

## 高等植物转座元件功能研究进展\*

崔颢奎 曹晓风\*\*

(植物基因组学国家重点实验室, 国家植物基因研究中心, 中国科学院遗传与发育生物学研究所, 北京 100101)

**摘要** 转座元件是指在基因组中能够移动、复制并重新整合到基因组新位点的 DNA 片段。转座元件一度被视为基因组内的“垃圾”或“自私 DNA”，长期以来，转座元件的研究主要集中于阐释转座元件在宿主中的复制或表观沉默机制，而转座元件的调控功能并未得到全面探讨。已有研究表明，转座元件的比例与物种基因组大小存在正相关性，从而为 C 值悖论的解释提供了依据。近年来，越来越多的证据表明转座元件可以作为宿主基因组的“控制元件”发挥重要的调控作用。在作物中研究发现，转座元件既可以通过顺式或反式作用方式调控基因表达，也可以诱导表观等位基因的产生，从而促使固着生长的植物更好地适应外界环境的变化。本文拟就高等植物转座元件的作用及其对未来作物育种的意义进行总结。

**关键词** 转座子, 逆转座子, 调控元件, 表观等位基因

**学科分类号** Q344

**DOI:** 10.16476/j.pibb.2015.0328

真核生物的基因组含有大量的转座元件，转座元件是指在基因组中能够移动、复制并重新整合到基因组新位点的 DNA 片段。20 世纪 50 年代，美国遗传学家 Barbara McClintock 在研究玉米籽粒时发现种皮的颜色斑点呈现不稳定遗传和恢复突变，最初她将其描述为解离与激活突变位点，并根据其对基因的调节作用而把它们称为“控制元件”(controlling elements)，该现象的发现打破了传统观念认为真核生物基因组固定不变的属性，揭示出转座元件能够从染色体的一个位置移动到另一个位置的规律，她将这种规律称为转座(transposition)，并首次提出了“转座因子学说”理论<sup>[1-2]</sup>，Barbara McClintock 的这一重要发现也使她于 1983 年获得诺贝尔生理学奖。

转座元件是基因组中控制元件的观点在很长一段时间内并未得到普遍认可。然而由于转座元件在宿主基因组中只能通过“劫持”细胞内的复制机制使其自身被不停扩增，使得人们将其视为“自私 DNA”(selfish DNA)或“垃圾 DNA”(junk DNA)<sup>[3-4]</sup>。这些转座元件则被认为是存在于基因组的小分子寄生物(molecular parasites)。如果转座元件的存在只对基因组有害，往往会受到自然选择的淘汰，但事

实上很多转座元件被保留在基因组中。越来越多的研究也表明转座子的存在是与其宿主基因组共进化的结果，正因为它们为基因组进化和基因表达提供了驱动力，选择作用才将其保留并加以利用<sup>[5-9]</sup>。

在 McClintock 研究的基础上，近年来人们对转座元件在高等植物中的结构组成、复制过程和进化意义进行了充分研究<sup>[5-11]</sup>。随着表观遗传研究的深入，人们发现宿主基因组通过表观遗传修饰稳定维持这类元件的沉默，以降低转座发生对其基因组稳定性的威胁<sup>[12]</sup>。最新研究表明，宿主除了尽可能降低转座发生对其基因组稳定性造成的威胁外，转座元件还可以在转录水平(transcriptional level)或转录后水平上(post-transcriptional level)参与邻近基因的表达调控，并能以“顺式”(in cis)或“反式”(in trans)方式调控内源基因表达<sup>[11-13]</sup>。这些研究从多方面证实了 Barbara McClintock 提出的转座元件是“控制元件”的理论。本文拟对高等植物转座元件

\* 转基因重大专项(2015ZX08010-002)和第 55 批博士后基金(2014M550874)资助项目。

\*\* 通讯联系人。

Tel: 010-64869203, E-mail: xfcao@genetics.ac.cn

收稿日期: 2015-10-19, 接受日期: 2015-10-25

的调控作用及其对未来作物育种的意义进行综述。

## 1 真核生物中转座元件的结构与类型

根据转座元件在体内发生转座的方式, 它们可以被分为三大类(图 1)。第一类被称为 RNA 类型的逆转座子(retrotransposon), 它们的转座由 RNA 介导完成, 即 RNA 中间分子经反转录形成 cDNA 后被整合到基因组新位点(图 1a)。该类逆转座子根据其末端结构特征, 又被进一步分为含有长末端重复(long terminal repeat, LTR)和不含长末端重复(non-LTR retrotransposon)两大类。其中, *copia* 和 *gypsy* 两个亚类是 LTR 类型逆转座子的重

要组成部分, 它们在植物中以高拷贝形式存在。从转座活性角度, LTR 类型逆转座子又可分为自主型和非自主型两种: 自主型元件通常含有 *gag* 和 *pol* 两个基因, 前者负责编码衣壳蛋白, 后者负责编码多功能蛋白, 其具有蛋白酶、反转录酶、RNase H 以及整合酶的活性功能域; 非自主型元件缺少完整或大部分转座所需蛋白的编码基因, 其对应于自主元件的区域由不相关的序列或宿主序列组成<sup>[5, 9-10]</sup>。non-LTR 类型的逆转座子则分为长散在核重复元件(long interspersed nuclear element, LINE)和短散在核重复元件(short interspersed nuclear element, SINE)。LINE 元件的编码区由 *ORF1* 和

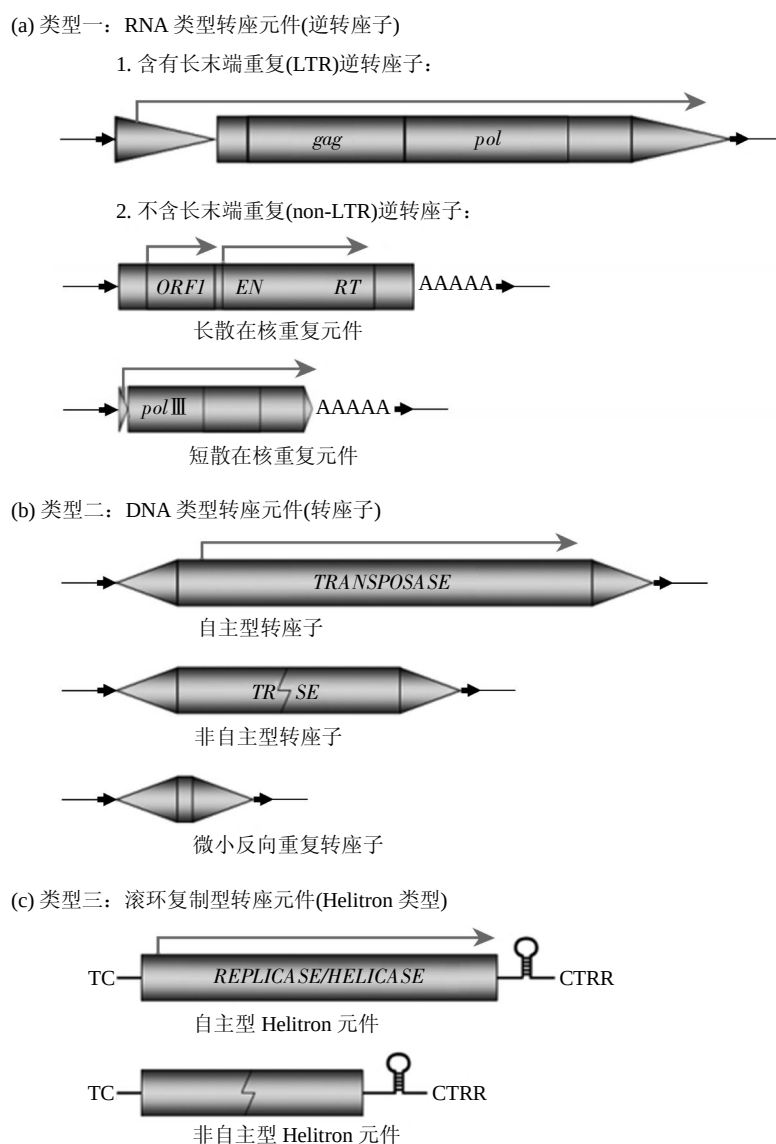


Fig. 1 Structure representation of transposable elements

图 1 不同转座元件的结构类型

*ORF2* 共同构成, *ORF1* 与 *gag* 基因编码的产物类似, *ORF2* 则含有内切酶(EN)和反转录酶(RT)的编码基因. LINE 和 SINE 均以简单的序列重复结尾, 通常有 poly(A). 对所有已知 SINE 分析发现, 它们的近 5' 端都含有一个潜在的 RNA pol III 启动子, 而除了 3' 端的序列与 LINE 同源外, 其余部分的特征还不清楚, 暗示 SINE 在基因组中作为非自主元件, 可能借助 LINE 的自主转座机制进行自我复制<sup>[5, 9-10, 14]</sup>. LINE 在植物中的比例较低, 而 SINE 则以高拷贝形式存在<sup>[15]</sup>.

