^[17]认为不同基因之间的互作是造成关联位点丢失的一个重要原因。随着对GWAS研究的不断深入,Wei等^[18]提出了鉴定上位基因间互作的方法。但是,要通过几百万甚至是上亿次的显著性检验鉴定关联位点,对统计方法和计算机计算能力的选择甚为重要。目前,研究GWAS遗传变异位点的上位基因间的互作方法通常分为分为两步。首先,基于全基因组水平运用一种全局搜索方法对成对的互作基因进行快速搜索,结果为成对基因的1个子集;然后,鉴定子集内标记遗传位点之间的互作关系,最终得到具有显著性的互作位点^[18]。

复杂性状不仅受到基因位点之间互作的影响同时也会受到不同环境条件的影响,在GWAS研究中未能通过遗传相关性解释的关联位点可能是由于没有考虑基因和环境之间的互作。因此,需要结合样本大小、试验设计和基因型等众多因素来考虑基因与环境之间的互作对于GWAS研究的影响。在GWAS的相关研究中,虽然已经提出有关研究基因与环境之间互作的研究方法和实验方案,并取得一定的成果^[19],但是,已提出的基于上位效应和环境效应背景下的研究方法并不能很好地解决关联性位点丢失的问题,也无法有效地估计关联位点的 P 值^[20]。因此,未来需要进一步研究解决GWAS中出现基因之间和基因与环境之间互作丢失关联性位点的

问题。

3)稀有等位频率。GWAS鉴定关联性位点是以等位基因频率(allele frequency)为基础,低频率的等位基因遗传变异对研究群体的影响较小,但还有部分低频率的等位基因对性状有很大的影响^[21]。目前,与变异显著相关的遗传变异位点容易被识别,但是这些位点仅与少量表型性状相关联。然而,从遗传学和进化学的角度来看,绝大多数等位基因突变都是低频的,控制复杂性状的遗传变异关联位点一般也是低频的,即稀有变异(rare variant)^[22]。相关研究得出了稀有变异与复杂性状之间的关联^[23],也提出了鉴定稀有变异的研究方法^[24]。

1.2 关联分析的优势与不足

QTL和GWAS都是基于群体的遗传图谱来研究群体间个体的复杂性状,二者有许多异同点。数量性状位点(QTL)定位的连锁分析是GWAS等关联研究的直接先驱,这就意味着GWAS比QTL研究结果更为精确。连锁图谱不是将个体组合成一个不同的GWAS面板,而是研究具有已知关系的个体^[25]。例如,对作物的连锁定位分析通常是使用双亲本杂交产生的后代,无论是 F_2 个体还是重组自交系(RILs)。由于QTL研究的个体在家系中亲缘关系较近,发生的重组较少,因此,存在较大的连锁块,意味着所使用的遗传标记不必像GWAS中使用的遗传标记那样密集,以确保检测到含有研究位点的基因组区域,同时也意味着QTL对群体选择要比GWAS苛刻。一旦发现并验证了一个QTL,该区域就可以被用于精细定位和QTL克隆。在下一代测序技术出现之前,这是1个相当大的优势。1988年,在番茄中进行了第1次全基因组QTL分析^[26],比第1个GWAS发表早了14 a。发展到今天,连锁分析和GWAS互补,可以用来解析不同群体中的复杂性状。

尽管连锁分析是解析复杂性状的主要方法,但关联研究的潜在价值也受到了重视。与连锁分析相比,关联分析的优势包括不需要进行杂交实验^[27],增加了检测效应较小基因的能力,并提高较小LD块的分辨率。然而,对高密度基因型的需求,意味着在三代测序发展之前进行的第1个关联研究只能关注于基因组的小子集,只能研究已经被其他方法确定的区域。例如,1项关于玉米株高和开花时间的早期关联研究集中在之前研究已经确定的候选基因*dwarf8*上,检测到*dwarf8*及其附近的123个多态性位点以及141个全基因组标记^[28]。但是,随着测序技术的

