

第二章 基因概念的演化

- 基因的演化
- C值矛盾
- 细菌基因特点
- 病毒基因特点
- 真核基因组特点
- 真核基因组中DNA序列特征
- 基因家族

基因的概念及发展

- 1909年:丹麦遗传学家W. Johansen将孟德尔遗传因子命名为基因(gene)
- 1910-1925年摩尔根利用果蝇做研究材料, 证明基因是呈直线排列的遗传单位
- 1941年George 等通过对粗糙链孢霉营养缺陷型的研究, 认为基因是通过控制酶和其他蛋白质合成来控制细胞代谢--“一个基因一个酶”学说

One Gene, One Enzyme

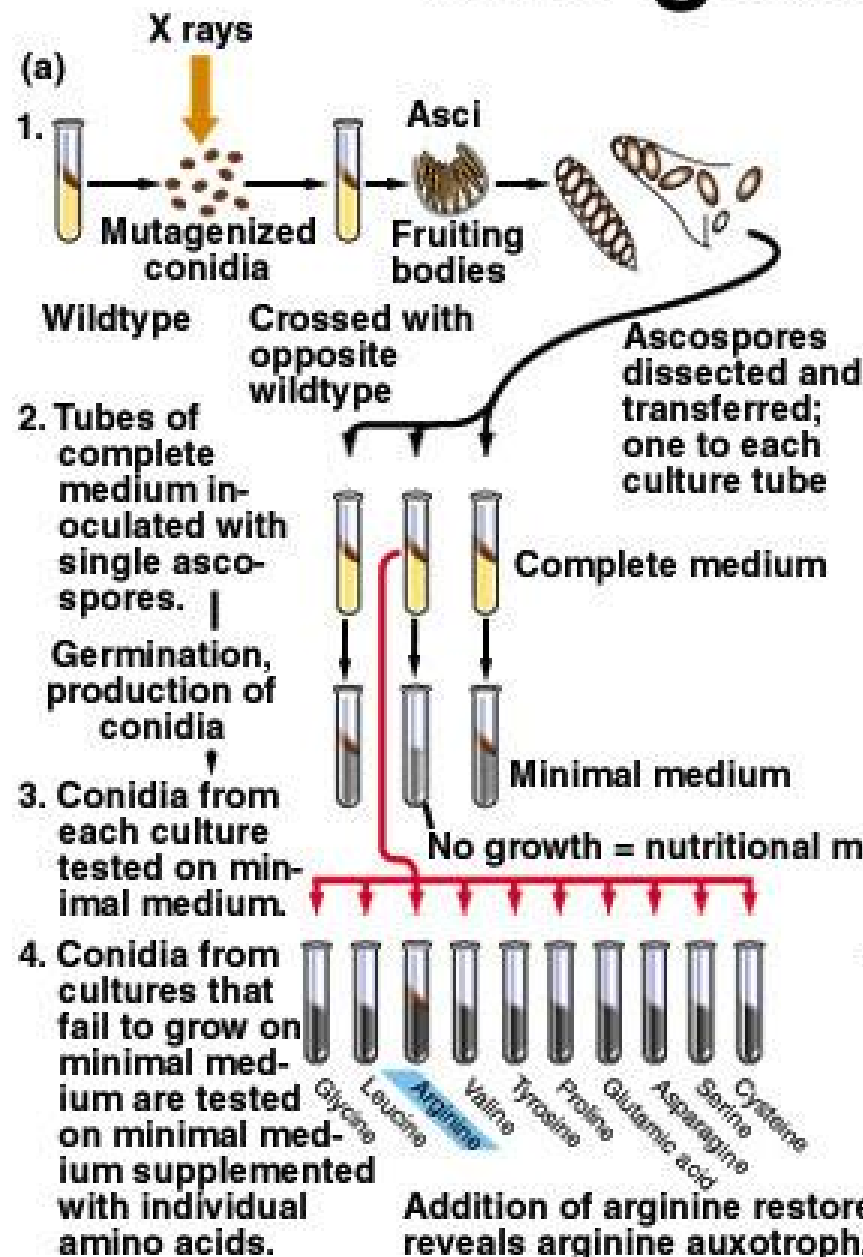
链孢霉营养缺陷型：精氨酸依赖型

前体 \longrightarrow 鸟氨酸 \longrightarrow 瓜氨酸 \longrightarrow 精氨酸

在基本培养基上分别添加三种氨基酸，结果如下：

	鸟氨酸	瓜氨酸	精氨酸
突 1	—	—	+
变 2	—	+	+
体 3	+	+	+
说明	酶1	酶2	酶3

One gene, one enzyme



(b) Growth response if nutrient is added to minimal medium

Mutant strain	Supplements				
	Nothing	Ornithine	Citrulline	Arginino-succinate	Arginine
Wildtype: Arg^+	+	+	+	+	+
$Arg-E^-$	-	+	+	+	+
$Arg-F^-$	-	-	+	+	+
$Arg-G^-$	-	-	-	+	+
$Arg-H^-$	-	-	-	-	+

(c)

Gene:

$ARG-E$

$ARG-F$

$ARG-G$

$ARG-H$

Enzymes:

Acetylornithinase

Ornithine transcarbamylase

Argininosuccinate synthetase

Argininosuccinate lyase

Reactions:

N -Acetylornithine

Ornithine

Citrulline

Argininosuccinate

Arginine

Carbamyl phosphate

Aspartate

基因的概念及发展

- 1957年S. Benzer用大肠杆菌T4噬菌体为材料，在DNA分子结构水平上，通过互补实验，分析了基因内部的精细结构，提出顺反子(cis trans)的概念：
独立的功能单位
- 基因是一个具有特定功能的，完整的，不可分割的最小的遗传单位
- 顺反子概念：推翻了基因是三位一体的概念

T4噬菌体野生型与突变型的区别

- T4r⁺野生型: E. coli K可以生长
T4r^{II}突变型: E. coli K不能生长
- 1957年S. Benzer用大肠杆菌T4噬菌体为材料, 在r^{II}区通过大量成对突变/杂交: (400个突变)
两种r^{II}区突变体的混合物能否在K菌株上生长?

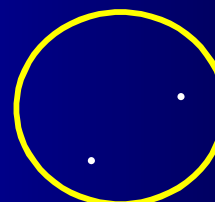
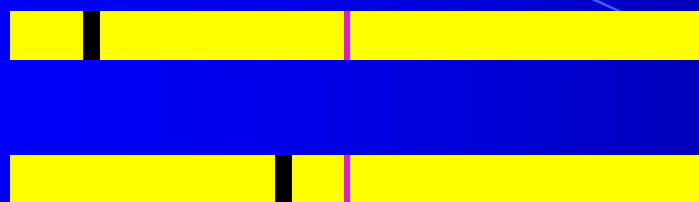
类似于肺炎双球菌

突变位点

A

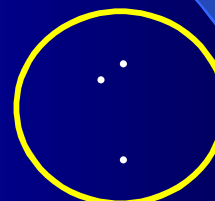
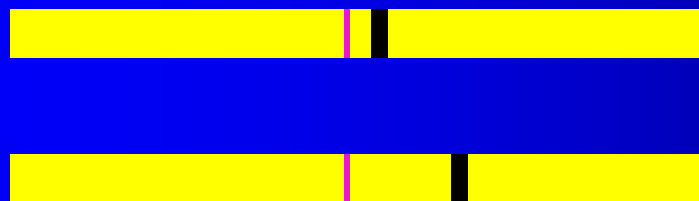
B

无互补



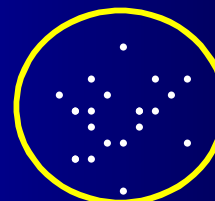
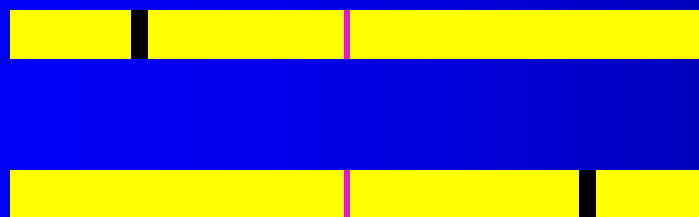
没有或极少噬菌斑

无互补



没有或极少噬菌斑

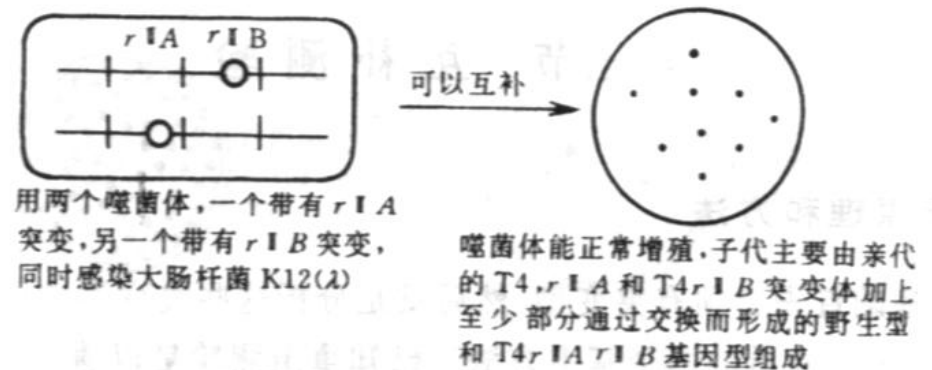
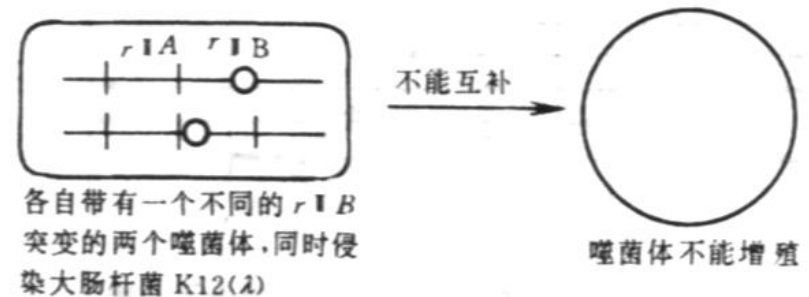
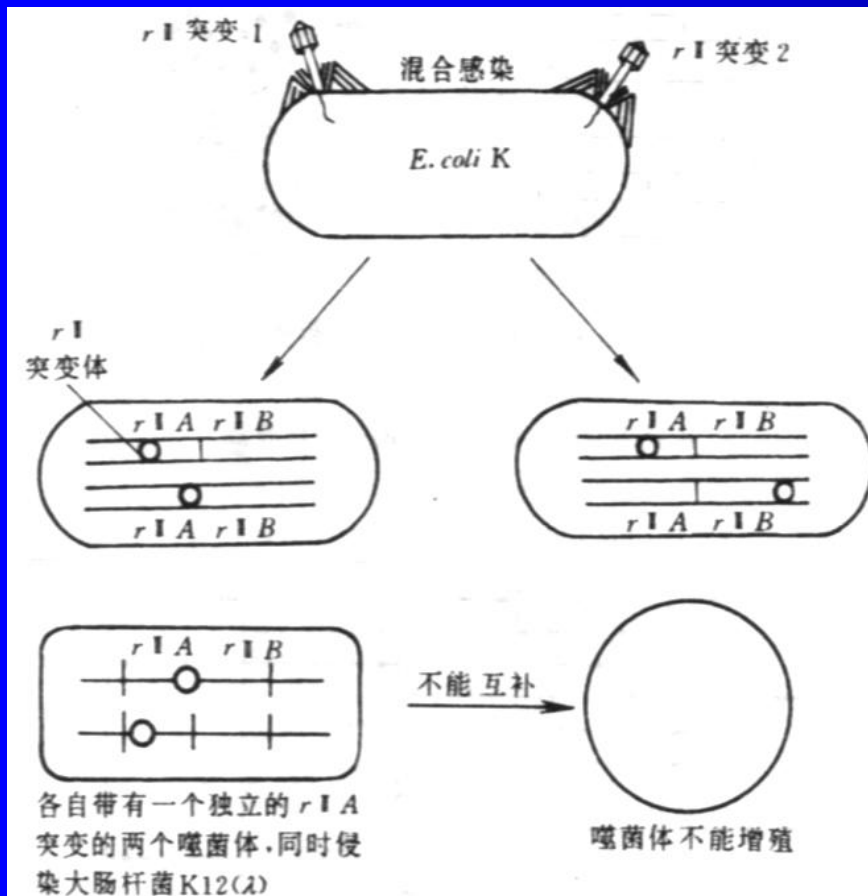
互补



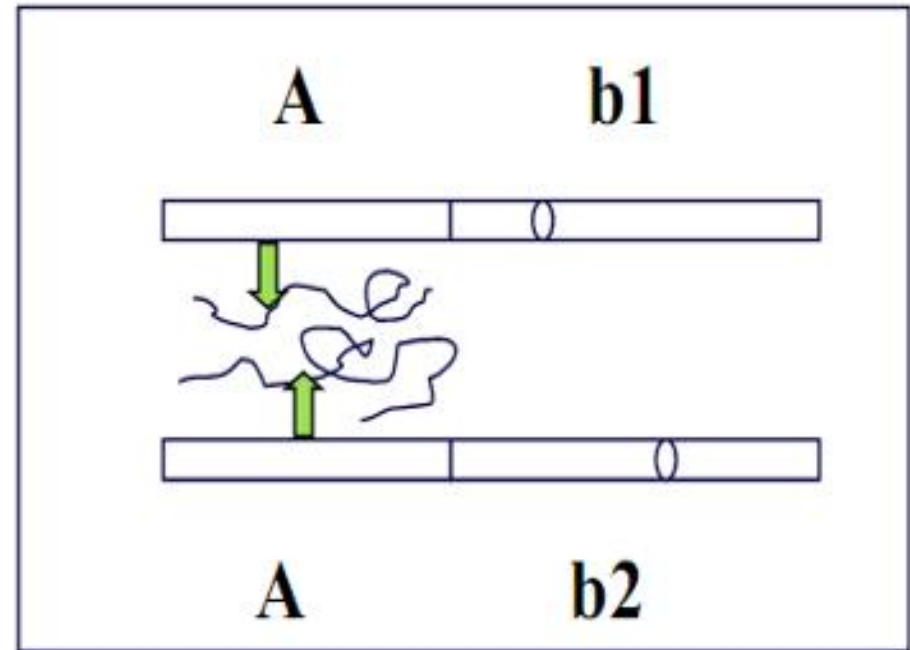
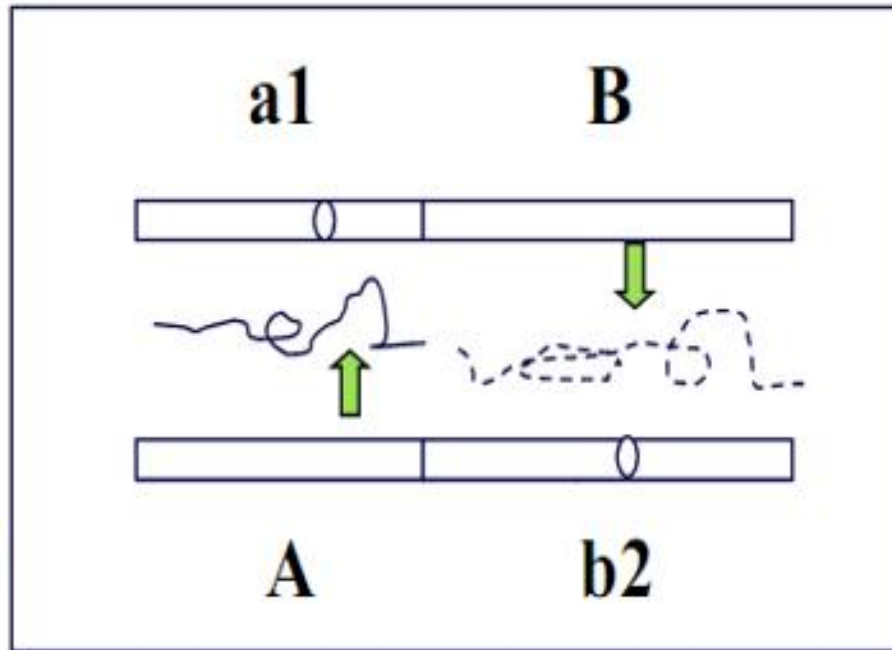
许多大的噬菌斑

