

植物基因表达调控与进化机制研究进展

王子韵^{1,2}, 吕燕文^{1,2}, 肖钰^{1,2}, 吴超^{1,2}, 胡新生^{1,2*}

¹华南农业大学林学与风景园林学院, 广州 510642; ²广东省森林植物种质创新与利用重点实验室, 广州 510642

摘要 功能基因表达是连接基因编码信息与蛋白质产物的一个基本生命过程, 基因表达水平被视为介于基因型与表现型之间的一种数量性状, 在植物应对气候和环境变化时发挥重要作用。该文首先系统综述了植物基因表达调控因子研究进展, 包括转录因子和小RNA等在基因表达调控中的作用。其次, 探讨了基于基因表达数据进行全基因组关联分析(GWAS)估计调控因子基因的表达数量性状基因座(eQTLs)位置以及该方法的局限性。随后从理论上分析了在突变、漂变、选择和迁移过程中的种内基因表达变异与检验方法, 在突变-漂变过程中以及在基于系统发育树的漂变-选择过程中的种间基因表达进化与检测方法。最后, 探讨了植物交配系统对基因表达进化的调控, 自交降低了有效群体大小、突变率、基因重组及外源花粉竞争, 改变了配子与合子阶段的自然选择功效等, 从而间接调控种内基因表达变异和种间基因表达进化。全文综合评述了目前的理论和实际研究进展及存在的问题, 有助于深入理解植物基因表达调控和进化机制。

关键词 基因表达, 调控因子, 自然选择, 数量性状, 交配系统

王子韵, 吕燕文, 肖钰, 吴超, 胡新生 (2025). 植物基因表达调控与进化机制研究进展. 植物学报 60, 621–639.

基因表达是连接基因贮存的遗传信息与基因产物(如蛋白质和非编码RNA)的一个基本生命过程。从基因表达到蛋白质产物的形成涉及转录、mRNA降解、翻译和蛋白质修饰等过程, 每一步都受到基因调控(Schwanhäusser et al., 2011)。早在1961年, Jacques Monod通过大肠杆菌乳糖代谢实验, 发现底物诱导基因表达调控, 并提出了操纵子学说(Jacob and Monod, 1961)。此外, Andre Lwoff、Elie Wollman及Francois Jacob等发现λ噬菌体的溶源现象(Barton et al., 2007)。这两类实验现象均揭示了原核生物基因表达的诱导调控机制。随后, 科学家开展了原核生物其它类型的操纵子和基因表达调控机制研究。真核生物的基因表达调控机制也逐渐被揭示, 其过程涉及基因表达的每一个环节(Singh et al., 2018), 最终使得每个基因在适当的时间和空间展现出合适的表达水平, 构建多细胞生物有机体, 调节特定的组织分化和发育, 以及对生态环境变化的响应和适应。

植物基因时空表达的启动或抑制以及表达水平受多种因素调控, 包括在转录过程中的转录因子和酶

的蛋白质调控, 转录过程和转录后的非编码RNA调控, 以及参与翻译过程和翻译后蛋白质水平的相关因子调控等, 这里将这些调控因素统称为调控因子, 确定这些调控因子及其作用机制是基因表达研究的核心问题。目前已有许多关于调控因子的研究报道(Mattick and Makunin, 2006; 陈玲玲等, 2019; Strader et al., 2022)。随着高通量测序技术的发展和运用, 已建立了多种数据库和分析方法用于鉴定和识别调控因子(李霞和雷健波, 2015), 但这些调控因子的变异和进化最终由调控因子基因或相关DNA序列决定。那么, 如何认识与这些调控因子基因或DNA序列相关的表达调控机制? 调控因子基因在物种内(或群体内)和物种间(或群体间)的进化如何影响靶基因的表达进化? 或反之, 如何从基因表达变异或基因表达进化角度推测调控因子在种内和种间的进化规律? 探讨这些问题有助于深入理解植物基因表达变异与进化机制。

基因表达水平类似于数量性状, 与基因本身的分子进化不同, 其表达水平的变异反映了影响该基

收稿日期: 2024-11-22; 接受日期: 2025-02-09

基金项目: 国家自然科学基金(No.32171819)和华南农业大学人才引进启动项目(No.4400-K16013)

* 通讯作者。E-mail: xinsheng@scau.edu.cn

因表达的调控因子的遗传变异(Mostafavi et al., 2023)。如何准确测定基因的表达水平以及如何利用基因表达水平的变异来推测基因表达进化? 基因表达进化与基因序列变异或其翻译的蛋白质序列变异有怎样的联系? 这些问题有待深入探讨。广义上说, 基因表达水平测定不仅是理解基因功能和调控机制的基础(Gómez-Soto et al., 2021; Irving et al., 2022; Sezen et al., 2024), 也是疾病诊断和药物研发的关键环节(Wu et al., 2020; Parikh et al., 2020; Bernabè et al., 2021; Feng et al., 2021)。近十几年, 高通量测序技术的应用为基因表达测定及其变异分析提供了有效手段。本文将概述基因表达水平的测定方法, 探讨如何利用基因表达水平变异分析调控因子基因的进化机制, 从而理解基因表达进化与基因分子进化之间的关系。

植物与动物的生活史和繁殖方式不同, 在由无性繁殖演化到有世代交替的有性繁殖过程中, 植物交配系统作为重要的生活史性状之一, 可调整亲本基因在子代群体中的基因型分配, 与配子体和孢子体世代选择互作(Zhang et al., 2020), 调整突变负荷在配子体和孢子体阶段的分配(Xiao et al., 2024)。植物繁衍方式演化也会改变其对基因表达的调控能力, 如开花植物从异花受精到自花受精的演化会影响基因分子进化和基因表达进化。此外, 植物繁殖方式与4种基本进化动力(选择、漂变、突变及迁移)互作也会改变基因频率(胡新生等, 2021), 影响基因表达在群体内和群体间的进化。因此, 探讨植物繁殖方式演化如何调控基因表达进化具有重要的理论和实际意义。

本文首先概述基因表达启动和抑制的调控作用, 概括转录因子和非编码RNA的调控机制; 其次, 讨论基因表达水平的测定方法, 探讨利用基因表达水平推测基因表达的调控因子, 重点评价利用基因表达数据进行全基因组关联分析(genome-wide association study, GWAS)获得表达数量性状基因座(expression quantitative trait loci, eQTLs)位点的方法; 然后讨论基于基因表达水平性状分布特征推测调控因子基因在物种内和物种间的潜在进化机制, 综述近期的研究成果; 最后探讨植物交配系统对基因表达的调控作用, 分析交配系统如何调控基因表达在物种内和物种间的进化, 包括自交如何调控基因在配子、合子以及两个阶段都表达的变异与进化机制, 以期

深入理解植物繁衍方式在基因表达变异和进化中的作用。

1 基因表达调控因子与调控机制

依据分子生物学中心法则, 基因表达调控可以发生在转录过程、转录后水平、翻译过程及翻译后水平等阶段。理论上酶蛋白和非编码RNA干扰调控都有可能作用于多个阶段。表1概括了针对不同阶段的调控因子及相关的调控机制(戴灼华和王亚馥, 2016)。下文将概述转录因子和非编码RNA对植物基因表达调控的研究进展。

1.1 转录因子

转录因子是一类蛋白质分子, 在基因表达调控中的作用机制涉及多个途径, 可参与信号转导、与DNA相互作用以及在细胞特异性和发育阶段特异性表达中发挥调控作用等。许多转录因子参与细胞内外的信号转导途径, 如激素信号和环境胁迫反应(戚莹雪等, 2019; 任永娟等, 2021)。例如, 在非生物胁迫下, 通过分析HvNF-YC转录因子家族基因在耐盐型和盐敏感型大麦(*Hordeum vulgare*)品种中的表达量变化, 推测HvNF-YC9、10和11基因簇可能与盐胁迫响应密切相关(王菲菲等, 2023)。陈思思等(2022)对构树(*Broussonetia papyrifera*)转录因子BpbZIP1进行镉胁迫响应分析, 结果表明BpbZIP1基因能积极响应镉胁迫, 其表达可改善构树对镉胁迫的耐受力, 是构树响应镉胁迫的重要基因。

细胞感知胞外环境信号后, 通过信号转导途径将信息传递到胞内。转录因子接收信号并被激活, 通过结合特定的DNA序列, 招募或阻止转录复合体组装, 从而激活或抑制靶基因的转录, 上调或下调靶基因表达水平。例如, SATB1是一种通过其特殊的结构域(ULD和CUTL结构域)特异性结合DNA的转录因子(王峥, 2014)。转录因子能够感知并响应DNA的形状变化, 这在一定程度上解释了为什么某些转录因子能够识别非典型的或变异的DNA序列(Dror et al., 2014)。还有些转录因子(如FOXO1)能够以单体或二聚体的形式与DNA结合, 多聚化后增强转录因子与DNA的结合亲和力(Li et al., 2021)。通过参与信号转导并与DNA相互作用, 最终导致一些转录因子在特定的

表1 基因表达相关调控因子与调控机制

Table 1 Regulatory factors and regulatory mechanisms of gene expression

调控过程	调控因子	调控机制
转录过程	转录因子、激活因子和阻碍物(蛋白质)	结合在靶基因启动子序列的结合位点, 将RNA聚合酶定位于启动子上, 控制基因的转录或转录效率, 或抑制转录启动
	内、外源激素	甾类和多肽激素与细胞质内或核内受体形成复合物, 调控某些基因转录
	lncRNA (>200 bp)	可与蛋白质(转录因子和组蛋白等)互作; 与DNA互作
	siRNA (20–25 bp)	干扰特定基因表达
转录后水平	miRNA (21–23 bp)	与靶mRNA互补配对, 抑制mRNA的翻译或降解靶mRNA
	lncRNA (>200 bp)	与RNA (mRNA、miRNA及其它lncRNA)互作
翻译过程	酶蛋白	mRNA选择性和反剪接; RNA编辑
翻译后水平	翻译起始因子等	翻译起始因子的可逆磷酸化; 非翻译区结构修饰变异
	肽链剪接酶、化学修饰酶等	新生肽链剪接, 化学修饰; 肽链折叠; 蛋白质更换等

细胞类型或发育阶段特异性表达, 从而调控细胞特异性的基因表达模式(Mills and Taghert, 2012), 使不同细胞类型执行各自独特的功能。还有一些转录因子在特定的发育阶段调控植物生长发育的时序性, 涵盖从胚胎发育到器官形成, 再到生殖生长等各个重要阶段。例如, 孟令桐等(2024)对毛果杨(*Populus trichocarpa*) ANT转录因子进行研究, 发现PtrANT-4基因在形成层中特异性表达, 这表明ANT转录因子可能在维管形成层干细胞分裂分化过程中扮演重要角色。李佳祯等(2024)通过分析黍子(*Panicum miliaceum*)转录组数据, 发现转录因子PmSRS表现出组织特异性, 其在8周龄的茎、花序和根中高表达, 但在8周龄的叶片中几乎不表达。

植物转录因子包括多个家族, 每个家族包含多种成员, 它们可以识别相似但略有不同的DNA序列, 从而在不同的细胞类型和生理条件下发挥不同的调控作用。以植物WRKY转录因子家族为例, 不同种类植物中WRKY基因家族成员的数量存在差异, 它们通过特异性结合启动子区域的W-box来调控基因表达。在拟南芥(*Arabidopsis thaliana*)中共发现90个WRKY成员, 而在沟叶结缕草(*Zoysia matrella*)中则发现269个WRKY成员(Tian et al., 2020)。近期对植物WRKY基因家族进行全基因组鉴定, 并分析其结构、功能以及在不同组织或生长发育过程中的表达模式(黄诗宇等, 2024; 黄金龙等, 2024)。有研究表明, WRKY转录因子在病害抗性(潘钰, 2019; 于晓跃, 2021)、发育和激素控制过程(陈宇翔, 2020; Chen et al., 2021; 柳金华, 2023), 尤其是非生物胁迫响应(张艺腾, 2022;

