

DOI: 10.13376/j.cbls/2015137

文章编号: 1004-0374(2015)08-0986-09



罗杰, 教授, 博士生导师。长期从事植物次生代谢调控及代谢组学研究, 博士毕业后赴英国 John Innes Centre 实验室从事访问及博士后研究。2009 年起在华中农业大学作物遗传改良国家重点实验室从事水稻、玉米等主要作物的代谢组学研究。近年来以通讯作者或第一作者在 *Nature Genetics*、*Nature Communications*、*PNAS*、*Plant Cell* 等国际主流杂志发表多篇高水平论文, 具有较大的影响。

基于连锁与关联分析的植物代谢组学研究进展

刘贤青, 董学奎, 罗 杰*

(华中农业大学生命科学技术学院, 武汉 430070)

摘要: 代谢物作为生物体表型的基础能够帮助我们更直观有效地了解生物学过程及其机理。植物代谢组研究对理解植物代谢途径, 改善植物对环境胁迫的应答, 提高作物产量与品质有着非常重要的作用。代谢物的种类和数量在不同品种与组织中存在很大差异, 利用这些差异进行遗传基础的解析将有助于我们深入了解代谢生物学过程。随着代谢组学分析技术和方法的发展与第二代测序技术的应用, 植物代谢组学的研究内容不断地扩增, 代谢遗传机制的研究不断深入。概述了植物代谢组学的研究内容、基因型分型、植物代谢组遗传连锁分析与植物代谢组全基因组关联分析及其应用。

关键词: 代谢组学; 广泛靶向代谢组学; 基因型分型; 数量性状座位; 全基因组关联分析; 基因功能

中图分类号: Q591; Q945.1 **文献标志码:** A

Progresses in plant metabolomics using linkage and association mapping strategies

LIU Xian-Qing, DONG Xue-Kui, LUO Jie*

(College of Life Science and Technology, Huazhong Agricultural University, Wuhan 430070, China)

Abstract: It has been demonstrated that metabolites, as the basis of biological phenotypes, help us to better understand biological processes and provide insights into the biochemical mechanisms of the processes. Research on plant metabolomics plays a vital role in understanding plant metabolic pathways, manipulating plant responses to environmental stresses, improving crop yields and enhancing food quality. Significant difference could be found in the number and content of metabolites in plants. Dissection of the genetic basis of metabolome thus helps us to extensively understand metabolically biological processes. Moreover, with the advancement of metabolomics and second-generation sequencing, researches in plant metabolomics with its genetic basis are

收稿日期: 2014-12-31; 修回日期: 2015-01-11

基金项目: 国家自然科学基金项目(31070267, 31100962)

*通信作者: E-mail: jieluo@mail.hzau.edu.cn, Tel: 027-87281960

increasing. This review summarizes progresses in plant metabolomics, genotyping, and the understanding of genetic basis of plant metabolome by linkage and association mapping, which can be used as reference for related research.

Key words: metabolomics; widely-targeted metabolomics; genotyping; quantitative trait loci; genome-wide association study; gene function

代谢组学是指在整体水平上研究生物体的代谢物, 是对某一生物、组织或细胞中的所有低分子量化合物进行定性与定量分析的一门学科。代谢组学以指标分析为基础, 高通量检测为手段, 通过研究生物体内代谢物的种类与数量及其变化规律来阐述机体在正常生命状态下及环境变化后的代谢过程^[1]。代谢组学定义起源于20世纪90年代, Jeremy Nicholson等首次提出的“metabonomics”概念, 以及21世纪初Oliver Fiehn提出的“metabolomics”一词^[2-3]。

植物产生的大量化学性质与生物学功能差异的代谢物质是人类营养、能量与医药的来源。植物由于固着生长, 面对多样的环境, 形成了一套复杂的代谢物调控系统来弥补这一缺陷^[4]。植物代谢物总数已超过20万种, 植物代谢物的丰富与复杂程度体现于代谢物在结构与功能上的多样性^[5]。一直以来, 植物的化学组成吸引着生物研究人员, 因为生化表型很大程度上反映了植物与环境的相互作用^[6]。此外, 不同的物种之间以及同一物种的不同组织中, 代谢物的种类与数量存在着很大的差异^[7]。利用代谢物的差异来挖掘代谢途径及其调控机理, 对于提高植物对环境胁迫的适应性、提高作物产量, 并改善食物品质有着至关重要的作用^[8-10]。

随着植物全基因组测序的完成、突变体库的建立、转基因及其他各种生物技术的发展, 以模式植物拟南芥和水稻为代表的植物功能基因组的研究获得了巨大的进步^[11-12]。继基因组学、转录组学和蛋白质组学之后, 代谢组学作为一门新兴学科, 随着现代分析技术的快速发展, 其研究内容在不断扩展^[13]。同时, 第二代测序技术的出现, 为植物人工杂交群体与自然群体的基因型鉴定和遗传作图带来了方法学上的跨越^[14]。近年来, 利用这些高分辨遗传标记的群体材料, 代谢组学结合连锁分析与全基因组关联分析 (genome-wide association studies, GWAS) 的研究取得了很大的进展, 对植物代谢组的遗传基础解析与分子育种提供了重要的参考^[8,15-16]。本文将简要地综述植物代谢组学、基于遗传连锁分析与全基因组关联分析的植物代谢组学的研究进展。

