

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



内 容

第一节 基因 (Gene)

第二节 基因组 (Genome)

病毒基因组

原核基因组

真核基因组

第三节 基因组学&人类基因组计划



基本概念

- 基因 (**Gene**)
- 基因组 (**Genome**)
- 基因组学 (**Genomics**)



第一节 基因

一、基因的概念

二、基因的特点和鉴定标准

三、基因的大小



一、基因(gene)的概念

□ 早期的基因概念

- **Morgan**等把遗传性状与特定的染色体区段联系起来，并提出“一个基因，一个酶”和“一个基因，一条多肽链”的概念。

□ 分子生物学的基因概念

- 基因是一段携带功能产物（多肽，蛋白质，**tRNA**和**rRNA**和某些小分子**RNA**）信息的**DNA**片段，是控制某种性状的遗传单位。



一、基因的概念

基因 (gene)

- ❑ 编码蛋白质多肽链或RNA所必需的核酸片段。
- ❑ 一般是DNA序列（或RNA病毒中的RNA）。

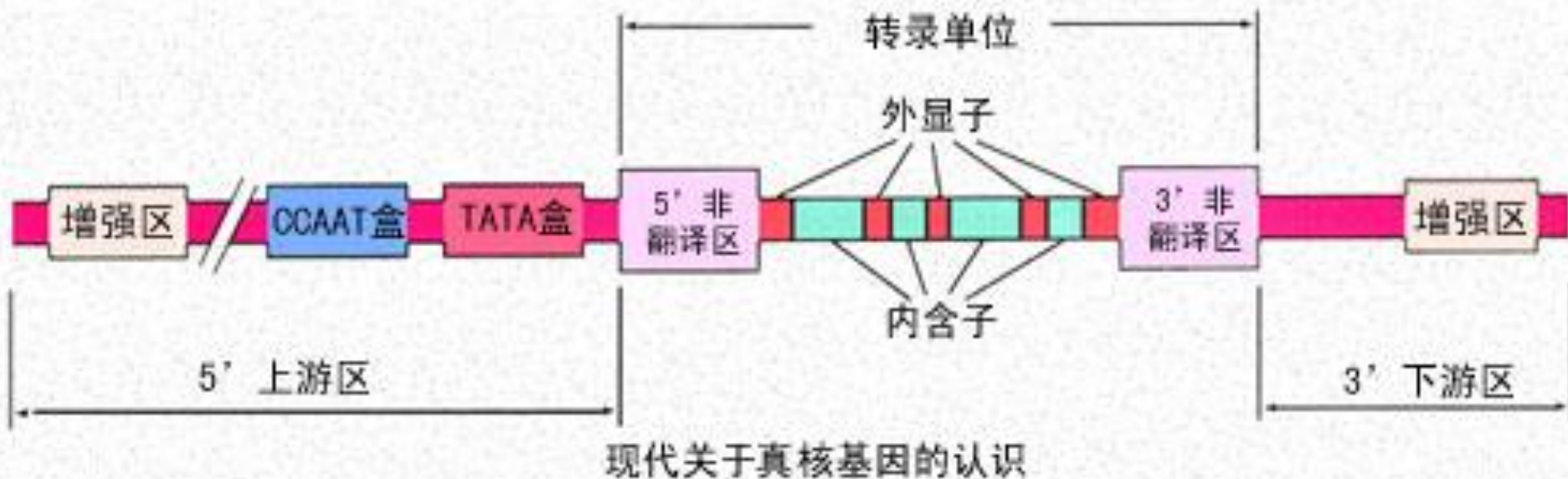


一、基因的概念

基因 = 结构基因 + 调控序列

- 结构基因：编码多肽链或RNA的部分。
- 调控序列：基因编码区的两侧，与转录调控相关。





一个典型的真核基因

- ① **编码序列**：外显子(exon)
- ② 插入外显子之间的**非编码序列**：内含子(intron)
- ③ **5'-端和3'-端非翻译区(UTR)**
- ④ **调控序列**(可位于上述三种序列中)



二、基因的特点和鉴定标准

1. 开放阅读框(ORF)
2. 密码子偏爱
3. 序列保守性
4. 转录产物
5. 基因失活



open reading frame, ORF(开放阅读框架)