如果把第一类逆转座子的转座方式看作“复制 - 粘贴”(copy and paste), 那么第二类转座子则以 DNA 介导的“剪切 - 粘贴”(cut and paste)方式进行转座, 这类元件被称为 DNA 类型的转座子(transposon) (图 1b). 第二类转座子具有末端反向重复序列(terminal inverted repeat, TIR)和靶位点重复序列(target site duplication, TSD), 其中非自主元件也被看作是自主型转座子发生在编码序列缺失的形式. 微型反向重复转座元件(miniature inverted-repeat transposable element, MITE)是非自主元件中拷贝数最多的转座子, 尽管它们不能自主转座, 但在动、植物物种均以高拷贝形式存在<sup>[9-10]</sup>. 非自主元件的另一个重要特征是它们能够携带宿主的基因片段发生转座<sup>[9]</sup>.

除了转座子和逆转座子外, Helitron 是另一类特殊的移动元件, 它们依赖滚环复制(rolling cycle)方式进行转座. 该类型转座子精确插入 AT 位点, 其末端具有保守的 5'-TC 和 3'-CTRR (R = A/G) 结构, 且 3' 端前 10~12 碱基处具有一个 16~20 个

碱基的发卡结构<sup>[16]</sup>(图 1c). Helitron 也包括自主和非自主两种类型, 前者携带编码自身转座所需的复制酶(REPLICASE)以及解旋酶(HELICASE)的基因; 后者在植物中以高丰度形式存在并常携带宿主的基因片段. 由于 Helitron 类型中自主型转座子与细菌中依赖滚环复制的结构相似, 人们也将 Helitron 看作 DNA 类型的转座子<sup>[17]</sup>. 此外, 由于 Helitrons 在基因组中的序列特征不明显, 而且不产生转座后的靶位点重复序列(TSD), 对其深入研究增加了难度<sup>[9]</sup>.

2 转座元件在动植物基因组中的组成与差异

转座元件虽然属于动、植物基因组的有效功能单元, 但在不同物种或近缘物种中, 它们的分布及数量有明显差异<sup>[15, 17]</sup>, 而造成这种差异的原因是有待人们去解决的重要科学问题.

在已完成基因组测序的一些植物物种中(图 2), 拟南芥(*Arabidopsis thaliana*) (125 Mb)<sup>[18]</sup>、短花药野生稻(*Oryza brachyantha*) (261 Mb)<sup>[19]</sup>、白菜(*Brassica rapa*) (283.8 Mb)<sup>[20]</sup> 和栽培水稻 (*Oryza sativa*) (430 Mb)<sup>[21-22]</sup> 具有相对较小的基因组, 转座元件所占各基因组比例分别为 10%、29.2%、39.5%和 46%; 而在基因组较大的玉米(*Zea mays*) (2 300 Mb)<sup>[23]</sup>、乌拉尔图小麦(*Triticum urartu*) (4 940 Mb)<sup>[24]</sup>、挪威云杉(*Picea abies*) (19 600 Mb)<sup>[25]</sup> 和大麦 (*Hordeum vulgare* L.) (51 000 Mb)<sup>[26]</sup> 中, 转座元件占全基因组的比例分别为 85%、66%、72%和 82%. 同时比较分析其他多种已测序物种<sup>[20, 22, 27-36]</sup>, 我们可以认为, 转座元件所占比例与基因组大小存在着一定的正相关性(图 2).

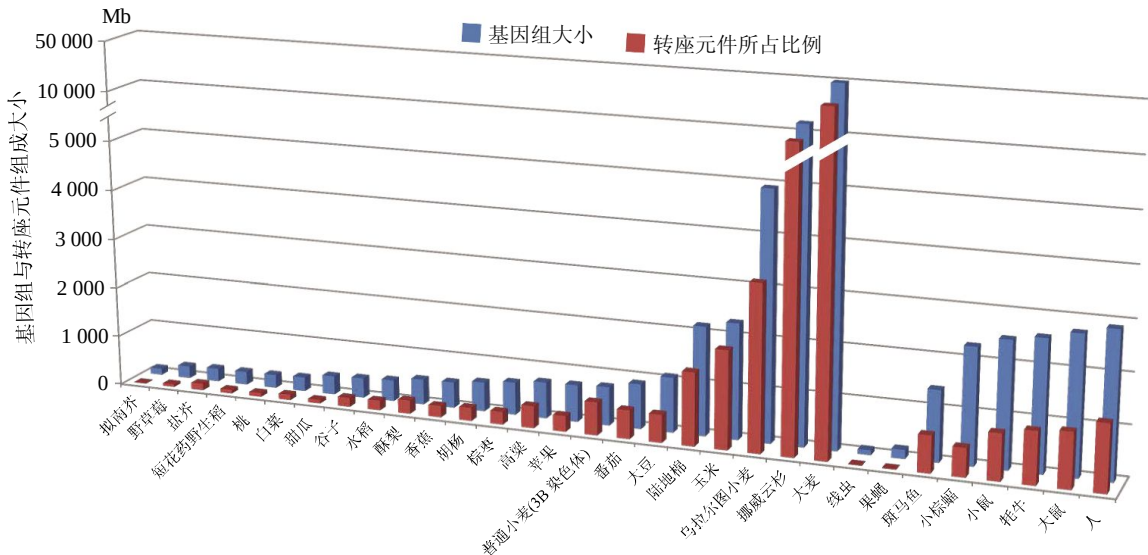


Fig. 2 Content of transposable elements in various genomes  
图 2 不同物种基因组中转座元件的含量

分析还发现,作物基因组中比例最富集的转座元件通常为 LTR 类型的逆转座子,而 non-LTR 类型的逆转座子 LINE 则比例很低<sup>[15]</sup>。例如,水稻基因组包含了几乎所有类型的转座元件<sup>[21]</sup>,LTR 和 non-LTR 类分别占 14.75%和 1.18%,其中 LINE 和 SINE 分别占基因组的 1.12%和 0.06%,表现出以 LTR 类逆转座元件为主导的分布模式。异源六倍体的普通小麦(*Triticum aestivum*)是人类粮食的重要来源,普通小麦具有较大的基因组(17 000 Mb),约为水稻基因组的 40 倍大小,其中,85%以上序列为重复序列,造成了小麦基因组测序和组装的困难<sup>[37]</sup>。普通小麦来自于 3 个二倍体基因组,即通过乌拉尔图小麦(*Triticum urartu*, AA 基因组)、拟斯卑尔托山羊草(*Aegilops speltoides*, BB 基因组)以及粗山羊草(*A. tauschii*, DD 基因组)三者先后杂交,再经历自然与人工选择产生了目前广泛种植的异源六倍体普通小麦(AABBDD 基因组)<sup>[38]</sup>。二倍体乌拉尔图小麦作为普通小麦的核心基因组供体,其基因组大小为 4 900 Mb,分别是水稻和人类基因组的 11 倍和 1.7 倍,其基因组的 67%由转座元件构成,LTR 逆转座子占将近 50%,而 DNA 类型和其他类型的转座元件分别为 9.77%和 8.04%。基因组比较研究发现,在进化过程中由于大量逆转座子在基因间的插入,导致了小麦 A 基因组的剧烈扩增<sup>[24]</sup>。对含有 D 基因组供体的粗山羊草测序发现,其基因组的 66%由转座子组成,与乌拉尔图小麦基因组类似,这些转座子绝大多数属于 LTR 的 *gypsy* 亚家族。在粗山羊草基因组中,含有抗病、抗非生物胁迫的基因数量都发生了显著扩张,从而为小麦品质、性状改良提供了遗传基础,而这些基因的扩张是否与 LTR 类型的逆转座子的增加有关,则值得深入研究<sup>[39]</sup>。雷蒙德氏棉(*Gossypium raimondii*)是首个被测序 D 基因组的棉花种类<sup>[40]</sup>,包括 40 976 个蛋白编码基因,而转座元件则占基因组大小的 57%,其中逆转座子比例为 78.6%;A 基因组的亚洲栽培棉(*Gossypium arboreum* L.)基因组包含 41 330 个蛋白编码基因,转座元件占基因组的 68.5%,其中,逆转座子比例为 68%<sup>[41]</sup>。在长期进化中,二倍体 A 基因组和 D 基因组棉花杂交形成了新的异源四倍体(A<sub>1</sub>D<sub>1</sub>)——陆地棉(*Gossypium hirsutum*),并逐渐成为人类的重要经济作物,它含有 76 943 个蛋白编码基因,基因组大小为 2 173 Mb,研究表明,陆地棉中 66%为转座元件,并以 LTR 类型的 *copia* 和 *gypsy* 逆转座子占主导,