1 植物代谢组学研究进展

代谢组是系统生物学的分支学科之一。目前, 其研究内容主要包括代谢物的定性或定量分析, 不同基因型的代谢组学表型, 不同生长环境及施加物理、化学或生物性刺激后代谢产物的应答, 代谢途径或代谢网络以及筛选优良品种与转基因植物评价^[13,17]。

生物信息从基因组到代谢组的传递实现了遗传物质和环境因素对个体生长、发育、衰老的调控。作为生物体表型的体现者, 代谢物能够帮助人们有效地了解生命现象, 揭示生命本质。植物代谢物主要分为初生代谢物与次生代谢物两大类^[18], 参与植物对生物胁迫 (病虫胁迫等)、非生物胁迫 (旱涝胁迫、盐碱胁迫、紫外胁迫等) 的应答^[19-20]。其中, 植物激素在植物生长发育及胁迫应答过程中起着关键的调节作用^[21]。它们的合成与调控、尤其是信号途径的研究在近几年取得了突破性的进展^[22-23], 如精氨酸的代谢能够形成一类含氮的脂肪族小分子多胺 (胍丁胺、腐胺、亚精胺、精胺等)。大量研究表明, 多胺及其羟基肉桂酰化的产物 (酚胺 / 羟基肉桂酰胺) 参与调控植物生长、生物与非生物胁迫等生理过程^[24-25]。此外, 越来越多的类黄酮、硫代糖苷、萜类以及生物碱类等化合物被鉴定, 代谢途径得到解析, 尤其是随着检测手段的提高, 越来越多的衍生物得以鉴定^[26-27]。

代谢物种类多、性质差异大, 单靠一种分离分析手段难以进行无偏向的全面分析^[13]。目前, 在代谢组学中较为常见的分析手段有色谱、质谱、磁共振、紫外吸收、放射性检测、库仑分析及红外光谱等 (图1)。磁共振是进行化合物成分结构定性分析最强有力的工具之一^[28]。应用最广泛、最有效的是气相色谱-质谱 (GC-MS) 和液相色谱-质谱 (LC-MS)^[29]。色谱与质谱联用实现了从利用色谱进行物质分离到利用质谱进行物质鉴定的整个流程。其中, LC-MS能分析极性与相对分子质量更高及热稳定性差的化合物, 而且, 大多数情况下无需对非挥发性代谢物进行化学衍生化处理^[30]。毛细管电泳、电化学、飞行时间质谱 (TOF)、傅立叶变化离子共振质谱 (FT-

ICR-MS) 等分离与分析方法也逐步应用到代谢物的检测中^[31]。此外, 植物代谢物检测方法已由传统的靶向代谢组学 (targeted metabolomics) 和非靶向代谢组学 (non-targeted metabolomics) 扩展到广泛靶向的代谢组分析 (widely-targeted metabolomics)。广泛靶向的代谢组分析能够一次定量检测近千种代谢产物, 更加适合大量样品的高通量、广覆盖的代谢组学分析, 有利于全面有效地比较代谢物差异与解析代谢途径^[32-33]。随着分析平台的发展, 许多物质结构得以鉴定。然而, 在非靶向与广泛靶向代谢分析过程中, 大部分代谢物化学结构仍然未知, 代谢物的结构鉴定仍然是目前代谢组学分析的一个重难点^[34]。最近, 我们利用正向遗传学手段结合代谢网络分析或标准品对比分析这种新的方法成功地鉴定或注释了一系列未知代谢物^[8]。尽管如此, 代谢物结构鉴定的发展仍需要人们综合利用生物信息手段, 建立一系列公共代谢组学数据库, 并实现数据共享模式的标准化。

植物代谢物存在大量的修饰, 而且大多以修饰化的形式广泛存在, 包括羟基化、甲基化、糖基化和酰基化等^[8]。不同修饰的代谢产物功能各异。分析这些代谢物的积累模式与非生物胁迫条件下的变化有利于解析它们的生物合成调控途径。苯丙氨酸代谢途径是一条重要的次生代谢物合成途径, 在植物体内, 存在几大类苯丙氨酸衍生物, 如类黄酮 (flavonoids)、羟基肉桂酸、白藜芦醇、木质素、硫代糖苷、植保素与酚胺^[35]。其中, 类黄酮是一种大家熟知的次生代谢物, 根据其结构分为6大类。模式植物拟南芥 (双子叶植物) 中只积累黄酮醇 (flavonol)、花青素 (anthocyanins)、原花青素 (proanthocyanindins)。2013年, Saito 等^[36]揭示了拟南芥至少含有多种不同修饰的类黄酮物质。水稻 (单子叶植物) 中也积累了大量的类黄酮物质, 且主要以黄酮 (flavone) 及其衍生物的形式存在, 而拟南芥中因为缺乏黄酮的合成途径而检测不到黄酮的积累^[37]。类黄酮在拟南芥与水稻中的物质积累模式的研究展现了类黄酮的积累在物种中的组织特异性与时空特异性。我们利用水稻自然资源品种, 分析了不同类型修饰的黄酮含量差异, 发现了不同基团修饰的黄酮在水稻亚种中的积累模式各有不同^[37], 而探索这些自然品种叶子中酚胺 (phenolamides) 的积累模式之后发现, 一些酚胺在亚种间的积累模式与黄酮的积累模式相反^[38]。这些结果表明, 水稻亚种中不同修饰的黄酮与酚胺在进化上存在一些差异

