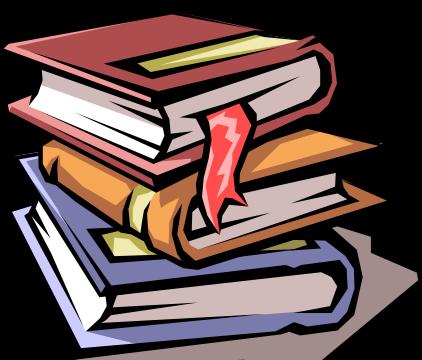


第2章

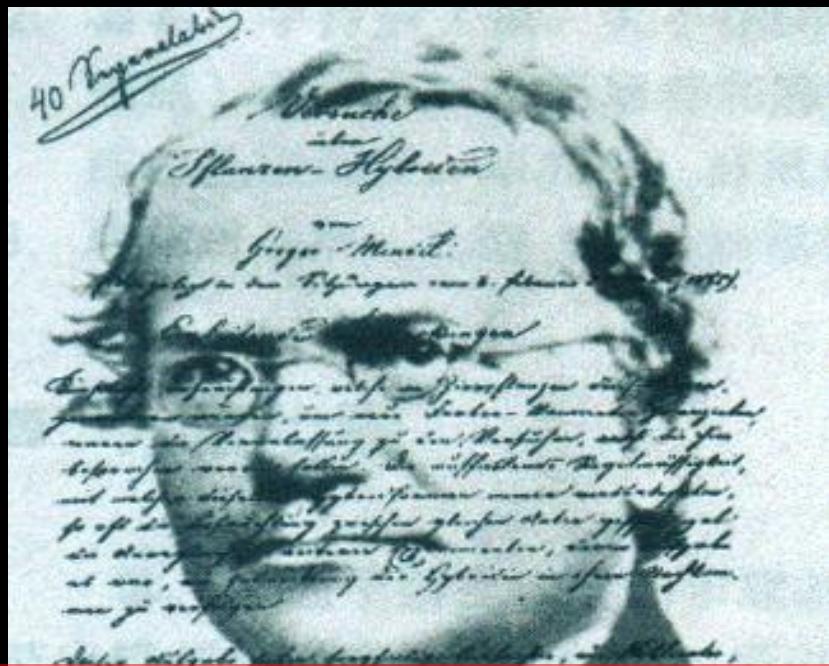
基因的概念

Concept of the Gene



- 
- 2.1. 基因的概念、演变与发展
 - 2.2. 基因的分子结构
 - 2.3. 基因概念的多样性-重叠基因
 - 2.4. 基因概念的多样性-重复基因
 - 2.5. 基因概念的多样性-间隔基因
 - 2.6. 基因概念的多样性-跳跃基因

(Source:http://es.wikipedia.org/wiki/Gregor_Mendel)



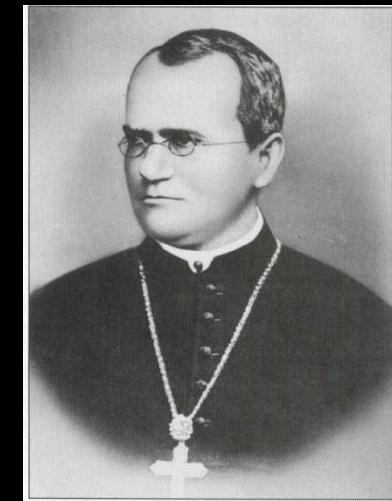
2.1.1 经典的基因概念

(Source:National Library of Medicine)

1866. -1926.

遗传因子假说

(Hypothesis of the inherited factor G. J. Mendel 1866.)



- 生物性状由遗传因子控制
- 亲代传给子代的是遗传因子(A,a....)
- 遗传因子在体细胞内成双(AA,aa)
在生殖细胞内为单 (A,a)
- 杂合子体细胞内具有成双的遗传因子(Aa)
- 等位的遗传因子独立分离
- 非等位遗传因子间自由组合地分配到配子中

否定了 **Hypothesis of the Pangenesis**

奠定了 **Theory of particulate inheritance**

提出了 **Law of segregation**

Law of independent assortment

缺乏细胞在有丝分裂和减数分裂过程中

染色体行为的有力证据

1900.

Hugo de vires 1848-1935 荷兰 阿姆斯特丹大学

Erich von Tchermark 1871-1962 德国 土宾根大学

Carl Correns 1864-1933 奥地利 维也纳农业大学

同时发现并证实 Mendel 的两大规律

Law of segregation

Law of independent assortment

Walter S. Sutton “The Chromosome in Heredity” (1903)

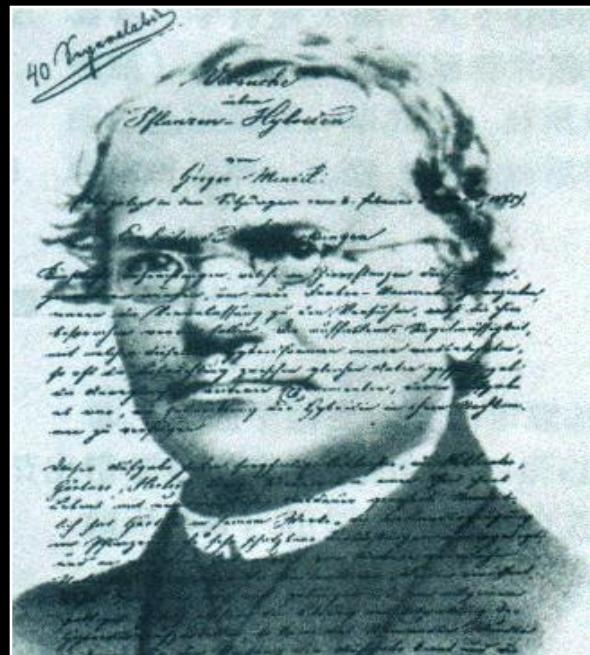
- 二倍体生物减数分裂后每个配子仅接受成对染色体中的一个
- 假定基因是染色体的一部分

第一次将遗传学和细胞学科学地结合起来
为Mendel遗传规律的解释提供了细胞学证据

1906 Batson F2 ≠ 9: 3: 3: 1 ?

1910 T. H. Morgan → The law of linkage

(Source:http://es.wikipedia.org/wiki/Gregor_Mendel)



(Source:National Library of Medicine)

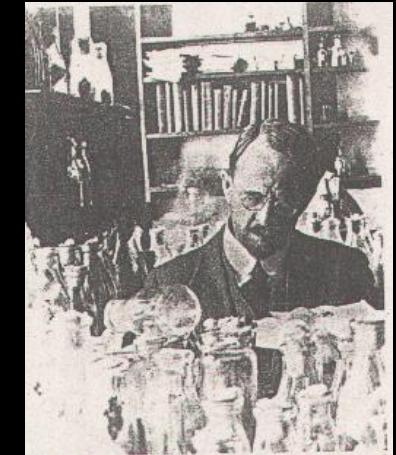
Theory of the Gene

Thomas Hunt Morgan

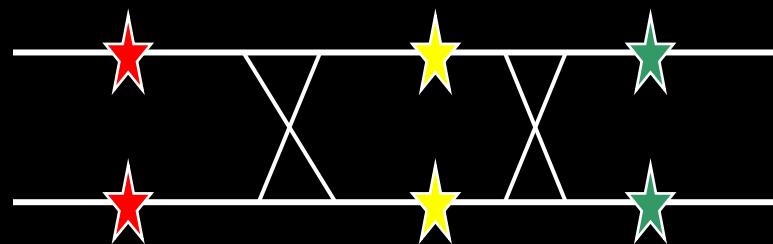
1926

Theory of the gene

- 基因是染色体上的实体
- 基因象链珠(bead)一样，孤立地呈线状地排列在染色体上
- 基因是 **(Three in one)**；



(1926 T. H. Morgan)



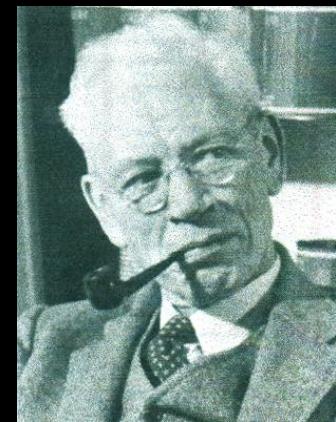
功能(function unit)
突变(mutation unit)
交换(crossing-over unit)

“三位一体”的

最小的
不可分割的
基本的
遗传单位

2.1.2. 基因的位置效应

Sturtevant



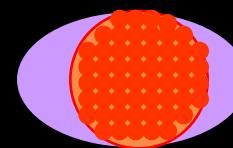
Position effect

X - chromosome

16A



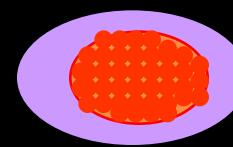
Wild type



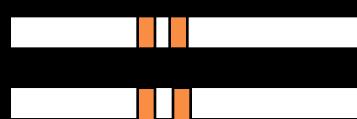
780个



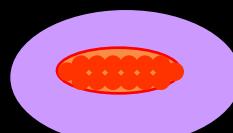
Duplication



385个



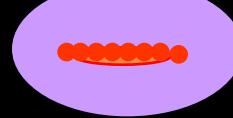
Dosage effect



68个



Position effect



45个

Bar eye

来源：分子生物学（2012），郑用琏，第19页

(Source:http://www.nobelprize.org/nobel_prizes/medicine/laureates/1933/morgan-article.html)

W



Ww Red eye

(W>w in euchromatin)

w

Epigenetics Why ?

W



Ww white eye

(W gene be silenced

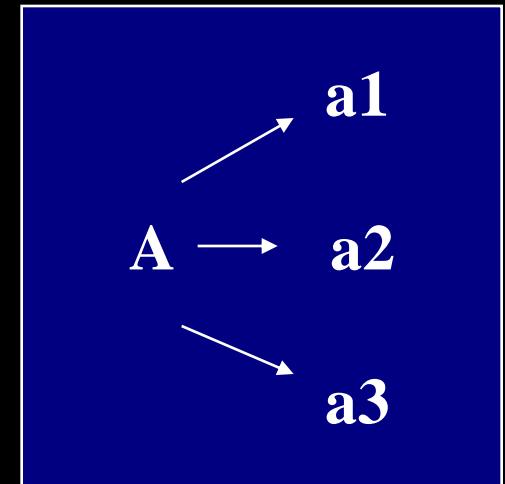
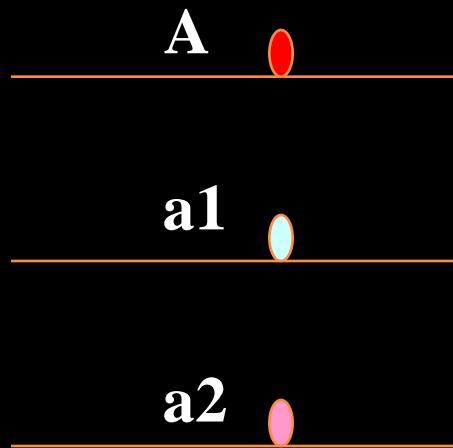
in heterochromatin)

w

Position effect

2.1.3. 拟等位基因概念的提出 (pseudo alleles)

Multiple alleles

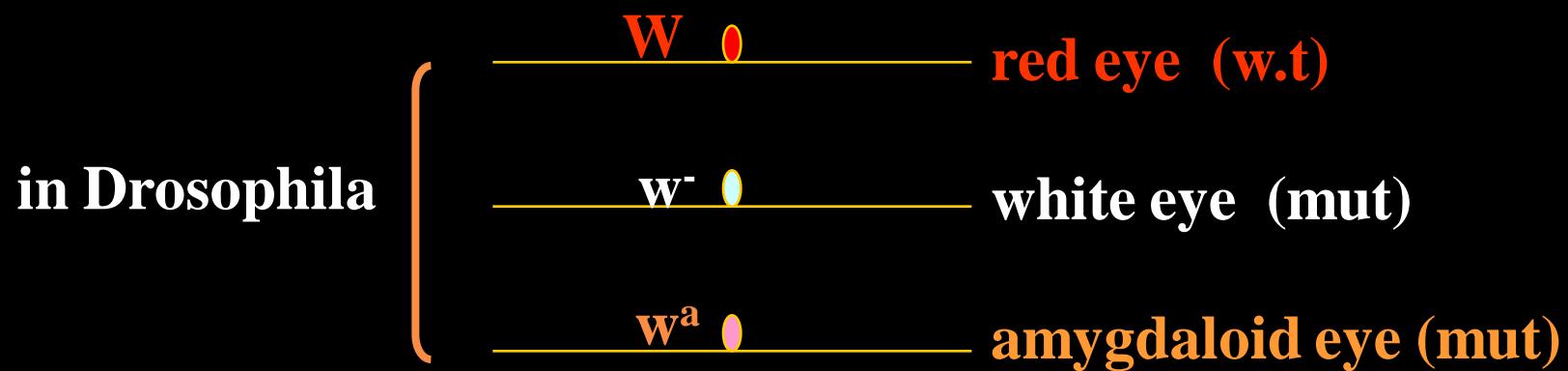


a1a1 × a2a2

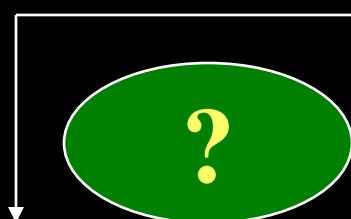


a1a2 (no wild type)

But !



$X^{w^-} X^{wa} \times X^{w^-} Y$



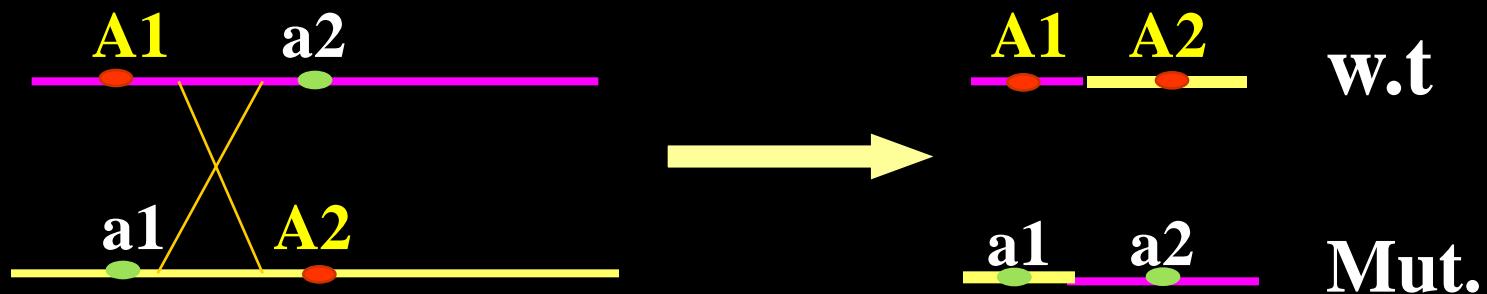
1/1000
W.T
red eye

| | | | |
|----------|-----------|-------------------|-------------|
| | | X^{w^-} | Y |
| | X^{w^-} | $X^{w^-} X^{w^-}$ | $X^{w^-} Y$ |
| X^{wa} | | $X^{w^-} X^{wa}$ | $X^{wa} Y$ |

Pseudo Alleles

紧密连锁 (交换率极低)

功能相同 (表型相似)



From 1940's

Microbe as genetic research material



Phage Group

Single
Simple
Propagation
Population

Genetics developing

(source:http://www.nobelprize.org/nobel_prizes/medicine/laureates/1969/hersey-photo.html)

顺反子理论
Theory of cistron
(S.Benzer 1955)

对经典的基因概念的
第一次重要修正与发展

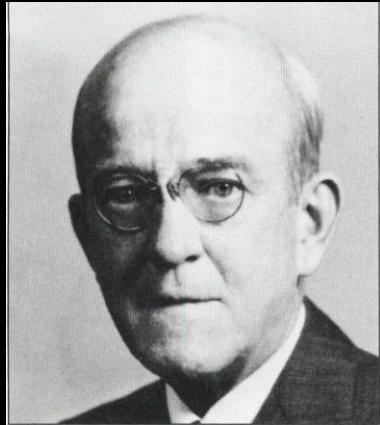


理论基础

- 1941. Beadle & Tatum

Nuraspora crassa One gene ----one enzyme

- 1944 Oswald Avery (Canada)



Diplococcus pneumoniae

DNA as genetic material

(Source:National Academy of Sciences)

Jacob & Monod

Escherichia coli Lactose operon

- 1955. S.Benzer

E. coli T4 phage Cistron

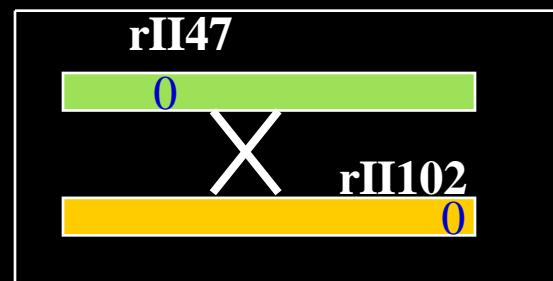
2.1.4. Theory of cistron (S.Benzer 1955)

Plaque

| Phage | E.coli B | E.coli k12 |
|------------|------------|------------|
| W.t T4 | 白, 小, 边缘模糊 | 白, 小, 边缘模糊 |
| Mut.T4 rII | 大, 圆, 边缘清晰 | ! |

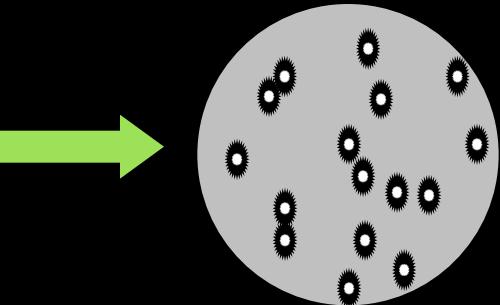
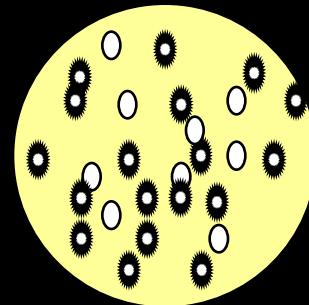
Mut.T4 rII : rII107, rII105, rII51, rII47.....(400)

Recombination assay

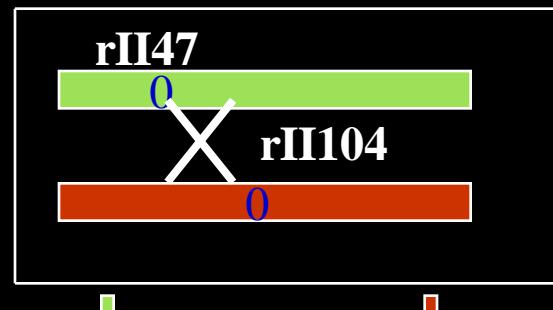


E.coli B

plane *E.coli B*

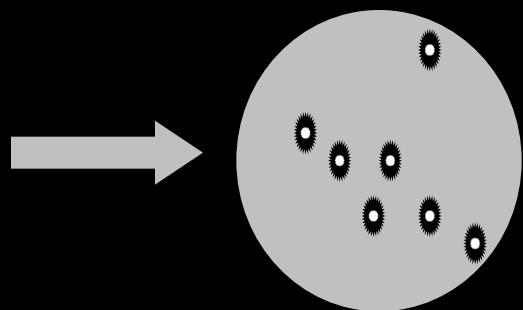
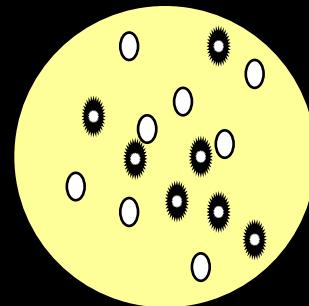


Blotting into plane
E.coli K12

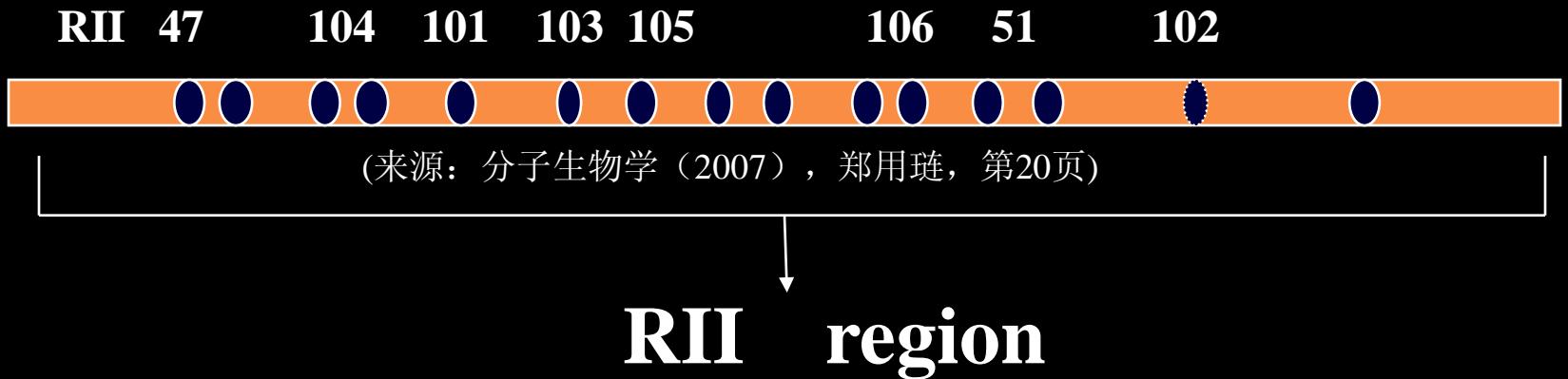


E.coli B

plane *E.coli B*



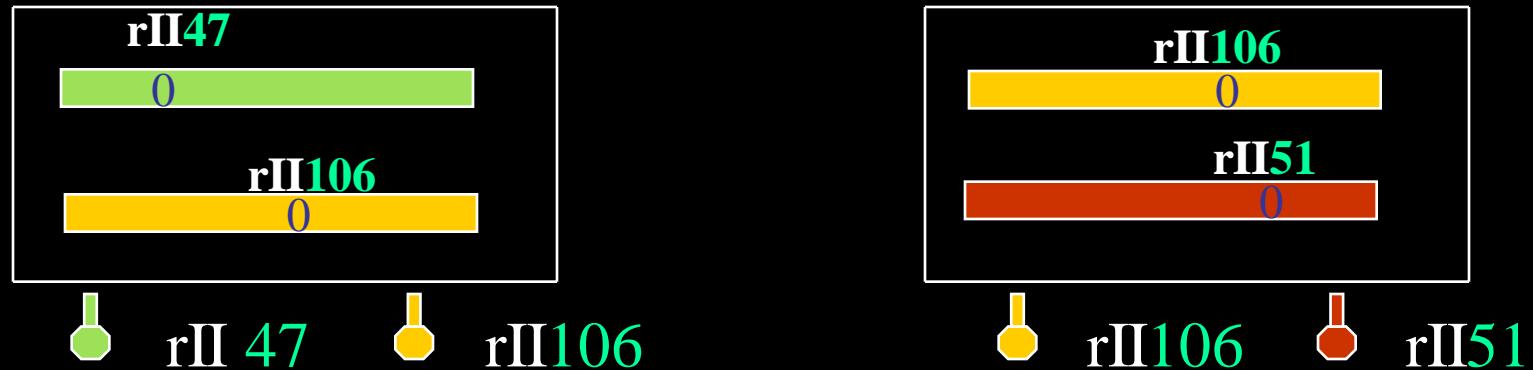
(来源: 分子生物学 (2007), 郑用琏, 第20页)



Three in one !

How many genes ? !

Complementary assay

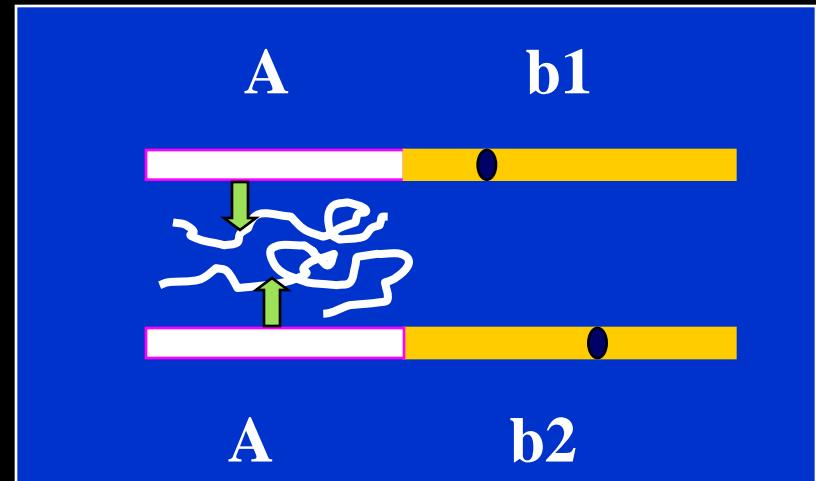
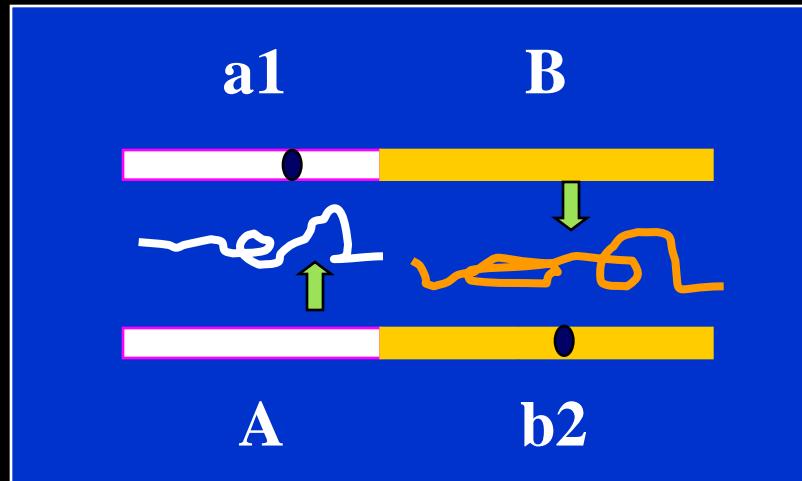


plane
recombination ? !
How?

Functional complementary → propagation → recombination

(来源：分子生物学（2007），郑用琏，第21页)

依据; One gene → one enzyme



Wild type

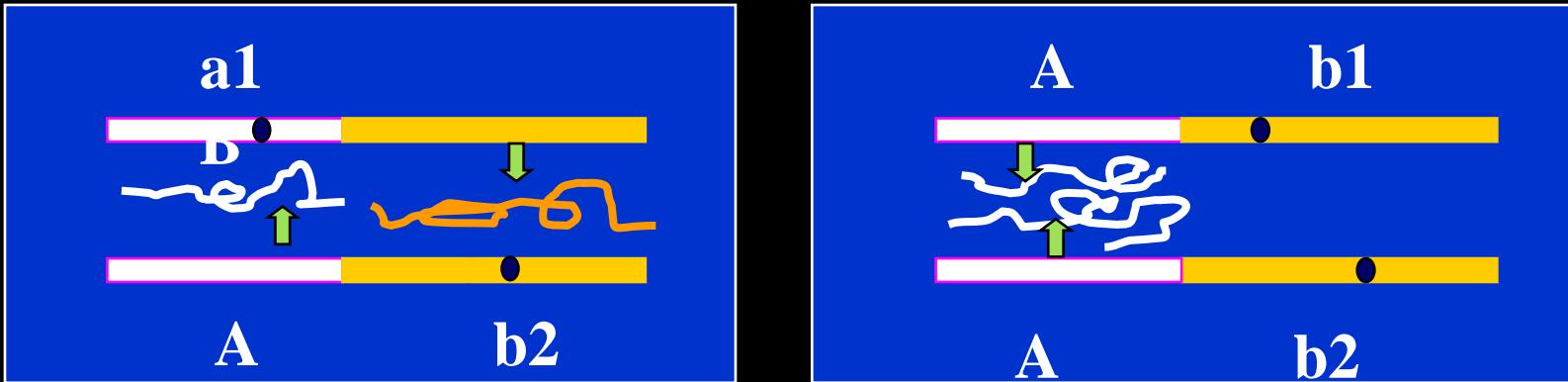


Mutant

相依为命 !

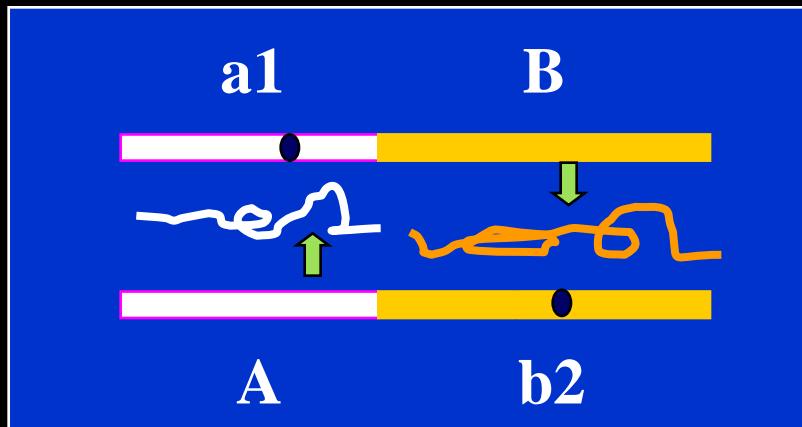
无能为力 !

(来源: 分子生物学 (2007), 郑用琏, 第21页)



(来源：分子生物学（2007），郑用琏，第21页)

- 杂合二倍体内， w.t. 基因对mut. 等位基因可以发生功能的补偿，产生**功能互补效应**
- 带有不同突变位点的噬菌体同时感染一个细菌，构成双突变杂合二倍体，组成互补测验体系，以测定各突变位点所在基因的等位性

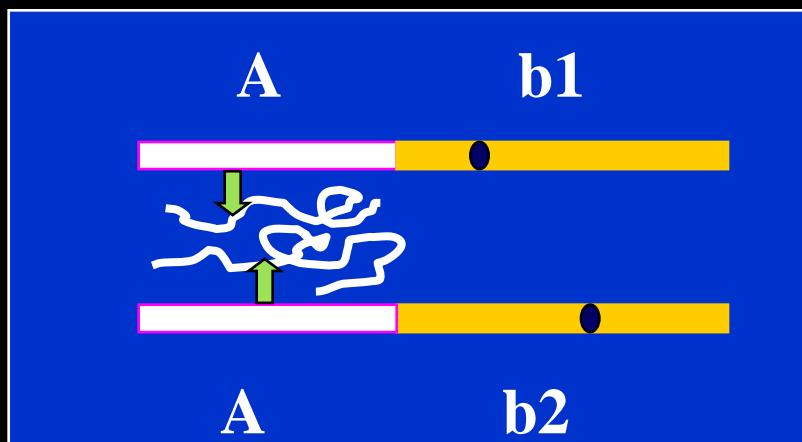


(来源: 分子生物学 (2007), 郑用琏, 第21页)

具有
功能互补效应
的测验体系
不具有

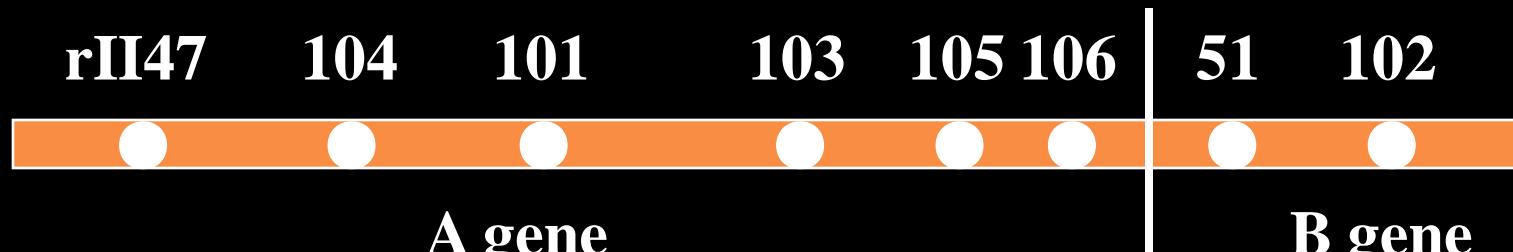
突变位点
处于

不同的非等位基因
同一等位基因



(来源: 分子生物学 (2007), 郑用琏, 第21页)

rII of T4 phage including two genes



(来源: 分子生物学 (2007) , 郑用琏, 第22页)

顺反子假说 (Theory of cistron)

- Cistron 是基因的同义词
- 在一个顺反子内，有若干个突变单位, 突变子 (muton)
- 在一个顺反子内，有若干个交换单位, 交换子 (recon)
- 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 three in one ——————> one in one
- 基因内可以较低频率发生基因内的重组，交换
- pseudo alleles 是基因内的突变体
- mut1 X mut2 ——————> W.t 是基因内发生交换的结果
- cistron 概念的提出是对经典的基因概念的动摇
是对pseudo alleles概念的修正

2.1.5. 等位基因 (Allele, Allelomorph) 概念的发展

Allele 同一座位存在的两个以上不同状态的基因, 其总和称之为复等位基因 (multiple alleles) (A, a1, a2..)



Pseudo allele

? 非全同等位基因!



Allele 具有相对差异的DNA区域

TAAAGTAAT TAAAGCAAT



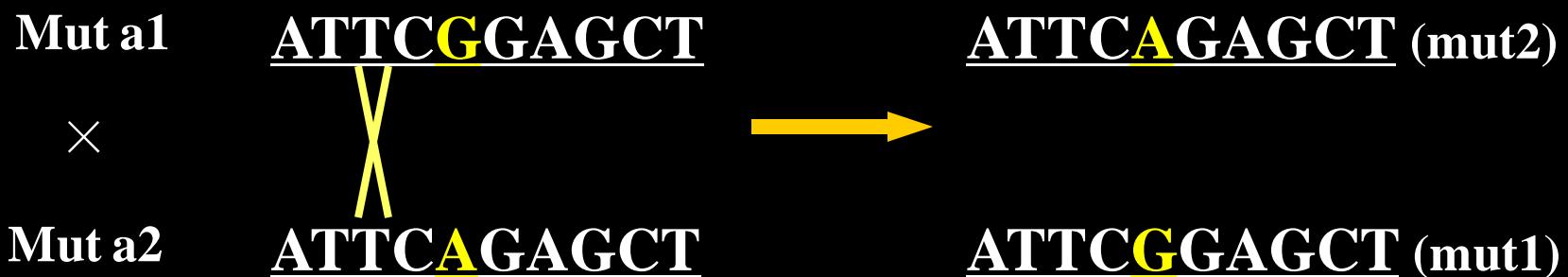
全同等位基因 (homoallele)

非全同等位基因 (heteroallele)

全同等位基因 在同一基因座位(**locus**)中，同一突变位点 (**site**) 向不同方向发生突变所形成的等位基因(**homoallele**)

| | site ↓ |
|--------|----------------------|
| W.t A | ATTCT G A GCT |
| Mut a1 | ATT C GGAGCT |
| Mut a2 | ATT C A GAGCT |

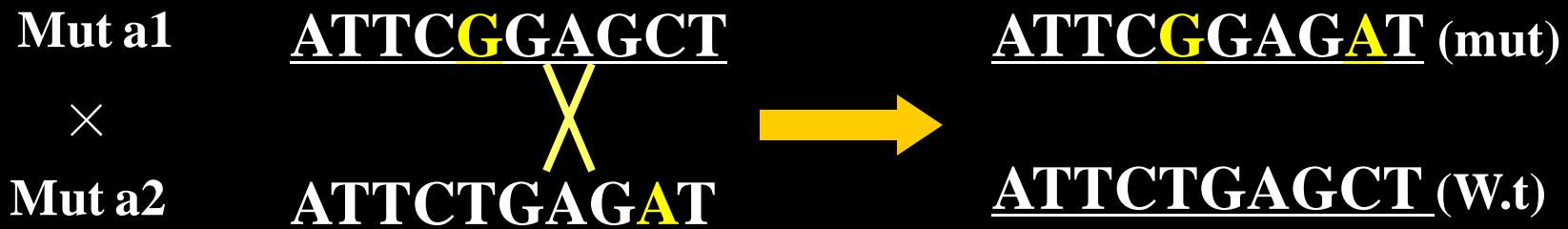
Gene locus



非全同等位基因；在同一基因座位（locus）中，不同突变位点（site）发生突变所形成的等位基因。（heteroallele）

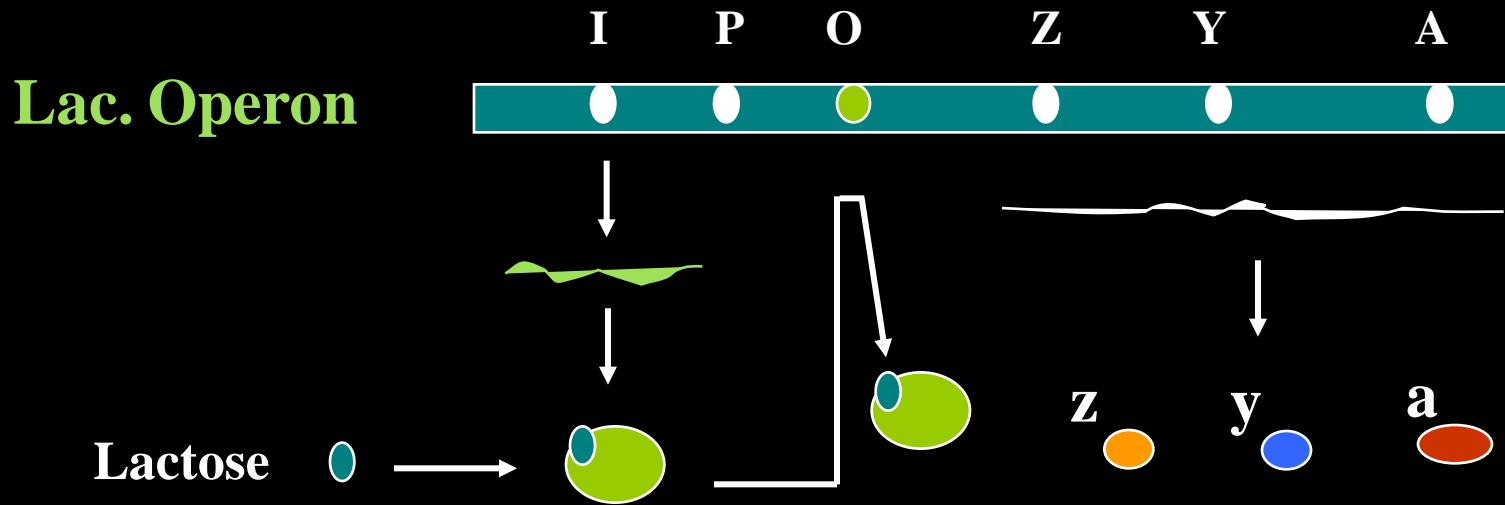
| | site site |
|--------|---|
| W.t A | ↓ ↓ |
| | ATTCTGAGCT ATT C G GAGCT |
| Mut a1 | |
| Mut a2 | ATTCTGAG A T |

Gene locus



2.1.6. 操纵子理论

(Lactose operon 1961. Jacob, Monod)



生物体性状的表现往往具有上位性互作的遗传效应

- 某一基因功能的表现

是若干基因组成的信息表达的整体行为

- **one gene → one enzyme**

tDNA, rDNA, polymer

one gene → one peptide

one gene → one function

(Ribozyme, Abzyme, rDNA, tDNA..)

2.1.7 基因的类型

★ transcribed, translatable gene

(Z,Y,A)

transcribed but non-translatable gene

(tDNA, rDNA, small RNA.....)

Non- transcribed, non-translatable gene (?)

(promoter, operator)

★ cis action factor (基因表达调控因子)

Affects the activity of DNA sequences only on its own molecular of DNA, this property usually implies that the factor does not code for protein

通过核苷酸自身的特异二级结构控制与它紧密连锁的结构基因的表达

一般不编码蛋白质(无基因产物的DNA功能区)

★ trans action factor (基因表达调控因子)

Affects the activity of any gene located on genome by its translated product.

通过扩散自身表达产物（酶，调节蛋白）
控制其他基因的表达

可转录，可翻译调节蛋白的DNA功能区

可通过互补测验体系确定其功能区域

2.1.8. DNA 是主要的遗传物质

DNA编码两类遗传信息

I类; DNA seq. → RNA seq. (codon) →

→ aa seq. → protein

phenotype (central dogma)

II类; 特定DNA seq. + 特定蛋白质/核酸结合



基因表达的指令 → gene on / off

I类

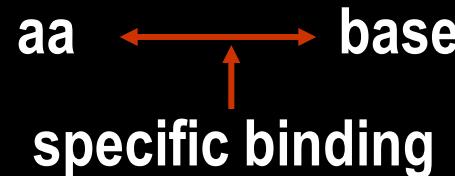
- 内在信息
- ORF only
- expression by
DNA → RNA → protein
- 遗传信息存在于模板链的一级结构上
- 三联体密码
- 简并 (degeneration)

Wobble
Iso-acceptor

II类

内, 外(信号分子)结合信息

Helix , Nt seq....



三维空间结构 / DNA序列
(IR, Box, paracodon...)

空间, 调控密码 (key & lock)

简并 (degeneration)



- DNA作为遗传物质的优点（自然选择的优势）

★ 储存遗传信息量大

1kb DNA序列 → 4^{1000} 种遗传信息

★ A / T, C / G 互补 → 双螺旋结构 → 复制, 转录
遗传稳定 ↘

★ 核糖的2'-OH脱氧 → 在水中的稳定性高于RNA

★ 可以突变 → 以求不断进化
方便修复 → 以求稳定遗传

★ DNA中有T无U是进化的结果

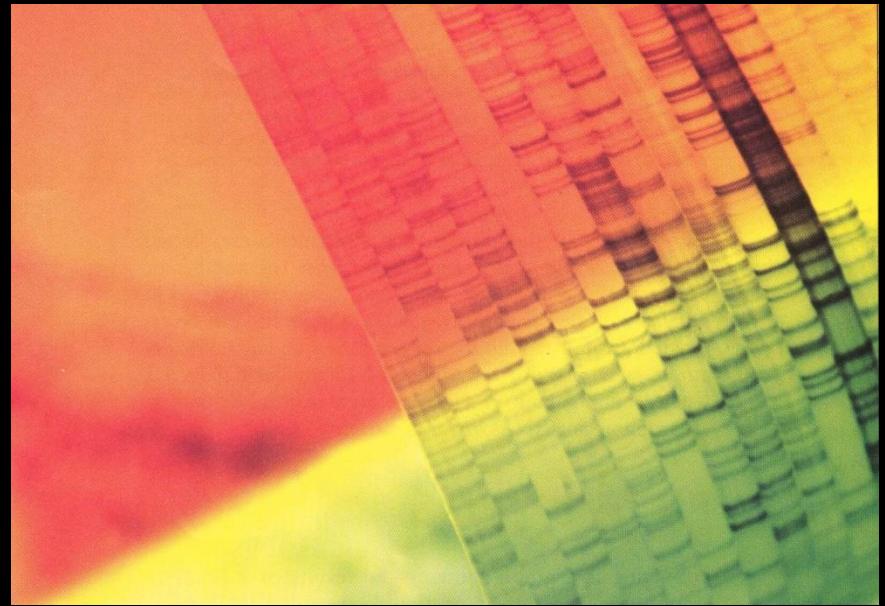
G/C → G/U $\xrightarrow[\text{倾向除U插C}]{\text{错配修复系统}}$ G/C

C → U 突变

DNA中如有U 错配修复系统不能
识别两种U，难以
对突变进行修复 $\xrightarrow{?}$ 潜在的遗
传危险

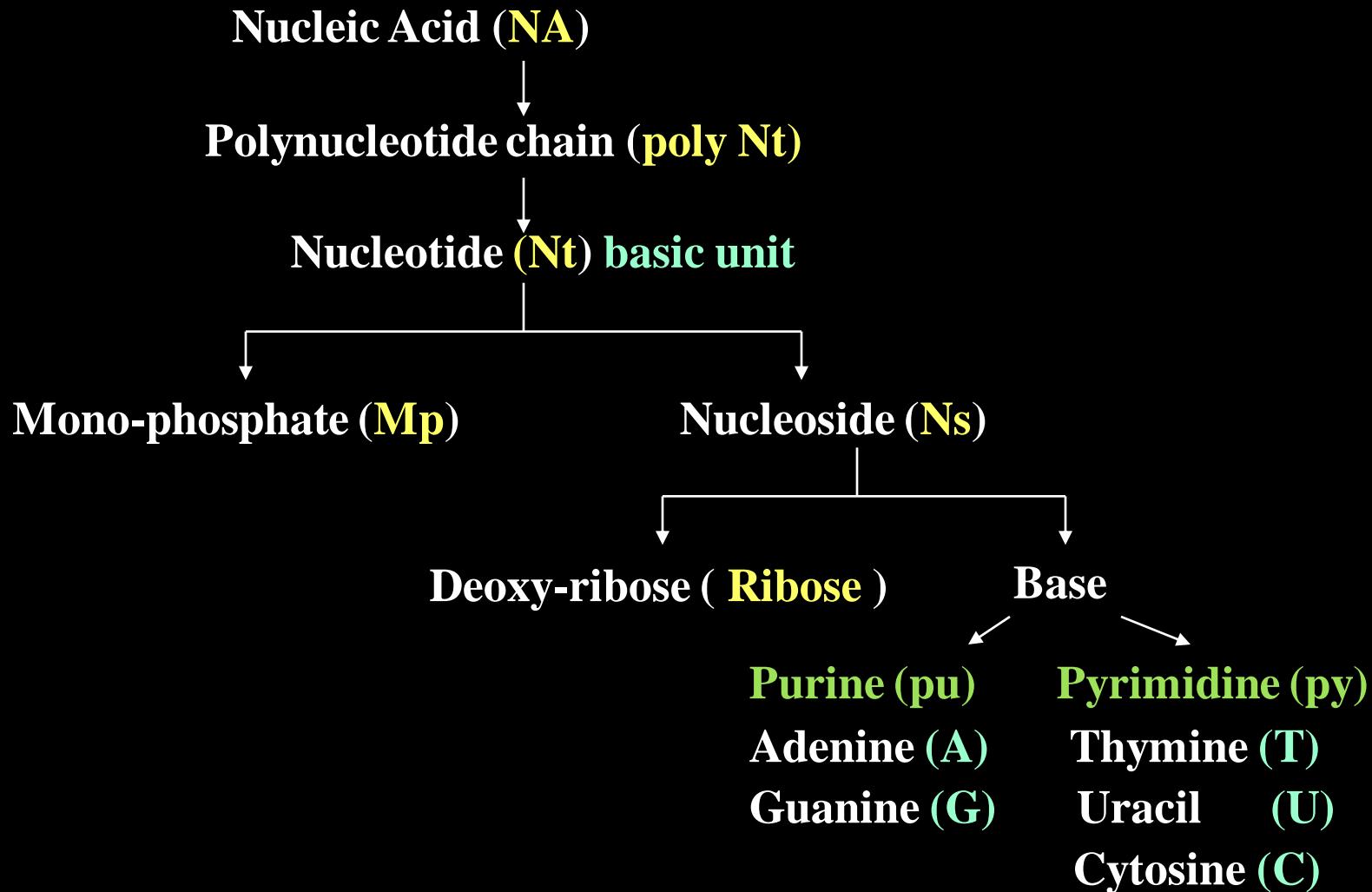
★ DNA中有T无U → 无后顾之忧 → 进化
DNA扩增

2. 2. 基因的分子结构



(Source:Molecular Biology(2002),Robert F.Weaver,Page84)

2.2.1. 基因是DNA分子的片段



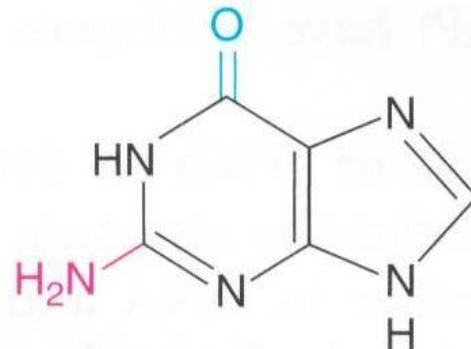
Nucleotide (Nt) basic unit



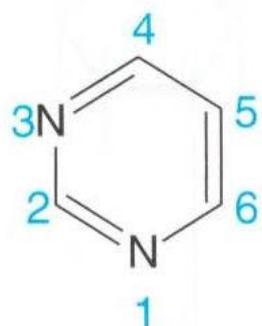
Purine



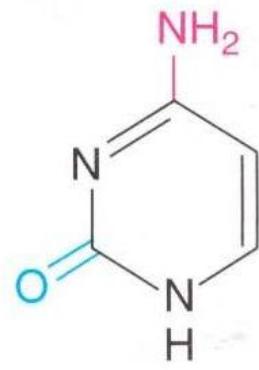
Adenine



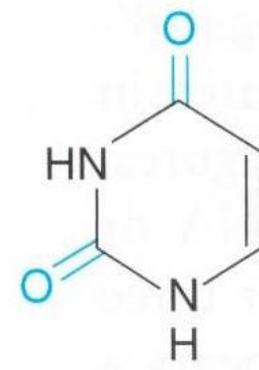
Guanine



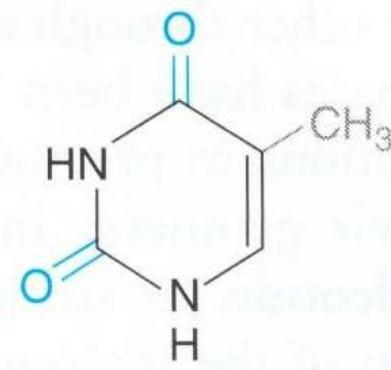
Pyrimidine



Cytosine



Uracil

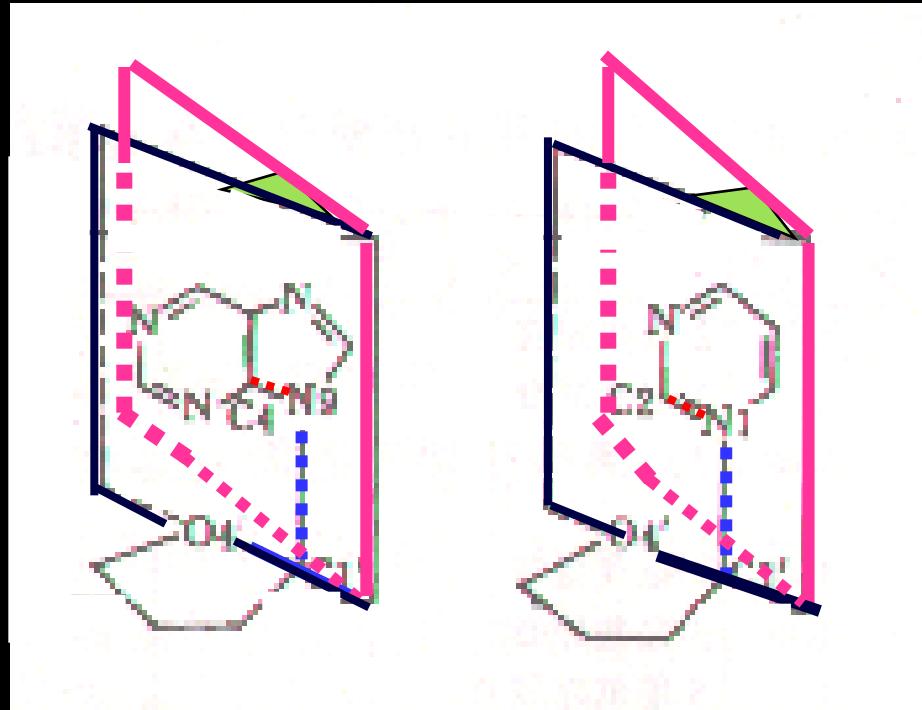


Thymine

(Source:Molecular Biology(2002),Robert F.Weaver,Page17)

2.2.2 核苷的构象 (conformation of nucleoside)

purine

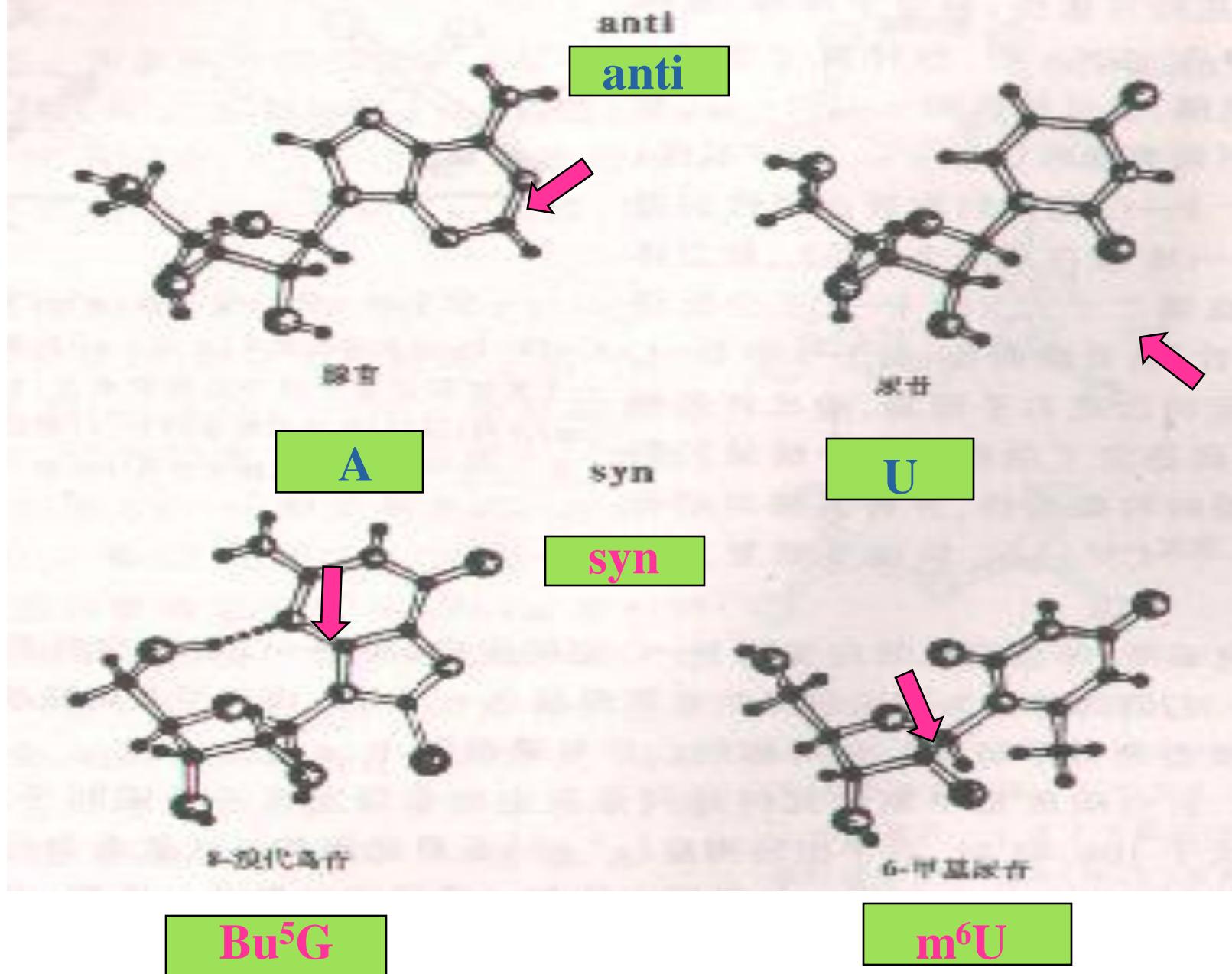


pyrimidine

$$\chi = \text{C4-N9-C1}'-\text{O4}' \\ 0^\circ$$

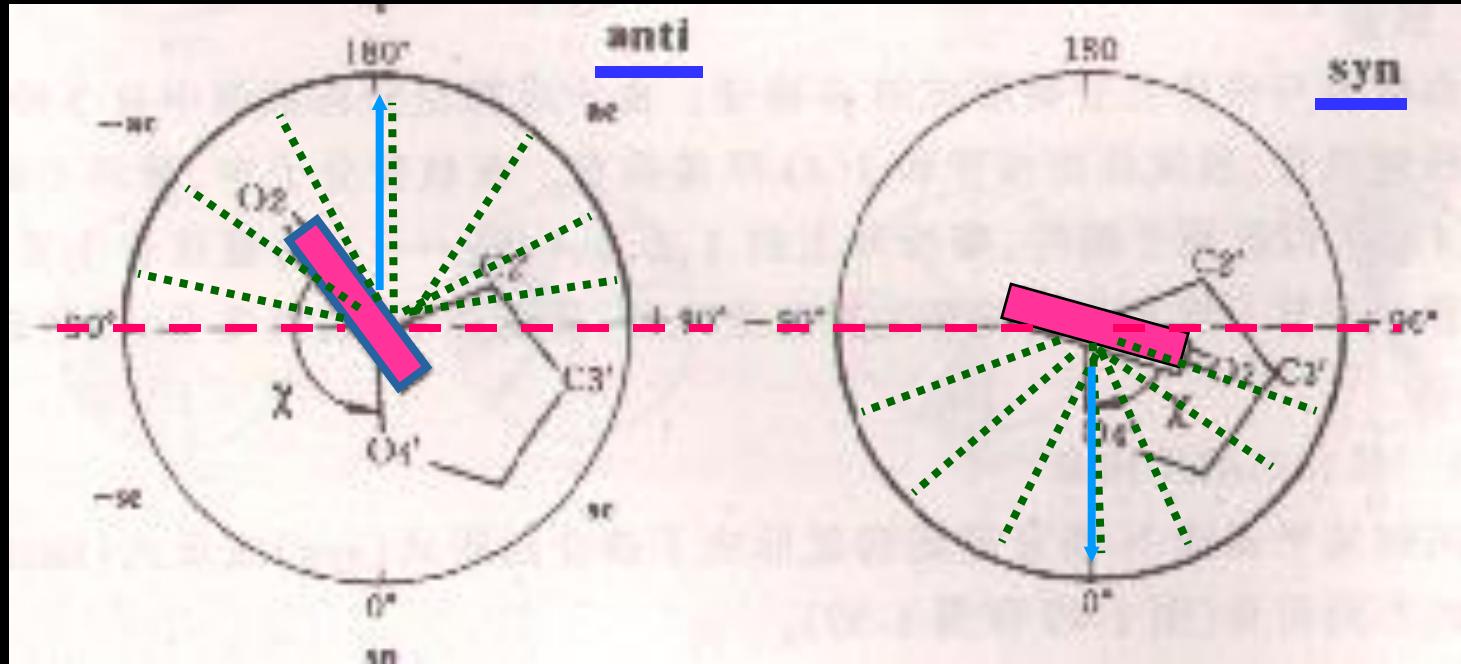
$$\chi = \text{C2-N1-C1}'-\text{O4}' \\ 0^\circ$$

(来源: 分子生物学 (2007) , 郑用琏, 第24页)



(来源: 分子生物学 (2007), 郑用琏, 第26页)

Pyrimidine (py)



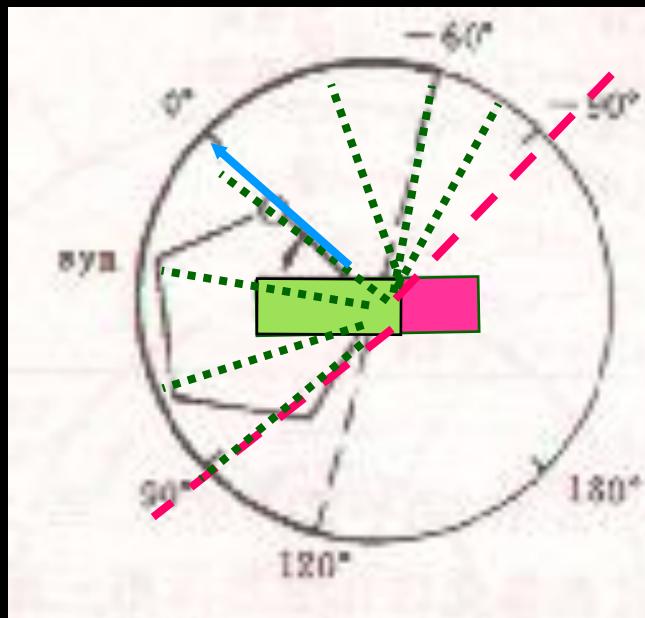
$$\chi = 180^\circ \pm 90^\circ$$

$$\chi = 0^\circ \pm 90^\circ$$

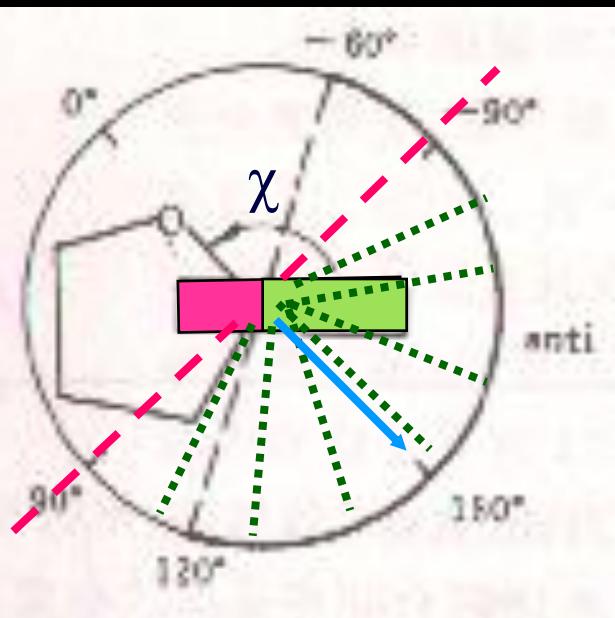
(来源: 分子生物学 (2007), 郑用琏, 第25页)

Purine (pu)

Syn



Anti



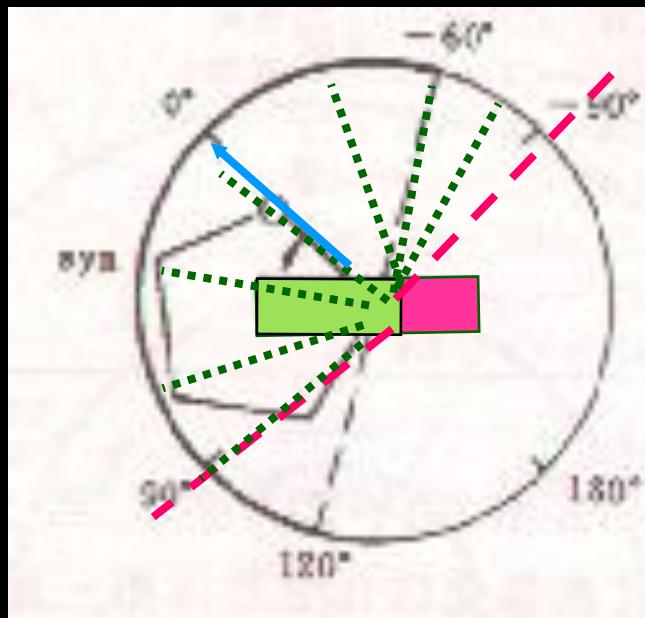
$$\chi = 0^\circ \pm 90^\circ$$

$$\chi = 180^\circ \pm 90^\circ$$

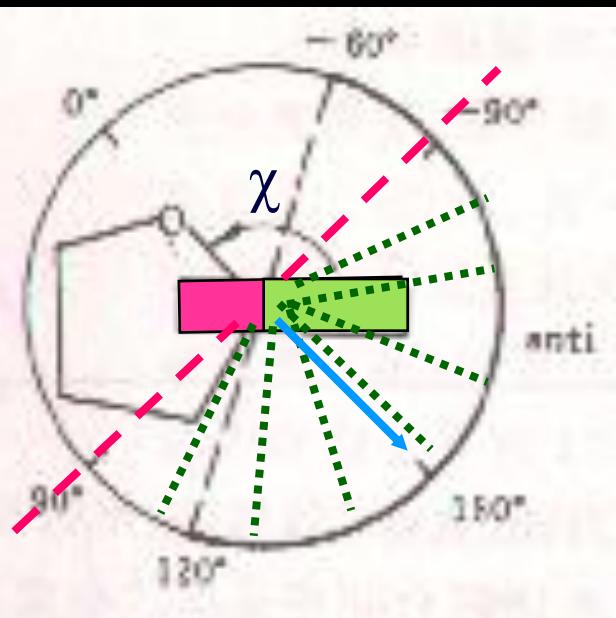
(来源: 分子生物学 (2007), 郑用琏, 第25页)

Purine (pu)

Syn



Anti



$$\chi = 0^\circ \pm 90^\circ$$

$$\chi = 180^\circ \pm 90^\circ$$

(来源: 分子生物学 (2007), 郑用琏, 第25页)

2.2.3. DNA 双螺旋结构模型 (DNA Double Helix Model)

1938. W. T. Astbury

X-ray photograph of DNA



Erwin Chargaff b. 11 Aug
1905, Czernowitz, Austria
(now Chernovtsi, Soviet
Union)

1950. Chargaff

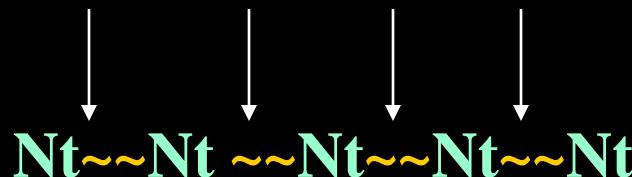
$$A + G / T + C = 1$$

$$A + T \neq G + C$$

Rich AT form & rich GC form

1952. Alexander Todd

3', 5' phosphodiester bond



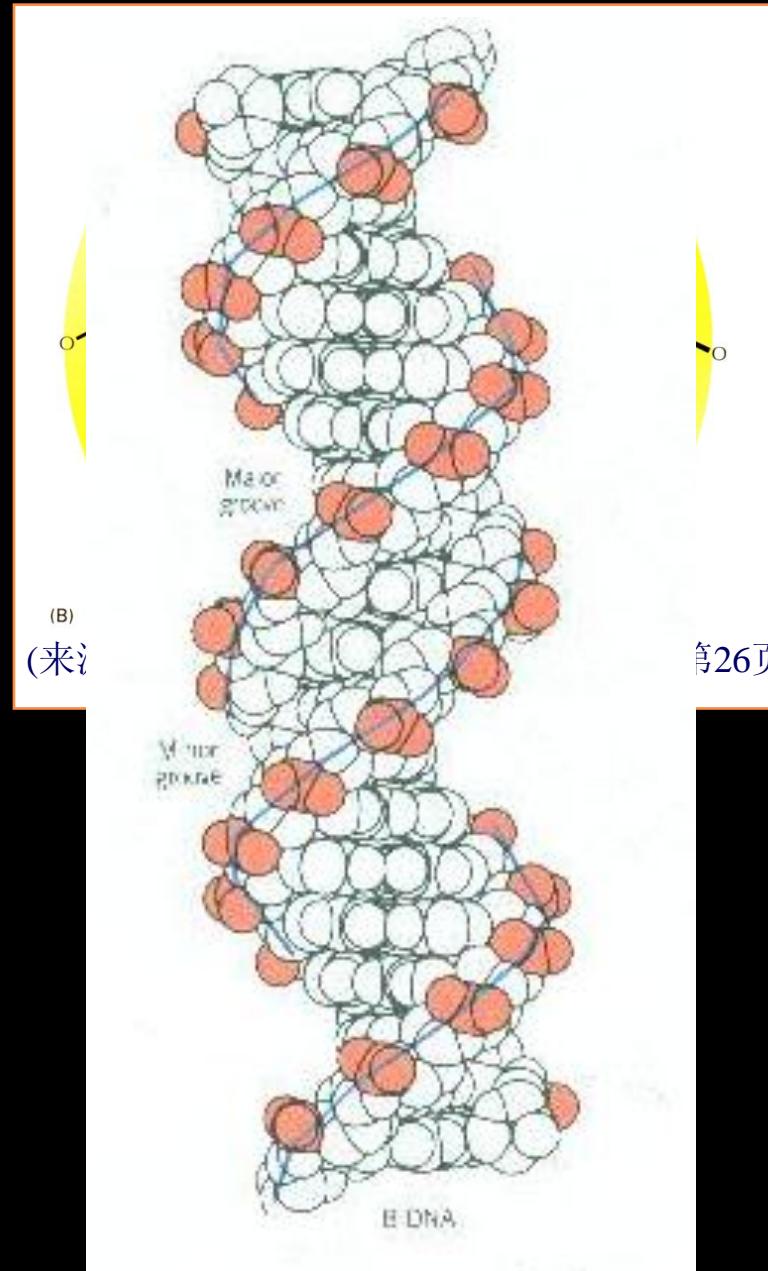
1953. Watson & Crick



(Source:Corbis/Bettmann Archive.)