杨建坤, 2023; 穆宣邑等, 2024)等方面发挥重要作用。苏文娟等(2023)研究发现, 油茶(*Camellia oleifera*)大部分CoWRKYs基因正向调控抗逆胁迫响应, 其中部分基因受干旱和盐胁迫诱导上调表达, 表明WRKY转录因子参与调控非生物胁迫响应。

1.2 非编码RNA

非编码RNA (non-coding RNA, ncRNA)参与调控基因表达、细胞分化和发育等多个生物学过程。根据长度和功能的不同, 非编码RNA可分为小RNA (small RNA, sRNA)和长链非编码RNA (long non-coding RNA, lncRNA)等(陈玲玲等, 2019; Zheng et al., 2019)。这些非编码RNA分子通过与DNA、mRNA、蛋白质以及miRNA等相互作用, 在表观遗传、转录、转录后、翻译及翻译后等多个水平上调控基因表达(Zhang et al., 2019) (表1)。

1.2.1 小RNA

植物中主要有2类小RNA参与基因表达调控, 即微小RNA (microRNA, miRNA)和小干扰RNA (small interfering RNA, siRNA) (Song et al., 2021)。snoRNA (small nucleolar RNA)属于核仁小RNA, 主要参与核糖体RNA的修饰。miRNA通过互补配对原则作用于靶基因, 进而调控植物的多种生物学过程(Bonnet et al., 2006; Mallory and Vaucheret, 2006)。其通过生物合成、装载到RNA诱导沉默复合体(RNA-induced silencing complex, RISC)以及识别靶mRNA来调控基因表达。该过程主要涉及mRNA的降解或翻译抑制,

影响植物的生长发育、环境适应性以及对生物/非生物胁迫的响应(Hausser et al., 2013; Wu et al., 2023)。miRNA与靶mRNA结合后,可阻止核糖体移动,从而抑制蛋白质的翻译。在某些情况下,如果miRNA与靶mRNA的配对非常完美, RNA诱导沉默复合体(RISC)中的活性核酸酶可以切割mRNA,导致mRNA降解。

近期研究发现,一些miRNA在细胞核中也具有功能,参与调控靶基因的转录激活或沉默(Stavast and Erkeland, 2019)。miRNA在植物生长发育、激素信号转导以及应对环境胁迫等多个方面发挥重要作用,能够参与调控植物的花期、器官形态建成、应对干旱和盐碱等非生物胁迫以及病原体防御等过程(Shriram et al., 2016; Waheed and Zeng, 2020; Yang et al., 2021)。

miRNA也通过与其它分子(如转录因子和lncRNA)相互作用,构建复杂的调控网络(Samad et al., 2017; Yang et al., 2021)。有些miRNA通过影响特定转录因子的表达或活性来间接调控下游基因的表达(Samad et al., 2017),或与激素信号互作(如与生长素信号途径的TIR1/AFB互作),调控植物的生长发育和应对环境变化(Luo et al., 2022)。

siRNA主要由类似于Dicer酶作用产生,这些酶将双链RNA切割成短小的干扰RNA片段,通过与Argonaute蛋白形成复合物(RISC)作用于目标mRNA,导致其降解或翻译被抑制(Shamandi et al., 2015)。利用经过筛选的siRNA可以特异性下调特定的基因或基因组表达,还可以调控基因沉默的程度、时间和范围。这在植物的基因表达调控中发挥重要作用,尤其是在对抗转座元件和病毒的防御机制中,是植物防御病毒的重要机制之一。但病毒也会形成逃避干扰机制,一些病毒编码的抑制蛋白能够结合并消耗siRNA,从而减少植物体内有效的抗病毒siRNA数量(Pertermann et al., 2018)。在植物与病毒的相互作用中,除了miRNA直接介导的基因沉默外,还有复杂的分子机制参与调控siRNA的生成和功能。

尽管miRNA比siRNA可以干扰更多的靶mRNA,但两者可以通过多种机制相互作用,影响彼此的功能。它们都可以通过与靶mRNA的互补配对来实现基因沉默,但miRNA可以通过影响siRNA-PEI聚集体的稳定性间接影响siRNA的功能。miRNA能够与siRNA聚集体形成复合物,影响siRNA的传递效率和亚细胞

定位(Meneksedag-Erol et al., 2015),这表明miRNA可以通过影响siRNA的物理化学性质间接调控基因表达,但二者的具体调控机制有待进一步研究。

1.2.2 长链非编码RNA

长链非编码RNA是一类长度超过200 bp的非编码RNA分子,其在转录、转录后和翻译水平等阶段调控植物基因表达。环状RNA(circRNA)属于一类特殊的lncRNA。lncRNA通过与DNA、RNA或蛋白质相互作用,在转录水平上调控基因表达。在小麦(*Triticum aestivum*)中,lncRNA GARR2能够调控与赤霉素和生长素相关的差异表达基因,从而影响小麦的株高和叶鞘长度等性状(胡卫国, 2020)。在白桦(*Betula platyphylla*)中,lncW20通过结合BpPOD59和BpCY-P76G1基因的启动子,上调其表达水平,提高植物的抗氧化能力,从而增强抗旱性(赵慧敏, 2022)。

lncRNA在调控植物开花时间、器官发生以及应对生物和非生物胁迫等方面发挥作用。Shea等(2019)通过对春化过程中的蔓菁(*Brassica rapa*)转录组进行分析,发现许多lncRNA在冷处理后表达发生显著改变,部分差异表达的mRNA与其对应的lncRNA在表达上呈现协同变化的趋势。Wang等(2021a)发现了多个与次生细胞壁生物合成相关的lncRNA及其调控网络,明确了其在毛竹(*Phyllostachys edulis*)快速生长过程中的表达模式和潜在功能。Palos等(2022)对十字花科(Brassicaceae) 4种植物的研究表明,约10%的lncRNA响应胁迫。这些响应多与温度胁迫有关,部分与脱落酸(ABA)响应相关,且呈现特异性表达模式。综上所述,lncRNA在植物生长发育及应对环境胁迫过程中具有重要作用,其功能多样且调控机制复杂,未来还需深入研究其作用机制。

2 基因表达水平与调控因子估计

受测定技术成本高等因素限制,蛋白质组学技术尚未广泛应用于鉴定基因表达,目前更多是以转录后的mRNA拷贝数来度量基因表达水平,而不是以基因翻译后蛋白质产物来表示。基因表达水平反映了转录启动后植物生长代谢或对环境变化响应的需求程度,虽然其生物学意义早已被认知,但其准确且大规模测定技术是在PCR技术发展后建立起来的。在实际测定

时, 通常用qRT-PCR技术测定RNA分子含量(Bustin, 2000; Gachon et al., 2004), 该技术结合了RT-PCR和qPCR的原理, 通过反转录酶将RNA反转录成cDNA, 随后利用qPCR技术对cDNA进行定量分析(Vanguilder et al., 2008)。但样本量、RNA完整性、反转录效率和cDNA质量等因素均会影响基因表达测定结果的可靠性(Vandesompele et al., 2002; Nolan et al., 2006), 通常选择稳定表达的内参基因作为对照, 对目的基因表达量进行校正以提高测定结果的准确性(Zhang et al., 2017)。

基因表达水平可以用绝对和相对表达量来表示, 前者指在特定条件下, 基因的表达量可以被直接量化为一个具体的数值, 这种方法通常需要使用标准样品或已知浓度的样品做参考(Bustin, 2000; Eppler et al., 2005), 实验要求较高, 步骤较烦琐。后者则是通过比较目标RNA分子在不同样品或不同条件下的表达水平变化来估计其相对表达量(Fleige et al., 2006), 这种方法通常选择1个或多个内参基因作为对照, 以校正样品间的差异。衡量基因相对表达量常用的方法有 $2^{-\Delta\Delta CT}$ 法(Livak and Schmittgen, 2001)和Pfaffl法(Pfaffl, 2001)。应用RNA-seq分析时, 同时测定多基因表达水平, 为消除基因长度和测序深度对基因表达水平的影响, 需将数据进行归一化处理(Evans et al., 2018), 常用RPKM (reads per kilobase per million)、FPKM (fragments per kilobase per million) (Mortazavi et al., 2008)以及TPM (transcripts per million)表示(Wagner et al., 2012)。

基因表达在生态与进化生物学方面有广泛应用。例如, Palma-Silva等(2024)利用基因表达评价竹芋科细穗竹芋属植物*Ischnosiphon colombianus*在亚马逊河岸森林系统对2种环境的响应, 认为不同环境条件和洪涝灾害是种群遗传分化的驱动因素。Andrew等(2024)研究了澳大利亚17个相思树属(*Acacia*)树种响应热胁迫后的基因表达与环境关联性。基因表达测定也常用于分析特定基因在植物不同组织器官中的表达定位, 以及比较物种间或不同环境条件下的基因表达模式差异等(Wu et al., 2024), 依据统计学检验筛选表达差异显著的基因并进行注释, 但这类研究一般不探究基因表达调控因子。

基因表达数据的另一重要应用是推测基因表达的调控因子, 利用基因表达数据进行GWAS, 检测基

因表达与全基因组SNP的关联, 并寻找*cis*-和*trans*-eQTLs (Nica and Dermitzakis, 2013), 该分析与传统的基于表型数量性状进行GWAS不同。有研究显示, *cis*-eQTLs更倾向于转录起始位点处(Mostafavi et al., 2023), 与利用表型数量性状进行GWAS所揭示的变异区域不同。近年来, 利用基因表达量进行GWAS, 并由eQTLs定位推测控制基因表达的调控因子, 在人和动植物GWAS研究中均取得了一定进展(Mai et al., 2023)。例如, Vösa等(2021)利用人血液基因表达分析了大量的*cis*-以及*trans*-eQTLs, 利用单细胞RNA-seq数据分析*trans*-eQTLs, 证实了*trans*-eQTLs主要通过转录因子调控基因的表达。Brümmer和Bergmann (2024)分析了基因外显子和内含子转录水平及其在人淋巴母细胞系中的比例, 结果显示外显子*cis*-QTLs优先捕获转录调控的遗传效应, 而基因外显子和内含子表达水平的*cis*-eQTLs比率可以更好地检测到共转录和转录后过程。以动物组织为研究材料也有报道。Cai等(2023)利用eQTL定位分析与肉牛肌肉、背膘和肝组织关联的基因型变异, 获得了许多*cis*-eQTLs位点, 这些位点靠近基因的转录起始位点, 富集在功能调控区域。Farhangi等(2024)利用猪肝、肺、脾及肌肉组织的转录组数据进行GWAS, 鉴定eQTL调控区域和转录因子。关于植物基因表达的GWAS报道较少。Liu等(2022)对287个水稻(*Oryza sativa*)品种的叶片转录组进行分析, 利用基因表达数据, 通过GWAS鉴定出调控多个基因表达的eQTLs, 筛选出21个与抽穗期旗叶淀粉含量相关的候选基因, 并发现一个二萜类抗毒素合成的重要调控因子, 可增强水稻的抗病性。Zhang等(2023)利用美洲黑杨(*Populus deltoides*)和小叶杨(*P. simonii*)及其F₁杂种木质部的转录组, 鉴定出38个与木材基本密度关联的eQTLs, 其中1个参与调节次生细胞壁相关基因的表达。

理论上, 利用基因表达数据进行GWAS只能找到一部分基因表达的调控因子, 在实际分析时, 以下几个问题值得关注。

(1) 类似于用表型数量性状进行GWAS, 用基因表达数据进行GWAS也受样本容量影响, 这由所用的统计检测功效决定。当转录组分析样本小时, 效应较小的*trans*-和*cis*-eQTL容易被漏检。

