

综 述 Reviews

全基因组关联分析(GWAS)在林木育种中的应用

阙青敏, 欧阳昆唏, 李培, 陈晓阳*

华南农业大学林学与风景园林学院, 广东省森林植物种质创新与利用重点实验室, 广州510642

摘要: 全基因组关联分析(genome-wide association studies, GWAS)是近几年发展起来的解析人类、动物或植物表型多样性遗传基础的有效分析手段。GWAS具有高通量、高精度和高速度等显著优点, 其在林木遗传育种中的作用日益凸显。本文从种质材料、目标性状的选择和表型鉴定、全基因组SNP位点的获取、群体结构、连锁不平衡分析以及关联作图与候选基因发掘等方面对GWAS在林木育种中的应用进行了综述, 并提出后续研究展望。以期为进一步利用GWAS技术进行林木育种中各种性状遗传基础的研究提供参考。

关键词: 全基因组关联分析(GWAS); 连锁不平衡; 林木育种

林木不仅是重要的原材料, 而且是陆地生态系统最重要的组成部分。同时, 林木还承担着水源涵养、气候调节、生物多样性保护、碳沉积等重要的生态服务功能, 并且在塑造人类文化的多样性方面做出了一定的贡献(Jiang等2001)。由于全球人口数量的增加, 对木材的需求急剧上升, 导致全球森林退化速度加快。全球需要恢复的森林面积多达20亿公顷, 并且还在不断增加(Stanturf等2014), 森林生态系统亟需保护。我国已经实施了多项天然林保护工程, 此项措施对保护生物多样性及改善生态环境具有重要意义。天然林保护工程实施之后效果显著, 我国的植被增加量占2000至2017年间全球植被总增量的25%以上(Chen等2019)。但天然林保护工程的实施不可避免地加剧我国木材供求矛盾, 而林木育种为解决木材供求矛盾提供了出路。林木育种是通过遗传学的方法对林木的自然生产潜力进行充分的利用, 以期提高林产品的产量和质量, 充分发挥林木的经济效益和生态效益。然而, 由于林木的生长周期长、基因组较大和遗传评价困难等特点, 与玉米(*Zea mays*)、水稻(*Oryza sativa*)等其他农作物的高速育种进程相比, 林木育种研究工作进展缓慢(田斌等2013)。

林木的生长性状、材性性状以及抗逆性状决定着林木的经济价值, 是林木育种工作者主要关注的对象。林木的生长量、材性以及抗逆性等数量性状是微效多基因联合作用的结果, 且同时受环境以及基因与环境互作的影响(Zinkgraf等2017)。与质量性状相比, 数量性状的遗传基础更为复杂。

近年来, 已有众多研究表明, 全基因组关联分析(genome-wide association studies, GWAS)对揭示林木产量、品质和抗性性状的遗传基础、挖掘重要的主效基因, 从育种技术上快速改良并选育优良品种具有重要意义(Du等2018; Wisniewski等2018)。

GWAS是由Risch和Merikan-gas (1996)在研究复杂疾病的遗传学问题时提出的, 并于2005年被Klein等人在视网膜黄斑变性的研究中所证实该方法切实可行, 引起遗传学界的轰动。GWAS是一种对研究对象全基因组范围内的单核苷酸多态性(single nucleotide polymorphism, SNP)及其拷贝数进行总体关联分析的方法, 所研究对象为自然群体, 理论基础是生物进化过程中因长期重组而存在基因(位点)间连锁不平衡(Hirschhorn和Daly 2005)。GWAS的基本方法是先通过检测研究对象全基因组的SNP标记并进行分型, 然后统计目标性状的变异, 最后基于SNP位点与目标性状的连锁不平衡衰减关系, 通过统计学方法, 精确鉴定出与表型变异密切相关的具有特定功能的SNP位点, 从而获得与目标性状相关联的候选基因或基因区域(Gajardo等2015)。GWAS不仅成功应用在水稻(Chen等2014; Huang等2010; Huang等2012)、玉米(Farfan等2015; Tian等2011; 邵晓宇等2017)、油菜

