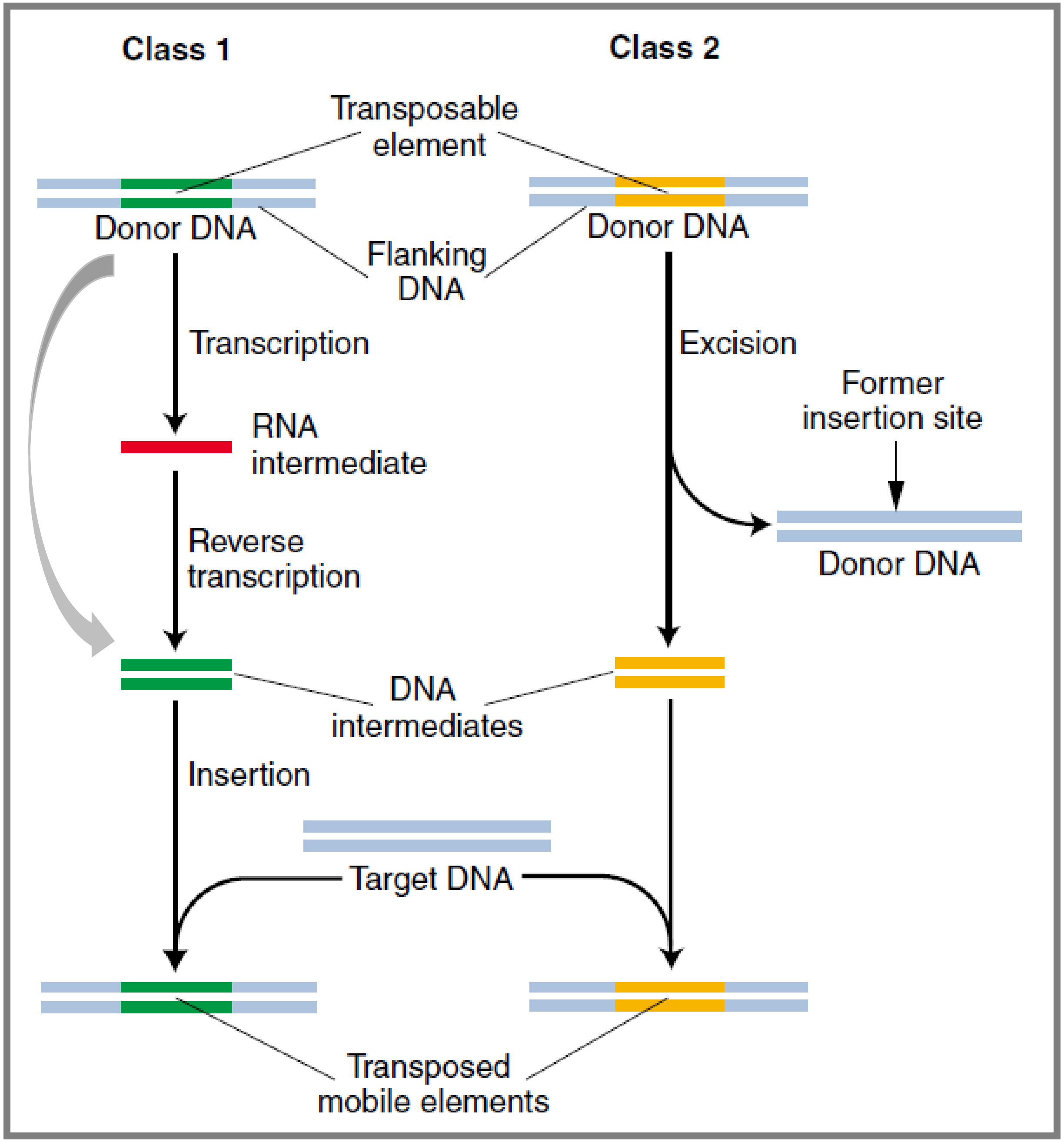
第五章　转座子和逆转座子
Transposons and retrotransposons

- **Transposable Elements (transposons)**

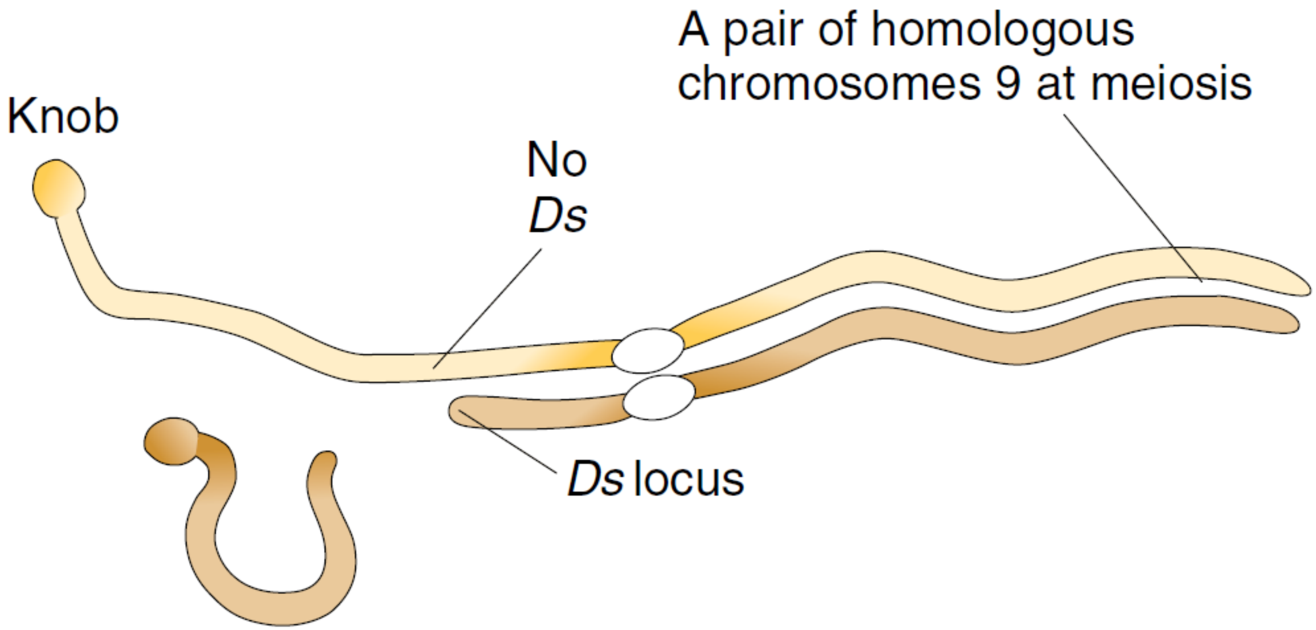
A DNA sequence capable of moving (transposing) from one location to another in a genome.



Discovery of Transposable Elements

- Barbara McClintock (1902-1992)
- 1940s ~ 1950s

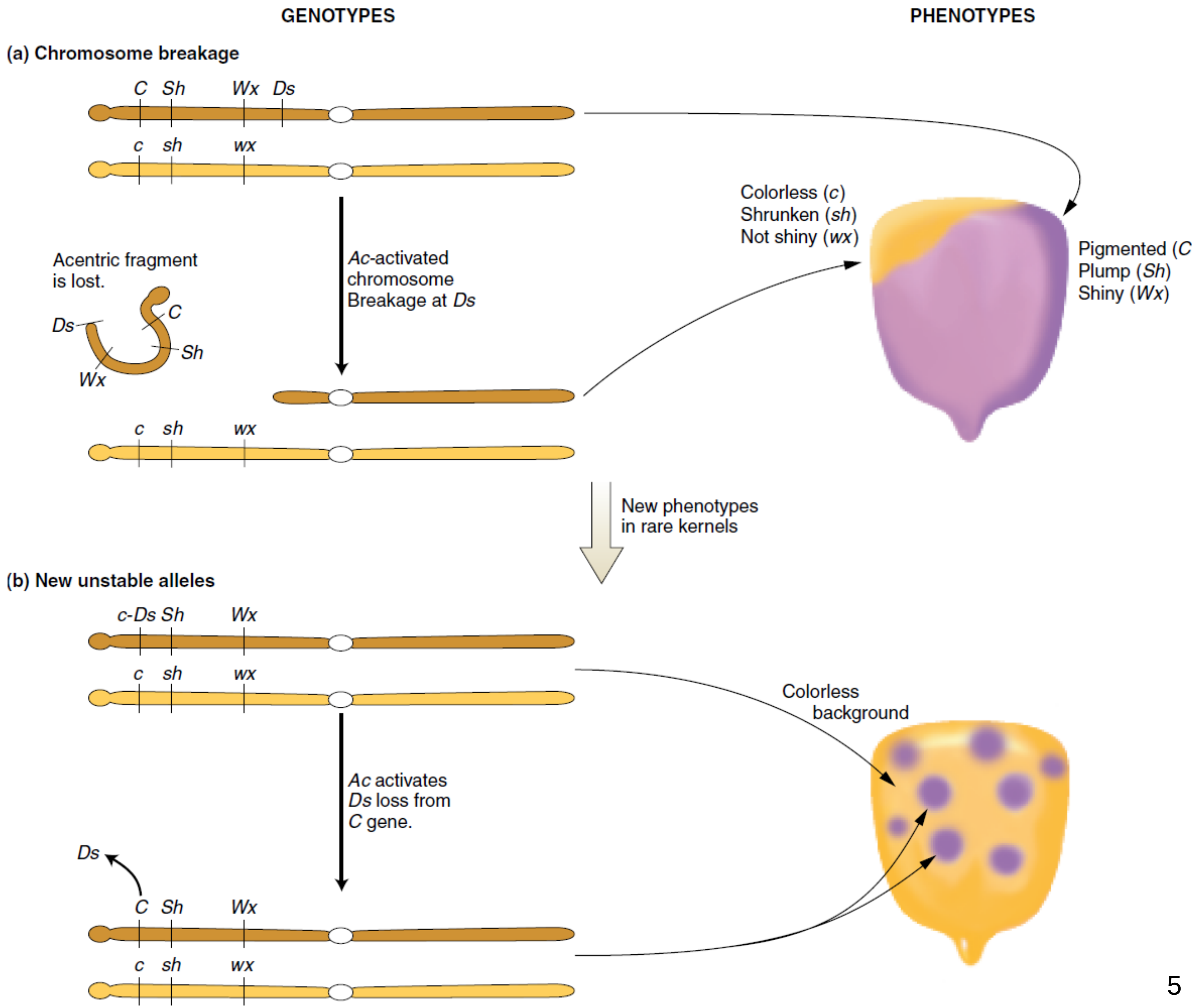
Indian corn, 10 chromosomes numbered from largest (1) to smallest (10)



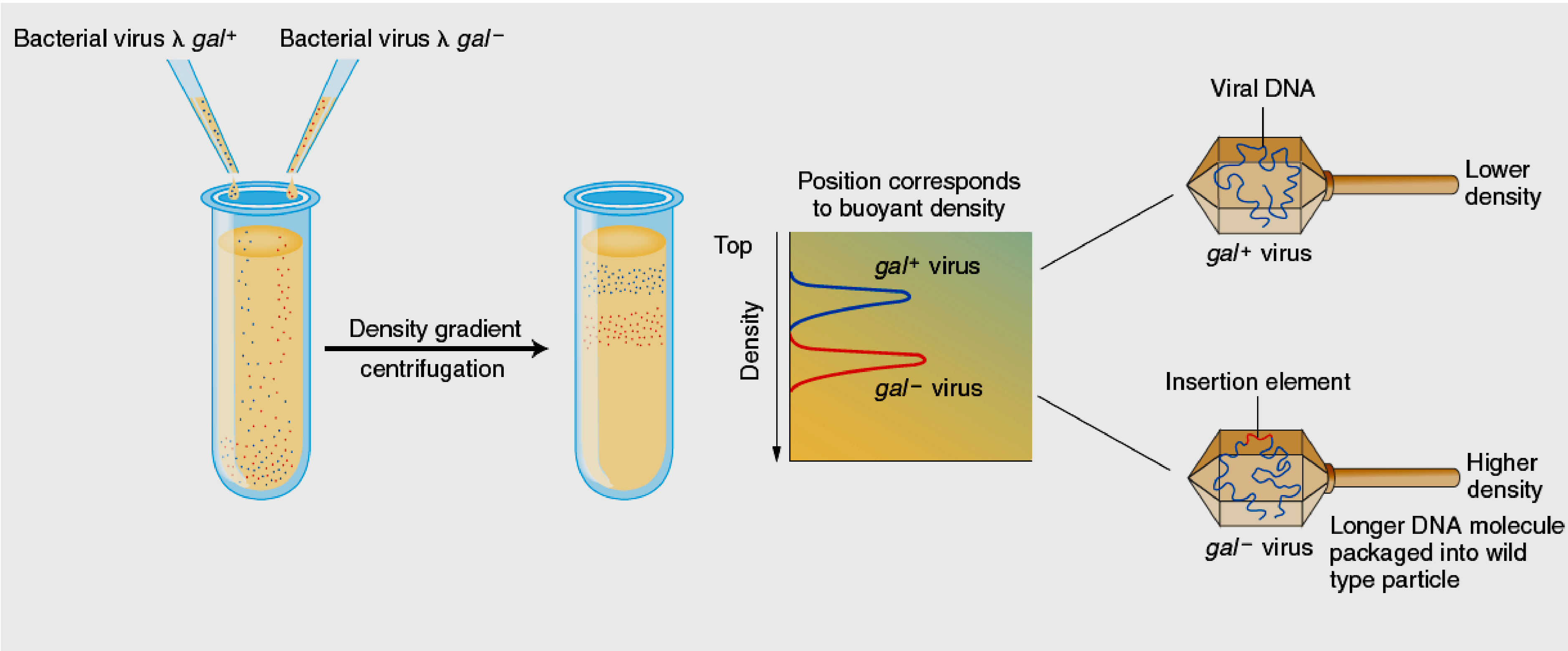
Ds (for ***Dissociation***) located at the site of the break

Ac (for ***Activator***) unlinked genetic factor, present in some maize stocks but absent in others

McClintock found it is impossible to map ***Ac***.


GENOTYPES
PHENOTYPES
(a) Chromosome breakage
C Sh Wx Ds
c sh wx
Acentric fragment is lost.
Ds
C
Sh
Wx
Ac-activated chromosome Breakage at Ds
c sh wx
Colorless (c)
Shrunken (sh)
Not shiny (wx)
Pigmented (C
Plump (Sh)
Shiny (Wx)
New phenotypes in rare kernels
(b) New unstable alleles
c-Ds Sh Wx
c sh wx
Ac activates Ds loss from C gene.
Ds
C Sh Wx
c sh wx
Colorless background

■ Transposable element in *E. coli* , 1960



McClintock received the **Nobel Prize** in 1983

第一节 细菌的转座因子 Transposable Elements in Bacteria

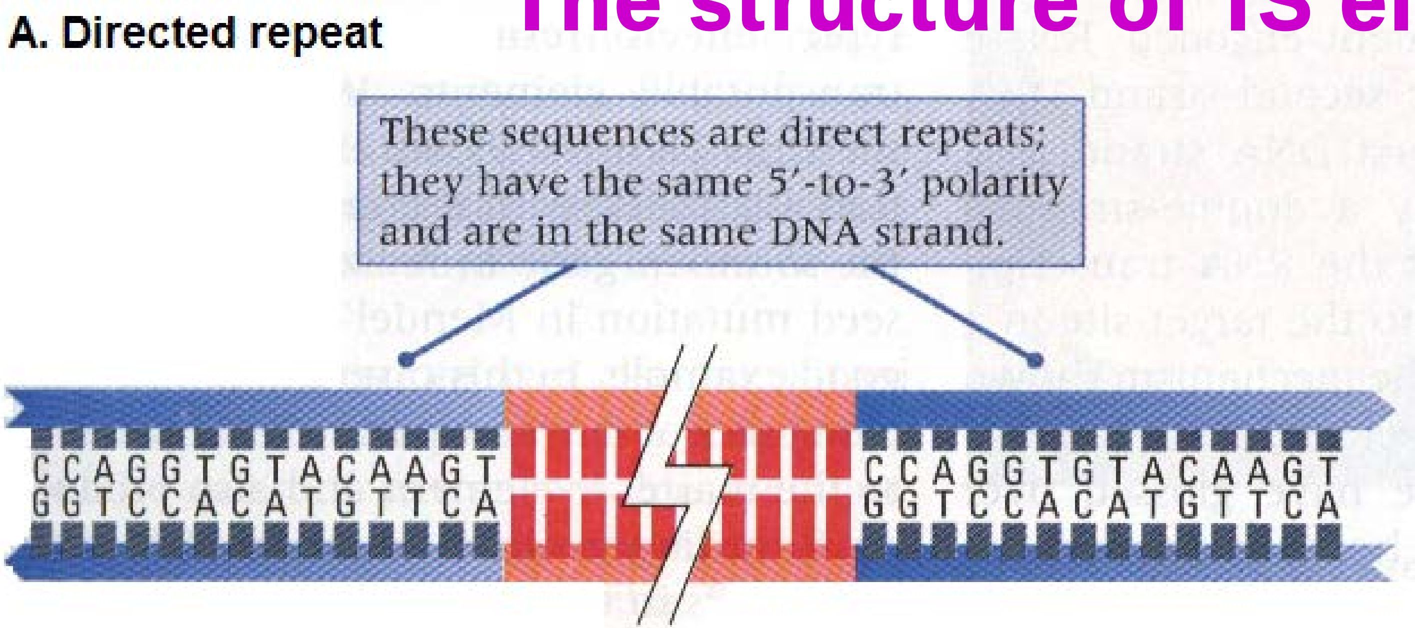
Two major groups:

