

# 果蝇转座因子对基因组进化的影响

解生勇

(中国农业大学动物科技学院 北京 100094)

**摘要:** 真核生物基因组含有很多可移动 DNA 片段称为转座因子。果蝇是大量系统研究的最好实验材料之一,其基因组的 10%~12% 是由转座因子组成。在宿主中,TEs 也许改变基因表达模型,也许改变 ORFs 编码序列,也许对细胞功能产生影响。这些因子遗传的可动性也可能使它们适于建造载体产生转基因生物。因此,对 TEs 进化的动态研究以及对宿主基因组进化影响的探索将有助于用 TEs 作为载体的细胞工程研究。

**关键词:** 果蝇;转座因子

**中图分类号:** Q953 **文献标识码:** A **文章编号:** 0250-3263(2001)01-38-05

---

**作者简介** 解生勇,男,57岁,副教授;研究方向:遗传学及分子细胞遗传学;E-mail: XieF@mail.cau.edu.cn

**收稿日期:** 2000-03-06, **修回日期:** 2000-08-30

# The Impact of Transposable Elements on Genome Evolution in *Drosophila*

XIE Sheng-Yong

(College of Animal Science and Technology, China Agricultural University Beijing 100094, China)

**Abstract:** Eukaryotic genomes harbour a considerable fraction of mobile DNA that are called transposable elements (TEs). *Drosophila* is one of the best experimental material. Its genome consists of 10% ~ 12% transposable elements. In the host, TEs may alter genes expression pattern, change ORFs coding sequence, influence cellular functions. The genetic mobility of these elements can also make them suitable for the construction of vectors to create transgenic organisms. Therefore, the study on the evolutionary dynamics of TEs and the probe on the evolutionary impact of the host genome should be helpful in a cell engineering study of the use TEs as vectors.

**Key words:** *Drosophila*; Transposable element

在很多动物中突变的主要原因是转座因子的插入,其结果是宿主基因组从点突变到染色体重排。虽然多数转座因子及其突变在其宿主中几乎都产生阴性结果,但转座因子的变化对其宿主基因组的进化将会产生重要影响。也许改变基因表达模型,也许改变 ORFs 的编码序列形成新基因,也许对细胞功能产生影响。在真核生物中存在的这一重要的可移动 DNA 片段,果蝇 (*Drosophila*) 是对这一系统研究最好的实验材料之一。这些因子遗传的可动性也使它们适于建造载体产生转基因生物。若如此,载体导入外源基因的遗传稳定性以及不产生负效应的靶位定位将是重要的研究课题。因此对转座因子进化的动态观察以及对其宿主基因组进化影响的探索也许有助于这一系统研究。

## 1 转座因子的结构进化

转座因子与不动的基因相比有很多偶然性成为整合到一新基因组环境,由于它们的可动性可能散布到新宿主的基因库中,与转座突变一起引起各方面的影响。这些转座因子在基因组进化上所担负的这一作用在不同程度上成为一种变异附加的源泉。

转座因子家系的进化可分为连续的三个阶段:即动态复制、失活和退化。对这一进化的最好实例是果蝇不同谱系的 P 转座因子从 *D. willistoni* 到黑腹果蝇 (*D. melanogaster*) 的转移。即 P 转座因子通过水平感染侵入新基因库 (*D. melanogaster*),继之以在新基因库进入活化转座和迅速增殖阶段。调节机制的作用致使转座因子固定失活。然而,只要有功能的重要序列仍

是完整的,抑制的丧失能导致休眠的转座因子复活,从而起始增殖和在基因组可动性的新阶段。经过长期固定失活后,转座因子将储积有害突变。若有功能重要序列片段损伤将引起永久的固定失活,最后转座因子从基因组中消失退化。只有通过水平逃脱到未感染的基因库中,起始新周期或通过分子驯化稳定的整合到基因组与功能的饰变相结合方能避开这种命运。所以转座因子若保持长期的休眠维持在基因库中,则要设法捕获宿主的控制机制获得新的功能,在相互适应过程中稳定地整合到基因组,成为有用的组分<sup>[1]</sup>。这在进化上比罕见有利点突变以及随机突变引起的转座突变更有价值。