(2) 用基因表达数据进行GWAS的遗传基础是控

制基因表达水平的eQTL与SNP存在连锁不平衡(LD), 因此, 统计检验为显著关联的eQTLs有可能为非功能上互作的位点, 导致产生假阳性eQTL位点。当样本来源于不同群体结构的单株混合时, 容易产生过多的非功能上互作位点间LD, 类似于基因流产生的LD (胡新生等, 2021)。当样本为随机交配群体时, 只有那些与eQTLs紧密连锁的SNPs才有可能存在显著的LD, 容易检测到真正的cis-eQTLs。在实际分析时, 要清除Hardy-Weinberg不平衡的SNPs以及多态性较小的SNPs (如MAF<5%), 同时控制整个测验的显著水平。

(3) 当测定的基因表达源于临时的不同外显子区域组合 (如断裂基因) 时, 利用基因表达数据进行GWAS, 估计基因内含子的调控作用仍具挑战 (Brümmer and Bergmann, 2024)。由于内含子的调控作用也可能是不同内含子或内含子-外显子组合作用, 这些效应很难由基因表达数据估计, 这类似于估计数量性状变异的上位性效应。

(4) 目前获得的样本基因表达数据多数是基因所有细胞的表达均值, 不区分纯合子或杂合子基因表达, 不区分细胞类型调控差异。近期有学者建议利用单细胞组学研究特定细胞类型的基因表达, 然后进行eQTL分析 (Luo et al., 2023)。

总之, 基于基因表达数据进行GWAS估计调控因子仍然是一个重要途径, 是对现有转录因子进行筛选和识别分析的补充, 筛选出的调控因子通常需要进行一步实验验证 (Wang et al., 2021b; Peleke et al., 2024)。

3 基因表达的进化机制

3.1 基因表达水平变异特征

与一般数量性状相比, 基因表达水平也呈连续分布, 易受遗传和环境影响, 但仍有其自身变异特点。

(1) 单个基因表达已有明确的基因类型, 表达水平受调控因子和基因本身序列进化影响, 后者在基因表达进化研究中常被忽略, 但可以从基因表达进化、选择程度($4N_e s$, s 为突变等位基因选择系数)及基因分子进化速率(K_a)三者关系中推测 (Kimura, 1962): 对于有利突变等位基因($s>0$), $K_a = \mu \cdot \frac{4N_e s}{1 - e^{-4N_e s}}$ (μ 为突变

率); 对于有害突变等位基因 ($s<0$), $K_a = \mu \cdot \frac{4N_e s}{e^{4N_e s} - 1}$ 。有利突变等位基因在群体中固定概率大,

分子进化速率高, 当群体中有利等位基因比有害等位基因频率高时, 有利等位基因在群体基因表达水平的变异中会占主要部分, 且比有害等位基因表达水平高, 反之亦然。

(2) 一般数量性状受多个基因表达控制, 包含细胞、组织及更高层次上的基因与基因互作效应, 基因间互作产生更复杂的网络。与此类似, 一个基因的表达水平也受到多种调控因子基因作用, 即由多基因效应组成 (图1)。这些调控因子也可形成复杂的互作网络, 但更多倾向于产生细胞或组织水平上的变异。

(3) 通常将基因表达水平的变异归因于顺式和反式调控因子作用的遗传效应 (Hill et al., 2021), 如cis-和trans-eQTL效应, 很少与基因表达变异隐含的加性、显性和上位性效应联系起来。Cis-eQTL效应可能以加性和加性×加性为主, 而trans-eQTL效应主要为加性×加性和加性×非加性效应。顺式和反式作用效应的相对贡献在不同基因表达中不同 (Hill et al., 2021)。一般 (或复杂) 数量性状遗传变异主要由加性效应控制 (Hill et al., 2008), 基因表达变异是否具有相同属性仍需要实验验证。

(4) 与一般数量性状 (如树高和胸径) 不同, 基因表达对时空或环境条件变化更敏感, 时间短, 需要严格的环境试验设计。如同质园试验, 取样考虑同一背景误差或植物相同发育阶段等因素。同样, 若要估计基因表达的遗传方差 (如加性和非加性) 或遗传力等参数, 还要考虑遗传交配设计或亲缘关系明确的谱系材料, 目前这方面研究很少。

(5) 与一般数量性状不同, 高通量测序使得RNA-seq分析产生大量的基因表达数据, 给数据分析带来挑战 (如高维统计分析), 存在受样本容量限制以及需注意控制统计分析误差等问题。

3.2 种内基因表达变异与检测

类似于数量性状适应性检测, 可通过多种途径检测单个基因表达是否为中性。在短期漂变过程下, 单个基因表达的调控因子基因在群体内的遗传多样性和加性方差会随世代而逐渐减少 (二倍体杂合子频率或加

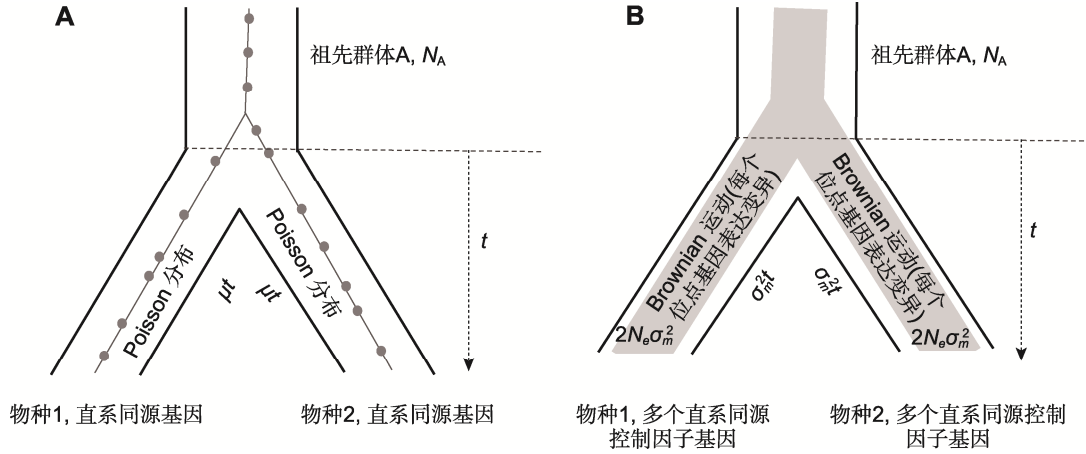


图1 直系同源基因序列进化与直系同源基因表达进化比较

(A) 一对直系同源基因序列进化: 在中性过程下, 物种内基因序列上的突变数用Poisson分布描述, 每个子代物种累积突变数为 μt , μ 为每个碱基的突变率, t 为两物种分化时间。在突变-漂变作用平衡时, 两直系同源基因的总突变数 $K=2\mu t$; **(B)** 一对直系同源基因表达进化: 每个子代物种内, 1个基因表达受许多控制因子基因的影响, 在漂变过程中, 基因表达水平在物种内用Brownian运动分布表示, 在突变-漂变作用平衡时, 物种内基因表达的加性遗传方差为 $2N_e\sigma_m^2$, σ_m^2 为单世代突变产生的方差, 每个子代物种内积累的突变方差为 $2N_e\sigma_m^2 t$, 两子代物种间基因表达水平平均值的加性方差为 $\sigma_B^2=2\sigma_m^2 t$ 。

Figure 1 Comparison of the gene sequence evolution versus the gene expression evolution in a pair of orthologous genes **(A)** Gene sequence evolution of a pair of orthologous genes: under neutral processes, the number of mutations in a gene sequence within a species can be described by a Poisson distribution, the cumulative number of mutations in the gene sequence is approximated by μt , where μ is the mutation rate per base and t is the divergence time between two species. When mutation-drift equilibrium is reached, the total number of mutations in two orthologous genes is $K=2\mu t$; **(B)** Gene expression evolution of a pair of orthologous genes: gene expression in each descendant species is controlled by multiple orthologous genes for regulatory elements, under the process of genetic drift, the distribution of gene expression levels within a species can be described by the Brownian motion distribution, under equilibrium between mutation and drift effects, the additive genetic variance of gene expression within a species is equal to $2N_e\sigma_m^2$, where σ_m^2 is the variance generated by mutations per generation, the cumulative mutation variance within each descendant species is $2N_e\sigma_m^2 t$, and the additive genetic variance of the mean gene expression levels between two descendant species is $\sigma_B^2=2\sigma_m^2 t$.

性方差衰减速率 $= \frac{1}{2N_e}$), 直到等位基因固定或丢失, 群体内个体间基因表达方差也逐渐减小, 多位点基因表达方差均呈现类似的变异模式:

$$\sigma_A^2(t) = \sigma_A^2(0) \left(1 - \frac{1}{2N_e}\right)^t \quad (1)$$

因此, 在短期内, 通过测验各基因表达的加性方差 $\sigma_A^2(t)$ 的差异可以推测各表达基因的初始方差是否相等。若调控因子基因受定向选择影响, 该基因的表达水平方差也会呈现减小模式, 与漂变过程结果类似。

在长期漂变与突变两过程作用下, 群体内基因表达的加性方差会趋于稳定, 达到平衡时, $\sigma_A^2=2N_e\sigma_m^2$,

σ_m^2 为突变产生的突变方差 (Walsh and Lynch, 2018), 基因表达的非加性方差趋于零。基因表达水平趋于稳定正态分布, 基因序列的遗传多样性也趋于稳定, 如平衡时杂合子频率期望值 $=\theta/(1+\theta)$, 式中 $\theta=4N_e\mu$ 。如群体分化成 L 个亚群体, 每个亚群体有效群体大小为 $N_{e(l)}$ 时, 整个群体(种群)的有效群体增加, 加性方差增大, 依据Wright (1943)以及Walsh和Lynch (2018)的研究方法, 加性方差为:

$$\sigma_A^2 = \frac{2LN_{e(l)}}{1-F_{st}} \sigma_m^2 \quad (2)$$

F_{st} 为亚群体间的遗传分化系数。该基因在整个群体内的表达变异加大, 总的加性方差趋于稳定。

上述为在中性过程下(假设迁移的作用可忽略),

每个基因表达水平的期望变异特征。假设每个调控因子基因的效应是微效且近似相等,且每个基因的控制因子基因突变引起相同的突变方差,则不同基因表达的加性方差应该相等(H_0 假设: $\sigma_{A(i)}^2 = \sigma^2$, $i=1, 2, \dots, n$),通过多个基因表达水平变异的卡方或F测验,推测群体内各表达基因的调控因子基因作用是否相等。

当基因表达受自然选择影响时,对生物个体生存和繁殖有利的基因及其表达模式得以保留和传播,而不利的基因表达模式则逐渐被淘汰,从而导致种群基因表达呈现非中性变异特征。与检验基因序列变异是否为中性变异(如Tajima's D 测验)不同,要测定各基因表达是否为非中性,需借助数量和群体遗传学方法(胡新生等, 2021),可以采用以下途径进行。

(1) 利用同质园试验。当试验材料是1个群体时,通过基因表达量与适应性表型性状相关分析,鉴定自然选择是否参与基因表达;当试验材料是多个群体(如林木种源试验),可以测试基因表达与种源地气候或环境因子的关联性来推测是否存在选择(Andrew et al., 2024);当试验设计为多年多点试验时,通过测验基因表达与地点交互效应,推测表达基因是否存在选择。

(2) 利用两世代的调查数据进行回归分析,测验基因表达与有关适应性数量性状的关联性来推测是否存在选择。基于适合度的自然选择变化估计选择差 S , 相应的群体基因表达量的均值变化(跨越1个世代的选择响应) R , 育种学上又称为遗传增益,根据育种者方程(breeder's equation)估计基因表达的选择梯度参数 β , 即 $\beta=R/S$, 选择梯度 β 可用于指示表达基因的选择强度, β 数值大小反映了该基因表达的选择强度。该方法在水稻育种中得到应用,发现在湿稻田条件下,多数基因表达处于中性或弱稳定选择(stabilizing selection),但在干旱条件下处于强选择,且基因表达的选择强度与 *cis*-调控因子表达水平及网络连接性存在弱负相关(Groen et al., 2020)。