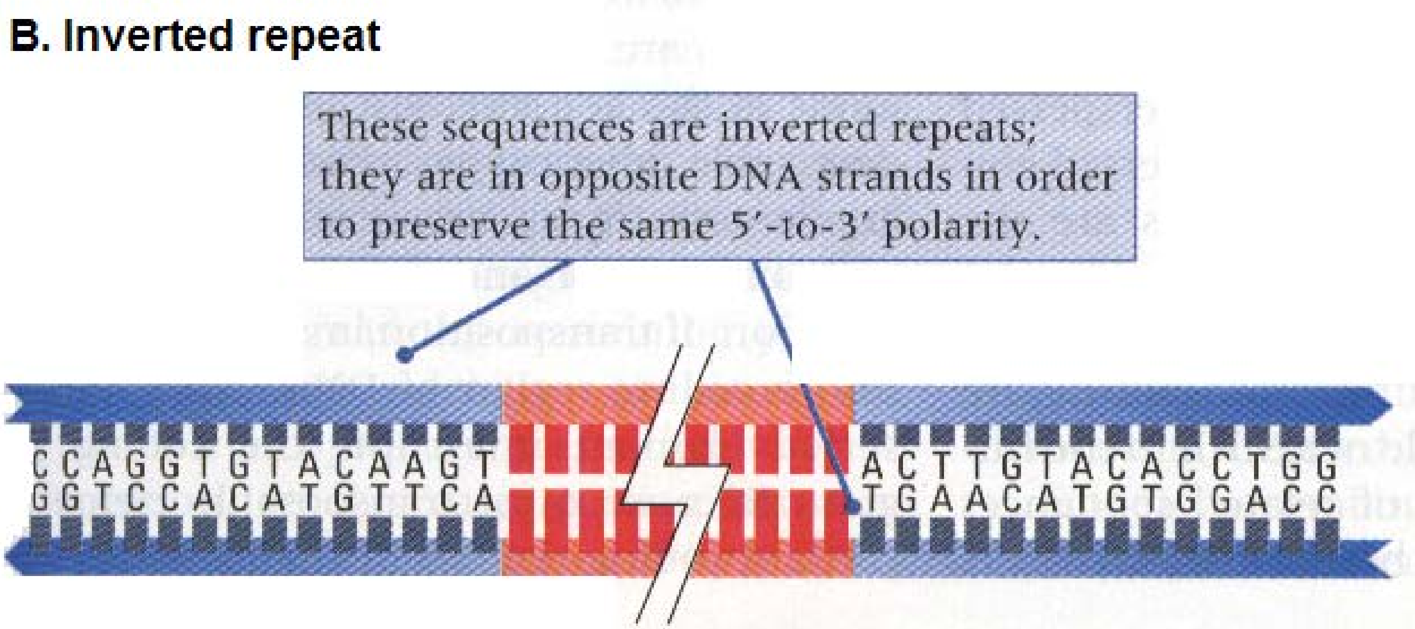
- Simple transposable elements which carry only the information required for movement
- Complex transposable elements which contain DNA sequences not directly related to transposition

1. Insertion sequences (IS) (插入序列)

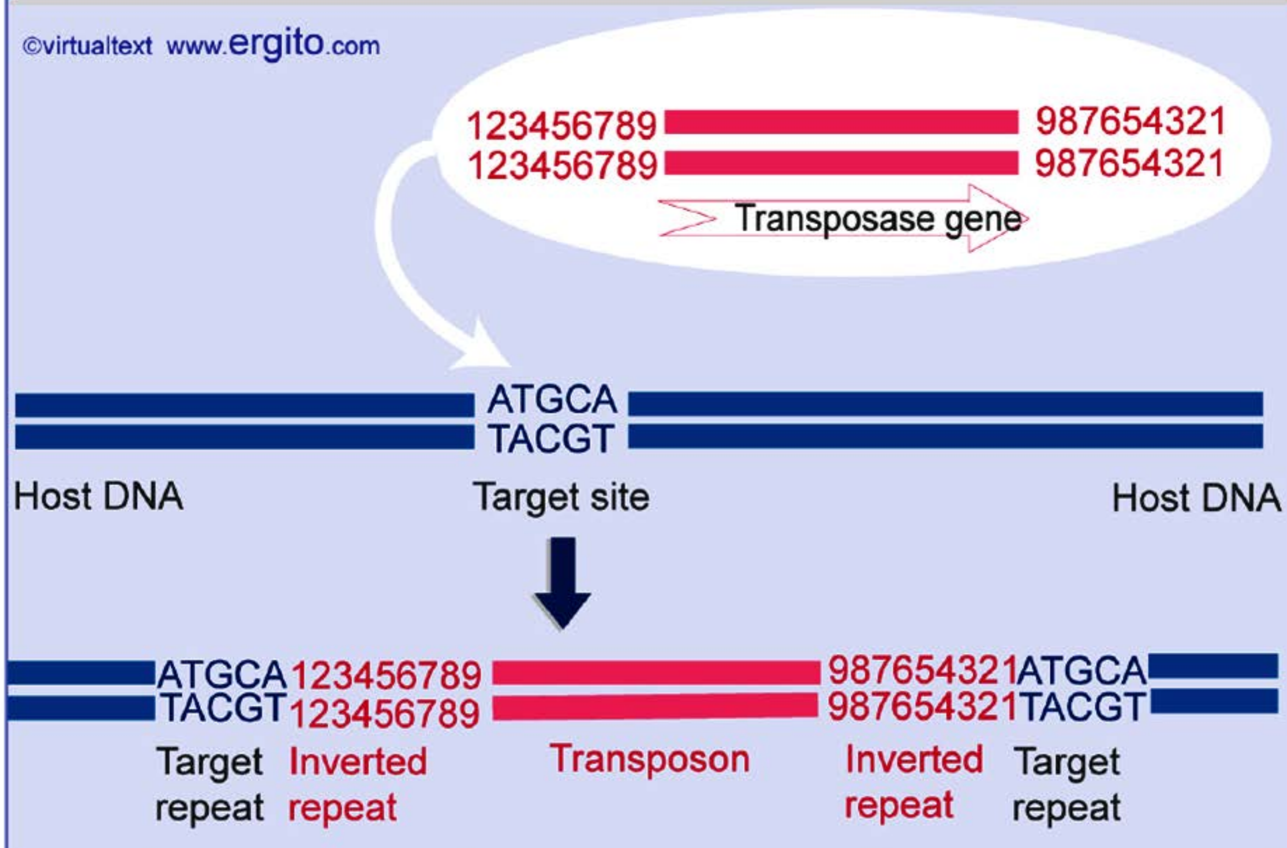
Insertion sequences (IS) are the simplest transposons

The structure of IS element





Transposons have inverted repeats and generate target repeats

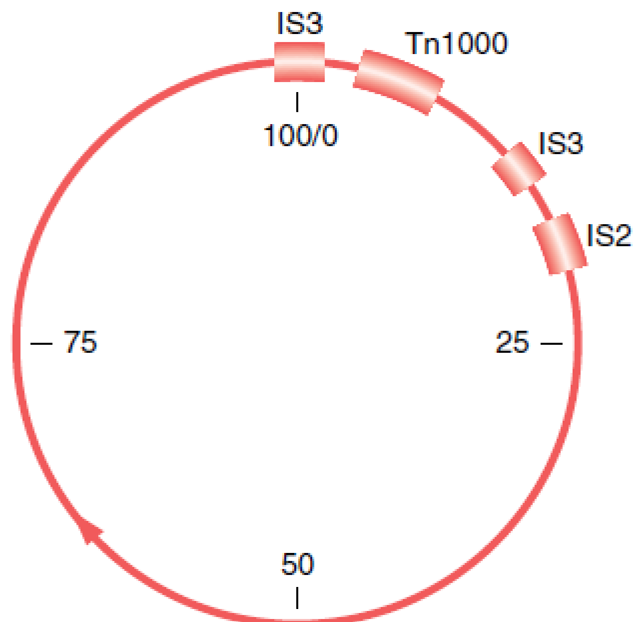


			Overall length	Target selection
IS1	9 bp	23 bp	768 bp	random
IS2	5 bp	41 bp	1327 bp	hotspots
IS4	11-13 bp	18 bp	1428 bp	$AAAN_{20}TTT$
IS5	4 bp	16 bp	1195 bp	hotspots
IS10R	9 bp	22 bp	1329 bp	NGCTNAGCN
IS50R	9 bp	9 bp	1531 bp	hotspots
IS903	9 bp	18 bp	1057 bp	random

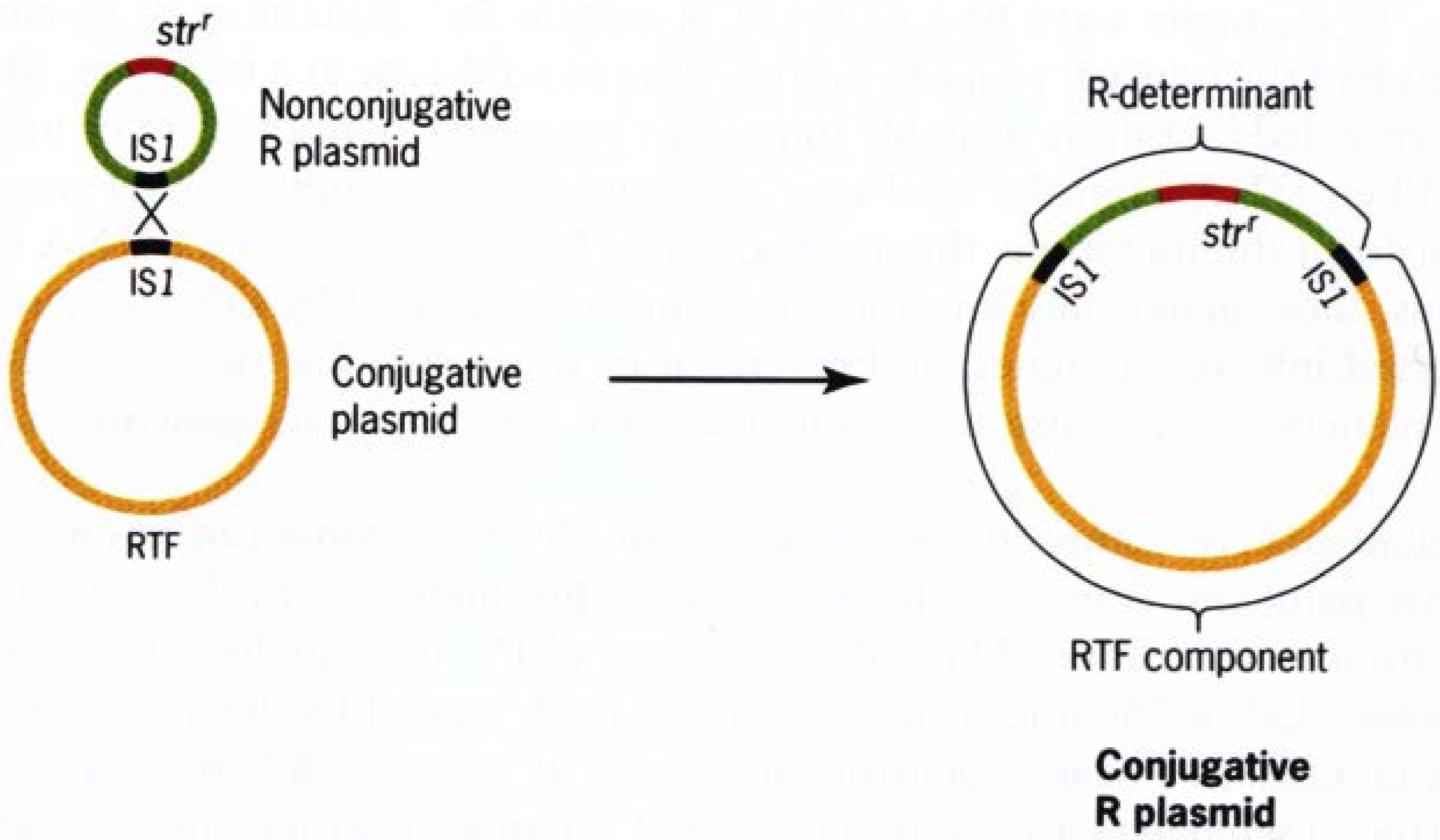
- An IS element ends in short *inverted terminal repeats* ; usually the two copies of the repeat are closely related rather than identical.
- IS elements code for transposase

Insertion sequence	Normal occurrence in *E. coli*
IS1	5–8 copies on chromosome
IS2	5 copies on chromosome; 1 copy on F
IS3	5 copies on chromosome; 2 copies on F
IS4	1 or 2 copies on chromosome
IS5	Unknown

The IS elements are normal constituents of bacterial chromosomes and plasmids.



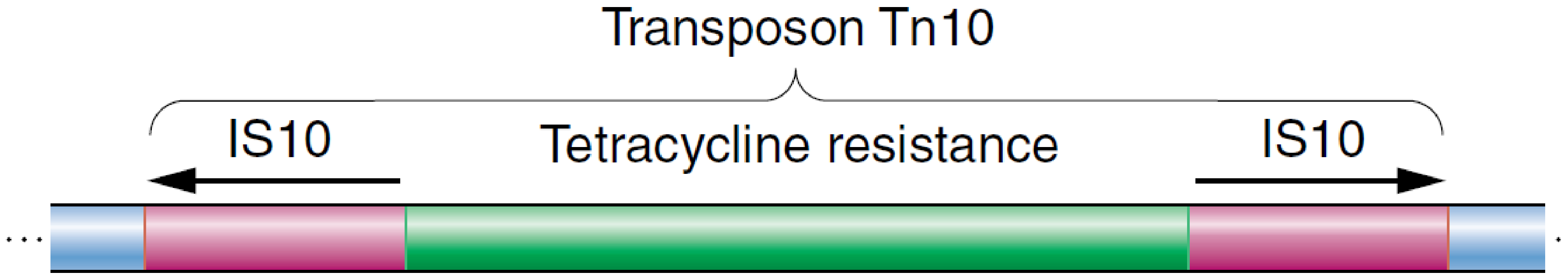
Formation of a conjugative R plasmid by recombination between IS elements

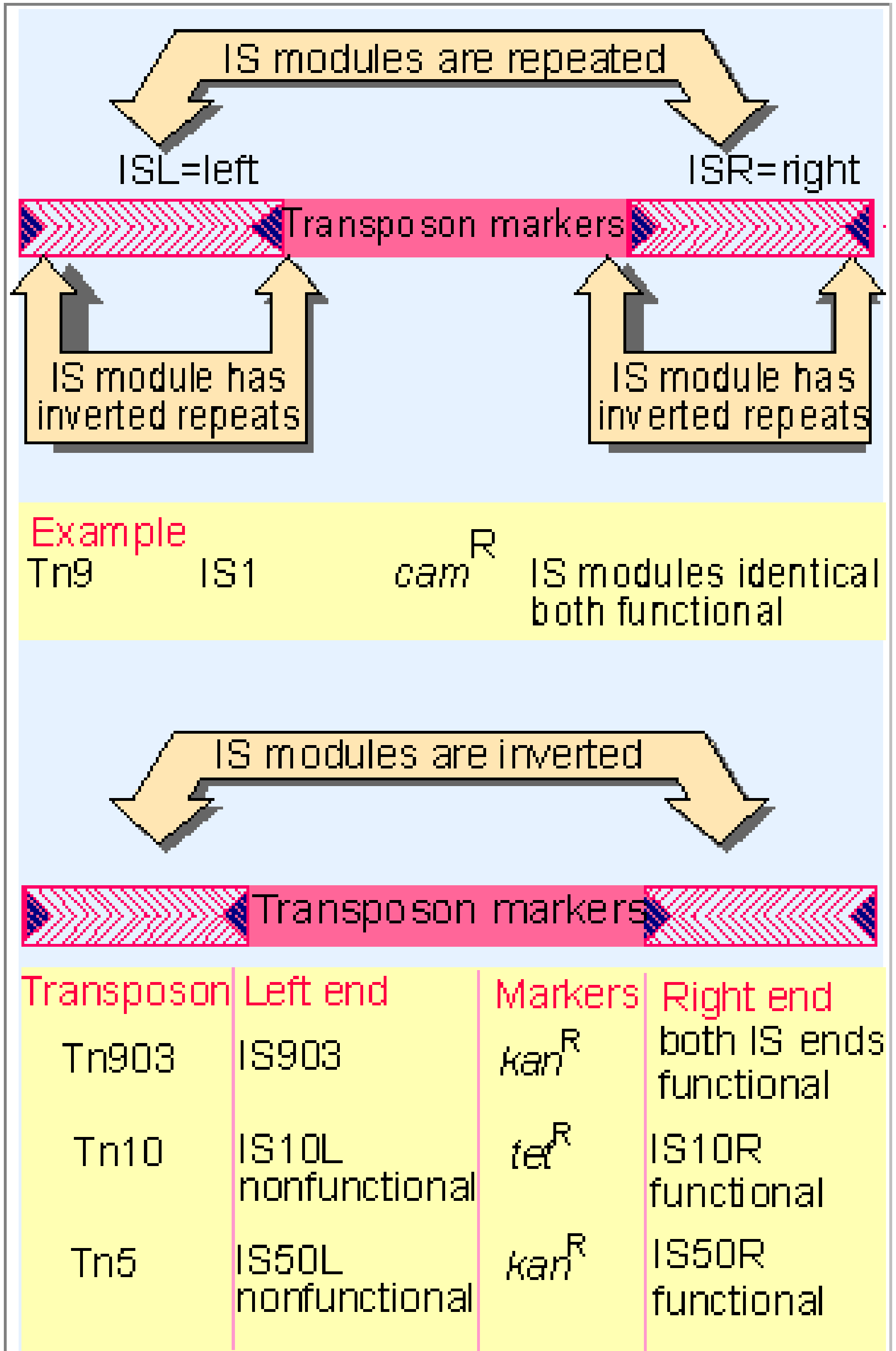


2. Composite Transposons

Composite transposons contain a variety of genes that reside between two nearly identical **IS elements**