收稿 2019-08-30 修定 2019-10-26

资助 广东省科技计划项目(2017B020201008)、国家自然科学基金(31600525)、中央财政林业科技推广示范项目[(2018)GDTK-08]。

* 通讯作者(xychen@scau.edu.cn)。

(*Brassica campestris*) (Li等2016)、大麦(*Hordeum vulgare*) (Alqudah等2014)、小麦(*Triticum aestivum*) (Zhou等2017)等重要农作物中,在果树育种中也得到了很好的应用(Iwata等2016)。随着基因分型和测序技术的不断发展, GWAS已经成为研究自然表型变异遗传机制的常用方法,在林木育种中的应用也日益广泛(Du等2018)。本文从种质材料与目标性状的选择及表型鉴定、全基因组SNP位点的获取、群体结构和连锁不平衡分析以及关联作图与候选基因发掘等方面对GWAS在林木育种中的应用进行了综述,并提出后续研究展望。

1 种质材料与性状的选择及表型鉴定

林木群体具有较高的遗传多样性,驯化程度低,同时具有连锁不平衡快速衰减等特点(Du等2018),为林木育种工作者提供了GWAS研究和育种的理想体系。但群体规模过大时,基因型检测成本和性状测量成本都会相应增加。所以,选择表型齐全和遗传变异丰富的作物研究群体可减少研究时所需群体内的个体数目,提高关联分析的分辨率(姜洪真等2018)。在种质材料的选择时,应该以最少的种质资源在最大程度上代表该物种的遗传多样性,从而提高GWAS的效率。

从表1中可看出,在林木GWAS研究中,大部分(60%)研究所选种质材料为相关林木全分布区采样的无关个体构成的自然群体,只有那些在前期具有育种基础(交配设计)的树种所用的种质材料为半同胞家系或全同胞家系。这是由于林木育种进展相对缓慢造成的普遍现象。但自然群体进化历史的各种变异和重组,使数量性状基因座(quantitative trait locus, QTL)的定位精度有明显的提升,对GWAS研究更加有利。也有研究指出,自然群体在做GWAS时容易出现假阳性的结果(Korte和Farlow 2013)。但随着DNA测序技术的不断发展,高通量基因分型技术的进一步提高,以及群体结构统计方法的不断更新,促使自然群体在GWAS研究中的重要性逐步显现。从表1可以看出,研究中所用样本大小从100个到3 373个,样本大小不尽相同。在样本大小选取方面,并没有一定之规,关键在于所取样本是否能够代表研究对象的绝大部分基因型,

这就要求样本选取的地理范围要包括研究对象的全分布区(Korte和Farlow 2013)。在早期林木育种中,由于育种工作者主要关注林木的产量,通常考虑的目标性状为树高、胸径、材积等生长性状。随着育种工作逐步深入,育种工作者发现林木的木材性状和抗逆性状与生长性状同样重要。由表1可以看出,目前关于林木育种的GWAS研究中,仍然以生长性状(60%)、木材性状(46.7%)和抗逆性状(26.7%)为主,少数有特殊需求的研究主要为生理生态性状或物候性状。

在进行目标性状的选择时,除了考虑该性状的重要经济效益及生态效益,还应考虑性状表型鉴定的可行性以及准确性,相关数据采集的简易性以及可重复性。例如在木材性状的相关研究中,为使研究材料可以持续生长并保留种质资源,在测量木材密度时,应选择无损检测技术(Cappa等2013; 张帅楠等2017),而不应杀鸡取卵式的将树伐倒以达到研究目的。

目前,代谢组学的兴起也使目标性状的选择更加多样化,表型鉴定的方法越来越多样化和精确化,这为林木育种的GWAS研究带来不小的便利。例如,De La Torre等(2019)在火炬松(*Pinus taeda*) GWAS研究中发现,与代谢性状相关的SNPs广泛分布于整个基因组中,大多数代谢性状与位于不同连锁群的多个QTLs表现出相关性。