基因的顺反子概念

- 1955年, **Benzer**的互补实验



功能互补测试体系

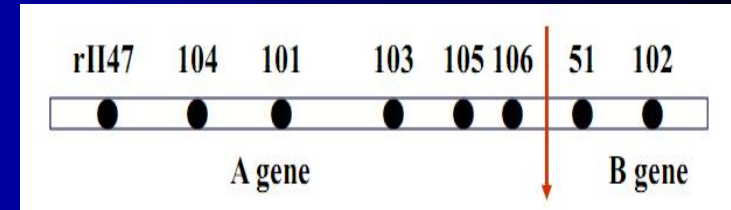


野生型

突变体

- 顺反互补型突变: 2个独立的功能区
- 顺反非互补型突变: 1个独立的功能区—基因

顺反子

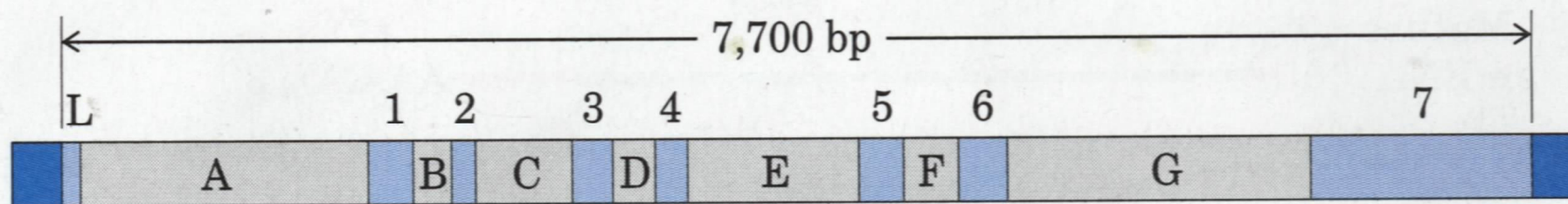
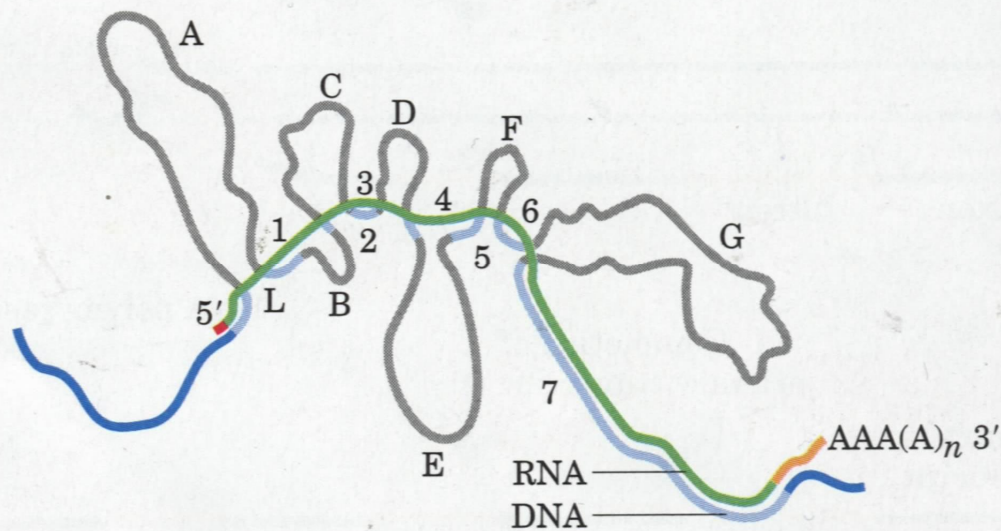
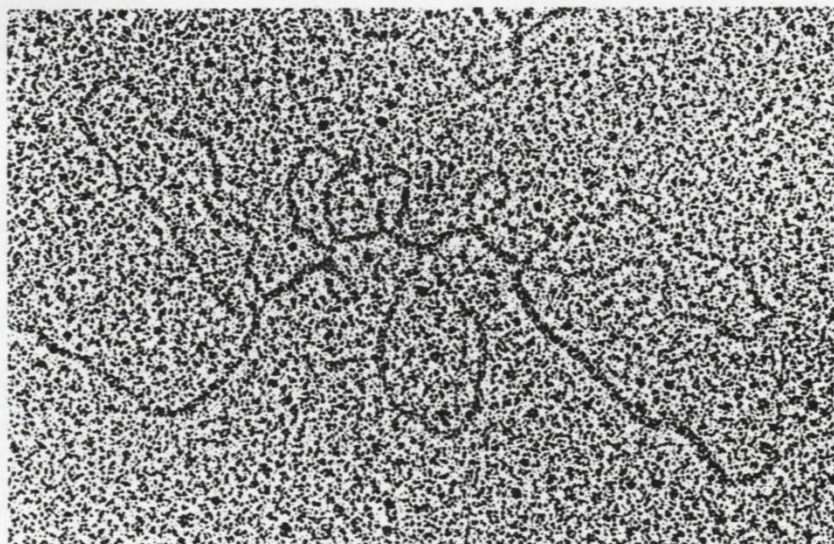


- 顺反子：通过互补实验定义
- 不同的亚区域可以发生互补，而同一亚区域不能发生互补，所以rIIA和rIIB为两个不同的功能单位，这样的亚区域称为**顺反子**：
- 一段核苷酸序列，能编码一段完整的多肽链
- **是一个有功能、可以重组、可以突变的遗传单位**
- 三位一体：基因是最小的功能单位, 最小的重组单位, 最小的突变单位

DNA与cDNA杂交

- 1977年：鸡管状细胞DNA与红细胞DNA分别用HindIII和EcoRI酶切，与卵清蛋白cDNA杂交
- 结果两种组织中均有3条杂交条带
- 但cDNA中无酶切位点？
- cDNA的酶切位点与 ?? 一致





内含子检测

限制性内切酶图谱

DNA分子杂交

Splitting gene

间隔基因，断裂基因

Interrupted gene

真核生物基因的转录物又称为

Precursor mRNA (pre-mRNA)

Heterogeneous nuclear RNA (HnRNA)

Introns and exons:

Eukaryote pre-mRNAs often have intervening introns that must be removed during RNA processing.

intron = non-coding DNA sequences between exons in a gene.

exon = expressed DNA sequences in a gene, code for amino acids.



1993: Richard Roberts (New England Biolabs) & Phillip Sharp (MIT)

真核基因概念的发展

基因概念及发展

断裂基因：内含子可以将基因分割

1977年：Gilbert预言真核基因是外显子重组产生

基因与酶 → 基因与多肽链（顺反子） → 外显子组合

基因(gene)

合成一种蛋白或有功能的RNA分子所需要的DNA序列

有功能的DNA？

- Splitting gene 概念的相对性

- a) Intron 并非“含而不露”

- Yeast Intron II of cyt.b gene coding a maturase

- b) Exon 并非“表里如一”

- 人类尿激酶原基因 Exon I 不编码 氨基酸序列

- c) 并非真核生物所有的结构基因均为splitting gene

- Histone gene family

- Interferon gene

- Yeast 中多数基因 (ADH...)

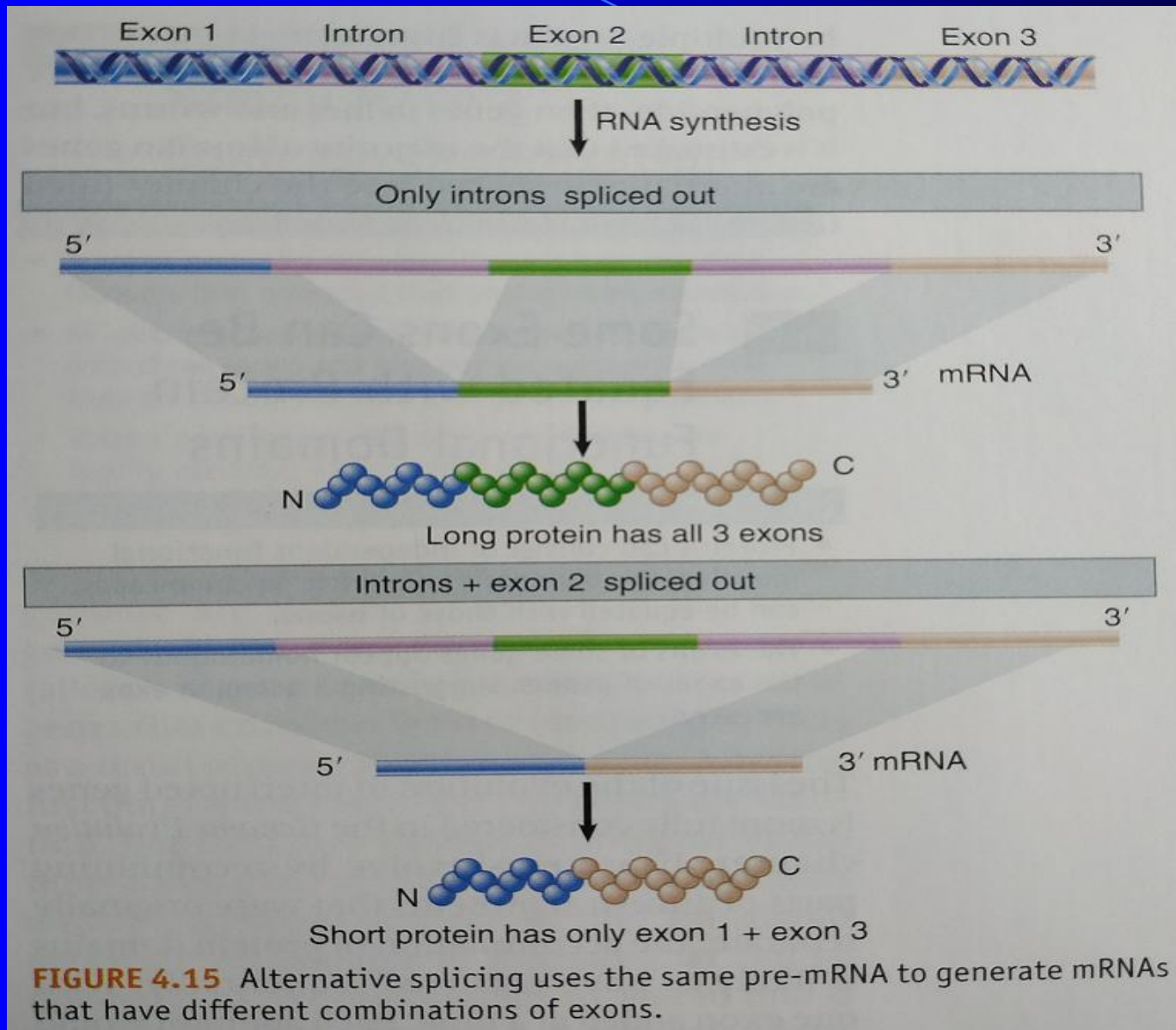
- (果蝇 ADH 基因为间隔基因)

- } 不是splitting gene

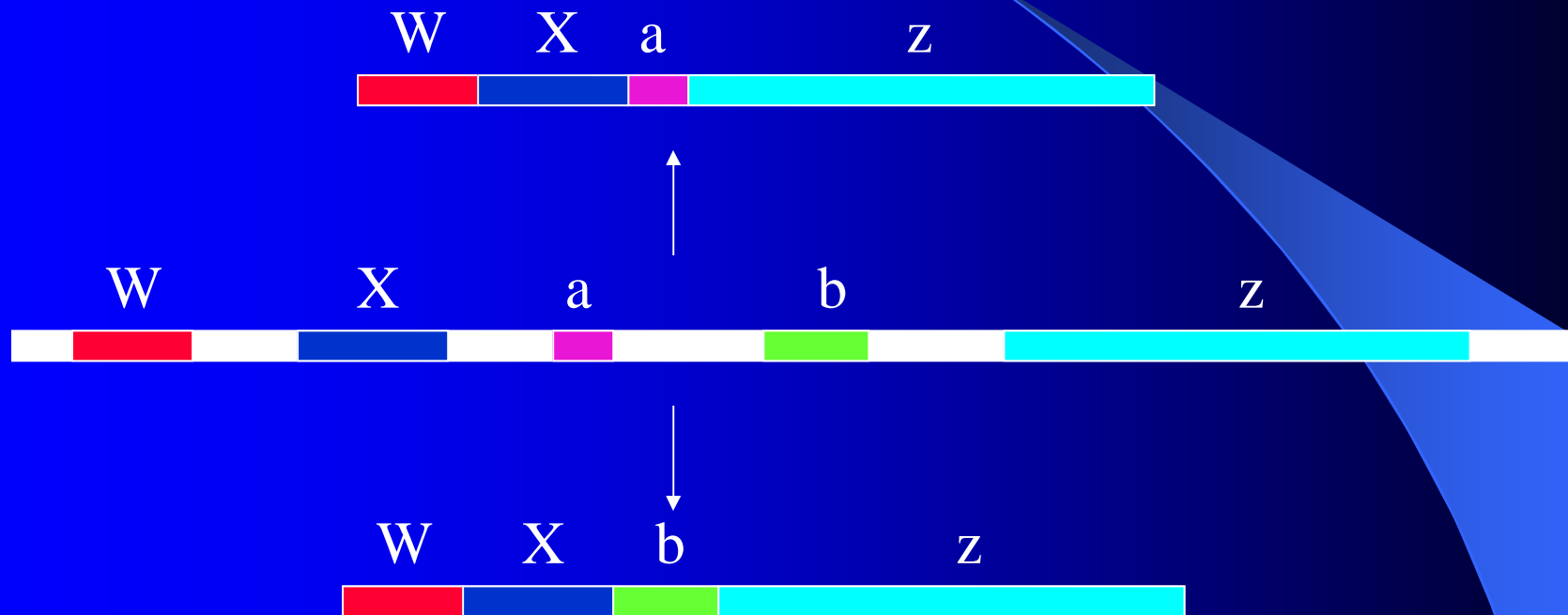
无内含子的基因

- 约有687个，约占蛋白质基因（2万2百多个）总数的3%
- 第一类：编码的蛋白质参与信号转导，调节细胞的生长、增殖和发育。这一类IGs约占53%，其中大多数属于与G蛋白偶联的受体（GPCR，大约50%的GPCR无内含子）
- 第二类：编码各种组蛋白
- 第三类：编码各种干扰素编码基因，热休克蛋白等

为何存在高度耗能的剪接过程???

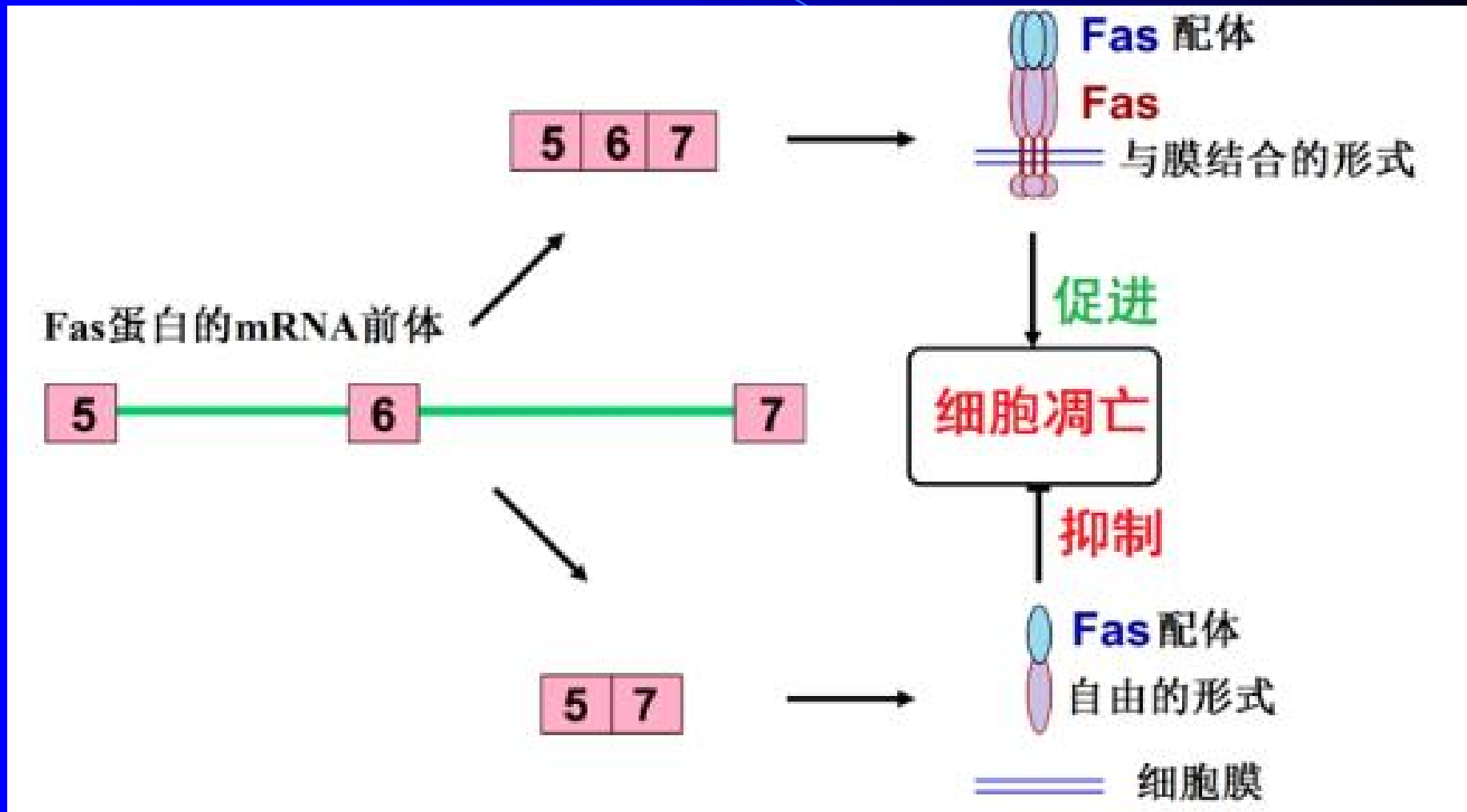


■ 拼接方式不同（选择性剪接）



两种拼接方式分别产生肌钙蛋白a, b

Fas通过选择性剪接的两种变体具有相反的功能



保留第6个外显子，促进细胞凋亡；切除第6个外显子，抗凋亡

外显子 1 | 内含子 | 外显子 2

外显子 3 | 内含子 | 外显子 4

外显子 1 | 外显子 2

外显子 3 | 外显子 4

内含子中引入互补序列后

外显子 1 | 内含子 | 外显子 2

外显子 3 | 内含子 | 外显子 4

外显子 1 | 外显子 2

外显子 3 | 外显子 4

顺式剪接

外显子 1 | 外显子 2

外显子 3 | 外显子 4

外显子 1 | 外显子 4

外显子 3 | 外显子 2

反式剪接

基因组定义

- 基因组(genome)

