

· 专题论坛 ·

植物次生代谢产物生物合成基因簇研究进展

朱璐, 袁冲, 刘义飞*

湖北中医药大学, 武汉 430074

摘要 植物产生的次生代谢产物为人类提供了丰富的药物、香料和工业原料。随着分子生物学和基因组学研究的快速发展, 目前已解析了多种植物的次生代谢产物生物合成基因簇(BGCs)。这为我们快速获取目标产物的生物合成通路和发掘新颖的天然产物开辟了新路径。该文重点围绕植物次生代谢产物生物合成基因簇的定义和特点、基本结构模型与演化以及调控机制等进行综述, 以期对相关研究提供理论依据和借鉴。

关键词 次生代谢产物, 生物合成基因簇, 基因簇调控, 基因簇特征

朱璐, 袁冲, 刘义飞 (2024). 植物次生代谢产物生物合成基因簇研究进展. 植物学报 59, 134–143.

植物产生的丰富天然产物, 特别是多样化的次生代谢产物(secondary metabolites)为其自身适应生态环境, 防御天敌以及病虫害侵扰提供了物质基础(Nützmann and Osbourn, 2014)。不同植物在其进化过程中可产生物种特异或新的代谢产物, 这些产物为人类提供了丰富的药物、香料和工业原料, 对农业、医药和工业生产均具有极其重要的价值。在我国的传统中医药实践和现代医学发展过程中, 源于各类中草药的次生代谢产物已广泛应用于治疗各种疾病, 或用于新药研发(Verpoorte and Memelink, 2002)。

随着现代分子生物学技术的飞速发展, 特别是基因组、转录组和代谢组等多组学研究的不断深入, 大量植物天然产物生物合成的关键途径和机制得以解析(Zhu et al., 2018)。其中, 植物次生代谢产物生物合成基因簇(biosynthetic gene clusters, BGCs)的发现为我们快速获取目标产物的生物合成通路和发掘新颖的天然产物开辟了新路径(杨谦等, 2021)。基于基因组数据, 通过生物合成基因簇预测及后续的功能验证, 可全面获取相关通路基因及调控元件。相比传统的基因挖掘方法极大地提高了通路解析效率和精准度。同时, 随着植物基因组测序的广泛开展, 多个植物基因组的比较分析为批量挖掘及应用生物合成通路相关功能基因和酶提供了可能(Boutanaev et al.,

2015)。

本文从植物次生代谢产物生物合成基因簇的定义和特点、基本结构模型、进化和表达调控机制、部分重点研究的基因簇参与合成的化合物以及微生物和植物次生代谢产物合成基因簇的比较等方面进行综述, 为更好地发掘和利用植物多样化的次生代谢产物提供理论依据。

1 植物次生代谢产物生物合成基因簇的定义和特点

自1997年在玉米(*Zea mays*)中发现一种防御性化合物DIMBOA的生物合成基因簇以来, 截至2022年5月, 已有40多篇文献报道了30多个不同类型的植物次生代谢产物生物合成基因簇(Frey et al., 1997; Park et al., 2018; Bharadwaj et al., 2021; Li et al., 2022; Wu et al., 2022; Zhan et al., 2022), 其中包括在药用植物罂粟(*Papaver somniferum*)和丹参(*Salvia miltiorrhiza*)中发现的基因簇(Guo et al., 2018; Park et al., 2018)。这些基因簇参与合成的化合物中一半以上为萜类, 其次是糖苷类和生物碱。近年来, 植物生物合成基因簇研究呈上升趋势(图1), 且基因簇参与合成的化合物种类日益丰富。

收稿日期: 2022-10-03; 接受日期: 2023-04-18

基金项目: 国家自然科学基金(No.32270231)和湖北省杰出青年基金(No.2019CFA097)

* 通讯作者。E-mail: liuyifei@hbtcm.edu.cn

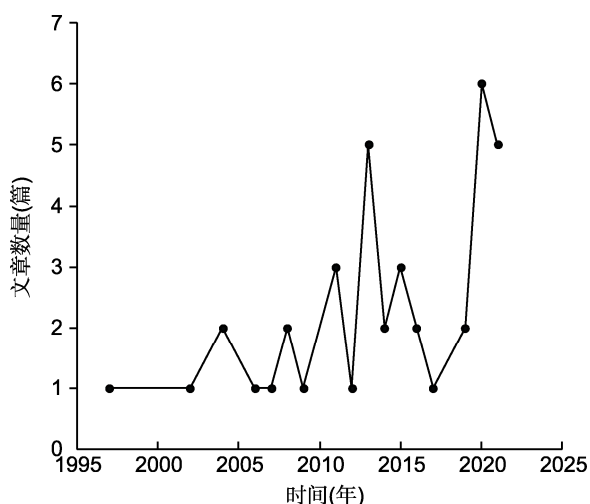


图1 近年来植物次生代谢物生物合成基因簇研究的文章数量

Figure 1 Number of published articles on plant secondary metabolic biosynthetic gene clusters in recent years

次生代谢产物生物合成基因簇是指在植物和细菌(或真菌)等生物基因组的一定范围内(通常几十到几百kb)连续出现的至少包括3个非同源(non-homologous)类型的基因合集(Nützmann and Osbourn, 2014)。这些基因隶属同一或者关联的生物合成通路,其中1个关键基因编码的酶催化代谢产物骨架的形成,其它基因编码多样化修饰中间产物骨架的酶,形成终端产物(Osbourn, 2010)。细胞色素P450超家族(cytochromes P450, CYP)基因在植物基因簇中广泛存在(吕海舟等, 2017)。基于氨基酸序列相似性(>40%), CYP基因可分为若干家族,源于不同家族的基因常被视作非同源类型基因(Nelson and Werck-Reichhart, 2011)。

植物基因簇具有一些共性和差异化特征。植物生物合成基因簇通常包括某个终端次生代谢产物合成通路上编码第1步催化反应酶的基因,称为信号基因(signature gene)。同时,还附加2个或更多基因编码通路下游的修饰酶(亦称剪裁酶(tailoring enzymes))。以玉米的Bx基因簇为例,该基因簇中的Bx1基因是玉米中色氨酸合成酶 α -亚基的同源基因,可将吲哚-3-甘油磷酸转化为吲哚,而Bx2-Bx5均为CYP71C家族基因,可逐步催化修饰吲哚转化为DIBOA (DIMBOA前体物质)(Frey et al., 1997)。番茄(*Solanum lycopersicum*)的单萜合成酶基因簇则由TPS、CPT和CYP等非同源基因组成,其中TPS基因编码信号酶,