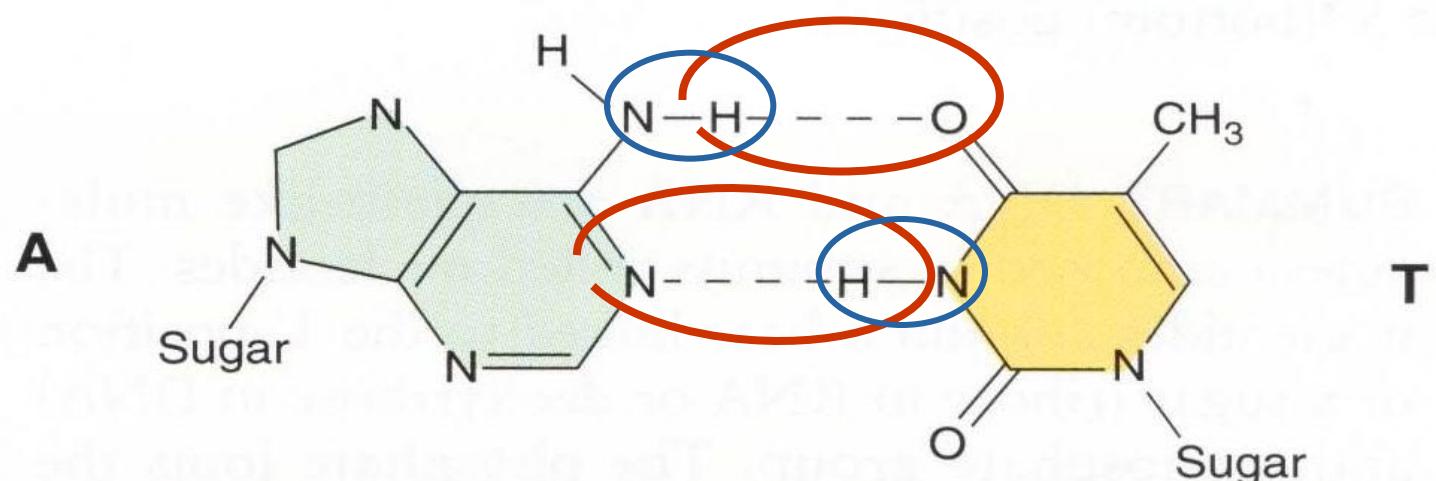
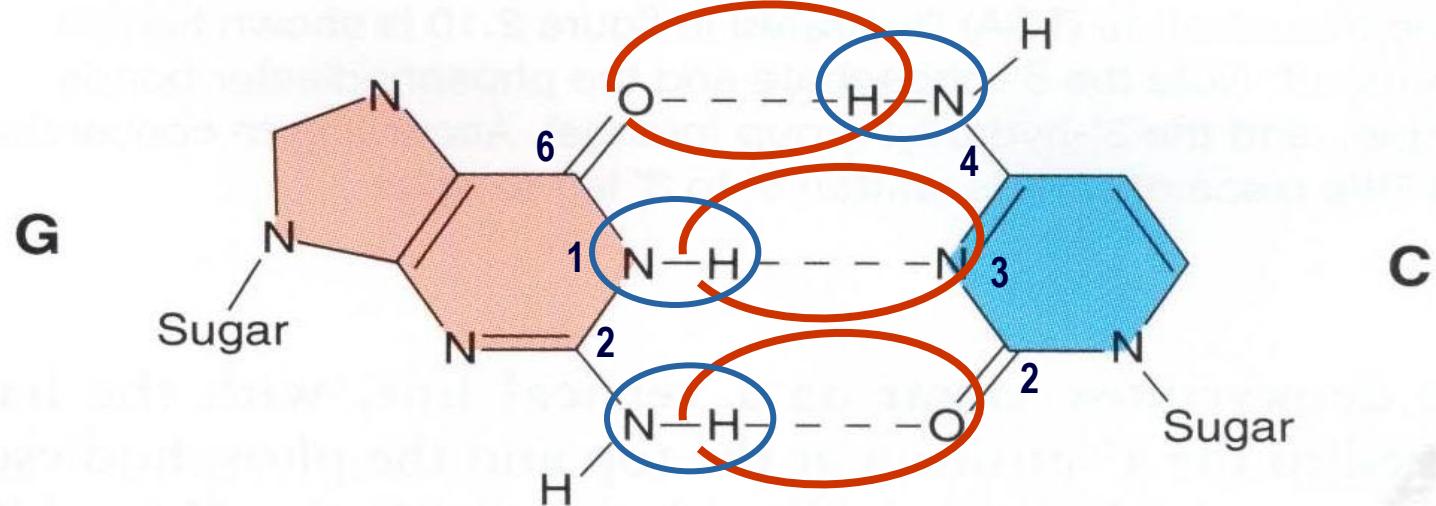
Right handed B-form DNA

Double helix Model



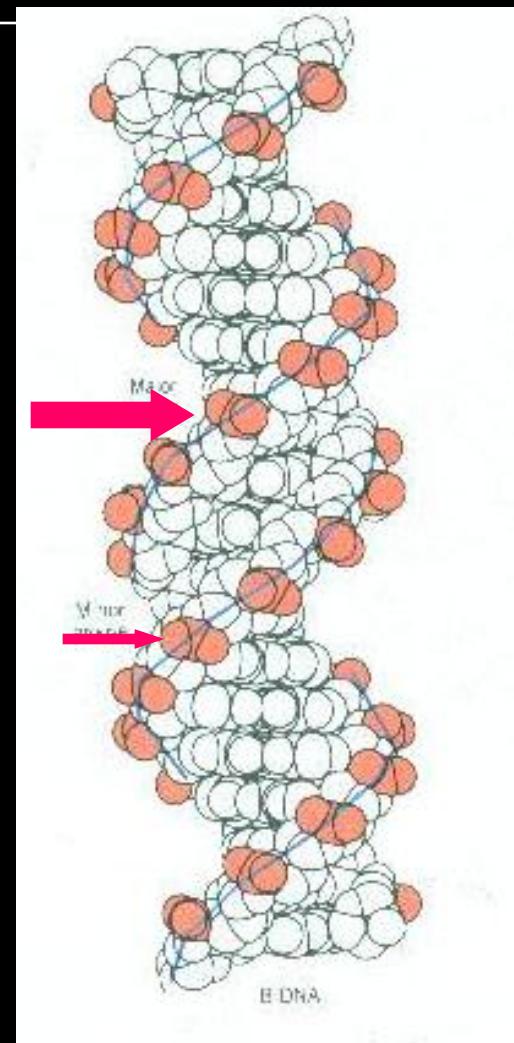
(来源: 分子生物学 (2007), 郑用琏, 第28页)

G-C pair, held together by three hydrogen bonds
A-T pair , held together by two hydrogen bonds

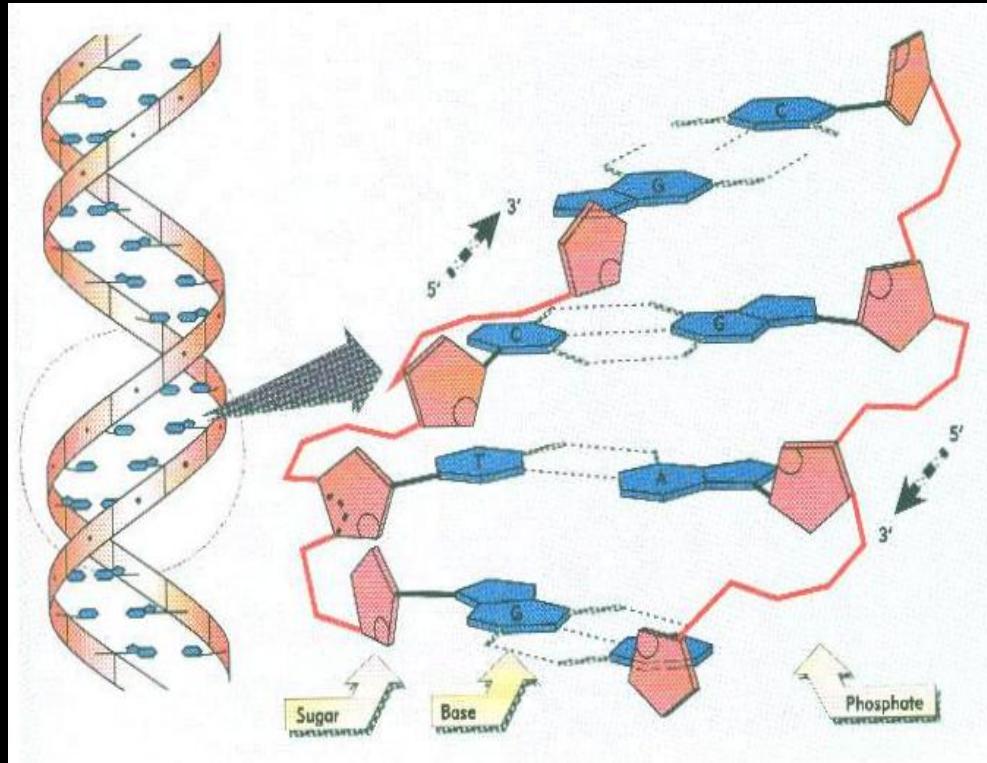


DNA双螺旋的结构特点

- 碱基顶部基团裸露在DNA 大沟内
- 蛋白质因子与DNA 的特异结合依赖于氨基酸与DNA 间的氢键的形成
- 蛋白质因子沿大沟与DNA形成专一性结合的机率与多样性高于沿小沟的结合
- 大沟的空间更有利于与蛋白质的结合



(来源：分子生物学（2007），郑用琏，第28页)



(来源：分子生物学（2007），郑用琏，第28页)

Right handed B-form DNA Double helix Model

- 每一单链具有 $5'$ → $3'$ 极性
- 两条单链间以氢键连接
- 两条单链，极性相反，反向平行
- 以中心为轴，向右盘旋 (B-form)
- 双螺旋中存在 大沟 (2.2nm)
小沟 (1.2nm)

2.2.4. 影响双螺旋结构稳定性的因素



碱基堆积的棒状实体

1 氢键 (Hydrogen bond 4~6 kc / mol)

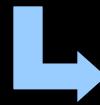
弱键, 可加热解链

氢键堆积, 有序排列(线性, 方向)

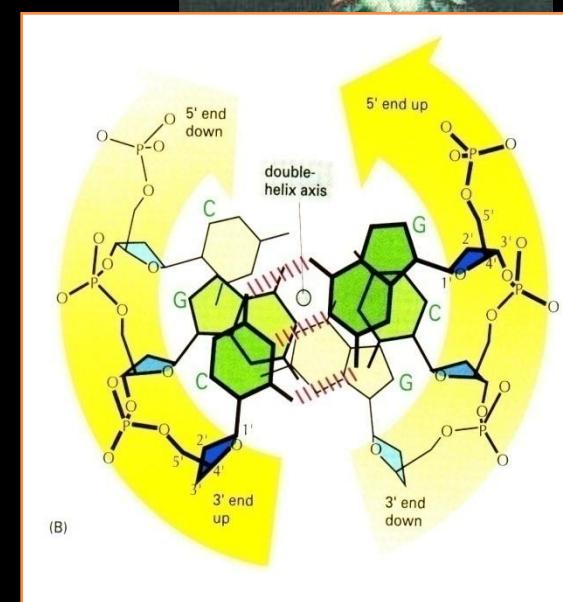
1 磷酸酯键 (phosphodiester bond 80~90 kc / mol)

强键, 需酶促解链

1 0.2 mol / L Na⁺ 生理盐条件



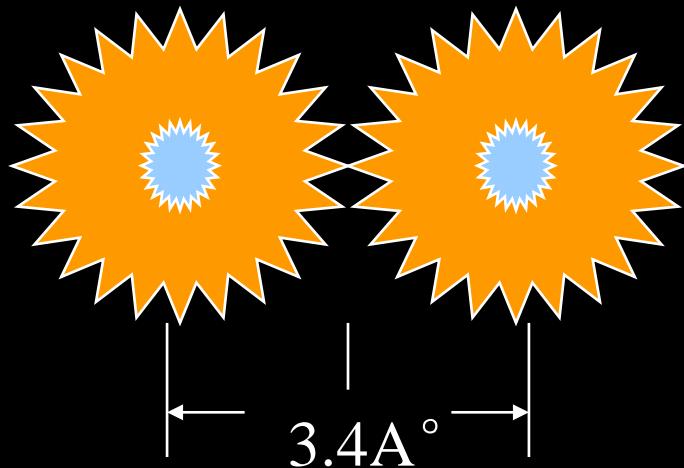
消除DNA单链上磷酸基团间的静电斥力



(来源: 分子生物学 (2007), 郑用琏, 第26页)

1 碱基堆积力(非特异性结合力)

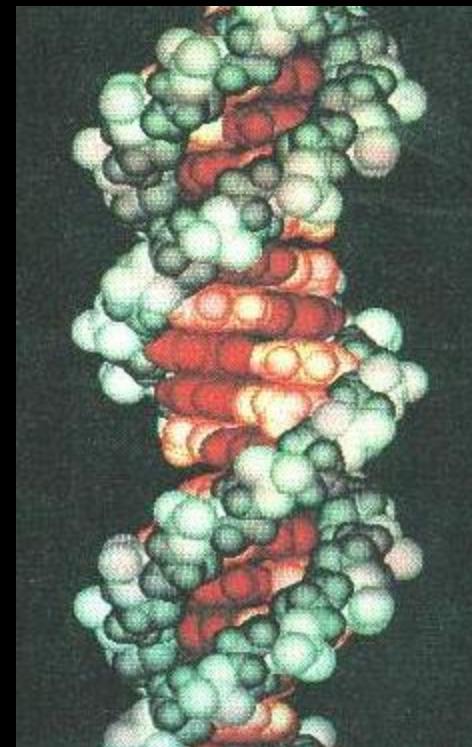
- ☆ 磷酸骨架, 氨基, 酮基周围水分子间的有序排列
- ☆ Van de waals force (1.7A° / 嘧啶环与嘌呤环作用半径)



(0.34 nm/碱基对间距)

$$(1 \text{ kc/mol} - 0.6 \text{ kc/mol}) \times n$$

(热运动消耗)



来源：不详

☆ 疏水作用力 (Hydrophobic interaction / π ~ π interaction)

不溶于水的非极性分子在水中相互联合，
成串结合的疏水作用力，即为熵 Entropy (ΔS)

DNA分子中非极性碱基的重叠，聚集。产生疏水作用力



成为碱基间的部分堆积力

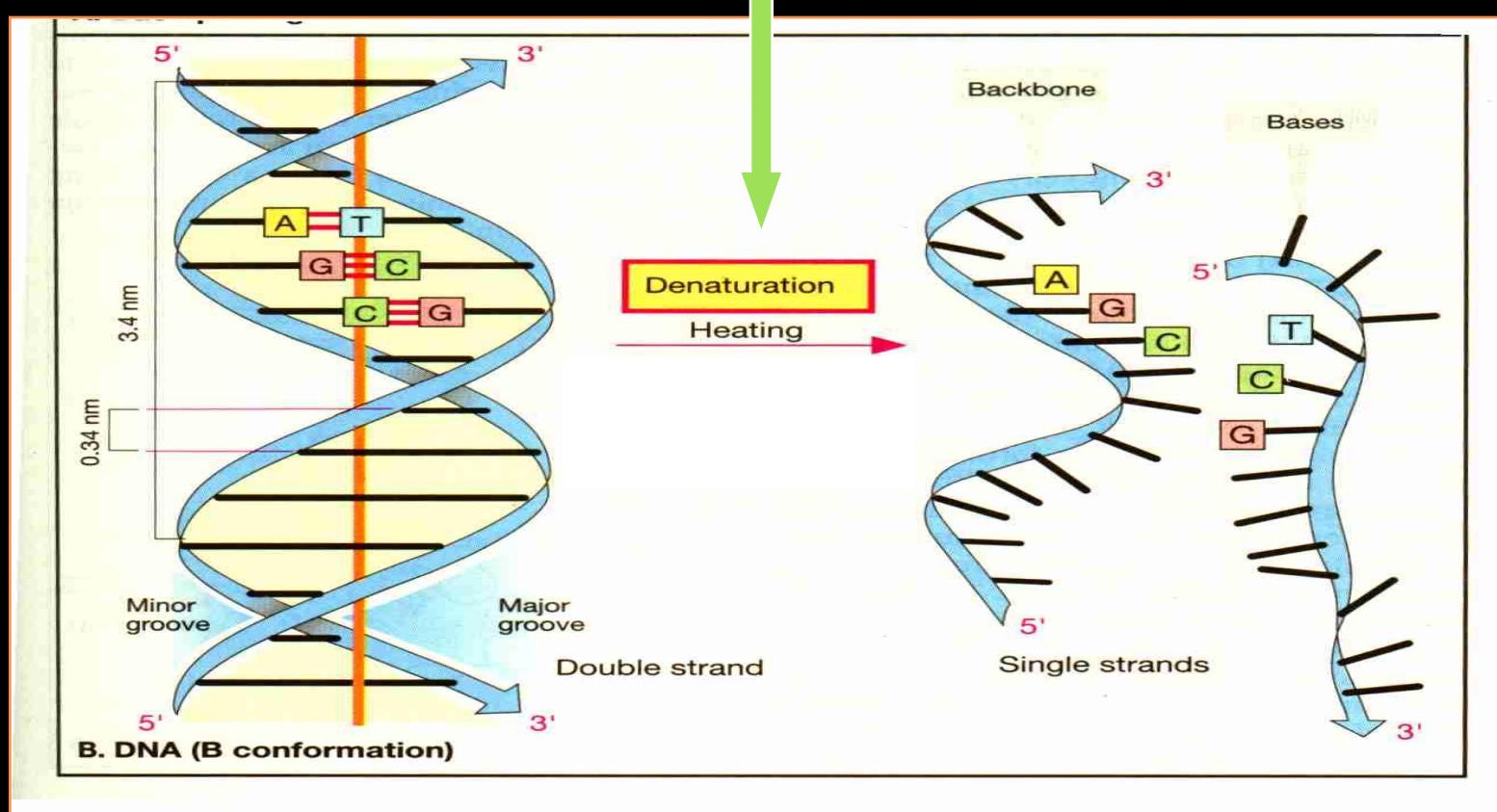
- 磷酸基团间的静电斥力
- 碱基间的挤压、抵御使其内能增加，碱基间有序排列的状态破坏（氢键作用力被减弱）

影响双螺旋结构稳定性的因素

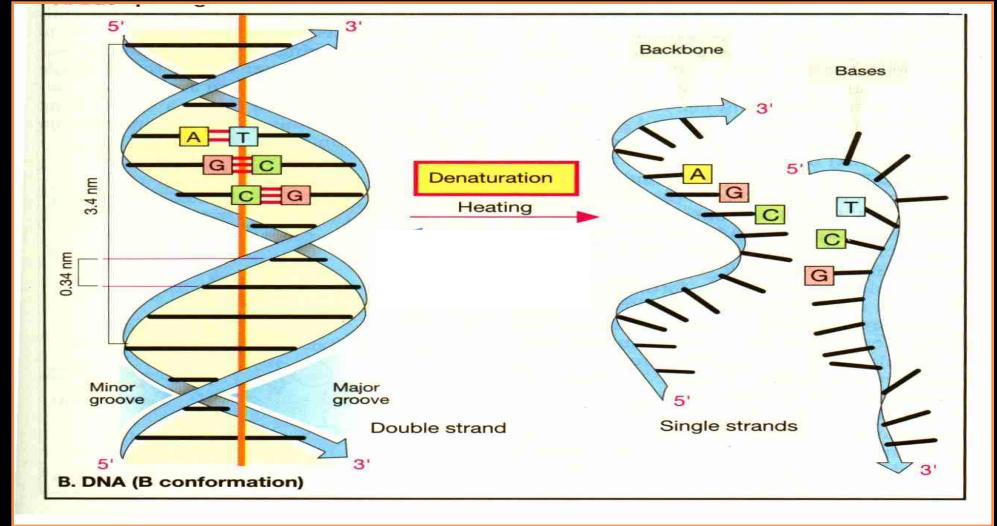
- 1 氢键 (Hydrogen bond 4~6 kc / mol)
- 1 磷酸酯键 (phosphodiester bond 80~90 kc / mol)
- 1 0.2 mol / L Na⁺ 生理盐条件
- 1 碱基堆积力 (非特异性结合力)
- 磷酸基团间的静电斥力
- 碱基间的挤压、抵禦使其内能增加, 碱基间有序排列的状态破坏 (氢键作用力被减弱)

2.2.5. DNA分子变性(DNA denaturation)

- D.S. DNA → S.S. DNA
(加温, 极端pH, 尿素, 酢胺)



(来源: 分子生物学(2007), 郑用琏, 第30页)



(来源：分子生物学（2007），郑用琏，第30页)

变性过程的表现

★ S.S. DNA粘度降低

D.S DNA → S.S DNA 粘度降低 ?



溶液粘度取决于分子流动过程中的内摩擦和阻力

高分子溶液 > 普通溶液

线状分子 > 不规则线团 > 球形分子

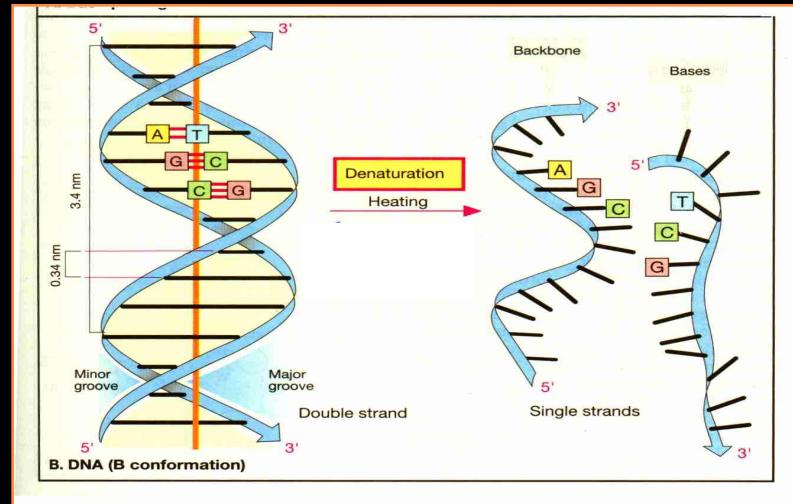
D.S. DNA 钢性较强，结构较为舒展的 Double helix

S.S. DNA 没有氢键的支撑

由螺旋结构向折叠和线团结构转变

D.S DNA → S.S DNA 粘度降低

变性过程的表现



(来源：分子生物学（2007），郑用琏，第30页)

☆ S.S.DNA 沉降速度加快

☆ S.S.DNA分子的A 260 nm UV 值上升
(Hyperchromicity)

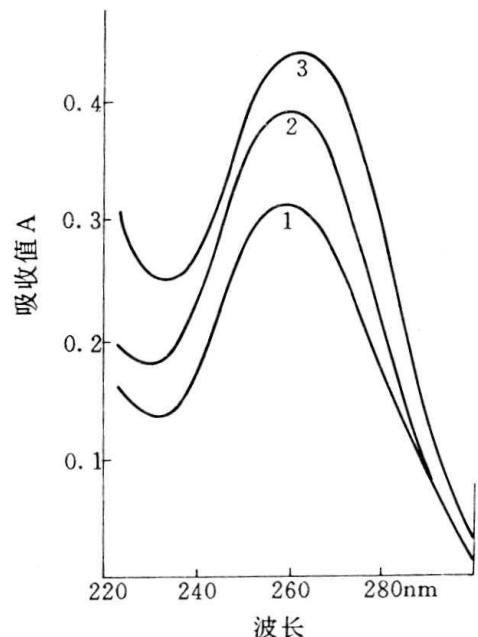
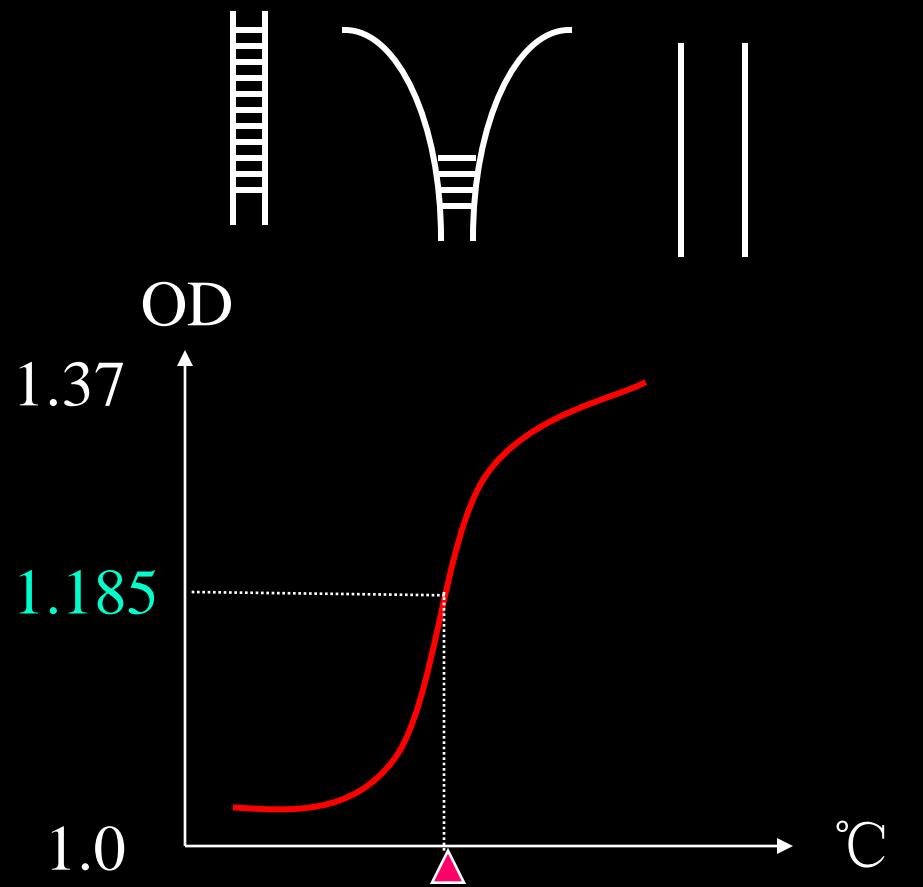


图 5-22 1. 双链 DNA, 2. 变性 DNA,
3. 核苷酸的紫外吸收

(来源: 分子生物学 (2007) ,
郑用琏, 第30页)



(来源: 分子生物学 (2007) , 郑用琏, 第31页)

Tm (melting temperature) = midpoint of the temperature range over which DNA is denatured

= OD增加值的中点温度(一般为85-95°C)

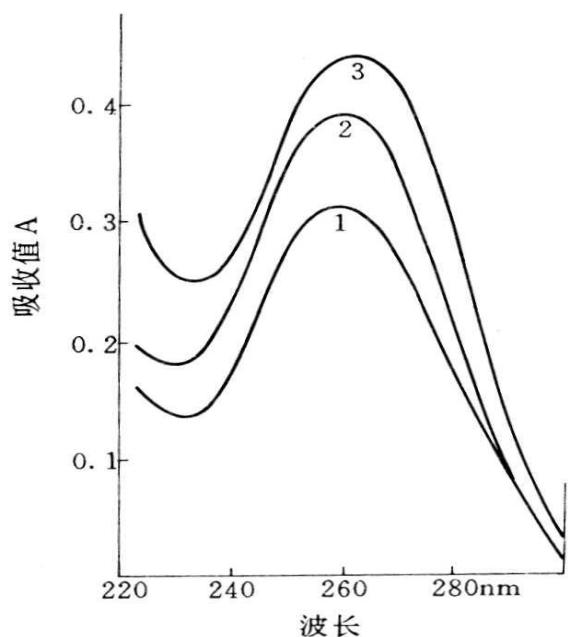
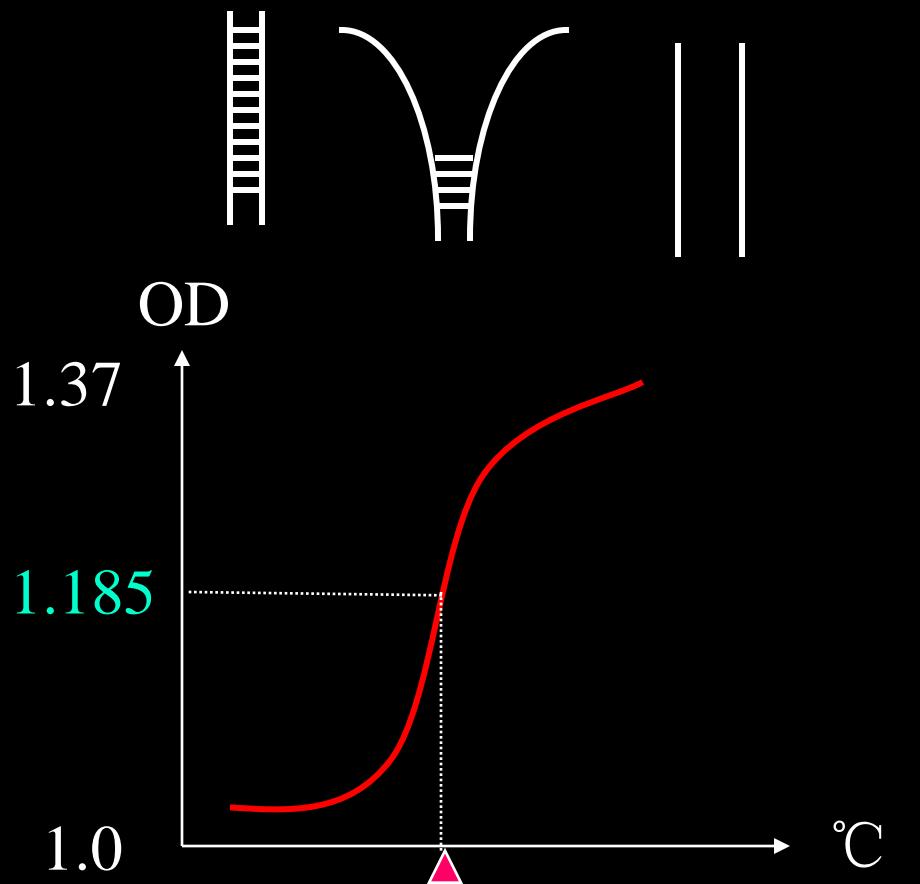


图 5-22 1. 双链 DNA, 2. 变性 DNA,
3. 核苷酸的紫外吸收

(来源: 分子生物学 (2007), 郑用琏,
第30页)



(来源: 分子生物学 (2007), 郑用琏, 第31页)

T_m (melting temperature) = midpoint of the temperature range over which DNA is denatured

= OD增加值的中点温度(一般为85-95°C)

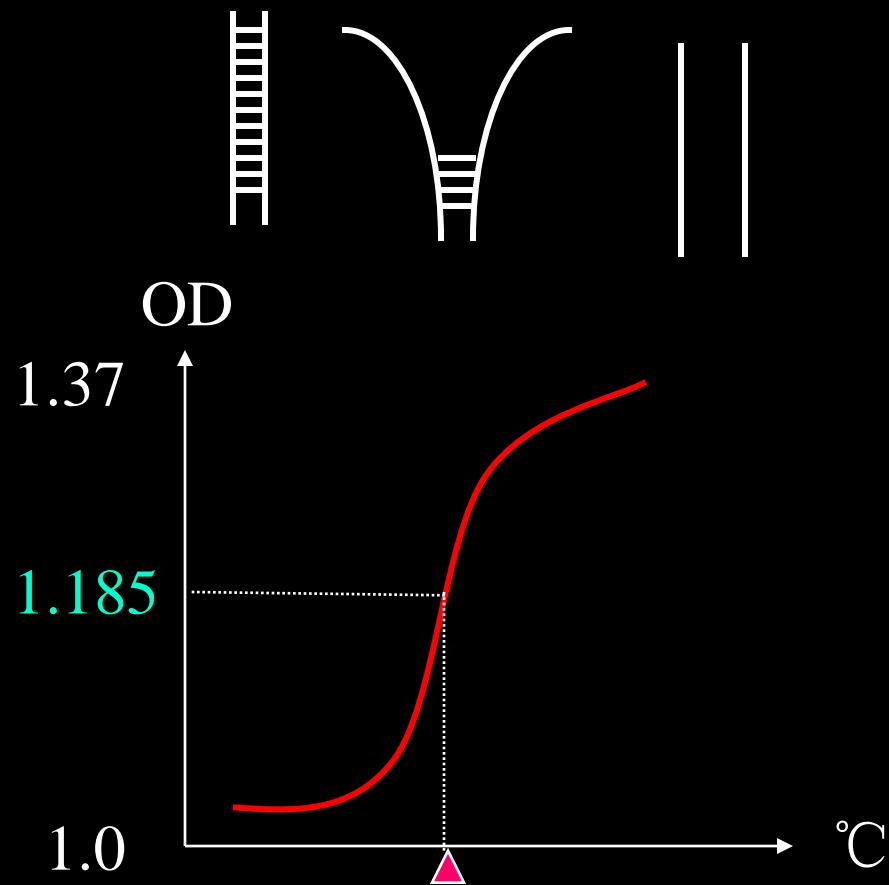
Concentration 50 μ g/ml

Optical Density
dNTPs $A_{260} = 1.60$

S.S DNA $A_{260} = 1.37$

D.S DNA $A_{260} = 1$

芳香环的不断暴露



(来源: 分子生物学 (2007), 郑用琏, 第31页)

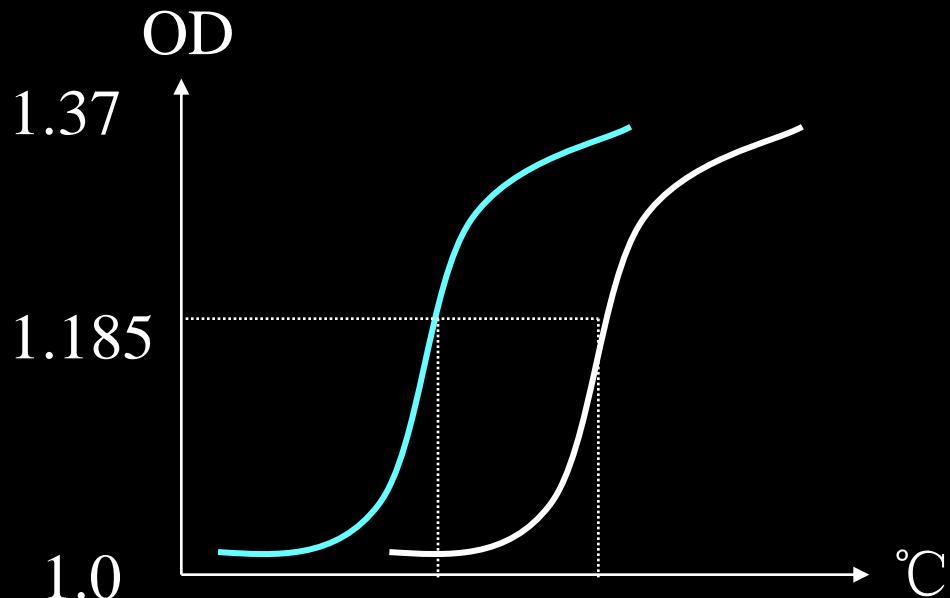
T_m (melting temperature) = midpoint of the temperature range over which DNA is denatured

= OD增加值的中点温度(一般为85-95°C)

Marmur-Doty formula

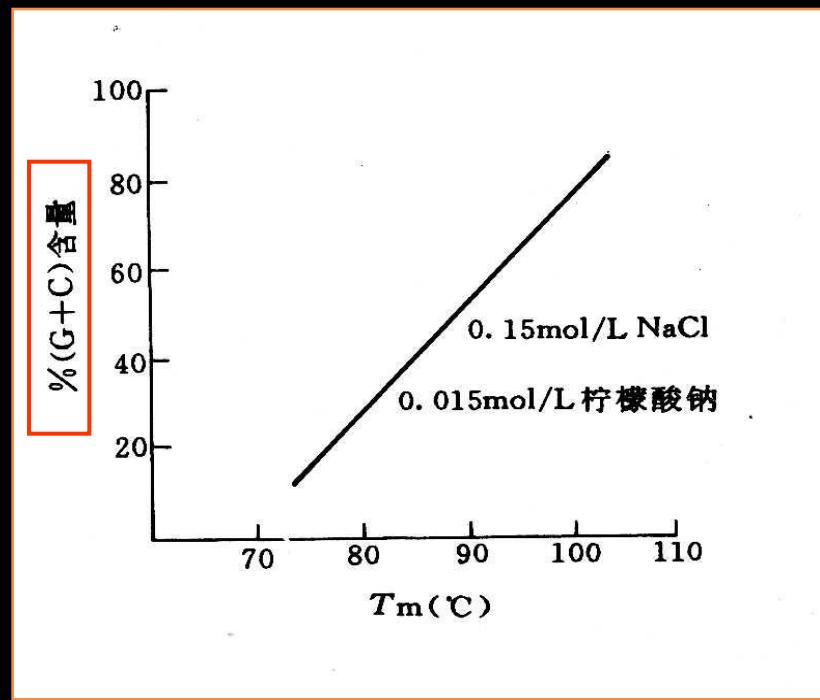
1x SSC (0.15 M NaCl + 0.015 M Sodium citrate)

GC% = 30-70%



(来源: 分子生物学 (2007), 郑用琏, 第31页)

$$Tm = 69.3 + 0.41 \times (G+C)\%$$

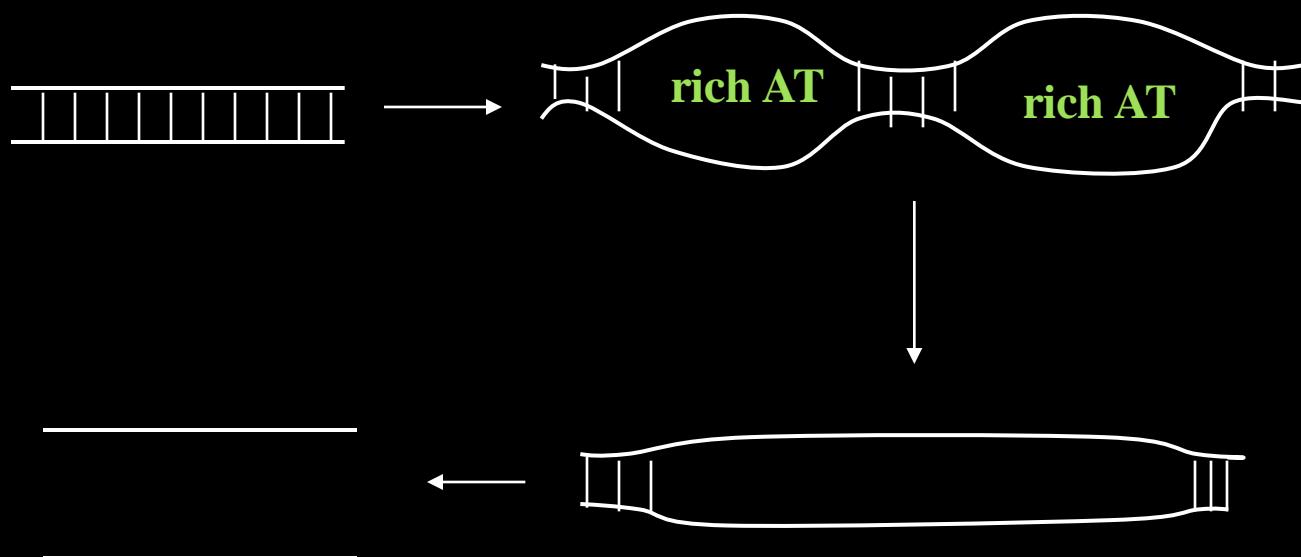


Evaluation GC% of DNA
Tm of PCR design

(来源: 分子生物学 (2007), 郑用琏, 第32页)

- 增色效应的跳跃现象 (Jump of Hyperchromicity)

高分子量的DNA分子在热变性过程中, 富含AT区域首先发生变性, 然后逐步扩展, 表现增色效应的跳跃现象, 使变性过程加快.



(来源: 分子生物学 (2007), 郑用琏, 第31页)

● 影响 Tm值的因素

☆ 在 A, T, C, G 随机分布的情况下

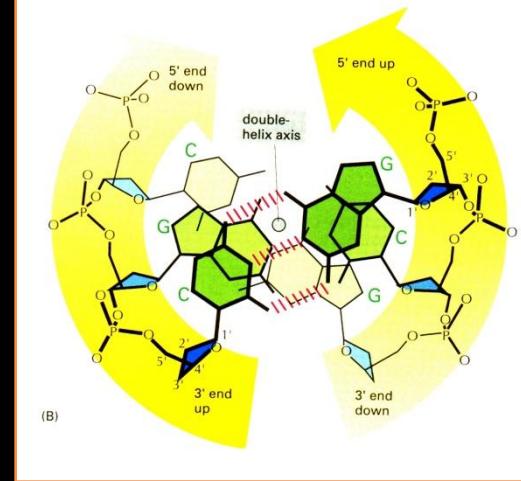
GC%愈高 → Tm值愈大

GC%愈低 → Tm 值愈小

☆ GC%含量相同的情况下

AT形成变性核心，变性加快，Tm 值小

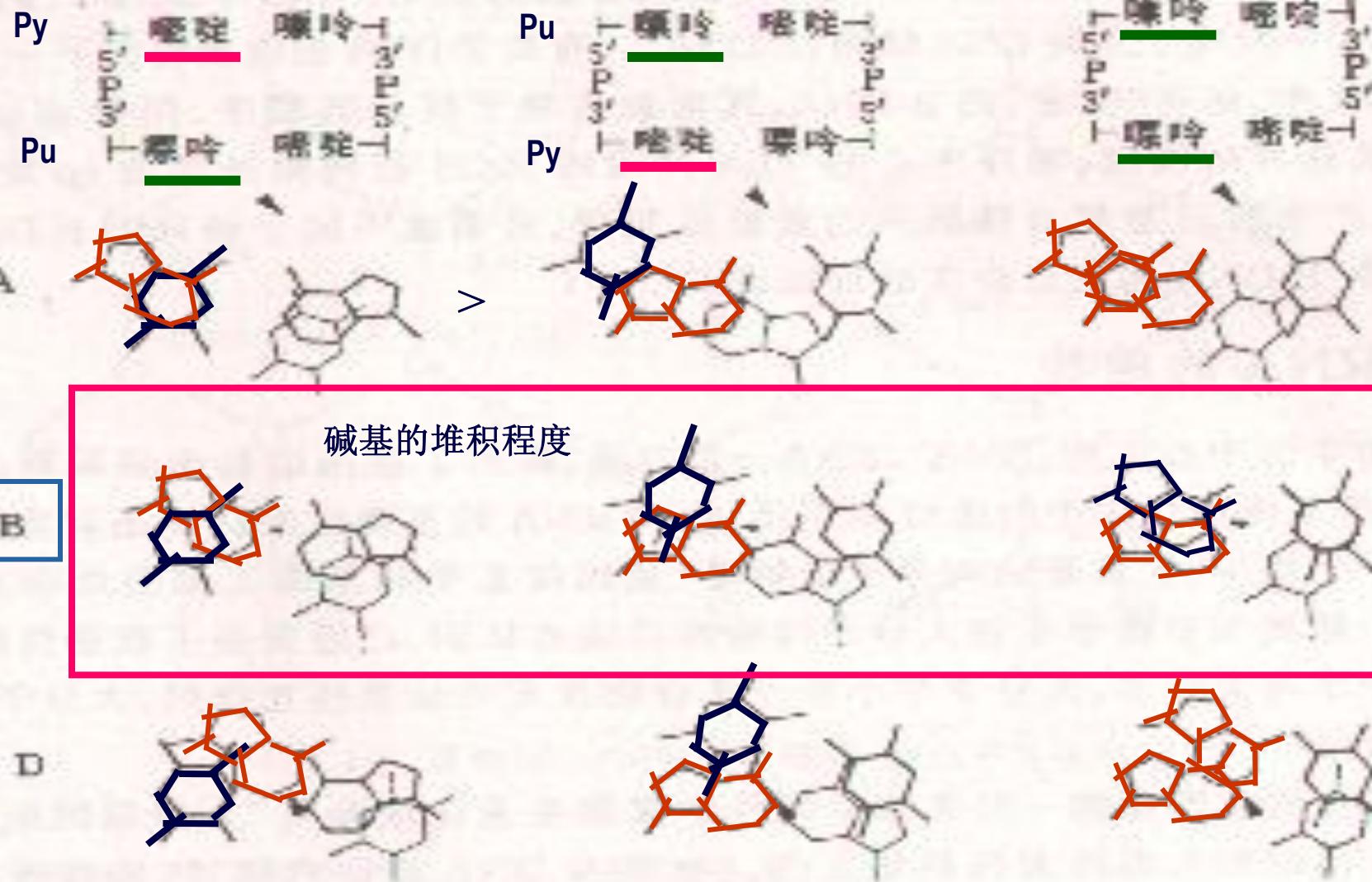
碱基排列对Tm值具有明显影响（除变性核心外）
(相同碱基组成,但不同排列,堆积力的差异)



(来源：分子生物学（2007），
郑用琏，第31页)

碱基的堆积 → 使官能团发生**重叠**、
挤压的程度有异
↑

因核苷酸序列不同而异

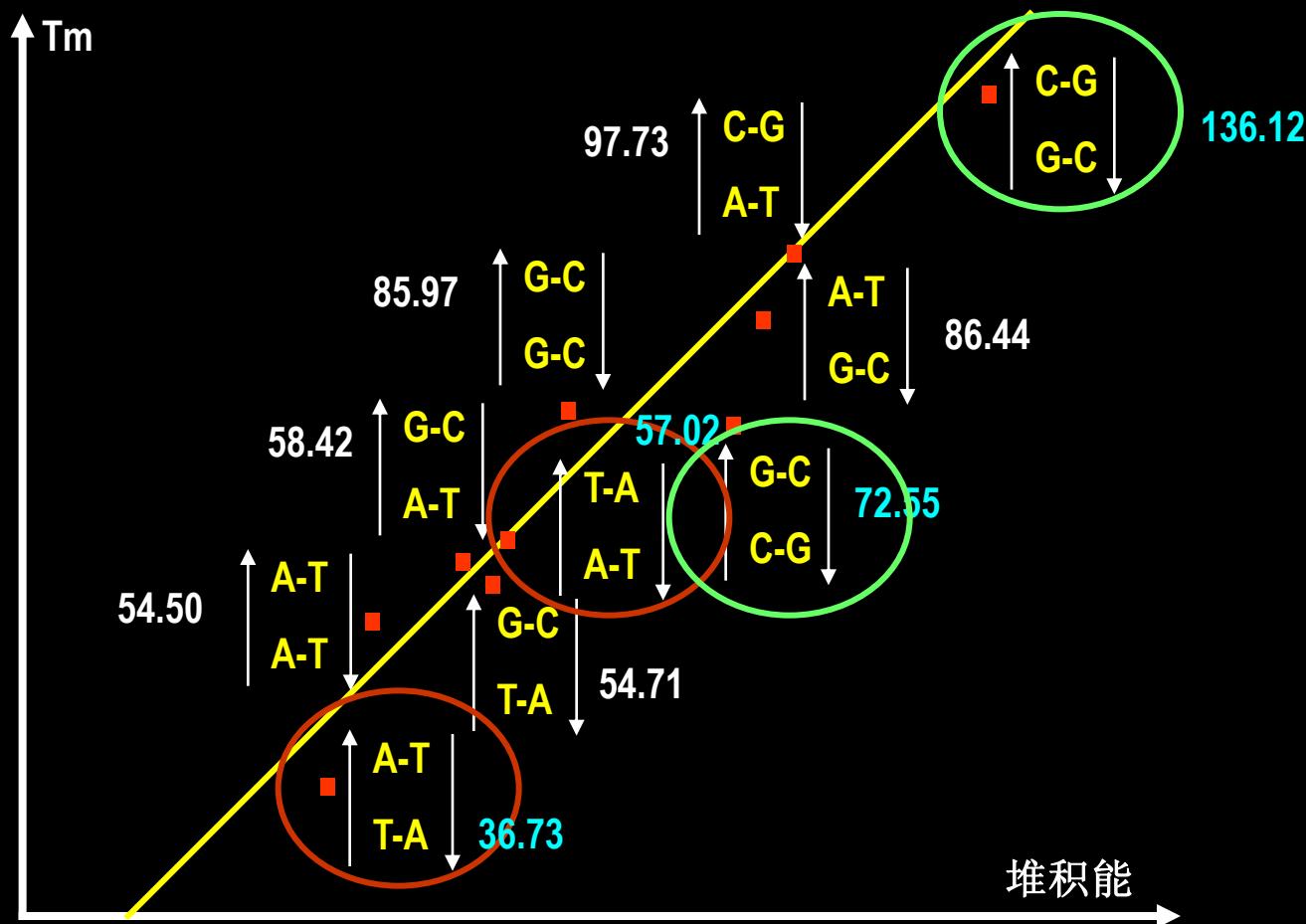


(来源: 分子生物学 (2007), 郑用琏, 第33页)

不同构向DNA分子中的碱基堆积的程度

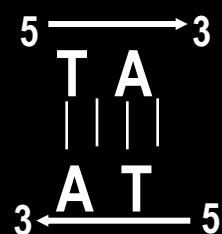
从5'到3'方向，碱基堆积分力的大小，与重叠面积正相关

与一个碱基的-NH₂, -N=, =O, 卤素基团与邻近碱基的芳香环互作相关

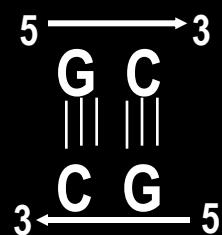


DNA的呼吸现象

TATA Box



UAA



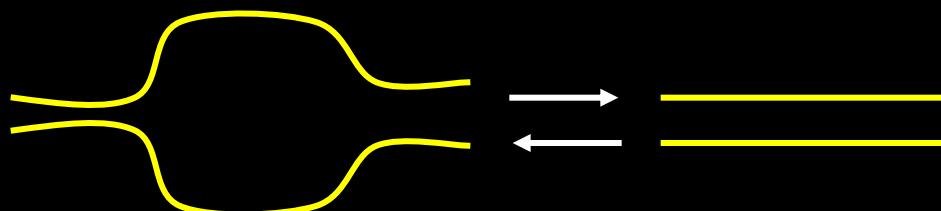
小

大

堆积力

稳定性

Tm值



☆ 大片段D.S. DNA分子之间比较

片段长短对Tm值的影响较小,与组成和排列相关

☆ 小于100bp 的 D.S DNA分子比较

片段愈短, 变性愈快, Tm值愈小

☆ 变性液中含有尿素, 酰胺等 (变性凝胶)

尿素, 酰胺与碱基间形成氢键



改变碱基对间的氢键



Tm值可降至40℃左右

☆ 盐浓度的影响

单链DNA主链的磷酸基团



负电荷的静电斥力

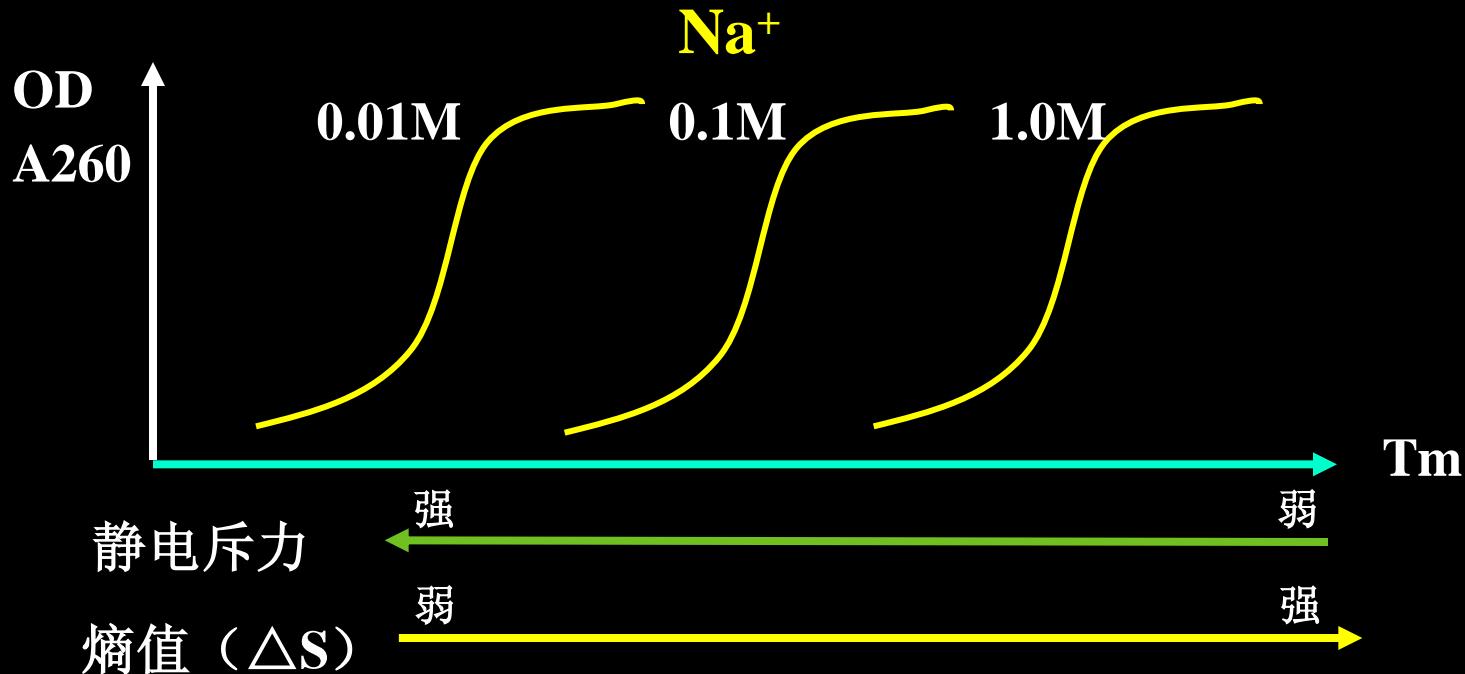


两条单链DNA的分离

Na^+ 在磷酸基团周围形成的电子云
对静电斥力产生屏蔽作用



减弱静电斥力



(来源: 分子生物学 (2007), 郑用琏, 第34页)

当Na⁺浓度低 → 屏蔽作用小 → 斥力加强 → Tm ↓

当Na⁺浓度高

→ { 屏蔽作用大 → 斥力减弱
碱基溶解性降低 → 熵值(ΔS)上升
疏水作用力增加 } → Tm ↑

☆极端pH条件的影响

pH ~ 12 酮基 → 烯醇基



改变氢键的形成与结合力

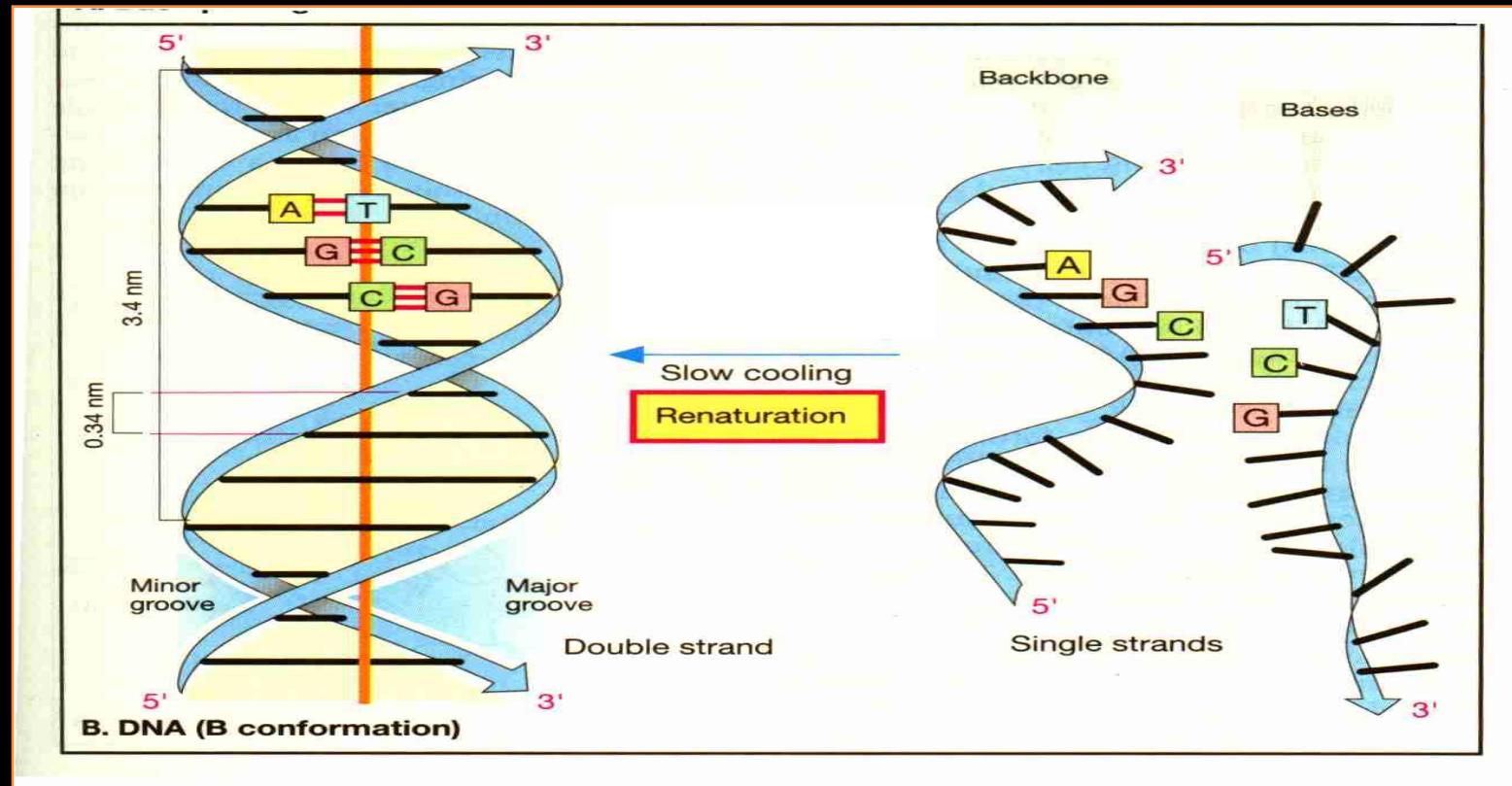


pH ~ 2-3 NH_2 → NH_2^+ (质子化)

一切减弱氢键， 碱基堆积力的因素

均将使Tm 值降低

2.2.6. DNA 分子的复性 (anneal or renaturation)



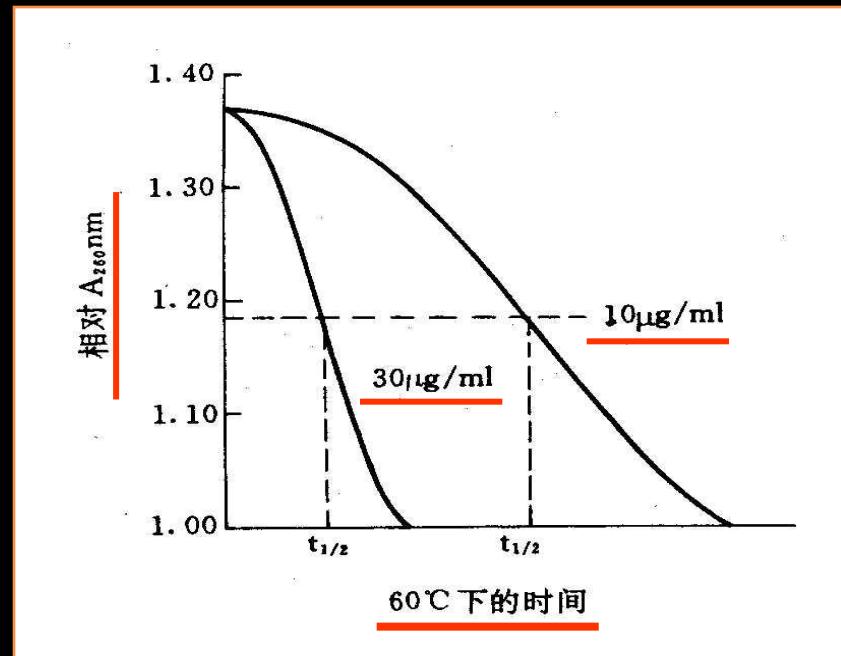
(来源: 分子生物学 (2007), 郑用琏, 第30页)

影响DNA复性过程的因素

- 阳离子浓度
 $0.18 \sim 0.2M\ Na^+$ 可消除poly-dNt 间的静电斥力
- 复性反应的温度 $T_m - 25^\circ C$ ($60-65^\circ C$)
 以消除S.S. DNA 分子内的部分二级结构
- S.S. DNA分子的长度
 S.S. DNA愈长 → 分子扩散愈慢 → 复性愈慢
 S.S. DNA愈短 → 分子扩散愈快 → 复性愈快

影响DNA复性过程的因素：

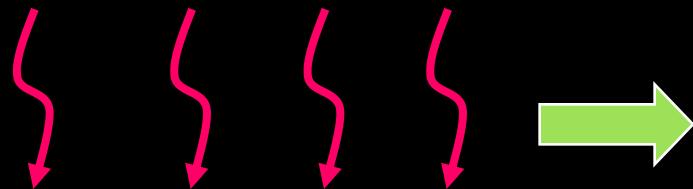
- S.S, DNA 的初始浓度 C_0



(来源：分子生物学（2007），郑用琏，第34页)

DNA 分子中, dNt 的排列状况
(随机排列, 重复排列)

3'-ATCTATGCTGTCAT-5'



3'-ATCTATGCTGTCAT-5'
5'-TAGATACGACAGTA-3'

3'-ATATATATATAT-5'

5'-TATATATATATA-3'



3'-ATATATATATAT-5'
5'-TATATATATATA-3'

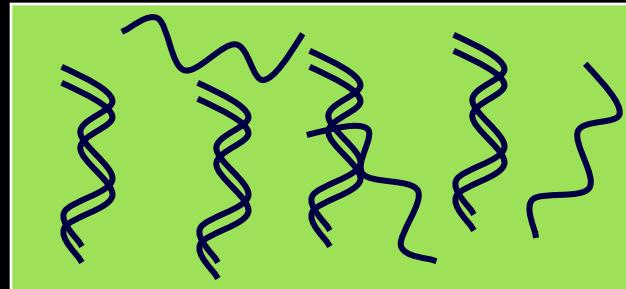
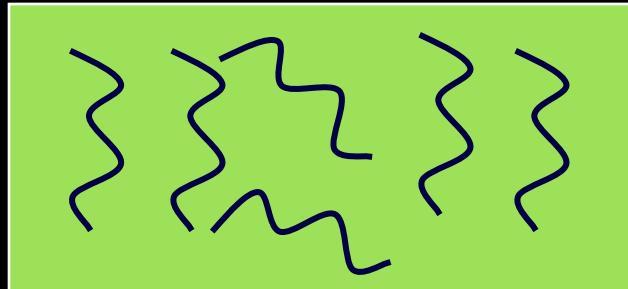
3'-ATATATATATAT-5'
5'-TATATATATATA-3'

3'-ATATATATATAT-5'
5'-TATATATATATA-3'

(来源：分子生物学（2007），郑用琏，第35页)

复性动力学的研究方法

“拉链效应” (大于20Nt配对)



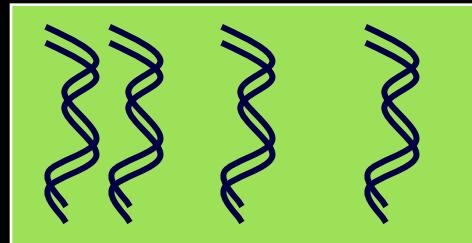
0.5M 磷酸盐
缓冲液



release D.S. DNA

Hydroxyapatite column

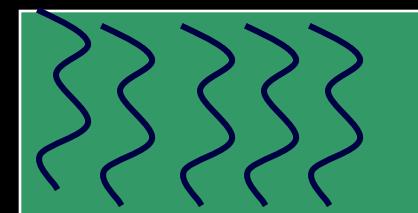
羟基磷灰石柱



0.12M 磷酸
盐缓冲液



absorb
D.S. DNA



复性发生过程的讨论

- 两条部分同源(小于 $20dNt$)的S.S. DNA, 在复性过程中形成的部分双链区是不稳定的
- 单链DNA的随机碰撞过程 (**randomly collision**)
拉链过程的时间长短

遵循*second-order kinetics formula* (二级反应动力学)

反应初始 $t = 0$

$$dC_t / dt = -kC_t^2$$

单链 DNA 浓度 = C_0

反应达 t 时

单链DNA浓度 = C_t

$$dC_t / dt = -k C_t^2 \quad \text{积分}$$



$$C_t / C_0 = 1 / (1 + k C_0 t)$$

当 $C_t / C_0 = 1/2$ 时



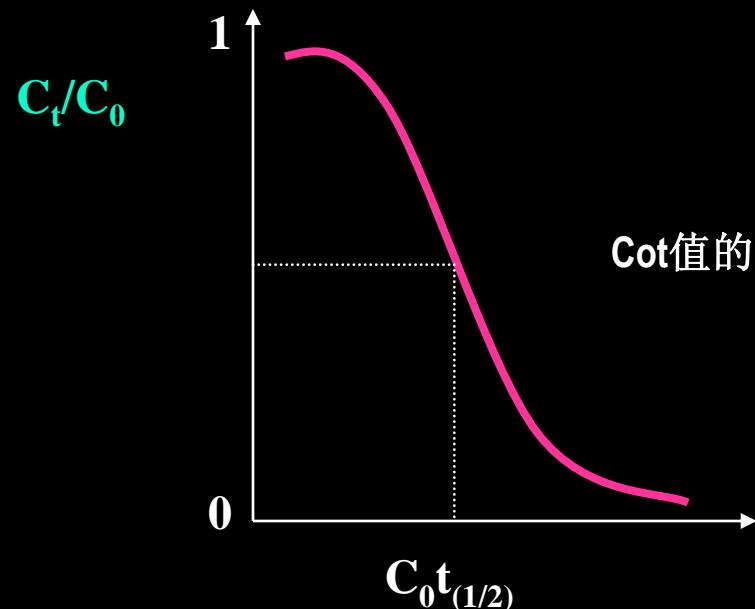
$$C_t / C_0 = 1/2 = 1 / (1 + k C_0 t_{(1/2)})$$

$$k (\text{L} / \text{mol. Sec}) = 1 / C_0 t_{(1/2)}$$

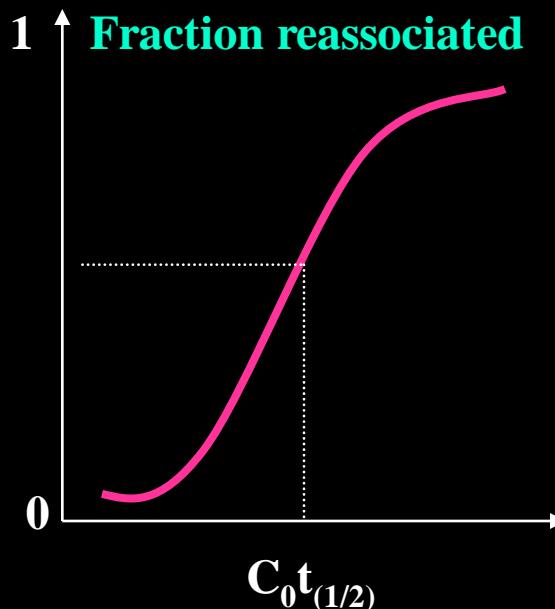


$$C_0 t_{(1/2)} = 1 / k (\text{mol. Sec} / \text{L})$$

任一DNA分子达到
 $C_t / C_0 = 1/2$ 的速率
是定值



Cot值的对数函数图



(来源：分子生物学（2007），郑用琏，第31页)

在控制反应条件相同的情况下，两种DNA分子的 $C_0 t_{(1/2)}$ 值，

取决于dNt 的排列复杂性：（复性动力学的复杂性，*Kinetic complexity, K.C.*）单一排列核苷酸的长度

AAAAAA

K. C. = 1

$C_0 t_{(1/2)} = 2 \times 10^{-6}$

ATCGATCGATCG

K.C. = 4

K.C. = 5×10^5 $C_0 t_{(1/2)} = 1$

2.2.7. 核酸分子的空间结构

一级结构； DNA分子的核苷酸组成与排列（sequence）

1977 F. Sanger $\varphi X174$ 的全序列分析

二级结构； D. S. DNA Double Helix model

A, B, Z...., Linear , circle...

