

7 由转座造成的进化

基因组曾经被认为是相当静止的实体，在其中可以给基因指定一个界限明确的基因座位。因此，基因被想象成在漫长的进化时期里一直停留在它们正好所处的染色体位置上。当芭芭拉·麦克林托克 (Barbara McClintock) 在十九世纪四十年代发现玉米中的某些遗传成份能从一个基因组座位“跳到”另一个上，不时地改变着结构基因的表达时，基因组的这种静态图景开始崩溃。然而，这种凝固静止的图景在科学思想界中已积习难改，以至于人们花了近 40 年的时间才认识到麦克林托克这一开创性发现的意义。今天我们认识到，基因组的结构组织比以前所想象的要更富于流动性、且更易于发生进化变化。在这一章里，我们将描述大量有助于遗传物质从一个基因组处位向另一个处位运动的可转座因子，并且讨论这类因子对进化过程可能会产生的影响。

7.1 转座与反录转座

转座 (transposition) 的定义是，遗传物质从一个染色体位置向另一个位置的运动。具有能改变其基因组位置这种内在潜能的 DNA 序列，被称为易动因子 (mobile elements) 或可转座因子 (transposable elements)。根据可转座因子是否被复制来区分，则有两种类型的转座。在保守型 (conservative) 转座中，该因子从一个位点移到另一个位点 (图 7-1 a)，供体位点上发生了什么则不清楚。有一种模型提出，供体 DNA 的末端不是相互连结的，因而转座后剩余下来的分子就瓦解了。然而，如果细胞含有该供体序列的重复，则该供体 DNA 种即可避免从细胞谱系中丢失。在这种情况下，虽然消耗了一个拷贝，但另一个却幸存了下来，由此而产生的谱系将在原始位点上有一个因子，而在新位点上有第 2 个因子 (Berg 等, 1984)。另一种模型提出，该双链断裂是由宿主的修复系统所修复的。

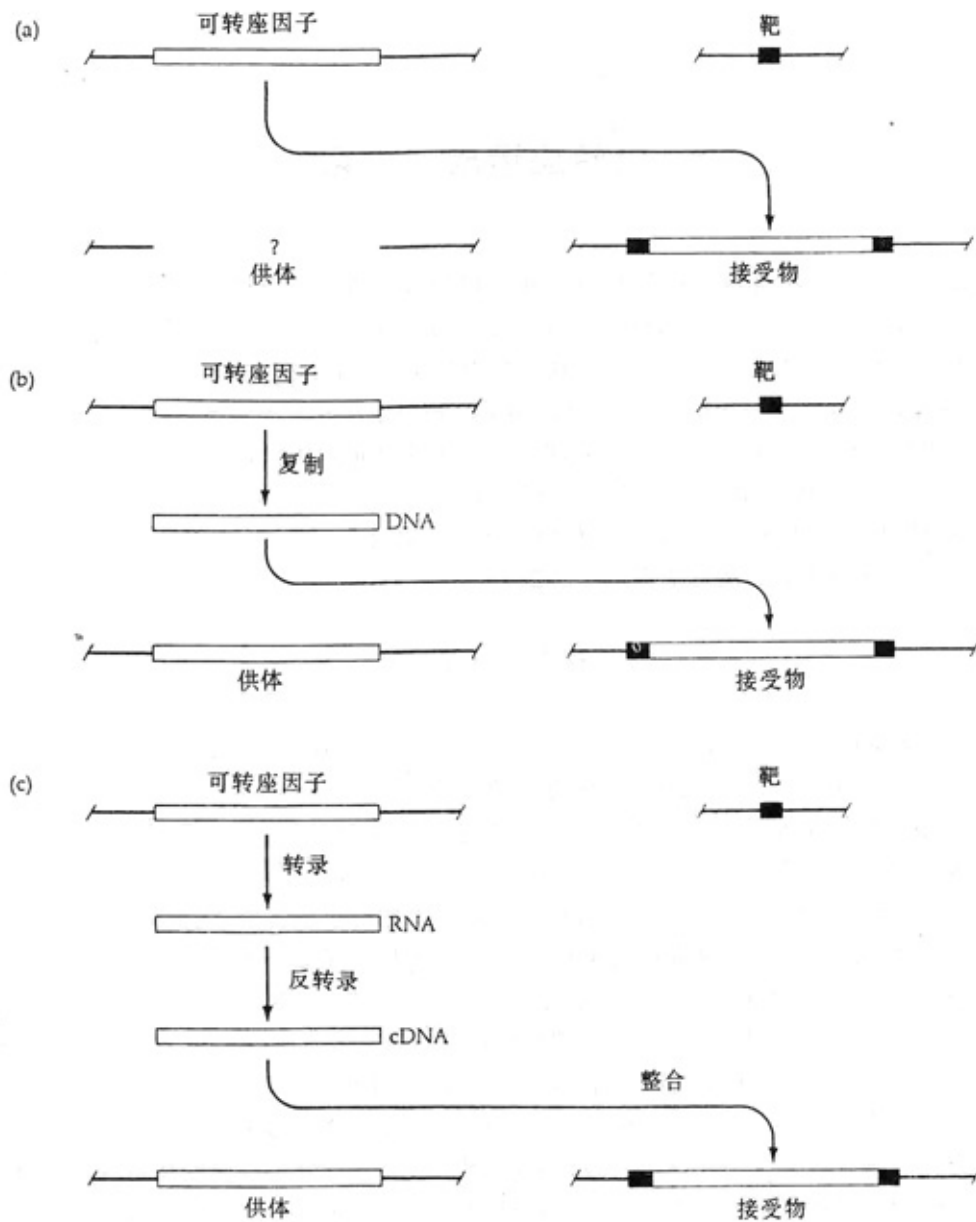
在复制 (型) (replicative) 或重复 (型) (duplicative) 转座中，可转座因子被拷贝，并且其中一个拷贝留在原位点处，另一个则插入一个新位点中 (图 7-1 b)。所以，复制型转座的特征是可转座因子的拷贝数的增加。有些可转座因子只采用一种类型的转座，而另一些则对保守型和复制型两种方式全采用。

在以上转座类型中，遗传信息由 DNA 所携带。已知遗传信息也能通过 RNA 而转座。在这种模式中，DNA 被转录成 RNA，然后 RNA 再反转录成 cDNA (图 7-1 C)。为了将这两种模式加以区别，我们把由 RNA 中介的模式称为反录转座 (retroposition)。转座与反录转座都曾既在真核生物中又在原核生物中发现 (见 Weiner 等, 1986; Temin, 1989)。与由 DNA 中介的转座相比，反录转座总是复制型的，因为被转座的是该因子的反转录拷贝，而不是该因子本身。

当一个可转座因子插入一个宿主基因组中时，在该插入位点处的一小段宿主 DNA (通常为 4-12 bp) 被复制 (图 7-1)。这种经复制的重复有同一取向，因而被称为顺向重复 (direct repeats)。这是转座和反录转座的特征标记。

有些可转座因子能转座到所有细胞中去；而另一些则是高度特异的。例如，果蝇 *D. melanogaster* 的 P 因子通常仅在生殖细胞中是易动的。转座的接受位点的基因组位置图在不同可转座因子中也表现出有变异。有些因子对某一特别的基因组位置表现出不寻常的偏爱。例如，IS4 总是精确地将自己掺入到大肠杆菌的半乳糖苷酶操纵子中的同一点上，所以，每个细菌只能含有一个拷贝的 IS4 (Klaer 等, 1981)。另外一些，象噬菌体 Mu，则能随机地转座到几乎任何基因组位置上。许多可转座因子则表现出中等程度的基因组位置偏向性。例如，E. coli 的 Tn10 转座子，其 40% 被发现位于 lac Z 基因中，成为基因组的精细结构部分；P 因子对 X 染色体有亲合性，且偏向于插入从 5' 端

靠近基因编码区的序列中，而不是编码区之中。有些可转座因子对某种特别的核苷酸组成展示出有更高的亲合性。例如，IS1 偏爱富含 A T 的插入位点 (Devos 等, 1979) 即是如此。



7 - 1 (a) 保守型转座。该因子从供体位点转座到靶位点。供体位点处发生了什么则不清楚。该供体分子可能会瓦解，这对于具有一个以上染色体拷贝的细菌而言是可以忍受的。另一种可能性是，该双链的断裂处被宿主的修复系统所修复。(b) 复制型转座。该因子被复制，并且有一个拷贝插入一个靶位点处，而另一个拷贝则保留在供体位点上。关于保守型和复制型转座的更详细解释可参阅 Lewin (1990)。(c) 反录转座。该因子转录成 RNA，然后 RNA 再反转录成 DNA。该 DNA 拷贝插入宿主基因组。作为反录转座的一个例子可见图 7 - 4。注意，转座与反录转座都会在新插入成份的两端各产生一小段重复 (黑矩形块)。

7.2 可转座因子

可转座因子，根据其转座模式和其所含基因的数目与类型，可分成三类：插入序列、转座子和反录因子。

插入序列

插入序列 (insertion sequences) 是最简单的可转座因子。它们除具有为转座所必需的部分外不带任何遗传信息。插入序列在长度上通常为 700—2500 bp, 已在细菌、噬菌体、质粒和玉米中发现。细菌的插入序列由前缀 IS 加上后随的类型号数来表示。一个来自肠道细菌 *E. coli* 和 *Shigella dysenteriae* 的插入序列, IS1 的结构, 如图 7-2 a 所示。IS1 在长度上约为 770 个核苷酸, 包括两个颠倒的非等同末端重复, 每个各 23 bp。它含有两个阅读框架, *insA* 和 *insB*, 它们为一种或两种形式的转座酶 (transposase) 编码。转座酶是一种催化可转座因子向插入位点插入的酶。在 *E. coli* 中有几十种不同类型的插入序列, 从自然界分离出的品系, 其大多数的基因组中所含每种序列的数是可变的 (Sawyer 等, 1987)。