比较分析发现,前者在 D<sub>1</sub> 亚基因组中较高;而后者则在 A<sub>1</sub> 亚基因组中较高<sup>[42]</sup>。

水稻(*O. sativa*)与同属的其他 23 个稻种组成了稻属,其中 AA 基因组类型最为复杂,它包含水稻,即亚洲栽培稻和其他 7 个稻种(普通野生稻 *O. rufipogon*、尼瓦拉野生稻 *O. nivara*、非洲栽培稻 *O. glaberrima*、短舌野生稻 *O. barthii*、展颖野生稻 *O. glumaepatula*、长雄蕊野生稻 *O. longistaminata* 和南方野生稻 *O. meridionalis*)<sup>[43-44]</sup>。最近一项对 AA 基因组 5 个稻种(*O. nivara*、*O. glaberrima*、*O. barthii*、*O. glumaepatula*、*O. meridionalis*)的研究表明,它们的基因组大小从 334.7 Mb(*O. glumaepatula*)到 375.0 Mb(*O. nivara*)变化不等,其中重复元件的比例分别为:约 27.6%、29.4%、29.8%、29.9%和 29.8%,尽管这些组成略低于水稻中的水平,但绝大多数转座元件都属于 LTR 类型,分别为:18.0%、18.7%、19.3%、18.9%以及 19.6%<sup>[45]</sup>。进一步对水稻与这 5 个稻种的基因组比较分析发现,转座元件的插入和丢失所带来的基因组动态变化是产生物种分化的重要原因,尤其是一些较短 DNA 类转座元件的重新整合是造成相关基因组结构变异的重要因素<sup>[46]</sup>。同时,这也暗示着相关变异可能与稻属物种间的表型变异、开花发育多样、生殖隔离形成之间关系密切,揭示了物种特有基因家族扩增、丢失、重要基因缺失与转座元件之间的特殊关系。短花药野生稻是稻属中与水稻亲缘关系较远的一类野生稻,具有对多种水稻病原菌和一些非生物胁迫的抗性。同时,短花药野生稻具有稻属最小的基因组(FF, 261 Mb),仅 29.2%的序列由转座子组成并且 LTR 类型反转座子活性较低,因此该基因组比较稳定。另外,不断发生的转座子删除,使得短花药野生稻基因组远小于水稻基因组<sup>[49]</sup>。根据上述研究结果,我们认为转座元件所占的比例与基因组大小存在着正相关性,同时也暗示着此类元件,尤其是 LTR 类型反转座子,在作物驯化中可能具有重要作用。

在动物中,目前也已经报道了多个物种的转座元件与基因组构成<sup>[46]</sup>(图 2)。例如,果蝇(*Drosophila melanogaster*)中含有约 20 个 LTR 逆转座子家族的成员,构成其 1%左右的基因组<sup>[47]</sup>,而玉米中,400 多个 LTR 逆转座子家族成员构成了基因组约 75%的区域<sup>[23]</sup>。小鼠(*Mus musculus*)基因组含有多个激活形式的 LTR 类元件,其转座导致 10%~12%的小鼠基因组突变<sup>[48]</sup>。相比而言,人类(*Homo sapiens*)

基因组中含有非常少的 LTR 逆转座子<sup>[49]</sup>, 但 non-LTR 逆转座子则在哺乳动物中尤其丰富<sup>[50]</sup>, 例如, 长散在重复元件 L1 及其非自主 SINE 元件 *Alu* 和 SVA (SINE-R-VNTR-*Alu* element)<sup>[51-52]</sup>, 通过转座和复制构成了接近 30% 的人类基因组<sup>[49]</sup>. 研究表明, L1 的重组造成了人类个体间基因组的结构变异, 表明 L1 的复杂性及其对人类基因组构成的深远影响<sup>[53]</sup>. 最近, 对中国青藏高原地区分布的特殊类群牦牛(*Bos grunniens*)基因组的研究表明, 40% 的牦牛基因组由转座元件构成, 其高海拔和低氧环境的适应性进化与基因家族扩张存在着相关性, 例如, 包含 HIG 结构域的低氧诱导基因家族在该物种中发生了扩张, 使得该家族在牦牛中有 13 个拷贝, 多于任何其他哺乳动物. 因此, 阐明转座元件是否参与新基因的产生并造成牦牛对低氧环境适应性将会非常有意义<sup>[54]</sup>. 另外, 在作物中, 表现出以 LTR 类型为主导的分布模式, 与人类基因组中的分布规律明显不同. 目前我们还不了解为何存在这种分布差异, 推测认为, non-LTR 和 LTR 逆转座子在不同物种的基因组进化中有不同的作用.

全基因组 DNA 甲基化的研究表明, 玉米近交系<sup>[55]</sup>、番茄<sup>[56]</sup>、大豆<sup>[57]</sup>中的 DNA 甲基化发生位点和模式与拟南芥极其类似, 其甲基化都发生在 CG、CHG 和 CHH 序列上; 同时, 它们都在近着丝粒区域显著富集; 此外, 存在于转座元件上的 DNA 甲基化既可发生在 CG, 也可在 CHG 和 CHH 上; 对于蛋白编码基因, DNA 甲基化更倾向于发生在 CG 类型上<sup>[58-59]</sup>. 目前, 多数作物或其他被子植物与拟南芥都有相似的 DNA 甲基化修饰, 并且在水稻、玉米和拟南芥中, 这些 DNA 甲基化区域也是小分子干扰 RNA (small interfering RNA, siRNA) 产生和富集的位置<sup>[55, 58, 60]</sup>. 虽然水稻、玉米和拟南芥中的 DNA 甲基化整体变化趋势相似, 但它们之间仍然存在差异, 这种差异与转座元件的组成密切相关. 其中最明显的特征是基因组大小的增加伴随着 DNA 甲基化修饰的增加, 而转座元件的比例正是造成基因组水平上 DNA 甲基化程度变化的直接因素; 同时, 与拟南芥相比, 水稻中的转座元件则以更加散在的形式分布于基因组, 这与 DNA 甲基化在水稻中的分布模式非常一致, 暗示转座元件的扩增可能是造成 DNA 甲基化修饰变化的潜在原因. 此外, 在拟南芥中, 转座元件和蛋白编码基因的长度与 DNA 甲基化强度呈正相关<sup>[58]</sup>, 而这一特征在水稻中只存在于蛋白编码基因上<sup>[60]</sup>,

表明水稻基因组中的转座元件序列被高度甲基化, 并与序列的长度无关<sup>[61]</sup>. 由此我们推测, 在不同植物中, 转座元件差异所带来的基因组结构变异, 可能是塑造植物表观基因组(epigenomics)的重要原因. 在人类基因组中, 约 97% 的序列由转座元件、假基因、重复序列或某些未知序列构成, 而编码序列低于 1.5%, 这些编码基因序列散在地分布在染色体上, 呈现出每条染色体在某些区域富含基因而另一些区域基因较少的特征<sup>[62]</sup>, 基因密度所呈现的非随机模式的重要性仍然未知, 这一现象与水稻基因组有一定的相似性, 暗示不同区域的蛋白编码基因相间分布与非编码序列的散在分布可能存在着相关性; 此外, 人类基因组中这些非编码序列的大量转录也暗示它们可能具有更加重要的作用, 是否也影响表观基因组修饰则需要更进一步的研究<sup>[49]</sup>.

### 3 转座元件在植物中的表观遗传调控

虽然动植物中存在着大量转座元件, 但表观遗传修饰参与维持其沉默状态, 其中, DNA 甲基化、组蛋白修饰和小分子 RNA 是抑制转座发生的重要方式, 通常在转录或转录后水平进行沉默.

在拟南芥中, *DDM1* (decrease in DNA methylation 1) 和 *DDM2* (也称为 *MET1*) 分别编码染色质重塑蛋白和 CpG 甲基转移酶<sup>[63]</sup>. 在 *ddm1* 和 *met1* 突变体中, 以 CG 类型为主的胞嘧啶甲基化降低了约 70%. 而 Chromomethyltransferase 3 (CMT3)、DRM2 (domains rearranged methyltransferase 2) 和 CMT2 是负责拟南芥中 non-CG DNA 位点的 DNA 甲基转移酶<sup>[63-64]</sup>. 在 *cmt3* 和 *met1* 双突变体中, 转座子 *CACAT* 的转录水平明显升高, 并伴随着它的转座<sup>[65]</sup>. 另外, *CACAT* 的转座在 *ddm1* 突变体中也被激活, 还发现其他类型转座元件也被激活<sup>[66]</sup>. 近期对拟南芥 *DDM1* 的研究表明, 较短的转座元件易于被 CHH 类型甲基化抑制, 而较长的转座元件则主要被 CG 和 CHG 类型的 DNA 甲基化抑制<sup>[67]</sup>. 通常情况下, 转录的升高是转座发生的必要前提, 但仅有转录的升高不足以激活转座, 说明转录后的机制在另一水平参与转座元件的抑制<sup>[59, 68-69]</sup>. 对拟南芥 *copia* 类逆转座子 *Évadé* (*EVD*) 的研究就证明了这一假设, 即 *Évadé* 的转座需要同时缺失 RNA 指导的 DNA 甲基化和 H3K9me2 修饰<sup>[70]</sup>. 现在我们已经清楚地认识到, DNA 甲基化与组蛋白甲基化修饰之间的相互作用为转座的抑制增加了双重保证. 研究表明, H3K9 甲基转移酶 Kryptonite

(KYP)对 CHG 类型 DNA 甲基化的识别引导了 H3K9me2 抑制修饰,而 CMT3 能够结合 H3K9me2 并造成其附近 DNA 上的 CHG 甲基化,从而产生正反馈协同作用抑制转座<sup>[63,71]</sup>.此外,对表观重组自交系(epigenetic recombination inbred lines, epiRILs)的利用<sup>[72-73]</sup>,为人们深入认识转座元件在宿主基因组中的入侵、扩增以及沉默提供了有利条件.例如,在 *met1* 或 *ddm1* 诱导的 *MET1*-或 *DDM1*-epiRILs 中, siRNA 介导的转录后水平沉默途径首先降解 *EVD* RNA,然而经过数代后,其拷贝数不断增加,不足以被降解的 *EVD* RNA 则由其自身翻译的 GAG 蛋白结合形生 RNP,而不受 siRNA 途径切割降解,但是, *EVD* RNA 转录本在后代中的增加则产生 siRNA 介导的 DNA 甲基化,从而引导 RNA 聚合酶 Pol IV-Pol V 介导的反义转录,最终产生 5' LTR 区域的 DNA 甲基化,并转换为转录水平的长期稳定抑制<sup>[74]</sup>.因此, DNA 甲基化对转座元件的抑制发挥着重要作用.