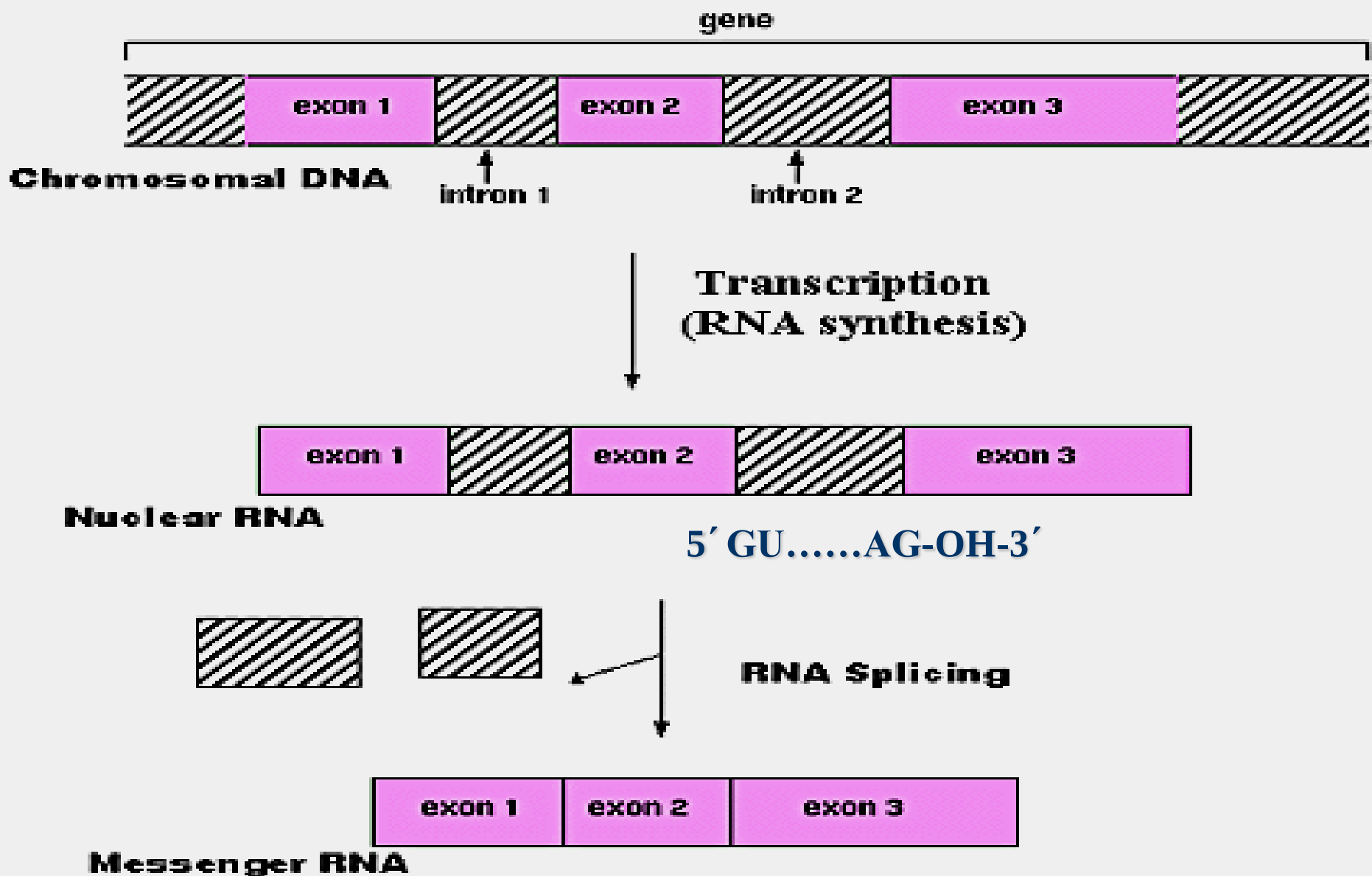
从**mRNA 5'端**起始密码子**AUG**到**3'端**终止密码子之间的核苷酸序列，各个三联体密码连续排列编码一个蛋白质多肽链，称为开放阅读框架。



Genetic code

1 st base		U				C				A				G				3 rd base	
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U										
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C										
	UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop	A										
	UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp	G										
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U										
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	C										
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A										
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G										
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U										
	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C										
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A										
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G										
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U										
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C										
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A										
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G										