(3) 利用中性分子标记和基因表达数据分析群体间分化程度,推测基因表达是否存在自然选择。采用相同群体样本,用分子标记(如卫星DNA标记)估计群体分化系数 F_{st} , 同时采集有关组织样本进行RNA-seq分析,获得基因表达水平数据并估计群体间的分

化系数 Q_{st} , 根据统计检验 F_{st} 与 Q_{st} 差异是否显著。当 $Q_{st} > F_{st}$, 该基因表达在群体间存在分化选择;当 $Q_{st} = F_{st}$, 该基因表达趋于中性,与漂变产生的结果一致;当 $Q_{st} < F_{st}$, 该基因表达在群体间存在相同选择。实际分析时,所研究的群体个体间的亲缘关系通常不确定(如林木天然群体),可用基因表达水平计算群体分化来近似 Q_{st} , 同时也需要注意选择合适的分子标记以及 Q_{st} 和 F_{st} 差异的统计检验问题等因素的影响(Walsh and Lynch, 2018)。应用该方法分析基因表达群体分化,在动物中已有研究报道(Roberge et al., 2007),但在植物中尚未见报道。

此外,与上述分析相关的基因表达遗传力估计问题报道较少,该参数反映各基因表达的遗传稳定性,与 *cis*-和 *trans*-eQTL加性效应关联。理论上根据谱系材料亲属间关系和试验的环境设计,可以估算出基因表达的加性方差和遗传力。

3.3 种间基因表达进化与检测

通常用统计分析检测种间基因表达差异是否显著,这些分析均基于假设基因表达在种间相互独立,无种间系统发育关系带来的基因表达协方差影响。当种间谱系分化时间短时,祖先群体的基因多态性遗留在子代种群中,如红椿(*Toona ciliata*)变种保留着大量祖先的多态性位点(王茜, 2023),这自然会增加种内基因表达方差,减少种间基因表达方差,使得基因表达受稳定选择影响这一推测的可信度增加。当种间分化时间足够长时,祖先群体基因多态性位点遗留殆尽或在祖先群体中等位基因已固定,种内基因表达变异和种间基因表达进化主要由物种分化后所经历的进化过程产生。在此基础上进行以下探讨。

假设每个基因表达的控制因子基因在祖先群体中为纯合子,或祖先遗留下来的多态性位点很少,这些控制因子基因在物种分化后经历各种进化过程。种间基因表达的分化程度实际上反映了控制该基因表达的调控因子基因的种间分化,同时也与基因序列本身的进化有关。理论上,4种基本进化动力(漂变、突变、选择和迁移)之一或联合作用均可影响种间基因表达进化。

通过统计检验短期内漂变促进基因表达进化是否显著来检测种间基因表达进化是否由纯漂变产生。

理论上, 假设两物种的有效群体大小均为 N_e , 基因表达水平在起始(祖先)群体的加性方差为 $\sigma_A^2(0)$, 只考虑调控因子的加性效应, 种间基因表达的均值方差可以表示为(Walsh and Lynch, 2018):

$$\sigma_B^2(t) = 2(1 - e^{-t/2N_e})\sigma_A^2(0) \quad (3)$$

随着物种分化时间(t)的增长, 基因表达均值的方差就越大。实际测验时, 可应用单因素方差分析(ANOVA)估计种群间基因表达方差 $\hat{\sigma}_B^2(t)$, 应用Lande (1976)提出的 F 测验:

$$F = \frac{\hat{\sigma}_B^2(t)}{t \times \hat{\sigma}_A^2(0) / N_e} \sim F_{1, df} \quad (4)$$

df 为估计 $\sigma_A^2(0)$ 时的自由度(Walsh and Lynch, 2018)。该分析方法适合测验短期内($t < N_e$)漂变对基因表达进化的影响。该测验的严格假设就是基因表达仅受加性影响, 当 $\hat{\sigma}_B^2(t)$ 也包含上位性效应时, 上述测验会产生偏差。

也可用布朗运动(Brownian-motion, BM)模型来测验种间基因表达进化, 基因表达水平(x)概率密度服从正态分布:

$$f(x | x_0, \sigma, t) \sim N(x_0, t\sigma^2) \quad (5)$$

式中 x_0 为基因在祖先群体中的表达水平或物种分化初始时的表达水平, σ^2 为BM过程中基因表达方差, t 为种群分化时间。Lande (1976)在分析时, 采用基因表达均值分布, 即 $f(\bar{x}) \sim N(x_0, t\sigma_A^2 / N_e)$, 方差随着物种分化时间 t 而增加。依据该分布可以构建出有效群体大小(N_e)置信区间, 并与实际估值比较来判断是否接受漂变效应。理论上, 可同时考虑基因在祖先群体中的表达水平(x_0)分布和子代物种间的表达分布(x_0), 以估计相关的基因表达参数(Gu, 2010)。

可以用不同模型构建统计测验以检测突变对基因表达进化的影响。当种群分化时间较长时, 种间基因表达水平均值的加性方差可表示为:

$$\sigma_B^2(t) = 2\sigma_m^2 t + 2(\sigma_A^2(0) - 2N_e\sigma_m^2)(1 - e^{-t/2N_e}) \quad (6)$$

当分化时间 t 趋于无穷大时, 种间基因表达均值的加性方差为 $2\sigma_m^2 t$, 该数值类似于两直系同源基因序列在物种分化后积累的突变数 $2\mu t$ (图1)。类似于公式(4)测验, 也可以建立 F 分布测验种间基因表达进化是

否显著 ($F = \hat{\sigma}_B^2(t) / (2t\hat{\sigma}_m^2)$)。也可以利用基于漂变-突变中性过程来间接判断突变影响, 采用BM模型构建突变遗传力(h_m^2) (定义为突变方差 σ_m^2 与环境方差 σ_m^2 之比)的置信区间, 来判断突变效应是否显著(Walsh and Lynch, 2018)。

目前已有一些方法检测种间基因表达进化是否受自然选择影响(表2)。

(1) 类似于研究分子进化中的MK (McDonald-Kreitman) (McDonald and Kreitman, 1991)方法(检测种间基因序列进化是否存在选择作用), 假设1个基因表达的所有调控因子基因在种内变异和种间进化完全由突变产生, 理论上种内基因表达方差和种间基因表达方差相等(H_0 假设), 备选假设(H_1)为种间基因表达进化存在稳定选择或歧化选择(disruptive selection), 通过测验推测是否存在选择效应。该方法的缺点是假设种间是独立的, 忽视了种间系统发育关系对基因表达进化的影响, 其次, 基因表达进化随时间增加且无上限, 漂变导致种间基因表达进化可能会被误认为是定向选择(Price et al., 2022)。

(2) 用BM模型(漂变-突变中性过程)作为 H_0 假设, OU (Ornstein-Uhlenbeck)模型为备选假设(H_1), 测验种间基因表达进化是否由稳定选择产生。基因表达水平(x)密度分布函数 f 为:

$$f(x | x_0, O, \lambda, \sigma, t) \sim N\left(x_0 e^{-\lambda t} + O(1 - e^{-\lambda t}), \frac{\sigma^2}{2\lambda}(1 - e^{-2\lambda t})\right) \quad (7)$$

式中 O 为表达水平在种内的最优表达水平值, λ 为选择强度(Bedford and Hartl, 2009)。随着分化时间 t 的延长, 基因表达密度分布趋于 $\lim_{t \rightarrow \infty} f(x | x_0, O, \lambda, \sigma, t) \sim N\left(O, \frac{\sigma^2}{2\lambda}\right)$ 分布, 方差趋于稳定, 均值趋于最优表达值, 通过建立似然函数可以获得参数 O 、 λ 及 σ 的极大似然估计值。该模型也可以拓宽到分析特定的进化分支, 允许有不同的最佳表达值, 类似于定向选择模型。目前该检测方法已得到一定应用(Price et al., 2022), 有开发的R程序包OUwie (Beaulieu et al., 2012)和geiger (Pennell et al., 2014)函数。该方法的缺点是没有利用种内基因表达变异信息, 未考虑环境差异等因素带来的基因表达变异。

表2 基于种间基因表达进化的选择检验

Table 2 Selection tests based on the evolution of gene expression among species

检测方法	假设	参考文献
种内变异-种间分化比较 (纯突变过程)	H ₀ : 种间基因表达进化等于种内基因表达变异 H ₁ : 稳定选择或歧化选择	Oleksiak et al., 2002; Rifkin et al., 2003; Khaitovich et al., 2004; Lemos et al., 2005; Gilad et al., 2006
基于系统发育树 (考虑种间基因表达均值变异)	H ₀ : 种间基因表达进化遵循BM (Brownian-motion)过程, 类似于漂变过程; 只考虑基因表达均值特征 H ₁ : 种间基因表达进化遵循OU (Ornstein-Uhlenbeck)过 程, 受稳定选择或定向选择	Butler and King, 2004; Bedford and Har- tl, 2009; Gu, 2010
基于系统发育树 (考虑种内和种间基因表达变异)	H ₀ : 种内和种间基因表达进化遵循OU过程, 受稳定选择 H ₁ : 种内和种间基因表达进化遵循OU过程, 受定向选择	Rohlf and Nielsen, 2015

(3) 在上述OU模型基础上, 同时考虑种内基因表达方差和种间基因表达进化方差, Rohlf和Nielsen (2015)提出了基因表达变异和进化(expression variance and evolution, EVE)模型(表2), 该模型假设第*i*个物种的第*k*个体某个基因表达水平 x_{ik} 服从正态分布:

$$x_{ik} \sim N\left(a_i, \beta_i \frac{\sigma_i^2}{2a_i}\right) \quad (8)$$

式中, a_i 为该基因在物种 *i* 的平均表达水平, 由OU过程决定, σ_i^2 为种间基因表达进化方差, a_i 为对于基因表达的选择适应速率(与自然选择强度相关), β_i 为种内表达方差与种间表达进化方差之比, 在稳定选择下 β_i 趋于一个常数。根据多基因表达数据以及种间系统发育树(各分支的时间), 建立似然函数并估计每个基因的所有参数。通过检测每个基因表达的 β_i 估计值是否与多基因共同的 β_{shared} 有显著差异(H₀: $\beta_i = \beta_{shared}$), 来推测选择类型(β_{shared} 是利用所有基因表达水平值估计的参数值)。少数案例显示EVE模型得出基因表达较多受定向选择, 而OU模型分析结果更倾向于稳定选择类型(Price et al., 2022)。注意该模型依据 β_i 大小来判断选择类型($\beta_i > \beta_{shared}$ 为物种或特定物种分支定向选择; $\beta_i < \beta_{shared}$ 为环境因素引起变异或种内多样化选择), 未用中性过程作为参考背景。

目前还没有合适的方法检测种间基因表达进化是否受基因迁移影响。尽管已有许多检测种间基因渐渗的方法(程祥等, 2020), 但鲜有案例报道基因渐渗对种间基因表达进化的影响。植物种间杂交频繁发生(~25%的植物种存在这种情况) (Mallet, 2005), 尤其

是快速辐射形成物种或近缘种之间, 基因渐渗现象较为普遍(吕燕文等, 2024)。理论上, 基因渐渗改变了物种基因组的部分组成, 对种间基因表达与基因本身的适应性都有影响(Li et al., 2023)。对于中性或近中性基因及其调控因子基因, 基因渐渗的作用类似于突变, 有助于产生种内基因多态性, 提高种内基因表达的方差, 同时降低种间基因表达进化方差, 应用EVE或BM-OU模型检测可能产生过多的稳定选择类型; 对于生殖隔离基因或适应性基因及其调控因子基因, 基因渐渗的可能性低, 这些基因的表达仍维持选择过程产生的种间基因表达进化。