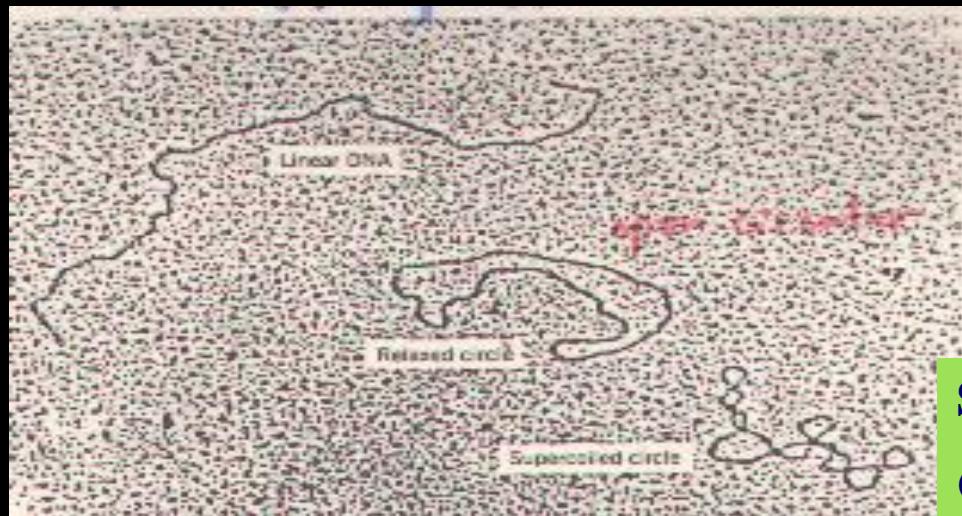
T. S. DNA (Triplex Strands)

Tet. S DNA (Tetraplex Strands)

三级结构； （高级结构）

superhelix DNA (Negative , Positive)

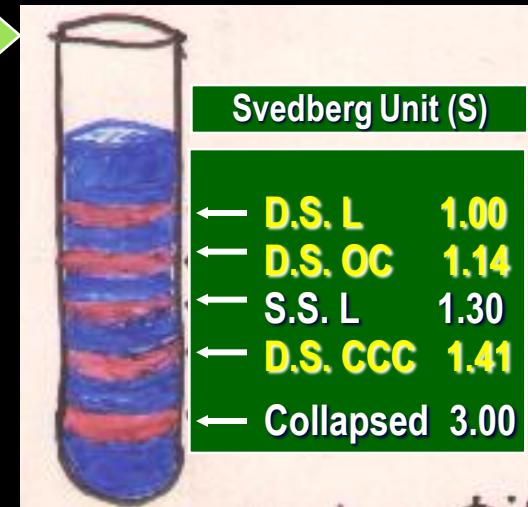
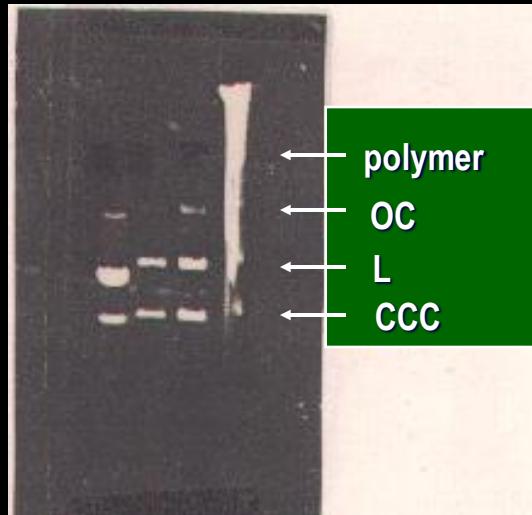
● 二级结构的形态



OC relaxed form

Supercoiled circle(高级结构)
Covalent Closed Circle CCC

来源: 不详



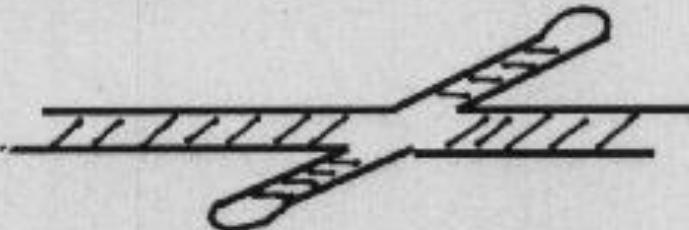
来源: 不详

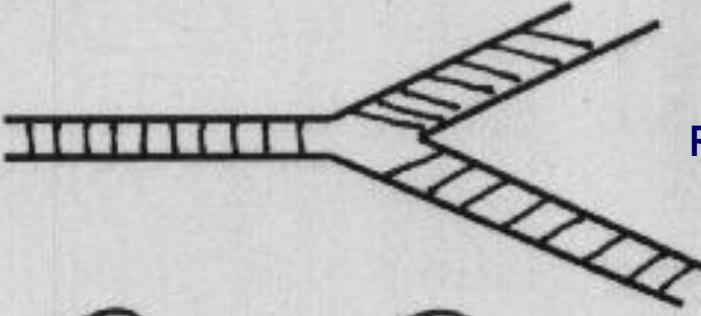
来源: 不详

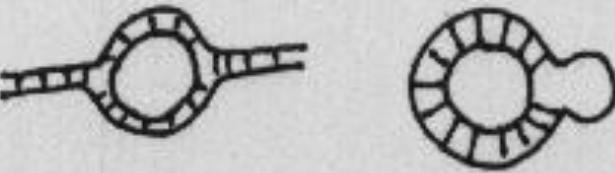
核酸分子的二级结构 (分别出现在DNA复制, 转录, 重组等阶段)

核酸分子的二级结构

L type (线型) DNA =  RNA = 

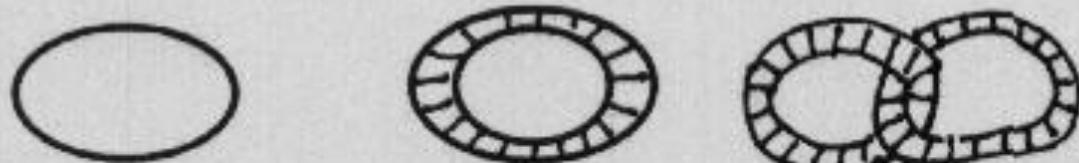
B type (分支型)  Inverted repeat

Y type (叉型)  Replication

D type (D 环)  Replication

核酸分子的二级结构 分别出现在DNA复制， 转录， 重组等阶段

C type (环型)



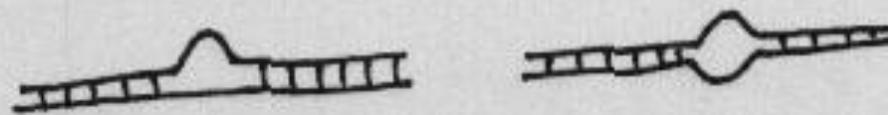
H type (发夹型)



K type (纽结型)



R type (环突)



三股螺旋， 四股螺旋



(来源：分子生物学（2007），郑用琏，第38页)

超螺旋结构DNA

绝大多数原核生物的DNA → 共价闭合环的分子结构

真核生物染色体 → 线形分子DNA + 组蛋白

多级螺旋

多个类似环型的结构

● 超螺旋结构DNA

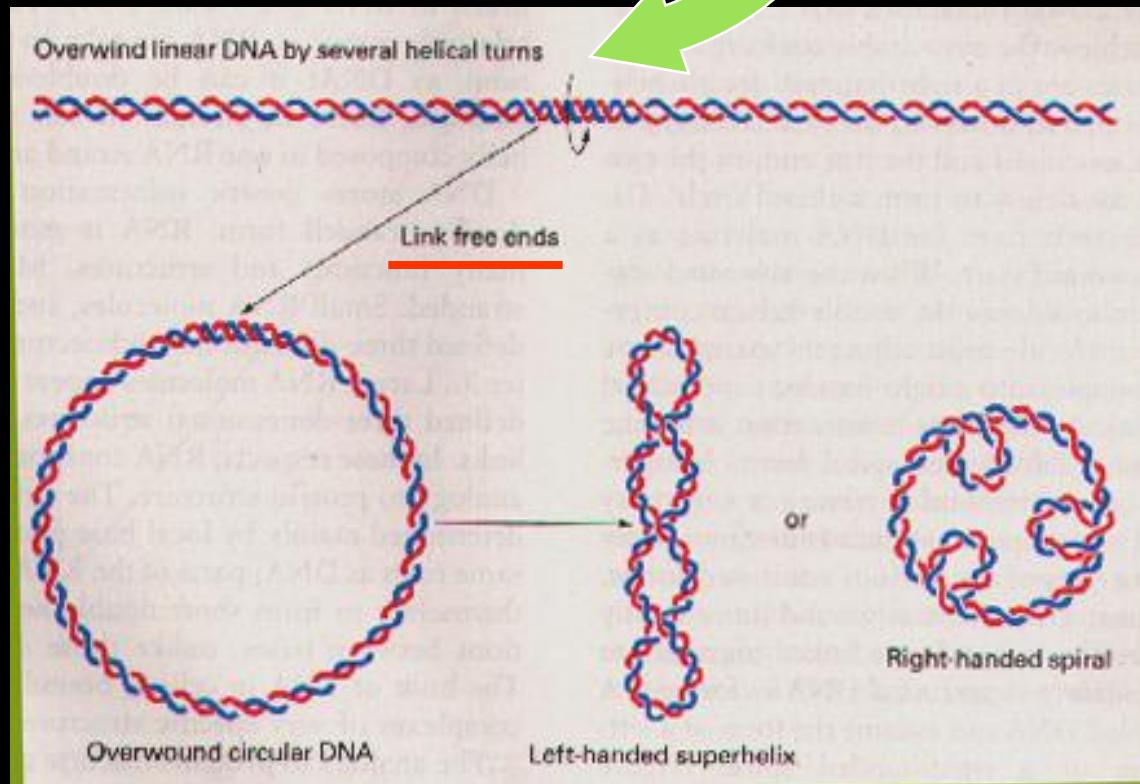
B-DNA



overwinding (右旋)

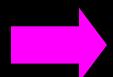


leads to
left-handed
superhelix



(来源：分子生物学（2007），郑用琏，第46页)
positive supercoiled

B-DNA

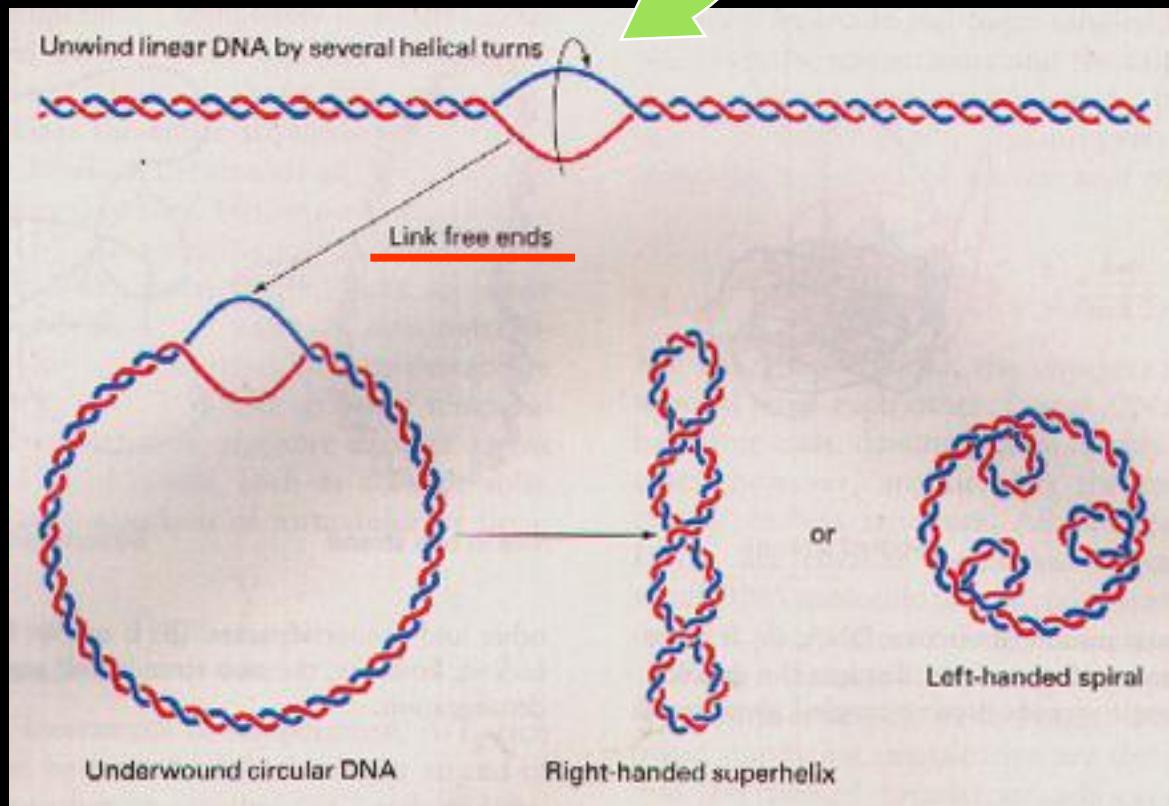


unwinding (左旋)



Leads to
right-handed
superhelix

Negative
Supercoiled



(来源: 分子生物学 (2007), 郑用琏, 第47页)

所有生物的DNA几乎有5%为 Negative Superhelix



超螺旋发生的规律

Vinograd. J (1968)

Vinograd equation

$$L = T + W$$

L Linking number

(双链DNA的交叉数)

Non-breaking

Non-unwinding

Non-overwinding

L为定值，整数

T Twisting number

(双链DNA的缠绕数，初级螺旋圈数，可为小数)

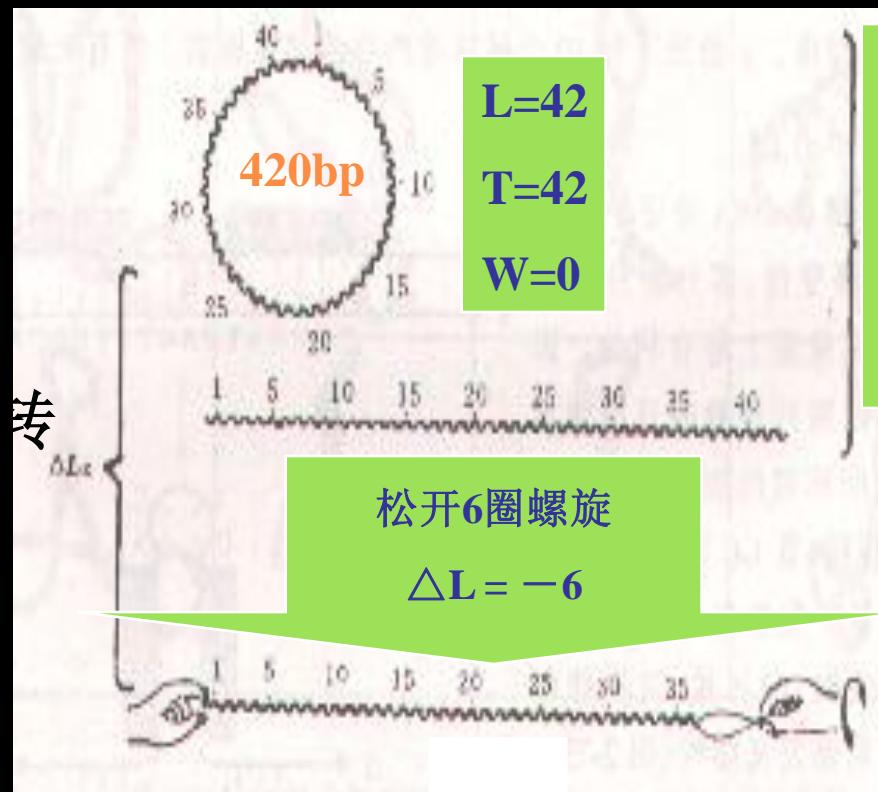
W Writhing number

(直观上为超螺旋数
可为小数)

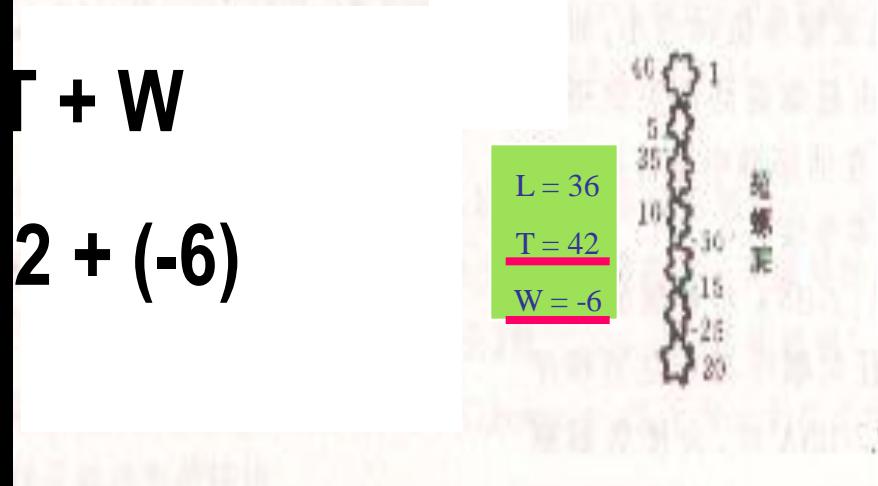
W = 负值 (negative superhelix)

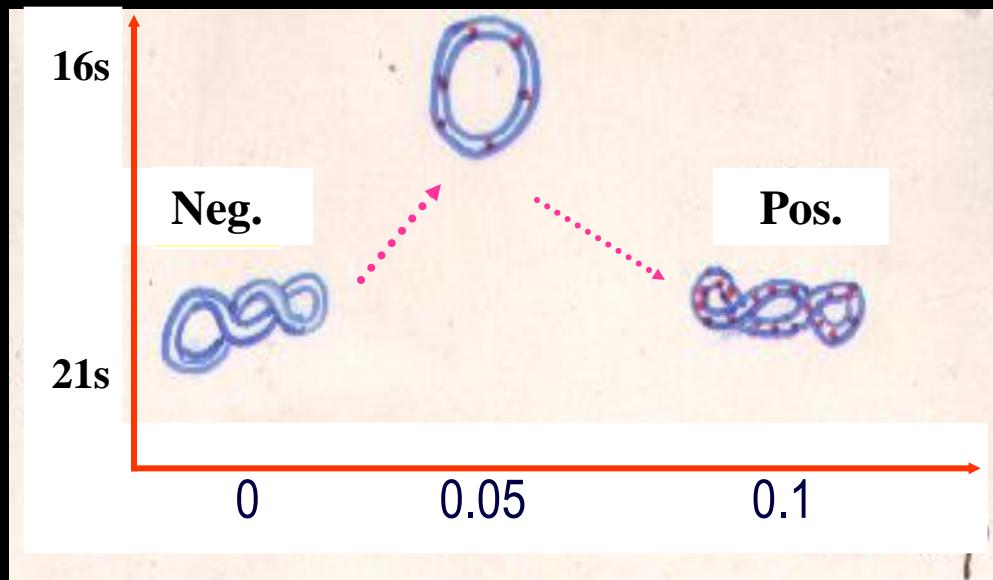
W = 正值 (positive superhelix)

无应力松弛状态



应力的分配





In vitro

EB inserted

Neg. → Pos.

Superhelix change

(来源: 分子生物学 (2007) , 郑用琏, 第47页)

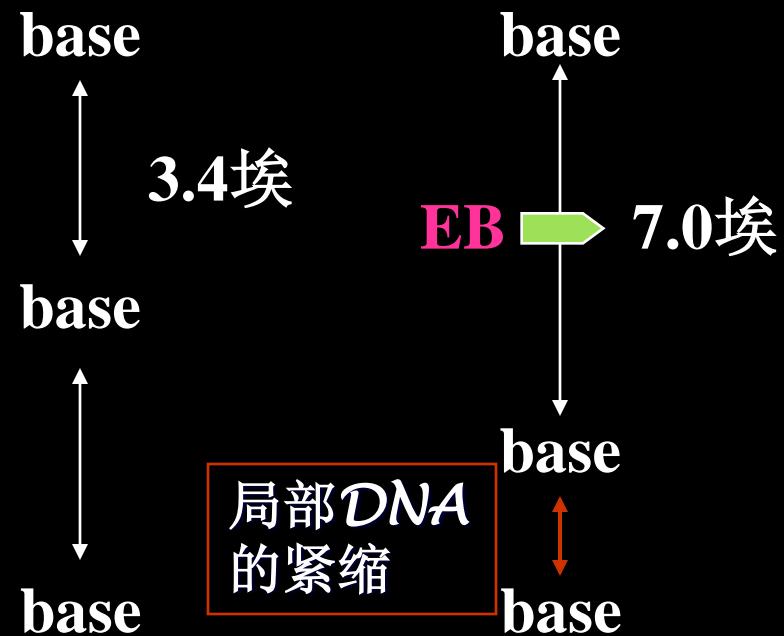
L (linking Num.) 不变

插入点T (Twisting Num.) 减少
(-26°/helix / One of EB inserted)

$$L = T + W$$

$$W = L - T = \text{正值}$$

局部DNA的紧缩



conclusion

- DNA在水溶液中,构型偏B型状态
- DNA以10.5 bp/helix为最稳定构型
- 小于10.5bp/helix 向正超螺旋发展(紧缩态)
- 大于10.5bp/helix 向负超螺旋发展(松弛态)

在细胞内(**in vivo**)

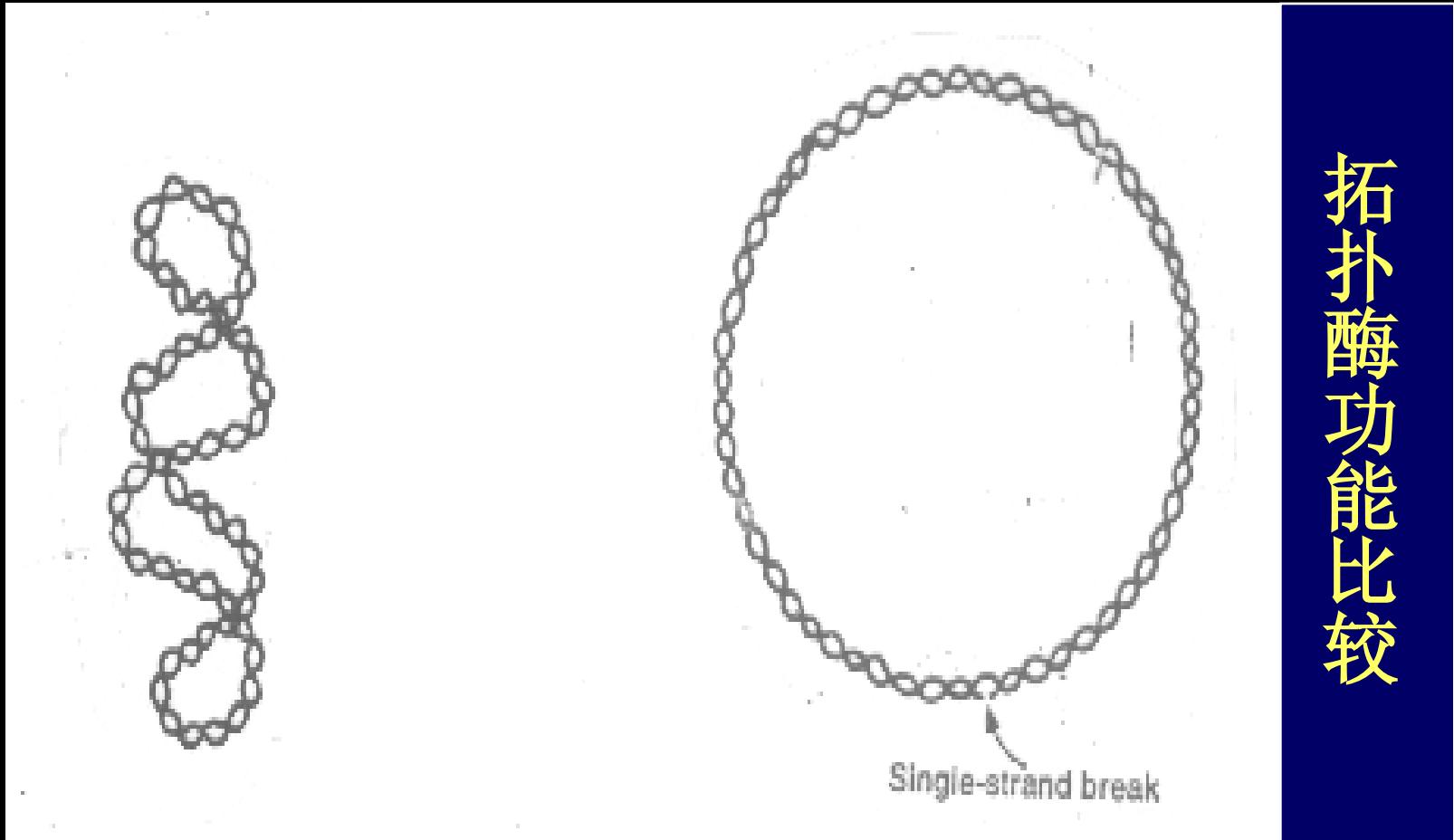
- DNA分子需以高度致密的超螺旋状态, 形成核小体
结构压缩在细胞核内
- 富含 **AT** 及 **IR** 区域易于解链, 形成十字型的负超螺旋,
以消除解链产生的应力, 维持 DNA分子的稳定状态
- B-DNA → Z-DNA → B-DNA 调控基因的表达

DNA复制
RNA转录



需 D. S. → S. S. 状态

- 拓扑异构酶
(topoisomerase I, II) 参与构型的改变



来源：不详

1 拓扑异构酶

(topoisomerase I, II) 参与构型的改变

Top I (swivelase or ω enzyme

or nicking-closing enzyme)

Top II (gyrase)

Cutting & ligation

of D. S. DNA

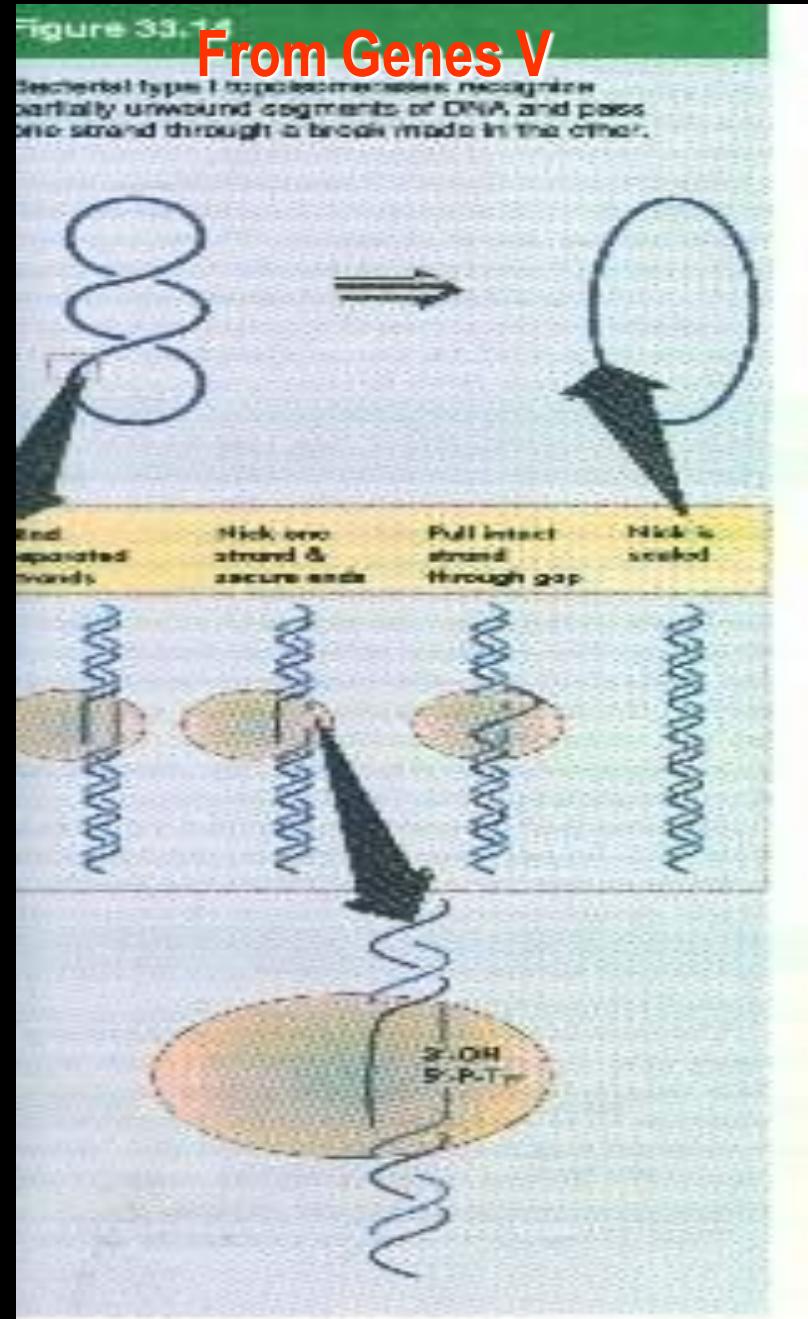
● 拓扑异构酶

(topoisomerase I) 参与构型的改变

The reaction of **Top-1 changes the linking Number in steps of 1**. Each time one strand is passed through the break in the other, there is a ΔL of +1...the interchange ability of W and T should let the change in **Linking number be taken up by change of $\Delta W=+1$, that is, by one less turn of negative supercoiling**

$$L=T+W \quad 36 = 42 + (-6)$$

$$\text{top-1} \quad \xrightarrow{\text{L}} \quad 36+1=42+(-6+1) \\ \text{II} \\ -5$$



来源：不详

Top I (swivelase, nicking-closing enzyme)

Breakage & rejoining of S.S. DNA at phospho-diester bonds

每次 ΔL of +1 (在酶的作用下, DNA单链断裂)

attach
Negative
supercoil

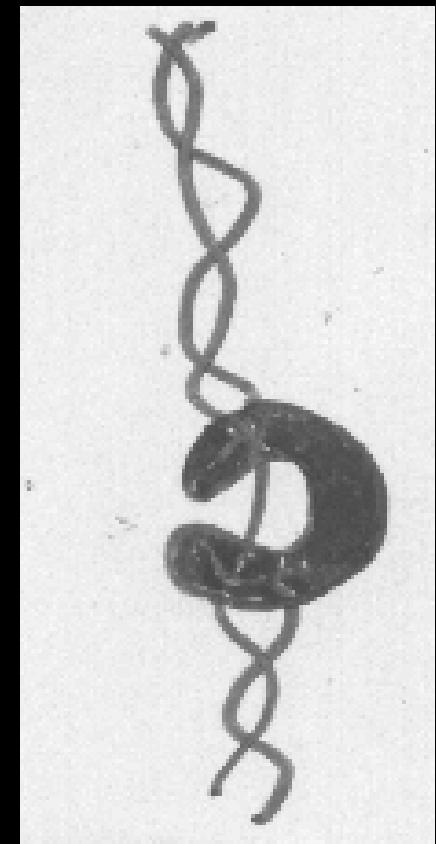
消除负超螺旋

Get energy
from
Negative
supercoil

松弛负超状态
的B-双螺旋

$\text{CCC} \rightarrow \text{OC}$
only

No ATP, NAD



来源: 不详

Top II (gyrase)

ATP needed

tetramer



Cutting & ligation
of D. S. DNA

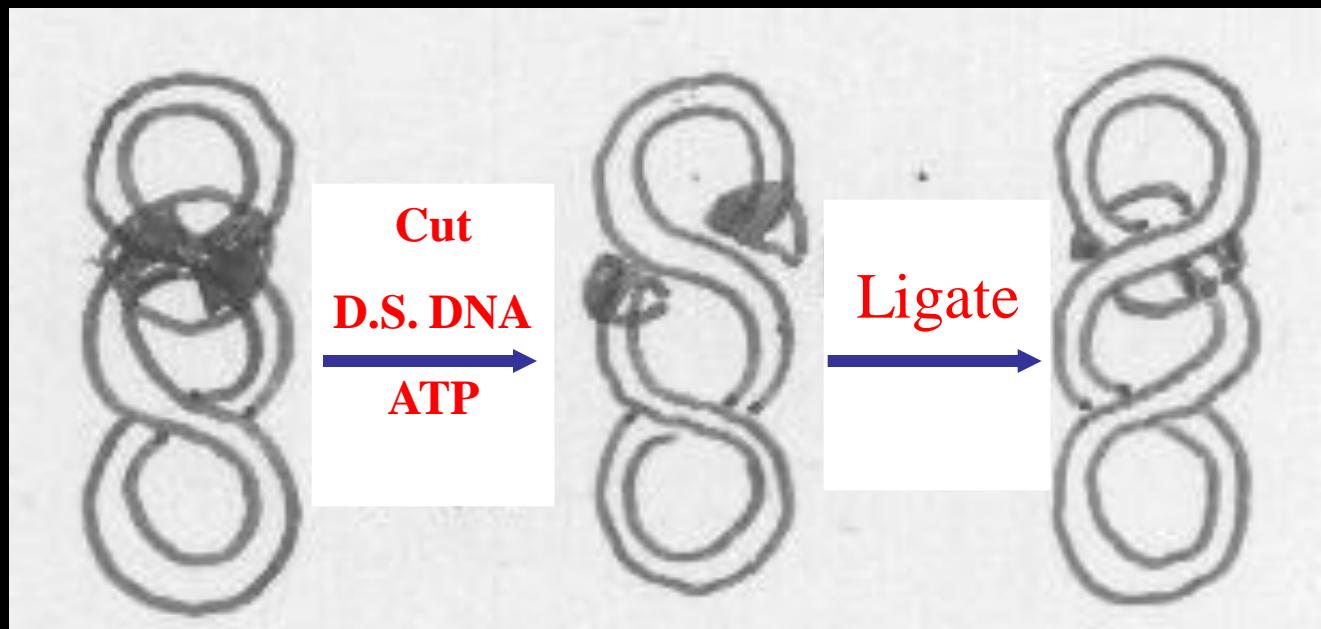
每次 ΔL of -2



引入负超双螺旋

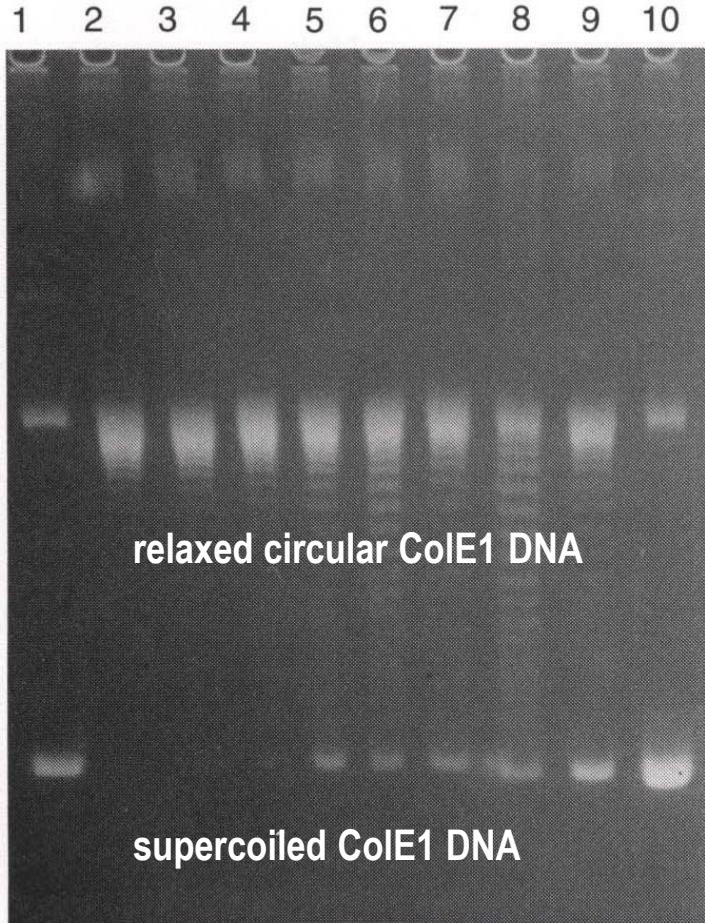


紧缩B-双螺旋



(来源: 分子生物学 (2007), 郑用琏, 第49页)

Assay for a DNA topoisomerase.



relaxed circular ColE1 DNA with varying amounts of E. coli DNA gyrase, plus ATP, spermidine, and MgCl₂,

- 1; supercoiled ColE1 DNA
- 2; no DNA gyrase
- 3-10; DNA gyrase increasing
24 ng, 48 ng, 72 rig, 96 ng,
120 ng, 240 ng, and 360 ng.

DNA gyrase increasing

Gelled et al.

DNA gyrase: An enzyme that introduces superhelical turns into DNA. PNAS 73 (976) p. 3873

● 细胞内

精细调控机制维持

Top I ~ Top II 含量的平衡

严格控制体内负超螺旋

维持在 5% 水平

保证DNA的各种遗传活动

Top II up mut.

More Neg.supercoiling

Top I up mut.

5% Neg. Supercoiling

Top I mut ~ Top II mut.

superhelix density

$\sigma = \gamma / \beta$ = 每一圈初级螺旋（10bp, 360°）出现超螺旋数

不同生物的DNA分子的大小不同

形成几乎相同的超螺旋密度

SV40 5226bp T= 522 实际测定 W = -26 (L=496) $\sigma = -0.05$

E.Coli 4.2×10^6 bp T= 4.2×10^5 分析推测 W = $4.2 \times 10^5 \times -0.05$
= -2×10^4

一般天然DNA分子中 $\sigma = -0.05 = 5\%$ 负向超螺旋

2.2.8. 左旋双螺旋DNA Z-DNA

(Left-Handed Double Helix)

1953. Watson , Crick & Wilkins

- 技术条件本身的局限 (分辨率)
- 研究的精确性 (原子间结构的特点?)
- 不能人工合成寡聚核苷酸片段 (oligo-dNt fragment)
- 仅获得了 DNA fiber graphic of X-ray

1979. Alecxander Rich (MIT)

- 人工合成6 bp $(G/C)_6$ 核昔酸片段
- 获得寡聚核昔酸的crystal graphic of x-ray
- 图象的分辨度达到0.9埃 (0.09 nm)
- 研究了原子间的Bond angle

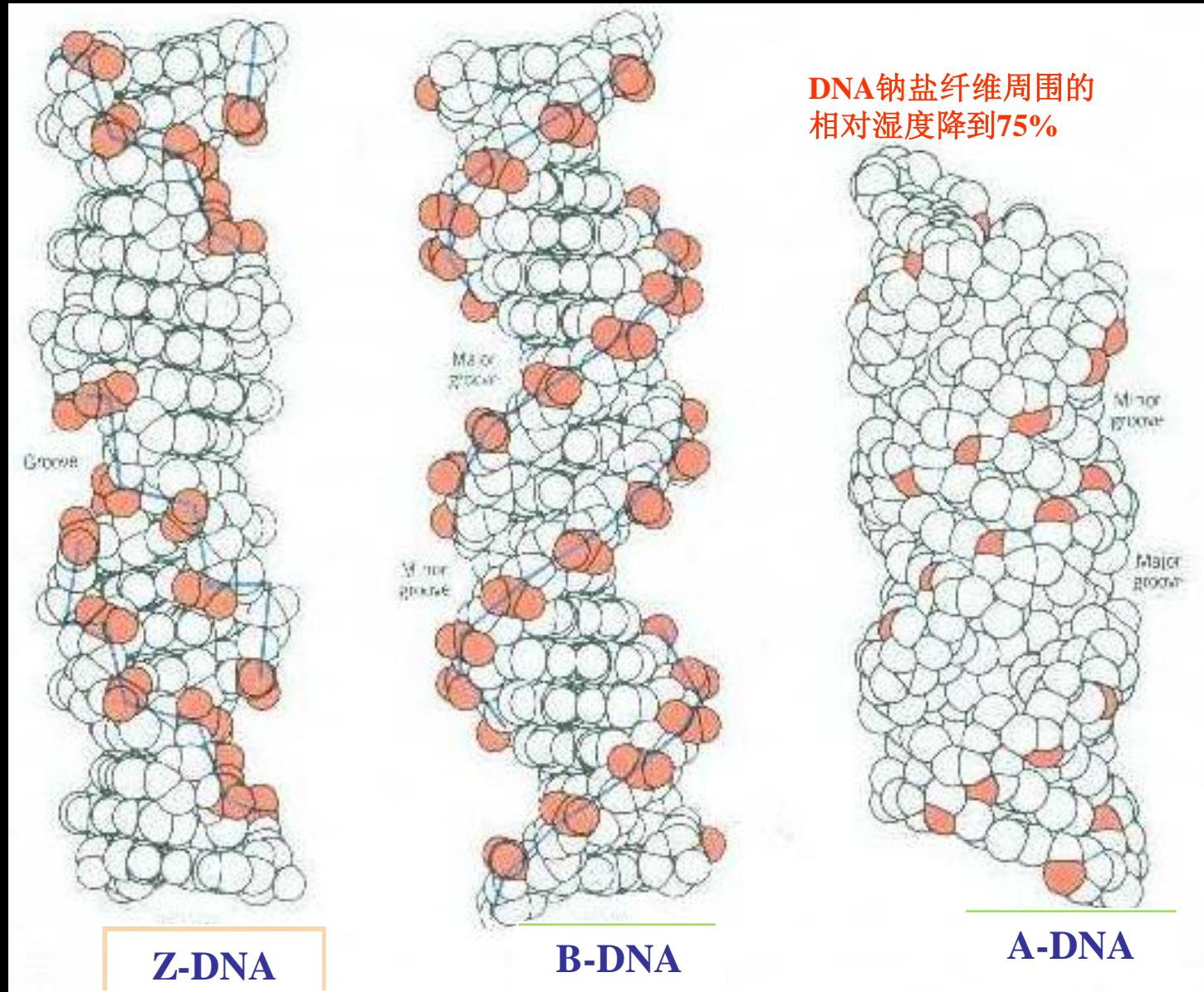
Distance

Ring pucker

$(G/C)_6 \longrightarrow$ *Left-handed double helix*

任何 > 6 bp py-pu 交替排列序列的DNA片段均可
形成Z-DNA, 但难易程度不同

DNA的分子构型 (B, Z, A) 比较

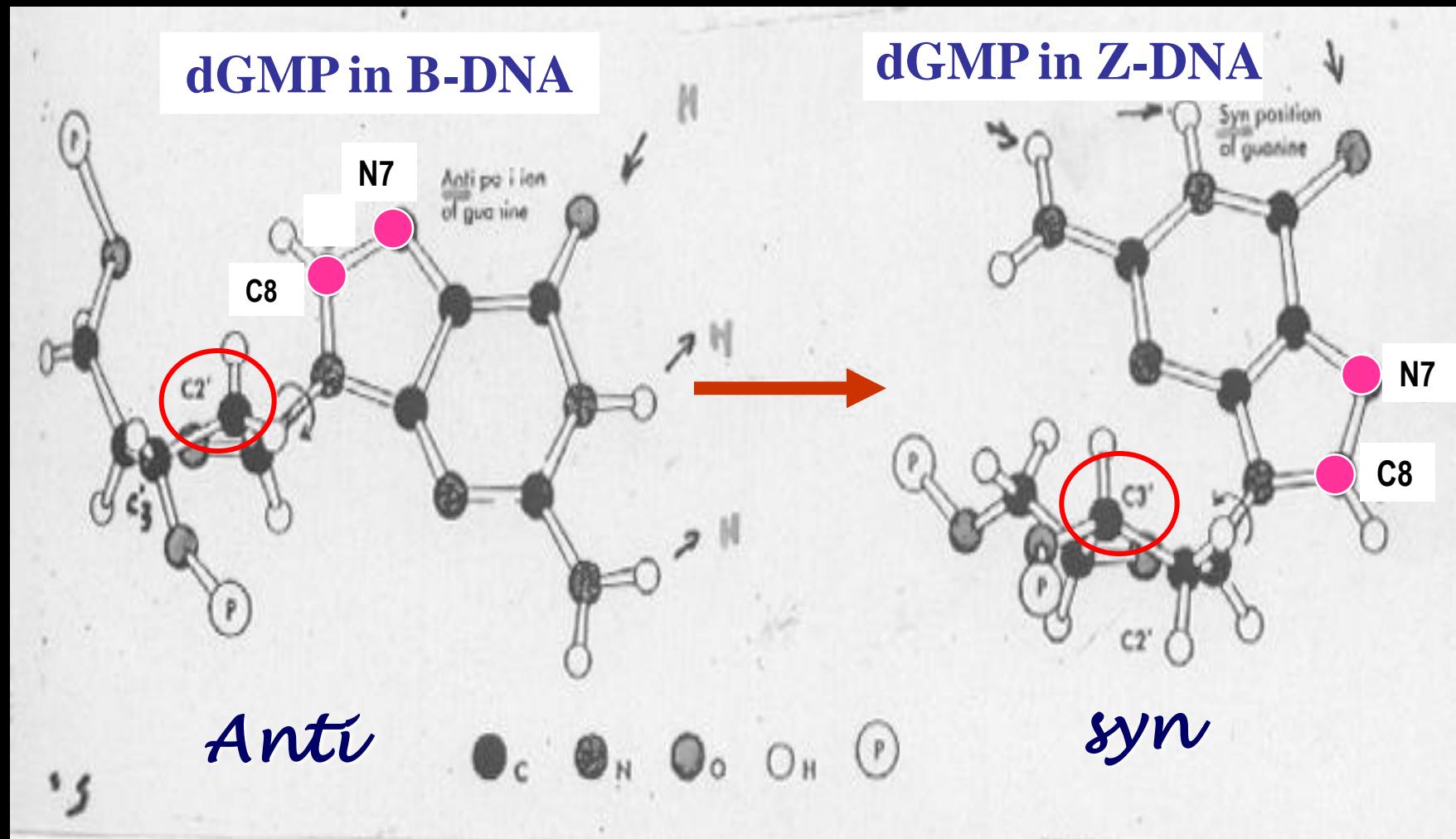


DNA的分子构型 (B, Z, A) 比较

| Form | B | Z | A |
|-------------------------|---------|--|---------|
| Helix Direction | Right | Left | Right |
| Bp / circle | 10 | 12 | 10.7 |
| Distance / bp | ~0.34nm | ~0.38nm | ~0.25nm |
| Distance / helix | 3.4nm | 4.46nm | 2.8nm |
| Diameter / helix | 2.0 nm | 1.8nm | 1.9nm |
| Sequence | Any | Poly G-C Poly C-A Poly T-G Poly T-A | Any |

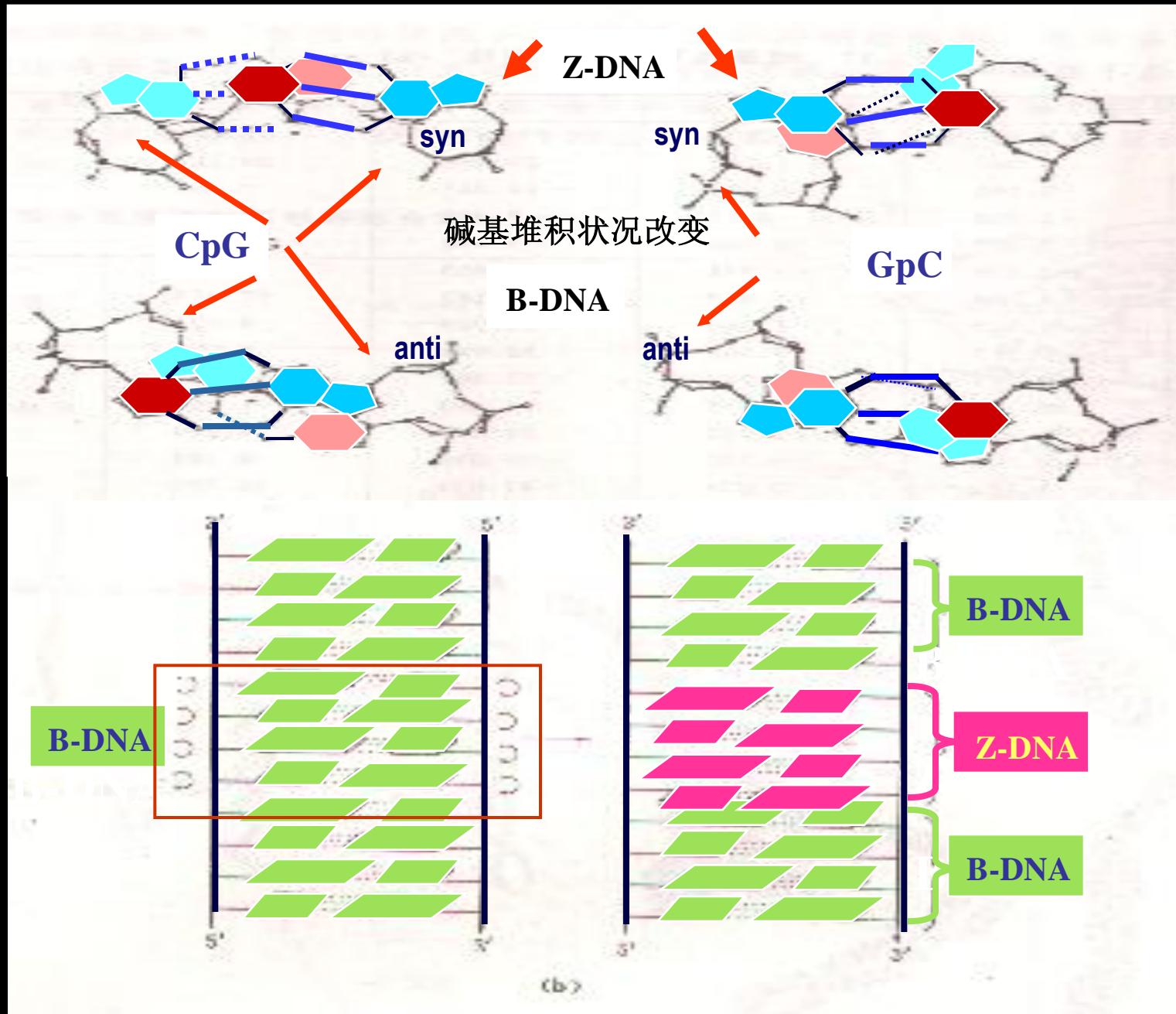
| Form | B | Z | A |
|--|------------------------|---|--|
| Condition | Normal | 4M Na ⁺ | DNA-Na Crytic D.S RNA DNA/RNA |
| Conformation (dGMP, dAMP) | A, G-anti C, T-anti | A, G-syn T, C-anti | A, G-anti T, C-anti |
| C₂, C₃ / Pu | C ₂ -endo | C ₃ -endo | |
| Groove | Minor Major | N ₇ ,C ₈ outer Minor deeper | Minor Major deeper |
| Antigen | No | Yes | Yes (check RNA virus) |

DNA的分子构型 (B, Z) 比较



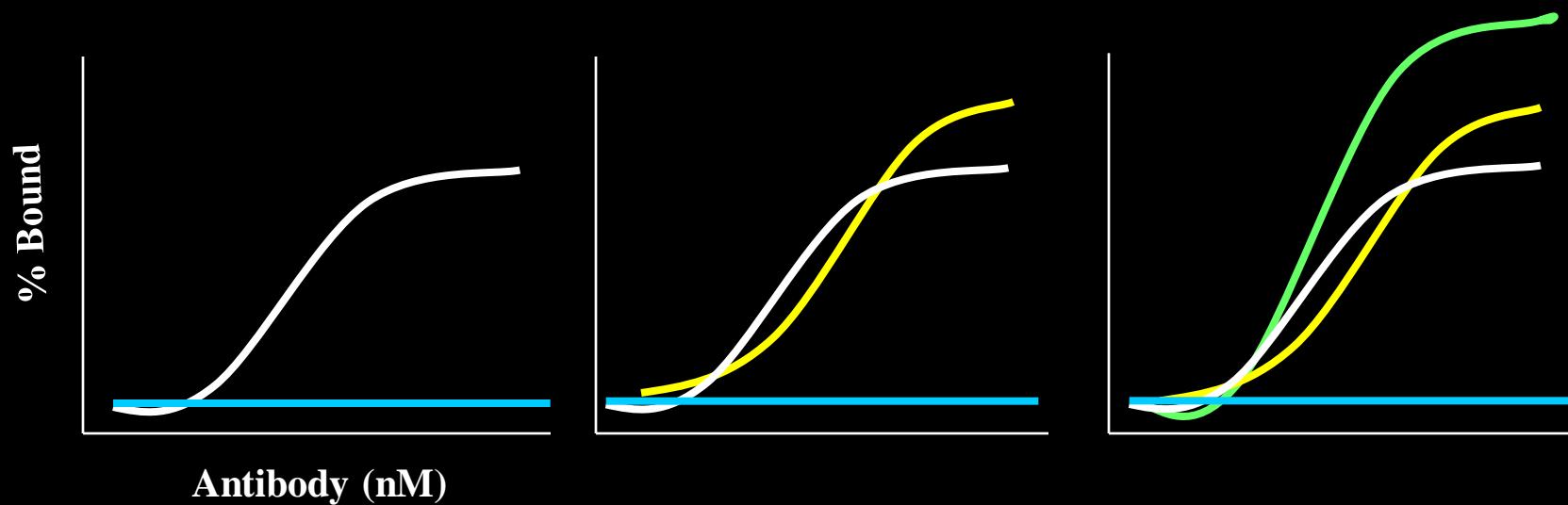
来源：不详

DNA 分子构型的比较



(来源: 分子生物学 (2007), 郑用琏, 第41页)

Z-DNA 的检测



0.2 M Nacl

Br.poly(dG-dC)

1.5M Nacl

Br.poly(dG-dC)
poly(dG-dm⁵C)

4.0M Nacl

Br.poly(dG-dC)
poly(dG-dm⁵C)
poly(dG-dC)

Z-DNA 的检测

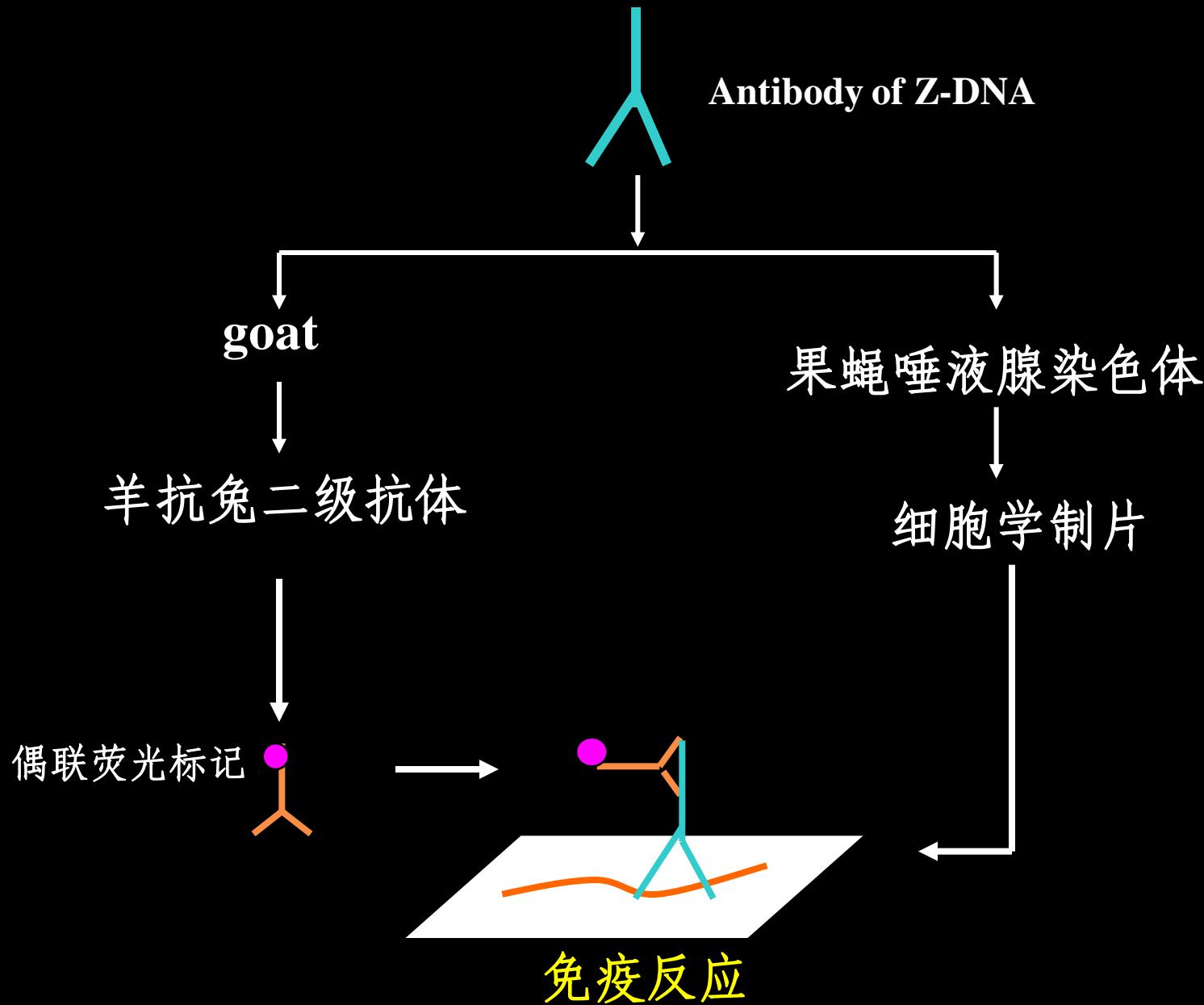
B-DNA $\xrightarrow{4M\ NaCl}$ Z-DNA $\xrightarrow{Br^+}$ Z-DNA(Br^8G)

0.2 M Nacl Z-DNA (Br^8G) 趋于稳定

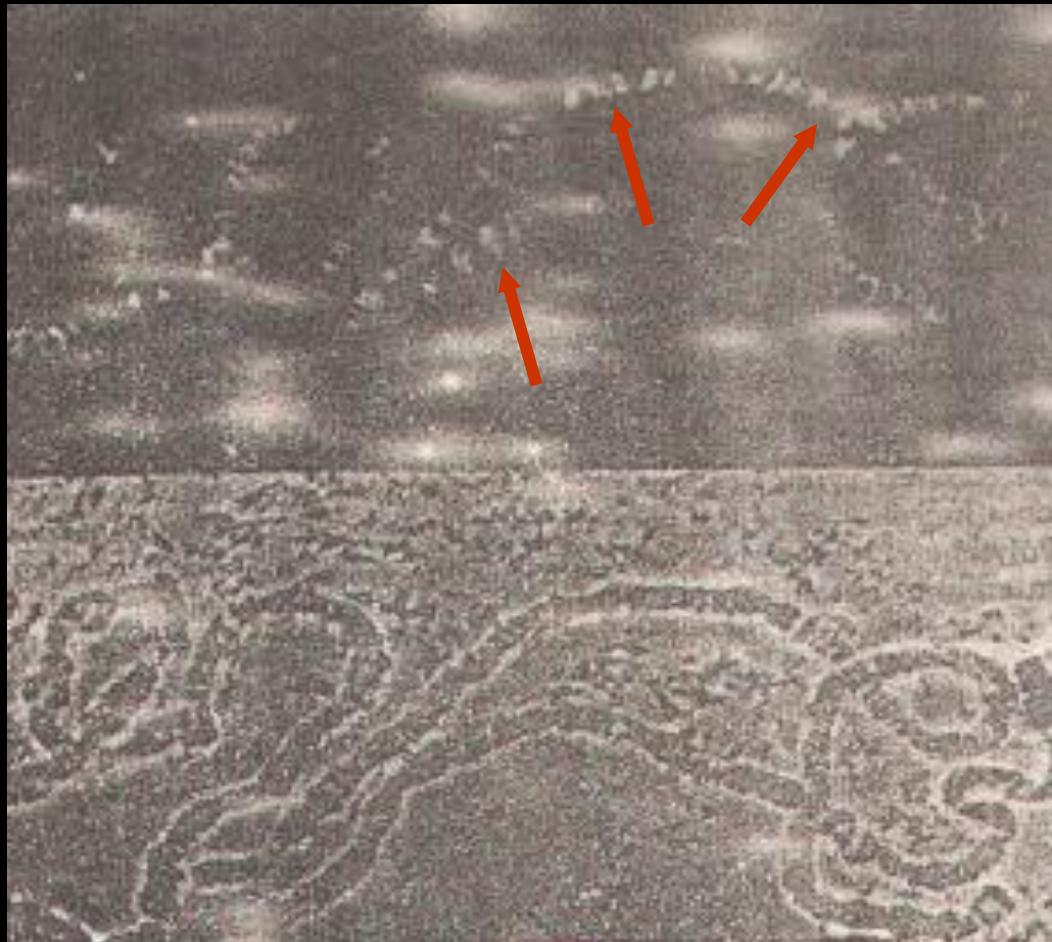
rabbit

Antibody of Z-DNA





Z-DNA in *Drosophila* chromosome proved by anti-Z antibody



Immunological
slide

Cytological
slide

(来源：分子生物学（2007），郑用琏，第40页)

影响Z-DNA稳定的因素(in vivo)

1, m⁵C



(G--C)_n排列不是Z-DNA形成的必要条件

2, Z-DNA 中 G 的 C₂-NH₂ 在大沟内与 H₂O--H₂O 形成氢键连接

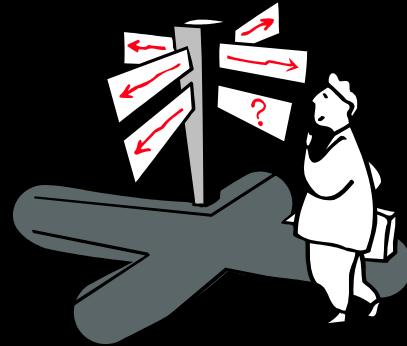
3, Z-DNA 中 G 的 N₇, C₈ 外露, 易与特异蛋白(或 Br)结合

Histone H1 → 稳定B-DNA

H2A, H2B, H3, H4 → 稳定Z-DNA

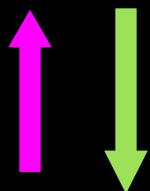
4, B-DNA → 负向超螺旋 → 应力 → Z-DNA

可能的功能



基因表达调控

Z-DNA → (小沟, 信息少) → 基因关闭



B-DNA → (大沟, 信息多) → 基因表达



可能的功能

- 转录过程中模板DNA双链的Unwinding → 引发负向超螺旋的力量 → 导致RNA聚合酶之后的Z - DNA形成 → 调控基因两次转录过程的间隙

Z - DNA在活体中是一种依赖转录的结构，基因的转录导致了Z - DNA的形成，而不是形成Z - DNA后启开基因！？

2.2.9. 三股螺旋DNA (Triple Helix DNA, T.S DNA)

※ T.S DNA 的发现与证实

DNA结构领域与Watson竞争，支流？主流？

- 1953 年以前 Pauling (Chemist) 提出 T.S DNA 存在的可能性
- 1953 年 Watson & Crick D.S DNA model 证明沿大沟存在多余的氢键给体与受体



潜在的专一与DNA (蛋白质)结合的能力



形成T.S DNA 可能性

- 1957年Davis , Felsenfeld , Rich 发现

$\text{poly(U)} + \text{poly(U)} + \text{poly(A)}$ \longrightarrow T.S RNA



T.S DNA的概念

•••

但由于D.S
DNA的提出
而被忽视

- 1966年Miller & Sobell

实现 RNA + D.S DNA



Triple polyNt



as Repressor 关闭基因

但因证明 LacI
产物为
Repressor 而被
忽视

- 1975年 Perlmut

人工合成T.S DNA并证明其 Tm 值, 沉降系数(S)

- 1987 年 Mirkin . S . M *Nature* 330 (495)

证明 plasmid DNA 在 pH= 4.3 的溶液中,
有 T.S DNA 的存在

- 1987年 Dervan . Moser *Science* 238 (645)

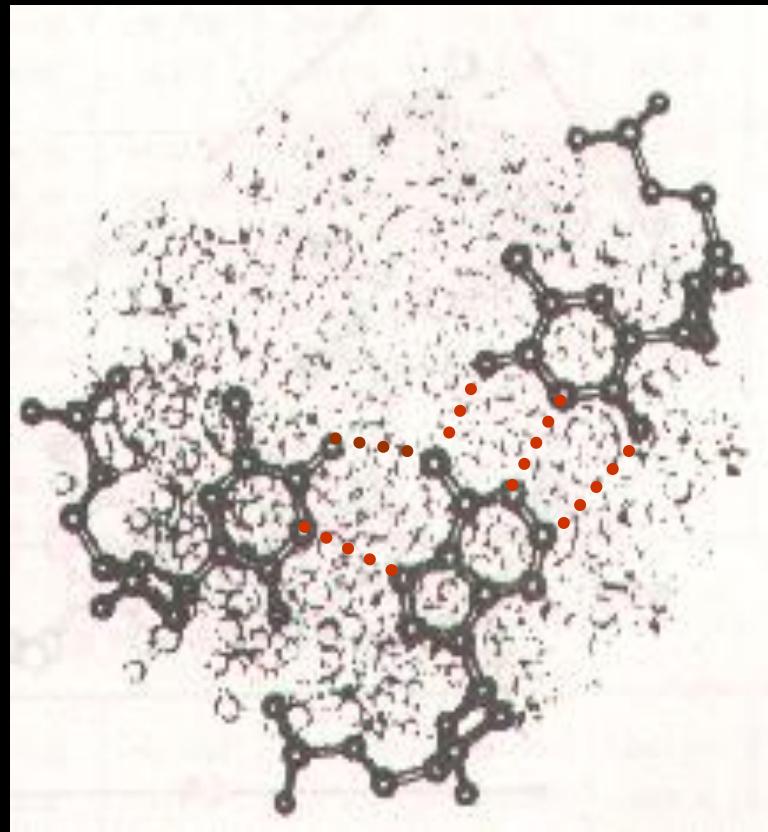
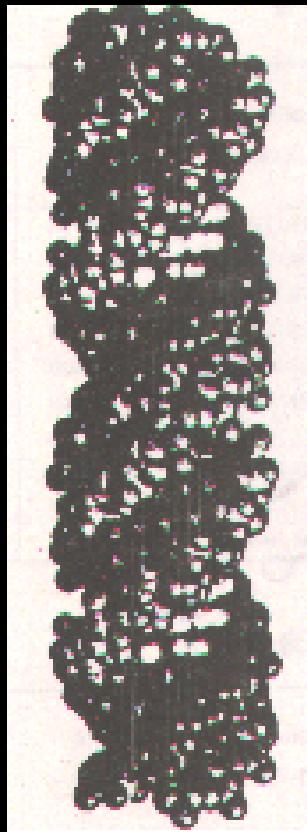
合成 S.S DNA + D.S DNA → T.S DNA

- 实现DNA的定点切割
- 研究X-ray photograph
- 核磁共振 → 结构功能



继Davis (1957) 后 30年

第一次证明T.S DNA 在生物体内的存在



(来源: 分子生物学 (2007), 郑用琏, 第42页)

※ T.S. DNA的类型

- D.S. DNA + D.S.DNA



T.S. DNA + S.S. DNA

PolyT/A

TTTTTTTTTT
||| ||| ||| |||
AAAAAAA
AAAAAAA

PolyT/A

TTTTTTTTTT
||| ||| ||| |||
AAAAAAA
AAAAAAA



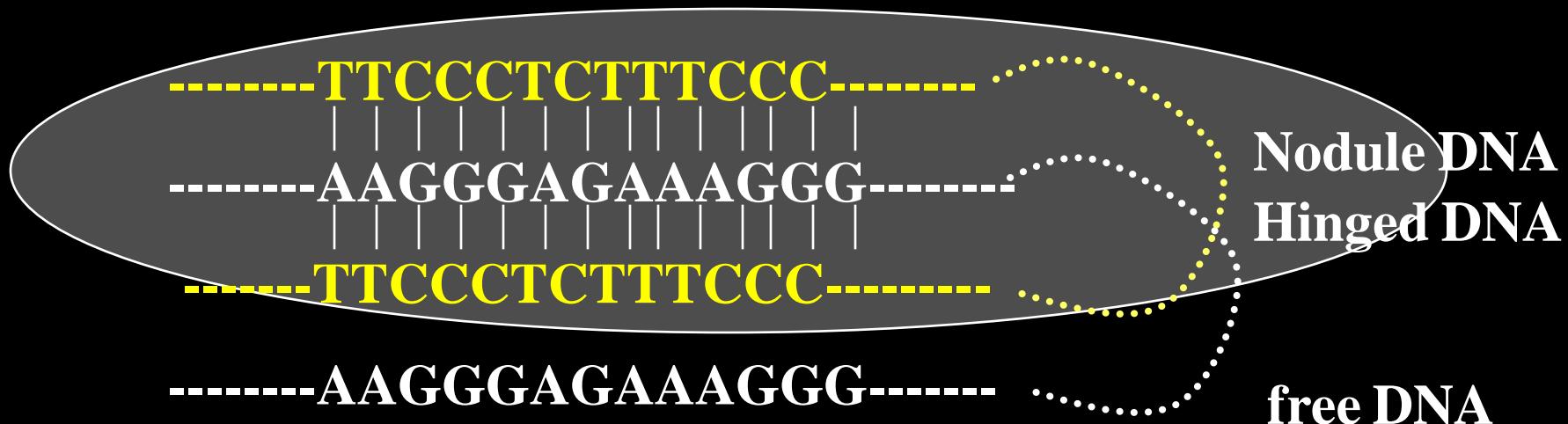
TTTTTTTTTT
||| ||| ||| |||
AAAAAAA
TTTTTTTTTT

AAAAAAA

- Homologous palindromic sequence in a D. S. DNA
(mirror image structure)

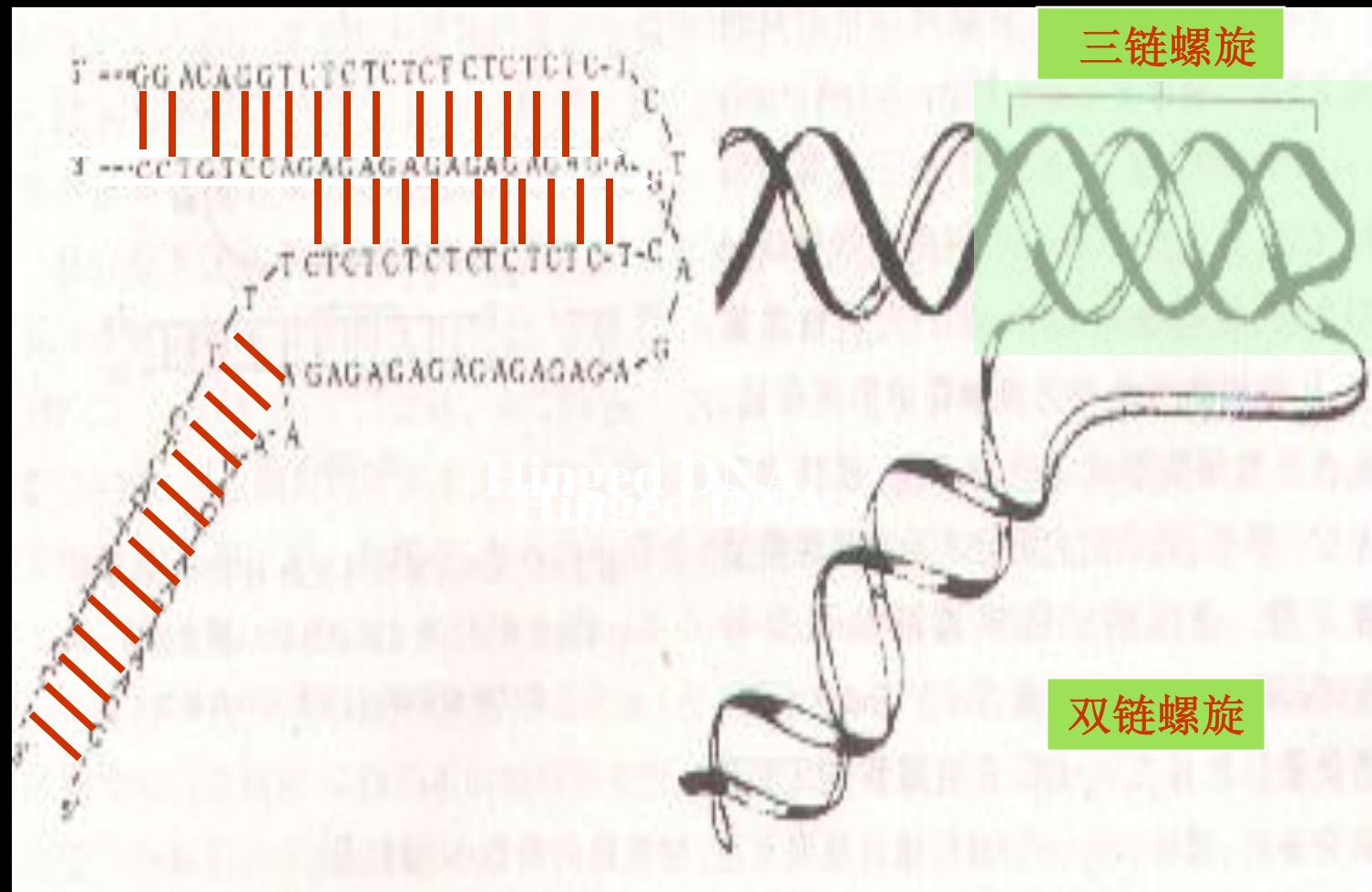
Mirkin (1987) pGG332 plasmid DNA

-----TTCCCTCTTCCC-----CCCTTTCTCCCT-----
-----AAGGGAGAAAGGG---GGGAAAGAGGGAA----



Homologous palindromic sequence in a D. S. DNA

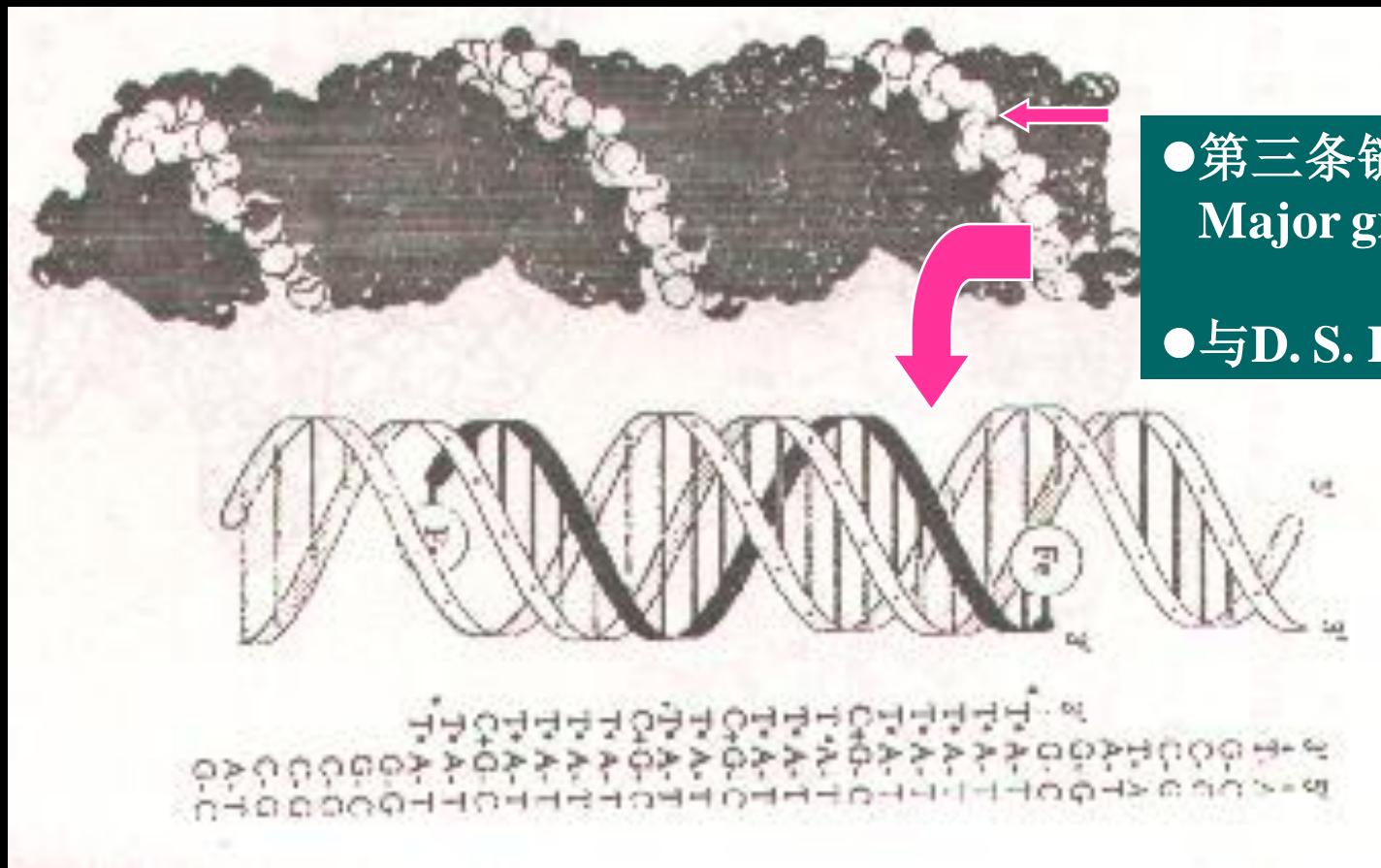
Nodule DNA or Hinged DNA



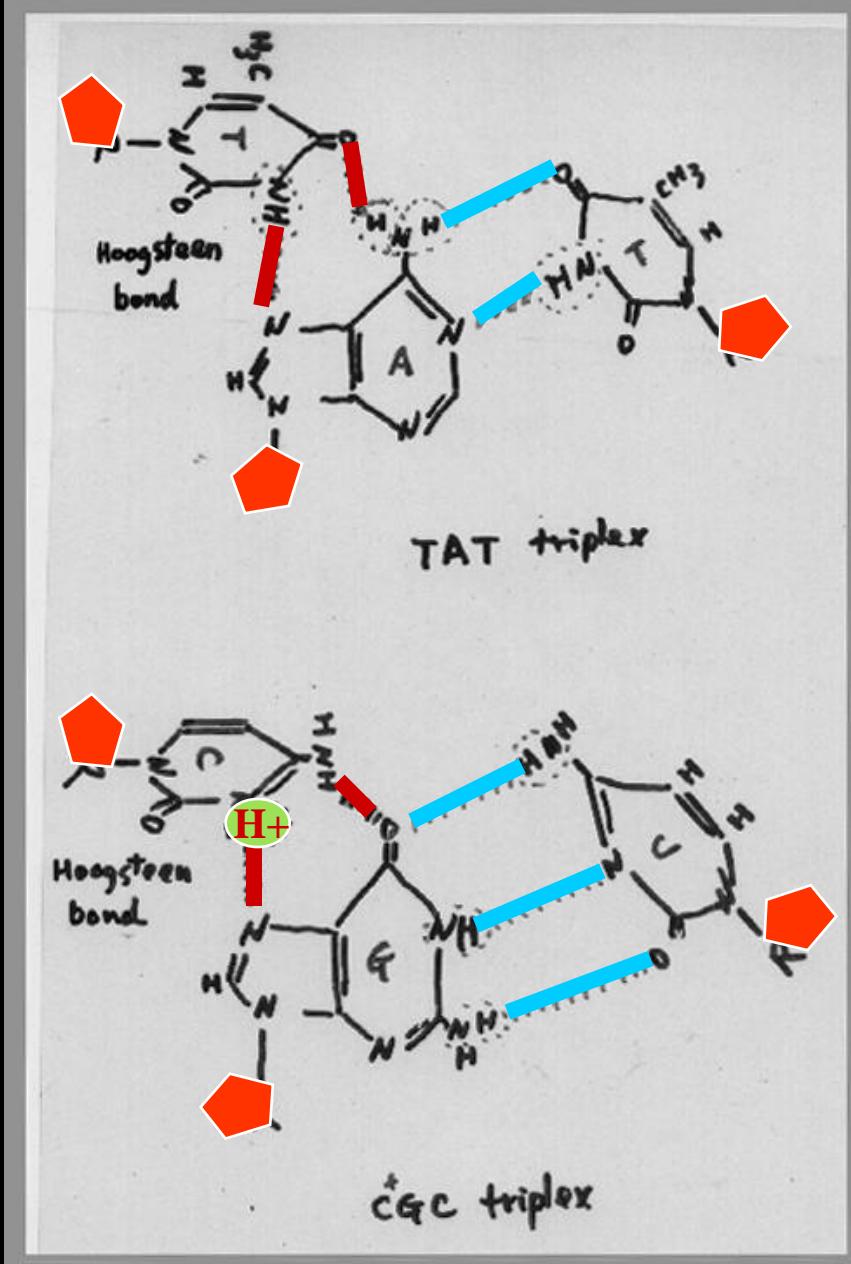
(来源: 分子生物学 (2007), 郑用琏, 第43页)