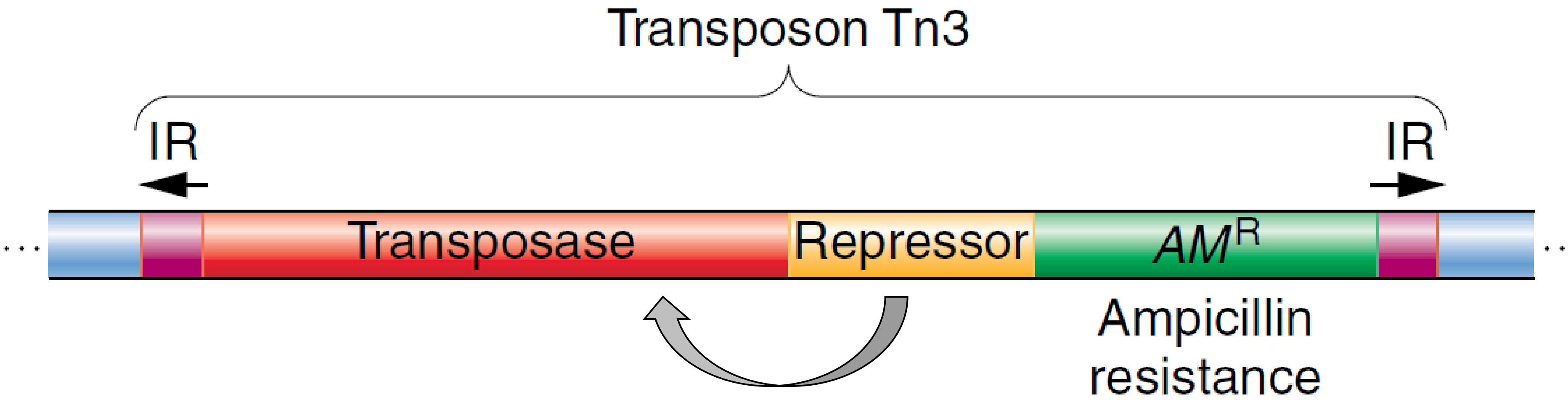
Composite transposon is designated by **Tn,** followed by a number



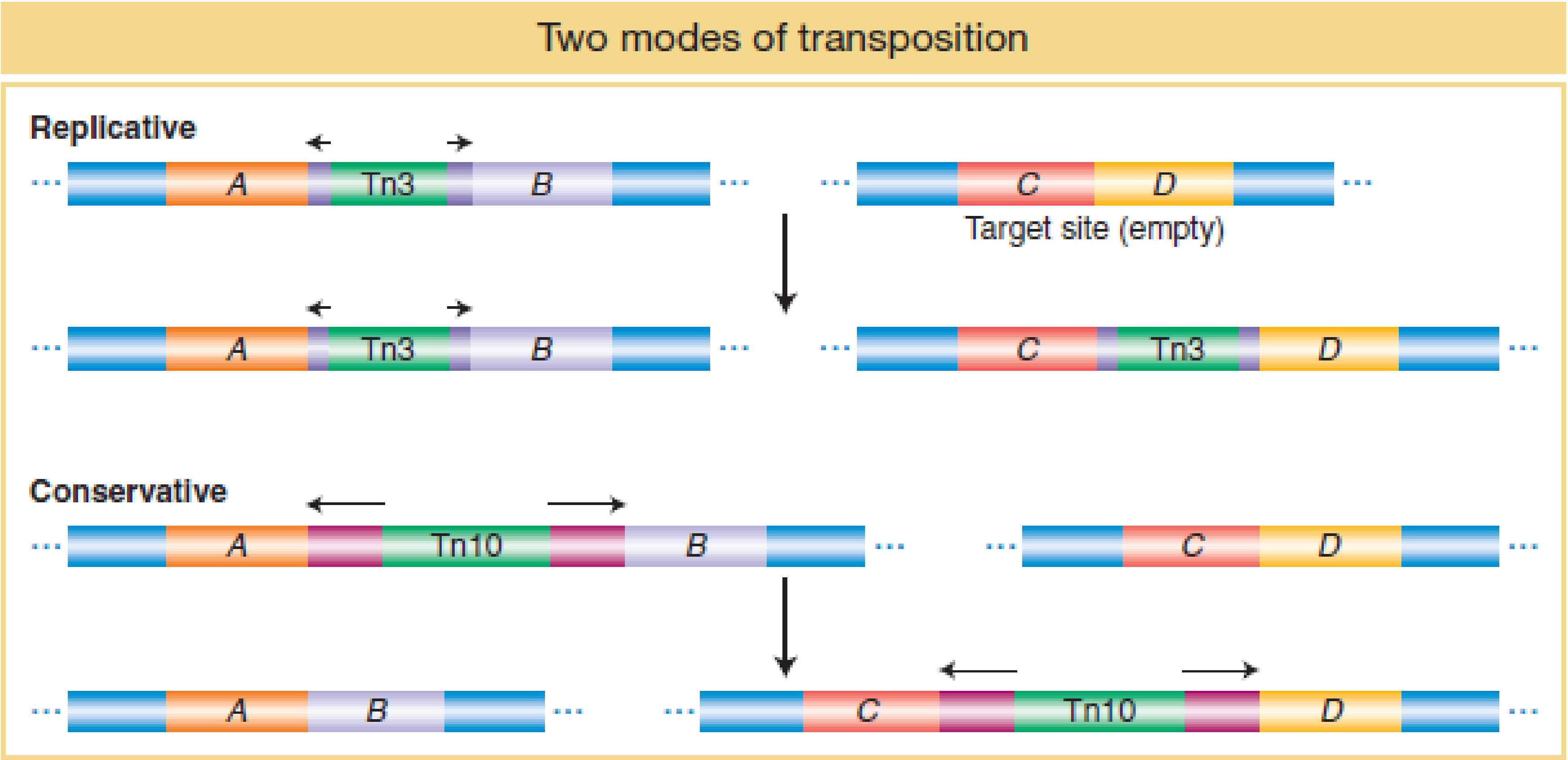


- In some cases, IS modules are identical. In other cases, the IS modules are closely related, but not identical.
- A functional IS module can transpose either itself or the entire transposon.
- When the two IS modules are identical, either module can sponsor the transposition.
- When the modules are different, transposition can depend on one of the modules

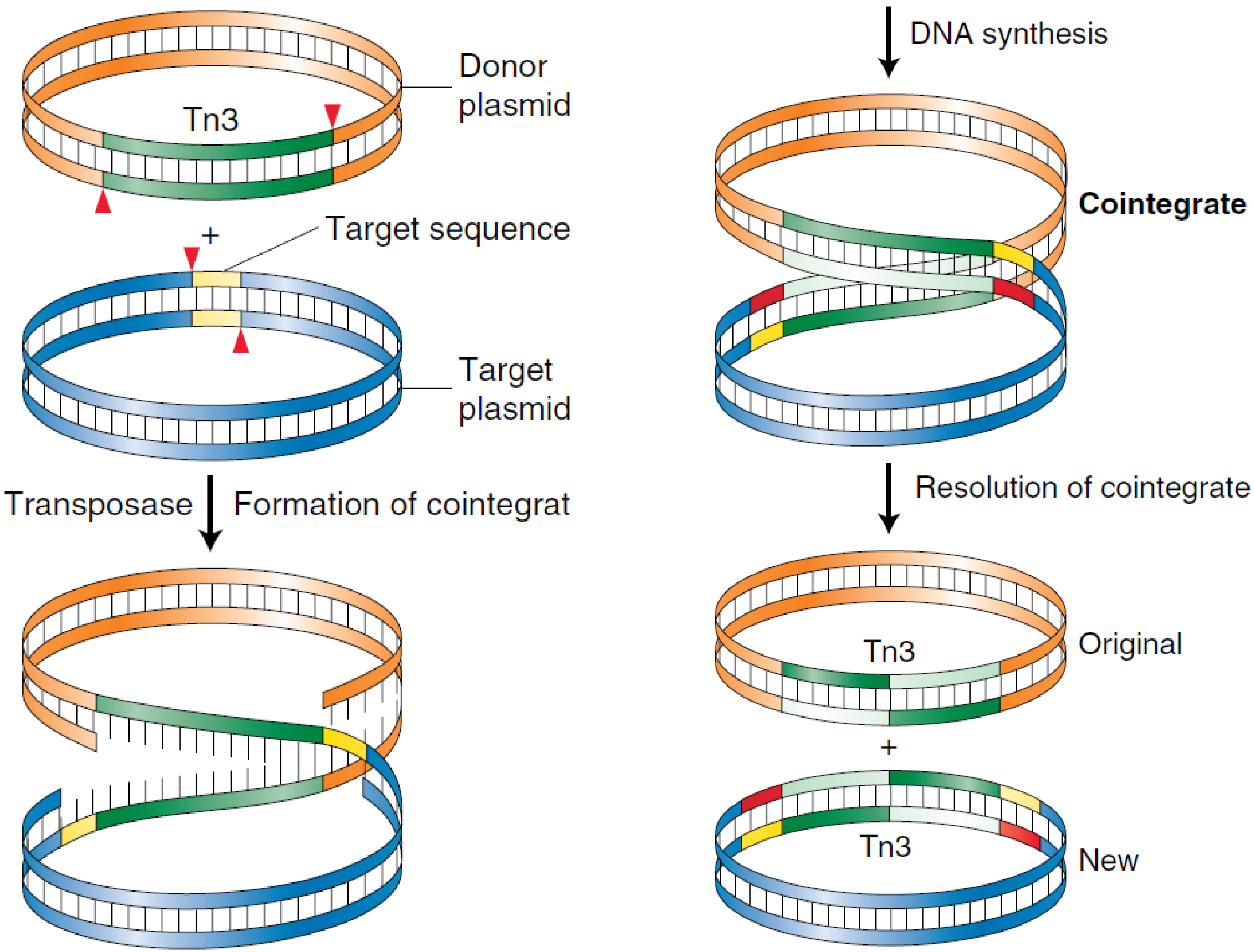
3. Noncomposite / Simple transposons



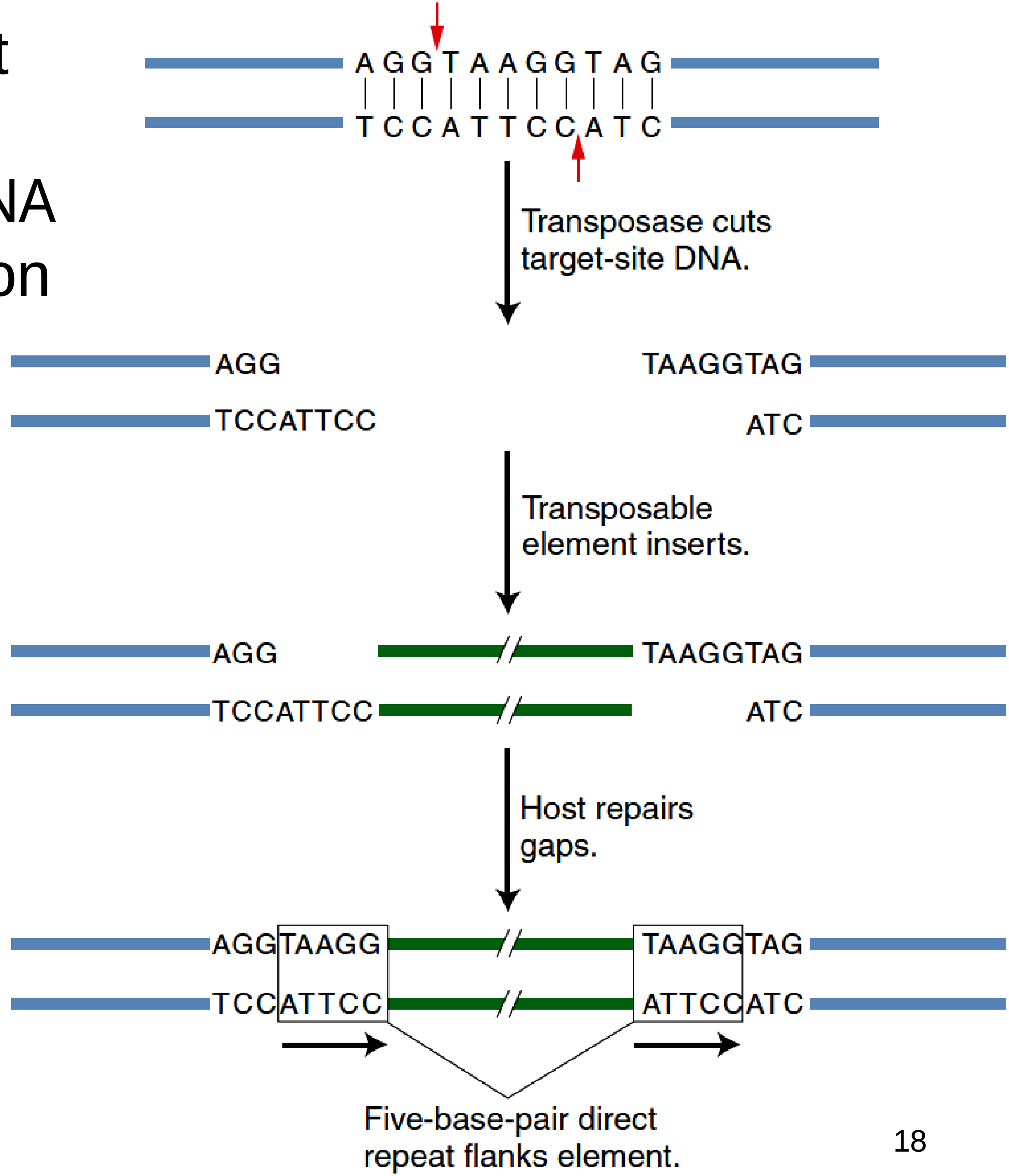
4. Mechanism of transposition



Replicative transposition of Tn3



When an IS element transposes, a sequence of host DNA at the site of insertion is duplicated.



第二节 真核生物的转座因子
Transposable Elements in Eukaryotes

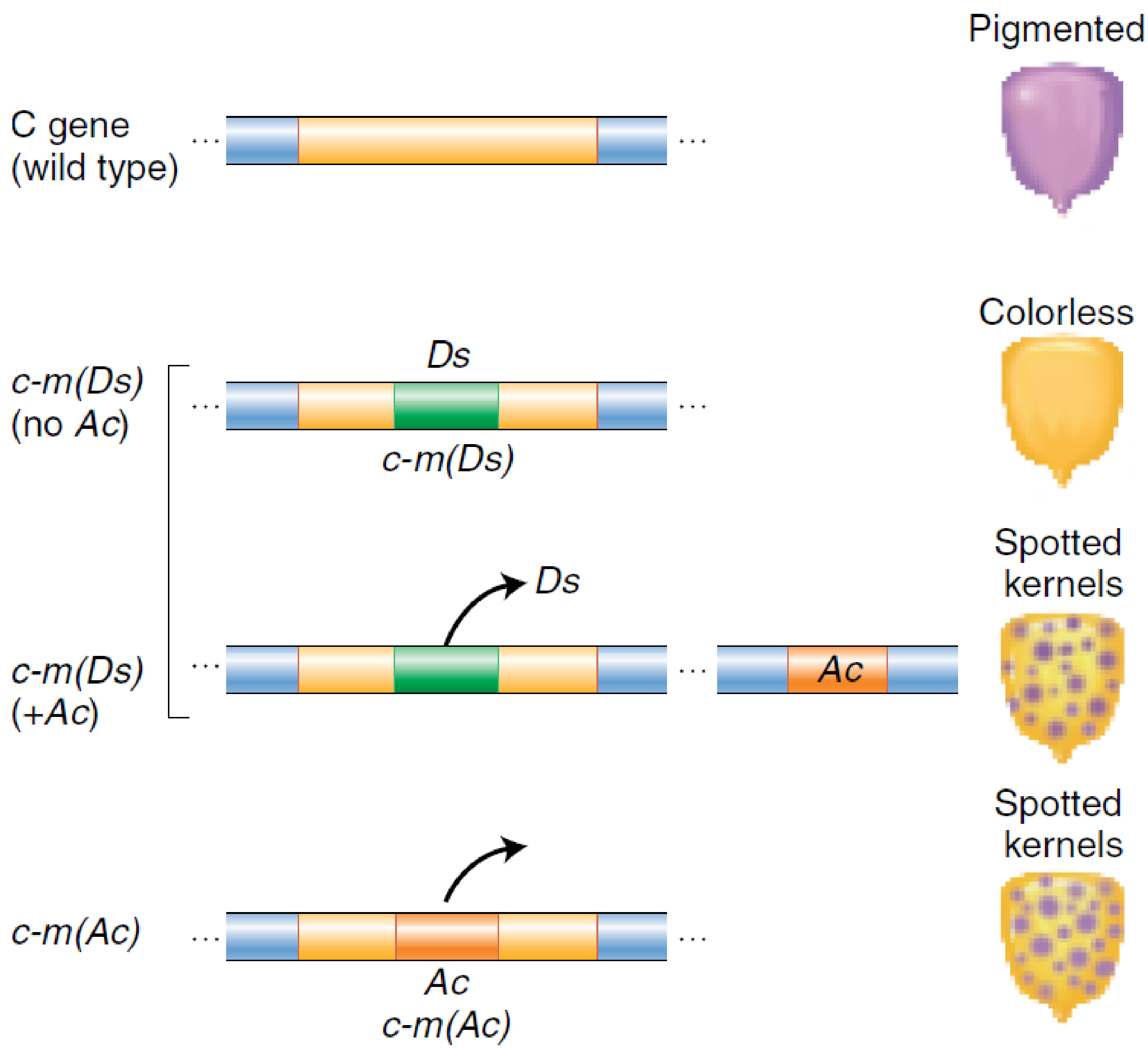
- DNA transposons (cut and paste transposons)
  - ***Ac*** and ***Ds elements*** in maize
  - ***P elements*** in *Drosophila*
- Retrotransposons