4 交配系统对基因表达的调控

已知交配系统演化对植物生态与进化以及物种形成等方面均有重要影响, 在植物遗传育种、资源保护和遗传多样性等研究中具有重要作用。前文讨论了4种基本动力对种内基因表达变异和种间基因表达进化的影响以及相关检测方法, 交配系统可与任一进化动力互作(Muyle and Marais, 2016; 胡新生等, 2021), 间接对种内和种间遗传变异、分子进化、基因组大小和结构等进行调控(李玲玲等, 2023)。以下简要从理论上探讨自交与不同进化过程互作对基因表达在种内变异和种间进化的潜在调控模式。

4.1 自交与漂变联合作用

相对于异交物种, 自交降低了有效群体大小(Caballero and Hill, 1992)。 $N_e = N / (1 + F_{is})$, N 为实际群体大小, F_{is} 为近交系数。对于每个调控因子基因, 亚群体内基因加性方差减小, 但亚群体间基因的加性方差进

一步加大, 多调控因子基因的累积加性方差近似为:

$$\sigma_A^2 = 2(1+F_{is}) \sum_{i=1}^n p_i \alpha_i^2 \quad (9)$$

p_i 为第*i*个基因的基因频率, α_i 为加性效应。调控因子基因加性方差减小将导致亚群体内基因表达方差变小, 亚群体间基因表达方差增大。对于多个调控因子基因, 有效群体规模变小, 位点间连锁不平衡(LD)加快减弱, 位点间上位性效应变小, 调控因子基因间的互作网络可能变弱。相反, 种间调控因子基因变异加大, 种间基因表达进化加快。

4.2 自交与突变联合作用

自交降低了调控因子基因的点突变率以及由重组产生的突变率, 相关案例见李玲玲等(2023)的研究。因此, 减小了单位时间内调控因子基因的突变方差(σ_m^2)输入, 种群内的累积加性方差变小(漂变与突变平衡时, 加性方差 $\sigma_A^2 = 2N_e \sigma_m^2$), 种内基因表达方差变异减小。同样, 种间基因表达的加性方差减小, 种间基因表达进化的速率变慢。

4.3 自交与选择联合作用

自交与选择互作比较复杂, 在配子阶段和合子阶段表达的基因具有不同的分子进化速率(Li et al., 2023; Xiao et al., 2024), 这些基因表达在种内变异和种间进化也不同。开花植物通常有配子和合子世代交替现象, 自交与两阶段有不同的互作(Hu, 2015)。有7%–11%的基因只在配子体阶段表达(Arunkumar et al., 2013), 自交可以降低自然选择对配子阶段表达基因或调控因子基因的选择功效, 无论是定向选择还是稳定选择, 功效都会下降, 从而减缓它们的进化速率, 减弱调控因子的限制, 导致种内基因表达变异加大。由于选择功效降低, 种群间直系同源基因表达进化速率亦减缓。

目前的研究主要集中在花粉的基因表达谱上, 因为雌性配子体分离比较难(Beaudry et al., 2020)。由于在配子体阶段, 等位基因直接暴露在选择中, 更利于清除配子中的有害突变基因, 提高自然选择效率(Immler, 2019), 但自交降低了选择功效。例如, 柳兰(*Chamerion angustifolium*)自交效应在四倍体花粉中显著, 在二倍体花粉中不显著(Husband, 2016)。通过

分析拟南芥的基因组数据, 发现在以自交为主的群体中, 随着自交率的增加, 遗传多样性减少, 其配子细胞表达基因的纯化选择明显减弱(Gutiérrez-Valencia et al., 2022)。

自交间接地与合子阶段选择互作, 增加了基因和调控因子基因的纯合子频率, 减少杂合子频率。当处于定向选择模式下, 自交加快了有利等位基因固定, 从而减小了种内基因表达方差; 当处于稳定选择模式下, 自交降低了调控因子基因的杂合子频率和遗传变异, 减弱了对基因表达的限制, 增加了种内基因表达的方差。而种间基因表达进化呈相反的模式: 一方面, 自交使得歧化选择在种间差异加大, 促进了基因表达进化; 另一方面, 自交使得稳定选择功效减弱, 从而阻碍了种间基因表达进化。

当基因在配子和合子阶段都表达时(一因多效基因), 如植物中超过60%的基因在两阶段都表达(Arunkumar et al., 2013), 相应地, 调控因子基因在两阶段也可能都表达或部分表达。当基因在两阶段经受拮抗选择(两阶段呈相反的定向选择)模式时, 两阶段选择效应有一定程度的相互抵消, 自交减弱了配子选择而强化合子选择, 或反之, 其结果最终取决于两者的相对强弱(Li et al., 2023)。与异交植物相比, 自交使得基因表达受到更多限制, 种内基因表达变异减小, 这也使得种间基因表达进化增大。相反, 当基因在两阶段经受协同选择(为相同的定向选择或稳定选择)模式时, 两阶段选择效应积累, 自交减弱了配子阶段对有害突变的清除功效, 同时也弱化了合子阶段有害突变的选择, 最终使得基因表达比异交物种受到更小的限制, 这样种内基因表达变异增大, 与异交物种相比, 自交使得种间基因表达进化变小。

4.4 自交与迁移联合作用

与种内亚群体间基因流互作, 自交拒绝部分外源花粉流入接受群体, 依据Hu和Ennos (1999), 迁移后基因频率 p' 可表示为:

$$p' = \left(1 - m_s - \frac{1-\alpha}{2} m_p\right) p + \left(m_s + \frac{1-\alpha}{2} m_p\right) Q \quad (10)$$

式中, Q 为迁移基因频率, m_s 和 m_p 依次为种子和花粉迁移率, α 为自交率。花粉流直接影响花粉基因表达, 而种子迁移则影响全基因组基因表达, 由于植物子房不迁移, 因此不受外来因素影响。在自交条件

下,花粉基因表达可能不会受到强烈的选择压力(Peters and Weis, 2018),因为自交植物的花粉直接与自身的卵细胞结合,减少了与其它花粉竞争的机会。与异交植物相比,自交植物迁入率降低,基因频率趋于迁移基因频率(Q)速度慢,调控因子基因也有类似的模式。这样种群内基因表达方差趋于迁移群体的基因表达方差,比异交物种基因表达方差变化慢,但亚群体间基因表达方差比异交物种大。自交与种间基因流(基因渐渗)互作,减少基因及其调控因子基因在种间的交换,加快了种间基因表达进化。

以上是自交与单一进化过程互作的理论预测,自交可与多个进化过程互作,共同参与基因表达调控(李玲玲等, 2023),但自交与不同过程互作效应的相对大小因物种而异。实际研究时可通过比较自交与异交物种间基因表达差异来分析交配系统作用,动物中已有此方面的报道(Renn et al., 2018; Meléndez-Rosa et al., 2019),而在植物中报道较少。Frazee等(2021)利用锦龙花属(*Collinsia*)的姐妹种证实了自交和异交种花性状基因表达差异。Zhang等(2022)比较了异交物种*Capsella grandiflora*(二倍体)、自交物种*Capsella orientalis*(二倍体)和*Capsella bursa-pastoris*(四倍体)的花、叶片及根系基因表达模式,证实交配系统从异交到自交演化会影响3个组织的基因表达,花中更明显,且自交物种花性状基因表达呈趋同选择适应性,叶片和根系基因表达的变化可能并非直接与繁殖特征相关,而是由自交带来的有效种群减小和重组率降低等导致。该案例表明自交与趋同选择互作影响花性状基因的表达进化,而自交与漂变互作效应为影响叶片和根系基因表达进化的主要机制,今后可利用不同交配系统物种研究种间基因表达进化机制,为研究物种形成机制提供了重要视角。

5 结论

功能基因表达是连接基因编码信息与蛋白质产物的基本生命过程,调控因子通过顺式和反式作用于靶基因表达,进一步决定植物表型。本文系统综述了基因表达调控与进化机制研究进展,结论概括如下。

(1) 调控因子是控制基因表达的核心部分,涉及转录过程的转录因子、转录后非编码sRNA(miRNA和siRNAs)、翻译过程调控以及翻译后水平调控等复

杂的分子间相互作用。这些分子间的相互作用形成不同的调控机制,构成了复杂的调控网络,对植物生长发育、代谢和环境响应与适应等起重要作用。至今,植物基因表达在调控因子识别和分子机制研究方面已取得长足进展。

(2) 目前仍以mRNA拷贝数量来度量基因表达水平,已提出了多种基因表达的计算方法,包括 $2^{-\Delta\Delta CT}$ 和Pfaffl等度量方法以及数据标准化方法(RPKM、FPKM和TPM)。利用这些度量方法分析植物生态与进化问题是生物学研究的热点之一,但是如何准确地利用基因表达水平变异定位隐含的调控因子基因仍是亟待解决的关键问题,目前利用基因表达数据进行GWAS,进而估计调控因子基因eQTL仍受许多因素限制,包括样本容量、eQTL与基因组SNP连锁不平衡以及植物特定细胞或组织混合基因表达等。

(3) 基因表达与一般数量性状具有类似的遗传特征,理论上可以用基因加性、显性及上位性等数量遗传学理论来解析调控因子基因的作用模式,利用同质园等试验检测种内基因表达的选择适应性,也可依据基因表达性状变异特征,在不同进化模型假设下,如种内基因表达变异-种间基因表达进化比较,基于系统发育树的漂变(BM模型)-稳定选择或定向选择(OU模型)等进行统计检验,识别种内基因表达变异和种间基因表达进化的机制。

(4) 交配系统作为一种重要的生活史性状可以调控基因表达,自交通过降低有效群体大小、突变率及外源花粉竞争等影响种内和种间基因表达变异,通过影响配子和合子阶段自然选择功效,调控相应阶段的基因表达,改变种内基因表达变异和种间基因表达进化。实际研究时,采用不同交配系统物种分析特定组织基因表达差异,可揭示交配系统在调控基因表达进化机制中的作用。

作者贡献声明

王子韵: 文献检索、归纳和撰写论文; 吕燕文、肖钰和吴超: 参与文献收集; 胡新生: 修改和审阅论文。

参考文献

Andrew SC, Simonsen AK, Coppin CW, Arnold PA, Briceño VF, McLay TGB, Jackson CJ, Gallagher RV,