2 全基因组SNP位点的获取

通常来说,在某个生物体的全基因组中,单个碱基发生突变的频率大于1%的位点称为SNP,其具有较高的遗传稳定性(Seki等2005)。目前常用的SNP检测方法主要有SNP芯片技术和全基因组重测序技术。SNP芯片技术已广泛应用于动植物的遗传研究和生产实践,该技术是预先将寡核苷酸探针阵列固定在载体上,随后通过目标DNA与探针进行特异性反应,通过电脑显示接收反应信号的强弱或者有无从而确定SNP位点。SNP在植物中的分布具有数量多、密度高的特点(郝岗平等2004),例如水稻中每268个碱基就有1个SNP出现(Shen等2004)。早期的林木GWAS研究中,主要使用的是Illumina公司的Infinium和GoldenGate 两大

表1 主要林木GWAS研究综述

Table 1 A review of genome-wide association analysis about forest tree breeding

树种	表型性状	样本来源	样本大小/个	标记位点数/个	标记位点获取方式	关联分析模型	文献来源
蓝桉 (<i>Eucalyptus globulus</i>)	生长性状; 木材性状	家系	303	7 680	分集阵列技术(DArT)	统一混合模型(unified mixed model, UMM)	Cappa等2013
日本柳杉 (<i>Cryptomeria japonica</i>)	木材性状; 雄花(花粉)产量	无关个体(优树嫁接无性系)	367	1 032	基因芯片	有效混合模型分析(efficient mixed model analysis, EMMA)	Uchiyama等2013
三角叶杨 (<i>Populus trichocarpa</i>)	木材性状	无关个体(无性系林)	334	29 233	基因芯片	广义线性模型(general linear models, GLM)	Porth等2013
三角叶杨 (<i>Populus trichocarpa</i>)	生长量性状; 生理生态性状; 物候性状	无关个体	448	34 131	基因芯片	广义线性模型(general linear models, GLM)	McKown等2014
白云杉(<i>Picea glauca</i>)	木材性状	半同胞家系	1 694	7 437	基因芯片	混合线性模型(mixed linear model, MLM)	Lamara等2016
巨桉 (<i>Eucalyptus grandis</i>)	生长性状; 木材性状; 抗虫性	无关个体(种源林)	100	188 591	重测序	统一混合模型(unified mixed model, UMM)	张照远2017
山毛榉 (<i>Fagus grandifolia</i>)	树皮抗病性	无关个体(天然林)	514	5 838	转录组测序	逻辑斯蒂回归模型(logistic regression model, LRM)	Calic等2017
美洲黑杨 (<i>Populus deltoides</i>)	生长性状和木材形状	无关个体(种源林)	391	18 153	重测序	广义线性模型(general linear models, GLM)	Fahrenkrog等2017
粗皮桉 (<i>Eucalyptus pellita</i>)	生长性状(树高、胸径、材积)	半同胞家系	732	19 506	重测序	混合线性模型关联分析(mixed linear model association, MLMA)	Muller等2017
边沁桉 (<i>Eucalyptus benthamii</i>)	生长性状(树高、胸径、材积)	半同胞家系	505	13 787	重测序	混合线性模型关联分析(mixed linear model association, MLMA)	Muller等2017
巨尾桉 (<i>Eucalyptus grandis</i> * <i>Eucalyptus urophylla</i>)	生长性状; 木材性状; 抗病性	F2代全同胞家系	768	2 191	基因芯片	分层遗传作图(regional heritability mapping)	Resende等2017
火炬松(<i>Pinus taeda</i>)	生长性状; 抗旱性; 抗病性; 代谢组; 基因表达量	无关个体(种源林)	377	87 825	重测序	混合线性模型(mixed linear model, MLM)	De La Torre等2019
三角叶杨 (<i>Populus trichocarpa</i>)	树皮纹理	无关个体(无性系)	917	8 253 066	重测序	有效混合模型关联分析(efficient mixed model association, EMMA)	Bdeir等2019
巨尾桉 (<i>Eucalyptus grandis</i> * <i>Eucalyptus urophylla</i>)	生长性状	F2代全同胞家系	3 373	59 222	基因芯片	线性模型关联分析(linear model-based association, LMA)/混合线性模型关联分析(mixed linear model-based association, MLMA)	Mueller等2019
三角叶杨 (<i>Populus trichocarpa</i>)	形态性状; 生理性状	无关个体(扦插无性系)	882	6 781 211	重测序	全基因组有效混合线性模型关联分析(genome-wide efficient mixed-model association, GEMMA)	Bdeir等2019

