

转座子衍生基因的演变途径及对生物生长发育的影响

刘珂珂 汤定钦 周明兵*

(浙江农林大学林业与生物技术学院, 亚热带森林培育国家重点实验室, 杭州 311300)

摘要 转座子(transposable elements, TEs)是指在基因组上能从同一条染色体的一个位置转移到另一个位置或者从一条染色体转移到另一条染色体上的一段DNA序列。广泛存在于基因组中的转座子通过复制、动员、重组基因片段以及修改原基因结构形成的新基因,被称为转座子衍生基因。该文综述了转座子衍生基因与转座子和常规基因的异同以及转座子衍生基因的演变途径,归纳了转座子衍生基因对宿主基因进化,以及对生物生长发育的影响。

关键词 转座子; 转座子衍生基因; 基因演化

The Evolution Pathway of Transposon-Derived Genes and Its Effect on Biological Growth and Development

LIU Keke, TANG Dingqin, ZHOU Mingbing*

(State Key Laboratory of Subtropical Silviculture, College of Forestry and Biotechnology, Zhejiang A&F University, Hangzhou 311300, China)

Abstract Transposon is a segment of DNA sequence that can be translocated in one chromosome or from one chromosome to another chromosome in the host genome. Transposons can form new genes through duplication, mobilization, reorganization of gene fragments, and modification of the original gene structure, Which are named transposon-derived genes. This article reviews the similarities and differences between transposon derived genes, transposon and conventional genes. The evolution pathways of transposon derived genes is also concluded. It also summarizes the effects of transposon derived genes on host gene evolution, as well as the effects on the growth and development.

Keywords transposon; transposon derived gene; gene evolution

转座子是1951年由美国遗传学家MCCLINTOCK^[1]在玉米(*Zea mays* Linn.)中观察到异常的颜色模式后发现的,此后在高等动植物、细菌以及酵母中,越来越多的转座子被鉴定发现。转座子在真核生物基因组中所占比例很高,如占香蕉(*Musa nana* Lour.)基因组的50.0%^[2]、果蝇(*Drosophila melanogaster*)的22%^[3]、水稻(*Oryza sativa* Linn.)的39.5%^[4]、玉米的

84.2%^[5]、马铃薯(*Solanum tuberosum* Linn.)的62.0%^[6]、毛竹(*Phyllostachys heterocycla*)的59.0%^[7],是真核生物基因组的重要组成部分^[8]。

根据转座机制可以把转座子分为两类。第I类元件即逆转录转座子,遵循复制-粘贴机制,通过RNA中间体来实现转座,每发生一次转座,即可以增加一个拷贝,导致在真核基因组中形成很大一部

收稿日期: 2020-07-14 接受日期: 2020-10-28

浙江自然科学基金重点项目(批准号: LZ19C160001)和国家自然科学基金(批准号: 31870656、31470615)资助的课题

*通讯作者。Tel: 13588152716, E-mail: zhoumingbing@zafu.edu.cn

Received: July 14, 2020 Accepted: October 28, 2020

This work was supported by Zhejiang Natural Science Foundation (Grant No.LZ19C160001) and the National Natural Science Foundation of China (Grant No.31870656, 31470615)

*Corresponding author. Tel: +86-13588152716, E-mail: zhoumingbing@zafu.edu.cn

URL: <http://www.cjcb.org/arts.asp?id=5438>

分重复序列。逆转录转座子结构比较复杂, 包括长末端重复序列(long terminal repeats, LTRs)、长散在核元件(long interspersed nuclear elements, LINEs)和短散在核元件(short interspersed nuclear elements, SINEs)。LINEs和SINEs被称为非LTR逆转座子, 在植物基因组中数量较多。LTR逆转录转座子主要包括Ty1-copia、Ty3-gypsy和LTR-ERV反转录病毒类转座子^[9]。而第II类转座子即DNA转座子以“剪切-粘贴”的方式转座, 它含有末端反向重复(terminal inverted repeat, TIR)和编码转座酶的基因, 结构较简单^[10]。又根据转座时切除的DNA链的条数分为亚类I和亚类II。亚类I的DNA类转座子转座时转座酶识别TIR并切断末端的两条链, 根据其TIR内的序列基序和插入时在转座子两端两侧产生的靶位点重复(target site duplications, TSD)的长度来可将其分为Tc1-Mariner、hAT、Mutator、Merlin、Transib、P、PiggyBac、PIF-Harbinger和CACTA等9个超家族^[11]。Helitron和Maverick转座子家族属于DNA转座子的第二个亚类, 因为它们的转座过程需要复制, 不会导致双链DNA断裂^[12]。DNA转座子和逆转录转座子又都作为非自主和自主转座子存在。自主转座子编码移动所需的所有酶, 而非自主转座子没有编码转座酶的能力, 因此它们的移动性完全依赖于其他转座酶。转座子在不断插入、转座过程中, 一些转座子缺失部分片段而失去转座能力, 稳定地保留在基因组内, 经过分子驯化, 进化为基因的一部分, 参与基因表达, 这些基因被称为转座子衍生基因。现已报道的转座子衍生基因的结构具有多样性, 但都含有部分的转座子保守序列和转座酶DNA结合域。这些转座子衍生基因在脊椎动物适应性免疫^[13]、植物的系统发育^[14-15]、细胞周期与进程的调控^[16]过程中都发挥着关键作用。已有部分转座子衍生基因的功能已被验证, 在这里, 我们重点综述了转座子衍生基因与转座子和常规基因的异同, 归纳了转座子衍生基因的演变途径, 以及对生物生长发育的影响。

1 转座子衍生基因与转座子和常规基因的异同

转座子通过转座插入到宿主基因中, 形成嵌合基因, 在某些情况下可以产生功能蛋白。大多数转座子可以通过剪接募集到编码区的不同外显子, 它和宿主基因组已经进化出一种互利的关系, 平衡了