合成单萜骨架,CPT和CYP基因编码剪裁酶,催化修饰中间产物形成单萜目标化合物(Matsuba et al., 2013; 吕海舟等, 2017)。Gao等(2018)分析了罂粟基因组,发现天然药物诺司卡品(Noscapine)和吗啡(Morphine)的生物合成基因簇也具有类似的信号和修饰基因(Guo et al., 2018);并且发现在诺司卡品和燕麦根皂苷(avenacin)的生物合成基因簇中,基因排序与其在代谢通路上的反应步骤存在明显的共线性关系(Winzer et al., 2012)。

2 生物合成基因簇: 从微生物到植物

微生物次生代谢产物生物合成基因簇的研究早于植物。Smith等于1990年就实现了在粗糙链孢菌(*Neurospora crassa*)和黑曲霉(*Aspergillus niger*)中转入青霉素合成基因簇,异源表达抗生素(青霉素)基因(Smith et al., 1990)。在目前已测定的80多个链霉菌全基因组中,每个链霉菌属基因组可编码约30个次生代谢产物生物合成基因簇(Komaki et al., 2018)。微生物的生物合成基因簇通常也由主干基因(backbone gene)和修饰酶基因组成。主干基因编码聚酮合酶(polyketide synthases)、非核糖体肽合成酶(nonribosomal peptide synthetases)、二甲基烯丙基色氨酸合成酶(dimethylallyl tryptophan synthases)和萜烯环化酶(terpene cyclases)等;聚集在主干基因周围的修饰酶基因(主要包括氧化还原酶、加氧酶和脱氢酶等)则负责修饰骨架多样化的次生代谢产物。随着研究的不断深入,批量化发掘工具(如antiSMASH)(antibiotics & Secondary Metabolite Analysis Shell)的发布极大地提升了微生物生物合成基因簇的发掘效率(Blin et al., 2019)。其内置的聚类分析算法能鉴定45种不同类型次生代谢产物的生物合成途径,并能详细预测其中多种类型基因簇可合成的次生代谢产物(Weber and Kim, 2016)。目前,antiSMASH已应用于植物基因簇研究。

Netzker等(2015)和Nützmann等(2018)研究表明,植物和微生物中大多数能产生类药分子的次生代谢产物生物合成基因簇常处于“沉默状态”,需特殊的环境压力刺激唤醒其表达,从而催化相关次生代谢产物的合成。水稻(*Oryza sativa*)在受到病原菌攻击时,会激活“沉默状态”的基因簇,合成一些二萜(Kariya et al., 2020)。低浓度的甲氧苄氨嘧啶可作为

诱导剂唤醒沉默的生物合成基因簇, 促进泰国伯克霍尔德菌(*Burkholderia thailandensis*)产生100多种正常生长条件下不会产生的次生代谢产物(Okada et al., 2016)。目前, 已有大量关于微生物基因簇激活的研究, 而植物的相关研究较少。

微生物与植物次生代谢产物生物合成基因簇的联系至今仍是未解之谜。Yue等(2013)研究表明, 植物基因组中的一些核基因家族是从原核生物、真菌或病毒中获取, 但在基因簇研究领域, 基因水平转移并未被广泛证实, 且有研究表明植物基因簇是独立起源进化的(Nützmann and Osbourn, 2014; Nützmann et al., 2018; Rokas et al., 2018, 2020)。例如, 在植物基因簇中, 真菌基因簇通路特异性转录调控因子和转运子(transcriptional regulators and transporters)并不常见(Rokas et al., 2020)。另外, 有研究发现植物基因簇也含有转运体和协调激活的调控基因(Darbani et al., 2016; Hen-Avivi et al., 2016; Shen et al., 2021)。因此, 微生物次生代谢产物合成基因簇研究可为植物中的相关研究提供借鉴与参考。

3 植物次生代谢产物生物合成基因簇的形成与演化

目前, 关于植物次生代谢产物生物合成基因簇形成的进化遗传机制尚不清楚。假设植物产生大量特异且丰富的次生代谢产物与其适应多样化的生态环境相关, 基因簇的产生则可能有利于植物适应特殊的生境, 进

而生存和繁衍, 而不是利于代谢通路本身的形成与进化(Nützmann et al., 2016)。许多植物基因簇是在一个相对较短的时间内形成, 且仅出现在特定的植物分类单元中(Field et al., 2011; Matsuba et al., 2013)。假设植物基因组具有高度的可塑性, 基因簇的出现则可保持通路基因的紧密连锁和稳定遗传, 避免基因组重组带来表达和功能上的干扰, 呈现出一定的选择性优势(selective advantage)。在较长的时间尺度上, 基因簇本身也可能经历从生到死的动态历程。此外, 基因簇的产生可保证植物特异性代谢产物生物合成通路的完整性, 避免产生有毒或有生物活性的中间产物, 不利于植物的生长适应。人工突变、RNA干扰或过表达基因簇相关基因会导致积累大量有毒的中间产物, 从侧面验证了这一假设(Field et al., 2011; Xu et al., 2012)。Ghanbarian和Hurst (2015)研究表明, 基因组上物理位置邻近的基因表达具有重要的相关性, 物理位置邻近可强化基因簇基因的共表达特性, 为基因簇整体调控提供更多的机会。此外, 基因簇的形成还可能具有一定的物理廊道效应, 促进底物依次级联反应, 使代谢产物的合成效率最大化。

目前推测的几种植物基因簇的起源方式包括从头组装(图2A), 以及基于基因对(图2B)和基因簇(图2C)的复制等(Nützmann et al., 2016; Bharadwaj et al., 2021)。在基因簇的从头组装中, 关键信号基因可能直接或间接来源于初生代谢(primary metabolism)途径, 进而作为基因簇的“种子”基因。例如, 由环阿屯醇合成酶(cycloartenol synthase)基因分化而来