RNA synthesis and processing



三、基因的大小

- 不同物种相差较大。
- 主要取决于内含子的有无、大小及数量。

病毒、原核、低等真核生物：不超过5Kb

高等真核生物：高达100Kb



第二节 基因组

一、基因组的概念

二、基因组大小与C值矛盾

三、基因总数与N值矛盾



一、基因组的概念

基因组(genome):

- ❑ 1920年Winkles从GENes和chromosOMEs组成
- ❑ 指生物体的一套完整的单倍体遗传信息。

原核生物和真核生物的基因组： DNA

- ❑ 包括：染色体基因组和染色体外基因组
质粒DNA；
线粒体DNA及叶绿体DNA





人类基因组

人类基因组是指人的22条常染色体与2条性染色体的全部DNA和线粒体DNA。



二、基因组大小与C值矛盾

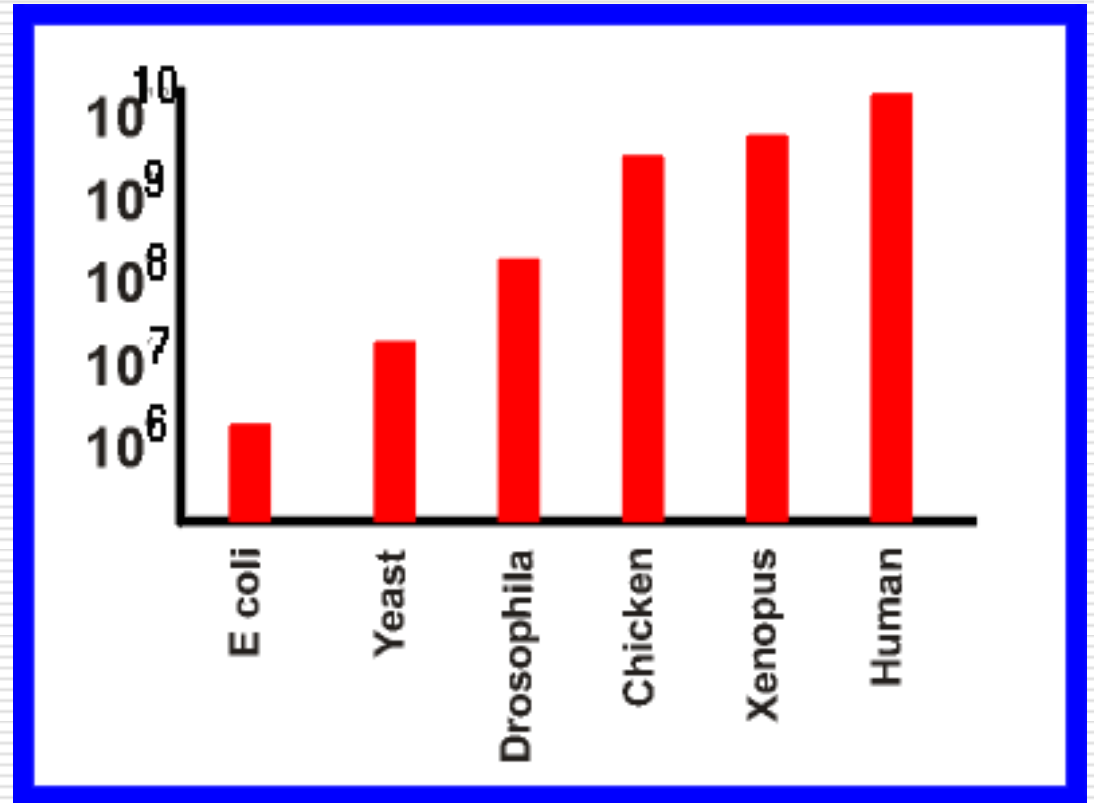
1. 基因组的大小:

- 通常以其DNA含量（全部DNA的碱基对总数）来表示。
- 单倍体基因组中的全部DNA量称为C值 (C-Value)。



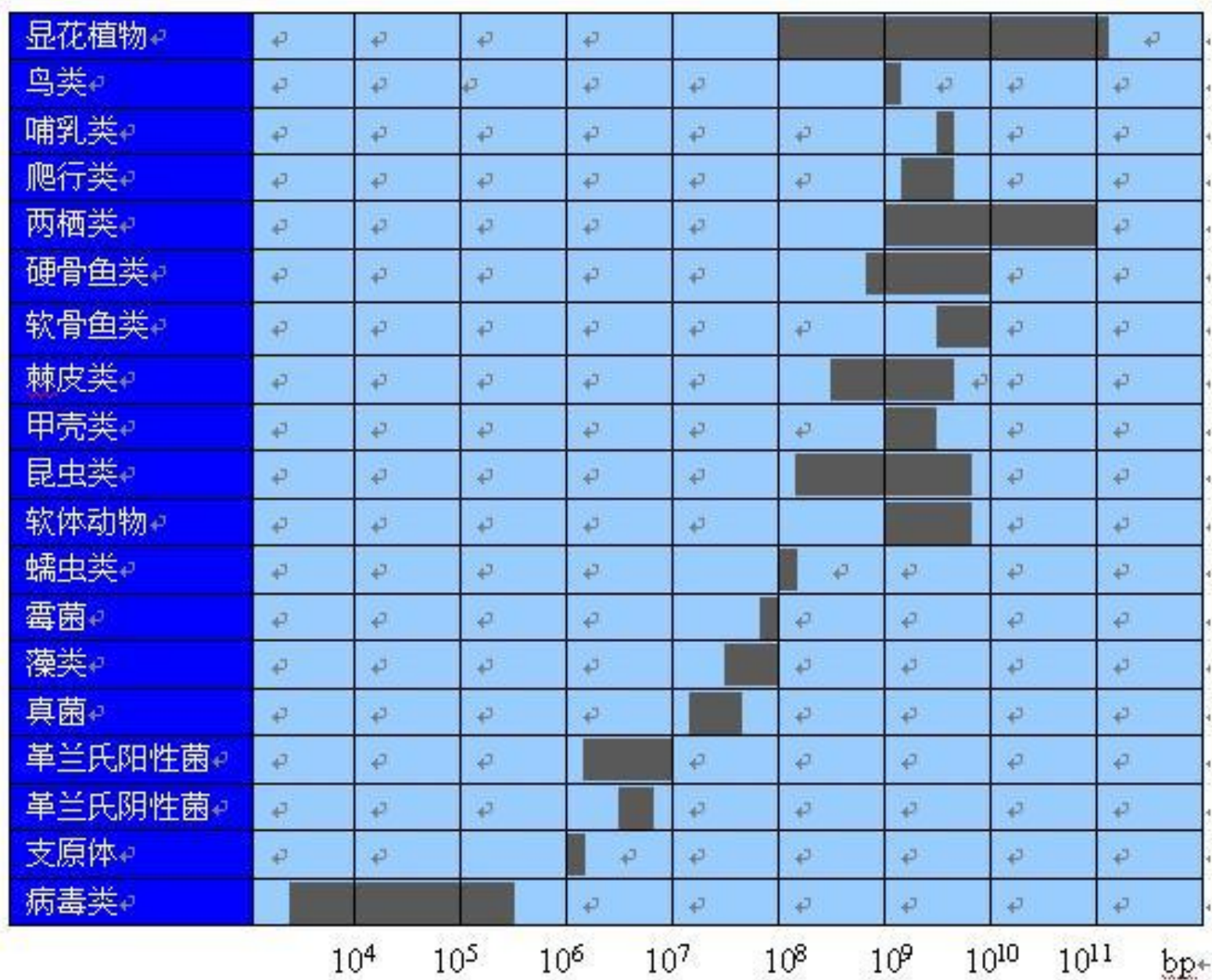
二、基因组大小与C值矛盾

C值 (C-Value)



- C值大小基本上反应生物进化程度的差异。





二、基因组大小与C值矛盾

2. C值矛盾

- 生物体的进化程度与基因组大小不完全成比例的现象。
- C值矛盾主要是由于各种生物基因组的结构特征的差异造成的，低等和高等生物差别明显。



C值矛盾的表现方面

- ❑ C值不随生物的进化程度和复杂性而增加,如肺鱼的C值为112.2G, 而人的是3.2G, 与牛相近.
- ❑ 亲缘关系密切的生物C值相差甚大, 如豌豆为14, 而蚕豆为2.
- ❑ 高等真核生物具有比用于表达高得多的C值, 如人的染色体组DNA含量在理论上包含300万个基因, 但实际有用途的基因只有4万左右.



三、基因总数与N值矛盾

□ 人类基因总数：2.1万

水稻基因总数：5万

□ N值矛盾：基因组中的基因数目与生物进化程度或复杂程度的不对称性。

