

8 基因组的组织化和进化

关于基因组的组织化和进化的讨论应该包括 3 个不同的课题。第一个是基因组大小，这在不同生物间变化极大。这种变化是如何形成的？增加和降低基因组大小以产生这种变化的机制是什么？第二个是基因组中的遗传信息。基因组主要是由基因的 D N A 构成的吗？还是主要是由非基因的序列构成的？基因组中有许多重复序列吗？而如果是这样那么重复序列的染色体分布模式是什么？非基因部分有功能吗？或者它只是“废物”？第三个课题涉及基因组的核苷酸组成。基因组的不同区域中在组成上有异质性吗？什么机制使核苷酸在组成方面造成区域性差异？

8.1 C 值

单倍体基因组，象在精子细胞核中的那样，其 D N A 的量称基因组大小 (genome size) 或 C 值 (C value)，这里 C 来自 “constant” (常数) 或 “characteristic” (特征)，表示单倍体基因组的大小在任何物种中都是相当恒定的这样一个事实。反之，C 值在原核生物和真核生物的各物种间则变化幅度极大。

表 8—1 常用于测试基因组大小的单位换算

单位	换算因子		
	微微克	道尔顿	碱基对
微微克	1	6.02×10^{11}	0.98×10^9
道尔顿	1.66×10^{-12}	1	1.62×10^{-3}
碱基对	1.02×10^{-9}	618	1

真核生物中细胞核基因组的大小通常用微微克 (pg) D N A 来度量 ($1\text{pg}=10^{-12}\text{g}$)。较小一些的原核生物的基因组则通常用道尔顿来度量，道尔顿即一个原子或分子的相对质量单位。还有更小的基因组，象细胞器和病毒等的基因组，以及某些特殊的 D N A 延伸物，它们的大小则常用双链 D N A 或 R N A 的碱基对 (b p) 或千碱基对 (k b) 来表示 ($1\text{kb}=1000\text{bp}$)。为避免混乱，我们将只用 b p 和 k b。单位换算因子可见表 8—1。

8.2 细菌的基因组大小的进化

细菌的基因组大小其变化幅度超过了 2 0 倍，从某些限制性细胞间寄生物中的约 $6 \times 10^5 \text{bp}$ 到几种蓝细菌中的 10^7bp 以上 (表 8—2)。最小的游离生活的原核生物，枝原体，含有包括 5 0 种核糖体蛋白在内的 3 5 0 个为蛋白质编码的基因，两组 rRNA 基因 (5 S, 1 6 S 和 2 3 S)，和大约 4 0 个 t R N A 基因。因此，一个枝原体的基因组由大约 4 0 0 个基因所构成，这接近于所猜测足够维持机体自主性生活的最小数目 (Muto 等, 1986)。其他细菌中的基因数大致在 5 0 0 到 8 0 0 0 的范围内变化，换句话说，在特征细菌的种类中，基因数上的变异与 C 值方面的变异相同。因此，细菌看来不会含有大量的非基因的 D N A。

细菌的基因组可被划分成 3 个部分：(1) 染色体 D N A，(2) 源于质粒的 D N A，和 (3) 可转座因子 (Hartl 等, 1986)。染色体部分含有为生长和代谢功能所必需的蛋白质编码基因 (7 0—8 0 %)、间隔序列和各种信号序列 (2 0—3 0 %)、编码 R N A 的基因 ($\sim 1\%$)，以及一些通常长度在几十个碱基对这一级别上的短重复序列。细菌常带有许多作为染色体外遗传成份的质粒，然而，在有些例子中，一些来自质粒的基因却发现被整合在细菌染色体中。可转座因子是某些细菌基因组的普通组分。例如，E. coli 的染色体，至少含有 6 个不同类型的插入序列各 1—1 0 个拷贝 (第七章)。基因组的非染色体

部分（包括染色体中的插入序列和来自质粒的基因）看来要比染色体部分小一个数量级。

表 8—2 细菌中 C 值的范围

分类单位	基因组大小的范围(kb)	比例(最高/最低)
真细菌	650—13200	20
革兰氏阴性	650—7800	12
革兰氏阳性	1600—11600	7
蓝细菌	3100—13200	4
枝原体	650—1800	3
古细菌	1600—4100	3

资料自 Cavalier Smith(1985)

细菌中基因组大小的分布是不连续的，在模值约 0.8×10^6 , 1.6×10^6 和 4.0×10^6 b p 处呈现出主峰，而在 7.2×10^6 和 8.0×10^6 bp 处有几个副峰 (Herdman, 1985)。这种分布引出了这样的看法：较大的基因组是从较小的基因组经连续循环的基因组重复进化而来的。由于基因组大小和细菌的系统发育间看来没有显著关系，所以，已经出现了这样一个提议：基因组重复在细菌谱系的进化中是频频出现的 (Wallace 和 Morowitz, 1973)。

应用根据 r R N A 顺序比较而得出的尝试性细菌系统发育，赫德曼 (Herdman, 1985) 曾找出了基因组大小变化与系统发育史的关系。其结果指示，基因组重复在不同细菌谱系中是独立地发生的。有趣的是，许多重复看来曾在进化中的某一相当特别的时刻，即氧在大气层中出现后不久，大约在 1 8 亿年前，同时出现在几种不同的细菌谱系中的。

概括起来，细菌中基因组大小的分布能用几个过程的结合来加以解释：（1）某些谱系中呼吸代谢独立进化期间的基因组重复，（2）独立地出现在许多谱系中的，随后的基因组重复。（3）小规模缺失和插入，（4）复制性转座，（5）主要来自质粒的基因的水平转移，和（6）许多寄生性谱系中大量 D N A 丢失。

8.3 真核生物的基因组大小和 C 值悖论

真核生物中的 C 值通常比原核生物中的要大得多，但也有例外。例如，酵母 *Saccharomyces cerevisiae* 的基因组就比许多革兰氏阳性菌和大多数蓝细菌的小。因为真核生物的基因组有多重的复制原点，而大多数原核生物看来只有一个，所以，在复制出较小原核生物基因组所需的相同的时间里，真核生物复制出的 D N A 的量就要大得多。

真核生物中的 C 值变化要比细菌中的大得多，从 8.8×10^6 bp 到 9.9×10^{11} bp, 范围接近 80000 倍 (表 8-3)。单细胞的原生生物，特别是肉足类变形虫，在 C 值上表现出的变化最大。3 个羊膜类的纲（哺乳类、鸟类和爬行类）则相反，在真核生物中以它们在基因组大小上的变化小（最高相差 4 倍）而显得不同寻常。其他的纲，对那些 C 值数据大体上存在者而言，表现出至少 1 0 0 倍的变异。

有趣的是，真核生物间基因组大小上的巨大种间差异看来与生物体的复杂性和生物体所编码的可能基因数没有什么关系。例如，几种单细胞的原生动物具有比哺乳动物要多得多的 D N A（表 8—4），而后者据推测应要更复杂一些。而且，一些复杂性看来相似的生物（例如，果蝇与蝗虫，洋葱与百合，双核草履虫（*Paramecium aurelia*）与尾草履虫（*P. caudatum*））却展示出极大的 C 值差异（表 8—4）。这种 C 值与基因组中所含的遗传信息的认定数量间缺乏对应的现象，以 C 值悖论或 C 值矛盾 (C value paradox) 而著称于文献之中。在姐妹种（即一些相互在形态学上极为相似，以至表型上不可区分的种）的比较中 C 值矛盾也是明显的。在原生生物、真骨鱼类，两栖类和显花植物中，许多姐妹种，尽管根据姐妹种的定义这些生物间的复杂性没有什么差异，但在它们的 C 值上却差别极大。我们可以假定一个物种不可以具有少于保证其生命功能所必需的 D N A 的量，但我们必须解释为什么有那么多种含有远远过剩的 D N A 的量。

表 8—3 不同真核生物类群中的 C 值范围

分类单位	基因组大小范围(kb)	比例(最高/最低)
原生生物	23,500—686,000,000	29,191
眼虫类	98,000—2,350,000	24
纤毛(亚门)类	23,500—8,620,000	367
肉足(总纲)类	35,300—686,000,000	19,433
真菌	8,800—1,470,000	167
动物	49,000—139,000,000	2,837
海绵	49,000—53,900	1
环节动物	882,000—5,190,000	6
软体动物	421,000—5,290,000	13
甲壳类	686,000—22,100,000	32
昆虫	98,000—7,350,000	75
棘皮动物	529,000—3,230,000	6
无颌类	637,000—2,790,000	4
蜜与苔	1,470,000—15,800,000	11
真骨鱼类	382,000—139,000,000	364
两栖类	931,000—84,300,000	91
爬行类	1,230,00—5,340,000	4
鸟类	1,670,000—2,250,000	1
哺乳类	1,420,000—5,680,000	4
植物	50,000—307,000,000	6,140
藻类	80,000—30,000,000	375
蕨类植物	98,000—307,000,000	3,133
裸子植物	4,120,000—76,900,000	17
被子植物	50,000—125,000,000	2,500