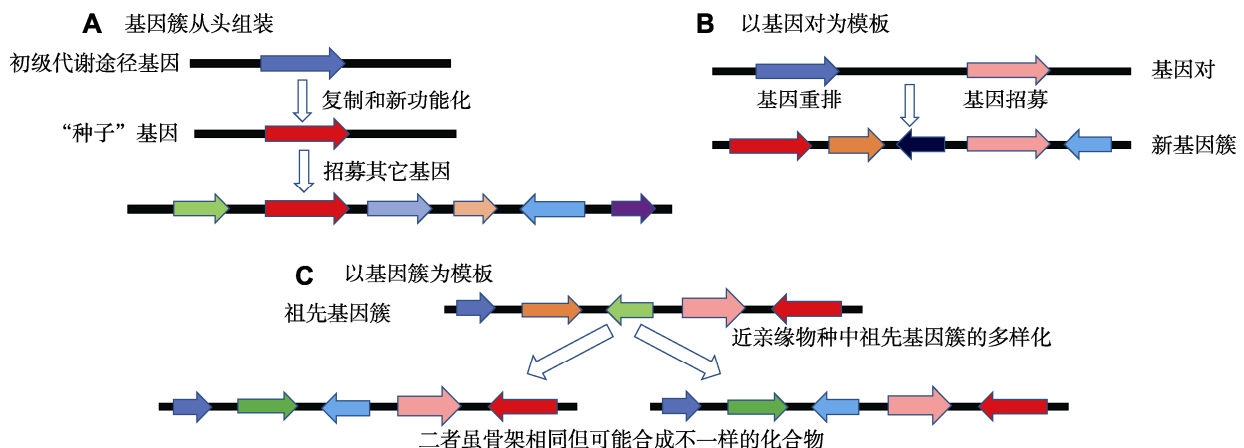


图2 植物生物合成基因簇起源的3种可能方式(改自Nützmann et al., 2016)

Figure 2 Three possible scenarios for the origin of biosynthetic gene clusters in plants (modified from Nützmann et al., 2016)

的燕麦根皂苷基因簇的关键信号基因*AsbAS1*招募于甾醇合成途径(Qi et al., 2004)。拟南芥(*Arabidopsis thaliana*)基因簇则是通过对氧化鲨烯环化酶(oxidosqualene cyclase, OSC)以及*CYP705*基因对的原始复制及其它新基因的招募形成(Field et al., 2011)。拟南芥与其近缘种琴叶拟南芥(*A. lyrata thaliana*)基因簇的核苷酸序列高度同源, 且具有类似的基因组排列顺序和根组织的特异性表达, 因而二者可能拥有共同的祖先进化起源(Liu et al., 2020)。最近, Yang等(2021)比较分析了3个罂粟基因组, 发现诺司卡品以及吗啡的生物合成通路基因可能通过补丁式(patchwork)跳跃演化聚集成簇。

4 植物次生代谢物生物合成基因簇的基本结构模型

Nützmann等(2016)总结了植物次生代谢物生物合成基因簇的4种基本结构模型(图3)。除类似拟南芥基因簇(Field and Osbourn, 2008)的经典简约模式(图3A)外, 一些基因簇还包括编码其它酶的附加基因(peripheral genes), 这些基因与基因簇的核心基因功能相关或无关(图3B)。例如, 高粱蜀黍氰苷(dhurrin)生物合成基因簇中包含了共表达的转运蛋白基因*SbMATE* (Darbani et al., 2016)。一些基因簇除分布于同一染色体区域外, 同一通路的基因还会在其它染色体上形成卫星簇(satellite subgroups) (包括1个或2个剪裁酶基因) (图3C)。例如, 黄瓜(*Cucumis sativus*)中负责葫芦苦素(cucurbitacin)生物合成的5个基因主要以基因簇的方式存在于其6号染色体上, 其它4个CYP通路基因则以卫星簇或单基因的形式分别存在于第1和3号染色体上(Shang et al., 2014)。此外, 关键信号基因在基因簇中的位置也可能被“边缘”化, 出现在远离基因簇的地方(图3D)。例如, 番茄甾体类生物碱 α -番茄碱(α -tomatine)生物合成中的关键信号基因*GAME7*虽与核心基因簇基因共处于7号染色体, 但位置在远离核心基因约7 880 kb处(Itkin et al., 2013)。类似的, 诺司卡品生物合成关键基因*BBE*和*TNMT*也不在其核心基因簇中(Guo et al., 2018)。

5 植物基因簇的表达调控机制

植物次生代谢产物生物合成基因簇的表达符合时空

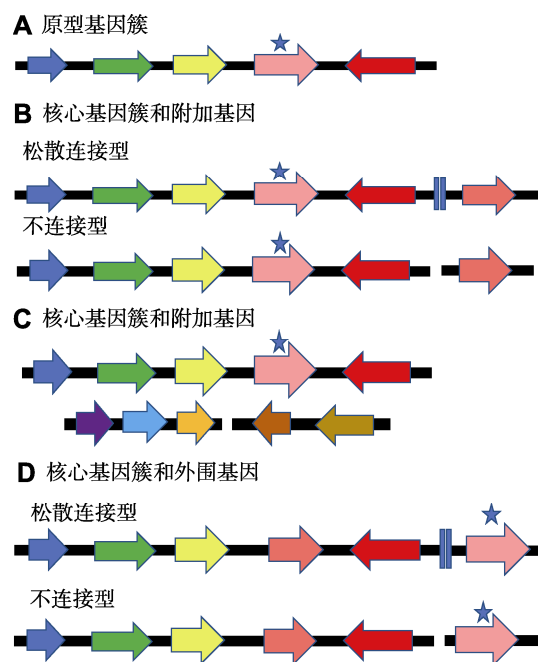


图3 植物次生代谢物生物合成基因簇的4种结构模型(改编自Nützmann et al., 2016)

*信号基因

Figure 3 Four basic structural models of plant secondary metabolite biosynthesis gene clusters (modified from Nützmann et al., 2016)

*Signature genes

调节的一般规律, 同时一些转录因子和染色质重塑(chromatin remodeling)也可能在基因簇的基因表达调控中具有重要作用(方荣俊等, 2014; Nützmann et al., 2016; Bharadwaj et al., 2021)。燕麦根皂苷生物合成基因簇中的信号基因*AsbAS1*和其它修饰基因的表达主要集中在根尖, 表明相关次生代谢产物的合成具有明显的组织特异性(Haralampidis et al., 2001; Li et al., 2022)。类似的, 玉米DIMBOA的生物合成受到生长发育调控, 主要局限于叶和根(Frey et al., 1997)。水稻中, 一些二萜生物合成基因簇的表达受几丁质寡糖诱导和紫外线辐射激活, 也主要存在于叶和根中(Bhara-dwaj et al., 2021)。

