

## 第二章 基因组的结构与功能

### Chapter 2 Structure and Function of Genome

制作：李娟

- ④ 基因组（genome）是指携带某一生物体全部遗传信息的核酸（DNA或RNA）。
- ④ 基因组的结构是指具有一定生物学功能的片段在核酸分子上的分布与排列情况。
- ④ 基因组的功能是贮存和表达遗传信息。

#### 第一节 真核生物基因组的结构与功能

##### Section 1 Structure and Function of Eukaryotic Genome

3. 基因组中存在大量的重复序列以及非编码序列。真核生物基因组内非编码序列占90%以上，是与细菌、病毒的重要区别，在一定程度上也是生物进化的标尺。

#### 真核生物基因组的特点

1. 真核生物基因组都是由大分子双链线状DNA构成，这些DNA通常与组蛋白、非组蛋白组成核小体、染色质及染色体等复合体而存在。染色体通常成对出现（双倍体）。
2. 基因组非常庞大，结构非常复杂，有多个复制起始位点。

4. 真核生物基因组中也存在一些可移动的DNA序列（转座元件），但这些序列无明确的生物学功能，似乎为自己的目的而组织，故有自私基因（selfish DNA）之称。

#### 一、真核生物基因组中重复序列的结构与功能

- ④ 真核生物基因组中通常存在大量的重复序列，可占整个基因组DNA的90%以上。
- ④ 采用复性动力学方法来研究真核生物基因组时，可按其复性速度的快慢或出现频率的高低，将这些重复序列分为高度重复序列，中度重复序列和单拷贝序列三大类。[www.med126.com](http://www.med126.com)

##### (一) 高度重复序列：

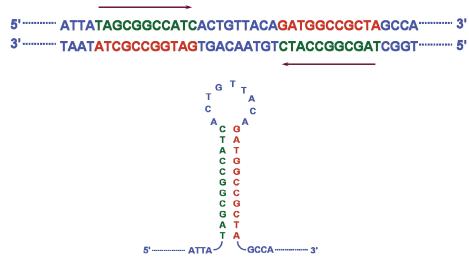
- 高度重复序列在基因组中重复频率高，可达百万次，因此复性速度很快。高度重复序列在基因组中所占比例随种属而异，一般在1~30%范围内。
- 人的高度重复序列约占整个基因组的20%左右。

➤ 高度重复序列按其结构特点可分为三种：

1. 反向重复序列（inverted repeats）：
  - 反向重复序列的长度一般在300~600 bp范围内，分散于整个基因组中。

- 此序列的结构特点是由两个相同顺序的互补拷贝在同一DNA双链上反向排列而成。
- 当反向重复序列的双链DNA分离后，每一条单链内的回排部分可通过链内氢键相连，形成发卡结构或十字架结构。

### 反向重复序列及其发卡结构



### 2. 卫星DNA (satellite DNA) :

- 卫星DNA的碱基组成不同于其它部分，故可用等密度梯度离心法从基因组中分离出来。

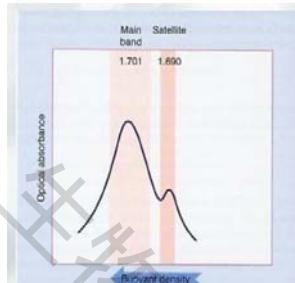


图10-25 小鼠DNA经CsCl密度梯度离心显示出生带和卫星DNA带(转引自Russet,L,1992)

