

基因、基因组和基因组学

生物化学与分子生物学教研室

赵蕊



第二节 基因组 (Genome)

病毒基因组

原核基因组

真核基因组



病毒基因组

病毒（virus）的概念

是一类比较原始的、有生命特征的、能够自我复制的、严格细胞内**寄生的非细胞生物**，是结构最简单、最微小的生命形式。

病毒的四种基本结构成份：

- ①病毒基因组
- ②衣壳
- ③包膜
- ④病毒颗粒中的其它内容物

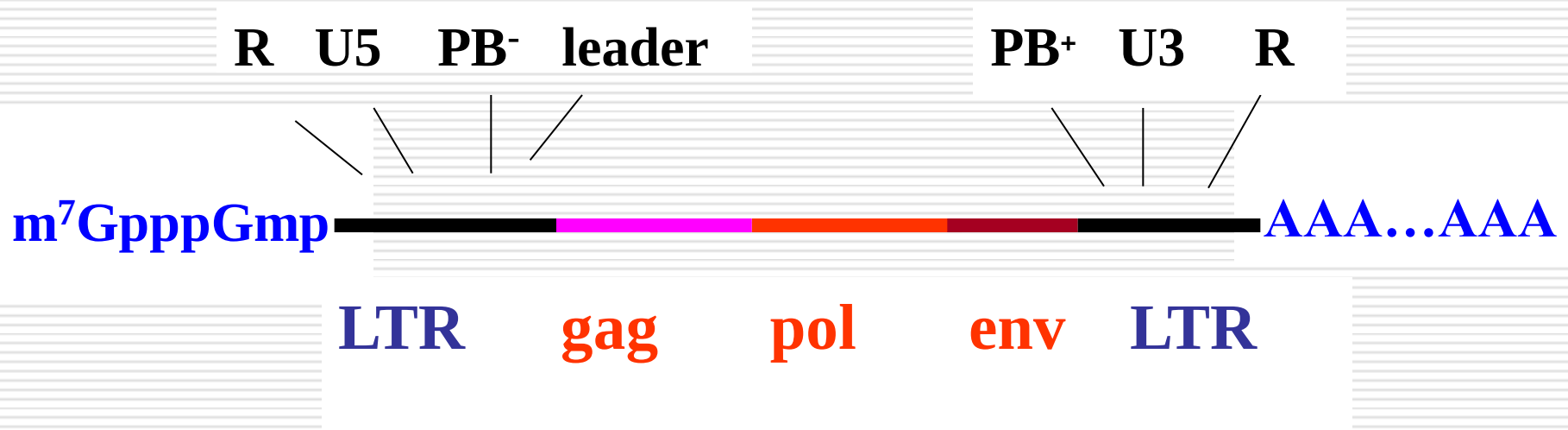


病毒基因组的特点

1. 每种病毒核酸只有一种，或者DNA，或者RNA；
2. 病毒核酸大小差别很大，3kb—300kb；
3. 大部分病毒核酸是由一条双链或单链分子(RNA或DNA)，仅少数RNA病毒由几个核酸片段组成。
4. 具有操纵子的结构，具有重叠基因的结构。
5. 真核病毒基因有内含子，而噬菌体（感染细菌的病毒）基因中无内含子。
6. 除逆转录病毒外，所有病毒基因都是单拷贝的。