性。另外, 2013年, Park 等^[39]利用代谢组学与转录组学分析发现, 水稻植株在紫外照射的条件下, 叶子中黄酮与酚胺合成的结构基因与相关的代谢途径合成调控基因表达受到强烈地诱导, 物质积累也大量上升。

某个或某一类代谢物在水稻亚种间积累模式存在显著性差异的时候, 这个或这类代谢物就可以被用作生物标志物 (biomarker)。通过代谢组学的研究, 我们可以发现很多代谢物在不同组织、不同品种、不同亚种之间存在差异, 如前所述的黄酮与酚胺^[37-38]。我们发现丙二酰基化的黄酮在籼稻亚种中的含量约是粳稻亚种中的10倍, 五碳糖糖基化及六碳糖糖基化的黄酮在籼稻中含量则远远高于粳稻, 一些苯环酰基化的亚精胺在籼稻中含量却远远低于粳稻。另外, 一些报道指出, 磷缺乏时, 脂类物质的平衡则发生重构; 脯氨酸的含量差异与抗旱等逆境胁迫直接相关^[40]。这些表明, 代谢物的在亚种间差异与其所处的环境非常相关, 可能是进化与人工选择栽培的结果。此外, 越来越多的研究关注与人类健康相关的具有抗氧化、抗癌活性的物质, 如花青素、白藜芦醇等^[13,41]。

2 高通量的基因型鉴定

作物性状的遗传多样性是人们有效地从事作物改良的前提^[42]。自20世纪80年代以来, RFLP、SSR 与 Indel 等 DNA 分子标记 (DNA marker) 技术的出现和发展, 为数量遗传学家提供了重要工具。与形态标记、细胞标记、生化标记相比, DNA 分子标记是 DNA 水平上遗传多态性的直接反映, 具有明显的优点。单核苷酸多态性 (single nucleotide polymorphisms, SNPs) 是指由于单个核苷酸的变异所引起的 DNA 序列多态性。近年来, 随着 DNA 测序技术与 SNPs 检测与分析方法的飞速发展, SNPs 成为最广泛应用的第三代分子标记 (图 1)。

植物全基因组序列是遗传研究的基础, 对于研究自然资源的遗传变异至关重要。自从10年前水稻全基因组序列公布以来, 很多作物的全基因组序列也相继问世^[43]。同时, 第二代测序技术与新的计算生物学方法的发展, 使全基因组遗传作图变得更为快捷准确。利用第二代测序技术进行的基因型鉴定方法具有快速价廉、高密度标记覆盖、高精度度和高分辨率等特点^[14]。大量品种的重测序, 使更多的作图群体和物种之间的基因组比较和遗传图谱构建成为可能。近年来, 基于第二代测序技术的植物

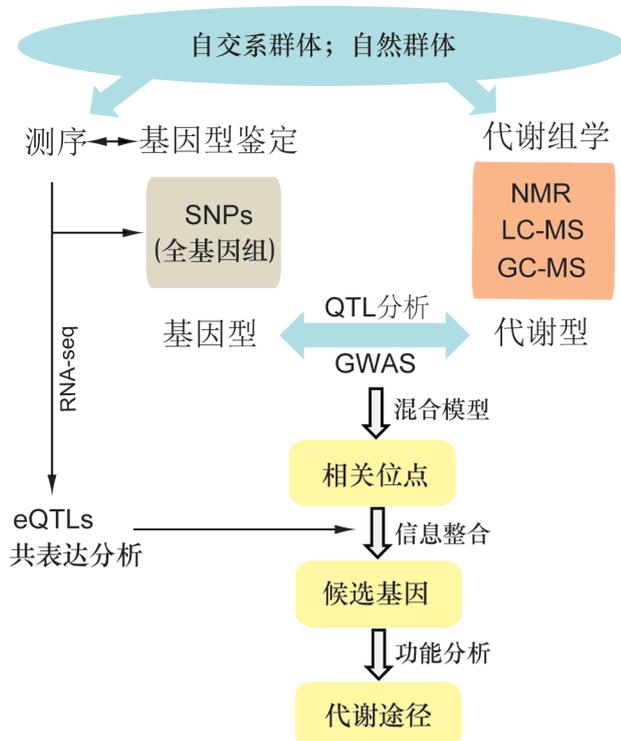


图1 代谢组学在基于遗传连锁分析与全基因组关联分析的植物生物学中的应用

基因型分型得到广泛的应用。2009年, Huang等^[44]通过对150份水稻籼稻品种9311与粳稻品种日本晴的重组自交系的群体(RILs)进行重测序, 构建了基于SNPs的超高密度连锁图谱。与之前仅有几百个基于PCR的分子标记图谱相比, 利用测序进行连锁谱图的构建更加快捷、准确。对群体一些农艺性状进行连锁分析, 发现了多个数量性状位点(quantitative trait locus, QTL), 遗传效应值达到46%。2010年, Huang等^[45]对500多个水稻地方品种进行低丰度的全基因组重测序(low-coverage whole-genome resequencing), 获得了上百万个有效SNPs, 构建了超高密度基因型图谱, 进一步对群体的14个重要农艺性状进行了全基因组关联分析, 获得了几个大效应的位点, 并找到了相关的候选基因。