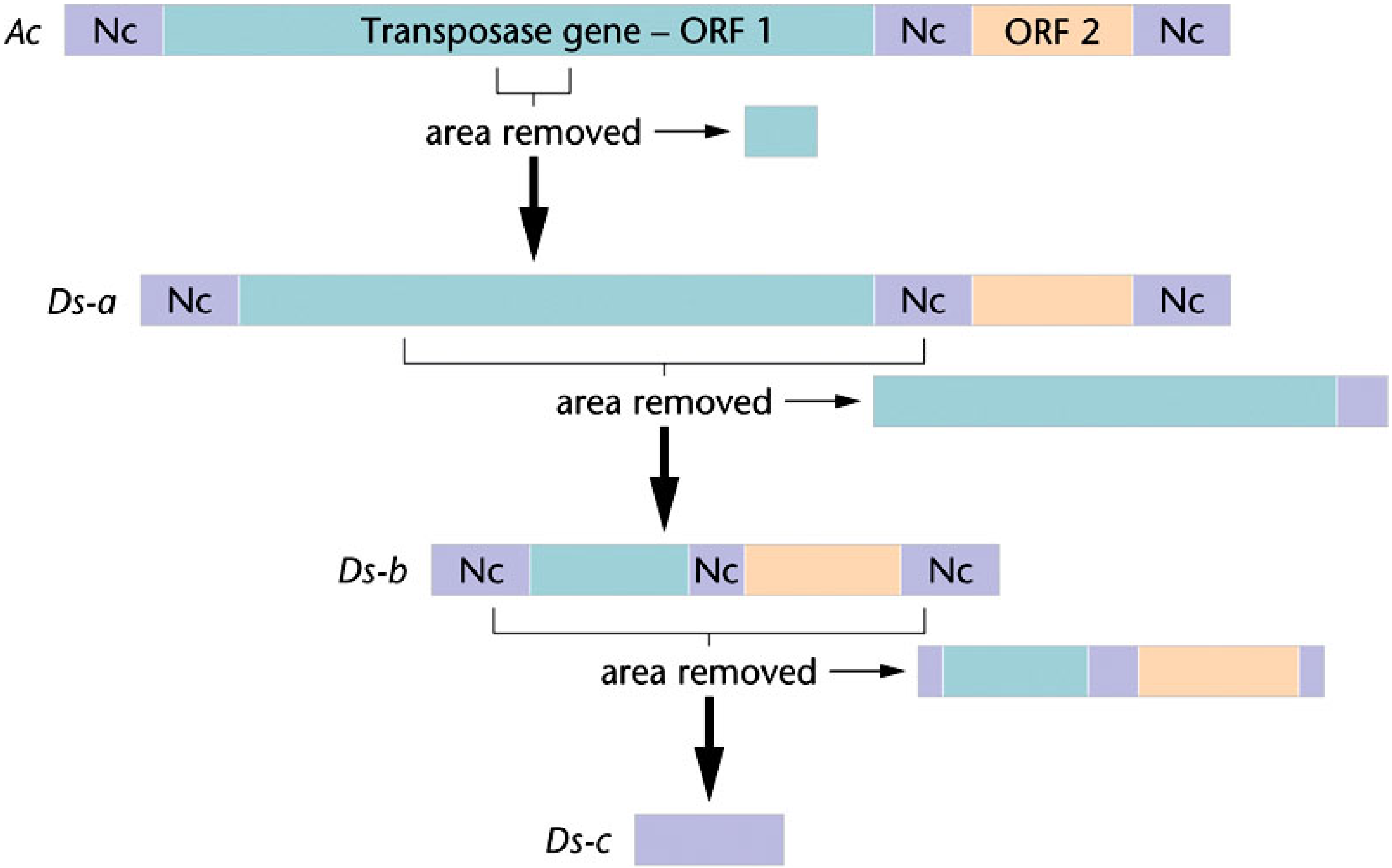
1. The Ac-Ds system in maize
（玉米的Ac-Ds 转座系统）

In absence of *Ac*, *Ds* is not transposable

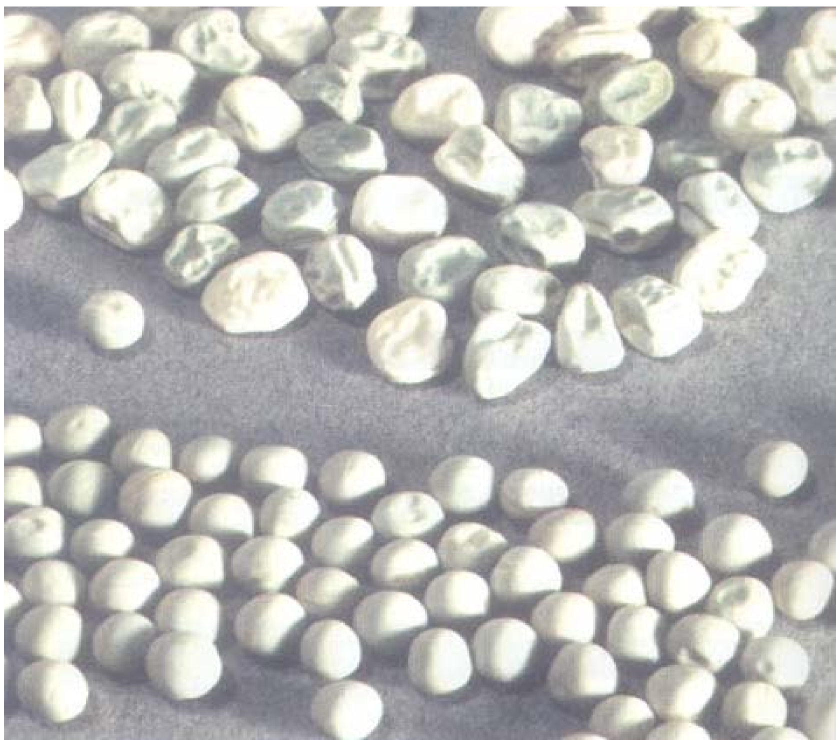
When *Ac* is present, *Ds* may be transposed

See movie




Ac
Nc
Transposase gene – ORF 1
Nc
ORF 2
Nc
area removed
Ds-a
Nc
Nc
Nc
area removed
Ds-b
Nc
Nc
Nc
area removed
Ds-c

Transposable element in other plants

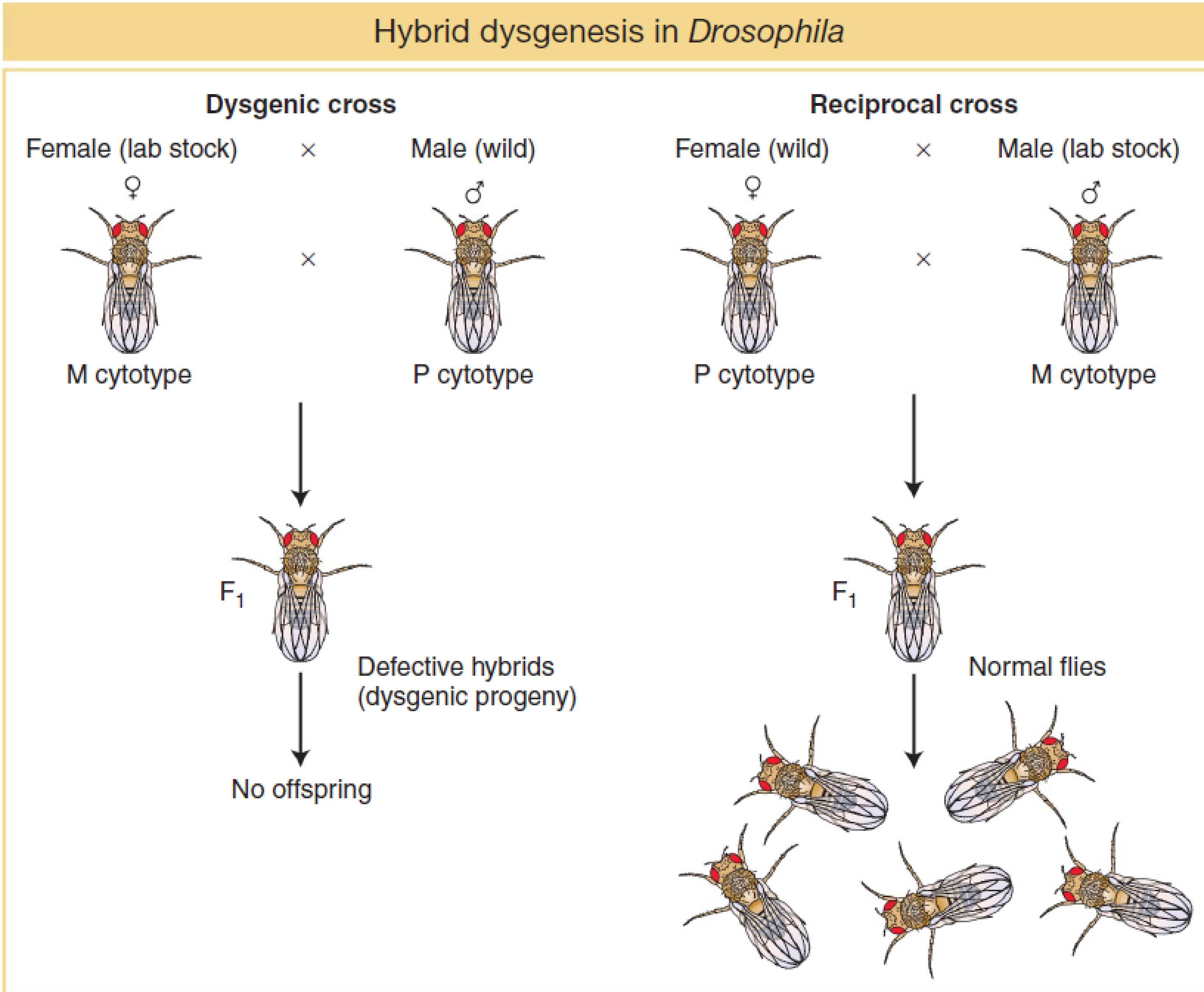


peas

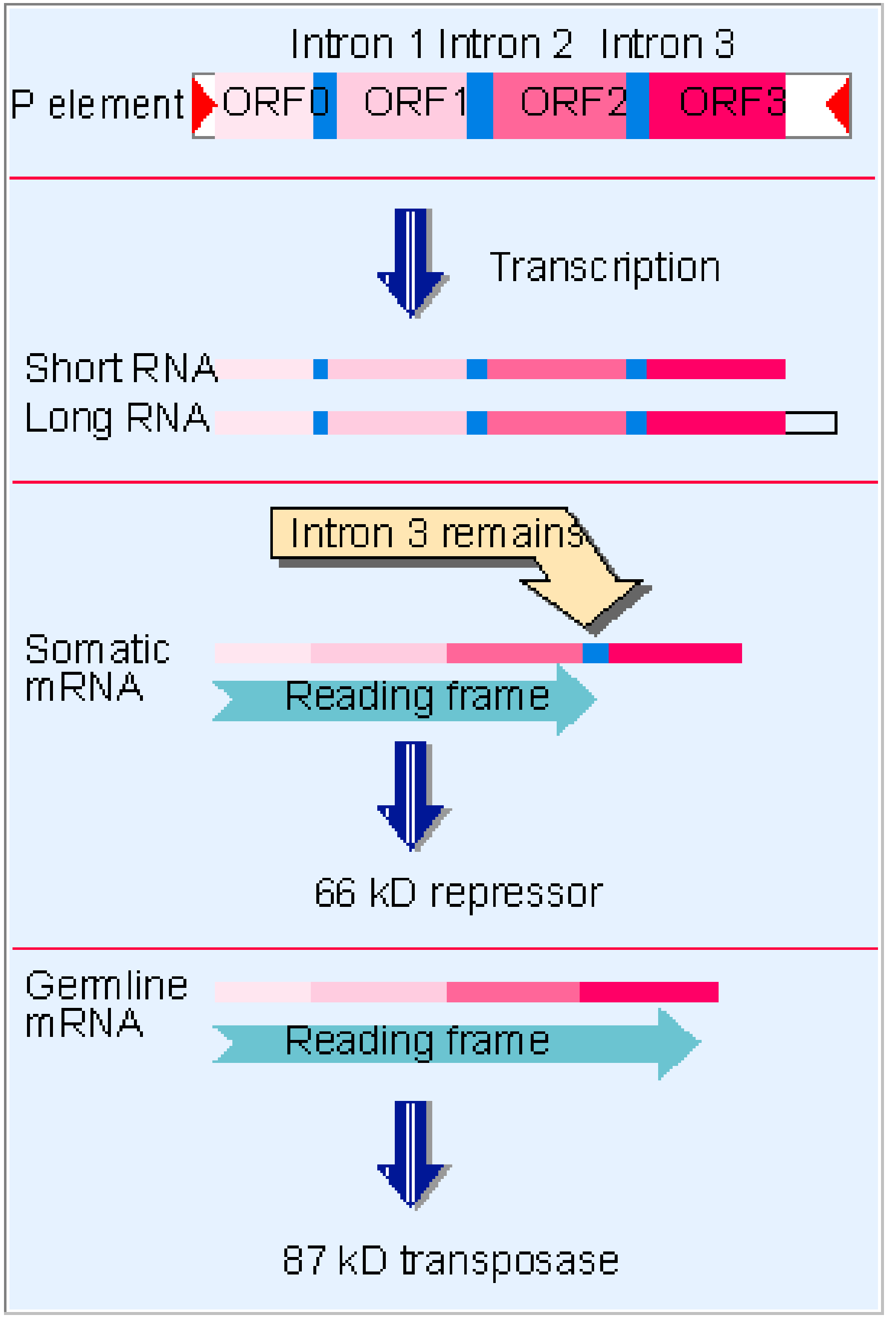
snapdragons

2. P element transposons in *Drosophilia* 果蝇的P因子

Hybrid dysgenesis: a condition causing sterility, elevated mutation rate, and a chromosome rearrangement in the offspring of crosses between certain strains of fruit flies.

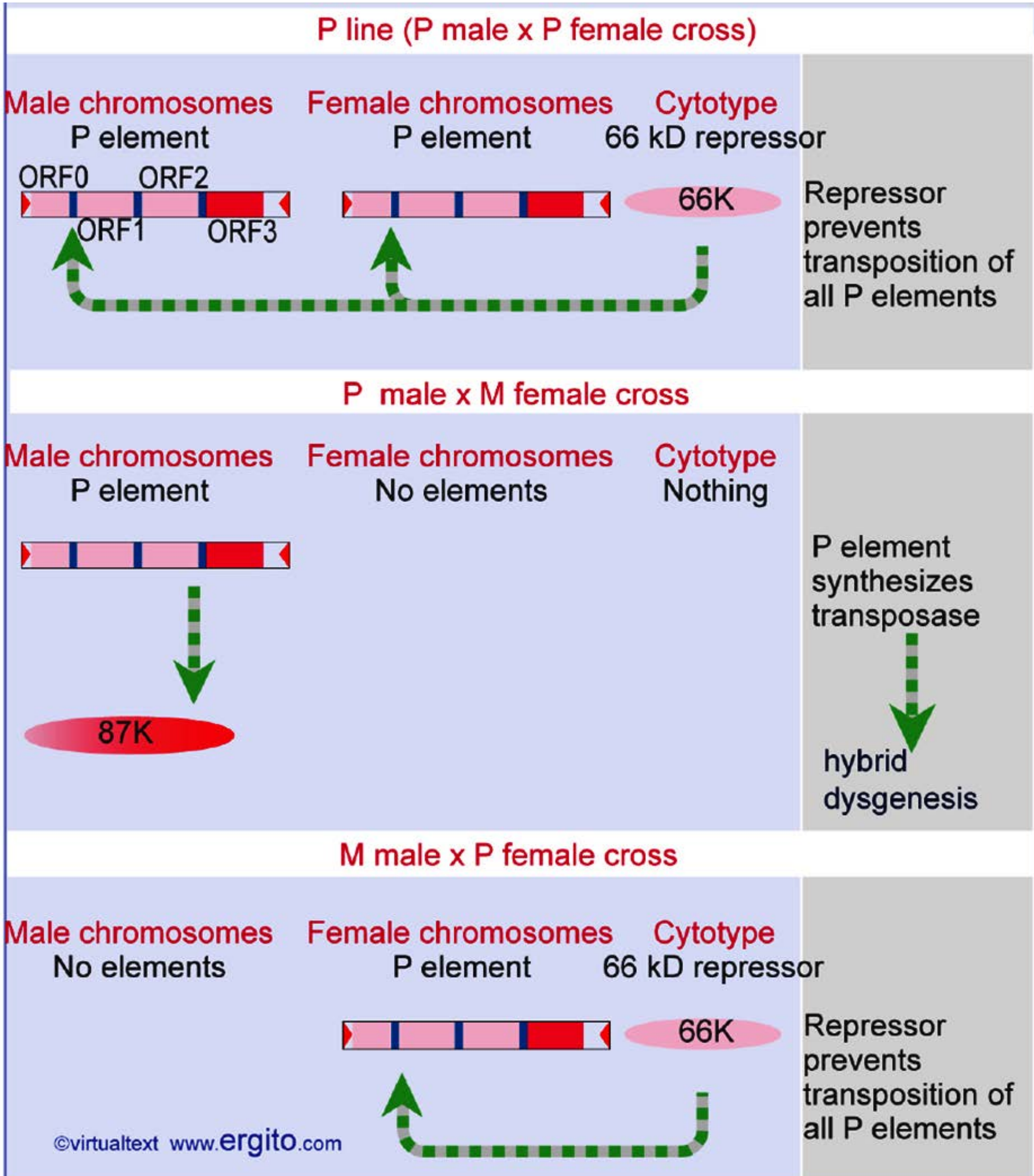


Hybrid dysgenesis in Drosophila
Dysgenic cross
Female (lab stock)
×
Male (wild)
M cytotype
P cytotype
F1
Defective hybrids (dysgenic progeny)
No offspring
Reciprocal cross
Female (wild)
×
Male (lab stock)
P cytotype
M cytotype
F1
Normal flies