目前, 在植物中已发现一些转录因子特异性调控次生代谢产物生物合成基因簇的表达。黄瓜中, 3个bHLH类转录因子基因(*Bt*, *Bl*和*Br*)分别调控果实、叶和根中的葫芦苦素生物合成基因簇的表达, 从而影响其苦味产生(Shang et al., 2014)。一些驯化的黄瓜品种由于*Bt*基因突变, 导致其苦味消失(Chomicki et

al., 2020)。植物代谢相关基因簇也受到表观遗传广泛调控。Yu等(2016)对拟南芥、玉米、燕麦(*Avena sativa*)和水稻的研究均发现,组蛋白H3赖氨酸27的三甲基化(H3K27me3)水平与相关基因簇的表达呈负相关。位于水稻7号染色体上生物合成5,10-diketo-casbene抗病二萜的DGC7基因簇可以被组蛋白去甲基化酶JMJ705通过茉莉酸甲酯介导去甲基化(Zhan et al., 2020)。拟南芥中和 $marneral$ 成簇基因也受到SWR1染色质重塑复合物的正调控,且活跃的基因簇通常出现在远离异染色质和细胞核边缘的热点区域(Nützmann et al., 2020)。

6 植物次生代谢产物的种类和相关基因簇

6.1 萜类

萜类(terpenoids)是自然界中结构和功能最丰富的天然产物,也是现有研究中发现的由基因簇参与合成最多的植物代谢产物。目前已知的萜类有5 000多种,约占植物天然产物的40% (Vranová et al., 2013; Liu et al., 2020)。根据结构中包含异戊二烯单元的数量不同,萜类物质可分为单萜、倍半萜、二萜和三萜等。植物中萜类生物合成途径主要包括甲羟戊酸(mevalonate pathway, MVA)和甲基赤藓糖醇-4-磷酸(2-C-methyl-D-erythritol-4-phosphate, MEP)途径。萜类生物合成通路中,萜类合成酶(terpene synthases, TS)基因和不同类型的CYP基因通常形成基因簇。前者作为信号酶负责不同萜类骨架的产生,后者则执行修饰任务,促进多样化萜类终端产物的形成。Boutanaev等(2015)系统分析了17种植物的基因组,发现TS/CYP基因簇的结构和功能具有丰富的多样性(图4)。在双子叶和单子叶植物中,该类基因簇可能具有不同的起源和组装机制。双子叶植物的基因对起源方式(图2B)可能是TS/CYP基因簇进化多样化的主要方式;单子叶植物中,新基因簇的形成则通过已有基因簇的基因组重排产生。对双子叶植物拟南芥和单子叶植物燕麦相关基因簇的分析也验证了这一假说(Qi et al., 2004, 2006; Field et al., 2011)。Liu等(2020)基于十字花科多个基因组,系统挖掘了OSC基因相关的三萜生物合成基因簇,发现这些基因簇多位于植物基因组的高变区域,其进化相对独立,故与其它基因组区

域缺乏共线性。

6.2 生氰糖苷类

植物中含有60多种生氰糖苷(cyanogenic glycoside, CG) (Jones, 1998; Takos et al., 2011) (图4)。早期研究发现,生氰糖苷存在于拟南芥和双色拟南芥中且具有阻碍食草动物取食的作用(Jones, 1998; Tattersall et al., 2001)。同时,生氰糖苷水解释放的氢氰酸(hydrogencyanide)在亚麻(*Linum usitatissimum*)抵抗枯萎病过程中也发挥重要作用(Biere et al., 2004)。目前,由基因簇参与合成的生氰糖苷类化合物主要有亚麻苦甙(linamarin)、百脉根甙(lotaustralin)和蜀黍氰苷,参与这3类化合物合成的3种前体氨基酸(L-缬氨酸、L-异亮氨酸以及L-酪氨酸)的生物合成基因也聚集在一起(Takos et al., 2011)。由CYP79D3编码的生氰糖苷合成酶催化生氰糖苷合成的第1步反应,而CYP736A2和UGT85K3位于CYP79D3的周围,共同参与整个生物合成途径。高粱(*Sorghum bicolor*)中,编码转运蛋白的基因SbMATE2也存在于该基因簇中,并与其它基因共表达(Darbani et al., 2016)。已知生氰糖苷具有抵抗病原体以及食草动物取食的作用(Kakes, 1989),这一共表达是否通过未知机制进行尚待进一步研究。

6.3 生物碱类

生物碱(alkaloid)是多种中药的有效成分,其药理作用包括抗病毒、抗菌和抗肿瘤等(Tao et al., 2020)。在植物中发现了2种由基因簇参与合成的生物碱,一种是罂粟中的苜基异喹啉类生物碱诺司卡品和吗啡(Winzer et al., 2012),另一种是马铃薯(*Solanum tuberosum*)和番茄中的甾族生物碱 α -番茄素(Darbani et al., 2016)。罂粟中,编码O-甲基转移酶和CYP酶合成诺司卡品的基因簇由10个基因构成。进一步研究发现,这一基因簇实际上隶属于1个由15个基因组成的更大基因簇,其包含的其它5个酶参与可待因和吗啡等止痛剂的合成(Guo et al., 2018)。茄科植物中的甾体生物碱(steroidal glycoalkaloids, SGAs)具有抗营养物质吸收的作用,能破坏或阻碍营养物质在动物体内的消化吸收(Roddick et al., 2001)。Itkin等(2013)对马铃薯和番茄进行研究,发现了10个编码SGAs生物合成酶的基因序列,其中6个基因位于7号染色体