可见,转座因子结构的进化重要的是其转座因子被定位在基因组。一旦定位于基因组,则多数突变起因于转座因子。例如,转座因子插入到编码区,则是有害的,尽管有利突变也会发生。若如此,则选择会迅速加以淘汰。这已通过 *D. melanogaster* 基因组区域的限制图分析得以证实。若转座因子插入到非编码区,则只有较小影响,选择几乎减少到零。显然,即使转座因子在异染色质区插入频繁发生,然而其重组频率很低,插入异染色质区的转座因子是失活的<sup>[3]</sup>。例如插入黑腹果蝇 R 品系异染色质中缺陷的 I 因子是没有长末端重复失活的反转录转座子。缺陷的 I 因子主要退化方式是积累核苷酸代换和插入缺失。多数插入缺失是影响 I 因子有功能的 ORF2。该处是编码一种具有重要同源性的病毒反转录酶蛋白质。该序列变化范围从一个碱基对到几千个碱基对。然而,全面相似的退化的 I

因子仍超过 90% 与它们相似的有活性 1 因子处于同一数量级。

另外,从黑腹果蝇 X 染色体末端异染色质中发现的 SCLR 重复<sup>[6]</sup>。经克隆 SCLR 重复发现不同于相互间插入、缺失、点突变以及其中整合丢失功能的可动因子。显现了转座因子结构的进化而稳定的整合在宿主基因组中。SCLR 重复长 60 kbp。经分析发现,多数 SCLR 重复像是由 copia 反转录转座子 aurora,mdgl<sup>-het</sup>, GATE 和 LINE 组分 G 以及 I 型核糖体插入组成,少数是由异染色质特有的 stellate 基因的变异体和 rDNA 片段组成。说明 SCLR 重复像是可动因子插入 stellate 和 rDNA 基因,加之内部重复导致串联排列,随后的重组使 stellate 和 rDNA 混合重复以及不对等姊妹染色单体交换导致不同品系间 SCLR 重复变异的结果。

P 转座因子是侵入 *D. melanogaster* 基因组的 DNA 转座子。经研究发现<sup>[8]</sup>,在其相关的 *D. subobscura* 等物种中,存在于异染色质的 P 序列比存在于常染色质的 P 序列高度退化。序列的相似性只有 85%~87%,并且只存在于 425 bp 和 37 bp 区域内外显子 1 和内含子 2 中。而多数显现长度的多态性和碱基替换的插入和缺失。而果蝇中的 Bari-1 因子多数存在于 *D. melanogaster* 的 2 号染色体着丝粒异染色质 h39 区,并呈现非编码卫星 DNA 的串联结构。少数 Bari-1 因子存在于常染色质并形成不同品系。经检测<sup>[9]</sup>,常染色质的 Bari-1 编码一种 339 个氨基酸的类似于 Tc1 转座子的转座酶,而异染色质的 Bari-1 因子转座酶的 ORFs 并未退化,仍具有 Bari-1 转座子的功能。可见,活性转座子也可能发生在异染色质区,表现出转座因子的结构进化。

果蝇 412 因子是 7.2 kb 反转录转座子具有两个 ORFs。它的序列在结构上类似于 gypsy、17.6 和 297 因子,并具有很长的 ORF2,该读框可能起因于 pol 和 env 基因的融合。研究表明,412 反转录转座子在 *D. melanogaster* 品系中多数具有大小完整的拷贝,而 *D. simulans* 多数具有重排拷贝。通过限制酶分析 *D. melanogaster* 和 *D. simulans* 自然群体中反转录转座子 412 的结构发现,虽然在两个物种之间在宿主基因组的常染色质和异染色质区保存有相同的 12 因子的标准结构,但是在很多重排拷贝的群体中存在高度的结构多态性<sup>[10]</sup>。在果蝇 P 和 hobo 转座因子中也存在这种共有序列的多态现象。