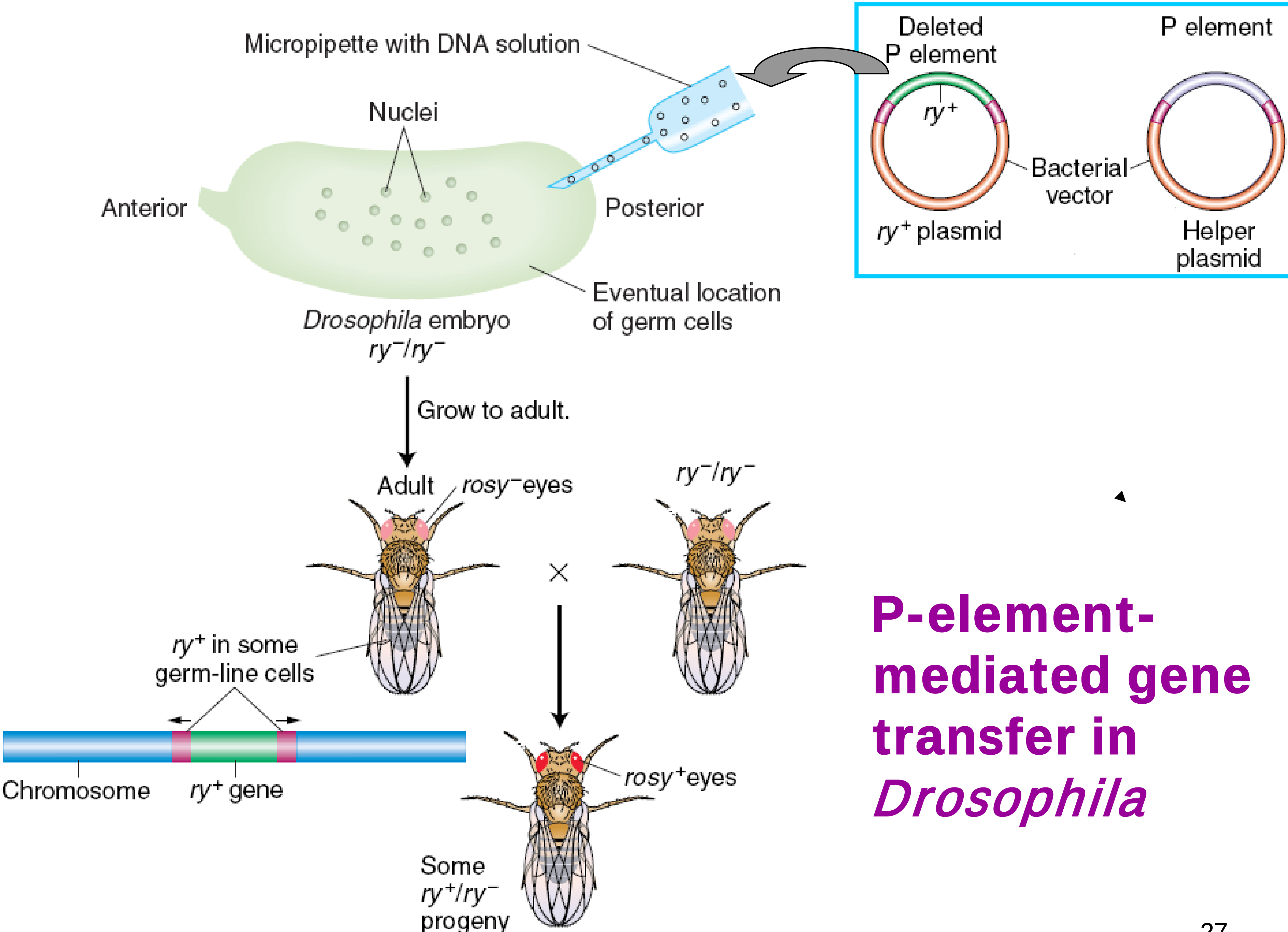


Somatic cells contain a protein that binds to sequences in exon 3 to prevent splicing of the last intron

The absence of this protein in germline cells allows splicing to generate the mRNA that codes for the transposase



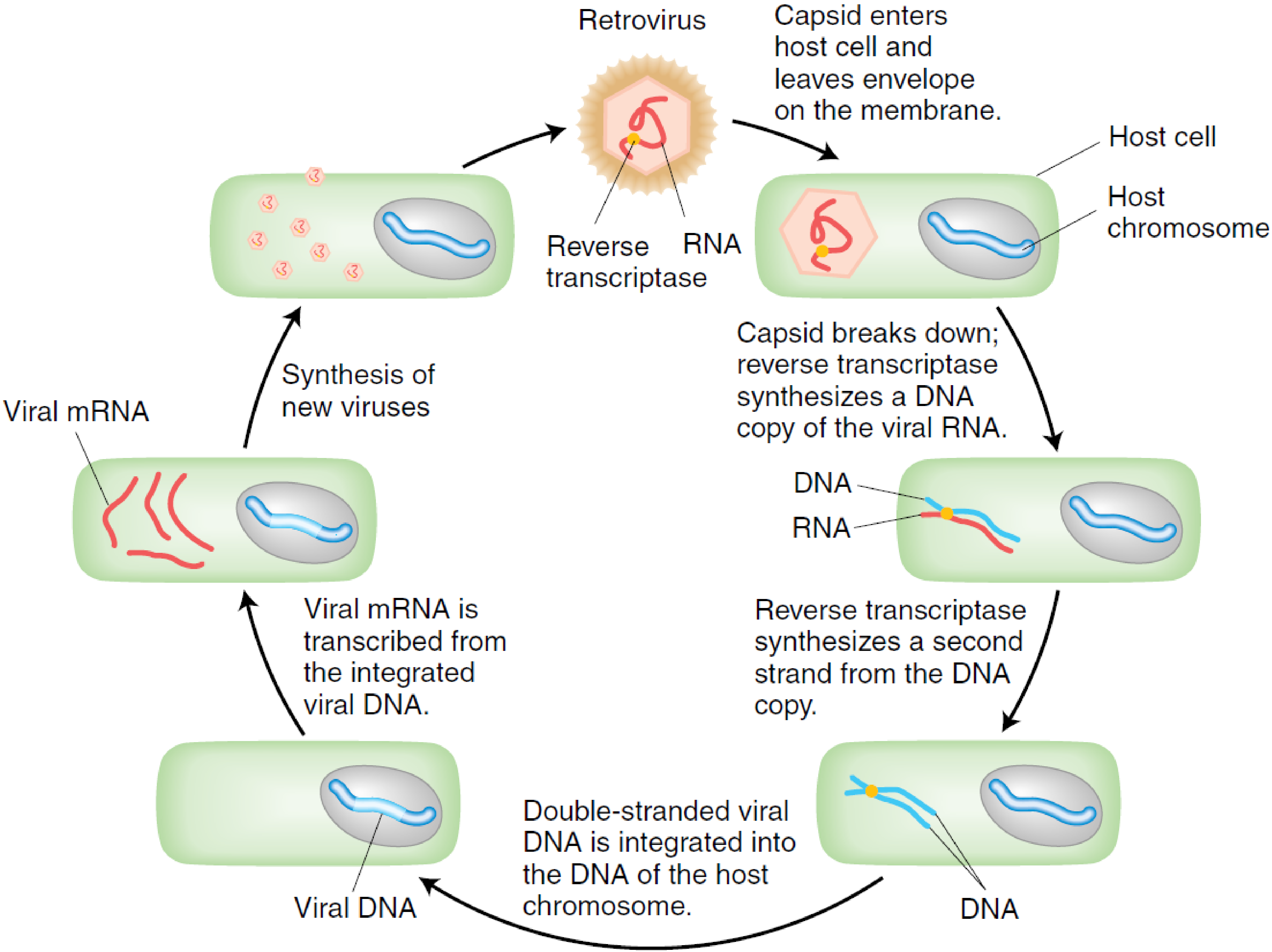
The 66KD repressor protein is provided as a maternal factor in the egg.



P-element-mediated gene transfer in *Drosophila*

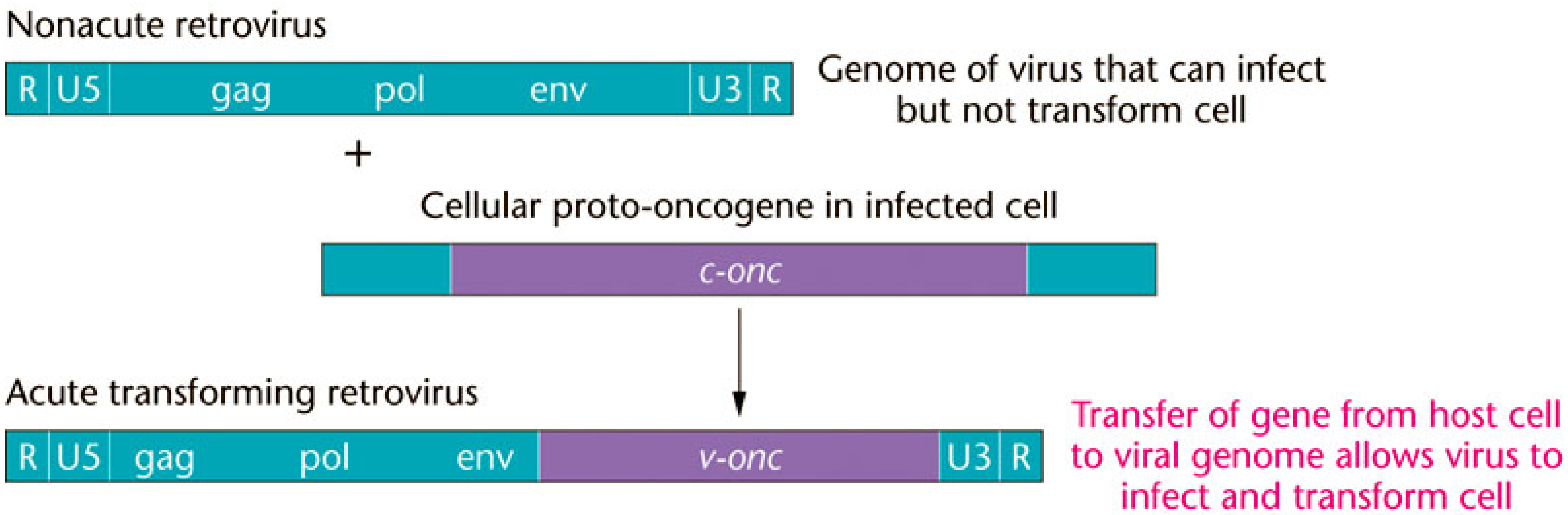
3. Retroviruses and Retroposons
逆转录病毒与逆转座子

Retroviruses are single-stranded RNA animal viruses that employ a double-stranded DNA intermediate for replication.



※The life cycle of a retrovirus※

- The viral DNA integrates into the host genome at randomly selected sites.
- Sometimes (probably rather rarely), the integrated retrovirus can convert a host cell into a tumorigenic state through activating certain types of host genes.

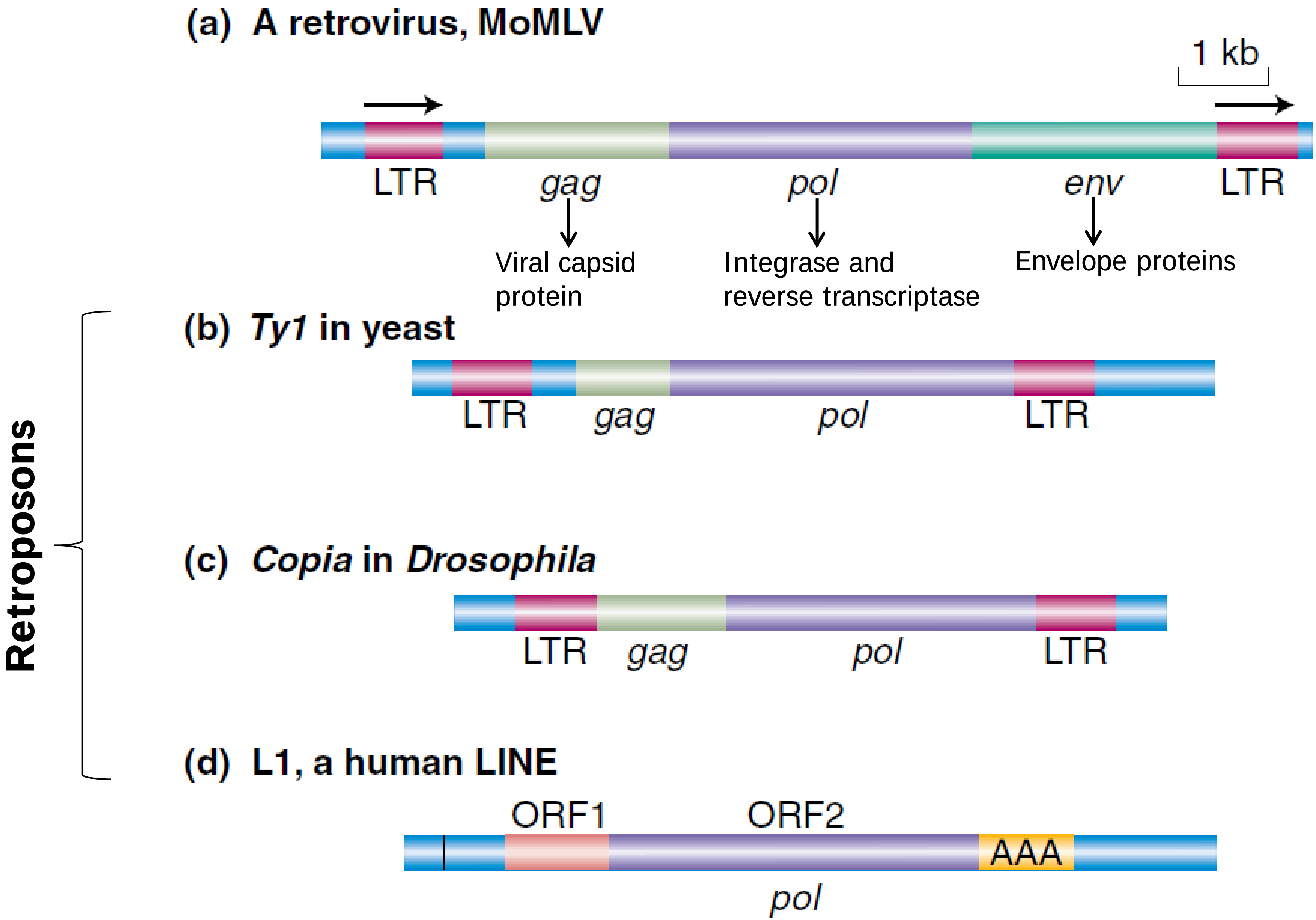


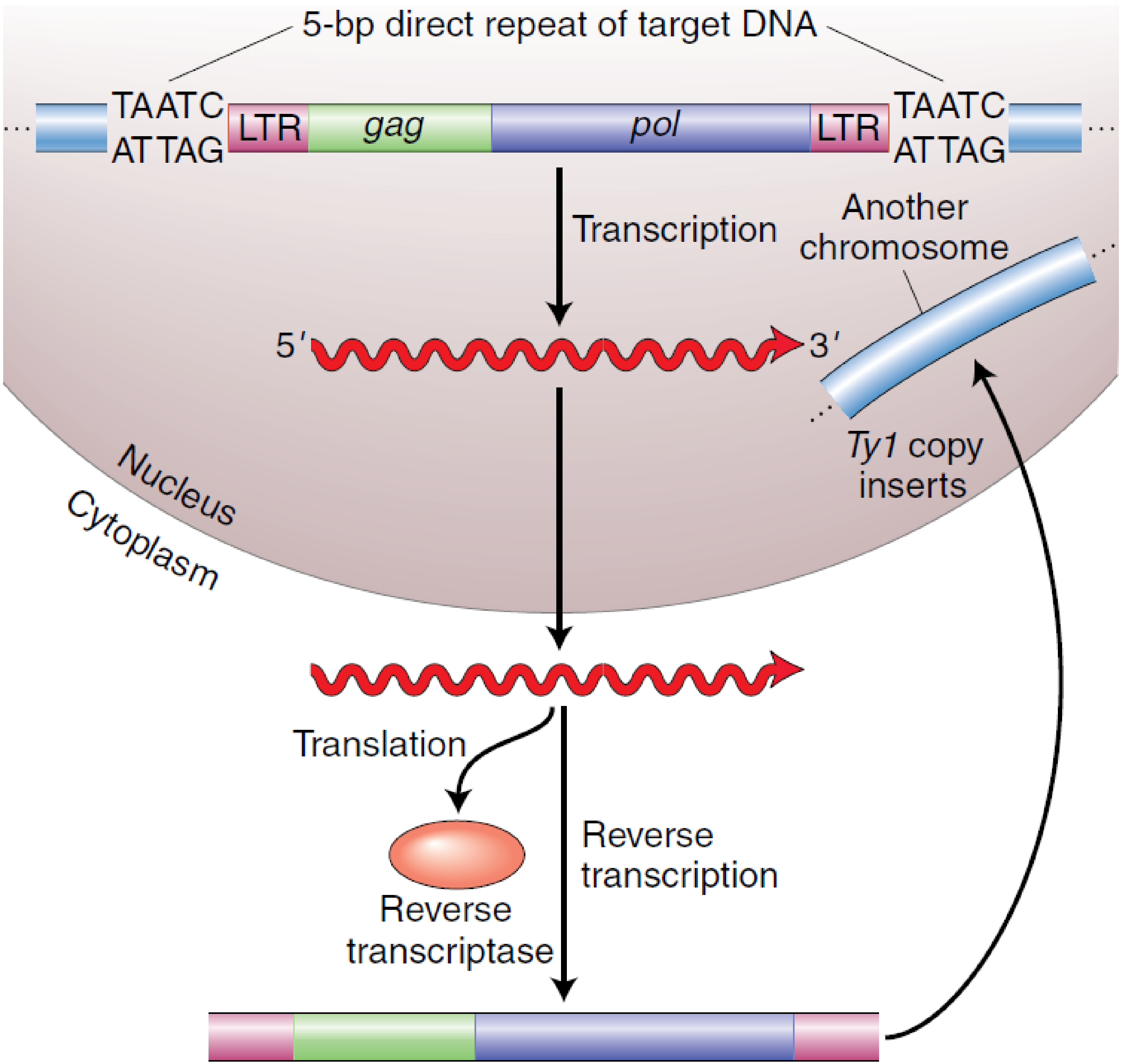
Retroposons

Transposable elements that utilize reverse transcriptase to transpose through an RNA intermediate are termed **retrotransposons.**

