

11

转座因子的遗传分析

20世纪40至50年代 **B. McClintock** 通过对玉米花斑糊粉层和植株色素形成的遗传研究,发现色素的变化与一系列染色体断裂重组有关,并首次报道了在基因组的不同区域内存在可移动的遗传因子(**mobile genetic element**)。可是由于当时人们受基因固定排列于染色体的传统观念的束缚,加之分析手段的限制,因此这一划时代的发现在当时几乎不能被科学界所接受。直到20世纪70年代,在大肠杆菌半乳糖操纵子的突变型研究中第一次在细菌中发现了可转移座位的插入序列(**insertion sequence, IS**),之后才被重视,此后近30年来在许多生物中也发现各种类型的可移动的遗传因子,并在分子水平上得到证实。转座因子不像某类噬菌体或质粒那样,既可插入基因组,又可采取独立形式而存在。转座因子是从基因组中一个位置直接移动到另一个位置,而且它不依赖于供体位点与受体位点之间的任何关系。转座是基因组突变与进化的主要来源之一,也是表观遗传的信号与调节因子。因此,可转座遗传因子的发现是遗传学发展史上重要的里程碑之一,**McClintock** 因而获得1983年诺贝尔奖。本章仅就这30余年研究成果与进展讨论转座因子(**transposable element**)——一类可改变位置的遗传因子的总称,其分类与结构特征,原核生物的转座因子与真核生物的转座因子,转座作用的机制和遗传学效应。

11.1 转座因子的发现与分类

11.1.1 转座因子的发现

早在 20 世纪初, Emerson 于 1914 年在研究玉米果皮色素遗传的过程中, 发现一种花斑果皮的突变类型。这种突变可发生多次回复突变, 从而产生宽窄不同、红白相间的花斑。在金鱼草植株叶片以至花瓣上都可见花斑表型。他意识到这种花斑产生在于突变基因的不稳定性, 但如何不稳定不得其解。到 1938 年 Rhoades 在研究玉米籽粒糊粉层色素遗传时, 发现有色籽粒纯种自花授粉的后代中, 表现出一种意外的修饰的孟德尔分离比: 有色: 斑点: 白色 = 12: 3: 1, 显然这两个基因是不连锁的。他认为控制色素的基因 A_1 突变为等位基因 a 时表现为无色; 另一基因是 D (斑点) 表型为有色斑点。这样原品系的基因型为 $A_1 A_1 dtdt$ 突变后产生了 $A_1 a Dtdt$ 的植株, 这种双突变植株自交就产生了上述比例(图 11-1)。

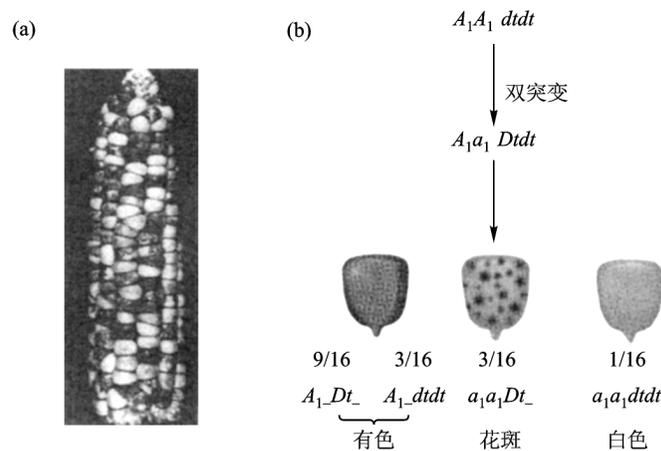


图 11-1 玉米花斑表型的遗传学解释

(a) 由于转座因子而引起的玉米籽粒花斑的表型 (b) 最初认为玉米籽粒花斑形成是双突变的结果

但是什么因素导致或产生花斑呢? 一种可能是在体细胞中产生了回复突变 $a \rightarrow A_1$, 但大量的斑点需要很高频率的回复突变。Rhoades 在 $a a D _$ (花斑) 特殊无性生殖植物花中找到相应的花药, 其花粉应携带回复突变产生色素的基因, 而他用这些花粉与 $a a$ 的植株测交, 结果有的后代完全是有颜色的。表明在亲本中每个斑点实际上是回复突变的表型效应。这样 a 成为首次发现的不稳定突变等位基因的例子 即一种回复突变率很高的基因。然而这种等位基因的不稳定性取决于不连锁 D 基因的存在。一旦回复突变发生, 它们就变得稳定了; 即 D 基因能离开 A_1 基因, 这时 A_1 的表型不再改变。 a 的表型是由一个缺陷型转座因子的插入而产生的, 而缺陷的转座因子自身不能移动。因而 D 的缺乏使得表型保持稳定。显然 Rhoades 这时已发现了某些基因的不稳定性, 而且这种不稳定性是由另一个独立的因子所控制。但仍未揭示这种不稳定性的遗传学机制, 也缺乏实验证据。

M cC linto ck 于 1940 年至 1950 年在美国康奈尔大学和冷泉港实验室工作期间, 研究玉米花斑糊粉层和植株色素产生的遗传基础时, 也发现玉米籽粒上有色素斑点的变化。当时已知许多基因共同控制红色花青素的合成使玉米胚乳成紫色。这些基因中任何一对基因发生突变都会影响色素的合成使

胚乳呈白色。人们已经发现控制玉米糊粉层颜色至少有 5 对相关基因:① A 代表花色素 (anthocyan), 如 A 变为 a 则什么颜色都不会产生。② C 代表颜色 (color), 决定红色和紫色的发生。③ R 代表红色 (red), 其作用是以 A、C 两基因的存在为先决条件。④ Pr 代表紫色 (purple), 它是在红色基础上辅助成紫色的, 没有 A、C、R 基因存在, 即使它是显性也不能使玉米籽粒产生任何颜色。⑤ I 称为抑制基因 (inhibitor), 它抑制 C 基因的作用, C 基因如被抑制, 则红色与紫色都不会形成。

M cC lintock 研究了玉米胚乳的紫色、白色以及白色背景上带有紫色斑点这些表型之间的相互关系。她发现花斑表型是不稳定的, 并根据自己的遗传学和细胞学研究结果推断“花斑”这种表型并不是一般的基因突变产生的, 而是由于一种控制因子的存在所导致的。她认为若玉米带有野生型 C 基因, 则胚乳呈紫色, C 基因的突变阻断了紫色素的合成, 那么胚乳呈白色。在胚乳发育的过程中, 突变发生回复导致斑点的产生。而回复突变的遗传学性质受到这样一种事实的支持, 即细胞的后代经过回复突变也能产生色素。回复突变发生在早期发育阶段, 紫斑就比较大。M cC lintock 认为原来的 C 突变(无色)是由一个“可移动的遗传因子”, 当时称为抑制基因, 即 I 基因, 现在称为 Ds 即解离因子 (dissociator)。它可以插入到 C 基因中。另一个可移动的因子是 Ac 称激活因子 (activator), 它的存在激活 Ds 转座进入 C 基因或其他基因中, 也能使 Ds 从基因中转出, 使突变基因回复, 这就是著名的 Ac-Ds 系统。M cC lintock 还发现 Ds 可导致所在位置的染色体断裂, 这种断裂可以通过细胞学和遗传学的方法加以检测。Ds 存在的玉米的 9 号染色体的一条臂上, 该染色体一端带有结节 (knob), 这一特征性结构极易辨认, 在 Ds 处易发生断裂 (图 11-2)。当 Ds 插入 C 基因后籽粒为无色, 当 Ac 激活 Ds 从 C 基因切离后, 则籽粒出现有色斑点, 而且斑点大小取决于产生它的细胞分裂次数。

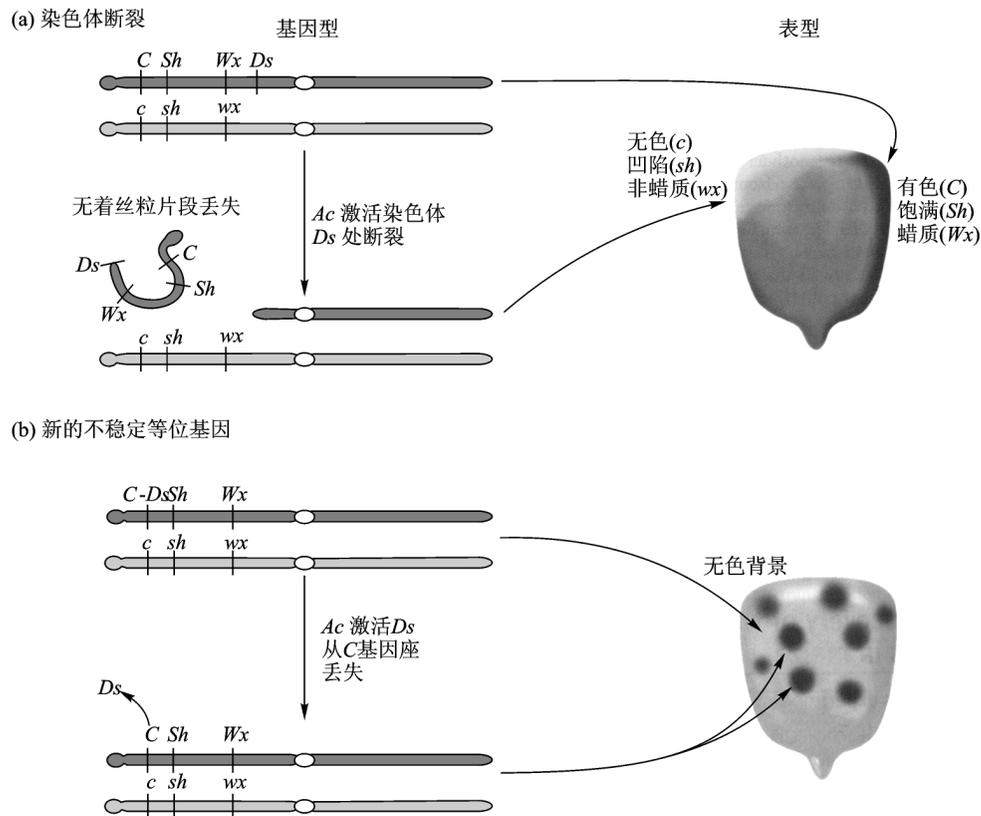


图 11-2 Ds 转座导致籽粒斑点表型 (引自 Griffiths 等, 2005)

(a) Ds 位于杂合的一条同源染色体的着丝粒与一系列显性标记之间, 另一条同源染色体缺乏 Ds 而有几个隐性基因, 在 Ds 处断裂产生携带显性标记的无着丝粒片段而丢失, 故同源染色体上隐性基因表达, 籽粒产生无色扇形 (b) Ds 插入 C 基因后籽粒为无色; 当 Ac 激活 Ds 从 C 基因切离后, 则籽粒出现有色斑点, 而且斑点的大小取决于产生它的细胞的分裂次数

染色体的断裂结果还会形成断裂融合桥(breakage fusion bridge)。这是由于 Ds的存在可导致其所在位点的染色体断裂后丢失端粒。经复制后缺少端粒的染色体彼此连结,形成双着丝粒染色体,在有丝分裂中两个着丝粒彼此分离向两极移动形成染色体桥。桥的断裂使一端产生重复,一端产生缺失。这样周而复始形成断裂融合桥循环(breakage fusion bridge cycle)。

M cC lintock根据大量遗传学和细胞学研究结果,于 1951年提出了生物的基因组中存在转座因子学说。这些转座因子既可以沿染色体移动,也可以在不同染色体之间跳跃。因此转座因子又可称为跳跃基因(jumping gene)。这是遗传学发展史中划时代的重大发现,将基因概念向前推进了一大步。但这项划时代的成果并未受到当时同行们重视。

直到 20世纪 60年代 P. A. Jacob和 L. M onod的乳糖操纵子模型和基因调控理论发表后,特别是 J Shapiro在细菌中也发现了可转座的遗传因子后,这一成果才被接受。Shapiro在大肠杆菌半乳糖操纵子中发现了半乳糖突变基因 gal^- ,由于 λ 噬菌体的整合位置在 gal^- 基因的旁边,易于得到带有 gal^- 基因的转导噬菌体 $\lambda dgal^-$ 。实践证明,这种突变不能被核酸置换所回复,说明它不是一般的点突变。又由于这种突变可回复,因此也不是缺失突变。最终通过密度梯度离心实验证明 $\lambda dgal^-$ 的密度大于 $\lambda dgal^+$ 的密度(图 11-3)。进一步将两者 DNA变性并相互复性,在电镜下可观察到一部分杂合双链上出现一个多余的 DNA 环,从而证明这种突变体是由于 DNA 片段插入而产生的,这一插入序列是最先发现的最小的一种转座因子,称为 IS I。IS 序列就是在细菌中首次发现的跳跃基因,并在分子水平上得到证实。至此,转座因子的概念才被遗传学界所公认。接着在利用细菌的抗药性质粒(R 质粒)构建载有卡拉霉素(kanamycin)抗性基因的 λ 噬菌体的过程中,发现抗药性基因的传播非常快,它可以从一种 R 质粒跳跃到另一质粒上,也可以跳跃到另一些噬菌体或细菌染色体上。根据 DNA 分子杂交,与卡拉霉素抗性有关的 DNA 序列长 5.2 kb,两端有长 1.5 kb 的反向重复序列,该结构称为末端反向重复序列(inverted terminal repeat)。当抗药性基因跳跃时,反向重复序列也随之转移。具有这种结构和特性的 DNA 序列就称作转座子(transposon, $T\alpha$ 或 Tn 因子,如 Tn3 和 Tn5 等。至此,几乎在各类生物基因组中都发现有不同类型的转座因子。

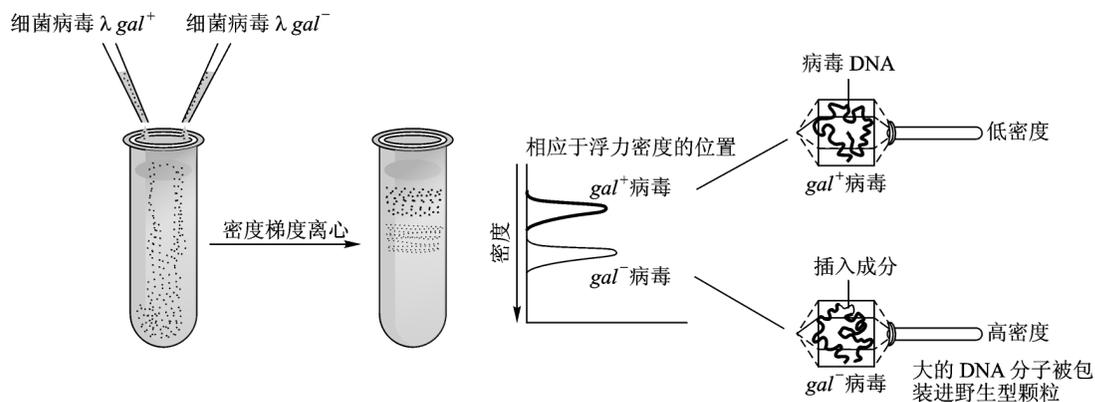


图 11-3 外源 DNA 片段插入引起基因失活

知识窗 11-1

Barbara M cC lintock, 转座因子的发现者

B M cC lintock 1902年出生于新英格兰,生长在纽约,就读于康奈尔大学,获博士学位后又在本校著名的 R. A. Emerson 玉米遗传学研究室工作,与 George Beadle, R. A. Emerson, Charles Burnham 和 Marcus Rhoades 等合作。他们对玉米的遗传,包括籽粒颜色变异的遗传基础做了大量而深入的研究。