产品系列。Porth等(2013)采用了Illumina Infinium 34K SNP基因分型阵列, 该阵列从334个无关的毛白杨(*Populus trichocarpa*)个体的群体中的29 233个高质量的SNP中筛选出约3 500个可用的SNPs;

McKown等(2014)在毛白杨的生态性状GWAS研究中同样使用了Illumina Infinium 34K SNP基因分型阵列, 并获得34 131个候选SNPs; Lamara等(2016)则使用的是Illumina Infinium平台的HD iSelect珠片

进行SNP基因分型; Uchiyama等(2013)在对日本柳杉(*Cryptomeria japonica*)的GWAS研究中利用Illumina GoldenGate 平台的SNP基因分型阵列, 选择SNP的策略为每1 536个序列中选择1个SNP; 而Sansaloni等(2010)则专门针对桉树(*Eucalyptus globulus*)开发了一种用于全基因组基因分型的高密度多样性阵列技术(DArT)基因芯片。

全基因组重测序的发展为快速精确的QTL定位提供了技术基础(袁金红等2015)。全基因组重测序不仅可以直接获取高达百万甚至千万级别的可信度较高的SNP位点供关联分析研究所需, 且对未知变异点中新的SNP位点和稀有SNP位点的检测非常灵敏。Du等(2018)在棉花(*Anemone vitifolia*) GWAS研究中利用重测序技术获得17 883 108个高质量的SNPs; Fang等(2017)使用直接深度测序比较并基于PCR的测序策略验证SNP调用的质量, 并获得2 167 186个常见的SNPs; SMR等(2016)在非洲水稻的GWAS研究中利用成对端部Illumina测序技术对98份非洲水稻进行重测序, 获得2 317 937个SNPs。随着测序技术不断提高, 全基因组测序的价格也在日益下降。林木育种中研究中越来越多的关于GWAS的研究采用直接高通量重测序的方法获取全基因组的SNP位点。例如Fahrenkrog等(2017)通过对18 153个基因进行定向重测序的方法研究生长和木材组成性状之间的关联性; Bdeir等(2019)在对毛白杨的树皮性状GWAS研究中使用重测序技术获得8 253 066个SNPs用于后续分析; De La Torre等(2019)在火炬松复杂性状的GWAS研究中, 利用重测序获得5 200 000个SNPs, 并从中选择6 350个SNPs标记进行基因分型。

3 群体结构和连锁不平衡分析

林木GWAS研究中样本量大, SNP标记位点多, 需要避免假阳性和I型错误才能成功筛选出与目标性状相关联的SNP位点(Chanock等2007)。由于林木GWAS研究所选群体要求遗传背景一致或相似的自然群体, 会有群体分层的情况, 不同种源或不同区域的个体之间常常存在遗传差异, 在关联分析时会导致虚假关联(Cappa等2013)。因此要对研究群体进行群体结构研究, 以期消除

或减少虚假关联。在林木GWAS研究中, 群体结构研究常用的有主成分分析法和混合线性模型分析法。

主成分分析法主要是通过构建亲缘关系矩阵计算出特征向量值, 以此反映不同个体间的遗传差异。该方法主要应用于亚群分界不明显的自然群体中。虽然主成分分析法可以有效地减少群体结构对关联分析的影响, 但是该方法受样本大小的影响较大。Porth等(2013)在研究中为避免样本大小对结果的影响, 采用了后向逐步回归(即基于PCA的模型, 通过贝叶斯信息准则选择主成分), 并只选取具有解释基因组数据中90%总变异的特征值的主成分进行分析; Lamara等(2016)利用主成分分析评估了1 694个样本的群体中是否存在群体亚结构; Calic等(2017)也利用主成分分析对试验群体进行了结构评估, 结果从506个个体和5 838个SNPs中筛选出有效的327个个体和3 220个SNPs用于后续分析。