● S. S. DNA + D. S. DNA → T. S. DNA

- ☆ PU + PU/PY (偏碱性介质中稳定)
- ☆ PY + PU/PY (偏酸性介质中稳定) 常见类型



(来源: 不详)



T. S. DNA 的连接键

Watson bonding

$$A = T \quad G \equiv C$$

(D. S. DNA)

Hoogsteen bonding

第二链的 pu 6',7' 与第三链的 py4',3' 形成 H 键

$$G = C^+ \text{ (pH 小于 7)}$$

第三链 质子化

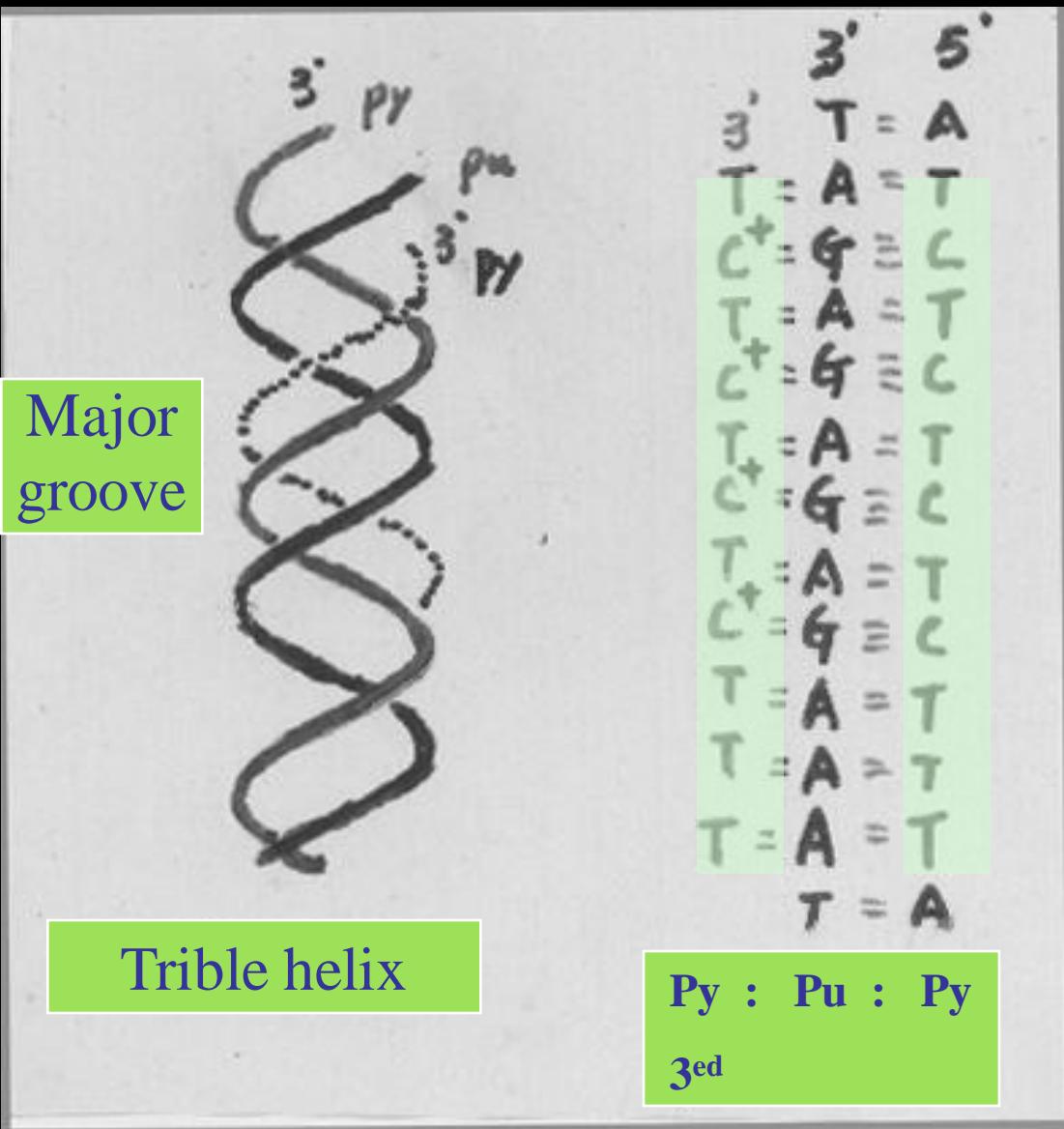
$$G_{(6,7)} = G_{(1,2)} \quad A = A$$

(同向平行, 反向平行)

$$G = A^+ ? A = T$$

反向平行 (同向平行, 反向平行)

三股螺旋DNA形成的条件及结构特点



第三条单链DNA分子

- 位于B-DNA大沟内
- 与B-DNA以 Hoogsteen 键连接
- 在 Py / Pu : Py 结构中
A = T , G = C 两氢键配对
- C质子化 } 必需条件
镜相结构 }
- 在 Py / Pu (A,G): Pu (A,G)
结构中有多种配对方式,
镜向结构, 非必需条件

- 无论任何形式的三螺旋 DNA, 第二股中间链必须是 Purine 链
- 第三股链至少长于 8 dNt
- 真核生物基因组内, 约 1% 左右的序列为大于 100 bp 的 homologous cluster (重复序列与调控序列) T. S. DNA 多存在于其中

T. S. DNA可能的功能

- a) 可阻止调节蛋白与 DNA结合,关闭基因转录过程
micro-RNA Epigenetic!
- b) 与基因重组,交换有关
- c) 加入第三条S. S. DNA 作为分子剪刀(molecular scissors),
定点切割DNA分子
- d) 加入反义的第三条链(anti-sence polydNt) 终止基因的表达

2.2.10. 四股螺旋DNA

(tetraplex DNA, tetrable Helix DNA, quadruplex DNA)

发现

1958. Poly(I) X-ray photograph

碱基形成环状氢键连接结构



Tetrable Helix DNA

- Poly (G), 4 (dG)
- 染色体端粒高度重复的 DNA 序列

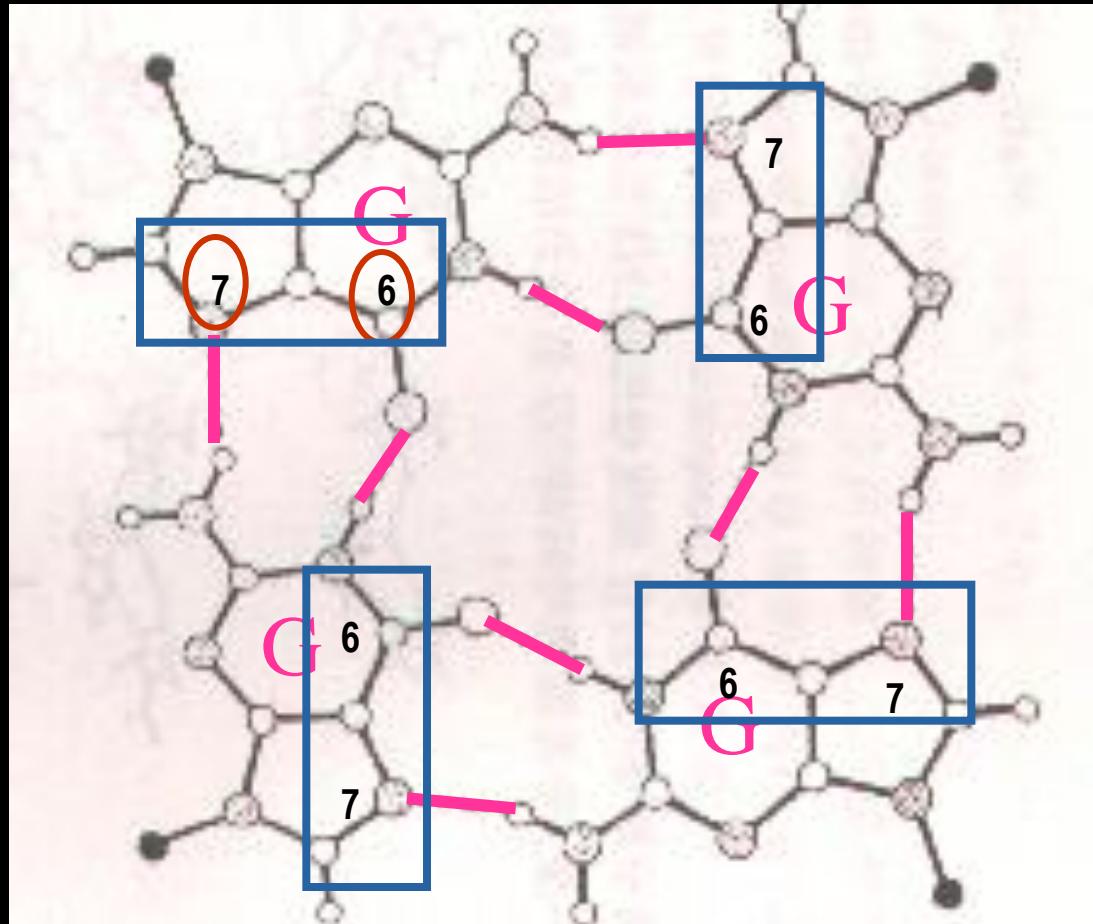
5'---TTAGGGTTAGGGTTAGGG-3'

3'---AATCCCAATCCC-5'

- 着丝点附近的高度重复序列

均有形成
四股螺旋DNA
的可能

结构特点



Linked by
Hoogsteen
Bonding

6—1

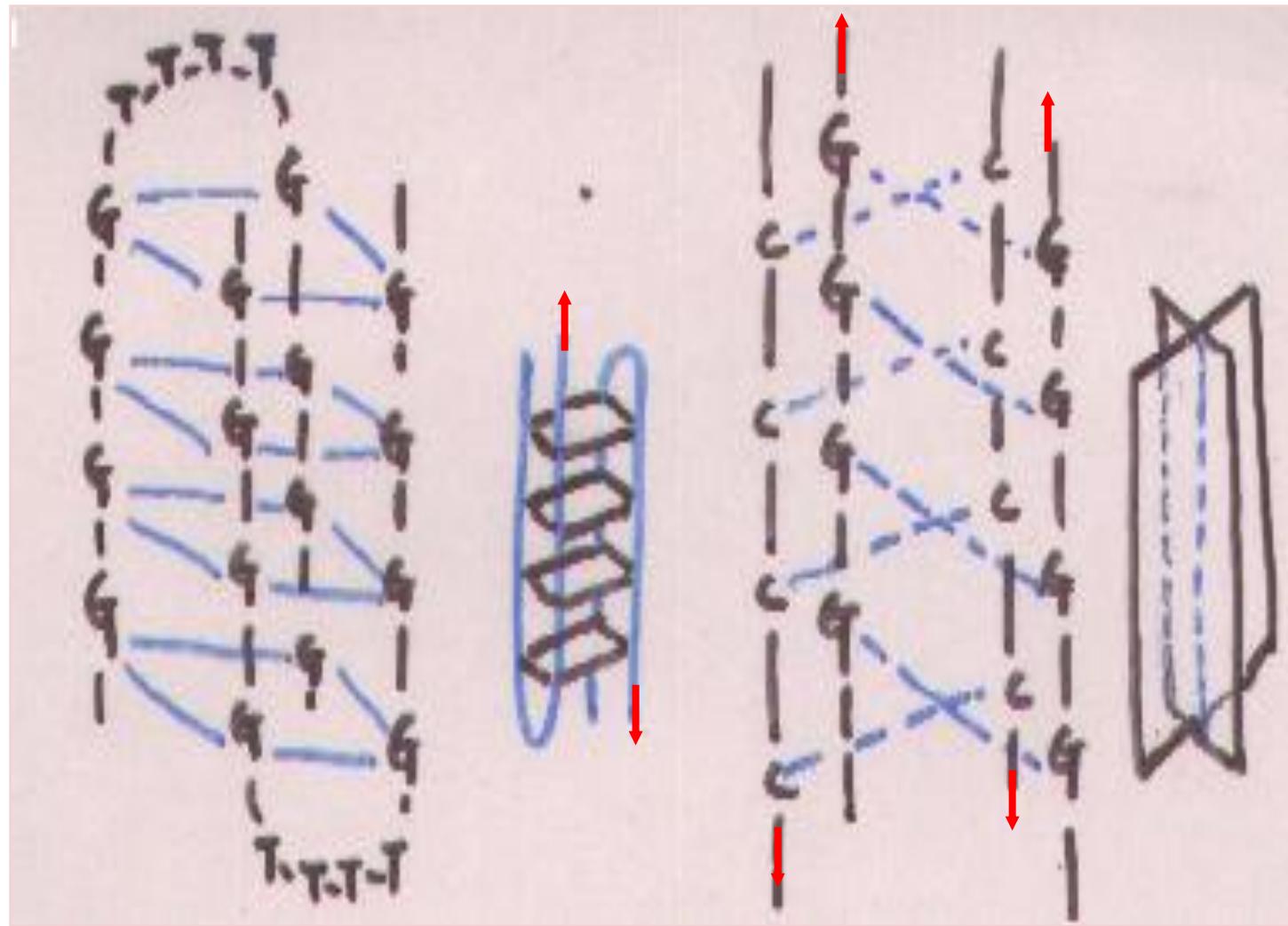
7—2

(来源: 分子生物学 (2007), 郑用琏, 第45页)

结 构 特 点

$2 \times \text{poly}(\text{T}_4\text{G}_4)$

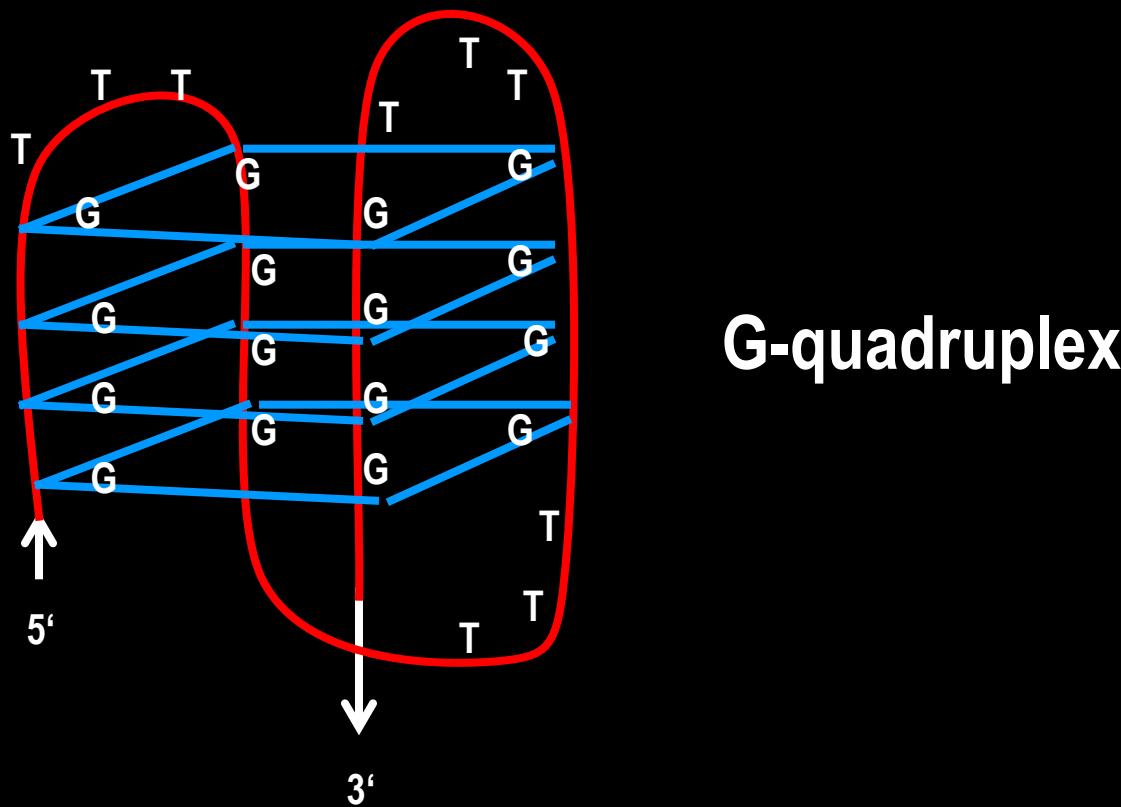
$2 \times \text{poly}(\text{G}_4\text{C}_4)$



(来源：分子生物学（2007），郑用琏，第45页)

真核生物染色体端粒DNA结构

GGGGTTTGGGGTTTGGGGTTT



G-quadruplex

(来源：分子生物学（2007），郑用琏，第45页)

可能的功能

A 稳定真核生物染色体结构



Hoogsteen Bonding

B 保证DNA末端准确复制

C 与DNA分子的组装有关

D 与染色体的meiosis & mitosis 有关

可能的功能

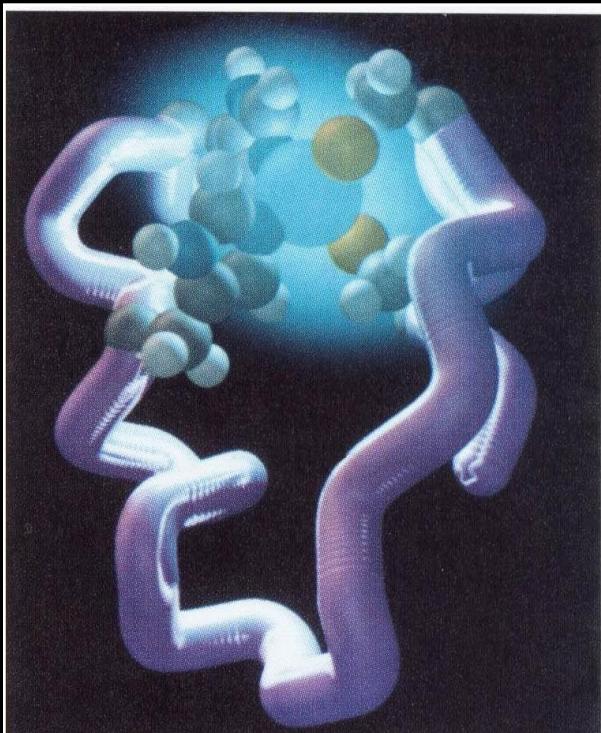
E. G-quadruplex阻止端粒酶对端粒DNA的延伸
(因为G-quadruplex不是端粒酶的底物)



G-quadruplex 的边序直接
影响其结构的稳定性！

2. 3. 基因概念

的多样性重叠基因



生物进化的C值矛盾 (C value paradox of nucleotide)

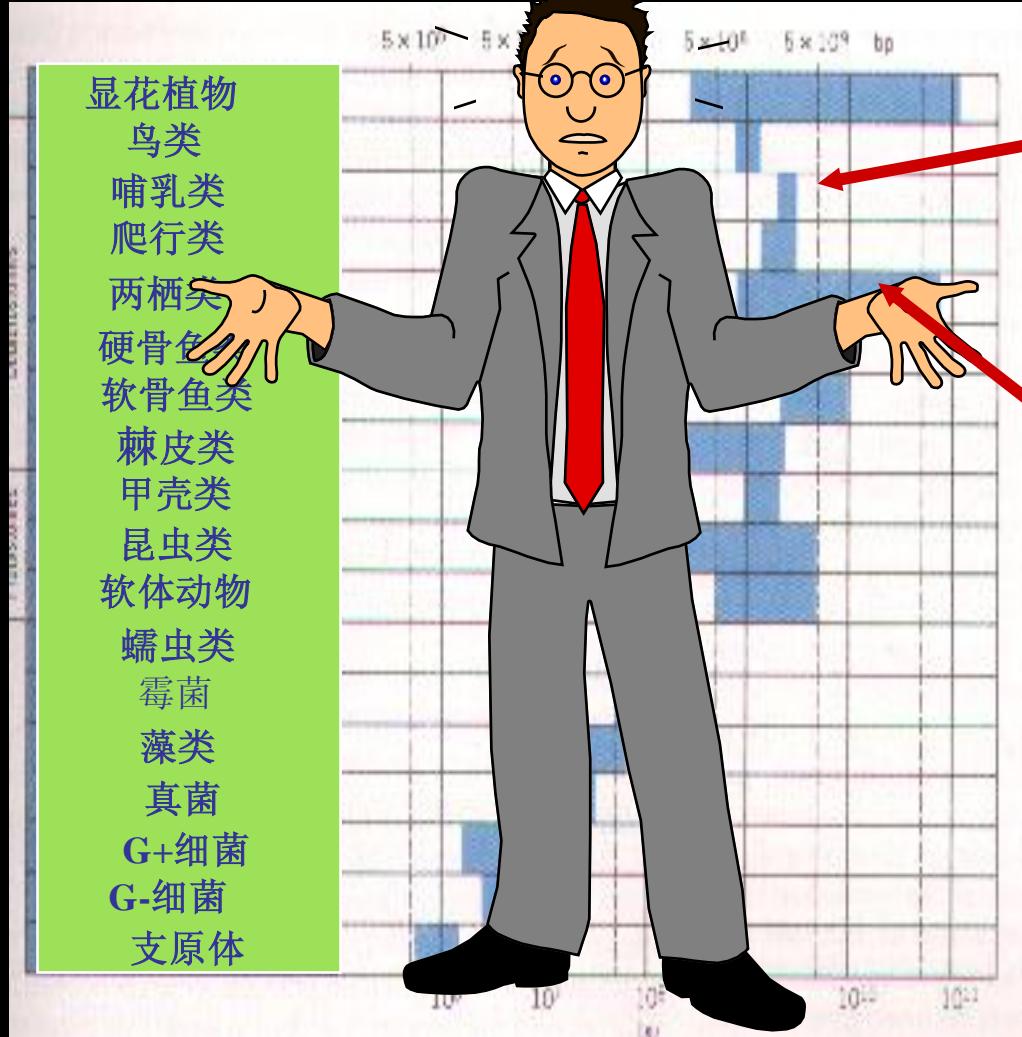
最大C值 (Maximum C value)

The total amount of DNA in the genome of haploid
Is a characteristic of each living species known as its
Maximum C value (单倍体基因组总DNA的含量)

最小C值 (Minimum c value)

The total amount of DNA for encoding the genes
information is termed its Minimum c value (编码基因
信息的总DNA含量)

C value paradox of nucleotide



- A 生物体进化程度高低与大C值不成明显相关（非线性）
- B 亲缘关系相近的生物大C值相差较大
- C 一种生物内大C值与小c值相差极大
(Euk. 人体 $c = C/10$)
(Prok. $\Phi \times 174 c > C$)

重叠基因

(overlapping gene)

F. Sanger 1977

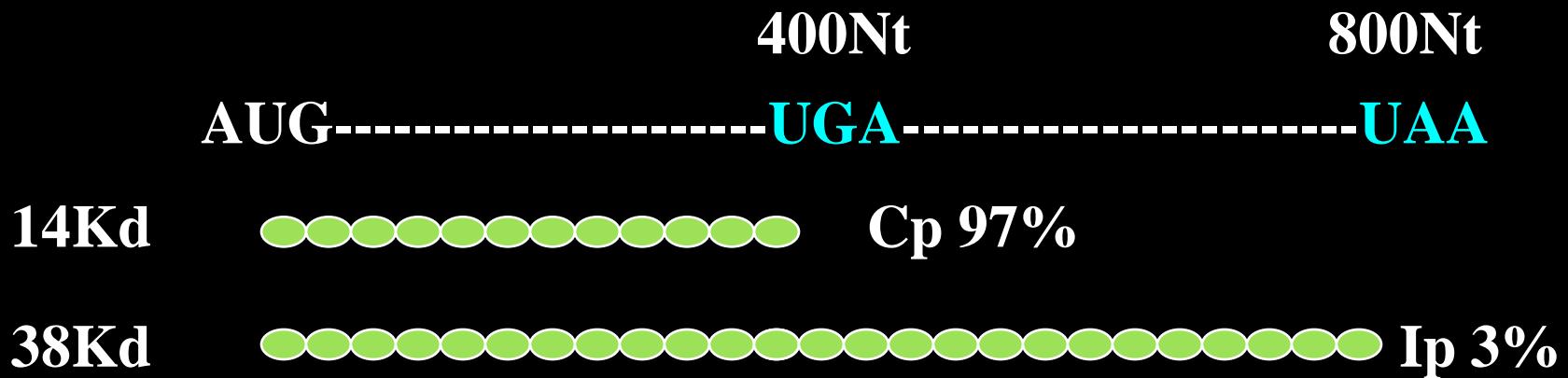
ϕ X174 dnt sequencing

$$C = 5387 \text{ bp} \quad c = 11 \times 2000 \text{ bp}$$

2.3.1 基因重叠方式

- Mis-reading for stop codon

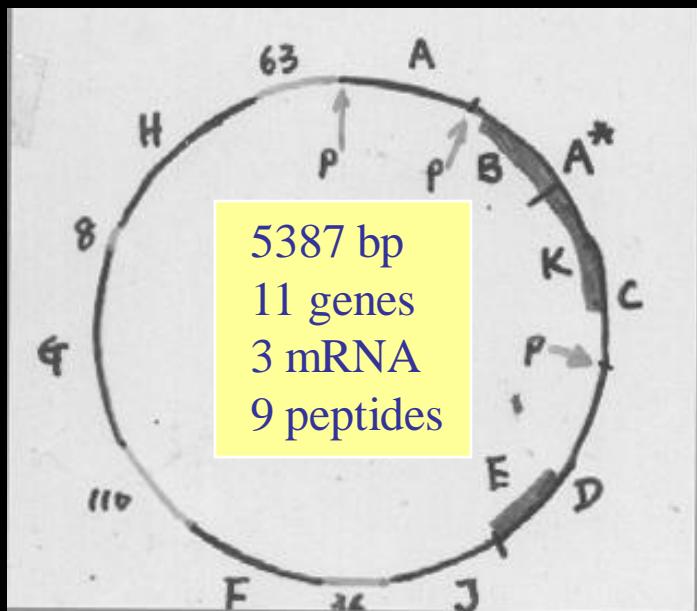
(*QβRNA virus* 1973. A. Weiner)



UGA, UAG 易被漏读，错读

UAA 能严格终止

- Choose different reading frame



ϕ X174 (F. Sanger, 1977)

$$C = 5387 \text{ bp}$$

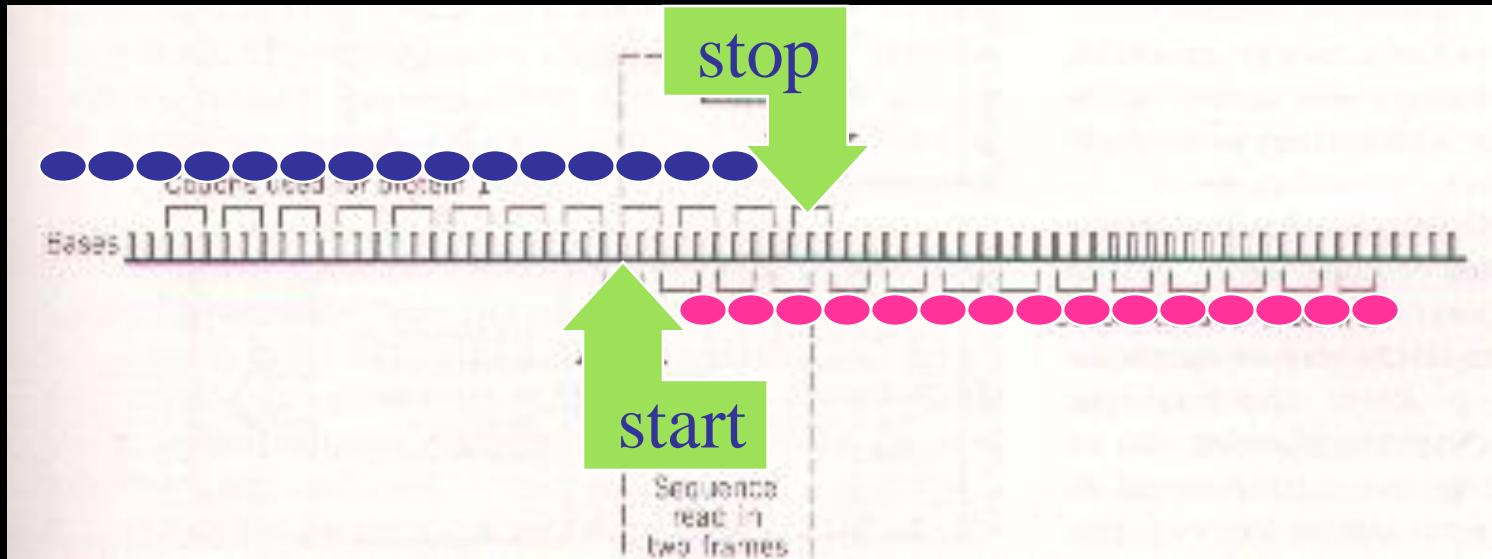
$$c = 11 \times 2000 \text{ bp}$$

(来源: 分子生物学 (2007), 郑用琏,

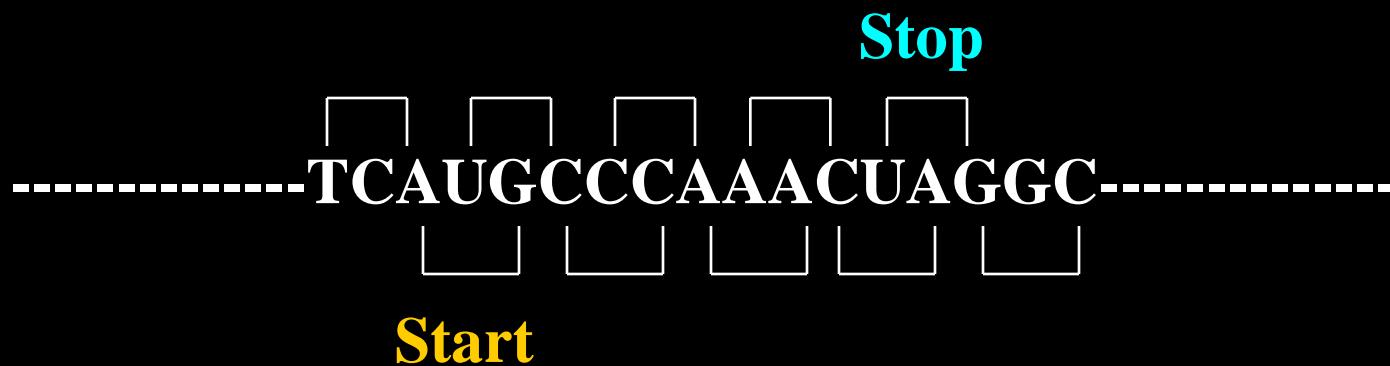
第50页)



Choose different reading frame

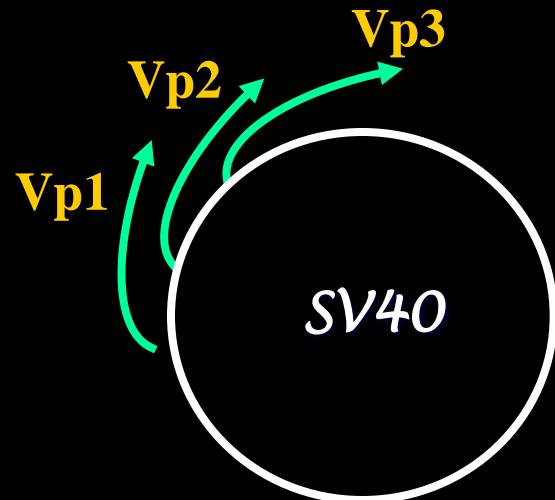


(来源: 分子生物学 (2007), 郑用琏, 第51页)



- Selection different start codon or stop codon

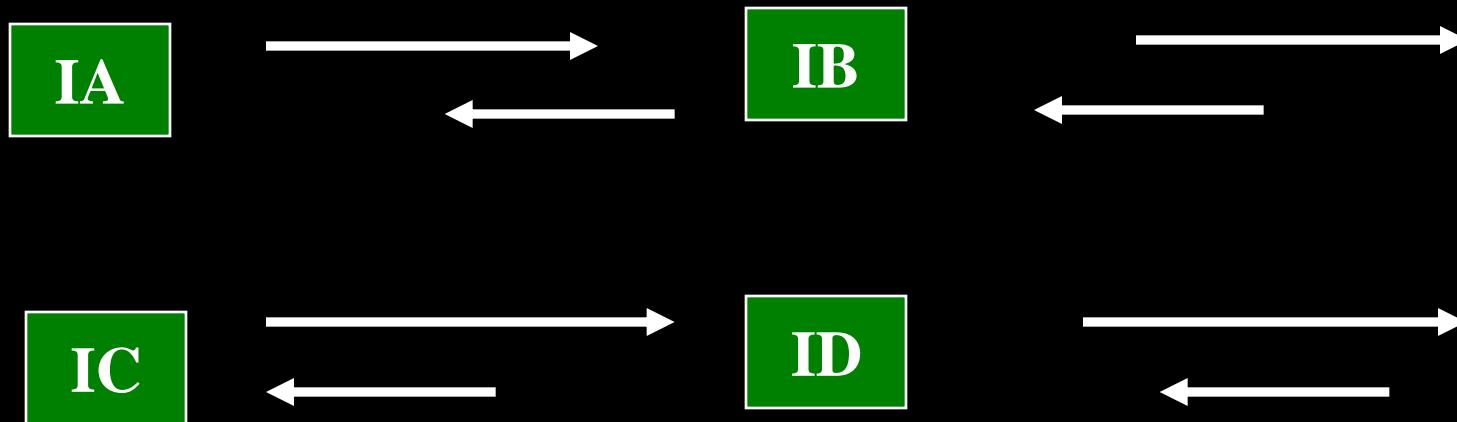
(Simian Virus 40 SV40)



2.3.2. 种类

※ I类： 反向重叠基因

(重叠基因分布在同一DNA区域的不同单链上)？！



方式？！ 功能？！

※ II类; 同向重叠基因

(重叠基因分布在同一DNA区域的同一单链上)

IIA



IIB



方式?

IIC



IID



2.3.3. 重叠基因的生物学意义

a) 原核生物进化的经济原则

(较小的C值编码较多的基因信息)

b) 遗传信息量的估算

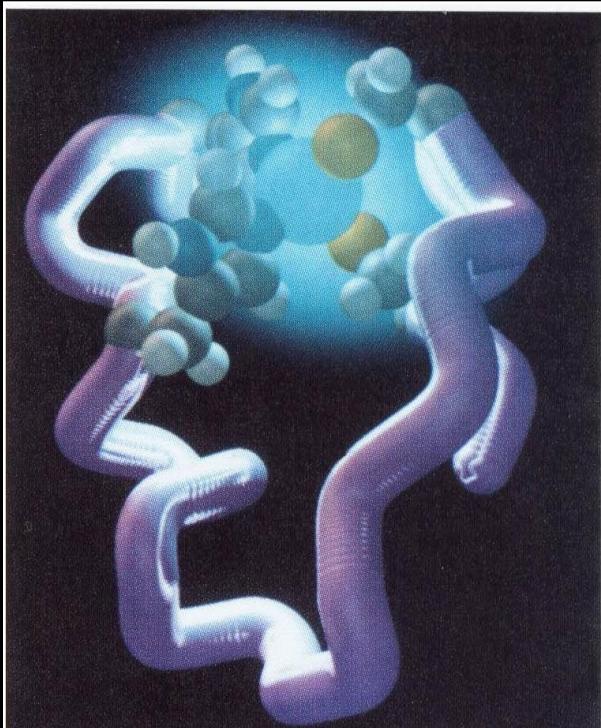
突变效应的鉴定 表达调控的理论发展

c) 丰富和发展了基因的概念

(部分回答 C = c)

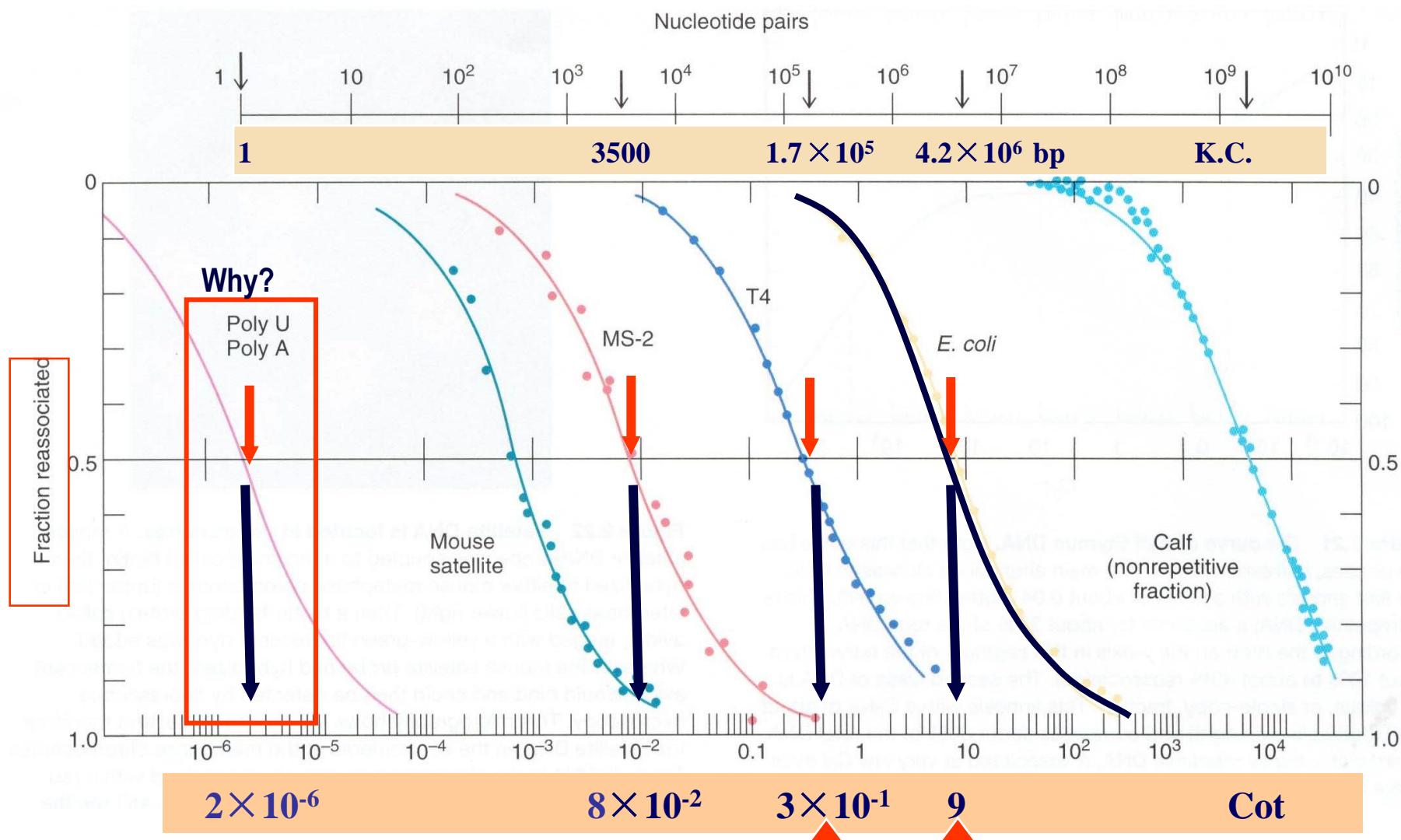
2. 4. 基因概念

的多样性重复基因



重复基因(Repetitive gene)

2.4.1. 重复序列的发现与证实



(来源: 不详)

单一序列长度愈小（重复度愈大）

kinetic complexity(K.C.) 愈小

$C_0 t_{(1/2)}$ 值小（复性愈快）

poly(A) K.C. = 1

$C_0 t_{(1/2)} = 2 \times 10^{-6}$

T₄ DNA K.C. = 1.7×10^5

$C_0 t_{(1/2)} = 0.3$

E.coli DNA K.C. = 4.2×10^6 bp

$C_0 t_{(1/2)} = 9$

K.C. 与 $C_o t_{(1/2)}$ 值呈正比 K.C. = $K \times C_o t_{(1/2)}$

- poly(A) K.C. = 1 ; $C_o t_{(1/2)} = 2 \times 10^{-6}$

比例常数 $K = K.C. / C_o t_{(1/2)} = 1 / 2 \times 10^{-6}$
= 5×10^5

- T₄ DNA K.C. = 1.7×10^6 ; $C_o t_{(1/2)} = 0.3$

比例常数 $K \approx 1.7 \times 10^6 / 0.3$
= 5×10^5

- E.coli DNA K.C. = 4.2×10^6 bp ; $C_o t_{(1/2)} = 9$

比例常数 $K \approx 4.2 \times 10^6 / 9$
= 5×10^5

K.C. 与 $C_0 t_{(1/2)}$ 值呈正比

$$K.C. = K \times C_0 t_{(1/2)}$$

$$Bp = (bp \times L / M \times S) \times (M \times S / L)$$

在特定的实验条件下
尽管不同的DNA分子K.C.值不同， $Cot_{(1/2)}$ 也不相同，
但他们的比例常数是相同的

$$K = 5 \times 10^5$$

任一DNA分子

$$K.C. = 5 \times 10^5 \times C_0 t_{(1/2)}$$

K.C. 与 $C_0 t_{(1/2)}$ 值呈正比

Kinetic Complexity of *E.coli*

==

$C_0 t_{(1/2)}$ of *E.Coli* DNA

Kinetic Complexity of any genome DNA

$C_0 t(1/2)$ of any genome DNA

$C_0 t(1/2)$ of any genome DNA

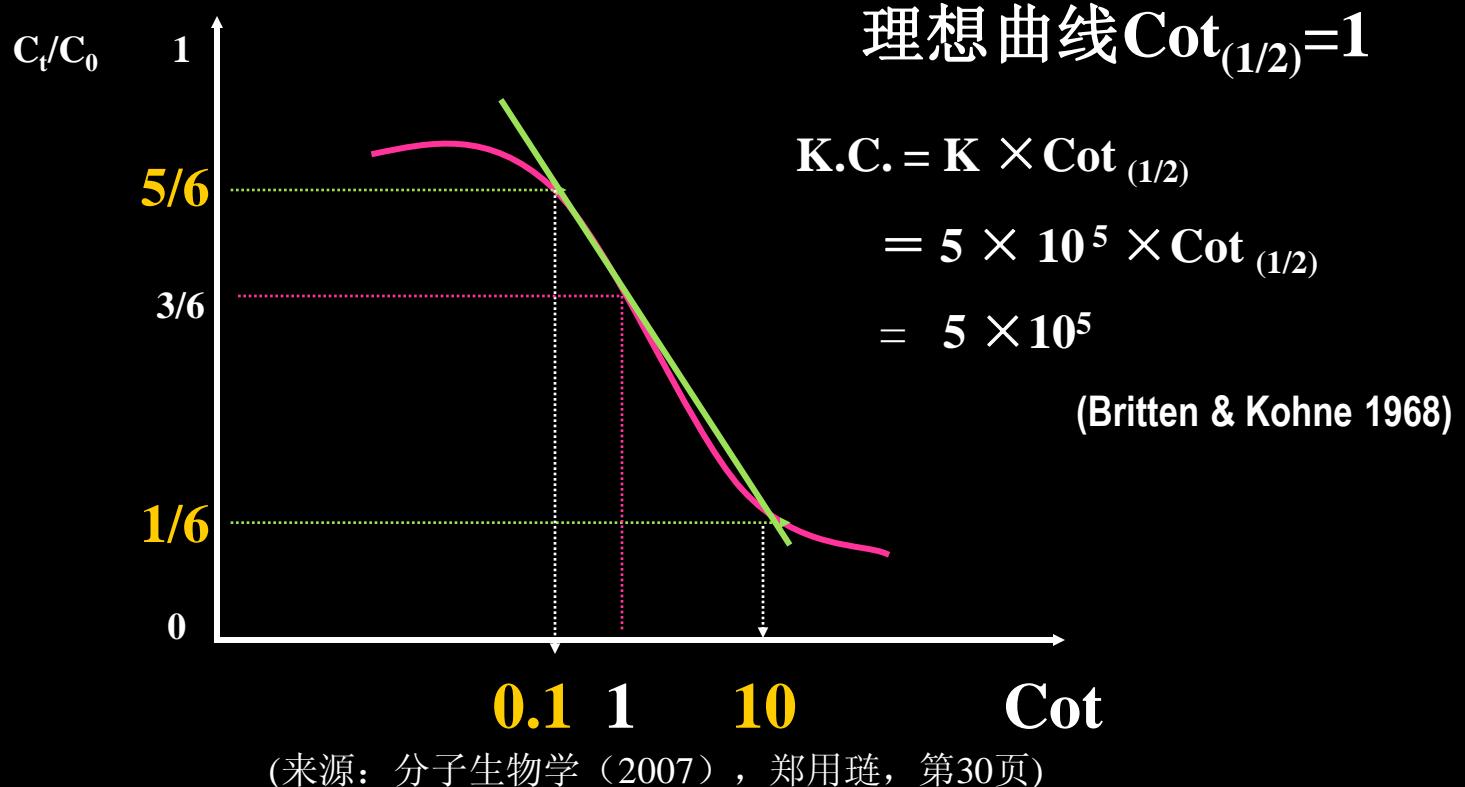
==

$C_0 t_{(1/2)}$ of *E.Coli* DNA

Kinetic Complexity of any genome DNA

Kinetic Complexity of *E.coli*

● 真核生物复性动力学研究

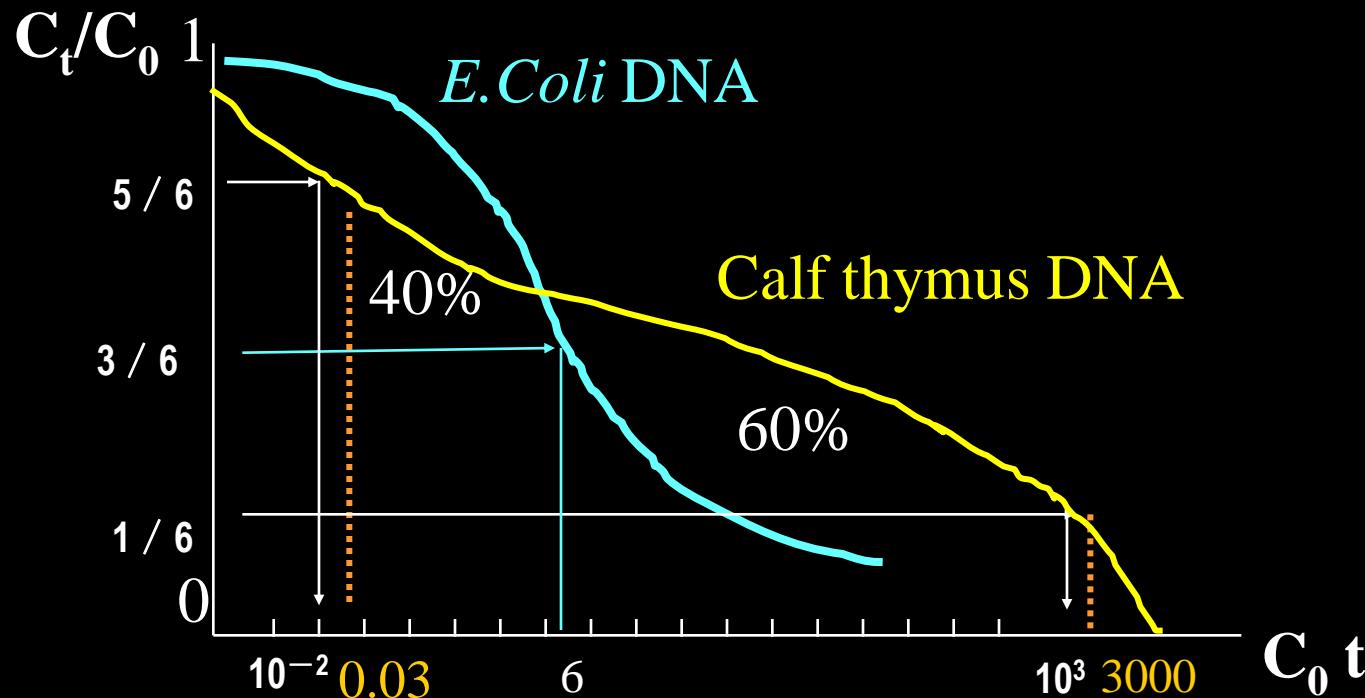


(来源: 分子生物学 (2007), 郑用琏, 第30页)

$$\frac{\frac{1}{6} \text{Ct/C}_0 = \text{Cot}_{(1/6)}}{\frac{5}{6} \text{Ct/C}_0 = \text{Cot}_{(5/6)}} \asymp 10^2$$

含有多复性蛋白组分的以上的组分

● 真核生物复性动力学研究



$$\frac{\frac{1}{6}C/C_0 = \text{Cot}_{(1/6)}}{\frac{5}{6}C/C_0 = \text{Cot}_{(5/6)}} > 10^2$$

Calf thymus DNA 含有两个以上的
K.C. 不同的组分

$C_0 t_{(1/2)}$ 值矫正

$$40\% \times 0.03 = 0.012 \quad 60\% \times 3000 = 1800$$

$C_0 t_{(1/2)}$ of any genome DNA

$C_0 t_{(1/2)}$ of *E.coli* DNA

=

Kinetic Complexity of any genome DNA

Kinetic Complexity of *E.coli*

$C_0 t_{(1/2)}$ of *E.coli* DNA = 6

Kinetic Complexity of *E.coli* = 4.6×10^6

60%DNA的总长度 = 单一序列的Kinetic complexity

$$\frac{1800}{6} = \frac{X}{4.6 \times 10^6} \quad X = 13.8 \times 10^8 \text{ bp}$$

40%DNA 的总长度 (chemical complexity)

$$13.8 \times 10^8 \text{ bp} \times 4/6 = 9.2 \times 10^8 \text{ bp}$$

40%DNA中单一序列的 Kinetic complexity

$$\frac{0.012}{6} = \frac{Y}{4.6 \times 10^6}$$

Y = 9200 bp

40%DNA序列中单一序列的重复次数 (F)

Repetitive frequency = C.C / K.C = 9.2×10^8 bp / 9200 bp

= **100,000 copies**

$$K.C. = K \times C_0 t_{(1/2)}$$

$C_0 t_{(1/2)}$ 值矫正

$$40\% \times 0.03 = 0.012 \quad 60\% \times 3000 = 1800$$

60%DNA的总长度 = 单一序列的 Kinetic complexity

$$X_{K.C} = 1800 \times 5 \times 10^5 \text{ bp} = 9 \times 10^8 \text{ bp}$$

40%DNA 的总长度 (chemical complexity)

$$9 \times 10^8 \text{ bp} \times 4/6 = 6 \times 10^8 \text{ bp}$$

$C_0 t_{(1/2)}$ 值矫正

$$40\% \times 0.03 = 0.012 \quad 60\% \times 3000 = 1800$$

40%DNA中单一序列的 Kinetic complexity

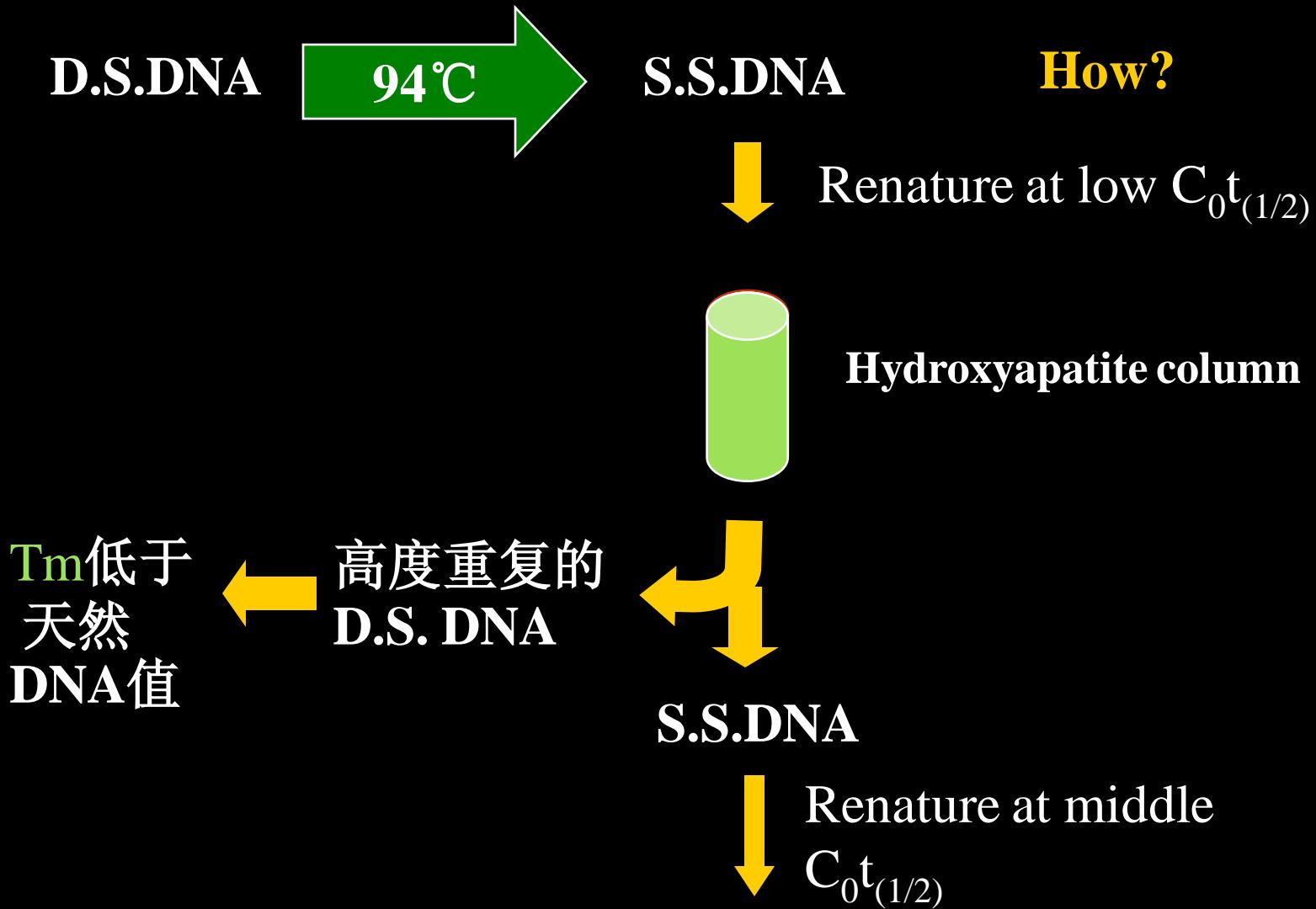
$$Y_{K.C} = 0.012 \times 5 \times 10^5 \text{ bp} = 6000 \text{ bp}$$

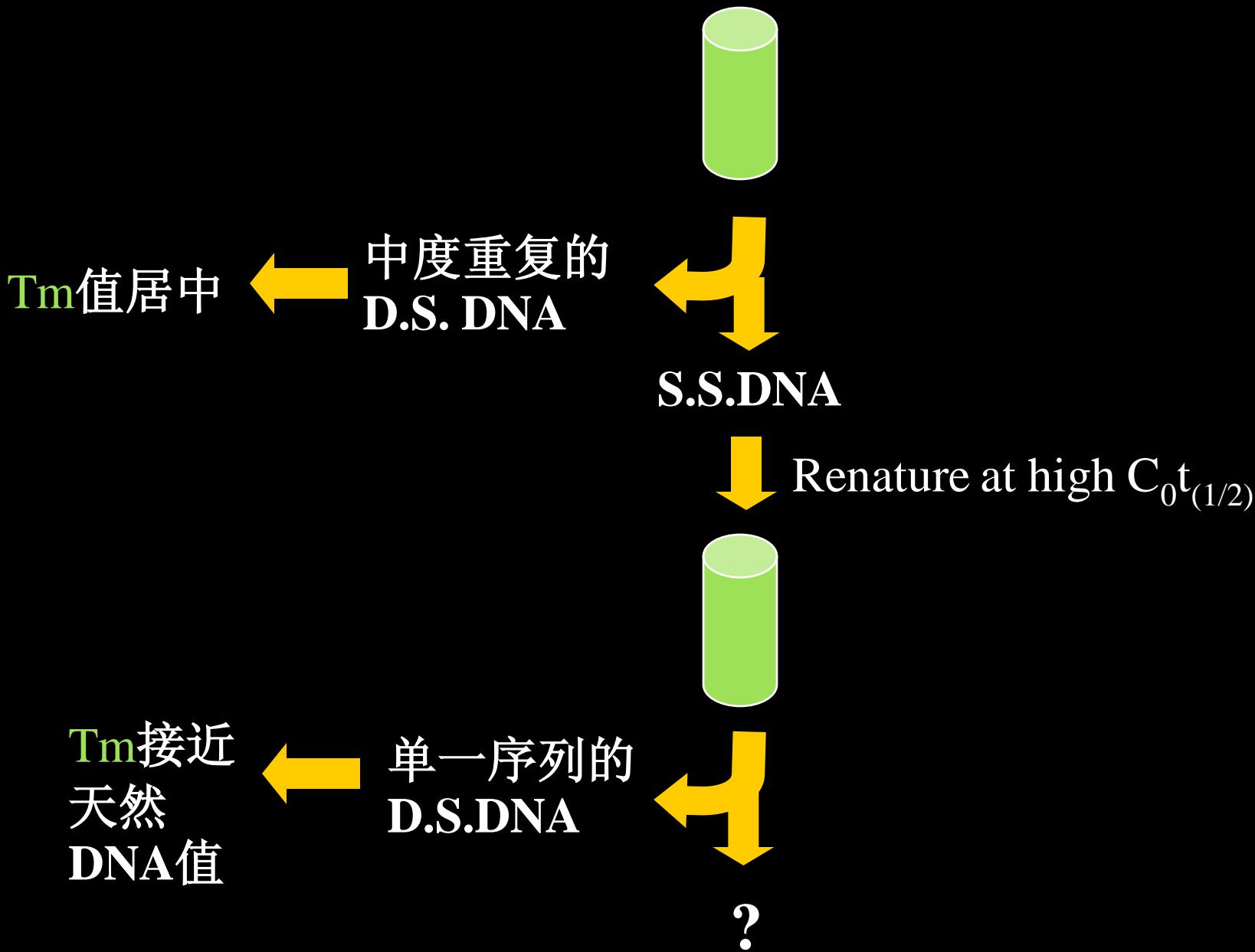
40%DNA序列中单一序列的重复次数 (F)

$$\text{Repetitive frequency} = C.C / Y_{K.C} = 6 \times 10^8 \text{ bp} / 6000 \text{ bp}$$

$$= 100,000 \text{ copies}$$

2.4.2 重复序列复性的相对性



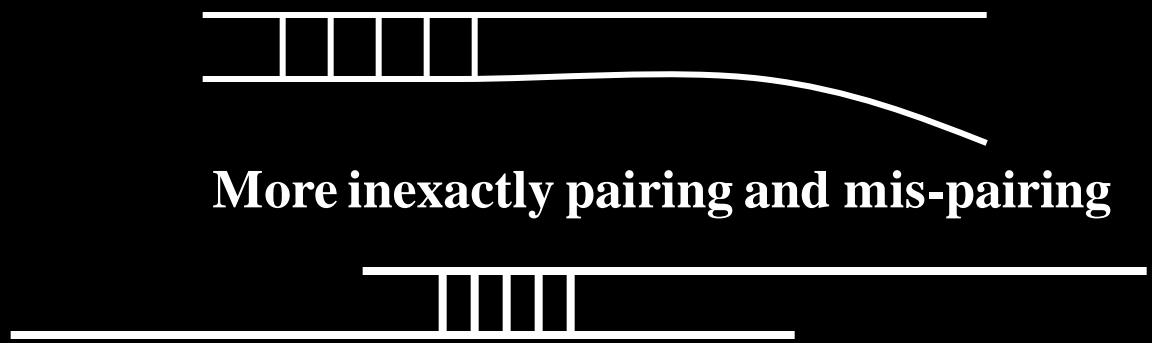




结果表明：

a) Non-repetitive S.S. DNA → Renatured by exactly pairing

b) Repetitive S.S. DNA → More renatured by partially pairing



天然DNA与复性DNA之间比较

ΔT_m 相差1—1.5度时，预计约有1%的碱基错配

部分配对，错误配对愈多，变性时 T_m 值愈低

PCR, 分子杂交中，杂交温度（复性）一般为

$$T_m - (25 \sim 30^\circ\text{C}) = 55 \sim 65^\circ\text{C}$$

高于**55~65°C**，为高严谨杂交条件

(用于单一序列探针)

低于**55~65°C**，为低严谨杂交条件

(用于重复序列探针)

} Why ?

2.4.3. 重复序列分类

a) 高度重复序列 (High repetitive sequence)

2-10 bp / copy

10^5 - 10^6 copies / genome

$C_0 t_{(1/2)} < 0.001$ 多为串联重复排列

分布于着丝点, 端粒区, 结构基因两侧

} Microsatellite DNA

heterochromatin

5-50 copies / genome → Minisatellite DNA

(Variable number tandem repeats . VNTR)

10^4 bp fragment  CsCl gradient centrifugation



Sequence in satellite DNA

bp of repeat unit sequence

| | | |
|------------|---|-----------------------------|
| Sea crab | 2 | <u>ATATAT.....</u> |
| Drosophila | 5 | <u>ATAATATAAT.....</u> |
| Mouse | 9 | <u>GAAAAAATGAGAAAAAATGA</u> |

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)

CNV (copy number variation) !

b) 中度重复序列

(middle repetitive sequence)

0.1-1 Kb / copy

10-10⁴ copies / genome

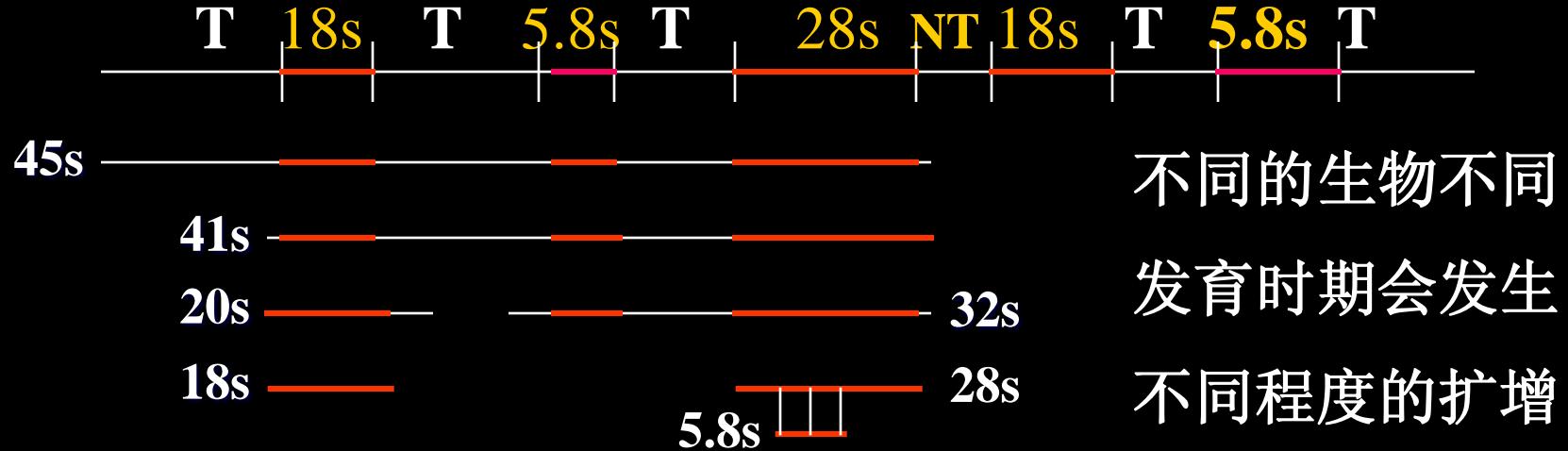
$C_0 t_{(1/2)} \sim 0.001 - 0.1$

rDNA tDNA

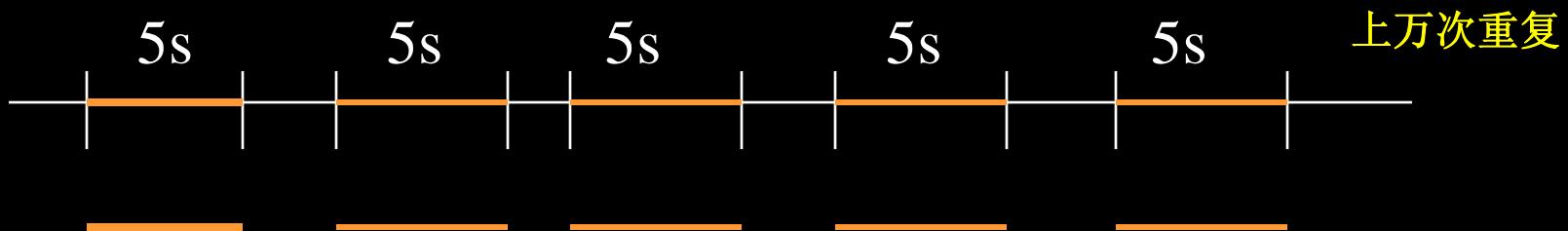
Alu family

Histone gene cluster

rDNA gene family



5s rDNA



middle repeat gene

数量性状基因QTG (量多, 微效? 效等? 累加)

不等于中度重复基因

特点

- {
 - 拷贝重复, 多量
 - 序列多为相似
 - 排列成束
 - 功能完全相同
 - 具有进化的整体性, 累积突变

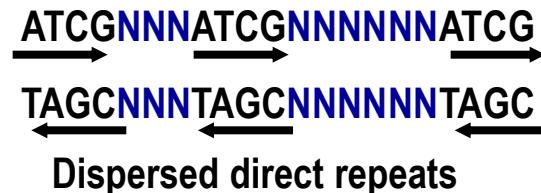
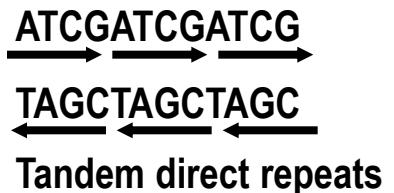
rDNA

tDNA

Histone

Structure of repetitive sequence of DNA

DR direct repeats



IR Inverted repeats

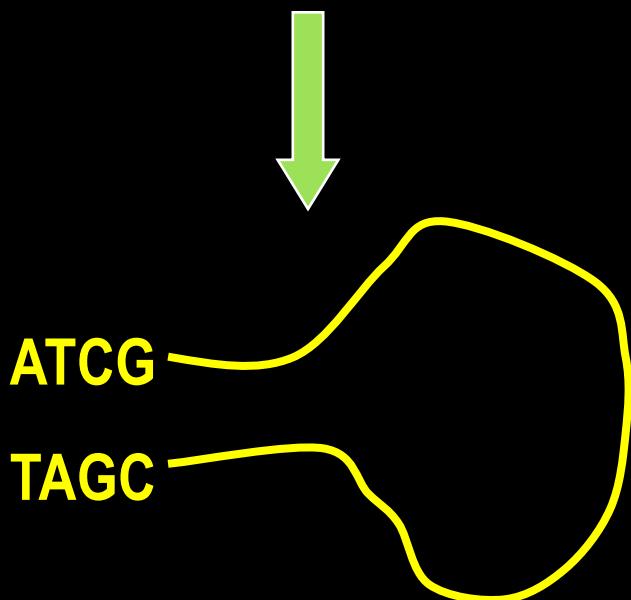


Mirror Structure



Structure of Inverted repetitive sequence (IR) of DNA

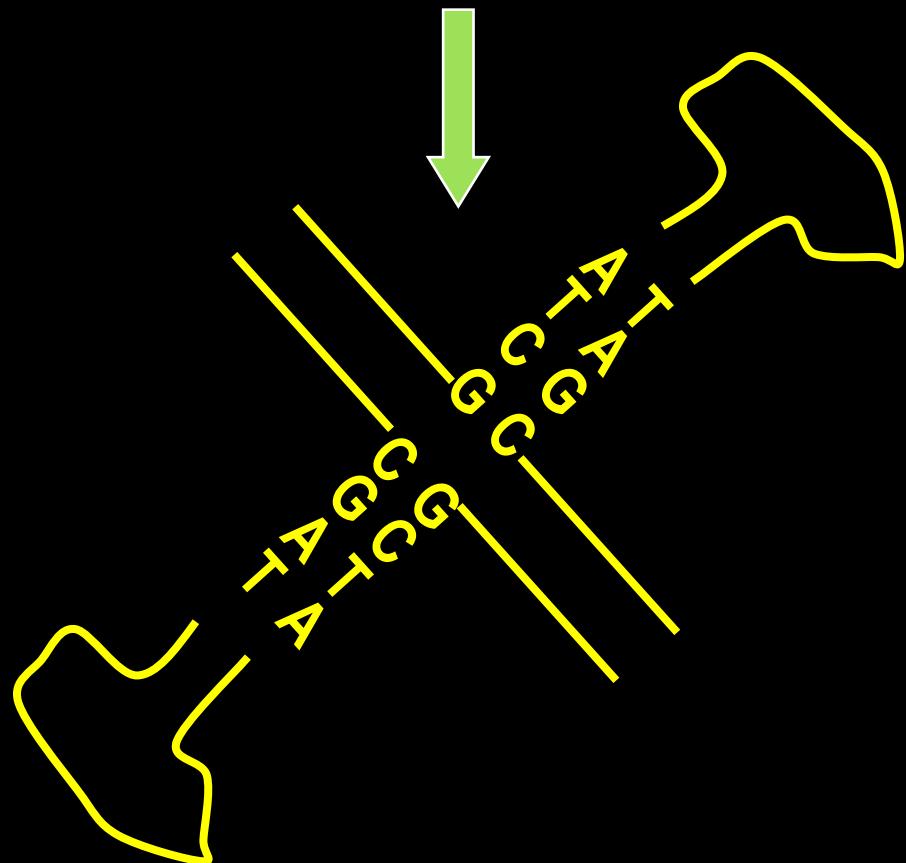
ATCGNNNNNNNNCGAT



When S.S. DNA dNts contains two sequences that are complementary, and form a hairpin or stem-loop structure.