生物体的整套(单倍体)遗传物质的总和

- 基因组的大小

用全部DNA的碱基对总数表示

- 人类基因组大小

30亿bp

各类生物最小基因数

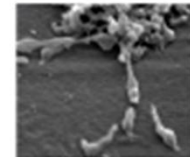
- 各类生物最小基因数随复杂度增加而增加
- 几种真核生物基因总数：

酵母	6000
线虫	18500
果蝇	13600
拟南芥	25000
小鼠	30000
人	20000
模式生物	

Minimum gene numbers range from 500 to 30,000

500 genes

Extracellular (parasitic)
bacterium



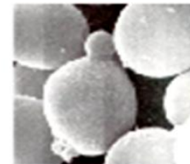
1,500 genes

Free-living bacterium



5,000 genes

Unicellular eukaryote



13,000 genes

Multicellular eukaryote



25,000 genes

Higher plants



30,000 genes

Mammals



C值矛盾

■ C值（大C值）

单倍体基因的全部DNA含量

小C值：编码基因信息的DNA含量

■ C值矛盾 (C-value Paradox)

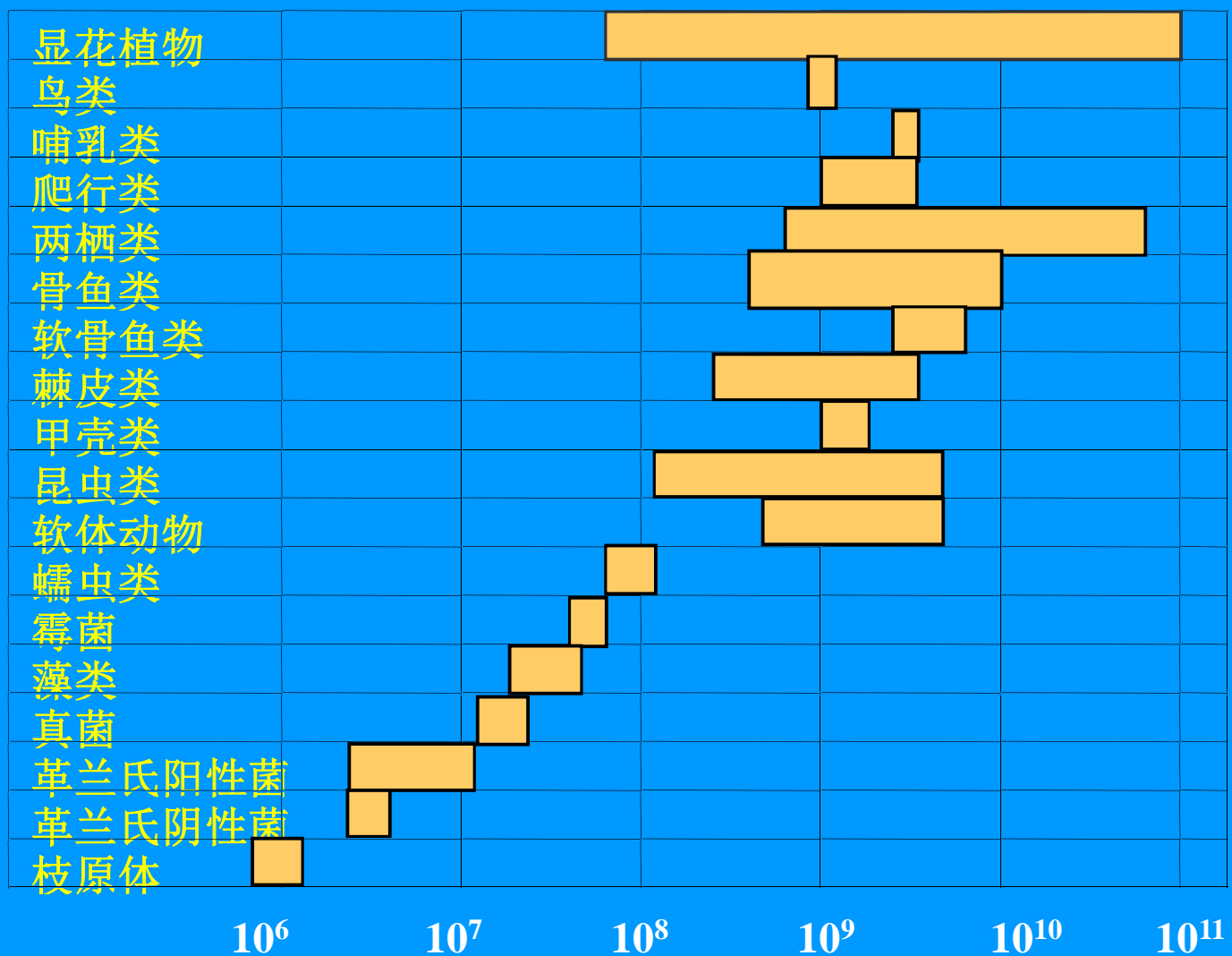
生物基因组大小与其进化地位

复杂性不相关

种类	Mb
大肠杆菌	4.64
啤酒酵母	12.1
线虫	100
果蝇	140
蝗虫	5000
小鼠	3300
豌豆	4800
玉米	5000
小麦	17000
人	3000

某些物种与预期相比, C值明显过大

同一物种, 不仅C值相差很大, 大小C值也相差很大

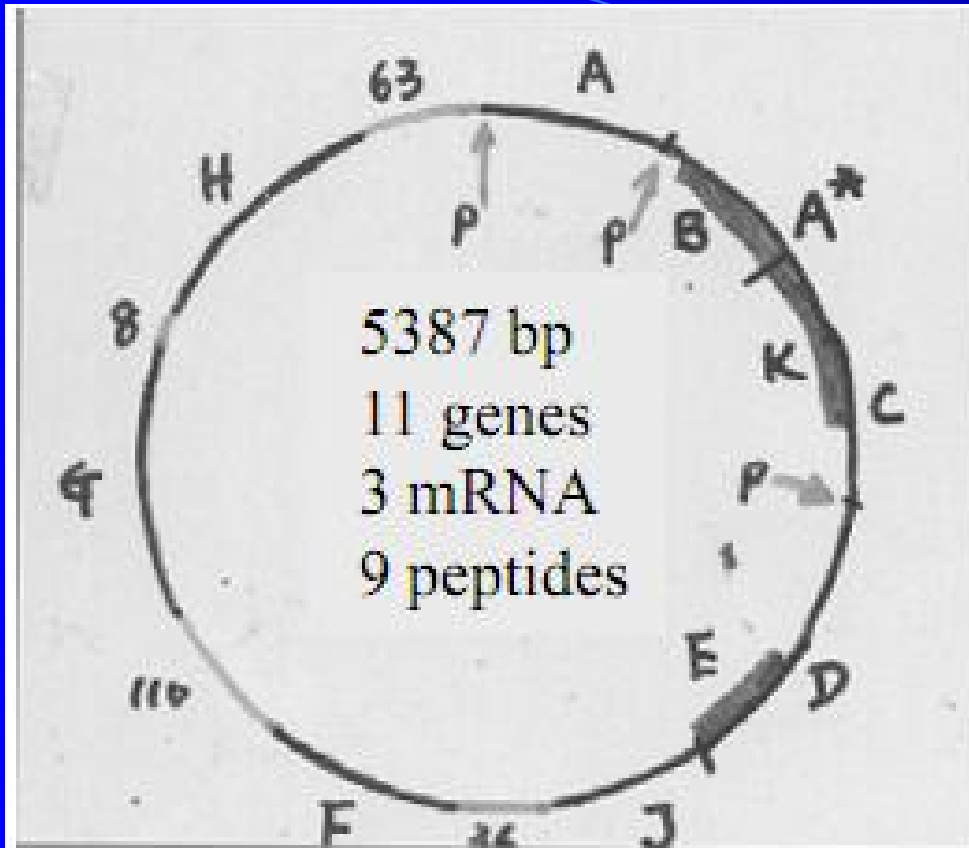


不同种类生物基因组C值分布



C值矛盾续-基因组序列特征

- 原核生物基因组中基本上不含重复序列；
- 低等真核生物基因组中，重复序列的组成不超过20%，且多半是中度重复序列；
- 动物细胞的基因组中，中度和高度重复序列约占50%；
- 在一些显花植物和两栖类基因组中，中度和高度重复序列几乎可以高达80%



ΦX174 (F. Sanger, 1977)

$C = 5387 \text{ bp}$

$c = 11 \times 2000 \text{ bp}$

高等生物小C比大C小很多
低等生物则很可能反过来

判断以下描述是否正确？

- 基因组大小与遗传复杂性之间有必然关系
- 生物越复杂，所需要的最小基因组越大
- 两栖类动物C值变化很大，说明它们基因数相差很大
- 基因数目可以从基因组总的大小来推测
- 低等生物小C可以比大C大
- 小C和大C都不能解释生物的复杂性

细菌基因

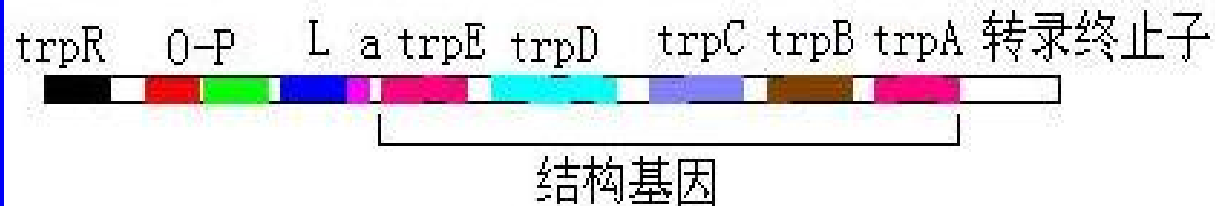
■ 操纵子结构:

功能相关的几个结构基因往往串联在一起，受上游共同调控区控制



乳糖操纵子结构

与乳糖代谢相关的3个酶的基因串联排列在一起，受同一个启动子调节控制。



色氨酸操纵子结构 原核基因组织

细菌基因特点

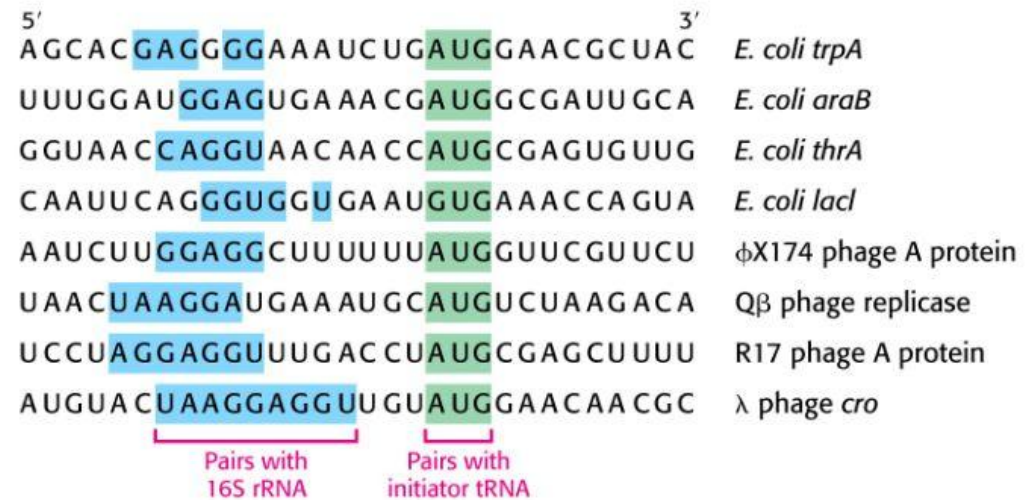
- 结构基因中没有内含子，也无重叠现象

- 细菌DNA大部分为编

码序列

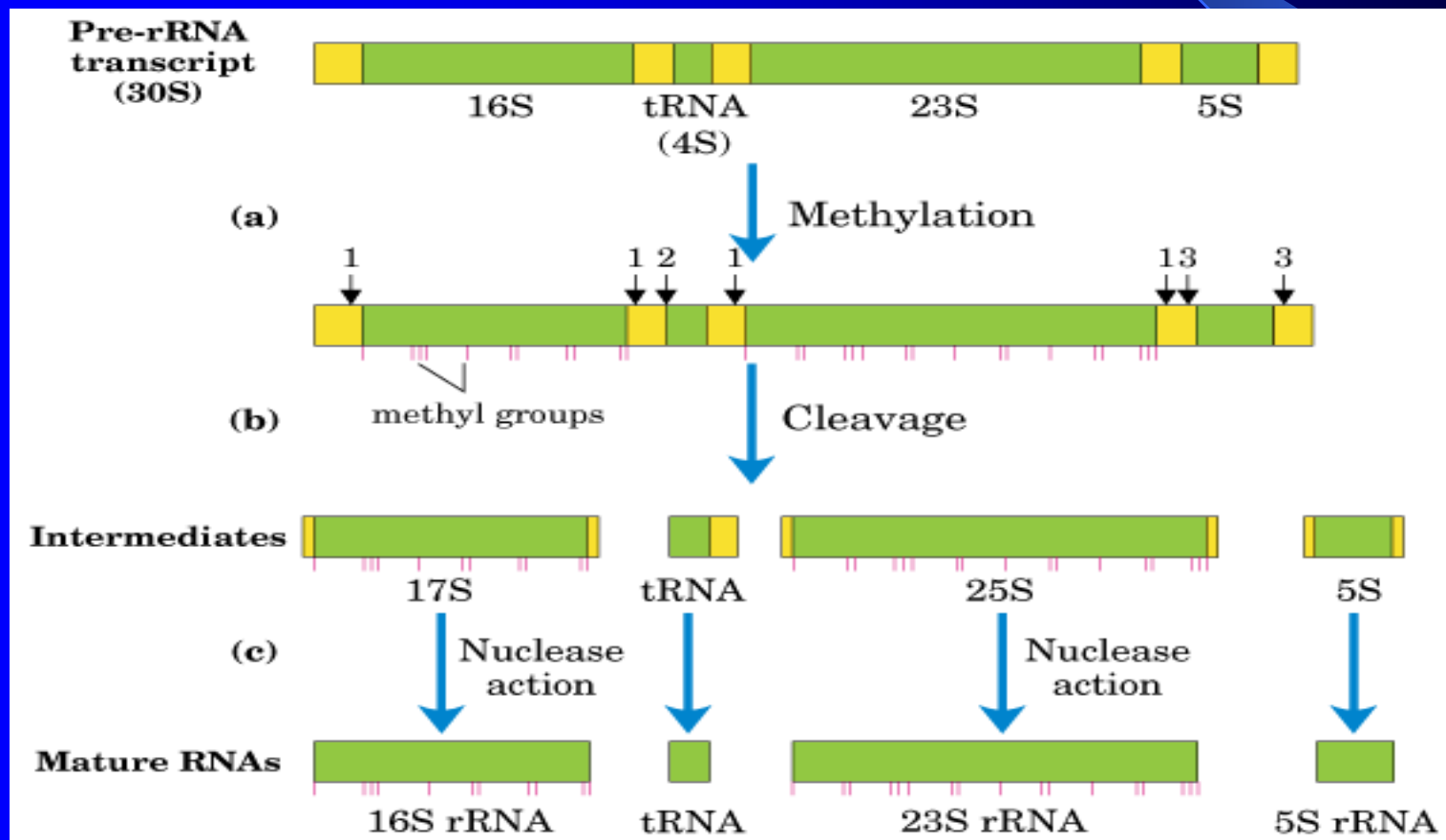
- SD序列

在翻译起始AT (U) G
前3--12bp处有一个
保守序列称SD顺序，
是核糖体RNA结合区



细菌基因特点续

- rRNA基因往往是多拷贝的，蛋白质基因为单拷贝
有利于核糖体的快速组装（15AA/秒，2AA/秒）



病毒基因特点

- 原核生物和真核生物概念
- 每种病毒只有一种核酸：DNA，或RNA
大小仅为一般细菌1%左右
- 除逆转录病毒外，所有病毒基因都是单拷贝
- 真核病毒基因有内含子-类似于真核生物
- 含复制、转录等必需基因 -机器为宿主