- Mokany K** (2024). Expression-environment associations in transcriptomic heat stress responses for a global plant lineage. *Mol Ecol* **33**, e17473.
- Arunkumar R, Josephs EB, Williamson RJ, Wright SI** (2013). Pollen-specific, but not sperm-specific, genes show stronger purifying selection and higher rates of positive selection than sporophytic genes in *Capsella grandiflora*. *Mol Biol Evol* **30**, 2475–2486.
- Barton NH, Briggs DEG, Eisen JA, Goldstein DB, Patel NH** (2007). Evolution. Cold Spring Harbor: Cold Spring Harbor Laboratory Press. pp. 49–53.
- Beaudry FEG, Rifkin JL, Barrett SCH, Wright SI** (2020). Evolutionary genomics of plant gametophytic selection. *Plant Commun* **1**, 100115.
- Beaulieu JM, Jhwueng DC, Boettiger C, O'Meara BC** (2012). Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution* **66**, 2369–2383.
- Bedford T, Hartl DL** (2009). Optimization of gene expression by natural selection. *Proc Natl Acad Sci USA* **106**, 1133–1138.
- Bernabè G, Dal Pra M, Ronca V, Pauletto A, Marzaro G, Saluzzo F, Stefani A, Artusi I, De Filippis V, Ferlin MG, Brun P, Castagliuolo I** (2021). A novel aza-derivative inhibits *agr* quorum sensing signaling and synergizes methicillin-resistant *Staphylococcus aureus* to clindamycin. *Front Microbiol* **12**, 610859.
- Bonnet E, Van de Peer Y, Rouzé P** (2006). The small RNA world of plants. *New Phytol* **171**, 451–468.
- Brümmer A, Bergmann S** (2024). Disentangling genetic effects on transcriptional and post-transcriptional gene regulation through integrating exon and intron expression QTLs. *Nat Commun* **15**, 3786.
- Bustin SA** (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* **25**, 169–193.
- Butler MA, King AA** (2004). Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am Nat* **164**, 683–695.
- Caballero A, Hill WG** (1992). Effects of partial inbreeding on fixation rates and variation of mutant genes. *Genetics* **131**, 493–507.
- Cai WT, Zhang YP, Chang TP, Wang ZZ, Zhu B, Chen Y, Gao X, Xu LY, Zhang LP, Gao HJ, Song JZ, Li JY** (2023). The eQTL colocalization and transcriptome-wide association study identify potentially causal genes responsible for economic traits in Simmental beef cattle. *J Anim Sci Biotechnol* **14**, 78.
- Chen LG, Zhang LP, Xiang SY, Chen YL, Zhang HY, Yu DQ** (2021). The transcription factor WRKY75 positively regulates jasmonate-mediated plant defense to necrotrophic fungal pathogens. *J Exp Bot* **72**, 1473–1489.
- Chen LL, Feng SS, Fan ZS, Gong C, Liu BY, Liu ZH, Li CW, Song EW, Sun SH, Wu GZ, Wu H, Wu M, Xu G, Yuan JX, Zeng CY, Zhu YM** (2019). Progress in non-coding RNA research. *Sci Sin Vitae* **49**, 1573–1605. (in Chinese)
- 陈玲玲, 冯珊珊, 范祖森, 龚畅, 刘本宇, 刘子豪, 李传伟, 宋尔卫, 孙树汉, 吴庚泽, 吴煌, 吴缅, 许光, 袁继行, 曾春雨, 朱友明 (2019). 非编码RNA研究进展. 中国科学: 生命科学 **49**, 1573–1605.
- Chen SS, Xie MH, Cui MK, Li WK, Xu ZG, Jia CX, Yang GY** (2022). Identification of *Broussonetia papyrifera* transcription factor BpbZIP1 and analysis of its response to cadmium stress. *Bull Bot Res* **42**, 394–402. (in Chinese)
- 陈思思, 谢牧洪, 崔茂凯, 李文凯, 徐正刚, 贾彩霞, 杨桂燕 (2022). 构树转录因子BpbZIP1的鉴定及镉胁迫响应分析. 植物研究 **42**, 394–402.
- Chen YX** (2020). Identification of WRKY Gene Family in *Phyllostachys edulis* and Their Correlation With Leaf Senescence Regulation. Master's thesis. Fuzhou: Fujian Agriculture and Forestry University. pp. 1–72. (in Chinese)
- 陈宇翔 (2020). 毛竹WRKY基因家族鉴定与叶片衰老调控的相关性分析. 硕士论文. 福州: 福建农林大学. pp. 1–72.
- Cheng X, Li LL, Xiao Y, Chen XY, Hu XS** (2020). Advances in the methods of detecting interspecific gene introgression and their applications. *Sci Sin Vitae* **50**, 1388–1404. (in Chinese)
- 程祥, 李玲玲, 肖钰, 陈晓阳, 胡新生 (2020). 种间基因渐渗检测方法及其应用研究进展. 中国科学: 生命科学 **50**, 1388–1404.
- Dai ZH, Wang YF** (2016). Genetics, 3rd edn. Beijing: Higher Education Press. pp. 338–346. (in Chinese)
- 戴灼华, 王亚馥 (2016). 遗传学(第3版). 北京: 高等教育出版社. pp. 338–346.
- Dror I, Zhou TY, Mandel-Gutfreund Y, Rohs R** (2014). Covariation between homeodomain transcription factors and the shape of their DNA binding sites. *Nucleic Acids Res* **42**, 430–441.
- Eppler E, Caelers A, Berishvili G, Reinecke MN** (2005). The advantage of absolute quantification in comparative hormone research as indicated by a newly established real-time RT-PCR: GH, IGF-I, and IGF-II gene expression in the tilapia, *Oreochromis niloticus*. *Ann NY Acad Sci* **1040**, 301–304.
- Evans C, Hardin J, Stoebel DM** (2018). Selecting between-sample RNA-Seq normalization methods from the

- perspective of their assumptions. *Brief Bioinform* **19**, 776–792.
- Farhangi S, Gòdia M, Derks MFL, Harlizius B, Dibbits B, González-Prendes R, Crooijmans RPMA, Madsen O, Groenen MAM** (2024). Expression genome-wide association study identifies key regulatory variants enriched with metabolic and immune functions in four porcine tissues. *BMC Genomics* **25**, 684.
- Feng K, Liu YX, Sun J, Zhao CL, Duan YJ, Wang WJ, Yan KJ, Yan XJ, Sun H, Hu YH, Han JH** (2021). Compound Danshen Dripping Pill inhibits doxorubicin or isotretinoin-induced cardiotoxicity. *Biomed Pharmacother* **138**, 111–531.
- Fleige S, Walf V, Huch S, Prgomet C, Sehm J, Pfaffl MW** (2006). Comparison of relative mRNA quantification models and the impact of RNA integrity in quantitative real-time RT-PCR. *Biotechnol Lett* **28**, 1601–1613.
- Frazer LJ, Rifkin J, Maheepala DC, Grant AG, Wright S, Kalisz S, Litt A, Spigler R** (2021). New genomic resources and comparative analyses reveal differences in floral gene expression in selfing and outcrossing *Collinsia* sister species. *G3-Genes Genom Genet* **11**, jkab177.
- Gachon C, Mingam A, Charrier B** (2004). Real-time PCR: what relevance to plant studies? *J Exp Bot* **55**, 1445–1454.
- Gilad Y, Oshlack A, Rifkin SA** (2006). Natural selection on gene expression. *Trends Genet* **22**, 456–461.
- Gómez-Soto D, Ramos-Sánchez JM, Alique D, Conde D, Triozzi PM, Perales M, Allona I** (2021). Overexpression of a *SOC1*-related gene promotes bud break in ecodormant poplars. *Front Plant Sci* **12**, 670497.
- Groen SC, Čalić I, Joly-Lopez Z, Platts AE, Choi JY, Natividad M, Dorph K, Mauck III WM, Bracken B, Cabral CLU, Kumar A, Torres RO, Satija R, Vergara G, Henry A, Franks SJ, Purugganan MD** (2020). The strength and pattern of natural selection on gene expression in rice. *Nature* **578**, 572–576.
- Gu X** (2010). *Statistical Theory and Methods for Evolutionary Genomics*. Oxford: Oxford University Press. pp. 95–112.
- Gutiérrez-Valencia J, Fracassetti M, Horvath R, Laenen B, Désamore A, Drouzas AD, Friberg M, Kolář F, Slotte T** (2022). Genomic signatures of sexual selection on pollen-expressed genes in *Arabidopsis thaliana*. *Mol Biol Evol* **39**, msab349.
- Hausser J, Syed AP, Bilen B, Zavolan M** (2013). Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation. *Genome Res* **23**, 604–615.
- Hill MS, Vande Zande P, Wittkopp PJ** (2021). Molecular and evolutionary processes generating variation in gene expression. *Nat Rev Genet* **22**, 203–215.
- Hill WG, Goddard ME, Visscher PM** (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet* **4**, e1000008.
- Hu WG** (2020). Preliminary Identification of Long Non-coding RNA in Wheat Powdery Mildew Resistance and Preliminary Analysis of its Regulatory Mechanism. PhD dissertation. Yangling: Northwest A&F University. pp. 1–146. (in Chinese)
- 胡卫国** (2020). 小麦白粉病抗性反应中链非编码RNA的鉴定及其调控机制的解析. 博士论文. 杨凌: 西北农林科技大学. pp. 1–146.
- Hu XS** (2015). Mating system as a barrier to gene flow. *Evolution* **69**, 1158–1177.
- Hu XS, Chen XY, Yeh FC** (2021). *Forest Population Genetics*. Beijing: China Forestry Publishing House. pp. 6–54. (in Chinese)
- 胡新生, 陈晓阳, Yeh FC** (2021). *林木群体遗传学*. 北京: 中国林业出版社. pp. 6–54.
- Hu XS, Ennos RA** (1999). Impacts of seed and pollen flow on population genetic structure for plant genomes with three contrasting modes of inheritance. *Genetics* **152**, 441–450.
- Huang JL, Chen H, Chen YH, Yang ZQ** (2024). Bioinformatics analysis and expression of *PmWRKY2* and *PmWRKY6* in *Pinus massoniana*. *Guihaia* <http://kns.cnki.net/kcms/detail/45.1134.Q.20240705.1521.002.html>. (in Chinese)
- 黄龙龙, 陈虎, 陈颖豪, 杨章旗** (2024). 马尾松 *PmWRKY2* 和 *PmWRKY6* 基因生物信息学及表达分析. 广西植物 <http://kns.cnki.net/kcms/detail/45.1134.Q.20240705.1521.002.html>.
- Huang SY, Yang TW, Zhang XJ, Tian SS, Gao MR, Li T, Shi Q, Zhang SW** (2024). Genome-wide identification and expression pattern analysis of WRKY genes family in *Erythralum scandens*. *Southwest China J Agr Sci* **37**, 1705–1714. (in Chinese)
- 黄诗宇, 杨天为, 张向军, 田姗姗, 高曼熔, 李婷, 石前, 张尚文** (2024). 赤苍藤 WRKY 基因家族的全基因组鉴定及表达模式分析. 西南农业学报 **37**, 1705–1714.
- Husband BC** (2016). Effect of inbreeding on pollen tube growth in diploid and tetraploid *Chamerion angustifolium*: do polyploids mask mutational load in pollen? *Am J Bot* **103**, 532–540.
- Immler S** (2019). Haploid selection in “diploid” organisms. *Annu Rev Ecol Evol S* **50**, 219–236.
- Irving TB, Chakraborty S, Maia LGS, Knaack S, Conde D, Schmidt HW, Triozzi PM, Simmons CH, Roy S, Kirst M,**