资料自 Cavalier Smith(1985)及其他来源

要搞清楚的第一个问题是，基因组大小和基因的数目间是否存在相关。换言之，基因组大小上的差异应归因于基因的 D N A 还是非基因的 D N A？真核生物在为蛋白质编码的基因数上表现出约有 5 0 倍的变异，从酵母中的约 3000 个到哺乳动物中的 150000 个左右 (Cavalier-Smith, 1985)。这种 5 0 倍的差异显然不足以解释细胞核 D N A 在含量上的 80000 倍的变异。再者，基因数是与结构复杂性正相关的，而基因组大小则不是。m R N A 分子长度上的种间差异也不能解释 C 一值矛盾。在不同生物间编码区和非编码区的平均长度都有着差异的情况下，基因长度和基因组的大袖就不会存在相关。

几种编码 R N A 的基因的重复程度与基因组大袖的正相关业已被发现 (第六章)。类似地，象端粒、着丝粒和复制器基因，这样一些在有丝分裂和减数分裂期间为染色体复制、分离和重组所必需的调控序列，它们的拷贝数与基因组大袖也存在相关性。不过，所有这些基因只构成基因组的一小部分，以至 R N A 基因和调控序列等的在数量上的差异并不能解释基因组大小上的差异。

总之，我们只剩下非基因 D N A 部分作为造成 C 一值矛盾的唯一嫌疑犯。换句话说，真核基因组的主要部分是由不含遗传信息的 D N A 所构成的。曾有估计认为，每基因组非基因 D N A 的量在真核生物中从约 $3.0 \times 10^6 \text{bp}$ 到 $1.0 \times 10^{11} \text{bp}$ 以上 (10 万倍范围) 之间变化，且构成了基因组的量的不到 30% 至几乎 100% (Cavalier-Smith, 1985)。

表 8—4 按基因组大小等级的真核生物的 C 值

物种	C 值(kb)
<i>Navicola pelliculosa</i> (硅藻)	35,000
<i>Drosophila melanogaster</i> (果蝇)	180,000
<i>Paramecium aurelia</i> (纤毛虫)	190,000
<i>Gallus domesticus</i> (鸡)	1,200,000
<i>Erysiphe cichoracearum</i> (真菌)	1,500,000
<i>Cyprinus carpio</i> (鲤)	1,700,000
<i>Lampræta planeri</i> (七鳃鳗)	1,900,000
<i>Boa constrictor</i> (蛇)	2,100,000
<i>Parascaris equorum</i> (蚯蚓)	2,500,000
<i>Carcarias obscurus</i> (鲨)	2,700,000
<i>Rattus norvegicus</i> (大鼠)	2,900,000
<i>Xenopus laevis</i> (蟾)	3,100,000
<i>Homo sapiens</i> (人)	3,400,000
<i>Nicotiana tabaccum</i> (烟草)	3,800,000
<i>Paramecium caudatum</i> (纤毛虫)	8,600,000
<i>Schistocerca gregaria</i> (蝗虫)	9,300,000
<i>Allium cepa</i> (洋葱)	18,000,000
<i>Coscinodiscus asteromphalus</i> (硅藻)	25,000,000
<i>Lilium formosanum</i> (百合)	36,000,000
<i>Amphiuma means</i> (蝾螈)	84,000,000
<i>Pinus resinosa</i> (松)	68,000,000
<i>Protopterus aethiopicus</i> (肺鱼)	140,000,000
<i>Ophioglossum petiolatum</i> (蕨类)	160,000,000
<i>Amoeba proteus</i> (变形虫)	290,000,000
<i>Amoeba dubia</i> (变形虫)	670,000,000

资料自 Cavalier-Smith(1985), Sparrow 等(1972)和其他参考文献

8.4 真核生物基因组的重复结构

真核生物的基因组有两大特征: (1) 序列的重复, 和 (2) 组成分隔化, 分隔成以特别的核苷酸组成而相互区别的特征片段。重复 DNA (repetitive DNA) 是由不同长度和组成的核苷酸序列所构成, 这些序列以串接或散在的形式在基因组中几次出现。不发生重复的 DNA 片段被称为 **单拷贝 DNA** (single copy DNA) 或 **单一 DNA** (unique DNA)。基因组中由重复序列占据的部分在各分类单位间变化极大。在酵母中, 该部分的量约为 20%; 而在哺乳类中, 则高达 60% 的 DNA 是重复性的。在植物中, 该部分可超过 80%, 而且比它更高的值也曾被记录到 (Flavell, 1986)。

布里滕和科恩 (Britten 和 Kohne, 1968) 所做的关于 DNA 双链重新结合的经典研究表明, 较高等的真核生物, 其基因组大致分成 4 部分: **折回 DNA** (foldback DNA), **高度重复 DNA** (highly repetitive DNA), **中度重复 DNA** (middle repetitive DNA) 和 **单拷贝 DNA** (图 8—1)。折回 DNA 由回文 DNA 顺序所构成, 一旦这种变性的 DNA 被允许复性时, 该顺序可以形成发夹式的双链结构。高度重复部分是由长度从几个到成百个核苷酸的短序列所构成, 其重复数平均为 50 万次。中

度重复部分由平均达成百或成千碱基对的长序列所构成，它们在基因组中以成百的次数出现。

根据重复在基因组中散在的模式，重复部分已被发现是由两类重复家族所构成的：区域性重复家族和散在的重复家族。

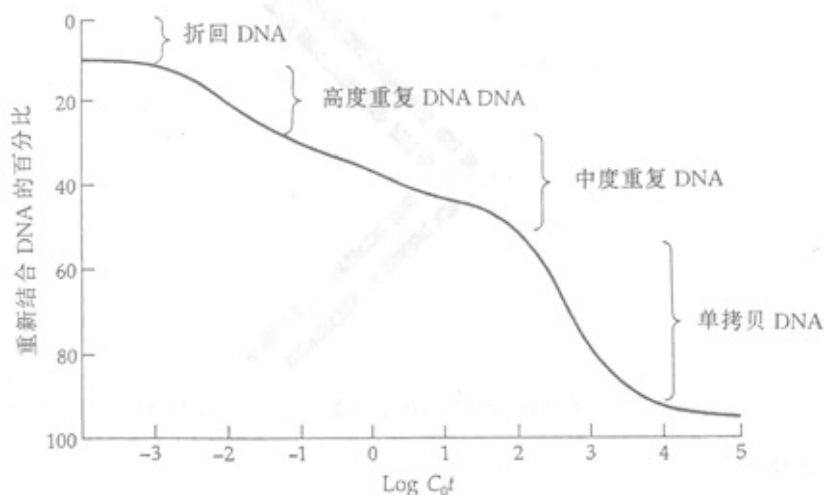


图 8 - 1 哺乳类的 DNA 重新结合图。DNA 被纯化、裁剪，加热熔解成单链，然后通过逐渐冷却而使之重新结合。纵轴上的重新结合的双链 DNA 百分比，显示出是 DNA 的浓度与横轴上的时间 (Cot) 之积的函数。自 Schmid 和 Deiniger (1975) 修改而成。

区域化的重复序列

大多数真核生物的基因组含有串联排列的、高度重复的 DNA 序列。在有些物种中，这些区域化的重复 DNA 序列可构成基因组中 DNA 的主要部分。例如，在袋鼠 *Dipodomys ordii* 中，基因组的 50% 以上由 3 个重复序列构成：AAG (2.4 亿次) TTAGGG (22 亿次) 和 ACACAGCGGG (1.2 亿次) 见 (Widgren 等, 1985)。当然，这些家族并不是完全同质的，而是含有许多变异型，它们与一致的顺序有一个或两个核苷酸差异。例如，TTAGGG 家庭中，有些序列实际上是 TTAGAG。虽然如此，但许多区域化的高度重复序列仍有着相当一致的核苷酸组成，以至根据基因组 DNA 的部分化和密度梯度分离，它们能形成一条或多条粗带。这些带明显地与 DNA 的主带和由其他组成更异质的片段产生的带有区别，被称为 **卫星 DNA** (satellite DNA)。