混合线性模型将谱系关系作为随机效应, 将性状数据作为固定效应, 并将估算的模型误差考虑进去, 从而可以消除或淡化群体结构中亲缘关系对关联分析的影响。常规育种中还常常用该方法进行无偏估计(Yang等2014)。在林木GWAS研究中, 混合线性模型的应用具有突出的优势, 也被大部分林木育种工作者所采用。Uchiyama等(2013)比较了三个模型: 没有结构或亲属关系效应的混合模型(naïve模型)、带有协变量以说明种群结构的模型(Q模型)以及包含种群结构和基于标记的亲属关系估计的混合模型(Q+K模型), 结果显示naïve模型的估计偏差最大, Q+K模型的误报可能性最小; McKown等(2014)则使用统一混合线性模型(UMM)评估了群体中遗传结构对关联性的影响, 并使用BIC比较了模型之间的对数似然值。

4 关联作图与候选基因挖掘

关联作图与候选基因挖掘是林木GWAS研究中的核心工作。在关联分析中, 虽然有少部分研究仍然采用广义线性模型(McKown 2013), 但随着强大的分析模型-混合线性模型在动物育种中的广泛应用(Korte和Farlow 2013), 林木育种工作中也

越来越多地使用混合线性模型进行关联分析,并逐渐衍生出针对不同侧重点的混合线性模型。如表1所示,67%的研究中所用关联分析模型为各式各样的混合线性模型。

Porth等(2013)在杨树的木材性状GWAS研究中分离出141个SNPs位点与16个木材性状显著相关,并发现候选基因中41%的位点参与木材次生生长的调控;Uchiyama等(2013)在研究日本柳杉时,利用混合线性模型关联分析找到6个新的SNPs位点;Cappa等(2013)在蓝桉的育种研究中,检测到18个与性状相关联的标记位点,其中16个与胸径生长相关,2个与木质素单体组成(S:G比率)显著相关,并且将5-羟化酶(F5H)基因物理定位到1 Mb内;McKown等(2014)在对杨树的GWAS研究中,在19条染色体上分离出410个SNPs位点与性状显著相关,并根据基因注释得到与物候性状相关的基因240个、与生物量相关的基因53个以及与生理生态性状相关的基因25个;Lamara等(2016)在白云杉(*Picea glauca*)的育种研究中利用关联分析鉴定出每个木材性状的关联基因多达229~292个,并利用180个木质部相关基因构建了木质部共表达网络;Resende等(2017)在桉树育种研究中,利用区域遗传力作图(regional heritability mapping)检测到包含2 191个SNPs位点的26个数量性状位点,利用GWAS检测到13个SNPs位点与性状相关;Calic等(2017)在对山毛榉(*Fagus grandifolia*)的树皮抗病性的研究中,检测出4个SNPs与性状极显著相关,并将标记位点定位到5号染色体的一个编码金属螯合蛋白mRNA的单基因上。Fahrenkrog等(2017)在对美洲黑杨(*Populus deltoides*)做GWAS研究中发现,所有表型都有对应的SNPs与之相关;而Muller等(2017)则在桉树育种群体中,比较了基因组选择与GWAS的优劣,研究认为:在群体结构清晰的育种群体中,基因组选择要优于GWAS。张照远(2017)在对巨桉(*Eucalyptus grandis*)重要性状的GWAS研究中,鉴定出15个SNPs与桉树枝瘿姬小蜂危害显著相关,并检测到定位在8条染色体框上的51个基因与植物抗性、病害防御、呼吸及光合作用、适应胁迫环境等有关;Bdeir等(2019)在对杨树的树皮纹理研究中,检测到42个与性状显著相

关的标记位点,并鉴定出98个与树皮纹理相关的候选基因。Chhetri等(2019)在杨树多性状GWAS研究中发现,多性状关联分析比单性状关联分析更加有效,对于基因共表达网络的研究尤其有用。Mueller等(2019)则比较了独立的GWAS和联合的GWAS的效率,研究认为:联合的GWAS能够更加精确地捕获与性状相关的标记位点;De La Torre等(2019)在火炬松的研究中鉴定出2 335个新的与性状相关的SNPs位点,并在研究中指出突变选择平衡会导致代谢性状的遗传变异。