## 2 转座因子对宿主基因表达的调节

50 年前,当 B. McClintock 首次发现转座因子时就

指出,这种序列可能起基因表达的调控作用。后来人们又提出,在进化过程中转座因子引起的突变也可能对宿主表达模型的持久变化或修饰起着重要作用,这一设想被新近的发现得以证实。

据报道<sup>[12]</sup>,对黑腹果蝇正常发育期间含 LTR 反转录转座子的 15 个不同家系,如 copia 因子的 17.6,297,412,gypsy,blood,copia,mdgl,mdg3,roo 等,对每一类型转座因子在不同野生型品系染色体插入位置的分析结果证实,15 个转座因子家系的每一个都显现出胚胎发育期间时间和空间表达的典型模型。虽然这些品系的插入位置有很大不同,但这些特异模型得到严格保留。可以推断,每一转座子内部含顺式调节序列,它们与特异的宿主编码的转录因子相互影响。例如,黑腹果蝇反转录转座子 412 其顺式调节序列位于 LTR<sub>2</sub> 内,通过体内外试验证明,LTR 特异的调节序列和 ubx 基因编码的同源框蛋白 UBX 之间直接的 DNA-蛋白质相互作用进行特异调节。类似的情况,转座因子 copia 所拥有的同源蛋白质结合位置位于 LTR 内,靠近增强子。

另据实验证明,组织特异表达模型是分配在各个反转录转座子家系中。例如,转座因子 297,mdg3 和 copia 是在胚胎发育晚期的中枢神经系统中表达,而 blood,roo 和 gypsy 是在早期胚胎的卵黄核中表达。说明果蝇基因组中的反转录转座子富含全部的转录调节因子,在宿主基因的两侧潜在比较新的时间和空间模型。因此,在重复基因的顺式调节区域转座因子诱发的突变对于它们功能多样化比编码序列突变也许更重要。据报道<sup>[14]</sup>,果蝇的配对控制基因(paired,PRD),部分极性基因(segment-polarity,gooseberry,GSB)和神经特异基因(neural specific,gooseberry neuro,GSBN)在胚胎发育期间有独特的发育功能。其相应蛋白质 PRD,GSB 和 GSBN 起转录因子的作用,其 N 端一半很保守,而 C 端区域及其顺式调节区域决定其特异的时间和空间表达,显现高度的变异。经证实,这三种蛋白质的调节功能可互换。例如,将 gsb 基因调节区与 prd 基因蛋白质编码区融合可使 gsb<sup>-</sup> 基因表达。说明 PRD 蛋白取代了 GSB 蛋白的作用从而调节 gsb<sup>-</sup> 控制区的表达。另外,对三种蛋白质的足迹分析证实,三种蛋白质都准确地结合到同一富含(A+T)DNA 靶位。它们之间序列的相似性及其 DNA 结合功能的保守性表明,是从一个祖先的单拷贝基因经过反复的重复进化而来。原始基因的重复编码序列可置成新的顺式调节因子进而改变其表达模型并从而获得新的场所和组织特异功能。可见,转座因子能通过提供顺式调节区域改变表达模型使基因功能多样化。