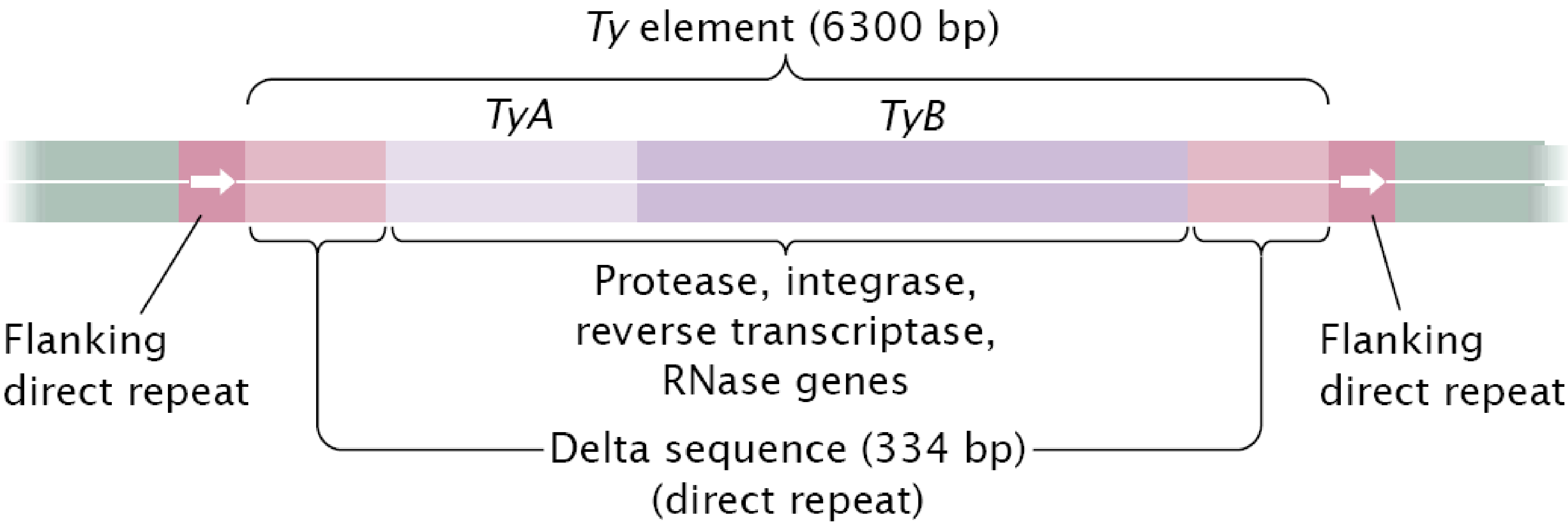
（逆转录转座子是指通过**RNA**为中介,反转录成**DNA**后进行转座的转座因子。逆转座作用出现在真核生物）

Structural comparison of a retrovirus to retrotransposons found in eukaryotic genomes



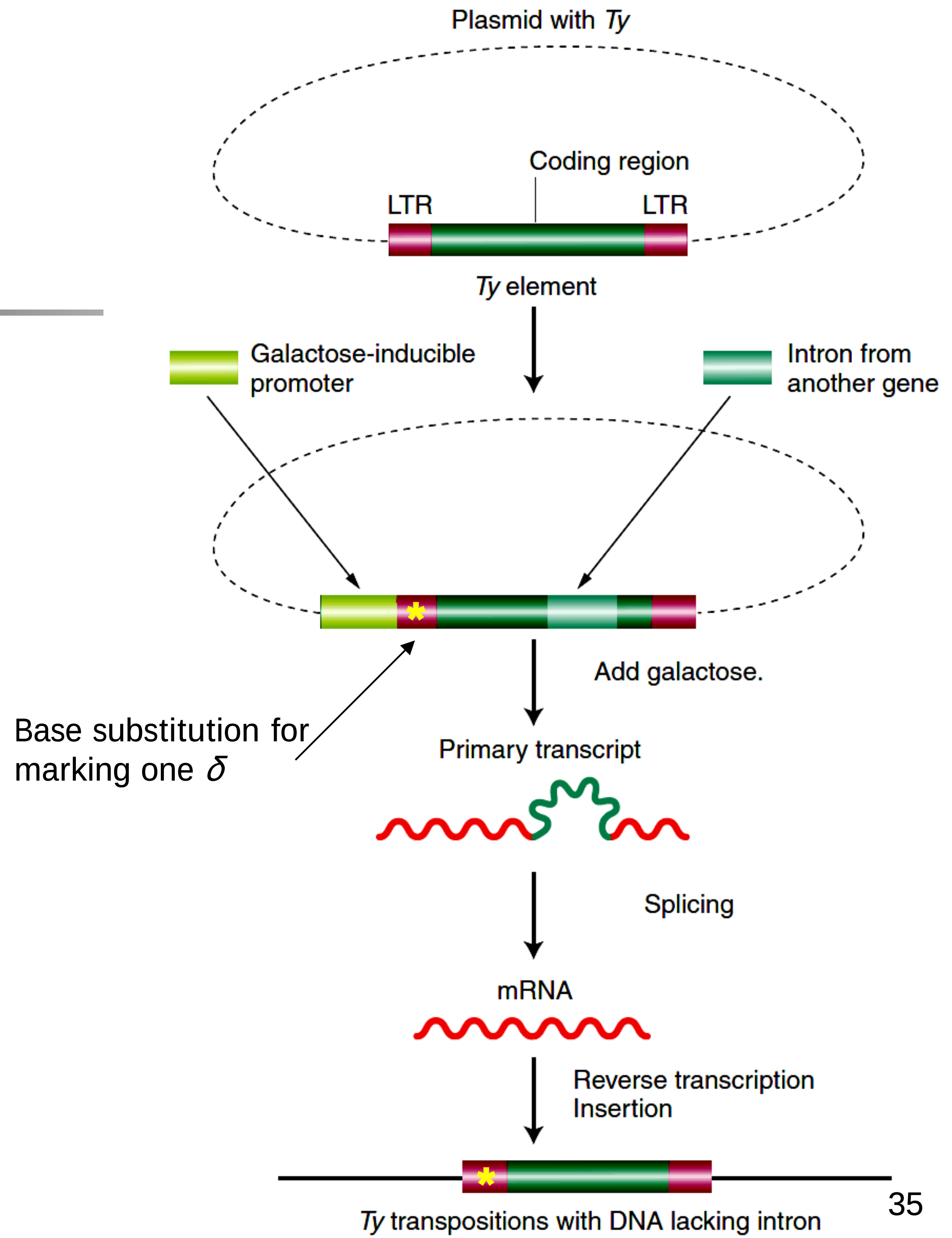

5-bp direct repeat of target DNA
TAATC
ATTAG
LTR
gag
pol
LTR
TAATC
ATTAG
Transcription
Another chromosome
5′
3′
Ty1 copy inserts
Nucleus
Cytoplasm
Translation
Reverse transcriptase
Reverse transcription

Yeast *Ty* elements

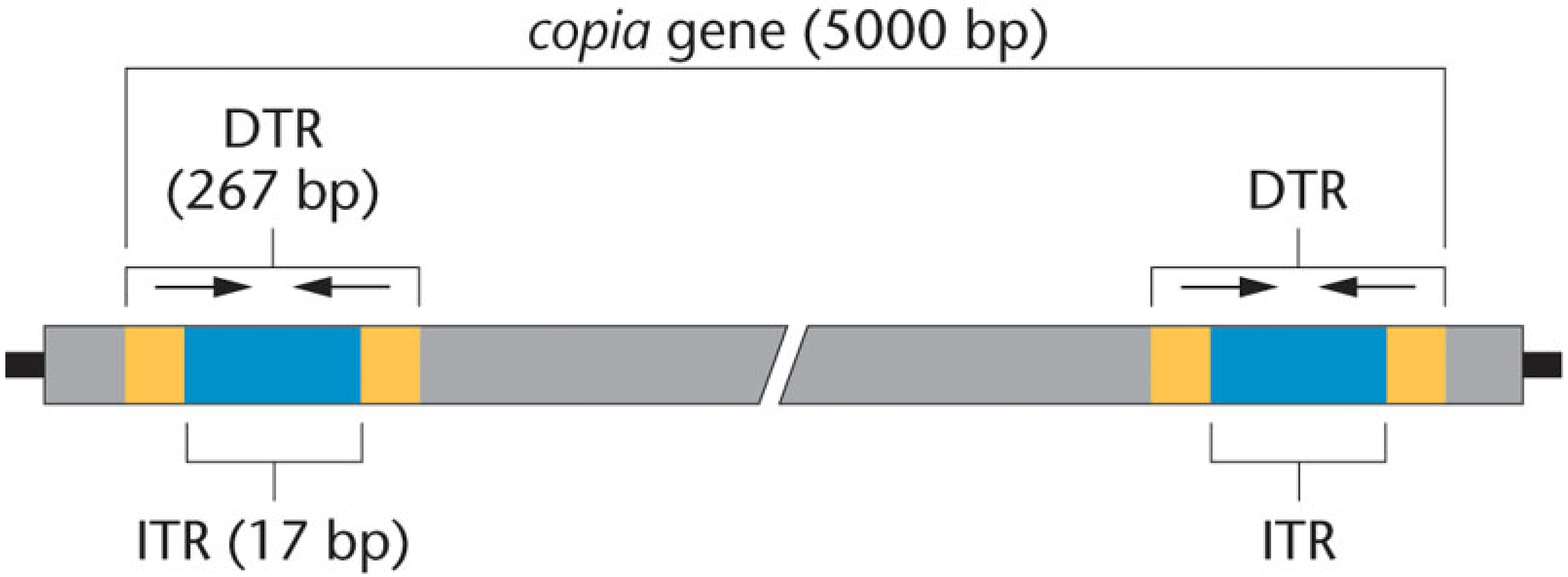


- About 35 Ty1 copies in the yeast genome.
- *δ sequences* : 330 bp, about 100 copies.
- solo δ element

Transposition through an RNA intermediate



- *Drosophila* ***copia*** elements



第三节 转座的遗传学效应

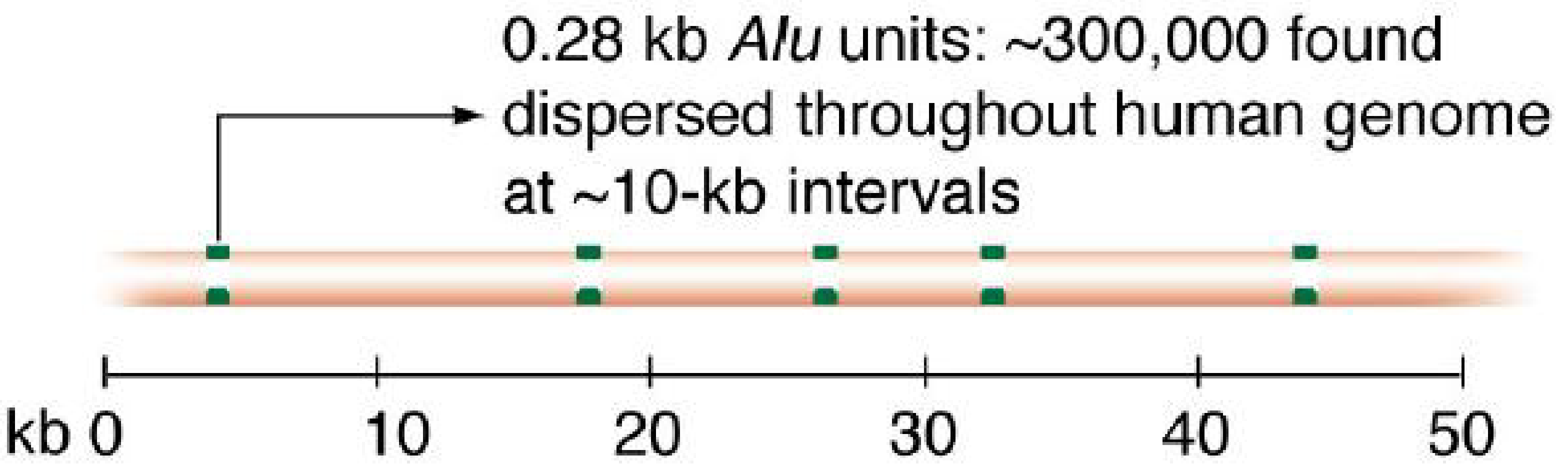
Genetic Effects of Transposition

1. Transposable elements shape the genomes of many organisms

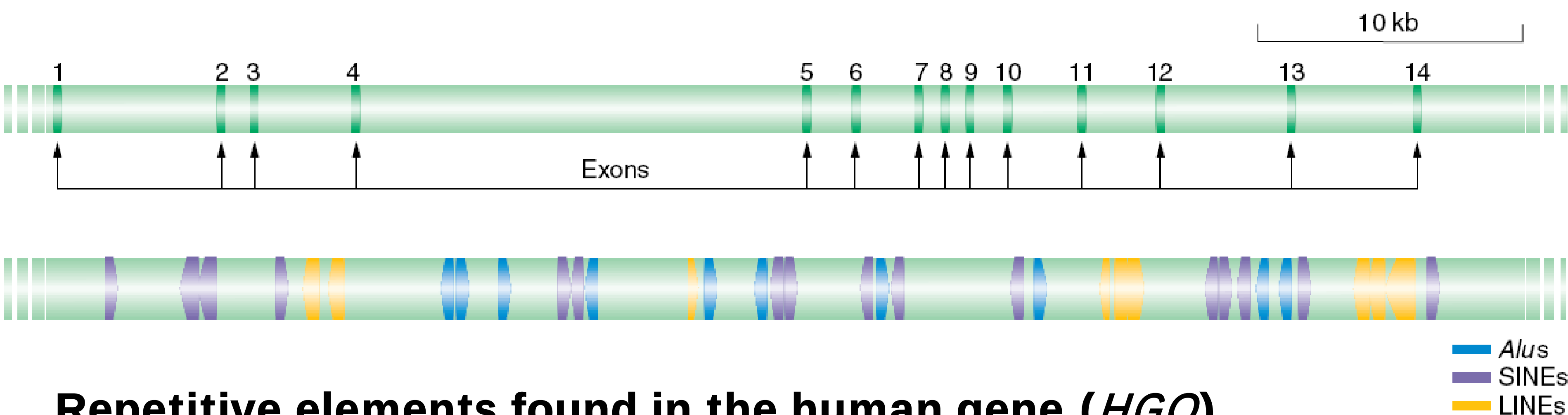
Transposable Elements in the Human Genome

Element	Structure	Length (kb)	Number	Genome fraction
Retrotransposons				
LINEs	*ORF1* · *pol* · AAA	6-8	1,000,000	20%
SINEs	AAA	<0.3	2,000,000	13%
HERVs	LTR · *gag* · *pol* · *env* · LTR	1-11	600,000	8%
DNA transposons				
	→ *transposase* ← → ←	2-3	400,000	3%
		Total	4,000,000	44%

LINEs (long interspersed elements) , **SINEs** (short interspersed elements), **HERV** (human endogenous retrovirus)

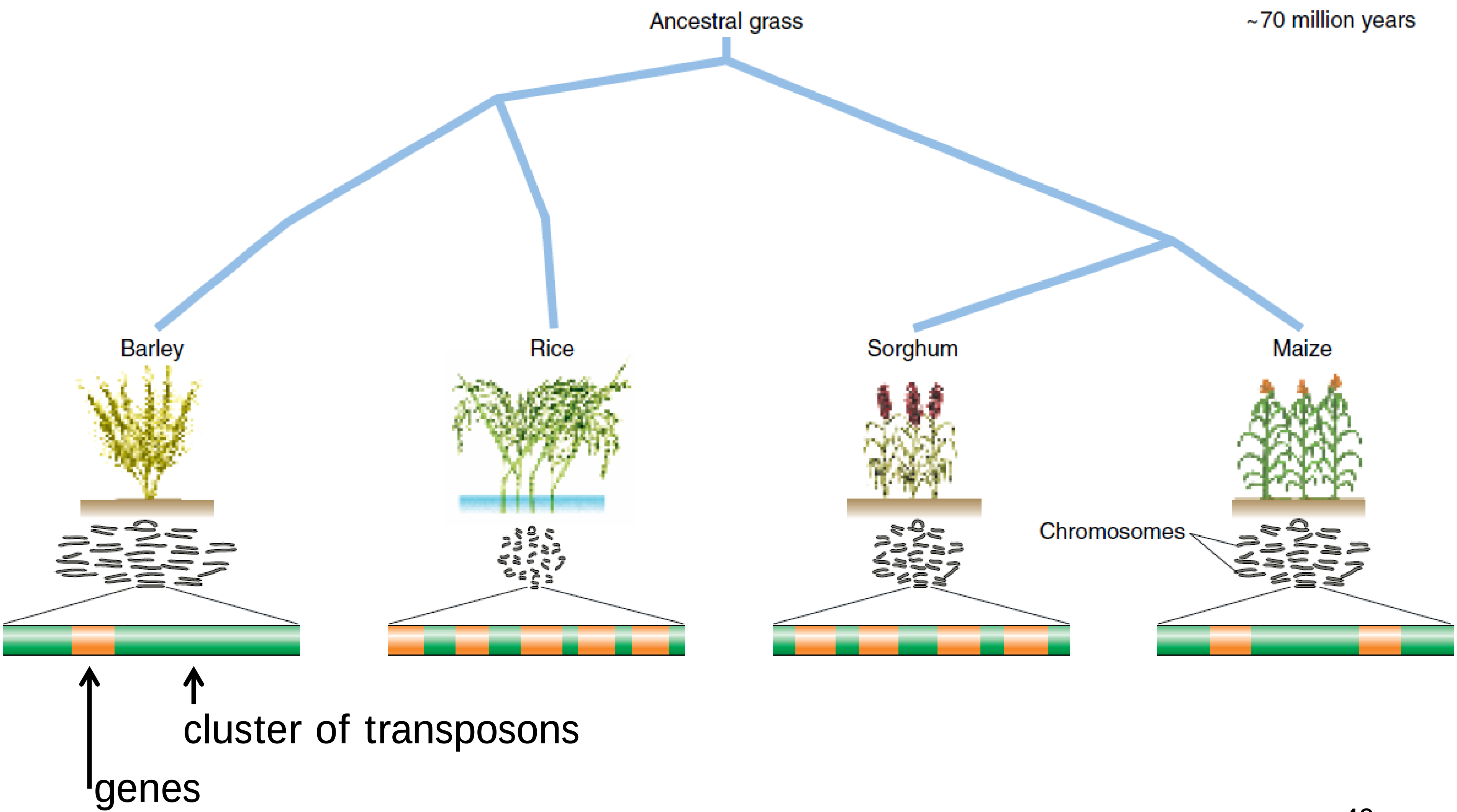


The *Alu* sequences make up more than 10 % of the human genome.