- 卫星DNA的重复单位一般由2~10 bp组成，成串排列。

- 卫星DNA占基因组的比例随种属而异，在0.5~31%范围内。

- 人类基因组中可分离出三类卫星DNA，共占人类基因组的5~6%：

#### ① 大卫星DNA (macrosatellite DNA) :

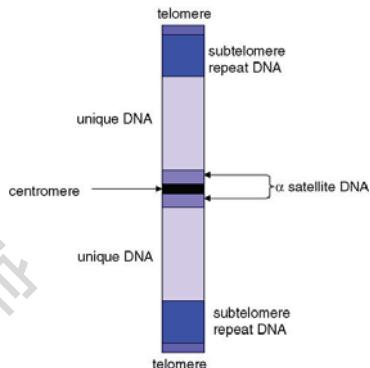
- 其重复单位为5~171 bp，主要分布于染色体的着丝粒区。

#### ② 小卫星DNA (minisatellite DNA) :

- 其重复单位为15~70 bp，存在于常染色体。

#### ③ 微卫星DNA (microsatellite DNA) :

- 其重复单位为2~5 bp，存在于常染色体。



#### 高度重复序列的功能:

- ① 参与复制水平的调节。
- ② 参与基因表达的调控。
- ③ 参与染色体配对。
- ④ 参与转位作用。
- ⑤ 与进化有关。
- ⑥ 同一种属中不同个体的高度重复顺序的重复次数不一样，这可以作为每一个体的特征，即DNA指纹。

[www.med126.com](http://www.med126.com)

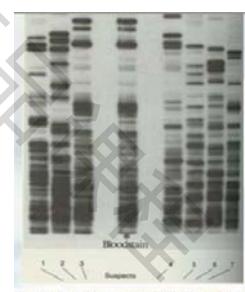


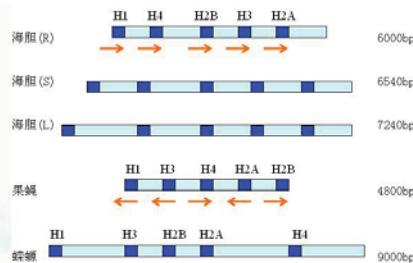
图10-36 人类DNA指纹。来自犯罪现场的血迹指认\*和7个嫌疑者的血液指认，其中标本3和5是犯人的血迹指认相同。(引自 Griffiths, 1999)

### (二) 中度重复序列：

- 中度重复序列是指在基因组中重复十几次至几十万次的部分，其复性速度快于单拷贝序列，但慢于高度重复序列。
- 中度重复序列在基因组中所占比例在不同种属之间差异很大，一般为12~35%，人类基因组中约占12%。

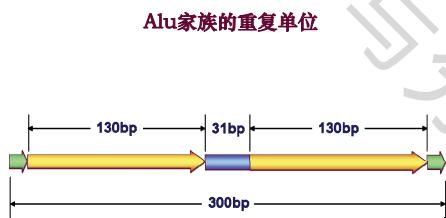
- 中度重复序列中有些属于结构基因，如rRNA基因、组蛋白基因、免疫球蛋白基因、tRNA基因等，这些结构基因常常以串联形式排列在基因组中，故也属于串联重复序列；另一些中度重复序列则属于转座元件。

- 中度重复序列中，典型的编码蛋白质的结构基因是组蛋白基因。
- 各种生物组蛋白基因在基因组中重复频率不同。如海胆中的组蛋白基因重复300~1000次，果蝇的重复约110次，人的组蛋白基因重复30~40次。



- 真核基因组中，中度重复序列主要包括短散在核元件（short interspersed nuclear elements, SINEs）和长散在核元件（long interspersed nuclear elements, LINEs）。

- 典型的短散在核元件（SINEs）是灵长类所特有的Alu重复序列家族，因在其序列中存在限制酶Alu的切点而得名。
- Alu家族的重复单位为300bp，由两段130bp的重复序列与一段31bp的间隔序列构成，两侧各有一段6~20bp的正向重复序列，类似于转座元件的靶位点。因此，Alu家族可通过逆转录机制进行转座。



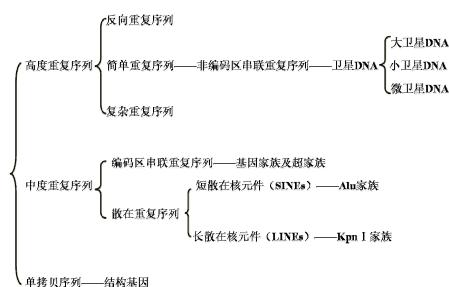
- Alu家族的功能是多方面的，可能参与hnRNA的加工与成熟也与遗传重组及染色体不稳定性有关。最近研究表明，Alu顺序可能具有转录调节作用。