转座子的生存和宿主的进化利益。转座子最显著的贡献是一种被称为“分子驯化”的进化过程, 通过这种进化过程, 转座子衍生的编码序列产生了一个功能性宿主基因。因此, 驯化的基因代表了基因组中稳定的功能成分^[17]。这种转座子衍生的基因首先在果蝇中被鉴定为驯化的P元件^[18], 并进一步扩展到植物和动物基因组, 包括人类基因组^[19]。与转座子序列的重复性不同, 转座子衍生基因在基因组中以单拷贝的形式存在, 并且在远亲物种中可以检测到同源基因。在结构上, 这些基因缺乏转座的分子特征, 如侧翼TIR和TSD, 并且许多转座子衍生基因只保留了祖先转座子编码蛋白的DNA结合域(DNA-binding domain, DBD), 具有结构多样性。它们在体内扮演着重要的生物学角色, 但总的来说, 它们已经失去了介导转座的能力。

大多数转座子衍生基因都是在正向遗传的筛选中偶然被发现的, 很少有系统的研究报告。迄今的研究结果表明: 转座子衍生基因的进化功能更类似于常规基因(表1)。首先, 转座子与转座子衍生基因进化不同: 功能性的转座子是移动的, 而已知的转座子衍生基因是固定的。转座子衍生基因在基因组的位置稳定, 其中与移动性相关的侧翼序列很大程度上不再受选择的影响, 逐渐分解; 并且因为转座子衍生基因不再转座, 所以它的拷贝数减少。其次, 转座子和转座子衍生基因的功能也不同, 在大多数组织和条件下很多转座子是沉默的, 而大多数常规基因能正常表达。在植物中, 转座子沉默的主要途径是RNA介导的DNA甲基化, 其中小干扰RNAs(small interfering RNAs, siRNAs)指导抑制性表观遗传修饰^[20]; 而与转座子相比, 转座子衍生基因往往产生更高水平的mRNAs, 并被较低水平的siRNAs作为靶标^[21-23]。再次, 转座子衍生基因和常规基因一样具有微同线性(micro-syteny), 即转座子衍生基因进化过程中受到表型选择的影响, 其基因序列的数量和位置具有保守性。而特定同源基因组位置的单个转座子是不保守的, 在长时间的进化过程中, 不保持开放阅读框(open reading frame, ORF), 不具有微同线性。因此, 可以根据转座子衍生基因的这一特征来区分转座子和转座子衍生基因。然而转座子衍生基因和常规基因一样, 都会受到染色体重排的影响, 这可能会侵蚀微同线性^[24-26]。

基于以上转座子衍生基因的特性, 使用单独的

表1 转座子衍生基因与转座子和常规基因的比较(根据参考文献[21-26]修改)

Table 1 Comparison of transposon-derived genes with transposon and conventional genes (modified from references [21-26])

类型 Types	特征 Characteristics			
	位置 Position	结构 Structure	功能 Function	保守性 Conservation
Transposon derived gene	Its genomic position stabilizes	Transposase coding sequence without flanking TIR and TSD	Normal expression	Conservative
Transposon	Its genomic position unstabilizes	Transposase coding sequence with flanking TIR and TSD	Most TEs (in most tissues and conditions) are silenced	Not conservative
Conventional gene	Position is fixed	No structural characteristics related to transposons	Normal expression	Conservative

遗传属性来识别转座子衍生基因可能是局限的, 可以通过结合多个遗传属性的方法提高识别转座子衍生基因的精确度。HOEN等^[24]开发了一个全面系统地发现转座子衍生基因的方法: 使用转座子特异的保守域高精度地识别转座子衍生的基因, 然后, 将这些基因与已知转座子、转座子衍生基因和常规基因进行遗传属性的比较, 包括重复性、微同线性以及表达和抑制小RNA的水平。在整合了所有属性之后, 根据基因与参考基因的相似性或差异, 对基因进行分类。利用这种方法在模式植物拟南芥中发现了36个新的高信度转座子衍生基因。