进步,全基因组序列的公布显著加快了GWAS的研究进程。

GWAS一个明显的缺陷是会出现较高的假阳性,即把不与性状相关联的位点划分到关联位点内,使结果中包含有与研究目的不相关的位点。群体结构和个体间的亲缘关系都会导致高假阳性率,降低假阳性的方法包括调整多重测试校正的限制错误发现率、调整每个标记的 P 值以及使用合适的结构化关联方法^[29]。通用线性模型(*generalized linear model*, GLM)和混合线性模型(*mixed linear model*, MLM)是最基础的结构化关联方法,混合线性模型可以通过亲缘关系和群体结构分析的结果减少假阳性。为了提高混合线性模型的运算速率,研究者开发了混合模型关联(EMMA)、分解谱变换线性混合模型(*factored spectrally transformed linear mixed models*, FaSTLMM)、全基因组高效混合模型分析(*genome-wide efficient mixed model association*, GEMMA)、压缩混合线性模型(*compressed mixed linear model*, CMLM)和富集压缩混合线性模型(*enriched compressed mixed linear model*, ECMLM)。虽然上述混合线性模型提高了MLM的运行速度,但是增加了运行的负荷,同时也可能增加结果的假阴性,所以要根据实际情况选择低假阳性、低假阴性和高运算速度的最佳模型^[30-31]。

2 GWAS在茶叶饮用人群特征中的应用

茶作为三大无酒精饮料之一,具有良好的养生保健功能,其饮用价值一直是研究的热点。对于茶叶消费习惯的GWAS研究是社会研究的热点。虽然有较多关于茶叶消费习惯的GWAS研究,但是并没有鉴定出与茶叶饮用习惯相关的遗传位点。Taylor等^[32]、Cornelis^[33]和Zhong等^[34]都对酒精和非酒精饮料(咖啡、茶)的消费习惯与基因型数据进行了GWAS研究,虽然能够鉴定出与酒或咖啡消费相关联的遗传位点,但是并未鉴定出与茶叶消费习惯相关的遗传位点,可能是调查的人群数量过少或是人群过于集中不具有代表性。然而,仍有部分研究鉴定出与饮茶习惯相关的遗传位点。例如,Furukawa等^[35]和Matoba等^[36]分别收集了12 258和165 084个日本参与者的茶叶消费习惯和基因型数据,对茶叶消费习惯和基因型数据进行GWAS分析,都鉴定出12q24这一遗传变异位点与日本人群的茶叶饮用习

惯相关,并鉴定出调控饮茶习惯相关的基因 $ALDH2$ 。另外,Cole等^[37]对449 210名个体进行问卷调查,研究143个显著遗传的饮食习惯,共鉴定出与饮食习惯相关联的814个SNPs位点,其中rs1453548位点与饮茶习惯相关,并在位点附近鉴定出6个与饮茶相关的基因。通过GWAS鉴定出与茶叶饮用习惯相关的遗传位点和调控基因,有利于加深对人类饮茶习惯基因层面的了解,并能发掘出人类基因新功能。

人类健康问题一直是研究者关注的重心,由于茶有其独特的养生保健功能,所以也一直成为医药和营养研究关注的热点。Malik等^[38]研究性别、吸烟、饮用盐茶与食管癌之间的全基因组关联分析,发现与食管癌相关的遗传变异位点,鉴定出PLCEL(rs2274223、rs7922612和rs3765524位点)基因型与吸烟和盐茶摄入以及患食管癌(EC)风险之间的关联。这些饮用盐茶与食管癌的关联位点和调控基因,为医学上研究饮用盐茶和食管癌之间的关系提供了基因层面的证据,并为进一步了解饮茶和疾病之间的关系提供了新的研究方向。

3 GWAS在茶树生物学研究中的应用

3.1 GWAS在茶树表型性状研究中的应用

1)GWAS在茶树树型和叶片表型研究中的应用。茶树主要生长在热带、亚热带,不同地区、不同品种的茶树树型和叶片表型差异很大,即使是生长在同一地区的不同茶树品种的叶片表型差异也较大。茶树的树型主要分为灌木、小乔木和乔木,树姿主要分为直立、半开张和开张,不同地区、不同品种的茶树树型和树姿有明显的差异,因此,树型和树姿可作为茶树分类和品种选育的重要依据。Lu等^[39]测定了云贵高原古茶树的表型和遗传数据,将茶树树型、叶形、叶色、叶身等表型数据与遗传数据进行全基因组关联分析,揭示了云贵高原古茶树的多样性和进化、驯化机制,系统进化分析结果与主成分分析(PCA)结果一致,将120株古茶树主要分为3类和5个单分支;遗传结构分析进一步将古茶树分为7个亚群;研究发现,由于外部自然环境或人工育种的压力,古茶树的变异(非同义/同义=1.05)并没有减少;最后,通过整合GWAS、选择信号和基因功能预测结果,鉴定出4个候选基因与3个叶片性状显著相关,2个候选基因与植物株型显著相关,这些候选基因未来可用于茶树的功能鉴定和遗传改良。

