



## 真核生物的遗传分析

**真**菌、原生动物、藻类、植物和动物都是真核生物(**eucaryote**),它们的遗传物质主要集中在有核膜包围的细胞核中,**DNA**和特殊的蛋白质相结合组装成为染色体。通常一个基因组包含若干个染色体,每条染色体具有多个复制起点,含有大量重复序列和很多不编码序列。功能上相关的基因可以位于不同的染色体上,没有明显的操纵子结构,但存在不同类型的基因家族(**gene family**)。真核生物基因常常被内含子(**intron**)序列隔开,在剪接(**splicing**)过程中这些内含子被切除,剩余编码氨基酸的外显子(**exon**)连接起来构成一个成熟的 **mRNA**,随后 **mRNA** 翻译成蛋白质。无论是单细胞真核生物,还是多细胞真核生物,其细胞都是通过有丝分裂方式进行无性生殖,大多数真核生物通过减数分裂形成配子进行某种形式的有性生殖,随后配子结合形成合子,发育成为成熟的生物体。真核生物基因组同样呈现出不固定性,不仅有基因丢失、扩增与重排,而且转座成分首先是在真核生物中发现的。随着人类基因组计划的完成,迄今已有几十种真核生物完成基因组测序,尚有几百个物种的基因组计划正在进行。所取得的一系列令人瞩目的成就为破解生命密码、揭示其规律奠定了坚实的基础。本章重点是讨论真核生物的基因组、基因定位与染色体作图和真核生物同源重组的分子机制,并介绍基因丢失、扩增与重排及其遗传学效应。

## 6.1 真核生物基因组

### 6.1.1 C 值悖理

一个物种单倍体的染色体数目及其所携带的全部基因称为该物种的基因组 (**genome**)。因此,基因组是生物体内遗传信息的集合。一个物种基因组的 **DNA** 含量是相对恒定的,它通常称为该物种 **DNA** 的 **C 值 (C value)**,即单倍体所含 **DNA** 量。不同物种的 **C 值** 差异极大,最小的 **C 值** 是支原体 (**mycoplasma**),小于  $10^6$  **bp**,最大的是某些显花植物和两栖动物的 **C 值**,可达  $10^{11}$  **bp**。从原核生物到真核生物,其基因组大小和 **DNA** 含量是随着生物进化复杂程度的增加而稳步上升的。随着生物结构和功能复杂程度的增加,需要的基因数目和基因产物种类越多,因而 **C 值** 也越大(图 6-1)。

尽管 **C 值** 大小随着生物的复杂程度而增加,然而在结构与功能相似的同类生物中,以至亲缘关系很近的物种之间,它们的 **C 值** 差异仍可达 10 倍乃至上百倍。如两栖类、被子植物不同物种之间,其 **C 值** 小的低于  $10^9$  **bp**,大的达到  $10^{11}$  **bp**。如豌豆与蚕豆均属豆科,又都只有 12 条染色体,而 **DNA** 含量却相差 7 倍。特别是万物之灵的人类,其 **C 值** 只有  $10^9$  **bp** 而肺鱼的 **C 值** 则为  $10^{11}$  **bp**,居然比人高出 100 倍,很难设想两栖类、肺鱼的结构与功能比哺乳动物包括人类更复杂。由此表明 **C 值** 的大小并不能完全说明生物进化的程度和遗传复杂性的高低,也就是说,物种的 **C 值** 及其进化复杂性之间没有严格的对应关系,这种现象称为 **C 值悖理 (C value paradox)**,或 **C 值佯谬**(图 6-2)。

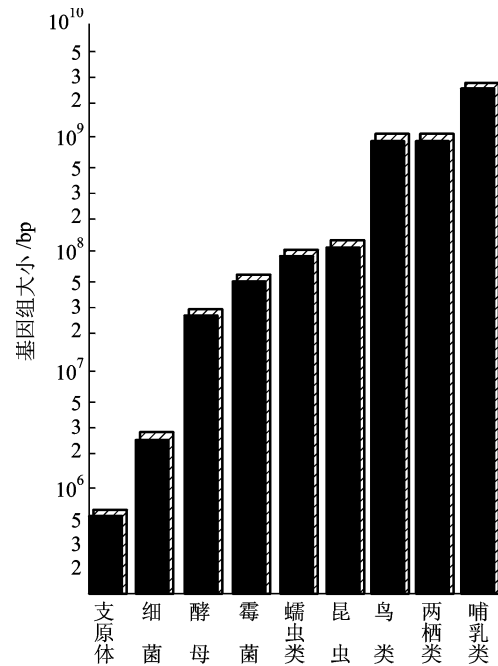


图 6-1 各类生物基因组大小

物种的 **C 值** 及其进化复杂性之间没有严格的对应关系,这种现象称为 **C 值悖理 (C value paradox)**,或 **C 值佯谬**(图 6-2)。

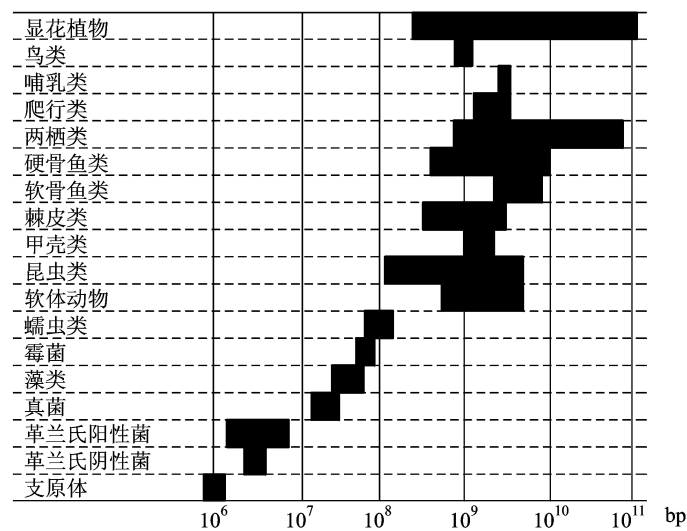


图 6-2 各类生物单倍体 DNA 含量

表 6-1 真核生物(2C)的 DNA 含量

单位: pg

真核生物	DNA(2C)
两栖鲵( <i>Amphimè</i> )	168 0
肺鱼( <i>Protopterus</i> )	100 0
蝾螈( <i>Salamandra</i> )	85 3
金线蛙( <i>Rana esculenta</i> )	16 8
牛[ <i>Bos taurus</i> (ox)]	6 4
人( <i>Homo sapiens</i> )	6 4
土耳其盘羊( <i>Ovis aries</i> )	5 7
黑腹果蝇( <i>Drosophila melanogaster</i> )	0 2
豌豆( <i>Pisum sativum</i> )	28 0
蚕豆( <i>Vicia faba</i> )	4 0
玉米( <i>Zea mays</i> )	11 0
酿酒酵母( <i>Saccharomyces cerevisiae</i> )	0 026(C)

从图 6-2 和表 6-1 可见,牛与人的 DNA 含量相等,包括人类在内的哺乳动物基因组 DNA 的最高含量与爬行动物相同。众所周知,鸟类是由爬行动物进化来的,而软骨鱼类是脊椎动物中最低的类群,但鸟类 DNA 的最高含量竟然与爬行动物 DNA 的最低含量一样,而且还远不如软骨鱼类 DNA 的最低含量,特别是软骨鱼类 DNA 含量比硬骨鱼类、爬行类、鸟类、哺乳类都高,在这里 C 值与进化趋势完全相反。

人们对 C 值悖理已经提出许多解释:包括基因组的部分或完全加倍、转座(transposition)、反转录已加工假基因(retroprocessed pseudogenes)、DNA 复制滑动(replication slippage)、不等交换(unequal crossover)和 DNA 扩增(DNA amplification)等,Petrov 等又提出一个解释是:各种生物基因组的大小是由于基因组中长期积累起来的过量的非编码 DNA 被清除的速率不同所造成的结果,即 DNA 丢失的速率愈慢,那么基因组 DNA 含量愈高。这种说法与以前一些解释大同小异。真核生物基因组中是存在大量的不编码基因产物的 DNA 序列。一般而言,愈是简单的生物,基因组中不编码蛋白质的 DNA 序列愈少,它们的结构基因的数目愈接近于相应 DNA 含量所估计的基因数。如 MS2 的遗传物质是 RNA,其含量为 3 000 个核苷酸单链 RNA,按每个基因平均为 1 000 个核苷酸计算,刚好具有 3 个结构基因,编码 3 种蛋白质;又如  $\lambda$  噬菌体中除去附着位点(attD)、复制起点、黏性末端、基因的启动区和操作区外,其余全部为结构基因编码区。同样在大肠杆菌基因组中除控制区外,也都是编码蛋白质的结构基因的 DNA 序列。然而真核生物的情况就大不相同了,就哺乳动物而言,由于结构基因有内含子,因而基因长度可比原核生物增加 5~8 倍,即使按如此大小的基因进行推算,那么哺乳动物的基因组具有 400 000~600 000 个基因,这显然是不可能的,按各种实验结果估计,有功能的基因数不会超过这个数字的 10%。随着人类基因组计划的完成,发现人类基因组含 20 000~25 000 个基因,那么在生物进化中为什么要积累并保留愈来愈多的冗余 DNA? 而它们之间的 C 值悖理仅仅是由于删除这些冗余 DNA 速率的缘故? 再说这些非编码序列或冗余 DNA 都是不含遗传信息的 DNA 吗? 这些序列的结构与功能如何? 非编码序列除了调控序列、rRNA 基因、tRNA 基因外,它们还可能对基因多样性形成、对基因组进化(genome evolution)等具有重要的生物学意义。这些问题都有待进一步阐明。

### 6.1.2 N 值悖理

随着各类生物基因组测序工作的完成,特别是人类基因组计划的研究报告发表后,认为人类基因组只含大约 25 000 个基因,而只有 1 000 多个体细胞的线虫却含有近 20 000 个基因,人类真正的遗传

信息只比低等线虫多 1/5, 更难以理解的是比线虫高级的果蝇则只有 14 000 个基因, 其遗传信息只有线虫的 70%。人们将生物的基因数目与生物在进化树上的位置不存在正相关的事实称为 N 值悖理 (N value paradox) 或 N 值伴谬。

从 N 值悖理说明, 生物体的复杂性不仅仅是基因数目的函数, 随着生物体复杂性的增加, 基因的大小和基因结构的复杂性亦增加。较为复杂的生物体有一些机制使一个基因产生多个蛋白质分子, 如内含子的可变剪接 (alternative splicing), 不同的剪接方式产生不同的 mRNA, 随后通过翻译产生不同的蛋白质产物。据估计, 人类基因组中, 一大半基因具有可变剪接的调节而形成大量不同蛋白质的异构体。

另外, 随着生物体复杂程度的提高, 其基因组中的基因重复程度越高, 大肠杆菌含有约 1 345 个重复基因, 而果蝇含有 5 536 个。基因结构复杂性的增加还体现在结构域 (domain) 的数目上。域是一种特殊的蛋白质序列元件, 一个域一般与一种功能有关, 如 DNA 结合域。果蝇中含有 5 个以上结构域相同或不同的蛋白质有 100 个左右, 而酵母中仅有 20 个蛋白质含有相同数目的域。内含子的数目也是随着生物体复杂性的增加而增加的。酿酒酵母 (*S. cerevisiae*) 基因组中含有 220 个内含子, 而果蝇保有 41 000 个内含子。有学者提出生物的复杂性不能仅仅用基因的数目来衡量, 而应该用整个基因组的理论上的转录组 (transcriptome) 来定义。何况真核生物基因依据序列的相似性组成许多基因家族, 其序列的相似性一般又转化为功能的保守性, 在每一个基因家族中可以有成千上万个基因, 但基因组序列鉴定出的全部基因中有 30% 到 50% 不具有可测的功能, 特别是在高等真核生物中基因家族中大量的基因功能无法确定。总之, 对 N 值悖理现象尚待作出更为合理的解释。

### 6.1.3 真核生物基因组 DNA 序列的复杂度

真核生物基因组 DNA C 值和 N 值悖理现象都表明其 DNA 序列的复杂度, 为此可通过复性动力学来检测基因组 DNA 序列的复杂性。也就是通过 DNA 的变性 (denaturation) 和复性 (renaturation) 反应的动力学过程分析 DNA 序列的性质, 由于复性的速率取决于互补的 DNA 序列之间的随机碰撞, 所以 DNA 复性是一个双分子二级反应。单链消失速度的微分方程为:

$$-\frac{dC}{dt} = kC^2$$

其中: C 为单链 DNA 的浓度 (单位是每升的核苷酸摩尔数); t 为时间 (单位为 s); k 为重组速率常数 (单位是  $L \cdot mol^{-1} \cdot s^{-1}$ ), k 取决于阳离子浓度、温度、片段大小和 DNA 序列的复杂性。上式可改写为:

$$-\frac{dC}{C^2} = k dt$$

当  $t = 0$  时,  $C = C_0$ , 将上式积分:

$$-\left[ \frac{1}{C_0} - \frac{1}{C} \right] = k t$$

即:  $\frac{1}{C} - \frac{1}{C_0} = k t$  改写为:

$$\frac{C}{C_0} = \frac{1}{1 + k C_0 t}$$

当  $t = 0$  时,  $C = C_0$ , 表明所有 DNA 都是单链,  $C_0$  为 DNA 总浓度。复性分数  $C/C_0$  是起始浓度和经过时间的乘积  $C_0 t$  的函数, 这样的函数绘成图称为  $C_0 t$  曲线 (图 6-3)。

从方程式可见控制复性反应的参数是  $C_0 t$  如当  $t = t_2$  时, 即  $C/C_0 = 1/2$  时, 也就是 50% 单链复性时, 则方程为:  $\frac{C}{C_0} = \frac{1}{2} = \frac{1}{1 + k C_0 t_2}$ , 因此  $C_0 t_2 = \frac{1}{k}$ 。

如果基因组中每一种基因只有一个, 即都是单拷贝序列, 那么基因组愈大则基因组的复杂性愈大, 复性速率愈小, k 也愈小, 所以  $C_0 t_2$  与非重复序列的基因组大小呈正比。即

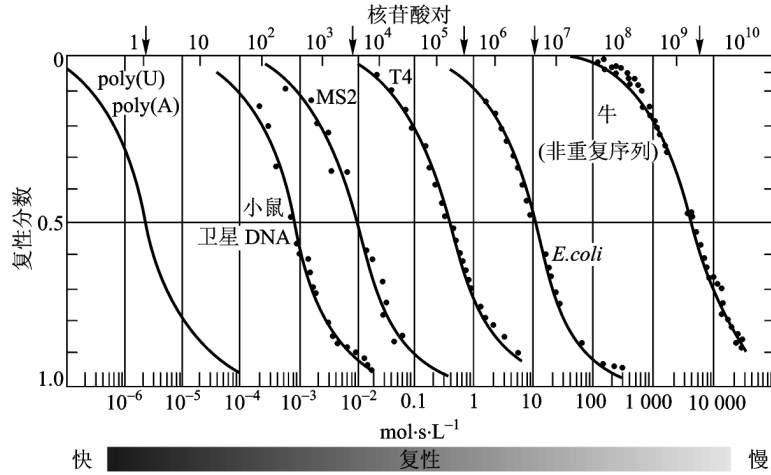


图 6-3 不同物种核酸的  $C_0t$  曲线

$$\frac{\text{基因组 A 的 } C_0t_{1/2}}{\text{基因组 A 的核苷酸对数}} = \frac{\text{基因组 B 的 } C_0t_{1/2}}{\text{基因组 B 的核苷酸对数}}$$

图 6-3 表明,  $C_0t_{1/2}$  与基因组的大小成正比。其中 poly(U) + poly(A), 其  $kC_0t_{1/2} = 1$  个核苷酸对, 因而复性最快; MS2 是 RNA 噬菌体, T4 为 DNA 噬菌体, 每个基因组的大小用箭头标于图的上方。

不同生物基因组的  $C_0t_{1/2}$  是不相同的,  $C_0t_{1/2}$  的位置除了决定于基因组的大小以外, 还取决于每个基因的核苷酸序列的重复次数, 重复次数愈少则复性愈慢,  $C_0t_{1/2}$  的位置愈后; 重复次数愈多,  $C_0t_{1/2}$  位置愈前。真核生物基因组的复性曲线往往出现 2 个或 3 个明显不同的  $C_0t_{1/2}$  位置, 说明这类基因组中包含着重复次数显然不同的几个成分(图 6-4), 图 6-4 是假设的一个真核生物基因组复性曲线。

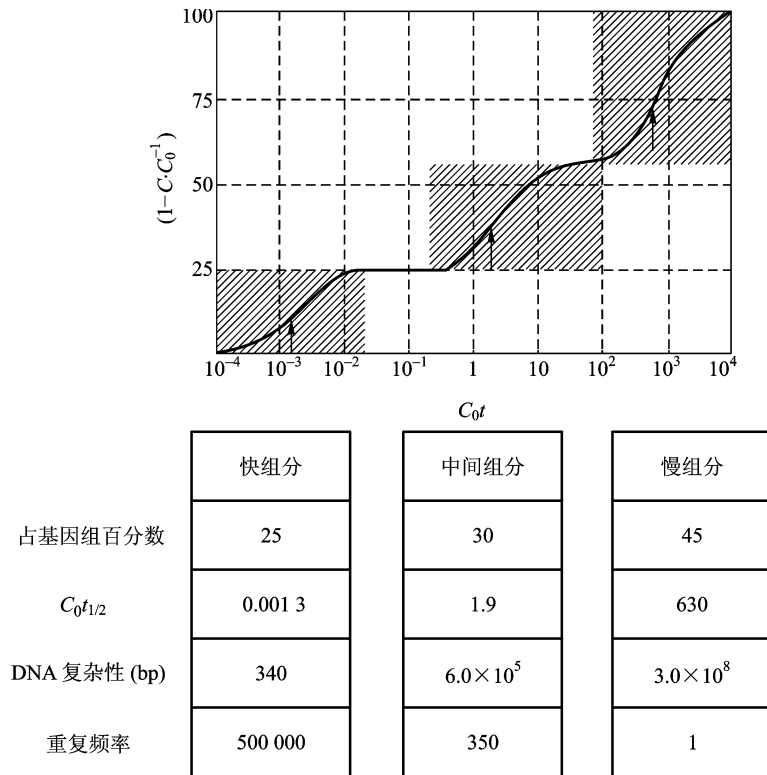


图 6-4 真核生物 DNA 复性动力学模式图

从图 6-4 可以看出,最后复性这部分 DNA 的  $C_0 t_2 = 630$  它占全部 DNA 的 45%, 相当于  $3.0 \times 10^8$  bp, 重复频率为 1; 中间这部分 DNA 的  $C_0 t_2 = 1.9$  占全部 DNA 的 30%, 相当于  $6.0 \times 10^7$  bp, 该 DNA 序列重复频率为 350; 复性最快 DNA 的  $C_0 t_2 = 0.0013$  占基因组 DNA 的 25%, 相当于 340 bp, 重复频率为  $5 \times 10^5$ 。显然第一部分, 即复性最慢的部分是单拷贝 DNA 序列, 第二部分为少量重复或中度重复 DNA 序列, 第三部分为高度重复 DNA 序列。因此, 真核生物基因组序列大致可分为 3 种类型。

(1) 单拷贝序列 (**unique sequence**) 亦称非重复序列 (**nonrepetitive sequence**)

在一个基因组中只有一个拷贝或 2~3 个拷贝。真核生物的大多数基因在单倍体中都是单拷贝的。不同生物基因组中单拷贝序列所占的比例是不同的(图 6-5)。原核生物中一般只含有非重复序列, 较低等的真核生物中大部分 DNA 也是单拷贝的。动物细胞将近 50% DNA 是中度或高度重复的, 特别是植物和两栖类生物中单拷贝 DNA 序列降低, 而中度和高度重复序列增加。

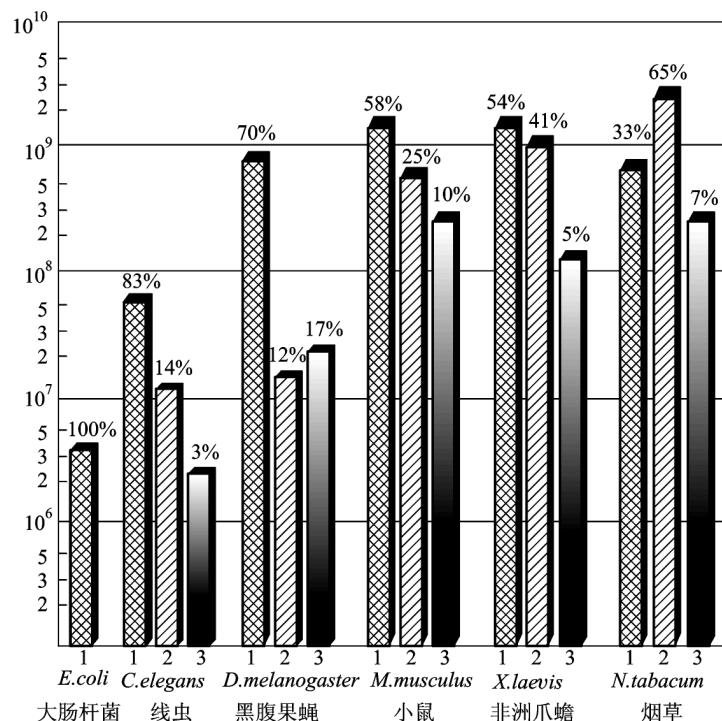


图 6-5 不同生物基因组中单拷贝序列所占比例

1 非重复序列 2 中度重复序列 3 高度重复序列

(2) 中度重复序列 (**moderately repetitive sequence**)