2 转座子衍生基因演变途径

2.1 转座子插入

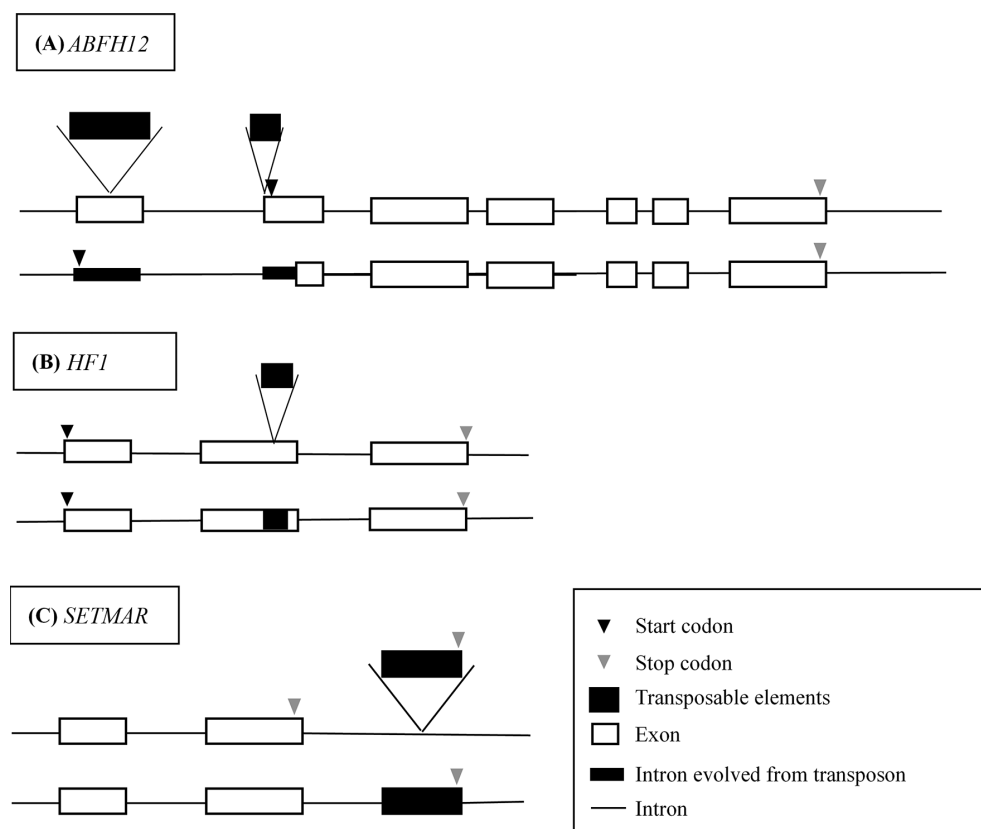
部分转座子是在宿主基因上插入转座子元件, 将自身整合到基因内部, 参与基因的表达。基因上游的TE插入和整合能够改变基因的表达模式; 转座子能够在外显子中形成新的蛋白结构域, 例如在黄瓜(*C. sativus*)基因组的*csgl3*基因的第四个外显子中插入了一个5 Kb的LTR反转录转座子, 导致*csgl3*功能丧失^[27]; 当将其插入到内含子中时, 可以导致从头蛋白的产生。在人类基因中有4%的CDS与转座子相关^[28], 在水稻中发现了439个(大于2%) CDS与TE相关。虽然这种新编码区的偶然招募似乎并不常见, 但在基因区插入转座子可能是在高等真核生物中创造新生物功能的进化推动力。有些转座子插入经过分子驯化被保留了下来, 水稻中, 转座子在调控基因表达和创造新基因方面发挥着重要作用^[29]。水稻基因组中含有丰富的转座子衍生基因, 例如, 在水稻基因组中, Pack-Mutator-like转座子(Pack-MULE)插入基因*abfh12*的5'端, 形成独立的、富含GC的转

录本。在该转录本中, 原始外显子1和部分外显子2被剪接, 成为新的内含子序列(图1A)。因此, 12号染色体上的Pack-MULE似乎插入了基因的上游区域, 从而改变了基因的剪接模式, 这种修改导致了一个新的转录本, 其中原来的外显子1被替换了^[30]。富含GC的Pack-MULE在基因5'端的存在可能会产生进一步的进化影响, 而不仅仅是改变基因的5'端。Pack-MULE的插入可能会改变局部重组率, 提高基因的表达水平^[31], 为DNA甲基化提供了更多靶标, 有助于调节基因表达^[32], 从而有助于表型多样性和对环境条件的适应。另外, 在矮牵牛(*Petunia hybrida* Vilm.)基因组内的*hfl*基因的第二内含子中发现有Psl(petunia spm-like)转座子插入^[33](图1B)。*setmar*也属于这种机制, 不过转座子插入的位置是在原有基因的下游^[34](图1C)。

2.2 形成嵌合基因

还有一部分的转座子衍生基因来源于原转座子的转座酶基因, 进化过程中源于原转座子的转座酶基因经过分子驯化过程进化成宿主功能基因。

Harbinger转座子是DNA转座子, 通常编码DDE转座酶和含有SANT/Myb/Trihelix结构域的DNA结合蛋白^[35]。据报道, 在哺乳动物、果蝇和拟南芥中已经驯化了Harbinger转座子^[36-37], 这表明了Harbinger转座子衍生基因在进化上的重要性。例如在拟南芥中鉴定出的一对由Harbinger转座子衍生的抗沉默因子HDP1和HDP2。HDP1是拟南芥中由Harbinger转座子衍生的宿主蛋白。HDP1蛋白有一个注释为“Harbinger转座酶衍生的核酸酶结构域”^[38], 表明它可能是Harbinger转座酶的“驯化”蛋白。使用Harbinger转座子数据库(Repbase)^[39]进行的BLASTP搜索显示, HDP1与一些Harbinger转座酶序列有大约



在图A、B、C中, 第一条代表原始基因, 第二条代表转座子衍生基因。

In figure A, B, C, the first is the original gene, the second is the transposon derived gene.

图1 插入型转座子衍生基因产生机制

Fig.1 Generating mechanism of transposon derived gene

30%的同源性。Harbinger转座酶属于“DDE”超家族核酸内切酶, 含有一个保守的催化三联体氨基酸序列“DDE”^[40]。然而, HDP1与选定的转座酶的比对表明, 三联体在HDP1中不是完全保守的, 这表明HDP1可能已经失去了核酸酶活性。值得注意的是, HDP1不是一个转座子残留物, 因为尽管它与Harbinger转座酶同源, 但它缺乏转座子的特征, 如TIR和TSD。系统发育分析表明, HDP1起源可能发生在被子植物出现之前^[41]。

HDP2是一种来源于Harbinger转座子的DNA结合蛋白。HDP2蛋白含有SANT/Myb/三螺旋DNA结合基序。有趣的是, 除了典型的转座酶, Harbinger转座子通常还有第二个ORF编码这个基序^[34]。HDP2和Harbinger转座子编码蛋白的SANT/Myb/三螺旋结构域的比对表明, 对DNA结合至关重要的三个色氨酸残基在拟南芥HDP2和Harbinger蛋白之间是保守的。系统发育分析表明, HDP2起源可能发生在被子植物分化之前。另外, 这两个基因的功能丧失突变能够触发转基因和一些内源性转座子的沉默, 而且