在水稻中, *Tos17* 的转录升高和转座发生与 DNA 甲基化的降低直接相关<sup>[75]</sup>,而目前的研究进一步指出,编码 H3K9me2 甲基转移酶 *SDG714* 的缺陷,或编码 DNA 去甲基化酶 *DNG701* 的过表达均会导致 *Tos17* 的转座激活<sup>[76-77]</sup>,说明组蛋白 H3K9 甲基化修饰在水稻基因组中对转座元件也具有抑制作用. *Karma* 是水稻中第一个被发现的 LINE 元件,它的转座只发生在培养多代的再生苗中,研究表明, *Karma* 的转录只在培养的愈伤和再生的小苗中有所升高,但这不足以激活它在愈伤和再生当代小苗中的转座,分析 DNA 甲基化发现,它在愈伤和再生苗中的水平均比正常情况低,说明转录后修饰对它的转座具有更重要的调节作用;此外,在再生多代的苗中发生转座也从另外一个角度说明 *Karma* 转座还受发育阶段的调控<sup>[78]</sup>.更进一步的研究表明,水稻中组蛋白 H3K4 去甲基化酶 JMJ703 对 *Karma* 及其 N 端缺失形式 *LINE1* 的转座活性具有控制作用<sup>[79]</sup>.在 *jmj703* 突变体中, H3K4me3 在 *Karma* 5'端异常升高,并抑制了该区域本身的 DNA 甲基化修饰,导致 *Karma* 在突变体中转录升高并激活它自身与 *LINE1* 的转座.同时,也发现 *Tos17* 的沉默抑制不依赖于 JMJ703 介导的 H3K4me3 去甲基化,而受到 H3K9me2 与 DNA 甲基化的抑制.对水稻基因组的深入分析发现, *Karma* 和 *LINE1* 分别处于不同染色体上 H3K4me3 较低的异染色质区,其上下游富集转座元件,而

*Tos17* 上下游均为蛋白编码基因,其上存在正常的 H3K4me3 修饰,表明 *Tos17* 处于常染色质区<sup>[79-80]</sup>.这一结果暗示,水稻基因组利用不同的表观遗传修饰来区分标记不同染色质微环境(chromatin niche)中的转座元件,它们被相应地去除激活修饰(例如, H3K4me3)或者增加抑制修饰(例如, H3K9me2 或 DNA 甲基化修饰),持续处于沉默状态,最终达到维持基因组稳定性的目的.

在动植物中,小分子 RNA 介导的转座抑制是另一种重要的方式,其中包括内源 small interfering RNAs (endo-siRNAs) 和 PIWI-interacting RNAs (piRNAs)两种途径<sup>[50]</sup>. siRNA 由转座子末端反向重复序列形成的 dsRNA 产生,长度约为 21~24nt,在 RNA 诱导沉默复合体 RISC 的帮助下与靶位点互补,使其降解,完成对转座子的抑制. piRNA 的产生基于含有转座元件的较长 RNA 前体,其长度约为 24~35nt,最终在 PIWI 蛋白的指导下降解目标转座元件.在植物中,现在研究比较清楚的是拟南芥花粉中小 RNA 对转座子的抑制,其营养核中的 LTR 逆转座子 *Athila* 可产生 21nt 的 siRNA 沉默信号,它被传递至两个精核中抑制其中的转座元件,保证植物的遗传信息正确传递<sup>[81]</sup>.另外也发现,拟南芥受热激后, siRNA 途径抑制了 *ONSEN* 逆转座子在代际间的转座传递<sup>[82]</sup>.最近有研究报告,尽管并未发现转座被激活的元件,但拟南芥组蛋白 H3K4 去甲基化酶 JMJ14 与 RNA 介导的转座元件沉默存在一定联系<sup>[83-85]</sup>.以上结果表明,组蛋白上发生的翻译后共价修饰(组蛋白密码)、DNA 甲基化或小分子 RNA 是表观遗传调控的重要方式,这些修饰在体内的平衡是一种复杂而精细的动态过程,它们在基因转录调控和转座子沉默等过程中都发挥着重要作用.

#### 4 转座元件在植物基因组中的调控作用

转座元件在数量上的扩张和收缩伴随着宿主基因组的防御机制,由此不仅改变了基因组的整体结构,而且也造成基因表达和功能的变化.转座元件对插入位点基因的影响主要表现为:基因自身功能突变以及新功能化、基因结构变异、核酸序列和表观遗传修饰的重新编排等(图 3),这些影响最终可能造成表型变异.

转座元件插入基因的位置对其表达具有不同作用<sup>[8,11-13]</sup>.转座子整合到基因的增强子上,则造成基因表达的抑制,而整合在抑制子内,则开启基因表



达(图 3a). 转座元件的启动子区可从不同角度影响邻近基因表达, 当转座元件携带新的增强子或抑制子时, 它不仅可增强或抑制转录, 而且也可能使这一转录本具有新的转录后修饰能力. 此外, 转座元件的 5' 及 3' 区除了包含潜在的顺式作用元件外, 其 5' 区的一些序列元件也可作为启动子起始转录, 通过这些潜在顺式元件的改变, 可以使下游基因产生异位表达或对外界环境做出快速响应(图 3b). 在很多情况下, 转座元件在基因组中的异位整合可以使编码基因发生转录提前终止, 这主要是由于它所携带的遗传信息在新的基因位置产生 poly(A) 信号的结果. 转座的发生也能引入新的可变拼接位点或使转座子本身的序列外显子化等, 这些功能单元的出现使基因表达被重新调整并建立新的转录模式(图 3b). 此外, 转座元件也可造成宿主基因组的结构变化, 例如将基因的片段通过转导的方式置换到新的位置; 通过介导非等位基因间的同源重组, 产生序列删除、重复或重排等结构性改变; 通过转座介导的假基因掺入使其外显子化并产生新功能等(图 3c). 最后, 当转座元件在整合位点具有反义启动子效应时, 它能同时启动上下游的转录.

这一系列的调控方式随着研究的深入得到了证实: 第一, 转座子插入引起基因失活或者增强进而造成突变表型. 例如: 从最早发现影响玉米籽粒种皮色斑的转座子 *Ac* (activator)-*Ds* (*Dissociation*) 开始<sup>[1, 10]</sup>; 到后来发现, 谷子(*Setaria italica*)的糯性表型由转座子插入到颗粒结合淀粉合成酶基因 *GBSSI* (gene for granule-bound starch synthase 1) 导致<sup>[86]</sup>; 以及最近的研究表明, LTR 类反转座子 *Gret1* (grapevine retrotransposon1) 插入黑皮葡萄 *Cabernet* 的 *VvmybA1* 基因上游后, 抑制它的表达而产生白色果皮, 并产生白皮葡萄变种 *Chardonnay*, *Gret1* 在插入位点又造成 DNA 序列重排, 引起 *VvmybA1* 基因的恢复突变, 形成红皮葡萄变种 *Ruby Okuyama*<sup>[9, 87]</sup>. 在栽培玉米中, LTR 类反转座子 *Hopscotch* 整合到玉米驯化基因 *tb1* (teosinte branched 1) 上游形成潜在的增强子, 使得 *tb1* 转录升高并产生无分蘖或少分蘖表型, 最终通过增加植株的顶端优势被固定下来<sup>[88]</sup>. 近期在油棕榈中的研究进一步揭示了转座元件产生的表观等位基因在调控重要农艺性状中的作用. 尽管油棕榈树通过无性繁殖产生后代, 但仍有一部分被称为 “mantled” 的油棕榈容易长出异常花朵, 产生极少的油脂, 并且只能在植株发育的晚期检测到这种异常, 因而造