逆转录病毒基因组RNA的结构

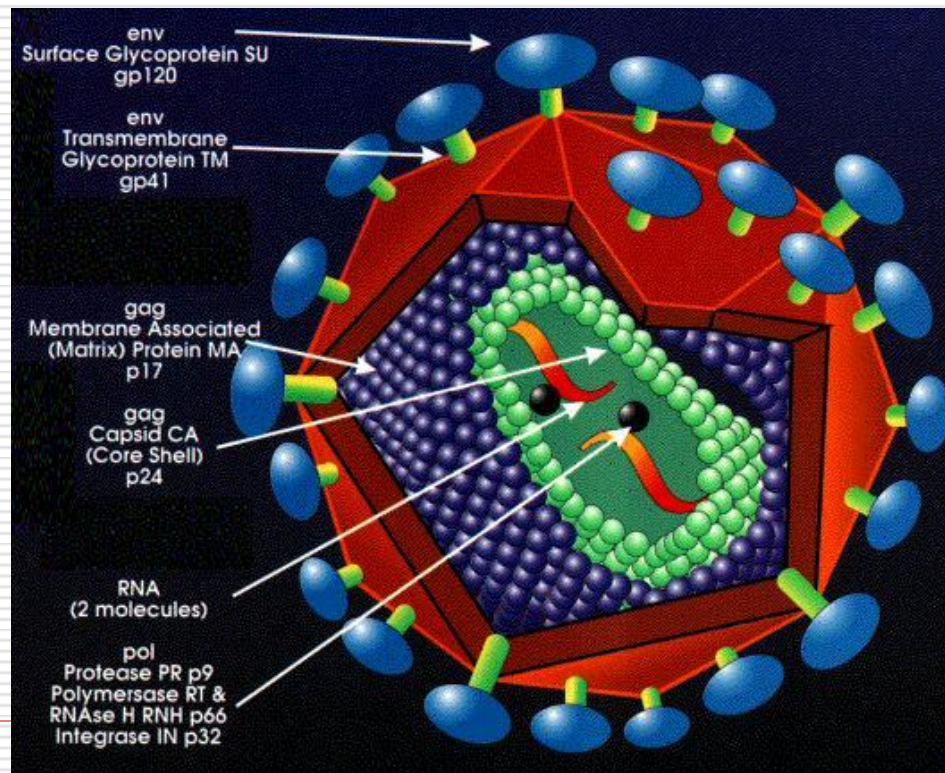
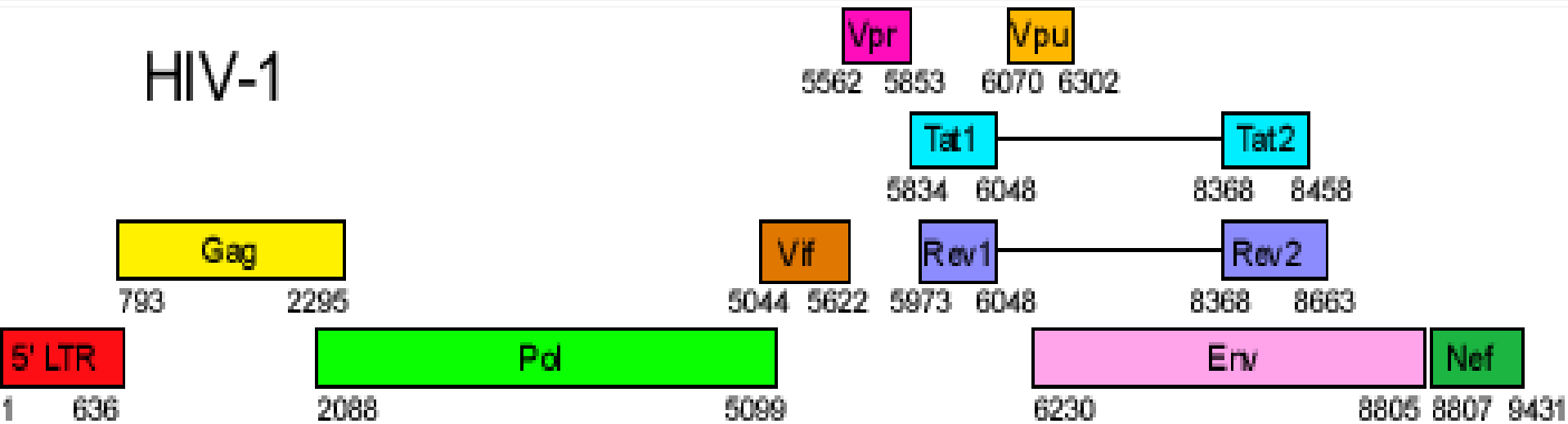