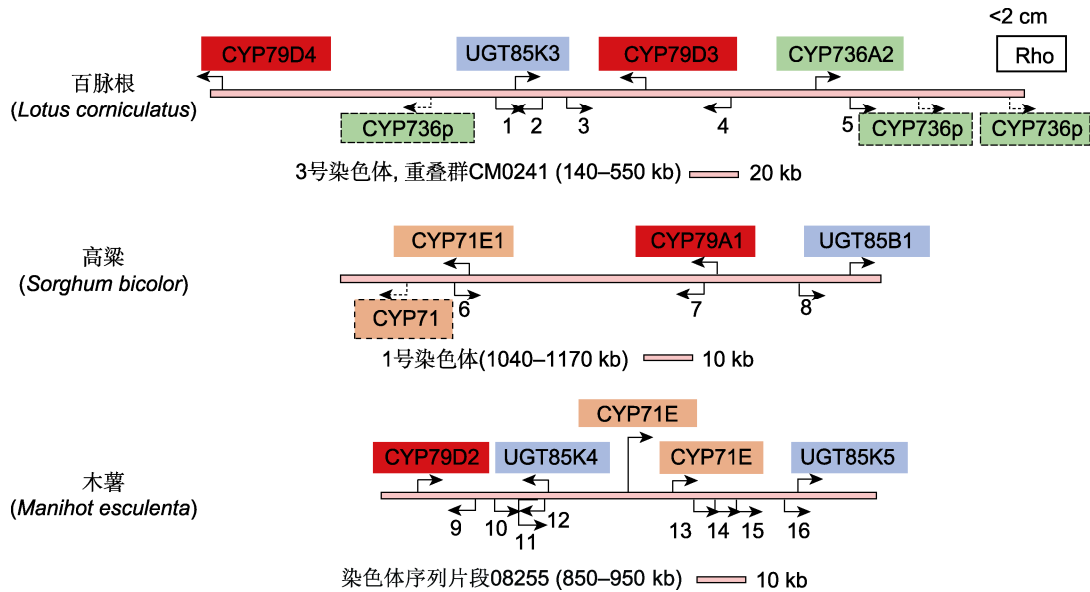


图4 3种典型植物中生氰糖苷生物合成基因簇(改编自Takos et al., 2011)

图中箭头代表基因的方向, 方框表示已确认的生氰糖苷生物合成基因, 其中CYP79基因标记为红色方框, CYP71和CYP736基因分别标记为橙色和绿色方框, UGT85基因为蓝色方框。百脉根染色体下方为与CYP736A2相似的3个假基因, 高粱染色体下方显示的为假基因CYP71。Rho基因座位于CM0241等位基因的2 cM范围内。其余基因的编号和注释如下: (1) 顺式还原酮加双氧酶; (2) 核酸结合, OB折叠; (3), (4) 假设蛋白; (5) 核糖核酸酶H; (6) 短链脱氢/还原酶; (7), (8) 假定转座酶; (9) 富亮氨酸重复受体样蛋白激酶; (10), (11) α/β -折叠水解酶; (12) 假设蛋白; (13)–(15) 假定羟腈裂解酶; (16) 假设蛋白。

Figure 4 Genomic clustering of cyanogenic glucoside biosynthetic genes in three typical plants (modified from Takos et al., 2011)

Functional genes are presented by arrows indicating their orientation. Confirmed genes in cyanogenic glycoside biosynthesis are labelled above each bar, with CYP79 genes in red, CYP71 genes in orange, CYP736 genes in green, and UGT85 genes in blue. The three CYP736A2-like pseudo-genes are indicated below the *Lotus corniculatus* bar, as is the additional CYP71 in *Sorghum bicolor*. The Rho locus is within 2 cM of the CM0241 contig. The remaining genes are numbered and annotated as follows: (1) Acireductone dioxygenase; (2) Nucleic acid binding, OB-fold; (3), (4) Hypothetical proteins; (5) Ribonuclease H; (6) Short-chain dehydrogenase/reductase; (7), (8) Putative transposases; (9) Leucine-rich repeat receptor-like protein kinase; (10), (11) α/β -fold hydrolases; (12) Hypothetical protein; (13)–(15) Putative hydroxynitrile lyases; (16) Hypothetical protein.

的相邻区域, 而另外2个(CYP88D亚家族的GAME4和转氨酶基因GAME12)位于12号染色体上。

6.4 其它

除上述类型的次生代谢产物外, 异羟肟酸、酚胺、聚酮化合物以及脂肪酸等脂肪族化合物的合成也有基因簇参与。例如, Guo等(2017)在稗草(*Echinochloa crusgalli*)中发现可合成异羟肟酸类次生代谢产物DIMBOA的3个基因簇, 同时当稗草与水稻混种时, 该基因簇会快速启动DIMBOA的合成, 从而明显抑制水稻生长。另外, 玉米中发现的Bx基因簇也参与该类物质的合成(Frey et al., 1997; von Rad et al., 2001;

Jonczyk et al., 2008)。

酚胺类物质是一种多样化的特殊代谢物, 包括羟基肉桂酰普特卡因、羟基肉桂酰拉马汀和羟基肉桂酰色胺, 在植物抵抗生物胁迫方面具有重要作用。Shen等(2021)研究发现, 水稻中的羟基肉桂酰酪胺基因簇参与酚胺类化合物的合成, 该基因簇包含1个产生辅助因子5'-磷酸吡哆醇(PLP)的5'-磷酸吡哆醇氧化酶(OsPDX3)、1个依赖PLP的酪氨酸脱羧酶(OsTyDC1)以及2个重复的羟基肉桂酰转移酶(OsTHT1和OsTHT2)。大麦(*Hordeum vulgare*) Cer-cqu基因簇中3个Cer (Eceriferum)基因(Cer-c、Cer-q和Cer-u)紧密相连, 参与合成 β -二酮和羟基 β -二酮(Schneider et

al., 2016)。番茄中除萜类化合物的合成有基因簇参与外, 脂肪族化合物(法卡林二醇和中链酰基糖)的合成也与基因簇有关(Jeon et al., 2020)。

7 展望

测序技术的发展极大地提高了识别和解析植物生物合成基因簇的能力, 且相关研究也逐渐拓展到生物合成和调控基因的挖掘(Shang et al., 2014)、转录和表观遗传调控机制的揭示(Yu et al., 2016; Zhan et al., 2020)及初生和次生代谢物的多样化研究(Jeon et al., 2020)等不同方面和层次。越来越多新工具的应用加速了植物生物合成基因簇的研究(Gaquereel et al., 2014; Yamamuro et al., 2016; Luo et al., 2020; Yang et al., 2020; Zhan et al., 2020; Li and Gaquerel, 2021; Varshney et al., 2021; Wu et al., 2021)。例如, 以深度学习(deep learning)为代表的人工智能技术开始应用到生物学研究中, 也必将应用到批量生物合成基因簇的发掘(Tunyasuvunakool et al., 2021)。未来次生代谢产物基因簇研究可能会发展为基于大数据计算的生物信息学分析预测, 及结合现有代谢网络和精准实验验证的模块化工作。随着数据量的积累、研究速度的加快及植物基因簇的大量挖掘和功能研究, 将会发现更多的稀有天然产物, 为人类健康生活提供更多的物质基础和医药保障。

作者贡献声明

朱璐: 收集文献和撰写论文; 袁冲: 分析数据和作图; 刘义飞: 构思和撰写论文。

参考文献

方荣俊, 赵华, 廖永辉, 汤程贻, 吴凤瑶, 朱煜, 庞延军, 陆桂华, 王小明, 杨荣武, 戚金亮, 杨永华 (2014). 乙烯对植物次生代谢产物合成的双重调控效应. *植物学报* **49**, 626–639.