In D.S. DNA
such structure
consist of two
copies of an
identical
sequence
present in the
inverted
orientation

ATCGNNNNNCGAT
TAGCNNNNNGCTA



c) 单拷贝序列

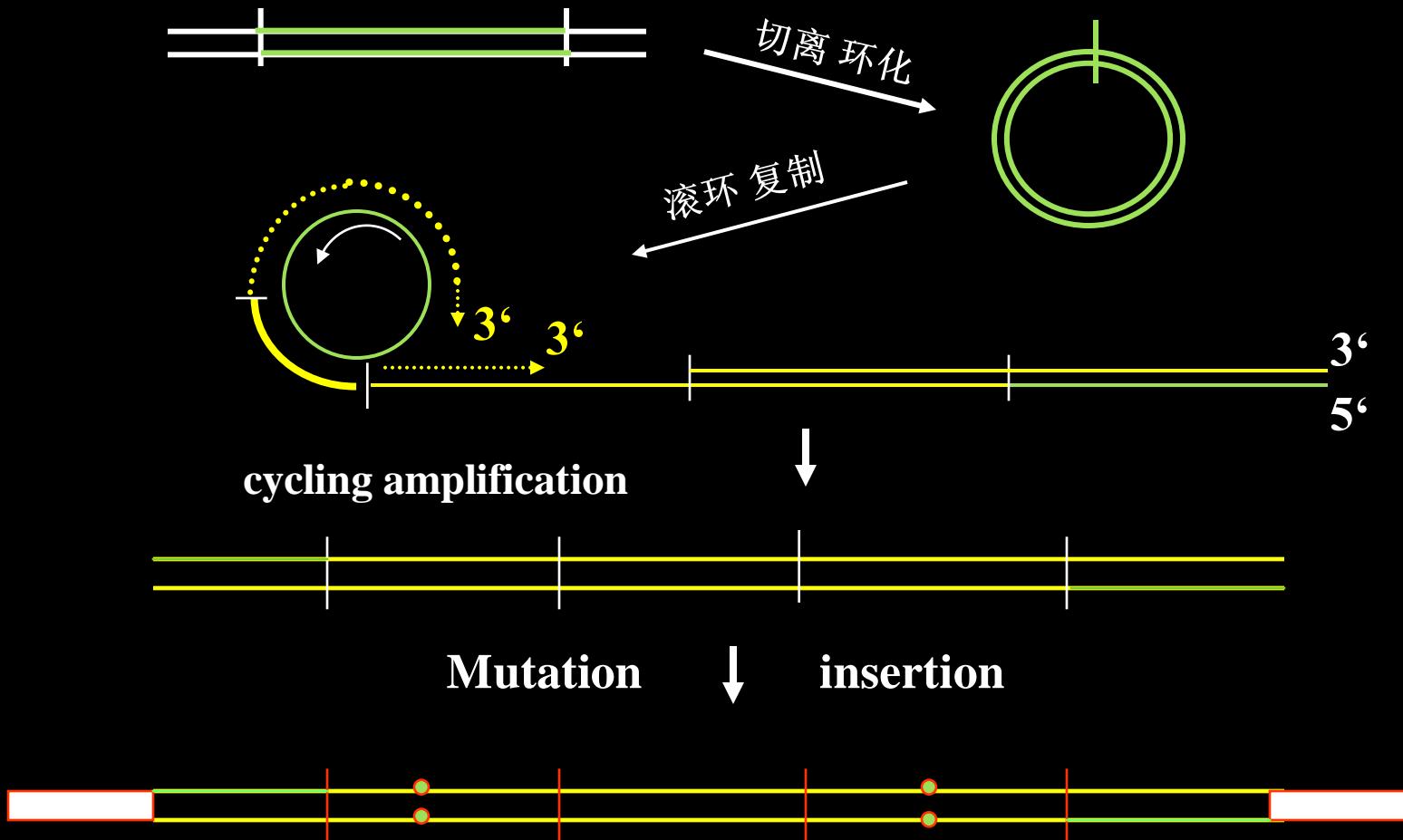
(single copy sequence, 1-3 copies)

多为结构基因

| | | | |
|--------|--------|---|---------------------|
| 低等真核生物 | 10-20% | } | repetitive sequence |
| 高等植物 | 80% | | |
| 高等动物 | 50% | | |

2.4.4. 重复序列形成的理论

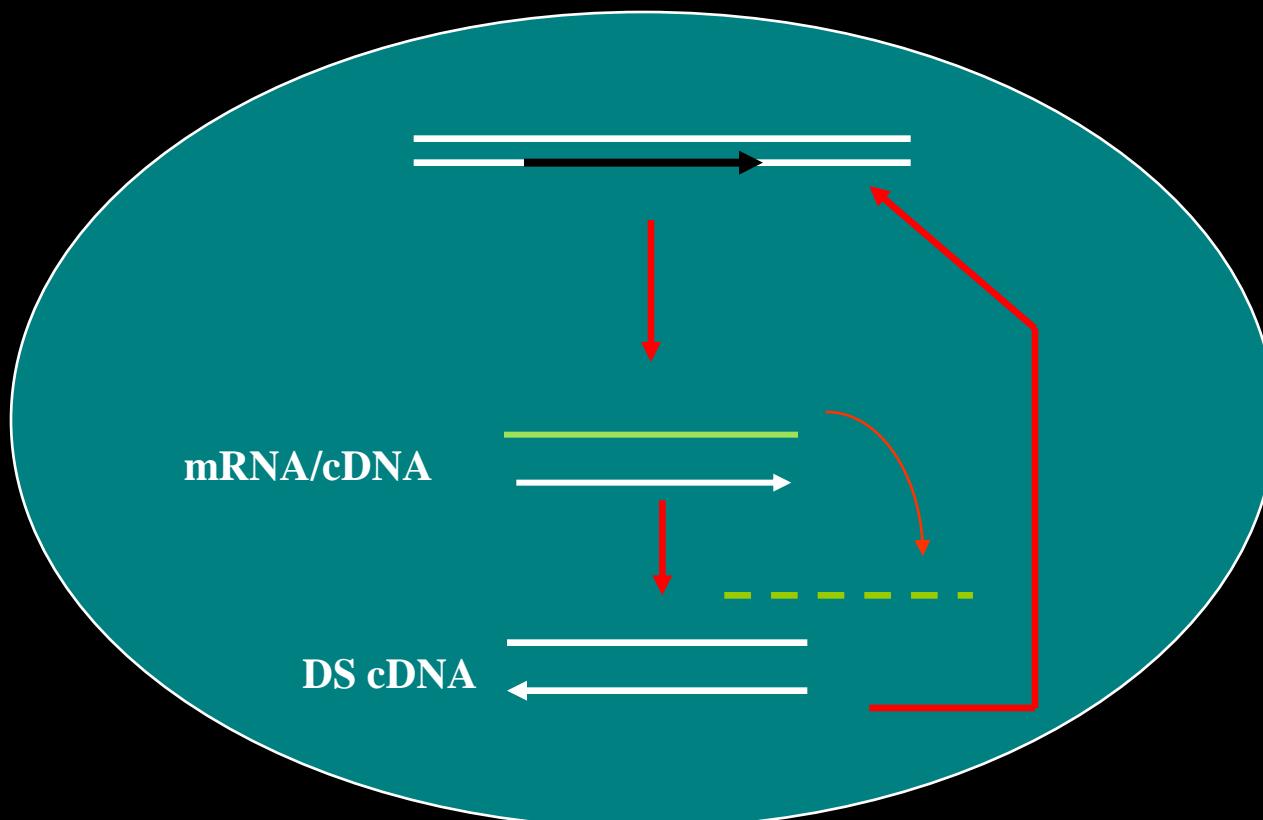
a) 滚环扩增—突变 (Amplification-Mutation)



(来源: 分子生物学 (2007), 郑用琏, 第55页)

b) 反转座插入 (retro-transposition)

DNA → RNA → DNA

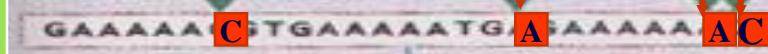


c) 跳跃复制 saltatory replication

9 bp repeat



27 bp repeat



58 bp repeat



58 bp repeat



mutation

Saltatory replication

116 bp repeat



Mutations:



Amplify 23.4 bp unit for present satellite

(来源: 不详)

基因的重复可以通过复制产生
(**duplication**)

复制后的拷贝发生突变和突变累积
(**divergence**)

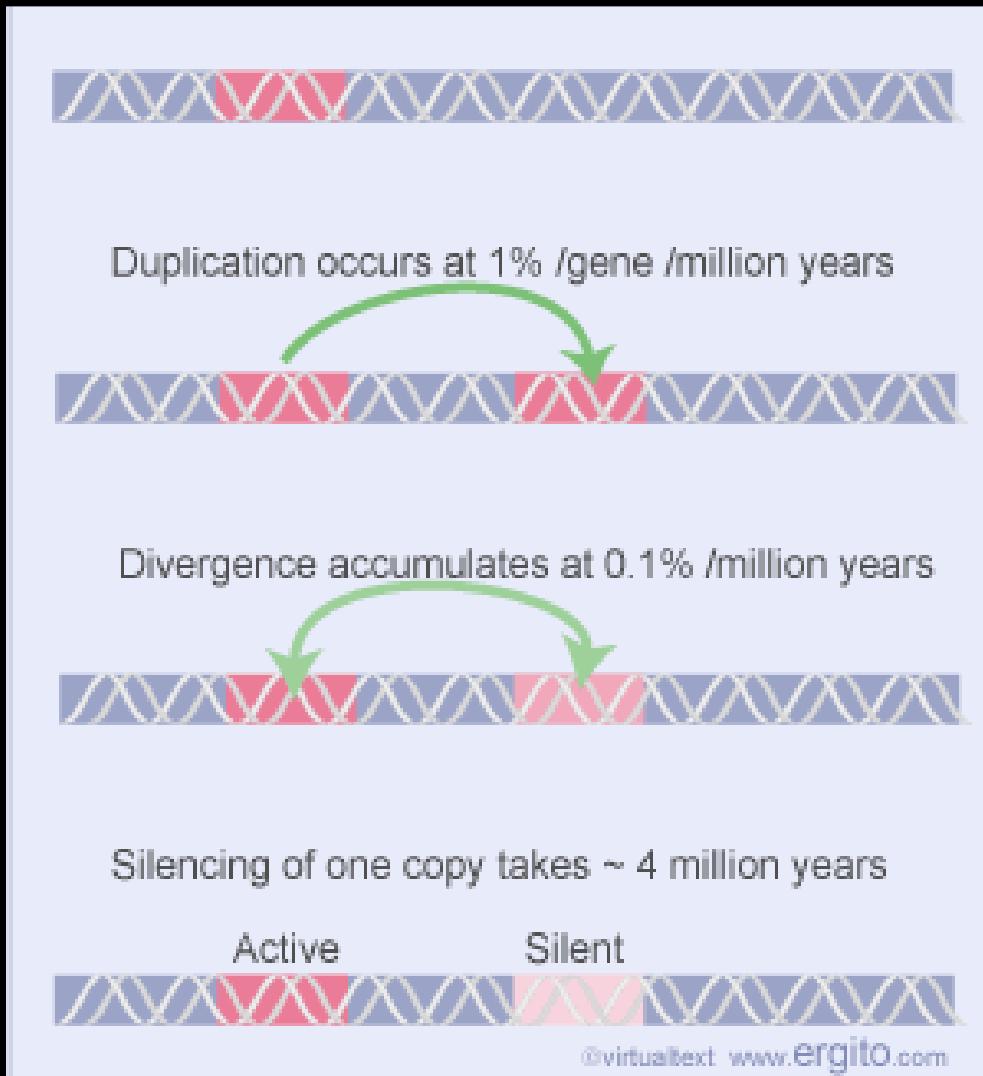


形成基因簇 **gene cluster**



分化成执行不同功能的基因

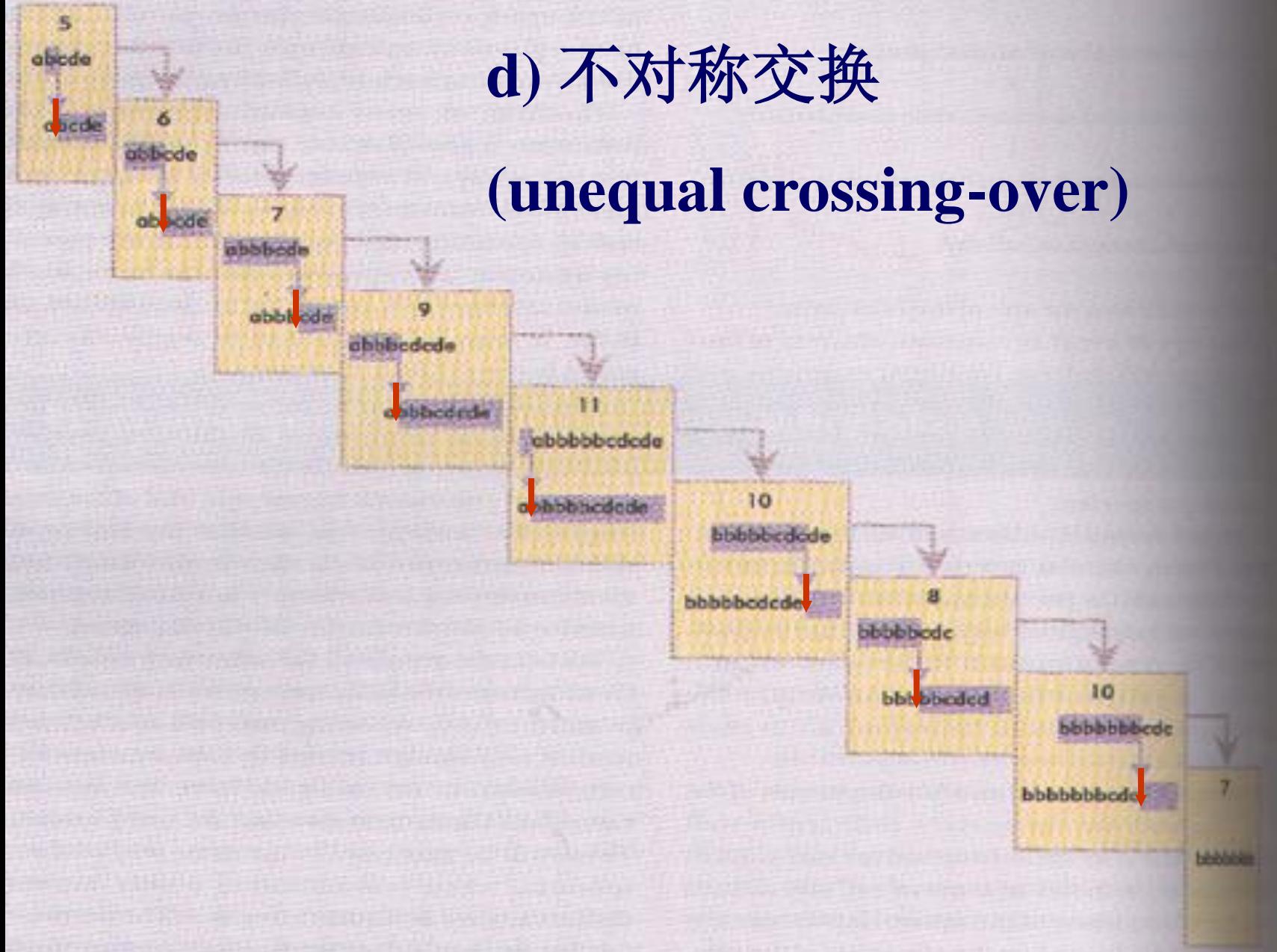
抗病基因簇的形成



**Gene duplication
is a major force in
evolution**

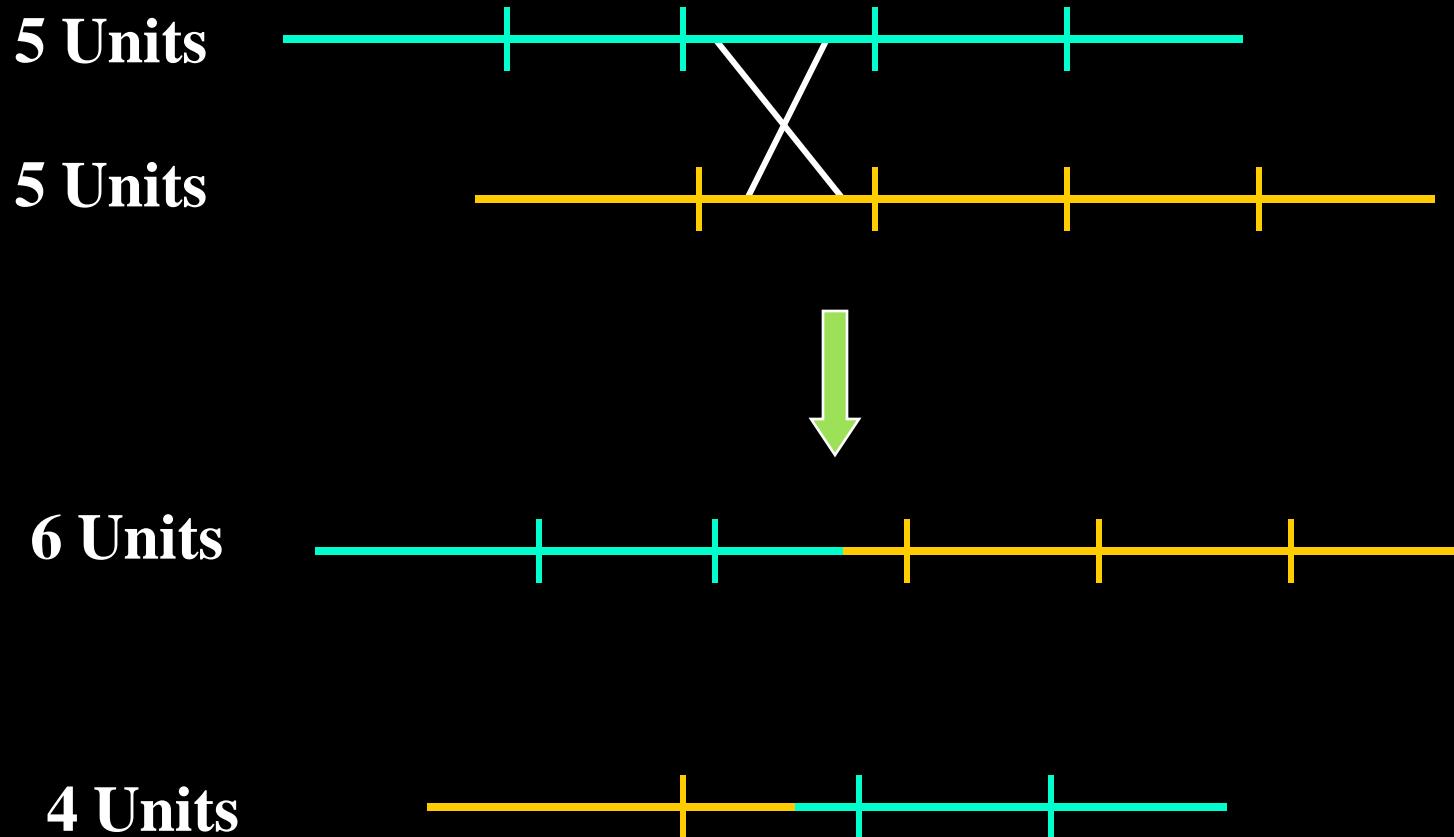
(来源: 不详)

d) 不对称交换 (unequal crossing-over)

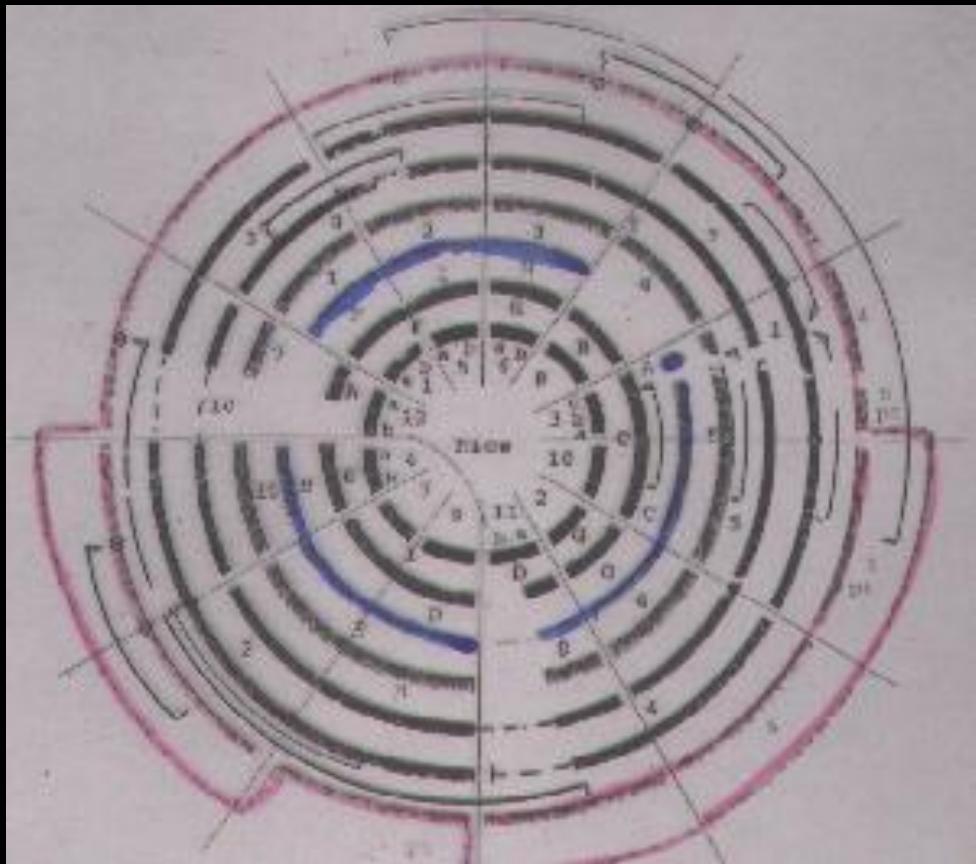


(来源: 不详)

unequal crossing-over



基因组比较研究



禾本科植物
基因组比较

(来源: 不详)

- 基因组比较研究表明；

$5 \times \text{Rice genome} < \text{Maize genome}$

$40 \times \text{Rice genome} < \text{Wheat genome}$

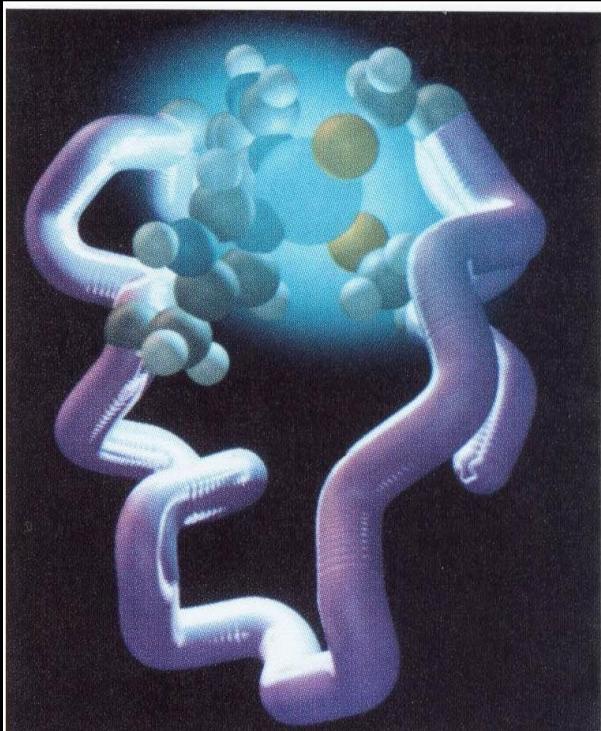
但基因组成，排列表现惊人的相似

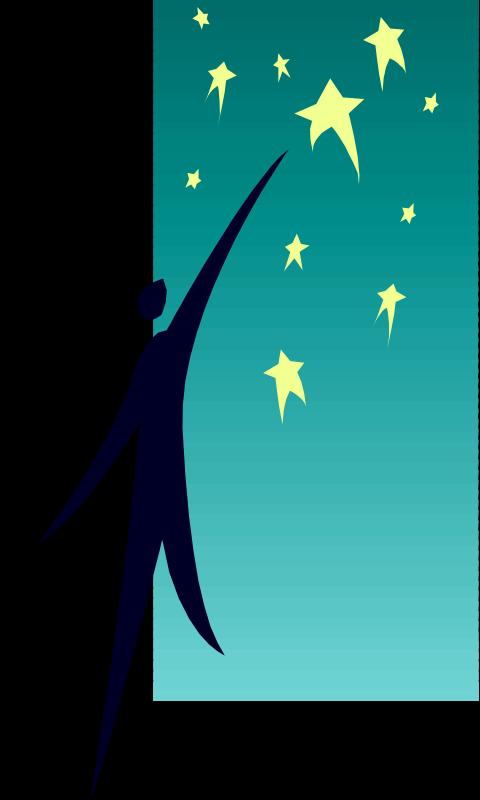
差异多存在于repetitive sequence

2. 5. 基因概念

的多样性-间隔基因

间隔基因 (splitting gene)





间隔基因的发现 (mini revolution)

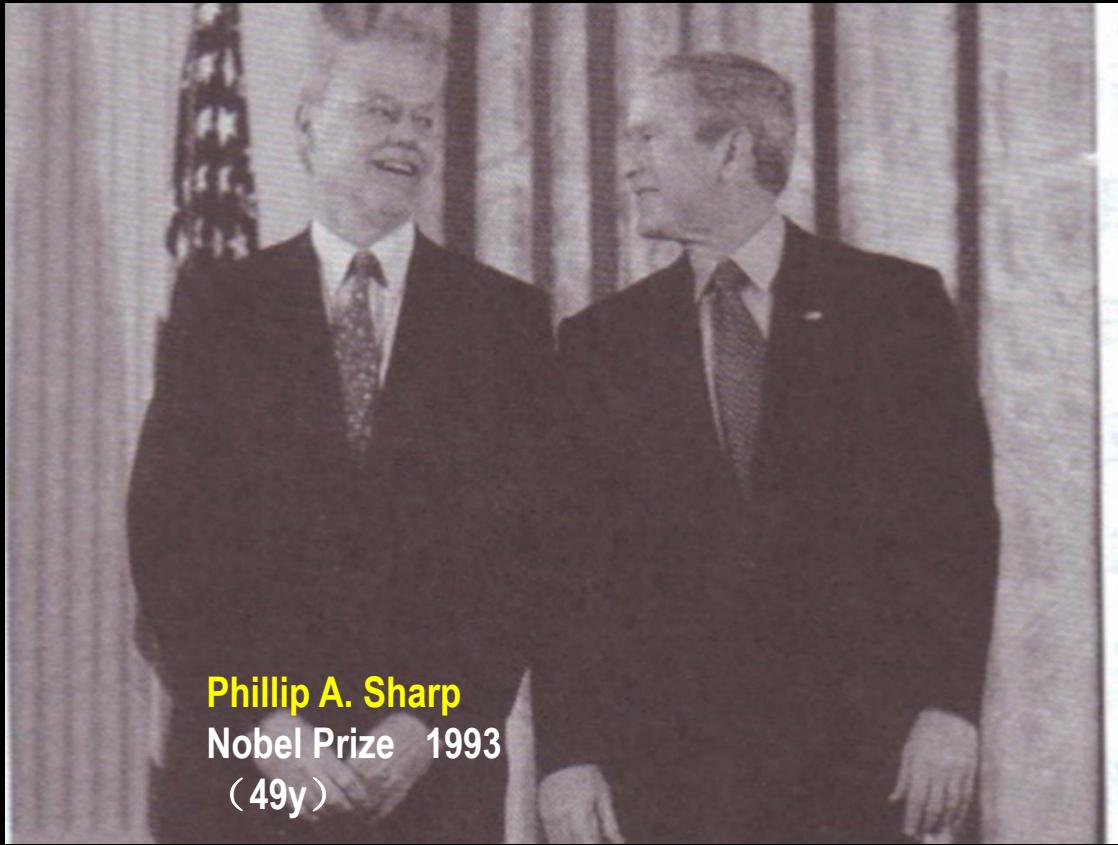
(Source:
<http://www.nndb.com/people/128/000133726/>)



Richard J. Roberts Phillip A. Sharp

(Source:
<http://www.nndb.com/people/088/000133686/>)

Nobel Prize 1993



2006.2.13 布什给 Philip Sharp 颁发国家科技奖章

曾获得40项奖励，拥有4个院士称号（美国科学院、美国文理科学院、
美国医学研究院、美国哲学研究院）

(Source:
<http://www.nndb.com/people/478/000131085/>)



● 限制性核酸内切酶的发现

1965 Arber. W



Werner Arber
Nobel Prize 1978



- a) *E.coli* restriction alien DNA
with restriction endonuclease (RE)
- b) Modification enzyme in host
- c) Alien DNA fate of be restricted or
be modified
- d) There are different RE
in different host of bacterial

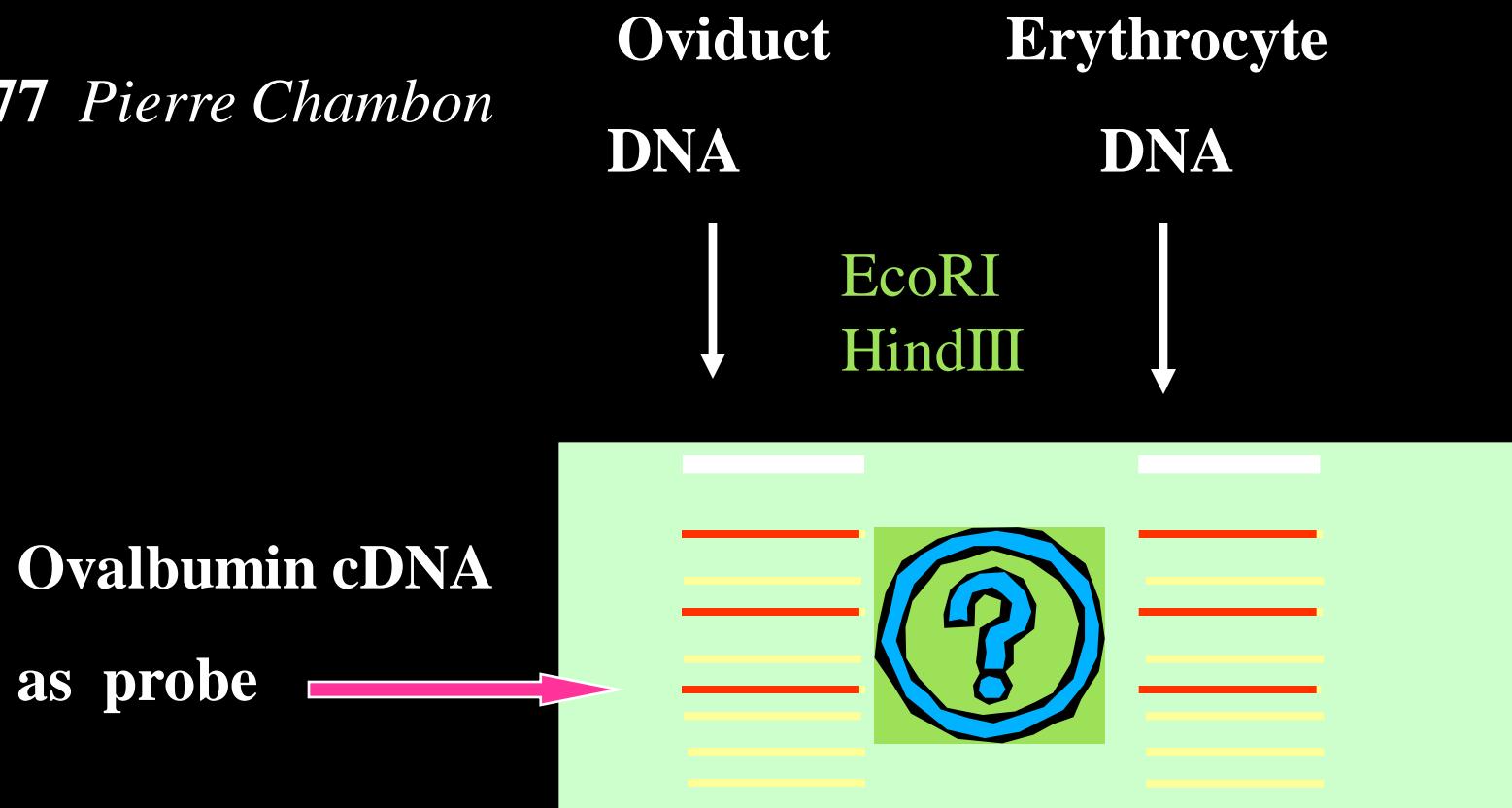
1 Restriction Endonuclease 的种类

| | Type II Enzyme | Type III Enzyme | Type I Enzyme |
|-----------------------------|--------------------------------------|---|---|
| Protein structure | Separate endonuclease & methylase | Bifunctional enzyme of 2 subunits | Bifunctional enzyme of 3 subunits |
| Recognition site | 4-6 bp sequence, often palindromic | 5-7 bp Asymmetric sequence | Bipartite & asymmetric |
| Cleavage site | Same as or close to recognition site | 24-26 bp Downstream of recognition site | Nonspecific >1000 bp from recognition site |
| Restriction & methylation | Separate reactions | Simultaneous | Mutually exclusive |
| ATP needed for restriction? | No | Yes | Yes |

(来源: 不详)

● Splitting gene 的证实

1977 *Pierre Chambon*



cDNA 中没有 HindIII, EcoRI 切点！？



(Source: <http://www.nndb.com/people/128/000133726/>)

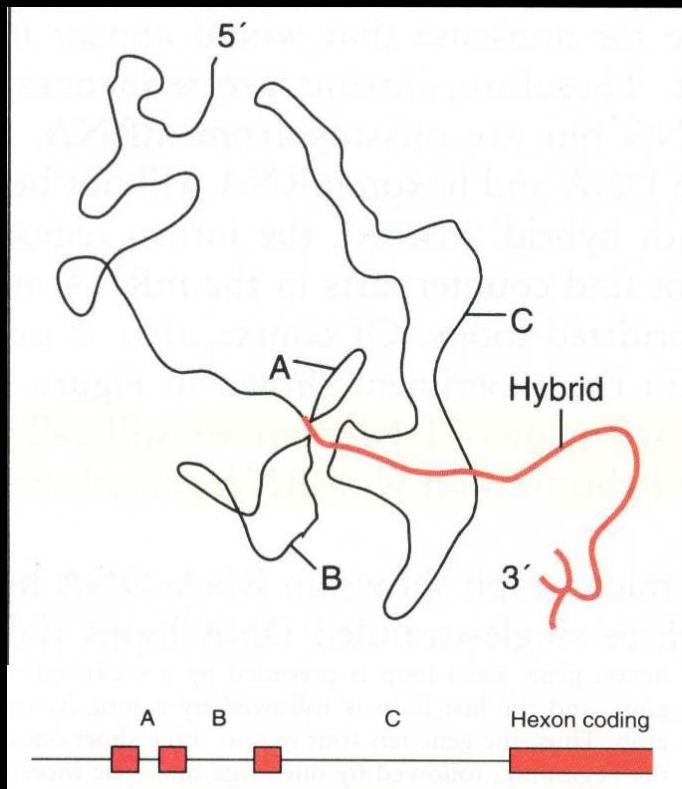
Sharp group

Roberts group

1977

**Splitting seq. in the 5' end of
Hexon cp gene of Ad2**

(Source: <http://www.nndb.com/people/088/000133686/>)



DNA of Ad2

mRNA of Hexon cp

And

**presentation in CSH
PNAS 74: 3173, 1977**

(来源: 分子生物学 (2007), 郑用琏, 第59页)

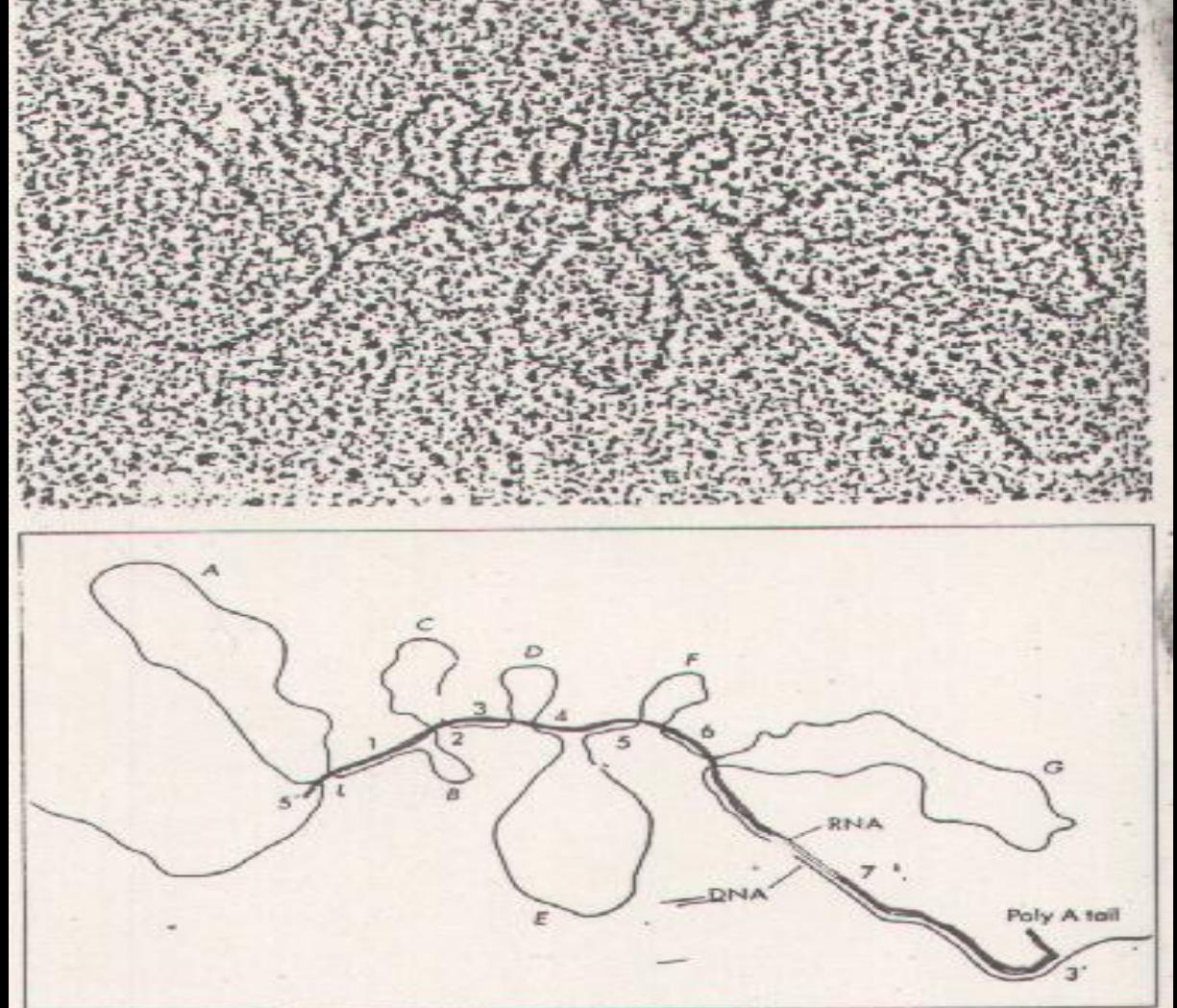
Total DNA of chicken



cDNA



7 DNA loop



(来源：分子生物学（2007），郑用琏，第59页)

Chambon was inspired by Berget's report

Berget, S. M., C. Moore, and P. Sharp. 1977. PNAS 74; 3171-3175

Ovalbumin cDNA RE map

Hinf I Hinf II Hea III **Hinf III** Pst I Hinf IV



(No EcoRI Hind III RE site)

the cDNA probe be digested by **Hinf III and PstI**

Labeling with p32

Probe A

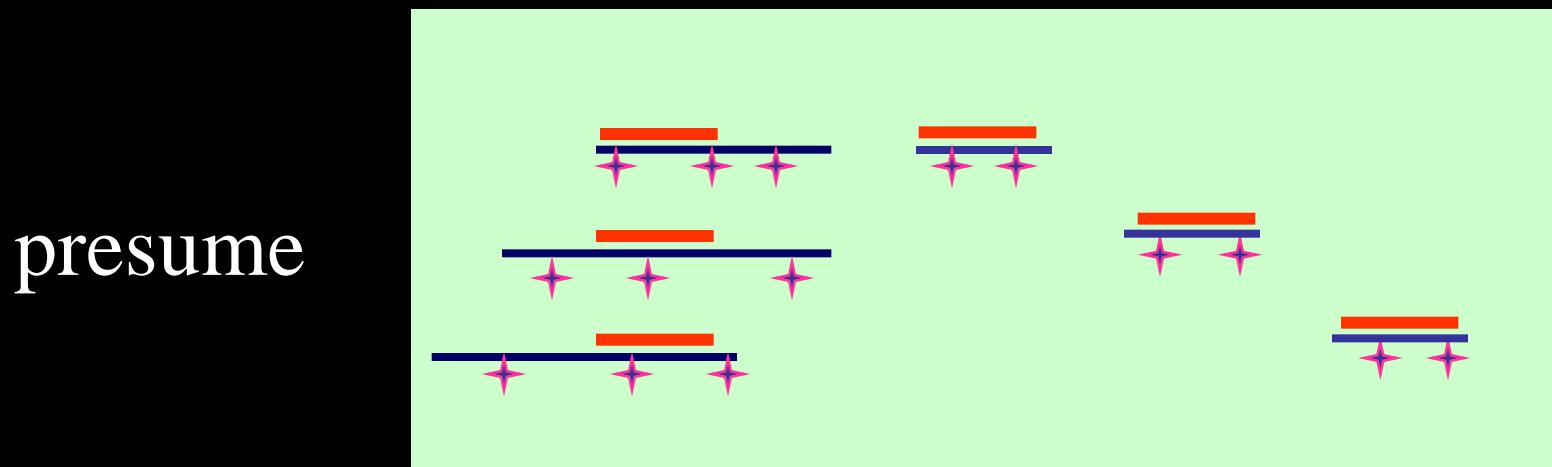
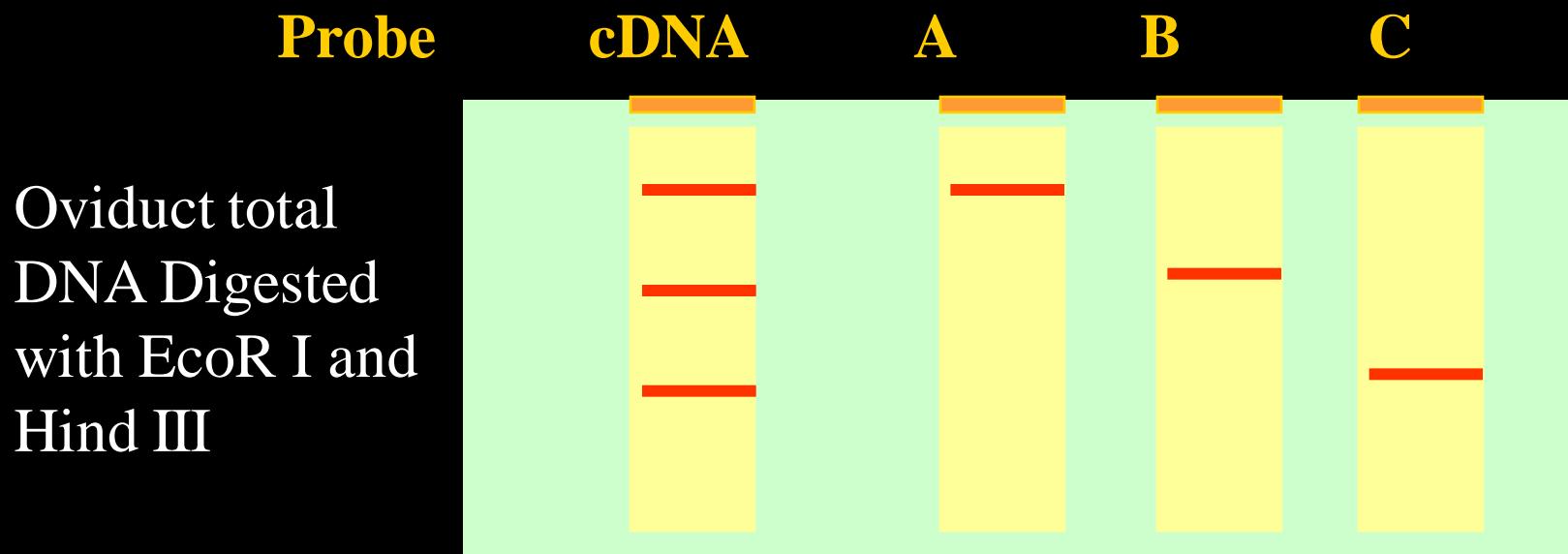


Probe B

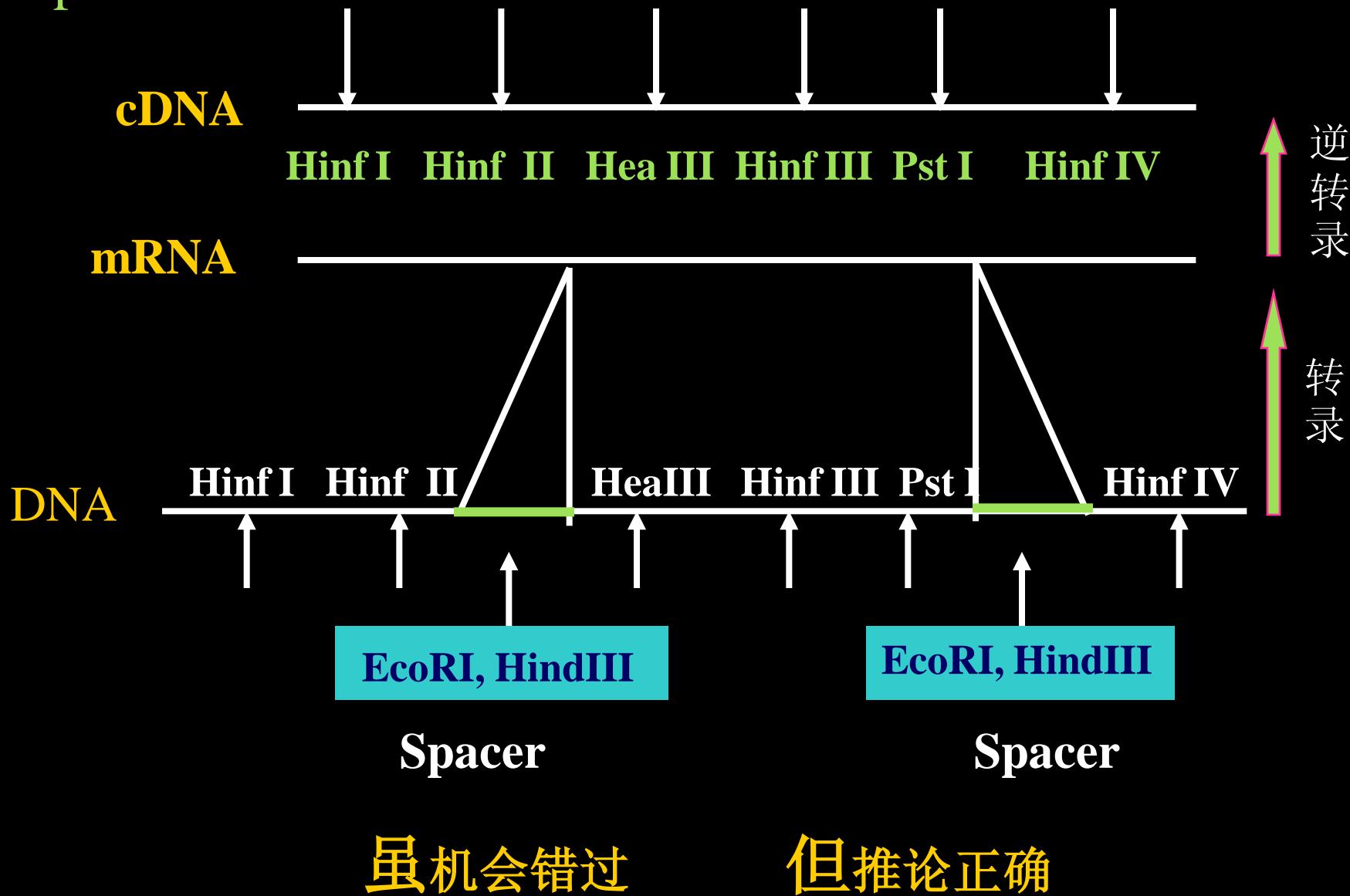


Probe C





presume



1978 Gilbert 真核生物基因的新概念

Exon (外显子)

is any segment of an interrupted gene that is represented in the mature RNA product

DNA 与成熟RNA间的对应区域

氨基酸的编码区 (amino acid coding region)

非间隔区 (unspacer)

Intron (内含子)

is a segment of DNA that is transcribed, but removed from within the transcript by splicing together the sequences (exons) on either side of it.

DNA 与成熟RNA间的非对应区域

氨基酸的非编码区 (non-coding region)

间隔区 (spacer)

Ovalbumin

DNA X cDNA

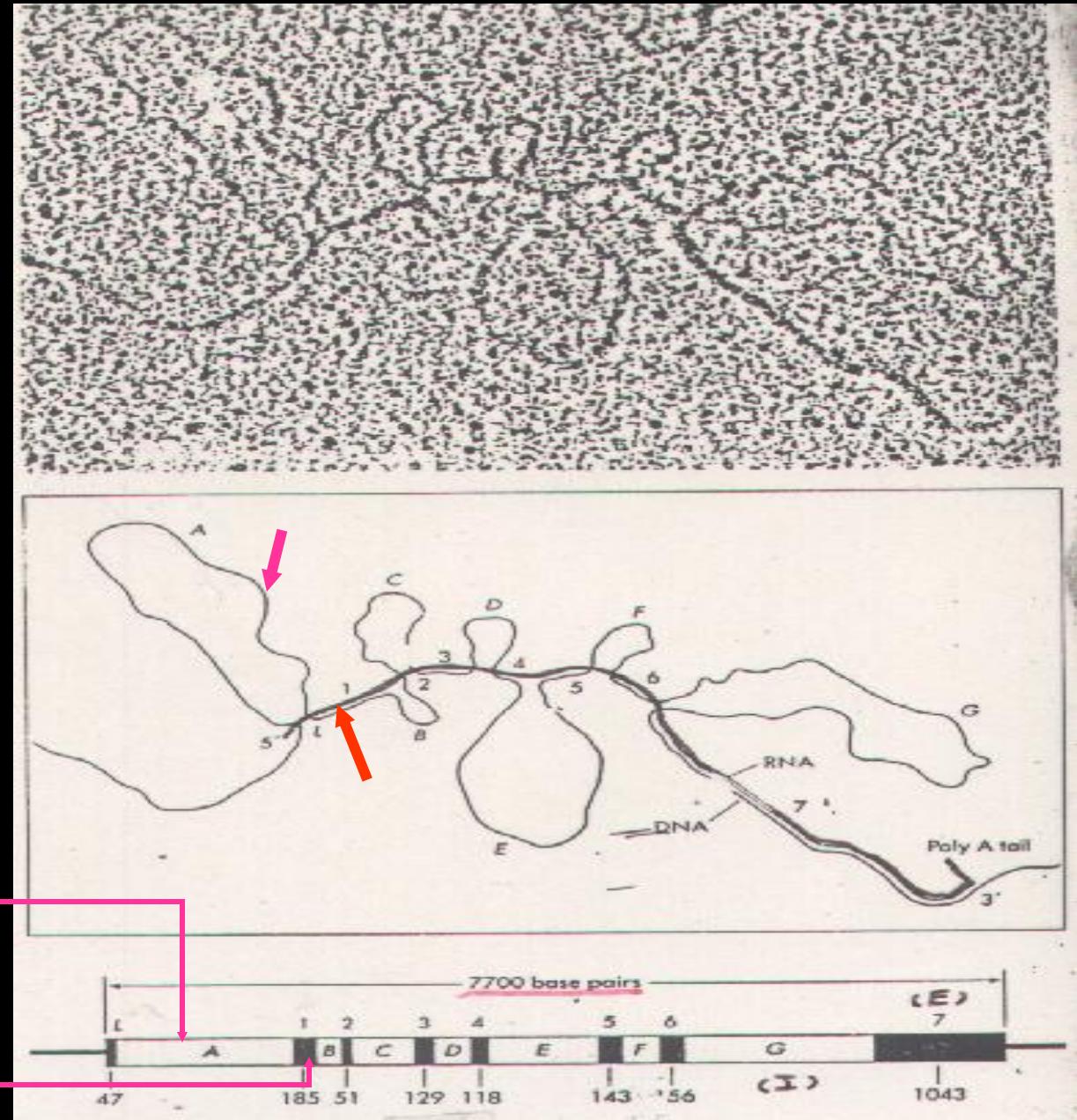


Electro-microscope



7 introns

8 exons



(来源：分子生物学（2007），郑用琏，第59页)

真核生物基因又称为

Splitting gene

Interrupted gene

间隔基因， 断裂基因

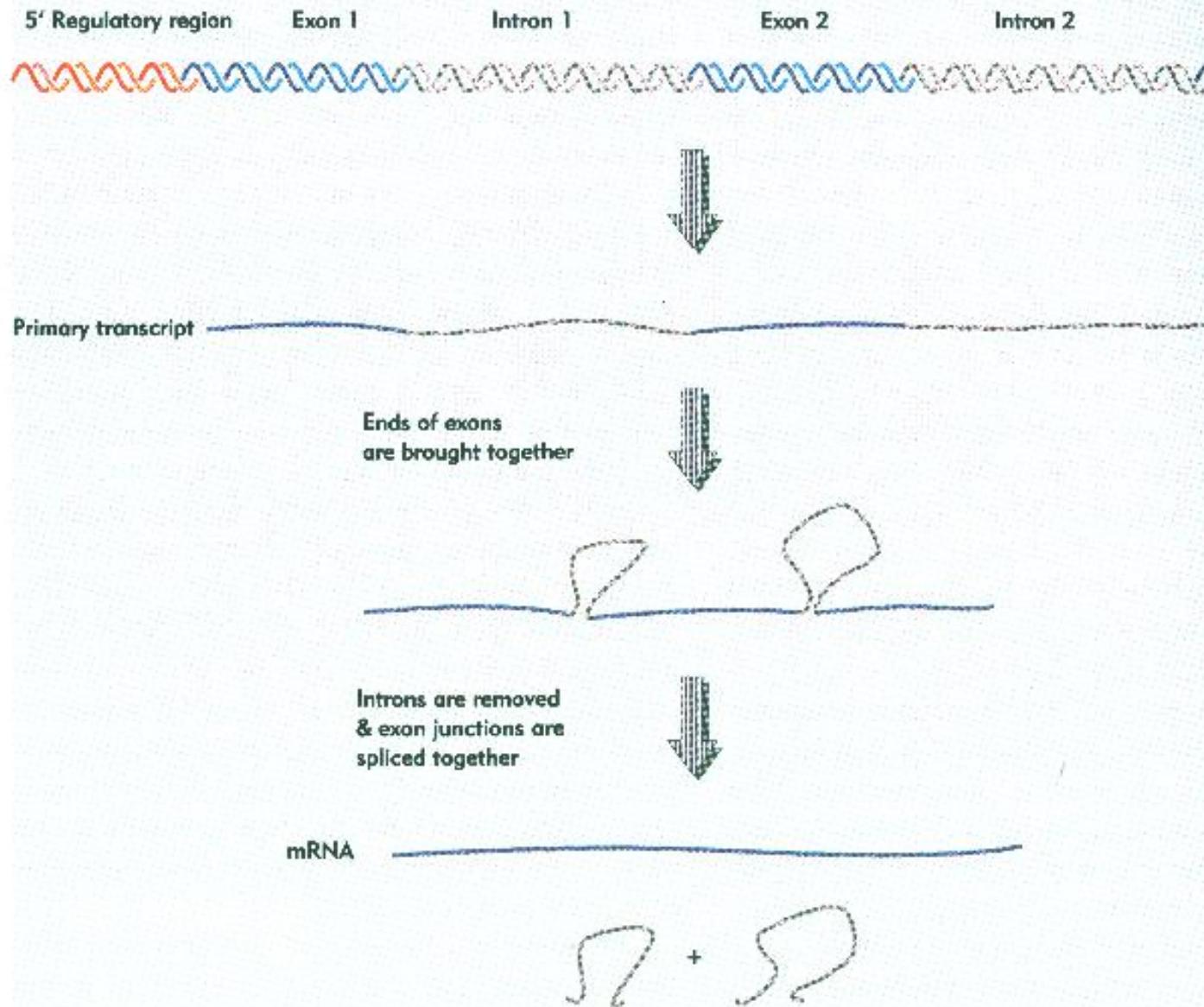
真核生物基因的转录物又称为

Precursor mRNA (pre-mRNA)

Heterogeneous nuclear RNA (HnRNA)

前体mRNA， 核内不均一 RNA

Splitting gene



Pre-mRNA

Introns

Removed

Exons
junctions

Splitting Gene

真核生物的结构基因是由若干exon 和intron 相间隔排列的序列组成的间隔基因

1 Splitting Gene 的普遍性

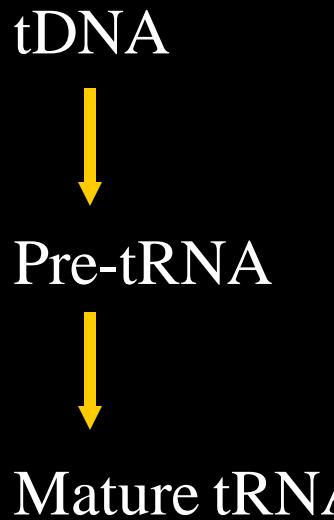
a) in Eukaryotes

most of structural gene

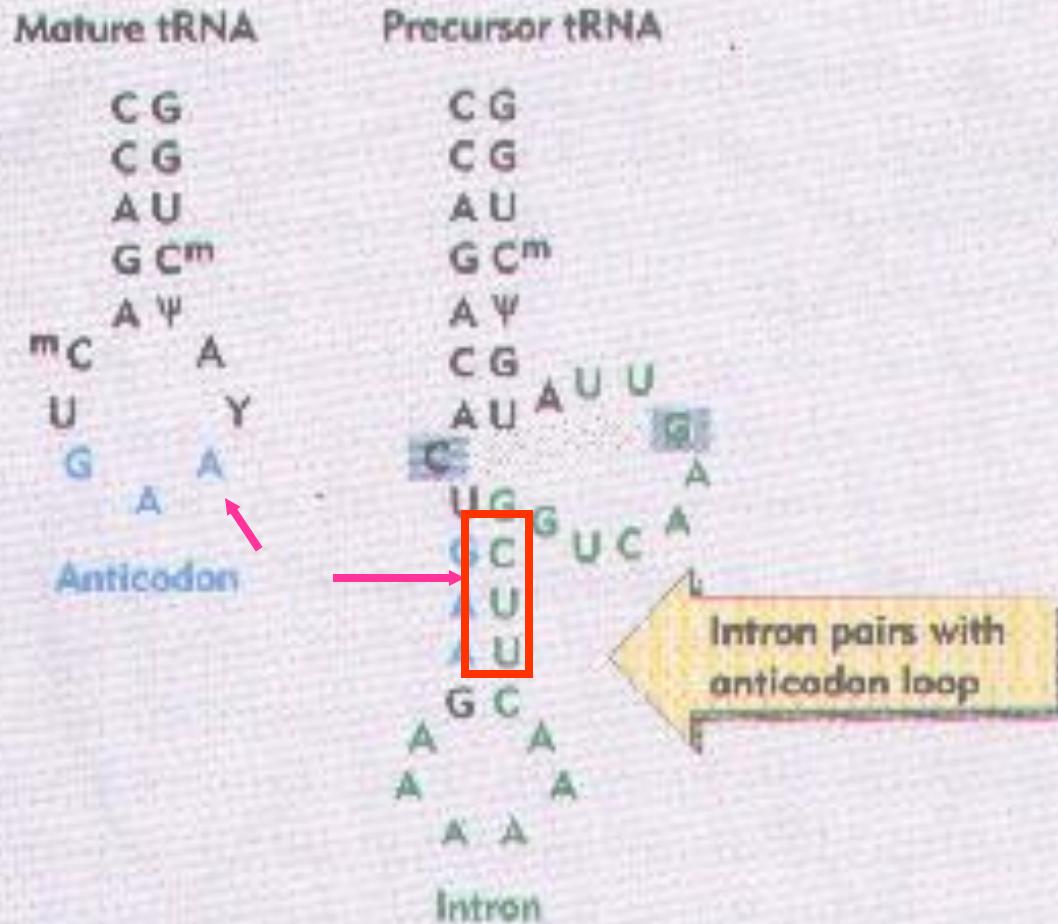
tDNA, rDNA

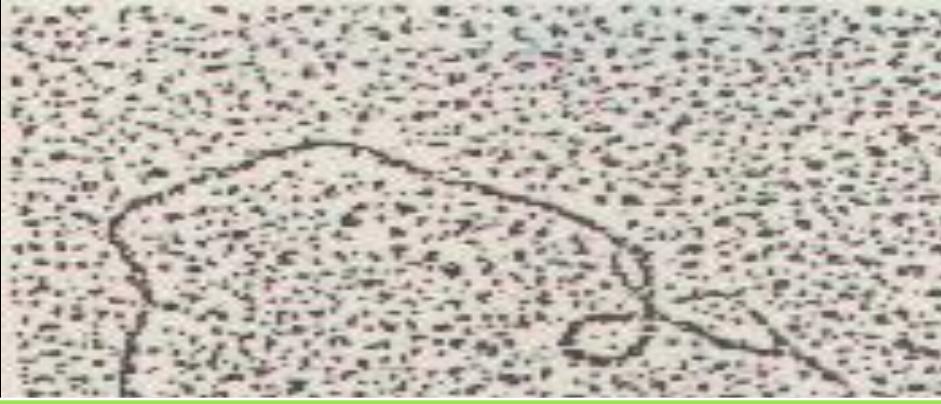
mtDNA, cpDNA

tDNA also is interrupted gene



The intron in yeast tRNA^{Phe} base pairs with the anticodon to change the structure of the anticodon arm. Pairing between an excluded base in the stem and the intron loop in the precursor may be required for splicing.

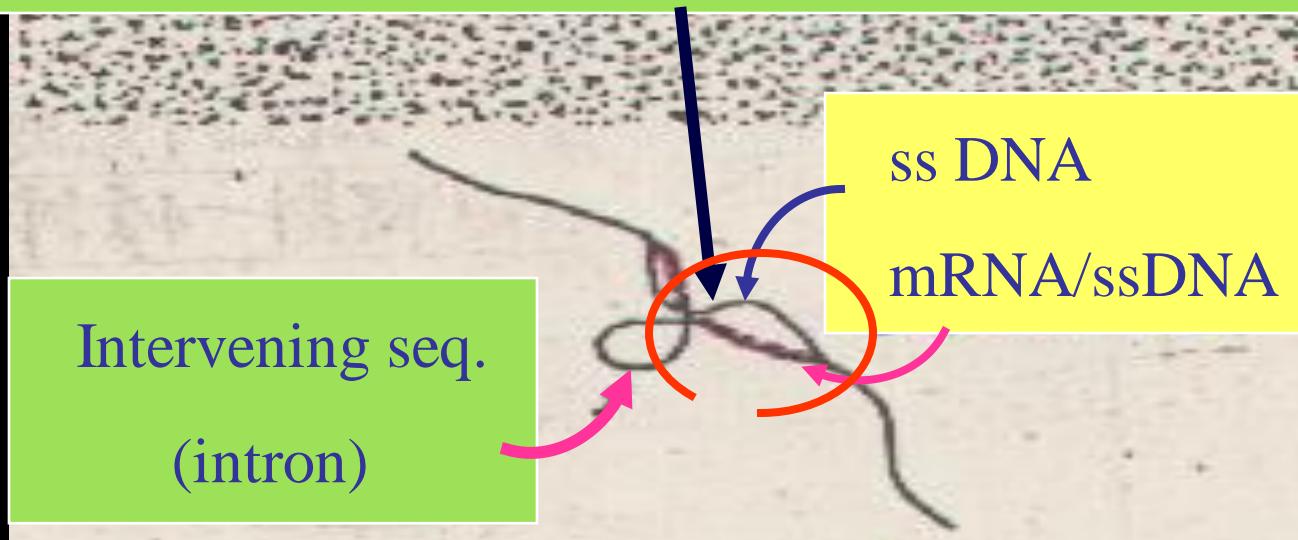




R-loop mapping of the mouse minor β -globin gene

mRNA \times D.S DNA

R-loop is the structure formed when a RNA strand hybridized with its complementary strand in a DNA duplex, thereby displacing the original strand of DNA in the form of a loop extending over the region of hybridization (Genes VI)

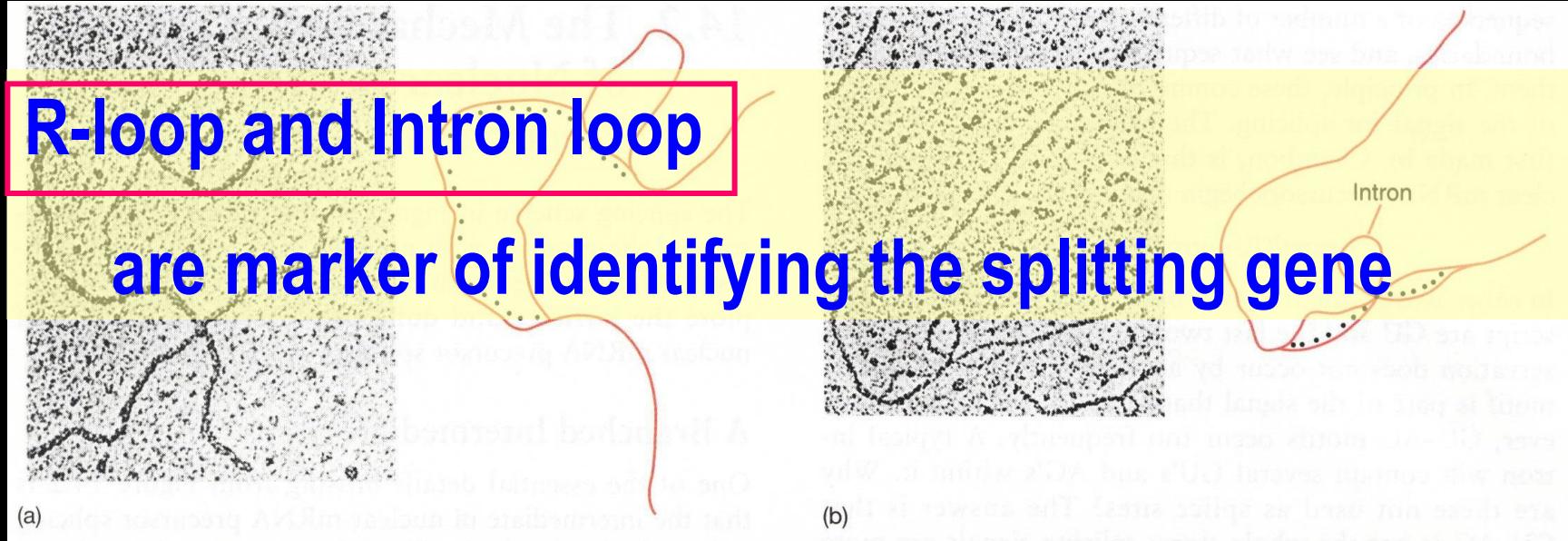


The hybrid duplex and the displaced stretch of single-strand DNA are called an **R-loop**

From Mol. Cell biology

Introns are transcribed

Introns are removed



(来源: 分子生物学 (2007) , 郑用琏, 第61页)

a; R-looping in which mouse globin pre-mRNA was hybridized to a clone of mouse β -globin gene . A smooth hybrid formed (introns are represented in pre-mRNA)

b; R-looping in which mature mouse globin mRNA was used to hybridized to a clone of mouse β -globin gene. A large intron in the gene looped out (the intron no longer present in the mRNA)

外显子序列保守

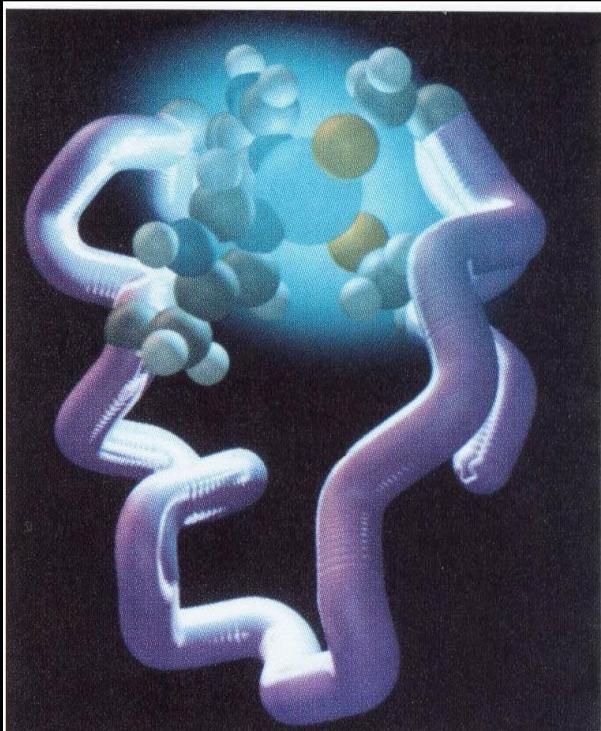
内含子序列多变

(但外显子与内含子基因发生突变的频率相同)

外显子序列承受选择压力!