转座子

转座子 (transposons) 是易动因子, 通常长为 2500—7000 bp 左右, 大多数在基因组中作为散在重复顺序的家族而存在。它们与插入序列的区别是还带有所谓外源性基因 (exogenous genes), 即一些为除与转座有关的功能之外的功能蛋白质编码的基因。(注意, 在有些文献中命名方式较混乱, 有时转座子这个术语被用来指所有可转座因子, 包括插入序列、反录转座子等。) 在细菌中, 转座子用前缀 Tn 和后随的类型号数来表示。有些细菌转座子是复合转座子 (complex transposons) 或混合转座子 (composite transposons), 这样命名是因为, 有两个完整而独立的可转座插入序列以任意一种排向从两个侧面夹拥着一个或多个外源性基因 (图 7-2 b)。有趣的是, 在复合转座子情况下, 不仅整个转座子能作为一个整体转座, 而且两侧的一个或两个插入序列还能独立地转座。因为转座的功能是由插入序列编码的, 所以, 复合转座子通常不含有一种独立的转座酶基因。

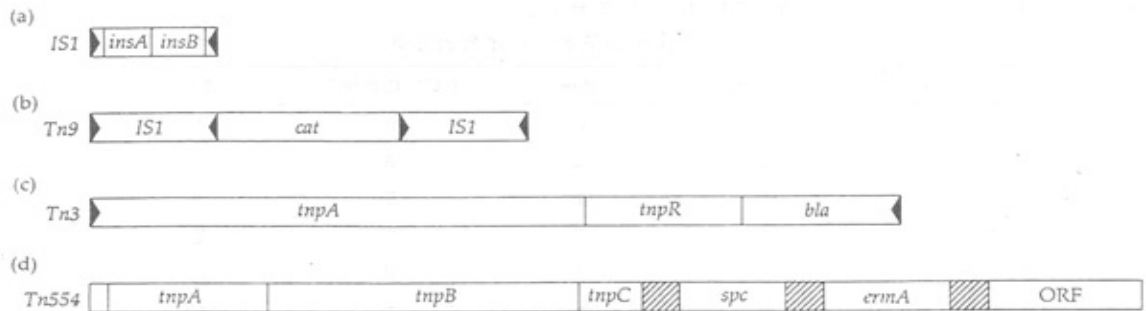


图 7-2 细菌中的四种可转座因子的模式图。黑三角形表示颠倒重复。(a) 来自 *E. coli* 和 *Shigella dysenteriae* 的插入序列 IS1 两侧由长为 23bp 的不完美颠倒重复所夹拥。(b) 来自 *E. coli* 的复合转座子 Tn9, 含有两拷贝 IS1, 分别位于 *cat* 基因两侧。*cat* 基因为具有氯霉素抗性的蛋白质编码。(c) 来自 *E. coli* 的转座子 Tn3, 能产生链霉素抗性, 含有 3 个基因, 其中两个 (*tnpR* 和 *bla*) 在一条链中转录, 第三个 (*tnpA*) 则在另一条链中转录。Tn3 两侧都有长为 38bp 的完美颠倒重复。(d) 来自 *Staphylococcus aureus* 的 Tn554 缺少末端重复, 含有 5 个基因和 1 个开放阅读框架 (ORF)。其中 3 个基因 (*tnpA*, *tnpB* 和 *tnpC*) 为转座酶编码, 且作为一个单位而被转录。*spc* 和 *ermA* 基因则分别提供对壮观霉素和红霉素的抗性。*pc* 基因为依赖 S-腺苷甲硫氨酸的甲基化酶编码, 是在与别的基因不同的链中转录的。ORF 被大量转录, 但是否被翻译则尚不得而知。画有斜条纹的矩形块不含开放阅读框架。

其他的细菌转座子, 以及许多真核生物的转座子, 其两侧只有较短的不同取向的重复顺序 (图 7-2 c), 并且不含插入序列。然而, 并非所有转座子都是在结构上对称的。有些具有不对称的末端, 缺少颠倒的或顺向的末端重复 (图 7-2d)。动物中有些转座子 (例如果蝇中的 P 因子) 的编码区被内含子所隔断 (图 7-3)。



图 7-3 果蝇 *D. melanogaster* 的 P 因子的模式结构图。该因子两侧有长为 31bp 的短颠倒重复, 其编码区含有 4

个外显子(白矩形块),由3个内含子(黑矩形块)所隔断。该因子长为2900bp左右。

细菌中的转座子常常带有这样一些基因,它们能给予携带者以抗生素抗性(如Tn554)、重金属抗性(如Tn21)或抗热性(如Tn1681)。质粒则可把这样一些转座子从一个细胞带到另一个细胞,结果,抗性就能迅速传遍到整个暴露于这样一些环境因子中的细菌群体中。

有几种噬菌体事实上是细菌中的转座子,或转座性噬菌体(transposing bacteriophages)例如,噬菌体Mu就是一个非常大的转座子(~38000bp),它不仅为那些调控其转座的酶编码,而且还为大量构成其DNA包装所必需的结构蛋白质编码。

许多类型的转座子在动物、植物和真菌的基因组中广泛分布。例如果蝇*D. melanogaster*就含有50—100种不同类型的转座子,而且都是多重拷贝的(Rubin, 1983)。

反录因子

反录因子(retroelements)是含有一个为反转录酶编码的基因的DNA序列或RNA序列,反转录酶则催化以RNA为模板的DNA的合成。由此而产生的DNA分子称互补的(DNA complementary DNA, cDNA)。这些的确转座的反录因子是通过反录转座过程而转座的。有一些不同类型的反录因子,我们采用由特明(Temin, 1989)提出的分类,并将它们列于表7-1。

反录病毒(retroviruses)是一些结构上类似于转座子的RNA病毒。虽然它们是所有反录因子中最复杂的,但我们还是最先讨论它们,因为反录转座概念起源于对反录病毒生命周期(图7-4)的发现。反录病毒的颗粒称为病毒粒子(virion),它侵入某一宿主细胞以后,其基因组RNA即被反转录成病毒DNA。该DNA能整合到宿主基因组中,变成一个前病毒(provirus)。接下来,前病毒DNA转录成RNA,它们既可作为合成病毒蛋白质的mRNA,又可作为病毒的基因组,并被包装到具有感染性的病毒粒子中。病毒粒子一旦形成,周期即可再次开始。

表7-1 反录因子和反录序列的分类

因子	反转录酶	转座	LTR* 是否存在	病毒粒子
反录子	是	否	否	否
反转录子	是	是	否	否
反录转座子	是	是	是	否
反录病毒	是	是	是	是
拟反录病毒	是	否	是	是
反录序列	否	否	否	否

自 Temin(1989) a、LTR, 长末端重复

反录病毒至少具有3个基因: gag, pol 和 env (图7-5)。这些基因分别为几种内部蛋白质、几种酶(包括一种反转录酶)和一种被膜蛋白编码。许多反录病毒还有别的基因:例如艾滋病毒至少有6个额外基因。反录病毒的编码区两侧有长末端重复(long terminal repeats, LTRs)。LTR含有与转录(在前病毒阶段)有关的启动子和与反转录(在病毒阶段)有关的启动子。

反转录子(retroposons)和反录转座子(retrotransposons)是不构成病毒粒子的可转座因子,所以,它们与反录病毒不同,不能独立地穿越细胞而转座。它们间的相互区别是,是否存在末端重复顺序(LTR)(表7-1)。注意,有些作者是将反转录子和反录转座子当作同义词来用的。果蝇中的copia因子代表一种典型的反录转座子;它的两端都具有LTR,并且还含有一段长的开放阅读框架,其中有类似于反录病毒的pol基因的区域。图7-5b所示的是另一个反录转座子的例子,即粘性霉菌*Dictyostelium discoideum*中的DIRS-1因子。*D. discoideum*平均含有40个左右的完整DIRS-1拷贝,以及大约200—300个的DIRS-1片段。有趣的是,DIRS-1有一种将自己插入别的DIRS-1序列中,从而抵消它们的作用的倾向,这或许是粘性霉菌基因组中存在许多有缺陷的DIRS-1片段的原因(Cappello等, 1984)。DIRS-1基因的转录受发育阶段诱导,也受热震惊的诱导。

与反录转座子不同,反转录子不含LTR。果蝇*D. melanogaster*中的G3A因子是一个反转录子(图7-5c)。这种反转录子含有两个ORF。ORF-1含有一个与反录病毒的pol基因类似的区