茶树叶片不仅是重要的营养器官也是茶叶产品的原材料,具有重要的经济价值。茶叶面积的大小对茶叶产量具有重要影响,以黑茶、乌龙茶、红茶以及抹茶尤为突出^[40]。根据茶叶叶面积的大小将茶树分为大叶种和小叶种,茶叶叶面积大小与茶树树型和地理分布密切相关。目前,对于调控叶面积的关键基因和遗传机制的研究相对较少。Tan等^[41]以‘龙井43’和‘白毫早’作为亲本,对其170个后代构建了基于SSR标记的连锁图谱,对成熟茶叶叶面积进行了初步QTL定位,鉴定出1个与茶叶面积显著相关的QTL位点;但是与SNPs连锁图谱相比,构建的连锁图谱准确性较低。An等^[40]以‘金萱’和‘云茶1号’作为亲本,对其96个后代进行重测序,构建了1个包含15个连锁群的高密度遗传图谱,其中包括8 956个高质量SNPs,共鉴定出25个与叶面积相关的代表性标记(潜在的QTL)和2个与茶树叶面积相关的SNP,这些数据 and 图谱将进一步研究茶树性状分离、功能基因定位和标记辅助育种提供强有力的支持。

茶叶叶片形状具有较强的遗传多样性,不同品种的茶树叶片形状差异大,即使是同一品种的茶树其叶片形状也会有所不同。茶叶叶片特性是决定茶叶品质的重要因素,是育种选择的重要指标。Zhang等^[42]对贵州黔南都匀毛尖8个产区中的123株茶树进行特定位点扩增片段(the specific-locus amplified fragment, SLAF)测序,将测序数据与茶树叶尖和叶形等表型数据进行全基因组关联分析,结果表明,群体结构分析将123株茶树划分为3个群体,关联分析共鉴定出11个与叶尖相关的候选基因和7个与叶片形状相关的候选基因。这些候选基因更新了对于茶树叶片表型调控基因的认识,未来可用于茶树叶片表型的分子生物学研究和茶树遗传育种研究。

2)GWAS在茶树春芽萌发时间研究中的应用。茶树春梢是茶叶生产加工的直接原材料,叶片是进行光合作用制造有机营养物质的器官,其着生于新梢^[43],充分了解新梢生长发育的特点,对于茶树生长规律的深入了解具有极其重要的意义。尤其是对于早春温度相对较低的地区,早春芽萌发(spring bud flush, TBF)的茶树在第一轮收获中可以避免一些病虫害发生,并积累大量的有机天然化合物,确保茶的最佳品质,因此,了解茶树春芽萌发时间的遗传背景和多样性对于选育早春芽萌发的茶树良种具有重要意义。Wang等^[44]以151个茶树种质资源为研究对

象,利用确定位点扩增片段测序(specific-locus amplified fragment sequencing, SLAF-seq)进行GWAS分析,鉴定大量的SNP变异位点以及与茶树TBF相关的候选基因。3 a中GWAS分析检测到26个与TBF相关的SNP,最终确定了1个与茶树TBF显著相关的SNP位点。为了鉴定可能与TBF相关的候选基因,在与TBF密切相关的SNP位点100 kb区域范围内筛选了7个候选基因。此外,还在SNP位点中发现了有利的等位基因变异的“TT”基因型,设计与TBF共分离的裂切扩增多态性(derived cleaved amplified polymorphism, dCAPS)标记,此标记可用于茶树标记辅助选择育种。鉴定出的7个候选基因与调控春芽萌发密切相关,将为之后春芽萌发的调控机制研究奠定基础。

3.2 GWAS在茶树品质性状研究中的应用

茶叶品质一直是茶叶研究的重点。茶叶的品质由茶叶中多种生化成分共同决定,其中对茶叶品质影响显著的主要是次生代谢物,如儿茶素、茶多酚和咖啡碱等。Yamashita等^[45]评估了基因组预测(genomic prediction, GPs)和GWAS对茶品质相关代谢物遗传育种的潜力,使用SNP检测来自150个茶树种质的DNA测序结果的限制性相关位点。GP有效地预测了与EC、ECG、EGCG、总儿茶素和咖啡碱含量相关的遗传位点,但没有预测到与游离氨基酸(FAAs)含量相关的位点。采用genomic best linear unbiased prediction (GBLUP) with linear ridge kernel regression (RR)或GBLUP with non-linear Gaussian kernel regression (GAUSS)、Ridge、Lasso、Elastic Net和随机森林模型等6种分析模型对GP进行评估,结果显示这些预测值是稳定的。此外,通过GWAS和GP的综合分析可以鉴定与预测代谢物相关的潜在候选基因,在150份材料中,每个代谢物通过GWAS分析鉴定出80~160个SNPs,最终鉴定到13个与EGCG和咖啡碱含量相关的常见候选基因。这些候选基因因为研究茶叶品质成分相关的基因奠定了基础,有利于培育出高品质的茶树新品种。