在有些物种中，所有染色体上都发现有串接排列的高度重复序列，而在另一些物种中它们则被限制在某一特别的染色体位置中。例如，果蝇 *D. nasutooides* 基因组的 60% 由卫星 DNA 构成，而这类 DNA 全部都处在 4 对染色体中的一对上 (图 8 - 2)，而这种染色体上看来几乎不含别的 DNA (Miklos, 1985)。

根据当前可得到的证据，区域化的高度重复序列很有可能是无功能的。而且，区域化的高度重复序列可能既不会降低、也不会提高个体的适合度。因此，这类序列的进化不受自然选择的影响，而主要是由基因转变和不等价交换 (第六章) 所决定的。这些机制将产生两种结果：(1) 序列同质性，和 (2) 拷贝数随进化时间而大幅度波动 (见 Charlesworth 等, 1986)。还曾经有这样的提法：区域化的重复序列的翻转速率是相当高的，即，现存的排列可能会通过不等价交换而被排除，而新的排列则可能会通过 DNA 重复过程而不断地产生 (Walsh, 1987)。

大多数串接重复序列只不过是“废物”DNA 的说法，本质上是说，它们没有表型效应或者它们对它们的携带者的适合度没有影响。这在大多数情况下可能是真实的，但也有证据表明高度重复序列的某种特殊排列并不总是如此。果蝇 *D. melanogaster* 的自然群体中，应答者 Responder (Rsp) 基因座位由 20-2500 个拷贝的富含 AT、长 120 bp 的序列所构成 (Wu 等, 1988)。有一次竞争性实验涉及某一混合群体，该群体由含 700 个重复拷贝的果蝇和只含 20 个拷贝的果蝇构成，实验中观察到只含 20 个重复的果蝇频率随时间推移而降低 (Wu 等, 1989)。因此，已有结论是，含 700 个拷贝的果蝇有比只含 20 个拷贝

的果蝇更高的适合度。目前，R s p 的功能尚不知，但它显然不是“废物”DNA，因为它的缺乏会降低适合度。这些结果能否用于其他串联重复的序列，特别是用于主要卫星DNA家族，也是未知的。

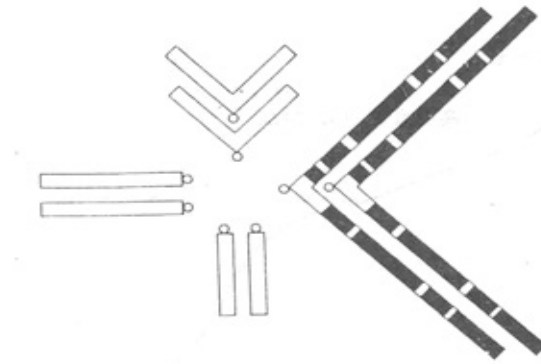


图 8 - 2 果蝇 *D. nasutooides* 中高度重复DNA序列（黑色区）位于一条染色体上。自 Miklos(1985)修改而成。

散在的重复序列

第二类高度重复的DNA由散在于整个基因组中的序列所构成。散在的高度重复序列的拷贝已在内含子、基因的侧翼区域、基因间区域、和非基因DNA中发现。有两大类散在的高度重复序列：短的散在重复序列，缩写成SINE，和长的散在重复序列，或LINE（Singer, 1982）。

SINE中的典型代表为短于500bp，且在单倍体基因组中出现拷贝数在 10^5 或 10^5 以上者。与确定tRNA的基因一样（第一章），许多SINE含有内部转录信号且被RNA多聚酶III转录。大多数SINE是反录序列。人的基因组中最著名的SINE家族是Alu家族（第七章）。

LINE原来曾被描绘成长于5kb、且每基因组以 10^4 或其以上的拷贝数存在的DNA序列（Singer, 1982）。人的基因组中显然只含有一个LINE家族、即L1（Hutchison等, 1989）。这种一致性的L1序列长约6kb，在一端有一个poly-A尾巴，且一般由长度短于200bp的短顺向重复构成侧翼。拷贝数大约为每单倍体基因组10万个。大多数L1序列在它们的5'端处被截断。完整的L1序列含有2个长开放阅读框架，ORF-1和ORF-2，分别有约375和1300个密码子。ORF-2含有的氨基酸序列具有反转录酶的主要特征。以上特点表明，L1因子是通过多聚腺苷化的mRNA的反转录，和该反转录本随后插入基因组中而产生的。由于L1序列不具长末端重复（LTR），所以它们很可能是反转录子（第七章）。大多数L1序列不被转录。不过其中少数则用RNA多聚酶II进行转录。

与人的L1家族同源的LINE已在所有哺乳类，包括有袋哺乳类中发现（见Hutchison等, 1989）。与L1相关的非哺乳类因子包括，果蝇中的I、F、G和D因子，Trypanosoma brucei中的Ingi，家蚕Bombyx mori中的R2，和玉米中的Cin4。物种间的L1序列分歧程度远大于同物种的L1拷贝间的分歧程度。例如，来自小鼠和人类的L1序列相互平均约有30%的差异。相比之下，小鼠中L1因子间的顺序分歧量只有4%（Hutchison等, 1989）。

由于大多数L1序列被截去尾部，所以它们不含有完整的阅读框架。结果，它们也许不能转座。这类有缺陷的因子被发现比完整的因子进化得更迅速。而且，有缺陷L1序列的进化谱系已被发现不含分枝，这指示这些因子不能复制性转座。于是，它们可被看成是反转录子的假基因，对它们而言功能限制已不再起作用。大多数L1序列是有缺陷的，这一事实的含义是，基因组中的L1因子的传播仅依赖于少数的源因子。结果，基因组中的L1因子是高度同质的，且序列翻转的速率非常高。事实上，曾有估计，啮齿类中一半以上的L1因子只有300万年或者更短一些的历史（Hardies等, 1986）。哈奇森等（Hutchison等, 1989）曾建议，SINE和LINE应该重新加以定义。即，不用长度和拷贝数作诊断特征，而把LINE看成是，应包括所有具有为中介它们自己和它们后代的转座的蛋白质编码活性的反转录子和反录转座子。另一方面，SINE则应包括不为这类功能编码的反录序列。

可转录基因的基因组位置

RNA-DNA杂交实验表明,只有很小一部分的可转录基因位于基因组的重复部分中。大多数可转录基因处在单一DNA部分内。即使在单一部分中,大多数序列也是不转录的。事实上,人类中只有3%的非重复DNA序列被转录(见Lewin, 1990)。这些实验进一步地支持真核生物基因组中大多数是不具遗传信息的观点。

8.5 增加基因组大小的机制

为了解释真核生物的基因组中非基因DNA大量存在的现象,我们首先必须探讨可能会导致基因组大小增加的过程。设想出来的机制应该不仅能解释非基因DNA本身的现象,而且能解释它的重复性和特别的基因组分布。

我们把基因组增加分成两种类型:(1)全局性增加,即整个基因组或它的主要部分,如染色体,重复,和(2)局部性增加,即某一具体序列增殖而产生重复DNA。在后一种情况中,我们又分散在的重复序列产生的机制和区域化重复序列产生的机制。

基因组重复

由于真核生物的基因组显著地大于真细菌的基因组,所以,真核生物从原核生物的祖先进化而来的过程必须涉及基因组大小上的增加。有几种分子机制可以导致基因组大小上的增加。造成基因组增长的一个重要因素是基因组重复(genome duplication)或基因组加倍(genome doubling)。基因组重复是在DNA复制之后,由于缺少所有染色单体间的分裂过程而产生的结果。

假定哺乳类基因组比细菌的基因组大1000倍左右,并假定基因组重复是造成基因组增大的唯一原因,则我们可以推出,基因组从原始的细菌大小增加到现在哺乳类中的大小,大约需要10轮基因组重复。从另一个角度看,即基因组重复平均每3亿年出现一次。另一方面,如果DNA含量是以小段DNA连续增加的方式增长,比如说通过转座或不等价交换,则从细菌到哺乳类基因组增长的速率近似为每年7个核苷酸(Nei, 1969)。