成严重经济缺陷. 由于该缺陷不遵循孟德尔遗传, 研究者利用全基因组甲基化进行了分析, 结果发现一个基因组位点的 DNA 甲基化变化与造成该缺陷表型的相关基因 *MANTLED* 存在关联. 进一步分析发现, 这一位点是嵌合在 *MANTLED* 基因内的一个反转座子, 并与水稻中的 *Karma* 同源, 该元件上 DNA 甲基化的缺陷导致编码基因不能正常拼接, 而将 *Karma* 作为新的拼接位点并翻译出突变蛋白, 造成植物产生无价值的果实, 研究者将这一缺陷基因状态称作 “Bad Karma”; 而在正常发育的油棕榈中, *Karma* 则具有高水平的 DNA 甲基化修饰, 拼接能准确完成, 并产生正常果实, 研究者称之为 “Good Karma”<sup>[89]</sup>. 在动植物中, 端粒酶在细胞增殖过程中发挥着必要作用, 而一旦存在 DNA 双链损伤, 端粒酶的活性锐减并让细胞进入忠实 DNA 修复过程以此让有机体达到 DNA 复制与保真的平衡, 一项在拟南芥中的新研究证实了端粒酶的这种活性是如何被快速可逆地调节. 研究表明, 端粒酶的活性受到一个被称为 *TER2* 的长非编码 RNA (lncRNA) 的调控, 这一非编码 RNA 的基因组位点在长期进化中经历了转座元件的整合插入, 并且这种整合只存在于某些拟南芥生态型中. 分析不同生态型发现, 转座元件在 *TER2* 基因组位点的整合增加了端粒酶催化亚基 *TERT* 对 *TER2* lncRNA 的亲合性, 一旦出现 DNA 双链损伤, *TER2* 被诱导并在含有 *TERT* 结构域的端粒酶复合体中富集, 从而快速地 “封闭” 端粒酶活性, 保证 DNA 修复进行. 因此, 在长期进化中, 转座元件作为细胞内的 “工具”, 通过整合插入方式 “优化” 基因组 “平台” 的结构或功能, 最终促进细胞增殖的稳定发生<sup>[90]</sup>. 第二, 转座子影响植物的环境适应性. *mPing* 是第一个被发现具有转座活性的水稻 MITE 类转座子<sup>[91-93]</sup>, 研究发现它的拷贝数会在一些水稻品系中发生快速扩增, 每一代增加约 40 个新的拷贝<sup>[94]</sup>; 分析表明, *mPing* 更易于整合到基因的 5' 区域, 其中一些旁侧含有 *mPing* 的基因能够在冷和盐胁迫下被诱导表达. 因此, *mPing* 在水稻基因组中的快速扩增及整合增加了遗传变异的多样性, 提高水稻对环境的适应能力<sup>[95]</sup>. 这些例子均说明, 转座元件通过整合基因组的方式在很大程度上影响了基因表达, 进而导致功能的变化并产生新的表型.

转座元件也能造成动、植物中基因组的表观遗传修饰改变<sup>[12-13]</sup>(图 3d). 通过影响邻近基因的转录, 产生表观等位基因(epialleles), 尽管它们可以在代

际间遗传,但由于其侧翼调控元件表观修饰的可变性,从而产生表观等位基因的亚稳定状态(metastable),并表现出不同细胞的表达差异特征<sup>[96]</sup>. 目前已知,转座元件对编码基因的表观遗传学效应具体包括三方面:组蛋白甲基化修饰的扩散(例如,动物中 H3K9me3/H4K20me3,植物中 H3K9me2)<sup>[97-100]</sup>、DNA 甲基化修饰的扩散<sup>[97-98]</sup>、转座元件产生的小分子 RNA 对基因的影响<sup>[101-104]</sup>. 这些元件既可通过“顺式作用”方式,也可通过“反式作用”方式调节基因转录<sup>[11-13]</sup>. 这些研究已经在拟南芥中得到证实,其中最经典的例子包括:*FWA* 基因<sup>[105-106]</sup>、*FLOWERING LOCUS C (FLC)* 基因<sup>[107-109]</sup>和 *BONSAI (BNS)* 基因<sup>[110]</sup>. 它们的共同特征是,在这些基因的旁侧都存在着转座元件,这些元件通过 DNA 甲基化、小分子 RNA 或 H3K9me2 修饰影响三者本身的表达模式,从而产生表型变异.

最近,在玉米和水稻中的三项研究,为转座元件如何通过表观遗传信息的传递来重塑作物基因表达提供了新证据. 在玉米中,*ZmCCT* 是影响开花的关键基因,*CACTA* 类转座子在该基因启动子区的插入,增加了这一位置的 DNA 甲基化修饰并降低 *ZmCCT* 的转录,从而减弱玉米对光周期的敏感性使其提前开花,而玉米祖先品种中并没有这一转座子的插入,因此,该转座子的插入可能是玉米从短日照到长日照驯化的重要原因<sup>[111]</sup>. 与此类似的另一个例子也证明了转座子插入在玉米驯化过程中的作用,研究者对全球不同地区的玉米自然变异群体进行了全基因组关联分析,结果发现位于 10 号染色体上一个 NAC 类转录因子编码基因 *ZmNAC111* 对玉米耐旱性具有重要作用. 在干旱敏感材料中,*ZmNAC111* 启动子区域整合插入了一个 82 bp 的 MITE 转座子,研究发现,该转座子通过产生小分

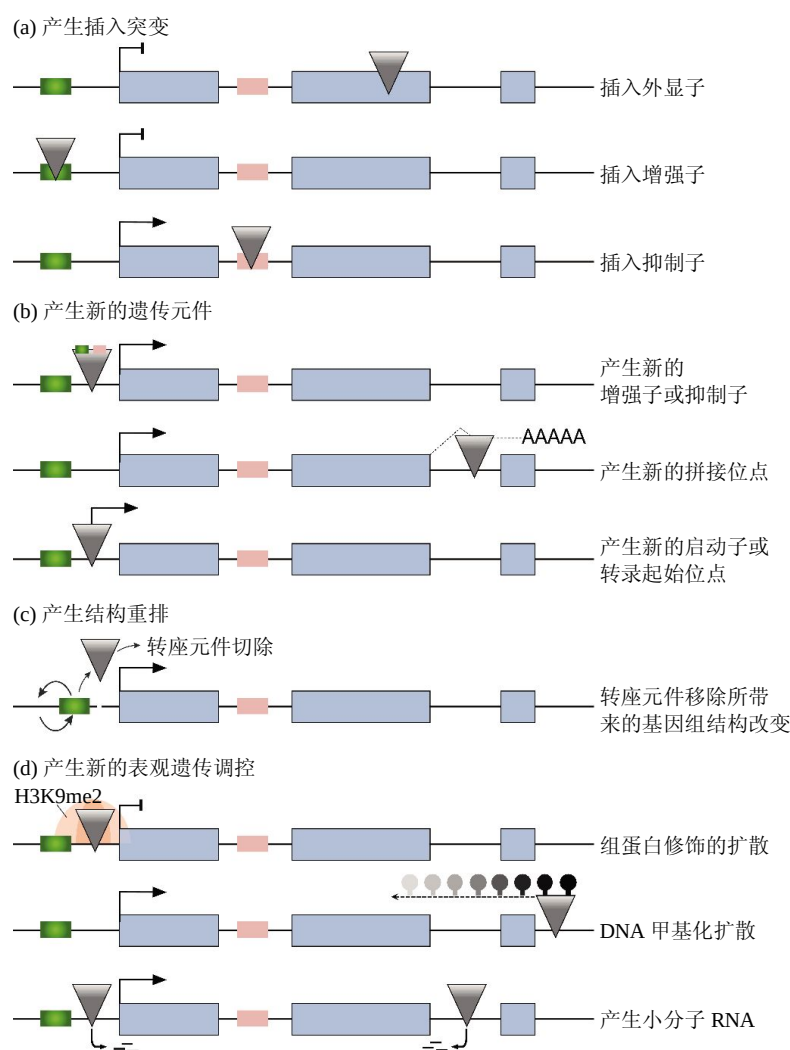


Fig. 3 Functional effects of transposable elements in host genome

图 3 转座子对基因组的影响



子 RNA 诱导该基因启动子区的 DNA 和组蛋白甲基化, 从而抑制 *ZmNAC111* 的表达, 促进了玉米由热带向温带的驯化; 而在耐旱材料中, *ZmNAC111* 启动子区域不含该 MITE 转座子。因此, 通过提高玉米内源 *ZmNAC111* 表达, 则可以促进叶片气孔在干旱胁迫下的关闭、提高水分利用效率, 从而增强玉米的耐旱能力<sup>[112]</sup>。OsDCL3a 是水稻中负责产生小分子 RNA 的核酸酶, 它的 RNAi 株系表现出矮化、叶夹角增大和穗长变短的多效缺陷表型。通过研究发现, OsDCL3a 主要以 MITE 类转座子为靶标加工产生 24nt 的小分子 RNA, 进一步分析表明, 这一类 24nt 小 RNA 则通过介导 H3K9me2 等异染色质修饰, 调控 MITEs 旁侧基因的表达, 并以赤霉素和油菜素内酯途径的关键基因作为靶基因影响水稻发育, 而这些基因在水稻株高、叶夹角等重要农艺性状建成中发挥着关键作用<sup>[104]</sup>。MITE 类转座子在动、植物中广泛存在, 这一机制可能在水稻、玉米等作物中存在保守性, 也可能会对肿瘤发生等研究提供一定的线索。