中度重复序列中的重复单位平均长度约 300 bp, 重复次数为  $10^3 \sim 10^5$ , 人的珠蛋白(红血蛋白)基因属于这种重复序列, 该基因中除了包括已确定的 8 个珠蛋白功能基因和 3 个珠蛋白假基因外, 还有一个近年发现的基因。假基因也属于中度重复序列。另一类中度重复序列的重复次数为  $10^3 \sim 10^6$ , 该序列常以回文序列方式出现在基因组的许多位置上, 一些回文序列中间隔着单拷贝序列, 另一些中间不存在单拷贝序列, 所以经变性并复性后, 前者可观察到茎-环结构, 后者则形成发夹图像(图 6-6)。

(3) 高度重复序列 (**highly repetitive sequence**)

顾名思义, 高度重复序列就是在基因组中存在大量拷贝的序列, 一般重复次数在  $10^6$  以上。通常这些序列的长度为 6~200 bp, 如卫星 DNA。这些重复序列大部分集中在染色体的异染色质区, 特别是在着丝粒和端粒附近。高度重复序列中常有一些 AT 含量很高的简单串联重复序列。因序列简单, 缺乏转录所必需的启动子, 故没有转录能力。然而 DNA 复制能力却和单一序列一样, 复制速度

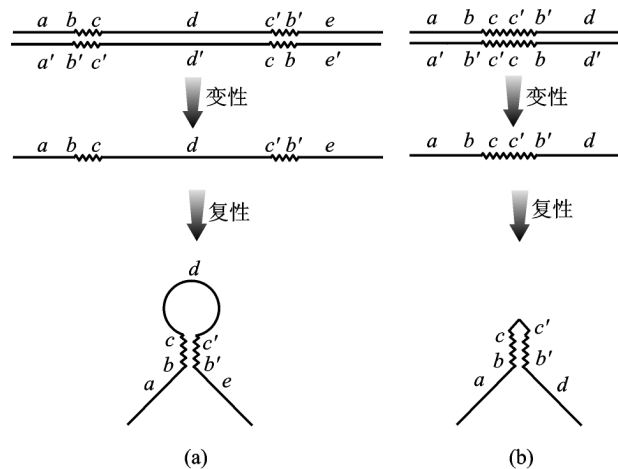


图 6-6 一些中度重复序列的排列方式和变性并复性后的模式图  
(a) 茎-环结构 (b) 发夹图像

很快。大多数高等真核生物 DNA 都有 20% 以上的高度重复序列,而且数目变化很大,这类序列的多少对 C 值的影响可能最大。重复序列的确切生物学意义尚有待阐明。

卫星 DNA (satellite DNA) 是一类高度重复的 DNA 序列。各种 DNA 在氯化铯梯度离心中,平衡时的浮力密度决定于它的 GC 含量,GC 含量越高,浮力密度越大。真核生物的 DNA 一般含有 30%~50% 的 GC,在 DNA 的不同区段,GC 含量约相差 10%。对一个物种来说,当基因组 DNA 切断成数百个碱基对的片段进行超离心时,其浮力密度曲线是覆盖一定浮力密度范围的一条宽带,但是有些 DNA 片段都有异常高或低的 GC 含量,常在主要 DNA 带的前面或后面有一个次要的 DNA 带相伴随,这些小的区带就像卫星一样围绕着 DNA 主带,故称卫星 DNA (表 6-2)。果蝇 (*D. v. irilis*) 基因组中有 3 种主要的卫星 DNA,它们的核心序列非常相似,改变一个碱基,足以从 I 型卫星序列变成 II 或 III 型卫星序列。II 或 III 型卫星序列为 *D. v. irilis* 特有, I 型卫星序列存在于与上述果蝇相似的其他果蝇类中。这 3 种卫星序列的主要特征是它们的重复单位只有 7 个碱基对。在其他物种中也存在类似的卫星 DNA。

表 6-2 果蝇 (*D. v. irilis*) 的卫星 DNA

卫星	核心序列	总长度/bp	占基因组比例
I	ACAAACT	$1 \times 10^7$	25%
	TGTTTGA		
II	ATAAACT	$3.6 \times 10^6$	8%
	TATTTGA		
III	ACAAATT	$3.6 \times 10^6$	8%
	TGTTTAA		
隐蔽卫星*	AATATAG		
	TTATATC		

\* 隐蔽卫星 DNA (cryptic satellite DNA) 是用密度梯度离心分析出一条卫星带,但仍存在于 DNA 主带中的高度重复序列。

根据卫星 DNA 的浮力密度可以分成 I、II、III、IV 4 类,它们在染色体上的位置可以用放射性标记探针作 DNA 分子原位杂交 (in situ hybridization) 来鉴定,发现 4 类卫星 DNA 都能与人的各条染色体杂交,而且杂交的带型也很相似。说明这 4 类卫星 DNA 也存在于人的染色体上,而且分布的状况也相同。卫星 DNA 分布于着丝粒附近的异染色质区。由于异染色质区是高度螺旋化的, DNA 是不表达的。卫星 DNA 在着丝粒处的集中分布可能与细胞分裂时染色体的运动有关。

## 6.2 真菌类的四分子分析与作图

### 6.2.1 顺序四分子的遗传分析

真核生物的遗传分析的原理与方法在第 5 章侧重讨论了果蝇和玉米以及人类的连锁遗传及基因定位与作图,在此我们将介绍真菌类的四分子分析与作图方法。粗糙脉孢菌(*Neurospora crassa*)和酵母菌(*Saccharomyces*)都是子囊菌(ascmycetes)的真菌,属于低等真核生物。它们可以在一个共有的子囊(ascus)中产生子囊孢子(ascospore),并同时具有无性生殖和有性生殖过程,特别是粗糙脉孢菌中,子囊孢子在子囊中的线性排列顺序与处在减数分裂中期赤道板上的染色单体的排列顺序完全一致,因而成为遗传学分析的经典材料。

脉孢菌减数分裂的 4 个产物保留在一起,称为四分子(tetrad),但是,子囊的外形是如此的狭窄,以致分裂的纺锤体不能重叠,只能纵立于它的长轴之中,这样一来,所有分裂后的核——8 个子囊孢子都从上到下顺序排列成行,可以推知,第一对子囊孢子来自一条染色单体,第二对子囊孢子则是来自这条染色单体的姊妹染色单体;第三和第四对子囊孢子是来自前一条染色体的同源染色体的姊妹染色单体。所以,脉孢菌减数分裂所产生的四分子属于顺序四分子(ordered tetrad)[图 6-7 a]。

这种顺序四分子在遗传分析上有很多优越性:① 可以将着丝粒视作一个座位,计算某一基因与着丝粒的重组率。② 子囊中子囊孢子的对称性,证明减数分裂是一个交互过程。③ 可以检验染色单体的交换是否存在干涉现象,还可利用它来研究基因转变(gene conversion)。④ 证明双交换不仅可以包括 4 线中的两线,而且还可以包括 3 线或 4 线。

#### (1) 着丝粒作图(centromere mapping)

利用四分子分析法,测定基因与着丝粒间的距离称为着丝粒作图。着丝粒作图是基于这样的事实:在一对非姊妹染色单体间没有发生着丝粒和某杂合基因座交换的减数分裂和发生了交换的减数分裂,其产物(四分子)中的等位基因在排列方式上是不同的。前者称为第一次分裂分离(first division segregation),又称  $M_I$  模式;后者称为第二次分裂分离(second division segregation),或称为  $M_{II}$  模式。以粗糙脉孢菌接合型 A、a 为例分别说明上述概念的由来[图 6-7 b)、c)]。

从图 6-7 b)中可知 8 个子囊孢子呈 AAAAaaaa(或 aaaaAAAA)的排列是如何产生的。因为没有纺锤体的重叠,携带 A 和 a 的核在子囊中互不干扰,同时,尽管等位基因 A 和 a 同处于二倍体的性母细胞中,减数分裂的第一次分裂就将 A 和 a 分离开,并且在减数分裂 II 中依然保持这种分离状态。这种情况就是第一次分裂分离,或  $M_I$  模式。图 6-7 c)显示的是第二次分裂分离的染色体行为:当非姊妹染色单体在着丝粒和基因座(A, a)间发生交换时,等位基因 A 和 a 在减数分裂第一次分裂后仍同时存在于同一核中,因为在这个过程中,只有同源染色体的分离,而染色体各自的着丝粒没有分开,待到第二次分裂(减数分裂 II)时, A 和 a 才进入不同的核中,这种情况称为第二次分裂分离。由此得到的等位基因的排列模式 AAaaAAaa(或它的镜像排列 aaAAaaAA)称为  $M_{II}$  模式。这种模式的特征是在一半的四分子产物中发生了重组。这种交换型排列模式还有可能出现 AAaaaaAA(或 aaAAAAaa),这是由于纺锤体的随机趋向的机制,使得“A 上 a 下”和“A 下 a 上”或“A 上 a 下”和“a 下 A 上”,而且二者的频率也大致相当。

现在用实验说明着丝粒作图:有两种不同接合型的脉孢菌菌株,一种是能合成赖氨酸(lysine)的野生型菌株(记作  $lys^+$  或 +),该菌株成熟的子囊孢子呈黑色。另一种是赖氨酸缺陷型(记作  $lys^-$  或 -),这种菌株的子囊孢子成熟较迟,呈灰色。将这两种菌株进行杂交  $lys^+ \times lys^-$ 。在杂种子囊中减数分裂的产物根据黑色孢子和灰色孢子的排列次序可有 6 种子囊型,每种子囊型中有 8 个子孢子,为



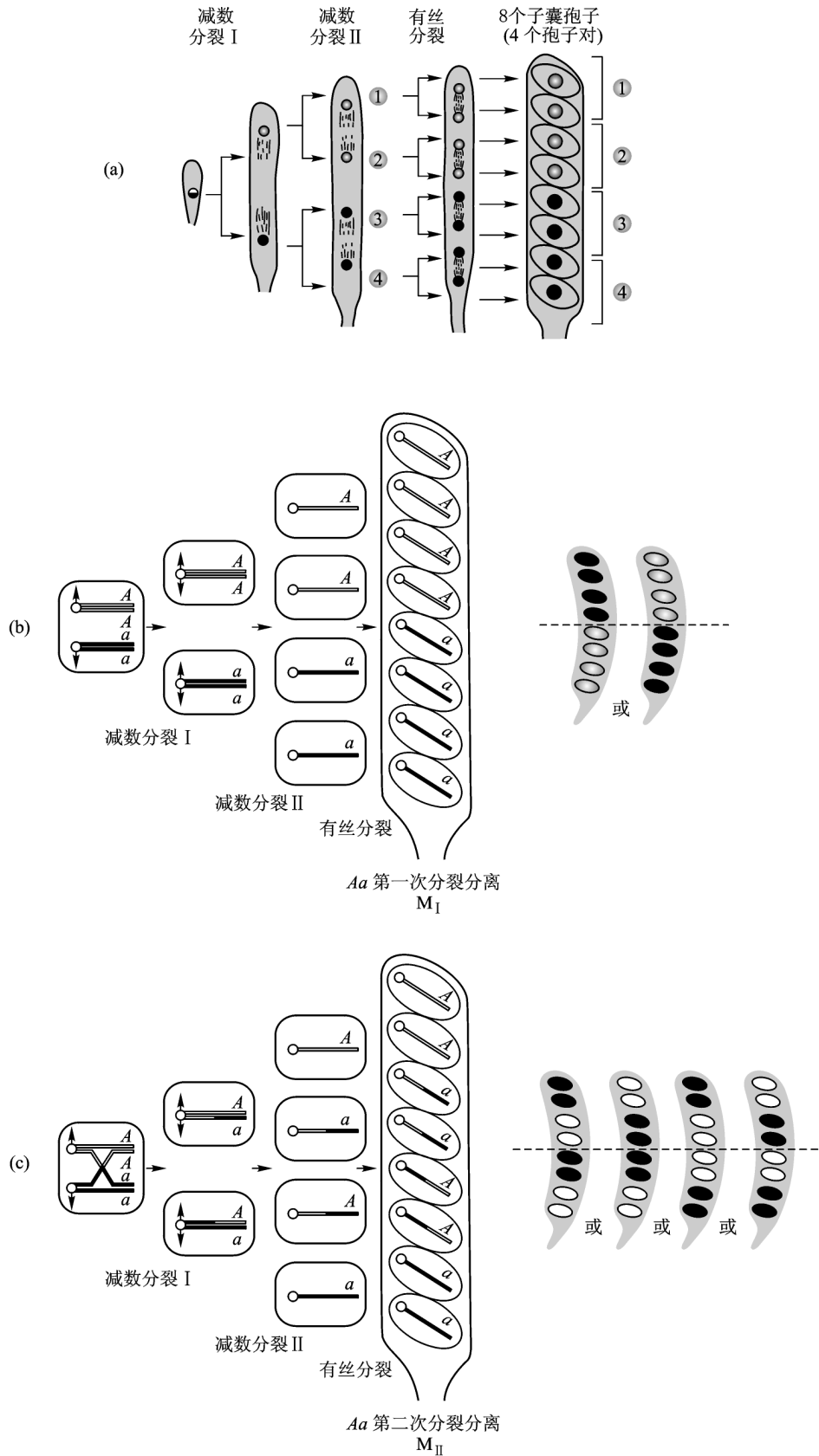


图 6-7 粗糙脉孢菌减数分裂中等位基因的分离

(a) 顺序四分子的形成 (b) 第一次分裂分离( $M_I$ ) (c) 第二次分裂分离( $M_{II}$ )

方便起见只写出其中的 4 个孢子对 (spore pairs), 其计数的结果在表 6-3 中。

表 6-3 粗糙脉孢菌  $lys^+ \times lys^-$  杂交子代子囊类型

	(1)	(2)	(3)	(4)	(5)	(6)
子囊类型	+	-	+	-	+	-
子囊数	105	129	9	5	10	16
分裂类型	$M_I$	$M_I$	$M_{II}$	$M_{II}$	$M_{II}$	$M_{II}$
	非交换型		交换型			

表 6-3 中的 (1) 和 (2) 两种类型的子囊中其子囊孢子的排列方式互为镜影, 同样 (3) 和 (4), (5) 和 (6) 也互为镜影, 说明减数分裂是一个交互过程。从图 6-7 可以推知, 第一次减数分裂时, 带有  $lys^+$  的两条染色单体移向一极, 而带有  $lys^-$  的两条染色单体移向另一极。这样, 就  $lys^+ / lys^-$  这一对基因而言, 在第一次减数分裂时就发生了分离, 所以子囊型 (1) 和 (2) 属于第一次分裂分离 ( $M_I$ ) 类型。进入第二次减数分裂时, 每一染色单体相互分开, 所以在每一子囊中两个  $lys^+$  的孢子排列在一起, 两个  $lys^-$  的孢子排列在一起, 再经过一次有丝分裂, 最后形成 4 个孢子对, 排列顺序自然是 + + - - 或 - - + +, 因为凡属第一次分裂分离的子囊, 均是着丝粒和所研究的基因间未发生过交换, 所以称为非交换型。毫无疑问, 在此,  $lys^+ / lys^-$  基因与着丝粒间没有发生过交换。子囊型 (3)、(4)、(5) 和 (6) 均属于第二次分裂分离 ( $M_{II}$ ) 的子囊。由于  $lys^+ / lys^-$  基因与着丝粒之间发生了一次单交换, 带有  $lys^+$  的染色单体和带有交换而来的  $lys^-$  的染色单体相互分开, 必然发生在第二次减数分裂 ( $M_{II}$ ) 时。就  $lys^+ / lys^-$  这一对基因对而言, 分离发生在第二次减数分裂。所以它们在有丝分裂后最终形成的子囊孢子的排列顺序是 + - + - (或 - + - +) 以及 + - - + (或 - + + -)。由此可见, 凡属第二次分裂分离型的子囊, 一定是在着丝粒与所研究的基因之间发生过交换。所以  $M_{II}$  型也称作交换型。

我们已知, 染色体上两个基因座之间的距离愈远, 则它们之间发生交换的频率愈高。因此, 第二次分裂分离的子囊愈多, 则说明有关基因和着丝粒的距离愈远。所以由第二次分裂分离子囊的频数, 可以计算某一基因和着丝粒间的距离, 这距离称为着丝粒距离。由于交换发生在 4 个染色单体的两个染色单体中, 所以有关的基因和着丝粒之间的每一次交换产生两个交换型和两个非交换型的染色单体, 也就是说, 每发生一次交换, 一个子囊中只有半数孢子发生重组。从子囊中子囊孢子排列的形式来讲, 则每一次交换产生一个第二次分裂分离的子囊。因此为了使着丝粒与有关基因间的距离和用杂交子代中按重组率所表示的距离一致起见, 着丝粒与有关基因间的重组率为:

$$\text{着丝粒与有关基因的 RF} = \frac{\text{第二次分裂分离子囊数(或交换型子囊数)}}{\text{子囊总数}} \times \frac{1}{2} \times 100\%$$

$$\text{或者} \quad \text{RF(着丝粒—基因)} = \frac{M_{II} \times \frac{1}{2}}{M_I + M_{II}} \times 100\%$$

将表 6-3 中所观察到的各类子囊数代入上式可得:

$$\text{着丝粒—} lys^+ \text{ 基因的 RF} = \frac{M_{II} \times \frac{1}{2}}{M_I + M_{II}} \times 100\% = \frac{(9+5+10+16) \times \frac{1}{2}}{105+129+9+5+10+16} \times 100\% = 7.3\%$$

换言之,  $lys^+$  基因与着丝粒间的图距是 7.3 cM。

## (2) 两个连锁基因的作图

利用顺序四分子分析也可以对两个基因进行连锁分析和作图。粗糙脉孢菌的烟酸依赖型  $nic$  (简化代号为  $a$ ), 需要在培养基中添加烟酸才能生长,  $ade$  (简化代号为  $n$ ) 是腺嘌呤依赖型, 在培养基中添加腺嘌呤才能生长。将两个菌株杂交 ( $n^+ \times + a$ )。由前可知, 一对基因杂交, 有 6 种不同的子囊型, 两对基因杂交必有  $6 \times 6 = 36$  种不同的子囊类型; 但是因为半个子囊内的基因型次序可以忽视, 不论是  $n^+$  孢子对在“上面”,  $+ a$  孢子对在“下面”, 还是  $+ a$  在“上面”,  $n^+$  孢子对在“下面”, 都不过是反映着丝粒在减数分裂过程中的随机趋向而已, 所以可以将 36 种不同的子囊型归纳为 7 种基本子囊型。表 6-4 显示的是 7 种子囊型及其实得子囊数。

表 6-4 粗糙脉孢菌  $n^+ \times + a$  杂交结果

子囊型	①	②	③	④	⑤	⑥	⑦
四分子基因型次序	$+ a$ $+ a$ $n^+$ $n^+$	$++$ $++$ $na$ $na$	$++$ $+ a$ $n^+$ $na$	$+ a$ $na$ $++$ $n^+$	$+ a$ $n^+$ $+ a$ $n^+$	$++$ $na$ $++$ $na$	$++$ $na$ $+ a$ $n^+$
分离发生的时期	$M_I M_I$	$M_I M_I$	$M_I M_{II}$	$M_{II} M_I$	$M_{II} M_{II}$	$M_{II} M_{II}$	$M_{II} M_{II}$
四分子类别	PD	NPD	T	T	PD	NPD	T
实得子囊数	808	1	90	5	90	1	5

两对基因杂交时, 如不考虑孢子排列, 只考虑性状组合时, 子囊可以分作 3 种四分子类型: 亲二型 (parental ditype, PD), 2 种基因型, 而且跟亲代一样, 包括子囊型①和⑤; 非亲二型 (non parental ditype, NPD), 有 2 种基因型, 都跟亲代不同, 是重组型, 包括子囊型②和⑥; 四型 (tetatype, T), 有 4 种基因型, 2 种与亲代相同, 2 种重组型, 包括子囊型③、④和⑦。从染色体交换和重组规律说明 3 种四分子类型和各类子囊的形成原因 (图 6-8)。

交换类型	染色体图像	重组	四分子类型	子囊型
无交换		0	(PD) $+ a + a n^+, n^+$	①
四线双交换 (1-4)(2-3)		100%	(NPD) $++, ++, na, na$	②
单交换 (1-4)		50%	(T) $++, + a n^+, na$	③
二线双交换 (2-3) 着丝粒—基因—基因		50%	(T) $+ a na ++, n^+$	④

交换类型	染色体图像	重组	四分子类型	子囊型
单交换 (2-3) 着丝粒—基因		0	(PD) + a n +, + a n +	⑤
四线多交换 (1-4)(2-3)		100%	(NPD) ++, na ++, na	⑥
三线双交换 (1-3)(2-3) 着丝粒—基因—基因		50%	(T) ++, na + a n +	⑦