基因组大小的多峰态值的分布也已在许多真核生物类群中被记录到(Rees和Jones, 1972; Grime和Mowforth, 1982),这与细菌中的情形类似(见第119-120页)。这在单子叶植物中特别明显,它们的基因组大小在0.60, 1.18, 2.16, 4.51和 8.53×10^9 bp处出现峰值,表现出一种多峰态值的分布(Sparrow和Nauman, 1976)。类似的分布也曾别的类群,如棘皮动物、昆虫和真菌中观察到,以及在两栖类和真骨鱼类中,较小范围地观察到。基因组重复或多倍性重复(polyploidy)看来是真核生物基因组大小进化的主要机制。有趣的是,每一轮基因组重复都牵涉DNA的少量丢失,所以,每轮重复后DNA的量都是以一个略小于2的因子增加(Sparrow和Nauman, 1976)。在新近形成的多倍体中,还谈不上C值方面的增加,因为该值是指单倍体的大小,而这两个基因组经历过突变、易位、染色体重排和染色体数目上的变化后,它们将最终会变成一个基因组,到那时才谈得上C值增加的问题。换言之,一个古老的多倍体与一个二倍体将是不可区别的(Cavalier-Smith, 1985)。

多倍性重复是自然界中的一个普遍现象;然而,在进化的历史中多倍体看来很少能生存下来。在许多情况下,其原因是,多倍性重复是有害的,因而会受到选择的强烈对抗。多倍性重复的有害效应包括:(1)细胞分裂时间大大延长,(2)细胞核的体积增大,(3)在减数分裂的分离期间染单体断裂的数目显著增加,(4)遗传不平衡,和(5)在生物的性别由性染色体和常染色体间的数目比(例如果蝇),或性染色体与倍数性间的数目比(例如膜翅目)来决定的情况下,干扰性别分化。在有些情况下,多倍性重复可能是有利的,象在显花植物中,它可以减少或消除杂种不育性(见Cavalier-Smith, 1985)。

染色体重复

一个染色体的重复,或非整数倍重复,大多数是有害的。在哺乳类中,它常伴随着致死性或不育性。在人类中,众所周知的例子包括象唐氏(Down)综合症(21三体)和第18染色体三体等频频出现的病症。类似的有害症状也会与染色体的部分重复相关联。因此,染色体重复预期对基因组大小上的增加不会有显著贡献。

在许多生物中,特别是在那些被调查得最彻底的植物群体中,已经观察到几个建立了额外染色体,即

所谓 B-染色体 (B-chromosomes) 的例子 (Jones, 1985)。B 染色体不是整个染色体精确地重复, 而可被看成是部分染色体重复的产物。这些额外染色体看来不含太多基因, 而且它们大多数是由短重复序列所构成的 (即它们是高度异染色质的)。因此, 部分非整数倍重复可能在进化期间的基因组大小增长过程中起作用, 但对于基因数的进化或许并无重大贡献。

基因组大小的区域性增加

基因组大小的区域性增加可由转座 (第七章) 和不等价交换 (第一和第六章) 所引起。前一种模式将产生散在的重复序列; 后一种则产生串联的重复序列。基因组大小的区域性增加也可通过获得外来 DNA (第七章) 而实现; 不过, 此过程的贡献与整个 DNA 大小比起来或许可忽略不计。

曾有建议认为, 真核生物中所有中度重复的 DNA 部分都起源于可转座因子。不过, 这些因子中大多数已不再可动, 它们的转座能力已被突变或别的因子的插入摧毁了。例如, 果蝇中的异染色质的大多数, 实际上也许是这类已经死亡了的因子的墓地。不过, 非活性的易动因子能通过不等价交换那样的过程而局部地增殖。

许多区域性重复序列曾经是由不等价交换而产生的。然而, 看起来不等价交换更常做的是打散串联排列、而不是增加它们的大小和拷贝数, 所以, 该过程并不能解释所有区域性重复 DNA 存在的原因 (Walsh, 1987)。再者, 不等价交换通常产生的是由相对而言较长的重复单位构成的序列。相比之下, 许多区域性重复序列, 象卫星 DNA, 则是由非常短的, 简单地重复的主体所构成的 (见第 123 页)。为了解释这类序列存在的原因, DNA 扩增被提了出来。DNA 扩增 (DNA amplification) 是指, 任何一种将某一基因或 DNA 序列的拷贝数增加到远超出一个生物的特征水平之上的事件。狭义地, DNA 扩增指在一个生物的寿命期间中发生, 且使某一 DNA 序列的拷贝数突然增加的事件。

扩增的最有效方法之一是 DNA 复制的滚环模型 (rolling circle mode) (图 8-3; Bostock, 1986)。这种复制类型在两栖类卵母细胞中被用来进行 rRNA 基因的扩增。在这种情

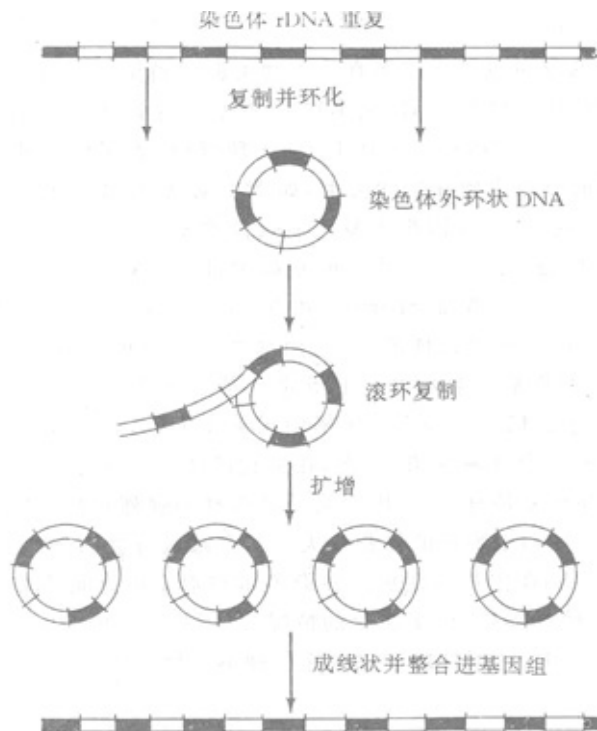


图 8-3 两栖类卵母细胞中基因扩增的滚环模型。染色体的 rRNA 基因呈串联排列, 含有可转录部分 (黑色) 和不转录部分 (白色)。扩增涉及一个染色体外环状拷贝的形成, 该拷贝含有数目可变的重复, 通过多轮的滚环复制而扩增。注意, 重复的周期性可能随滚环扩增而变化。自 Bostock (1986) 修改而成。

况下的扩增包括一个 DNA 序列的染色体外环状拷贝的形成, 然后可产生许多新增加的含原串接重复序列的染色体外单位。如果这样一些单位整合回染色体中, 则该基因组将增加了由等同的重复序列构成的部分。

复制滑脱 (replication slippage) 或 滑脱链误配 (slipped strand mispairing) 是这样一种

过程,即DNA多聚酶转回来并用同一个模板以产生一个重复(图1—8)。现存的串接重复序列特别容易造成复制滑脱,因而此过程能产生短重复的非常长串的排列。滚环复制和复制滑脱都能提供关于串接重复序列在基因组中迅速增生的机制。然而,对于这些过程经验性证据仍然是有限的。

8.6 非基因DNA的维持

解决C—值矛盾问题以及说明真核生物基因组的结构,需要我们为大量非基因的、看起来完全是多余的DNA的长期维持提出一个进化机制。这反过来又与这样的问题紧密联系在一起:这种DNA如果有功能的话、那么可能会是什么样的功能呢?为对这一现象作出进化论的解释,已有不少尝试。下面我们将给出4种这样的假说。

1、非基因DNA行使着至关重要的功能,象基因表达和全局性调控之类(Zuckerlandl, 1976)。根据这一假说,DNA的过剩只是表面上的,而这种DNA是全部都有功能的。因此,这类DNA的缺失或去除将对适合度具有害影响。

2、非基因DNA是无用的废物DNA(Ohno, 1972),它被染色体所被动地携带仅仅只是因为它与有功能的基因在实体上是连续的。根据这一观点,则过剩的DNA并不影响该生物的适合度,于是将无限制地从一代传给另一代。

3、非基因DNA是一种无功能的“寄生物”(Ostergren, 1945)或“自私DNA”(Orgel 和 Crick 1980;Doolittle 和 Sapienza, 1980),是通过基因组间选择而累积且被积极地维持的。

4、DNA有一种结构的或核类型的(nucleotypic)功能,即一种与携带遗传信息的任务无关的功能(Cavalier Smith, 1978)。

关于第一种假说的证据极少。事实上,大多数资料倒是指示出,现在被认为是非基因的DNA事实上就是无功能的,它们中大多数可以缺失而不造成可查觉的表型效应。真核生物中的过剩DNA看来也不会加重代谢系统的负担,在维持和复制大量非基因DNA时能量和营养方面的耗费看来也不是过分的。所以,有可能大多数非基因DNA事实上就是废物或自私DNA。