域，而ORF-2则含有7个由非常短的间隔顺序隔开的外显子。

反录子(retrons)是最简单的反录因子(图7-5d)。它们曾在某些细菌基因组中发现(Inouye等, 1989; Lampson等,

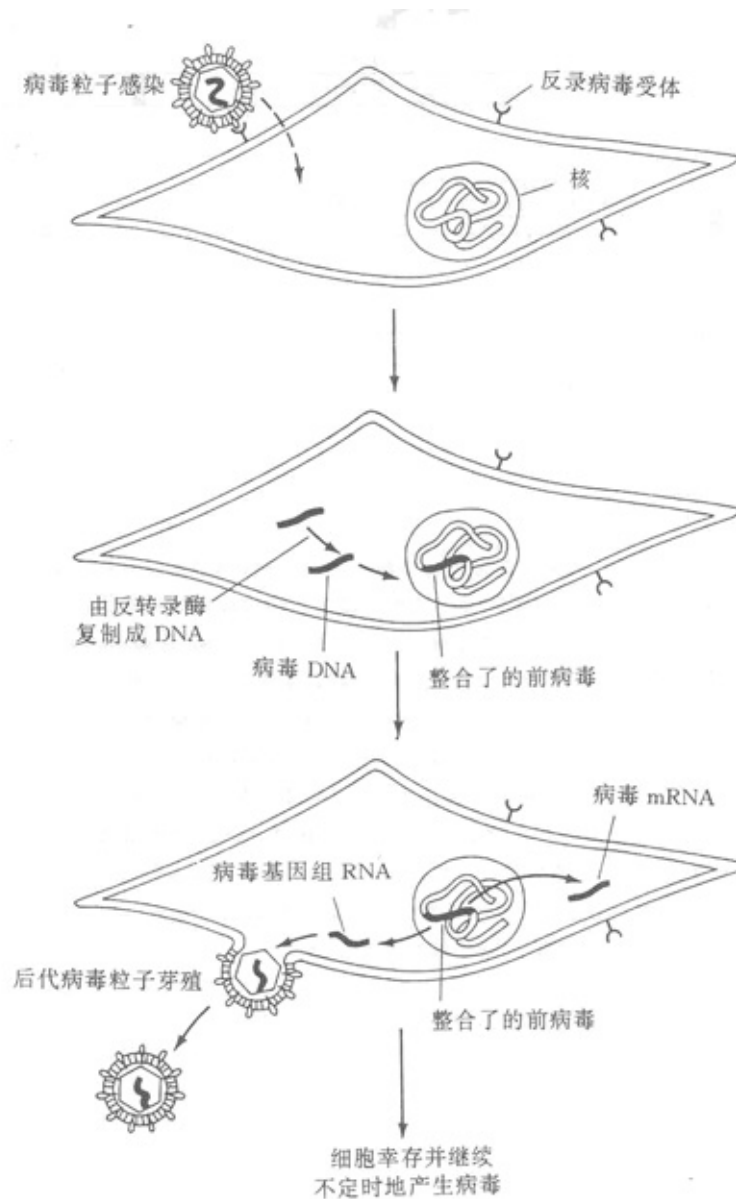


图7-4 一个反录病毒生命周期的概貌。反录病毒的病毒粒子附着到细胞表面的一个受体上。基因组RNA注入细胞质中，在那里通过反转录酶的作用而被反转录。cDNA穿入细胞核，并且可能会整合到宿主细胞的基因组之中。整合进去的前病毒被转录成：(1)用于合成病毒蛋白的mRNA，和(2)基因组RNA。该基因组RNA与结构蛋白质和酶性蛋白质装配成有感染性的病毒粒子，以出芽的形式穿出细胞膜。自Watson等(1987)。

1989)，也曾在植物 *Oenothera berteriana* 的线粒体基因组中发现(Schuster和Brennicke, 1987)，它们的开放阅读框架中有与其他反转录酶基因类似的顺序。然而，反录子不切离，因而是构成基因组整体的必要部分。与前病毒不同，反录子没有LTR，也不能构成病毒粒子。

拟反录病毒(pararetroviruses)，如乙型肝炎病毒，是结构上类似于反录病毒，但已失去了将自己插入宿主基因组中的能力的一类病毒。由于这个缘故它们是不够作为可转座因子的资格的，虽然它们显然与反录病毒有着共同的进化起源。

所有反录因子的反转录酶都有某些氨基酸等同性，这一事实表明了这些因子有一种共同的进化起源。由于反录子与反录病毒的复杂结构相反具有简单性，又由于细菌的古老性，所以，特明(Temin, 1989)认为，进化的途径是从反录子到反转录子、到反录转座子、到反录病毒、到拟反录病

毒(图7-6)。当然,某些现存的反录转座子有可能是从反录病毒衍生而来,而不是从它周围的另一通路衍生而来。

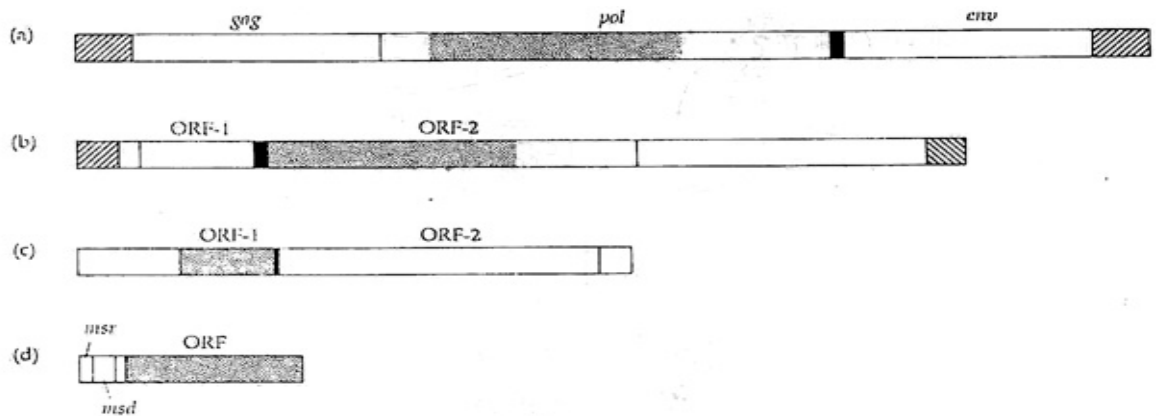


图7-5 反录因子的模式结构。打点的区域表示为反转录酶编码的区域;相邻基因间的重叠区则用黑矩形块表示;画有斜条纹的区域表示长末端重复(LTR)。ORF,即开放阅读框架。(a)猫白血病反录病毒。编码区的两侧有两个不等长的LTR,各长482和472bp。该编码区为两个多重蛋白前体编码。由gag-pol区编码的多重蛋白前体将分裂成两个多重蛋白,分别对应于gag和pol。gag多重蛋白产生4种内部病毒蛋白质,分别用p15,p12,p27,p10表示。pol多重蛋白则分裂成3种酶,一种蛋白酶,一种反转录酶,和一种内切核酸酶/整合酶。由env编码的多重蛋白前体分裂成两种被膜蛋白,分别用p70和p15表示。(b)粘性霉菌(Dictyostelium discoideum)的反录转座子DIRS-1。颠倒的LTR长为200-350bp。ORF-2含有一个顺序与反录病毒的pol基因类似的区域;(c)果蝇D.melanogaster的反录转座子G3A。ORF-1含有一个顺序与反录病毒的pol基因类似的区域。注意LTR缺乏这一特征。(d)来自粘球菌Myxococcus xanthus的反录转座子。其中,msr基因被转录成RNA;而msd基因则从互补的链上转录,然后再经由该反录转座子的ORF编码的反转录酶反录成DNA。反录产生的两个分子接着通过2'、5'磷酸二酯键而相互连接,从而形成被称为复拷贝单链DNA(msDNA)的分枝形分子。注意这里没有LTR。

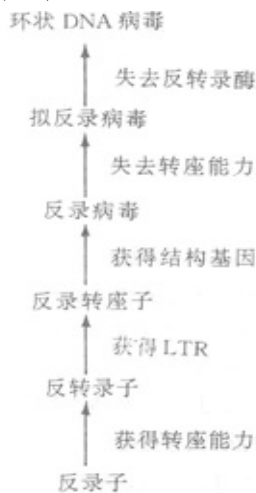


图7-6 反录因子的一种可能进化路线的模式图。

7.3 反录序列

反录序列(retrosequences或retrotranscripts),是通过RNA的反转录而得到、接着整合到基因组中、但缺乏产生反转录酶能力的基因组的序列(表7-1)。产生反录序列的模板通常是某一基因的RNA转录本。有些作者把反录序列称为“非病毒的反转录子”(例如Weiner等,1986)。一种产生反录序列的过程如图7-7所示。如果某一基因不在任何种系细胞中转录,那么,产生反录序列就需要RNA跨越细胞障碍。这可以通过以下方式实现,即RNA分子被包进某一反录病毒的病毒粒子

中，然后传送到种系细胞，并在那里被反转录 (Linial, 1987)。这一过程被命名为反录转染 (retrofection)。

由于反录序列起源于 RNA 序列，所以它们带有一些进行过 RNA 加工的标记，因而又称为加工后序列 (processed sequences)。反录序列的特征包括：(1) 缺少内含子，(2) 与基因的转录区域有精确一致的边界，(3) 在 3' 端有 poly A 延伸物，(4) 在两端都有短的顺向重复，这指示可能已涉及转座，(5) 各种转录后的修饰，如短核苷酸延伸物的加入或去除，以及 (6) 序列所在染色体上的位置已不同于转录产生 RNA 的原始基因的基因座位。

存在有两种类型的反录序列：加工后基因 (或反录基因) 和加工后假基因 (或反录假基因)

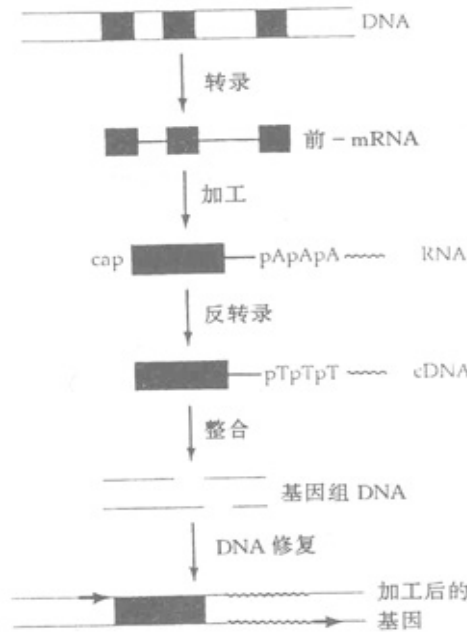


图 7-7 加工后反录序列的产生。黑块状体表示外显子。波纹线表示 mRNA 中的多聚 A 尾巴和 cDNA 中的互补多聚 T。DNA 被转录成前体 mRNA，然后被加工成 mRNA。mRNA 再被反转录成 cDNA，cDNA 则整合进基因组 DNA 中。裂缝被修复，这样在插入的反录序列两段侧就产生了两段顺向的短重复 (黑水平箭头)。如果 cDNA 从合成 RNA 的细胞转插入到一个不同细胞的基因组中，则产生反录序列的过程就需要 mRNA 包装进一个反录病毒粒子中，并且转移到那个靶细胞里。这样的过程称反录转染。