图 6-8 3种四分子类型的形成

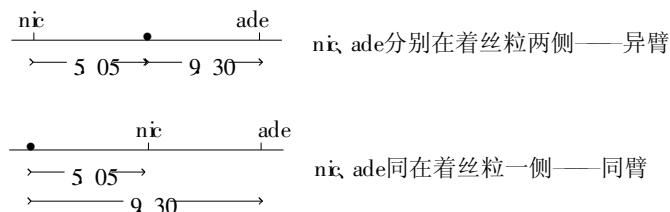
首先判断  $n$ 、 $a$  基因是独立分配还是连锁。根据表 6-4 中的数字,  $PD = 808 + 90 = 898$ ,  $NPD = 1 + 1 = 2$ 。如果这两个基因是自由组合的, 可以预期  $PD:NPD = 1:1$ , 而实验结果是  $PD \gg NPD$ 。说明这两个基因不是自由组合, 而是相互连锁的。在判断这两个基因是否连锁时, 四分子类型中的 T 型(四型)的数据不能带来有价值的信息, 因为 T 型四分子中有 4 种基因型, 其中 50% 为重组型, 50% 为亲本型, 与独立分配的遗传规律不能相区别, 因此 T 型四分子说明不了问题, 故只需比较  $PD/NPD = 1$  或  $\approx 1$ , 则非连锁,  $PD/NPD > 1$ , 则连锁。如果 NPD 的子囊数极少, 则连锁很紧密。

其次计算着丝粒(•)与基因  $nic^-(n)$ 、 $ade^-(a)$  间的重组率:

$$RF(\bullet - a) = \frac{M_{II} \times \frac{1}{2}}{M_I + M_{II}} = \frac{\frac{1}{2}(5 + 90 + 1 + 5)}{1000} = 5.05\%, 5.05 \text{ cM}$$

$$RF(\bullet - n) = \frac{M_{II} \times \frac{1}{2}}{M_I + M_{II}} = \frac{\frac{1}{2}(90 + 90 + 1 + 5)}{1000} = 9.30\%, 9.30 \text{ cM}$$

由于  $5.05 < 9.30$ , 所以  $n$  基因比  $a$  基因离着丝粒近。但是  $n$  与  $a$  基因在着丝粒同侧还是不同侧? 也就是说着丝粒、 $n$ 、 $a$  在染色体上的排列可能有两种方式:



对上述问题的分析可以有以下两种方法:

① 利用连锁基因  $n$  和  $a$  都处在  $M_{II}$  状态下的 PD 和 NPD 四分子类型出现的频率来判断这两个基因在同臂还是异臂上(图 6-9)。

如果  $n$ 、 $a$  在异臂, 则 PD 与 NPD 都是由双交换形成, 且机会应相等, 因而 PD 与 NPD 的频率相等。但是, 表 6-4 的实验数据显示, 在⑤中, 同处  $M_{II} M_{II}$  的 PD 子囊数为 90, 在⑥中, 同处  $M_{II} M_{II}$  的

NPD子囊数为1。显然,PD多于NPD。故这种排列的可能性不存在。于是,我们可以判断  $n$ 、 $a$  在同臂。

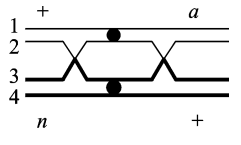
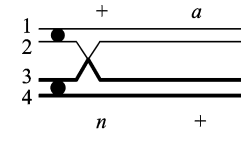
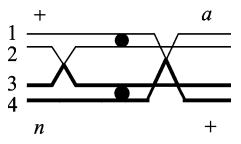
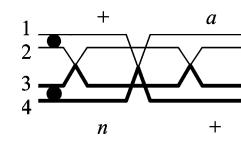
	同一染色体上标记基因		交换产物
	在异臂上	在同臂上	
I PD $M_{II} M_{II}$	 <p>二线双交换</p>	 <p>单交换</p>	<p>+ a n + + a n +</p>
II NPD $M_{II} M_{II}$	 <p>四线双交换</p>	 <p>四线三交换</p>	<p>+ + n a + + n a</p>

图 6-9 连锁基因的同(异)臂分析

② 在  $n$  和  $a$  分别为  $M_I$  和  $M_{II}$  的情况下,对子囊数与  $RF(\bullet - n), RF(\bullet - a)$  进行比较分析,得下表(表 6-5)。

表 6-5 按分离时期重新整理表 6-4 数据

$n$ 与 +	+ / ade	子囊数
$M_I$	$M_I$	(808+1) 809
$M_I$	$M_{II}$	(90) 90
$M_{II}$	$M_I$	(5) 5
$M_{II}$	$M_{II}$	(90+5+1) 96

已知  $RF(\bullet - n) = 5.05\%$ ,  $RF(\bullet - ade) = 9.30\%$ ,两者相差不到一倍,若  $n$  和  $a$  各自独立地与着丝粒发生重组,则  $M_{II}$  的子囊数也应相差不到一倍。但实际上,交换发生在着丝粒与  $a$  间, $n$  是  $M_I$ ,  $a$  是  $M_{II}$  的子囊有 90 个[表 6-4 中的③];交换发生在着丝粒与  $n$  间时, $n$  是  $M_{II}$ ,  $a$  是  $M_I$  的子囊只有 5 个[表 6-4 中的④]。两者相差悬殊,与计算的重组值不相符合,所以,这两个基因不可能独自与着丝粒发生交换。另外,从表 6-5 可见,在  $n$  与着丝粒发生交换时, $a$  基因也一道与着丝粒交换,即  $n$  是  $M_{II}$ ,  $a$  也是  $M_{II}$  共计  $96 = 90 + 5 + 1$  个子囊。也就是说,同一交换使  $+ / n$  出现  $M_{II}$  型分离,也使  $+ / a$  出现  $M_{II}$  型分离,101 次中有 96 次,强有力地证明了  $n$  和  $a$  的连锁关系是  $n$ 、 $a$  在着丝粒同一侧。因为 T 型四分子中只有  $1/2$  重组产物, NPD 型四分子中,4 个孢子对全部都是重组基因型,所以我们计算  $n$ — $a$  间的重组率可用公式:

$$RF(n-a) = \frac{\frac{1}{2}T + NPD}{T + NPD + PD} = \frac{\frac{1}{2}(90 + 5 + 5) + (1 + 1)}{1000} = 5.20\%, 5.20 \text{ cM}$$

显然,已算得的  $RF(\bullet - a) = 9.30\%$  (9.30 cM),但实际上  $(\bullet - n)$  与  $(n-a)$  重组率之和 =  $5.05\% + 5.20\% = 10.25\%$ 。在此,又一次表现出末端两个基因间的重组值低估了实际图距的现象。由果蝇的三点测交的遗传分析中可知,造成这一结果的原因是着丝粒— $a$  基因间发生过双交换。当我们回

顾先前所述的计算过程时可以发现,在求  $RF(\bullet - a)$  时,是将所有  $M_{II}$  的子囊相加除以 2 倍的总子囊数,这样使少量的在  $(\bullet - a)$  间起过交换的子囊遗漏了。例如子囊型④,对于  $+ / a$  来讲,是属于  $M_I$ ,但是我们决定了  $(\bullet - n - a)$  顺序之后,得知这些子囊中发生了双交换,如子囊型④中的二线双交换及其产物[图 6-10(a)],可是在计算  $(\bullet - a)$  间的重组值时,却没有把这两次单交换计算在内,因而使  $(\bullet - a)$  间的重组值的估计偏低。被低估的重组值从表 6-6 的分析可以使  $(\bullet - a)$  间的重组值得到校正。

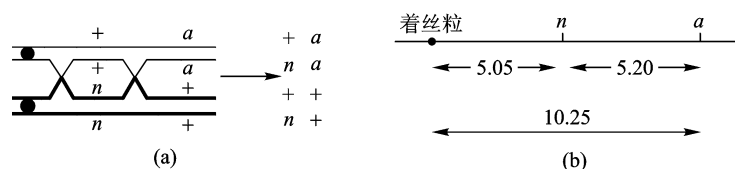


图 6-10 粗糙脉孢菌 3 个基因座的连锁遗传分析

(a) 发生双交换,则  $+ / a$  为  $M_I$ ,  $+ / n$  为  $M_{II}$  (b) 着丝粒、 $n$ 、 $a$  3 个基因座的连锁图

表 6-6 着丝粒与基因  $a$  间的重组值被低估的分析

子囊型	每一子囊被计算为重组子的染色单体数			子囊数	在所有子囊中被计算为重组子的染色单体数		
	$\bullet - n$	$n - a$	$\bullet - a$		$\bullet - n$	$n - a$	$\bullet - a$
2	0	4	0	1	0	4	0
3	0	2	2	90	0	180	180
4	2	2	0	5	10	10	0
5	2	0	2	90	180	0	180
6	2	4	2	1	2	4	2
7	2	2	2	5	10	10	10
总数					202	208	372

由表 6-6 可以看到  $202 + 208 \neq 372$ , 低估的重组值为  $\frac{202 + 208 - 372}{4000} \times 100\% = 0.95\%$ 。将这一数值加到由重组频率计算得来的图距 9.30 cM 上,就完全符合基因的直线排列原理了[图 6-10(b)]。

### 6.2.2 非顺序四分子的遗传分析

酿酒酵母、构巢曲霉和单细胞藻类中的衣藻的每一个子囊中 8 个子囊孢子的排列是杂乱无序的。这类真菌的遗传分析可采用非顺序四分子分析( unordered tetrad analysis)方法。以酿酒酵母为例,如果要研究  $A, B$  基因是否连锁,并计算图距,首先要明了当  $AB \times ab$  时,无论有无连锁,只产生下列 3 种可能的无序四分子。

	AB	aB	ab
孢子	AB	aB	aB
	ab	Ab	Ab
	ab	Ab	AB
	PD	NPD	T

因为这些子囊是无序排列的,所以尽管第一种类型(亲二型)看起来似乎是两个座位都属  $M_I$  模式,实际上不是这样,这些孢子以任意顺序写出来都是等效的。这些子囊仅仅是按照它们是包含二型(ditype)还是四型(tetatype)来划分的。在二型中,两种基因型要么是亲二型(PD)要么是非亲二型(NPD),四型子囊则是由 T 所表示的。读者可以试着写一些子囊,证明除了 PD、NPD 和 T 之外,没有

别的子囊型。观察这 3 种子囊类型(即四分子类型),可以发现只有 NPD 和 T 有重组型,故这两种四分子类型是决定重组率的关键。由于 T 只有  $1/2$  重组型,所以当我们已知 NPD 及 T 的百分数后,用下列公式求:

$$RF = \frac{1}{2} T + NPD$$

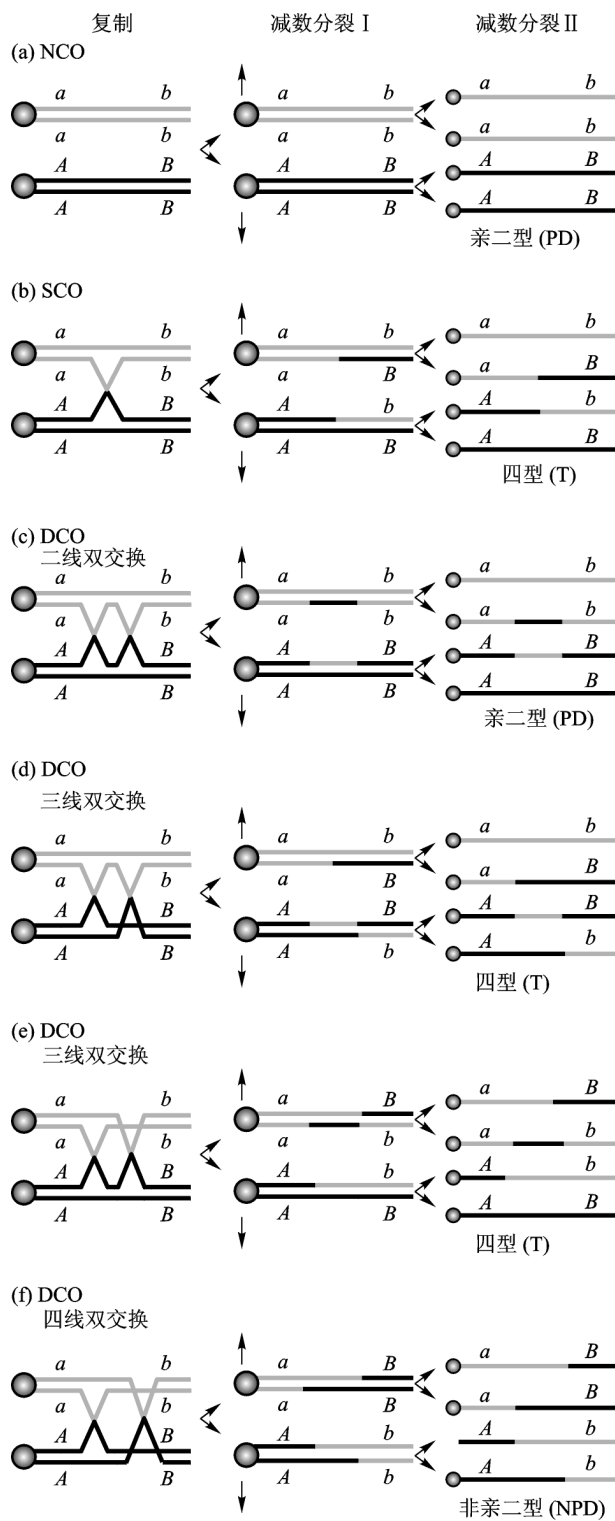


图 6-11 3 种无序四分子类型的形成

若求出  $RF=0.5$ , 可以断定  $A, B$  基因无连锁。因为  $A$  与  $B$  位于不同染色体上, 一对染色体与另一对染色体分别向两极移动是随机的, 所以  $PD=NPD$ , 各占  $50\%$ ,  $T$  的频率则取决于基因与着丝粒之间距离而定。如果  $RF<0.5$ , 则两基因座连锁。那么在减数分裂过程中  $A, B$  间可以发生非交换 (no crossover, NCO)、单交换 (single crossover, SCO) 或双交换 (double crossover, DCO), 图 6-11 表示减数分裂形成的这些无序子囊。

显然  $T$  型四分子既可来自单交换 (SCO), 也可来自双交换 (DCO)。NPD 是四线双交换产物。如果假设双交换在 4 条染色单体间随机发生, 那么, 可以推论 4 个 DCO 的频率是相同的。这就意味着 NPD 类型包括了  $1/4$  的 DCO, 则可以表示为:  $DCO=4NPD$ 。而  $T$  型中来自 DCO 减数分裂的部分是  $2NPD$ , 故 SCO 的大小可以用下式表示:  $SCO=T-2NPD$ 。那么, NCO 的大小可依公式:  $NCO=1-(SCO+DCO)$  求得。

当我们已估算出这一区域上 NCO、SCO、DCO 的值之后, 可以由这些数值推导出  $m$  值 (这一区域平均每次减数分裂的交换数), 将 SCO 数值加上两倍的 DCO 数值来求  $m$ :

$$m=(T-2NPD)+2(4NPD)=T+6NPD$$

在第 5 章作图函数一节中已知, 将  $m$  换算为图距时要乘以  $0.5$  (因为每个交换只产生  $50\%$  的重组)。所以: 图距 =  $50(T+6NPD)$  cM。

现举例说明: 若在  $AB \times ab$  中,  $PD=0.56$ ,  $NPD=0.03$ ,  $T=0.41$ 。由上述公式可得到  $ab$  间的图距 =  $50 \times (0.41 + 6 \times 0.03) = 29.5$  cM。

如果让我们将这个数值同由 RF 直接得到的数值比较时,  $RF=1/2T+NPD=1/2 \times 0.41 + 0.03 = 0.235 = 23.5\%$ , 即  $23.5$  cM。可以发现, 用 RF 值算出的图距比用图距公式 (作图函数校正) 算出的数值少了  $6$  cM。这是因为 RF 无法校正双交换的影响所致。该公式也适用于顺序四分子分析。当顺序四分子依次逐个取出培养, 在操作上比较困难时, 同样可采用非顺序四分子分析方法。

## 6.3 真核生物重组的分子机制

### 6.3.1 同源重组发生在减数分裂前期

同源重组 (homologous recombination) 又称普遍性重组 (generalized recombination), 它的发生是依赖于较大范围的 DNA 同源序列的联会。由前已知真核生物的遗传重组发生在减数分裂时期的同源染色体的非姊妹染色单体之间, 而且染色体或 DNA 分子之间相互交换对等的部分。同源重组中负责 DNA 配对和重组的蛋白质因子无碱基序列的特异性, 只要两条 DNA 序列相同或接近, 重组就可以在此序列中任何一点发生。当然也存在重组热点, 即某类序列发生重组的概率高于其他序列。此外, 真核生物的染色质状态影响重组, 如异染色质及其附近区域很少发生重组。而且两个 DNA 分子序列同源区越长越有利于同源重组, 大肠杆菌活体重组至少要求  $20 \sim 40$  bp 同源序列; 大肠杆菌与  $\lambda$  噬菌体或质粒之间的重组, 其同源序列要求  $\geq 13$  bp; 枯草杆菌基因组与质粒之间重组, 同源区的长度应  $\geq 70$  bp; 哺乳动物基因间的重组要求同源序列在  $150$  bp 以上。这个长度显然远远大于形成互补碱基对所需的  $10$  bp 区段, 这说明重组除了碱基互补复性这个条件之外, 还可能有更多的要求。但只要同源序列达到一定的长度, 那么即使相差  $1$  bp 的不同遗传标记仍然有发生重组的可能性。

在真核生物中, 双链 DNA 分子间基因重组是完成减数分裂所必需的, 而且基因重组发生在减数分裂前期 (图 6-12)。染色体变得可见就标志着减数分裂开始, 每条染色体都是已经复制的, 由两条姊妹染色单体组成, 每条染色单体都含有一条双链 DNA 分子。同源染色体相互接近, 在一个或多个



区域配对形成二价体。配对一直进行到整个染色体的同源部位并排在一起,该过程称为联会或染色体配对。联会结束时同源染色体以联会复合体形式紧靠在一起。尽管不同生物中联会复合体在细节上有很大差异,但同一物种内它们都有共同的特征性结构(图 6-12)。

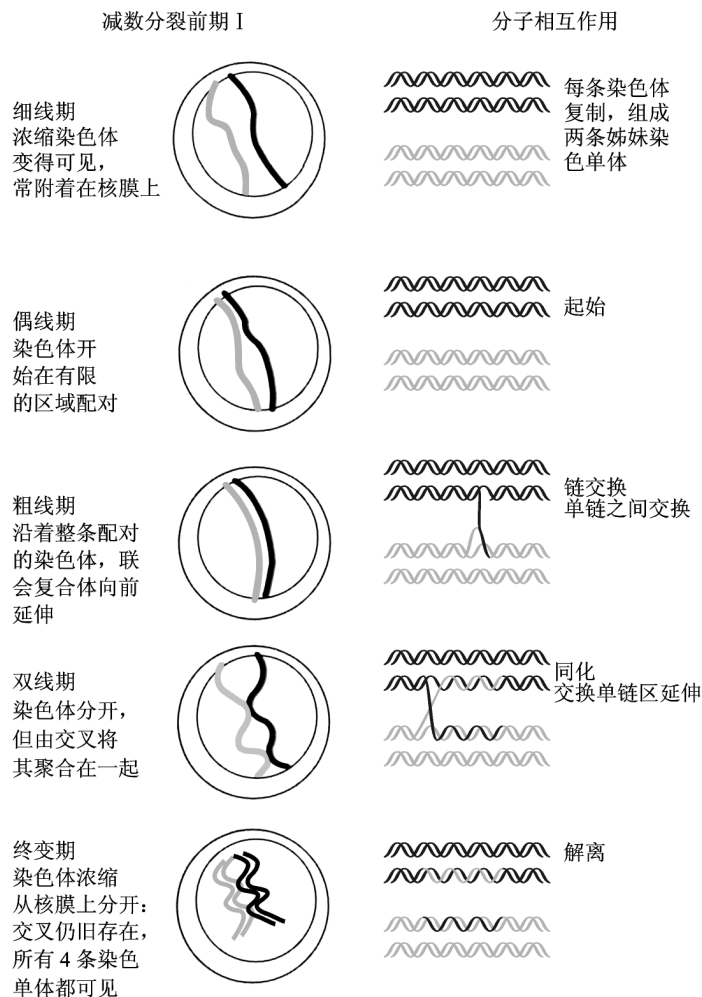


图 6-12 染色体交换与 DNA 分子相互作用的对应关系

染色体间的重组涉及部分遗传物质交换,通常以断裂和重新连接为代表,当染色体开始分开时,可以观察到交叉。交叉点的数量和分布与遗传重组的频率相平行。一般认为一个交叉代表发生了一次交换。重组过程发生的分子基础是每一条姊妹染色单体含有一条 DNA 双链分子,所以每个二价体含有 4 条双链 DNA 分子。在重组过程中,一条姊妹染色单体的双链 DNA 分子与另一条染色体上的一条染色单体之间(即非姊妹染色单体间)相互作用。这种反应可能在两个分子中任何配对的相应序列之间,可使遗传物质交换在单个碱基对水平上精确地发生。众所周知,单链之间的互补性是核苷酸相互识别的基础。迄今人们还不能准确地将染色体水平的改变与其分子机制联系起来。高等真核生物重组的分子机制的一些细节尚不很清楚,在分子水平上解释真核生物染色体的联会仍然很困难。

### 6.3.2 同源重组的分子模型

#### (1) Holliday 模型

1964年 R. Holliday提出了一个重组的杂合 DNA 模型(hybrid DNA model),又称为 Holliday模型(图 6-13),该模型基本内容是:

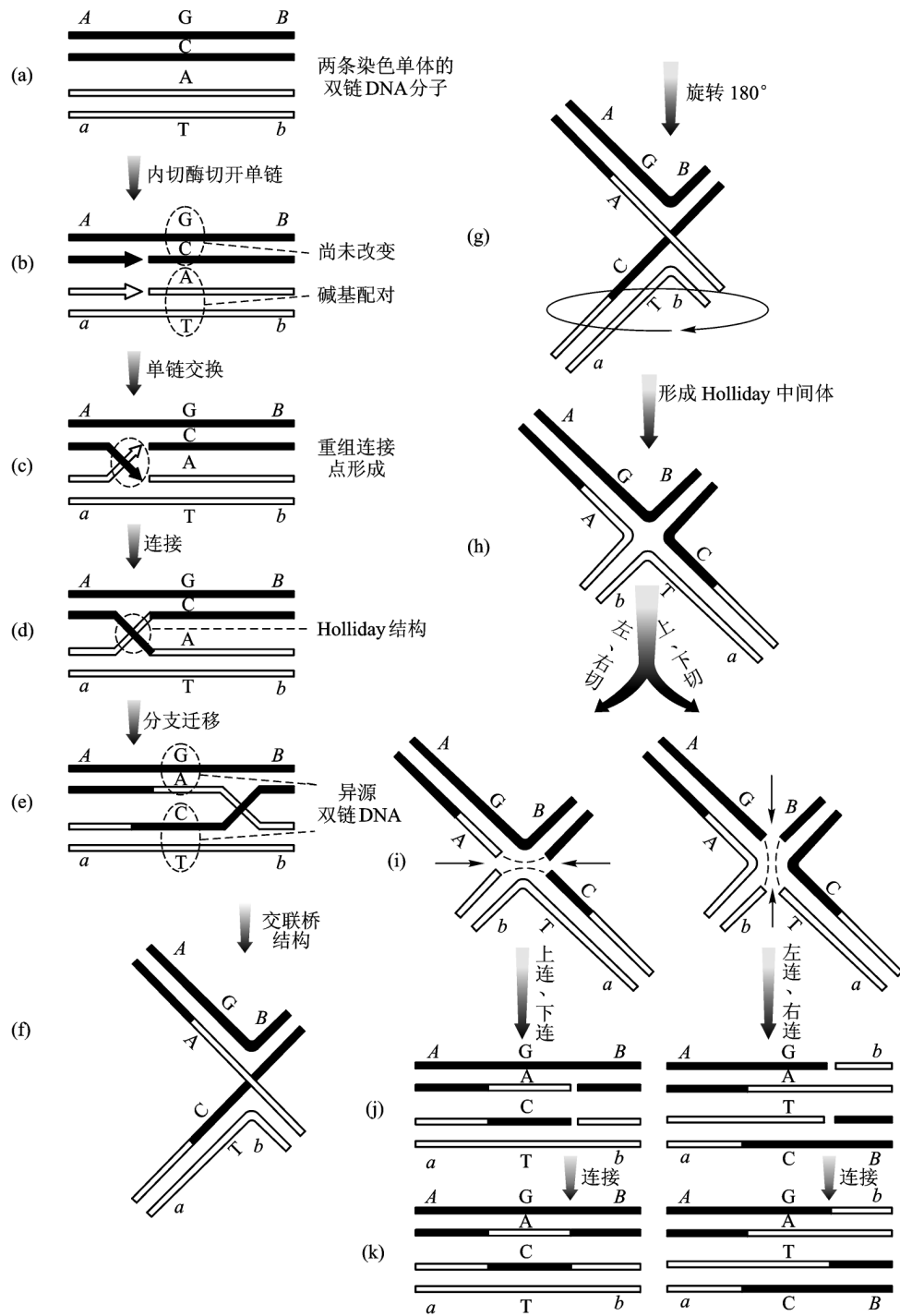


图 6-13 同源重组的 Holliday 模型

图中只显示 4 条染色单体中的两条单体的双链 DNA 分子；DNA 分子的中央基因区以 G—C、A—T 表示，AB, ab 表示旁侧标记基因

① 重组始于两个配对双链 DNA 的同源链中相应位点发生断裂。断裂使切口产生的自由末端可以移动，每条链离开其配对链与另一双链上的互补链交叉配对，相互交换在两个 DNA 双链之间形成了连接，这样所连接的一对双链 DNA 分子称为接合分子 (joint molecule)。一条 DNA 链从自身双链 DNA 交叉到另一双链 DNA 分子的位点就称为重组连接点 (recombinant joint)。在重组连接点，每个 DNA 双链分子都含有一个来自父母双方的 DNA 区域，即杂合 DNA (hybrid DNA) 或异源双链

DNA (heteroduplex DNA)。

② 交叉迁移 重组连接点的一个重要特性是它可沿双链 DNA 分子滑动,这种滑动称为分支迁移 (branch migration)。当双链分子中的单链分支迁移,一条链被另一条链置换时,分支点可以向两边任意一个方向迁移。分支迁移具有重要理论和实际意义,在理论上它赋予了重组结构的一种动态性质;从实际而论,尽管分支点确实存在,但无法通过体外实验来证实它,因分离 DNA 后,分支可能已经移动了。分支迁移速率还不确定,体外观察发生的频率远低于自然状态下大范围异源双链 DNA 的形成频率。因此在体内任何大范围的分支迁移都必定由重组酶 (recombination enzyme) 催化。

③ Holliday 连接体形成 由链交换形成的接合分子可以解离,并形成两条分开的双链 DNA 分子,接合分子相对于一个双链来旋转 180° 另一个双链,则可在一个平面上见到一个称为 Holliday 中间体的结构。链交换形成的接合分子解离而形成两个独立的双链 DNA 分子,这需要再产生两个切口,这一反应的结果是取决于哪一对链被切开。

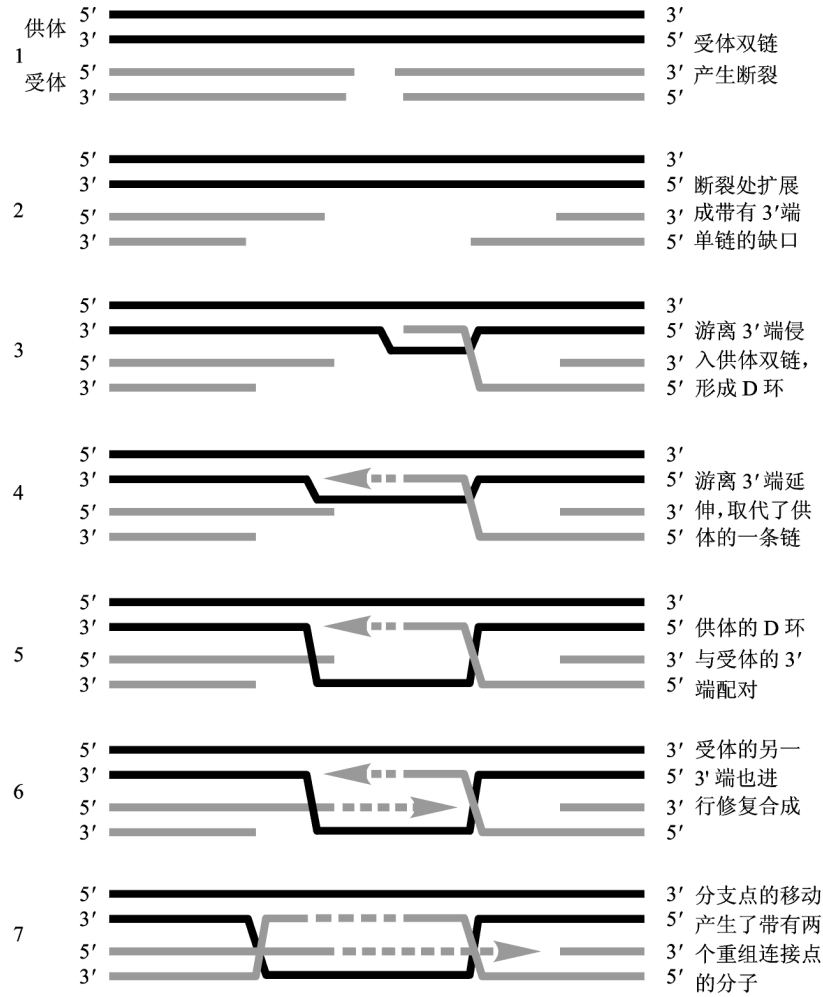
④ 切开的方向决定重组结果 如果第二次切口发生在原初未切开过的两条单链上,即未曾起始链交换的那一对链[图 6-13 ]右方],那么 4 条单链都被切开了,这样就释放出剪接重组 DNA (splice recombinant DNA) 分子。在该双链 DNA 分子中的一条亲体 DNA 链与另一条亲体 DNA 链以异源双链 DNA 的形式共价连接,在异源双链区域两端的分子标记 (A, B) 进行常规的重组事件。但如第二次切口发生在原来被切开过的两条单链上[图 6-13 ]左方],那么另两条单链将是完整的,这样切口释放出原来的两个亲体双链 DNA 分子,仅留下一段异源双链 DNA 区域,这种重组称为补丁重组体 (patch recombinant)。

⑤ 重组产生一段异源双链 DNA 区域[图 6-13 ](k)中 G-A, C-T 区] 从不同的解离方式决定重组的结果可见,双链 DNA 分子间的链交换总会产生一段异源双链 DNA,但是交换不一定伴有两侧区域的重组。那么,在重组双链 DNA 分子之间能建立起联系的最短区域长度是多少呢? 通过质粒或噬菌体的同源序列引入细菌的试验表明,如果同源区段小于 75 bp 会使重组速率显著下降,这个长度显然远远大于形成互补碱基对所需的 10 bp 区段,这说明重组除了碱基互补复性这个条件之外,还可能有更多的要求。

## (2) 双链断裂起始重组模型

重组 Holliday 模型显示了两个双链 DNA 分子的单链依次被切开并参与遗传重组。然而实际遗传交换大多是由双链断裂 (double strand break, DSB) 引发的(图 6-14)。该模型主要步骤如下:① 核酸内切酶在一条染色单体的一个双链 DNA 分子,即“受体”的同一位置切开造成一个双链切口,出现双链断裂后启动同源重组。② 这个切口在核酸外切酶的作用下被扩大成间隙,核酸外切酶从切口的任一端开始切割产生单链 3'端。③ 其中一条 3'游离端侵入另一条双链的同源区(“供体”),这被称为单链侵入 (single strand invasion)。④ 异源双链 DNA 形成时产生一个 D 环(即供体双链 DNA 中的一条链被置换)。⑤ D 环随着 DNA 修复合成而延伸,以 3'端为引物产生双链 DNA 分子。⑥ D 环逐渐扩大,达到相当于受体染色体中整个缺口的长度,当突出的单链到达缺口的远端时,互补的单链序列即发生复性。于是缺口的两端都有异源双链 DNA 分子,而缺口本身只是被单链 D 环补上。缺口以左侧的 3'端为引物,通过修复合成将缺口区恢复成双链。总之,缺口是由两次单独的单链 DNA 合成所修复。分支迁移将此结构转变为一个具有两个重组位点的分子,重组位点必须通过切割来解离。如果两个重组位点以同样方式解离,原来没有发生交换的分子被释放,各自保留一段改变了遗传信息的区域,这是交换过程留下的痕迹。如果两个重组位点以相反的方式解离,结果就发生了遗传重组。

在解离前含有两个重组位点的分子结构显示了双链断裂模型与仅有一条单链发生交换的模型之间存在着显著差异:① 随着双链断裂,发生交换的区域两端都形成了异源双链 DNA 区。两个异源双链体片段之间是相对应于缺口的区域,两个分子都含有供体 DNA 序列。所以异源序列的排列是不对称的,一个分子的一部分被转换到另一个分子的序列中去了,因而将启动的染色单体被称为受体。② 随着单链的相互交换,每条 DNA 双链序列含有从交换起始区域到分支迁移区的异源双链 DNA

图 6-14 双链断裂起始重组模型(仿自 Lew *et al.* 2004)

序列。在不同的单链交换模型中,一些 DNA 被降解,一些 DNA 被合成,而起始的染色单体是遗传信息的供体。③ 双链断裂模型同样说明异源双链 DNA 形成的机制和重要性,而且可以合理解释两个双链 DNA 是如何相互作用的。机体把起始重组的任务从单链转移到双链断裂,这种双链断裂机制而不致细胞死亡,说明细胞对 DNA 操纵的能力。④ 一般是 DNA 分子一旦完全被切开,它就是进入一个不可逆过程。因为在单链断裂模型中没有任何信息丢失,但在双链断裂模型中,切口形成后信息就丢失了。在信息恢复过程中所发生的任何错误都将可能是致命的,但细胞可通过另一条 DNA 双链的重新合成以恢复所失去的信息,这一特殊能力又为细胞构成了安全网。

### 6.3.3 联会复合体与重组

在细胞的减数分裂过程中,同源染色体配对并且形成联会复合体。多年来,此过程被认为与重组有关,有可能是 DNA 重组的必要前提。然而近年来研究表明,联会复合体是重组的结果,而不是原因。

#### (1) 联会复合体在双链断裂后形成

酵母中研究结果证明,不论是同源重组还是位点专一性重组,只有双链断裂才能起始重组。双链断裂也发生在减数分裂早期,而且是在联会复合体形成之前[图 6-15(a)],发生在一些重组的热点上,即重组的起始位点,重组频率从起始位点一侧或两侧开始呈梯度下降,暗示了重组事件可能从这个位点开始向两边扩散。双链断裂形成的平齐末端在两边迅速地转变为长 3 单链末端。但是在一

种酵母突变(*rad 50*)中,由于突变阻止平齐末端转变为单链突出端,导致重组无法进行,可见双链断裂对于重组是必需的前提。离双链断裂位点越远,形成单链区域的可能性越小,所以重组频率就呈梯度下降。在*rad 50*突变体中,双链断裂后的5'端与Spoll蛋白相连[图6-15b]。只有Spoll被移走后,核酸酶才能发挥作用,并证明至少有9种其他蛋白质共同参与双链断裂过程,一组蛋白质将双链断裂端转变为3'-羟基突出的单链末端;另一组蛋白质使单链末端侵入同源双链DNA分子。

重组与联会复合体的形成关系密切,果蝇和酵母中研究结果表明,凡是不能进行染色体配对的突变体都不能发生重组。由此可见产生双链断裂并进一步引发重组的系统是很保守的。Spoll的同源序列已在一些高等真核生物中被鉴定出来,一种果蝇基因的突变型阻断了所有减数分裂中的重组。但将重组过程中分子水平上发生的事件与细胞水平发生的事件能直接进行比较的系统极少,最近在酿酒酵母减数分裂中研究取得一些进展。

双链断裂出现到消失这段时间超过60 min,第一个接合分子,即中间体在双链断裂后很快出现。双链断裂在中央成分(中体)形成时发生,它们在联会复合体形成时消失,其相对时相为:双链断裂,重组中体形成,引发重组过程中联会复合体形成。根据是*rad 50*突变型不能将中央成分转变为联会复合体的现象,显然认为在减数分裂中,联会复合体是染色体配对所必需的,进而使分子水平引起重组这一传统观点受到了挑战。在酵母的研究中发现,联会复合体是在双链断裂起始重组之后形成,它一直持续到重组体形成,故联会复合体对于重组的形成不是必需的。而在一些缺失正常联会复合体的突变体中重组型仍可以形成,不能进行重组的突变型则不形成联会复合体。这充分表明,联会复合体仅仅是随染色体配对之后的重组结果[图6-15a]。

由Holliday重组模型人们一直认为:Holliday结构的解离方式决定了最终是产生非交换产物还是交换产物,也就是说,取决于解离发生在哪些链上。但是最近对于非交换和交换产物生成时间的分析表明,交换产物直到接合分子第一次产生后的相当一段时间才会出现,而非交换产物几乎与接合分子同时出现。如果两种产物由同一解离过程产生,那么它们就应在同一时间出现。这一时间上的差异表明,交换产物不是按原先设想的是通过接合分子解离产生的,而是由另一条途径产生。

## (2) 同源染色体配对与联会复合体的形成是两个独立的过程

突变可以发生在染色体配对或是联会复合体形成的任一过程,并且彼此互不干涉。*Zip2*突变型中染色体可以配对,但不能形成联会复合体,所以同源染色体之间的识别不依赖于重组或是联会复合体的形成[图6-15b]。在酿酒酵母中,同源染色体之间的特异性联系是由基因*Hop2*控制的,*Hop2*的作用是防止非同源染色体的相互作用。在*hop2*突变型中,减数分裂虽然产生正常数量的联会复合体,但每个复合体含有非同源染色体,这表明,这类联会复合体的形成不依赖于同源性。*hop2*突变型中联会复合体内错配(*mismatch*)的染色体发生双链断裂,但是它们不能被修复,这表明联会复合体的形成需要双链断裂,但复合体的形成并不要求这些断裂进一步与同源DNA相互作用。

在DNA重组体被观察到之前,推测在粗线期可能经历着重组之后的一系列重要事件,包括链的交换、DNA合成和解离。到双线期,联会复合体解聚,染色体脱去联会复合体形式,便可见染色体相互连接的交叉,这被认为是发生了遗传交换的标志。但交叉在分子水平上的本质尚不清楚,有可能它是交换完成的残留痕迹或是代表同源染色体之间的联系,此时遗传交换还未完成。直到减数分裂后期,交叉向染色体两端移动,这一动态变化表明,交叉代表重组事件的残留步骤,而不是作为重组的中间体。

重组可能是受某种交换控制(*crossover control*)的过程,如重组发生在减数分裂染色体的不同位置上,只有一小部分相互作用真正地发展为交换,而且每对同源染色体上分布1~2个交换,完全不发生交换的概率也是很低的,同时某些突变型中交换的非随机性普遍受到干扰。此外,重组的发生对减数分裂过程是重要的,它存在着一个“关卡”(checkpoint)系统,以确保细胞不会在重组尚未发生前就分离染色体。

迄今还不能完全揭示重组的发生与所观察到的不同结构——重组结与交叉的联系,且在分子水平上阐明其本质。尽管近年来在酵母重组中,鉴定出一些与重组相关的蛋白质,它们分布在不同位点,为研究重组中一些有关形成不连续结构的机制奠定了分子基础,其中有些蛋白质与原核生物重组相关蛋白质同源,但这些蛋白质在重组中的确切功能仍有待进一步研究。

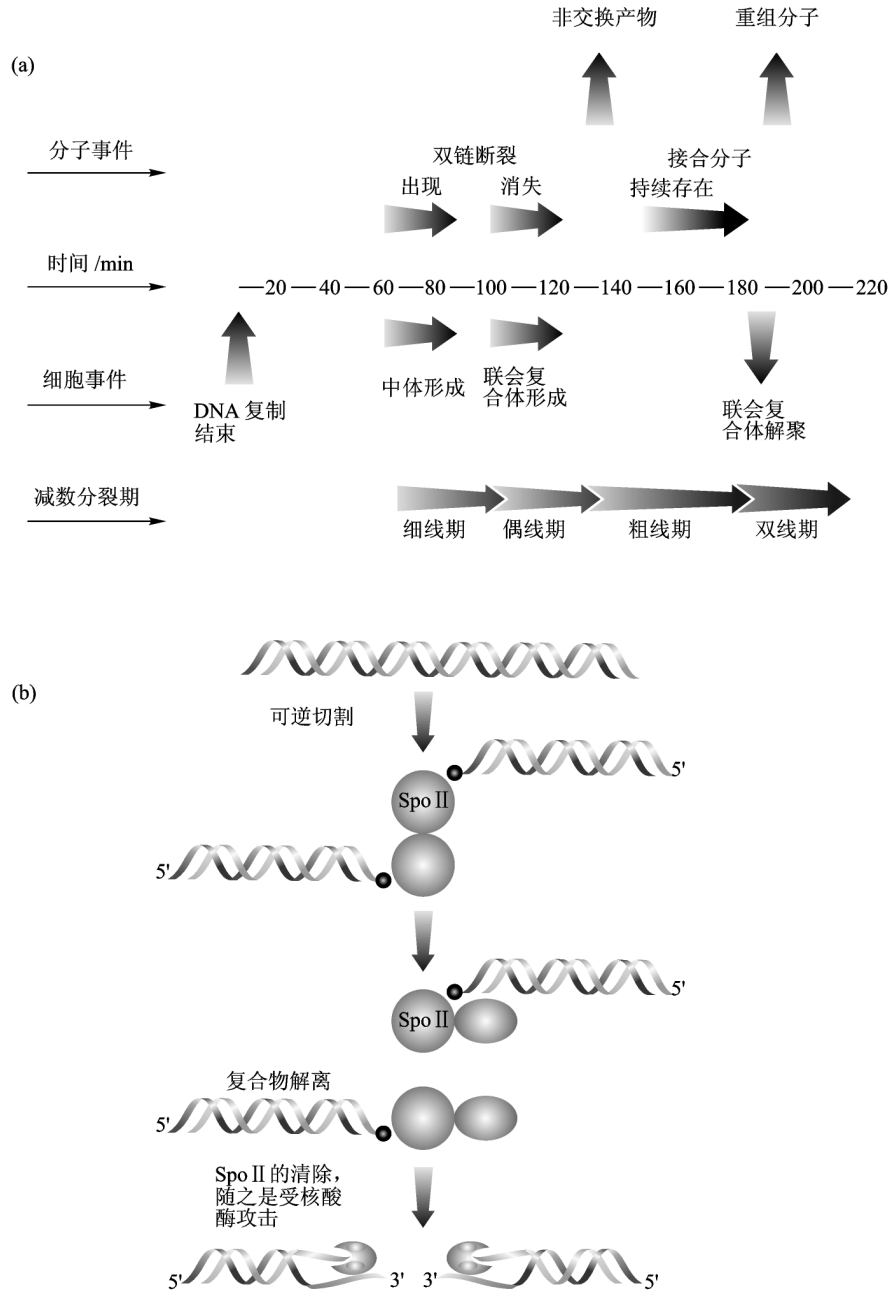


图 6-15 酵母细胞减数分裂重组中有关分子的作用

(a) 减数分裂中分子交换与细胞学元素的关系 (b) SpoII 产生双链断裂

## 6.4 基因转变及其分子机制

### 6.4.1 异常分离与基因转变

遗传重组导致异源双链 DNA 片段的形成,即异源双链 DNA 含有错配碱基序列,由此揭示了等位基因间重组特性,而等位基因重组又为异源双链模型的发展提供了证据。等位基因重组时,基因座内染色体发生断裂和重接产生了交互的重组染色体对。然而在一对染色体的四分体时期,异源双链 DNA 分子本身通常会引起重组事件的发生。子囊菌类(ascomyetes fung)可用来研究等位基因重组事件。由 6-2-1 已知一对等位基因杂交子代的子囊不外乎形成 6 种正常分离类型(AAAAaaaa 或 aaaaAAAA, AAaaAAaa 或 aaAAaaAA, AaaaaaaaAA 或 aaAAAAaa),这只是四分子中是否发生重组和纺锤体随机取向而形成,两种孢子的比例相等,即等位基因 A: a= 4: 4。