然而,维持大量的非基因DNA也可能会有些不利之处。首先,大基因组对诱变剂的敏感性表现得比小基因组的高(Heddle 和 Athanasiou, 1975)。其次,维持和复制大量的非基因DNA可能会给该生物带来一定的负担,特别是当基因组的绝大部分都是非基因的时。

卡瓦利耶—史密斯(Cavalier-Smith, 1978, 1985)争辩说,一定有一种维持大基因组的“主要的进化力量”。他的假说是,这种DNA起着决定细胞核体积的核骨架作用。因为细胞越大要求细胞核也越大,所以,导致某种特别的细胞体积的选择,将会次生出导致某种特别的基因组大小的选择的结果。根据这一蓝图,过剩DNA是由选择来维持的,而它的核苷酸组成则可随机地改变。非基因DNA的附加的核类型功能可能是机械的。例如,一种高度重复、串接排列的卫星DNA,可能会影响染色体建筑样式,特别是它的曲度和核粒的相位。有些卫星DNA已知还与特异性的染色体蛋白结合,后者则可能在减数分裂和有丝分裂期间影响染色体的凝聚,因而也许有基因调控方面的效应(Levinger 和 Varshavsky, 1982;James 和 Elgin, 1986)。

没有一种解释看来是已解决了C—值矛盾问题的。以上所有机制,以及许多另外的机制,都可能只对维持“过度的”基因组大小有贡献,而我们将来的任务则是决定每种机制的相对贡献。

8.7 细菌中的GC含量

在真细菌中,基因组序列中鸟嘌呤和胞嘧啶的平均百分比,或GC含量(GC content),在从约25%到75%范围内变化。在许多情况下,细菌的GC含量看来与系统发育有关,亲缘关系接近的细菌有着类似的GC含量(图8—4)。

解释细菌中GC含量的变异本质上有两类假说。选择论者把GC含量看成是对环境条件的一种适应形式。例如,生活在非常热的生态位中的嗜热细菌,强烈地倾向于应用由富GC的密码子编码的对热稳定的氨基酸(例如丙氨酸和精氨酸),且极力避免用由少GC的密码子编码的对热不稳定的氨基酸(例如丝氨酸和赖氨酸),这一现象已有若干报导(例如,Argos等, 1979;Kagawa等, 1984;Kushiro等, 1987)。因

此，G C 含量可能是一种由选择决定的性状。另一个涉及选择论的事件要借助紫外线辐射作为选择力量。因为 T-T 二聚体对辐射敏感，所以，土壤上层的微生物，由于它们常暴露在阳光下，就应该比不暴露在阳光下的，比如说象 *E. coli* 那样的肠道细菌，含有更高的 G C 含量 (Singer 和 Ames, 1970)。

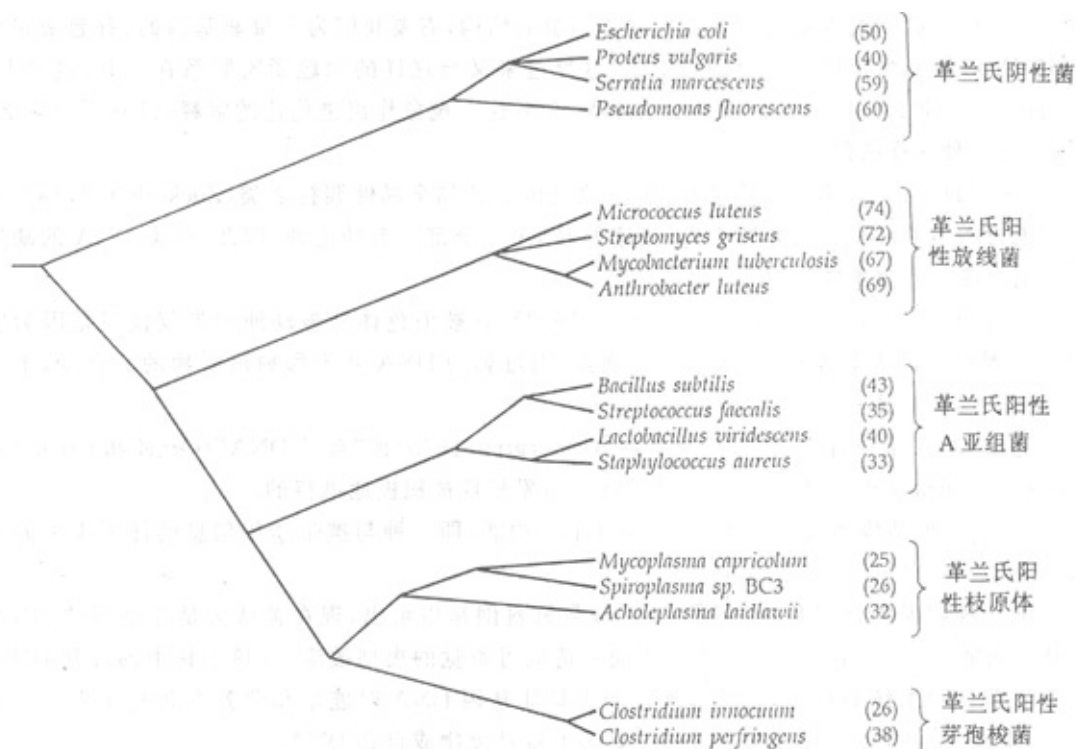


图 8-4 几种真细菌中基因组的 G C 含量 (括号中的数字)。该系统树是根据 5 S rRNA 序列构建的。分枝是无尺度的。自 Hori 和 Osawa (1986), 和 Muto 等 (1986, 1987) 的资料结合而成。

突变论者的观点则以突变模式中的倾向性来解释 G C 含量上的变异 (Suevka, 1964; Muto 和 Osawa, 1987)。根据这一观点，某一给定细菌种的 G C 含量是由以下两个因素的平衡所决定的：(1) 从 G 或 C 到 T 或 A 的替换速率，用 u 表示，和 (2) 从 A 和 T 到 G 或 C 的替换速率，用 v 表示。处于平衡时，G C 含量预期为

$$P_{GC} = \frac{v}{v + u} \quad (8.1)$$

因此，

$$\frac{u}{v} = \frac{1 - P_{GC}}{P_{GC}} \quad (8.2)$$

比率 u/v 又称 **GC 突变压力** (GC mutational pressure)。当 u/v 为 3.0 时，平衡点处的 G C 含量将为 25%。枝原体 *Mycoplasma capricolum* 中的情形就是如此。当该比率为 1 时，G C 含量将为 50%，象在 *E. coli* 中那样。当它为 0.33 时，G C 含量将为 75%，如在 *Micrococcus luteus* 所示。不过，要估计 G C 突变压力，最好选用没有选择限制的部位而不要用总 G C 含量。例如，*Mycoplasma capricolum* 的为蛋白质编码基因中，四重简并位点上的 G C 含量低于 10% (图 8-5)。因而， u/v 一定高于 9。类似地，*Micrococcus luteus* 中简并位点上的 $P_{GC} > 0.9$ ，所以 $u/v < 0.11$ 。

除 G C 突变压力外，突变引起的变化也受选择限制。换句话说，替换的模式是由突变模式和对某些突变的纯洁化选择的模式所决定的 (第四章)。某一特定区域中的选择限制越弱，G C 突变压力对 G C 组成的影响就越强。图 8-5 展示了总 G C 含量与 3 个密码子位置上 G C 含量间的相关，数据来自 11 种细菌，覆盖了范围相当广的 G C 含量值。我们看到，第三密码子位置上的相关与无选择情况下的预