反录基因

加工后基因 (processed gene) 或反录基因 (retrogene) 是一种有功能的反录序列，它产生的蛋白质与产生该反录基因的原基因所产生的蛋白质等同或接近等同。有几个原因可以说明为什么一个反转录的基因保留其功能的可能性很小。第一，反转录过程是很不精确的，以至 RNA 模板与 cDNA 之间可能会出现许多差异 (突变)。第二，除非加工后基因是从一个由 RNA 多聚酶 III 转录的基因衍生而来，否则它通常不含有那些位于不转录区中的必需调控序列。第三，加工后基因有可能被插入到不适于其正确表达的基因组位置上。事实上，在绝大多数情况下，一个加工后基因就是一个“垂死者”。

令人惊奇的是，加工后的有功能基因已被发现，虽然它们看来是非常罕见的。人的磷酸甘油酸激酶 (PGK) 多家族由一个活性的 X-连锁基因、一个加工后的 X-连锁基因和一个额外的常染色体基因所组成。该 X-连锁基因含 11 个外显子和 10 个内含子。另一方面，它的常染色体同源物却是不寻常的，它没有内含子并且它的 3' 端由一个 poly-A 尾巴的剩余物构成侧翼，这极有力地表明曾发生过涉及 mRNA 的反转录过程。有趣的是，该常染色体的 PGK 基因几乎独特地在睾丸中表达。于是，反转录而成的 PGK 基因不仅保留了完整的阅读框架和转录并产生有功能的多肽的能力，而且还获得了一种新的组织特异性 (McCarrey 和 Thomas, 1987)。鸡中的肌肉特异的钙调素 (calmodulin) 基因也是无内含子的，显然也是经由反转录酶中介的事件而产生的 (Gruskin 等, 1987)。

大鼠和小鼠的前胰岛素原 I 基因可能是半加工反录基因 (semiprocessed retrogene) 的一个代表性例子。该基因在 5' 不翻译区中有一个长 119bp 的内含子。相比之下, 它的同源物前胰岛素原 II, 含有与它同样的小内含子, 另外还有一个存在于 c 肽编码区中的较大 (499bp) 内含子。来自其他哺乳类, 包括其他啮齿类在内的所有前胰岛素原基因也含有两个内含子。而且, 前胰岛素原 I 基因两侧由短重复夹拥, 且在多聚腺苷化信号后有一段短多聚 A (poly-A) 区 (Soares 等, 1985)。这些特征表明, 前胰岛素原 I 基因可能是从被部分加工的前胰岛素原 II 基因的前体 mRNA 衍生而来的半加工反录基因。事实上, 根据这两个前胰岛素原基因间的比较, 前胰岛素原 I 看来是从一个偏离了正轨的前体 mRNA 转录本衍生而来的, 该转录本起始于正常帽子部位上游的 500 个碱基对处, 且只有第一个内含子从它里面切离了。正是因为该偏离了正轨的转录本含有未被正常转录的 5' 调控序列, 反录基因在整合到一个新基因组位置后才保留了它的功能。

加工后假基因

加工后假基因 (processed pseudogene) 或反录假基因 (retropseudogene) 是一个失去其功能的反录序列。它带有一切有功能的反录序列的标记, 但却有一些妨碍其表达的分子缺陷。一个有功

SOD-1	M	A	T	K	A	V	C	V	L	K	G	D	G	P	V	
	ATG	GCG	ACG	AAG	GCC	GTG	TGC	GTG	CTG	AAG	GGC	GAC	GGC	CCA	GTG	
ψ69.1	ATA	ATG	ATG	AAG	GTC	ATG	TAC	ATG	TTG	AAG	GGC	CAG	AGC	CCG	GTG	
	I	M	M	K	V	M	Y	M	L	K	G	Q	S	P	V	
SOD-1	Q	G	I	I	N	F	E	Q	K			E	S	N	G	
	CAG	GCG	ATC	ATC	AAT	TTC	GAC	CAG	AAG	G	intron	AA	AGT	AAT	GGA	
ψ69.1	CAG	GCG	A	C	ATC	CAT	TT	GAG	CAG	AAG	G		AA	AAT	GAA	
	Q	V	T	S	I	*	**									
SOD-1	P	V	K	V	W	G	S	I	K	G	L	T	E	G	L	
	CCA	GTG	AAG	GTG	TGG	GGA	A	GC	ATT	AAA	GGA	CTG	ACT	GAA	GGC	CTG
ψ69.1	CCA	TTT	ATG	GTG	T	C	AGA	ATGC	ATT	ACA	GGA	TTG	ACT	GAA	CGC	CAG
SOD-1	H	G	F	H	V	H	E	F	G	D	N	T	A			
	CAT	GGA	TTC	CAT	GTT	CAT	GAG	TTT	GGA	GAT	AAT	ACA	GCA	intron		
ψ69.1	CAC	AGA	TTC	CAT	GTT	CAT	CAG	TTT	GGA	G	T	A	T	AAC	ACA	

图 7-8 人的 Cu/Zn 超氧化物歧化酶基因 (SOD-1) 的前两个外显子与一个加工后假基因 (ψ69.1) 的同源部分间的比较。圆点表示替换, “-”表示缺失, “+”号表示插入。注意内含子的缺乏和成熟前终止密码子 (用星号指示出)。关于单字母氨基酸缩写可参阅表 1-1。资料自 Danciger 等 (1986)。

能的基因和一个加工后假基因间的比较如图 7-8 所示。许多加工后假基因在反录转染期间被截去尾部; 加工后 mRNA 的 5' 截尾特别普遍, 但 3' 截尾也是曾有所闻的。5' 端被截去可能发生在: (1) 转录期间 (例如从正常位置的下游起始转录), (2) RNA 加工期间 (例如错误地拼接), 或 (3) 反转录期间 (例如酶在反转录进行到 RNA 分子的 3' 端之前失效, 该部位对应于 cDNA 的 5' 端)。

已经知道了从所有类型的 RNA (例如 mRNA, tRNA, rRNA, snRNA 和 7SL RNA) 衍生而来的加工后假基因。转运 RNA 特别有趣, 因为它们提供了一个最有说服力的证据, 表明加工后假基因事实上是通过 RNA 的反转录而派生出来的。所有细胞核 tRNA 都在 3' 端有一个 CCA 序列 (图 7-9)。该序列不由确定该 tRNA 的基因编码, 而是转录后经酶作用加上去的。对比之下, 基因组的加工后 tRNA 假基因则常常在 3' 末端有 CCA 顺序。

加工后假基因已在动物、植物、甚至细菌中发现。不过, 虽然加工后假基因在哺乳类中大量存在, 但它们在其它生物, 象鸡、两栖类和果蝇中却相对地较为稀有。表 7-2 列出了人类和啮齿类中的某些加工后假基因, 这些假基因以及它们的功能基因的数目都是已知的, 或已估出的。平均下来, 在这

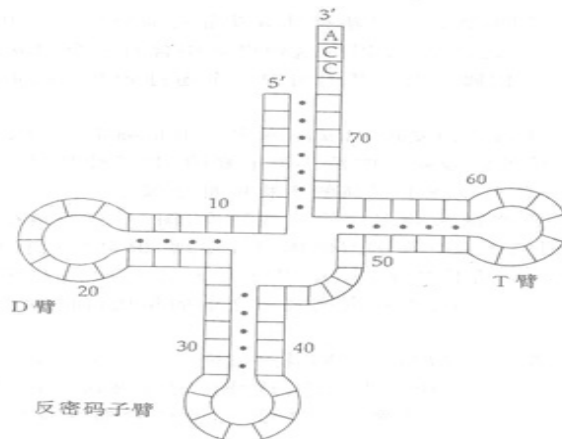


图 7-9 一个 tRNA 分子的三叶草式结构。它的 3' 端上的顺序 CCA 是转录后添加到功能分子中的，但它常在 tRNA 假基因的基因组顺序中出现。

些物种中加工后假基因的数目比功能基因的数目要多。事实上，在许多情况下加工后假基因的数目甚至有可能是被低估了的，因为古老的加工后假基因可能在顺序上与其亲本基因发生了较大程度的歧化，以至用从其功能基因同源物制造出的分子探针已不再能将它们检出来了。

表 7-2 反录假基因和其亲本功能基因的数目

物种	基因	基因的数目	反录假基因的数目
人	精氨酸琥珀酸合成酶	1	14
	β -肌动蛋白	1	~20
	β -微管蛋白	2	15-20
	Cu/Zn 超氧化物歧化酶	1	≥ 4
	细胞色素 C	2	20-30
	二氢叶酸还原酶	1	~5
	非肌的原肌球蛋白	1	≥ 3
	甘油醛-3-磷酸脱氢酶	1	~25
	磷酸甘油酸激酶	2*	1
	核糖体蛋白 L32	1	~20
	磷酸丙糖异构酶	1	5-6
	α -珠蛋白	2	1
	细胞角蛋白内 A (cytokeratin endo A)	1	1
小鼠	甘油醛-3-磷酸脱氢酶	1	~200
	肌球蛋白轻链	1	1
	鸦片黑素皮质激素原	1	1
	核糖体蛋白 L7	1-2	≥ 20
	核糖体蛋白 L30	1	≥ 15
	核糖体蛋白 L32	1	16-20
	肿瘤抗原 P53	1	1
大鼠	α -微管蛋白	2	10-20
	细胞色素 C	1	20-30

自 Weiner 等, (1986) a、其中一个反录基因。

在有些情况下，加工后假基因的数目可以超过其有功能对应物数目达几个数量级。Alu 家族就是一个这样的例子，该家族之所以如此命名，是因为这种序列中含有一个 Alu I 内切核酸酶的特征性限制位点。Alu 序列长约为 300 bp，它们属于人类基因组中一个超过 50 万次的重复序列家族，以构成基因组的