还能增加DNA甲基化。

2.3 分子驯化

DNA转座子携带着十分精细的酶催化机制, 它在几种情况下被宿主基因组通过被称为分子驯化的进化过程增选为宿主基因组的一部分。第一种情况是反复驯化(图2A)。P元件驯化是同一转座子来源的编码序列在不同果蝇谱系中发生的多次独立获取的独特例子。区分了两类不同的功能性P元件^[42]。第一类代表果蝇的典型P DNA转座子。典型的P转座子通过内含子3的组织特异性被选择性剪接, 在果蝇的生殖细胞中产生87 kDa的转座酶, 在果蝇的体细胞中产生66 kDa的蛋白阻遏因子^[43]。第二类包括不同固定形式的P元素, 包括由于失去TIR和转座酶最后一个外显子而变得不活动的新生细胞P *obscura*和P *montium* ^[44]。P *obscura*新基因最初被发现于暗果蝇种亚群(*Drosophila obscura subgroup*), 它产生G或A-Thap蛋白; P *montium*基因存在于山果蝇种亚群(*Drosophila montium subgroup*)中, 是单拷贝基因, 它包括产生65.9 kDa RL蛋白(repressor-like protein)的P

基因和获得一个未翻译外显子E-1和附加外显子E0的*P-BOC*新基因。*P-BOC*新生蛋白的选择性剪接产生两种蛋白, RL1和RL2。*D. vulkana*基因组中含有第三种形式的*P montium*新基因, 该类型增加了一个外显子E0', 位于外显子E0的上游。RL、RL1和RL2蛋白已经被证明在体内结合染色质, 并且不抑制*P*元件转座子的转座或转录^[45]。推测这些蛋白可能参与了许多不同常染色质区域的表达调控和或染色质结构的修饰。

另一种情况是转座酶的趋同驯化(图2B)。许多研究已经指出, CENP-B蛋白与几种POGO转座酶之间可能的进化关系, 包括来自黑腹果蝇的POGO转

座酶, 以及人的Tigger1和Tigger2转座酶。事实上, CENP-B的DBD和POGO转座酶DDE/D催化核心有惊人的序列相似性。此外, CENP-B的结合位点类似于Tigger2的TIR^[46]。不同的POGO转座酶来源独立地产生了哺乳动物谱系中的CENP-B蛋白, 以及存在于分裂酵母中的Ars结合蛋白(Ars-binding protein, Abp1)、CENP-B同源物1(CENP-B homolog 1, Cbh1)和CENP-B同源物2(CENP-B homolog 2, Cbh2)。所有这些蛋白都与着丝粒结合活性相关^[47-48]。

还有一种情况是共驯化(图2C)。已经报道了很多PIF/Harbinger分子驯化的例子, CASOLA等^[36]已经鉴定出七个不同的转座酶衍生基因被称为*DPLGI-*