- 典型的长散在核元件（LINEs）是Kpn I 重复序列家族，因在其序列中存在限制酶Kpn I 的切点而得名。
- Kpn I 家族的重复单位一般为6~7 kb或更长，其两侧也各有一段正向重复序列，功能上与Alu家族相似。

### (三) 单拷贝序列：

- 单拷贝序列在基因组中只出现一次或几次，因此复性速度很慢。
- 单拷贝序列属于结构基因，它代表了巨大的遗传信息。

## 真核生物基因组中的重复序列

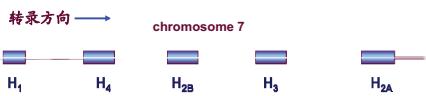


## 二、多基因家族与假基因

- 多基因家族（multigene family）——指DNA序列具有较高的同源性（通常大于50%），并且其编码产物具有相同或相似生理功能的一组结构基因。
- 多基因家族中的基因通常是由同一祖先基因经进化或变异而来。

多基因家族可分为两类：

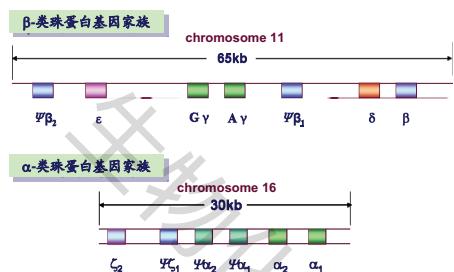
1. 基因家族成簇地分布在同一染色体上并同时进行转录，如组蛋白基因家族；



2. 基因家族成簇地分布于不同的染色体上并分

别进行转录，且不同基因编码的蛋白质在功能上相关，如珠蛋白基因家族。

### 珠蛋白多基因家族的组织结构



### 人血红蛋白基因家族中的假基因 (ψ)

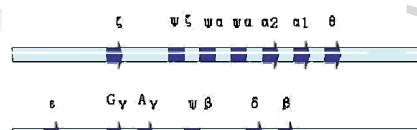


图 10-31 人类血红蛋白的  $\alpha$  和  $\beta$  基因簇（引自 B.Lewin 2000）

假基因 (pseudogene) —— 又称为加工基因或非功能基因。这类基因的核苷酸顺序虽然与正常的结构基因很相似，但基本上不能表达。

### 三、线粒体DNA的结构

线粒体DNA (mitochondrial DNA, mtDNA)

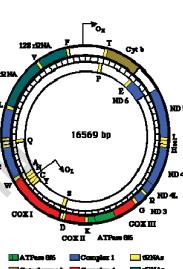
www.med126.com 属于真核细胞核外遗传物质，可独立编码存在于线粒体中的多肽链、rRNA或tRNA。

mtDNA为双链环状DNA，其分子结构特点与原核生物DNA相同。

假基因是由于在进化过程中，某些DNA片段发生了缺失、倒位或点突变，导致调控基因丢失；或无剪接加工信号；或编码区出现终止信号；或编码无功能或不完整的基因。

mtDNA长16,569 bp，含

37个结构基因，编码13种蛋白、22种tRNA和2种rRNA。这13种蛋白质为呼吸链酶复合物的单位。

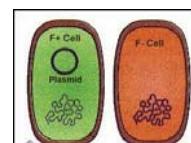


### 第二节 原核生物基因组的结构与功能

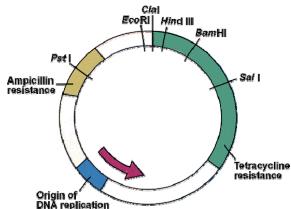
#### Section 2 Structure and Function of Prokaryotic Genome

### 原核生物基因组的特点

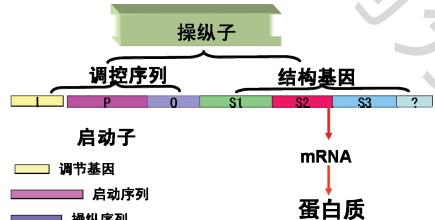
1. 基因组相对较小，由DNA组成，包括染色体DNA和质粒DNA两种DNA分子，均为共价闭环双链。染色体DNA为单拷贝。