病毒基因特点续

■ 有重叠基因

1973: RNA病毒QB噬菌体

1977: Sanger第二次Nobel奖 (1958 insulin)

ϕ X174

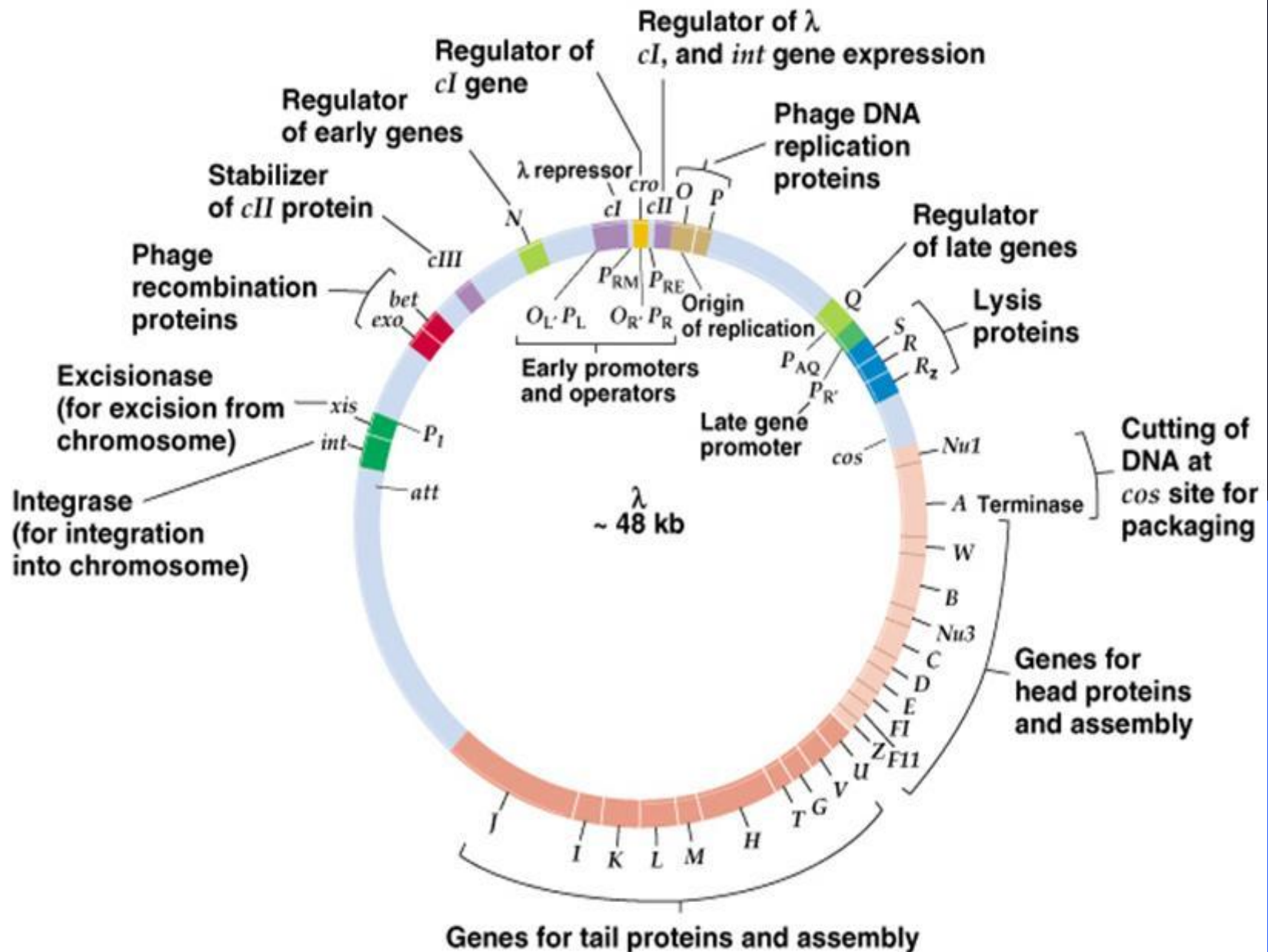
重叠方式多样化



重叠基因 (overlapping gene)

- 莲人在绿杨津
采 一
玉漱声歌新阙
- 采莲人在绿杨津，在绿杨津一阙新；
一阙新歌声漱玉，歌声漱玉采莲人。

λ 噬菌体



线粒体基因组

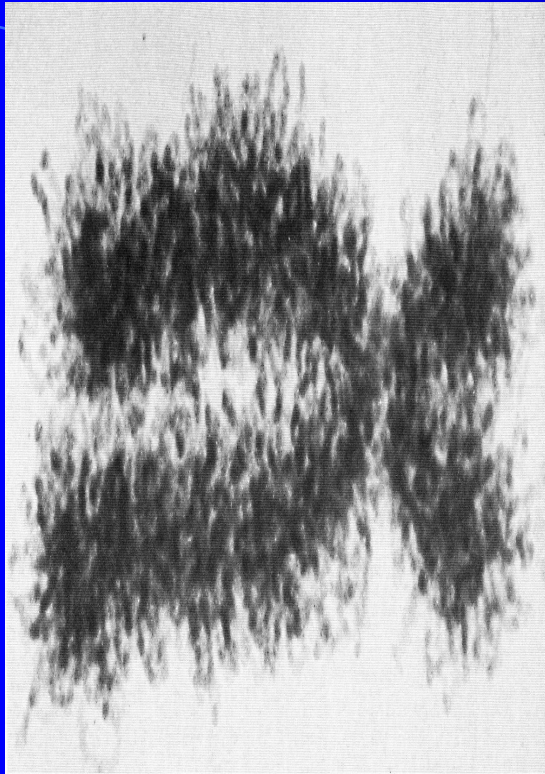
人线粒体基因组为16,569bp的双链闭环分子

- 一条链为重链(H链)，一条链为轻链(L链)
两条链均有编码功能
- 部分mtDNA的密码子不同于核内DNA的密码子
- 结构基因不含内含子，部分区域有基因重叠
任何突变都会累及到基因组中重要功能区
- mtDNA突变频率更高（无修复系统）
- 有些遗传病如Leber遗传性视神经病，肌阵挛性癫痫等与线粒体基因突变有关

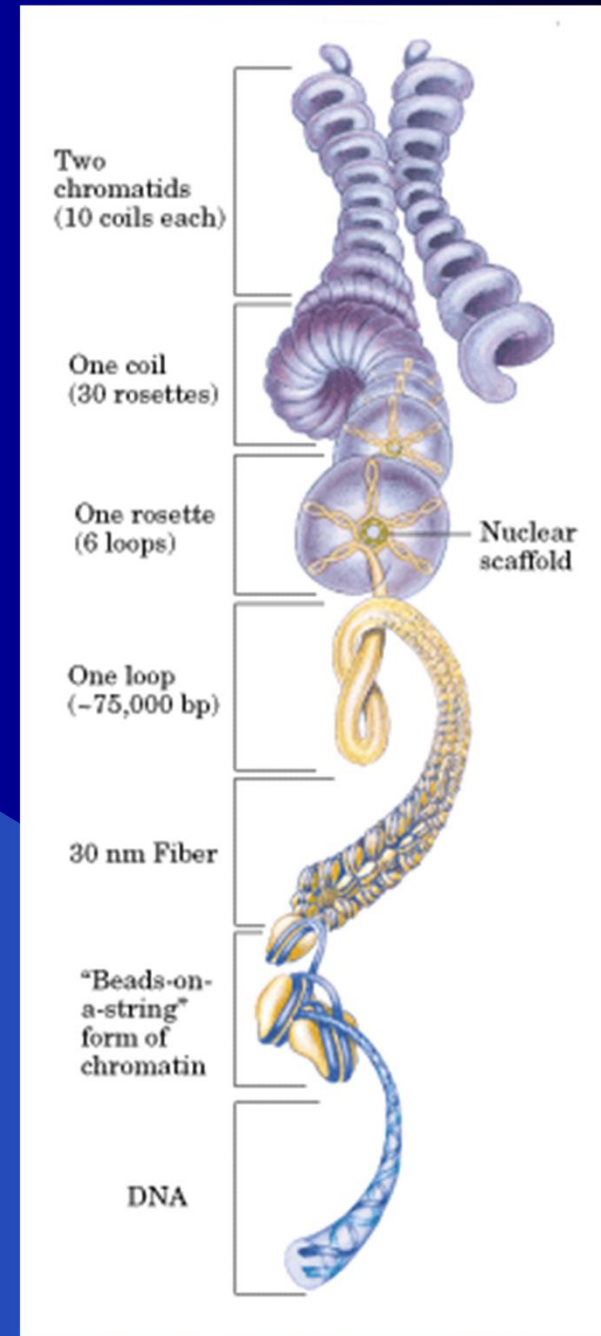
线粒体基因

- 一条链为重链(H链)，一条链为轻链(L链)
两条链均有编码功能
- UGA不是终止密码子，而是色氨酸的密码子
- AGA、AGG不是精氨酸的密码子而是终止密码子
- 加上通用密码UAA和UAG，线粒体共有4个终止密码子
- 内部甲硫氨酸密码子有2个，即AUG和AUA
- 起始甲硫氨酸密码子有4个，即AUN
- 线粒体DNA为母系遗传

染色质结构

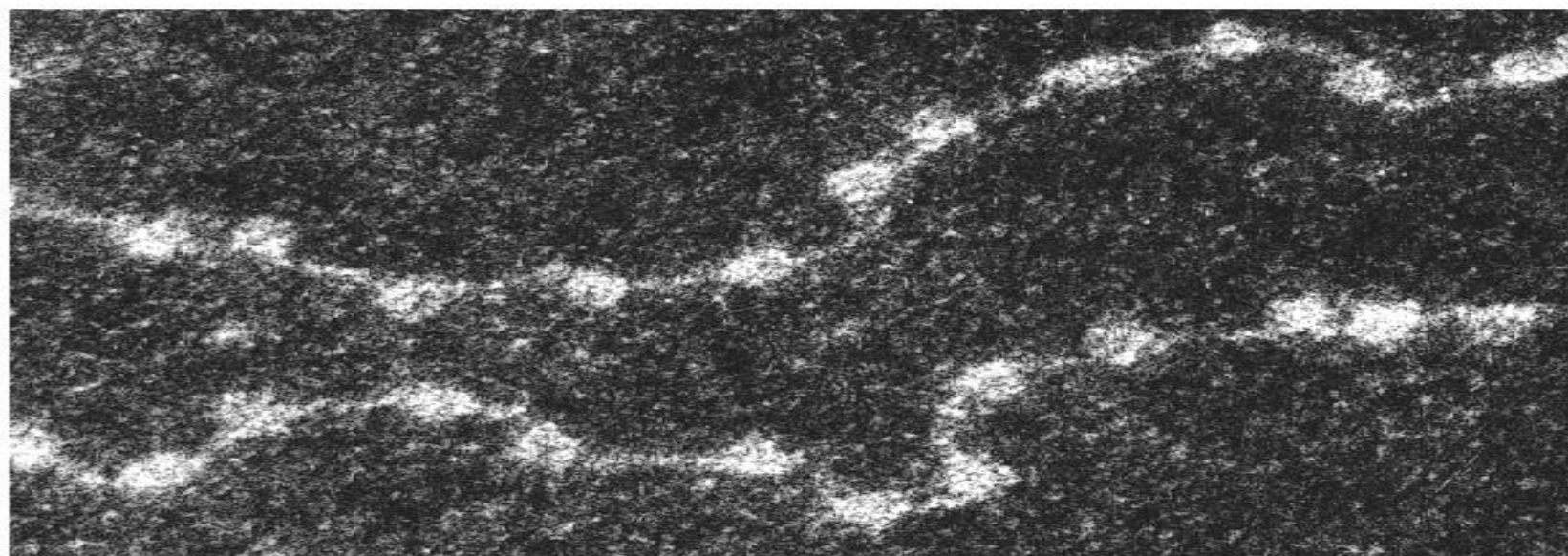
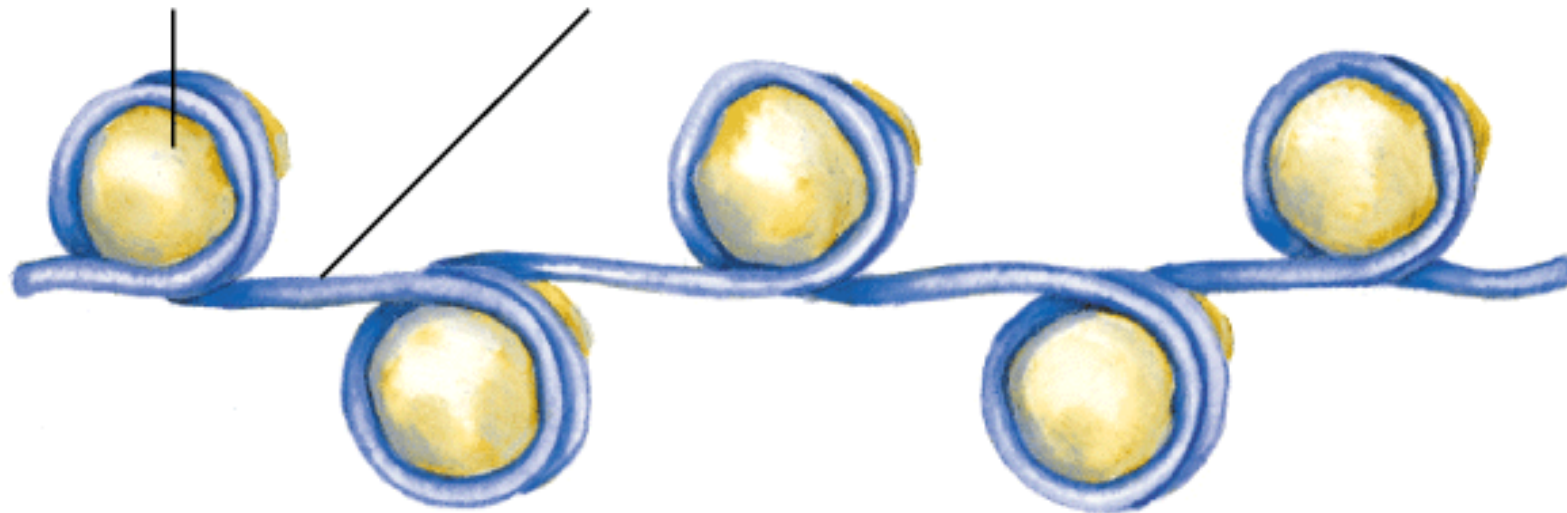


DNA length: 4.6×10^7 bp = 1.5 cm
Chromosome length: $2 \mu\text{m}$



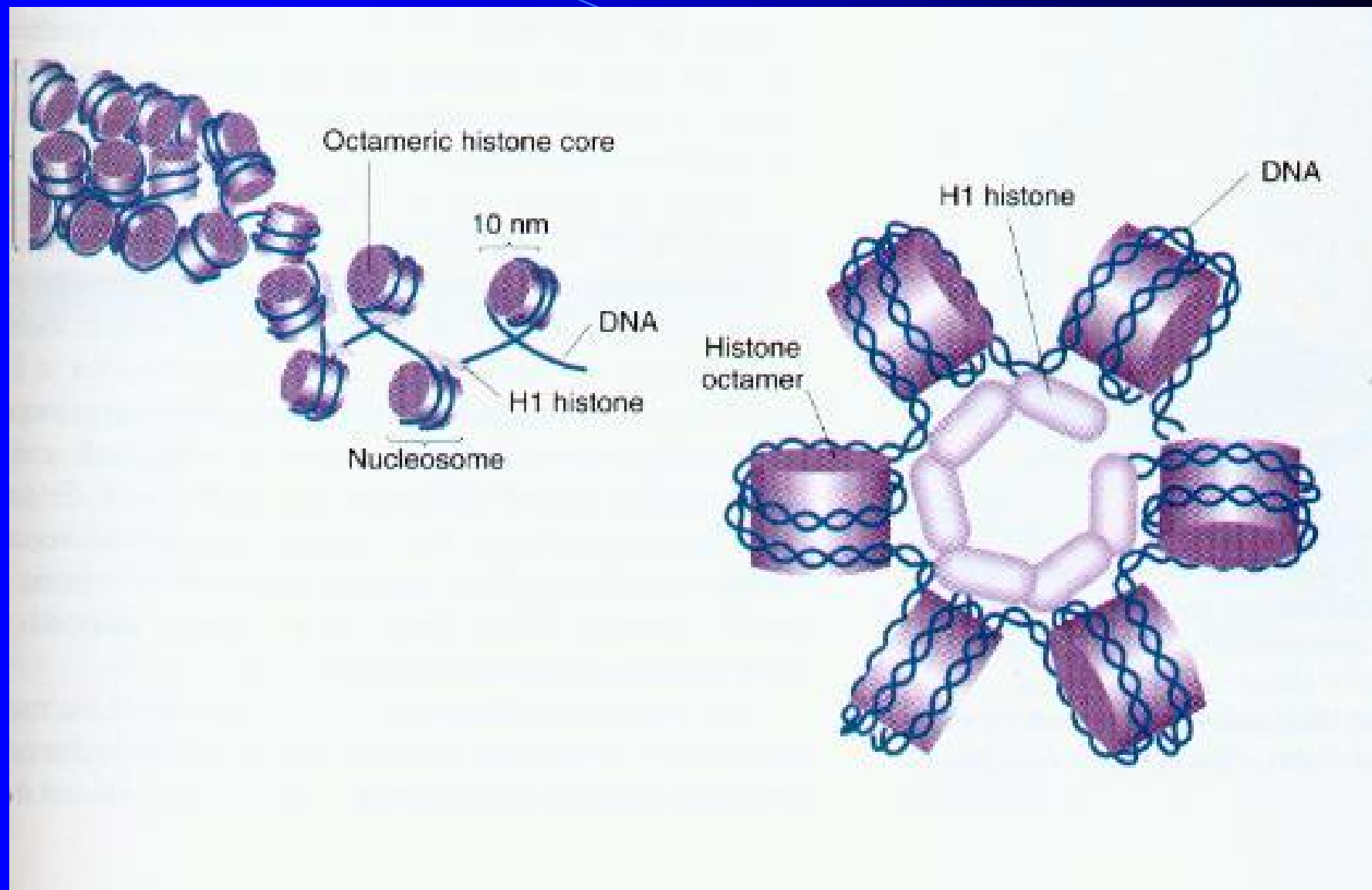
Histone core
of nucleosome

Linker DNA
of nucleosome



(b)