近年来, 多个研究组陆续报道了转座元件在动物与植物中的调控功能, 这些研究已不再从单个转座元件与邻近基因为例证实其功能, 而是从全基因组水平为我们阐释“移动元件组”(mobilome)与“转录组”(transcriptome)之间的联系<sup>[102, 113-115]</sup>。这些研究共同表明, 转座元件可在遗传学与表观遗传学水平对宿主转录组进行微调, 为基因组带来更丰富的转录式样, 增加其在不同环境或胁迫条件下的应激能力, 从而提高宿主的表型适应性。同时, 鉴于“移动元件组”所具有的潜在作用, 人们可以在未来尝试通过修饰这类元件的特性来达到调控基因表达的目的, 最终快速固定物种的优势特征或生理过程。

## 5 转座元件与多倍化和杂种优势

多倍化(polyploidy)是被子植物中常见的全基因组加倍现象, 其目的是进行自我进化和适应新环境, 研究表明大多数被子植物都经历过一次或者多次多倍化事件<sup>[116-117]</sup>。转座元件的复制与多倍化是造成植物基因组扩张的两种方式, 而多倍化造成的全基因组加倍却往往伴随着全基因组的结构变化、基因表达激活和沉默等多方面。其中, “亚基因组优势”或“亚基因组显性”(subgenome dominance)是多倍化事件中最明显的现象<sup>[118]</sup>。植物基因组加倍后, 染色体内的重组(intrachromosomal

recombination)会造成重复基因的删除, 并在漫长进化中逐步恢复到二倍体减数分裂行为(也即二倍化过程), 使得来源于一套亲本基因组的基因表达较高, 呈现显性特征, 而另一套亲本基因组则发生不对称丢失, 同源基因表达较低, 呈现隐性特征, 最终导致多倍化后亲本亚基因组组分的偏差(subgenome fractionation bias)<sup>[119]</sup>, 即所谓的亚基因组优势, 这种现象在异源多倍体中尤为显著<sup>[120]</sup>。而长期以来, 亲本亚基因组在多倍化后的不均等表现一直未被解析。最近, 以六倍体白菜<sup>[121-122]</sup>为研究对象, 对植物基因组多倍化进化过程中基因分化和多基因组分化机理进行了深入研究<sup>[123]</sup>。研究发现, 六倍体白菜基因组由 3 个亚基因组构成, 其中一个亚组的基因数量显著超过其他两个亚组, 并且在基因表达上也占优势。通过比较基因组学手段, 发现造成这种优势亚基因组的原因主要是 24nt 长小分子 RNA 抑制了转座子的活性, 而转座子的沉默过程则进一步造成邻近基因的抑制, 产生细胞内基因组层面的分化, 最终促进了优势亚基因组的稳定形成<sup>[123]</sup>。简言之, 亚基因组优势的产生是转座子抑制和基因表达之间折衷(tradeoff)的结果<sup>[124]</sup>。如果亚基因组优势是一种自然界的潜在机制, 那么这种机制则反映了近交衰退(inbreeding depression)的本质, 也就是说, 不同亲本间(例如玉米的不同品系)的杂交 F<sub>1</sub> 代能够保持杂种优势(hybrid vigor), 但子一代自交的 F<sub>2</sub> 代很多有害基因的作用都不再被有利基因的显性作用所掩盖, 出现近交衰退, 与多倍化后的基因组优势现象类似, 转座子抑制与基因表达之间的平衡过程最终使杂种基因组中来自一方亲本的基因组呈显性, 另一方呈隐性, F<sub>2</sub> 代则出现优势性状的衰退, 杂交优势无法在后代中得到保持<sup>[118]</sup>。

因此, 转座元件沉默抑制造成的基因组水平的分化, 不仅为阐明杂交优势现象提供依据, 也为作物遗传育种奠定新的理论基础。

## 6 展 望

随着表观遗传研究的不断深入, 人们对转座元件在真核生物中所扮演的角色有了愈加清楚的认识: 从最开始的结构特征、转座机制到后来的表观抑制修饰, 直至今天对其调节基因表达功能的进一步认识。在高等植物中, 由于转座元件的这些特征, 我们可以把它们看作是连接基因组变异和表观基因组变异的枢纽<sup>[60]</sup>, 即它们在受到表观抑制的同

时,也将携带的修饰信息传递到邻近的区域,诱导表观等位基因(epiallele)的形成,而这些等位基因可以在子代中进一步传递。

与其说由转座元件造成的表观等位基因具有不稳定性,不如说它具有遗传多样性,它赋予植物更丰富的表型可塑性,促使植物基因组具有潜在地快速响应外环境的能力,这一点在未来作物育种中尤为重要。尤其当有限的遗传多样性在育种中被充分利用时,这种潜藏在作物基因组中的表观遗传变异则可能会发挥更重要的作用<sup>[125]</sup>。同时,发掘更多受转座元件调节的表观等位基因,甚至表观数量性状位点(QTL<sup>epi</sup>)<sup>[126-127]</sup>,将会使我们更全面地认识并利用植物中的转座元件。

### 参 考 文 献

- [1] Mc C B. Chromosome organization and genic expression. Cold Spring Harbor Symposia on Quantitative Biology, 1951, **16**: 13-47
- [2] McClintock B. Controlling elements and the gene. Cold Spring Harbor Symposia on Quantitative Biology, 1956, **21**: 197-216
- [3] Doolittle W F, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. Nature, 1980, **284**(5757): 601-603
- [4] Orgel L E, Crick F H. Selfish DNA: the ultimate parasite. Nature, 1980, **284**(5757): 604-607
- [5] Bennetzen J L, Wang H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annual Review of Plant Biology, 2014, **65**: 505-530
- [6] Cordaux R, Batzer M A. The impact of retrotransposons on human genome evolution. Nature Reviews Genetics, 2009, **10** (10): 691-703
- [7] Fedoroff N V. Presidential address. Transposable elements, epigenetics, and genome evolution. Science, 2012, **338** (6108): 758-767
- [8] Feschotte C. Transposable elements and the evolution of regulatory networks. Nature Reviews Genetics, 2008, **9**(5): 397-405
- [9] Lisch D. How important are transposons for plant evolution? Nature Reviews Genetics, 2013, **14**(1): 49-61
- [10] Feschotte C, Jiang N, Wessler S R. Plant transposable elements: where genetics meets genomics. Nature Reviews Genetics, 2002, **3**(5): 329-341
- [11] Rebollo R, Romanish M T, Mager D L. Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annual Review of Genetics, 2012, **46**: 21-42
- [12] Lisch D. Epigenetic regulation of transposable elements in plants. Annual Review of Plant Biology, 2009, **60**: 43-66
- [13] Cui X, Cao X. Epigenetic regulation and functional exaptation of transposable elements in higher plants. Current Opinion in Plant Biology, 2014, **21C**: 83-88
- [14] Ogiwara I, Miya M, Ohshima K, *et al.* Retropositional parasitism of SINEs on LINEs: identification of SINEs and LINEs in elasmobranchs. Molecular Biology and Evolution, 1999, **16** (9): 1238-1250
- [15] Kumar A, Bennetzen J L. Plant retrotransposons. Annual Review of Genetics, 1999, **33**: 479-532
- [16] Kapitonov V V, Jurka J. Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci USA, 2001, **98**(15): 8714-8719
- [17] Kidwell M G. Transposable elements and the evolution of genome size in eukaryotes. Genetica, 2002, **115**(1): 49-63
- [18] Arabidopsis Genome I. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature, 2000, **408** (6814): 796-815
- [19] Chen J, Huang Q, Gao D, *et al.* Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying *Oryza* genome evolution. Nature Communications, 2013, **4**: 1595
- [20] Wang X, Wang H, Wang J, *et al.* The genome of the mesopolyploid crop species *Brassica rapa*. Nature Genetics, 2011, **43**(10): 1035-1039
- [21] International Rice Genome Sequencing P. The map-based sequence of the rice genome. Nature, 2005, **436**(7052): 793-800
- [22] Paterson A H, Bowers J E, Bruggmann R, *et al.* The Sorghum bicolor genome and the diversification of grasses. Nature, 2009, **457**(7229): 551-556
- [23] Schnable P S, Ware D, Fulton R S, *et al.* The B73 maize genome: complexity, diversity, and dynamics. Science, 2009, **326** (5956): 1112-1115
- [24] Ling H Q, Zhao S, Liu D, *et al.* Draft genome of the wheat A-genome progenitor *Triticum urartu*. Nature, 2013, **496** (7443): 87-90
- [25] Nystedt B, Street N R, Wetterbom A, *et al.* The Norway spruce genome sequence and conifer genome evolution. Nature, 2013, **497**(7451): 579-584
- [26] International Barley Genome Sequencing C, Mayer K F, Waugh R, *et al.* A physical, genetic and functional sequence assembly of the barley genome. Nature, 2012, **491**(7426): 711-716
- [27] Al-Dous E K, George B, Al-Mahmoud M E, *et al.* *De novo* genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). Nature Biotechnology, 2011, **29**(6): 521-527
- [28] Dhont A, Denoeud F, Aury J M, *et al.* The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature, 2012, **488**(7410): 213-217
- [29] Garcia-Mas J, Benjak A, Sanseverino W, *et al.* The genome of melon (*Cucumis melo* L.). Proc Natl Acad Sci USA, 2012, **109**(29): 11872-11877
- [30] Ma T, Wang J, Zhou G, *et al.* Genomic insights into salt adaptation in a desert poplar. Nature Communications, 2013, **4**: 2797
- [31] Shulaev V, Sargent D J, Crowhurst R N, *et al.* The genome of woodland strawberry (*Fragaria vesca*). Nature Genetics, 2011, **43**(2): 109-116
- [32] Tomato Genome C. The tomato genome sequence provides insights into fleshy fruit evolution. Nature, 2012, **485**(7400): 635-641
- [33] Vitte C, Fustier M A, Alix K, *et al.* The bright side of transposons in crop evolution. Briefings in Functional Genomics, 2014, **13** (4): 276-295