吕海舟, 刘琬菁, 何柳, 徐志超, 罗红梅 (2017). 植物次生代谢基因簇研究进展. *植物科学学报* **35**, 609–621.

杨谦, 程伯涛, 汤志军, 刘文 (2021). 基因组挖掘在天然产物发现中的应用和前景. *合成生物学* **2**, 697–715.

Bharadwaj R, Kumar SR, Sharma A, Sathishkumar R (2021). Plant metabolic gene clusters: evolution, organization, and their applications in synthetic biology. *Front Plant Sci* **12**, 697318.

Biere A, Marak HB, van Damme JMM (2004). Plant chemical defense against herbivores and pathogens: generalized defense or trade-offs? *Oecologia* **140**, 430–441.

Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* **47**, W81–W87.

Boutanaev AM, Moses T, Zi JC, Nelson DR, Mugford ST, Peters RJ, Osbourn A (2015). Investigation of terpene diversification across multiple sequenced plant genomes. *Proc Natl Acad Sci USA* **112**, E81–E88.

Chomicki G, Schaefer H, Renner SS (2020). Origin and domestication of Cucurbitaceae crops: insights from phylogenies, genomics and archaeology. *New Phytol* **226**, 1240–1255.

Darbani B, Motawia MS, Olsen CE, Nour-Eldin HH, Møller BL, Rook F (2016). The biosynthetic gene cluster for the cyanogenic glucoside dhurrin in *Sorghum bicolor* contains its co-expressed vacuolar MATE transporter. *Sci Rep* **6**, 37079.

Field B, Fiston-Lavier AS, Kemen A, Geisler K, Quesneville H, Osbourn AE (2011). Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc Natl Acad Sci USA* **108**, 16116–16121.

Field B, Osbourn AE (2008). Metabolic diversification-independent assembly of operon-like gene clusters in different plants. *Science* **320**, 543–547.

Frey M, Chomet P, Glawischnig E, Stettner C, Grun S, Winklmair A, Eisenreich W, Bacher A, Meeley RB, Briggs SP, Simcox K, Gierl A (1997). Analysis of a chemical plant defense mechanism in grasses. *Science* **277**, 696–699.

Gaquereel E, Gulati J, Baldwin IT (2014). Revealing insect herbivory-induced phenolamide metabolism: from single genes to metabolic network plasticity analysis. *Plant J* **79**, 679–692.

Ghanbarian AT, Hurst LD (2015). Neighboring genes show correlated evolution in gene expression. *Mol Biol Evol* **32**, 1748–1766.

Guo L, Winzer T, Yang XF, Li Y, Ning ZM, He ZS, Teodor R, Lu Y, Bowser TA, Graham IA, Ye K (2018). The opium poppy genome and morphinan production. *Science* **362**, 343–347.

Guo LB, Qiu J, Ye CY, Jin GL, Mao LF, Zhang HQ, Yang XF, Peng Q, Wang YY, Jia L, Lin ZX, Li GM, Fu F, Liu C, Chen L, Shen EH, Wang WD, Chu QJ, Wu DY, Wu SL, Xia CY, Zhang YF, Zhou XM, Wang LF, Wu LM, Song WJ, Wang YF, Shu QY, Aoki D, Yumoto E, Yoko-

- ta T, Miyamoto K, Okada K, Kim DS, Cai DG, Zhang CL, Lou YG, Qian Q, Yamaguchi H, Yamane H, Kong CH, Timko MP, Bai LY, Fan LJ (2017). *Echinochloa crusgalli* genome analysis provides insight into its adaptation and invasiveness as a weed. *Nat Commun* **8**, 1031.
- Haralampidis K, Bryan G, Qi X, Papadopoulou K, Bakht S, Melton R, Osbourn A (2001). A new class of oxidosqualene cyclases directs synthesis of antimicrobial phytoprotectants in monocots. *Proc Natl Acad Sci USA* **98**, 13431–13436.
- Hen-Avivi S, Savin O, Racovita RC, Lee WS, Adamski NM, Malitsky S, Almekias-Siegl E, Levy M, Vautrin S, Bergès H, Friedlander G, Kartvelishvili E, Ben-Zvi G, Alkan N, Uauy C, Kanyuka K, Jetter R, Distelfeld A, Aharoni A (2016). A metabolic gene cluster in the wheat *W1* and the barley *Cer-cqu* loci determines β -diketone biosynthesis and glaucousness. *Plant Cell* **28**, 1440–1460.
- Itkin M, Heinig U, Tzfadia O, Bhide AJ, Shinde B, Cardenas PD, Bocobza SE, Unger T, Malitsky S, Finkers R, Tikunov Y, Bovy A, Chikate Y, Singh P, Rogachev I, Beekwilder J, Giri AP, Aharoni A (2013). Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science* **341**, 175–179.
- Jeon JE, Kim JG, Fischer CR, Mehta N, Dufour-Schroif C, Wemmer K, Mudgett MB, Sattely E (2020). A pathogen-responsive gene cluster for highly modified fatty acids in tomato. *Cell* **180**, 176–187.
- Jonczyk R, Schmidt H, Osterrieder A, Fiesselmann A, Schullehner K, Haslbeck M, Sicker D, Hofmann D, Yalpani N, Simmons C, Frey M, Gierl A (2008). Elucidation of the final reactions of DIMBOA-glucoside biosynthesis in maize: characterization of *Bx6* and *Bx7*. *Plant Physiol* **146**, 1053–1063.
- Jones DA (1998). Why are so many food plants cyanogenic? *Phytochemistry* **47**, 155–162.
- Kakes P (1989). An analysis of the costs and benefits of the cyanogenic system in *Trifolium repens* L. *Theor Appl Genet* **77**, 111–118.
- Kariya K, Ube N, Ueno M, Teraishi M, Okumoto Y, Mori N, Ueno K, Ishihara A (2020). Natural variation of diterpenoid phytoalexins in cultivated and wild rice species. *Phytochemistry* **180**, 112518.
- Komaki H, Sakurai K, Hosoyama A, Kimura A, Igarashi Y, Tamura T (2018). Diversity of nonribosomal peptide synthetase and polyketide synthase gene clusters among taxonomically close *Streptomyces* strains. *Sci Rep* **8**, 6888.
- Li DD, Bi XY, Ma JJ, Zhang XH, Jiang KN, Zhu XZ, Huang JG, Zhou LJ (2022). Natural herbicidal alkaloid berberine regulates the expression of *thalianol* and *marneral* gene clusters in *Arabidopsis thaliana*. *Pest Manag Sci* **78**, 2896–2908.
- Li DP, Gaquerel E (2021). Next-generation mass spectrometry metabolomics revives the functional analysis of plant metabolic diversity. *Annu Rev Plant Biol* **72**, 867–891.
- Liu ZH, Suarez Duran HG, Harnvanichvech Y, Stephenson MJ, Schranz ME, Nelson D, Medema MH, Osbourn A (2020). Drivers of metabolic diversification: how dynamic genomic neighbourhoods generate new biosynthetic pathways in the Brassicaceae. *New Phytol* **227**, 1109–1123.
- Luo C, Fernie AR, Yan JB (2020). Single-cell genomics and epigenomics: technologies and applications in plants. *Trends Plant Sci* **25**, 1030–1040.
- Matsuba Y, Nguyen TTH, Wiegert K, Falara V, Gonzales-Vigil E, Leong B, Schäfer P, Kudrna D, Wing RA, Bolger AM, Usadel B, Tissier A, Fernie AR, Barry CS, Pichersky E (2013). Evolution of a complex locus for terpene biosynthesis in *Solanum*. *Plant Cell* **25**, 2022–2036.
- Nelson D, Werck-Reichhart D (2011). A P450-centric view of plant evolution. *Plant J* **66**, 194–211.
- Netzker T, Fischer J, Weber J, Mattern DJ, König CC, Valiante V, Schroeckh V, Brakhage AA (2015). Microbial communication leading to the activation of silent fungal secondary metabolite gene clusters. *Front Microbiol* **6**, 299.
- Nützmans HW, Doerr D, Ramírez-Colmenero A, Sotelo-Fonseca JE, Wegel E, Di Stefano M, Wingett SW, Fraser P, Hurst L, Fernandez-Valverde SL, Osbourn A (2020). Active and repressed biosynthetic gene clusters have spatially distinct chromosome states. *Proc Natl Acad Sci USA* **117**, 13800–13809.
- Nützmans HW, Huang AC, Osbourn A (2016). Plant metabolic clusters—from genetics to genomics. *New Phytol* **211**, 771–789.
- Nützmans HW, Osbourn A (2014). Gene clustering in plant specialized metabolism. *Curr Opin Biotechnol* **26**, 91–99.
- Nützmans HW, Scazzocchio C, Osbourn A (2018). Metabolic gene clusters in eukaryotes. *Annu Rev Genet* **52**, 159–183.
- Okada BK, Wu YH, Mao DN, Bushin LB, Seyedsayamdost MR (2016). Mapping the trimethoprim-induced secondary metabolome of *Burkholderia thailandensis*. *ACS Chem Biol* **11**, 2124–2130.
- Osbourn A (2010). Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends Genet* **26**, 449–457.