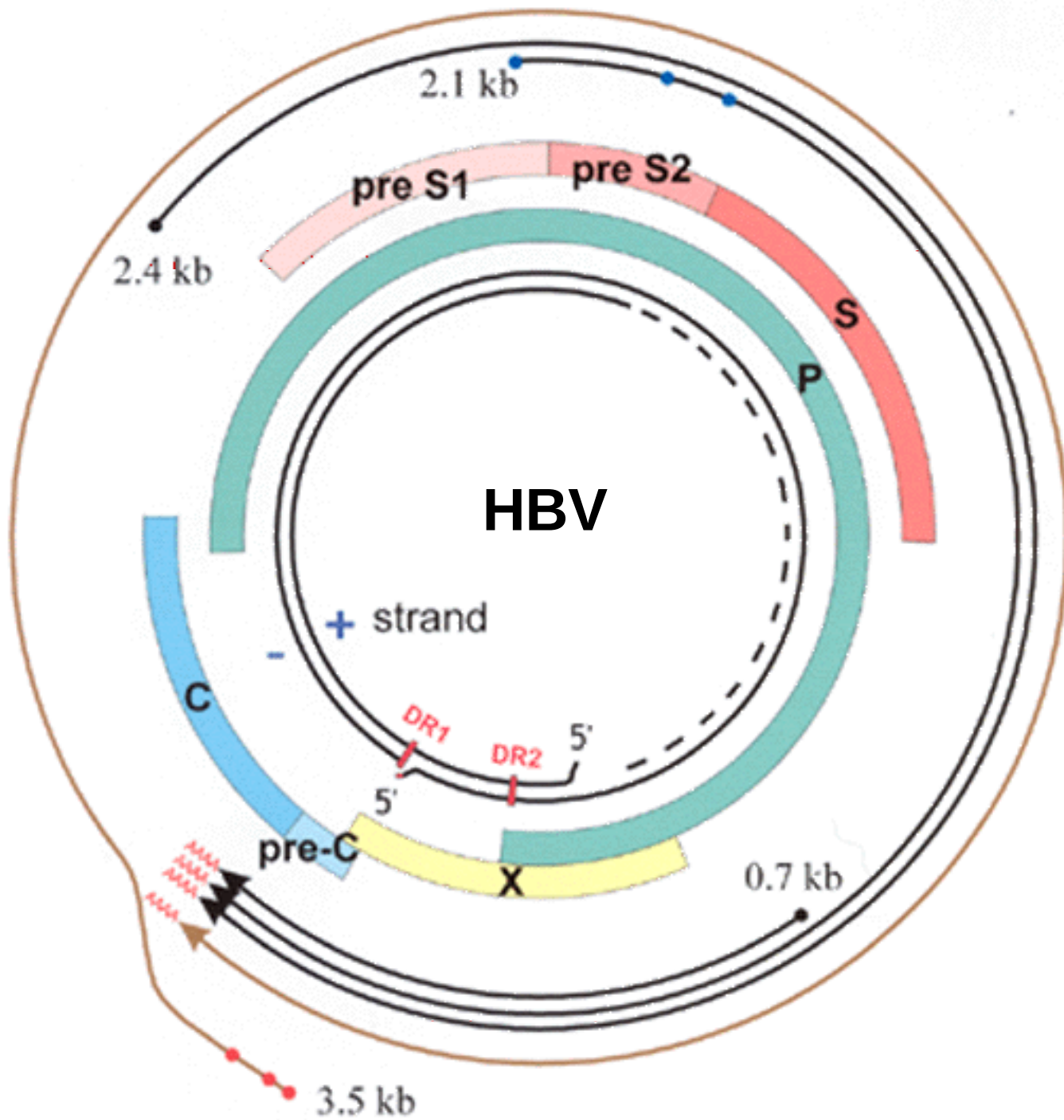
5' 帽—R—U5—PB⁻—DLS—ψ gag—pol—env (onc—) —