- [34] Wu H J, Zhang Z, Wang J Y, *et al.* Insights into salt tolerance from the genome of *Thellungiella salsuginea*. *Proc Natl Acad Sci USA*, 2012, **109**(30): 12219–12224
- [35] Wu J, Wang Z, Shi Z, *et al.* The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Research*, 2013, **23**(2): 396–408
- [36] Zhang G, Liu X, Quan Z, *et al.* Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nature Biotechnology*, 2012, **30**(6): 549–554
- [37] Brenchley R, Spannagl M, Pfeifer M, *et al.* Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature*, 2012, **491**(7426): 705–710
- [38] Petersen G, Seberg O, Yde M, *et al.* Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). *Molecular Phylogenetics and Evolution*, 2006, **39**(1): 70–82
- [39] Jia J, Zhao S, Kong X, *et al.* *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature*, 2013, **496**(7443): 91–95
- [40] Wang K, Wang Z, Li F, *et al.* The draft genome of a diploid cotton *Gossypium raimondii*. *Nature Genetics*, 2012, **44**(10): 1098–1103
- [41] Li F, Fan G, Wang K, *et al.* Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nature Genetics*, 2014, **46**(6): 567–572
- [42] Li F, Fan G, Lu C, *et al.* Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nature Biotechnology*, 2015, **33**(5): 524–530
- [43] Zhu Q, Ge S. Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. *The New Phytologist*, 2005, **167**(1): 249–265
- [44] Zou X H, Zhang F M, Zhang J G, *et al.* Analysis of 142 genes resolves the rapid diversification of the rice genus. *Genome Biology*, 2008, **9**(3): R49
- [45] Zhang Q J, Zhu T, Xia E H, *et al.* Rapid diversification of five *Oryza* AA genomes associated with rice adaptation. *Proc Natl Acad Sci USA*, 2014, **111**(46): E4954–4962
- [46] Huang C R, Burns K H, Boeke J D. Active transposition in genomes. *Annual Review of Genetics*, 2012, **46**: 651–675
- [47] Bowen N J, McDonald J F. *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. *Genome Research*, 2001, **11**(9): 1527–1540
- [48] Maksakova I A, Romanish M T, Gagnier L, *et al.* Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. *PLoS Genetics*, 2006, **2**(1): e2
- [49] Lander E S, Linton L M, Birren B, *et al.* Initial sequencing and analysis of the human genome. *Nature*, 2001, **409**(6822): 860–921
- [50] Levin H L, Moran J V. Dynamic interactions between transposable elements and their hosts. *Nature Reviews Genetics*, 2011, **12**(9): 615–627
- [51] Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. *Nature Genetics*, 2003, **35**(1): 41–48
- [52] Hancks D C, Goodier J L, Mandal P K, *et al.* Retrotransposition of marked SVA elements by human L1s in cultured cells. *Human Molecular Genetics*, 2011, **20**(17): 3386–3400
- [53] Beck C R, Garcia-Perez J L, Badge R M, *et al.* LINE-1 elements in structural variation and disease. *Annual Review of Genomics and Human Genetics*, 2011, **12**: 187–215
- [54] Qiu Q, Zhang G, Ma T, *et al.* The yak genome and adaptation to life at high altitude. *Nature Genetics*, 2012, **44**(8): 946–949
- [55] Regulski M, Lu Z, Kendall J, *et al.* The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. *Genome Research*, 2013, **23**(10): 1651–1662
- [56] Zhong S, Fei Z, Chen Y R, *et al.* Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. *Nature Biotechnology*, 2013, **31**(2): 154–159
- [57] Schmitz R J, He Y, Valdes-Lopez O, *et al.* Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. *Genome Research*, 2013, **23**(10): 1663–1674
- [58] Cokus S J, Feng S, Zhang X, *et al.* Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature*, 2008, **452**(7184): 215–219
- [59] Lister R, O'malley R C, Tonti-Filippini J, *et al.* Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell*, 2008, **133**(3): 523–536
- [60] Li X, Zhu J, Hu F, *et al.* Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. *BMC Genomics*, 2012, **13**: 300
- [61] Mirouze M, Vitte C. Transposable elements, a treasure trove to decipher epigenetic variation: insights from *Arabidopsis* and crop epigenomes. *Journal of Experimental Botany*, 2014, **65**(10): 2801–2812
- [62] Korenberg J R, Rykowski M C. Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. *Cell*, 1988, **53**(3): 391–400
- [63] Law J A, Jacobsen S E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, 2010, **11**(3): 204–220
- [64] Zhong X, Du J, Hale C J, *et al.* Molecular mechanism of action of plant DRM *de novo* DNA methyltransferases. *Cell*, 2014, **157**(5): 1050–1060
- [65] Kato M, Miura A, Bender J, *et al.* Role of CG and non-CG methylation in immobilization of transposons in *Arabidopsis*. *Curr Biol*, 2003, **13**(5): 421–426
- [66] Tsukahara S, Kobayashi A, Kawabe A, *et al.* Bursts of retrotransposition reproduced in *Arabidopsis*. *Nature*, 2009, **461**(7262): 423–426
- [67] Zemach A, Kim M Y, Hsieh P H, *et al.* The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell*, 2013, **153**(1): 193–205
- [68] Saze H, Mittelsten Scheid O, Paszkowski J. Maintenance of CpG methylation is essential for epigenetic inheritance during plant