但是后来发现有些孢子分离比例异常,如 3: 5 和 6: 2 以及异常的 4: 4 分离。如在酵母(*S cerevisiae*)不同交配型 A× a 中,有些子囊所含的孢子为(3A+ 1a),而不是预期的(2A+ 2a)。在脉孢菌和粪生粪壳菌(*Sodaria fimicola*)等子囊菌中也发现了一些异常分离(abnormal segregation)现象。Mitchell 将与维生素 B<sub>6</sub> 合成有关的不同的突变型进行杂交,吡哆醇(pyridoxine, 维生素 B<sub>6</sub>)的突变基因 pdxp 需要在培养基上添加吡哆醇后才能生长,而且对 pH 敏感,改变酸度后,可不添加吡哆醇。其邻近的另一突变基因 pdx 也是吡哆醇依赖型突变,但对酸度不敏感。将两个吡哆醇突变株杂交, + pdxp× pdx+, 取得子一代子囊后,对 585 个子囊中的孢子依次解剖,进行培养和鉴定。他发现其中 4 个子囊中有野生型的孢子对出现,可是跟预期的相反,如果这两个基因间发生了重组,重组后应该同时出现的双突变型(pdx pdxp)孢子对却没有发现(表 6-7)。

表 6-7 粗糙脉孢菌中 + pdxp× pdx+, 其中 4 个子囊的结果

孢子对	子 囊								
	1		2		3		4		
第一对	+	pdxp	pdx	+	+	+	+	pdx	+
第二对	+	+	pdx	+	+	pdxp	+	+	pdxp
第三对	+	pdxp	+	+	pdx	+	+	+	+
第四对	pdx	+	+	pdxp	pdx	+	pdx	+	+

怎样解释这一现象? 不可能是突变,因为它们的频率远比这些基因的正常突变率高得多。分析的方法是灵敏的,如果有双突变的话,就能够检出。Mitchell 的上述实验结果用图 6-16 来表示。

由于重组,出现了完全野生型的孢子对(+, +),但没有重组的对应产物——双突变型的孢子对(pdx pdxp),结果应如图 6-16 的左侧所示。然而右侧发生的情况说明尽管 pdxp 出现异常的 3: 1 分离,但紧密连锁的 pdx 基因却显示出正常 2: 2 分离。这些反常的情况,好像是一个基因转变为它的等位基因,这种现象称为基因转变(gene conversion)。图 6-16 中,基因 pdxp 转变为 pdxp<sup>+</sup>(或 +),图中以 ⊕ 表示该基因发生了基因转变。所以双突变型的孢子对没有出现。

此外,基因转变往往伴有转变区外基因的重组,但区外基因的重组是正常的交互方式,所以虽然 pdxp 基因座出现了异常的 3: 1 或 6: 2 分离,而邻接的基因 pdx 仍显示正常的 2: 2 或 4: 4 分离。

稀有的异常分离现象也在粪生粪壳菌中发现。Olive 等在粪生粪壳菌的野生型黑色孢子(g<sup>+</sup>)与突变型灰色孢子(g<sup>-</sup>)杂交中,曾观察了 200 000 个子囊,其中绝大部分属于正常的 4g<sup>+</sup>: 4g<sup>-</sup> 类型。但

却有少数子囊是例外的比例,其中包括  $5g^+ : 3g^-$  120个,占 0.06%;  $6g^+ : 2g^-$  100个,占 0.05%; 以及不规则的 4:4或是 3:1:3:1 16个,占 0.008%。

### 6.4.2 基因转变的类型

基因转变的类型分为:①染色单体转变(chromatid conversion),即减数分裂的4个产物中有一个产物发生了基因转变,故出现  $6g^+ : 2g^-$  或  $2g^+ : 6g^-$  类型的子囊[图 6-17(a)]。②半染色单体转变(half chromatid conversion),即减数分裂的4个产物中,有1个产物的一半或两个产物的各一半出现基因转变,因而形成 5:3或 3:5和异常 4:4即 3:1:3:1类型的子囊,基因转变只影响半个染色单体分离显然是发生在减数分裂后的有丝分裂中,所以又称为减数后分离(postmeiotic segregation)[图 6-17(b),(c)]。

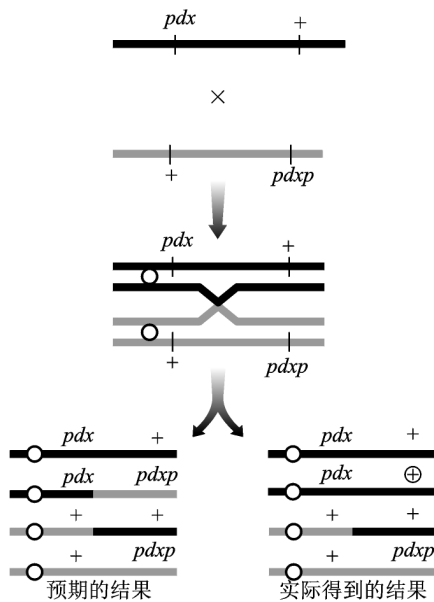


图 6-16 粗糙脉孢菌的基因转变

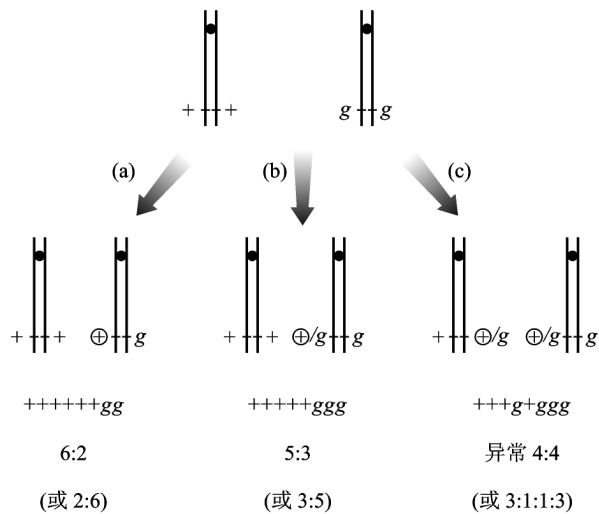


图 6-17 基因转变的类型

### 6.4.3 基因转变的分子机制

由 Holliday模型和 DSB起始重组模型都已清楚表明在对称的异源双链区存在着不配对碱基(如 G—A, C—T),形成两个杂种分子。异源 DNA是不稳定的,细胞内的修复系统能够识别不配对碱基,并以切除修复的方式完成修复过程(详见第 13章),杂种分子得到校正。根据切除修复原理,基因转变的几种类型产生的分子机制可以归纳如下(图 6-18)。

(1) 两个杂种分子均未校正,复制后出现异常的  $4+ : 4g$ (或  $3 : 1 : 1 : 3$ )的分离。由于不配对的碱基没有得到修复校正,杂种 DNA分子保留到下一次复制。在下次复制即减数分裂以后的复制中,异源双链 DNA复制形成两个不同的等位基因,通过有丝分裂发生分离[图 6-18(a)]。减数分裂的4个产物中的两个染色单体的各 1/2出现基因转变,即仅仅影响 DNA双链中的一条链,实际只影响两个单体中的半个染色单体,即 DNA双链中的各一条链。显然转变的链与未转变的链的分离一定是在减数分裂后的有丝分裂中发生。因而这类基因转变属于半染色单体转变,或减数后分离,子囊孢子的异常分离比为 3:1:1:3。

(2) 一个杂种分子校正为+,或校正为g时,则发生另一种类型的半染色单体转变,前者修复后出现 5:3的分离,后者子囊孢子的异常分离比为 3:5[图 6-18(b)]。由此可见,减数分裂4个产物中



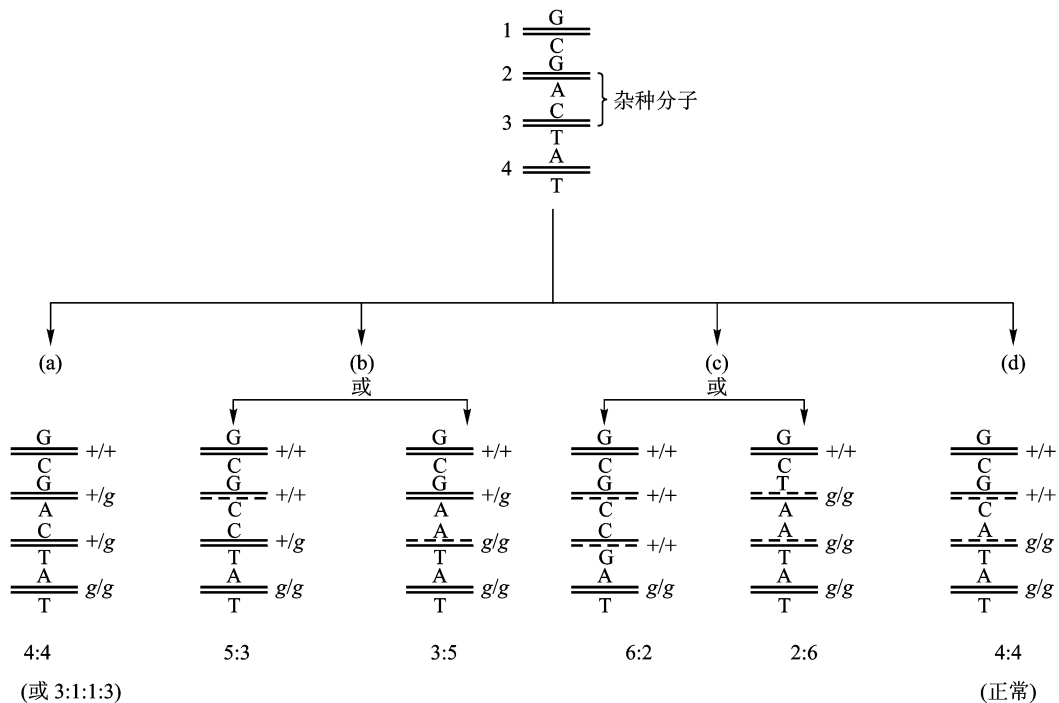


图 6-18 基因转变的分子机制示意图

的一个产物(第 3 条染色单体或第 2 条染色单体)的  $1/2$  出现转变。由于只影响半个染色单体,所以转变的链与未转变的链将在减数分裂后的有丝分裂中发生分离。

综合上述两种修复情况可见,半染色单体转变是当两个杂种分子都未进行修复校正或一个杂种分子校正到+ (或 g) 时,使得减数分裂的 4 个产物中的两个产物的各  $1/2$  或一个产物的  $1/2$  发生了基因转变。其子囊孢子的典型异常分离比分别为 3:1:1:3 与 5:3 或 3:5。

(3) 两个杂种分子都被校正到+ (或 g) 时,修复后出现 6+:2g (或 2+:6g) 的异常分离[图 6-18 c]。这便是出现染色单体转变的起因。

(4) 当两个杂种分子都按原来两个亲本的遗传结构进行修复时,则减数分裂 4 个产物恢复成 G-C、G-C、A-T、A-T 的正常配对状态,子囊孢子分离正常,呈现 4+:4g 的结果[图 6-18 d]。

切除修复系统识别异源双链 DNA 中错配的碱基对时,切除其中的一条链,使其恢复互补性。于是,使得代表一个等位基因的 DNA 链变成了代表另一个等位基因的序列。如果这种校正变化只发生在一个异源双链 DNA 分子上,就会产生 5:3 或 3:5 的异常分离。如果两个异源双链以同样方式校正,就产生了 2:6 或 6:2 异常分离。所以基因转变实质上是异源双链 DNA 错配的核苷酸对在修复校正过程中所发生的一个基因转变为它的等位基因的现象。

由于某些不配对碱基切除修复校正的结果与该区发生双交换的效应一致,其实是基因转变的结果,因此出现 C 值(并发系数)大于 1,这称为负干涉(negative interference),即一个区域的交换引起邻近区域另一次交换频率的增大的现象。已知基因转变可以伴随或不伴随两侧遗传标记的重组,这取决于 Holliday 结构的切断方向(图 6-13)。大量实验结果表明,大约有一半的子囊发生基因转变的同时伴随两侧遗传标记的重组,而“负干涉”形成的原因显然是基因转变的同时并不伴随两侧遗传标记的重组。基因转变不仅是专一的,而且是有方向的,它不仅涉及单个位点(或基因座),而且涉及染色体的一个区段,如一对含有两个基因差异的突变型杂交时,在某些子囊中可以发生几个基因同时发生转变的现象,称为共转变(coconversion)。共转变的频率大于独立转变的预期频率,两个基因距离越近,共转变频率越大,即在异源双链的同一区域内沿相同方向同时被修复校正的概率越大。一般说,愈接近单链断裂点的位置愈容易发生基因转变,愈远则愈少发生基因转变,这也说明基因转变是有极

性(polarity)的,并由此提出极化子模型(polaron model)。该模型假定内切酶首先作用于基因的一端,从起点开始,基因转变频率由高到低形成一个梯度,在染色体上呈现基因转变极化现象的这样一个区域称为一个极化子(polaron),有时一个极化子就相当于一个基因。

## 6.5 体细胞交换与基因定位

### 6.5.1 单倍体化与体细胞交换

在真核生物的体细胞中也会发生基因重组。有些真菌是研究体细胞染色体重组的好材料。一种绿色霉菌——构巢曲霉(*Aspergillus nidulans*)在自然界一般以单倍体状态存在,其分生孢子是单核的,可是在实验室中可以获得稳定的二倍体菌株。两个不同营养缺陷型菌株所产生的分生孢子大量混合接种在基本培养基的表面,可以得到少量的原养型菌落。由于这些原养型菌落细胞含有来自两亲本的细胞核,因此称为异核体(heterokaryon)。异核体在自然界里广泛存在。如两种不同基因型真菌的菌丝接触时,菌丝之间连接起来,细胞质和细胞核就混合在一起。在菌丝连接处长出的菌丝细胞中便含有两个不同基因型的细胞质和细胞核,这种异核体具有生长优势,类似于高等植物的杂种优势。同时异核体中包含不同基因型的细胞核,因此增强了对环境的适应能力。

大量异核体中的核仍保持单倍体状态,但有少数单倍体细胞核融合成为二倍体细胞核或合核体(synkaryon)。利用分生孢子颜色突变的遗传标记来鉴别一种菌落是二倍体还是异核体。如构巢曲霉的野生型分生孢子是绿色的( $w^+ y^+$ ),其突变型有黄色分生孢子( $w^+ y^-$ )和白色分生孢子( $w^- y^+$ )。由产生黄色孢子的某一营养缺陷型( $A^- B^+ w^+ y^-$ )和产生白色孢子的另一缺陷型( $A^+ B^- w^- y^+$ )组成的异核体的基因型为: $\left[ \frac{A^- B^+ w^+ y^-}{A^+ B^- w^- y^+} \right]$ 。异核体所产生的分生孢子就是两个亲本所产生的孢子,因为这些单倍体在异核体中是各自分开的。由异核体产生的分生孢子则有黄、白两种。两个单倍体核偶尔也可以融合到一起形成上述二倍体菌株,由这种二倍体菌株所产生的分生孢子则都是绿色的。因为黄色和白色突变型对于野生型的绿色而言都是隐性的,所以孢子呈绿色。

二倍体比异核体较为稳定,但是这种稳定性也是相对的,因为从大量二倍体分生孢子中也可以得到少数体细胞分离体(segred),所谓分离体是重组体和非整倍体或单倍体的总称。在这里产生非整倍体或单倍体的过程称为单倍体化(haploidization),产生重组体的过程称为体细胞交换(somatic crossing over)。

单倍体化是在有丝分裂过程中染色体不分离的结果。在正常的体细胞有丝分裂中,一条染色体分裂为两条,这两条染色体各自趋向一极,使两个子细胞中各有一条。最终导致了在表型上的分离现象。最初是 Bridges 观察果蝇杂合基因型  $M^+ M$  ( $M$  是决定细长刚毛性状的显性基因)的雌蝇时,发现有些具有  $M$  基因的雌蝇长出了一块扇形的野生型刚毛。表明杂合体的等位基因在表型上发生了分离。Bridges 证明这是有丝分裂不分离(mitotic nondisjunction)造成的[图 6-19 a]。 $2n+1$  也称为三体,它常失去一条染色体而成为二倍体。在此过程中,可以由杂合体转变为纯合体; $2n-1$  也称为单体,生长迟缓,且不稳定,常进一步失去其他染色体而最后成为一个稳定的单倍体[图 6-19 b]。在单倍体化过程中,如果排除染色体携带的是某一个显性基因,那么相应的隐性基因所决定的性状就得以表现,这正像高等植物杂交在子一代中不分离,而在子二代中分离出隐性性状个体的原理一样。单倍体化过程使隐性性状表现,这种现象也称分离。此外,杂合体细胞中在有丝分裂后的子细胞核重建时,发生一条染色体丢失的现象称为有丝分裂染色体丢失(mitotic chromosome loss),从而也导致表型的分离[图 6-19 c]。

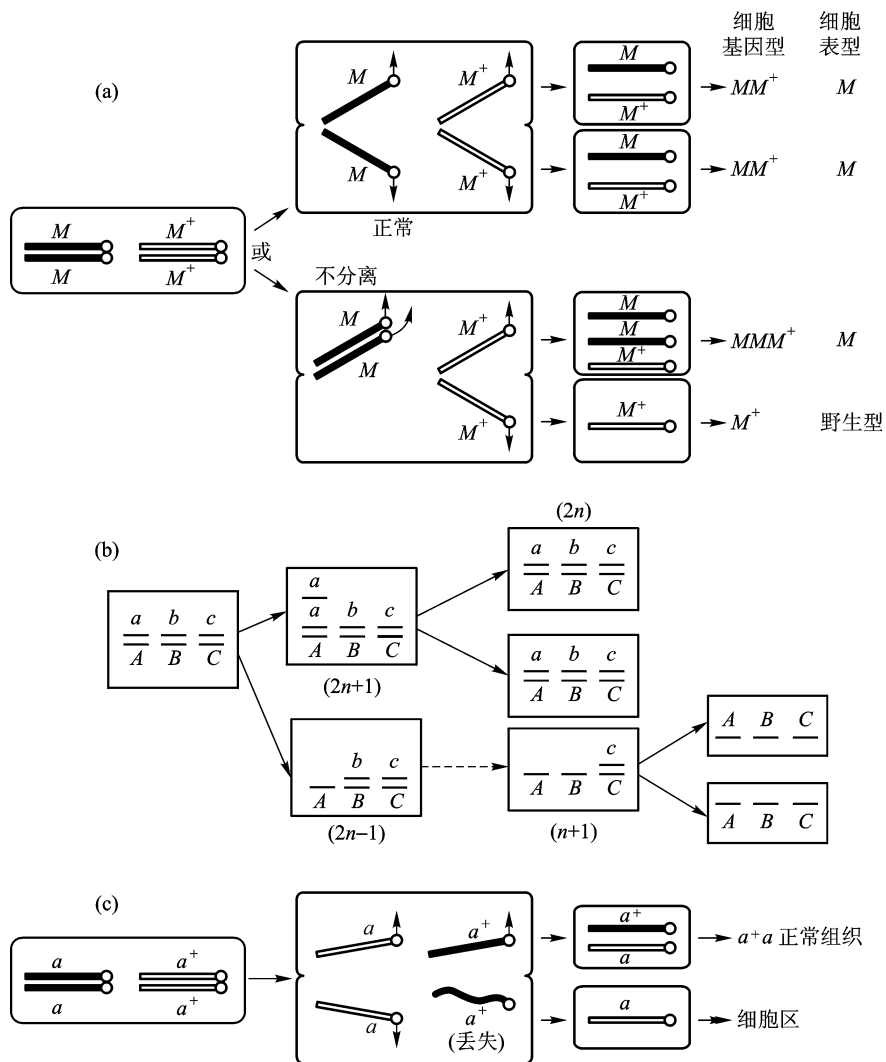


图 6-19 体细胞染色体重组

(a) 有丝分裂不分离导致表型分离 (b) 单倍体化过程示意图 (c) 有丝分裂染色体丢失

杂合二倍体的体细胞交换是产生分离子的第二个途径。一般在有丝分裂中同源染色体并不像在减数分裂时那样进行有规则的联会与交换,如杂合二倍体  $\begin{pmatrix} d & a & b & c \\ + & + & + & + \end{pmatrix}$  细胞有丝分裂后的子细胞都是  $\frac{d}{+} \frac{a}{+} \frac{b}{+} \frac{c}{+}$  基因型,杂合二倍体  $\frac{d}{+} \frac{a}{+} \frac{b^+}{+} \frac{c}{+}$  的子细胞的基因型都是  $\frac{d}{+} \frac{a}{+} \frac{b^+}{+} \frac{c}{+}$ 。但构巢曲霉的体细胞在有