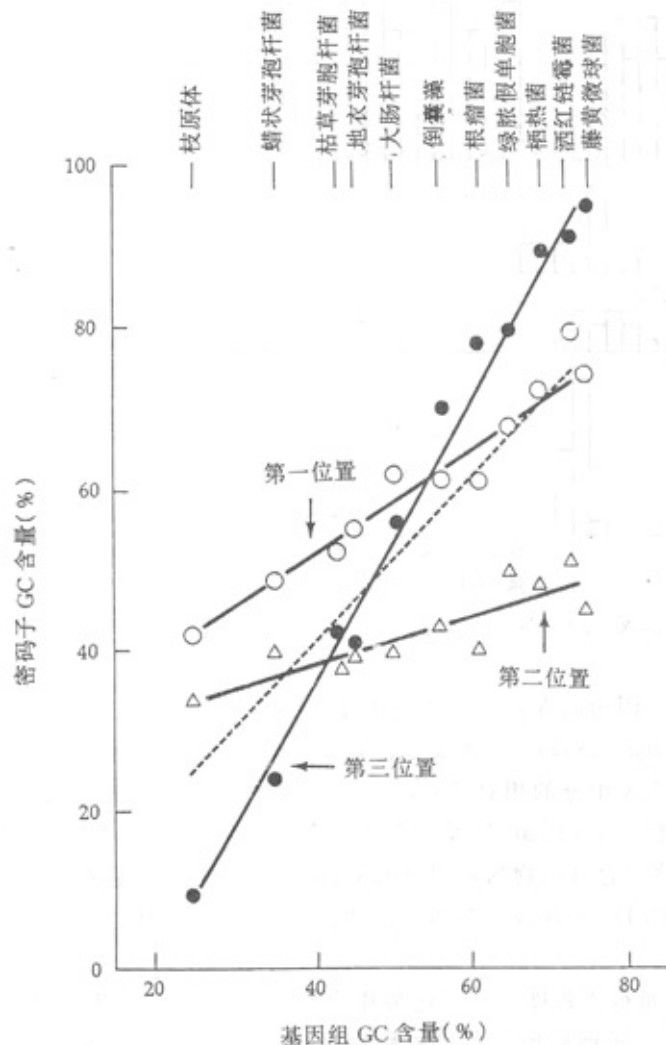


图 8 - 5 总基因组 DNA 和第一、第二以及第三密码子位置间 GC 含量的相关。虚线表示基因组中与密码子中 GC 含量间完全对应时的理论期望关系。自 Muto 和 Osawa (1987)。

期相关类似。另一方面，第一和第二密码子位置上的相关，在正相关的情况下，则表现出平缓得多的坡度。根据下面所说的事实这很容易得到解释，即作用在大多数简并的密码子第三位置上的选择，远不如第一和第二位上的那样严峻（第四章），以致于第三位上的 GC 水平很大程度上是由突变压力来决定的。

8.8 脊椎动物基因组的组成上的组织化

图 8 - 6 展示了不同生物类群中的 GC 含量。有趣的是，虽然多细胞的真核生物的基因组大小一般要比原核生物的大，但真核生物中的 GC 含量却表现出了小得多的变异。特别地，脊椎动物的基因组显示出有相当一致的 GC 含量，处在从约 40% 到 45% 的范围中 (Sueoka, 1964)。脊椎动物中的 GC 含量处于一个较小的范围内的部分原因也许是，脊椎动物与细菌不同，它们相互分歧的时间还不够长，所以还不能在 GC 含量上积累较大的差异。

虽然脊椎动物在基因组的 GC 含量上有一致性，但它的基因组却有着比原核生物基因组更复杂的组成上的组织化。当脊椎动物的基因组被随机地裁剪成长 30—100 kb 的片段，并且将这些片段按其碱基组成分离时，这些片段即聚类成为数不多的、可通过其 GC 含量而相互区别的（富 GC 片段比富 AT 片段重）类型。每一类都以有虽不同但却相似的碱基组成带为特征 (Bernardi 等, 1985; Bernardi, 1989)。

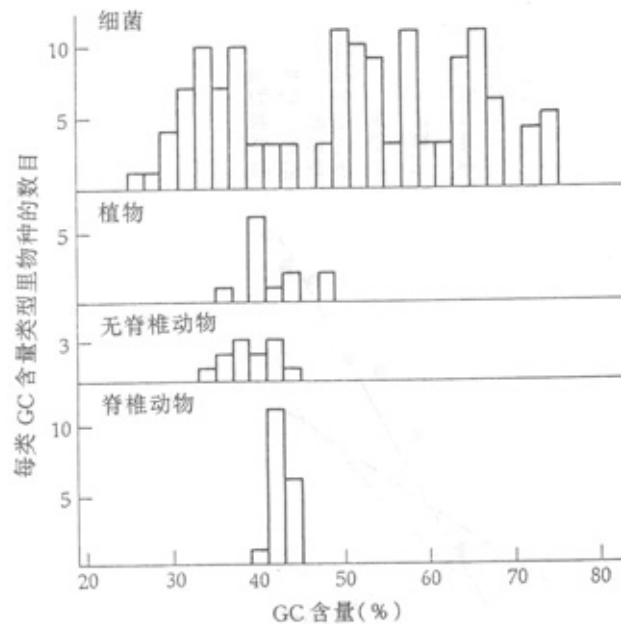


图 8 - 6 不同生物类群中的 G C 组成。自 Sueoka (1964)。

温血和冷血脊椎动物的基因组间在组成上的组织化方面有明显差异 (Bernardi 等, 1985, 1988)。图 8 - 7 a 展示了, 来自鲤 *Cyprinus carpio* 和两栖类 *Xenopus laevis* (左图), 以及来自 3 种温血脊椎动物: 鸡、小鼠和人 (右图) 的主要 D N A 组分相对量和浮力密度。在鸡、小鼠和人的基因组中, 有两个轻组分 (L_1 和 L_2), 代表了约三分之二的基因组, 以及两个或三个重组分 (H_1 , H_2 , 和 H_3), 代表余下的三分之一。相比之下, 来自大多数冷血脊椎动物的基因组 D N A 则主要为轻组分 (图 8 - 7 a)。例如, 在 *Xenopus* 中, 密度高于 $1.704\text{g}/\text{cm}^3$ 的 D N A 片段占基因组的比例少于 10%, 相比之下温血脊椎动物则占 30 - 40%。

D N A 片段的组成上的分布很大程度上是与这类片段的大小独立的, 这指示出在非常长的 D N A 链中有某种组成上的同质性。这种同质的长段术语称为同质段 (isochore)。图 8 - 7 b 展示了来自温血脊椎动物的细胞核 D N A 的镶嵌式组织化 (即, 轻与重同质段轮流出现)。当这些同质段在 D N A 裁剪期间断裂时, 4 种较大的有不同 G C 含量的分子家族即产生了。贝尔纳迪等 (Bernardi 等, 1985) 的结论是, 富 G C 的 (重的) 同质段约占温血脊椎动物基因组的三分之一, 而在冷血脊椎动物中则几乎没有。

温血脊椎动物的基因组是镶嵌式的, 这一发现与染色体分带研究的结果是一致的。温血脊椎动物的中期染色体, 当用荧光染料、蛋白酶水解, 或有差别的变性条件处理时, 将显现出清晰的吉姆萨 (Giemsa) 暗带 (G-带) 和亮带 (R-带)。与之成对照的是, 冷血脊椎动物的中期染色体则只显出很少的分带, 或完全不分带。因此, 曾有建议认为, 少 G C 和富 G C 同质段大致上分别对应于 G-带和 R-带 (Comings, 1978; Cuny 等, 1981)。关于基因的复制时序的研究表明, 位于富 G C 同质段 (R-带) 中的基因在细胞周期的早期复制, 而位于少 G C 同质段 (G-带) 中的基因则在晚期复制 (Goldman 等, 1984; Bernardi 等, 1985; Bernardi, 1989)。

同质段中基因的位置

很多来自人和其他温血脊椎动物的基因组中的基因, 曾用与适当的探针杂交的方法而被定位在这类组成上的同质段上 (Bernardi 等, 1985)。这些研究指出, 基因在整个基因组中的分布是明显地非随机的, 大多数基因位于仅占基因组的 3 - 5% 的最重组分 (H_3) 中。

在大多数情况下, 基因是处在有与其自身类似的 G C 含量的 D N A 片段中。这一发现为同质段的存在提供了独立的证据, 同时也为同质段有较大的尺寸提供了独立的证据。事实上, 因为构成 D N A 制备物的这些片段是经随机降解而产生的, 所以, 这些带有基因的片段在组成上处于较狭窄的范围中的现象表明, 它们在约两倍于这些片段自身大小的区域内 (即高到 200 kb), 有着非常同质的碱基组成。