- Ané JM (2022). An LCO-responsive homolog of *NODULE INCEPTION* positively regulates lateral root formation in *Populus* sp. *Plant Physiol* **190**, 1699–1714.
- Jacob F, Monod J (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* **3**, 318–356.
- Khaitovich P, Weiss G, Lachmann M, Hellmann I, Enard W, Muetzel B, Wirkner U, Ansorge W, Pääbo S (2004). A neutral model of transcriptome evolution. *PLoS Biol* **2**, e132.
- Kimura M (1962). On the probability of fixation of mutant genes in a population. *Genetics* **47**, 713–719.
- Lande R (1976). Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**, 314–334.
- Lemos B, Meiklejohn CD, Cáceres M, Hartl DL (2005). Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. *Evolution* **59**, 126–137.
- Li J, Dai SY, Chen XJ, Liang XJ, Qu LZ, Jiang LY, Guo M, Zhou Z, Wei HD, Zhang HJ, Chen ZC, Chen L, Chen YH (2021). Mechanism of forkhead transcription factors binding to a novel palindromic DNA site. *Nucleic Acids Res* **49**, 3573–3583.
- Li JZ, Yang MK, Song XH, Wang Y, Wang YX, Wang SJ, Zhang MC, Deng X, Wang HY (2024). Genome-wide identification and expression analysis of the *SRS* gene family in *Panicum miliaceum*. *Pratacult Sci* **41**, 2035–2045. (in Chinese)
- 李佳祯, 杨梦珂, 宋晓涵, 王月, 王禹茜, 王世纪, 张美春, 邓欣, 王红艳 (2024). 黍子 *SRS* 基因家族的全基因组鉴定与表达分析. *草业科学* **41**, 2035–2045.
- Li LL, Wang X, Xiao Y, Cheng X, Chen XY, Hu XS (2023). On the theories of plant mating system and molecular evolution and their applications. *Sci Sin Vitae* **53**, 50–63. (in Chinese)
- 李玲玲, 王茜, 肖钰, 程祥, 陈晓阳, 胡新生 (2023). 植物交配系统与分子进化理论及其应用研究. *中国科学: 生命科学* **53**, 50–63.
- Li LL, Xiao Y, Wang X, He ZH, Lv YW, Hu XS (2023). The K_a/K_s and π_a/π_s ratios under different models of gametophytic and sporophytic selection. *Genome Biol Evol* **15**, evad151.
- Li X, Lei JB (2015). *Bioinformatics*, 2nd edn. Beijing: People's Medical Publishing House. pp. 236–257. (in Chinese)
- 李霞, 雷健波 (2015). *生物信息学(第2版)*. 北京: 人民卫生出版社. pp. 236–257.
- Liu C, Zhu XY, Zhang J, Shen M, Chen K, Fu XK, Ma L, Liu XL, Zhou C, Zhou DX, Wang GW (2022). eQTLs play critical roles in regulating gene expression and identifying key regulators in rice. *Plant Biotechnol J* **20**, 2357–2371.
- Liu JH (2023). Genome Wide Identification of WRKYs from *Caragana korshinskii* and Function Analysis of Four Group II Members. PhD dissertation. Hohhot: Inner Mongolia Agricultural University. pp. 1–115. (in Chinese)
- 柳金华 (2023). 柠条锦鸡儿WRKYs的全基因组鉴定及四个II组成员的功能探究. 博士论文. 呼和浩特: 内蒙古农业大学. pp. 1–115.
- Livak KJ, Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408.
- Luo J, Wu XY, Cheng Y, Chen G, Wang J, Song XJ (2023). Expression quantitative trait locus studies in the era of single-cell omics. *Front Genet* **14**, 1182579.
- Luo P, Di DW, Wu L, Yang JW, Lu YF, Shi WM (2022). MicroRNAs are involved in regulating plant development and stress response through fine-tuning of TIR1/AFB-dependent auxin signaling. *Int J Mol Sci* **23**, 510.
- Lv YW, Wang ZY, Xiao Y, He ZH, Wu C, Hu XS (2024). Advances in lineage sorting theories and their detection methods. *Biodivers Sci* **32**, 121–138. (in Chinese)
- 吕燕文, 王子韵, 肖钰, 何梓晗, 吴超, 胡新生 (2024). 谱系分选理论与检测方法的研究进展. *生物多样性* **32**, 121–138.
- Mai J, Lu MM, Gao QW, Zeng JY, Xiao JF (2023). Transcriptome-wide association studies: recent advances in methods, applications and available databases. *Commun Biol* **6**, 899.
- Mallet J (2005). Hybridization as an invasion of the genome. *Trends Ecol Evol* **20**, 229–237.
- Mallory AC, Vaucheret H (2006). Functions of microRNAs and related small RNAs in plants. *Nat Genet* **38** (Suppl), S31–S36.
- Mattick JS, Makunin IV (2006). Non-coding RNA. *Hum Mol Genet* **15**, R17–R29.
- Mcdonald JH, Kreitman M (1991). Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654.
- Meléndez-Rosa J, Bi K, Lacey EA (2019). Differential gene expression in relation to mating system in Peromyscine rodents. *Ecol Evol* **9**, 5975–5990.
- Meneksedag-Erol D, Tang T, Uludağ H (2015). Probing the effect of miRNA on siRNA-PEI polyplexes. *J Phys Chem B* **119**, 5475–5486.
- Meng LT, Su LW, Li XX, Xiong TS, Chang PP, Liu MZ, Zhou CG (2024). Functional analysis of ANT transcription factor of *Populus trichocarpa* based on CRISPR-dCas9 transcription activation system. *Bull Bot Res* **44**, 431–440.

- (in Chinese)
孟令桐, 苏丽伟, 李祥欣, 熊天圣, 常攀鹏, 刘孟卓, 周晨光 (2024). 基于CRISPR-dCas9转录激活系统的毛果杨ANT转录因子功能分析. *植物研究* **44**, 431–440.
- Mills JC, Taghert PH** (2012). Scaling factors: transcription factors regulating subcellular domains. *Bioessays* **34**, 10–16.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B** (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621–628.
- Mostafavi H, Spence JP, Naqvi S, Pritchard JK** (2023). Systematic differences in discovery of genetic effects on gene expression and complex traits. *Nat Genet* **55**, 1866–1875.
- Mu XY, Yu LJ, Cheng ZH, Fan GF, Wang N, Cai JQ, Jia YY, Liu HZ** (2024). Cloning and salt stress function analysis of the transcription factor *WRKY23* gene of *Populus simonii* × *P. nigra*. *Mol Plant Breeding* <http://kns.cnki.net/kcms/detail/46.1068.S.20240717.1719.014.html>. (in Chinese)
穆宣邑, 于莉婧, 程紫涵, 樊高峰, 王娜, 蔡静秋, 贾玉莹, 刘焕臻 (2024). 小黑杨转录因子*PsnWRKY23*基因的克隆及盐胁迫功能分析. *分子植物育种* <http://kns.cnki.net/kcms/detail/46.1068.S.20240717.1719.014.html>.
- Muyle A, Marais G** (2016). Genome evolution and mating systems. In: Kliman RM, ed. *Encyclopedia of Evolutionary Biology*. Oxford: Academic Press. pp. 480–492.
- Nica AC, Dermitzakis ET** (2013). Expression quantitative trait loci: present and future. *Phil Trans R Soc B* **368**, 20120362.
- Nolan T, Hands RE, Bustin SA** (2006). Quantification of mRNA using real-time RT-PCR. *Nat Protoc* **1**, 1559–1582.
- Oleksiak MF, Churchill GA, Crawford DL** (2002). Variation in gene expression within and among natural populations. *Nat Genet* **32**, 261–266.
- Palma-Silva C, Mortati AF, Chaves CJN, Simões Santos Leal B, Ribeiro RV, Pinheiro F, Ferro M, Riaño-Pachón DM, de Mattos JS, Tavares MM, Aecyo P, da Costa Cacossi T, Schöngart J, Piedade MTF, André T** (2024). Ecological transcriptomics reveals stress response pathways of a ground-herb species in a waterlogging gradient of Amazonian riparian forests. *Mol Ecol* **17**, e17437.
- Palos K, Nelson Dittrich AC, Yu LA, Brock JR, Railey CE, Wu HYL, Sokolowska E, Skiryicz A, Hsu PY, Gregory BD, Lyons E, Beilstein MA, Nelson ADL** (2022). Identification and functional annotation of long intergenic non-coding RNAs in Brassicaceae. *Plant Cell* **34**, 3233–3260.
- Pan Y** (2019). Cloning and Functional Identification of Transcription Factor *PeWRKY31* in *Populus × euramericana*. Master's thesis. Baoding: Hebei Agricultural University. pp. 1–43. (in Chinese)
潘钰 (2019). 欧美杨转录因子*PeWRKY31*的克隆及功能鉴定. 硕士学位论文. 保定: 河北农业大学. pp. 1–43.
- Parikh TP, Malik M, Britten J, Aly JM, Pilgrim J, Catharino WH** (2020). Steroid hormones and hormone antagonists regulate the neural marker neurotrimin in uterine leiomyoma. *Fertil Steril* **113**, 176–186.
- Peleke FF, Zumkeller SM, Gültas M, Schmitt A, Szymański J** (2024). Deep learning the *cis*-regulatory code for gene expression in selected model plants. *Nat Commun* **15**, 3488.
- Pennell MW, Eastman JM, Slater GJ, Brown JW, Uyeda JC, FitzJohn RG, Alfaro ME, Harmon LJ** (2014). geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* **30**, 2216–2218.
- Pertermann R, Tamilarasan S, Gursinsky T, Gambino G, Schuck J, Weinholdt C, Lilie H, Grosse I, Golbik RP, Pantaleo V, Behrens SE** (2018). A viral suppressor modulates the plant immune response early in infection by regulating microRNA activity. *mBio* **9**, e00419-18.
- Peters MAE, Weis AE** (2018). Selection for pollen competitive ability in mixed-mating systems. *Evolution* **72**, 2513–2536.
- Pfaffl MW** (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**, e45.
- Price PD, Palmer Drogue DH, Taylor JA, Kim DW, Place ES, Rogers TF, Mank JE, Cooney CR, Wright AE** (2022). Detecting signatures of selection on gene expression. *Nat Ecol Evol* **6**, 1035–1045.
- Qi YX, Wang L, Wang YY, Pu GB, Liu Q, Zhang YQ** (2019). Function and mechanism of WRKY transcription factors of plants under abiotic stress. *Mol Plant Breeding* **17**, 5973–5979. (in Chinese)
戚莹雪, 王蕾, 王尧尧, 蒲高斌, 刘谦, 张永清 (2019). 植物WRKY类转录因子在非生物胁迫下的功能与作用机制. *分子植物育种* **17**, 5973–5979.
- Ren YJ, Wang DJ, Su YC, Wang L, Zhang X, Su WH, Que YX** (2021). Structure, classification, evolution and function of plant WRKY transcription factors. *J Agr Biotechnol* **29**, 105–124. (in Chinese)
任永娟, 王东姣, 苏亚春, 王玲, 张旭, 苏炜华, 阙友雄 (2021). 植物WRKY转录因子: 结构、分类、进化和功能. *农业生物技术学报* **29**, 105–124.

- Renn SCP, Machado HE, Duftner N, Sessa AK, Harris RM, Hofmann HA (2018). Gene expression signatures of mating system evolution. *Genome* **61**, 287–297.
- Rifkin SA, Kim J, White KP (2003). Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat Genet* **33**, 138–144.
- Roberge C, Guderley H, Bernatchez L (2007). Genome-wide identification of genes under directional selection: gene transcription Q_{ST} scan in diverging Atlantic salmon subpopulations. *Genetics* **177**, 1011–1022.
- Rohlf RV, Nielsen R (2015). Phylogenetic ANOVA: the expression variance and evolution model for quantitative trait evolution. *Syst Biol* **64**, 695–708.
- Samad AFA, Sajad M, Nazaruddin N, Fauzi IA, Murad AMA, Zainal Z, Ismail I (2017). MicroRNA and transcription factor: key players in plant regulatory network. *Front Plant Sci* **8**, 565.
- Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011). Global quantification of mammalian gene expression control. *Nature* **473**, 337–342.
- Sezen UU, Shue JE, Worthy SJ, Davies SJ, McMahon SM, Swenson NG (2024). Leaf gene expression trajectories during the growing season are consistent between sites and years in American beech. *Proc R Soc B* **291**, 20232338.
- Shamandi N, Zytynicki M, Charbonnel C, Elvira-Matlot E, Bochnakian A, Comella P, Mallory AC, Lepère G, Sáez-Vásquez J, Vaucheret H (2015). Plants encode a general siRNA suppressor that is induced and suppressed by viruses. *PLoS Biol* **13**, e1002326.
- Shea DJ, Nishida N, Takada S, Itabashi E, Takahashi S, Akter A, Miyaji N, Osabe K, Mehraj H, Shimizu M, Seki M, Kakizaki T, Okazaki K, Dennis ES, Fujimoto R (2019). Long noncoding RNAs in *Brassica rapa* L. following vernalization. *Sci Rep* **9**, 9302.
- Shriram V, Kumar V, Devarumath RM, Khare TS, Wani SH (2016). MicroRNAs as potential targets for abiotic stress tolerance in plants. *Front Plant Sci* **7**, 817.
- Singh KP, Miaskowski C, Dhruva AA, Flowers E, Kober KM (2018). Mechanisms and measurement of changes in gene expression. *Biol Res Nurs* **20**, 369–382.
- Song L, Fang Y, Chen L, Wang J, Chen XW (2021). Role of non-coding RNAs in plant immunity. *Plant Commun* **2**, 100180.
- Stavast CJ, Erkeland SJ (2019). The non-canonical aspects of microRNAs: many roads to gene regulation. *Cells* **8**, 1465.
- Strader L, Weijers D, Wagner D (2022). Plant transcription factors—being in the right place with the right company. *Curr Opin Plant Biol* **65**, 102136.
- Su WJ, Cao RL, Zhou ZL, Zhao NH, Zhang YY, Hu DN, Liu J (2023). Identification and expression analysis of the WRKY gene family in *Camellia oleifera* under stress conditions. *J Central South Univ Forestry Technol* **43**(3), 155–166, 174. (in Chinese)
- 苏文娟, 曹瑞兰, 周增亮, 赵娜红, 张运煜, 胡冬南, 刘娟 (2023). 油茶WRKY基因家族鉴定及逆境胁迫表达分析. 中南林业科技大学学报 **43**(3), 155–166, 174.
- Tian F, Yang DC, Meng YQ, Jin JP, Gao G (2020). PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res* **48**, D1104–D1113.
- Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**, Research0034.1.
- Vanguilder HD, Vrana KE, Freeman WM (2008). Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques* **44**, 619–626.
- Vösa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B, Kirsten H, Saha A, Kreuzhuber R, Yazar S, Brugge H, Oelen R, De Vries DH, Van Der wijst MGP, Kasela S, Pervjakova N, Alves I, Favé MJ, Agbessi M, Christiansen MW, Jansen R, Seppälä I, Tong L, Teumer A, Schramm K, Hemani G, Verlouw J, Yaghootkar H, Sönmez Flitman R, Brown A, Kukushkina V, Kalnapienkis A, Rüeger S, Porcu E, Kronberg J, Kettunen J, Lee B, Zhang FT, Qi T, Hernandez JA, Arindrarto W, Beutner F, Dmitrieva J, Elansary M, Fairfax BP, Georges M, Heijmans BT, Hewitt AW, Kähönen M, Kim Y, Knight JC, Kovacs P, Krohn K, Li S, Loeffler M, Marigorta UM, Mei HL, Momozawa Y, Müller-Nurasyid M, Nauck M, Nivard MG, Penninx BWJH, Pritchard JK, Raitakari OT, Rotzschke O, Slagboom EP, Stehouwer CDA, Stumvoll M, Sullivan P, 't Hoen PAC, Thiery J, Tönjes A, Van Dongen J, Van Iterson M, Veldink JH, Völker U, Warmerdam R, Wijmenga C, Swertz M, Andiappan A, Montgomery GW, Ripatti S, Perola M, Kutalik Z, Dermizakis E, Bergmann S, Frayling T, Van Meurs J, Prokisch H, Ahsan H, Pierce BL, Lehtimäki T, Boomsma DI, Psaty BM, Gharib SA, Awadalla P, Milani L, Ouwehand WH, Downes K, Stegle O, Battle A, Visscher PM, Yang J, Scholz M, Powell J, Gibson G, Esko T, Franke L (2021). Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet* **53**, 1300–1310.

