



植物代谢基因簇的研究进展

沈双欠^{1†}, 占传松^{2†}, 杨陈坤¹, 罗杰^{2*}

1. 华中农业大学生命科学技术学院, 作物遗传改良国家重点实验室, 武汉 430070;

2. 海南大学热带作物学院, 海南 572208

† 同等贡献

* 联系人, E-mail: jie.luo@hainanu.edu.cn

收稿日期: 2021-06-11; 接受日期: 2021-08-04; 网络版发表日期: 2021-11-23

国家杰出青年科学基金(批准号: 31625021)和国家自然科学基金创新研究群体科学基金(批准号: 31821005)资助

摘要 植物产生大量结构和功能多样的次生代谢物, 它们在植物生长发育和植物响应非生物和生物胁迫过程中发挥着不可或缺的作用. 近年来, 植物代谢基因聚集成簇现象受到了广泛关注, 而基因组、转录组和代谢组等多组学技术的快速发展使得高效挖掘植物代谢基因簇成为可能. 另一方面, 植物代谢基因簇研究也将为发掘和利用植物次生代谢物多样性提供新的见解. 本文不仅系统总结了植物基因簇的结构特征和调控方式, 同时也对相关新兴技术在代谢基因簇发掘及鉴定等方面的应用进行了综述.

关键词 植物代谢基因簇, 正向遗传学, 基因簇特征, 基因簇挖掘方法, 基因簇功能验证, 染色质调控

植物次生代谢物不仅在植物本身的正常生命活动中发挥重要作用, 对于人和动物的营养和健康也具有巨大价值. 固着生长的植物面临复杂多样的生长环境, 为了获得稳定的生存, 它们进化出丰富多样的化学物质库^[1,2]. 基于目前生物学技术, 植物中各类生物活性物质被发现和鉴定, 对其生物合成途径的解析和对植物化合物宝库改造并利用, 使其在现代医学、人类和动物健康、生物能源和农业等方面被充分应用, 是目前生物研究的热点领域.

近年来, 后基因组时代多组学技术的协同发展使得植物代谢物的研究取得突破性进展. 首先, 随着代谢组学检测技术的进步, 尤其是基于色谱质谱联用技术的迅速发展, 植物次生代谢物库得到极大丰富, 例

如, 在传统靶向和非靶向检测技术的基础上新发展的广泛靶向代谢组学(widely targeted metabolomics)^[3,4]. 此外, 核磁共振(nuclear magnetic resonance, NMR)^[5]、X-ray^[6]和电子冷冻显微镜法微晶电子衍射^[7]等技术也越来越多地应用于植物代谢物的结构解析. 其次, 新一代测序技术的突破和测序成本的降低, 使得通过识别潜在基因来发现新的自然代谢产物途径的能力增强^[8].

植物全基因组序列的测定和生物信息学技术的发展能够帮助建立植物表型和基因型之间的有效联系, 从而使得对大量性状全基因组关联分析(genome-wide association study, GWAS)成为可能^[9,10]. 同时, 比较基因组学在挖掘重要代谢途径的过程中也被有效使用. 例如, 研究人员利用水稻和玉米在遗传定位效力及定位

引用格式: 沈双欠, 占传松, 杨陈坤, 等. 植物代谢基因簇的研究进展. 中国科学: 生命科学, 2021, 51: 1734–1746

Shen S Q, Zhan C S, Yang C K, et al. Research progress on plant biosynthetic gene cluster (in Chinese). Sci Sin Vitae, 2021, 51: 1734–1746, doi: 10.1360/SSV-2021-0198

精度方面互补的特点, 通过比较全基因组关联分析, 同时提高了基因定位的效力和定位精度, 确定了近20个新的控制水稻种子代谢物含量的候选基因^[11]. 另外, 基因组、转录组学、蛋白组和代谢组学分析等多组学结合的研究策略也被广泛用于鉴定植物代谢物合成的相关基因^[12].

近年来的研究表明, 某些次生代谢物生物合成酶的编码基因在植物基因组中成簇分布, 形成代谢基因簇(表1)^[13~50]. 从目前已经报道的研究结果来看, 代谢基因簇在双子叶和单子叶植物的不同种类次生代谢物合成途径中广泛存在^[51]. 这种功能相关基因紧凑分布的现象改变了传统观念中植物代谢物合成途径编码基因在基因组中分散分布的认识, 为从植物基因组序列直接预测生物合成途径提供了新的可能. 本文将从植物代谢基因簇的定义、结构特点、预测和寻找方法、功能鉴定、调控模式等方面系统阐述植物代谢基因簇的研究进展.

1 植物代谢基因簇的特征

1.1 真核基因簇与原核操纵子的比较

细菌基因组约有二分之一是由操纵子组成. 操纵子是相关基因特别是代谢途径相关基因聚集的典型形式. 与此相反, 在真核生物中, 绝大多数功能相关基因并不会聚集而是分散在相应的基因组中. 有意思的是, 在越来越多的真核生物中发现了基因簇. 它们大部分属于非同源基因代谢基因簇. 尽管这些代谢基因簇与细菌操纵子表面上看都是由功能相关的、共表达的、具有物理连接的基因群组成, 但真核生物中的基因簇通常并非由细菌的水平基因转移引起, 而极有可能是物种演化过程中通过基因的复制和新功能化以及基因组的动态重组而产生的, 并且与细菌操作子根本性的不同之处是这些簇中的基因是单独转录的, 而不是作为单一的多顺反子^[52].

1.2 基因簇的组成成分

信号酶(也称之为特征酶, *signature enzymes*), 通常是基因簇负责合成次生代谢物骨架的酶. 这些酶通常可以清晰地表明基因簇合成化合物的类别. 例如, 氧化氧化角鲨烯环化酶用于合成三萜, I类和II类二萜合成酶用于合成二萜类物质的支架. 另外, 催化第一

步的酶也可能是将代谢产物从初生代谢途径转移到专门次生代谢途径的酶. 例如, 色氨酸合酶 α 同系物Bx1, 催化DIBOA/DIMBOA生物合成的第一步^[43]; 属于CYP79家族的细胞色素P450酶在催化生物合成cyano-genic glycoside的第一步反应中, 将氨基酸转化为胍^[40]; 水稻中酪氨酸脱羧酶OsTyDC1(*tyrosine decarboxylase*)催化酪氨酸合成酪胺, 催化合成酰基化的羟基肉桂酸的起始步骤^[46]. 此外, 基因簇的起始酶还有可能是从复杂的代谢网络中吸取中间产物用于专门合成特定终产物的途径^[41]. 值得注意的是, 也存在一些特殊情况, 编码代谢途径起始步骤的酶与该途径其他的基因并没有紧密聚集. 例如, 番茄甾体糖生物碱生物合成途径中, CYP72家族的基因GAME7编码的酶负责催化该代谢途径的第一步反应, 其与基因簇的其他组分位于同一染色体上, 但并未紧邻而是相距甚远^[42].

次生代谢物的多样性在很大程度上是由于化合物骨架的化学修饰造成的, 而代谢基因簇中除了编码代谢途径起始的信号酶的基因外, 还通常包含编码各种修饰酶的基因, 如细胞色素氧化酶、糖基转移酶、酰基转移酶、甲基转移酶、双加氧酶、羧酸酯酶、脱氢酶/还原酶和转氨酶等. 修饰酶的种类和数量往往可以影响基因簇的大小. 通常基因簇中修饰酶的种类越丰富, 数量越多, 该基因簇往往就越大. 迄今为止, 已知的植物代谢基因簇中既有30多kb的, 也有上百kb的. 基因簇中包含的基因数量和酶的种类也不尽相同. 基因簇中包含的基因数量少则3个多则12个, 而酶的种类则通常多于3种^[51].