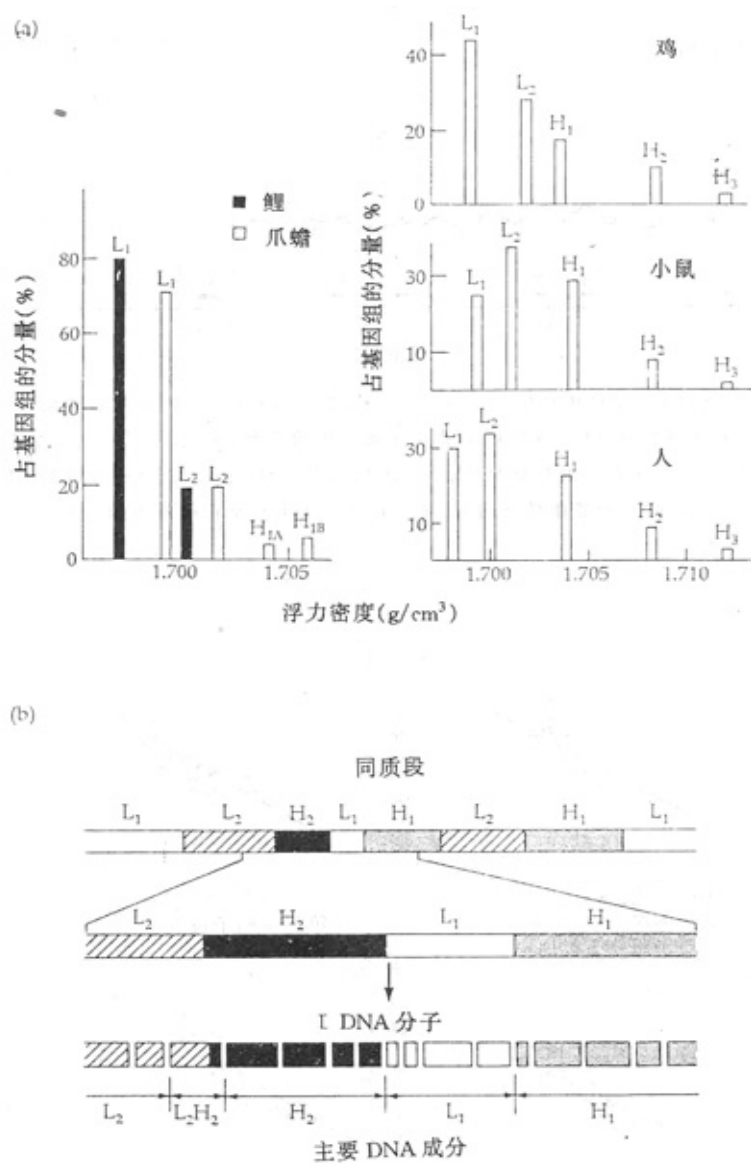


图 8-7 (a) 直方图展示, 来自 *Cyprinus carpio* (鲤) 和 *Xenopus laevis* (爪蟾) (左图) 和来自鸡、小鼠与人 (右图) 的主要 DNA 同质段的相对量和浮力密度。(b) 描绘温血脊椎动物细胞核 DNA 的镶嵌组织化的模式图。当这些同质段在 DNA 制备期间随机断裂时, 4 种较大的具不同 GC 含量的分子家族即产生了。还有几种较小的杂种家族产生。自 Bernardi 等 (1985)。

DNA 序列数据分析已揭示出, 在基因、外显子和内含子的 GC 水平, 与它们处于其中的大段 DNA 区域的 GC 水平间存在正相关 (Bernardi 和 Bernardi, 1985; Bernardi 等, 1985; Ikemura, 1985; Aota 和 Ikemura, 1986)。图 8-8 将人类中的 α -和 β -珠蛋白簇进行了对比。 β 和类 β 珠蛋白基因是低 GC 含量的, 且它们处在低 GC 区域中。反之, α 和类 α 珠蛋白基因是富含 GC 的, 则它们处在富含 GC 的区域中。同样的情形也曾在兔、山羊和小鼠中发现。在鸡中, β -和 α -珠蛋白基因都是富含 GC 的, 而且两者都处在富含 GC 的区域中。相比之下, *Xenopus* (爪蟾) 中的 α -和 β -珠蛋白基因则是少 GC 的, 并且两者都处在少 GC 的区域中。

在绝大多数情况下, 编码区中的 GC 含量有高于侧区中含量的倾向 (图 8-9)。我们还看到, 第三密码子位置上的 GC 水平, 平均下来高于内含子中的水平, 而后者则又比 5' 和 3' 侧区域中的水平高。5' 侧区中的 GC 水平又高于 3' 侧区中的, 这或许是因为启动子及其周围区域倾向于富含 GC 的缘故。

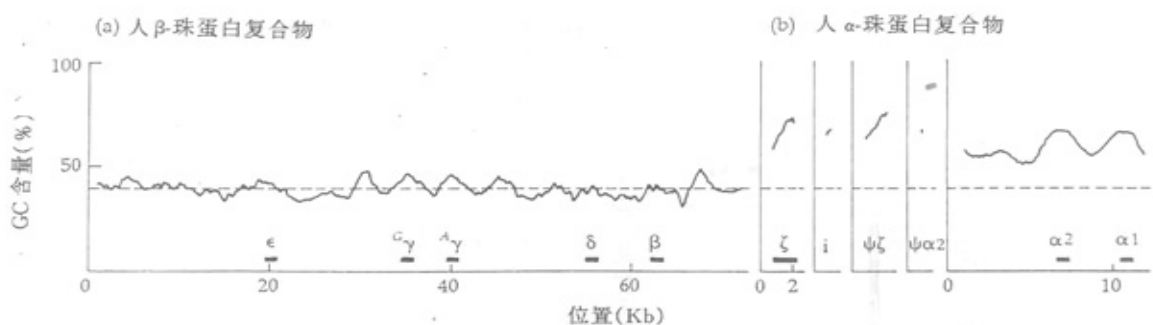


图 8 - 8 GC 含量沿人的珠蛋白 DNA 顺序的分布: (a) β - 珠蛋白基因簇; (b) α - 珠蛋白基因簇 (不完全)。这些基因 (粗线) 以与图 6 - 7 中同样的次序排列。这些基因的名称显示在该图的底部; 区域 i 是 ζ 和 $\psi \zeta$ 之间的基因间区。在 β - 珠蛋白簇和覆盖 $\alpha 1$ - 和 $\alpha 2$ - 珠蛋白基因的区域中, 每一点代表围绕该点的 2 0 0 1 个核苷酸的 GC 组成的平均值, 而其他区域中的每一点则代表 1 4 0 1 个核苷酸的平均值。水平的虚线表示整个人基因组的 GC 含量 (4 0 %)。自 Ikemura 和 Aota (1988) 修改而成。

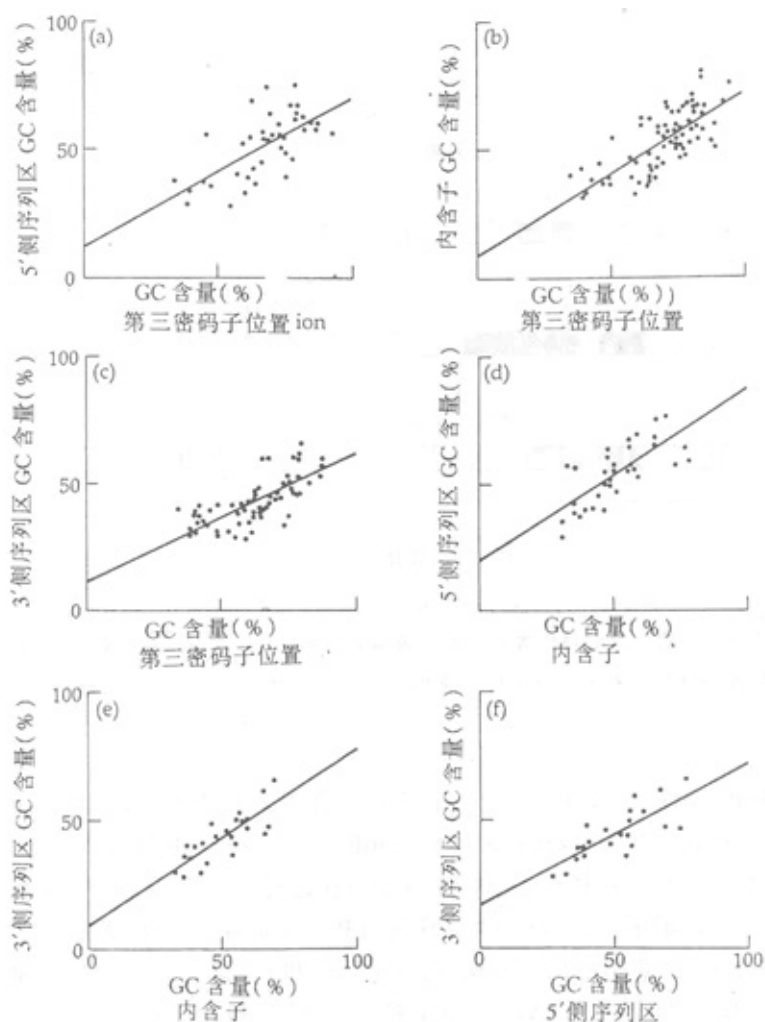


图 8 - 9 基因的各个区域中的 GC 百分比含量间的关系。(a) 第三密码子位置与 5' 侧区。(b) 第三密码子位置与内含子。(c) 第三密码子位置与 3' 侧区。(d) 内含子与 5' 侧区。(e) 内含子与 3' 侧区。(f) 5' 侧区与 3' 侧区。自 Aota 和 Ikemura (1986)。