当时没有分子生物学技术,而是采用细胞遗传学的方法对细胞的有丝分裂和减数分裂过程中染色体行为,包括重组和染色体断裂、倒位等做了深入细致的工作。在该项研究中 M cC lintock起着重要的作用,她熟练的细胞遗传学分析技术可准确地鉴定玉米的 10条染色体和相应的连锁群,率先提出染色体交换机制和核仁起源等观点。然而她最著名的成果是转座因子的遗传性质的研究。她到纽约冷泉港 Carnegie 研究所遗传学研究室工作期间仍然坚持不懈地研究追踪转座子介导玉米籽粒颜色变异和染色体断裂融合桥形成之间关系, M cC lintock自 1948年发表了第一篇论文,随后又发表了几篇 其中最主要的一篇 1951年刊登在 Cold Spring Harbor Symposium on Quantitative Biology。只是由于转座因子的概念与当时基因在染色体上具有固定位置的传统观点相互矛盾,因而基因可以转座的观点未被同行接受。何况当时也没有人在其他生物中发现有转座现象。直到 20世纪 60至 70年代,在细菌和果蝇中都发现有转座因子后从而改变了这一状况,科学界意识到 M cC lintock所提出转座因子的重大意义,属于划时代的成果。因而在 1983年,也就是她发表关于转座因子的第一篇论文后的 35年获得诺贝尔奖。M cC lintock的伟大在于她的坚韧不拔的科研精神,顽强拼搏的工作态度,扎实而富于创新的思维方法,精湛而熟练的细胞遗传学技术。她的这种为科学事业而献身的精神是我们学习的榜样。

11.1.2 DNA 转座

可移动位置的遗传因子包括两种类型:一种类型是直接以 DNA 序列某些区段作为转座成分的,如 M cC lintock在玉米中发现的转座因子和 Shapiro在大肠杆菌半乳糖操纵子中发现的 IS都是涉及 DNA直接转座,因而称为 DNA转座;另一种类型是以 RNA介导进行转座,称为反(逆)转座子(retroposon)或反转录转座子(retrotransposon)。首先讨论 DNA转座,根据转座机制, DNA转座可分为 3种类型:

(1) 复制型转座 (replicative transposition)

在转座反应过程中,转座子被复制,转座的 DNA 序列实际上是原转座子的一个拷贝[图 11-4 (a)]。转座子中移动的部分被拷贝,一个拷贝保留在原位点,而另一份拷贝则插入到新的位点,因此其转座过程伴随着转座子拷贝数的增加。复制型转座涉及两种类型的酶:一是转座酶(transposase),该酶作用在原转座子的末端;另一种为解离酶(resolvase)或称拆分酶,它作用于复制的转座子拷贝。例如与 TnA有关的一组转座子只通过复制型转座机制移动。

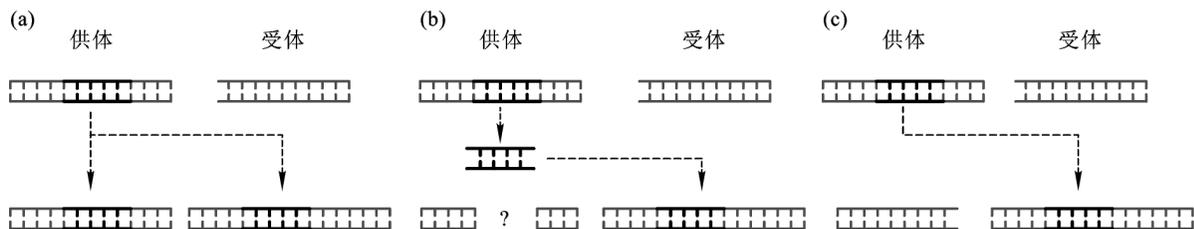


图 11-4 3种不同的 DNA转座类型(引自 Lew in,2004)

- (a) 复制型转座:转座子产生的一份拷贝插入到受体位点,供体位点序列不变,供体和受体位点都有转座子的一份拷贝 (b) 非复制型转座:转座子从供体位点移动到受体位点,在供体位点造成断裂,若断裂不被修复,后果是致死的 (c) 保守型转座:转座只是序列的直接移动,没有核苷酸键的损失,其与 λ 噬菌体的整合和切除机制相似

(2) 非复制型转座 (nonreplicative transposition)

转座因子作为一个物理性实体,直接从一个位点移动到受体 DNA 一个靶位点,并且保守性很强。这可以通过两种机制实现:一种是利用供体与靶 DNA 的连接,并且有些步骤与复制型转座相同。插入序列和复合转座子 Tn10和 Tn5便是利用这种转座机制进行转座的[图 11-4 (b)]。另一种是转座过程中转座子必须从供体 DNA 上释放,这种机制只需要转座酶。非复制型转座的两种机制

皆导致转座子离开供体位点并插入到靶位点。在发生非复制型转座之后,供体分子的存活可能需要宿主修复系统识别双链缺口并对其进行修复。

(3) 保守型转座 (conservative transposition)

这种转座实质是另一种非复制型转座过程。在此过程中,转座元件从供体位点上切除并通过一系列过程插入到靶位点,其所有碱基均被保留[图 11-4 c]。保守型转座与 λ 噬菌体整合机制非常相似,其转座酶与 λ 整合酶也相关。应用保守型转座机制的转座子较大,并且除了介导转座子本身转移外,还能将供体细菌的 DNA 转移到另一种细菌。虽然这种元件最初被归类为转座子,但称为附加体(episome)更加确切。

某些转座子只采用一种转座机制,而另一些转座子可能具备两种途径。如 IS 1 和 IS 903 就采用复制和非复制两种途径。Mu 噬菌体能从共同的中间体转变为两种途径中的一种。还有些转座子存在着不同的转座机制交替使用,据此特征难以划分它们的类型。

DNA 区段无论以什么机制进行转座,其转座过程是有别于同源重组过程。这些 DNA 序列的转座往往发生在非同源序列之间,也不需要像细菌同源重组中 Rec A 等蛋白质的参与作用,只依赖于转座区域 DNA 复制、转座酶和特定的反向重复序列等因子而完成重组过程。

11.1.3 反转录转座子

反转座子的转座过程均由 RNA 介导,通过反转录酶将转座子 RNA 拷贝为 cDNA 后再整合到宿主基因组中(图 11-5),因此称为反转座子,包含反转录转座子和反转录病毒(retrovirus)。

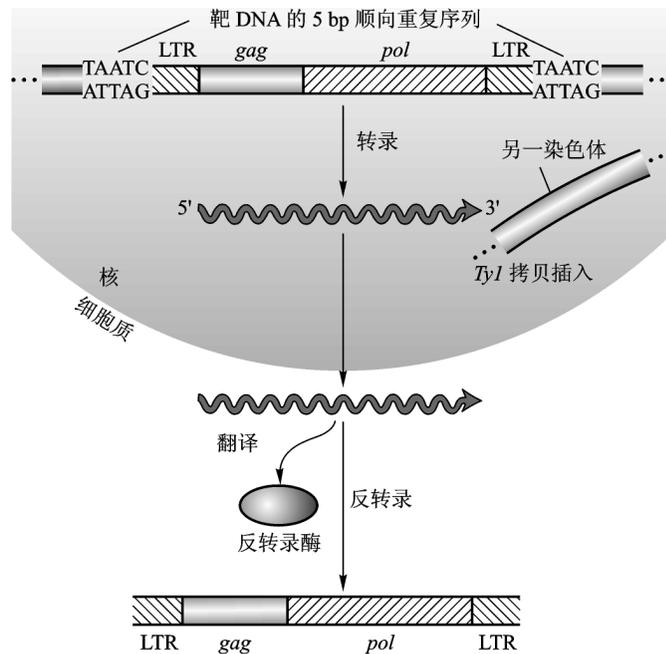


图 11-5 反转录转座子的转座作用(引自 Griffiths 等, 2005)

通过反转录转座子编码的反转录酶将其转录的 RNA 反转录为 DNA 再插入基因组的新位点

反转座子只在真核生物基因组中发现,迄今尚未在原核生物中找到。一些反转录转座子与细胞中游离的反转录病毒基因组之间有许多相似性(图 11-6)。① 反转录病毒的基因组为 RNA 分子,可感染许多脊椎动物,该病毒感染细胞后,由病毒基因组编码的反转录酶将其 RNA 拷贝为 DNA,然后整合到宿主基因组中。新的病毒产生必须由整合的 DNA 转录为 RNA,合成由病毒基因组编码的蛋白质外壳,然后再进行包装[图 11-6 a]。已经整合到脊椎动物染色体中的反转录病毒基因组称为内源性反转录病毒(endogenous retrovirus, ERV)。这些 ERV 有些仍有活性,在活细胞的一定生活时期

可指导内源病毒的合成,但大多数 ERV 已无功能。这些无功能的病毒基因组拷贝虽然分散在基因组中,但不能扩增,人类基因组中约有 10 000 个残缺的内源病毒拷贝。还有一类称为类反转录因子

(retroviral like element RTVL),人类基因组中有近 20 000 拷贝。② 反转录转座子具有类似 ERV 的顺序,只是没有编码外壳蛋白的基因(*env*),故不会包装形成具有蛋白质外壳的颗粒。这类反转录转座子多分布在植物、真菌和无脊椎动物中,脊椎动物中少见。反转录转座子在某些生物基因组中有很多的拷贝,有许多不同的类型,玉米基因组中大多数散在重复序列都是反转录转座子,几乎占基因组一半,它与反转录病毒结构类似,而且都含有负责转座的长末端重复序列(long terminal repeat, LTR) [图 11-6 (b)]。③ 长散在核元件(long interspersed nuclear element LINE)又称长散在重复序列(long interspersed repeated sequence)含有与反转座有关的类反转座基因 [图 11-6 (c)]。人类基因组中的 LINE 1 是典型的长散在核元件,长 6.1 kb,含 3 500 个全

长的拷贝,另外有数万个残缺的拷贝。④ 短散在核元件(short interspersed nuclear element SINE)又称短散在重复序列(short interspersed repeated sequence)本身无反转座酶基因,但可借用宿主中的反转录酶实现转座 [图 11-6 (d)]。人类基因组中最普遍的 SINE 为 Alu,它有上百万份拷贝。最初的 Alu 可能是 7SL RNA 偶尔的反转录产物,并整合到基因组中。因 Alu 拷贝含有内启动子,具有转录活性,细胞可转录大量的 AluRNA,从而提供了该成分大量扩增的机会。

此外,还有只分布在哺乳动物中的哺乳动物散在重复顺序(mammalian interspersed repeat MIR)是起源于 rRNA,在灵长类基因组中的拷贝数估计为 120 000~300 000。由于 rRNA 基因含有内部启动子,如同 Alu 重复序列一样, MIR 也可经反转录大量拷贝整合到基因组中。

反转录病毒只在脊椎动物中发现,它们可能起源于 Gypsy 因子(吉卜赛因子),一种反转录转座子,获得编码衣壳蛋白基因(*env*)后成为可感染动物细胞的反转录病毒。植物中含有除反转录病毒之外的所有反转座子。

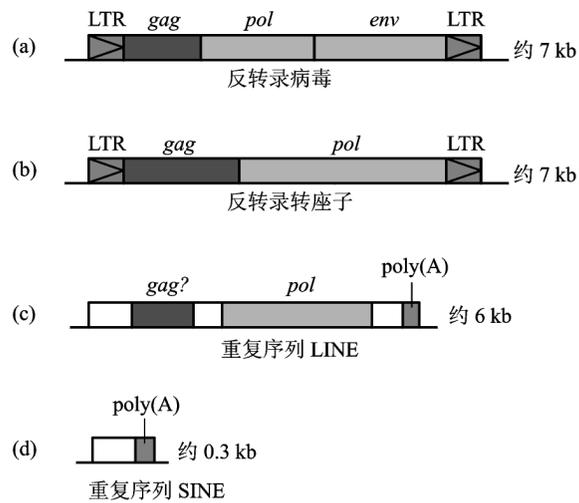


图 11-6 4种反转录转座因子结构比较

反转录病毒(a)和反转录转座子(b)均含有长末端重复序列,归于LTR一类,只是后者缺编码外壳蛋白的基因(*env*)。LINE(c)和SINE(d)为非LTR反转座因子,两者在3端均有poly(A)区

11.2 原核生物中的转座因子

11.2.1 插入序列

根据分子结构和遗传性质可将原核生物中的转座因子分为插入序列(insertion sequence, IS)、转座子(transposon, Tn)、Mu噬菌体(Mu phage)。原核生物中的转座因子是存在于染色体DNA上可自主复制和转座的基本单位,其中最简单的转座因子,不含任何宿主基因的可转位的DNA序列就称为插入序列。它们是细菌染色体或质粒DNA的正常组成部分。如大肠杆菌基因组中约有20种不同的DNA转座因子,都含有编码转座酶的基因。一个细菌细胞常带有多个IS序列。它们都是可以自主转座的单元,因为带有介导自身转座的转座酶。大多IS序列已被鉴定,如IS1的全长为768 bp,它

的两端有 18/23 bp 的反向重复序列。IS 本身没有任何表型效应,只携带和它转座作用有关的基因,称为转座酶基因。它们是一类最小的转座因子,可以从染色体的一个位置转移到另一位置,或从质粒转移到染色体上,这种改变位置的行为称为转座(transposition)。如 F 因子和大肠杆菌的染色体上有一些相同的插入序列,即 IS₂、IS₃等。通过这些同源序列之间的重组,F 因子便插入染色体中成为 Hfr 菌株。目前已知的 IS 至少有 10 余种,如 IS₁、IS₂、IS₃、… ,虽然它们的大小不同,但有某些共同的结构特征。如每种 IS 两端的短核酸序列完全相同,但方向相反,所以称为反向重复序列。已知各种 IS 的长度在 768~5 700 bp 之间。如 IS₁ 的两端 IR 长 23 bp,其中 18 bp 在两端是相同的(图 11-7)。而 IS₂ 的两端是 41 bp 的 IR。

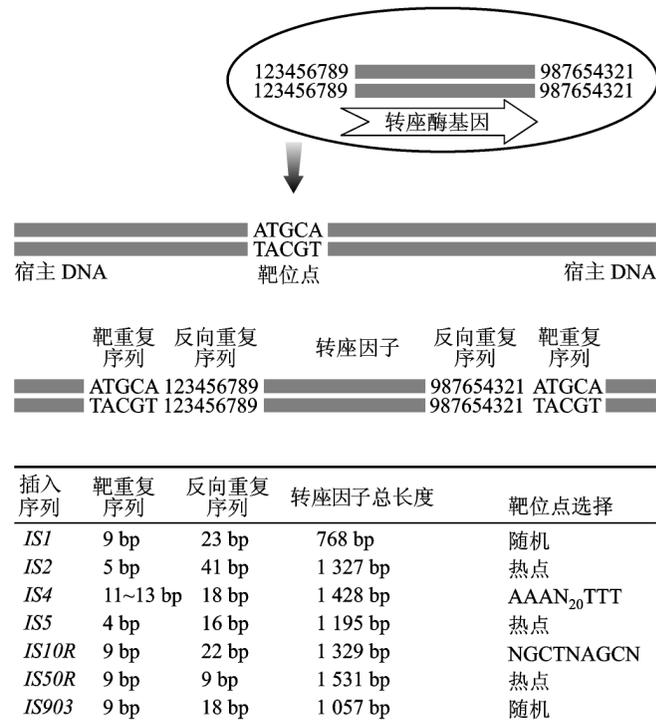


图 11-7 IS 结构模式图(引自 Lewin, 2004)

IS 末端的反向重复序列为 9 bp, 数字 1~9 示碱基序列

所以含有 IS 的质粒经变性后,分别以单链复性,于是在电镜下出现颈环结构,颈的部分是 IS 的 IR,大环是质粒 DNA,小环是 IS 的中间序列(图 11-8)。而且当一个 IS 插入“靶”DNA 后,在插入片段的两端出现一小段靶 DNA 的序列正向重复,每种 IS 插入形成这种正向重复序列的长度是不同的,一般为 5~11 bp。

除 IS₁ 以外,所有已知 IS 序列都只有一个开放阅读框(open reading frame, ORF),翻译起始位点紧接第一个反向重复区,终止点位于第二个反向重复区或重复区附近。IS₁ 含有两个分开的读码框,只有移码通读才能产生功能型转座酶。一般情况下,每个 IS 转座频率是 $10^{-4} \sim 10^{-3}$ /世代,恢复频率则低得多,为 $10^{-7} \sim 10^{-6}$ /世代。