基于RNA测序(RNA sequencing, RNA-seq)的基因型分型已引起大家的关注。基于第二代测序技术的RNA测序为转录组学的作图与定量分析提供了新方法, 同时为高通量、定量地研究基因转录与可变剪接打开新局面^[46]。不同的物种虽然在基因组大小上差异非常大, 但是在基因数量上差别不大。基因组上存在许多重复区域, 所以进行转录组的测序挖掘SNPs比全基因组的重测序要显得更加实惠。

2013年, Li等^[47]通过对玉米368份品种进行RNA测序, 分析出了大量的SNPs, 用于构建基因型图谱; 同时, RNA测序的基因表达丰度被用于进一步的表达量QTL(expression QTL, eQTL)分析, 来挖掘油份合成调控的功能基因。高通量测序使得遗传的分子标记资源变得更加丰富, 水稻、玉米、高粱、大豆、大麦、番茄等主要农作物均已构建了高密度SNP遗传图谱, 进而实现QTL的精细定位和相关基因的分离克隆^[48]。

3 基于连锁分析的代谢组学遗传基础研究进展

在过去10年里, 将代谢组学手段整合到表型多样性及其遗传变异的研究中, 一直吸引着科学家的兴趣, 尤其是植物生物学家们的关注^[19]。利用代谢组学大规模地检测不同物种或者品种间代谢物的种类与含量, 在此基础上运用QTL作图可对各种代谢物含量进行遗传分析。植物组织中代谢物的水平也可以看作为一个数量性状, 因此, 我们也可以对代谢物性状(m-trait)进行相关的QTL分析^[49](图1)。近年来, 得益于高密度遗传连锁图谱的构建, 代谢物差异的遗传基础解析取得了非常大的进展。

代谢物遗传基础的分析首先锁定在拟南芥与一些作物中某些特定物质, 即靶向的mQTL分析。利用传统的分子标记遗传群体, 鉴定出了一些控制西红柿中可溶性糖、玉米中防御性物质与拟南芥、油菜中硫代糖苷类物质的QTL^[50-52]。随着代谢组学的兴起, 大规模的非靶向代谢物检测与QTL分析结合起来, 深化了植物代谢组学的遗传研究。2007年, Keurentjes^[27]通过非靶向的LC-QTOF MS检测了叶子中的代谢物, 发现群体里物质种类、含量差异非常大。利用了160份的自交系材料与基于PCR的分子标记进行进一步的连锁分析阐述了拟南芥中脂肪族的硫代糖苷的遗传途径。2007年, Meyer等^[53]通过重组自交系群体与导入系群体(introgression line, IL), 分析了拟南芥中的初生代谢物, 发现了6个生物量QTL与157个代谢物QTL(metabolic QTLs, mQTLs)。2012年, Matsuda等^[54]大规模地分析了籼粳稻自交系群体的水稻种子中的初生和次生代谢物, 构建了包含759个代谢产物信号的代谢库, 通过连锁分析进一步分析了黄酮合成途径中的糖基转移酶基因。由于标记资源的限制, 所用遗传作图群体往往因分子标记覆盖密度较低而难以准确地提供目标QTL相关信息。超高密度SNPs基因型图谱的构建使mQTL显示出更大的遗传效应,

定位区域更为狭窄。2011年, Yu等^[55]通过群体高通量测序构建了200多个家系水稻重组自交群体超高密度的SNP图谱。我们利用广泛靶向的代谢组学方法分析了这个群体剑叶与发芽种子两个组织共900个代谢物, 其中多数代谢物在两个组织中的广义遗传力超过0.6。对这些代谢组性状进行QTL作图后共获得了超过2800个mQTL, 这些QTL定位区间精度较高(低于100 kb), 遗传效应很大(高达90%)。其中丙二酰基化的黄酮物质QTL定位区间效应值达到95%, LOD值高达146, 其合成的候选基因通过转基因也得以验证^[15]。

2008年, Schauer等^[56]利用导入系材料, 对两年的西红柿的mQTLs进行了遗传性与稳定性的比较, 发现多数的mQTLs显示出一定的遗传显性或加性效应, 而且化学结构较为相似的代谢物, 它们的遗传模式比较类似。大多代谢物合成调控受到多位点的控制, 对水稻剑叶与发芽种子mQTL位点进行成对上位性互作分析(pairwise epistatic interaction analysis)揭示许多代谢物的定位位点, 如丙二酰基化的黄酮的两个QTL位点, 在遗传上存在明显上位性关系。此外, 通过对两个组织的mQTLs在全基因组上进行比较发现, 在剑叶与发芽种子中分别有44个和16个热点区域, 这些热点区域表明可能存在一些主调控因子控制着这些物质的积累^[15]。通过mQTL对遗传互作的分析, 将有助于对代谢网络模型重构与评估, 并对一些特定的代谢途径进行深入研究。

mQTL的研究也为解决基因复制的进化问题提供了思路, 有助于我们了解新基因与代谢物多样性之间的遗传关系^[57]。当一个基因复制之后, 通常会通过遗传保存下来, 同时也会产生一定的变异, 这些变异会发生些新的功能, 从而引起代谢物上的差异以适应复杂多变的生存环境^[58]。尤其是一些次生代谢物mQTL的克隆研究揭示了复制基因的进化途径与特异的生化活性, 如Kliebenstein等^[59]通过mQTL分析发现基因复制产生AOP2、AOP3两个基因, 进化后使用相同底物而进行了不同生化反应, 从而控制拟南芥硫代糖苷合成。