5 研究展望

GWAS是复杂数量性状定位的有效手段,该分析不需要构建家系群体,使用自然群体即可同时对多个性状进行分析,找到与性状关联的SNP位点,进而找到与性状关联的候选区间。但是该方法也有其局限性,即对微效多基因控制的数量性状的检测能力不足;而林木的绝大部分性状均为微效多基因控制的数量性状(De La Torre等2019),每个基因对性状的贡献度都很小,在GWAS中往往达不到显著水平,从而不能被准确挖掘出来。在动物育种中,已经有针对该缺点的应对之策。Deng等(2019)在对水牛(*Bubalus bubalis*)产奶性状的研究中,首次应用了GWAS与权重基因共表达网络分析(weighted gene co-expression network analysis)相结合的方法,结合转录组数据,鉴定到了12个与产奶量有关的枢纽基因。此外,随着测序技术的发展和价格的平民化,多组学联合分析的GWAS方法逐渐成为热门,并且针对转录组学、蛋白组学和代谢组学的综合分析已经在烟草(*Nicotiana tabacum*)中获得成功(Zhou等2015)。在林木育种中,应及时学习和利用对人类、动物及作物研究中的最新研究方法和分析手段,以达到更加高效的育种目的。

GWAS虽然已经应用到林木育种工作中,但大部分研究还停留在寻找与目的性状相关的SNPs位点/候选区间/候选基因上,极少数研究对所获得的目的基因进行过实验验证。在后续的研究中还需加强对GWAS的结果进行实验验证和深入研究。

参考文献(References)

- Alqudah AM, Sharma R, Pasam RK, et al (2014). Genetic dissection of photoperiod response based on GWAS of pre-anthesis phase duration in spring barley. *PLoS ONE*, 9 (11): e11312011
- Bdeir R, Muchero W, Yordanov Y, et al (2019). Genome-wide association studies of bark texture in *Populus trichocarpa*. *Tree Genet Genomes*, 15: 14
- Calic I, Koch J, Carey D, et al (2017). Genome-wide association study identifies a major gene for beech bark disease resistance in American beech (*Fagus grandifolia* Ehrh.). *BMC Genomics*, 18: 547
- Cappa EP, El-Kassaby YA, Garcia MN, et al (2013). Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: A case study in *Eucalyptus globulus*. *PLoS ONE*, 8 (11): e8126711
- Chanock SJ, Manolio T, Boehnke M, et al (2007). Replicating genotype-phenotype associations. *Nature*, 447 (7145): 655–660
- Chen C, Park T, Wang X, et al (2019). China and India lead in greening of the world through land-use management. *Nat Sustainability*, 2 (2): 122–129
- Chen W, Gao Y, Xie W, et al (2014). Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat Genet*, 46 (7): 714–721
- Chhetri HB, Macaya-Sanz D, Kainer D, et al (2019). Multitrait genome-wide association analysis of *Populus trichocarpa* identifies key polymorphisms controlling morphological and physiological traits. *New Phytol*, 223 (1): 293–309
- De La Torre AR, Puiu D, Crepeau MW, et al (2019). Genomic architecture of complex traits in loblolly pine. *New Phytol*, 221 (4): 1789–1801
- Deng T, Liang A, Liang S, et al (2019). Integrative analysis of transcriptome and GWAS data to identify the hub genes associated with milk yield trait in buffalo. *Fron Genet*, 10: 36
- Du Q, Lu W, Quan M, et al (2018). Genome-wide association studies to improve wood properties: challenges and prospects. *Fron Plant Sci*, 9: 1912
- Du X, Huang G, He S, et al (2018). Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nat Genet*, 50 (6): 796–802
- Fahrenkrog AM, Neves LG, Jr. Resende MFR, et al (2017). Genome-wide association study reveals putative regulators of bioenergy traits in *Populus deltoides*. *New Phytol*, 213 (2): 799–811
- Fang L, Wang Q, Hu Y, et al (2017). Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat Genet*, 49 (7): 1089–1098
- Farfan IDB, De La Fuente GN, Murray SC, et al (2015). Genome wide association study for drought, aflatoxin resistance, and important agronomic traits of Maize hybrids in the sub-tropics. *PLoS ONE*, 10 (2): e01177372
- Gajardo HA, Wittkop B, Soto-Cerda B, et al (2015). Association mapping of seed quality traits in *Brassica napus* L. using GWAS and candidate QTL approaches. *Mol Breeding*, 35: 143
- Hao GP, Yang Q, Wu ZY, et al (2004). Single nucleotide polymorphism (SNP) and its applications in crop genetics and breeding. *Chin Bull Bot*, 21 (5): 618–624 (in Chinese with English abstract) [郝岗平, 杨清, 吴忠义等(2004). 植物的单核苷酸多态性及其在作物遗传育种中的应用. *植物学通报*, 21 (5): 618–624]
- Hirschhorn JN, Daly MJ (2005). Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*, 6 (2): 95–108
- Huang X, Wei X, Sang T, et al (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet*, 42 (11): 961–967
- Huang X, Zhao Y, Wei X, et al (2012). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat Genet*, 1 (44): 32–39
- Iwata H, Minamikawa MF, Kajiya-Kanegae H, et al (2016). Genomics-assisted breeding in fruit trees. *Breeding Sci*, 66 (1): 100–115
- Jiang HZ, Ma BJ, Qian Q, et al (2018). The application of genome-wide association study (GWAS) in crop agronomic traits. *J Agr Biotech*, 26 (7): 1244–1257 (in Chinese with English abstract) [姜洪真, 马伯军, 钱前等 (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]
- Klein RJ, Zeiss C, Chew EY, et al (2005). Complement factor H polymorphism in age-related macular degeneration. *Science*, 308 (5720): 385–389
- 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 F, Chen B, Xu K, et al (2016). A genome-wide association study of plant height and primary branch number in rapeseed (*Brassica napus*). *Plant Sci*, 242: 169–177
- McKown AD, Klapste J, Guy RD, et al (2014). Genome-wide