目前很少同时将茶树表型性状、茶叶品质成分与茶树进行全基因组关联分析,但Hazra等^[46]将23株大吉岭茶树种质与农艺性状和风味相关成分进行全基因组关联分析,最终检测到57个SNPs标记位点,其中12个SNPs与表没食子儿茶素没食子酸酯(EGCG)密切相关,8个SNPs与茶叶风味密切相关,3个SNPs与酚类物质密切相关,8个SNPs与活性氧

清除密切相关,6个SNPs与气孔指数密切相关,5个SNPs与单宁密切相关,7个SNPs与产量密切相关。这些SNPs位点的鉴定为茶树表型性状和品质调控相关研究奠定了基础,有利于推动茶树分子育种的发展。

由于茶氨酸、咖啡碱和儿茶素对茶品质及其生理功能有重要影响,其合成和调控相关基因的鉴定对于了解茶叶品质成分的合成和调控途径、培育高品质的茶树具有重要意义。Fang等^[47]对289株茶树在3个季节的茶氨酸、咖啡碱、儿茶素进行全基因组关联分析,基于SNP数据将茶树分为2个群体,群体1(191株)鉴定到307个与茶氨酸、咖啡碱、儿茶素相关的SNP标记,通过群体2(98株)验证候选的30个SNP标记,其中有17个与特定的次生代谢物显著或极显著相关,这些标记位点的确定将为茶叶风味相关代谢物的研究提供基础,并有助于加快茶树新品种的培育。

游离氨基酸(尤其是茶氨酸)是茶叶中最主要的化学成分之一,有助于提高茶的口感、功能和质量。人工杂交能够有效选育出高氨基酸含量且综合性状优良的茶树新种质。Huang等^[48]以‘龙井43’和‘白鸡冠’作为亲本,基于简化基因组测序(genotyping-by-sequencing,GBS)技术对其198个杂交F₁植株的游离氨基酸含量进行GWAS分析,发掘出2688个多态性SNP标记,构建了具有15个连锁群(linkage groups,LGs)的高密度图谱,共鉴定出4个与游离氨基酸(茶氨酸、谷氨酸、谷氨酰胺、天冬氨酸和精氨酸)含量相关的QTL,分别定位到LG03、LG06、LG11和LG14。该研究对于阐明茶树优异种质资源高氨基酸性状的遗传机制以及优质茶树品种选育和改善茶叶品质具有重要意义。

4 总结与展望

传统的基于表型评价的育种方法存在茶树世代长、表型测量准确性差,无法在茶树达到生理成熟之前对茶叶品质和产量进行评估等问题。然而,传统的茶树育种方法仍处于主导地位^[49]。快速且高效的育种方法对于推动茶产业的发展和满足茶叶市场的需求具有重要意义。基因组学的发展能够推动植物育种进程,提高植物的育种效率,如分子标记辅助育种技术能够提高植物的育种效率,缩短木本植物的育种进程。从育种群体的种子或幼苗中获得的基因型数据可用于预测成熟个体的表型性状,而不需要

在不同的年份和环境中进行广泛的表型评估。由于茶树中的大多数性状都是复杂性状,目前有关茶树GWAS研究检测到的基因位点只能解释一小部分表型变异(远小于性状的遗传力),所以未来需要对茶树复杂性状进行更深入的GWAS研究,尽可能减少由于上位效应和环境效应造成遗传位点丢失的现象发生,从而能够检测到更为有效的基因位点。

成功的GWAS需要来自群体的全面而又准确的基因组和表型数据。虽然有多种高通量DNA测序方法可用于生成基因组数据,但生成高质量表型数据的技术却远远落后于基因组测序技术。传统的植物表型测量方法大多依赖于人工,具有费力、耗时及不准确等缺点,不利于从群体中快速而准确地获取表型数据。高通量表型技术克服了传统方法的上述缺点,已成为评价植物表型的有力工具^[50]。高通量表型技术,如可见光成像、高光谱成像和荧光成像,已成功应用于评估植物生长、生物量和营养状况,主要用于测量细胞、种子、芽、叶、根、单个植物和冠层等植物表型^[51-52]。高质量表型数据测量技术已经广泛运用在水稻、玉米、小麦、大豆等农作物上,其表型主要包括株高、粒形、芽长、分蘖数等植物常见表型和植物病害等非生物胁迫^[53-54]。茶树叶身、叶面、叶缘、叶锯齿等表型数据和病虫害测量需要克服人工测量精准度差、主观意识强等缺陷。所以未来对茶树表型数据的测量可以应用高通量表型测量技术检测出更多、更有效的与茶树表型调控相关的遗传位点和基因。