### 3 转座因子对宿主编码序列的改变<sup>[15]</sup>

转座因子中含 P 因子家系是典型的分散重复序列。可是,与其密切相关的物种 *D. guanche*, *D. subobscura* 以及 *D. madeirensis* 只含不动 P 因子衍生物,它们成串分布在单个染色体上<sup>[16]</sup>。虽然这些不完整的转座因子衍生物丢失其末端和一个外显子,但仍能保持像 P-阻遏物 66 ku 蛋白质的编码能力。RNA 印迹法分析表明,在 *D. guanche* 和 *D. subobscura* 两者中具有 2 100 b 的转录物。而且,*D. guanche* 的 P 同源染色体甚至按时期和组织特异模型转录。ORFs 和转录活性的保存说明这些序列已采用新功能。因为这些 P 同源染色体已丢失包括起始启动子序列的两个末端,它们的控制区已遭到破坏。这一间隔区位于不同丢失片段的可动 P 转座因子串联排列的相邻拷贝之间。从 *D. guanche* 和 *D. subobscura* 独立 P 同源染色体非编码上游区的序列分析表明,存在两种不同类型的成串单位,称 A 型和 B 型。虽然这两种类型具有典型的真核生物转录起始的基本特征,即具有 GC 框、八聚体基本结构、CAAT 框和 TATA 框,不存在典型的 P 因子,有不同的进化起源。但是 A 型启动子是由足迹组成至少有两个不同转座因子插入,一个序列表现出与 *D. miranda* 的 ISY3, LINE 插入序列有很强的相似性,另一个类似 *D. subobscura* 的分散重复因子(UDRE-812)的反向重复。192 bp 片段类似于 ISY3 的亚区域包括推定的转录因子的结合位置。即 GC 框、CAAT 框和启动子相联系的八聚体基本结构 ATTTGCAT。由 UDRE-812 反向重复延伸的 179 bp 同源片段装置有 TATA 框。因此,获得活性的启动子可能导致不同表达模型产生一新功能的蛋白质。这是果蝇野生型基因中转座因子衍生启动子的例证。不仅如此,甚至编码区都不是源于宿主基因而是 P 转座因子。通过 *D. guanche* 和 *D. subobscura* 比较表明,P 因子衍生的有功能的阻遏蛋白其氨基酸序列是由各自成串单位编码。形成的一串基因是由分散的重复单位组成,每一基因由三个不同转座因子组成。表明由形成的可动因子到调节因子功能的转变是通过宿主基因组到综合调节网络的结合。

### 4 转座因子对细胞功能的影响

研究表明,多数真核细胞染色体异质端粒 DNA 含 2~8 bp 的串联重复序列,称末端重复或端粒重复<sup>[17]</sup>。对不同谱系果蝇异质端粒的检测表明,末端序列的丢失可通过转座因子的转座致使重复序列增加来补偿。例如,果蝇的 HeT-A 序列即能自发转座到破裂的染色

体末端。

HeT-A 因子<sup>[18]</sup>是 LINE 家族无 LTR,其 3'端有典型寡 A 区的反转录转座子。它有两个开放读框,分别编码 gag 蛋白质和反转录酶。在自然群体的果蝇端粒中,还发现另一种端粒特征象 LINE 因子的 TART 序列<sup>[19]</sup>。HeT-A 和 TART 两者只 ORF2 相似,同属于 LINE 一类的反转录转座子。据报道,在 TART 因子近侧另有两个 HeT-A 和一个额外的 TART 序列的转座因子插入。这些转座因子都依同一的 5'-3'方位而有序地截去它们的 5'末端。通过用 TART 因子作为探针与其多线染色体的原位杂交表明,在果蝇染色体末端串联排列的两个不同类型的 LINE 因子是起因于 4 个转座因子到末端的连续转座。这一对准方位的转座表明,果蝇端粒其中即是通过这一基本过程来维持。经测定,HeT-A 因子在 X 染色体上的转座频率每世代为 0.7%~1.0%。而渐次端粒 DNA 的丢失是由于染色体末端以每世代 70~80 bp 的不完全复制。所以,端粒 DNA 的丢失是通过平均约 7 kb 的一个 LINE 因子以每世代 0.7%~1.0%的转座导致每一染色体末端 50~70 bp 的净增加而达到动态平衡。可见,这种特定转座因子的转座频率必须适应端粒丢失的相应平均频率以保证染色体大小的恒定。从而显现出,宿主基因组的可动性及其与可移动因子之间精细的调节系统对细胞功能进化的影响。