同质段的起源

同质段的起源一如其被争论的那样神秘莫测。注意，问题在于DNA的长区段(30 kb或更长一些)，它的普遍趋势是或者是富GC的或者是少GC的，而不是在GC含量上发生局部变异——象我们在基因的不同区域中所看到的那样。贝尔纳迪等(Bernardi 等, 1985; 1988)提出，同质段是因功能上的(即选择上的)优势而产生的。他们的主要论据是，在温血生物中，GC含量上的增加可以保护DNA、RNA和蛋白质免受热的降解(见下面)，因为G—C键是比A—T键强的化学键(第一章)。我们称这个观点为选择论者的假说(selectionist hypothesis)。

沃尔夫等(Wolfe 等, 1989a)提出，同质段是种系DNA复制期间，由于前体核苷酸库中的组成变化而造成的突变倾斜所引起的。富含GC的同质段由在种系细胞周期的早期复制的DNA区域所携带，而在那一期间前体库有较高的GC含量，因而就有一种突变成GC的偏向。反之，富含AT的同质段在细胞周期的晚期复制，那时前体库有较高的AT含量，故而有一种突变成AT的倾向。我们称它为突变论者的假说(mutationist hypothesis)。此假说是根据这样的观察而作出的：核苷酸前体库的组成在细胞周期内会发生变化，而这些变化事实上能导致新合成的DNA中碱基比改变(Leeds 等, 1985)。必须清楚，哺乳动物基因组的复制是一个相当漫长的过程，要花8小时或者更多时间(Holmquist, 1987)。

一个支持选择论者假说的理由是，第一和第二密码子位置上GC含量的增加可给予蛋白质以热稳定性，而内含子，第三密码子位置和不翻译区中GC含量的增加，则既能增加原始mRNA转录本的热稳定性，又能稳定染色体的结构，后者或许是通过影响DNA—蛋白质的相互作用而实现的。事实上，嗜热细菌强烈地倾向于采用富GC密码子的情况已有报导(见第127页)。然而，温血脊椎动物的体温要比嗜热细菌所经受的温度低得多，所以，对脊椎动物的蛋白质和DNA序列的进化来说，温度可能并不是一个非常重要的因素。

选择论者假说所面临的一个困难是，大部分哺乳动物和鸟类的基因是低GC含量的这样一个事实。此假说也不能解释为什么有些重复基因有着相反的GC含量。例如，在哺乳类中， β —珠蛋白簇是低GC的，而 α —珠蛋白簇则是富GC的(图8—8)，尽管这两类珠蛋白基因是在同一细胞内、同一时间里表达的，且有着同样的功能。类似地，有些免疫球蛋白基因位于富GC区，而另一些则位于富AT区(见Aota和Ikemura, 1986)。贝尔纳迪等(Bernardi 等, 1985, 1988)的解释是，同质段代表选择的单位，而哺乳动物中的 α 簇已被易位到了富GC同质段， β 簇则被留在低GC同质段中。根据这一理由，鸡中的 α 和 β 簇应该都被易位到富GC同质段了。然而，这一理由却引出了一个富GC同质段的功能优势问题。如果这一优势不是来自它所含有的基因，那么它来自何方呢？

突变论者的假说能解释哺乳类基因组中 α —和 β —珠蛋白簇间GC含量上有较大差异的成因，即假定它们分别位于早期和晚期复制的区域内。不过，它也面临着一些困难(Bernardi 等, 1988)。例如，结构异染色质，如卫星DNA，大多数是富GC的；而功能异染色质，如非活性的X染色体，在细胞周期的尾声阶段复制；在这些情况下，核苷酸库和DNA组成上的变化间看不到有什么联系。

结论是，现有的资料看来还不足以对这两种假说加以评判。也有可能突变压力和自然选择两者，都在形成温血脊椎动物基因组的组成上的组织化中起着作用。

习题

1、表8—5列出了4种原生动物的基因数，平均基因大小的大致估值和C值。(a)假定基因组的基因部分仅由为蛋白质编码的基因构成，那么，每一种生物中非基因DNA的比例是多少？(b)基因组大小和非基因DNA的比例间存在某种关系吗？(注意C值以pg—微微克为单位)。

表8—5 4种原生动物的前体mRNA的平均大小和数目的大致估值和C值

物种	基因组大小(pg)	基因大小(核苷酸)	基因数
<i>Physarum polycephalum</i>	0.57	1500	20000
<i>Oxytricha nova</i>	0.4	2200	24000
<i>Euplotes aediculatus</i>	0.3	1800	40000
<i>Dictyostelium discoideum</i>	0.036	1500	6500

资料取自 Cavalier Smith(1985)

2、表8—6列出了4种生物的前体mRNA和成熟mRNA的平均大小的大致估值和C值。C值上的差异，能用这些生物的基因间在非编码区的大小上存在差异来加以解释吗？（注意C值以pg为单位给出）。

表8—6 4种动物的平均基因大小和mRNA大小的大致估值和C值

属	基因组大小(pg)	前体 mRNA 大小(核苷酸)	mRNA 大小(核苷酸)
<i>Drosophila</i> (果蝇)	0.18	4200	2100
<i>Aedes</i> (蚊子)	0.83 ^a	8400	2100
<i>Strongylocentrotus</i> (海胆)	0.89	8800	2100
<i>Homo</i> (人)	3.50	10000	2100

资料取自 Cavalier-Smith(1985) a、不同物种平均后得到的值。

3、假定在某一生物中，整个基因组都是由为蛋白质编码的基因所构成，20种氨基酸以相等的频率使用，（a）如果同义密码子的选取由尽可能导致最高GC含量的选择所决定，那么这种生物的GC含量将是多少？（b）如果同义密码子的选取由尽可能导致最低GC含量的选择所决定，那么这种生物的GC含量又是多少。

4、假定在某一DNA序列中，每核苷位点每世代的替换速率从G或C到T或A为u，从A或T到G或C为v。设P_t是世代t时G和C核苷酸在该序列中的比例。那么，GC在下一世代中的比例为

$$P_{t+1} = (1-u)P_t + v(1-P_t)$$

根据这一等式，证明平衡点GC比例由等式8.1给出。

5、如果某一序列的起始GC含量为P₀，则可证明世代t时的GC比例由下式给出：

$$P_t = \frac{u}{u+v} + (P_0 - \frac{v}{u+v})e^{-(u+v)t}$$

假定u=3×10⁻⁸，v=5×10⁻⁸，且P₀=0.20，计算t=10⁷时的P_t值。在这些条件下，GC含量上的变化是一个缓慢过程还是一个快速过程？

后继阅读文献

Bernardi, G., B. Olofsson, J. Filipski, M. Zerial, J. Salinas, G. Cuny, M. Meunier Rotival and F. Rodier. 1985. The Mosaic genome of warm blooded vertebrates. Science 228:953-958.

Blake, R. D. and S. Early. 1986. Distribution and Evolution of sequence characteristics in the E. coli genome. J. Biomol. Struct. Dynamics 4: 291-307.

Britten, R. J. and D. E. Kohne. 1968. Repeated sequences in DNA. Science 161:529-540.

Brutlag, D. L. 1980. Molecular arrangement and evolution of heterochromatic DNA. Annu. Rev. Genet. 14:121- 144.

Cavalier Smith, T. (ed). 1985. The Evolution of Genome Size. Wiley, New York. Cold Spring Harbor Symposia on Quantitative Biology. 1986. Molecular Biology of Homo sapiens .Vol. 51. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Deininger, P. L. and G. R. Daniels. 1986. The recent evolution of mammalian repetitive DNA elements. Trends Genet. 2:76-80

Dover, G. A. and R. B. Flavell (eds.). 1982. *Genome Evolution*. Academic Press, New York.

Jelinek, W. R. and C. W. Schmid. 1982. Repetitive sequences in eukaryotic DNA and their expression. *Annu. Rev. Biochem.* 51:813-844.

John, B. and G. Miklos. 1988. *The Eukaryotic Genome in Development and Evolution*. Allen & Unwin, London.

Singer, M. F. 1982. Highly repeated sequences in mammalian genomes. *Int. Rev. Cytol.* 76:67-112