11.2.2 转座子

Tn 是一类较大的转座子,除了含有与转座作用有关基因外,还带有抗药基因以及其他基因,如乳糖发酵基因。因此 Tn 的转座能使宿主菌获得有关基因的特性,已发现有多种 Tn 如 Tn₁、Tn₂、Tn₃ 等。Tn 分子大小一般在 2 000~25 000 bp,两端有相同序列,如 IR。某些 Tn 的 IR 便是已知的 IS,带有 IS 的 Tn 也称为复合转座子(composite transposon)。Tn₅、Tn₁₀ 和 Tn₉₀₃ 都属于一类复合转座子,结

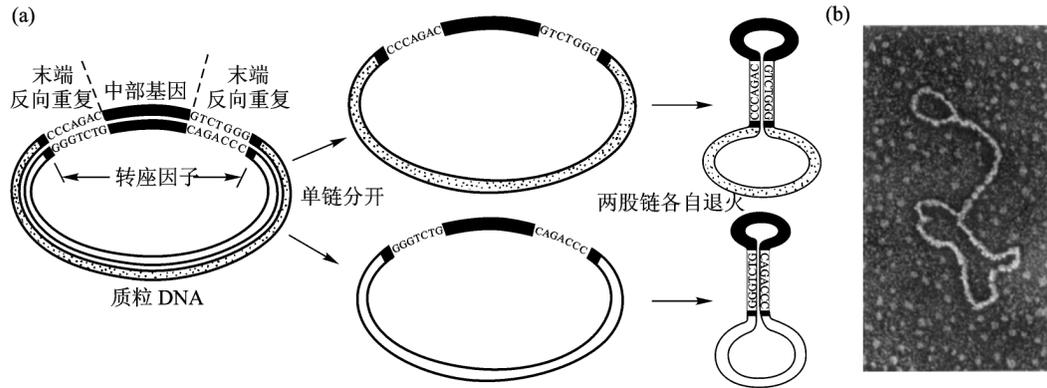


图 11-8 颈环结构的形成

含 IS 质粒经变性复性形成颈环结构 (a) 及其电镜照片 (b), 大环是质粒 DNA, 小环是 IS 的中间序列, 颈的部分是 IS 的 IR

构比较特殊, 其主序列长 2.7 kb 左右, 位于中央, 两个 IS 以相反的极性位于主序列两侧 [图 11-9 (a)]。不含 IS 的 Tn 称为简单转座子 (simple transposon), 如 Tn3 等 [图 11-9 (b)]。

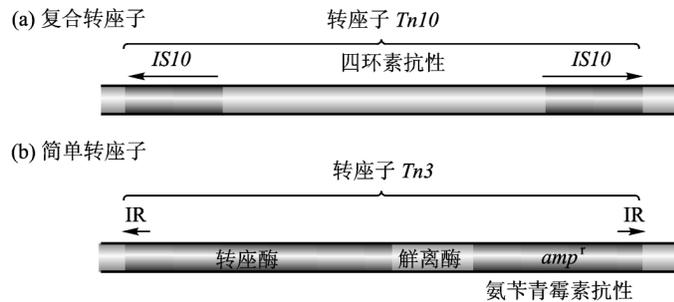


图 11-9 复合转座子与简单转座子的结构 (引自 Griffiths 等, 2005)

(a) 复合转座子, 如 Tn10 含转座酶基因 IS10 以相反方向插入形成 IR (IS10L 不含转座酶基因) (b) 简单转座子, 如 Tn3, 有短 IR, 不含转座酶, Tn3 序列编码自己的转座酶, 解离酶将负责分离共合体为供体和受体

Tn5 主序列含有几个抗生素抗性基因, 包括卡拉霉素 (kanamycin)、链霉素 (streptomycin) 和博来霉素 (bleomycin) 一类的抗生素基因。Tn5 主序列两侧的 IS 既是 Tn5 的重要组成部分, 也是一种自主性的转座因子。位于 Tn5 右侧的 IS50R 编码两种与转座有关的酶, 其中一种 (蛋白 1) 为转座酶, 另一种 (蛋白 2) 为转座抑制蛋白, 这两种蛋白质都由同一段 DNA 序列编码。Tn5 左侧的 IS50L 与右侧的 IS50R 只是一对碱基的差别, 但这一微小差别会产生两种重要影响, ① 在转座酶编码区内 IS50L 处为 CAA (谷氨酰胺) 造成琥珀突变 (amber mutation) 成为终止密码子 UAA, 产生翻译的终止信号, 所以这种突变蛋白不具有转座功能。即 IS50R 含转座酶基因, 而 IS50L 则不含转座酶基因。② 形成一个 Tn5 主区中卡拉霉素抗性基因的启动子 (P2), 所以 IS50L 对主序列 Tn 的转录具有重要影响 (图 11-10)。Tn5 的这种结构特征表明, 两个 IS 夹住一个 Tn5 主序列并促使其转座到新的位点上。Tn5 转座时, 首先从染色体上切割下来, 形成环状结构, 然后再插入到新的位置上。因此, Tn5 是非复制型转座, 只是从一个位点转移到另一个位点, 在原位并不留下一个拷贝。Tn5 转座到新的位点上后, 两端都形成一段同向重复 (direct repeat DR) 序列。

除了末端带有 IS 序列的复合转座子以外, 还有一些没有 IS 序列的、体积庞大的转座子 TnA 家族。目前已发现近 40 种不同的 Tn 分别带有不同的抗性基因、乳糖基因、热稳定肠毒素基因或接合转移基因等 (表 11-1)。例如 Tn3 是不含 IS 序列, 只是在两端接有短的反向重复序列, 其中有 3 个编码区: 编码 β -内酰胺酶的氨苄青霉素 (ampicillin) 抗性基因 (amp^r)、转座酶基因 (tpA) 和编码一种阻遏物

的调节基因(*tnpR*)。整个分子长为 5 000 bp,两端各有 38 bp 的 IR(图 11-11)。显然,无论是 IS 或 Tn 的两端都有反向重复序列,似乎两端反向重复序列的存在与它们的转座功能密切相关。如 Tn3 的两个 IR 中任一个顺式作用元件缺失都会阻止转座。此外与转座有关的基因当然也是通过突变鉴定出来的,如 Tn3 中的 *tnpA* 突变体是不能转座的。该基因的产物是一种转座酶。*tnpR* 突变将增加转座频率,因为该基因表达产物是阻遏物,阻遏 *tnpA* 和它自己基因转录,如 *tnpR* 蛋白失活就使 *tnpA* 合成增加而提高了转座频率。这也表明 *tnpA* 转座酶的量是转座的限制因子。

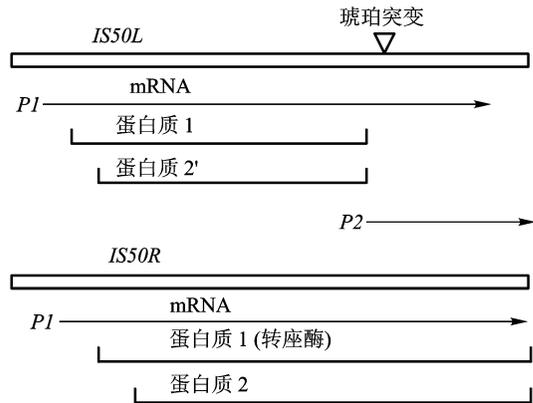


图 11-10 Tn5 末端重复序列的组成

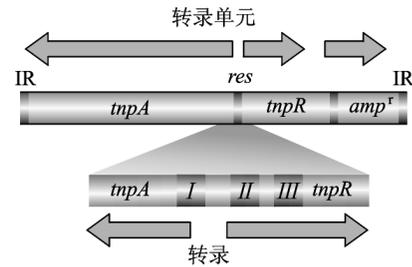


图 11-11 TnA 转座子家族成员结构(引自 Lewin, 2005)

TnA 家族的转座子末端有 IR, 3 个已知的基因和一个内部的解离位点(resolution site, *res*)

表 11-1 Tn 的特征

转座子	抗性标记	长度/ bp	反向重复序列中共同的序列/ bp	靶 DNA 中产生的重复序列的大小/ bp
Tn1, Tn2, Tn3	氨苄青霉素	4 975	38	5
Tn4	氨苄青霉素、链霉素、磺胺	205 000	短	含 Tn3
Tn5	卡拉霉素	5 400	8/ 9	α Tn5 的每一端都由插入序列 IS50 按相反方向构成)
Tn6	卡拉霉素	4 200		
Tn7	三甲氧苄二氨嘧啶、链霉素	14 000		
Tn9	氯霉素	2 638	18/ 23	α Tn9 的每一端都是同向插入序列 IS1)
Tn10	四环素	9 300	17/ 23	α Tn10 的每一端都按相反方向构成的插入序列 IS10)

转座子赋予宿主细菌一定的表型,如带有抗药性基因的转座子质粒便是抗药性质粒,根据这一特性采用相应的方法来检测某质粒上是否存在转座子。如该质粒是一个非转座性的,上面有一个 A^rTn 将另一个带有抗性基因 B^r的转移质粒引进同一细菌,然后把敏感细胞和带有这两个质粒的细菌进行接触转移试验,在特定的培养条件下使后者不能生长。如果一部分敏感细胞中出现 A^rB^r表型,这就说明 A^r属于某一转座子。如某一质粒上带有一个 Tn 那么变性并复性后,在电镜下可见到一部分杂合 DNA 双链上出现颈环结构。

11.2.3 转座噬菌体

Taybr 于 1963 年发现了一种特殊的噬菌体,称为 M μ (mutator phage),即突变者的意思。它是大肠杆菌的一种温和噬菌体,按理每一种温和噬菌体应整合到宿主染色体的一定位置上,如像 λ 噬菌体

的整合的位置是一定的。可是 *Mu* 几乎可插入宿主染色体任何一个位置上。而且游离 *Mu* 和已经插入的 *Mu* 基因次序是相同的。另外它的两端没有黏性末端,插入某基因中就引起该基因突变。这些都说明它的整合方式不同于 λ 噬菌体,而类似于转座因子的作用。

Mu 是一种 DNA 噬菌体,它是含有 38 000 bp 的线状 DNA,两端各带一小段大肠杆菌的 DNA,这与该噬菌体插入大肠杆菌染色体上有关。距末端不远处也有类似于 IS 的序列,靠近一端处存在与转座有关的整合与复制 A、B 基因。它们分别编码相对分子质量为 70 000 与 33 000 两种蛋白,位于 *Mu* DNA 一端。在 A、B 与末端之间有一 C 区,对 A、B 有负调节作用。*Mu* 的转座频率比一般的转座子要高。*Mu* 的复制能力和它的转座能力是密切相关的,*Mu* 的生存依靠转座,复制转座是其正常生活史中的一种方式。在转座过程中,它摆脱两端原有的细菌 DNA 而转座到新的某个位点上。*Mu* DNA 右侧还含有一段 3 kb 的序列,称为 G 区〔图 11-12 a〕。G 区两端各有一个 34 bp 的反向重复序列,其中含有两套基因,即 *Sv* 和 *U* 或 *Sb* 和 *U'*,后两个基因的转录方向与前两个基因的转录方向相反〔图 11-12 b) (c)〕。这 4 个基因都编码组成尾丝的部件。G 区的 4 个基因可以进行位点特异性的倒置,倒置后 *U* 和 *Sb* 基因转录,因为倒置后使这两个基因与其相邻的启动子位于同一条 DNA 单链上,另外两个基因则不能转录,因为它们的转录方向与启动子的作用方向相反。发生倒置的区域位于两个长 34 bp 的反向重复序列之间,倒置过程主要由噬菌体的 *gin* 基因编码产物即转化酶 (invertase) 催化,但还需要一种宿主细胞基因的产物。通过 DNA 片段的倒置来控制基因表达的现象还见于沙门氏菌 (*Salmonella typhimurium*) 的 P1 和 P7 噬菌体。

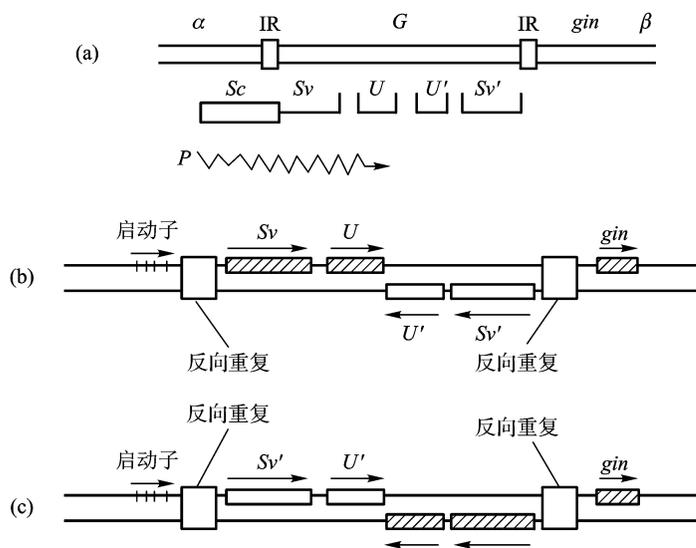


图 11-12 *Mu* 位点特异性的倒置

(a) *Mu* 噬菌体 G 区的结构 (α 区中的启动子 *P* 起始 *Sv* 和 *U* 或 *Sb* 和 *U'* 基因的转录, *gin* 基因的产物作用于反向重复,使 G 区颠倒) (b) 顺方向 (*Sv* 和 *U* 转录) (c) 反方向 (*Sv'* 和 *U'* 转录)

11.3 真核生物中的转座子

11.3.1 酵母菌基因组中的转座子

酵母菌是低等真核生物,它的基因组虽然在许多方面不同于原核生物,可是它的转座子却在许多

方面和细菌的转座子相似。目前在酵母中研究得较清楚的转座子是 Ty(transposon yeast Ty) 系列。它们的一般长度约为 5 900 bp, 两端含有两个称为 δ 的正向长末端重复序列(LTR), 正向重复序列的长度约 340 bp。 δ 因子大约由 70% 的 AT 组成, 每一个 δ 因子都含有一个启动子和一段被转座酶识别的序列[图 11-13 a]。

Ty1 插入酵母菌染色体后, 受体上就会出现 5 bp 的正向重复序列, 这和细菌中转座子作用相似。另外由于 Ty1 插入后, δ 为正向重复序列, 所以也有可能发生类似于细菌中复制重组过程形成的小环, 丢失一个 δ , 而留在酵母中的 δ 称为 Solo δ , 这时酵母细胞表型则恢复正常, 而人们可以通过对 Solo δ 的分析来判断此处曾插入过 Ty[图 11-13 b]。据研究, 每一个酵母细胞中有数百个 δ 因子。这些 δ 不能转座, 但由于其含有转录启动信号, 它们可能会影响其附近基因的表达。另外由于每一个细胞中有多个拷贝的 Ty1 因子, 所以基因组中不同位置的 Ty1 因子也可能发生重组, 从而导致染色体的易位、缺失和倒位等结构变异。

Ty1 因子只编码一条长 5 700 nt 的 mRNA, 转录的启动子位于 Ty1 因子左端的 LTR 之中。转录物含有两个可读框, 所以这条 mRNA 可能编码两种蛋白质。在某些酵母品系中, Ty1 因子可多达 35 个拷贝, 但拷贝数因品系不同而有差别。

Ty1 因子转座是通过一种 RNA 中间产物进行的, 其过程类似反转录病毒的复制与整合, 所以 Ty1 因子也称作反转录转座子。一般认为 Ty1 因子转座时, 首先以其 DNA 为模板合成一个拷贝的 RNA, 然后再通过反转录合成一条新的 Ty1 因子, 最后这条新的 Ty1 转座子再插入到新的位点上(图 11-14)。