### 参 考 文 献

- [1] Von Sternberg, R. M., G. E. Novick, G. P. Gao *et al.* Genome canalization: The coevolution of transposable and interspersed repetitive elements with single copy DNA. *Genetica*, 1992, **86**: 215~246.
- [2] Voelke, R. A., J. Graves, W. Gibson *et al.* Mobile element insertions causing mutation in the *Drosophila* suppressor of sable locus occur in DNase I hypersensitive subregions of 5'-transcribed nontranslated sequences. *Genetics*, 1990, **126**: 1 071~1 082.
- [3] Charlesworth, B., P. Sniegowski, W. Stepkan. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*, 1994, **371**: 215~220.
- [4] Houck, M. A., J. B. Clark, K. R. Peterson *et al.* Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. *Science*, 1991, **253**: 1 125~1 129.
- [5] Steinemann, M., S. Steinemann. A duplication including the Y allele of Lcp2 and the TRIM retrotransposon at the Lcp locus on the degenerating neo-Y-Chromosome of *Drosophila miranda*: Molecular structure and mecha-

- nisms by which it may have arisen. *Genetics*, 1993, **134**: 497~505.
- [ 6 ] Nurmisky, D. I., Y. Y. Shevelyov, S. V. Nuzhdin *et al.* Structure, molecular evolution and maintenance of copy number of extend repeated structures in the X-heterochromatin of *Drosophila melanogaster*. *Chromosoma*, 1994, **103**: 277~285.
- [ 7 ] Calvi, B., T. J. Hong, S. D. Findley *et al.* Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: hobo, Activator, and Tam3. *Cell*, 1991, **66**: 465~471.
- [ 8 ] Paricio, N., M. J. Martinez-Sebastian, R. De frutos. A heterochromatic P sequence in the *D. subobscura* genome. *Genetica*, 1994, **92**: 177~186.
- [ 9 ] Caizzi, R., C. Caggese, S. Pimpinelli. Bari-1 a new transposon-like family in *D. melanogaster* with a unique heterochromatic organization. *Genetics*, 1993, **37**: 675~689.
- [ 10 ] Cizeron, G., C. Biemont. Polymorphism in structure of the retrotransposable element 412 in *Drosophila simulans* and *D. melanogaster* population. *Gene*, 1999, **232**: 183~190.
- [ 11 ] Brookfield, J. F. Y. Models for repression of transposition in P-M hybrid dysgenesis by P Cytotype and by zygotically encoded repressor protein. *Genetics*, 1991, **128**: 471~486.
- [ 12 ] Daniels, S. B., K. R. Peterson, L. D. Strausbaugh *et al.* Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics*, 1990, **124**: 339~355.
- [ 13 ] Coen, D. P-element regulatory products enhance zeste repression of P [ white<sup>duplicate</sup> ] transgene in *Drosophila melanogaster*. *Genetics*, 1990, **126**: 949~960.
- [ 14 ] Li, X., M. Noll. Evolution of distinct developmental function of three *Drosophila* genes by acquisition of different cis-regulatory regions. *Nature*, 1994, **367**: 83~87.
- [ 15 ] Miller, W. J., N. Paricio, S. Hagemann *et al.* Structure and expression of the clustered P element homologues in *Drosophila subobscura* and *D. guanche*. *Gene*, 1995, **156**: 167~174.
- [ 16 ] Paricio, N., M. J. Martinez-Sebastian, R. DeFrutos. A heterochromatic P sequence in the *D. subobscura* genome. *Genetica*, 1994, **92**: 177~186.
- [ 17 ] Blackburn, E. H. Structure and function of telomeres. *Nature*, 1991, **350**: 569~573.
- [ 18 ] Biessmann, H., B. Kasravi, K. Jakes *et al.* The genomic organization of HeT-A retroposons in *D. melanogaster*. *Chromosoma*, 1993, **102**: 297~305.
- [ 19 ] Levis, R. W., R. Ganesan, K. Houtchens *et al.* Transposon in place of telomeric repeats at a *Drosophila* telomere. *Cell*, 1993, **75**: 1 083~1 093.