生物/非生物胁迫是影响植物产量、品质和生长发育的主要因素,水稻、玉米、小麦等禾本科植物和梨、桃、苹果、茶树等木本植物都将生物/非生物胁迫作为研究的热点。基于病虫害的GWAS最初是在拟南芥、水稻、小麦等模式作物中开展,因木本植物基因组问世比较晚,木本植物的GWAS也进展缓慢。随着木本植物基因组的公布,对木本植物基于生物/非生物胁迫的GWAS也在不断发展,如在苹果、梨、橙子和杏等木本植物中开展了基于苹果斑枯病、白粉病、病原菌抗性和低温胁迫等的GWAS^[55-56],鉴定出一系列与植物抗性相关的基因,有利于今后选育抗性强的优良品种。目前,几乎没有关于茶树抗性鉴定的GWAS研究,比较茶树在不同胁迫下的抗性指标及不同茶树品种在胁迫下抗逆性指标变化的差异仍是茶树抗性鉴定的重要手段。基于全基因组的QTL和SNP关联作图,可同时多个等位基因进行

标记分析,能够准确鉴定出与表型和抗性性状相关的许多遗传变异位点和目的基因,发现参与抗性调控的新遗传变异位点和新基因。总的来说,由于GWAS能够为茶树育种提供分子标记,可以加快茶树抗性基因鉴定的研究进程,并最终提高茶树抗性品种的育种效率,基于茶树抗性的GWAS将会逐渐成为茶树育种研究的热点。

一般来说,表型性状的变异有限,而大量代谢物的含量存在巨大的差异。GWAS鉴定的作物表型性状QTL往往具有中或低等效应^[57],部分原因可能是基于目标表型的有限变异。调控代谢物积累的遗传变异更容易与全基因组进行关联分析,mGWAS (metabolome-based genome-wide association study) 在水稻、玉米、番茄等作物中已经被广泛使用,鉴定出与独角金内酯、类黄酮等代谢物相关的调控基因^[23,58-59]。此外,结合转录组、代谢组和基因组的多组学关联分析已经在拟南芥、番茄等植物中普遍应用,鉴定出与植物非生物胁迫和调控类黄酮合成相关的基因^[59-60]。茶树多组学GWAS将会成为今后研究的热点。