5—6%而引人注目。

乌卢和楚迪 (Ullu 和 Tschudi, 1984) 发现, Alu 序列实际上是确定 7S L RNA 的基因的加工后假基因。7S L RNA 在切除分泌蛋白质的信号顺序中是至关重要的。其活性基因受着严格限制, 而且它的顺序在象人、爪蟾以及果蝇这样一些分歧程度很高的生物中都是保守的。人的 Alu 序列是从 7S L 序列经过一系列步聚, 包括一次重复, 两次缺失和多次核苷酸替换 (图 7-10 a) 衍生而来的。大多数的人 Alu 序列有一个二聚体结构。人基因组还含有许多四聚体 Alu 序列, 但至今只有几个单体 Alu 因子曾在人类中发现。相比之下, 啮齿类的 Alu 等价物 B1 家族, 几乎是绝对地单体的 (图 7-10 b)。布里滕等 (Britten 等, 1988) 将第一个单体出现的时间定在哺乳类辐射之前的年代, 而把产生二聚体的重复所处的年代定在灵长类谱系建立以后。

如果一个反录假基因保留着被转录的能力, 则其后可能会有一个倾落过程跟随, 藉此新的反录假基因即从现存反录假基因的 RNA 转录中产生出来。这种情况曾被认为在 Alu 家族中发生过 (Bains, 1986)。我们注意到, 7S L 基因是被 RNA 多聚酶 III 转录的, 而该酶则并不需要转录区外的启动子。因此, 某些 Alu 序列保留了完整的启动子并且连续不断地被转录, 这是可以想象到的。不过, 威拉德等 (Willard 等, 1987) 和布里滕等 (Britten 等, 1988) 仅认出了 Alu 序列的几个亚家族。布里滕等 (Britten 等, 1988)

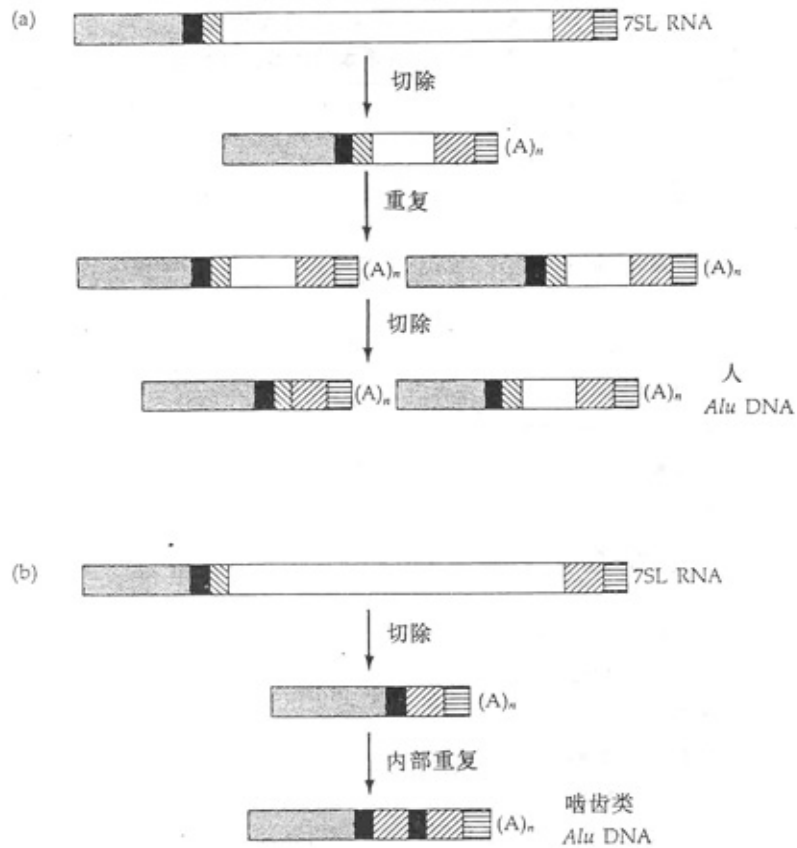


图 7-10 Alu 序列的起源, (a) 在人和其他灵长类中, (b) 在啮齿类中。7S L RNA 基因中的不同区域用不同的阴影图案区别, 以强调 Alu 序列中的缺失和重排。(A)_n 表示 A 被重复 n 次。注意在 (a) 中的二聚体结构和在 (b) 中的单体结构。

提出, 这些亚家族是从 4 个源基因逐步衍生的 (图 7-11)。更近期衍生出来的亚家族, 其与祖先 7S L 序列在顺序上的分歧比古老一些亚家族的更大。所以结论是, Alu 序列不是直接从 7S L 功能基因衍生而来, 而是从少数源序列衍生而来的, 这些源序列则是从 7S L RNA 序列经过许多变化步骤而产生的。这四种源基因各在这一时刻或那一时刻起着 Alu 序列的主导源作用, 并各被某一后代谱系所替代。固定的连续波虽不在突然的爆发后出现, 但连续的亚家族却继续在基因组中长期共存 (另见 Quentin, 1988)。

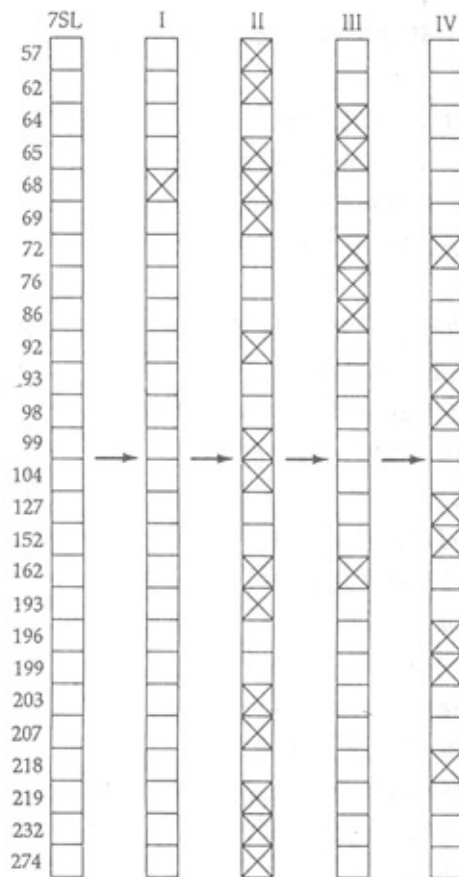


图 7 - 1 1 Alu 序列的不同亚家族间在已判断出的位置 (有阿拉伯数字) 上的突变次序。罗马数字指示在不同时期起着 Alu 序列的主导源作用的连续的源基因。将每一个亚家族与其前面一个区别开来的替换用 X 表示。资料自 Britten 等 (1988)。

由于反转录的泛有性，哺乳动物的基因组简直是处在反录序列的拷贝的困扰之中。绝大多数的这类拷贝从它们整合到基因组中时起就是无功能的。而且，这样一些序列也不能通过基因转变而轻易地复苏，因为它们大多数位于距其亲本功能基因很远的染色体位置上 (第六章)。一个有功能的基因座位泵出它自己的有缺陷拷贝、并将它们散布在整个基因组中的现象，可以与火山产生岩浆作类比，因此该过程曾被命名为进化的维苏威模式 (Vesuvian mode of evolution) (P. Leder, 在 Lewin, 1981 中被引用)。

加工后假基因的进化

正如前面所说的，一旦加工后假基因作为基因组中的一种染色体序列而存在后，它就是无功能的并不受一切选择的限制。由于缺少功能，所以假基因受两种进化过程的影响 (Graur 等, 1989b)。第一种涉及点突变的极迅速的积累。这种积累最终会抹去假基因与其功能同源物间的顺序相似性，因为后者的进化要缓慢得多。假基因的核苷酸组成将变得与其无功能的近邻越来越相象，以至最终它将“混入”其中。这一过程曾被称为组成同化 (compositional assimilation)。

第二种进化过程的特征是，与其功能基因相比，假基因会变得越来越短。这种长度缩短 (length abridgement) 是由于缺失超过插入而造成的。曾有人估计，哺乳动物的加工后假基因在大约 4 亿年的时间里失去了约一半的 DNA。这一过程是如此地缓慢，以至在人的基因组中还含有那些在其非常远古的祖先中发现的假基因 DNA 的主要部分就是一个很好的例子。显然，这些古老的假基因时至今日几乎已经失去了与其从之发端的功能基因的一切相似性。

总而言之，看来加工后基因产生的速率比它们通过缺失而被抹去的速率要快得多。所以结论是，缩短的过程进行得太慢了，以至于无法抵消继不断的维苏威式的轰击之后出现的基因组大小上的增加 (第八章; Graner 等, 1989b)。

7.4 转座对宿主基因组的影响

转座和反录转座可能对基因组的大小和结构有着深远的影响。尤其是，可转座因子已被看成是“自私DNA”(selfish DNA)的最好例子；这里自私DNA可能不带给宿任何利益，却因为它比基因组的序列增殖迅速而能在基因组中传播(Doolittle 和 Sapienza, 1980; Orgel and Crick, 1980)。为此缘故，转座能极大地增加基因组的大小。这种效应将在第八章中探讨。这里，我们将只关心可转座因子影响基因进化和表达的方式。

第一，如以上所讲到的，细菌中的转座子常常带有能给予携带者以抗生素抗性或其他抗性的基因。于是，转座子也许能使该宿主物种在某种逆境中生存。

第二，一个基因的表达也许会因为可转座因子存在于该基因之中或其邻近而改变。情况最简单的是，可转座因子插入到一个为蛋白质编码基因的编码区中，这将极有可能改变阅读框架，因而可能会有激烈的表型效应。类似地，可转座因子的切离可能是不精确的，结果就会导至碱基的增加或缺失。然而，也有预料不到的效应。例如，一个可转座因子也许会含有调控因子，象启动子，这就会影响邻近基因的转录频率。事实上，反录病毒的LTR中常常含有很强的增强子，它们将对邻近基因的表达产生很大的影响。类似地，酿酒酵母 *Saccharomyces cerevisial* 中的Ty(表示“transposon yeast”)因子已知能增加下游基因的表达。这在某些特殊环境中可能是有利的，虽然在大多数情况下，由这类变化造成的代谢不平衡极有可能是有害的。含有拼接给体或受体的可转座因子，即使被掺入到基因的非编码区，如某一内含子之中，也有可能影响原初RNA转录本的加工。