通过mQTL分析, 水稻黄酮类物质与拟南芥硫代糖苷等代谢途径得以重构^[15]。利用多平台的代谢组学分析技术结合mQTL的研究对于功能基因的鉴定与代谢网络的构建有着重要的意义。QTL的研究大多是基于两个亲本进行杂交的遗传群体, 两个亲本性状的差异与遗传重组事件的有限性使性状差

异的检测与群体分辨率的提高都受到限制。选择性状与群体结构差异更大的遗传群体将有助于进一步提升mQTL分析能力。

4 基于关联分析的植物代谢组学遗传基础研究进展

GWAS是在全基因组范围内筛选不同遗传差异个体分子标记的基础上, 分析扫描群体中这些分子标记数据与表型性状之间关联关系的方法。GWAS是一种综合、系统的分析方法, 除了全基因组分子标记之外, 还需要考虑群体材料的群体结构与遗传代表性、表型的选择与鉴定以及关联分析模型的选择等^[60-61]。基于基因芯片与第二代测序的高通量全基因组基因型分型技术的迅速发展, 全基因组关联分析(GWAS)已被广泛地应用于人类疾病与植物复杂农艺性状遗传基础的解析^[62]。然而, 植物的农艺性状, 如产量、开花时间和粒宽等性状, 由于受到较多数量性状位点控制且受环境的影响非常大, 所以这些性状GWAS位点的效应值通常较小^[45,63]。GWAS结合代谢组学分析(metabolic GWAS, mGWAS)在近几年已受到很大的关注, 以阐述代谢物多样性及代谢相关性状的遗传机制(图1)。代谢物在植物中的巨大差异与高遗传力使植物中的mGWAS更加准确^[8]。

基于植物全基因组SNP遗传图谱, 植物中的mGWAS首先在拟南芥中得以开展, 最近mGWAS在水稻、玉米等作物中也有了深入的研究。在代谢组学分析中, 为了理解某些特定代谢途径, 靶向代谢组学通常显示出一定的优势, 所关注途径的基因与物质有一定的研究基础, 靶向代谢物的mGWAS也更便于相关功能基因研究与代谢途径的构建^[13]。硫代糖苷作为拟南芥中一种重要的抗虫抗病物质, 它与环境互作的关系及自然变异的研究仍然缺乏。2011年, Chan等^[64]利用了96个拟南芥品种(约23万SNPs), 对两个组织中43个硫代糖苷代谢性状进行了mGWAS, 鉴定了群体里面硫代糖苷自然变异的两个主要多态性位点。通过比较不同组织(叶子与幼苗)与不同胁迫(模拟病菌侵害)处理条件下的mGWAS位点, 发现发育阶段对GWAS结果有非常大的影响, 而胁迫处理条件下的mGWAS位点与正常条件下类似。这表明发育阶段相比病菌胁迫而言, 对硫代糖苷的影响更大。2013年, Angelovici等^[65]对313份拟南芥群体(17.0344万SNPs)中的一些支链氨基酸性状进行mGWAS, 发

现了一个支链氨基酸转移酶 (BCAT2), 该基因在群体中位点的自然变异与支链氨基酸的自然变异、谷氨酸与其他自由氨基酸的动态平衡显著相关。2013年, Li 等^[47]利用了 RNA 测序构建了 368 份玉米自交系群体的 SNP 标记遗传图谱 (103 万 SNPs), 对群体中 10 个油份物质代谢性状进行 mGWAS, 鉴定了 74 个与玉米籽粒中油份物质含量与脂肪酸组成相关的显著性位点。同时, 也通过 eQTL 对其中一些位点进行了候选基因的分析, 该研究对今后玉米籽粒中油份物质遗传基础解析与油份性状的分子育种提供了重要的参考。通过靶向代谢组学, 我们对水稻叶子中的酚胺物质也进行了 mGWAS, 得到了羟基肉桂酰基化亚精胺的显著性位点 ($P=6.0 \times 10^{-23}$) 与该物质合成的候选基因, 并通过物质积累模式分析和转基因予以了验证^[38]。

随着非靶向与广泛靶向代谢组学分析方法的应用, 非靶向与广泛靶向的 mGWAS 为全面系统地构建植物代谢途径以及解析其遗传机制提供了强有力的手段。2012 年, Riedelsheimer 等^[66]通过对 289 份玉米自交系幼叶中的 118 个代谢性状进行 mGWAS, 其中 26 个代谢物性状得到了显著的 SNPs 位点, 遗传效应值达到 32%。2014 年, Matsuda 等^[67]利用基于 LC-MS 的非靶向代谢检测方法对 175 份水稻自然品种进行 mGWAS, 鉴定了其中 89 个代谢物的显著性位点。最近, 我们利用 LC-MS 对玉米、水稻进行了广泛靶向的代谢谱分析, 并通过 mGWAS 获得了大量代谢物相关的位点, 对其代谢途径的构建及其遗传与生化基础进行了深入的阐述。在 368 份玉米自交系群体中, 鉴定了 983 个代谢性状, 并利用 GWAS 定位到 1 459 个显著性关联位点 ($P < 1.8 \times 10^{-6}$)^[68]。进一步通过自交系群体的 eQTL 与其他 RIL 群体 (334 个家系) 的 mQTL 分析对 mGWAS 显著位点进行比较验证, 阐述了酰基化胍丁胺、酰基化腐胺、色胺与黄酮等物质的代谢途径与遗传机理^[68]。水稻作为单子叶的模式植物, 拥有相对较小的基因组, 且存在丰富的品种资源, 经过长久的自然进化与人工驯化, 水稻遗传与生物生理性状差异较大^[69]。我们利用 524 份自然栽培稻品种资源 (642.877 万 SNPs) 结合广泛靶向代谢组方法, 构建叶片代谢数据库并进行了 mGWAS 分析, 检测到了 2 947 个主效 SNPs, 共 634 个遗传位点^[8]。数据的挖掘分析注释了 36 个影响生理、营养相关代谢物的候选基因, 重构了相关的代谢途径, 进一步的遗传与生化分析鉴定了其中 5 个候选基因^[8]。该