50 nm



真核生物基因组

■ 核小体(nucleosome)结构

DNA分子一般有100厘米左右（人类）

DNA分子核小体链螺线管结构

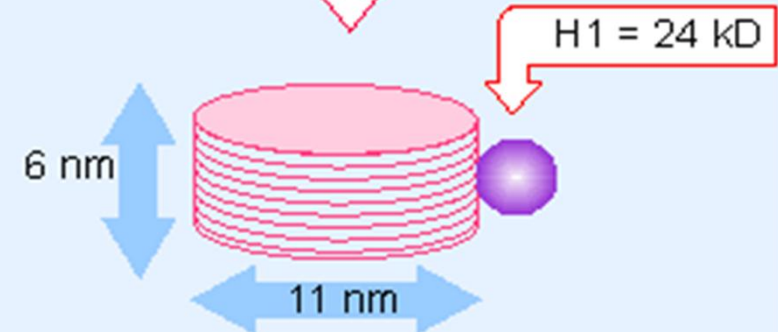
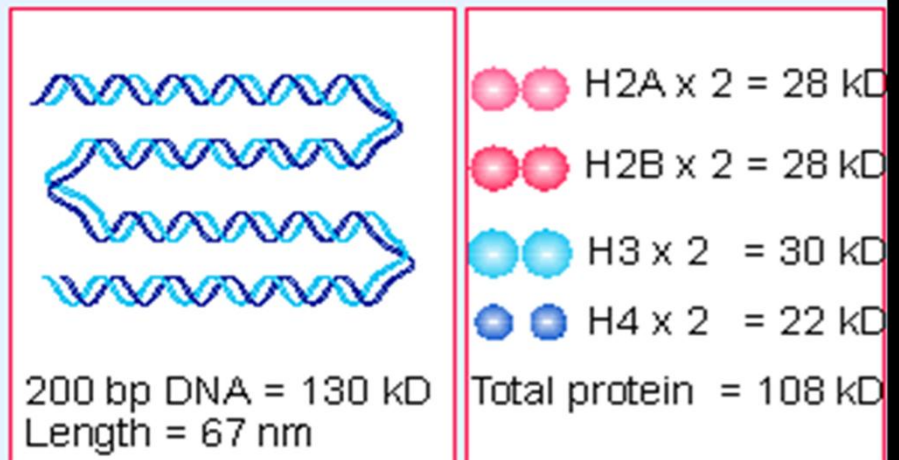
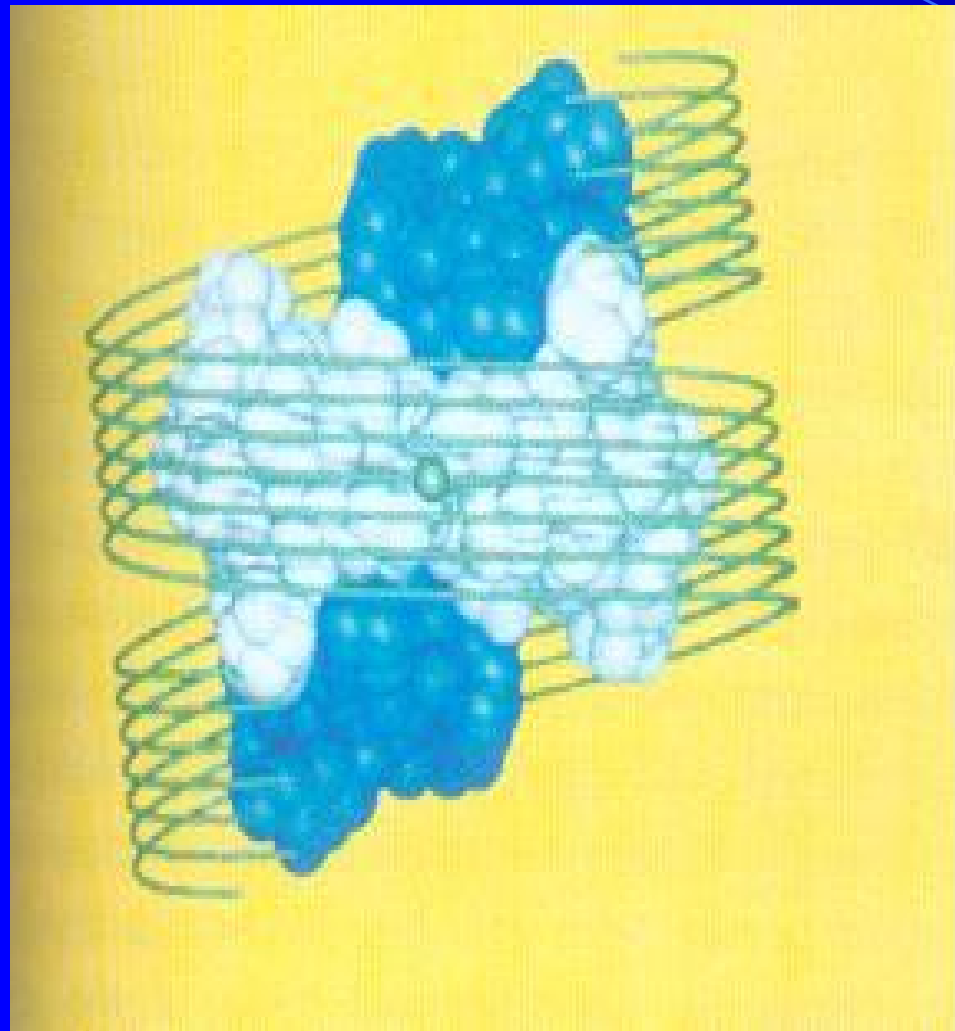
—Kornberg模型：

DNA绕在组蛋白八聚体(H2A、H2B、H3、H4各一对)

核心外1.8周(146bp), 形成核小体核心颗粒

两个核心颗粒之间有Linker (H1蛋白) DNA(0-80bp),

核小体核心颗粒+Linker=核小体DNA(长180-210bp)



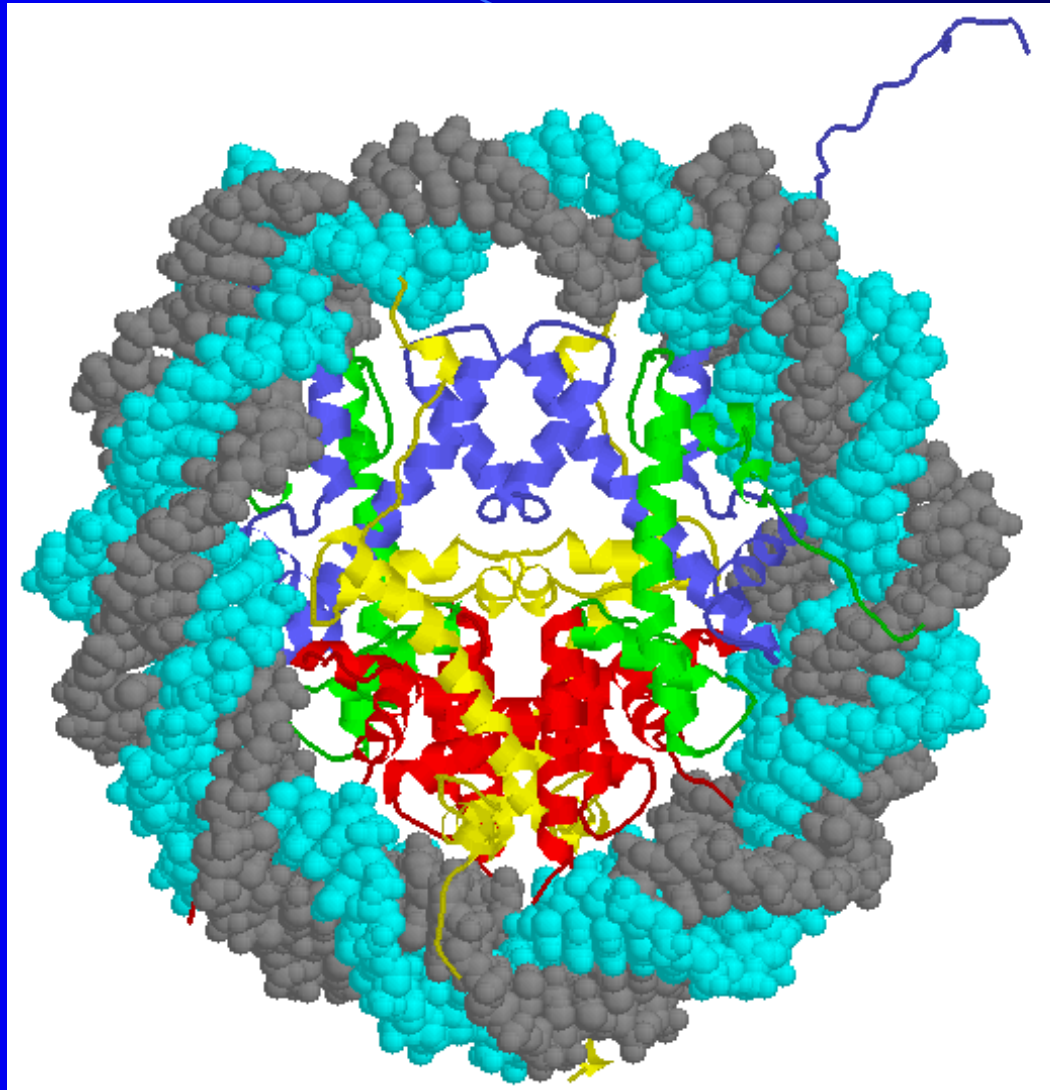
Diameter = 110 Å

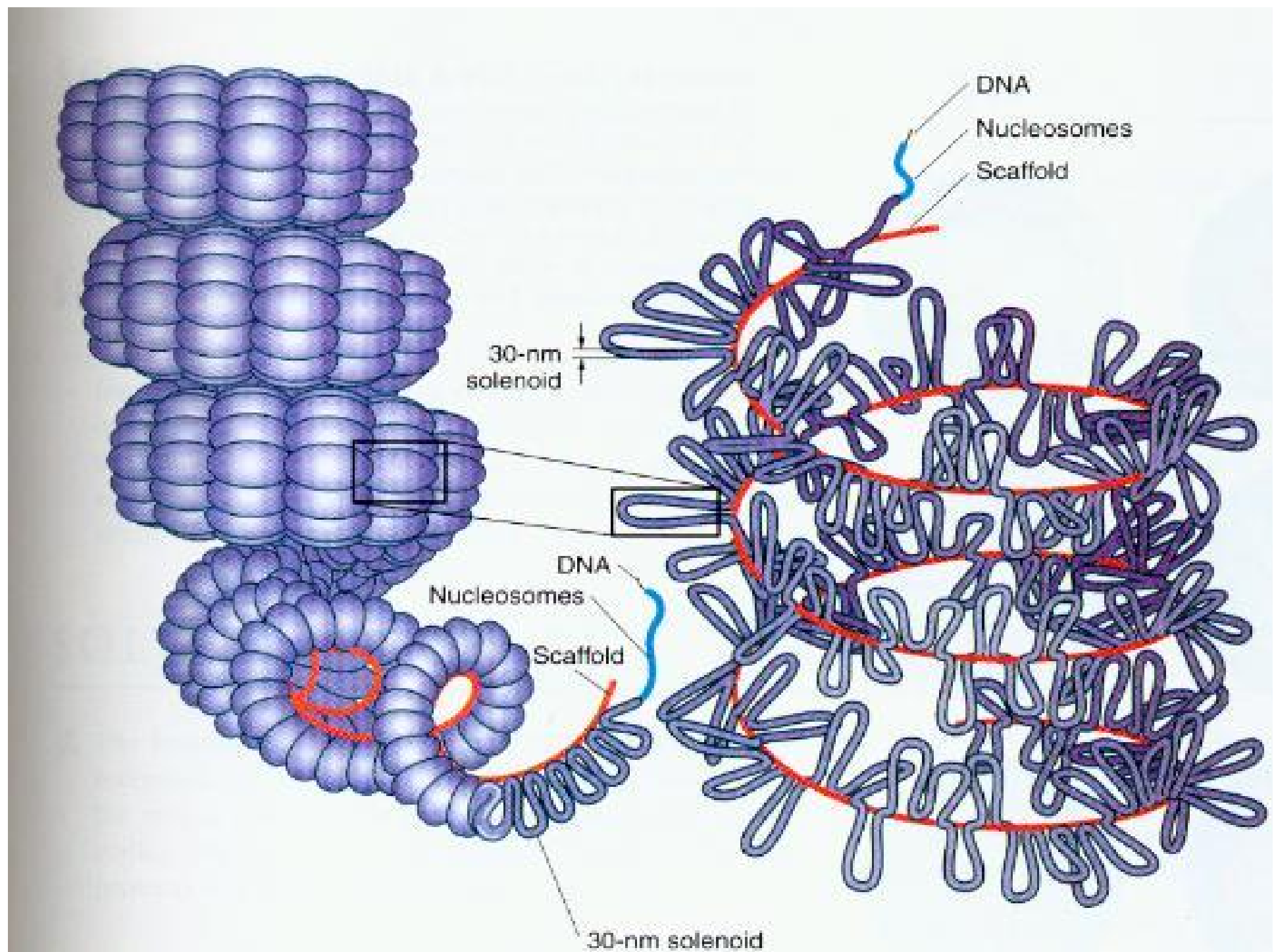
H3

H4

H2A

H2B





组蛋白性质

- 组蛋白(hi stone)