植物次生代谢基因簇也可能包含非核心酶类的基因, 包括编码负责代谢产物转运的转运子基因和辅酶合成酶等. 研究人员近期也发现, 高粱氰化糖苷dhur-rin的基因簇也包含一个定位在液泡膜上的基因SbMATE2, 它编码多药和有毒化合物外排转运蛋白(*multidrug and toxic compound extrusion*, MATE). 爪蟾卵母细胞的转运研究表明, SbMATE2能够转运dhur-rin^[53]. 5'-磷酸吡哆醛(*pyridoxal 5'-phosphate*, PLP)辅助因子在转氨酶、异构酶、脱羧酶、消旋酶、醛缩酶、脱氨基酶、裂解酶和氨基变位酶等核心酶发挥催化作用时都是必需的. 最近研究表明, 苯氧胺5'-磷酸盐氧化酶OsPDX3(*pyridoxamine 5'-phosphate oxidase*)位于水稻催化生成芳香族酚胺的基因簇中, 它催化生成的PLP作为酪氨酸脱羧酶发挥催化功能依赖的辅助因

表 1 已报道的植物代谢基因簇

Table 1 Reported plant biosynthetic gene clusters

物质类别	物质名称		来源物种		参考文献
	中文名称	英文名称	中文名称	拉丁文名称	
单萜	—	β -Phellandrene	番茄	<i>Solanum lycopersicum</i>	[13]
二萜	—	Lycosantalanol	番茄	<i>Solanum lycopersicum</i>	[14,15]
二萜	Casbene二萜	Casbene diterpenoids	南欧大戟 麻疯树	<i>Euphorbia peplus</i> <i>Jatropha curcas</i>	[16]
二萜	Casbene二萜	Casbene diterpenoids	蓖麻	<i>Ricinus communis</i>	[17]
二萜	Casbene二萜	Casbene diterpenoids	水稻	<i>Oryza sativa</i>	[18,19]
二萜	稻壳酮	Momilactones	大灰藓 稗草 水稻	<i>Calohypnum plumiforme</i> <i>Echinochloa crus-galli</i> <i>Oryza sativa</i>	[20~23]
二萜	—	Phytocassanes Oryzalides	水稻	<i>Oryza sativa</i>	[24]
二萜	—	Zealexin	玉米	<i>Zea mays</i>	[25]
二倍半萜	—	Sesterterpenoids	拟南芥	<i>Arabidopsis thaliana</i>	[26]
三萜	燕麦素	Avenacins	燕麦	<i>Avena strigosa</i>	[27~30]
三萜	—	Thaliano Marneral	拟南芥	<i>Arabidopsis thaliana</i>	[31,32]
三萜	—	Tirucalla-7,24- dien-3b-ol	拟南芥	<i>Arabidopsis thaliana</i>	[17]
三萜	—	Arabidiol	拟南芥	<i>Arabidopsis thaliana</i>	[33,34]
三萜	葫芦素C,B,E	Cucurbitacins C,B,E	黄瓜	<i>Cucumis sativus</i>	[35,36]
三萜	—	Thalianol	<i>Lyrata</i> 拟南芥	<i>Arabidopsis lyrata</i>	[37]
三萜	表大戟 二烯醇	Tirucallol	芥菜	<i>Capsella rubella</i>	[38]
三萜	20-羟基 白桦酸	20-Hydroxy -betulinic acid	百脉根	<i>Lotus japonicus</i>	[39]
糖苷类化合物	亚麻苦甙 百脉根甙	Linamarin lotaustralin	百脉根 木薯	<i>Lotus japonicus</i> <i>Manihot esculenta</i>	[40]
	蜀黍苷	Dhurrin	高粱	<i>Sorghum bicolor</i>	
苄基异喹啉类生物碱	诺司卡品	Noscapine	罂粟	<i>Papaver somniferum</i>	[41]
甾体生物碱	α -茄碱	α -Tomatine	番茄	<i>Solanum lycopersicum</i>	[42]
		α -Chaconine, α -Solanine	马铃薯	<i>Solanum tuberosum</i>	
异羟肟酸	丁布	DIMBOA	玉米	<i>Zea mays</i>	[43~45]
	丁布	DIMBOA	稗草	<i>Echinochloa crus-galli</i>	[22]
酚胺	羟基肉桂酰酪胺	Hydroxycinnamoyl-tyramine	水稻	<i>Oryza sativa</i>	[46]
酚胺	羟基肉桂酰腐	Hydroxycinnamoyl-putrescine	水稻	<i>Oryza sativa</i>	[47]
聚酮化合物	β -二酮	β -diketones	大麦	<i>Hordeum vulgare</i>	[48]
修饰脂肪酸	法卡林二醇	Falcarindiol	番茄	<i>Solanum lycopersicum</i>	[49]
糖脂脂肪族酯类	中链酰基糖	Medium chain acylsugar	番茄	<i>Solanum lycopersicum</i>	[50]

子^[46]。随着越来越多的代谢基因簇被发掘, 基因簇中的修饰酶的类型将会变得更加丰富^[54]。

1.3 基因簇的结构特点

基于植物代谢基因簇中基因的排列方式, 本文将植物代谢基因簇的结构特点归纳成如下两种类型(图1)。

(1) 完全紧密排列的基因簇. 这类基因簇包含催化第一步反应的特征酶和负责下游催化步骤的修饰酶, 并且这些组分相邻排布在基因组上. 这些紧凑型基因簇中, 代谢途径中的绝大多数基因是连续的, 没有其他非酶类基因. 例如, 80%拟南芥生态型中的Thalianol植物代谢基因簇^[38]、水稻中酰基化酪胺基因簇^[46]和二萜casbene基因簇^[18]都属于紧凑型排布的基因簇。

(2) 部分紧密相邻的核心基因簇组分与分散组分. 与常规的紧密排列的基因簇不同, 部分基因簇组分的分布并非完全聚集在一起, 也有可能是其中大部分聚集, 而另外也有一些起始酶、修饰酶或是转录调控因子的组分分布在较远的区域. 例如, 在黄瓜中, 用于合成葫芦二烯醇的氧化喹啉环化酶基因、3种不同类型的CYP基因(CYP81, CYP87, CYP89)和一个酰基转移酶基因, 这5个编码葫芦素生物合成所需酶(产生苦味的三萜类)的基因聚集在6号染色体上. 而葫芦素生物合成所需的另外四个CYP基因位于其他染色体上(3号染色体上有一个CYP71基因和两个CYP88基因簇, 1号染色体上有一个CYP87基因簇). 另外, 分别调节叶片和果实中葫芦素C合成的两个bHLH转录因子Bl(苦叶)和Bt(苦果)与这些葫芦素的核心生物合成簇和外周代谢基因分别位于不同的染色体上^[35]. 大多数番茄中合成甾醇生物碱 α -茄碱的基因在7号染色体上聚集, 其中包括两个早期途径基因-双加氧酶基因GAME11和CYP72基因GAME6以及编码四种糖基转移酶的基因GAME1, GAME17, GAME18和GAME2, 而另外两个结构基因——编码氨基转移酶GAME12和CYP88D基因GAME4在12号染色体上彼此相邻. α -茄碱代谢途径的第一步——胆固醇转化为22-羟基胆固醇被预测是CYP72基因GAME7催化的, GAME7也位于7号染色体上, 但距离最近的簇状基因8 Mb, 与核心基因簇形成较为松散的基因聚集结构^[42]. 另外, 在马铃薯中合成甾醇生物碱的相关基因也聚集分布在7号和12号染色体上, 与番茄中的基因簇具有相似的组织结构, 近源物