参考文献 References

- [1] 叶乃兴. 茶学研究法[M]. 北京: 中国农业出版社, 2011. YE N X. Research methods of tea science [M]. Beijing: China Agriculture Press, 2011 (in Chinese).
- [2] ZHANG X T, CHEN S, SHI L Q, et al. Haplotype-resolved genome assembly provides insights into evolutionary history of the tea plant *Camellia sinensis* [J/OL]. *Nature genetics*, 2021, 53 (6): 1250 [2022-03-07]. <https://doi.org/10.1038/s41588-021-00895-y>.
- [3] LIN X H, SUN D W. Recent developments in vibrational spectroscopic techniques for tea quality and safety analyses [J]. *Trends in food science & technology*, 2020, 104: 163-176.
- [4] ZHANG L, HO C T, ZHOU J, et al. Chemistry and biological activities of processed *Camellia sinensis* teas: a comprehensive review [J]. *Comprehensive reviews in food science and food safety*, 2019, 18(5): 1474-1495.
- [5] XIA E H, TONG W, WU Q, et al. Tea plant genomics: achievements, challenges and perspectives [J/OL]. *Horticulture research*, 2020, 7 (1): 7 [2022-03-07]. <https://doi.org/10.1038/s41438-019-0225-4>.
- [6] XIA E H, ZHANG H B, SHENG J, et al. The tea plant genome provides insights into tea flavor and independent evolution of caffeine biosynthesis [J]. *Molecular plant*, 2017, 10(6): 866-877.
- [7] CHEN J D, ZHENG C M, JIAN Q J, et al. The chromosome-scale genome reveals the evolution and diversification after the recent tetraploidization event in tea plant [J]. *Horticulture research*, 2020, 7 (1): 63-73.
- [8] XIA E H, TONG W, HOU Y, et al. The reference genome of tea plant and resequencing of 81 diverse accessions provide insights into its genome evolution and adaptation [J]. *Molecular plant*, 2020, 13(7): 1013-1026.
- [9] ZHANG Q J, LI W, LI K, et al. The chromosome-level reference genome of tea tree unveils recent bursts of non-autonomous LTR retrotransposons in driving genome size evolution [J]. *Molecular plant*, 2020, 13(7): 935-938.
- [10] WANG X C, FENG H, CHANG Y X, et al. Population sequencing enhances understanding of tea plant evolution [J/OL]. *Nature communications*, 2020, 11 (1): 4447 [2022-03-07]. <https://doi.org/10.1038/s41467-020-18228-8>.
- [11] ZHANG W Y, ZHANG Y J, QIU H J, et al. Genome assembly of wild tea tree *dazl* reveals pedigree and selection history of tea varieties [J/OL]. *Nature communications*, 2020, 11 (1): 3719 [2022-03-07]. <https://doi.org/10.1038/s41467-020-17498-6>.
- [12] WANG P J, YU J X, JIN S, et al. Genetic basis of high aroma and stress tolerance in the oolong tea cultivar genome [J/OL]. *Horticulture research*, 2021, 8 (1): 107 [2022-03-07]. <https://doi.org/10.1038/s41438-021-00542-x>.
- [13] ERSOZ E S, YU H M, BUCKLER E S. Applications of linkage disequilibrium and association mapping in crop plants [M]. Dordrecht: Springer, 2008: 97-119.
- [14] 赵振卿, 顾宏辉, 盛小光, 等. 作物数量性状位点研究进展及其育种应用 [J]. *核农学报*, 2014, 28(9): 1615-1624. ZHAO Z Q, GU H H, SHENG X G, et al. Advances and applications in crop quantitative trait locus [J]. *Journal of nuclear agricultural sciences*, 2014, 28(9): 1615-1624 (in Chinese with English abstract).
- [15] FREEMAN J L, PERRY G H, FEUK L, et al. Copy number variation: new insights in genome diversity [J]. *Genome research*, 2006, 16(8): 949-961.
- [16] KORTE A, FARLOW A. The advantages and limitations of trait analysis with GWAS: a review [J]. *Plant methods*, 2013, 9 (4): 749-764.
- [17] ZUK O, HECHTER E, SUNYAEV S R, et al. The mystery of missing heritability: genetic interactions create phantom heritability [J]. *PNAS*, 2012, 109(4): 1193-1198.
- [18] WEI W H, HEMANI G, HALEY C S. Detecting epistasis in human complex traits [J]. *Nature reviews genetics*, 2014, 15 (11): 722-733.
- [19] WINHAM S J, BIERNACKA J M. 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.
- [20] THOMAS D. Gene-environment interactions in human diseases [J]. *Nature reviews genetics*, 2005, 6 (4): 287-298.
- [21] MYLES S, PEIFFER J, BROWN P J, et al. Association mapping: critical considerations shift from genotyping to experimental design [J]. *The plant cell*, 2009, 21(8): 2194-2202.
- [22] GIBSON G. Rare and common variants: twenty arguments [J]. *Nature reviews genetics*, 2012, 13 (2): 135-145.
- [23] HAZELETT D J, COETZEE S G, COETZEE G A. A rare variant, which destroys a FoxA1 site at 8q24, is associated with prostate cancer risk [J]. *Cell cycle*, 2013, 12(2): 379-380.
- [24] GOURAB D, YIP W, IULIANA I, et al. Rare variant analysis for family-based design [J/OL]. *PLoS One*, 2017, 8 (1): e48495