丝分裂过程中,同源染色体间却可发生染色体交换,这称为体细胞交换。由于体细胞交换可导致原杂合二倍体的部分基因纯合化,这种现象也称为有丝分裂交换(mitotic crossing over)。以构巢曲霉 5 个隐性基因 *paba*(对氨基苯甲酸)、*y*(黄色孢子)、*ad<sub>6</sub>*(嘌呤<sub>16</sub>)、*ad<sub>8</sub>*(嘌呤<sub>8</sub>)、*bi*(生物素)的杂合二倍体  $paba y^+ ad_8^+ / ++ ad_6^+ b$  为例,说明在发生体细胞重组后形成末端 *ad<sub>8</sub>* 与 *bi* 基因纯合二倍体的过程(图 6-20)。

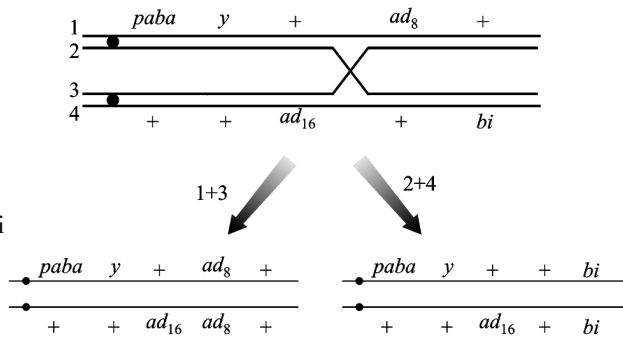


图 6-20 构巢曲霉的体细胞重组

从图 6-20 可见,在杂合二倍体  $\frac{paba\ y\ +\ +\ ad_6\ +}{+\ +\ ad_6\ +\ bi}$  的有丝分裂中,出现  $ad_6$  和  $bi$  两基因纯合的二倍体  $\frac{paba\ y\ +\ +\ bi}{+\ +\ ad_6\ +\ bi}$  以及  $\frac{paba\ y\ +\ +\ ad_6\ +}{+\ +\ ad_6\ +\ ad_6\ +}$ 。在杂合二倍体的无性繁殖子代中,由于隐性基因纯合状态的出现,而表现隐性性状的现象也称为分离,该体细胞重组体 (somatic recombinant) 亦称为分离子。由此可见,分离子可以是二倍体、非整倍体,也可以是单倍体;它既可来源于体细胞重组,又可来源于单倍体化。不过有丝分裂交换和单倍体化是两个彼此独立的事件,它们同时发生在一个细胞中的概率很小。

Stem 在果蝇有丝分裂中发现了连锁基因的交换。他使用了易于辨别的两个伴性遗传的隐性性状:一个是黄体 ( $y$ ), 一个是焦刚毛 ( $sn$ )。前者影响体色,后者改变刚毛的形状。既然这些基因都是隐性的,其基因型为杂合体的雌果蝇 ( $y^+/+ sn$ ) 照例应表现出野生型的灰体与直刚毛。但是 Stem 发现,部分雌蝇中的灰体上单独有一小块黄体,或少数个体的直刚毛中嵌合出一块单独的焦刚毛,更有趣的是出现一小块黄体与一小块焦刚毛的孪生斑 (twin spots)。所谓孪生斑指的是两块互相靠近而面积大小相当的斑点 [图 6-21 (b)]。Stem 注意到孪生斑发生频率较高,这不能视作两个单独斑点的概率偶然并存,而有其必然性。这种现象无法用体细胞突变或其他原因解释。Stem 认为,这是由于体细胞在有丝分裂过程中发生同源染色体间的交换造成的 [图 6-21 (a)]。

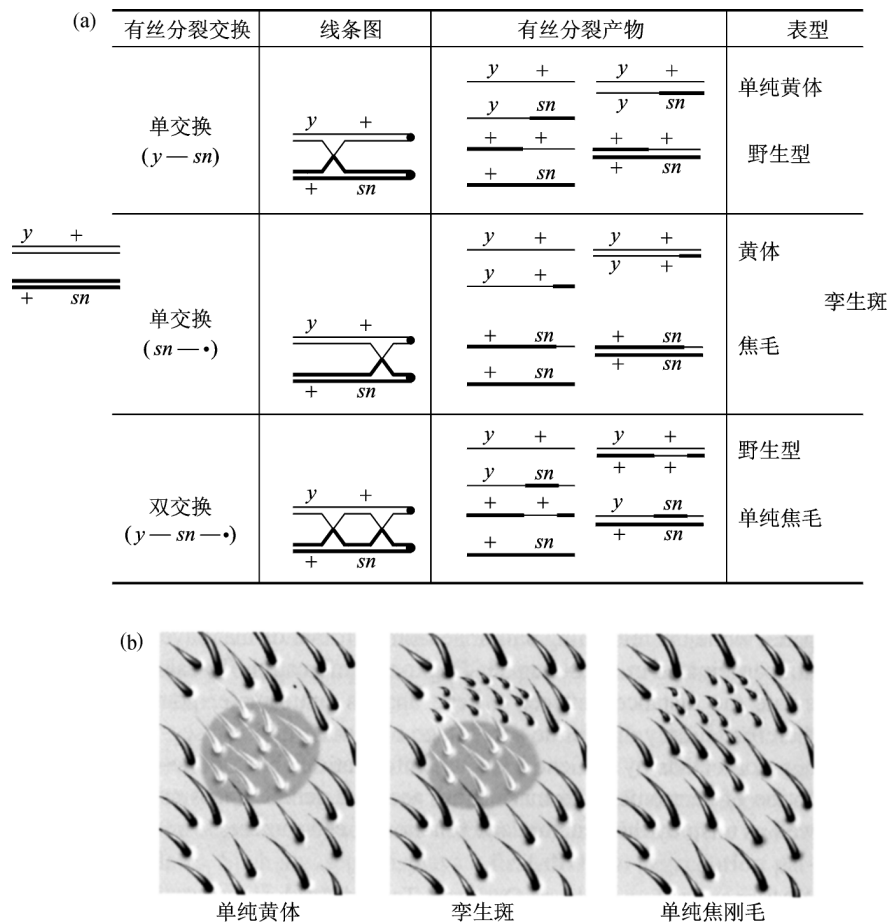


图 6-21 体细胞有丝分裂交换

(a) 体细胞有丝分裂交换 (b) 果蝇中黄体、焦刚毛孪生斑

### 6.5.2 有丝分裂交换与基因定位

基因定位一般是通过有性过程中细胞的减数分裂、同源染色体联会、基因间的重组率来确定基因的位置的。由于在有丝分裂过程中一般不发生基因重组,所以在无性生殖的生物中无法通过遗传重组进行基因定位。自从在构巢曲霉中发现体细胞重组以后,就可通过有丝分裂交换进行基因定位。其基因定位的原理是依据体细胞同源染色体的交换使得染色体远端的杂合基因纯合化的规律,这个规律可用来确定基因的排列位置和距离(图 6-22)。

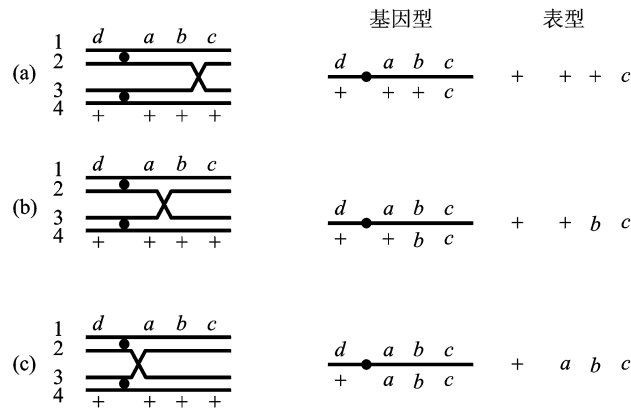


图 6-22 有丝分裂基因定位的原理

从图 6-22可以看出,交换如发生在 b和 c之间,那么带来 c的纯合。如发生在 a和 b之间,使 b和 c纯合。如交换发生在着丝粒和 a之间,则 a b和 c纯合。由此可见,离着丝粒愈近的基因纯合的机会愈小,愈远的则愈大,而且着丝粒一端的基因纯合影响着丝粒另一端基因 d的纯合。因此如果发现两个染色体臂上的基因同时出现纯合化现象,就有可能是一条染色体丢失的结果。对于一个未知属于哪一个连锁群的基因而言,如果它是随着某条染色体两个染色体臂上的基因的纯合而呈现纯合化现象,那么就可初步判断该基因属于这一连锁群。真菌在二倍体的有丝分裂过程中,偶尔发生同源染色体交换,导致连锁基因的重组,这一遗传变异过程称为准性生殖(parasexuality)。准性生殖产生的重组体细胞一般和营养体细胞没有形态、生理上的差异,而且不产生在特殊的囊器中。准性生殖过程中染色体交换和染色体的减少是不规律且不协调的,这便是准性生殖过程与有性生殖过程中基因重组的主要区别。现以构巢曲霉中一个具有  $y$ ,  $ad$ ,  $pab$ 和  $pro$ 以及它们的显性等位基因的杂合二倍体为例来说明。先检出二倍体黄色分生孢子( $\bar{y} / \bar{y}$ ),随后按它们所产生的表型来确定分离子的基因型。在  $y$ 的这些隐性纯合子中有 55%为  $pab$ 和  $pro$ 的纯合子,72%是  $pab$ 纯合子,22.5%为原养型。由于没有一个  $y$ 分离子是需腺苷酸的,所以可推断  $ad$ 位于染色体的另一臂上。染色体一个臂上的有丝分裂交换对于其另一臂上等位基因的分离无效,而在一条染色体的两臂同时发生有丝分裂交换的情况又极其罕见。由此通过有丝分裂定位得到的构巢曲霉的第一染色体右臂的各基因之间的图距如下:

$$\frac{ad \quad 69}{(32)} \cdot \frac{55 \quad pro \quad 72}{(41)} \quad \frac{pab \quad 22.5 \quad y}{(39)}$$

染色体上面是有丝分裂数据,下面是减数分裂定位数据,两者定位数据相差甚远,这正是有丝分裂与减数分裂交换过程的基本差别,但二者在顺序上是一致的。构巢曲霉既可进行准性生殖,也可进行有性生殖。因而基因定位可通过有丝分裂和减数分裂定位数据对照分析。但是黑曲霉(*Aspergillus niger*)、产黄青霉(*Penicillium chrysogenum*)和酱油曲霉(*Aspergillus sojae*)等真菌都不进行有性生殖,对于这些真菌而言,就只能通过准性生殖过程来进行遗传学分析。

## 6.6 体细胞融合与基因定位

### 6.6.1 细胞融合与基因定位

细胞融合 (cell fusion) 是体细胞遗传学的核心内容,也是应用体细胞遗传学技术进行基因定位的基础。1958年 Okada 第一次观察到两个不同的肿瘤细胞在注射了日本血凝病毒,即仙台病毒 (hemagglutinating virus of Japan, HVJ) 后可融合在一起。以后 Barsk 等报道,两个具有不同标记染色体的细胞系经过共同培养后可形成杂种细胞 (hybrid cell)。特别是 Harris 和 Watkins 证明,用 UV 灭活的仙台病毒介导细胞融合过程,该病毒可进入细胞膜内,在邻近的细胞之间形成细胞质桥 (cytoplasmic bridge),从而大大地提高了细胞融合频率。此外,用聚乙二醇 (polyethylene glycol PEG) 也可显著地提高细胞融合频率,将一定浓度的 PEG 加入到培养液内,就可使细胞发生凝集。由于 PEG 是非生物试剂,融合效率高,易于标准化,且价格便宜,因此 PEG 已成为广泛采用的融合剂。大量研究证明,许多动物的不同类型细胞之间都可进行融合,如小鼠、中国仓鼠、家兔和人的成纤维细胞、巨噬细胞、淋巴细胞或白细胞之间的融合。细胞融合过程可分为异核体 (heterokaryon) 和杂种 (hybrid) 细胞形成两个阶段。

在异核体阶段融合的细胞内含有来自两个亲本的细胞核。随后,异核体同步进入有丝分裂,核膜崩溃,来自两个亲本细胞的基因组合在一起形成只含有一个细胞核的杂种细胞。杂种细胞对于遗传学最有价值的贡献是在基因定位研究中的应用。

将两个小鼠细胞 (如  $A_9$  和  $B_{82}$ ) 或两个中国仓鼠细胞 (如突变型  $gly^-$  和  $ade^-$ ) 融合在一起,形成的杂种细胞一般都保留来自两个亲本细胞的染色体。但是,如果实验是将一个经长期传代培养的小鼠细胞株  $B_{82}$  ( $TK^-$ ) 和人的二倍体正常细胞进行体细胞融合,可在 HAT 选择性培养基上筛选到杂种细胞。在杂种细胞传代过程中,人的染色体被不断排斥。经过较长时间培养,杂种细胞约丢失 10% 染色体,从而使杂种细胞中的染色体达到并保持一个稳定的数目。经鉴定,在这一杂种细胞内保留了全套小鼠基因组,丢失了几乎所有人的染色体,在众多的杂种细胞克隆中只保留了人的第 17 号染色体。由于这些杂种细胞株 (系) 在 HAT 培养基中生长依赖于培养基中有 TK 酶的存在,因此可推断编码 TK 的基因位于人的第 17 号染色体上。这是人类第一次应用体细胞遗传学技术将某一基因定位于某一特定染色体上的实例 (图 6-23)。

大量实验证明,杂种细胞中一个亲本染色体的被排除是由不同亲本细胞相对生长速率所决定的,而不是亲本细胞的亲缘关系所决定的。如在人-鼠杂种细胞中一般排斥人的染色体,但当用快速分裂永久性的人体细胞和新鲜得到的小鼠细胞进行细胞融合,杂种细胞染色体首先排除小鼠染色体。一般说,染色体的丢失主要是在融合后早期进行,然后就趋向稳定。如人-鼠杂种细胞进行约 30 轮有丝分裂之后,杂种细胞内基本含有完整的小鼠染色体组,而人的染色体在不同杂种细胞系中含 1~7 条不等,而且所含的人染色体类型也不完全相同。显然人的各条染色体是以随机方式丢失的,因此也可得到多种不同染色体组合的杂种细胞株 (表 6-8)。通过分析某一基因产物与某一人的染色体是否存在而进行基因定位。这一方法的应用,改变了人类遗传学过去传统的依靠家系调查进行基因定位的方法,大大地加速了基因定位研究的进展。

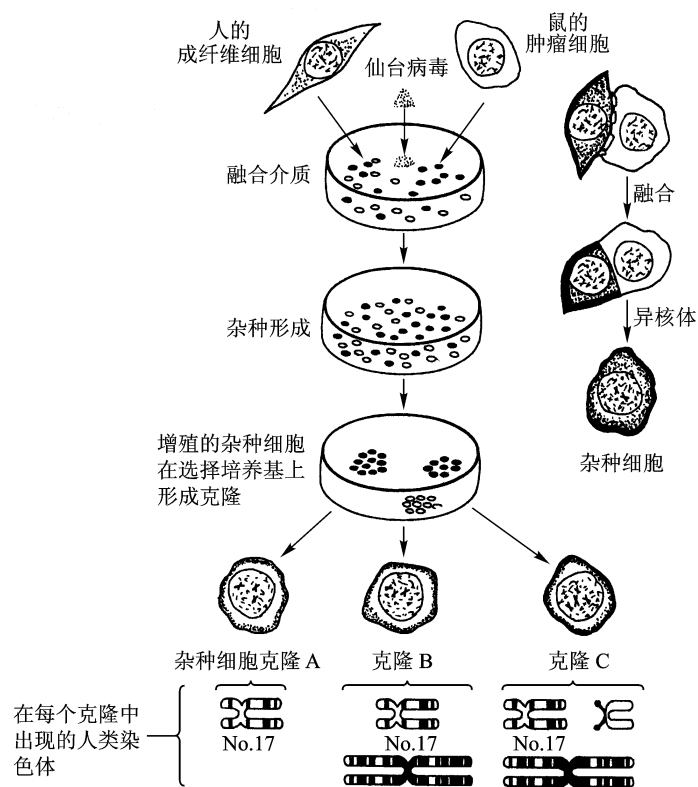


图 6-23 体细胞融合与基因定位

表 6-8 杂种细胞中标记基因的存在与人染色体间的关系

	人-鼠杂种细胞株编码						
	A	B	C	D	E	F	G
所出现的人染色体号码:							
1	+	-	-	-	+	-	-
2	-	-	+	-	-	-	+
3	+	-	+	-	-	+	-
4	-	+	-	-	-	+	-
5	-	-	-	+	-	-	+
人类基因编码的酶							
葡萄糖磷酸变位酶-1	+	-	-	-	+	-	-
异柠檬酸脱氢酶	-	-	+	-	-	-	+
肽酶 C	+	-	-	-	+	-	-
苹果酸氧化还原酶	-	-	+	-	-	-	+
磷酸葡萄糖酸脱氢酶	+	-	-	-	+	-	-
核苷磷酸化酶	-	-	-	-	-	-	-

尽管灭活的仙台病毒或聚乙二醇都能促进细胞融合,然而杂种细胞形成的频率仍然很低。为此必须提供鉴定和分离杂种细胞的有效方法。常用的方法有:

#### (1) HAT 选择系统

采用遗传互补 (genetic complementation) 的方法,将两类分别为  $\text{HGPRT}^-$  或  $\text{TK}^-$  的细胞混合培养后,在 HAT 培养基上进行选择。凡能够在 HAT 培养基上生长的细胞就是经过融合后两个突变基因得到互补的杂种细胞 ( $\text{HGPRT}^+ \text{TK}^+$ )。在进行细胞融合时,如果只有一个亲本具有选择标记,分离杂种细胞可采用半选择系统 (half selective system)。在人-鼠杂种细胞的选择中,最常采用的是将小鼠

突变型细胞 HGPRT<sup>-</sup> 和正常人体成纤维细胞或淋巴细胞进行融合,在 HAT 选择培养基上存活的细胞包括杂种细胞和未融合的人体细胞,虽然人体亲代细胞不具有专一性选择标记,但可采用两种方法除去未融合的人体亲本细胞。第一种,如人体亲本细胞是淋巴细胞,可通过更换培养基洗脱的方法去除;第二种,如人体亲本细胞是成纤维细胞,可用乌本苷处理去除。因为一定浓度的乌本苷能专门杀死成纤维细胞,而对杂种细胞没有影响。在 HAT 和乌本苷选择条件下,细胞融合后 2~3 周就能分离到杂种克隆。

### (2) 营养缺陷突变型标记系统

CHO 细胞已产生了许多营养缺陷突变型,已知在人-中国仓鼠细胞突变型杂种细胞中,人染色体也优先丢失。由于这种中国仓鼠细胞突变型属于营养缺陷突变型,因此只要将融合细胞群培养在不含有这一突变型所需成分的培养基上,就可去除未被融合的突变型细胞。在这种条件下选出的杂种细胞不仅可保留能够与 CHO 细胞营养缺陷突变型发生互补的人类特异性染色体,而且还能保留其他人体染色体。利用含多条人染色体的杂种细胞更便于快速、系统地进行基因定位。

## 6.6.2 同线分析

用体细胞杂种进行基因定位有下列两个层次:第一,将某个基因定位于某一条染色体上,这称为染色体定(配)位(chromosomal assignment);第二,将该基因定位于某一染色体的具体位置上,这称为区域定(配)位(regional assignment)。染色体定位的基础是同线分析(synteny analysis)。连锁分析原理用于体细胞杂种染色体分析的方法称为同线分析。连锁分析是观察子代中基因标记性状的分离情况,而同线分析则是注重杂种细胞中整条染色体的分离情况。同线分析的基点是:如果两个基因在一条染色体上,它们总是共同分离的;如果两个基因位于不同的染色体上,它们之间或多或少会发生自由组合。例如有一人-鼠杂种细胞,是由一个具有两个突变基因(A<sup>-</sup>B<sup>-</sup>)的小鼠细胞系和一个具有两个相应的野生型基因(A<sup>+</sup>B<sup>+</sup>)的人体细胞系融合而成。如果 A 和 B 基因在同一条染色体上,其子代克隆或因为存在该条染色体而表现为 A<sup>+</sup>B<sup>+</sup>,或因为丢失这条染色体而表现为 A<sup>-</sup>B<sup>-</sup>。由于杂种细胞是无性生殖的,不会发生交换而产生 A<sup>+</sup>B<sup>-</sup> 或 A<sup>-</sup>B<sup>+</sup> 这两种重组类型[图 6-24(a)]。相反,如 A 或 B 位于人的不同染色体上,它们的子代克隆中,有的克隆中只含两条人体染色体中的一条,或表现为 A<sup>+</sup>B<sup>-</sup>,或表现为 A<sup>-</sup>B<sup>+</sup>。因此在这种情况下,子代克隆会出现 A<sup>+</sup>B<sup>+</sup>、A<sup>+</sup>B<sup>-</sup>、A<sup>-</sup>B<sup>+</sup> 和 A<sup>-</sup>B<sup>-</sup> 4 种类型[图 6-24(b)]。用这一方法就可以判断两个基因是否在同一条染色体上,也就是说它们是否为同线的。