种具有类似结构基因簇的现象说明这些物质的代谢合成途径似乎具有共同的进化起源, 且在西红柿和土豆中获得了不同的功能, 因而这些基因簇的揭示将有助于确定植物代谢基因簇在拆分、组装和多样化过程中的进化规律。

2 识别基因簇的方法和策略

2.1 基于基因簇特征的挖掘

(1) 结合植物基因组和转录组数据挖掘代谢基因簇. 测序技术的发展使得大量的植物基因组序列被注释与发布, 据初步统计, 目前已经约有400多种植物的基因组(<https://www.plabipd.de/index.ep>)公布. 大量的植物基因组信息为解析植物代谢途径提供了强有力的原始数据支持. 例如, 在拟南芥基因组信息公布后, Osbourn和Matsuda等科学家就成功通过对拟南芥基因组基因信息挖掘, 找到并鉴定出4个不同的三萜基因簇^[17,31~32]. 共表达是代谢基因簇共有的特征之一. RNA-seq技术的发展使得人们可以获得多种组织类型、不同发育阶段或不同处理条件下的多个基因的表达量数据^[55]. 因此, 在有参考基因组的植物中通过转录共表达分析来筛选有效候选基因以及代谢基因簇已成为筛选生物合成基因和代谢基因簇的最有力方法之一. 例如, 酰基糖是茄科植物的腺毛中特异积累的防御性物质, 通过在番茄中相应组织的特进行共表达分析, 发现在番茄7号染色体上的酰基辅酶A合成酶基因(*Solyc07g043680*)和烯醇辅酶A水合酶基因(*Solyc07g043630*), 与12号染色体上已经报道的BAHD酰基转移酶基因*Sl-ASAT1*构成代谢基因簇协同作用, 共同参与中链长度的酰基糖积累^[50]. 类似的茄科植物中的 α -茄碱, 是人体的抗营养因子. 科学家们通过结合番茄和马铃薯的基因比较共表达分析和酶学功能验证, 揭示了参与甾体甾生物碱(solanum steroidal glycoalkaloids, SGA)生物合成的10个合成基因, 其中6个在第7号染色体上以基因簇的形式存在, 而另外两个在第12号染色体的基因组同源区段中相邻. 其中, GAME4可以调控SGAs的积累, 抑制GAME4(glycoalkaloid metabolism 4)的表达能够显著抑制马铃薯块茎和番茄果实中SGAs的积累. 这为人们通过分子手段减少茄科植物广泛存在的不安全、抗营养物质——茄碱提供了可能^[42]。

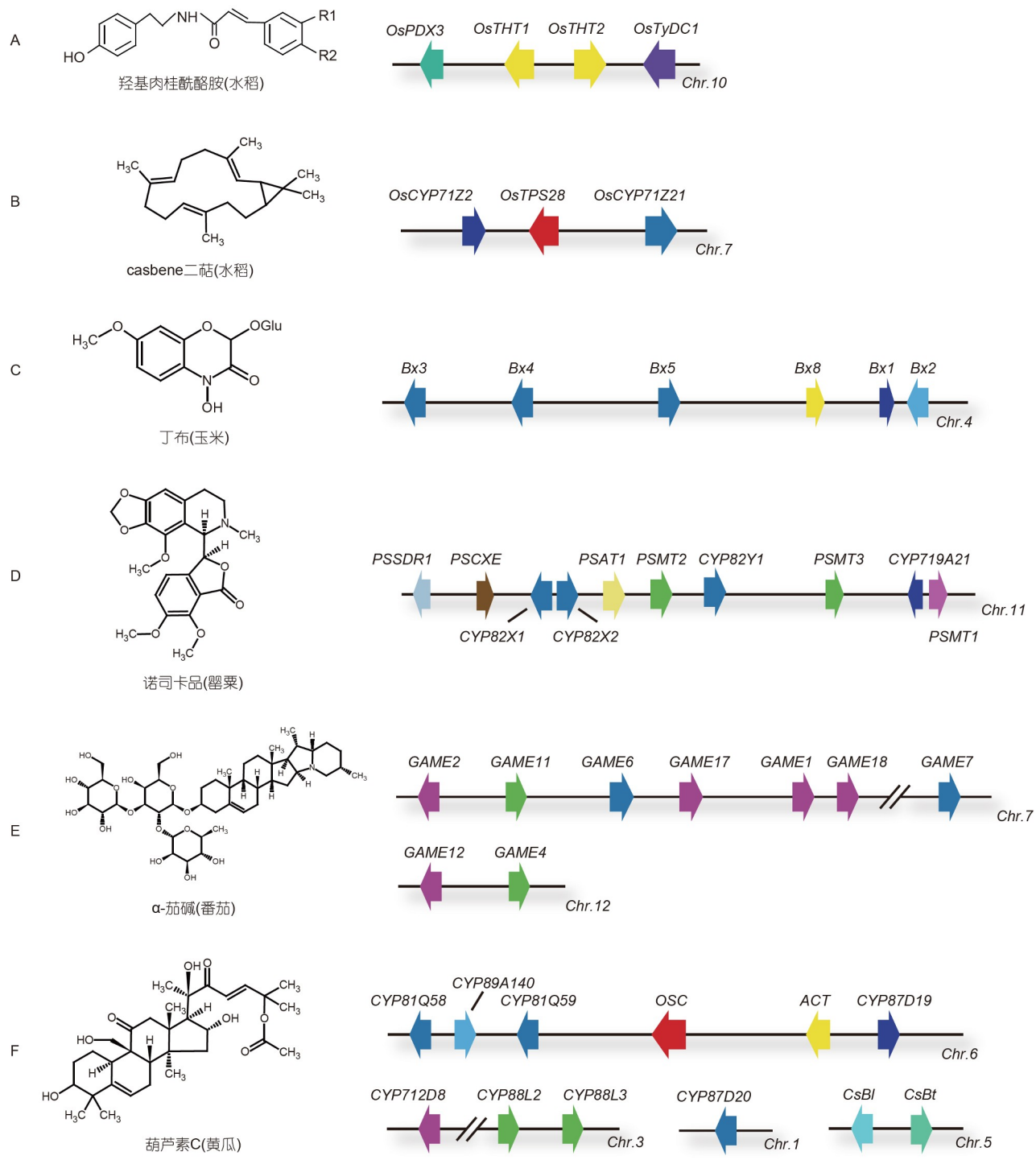


图 1 基因簇的结构特点。A~D: 完全紧密排列的基因簇——(A)水稻的羟基肉桂酰胺基因簇, (B)水稻的casbene二萜基因簇, (C)玉米的丁布基因簇, (D)罂粟的诺思卡品基因簇; E和F: 部分紧密相邻的核心基因簇组分与分散组分——(E)番茄的茄碱基因簇, (F)黄瓜的葫芦素C基因簇。Chr, 染色体