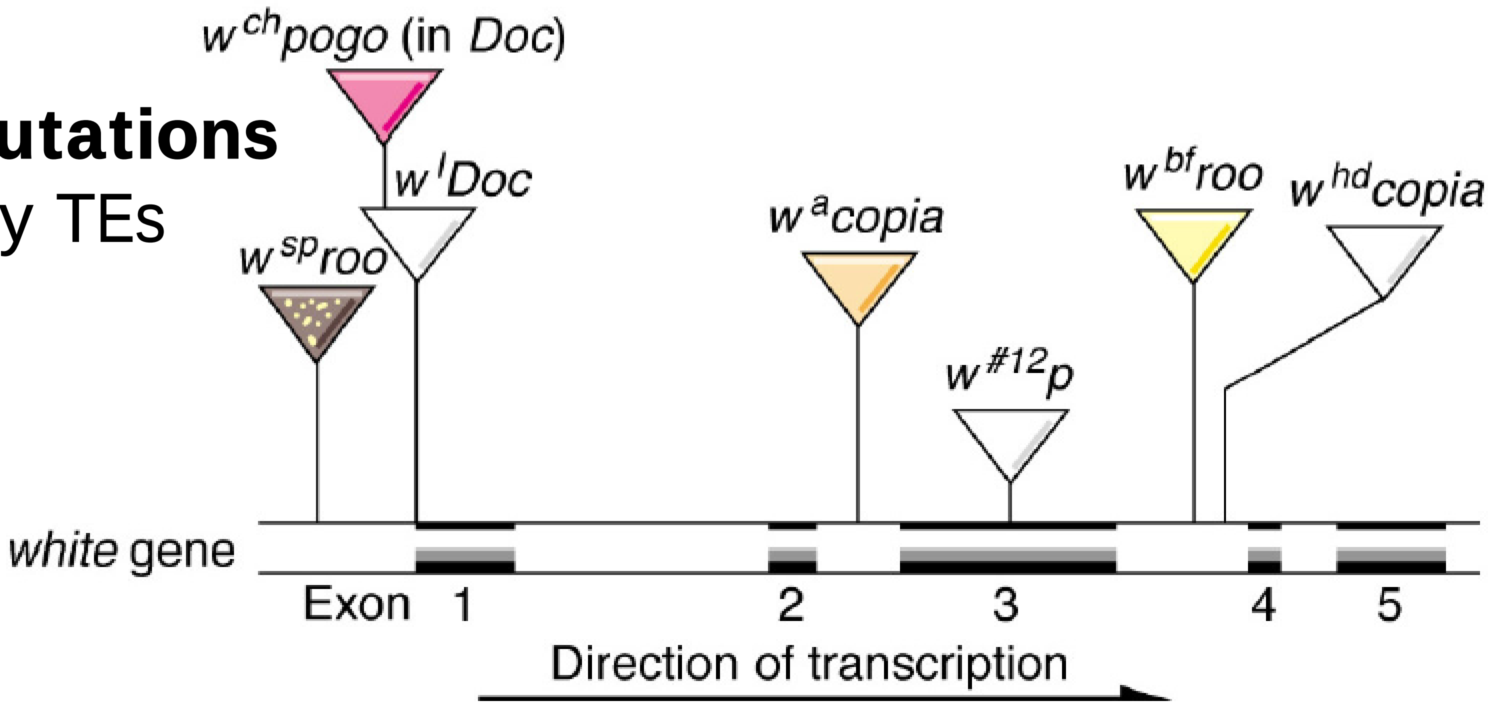
Repetitive elements found in the human gene (*HGO*) coding for homogentisate 1,2-dioxygenase, the enzyme whose deficiency causes alkaptonuria （黑尿病）

Transposable elements in grasses are responsible for genome size differences



2. Transposable elements can disrupt genes and alter genomes

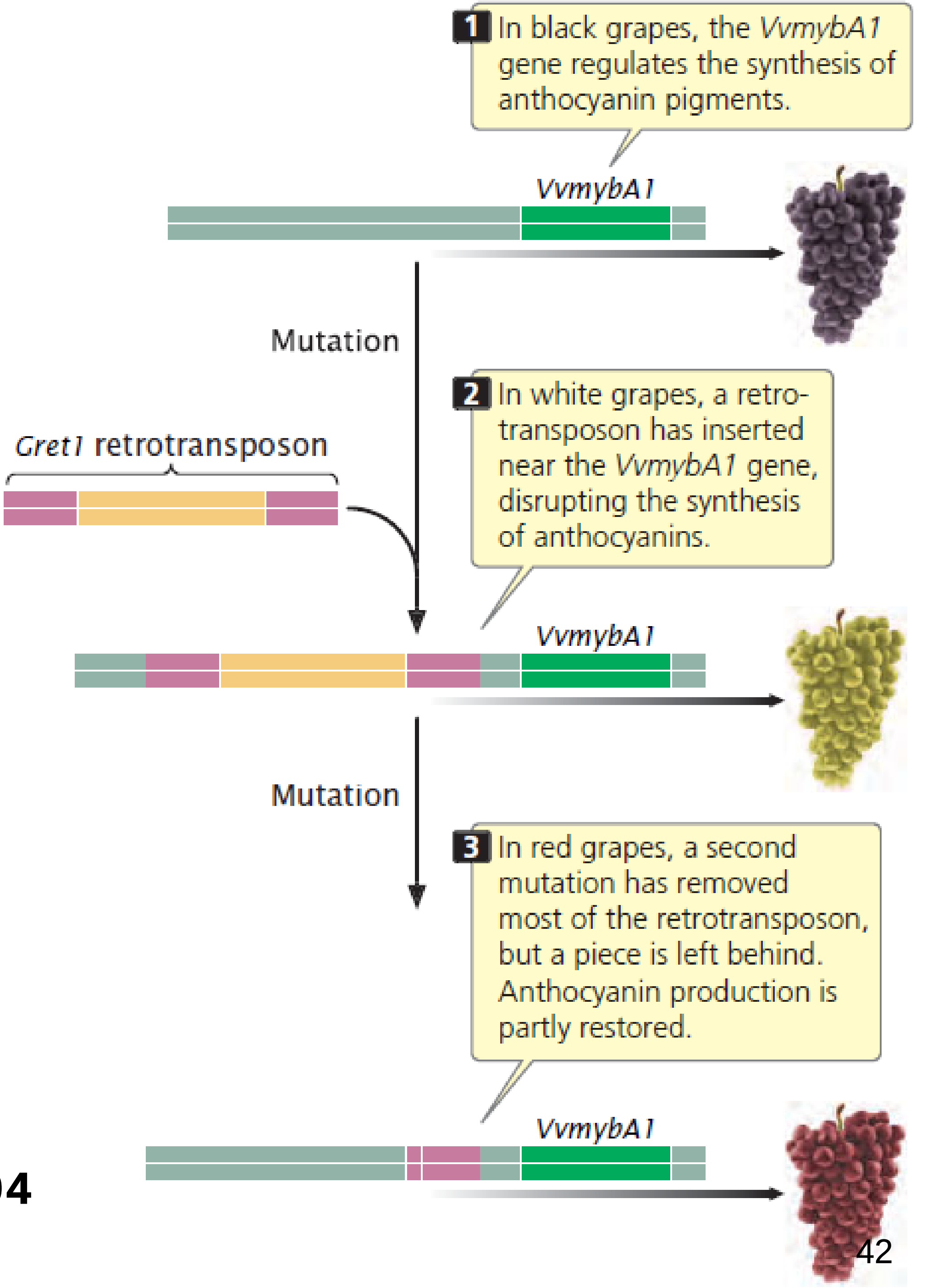
- **Gene mutations** caused by TEs



Many spontaneous mutations in the *white* gene of *Drosophila* arise from insertions of transposable elements such as *copia, roo, pogo,* or *Doc.*

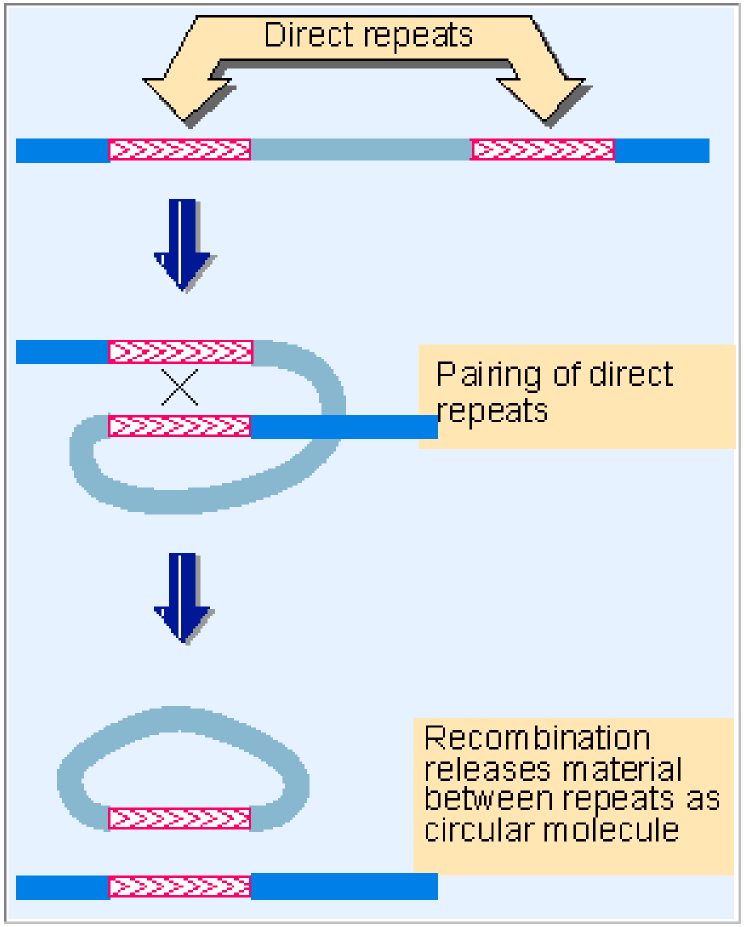
About 50% of all spontaneous mutations in ***Drosophila*** are due to transposition, 10%(mouse), 0.2%(human)

Grape skin color

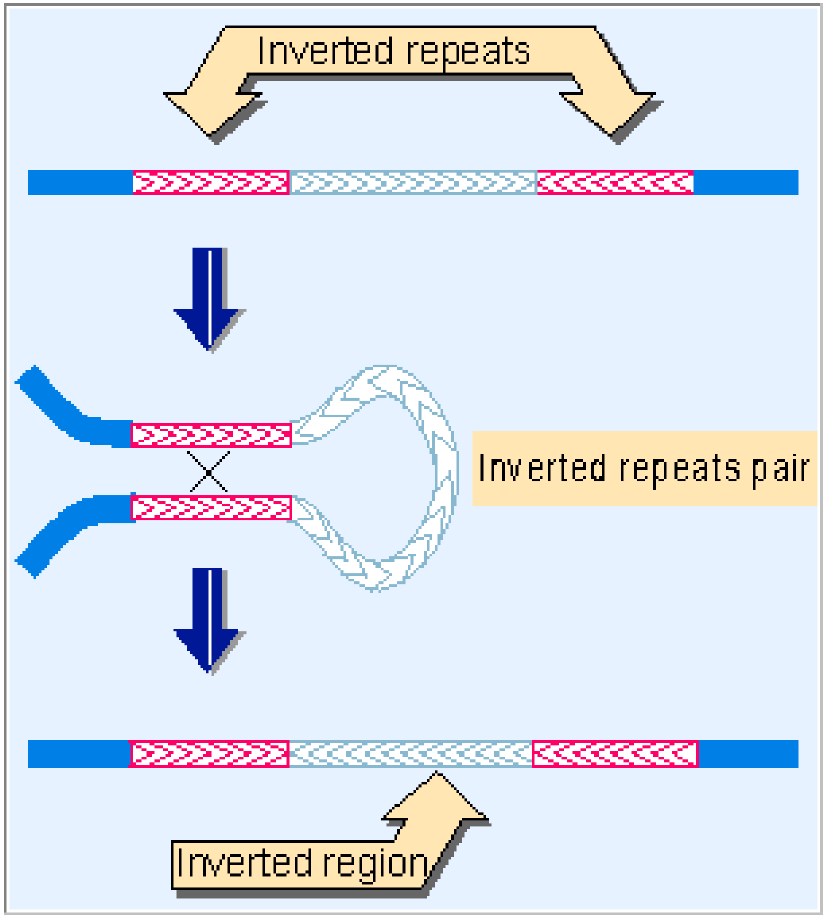


SCIENCE, 14 MAY 2004, VOL 304

Chromosomal rearrangements caused by TEs

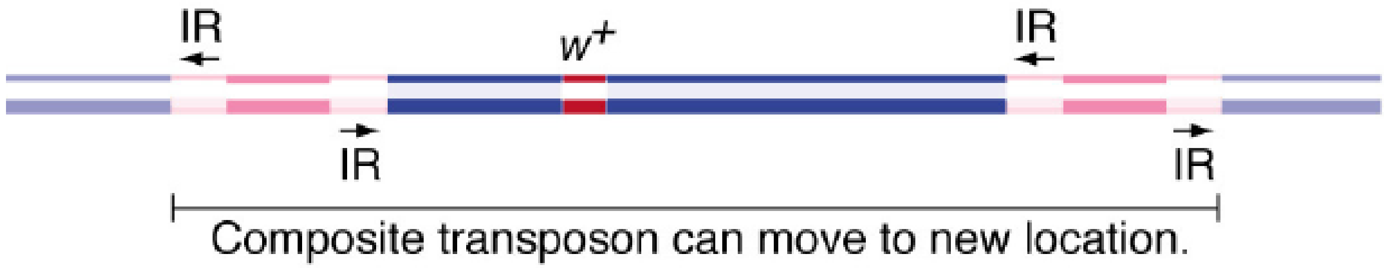


deletion



inversion

- **Gene relocation** due to transposition



Imprecisely excise of transposable elements:

- MULEs (*Mutator*-like transposable elements) and Pack-MULEs
- More than 3000 Pack-MULEs in rice genome, containing multiple fragments of genes from different chromosomal locations

3. The evolution of new genes through transposons

- **Exon shuffling**
  - Origin of "*jingwei*" gene
- **Domestication** of transposable elements (转座子驯化)
  - *A. thaliana* transcription factors
  - Human recombinase enzymes essential to the immune system

Further Discussion

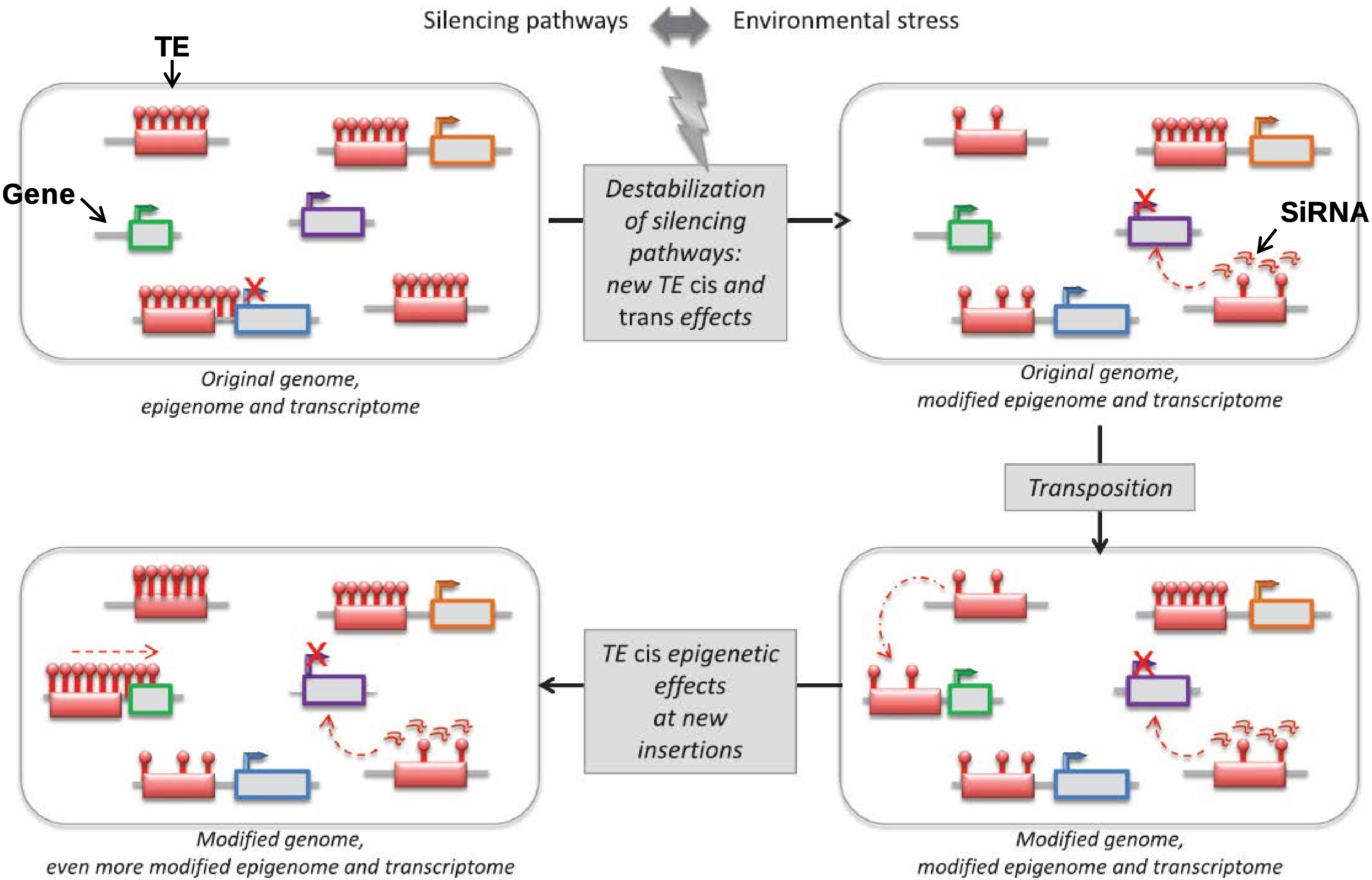
Transposable elements: **Genomic parasites? Selfish DNA?**

- The insertion of TEs into a gene will often destroy its function, with harmful consequences for the cell
- Replication and spread of TEs may serve no purpose for the cell
- The time and energy required to replicate large numbers of TEs are likely to place a metabolic burden on the cell

Or a treasure trove?