C—PB⁺—U3—R—poly(A)_n



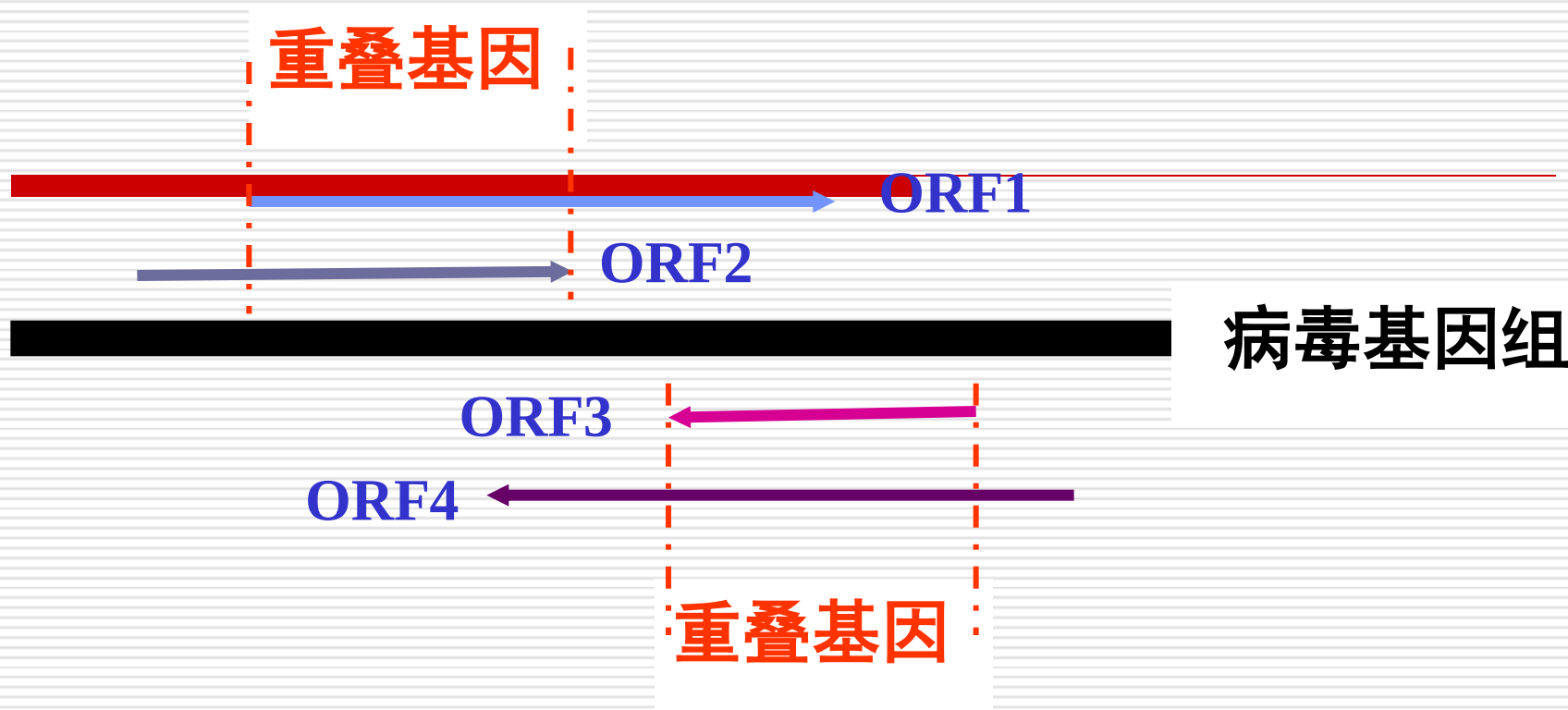
HIV-1





乙型肝炎病毒的基因组结构



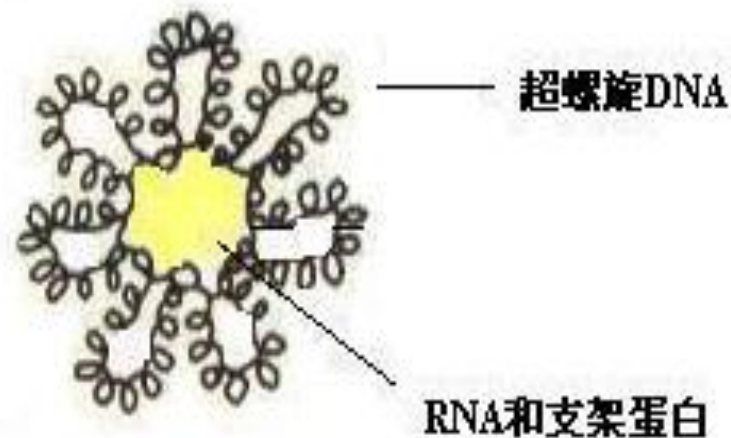


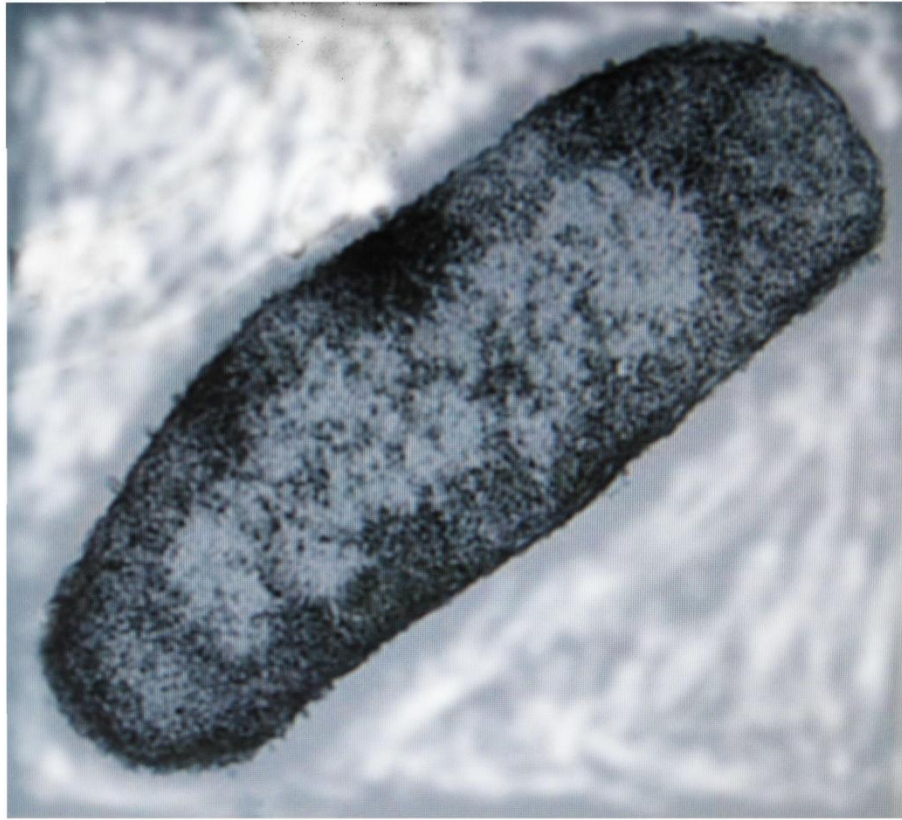
重叠基因： 病毒基因组的一段DNA序列有两个或两个以上的开放读码框架，可以编码两种或两种以上的多肽链，称为重叠基因。