Figure 1 Types of biosynthetic gene cluster organization. A~D: Completely tightly packed gene clusters—(A) the hydroxycinnamyl tyramine gene cluster in rice, (B) the diterpenoid casbene cluster in rice, (C) the DIMBOA gene cluster in maize, (D) the noscapine cluster in poppy; E and F: loose gene clusters—(E) the α -solanine cluster in tomato, (F) the Cucurbitacins C gene cluster in cucumber. Chr, chromosome

除了基于基因时空表达模式上的相似性, 另一种基于代谢途径基因诱导共表达挖掘基因簇的方法, 近期被用于阐明高度修饰的脂肪酸Falcarindiol的生物合成途径. Falcarindiol是一种存在于番茄、胡萝卜和人参中的乙酰脂质. 用不同的真菌和细菌病原体或诱导子诱导Falcarindiol的产生, Falcarindiol的积累与相应基因的转录水平密切相关. 这一分析揭示了一个新的控制Falcarindiol生物合成的基因簇, 并鉴定了具有罕见乙酰化酶活性的ACET1a/b^[49]. 在这三个代表性的研究中, 对茄科酰基糖、茄碱和Falcarindiol生物合成的阐述说明了共表达分析方法在植物中发现代谢基因簇的可行性.

(2) 利用算法与程序批量挖掘代谢基因簇. 研究人员基于大量已经发表的植物基因组和转录组信息, 开发出一些系统的算法来有效识别植物中可能的代谢基因簇^[56,57]. 这些算法的检索规则大多是运用典型的基因簇起始酶和修饰酶的组合形式来筛选候选代谢基因簇, 例如, 萜烯合成酶(terpene synthase, TPS)和细胞色素P450酶(cytochrome P450 enzyme system, CYP450)组合^[58]. 如果在某个物种的基因组邻近区段同时出现一个或者多个TPS/CYP450组合且这些基因在转录水平上共表达, 那就可以将它们视为基因簇的候选成员. 这种方法的优点主要在于能够较快地找到候选代谢基因簇并进行相关验证. 为了确定代谢途径基因成簇分布的现象是否普遍存在于开花植物和低等植物中, Schl pfer等人^[59]开发了一种新的预测软件Plant Cluster Finder分析了18种植物, 包括一种绿藻、两种低等植物和15种高等植物, 检测到11969个候选的代谢基因簇, 其中1700个以上的基因簇可能与特定代谢物的产生有关. 类似地, PlantSMASH^[60]和PhytoClust^[58]等算法也相继被开发出来, 这两种工具使用更为精确的隐马尔科夫模型(profile Hidden Markov Models, pHMMs)^[61]判断不同的生物合成酶并结合基因组位置来预测候选基因簇. 另外, 当预测模型中结合了基因共表达数据时, 候选基因簇的数量会大大减少, 同时预测的准确性也会显著提高^[60]. 总之, 相信随着更多的基因簇特征被解析和对计算机算法的不断更新, 人们对于基因簇的预测能力也会逐渐增强. 这些发展必将为天然代谢产物代谢途径关键酶的挖掘提供更多的便利.

2.2 基于已知来挖掘邻近未知

自然界分布着多种多样的代谢物, 其生物合成往往需要许多催化步骤. 但是由于实验技术和基因组信息的局限性, 人们对代谢物代谢途径的解析通常是一个循序渐进的过程, 因此, 可以根据已经报道的一个或多个代谢途径, 结合逐步更新的数据来挖掘潜在的基因簇. 大戟科植物产生多种二萜类化合物, 其中许多具有药理活性. 蓖麻的casbene合成酶是第一个被克隆的萜烯环化酶, 随后在其他大戟科植物中被鉴定, 然而casbene衍生二萜的生物合成还需要一系列的氧化反应、环闭合反应以及酰基化修饰反应^[16]. 研究人员根据这些生化反应特征, 报道了蓖麻属植物基因组中的细胞色素P450s的生化功能, 包括CYP726A14, CYP726A17和CYP726A18能够催化casbene的5号位氧化, CYP726A16催化5-keto-casbene的7,8-环氧化反应, CYP726A15催化neocembrene的5号位氧化反应. 这些基因与casbene和neocembrene合成酶、一个BAHD酰基转移酶和两个类似醇脱氢酶的酶共同构成了蓖麻中二萜casbene生物合成的代谢基因簇. 类似地, 水稻中稻壳酮途径最先只解析了下游途径中的两步催化反应, 分别是由syn-CDP synthase(OsCPS4)催化牻牛儿基牻牛儿基焦磷酸(geranylgeranyl pyrophosphate, GGPP)形成syn-Copalyl diphosphate, 进而由9 H-pimara-7,15-diene synthase(OsKS4)催化形成9 H-Pimara-7,15-diene, 并且这两个基因位于水稻4号染色体上的相邻位置. 结合稻壳酮的结构特点分析发现, 还需要脱氢酶和氧化酶才能催化其完整的生物合成. 详细分析水稻4号染色体的基因信息发现一个脱氢酶基因(OsMAS)和两个P450基因(CYP99A2和CYP99A3)恰好位于OsKS4和OsCPS4附近, 最终通过实验验证了这些基因的生化功能, 并且这些基因共同组成了稻壳酮代谢基因簇^[20]. 另外, 近期有研究表明, 稻壳酮基因簇不仅存在于水稻中, 它也保守地存在于稗草和大灰藓基因组中^[21].

2.3 基于正向遗传学策略挖掘基因簇

除了上文提及的筛选和鉴定植物代谢基因簇的方法外, 运用自然变异以及其他基于相关性的正向遗传学的方式来寻找和进一步确认植物代谢基因簇也是一种有效手段. 采用此种研究方式来寻找植物代谢基因

簇的通常都是基于两个或多个亲本间杂交产生的重组自交系F₂等群体的连锁图谱, 如限制性片段长度多态性和简单序列重复标记。当然, 在条件许可的情况下使用由下一代测序产生的高密度图谱进行这类研究, 可以显著提高基于代谢产物的定量位点(metabolites quantitative trait loci, mQTLs)图谱分辨率^[62]。例如, Schneider等人^[48]在大麦中利用群体定位的方法将Cercu位点定位到染色体臂上的一个离散位置, 结合BAC文库数据和突变体测序确定了3个基因*Cer-c*、*Cer-q*和*Cer-u*。进一步分析表明, 它们是催化 β -diketone生物合成的代谢基因簇。类似地, 研究人员在一个丰富积累诺司卡品的罂粟品种HN1中发现10个编码5种不同酶类的基因特异表达。对F₂作图群体的分析表明, 这些基因在HN1中紧密相连, BAC测序进一步证实了它们作为一个复杂的植物生物碱基因簇存在^[41]。

除了mQTLs定位外, 正向代谢物突变体筛选也是另一种有效的应用于代谢物合成基因和代谢基因簇的挖掘的手段。燕麦素(avenacin A-1)在紫外线照射下, 燕麦根会发出明亮的蓝色荧光, 研究人员利用这种荧光特征识别了在燕麦素合成受损的二倍体燕麦突变体, 并且利用这个突变体发现并鉴定出燕麦素合成的代谢基因簇^[27,63]。