根据同线分析原理,利用人体染色体和标记基因产物是否同时存在于某个人-鼠杂种细胞株系进行染色体定位(表 6-8)。如果两个基因产物和某条染色体的存在有平行关系,表明它们具有同线性而位于同一染色体上。如表 6-8 中杂种细胞株的葡萄糖磷酸变位酶-1、肽酶 C 和磷酸葡萄糖酸脱氢酶都与人的 1 号染色体存在平行关系,它们应是同线的,就可定位在 1 号染色体上。此外,异柠檬酸脱氢酶与苹果酸氧化还原酶则与 2 号染色体平行,可定位于人的 2 号染色体上。

通过检测基因产物进行基因的染色体定位方法特别适用于具有调节功能的基因以及必须通过诱导才能产生活性的基因的定位。但是应用这一方法除了要求两种细胞融合后,其中一种细胞染色体要优先丢失,而又不能全丢失,对特殊的染色体的识别做到有效和精确以外,还必须具备几个基本条件,如明确被检基因产物的理化性质或免疫学性质,被检基因产物在人和鼠之间可以明显区分,被检基因产物在杂种细胞中能达到可供检测的量等。

基因的染色体定位还可根据选择性标记互补原理构建含有单条人染色体杂种细胞系,如将 CHO-R1 营养缺陷突变型细胞和人体细胞融合,然后通过不断在选择性培养液中筛选,凡能在选择性培养液中存活的杂种细胞,必然保留着带有小鼠营养缺陷互补基因的人染色体。这样既可筛选到在选择性培养液中存活,又只含有一条人染色体的杂种细胞,也可知互补基因就位于这条人的染色体上。应用此法已将多个基因定位于相应的染色体上。



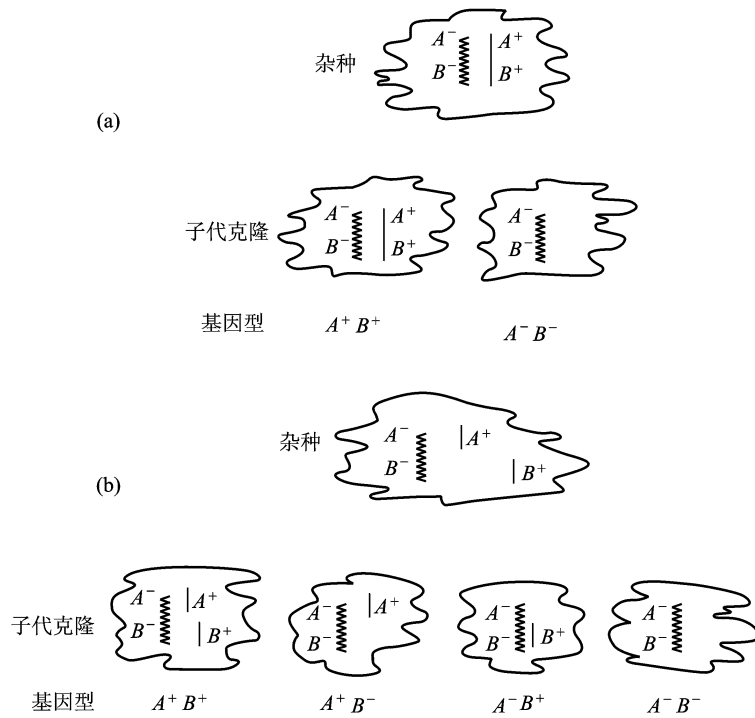


图 6-24 同线分析原理

(a) 两个基因位于人的同一条染色体上 (b) 两个基因位于人的不同染色体上

将某一基因定位于某一染色体之后,接下来的工作就是确定这一基因在染色体上的具体位置,即区域定位。一个基因在染色体上只占一个很小区域,因此很难通过一次定位就确定它的精确位置。一般是将有关这一基因区域定位的资料汇集起来,得出一个定位一致的区域,称为最小重叠区(smallest region of overlap, SRO)。SRO 是某一基因较为精确的位置, SRO 的范围还会随着资料掌握的程度而进一步缩小,从而使区域定位更加精确。

在杂种细胞中可根据易位染色体长度、带型的改变来进行区域定位,如人类从 Xq15处断裂后的一个大片段易位到人 14号染色体上,具有这种 X 染色体的杂种细胞由于这个片段的丢失,同时停止表达人的 4个遗传标记,即 HGPRT(HGPRT 酶)、PGK(磷酸甘油酸激酶, PGK)、G6PD(葡萄糖-6-磷酸脱氢酶, G6PD)和 NP(核苷磷酸化酶, NP)。根据染色体定位已知前 3个标记是位于 X 染色体上,而 NP 位于 14号染色体上。通过易位的遗传分析表明这 3个 X 连锁标记进一步定位于 Xq15的远端。在此基础上还可利用小片段易位将这些基因更精确地定位。又如在某杂种细胞中失去 Xq24远端,结果有 PGK 基因表达产物,而失去了 HGPRT 和 G6PD 标记,由此推断 PGK 基因位于 Xq15~q24 而 HGPRT 和 G6PD 位于 Xq24~末端。

以上的染色体定位和区域定位方法都是依赖于检测细胞内的基因产物,但是许多基因,如调节基因、某些疾病基因等并没有相应的基因产物。在这种情况下,用该基因作为探针,可通过 Southern 印迹杂交法加以鉴定,近来更多地用 PCR 方法来识别人类基因的存在。该技术的应用在 20 世纪末对人类遗传学有着深刻的影响。同样的程序也应用于小鼠的基因定位,值得注意的是小鼠和仓鼠杂合体细胞是小鼠的染色体丢失。

种间杂交细胞中,产生人类染色体片段的方法还有染色体介导的基因转移(chromosome mediated gene transfer, CMGT),是将分离到的中期染色体转入细胞中,在此过程中染色体被破坏,只有少数片段整合到杂交细胞中。也可制备携带单个人类染色体的杂合体细胞,并分离这条染色体的某些片段已丢失的克隆,在这个过程中丢失的基因也必须定位到缺失的染色体区域。

体细胞融合已证明是绘制人类和其他一些物种基因图的一种重要方法。然而必须注意这种技术只适用于在培养细胞中具有基因产物或具备分子探针的基因,如一个 cDNA 克隆或 PCR 所需的足够的 DNA 序列。如果一个与遗传综合征相关的基因,其基本生化缺陷尚未确定,则不能运用此方法绘制基因图。

## 6.7 真核生物基因的删除与扩增及重排

### 6.7.1 基因删除

某些原生动动物,如线虫、昆虫和甲壳类动物等在个体发育中,许多体细胞常常丢掉整条或部分的染色体,只有将要分化形成生殖细胞的那些细胞一直保留全部染色体。通过丢失染色体,丢失某些基因而删除这些基因的活性的现象称为基因删除(gene elimination)。如马蛔虫(*Ascaris megabcephala*)受精卵细胞内只有一对染色体( $2n=2$ ),这对染色体上具有若干个着丝粒,在受精卵发育为成体的早期阶段,只有其中一个着丝粒行使功能,保证了有丝分裂的正常进行。当个体发育到一定阶段后,在将要分化形成体细胞的那些细胞中,这对染色体破碎,形成许多小的染色体,而其中一些小染色体并不含着丝粒,因而在以后的细胞分裂过程中丢失,但是在那些将要分化形成生殖细胞的细胞中则不存在染色体破碎现象。又如在小麦瘦蚊(*Mayetiole destructo*)的个体发育过程中,由于瘦蚊卵的后端含有一种特殊细胞质,称为极细胞质。在极细胞质中的核保留了全部 40 条染色体,但位于其他细胞质区域的核丢失了 32 条染色体,只保留了 8 条,这种现象称为染色体消减(chromosomal elimination)。有 40 条染色体的细胞就分化为生殖细胞,而只有 8 条染色体的细胞继续增殖成为体细胞。如用尼龙线结扎卵,使细胞核不能向极细胞质移动,或用紫外线照射极细胞质,那么所有的核都丢失 32 条染色体,其结果体内没有生殖细胞而发育成为不育的瘦蚊(图 6-25)。

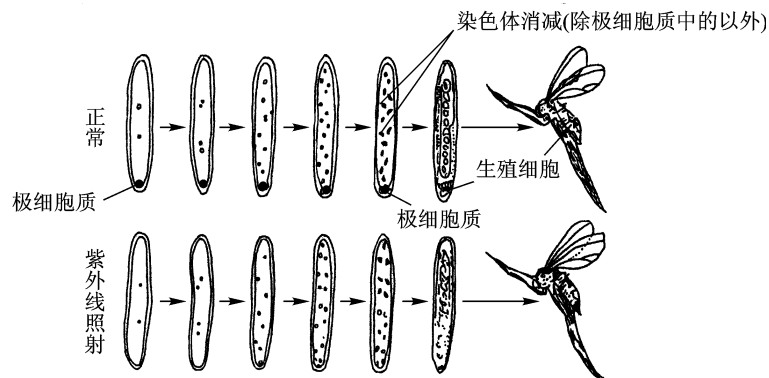


图 6-25 小麦瘦蚊的染色体消减

正常情况下,在第 3~4 次分裂之间,两个核向极细胞质移动,不久成为生殖细胞,其余的核放弃部分染色体变成体细胞。如用紫外线照射极细胞质,所有核都放弃部分染色体发育成不育的瘦蚊

极细胞质中何种物质可控制染色体丢失,从而使生殖细胞的分化成为可能?原来在极细胞质中有许多颗粒,而这些颗粒中含有丰富的 RNA,实际上它是核糖体颗粒的集合体。用离心的方法使这种颗粒移动,让极细胞质以外的核也接触到这种颗粒,结果这些细胞核的染色体就不再发生丢失。此外,紫外线照射可使极细胞质中这类颗粒性物质失活,使染色体消减而无法分化为生殖细胞。说明调

控染色体丢失的主要物质是核酸。

在高等植物中,目前尚未发现类似的基因丢失现象。因为高等植物各种组织和器官的细胞都具有脱分化和再分化而发育成完整个体的潜在能力,这称为细胞分化的全能性(totipotency)。说明这些高度分化的细胞保存了个体发育的全部基因,并未发生基因删除。

### 6.7.2 基因扩增

基因扩增(gene amplification)是指基因组内某些基因的拷贝数专一地大量增加的现象。它是细胞在短期内为满足发育或生理适应的需要而产生足够多的基因产物的一种调控手段,其实质是通过差别的基因复制(differential gene replication)而完成的。如两栖类和昆虫的卵母细胞 rRNA 基因( $\text{rDNA}$ )的扩增,由于卵母细胞成长过程中需要合成大量的蛋白质以满足它在受精后发育的需要,这就需要卵母细胞中积累大量的核糖体,为此首先必须合成大量的 rRNA。爪蟾(*Xenopus*)的卵母细胞便是通过  $\text{rDNA}$  的扩增而大量合成 rRNA 的。爪蟾体细胞中  $\text{rDNA}$  的拷贝数只有 500 个,而在卵母细胞中的拷贝数为约 2 000 000 个,两者相比整整增加了 4 000 倍,这些  $\text{rDNA}$  约占整个卵母细胞 DNA 的 75%,以供转录合成卵裂期所需要的  $10^{12}$  个核糖体。在扩增前爪蟾的  $\text{rDNA}$  区包含 500 个首尾相联的  $\text{rDNA}$  单位,每一单位中包含 18 S rRNA 基因、5 S rRNA 基因和 28 S rRNA 基因,它们成簇存在,重复串联在一起形成核仁组织区(nucleolar organized)。不同的真核生物中这一基因簇的重复单位的排列基本类似。但 rRNA 基因簇的串联重复单位的长度、转录区的长度和非转录间隔区长度在不同生物中均有所不同(表 6-9)。

表 6-9 rRNA 基因簇串联重复单位的不同长度

单位: bp

生物	重复单位长度	非转录间隔区长度	转录区长度
酿酒酵母	8 950	1 750	7 200
黑腹果蝇	11 500~14 200	3 750~6 450	7 750
非洲爪蟾	10 500~13 500	2 300~5 300	7 875
小鼠	4 400	30 000	13 400

高等真核生物中有 4 种 rRNA,其中 18 S、5 S 和 28 S rRNA 为主体 rRNA,它们的基因组成重复单位;而 5 S rRNA 基因与主体 rRNA 基因分隔开,独立成为串联重复单位,每个重复单位由 5 S rRNA 基因和转录区之前的非转录区组成。重复单位拷贝数差异很大,如非洲爪蟾有卵母细胞和体细胞两种类型,两种 5 S rRNA 有 6 个碱基差异,约有 20 000 个 5 S rRNA 基因拷贝,几乎所有染色体上都有 5 S rRNA 基因。这么大的数目可能是为了与主体 rRNA 基因扩增机制相适应,因为 5 S rRNA 基因没有基因扩增现象。

主体 rRNA 基因在卵母细胞中扩增后新形成的  $\text{rDNA}$  并不完全是以一条长链  $\text{rDNA}$  串联重复形式出现,而是以独立的染色体外分子形式出现,形成一个个环状  $\text{rDNA}$  小分子,每个环状  $\text{rDNA}$  分子可能来自基因组上的一个  $\text{rDNA}$  拷贝并含间隔区在内。 $\text{rDNA}$  扩增就以此环状  $\text{rDNA}$  重复单位作为模板,以滚环方式进行复制。扩增后的  $\text{rDNA}$  虽然不在染色体上,但仍包含在核仁中,每一个核仁中包含一个复制单位,因此卵细胞核中的核仁数往往增加到几百个。当卵细胞成熟后, $\text{rDNA}$  便失去了继续存在的意义,因而非但扩增停止,而且所扩增的  $\text{rDNA}$  也开始逐渐降解。当受精卵分裂至数百个细胞后,这类染色体外的  $\text{rDNA}$  分子便完全消失了。

由于发育过程的需要而出现基因扩增现象。在一些昆虫、两栖动物、鱼类的卵母细胞发育中和原生动物的核形成过程中都观察到  $\text{rDNA}$  扩增;在果蝇的卵细胞成熟前卵泡细胞中合成大量的卵壳蛋白,在这之前卵壳蛋白基因大量扩增。在鸡胚胎中也曾发现  $\alpha$ -肌动蛋白基因的扩增现象。

除了发育中的基因扩增外,外界环境条件的改变,也会造成基因扩增。在哺乳动物离体培养的细胞系中加入二氢叶酸还原酶(dihydrofolate reductase, DHFR)的抑制剂氨甲蝶呤(methotrexate, MTX)

后,可以使该酶的结构基因 *dhfr* 扩增达到 40~400 个拷贝,于是细胞产生更高的酶活性来增强对氨甲蝶呤的抗性。再将具有抗性的细胞系培养在不含药物的条件下,结果一些细胞株抗性保持不变,称为稳定系(stable lines),而另一些则抗性逐渐消失,称为不稳定系(unstable lines)。用原位杂交实验证明,稳定系无论是否除去氨甲蝶呤,扩增的 *dhfr* 基因仍然存在于染色体上,其同源染色体的 *dhfr* 通常不扩增(图 6-26)。这些扩增区域染色后表现为均匀染色区(homogeneously staining region, HSR),在光学显微镜下可观察到[图 6-27 a]。中国仓鼠的均一染色区是在第 2 号染色体的长臂上,这正是 *dhfr* 基因所在的位置。相反,不稳定系中扩增的 *dhfr* 基因是在染色体外,它们常以双小染色体或双小体(double minute chromosomes, DM 或 double minute, DM)的形式出现[图 6-27 b]。双小染色体可以自主复制,每个 DM 含有 2~4 个 *dhfr* 基因。在有氨甲蝶呤选择压存在时,DM 可自行复制,可在细胞中保持一定数量,只有保持一定数量的 DM 细胞才能成活。在没有选择作用下,由于含有 DM 的细胞生长缓慢,细胞增殖率低,而不含 DM 的细胞生长快而逐渐占主导地位,并很快地占据了整个细胞群体,加之 DM 又无着丝粒,在细胞分裂中也无法保证可均等地分配到子细胞中去,因此也会逐渐丢失。

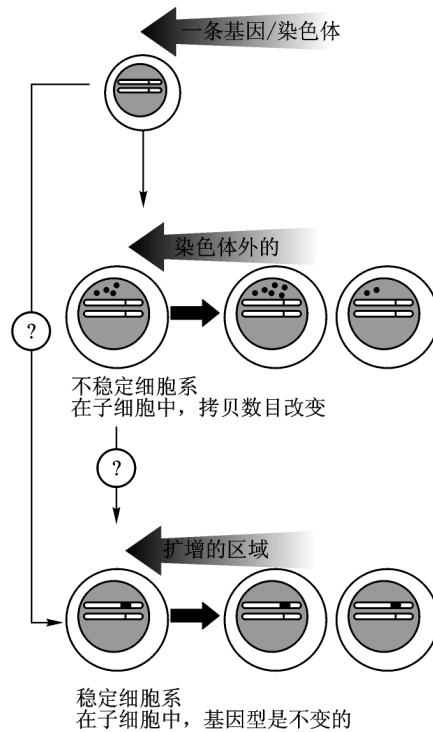


图 6-26 *dhfr* 基因在不稳定系和稳定系细胞中的扩增

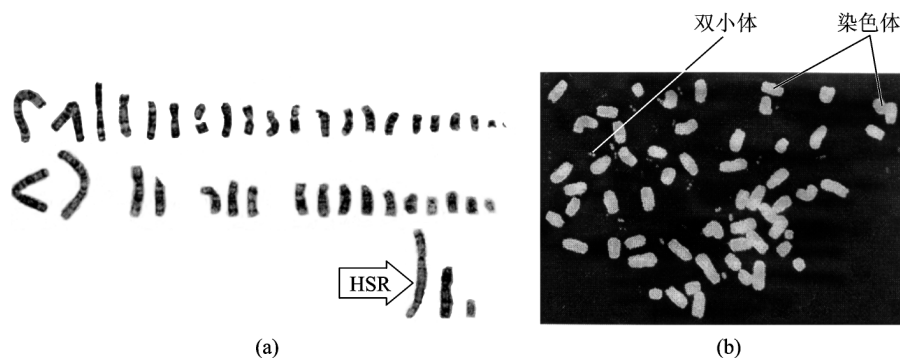


图 6-27 *dhfr* 基因扩增拷贝形成 HSR(a) 或 DM(b)

经 MTX 长期选择后既有抗药性的稳定细胞系,也有不稳定细胞系;但短期选择后则多为不稳定的抗性细胞系。显然染色体外拷贝扩增频率高于染色体内的扩增频率。那么染色体外 *dhfr* 基因拷贝又是如何形成的? 据推测有两种可能性:① 在 *dhfr* 基因附近启动复制后再进行拷贝间的非同源重组。② 通过等位基因间的非同源重组启动这一过程。通过类似的重组机制,这些额外的拷贝就从染色体上释放出来。这一解释已通过相应突变系的研究得到部分证实。*dhfr* 基因大小为 31 kb,然而扩增区在染色体上 HSR 区域平均长度为 500~1 000 kb,而且每个细胞系扩增区范围是不同的。一个 DM 含 DNA 量为 100~1 000 kb。至于染色体中 *dhfr* 基因拷贝数的增加是独立事件,还是依赖于染色体外扩增后再重组的结果,迄今尚无定论。基因扩增采取的形式受细胞基因型的影响,一些细胞系多产生 DM,另一些细胞系快速呈现染色体 HSR 构象。扩增的类型也依赖于特定的基因座位,如对氨甲酰基转移酶抑制剂的抗性 *cad* 基因(CAD 为 UMP 合成途径中前 3 个酶活性的蛋白质)的扩增总是在染色体内,而且有几个分散的扩增区,常涉及几条染色体。

## 知识窗 6-1

## 扩增基因通过趋异产生不同基因

一个基因被扩增时,在一份拷贝中可积聚突变而未遭自然选择的淘汰,那么这份拷贝将可形成新的功能;或者它的表达出现时空差异;或是获得不同活性的基因。而基因扩增的速率大约是每百万年内一个基因扩增一次的概率为 1%,基因扩增后通过趋异积聚突变的速率为每百万年约 0.1% [图 1 (a)]。珠蛋白基因簇就是通过扩增和趋异形成的,所有的珠蛋白基因都是从一个具有 3 个外显子的祖先基因通过扩增和突变而产生的[图 1 (b)]。