3. 原核生物基因组

- 类核结构:
- 中央部分—RNA和支架蛋白
- 外围部分—双链闭环的超螺旋DNA





大肠杆菌类核结构照片图



已完成测序的部分病原原核生物

- 金黄色葡萄球菌 (*Staphylococcus aureus*, SaurN) 3
- 肺炎链球菌 (*Streptococcus pneumoniae*, Spneu) 2
- 化脓性链球菌 (*Streptococcus pyogenes*, Spyo) 2
- 脑膜炎奈瑟菌 (*Neisseria meningitidis*, Nmen) 2
- 沙门氏菌 (*S. typhi*; *S. typhimurium*) 2
- 空肠弯曲菌 (*Campylobacter jejuni*, Cjej) 1
- 大肠杆菌 (*E. coli* O157:H7, Ecoli_O157) 2
- 幽门螺杆菌 (*Helicobacter pylori*, Hpyl) 2
- 霍乱弧菌 (*Vibrio cholerae*, Vcho) 1



已完成测序的部分病原原核生物

□ 鼠疫耶氏菌 (<i>Yersinia pestis</i>)	3
□ 布鲁氏菌 (<i>Brucella melitensis</i> 16M)	1
□ 梭菌 (<i>Fusobacterium nucleatum</i>)	1
□ 流感嗜血杆菌 (<i>Haemophilus influenzae</i> , Hinf)	1
□ 麻风分枝杆菌 (<i>M. leprae</i> , Mlep)	1
□ 结核分枝杆菌 (<i>M. tuberculosis</i> , Mtub)	2
□ 巴斯德氏菌 (<i>Pasteurella multocida</i> , Pmul)	1
□ 李斯特杆菌 (<i>Listeria monocytogenes</i>)	1
□ 绿脓假单胞菌 (<i>P. aeruginosa</i> , Paer)	1



已完成测序的部分病原原核生物

□ 沙眼衣原体 (<i>Chlamydia trachomatis</i> , Ctra)	2
□ 肺炎衣原体 (<i>Chlamydia pneumoniae</i> , Cpneu)	3
□ 生殖道枝原体 (<i>M. genitalium</i> , Mgen)	1
□ 肺炎枝原体 (<i>M. pneumoniae</i> , Mpneu)	1
□ 支气管枝原体 (<i>M. pulmonis</i> , Mpul)	1
□ 溶脲脲原体 (<i>Ureaplasma urealyticum</i> , Uure)	1
□ 立克次体 (<i>Rickettsia prowazekii</i> , Rpxx)	2
□ 梅毒螺旋体 (<i>Treponema pallidum</i> , Tpal)	1
□ 伯氏疏螺旋体 (<i>Borrelia burgdorferi</i> , Bbur)	1