研究为水稻代谢组的差异的遗传与生化基础的解析提供了思路, 为其他表型性状的自然变异分析、水稻的品种改良提供了互补。代谢相关的基因在基因组中占有很大的比例, 基于 mQTL 与 mGWAS 的分析显示, 代谢组与基因组学的整合为代谢相关基因的功能鉴定与代谢途径的重构奠定了良好的基础。

解析遗传变异如何控制性状变异是现代生物学的目标之一^[62]。全基因组分析代谢物相关的显著性位点揭示很多位点的分布也并不是随机的, 而是存在一些热点区域。这些区域里可能存在主效基因, 能够控制一系列物质的积累^[8,70]。拟南芥中一些深入研究表明这些热点可能是选择压力的结果, 选择压力影响了植物的代谢进程^[70]。此外, 利用遗传连锁分析与 mGWAS 来分析自然群体中代谢物的自然变异与代谢相关基因的自然变异也更有助于理解这些变异的机理以及有目的地重构代谢途径。

mGWAS 分析也可以用于研究代谢性状、其他重要农艺性状与生物学性状, 如产量、口感、生物量, 因为这些性状与代谢物可能直接相关^[56,68]。2008 年, Schauer 等^[56]通过分析西红柿果实性状与代谢性状发现这两个性状之间有一定的关系。单一的通过简单的相关性分析与回归模型难以解释它们之间的关系, 但是代谢物组成与其他性状之间的相互关系在今后是可以透过深入的代谢组学分析来预测的。

5 展望

代谢组学是一门交叉学科, 与有机化学、分析化学、信息科学及生物科学等学科密切相关。植物代谢物成分和水平的分析有助于全面了解植物响应细胞内外因子的真实反应, 更直观地揭示遗传背景和环境条件对其决定作用。结合 QTL 与 GWAS 进行代谢组学的分析将大大促进植物代谢组学与植物功能基因组学的研究。QTL 分析与 GWAS 在应用上各有利弊, 将两者结合起来, 并整合转录组学、蛋白质组学无疑更有助于植物代谢生物学的研究。GWAS 主要依赖统计分析, 在研究上可能会出现假阳性和假阴性, 变异重复性不好, 难以确定候选基因, 对稀有变异不敏感等问题。精准的性状鉴定, 设计更加准确的群体结构分析模型, 增加分析群体容量与改进关联分析的数理统计方法将有助于发现真正与性状关联的位点。此外, 连锁作图对于稀有变异非常有效, 综合连锁作图和关联分析作图的巢式关联分析 (nested association mapping, NAM) 群体和 MAGIC (multiparent advanced generation intercross)

群体也受到研究者的重视^[71-72]。使用代谢物与代谢物之间含量数据的比率 (ratio) 作为性状应用到 mGWAS, 也可以增强分析的关联性, 逐渐被应用到代谢物的关联分析^[16]。然而, 代谢组学相关的研究首要任务是代谢物的提取、检测、分析与鉴定。代谢物检测分辨率的提高, 多平台技术的综合应用, 检测与分析新方法的开发与数据库的共享是代谢组学发展的趋势, 也将为植物代谢组学的深入研究提供铺垫^[73]。

[参 考 文 献]