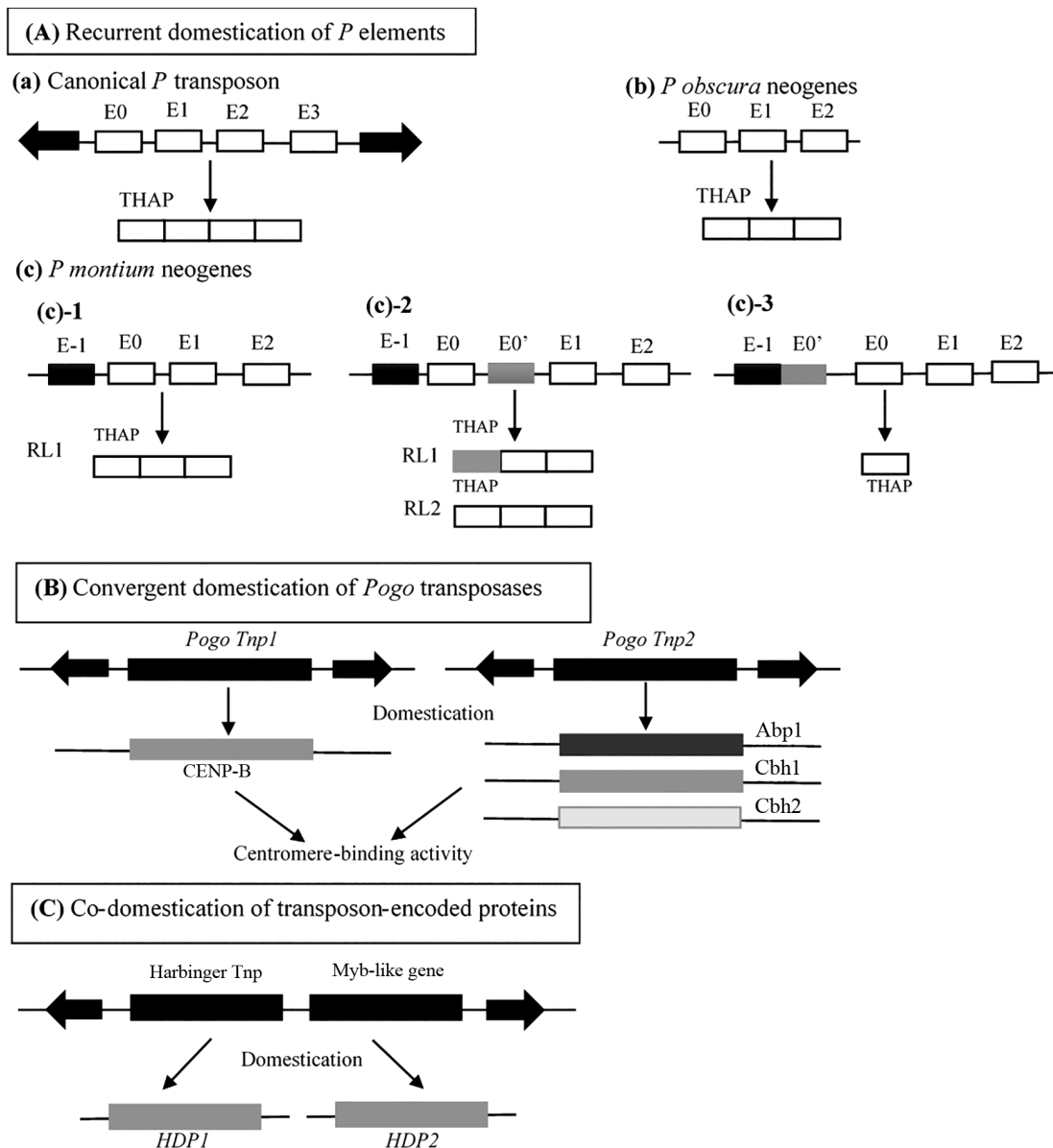


图2 转座子衍生基因的驯化机制

Fig.2 Domestication mechanism of transposon derived genes

7(*Drosophila* PIF-like genes 1-7), 它们可能来自至少三个独立的驯化事件。此外, 作者还在三种果蝇DPLG7A同源基因附近的区域发现了一个驯化的Myb-like蛋白DPMG7(*drosophila* PIF MADF-like protein-encoding gene 7)。上文提到的HDP1和HDP2是Harbinger转座子共驯化来的。这些发现都进一步支持了共驯化理论。

3 转座子衍生基因对生物生长发育的影响

转座子衍生基因通过调控基因表达来影响生物的生长发育, 表2列出了部分转座子衍生基因在生物生长发育过程中的作用。*daysleeper*是拟南芥中第一个被鉴定的重要的转座酶衍生基因, 编码一个转座酶, 该转座酶对拟南芥的发育至关重要。它来源于hAT超家族转座子, 包含这些元件编码序列中的许多特征, 是通过在体外结合DNA损伤反应基因*ku70*启动子的能力来鉴定的, 并被认为是可影响其他基因的转录^[49]。*Daysleeper*是一种主要的核蛋白, 该蛋白主要定位于细胞核, 位于跨高尔基体网络、晚期内小体和多泡小体(multivesicular body's, MVB's)中, 它主要在分生组织、发育中的花和果实中表达, 过度表达*daysleeper*的植物的生长和花型迟缓^[50]。此外, *daysleeper*基因在物种间结构上是保守的, 当这些结构在拟南芥中表达时, 会导致显性表型。*ricesleeper1*或*ricesleeper2*位点插入的水稻株系表现出表型异常, 表明这些基因对水稻的正常发育具有重要的功能^[16]。

异源多倍化是禾本科植物基因组进化的主要因素之一^[51], 在单子叶植物和双子叶植物分化之前, 被子植物中存在四个Mudra-like转座子序列分支。这些分支被命名为I类、II类、III类和IV类。进一步的结构特征和基因组分布分析表明, I类和II类包含具有转座子特征的元件, 但III类和IV类不具有TIR, 与驯化的转座酶相对应。这些所谓的“驯化的转座酶”已在宿主基因组中获得新功能^[52]。COWAN等^[21]在水稻和拟南芥中发现了一个新的驯化的Mudra-like转座酶家族*mustang*。在甘蔗多倍体基因组中, 分别鉴定出6个和26个拷贝的*mustang*基因^[52]。从III类中选择了两个单倍型*scmug266bac095*和*scmug266bac148*, 从IV类中选择了*scmug148bac249*基因。启动子-报告基因融合瞬时表达分析表明, *scmug266bac095*、*scmug266bac148*和*scmug148bac249*

含有活性启动子。此外, 这三个启动子都受到植物激素的负调控, 没有观察到特异性反应。*scmug266bac148*和*scmug148bac249*受生长素抑制, *scmug266bac095*经细胞分裂素和脱落酸处理后mRNA水平降低。生长素的抑制作用最强^[53]。这些结果表明, 功能多样化有助于不同*mustang*基因的正向选择, 并有助于将它们保留在甘蔗基因组中^[54]。将在异源系统中看到的*mustang*转录过程与在甘蔗中看到的表达模式进行比较, 可以得出这三个基因都能够表达, 并且表达强度差异不大; *scmug148bac249*调节区转录水平最高, 并在茎尖分生组织中转录最丰富。这一发现与*mustang1*水稻*mustang*基因的表达模式一致, *mustang1*基因在茎尖分生组织和花中表达水平最高^[55]。

我们知道光是植物生长和发育的主要信息来源, 光感受器光敏色素A(PhyA)介导各种远红光诱导的反应。*Fhy3*(far-red elongated hypocotyl 3)和*far1*(far-red impaired response 1)这两个转座酶衍生的转录因子编码Mutator-like转座酶相关的蛋白, 通过直接激活*fhy1*和*fhl*(*fhy1*-like)的转录, 共同调节PhyA信号介导的反应, 它们的产物对于光诱导的PhyA核积累和随后的光响应是必不可少的^[14]。*fhy3*和*far1*也是拟南芥叶绿素生物合成的正调控因子, *fhy3*和*far1*可以直接与启动子结合, 并通过激活编码丙氨酸脱水酶的*hemb1*的表达来促进叶绿素的生物合成。*fhy3*与光敏色素介导的信号反应的负调控因子PIF1在物理上相互作用, 以协调原叶绿素的合成和幼苗的绿化^[56-57]。此外, *fhy3*或*far1*的破坏减少了*fhy3-4*、*far1-2*和*fhy3-4 far1-2*突变植物中的淀粉积累并改变了淀粉颗粒结构。编码淀粉脱支酶(isoamylase2, ISA2)的基因是*fhy3*和*far1*的直接靶标, *fhy3*和*far1*通过与FBS(*fhy3/ far1* binding site)顺式元件在体内和体外直接与ISA2的启动子和5'编码区结合, 正调控ISA2的转录水平, 从而调控淀粉合成, 影响碳代谢和植物生长^[58]。