细菌基因组的特点

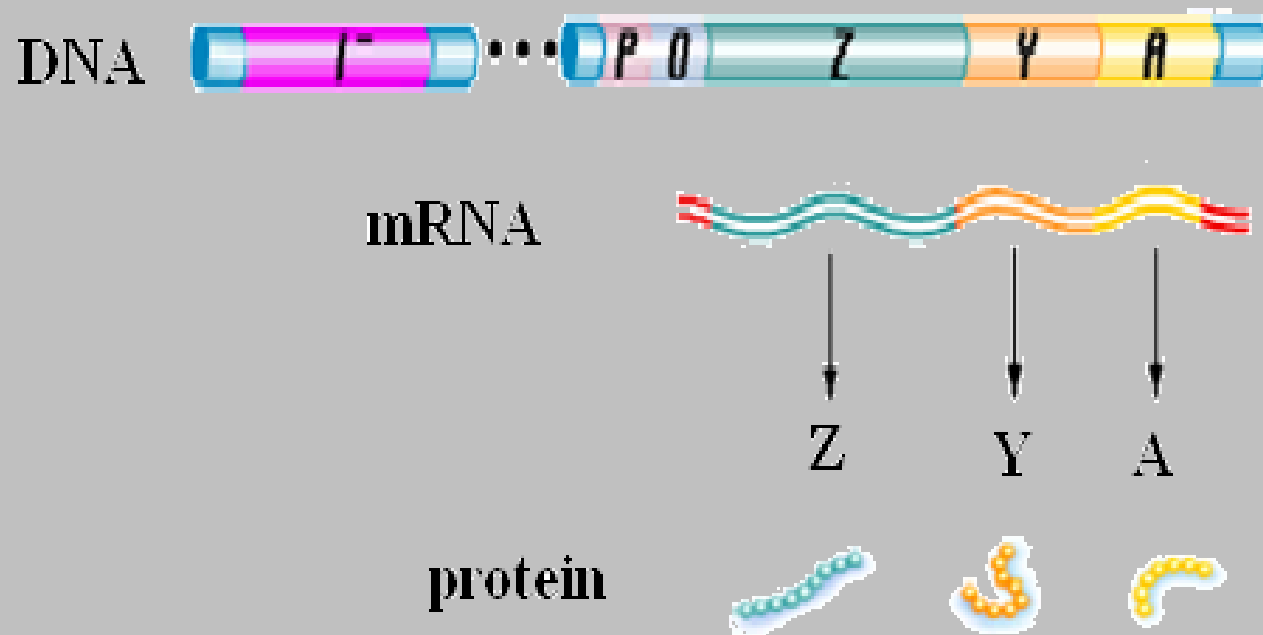
□ 细菌染色体基因组：大多仅由一条环状双链DNA分子组成

1. 细菌DNA大部分为编码序列。
2. 结构基因中没有内含子，也无重叠现象。
3. 操纵子结构普遍

□ 质粒

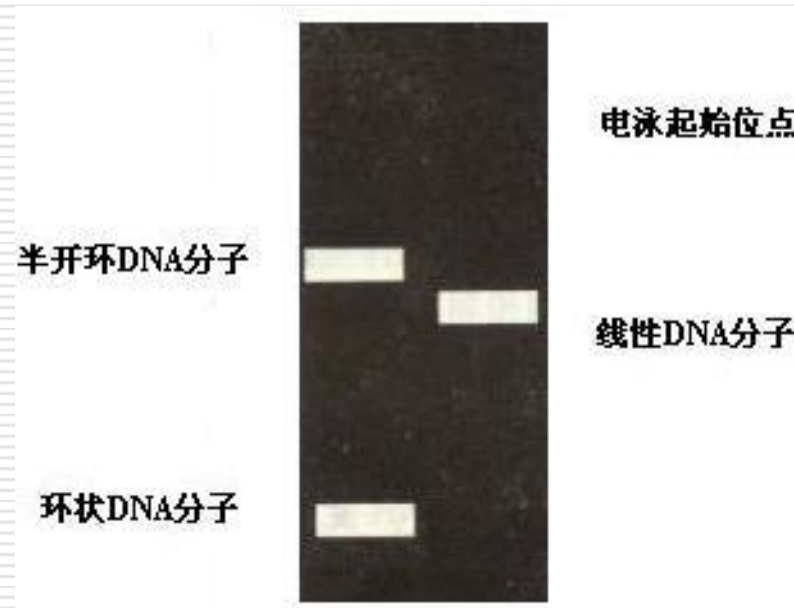


乳糖操纵子



质粒

- ◆ **质粒**是细菌中染色体以外的能独立复制并稳定遗传的共价闭合环状DNA分子。
- ◆ **质粒可以移动**
- ◆ 质粒通常有三种分子构型，即共价闭合环状DNA分子、半开环DNA分子和线性DNA分子。