- [2022-03-07].<https://doi.org/10.1371/journal.pone.0048495>.
- [25] CORTES L T, ZHANG Z W, YU J M. Status and prospects of genome-wide association studies in plants[J/OL]. *Plant genome*, 2021, 14 (1): e20077 [2022-03-07]. <https://doi.org/10.1002/tpg2.20077>.
- [26] PATERSON A H, LANDER E S, HEWITT J D, et al. Resolution of quantitative traits into mendelian factors by using a complete linkage map of restriction fragment length polymorphisms[J]. *Nature*, 1988, 335 (6192): 721-726.
- [27] LANDER E, KRUGLYAK L. Genetic dissection of complex traits - guidelines for interpreting and reporting linkage results[J]. *Nature genetics*, 1995, 11 (3): 241-247.
- [28] THORNSBERRY J M, GOODMAN M M, DOEBLEY J, et al. Dwarf8 polymorphisms associate with variation in flowering time[J]. *Nature genetics*, 2001, 28 (3): 286-289.
- [29] YU J M, PRESSOIR G, BRIGGS W H, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness[J]. *Nature genetics*, 2006, 38 (2): 203-208.
- [30] KLASEN J R, BARBEZ E, MEIER L, et al. A multi-marker association method for genome-wide association studies without the need for population structure correction[J/OL]. *Nature communications*, 2016, 7: 13299 [2022-03-07]. <https://doi.org/10.1038/ncomms13299>.
- [31] KANG H M, ZAITLEN N A, WADE C M, et al. Efficient control of population structure in model organism association mapping[J]. *Genetics*, 2008, 178(3): 1709-1723.
- [32] TAYLOR A E, DAVEY SMITH G, MUNAFÒ M R. Associations of coffee genetic risk scores with consumption of coffee, tea and other beverages in the UK Biobank[J]. *Addiction*, 2018, 113 (1): 148-157.
- [33] CORNELIS M C. Genetic determinants of beverage consumption; implications for nutrition and health[J]. *Advances in food and nutrition research*, 2019, 89: 1-52.
- [34] ZHONG V W, KUANG A L, DANNING R D, et al. A genome-wide association study of bitter and sweet beverage consumption[J]. *Human molecular genetics*, 2019, 28(14): 2449-2457.
- [35] FURUKAWA K, IGARASHI M, JIA H, et al. A genome-wide association study identifies the association between the 12q24 locus and black tea consumption in Japanese populations[J/OL]. *Nutrients*, 2020, 12 (10): 3182 [2022-03-07]. <https://doi.org/10.3390/nu12103182>.
- [36] MATOBA N, AKIYAMA M, ISHIGAKI K, et al. GWAS of 165,084 Japanese individuals identified nine loci associated with dietary habits[J]. *Nature human behaviour*, 2020, 4 (3): 308-316.
- [37] COLE J B, FLOREZ J C, HIRSCHHORN J N. Comprehensive genomic analysis of dietary habits in UK Biobank identifies hundreds of genetic associations[J/OL]. *Nature communications*, 2020, 11 (1): 1467 [2022-03-07]. <https://doi.org/10.1038/s41467-020-15193-0>.
- [38] MALIK M A, UMAR M, GUPTA U, et al. Phospholipase C *Epsilon* 1 (PLCE1 rs2274223A>G, rs3765524C>T and rs7922612C>T) polymorphisms and esophageal cancer risk in the Kashmir Valley[J]. *Asian Pacific journal of cancer prevention: APJCP*, 2014, 15(10): 4319-4323.
- [39] LU L T, CHEN H F, WANG X J, et al. genome-level diversification of eight ancient tea populations in the guizhou and yunnan regions identifies candidate genes for core agronomic traits[J/OL]. *Horticulture research*, 2021, 8 (1): 190 [2022-03-07]. <https://doi.org/10.1038/s41438-021-00617-9>.
- [40] AN Y L, MI X Z, ZHAO S Q, et al. Revealing distinctions in genetic diversity and adaptive evolution between two varieties of *Camellia sinensis* by whole-genome resequencing[J/OL]. *Frontiers in plant science*, 2020, 11: 603819 [2022-03-07]. <https://doi.org/10.3389/fpls.2020.603819>.
- [41] TAN L Q, WANG L Y, XU L Y, et al. SSR-based genetic mapping and QTL analysis for timing of spring bud flush, young shoot color, and mature leaf size in tea plant (*Camellia sinensis*) [J]. *Tree genetics & genomes*, 2016, 12(3): 1-13.
- [42] ZHANG F, TIAN W L, CEN L, et al. Population structure analysis and genome-wide association study of tea (*Camellia Sinensis* (L.) Kuntze) germplasm in Qiannan, China, based on SLAF-Seq technology[J]. *Phyton-international journal of experimental botany*, 2022, 91 (4): 791-809.
- [43] 史春彦, 张前东, 张晓平, 等. 济南市长清区茶树种植适宜性农业区划[J]. *山东农业科学*, 2016, 48(10): 81-85. SHI C Y, ZHANG Q D, ZHANG X P, et al. Suitable agricultural regionalization for tea planting in Changqing district of Ji'nan City[J]. *Shandong agricultural sciences*, 2016, 48(10): 81-85 (in Chinese with English abstract).
- [44] WANG R J, GAO X F, YANG J, et al. Genome-wide association study to identify favorable SNP allelic variations and candidate genes that control the timing of spring bud flush of tea (*Camellia sinensis*) using SLAF-seq[J]. *Journal of agricultural and food chemistry*, 2019, 67(37): 10380-10391.
- [45] YAMASHITA H, UCHIDA T, TANAKA Y, et al. Genomic predictions and genome-wide association studies based on RAD-seq of quality-related metabolites for the genomics-assisted breeding of tea plants[J]. *Scientific reports*, 2020, 10 (1): 17480 [2022-03-07]. <https://doi.org/10.1038/s41598-020-74623-7>.
- [46] HAZRA A, KUMAR R, SENGUPTA C, et al. Genome-wide SNP discovery from Darjeeling tea cultivars - their functional impacts and application toward population structure and trait associations[J]. *Genomics*, 2021, 113(1): 66-78.
- [47] FANG K X, XIA Z Q, LI H J, et al. Genome-wide association analysis identified molecular markers associated with important tea flavor-related metabolites [J/OL]. *Horticulture research*, 2021, 8 (1): 42 [2022-03-07]. <https://doi.org/10.1038/s41438-021-00477-3>.
- [48] HUANG R, WANG J Y, YAO M Z, et al. Qe trait loci mapping for free amino acid content using an albino population and SuantativNP markers provides insight into the genetic improvement of tea plants[J/OL]. *Horticulture research*, 2022, 9: uhab029 [2022-03-07]. <https://doi.org/10.1093/hr/uhab029>.
- [49] IWATA H, MINAMIKAWA M F, KAJIYA-KANEGAE H, et al. Genomics-assisted breeding in fruit trees[J]. *Breeding science*, 2016, 66(1): 100-115.
- [50] XIAO Q L, BAI X L, ZHANG C, et al. Advanced high-throughput plant phenotyping techniques for genome-wide association studies: a review [J]. *Journal of advanced research*, 2022, 35: 215-230.
- [51] YANG W N, FENG H, ZHANG X H, et al. Crop phenomics and high-throughput phenotyping: past decades, current challenge