GWAS是一种新兴的可以同时大量性状结合植物群体单核苷酸多态性进行关联分析以寻找候选基因的有效方法。例如, 科研人员通过对115个黄瓜品种的苦味性状进行关联分析, 定位到参与控制黄瓜苦味素生物合成的第一步的催化酶*Bi*基因, 之后利用基因组学数据并结合代谢组学、遗传学、分子生物学等多种研究手段, 揭示了9个负责黄瓜苦味物质生物合成的基因, 其中包含*Bi*在内的6个基因, 它们在6号染色体上聚集排列, 形成代谢基因簇^[35]。与其他性状相比, 代谢性状往往由相对较少且影响较大的基因座控制, 因此基于代谢物含量的全基因组关联分析(metabolome-based genome-wide association studies, mGWAS)方法对于识别潜在途径的效应位点通常效果显著, 在这些位点下进一步结合已有的基因组信息, 分析候选基因所在的染色体区段来识别候选基因簇则是另外一种寻找植物代谢基因簇的有效方式。例如, 研究人员对600余份番茄材料的代谢物含量进行全基因组关联分析, 发现茄碱的自然变异受5个主要遗传位点控制, 且这些位点在驯化及改良过程中受到强烈选择。

对其中效应最大的10号染色体的位点进行分析, 发现一个潜在的茄碱代谢基因簇, 且在这个潜在的代谢基因簇的糖基转移酶基因(*Solyc10g085230*)外显子上发现一个可以造成提前终止的自然变异^[12]。关联分析的方法优点在于能够有效定位复杂性状的遗传位点, 从而较为准确地定位到候选基因^[3]。相信随着基因组信息的丰富, 全基因组关联分析的方法与其他研究方法的结合使用将在植物次生代谢物的合成途径解析方面发挥重要作用。

3 植物代谢基因簇中合成基因的功能验证

基因组信息的大量涌现意味着更大数量的基因(多为代谢合成基因)功能有待解析。总体而言, 验证代谢合成基因功能的方法可以分为两类: 一是通过异源表达的方式验证基因的功能, 二是通过转基因或者构建突变体材料的方式验证基因的功能。对于基因的功能验证通常需要结合以上两种方法, 而对于一些暂时没有成熟的稳定转基因体系的植物而言, 第一种方式将是首要的选择, 同时可以结合在模式植物中异源稳定转化的方法共同验证。

3.1 异源表达验证

异源表达验证常见的宿主有以下几种: 大肠杆菌、酵母、本氏烟草^[64-67]。

大肠杆菌作为最常见的异源表达微生物, 被广泛应用于在植物基因簇基因功能验证过程中, 如玉米中的DIMBOA-Glu代谢基因簇^[43,44], 水稻中的植物卡生A-E(phytocassanes A-E)基因簇^[24]、稻壳酮基因簇^[16]和Oryzalide A基因簇^[68], 稗草中的稻壳酮基因簇^[21], 柴胡中的氰苷基因簇^[40], 蓖麻中的casbene二萜基因簇^[16]等。

酵母也已被证明是另一种有效的外源宿主, 特别是在真核生物膜蛋白的异源表达和基因功能验证中具有不可替代的优势。因此, 编码细胞色素P450家族酶和异戊烯基氧化酶类膜蛋白的基因簇基因, 其功能验证都会优先选择酵母表达体系, 如拟南芥中Thalianol三萜基因簇中的THAD(*CYP705A5*)和THAH(*CYP708A2*)^[31], 黄瓜葫芦素代谢合成基因簇中的*Csa3g903540*, *Csa6g088160*和*Csa6g0887007*^[35]以及葫芦科中其他三萜基因簇的CYP450基因的功能验证^[36],

水稻二萜基因簇中的 *CYP99A2&3*^[69], 水稻 5,10-diketo-casbene 基因簇^[18]. 另外, 包括罂粟碱在内的多种植物天然产物都被验证可以在酵母中异源合成^[35,70-74].

烟草瞬时表达体系是英国 Lomonosoff 教授团队^[65,67]开发的一套能在烟草中瞬时表达异源基因蛋白的蛋白瞬时表达体系. 作为异源蛋白表达宿主, 烟草具有多种优势, 其生长周期较短, 同时其支持 mRNA 和蛋白质加工且具有许多不可或缺的代谢前体与辅酶, 因而与微生物相比其更合适植物基因的异源表达. 这套体系目前已被广泛应用于基因功能验证和合成生物学等多个领域^[37,75-77], 植物中代谢基因的功能验证也通常会用到该体系. 例如, 大戟科植物中单环二萜基因簇中的 *CYP450* 基因^[16,17,78], 拟南芥和黄瓜中三萜基因簇中的 *CYP450* 基因^[17], 稗草中的稻壳酮基因簇基因^[21], 柴胡中的氰苷基因簇中的 *BGD2&4*^[79].

除烟草瞬时表达外, 一种基于在植物体内诱导基因沉默(virus-induced gene silencing, VIGS)的方式来验证基因功能的方法也常常在基因簇基因功能验证中发挥不可替代的作用^[80]. 例如, 罂粟碱和酰基糖基因簇的基因的功能验证就有效地使用了 VIGS 这种方式^[41,50].

3.2 遗传材料验证法

构建相应基因的遗传材料是在植物体内验证基因功能的可靠方式. 这种方法已在拟南芥^[51,32]、燕麦^[27]、玉米^[45]、百脉根^[40]、番茄^[42]、芥菜^[81]等常见作物的基因簇功能验证过程中发挥了重要作用. 随着相应转基因技术体系的成熟^[82], 相信不久的将来这种方法将在基因簇的基因功能验证过程中越来越受重视.

4 植物代谢基因簇的调控

研究表明, 基因簇基因在植物特定部位或植物处于某种环境中的共转录现象, 因此对植物基因簇协同调控机制的研究尤为重要. 但是, 相比于植物代谢基因簇的生化功能的解析, 这些植物代谢基因簇的调控研究就明显相对滞后. 到目前为止, 共有约 30 多个植物代谢基因簇的代谢途径被解析, 但是其中仅有个位数的基因簇的调控机制研究深入到了转录因子水平或者染色质水平.

4.1 转录因子调控

稻壳酮 A&B(momilactone A&B), 植物卡生 A-E, Oryzalide A-C, 稻叶素 A-F(Oryzaalexin A-F)是水稻中的二萜类植保素, 它们的合成与水稻中两个经典的二萜基因簇紧密相关^[24,69]. Oryzalides 在未受到病虫害感染的水稻植株中广泛存在, 稻壳酮仅在水稻根中结构性合成, 其余的都是在病原体感染或者诱导子处理之后大量合成^[83]. 这就暗示着水稻二萜的合成可能受到多种不同因素的调控. OsTGAPI 是一个 bZIP 家族的转录因子, 它可以间接协同调控稻壳酮、植物卡桑以及 Oryzalides 二萜基因簇基因转录进而调节相应二萜的合成^[84]. bHLH 和 WRKY45 也在水稻二萜基因簇基因的转录调控过程中发挥着重要的功能^[85,86]. 转录因子 Bt 和 BI 可以分别在黄瓜果实和叶片中协同调控葫芦素 C 基因簇基因的表达进而调控黄瓜的苦味. Bt 基因和 BI 基因相邻但仅 Bt 基因在黄瓜驯化过程中受到了选择, 使得现在的黄瓜栽培品种的果实吃起来不苦^[35]. 类似的 bHLH 转录因子调控葫芦素的合成不仅出现在黄瓜中, 同时也出现在甜瓜和西瓜中^[36]. 茄科中茄碱基因簇的调控是另一个研究得比较深入的事例. GAME9 (AP2) 转录因子可以调控茄碱基因簇基因的转录. GAME9 不仅可以调控 GMAE4&7, 而且也可以调控茄碱合成通路上游的 HMGR 以及 CS-SD 基因, 更有意思的是, GAME9 可以与另外一个转录因子 MYC2 结合, 共同完成相关调控^[87].