2. 6. 基因多样性

的多样性-跳跃基因



跳跃基因
(Jumping gene /
Transposable element)

跳跃基因

(Jumping gene / Transposable element)



(来源：不详)

2.6.1. 基因转座现象的最初发现

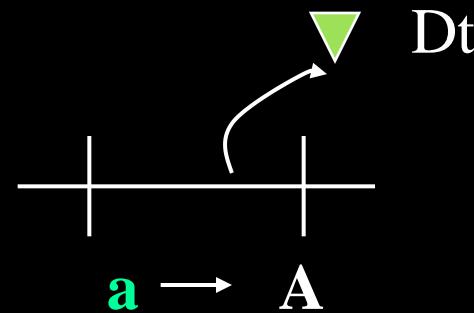
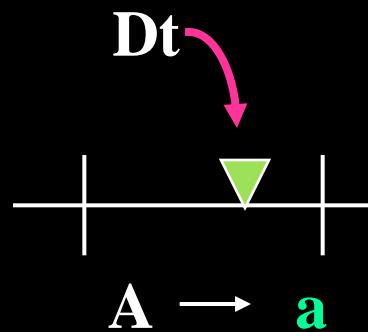
- 1914 A. Emerson Cornell university

玉米果皮花斑突变



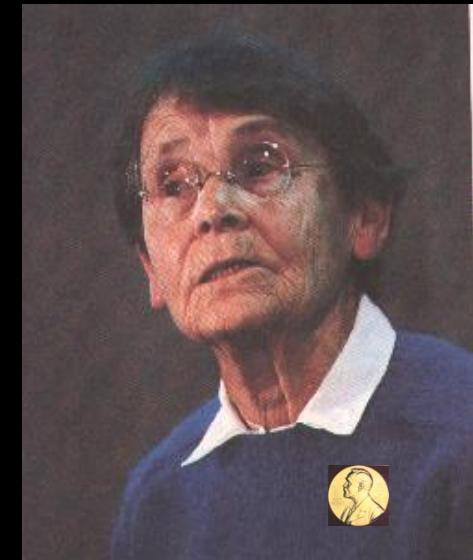
- 1936 Marcus . M. Rhoades Indiana University

玉米糊粉层斑点 (Dotted D_t) 突变



- 1947 Barbara McClintock CSH

玉米糊粉层花斑不稳定现象伴随的遗传事件



a) 系列染色体遗传重组事件

b) 在 **Ac** (Activator) 因子存在时

$C^I \rightarrow c^i \rightarrow$ spots in aleurone layer

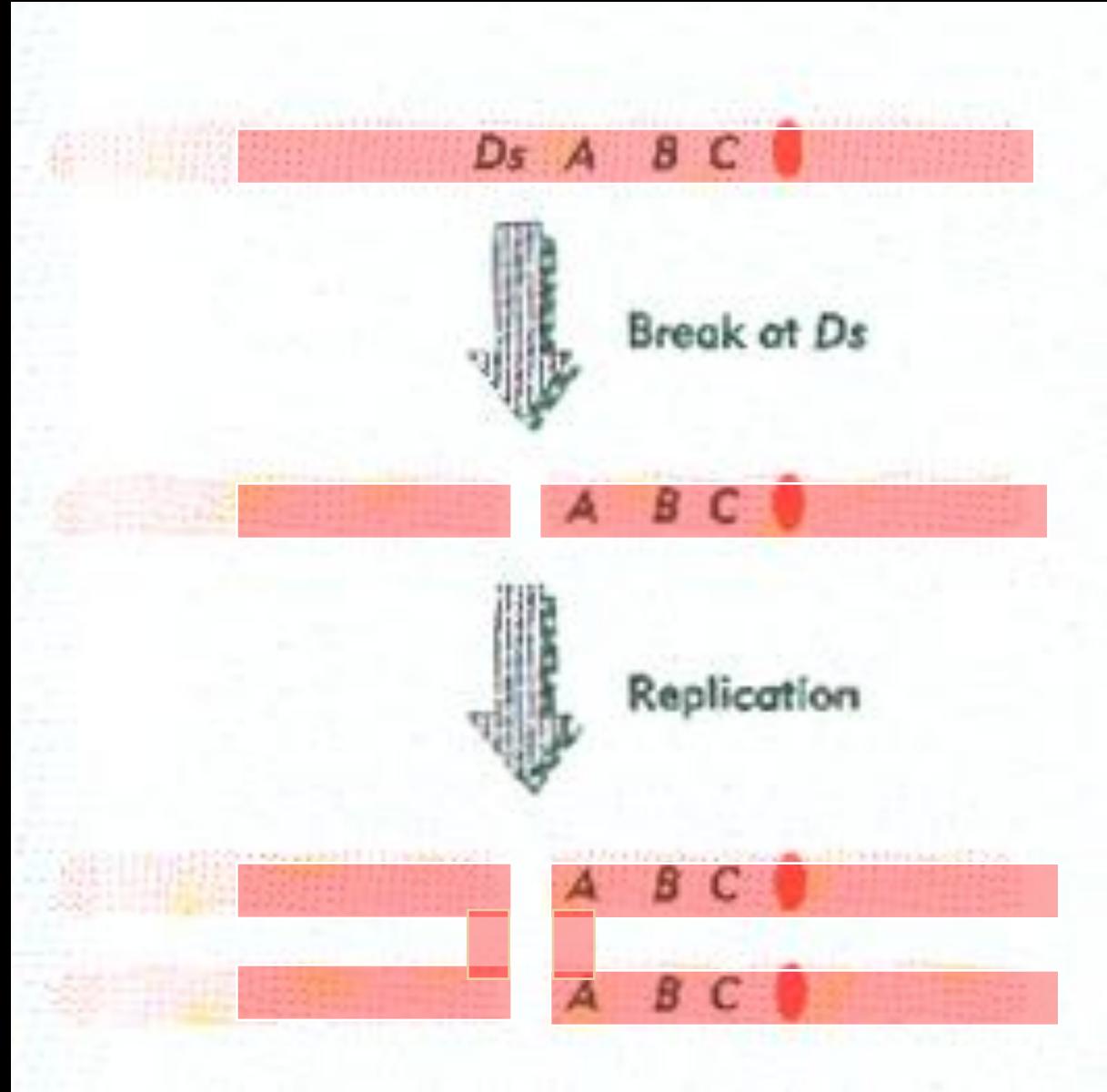
染色体 → BFBC

(Breakage – Fusion – Bridge – Cycle)

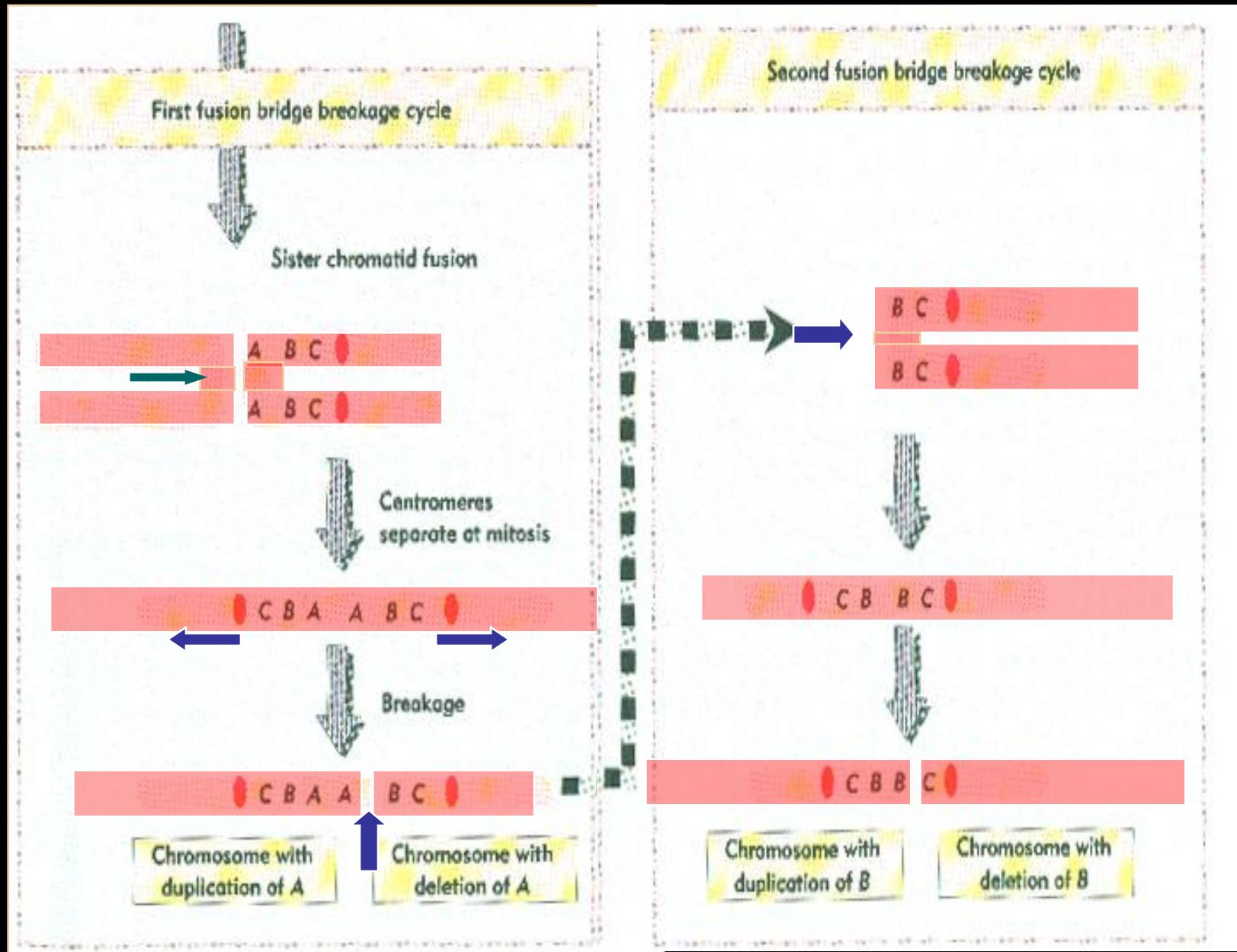


site named **Ds** (Dissociator)

BFBC 的形成

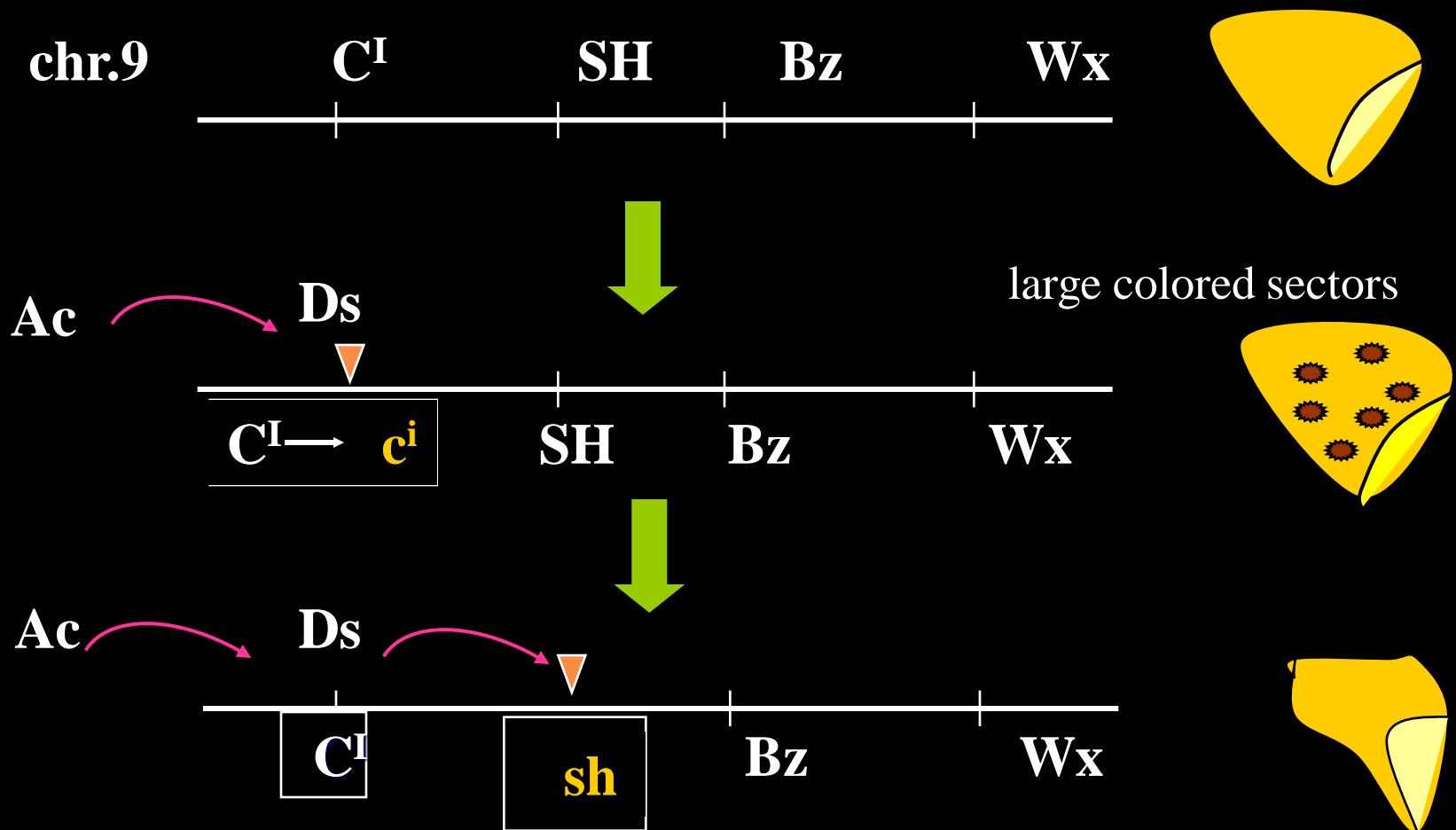


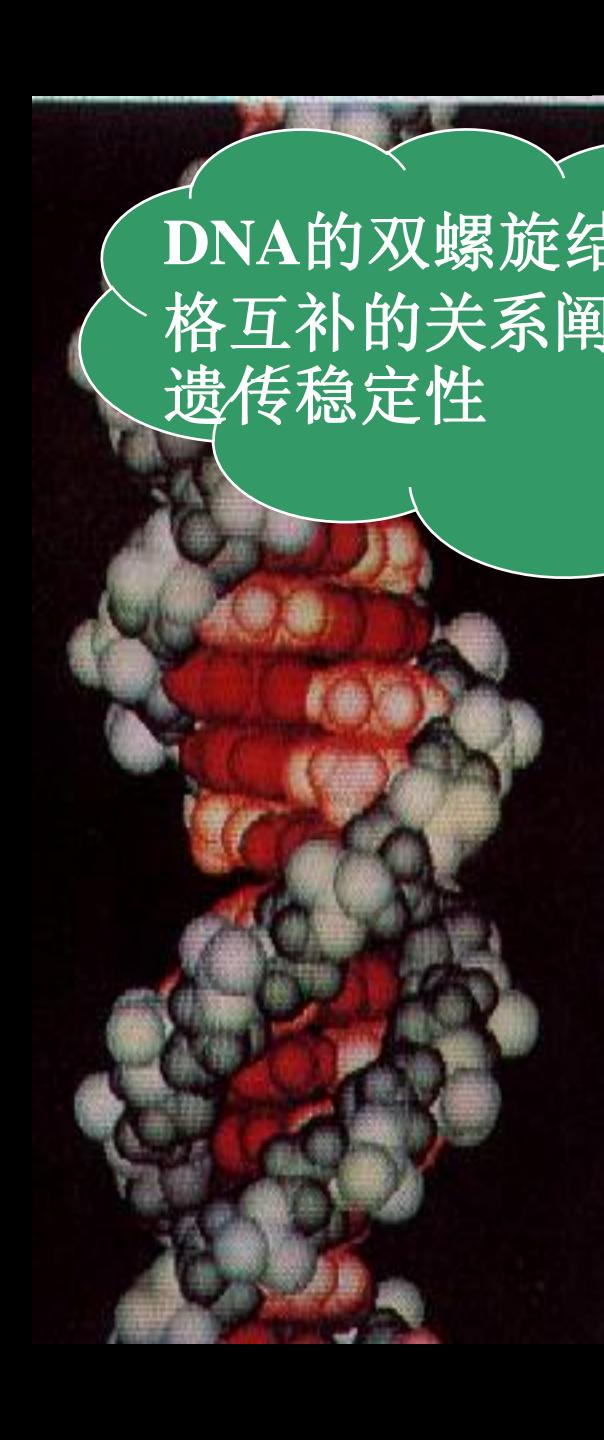
(来源: 分子生物学 (2007), 郑用琏, 第68页)



(来源: 分子生物学 (2007), 郑用琏, 第68页)

c) transposable element 是引起玉米糊粉层花斑不稳定
现象的遗传因子





DNA的双螺旋结构中双链严格互补的关系阐明了基因的遗传稳定性

-1953



2.6.2. 基因转座现象的再次发现与证实

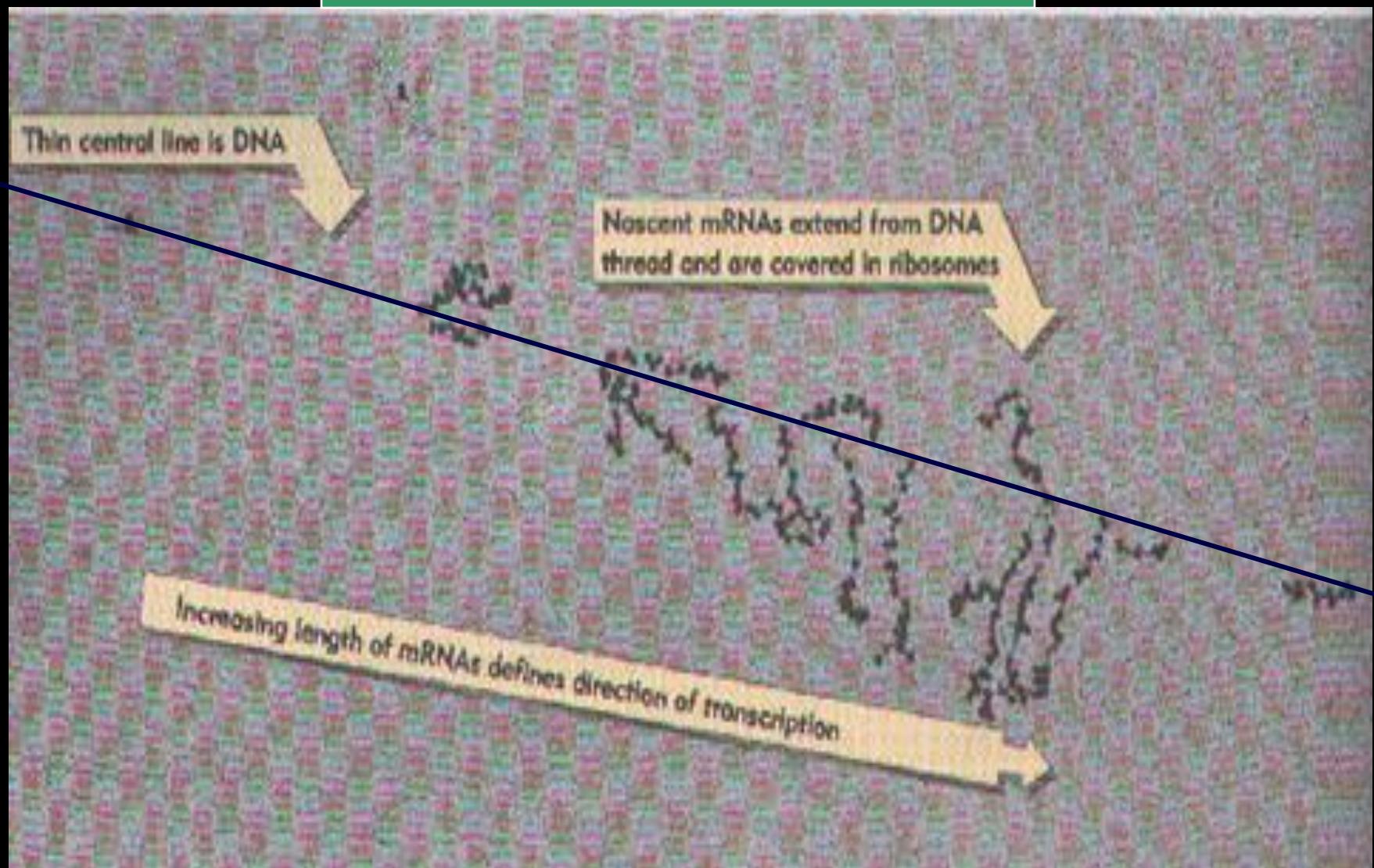
插入型极性突变的发现与机理

Operon & Polarity Mutation

极性突变的基本概念

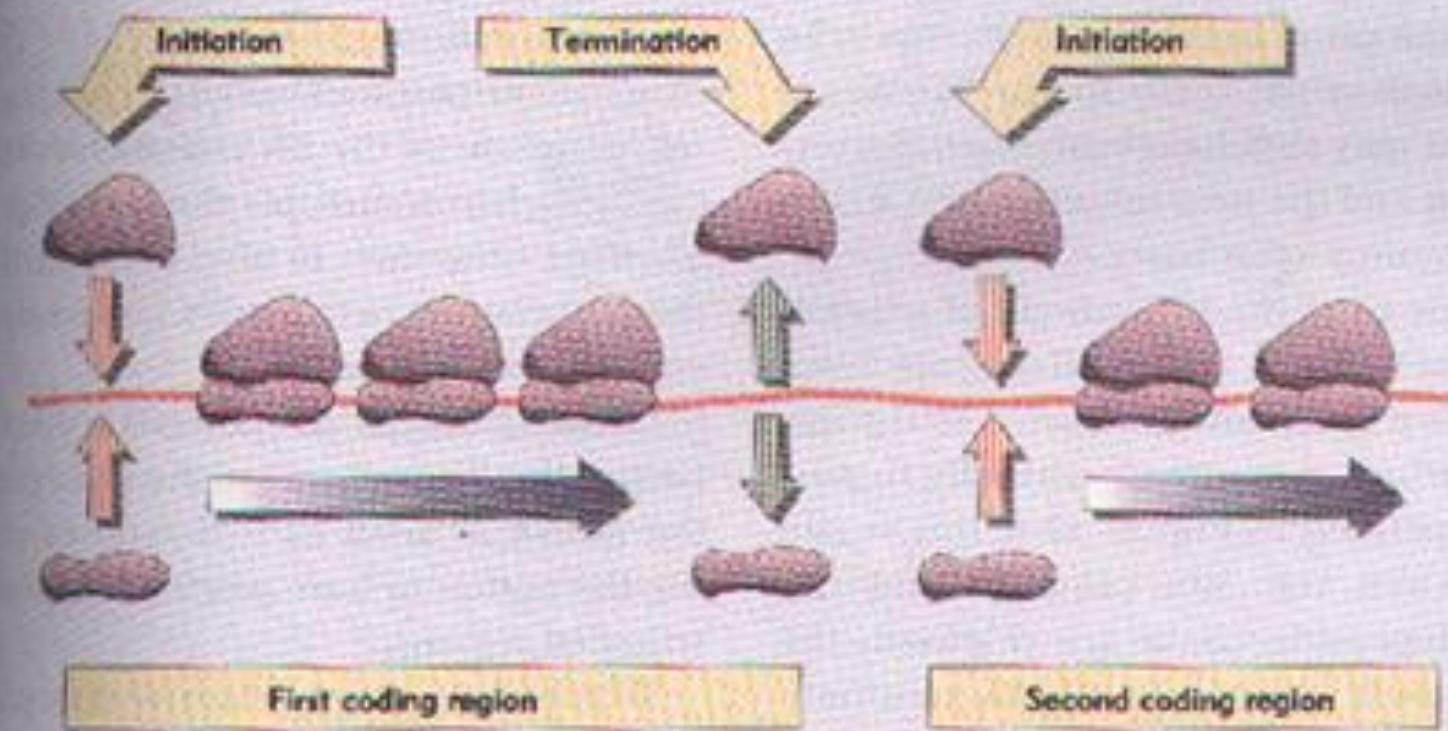
在一个操纵子中，与操纵基因毗连的结构基因发生终止突变后，它除了影响该基因本身产物的翻译外，还影响其后结构基因多肽的翻译，并且具有极性梯度的特征。

原核生物的多聚核糖体结构



(来源：不详)

操纵子结构中蛋白质合成过程的核糖体形成与解体



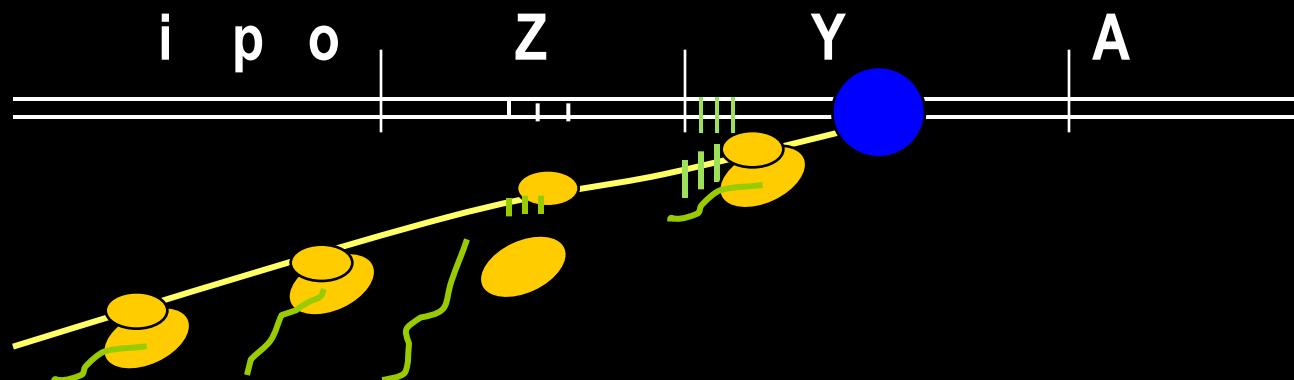
(来源: 分子生物学(2007), 郑用琏, 第68页)



A酶活性

O gene

z 基因终止突变位
点离 O 基因的距离



(来源：分子生物学（2007），郑用琏，第67页)

Polarity Mutation discovered in galactose operon

James. Shapiro 1966, Jodan E. 1968

E. coli galactose operon



Galactose 代谢

gal

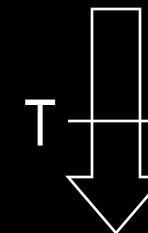
K → kinase

p-1-galactose

t mut. → U-transferase 缺乏

P-1-galactose 积累 → 半乳糖症

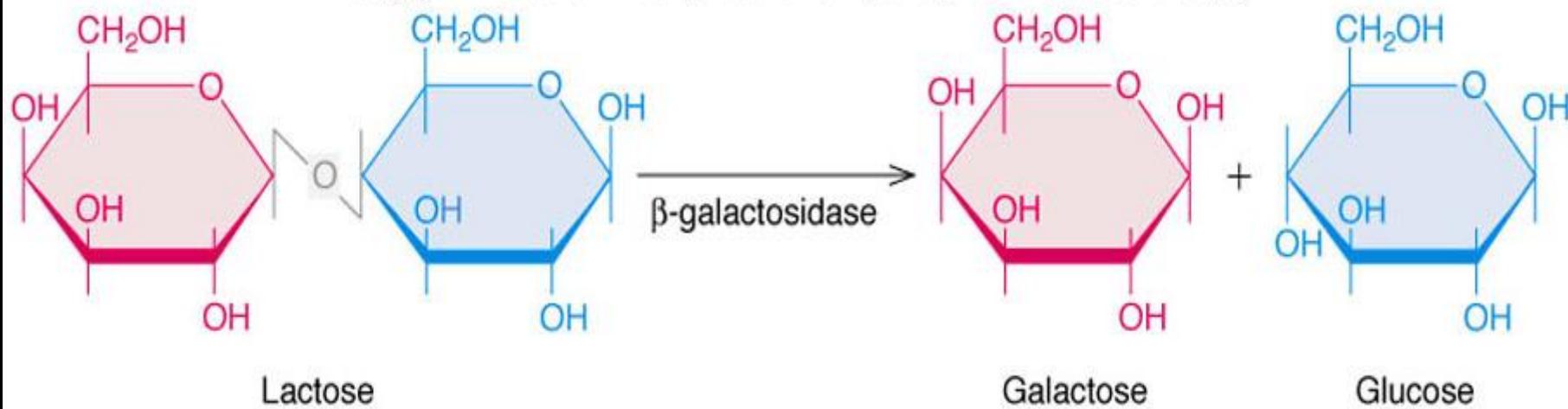
UDP-glu 肝大, 黄疸, 痴呆婴儿不能成长



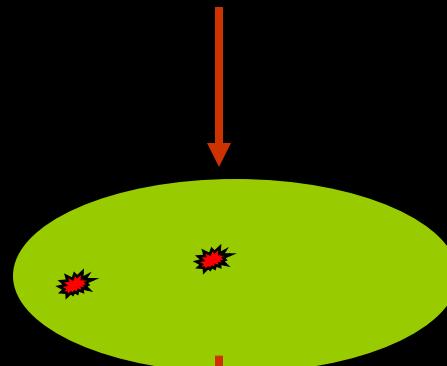
U - transferase

UDP- galactose

Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.



E. coli [t⁻]



lactose
medium

genotype of survival

E t⁻ K → E T K

E t⁻ K → E t⁻ k⁻

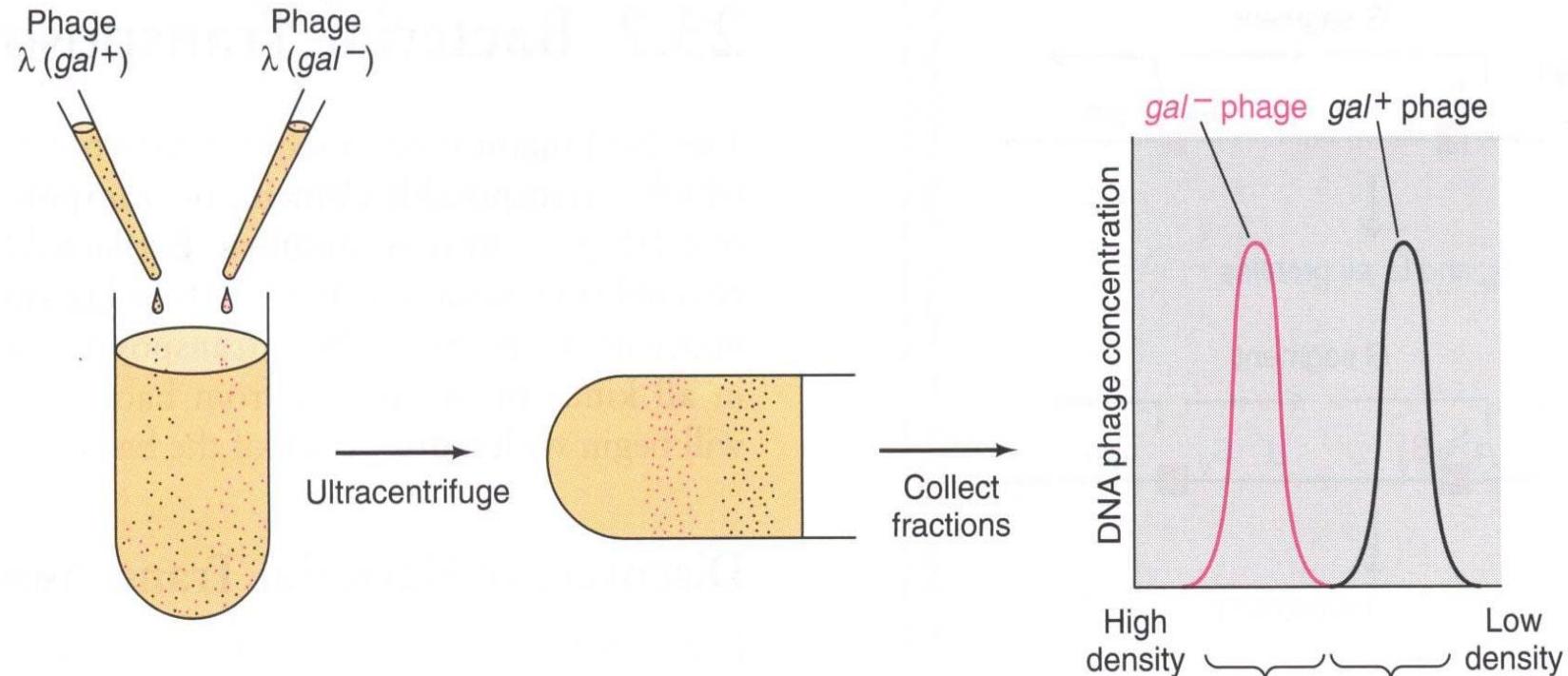
E t⁻ K → e⁻ t⁻ K (e⁻ polarity mutation)



转导实验证明：

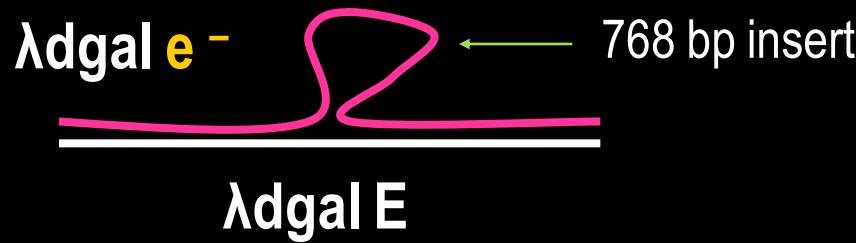
- $E \text{ t}^- K \xrightarrow{\times} E \text{ T} K$
- $E \text{ t}^- K \xrightarrow{\times} E \text{ t}^- k^-$
- e^- gene polarity mutation \longrightarrow Kinase 减少

- e^- 回复突变频率高 $\longrightarrow e^-$ 不是缺失突变
- 诱变剂不能提高 e^- 突变频率 $\longrightarrow e^-$ 不是终止突变
- λ dgal e^- 和 λ dgal E 分子 CsCl 梯度离心



λ dgale⁻片段有所增长

| λ dgal e⁻ / λ dgal E 分子杂交



(来源: 分子生物学 (2007), 郑用琏, 第72页)

表明外来DNA片段的 转座可引起
插入型极性突变

2.6.3. DNA转座现象的一般遗传特点

- a) 不依赖 **Donor site** 与 **Target site** 间序列的同源性
(非同源重组过程，不依赖 **recA** 酶)
- b) 转座插入的靶位点并非完全随机 (插入专一型)
Hotspots (热点)
Regional preference (在3kb区域内的随机插入)
- c) 某些转座因子 (Tn3) 对同类转座因子的插入具有排他性 (免疫性)
- d) 靶序列在转座因子两侧会形成正向重复(**DR**)
- e) 转座因子的切除与转座将产生复杂的遗传学效应

f) 与倒位和易位染色体结构变异现象的区别

Transposition

特定的基因片段

U.V 或诱变剂无效

高频率的回复突变

形成 **mutable gene**

转座需 **transposase**

inverted repeat (IR)

与复制有关的过程

Inversion & Translocation

断点随机的任意DNA片段

可能破坏某结构基因

U.V 或 诱变剂可提高频率

不能产生回复突变

Rec.A, B, D

X(chi) site,

DNA链的断裂与错接的过程

2.6.4 转座因子的种类及机制

- 原核生物的转座因子的种类

IS

TnA family

Composite transposon

Transposable phage

复制型转座子

(replicating transposable element)

- 真核生物的转座因子的种类

剪贴式转座因子（cut - paste ）

反转录转座子（Retro-transposon ）

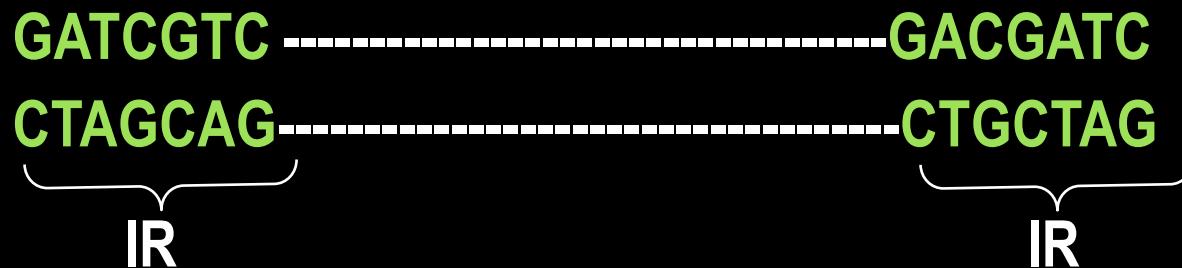
类似反转录病毒

2. 6. 4. 1

原核生物转座因子
的种类及转座机制

a) 插入序列 (insertion sequence IS)

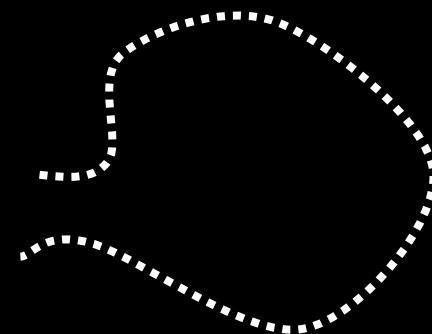
| 700 ————— 2000 base pair (bp)



● **GATCGTC** ————— **GACGATC**

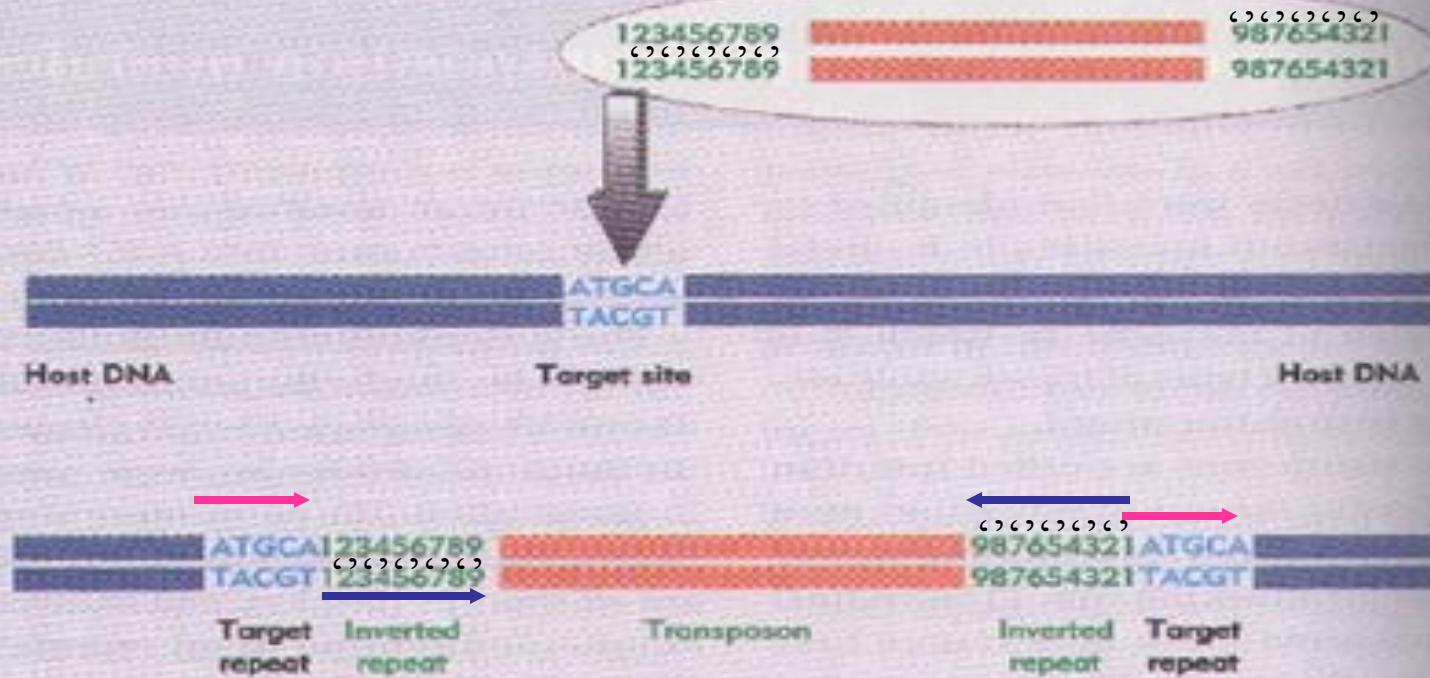
Stem-loop

5' **GATCGTC**
3' **CTAGCAG**



(来源: 分子生物学 (2007), 郑用琏, 第73页)

- IS插入target site后 → target sequence repeats directly flanked the IS



(来源: 分子生物学 (2007), 郑用琏, 第73页)

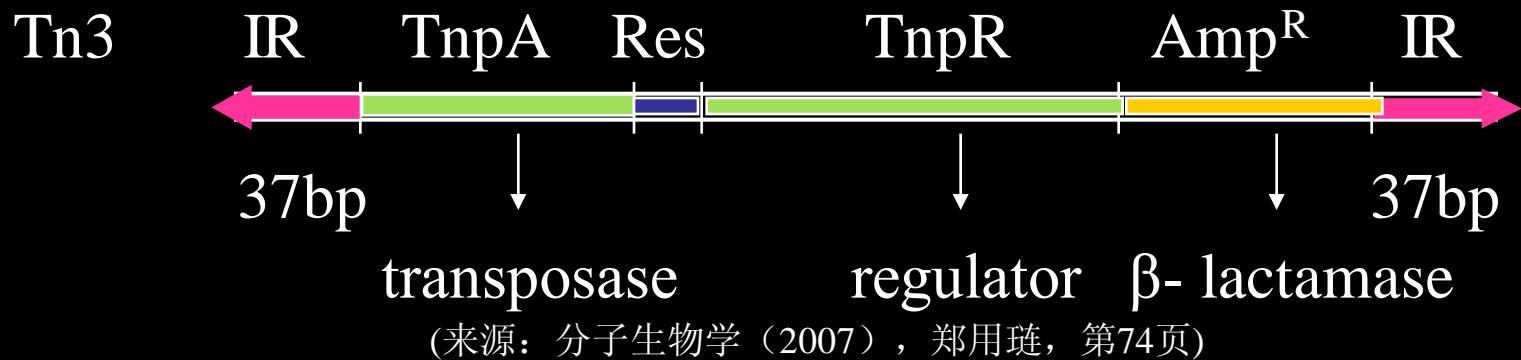
Individual elements comprise transposable units or modules of composite transposons

| Element | Length | IR | DR | Target Selection |
|---------|---------|-------|-------------|------------------------|
| IS1 | 768 bp | 23 bp | 9 bp | Random |
| IS2 | 1327 bp | 41 bp | 5 bp | Hotspots |
| IS4 | 1428 bp | 18 bp | 11 or 12 bp | AAAN ₂₀ TTT |
| IS5 | 1195 bp | 16 bp | 4 bp | Hotspots |
| IS10R | 1329 bp | 22 bp | 9 bp | NGCTNAGCN |
| IS50R | 1531 bp | 9 bp | 9 bp | Hotspots |
| IS903 | 1057 bp | 18 bp | 9 bp | Not known |

(来源：不详)

b) Transposon (Tn / TnpA family)

- 2.5 kb → 20 kb
- 具有IR、转座酶基因、调节基因、抗生素抗性基因

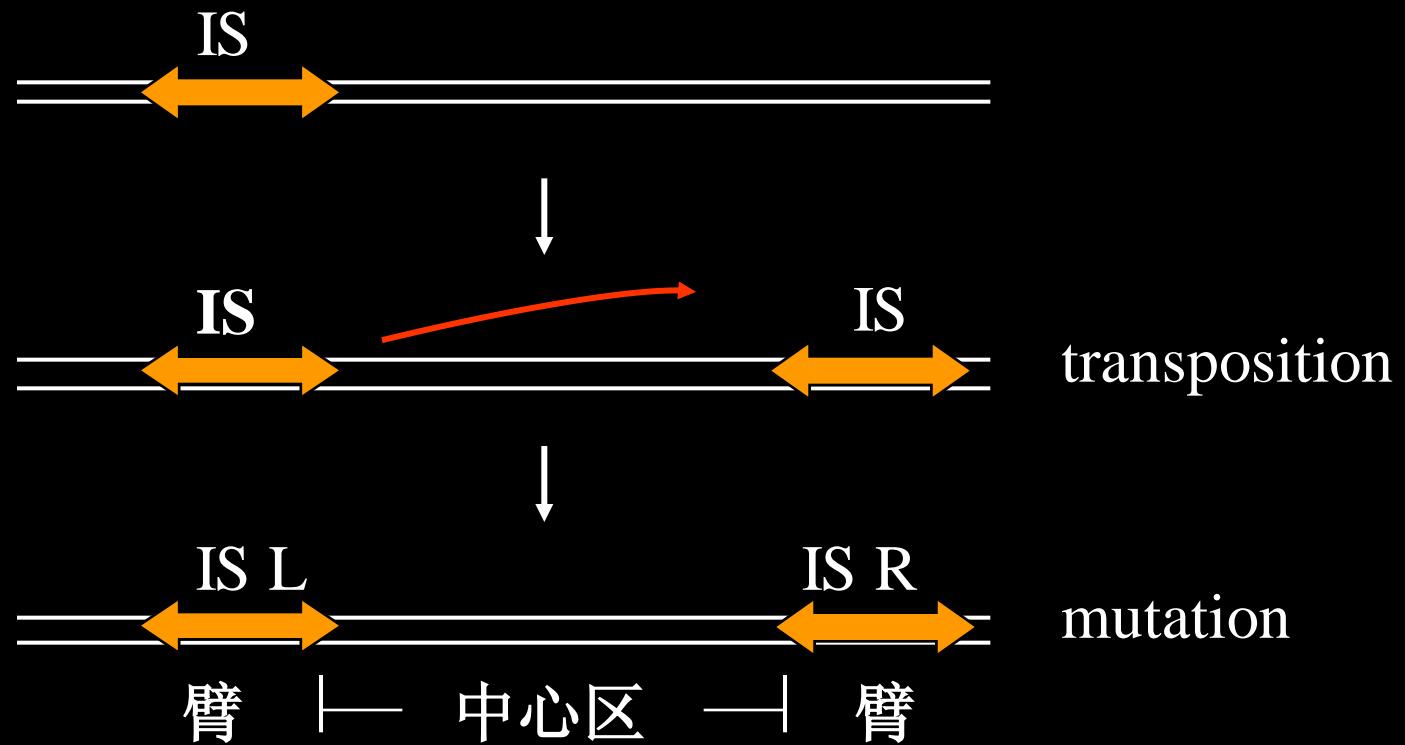


- Tn1 (Amp^R) Tn2 (Amp^R)
Tn3 (Amp^R) Tn4 (Amp^R Str^R)
Tn5 (Kan^R) Tn6 (kan^R)
Tn7 (Str^R Tmp^R) Tn9 (Cam^R)
Tn10 (Tet^R)

C) 复合因子

(composite element / composite transposon)

- IS插入到功能基因两端，可能形成复合转座因子



(来源: 分子生物学 (2007), 郑用琏, 第75页)

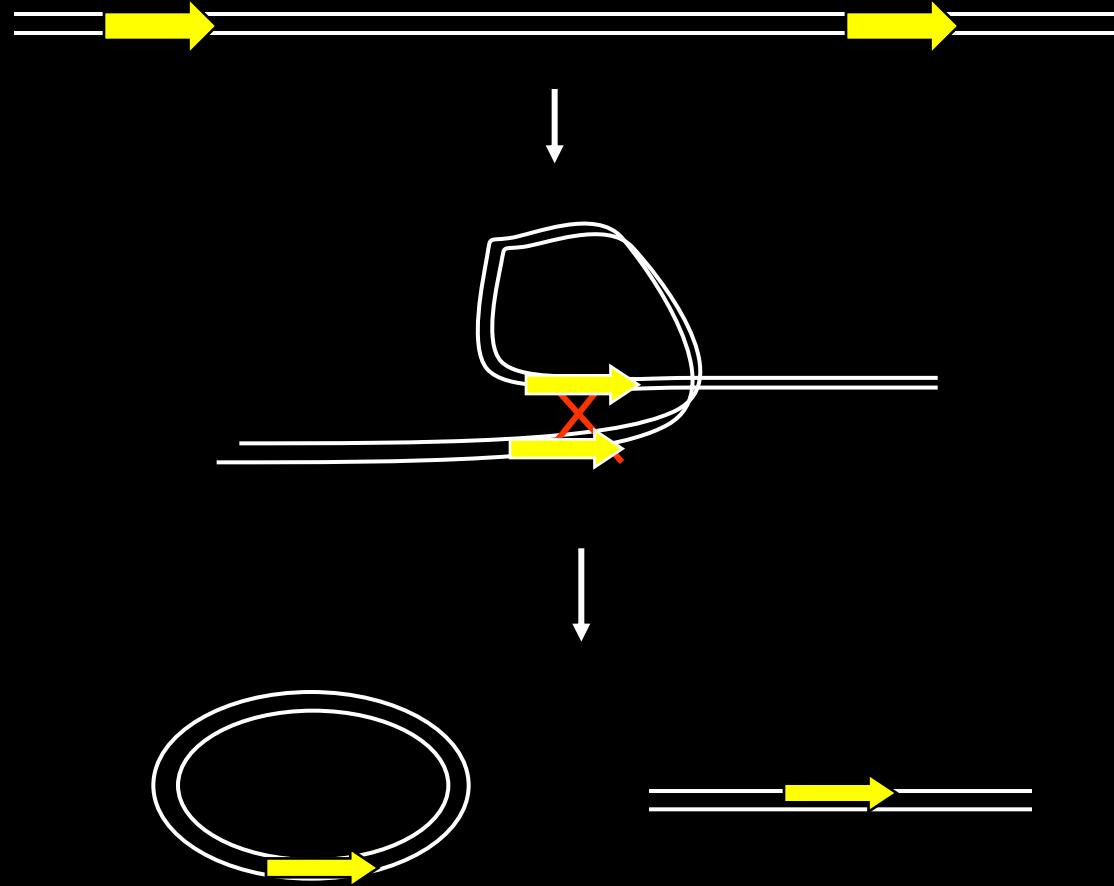
- 复合转座子一旦形成，转座功能受较多因子调控
- 结构特点



(来源：分子生物学（2007），郑用琏，第75页)

- 稳定性

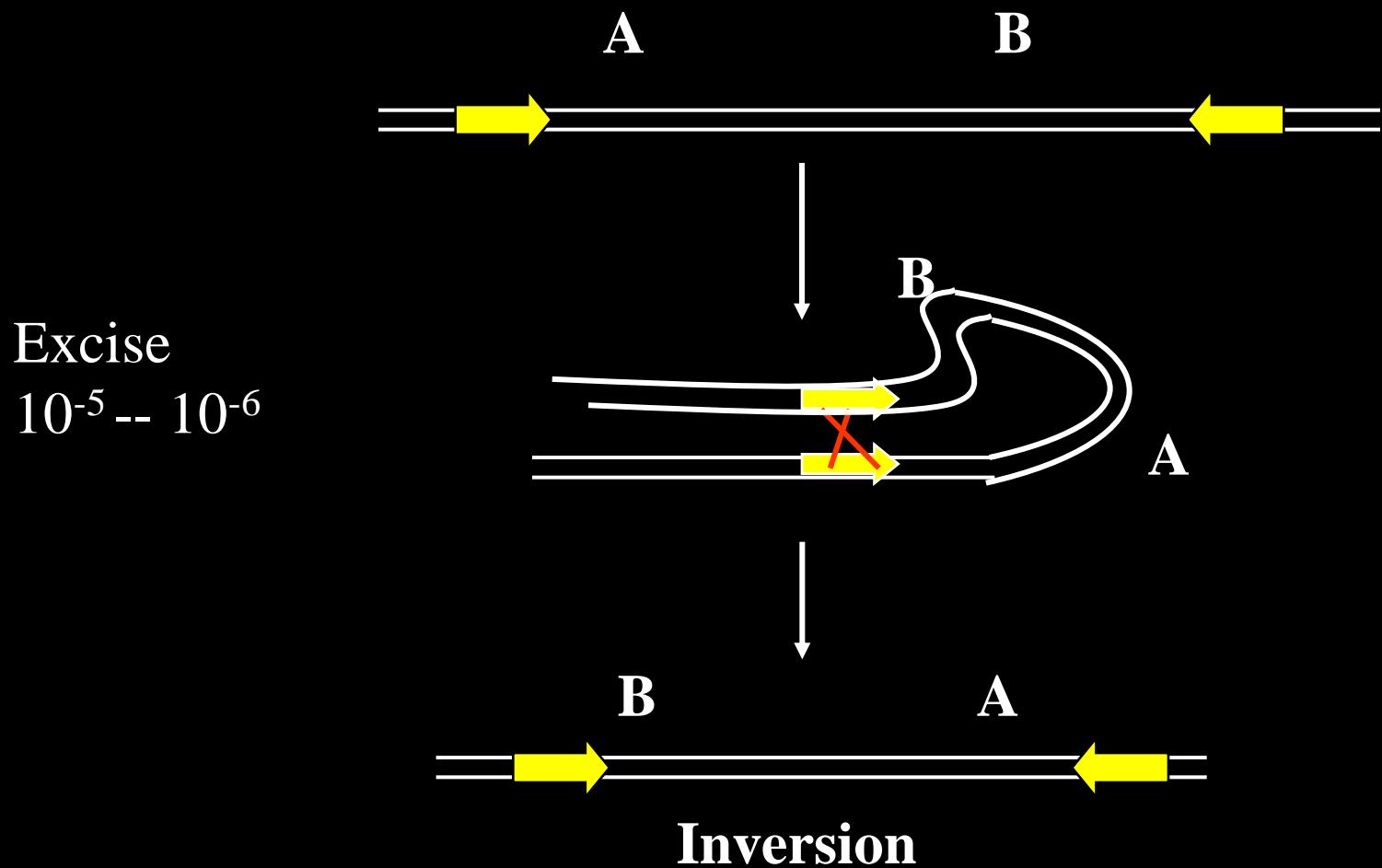
DR型 (Instable)



Excise or
Deletion

(来源: 分子生物学 (2007), 郑用琏, 第76页)

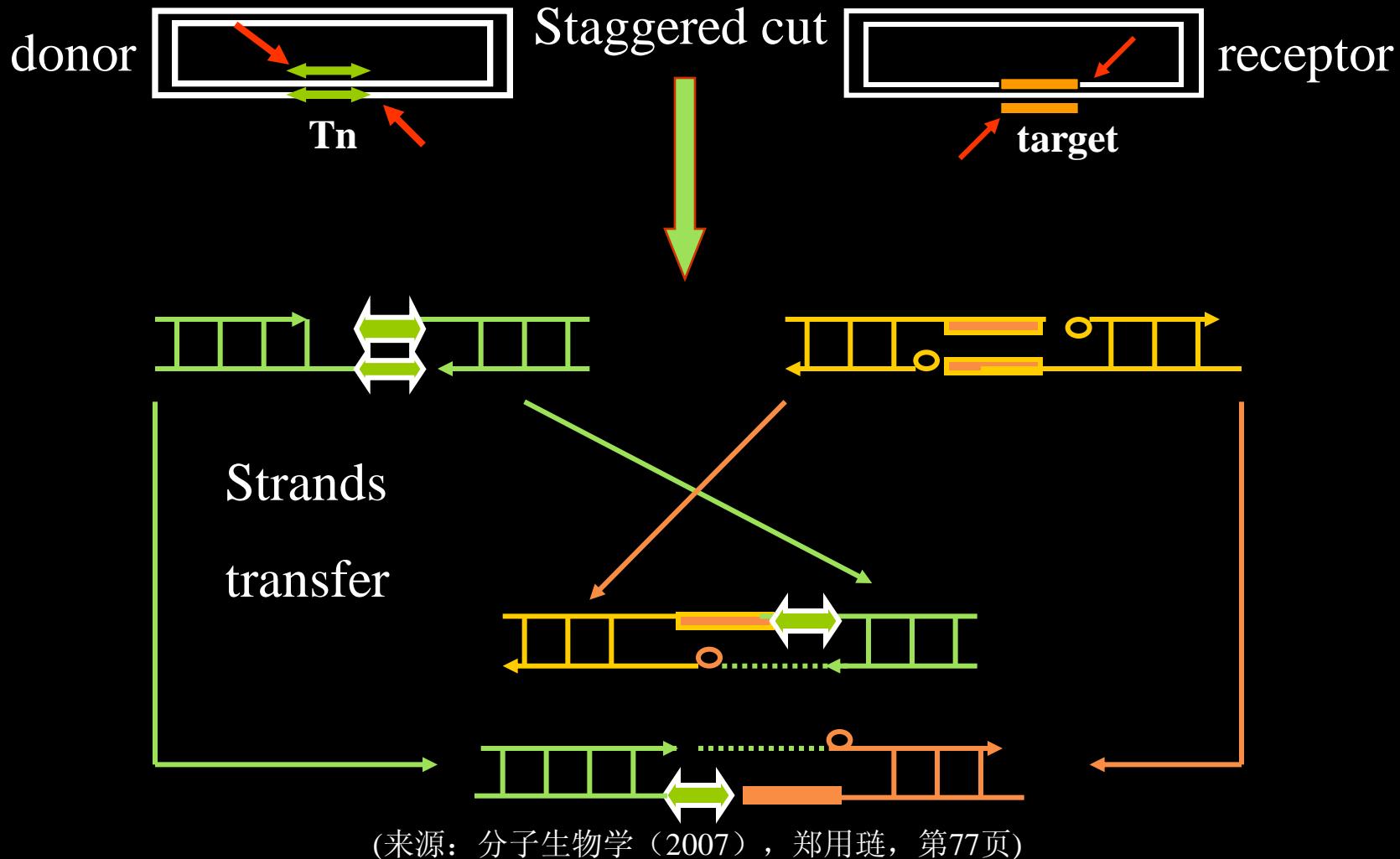
IR型 (Stable)



(来源: 分子生物学 (2007), 郑用琏, 第76页)

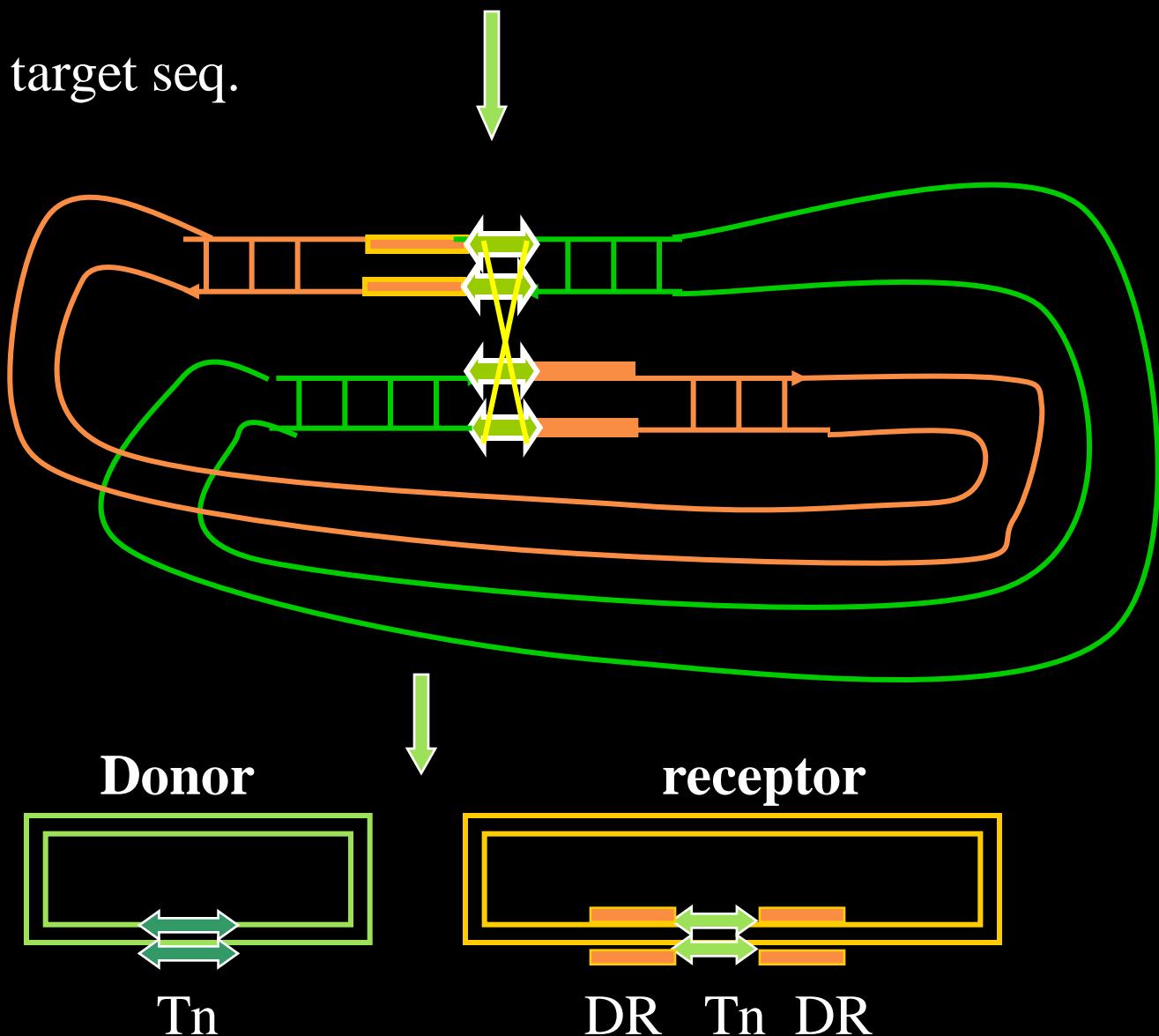
转座机制

replicating transposition in prokaryote



copying of Tn & target seq.

cointegrator
&
resolution



Conclusion

- a) 转座过程是由**Donor** 提供Tn copy到 target site， 涉及酶切、复制、重组的遗传学过程。
- b) 转座完成后，在Tn的两端出现target site序列的正向重复，其长度取决于**staggered cutting**的长度。
- c) 转座因子的 **IR** 序列是转座酶的重要识别位点，与转座，切除有关

- d) Cointegrator 是转座过程的中间体(具有两个 Tn 和两个 replicon), 其稳定性依 Tn 不同而异, 或 resolution 完成转座过程.
- e) Cointegrator 可能导致 Tn 和抗性的积累。

The evolution of plasmids with multiple-resistance to antibiotic is made easy by the ability of transposons to recombine without homology and thus to gather together in a single plasmid



2.6.4.2 真核生物的转座因子及转座机制



I型： 剪贴式转座因子（cut - paste ）

Two components system

Ac/Ds, Spm/dspm, Dt Mu in maize

IR, transposase, needed

II型： 反转录转座子（Retro-transposon 类似反转录病毒）

Copia in drosophila

Ty1 in yeast

Alu family

Two components system of type I

- Autonomous controlling element & Non-autonomous controlling element
- Transposition with chromosome replication
- Lead to insertion mutation ($a :: Tn$) & back mutation, but cause BFBC under **Ac/Ds** only
- Autonomous element create **autonomous mutable gene**

e.g a^{-m} , bz^{-m}

non-autonomous element create **non-autonomous mutable gene** under autonomous element activation only

- Autonomous element ----- Trans-action factor

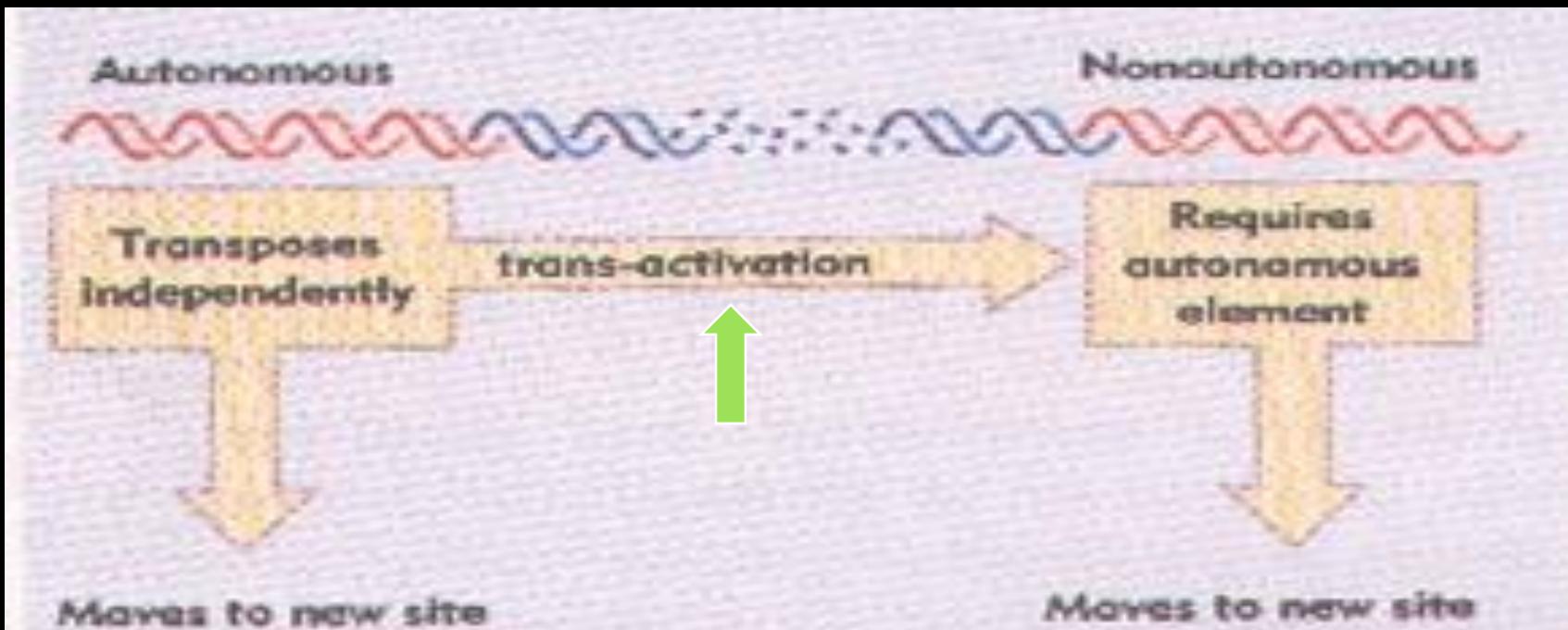


Regulator
Signal

Deleted mutant from Auto. Ele.