- es, and future perspectives [J]. *Molecular plant*, 2020, 13 (2) : 187-214.
- [52] JIANG L B, SUN L D, YE M X, et al. Functional mapping of N deficiency-induced response in wheat yield-component traits by implementing high-throughput phenotyping [J]. *The plant journal*, 2019, 97(6) : 1105-1119.
- [53] RASHEED A, XIA X C, OGBONNAYA F, et al. Genome-wide association for grain morphology in synthetic hexaploid wheats using digital imaging analysis [J/OL]. *BMC plant biology*, 2014, 14 : 128 [2022-03-07]. <https://doi.org/10.1186/1471-2229-14-128>.
- [54] SHAKOOR N, LEE S, MOCKLER T C. High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field [J]. *Current opinion in plant biology*, 2017, 38 : 184-192.
- [55] NOH J, DO Y S, KIM G H, et al. A genome-wide association study for the detection of genes related to apple Marssonina Blotch disease resistance in apples [J/OL]. *Scientia horticulturae*, 2020, 262 : 108986 [2022-03-07]. <https://doi.org/10.1016/j.scienta.2019.108986>.
- [56] PENG Z, BREDESON J V, WU G A, et al. A chromosome-scale reference genome of trifoliolate orange (*Poncirus trifoliata*) provides insights into disease resistance, cold tolerance and genome evolution in *Citrus* [J]. *The plant journal*, 2020, 104 (5) : 1215-1232.
- [57] HUANG X, ZHAO Y, WEI X, et al. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm [J]. *Nature genetics*, 2012, 44 : 32-39.
- [58] CHEN W, WANG W S, PENG M, et al. Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals [J/OL]. *Nature communications*, 2016, 7 : 12767 [2022-03-07]. <https://doi.org/10.1038/ncomms12767>.
- [59] ZHU G T, WANG S C, HUANG Z J, et al. Rewiring of the fruit metabolome in tomato breeding [J]. *Cell*, 2018, 172 (1/2) : 249-261.
- [60] LEE T, KIM H, LEE I. Network-assisted crop systems genetics: network inference and integrative analysis [J]. *Current opinion in plant biology*, 2015, 24 : 61-70.

Application of genome-wide association analysis in studying tea

FANG Wanping, LEI Xiaogang, YANG Bin, WANG Ya, MA Yuanchun

College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China

Abstract Genome-wide association analysis (GWAS) is based on high-throughput sequencing technology, combined with bioinformatics and statistical methods, to identify genetic variants that regulate complex traits at the genome-wide level. It is the most powerful and effective method to study the genetic variation of complex agronomic traits and diseases, and its core is to study the association between genetic variation and target traits. This review briefly introduces the basic principles, processes and influencing factors of GWAS analysis. A series of progress made by GWAS in studying the tea beverage consumption, important agronomic traits of tea trees, tea quality and the population structure of tea plant were summarized. Finally, the problems encountered and the direction of development in the GWAS research of tea plant were discussed. It will provide a basis for using GWAS in further researches on the selection of tea consumers, the genetics and breeding of tea plant.

Keywords tea plant; genome-wide association analysis (GWAS); agronomic traits; tea consumption; quality; breeding

(责任编辑:张志钰)