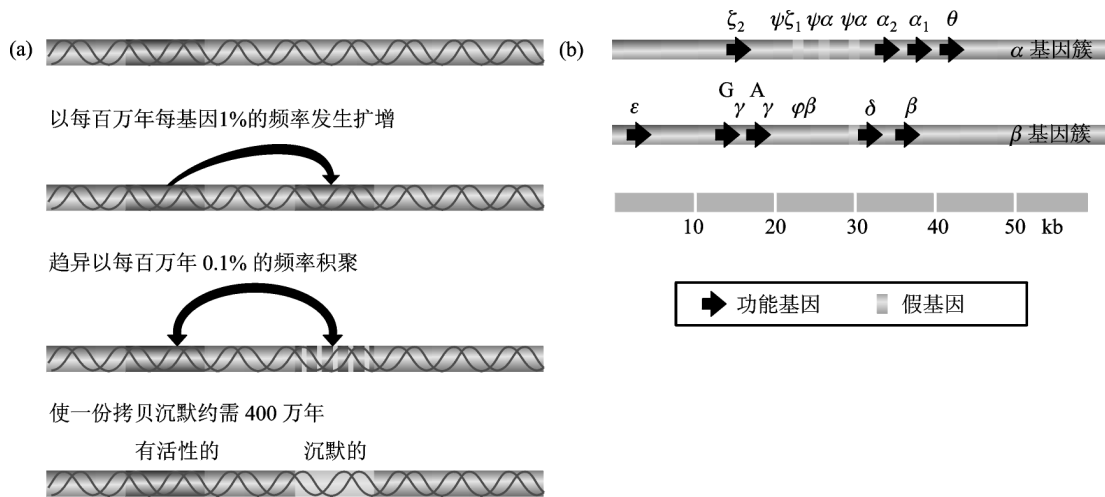


图 1 基因扩增与趋异形成

(a) 扩增的基因可趋异或沉默 (b) 通过扩增和趋异形成珠蛋白基因簇

## 6.7.3 基因重排

基因重排 (gene rearrangement) 是指 DNA 分子的核苷酸序列的重新排列,序列的重排不仅可形成新的基因,还可调节基因的表达。酿酒酵母 (*Saccharomyces cerevisiae*) 中接合型 (mating type) 的决定属于基因重排,单倍体酵母有 a 和  $\alpha$  两种接合型,分别由 *Mat a* (mating 的缩写) 和 *Mat  $\alpha$*  两个基因所控制。一个单倍体细胞不是 a 型就是  $\alpha$  型。两个不同接合型细胞可以接合,而相同接合型细胞不能接合,这是因为产生了相互抑制物质。对不同接合型细胞的识别是根据分泌物质的不同,*Mat a* 细胞产生 a 因子,*Mat  $\alpha$*  细胞产生  $\alpha$  因子,它们是促进不同接合型细胞相互接合的物质基础。接合后的二倍体细胞基因型为 a/ $\alpha$ ,此杂合二倍体可产生 *Mat a* 和 *Mat  $\alpha$*  (图 6-28)。但是 a 型可以转变成  $\alpha$  型, $\alpha$  型也可以转变成 a 型。这种相互转化的现象称为接合型互变 (mating type interconversion)。接合型基因 *Mat* 位于酵母细胞的第 3 号染色体上,而接合型互变又依赖于另一染色体上的一个显性基因 *HO* (homothallism, *HO*)。如果 *HO* 基因突变成为 *ho*,互变频率下降到  $10^{-6}$ 。由于 *HO* 基因的存在使酵母菌居群基因型发生变化,不管起始为哪种接合型,经几代后出现两种接合型细胞,并导致大量二倍体 *Mat a*/*Mat  $\alpha$*  形成,于是由单倍体居群转变为二倍体居群。由此也可判断 *HO* 基因对接合型转变的作用。

一个单倍体细胞同时存在着 *Mat a* 和 *Mat  $\alpha$*  的遗传信息,但在特定时刻只有其中一个表达。当

MAT基因发生突变而使细胞失去接合能力的情况下,另一基因 MAR<sub>1</sub>(mating type regulation)或 MAR<sub>2</sub>发生突变成为 mar 或 mar 时接合能力重新出现。重新出现的接合功能并不在原来的 MAT位置上,而是在 MAT基因的两侧座位上。位于左侧的称为 HML<sub>a</sub>,它是使细胞由 α型转变为 a型所必需的;位于右侧的称为 HMR<sub>α</sub>,它是 a型转变为 α型所必需的。HML<sub>a</sub>距离 MAT为 180 kb, HMR<sub>α</sub>距离 MAT为 150 kb。此外, MAT两边还有一些其他基因: HML his leu MAT thr HMR。现有一种盒子模型(cassette model)来解释接合型互变的机制。

该模型认为在 MAT位置上存在着活跃盒(active cassettes),原来位置上的 HML<sub>a</sub>和 HMR<sub>α</sub>基因都不表达,称为沉默盒(silent cassettes),当它们转移到 MAT活跃盒时便表达,如 HML<sub>a</sub>转移到 MAT座位上后细胞便呈 a型,但是当 HMR<sub>α</sub>转移到了 MAT座位而取代了 HML<sub>a</sub>后,细胞便由 a型转变为 α型。MAT位置出现 HML<sub>a</sub>或 HMR<sub>α</sub>后,原来位置上的 HML<sub>a</sub>或 HMR<sub>α</sub>并不消失,仍然保留一个拷贝的沉默盒。这就是说在取代的过程中必然伴随着 DNA的复制,该机制有些类似于转座行为,但又不完全相同。这一取代只发生在特定的位点,并不涉及在 DNA中插入一段序列,在 HML<sub>a</sub>或 HMR<sub>α</sub>两端未发现反向重复序列,而且接合型的转换是有方向性的,受体为 MAT一个,供体有 HML<sub>a</sub>和 HMR<sub>α</sub>两个。由此认为这种取代是属于位点专一性重组(图 6-29)。

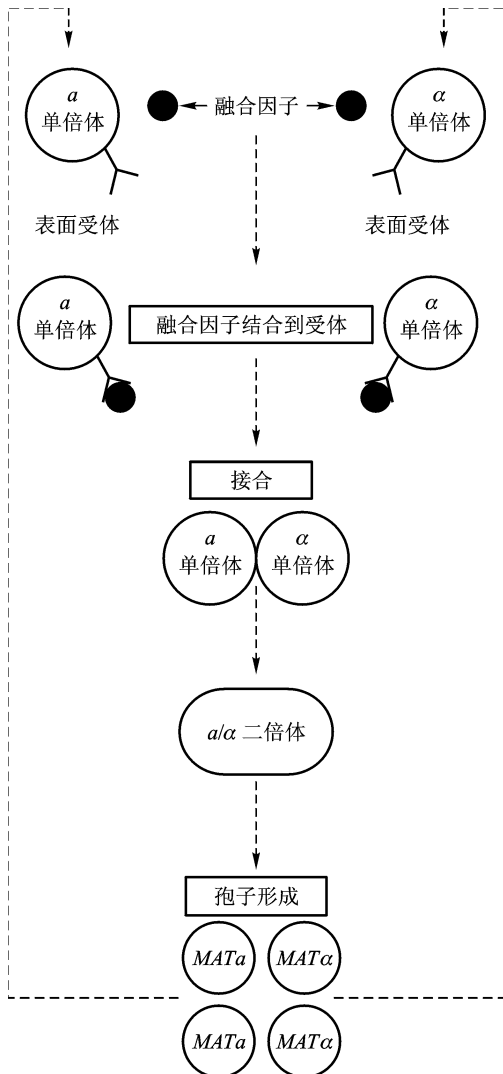


图 6-28 酵母产生单倍体孢子的生活周期

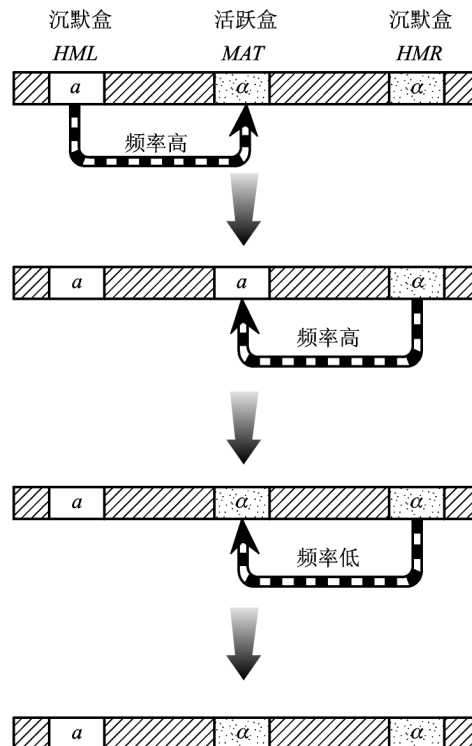


图 6-29 酵母接合型转变

表 6-10 酵母 4 种与接合型相关盒序列组成及长度

单位: bp

盒	W	X	Y	Z <sub>1</sub>	Z <sub>2</sub>	总长
HML <sub>a</sub>	723	704	747	239	88	2 501
MAT <sub>a</sub>	723	704	747	239	88	2 501
MAT <sub>α</sub>	723	704	642	239	88	2 396
HMR <sub>α</sub>		704	642	239		1 585

将两个沉默盒(HML<sub>a</sub>和 HMR<sub>α</sub>)与两个活跃盒(MAT<sub>a</sub>和 MAT<sub>α</sub>)的序列加以比较分析,结果表明,HML<sub>a</sub>、MAT<sub>a</sub>、MAT<sub>α</sub>均由 W、X、Y、Z<sub>1</sub>、Z<sub>2</sub> 5部分组成,而 HMR<sub>α</sub>仅由 X、Y、Z<sub>1</sub> 3部分组成(表 6-10)。Y为中心区,两个活跃盒的差异就在 Y区,因此基因产物也不同。MAT<sub>a</sub>编码 a<sub>1</sub>和 a<sub>2</sub>两种调节蛋白;MAT<sub>α</sub>仅编码 α<sub>1</sub>一种调节蛋白。同时在 Y区的左端,4种盒都有序列完全相同的启动子,但只有两个活跃盒表达,而且有相同的 Y区的沉默盒都不能表达(图 6-30)。这就意味着沉默盒的表达调节不是直接依赖于识别启动子位点来完成的。事实上接合型的建立和转换功能的实现涉及一系列基因,不仅包括控制接合型的基因(MAT)、控制接合型转换的基因(HO)和抑制HO表达的基因(SIR<sup>1-5</sup>),还有阻遏沉默盒基因表达的阻遏蛋白4个亚基的基因 sir1~4 silent information repressor, sir<sup>1-4</sup>。通过缺失突变分析表明,在两个沉默盒启动子上游 1.5 kb有

抑制表达的靶位点,位于 HML上游的靶位点称为 E<sub>L</sub>,位于 HMR上游的称为 E<sub>R</sub>。E位点的行为类似负增强子(negative enhancers),可在离启动子 2.5 kb远距离行使功能,且无方向性。有时称它为沉默子(silence)。而活跃盒上游无 E位点。阻遏蛋白(SIR)的结合位点正是 E位点,它们可能是通过改变染色质结构而阻止基因表达的。此外,两个沉默盒又缺乏 DNA酶 I超敏感点,而活跃盒的两个转录单位上游各含有一个超敏感点。这也可能是造成活跃盒和沉默盒表达活性差异的原因。

接合型转换说明发生了基因转变,即受体位点(MAT)转变成供体类型(HML或 HMR)的序列,而且此过程是单向的。通过点突变已鉴定出在 MAT的 Y右侧一位点是结合型转换的必需位点,Y-Z边界是一个 DNA改变位点,它标志着转换事件的起始位点,在经历转换的细胞群体中,MAT座位有 1%~3% DNA在此位置上有双链切口,此切口靠近边界且与 DNase超敏感位点一致。已知 HO基因编码一种内切酶,该酶作用于 MAT盒上 Y区的右侧,其识别序列为 TTTTCAGCTTTCCGCCAACAAGTATA。切开双链缺口,分裂产生一个 4 bp的单链末端[图 6-31(a)]。正是这种双链断裂起始了 MAT序列的转变过程[图 6-31(b)]。由此可见酵母 MAT序列的转换是一种具有多方面特性的重组过程。

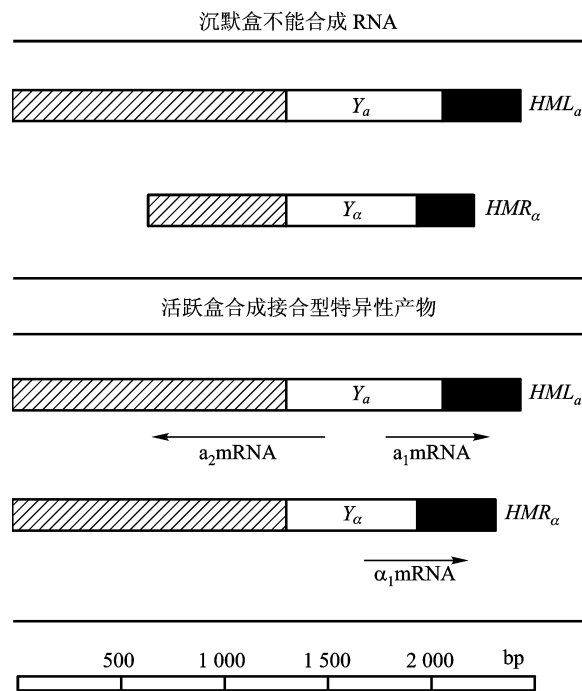


图 6-30 a与α之间 Y区的差异

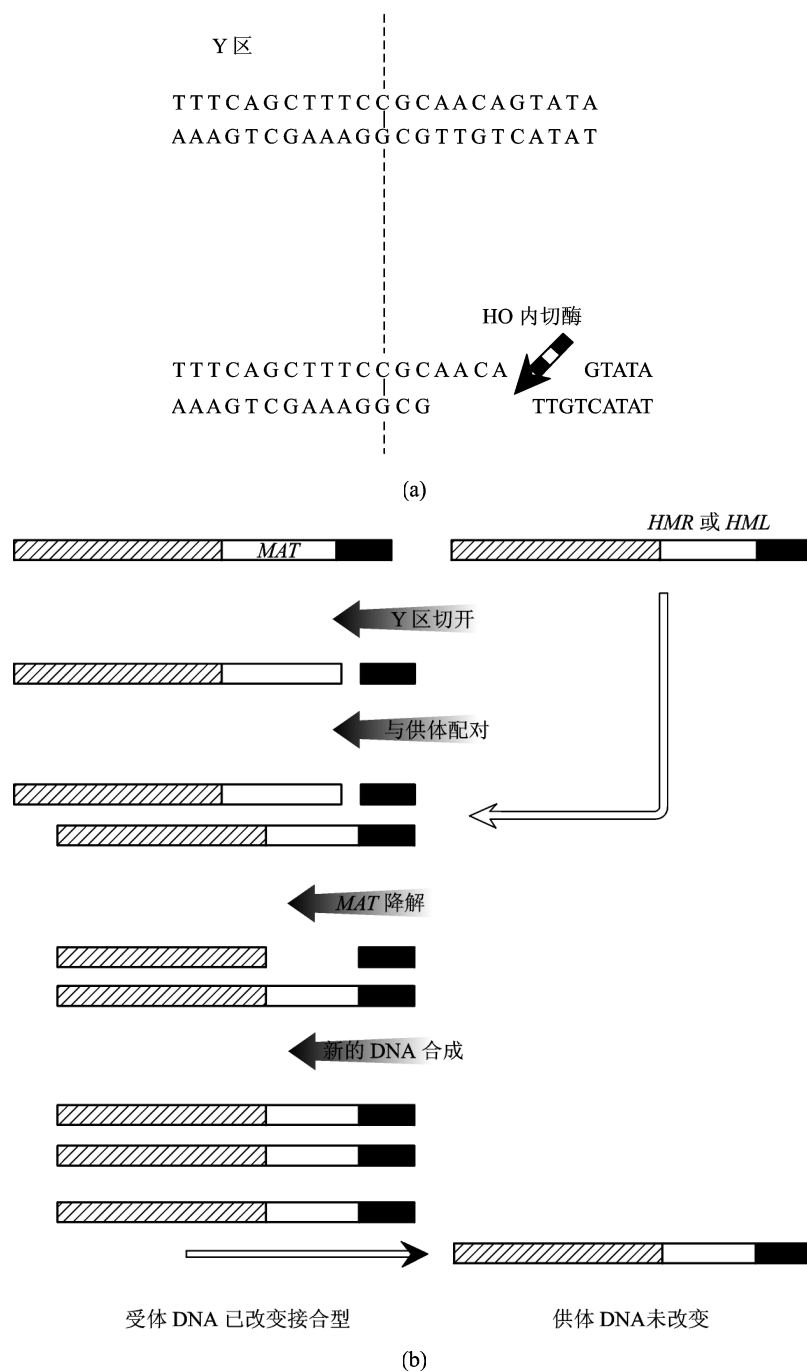


图 6-31 在受体 MAT 双链断裂起始接合型转变  
(a) HO 内切酶酶切 Y 区右侧 (b) MAT 序列转变过程

❓ 思考题

1. 链孢霉  $nic^+ \times + ade$  的 7 种基本子囊型和相应子囊数是:

+ ade	+ ade	$nic^+$	$nic^+$	808
++	++	$nic^+ ade$	$nic^+ ade$	1
++	+ ade	$nic^+$	$nic^+ ade$	90



续表

+ ade	nic ade	+ +	nic +	5
+ ade	nic +	+ ade	nic +	90
+ +	nic ade	+ +	nic ade	1
+ +	nic ade	+ ade	nic +	5

试对 *nic* 和 *ade* 作图。

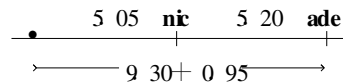
2. 根据上题计算的重组值, 得到  $\bullet$ -*nic* 5.05%, *nic*-*ade* 5.20%,  $\bullet$ -*ade* 9.30%。

为什么  $5.05\% + 5.20\% > 9.30\%$  呢? 因为  $\bullet$ -*ade* 间的重组值低估了。我们看下表:

子囊型	每一子囊被计算为重组子的染色单体数			子囊数	在所有子囊中被计算为重组子的染色单体数		
	$\bullet$ - <i>nic</i>	<i>nic</i> - <i>ade</i>	$\bullet$ - <i>ade</i>		$\bullet$ - <i>nic</i>	<i>nic</i> - <i>ade</i>	$\bullet$ - <i>ade</i>
2	0	4	0	1	0	4	0
3	0	2	2	90	0	180	180
4				5			
5				90			
6				1			
7				5			

$$202 + 208 \neq 372$$

这儿,  $(202 + 208 - 372) / 4000 = 0.95\%$ , 是低估的重组值。把这低估的数值加上, 就完全符合了。



请在表中空白处填上数字。

3. 重组极性杂合 DNA 模型中的异常分离现象最早是在酵母不同交配型  $A \times a$  的杂交中发现的。合子减数分裂产生的 4 个子囊孢子除了正常的  $2A: 2a$  分离外, 出现了  $3A: 1a$  或  $1A: 3a$  的分离, 试用某种 DNA 重组模型加以说明, 并附图解。

4. 子囊菌纲的一种真菌 *Ascochalis* 某座位上的一些突变产生浅色子囊孢子, 称为 *a* 突变体。不同 *a* 突变体进行了以下的杂交, 观察具有黑色野生型子囊孢子的子囊, 杂交中出现的黑色孢子的基因型如下:

$a \times a$	$a_1 \times a_2$	$a_2 \times a_3$
↓	↓	↓
$a+$	$a_1+$	$a_2+$
$++$	$a_1+$	$a_2+$
$+a$	$++$	$++$
$+a$	$+a_2$	$+a_3$

试说明这些结果。  $a_1$ ,  $a_2$  和  $a_3$  这 3 个突变位点的可能次序如何?

5. 下表表示人-鼠杂种无性系以及它们所包含的人类染色体。检验了 4 种人的酶: **TK** (**thymidine kinase** 胸腺激酶), 只在无性系 **C** 中有活性; **LDH** (**lactate dehydrogenase** 乳酸脱氢酶), 在 **A**, **B**, **C** 3 个无性系中都有活性; **PGK** (**phosphoglycerate kinase**, 磷酸甘油酸激酶), 在无性系 **A** 和 **C** 中有活性; **AHH** (**arylhydrocarbon hydroxylase** 芳烃氢氧化物), 在无性系 **B** 中有活性。你能说出有关人的酶的这些基因所在的染色体吗?

无性系	人的染色体			
	X	2	11	17
A	+	-	+	-
B	-	+	+	-
C	+	-	+	+

6. 用体细胞杂交技术研究杂种克隆中染色体与人类 4种酶的关系,得下列结果。这 3个染色体上的基因与 4种酶有什么关系?

		杂种克隆				
		A	B	C	D	E
4种酶	I	-	+	-	-	-
	II	+	-	-	+	-
	III	+	+	-	-	-
	IV	+	-	-	+	-
人类染色体	1	-	+	-	+	-
	2	+	-	-	+	-
	3	-	-	+	+	+

7. 对你的两个  $\beta$ -珠蛋白基因(每个来自 11号染色体中的一条)的 DNA测序表明你的一个基因有突变。如果仅有这个信息,你是否担心成为遗传病的携带者并遗传给你的孩子? 你还想得到哪些信息来判断你的危险性?

8. 在链孢霉的单倍体中,假设两个 Val合成的基因  $\text{Val}^1$  和  $\text{Val}^2$ , 它们的突变等位基因是  $\text{Val}^1$  和  $\text{Val}^2$  这两个基因连锁,平均两次减数分裂中在它们之间发生一次交换。那么:① 该两基因之间未发生交换的减数分裂的比例是多少? ② 两个基因之间重组率是多少? ③  $\text{Val}^1 \text{Val}^2 \times \text{Val}^1 \text{Val}^2$  杂交后代放在 Val 的平板上,有多少比例的后代能生长? ④  $\text{Val}^1 \text{Val}^2$  品系积聚的中间成分是 B, 而  $\text{Val}^1 \text{Val}^2$  品系积聚的中间成分是 A。  $\text{Val}^1 \text{Val}^2$  品系在有 Val 或 A 的平板上生长,但  $\text{Val}^1 \text{Val}^2$  品系仅在有 Val 平板上生长,不能在有 B 的平板上生长。说明 Val 的合成途径及基因控制的位置。