4 结论与展望

转座子在生物进化过程中扮演着重要角色, 本文综述了转座子衍生基因与转座子和常规基因的异同, 归纳了转座子衍生基因的演变途径, 以及对生物生长发育的影响。这些转座子衍生基因的例子证明了转座子具有深刻影响基因组功能的能力。

表2 转座子衍生基因的功能
Table 2 Function of transposon derived gene

基因名称 Gene name	转座子家族 TE-family	物种 Species	功能 Functions	参考文献 References
<i>daysleeper</i>	hAT	<i>Arabidopsis thaliana</i>	Development	[49]
<i>mustang1-8</i>	Mutator	<i>Arabidopsis thaliana</i>	Growth and flowering	[22,59]
<i>fhy3</i>	Mutator	<i>Arabidopsis thaliana</i>	Light-induced reaction	[14]
<i>far1</i>	Mutator	<i>Arabidopsis thaliana</i>	Light-induced reaction	[14]
<i>CENP-B</i>	Pogo	<i>Homo sapiens</i>	Centromeric chromatin assembly	[48]
<i>DPLG1-7</i>	PIF	<i>Drosophila melanogaster</i>	Unknown	[36]
<i>flc</i>	LTR (copia-like), MITE, hAT	<i>Arabidopsis thaliana</i>	Inhibit plant flowering	[60]
<i>thap</i>	P	<i>Drosophila melanogaster</i>	Induce apoptosis	[44]
<i>setmar</i>	Tc1/Mariner	<i>Homo sapiens</i>	Histone methylation	[34]
<i>tamrsi</i>	CACTA	<i>Antirrhinum majus</i>	DNA binding factor	[25]
<i>sun</i>	LTR (copia-like)	<i>Solanum lycopersicum</i>	Elongate fruit	[61]
<i>bomyb2</i>	Harbinger	<i>Brassica oleracea</i>	Regulates anthocyanin biosynthesis	[62]
<i>abfh12</i>	Pack-MULEs	<i>Oryza sativa</i>	Change GC content	[30]
<i>RAG1</i>	Transib	<i>Homo sapiens</i>	V(D)J recombination	[13]
<i>RAG2</i>	Transib	<i>Homo sapiens</i>	V(D)J recombination	[13]
<i>HDP1</i>	Harbinger	<i>Arabidopsis thaliana</i>	DNA demethylation	[41]
<i>HDP2</i>	Harbinger	<i>Arabidopsis thaliana</i>	DNA demethylation	[41]
<i>TIGD3</i>	Tc1/Mariner	<i>Latimeria chalumnae</i>	Unknown	[63]
<i>TIGD4</i>	Tc1/Mariner	<i>Latimeria chalumnae</i>	Unknown	[63]
<i>TIGD5</i>	Tc1/Mariner	<i>Latimeria chalumnae</i>	Unknown	[63]
<i>POGZ</i>	Pogo	<i>Latimeria chalumnae</i>	Unknown	[63]
<i>Tpb2</i>	PiggyBac	<i>Tetrahymena thermophila</i>	DNA excision	[64]
<i>Tpb1</i>	PiggyBac	<i>Tetrahymena thermophila</i>	Transposon-like excision	[65]
<i>Tpb6</i>	PiggyBac	<i>Tetrahymena thermophila</i>	Transposon-like excision	[65]

分子驯化导致新的宿主基因出现, 即转座子衍生基因, 这些基因具有重要的细胞功能, 包括转录调控、基于染色质的细胞周期控制、细胞增殖和凋亡。尽管越来越多的转座子衍生基因被发现, 但只有很少的转座子衍生基因具有功能特征。未来对转座子转座机制和进化的研究对新的转座子衍生基因的发现和功能鉴定具有重要的意义。由于功能域的保守, 一些转座子衍生基因可能保留了祖先转座酶的某些特定活性。因此, 基于与转座子转座反应机制的相似性, 可以潜在地阐明它们的生物学作用。毫无疑问, 在未来的几年里, 可能会发现更

多转座子衍生基因。其作为由转座子衍生的新基因, 或比原来的基因更加适应环境, 或者进化出新的功能, 具有巨大的研究价值。

参考文献 (References)

- [1] MCCLINTOCK B. Chromosome organization and gene expression [J]. Cold Spring Harbor Symp, 1951, 16: 131.
- [2] FRANCE, DENOEUDE, FRANCOISE, et al. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants [J]. Nature, 2012, 488: 213-7.
- [3] KAPITONOV V V, JURKA J. Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome [J]. Proc Natl Acad Sci USA, 2003, 100(11): 6569-74.