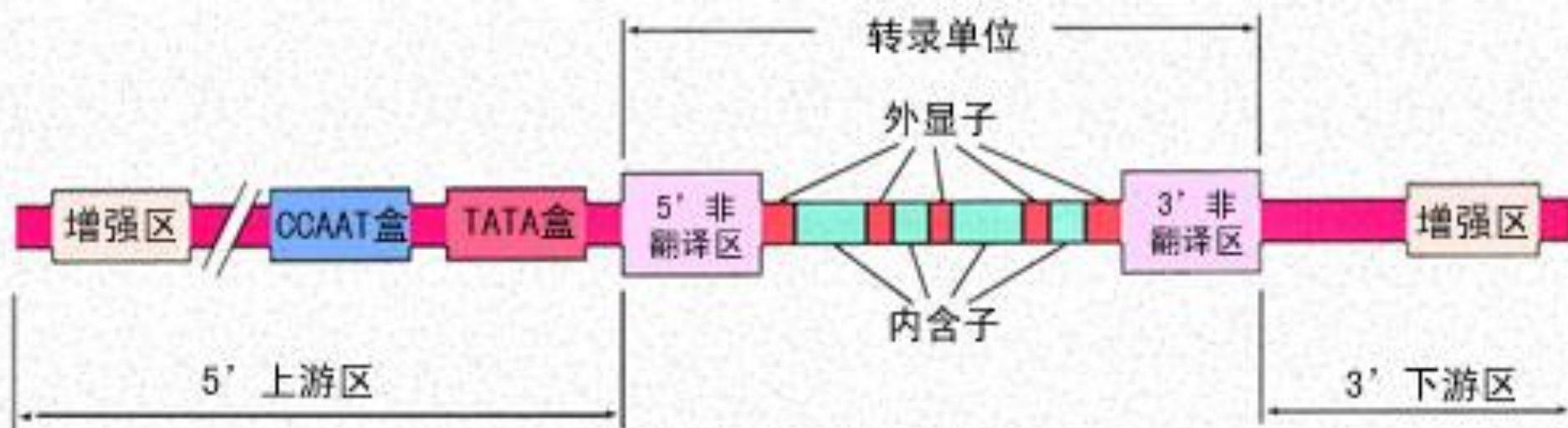


真核生物基因组

真核生物基因组的特点

1. 体细胞内的基因组为二倍体。
2. 基因组远远大于原核生物的基因组，具有多个复制起点。
3. 基因组中不编码的区域多于编码区域。





现代关于真核基因的认识



-
4. 大部分基因含有内含子，因此，基因是不连续的（断裂基因）。
 5. 真核细胞基因转录产物为单顺反子。一个结构基因经过转录和翻译生成一个mRNA分子和一条多肽链。
 6. 存在大量重复序列。
 7. 存在多基因家族和假基因。



重复序列

（一）高度重复序列

重复频度 $>10^5$

卫星DNA (satellite DNA)

小卫星DNA (mini-satellite DNA)

微卫星DNA (micro-satellite DNA)

（二）中度重复序列

$10^2 - 10^5$ 次

1. rRNA基因

2. tRNA基因

3. Alu家族



卫星DNA (satellite DNA)

1. 卫星DNA (macro-satellite DNA)

重复单位5—100bp

2. 小卫星DNA (mini-satellite DNA)

重复单位6—70bp

又称可变数目串联重复序列

(Variable number tandem repeat, **VNTR**)

3. 微卫星DNA (micro-satellite DNA)

重复单位2—6bp

又称短串联重复 (Short tandem repeat, **STR**)



基因家族(gene family)

- 一组功能相似、核苷酸序列具有同源性的基因。
- 可能由某一共同祖先基因经重复和突变产生。



基因家族的特点

- ① 基因家族的成员可以串联排列在一起，形成基因簇(gene cluster)或串联重复基因，如rRNA、tRNA和组蛋白的基因；
- ② 有些基因家族的成员也可位于不同的染色体上，如珠蛋白基因。
- ③ 有些成员不产生有功能的基因产物，这种基因称为假基因 (Pseudogene)。



基因家族分类

□ 编码RNA的:

❖ rRNA、tRNA、snRNA等;

□ 编码蛋白质的:

❖ 组蛋白基因

❖ 珠蛋白基因

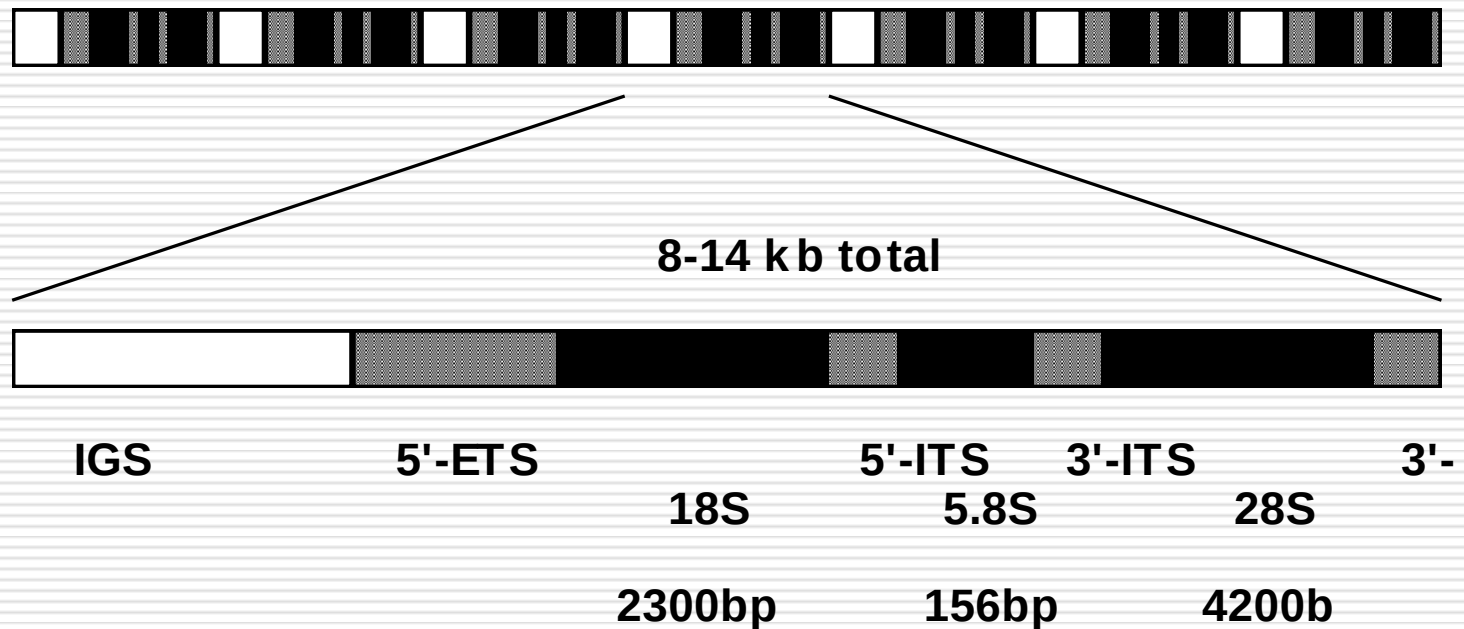
❖ 生长激素



rRNA 基因

□ >100copy.

□ rRNA基因簇(重复单元18S - 5.8S - 28S RNA)



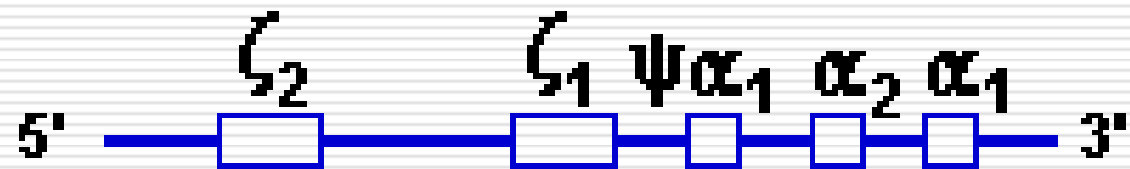
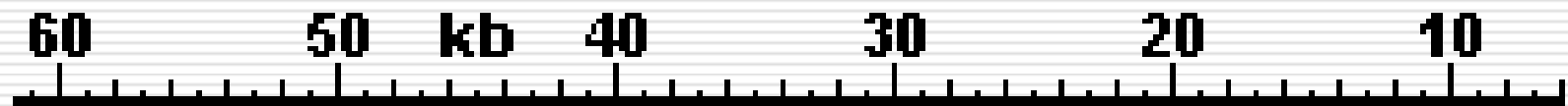
IGS: intergenic spacer

ETS: extragenic transcribed spacer

ITS: intragenic transcribed spacer



珠蛋白基因



Chromosome #16



Chromosome #11



假基因

- 有些成员不产生有功能的基因产物，这种基因称为假基因 (Pseudogene)。

$\Psi a1$ 表示与a1相似的假基因。



线粒体基因组

线粒体DNA (mtDNA)遗传特性：

- 1、母系遗传
- 2、高突变率
- 3、结构基因不含内含子，部分区域有基因重叠现象
- 4、受双重遗传控制
- 5、具有自我复制、转录和编码的功能。
- 6、线粒体遗传密码与通用密码不同



小结

真核生物基因组与原核生物基因组的主要区别：

- (1) 真核生物基因组DNA与蛋白质结合形成**染色体**，储存于**细胞核内**，除配子细胞外，体细胞内的基因组是双份的（即**双倍体**）。细菌染色体基因组通常由一条环状双链DNA分子组成，染色体形成类核，无核膜与胞浆分开。
 - (2) 基因组远远大于原核生物的基因组，具有许多复制起点，而每个复制子的长度较小。
 - (3) 真核细胞基因转录产物为单顺反子。一个结构基因经过转录和翻译生成一个mRNA分子和一条肽链。原核生物基因转录产物为多顺反子，功能上相关的几个基因往往在一起组成操纵子结构。
-



真核生物基因组与原核生物基因组的主要区别

- (4) 真核基因组大部分基因含有内含子，因此，基因是不连续的，称为断裂基因，需要进行转录后加工；原核基因组没有内含子结构，不需进行转录后剪接加工。
- (5) 真核基因组中不编码的区域多于编码区域。原核基因组大部分为编码序列，不编码区域仅占一小部分。
- (6) 真核生物基因组存在重复序列，重复次数可达百万次以上基因组远远大于原核生物的基因组。
- (7) 真核生物基因组存在多基因家族和假基因。