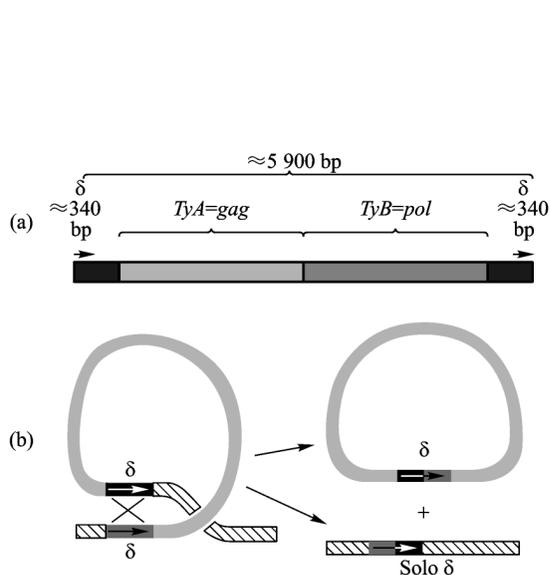


图 11-13 酵母 Ty1 转座子的结构 (a) 与 Solo δ 的形成 (b) (引自 Snustad, 2003)

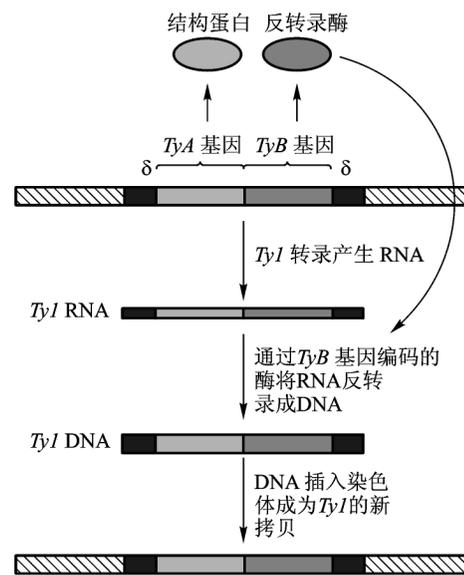


图 11-14 酵母 Ty1 的转座(引自 Snustad, 2003)

酿酒酵母中有控制 a 与 α 两种接合型的两个基因 a 与 α 。这两个基因都能转座, 当 a 基因从它的位置 HMR 转座到 MAT 位置后便能表达, 细胞就成为 a 接合型; 当 α 基因从它的位置 HML 转座到 MAT 后, 原来在 MAT 上的 a 基因消失, α 基因得以表达, 细胞便能转换成 α 接合型。表明这两个转座子具有明显的生理功能, 它们与其他转座子不同之处是只能转座到 MAT 这一个位置上, 而其他的转座子几乎可以转座到该基因组中任何位置上。同时酵母接合型相互转换是属于复制型转座。

11.3.2 果蝇基因组中的转座子

从 20 世纪 70 年代以来,在黑腹果蝇中一些品系间杂交子代出现某些异常现象,例如卵巢发育不全,分离比异常,雄性个体的减数分裂中出现重组,高突变率,染色体畸变等。这种现象统称为杂种劣育(hybrid dysgenesis)综合征。1977 年, M. G. Kilwell 等第一次证实杂种劣育只出现在特定的杂交组合中。黑腹果蝇中作为父方的,造成杂种劣育的品系称为父方品系(paternal strains)简称 P 品系作为母方的与 P 品系杂交能造成杂种劣育的品系,则称为母方品系(maternal strains)简称 M 品系。P 雌×P 雄、M 雌×M 雄、P 雌×M 雄杂交 F_1 正常。唯独 M 雌×P 雄杂交 F_1 出现杂种劣育。深入研究证明,这是因为在 P 品系的细胞中有导致杂种劣育的遗传因子,称为 P 因子(P element)。P 因子和细菌中的转座子亦有许多相似之处:① P 因子的全长是 2907 bp,两端有 31 bp 的反向重复序列。② P 因子有 4 个编码区(ORF 0、1、2、3)、3 个内含子(1、2、3)。体细胞中,只有内含子 1、2 能被顺利切除,产生包括外显子 0、1、2 的功能型 mRNA,并被翻译成相对分子质量为 66×10^3 的转座阻遏蛋白——一种转座活性的抑制因子[图 11-15 a]。在生殖细胞中这种蛋白质缺失,使内含子 3 被切除,产生的成熟 mRNA 包括全部 4 个外显子并被翻译成相对分子质量为 87×10^3 转座酶,才能导致 P 因子转座和配子败育。因为在体细胞中存在一个与外显子 3 特异性结合的蛋白,这种蛋白与 RNA 的结合妨碍了内含子 3 的剪接。

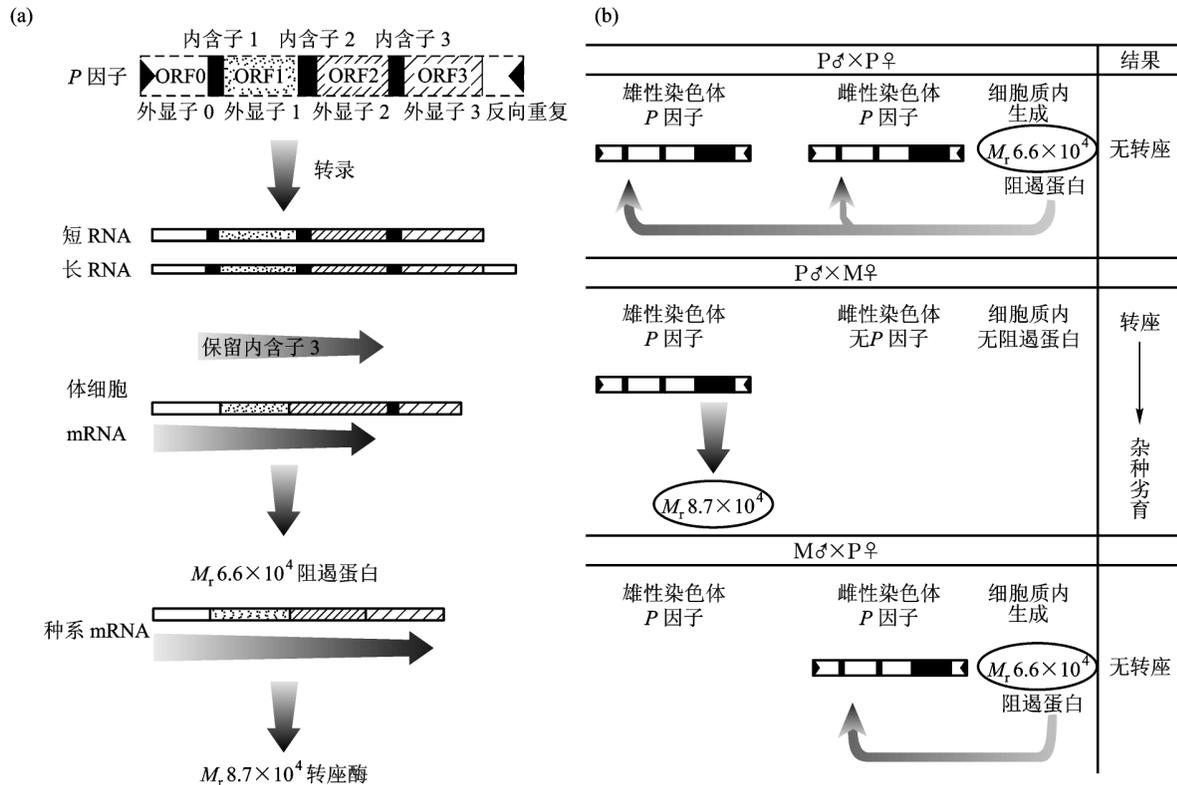


图 11-15 果蝇的 P 因子与杂种劣育

研究发现,果蝇的细胞质因子与 P 因子转座有关。只有在 P 雄×M 雌的 F_1 劣育是因为 M 品系雌性细胞质内缺失一个相对分子质量为 66×10^3 的转座阻遏蛋白,因此引起了 P 品系的雄性细胞核染色体上的 P 转座子自由转座,后代劣育。而在其他 3 组杂交中,由母本提供的正常细胞质因子都能抑制 P 因子的转座,后代可育。由此可见 P 因子与杂种劣育取决于基因组中的 P 因子与细胞质中阻遏蛋白之间的关系[图 11-15 b]。

P因子插入后在靶 DNA 序列形成 8 bp 的正向重复序列。如缺失了中间序列而保留两端的 31 bp 反向重复序列的 P 因子,只要有相应的外显子提供转座酶,那么这些缺失的 P 因子仍然可以转座。P 因子可以从原来的位置上消失,这一过程称为切离(excision)。准确的切离可以使得由于 P 因子的插入而引起的突变型发生回复突变,同时 P 因子也全部从插入位置上消失。不准确的切离可以带来染色体畸变,P 因子或是全部消失或是有一部分残留在染色体上。P 因子的位置和数目分析是用同位素标记的 P 因子克隆作为探针进行检测,结果表明不同果蝇品系的基因组中 P 因子的位置、数目均不相同。而且 P 因子插入引起突变的基因有白眼(*w*)、焦刚毛(*su*)、黄体(*y*)等。许多这类突变基因中都可以通过原位杂交的方法来证明 P 因子的存在。

果蝇中的转座子除了 P 因子外还有 *cop*₁、*412*、*279*、*T₁*、*FB* 等。它们的结构虽有所不同,但两端都有反向重复序列(图 11-16)。*cop*₁ 在果蝇中广泛地存在,已发现有 20~30 族,其成员的长度从 5 000 bp 到 9 000 bp 不等,各个成员的两端都接有长 200~500 bp 的正向重复序列,每个正向重复序列两端具有反向重复序列,这一特征与酵母的 Ty 转座子相同。每一成员两端正向重复序列并非高度同源,其中存在一部分碱基不能互补的区域。在果蝇基因组中,*cop*₁ 约有 30 个拷贝,分布在几条不同染色体上,各个成员之间存在高度结构保守性。当 *cop*₁ 插入到某个基因座位上以后,也在其两侧形成 5 bp 的染色体正向重复 DNA。在成体的果蝇细胞中以及在胚胎细胞中还存在多个拷贝的、呈小环状的、游离于染色体外的 *cop*₁ 转座子。与酵母的 Ty 转座子相似,*cop*₁ 也是通过一种 RNA 中间产物进行转座。*cop*₁ 序列含单一的长为 4 227 bp 的阅读框,其阅读框部分序列与反转录病毒的 *gag* 和 *pol* 具有同源性,但与 *env* 缺乏同源性,这意味着 *cop*₁ 不能编码外壳蛋白而产生病毒样颗粒。*cop*₁ 转录物是一类富含 poly(A)⁺ 的 mRNA,这些 mRNA 有相同的 5' 端,都是从一个末端重复序列的中部起始转录而形成的。转录物能翻译产生出多种蛋白质。它们可能涉及 RNA 剪接和多聚蛋白质的切割等过程。

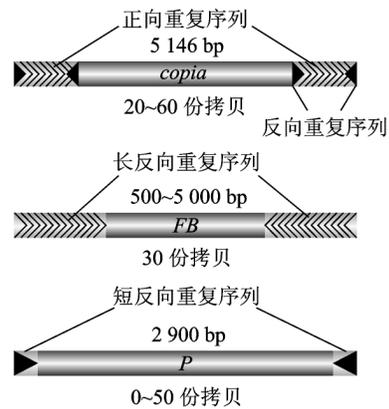


图 11-16 果蝇中的 3 种转座子结构

11.3.3 玉米基因组中的转座子

在著名的 Ac-Ds 体系中,自主转座的 Ac 因子为一种结构和功能都比较完整的转座子,除了含有自动转座的信息外,还具有一些与切割和转座有关的基因或 DNA 序列,所以 Ac 因子的结构比较复杂,相对分子质量也比较大。根据对玉米非蜡质基因(*W₂* 座位中 Ac 因子的序列分析,Ac 因子长 4 563 bp,转录生成单一 RNA。成熟 mRNA 长 3 500 nt 并含一个长 807 个密码子的开放阅读框。在阅读框前有 TATA 盒,其后接有多聚 A 加尾信号序列,并被 4 个内含子分隔成 5 个外显子,可能编码与转座有关的转座酶(transposase)。Ac 转座子两端具有长 11 bp 的反向重复序列,在 Ac 因子的插入位点上,两段反向重复序列之外各接有一段 8 bp 的染色体 DNA 正向重复序列(图 11-17)。

Ds 为一种结构和功能都不完整的转座子,只具有与切割有关的识别序列,缺乏与转座有关的功能,所以不能自动转座,容易在基因组中固定下来。根据各种 Ds 与 Ac 因子的序列同源程度差异,可以将 Ds 分成 3 种不同类型,第一类为 Ac 因子的缺失体,如图 11-17 中的 Ds-a。第二类 Ds 只在总长不足 1 kb 的末端区同 Ac 因子存在同源性,而中部序列则与 Ac 因子无关,如图 11-17 中的 Ds-b。第三类 Ds 只是在末端反向重复序列以及 3' 端约 20 bp 的序列与 Ac 因子存在同源性,如图 11-17 中的 Ds-c。同 Ac 因子一样,Ds 也具有 11 bp 的反向重复序列和 8 bp 的染色体 DNA 正向重复序列。Ac-Ds 的转座机制属于非复制型转座,直接从原来位置切离后插入到新的靶位点。

在玉米或其他植物中另一类控制因子为抑制子-启动子-突变子系统(suppressor promoter mutation system, Spm)与缺陷型的 Spm(deficient Spm, dSpm)组成的 Spm-dSpm 系统。这类控制因子也由两个

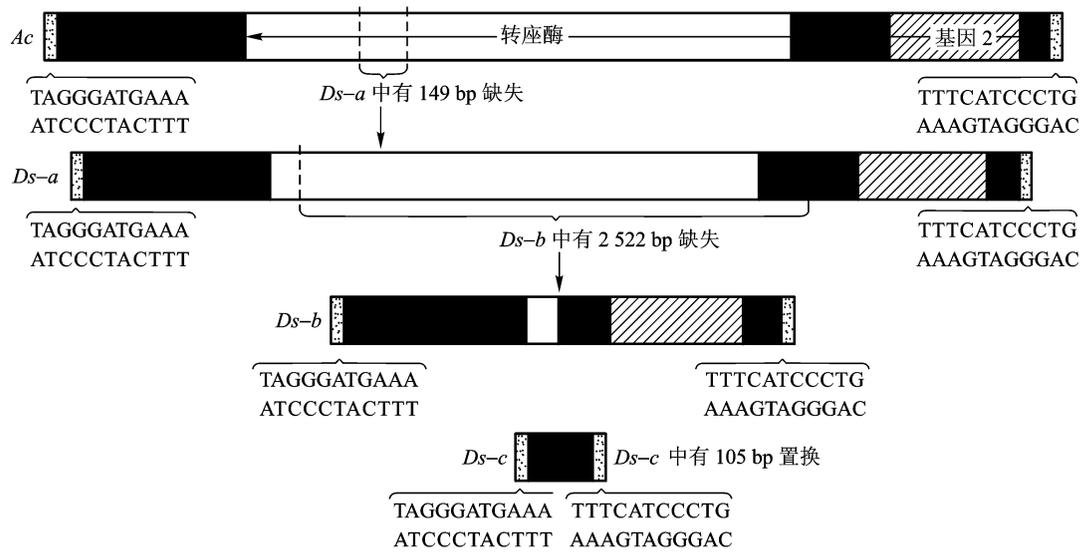
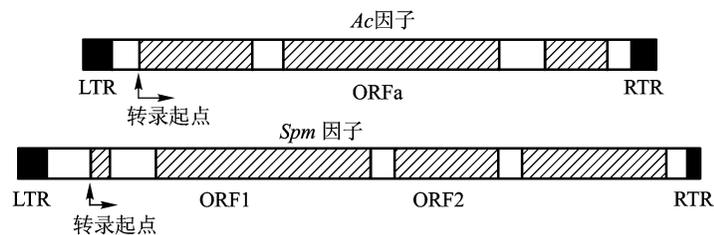