- association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa*. *New Phytol*, 203 (2): 535–553
- Mueller BSF, de Almeida Filho JE, Lima BM, et al (2019). Independent and Joint-GWAS for growth traits in *Eucalyptus* by assembling genome-wide data for 3373 individuals across four breeding populations. *New Phytol*, 221 (2): 818–833
- Muller BSF, Neves LG, de Almeida Filho JE, et al (2017). Genomic prediction in contrast to a genome-wide association study in explaining heritable variation of complex growth traits in breeding populations of *Eucalyptus*. *BMC Genomics*, 18: 524
- Porth I, Klapste J, Skyba O, et al (2013). Genome-wide association mapping for wood characteristics in *Populus* identifies an array of candidate single nucleotide polymorphisms. *New Phytol*, 200 (3): 710–726
- Resende RT, Vilela Resende MD, Silva FF, et al (2017). Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in *Eucalyptus*. *New Phytol*, 213 (3): 1287–1300
- Risch N, Merikangas K (1996). The future of genetic studies of complex human diseases. *Science*, 273 (5281): 1516–1517
- S MR, Young CJ, Michelle S, et al (2016). Domestication history and geographical adaptation inferred from a SNP map of African rice. *Nat Genet*, 48 (9): 1083–1088
- Sansaloni CP, Petrolis CD, Carling J, et al (2010). A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*. *Plant Methods*, 6: 16
- Seki S, Kawaguchi Y, Chiba K, et al (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
- Shao XY, Song XY, Pan SX, et al (2017). Genome-wide association study and Meta-QTL analysis of ear diameter trait in maize. *Plant Physiol J*, 53 (12): 2091–2102 (in Chinese with English abstract) [邵晓宇, 宋希云, 潘顺祥等(2017). 玉米穗粗性状的全基因组关联分析及QTL元分析. *植物生理学报*, 53 (12): 2091–2102]
- Shen YJ, Jiang H, Jin JP, et al (2004). Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. *Plant Physiol*, 135 (3): 1198–1205
- Stanturf JA, Palik BJ, Dumroese RK (2014). Contemporary forest restoration: A review emphasizing function. *Forest Ecol Manag*, 331: 292–323
- Tian B, Xin PX, Zhang XJ, et al (2013). Applications of high-throughput sequencing in forest tree breeding. *Modern Agr Sci Tec*, (12): 130–132 (in Chinese with English abstract) [田斌, 辛培尧, 张雪娟等(2013). 高通量测序技术在林木育种中的应用. *现代农业科技*, (12): 130–132]
- Tian F, Bradbury PJ, Brown PJ, et al (2011). Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat Genet*, 43 (2): 159–162
- Uchiyama K, Iwata H, Moriguchi Y, et al (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
- Wisniewski M, Nassuth A, Arora R (2018). Cold hardiness in trees: A mini-review. *Front Plant Sci*, 9: 1394
- Yang J, Zaitlen NA, Goddard ME, et al (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet*, 46 (2): 100–106
- Yuan JH, Li JH, Huang XC, et al (2015). Advance of SNP analysis based on whole genome resequencing in crop gene mapping. *Plant Physiol J*, 51 (9): 1400–1404 (in Chinese with English abstract) [袁金红, 李俊华, 黄小城等(2015). 基于全基因组重测序的SNP分析在作物基因定位中的研究进展. *植物生理学报*, 51 (9): 1400–1404]
- Zhang SN, Jiang JM, Xu YQ, et al (2017). Study on the modulus of elasticity non-destructive evaluation technique of slash pine standing tree. *Forest Res*, 30 (1): 75–80 (in Chinese with English abstract) [张帅楠, 姜景民, 徐永勤等(2017). 湿地松活立木弹性模量无损评估技术研究. *林业科学研究*, 30 (1): 75–80]
- Zhang ZY (2017). Genome-wide association analysis with significant characters and integration on resistance to *Leptocybe invasa* in *Eucalyptus grandis* (dissertation). Beijing: Chinese Academy of Forestry (in Chinese with English abstract) [张照远(2017). 巨桉重要性状GWAS分析及抗桉树枝瘿姬小蜂整合研究(学位论文). 北京: 中国林业科学研究院]
- Zhou L, Li R, Fan L, et al (2013). Mapping epistasis and environment x QTX interaction based on four-omics genotypes for the detected QTX loci controlling complex traits in tobacco. *Crop J*, 1: 151–159
- Zhou Y, Tang H, Cheng M, et al (2017). Genome-wide association study for pre-harvest sprouting resistance in a large germplasm collection of Chinese wheat landraces. *Front Plant Sci*, 8: 401
- Zinkgraf M, Liu L, Groover A, et al (2017). Identifying gene coexpression networks underlying the dynamic regulation of wood-forming tissues in *Populus* under diverse environmental conditions. *New Phytol*, 214 (4): 1464–1478