- [4] MATSUMOTO T, WU J Z, KANAMORI H. The map-based sequence of the rice genome [J]. *Nature*, 2005, 436(7052): 793-800.
- [5] SANMIGUEL P, GAUT B S, TIKHONOV A, et al. The paleontology of intergene retrotransposons of maize [J]. *Nat Genet*, 1998, 20(1): 43-5.
- [6] XU X, PAN S K, CHENG S F, et al. Genome sequence and analysis of the tuber crop potato [J]. *Nature*, 2011, 475: 189-95.
- [7] PENG Z, LU Y, LI L, et al. The draft genome of the fast-growing non-timber forest species Moso bamboo (*Phyllostachys heterocycla*) [J]. *Nat Genet*, 2013, 45(4): 456-63.
- [8] MAHBOD S, MOHAMED M H, ANDRE J, et al. Contribution of transposable elements in the plant's genome [J]. *Gene*, 2018, 665: 155-66.
- [9] MARIEANGELE G. Activation of plant retrotransposons under stress conditions [J]. *Trends Plant Sci*, 1998, 3(5): 181-7.
- [10] FESCHOTTE, CEDRIC, PRITHAM E J. DNA transposons and the evolution of eukaryotic genomes [J]. *Annu Rev Genet*, 2007, 41(1): 331-68.
- [11] Thomas W, FRANCOIS S, AURELIE H V, et al. A unified classification system for eukaryotic transposable elements [J]. *Nat Rev Genet*, 2007, 8(12): 973-82.
- [12] PRITHAM E J, PUTLIWALA T, CEDRIC F. Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses [J]. *Gene*, 2007, 390(1/2): 3-17.
- [13] KAPITONOV V V, KOONIN E V. Evolution of the RAG1-RAG2 locus: both proteins came from the same transposon [J]. *Biol Direct*, 2015, 10: 20-7.
- [14] LIN R, DING L, CASOLA C, et al. Transposase-derived transcription factors regulate light signaling in *Arabidopsis* [J]. *Science*, 2007, 318(5854): 1302-5.
- [15] MARIJN K, SYLVIA, PAUL J H. The SLEEPER genes: a transposase-derived angiosperm-specific gene family [J]. *BMC Plant Biol*, 2012, 12(1): 192.
- [16] SAITO R M, PERREAUIT A, PEACH B, et al. The CDC-14 phosphatase controls developmental cell-cycle arrest in *C. elegans* [J]. *Nat Cell Biol*, 2004, 6(8): 777-83.
- [17] SINZELLE L, IZSVAK Z, IVICS Z. Molecular domestication of transposable elements: from detrimental parasites to useful host genes [J]. *Cell Mol Life Sci*, 2009, 66(6): 1073-93.
- [18] Miller W J, HAGEMANN S, REITER E, et al. P-element homologous sequences are tandemly repeated in the genome of *Drosophila Guanche* [J]. *Proc Natl Acad Sci USA*, 1992, 89(9): 4018-22.
- [19] VOLLF J N. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *BioEssays: news and reviews in molecular* [J]. *Cell Dev Biol*, 2006, 28(9): 913-22.
- [20] ZHANG H M, ZHU J K. RNA-directed DNA methylation [J]. *Curr Opin Plant Biol*, 2011, 14(2): 142-7.
- [21] COWAN R K. MUSTANG is a novel family of domesticated transposase genes found in diverse Angiosperms [J]. *Mol Biol Evol*, 2005, 22(10): 2084-9.
- [22] HOLLISTER J D, SMITH L M, GUO L, et al. Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata* [J]. *Proc Natl Acad Sci USA*, 2011, 108(6): 2322-7.
- [23] MATZKE M A, MOSHER R A. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity [J]. *Nat Rev Genet*, 2014, 15(8): 570.
- [24] HOEN D R, BUREAU T E. Discovery of novel genes derived from transposable elements using integrative genomic analysis [J]. *Mol Biol Evol*, 2015, 32(6): 1487-506.
- [25] ROCCARO M, LI Y, SOMMER H, et al. ROSINA (RSI) is part of a CACTA transposable element, TamRSI, and links flower development to transposon activity [J]. *Mol Genet Genomics*, 2007, 278(3): 243-54.
- [26] HOEN D R, BUREAU T E. Transposable element exaptation in plants Topics in current genetics [J]. Berlin (Germany): Springer, 2012: 219-51.
- [27] PAN Y P, BO K L, CHENG Z H, et al. The loss-of-function GLABROUS 3 mutation in cucumber is due to LTR-retrotransposon insertion in a class IV HD-ZIP transcription factor gene CsGL3 that is epistatic over CsGL1 [J]. *BMC Plant Biol*, 2015, 15: 302-16.
- [28] NEKRUTENKO A, LI W H. Transposable elements are found in a large number of human protein-coding genes [J]. *Trends Genet*, 2001, 17(11): 619-21.
- [29] SONG X and CAO X. Transposon-mediated epigenetic regulation contributes to phenotypic diversity and environmental adaptation in rice [J]. *Curr Opin Plant Biol*, 2017, 36: 111-8.
- [30] JIANG N, FERGUSON A A, SLOTKIN R K, et al. Pack-Mutator-like transposable elements (Pack-MULEs) induce directional modification of genes through biased insertion and DNA acquisition [J]. *Proc Natl Acad Sci USA*, 2011, 108: 1537-42.
- [31] KUDLA G, LIPINSKI L, CAFFIN F, et al. High guanine and cytosine content increases mRNA levels in mammalian cells [J]. *PLoS Biol*, 2006, 4(6): 933-42.
- [32] TATARINOVA T V, ALEXANDROV N N, BOUCK J B, et al. GC3 biology in corn, rice, sorghum and other grasses [J]. *BMC Genomics*, 2010, 11: 308-25.
- [33] SNOWDEN K C, NAPOLI C A. Psl: a novel Spm-like transposable element from *Petunia hybrida* [J]. *Plant J*, 1998, 14(1): 43-54.
- [34] CORDAUX R, UDIT S, BATZER M A, et al. Birth of a chimeric primate gene by capture of the transposase gene from a mobile element [J]. *Proc Natl Acad Sci USA*, 2006, 103(21): 8101-6.
- [35] KAPITONOV V V, JURKA J. Harbinger transposons and an ancient HARBII gene derived from a transposase [J]. *DNA Cell Biol*, 2004, 23: 311-24.
- [36] CASOLA C, LAWING A M, BETRAN E, et al. PIF-like transposons are common in *drosophila* and have been repeatedly domesticated to generate new host genes [J]. *Mol Biol Evol*, 2007, 24(8): 1872-88.
- [37] HOEN D R, BUREAU T E. Discovery of novel genes derived from transposable elements using integrative genomic analysis [J]. *Mol Biol Evol*, 2015; 32(6): 1487-506.
- [38] JONES P, BINNS D, CHANG H Y, et al. InterProScan 5: genome-scale protein function classification [J]. *Bioinformatics*, 2014, 30(9): 1236-40.
- [39] BAO W, KOJIMA KCK, KOHANY O. Repbase update, a database of repetitive elements in eukaryotic genomes [J]. *Mobile DNA*, 2015; 6(1): 11-7.
- [40] YUAN Y W, WESSLER S R. The catalytic domain of all eukary-