4.2 染色质调控

人们发现, 真菌中基因簇基因的激活或抑制受染色质状态的影响^[88,89]. 随后发现, 这种现象不仅存在于真菌中, 在植物中也同样如此^[90]. 拟南芥、玉米、水稻和燕麦中的萜类基因簇存在明显的甲基化修饰(H3K27me3)和乙酰化修饰(H2A.Z), H3K27me3 的修饰水平与基因簇的表达呈现明显的负相关而与乙酰化(H2A.Z)呈现明显的正相关^[90,91]. 此外, 另一个关于植物基因簇表观调控的研究表明, 水稻中的去甲基化酶 JMJ705 可以通过茉莉酸甲酯的介导直接调控水稻中蓖麻烯二萜基因簇相关基因的甲基化修饰水平, 从而调控基因簇基因的表达^[23]. 最近的研究从三维基因组水平上揭示了拟南芥三萜基因簇(thalianol cluster, marneral cluster 和 arabidiol cluster)、水稻二萜基因簇

(momilactone cluster和phytocassane cluster)、玉米DI-BOA基因簇、番茄茄碱基因簇在激活和抑制(沉默)状态时, 它们相关结构域的空间构象不同^[92]。

5 总结与展望

在植物特异性代谢研究中, 通常是先确定目标化合物, 在研究代谢途径构成过程中通过搜索基因组信息发现对应的植物代谢基因簇。然而, 在基因簇挖掘的早期, 由于许多植物的基因组序列信息还未公布, 科学家们主要通过传统的构建近等基因系结合筛选文库的方法进行识别和鉴定^[43], 但是这种方法成本高且效率低。组学技术的协同快速发展以及人们对于植物代谢途径知识越来越多地了解, 使得近20年来植物代谢基因簇的研究进入新时代^[49-52,93](表1)。可以想象, 在不久的将来包含GWAS在内的关联分析技术^[35,94]、烟草瞬时表达技术^[67,76,77]、快速高效转基因技术(利用磁性纳米粒子作为基因载体进行植物转

化)^[82]以及合成生物学技术^[35,71]的交叉融合必将使植物代谢基因簇的解析、开发与利用进入高速发展的快车道。目前, 这一多种技术融合发展领域已有一些研究报道。例如, 斯坦福大学的Jeon等人^[49]结合代谢组与转录组数据成功鉴定了番茄中一个负责合成高度修饰脂类的代谢基因簇。越来越多的基因簇及其特征被揭示, 也将进一步为生物信息学开发相应算法和预测程序提供宝贵的基础数据, 也许在不久的将来, 可以运用生物信息学技术对植物代谢甚至是非代谢基因簇进行准确预测。但是, 随着预测植物代谢基因簇方法的开发, 越来越多的终产物未知的基因簇被发现且尚未进行功能验证, 表明以后相关研究可以多关注代谢基因簇的功能解析, 从而加深人们对不同化合物合成途径的认识。这种从基因簇到化合物功能的研究策略被广泛使用, 十字花科植物中二倍半萜基因簇的研究就是一个典型的样例^[26]。相信随着化合物结构的广泛鉴定和化合物生理活性筛选方法的开发, 这方面的研究会越来越多。

参考文献

- Jacobowitz J R, Weng J K. Exploring uncharted territories of plant specialized metabolism in the postgenomic era. *Annu Rev Plant Biol*, 2020, 71: 631-658
- Fang C, Fernie A R, Luo J. Exploring the diversity of plant metabolism. *Trends Plant Sci*, 2019, 24: 83-98
- Luo J. Metabolite-based genome-wide association studies in plants. *Curr Opin Plant Biol*, 2015, 24: 31-38
- 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: 1769-1780
- Boiteau R M, Hoyt D W, Nicora C D, et al. Structure elucidation of unknown metabolites in metabolomics by combined NMR and MS/MS prediction. *Metabolites*, 2018, 8: 8
- Wagner G, Hyberts S G, Havel T F. NMR structure determination in solution: a critique and comparison with X-ray crystallography. *Annu Rev Biophys Biomol Struct*, 1992, 21: 167-198
- Jones C G, Martynowycz M W, Hattne J, et al. The cryoEM method microED as a powerful tool for small molecule structure determination. *ACS Cent Sci*, 2018, 4: 1587-1592
- Bevan M W, Uauy C, Wulff B B H, et al. Genomic innovation for crop improvement. *Nature*, 2017, 543: 346-354
- Edwards D, Batley J. Plant genome sequencing: applications for crop improvement. *Plant Biotech J*, 2010, 8: 2-9
- Hu L, Xu Z, Wang M, et al. The chromosome-scale reference genome of black pepper provides insight into piperine biosynthesis. *Nat Commun*, 2019, 10: 4702
- 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
- Zhu G, Wang S, Huang Z, et al. Rewiring of the fruit metabolome in tomato breeding. *Cell*, 2018, 172: 249-261.e12
- Falara V, Akhtar T A, Nguyen T T H, et al. The tomato terpene synthase gene family. *Plant Physiol*, 2011, 157: 770-789
- Matsuba Y, Nguyen T T H, Wiegert K, et al. Evolution of a complex locus for terpene biosynthesis in *Solanum*. *Plant Cell*, 2013, 25: 2022-2036
- Matsuba Y, Zi J, Jones A D, et al. Biosynthesis of the diterpenoid lycosantalanol via nerylneryl diphosphate in *Solanum lycopersicum*. *PLoS*