Application of genome-wide association analysis (GWAS) in forest tree breeding

QUE Qing-Min, OUYANG Kun-Xi, LI Pei, CHEN Xiao-Yang*

College of Forestry and Landscape Architecture, South China Agricultural University; Guangdong Key Laboratory for Innovative Development and Utilization of Forest Plant Germplasm, Guangzhou 510642, China

Abstract: Genome-wide association studies (GWAS) have been developed in recent years as an effective method for analyzing the genetic basis of phenotypic diversity in humans, animals or plants. It has many advantages, such as high throughput, high precision and high speed, etc. This paper reviews the application of genome-wide association analysis in tree breeding from the aspects of selection of germplasm materials, selection of target traits and phenotypic identification, acquisition of genome-wide single nucleotide polymorphism (SNP) markers, analysis of population structure and linkage imbalance, association mapping and discovery of candidate genes, and puts forward the prospect of future research. It is expected to provide basis and reference for further research on genetic basis of various traits in tree breeding by using genome-wide association analysis.

Key words: Genome-wide association analysis; GWAS; linkage disequilibrium; forest tree breeding

Received 2019-08-30 Accepted 2019-10-26

This work was supported by the Science and Technology Program of Guangdong Province (2017B020201008); National Natural Science Foundation of China (31600525); Extension and Demonstration Project of Forest Science and Technology from China State Financial Budget [(2018)GDTK-08].

*Corresponding author(xyichen@scau.edu.cn).