图 11-17 玉米中的几个 Ds-Ac 系统

成员组成,即一个长 8.3 kb 的 Spm 因子和一个较小的称为 dSpm 的成员。dSpm 为 Spm 因子内部发生缺失后形成的转座子,但能够独立地转座。所谓抑制子(suppressor)即 Spm 插入到某个基因后,往往抑制该基因表达。而突变子(muton)则是指 Spm 转座所需的功能,它可能是指转座酶。Spm 因子具有上述两种功能,而 dSpm 则不具有抑制基因表达的功能,只有植物基因组中存在 Spm 因子时,dSpm 才能抑制插入位点基因的表达,即 Spm 以反式互补的方式为 dSpm 提供抑制功能。

各种 Spm 和 dSpm 因子的末端都具有 13 bp 的反向重复序列,转座以后,都能在插入位点上产生 3 bp 的染色体正向重复。同其他植物转座子一样,Spm 转座以后不能使插入位点的基因完全恢复到原来状态,而是在该位点造成一小段重复序列。Spm 的主要转录子只有 2.5 kb 长,编码一种反作用抑制子(trans-acting suppressor)蛋白,这种蛋白质也可能行使与 Spm 转录有关的一种正调节子的功能。Spm 也编码一些小转录子,这些小转录子含有主转录子的第一个内含子中的两个阅读框,即 ORF1 和 ORF2 (图 11-18)。所有这些转录子都可能是前体 mRNA 通过不同加工途径形成的。当这两个阅读框中发生缺失突变或移码突变之后,Spm 往往失去转座功能,而只具有抑制功能,因此,可能二者的编码产物都是转座酶的组成部分。

图 11-18 转座子 Ac 与 Spm 的结构比较
LTR、RTR 分别为左、右两个反向重复

11.3.4 人类基因组中的转座子

已知重复序列占了人类基因组 50% 以上,其中转座子占重复序列的 45%。所有的转座子都是多拷贝的,这些转座子分为 4 种类型(图 11-19)。最常见的为 LINE 称为长散在重复序列,如 L1 长为

6 500 bp,以富含 A 序列终止,它的全长含两个阅读框,分别称 ORF1与 ORF2。L1是唯一在人类和小鼠中都具有活性的一类成员。在哺乳动物基因组中的拷贝数可多达十万份。短散在重复序列 SINE,如 Alu元件长度在 100~300 bp之间,是非自主转座子,其 3'端与 L1有同源性,因此能依靠 L1进行转座。

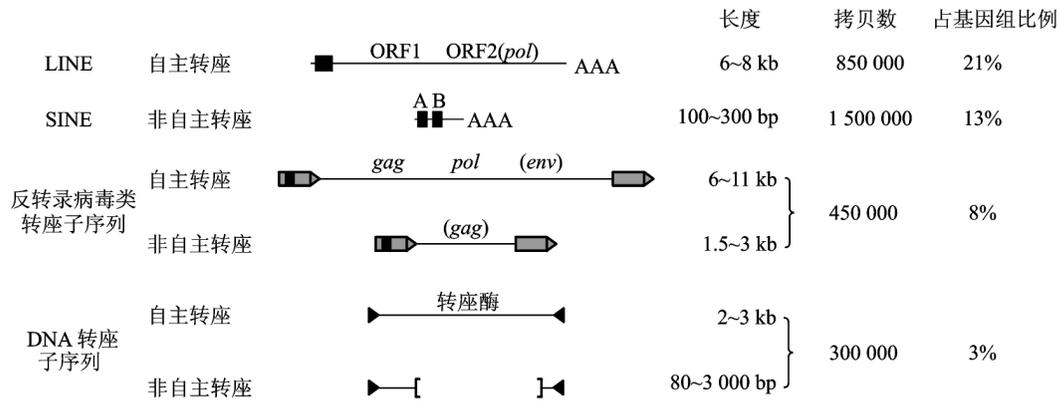


图 11-19 人类基因组中的几种主要转座子

人类基因组中存在大量的长约 300 bp 的中度重复序列,广泛地分布在非重复 DNA 序列之间。在这 300 bp 的序列中有一个限制性内切酶 AluI 的特异性识别位点 AGCT,由此将 300 bp 的序列切割为两个片段,一个片段为 170 bp,另一个片段为 130 bp,说明这是一类长度和性质相似的重复序列,因而称之为 Alu 家族(Alu family)。Alu 家族有多个成员,总拷贝数有 30 万~50 万个,大约平均每 6 kb DNA 序列中就有 1 个,占人体基因组的 3%~6%。在小鼠中与 Alu 序列相关序列称为 B1 家族,长度为 130 bp。在中国仓鼠中则被称为 Alu 相应家族(Alu-equivalent family),该家族也存在于其他哺乳动物的基因组中。

Alu 序列与 7SL RNA 序列有关,它是信号识别系统的一个组分,其序列与 Alu 序列的左半部类似,只是在中部有一个插入序列,所以 7SL RNA 5'端的 90 个碱基和 Alu 的左侧序列同源,而 3'端的 40 个碱基与 Alu 的右侧序列同源。编码 7SL RNA 的基因是由 RNA 聚合酶 III 转录,因此非活性的 Alu 序列可能是这些基因或者相关基因所衍生的。Alu 家族成员和转座子类似,其两端都具有短正向重复序列。由于它是来源于 RNA 聚合酶 III 的转录物,所以某些成员可能携带内源性活性启动子。Alu 家族部分成员被转录为单独的 RNA,有些成员可能存在于其他结构基因的转录单位内部。这也许正是 Alu 家族在哺乳动物基因组中含量如此丰富,分布如此广泛的重要原因。

11.4 转座作用的分子机制

11.4.1 DNA 转座机制

(1) 复制型转座

复制型转座机制的模型至少要说明下列现象:① 转座以后原来位置上的转座子保持不变。② 在新的位置上的转座子的两侧出现正向重复序列。③ 转座过程出现共联体。研究得比较清楚的还是细菌中的转座子,这里以细菌的转座子为例说明转座的一般过程。1979 年 J. A. Shapiro 提出了一个 Tn3 转座模型,大体分 4 步:① 切开(cutting):含有 Tn3 转座子的供体质粒,其 Tn3 中转座酶有两种功能,第一可以识别受体质粒上的靶(target)序列,并在该序列两侧各一条单链上造成一个切口,切口

之间的距离决定了转座后两侧正向重复序列的长度。第二可以识别自身两边的反向重复序列,并在3'端切开。② 连接(rejoining):供体和受体结合成为共联体(co-integrate),所谓共联体,就是两个或两个以上的复制子(replicon)通过共价连接起来的一个复制子。其过程是使供体切下IS或Tn3反向重复序列末端和受体黏性末端以共价链齐头相连,形成两个“缺口”。③ 复制:由DNA多聚酶进行修补复制补上缺口,由连接酶连接。于是在IS两端形成了两个正向重复序列(DR),一般为5~9 bp,最长的12 bp[图 11-20 a]。④ 重组:在特定位点进行重组,结果共联体分离形成两部分,一个是原来含有转座子的序列,另一个是通过转座插入了转座子的序列。由此可见,转座子是以它的一个复制品转移到另一位置的,而在原来位置上仍然保留着原有的转座子[图 11-20 b]。

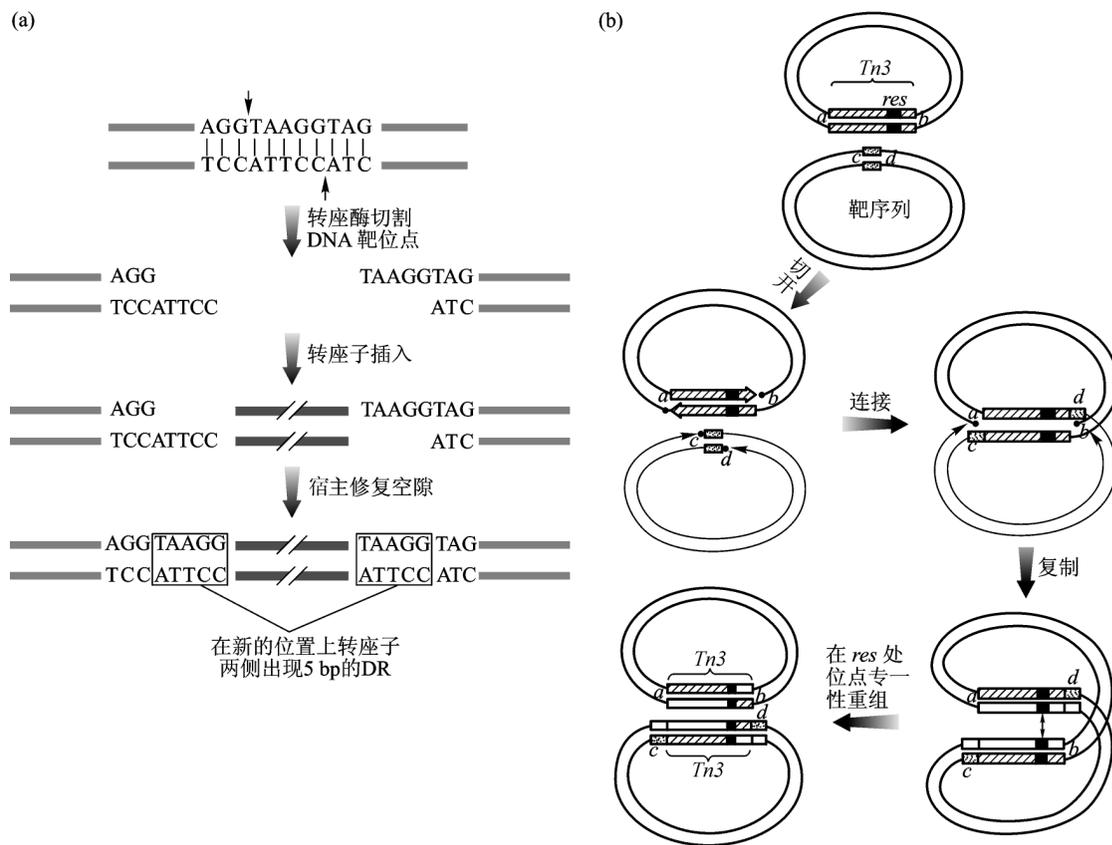


图 11-20 Tn3复制型转座模型
(a) 转座子两侧 DR 形成机制 (b) Tn3转座模型

Shapiro所提出的转座模型不仅可以说明上述3方面的现象,而且说明转座过程是遗传重组过程。这种重组发生在特定的位点,从这一层意义上说,转座过程是一种位点专一性重组;但是这种重组伴随着DNA的合成,所以是复制式的位点专一性重组。

(2) 非复制型转座

交换复合体也可用于非复制型转座,只是非复制型转座的原理是断裂和重接反应使靶序列重构,只有靶位点发生重连(reunion),而供体链仍保持裂缺,不形成共联体。非复制转座酶也可在靶DNA上产生剪接。转座子两侧产生双链断裂,完整的转座子从供体中释放出来。现将含有一个错配碱基的Tn10人工异源双链进行非复制型转座试验显示于下(图 11-2D)。若转座涉及复制的话,那么在新的位点的转座子应仅含来自Tn10亲体的一条链上的信息,若转座是一种被切下的转座子的物理转移,那么在新的位点上将会产生基因的转变。错配碱基的转变表明转座子保留了原来两条链的信息,

结果证明, Tn10的两条链的剪切发生在与靶序列连接之前。在反应的第一步是转座酶对转座子末端的识别,在转座子的两端的单链按一定顺序被剪切,首先是与靶位点连接的转移链被切,然后另一条链被切,在此位点释放被切的供体 DNA[图 11-21(a)]。同时受体也被交错切割,转座子与靶位点的切割端连接,保持在由转座酶产生的复合体中[图 11-21(b)]。在转座子两端的双链剪切排除任何复制型转座,结果产生非复制型转座。Tn10转座酶是一个二聚体,每一个单体都有活性位点,不仅能催化转座子两条链的断裂,并能交错地切割靶位点。Ac-Ds的转座也属非复制型转座,其转座机制与Tn10相似(图 11-22)。

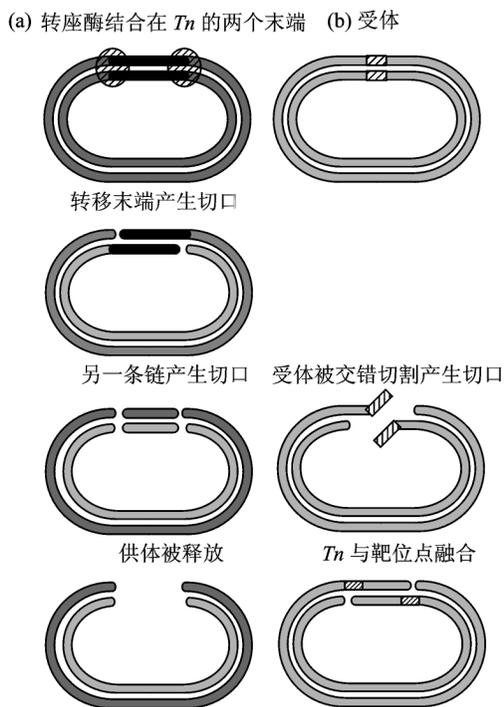


图 11-21 Tn10非复制型转座机制

(仿自 Lewin, 2005)

(a) Tn10的两条链先后被切割

(b) 转座子与切开的靶位点连接

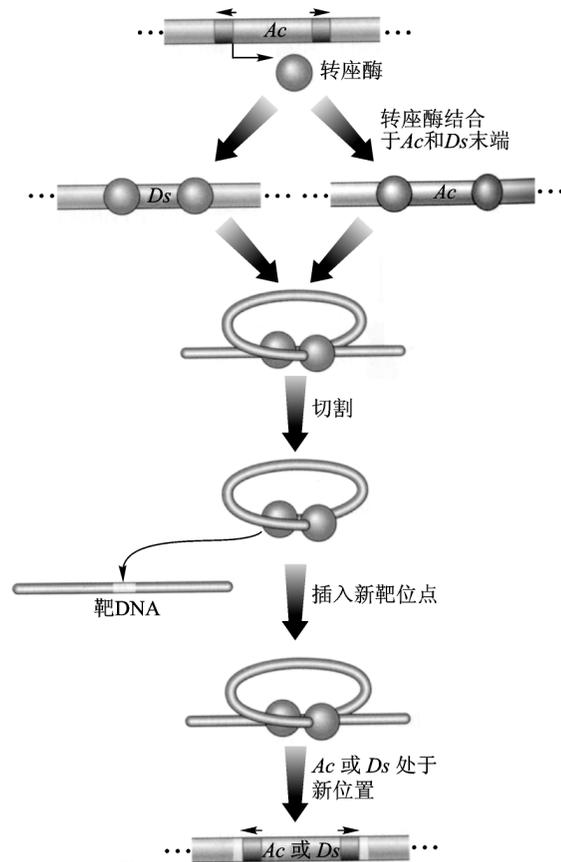


图 11-22 玉米中 Ac-Ds非复制型

转座机制(引自 Griffiths, 2005)

11.4.2 反转录转座子的转座机制

(1) 反转录病毒的转座机制

由 RNA 介导的转座仅发生在真核生物中,它们是由反转录病毒以其 RNA 基因组的 DNA 拷贝插入宿主细胞的染色体中而产生的。有些生物的转座子和它们组织中的反转录病毒的原病毒(provirus)有关,而且它们的转座也是通过 RNA 介导的。还有一些反转录成分,如 Ty1/copia 和 Ty3/gypsy 家族,它们虽不是病毒,但转座机制类似于反转录病毒。反转座子和反转录病毒都与其他类型的转座子相似,具有转座子的一些重要特征:如病毒 DNA 的两端的 LTR,在插入位点的靶 DNA 上产生正向重复序列(图 11-23)。