- Park HL, Kim TL, Bhoo SH, Lee TH, Lee SW, Cho MH** (2018). Biochemical characterization of the rice cinnamyl alcohol dehydrogenase gene family. *Molecules* **23**, 2659.
- Qi X, Bakht S, Qin B, Leggett M, Hemmings A, Mellon F, Eagles J, Werck-Reichhart D, Schaller H, Lesot A, Melton R, Osbourn A** (2006). 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* **103**, 18848–18853.
- Qi XQ, Bakht S, Leggett M, Maxwell C, Melton R, Osbourn A** (2004). A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. *Proc Natl Acad Sci USA* **101**, 8233–8238.
- Roddick JG, Weissenberg M, Leonard AL** (2001). Membrane disruption and enzyme inhibition by naturally-occurring and modified chactriose-containing *Solanum* steroidal glycoalkaloids. *Phytochemistry* **56**, 603–610.
- Rokas A, Mead ME, Steenwyk JL, Raja HA, Oberlies NH** (2020). Biosynthetic gene clusters and the evolution of fungal chemodiversity. *Nat Prod Rep* **37**, 868–878.
- Rokas A, Wisecaver JH, Lind AL** (2018). The birth, evolution and death of metabolic gene clusters in fungi. *Nat Rev Microbiol* **16**, 731–744.
- Schneider LM, Adamski NM, Christensen CE, Stuart DB, Vautrin S, Hansson M, Uauy C, von Wettstein-Knowles P** (2016). The *Cer-cqu* gene cluster determines three key players in a β -diketone synthase polyketide pathway synthesizing aliphatics in epicuticular waxes. *J Exp Bot* **67**, 2715–2730.
- Shang Y, Ma YS, Zhou Y, Zhang HM, Duan LX, Chen HM, Zeng JG, Zhou Q, Wang SH, Gu WJ, Liu M, Ren JW, Gu XF, Zhang SP, Wang Y, Yasukawa K, Bouwmeester HJ, Qi XQ, Zhang ZH, Lucas WJ, Huang SW** (2014). Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* **346**, 1084–1088.
- Shen SQ, Peng M, Fang H, Wang ZX, Zhou S, Jing XY, Zhang M, Yang CK, Guo H, Li YF, Lei L, Shi YH, Sun YY, Liu XQ, Xu CP, Tohge T, Yuan M, Fernie AR, Ning YS, Wang GL, Luo J** (2021). An *Oryza*-specific hydroxycinnamoyl tyramine gene cluster contributes to enhanced disease resistance. *Sci Bull* **66**, 2369–2380.
- Smith DJ, Burnham MKR, Edwards J, Earl AJ, Turner G** (1990). Cloning and heterologous expression of the penicillin biosynthetic gene cluster from *Penicillium chrysogenum*. *Biotechnology* **8**, 39–41.
- Takos AM, Knudsen C, Lai D, Kannangara R, Mikkelsen L, Motawia MS, Olsen CE, Sato S, Tabata S, Jørgensen K, Møller BL, Rook F** (2011). 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* **68**, 273–286.
- Tao HY, Zuo L, Xu HL, Li C, Qiao G, Guo MY, Lin XK** (2020). Alkaloids as anticancer agents: a review of Chinese patents in recent 5 years. *Recent Pat Anticancer Drug Discov* **15**, 2–13.
- Tattersall DB, Bak S, Jones PR, Olsen CE, Nielsen JK, Hansen ML, Høj PB, Møller BL** (2001). Resistance to an herbivore through engineered cyanogenic glucoside synthesis. *Science* **293**, 1826–1828.
- Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Židek A, Bridgland A, Cowie A, Meyer C, Laydon A, Velankar S, Kleywegt GJ, Bateman A, Evans R, Pritzel A, Figurnov M, Ronneberger O, Bates R, Kohl SAA, Potapenko A, Ballard AJ, Romera-Paredes B, Nikolov S, Jain R, Clancy E, Reiman D, Petersen S, Senior AW, Kavukcuoglu K, Birney E, Kohli P, Jumper J, Hassabis D** (2021). Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596.
- Varshney RK, Bohra A, Yu JM, Graner A, Zhang QF, Sorrells ME** (2021). Designing future crops: genomics-assisted breeding comes of age. *Trends Plant Sci* **26**, 631–649.
- Verpoorte R, Memelink J** (2002). Engineering secondary metabolite production in plants. *Curr Opin Biotechnol* **13**, 181–187.
- von Rad U, Hüttel R, Lottspeich F, Gierl A, Frey M** (2001). Two glucosyltransferases are involved in detoxification of benzoxazinoids in maize. *Plant J* **28**, 633–642.
- Vranová E, Coman D, Gruissem W** (2013). Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annu Rev Plant Biol* **64**, 665–700.
- Weber T, Kim HU** (2016). The secondary metabolite bioinformatics portal: computational tools to facilitate synthetic biology of secondary metabolite production. *Synth Syst Biotechnol* **1**, 69–79.
- Winzer T, Gazda V, He ZS, Kaminski F, Kern M, Larson TR, Li Y, Meade F, Teodor R, Vaistij FE, Walker C, Bowser TA, Graham IA** (2012). A *Papaver somniferum* 10-gene cluster for synthesis of the anticancer alkaloid noscapine. *Science* **336**, 1704–1708.
- Wu X, Feng H, Wu D, Yan SJ, Zhang P, Wang WB, Zhang J, Ye JL, Dai GX, Fan Y, Li WK, Song BX, Geng ZD, Yang WL, Chen GX, Qin F, Terzaghi W, Stitzer M, Li L, Xiong LZ, Yan JB, Buckler E, Yang WN, Dai MQ** (2021). Using high-throughput multiple optical phenotyping to de-