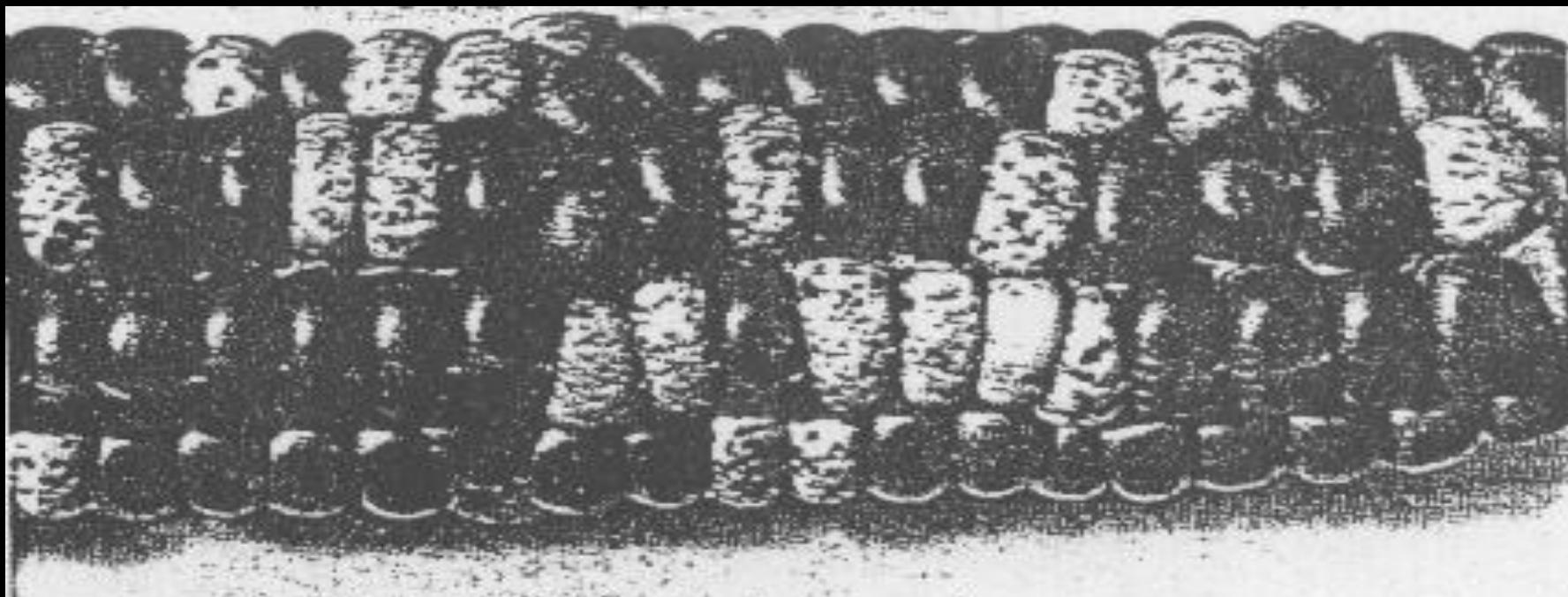
Non-autonomous element ----- Receptor

- Identifying the two components system by interaction between autonomous & non-autonomous elements



Identifying the two components system

| | |
|---------------------------------|-----------------------------|
| <u>Ac (activator)</u> | <u>Ds (dissociation)</u> |
| <u>Mp (modulator)</u> | |
| <u>Spm (suppressor-mutator)</u> | <u>dSpm (defective Spm)</u> |
| <u>En (enhancer)</u> | <u>I (inhibitor)</u> |
| <u>Dotted</u> | <u>Unnamed</u> |
| <u>Mu (mutator)</u> | <u>Not known</u> |

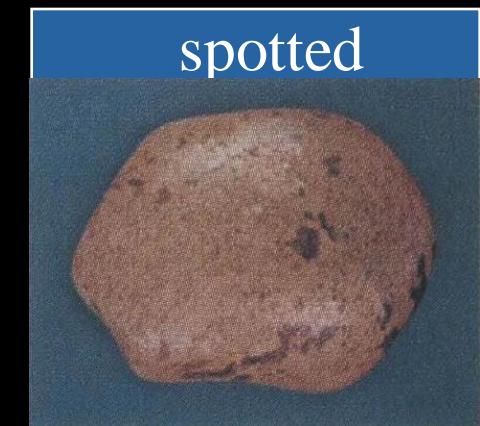
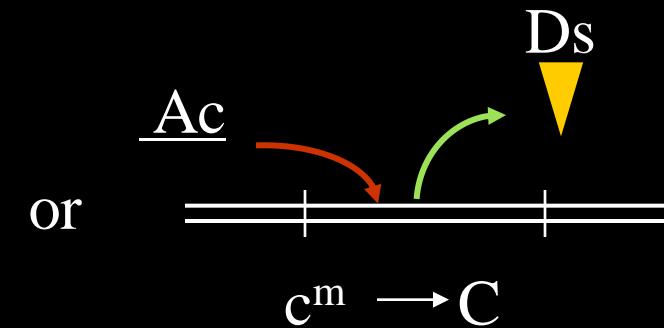
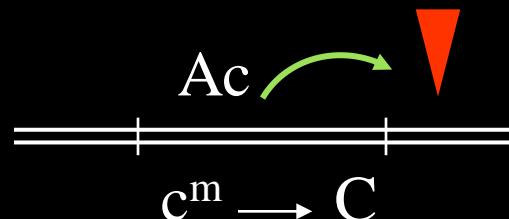
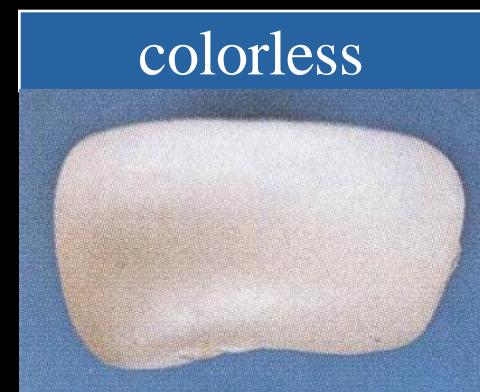
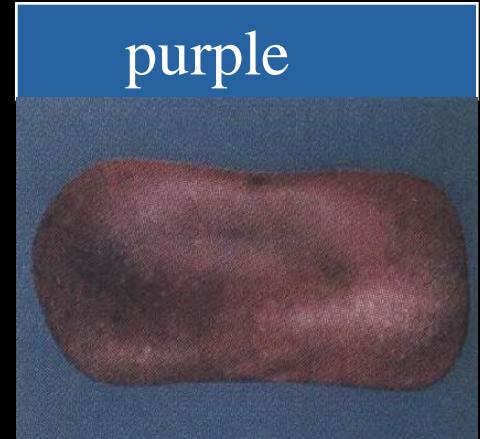
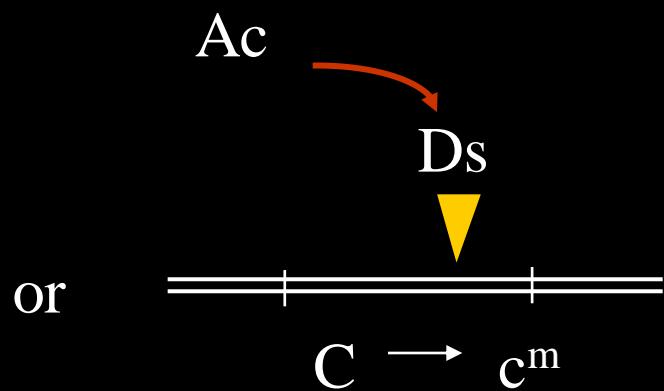
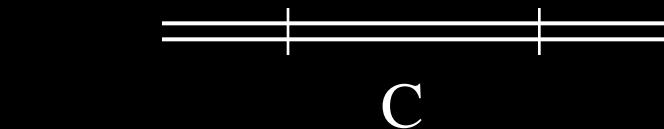
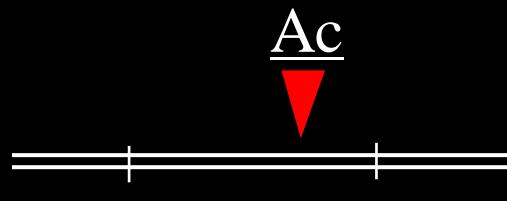


| Element | Copies /Genome | Length | mRNA | Protein Products | Inverted Repeats | Target Repeats |
|------------|----------------|---------|---------|------------------|------------------|----------------|
| ✓ Ac | ~9 | 4563 bp | 3500 bp | 1 | 11 bp | 8 bp |
| Mu1 | >10 | 1367 bp | ? | 4? | 213 bp | 9 bp |
| ✓ Spm (±n) | ? | 8287 bp | 2400 bp | 1 (+2?) | 13 bp | 3 bp |

(来源：分子生物学（2007），郑用琏，第80页)

b) Ac/Ds system

- 作用方式



(Source: F.W.Goro, from Fedoroff,N., Transposable genetic elements in maize. Scientific American 261(June 1984))



Bz

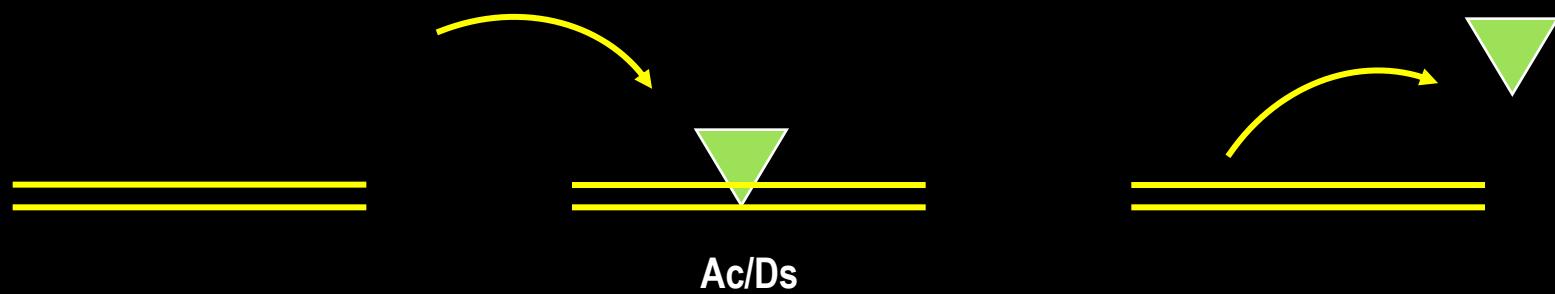


bz



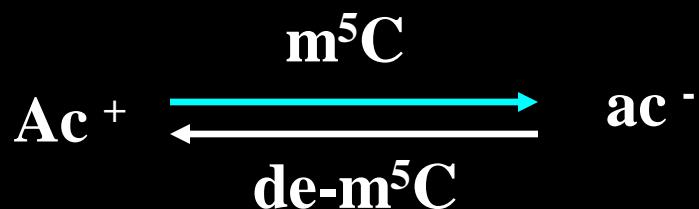
bz-m

(来源: 分子生物学 (2007) , 郑用琏, 第81页)



- 遗传特点

Ac 具有活化周期效应



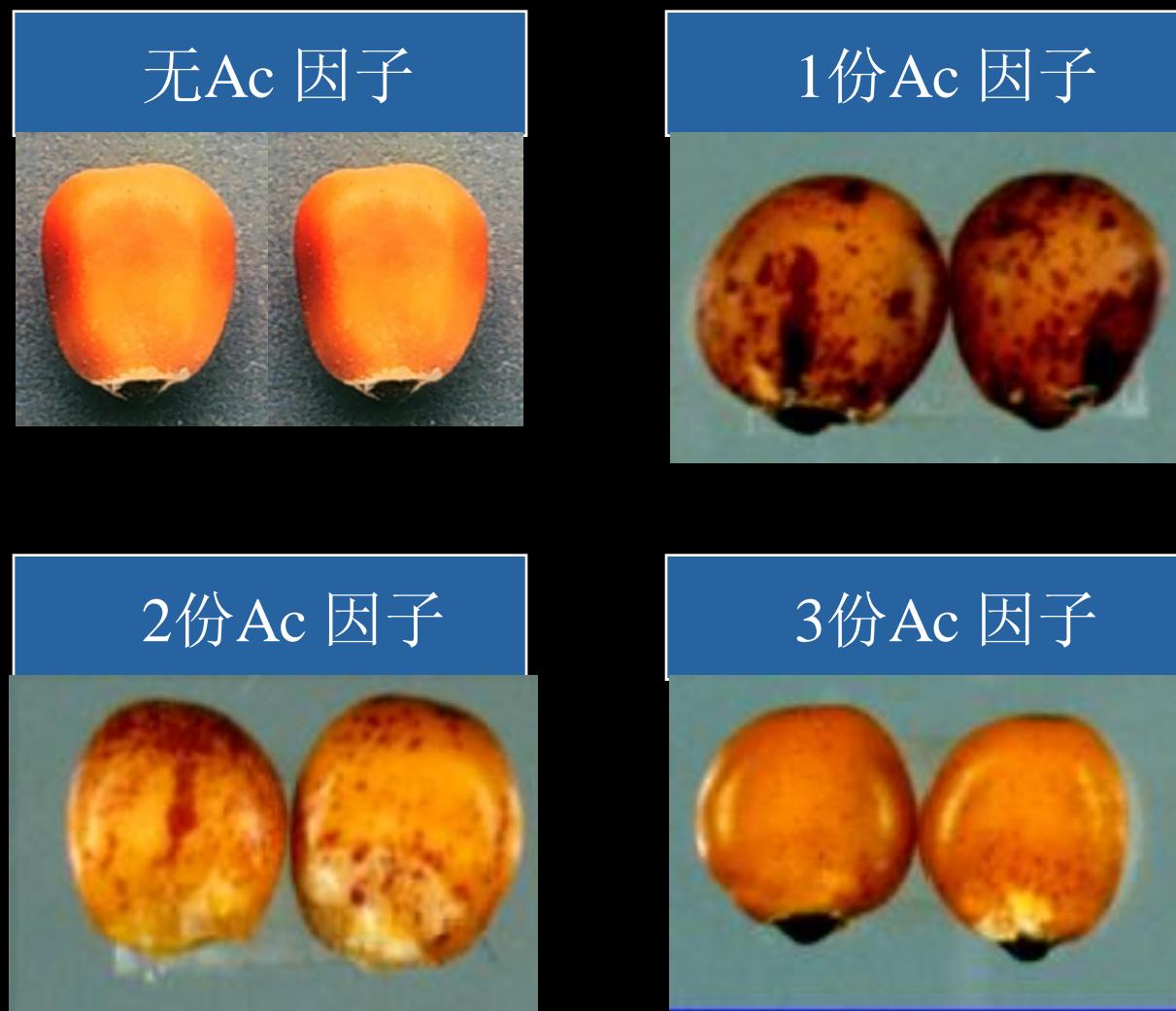
Ac与Ds连锁遗传但更多表现独立遗传

Ac  Ds 

或

Ac  _____
Ds 

Ac
对 Ds 的控制
具有负剂量效应



(来源：分子生物学（2007），郑用琏，第81页)

Ac/Ds引发的突变类型及表型

插入钝化，活性改变



表达水平改变



缺失突变

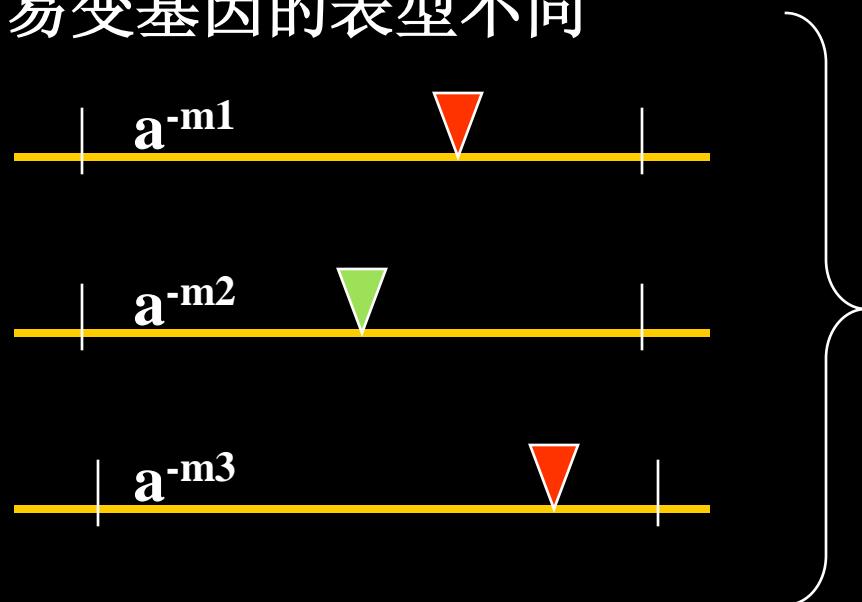


Ds的结构不同 (Ds1, Ds2...)

插入同一靶基因的位点不同



易变基因的表型不同

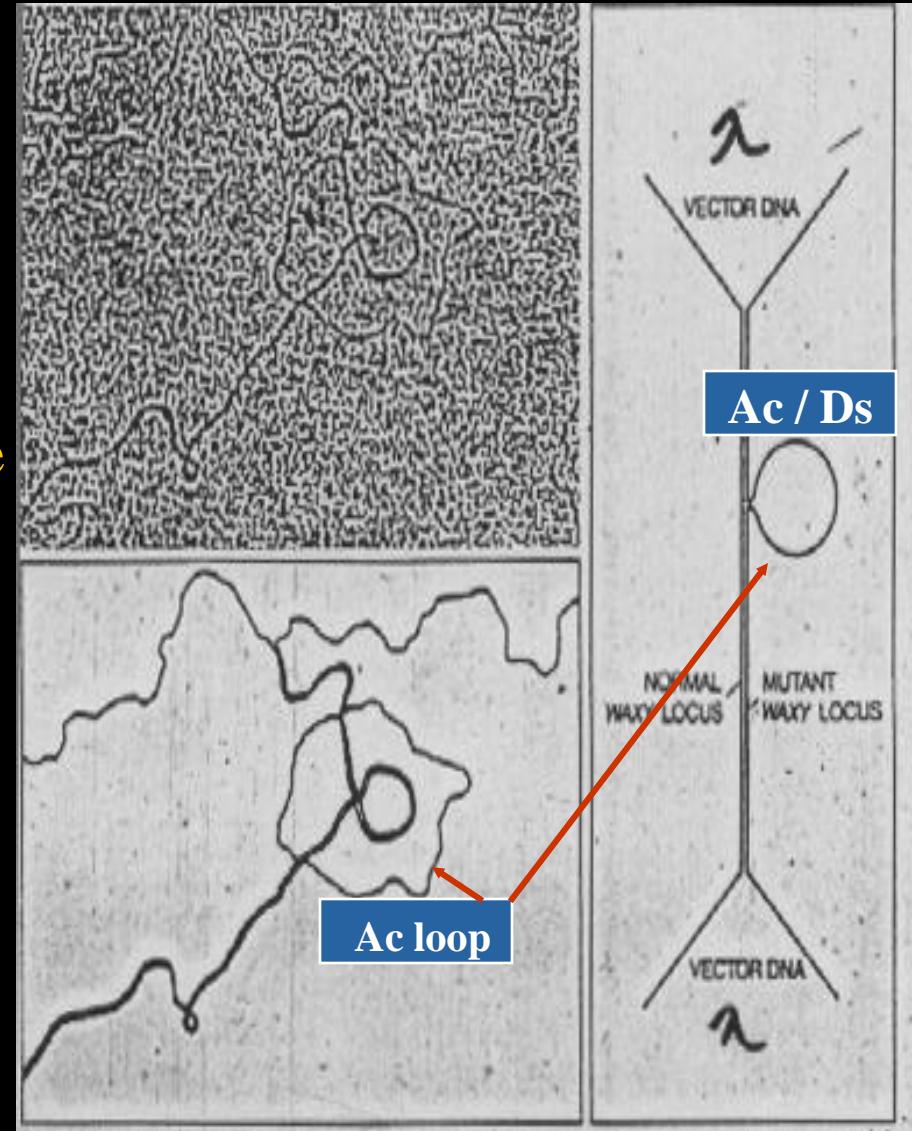
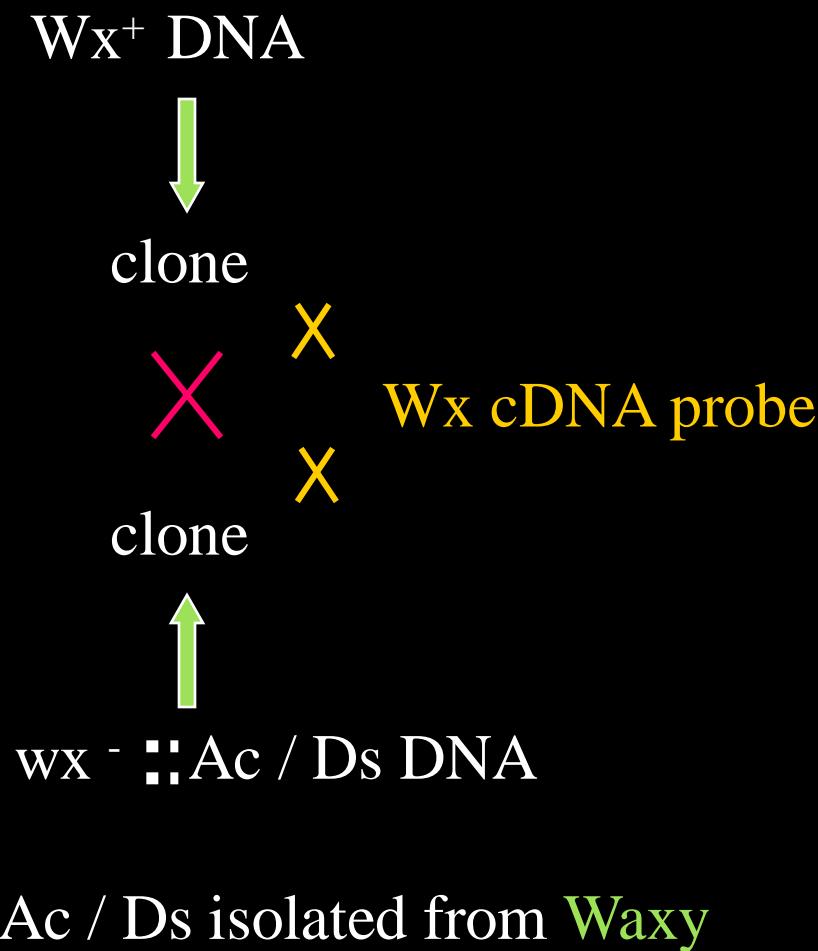


- 对靶基因的控制效应不同
- 引起回复突变的时间不同
(斑点大小差异)
- 引起回复突变的频率不同
(斑点多少差异)

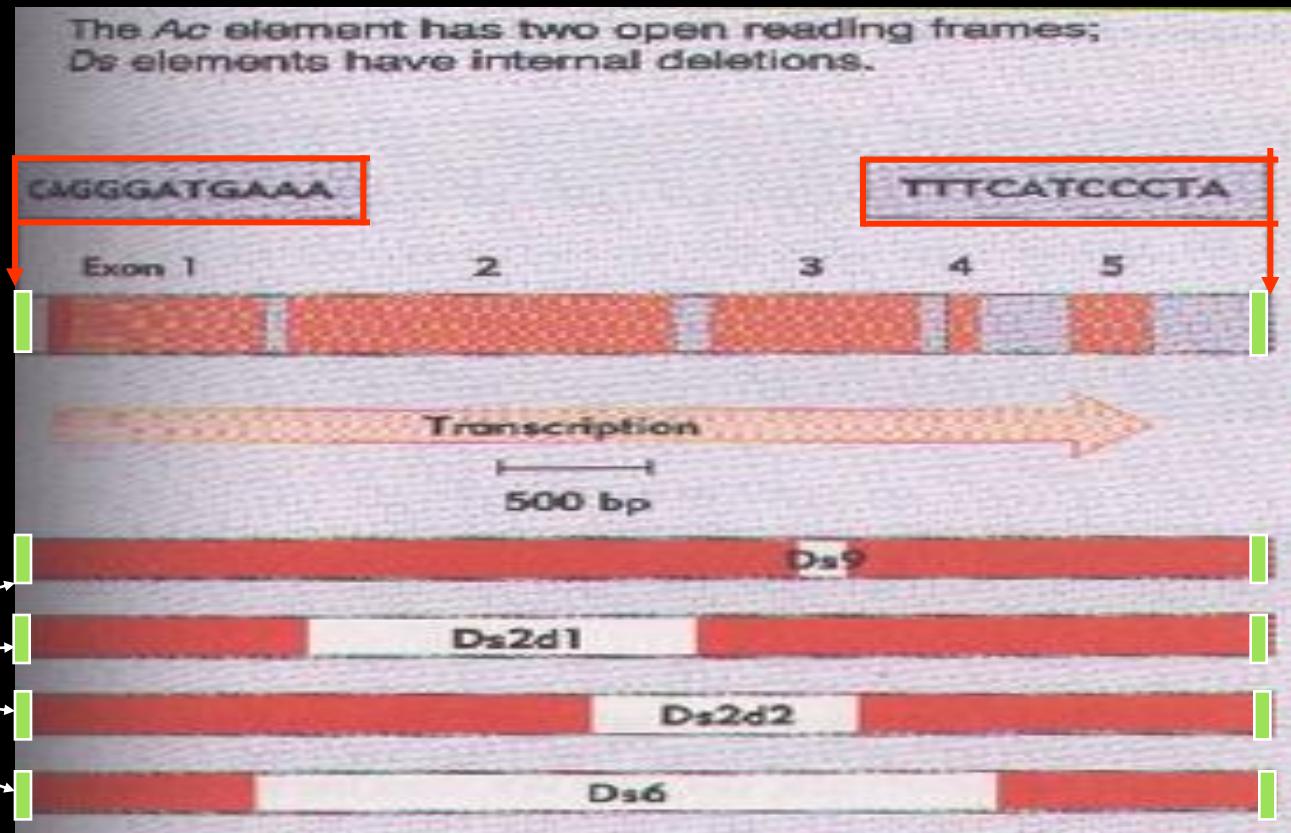
非自主型

复等位易变基因 multiple allele

Ac/Ds 因子的克隆及序列分析



(来源: 分子生物学 (2007), 郑用琏, 第80页)

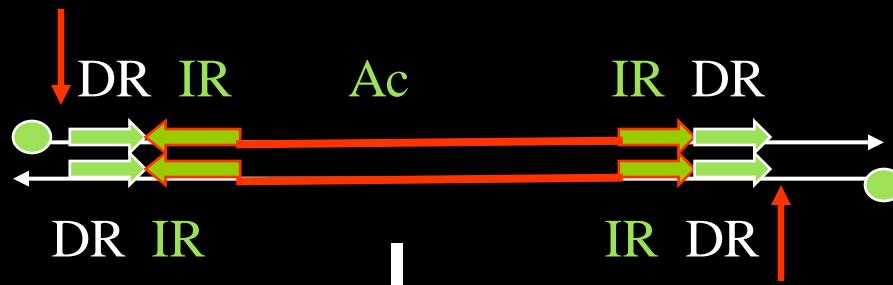


(来源：分子生物学（2007），郑用琏，第81页)

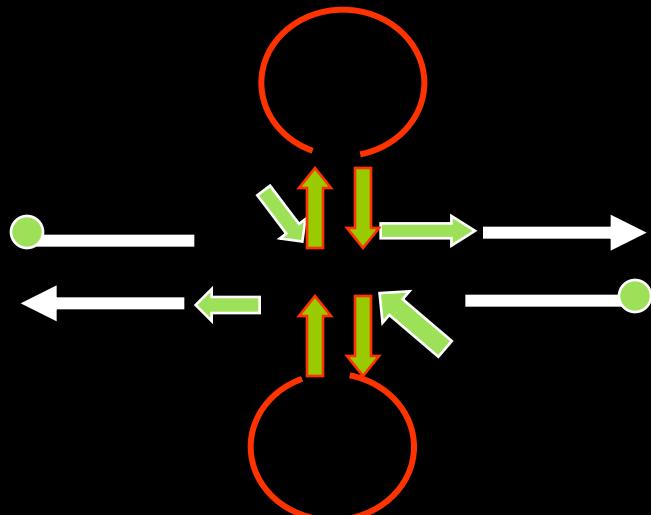
Ds 是 Ac 因子中心区发生不同缺失而形成的突变体
所有丙因子转座系统共
 丧失合成转座酶功能
同的分子生物学特征！
 具有IR序列 → 获得被动转座功能

转座机制

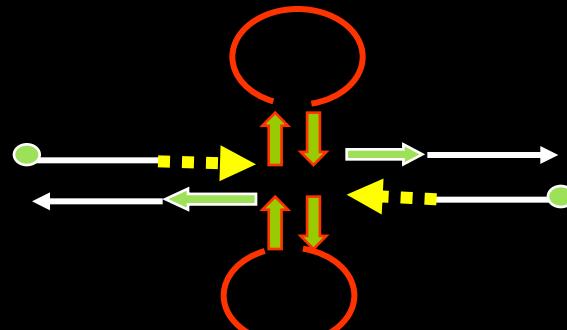
Cut-paste transposition in eukaryote



Excision by
Staggered cut



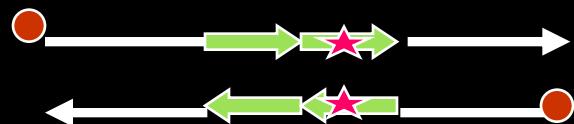
replication



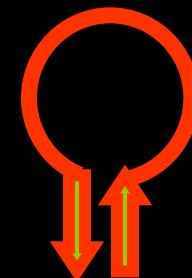
nuclease digestion

footprint

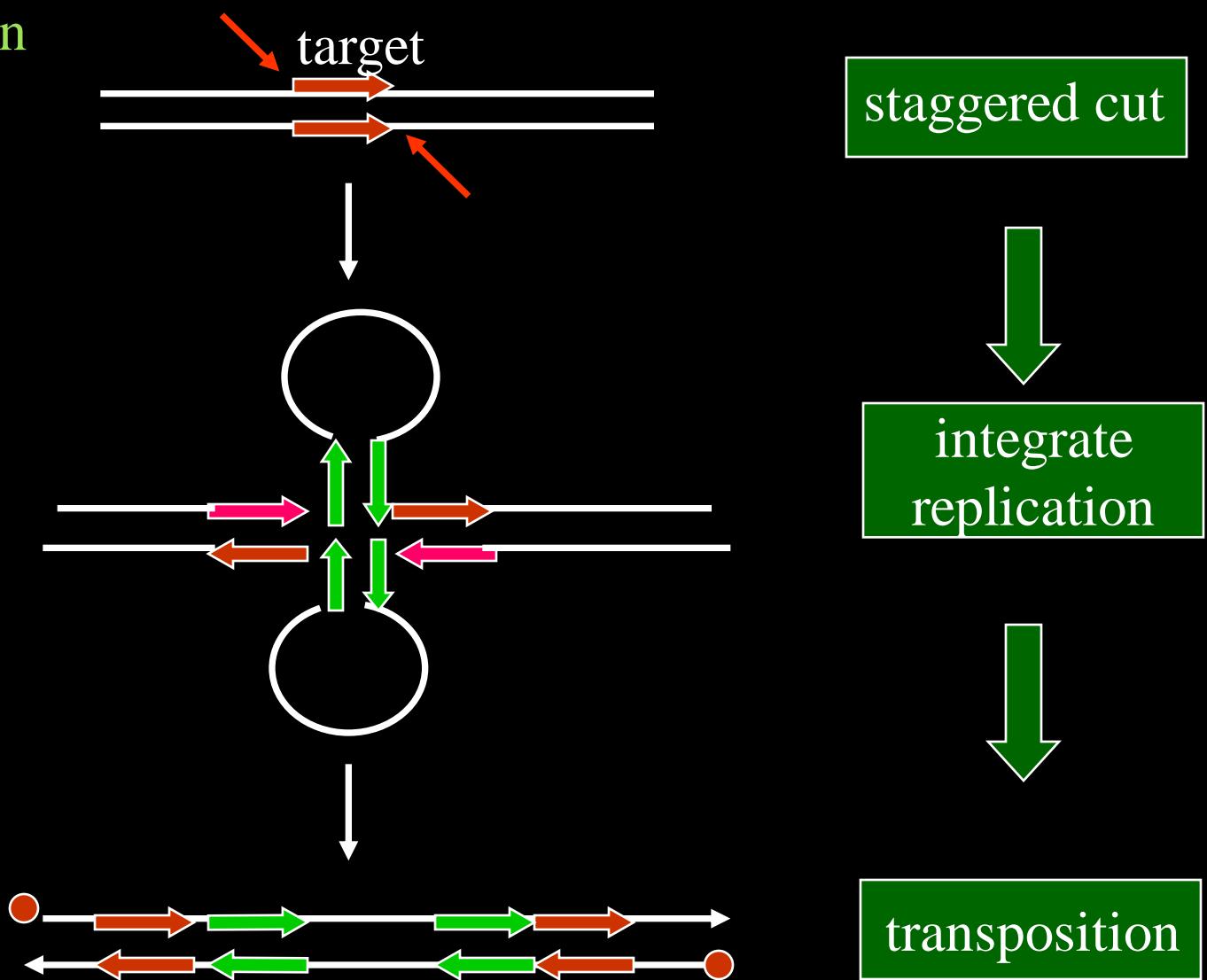
Tn without drag



DR



Transposition by Integrate



Euk. 与 Prok. 转座模式的异同比较

相似处：

- * 靶位点中 **stagger cut** 序列以**DR**形式出现在转座因子两侧
- * 转座因子的**IR**序列是转座酶的识读和作用的位点
- * 转座子的准确切除可能会引起靶基因的回复突变

不同点：

原核生物

(replication form)

copy 的复制与转移

转座与切除是不同的事件

非准确切除表现缺失，倒位效应

真核生物

(cut & paste form)

多数情况为切除后再整合

切除 → 整合 → 转座为同一事件的不同过程

非准确切除主要表现
footprint 效应

反转录转座子 (retro-transposon)

类似反转录病毒

原核生物与真核生物中均有存在

● 反转录转座子的分类

依其DNA结构及反转录酶的DNA的序列（分为3类）

☆ Ty1-copia

最早在果蝇，酵母中发现

transposable element in Yeast Ty

（与反转录病毒结构非常相似）

- 具有长末端重复 结构 (long terminal repeats **LTRs**)
- 核心蛋白基因 (**group-specific antigen gene-like gag**)
- 酶基因区域 (**polyprotein-like 多基因区域 pol**)
- 但无病毒包膜蛋白基因 (**envelop protein env**)

☆ Ty3-gypsy 类反转录转座子
(retro-transposon-like)

结构与 Ty1-copia 相似

具有 LTRs, gag-like, pol-like

但 pol 区域中酶基因排列序列与 Ty1-copia 不同

☆ LINE (long interspersed nuclear element)

无 LTRs 结构的类反转录转座子

有 gag-like 和 pol-like 区域中的部分酶基因

● 高等植物中的反转录转座子

植物中随物种和反转录转座子类型不同

反转录转座子的分子大小

从**1000 ~ 9300 bp**

LTRs 大小

从**120 ~ 2400 bp**

分布

Ty1-copia ; 茅藓门 (苔纲, 蕚纲)

蕨类门 (石松纲, 松叶蕨纲, 楔叶纲, 真蕨纲)

裸子植物门 (苏铁纲, 银杏纲, 松柏纲, 买麻藤纲)

被子植物门 (单子叶植物纲, 双子叶植物纲)

近100多种, 几乎覆盖所有植物种类

Ty3-gypsy ; 玉米和一种百合植物

LINE ; 玉米、甜菜、拟南芥、和一种百合植物



反转录转座子的转座方式及特点

DNA → RNA → DNA

高度的遗
传稳定性

高拷贝性

玉米；基因间几乎由具LTRs 反转录转座子填充
多数DNA 结构完整
部分呈相嵌 (chimera) 方式



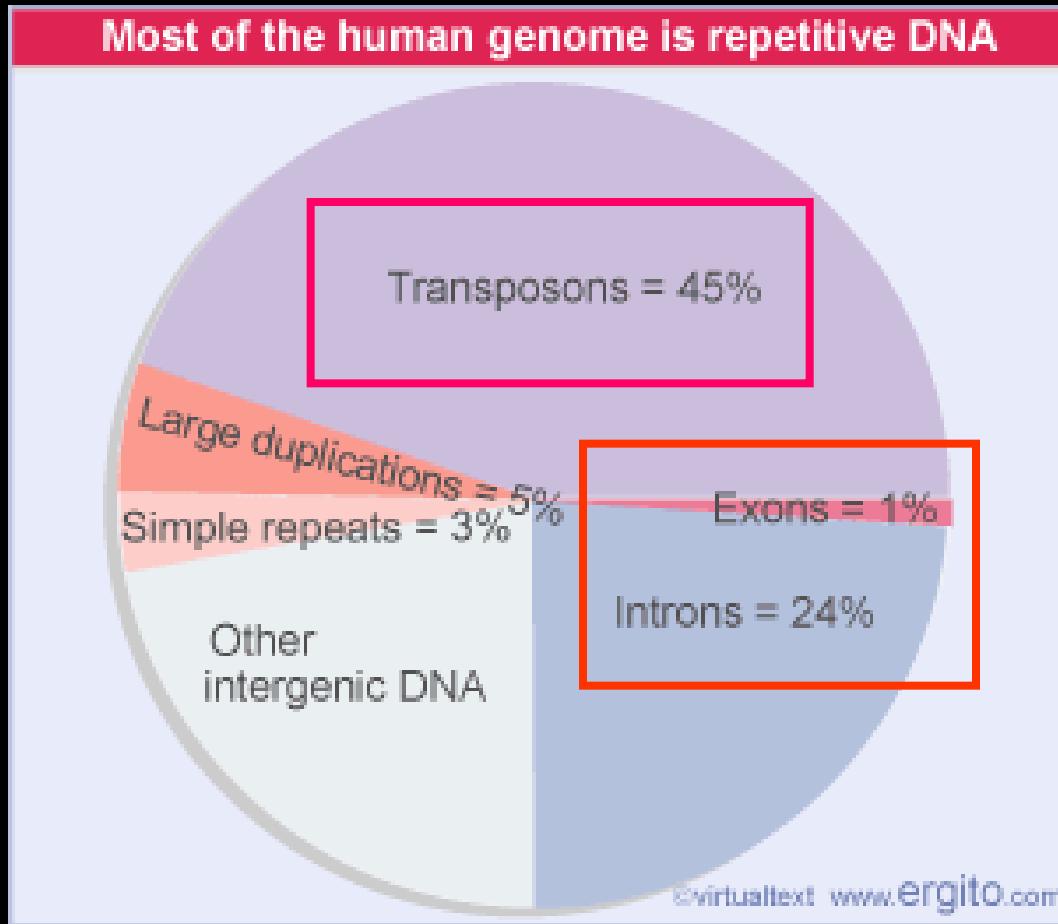
高拷贝性 ?

| species | retrotransposon class | copy number |
|--------------------------|-----------------------|-------------|
| <i>Zea mays</i> | LINE | 50-100 |
| <i>Nicotiana tabacum</i> | Ty1-copia | >100 |
| <i>Oryza sativa</i> | Ty1-copia | 1000 |
| <i>Arabidopsis</i> | Ty1-copia | 0.1% of C |
| <i>Zea mays</i> | Ty1-copia Ty3-gypsy | 50% of C |

推测；高等植物基因组大小(**C**)的差异

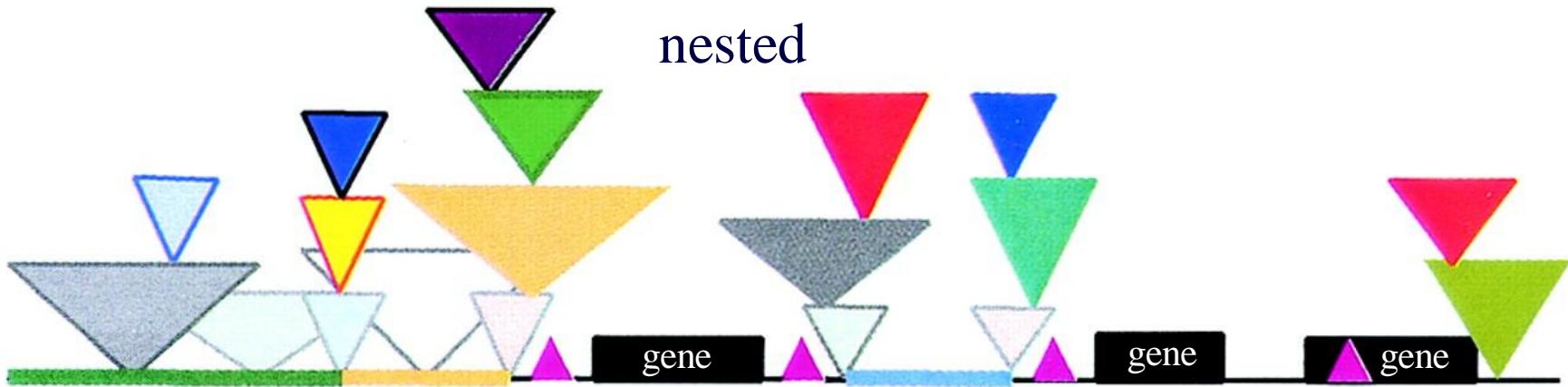
→ 主要由 **retro-transposons** 拷贝数所造成

Most of the human genome is repetitive DNA



(来源：不详)

The largest
component of
the human
genome
consists of
transposons



(来源：分子生物学（2007），郑用琏，第87页)

Tens of Mbp of satellite repeat sequence interspersed with RTE

(PNAS,1998,p395-406)

RTE prefer to insert within each other, resulting in nested groups of transposons in between genes.

(Science 1996,274:p765-768)

Mutator(*Mu*), Miniature inverted-repeat tranposable elements (MITEs), insert preferentially into genes.

(Genome Res. 2000,10:p982-990)

高异质性

- 同类群反转录转座子间（除类群特征的保守序列外）具有高度异质性

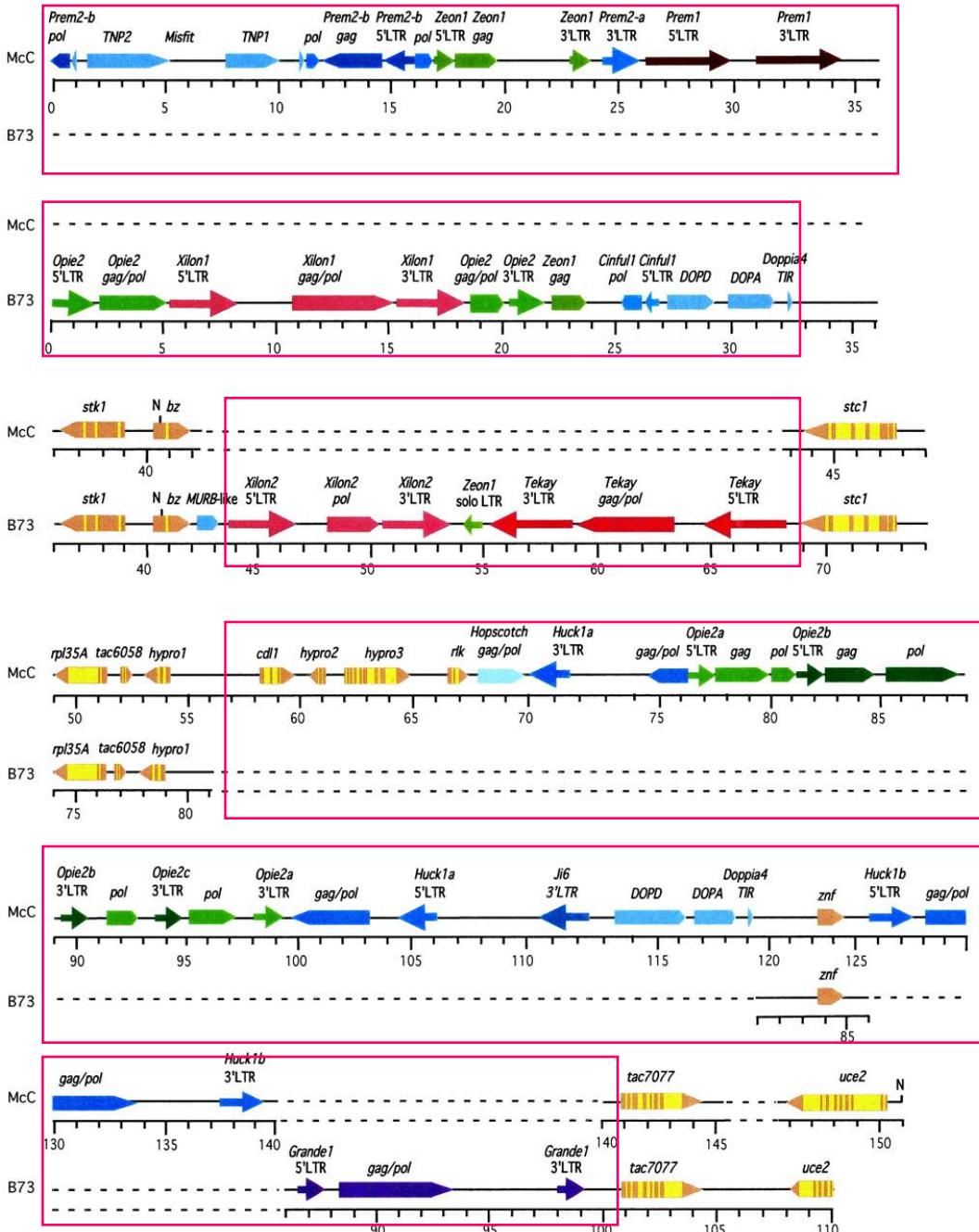
表现在不同植物间

同一物种不同品种间

同一基因组内

Fu H and Dooner HK

**Analysis 2 BACs
sequence isolated from
a commercial library
by using a *bz* probe
between two inbred
McC and B73**



除黄色箭头标记的
基因外,两个自交系
间存在大量
*LTREs, gag/pol*差异

(Source:July 9, 2002
PNAS.vol
99/no.14/9573-9578)

高异质性

原因有三；

纵向传递(vertical transmission)

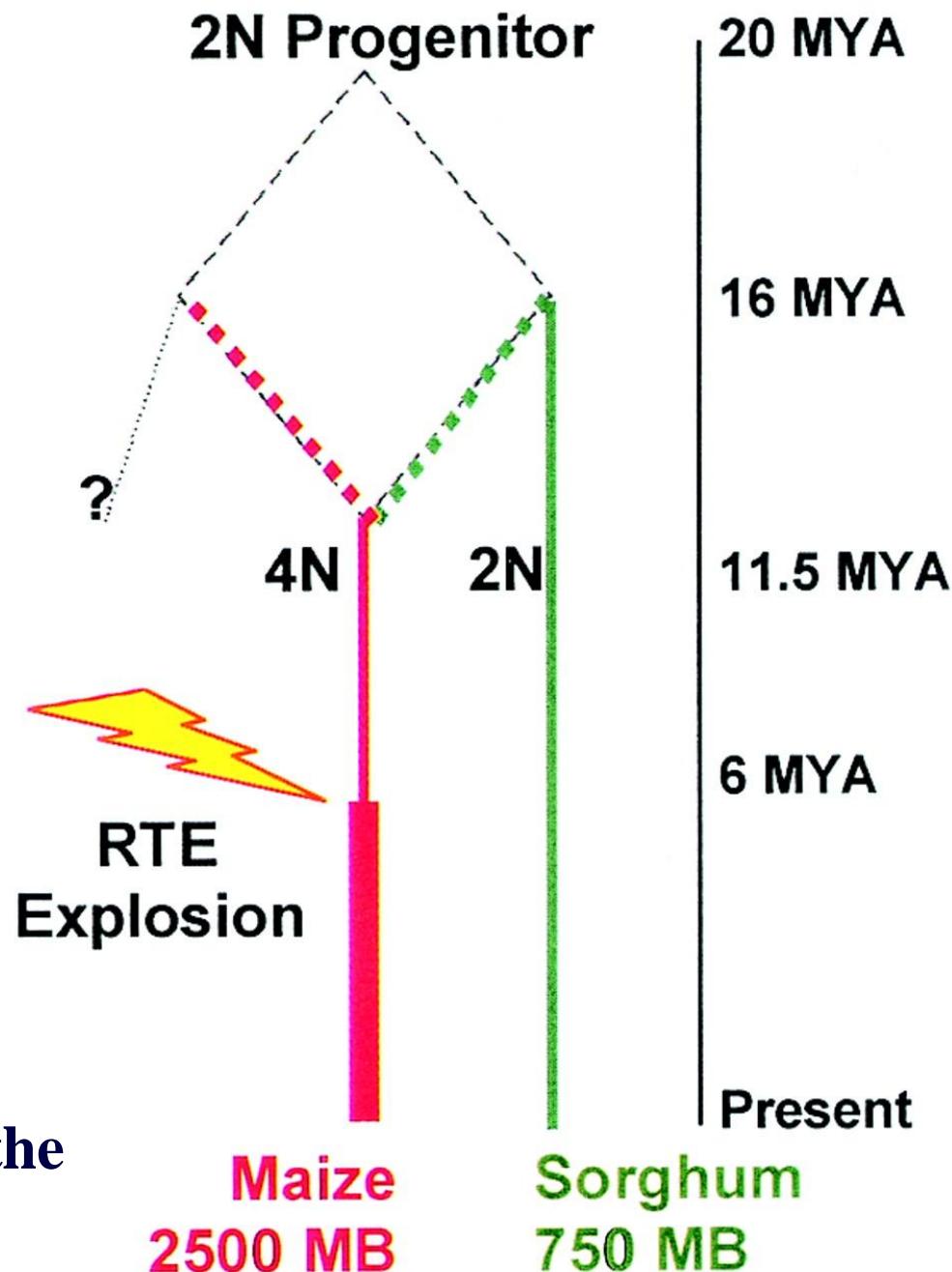


Source:PNAS July 17.2001.
vol.98.no.15 p8163-8164

form a tetraploid

repeated bursts of
retrotranspison

RTE about 80% of the
maize genome

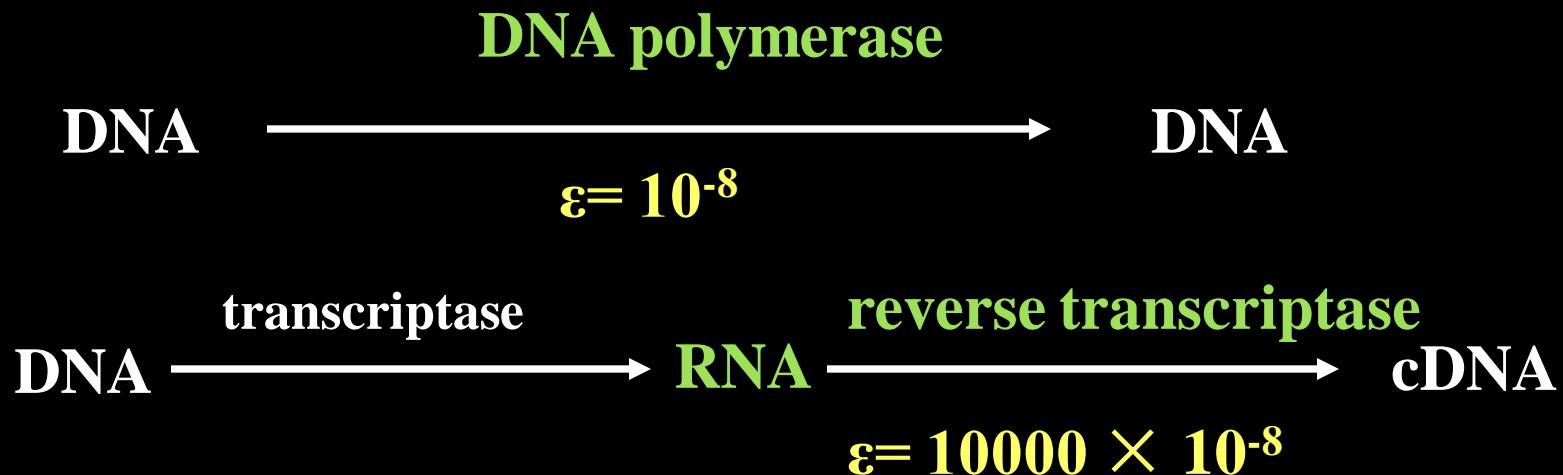


横向传递(horizontal transmission)

植物病原物的媒介传递
小麦 —————→ 马铃薯

证据：分化约在2亿年前, Ty1-copia相似性高达57%

反转录过程的错读



一种新型转座子Helitron 的发现与特点



(source:Kapitonov V.V. and Jurka (2001))



(source:Kapitonov V.V. and Jurka (2001))

described a new type of DNA transposon, called a *Helitron* , by computational analysis of genomic sequences from *Arabidopsis*, *Maize*, *barley*, *wheat*, *rice*, *C . elegans*, *droshophila*, *fungi*, *fish*



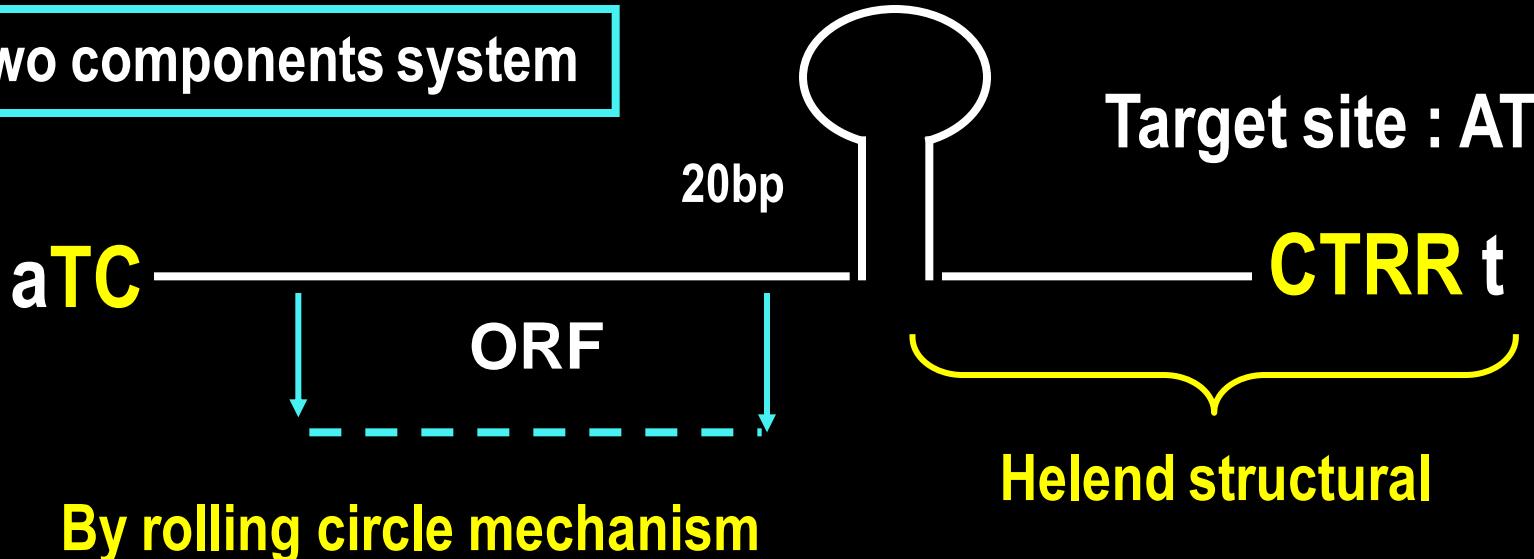
(source:Kapitonov V.V. and Jurka (2001))

Helitron

insertion hotspot: btw A and T,

Structure: 5' TC and 3' CTRR termini, and a palindromic sequence capable of forming a hairpin loop near the 3 terminus (11Nt) .

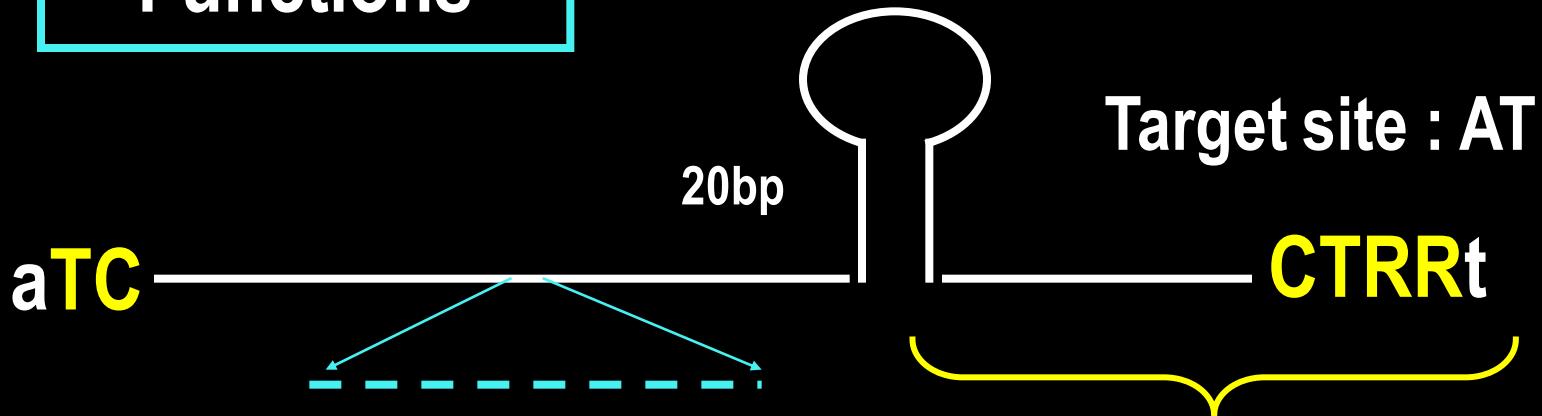
Two components system



By rolling circle mechanism

1. Autonomous *Helitrons* encode a DNA / RNA helicase and one or two copies of a nuclease / ligase, similar to genes encoded by known prokaryotic rolling circle replicons.
2. Most *Helitrons* were classified as non-autonomous because they did not encode the complete set of proteins found in the consensus elements and they shared only the common structural hallmarks with the “autonomous” *Helitrons*

Functions

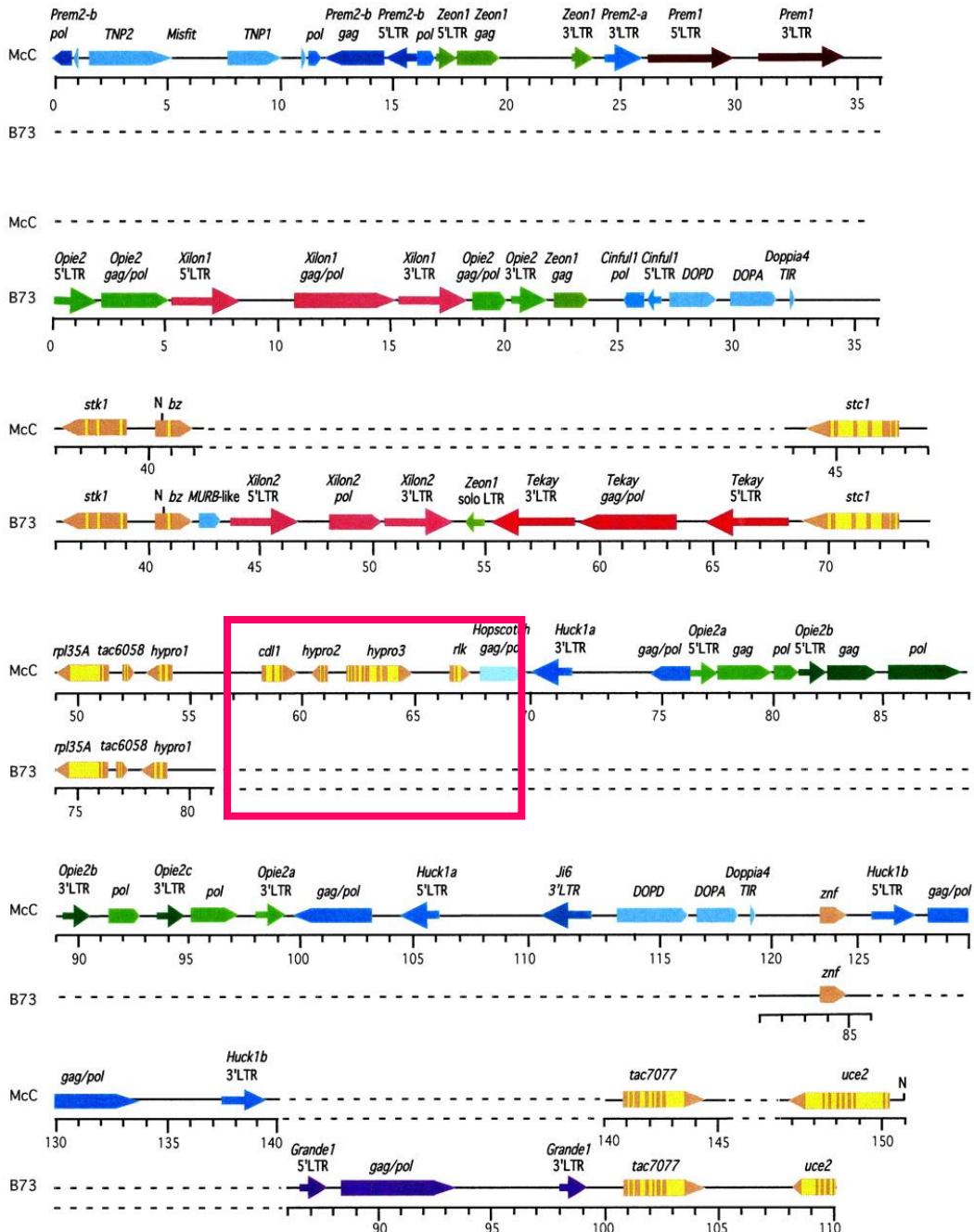


By rolling circle mechanism

Gene amplification

Translocation with structural gene

Exon shuffling...



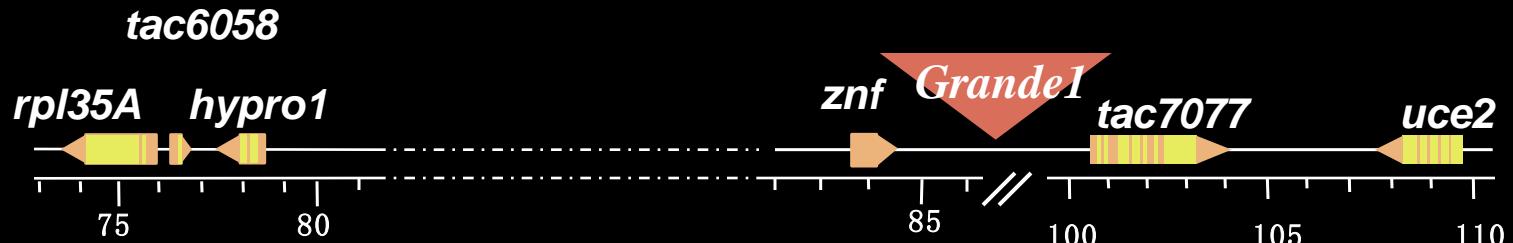
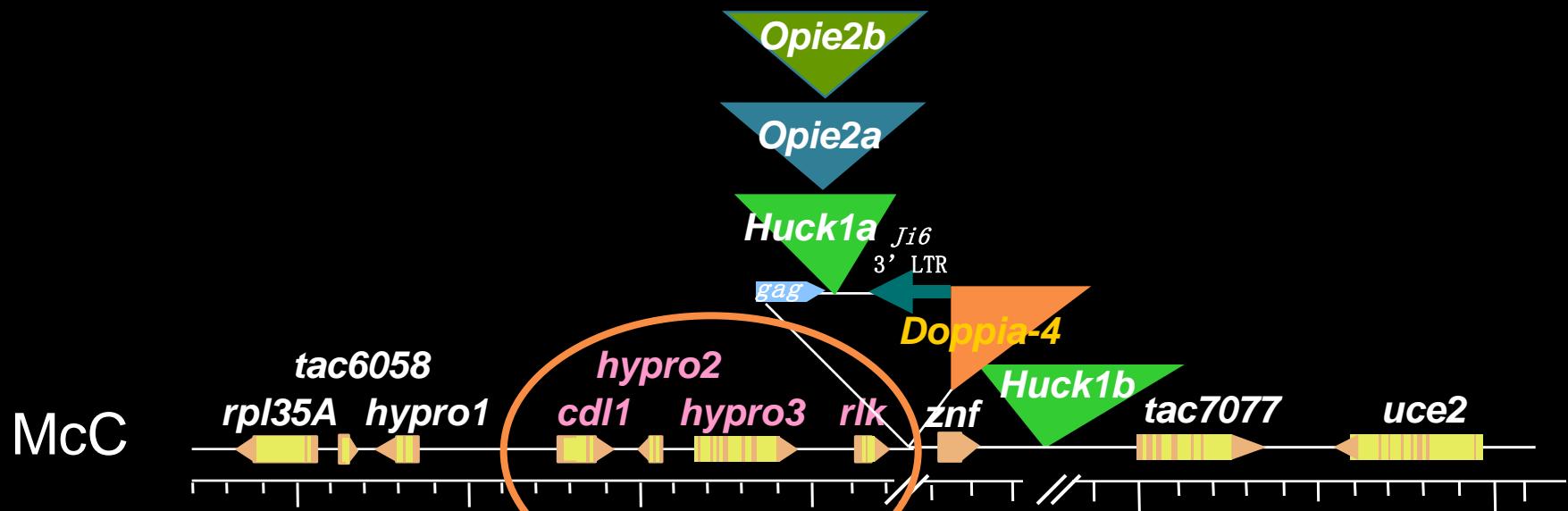
Fu H and Dooner HK

Analysis 2 BACs sequence isolated from a commercial library by using a *bz* probe between two inbred McC and B73

10-kb fragment containing the 4 genes (*cdl1*, *hypro2*, *hypro3*, and *rlk*) that are present in the *bz* region of McC, but not of B73,

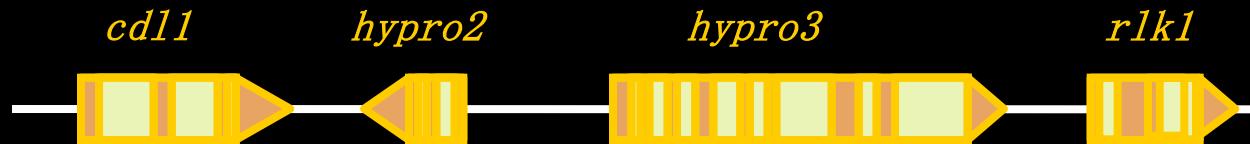
Souce:July 9, 2002
PNAS.vol
99/no.14/9573-9578

Gene islands also vary among lines



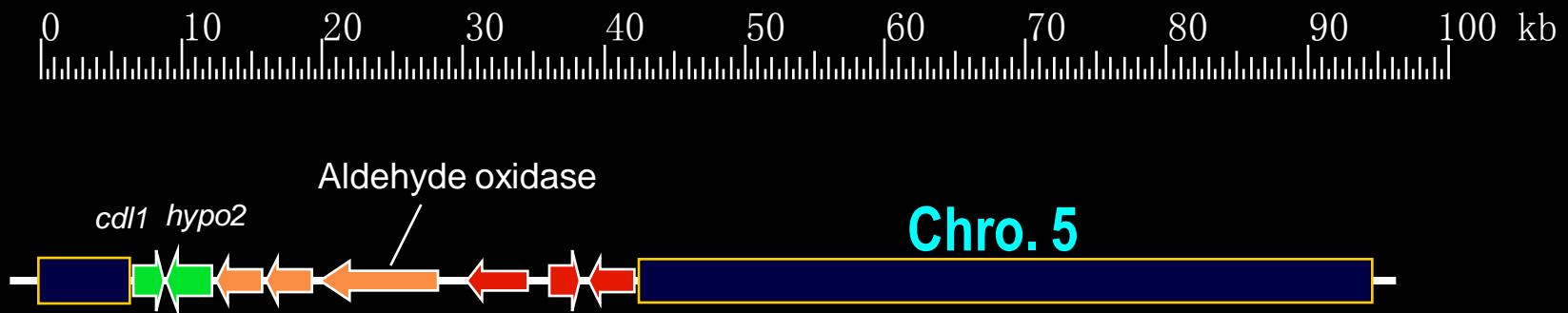
(来源: 分子生物学 (2007), 郑用琏, 第85页)

Where are the “missing” genes in B73?