- [1] Fiehn O. Metabolomics-the link between genotypes and phenotypes. *Plant Mol Biol*, 2002, 48(1-2): 155-71
- [2] Nicholson JK, Lindon JC, Holmes E. Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, 1999, 29(11): 1181-9
- [3] Fiehn O, Kopka J, Dormann P, et al. Metabolite profiling for plant functional genomics. *Nat Biotechnol*, 2000, 18(11): 1157-61
- [4] Fernie AR, Schauer N. Metabolomics-assisted breeding: a viable option for crop improvement? *Trends Genet*, 2009, 25(1): 39-48
- [5] Dixon RA, Strack D. Phytochemistry meets genome analysis, and beyond. *Phytochemistry*, 2003, 62(6): 815-6
- [6] Saito K. Phytochemical genomics-a new trend. *Curr Opin Plant Biol*, 2013, 16(3): 373-80
- [7] D'Auria JC, Gershenzon J. The secondary metabolism of *Arabidopsis thaliana*: growing like a weed. *Curr Opin Plant Biol*, 2005, 8(3): 308-16
- [8] Chen W, Gao Y, Xie W, et al. Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat Genet*, 2014, 46(7): 714-21
- [9] Sumner LW, Lei Z, Nikolau BJ, et al. Modern plant metabolomics: advanced natural product gene discoveries, improved technologies, and future prospects. *Nat Prod Rep*, 2015, 32(2): 212-29
- [10] Alonso-Blanco C, Aarts MG, Bentsink L, et al. What has natural variation taught us about plant development, physiology, and adaptation? *Plant Cell*, 2009, 21(7): 1877-96
- [11] Jiang Y, Cai Z, Xie W, et al. Rice functional genomics research: progress and implications for crop genetic improvement. *Biotechnol Adv*, 2012, 30(5): 1059-70
- [12] Atwell S, Huang YS, Vilhjalmsson BJ, et al. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature*, 2010, 465(7298): 627-31
- [13] Saito K, Matsuda F. Metabolomics for functional genomics, systems biology, and biotechnology. *Annual Rev Plant Biol*, 2010, 61: 463-89
- [14] Huang X, Lu T, Han B. Resequencing rice genomes: an emerging new era of rice genomics. *Trends Genet*, 2013, 29(4): 225-32
- [15] Gong L, Chen W, Gao Y, et al. Genetic analysis of the metabolome exemplified using a rice population. *Proc Natl Acad Sci USA*, 2013, 110(50): 20320-5
- [16] Adamski J, Suhre K. Metabolomics platforms for genome wide association studies-linking the genome to the metabolome. *Curr Opin Biotechnol*, 2013, 24(1): 39-47
- [17] Bino RJ, Hall RD, Fiehn O, et al. Potential of metabolomics as a functional genomics tool. *Trends Plant Sci*, 2004, 9(9): 418-25
- [18] Koch K. Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. *Curr Opin Plant Biol*, 2004, 7(3): 235-46
- [19] Carreno-Quintero N, Bouwmeester HJ, Keurentjes JJ. Genetic analysis of metabolome-phenotype interactions: from model to crop species. *Trends Genet*, 2013, 29(1): 41-50
- [20] Mitchell-Olds T, Schmitt J. Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*. *Nature*, 2006, 441(7096): 947-52
- [21] Wolters H, Jurgens G. Survival of the flexible: hormonal growth control and adaptation in plant development. *Nat Rev Genet*, 2009, 10(5): 305-17
- [22] Tsuchiya Y, McCourt P. Strigolactones: a new hormone with a past. *Curr Opin Plant Biol*, 2009, 12(5): 556-61
- [23] Miyazono K, Miyakawa T, Sawano Y, et al. Structural basis of abscisic acid signalling. *Nature*, 2009, 462(7273): 609-14
- [24] Bassard JE, Ullmann P, Bernier F, et al. Phenolamides: bridging polyamines to the phenolic metabolism. *Phytochemistry*, 2010, 71(16): 1808-24
- [25] Fellenberg C, Ziegler J, Handrick V, et al. Polyamine homeostasis in wild type and phenolamide deficient *Arabidopsis thaliana* stamens. *Front Plant Sci*, 2012, 3(180): 1-8
- [26] Buer CS, Imin N, Djordjevic MA. Flavonoids: new roles for old molecules. *J Integr Plant Biol*, 2010, 52(1): 98-111
- [27] Keurentjes JJ. Genetical metabolomics: closing in on phenotypes. *Curr Opin Plant Biol*, 2009, 12(2): 223-30
- [28] Yang Z, Nakabayashi R, Okazaki Y, et al. Toward better annotation in plant metabolomics: isolation and structure elucidation of 36 specialized metabolites from *Oryza sativa* (rice) by using MS/MS and NMR analyses. *Metabolomics*, 2014, 10(4): 543-55
- [29] Forcat S, Bennett MH, Mansfield JW, et al. A rapid and robust method for simultaneously measuring changes in the phytohormones ABA, JA and SA in plants following biotic and abiotic stress. *Plant Methods*, 2008, 4(16): 1-8
- [30] Lin LZ, Harnly JM. A screening method for the identification of glycosylated flavonoids and other phenolic compounds using a standard analytical approach for all plant materials. *J Agric Food Chem*, 2007, 55(4): 1084-96
- [31] Monton MR, Soga T. Metabolome analysis by capillary electrophoresis-mass spectrometry. *J Chromatogr A*, 2007, 1168(1-2): 237-46
- [32] Chen W, Gong L, Guo Z, et al. A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. *Mol Plant*, 2013, 6(6): 1769-80