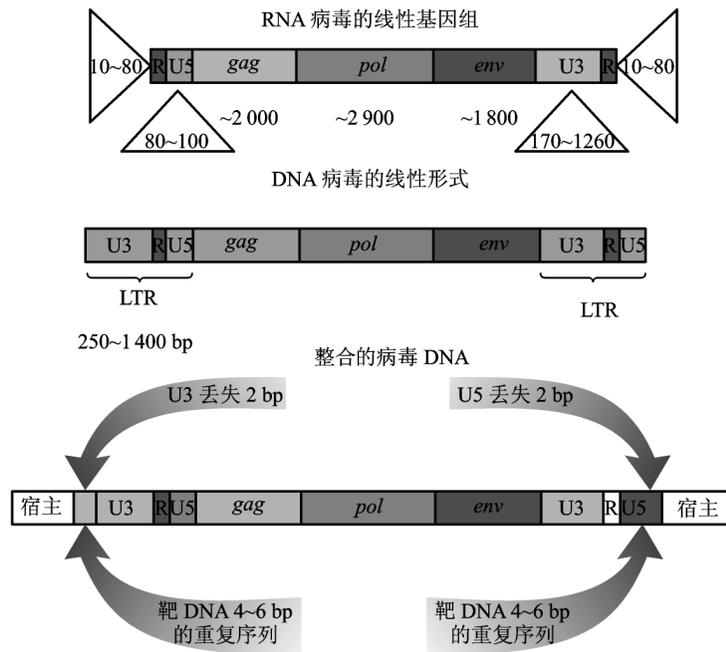


图 11-23 反转录病毒的基因组及其整合(引自 Lewin, 2005)

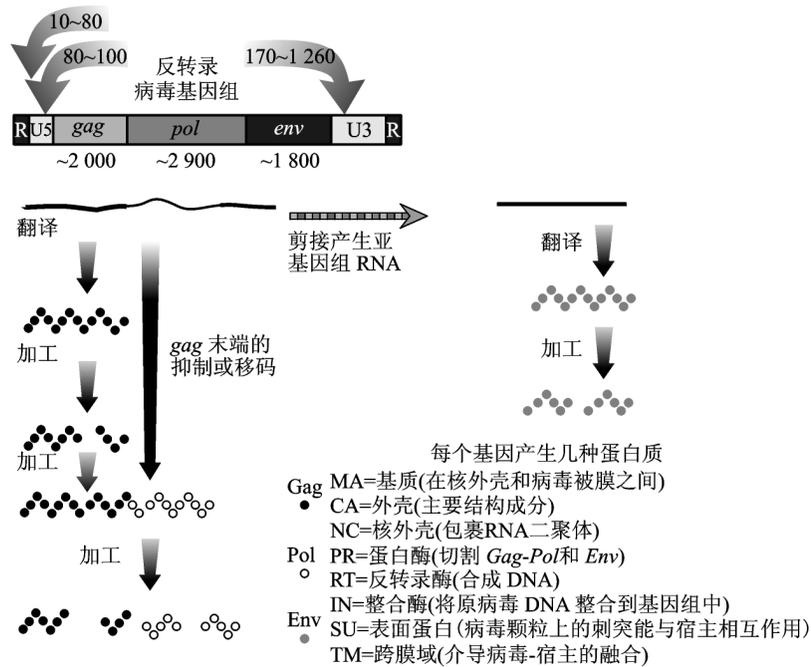
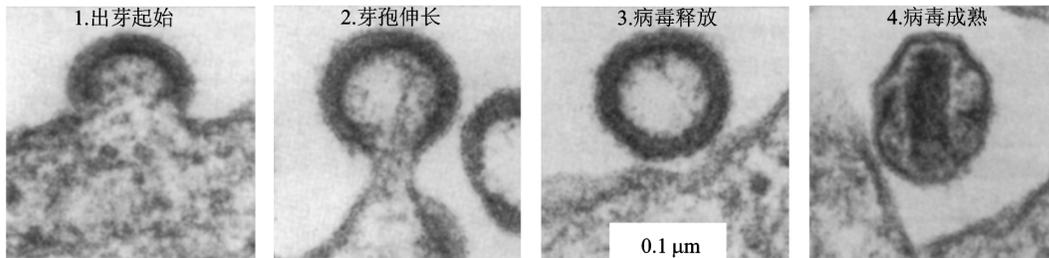
反转录病毒整合在细胞的基因组中,作为一种内源性前病毒(endogenous provirus),类似一个溶源性噬菌体。反转录病毒的增殖周期中,关键的步骤是借助反转录酶(reverse transcriptase)将病毒的RNA反转录成DNA,并整合到宿主基因组中。然后该DNA又转录成病毒的RNA,此RNA有两个用途:①作为翻译模板的mRNA。②作为病毒的基因组,典型的反转录病毒含2~4个基因。其mRNA具有常规的结构:在5'端有帽结构,在3'端有多聚腺苷的尾。它存在两种mRNA:长的mRNA翻译Gag和Pol多蛋白(polyprotein),从起始密码子到第一个终止密码的阅读框翻译成Gag产物。而Pol表达一定要越过第一个终止密码。不同的病毒翻译时采用不同的机制来越过第一个终止密码子,它取决于gag和pol两个阅读框之间的关系。当gag和pol需要连读的时候,通过可识别此终止密码子的G₁-rRNA来抑制终止,产生单个的蛋白。当gag和pol在不同的阅读框中,核糖体发生移框来产生各个不同的蛋白。通常连读率约在5%,因此Gag蛋白在数量上胜过Gag-Pol蛋白约20倍。

Env多蛋白通过另一种机制进行表达,它是将mRNA重新拼接成一个短的亚基因组的(subgenomic)信使,翻译成Env产物。gag或gag-pol和env产生的是多蛋白(polyprotein),它可能被蛋白酶水解酶切成多种单个的蛋白,存在于成熟的病毒中,这种蛋白酶的活性是由病毒编码的。它可能是gag或pol的一部分,有时又可以一个独立的附加阅读框存在(图11-24)。总之,反转录病毒基因表达产物为多聚蛋白质,经加工后成为单一蛋白质。

反转录病毒颗粒的产物涉及将RNA包装进核心蛋白中,外周包以衣壳蛋白和从宿主细胞膜上取下的片段,最后病毒颗粒从细胞中释放(图11-25)。感染正是这一过程的逆向反应。病毒是通过和一个新的宿主细胞质膜的融合来感染的,然后将病毒粒子的内容物释放到宿主细胞中。

①反转录病毒基因组RNA经反转录形成LTR 反转录病毒RNA可作为模板由依赖RNA的DNA多聚酶合成互补的DNA,这一步反应由转座因子中pol基因编码的反转录酶负责。因为DNA的合成需要引物,正常细胞的DNA合成中引物是重新合成的,由RNA多聚酶完成。但在反转录因子RNA的cDNA合成中,并无合成RNA引物的RNA多聚酶,因此该DNA的合成引物来源不同。这类DNA合成中采用的引物是细胞中的rRNA分子,何类rRNA分子用于引物依反转座因子的类型而异,如Ty1/copia家族总采用rRNA^{M^c}作为引物。

rRNA引物与反转录病毒RNA 5'端一段序列互补(图11-26)。引物结合的位置初看有点反常,

图 11-24 反转录病毒基因表达产物的加工(引自 Lew *in* 2005)图 11-25 HIV从受感染细胞的质膜出芽(引自 Lew *in* 2005)

因为它的合成方向不是指向分子内部而是朝向只有很短一段序列的 5'-LTR。当 cDNA 拷贝延伸到 LTR 的 5'端时,作为模板的这段 RNA 开始部分降解,从而向外延伸出一段 cDNA 单链。因为 3'-LTR 和 5'-LTR 为重复序列,合成的 cDNA 单链 3'端可与反转录病毒 RNA 分子的 3'-LTR 复性。随后 cDNA 合成继续沿转录物延伸,直到取代 RNA 引物。由此产生的 cDNA 拷贝是完整的,包括引物序列。事实上这里有一个模板转换,即从 5'-LTR 转到 3'-LTR 的过程,这一策略可解决线性 DNA 分子拷贝时末端缩短的问题。

第一个 cDNA 链完成合成后产生 1 个 DNA-RNA 杂种分子,该杂种分子 RNA 部分由 pol 基因编码的 RNaseH 酶降解。模板 RNA 分子并不是完全被降解,它会在接近 3'-LTR 的位置留下一小段作为引物启动下一轮第二个 cDNA 单链的合成。第二个 cDNA 单链合成仍由反转录酶负责,这是一个依赖 RNA/DNA 的 DNA 多聚酶。第一条单链 cDNA 合成后,第二条互补 DNA 单链开始合成,也在近 3'-LTR 处开始,到达末端时仍要进行一次模板转换,使 DNA 与 5'-LTR DNA 端退火以完成整个 DNA 分子复制。由此产生的双链 DNA 分子既含有反转座成分内部顺序,也有 2 个末端的 LTR(图 11-26)。

② 反转录 DNA 分子的整合过程 当完整的 DNA 分子合成后,下一步就是将新的反转座成分整合到基因组中。整合酶在 DNA 靶位交错产生单链缺口,使反转座因子与靶位点的 5'突出单链整合。插入成分与靶位点 5'突出序列之间并不互补,而是采用断裂重组方式使反转座子整合。整合时

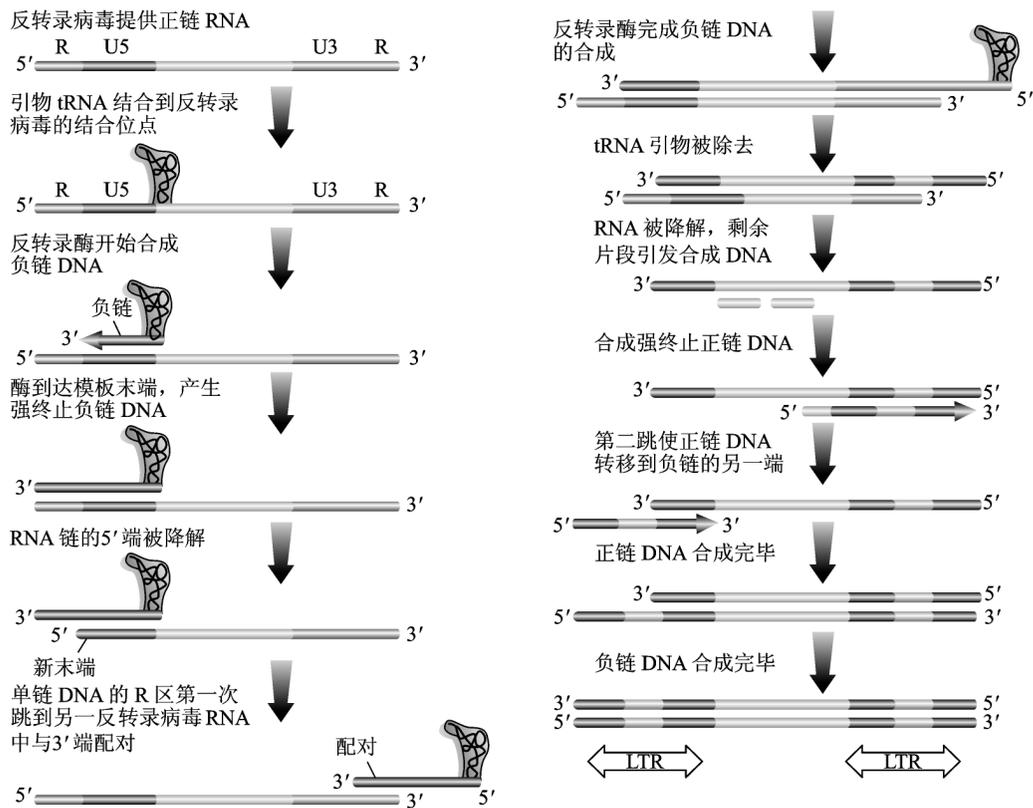


图 11-26 反转录病毒基因组中 LTR 的形成(引自 Lew 2005)

反转座子末端的突出序列会丢失一些碱基,留下的空缺将被填补,最终在插入位点的两端产生若干对碱基的正向重复序列(图 11-27)。

(2) Ty/copia 类反转座子的转座机制

Ty1 和 copia 因子是典型的反转座子,均由 RNA 介导转座,只是这类反转录子由于缺少编码外壳蛋白的基因 (*env*),因此不能形成病毒颗粒,它们只能在细胞内生活,亦可转录合成 RNA。因其基因组中有编码反转录酶的基因 (*pol*),所以能将它的 RNA 反转录为 cDNA,并以类似反转录病毒 cDNA 的整合机制插入到宿主基因组中。用基因工程的方法将内含子插入到某一个 Ty 序列中,此序列由质粒上的 *gal* 启动子来控制,然后导入到酵母细胞中,转座的结果酵母的基因组中存在多个拷贝的转座子,但它们没有内含子(图 11-28)。

上述结果说明, Ty 因子的转座是由 RNA 介导的,在 Ty 因子转座过程中,由于 mRNA 加工,内含子被切除后 RNA 拼接,剪接的 RNA 被反转录酶所识别,以其为模板合成了双链 DNA 拷贝而后转座。这表明 Ty 的转座机制与反转录病毒的整合机制类同。此外, Ty 因子与反转录病毒还

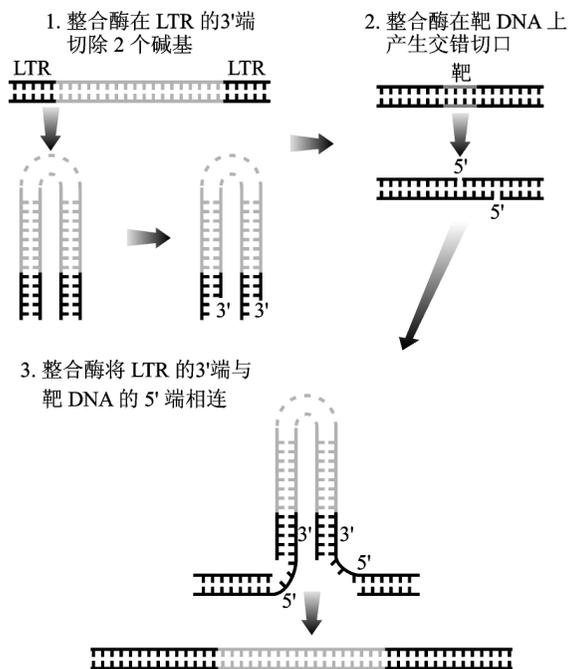


图 11-27 反转录病毒 DNA 的整合

(引自 Lew 2005)

有一些相似之处:① 原来的 Ty 因子两端的 δ 序列是不同的,但转座后的 Ty 两端具有相同的 δ 序列,它们来自原来的 5'端 δ 因子。如果我们把 δ 看做是一个由 U3-R-U5组成的 LTR 的话,那么 TyRNA 的延伸是由 R 区到 R 区。即 LTR 是由于在 3'加 U5在 5'端加 U3形成的。② 转座是由 Ty 因子内的基因控制的。gal 启动子用来控制被标记的 Ty 因子的转录,这个启动子可以被诱导,它可以通过加 Gal 来打开。启动子的诱导有两种作用:可促进目标 Ty 因子的转座 观察启动子的活性对 Ty 因子在酵母染色体上转座频率的影响。这意味着 Ty 因子的产物能反式作用于其他的 Ty 因子(实际是作用于它们的 RNA)。③ 虽然 Ty 因子不产生感染颗粒,但在经诱导发生转座的细胞中存在着 Ty 病毒样颗粒(Ty virus like particles, VLP)(图 11-29)。它们含有全长的 RNA,双链 DNA,具反转录酶活性的产物以及带有整合酶活性的 TyB 的产物。TyA 的产物像 gag 产物一样被剪切成 VLP 的核心蛋白。这显示了 Ty 转座子和反转录病毒十分相似。不同之处在于 Ty 因子没有 *env* 基因,因此不能包装它的基因组。④ 仅有某些 Ty 因子在任何酵母基因组中都有活性,大部分没有转座能力,这与惰性的内源性原病毒相似。由于这些无活性的 Ty 因子保留了 δ 重复序列,这些序列在对活性 Ty 因子的产物作出反应过程中提供了转座的靶序列。