- gametogenesis. *Nature Genetics*, 2003, **34**(1): 65–69
- [69] Zhang X, Yazaki J, Sundaresan A, *et al.* Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell*, 2006, **126**(6): 1189–1201
- [70] Mirouze M, Reinders J, Bucher E, *et al.* Selective epigenetic control of retrotransposition in *Arabidopsis*. *Nature*, 2009, **461** (7262): 427–430
- [71] Du J M, Zhong X H, Bernatavichute Y V, *et al.* Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell*, 2012, **151**(1): 167–180
- [72] Johannes F, Porcher E, Teixeira F K, *et al.* Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genetics*, 2009, **5**(6): e1000530
- [73] Reinders J, Wulff B B H, Mirouze M, *et al.* Compromised stability of DNA methylation and transposon immobilization in mosaic *Arabidopsis* epigenomes. *Genes & Development*, 2009, **23** (8): 939–950
- [74] Mari-Ordonez A, Marchais A, Etcheverry M, *et al.* Reconstructing *de novo* silencing of an active plant retrotransposon. *Nature Genetics*, 2013, **45**(9): 1029–1039
- [75] Cheng C, Daigen M, Hirochika H. Epigenetic regulation of the rice retrotransposon Tos17. *Molecular Genetics and Genomics: MGG*, 2006, **276**(4): 378–390
- [76] Ding Y, Wang X, Su L, *et al.* SDG714, a histone H3K9 methyltransferase, is involved in Tos17 DNA methylation and transposition in rice. *The Plant Cell*, 2007, **19**(1): 9–22
- [77] La H, Ding B, Mishra G P, *et al.* A 5-methylcytosine DNA glycosylase/lyase demethylates the retrotransposon Tos17 and promotes its transposition in rice. *Proc Natl Acad Sci USA*, 2011, **108**(37): 15498–15503
- [78] Komatsu M, Shimamoto K, Kyoizuka J. Two-step regulation and continuous retrotransposition of the rice LINE-type retrotransposon Karma. *The Plant Cell*, 2003, **15**(8): 1934–1944
- [79] Cui X, Jin P, Cui X, *et al.* Control of transposon activity by a histone H3K4 demethylase in rice. *Proc Natl Acad Sci USA*, 2013, **110**(5): 1953–1958
- [80] Cheng Z K, Buell C R, Wing R A, *et al.* Toward a cytological characterization of the rice genome. *Genome Research*, 2001, **11**(12): 2133–2141
- [81] Slotkin R K, Vaughn M, Borges F, *et al.* Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell*, 2009, **136**(3): 461–472
- [82] Ito H, Gaubert H, Bucher E, *et al.* An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature*, 2011, **472**(7341): 115–119
- [83] Deleris A, Greenberg M V, Ausin I, *et al.* Involvement of a Jumonji-C domain-containing histone demethylase in DRM2-mediated maintenance of DNA methylation. *EMBO Rep*, 2010, **11**(12): 950–955
- [84] Miura A, Nakamura M, Inagaki S, *et al.* An *Arabidopsis* jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. *The EMBO Journal*, 2009, **28**(8): 1078–1086
- [85] Searle I R, Pontes O, Melnyk C W, *et al.* JMJ14, a JmjC domain protein, is required for RNA silencing and cell-to-cell movement of an RNA silencing signal in *Arabidopsis*. *Genes & Development*, 2010, **24**(10): 986–991
- [86] Kawase M, Fukunaga K, Kato K. Diverse origins of waxy foxtail millet crops in East and Southeast Asia mediated by multiple transposable element insertions. *Molecular Genetics and Genomics: MGG*, 2005, **274**(2): 131–140
- [87] Kobayashi S, Goto-Yamamoto N, Hirochika H. Retrotransposon-induced mutations in grape skin color. *Science*, 2004, **304**(5673): 982
- [88] Studer A, Zhao Q, Ross-Ibarra J, *et al.* Identification of a functional transposon insertion in the maize domestication gene tb1. *Nature Genetics*, 2011, **43**(11): 1160–1163
- [89] Ong-Abdullah M, Ordway J M, Jiang N, *et al.* Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature*, 2015, **525**(7570): 533–537
- [90] Xu H, Nelson A D, Shippen D E. A transposable element within the Non-canonical telomerase RNA of *Arabidopsis thaliana* telomerase in response to DNA damage. *PLoS genetics*, 2015, **11**(6): e1005281
- [91] Jiang N, Bao Z, Zhang X, *et al.* An active DNA transposon family in rice. *Nature*, 2003, **421**(6919): 163–167
- [92] Kikuchi K, Terauchi K, Wada M, *et al.* The plant MITE mPing is mobilized in anther culture. *Nature*, 2003, **421**(6919): 167–170
- [93] Nakazaki T, Okumoto Y, Horibata A, *et al.* Mobilization of a transposon in the rice genome. *Nature*, 2003, **421**(6919): 170–172
- [94] Naito K, Cho E, Yang G, *et al.* Dramatic amplification of a rice transposable element during recent domestication. *Proc Natl Acad Sci USA*, 2006, **103**(47): 17620–17625
- [95] Naito K, Zhang F, Tsukiyama T, *et al.* Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature*, 2009, **461**(7267): 1130–1134
- [96] Rakyen V K, Blewitt M E, Druker R, *et al.* Metastable epialleles in mammals. *Trends in Genetics: TIG*, 2002, **18**(7): 348–351
- [97] Eichten S R, Ellis N A, Makarevitch I, *et al.* Spreading of heterochromatin is limited to specific families of maize retrotransposons. *PLoS Genetics*, 2012, **8**(12): e1003127
- [98] Gent J I, Ellis N A, Guo L, *et al.* CHH islands: *de novo* DNA methylation in near-gene chromatin regulation in maize. *Genome Research*, 2013, **23**(4): 628–637
- [99] Rebollo R, Karimi M M, Bilenky M, *et al.* Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. *PLoS Genetics*, 2011, **7**(9): e1002301
- [100] Sienski G, Donertas D, Brennecke J. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell*, 2012, **151**(5): 964–980
- [101] Mccue A D, Nuthikattu S, Reeder S H, *et al.* Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. *PLoS Genetics*, 2012, **8** (2): e1002474
- [102] Mccue A D, Nuthikattu S, Slotkin R K. Genome-wide identification of genes regulated in trans by transposable element

- small interfering RNAs. *RNA Biology*, 2013, **10**(8): 1379–1395
- [103]Nosaka M, Itoh J, Nagato Y, *et al.* Role of transposon-derived small RNAs in the interplay between genomes and parasitic DNA in rice. *PLoS genetics*, 2012, **8**(9): e1002953
- [104]Wei L, Gu L, Song X, *et al.* Dicer-like 3 produces transposable element-associated 24-nt siRNAs that control agricultural traits in rice. *Proc Natl Acad Sci USA*, 2014, **111**(10): 3877–3882
- [105]Chan S W, Zilberman D, Xie Z, *et al.* RNA silencing genes control *de novo* DNA methylation. *Science*, 2004, **303**(5662): 1336
- [106]Kinoshita Y, Saze H, Kinoshita T, *et al.* Control of FWA gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. *The Plant Journal: for Cell and Molecular Biology*, 2007, **49**(1): 38–45
- [107]Baurle I, Dean C. The timing of developmental transitions in plants. *Cell*, 2006, **125**(4): 655–664
- [108]Liu J, He Y, Amasino R, *et al.* siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in *Arabidopsis*. *Genes & Development*, 2004, **18**(23): 2873–2878
- [109]Zhai J, Liu J, Liu B, *et al.* Small RNA-directed epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Genetics*, 2008, **4**(4): e1000056
- [110]Saze H, Kakutani T. Heritable epigenetic mutation of a transposon-flanked *Arabidopsis* gene due to lack of the chromatin-remodeling factor DDM1. *The EMBO Journal*, 2007, **26**(15): 3641–3652
- [111]Yang Q, Li Z, Li W, *et al.* CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. *Proc Natl Acad Sci USA*, 2013, **110**(42): 16969–16974
- [112]Mao H, Wang H, Liu S, *et al.* A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. *Nature Communications*, 2015, **6**: 8326
- [113]Cowley M, Oakey R J. Transposable elements re-wire and fine-tune the transcriptome. *PLoS genetics*, 2013, **9**(1): e1003234.
- [114]Springer N M. Epigenetics and crop improvement. *Trends in genetics : TIG*, 2013, **29**(4): 241–247
- [115]Wang X, Weigel D, Smith L M. Transposon variants and their effects on gene expression in *Arabidopsis*. *PLoS Genetics*, 2013, **9**(2): e1003255
- [116]Blanc G, Wolfe K H. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *The Plant Cell*, 2004, **16**(7): 1667–1678
- [117]Cui L, Wall P K, Leebens-Mack J H, *et al.* Widespread genome duplications throughout the history of flowering plants. *Genome Research*, 2006, **16**(6): 738–749
- [118]Freeling M, Woodhouse M R, Subramaniam S, *et al.* Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. *Current Opinion in Plant Biology*, 2012, **15**(2): 131–139
- [119]Freeling M, Thomas B C. Gene-balanced duplications, like Tetraploidy, provide predictable drive to increase morphological complexity. *Genome research*, 2006, **16**(7): 805–814
- [120]Garsmeur O, Schnable J C, Almeida A, *et al.* Two evolutionarily distinct classes of paleopolyploidy. *Molecular Biology and Evolution*, 2014, **31**(2): 448–454
- [121]Subramaniam S, Wang X, Freeling M, *et al.* The fate of *Arabidopsis thaliana* homeologous CNSs and their motifs in the *Paleohexaploid Brassica rapa*. *Genome Biology and Evolution*, 2013, **5**(4): 646–660
- [122]Tang H, Woodhouse M R, Cheng F, *et al.* Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy. *Genetics*, 2012, **190**(4): 1563–1574
- [123]Woodhouse M R, Cheng F, Pires J C, *et al.* Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. *Proc Natl Acad Sci USA*, 2014, **111**(14): 5283–5288
- [124]Hollister J D, Gaut B S. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Research*, 2009, **19**(8): 1419–1428
- [125]Mirouze M, Paszkowski J. Epigenetic contribution to stress adaptation in plants. *Current Opinion in Plant Biology*, 2011, **14**(3): 267–274
- [126]Cortijo S, Wardenaar R, Colome-Tatche M, *et al.* Mapping the epigenetic basis of complex traits. *Science*, 2014, **343**(6175): 1145–1148
- [127]Schmitz R J. Genetics. The secret garden——epigenetic alleles underlie complex traits. *Science*, 2014, **343**(6175): 1082–1083

## Overview of The Function of Transposable Elements in Higher Plants\*

CUI Xie-Kui, CAO Xiao-Feng\*\*

(State Key Laboratory of Plant Genomics and National Center for Plant Gene Research, Institute of Genetics  
and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China)

**Abstract** Transposable element (TE) is a fragment of DNA sequences that can move and integrate into new sites to proliferate itself in the host genome. Based on this proliferation, TEs have been considered as "junk" or "selfish DNA". Most research has focused on investigating mechanisms of TE amplification or silencing whereas the regulatory function of TEs has not been completely unraveled. Previous studies have suggested a positive correlation between genome size and TE content, which provides a basis for the interpretation of the C-value paradox. Emerging studies on the regulatory functions of TEs in the host genome have updated the concept of TEs as junk DNA. In plants, especially in crops, these studies have demonstrated that TEs can reshape the transcription of adjacent genes in *cis* or *trans*, consistent with the first characterization of TEs as "controlling elements". TEs can also create epialleles, which confer stress-inducible gene expression that can increase the sessile plant's environmental fitness. In this review, we summarize recent progress on functional studies of TEs and their significance for crop breeding in the future.

**Key words** transposons, retrotransposons, controlling elements, epialleles

**DOI:** 10.16476/j.pibb.2015.0328

---

\* This work was supported by grants from The National Natural Science Foundation of China (2015ZX08010-002) and the 55th Post-Doctoral Fellowship (2014M550874).

\*\*Corresponding author.

Tel: 86-10-64869203, E-mail: xfcao@genetics.ac.cn

Received: October 19, 2015 Accepted: October 25, 2015