含丰富正电荷（富含Lys, Arg)的蛋白

- 核心组蛋白: H2A, H2B, H3, H4

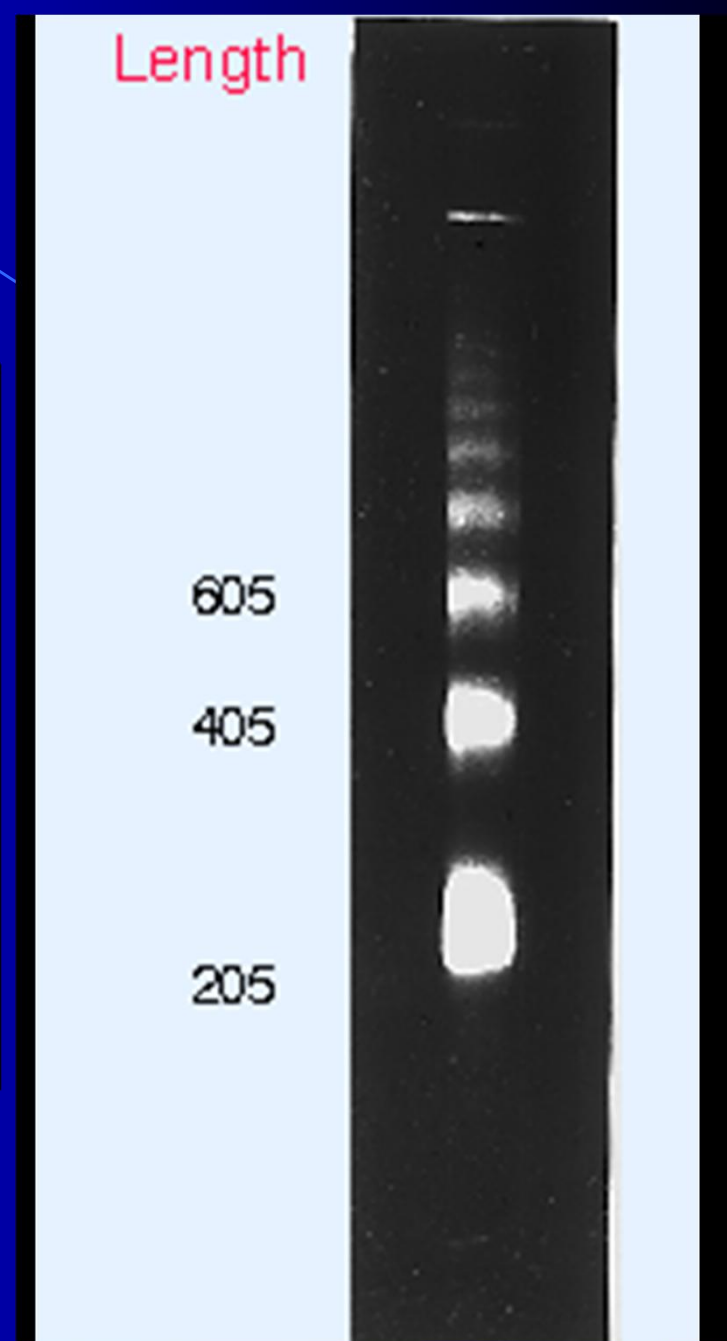
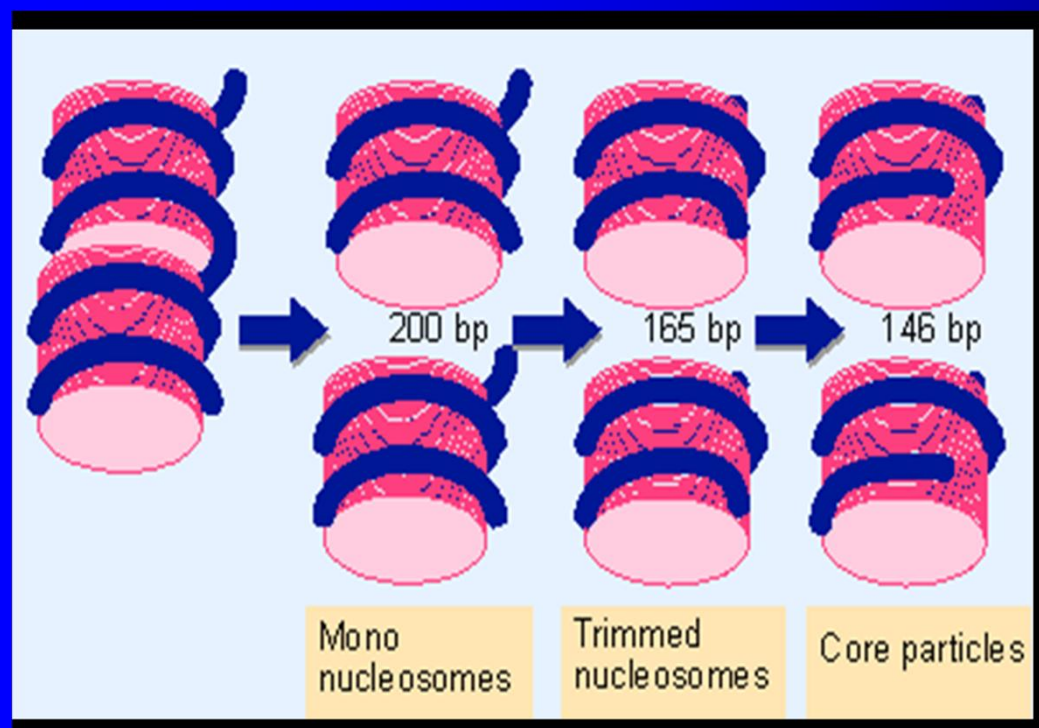
分子量 较小(102-135aa)，盘绕DNA形成核小体

- H1组蛋白：分子量较大(220aa)，作为Li nker

有利于核小体稳定和更高级结构的形成

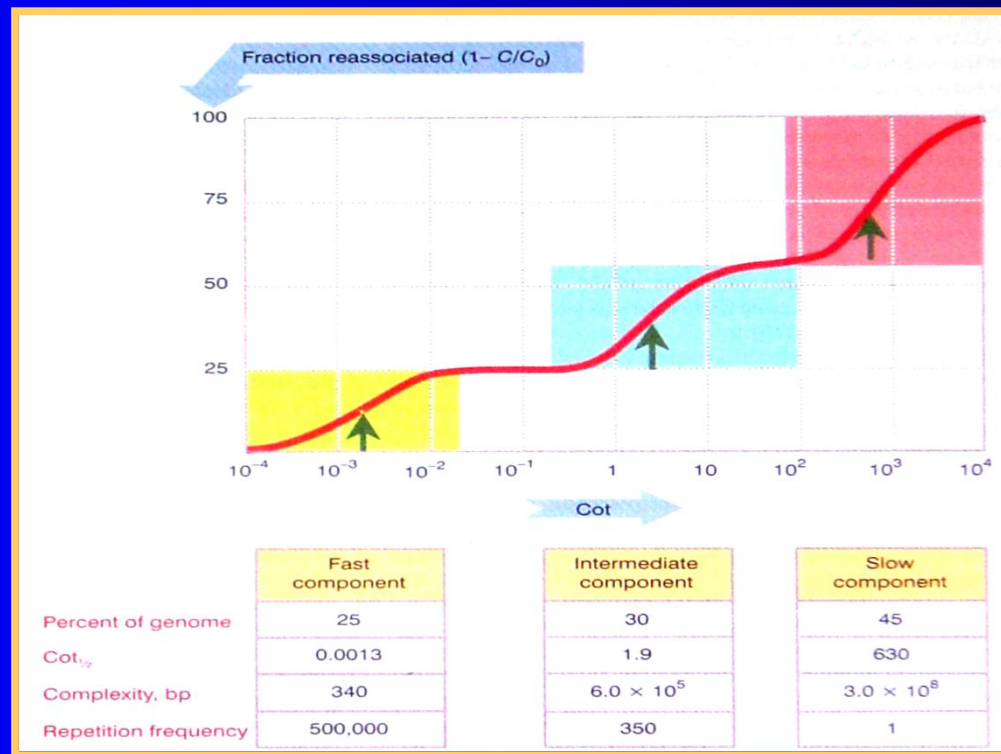
- 进化上极端保守性

- 无组织特异性



真核生物基因组序列特征*

- 具有多个复制起始位点.
- 编码序列仅占一小部分(1%), 大部分为非编码序列.



真核基因组中DNA序列的分类

- 慢复性组分

非重复序列 单拷贝序列 基因

- 中间复性组分

中度重复序列 调控作用

- 快复性组分

高度重复序列 一般不转录

- 不同生物中重复序列比例差异很大

1 高度重复序列

■ 卫星DNA (Satellite DNA)

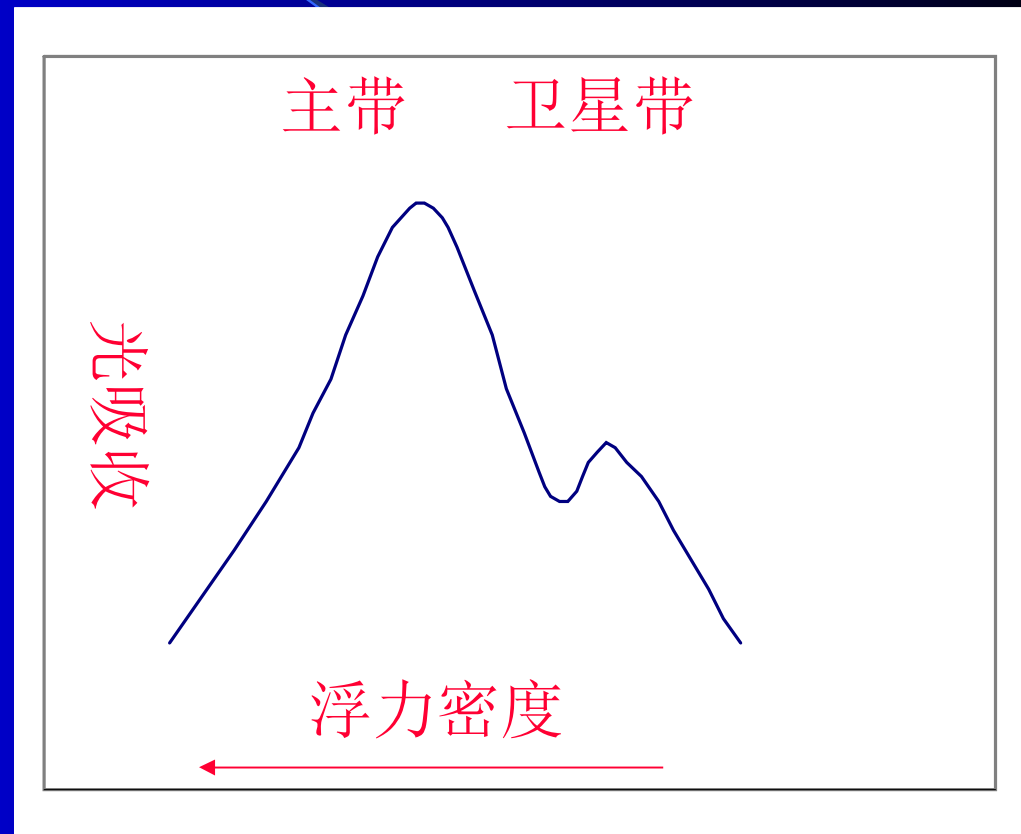
■ 隐蔽卫星DNA

DNA剪切成若干片段后

离心一般得到1个峰

而某些DNA在主峰旁边还有
1个小峰，其中所含DNA称

sat-DNA



Satellite DNAs often lie in heterochromatin

Cytological
hybridization shows
that mouse satellite
DNA is located at
the centromeres.



卫星 DNA在着丝粒处定位暗示着可能与染色体的分离过程有关

高度重复序列续

小卫星DNA(mini satellite DNA)

由短重复单位(6-40bp)串联重复(6-100次以上)而成，位于基因的非编码区，广泛分布。

不同个体串联重复的数目和位置存在差异*

符合孟德尔遗传规律 -DNA指纹图(fingerprint)。

微卫星DNA(microsatellite DNA)

短串联重复(short tandem repeat, STR)

2-6个核苷酸组成的重复单位串联重复(10-60次)

人基因组中每10kb DNA序列至少一个STR序列

短串联重复序列

(Short Tandem Repeats, STR)

- 广泛分布在人类染色体上
- 每个重复单位为2到6核苷酸

例如: tetra-nucleotides (AGAT)_n
penta-nucleotides (TTTTC)_n

- DNA位点的多态性可以作为遗传标记，用于基因诊断或个体鉴定，其与疾病的相关研究正在进之中。
- 短串联重复序列长度为一百到几百个核苷酸长度
新一代遗传标记，人类基因组研究，肿瘤，遗传病