- The evolution of new genes
- Potentially regulating gene expression and rewiring gene networks
  - TEs maintain the length of *Drosophila* chromosomes
- TEs represent a rich and constantly changing pool of genetic and epigenetic variation on which selection can operate----TEs for adaptive evolution?

Model for TE-derived genomic and epigenomic modifications



How do plants and animals survive and thrive with so many insertions in genes and so much mobile DNA in the genome?

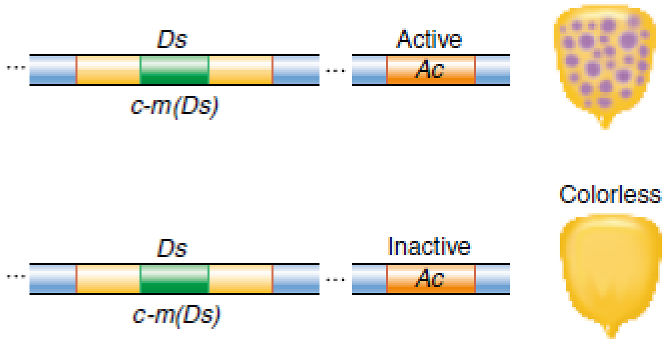
- Insertions into exons are negative selection. Successful transposable elements insert into so-called **safe havens** in the genome (retrotransposons, centric heterochromatin, introns)
- Most TEs in the genome are **inactive**, being the relics that have accumulated inactivating mutations over evolutionary time
- Others are still capable of movement but are rendered inactive by **host regulatory mechanisms**.

Transposable elements in humans

- **LINEs**
  - **L1**: 3000-5000 complete L1 (a small number are active), more than 500,000 truncated L1 (inactive)
  - **L2**: 315,000 copies, inactive
  - **L3**: 37,000 copies, inactive
- **SINEs** depend on LINEs for transposition
  - ***Alu*** (active), ***MIR*** and ***Ther2/MIR3***
- 400,000 retroviruslike elements, only a few are active
- DNA transposons

Regulation of Transposable Element Movement by the Host

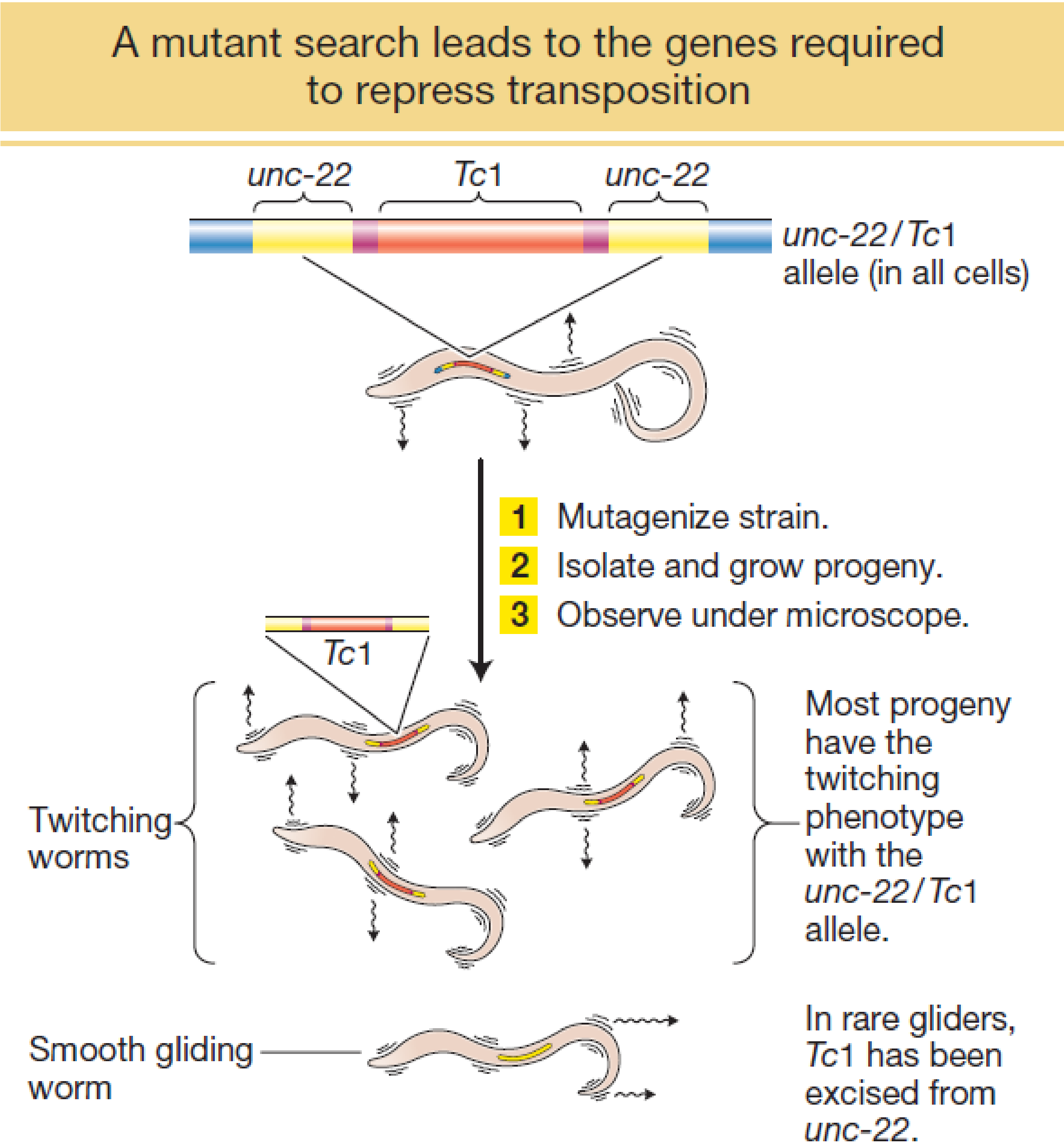
- Reversible changes in *Ac* activity



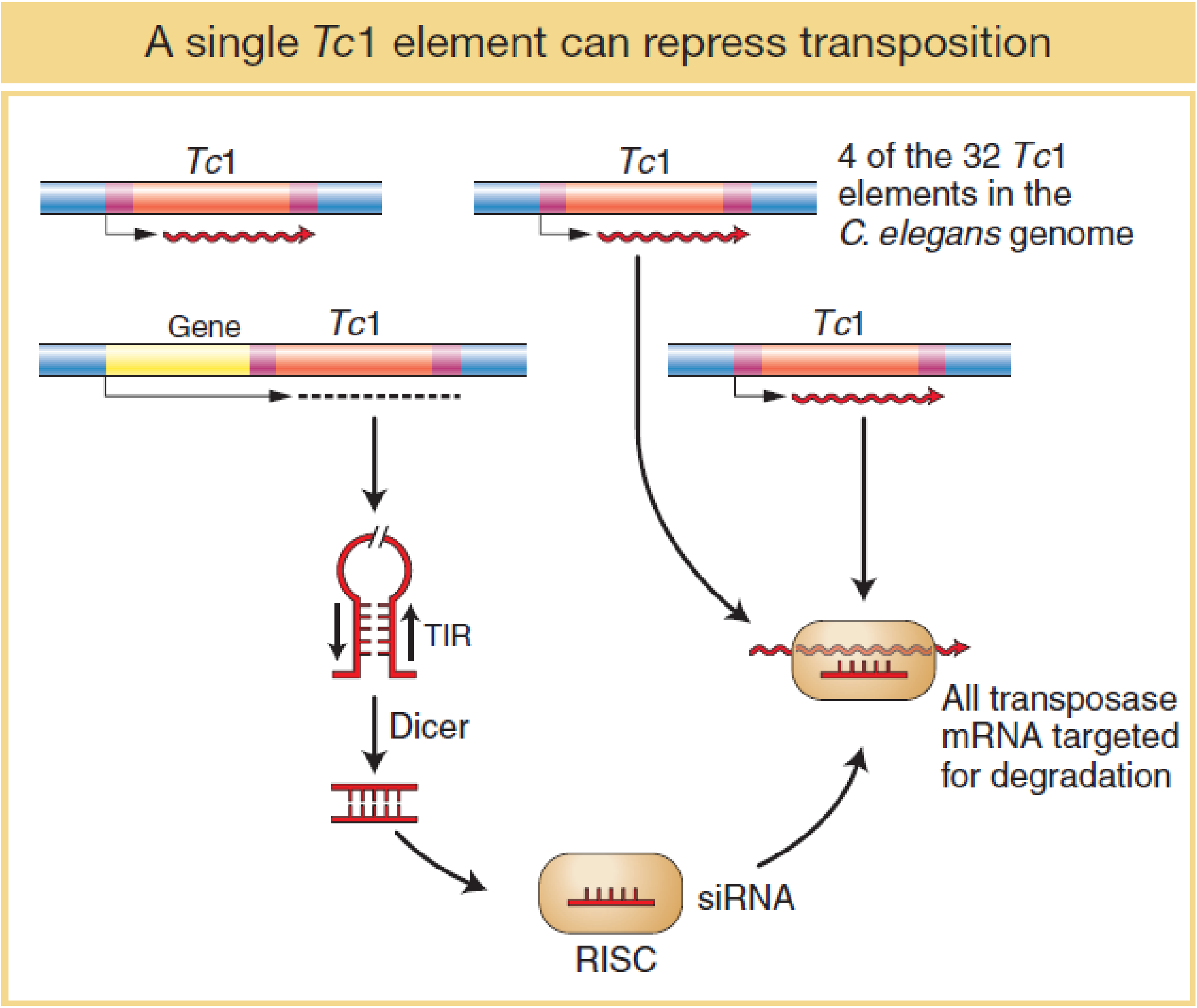
- Transgene silencing

Epigenetic regulation

Identify genes required to repress transposition



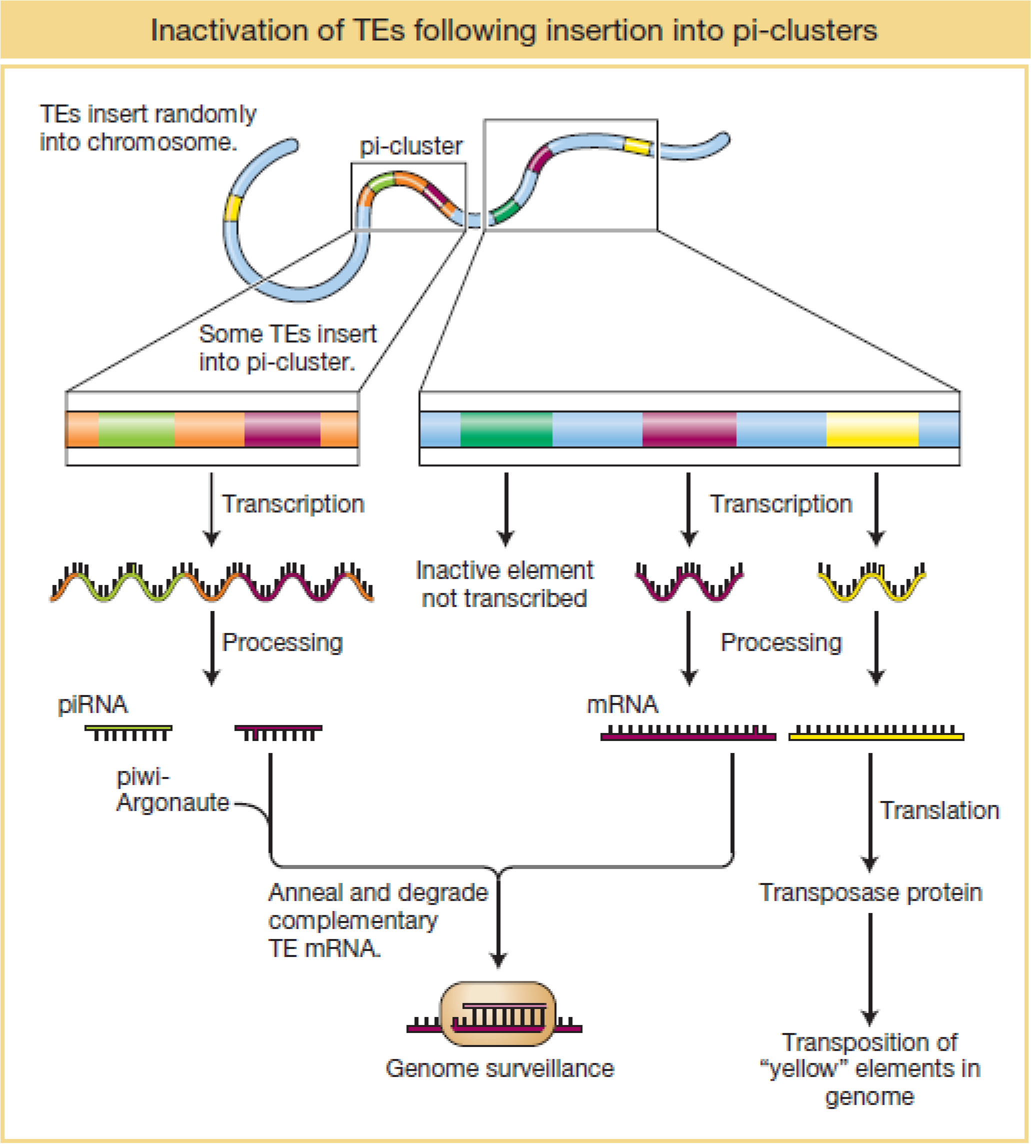
- 25 *C. elegans* genes were identified, whose mutations allowed the host to excise Tc1 in the germ line
- Many are integral components of the **RNAi** silencing pathway, including proteins found in Dicer and RISC



Eukaryotic hosts use RNAi to repress the expression of active transposable elements in their genomes. In this way, a single element that inserts near a gene can be transcribed to produce dsRNA that will trigger the silencing of all copies of the element in the genome.

Genome Surveillance

- **SiRNA**
- **piRNAs** in animals
  - Single-stranded RNAs, 26–30 nt in mammals
  - Originate from **pi-cluster**, comprised of remnants of many transposons
- **crRNAs** in bacteria
  - CRISPR loci is composed of invading virus sequence.



How do some transposons escaped from the surveillance of the host?

A genomic battleground

A constant battle between the proliferation of transposable elements and host attempts to silence or otherwise inactivate them.

The Frequency of Transposition

- The frequency varies among different elements.
- The overall rate of transposition is **10^{-3} ~ 10^{-4}** per element per generation.
  - New germ-line transpositions are estimated to occur once in every 50 to 100 human births
- Insertions in individual targets occur at a level comparable with the spontaneous mutation rate, usually **10^{-5} ~ 10^{-7}** per generation.

Human diseases caused by TEs

- To date, at least 14 cases of human diseases are determined due to insertions of L1 element
  - *dystrophin* gene, *APC* and *c-myc* genes
- SINE insertions are also responsible for more than 30 cases of human disease
  - *BRCA2* gene, *factor IX* gene, *ChE*, *NF1* gene, etc.
- A transposon "caught in the act”
  - A male child with hemophilia, blood-clotting factor VIII on X chromosome

The figures and tables are cited from:

- **Genetics (From genes to genomes)**, Leland Hartwell, Mcgraw-Hill Companies, Inc
- **Concept of Genetics**, William S.Klug, Prentice Hall, Inc
- **Introduction to Genetics Analysis**, Anthony J.F. Griffiths, W.H.Freeman, Inc
- **Principle of Genetics**, D.Peter Snustad, John Wiley & Sons, Inc
- **Genetics-A Conceptual Approach**, Benjamin A. Pierce, W. H. Freeman