- otic cut-and-paste transposase superfamilies [J]. *Proc Natl Acad Sci USA*, 2011, 108(19): 7884-9.
- [41] DUAN C G, WANG X G, XIE S J, et al. A pair of transposon-derived proteins function in a histone acetyltransferase complex for active DNA demethylation [J]. *Cell Res*, 2017, 27(2): 226-40.
- [42] MILLER W J, MCDONALD J F, NOUAUD D. Molecular domestication-more than a sporadic episode in evolution [J]. *Genetica*, 1999, 107(1/2/3): 197-207.
- [43] RIO D C, LASKI F A, RUBIN G M. Identification and immunochemical analysis of biologically active *Drosophila* P element transposase [J]. *Cell*, 1986, 44: 21-32.
- [44] QUESNEVILLE H, NOUAUD D, ANXOLABEHRE D. Recurrent recruitment of the THAP DNA-binding domain and molecular domestication of the P-transposable element [J]. *Mol Biol Evol*, 2005, 22(3): 741-6.
- [45] REISS D, NOUAUD D, RONSSERAY S, et al. Domesticated P elements in the *Drosophila* montium species subgroup have a new function related to a DNA binding property [J]. *J Mol Evol*, 2005, 61(4): 470-80.
- [46] KIPLING D, WARBURTON P E. Centromeres, CENP-B and Tigger too [J]. *Trends Genet*, 1997, 13(4): 141-5.
- [47] CASOLA C, HUCKS D, FESCHOTTE C. Convergent domestication of pogo-like transposases into centromere-binding proteins in fission yeast and mammals [J]. *Mol Biol Evol*, 25(1): 29-41.
- [48] OKADA T, OHZEKI J, NAKANO M, et al. CENP-B controls centromere formation depending on the chromatin context [J]. *Cell*, 2007, 131: 1287-300.
- [49] BUNDOCK P, HOOYKAAS P. An *Arabidopsis* hAT-like transposase is essential for plant development [J]. *Nature*, 2005, 436(7048): 282-4.
- [50] KNIP M, HIEMSTRA S, et al. DAYSLEEPER: a nuclear and vesicular-localized protein that is expressed in proliferating tissues [J]. *BMC Plant Biol*, 2013, 13: 211.
- [51] YAAKOV B, KASHKUSH K. Methylation, Transcription, and rearrangements of transposable elements in synthetic allopolyploids [J]. *Int J Plant Genom*, 2011, 2011: 1-7.
- [52] SACCARO N L, Sluys MAV, Varani ADM, et al. Mudra-like sequences from rice and sugarcane cluster as two bona fide transposon clades and two domesticated transposases [J]. *Gene*, 2007, 392(1/2): 117-25.
- [53] DANIELA K, FABIANA G, Thais AH, et al. Functional characterization of sugarcane mustang domesticated transposases and comparative diversity in sugarcane, rice, maize and sorghum [J]. *Genet Mol Biol*, 2012, 35(3): 632-9.
- [54] ADAMS K L, WENDEL J F. Novel patterns of gene expression in polyploid plants [J]. *Trends Genet*, 2005, 21(10): 539-43.
- [55] KWON S J, PARK K C, SON J H, et al. Sequence diversity of a domesticated transposase gene, MUG1, in *Oryza* species [J]. *Mol Cells*, 2009, 27(4): 459-65.
- [56] LEIVAR P, QUAIL P H. PIFs: pivotal components in a cellular signaling hub [J]. *Trends Plant Sci*, 2011, 16(1):19-28.
- [57] TANG W J, WANG W Q, CHEN D, et al. Transposase-derived proteins FHY3/FAR1 interact with PHYTOCHROME-INTERACTING FACTOR1 to regulate chlorophyll biosynthesis by modulating HEMB1 during deetiolation in *Arabidopsis* [J]. *Plant Cell*, 2012, 24(5): 1984-2000.
- [58] MA L, XUE N, FU X Y, et al. *Arabidopsis thaliana* FAR-RED ELONGATED HYPOCOTYLS3 (FHY3) and FAR-RED-IMPAIRED RESPONSE1 (FAR1) modulate starch synthesis in response to light and sugar [J]. *New Phytol*, 2017, 213(4): 1682-96.
- [59] JOLY-LOPEZ Z, EWA F, EMILIO V, et al. Abiotic stress phenotypes are associated with conserved genes derived from transposable elements [J]. *Front Plant Sci*, 2017, 8: 2027-43.
- [60] ZHAI J X, LIU J, LIU B, et al. Small RNA-directed epigenetic natural variation in *Arabidopsis thaliana* [J]. *PLoS Genet*, 2008, 4(4): 1-11.
- [61] XIAO H, JIANG N, SCHAFFNER E, et al. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit [J]. *Science*, 2008, 319: 1527-30.
- [62] CHIU L W, ZHOU X, BURKE S, et al. The purple Cauliflower arises from activation of a MYB transcription factor [J]. *Plant Physiol*, 2010, 154(3): 1470-80.
- [63] GAO B, SANG Y, ZONG W, et al. Evolution and domestication of Tc1/mariner transposons in the African coelacanth (*Latimeria chalumnae*) genome [J]. *Genome*, 2020, 8: 1-12.
- [64] CHENG C Y, ALEXANDER V, KAZUFUMI M, et al. A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in *Tetrahymena thermophila* [J]. *Mol Biol Cell*, 2010, 21(10): 1753-62.
- [65] CHENG C Y, YOUNG J M, LIN C Y G, et al. The piggyBac transposon-derived genes TPB1 and TPB6 mediate essential transposon-like excision during the developmental rearrangement of key genes in *Tetrahymena thermophila* [J]. *Genes Dev*, 2016, 30(24): 2724-36.