Blast-N analysis of maize GSS database (mostly B73)

Annotation of sequenced b0511l12 clone



- Present elsewhere in the genome
- Some are adjacent to each other

2.6.5 转座的遗传效应

诱变效应（提高重组频率、形成易变基因...）

切除效应(倒位、缺失、重复、foot printting)

外显子改组（exon shuffling）

位置效应（启动表达、增强表达...）

转座爆炸（激活表达、基因内重排突变基因形成）

诱变效应

- 高频率的回复突变，形成易变基因（mutable gene）
- 不同复制子间的融合，抗性积累
- 同一复制子中多个转座因子间的同源重组，导致基因组序列的重排
- 就个体而言，转座形成突变，引起生理负担

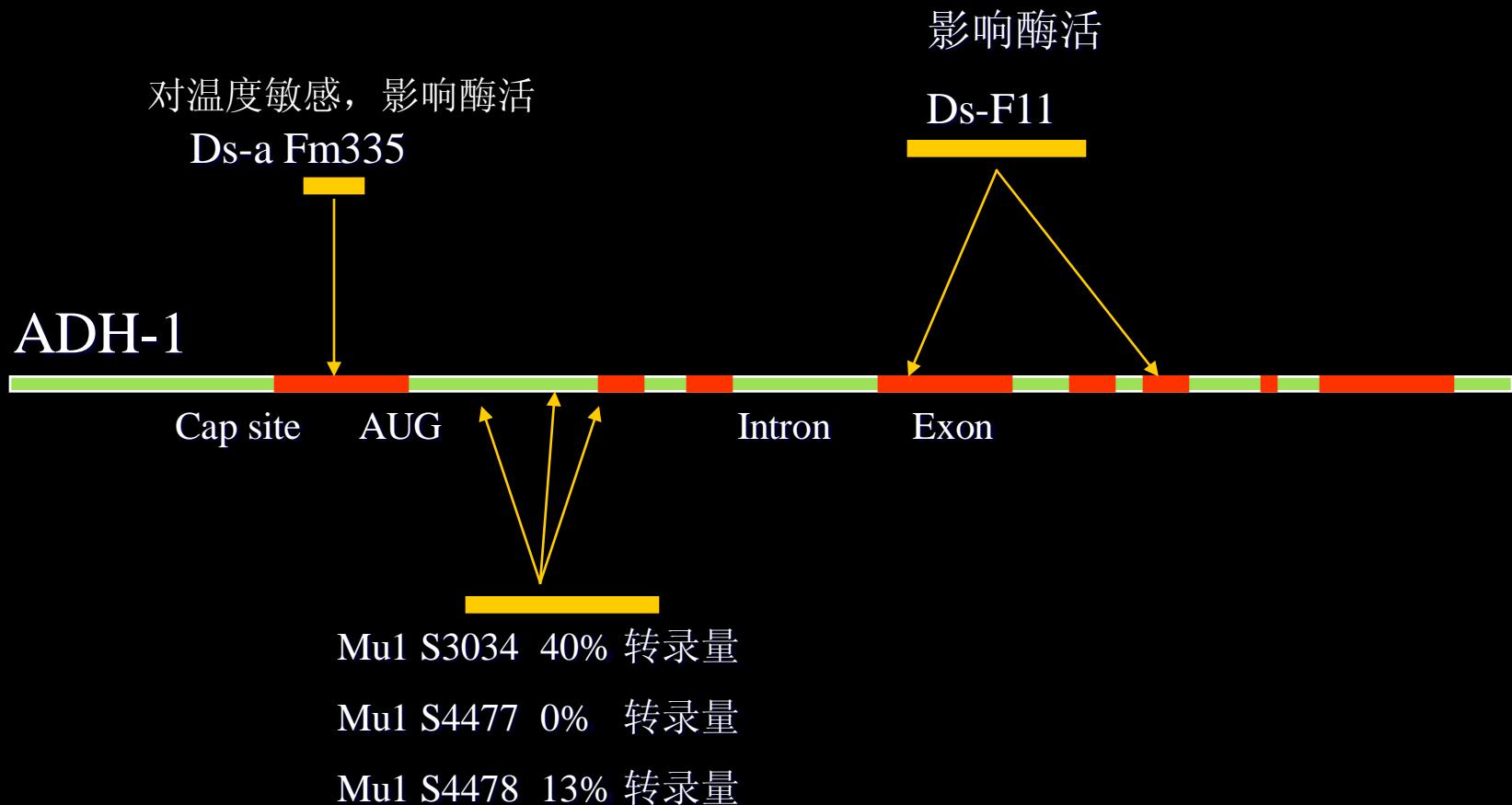
—————> 进化选择劣势

就整体而言，转座引起突变增加

—————> 进化选择优势

- 基因插入钝化（调节基因突变，结构基因突变或极性突变）

Ac, Ds, Mu插入不同的位点， 产生具有不同遗传效应的复等位基因



(Source: Osterman JC , 1981, Genetics,
99)

切除效应(excision)

a) precisely excision

→ back mutation

b) imprecisely excision

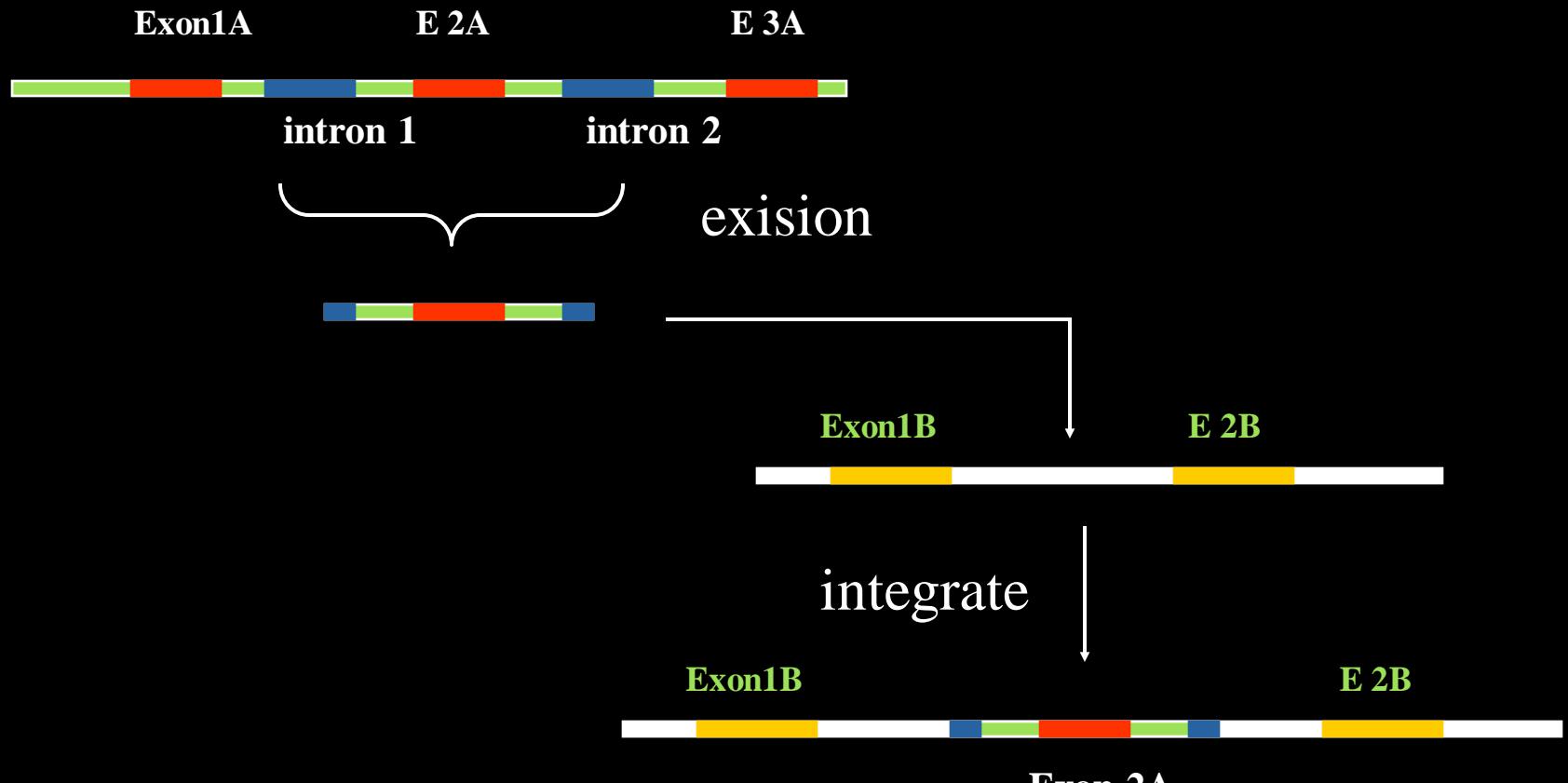
→ DNA deletion & inversion

c) excision

→ antibiotic-resistance lose
footprinting

| genotype | events | DNA seq. | Protein |
|--------------------------|-----------|---|------------------------|
| Wx (maize) | w.t | aaG TTc <u>aac</u> | w.t |
| wx-8: : spm18 | mut. | aa <u>G TT</u> ---Spm--- <u>GTT</u> caa c | mut. |
| Germinal excision | | | |
| Wx ⁺ -1 | back mut. | aaG TTA TTc aac | +leu |
| Wx ⁺ -2 | back mut. | aaG TCG TTc aac | +Ser |
| Somatic excision | | | |
| In leaf DNA | | aaG TAC GTT caa c | framshift |
| | | aaG TGT TTc aac | +Cys |
| In endosperm DNA | | aaG TTG TTc aac | +Leu |
| | | aaG TTc aac | w.t |
| | | aaG GTT caa c | framshift |
| | | aa c | △Lys, △Phe Thr mut. |

外显子改组 exon shuffling



exon shuffling

(来源: 分子生物学 (2007), 郑用琏, 第88页)

位置效应（启动表达、增强表达...）

- Retro-transposon with Enhancer

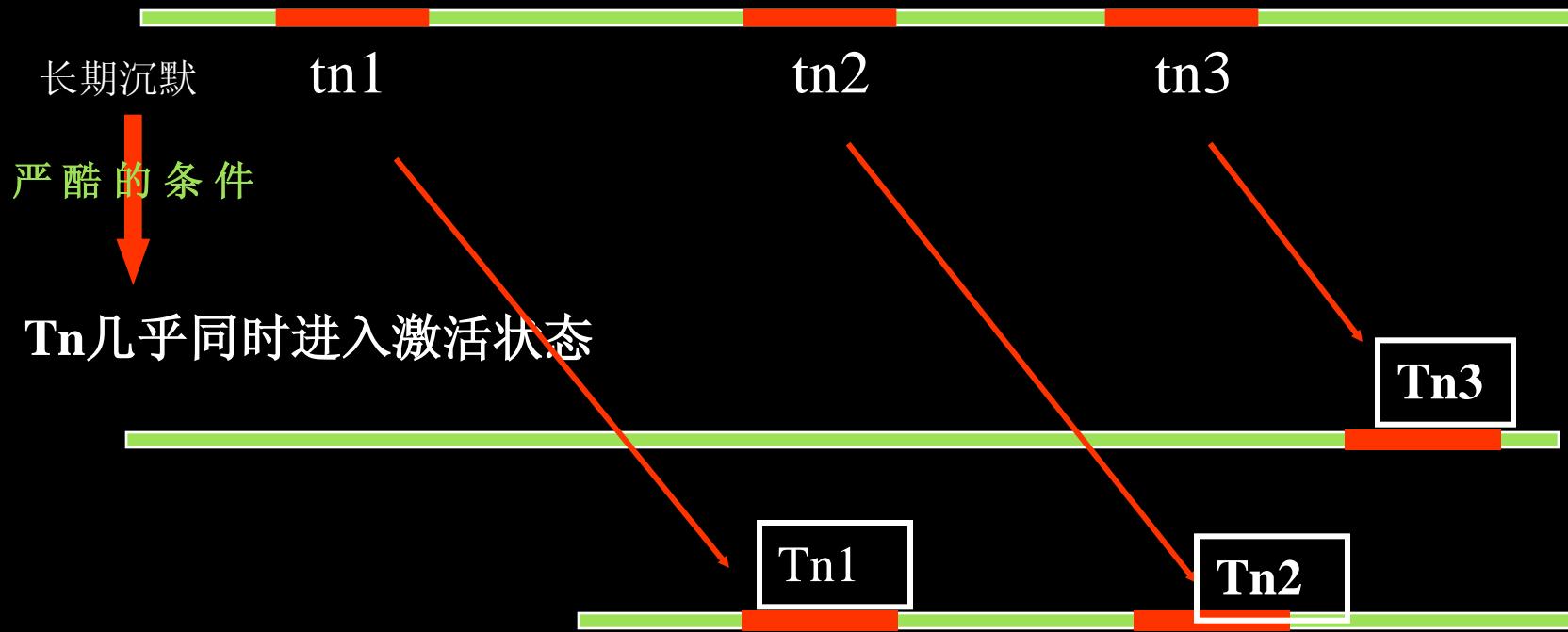
增强靶位点寄主基因的表达

- Composite transposon with promoter

启动靶位点寄主基因的表达

转座爆炸 transposition burst

(激活表达、基因内重排、突变基因形成)



大量变异个体产生 → 物种进化

植物生殖细胞发生转座爆炸 → 雄性不育

● 插入位点

插入热点 (hotspot)

重复序列区域

区域优先 (regional preference)

在特异序列的 $3\text{kb}\pm$ 的区域内，转座子优先插入。

而且与 DNA superhelix , DNA 与 protein 间的结合状态有关

Insertion hotspot

Adh1

---- CGTAACTGGTGAGGGACTGAGGTCTCGG-----

8 bp staggered cut

Ds-c (Adh1-Fm335)

-- CGTAACTGGTGAGGGACTGA**Ds-c...**

.....TTTCATCCCTAGGGACTGAGGTCTCGG----

11 bp IR

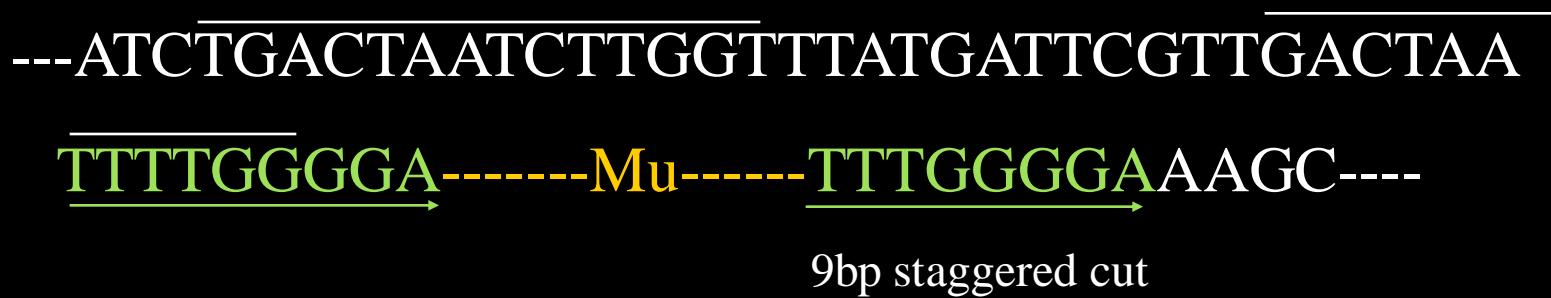
8 bp DR

Insertion hotspot

Ac (wt-m7)



Mu1 (Adh-1)



Insertion hotspot

Spm (wx-m8)

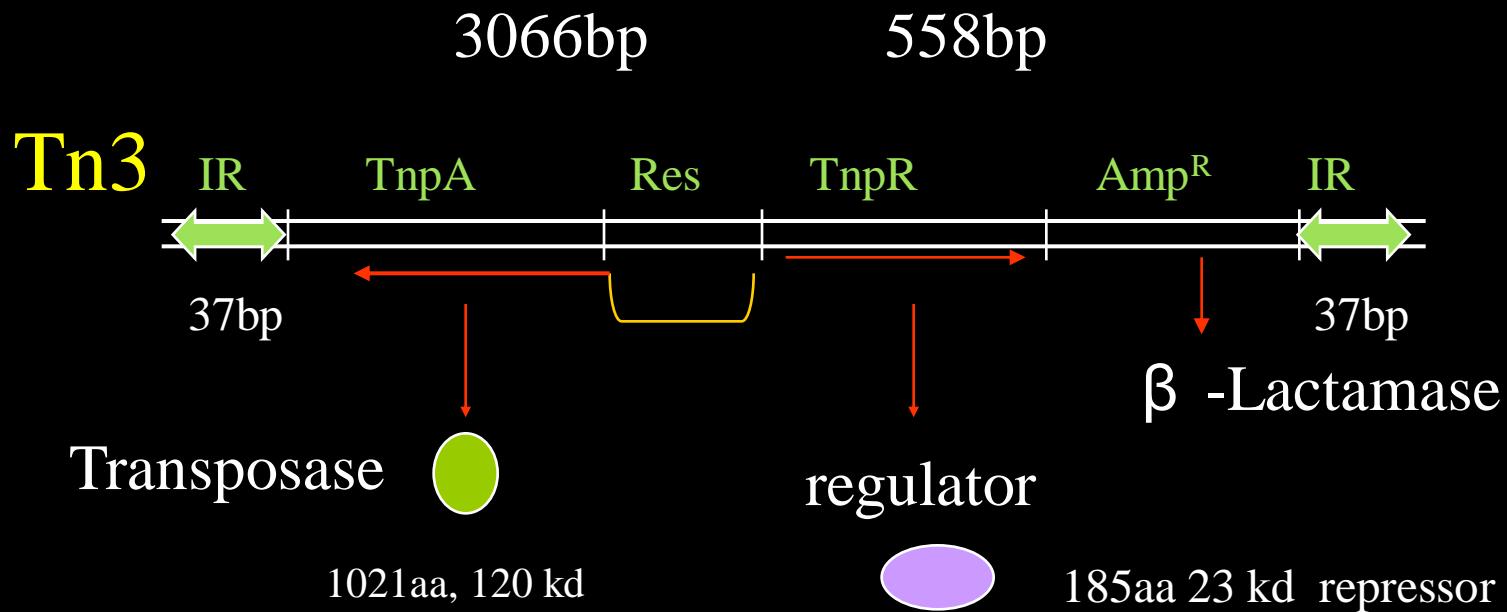
---ATGAGCGCCGAGAAAGTTCCCTAGGCAAGGTGAG

CGCCGTGGTCAAGTT-----Spm-----GTC-----

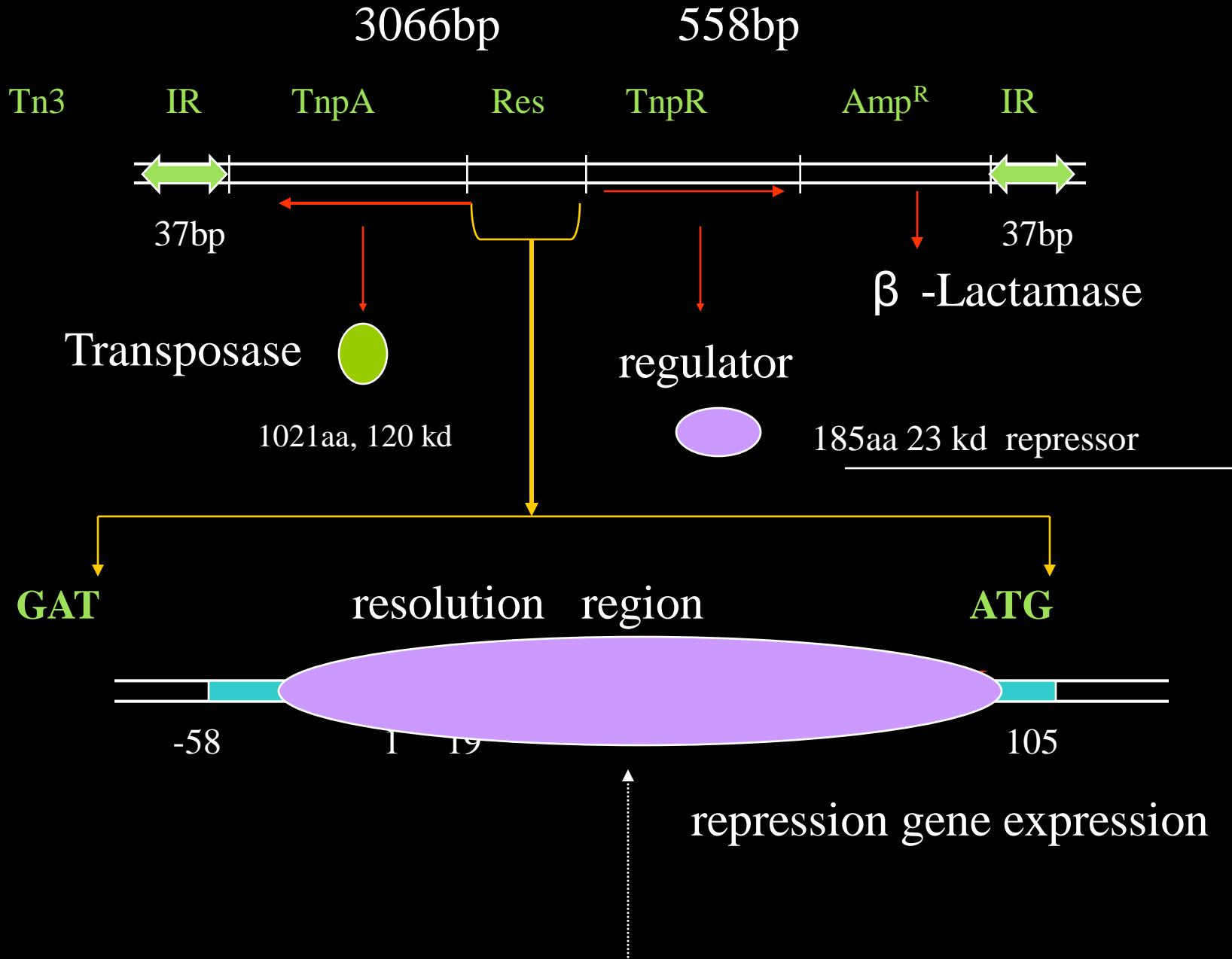
3bp staggered cut

2.6.6. 转座调控机理

a) Tn family 转座调控机理

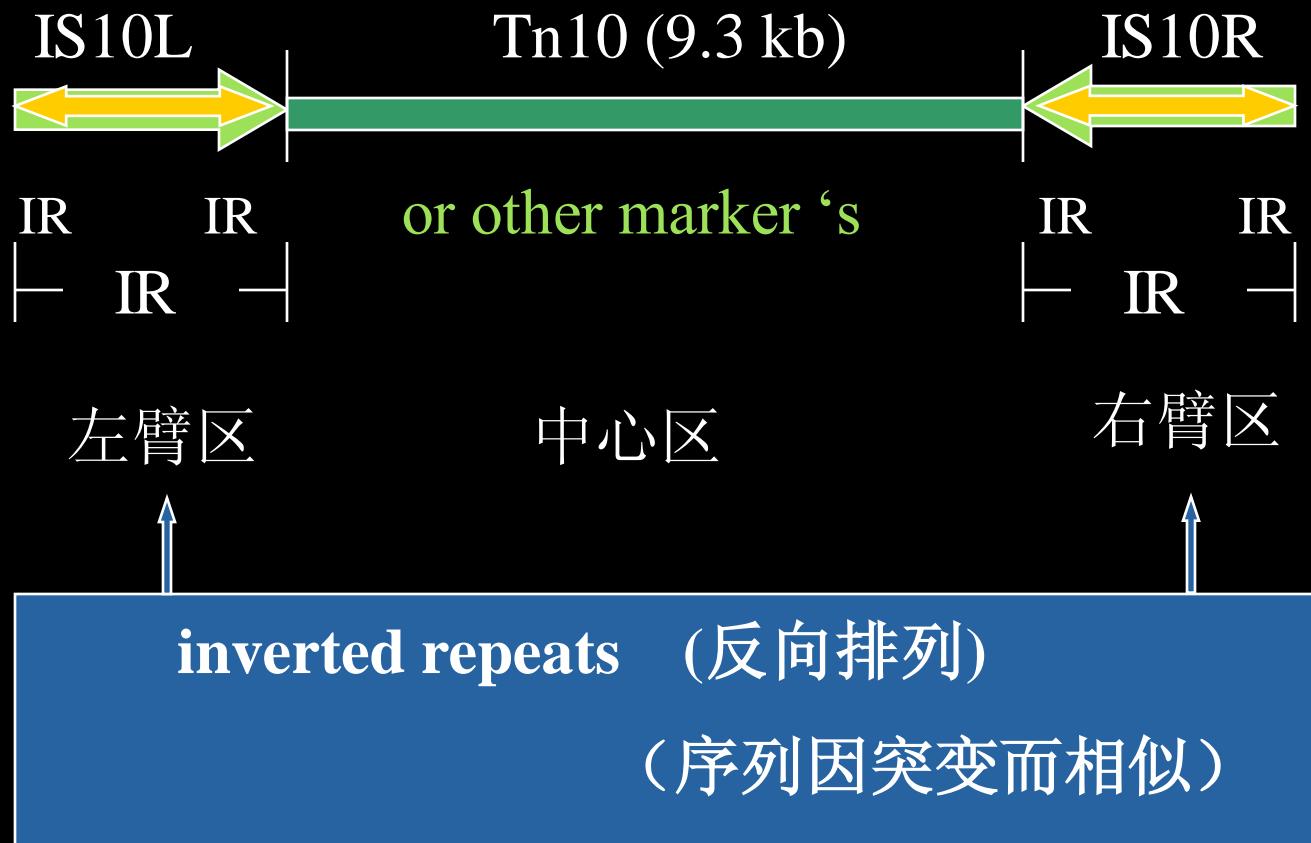


(来源: 分子生物学 (2007), 郑用琏, 第74页)



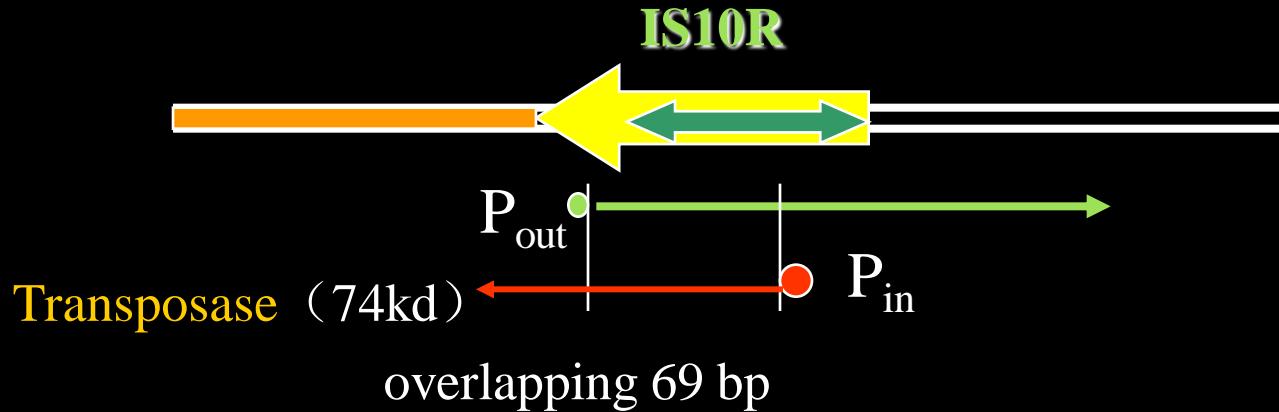
b) 复合因子转座调控机理

e.g. 1 Tn10



(来源：分子生物学（2007），郑用琏，第75页)

● 转座频率调控 (composite element of Tn10)



- ☆ IS10L、IS10R中13bp的IR序列也成为复合因子Tn10转座必须的 IR
- ☆ IS10R启动转录transposase (trans-action factor)具强互补效应并与转座频率明显正相关.
具有**cis-preference**效应（对靶位点序列的识别与结合能力较其他位点更为优先）

● IS10R对转座表现反义RNA效应

P_{out} 向外转录 IR 及 Host 基因序列

P_{in} 向内转录 transposase

P_{out} 转录效率 (100X) > P_{in} 转录效率

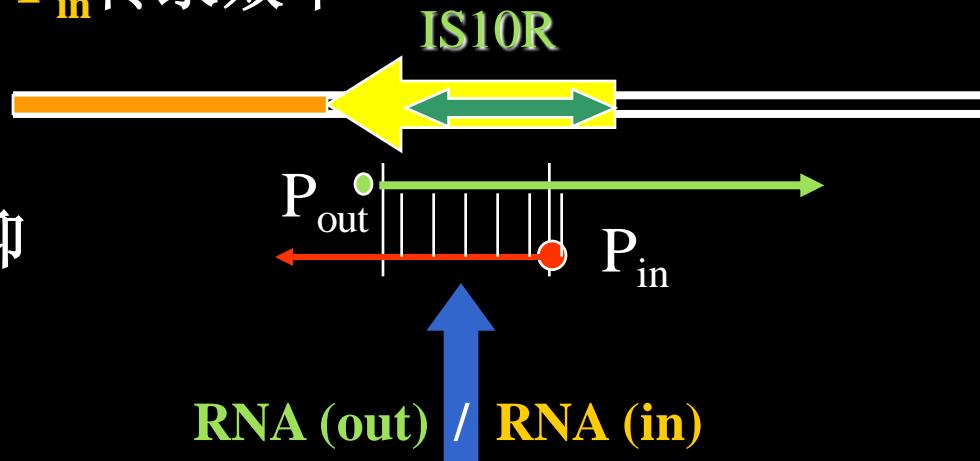
在翻译水平上

P_{out} RNA as anti-RNA 抑

制 Transposase 合成



转座率下降



69 bp overlapping

- Dam (DNA Adenine methylase) 甲基化作用的调控

dam⁻ mutant

Tn10 transposition up 1000 X

?

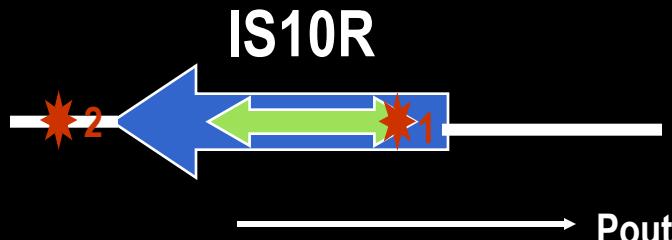
复制后短暂时

DNA半甲基化状态

转座酶上升

IR结合增强

转座被激活



- 位于IR中的甲基化位点1抑制转座酶的结合
- 位于Pin RNA编码区的甲基化位点2影响转座酶的合成

c) conclusion

- 不同转座因子的转座调控机制不尽相同
- 但均具有共同的

限制转座频率的负控制机制

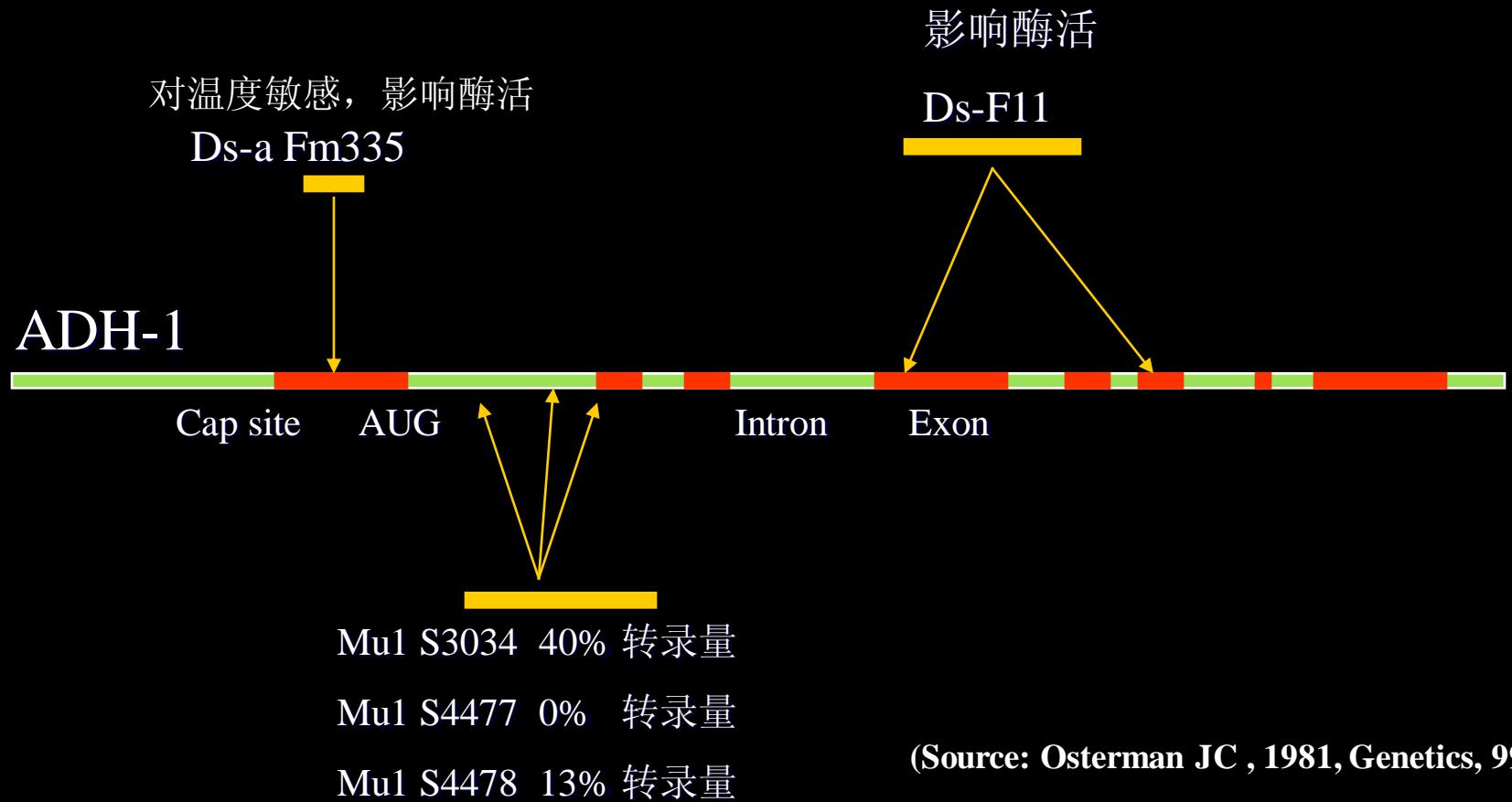
- 在进化过程中，严格控制转座频率

→ 减少突变，避免死亡 **Tn10 转座率
 $=10^{-7}$**

→ 保护自身，扩增衍繁

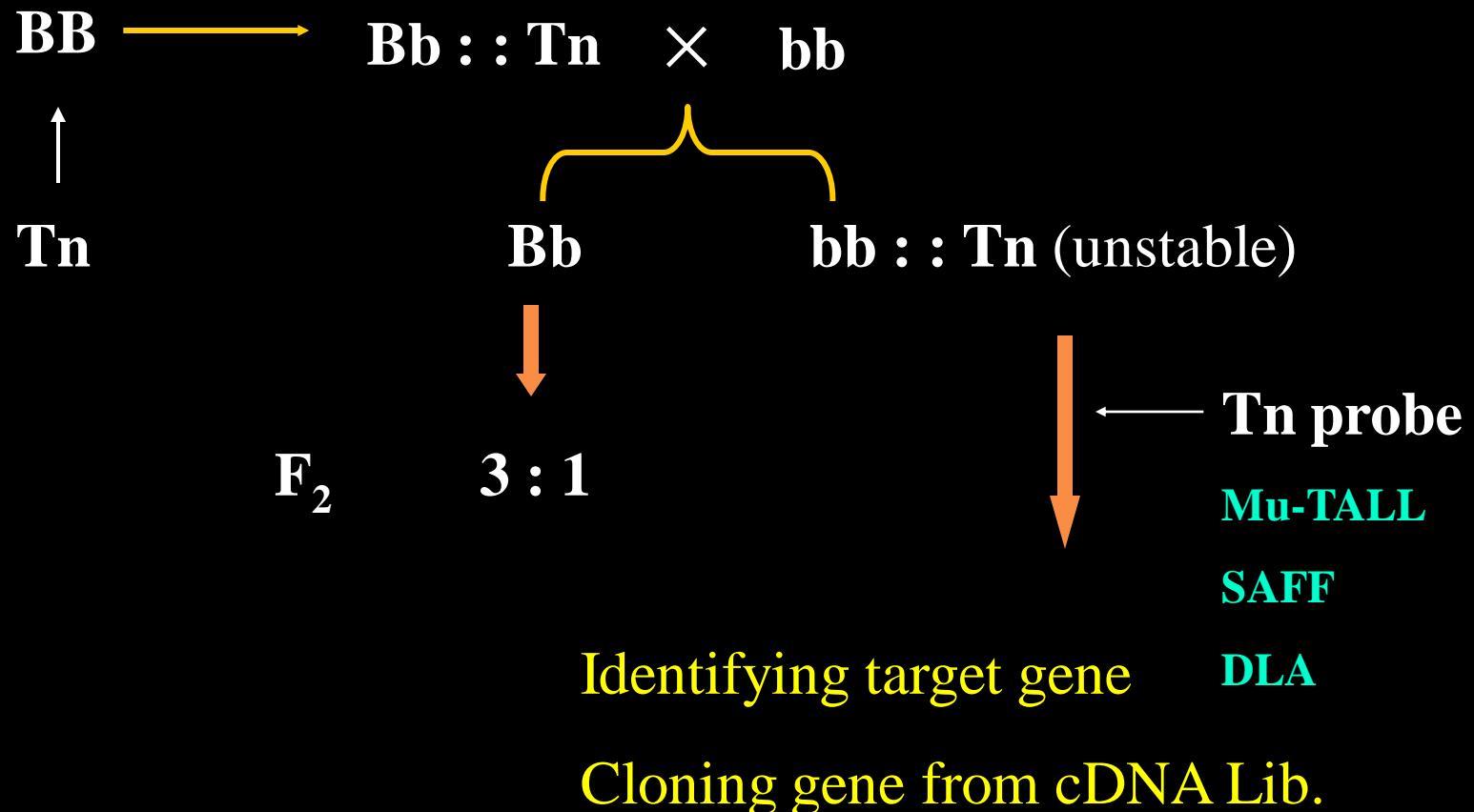
d) 转座因子的应用研究

- 利用转座子的插入突变效应，研究基因的结构与功能区



d) 转座因子的应用研究

- 利用转座子分离，克隆基因



玉米T-CMS Rf2 基因的克隆

178300 Rf2 plants

↓
Mu1

6 of rf2-mutant

↓
probe

cDNA Lib.

↓

2.2kb cDNA

↓

Sequencing(ALDH)



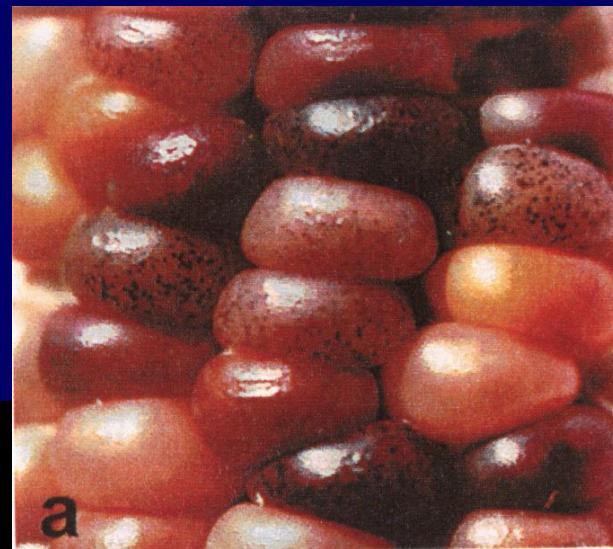
(来源: 不详)

Mutant Information:

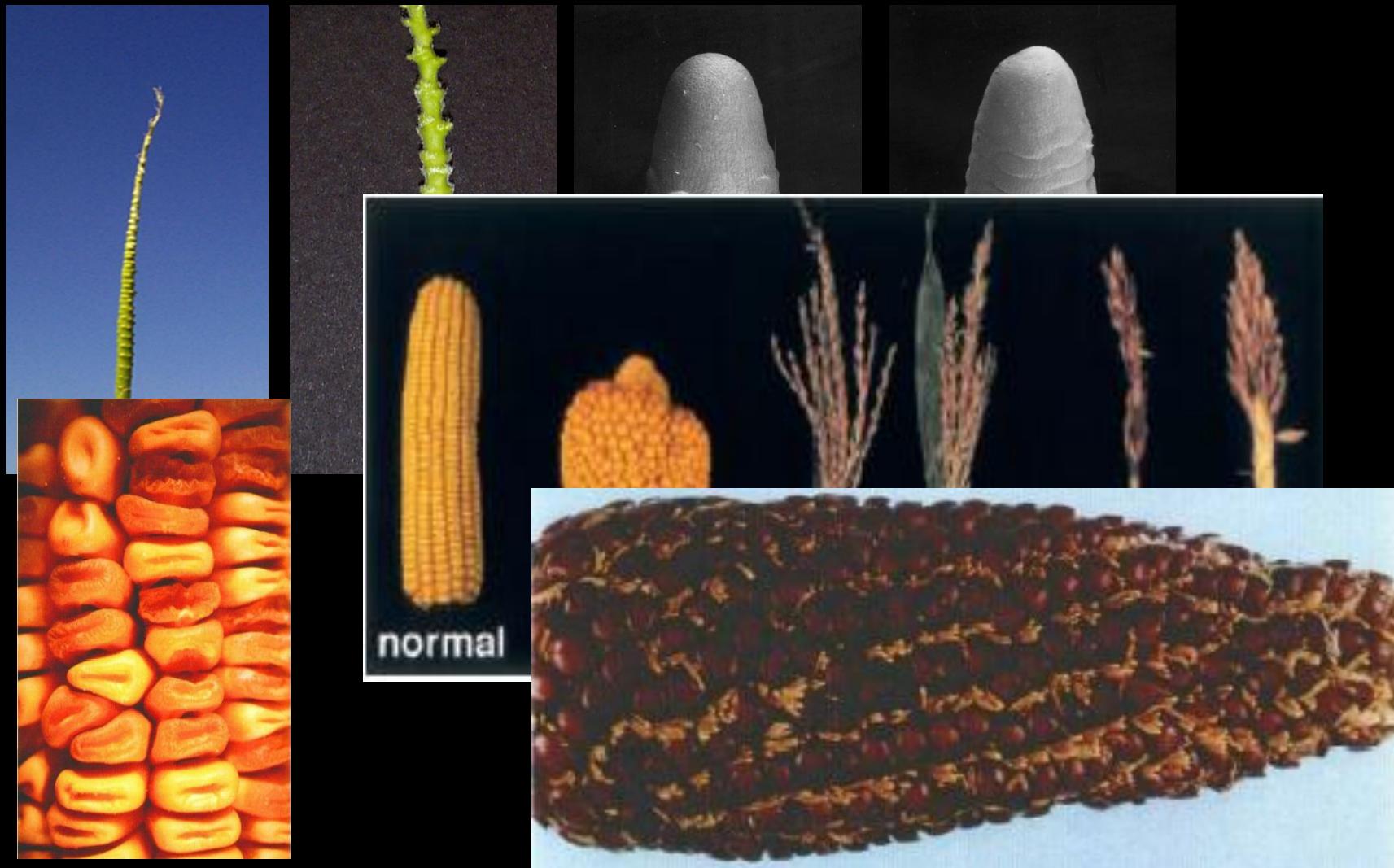
Result: 01S0454-07



BZ2 × Mu Activator



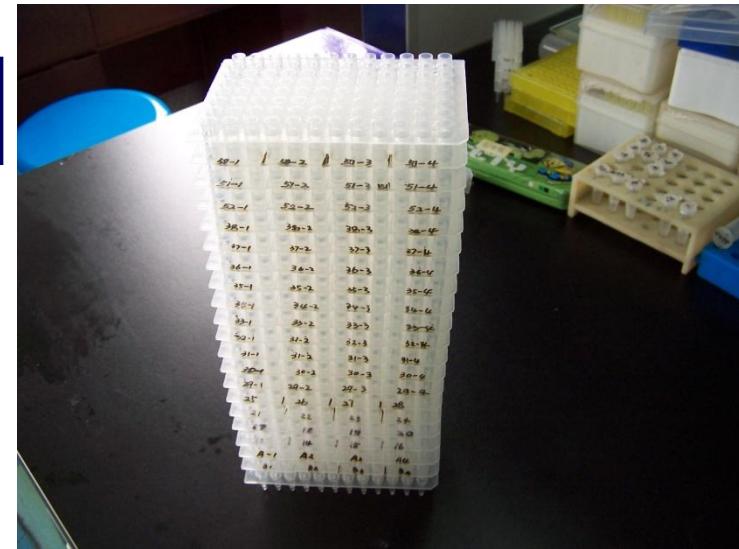
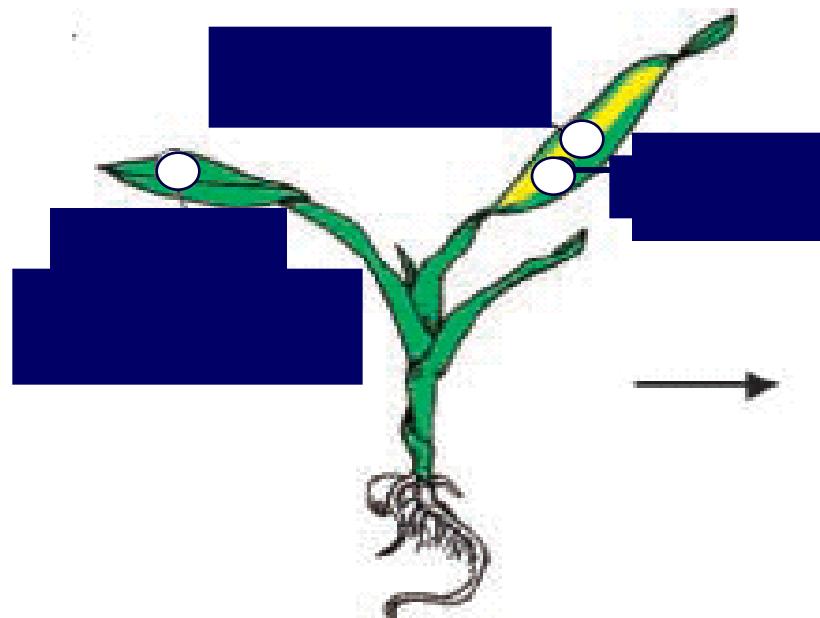
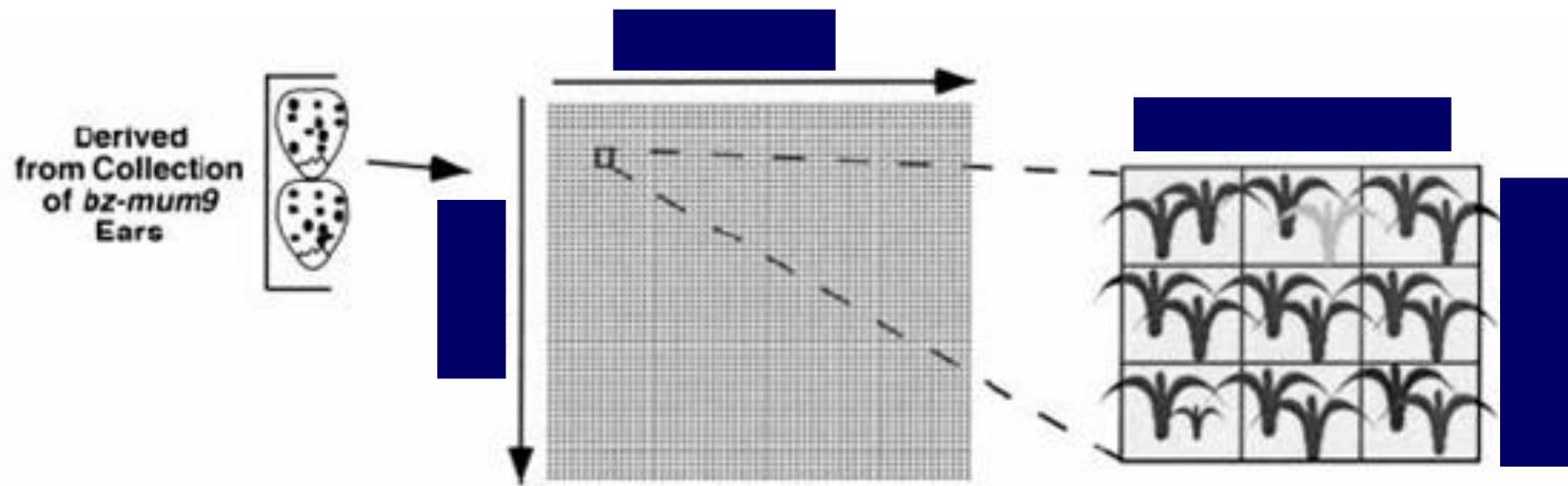
穗状突变体



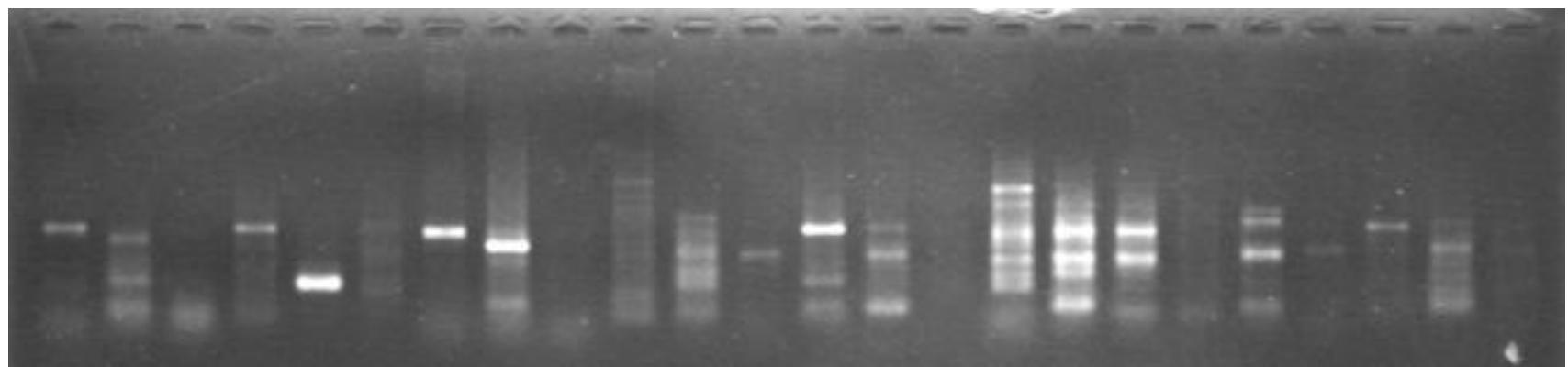
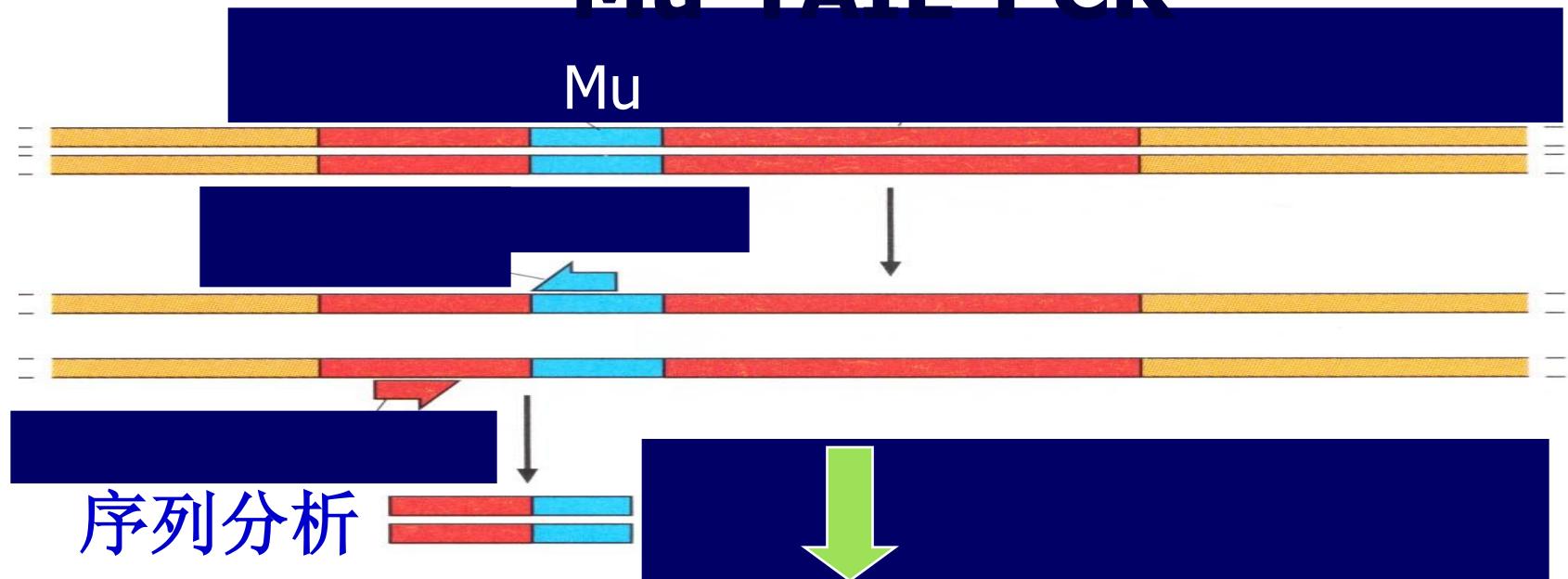


植株突变体

Mu转座子插入突变体鉴定策略

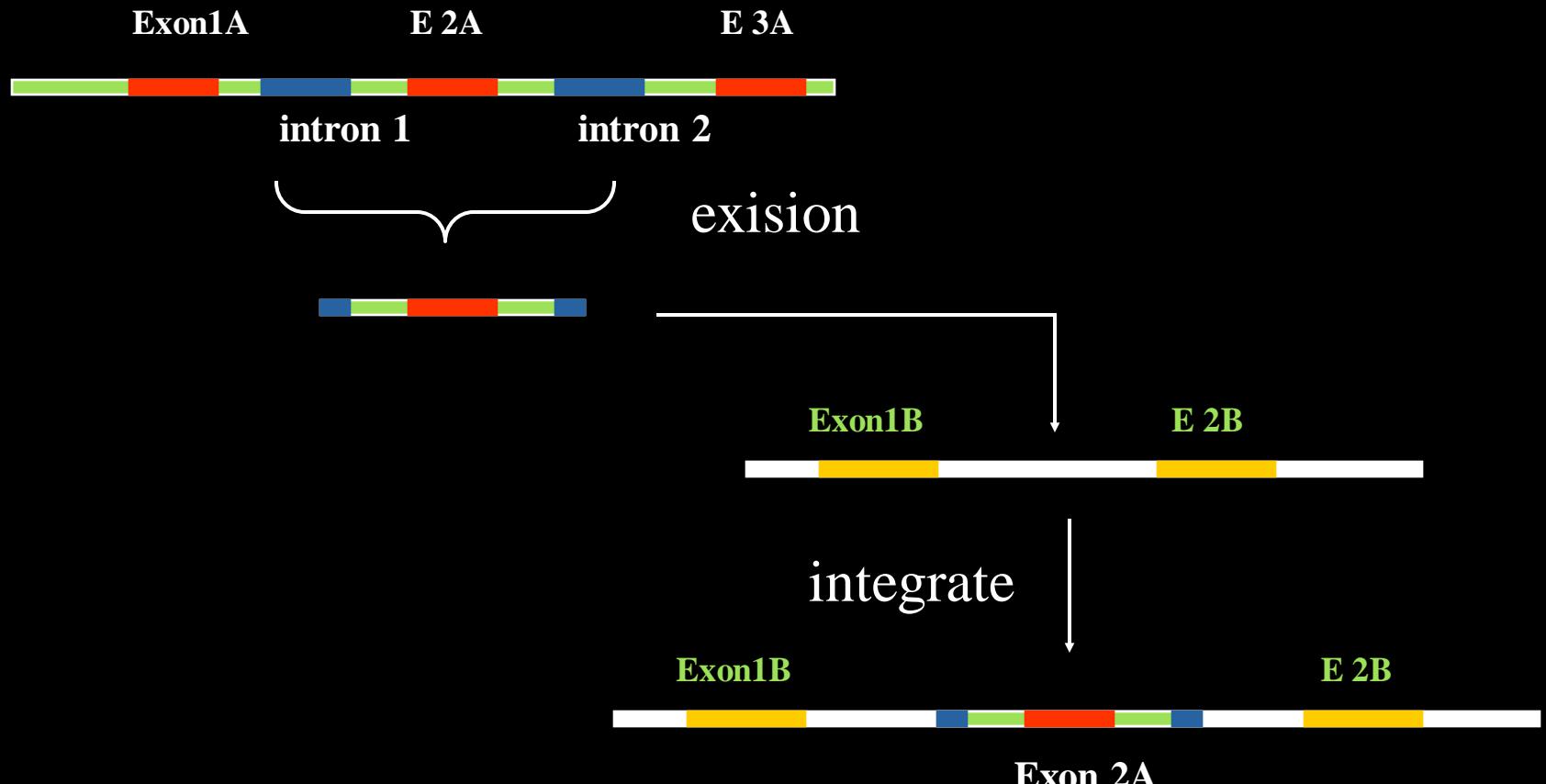


Mu-TATL-PCR



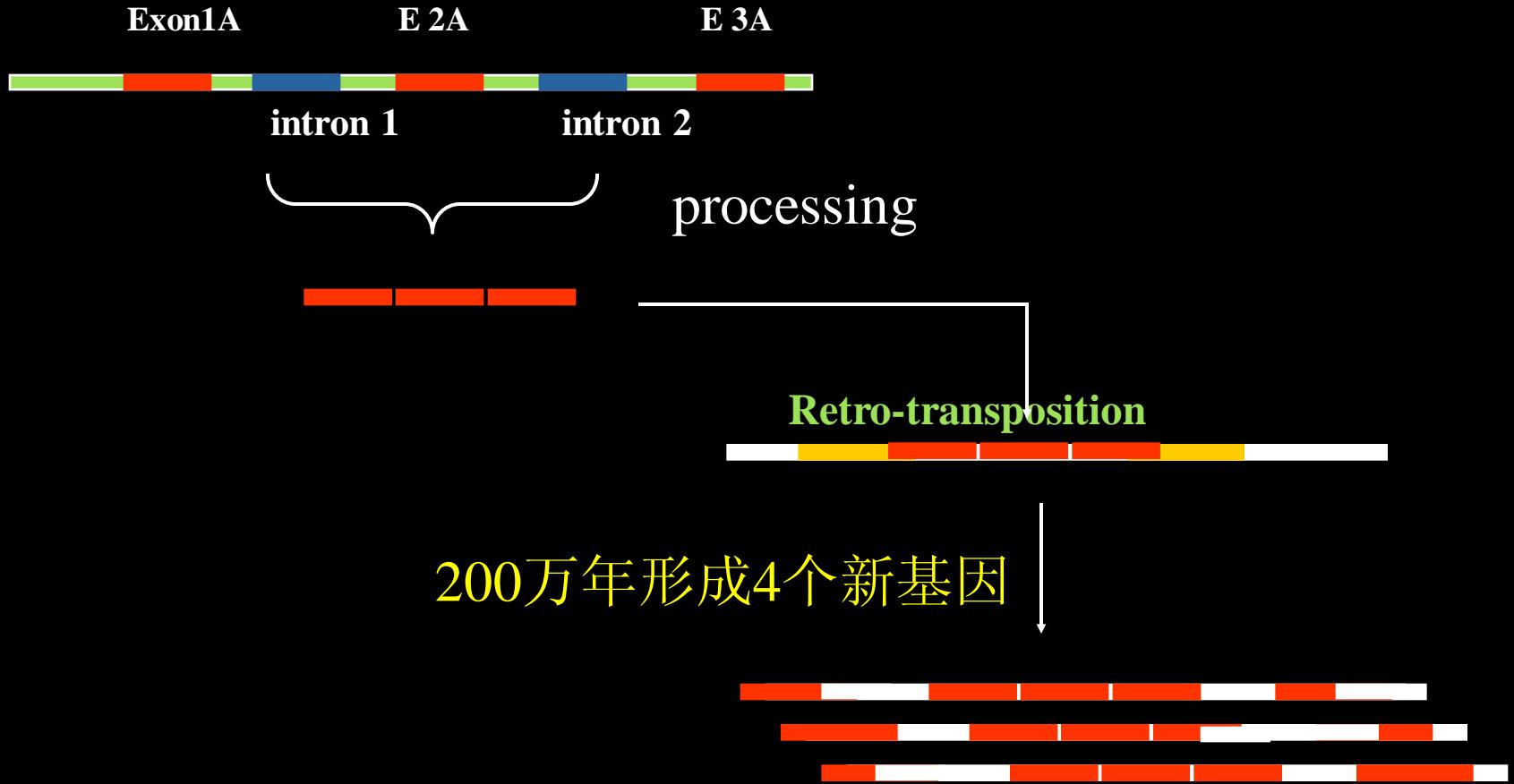
I II AD I II AD

外显子改组 exon shuffling形成新基因

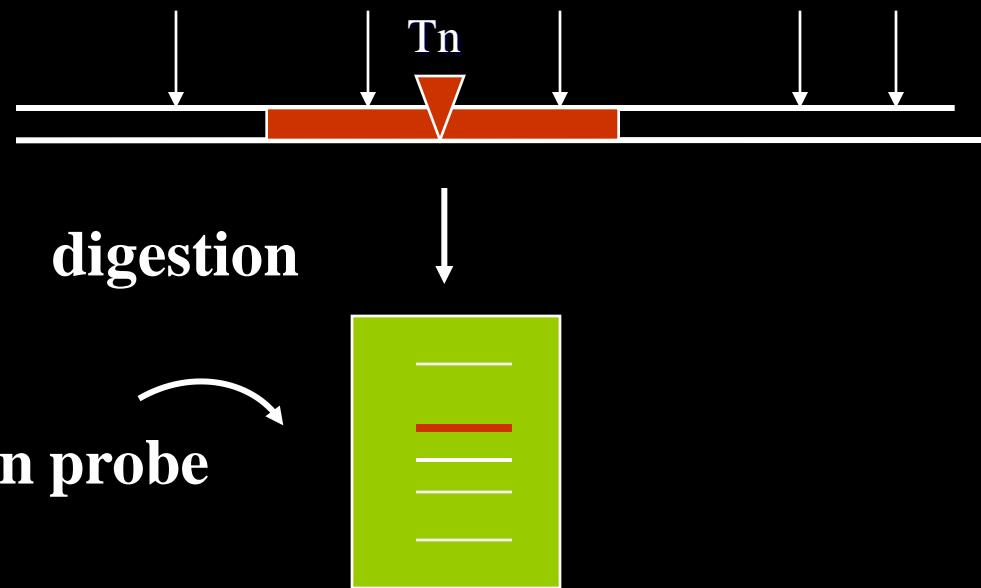


exon shuffling

Retro-transposon create chimeric new gene (jingwei基因, 猴王基因)



- 利用 Tn 进行基因定位



- 利用IR作为基因转移载体



2.6.7. 真核生物的假基因 (pseudo gene)

a) 概念

与正常基因结构相似

但丧失正常功能的DNA序列

往往存在于真核生物的多基因家族中

b) 种类

- 功能基因累积突变型

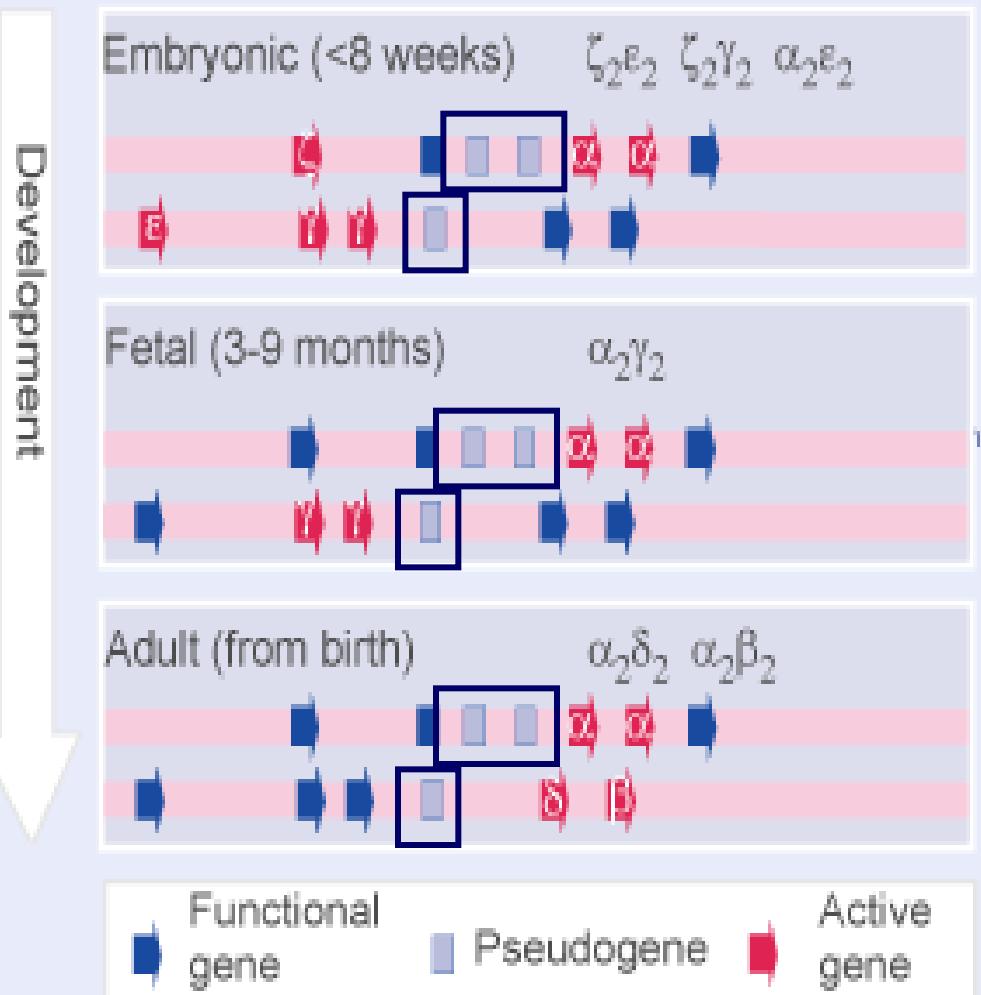
非洲爪蟾 5s rDNA family 人 α -globin $\psi\alpha-1$

3个缺失区段

多个移码突变

intron 及拖尾序列改变

Hemoglobin expression changes during development



Pseudogene in
Hemoglobin
family

- 加工基因 (processed gene, retrogene)

归类于非病毒类retro-transposon

1980. Leder 发现 rat α -globin, $\psi\alpha$ -3 基因中无 intron

1982. Temin 证实

chicken β -tubulin family 中存在 processed gene

正常基因-----promoter----GCTGA
codon-446

Surprising

加工基因
446codon-
GCTGAGGTGTC-lead seq.---ATG AGC---

正常基因----some introns----AATAAA----terminator

加工基因----no introns-----AATAAA----poly(A) GCTGAGGTGTC

c) 加工基因的形成

● DNA



pre-RNA



RNA



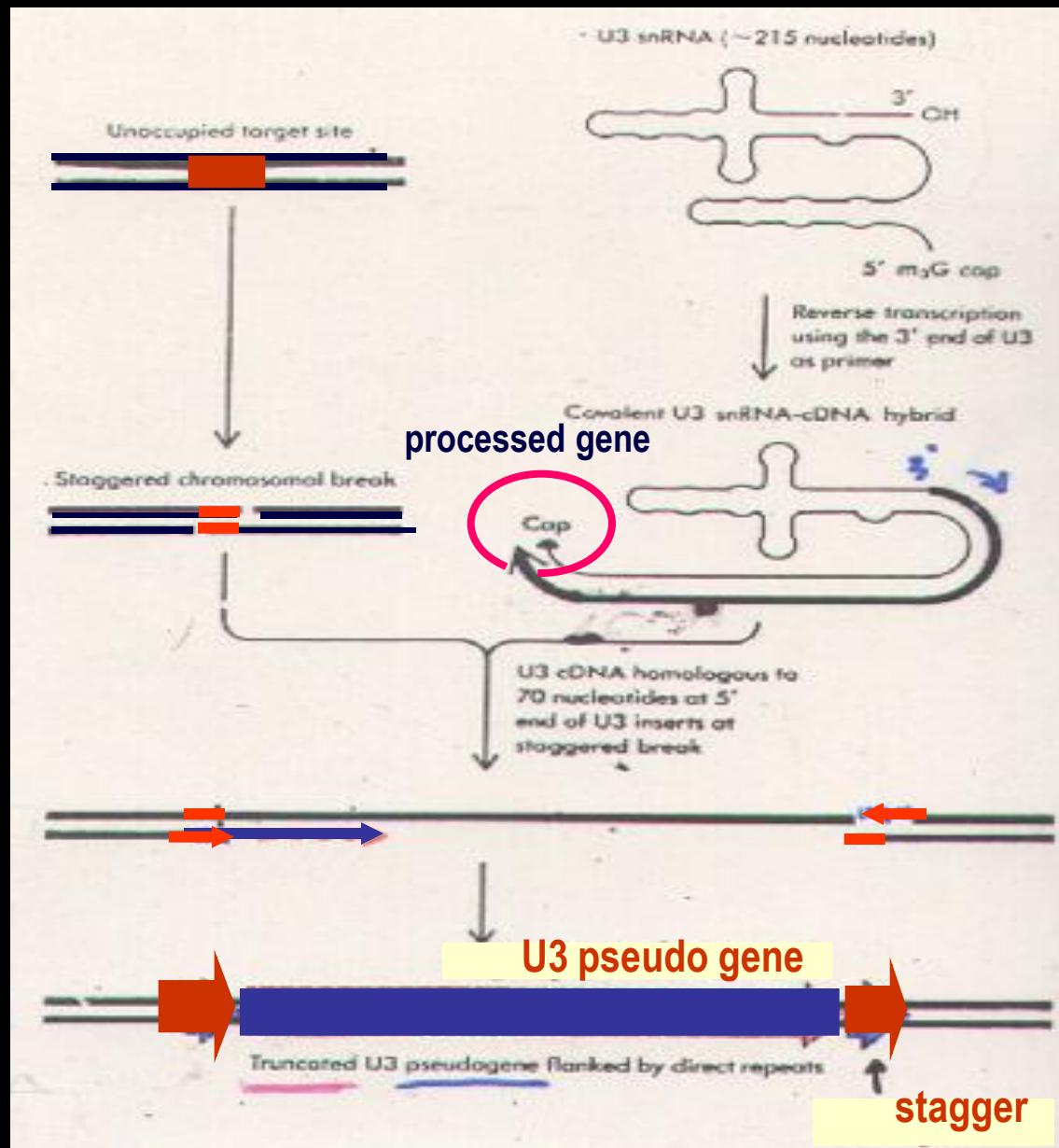
cDNA



Insertion



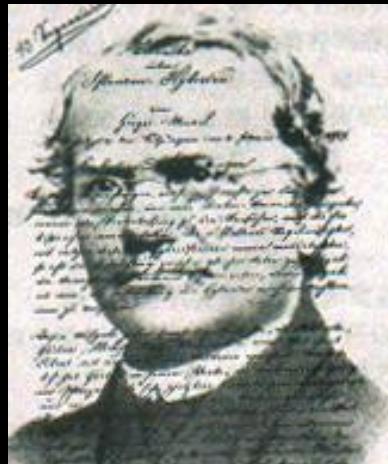
DNA



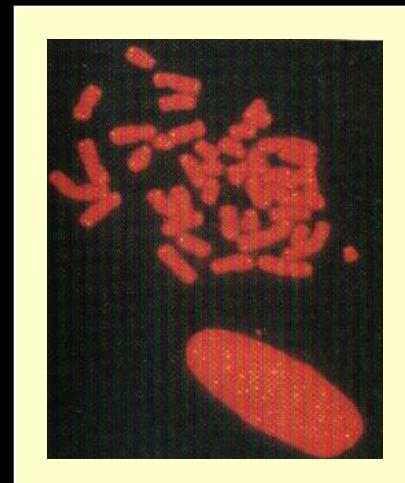
d) 遗传特征

- 由 RNAPol II 介导的 mRNA 引起
- 遗传稳定
- 进化的整体性，无选择压力
- 数目大，占活性基因数的1%（人类，鼠）
- dead end of evolution ?

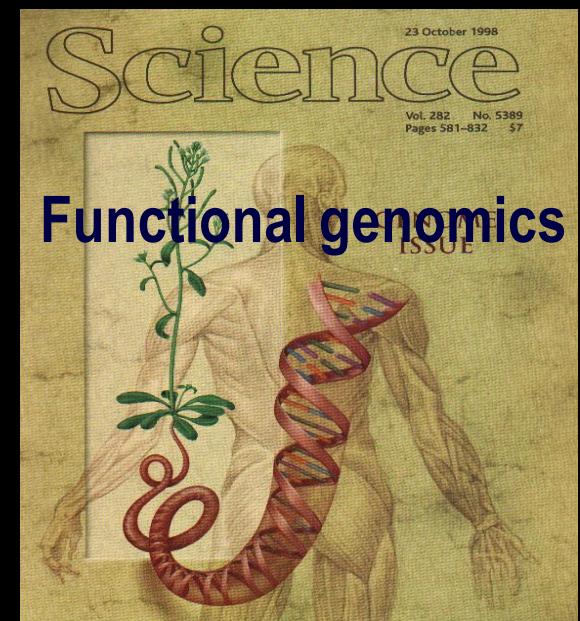
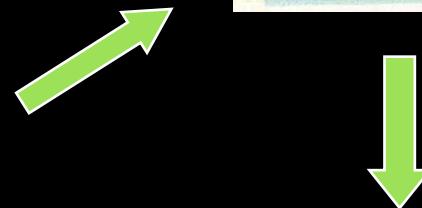
基因由“唯DNA”的观
点趋向于“基因是
DNA与蛋白质的结合
体”



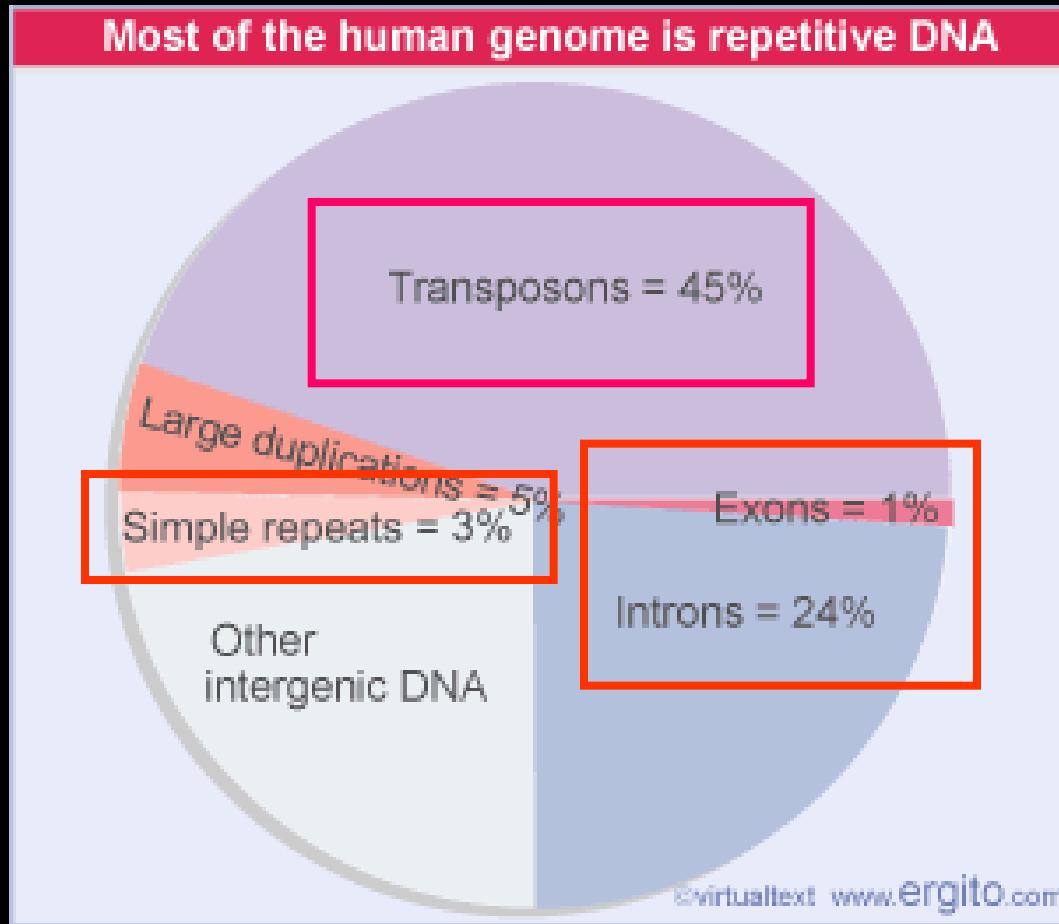
Genetics



Genomics

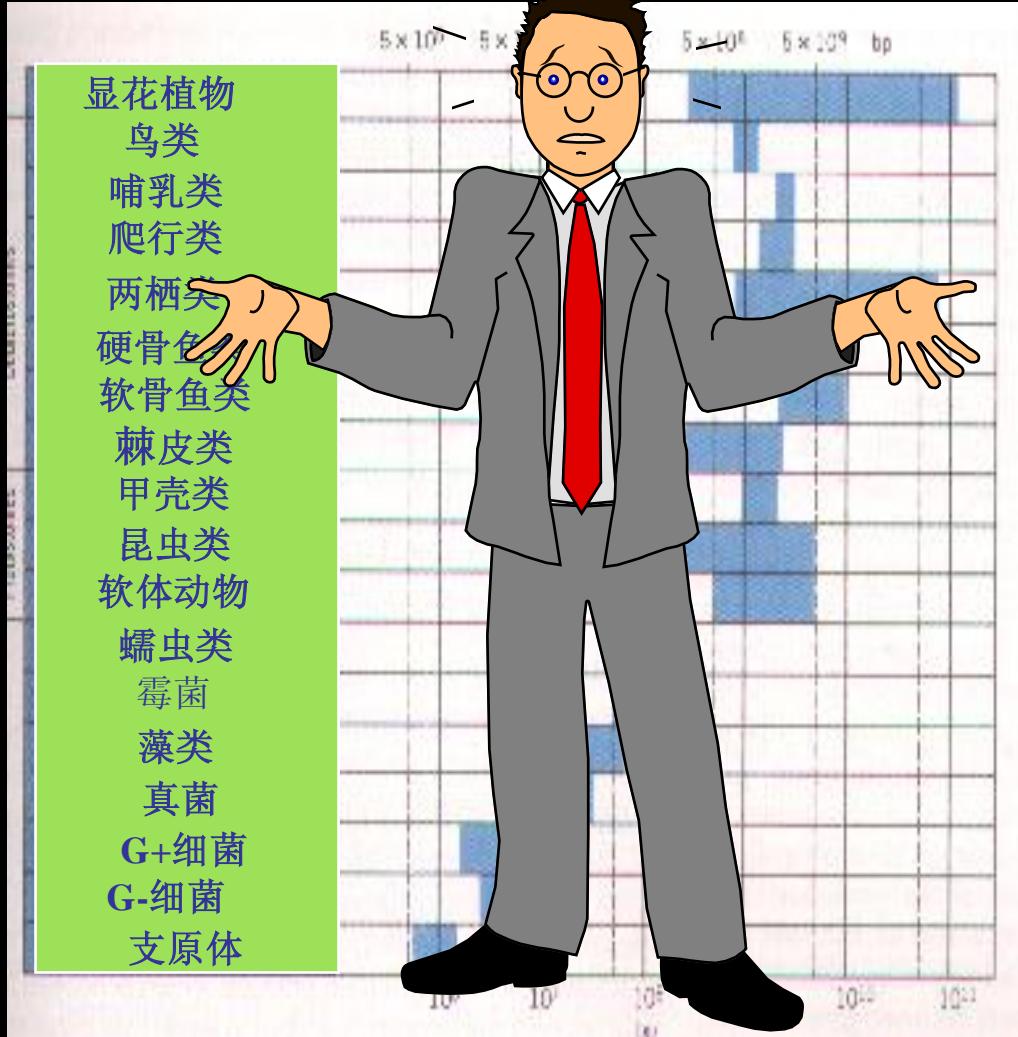


Most of the human genome is repetitive DNA



The largest component of the human genome consists of transposons

N value paradox of nucleotide



基因数目与生物进化程度和生物复杂性之间的不对称现象成为

N值矛盾