第三，许多可转座因子能促成大型的基因组重排。倒位、易位，重复以及大段的缺失和插入可以经可转座因子的中介而发生。这些重排能以转座的直接后果发生(即，通过DNA片段从一个基因组位置向另一位置的运动)，并且能改变它们进行表达时的环境。作为转座的结果，若两个先前相互间很少相似性的序列现在能共享某一相似的可转座因子，以至它们间有可能进行不等价交换，则随后将会出现较间接的效应。图7-12示意了，由于复本Alu序列存在于低密度脂蛋白受体基因的外显子5侧面的内含子中，从而促成了一次不等价交换事件，而这样一次事件又是怎样导至一个缺少该外显子的突

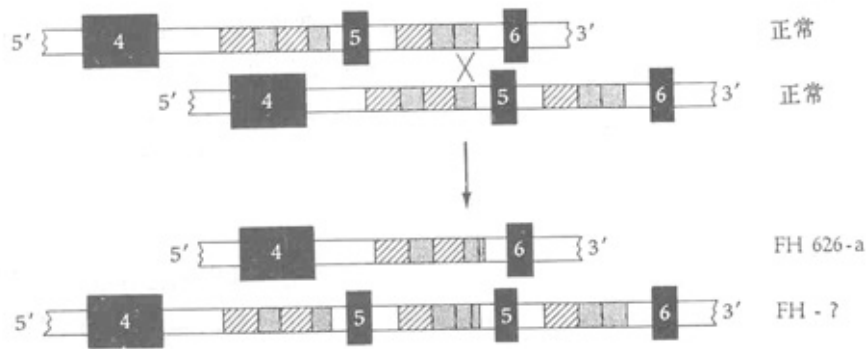


图7-12 低密度脂蛋白受体基因中的不等价交换。外显子用黑矩形块表示并加以编号。内含子中的Alu序列用画斜条纹(右臂)和加点(左臂)来表示；一个Alu序列可以由两个臂(二聚体的)也可以是由一个臂(单体的)所构成的。设想的交换位置用x指示。重组后的缺失(观察到的)和插入(推测的)产物被描绘成FH626-a和FH-?,位于箭头之下。自Hobbs等(1986)。

变型基因的产生的(Hobbs等, 1986)。在低密度脂蛋白基因中，由同样机制造成的外显子1-4的缺失也已被观察到(Lehrman等, 1986)。对这些缺失呈纯合的病人在血液中有高水平的胆固醇(高胆固醇血症)。两个Alu因子间的重组还被证明是造成患腺苷脱氨酶缺陷症病人的腺苷脱氨酶基因缺失启动子和第一个外显子的原因(Markert等, 1988)。一般而言，对于所有含Alu重复序列的区域，基因组不稳定性都已得到证实(Calabretta等, 1982)。

第四，有证据表明，某些可转座因子可能会造成突变率增加。例如，含有可转座因子Tn10的E. coli品系被发现插入速率升高了(Chao等, 1983)。在大多数情况下，这种性状对携带者来说将是有害的。然而，在严峻的环境压力下，突变率提高也有可能是有利的，因为有些突变也许能更好地适应新的环境，而

它们的携带者也将比非携带者有更高的适合度。

7.5 杂种劣势

果蝇中的杂种劣势 (hybrid dysgenesis) 是由一些相关的异常遗传性状造成的症候群, 这些性状在某些相互作用的品系间的一个杂种类型中自发地诱生出来, 但在相反方向交配的杂种中则通常不会出现 (Sved, 1976; Kidwell 和 Kidwell, 1976)。杂种劣势曾引起分子生物学家和进化生物学家们的浓厚兴趣, 因为已经发现它是由可转座因子造成的, 且它的主要特征是产生阻止品系间或群体间杂交的屏障, 而这种屏障则曾被推测是物种形成的一个原因 (见后)。

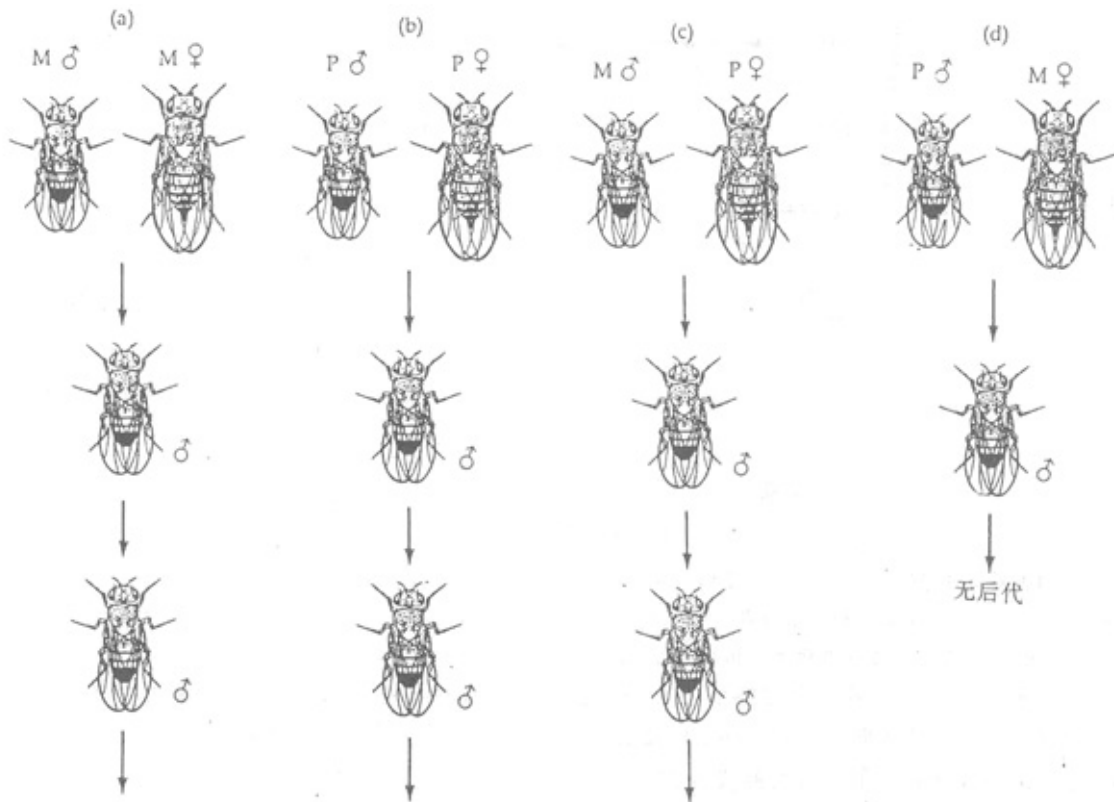


图 7-13 果蝇中的杂种劣势。(a) M品系内交配和 (b) P品系内交配产生正常后代。(c) 正常后代也能在 M雄体和 P雌体的杂交中产生。(d) 在 P雄体和 M雌体的杂交中, 产生了劣势的后代, 它们中许多是不育的。

果蝇中有几种杂种劣势系统, 下面我们将仅对其中之一, P—M系统进行探讨。杂种劣势的不对称性如图 7-13 所示。当一个 P品系的雄体与一个 M品系的雌体交配时, 其后代是劣势的; 而在相反方向的交配中, 其后代则是正常的。P—M系统的劣势性状包括: (1) 某些个体在一定的条件下生殖腺发育不全, (2) 在雄体中出现重组 (在果蝇中重组通常被限制在雌体中, 所以这是一种非自然的现象), (3) 染色体发生断裂, (4) 偏离了孟德尔传递比率 (即, 相对于 P染色体而言 M染色体传给后代的倾向性要大一些), 和 (5) 出现高突变频率。

P—M系统劣势的原因来自一个被称为 P因子的可转座因子的家族 (图 7-3)。在 P品系中, 基因组中有 30—50 个 P因子, 但它们中许多都可能含有缺失。它们分布于所有染色体中。虽然在有些品系中, 转座表现得有点偏爱 X染色体。M品系则不带 P因子。杂种劣势系统的不对称性被认为是这样造成的, 由于母性遗传, P雌体×M雄体交配的 F₁子代中存在着 P因子编码的抑制物, 而相反方向的交配 M雌体×P雄本, 其 F₁子代中则缺少这种抑制物。P因子可以在从母系的细胞质中得到缺乏该抑制物的种系细胞中转座, 杂种劣势是与这种转座相关联的。在杂种劣势的背景下, 细胞质中抑制物的存在与否定义了合子形成后的反应类型, 因而被命名为细胞型 (cytotype)。在抑制物存在下, P因子的转座可能被完全或部分地抑制。正常情况下体细胞中不发生 P因子的转座, 因为体细胞中编码区中的第 3 个内含子不象在种系细胞中那样会被切离。

基德韦尔 (Kidwell, 1983) 观察到的一个有趣现象涉及带有 P 的品系的分布。在 1950 年以前收集的任何 *D. melanogaster* 果蝇品系中都未曾观察到 P 性状，而随后收集到的品系则表现出 P 的频率增