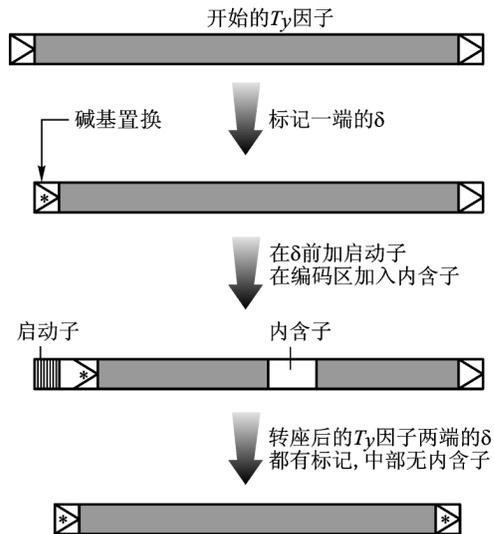


图 11-28 Ty 经 RNA 介导转座的证明(引自 Lewis 2005)

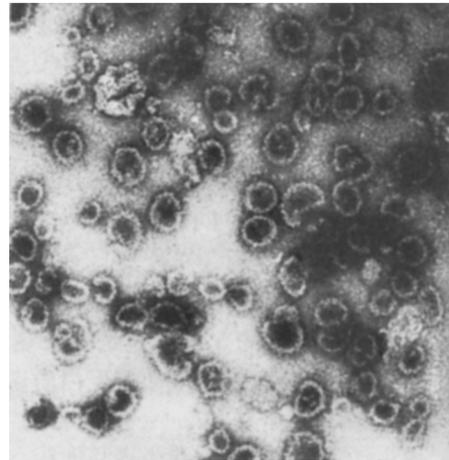


图 11-29 Ty 因子形成病毒样颗粒(电镜照片)(引自 Lewis 2005)

(3) 人类 LINE 反转座子的转座机制

缺少末端重复序列的反转座子的整合方式与反转录病毒略有不同。目前在哺乳动物基因组中发现的 LINE 和 Alu 反转座子的结构有 2 个共同特点:① 3'端有一串腺苷酸(—AAAAAA—)。② 转座子两侧有几对正向重复序列。但这类反转座子由于缺失了末端重复序列 LTR,它们不可能采取类似反转录病毒的转座方式。已经提出一些模型解释非 LTR 反转座子的转座机制,其中一种模型认为,① LINE 反转座子的转座首先在染色体 DNA 连续的 $(T)_n$ 位置产生缺口,② 以反转录转座子 3'端的多聚(A)与之配对,经反转录酶拷贝成 DNA 后再与 3'断头连接,③再以反转录的 DNA 为模板合成互补 DNA(图 11-30)。

由图 11-30 可见,人类基因组中 L1 RNA 转座机制步骤如下:① 在细胞核中,插入在染色体中 L1 转录为 L1 RNA。② L1 RNA 3'端加 poly(A)。③ poly(A) L1 RNA 转移到细胞质中。④ L1 RNA 的两个 ORF 翻译为两条多肽。⑤ L1 RNA 与其翻译的多肽转移至核中。⑥ ORF2 多肽切割染色体 DNA 分子单链,L1 RNA poly(A) 3'端与切开染色体 DNA 5'端连接。⑦ ORF2 多肽激活反转录酶,以 DNA 3'为引物,L1 RNA 为模板合成 DNA 单链。⑧ 新合成 DNA 单链插入 DNA 切口之中,同时

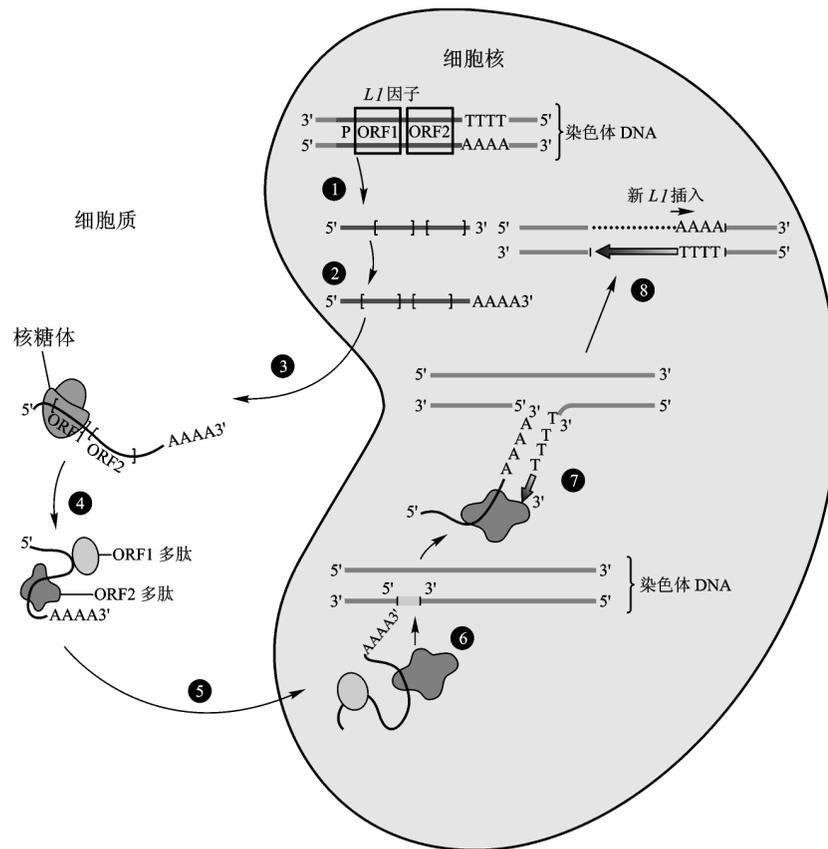


图 11-30 L1 类反转座子可能的转座机制(引自 Smustad, 2003)

L1 类反转座子的 ORF2 编码反转录酶, 在细胞质中翻译后即与编码 mRNA 结合成 RAP 复合物转移到细胞核中。在染色体 DNA 的一段串联 T 的位置产生切口, 以多聚 T 为引物合成与 mRNA 互补的第一链 DNA。合成的 DNA 与断口 5' 端连接, 然后在 5' 链合成互补 DNA

L1 RNA 消除, 染色体 DNA 另一链被切开, 合成与 L1 序列互补的第二链(图中点线处), 所有切口被修复并连接。

11.5 转座因子的遗传学效应及其应用

11.5.1 引起染色体结构变异

当转座子插入后而引起受体位点 DNA 一段短的正向重复序列(DR), 即靶位加倍(target site duplication)现象。如转座子切离在 DR 之间重组, 结果是夹在两个 DR 之间的 DNA 序列被切离而缺失(deletion), 中间的 DNA 序列形成一个环, 将从细胞中丢失, DR 在染色体上只留下一个拷贝[图 11-31(a)]; 如重组发生在 R 之间, 结果是夹在两个 R 之间的 DNA 发生倒位(inversion)[图 11-31(b)], 而反向重复得到保留, 这将会导致下一次的倒位。

转座子附近序列的缺失是两步反应的结果; 转座产生了 DR, 而重组又发生在此 DR 之间的转座子分子内。当新拷贝的方向和原拷贝方向相反时, 重组会导致倒位; 当方向相同时重组的结果产生了两个小的环状 DNA, 其中一个无复制原点而被丢失, 另一个有复制原点的可复制。因丢失环带有转座子附近序列结果造成转座子附近序列和一个重复序列的缺失。有的转座子可以促进复制倒位

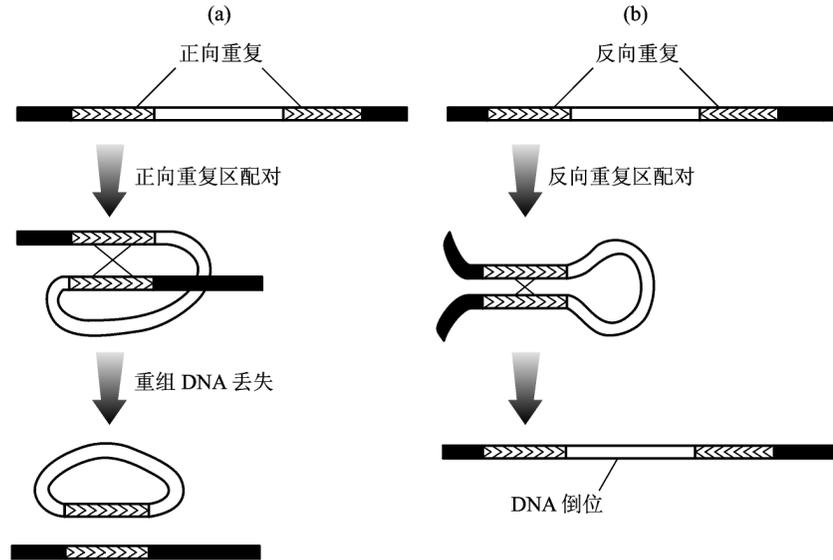


图 11-31 由转座子引起的染色体 DNA 缺失 (a) 或倒位 (b)

(*duplicative inversion*)。这些转座子以相反的方向位于中心区的任一侧,中心区发生倒位时,它们的方向也随之改变。

11.5.2 诱发基因突变与启动外显子混编

当转座子插入到某个基因中往往导致该基因失活,在某些情况下,插入位点的基因仍然能够正常转录,只是转座子中的插入序列通过转录后的剪接过程而被除掉,因此插入位点的基因仍表现出显性性状,这种现象称为渗漏突变(*leaky mutation*)。也就是仍有一些残余基因表达的突变。这类基因称为渗漏基因(*leaky gene*),又称亚效等位基因(*hypomorph*),即一种突变基因与其野生型有相似的效应,但效应较弱(详见第 13 章)。

插入失活和渗漏突变与转座子在插入位点中的方向有关。例如,当 *dSpm* 和 *Ds* 因子插入与所在基因的方向相反时,前体 mRNA 中的 *dSpm* 或 *Ds* 序列则可通过转录后的加工过程而被除掉,所以含有转座子的基因仍然编码野生型蛋白质和 mRNA,即所谓的渗漏突变。如果转座子在外显子中的转录方向与所在基因的转录方向相同,转录过程常终止在转座子的多聚 A 化信号位点上,形成半截子 mRNA,因此也造成插入失活。

对于 *Spm/dSpm* 控制因子系统来说,由 *dSpm* 造成的渗漏突变只是在基因组中不存在 *Spm* 因子时才能发生,如果植物基因组中存在 *Spm* 因子,无论 *dSpm* 的插入方向如何,含有转座子的基因都不能正常表达。这种现象进一步说明,*Spm* 以反式互补方式为 *dSpm* 提供抑制功能。

在细菌中,根据对乳糖操纵子和半乳糖操纵子的研究,当 *IS* 因子插入到这两种操纵子的 5' 端顺反子中以后,往往严重降低其下游顺反子的表达水平,这就是所谓的极性突变(*polar mutation*)或突变的极性效应(*polar effect*)。

当两个转座子被同一转座酶识别而整合到染色体的邻近位置时,则它们之间的 DNA 将变得易于被转座酶作用而转座。如果它们之间的 DNA 中含有外显子,则该外显子将被切离,并可能插入另一基因之中。这种效应称为外显子混编(*exon shuffling*)。所谓外显子混编,即源自一个或几个基因的若干个外显子像“洗牌”那样地进行重排。这也是生物体产生新基因和基因进化多样性的途径之一(图 11-32)(详见知识窗 11-2)。

转座子在染色体中经过一段长期沉寂之后,使几种转座类型的转座子几乎同时地进入一个活跃转座的时期,它们的转座、诱变效应将在群体的少数个体中不时地被激活。这种突发的变化称之为转

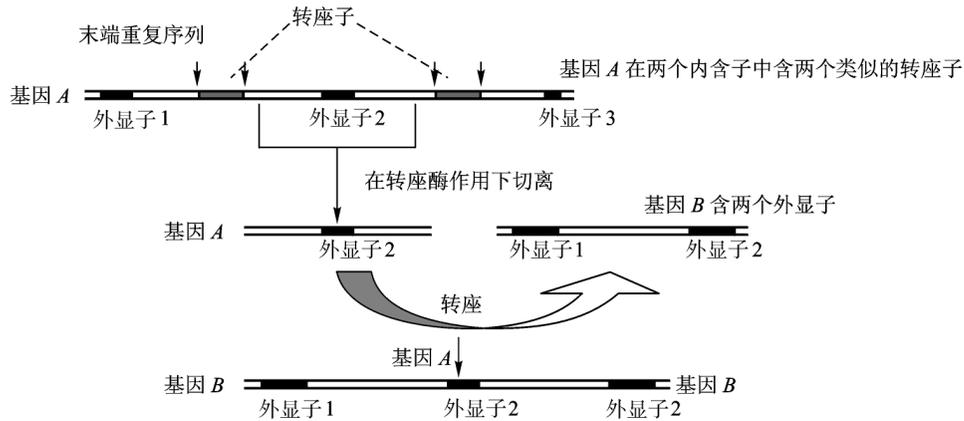


图 11-32 双转座启动不同基因间的外显子混编 (引自李振刚, 2002)

座爆炸 (transposon bursts)。转座爆炸首先发现于发育中的玉米,使其染色体反复地断裂而引发许多突变。另外,在果蝇某些品系的杂交中也发生转座爆炸。如果这种转座爆炸发生于生殖系的细胞中,则引发杂种劣育综合征。这种现象与通常的杂种优势 (hybrid vigor) 现象形成鲜明的对比。此外,若干植物类型证明,转座爆炸可以因严酷的生存条件引发而产生众多的变异子代,这对于增强物种的生存能力,扩大物种的适应性是有利的。因此,转座因子不仅仅导致生物基因组或基因的变异,有时它们也是一种增强生物生存能力的调控元件。

11.5.3 调节基因表达

已知酵母的 Ty1 因子是反转录转座子,其转座过程与 RNA 病毒的反转录插入宿主染色体的过程一样,只是 Ty1 转座子不像 RNA 病毒那样还制造外壳蛋白,因为它缺乏 *env* 基因。因此, RNA 病毒称为反转录病毒,而 Ty1 一类的转座子则称之为反转录转座子。正是这类反转录转座子不产生蛋白外壳,故只限于细胞内活动。众所周知,反转录病毒是带有增强子 (enhancer) 序列,而很多转座子也带有增强子,它们像 RNA 病毒一样,能使其插入部位附近的基因活性增加。转座子除了含有增强子外,有的转座子还含有启动子,也能促进基因的转录活性。如 Tn10 右侧的 IS10R 以某一方向插入到由于缺失了启动子而不能表达的 *argE* 基因的 5' 端时,结果使沉默的 *argE* 基因重新表达,对该序列分析表明,其末端含有一个外向的启动子,这可能是它启动沉默基因表达的原因 (图 11-33)。

当然也存在转座子插入某个基因的内含子中以后而影响该基因表达,通常表现为降低基因的转录水平。但在有些情况下也会改变内含子的剪接位点,使某些外显子丢失,因此也导致基因失活。如果转座子插入到某个基因的转录控制区或启动子中,则导致该基因完全失活。当转座子从该座位上切离后,该基因又可恢复到显性性状。但是由于切离后往往造成插入位点发生缺失、重复或倒位等结构的变异,因而其表达活性就不可能完全恢复到原来的状态,而使基因的转录活性和转录时空发生各种变化。如降低基因的表达水平,或改变基因的组织或器官特异性表达等。

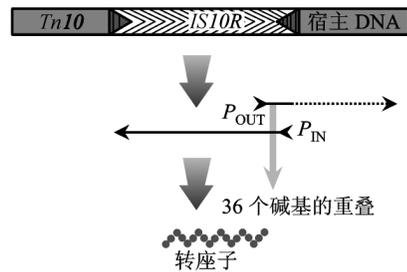


图 11-33 转座子调节宿主基因表达 (仿自 Lewis, 2008)