2. 每个DNA分子（染色体DNA和质粒DNA）只有一个复制起始点。



4. 结构基因通常为连续基因，少量重叠基因，非编码区和重复序列少。  
5. 功能相关的基因常常组织形成操纵子结构，操纵子结构是原核基因组的一个突出的结构特点。



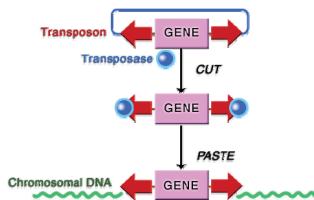
### 原核生物基因组中的质粒DNA



### (二) 质粒的遗传控制

- 由四个系统进行精细调控。
  - 复制控制系统:**
    - 由质粒上的复制起始点（ori）、rep基因和cop基因组成，主要参与调控细胞中质粒的拷贝数。
    - Rep蛋白启动复制，而Cop蛋白则抑制复制。

3. 含可转移的序列，如插入序列、转座子等。



### 一、原核生物基因以操纵子方式结构

- 功能上相关的若干结构基因串联在一起，由一套调控序列控制其转录表达，构成基本转录单位，称为**操纵子（operon）**。
- 其转录产生的mRNA通常为**多顺反子（multicistron）**：一个mRNA可编码多个蛋白质。

### 二、原核生物中的质粒DNA

#### （一）质粒的基本特性

- 质粒（plasmid）是指存在于细菌中的双链环状小分子DNA，可独立于染色体DNA进行自主复制，一般仅携带抗药基因等与细菌生长繁殖无直接关联的基因。

- 质粒分子量一般为 $10^6\sim 10^8$ ，小型质粒的长度为 $1.5\sim 15$  kb。
- 质粒只有在宿主细胞内才能完成自己的复制，一旦离开宿主就无法复制和扩增。但质粒对宿主细胞的生存不是必需的，宿主细胞丢失了质粒依旧能够存活。

#### 2. 质粒分配系统：

- 由质粒分配区序列组成，决定质粒在细胞分裂过程中精确分配到子细胞中的数目。
- 3. 细胞分裂控制系统:**
  - 能抑制细胞分裂，使细胞分裂与质粒复制协调，避免产生不含质粒的子细胞。
- 4. 位点特异重组系统:**
  - 该系统主要使质粒多聚体拆分为单体，以保证质粒在子细胞中的平均分配。

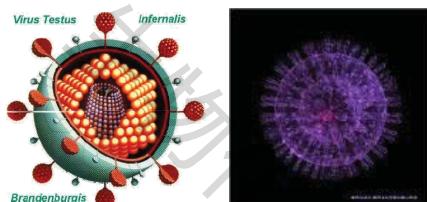
### (三) 质粒的不相容性

- 具有相同复制起始位点和分配区的两种质粒DNA不能在同一宿主细胞中共存的现象，称为质粒的不相容性。
- 具有相同复制起始位点的两种质粒，一般共用同一分配系统来对质粒进行分配，从而产生竞争。对单个细胞而言，经数代繁殖后，这种不均衡的分配就会造成子细胞丢失其中一种质粒。

### 第三节 病毒基因组的结构与功能

#### Section 3 Structure and Function of Virus Genome

病毒是由一个或几个核酸分子组成的基因组，有一层蛋白或脂蛋白保护性外壳，且可在一定宿主细胞中自我复制的感染性因子。



#### 一、病毒基因组的类型



#### 二、病毒基因组的特点

1. 基因组大小变化较大:
  - 一般为3~300 kb。
2. 基因组的组成与结构变化较大:
  - DNA或RNA，双链或单链，线状或环状，连续或不连续，单倍体或多倍体。

#### 3. 基因的结构变化较大:

- 连续基因，断裂基因，串联基因，mRNA缺失5'-端帽子结构或翻译起始序列等。

#### 4. 基因重叠现象较普遍:

- 核酸分子中的一部分核苷酸序列为两个或两个以上的基因所共用，即一个基因的序列与另一个基因的序列发生重叠，这些基因称为重叠基因。
- 按重叠方式不同，可分为完全重叠、部分重叠和读码重叠。