- ONE, 2015, 10: e0119302
- 16 King A J, Brown G D, Gilday A D, et al. Production of bioactive diterpenoids in the euphorbiaceae depends on evolutionarily conserved gene clusters. *Plant Cell*, 2014, 26: 3286–3298
- 17 Boutanaev A M, Moses T, Zi J, et al. Investigation of terpene diversification across multiple sequenced plant genomes. *Proc Natl Acad Sci USA*, 2015, 112: E81–E88
- 18 Zhan C, Lei L, Liu Z, et al. Selection of a subspecies-specific diterpene gene cluster implicated in rice disease resistance. *Nat Plants*, 2020, 6: 1447–1454
- 19 Liang J, Shen Q, Wang L, et al. Rice contains a biosynthetic gene cluster associated with production of the casbane-type diterpenoid phytoalexin *ent*-10-oxodepressin. *New Phytol*, 2021, 231: 85–93
- 20 Shimura K, Okada A, Okada K, et al. Identification of a biosynthetic gene cluster in rice for momilactones. *J Biol Chem*, 2007, 282: 34013–34018
- 21 Mao L, Kawaide H, Higuchi T, et al. Genomic evidence for convergent evolution of gene clusters for momilactone biosynthesis in land plants. *Proc Natl Acad Sci USA*, 2020, 117: 12472–12480
- 22 Guo L, Qiu J, Ye C, et al. *Echinochloa crus-galli* genome analysis provides insight into its adaptation and invasiveness as a weed. *Nat Commun*, 2017, 8: 1031
- 23 Wilderman P R, Xu M, Jin Y, et al. Identification of *Syn-pimara*-7,15-diene synthase reveals functional clustering of terpene synthases involved in rice phytoalexin/allelochemical biosynthesis. *Plant Physiol*, 2004, 135: 2098–2105
- 24 Swaminathan S, Morrone D, Wang Q, et al. CYP76M7 is an *ent*-cassadiene C11 α -hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. *Plant Cell*, 2009, 21: 3315–3325
- 25 Ding Y, Weckwerth P R, Poretsky E, et al. Genetic elucidation of interconnected antibiotic pathways mediating maize innate immunity. *Nat Plants*, 2020, 6: 1375–1388
- 26 Chen Q, Jiang T, Liu Y X, et al. Recently duplicated sesterterpene (C25) gene clusters in *Arabidopsis thaliana* modulate root microbiota. *Sci China Life Sci*, 2019, 62: 947–958
- 27 Qi X, Bakht S, Leggett M, et al. A gene cluster for secondary metabolism in oat: Implications for the evolution of metabolic diversity in plants. *Proc Natl Acad Sci USA*, 2004, 101: 8233–8238
- 28 Qi X, Bakht S, Qin B, et al. A different function for a member of an ancient and highly conserved cytochrome P450 family: from essential sterols to plant defense. *Proc Natl Acad Sci USA*, 2006, 103: 18848–18853
- 29 Mugford S T, Louveau T, Melton R, et al. Modularity of plant metabolic gene clusters: a trio of linked genes that are collectively required for acylation of triterpenes in oat. *Plant Cell*, 2013, 25: 1078–1092
- 30 Li Y, Leveau A, Zhao Q, et al. Subtelomeric assembly of a multi-gene pathway for antimicrobial defense compounds in cereals. *Nat Commun*, 2021, 12: 2563
- 31 Field B, Osbourn A E. Metabolic diversification—-independent assembly of operon-like gene clusters in different plants. *Science*, 2008, 320: 543–547
- 32 Field B, Fiston-Lavier A S, Kemen A, et al. Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc Natl Acad Sci USA*, 2011, 108: 16116–16121
- 33 Castillo D A, Kolesnikova M D, Matsuda S P T. An effective strategy for exploring unknown metabolic pathways by genome mining. *J Am Chem Soc*, 2013, 135: 5885–5894
- 34 Sohrabi R, Huh J H, Badieyan S, et al. In planta variation of volatile biosynthesis: an alternative biosynthetic route to the formation of the pathogen-induced volatile homoterpene DMNT via triterpene degradation in *Arabidopsis* roots. *Plant Cell*, 2015, 27: 874–890
- 35 Shang Y, Ma Y, Zhou Y, et al. Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science*, 2014, 346: 1084–1088
- 36 Zhou Y, Ma Y, Zeng J, et al. Convergence and divergence of bitterness biosynthesis and regulation in Cucurbitaceae. *Nat Plants*, 2016, 2: 16183
- 37 Huang A C, Jiang T, Liu Y X, et al. A specialized metabolic network selectively modulates *Arabidopsis* root microbiota. *Science*, 2019, 364: eaau6389
- 38 Liu Z, Cheema J, Vigouroux M, et al. Formation and diversification of a paradigm biosynthetic gene cluster in plants. *Nat Commun*, 2020, 11: 5354
- 39 Krokida A, Delis C, Geisler K, et al. A metabolic gene cluster in *Lotus japonicus* discloses novel enzyme functions and products in triterpene

- biosynthesis. *New Phytol*, 2013, 200: 675–690
- 40 Takos A M, Knudsen C, Lai D, et al. Genomic clustering of cyanogenic glucoside biosynthetic genes aids their identification in *Lotus japonicus* and suggests the repeated evolution of this chemical defence pathway. *Plant J*, 2011, 68: 273–286
- 41 Winzer T, Gazda V, He Z, et al. A *Papaver somniferum* 10-gene cluster for synthesis of the anticancer alkaloid noscapine. *Science*, 2012, 336: 1704–1708
- 42 Itkin M, Heinig U, Tzfadia O, et al. Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science*, 2013, 341: 175–179
- 43 Frey M, Chomet P, Glawischnig E, et al. Analysis of a chemical plant defense mechanism in grasses. *Science*, 1997, 277: 696–699
- 44 Jonczyk R, Schmidt H, Osterrieder A, et al. Elucidation of the final reactions of DIMBOA-glucoside biosynthesis in maize: characterization of *Bx6* and *Bx7*. *Plant Physiol*, 2008, 146: 1053–1063
- 45 von Rad U, Hüttl R, Lottspeich F, et al. Two glucosyltransferases are involved in detoxification of benzoxazinoids in maize. *Plant J*, 2001, 28: 633–642
- 46 Shen S, Peng M, Fang H, et al. An *Oryza*-specific hydroxycinnamoyl tyramine gene cluster contributes to enhanced disease resistance. *Sci Bull*, 2021, doi: 10.1016/j.scib.2021.03.015
- 47 Fang H, Shen S, Wang D, et al. A monocot-specific hydroxycinnamoylputrescine gene cluster contributes to immunity and cell death in rice. *Sci Bull*, 2021, doi: 10.1016/j.scib.2021.06.014
- 48 Schneider L M, Adamski N M, Christensen C E, et al. The *Cer-cqu* gene cluster determines three key players in a β -diketone synthase polyketide pathway synthesizing aliphatics in epicuticular waxes. *J Exp Bot*, 2016, 67: 2715–2730
- 49 Jeon J E, Kim J G, Fischer C R, et al. A pathogen-responsive gene cluster for highly modified fatty acids in tomato. *Cell*, 2020, 180: 176–187.e19
- 50 Fan P, Wang P, Lou Y R, et al. Evolution of a plant gene cluster in Solanaceae and emergence of metabolic diversity. *eLife*, 2020, 9: e56717
- 51 Nützmann H W, Huang A, Osbourn A. Plant metabolic clusters—from genetics to genomics. *New Phytol*, 2016, 211: 771–789
- 52 Nützmann H W, Scazzocchio C, Osbourn A. Metabolic gene clusters in eukaryotes. *Annu Rev Genet*, 2018, 52: 159–183
- 53 Darbani B, Motawia M S, Olsen C E, et al. The biosynthetic gene cluster for the cyanogenic glucoside dhurrin in *Sorghum bicolor* contains its co-expressed vacuolar MATE transporter. *Sci Rep*, 2016, 6: 37079
- 54 Takos A M, Rook F. Why biosynthetic genes for chemical defense compounds cluster. *Trends Plant Sci*, 2012, 17: 383–388
- 55 Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 2009, 10: 57–63
- 56 Medema M H, Kottmann R, Yilmaz P, et al. Minimum information about a biosynthetic gene cluster. *Nat Chem Biol*, 2015, 11: 625–631
- 57 Chae L, Kim T, Nilo-Poyanco R, et al. Genomic signatures of specialized metabolism in plants. *Science*, 2014, 344: 510–513
- 58 Töpfer N, Fuchs L M, Aharoni A. The *PhytoClust* tool for metabolic gene clusters discovery in plant genomes. *Nucleic Acids Res*, 2017, 45: 7049–7063
- 59 Schläpfer P, Zhang P, Wang C, et al. Genome-wide prediction of metabolic enzymes, pathways, and gene clusters in plants. *Plant Physiol*, 2017, 173: 2041–2059
- 60 Kautsar S A, Suarez Duran H G, Blin K, et al. plantSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Res*, 2017, 45: W55–W63
- 61 Banf M, Zhao K, Rhee S Y. METACLUSTER—an R package for context-specific expression analysis of metabolic gene clusters. *Bioinformatics*, 2019, 35: 3178–3180
- 62 Kliebenstein D J. Advancing genetic theory and application by metabolic quantitative trait loci analysis. *Plant Cell*, 2009, 21: 1637–1646
- 63 Mylona P, Owatworakit A, Papadopoulos K, et al. *Sad3* and *Sad4* are required for saponin biosynthesis and root development in oat. *Plant Cell*, 2008, 20: 201–212
- 64 O'Connor S E. Engineering of secondary metabolism. *Annu Rev Genet*, 2015, 49: 71–94
- 65 Sainsbury F, Thuenemann E C, Lomonossoff G P. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol J*, 2009, 7: 682–693
- 66 Sainsbury F, Lomonossoff G P. Extremely high-level and rapid transient protein production in plants without the use of viral replication. *Plant Physiol*, 2008, 148: 1212–1218
- 67 Peyret H, Lomonossoff G P. The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. *Plant Mol Biol*, 2013, 83: 51–58