IS10R 两端具 22 bp 的反向重复序列。在它的右侧反向重复序列内侧有两个方向相反的启动子,其中 P_{OUT} 是外向启动子,它能激活邻近下游的宿主 DNA 中的基因。P_{IN} 是内向启动子,它们之间有 36 bp 重叠

11.5.4 产生新的变异

由于转座插入位点可能出现新的基因,如像 Tn携带的抗药性基因,它的转座不仅造成某个基因的插入突变,同时在此位点上出现一个新的抗药性基因。由于转座作用,使某些原来在染色体相距甚远的基因组合在一起,构建成一个操纵子或表达单元,也有可能产生一些具有新的生物学功能的基因和编码新的蛋白质分子。例如一些细菌的质粒由于转座而携带多种基因,其中一些基因为人们提供了丰富而有用的蛋白质资源,有的蛋白质或酶可使细菌分解代谢多种有害的化学物质或元素,如分解甲苯、萘、石油并抗有毒金属汞等物质。因而含这种质粒的菌株可清除海水中石油等物质的泄漏和环境的污染物而成为环境保护中有价值的“超级菌种”。另一方面在这类质粒中带有引起动植物疾病的致病基因,加之由于转座又存在对多种药物产生抗性的基因,因而该致病微生物所引起的疾病难以治疗,这已成为公害问题。

由于转座而增加了同源序列的整合,如 IS既可插入细菌染色体的不同位置,又可插入质粒中。大肠杆菌 F因子和染色体上都发现 IS,由此构成了 F因子和染色体上存在一些相同的插入序列,如 IS1和 IS2等。通过这些同源序列间重组,因而扩充了 F因子插入染色体而成为 Hfr菌株的种类。此外由于转座子也可以作为基因工程中的载体携带其他基因进行转座,形成重新组合的基因组或产生新的类型,有利于生物的进化。

在某些情况下,插入到某个基因座位中的转座子也可能失活,转座子失活过程主要与 DNA分子的甲基化作用(methylation)有关。当转座子中的某些甲基化位点去甲基化后,转座子又可以恢复转座功能,使转座子表现出表观遗传变异(epigenetic variation),即某基因的 DNA序列并未改变,而基因表达出现变化。转座子 DNA序列的甲基化不仅影响转座子自身的表达,而且还影响邻近基因也呈现出表观遗传变化。表观遗传学(epigenetics)正是研究在 DNA序列不出现变化的情况下基因表达出现可遗传变化的科学。表观遗传学的研究对象如 DNA甲基化、基因组印记、RNA编辑和 RNA干涉以及转座子调节等。这些表观的遗传变化在转座过程中可能出现,从而导致各类表型变化。

11.5.5 转座子标记目的基因

基因克隆是研究基因结构与功能和基因工程的必要前提。目前常用的方法是构建基因组文库或 cDNA文库,然后从中筛选目的基因。可用的筛选方法大多是建立在已经知道该基因的表达产物的基础上进行的,然而对于发育、生理和行为有关的一些调节基因及其产物往往是微量而且表达时间短,不易分离,更难以提纯,何况调节基因的调控作用也不一定是通过其基因产物参与调控。因此,分离这些基因无法从蛋白质入手。其实直到今日我们对很多基因表达的产物知之甚少。在基因产物未知的情况下一般采用两种方法来分离控制发育的基因,一种是在特定的生理过程或发育时期采用差别筛选(differential screening)原理来分离目的基因,称之为消减杂交与差别筛选(subtraction hybridization and differential screening)。另一种是转座子标记(transposon tagging)方法。该技术的原理是由于转座子序列可以在基因组中转座,如该序列转座正好插入某一基因的外显子区域时,导致该基因失活,使其表型改变而成为突变体,如该突变是由于转座子的 DNA克隆,其中必定含有与该突变体有关的基因。也就是说,用转座子给未知的目的基因加以标记,便于对该基因的认识与分离。用该突变型提取 DNA构建基因组 DNA文库,用标记的转座子序列作为探针,筛选出的克隆中,再对转座子两端序列进行亚克隆,这便是被转座子插入的基因序列。这些亚克隆又反过来作为探针,用于筛选野生型的基因文库,获得完整的目的基因。应用转座子标记技术,在植物中已成功地克隆到几种调节基因,如调节花色苷合成的 cI基因抑制种子蛋白积累的 Opaque-2基因。利用玉米的转座子标记已克隆出雄性不育基因和抗病基因等。

从以上克隆目的基因程序可见,应用该技术的条件是:①亲本之一必须含有活跃的转座子。②所用的转座子必须已经分离和鉴定,否则无法对其进行标记作为探针。③转座子插入后要有目

的性状的突变型。由于转座子在突变型基因组中插入的不定向性,所以用转座子作为探针得到的基因并不只是预期的目的基因,一般含有多种性状的突变类型,必须进行大量筛选。因此,该项技术目前还只能应用于少数富含活跃转座子的植物,如玉米和金鱼草等。

11.5.6 作为基因工程的载体

利用 P 因子作为载体,将外源基因转移到果蝇胚胎种系细胞中,对果蝇进行遗传操作。将携带目的基因的缺陷 P 因子和完整的 P 因子同时注入到果蝇的胚胎中,完整的 P 因子不仅能识别自身的末端序列,也能识别缺陷 P 因子的末端而进行转座,结果两种 P 因子都被插入到基因组中。只有 P 因子两末端之间的 DNA 序列才能被插入,两端外侧序列不是转座因子的组成部分,不会被插入到基因组。因此该技术的优点是只插入一个外源基因的拷贝。也就是说,所有转基因果蝇只携带一个拷贝的外源基因,因而便于对其结构与功能研究(图 11-34)。

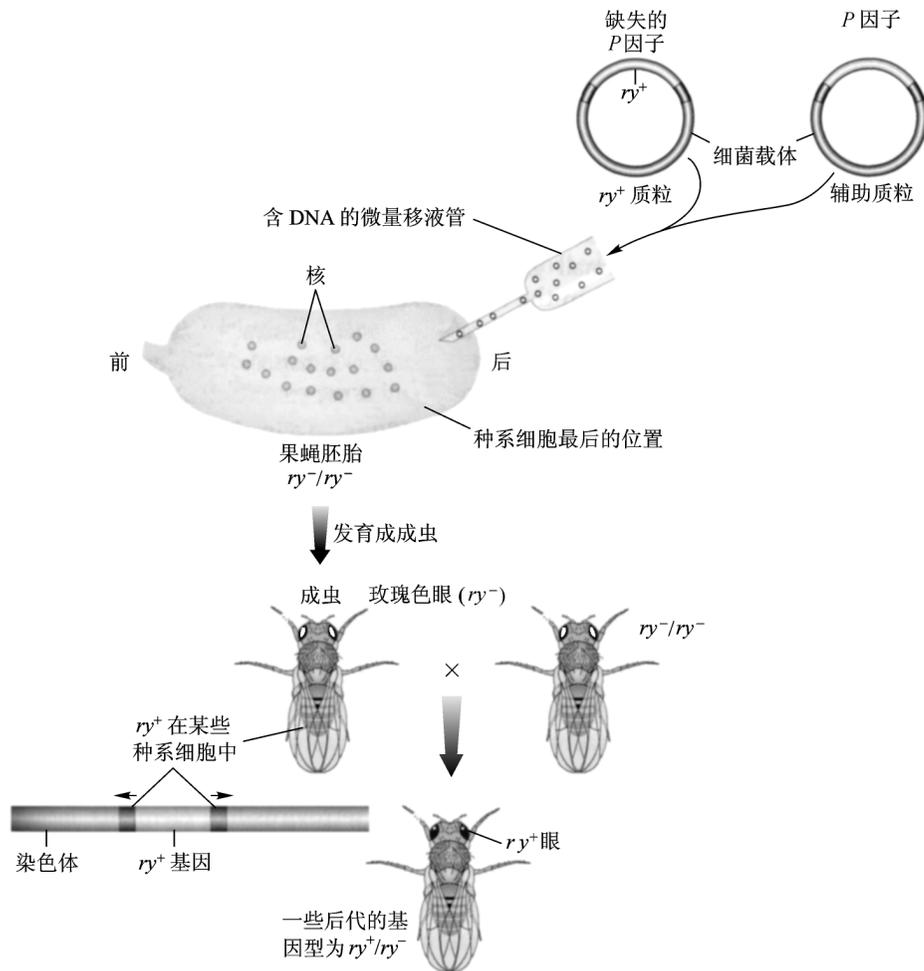


图 11-34 果蝇中 P 因子介导基因转移(引自 G riffs 等, 2005)

玫瑰色眼基因 ry^+ 被整合在细菌的带有缺陷的 P 因子中,同时与携带有完整的 P 因子的辅助质粒注入 ry^- 果蝇胚胎,结果在其后代中出现转基因 ry^+/ry^- 子代,原位杂交证明, P 因子介导 ry^+ 转移到一些果蝇的种系细胞的染色体中

利用 P 因子作为载体导入外源基因也只在适合的组织和适当的发育时期表达,并且表达与插入位点无关。因此对果蝇而言,基因自身含有调控基因表达所需的全部信息,基本不受外来因素干扰。由此可见,利用 P 因子将外源 DNA 导入基因组,或者引入在体外构建或修饰过的新基因或关闭内源性基因,研究它们的结构与功能和那些负责组织特异性表达的机制与特征,并最终采用定位的方

式替代基因组中的缺陷基因。

知识窗 11-2

外显子混编产生新基因

在生物进化的历史长河中,基因的年龄以 100 万年为一岁来计算。古老的基因年龄一般都在 1 000~3 000 岁之间。在自然界的物种中,新近产生的、年幼的基因才可称为新基因。探究新基因的起源已成为当代进化遗传学的前沿课题之一。关于新基因的起源的分子机制有如下几种:整个基因组的加倍,可最迅速地使基因的数目倍增,多倍化在植物的进化过程中起着重要的作用;其次是由于 DNA 复制错误,或是转座因子的拷贝插入基因组的另一位置所带来的基因重复(gene duplication),使新的重复片段发生戏剧性的变化产生全新的功能基因;产生新基因的第三种途径是从其他物种的基因组中获得,通过反转录病毒可以在具有巨大差异的物种间转移基因,从而获得具有新功能的基因,这种机制称为基因的水平转移(horizontal transfer);此外,真核生物中新基因起源的一种不可忽视的且已证明具普遍性的途径是在反转录转座或转座子的作用下,启动不同基因间的外显子混编而构成新的外显子-内含子的重组结构导致新的基因出现。

一个典型的由外显子混编产生的,在果蝇物种中最近出现的新基因——精卫基因(jingwei jgw)其年龄只有 2~2.5 岁。这是美国芝加哥大学教授、进化遗传学家龙漫远所领导的、多年从事新基因起源研究的重要发现之一。据报道,精卫基因只存在于 *D. yakuba*, *D. santomea* 和 *D. teissieri* 这 3 个非洲果蝇的基因组中。他们对于 jgw 基因起源的分子机制研究指出:位于果蝇第 2 染色体的、编码乙醇脱氢酶(alcohol dehydrogenase, Adh)基因,经反转录作用插入位于果蝇第 3 染色体的炎帝基因(yande gene)的第 3 个内含子中,从而由 5 端 yande 基因的 3 个外显子和 3 端 Adh 基因的 4 个外显子产生具有新的外显子-内含子结构的融合基因(图 1)。图中可见, Adh 基因的 mRNA 在反转录转座的介导下插入 yande 基因的第 3 内含子中。新形成的 jgw 基因的上游有 yande 基因的 3 个外显子(以深灰色显示),其下游带有 yande 基因的 9 个退化的外显子(以波状结构显示)。对 jgw 基因的功能研究表明,其表达产物是一种嵌合蛋白,由其 N 端来自 yande 基因编码产生的多肽序列, C 端则是 Adh 基因的多肽序列所组成。果蝇 *D. teissieri* 中的 jgw 基因在雄性睾丸中特异表达。jgw 基因的产物——精卫酶可能参与了雄性果蝇的激素和外激素的代谢过程。龙漫远发现,至少有一半以上的真核生物的基因,曾经经历了类似于 jgw 基因的起源过程,从而证明了外显子混编以及基因重复在新基因起源过程中的普遍意义。

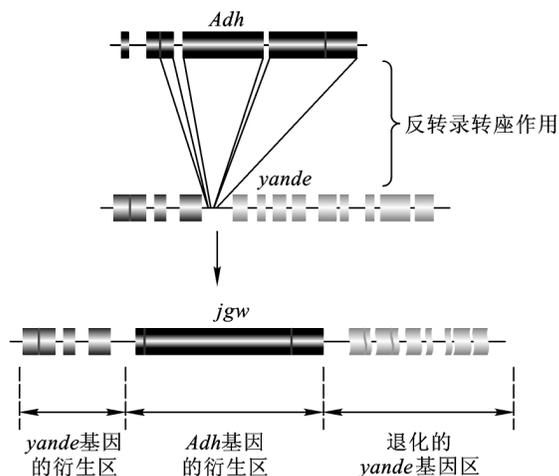


图 1 精卫基因形成的分子机制示意图

注：“精卫”基因的命名源自我国古代一个美丽的神话！

参考文献

- [1] Griffiths A J W, Essler S R, Lewontin R C, et al. *Introduction to Genetic Analysis*, 8th ed. New York: W. H. Freeman and Company, 2005.
- [2] Long M, Langley C H. Natural selection and the origin of jingwei: a chimeric processed functional gene in *Drosophila*. *Science*, 2004, 260: 91–95.
- [3] Long M Y, Betran E, Thomson K, et al. The origin of new genes: Glimpses from the young and old. *Nature Reviews Genetics*, 2003, Vol 4: 865–875.
- [4] Zhang JM, Dean A M, Brunet F, et al. Evolving protein functional diversity in new genes of *Drosophila*. *Proc Natl Acad Sci USA*, 2004, 101: 16 246–16 250.

2 思考题

- 如何证明 **Ty** 元件是典型的反转录转座子？它与反转录病毒有哪些相似的性质？
- 一个特定的转座因子以一个相当恒定的速率(尽管这一速率通常不高)进行复制加倍。因此,在进化过程中,预期一个细菌细胞的后代可能含有几千个这种转座子的拷贝。然而,细菌转座子的拷贝数却很低(通常每个细胞只有一个或两个)。
 - 为这种低拷贝数提供一个解释。
 - 为什么大多数细菌转座子是从质粒中,而不是从细菌染色体中分离出来的？
- 发现一个特定转座因子的转座依赖于反转录酶的活性。请提供一个机制,解释这种转座过程。
- 列举转座事件导致癌症发生的可能机制。
- 给出至少两种 **RNA** 病毒产生 **mRNA** 的机制。
- 关于反转录病毒:(a) 明确它们的定义特征。
 - 给出它们的病毒体中所包含的酶的名称,并列举该酶进行的 3 种生化活动。
 - 确定合成反转录病毒 **mRNA** 的模板。
 - 确定反转录病毒复制的胞内定位。
 - 指出反转录病毒中提示 **DNA** 插入机制与转座相关的特征。
- 在哺乳动物基因组中存在很多拷贝高度重复的 **Alu** 序列家族成员。试问:
 - Alu** 家族的祖先是什么结构? 该序列又是如何在进化中产生的?
 - Alu** 序列可通过什么途径去影响邻近基因?
- 反转座子与反转录病毒的主要区别是什么?
- 如果动物细胞的细胞质中含有高水平的反转录酶,你认为将对 **mRNA** 产生什么后果? 该细胞的命运又如何?
- 假设动物细胞中的 **mRNA** 被反转录成 **DNA**(由于反转录病毒感染时,或反转座子产生的反转录酶),接着形成双链 **DNA** 并整合到染色体上和原先基因不同位置。那么这个 **DNA** 序列和原先产生 **mRNA** 的那个基因有什么不同? 这个 **DNA** 序列能被转录吗?