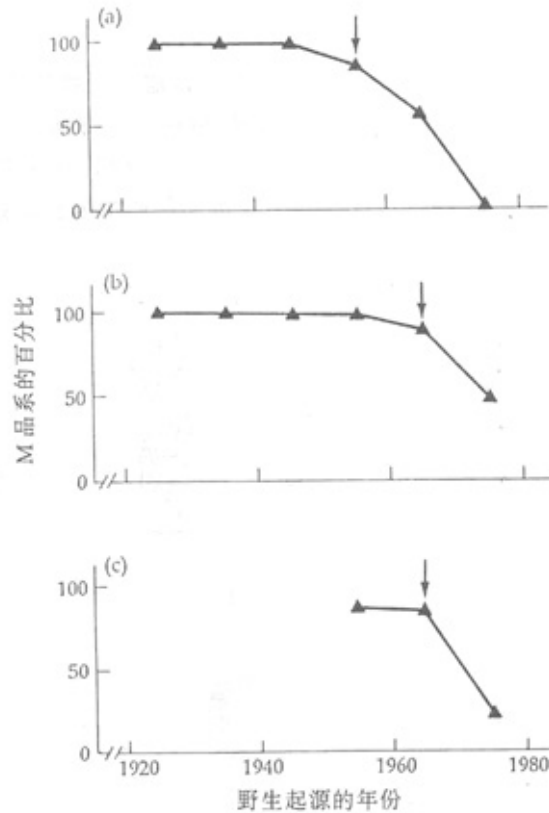


图 7-14 自然群体中 M 品系 (即无 P 因子的品系) 频率变化, (a) 在北美和南美的自然群体中, (b) 在欧洲、非洲和中东的自然群体中, (c) 在澳洲和远东的自然群体中。箭头指示 P 品系的第一次出现。注意, M 群体的比例下降最初出现在美洲大陆, 只是后来才在其他大陆出现下降。自 Kidwell (1983) 修改而成。

加和寿命降低 (图 7-14)。另一个劣势系统 I-R, 已得到类似的观察结果。已有两种假说被提出来, 以解释 P 因子的分布。恩格斯 (Engels, 1981b) 提出, 自然界中大多数品系是 P 型的, 但它们在实验室群体中有失去该转座子的倾向。第二种假说则设想, P 转座子是近来才导入 *D. melanogaster* 群体的, 随后 P 因子在以前的 M 群体中迅速传播 (Kidwell, 1979)。有几个原因可以说明为什么第二种假说看来要更合理一些。首先, 带有 P 的实验室品系曾被监测了近 15 年, 并未失去它们的 P 特性。其次, 在 P 品系的分布上看来有一个地理渐变群, 北美群体表现得比某些欧洲、非洲和亚洲品系更早地带有 P 特性。最后, 现在已有证据表明, *D. melanogaster* 近来已从一个亲缘关系疏远的物种那里获得了 P 因子 (见第 117 页)

7.6 转座与物种形成

物种形成 (speciation) 或**分枝进化** (cladogenesis) (即从一个亲本物种产生两个或多个物种) 是最重要的进化过程之一。遗憾的是, 在分子水平上, 它也是我们了解得最少的进化过程之一。我们不知道新物种是通过什么方式而从老物种中产生的。我们所知道的只是, 物种形成的过程需要在两个属于同一物种的群体间产生生殖障碍, 使得它们就不再能杂交。杂种劣势有一阵子曾被认为是物种形成过程的早期阶段, 起着属同一物种的不同群体间交配后产生生殖隔离的机制的作用。事实上, 姐妹种 *D. melanogaster* 和 *D. simulans* 间杂交产生的杂种不育就与劣势非常类似 (例如, 性腺发育不良, 分离扭曲等)。

不过, 此观点有几个问题有待解决。第一, 虽然杂种表现出适合度降低, 因而从生殖角度看是部分被隔离的, 但 P 因子在种系细胞中的转座实际上保证了大多数传给杂种的染色体将带有 P 因子, 且细胞型也最终会变成 P 型。于是, 倘若杂种的适合度降低不是太大的话, 则 P 因子将会传遍整个群体。事实上, 因

为有效的生殖隔离，杂种最终将是几乎完全不育的。第二，P因子有作为传染因子而从一个个体向另一个个体水平转座的能力（见第117页原版）。于是，整个群体可能会被P迅速接管，这样杂种劣势看来在自然界中只能维持非常短的时期。事实上，许多果蝇种已知其所有个体和所有群体都带有P因子或类P因子，因而杂种劣势将不会在这些种中的任何一种中出现。最后，就我们所知，杂种劣势限制在果蝇范围，因而可能并不代表自然界中的普遍现象。即使在果蝇中，也没发现易动因子与阻碍姐妹种间基因流动有关的证据。

由于可转座因子的发现，许多其他关于由转座导致物种形成的机制也在文献中出现。例如，曾有人提出，一个群体中的含有调控序列的因子，它的大规模复制型转座可能会引起所谓基因组的遗传重排（genetic resetting），藉此许多基因将会经受一种新的调控形式。这样一个群体显然将变得与保留着旧调控形式的群体发生生殖隔离。另一种说法则求助于机制不相容性（mechanical incompatibility），它也是由大规模复制型转座所造成的。在这种情况下，假定一个群体中的可转座因子曾发生增殖，并达到使染色体大小出现显著增大的程度。于是，从一个亲本那里继承大染色体而从另一个那里得到小染色体的杂种生物体，在减数分裂期间将会经历染色体配对上的困难，而这极有可能造成不育。遗憾的是，迄今为止所提及的物种形成模型，没有一个曾得到经验数据的支持。

7.7 可转座因子拷贝数的进化动力学

一个基因组中的可转座因子的拷贝由3个因素所决定。（1） u ，一个可转座因子产生一个新基因组拷贝的概率（即复制型转座的概率），（2） v ，该因子被切离的概率，以及（3）对抗增加可转座因子在基因组中的数目的选择强度。关于果蝇 *D. melanogaster* 群体中几个可转座因子的 u 和 v 值，已经由实验得出，转座频率被发现在不同可转座因子间有些变化，但平均起来则处在每因子每世代 10^{-4} 这一数量级。切离频率大约要低一个数量级（Charlesworth 和 Langley, 1989）。因此，在没有对抗可转座因子转座的选择的情况下，基因组中的拷贝数预期将无限制地增加。

如果可转座因子的数目维持在一个平衡位置上，这可能是一个在自然界不成立的假定，那么，选择就必须发生作用以对抗拷贝数的增加。在最简单的决定性模型中，我们假定，个体的适合度 w 将随拷贝数 n 的增加而降低。此假定的合理性是，可转座因子的插入频频地改变着邻近基因的表达；随着可转座因子的数目的增加，基因表达出现有害变化的概率也会增加。可以证明，只要 w 值随着 n 的增加而降低，则不管 n 和 w 间的精确关系如何，群体在平衡位置相对于缺少可转座因子的个体而言的平均适合度为

$$\bar{w} = e^{-n(u-v)} \quad (7.1)$$

（Charlesworth, 1985）。

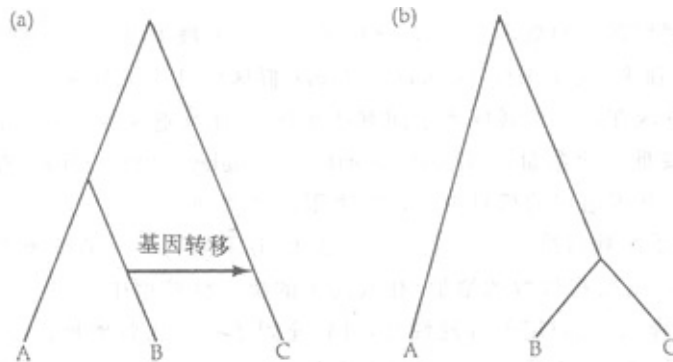
在果蝇 *D. melanogaster* 的情况中，有约50个转座子家族，每个家族平均在基因组中出现10次（Finnegan 和 Fawcett, 1986），于是 $n=500$ 。因为 v 比 u 至少小一个数量级，所以 $u-v \approx u = 10^{-4}$ 。解等式7.1，我们得 $\bar{w}=0.95$ 。于是，适合度上的降低为 $s=1-0.95=0.05$ 。由等式7.1给出的平衡的稳定性，要求适合度的对数衰减比 n 的线性增加更陡峭（Charlesworth, 1985）。不过，为了运算简单，我们假定呈线性，那么，每增加一个转座子适合度的降低近似为 $0.05/500=10^{-4}$ 。这样小的选择系数本质上意味着，一个生物体中的可转座因子的拷贝数主要是由随机遗传漂变所决定的。如果一个生物体含有数目更大的可转座因子，甚至在果蝇中可转座因子数大大超过500的可能性也是存在的（Rubin, 1983），那么，一个转座子对适合度的效应也许要比上面所得到的更小。

关于对抗拷贝数增加的选择的一个可采用形式是自我调控转座机制，即转座的频率随拷贝数的增加而降低，或切离的频率随拷贝数的增加而增加（见 Charlesworth 和 Langley, 1989）。

7.8 水平基因转移

水平基因转移 (horizontal gene transfer) 被定义成遗传信息从一个基因组向另一个基因组，特别是在两个物种之间的转移。这个术语的提出是为了把这类转移与通常的“垂直转移”区别开来，后者则是亲代把遗传信息传给子代的转移。水平基因转移要求，(1) 在各生物体和各细胞间运输遗传信息的交通工具，和(2) 将外源 DNA 片段插入宿主基因组中的分子机制。反录病毒能完成这两个任务，因为它们既能把染色体的 DNA 掺和到它们的基因组中，又能跨越物种界线 (Benveniste 和 Todaro, 1976; Bishop, 1981)。在转座子以及其他类型的 DNA 中介的转座中，跨越细胞的传输必须由一个传染因子来提供，如质粒。事实上，许多自然界中出现的质粒都含有可转座因子，这些因子可以从质粒上离开而向细菌染色体移动，也可相反方向地移动。