短串联重复序列

复等位基因的不同在于重复单位数目的不同（群体分布）

Polymorphism of STR Locus D5S818

Forward Primer

5' GGGTGATTTCCTCTTTGGT [AGAT]_n TGTGGCTATGATTGGAATCA 3' Reverse Primer

Allele7

→ **AGATAGATAGATAGATAGATAGAT** ←

Allele8

→ AGATAGATAGATAGATAGATAGATAGAT ←

• • •

Allele16

→ AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT ←

高度重复序列（**micro-satellite**）

不编码基因（功能？）

无选择压力

（**multiple allele** 可保留在群体中）

（有效的分子标记，

SSR simple sequence repeat）

中度重复序列

❖ 中度重复序列的特点

散布于基因组中

序列的长度和拷贝数非常不均一

中度重复序列一般具有种属特异性，可作为DNA标记

中度重复序列可能是转座元件

rDNA tDNA

Alu family

Histone gene cluster

中度重复顺序

- 轻度重复顺序

在基因组中含有2-10拷贝，

酵母tRNA基因、人和小鼠的珠蛋白基因等

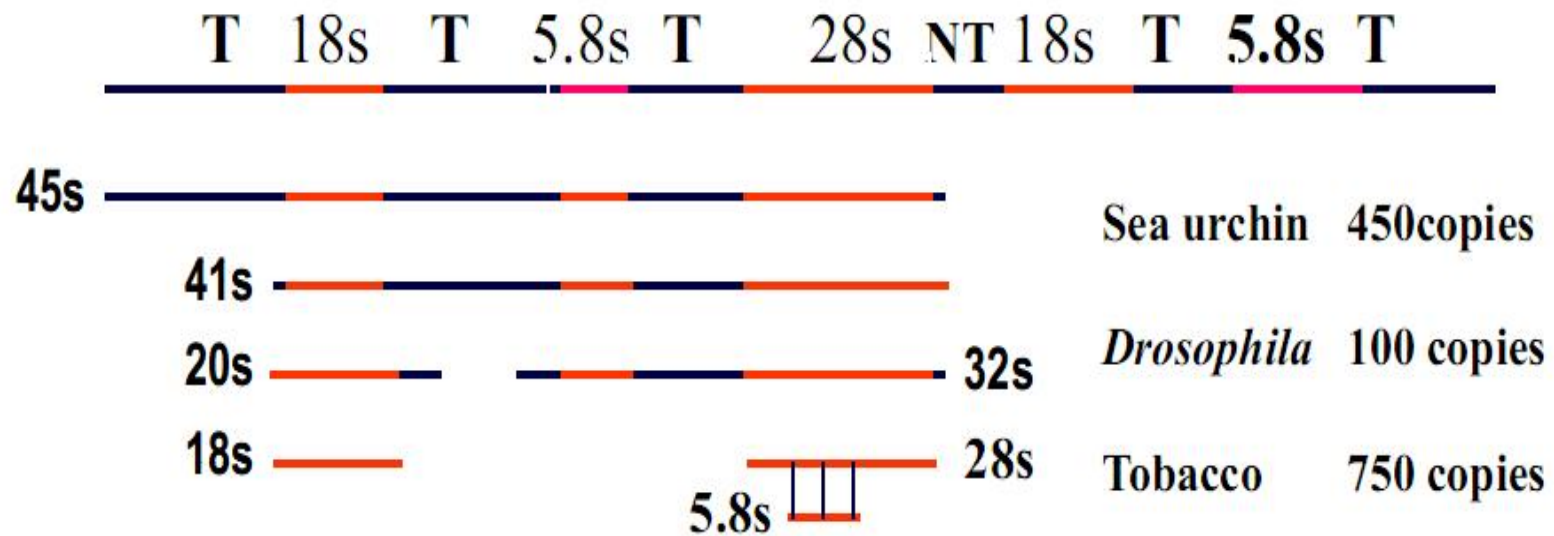
- 中度重复顺序

长约300bp

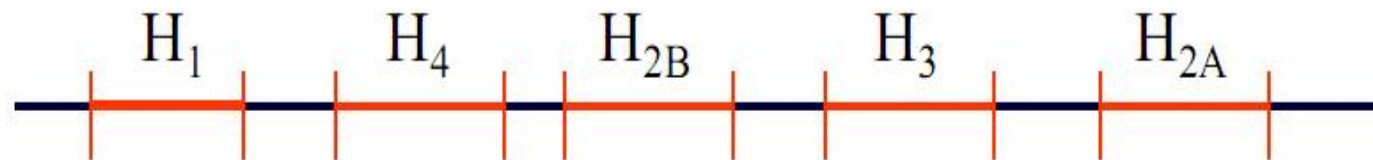
基因组中约有10-几千个拷贝的顺序

如rRNA和tRNA基因

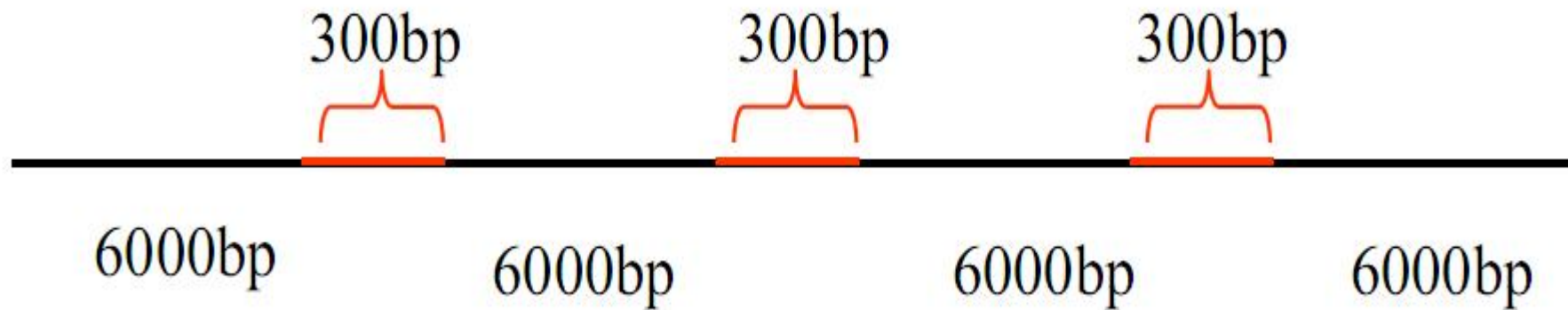
rDNA gene family



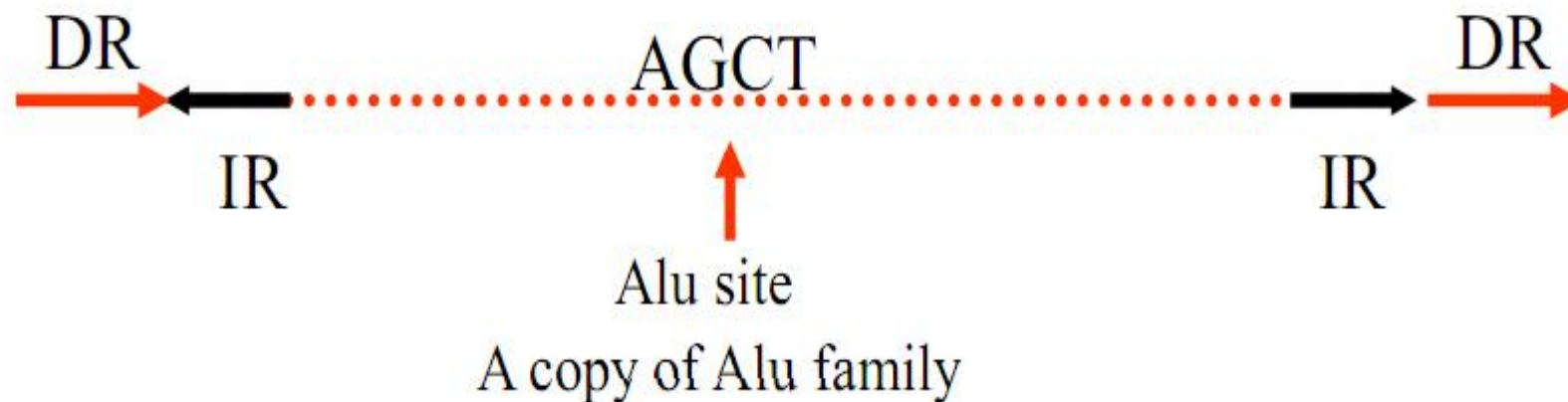
Histone gene family



真核生物的 Alu family



300,000 copies 广泛分布于非重复序列间



Alu重复序列

- ❖ 单倍体人基因组50万-100万拷贝，平均每隔3-6Kb就有一个Alu序列 散布
- ❖ Alu序列广泛散布于人基因组，约90%已克隆的人基因含有Alu序列
- ❖ 两侧7-21bp正向重复(direct repeats)，转座子？
Alu序列长300bp
- ❖ AG ↓ CT
- ❖ **SINEs**(short interspersed nuclear elements)

LINES(long interspersed nuclear elements)

- ❖ 序列中含Kpn I 限制酶位点
- ❖ 全长6-7kb
- ❖ Knp I 消化后可见4条带
- ❖ 集中分布在染色体G或Q带

.....

人类基因组大部分是重复DNA序列

- 重复序列占人类基因组50%以上
- 重复序列中有约45%为转座子
- 染色体上约10—25%区域为“沙漠化”
- 分布：

外显子	1%
内含子	24%
重复序列	8%
转座子	45%
基因间的序列	

人类基因数目比预计的要少

- 每条基因平均总长27Kb，由9个外显子组成，平均长约1400bp
- 与黑猩猩序列相似性超过99%
- 已鉴定约20000条基因，约90%基因为可变剪接
- 不同组织的比较显示： 哺乳动物中管家基因 (housekeeping) 约10000条
- “新基因从何而来？” 一些新基因是由非编码DNA以比预想更快的速度生成 (Science) 期刊报导

单拷贝序列

- 大多数编码蛋白质的结构基因和基因间隔序列

- 5' 端为GT, 3' 端为AG, GT/AG 规则

—— GT —— AG ——

- | | | |
|------|--------|-----|
| 外显子 | 内含子 | 外显子 |
| Exon | Intron | |



真核基因结构特征

- 绝大多数真核基因是断裂基因(split-gene)
- 编码序列—外显子(exon)
- 插入外显子之间的非编码序列—内含子(intron)
- 5' -端和3' -端非翻译区(UTR)
- 调控序列



蛋白质编码基因的精细结构

功能多样性

- 氨酰-tRNA合成酶(Aminoacyl-tRNA synthetase)是一类起源非常古老的酶家族，广泛存在
- 该酶家族在生命起源和进化过程中位于至关重要的位置，参与了由以核糖核酸为主导的生物界到以蛋白质为主导的生物界的转变
- 原本功能应该相当保守、分布普遍的生物酶在漫长的进化过程中获得了许多新的生物学功能，形成了复杂的信号调控网络

一种蛋白可以行使多种生物功能

➤许多新的细胞生物学功能：

tRNA的运输、核糖体RNA的合成加工、RNA剪切、运输、转录和翻译调控、细胞凋亡、炎症反应、血管生成等

➤追踪AARS的进化途径：研究自然界不同物种之间的进化关系

➤利用原核生物、古细菌和真核生物AARS结构和功能的显著差别，以AARS为靶点设计抗菌和抗肿瘤药物

基因家族



■ 基因家族(gene family)

一组功能相似且核苷酸序列具有同源性的基因。

可能由某一共同祖先经多次重复和突变产生

■ 排列方式

基因家族的成员可以串联排列在一起，形成基因簇：

如rRNA、tRNA和组蛋白的基因

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

典型的基因家族



■ tRNA基因

人类基因约有1300个tRNA基因，编码50多种tRNA。每种tRNA可有10-几百个基因拷贝。同种tRNA往往串联在一起形成基因簇，但基因间有非转录间隔区分隔，常常比结构基因长近10倍

• rRNA基因 （总RNA80%-90%）

酵母: 140个copy. E. coli: 7次

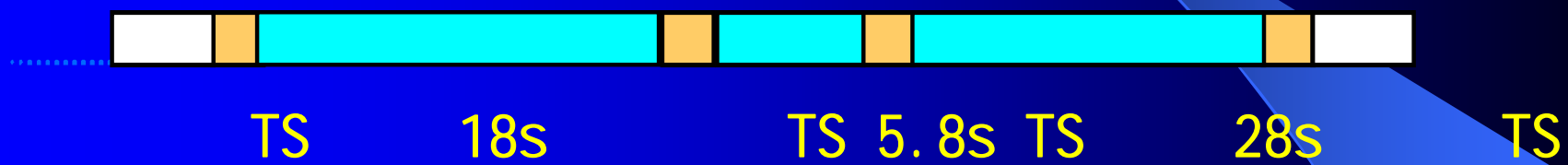
果蝇: 130—250次.

人类: 300次.



NTS

NTS



非洲爪蟾 rRNA 基因的串联重复排列

假基因 (Pseudogene)

- 假基因 类似于基因但不表达的DNA序列
不表现任何功能，是基因的退化形式
- 假基因常用符号 ψ 表示，如 $\psi \alpha_1$ 表示与 α_1 相似的假基因
- 假基因与有功能的基因同源，原来也可以是有功能的基因，由于发生缺失、移位或突变等
- 哺乳动物基因组中的1/4基因为假基因，可能为进化的痕迹??

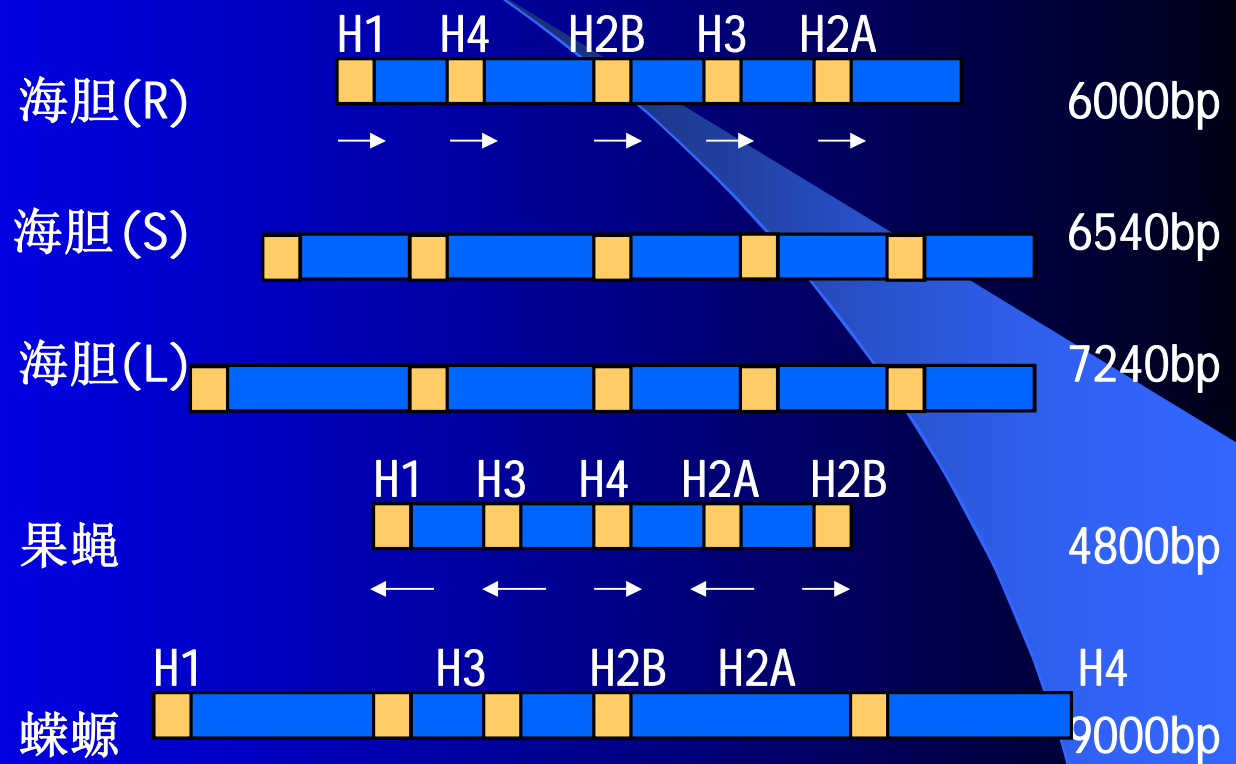
组蛋白基因家族——核苷酸序列高度同源

■ 重复单位

H1, H2A, H2B,
H3、H4

■ 特点

无intron



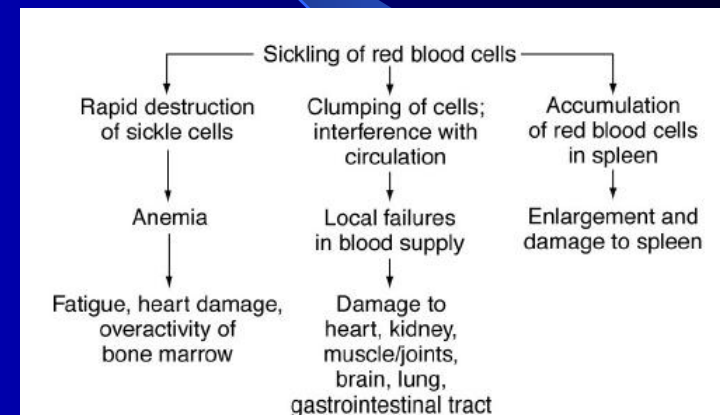
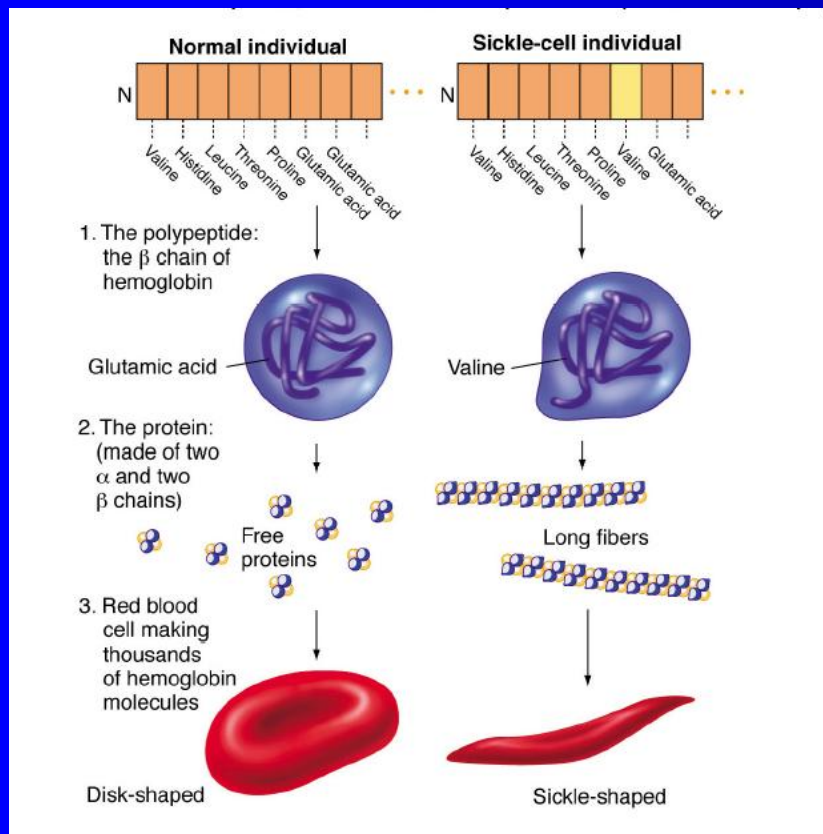
组蛋白基因簇的重复单位

生长激素基因家族——核苷酸序列高度同源

- 人生长激素（hGh）、人胎盘促乳素和催乳素之间同源性很高，尤其是hGh和hcs之间，蛋白质氨基酸序列有85%的同源性，mRNA上序列上有92%的同源性，说明它们是来自一个共同祖先基因
- 3种基因并不都排列在一起，hGh和hcs基因位于第17号染色体长臂，催乳素基因位于第6号染色体

One Gene, One Protein

镰刀形细胞贫血症：正常HbA四条多肽链（2条 α 链，两条 β 链）
Vernon M. Ingram证明 β 链第六位氨基酸谷氨酸突变为缬氨酸



(c) β -chain substitutions/variants

	Amino acid position													
	1	2	3	...	6	7	...	26	...	63	...	67	...	146
Normal (HbA)	Val	His	Leu		Glu	Glu		Glu		His		Val		His
HbS	Val	His	Leu		Val	Glu		Glu		His		Val		His
HbC	Val	His	Leu		Lys	Glu		Glu		His		Val		His
HbG San Jose	Val	His	Leu		Glu	Gly		Glu		His		Val		His
HbE	Val	His	Leu		Glu	Glu		Lys		His		Val		His
HbM Saskatoon	Val	His	Leu		Glu	Glu		Glu		Tyr		Val		His
Hb Zurich	Val	His	Leu		Glu	Glu		Glu		Arg		Val		His
HbM Milwaukee 1	Val	His	Leu		Glu	Glu		Glu		His		Glu		His
HbD β Punjab	Val	His	Leu		Glu	Glu		Glu		His		Val		Gln