- [33] Sawada Y, Akiyama K, Sakata A, et al. Widely targeted metabolomics based on large-scale MS/MS data for elucidating metabolite accumulation patterns in plants. *Plant Cell Physiol*, 2009, 50(1): 37-47
- [34] Wishart DS, Knox C, Guo AC, et al. HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res*, 2009, 37(Database issue): D603-10
- [35] Dixon RA, Paiva NL. Stress-induced phenylpropanoid metabolism. *Plant Cell*, 1995, 7(7): 1085-97
- [36] Saito K, Yonekura-Sakakibara K, Nakabayashi R, et al. The flavonoid biosynthetic pathway in *Arabidopsis*: structural and genetic diversity. *Plant Physiol Biochem*, 2013, 72: 21-34
- [37] Dong X, Chen W, Wang W, et al. Comprehensive profiling and natural variation of flavonoids in rice. *J Integr Plant Biol*, 2014, 56 (9): 876-86
- [38] Dong X, Gao Y, Chen W, et al. Spatio-temporal distribution of phenolamides and the genetics of natural variation of hydroxycinnamoyl spermidine in rice. *Mol Plant*, 2015, 8(1): 111-21
- [39] Park HL, Lee SW, Jung KH, et al. Transcriptomic analysis of UV-treated rice leaves reveals UV-induced phytoalexin biosynthetic pathways and their regulatory networks in rice. *Phytochemistry*, 2013, 96: 57-71
- [40] Bowne JB, Erwin TA, Juttner J, et al. Drought responses of leaf tissues from wheat cultivars of differing drought tolerance at the metabolite level. *Mol Plant*, 2012, 5(2): 418-29
- [41] De Luca V, Salim V, Atsumi SM, et al. Mining the biodiversity of plants: a revolution in the making. *Science*, 2012, 336(6089): 1658-61
- [42] Huang X, Kurata N, Wei X, et al. A map of rice genome variation reveals the origin of cultivated rice. *Nature*, 2012, 490(7421): 497-501
- [43] Yu J, Hu S, Wang J, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science*, 2002, 296(5565): 79-92
- [44] Huang X, Feng Q, Qian Q, et al. High-throughput genotyping by whole-genome resequencing. *Genome Res*, 2009, 19(6): 1068-76
- [45] Huang X, Wei X, Sang T, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet*, 2010, 42(11): 961-7
- [46] Lu T, Lu G, Fan D, et al. Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. *Genome Res*, 2010, 20(9): 1238-49
- [47] Li H, Peng Z, Yang X, et al. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat Genet*, 2013, 45(1): 43-50
- [48] Zhang G, Guo G, Hu X, et al. Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res*, 2010, 20(5): 646-54
- [49] Suhre K, Gieger C. Genetic variation in metabolic phenotypes: study designs and applications. *Nat Rev Genet*, 2012, 13(11): 759-69
- [50] Fridman E, Pleban T, Zamir D. A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. *Proc Natl Acad Sci USA*, 2000, 97(9): 4718-23
- [51] Byrne PF, McMullen MD, Snook ME, et al. Quantitative trait loci and metabolic pathways: genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks. *Proc Natl Acad Sci USA*, 1996, 93(17): 8820-5
- [52] Kliebenstein DJ, Kroymann J, Brown P, et al. Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. *Plant Physiol*, 2001, 126(2): 811-25
- [53] Meyer RC, Steinfath M, Lisek J, et al. The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA*, 2007, 104(11): 4759-64
- [54] Matsuda F, Okazaki Y, Oikawa A, et al. Dissection of genotype-phenotype associations in rice grains using metabolome quantitative trait loci analysis. *Plant J*, 2012, 70(4): 624-36
- [55] Yu H, Xie W, Wang J, et al. Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. *PLoS one*, 2011, 6(3): e17595
- [56] Schauer N, Semel Y, Balbo I, et al. Mode of inheritance of primary metabolic traits in tomato. *Plant Cell*, 2008, 20(3): 509-23
- [57] Kliebenstein D. Advancing genetic theory and application by metabolic quantitative trait loci analysis. *Plant Cell*, 2009, 21(6): 1637-46
- [58] Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science*, 2000, 290(5494): 1151-5
- [59] Kliebenstein DJ, Lambrix VM, Reichelt M, et al. Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in *Arabidopsis*. *Plant Cell*, 2001, 13(3): 681-93
- [60] Rafalski JA. Association genetics in crop improvement. *Curr Opin Plant Biol*, 2010, 13(2): 174-80
- [61] Han B, Huang X. Sequencing-based genome-wide association study in rice. *Curr Opin Plant Biol*, 2013, 16(2): 133-8
- [62] Huang X, Han B. Natural variations and genome-wide association studies in crop plants. *Annu Rev Plant Biol*, 2014, 65: 531-51
- [63] 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. *Nat Genet*, 2012, 44(1): 32-9
- [64] Chan EK, Rowe HC, Corwin JA, et al. Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. *PLoS Biol*, 2011, 9(8): e1001125
- [65] Angelovici R, Lipka AE, Deason N, et al. Genome-wide analysis of branched-chain amino acid levels in *Arabidopsis* seeds. *Plant Cell*, 2013, 25(12): 4827-43
- [66] Riedelsheimer C, Lisek J, Czedik-Eysenberg A, et al.

- Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *Proc Natl Acad Sci USA*, 2012, 109(23): 8872-7
- [67] Matsuda F, Nakabayashi R, Yang Z, et al. Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. *Plant J*, 2015, 81(1): 13-23
- [68] Wen W, Li D, Li X, et al. Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat Commun*, 2014, 5: 3438
- [69] Khush GS. Origin, dispersal, cultivation and variation of rice. *Plant Mol Biol*, 1997, 35(1-2): 25-34
- [70] Chan EK, Rowe HC, Hansen BG, et al. The complex genetic architecture of the metabolome. *PLoS Genet*, 2010, 6(11): e1001198
- [71] McMullen MD, Kresovich S, Villeda HS, et al. Genetic properties of the maize nested association mapping population. *Science*, 2009, 325(5941): 737-40
- [72] Cavanagh C, Morell M, Mackay I, et al. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr Opin Plant Biol*, 2008, 11(2): 215-21
- [73] Okazaki Y, Saito K. Recent advances of metabolomics in plant biotechnology. *Plant Biotechnol Rep*, 2012, 6(1): 1-15