通过发现某一具体基因的系统发育分布中的显著不连续性，也许能检出一个水平基因转移事件。例如，细菌 *Salmonella typhimurium* 含有一个类组蛋白基因，据我们所知它在其他细菌中没有对应物 (Higgins 和 Hillyard, 1988)。当发现基因系统发育和物种系统发育之间明显地存在矛盾时，特别是顺序类似性看来反映了地理上的近似而不是系统发育上的亲缘性时，或许也可以怀疑有水平基因转移。作为例子，我们考虑一下图 7-15 a 中的系统发育树。假定 B 从 A 那里分歧后，发生了一个 DNA 片段从物种 B 向物种 C 的转移。在除发生了水平转移的基因外的任何基因的顺序比较的基础上，我们可望得到一个正确的物种间系统发育关系。对比之下，如果我们采用发生了水平转移的 DNA 片段，则我们将得到如图 7-15 b 中那样的错误树。不过，我们也注意到，除水平基因转移外还有别的因素也能造成物种树和基因树之间的矛盾 (第五章)。



两类序列能被水平地转移：(1) 从可转座因子衍生出的图 7-15 水平基因转移的情况下系统树的构建。(a) 真实树，(b) 推测树。序列，和(2) 基因组序列。基因组序列的水平基因转移的例子中，很少有已被令人信服地证实了的。许多原来这样声称的例子，后来发现并不能得到分子证据的支持。而且，我们注意到，一个发生了水平转移的基因在宿主中保留其功能性的情况是很罕见的，预期比一个在同一物种内从一个基因组位置到另一个位置转移的基因 (见第 107 页) 更少出现。

病毒基因从狒狒到猫的水平转移

脊椎动物基因组中含有许多与反录病毒同源的序列。这些序列是真核生物细胞核 DNA 的正常组份，称为内源性反录病毒序列 (endogenous retroviral sequences) 或病毒基因 (virogene)。有几个内源性反录病毒序列在脊椎动物的种间转移的例子 (综述见 Benveniste, 1985)。其中一例与来自狒狒的 c 型病毒基因有关 (图 7-16)。

与狒狒的病毒基因同源的序列已在所有古世界猴的细胞 DNA 中检出。它们间的顺序类似性是物种间的分类学关系紧密相关的。于是，该病毒基因在灵长类中至少存在了 3 千万年。有趣的是 6 种与家猫 (*Felis catus*) 亲缘关系密切的猫也含有这一序列，虽然在亲缘关系较远的猫科动物、象狮、豹和短尾猫，以及任何其他食肉动物中，都不存在这一序列。所以，这一序列极有可能是在过去某一时刻在物种间水平转移的。水平传递的年代和方向可从两类资料中推出：(1) 顺序类似性和(2) 古地质学信息。

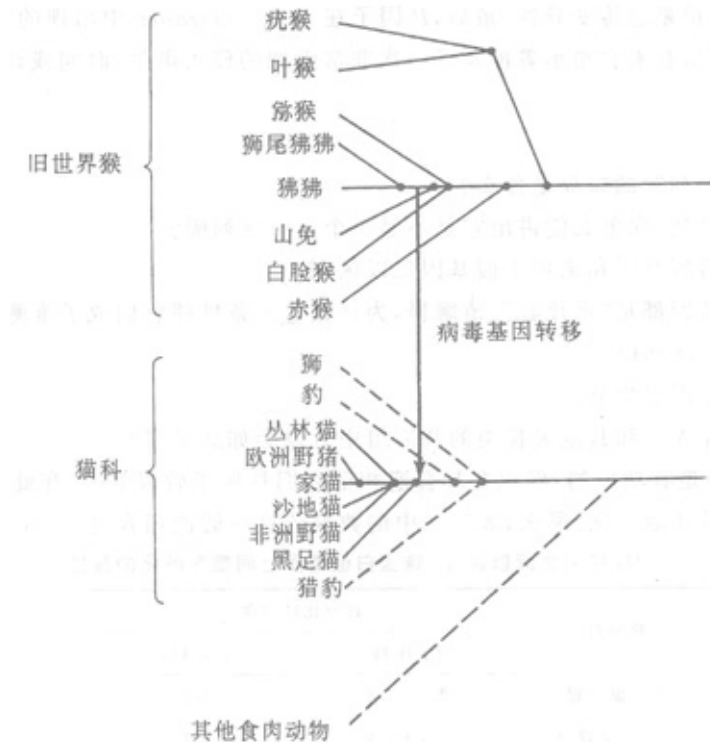


图 7-16 古世界猴和猫的系统发育树。含有 c 型病毒基因的物种用实线标出。来自狒狒和狮尾狒狒的 c 型病毒基因展现出与猫的 c 型病毒基因有最大的相似性。因此，在大约 1 千万前可能发生了一次水平基因转移，该转移是从狒狒和狮尾狒狒的祖先到与狮、豹和猎豹等的谱系分歧后的现代猫的祖先。自 Benveniste (1985) 修改而成。

所有含狒狒病毒基因的猫种都来自地中海地区，而来自东南亚、新世界和非洲的猫科动物则缺少该序列。因此，转移发生在猫科动物的大辐射之后，且被限制在一个地理区域中。这一结论指出，水平基因转移的年代在 500-1000 万年前。传递方向则可这样推出：一方面考虑该序列在灵长类中的分布，另一方面考虑它在猫类中的分布。因为所有古世界猴都具有该病毒基因，而只有几种猫才具有该基因。所以，只能假定猫类是从狒狒那里获得该序列，而不是相反，这才是合理的。这一结论在考虑到以下事实后得到强化：猫类与 3 种狒狒 (*Papio cynocephalus*, *P. papio* 和 *P. hamadryas*) 和亲缘关系密切的狮尾狒狒 (*Theropithecus gelada*) 中的病毒基因，它们的相似程度比猫类与任何其他灵长类序列的相似程度要更高一些。因此，该序列一定是狒狒和狮尾狒狒与山魈分歧后不久而从它们的祖先转移到猫类中去的 (图 7-16)。从对狒狒的研究中得出的年代与从对猫类的研究中得出的年代一致得相当好。

P 因子在果蝇的种间的水平转移

水平基因转移的另一个例子与 *D. melanogaster* 中的 P 因子有关。正如前面所讲的，在近 40 年的时间里 P 因子已迅速传遍了 *D. melanogaster* 的自然群体 (见第 114 页)。而 P 因子不存在于与 *melanogaster* 亲缘关系很近的一些种，象 *D. mauritania*, *D. sechellia*, *D. simulans* 和 *D. yakaba* 等之中。那么这些因子是从哪里来的呢？丹尼尔斯等 (Daniels 等, 1990) 曾对成百种果蝇进行了普查，结果表明，除 *D. melanogaster* 外，*melanogaster* 类亚组的任何其他种中均未发现 P 序列。相反，亲缘关系疏远的 *willistoni* 组和 *saltans* 组的所有种都含有 P 因子和类 P 因子。尤其是，*D. willistoni* 的 P 因子被发现除了一个碱基替换外完全等同于 *D. melanogaster* 中的因子，这指示 *D. willistoni* 在 P 因子转到 *D. melanogaster* 的水平基因转移中，起着供体物种的作用。

有几个原因使得我们怀疑这种水平基因转移是最近才发生的。首先，来自 *D. melanogaster* 和来自 *D. willistoni* 的 P 序列间的接近等同，表明发生分歧的时间非常短。其次，来自地理位置相距很远处的 *D. melanogaster*，它们的 P 序列间几乎没有遗传变异性，这指示自 P 因子导入 *D. melanogaster* 以来时间还很短，以至还来不及积累遗传变异性。最后，P 因子在 *D. melanogaster* 中出现的地理模式，以美洲

大陆中的群体最先获得它，看来它指示着涉及了一次非常近期的侵入事件，时间或许在近 50 年之内。

习题

- 1、反录因子和反录序列间的差异是什么？
- 2、给出一个基因组序列，你怎么能讲出它是不是一个反录序列呢？
- 3、怎样才能将加工后假基因和未加工假基因加以区别？
- 4、大多数加工后假基因都是“垂死者”。请解释，为什么这一条件使它们成了推测点突变模式的极好材料（见第四章和 Li 等, 1984）。
- 5、解释为什么反录基因很少见。
- 6、为什么 *Alu* 序列在人类和其他灵长类的基因组中含量会如此丰富？
- 7 *Alu* 序列曾被认为是有功能的，但现在却普遍相信它们是加工后假基因。在此假说下，*Alu* 序列应该象其他假基因那样迅速地进化。那么，表 7-3 中的数据与这一假说相容吗？

物种对	百分比纯化度	
	<i>Alu</i> 序列*	η 假基因
人对黑猩猩	2.2 ± 1.4	1.7
人对马来猩猩	3.7 ± 1.9	3.1

数据自 Koop 等(1986a)和 Li 等(1987a)

a. 7 个垂直相关的序列被用来计算平均值和标准偏差

- 8、试列举转座对宿主可能产生的有利和不利效应。

后继阅读文献

Berg, D. E. and M. M. Howe (eds). 1989. Mobile DNA. Academic Society for Microbiology, Washington DC.

Compbell, A. 1983. Transposons and their evolutionary significance. pp. 258-279. In M. Nei and R. K. Koehn (eds), Evolution of Genes and Proteins. Sinauer Associates, Sunderland, MA.

Charlesworth, B. and C. H. Langley, 1989. The population genetics of *Drosophila* transposable elements. *Annu. Rev. Genet.* 23:251-287.

Doolittle, R. F., D. F. Feng, M. S. Johnson and M. A. McClure. 1989. Origins and evolutionary relationships of retroviruses. *Quart. Rev. Biol.* 64:1-30.

Lewin, B. 1990. Genes IV. Oxford University press. Oxford.

Scaife, J., D. Leach and A. Galizzi (eds.). 1985. Genetics of Bacteria. Academic Press, New York.

Shapiro, J. A. (eds). 1983. Mobile Genetic Elements. Academic Press, New York.

Varmus, H. 1988. Retroviruses. *Science* 240:1427-1435.

Weiner, A. M., P. L. Deininger and A. Efstratiadis. 1986. Nonviral retroposons: Genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* 55:631-661