证明基因与氨基酸之间存在直接对应关系的第一个直接证据

珠蛋白基因

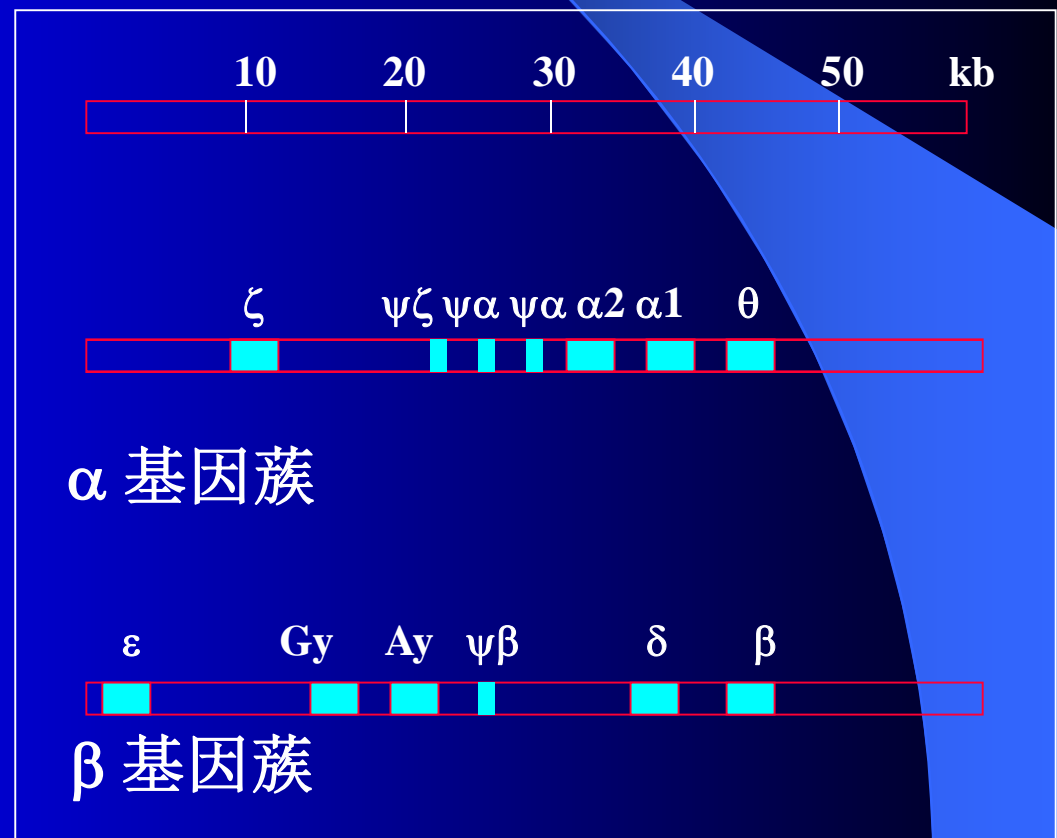
- 基因家族各个成员在DNA分子上的排列顺序按照发育不同阶段先后次序排列，也称“发育控制复合多基因家族”

- α 类

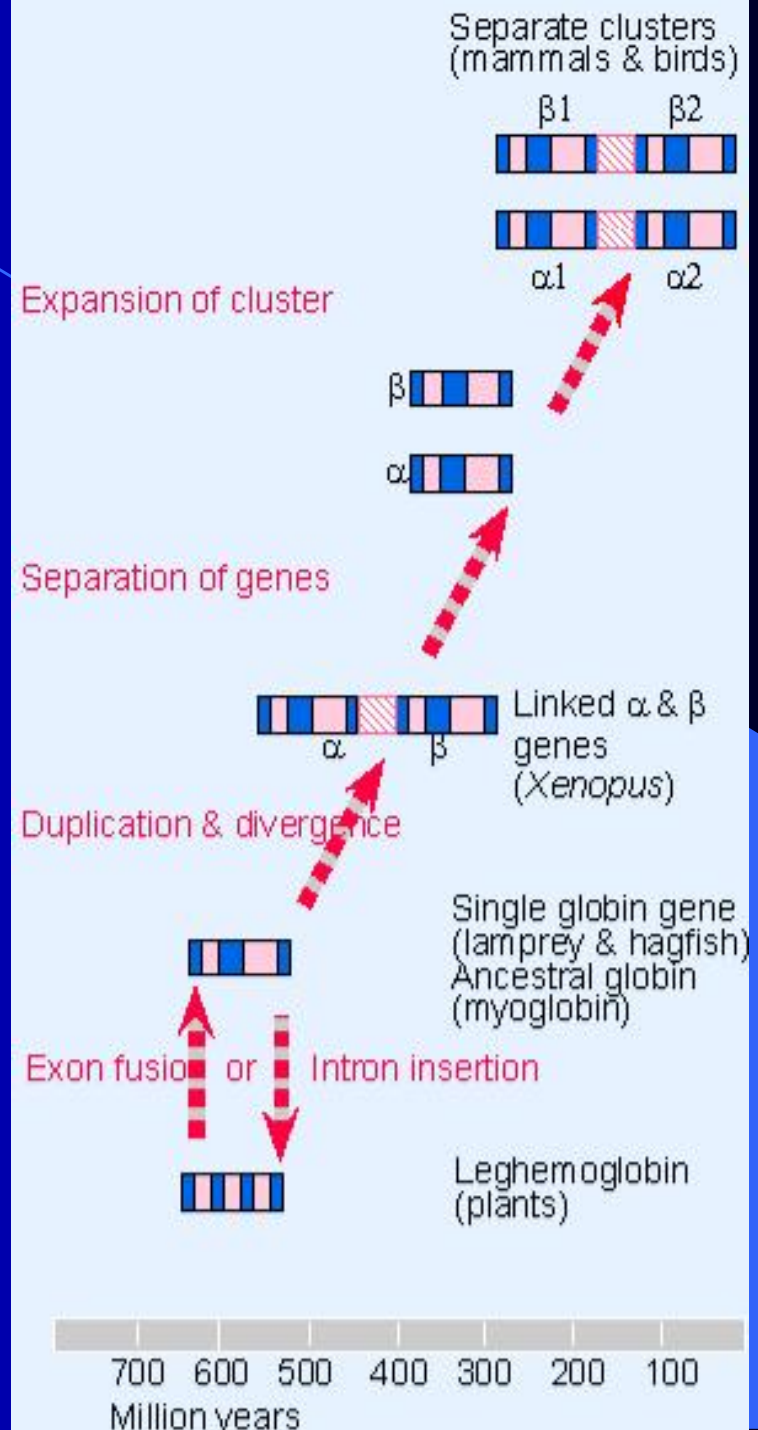
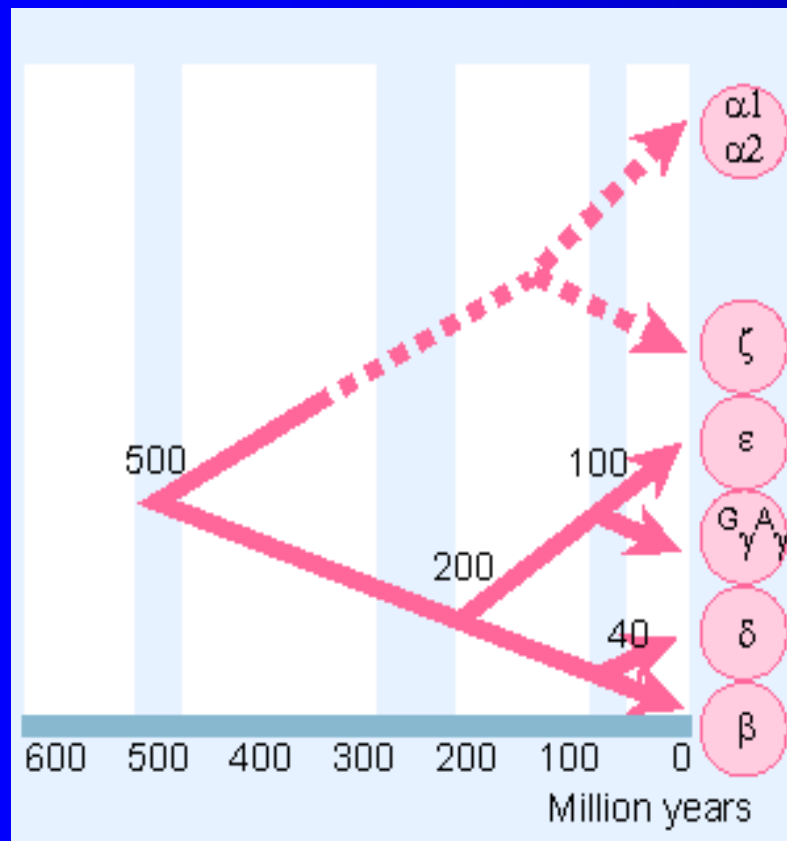
5' — ζ — Ψ ζ — Ψ $\alpha 1$
— $\alpha 2$ — $\alpha 1$ — 3'

- β 类

5' — ζ — Gr — Ar —
 Ψ β — δ — β — 3'



All globin genes have evolved by a series of duplications, transpositions, and mutations from a single ancestral gene.



蛋白质序列高度同源性

基因的核苷酸序列可能不同

- 如src癌基因家族

src, abl, fes, fgr, fps, fym, kck, lyn, ros, tk1, yes

- 家族中各基因的DNA序列没有明显的同源性
- 每个基因产物都含有250个氨基酸顺序的同源蛋白
激酶结构域

超基因家族

- 基因超家族 (gene superfamily)
- 是指一组由多基因家族及单基因家族组成的更大的基因家族
- 结构有程度不等的同源性，可能起源于相同的祖先，但其功能并不一定相同
- 这些基因在进化上也有亲缘关系，但亲缘关系较远，将其称为基因超家族

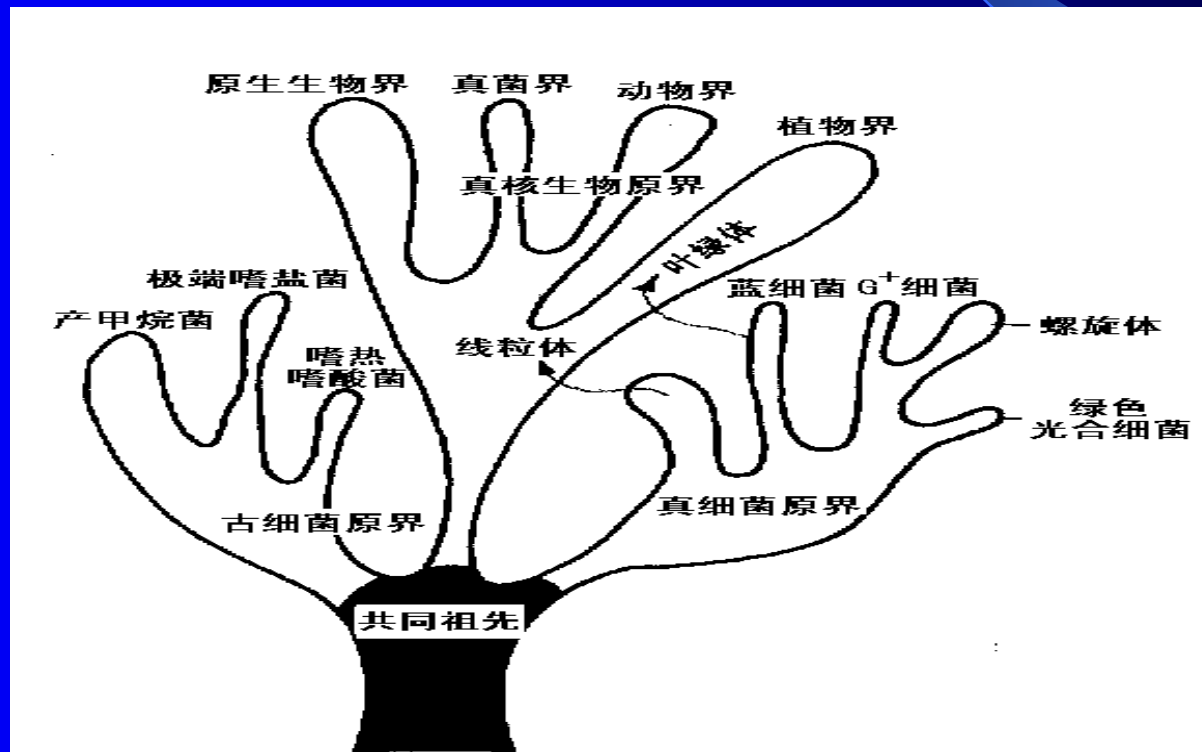
超基因家族

丝氨酸蛋白酶基因超家族

其基因产物都有一个特殊的功能区，丝氨酸是活性中心残基

很多新成员：如载脂蛋白（apolipoprotein），转移胆固醇蛋白颗粒中的成分而不具备任何水解蛋白质的酶功能

- (1) 古细菌原界 (Archaeobacteria) , 包括产甲烷细菌、极端嗜盐菌和嗜热嗜酸菌
- (2) 真细菌原界 (Eubacteria) , 包括蓝细菌和各种除古细菌以外的其它原核生物
- (3) 真核生物原界 (Eucaryotes) , 包括原生生物、真菌、动物和植物



同一物种或不同物种家族———衍生了进化树的概念

生物大分子作为进化标尺依据

蛋白质、RNA和DNA序列进化变化的显著特点是进化速率相对恒定，**分子序列进化的改变量(氨基酸或核苷酸替换数或替换百分率)与分子进化的时间成正比**

- a) 在两群生物中，如果同一种分子的序列差异很大时，
--进化距离远，进化过程中很早就分支了
- b) 如果两群生物同一起来源的大分子的序列基本相同，
--处在同一进化水平上

大量的资料表明：功能重要的大分子、或者大分子中功能重要的区域，比功能不重要的分子或分子区域进化变化速度低

RNA作为进化的指征

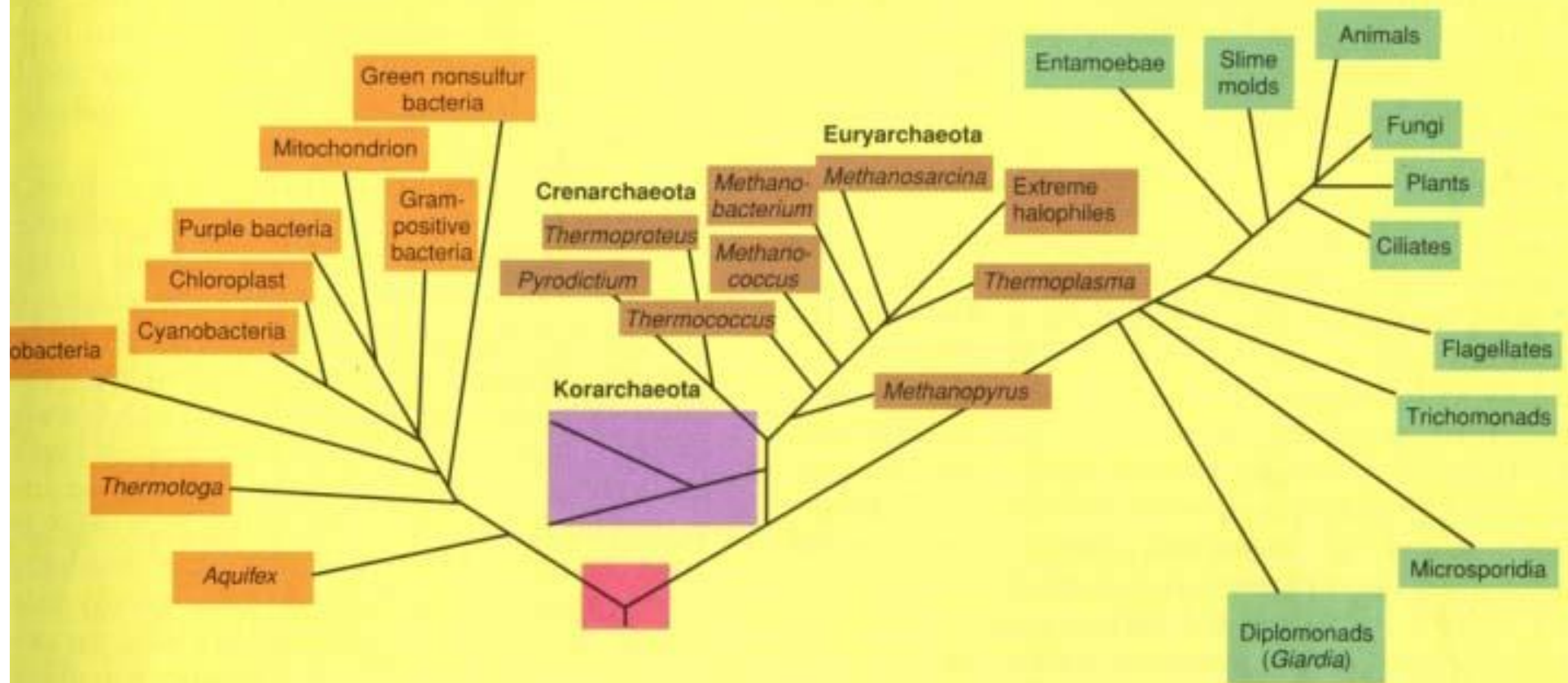
16S rRNA被普遍公认为是一把好的谱系分析的“分子尺”

- ❖ rRNA具有重要且恒定的生理功能
- ❖ 在16SrRNA分子中，既含有高度保守的序列区域，又有中度保守和高度变化的序列区域，因而它适用于进化距离不同的各类生物亲缘关系的研究
- ❖ 16SrRNA分子量大小适中，便于序列分析
- ❖ rRNA在细胞中含量大(约占细胞中RNA的90%)，也易于提取
- ❖ 16SrRNA普遍存在于真核生物和原核生物中(真核生物中其同源分子是18SrRNA)。因此它可以作为测量各类生物进化的工具

Eubacteria
(真细菌界)

Archaeobacteria
(古细菌界)

Eukarya
(真核生物界)



Carl Woese利用16SrRNA建立分子进化树



利用16SrRNA建立分子进化树的美国科学家
Carl Woese

本章要点

- C值矛盾
- 基因概念的演化
- 内含子功能
- 细菌/病毒基因特点
- 真核基因组中DNA序列的分类
- 基因家族
- 16SrRNA或18SrRNA的核酸序列分析