- cipher the genetic architecture of maize drought tolerance. *Genome Biol* **22**, 185.
- Wu S, Morotti AL, Wang SS, Wang Y, Xu XY, Chen JH, Wang GD, Tatsis EC** (2022). Convergent gene clusters underpin hyperforin biosynthesis in St John's wort. *New Phytol* **235**, 646–661.
- Xu MM, Galhano R, Wiemann P, Bueno E, Tiernan M, Wu W, Chung IM, Gershenzon J, Tudzynski B, Sesma A, Peters RJ** (2012). Genetic evidence for natural product-mediated plant-plant allelopathy in rice (*Oryza sativa*). *New Phytol* **193**, 570–575.
- Yamamuro C, Zhu JK, Yang ZB** (2016). Epigenetic modifications and plant hormone action. *Mol Plant* **9**, 57–70.
- Yang WN, Feng H, Zhang XH, Zhang J, Doonan JH, Batchelor WD, Xiong LZ, Yan JB** (2020). Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. *Mol Plant* **13**, 187–214.
- Yang XF, Gao SH, Guo L, Wang B, Jia YY, Zhou J, Che YZ, Jia P, Lin JD, Xu T, Sun JY, Ye K** (2021). Three chromosome-scale *Papaver* genomes reveal punctuated patchwork evolution of the morphinan and noscapine biosynthesis pathway. *Nat Commun* **12**, 6030.
- Yu N, Nützmann HW, MacDonald JT, Moore B, Field B, Berriri S, Trick M, Rosser SJ, Kumar SV, Freemont PS, Osbourn A** (2016). Delineation of metabolic gene clusters in plant genomes by chromatin signatures. *Nucleic Acids Res* **44**, 2255–2265.
- Yue JP, Hu XY, Huang JL** (2013). Horizontal gene transfer in the innovation and adaptation of land plants. *Plant Signal Behav* **8**, e24130.
- Zhan CS, Lei L, Liu ZX, Zhou S, Yang CK, Zhu XT, Guo H, Zhang F, Peng M, Zhang M, Li YF, Yang ZX, Sun YY, Shi YH, Li K, Liu L, Shen SQ, Wang XY, Shao JW, Jing XY, Wang ZX, Li Y, Czechowski T, Hasegawa M, Graham I, Tohge T, Qu LH, Liu XQ, Fernie AR, Chen LL, Yuan M, Luo J** (2020). Selection of a subspecies-specific diterpene gene cluster implicated in rice disease resistance. *Nat Plants* **6**, 1447–1454.
- Zhan CS, Shen SQ, Yang CK, Liu ZH, Fernie AR, Graham IA, Luo J** (2022). Plant metabolic gene clusters in the multiomics era. *Trends Plant Sci* **27**, 981–1001.
- Zhu GT, Wang SC, Huang ZJ, Zhang SB, Liao QG, Zhang CZ, Lin T, Qin M, Peng M, Yang CK, Cao X, Han X, Wang XX, van der Knaap E, Zhang ZH, Cui X, Klee H, Fernie AR, Luo J, Huang SW** (2018). Rewiring of the fruit metabolome in tomato breeding. *Cell* **172**, 249–261.

Research Progress on Plant Secondary Metabolite Biosynthetic Gene Clusters

Lu Zhu, Chong Yuan, Yifei Liu*

Hubei University of Chinese Medicine, Wuhan 430074, China

Abstract The secondary metabolites produced by plants provide human beings with a wealth of pharmaceutical, perfume and industrial raw materials. With the rapid development of molecular biology and genomics research, the biosynthetic gene clusters (BGCs) of secondary metabolites of various plants have been analyzed, which opens a new path for us to quickly obtain the biosynthetic pathways of target products and discover novel natural products. This paper focuses on the definition and characteristics of plant secondary metabolite biosynthesis gene clusters, and its basic structural models, evolution and regulatory mechanisms, in order to provide theoretical basis and reference for related research.

Key words secondary metabolites, biosynthetic gene cluster, BGCs regulation, BGCs characteristics

Zhu L, Yuan C, Liu YF (2024). Research progress on plant secondary metabolite biosynthetic gene clusters. *Chin Bull Bot* **59**, 134–143.

* Author for correspondence. E-mail: liuyifei@hbtcm.edu.cn

(责任编辑: 孙冬花)