- Wagner GP, Kin K, Lynch VJ** (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* **131**, 281–285.
- Waheed S, Zeng LH** (2020). The critical role of miRNAs in regulation of flowering time and flower development. *Genes* **11**, 319.
- Walsh B, Lynch M** (2018). *Evolution and Selection of Quantitative Traits*. Oxford: Oxford University Press. pp. 391–478.
- Wang FF, Zhou ZX, Hong Y, Gu YY, Lv C, Guo BJ, Zhu J, Xu RG** (2023). Identification of the *NF-YC* genes in *Hordeum vulgare* and expression analysis under salt stress. *Chin Bull Bot* **58**, 140–149. (in Chinese)
王菲菲, 周振祥, 洪益, 谷洋洋, 吕超, 郭宝健, 朱娟, 许如根 (2023). 大麦*NF-YC*基因鉴定及在盐胁迫下的表达分析. *植物学报* **58**, 140–149.
- Wang JL, Hou YG, Wang Y, Zhao HS** (2021a). Integrative lncRNA landscape reveals lncRNA-coding gene networks in the secondary cell wall biosynthesis pathway of moso bamboo (*Phyllostachys edulis*). *BMC Genomics* **22**, 638.
- Wang T, Liu YZ, Ruan JP, Dong XJ, Wang YD, Peng JJ** (2021b). A pipeline for RNA-seq based eQTL analysis with automated quality control procedures. *BMC Bioinformatics* **22**, 403.
- Wang X** (2023). *Phylogenomic Relationships Among Varieties of *Tonna ciliata* and the Strategy for Their Genetic Conservation*. PhD dissertation. Guangzhou: South China Agricultural University. pp. 1–173. (in Chinese)
王茜 (2023). 红椿变种基因组进化关系和遗传资源保护研究. 博士学位论文. 广州: 华南农业大学. pp. 1–173.
- Wang Z** (2014). *The Structural Basis for the Specific DNA Recognition of Transcriptional Factor SATB1*. PhD dissertation. Tianjin: NanKai University. pp. 1–126. (in Chinese)
王峥 (2014). 转录因子SATB1特异性结合DNA的结构基础研究. 博士学位论文. 天津: 南开大学. pp. 1–126.
- Wright S** (1943). Isolation by distance. *Genetics* **28**, 114–138.
- Wu JN, Del Duca E, Espino M, Gontzes A, Cueto I, Zhang N, Estrada YD, Pavel AB, Krueger JG, Guttman-Yassky E** (2020). RNA sequencing keloid transcriptome associates keloids with Th2, Th1, Th17/Th22, and JAK3-skewing. *Front Immunol* **11**, 597741.
- Wu SH, Zhang SX, Chao JQ, Li Y, Yang SG, Deng XM, Shi MJ, Tian WM** (2023). Differential miRNA expression profiling reveals a correlation between hbr-miR156 and laticifer differentiation in rubber trees. *Ind Crop Prod* **192**, 116067.
- Wu YQ, Sun RN, Huan T, Zhao YY, Yu DL, Sun YQ** (2024). An insight into the gene expression evolution in *Gossypium* species based on the leaf transcriptomes. *BMC Genomics* **25**, 179.
- Xiao Y, Lv YW, Wang ZY, Wu C, He ZH, Hu XS** (2024). Selfing shapes fixation of a mutant allele under flux equilibrium. *Genome Biol Evol* **16**, evae261.
- Yang JK** (2023). *Studies on the Stress Resistance of Transcription Factors PdPapERF34 and PdPapWRKY22 of *Populus davidiana* × *P. alba* var. *pyramidalis**. Master's thesis. Harbin: Northeast Forestry University. pp. 1–95. (in Chinese)
杨建坤 (2023). 山新杨转录因子PdPapERF34和PdPapWRKY22抗逆功能研究. 硕士学位论文. 哈尔滨: 东北林业大学. pp. 1–95.
- Yang XQ, Zhang LC, Yang YZ, Schmid M, Wang YW** (2021). miRNA mediated regulation and interaction between plants and pathogens. *Int J Mol Sci* **22**, 2913.
- Yu XY** (2021). *Cloning and Functional Identification of *PeWRKY41* Gene from *Populus* × *euramericana**. PhD dissertation. Baoding: Hebei Agricultural University. pp. 84–90. (in Chinese)
于晓跃 (2021). 欧美杨*PeWRKY41*基因克隆及功能鉴定. 博士学位论文. 保定: 河北农业大学. pp. 84–90.
- Zhang L, Lu DY, Ge XL, Du JJ, Wen SS, Xiang XD, Du CJ, Zhou XL, Hu JJ** (2023). Insight into growth and wood properties based on QTL and eQTL mapping in *Populus deltoides* 'Danhong' × *Populus simonii* 'Tongliao1'. *Ind Crop Prod* **199**, 116731.
- Zhang XP, Wang W, Zhu WD, Dong J, Cheng YY, Yin ZJ, Shen FF** (2019). Mechanisms and functions of long non-coding RNAs at multiple regulatory levels. *Int J Mol Sci* **20**, 5573.
- Zhang XX, Cheng X, Li LL, Wang X, Zhou W, Chen XY, Hu XS** (2020). The wave of gene advance under diverse systems of mating. *Heredity* **125**, 253–268.
- Zhang YT** (2022). *Functional Study on Drought and Salt-alkali Tolerance of *WRKY42* Gene in *Amorpha fruticosa**. Master's thesis. Harbin: Northeast Forestry University. pp. 27–34. (in Chinese)
张艺腾 (2022). 紫穗槐*WRKY42*基因抗旱耐盐碱性的功能研究. 硕士学位论文. 哈尔滨: 东北林业大学. pp. 27–34.
- Zhang YX, Han XJ, Chen SS, Zheng L, He XL, Liu MY, Qiao GR, Wang Y, Zhuo RY** (2017). Selection of suitable reference genes for quantitative real-time PCR gene expression analysis in *Salix matsudana* under different abiotic stresses. *Sci Rep* **7**, 40290.
- Zhang ZB, Kryvokhyzha D, Orsucci M, Glémin S, Milesi**

- P, Lascoux M** (2022). How broad is the selfing syndrome? Insights from convergent evolution of gene expression across species and tissues in the *Capsella* genus. *New Phytol* **236**, 2344–2357.
- Zhao HM** (2022). Molecular Mechanism of Long Noncoding RNA lncRNAW20 Regulating Drought Tolerance in *Betula platyphylla*. PhD dissertation. Harbin: Northeast Forestry

University. pp. 1–146. (in Chinese)

- 赵慧敏** (2022). 长链非编码RNA lncRNAW20调控白桦抗旱的分子机制. 博士论文. 哈尔滨: 东北林业大学. pp. 1–146.
- Zheng GZ, Li W, Liu ZY** (2019). Alternative role of noncoding RNAs: coding and noncoding properties. *J Zhejiang Univ Sci B* **20**, 920–927.

Advances in the Regulation and Evolutionary Mechanisms of Plant Gene Expression

Ziyun Wang^{1,2}, Yanwen Lü^{1,2}, Yu Xiao^{1,2}, Chao Wu^{1,2}, Xinsheng Hu^{1,2*}

¹College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou 510642, China

²Guangdong Key Laboratory for Innovative Development and Utilization of Forest Plant Germplasm, Guangzhou 510642, China

Abstract Functional gene expression is a basic life process that connects the coding information of a gene to protein products. The level of gene expression is considered as a quantitative trait between genotype and phenotype and plays an important role in response to climatic and environmental changes. First, we systematically summarize regulatory elements of gene expression in plant species and empirical evidence, including the effects of transcription factors and small RNAs on gene expression regulation. Second, this review discusses the eQTL mapping for regulatory elements of gene expression through gene expression-based genome-wide association study (GWAS) and the limitations of this method. This review analyzes the intraspecific variation in gene expression in theory under the processes of mutation, drift and selection and the testing methods. This review also analyzes the interspecific evolution of gene expression under the mutation and drift processes or under the phylogeny-based drift-selection processes and the testing methods. Finally, this review discusses the regulation of gene expression by the plant mating system. Selfing reduces the effective population size, mutation rate, recombination rate and competition from exogenous pollen, and changes the efficacy of natural selection in the gametophytic and sporophytic phases. Selfing regulates intraspecific gene expression variation and interspecific gene expression evolution. This review comprehensively comments on theoretical and practical research progress and existing questions, which aids in our deep understanding of plant gene expression regulation and evolution mechanisms.

Key words gene expression, regulatory element, natural selection, quantitative trait, mating system

Wang ZY, Lü YW, Xiao Y, Wu C, Hu XS (2025). Advances in the regulation and evolutionary mechanisms of plant gene expression. *Chin Bull Bot* **60**, 621–639.

* Author for correspondence. E-mail: xinsheng@scau.edu.cn

(责任编辑: 朱亚娜)

通讯作者/团队简介

胡新生, 男, 双博士学位, 华南农业大学教授, 博士生导师。团队研究交配系统在物种形成、分子进化和基因表达等过程中作用的理论与应用。以红椿和团花为材料, 利用基因组序列、多组学数据以及数量性状等, 研究森林植物遗传与进化、林木育种和遗传资源保护等问题。目前主持国家自然科学基金、国家十四五重点研发计划和华南农业大学人才引进等项目。