- 68 Wu Y, Hillwig M L, Wang Q, et al. Parsing a multifunctional biosynthetic gene cluster from rice: Biochemical characterization of CYP71Z6 & 7. *FEBS Lett*, 2011, 585: 3446–3451
- 69 Wang Q, Hillwig M L, Peters R J. CYP99A3: functional identification of a diterpene oxidase from the momilactone biosynthetic gene cluster in rice. *Plant J*, 2011, 65: 87–95
- 70 Galanie S, Thodey K, Trenchard I J, et al. Complete biosynthesis of opioids in yeast. *Science*, 2015, 349: 1095–1100
- 71 Paddon C J, Westfall P J, Pitera D J, et al. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, 2013, 496: 528–532
- 72 DeLoache W C, Russ Z N, Narcross L, et al. An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. *Nat Chem Biol*, 2015, 11: 465–471
- 73 Li Y, Smolke C D. Engineering biosynthesis of the anticancer alkaloid noscapine in yeast. *Nat Commun*, 2016, 7: 12137
- 74 Qu Y, Easson M L A E, Froese J, et al. Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proc Natl Acad Sci USA*, 2015, 112: 6224–6229
- 75 Reed J, Osbourn A. Engineering terpenoid production through transient expression in *Nicotiana benthamiana*. *Plant Cell Rep*, 2018, 37: 1431–1441
- 76 Reed J, Stephenson M J, Miettinen K, et al. A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. *Metab Eng*, 2017, 42: 185–193
- 77 Shamloul M, Trusa J, Mett V, et al. Optimization and utilization of *Agrobacterium*-mediated transient protein production in *Nicotiana*. *J Vis Exp*, 2014, 86: e51204
- 78 King A J, Brown G D, Gilday A D, et al. A cytochrome P450-mediated intramolecular carbon-carbon ring closure in the biosynthesis of multidrug-resistance-reversing lathyrane diterpenoids. *Chembiochem*, 2016, 17: 1593–1597
- 79 Takos A, Lai D, Mikkelsen L, et al. Genetic screening identifies cyanogenesis-deficient mutants of *Lotus japonicus* and reveals enzymatic specificity in hydroxynitrile glucoside metabolism. *Plant Cell*, 2010, 22: 1605–1619
- 80 Lu R, Martin-Hernandez A M, Peart J R, et al. Virus-induced gene silencing in plants. *Methods*, 2003, 30: 296–303
- 81 Liu Z, Suarez Duran H G, Harnvanichvech Y, et al. Drivers of metabolic diversification: how dynamic genomic neighbourhoods generate new biosynthetic pathways in the *Brassicaceae*. *New Phytol*, 2020, 227: 1109–1123
- 82 Zhao X, Meng Z, Wang Y, et al. Pollen magnetofection for genetic modification with magnetic nanoparticles as gene carriers. *Nat Plants*, 2017, 3: 956–964
- 83 Schmelz E A, Huffaker A, Sims J W, et al. Biosynthesis, elicitation and roles of monocot terpenoid phytoalexins. *Plant J*, 2014, 79: 659–678
- 84 Okada A, Okada K, Miyamoto K, et al. OsTGAP1, a bZIP transcription factor, coordinately regulates the inductive production of diterpenoid phytoalexins in rice. *J Biol Chem*, 2009, 284: 26510–26518
- 85 Akagi A, Fukushima S, Okada K, et al. WRKY45-dependent priming of diterpenoid phytoalexin biosynthesis in rice and the role of cytokinin in triggering the reaction. *Plant Mol Biol*, 2014, 86: 171–183
- 86 Yamamura C, Mizutani E, Okada K, et al. Diterpenoid phytoalexin factor, a bHLH transcription factor, plays a central role in the biosynthesis of diterpenoid phytoalexins in rice. *Plant J*, 2015, 84: 1100–1113
- 87 Cárdenas P D, Sonawane P D, Pollier J, et al. GAME9 regulates the biosynthesis of steroidal alkaloids and upstream isoprenoids in the plant mevalonate pathway. *Nat Commun*, 2016, 7: 10654
- 88 Shwab E K, Bok J W, Tribus M, et al. Histone deacetylase activity regulates chemical diversity in *Aspergillus*. *Eukaryot Cell*, 2007, 6: 1656–1664
- 89 Bok J W, Chiang Y M, Szewczyk E, et al. Chromatin-level regulation of biosynthetic gene clusters. *Nat Chem Biol*, 2009, 5: 462–464
- 90 Yu N, Nützmann H W, MacDonald J T, et al. Delineation of metabolic gene clusters in plant genomes by chromatin signatures. *Nucleic Acids Res*, 2016, 44: 2255–2265
- 91 Nützmann H W, Osbourn A. Regulation of metabolic gene clusters in *Arabidopsis thaliana*. *New Phytol*, 2015, 205: 503–510
- 92 Nützmann H W, Doerr D, Ramírez-Colmenero A, et al. Active and repressed biosynthetic gene clusters have spatially distinct chromosome states. *Proc Natl Acad Sci USA*, 2020, 117: 13800–13809
- 93 Osbourn A. Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends Genet*, 2010, 26: 449–457
- 94 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: 714–721

Research progress on plant biosynthetic gene cluster

SHEN ShuangQian¹, ZHAN ChuanSong², YANG ChenKun¹ & LUO Jie²

*1 National Key Laboratory of Crop Genetic Improvement and National Centre of Plant Gene Research (Wuhan),
Huazhong Agricultural University, Wuhan 430070, China;*

2 College of Tropical Crops, Hainan University, Haikou 572208, China

Plants produce a wide variety of structurally and functionally diverse metabolites, many of which play an indispensable role in plant growth and development and in the response to abiotic and biotic stresses. In recent years, plant biosynthetic gene clustering has attracted extensive attention from scientists, and the rapid development of multi-omics technologies such as genome, transcriptome and metabolome makes it possible to efficiently mine plant metabolic gene clusters. Further, the studies of plant biosynthetic gene clusters will provide new insights to explore and utilize the diversity of plant secondary metabolites. Here, we not only systematically summarize the structural characteristics and regulatory models of plant biosynthetic gene clusters, but also review the application of new technologies in the discovery and identification of biosynthetic gene clusters.

plant metabolic gene cluster, forward genetics, gene cluster characteristics, gene cluster mining methods, gene cluster function verification, chromatin regulation

doi: [10.1360/SSV-2021-0198](https://doi.org/10.1360/SSV-2021-0198)